Volume 247, number 1, 154–158 FEB 07007 April 1989

The complete islet amyloid polypeptide precursor is encoded by two exons

S. Mosselman*+, J.W.M. Höppener*+°, C.J.M. Lips° and H.S. Jansz*+

*Institute of Molecular Biology and Medical Biotechnology, Padualaan 8, 3584 CH Utrecht, +Laboratory for Physiological Chemistry, Vondellaan 24a, 3521 GG Utrecht and °Department of Internal Medicine, University Hospital, University of Utrecht, Utrecht, The Netherlands

Received 23 January 1989

Islet amyloid polypeptide (IAPP) is the 37-amino acid peptide subunit of amyloid found in pancreatic islets of type 2 diabetic patients and in insulinomas. Recently, we isolated the human gene encoding IAPP [(1988) FEBS Lett. 239, 227–232]. We now report the nucleotide sequences of a human insulinoma cDNA encoding a complete IAPP precursor, and of the corresponding parts of the IAPP gene. Two exons, which are approx. 5 kb apart in the human genome, encode the 89-amino acid pre-pro-IAPP. At least one additional exon is present further upstream in the IAPP gene. A putative signal sequence at the amino-terminus of the precursor suggests that IAPP is a secreted protein.

Islet amyloid polypeptide precursor; Genomic organization; Insulinoma; Diabetes mellitus type 2; Calcitonin gene family

1. INTRODUCTION

Localized amyloid formation is often associated with aging or with age-related diseases e.g. Alzheimer's disease and type 2 diabetes mellitus. In type 2 or non-insulin-dependent diabetes mellitus (NIDDM), amyloid is commonly found in pancreatic islets [1]. Islet amyloid polypeptide (IAPP) is a 37-amino acid polypeptide which was shown to be a major constituent of both islet and insulinoma amyloid [2–5]. Recently, we isolated the human IAPP gene [6]. The DNA sequence predicted an IAPP molecule identical to the peptides isolated from islet and insulinoma amyloid. Mature IAPP is processed from a larger precursor peptide and is probably amidated at its carboxy-terminus [6]. Furthermore, the resemblance of IAPP to human CGRP-II (calcitonin gene-related peptide-II) [3,5,7] and the demonstration of IAPP immunoreactivity in normal pancreatic β-cells [8,9], suggest that IAPP is a pancreatic islet polypeptide hormone.

IAPP is thought to play a role in NIDDM, since it inhibits both basal and insulin-stimulated rates of glycogen synthesis in rat skeletal muscle in vitro [10,11]. Glycolysis in the same tissue and glucose metabolism in adipocytes are unaffected [10,11]. Altered expression of the IAPP gene might thus contribute to the insulin resistance which is characteristic for NIDDM.

Northern blot analysis of human insulinoma RNA revealed the presence of two IAPP-specific polyadenylated RNAs of 1.6 and 2.1 kb, respectively [6]. From a cDNA library prepared from this tissue, two IAPP cDNAs were isolated. Corresponding genomic sequences show that the IAPP gene contains at least three exons, two of which encode the complete IAPP precursor. The evolutionary relationship between the IAPP gene and the calcitonin/CGRP (CALC) genes is discussed.

Correspondence address: S. Mosselman, Institute of Molecular Biology and Medical Biotechnology, University of Utrecht, Padualaan 8, 3584 CH Utrecht, The Netherlands

The nucleotide sequence presented here has been submitted to the EMBL/GenBank database under accession no. Y07495

Published by Elsevier Science Publishers B.V. (Biomedical Division)
00145793/89/$3.50

2. MATERIALS AND METHODS

2.1. *RNA isolation, cDNA synthesis and cloning*

RNA was isolated from a human insulinoma as described [12,13]. 3 μg poly A^+ RNA was used for the synthesis of cDNA according to the RNase H method (cDNA Synthesis System Plus, Amersham RPN.1256). Double-stranded cDNA was cloned in the λgt10 vector using *Eco*RI-linkers and in vitro packaged using the cDNA Cloning System λgt10 (Amersham RPN.1257B). A total of 1.2×10^6 different clones was obtained.

2.2. *Hybridizations and DNA probes*

Replica filters (Hybond-N, Amersham) were prepared as described by the supplier (density: 60000 clones/135 mm dish). A *Pst*I-*Bam*HI probe containing a genomic fragment encoding part of IAPP (4 nucleotides of the M13 polylinker and nucleotides 8-171 in fig.2 [6]) was labeled by random priming [14] using [α-^{32}P] dCTP. Hybridizations were performed at 42°C in the presence of 50% formamide, 1 M NaCl, 10% dextran-sulphate, 0.2% bovine serum albumin, 0.2% polyvinyl-pyrrolidone, 0.2% Ficoll, 50 mM Tris-HCl (pH 7.5), 0.1% sodium-pyrophosphate, 0.1% SDS and 100 μg/ml denatured herring sperm DNA. Filters were washed at 65°C in 2 × SSC (1 × SSC = 150 mM NaCl, 15 mM sodium citrate)/0.5% SDS and then exposed to Fuji RX films using intensifying screens.

2.3. *Subcloning and nucleotide sequence analysis*

λ DNA of purified plaques was isolated from plates according to [15]. The inserts of clones λhIAPP-c1 and -c2 and appropriate restriction fragments of the genomic clone λh201 [6] were subcloned in pEMBL 8. The nucleotide sequence analysis was performed according to the chemical modification technique [16] after labeling DNA fragments at their 5′- or 3′-termini.

3. RESULTS AND DISCUSSION

Hybridizations with the IAPP-specific *Pst*I-*Bam*HI probe [6] of Northern blots containing RNA from a variety of human tissues including placenta, liver, medullary thyroid carcinoma (MTC), glucagonoma and several smooth muscle tumours, were negative. IAPP-specific polyadenylated RNAs (1.6 and 2.1 kb) were detected only in a human insulinoma [6]. The insulinoma cDNA library contained one clone, λhIAPP-c1, which hybridized with the same probe. The 586 bp insert of λhIAPP-c1 terminates with a poly A tract of 18 residues, preceded by an AATAAA sequence (see figs 1 and 2). The IAPP-encoding reading frame in λhIAPP-c1 reveals an ATG codon (nucleotides 99–101), which must represent the site of initiation of translation since an in-frame stopcodon is present upstream. IAPP thus appears to be derived from an 89-amino acid precursor with a calculated molecular mass of 9808 Da. The amino-terminal 22 residues fit all criteria for signal sequences [17]: (i) length between 15 and 30 amino acids, (ii) a hydrophobic core (residues 8–17) and (iii) a positively charged residue in the amino-terminal part (Lys 5). Assuming that cleavage of this putative signal sequence occurs between Ala 22 and Thr 23, a 67-amino acid pro-IAPP is generated (7404 Da). Proteolytic processing of the IAPP-prohormone yields mature IAPP and amino- and carboxy-terminal peptides of 9 and 16 residues, respectively. Both the presence of a putative signal sequence (which directs pro-IAPP into the endoplasmic reticulum) and the occurrence of IAPP in extracellular amyloid strongly suggest the secretion of IAPP by the cells that produce it.

Two exons of the IAPP gene were characterized, which together encode the complete IAPP precursor as predicted by the cDNA sequence (see figs 1 and 2). The pre-pro-IAPP-encoding regions of the genomic and of the cDNA sequences are identical. Exon 2 contains 15 nucleotides of 5′-untranslated sequences and encodes the signal sequence and part of the amino-terminal peptide. Exon 3 encodes the remainder of the IAPP precursor and contains the 3′-untranslated region [6]. Exon 2 and exon 3 are separated by approx. 5 kb in the human genome. The 83 nucleotides of λhIAPP-c1 preceding exon 2 are derived from at least one exon further upstream (see figs 1 and 2). The nucleotide sequence of λhIAPP-c1 shows that the reported potential polyadenylation signal and the predicted splice acceptor signal of exon 3 [6] are actually used. Cleavage and subsequent poly A addition occur at an A residue (nucleotides 569–571 in fig.2), 22–24 nucleotides downstream of the AATAAA sequence.

Rescreening of the insulinoma cDNA library using the insert of λhIAPP-c1 as probe revealed λhIAPP-c2 (see figs 1 and 2). It encodes the 18 carboxy-terminal residues of the IAPP precursor and contains approximately 650 nucleotides of 3′-untranslated sequences, including an Alu-repeat [18]. The λhIAPP-c2 insert terminates with a stretch of 32 A residues which are not preceded by a typical polyadenylation signal and therefore are probably the result of oligo(dT) priming on an A-rich region of an IAPP (pre) mRNA.

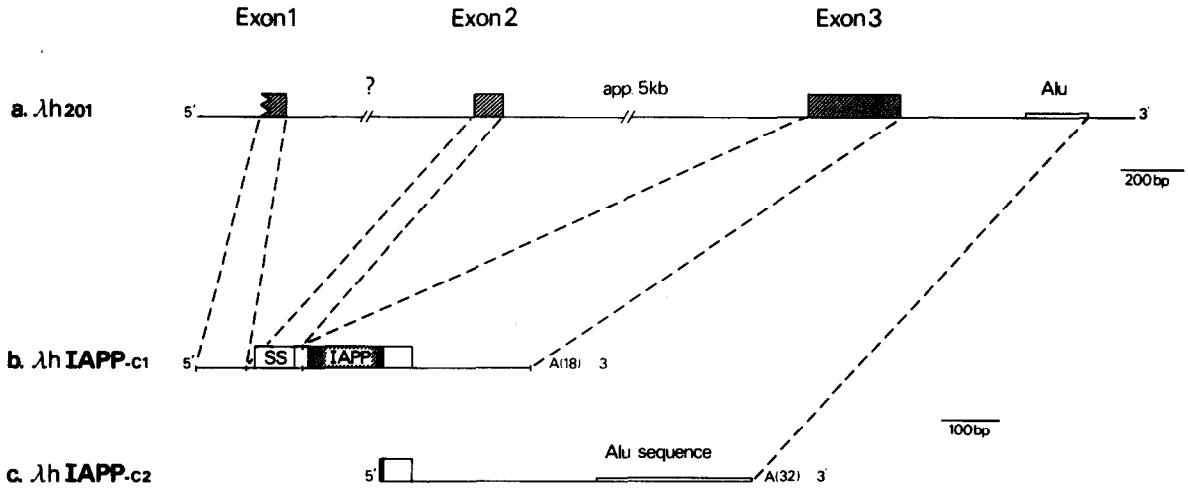

Fig.1. Schematic representation of the human insulinoma IAPP cDNAs λhIAPP-c1 (b) and λhIAPP-c2 (c) and of part of the human IAPP gene (a). The exons are designated exons 1, 2 and 3. The distance between exons 2 and 3 is approx. 5 kb, the distance between exons 1 and 2 and the 5′-end of exon 1 have not yet been determined. The translated regions of the cDNAs are boxed. Proteolytic processing sites are represented by solid boxes. The Alu sequence downstream of exon 3 and at the 3′-end of λhIAPP-c2 is also indicated.

```
                  20                  40                  60                  80
GGATACAAGCTTGGACTCTTTTCTTGAAGCTTTCTTTCTATCAGAAGCATTTGCTGATATTGCTGACATTGAAACATTAAAAG... - ? -  (120)tcttgatttcag

                  84                 100                 120                 140
tgctggattattctttgcagAAAATTTGAGAAGCA ATG GGC ATC CTG AAG CTG CAA GTA TTT CTC ATT GTG CTC TCT GTT GCA TTG AAC CAT
                                     Met Gly Ile Leu Lys Leu Gln Val Phe Leu Ile Val Leu Ser Val Ala Leu Asn His

   160                 178                                                                 180
CTG AAA GCT ACA CCC ATT GAA AG gttggtaacttaaaatcct(125)    - app.5kb -    (40)gttccatgttaccag T CAT CAG GTG GAA
Leu Lys Ala Thr Pro Ile Glu Se                                                             r His Gln Val Glu

          200                 220                 240                 260
AAG CGG AAA TGC AAC ACT GCC ACA TGT GCA ACG CAG CGC CTG GCA AAT TTT TTA GTT CAT TCC AGC AAC AAC TTT GGT GCC ATT
Lys Arg Lys Cys Asn Thr Ala Thr Cys Ala Thr Gln Arg Leu Ala Asn Phe Leu Val His Ser Ser Asn Asn Phe Gly Ala Ile

   280                 300                 320                 340
CTC TCA TCT ACC AAC GTG GGA TCC AAT ACA TAT GGC AAG AGG AAT GCA GTA GAG GTT TTA AAG AGA GAG CCA CTG AAT TAC TTG
Leu Ser Ser Thr Asn Val Gly Ser Asn Thr Tyr Gly Lys Arg Asn Ala Val Glu Val Leu Lys Arg Glu Pro Leu Asn Tyr Leu

         368
CCC CTT TAG AGGACAATGTAACTCTATAGTTATTGTTTTATGTTCTAGTGATTTCCTGTATAATTTAACAGTGCCCTTTTCATCTCCAGTGTGAATATATGGTC 463
Pro Leu End

TGTGTGTCTGATGTTTGTTGCTAGGACATATACCTTCTCAAAAGATTGTTTTATATGTAGTACTAACTAAGGTCCCATAATAAAAAGATAGTATCTTTTAAAATGAA 570

ATGTTTTTGCTATAGATTTGTATTTTAAAACATAAGAACGTCATTTTGGGACCTATATCTCAGTGGCACAGGTTTAAGAACGAAGGAGAAAAAAGGTAGTTTGAACCT 677

TGGTAAATTGTAAACAGCTAATAATGAAGTTATTCTTGACATGAGAAAATCAGTAATTGGACCAGGCGCGGTGGCTCTTGCCTGTAATCCCAGCACTTTGGGAGGCC 784

GAGGCAGGCAGATCACAAGGTCAGGAGTTCGAGACCAGCCTGACCAACATGGTGAAACCCTGTCTCTACTAAAAATACAAAAATTAGCCGGGGGTGGTGACATGTGC 891

CTGTAATCCCAGCTACTCAGGAGGCTAAGGCAGGAGAATCGCTTAAACCCAGGAGGCGGAGGTTGCAGTGAGCCGAGATTGCACCACTGCACTCCAGCCTGGGTGGC 998

AGAGTGAGACTCGTCTC 1015
```

Fig.2. Nucleotide sequences (in capitals) of exons 1, 2 and 3 of the human IAPP gene and deduced amino acid sequence of the IAPP precursor polypeptide. Intron sequences are indicated in small letters and numbers of additional nucleotides sequenced are indicated between parentheses. In λhIAPP-c1 nucleotides 1–568 are followed by 18 A residues, which indicates that the underlined AATAAA sequence (nucleotides 542–547) is a functional polyadenylation signal. λhIAPP-c2 corresponds to nucleotides 311–1015 with 32 A residues added to the 3′-end. The Alu sequence corresponds to nucleotides 73–1015. A few nucleotides in the 3′-untranslated region of the cDNAs differ from the reported genomic sequence [6], but they were all confirmed at the genomic level. Nucleotide sequence differences (underlined) with [19] are all A to G differences, and are all located within the Alu sequence.

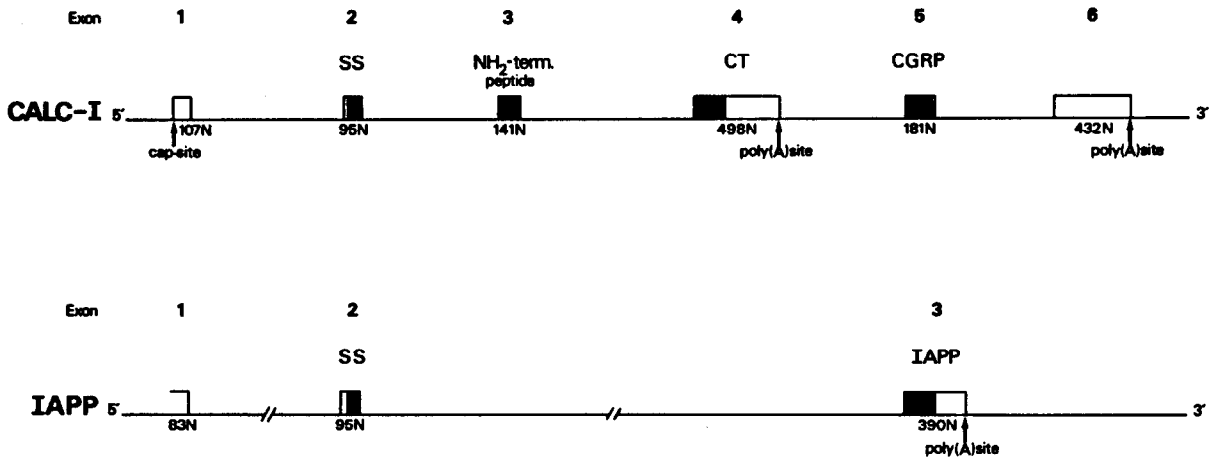

Fig.3. Comparison of the structural organization of the human CALC-I gene and part of the human IAPP gene. Exons and introns are drawn to scale, except for the introns in the IAPP gene. The distance between exons 2 and 3 in the IAPP gene is approx. 5 kb, whereas the distance between exons 2 and 5 in the CALC-I gene is approx. 3 kb. Polyadenylation sites are indicated. The coding regions are indicated by solid boxes. Calcitonin (CT) mRNA contains exons 1, 2, 3 and 4, whereas CGRP-I mRNA contains exons 1, 2, 3, 5 and 6 of the CALC-I gene. Exons 1 of the IAPP and CALC-I gene contain 5'-non-coding sequences and exons 2 encode the signal sequences (SS). In the CALC-I gene exon 3 encodes an amino-terminal peptide, exon 4 encodes the 32-amino acid calcitonin and a 16-amino acid carboxy-terminal peptide, exon 5 encodes CGRP and a carboxy-terminal tetrapeptide and exon 6 contains 3'-non-coding sequences. In the IAPP gene exon 3 encodes IAPP and a 16-amino acid carboxy-terminal peptide.

During the preparation of this paper, similar results were reported by Sanke et al. [19]. However, none of the 3 insulinoma cDNAs reported by these authors has the same structural organization as λhIAPP-c1. In one of their cDNAs, the same polyadenylation signal is used as in λhIAPP-c1, however, their cDNA is derived from an incompletely processed IAPP RNA. A further downstream variant polyadenylation signal ATTAAA is used in another of their cDNAs [19]. Alternative use of different polyadenylation signals in IAPP RNAs in human insulinomas may explain the size difference between the 1.6 and 2.1 kb IAPP RNAs, which we reported previously [6]. It will be important to see whether these two polyadenylated IAPP RNAs are insulinoma-specific or whether they are also found in normal pancreas.

IAPP has amino acid sequence homology with the human CGRPs, which are encoded by exon 5 of the CALC-I and -II genes [20]. At the nucleotide level there is 59% homology between the 111 nucleotides encoding the 37 amino acids of IAPP and CGRP-II, respectively. In addition, the location of the splice acceptor site of the exons encoding IAPP and the CGRPs is conserved. Furthermore, the exons 2 of both the IAPP and CALC genes are 95 nucleotides in length, contain the ATG start codon and encode the signal sequence. On the other hand, there is a striking length difference between the amino-terminal peptides of the IAPP precursor (9 residues) and of the CGRP precursors (54/55 residues). In the CALC genes the amino-terminal peptides are encoded for the larger part by exon 3 [20]. Unlike the CGRP mRNAs, where a 3'-untranslated exon (exon 6) is present, IAPP mRNA is polyadenylated at the 3'-end of the IAPP-encoding exon. We hypothesize that the IAPP gene and the CALC genes are evolutionarily related, but that the exon 3 and exon 6 equivalents of the CALC genes are absent in IAPP mRNA (see fig.3).

Acknowledgements: This research was supported by the Netherlands Organization for Chemical Research (SON), with financial aid from the Netherlands Organization for Scientific Research (NWO), and Ciba-Geigy AG, Basel, Switzerland. We would like to thank Dr A.D.M. van Mansfeld and Mrs H.A.A.M. van Teeffelen for the isolation of the insulinoma (poly A^+) RNA, and Drs P.D. Baas and A.D.M. van Mansfeld for discussions and critical reading of the manuscript. E.D. Kluis and A.A.X. Reintjens assisted in the preparation of the manuscript.

REFERENCES

[1] Westermark, P. and Wilander, E. (1978) Diabetologia 15, 417–421.

[2] Westermark, P., Wernstedt, C., Wilander, E. and Sletten, K. (1986) Biochem. Biophys. Res. Commun. 140, 827–831.

[3] Westermark, P., Wernstedt, C., Wilander, E., Hayden, D.W., O'Brien, T.D. and Johnson, K.H. (1987) Proc. Natl. Acad. Sci. USA 84, 3881–3885.

[4] Clark, A., Lewis, C.E., Willis, A.C., Cooper, G.J.S., Morris, J.F., Reid, K.B.M. and Turner, R.C. (1987) Lancet, 231–234.

[5] Cooper, G.J.S., Willis, A.C., Clark, A., Turner, R.C., Sim, R.B. and Reid, K.B.M. (1987) Proc. Natl. Acad. Sci. USA 84, 8628–8632.

[6] Mosselman, S., Höppener, J.W.M., Zandberg, J., Van Mansfeld, A.D.M., Geurts van Kessel, A.H.M., Lips, C.J.M. and Jansz, H.S. (1988) FEBS Lett. 239, 227–232.

[7] Steenbergh, P.H., Höppener, J.W.M., Zandberg, J., Lips, C.J.M. and Jansz, H.S. (1985) FEBS Lett. 183, 403–407.

[8] Westermark, P., Wilander, E., Westermark, G.T. and Johnson, K.H. (1987) Diabetologia 30, 887–892.

[9] Cooper, G.J.S., Willis, A.C., Reid, K.B.M., Clark, A., Baker, C.A., Turner, R.C., Lewis, C.E., Morris, J.F., Howland, K. and Rothbard, J.B. (1987) Lancet, 966.

[10] Leighton, B. and Cooper, G.J.S. (1988) Nature 335, 632–635.

[11] Cooper, G.J.S., Leighton, B., Dimitriadis, G.D., Parry-Billings, M., Kowalchuk, J.M., Howland, K., Rothbard, J.B., Willis, A.C. and Reid, K.B.M. (1988) Proc. Natl. Acad. Sci. USA 85, 7763–7766.

[12] MacDonald, R.J., Swift, G.H., Przybyla, A.E. and Chirgwin, J.M. (1987) Methods Enzymol. 152, 219–227.

[13] Aviv, M. and Leder, P. (1972) Proc. Natl. Acad. Sci. USA 69, 1408–1412.

[14] Feinberg, A. and Vogelstein, B. (1983) Anal. Biochem. 132, 6–13.

[15] Maniatis, T., Fritsch, E.F. and Sambrook, J. (1982) Molecular Cloning, Cold Spring Harbour.

[16] Maxam, A.M. and Gilbert, W. (1977) Proc. Natl. Acad. Sci. USA 74, 560–564.

[17] Von Heijne, G. (1983) Eur. J. Biochem. 133, 17–21.

[18] Jelinek, W.R. and Schmid, C.W. (1982) Annu. Rev. Biochem. 51, 813–844.

[19] Sanke, T., Bell, G.I., Sample, C., Rubenstein, A.H. and Steiner, D.F. (1988) J. Biol. Chem. 263, 17243–17246.

[20] Steenbergh, P.H., Höppener, J.W.M., Zandberg, J., Visser, A., Lips, C.J.M. and Jansz, H.S. (1986) FEBS Lett. 209, 97–103.