Identification of a Human Insulinoma cDNA Encoding a Novel Mammalian Protein Structurally Related to the Yeast Dibasic Processing Protease Kex2

(Received for publication, October 20, 1989)

Steven P. Smeeckens and Donald F. Steiner
From the Howard Hughes Medical Institute and the Department of Biochemistry and Molecular Biology, The University of Chicago, Chicago, Illinois 60637

We have identified a human insulinoma cDNA (PC2) that encodes a protein homologous to the precursor processing Kex2 endoprotease of yeast by using a polymerase chain reaction to detect and amplify conserved sequences within the catalytic site. The 638-residue amino acid sequence of PC2 begins with a cleavable signal peptide, indicating that it enters the secretory pathway, and contains a 282-residue domain that is homologous to the catalytic modules of both Kex2 and the related bacterial subtilisins. Within this region 49 and 27% of the amino acids are identical to those in the aligned Kex2 and subtilisin BPN' sequences, respectively, and the catalytically essential Asp, His, and Ser residues are all conserved. Northern blot analysis revealed the presence of 2.4- and 5.0-kilobase hybridizing bands in mRNA from the insulinoma. The PC2 protein also shows great similarity to the incomplete NH2-terminal sequence of the human furin gene product, a putative membrane-inserted receptor-like molecule. We propose that PC2 is a member of a family of mammalian Kex2/subtilisin-like proteases that includes members involved in a number of specific proteolytic events within cells, including the processing of prohormones.

Proteolytic processing at dibasic amino acids represents an important step in the maturation of a large number of prohormones, neuropeptides, and other biologically active peptides and proteins (1, 2). Despite the widespread occurrence of this mechanism in nature, little is known about the endoproteases involved in this process. Recently, two Ca2+-dependent proteolytic activities have been partially purified from an insulinoma and have been shown to cleave proinsulin appropriately at the B chain-C peptide and C peptide-A chain junctions (3). Designated types I and II, these activities specifically cleave on the carboxyl side of Arg-Arg and Lys-Arg residues, respectively. A similar Ca2+-dependent proteolytic activity has also recently been described in liver and appears to be involved in the processing of proalbumin (4, 5).

The yeast Saccharomyces cerevisiae also requires endoproteolytic cleavage at dibasic amino acids in the processing of proteins involved in its life cycle. The a-mating factor peptide is translated in tandem copies which must be cleaved at Lys-Arg residues to be released (6, 7). In addition, the yeast prokiller factor requires proteolytic processing at both Lys-Arg and Arg-Arg residues to be activated (8, 9). In both of these cases the endoprotease involved has been mapped to the KEX2 locus (10). Biochemical characterization of Kex2 has revealed it to be a Ca2+-dependent serine protease (11, 12). Upon cloning (13), Kex2 was found to be a subtilisin-like protease since it contained an active site domain homologous to the bacterial subtilisins (12, 13). Furthermore, Kex2 has been shown to process proinsulin expressed in yeast (14) and also to process proproiominelanocortin when transfected into proproiominelanocortin-secreting mammalian cells (15).

In order to explore the possible existence of processing enzymes related to the yeast Kex2 protease in pancreatic β cells, we have used polymerase chain reaction (PCR) in a strategy we have called amplification of homologous DNA fragments (16). We report here the identification and characterization of a cDNA from a human insulinoma which encodes a protein with a high degree of similarity to Kex2 that may represent a homologous mammalian converting protease.

EXPERIMENTAL PROCEDURES

Polymerase Chain Reactions—The degenerate nucleotide sequences of the primers used for PCR were derived from the amino acid sequences surrounding the aspartate and histidine active site residues of the yeast KEX2 gene and the bacterial subtilisins. The consensus amino acid and nucleotide sequences used for the aspartate site primer were

$$5' - CG G T G A T T G A G C G G A G C G G C T T - 3'$$

and the histidine active site region and the corresponding complementary primer used were as follows.

$$5' - CG G T G A T T G A G C G G A G C G G C T T - 3'$$

Reactions for PCR contained cDNA template, 100 pmol of each primer, and 2.5 units of Taq DNA polymerase (Intron) in 10 mM Tris (pH 8.3), 50 mM KCl, 1.5 mM MgCl2, 0.2 mM dNTPs, 0.01% (w/v) gelatin in a volume of 0.1 ml. Reactions were carried out in a Perkin-Elmer Cetus thermal cycler for 30 cycles of denaturation (94 °C, 1 min), annealing (55 °C, 2 min), and extension (72 °C, 3 min). The cDNA templates for PCR were prepared from human insulinoma total RNA (6 μg) using Moloney murine leukemia virus reverse transcriptase as suggested by the supplier (Bethesda Research Laboratories). One-fourth of the cDNA products of this reaction was used for each subsequent PCR reaction. Following fractionation of the PCR products on a polyacrylamide gel, the band of interest was electroeluted and then amplified further by 15 cycles of PCR before

* This work was supported in part by the Howard Hughes Medical Institute and by National Institutes of Health Grants DK 13914 and DK 20595. The costs of publication of this article were defrayed in part by the payment of page charges. This article must therefore be hereby marked "advertisement" in accordance with 18 U.S.C. Section 1734 solely to indicate this fact.

The nucleotide sequence(s) reported in this paper has been submitted to the GenBank™/EMBL Data Bank with accession number(s) J05252.

The abbreviations used are: PCR, polymerase chain reaction; kb, kilobase(s).
being blunt-ended with T4 DNA polymerase and ligated into the EcoRV site of pBluescript (Strategene).

Construction and Screening of the cDNA Library—Total RNA was isolated from human insulinoma using the guanidinium isothiocyanate/cesium trifluoroacetic acid procedure (17). Poly(A)+ RNA was selected by oligo(dT)-cellulose chromatography using a commercial kit (Stratagene). Double-stranded cDNA was synthesized with a DNA kit (Pharmacia LKB Biotechnology Inc.) followed by the addition of EcoRI adaptors and cloning into the EcoRI site of pBluescript. To screen the library, was used as the DNA template for 30 cycles of PCR carried out as described above. This was followed by reamplification of one-tenth (10%) of this reaction for 15 cycles in the presence of 0.2 mm dATP, dCTP, dGTP, dTTP, 0.01 mm d32P-dCTP, and 25 µCi of [α-32P]dCTP (3000 Ci/mmol) (Du Pont-New England Nuclear).

Northern Blotting—RNA for Northern blots was fractionated on 1% agarose/formaldehyde gels. Following transfer to nitrocellulose, the blot was hybridized (2x SSC, 50% formamide, 37°C for 18 h) and washed (0.1x SSC, 0.1% sodium dodecyl sulfate, 65°C for 1 h) as described (18). Following plaque purification, the cloned inserts were subcloned into the EcoRI site of pBluescript. To screen the library, labeled probe was prepared as follows. pPCR1 plasmid DNA (10 ng) was used as the DNA template for 30 cycles of PCR carried out as described above. This was followed by reamplification of one-tenth (10%) of this reaction for 15 cycles in the presence of 0.2 mm dATP, GTP, CTP, TTP, 0.01 mm d32P-dCTP, and 25 µCi of [α-32P]dCTP (3000 Ci/mmol) (Du Pont-New England Nuclear).

RESULTS AND DISCUSSION

Examination of a large number of proteolytic enzymes has shown that both the amino acid sequences surrounding the active site residues and the distance between the catalytic sites are highly conserved within any one family. In an effort to identify a mammalian gene homologous to the yeast KEX2 gene, we therefore designed two degenerate oligonucleotides based upon the consensus amino acid sequence surrounding the aspartate and histidine active site residues of Kex2 and the related bacterial subtilisins ("Experimental Procedures"). These oligonucleotides were used to prime PCR amplification of cDNA synthesized from human insulinoma total RNA. Analysis of the PCR products by polyacrylamide gel electrophoresis revealed a major band of 150 nucleotides (Fig. 1A). Since this was of a length consistent with the distance between the Asp and His catalytic residues encoded by the KEX2 gene, the DNA was subcloned for further analysis. Sequencing of the cloned DNA (designated pPCR1) revealed that one of the two potential open reading frames displayed extensive amino acid sequence similarity to the corresponding region of the KEX2 gene. In addition, Northern blot analysis indicated the presence of both 5' and 2.8 kb transcripts in PC2. Within this 246-amino acid segment 27% of the residues in this region were identical. Moreover, an additional 35% of the residues in this region were similar. In all of these comparisons, the Asp, His, and Ser residues of the catalytic triads aligned exactly.

ProteaseB is a vacuolar protease from yeast that is also related to the subtilisins (20). A comparison of PC2 to proteaseB, however, revealed little similarity outside of the amino acid sequences directly adjacent to the catalytic residues (Fig. 3). Comparison of Kex2 and proteaseB showed only

![Fig. 1. Polyacrylamide gel analysis of the products of PCR and Northern blot hybridization of the amplified cloned cDNA from human insulinoma mRNA. A, 6% polyacrylamide gel electrophoresis of the products obtained following polymerase chain amplification of human insulinoma cDNA using the Asp and His primers described under "Experimental Procedures." The position of the 150-base pair band electroeluted and cloned (pPCR1) is indicated by the arrow. Size markers are HaeIII-digested dX174 (Bethesda Research Laboratories). B, the cloned 150-base pair band from A was PCR-labeled and used to probe human insulinoma poly(A)-selected RNA fractionated on 1% agarose/formaldehyde gel (see "Experimental Procedures"). The position of ribosomal RNAs is indicated. Hybridization to insulinoma total RNA indicates that the higher molecular weight band does not correspond to 28 S RNA (not shown).
FIG. 3. Comparison of selected regions of PC2 to the amino acid sequences surrounding the active site catalytic residues of Kex2, proteaseB, and subtilisin BPN'. The relative position of the amino acid segment within each of the proteins is indicated by the numbers to the left of each column. Residues identical to those of PC2 are boxed. Catalytic residues are indicated by an asterisk. Amino acids are indicated by their single-letter abbreviations.

21% amino acid sequence identity over a 334-amino acid region that included the active site residues. It thus appears that within the active site domain, PC2 is more closely related to Kex2 than Kex2 is to proteaseB.

A comparison of the overall domain structures of PC2 and Kex2 is displayed in Fig. 4. Both proteins contain a putative signal peptide followed by the subtilisin-like domain. Although this region contains the highest levels of homology between the two proteins (see above), Fig. 4 shows that PC2 contains amino acid sequence similarity with Kex2 throughout its sequence. Of the first 594 amino acids of PC2, 34% are identical and 41% similar to those of the aligned Kex2 sequence. By comparison, 40-50% of the amino acids are identical between related mammalian members of the trypsin family of serine proteases (21). The end of the similarity between PC2 and Kex2 is punctuated by the end of PC2. Beyond this point, Kex2 contains a Ser/Thr-rich region followed by a transmembrane-spanning domain.
insulin and growth factor receptors as well as a transmembrane domain resembling those of the class II major histocompatibility complex antigens. Both of these domains are contained within the COOH-terminal half of the protein (22). Since the complete furin mRNA has not been cloned the nature of the amino-terminal region of this protein remains unknown. Interestingly, the similarity between furin and PC2 extends over the first 280 amino acids of the cloned furin fragment while spanning a 287-amino-acid segment near the COOH terminus of PC2 (Asp<sup>310</sup> to Ser<sup>385</sup>). Within this region, 48% of the amino acids are identical. Likewise, the same region of furin shows 36% sequence identity with the corresponding region of Kex2 (Gly<sup>315</sup> to Ser<sup>395</sup>). Given these high levels of similarity and the fact that this overlap includes a subtilisin-like serine active site region within furin that aligns to the putative active site serine residues of both PC2 (Ser<sup>383</sup>) and Kex2 (Ser<sup>386</sup>), we infer that the uncharacterized amino-terminal portion of furin may also encode a subtilisin/Kex2-like catalytic domain.

The high degree of similarity between PC2 and Kex2 make it tempting to ascribe a role for this mammalian gene in the endoproteolytic processing of prohormones. However, the functional specificity of PC2 must await biochemical characterization of the encoded protein, e.g. the presence of Asp rather than Asn at position 310 might lower the catalytic efficiency of this putative protease (23). Further studies are under way to rule out a possible cloning artifact or point mutation occurring in the insulinoma DNA.

Acknowledgments—We wish to thank Drs. Krishna M. Finnama-ni, Arthur O'Connor, and Bruce D. Ragsdale, Desert Samaritan Hospital and Health Center, Mesa, AZ, for providing the human insulinoma tissue, Anthony Avruch for expert technical assistance, Paul Gardner for oligonucleotide synthesis, and Florence Rozenfeld for assistance in preparing this manuscript.

REFERENCES

1. Docherty, K., and Steiner, D. F. (1982) Annu. Rev. Physiol. 44, 625-638
2. Andrews, P. C., Brayton, K., and Dixon, J. E. (1987) Experientia 43, 784-790
3. Davison, H. H. W., Rhodes, C. J., and Hutton, J. C. (1988) Nature 333, 93-96
4. Frimann, O., and Peach, R. J. (1988) FEBS Lett. 229, 167-170
5. Oda, K., and Ikehara, Y. (1988) J. Biochem. (Tokyo) 104, 159-161
6. Kurjan, J., and Herskowitz, I. (1982) Cell 30, 933-943
7. Julius, D., Blair, L., B advertisement in the alignment are indicated by hyphens. Active site catalytic residues are indicated by an asterisk.

domain in Kex2 is thought to be involved in O-linked glycosylation of the protein. Both PC2 (Fig. 2) and Kex2 (13) possess consensus sequences for N-linked glycosylation. While analysis of PC2 indicates that it does not possess a transmembrane-spanning domain and therefore may not be associated with a membrane in vivo, it is of interest that Northern blot analysis of the human insulinoma RNA using either pPCR1 (Fig. 1B) or PC2 (not shown) as a probe consistently demonstrated the presence of both 5- and 2.8-kb transcripts in these cells. Whether one of these encodes a form of PC2 containing a transmembrane-spanning domain or the larger transcript simply represents splicing intermediates accumulating in the insulinoma remains to be determined. Finally, both PC2 and Kex2 (12, 13) possess highly charged, though otherwise unrelated, COOH-terminal tails. In PC2 45% of the 38 carboxyl-terminal amino acids are either acidic or basic residues.

A search of the NBRF-PIR protein sequence library revealed that PC2 was also related to the human furin gene product. Furin was identified based upon its proximity to the c-fes/fps proto-oncogene and is transcribed as a 4.5-kb mRNA (22). Cloning and sequencing of 3.1 kb of the furin mRNA revealed a cystine-rich region with homology to the human
Identification of a human insulinoma cDNA encoding a novel mammalian protein structurally related to the yeast dibasic processing protease Kex2.
S P Smeekens and D F Steiner

J. Biol. Chem. 1990, 265:2997-3000.

Access the most updated version of this article at http://www.jbc.org/content/265/6/2997

Alerts:
- When this article is cited
- When a correction for this article is posted

Click here to choose from all of JBC's e-mail alerts

This article cites 0 references, 0 of which can be accessed free at http://www.jbc.org/content/265/6/2997.full.html#ref-list-1