Volume 300, number 1, 82–88 FEBS 10854 March 1992

Expression in human lung tumor cells of the proprotein processing enzyme PC1/PC3

Cloning and primary sequence of a 5 kb cDNA

John W.M. Creemers, Anton J.M. Roebroek and Wim J.M. Van de Ven

Molecular Oncology Section, Center for Human Genetics, University of Leuven, Herestraat 49, B-3000 Leuven, Belgium

Received 13 January 1992; revised version received 10 February 1992

Northern blot analysis of human lung tumors indicated that the gene, which encodes the subtilisin-like proprotein processing enzyme PC1/PC3, was highly expressed in almost all carcinoid tumors tested. In small cell lung carcinomas (SCLCs), expression varied. In non-SCLCs and normal lung, no expression was found. Analysis of SCLC cell lines revealed that expression was restricted preferentially to cell lines of the classical type. In lung tumor cells expressing the PC1/PC3 gene, transcripts of 3 kb and 5 kb were detected, the 5 kb mRNA always being the most abundant species. We isolated a cDNA corresponding to the 5 kb human PC1/PC3 transcript, determined the nucleotide sequence of it and deduced the amino acid sequence of the corresponding protein. Furthermore, we conclude that the two PC1/PC3 transcripts have 3' non-coding regions of different size and encode the same protein.

Subtilisin-like proprotein processing enzyme; PC1/PC3; Human lung tumor

1. INTRODUCTION

Recently, the discovery of the coding sequences for a number of new mammalian endoproteolytic processing enzymes was reported [1–8]. The enzymes were named furin [1–3], PC1 [4,7], PC2 [4,5] and PC3 [6,8]. Two groups independently discovered the coding sequences for the same mouse enzyme. As a result, different names were assigned, one group called the enzyme PC1 [4,7], the other named it PC3 [6]. To avoid confusion here, we will refer to this enzyme as PC1/PC3. The newly discovered enzymes were demonstrated to exhibit cleavage specificity for sites specified by particular sequences of basic amino acid residues; most frequently paired basic residues (Lys-Arg or Arg-Arg).

The furin enzyme, which can be considered as the mammalian prototype, is encoded by the *fur* gene [1] and was found widely expressed [9–11]. The coding sequence of the human [3], mouse [10] and rat [12] *fur* genes have been fully characterized and *fur*-like genes have been isolated from *Drosophila melanogaster*, D*fur*1 and D*fur*2 [13], and *Caenorhabditis elegans*, *bli*-4 [14]. Mammalian furin appeared capable to correctly process the precursor of von Willebrand factor [15,16], β-nerve growth factor [17], proalbumin [18] and complement pro-C3 (the third component of complement) [18]. As far as the two other newly discovered enzymes are concerned, the coding sequences of mouse PC1/PC3 [6–8] and those of human [5] and mouse [4] PC2 have also been fully characterized. The genes which encode the PC1/PC3 and PC2 proteins have recently been designated as *NEC*1 and *NEC*2, respectively [19]. They seem to have an expression pattern that is restricted to endocrine and neural tissue [4,6]. PC1/PC3 and PC2 appeared capable of correctly cleaving the multifunctional precursor protein proopiomelanocortin, each at distinct pairs of basic residues [20,21]. Structurally and functionally, the three new enzymes closely resemble the prohormone processing enzyme Kex2 which is encoded by the *KEX*2 gene of yeast *Saccharomyces cerevisiae*. Kex2 is a membrane-associated, Ca^{2+}-dependent, subtilisin-like serine endoprotease with a neutral pH optimum [22]. In the primary translational product of *KEX*2 gene, a number of domains have been defined: a 'prepro' domain, a subtilisin-like catalytic domain, a 'P' domain which is critical for catalytic activity, a serine/threonine-rich region, a transmembrane anchor and a small cytoplasmic tail [2,23]. Similar domains are also found in furin, except that furin has no serine/threonine-rich region but a cysteine-rich region just preceding a presumptive transmembrane domain [11,15]. Using computer-assisted molecular modelling with data of known members of the subtilisin-like family of serine proteases [24], a model for the catalytic domain of furin was proposed [15]. The enzymes PC1/PC3 and PC2 apparently lack a serine/threonine- and a cysteine-rich

Correspondence address: J.W.M. Creemers, Molecular Oncology Section, Center for Human Genetics, University of Leuven, Herestraat 49, B-3000 Leuven, Belgium.

Published by Elsevier Science Publishers B.V.

region. Furthermore, they also do not possess a carboxy-terminal hydrophobic region that could act as a transmembrane anchor; however, association with membranes might occur through the carboxy-terminal amphipathic α-helical structures in these two enzymes [6]. Altogether, the properties of the newly discovered enzymes support the presumption that these enzymes constitute the first representatives of the long-sought convertases of precursors of polypeptide hormones, neuropeptides, growth factors, membrane-associated growth factor receptors, plasma proteins and envelope glycoproteins of viral pathogens.

While the long mysterious identity of the mammalian enzymes responsible for proprotein processing is now readily being uncovered, it is of interest to establish the degree of enzyme diversity in this novel and still small enzyme family. Since some cleavage specificities are still unaccounted for, discovery of additional members of this gene family might be expected. However, some of the enzyme diversity might also be obtained via alternative splicing by which the presence or absence of particular domains is controlled and enzyme specifications determined. In Northern blot analysis of neuroendocrine tissues, PC1/PC3-specific mRNA species of 3 kb and 5 kb [5,8] were detected, apparently with the 3 kb transcript as the more abundant one. Similarly, expression of the D*fur*1 gene of *Drosophila melanogaster* revealed transcripts of 4.0, 4.5 and 6.5 kb [13]. Detailed analysis of these D*fur*1 transcripts revealed that the observed differences between them were partly due to alternative splicing which resulted in protein diversity (Roebroek et al. in preparation). In this report, we characterize the coding sequences of human PC1/PC3 and define the molecular basis for the differences between the 3 kb and 5 kb PC1/PC3 transcripts. To accomplish this, we have first searched for human cells that preferentially expressed the 5 kb transcript of *NEC*1. Subsequently, we isolated and molecularly characterized a cDNA clone corresponding to this 5 kb human *NEC*1 transcript and compared the data to those of a mouse PC1/PC3 cDNA clone corresponding to the 3 kb transcript.

2. MATERIALS AND METHODS

2.1. *Primary lung tumors and lung carcinoma cell lines*

Lung carcinoma specimens from 24 patients were selected from the files of the Pathology Department of the St. Antonius Hospital, Nieuwegein, The Netherlands. Specimens were processed as described before [25]. Tumors were characterized by routine microscopical and histopathological techniques and classified according to the criteria of the WHO [26] as SCLC (6 cases), carcinoid tumors (5 cases) or non-SCLC (13 cases).

The cell lines used in this study were all established human SCLC cell lines, which have been described before [25]. The SCLC cell lines GLC-1 and GLC-1-M13 were obtained from L. de Leij, and the cell lines SCLC-16HC, SCLC-16HV, SCLC-21H, and SCLC-22H from G. Bepler. NL-SCLC3 was obtained from J. Broers [27]. All other SCLC cell lines (National Cancer Institute series) were obtained from D.N. Carney.

2.2. *RNA isolation and Northern blot analysis*

Total cellular RNA was isolated from primary lung tumors or cell lines using the lithium-urea procedure described by Auffray and Rougeon [28]. 15 μg of total RNA was glyoxylated and size fractionated on a 1% agarose gel and transferred to Hybond-N (procedure as recommended by Amersham Corp.). For hybridization on nylon membranes the method of Church and Gilbert [29] was used.

2.3. *Construction and screening of a human carcinoid cDNA library*

Purified mRNA was isolated from total RNA of a primary human carcinoid tumor of the lung by oligo(dT)-cellulose affinity chromatography. An oligo(dT)-primed cDNA library was constructed in λgt11 using a Pharmacia/LKB cDNA synthesis kit. About 150,000 plaques obtained upon infection of *Escherichia coli* Y1090 were screened using the 2.6 kb mouse PC3 cDNA [6] as a molecular probe, which was kindly provided by D. Steiner, Chicago. Hybridization experiments on nitrocellulose membranes were performed as previously described [25].

2.4. *Nucleotide sequence analysis*

Nucleotide sequences were determined according to the dideoxy chain termination method using the T7 polymerase sequencing kit of Pharmacia/LKB. The DNA fragments to be sequenced were sub-

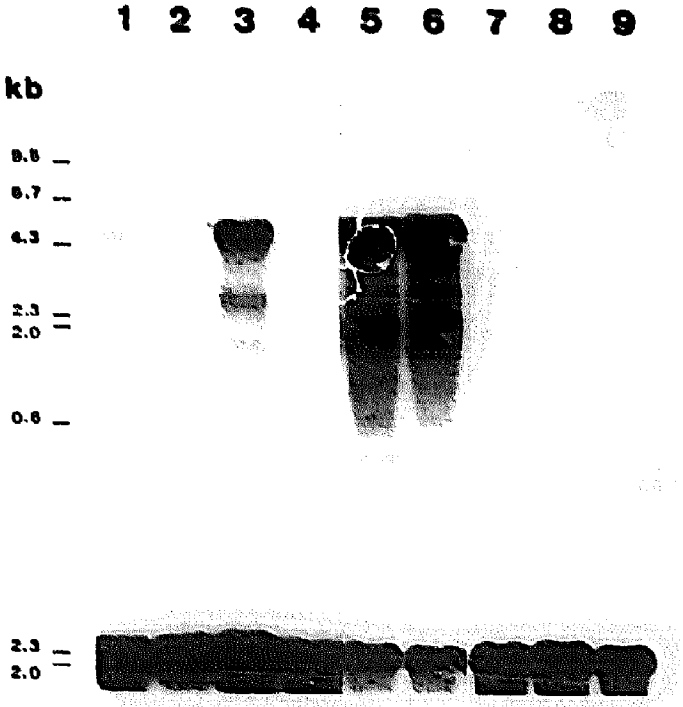
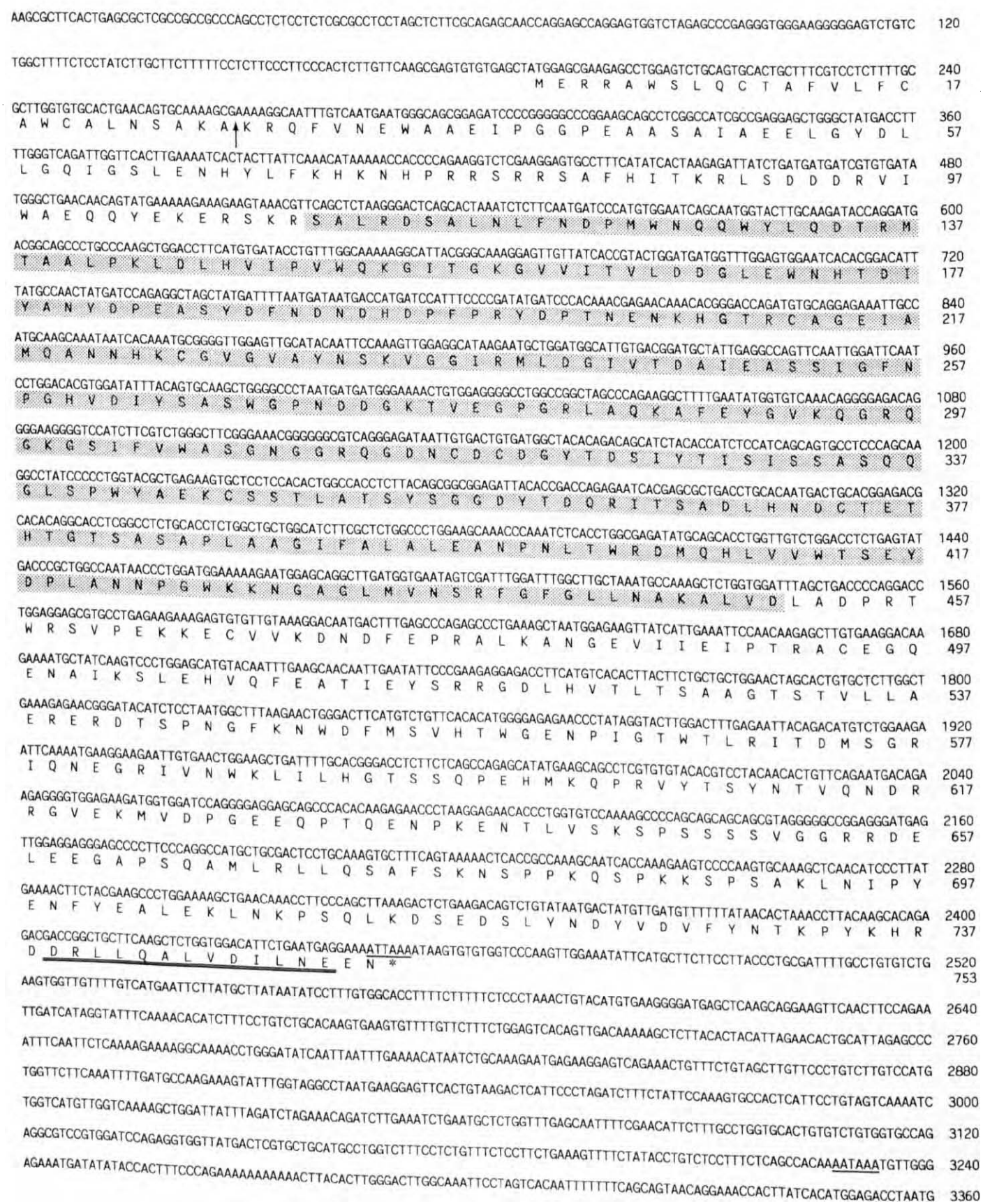


Fig. 1. Expression of the PC1/PC3 gene in cells of normal human lung and lung tumors. Northern blot analysis was performed with total RNA (amounts of 15 μg) isolated from the SCLC cell lines NCI-H249 (lane 1) and NL-SCLC3 (lane 2), two primary human SCLCs (lanes 3 and 4), two carcinoid tumors of the lung (lanes 5 and 6), two primary non-SCLCs (lanes 7 and 8), or normal human lung (lane 9). As a molecular probe, a 2.6 kb mouse PC1/PC3 cDNA was used. Using intensifying screens, exposure time of X-ray films was 40 h. The positions of the molecular weight markers, *Hin*dIII-digested λ DNA, are indicated. In control experiments, blots were dehybridized and subsequently hybridized with a hamster actin cDNA probe [25] under similar conditions as described above. Exposure time of X-ray films was 20 h. kb, kilobase.


AAGCGCTTCACTGAGCGCTCGCCGCCGCCCAGCCTCTCCTCTCGCGCCTCCTAGCTCTTCGCAGAGCAACCAGGAGCCAGGAGTGGTCTAGAGCCCGAGGGTGGGAAGGGGGAGTCTGTC 120
TGGCTTTTCTCCTATCTTGCTTCTTTTTCCTCTTCCCTTCCCACTCTTGTTCAAGCGAGTGTGTGAGCTATGGAGCGAAGAGCCTGGAGTCTGCAGTGCACTGCTTTCGTCCTCTTTTGC 240
M E R R A W S L Q C T A F V L F C 17
GCTTGGTGTGCACTGAACAGTGCAAAAGCGAAAAGGCAATTTGTCAATGAATGGGCAGCGGAGATCCCCGGGGGCCCGGAAGCAGCCTCGGCCATCGCCGAGGAGCTGGGCTATGACCTT 360
A W C A L N S A K A K R Q F V N E W A A E I P G G P E A A S A I A E E L G Y D L 57
TTGGGTCAGATTGGTTCACTTGAAAATCACTACTTATTCAAACATAAAAACCACCCCAGAAGGTCTCGAAGGAGTGCCTTTCATATCACTAAGAGATTATCTGATGATGATCGTGTGATA 480
L G Q I G S L E N H Y L F K H K N H P R R S R R S A F H I T K R L S D D D R V I 97
TGGGCTGAACAACAGTATGAAAAAGAAAGAAGTAAACGTTCAGCTCTAAGGGACTCAGCACTAAATCTCTTCAATGATCCCATGTGGAATCAGCAATGGTACTTGCAAGATACCAGGATG 600
W A E Q Q Y E K E R S K R S A L R D S A L N L F N D P M W N Q Q W Y L Q D T R M 137
ACGGCAGCCCTGCCCAAGCTGGACCTTCATGTGATACCTGTTTGGCAAAAAGGCATTACGGGCAAAGGAGTTGTTATCACCGTACTGGATGATGGTTTGGAGTGGAATCACACGGACATT 720
T A A L P K L D L H V I P V W Q K G I T G K G V V I T V L D D G L E W N H T D I 177
TATGCCAACTATGATCCAGAGGCTAGCTATGATTTTAATGATAATGACCATGATCCATTTCCCCGATATGATCCCACAAACGAGAACAAACACGGGACCAGATGTGCAGGAGAAATTGCC 840
Y A N Y D P E A S Y D F N D N D H D P F P R Y D P T N E N K H G T R C A G E I A 217
ATGCAAGCAAATAATCACAAATGCGGGGTTGGAGTTGCATACAATTCCAAAGTTGGAGGCATAAGAATGCTGGATGGCATTGTGACGGATGCTATTGAGGCCAGTTCAATTGGATTCAAT 960
M Q A N N H K C G V G V A Y N S K V G G I R M L D G I V T D A I E A S S I G F N 257
CCTGGACACGTGGATATTTACAGTGCAAGCTGGGGCCCTAATGATGATGGGAAAACTGTGGAGGGGCCTGGCCGGCTAGCCCAGAAGGCTTTTGAATATGGTGTCAAACAGGGGAGACAG 1080
P G H V D I Y S A S W G P N D D G K T V E G P G R L A Q K A F E Y G V K Q G R Q 297
GGGAAGGGGTCCATCTTCGTCTGGGCTTCGGGAAACGGGGGGCGTCAGGGAGATAATTGTGACTGTGATGGCTACACAGACAGCATCTACACCATCTCCATCAGCAGTGCCTCCCAGCAA 1200
G K G S I F V W A S G N G G R Q G D N C D C D G Y T D S I Y T I S I S S A S Q Q 337
GGCCTATCCCCCTGGTACGCTGAGAAGTGCTCCTCCACACTGGCCACCTCTTACAGCGGCGGAGATTACACCGACCAGAGAATCACGAGCGCTGACCTGCACAATGACTGCACGGAGACG 1320
G L S P W Y A E K C S S T L A T S Y S G G D Y T D Q R I T S A D L H N D C T E T 377
CACACAGGCACCTCGGCCTCTGCACCTCTGGCTGCTGGCATCTTCGCTCTGGCCCTGGAAGCAAACCCAAATCTCACCTGGCGAGATATGCAGCACCTGGTTGTCTGGACCTCTGAGTAT 1440
H T G T S A S A P L A A G I F A L A L E A N P N L T W R D M Q H L V V W T S E Y 417
GACCCGCTGGCCAATAACCCTGGATGGAAAAAGAATGGAGCAGGCTTGATGGTGAATAGTCGATTTGGATTTGGCTTGCTAAATGCCAAAGCTCTGGTGGATTTAGCTGACCCCAGGACC 1560
D P L A N N P G W K K N G A G L M V N S R F G F G L L N A K A L V D L A D P R T 457
TGGAGGAGCGTGCCTGAGAAGAAAGAGTGTGTTGTAAAGGACAATGACTTTGAGCCCAGAGCCCTGAAAGCTAATGGAGAAGTTATCATTGAAATTCCAACAAGAGCTTGTGAAGGACAA 1680
W R S V P E K K E C V V K D N D F E P R A L K A N G E V I I E I P T R A C E G Q 497
GAAAATGCTATCAAGTCCCTGGAGCATGTACAATTTGAAGCAACAATTGAATATTCCCGAAGAGGAGACCTTCATGTCACACTTACTTCTGCTGCTGGAACTAGCACTGTGCTCTTGGCT 1800
E N A I K S L E H V Q F E A T I E Y S R R G D L H V T L T S A A G T S T V L L A 537
GAAAGAGAACGGGATACATCTCCTAATGGCTTTAAGAACTGGGACTTCATGTCTGTTCACACATGGGGAGAGAACCCTATAGGTACTTGGACTTTGAGAATTACAGACATGTCTGGAAGA 1920
E R E R D T S P N G F K N W D F M S V H T W G E N P I G T W T L R I T D M S G R 577
ATTCAAAATGAAGGAAGAATTGTGAACTGGAAGCTGATTTTGCACGGGACCTCTTCTCAGCCAGAGCATATGAAGCAGCCTCGTGTGTACACGTCCTACAACACTGTTCAGAATGACAGA 2040
I Q N E G R I V N W K L I L H G T S S Q P E H M K Q P R V Y T S Y N T V Q N D R 617
AGAGGGGTGGAGAAGATGGTGGATCCAGGGGAGGAGCAGCCCACACAAGAGAACCCTAAGGAGAACACCCTGGTGTCCAAAAGCCCCAGCAGCAGCAGCGTAGGGGGCCGGAGGGATGAG 2160
R G V E K M V D P G E E Q P T Q E N P K E N T L V S K S P S S S S V G G R R D E 657
TTGGAGGAGGGAGCCCCTTCCCAGGCCATGCTGCGACTCCTGCAAAGTGCTTTCAGTAAAAACTCACCGCCAAAGCAATCACCAAAGAAGTCCCCAAGTGCAAAGCTCAACATCCCTTAT 2280
L E E G A P S Q A M L R L L Q S A F S K N S P P K Q S P K K S P S A K L N I P Y 697
GAAAACTTCTACGAAGCCCTGGAAAAGCTGAACAAACCTTCCCAGCTTAAAGACTCTGAAGACAGTCTGTATAATGACTATGTTGATGTTTTTTATAACACTAAACCTTACAAGCACAGA 2400
E N F Y E A L E K L N K P S Q L K D S E D S L Y N D Y V D V F Y N T K P Y K H R 737
GACGACCGGCTGCTTCAAGCTCTGGTGGACATTCTGAATGAGGAAAATTAAAATAAGTGTGTGGTCCCAAGTTGGAAATATTCATGCTTCTTCCTTACCCTGCGATTTTGCCTGTGTCTG 2520
D D R L L Q A L V D I L N E E N *
AAGTGGTTGTTTTGTCATGAATTCTTATGCTTATAATATCCTTTGTGGCACCTTTTCTTTTTCTCCCTAAACTGTACATGTGAAGGGGATGAGCTCAAGCAGGAAGTTCAACTTCCAGAA 2640
TTGATCATAGGTATTTCAAAACACATCTTTCCTGTCTGCACAAGTGAAGTGTTTTGTTCTTTCTGGAGTCACAGTTGACAAAAAGCTCTTACACTACATTAGAACACTGCATTAGAGCCC 2760
ATTTCAATTCTCAAAAGAAAAGGCAAAACCTGGGATATCAATTAATTTGAAAACATAATCTGCAAAGAATGAGAAGGAGTCAGAAACTGTTTCTGTAGCTTGTTCCCTGTCTTGTCCATG 2880
TGGTTCTTCAAATTTTGATGCCAAGAAAGTATTTGGTAGGCCTAATGAAGGAGTTCACTGTAAGACTCATTCCCTAGATCTTTCTATTCCAAAGTGCCACTCATTCCTGTAGTCAAAATC 3000
TGGTCATGTTGGTCAAAAGCTGGATTATTTAGATCTAGAAACAGATCTTGAAATCTGAATGCTCTGGTTTGAGCAATTTTCGAACATTCTTTGCCTGGTGCACTGTGTCTGTGGTGCCAG 3120
AGGCGTCCGTGGATCCAGAGGTGGTTATGACTCGTGCTGCATGCCTGGTCTTTCCTCTGTTTCTCCTTCTGAAAGTTTTCTATACCTGTCTCCTTTCTCAGCCACAAAATAAATGTTGGG 3240
AGAAATGATATATACCACTTTCCCAGAAAAAAAAAAACTTACACTTGGGACTTGGCAAATTCCTAGTCACAATTTTTTTCAGCAGTAACAGGAAACCACTTATCACATGGAGACCTAATG 3360

cloned in the pKUN19 or pGEM-3Zf(+) vector and sequenced using standard primers and primers synthesized based upon newly obtained sequences. The sequences were obtained from both strands and analyzed using the sequence analysis computer programs Genepro (Riverside Scientific), PC/Gene and Intelligenetics (IntelliGenetics, Inc.).

3. RESULTS AND DISCUSSION

The objective of this study was to characterize the coding sequences of the human PC1/PC3 gene, to define the molecular basis for the observed divergence in size of the major mRNA species transcribed from this gene, and to evaluate the possible implications of this divergence for the proteins that might be encoded by the PC1/PC3 mRNA s. Expression of the PC1/PC3 gene, which is apparently restricted to cells of endocrine and neural origin, apparently results in two major transcripts, as has been reported before [4,6]; one of about 3 kb, the other of about 5 kb. In studies described so far, the 3 kb transcript was very often the most abundant of the two PC1/PC3 transcripts. An exception, in this respect, constituted for example mouse pituitary corticotroph AtT-20 cells. In these cells, similar levels of the 3 kb and 5 kb transcripts were detected in one study [6]; low levels of the 5 kb transcript, however, were reported in another study [4]. The nucleotide sequence of mouse PC1/PC3 cDNAs of about 2.5 kb have been reported by several groups [6–8]. These cDNAs were isolated either from cDNA libraries of AtT-20 cells [6,8] or mouse insulinoma (βT3) cells [7]; βT3 cells preferentially express the 3 kb transcript. As far as the cDNA clones isolated from the AtT-20 cDNA library are concerned, no conclusion could be drawn as to which of the two PC1/PC3 transcripts they might correspond. cDNA clones isolated from the βT3 library, however, contained poly(dA)-stretches at their 3' ends and, therefore, it is reasonable to assume that they were derived from the PC1/PC3 transcripts of 3 kb.

In an attempt to obtain further insight in the structural differences between the two major mRNA species produced by the PC1/PC3 gene, we searched for human cells that preferentially expressed the 5 kb transcript. Therefore, we studied cells of human lung tumors of different histopathological types by Northern blot analysis. We selected these cells because some forms of SCLC and, more in particular, carcinoid tumors of the lung are known to display neuroendocrine features; therefore, they might express readily detectable levels of PC1/PC3 mRNAs. In Fig. 1, results of Northern blot analysis are shown to illustrate the levels of the 3 kb and 5 kb transcripts of PC1/PC3. As molecular probe, a mouse PC1/PC3 cDNA probe of about 2.6 kb was used as described in section 2. In the lower part of the figure, the actin mRNA levels in the various samples are given, as a control. The results of all Northern blot experiments are summarized in Table I and Table II. Initially, studies were performed with a variety of well-defined cell lines derived from human small cell lung carcinomas (Table I). Expression of PC1/PC3 transcripts of 3 kb and 5 kb were readily detectable in SCLC cell

```
TAATAATAGAAAAATACTCATAATAGGGAGAAACCAAGAGAAGTTTTGTTTTTGTTTTTTTCCAACTGTGTTCATTAGAACAGCGTGTTCTAAGTATTTGAAACTGAATGTTTATTCCTT  3480
GATACTAAAAGTTCTTCTCCAATCCTATCACTGATAGTGTCCAAATTCTCACCAAATTGCTCCTAAGCTTCAAATCAGAAGCAGAAACTGGCAGGCCATGGACCTTAATTGTCCCTCAGG  3600
TAGATTTTGTTTGGTATGCAGAATGTTTTTAAAATATGAGTGGTTATTGAAAATATGATGTTTCACATAAAACCTCATTCTCGGACCCATCTTTGCTCATGGCAACAGTTAGCTGGAGCT  3720
GAGTAGCAGCTGCCTGATTAGATGACTCTCAGTCCCCATGGCACCCTGCTCCATGTTACCTAGAGCAGGCACTTGATTCCTTGCTGGGCAGTATCCAATAGGCATTTGATTTTGCCCACT  3840
CCTACACTAAGCGAATGTGTACAAAGTGTAAATGCATTAGGAAAAACAAACTACCCGCATCTTCTGTTAGGCAGGATCTGTACAATAATAATTATGAGTTTGCTTATGTAATCTCACCTT  3960
ACCTGGATGATCACTAATACTAATTCATTTATTACTAACCTTCTGGCTTCCTTCTCTCAATATGCTTACAAAGTGTCCAGTCACCTACAATGCTGGCTTTCTCCCACTGAGTTTGCTGTT  4080
TGCAATTTTTCCATGAAGTTTGAACTTCATAAGGTAATTCATGGCATTGAACTGGTTCATGAAAAGAACACTAGAGTCTGTCATTTGCTTTGGCTTGAAGTATGGTTGGTAACACAAATT  4200
TTCACCTGCTCTTCTACCATTTGAATTTGTGTAGAGGGTGTTTGCAGAGCAATGCCCGTAATGCTTAGAGAATGTTCTCCTAAAAGACTTGCGGAATCACTCTGTCCTTGGAAGTTTCAT  4320
ATATTGTTTGATATGAAGTGTTAGATAGAATTTCCAATATTGGAGCATATCAAAAAGTATTAAAACTAAAAAGGACCAGAGAATTCTTAGATTGGCCCGGAAAGGCCAATAAAGAGTTAG  4440
AATGAAAACTCATTACTTTTCCATTCCCAATCTAGTGCTAGATGTATAAATCTTTCTTTTGATTCTTCCTAACAAAATATTTTCTGGGTTAAAACCCCAGCCAACTCATTGGGTTGTAGC  4560
CAAAGGTTCACTCTCAAGAAGCTTTAATATTTAAATAAAATCATATTGAATGTTTCCAACCTGGAGTATAATATTCAGATATAAAACAGTTTTGTCAGTCTTTCTTAGTGCCTGTGTGGA  4680
TTTTTGTGAAAATGTCAAAGAGAAAACTTATATACTATTTCCCTTGAAATTTTAAACTATATTTTCTTTACAGGTATTTATAATATACCAATGCTTTTATCAAACAGAATTTTAAAGAGC  4800
ATAATAAATTATATTAAAGAACCAAAAGTTTTCCTGAGAATAAGAAAGTTTCACCCAATAAAATATTTTTGAAAGGCATGTTCCTCTGTCAATGAAAAAAAGTACATGTATGTGTTGTGA  4920
TATTAAAAGTGACATTTGTCTAATAGCCTAATACAACATGTAGCTGAGTTTAACATGTGTGGTCTTGGTATTCTTAAGGGAACTTCCACATTATACATTTGATGTATTGACCAGAAT  5037
```

Fig. 2. Nucleotide sequence of the 5 kb transcript of the human PC1/PC3 gene and the predicted amino acid sequence of human PC1/PC3. Numbering of the amino acids and the nucleotides is indicated at the end of each line. The putative cleavage site for the signal peptide is indicated by an arrow. The amphipathic α-helical structure in the carboxy-terminus of the deduced PC1/PC3 protein is underlined with a double line. The predicted subtilisin-like catalytic domain is shaded and essential amino acid residues of its active site are D (167), H (208) and S (382). At position 309, the catalytically important asparagine residue (N) is present. The stop codon is marked with an asterisk. Consensus sequences for N-glycosylation are NHT (173–175) and NLT (401–403). An RGD sequence is present from amino acid position 518 to 520. Nucleotide sequences corresponding to the consensus signals (ATTAAA and AATAAA) for polyadenylation are underlined.

lines of the classical type (five out of five SCLC-C cell lines expressed PC1/PC3) but neither of these transcripts could be detected in most of the SCLC cell lines of the variant type; in seven out of eight SCLC-V cell lines, expression of PC1/PC3 was below the detection level. In the SCLC-V cell lines NCI-526 and NL-SCLC3, however, low levels of PC1/PC3 expression could be detected upon prolonged exposure of the X-ray films. All the SCLC-C cell lines tested here exhibited a clear neuroendocrine phenotype. They all expressed the neuroendocrine polypeptides 7B2 and gastrin-releasing peptide (GRP) [25]. The fact that the SCLC-C cell lines expressed PC1/PC3 is in agreement with earlier suggestions that expression of this gene is restricted to endocrine and neural tissue [4,6]. Furthermore, the lack of PC1/PC3 expression in most SCLC-V cell lines tested is in further support of this, since most of them display only limited neuroendocrine features, if any at all; the loss or absence of neuroendocrine features in SCLC-V cell lines is an established phenomenon [25]. Variant SCLC cell line NCI-H446 seemed to be an exception since it expressed PC1/PC3 but not the neuroendocrine markers 7B2 and GRP. We do not have a clear explanation for this apparent discrepancy yet, but we speculate that SCLC-V cell lines not invariably lose all their neuroendocrine features completely. In this context, the very low PC1/PC3 expression in the SCLC-V cell lines NCI-H526 and NL-SCLC3 might be explained in a similar manner, since NCI-H526 does express 7B2 but not GRP [25]. Finally, it is important to note here that of the two PC1/PC3 transcripts the one of 5 kb was the more abundant one in the SCLC-C cell lines. This is in contrast with results of PC1/PC3 expression studies in which other cell types were used [4,6,8].

Table I

Expression of the PC1/PC3 gene in human lung carcinoma cell lines

Lung cancer cells	Classification	PC1/PC3 expression
GLC-1-M13	SCLC-C	+
SCLC-21H	SCLC-C	+
SCLC-22H	SCLC-C	+
SCLC-16C	SCLC-C	+
NCI-H249	SCLC-C	+
GLC-1	SCLC-V	–
SCLC-16V	SCLC-V	–
NCI-H82	SCLC-V	–
NCI-N417	SCLC-V	–
NCI-H524	SCLC-V	–
NCI-H526	SCLC-V	–(*)
NCI-H446	SCLC-V	+
NL-SCLC3	SCLC-V	–(*)

+ Compares to expression levels of PC1/PC3 mRNAs as in Fig. 1, lane 1.

– No expression of PC1/PC3 gene, even not after prolonged exposure of the X-ray films.

(*) After prolonged (120 h) exposure of X-ray films a weak signal comparable to that in Fig. 1, lane 2, became visible.

We also studied the PC1/PC3 expression pattern in primary tumors of the lung (Table II). Six primary SCLCs, five carcinoid tumors of the lung and thirteen non-SCLCs were analyzed. Of the latter category, six tumors were classified as adenocarcinomas and seven as squamous cell carcinomas. In two of the primary SCLCs, PC1/PC3 expression was much higher than in the SCLC-C cell lines tested (for example, compare in Fig. 1, lanes 1 and 3). In two other SCLC tumor samples, PC1/PC3 expression appeared barely detectable. Finally, in two of the primary SCLCs, no PC1/PC3 expression could be demonstrated, not even upon prolonged exposure of the X-ray films. In contrast, four of the five carcinoid tumors of the lung appeared to exhibit very high levels of PC1/PC3 expression. In the fifth one, expression was lower but still comparable to the levels observed in the SCLC-C cell lines. All primary non-SCLCs were negative for PC1/PC3 expression. Like in the SCLC-C cell lines, both the 3 kb and 5 kb PC1/PC3 transcripts were expressed in primary SCLCs and carcinoid tumors of the lung, with the 5 kb mRNA always the more abundant.

Based upon the results of the expression studies, a primary carcinoid tumor of the lung was selected as mRNA source for experiments to obtain cDNAs corresponding to the 5 kb PC1/PC3 transcript. Using mRNA isolated from such a carcinoid tumor of the lung (tumor 10605), a cDNA library in λgt11 was constructed as described in section 2. Screening of about 150,000 plaques of this library with the 2.6 kb mouse PC1/PC3 cDNA probe resulted in the identification of about 50 positive bacteriophage clones. From these, clones were selected that hybridized to the 5' half as well as to the 3' half of the mouse PC1/PC3 cDNA of 2.6 kb. Six clones were identified on the basis of these selection criteria. They were plaque-purified and designated λJC1-λJC6. Upon isolation of DNA from these bacteriophage stocks, the insert DNAs were studied by

Table II

Expression of the PC1/PC3 gene in primary human lung tumors

Primary tumors	Tumors tested	Levels of PC1/PC3 expression				
		++++	++	+	+/–	–
SCLCs	5	0	2	0	2	2
non-SCLCs						
- adenocarcinomas	6	0	0	0	0	6
- squamous cell carcinomas	7	0	0	0	0	7
Carcinoid tumors	5	4	0	1	0	0

++++ Compares to expression levels of PC1/PC3 mRNAs as in Fig. 1, lanes 5 and 6.

++ Compares to expression levels of PC1/PC3 mRNAs as in Fig. 1, lane 3.

+ Compares to expression levels of PC1/PC3 mRNAs as in Fig. 1, lane 1.

– No expression of the PC1/PC3 gene.

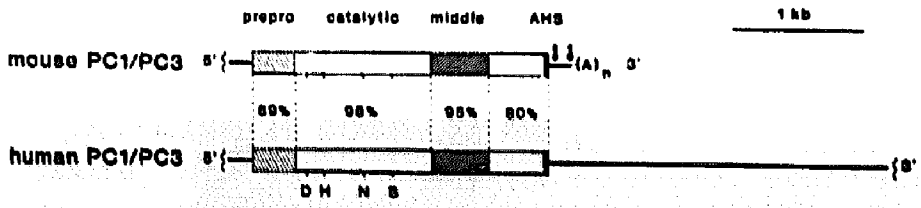


Fig. 3. Schematic representation of the human 5 kb and the mouse 3 kb transcripts of the PC1/PC3 gene and the corresponding deduced PC1/PC3 proteins. The coding regions are represented as boxes in which the various domains are indicated. Corresponding regions in both proteins are connected by dotted lines and percentages of amino acid sequence similarity are given. The carboxy-terminal amphipathic α-helical structures (AHS) are indicated as black boxes. Polyadenylation sites found in mouse PC1/PC3 transcripts [7] are indicated by arrows.

restriction enzyme analysis. λJC6 was selected for further studies, because it contained the largest insert. Restriction enzyme analysis revealed that the insert of λJC6 contained three internal *Eco*R1 sites; the four *Eco*R1 fragments were subcloned in pKUN19 or pGEM-3Zf(+) and the nucleotide sequence of these DNAs was determined. The order and orientation of the four *Eco*R1 fragments could be established by direct sequencing of the phage insert DNA. The nucleotide sequence data are shown in Fig. 2. The PC1/PC3 cDNA insert of λJC6 appeared to be 5037 nucleotides long. As it contained no polydA-tail, the corresponding transcript is presumably still a few hundred bases larger, including the poly(A)-tail. Analysis of the nucleotide sequence revealed an open reading frame for a protein of 753 amino acid residues; exactly the same number of residues as the protein deduced from the 3 kb mouse PC1/PC3 mRNA (7; see also 6,8) and with similar protein domains. In Fig. 3, we have indicated the 'prepro', catalytic and middle domain and the amphipathic α-helical structure (AHS). Overall amino acid sequence similarity between the predicted human and mouse PC1/PC3 proteins is about 92%. In the 'prepro' domain, the subtilisin-like catalytic domain, the middle domain and the remaining carboxy-terminal domain, the percentages are 89%, 98%, 95% and 80%, respectively (Fig. 3). The subtilisin-like domain of human PC1/PC3 contains five cysteine residues which are conserved in human furin and the prohormone processing enzymes encoded by the *KEX1* gene of *Kluyveromyces lactis* and the *KEX2* gene of *Saccharomyces cerevisiae*. In earlier computer-assisted molecular modelling studies, we have proposed a 3D model for human furin and the two yeast-processing enzymes and suggested that in furin two disulphide bridges may be formed and in the yeast enzymes three [15]. Based upon the sequence data presented here, we also predict two such disulphide bonds in human PC1/PC3. In contrast to the predicted amino acid sequence of mouse PC1/PC3 which contains three potential sites for *N*-glycosylation [6–8], that of human PC1/PC3 contains only two such potential sites. Both of these, NHT (173–175) and NLT (401–403) (Fig. 2), are located in the subtilisin-like catalytic domain. Like in mouse PC1/PC3 [6–8], human and mouse PC2 [4,5], human [3,11], mouse [10] and rat furin [12], the tripeptide consensus recognition sequence RGD, which is found in a number of extracellular matrix proteins and seems implicated in cell adhesion, is present in human PC1/PC3 at amino acid positions 518 to 520. The RGD sequence is not found in the two yeast processing enzymes.

In summary, we conclude from our studies that transcription of the human PC1/PC3 gene in lung tumor cells with neuroendocrine features results in the generation of two major transcripts; one of 3 kb and one of 5 kb, with a preference for the 5 kb species. Taking into account the published nucleotide sequence data of the 3 kb transcript of mouse PC1/PC3 and the nucleotide sequence data of human PC1/PC3 presented here, we conclude that the molecular basis for the size difference between the two PC1/PC3 mRNAs in both mammals is due to differences in the length of the 3' untranslated region and that both the 3 kb and 5 kb PC1/PC3 transcripts encode the same protein. Thus, the divergence observed in transcription of the PC1/PC3 gene does not generate a diversity in PC1/PC3 proprotein processing enzymes. Assuming that the genes for human and mouse PC1/PC3 are single copy genes [7], the 3 kb and 5 kb transcripts likely originate from the same gene. mRNAs transcribed from the same gene and varying in size due to divergence in length of their 3' untranslated region are not unusual and could affect mRNA stability in a positive way; it is conceivable that the higher levels of the 5 kb transcript in the human lung tumor cells as compared to the 3 kb species should be attributed to this, at least in part. Analysis of the rates according to which both transcripts are synthesized could be instrumental in resolving this matter more conclusively.

Acknowledgements: We thank G. Bepler, J. Broers, D. Carney and L. de Leij for providing the lung carcinoma cell lines used in our studies. We thank Sj. Sc. Wagenaar and E. Vaarkamp for providing the lung tumor specimens and D. Steiner for the mouse PC3 cDNA probe [6]. This work was supported by the Inter-University Network for Fundamental Research sponsored by the Belgian Government, by the 'Geconcerteerde Onderzoekacties 1992–1996' of the Belgian Government, by EC contract BIOT-CT91-0302 and by a grant from the 'Nationaal Fonds voor Wetenschappelijk Onderzoek' (NFWO, Levenslijn).

REFERENCES

[1] Roebroek, A.J.M., Schalken, J.A., Leunissen, J.A.M., Onnekink, C., Bloemers, H.P.J. and Van de Ven, W.J.M. (1986) EMBO J. 5, 2197–2202.

[2] Fuller, R.S., Brake, A.J. and Thorner, J. (1989) Science 246, 482–486.

[3] Van den Ouweland, A.M.W., Van Duijnhoven, J.L.P., Keizer, G.D., Dorssers, L.C.J. and Van de Ven, W.J.M. (1990) Nucl. Acids Res. 18, 664.

[4] Seidah, N.G., Gaspar, L., Mion, P., Marcinkiewicz, M., Mbikay, M. and Chrétien, M. (1990) DNA Cell Biol. 9, 415–424.

[5] Smeekens, S.P. and Steiner, D.F. (1990) J. Biol. Chem. 265, 2997–3000.

[6] Smeekens, S.P., Avruch, A.S., LaMendola, J., Chan, S.J. and Steiner, D.F. (1991) Proc. Natl. Acad. Sci. USA 88, 340–344.

[7] Seidah, N.G., Marcinkiewicz, M., Benjannet, S., Gaspar, L., Beaubien, G., Mattei, M.G., Lazure, C., Mbikay, M. and Chrétien, M. (1991) Mol. Endocrinol. 5, 111–122.

[8] Nakayama, K., Hosaka, M., Hatsuzawa, K. and Murakami, K. (1991) J. Biochem. 109, 803–806.

[9] Schalken, J.A., Roebroek, A.J.M., Oomen, P.P.C.A., Wagenaar, Sj.Sc., Debruyne, F.M.J., Bloemers, H.P.J. and Van de Ven, W.J.M. (1987) J. Clin. Invest. 80, 1545–1549.

[10] Hatsuzawa, K., Hosaka, M., Nakagawa, T., Nagase, M., Shoda, A., Murakami, K. and Nakayama, K. (1990) J. Biol. Chem. 265, 22075–22078.

[11] Barr, P.J., Mason, O.B., Landsberg, K.E., Wong, P.A., Kiefer, M.C. and Brake, A.J. (1991) DNA Cell Biol. 10, 319–328.

[12] Misumi, Y., Sohda, M. and Ikehara, Y. (1990) Nucleic Acids Res. 18, 6719.

[13] Roebroek, A.J.M., Pauli, I., Zhang, Y. and Van de Ven, W.J.M. (1991) FEBS Lett. 289, 133–137.

[14] Peters, K. and Rose, A. (1991) Worm Breed. Gaz. 11, 28.

[15] Van de Ven, W.J.M., Voorberg, J., Fontijn, R., Pannekoek, H., Van den Ouweland, A.M.W., Van Duijnhoven, J.L.P., Roebroek, A.J.M. and Siezen, R.J. (1990) Mol. Biol. Rep. 14, 265–275.

[16] Wise, R.J., Barr, P.J., Wong, P.A., Kiefer, M.C., Brake, A.J. and Kaufman, R.J. (1990) Proc. Natl. Acad. Sci. USA 87, 9378–9382.

[17] Bresnahan, P.A., Leduc, R., Thomas, L., Thorner, J., Gibson, H.L., Brake, A.J., Barr, P.J. and Thomas, G. (1990) J. Cell Biol. 111, 2851–2859.

[18] Misumi, Y., Oda, K., Fujiwara, T., Takami, N., Tashiro, K. and Ikehara, Y. (1991) J. Biol. Chem. 266, 16954–16959.

[19] Seidah, N.G., Mattei, M.G., Gaspar, L., Benjannet, S., Mbikay, M. and Chrétien, M. (1991) Genomics 11, 103–107.

[20] Benjannet, S., Rondeau, N., Day, R., Chrétien, M. and Seidah, N.G. (1991) Proc. Natl. Acad. Sci. USA 88, 3564–3568.

[21] Thomas, L., Leduc, R., Thorne, B., Smeekens, S.P., Steiner, D.F. and Thomas, G. (1991) Proc. Natl. Acad. Sci. USA 88, 5297–5301.

[22] Fuller, R.S., Brake, A.J. and Thorner, J. (1989) Proc. Natl. Acad. Sci. USA 86, 1434–1438.

[23] Wilcox, C.A. and Fuller, R.S. (1991) J. Cell Biol. 115, 297–307.

[24] Siezen, R.J., De Vos, W.M., Leunissen, J.A.M. and Dijkstra, B.W. (1991) Prot. Engineering 7, 719–731.

[25] Roebroek, A.J.M., Martens, G.J.M., Duits, A.J., Schalken, J.A., Van Bokhoven, A., Wagenaar, Sj.Sc. and Van de Ven, W.J.M. (1989) Cancer Res. 49, 4154–4158.

[26] World Health Organization (1982) Am. J. Clin. Pathol. 77, 123–136.

[27] Broers, J.L.V., Mijnheere, E.P., Klein Rot, M., Schaart, G., Sijlmans, A., Boerman, O.C. and Ramaekers, F.C.S. (1991) Cancer 67, 619–633.

[28] Auffray, C. and Rougeon, F. (1980) Eur. J. Biochem. 107, 303–314.

[29] Church, G.M. and Gilbert, W. (1984) Proc. Natl. Acad. Sci. USA 81, 1991–1995.