Gene, 163 (1995) 307–311

GENE 09082

Human peptidylglycine α-amidating monooxygenase transcripts derived by alternative mRNA splicing of an unreported exon

(Recombinant DNA; PCR; alternative splicing; PAM; PAL; PHM; lyase)

Michele D. Vos, John E. Jones and Anthony M. Treston

Intervention Section, Biomarkers and Prevention Research Branch, National Cancer Institute Division of Cancer Prevention and Control, Rockville, MD 20850-3300, USA

Received by J.L. Slightom: 30 January 1995; Revised/Accepted: 21 April/28 April 1995; Received at publishers: 22 May 1995

SUMMARY

We are characterizing the alternatively spliced human peptidylglycine α-amidating monooxygenase (hPAM)-encoding mRNA transcripts expressed by human cells. Reverse transcription coupled to the polymerase chain reaction (RT-PCR) has been used to identify four alternatively spliced variants that differ in the region joining the two catalytic domains. Two of the transcripts represent previously reported splice variants differentiated by the presence (h*PAM-A*) or absence (h*PAM-B*) of a 321-nucleotide (nt) linker (optional exon A) which in the rat produce functionally distinct enzymes. Different mRNAs represent two splice variants, h*PAM-C* and h*PAM-D*, that show the presence of an exon unreported for *PAM* in any other species. This new exon, designated exon C, is 54 nt in length, encodes an 18-amino-acid (aa) peptide containing a conserved dibasic aa endoproteolytic processing motif, and is located 3′ of exon A in human genomic DNA. We propose that cell-specific regulation of mRNA splicing would provide a mechanism for control of prohormone activation by these variants of the PAM enzyme.

INTRODUCTION

C-terminal amidation is a crucial posttranslational modification, required by half of all known gastrointestinal and neural peptides for full biological potency, accomplished by peptidyl-glycine α-amidating monooxygenase (PAM) (Mains et al., 1990; Eipper et al., 1992). PAM is encoded by a single gene with two catalytic activities, peptidylglycine α-hydroxylating monooxygenase (PHM) and peptidyl α-hydroxyglycine α-amidating lyase (PAL) (Perkins et al., 1990) (Fig. 1a). In human tumor cells both PHM and PAL are required for synthesis of α-amidated peptide growth hormones (Quinn et al., 1991; Treston et al., 1993).

In the rat, it has recently been shown that alternative mRNA splicing of exons in the 3′ region of the precursor and in the linker region between the catalytic domains generates tissue-specific forms with marked biochemical differences (at least eight mRNA forms) (Kato et al., 1990; Eipper et al., 1992). Two alternatively spliced forms of human *PAM* (h*PAM*) identified have been shown to differ by optional splicing of the 321 nt exon A linking the PHM and PAL domains (Glauder et al., 1990, h*PAM-A* and *B*) and a third form lacks the transmembrane domain, exon B (*PAM-15*; Tateishi et al., 1994). Studies with rat PAM (rPAM) demonstrate that linker exon A directs posttranslational glycosylation and sulfation at

Correspondence to: Dr. A.M. Treston, BPRB, DCPC, NCI, 9610 Medical Center Drive, Suite 300, Rockville, MD 20850-3300, USA. Tel. (1-301) 402-3128, ext. 312; Fax (1-301) 402-4422; e-mail: trestona@bprb.nci.nih.gov

Abbreviations: aa, amino acid(s); bp, base pair(s); cDNA, DNA complementary to RNA; EtdBr, ethidium bromide; h, human; r, rat; kb, kilobase(s) or 1000 bp; nt, nucleotide(s); PAL, peptidyl α-hydroxyglycine α-amidating lyase; *PAM*, gene (DNA, RNA) encoding PAM; PAM, peptidylglycine α-amidating monooxygenase (EC 1.14.17.3); PCR, polymerase chain reaction; PHM, peptidylglycine α-hydroxylating monooxygenase; RT, reverse transcription.

SSDI 0378-1119(95)00364-9

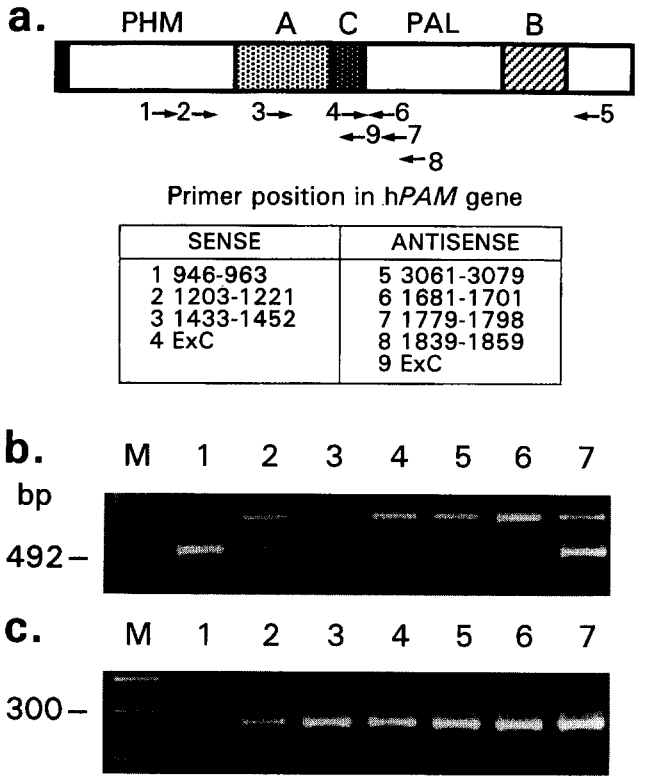

Fig. 1. PCR strategy for the identification and analysis of the linker region of h*PAM* mRNA transcripts by RT-PCR. (**a**) The two enzymatically active domains (PHM and PAL) are indicated. The light stippled box denotes exon A, the dark stippled box exon C and the hatched box exon B (not to scale). The blackened box denotes the signal peptide. All primers used in this study are shown; bp numbers correspond to the cDNA sequence of human *PAM* numbered by Glauder et al. (1990), lettered abbreviations are for new exon C. (**b** and **c**) RT-PCR analysis of *PAM* mRNAs. Lanes: **1**, GH3; **2**, AtT20; **3**, HTB103; **4**, HTB11; **5**, H1011; **6**, H889; **7**, H157. (**b**) Primer pair 1 and 7 amplified an 852-bp product representing h*PAM-A* and a 531-bp product, h*PAM-B*. The previously unidentified 585-bp product represents a form (h*PAM-C*) that lacks optional exon A but contains the new 54-nt sequence designated optional exon C. (**c**) Primers to human cyclophilin were used to confirm intact mRNA, the expected 254-bp product was demonstrated in all cell lines. The negative RNA control (water) and positive RNA control (*PAM*-non-expressing cell line, HTB103, lane 3) were negative. **Methods:** Total RNA was prepared from each cell line, DNase treated and reversed transcribed into cDNA using random hexamers and SUPERSCRIPT II RNase H$^-$ (Gibco-BRL, Gaithersburg, MD, USA). cDNA (1 µl) and 100 ng of synthesized primers were used in the PCR. PCR was carried out in the Perkin-Elmer 9600 thermal cycler. PCR products were analyzed on Tris/borate/EDTA-20% polyacrylamide gels (NOVEX, San Diego, CA, USA) and stained with EtdBr.

several sites and contains a conserved dibasic site required for proteolytic separation of rPAM into monofunctional rat PHM and PAL with an altered pH profile and increased K_m for lipophilic substrates (Yun et al., 1994).

We are in the process of identifying alternatively spliced forms of hPAM in order to characterize their biochemistry and enzymatic activities, as well as their pattern of expression in human normal and tumor cells.

EXPERIMENTAL AND DISCUSSION

(a) Identification of variant h*PAM* transcripts

Preliminary analysis of several human tumor cell lines indicated that, while splicing within the exon B region of the *PAM* primary transcript is compatible with reported forms for other species, splicing within the exon A region resulted in unexpected forms. To identify these forms we analyzed cDNA products obtained by RT-PCR of total RNA from human and rodent tumor cell lines. Primers used for the amplification (primer pairs 1 and 7, 2 and 7, 2 and 8) targeted the region linking the two catalytic domains of the PAM precursor (Fig. 1a). Polyacrylamide-gel electrophoresis revealed three major bands of varying intensities in all the human cell lines, and the predicted two bands in the rodent cell lines (Fig. 1b). Based on previously reported data (Glauder et al., 1990), we interpreted the highest molecular size band (852 bp) as corresponding to h*PAM-A*, the bifunctional form of *PAM* containing the PHM-PAL linker region, and the smallest band (531 bp) as representing h*PAM-B*, the form of *PAM* which deletes the 321 nt segment separating the two catalytic domains. This deletion compares in length (315 bp) and location to rat optional exon A (rat genomic exon 16) and will be referred to as such in this paper. Using both UV detection and Southern analysis the third band (585 bp) appeared only in the human cell lines and did not coincide with any splice variant reported for the *PAM* gene in any species analyzed to date.

(b) Analysis of amplified DNAs by cloning and sequencing

We cloned and sequenced each of the h*PAM* cDNA PCR amplification products generated from primer pair 2 and 8 using H889 and H157 cDNA (Fig. 2a). Sequencing data verified our identification of Glauder's forms h*PAM-A* and h*PAM-B*. No nt substitutions to the reported sequences were seen. The cDNA sequence of the unknown band from both H889 and H157 was identical to that of h*PAM-B* except for a 54-nt segment inserted within the region occupied by optional exon A in h*PAM-A* (Fig. 2a). This new sequence maintains the translational reading frame of the PAM protein, utilizes the same donor and acceptor splice sites as optional exon A, and encodes an 18-aa peptide (Fig. 2b). These results demonstrate the existence of a human mRNA transcript not previously described for the *PAM* gene. We have named the new form and sequence h*PAM-C* and exon C, respectively. The intensity of the bands seen in Fig. 1b suggests that the relative level of the new form is similar to that of h*PAM-A* or h*PAM-B* in different cell lines. PCR analysis using primer pair 4 and 5 (Fig. 1a, data not shown)

a.

```
1203 AGAACCATACCACCAGAGGCCAACATTCCAATTCCCGTGAAGTCTGATATGGTTATGATGCATGAACATCATAAAGAAACAGAATATAAAGATAAGATTCCTTTACTACAGCAGCCAAAA hPAM-A
     AGAACCATACCACCAGAGGCCAACATTCCAATTCCCGTGAAGTCTGATATGGTTATGATGCATGAACATCATAAAGAAACAGAATATAAAGATAAGATTCCTTTACTACAGCAGCCAAAA hPAM-B
     AGAACCATACCACCAGAGGCCAACATTCCAATTCCCGTGAAGTCTGATATGGTTATGATGCATGAACATCATAAAGAAACAGAATATAAAGATAAGATTCCTTTACTACAGCAGCCAAAA hPAM-C

1323 CGAGAAGAAGAAGAAGTGTTAGACCAGGGTGATTTCTATTCACTACTTTCCAAGCTGCTAGGAGAAAGGGAAGATGTTGTTCATGTGCACAAATATAATCCTACAGAAAAGGCAGAATCA hPAM-A
     CGAGAAGAAGAAGAAGTGTTAGACCAGG............................................................................................ hPAM-B
     CGAGAAGAAGAAGAAGTGTTAGACCAGG............................................................................................ hPAM-C
                                                                                                                   GGCAGAATCA hPAM-D

1443 GAGTCAGACCTGGTAGCTGAGATTGCAAATGTAGTCCAAAAAAAGGATCTTGGTCGATCTGATGCCAGAGAGGGTGCAGAACATGAGAGGGGTAATGCTATTCTTGTCAGAGACAGAATT hPAM-A
     ........................................................................................................................ hPAM-B
     ........................................................................................................................ hPAM-C
     GAGTCAGACCTGGTAGCTGAGATTGCAAATGTAGTCCAAAAAAAGGATCTTGGTCGATCTGATGCCAGAGAGGGTGCAGAACATGAGAGGGGTAATGCTATTCTTGTCAGAGACAGAATT hPAM-D

1563 CACAAATTCCACAGACTAGTATCTACCTTGAGGCCACCAGAGAGCAGAGTTTTCTCATTACAGCAGCCCCCACCTGGTGAAGGCACCTGGGAACCAGAACACACAGGAG........... hPAM-A
     ........................................................................................................................ hPAM-B
     .........................................................................................................CTAGAAGAGAA hPAM-C
     CACAAATTCCACAGACTAGTATCTACCTTGAGGCCACCAGAGAGCAGAGTTTTCTCATTACAGCAGCCCCCACCTGGTGAAGGCACCTGGGAACCAGAACACACAGGAGCTAGAAGAGAA hPAM-D

1672 .........................................ATTTCCACATGGAAGAGGCACTGGATTGGCCTGGAGTATACTTGTTACCAGGCCAGGTTTCTGGGGTGGCTCTAGAC hPAM-A
     .........................................ATTTCCACATGGAAGAGGCACTGGATTGGCCTGGAGTATACTTGTTACCAGGCCAGGTTTCTGGGGTGGCTCTAGAC hPAM-B
     GATATGAAATGTTTCCAGTGCAAAGAAATGATAAGTGTTCAAGATTTCCACATGGAAGAGGCACTGGATTGGCCTGGAGTATACTTGTTACCAGGCCAGGTTTCTGGGGTGGCTCTAGAC hPAM-C
     GATATGAAATGTTTCCAGTGCAAAGAAATGATAAGTGTTCAAG                                                                                 hPAM-D

1749 CCTAAGAATAACCTGGTGATTTTCCACAGAGGTGACCATGTCTGGGATGGAAACTCGTTTGACAGCAAGTTTGTTTACCAGCAAATAGGACTCGGACCAATTGAAGAAGAC        hPAM-A
     CCTAAGAATAACCTGGTGATTTTCCACAGAGGTGACCATGTCTGGGATGGAAACTCGTTTGACAGCAAGTTTGTTTACCAGCAAATAGGACTCGGACCAATTGAAGAAGAC        hPAM-B
     CCTAAGAATAACCTGGTGATTTTCCACAGAGGTGACCATGTCTGGGATGGAAACTCGTTTGACAGCAAGTTTGTTTACCAGCAAATAGGACTCGGACCAATTGAAGAAGAC        hPAM-C
```

b.

```
339 RTIPPEANIPIPVKSDMVMMHEHHKETEYKDKIPLLQQPKREEEEVLDQGDFYSLLSKLLGEREDVVHVHKYNPTEKAESESDLVAEIANVVQKKDLGRSDAREGAEHERGNAILVRDRI hPAM-A
    RTIPPEANIPIPVKSDMVMMHEHHKETEYKDKIPLLQQPKREEEEVLDQ....................................................................... hPAM-B
    RTIPPEANIPIPVKSDMVMMHEHHKETEYKDKIPLLQQPKREEEEVLDQ....................................................................... hPAM-C
                                                                              TEKAESESDLVAEIANVVQKKDLGRSDAREGAEHERGNAILVRDRI hPAM-D
                                        ▲▲                                                       ▲▲

459 HKFHRLLSTLRPPESRVFSLQQPPPGEGTWEPEHTG.................DFHMEEALDWPGVYLLPGQVSGVALDPKNNLVIFHRGDHVWDGNSFDSKFVYQQIGLGPIEED    hPAM-A
    .....................................................DFHMEEALDWPGVYLLPGQVSGVALDPKNNLVIFHRGDHVWDGNSFDSKFVYQQIGLGPIEED    hPAM-B
    ....................................ARREDMKCFQCKEMISVQDFHMEEALDWPGVYLLPGQVSGVALDPKNNLVIFHRGDHVWDGNSFDSKFVYQQIGLGPIEED    hPAM-C
    HKFHRLLSTLRPPESRVFSLQQPPPGEGTWEPEHTGARREDMKCFQCKEMISVQ                                                                  hPAM-D
                                         ▲▲      ^   ^
```

Fig. 2. Comparison of the nt and aa sequences of h*PAM* transcripts. (**a**) Primers used in the cloning reactions are bolded. h*PAM-A* and h*PAM-B* are identical to the forms reported by Glauder et al. (1990). h*PAM-C* and h*PAM-D* represent mRNA transcripts not previously reported for this gene. These sequences have been submitted to GenBank under accession Nos. U19965 and U19966, respectively. (**b**) The deduced aa sequence of the linker and flanking regions of the four forms of the PAM proenzyme; the dibasic site located upstream from the linker region is marked by ▲▲; pairs of basic aa located within the linker regions separating the PHM and PAL domains are bolded and marked by ▲▲; Cys residues in exon C are designated by ^. The dibasic site shown in hPAM-A, Lys[432]-Lys[433], is cleaved in both rat and bovine PAMs to separate the PHM and PAL catalytic domains. The Lys[378]-Arg[379] dibasic site upstream from the linker region is not cleaved in bovine or rat PAM. **Methods:** PCR products were directly ligated into the pCRII vector (Invitrogen, San Diego, CA, USA) followed by transformation into competent *E. coli* cells. Clones containing the inserts of interest were sequenced using the dsDNA Cycle Sequencing System (Gibco-BRL) and on the Applied Biosystems Model 373 DNA Sequencing System using the *Taq* DyeDeoxy™ Terminator Cycle Sequencing Kit. Sequence alignments were performed using Lasergene software by DNASTAR (Madison, WI, USA).

indicates that the new form continues 5′ to at least nt 3079 and so is not related to r*PAM-4* which contains part of a retained intron after exon 16 including a stop codon, resulting in no PAL domain in the translated protein (Eipper et al., 1992).

(c) Genomic organization of exons A and C, and identification of a transcript containing both exons A and C

There is nothing known about the exon/intron organization of the h*PAM* gene. In the r*PAM* gene, exon 16 (optional exon A) is flanked by 11.8 kb of 5′ intronic DNA and >23 kb of 3′ intronic DNA (Fig. 3a). No significance has been ascribed to these introns in the rat, but in humans one of these introns must include exon C. We followed two approaches to determine the placement of exon C in relation to exon A. Firstly, we performed PCR amplification using human genomic DNA. Primer pair 4 and 6 were used to span the region 3′ of exon C and 5′ of the human equivalent of rat exon 17 (Fig. 3a). If the human genomic structure is similar to that of the rat, and if exon C was positioned 5′ of exon A, then no product would be generated (predicted size >35 kb). However, this analysis produced a 2.8-kb amplification product which was cloned and partially sequenced (Fig. 3b). This product contains sequence from both exon C and exon 17. The splice junctions located at the 3′ end of exon C and 5′ end of exon 17 display homology to both donor and acceptor consensus sequences. We are currently in the process of determining the configuration of the 5′ acceptor splice junction of exon C.

Secondly, we used primer pair 3 and 7 to amplify cDNA of the expected size for h*PAM-A*. Gel analysis of the amplification product showed two bands, one of 365 bp representing h*PAM-A* and another of 419 bp

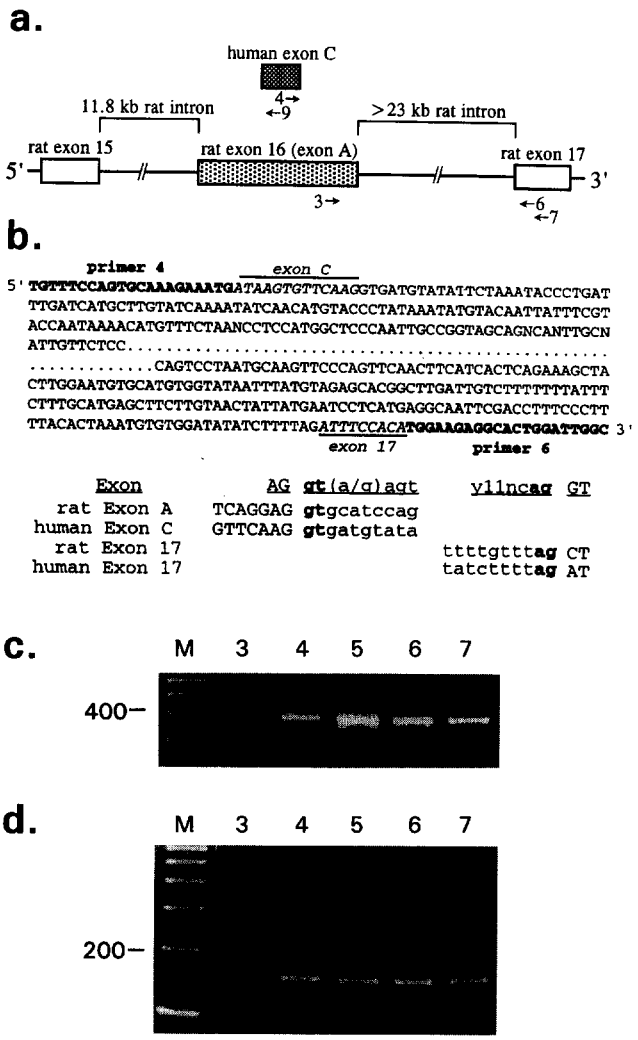

Fig. 3. PCR amplification of human genomic DNA containing exon C and identification of exon/intron boundaries. (**a**) Diagram of r*PAM* genomic DNA organization, not to scale. (**b**) DNA sequence analysis of the 2.8-kb PCR product of primer pair 4 and 6 (bolded) demonstrates that exon C is 2.8-kb 5′ of the human equivalent of rat exon 17. The sequence has been submitted to GenBank under accession Nos. U20089 and U20090. Exon C maintains a characteristic 3′ splice junction and utilizes the same 5′ acceptor splice site as h*PAM-A* and *B*. The nt which are italicized represent exonic sequences which are not included in the amplification primers. (**c**) Primer pair 3 and 7 amplified the expected 365-bp cDNA product for form h*PAM-A* and in addition a 419-bp product representing h*PAM-D*. (**d**) Primer pair 4 and 7 amplified a single 160-bp product representing h*PAM-C* and h*PAM-D* that can only result from exon C being positioned 3′ of exon A. Cell line mRNAs are as described in the legend to Fig. 1. **Methods:** Human genomic (2 μg) DNA (Promega, Madison, WI, USA) was amplified by optimizing PCR conditions using the Optitaq System (BIOS, New Haven, CT, USA). PCR products were cloned and sequenced using methods previously described.

(Fig. 3c). The implied sequence of exons A and C from the human genomic analysis, assuming similar organization to the rat, suggested that the unidentified cDNA product represents a form containing both exon A and exon C. To confirm this possibility we used cDNA from NCI-H889 and H157 and primer pair 3 and 9 to amplify and clone a 293-bp product, consistent with that predicted for a form containing part of both exons. Sequence analysis performed on this product conclusively shows the existence of a mRNA transcript which contains the 5′ end of exon A linked to the 3′ end of exon C (Fig. 2a). We named this new *PAM* form h*PAM-D*. Consistent with these findings, analysis of the cDNA amplification product generated using primer pair 4 and 7 revealed a single 160-bp product (Fig. 3d) representing a transcript that can only result from exon C being positioned immediately 5′ of the human equivalent of rat exon 17 in both h*PAM-C* and h*PAM-D*. In this analysis any form with exon A 3′ of exon C would have resulted in a product of 481 bp (Fig. 3d). The relative intensities in Figs. 1b and 3c demonstrate that h*PAM-D* is a quantitatively minor form, however it serves to confirm a genomic organization in which exon A is 5′ of exon C. The sequences of both exons maintain the translational reading frame of the hPAM-D protein (Fig. 2b).

(d) Conclusions

(*1*) We have analyzed approx. 75% of the coding sequence for h*PAM* to characterize alternatively spliced mRNAs. Amplification products generated were analyzed by nucleotide sequence determination or size fractionation and found to be in accordance with the published sequence with the exception of the novel exon reported here.

(*2*) We have identified two alternatively spliced transcripts of the human *PAM* gene, h*PAM-C* and h*PAM-D*, that maintain sequence consistent with that of previously reported forms (h*PAM-A* and h*PAM-B*) except for the inclusion of an apparently human-specific 54-nt exon.

(*3*) Exon C contains a dibasic site (Arg-Arg) for processing of hPAM-C to monofunctional PHM and PAL proteins (Fig. 2b). Both exons A and C are linker regions separating the monooxygenase and lyase domains. Exon A, which is highly conserved among species, contains a paired basic aa site (Lys^{432}-Lys^{433}, Fig. 2) that is utilized by endoproteases during posttranslational processing of bifunctional rPAM to yield monofunctional PHM and PAL proteins. The bifunctional and monofunctional enzymes have distinct pH activity profiles and K_m. To separate the two catalytic domains, the dibasic aa sites present in exons A and C must be exposed to the endoprotease. Computer assisted comparison of hydrophilicity (Kyte-Doolittle) and surface probability (Emini) algorithms for exons A and C revealed that the paired basic aa residues within both these exons are hydrophilic and charged, implying surface localization.

(*4*) The peptide encoded by exon C confers a variety of new features to the biochemistry of the PAM protein, for example, the two closely located Cys residues in the

middle of exon C can modify the conformation of the protein in this region by formation of a disulfide bridge. Further, the glycosylation sites present within exon A peptide are not conserved in exon C (Yun et al., 1994). In order to characterize the biochemistry of the new forms we have initiated studies to express the proteins encoded by these variants.

REFERENCES

Eipper, B.A., Stoffers, D.A. and Mains, R.E.: The biosynthesis of neuropeptides: Peptide α-amidation. Annu. Rev. Neurosci. 15 (1992) 57–85.

Eipper, B.A., Green, C.B.-R., Campbell, T.A., Stoffers, D.A., Keutmann, H.T., Mains, R.E. and Ouafik, L'H.: Alternative splicing and endoproteolytic processing generate tissue-specific forms of pituitary peptidylglycine α-amidating monooxygenase (PAM). J. Biol. Chem. 267 (1992) 4008–4015.

Glauder, J., Ragg, H., Rauch, J. and Engels, J.W.: Human peptidylglycine α-amidating monooxygenase: cDNA, cloning and functional expression of truncated form in COS cells. Biochem. Biophys. Res. Commun. 169 (1990) 551–558.

Mains, R.E., Dickerson, I.M., May, V., Stoffers, D.A., Perkins, S.N., Ouafik, L'H., Husten, E.J. and Eipper, B.A.: Cellular and molecular aspects of peptide hormone biosynthesis. Front. Neuroendocrinol. 11 (1990) 52–89.

Perkins, S.N., Husten, E.J. and Eipper, B.A.: The 108-kDa peptidylglycine α-amidating monooxygenase precursor contains two separable enzymatic activities involved in peptide amidation. Biochem. Biophys. Res. Commun. 171 (1990) 926–932.

Quinn, K., Treston, A.M., Scott, F.M., Kasprzyk, P.G., Avis, I., Siegfried, J.M., Mulshine, J.L. and Cuttitta, F.: α-Amidation of peptide hormones in lung cancer. Cancer Cells 3 (1991) 504–501.

Tateishi, K., Arakawa, F., Misumi, Y., Treston, A.M., Vos, M. and Matsuoka, Y.: Isolation and functional expression of human pancreatic peptidylglycine α-amidating monooxygenase. Biochem. Biophys. Res. Commun. 205 (1994) 282–290.

Treston, A.M., Scott, F.M., Vos, M., Iwai, N., Mains, R.W., Eipper, B.A., Cuttitta, F. and Mulshine, J.L.: Biochemical characterization of peptide α-amidation enzyme activities of human neuroendocrine lung cancer cell lines. Cell Growth Diff. 4 (1993) 911–920.

Yun, H.Y., Keutmann, H.T. and Eipper, B.A.: Alternative splicing governs sulfation of tyrosine or oligosaccharide on peptidylglycine α-amidating monooxygenase. J. Biol. Chem. 269 (1994) 10946–10955.