Volume 231, number 2, 347–351 FEB 05795 April 1988

Primary structure of sheep prostaglandin endoperoxide synthase deduced from cDNA sequence

Chieko Yokoyama, Toshiyuki Takai and Tadashi Tanabe

Laboratory of Cell Biology, National Cardiovascular Center Research Institute, Fujishiro-dai, Suita, Osaka 565, Japan

Received 1 February 1988

The complete amino acid sequence of prostaglandin endoperoxide synthase from sheep vesicular gland has been deduced by cloning and sequence analysis of DNA complementary to its messenger RNA. The results were confirmed by digestion of the enzyme with carboxypeptidase Y and by automated Edman degradation of the intact enzyme polypeptide and peptide fragments obtained by limited proteolysis of the enzyme with *Achromobacter* proteinase I. Mature sheep prostaglandin endoperoxide synthase is shown to be composed of 576 amino acids with an M_r of 66 175. The precursor peptide is predicted to contain a 24-residue signal peptide. The serine residue susceptible to acetylation by aspirin is found to be located near the C-terminus of the enzyme polypeptide.

Prostaglandin endoperoxide synthase; Cyclooxygenase; cDNA cloning; Nucleotide sequence; Primary structure; (Sheep vesicular gland)

1. INTRODUCTION

Prostaglandins (PGs) and thromboxane are potent biological mediators. The initial step of their biosyntheses is catalyzed by PG endoperoxide synthase. The enzyme has been purified from bovine [1] and sheep vesicular gland [2,3]. The enzyme is a membrane-bound glycoprotein with an M_r of 67000–72000 and exhibits both activities of fatty acid cyclooxygenase (arachidonic acid $\longrightarrow$ PGG_2) and PG hydroperoxidase ($PGG_2 \longrightarrow PGH_2$) [1–3]. Aspirin inhibits the cyclooxygenase activity by acetylating a single serine residue of the enzyme. The peptic fragment of the enzyme containing the acetylated serine residue was isolated from the enzyme inactivated by aspirin and the amino acid sequence was determined [4,5]. In order to elucidate the reaction mechanism of the enzyme it is important to determine its primary structure. Here, we describe the cloning of cDNA encoding PG endoperoxide synthase from sheep vesicular gland and the complete amino acid sequence of the enzyme deduced from the cDNA sequence.

Correspondence address: T. Tanabe, Laboratory of Cell Biology, National Cardiovascular Center Research Institute, Fujishiro-dai, Suita, Osaka 565, Japan

The nucleotide sequence presented here has been submitted to the EMBL/GenBank database under the accession number Y00750

2. MATERIALS AND METHODS

Partially purified PG endoperoxide synthase was prepared from sheep vesicular glands using DEAE-Cellulofine (Chisso, Tokyo) as described in [1,2], and subjected to preparative 7% SDS-polyacrylamide gel electrophoresis for further purification [6]. The purified enzyme (100 μg) which was pyridylethylated according to [7] was incubated with 5 μg *Achromobacter* proteinase I in 50 mM Tris-HCl buffer, pH 8.0, containing 0.5% SDS and 1 mM EDTA at 37°C for 12 h. The polypeptide fragments were separated by reverse-phase high-performance liquid chromatography with a 0–50% linear gradient of acetonitrile in 0.1% trifluoroacetic acid as described [8]. The C-terminal amino acid of the enzyme was determined by digesting it with carboxypeptidase Y according to [9]. The amino acid sequence and amino acid analyses were carried out as described in [8].

Total and poly(A)$^+$ RNAs were prepared from sheep vascular glands as described in [10]. 24- and 21-mer oligonucleotides, 5′-$^{A}_{G}$TATTCIGG$^{A}_{G}$CTT$^{T}_{C}$CIATIGG$^{A}_{G}$TT-3′ (oligonucleotide 1, I denotes inosine) and 5′-A$^{A}_{G}$AIGGIGCICCCATT$^{T}_{C}$CA$^{A}_{G}$T-3′ (oligonucleotide 2) were synthesized on the basis of the C- and N-terminal portions (in one-letter code, NPIESPEY and IEMGAPF, respectively) of the amino acid sequence of the aspirin-acetylation site [5]. cDNA libraries were constructed

Published by Elsevier Science Publishers B.V. (Biomedical Division)
00145793/88/$3.50

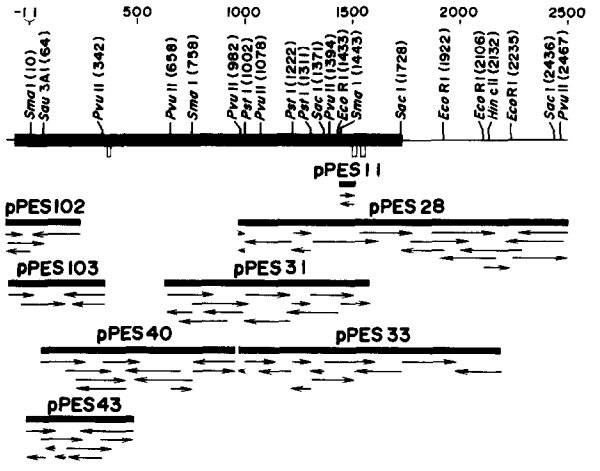


Fig.1. Restriction map and sequencing strategy for the cDNA clones. The restriction map shows only the relevant restriction sites, which are identified by numbers indicating the 5′-terminal nucleotide generated by cleavage. The poly(dA)·poly(dT) tract and poly(dG)·poly(dC) tails are not included in the restriction map. The protein coding region is indicated by a closed box, the regions used for synthesis of oligonucleotides by open boxes. The direction and extent of sequencing are illustrated by horizontal arrows under each clone used.

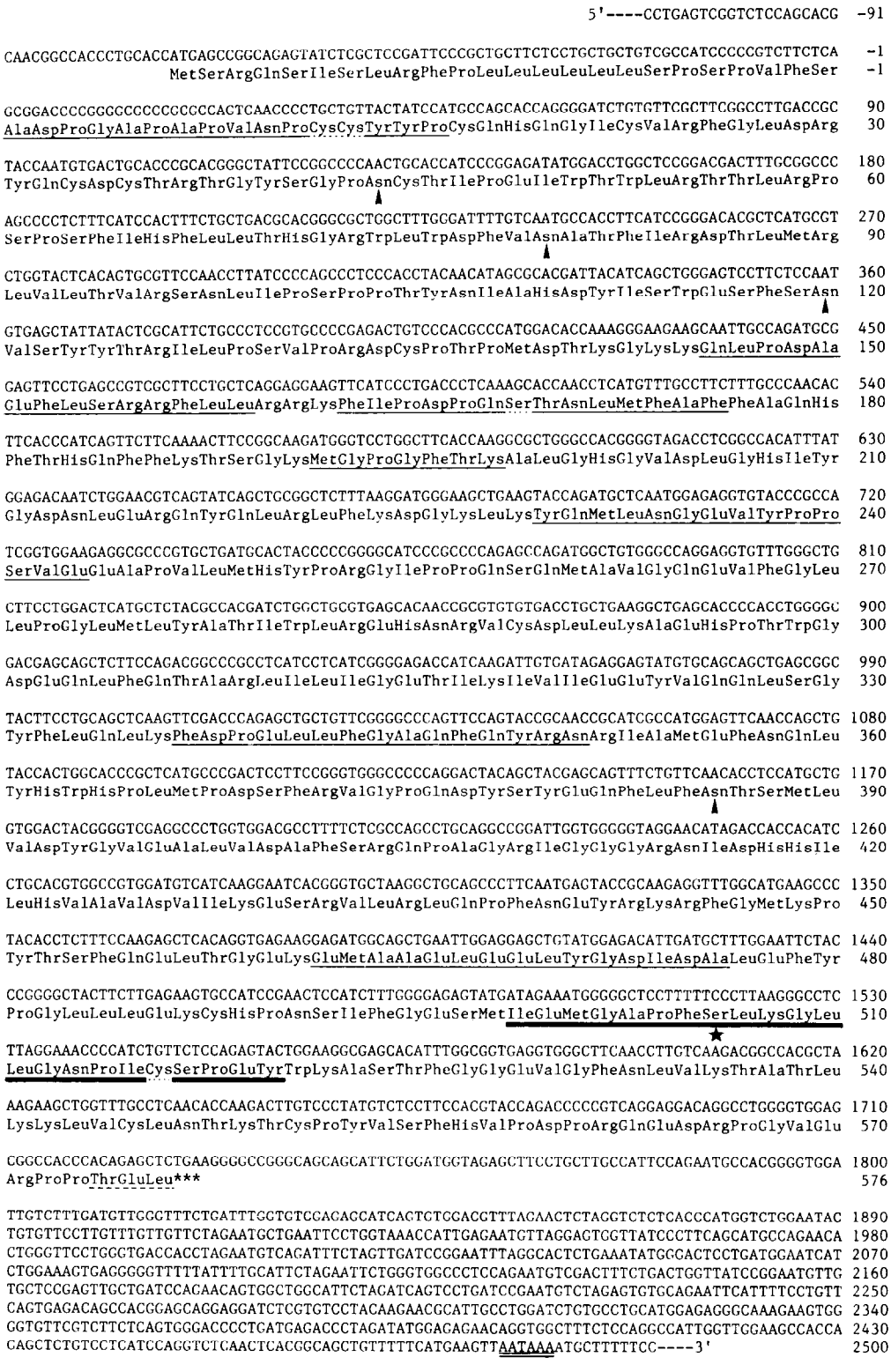
with oligo(dT)$_{12-18}$ or synthetic oligonucleotides as a primer using poly(A)$^+$ RNA from sheep vesicular glands as a template according to [11]. Transformation and screening procedures were as in [12]. Nucleotide sequence was determined by the dideoxy chain-termination method [13] as described [14]. RNA blot hybridization analysis was carried out according to [15]. Hydropathicity analysis using the procedure of Kyte and Doolittle [16] and prediction of protein secondary structure by Chou-Fasman procedure [17] were performed with the aid of the SDC-GENETYX program (Software Development, Tokyo).

3. RESULTS AND DISCUSSION

The N-terminal amino acid sequence of the purified PG endoperoxide synthase from sheep vesicular gland was the same as that of the previous report [4] (see fig.2). The enzyme polypeptide was also digested with *Achromobacter* proteinase I and the fragments were examined for amino acid sequence (see fig.2).

A cDNA library was constructed using poly(A)$^+$ RNA from sheep vesicular glands as a template and the oligonucleotide 1 as a primer. Screening of about 2×10^5 transformants from the cDNA library with the oligonucleotide 2 led to the selection of 3 positive clones including pPES11 (fig.1). Inserts of these clones contained the same nucleotide sequence. Using pPES11 as a probe for cloning longer cDNA sequences, clones pPES28, pPES31, pPES33, pPES40 and pPES43 were obtained from the oligo(dT)- and oligonucleotide 1-primed cDNA libraries. For isolation of clones containing the 5′-terminal region of cDNA, a synthetic oligonucleotide, 5′-AGTATAATAGCTCA-CAT-3′, complementary to the nucleotide sequence 359–375 was prepared and elongated using poly(A)$^+$ RNA from sheep vesicular glands as a template. Screening of 1×10^4 transformants from the primer-extension library with the 5′-terminal *Sau*3AI fragment of pPES43 insert yielded two clones, pPES102 and pPES103. DNA sequencing was performed on the inserts of 8 cDNA clones (fig.1), which covered a 2611-nucleotide sequence of cDNA for PG endoperoxide synthase from sheep vesicular gland.

Fig.2. Nucleotide sequence of cloned cDNA for sheep vesicular gland prostaglandin endoperoxide synthase and deduced amino acid sequence of the protein. Nucleotide residues are numbered beginning with the first residue of the GCG codon encoding the alanine of the N-terminus. The deduced amino acid sequence is shown below the nucleotide sequence and amino acid residues are numbered beginning with the N-terminal alanine. Fine underlining indicates amino acid sequences determined by Edman degradation. Ambiguously identified residues within a sequenced peptide are indicated by a dotted line. Broken underlining indicates C-terminal residues determined by carboxypeptidase Y digestion. The serine residue susceptible to acetylation by aspirin and the amino acid sequence of the peptic peptide containing the serine residue are indicated by an asterisk and thick underlining, respectively [5]. The potential asparagine-linked glycosylation sites are indicated by arrowheads. The nucleotide differences observed among the individual clones are as follows: G (pPES102 and 103) or A (pPES43) at nucleotide residues 3 and 218; C (pPES102 and 103) or A (pPES43) at 9; G (pPES103 and 43) or A (pPES102) at 165; A (pPES40) or G (pPES43) at 419; A (pPES40) or G (pPES31) at 819; G (pPES40) or A (pPES31) at 846; G (pPES33 and 28) or A (pPES31) at 1188; G (pPES33 and 31) or A (pPES28) at 1295; A (pPES33) or G (pPES31 and 28) at 1302; G (pPES33 and 31), A (pPES28) or C (pPES11) at 1486; G (pPES33, 28, and 11) or A (pPES31) at 1488; G (pPES33, 31 and 11) or A (pPES28) at 1503 and 1506; A (pPES33) or G (pPES28) at 1674; G (pPES33) or A (pPES28) at 1799 and 2028. The resulting amino acid substitutions are as follows: Arg or His at amino acid residue 73; Asp or Gly at 140; Arg or Gln at 432; Glu, Gln or Lys at 496; Met or Ile at 501. Wherever a nucleotide or amino acid difference occurs, the first named residue is shown in the sequences presented.


5'----CCTGAGTCGGTCTCCAGCACG -91
CAACGGCCACCCTGCACCATGAGCCGGCAGAGTATCTCGCTCCGATTCCCGCTGCTTCTCCTGCTGCTGTCGCCATCCCCCGTCTTCTCA -1
MetSerArgGlnSerIleSerLeuArgPheProLeuLeuLeuLeuLeuSerProSerProValPheSer -1
GCGGACCCCGGGCCCCCCCCCCCACTCAACCCCTGCTGTTACTATCCATGCCAGCACCAGGGGATCTGTGTTCGGCTTCGGCCTTGACCGC 90
AlaAspProGlyAlaProAlaProValAsnProCysCysTyrTyrProCysGlnHisGlnGlyIleCysValArgPheGlyLeuAspArg 30
TACCAATGTGACTGCACCCGCACGGGCTATTCCGGCCCCAACTGCACCATCCCGGAGATATGGACCTGGCTCCGGACGACTTTGCGGCCC 180
TyrGlnCysAspCysThrArgThrGlyTyrSerGlyProAsnCysThrIleProGluIleTrpThrTrpLeuArgThrThrLeuArgPro 60
AGCCCCTCTTTCATCCACTTTCTGCTCACGCACGGGCGCTGGCTTTGGCATTTTGTCAATGCCACCTTCATCCGGGACACGCTCATGCGT 270
SerProSerPheIleHisPheLeuLeuThrHisGlyArgTrpLeuTrpAspPheValAsnAlaThrPheIleArgAspThrLeuMetArg 90
CTGGTACTCACAGTGCGTTCCAACCTTATCCCCAGCCCTCCCACCTACAACATAGCGCACGATTACATCAGCTGGGAGTCCTTCTCCAAT 360
LeuValLeuThrValArgSerAsnLeuIleProSerProProThrTyrAsnIleAlaHisAspTyrIleSerTrpGluSerPheSerAsn 120
GTGAGCTATTATACTCGCATTCTGCCCTCCGTGCCCCGAGACTGTCCCACGCCCATGGACACCAAAGGGAAGAAGCAATTGCCAGATGCG 450
ValSerTyrTyrThrArgIleLeuProSerValProArgAspCysProThrProMetAspThrLysGlyLysLysGlnLeuProAspAla 150
GAGTTCCTGAGCCGTCGCTTCCTGCTCAGGAGGAAGTTCATCCCTGACCCTCAAAGCACCAACCTCATGTTTGCCTTCTTTGCCCAACAC 540
GluPheLeuSerArgArgPheLeuLeuArgArgLysPheIleProAspProGlnSerThrAsnLeuMetPheAlaPhePheAlaGlnHis 180
TTCACCCATCAGTTCTTCAAAACTTCCGGCAAGATGGGTCCTGGCTTCACCAAGGCGCTGGGCCACGGGGTAGACCTCGGCCCACATTTAT 630
PheThrHisGlnPhePheLysThrSerGlyLysMetGlyProGlyPheThrLysAlaLeuGlyHisGlyValAspLeuGlyHisIleTyr 210
GGAGACAATCTGGAACGTCAGTATCAGCTGCGGCTCTTTAAGGATGGGAAGCTGAAGTACCAGATGCTCAATGGAGAGGTGTACCCGCCA 720
GlyAspAsnLeuGluArgGlnTyrGlnLeuArgLeuPheLysAspGlyLysLeuLysTyrGlnMetLeuAsnGlyGluValTyrProPro 240
TCGGTGGAAGAGGCGCCCGTGCTGATGCACTACCCCCGGGGCATCCCGCCCCAGAGCCAGATGGCTGTGGGCCAGGAGGTGTTTGGGCTG 810
SerValGluGluAlaProValLeuMetHisTyrProArgGlyIleProProGlnSerGlnMetAlaValGlyGlnGluValPheGlyLeu 270
CTTCCTGGACTCATGCTCTACGCCACGATCTGGCTGCGTGAGCACAACCGCGTGTGTGACCTGCTGAAGGCTGAGCACCCCACCTGGGGC 900
LeuProGlyLeuMetLeuTyrAlaThrIleTrpLeuArgGluHisAsnArgValCysAspLeuLeuLysAlaGluHisProThrTrpGly 300
GACGAGCAGCTCTTCCAGACGGCCCGCCTCATCCTCATCGGGGAGACCATCAAGATTGTGATAGAGGAGTATGTGCAGCAGCTGAGCGGC 990
AspGluGlnLeuPheGlnThrAlaArgLeuIleLeuIleGlyGluThrIleLysIleValIleGluGluTyrValGlnGlnLeuSerGly 330
TACTTCCTGCAGCTCAAGTTCGACCCAGAGCTGCTGTTCGGGGCCCAGTTCCAGTACCGCAACCGCATCGCCATGGAGTTCAACCAGCTG 1080
TyrPheLeuGlnLeuLysPheAspProGluLeuLeuPheGlyAlaGlnPheGlnTyrArgAsnArgIleAlaMetGluPheAsnGlnLeu 360
TACCACTGGCACCCGCTCATGCCCGACTCCTTCCGGGTGGGCCCCCAGGACTACAGCTACGAGCAGTTTCTGTTCAACACCTCCATGCTG 1170
TyrHisTrpHisProLeuMetProAspSerPheArgValGlyProGlnAspTyrSerTyrGluGlnPheLeuPheAsnThrSerMetLeu 390
GTGGACTACGGGGTCGAGGCCCTGGTGGACGCCTTTTCTCGCCAGCCTGCAGGCCCGATTGGTGGGGGTAGGAACATAGACCACCACATC 1260
ValAspTyrGlyValGluAlaLeuValAspAlaPheSerArgGlnProAlaGlyArgIleGlyGlyGlyArgAsnIleAspHisHisIle 420
CTGCACGTGGCCGTGGATGTCATCAAGGAATCACGGGTGCTAAGGCTGCAGCCCTTCAATGAGTACCGCAAGAGGTTTGGCATGAAGCCC 1350
LeuHisValAlaValAspValIleLysGluSerArgValLeuArgLeuGlnProPheAsnGluTyrArgLysArgPheGlyMetLysPro 450
TACACCTCTTTCCAAGAGCTCACAGGTGAGAAGGAGATGGCAGCTGAATTGGAGGAGCTGTATGGAGACATTGATGCTTTGGAATTCTAC 1440
TyrThrSerPheGlnGluLeuThrGlyGluLysGluMetAlaAlaGluLeuGluGluLeuTyrGlyAspIleAspAlaLeuGluPheTyr 480
CCGGGGCTACTTCTTGAGAAGTGCCATCCGAACTCCATCTTTGGGGAGAGTATGATAGAAATGGGGGCTCCTTTTTCCCTTAAGGGCCTC 1530
ProGlyLeuLeuLeuGluLysCysHisProAsnSerIlePheGlyGluSerMetIleGluMetGlyAlaProPheSerLeuLysGlyLeu 510
TTAGGAAACCCCATCTGTTCTCCAGAGTACTGGAAGGCGAGCACATTTGGCGGTGAGGTGGGCTTCAACCTTGTCAAGACGGCCACGCTA 1620
LeuGlyAsnProIleCysSerProGluTyrTrpLysAlaSerThrPheGlyGlyGluValGlyPheAsnLeuValLysThrAlaThrLeu 540
AAGAAGCTGGTTTGCCTCAACACCAAGACTTGTCCCTATGTCTCCTTCCACGTACCAGACCCCCGTCAGGAGGACAGGCCTGGGGTGGAG 1710
LysLysLeuValCysLeuAsnThrLysThrCysProTyrValSerPheHisValProAspProArgGlnGluAspArgProGlyValGlu 570
CGGCCACCCACAGAGGCTCTGAAGGGGCCGGGCAGCAGCATTCTGGATGGTAGAGCTTCCTGCTTGCCATTCCAGAATGCCACGGGGTGGA 1800
ArgProProThrGluLeu*** 576
TTGTCTTTGATGTTCGGTTTCTGATTTGGTGTCGACAGCATCACTCTCGACGTTTAGAACTCTAGGTCTCTCACCCATGGTCTGGAATAC 1890
TGTGTTCCTTGTTTGTTGTTCTAGAATGCTGAATTCCTGGTAAACCATTGAGAATGTTAGGAGTCGTTATCCCTTCAGCATGCCAGAACA 1980
CTGGGTTCCTGGGTGACCACCTAGAATGTCAGATTTCTAGTTGATCCGGAATTTAGGCACTCTGAAATATGGCACTCCTGATGGAATCAT 2070
CTGGAAAGTGAGGGGGTTTTTATTTTGCATTCTAGAATTCTGGGTGGCCCTCCAGAATGTCGACTTTCTGACTGGTTATCCGGAATGTTG 2160
TGCTCCGAGTTGCTGATCCAGAACAGTGGCTGGCATTCTAGATCAGTCCTGATCCGAATGTCTAGAGTGTGCAGAATTCATTTTCCTGTT 2250
CAGTGAGACAGCCACGGAGCAGGAGGATCTCGTGTCCTACAAGAACGCATTGCCTGGATCTGTGCCTGCATGGAGAGGGCAAAGAAGTGG 2340
GGTGTTCGTCTTCTCAGTGGGACCCCTCATGAGACCCTAGATATGGAGAGAACAGGTGGCTTTCTCCAGGCCATTGGTTGGAAGCCACCA 2430
CAGCTCTGTCCTCATCCAGGTCTCAACTCACGGCAGCTGTTTTTCATGAAGTTAATAAAATGCTTTTTCC----3' 2500

The nucleotide and deduced amino acid sequence are presented in fig.2. Primary structure of sheep PG endoperoxide synthase was deduced from the cDNA sequence using the reading frame corresponding to the N-terminal amino acid sequence of the intact enzyme polypeptide, the amino acid sequences of 6 *Achromobacter* proteinase I fragments, and that of the aspirin-acetylation site. The deduced amino acid sequence presented in fig.2 contains all the amino acid sequences described above except the residue at 516. All the nucleotide sequences corresponding to the amino acid residue 516 in clones pPES28, pPES31 and pPES33 were TGT. The alanine residue at position 1 is the N-terminus of the enzyme determined from the protein [4]. The first ATG (nucleotide residues −72 to −69) was assigned as the translational initiation codon. The nucleotide sequence surrounding the putative initiation codon agrees with the consensus sequence for eukaryotic initiation codons described by [18]. The nucleotide residues −72 to −1 encodes a 24-amino acid polypeptide which shows a characteristic sequence of the signal peptide [19]. A termination codon was found in frame after the 576th codon. A typical polyadenylation signal, AATAAA, appeared at nucleotide residues 2484–2489. Although the poly(dA) tract is not shown in fig.2, the residue 2500 was followed by $(dA)_{16}$. Determination of the C-terminal amino acid sequence with carboxypeptidase Y indicated a sequence of Thr-Glu-Leu, which agreed with the amino acid sequence deduced from the nucleotide sequence on the 5′-side of the termination codon. Therefore, a mature form of the enzyme is composed of 576 amino acid residues, omitting glycosyl residues, with an M_r of 66175. Except tryptophan the amino acid composition of the enzyme determined with the purified enzyme [20] coincided well with that calculated from the deduced amino acid sequence (not shown). Total RNA from sheep vesicular glands was shown to contain a mRNA species of approximately 3300 nucleotides that hybridized to the *Sma*I(758)-*Sma*I(1443) fragment of pPES31 (fig.3).

The enzyme is a glycoprotein and contains 4–5 mol of *N*-acetylglucosamine and 12–16 mol of mannose per mol of the enzyme polypeptide [3]. Four potential asparagine-linked glycosylation sites (residues 44, 80, 120 and 386) were found in

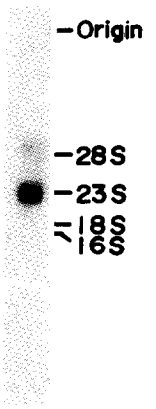


Fig.3. Autoradiogram of blot hybridization analysis of sheep vesicular gland RNA with a cDNA probe. 10 μg of total RNA was analyzed. The size markers used were sheep and *Escherichia coli* rRNA [21].

the deduced amino acid sequence. Hydropathicity profile of the deduced amino acid sequence of the enzyme showed several hydrophobic regions including the signal peptide (amino acid residues −19 to −8, 78–95, 261–282, 304–321 and 383–399). PG endoperoxide synthase is a membrane protein [1–3] and it has been shown by electron microscopic immunocytochemistry that the enzyme is associated with the endoplasmic reticulum and nuclear membrane in Swiss mouse 3T3 fibroblasts [22]. Thus, some of these hydrophobic segments possibly constitute membrane spanning portions and/or interact with hydrophobic substrates such as arachidonic acid. Cyclooxygenase activity of the enzyme is inhibited by acetylation with aspirin. The amino acid sequence of the aspirin-acetylation site is present near the C-terminus of the enzyme polypeptide: the serine residue specifically acetylated by aspirin was found to be the residue at 506. The portion around the serine residue is hydrophobic and is predicted to form an α-helical structure.

Acknowledgements: This work was supported in part by grants from Ministry of Health and Welfare, Ministry of Education of Japan and Cardiovascular Research Foundation, Japan.

REFERENCES

[1] Miyamoto, T., Ogino, N., Yamamoto, S. and Hayaishi, O. (1976) J. Biol. Chem. 251, 2629–2636.

[2] Hemler, M., Lands, W.E.M. and Smith, W.L. (1976) J. Biol. Chem. 251, 5575–5579.
[3] Van der Ouderaa, F.J., Buytehek, M., Nugteren, D.H. and Van der Dorp, D.A. (1977) Biochim. Biophys. Acta 487, 315–331.
[4] Van der Ouderaa, F.J., Buytenhek, M., Nugteren, D.H. and Van der Dorp, D.A. (1980) Eur. J. Biochem. 109, 1–8.
[5] Roth, G.J., Machuga, E.T. and Ozols, J. (1983) Biochemistry 22, 4672–4675.
[6] Laemmli, U.K. (1970) Nature 227, 680–685.
[7] Hermodson, M.A., Ericsson, L.E., Neurath, H. and Walsh, K.A. (1973) Biochemistry 12, 3146–3153.
[8] Wada, K., Takai, T. and Tanabe, T. (1987) Eur. J. Biochem. 167, 13–18.
[9] Botelho, L.H., Ryan, D.E. and Levin, W. (1979) J. Biol. Chem. 254, 5635–5640.
[10] Chirgwin, J.M., Przybyla, A.E., MacDonald, R.J. and Rutter, W.J. (1979) Biochemistry 18, 5294–5299.
[11] Maniatis, T., Fritsch, E.F. and Sambrook, J. (1982) in: Molecular Cloning: A Laboratory Manual, pp.230–241, Cold Spring Harbor Laboratory, Cold Spring Harbor, NY.
[12] Takai, T., Noda, M., Furutani, Y., Takahashi, H., Notake, M., Shimizu, S., Kayano, T., Tanabe, T., Tanaka, K., Hirose, T., Inayama, S. and Numa, S. (1984) Eur. J. Biochem. 143, 109–115.
[13] Sanger, F., Nicklen, S. and Coulson, A.R. (1977) Proc. Natl. Acad. Sci. USA 74, 5463–5467.
[14] Furutani, Y., Notake, M., Fukui, T., Ohue, M., Nomura, H., Yamada, M. and Nakamura, S. (1986) Nucleic Acids Res. 14, 3167–3179.
[15] Thomas, P.S. (1980) Proc. Natl. Acad. Sci. USA 77, 5201–5205.
[16] Kyte, J. and Doolittle, R.F. (1982) J. Mol. Biol. 157, 105–132.
[17] Chou, P.Y. and Fasman, G.D. (1978) Annu. Rev. Biochem. 47, 251–276.
[18] Kozak, M. (1984) Nucleic Acids Res. 12, 857–873.
[19] Steiner, D.F., Quinn, P.S., Chan, S.J., Marsh, J. and Tager, H.S. (1980) Ann. NY Acad. Sci. 343, 1–16.
[20] Roth, G.J., Siok, C.J. and Ozols, J. (1980) J. Biol. Chem. 255, 1301–1304.
[21] McMaster, G.K. and Carmichael, G.G. (1977) Proc. Natl. Acad. Sci. USA 74, 4835–4838.
[22] Rollins, T.E. and Smith, W.L. (1980) J. Biol. Chem. 255, 4872–4875.