Structure and expression of the human gene encoding plasminogen activator inhibitor, PAI-1

(Primer extension analysis; genomic library; serine protease inhibitor; glycoproteins; promoter regions; serpins; phage $\lambda$ vectors; introns; cat gene; $Alu$ repeat)

Marie Follo* and David Ginsburg*\textsuperscript{a.b}

Departments of \textsuperscript{a} Human Genetics and \textsuperscript{b} Internal Medicine, and \textsuperscript{b} Howard Hughes Medical Institute, University of Michigan Medical School, Ann Arbor, MI 48109-0650 (U.S.A.)

Received by J.L. Slightom: 1 May 1989
Revised: 18 July 1989
Accepted: 22 July 1989

SUMMARY

The gene ($pail$) encoding human plasminogen activator inhibitor, PAI-1 was cloned from a $\lambda$EMBL3 genomic library and was found to span approx. 12 kb and to contain eight introns. Primer extension and S1 nuclease analyses both showed the transcription start point to be located 142 nt upstream from the start codon. The 5'-flanking region was sequenced and found to contain a TATA box, but no CAAT sequence. When a fragment containing 730 nt of 5'-untranslated region was placed upstream from a promoterless cat gene, it was shown to function as a promoter when transfected into COS cells. Northern-blot analysis was consistent with low level expression of the endogenous $pail$ gene in COS cells. When the $pail$ gene structure was compared to those of other members of the serine-protease-inhibitor encoding gene family, little conservation of intron positions was observed.

INTRODUCTION

Plasminogen activators are serine proteases which are able to convert the inactive plasminogen into the active protease plasmin. Plasminogen activation is involved in many biological processes such as tumorigenesis, embryogenesis, and tissue repair. However, it has been most completely studied for its role as the central step in the activation of the fibrinolytic system (Erickson et al., 1985).

The major form of plasminogen activator inhibitor found in human plasma is PAI-1 (for review see...
Loskutoff et al., 1989). PAI-1 is also found in platelets, endothelial cells, hepatoma cells, fibrosarcoma cells, mesothelial cells, and recently as a major component of the extracellular matrix. Vascular endothelial cells may represent the major site of synthesis for plasma PAI-1 (Loskutoff et al., 1989).

Several groups have cloned full-length PAI-1 cDNA (Ginsburg et al., 1986, Ny et al., 1986, Pannekoek et al., 1986, Andreasen et al., 1986). In human cells PAI-1 mRNA occurs in two sizes, 2.3 kb and 3 kb, which differ only in the length of their 3'-untranslated regions. Both mRNAs encode a 42-45-kDa polypeptide which displays sequence similarity to other members of the serpin gene family. The human PAI-1 gene has been mapped to chromosome 7 (Ginsburg et al., 1986; Klinger et al., 1987).

The regulation of PAI-1 gene expression is complex. In HUVECs, PAI-1 activity is induced by substances such as thrombin, endotoxin, IL-1, TNF-α and TGF-β, and decreased by protein C and heparin with ECGF. Expression of PAI-1 has also been shown to be induced by glucocorticoids in hepatoma and fibrosarcoma cells (Loskutoff et al., 1989).

We have used a PAI-1 cDNA probe to screen a human genomic library and have isolated overlapping clones which contain the PAI-1 gene. The exon structure and the tsp were determined. A fragment containing 730 bp of 5'-flanking sequence was found to act as a promoter when transfected into COS cells. The gene structure of PAI-1 was compared to that of other serpins.

MATERIALS AND METHODS

(a) Genomic cloning

Approximately 2 × 10⁶ recombinant clones from an amplified partial Sau3A human genomic library (obtained from P. Venta, University of Michigan) constructed in the λEMBL3 vector (Frischauf et al., 1983) were screened by hybridization to the 32P-labeled 2-kb EcoRI insert fragment of cDNA clone PAIB6 (Ginsburg et al., 1986). Four overlapping clones containing PAI-1 gene sequences were identified. Restriction fragments were subcloned into pGEMblue (Promega) and a detailed restriction map was constructed by standard methods (Maniatis et al., 1982). Probes were radiolabeled using the random hexamer method (Feinberg and Vogelstein, 1983). Sequencing was by the dideoxy method (Sanger et al., 1977) using Sequenase (U.S. Biochemical Corporation). Oligo sequencing primers were synthesized based on the PAI-1 cDNA sequence designed to sequence across the splice junctions as reported by Loskutoff et al. (1987).

(b) Polymerase chain reaction

Reactions were performed as previously described (Ginsburg et al., 1989). Either 100 ng of the λEMBL3 phage clone 4A DNA, 100 ng of control plasmid DNA, or 100 ng of normal genomic DNA were added to each reaction mix in addition to 0.02 A₂₆₀ of each 30-mer oligo primer. Taq polymerase (Cetus) was used according to the manufacturer's instructions. PCR conditions were 94°C for 1 min, 15 s, followed by 72°C for 4 min repeated for 30 cycles. The oligos used for intron 3 were the following. (3A) GATGAGATCAGCACCACAGACGC, PAI-1 cDNA positions 334-363 (Ginsburg et al., 1986), and (3B) 5'-GCACAGAGCAGTGCTGCCGTCTGA-3', oligos for intron 4 were the following. (4A) 5'-ACTCCCTTC-CCGACTCTTCAGCACCAGCCG, located from nt 598-627; and (4B) 5'-GGTGTTCCCGTG- GTAGGGCAGTTCAGGAT, complementary to positions 638 through 667. Oligos for intron 4 were the following. (4A) 5'-ACTCCCTTC-CCGACTCTTCAGCACCAGCCG, located from nt 598-627; and (4B) 5'-GGTGTTCCCGTG-GTAGGGCAGTTCAGGAT, complementary to positions 638 through 667.

(c) Computer analysis

Alignment of serpin aa sequences was performed using the Needleman-Wunsch (1970) algorithm through the Sequence Analysis Software Package (Genetics Computer Group at the University of Wisconsin).

RESULTS AND DISCUSSION

(a) Gene cloning

A human genomic library in λEMBL3 was screened with a full-length human cDNA clone. Four overlapping phage clones were isolated and
Fig. 1. Structure and sequence of \( pail \) gene. (A) Gene structure. Restriction sites are marked above the map. E, EcoRI; H, HindIII; N, NcoI; S, SalI. Exons are numbered from below. Phage clone 4A obtained from a \( \lambda \)EMBL3 genomic library encompasses the entire region shown in detail, as well as greater than 8 kb of upstream sequence. The two flanking SalI sites have been contributed by the phage. (B) \( PAIL \)-intron/exon boundaries. Sequences of the intron/exon splice junctions are shown. Appropriate aa are listed below the codons. The mRNA splicing consensus sequences (Padgett et al., 1986) are shown in the bottom line for comparison. (C) Sequence of the \( pail \) promoter region. The TATA box is underlined and in bold type. The major \( tsp \) is the adenine residue located under the arrowhead. The possible Z-DNA forming region is also underlined. Met denotes the initiation methionine. The short vertical line indicates the location of intron 1.
subjected to restriction endonuclease mapping. Clone 4A was found to contain an insert of greater than 19 kb and was studied in further detail. Exons were located by restriction enzyme mapping and Southern-blot analysis. All of the clones isolated stop at the *Sau*3a site approx. 650 bp 5' of the distal polyadenylation signal used for the 3 kb PAI-1 mRNA species. Clone 4A extends approx. 8 kb upstream from the first exon. The structure of clone 4A was confirmed by comparison with Southern blots of human DNA (not shown).

(b) Gene structure

The human *pail* gene spans approx. 12 kb and contains eight introns (Fig. 1a). The locations of exon 4 was deduced using PCR to determine the sizes of introns C and D. All exons were sequenced in their entirety and the sequence was found to be identical to that of our previously reported cDNA sequence (Ginsburg et al., 1986). The intron/exon boundary sequences are shown in Fig. 1b and were found to conform to the known consensus sequences (Padgett et al., 1986). Sequencing differences were seen in the splice acceptor sequences of introns A, D, E and F when compared to Loskutoff et al. (1987), though our sequence agrees completely with that of Bosma et al. (1988) and Strandberg et al. (1988). The 300-bp *Eco*RI fragment located in intron 3 contains an *Alu* repeat and is not present in the gene structure of Loskutoff et al. (1987).

An optimal alignment of the PAI-1 protein sequence with that of other members of the serpin family was determined, and the relative locations of the introns were compared. Our results are consistent with those of Strandberg et al. (1988). As previously noted, conservation of intron positions among serpins correlates poorly with aa sequence similarity (Bao et al., 1987). The degree of aa sequence similarity has frequently been used to estimate evolutionary relatedness among proteins. We hypothesize that the presence or absence of introns should subject to weaker selective pressures than aa sequence, and thus may provide a more reliable measure of evolutionary relatedness within the serpin family than aa sequence. A similar hypothesis was suggested by Ragg and Preibisch (1988) in their analysis of the serpin hLS2.

The *tsp* for *pail* transcription was determined by primer extension and S1 nuclease analysis (Fig. 2). With both methods, the *tsp* was shown to be 142 bp 5' of the start codon (Figs. 1c and 2). 432 bp of upstream promoter sequence was determined on human DNA (not shown).
both strands (Fig. 1c) and consensus promoter elements were noted as previously observed (Bosma et al., 1988 and Fig. 1c).

The location of the pail tsp is consistent with the observation that A residues are the most common nt used at the tsp (Salditt-Georgieff et al., 1980) and that CA is the most common motif at the cap site (Bucher and Trifonov, 1986). Our results agree with those found by Riccio et al. (1988) in their S1 analyses on a variety of cell types and Strandberg et al. (1988) in S1 analysis of human placenta. Bosma et al. (1988) and Van Zonnefeld et al. (1988) have localized the start 3 nt upstream from this site at an A residue which follows a G. In both of our experiments secondary bands became visible upon longer exposure, one of which was located at this second A residue, possibly representing a weak alternative tsp.

(c) Promoter function

To test whether the region upstream from the pail gene could function as a eukaryotic promoter, an 800-bp Ncol-EcoRI fragment which contained approx. 730 bp of 5' flanking region was inserted 5' of the bacterial cat gene in both orientations in the vector pUMSVO. Plasmid pRSVCAT was used as a positive control. The 800-bp EcoRI-NcoI pail fragment was shown to function as a promoter in COS cells only in the correct orientation (Fig. 3). This segment is a much weaker promoter in COS cells than the strong viral promoter of RSV.

Northern-blot analysis of HUVEC and COS total RNA probed with radiolabeled pail cDNA revealed that COS cells contain the PAI-1 message (Fig. 4). The relative amount of PAI-1 in COS cells appears to be substantially lower than that in HUVECs. Since COS cells are derived from the kidney epithelium of the African green monkey (Jensen et al., 1964), these results are consistent with the previous detection of PAI-1 antigen in BSC-1, another monkey kidney epithelial cell line (Thalacker and Nilsen-Hamilton, 1987).

In human cells, PAI-1 mRNA is of two sizes, a smaller 2.3-kb message and a larger 3-kb message. This size difference has been postulated to be due to either differential splicing or to the use of different polyadenylation sites. The absence of an intron at
Fig. 4. Northern-blot analysis of total HUVEC and COS cell RNA was performed as described (Ginsburg et al., 1986). The final two washes were in 0.1 $\times$ SSC and 0.1% SDS at 68°C for 30 min each. Film was developed after a 16 h exposure at -80°C with an intensifying screen. The probe used was the same as that used to screen the genomic library. Equal amounts (10 $\mu$g) of total RNA was loaded on each lane. The quantity of RNA present in each lane appeared equal by the comparison of 18S and 28S rRNA bands visualized by ethidium bromide staining (not shown). The locations of the 28S and 18S rRNA bands are indicated.

the end of the 2.3-kb message confirms that this size difference is due to alternative polyadenylation (Loskutoff et al., 1987; Bosma et al., 1988; Strandberg et al., 1988). It is interesting to note that both messages are present in COS cells, as in human cells (Fig. 4). When PAI-1 mRNA has been examined in other species only the larger 3-kb mRNA species is seen [bovine (Sawdey et al., 1986), rat (Zeheb and Gelehrter, 1988), and porcine, (M.F. and D.G., unpublished data)]. Processing at the proximal polyadenylation signal producing the 2.3-kb mRNA species removes an A + T-rich palindromic sequence which has been hypothesized to play a role in regulating pail gene expression (Ginsburg et al., 1986). The significance of the shorter message observed only in primate species remains to be determined.

ACKNOWLEDGEMENTS

We would like to thank J. Leiden for his thoughtful review of the manuscript. This work was supported in part by NIH-NHLBI-No. 1-ROI-HL39137-01. DG is an investigator of the Howard Hughes Medical Institute.

REFERENCES


Gorman, C.M., Merlino, G.T., Willingham, M.C., Pastan, I. and Howard, B.H.: The Rous sarcoma virus long terminal repeat


