THE JOURNAL OF BIOLOGICAL CHEMISTRY

Vol. 263, No. 19, Issue of July 5, pp. 9129-9141, 1988
Printed in U.S.A.

Human Plasminogen Activator Inhibitor-1 Gene

PROMOTER AND STRUCTURAL GENE NUCLEOTIDE SEQUENCES*

(Received for publication, January 12, 1988)

Piter J. Bosma‡, Eva A. van den Berg, and Teake Kooistra

From the Gaubius Institute TNO, Herenstraat 5d, 2313 AD Leiden, The Netherlands

David R. Siemieniak and Jerry L. Slightom

From the Division of Molecular Biology, The Upjohn Company, Kalamazoo, Michigan 49007

We have determined the nucleotide sequence of the human plasminogen activator inhibitor-1 (PAI-1) gene and significant stretches of DNA which extend into its 5′- and 3′-flanking DNA regions; a total sequence of 15,867 base pairs (bp) is presented. The sequenced 5′-flanking DNA (1,520 bp) contains the essential eukaryotic *cis*-type proximal regulatory elements CCAAT and TATAA; the more distal 5′-flanking DNA region, as well as some introns, contain sequence elements which share identities with known eukaryotic enhancer elements. A major finding is the identification of a large region of shared nucleotides (comprising of about 520 bp) between the 5′-flanking DNAs of PAI-1 and tissue-type plasminogen activator genes. The length of the PAI-1 5′-untranslated region was found to be 145 bp as determined by nuclease analysis. The remaining PAI-1 structural gene consists of amino acid coding regions (containing a total of 1,206 bp, coding for the 23 amino acids of the signal peptide and 379 amino acids of the mature PAI-1 protein), 8 intron regions (a total of 8,978 bp), and a long 3′-untranslated region of about 1,800 bp which contains several polyadenylation sites. Two types of repetitive DNA elements are located within the PAI-1 structural gene and flanking DNAs: we have found 12 *Alu* elements and 5 repeats of a long poly (*Pur*) element. These *Alu-Pur* elements may represent a subset of the more abundant *Alu* family of repetitive sequence elements.

Plasminogen activators (PAs)[1] are serine proteases which convert the proenzyme plasminogen into another serine protease, plasmin, by specific cleavage of a single peptide bond. Plasmin, a powerful broad spectrum protease, is involved in a variety of physiological and pathological processes like fibrinolysis, tissue remodeling, tumor growth, and metastasis (Reich, 1978; Danø *et al.*, 1985; Emeis *et al.*, 1985). Thus, precise regulation of PA activity may play a crucial role in the control of these processes. Two distinct types of PA have been identified, tissue-type plasminogen activator (t-PA) and urokinase-type plasminogen activator (u-PA). t-PA has been primarily associated with fibrinolytic function, while u-PA is believed to also have a regulatory role in other forms of extracellular proteolysis. As a specific inhibitor of both t-PA and u-PA, plasminogen activator inhibitor-1 (PAI-1) is a protein of great interest because of its potential role in the control of plasminogen activation.

PAI-1 is a glycoprotein present in plasma and in the α-granules of blood platelets and is synthesized by a variety of cells *in vitro*, including endothelial cells, hepatocytes, smooth muscle cells, and several tumor cell lines (Erickson *et al.*, 1985; Knudsen *et al.*, 1987; Sprengers and Kluft, 1987). Molecular cloning of a human PAI-1 cDNA has been reported independently by several groups (Ny *et al.*, 1986; Pannekoek *et al.*, 1986; Ginsburg *et al.*, 1986; Wun and Kretzmer, 1987). The deduced amino acid sequence of PAI-1 reveals a signal peptide of 23 amino acid residues and a mature protein containing 379 amino acid residues with three potential sites of *N*-linked glycosylation. Comparison of the sequence of PAI-1 with other proteins indicates that PAI-1 is a member of the serine protease inhibitor (Serpin) family (Carrell and Boswell, 1986).

The expression of PAI-1 is enhanced many-fold in several situations. In plasma, PAI-1 behaves as an acute phase reactant: a rapid increase in PA inhibitory activity is seen after major surgery, severe trauma, and myocardial infarction (Juhan-Vague *et al.*, 1985; Kluft *et al.*, 1985a). PAI-1 activity in plasma is also induced by agents such as bacterial wall lipopolysaccharide (endotoxin), interleukin-1, and tumor necrosis factor (Colucci *et al.*, 1985; Emeis and Kooistra, 1986; Van Hinsbergh *et al.*, 1987a). Studies with cultured cells also suggest that stimulation of PAI-1 production can be induced by a number of cellular stimuli including thrombin (Gelehrter and Sznycer-Laszuk, 1986; Van Hinsbergh *et al.*, 1987b), endotoxin (Colucci *et al.*, 1985; Emeis and Kooistra, 1986), inflammatory mediators (Emeis and Kooistra, 1986; Van Hinsbergh *et al.*, 1987b), and glucocorticoids (Gelehrter *et al.*, 1983; Baumann and Eldredge, 1982; Andreasen *et al.*, 1987).

Recent reports suggest that PAI-1 may have functions other than intravascular control of PA activity. Several groups have recently presented evidence that the extracellular matrix produced by cells in tissue culture may contain PAI-1 (Pöllänen *et al.*, 1987; Rheinwald *et al.*, 1987; Knudsen *et al.*, 1987). (It should be noted that the amino acid sequence of mesosecrin (Rheinwald *et al.*, 1987) is identical to that of PAI-1.) In view of the well established role of PA in matrix remodeling and degradation, the presence of PAI-1 in the matrix may provide an important component in the regulation of a number of extravascular physiological and pathological processes.

* 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) J03764.

‡ This work partially fulfills the Ph.D. thesis requirement.

[1] The abbreviations used are: PAs, plasminogen activators; t-PA, tissue-type PA; PAI-1, PA inhibitor-1; kb, kilobase(s); bp, base pair(s); Pipes, 1,4-piperazinediethanesulfonic acid.

To learn more about how the expression of the PAI-1 gene is regulated, we need to know the nucleotide sequence of the complete gene and flanking DNAs so that potential enhancer-like sequence elements can be identified. Toward this goal we have isolated the PAI-1 gene from a human genomic cosmid library and determined its complete nucleotide sequence. Our sequence reveals the same exon/intron structure for the PAI-1 gene as seen in the partial characterization of the PAI-1 gene reported by Loskutoff *et al.* (1987). Our complete sequence also reveals the identity and location of possible *cis*-acting regulatory elements in the 5′-flanking DNA region and the location of repetitive DNA elements. In addition, a major finding is that the PAI-1 and t-PA 5′-flanking DNA regions contain sequences which share extensive identity. Although the identification of possible regulatory elements by sequence identity is speculative, the 5′-flanking sequence and its subclone will supply the necessary material and information needed to design biological experiments to test the authenticity of such potential regulatory elements. Such experiments may also elucidate the identity of other *cis*-type regulatory elements which are not apparent at this time. (Part of this work has been presented at the XIth International Congress of Thrombosis and Haemostasis (Bosma *et al.*, 1987).)

EXPERIMENTAL PROCEDURES

Materials—Restriction enzymes were purchased from Boehringer Mannheim GmbH, Promega Biotec, and New England Biolabs. Agarose ultrapure DNA grade was from Bio-Rad. DNA polymerase I, S1 nuclease, and bovine alkaline phosphatase were obtained from Boehringer Mannheim, and polynucleotide kinase was obtained from U.S. Biochemical Corp. T4 ligase was obtained from Collaborative Research, and mung bean nuclease was obtained from Pharmacia LKB Biotechnology Inc. Radioactive nucleotide [γ-^{32}P]ATP (2000–3000 Ci/mM; 1 Ci = 3.7×10^{10} Bq) was obtained from ICN Pharmaceuticals and [α-^{32}P]CTP (2000–3000 Ci/mM) and GeneScreen filter membranes were from Du Pont-New England Nuclear. Chemicals used for DNA sequencing were obtained from vendors recommended by Maxam and Gilbert (1980). X-ray roll film (Kodak XAR-351) and DNA sequencing gel stands and safety cabinets were obtained from Fotodyne, Inc. Intensifying screens (Quanta III: 35 cm × 1 m) were obtained from Du Pont.

Screening of Human Cosmid Library and Subcloning—A human genomic cosmid library was kindly provided by Dr. J. Hoeymakers (Erasmus University, Rotterdam). This library original contained 1.6×10^6 independent recombinant clones and was prepared by cloning partial *Mbo*I-digested human placenta DNA (size fractionated for fragments in the range of 40–50 kb) into the *Bam*HI-cut cosmid vector pTCF (Grosveld *et al.*, 1982). One equivalent of this cosmid library was screened with a ^{32}P-labeled (nick-translated) 1400-bp PAI-1 cDNA probe (van den Berg *et al.*, 1987) for the presence of PAI-1 containing cosmid clones according to the procedures described by Maniatis *et al.* (1982). Cloned DNA inserts used for making ^{32}P-labeled probes were isolated from ultrapure DNA grade agarose gels by electroelution. Five cosmid clones which contain PAI-1 sequence were isolated, and cosmid DNAs were purified as described by Maniatis *et al.* (1982). The location of the PAI-1 gene in each cosmid clone was identified by restriction enzyme site analysis and blot hybridization. Two of these cosmid clones, designated PAI-Cos1 and PAI-Cos2, were subsequently mapped in detail. Both cosmid clones contain the complete PAI-1 gene (see Fig. 1).

Restriction enzyme fragments of PAI-Cos2 were subcloned into pUC9 to facilitate DNA sequencing and later gene engineering experiments. This was accomplished by subcloning a 7.0-kb *Eco*RI fragment (which contained the promoter region) and three other *Eco*RI fragments (1.8, 6.9, and 5.2 kb) which contained the major portion of the PAI-1 structural gene (see Fig. 2). These non-overlapping clones are referred to as pPAI-E7.0, pPAI-E1.8, pPAI-E6.9, and pPAI-E5.2, respectively.

Nucleotide Sequencing—Chemical sequencing was performed essentially as described by Maxam and Gilbert (1980). Appropriate restriction enzyme digests were prepared using either PAI-1 subcloned plasmid DNAs (10–20 μg) or the parent cosmid PAI-Cos2 (30 μg). The 5′-phosphate groups were removed by treatment with bovine alkaline phosphatase, without removing the restriction enzyme; the pH of the restriction enzyme buffer was adjusted to 8.4 by the addition of 0.1 volume of 1 M Tris-HCl (pH 8.4). Then, 5–10 units of phosphatase were added, and the samples were incubated at 55 °C for 30 min. Restriction enzymes and phosphatase were removed by two extractions with a mixture of phenol/chloroform/isoamyl alcohol (1:1:0.04, by volume) followed by one extraction with chloroform/isoamyl alcohol (1:0.04). Endlabeling with [γ-^{32}P]ATP was done using conditions described by Slightom *et al.* (1980), and end-labeled fragments were isolated on glycerol (10–20%) polyacrylamide (5–7.5%) gels as described by Maxam and Gilbert (1980). Because long DNA sequencing reads were expected in the range of 600 bp, the specific chemical reaction times (at 20 °C) were reduced as follows: 1 μl of dimethyl sulfate for the G[2] reaction, incubate for 30 s; 30 μl of 95% formic acid for the A reaction, incubate for 2.5 min; and 30 μl of 95% hydrazine for the C + T and C reactions, incubate for 2.5 min (Chang and Slightom, 1984). Chemical reaction times were reduced even further (one-half from that described above) for reactions involving DNA fragments larger than 1.5 kb to ensure samples were not overreacted and thus could be read out to at least 600 bp. DNA sequence reactions were subjected to electrophoresis on gels measuring 20 cm wide, 104 cm long and with a wedge spacer thickness which varies from 0.4 mm (bottom) to 0.2 mm (top) (obtained from C.B.S. Scientific Co.). Because the wedge is not continuous through the complete gel, these gels are referred to as "bell-bottom." Additional information concerning the design, pouring, and electrophoresis running times has been described previously by Slightom *et al.* (1987). Generally, using a 4% polyacrylamide bell-bottom gel, nucleotide sequence reads extend out to about 500 bp, and if longer reads were desired (reads beyond 650 bp) the sequenced samples were run on a 4% polyacrylamide gel fitted with a uniform 0.2-mm spacer and electrophoresed for a total of 45,000 volt-hours. The capacity of these gels could be increased by using a comb with 3-mm slots (available from International Biotechnologies, Inc.) which allowed 32 loads to be made across a 20-cm wide gel; thus, from a single 4% bell-bottom gel loaded once with eight different sequenced fragments, as much as 4,000 bp of nucleotide sequence information could be obtained.

RNA Isolation—Total cellular RNA was isolated from cultured endothelial cells as described by Lizardi and Engelberg (1979) with two minor modifications. In the first modification, cells were washed in warm, 37 °C, phosphate-buffered saline (6.7 mM KH_2PO_4, 6.7 mM K_2HPO_4, and 150 mM NaCl) and incubated at 37 °C for 30 min in lysis buffer (50 mM Tris-HCl (pH 7.4), 100 mM NaCl, 7.5 mM EDTA, 0.5% sodium dodecyl sulfate, 150 μg/ml proteinase K) which was applied directly to the culture dish. The second modification was introduced after the ethanol precipitation step as the RNA was purified from contaminating DNA by precipitation in 2 M LiCl for 3 h at 0 °C followed by one additional ethanol precipitation. RNA samples were then dissolved in sterile H_2O, and the RNA concentration was determined spectrophotometrically.

Nuclease Analysis of 5′-Transcript Start Site—The PAI-1 transcription initiation site was determined by nuclease mapping with either S1 or mung bean nucleases, as described by Weaver and Weissmann (1979). This analysis was done by adding approximately 100,000 cpm of a *Nco*I-*Eco*RI restriction enzyme fragment (from positions −795 to +70 (see Fig. 3)), which was 5′-end-labeled at the *Eco*RI site, to 50 μl of hybridization buffer (0.08 M Pipes (pH 6.4), 2 mM EDTA, 80% formamide). This ^{32}P-labeled DNA probe solution was heated to 95 °C for 1.5 min (to denature the DNA duplex) and then chilled on ice, followed by the addition of 50 μl of a high salt solution (0.8 M NaCl, 80% formamide). Total RNA, approximately 15 μg, was dissolved in 20 μl of this DNA probe solution and then sealed in a capillary, or put into a small polypropylene tube under mineral oil, and allowed to anneal with the single-stranded DNA overnight at 50 °C. After hybridization, samples were added to 200 μl of nuclease digestion buffer (0.25 M NaCl, 0.03 M sodium acetate, 1 mM $ZnSO_4$) and 5 μg of sonicated calf thymus or salmon sperm DNA. Single-stranded DNA and RNAs were removed by the addition of 400 units of nuclease (either S1 or mung bean) and incubated at room temperature for 20 min. These nuclease-digested samples were then precipitated with ethanol. Reaction samples, control samples which contained either tRNA or no RNA and the sequenced *Nco*I-*Eco*I

[2] Letter codes for amino acids and nucleotide sequence ambiguities are in accordance with those proposed by IUB Nomenclature Committee (1985) *Eur. J. Biochem.* **150,** 1–5. The 15,867 bp of this human PAI-1 gene is available from GenBank nucleotide sequence database.

fragment, were analyzed by electrophoresis through 16 and 6% polyacrylamide sequencing gels. This procedure allows for the identification of protected fragments up to 400 bp 5′ of the *Eco*RI site.

Computer Analysis of Nucleotide Sequences—Computer-aided analyses of the PAI-1 nucleotide sequence were performed using programs supplied by the University of Wisconsin Genetics Computer Group. These programs were designed for use on the VAX computer, and some of the programs have been described by Devereux *et al.* (1984).

RESULTS AND DISCUSSION

Human PAI-1 Gene Isolation, Characterization, and Subcloning—We isolated five cosmid clones by screening a human recombinant cosmid library (1.6×10^6 independent clones) with a ^{32}P-labeled DNA probe isolated from a PAI-1 cDNA clone (van den Berg *et al.*, 1987). These cosmid clones were subjected to restriction enzyme site mapping followed by blot hybridization analysis of the resulting gels. These analyses showed that the five cosmid clones represented only two unique 43-kb insert types, designed PAI-Cos1 and PAI-Cos2 (Fig. 1), each of which contains the complete PAI-1 gene. This limited analysis demonstrated that these PAI-1 genes are probably identical, which is consistent with the presence of one PAI-1 gene copy per haploid genome (Ginsburg *et al.*, 1986; and confirmed by our analysis, data not shown). These two cosmid clones could contain non-allelic isolates of the PAI-1 gene, but our present analysis has not revealed any polymorphic restriction enzyme sites to support this possibility.

Nucleotide sequence analysis of the human PAI-1 gene was facilitated by subcloning restriction enzyme fragments of PAI-Cos2 into either the pUC-derived part of the original cosmid vector (pTCF) or pUC9 using *Eco*RI restriction enzyme sites. The resulting subclones are shown in Fig. 2, below a detailed map of PAI-Cos2, and they are designated as pPAI-E7.0 (an *Eco*RI subclone which contains the PAI-1 promoter region), and *Eco*RI subclones pPAI-E1.8, pPAI-E6.9, and pPAI-E5.2.

Nucleotide Sequence of the Human PAI-1 Gene—The complete nucleotide sequence of the human PAI-1 gene and considerable stretches of its 5′- and 3′-flanking DNA regions were determined using the chemical DNA sequencing procedure described by Maxam and Gilbert (1980). The strategy used to determine this nucleotide sequence is shown in Fig.

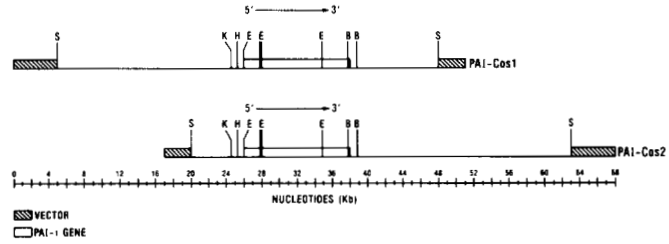


FIG. 1. **The physical location of the PAI-1 gene and overlapping regions of cosmids PAI-Cos1 and -Cos2.** The restriction enzyme maps and blot hybridization with the PAI-1 cDNA clone were used to locate the PAI-1 gene in each of the cosmid clones; enzyme sites present in the vector, pTCF (Grosveld *et al.*, 1982) are not shown. The restriction enzymes used are indicated as follows: *B*, *Bam*HI; *E*, *Eco*RI; *H*, *Hin*dIII; *K*, *Kpn*I; *S*, *Sal*I.

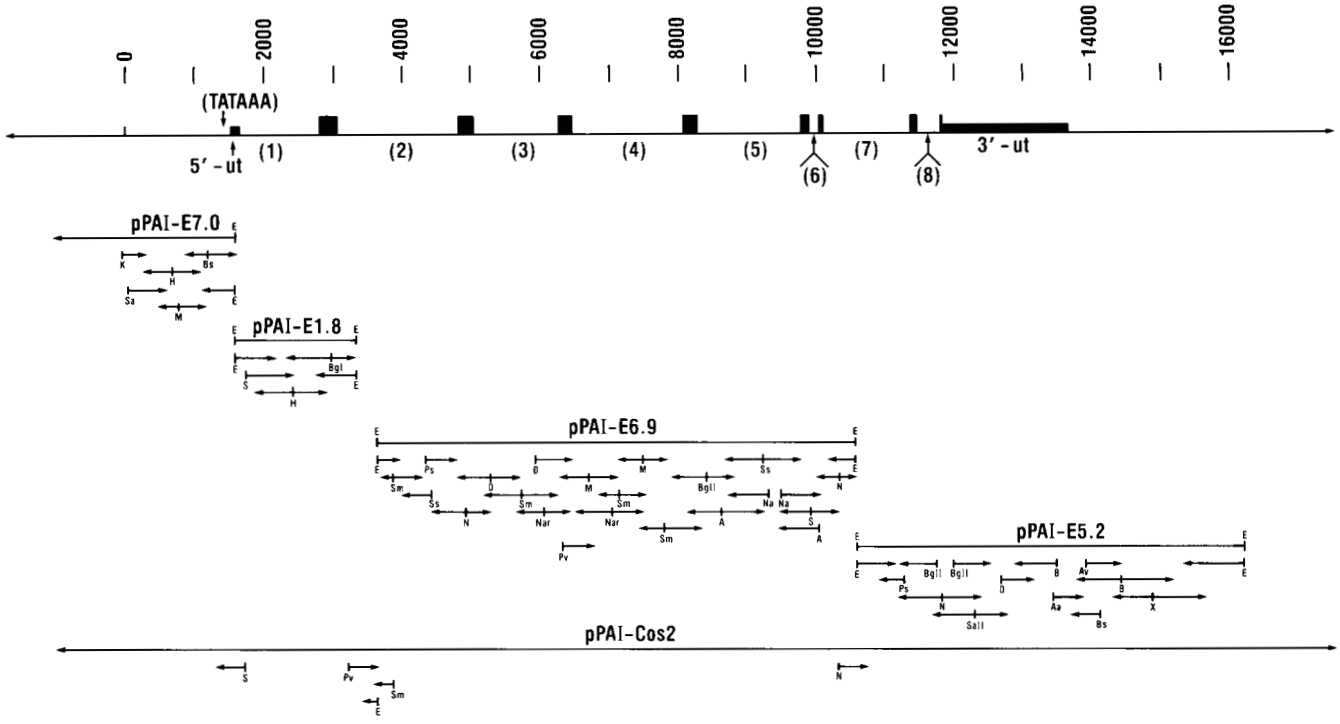


FIG. 2. **Structure of the PAI-1 gene, subclones, and sequencing strategy.** The *top line* presents the structure of the PAI-1 gene by showing the location of exons (*raised black bars*) and introns (*regions between raised bars*) as determined from the PAI-1 gene sequence presented in Fig. 3. The *thicker raised black bars* represent coding regions, while the *thinner raised bars* represent noncoding regions of the transcript. The introns are numbered from *1* to *8*. *5′-ut*, 5′-untranslated. The complete nucleotide sequence of the PAI-1 gene was determined from the four subclones shown below the gene structure. Nucleotide sequences crossing enzyme sites used for cloning were determined from cosmid clone pPAI-Cos2. The sequencing strategy used is shown *below* the clones; the enzyme site used and direction and distance sequenced are shown by the *horizontal arrows*. The restriction enzymes used for sequencing are indicated as follows: *A*, *Apl*I; *Aa*, *Aat*II; *Av*, *Ava*I; *B*, *Bam*HI; *BgI*, *Bgl*I; *BgII*, *Bgl*II; *Bs*, *Bst*EII; *D*, *Dra*I; *E*, *Eco*RI; *H*, *Hin*dIII; *K*, *Kpn*I; *M*, *Mst*II; *N*, *Nco*I; *Na*, *Nae*I; *Nar*, *Nar*I; *Ps*, *Pst*I; *Pv*, *Pvu*II; *S*, *Sal*I; *Sa*, *Sac*I; *SaII*, *Sac*II; *Sm*, *Sma*I; *Ss*, *Ssp*I; *X*, *Xba*I.

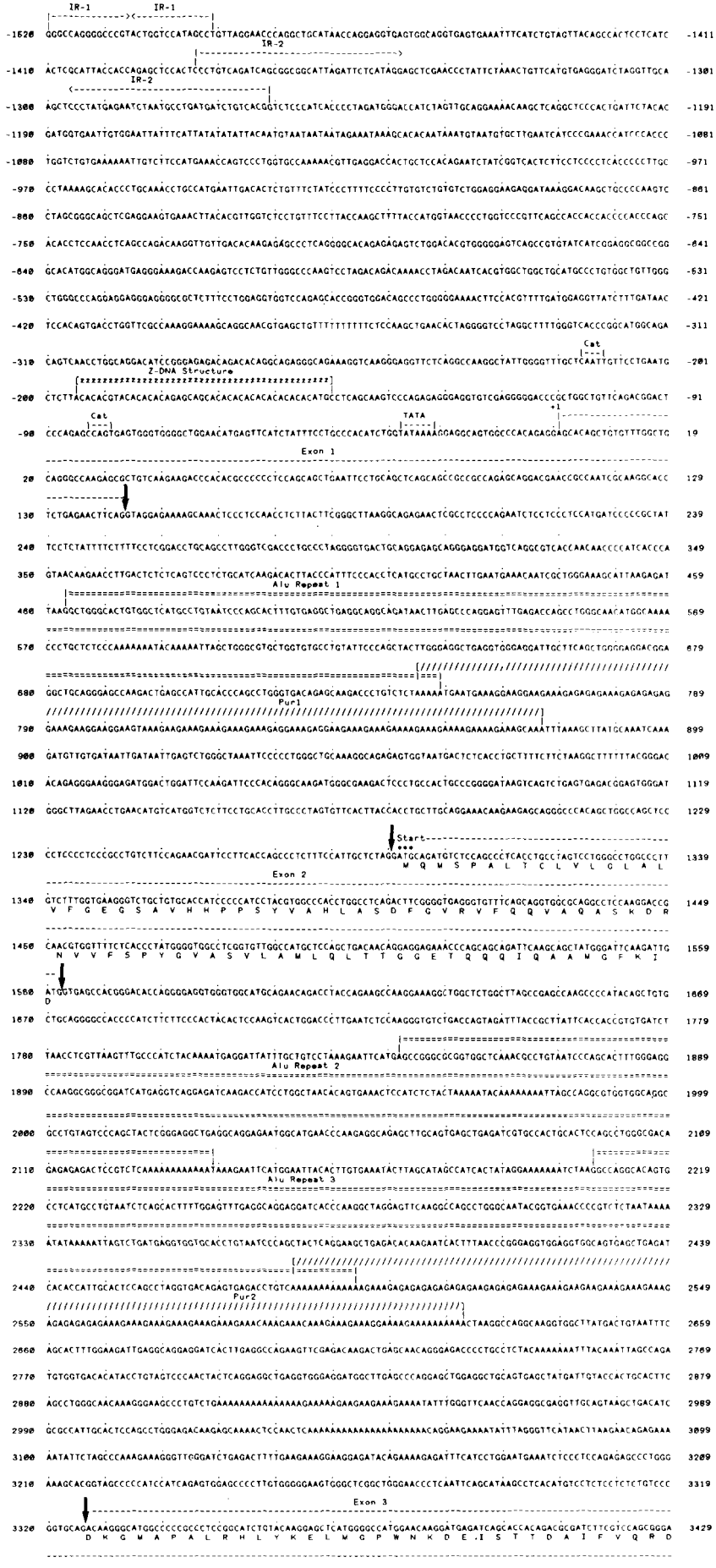


FIG. 3. **The complete nucleotide sequence of the human PAI-1 gene and flanking DNA regions.** The nucleotide sequence of the human PAI-1 gene and flanking DNAs was determined using the strategy shown in Fig. 2. The PAI-1 transcription initiation site was determined by S1 and mung bean nuclease mapping experiments (see Fig. 4), and the most probable initiation site (*A*) is marked as *position 1*. The organization of the PAI-1 exons (indicated by the *dashed line*) and introns was determined by comparison with the sequence of the many sequenced PAI-1 cDNA clones (see text). Translation initiation codon is indicated by *asterisks* (***) above the first *M* and the termination codon is designated *TER*; the complete PAI-1 amino acid sequence (single-letter code) is shown *below* the second nucleotide of each codon. Nucleotide sequences which may function as proximal promoter elements, because they share identity with the TATAA and CCAAT elements, are indicated *above* the sequence line. The location of two large DNA regions capable of forming Z-DNA are indicated by the letter (*z*) *above* the sequence line. *Vertical arrows* indicate exon-intron boundaries, and they all conform to the GT/AT rule (Breathnach *et al.*, 1978) and closely fit the consensus donor and acceptor sequences purposed by Mount (1982). The *arrows* marked *IR-1* and *IR-2* indicate the positions of the two largest inverted repeats found in the 5′-flanking DNA. In the 3′-untranslated DNA region (shown as a *continuous line*) sequences which share identity with the consensus polyadenylation signals (AATAAA) are indicated along with a nonconsensus polyadenylation signal sequence (AATAAT) which may be responsible for the shorter PAI-1 transcript (Pannekoek *et al.*, 1986; Ny *et al.*, 1986). The *double dashed line above* the sequence indicates the location of *Alu* elements, numbered *1–12*; poly purine (*Pur*) elements, numbered *1–5*, are indicated by ////// *above* the sequence.

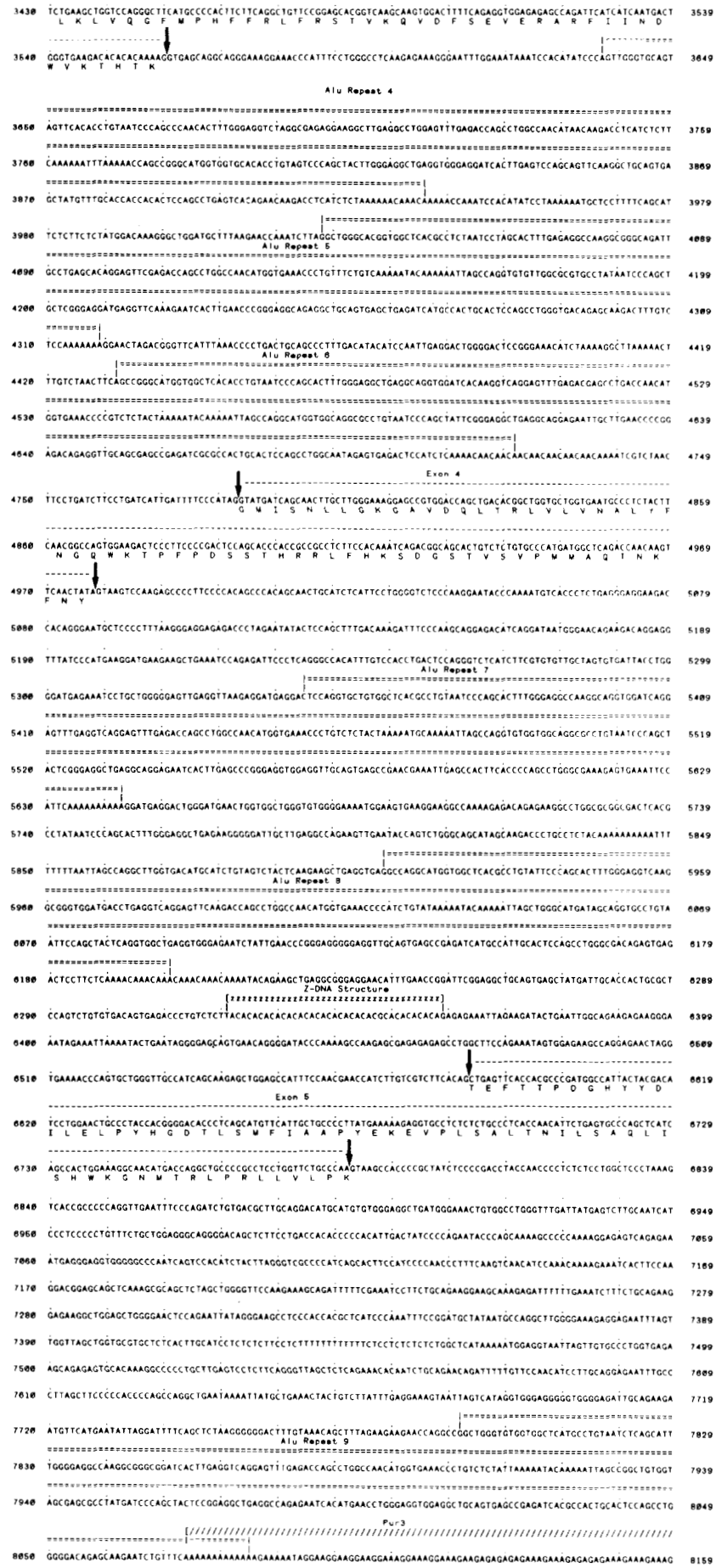


FIG. 3—*continued*

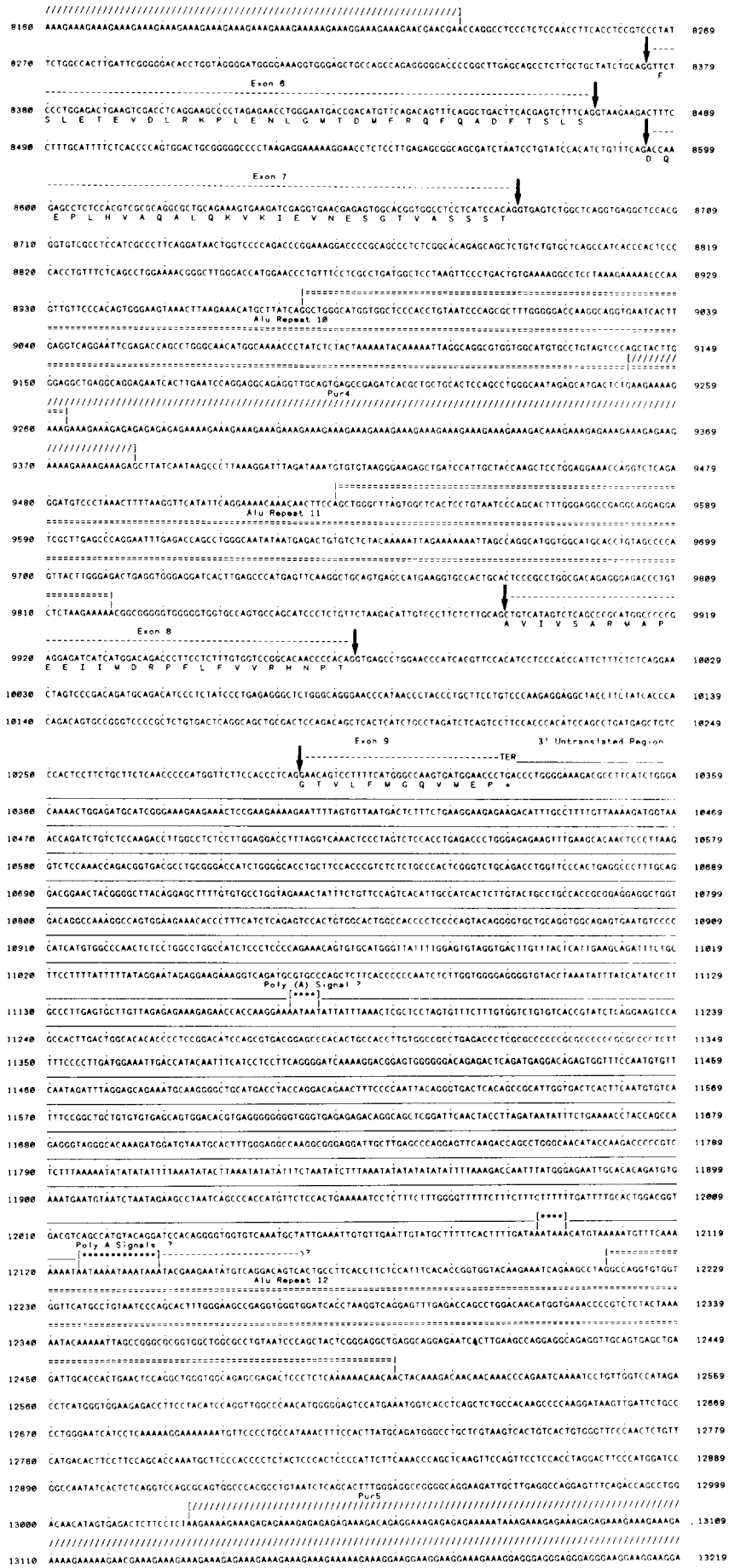
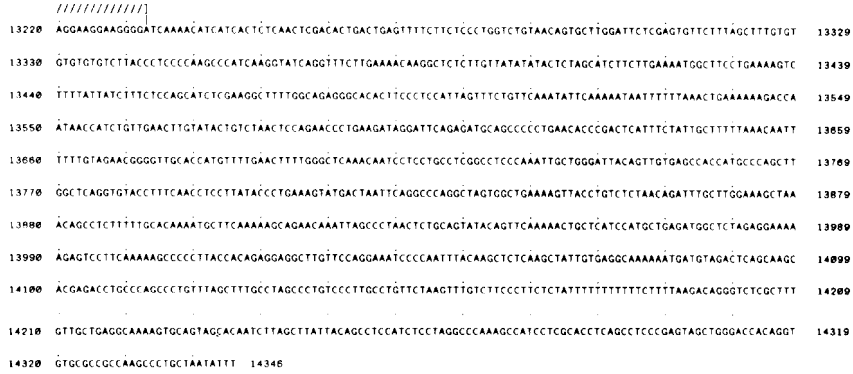


FIG. 3—*continued*

FIG. 3—*continued*



2. Most of the sequence was determined from the plasmid subclones except for that needed to cross the non-overlapping *Eco*RI cloning sites; these sequences were determined directly from the parent cosmid PAI-Cos2 (see Fig. 2). A total of 15,867 bp was determined, and a comparison with PAI-1 cDNA clone sequences (Pannekoek *et al.*, 1986; Ny *et al.*, 1986) shows that the entire PAI-1 structural gene is contained within about 12,165 bp, depending on the assignment of 5′- and 3′-untranslated DNA lengths (see below). In addition, we have sequenced 1,520 bp of 5′-flanking DNA and about 2,200 bp of 3′-flanking DNA; all of these sequences are presented in Fig. 3.

Structural Organization of the PAI-1 Gene—Because the sequences of several PAI-1 cDNA clones have previously been reported (Pannekoek *et al.*, 1986; Ny *et al.*, 1986; Ginsburg *et al.*, 1986; Wun *et al.*, 1987) determination of the organization of the PAI-1 gene was straightforward. The PAI-1 gene contains nine exons which are separated by eight introns. The location and size of these exons/introns are shown in Figs. 2 and 3, and are listed in Table I. The sites of both exons and intron regions vary considerably: exons range between 84 and 1823 bp, while introns range between 119 and 1764 bp in length. These intron sequences account for 74% of the PAI-1 structural gene length. The first exon contains 144 bp of 5′-untranslated DNA (see below). Exon 2 starts with 1 bp of 5′-untranslated DNA which is followed by the translation initiation site. The PAI-1 5′-untranslated region does not contain any false translation initiation start sites. The sequence flanking this translation initiation site, AGGATGC, only partially conforms to the consensus eukaryotic ribosome binding site (Kozak, 1987) because of the nonconsensus C-nucleotide in position +4.

Exon 9 is the largest exon, but encodes for only 12 amino acids (Fig. 3) as it consists mostly of the long 3′-untranslated DNA region. The remaining exons have an average length of 160 bp, which is consistent with the average length of 150 bp reported for the exons of genes of higher eukaryotes (Naora and Deacon, 1982). The exact length of this 3′-untranslated DNA region cannot be assessed at this time, as all the reported sequences of PAI-1 cDNA clones contain different 3′-untranslated DNA lengths. RNA blot analysis shows the presence of at least two distinct PAI-1 mRNA species, of about 3,500 and 2,350 nucleotides (Pannekoek *et al.*, 1986; Ny *et al.*, 1986; Ginsburg *et al.*, 1986). Restriction enzyme mapping and blot hybridization analyses of PAI-1 exon 9 region by Loskutoff *et al.* (1987) suggest that these different transcripts are not the result of alternative splicing but are due to the presence of multiple polyadenylation splice sites. Our analysis of the PAI-1 3′-untranslated and flanking DNAs confirms this observation because these 3′-untranslated regions of the cDNA and gene sequences are collinear. A check of the PAI-1 sequence (Fig. 3) reveals the presence of several potential

TABLE I

Location and size of exons and introns in the human PAI-1 gene

Exon	Nucleotide positions	Length	Amino acids[a]
		bp	
1	1–144	144	
2	1,292–1,563	272	−23–68
3	3,328–3,561	234	69–146
4	4,785–4,979	195	147–211
5	6,585–6,783	199	212–278
6	8,376–8,476	101	279–311
7	8,596–8,682	87	312–340
8	9,891–9,974	84	341–368
9	10,295–12,165	1,871	369–379 and 3′-untranslated region

Intron	Type[b]	Nucleotide positions	Length	No. of *Alu* repeats
			bp	
1		145–1,291	1,147	1
2	I	1,564–3,327	1,764	2
3	I	3,562–4,784	1,223	3
4	I	4,980–6,584	1,605	2
5	II	6,784–8,375	1,592	1
6	I	8,477–8,595	119	0
7	I	8,683–9,890	1,208	2
8	I	9,975–10,294	319	0

[a] Minus amino acids represent the PAI-1 signal peptide.

[b] Intron type is according to Sharp (1981); 0 indicates a splice between codons, I indicates a splice after the first nucleotide of a codon, and II indicates that the splice occurs after the second nucleotide.

polyadenylation signals which fit the consensus sequence AATAAA (Proudfoot and Brownlee, 1976). Four putative polyadenylation signals are located between nucleotides 12,092 and 12,141, but no consensus signal is present at an upstream position which could be responsible for the polyadenylation splice that produces the shorter mRNA species. However, the hexanucleotide sequence AATAAT, which has a 1-bp mismatch, is located at position 11,174 and is surrounded by an A-T-rich region. If this sequence functions as a polyadenylation signal, the resulting mRNA species would be about 900 bp shorter, close to that observed (Pannekoek *et al.*, 1986; Ny *et al.*, 1986; Ginsburg *et al.*, 1986). Additional analyses (nuclease mapping or primer extension experiments) are needed to determine the locations of the functional polyadenylation signals.

The eight introns found in the PAI-1 gene conform to the GT-AG rule (Breathnach *et al.*, 1978), and the sequences surrounding these splice junctions (Table II) conform, but are not identical to, the consensus sequence proposed by Mount (1982). However, the splice junction sequences presented here do differ from those reported by Loskutoff *et al.* (1987); these

TABLE II

Intron-exon splice junction sequences in the human PAI-1 gene

Consensus sequence (from Mount (1982)).

```
C      A                TTTTTTTTTTT T
 AG GT AGT. . .  . . .              N AG
A      G                CCCCCCCCCCC C
```

Intron	Splice junction sequences			
	Exon	Intron		Exon
1	TCAG	GTAGGA. . .	. . .TTTCCATTGCTCTAG	GATG
2	GATG	GTGAGC. . .	. . .TCTGTCCCGGTGCAG	ACAA
3	AAAG	GTGAGC. . .	. . .TTGATTTTCCCATAG	GTAT
4	TATA	GTAAGT. . .	. . .TTGTCGTCTTCACAG	CTGA
5	CCAA	GTAAGC. . .	. . .TGCTGCTATCTGCAG	GTTC
6	TCAG	GTAAGA. . .	. . .CCACATCTGTTTCAG	ACCA
7	ACAG	GTGAGT. . .	. . .TCCCTTCTCTTGCAG	CTGT
8	ACAG	GTGAGC. . .	. . .TTCTTCCACCCTCAG	GAAC

TABLE III

Differences in PAI-1 acceptor intron/exon boundary sequences reported by Loskutoff et al. (1987) (A) and those presented in Fig. 3 (B)

Gene	Intron	Sequence	
A	1	CTTGCTCTAG	GATG
B	1	ATTGCTCTAG	GATG
A	4	CTCTGTACAG	CTGA
B	4	GTCTTCACAG	CTGA
A	5	CCCGTTGCAG	GTTC
B	5	CTATCTGCAG	GTTC
A	6	TCTGTTTTAG	ACCA
B	6	TCTGTTTCAG	ACCA

TABLE IV

Differences in nucleotide sequence between the previously published PAI-1 cDNA (see cDNA numbers) and that presented in Fig. 3

The nucleotide at position 3,392 is in the coding region, the other nucleotides are in untranslated regions. Blanks indicate sequence not available.

Nucleotide position (from Fig. 3)	Gene	PAI-1 sequence from			
		cDNA 1[a]	cDNA 2[b]	cDNA 3[c]	cDNA 4[d]
19	T	G			
26	GCCA	GCA			
3,392	T	C	T	T	T
10,693	C		C	T	C
11,163	C		T	T	
11,166	C		T	T	

[a] Pannekoek *et al.* (1986).
[b] Ny *et al.* (1986).
[c] Ginsburg *et al.* (1986).
[d] Wun *et al.* (1987).

differences are listed in Table III. Interestingly, all of these differences are located within the intron part of the acceptor splice site. Whether these differences are due to nucleotide sequence polymorphisms or sequencing errors remains to be determined. A total of nine differences were found in the comparison of less than 100 bp, a level much higher than that expected for sequence polymorphisms (even in neutral DNA); the expected level is about 0.14% (Miyamoto *et al.*, 1987).[3] A recheck of our sequence gels shows no artifacts that could have been responsible for reading errors.

Several nucleotide sequence differences were found between our PAI-1 gene sequence and the cDNA sequences reported previously (Pannekoek *et al.*, 1986; Ny *et al.*, 1986; Ginsburg *et al.*, 1986; Wun *et al.*, 1987), and these differences are listed in Table IV. The 5′-untranslated region of the cDNA sequence presented by Pannekoek *et al.* (1986) differs at two positions from our sequence data and the 3′-untranslated regions the PAI-1 gene and PAI-1 cDNA clones differ at three positions (see Table IV). In the coding DNA region, these PAI-1 cDNA sequences differ only at gene nucleotide position 3,392 (Fig. 3 and Table IV), where either a T- or C-nucleotide substitution is observed. This nucleotide substitution is silent and thus does not influence the PAI-1 protein structure. The number of nucleotide sequence differences among these representatives of the PAI-1 transcribed regions is only five out of a total of about 12,000 nucleotides compared (assuming that each cDNA and the genomic clone represents a PAI-1 gene from five unrelated individuals), or a difference of 0.04%. Thus, these differences may represent PAI-1 gene polymorphisms in the human population.

Identification of the Transcription Initiation Site—The longest cloned PAI-1 5′-untranslated region, a length of 126 bp, was reported by Pannekoek *et al.* (1986). However, no additional evidence, such as nuclease mapping, was presented showing this to be the full length of the 5′-untranslated region. The precise location of the transcriptional start site is needed to support the identification of *cis*-acting DNA sequence elements. To clarify this, we mapped the PAI-1 transcription initiation site using both S1 and mung bean nuclease protection assays (Weaver and Weissmann, 1979) using a ^{32}P-end-labeled *Nco*I to *Eco*RI fragment (extending from position −795 to 70, see Fig. 3). The result of one of these nuclease protection assays is shown in Fig. 4; after a 1.5-bp correction for the difference in mobility found between identical fragments generated by DNA sequencing *versus* nuclease digesting (Sollner-Webb and Reeder, 1979) the major protected band corresponds to a A-nucleotide, as shown in Fig. 3. Some degree of ambiguity in mapping transcript initiation sites can be expected (Slightom *et al.*, 1985). This result indicates that the full length PAI-1 5′-untranslated region is approximately 145 bp.

Identification of Potential cis-Acting Regulatory Elements—

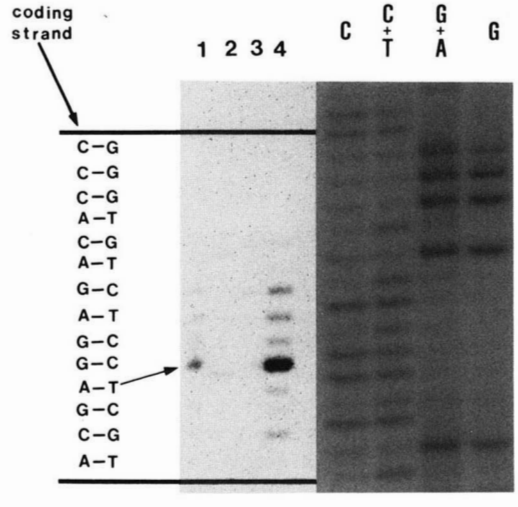


FIG. 4. **Localization of PAI-1 transcription initiation sites by nuclease mapping.** A 865-bp *Nco*I-*Eco*RI fragment spanning the expected PAI-1 promoter region was 5′-end-labeled at the *Eco*RI site, denatured, and then annealed with total cellular RNA from nontreated (*lane 1*) and tumor necrosis factor-treated (*lane 4*) endothelial cells. *Lanes 2* and *3* contain control samples annealed with tRNA or no RNA, respectively. After nuclease digestion, these samples were electrophoresed through 16 and 6% polyacrylamide sequencing gels along with the sequenced *Nco*I-*Eco*RI fragment. Nuclease-protected DNA fragments were only found at a position about 72 bp 5′ of the *Eco*RI site. The major transcription initiation site is indicated by the *arrow*, after correction (Sollner-Webb and Reeder, 1979).

[3] J. L. Slightom, unpublished data.

TABLE V

Putative regulatory elements found in the human PAI-1 gene

R stands for A or G. W stands for A or T. K stands for G or T. Y stands for C or T.

Enhancer element	Consensus sequence	Position	PAI-1 sequence
CAT	CCAAT	−218 to −214	CAATT
		−84 to −81	CCAGT
Inverted Cat	ATTGG	−217 to −213	AGTGG
TATAA	TATAA	−28 to −24	TATAA
SV40 enhancer core	GTGGWWWG	−479 to −469	GTGGACAG
		−227 to −220	GGGGTTTG
Acute phase signal	CTGGGA	439 to 444 (intron 1)	CTGGGA
Inverted acute phase	TCCCAG	−141 to −136	AGGGTC
		−92 to −86	ACGGGTC
		491 to 496 (intron 1)	AGGGTC
		625 to 630 (intron 1)	AGGGTC
Antithrombin III enhance	GTRRWTTG	−1186 to −1179	GTGAATTG
SP-1 binding site	KGGGCGGRRY	−534 to −525	TGGGCTGGGC
		−76 to −66	TGGGTGGGGC
Z-DNA	$(GT)_n$	−196 to −153	$(YR)_{23}$
		6321 to 6358	$(CA)_{18}$

A search of the DNA flanking the 5′-side of the human PAI-1 gene revealed the two common DNA sequence elements which are known to be important for proper transcription of many eukaryotic genes, the TATAA and CCAAT elements (Efstratiadis *et al.*, 1980; Maniatis *et al.*, 1987). A consensus TATAA element is located between positions −28 and −24 before the transcriptional start site (Fig. 3), a distance which is consistent with that found for most eukaryotic genes. No consensus CCAAT element was found, but two closely related sequences are located farther upstream; the sequence CCAGT at the expected position, −85 to −81, and the sequence CAATT (which is also present on the opposite strand) located farther upstream, at position −218 to −214 or −217 to −213 (see Fig. 3 and Table V). Which, if any, of these CCAAT-type elements is functional remains to be determined.

One of the important aspects of obtaining the nucleotide sequence of the 5′-flanking DNA region is to examine it for the presence of *cis*-acting regulatory elements, also referred to as enhancer elements. Such enhancer elements have been associated with modulating the activity of proximal promoter elements (Dynan and Tjian, 1985). However, the identification of enhancer-type elements by comparative analyses (even if computer aided) is difficult because there are many different types of enhancer sequences and their affect is not dependent on orientation or location. They can be located in either 5′ or 3′ directions and at distances far from the transcription start site. Although consensus sequences for many enhancers exist, these sequences are generally not large, involving less than 15 nucleotides. Thus, the probability of finding sequences which closely match any consensus enhancer sequence is high, and authenticity of any potential enhancer element requires *in vitro* or *in vivo* expression analysis.

With the aid of computer search programs we have examined the entire human PAI-1 gene for sequences similar to known enhancers. Many potential enhancer sequence elements were identified and most are listed, by sequence and position, in Table V. This list includes sequences which share identities with the following enhancers: SV40 enhancer (GTGGWWWG) (Gruss, 1984), SV40 Sp1 protein binding site (KGGGCGGRRY) (Dynan and Tjian, 1985), a sequence element common to acute-phase reactant genes (CTGGGA) (Fowlkes *et al.*, 1984; Adrian *et al.*, 1986), and the possible human antithrombin III or immunoglobulin light chain gene enhancer (CTRRWTTG) (Prochownik, 1985). In addition to these potential enhancer elements, a stretch (almost a pure 48 bp) of alternating purine and pyrimidines, Z-DNA, is located between positions −195 to −153 (Fig. 3). (Another Z-DNA stretch, 38 bp, is located in intron 4, see Fig. 3.) Such Z-DNA stretches have been found in the sequences of acute-phase reactant genes, such as human haptoglobin and human C-reactive protein (Cooper *et al.*, 1987). The presence of this Z-DNA region together with the two inverted copies of the acute-phase reactant regulatory sequences (CTGGGA, see Table V) in the PAI-1 5′-flanking DNA are interesting elements to be tested for their involvement in the acute-phase reactant-like behavior of PAI-1 expression. Three more copies of this probable acute-phase reactant regulatory hexanucleotide sequence, one in normal and two in inverted orientation, are found in the first intron (listed in Table V), and several other copies are located in its 3′-flanking DNA region (locations not listed in Table V). Whether any of these potential enhancer-type sequences are involved in regulating the level of human PAI-1 gene expression, or if there are other (yet to be discovered) unique PAI-1 gene enhancer elements, will have to await data from expression experiments.

Indirect Repeats and Shared Nucleotide Sequences between Human t-PA and PAI-1—With the aid of computer search programs, the 1520 bp of PAI-1 5′-flanking DNA was searched for the presence of direct and inverted repeats. Numerous short, imperfect direct and indirect repeats are located in this PAI-1 5′-flanking DNA region (data not shown), but only two imperfect indirect repeats, of at least 15 bp, were found, and their locations are shown in Fig. 3. One inverted repeat, IR-1, is located at the very end of the sequenced 5′-flanking DNA, positions −1520 to −1506 and −1505 to −1491, and contains matches of 11 out of 15 nucleotides. The IR-2 repeat is located nearby, positions −1383 to −1348 and −1295 to −1261, and contains matches of 31 out of 35 nucleotides (but requires two gaps).

Our search for PAI-1 5′-flanking DNA sequence identity in the GenBank database revealed a significant degree of shared identity with only one eukaryotic gene: the 5′-flanking region of the t-PA gene (Friezner Degen *et al.*, 1986). The presence of these regions of shared nucleotides is surprisingly quite extensive, involving an alignment of 521 positions, including only six gaps which were used to maximize identity. These sequence identities are located on opposite DNA strands, PAI-1 noncoding strand nucleotide positions −1520 to −1008 and t-PA coding strand nucleotide positions −3491 to −2977. The degree of shared identity is quite high (81%) and includes both IR-1 and IR-2 imperfect indirect sequence elements, see Fig. 5. It should be noted that the size of the shared PAI-1 and t-PA sequence region could be more extensive than we show because the region of identity, shown in

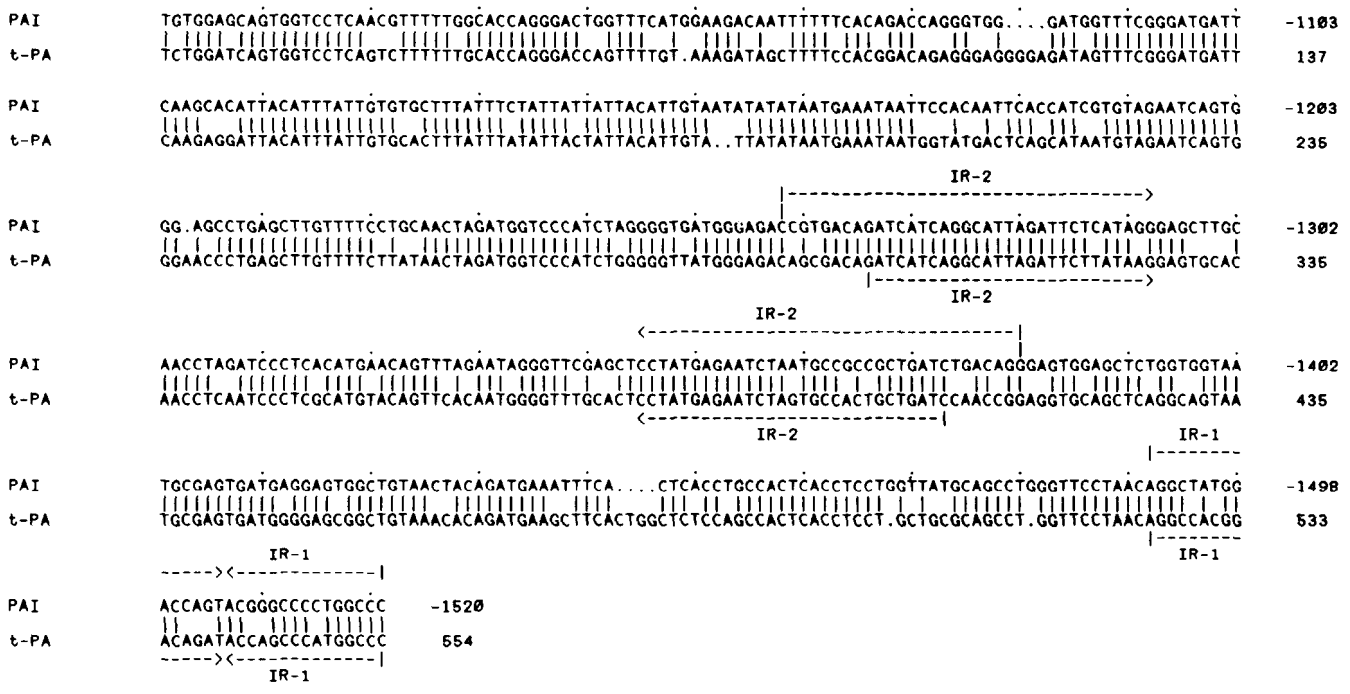


FIG. 5. **Alignment of shared nucleotide sequences between the human PAI-1 and t-PA genes.** A search of GenBank sequences showed that an extensive region sequence identity is shared between the 5′-flanking regions of the human PAI-1 and t-PA genes. The shared sequences are from opposite strands with the PAI-1 sequences (*top line*) coming from the noncoding DNA strand, positions −1520 to −1008; while the t-PA sequences (*bottom line*) are from the DNA strand which encodes the t-PA structural gene, positions −3491 to −2977 (Friezner Degen *et al.*, 1986). The alignment shows a match of 81%, which required the use of only six gaps (shown as *dots* in the sequence lines). This shared identity also includes the two inverted repeats which are found in both 5′-flanking DNAs.

Fig. 5, extends to the end of the PAI-1 5′-flanking DNA sequence presented in Fig. 3. Additional 5′ sequencing of this PAI-1 flanking DNA could reveal more shared sequences identity which would extend inward toward the t-PA promoter region. A check of the last 5′ sequences (38 bp) of the t-PA sequence (Friezner Degen *et al.*, 1986) shows a much-reduced level of shared identity, suggesting that this PAI-1 and t-PA identity region may not extend farther upstream in the t-PA 5′-flanking DNA, or in the reciprocal direction downstream toward the PAI-1 transcription initiation site. Clearly, delineation of the extent of this shared PAI-1 and t-PA nucleotide region will require additional sequencing of both genes. The t-PA indirect repeat IR-1 shares about the same degree of identity as that found for the IR-1 elements of PAI-1. However, the t-PA IR-2 repeats are shorter than those found in PAI-1; the t-PA repeat involves 29 nucleotides with a match of 23 out of 29, including one gap.

The existence of shared nucleotide sequence identity, even if on opposite strands, is interesting and suggests that these sequences, or shorter sequence elements within this region, may be important in regulating the coordinate expression of the t-PA and PAI-1 genes. In this respect it might be significant that changes in t-PA levels in plasma in several cases have been accompanied by corresponding changes in the activity of PAI-1. For example, an increase in plasma levels for both t-PA antigen and PAI-1 activity is observed after surgery, myocardial infarction, and severe trauma (Kluft *et al.*, 1985a); a decrease in both t-PA antigen and PAI-1 activity is observed after treatment of healthy volunteers with stanozolol (Verheijen *et al.*, 1984); a very similar diurnal fluctuation pattern was found for t-PA antigen and PAI-1 activity in plasma (Kluft *et al.*, 1985b). Recent *in vitro* studies also provide support for the coordinate expression of these genes. In cultured endothelial cells, thrombin and histamine stimulate the production of both t-PA and PAI-1 (Hanss and Collen, 1987; Van Hinsbergh *et al.*, 1987b), and in the human fibrosarcoma cell line HT1080, the increased production of t-PA in the presence of dexamethasone is also associated with increased production of PAI-1 (antigen) (Medcalf *et al.*, 1987). The latter authors demonstrated by Northern blot hybridization and nuclear "run-off" transcript assays that the dexamethasone-induced increase of t-PA and PAI-1 synthesis parallels equivalent changes of gene template activity. Thus, the functional significance of the shared t-PA and PAI-1 nucleotide sequence in their regulatory regions clearly deserves further investigation.

Repetitive Elements in and around the Human PAI-1 Gene—A search of the PAI-1 gene sequence for the presence and location of common repeat elements, such as *Alu* (Houck *et al.*, 1979), *Sau*3A (Kiyama *et al.*, 1986), and *Kpn* (Adams *et al.*, 1980), revealed only sequences which share identity with the *Alu* element. Higher primate *Alu* elements are short dimeric sequences (about 300 bp in length) which probably were derived by retroposition of a sequence originally derived from a 7SL RNA gene (Ullu and Tschudi, 1984; Kariya *et al.*, 1987). It has been estimated that the human genome may contain nearly 900,000 copies of the *Alu* repeat element (Hwu *et al.*, 1986). Fig. 6 presents the results of this *Alu* element search in the form of an alignment of the 12 PAI-1 gene *Alu* elements which share between 90 and 77% identity with the consensus *Alu* element (Koop *et al.*, 1986). None of these *Alu* elements are located in the 5′-flanking DNA or exon regions, and only one is located in the 3′-flanking DNA. Most are located in intron sequence regions (see Table I), all are located on the coding strand (see Fig. 3 for exact locations) and in many cases several *Alu* elements are located within close

A.

```
ConAlu      1  GGCCGGGCGCGGTGGCTCA..CGCCTGTAATCCCAGCAC.....TTTGGGAGG.CCGAGGCGGGCGGATCACCT....GAGGTCAGGAG....TTCGAGA   100
Alu1           GGCTGGGCACTGTGGCTCAT..GCCTGTAATCCCAGCAC.....TTTGTGAGGCT.GAGGCAGGCAGATAACTT....GAGCCCAGGAGTTTGAGACCAG
Alu2           AGCCGGGCGCGGTGGCTCAAACGCCTGTAATCCCAGCAC.....TTTGGGAGGCC.AAGGCGGGCGGATCA......TGAGGTCAGGAGATCAAGACCAT
Alu3           GGCCAGGCACAGTGCCTCAT..GCCTGTAATCTCAGCAC.....TTTGGAGTTT..GAGGCAGGAGGATCACCC....AAGGCTAGGAGTTCAAGGCCAG
Alu4           AGTTGGGGTGCAGTAGTTCACACC.TGTAATCCCAGCCCAACACTTTGGGAGGTCT.AGGCGAGAGGAAGGCTT....GAGGCCTGGAGTTTGAGACCAG
Alu5           GGCTGGGCACGGTGGCTCAC..GCCTCTAATCCTAGCAC.....TTTGAGAGG.CCAAGGCGGGCAGATTGCCT....GAGCACAGGAGTTCGAGACCAG
Alu6           AGCCGGGCATGGTGGCTCACA..CCTGTAATCCCAGCAC.....TTTGGGAGGCT.GAGGCAGGTGGATCAC......AAGGTCAGGAGTTTGAGACGAG
Alu7           CTCCAGGTGCTGTGGCTCAC..GCCTGTAATCCCAGCAC.....TTTGGGAGG.CCAAGGCAGGTGGATCAGGAGTTTGAGGTCAGGAGTTTGAGACCAG
Alu8           GGCCAGGCATGGTGGCTCAC..GCCTGTATTCCCAGCAC.....TTTGGGAGGTC.AAGGCGGGTGGATGACCT....GAGGTCAGGAGTTCAAGACCAG
Alu9           GGCTGGGTGTGGTGGCTCAT..GCCTGTAATCTCAGCA......TTTGGGGAGGCCAAGGCGGGCGGATCACTT....GAGGTCAGGAGTTTGAGACCAG
Alu10          GGCTGGGCATGGTGGCTCCCA..CCTGTAATCCCAGCGC.....TTTGGGGGA.CCAAGGCAGGTGAATCACTT....GAGGTCAGGAATTCGAGACCAG
Alu11          AGCTGGGCTTAGTGGCTCACT..CCTGTAATCCCAGCAC.....TTTGGGAGG.CCGAGGCAGGAGGATCGCTT....GAGCCCAGGAATTTGAGACCAG
Alu12          GGCCAGGTGTGGTGGTTCAT..GCCTGTAATCCCAGCAC.....TTTGGGAAG.CCGAGGTGGGTGGATCACCT....AAGGTCAGGAGTTTGAGACCAG

ConAlu    101  CCAGCCTGGCCAACATGGTGAAACCCCGTCTCTGCT.AAAAA..TACAAAAA...TTAGCTGGGCGTGGTGGCGCATG..CCTGTAATCCCAGCTACTCG   200
Alu1           CCTGGGCAACATGGCA....AAACCCTG.CTCTCCC.AAAAAAATACAAAAA...TTAGCTGGGCGTGCTGGTG....TGCCTGTATTCCCAGCTACTTG
Alu2           CCTGGCTAACACAGTG....AAACTCCATCTCTACT.AAAAA..TACAAAAAAAATTAGCCAGGCGTGGTGGCAGGCG..CCTGTAGTCCCAGCTACTCG
Alu3           CCTGGGCAATACGGTG....AAACCCCGTCTCTAAT.AAAAA..TATAAAAA...TTAGTCTGATGAGGTGGTGCA....CCTGTAATCCCAGCTACTCA
Alu4           CCTGGCCAACATAACAAGA....CCTCATCTCTTC..AAAAAATTTAAAAA....CCAGCCGGGCATGGTGGTGCACA..CCTGTAGTCCCAGCTACTTG
Alu5           CCTGGCCAACATGGTG....AAACCCTGTTTCTGTC.AAAAA..TACAAAAAA..TTAGCCAGGTGTGTTGGCGCGTG..CCTATAATCCCAGCTGCTCG
Alu6           CCTGACCAACATGGTG....AAACCCCGTCTCTACT.AAAAA..TACAAAAA...TTAGCCAGGCATGGTGGCAGGCG..CCTGTAATCCCAGCTATTCG
Alu7           CCTGGCCAACATGGTG....AAACCCTGTCTCTACT.AAAAA..TGCAAAAA...TTAGCCAGGTGTGGTGGCAGGCG..CCTGTAATCCCAGCTACTCG
Alu8           CCTGGCCAACATGGTG....AAACCCCATCTGTNAT.AAAAA..TACAAAAA...TTAGCTGGGCATGATAGCAGGTG..CCTGTAATTCCAGCTACTCA
Alu9           CCTGGCCAACATGGTG....AAACCCTGTCTCTATT.AAAAA..TACAAAAA...TTAGCCGGCTGTGGTAGCGAGCG..CCTATGATCCCAGCTACTCC
Alu10          CCTGGGCAACATGGCA....AAACCCTATCTCTACT.AAAA...TACAAAAA...TTAGGCAGGCGTGGTGG..CATGTGCCTGTAGTCCCAGCTACTTG
Alu11          CCTGGGCAATATAATGAG....ACTGTGTCTCT..ACAAAAA.TTAGAAAAAAA.TTAGCCAGGCATGGTGG..CATGCACCTGTAGCCCCAGTTACTTG
Alu12          CCTGGACAACATGGTG....AAACCCCGTCTCTACT.AAAAA..TACAAAAA...TTAGCCGGGCGCGGTGGCTGGCG..CCTGTAATCCCAGCTACTCG

ConAlu    201  GGAGGCTGAGGCAGGAGAA.TCGCTTGAACCCGGGAGG.CGGAGGTTG.CAGTGAGCCGAGA....TTG.GCCACTGCACTCCAGCCTGGGTAACAGAGC   300
Alu1           GGAGGCTGAGGTGGGAGGA.TTGCTTCAGCTGGGGAGGACGGAGGCTG.CAGGGAGCCAAGA....CTGAGCCATTGCAC.CCAGCCTGGGTGACAGAGC
Alu2           GGAGGCTGAGGCAGGAGAA.TGGCATGAACCCAAGAGG.CAGAGCTTG.CAGTGAGCTGAGA....TCGTGCCACTGCACTCCAGCCTGGGCGACAGAGA
Alu3           GGAAGCTGAGACACAAGAA.TCACTTTAACCCGGGAGG.TGGAGG.TGGCAGTGAGCTGAGA....TCACACCATTGCACTCCAGCCTAGGTGACAGAGT
Alu4           GGAGGCTGAGGTGGGAGGA.TCACTTGAGTCCAGCAGTT.CAAGGCTG.CAGTGAGCTATGTTTGCAC.CACCA...CACTCCAGCCTGAGTCACAGAAC
Alu5           GGAGGATGAGGTTCAAAGAATCACTTGAACCCGGGAGG.CAGAGGCTG.CAGTGAGCTGAGA....TCATGCCACTGCACTCCAGCCTGGGTGACAGAGC
Alu6           GGAGGCTGAGGCAGGAGAA.TTGCTTGAACCCCGGAGA.CAGAGGTTG.CAGCGAGCCGAGA....TCGCGCCACTGCACTCCAGCCT.GGCAATAGAGT
Alu7           GGAGGCTGAGGCAGGAGAA.TCACTTGAGCCCGGGAGGT.GGAGGTTG.CAGTGAGCCGAACGAAATTGAGCCACTTCACCCCAGCCTGGGCGAAAGAGT
Alu8           GGTGGCTGAGGTGGGAGAA.TCTATTGAACCCGGGAGGG.GGAGGTTG.CAGTGAGCCGAGA....TCATGCCATTGCACTCCAGCCTGGGCGACAGAGT
Alu9           GGAGGCTGAGGCCAGAGAA.TCACATGAACCTGGGAGG.TGGAGGCTG.CAGTGAGCCGAGA....TCACGCCACTGCACTCCAGCCTGGGGGACAGAGC
Alu10          GGAGGCTGAGGCAGGAGAA.TCACTTGAATCCAGGAGG.CAGAGGTTG.CAGTGAGCCGAGA....TCACGCTGCTGCACTCCAGCCTGGGCAATAGAGC
Alu11          GGAGACTGAGGTGGGAGGA.TCACTTGAGCCCATGAGT.TCAAGGCTG.CAGTGAGCCATGA....AGGTGCCACTGCACTCCCGCCT.GGCGACAGAGG
Alu12          GGAGGCTGAGGCAGGAGAA.TCACTTGAAGCCAGGAGG.CAGAGGTTG.CAGTGAGCTGAGA....TTGCACCACTGAACTCCAGGCTGGGTGGCAGAGC

ConAlu    301  GAGACTCC.....GTCTC.AAAAAAAA....  331
Alu1           AAGACCCT.....GTCTCTAAAAA.......
Alu2           GAGACTCC.....GTCTC.AAAAAAAAAAA
Alu3           GAGA..CC.....TGTC.AAAAAAAAAAA
Alu4           AAGA..CCTCATCTCT...AAAAAACAAACA
Alu5           AAGACTTT.....GTCTCCAAAAAAA.....
Alu6           GAGACTCCA.TC....TC.AAAACAACAACA
Alu7           GAAATTCC.A.....TTC.AAAAAAAAAA..
Alu8           GAGACTCCT.....TCTC.AAAACAAACAAA
Alu9           AAGAATCT.....GTTTC.AAAAAAAAAAA
Alu10          ATGACTCT.G.........AAGAAAAGAAAG
Alu11          GAGAC.CCT....GTCTCTAAGAAAAA....
Alu12          GAGACTCCC.....TCTC.AAAAAACAACAA
```

B.

```
Pur1        1  ..aa.aaaa.aaagaaaaAaaGaaaGaaAGaaAggaa...gAaaGAgAgaagAgaaa.aaaa.a.GAaAgAgagAgaaAgaaAgaaAgaaAaaaGAAA.   100
Pur2           ..............AAAAATGAATGAAAGGAAGGAA...GAAAGAGA...GAGAA.....AGA.GAGAGAGGAAAGAAGGAAGGAAGTAAAGAAGAAA.
Pur3           AAAAAAAAAAAAAGAAAGAGAGAGAGAAGAGAGAAAGAAAGAAGAAAGAAAA...GAAAGAGAGAGAGAGAAAGAAAGAAAGAAAGAAAGAAA.
Pur4           AAAAAAAAAAAAAGAAAAATAGGAAGGAAGGAAGGAAAGGAAAGGAAAGAAGAGAGAGAGAAA...GAAAGAGAGAGAAAGAAAGAAAGAAAGAAA.
Pur5           ............GAAGAAAAGAAAGAAAGAAAGAGA...GAGAGAGAGAAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAA..GAAA.
               ..AAGAAAAGAAAGAGAGAAAGAGAGAGAGAAAGACAGAGGAAAGAGAGAGAAAAATAAAGA.AAGAGAAAGAGAGAAAGAAAGAAAGAAAAA..GAAAA

               101  .GAAaGAAAgAAAGaaagAaa..GaAAGAaAgaaAaaGAaAaaaaa..GAAA..G.AAaGAAaGaa.g.....                        200
Pur1           .GAAAGAAAGAAAG.AGGAAAGAGGAAGAAAGAAA..GAAAAGAAA..GAAAA.GAAAAGAAAGCAAA.....
Pur2           .GAAAGAAACAAAGAAACAAA..GAAAGAAAGGAAAAGAAAAAAAAAA.....
Pur3           .GAAAGAAAGAAAGAAAGAAA..GAAAGAAA.G.AAAGAAAAA.....GAAA..GGAAAGAAAGAACGAACGAA.....
Pur4           .GAAAGAAAGAAAGACAAA....GAAAGAGAAAGAAAGAGAA......GAAAA.GAAAAGAAAGAG.....
Pur5           AGAACGAAAGAAAGAAAGAAAGAGAAAGAAAGAAA..GAAAGAAAAA.GAAA..GGAAGGAAGGAAGGAAAGAAAGGAGGGAGGGAGGGAGGGAAGGAAG

               201  .....  220
Pur1           .....  134 bp.  4 Pyrimidines
Pur2           .....  142 bp.  2 Pyrimidines
Pur3           .....  159 bp.  3 Pyrimidines
Pur4           .....  135 bp.  0 Pyrimidine
Pur5           GAAGGAAGGAAGGAAGGGGA  210 bp.  1 Pyrimidine
```

C.

5'Direct Repeat	Alu start site	[Alu]		poly A	Direct Repeat 3'
ATTAA	GGCTGG	Alu-1	Pur-1	AAAT	TTAAAGCT
TAAAGAATTCATG	AGCCGG	Alu-2		AAAA	TAAAGAATTCATGG
AAAAAAATCTAA	GGCCAGGCA	Alu-3	Pur-2	AAAA	CTAAGGCCAGGCAA
AAATCCACATATCCC	AGTTGG	Alu-4		AAAA	CCAAATCCACATATCC
no dir. repeat	GGCTGG	Alu-5		AAAA	no dir. repeat
GTCTAACTTC	AGCCGG	Alu-6		AAAA	TCGTCTAACTTC
AAGAGGATGAGGA	CTCCAG	Alu-7		AAAA	GGATGAGGACT
GAAGCTGAGGTGA	GGCCAG	Alu-8		AAAA	TACAGAAGCTGAGGCG
GAACCAGGCC	GGCTGG	Alu-9	Pur-3	CGAA	CCAGGCC
GCTTATCA	GGCTGG	Alu-10	Pur-4	AGAG	CTTATCA
no dir. repeat	GCTGGG	Alu-11		AAAA	no dir. repeat
no dir. repeat	GGCCAG	Alu-12		ACAA	no dir. repeat
ATCACTCTCA	GGTCCA	1/2 Alu	Pur-5	ACAT	ATCACTCTCA

FIG. 6. **Alignment of repetitive DNA elements found in the PAI-1 gene region and their association with each other.** The nucleotide sequence and possible alignment of *Alu* elements are presented in *A*, *Pur* elements in *B*, and *C* presents the direct repeats found flanking the *Alu* and closely associated *Alu-Pur* elements. The location of the *Alu* and *Pur* elements are shown in Fig. 3, and the numbering of these elements is consistent with that used in Fig. 3. The *top line* in *A* and *B* represents consensus sequences for the respective elements, the consensus *Alu* sequence is from Koop *et al.* (1986), while the consensus *Pur* element was derived from the alignment shown in *B*. *C* lists which of the *Alu* and *Pur* elements are found in close association; *Pur* element 5 is associated with a one-half *Alu* element (see text).

proximity to each other. Whether these *Alu* elements have any present regulatory function for PAI-1 expression is doubtful because of their presence throughout the human genome (Houck *et al.*, 1979). It has been suggested that these *Alu* elements may be important because they provide homologous DNA regions which could participate in unequal cross-over events (Jeffreys and Harris, 1982; Rogers, 1985). However, some examples of cross-overs involving *Alu* repeats have had deleterious results, such as that found in a defective low density lipoprotein receptor gene (Lehrman *et al.*, 1985). The determination of any evolutionary significance which can be assigned to any particular PAI-1 *Alu* elements will have to await the analysis of the orthologous genes of other primate and mammalian species.

Short direct terminal repeats have been found flanking many *Alu* elements (Van Arsdell *et al.*, 1981; Ruffner *et al.*, 1987), and these repeats, along with the A + T-rich composition of the surrounding DNA (Daniels and Deininger, 1985), may be important elements in controlling the targeting and orientation of integration of *Alu* elements. Several of the PAI-1 gene *Alu* elements are flanked by short perfect direct terminal repeats of at least 5 bp in length, while others are flanked by imperfect direct repeats as large as 14 bp (see Fig. 6). These direct terminal repeats do not appear to be related as they share little or no sequence identity. Direct repeats were not found for three *Alu* elements (Fig. 6), while the 3′-direct repeats for five other *Alu* elements appeared to be displaced by the presence of another sequence element, a poly purine sequence (see Figs. 3 and 6).

During the sequencing of the human PAI-1 gene, another type of DNA sequence repeat element became apparent because of its unusual sequence composition. These elements consist of long stretches of essentially pure G and A nucleotides, and they are found only on the coding strand, the same strand preference found for the *Alu* elements. We refer to these sequence elements as poly purine elements (*Pur* elements), and the PAI-1 gene contains 5 *Pur* elements which range in length from 134 to 210 bp (see Figs. 3 and 6). All of the *Pur* elements are located in intron sequences except *Pur* element 5 which is located in the 3′-flanking DNA. Alignment of these elements reveals a common core of about 120 bp (Fig. 6), which suggests that they may have a common evolutionary origin. This possibility is supported by the finding that each *Pur* element, except *Pur* element 5, overlaps with the 3′ end of an *Alu* element, see Fig. 3. A more detailed analysis of *Pur* element 5 shows that it overlaps with a one-half *Alu* element, positions 12,907–13,023 (Fig. 3). The close association of these *Alu* and *Pur* elements suggests a common link in their integration mechanism. Such a common link could be: (*i*) the A-rich composition of the 5′-end of the *Pur* element may be a target for *Alu* element integration, (*ii*) the A + T-rich composition of the 3′-end of the *Alu* element may be a target for *Pur* element integration, or (*iii*) these *Pur* elements may represent a different type of *Alu* repeat element which may have been deposited, as a single event, in the human genome by the *Alu* retroposition mechanism. If so, this would indicate that in some cases retroposition of *Alu* elements may involve more than the *Alu* element itself, the possible retroposition of sequence adjacent to an ancestral *Alu* element, in this case an adjacent *Pur* element. This information may be useful for the assignment of *Alu* element subclasses which may provide additional insight into their evolution. The degree of divergence among the individual *Alu* elements found in close association with *Pur* elements ranges between 89 and 77%, essentially the same range as found for the *Alu* elements (Kariya *et al.*, 1987). This finding suggests that these *Alu-Pur* elements are not the result of a single burst of retroposition of an ancestral *Alu-Pur* element. Additional insight in the evolution of these *Alu-Pur* elements and into the mechanism responsible for their close association can possibly be obtained with the determination of PAI-1 gene sequences from related species.

A search of GenBank sequences for the *Pur* element showed that this element is present much less frequently than the highly repetitive *Alu* sequence elements. This is consistent with our hypothesis that *Alu-Pur* elements represent a small subset of *Alu* elements. *Pur* elements in the range of at least 100 bp were found in the sequence of only a few genes, the human apolipoprotein CIII (Protter *et al.*, 1984), in a human actin pseudogene (Moos and Gallwitz, 1983), and even in DNA isolated from an ancient Egyptian mummy (Pääbo, 1985), but none are found in the sequenced t-PA gene region (Friezner Degen *et al.*, 1986). It is extremely interesting to note that the *Pur* elements found in the apolipoprotein CIII gene and the mummy DNA are also associated with an overlapping *Alu* element. All of these *Alu-Pur* element overlaps occur in the same orientation as those presented in Fig. 3, with respect to the *Alu* poly(A) region. This finding supports these *Alu-Pur* elements being a small subclass of *Alu* elements and supports the hypothesis that the *Alu-Pur* elements may have been incorporated in the genome as a result of a single retroposition event. Additional experiments, such as genomic blot hybridizations and copy number determination, are needed to accurately determine the repetitive nature of the *Pur* or *Alu-Pur* element. As in the case with *Alu* elements, it is doubtful that these *Alu-Pur* elements are important for the regulation of the PAI-1 gene, but this possibility certainly deserves to be tested.

Acknowledgment—We are grateful to Dr. J. Hoeymakers, Erasmus University, Rotterdam, for his gift of the human cosmid library which facilitated our isolation of the complete PAI-1 gene.

REFERENCES

Adams, J. W., Kaufman, R. E., Kretschmer, P. J., Harrison, M., and Nienhuis, A. W. (1980) *Nucleic Acids Res.* **8,** 6113–6128

Adrian, G. S., Korinek, B. W., Bowman, B. H., and Yang, F. (1986) *Gene (Amst.)* **49,** 167–175

Andreasen, P. A., Pyke, C., Riccio, A., Kristensen, P., Nielsen, L. S., Lund, L. R., Blasi, F., and Danø, K. (1987) *Mol. Cell. Biol.* **7,** 3021–3025

Baumann, H., and Eldredge, D. (1982) *J. Cell Biol.* **95,** 29–40

Bosma, P. J., van den Berg, E. A., and Kooistra, T. (1987) *Thromb. Haemostas.* **58,** 445 (abstr.)

Breathnach, R., Benoist, C., O'Hare, K., Gannon, F., and Chambon, P. (1978) *Proc. Natl. Acad. Sci. U. S. A.* **75,** 4853–4857

Carrell, R. W., and Boswell, D. R. (1986) in *Proteinase Inhibitors* (Barret, A. J., and Salvesen, G., eds) pp. 403–420, Elsevier Science Publishers, Amsterdam

Chang, L.-Y., and Slightom, J. L. (1984) *J. Mol. Biol.* **180,** 767–784

Colucci, M., Paramo, J. A., and Collen, D. (1985) *J. Clin. Invest.* **75,** 818–824

Cooper, R., Eckley, D. M., and Papaconstantinou, J. (1987) *Biochemistry* **26,** 5244–5250

Daniels, G. R., and Deininger, P. L. (1985) *Nucleic Acids Res.* **13,** 8939–8954

Danø, K., Andreasen, P. A., Grøndahl-Hansen, J., Kristensen, P., Nielsen, L. S., and Skriver, L. (1985) *Adv. Cancer Res.* **44,** 139–266

Devereux, J., Haeberli, P., and Smithies, O. (1984) *Nucleic Acids Res.* **12,** 387–395

Dynan, W. S., and Tjian, R. (1985) *Nature* **316,** 774–778

Efstratiadis, A., Posakony, J. W., Maniatis, T., Lawn, R. M., O'Connell, C., Spritz, R. A., DeRiel, J. K., Forget, B. G., Weissman, S. M., Slightom, J. L., Blechl, A. E., Smithies, O., Baralle, F. E., Shoulders, C. C., and Proudfoot, N. J. (1980) *Cell* **21,** 653–668

Emeis, J. J., and Kooistra, T. (1986) *J. Exp. Med.* **163,** 1260–1266

Emeis, J. J., Brommer, E. J. P., Kluft, C., and Brakman, P. (1985) in *Recent Advances in Blood Coagulation* (Poller, L., ed) Vol 4, pp. 11–

13, Churchill-Livingstone, Edinburgh
Erickson, L. A., Hekman, C. M., and Loskutoff, D. J. (1985) *Proc. Natl. Acad. Sci. U. S. A.* **82,** 8710–8714
Fowlkes, D. M., Mullis, N. T., Comeau, C. M., and Crabtree, G. R. (1984) *Proc. Natl. Acad. Sci. U. S. A.* **81,** 2313–2316
Friezner Degen, S. J., Rajput, B., and Reich, E. (1986) *J. Biol. Chem.* **261,** 6972–6985
Gelehrter, T. D., and Sznycer-Laszuk, R. (1986) *J. Clin. Invest.* **77,** 165–169
Gelehrter, T. D., Barouski-Miller, P. A., Coleman, P. L., and Cwikel, B. J. (1983) *Mol. Cell. Biochem.* **53/54,** 11–21
Ginsburg, D., Zeheb, R., Yang, A. Y., Rafferty, U. M., Andreasen, P. A., Nielsen, L., Dano, K., Lebo, R. V., and Gelehrter, T. D. (1986) *J. Clin. Invest.* **78,** 1673–1680
Grosveld, F. G., Lund, T., Murray, E. J., Mellor, A. L., Dahl, H. H. M., and Flavell, R. A. (1982) *Nucleic Acids Res.* **10,** 6715–6732
Gruss, P. (1984) *DNA (N.Y.)* **3,** 1–5
Hanss, M., and Collen, D. (1987) *J. Lab. Clin. Med.* **109,** 97–104
Houck, C. M., Rinehart, F. P., and Schmid, C. D. (1979) *J. Mol. Biol.* **132,** 289–306
Hwu, H. R., Roberts, J. W., Davidson, E. H., and Britten, R. J. (1986) *Proc. Natl. Acad. Sci. U. S. A.* **83,** 3875–3879
Jeffreys, A. J., and Harris, S. (1982) *Nature* **296,** 9–10
Juhan-Vague, I., Aillaud, M. F., De Cock, F., Philip-Joet, C., Arnaud, C., Serradimigni, A., and Collen, D. (1985) in *Progress in Fibrinolysis VII* (Davidson, J. F., Donati, M. B., and Coccheri, S., eds) pp. 146–149, Churchill-Livingstone, Edinburgh
Kariya, Y., Kato, K., Hayashizaki, Y., Himeno, S., Tarui, S., and Matsubara, K. (1987) *Gene (Amst.)* **53,** 1–10
Kiyami, R., Matsui, H., and Oishi, M. (1986) *Proc. Natl. Acad. Sci. U. S. A.* **83,** 4665–4669
Kluft, C., Verheijen, J. H., Jie, A. F. H., Rijken, D. C., Preston, F. E., Sue-Ling, H. M., Jespersen, J., and Aasen, A. O. (1985a) *Scand. J. Clin. Lab. Invest.* **45,** 605–610
Kluft, C., Verheijen, J. H., Rijken, D. C., Chang, G. T. G., Jie, A. F. H., and Onkelinx, C. (1985b) in *Progress in Fibrinolysis VII* (Davidson, J. F., Donati, M. B., and Coccheri, S. eds) pp. 117–118, Churchill Livingstone, Edinburgh
Knudsen, B. S., Harpel, P. C., and Nachman, R. L. (1987) *J. Clin. Invest.* **80,** 1082–1089
Koop, B. F., Miyamoto, M. M., Embury, J. E., Goodman, M., Czelusniak, J., and Slightom, J. L. (1986) *J. Mol. Evol.* **24,** 94–102
Kozak, M. (1987) *Nucleic Acids Res.* **15,** 8125–8148
Lehrman, M. A., Schneider, W. J., Sudhof, T. C., Brown, M. S., Goldstein, J. L., and Russell, D. W. (1985) *Science* **227,** 140–146
Lizardi, P. M., and Engelberg, A. (1979) *Anal. Biochem.* **98,** 116–122
Loskutoff, D. J., Linders, M., Keijer, J., Veerman, H., van Heerikhuizen, H., and Pannekoek, H. (1987) *Biochemistry* **26,** 3763–3768
Maniatis, T., Fritsch, E. F., and Sambrook, J. (1982) *Molecular Cloning, A Laboratory Manual,* Cold Spring Harbor Laboratory, Cold Spring Harbor, NY
Maniatis, T., Goodbourn, S., and Fischer, J. A. (1987) *Science* **236,** 1237–1245
Maxam, A. M., and Gilbert, W. (1980) *Methods Enzymol.* **65,** 499–560
Medcalf, R. L., van den Berg, E. A., and Schleuning, W.-D. (1987) *J. Cell Biol.* **106,** 971–978
Miyamoto, M. M., Slightom, J. L., and Goodman, M. (1987) *Science* **238,** 369–373
Moos, M., and Gallwitz, D. (1983) *EMBO J.* **2,** 757–761
Mount, S. M. (1982) *Nucleic Acids Res.* **10,** 459–472
Naora, H., and Deacon, N. J. (1982) *Proc. Natl. Acad. Sci. U. S. A.* **79,** 6196–6200
Ny, T., Sawdey, M., Lawrence, D., Millan, J. L., and Loskutoff, D. J. (1986) *Proc. Natl. Acad. Sci. U. S. A.* **83,** 6776–6780
Pääbo, S. (1985) *Nature* **314,** 644–645
Pannekoek, H., Veerman, H., Lambers, H., Diergaarde, P., Verweij, C. L., van Zonneveld, A.-J., and van Mourik, J. A. (1986) *EMBO J.* **5,** 2539–2544
Pöllänen, J., Saksela, O., Salonen, E.-M., Andreasen, P., Nielsen, L., Danø, K., and Vaheri, A. (1987) *J. Cell Biol.* **104,** 1085–1096
Prochownik, E. V. (1985) *Nature* **316,** 845–848
Protter, A. A., Levy-Wilson, B., Miller, J., Bencen, G., White, T., and Seilhammer, J. J. (1984) *DNA (N.Y.)* **3,** 449–456
Proudfoot, N. J., and Brownlee, G. G. (1976) *Nature* **263,** 211–214
Reich, E. (1978) in *Molecular Basis of Biological Degradative Processes* (Berlin, R. D., Herrmen, H., Lepow, I. H., and Tanzer, J. M., eds) pp. 155–169, Academic Press, New York
Rheinwald, J. G., Jorgensen, J. L., Hahn, W. C., Terpstra, A. J., O'Connell, T. M., and Plummer, K. K. (1987) *J. Cell Biol.* **104,** 263–275
Rogers, J. (1985) *Nature* **317,** 559
Ruffner, D. E., Sprung, C. N., Minghetti, P. P., Gibbs, P. E. M., and Dugaiczyk, A. (1987) *Mol. Biol. Evol.* **4,** 1–9
Sharp, P. A. (1981) *Cell* **23,** 643–646
Slightom, J. L., Blechl, A. E., and Smithies, O. (1980) *Cell* **21,** 627–638
Slightom, J. L., Drong, R. F., Klassy, R. C., and Hoffman, L. M. (1985) *Nucleic Acids Res.* **13,** 6483–6498
Slightom, J. L., Theisen, T. W., Koop, B. F., and Goodman, M. (1987) *J. Biol. Chem.* **262,** 7472–7483
Sollner-Webb, B., and Reeder, R. H. (1979) *Cell* **18,** 485–499
Sprengers, E. D., and Kluft, C. (1987) *Blood* **69,** 381–387
Ullu, E., and Tschudi, C. (1984) *Nature* **312,** 171–172
Van Arsdell, S. W., Denison, R. A., Bernstein, L. B., Weiner, A. M., Manser, T., and Gesteland, R. F. (1981) *Cell* **26,** 11–17
van den Berg, E. A., Sprengers, E., Burgess, W., and Van Hinsbergh, V. W. M. (1987) *Thromb. Haemostas.* **58,** 15 (abstr.)
Van Hinsbergh, V. W. M., Kooistra, T., Fiers, W., and Emeis, J. J. (1987a) *Thromb. Haemostas.* **58,** 15 (abstr.)
Van Hinsbergh, V. W. M., Sprengers, E. D., and Kooistra, T. (1987b) *Thromb. Haemostas.* **57,** 148–153
Verheijen, J. H., Rijken, D. C., Chang, G. T. G., Preston, F. E., and Kluft, C. (1984) *Thromb. Haemostas.* **51,** 396–397
Weaver, R. F., and Weissmann, C. (1979) *Nucleic Acids Res.* **7,** 1175–1193
Wun, T. C., and Kretzmer, K. K. (1987) *FEBS Lett.* **210,** 11–16