Multiple Members of the Plasminogen–Apolipoprotein(a) Gene Family Associated with Thrombosis†,‡

Akitada Ichinose
Department of Biochemistry, University of Washington, SJ-70, Seattle, Washington 98195
Received October 24, 1991; Revised Manuscript Received January 6, 1992

ABSTRACT: Plasminogen and apolipoprotein(a) [apo(a)] are closely related plasma proteins that are associated with hereditary thrombophilia. Low plasminogen levels are found in some patients who developed venous thrombosis, while a population with high plasma concentrations of apo(a) have a higher incidence of arterial thrombosis. Two different genes coding for human apo(a) have been isolated and characterized in order to study and compare these genes with four other closely related genes in the plasminogen–apo(a) gene family. These include the gene coding for plasminogen, two unique plasminogen-related genes, and a gene coding for hepatocyte growth factor. Nucleotide sequence analysis of these genes revealed that the exons and their boundaries of the genes for plasminogen and apo(a), and the plasminogen-related genes, differ only 1–5% in sequence. The types of exon/intron junctions and positions of introns in the molecules are also exactly identical, suggesting that these genes have evolved from an ancestral plasminogen gene via duplication and exon shuffling. By utilizing these results, gene-specific probes have been designed for the analysis of each of the genes in this gene family. The plasminogen and two apo(a) genes were all localized to chromosome 6 by employing the gene-specific primers and genomic DNAs from human–hamster cell hybrids. These data also make it possible to characterize the apo(a) and plasminogen genes in individuals by in vitro amplification.

Plasminogen is a key proenzyme in the fibrinolytic and thrombolytic systems. It contains a preactivation peptide, five kringle domains, and a serine protease domain. Apolipoprotein(a) [apo(a)][1] is a plasma glycoprotein which is present in a low-density lipoprotein (LDL)-like particle named lipoprotein(a) (Berg, 1963; Utermann, 1989; Scanu & Fless, 1990; Miles & Plow, 1990). Apo(a) shares homologous domains with plasminogen including a signal peptide, as many as 37 repeats of plasminogen kringle 4, a single plasminogen kringle 5, and a serine protease domain (McLean et al., 1987).

In the course of the study of characterization of normal and abnormal genes coding for human plasminogen (Petersen et al., 1990; Ichinose et al., 1991), several genes that are highly homologous to the gene for plasminogen have been found. In order to study and identify these closely related genes in the plasminogen–apo(a) gene family, their nucleotide sequences of regions corresponding to the exons and their boundaries of the plasminogen gene have been determined and compared to each other. Since cDNA probes of this gene family cross-hybridize and oligonucleotides designed from cDNAs anneal to other genes, it is essential to differentiate one gene from another.

Experimental Procedures

Southern Blot Hybridization of Genomic DNAs with a Radiolabeled cDNA Coding for Plasminogen. Venous blood was drawn after informed consent had been obtained from normal individuals and patients with abnormal plasminogens (Ichinose et al., 1991). Genomic DNA samples were prepared from the leukocytes by a standard technique (Bell et al., 1981). Five to ten micrograms of genomic DNA was digested with 20 units of *Fnu*4H1 or *Eco*RI endonuclease at 37 °C for 1.5 h and electrophoresed in a 0.4% agarose gel. After transfer to a nylon membrane (GeneScreen Plus, NEN Research Products), genomic DNA fragments were hybridized overnight with a ^{32}P-labeled cDNA probe, a *Pst*I–*Stu*I 610 bp fragment of the cDNA for plasminogen (Malinowski et al., 1984), at 65 °C, and the membrane was washed in 0.1 × SSC and 0.1% SDS at 70 °C.

Screening of Genomic Libraries and Sequencing of the Plasminogen–Apo(a) Gene Family. Genomic libraries for the patients with abnormal plasminogens (Ichinose et al., 1991) were constructed from the 5.0–7.0-kb *Eco*RI fragments electroeluted from 0.4% agarose gel, using λgtWES *Eco*RI arms (Bethesda Research Laboratories) and the Gigapack Gold packaging kit (StrataGene). Four other genomic libraries were obtained from Dr. Maniatis (Lawn et al., 1978), Dr. Yoshitake (Yoshitake et al., 1985), Clontech Lab. Inc., and StrataGene. Human genomic libraries were screened by in situ hybridization using the *Pst*I–*Stu*I cDNA fragment (Malinowski et al., 1984). Phage DNA was prepared by liquid culture lysis (Silhavy et al., 1984). Genomic DNA inserts were isolated by digestion of the phage DNA with *Eco*RI and/or *Sal*I endonuclease and then subcloned into plasmid pUC18 or pUC19. DNA fragments were also subcloned into M13mp18 or M13mp19 and then sequenced by the dideoxy termination method (Sanger et al., 1977) employing [^{35}S]-dATP and buffer gradient gels (Biggin et al., 1983). DNA sequences were analyzed by the Genepro program (version 4.1, Riverside Scientific Enterprises, Seattle, WA) employing an IBM PC-compatible computer.

Phage clones were also isolated employing the radiolabeled DNA fragment for exon I of the plasminogen gene. This DNA fragment was prepared by in vitro amplification of human leukocyte DNA (Saiki et al., 1988) employing 2.5–5.0 units of Taq DNA polymerase (New England Biolabs or Perkin Elmer Cetus) and oligonucleotides 5′-GTAACGTTT-

† This work was supported in part by a research grant from the NIH (HL 16919).

‡ The nucleotide sequence in this paper has been submitted to GenBank under Accession Number J05365.

[1] Abbreviations: apo(a), apolipoprotein(a); PRG, plasminogen-related gene.

0006-2960/92/0431-3113$03.00/0

GAACCCTGCTGAGCC [nucleotides −36 to −13 in Petersen et al. (1990) underlined] and 3′-CACAGAATTCCATGGCATATGTATTTTTACTAC (nucleotides 220–243) as primers. Oligonucleotides were prepared with a synthesizer purchased from Applied Biosystems. After 30 cycles of amplification, the reaction mixture was applied to a 1.6% agarose gel and electroeluted.

In Vitro Amplification of the Plasminogen and Apo(a) Genes. The genomic DNAs were amplified as described above by employing primers [listed in Pertersen et al. (1990)] designed from the cDNA and genomic sequences for plasminogen: for the first and second halves of the preactivation peptide region, 5′-TAGTAAGCTTCTTTATTTATGTCCAAATGCCCG (172–195 nucleotides upstream from exon II) and 3′-TATTAAGCTTACCGTTTGCTTGCTACTTGTAA (the 3′-end of the sequence of intron C); for the active Asp region, 5′-AAGCTTACCTGAAGGGGCTGGACCAT (32–53 nucleotides upstream from exon XVI) and 3′-AAGCTTGGGGTGAAGACCACAGGTGA (9–29 nucleotides downstream from exon XVI); for the so-called Met loop and the active site Ser and 3′-noncoding regions, 5′-GGTAGTCAAGAGGAGCTTCCTCCCTGCAGC (the 30 nucleotides of intro Q) or 4′-GTACTTTTGGAGCTGGCCTTCTCAAGGAAG (nucleotides 1–30 of exon XVIII) and 3′-GTTGAAGACCAAAACCAAAATTTACTCAAA (1–30 nucleotides downstream from the first polyadenylation site). The genomic DNAs were also amplified by employing primers designed from the cDNA sequence encoding apo(a): for the second and first halves of A and B types of kringle 4, 5′-CTGCAAGCTTCCAGATGCTGTGGCAGCTCCTTATTG [see McLean et al. (1987)] and 3′-AGCCAAGCTTTGGGTAGTATTCTGGGGTCCGACTATG; for the second half of the 37th kringle 4 and the first and second halves of kringle 5, 5′-AGATAAGCTTACAGGCCCTTGGTGTTTTACCATG- and 3′-GAGGGAATTCAGAGAGGGATATCACAGTAGTCAA; for the second half of kringle 5 and the activation site region, 5′-AAAAGGATCCTGCCGTAACCCTGATGGTGACATCAA and 3′-TGGGGGATCCACACACCCCCCTACAATGCTTCCA; for the activation site and active site His and Asp regions, 5′-CTCAGGATCCATCCTCTTCATTTGATTGTGGGAAG and 3′-TGCTAAGCTTTAGCAAGGCAATATCTGCTTG; the active site Asp and the connecting region between the A and B chains, 5′-CTGGGAATTCACCAAGAAGTGAACCTCGAATCTCA and 3′-AAGGAAGCTTGGGTTTCTCCCCAGCCAGTGATGT. A primer, 5′-CCCAGAATTCCTTTGGGACTGGCCTTCTCAAGG, was also used with the 3′-primer for the 3′-noncoding region of the plasminogen gene as listed above.

Chromosomal Localization of Plasminogen and Apo(a) Genes. The gene-specific primers were designed from the unique nucleotide sequences of each gene: for plasminogen, 5′-CAATGAATTCTGGTCCTTATTTCACATCTAAAA [nucleotides −851 to −826 in Petersen et al. (1990)] and 3′-CACAGAATTCCATGGCATATGTATTTTTACTAC (nucleotides 220–243); for apo(a)I, 5′-CCAGGATCCAGCATCTATTGACATTGCACT (nucleotides −90 to −69 in Figure 3) and 3′-TTTGAATTCATATACAAGATTTTGAACTGGGAA (nucleotides 285–310); for apo(a)II, 5′-ATCGGATCCTGAGGGATGGTTGGATTCAGAAA (nucleotides −80 to −57 in Figure 3) and 3′-TCCGAATTCTTATACAAGATTTTGAACTGGCGA (nucleotides 285–310). The genomic DNA samples (0.2 μg) from human, hamster, and the human–hamster hybrid cell lines (Panel I, Bios Corp., CT) were amplified with the above primers. The amplified products (10 μL) were electrophoresed at 80 V for 2 h in a 0.8 or 1.6% agarose gel. Buffer and water were also added as the negative controls without genomic DNA ("None" and "H_2O", respectively, in Figure 6). Another set of cell hybrids (Panel II, Bios Corp., CT) was also examined by the same method as Panel I. The DNA sequences of the amplified products were determined by the dideoxy method (Sanger et al., 1977).

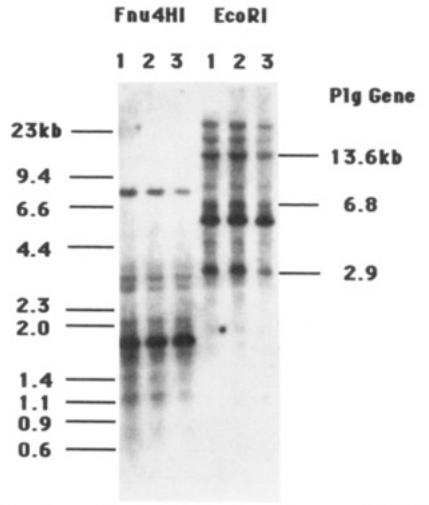

FIGURE 1: Southern blot hybridization of genomic DNAs with a radiolabeled DNA for human plasminogen. Each lane contains genomic DNA from a normal individual (lanes 1) and patients with abnormal plasminogens (lanes 2, PLG Nagoya I; lanes 3, PLG Kagoshima) digested with either *Fnu*4HI (on the left) or *Eco*RI (on the right) endonuclease. Only *Eco*RI fragments of 13.6, 6.8, and 2.9 kb were expected from the plasminogen gene, as indicated for the right panel.

Results and Discussion

Genomic Southern Blotting Analysis. Three *Eco*RI fragments of 13.6, 6.8, and 2.9 kb are predicted from the gene structure for plasminogen (Figure 1; Petersen et al., 1990) by Southern blot analysis; however, numerous bands of *Eco*RI fragments from human DNA hybridized with a probe from the cDNA coding for plasminogen. This probe included the kringle 4 region of the cDNA. It is very likely that the intense band of approximately 6.0 kb represents at least in part the multiple copies of the kringle 4 repeats in the apo(a) genes. These data suggested the presence of multiple homologous genes.

Identification of Apo(a) Genes. Because of the presence of the multile DNA fragments shown by Southern blotting analysis, it was not surprising that eight of the twelve phage clones originally isolated from the 5.0–7.0-kb genomic libraries employing the *Pst*I–*Stu*I fragment also contained DNA sequences coding for the kringle 4 regions of apo(a) (Figure 2). When exons XVI, XVIII, and XIX of the plasminogen gene were amplified, approximately one-third of the DNA sequences were found to code for the exons of apo(a) (data not shown).

Exons corresponding to the kringle 4 repeats, kringle 5, and the serine protease domain of apo(a) were also obtained by in vitro amplification employing specific oligonucleotides designed from the apo(a) cDNA (McLean et al., 1987). Furthermore, two genomic sequences for apo(a) (types I and II) were identified when clones containing the 5′-region were isolated from human genomic libraries employing the DNA fragment of exon I of the plasminogen gene. These genes share an almost identical sequence with exon I that codes for the

```
PLG     TAAGCTTGAT TTCTTTTATT TTAATTTCAG [ CACCACCTGA GCTAACCCCT GTGGTCCAGG ACTGCTACCA TGGTGATGGA CAGAGCTACC GAGGCACATC CTCCACCACC
4a-1    *C******** *****C*G** *******T** [ **G******* **A**G**A* ********** *T******** ********** *****T**T* *******G*A *********T
4a-A    *C******** ***CCG*G** **C******* [ ****GA**** **A**GG*** *G***G**** *G******** ****A***** *****T**T* *********A *********T
4a-B    *C**-**T** *****G*G** **C******* [ ****GA**** **AG*GG*** *G***G**** *G******** C***A***** *****T**T* *********A *********T
4a-30   *C**-**T** *****G*G** **C******* [ ****GA**** **AG*GG*** *G***G**** *G******** C***A***** *****T**T* *********A *********T
4a-31   *C**T***** *******C** **C******* [ *****A**** **A**GG*** *G***G**** *G******** C**AA***** *****T**T* A********A **T**TT**T

PLG     ACCACAGGAA AGAAGTGTCA GTCTTGGTCA TCTATGACAC CACACCGGCA CCAGAAGACC CCAGAAAACT ACCCAAATGC ] GTATGTCTTT GATTTTTACT GTAAGAGGGG
4a-1    GT******** G**CC**C** AG******** ********** ****T*AA** TA*T*G**** A********* ********** ] *******A** A**C**ACAG TA*GC*AAAC
4a-A    GT******** GA*CC**C** AG******** ********** *****TC*** TAGTCG**** ******T*** ********** ] ********** *T*C*****C A******AA*
4a-B    GT***T**** GA*CC**C** AG******** ********** *****TC*** TAGTCG**** ******T*** ********** ] ********** *T*C*****C A******AA*
4a-30   GT***T**** GA*CC**C** AG******** ********** *****TC*** TAGTCG**** ******T*** ********** ] ********** *T*C*****C A******AAT
4a-31   GT******** GA*CC**C** AG******** ********** *****TC*** TAGTCG**** ****C*T*** ********** ] ********A* TT*C*****C A****T*AA*
```

FIGURE 2: Nucleotide sequence of kringle 4 and the flanking regions of the plasminogen–apo(a) gene family. PLG, 4a-1, 4a-A, 4a-B, 4a-30, and 4a-31 represent the genomic sequences, respectively, for the first halves of the following: the kringle 4 region of plasminogen and the first, A type (2nd–23rd, 25th, and 26th), B type (24th, 27th–29th), 30th, and 31st kringle 4 repeats [numbered as in McLean et al. (1987)] in apo(a). Asterisks show nucleotides identical to those of the plasminogen gene. The 5′- and 3′-ends of exons are enclosed in brackets.

```
PLG         AGCATCTATT GCAGATTCCA CCCTCAAACA TTTTGTAAGG ACTCTTTATT CAAGGTAACG TTTGAACCCT GCTGAGCCAG TGGCATGGGT [ CTCTGAGAGA     10
PRG A       ********** ********** *******T** ********** **C******* ******T*T* ********** ********** ****G***** G*********
PRG B       ********** ********** *******T** ********** **C******* ******CGT* ********** ********** ****G***** G*********
Apo(a) I    ********** *-*C***G** *T******T** ****A***-* *****A**** ********T* ********** ********** **********  **********
Apo(a) II   TCTTAGGGGA CTGAGGGATG GTTGG*TT** GAAAAGGGCT T**AAAA*** *-*******T* *******T** ********** **********  **********

PLG         ATCATTAACT TAATTTGACT ATCTGGTTTG TGGATGCGTT TACTCTCATG TAAGTCAACA ACATCCTGGG ATTGGGACCC ACTTTCTGGG CACTGCTGGC     110
PRG A       *******C** ********** *******C** ********** ********** ********** ********** ********A* ********** **********
PRG B       ********** ********** *******C** ********** ********** ********** ********** ********A* ********** **********
Apo(a) I    ********** ********** ********** ***G****** ********** ********** **G******* ********A* ********** **********
Apo(a) II   ********** ********** ********** ********** ********** ********** ********** ********A* ********** **********

PLG         CAGTCCCAAA ATGGAACATA AGGAAGTGGT TCTTCTACTT CTTTTATTTC TGAAATCAG ] G TAAGACATAG TTTTTTTAAA TTATAATAAT TATTTTTTCT  210
PRG A       ********** ********** ********** ********** ********** ********* ] * ********** ********** ********** **********
PRG B       ********** ********** ********** ********** ********** ********* ] * ********** ********** ********** **********
Apo(a) I    ********** ********** ********** ********** ********** ********* ] * ********** ********** ******G*** **********
Apo(a) II   ********** ********** ********** ********** **G******* ********* ] * ********** *******TG- -*G******* **********

PLG         CCCACAATGT AGTAAAAATA CATATGCCAT GGCTTTATGT GCAATTC--- --ATTTAATT TTTGATTCAT GAAACTTCCA GTTGAAAATC TTGTATAAGA     310
PRG A       *********C ****C***** ********** ********** *******--- --********* ********** ******C*** ********** **********
PRG B       *********C ****C***** ********** ********** *******--- --********* ********** ******C*** ********** **********
Apo(a) I    ********** ********** ********** ********** *******--- --********* ********** ****T*C*** ***C****** ********T**
Apo(a) II   ********** ********** ********** AT******** *******ATT TA********* ********** A***TCG*** ***C****** **********
```

FIGURE 3: Nucleotide sequence of exon I and the flanking regions of the plasminogen–apo(a) gene family. PLG, PRG A, PRG B, Apo(a)I, and Apo(a)II represent the sequences of the gene for plasminogen, type A and B plasminogen-related genes, and type I and II genes for apo(a), respectively. Asterisks show nucleotides identical to those of the plasminogen gene. The 5′- and 3′-ends of exon I are enclosed in brackets. Regulator elements for expression, CCAAT boxes and TATAA sequences, are underlined. Nucleotide numbers are shown in the right margin.

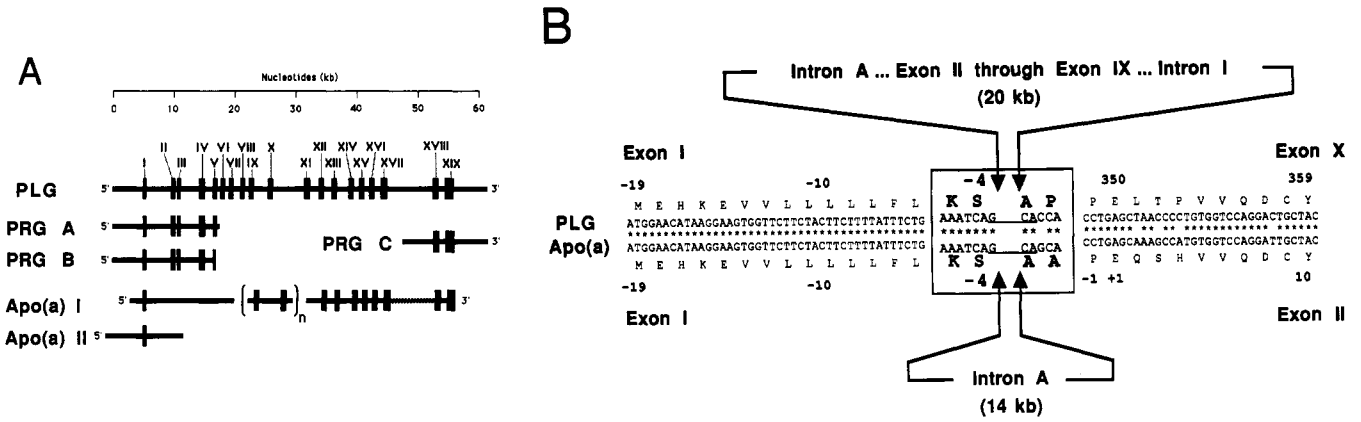

FIGURE 4: (A) Organization of the genes for the members in the plasminogen–apo(a) gene family. Each gene model is based on isoalted genomic phage clones except for the two 3′-exons of apo(a). All exons and their boundaries including the two 3′-exons of apo(a) were obtained also by in vitro amplification of genomic DNAs. Nucleotide sequences of regions corresponding to the exons and their boundaries of the plasminogen gene were actually determined by the dideoxy method. Exons are shown by wide vertical bars, and those for plasminogen are numbered with Roman numerals. The dotted line denotes an intron of unknown size. The "n" represents the number of kringle 4 repeats in apo(a), which varies among individuals up to 37. (B) Alignment of the intron/exon boundaries for the plasminogen and apo(a) genes. Introns A–I and exons II–IX were inserted between Ser at position −4 and Ala at position 347 in the plasminogen gene, while Ser at position −4 and Ala at −3 are separated by intron A in the apo(a)I gene.

signal peptide. The nucleotide sequences about 45 nucleotides upstream from the cytosine at position 1 [as defined in Petersen et al. (1990)], however, differed significantly (Figure 3). Thus far, 24 individuals have been examined for the presence of apo(a) type I and II genes, employing gene-specific primers for the chromosomal localization, and all were found to have both genes (data not shown).

At the moment, it is not clear whether either one or both of the apo(a) genes are expressed. The regulation of the type II gene may differ from the type I gene since it lacks a typical TATAA box. The type I gene, however, has typical regulator elements for expression (Figure 3). The organization of the human apo(a) gene in terms of the location of introns and type of splice junctions was identical to the gene coding for plasminogen (Figure 4A). All kringle 4 repeats in apo(a) were coded by two exons. Furthermore, the splice junctions of the exons coding for the signal peptide and the first kringle 4 repeat of apo(a) were located in exactly the same position as the splice junction for kringle 4 of plasminogen (Figure 4B). The plasminogen gene contains 20 kb of DNA between the signal peptide and kringle 4 (introns A–I and exons II–IX) (Petersen et al., 1990) while the gene for apo(a) has a single intron A of 14 kb. Therefore, apo(a) lacks the preactivation peptide and kringles 1–3 that are present in the plasminogen gene.

A

```
PLG GENE   ATTTATTCCA CTTGGATCTC CCACCTCTAG [ GTCAAGGAGA GCCTCTGGAT GACTATGTGA ATACCCAGGG GGCTTCACTG TTCAGTGTCA CTAAGAAGCA
PRG A      ********** ********** ********** [ ********** ********** ********A* **G******* ********** ********** **********
PRG B      ********** ********** ********** [ ********** ***C****** ********** ********** *C********* ********** **********

PLG GENE   GCTGGGAGCA GGAAGTATAG AAGAATGTGC AGCAAAATGT GAGGAGGACG AAGAATTCAC CTGCAG ] GTAT TTCCATTGTC GTTGCACCTA CGCAGGAATC
PRG A      ******G*** *****C*G** ********** ********** *********A ********** ****** ] **** **T***C*** ********** *C***A**** 
PRG B      ******G*** *****C*G** ********** ********** **A******A ********** ****** ] **** **T***C*** ********** *C***A****
```

B

```
Plg.GENE   TTCTTTAAAT AAAGAAAAAT ACTTATTGGA TTTCCTGCTT CGTTCTGCAG [ GGCATTCCAA TATCACAGTA AAGAGCAACA ATGTGTGATA ATGGCTGAAA
PRG A      ********** ********** ********** ********** ********** [ ********** ********** ********** ********** **********
PRG B      ********** ********** ********** ********** ********** [ ********** ********** ********** G********* **********

Plg.GENE   ACAGGAAGTC CTCCATAATC ATTAGGATGA GAGATGTAGT TTTATTTGAA AAGAAAG ] GTG AGTACATTTT CTTCCTCCTC CTCCTACTGT CCTCCCCATC
PRG A      ***A****** ********** ********** ********** ********** ***T**A ] *** ********** ********** ********** **********
PRG B      ********** ********** ********** ******C*** ********** ***T**A ] *** ********** ********** ********** **********
```

FIGURE 5: Nucleotide sequence of exons II (A) and III (B) and their flanking regions of the plasminogen gene and PRGs. An in-frame stop codon is underlined.

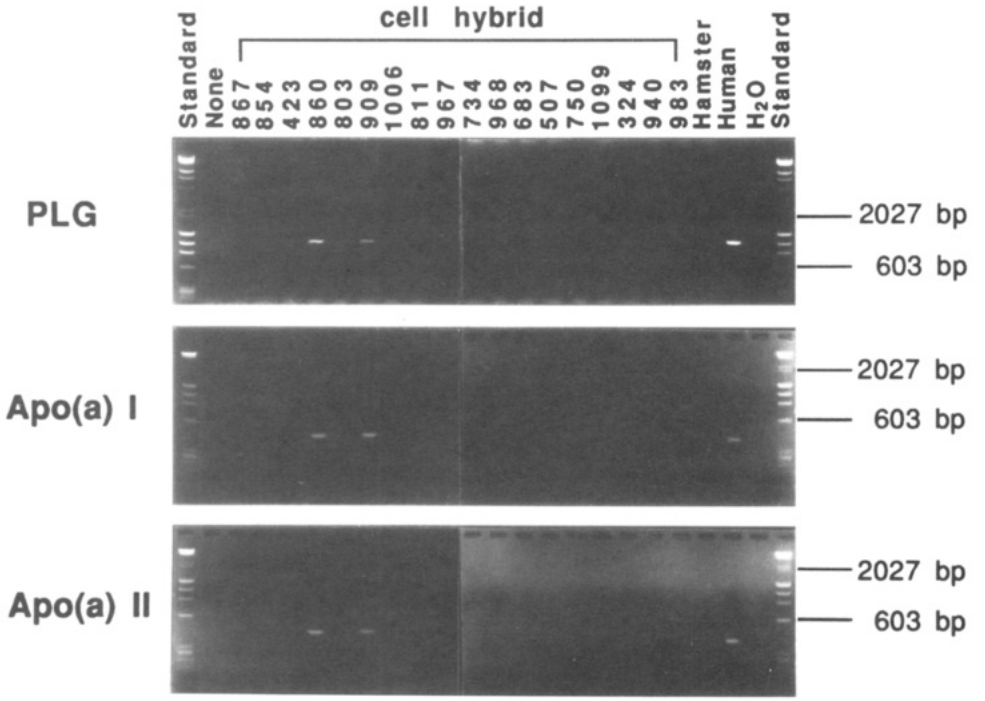

FIGURE 6: Chromosomal localization of the plasminogen and apo(a) genes. The amplified products of the genomic DNA samples from human, hamster, and the human–hamster hybrid cell lines (Panel I) were electrophoresed in a 0.8 (top) or 1.6% (middle and bottom) agarose gel. "None" and "H_2O" are the negative controls without added DNA (buffer or water, respectively).

Plasminogen-Related Genes. Two additional homologous sequences were also found when the amplified DNA and genomic clones corresponding to the 5′- and 3′-regions of the plasminogen gene were analyzed. Two 5′-phage clones encoding two distinct sequences were isolated from a single genomic library, and both of these sequences (corresponding to exons I–V of the plasminogen gene) were also found in each of the amplified DNAs from five individuals. These genes, called plasminogen-related genes (PRGs type A or B), have nucleotide sequences that are about 95% identical to those of the plasminogen gene (Figures 3 and 5). Also, they differ from each other by 1–2% in their nucleotide sequence. A DNA fragment (PRG C, corresponding to exons XVIII and XIX of the plasminogen gene) placed between PRG A and B in Figure 4A could be a part of PRG A or B; however, no genomic DNA (corresponding to exons VII–XVIII of the plasminogen gene) connecting this PRG C to PRG A or B has been obtained by screening of libraries and by in vitro amplification of genomic DNAs. Therefore, it also could be an independent gene but not a section of the apo(a) gene, because PRG C does not have the 27 bp deletion in the Met loop region and the 17 bp insertion in the 3′-noncoding region of the apo(a) cDNA (McLean et al., 1987).

At present, these genes are considered pseudogenes since the region corresponding to the exon coding for the second half of the preactivation peptide has an in-frame stop codon (Figure 5B), and the 3′-exon/intron boundary of the PRGs corresponding to that of the exon coding for the second half of kringle 1 of plasminogen has a significant mutation, GG instead of GT (data not shown). Alternative splicing, however, may bypass these regions and complete an open reading frame for an unknown protein. A 6.6-kb mRNA (McLean et al., 1987) of unidentified origin could be a candidate for the product of these genes.

An Additional Member of the Plasminogen–Apo(a) Gene Family. Recently, two independent groups (Miyazawa et al., 1989; Nakamura et al., 1989) determined the amino acid sequence of human hepatocyte growth factor and found that it contains four kringle structures and a pseudo serine protease domain homologous to plasmin. The amino acid identities were about 50% for the kringle domains, indicating that plasminogen, apo(a), and hepatocyte growth factor share the same basic structures. Also, each kringle in these genes was coded by two exons and the genomic structure of other regions were exactly identical to those of plasminogen (Seki et al., 1991; Miyazawa et al., 1991). Thus, these genes apparently

have evolved from an ancestral plasminogen gene via duplication and exon shuffling occurring at type I splice junctions (Petersen et al., 1990).

Chromosomal Localization of Plasminogen and Apo(a) Genes. Since the genes for apo(a) and plasminogen, and the plasminogen-related genes, differ only 1–5% in their nucleotide sequences of regions corresponding to the exons and their boundaries of the plasminogen gene, cDNA probes cross-hybridize to other genes (Figure 1; McLean et al., 1987; Murray et al., 1987), and oligonucleotides designed from cDNAs anneal to other genes as described above. Therefore, three pairs of gene-specific primers were employed in order to localize each gene on individual chromosomes by in vitro amplification. The single band of 1.1, 0.41, or 0.40 kb was obtained from the human DNA as were those of the 860 and 909 somatic cell hybrids which contain human chromosome 6 (Figure 6). The DNA sequences of 1.1-, 0.41-, and 0.40-kb bands have been determined, and they were confirmed to contain the genomic nucleotide sequences for plasminogen, apo(a)I, and apo(a)II, respectively. On the contrary, no bands were observed with hamster DNA and those of other cell lines. Another set of cell hybrids (Panel II) was also examined, and the same bands were obtained from the 756, 904, 909, and 860 cell lines which contain human chromosome 6 (data not shown), indicating 0% discordance (0 out of 25 different cell lines). Thus, the plasminogen and apo(a) type I and II genes were all localized to chromosome 6. It is probable that they are clustered at loci close to each other.

Discovery of multiple members in the plasminogen–apo(a) gene family has lead to a dramatic improvement in the genetic diagnosis of plasminogen abnormalities by eliminating coamplification of other homologous genes employing gene-specific primers (Ichinose et al., 1991). In addition, chromosomal localizations of the plasminogen and apo(a) genes have been reexamined employing gene-specific primers and somatic cell hybrids as described above. Since both plasma concentration and apparent molecular size of apo(a) are inherited in an autosomal codominant manner (Utermann et al., 1987), there must be differences in the genomic DNA among individuals. Thus, these data will also make it possible to characterize the regulatory nucleotide sequence in the 5′-flanking region and to determine the number of kringle 4 repeats (Gavish et al., 1989; Koschinsky et al., 1990; Lackner et al., 1991; Azrolan et al., 1991) of the apo(a) gene in individuals by in vitro amplification. Since a decrease in functional plasminogen is detected in some of the patients with venous thrombosis (Aoki et al., 1978; Kazama et al., 1981; Mannucci & Tripodi, 1987; Bauer & Rosenberg, 1987) and an increased plasma concentration of apo(a) leads to arterial thrombosis (Dahlen et al., 1975; Armstrong et al., 1986; Seed et al., 1990), it is important to explore the regulatory mechanism of these genes.

Added in Proof

Upon submission of this paper, other papers on the gene for hepatocyte growth factor-like protein have been published (Han et al., 1991; Degen et al., 1991). This protein also consists of a preactivation peptide region, four kringle domains, and a pseudo serine protease domain. Thus, it should be included in the plasminogen–apo(a) gene family.

Acknowledgments

I thank Prof. E. W. Davie for his support throughout this study, E. Espling for his excellent assistance, Dr. T. Petersen for sequencing the middle portion of the normal plasminogen gene, Dr. M. Martzen for isolation of PRG B and C, J. Harris for synthesis of oligonucleotides, Dr. J. Tait for helpful discussion, and L. Boba for her help in the preparation of the manuscript.

Registry No. Plasminogen, 9001-91-6.

References

Aoki, N., Moroi, M., Sakata, Y., Yoshida, N., & Matsuda, M. (1978) *J. Clin. Invest. 61*, 1186–1195.

Armstrong, V. W., Cremer, P., Eberle, E., Manke, A., Schulze, F., Wieland, H., Kreuzer, H., & Seidel, D. (1986) *Atherosclerosis 62*, 249–257.

Azrolan, N., Gavish, D., & Breslow, J. L. (1991) *J. Biol. Chem. 266*, 13866–13872.

Bauer, K. A., & Rosenberg, R. D. (1987) *Blood 70*, 343–350.

Bell, G. I., Karam, J. H., & Rutter, W. J. (1981) *Proc. Natl. Acad. Sci. U.S.A. 78*, 5759–5763.

Berg, K. (1963) *Acta Pathol. Microbiol. Scand. 59*, 369–382.

Biggin, M. D., Gibson, T. J., & Hong, G. F. (1983) *Proc. Natl. Acad. Sci. U.S.A. 80*, 3963–3965.

Dahlen, G., Berg, K., Gillnas, T., & Ericson, C. (1975) *Clin. Genet. 7*, 334–341.

Degen, S. J. F., Stuart, L. A., Han, S., & Jamison, C. S. (1991) *Biochemistry 30*, 9781–9791.

Gavish, D., Azrolan, N., & Breslow, J. L. (1989) *J. Clin. Invest. 84*, 2021–2027.

Han, S., Stuart, L. A., & Degen, S. J. F. (1991) *Biochemistry 30*, 9768–9780.

Ichinose, A., Espling, E. S., Takamatsu, J., Saito, H., Shinmyozu, K., Maruyama, I., Petersen, T. E., & Davie, E. W. (1991) *Proc. Natl. Acad. Sci. U.S.A. 88*, 115–119.

Kazama, M., Tahara, C., Suzuki, Z., Gohchi, K., & Abe, T. (1981) *Thromb. Res. 21*, 517–522.

Koschinsky, M. L., Beisiegel, U., Henne-Bruns, D., Eaton, D. L., & Lawn, R. M. (1990) *Biochemistry 29*, 640–644.

Lackner, C., Boerwinkle, E., Leffert, C. C., Rahmig, T., & Hobbs, H. H. (1991) *J. Clin. Invest. 87*, 2153–2161.

Lawn, R. M., Fritsch, E. F., Parker, R. C., Blake, G., & Maniatis, T. (1978) *Cell. 15*, 1157–1174.

Malinowski, D. P., Sadler, J. E., & Davie, E. W. (1984) *Biochemistry 23*, 4243–4250.

Mannucci, P. M., & Tripodi, A. (1987) *Thromb. Haemostasis 57*, 247–251.

McLean, J. W., Tomlinson, J. E., Kuang, W.-J., Eaton, D. L., Chen, E. Y., Fless, G. M., Scanu, A. M., & Lawn, R. M. (1987) *Nature 330*, 132–137.

Miles, L. A., & Plow, E. F. (1990) *Thromb. Haemostasis 63*, 331–335.

Miyazawa, K., Tsubouchi, H., Naka, D., Takahashi, K., Okigaki, M., Arakaki, N., Nakayama, H., Hirono, S., Sakiyama, O., Takahashi, K., Gohda, E., Daikuhara, Y., & Kitamura, N. (1989) *Biochem. Biophys. Res. Commun. 163*, 967–973.

Miyazawa, K., Kitamura, A., & Kitamura, N. (1991) *Biochemistry 30*, 9170–9176.

Murray, J. C., Buetow, K. H., Donovan, M., Hornung, S., Motulsky, A. G., Disteche, C., Dyer, K., Swisshelm, K., Anderson, J., Giblett, E., Sadler, E., Eddy, R., & Shows, T. B. (1987) *Am. J. Hum. Genet. 40*, 338–350.

Nakamura, T., Nishizawa, T., Hagiya, M., Seki, T., Shimonishi, M., Sugimura, A., Tashiro, K., & Shimizu, S. (1989) *Nature 342*, 440–443.

Petersen, T. E., Martzen, M. R., Ichinose, A., & Davie, E. W. (1990) *J. Biol. Chem. 265*, 6104–6111.

Saiki, R. K., Gelfand, D. H., Stoffel, S., Scharf, S. J., Higuchi, R., Horn, G. T., Mullis, K. B., & Erlich, H. A. (1988) *Science 239*, 487–491.

Sanger, F., Nicklen, S., & Coulson, A. R. (1977) *Proc. Natl. Acad. Sci. U.S.A. 74*, 5463-5467.
Scanu, A. M., & Fless, G. M. (1990) *J. Clin. Invest. 85*, 1709-1715.
Seed, M., Hoppichler, F., Reaveley, D., McCarthy, S., Thompson, G. R., Boerwinkle, E., & Utermann, G. (1990) *New Engl. J. Med. 322*, 1494-1499.
Seki, T., Hagiya, M., Shimohishi, M., Nakamura, T., & Shimizu, S. (1991) *Gene 102*, 213-219.
Silhavy, T. J., Berman, W. L., & Enquist, L. W. (1984) in *Experiments with Gene Fusions*, pp 140-141, Cold Spring Harbor Laboratory, New York.
Utermann, G. (1989) *Science 246*, 904-910.
Utermann, G., Menzel, H. J., Kraft, H. G., Duba, H. C., Kemmler, H. G., & Seitz, C. (1987) *J. Clin. Invest. 80*, 458-465.
Yoshitake, S., Schach, B. G., Foster, D. C., Davie, E. W., & Kurachi, K. (1985) *Biochemistry 24*, 3736-3750.

Deuterium Exchange of Operator 8CH Groups as a Raman Probe of Repressor Recognition: Interactions of Wild-Type and Mutant λ Repressors with Operator O_L1†

Kim E. Reilly, Renee Becka, and George J. Thomas, Jr.*
Division of Cell Biology and Biophysics, School of Basic Life Sciences, University of Missouri—Kansas City, Kansas City, Missouri 64110

Received November 1, 1991; Revised Manuscript Received January 7, 1992

ABSTRACT: The rate of deuterium exchange of a purine 8CH group in DNA is highly sensitive to both macromolecular secondary structure and intermolecular interactions which restrict solvent access to the major groove [Lamba, O. P., Becka, R., & Thomas, G. J., Jr. (1990) *Biopolymers 29*, 1465-1477]. We have exploited the sensitivity of the 8CH → 8CD reaction to probe DNA recognition by the helix-turn-helix (HTH) motif of phage λ cI repressor. We find that purine exchanges in the 19-base-pair O_L1 operator are strongly and specifically restricted by binding of the HTH N-terminal domain of the repressor fragment (RF) comprising residues 1-102. The kinetics indicate large-scale obstruction of solvent access to operator 7N-8C purine sites. Interpretation of the exchange kinetics using a simple model suggests that only 7 purine residues (5 of 10 adenines and 2 of 9 guanines) remain unrestricted with respect to 8CH exchange in complexes of O_L1 with the wild-type repressor. On the other hand, the 8CH exchange profile for the complex of O_L1 with the Tyr88→Cys mutant repressor indicates that 9 purines (7 adenines and 2 guanines) are exchangeable. These results suggest important differences in major groove recognition in the two complexes. The proposed 8CH labeling profiles are consistent with molecular models of related complexes determined by X-ray crystallography [Jordan, S. R., & Pabo, C. O. (1988) *Science 242*, 893-899] and indicate that the structures observed in the crystal are largely maintained in solution. The results also implicate both steric shielding of the 7N-8C locus and specific hyrogen bonding to the 7N acceptor by repressor donor groups as factors which contribute to retardation of operator 8CH exchange. Finally, present results confirm for D_2O solutions of repressors, operators, and complexes the same secondary structures and thermostabilities reported previously for corresponding H_2O solutions [Benevides, J. M., Weiss, M. A., & Thomas, G. J., Jr. (1991) *Biochemistry 30*, 4381-4388, 5955-5963].

The control of gene expression in prokaryotic and eukaryotic cells requires recognition of discrete DNA-binding domains by regulatory proteins. An intensively studied example is the cI repressor-operator system of bacteriophage λ (Ptashne, 1986). The N-terminal DNA-binding domain of λ cI repressor has been cocrystallized with its cognate operator, and the structure of the specific complex has been solved at 2.5-Å resolution by X-ray methods (Jordan & Pabo, 1988). The crystal structure reveals stereochemical complementarity between the helix-turn-helix (HTH) motif of λ repressor and the major groove of the cognate operator, mediated by many specific protein-DNA interactions. Related crystallographic studies of λ Cro (Brennan et al., 1990) and bacteriophage 434 repressor-operator complexes (Aggarwal et al., 1988; Wolberger et al., 1988; Mondragon et al., 1989a,b) have confirmed the generality of the HTH fold as a DNA-binding domain and have provided a detailed view of the HTH motif as a scaffold for DNA recognition. Such studies also reveal unifying features of the mechanism of HTH recognition and demonstrate the importance of detailed stereochemical interactions in lieu of a simple "recognition code" (Pabo et al., 1990; Harrison & Aggarwal, 1990; Steitz, 1990). A striking feature of the crystallographic results on phage 434 repressor-operator complexes is the different conformations of the operator in each complex, demonstrating that protein binding influences the orientation of functional groups in the major groove (Wolberger et al., 1988). This conclusion, deduced from the crystal structures, may apply also to repressor-operator complexes in solution, although direct evidence from comparison of solution structures has not yet been obtained. Since small differences in affinity between operator sites can be sufficient to define a genetic switch (Ptashne, 1986), these effects may

† Paper 43 in the series Raman Spectral Studies of Nucleic Acids. Supported by NIH Grant AI18758.
* To whom correspondence may be addressed.

0006-2960/92/0431-3118$03.00/0