Complete Nucleotide Sequence of the Antithrombin Gene: Evidence for Homologous Recombination Causing Thrombophilia†,ƒ

Robin J. Olds,*,‡ David A. Lane,§ Vijoy Chowdhury,‡ Valerio De Stefano,⊥ Guiseppe Leone,⊥ and Swee Lay Thein‡

Institute of Molecular Medicine, John Radcliffe Hospital, Oxford OX3 9DU, U.K., Department of Haematology, Charing Cross and Westminster Medical School, London W6 8RP, U.K., and Istituto di Semeiotica Medica, Universita Cattolica Del Sacro Cuore, Rome, Italy

Received October 20, 1992; Revised Manuscript Received January 20, 1993

ABSTRACT: Antithrombin is the principle regulator of thrombin and other blood coagulation proteinases. It is a member of the serpin family of proteinase inhibitors. The genomic sequence of the antithrombin locus has been completed, revealing a gene spanning 13 477 base pairs from the transcription start site to the poly(A) addition signal. Nine complete and one partial Alu repeat elements were identified within the introns of the gene, with all but one orientated in the reverse direction. Inherited deficiency of antithrombin is associated with a venous thrombotic tendency. Restriction fragment mapping of the antithrombin genes in an individual with type I antithrombin deficiency identified an intragenic deletion in one allele. Localization of the deletion breakpoints involved restriction analysis and direct sequencing of amplified DNA spanning the deletion site. The deletion removed 2761 base pairs, affecting exon 5 and flanking introns, with the deletion ends contained within the left components of two Alu elements. It is likely, therefore, that the deletion arose by homologous recombination between the two Alu elements.

[6]Antithrombin is a glycoprotein which acts as the major plasma inhibitor of thrombin (Abildgaard, 1969; Rosenberg & Damus, 1973; Lane et al., 1992b). In vitro inhibitory activity against several other of the activated serine proteinases of the coagulation system has also been documented. Acceleration of the proteinase inhibitory activity of antithrombin is promoted by heparin, a sulfated polysaccharide not found within the vasculature under normal circumstances. There is evidence that the proteoglycan heparan sulfate located on the endothelial cell surface may provide a physiological means by which these inhibitory reactions are accelerated in vivo (Marcum & Rosenberg, 1984; Marcum et al., 1986).

Antithrombin cDNA has been cloned and sequenced (Bock et al., 1982; Chandra et al., 1983; Prochownik, 1983), and the antithrombin structural gene has been assigned to human chromosome 1q23–25 (Bock et al., 1985). The gene consists of seven exons, 1, 2, 3A, 3B, 4, 5, and 6. Initially only six exons were identified, but subsequent analysis revealed a 1-kb[1] intron within exon 3 (Bock et al., 1988), which is now commonly referred to as exons 3A and 3B. In this report, we have chosen to continue with this terminology. The nucleotide sequences immediately flanking each of the seven antithrombin exons have been determined (Prochownik et al., 1985; Bock et al., 1988). Plasma antithrombin is a 432-residue polypeptide containing six cysteines, forming three intramolecular disulfide bonds, and four glycosylation sites. The protein is synthesized by the hepatocytes, with a 32-amino acid leader peptide which is cleaved before secretion of the mature inhibitor into the plasma. Sequence and structural homology suggests that antithrombin belongs to a large group of related molecules, the serine proteinase inhibitor, or serpin, family (Huber & Carrell, 1989). The family includes a number of physiologically important inhibitors, such as CI-inhibitor, heparin cofactor II, α_1-proteinase inhibitor, the plasminogen activator inhibitors, and α_2-antiplasmin.

Inherited antithrombin deficiency is associated with a venous thromboembolic tendency (Egberg, 1965). In a proportion of affected subjects immunological and functional plasma antithrombin levels may be reduced in concert, usually to about 50% of normal, suggesting that one mutant antithrombin allele fails to direct the synthesis of detectable protein; this has been termed type I deficiency. In other affected individuals the immunological level is (near) normal but results of functional assay are reduced, indicating the presence of a variant dysfunctional protein, and this is termed type II deficiency (Lane et al., 1991). Appreciable progress has been made identifying the amino acid substitutions responsible for type II antithrombin deficiency; this has been achieved initially by protein sequencing and more recently by analysis of the coding regions of the antithrombin gene. Following the demonstration of a point mutation in the gene coding sequence causing type I deficiency (Olds et al., 1990), several other point mutations have been described. However, progress in this area has been hindered by the incomplete knowledge of the antithrombin gene structure. Here we report the complete nucleotide sequence of the locus, and furthermore, we show that recombination between Alu repeat elements within the introns may cause type I antithrombin deficiency.

MATERIALS AND METHODS

Antithrombin-Deficient Kindred. The propositus (G052) suffered his first documented episode of thrombosis, affecting the dorsal vein of the penis, at the age of 23 years. A further episode of superficial venous thrombosis occurred in the right arm following injection of medication at the age of 46. The father of the propositus suffered deep leg vein thrombosis and pulmonary embolism after prostatectomy at age 68. Two sons, aged 17 and 20 years, have reduced levels of plasma antithrombin but as yet have had no clinically symptomatic thrombotic events. A reduction in plasma antithrombin of

† This work was supported by the Wellcome Trust.

ƒ The antithrombingene sequence in this paper has been submitted to the EMBL data bank under accession number X68793.

* To whom correspondence should be addressed.

‡ John Radcliffe Hospital.

§ Charing Cross and Westminster Medical School.

⊥ Universita Cattolica Del Sarco Cuore.

[1] Abbreviations: PCR, polymerase chain reaction; bp, base pairs; kb kilobase; nt, nucleotide.

0006-2960/93/0432-4216$04.00/0

the propositus was confirmed in the current investigation, and the plasma was examined by crossed immunoelectrophoresis to search for any variant form of the protein (Olds et al., 1990).

Southern Blot Analysis. Genomic DNA extracted from blood leukocytes was digested with restriction endonucleases, using conditions recommended by the manufacturer. After separation in 0.8% agarose gels, the fractionated DNA was blotted onto nylon membranes (Hybond N, Amersham, U.K.) and hybridized to antithrombin gene probes. The probes, which included two antithrombin gene fragments generated by the polymerase chain reaction (PCR) and spanning exons 1–2 and exons 3A–3B, and antithrombin cDNA were labeled with [^{32}P]dCTP by random priming (Megaprime, Amersham, U.K.). Following hybridization, the membranes were washed at high stringency and autoradiographed between intensifying screens at −70 °C.

DNA Amplification. To obtain the intron sequences of the antithrombin gene, regions spanning exons 1–2, 3A–3B, 3B–4, 4–5, and 5–6 were amplified from genomic DNA of normal individuals by PCR. Oligonucleotide primers were located within previously reported sequence flanking each of the exons (Bock et al., 1988). In each amplification reaction either the 5′ or the 3′ oligonucleotide primer was 5′-biotinylated. A similar strategy was adopted in subject G052; a series of PCR amplifications were performed using oligonucleotide primers located in the regions flanking the postulated deletion site. Restriction endonuclease digestion of PCR products was also performed with a variety of enzymes. The pattern of bands was resolved in 1% agarose and visualized by UV transillumination following ethidium bromide staining (0.5 μg/mL).

DNA Sequence Analysis. DNA amplified by PCR and a partial-length antithrombin gene clone were analyzed to derive the normal genomic sequence. The clone pAT5.0, a 5-kb *Pst*I fragment of the antithrombin gene which includes intron 2 (Prochownik et al., 1985), was obtained from Dr. E. Prochownik. In order to sequence this intron, the insert was isolated from the plasmid pBR322 by digestion with *Pst*I and subcloned into M13mp18 and 19. To obtain the remainder of the intron sequences, PCR products containing each of the introns were directly sequenced. Single-stranded (ss) template DNA was prepared by incubation of the PCR product in 0.15 M NaOH followed by isolation of the biotinylated primed ssDNA using the streptavidin–Dynabead system (Dynal U.K. Ltd.) (Thein & Hinton, 1991). Sequence analysis was performed by the dideoxynucleotide method (Sequenase, United States Biochemical) using [α-^{35}S]dATP.

RESULTS

Antithrombin Gene Sequence. The normal sequence of the antithrombin gene was derived by analysis of the clone pAT5.0, for intron 2, and by direct sequencing of amplified DNA for each of the other five introns. For each of the sequence regions previously unreported, we have confirmed our sequence by at least two independently repeated amplifications and sequence analyses. In the course of the sequence analysis we also confirmed the previously published exon sequences. The gene spans 13 477 base pairs, from the transcription start site (Prochownik & Orkin, 1984) to the last nucleotide of the AATAAA poly(A) addition signal (Bock et al., 1982; Chandra et al., 1983; Prochownik, 1983). Intron 5 is the longest at 3374 bp; in decreasing order of size the others are 2532 bp (intron 2), 2298 bp (intron 1), 2033 bp (intron 4), 905 bp (intron 3A), and 810 bp (intron 3B). The sequence with the positions of the exons is shown in Figure 1; in this figure the first 20 and the last 10 nucleotides have been added from the reports of Bock and co-workers (Bock & Levitan, 1983; Bock et al., 1988). Several polymorphisms within the antithrombin locus which had been described previously were localized: (1) the 5′-end length polymorphism, which consists of either a 108- or a 32-bp fragment (Bock & Levitan, 1983), extends 5′ to position −276 relative to the mRNA start site; (2) the *Pst*I cutting site polymorphism within codon 305 (Prochownik, 1983) is produced by either a G or an A at position 7626; (3) the third base of codon 295, which may be either G or A, which does not alter the normal Val (Olds et al., 1991b) is at position 7596; (4) an NheI polymorphism within intron 4 (Bock & Radziejewska, 1991) results from either T or C at position 7987; (5) a *Dde*I site polymorphism within intron 5 (Daly & Perry, 1990) is created by either C or G at position 9893. A number of sequence differences from previously published partial sequences were identified in our study and are listed in Table I. The sequence of Bock et al. (1988), which included short lengths of introns flanking exons, was confirmed in our study with the exception of four positions (Table I). At each of these we have confirmed our sequence by the repeated analysis of amplified DNA from several individuals, indicating that the differences are unlikely to be sequence polymorphisms.

Analysis of the genomic sequence revealed the presence of 10 Alu repeat elements within the antithrombin gene (Figures 1 and 2 and Table II). Of interest is the clustering of the Alu repeats between exons 4 and 5 where, within the 2-kb intron, four repeats contribute 53% of the sequence. Each of the Alu elements, with the exception of Alu 6 located within intron 4, are found in the reverse direction. Alu 2, within intron 2, is a partial copy, with the majority of the Alu-left component missing. Overall, the Alu repeats within the gene demonstrate a high degree of sequence homology (range 82–93%) to an Alu consensus sequence (Britten et al., 1988).

Plasma Antithrombin. The level of plasma antithrombin antigen in the propositus (G052) was reduced at 56% (normal 80–120%). Crossed immunoelectrophoresis of the plasma in the absence or the presence of heparin did not reveal any variant form of the inhibitor. This confirmed the classification of the deficiency as type I.

Identification of Antithrombin Locus Structural Rearrangement by Southern Blot Hybridization. Genomic DNA from G052 and a normal individual was digested with a variety of restriction enzymes and hybridized initially with labeled antithrombin cDNA. In addition to the expected normal fragments (Figure 2), restriction fragments of abnormal sizes were also present in G052 (Figure 3 and Table III). The pattern of bands indicated that the proband is a heterozygote with one normal antithrombin allele, the abnormal restriction fragments being derived from a mutant antithrombin allele. The *Pst*I site within exon 4 (Figure 2) is polymorphic in the normal population, and the absence of the site gives rise to a 10.4-kb band, while the presence of the site creates two fragments of 4.8 and 5.6 kb. Only the 10.4-kb fragment and the constant 2.0 and 2.5-kb bands are seen in the *Pst*I-digested DNA of the normal individual in Figure 3, who is homozygous for the absence of the *Pst*I site within exon 4. In addition to the 10.4-, 2.0-, and 2.5-kb bands in G052, a 4.8-kb band is observed, indicating that the individual is heterozygous for the exon 4 *Pst*I cutting site. However, instead of the expected 5.6-kb band, representing the other fragment derived from the allele with the *Pst*I site within exon 4, a 2.8-kb band was found (Figure 3 and Table III). Abnormal bands consistently

```
-534                                                                                                  CCAC AGGTGTAACA TTGTGTTTTC CTTGTCTGTG -501
-500 CCAGGGACAC CTTGGCATCA GATGCCTGAA GGTAGCAGCT TGTCCCTCTT TGCCTTCTCT AATTAGATAT TTCTCTCTCT CTCTCCCTCT CTCCATAAAG -401
-400 AAAACTATGA GAGAGGGAAT TACAGGTAGA GGGCTAGAAG TTTTTGGACA TTAACTATTT CTATCTTCTG ATTTAGTTAA CGAGAAACAA AAAATCCTGC -301
-300 AGACAAGTTT CTCCTCAGTC AGGTATTTCC TAACCAAGTT TGAGGGTATG AACATACTCT CCTTTTCCTT TTCTATAAAG CTGAGGAGAA GAGTGAGGGA -201
-200 GTGTGGGCAA GAGAGGTGGC TCAGGCTTTC CCTGGGCCTG ATTGAACTTT AAAACTTCTC TACTAATTAA ACAACACTGG GCTCTACACT TTGCTTAACC -101
-100 CTGGGAACTG GTCATCAGCC TTTGACCTCA GTTCCCCCTC CTGACCAGCT CTCTGCCCCA CCCTGTCCTC TGGAACCTCTG CGAGATTTA GAGGAAAGAA -1
   1 CCAGTTTTCA GGCGGATTGC CTCAGATCAC ACTATCTCCA CTTGCCCAGC CCTGTGGAAG ATTAGCGGCC ATGTATTCCA ATGTGATAGG AACTGTAACC 100
                                                                                  Exon 1   M  Y  S  N  V  I  G  T  V  T
 101 TCTGGAAAAA GGTAAGAGGG GTGAGCTTTC CCCTTGCCTG CCCCTACTGG GTTTTGTGAC CTCCAAAGGA CTCACAGGAA TGACCTCCAA CACCTTTGAG 200
      S  G  K  R
 201 AAGACCAGGC CCTCTCCCTG GTAGTTACAG TCAAAGACCT GTTTGGAAGA CGTCATTTCA AGTGCTCTCC CTCCCACCCC ACCTCTTGGG GTAAGGCCTT 300
 301 TCCTAAGCTA CCCCTTGGGT CCCTAGCCTA AGAAACAAGG GGGATGTCAT CCCTGGTGTA AAGATGCTGT GCAGGAAGTC AGCACTCACG GGATCCAGGG 400
 401 GACGCTCCAA GGGGAATCCC CAGGGCCTGC CATCCATCCG GGAAGAGAGC AAATGCTACC CATGAGGACC TCCTCACTCC CTTTTTGCTC TTTCTTCCAC 500
 501 TCAGATCCAC CCCACTCCAC CCCCACCCAA ATCCCAGTGA CCTTTGACTA AAGGGCCAAA ACTGCTTCCT TTTCTCACAA TGAGAGTTGT CCCTCCCTCA 600
 601 ATGCCACACA CACTCCCTTC TTCATCTGAG TTGTCACAGG AGGCTAGAAA CGGGGTGGTG GCACAACTGT CTTGGTTTTA ATTTGTGCTT CATAGCCCTC 700
 701 CCAGGTCCTC TCAGCCTCAA ATTGCATTTC CAAATGTAGT TGAAGGGACA GAGTGGGCAA CCGAAGCAGC AGTGGAGATG GGAAGATGAA TGGCAGGGTC 800
 801 CTCTCCTCTC TCTCTCTGCT TCTTCAGCCT GCCTTCCACA TCTCCCTTGG TGCCGCTGCT TCTCTCCGGC TTTGCACCTC TGTTCTTGAA AGGGCTGCAG 900
 901 AACTGGACTC AGACCACGCA AGAAGGCAAG TCCCCCTCAG CTGCCCCAGC TTCCAGCCAG CCCCAGGCTT GCCCAACGCA CCACGTCCGT GAATCTGCAC 1000
1001 TGGGTGCCTG TCTTTCTCTC CCAGGAGAAG ATGGGAAGAT CCAGTACCCA CACACAGACC CCCTTGTGTA CACGCAGGAA CCATAAACCA GCTGGAGGCA 1100
1101 GCCCCTGCCC CACCCTGTCT TATCTACAAA AAATATTACA AGAGACTTTA TCTCTTGATT TGCTTCATCG AGTGTCCCAA CTACCTCATT TTTTTAAAAT 1200
1201 GTGAAATTAG CTTCATTTAC CTTCATTGAA TCCATGTTGG CGACTATTAA AAATTCCAGG CAATAAAAAG GGATGAGAGC CTGAACTAAA GCAGTGGCAA 1300
1301 TAACTGGTGA AAGAGTAAAA AAACAGAACT GATTGACTCT GGGGTGAACT GATTGACTCT GGGGTTTGAC TAAATGAGGA GGAGAGAGGG AGAATCCAG 1400
1401 GGTGATTCTC AGGTTTCTGT ACGGGATTCA CTGAGCCCAC TCACAGGAGC AGGCCTGTGG GGGAGAATTA ATTACCAGTT CAGTTTGGTC CTGTTTCCCT 1500
1501 GAAGAACTTG TAGGAGTTCC TGGTGGAACT GTCCAGCAAA TAGTCAGTCT GGAGCTCAGT GGAAGGGTTA GGGCTGGAGC TAGAGATGTA GGAATCTTCA 1600
1601 GCACACAGAT ATTGCCATTG TTTTTGTTTG TTTGTTTGTT TGTTGTTGCT GTTTTGAAga cacagtctca ctttgtcacc caggttggag gtcagtggca 1700
1701 caatctcagc tcactgcaac cttcgcctcc tgggttcaag tgattcttct gcctcagcct ccctagtagc ttgggactac agtgtgcgcc accacaccca 1800
1801 gctaattttt gtatttttag tagagacagg gtttcaccat gttgtccagg ctgatctcga acacccaacc tcaagtgatc tgcatgcctc agcctccaaa 1900
1901 gtgctgggat tacagcgtga gccacgcacc cggccAGATA TTGCCTTTGC TCCATCCATT TCTTCTTTTT CTCTTGTGTT GCTGAAATCT CTCTGCCTGC 2000
2001 ATCTATCAGA GTCCTTCCCC AAACAGTTTC TGTAGATGGC TCCCCCTACC ACCCTGACTC TTCACTGGGC ACTAAAGCCG ATTTTTTAGG CATGCACATT 2100
2101 CCATGTCACA AACAGGAAGC TTCTCATTCT TTTTTCTCCC AGCGTGGGGA ATTGAGCACA TAATACTCCA AATAACCATC AGATGATTCT AATTCCAACA 2200
2201 TGACCACGTC CAGGCAACTG AACTGTCCCC TGGCAAGAAG TCTAGGACTG AACCTGTCCC GGGCCCCTGT ACTTGGTTCA AAGGATTTAG CCTTTCTCTT 2300
2301 GGCCACACCA GGTGGGCTGG AATCCTCTGC TTTACTGGGG CAACCCTGTG GTGGGCAGTG GGGCTAGGGG TTGCAGCCTA GCTTAACTTG GCATTTTGTC 2400
2401 TCCTTGCAGG AAGGTTTATC TTTTGTCCTT GCTGCTCATT GGCTTCTGGG ACTGCGTGAC CTGTCACGGG AGCCCTGTGG ACATCTGCAC AGCCAAGCCG 2500
     Exon 2   K  V  Y  L  L  S  L  L  L  I  G  F  W  D  C  V  T  C  H  G  S  P  V  D  I  C  T  A  K  P
2501 CGGGACATTC CCATGAATCC CATGTGCATT TACCGCTCCC CGGAGAAGAA GGCAACTGAG GATGAGGGCT CAGAACAGAA GATCCCGGAG GCCACCAACC 2600
     R  D  I  P  M  N  P  M  C  I  Y  R  S  P  E  K  K  A  T  E  D  E  G  S  E  Q  K  I  P  E  A  T  N  R
2601 GGCGTGTCTG GGAACTGTCC AAGGCCAATT CCCGCTTTGC TACCACTTTC TATCAGCACC TGGCAGATTC CAAGAATGAC AATGATAACA TTTTCCTGTC 2700
     R  V  W  E  L  S  K  A  N  S  R  F  A  T  T  F  Y  Q  H  L  A  D  S  K  N  D  N  D  N  I  F  L  S
2701 ACCCCTGAGT ATCTCCACGG CTTTTGCTAT GACCAAGCTG GGTGCCTGTA ATGACACCCT CCAGCAACTG ATGGAGGTAC GACCAAAGGT CTTCTGCCCA 2800
     P  L  S  I  S  T  A  F  A  M  T  K  L  G  A  C  N  D  T  L  Q  Q  L  M  E
2801 GCCACCTTGT TAGGAGCACC TTTGGGCTTC CATAGGCCCA AGTCCAATGA TTCCTCAACC AACACTGCAG CCACTAGGGG CGCTCATTAT GCATTACGAT 2900
2901 TCCCTTTGAA CATCACTGTG TTATAATTCC CTTTGAAAAT CATTTTTTAA AAAATTAGCC AAGGAATCTT GGCTATCTAC TTTTTAAATC CTGGTTTCCT 3000
3001 CTTTTGAGCA CCTTAAAATG GGGGAAGGCT TGTATCTTCT CTCAACTTCT TTTCAGTAAT TCTTTCATCT ATATGTTTAC TCATTAATTT GATCATTTAT 3100
3101 TTATTTATTC ATTCAGCACT TCCTCTGTGC CAGGCAATGT GTAGTGCCAG TCCCTCCTCT GGTGGAAGAA GAGTAGCTTT ACCATATGGT GACATCAGGC 3200
3201 ATATAGGCTC TCGTGGAAAA AAATTCTAGG ATAGTATTTT TTTTTTTTTg agatggaatc tcgctctatt gcccaggctg gagtgcagtg gtgcagtctc 3300
3301 ggctcactcc aaactctgcc tcccaggttc aagcaattct cccacctcag cctcctgagt agatgggatt acaggcacac gccatcacgc ccagctaatt 3400
3401 tctatatttt tagtagagat ggggtttcac cACGTGGCCA GACTGGTCTC AAACTTTTTT TTTTTTTTTT TTgagacgga gtctcgctct gtcacccagg 3500
3501 ctggagcgca gtgcacgatc tcagctcact gcaacctcta cctcccgggt tcaagcaatt ctcagcctca gcctcccgag tagctgggat tacaggcccc 3600
3601 cggcaccatg cctggctaat tttttttctt cttagtagag atggggtttc accatgttgg ccaggctagt cttgaactcc tgacctcgtg atccacctgc 3700
3701 cttggcctcc caaagtgctg tgattacagg cgtaacgacc gcgcctggcc TCAAACTCTT GACTTCAAGG GATCGGCCTG ACTTGGCTTC CCAAAGTGCT 3800
3801 GGGATTACAA GCATGAGCCA CTGCACGGGG CCTAGGATGG TATATTGAGA CCAGGGGCCC AGGAAAGCCA AGAGAAGCCT CAAGGACGTG AGAGTGTTTC 3900
3901 TGGCTCTGGG AAGTATGGAT CATTTCAGCT CAGTGACTTA GTTCCCACCC CCTTCCCCCC ACTGCCTTTT GTGGGAGGGA AGTAGGGCAT GATAAGATGA 4000
4001 AATGTCATAG ATTGATTGAT CACTGTTGGC CTCTGGGGCT ATGACAAGTC ATGGATGGAA ACACTAGATC TTTAATCTGT CCTTGGCTTG GCTGCATGAC 4100
4101 AGTCTTTCTT CAAGTTGGAT CACACTTTGG AAGCAGAGTT CATCAATAGG GAGGCATGAG TCCCTTCAAG ATGGTATACG GTGCTTATTT GAAACTTGGA 4200
4201 CACTAAAGTC TGTGGGTCTT AGGAGGGTTC CTTCTATTCT AGTGGTCAAT TTCCATGGAA CTTCATCACC TTTGCTCAGG GCTCTGGGGT GAGTTAACCC 4300
4301 AAGTCTTCAC TCTTTGAAAG AAATTGTAGA TTTAAAAACT CTGAAGACAC ATAATACTGC CTTCTCTGGG CCCTTCAGTC ATTTTTGTAT ACATTGGTAC 4400
4401 TGGTCTCAAA GTACTTCCAT ATCACTCATG TCTCTGTCCC CAGGTAAGAT CTTAAATGTA ACCCTTCCTA CAAGAAGAAC AGAACAGAAC ACTGCTTCCA 4500
4501 AACCACACAT GTTCCTTTGG TCCTCCCCTC TACACAAACG CCATGTGTTG GGAAAGCAGG GTGAGACTAA ATCTCTCTGG AGAAAAGAGA AATTCAGCAC 4600
4601 CAAGCTTTTG ATCAAAAGCA TAATCCCCCC CTAAAAAAAG TGCCTATTGG AGCAAAATCA GGAAAACCAA AGGCAGAGAA CAGATAAAAC CAAAAGGCCT 4700
4701 TTTGTAGCCT GAGGAGAGAG CATGGAAAGG GCAGGAGGGG AACAGCCTCA CCCATTTTGC CTTGGGGATG GTGAAGGTGG GCATTGGGGG ATTCCAACTT 4800
4801 CAAAGCATGG ATGACTTCTA AGTCCTTTTC AGCCCTGAGC TCTTAGATTC TGAGCCTGTT TAATCCCTTG CTGATAGATT CACTCTTCCT TTTTCACCCC 4900
4901 TACCACCAGT ATCCCAGAGC CTCCATGAGC AGCTGGCCCC AGTAGATGCC ACAAAAGTGT TTGTTACGAG AAGGACACCG TCTGATTCTC TTCTCTGTCC 5000
5001 AGAATCACCA AGAGGACTTT TCCCATTCCA GCAAGAAAAC GTCTGTGTGT TGATCTAGAG GCGTTTAGAG ACTTTAGGTG GCAACCTAGT CTCTCTTTTT 5100
5101 CCCTTTATCC TTCCTACCCT TCATTCTTCT TTTATCCTTT TATTCATCAG AACACAAGAG TTGAGCATTT ATGCTGTCCC AGGTACTGTG CTTGAAGGAG 5200
5201 TTAACAACTG AGGTGGCTAT TAGTCAGAGA CTGACCAGCA TGTGCTCACC ACCCATGTTA ACTAGGCAGC CCACCAAACC CACCACCATT TTTTTTGAC 5300
5301 TTCTATAGGT ATTTAAGTTT GACACCATAT CTGAGAAAAC ATCTGATCAG ATCCACTTCT TCTTTGCCAA ACTGAACTGC CGACTCTATC GAAAAGCCAA 5400
     Exon 3A  V  F  K  F  D  T  I  S  E  K  T  S  D  Q  I  H  F  F  F  A  K  L  N  C  R  L  Y  R  K  A  N
5401 CAAATCCTCC AAGTTAGTAT CAGCCAATCG CCTTTTTGGA GACAAATCCC TTACCTTCAA TGAGACCTAC CAGGACATCA GTGAGTTGGT ATATGGAGCC 5500
     K  S  S  K  L  V  S  A  N  R  L  F  G  D  K  S  L  T  F  N  E  T  Y  Q  D  I  S  E  L  V  Y  G  A
5501 AAGCTCCAGC CCCTGGACTT CAAGGTGAGT TGCAGATGTT ACCCCTGACC TCCGAGTTCT TCCTCTCCAC TCAGAGATTG AGGAGGTGGA GAAACAGCAT 5600
     K  L  Q  P  L  D  F  K
5601 CCAAATTCAC ACTGCTTTGC TGCTGAAGAC TGCTGGAGGG CTGACTAAAA GTTAGAACCC CTGCAATAGT TATTCTTACT TGAAACCTGA GAAATCAAAG 5700
5701 GTATCCATGC TTGGATTGTA GTGACTGCCC AGAAAACATG AATTAATAAT CAATTCTTCA TTCCATCCAC CAACTTCAAA TATATACCAA AGGGTGTTTT 5800
5801 GAAGATGCCA GTTCTACAAG ATATCTTACT TAATTTGAAC TGTTATCATG GTCAAATAAA GTTGGTACAT GATGCATGTT ACATTCTCCT CTTGGAGATT 5900
5901 CATGAAGCAC ATGGGCCTAT GAAGGTTCTG AGAAACTCTG CAACAAAGAA ATCTGTTGGC TTTATTCAAT CGGCATTCCT CAAATGTATT TGACTGCATG 6000
6001 GGCATTTCTC TCCTCCATAT AACCTGCCAA CCCCATATAA CCTGCCAACC CCATATAACC TGCAATCATT CATTGCTTCC CCTGGCACAT GCCTTGGAAA 6100
6101 TTCTACTTTT GTGAGTTAAG GTTTTCCAAA GTCAGAGAAA ATAATATTTT ATCTTCTTTT TCCCAGACTA TTTTCGCCTT CCTTCTTTTC ATTTATTTCT 6200
6201 TCCTATTTCT TTTTGTCTTT TTCTTCTGAT AATATTTATT AACTACAGGA AAGATTCATG GAACTATATT AGATATGTGA GGCTTCCCAA TTTGGGTTAG 6300
6301 AGCAATGGCT TCTTAATCAA ATGGTGGGAA AGGACAGAGG GATGGTGAGA AAAATAAAAT GCTGCCTGGG AAAATGGAGA AGCCAATTGA ATAGCACAGG 6400
6401 TGAGTAGGTT TATTTTCTGT TCTCCTCAGG GAATCCAGA GCAATCCAGA GCGGCCATCA ACAAATGGGT GTCCAATAAG ACCGAAGGCC GAATCACCGA 6500
                    Exon 3B  E  N  A  E  Q  S  R  A  A  I  N  K  W  V  S  N  K  T  E  G  R  I  T  D
6501 TGTCATTCCC TCGGAAGCCA TCAATGAGCT CACTGTTCTG GTGCTGGTTA ACACCATTTA CTTCAAGGTA CTCAGAATGG CCCTGGAGAG ACCCCAGGGA 6600
      V  I  P  S  E  A  I  N  E  L  T  V  L  V  L  V  N  T  I  Y  F  K
6601 CTTCCTCTTG CTCTTCAGCT TACCCCCTTT TTTTTTAAAT GGCGAGACCG AAGCCCTGAG AGGGCAAATG GACTGCCGAA AGCTACACAG GTACAGGTCA 6700
6701 GCAGGGCAGG TCAATCTATT ATTTATTTAT TTATTTATTT TTgacagagt ctcgctctgt cgcccaggct ggagtgcagt ggcgtgatct cggctcactg 6800
6801 caagcttcgc ctcctgggtt ctcggcattc tcctgcctca gcctcccaag tagctgggaa tacaggcacc caccaccatg cctggctaat tttttggttt 6900
6901 ttttttagta aagacggggt ttcaccgtgt tagccaggat agtcttgatc tcctgacctc gtgatctgcc cacctcggcc tcccaaagtg ctgggattac 7000
7001 aggcatgagc caccgcgccc ggcaGATTGG CTTCTTTCAC CTAGTAAAAT GCATTTACTG TTCCTTTGTG TTTTCGTGGC TTTGTCACTC ATTTCTTCTT 7100
```

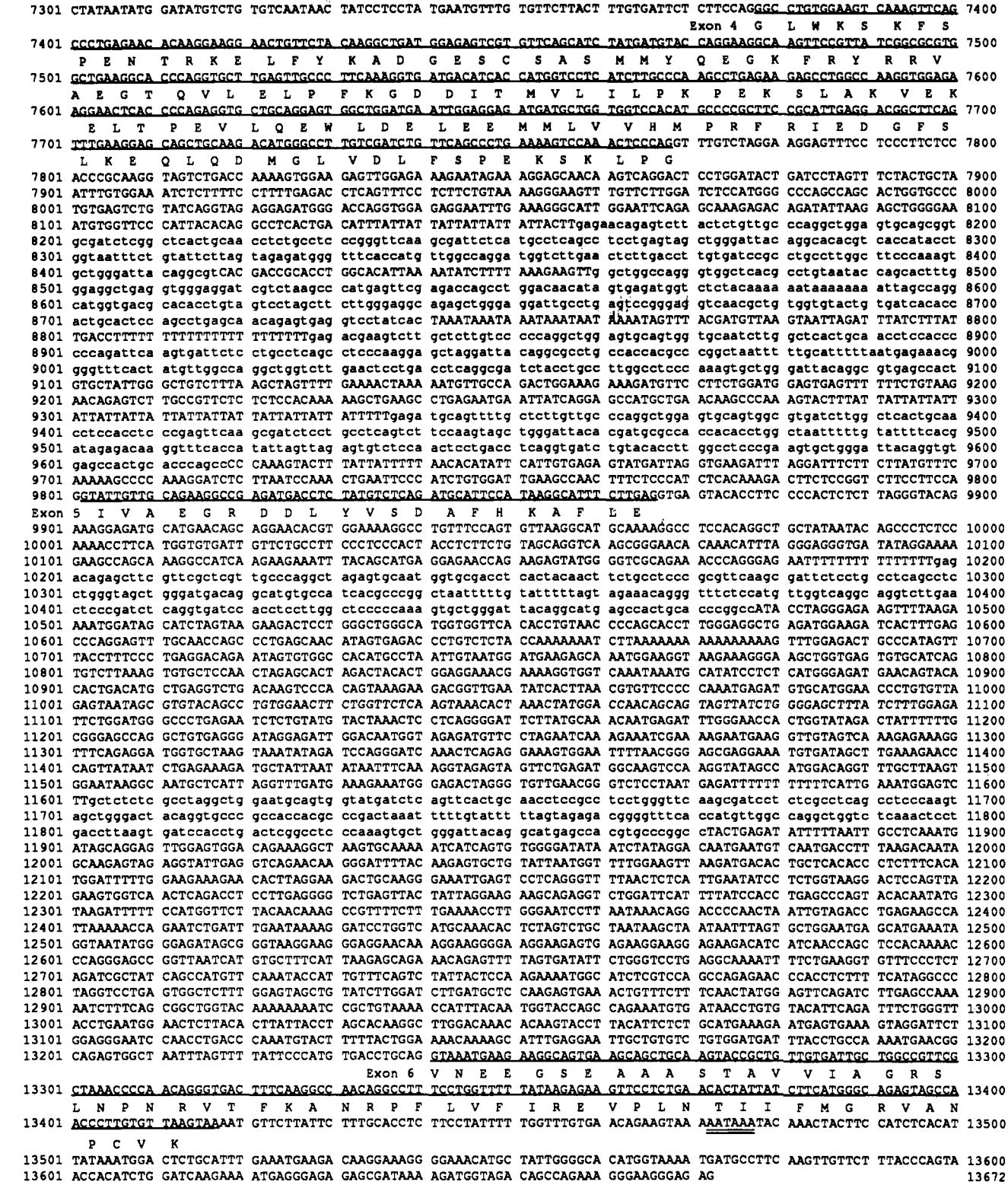

FIGURE 1: The antithrombin locus sequence. Coding regions are underlined, and the amino acids are shown underneath. Alu repeat elements are indicated in lowercase type. Nucleotide +1 represents the transcription start site as mapped by Prochownik and Orkin (1984). The poly(A) signal is indicated by double underlining. The sequence from −383 to −276 forms the long allele of the 5′-end length polymorphism (Bock & Levitan, 1983).

~2.8 kb shorter than the normal fragments were identified (Figures 2 and 3) by using restriction enzymes with sites flanking intron 4 and exon 5 (*Hin*d III, *Bgl*II, and *Bam*HI), suggesting that the abnormal allele of G052 contained an intragenic deletion affecting part of intron 4, exon 5, and part of intron 5. Furthermore, hybridization of DNA, digested with a variety of enzymes (Table III), with DNA probes specific for exons 3A–3B and exons 1–2 suggested that the region 5′ to the *Eco*RI site within intron 4 was intact. No abnormal fragments were observed in hybridizations of *Bam*HI, *Pst*I, and *Pvu*II digests with the *Pst*I-2.5-kb antithrombin probe which included exon 6, indicating that the 3′ region of the mutant allele up to the *Bam*HI site within intron 5 was intact. An abnormal band was not observed in

Table I: Variations in Sequence from Previous Reports[a]

	reference									
position	1	2	3	4	5	6	7	8	9	codon
−184	T	T				T	C			
−97	G	G				G	C			
−93	A	A				A	C			
−87	A	A				A	AA			
−34	C	C				C	T			
1	C	C				C	CC			
34	A	A	A			A	CA			
67	G	G	G			G	C			
68	G	G	G			G	C			
69	C	C	C			C	G			
122	G	G				G	C			
123	A	A				A	G	G		
2470	G	G	G		A					G2
2498	A	A	A	C	A					T9
2658	A	A	A	G	A					H65
2816	G*	A								
2827	C*	CC								
5306	A	A						G		
5307	T	T						C		
5419	A	A	A	A	G					V141
6444	A	A	A	G	A					Q181
6453	G	G	G	G	T					A184
6492	A	A	A	A	T					R197
6578	T	T						C		
6586	G	G						–		
6591	A	A						–		
7473	G	G	G	G	A					Q254
7485	C	C	C	T	C					F258
7665	C	C	T	C	C				C	V318
7795	TTCTCC	TTCTCC					–	TTCTCC		
9795	C*	–						–		
9928	C*	–								
13339	T	T	C	T	T				T	P407

[a] References: (1) this report; (2) Bock et al. (1988); (3) Bock et al. (1982); (4) Prochownik et al. (1983); (5) Chandra et al. (1983); (6) Bock and Levitan (1983); (7) Prochownik and Orkin (1984); (8) Prochownik et al. (1985); (9) Jagd et al. (1985). The table excludes the polymorphic sites listed in the results. Blank: not reported by the authors. (*) This has been confirmed by sequencing numerous samples of amplified genomic DNA. (–) Not present in this sequence.

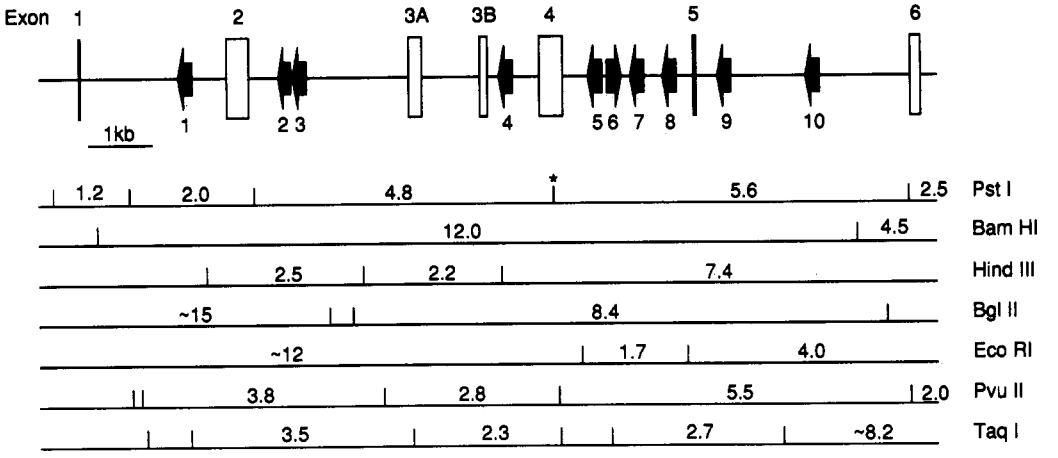

FIGURE 2: Scale diagram of the antithrombin gene, with the horizontal line representing the extent of the sequence information in Figure 1. Exons are shown as open boxes, and Alu repeats, as solid arrows. Cutting sites and normal fragment sizes are shown for several restriction endonucleases used to map the antithrombin locus (Table I). The *Pst*I site within exon 4 (*) is polymorphic, the absence of the site giving rise to a 10.4-kb fragment.

hybridization of *Pvu*II-digested DNA with the antithrombin cDNA since the 2.8-kb deletion is likely to remove exon 5. *Taq*I digestion also revealed only the normal pattern of bands; a 2.8-kb deletion around exon 5 would reconstitute an abnormal fragment of 8.1–8.2 kb which would comigrate with the normal 8.2-kb band. Overall, the pattern of restriction fragments observed in G052 was consistent with a deletion of ~2.8 kb between the *Eco*RI site in intron 4 and the *Bam*HI site in intron 5.

Localization of the Deletion Breakpoints by PCR Amplification and Restriction Digestion. The approximate breakpoints of the deletion having been identified, further characterization was carried out by analzying a series of amplified products produced by PCR. Pairs of oligonucleotide primers flanking the putative deletion breakpoints, as suggested by the gross mapping data, were synthesized and used to amplify DNA from normal individuals and G052 (Figure 4). The expected sizes of the amplification products in normal

Table II: Alu Repeat Elements within the Antithrombin Gene[a]

	location	first nt	last nt	orientation	% identity	class	specificity
Alu 1	intron 1	1649	1935	–	87	II	22/26
Alu 2	intron 2	3250	3431	–	89	II	11/14
Alu 3	intron 2	3473	3750	–	91	II	21/26
Alu 4	intron 3B	6743	7024	–	93	IV	24/26
Alu 5	intron 4	8158	8418	–	90	III	23/26
Alu 6	intron 4	8460	8740	+	82	I	17/26
Alu 7	intron 4	8828	9100	–	90	II	21/26
Alu 8	intron 4	9337	9618	–	84	II	21/26
Alu 9	intron 5	10198	10477	–	86	II	21/26
Alu 10	intron 5	11603	11871	–	90	II	19/25

[a] Orientation: (–) indicates that the Alu repeat is orientated in the reverse direction, and (+) indicates that the Alu sequence is oriented in the forward direction. % identity is calculated as the similarity to a consensus Alu sequence (Britten et al., 1988) excluding the diagnostic positions and counting insertions and deletions as 1 irrespective of size. The class of Alu is according to the same reference, with the specificity indicating how many of the nucleotides in diagnostic positions are the same as the dominant nucleotide for the class categorization.

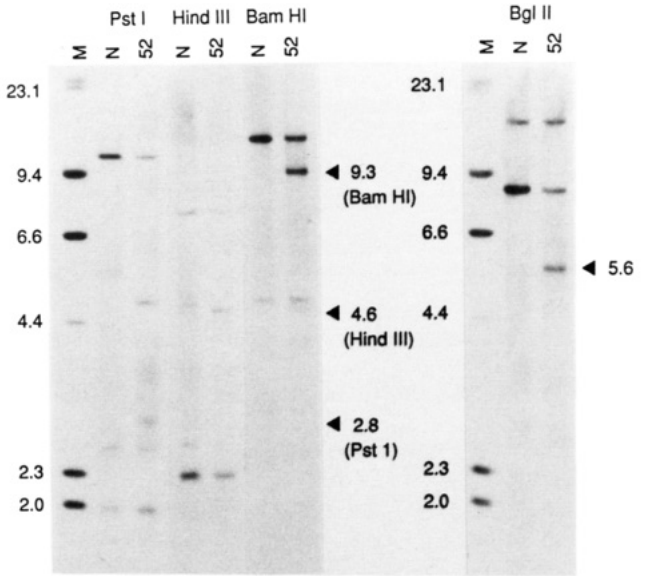

FIGURE 3: Southern blot hybridization of genomic DNA from a normal (N) individual and G052 with antithrombin cDNA probe. The sizes (in kb) of the abnormal fragments detected in G052 are indicated at the right of each panel. The marker is λ cut with *Hin*dIII. The *Pst*I, *Hin*dIII, and *Bam*HI digests were run on the same gel, and so the marker lane adjacent to the *Pst*I digest is applicable to the other lanes.

Table III: Restriction Fragment Sizes Observed in Normal and G052 Genomic DNA, Probed with Antithrombin cDNA[a]

enzyme	fragment sizes (kb) normal	G052
*Pst*I	2.0, 10.4,[b] 2.5	2.0, 10.4/4.8 + 2.8,[b] 2.5
*Eco*RI	12.0, 4.0	12.0, 4.0
*Bam*HI	12.0, 4.5	12.0, 9.3, 4.5
*Hin*dIII	2.5, 2.2, 7.4	2.5, 2.2, 7.4, 4.6
*Bgl*II	15, 8.4	15, 8.4, 5.6
*Pvu*II	3.8, 2.8, 5.5, 2.0	3.8, 2.8, 5.5, 2.0
*Taq*I	3.5, 2.3, 2.7, 8.2	3.5, 2.3, 2.7, 8.2

[a] Fragment sizes are listed in order from the 5′-end of the locus. Abnormal band sizes generated from the mutant allele are underlined. [b] Bands produced by the polymorphic *Pst*I site within exon 4 (see text).

individuals are shown in table IV; with the exception of primers AT17 and AT10, amplification produced bands of the predicted sizes in normal DNA (Figure 4). No product was observed in normal DNA with primers AT17 and AT10, presumably because under the conditions chosen amplification

Table IV: PCR Product Sizes Using Primers Spanning the Deletion Site

primers	PCR product sizes (kb) normal[a]	G052[b]
AT17-AT10	5.998	3.2
AT17-AT63	4.409	1.6
AT48-AT63	3.847	1.1
AT69-AT63	3.093	0.33
AT48-AT10	5.442	2.7

[a] Predicted from the normal sequence. [b] See Figure 4.

of the predicted 5998 bp was inefficient. Amplification of DNA from G052, with the same conditions and pairs of oligonucleotide primers used for normal DNA, produced abnormal bands consistently 2.8 kb shorter than normal (Table IV and Figure 4). In G052 preferential amplification of the abnormal (shorter) allele over the normal allele was observed; this probably reflects the greater efficiency of amplification of the shorter DNA sequence. No specific product could be amplified in DNA from G052 when primers located between AT69 and AT63 were used (not shown).

To further localize the ends of the deletion and to confirm the nature of the amplified product, the PCR product derived from G052 using primers AT48 and AT10 was analyzed. This 2.7-kb fragment was digested with restriction endonucleases that had cutting sites located in the vicinity of the predicted deletion (Figure 4). The site for *Dra*I, located 5′ to primer AT69, was intact, as indicated by bands of 370 bp and 2.3 kb. Similarly, the cleavage site for *Xba*I, which is located at the 3′-end of the fragment 3′ to the site for primer AT63, was shown to be intact. The *Hin*cII digest produced fragments compatible with the presence of the two predicted cutting sites, one immediately 5′ to AT69 giving rise to a band of approximately 600 bp and the other located immediately 3′ to AT63 giving rise to a 1.3-kb band. The third band of about 770 bp, representing the area including the deletion site, was about 2.8 kb shorter than that predicted from the sequence of a normal allele, confirming the size of the deletion. As can be seen in Figure 4, within the normal sequence between primers AT48 and AT10, there are two *Nco*I cutting sites. Presence of the 3′-site results in a fragment of 1.2 kb, which was observed, while the other site, located 443 bp 5′ to AT63, was missing in the amplified DNA from G052. This suggested that the 3′-end of the deletion was located between AT63 and the 5′ *Nco*I site. The 5′-end of the deletion could be further localized by the presence of a *Cfo*I site 3′ to AT48 which was found to be intact. A second *Cfo*I site found in normal DNA 1.4 kb 3′ to AT48 was deleted. The amplified DNA was not cut by either *Eco*RI or *Sca*I The 5′ *Sca*I site shown in Figure 4 is located 327 bp 3′ to the *Cfo*I site intact in G052, thereby localizing the 5′-end breakpoint to within this length of DNA. Therefore, by amplification of deletion-specific fragments and restriction analysis of the products, the 5′-breakpoint was localized to within positions 8957 and 9287 of the normal sequence (Figure 1), and the 3′-breakpoint, to between positions 11485 and 11904.

Sequencing of Amplified DNA Spanning the Deletion Breakpoints. A series of amplified products spanning the deletion was prepared, using the primers shown in Figure 4. The double-stranded products were then sequenced in both directions, and the sequences were compared to the previously determined normal sequence (Figure 5). In total the deletion removed 2761 bp, with the 5′ breakpoint lying within Alu 7 in intron 4 and the 3′ breakpoint located within Alu 10 in intron 5. Because of the high degree of sequence identity

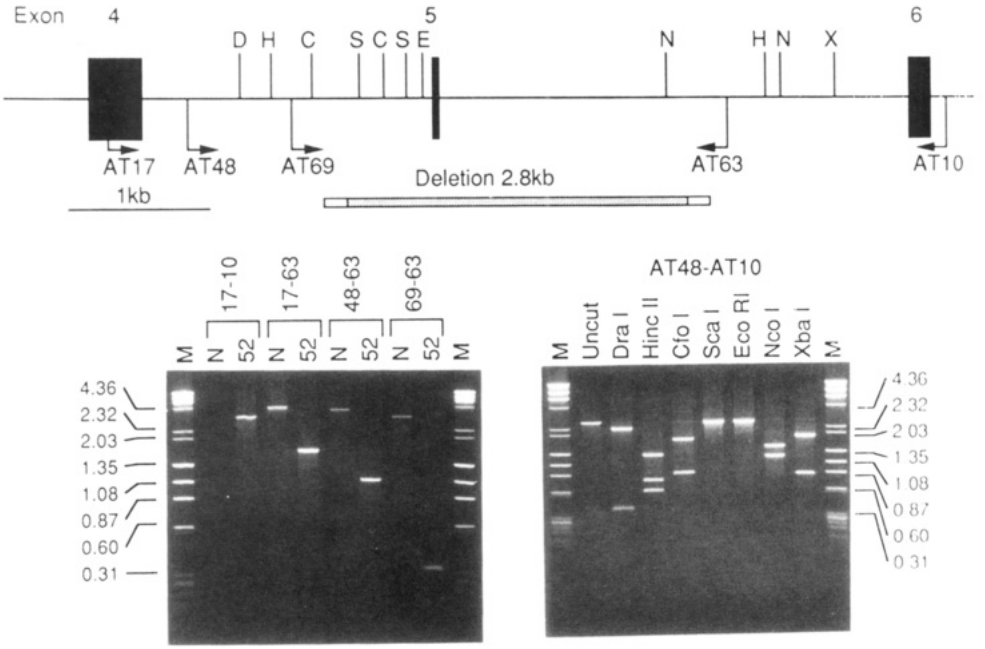

FIGURE 4: Mapping of the intragenic deletion in G052. Oligonucleotide primers used to amplify across the deletion site are shown as arrows. The left panel shows fragments resulting from the amplifications of DNA from both a normal individual (N) and G052. Restriction enzyme cutting sites are indicated on the gene: D, *Dra*I; H, *Hinc*II; C, *Cfo*I; S, *Sca*I; E, *Eco*RI; N, *Nco*I; X, *Xba*I. The right panel illustrates the fragments resulting from the digestion of the abnormal allele from G052 amplified using primers AT48 and AT10. Marker (M) is λ *Hin*dIII and *φX174 Hae*III, with sizes shown in kb. The 2.8-kb deletion derived from these mapping experiments is shown as a stippled line, with the open segments indicating the uncertainty of the deletion breakpoints.

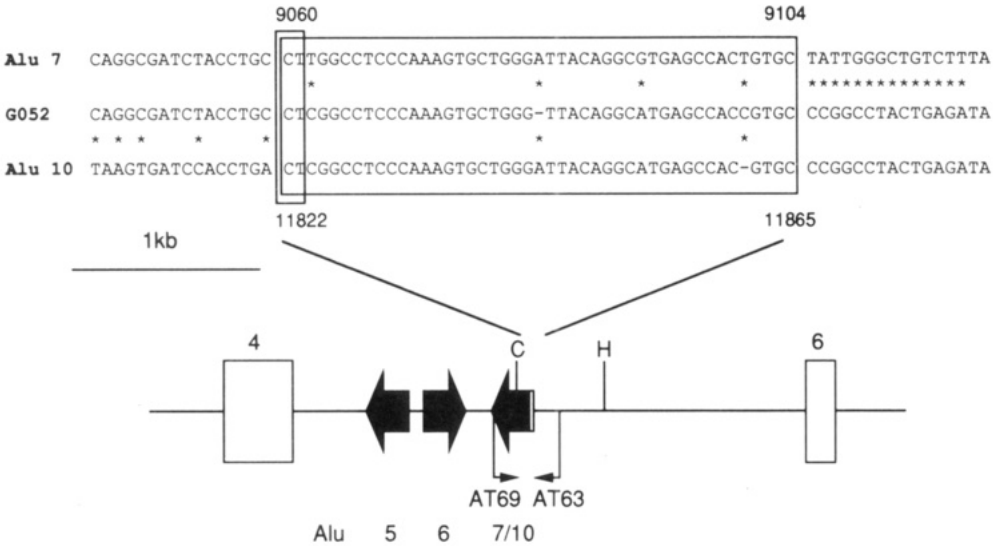

FIGURE 5: Sequence across the recombination junction in G052. The sequences of normal Alu 7 and normal Alu 10 are aligned above and below. Asterisks show sequence differences between G052 and the Alu repeats. The two boxes enclose the probable areas of recombination as described in Results. Nucleotide positions which form the limits of the breakpoints in Alu 7 and Alu 10 are as in Figure 1. The structure of the abnormal allele from G052 in the region of the deletion site is shown below, with the oligonucleotide primers and restriction enzyme sites which were mapped closest to the deletion (compare with Figure 4). The fragment of Alu 10 which has recombined with Alu 7 is unshaded.

between these two Alu elements (85% over 270 bp), it was not possible to determine the exact site of recombination. As is shown in Figure 5, the 3′ sequence from G052 in the region of the recombination junction was unequivocally derived from Alu 10. However, within a span of 45 nucleotides 5′ to position 11865 of the normal sequence, enclosed in the larger box in Figure 5, the origin of the sequence in G052 could not be assigned. In this area there is a high degree of identity to the sequence from positions 9060–9104 within Alu 7. Further 5′ to this region, proximal to normal position 9060, there was complete homology between normal sequence and G052, indicating that the breakpoint lay 3′ to position 9060. Within the 45-nucleotide area spanning the recombination junction the sequence of G052 differed at two positions from both normal Alu 7 and Alu 10. These sequence differences may have been generated during the recombination event or may represent sequence variations within antithrombin genes. Two further sequence differences to Alu 7 were observed in this area. If these represent true differences rather than point mutations associated with the recombination event, then the area of the breakpoint junction can be localized to within 2 nucleotides, enclosed within the smaller box in Figure 5.

DISCUSSION

Sequence analysis of the normal antithrombin locus revealed the presence of 10 Alu repeat elements, one within both introns 1 and 3B, two within intron 2, four within intron 4, and two within intron 5. Ten repeats represents about 22% of the

intron sequence, which is considerably greater than the estimated 5% of the human genome accounted for by Alu repeats. It is interesting to note, however, that Alu repeats are also frequent in the genes for CI-inhibitor (Carter et al., 1991), heparin cofactor II (Herzog et al., 1991), and plasminogen activator inhibitor I (Bosma et al., 1988), which, like antithrombin, are members of the serpin family. Alu elements are short interspersed repetitive DNA sequences, with perhaps as many as 5×10^5 copies per haploid genome and an average spacing of about 4 kb (Shen et al., 1991). In general, Alu repeats are about 280 bp in length, and although they are a unique feature of primate genomes, they show identity to the 130-bp mouse B1 repeat element (Schmid & Jelinek, 1982). Each element is composed of two homologous but nonidentical fragments of about 130 bp, Alu-left and Alu-right, separated by an A-rich region. The Alu-right fragment has a 31-bp insertion not found in Alu-left, and the 3′-end typically has an A-rich tail composed of short repeats. A consensus Alu sequence has been derived, and individual copies show a high degree of sequence identity; this led to the suggestion that Alu elements are derived from 7SL RNA (Ullu & Tschudi, 1984), which forms part of the signal recognition particle involved in translocation of proteins across the endoplasmic reticulum. Several classes of Alu repeats can be identified on the basis of sequences at diagnostic positions (Britten et al., 1988; Jurka & Smith, 1988), leading to the proposal that the elements have been incorporated into the genome during successive waves by retrotransposition. Representatives from each of the four classes of Alu repeat, according to the classification of Britten et al. (1988), were identified within the antithrombin introns (Table II). The majority belong to class II, which is the most commonly identified class within the human genome. Alu 4 within intron 3B, however, is of class IV, the most recent in evolutionary terms, and has probably been inserted within the last 20 million years (Britten et al., 1989). Alu 6 within intron 4 is a class I repeat, a representative of the oldest group with the least divergence from the human 7SL sequence. The mixture of Alu classes is consistent with the occurrence of several evolutionary steps in the generation of the current antithrombin gene structure.

The genomic basis for the vast majority of type I deficiencies reported to date has been point changes within the antithrombin gene, producing either termination codons, frameshifts with premature termination, RNA processing defects, or amino acid deletions or substitutions (Lane et al., 1991, 1992b). Rare instances of complete gene deletion have been reported (Winter et al., 1982; Bock & Prochownik, 1987), although the extent of the abnormalities in such cases has not been investigated. One other gene deletion has been reported in abstract form (Fernandez-Rachubinski et al., 1990). By screening affected individuals from 50 kindreds with type I deficiency, we have identified four cases of gene rearrangements (Olds et al., 1992), including the present case. On the basis of this data, gross structural rearrangements of the antithrombin locus account for less than 10% of inherited type I deficiencies.

In subject G052 reported here, data from restriction enzyme mapping was compatible with an intragenic deletion of about 2.8 kb removing exon 5 and parts of the flanking introns 4 and 5. All of the Southern blotting data were consistent with the predicted restriction enzyme sites based on the genomic sequence. By the use of PCR it was possible to amplify deletion-specific fragments from the affected allele and finally to localize the deletion endpoints by direct sequence analysis. It was found that the deletion had arisen by recombination between the repeat sequences presented by Alu 7 within intron 4 and Alu 10 within intron 5. Retrotransposition of Alu elements and subsequent homologous recombination has been suggested as a major factor in the generation of diversity during primate evolution (Britten et al., 1988; Britten et al., 1989). Alu–Alu recombinatin has also been identified as underlying several deletional events within the genes of the LDL receptor (Lehrman et al., 1987b; Rudiger et al., 1991), CI-inhibitor (Stoppa-Lyonnet et al., 1991), adenosine deaminase (Markert et al., 1988; Berkvens et al., 1990), apolipoprotein B (Huang et al., 1989), α-galactosidase A (Kornreich et al., 1990), and the α gene complex (Nicholls et al., 1987). The same mechanism has given rise to duplications in the LDL receptor gene (Lehrman et al., 1987a) and the dystrophin locus (Hu et al., 1991) and has been identified as the basis for an X,Y translocation (Rouyer et al., 1987) and evolution of glycophorin B from the glycophorin A gene (Kudo & Fukuda, 1989). Insertions of Alu elements, presumably by retrotransposition, have also been described, affecting the genes for ornithine δ-aminotransferase (Mitchell et al., 1991), cholinesterase (Muratani et al., 1991), and NF1 (Wallace et al., 1991). Within the antithrombin locus, nine of the ten Alu repeats, including both Alu 7 and Alu 10, are present in the reverse orientation. Recombination between repeats oriented in the same direction is most probably the result of mispairing of the homologous (but not identical) Alu elements derived from different chromatids during meiosis, followed by unequal crossing over. Although there remains some ambiguity concerning the exact location of the breakpoints in G052, the deletion left the majority of Alu 7 intact and removed almost all of Alu 10. The breakpoints within each of the Alu repeats lie between the sequences in the left monomer homologous to the A and B boxes of the split polymerase III promotor (Paolella et al., 1983). This is the region of Alu repeats most commonly involved in recombination (Lehrman et al., 1987a; Stoppa-Lyonnet et al., 1991). It has been suggested that recombination here may be encouraged in some way by more efficient transcription of the Alu-left fragment and the resultant unwinding of the DNA known to accompany this process (Lehrman et al., 1987b).

The deletion in subject G052 is associated with type I antithrombin deficiency and the absence of detectable variant protein in the plasma. The deletion predicts the removal of exon 5 from the coding sequence, along with flanking intron sequences which include the 3′-splice site of intron 4 and the 5′-splice site of intron 5. The intron 4 splice junction is a class 1 type, in that exon 4 contains the first nucleotide of codon 353 while the second and third nucleotides are encoded within exon 5. The deletion of exon 5, therefore, induces a shift in the frame of translation beyond P352. If the 3′-splice site of intron 5 is utilized during RNA processing, in the absence of the 3′-splice site in intron 4, the predicted mRNA then contains a termination signal in codon 356. This RNA would encode a variant protein with the 77 carboxy-terminal residues of the normal antithrombin protein deleted. Previously identified antithrombin mutations which affect the carboxy-terminal end of the protein are known to be associated with major perturbations in the structure of the inhibitor, leading to a reduction in the concentration or an absence from the plasma of antithrombin (Bock et al., 1988; Olds et al., 1991a; Lane et al., 1992a). It is not surprising, therefore, that the individual investigated in this study has a type I deficiency phenotype.

The completion of the genomic sequence for the antithrombin locus should allow the further characterization of

other partial gene deletions and indicate whether or not recombination between the high number of Alu repeat elements is an important factor in the generation of an apparently uncommon event as a basis for inherited antithrombin deficiency.

ACKNOWLEDGMENT

We thank Dr. E. Prochownik for the antithrombin cDNA probes.

REFERENCES

Abildgaard, U. (1969) *Scand. J. Clin. Lab. Invest. 24*, 23–27.
Berkvens, T. M., van Ormondt, H., Gerritsen, E. J. A., Meera Khan, P., & van der Eb, A. J. (1990) *Genomics 7*, 486–490.
Bock, S. C., & Levitan, D. J. (1983) *Nucleic Acids Res. 11*, 8569–8582.
Bock, S. C., & Prochownik, E. V. (1987) *Blood 70*, 1273–1278.
Bock, S. C., Wion, K. L., Vehar, G. A., & Lawn, R. M. (1982) *Nucleic Acids Res. 10*, 8113–8125.
Bock, S. C., Harris, J. F., Balazs, I., & Trent, J. M. (1985) *Cytogenet. Cell Genet. 39*, 67–69.
Bock, S. C., & Radziejewska, E. (1991) *Nucleic Acids Res. 19*, 2519.
Bock, S. C., Marrinan, J. A., & Radziejewska, E. (1988) *Biochemistry 27*, 6171–6178.
Bosma, P. M., van den Berg, E. A., Kooistra, T., Siemieniak, D. R., & Slightom, J. L. (1988) *J. Biol. Chem. 263*, 9129–9141.
Britten, R. J., Baron, W. F., Stout, D. B., & Davidson, E. H. (1988) *Proc. Natl. Acad. Sci. U.S.A. 85*, 4770–4774.
Britten, R. J., Stout, D. B., & Davidson, E. H. (1989) *Proc. Natl. Acad. Sci. U.S.A. 86*, 3718–3722.
Carter, P. E., Duponchel, C., Tosi, M., & Fothergill, J. E. (1991) *Eur. J. Biochem. 197*, 301–308.
Chandra, T., Stackhouse, R., Kidd, V. J., & Woo, S. L. C. (1983) *Proc. Natl. Acad. Sci. U.S.A. 80*, 1845–1848.
Daly, M. E., & Perry, D. J. (1990) *Nucleic Acids Res. 18, 5583.*
Egeberg, O. (1965) *Thromb. Diath. Haemorrh. 13*, 516–530.
Fernandez-Rachubinski, F., Rachubinski, R., & Blajchman, M. A. (1990) *Blood 76*, 506a.
Herzog, R., Lutz, S., Blin, N., Marasa, J. C., Blinder, M. A., & Tollefesen, D. M. (1991) *Biochemistry 30*, 1350–1357.
Hu, X., Ray, P. N., & Worton, R. G. (1991) *EMBO J. 19*, 2471–2477.
Huang, L. S., Ripps, M. E., Korman, S. H., & Deckelbaum, R. J. (1989) *J. Biol. Chem. 264*, 11394–11400.
Huber, R., & Carrell, R. W. (1989) *Biochemistry 28*, 8951–8966.
Jagd, S., Vibe-Pedersen, K., & Magnusson, S. (1985) *FEBS Lett. 193*, 213–216.
Jurka, J., & Smith, T. (1988) *Proc. Natl. Acad. Sci. U.S.A. 85*, 4775–4778.
Kornreich, R., Bishop, D. F., & Desnick, R. J. (1990) *J. Biol. Chem. 265*, 9319–9326.
Kudo, S., & Fukuda, M. (1989) *Proc. Natl. Acad. Sci. U.S.A. 86*, 4619–4623.
Lane, D. A., Ireland, H., Olds, R. J., Thein, S. L., Perry, D. J., & Aiach, M. (1991) *Thromb. Haemostasis 66*, 657–661.
Lane, D. A., Olds, R. J., Conard, J., Boisclair, M. Bock, S. C., Hultin, M., Abildgaard, U., Ireland, H., Thompson, E., Sas, G., Horrelou, M. H., Tamponi, G., & Thien, S. L. (1992a) *J. Clin. Invest. 90*, 2422–2433.
Lane, D. A., Olds, R. J., & Thein, S. L. (1992b) *Blood Coagulation Fibrinolysis 3*, 315–341.
Lehrman, M. A., Goldstein, J. L., Russell, D. W., & Brown, M. S. (1987a) *Cell 48*, 827–835.
Lehrman, M. A., Russell, D. W., Goldstein, J. L., & Brown, M. S. (1987b) *J. Biol. Chem. 262*, 3354–3361.
Marcum, J. A., & Rosenberg, R. D. (1984) *Biochemistry 23*, 1730–1737.
Marcum, J. A., Atha, D. H., Fritze, L. M. S., Nawroth, P., Stern, D., & Rosenberg, R. D. (1986) *J. Biol. Chem. 261*, 7507–7517.
Markert, M. L., Hutton, J. J., Wiginton, D. A., States, J. C., & Kaufman, R. E. (1988) *J. Clin. Invest. 81*, 1323–1327.
Mitchell, G. A., Labuda, D., Fontaine, G., Saudubray, J. M., Bonnefont, J. P., Lyonnet, S., Brody, L. C., Steel, G., Obie, C., & Valle, D. (1991) *Proc. Natl. Acad. Sci. U.S.A. 88*, 815–819.
Muratani, K., Hada, T., Yamamoto, Y., Kaneko, T., Shigeto, Y., Ohue, T., Furuyama, J., & Higashino, K. (1991) *Proc. Natl. Acad. Sci. U.S.A. 88*, 11315–11319.
Nicholls, R. D., Fischel-Ghodsian, N., & Higgs, D. R. (1987) *Cell 49*, 369–378.
Olds, R. J., Lane, D. A., Finazzi, G., Barbui, T., & Thein, S. L. (1990) *Blood 76*, 2182–2186.
Olds, R. J., Lane, D. A., Ireland, H., Leone, G., De Stefano, V., Cazenave, J. P., Wiesel, M. L., & Thein, S. L. (1991a) *Br. J. Haematol. 78*, 409–413.
Olds, R. J., Lane, D. A., Caso, R., Panico, M., Morris, H. R., Sas, G., Dawes, J., & Thein, S. L. (1991b) *Blood 79*, 1206–1212.
Olds, R. J., Lane, D. A., Chowdhury, V., Samson, D., De Stefano, V., Leone, G., Wiesel, M. L., Cazenave, J. P., Conard, J., & Thein, S. L. (1992) *Br. J. Haematol. Int. Soc. Haematol., Congr. 24th*, **1992**, *Abstr.* 40.
Paolella, G., Lucero, M. A., Murphy, M. H., & Baralle, F. E. (1983) *EMBO J. 2*, 691–696.
Prochownik, E. V., & Orkin, S. H. (1984) *J. Biol. Chem. 259*, 15386–15392.
Prochownik, E. V., Markam, A. F., & Orkin, S. H. (1983) *J. Biol. Chem. 258*, 8389–8394.
Prochownik, E. V., Bock, S. C., & Orkin, S. H. (1985) *J. Biol. Chem. 260*, 9608–9612.
Rosenberg, R. D., & Damus, P. S. (1973) *J. Biol. Chem. 248*, 6490–6505.
Rouyer, F., Simmler, M. C., Page, D. C., & Weissenbach, J. (1987) *Cell 51*, 417–425.
Rudiger, N. S., Hansen, P. S., Jorgensen, M., Faegeman, O., Bolund, L., & Gregersen, N. (1991) *Eur. J. Biochem. 198*, 107–111.
Schmid, C. W., & Jelinek, W. R. (1982) *Science 216*, 1065–1070.
Shen, M. R., Batzer, M. A., & Deininger, P. L. (1991) *J. Mol. Evol. 33*, 311–320.
Stoppa-Lyonnet, D., Duponchel, C., Meo, T., Laurent, J., Carter, P. E., Arala-Chaves, M., Cohen, J. H. M., Dewald, G., Goetz, J., Hauptmann, G., Lagrue, G., Lesavre, P., Lopez-Trascasa, M., Misiano, G., Moraine, C., Sobel, A., Spath, P. J., & Tosi, M. (1991) *Am. J. Hum. Genet. 49*, 1055–1062.
Thein, S. L., & Hinton, J. (1991) *Br. J. Haematol. 79*, 113–115.
Ullu, E., & Tschudi, C. (1984) *Nature 312*, 171–172.
Wallace, M. R., Andersen, L. B., Saulino, A. M., Gregory, P. E., Glover, T. W., & Collins, F. S. (1991) *Nature 353*, 864–866.
Winter, J. H., Bennett, B., Watt, J. L., Brown, T., San Roman, C., Schinzel, A., King, J., & Cook, P. J. L. (1982) *Ann. Hum. Genet. 26*, 29–34.