Kelly, R. C., Jensen, D. E., & von Hippel, P. H. (1976) *J. Biol. Chem. 251*, 7240–7250.
King, G. C., & Coleman, J. E. (1987) *Biochemistry 26*, 2929–2937.
Kowalczykowski, S. C., Paul, L. S., Lonberg, N., Newport, J. W., McSwiggen, J. A., & von Hippel, P. H. (1986) *Biochemistry 25*, 1226–1240.
Laemmli, U. K. (1970) *Nature (London) 227*, 680–685.
McGhee, J. D. (1976) *Biopolymers 15*, 1345–1375.
Moise, H., & Hosoda, J. (1976) *Nature (London) 259*, 455–458.
Monaco, H. L., Crawford, J. L. & Lipscomb, W. N. (1978) *Proc. Natl. Acad. Sci. U.S.A. 75*, 5276–5280.
O'Connor, T. P., & Coleman, J. E. (1983) *Biochemistry 22*, 3375–3381.
Prigodich, R. V., Casas-Finet, J., Williams, K. R., Konigsberg, W. H., & Coleman, J. E. (1984) *Biochemistry 23*, 522–529.
Prigodich, R. V., Shamoo, Y., Williams, K. R., Chase, J. W., Konigsberg, W. H., & Coleman, J. E. (1986) *Biochemistry 25*, 3666–3672.
Spicer, E. K., Williams, K. R., & Konigsberg, W. H. (1979) *J. Biol. Chem. 254*, 6433–6436.
Williams, K. R., & Konigsberg, W. H. (1978) *J. Biol. Chem. 253*, 2463–2470.
Williams, K. R. Sillerud, L. O., Schafer, D. E., & Konigsberg, W. H. (1979) *J. Biol. Chem. 254*, 6426–6432.
Williams, K. R., Spicer, E. K., LoPresti, M. B., Guggenheimer, R. A., & Chase, J. W. (1983) *J. Biol. Chem. 258*, 3346–3355.

Characterization of the M1(Ala^{213}) Type of α1-Antitrypsin, a Newly Recognized, Common "Normal" α1-Antitrypsin Haplotype

Toshihiro Nukiwa,* Mark Brantly,‡ Fumitaka Ogushi, Gerald Fells, Ken Satoh, Larue Stier, Michael Courtney, and Ronald G. Crystal

Pulmonary Branch, National Heart, Lung, and Blood Institute, Bethesda, Maryland 20892, and Transgene SA, 67000 Strasbourg, France

Received January 6, 1987; Revised Manuscript Received March 10, 1987

ABSTRACT: α1-Antitrypsin (α1AT) is a highly pleomorphic 52-kDa serum glycoprotein that functions as the major inhibitor of neutrophil elastase. Of these, the most common normal α1AT haplotypes identified by isoelectric focusing (IEF) of serum are those of the M family, including M1, M2, and M3. In the course of studying the α1AT type Z gene, we identified a restriction endonuclease *Bst*EII polymorphism in the M1 gene that predicted the existence of a previously unidentified, but relatively common, haplotype of M, referred to as M1(Ala^{213}) [Nukiwa, T., Satoh, K., Brantly, M. L., Ogushi, F., Fells, G. A., Courtney, M., & Crystal, R. G. (1986) *J. Biol. Chem. 261*, 15989–15994]. In this study we have cloned both α1AT genes from an individual heterozygous for the M1(Ala^{213}) and M1(Val^{213}) haplotypes. Sequencing of the coding exons of both demonstrated that they are identical except for the Ala–Val difference at residue 213. The codominant transmission of the M1(Ala^{213}) gene was demonstrated in a family study. Evaluation of 39 genomic samples of Caucasians with the IEF haplotype M1 demonstrated haplotype frequencies of 68% for M1(Val^{213}) and 32% for M1(Ala^{213}). α1AT serum levels of individuals inheriting the M1(Ala^{213}) gene in a homozygous fashion were in the same range as those for homozygous M1(Val^{213}) as was the rate of association of the M1(Ala^{213}) protein with neutrophil elastase. Interestingly, comparison of the M1(Ala^{213}) gene sequence to all of the known α1AT sequences at the gene, cDNA, and protein levels demonstrated that M1(Ala^{213}) is the closest to the baboon α1AT coding exons, suggesting that M1(Ala^{213}) is the "oldest" type human α1AT known.

α1-Antitrypsin (α1AT, also referred to as α1-antiprotease or α1-antiproteinase) is a 52-kDa glycoprotein produced and secreted by hepatocytes and mononuclear phagocytes (Gadek & Crystal, 1982; Travis & Salvesen, 1983; Carrell, 1986). The mature protein is comprised of a single polypeptide chain of 394 amino acids and three, *N*-asparaginyl-linked complex carbohydrate side chains (Carrell et al., 1982; Long et al., 1984; Mega et al., 1980). In normal individuals, α1AT comprises approximately 90% of the α1-globulin band of conventional serum protein electrophoretic pattern (Gadek & Crystal, 1982), and the serum levels of α1AT are 150–350 mg/dL (Gadek & Crystal, 1982; Keuppers, 1978). Although α1AT is a broad-spectrum antiprotease capable of complexing with and inhibiting a variety of serine proteases, its major function in the human is as an inhibitor of neutrophil elastase (EC 3.4.21.11) (Travis & Salvesen, 1983), an omnivorous protease capable of cleaving many proteins, including most proteins that comprise the structural backbone of tissues (Bieth, 1986). α1AT interacts with neutrophil elastase through the Met^{358}–Ser^{359} residues of α1AT (Travis & Salvesen, 1983; Johnson & Travis, 1978), a sequence that resides on the outside of the molecule (Loeberman et al., 1984). In the normal form of α1AT, this interaction occurs with an association rate constant (K_a) of approximately 10^7 M^{-1} s^{-1} (Beatty et al., 1980). Because the off-rate is so slow, this interaction is essentially irreversible, and neither the α1AT nor the elastase is capable of further function (Travis & Salvesen, 1983; Beatty et al., 1980).

α1AT is coded for by a single gene comprised of five exons and four introns encompassing approximately 10 kb of chromosome 14 (Long et al., 1984; Schroeder et al., 1985). More than 30 haplotypes of α1AT have been described; conven-

* Address correspondence to this author at the National Heart, Lung, and Blood Institute.
‡ M.B. is a Parker B. Francis Fellow in Pulmonary Research.

tionally these are defined by different patterns of α1AT bands observed on the isoelectric focusing (IEF) of serum between pH 4.0 and pH 5.0 (Cox et al., 1980). The haplotypes are named by letters according to their migration pattern; those in the middle are M type while the anodal haplotypes start with A and the cathodal haplotypes end with Z (Gadek & Crystal, 1982; Cox et al., 1980). The two parental haplotypes are inherited in an autosomal condominant fashion, and the phenotype is referred to as the Pi (protease inhibitor) type (Gadek & Crystal, 1982; Cox et al., 1980; Fagerhol & Cox, 1981). Studies of large populations in the USA and Europe have shown that the M-type haplotypes are the most common, comprising approximately 95% of all α1AT haplotypes (Gadek & Crystal, 1982; Kueppers, 1978; Fagerhol & Cox, 1981). On the basis of isoelectric focusing data, among the M haplotypes, three (M1, M2, and M3) are most frequent (Gadek & Crystal, 1982; Keuppers, 1978; Fagerhol & Cox, 1981; Dykes et al., 1984; Kueppers & Christopherson, 1978). Two other M haplotypes (M4, M5) have been identified by isoelectric focusing of serum, but these are relatively rare (Constans et al., 1980; Klasen et al., 1982; Weidinger et al., 1985). Thus, most individuals have the α1AT phenotype PiM_nM_n (where n = 1, 2, or 3). For all combinations of inheritance of the M1, M2, and M3 haplotypes, the serum α1AT levels are normal (Gadek & Crystal, 1982; Beckman & Beckman, 1980).

The importance of α1AT is highlighted by α1AT deficiency, a hereditary disorder characterized by serum levels of α1AT <35% of normal and associated, in the adult, with a high risk for the development of emphysema (Gadek & Crystal, 1982; Carrell, 1986; Laurell & Eriksson, 1963; Carrell & Owen, 1979). In the context of the function of α1AT as an inhibitor of neutrophil elastase and the capabilities of active neutrophil elastase to destroy the lower respiratory tract, the pathogenesis of the emphysema associated with α1AT deficiency is understood to be based on an imbalance of α1AT and neutrophil elastase in the lower respiratory tract such that there is insufficient α1AT to prevent destruction by the chronic burden of neutrophil elastase to which this tissue is exposed (Gadek & Crystal, 1982; Gadek et al., 1981). Most individuals with α1AT deficiency have the α1AT phenotype PiZZ (Gadek & Crystal, 1982; Carrell, 1986; Carrell & Owen, 1979).

During an investigation of the primary structure of the Z gene of α1AT, we discovered a restriction-site polymorphism for *Bst*EII in exon III coding for amino acid 213 (GTG in the M1 gene, GCG in the Z gene) (Nukiwa et al., 1986b). When oligonucleotides were constructed complementary to this region in the M1 and Z genes and used to confirm the universality of this sequence among individuals with the Z haplotype, we noted that a significant proportion of individuals identified by isoelectric focusing of serum as having the M1 haplotype were *Bst*EII negative; i.e., they had the apparent M1 haplotype but actually the sequence GCG (i.e., Ala^{213}) rather than GTG (i.e., Val^{213}). This led us to predict the existence of a different, relatively common haplotype of M which we referred to as M1(Ala^{213}) (Nukiwa et al., 1986b). In this context, this study is directed toward defining the primary structure, frequency, and function of M1(Ala^{213}). This has been accomplished by cloning the M1(Ala^{213}) gene, sequencing the structural exons, purifying the protein from the plasma of an M1(Ala^{213})M1(Ala^{213}) homozygote, and using a 0.95-kb *Pst*I–*Bst*EII fragment of genomic DNA to determine the frequency of M1(Ala^{213}) among the normal M haplotypes.

Materials and Methods

Study Population. The study population consisted of 69 unrelated individuals. Grouped by their α1AT serum phenotype (see below), the population included the following: M1M1 (n = 50), M1M2 (n = 5), M2M2 (n = 5), M2M3 (n = 5), M1M3 (n = 3), and M3M3 (n = 1). In addition, one family was evaluated to determine the inheritance pattern of the M1(Ala^{213}) gene. The α1AT serum phenotypes of the family members included M1M1 (n = 3), and M1M2 (n = 2).

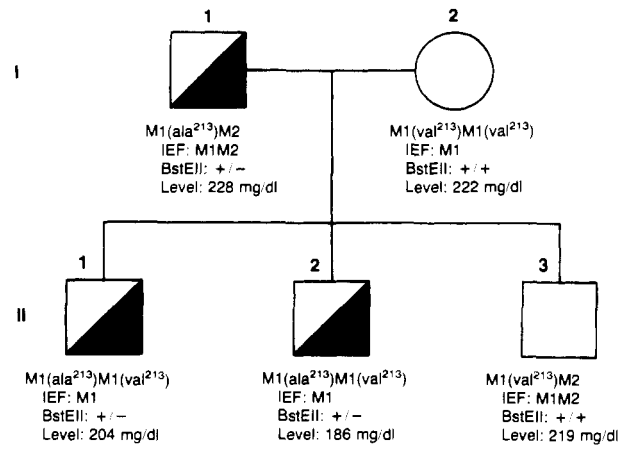

FIGURE 1: Family pedigree demonstrating the codominant inheritance of the α1AT M1(Ala^{213}) gene. The phenotype of each individual was determined with a combination of *Bst*EII polymorphism, isoelectrofocusing on polyacrylamide gel at pH 4–5, and the serum α1AT level. The two generations are indicated (I and II), as are the individual members of each generation (1–n). Below each individual (□, males; ○, females) are shown the serum IEF α1AT phenotype, *Bst*EII genotype ("+" indicates the presence of *Bst*EII site and "−" indicates its absence), and α1AT level (milligrams per deciliter). The presence of the α1AT M1(Ala^{213}) gene is indicated by the shading.

Cloning of M1(Ala^{213}) Gene and M1(Val^{213}) α1AT Genes. The M1(Ala^{213}) and M1(Val^{213}) α1AT genes were cloned from a genomic DNA library of an individual heterozygous for the two genes. This individual (family member II_1; see Figure 1 for family tree) had a serum α1AT isoelectric focusing (IEF) pattern of M1, a serum α1AT level of 204 mg/dL, and a *Bst*EII genomic DNA pattern of "+/−" (see below). Briefly, 100 μg of genomic DNA was partially digested with 8 units of the restriction endonuclease *Sau*3AI as recommended by the supplier (Bethesda Research Laboratories) in a 1-mL reaction mixture, and aliquots were removed at 3.5, 5, and 6.5 min. The aliquots were combined and subjected to 4–20% sodium chloride density gradient centrifugation (Beckman SW28; $5 \times 10^4 g$, 15 h). Fractions (0.5 mL) were collected from the bottom of the tube, and aliquots were electrophoresed (20 V, 16 h) on a 0.7% agarose gel. Fractions corresponding to the DNA size of larger than 15 kb were combined and ethanol precipitated; 1 μg of DNA was ligated with 2 μg of EMBL3 arms prepared by *Eco*RI and *Bam*HI serial digestion (Promega Biotech), and the ligated concatamers were packaged in vitro (Gigapack, Stratagene). With 5×10^5 λ phage plaque-forming units as a starting material, screening for full-length α1AT genes was carried out by using a 1.9-kb *Stu*I fragment encompassing the exon I area of the α1AT gene from α1AT clone λATMa14.4 (Nukiwa et al., 1986a) and a 1.1-kb *Pst*I fragment encompassing exon V of an α1AT Z gene clone (Nukiwa et al., 1986b). Three clones encompassing the entire α1AT gene were identified and evaluated by using the restriction endonuclease *Bst*EII (see below). Two of the clones demonstrated 0.95-kb *Bst*EII fragments (indicating amino acid 213 as alanine); one of these clones [λATM1(Ala^{213}), 15.5 kb in length], was chosen for further analysis. The third clone demonstrated 0.72- and

0.23-kb *Bst*EII fragments (indicating amino acid 213 as valine); this 16.0-kb clone was designated λATM1(Val^{213}).

Sequencing of the M1(Ala^{213}) and M1(Val^{213}) α1AT Genes. The promotor region, all exons, and all exon–intron junctions of λATM1(Ala^{213}) and λATM1(Val^{213}) were directly sequenced as double-stranded DNA of four pUC18 subclones by using synthetic bidirectional oligonucleotide primers. As the start codon (ATG) of α1AT gene is in exon II and the stop codon (TAA) in exon V (Long et al., 1984), the strategy for the sequencing of the protein coding region was designed as follows: A 10-kb *Eco*RI fragment encompassing α1AT exons II–V was isolated from λATM1(Ala^{213}) and λATM1(Val^{213}) and then digested with *Pst*I. Three fragments, including a 1.6-kb fragment encompassing exon II, a 2.4-kb fragment encompassing exons III and IV, and a 1.1-kb fragment encompassing exon V, were subcloned into pUC18. In addition, a 0.5-kb *Pst*I fragment encompassing the promoter region and exon I was subcloned into pUC18 (see Figure 2 for location of these *Pst*I sites in the α1AT gene). The four recombinant plasmids from λATM1(Ala^{213}) and λATM1(Val^{213}), respectively, were then subjected to large-scale preparation by two-step purification, including alkaline lysis and RNase A treatment (Chen & Seeburg, 1985) followed by CsCl equilibrium banding; the two steps were used to prevent RNA contamination and bacterial genomic DNA contamination. The double-stranded DNA plasmids were directly sequenced by the dideoxynucleotide chain termination method using bidirectional primers (Sanger et al., 1977; Nukiwa et al., 1986b).

Determination of Serum α1AT Phenotypes. α1AT phenotypes were determined by a combination of isoelectric focusing, serum α1AT levels, and family studies (Cox et al., 1980). The isoelectric focusing of the serum was carried out to achieve maximum separation of α1AT M subtypes according to a minor modification of the method of Constans et al. (1980) and Nukiwa et al. (1986b). In addition, isoelectric focusing was also performed with an immobilized pH gradient as described by Görg et al. (1985). Briefly, the pH gradient polyacrylamide gel (0.5 mm) was made up with two polyacrylamide solutions of pH 4.50 and 5.00 with immobilin of p*K* 4.6 and 9.3 in 200 mm × 260 mm glass plates. The gel was then polymerized at 50 °C for 1 h, washed in 20% glycerol for 1 h, and dried to the original weight. The gel was placed on the bed of LKB 2117 multiphor II cooled at 10 °C. Serum samples (20 μL) containing 10% dithiothreitol (DTT) were applied to the gel, which was run at 300 V, 15 mA, for 1 h and then reset to 5000 V, 15 mA, and run for 18 h.

α1AT serum levels were performed by radial immunodiffusion (Calbiochem-Behring Corp.). We, together with others (Jeppsson et al., 1978; Pannell et al., 1974), recognize that the quantification of amounts of α1AT is complicated by the fact that the commercially available standard (Calbiochem-Behring Corp.) commonly used for clinical studies yields values for amounts of α1AT that are higher than the true values. In this context, values for the α1AT concentration in the text and figures presented as milligrams per deciliter are based on the commercial standard and those given as micromolar are based on a true laboratory standard (Straus et al., 1985); multiplying the commercial standard values by 0.71 corrects them to the true values.

Determination of the True M1 Phenotype Using a BstEII Restriction Fragment Length Polymorphism. The existence of the M1(Ala^{213}) form of α1AT was discovered serendipitously during an evaluation of M1 and Z genotypes with the restriction endonuclease *Bst*EII and oligonucleotides directed around amino acid 213 (Nukiwa et al., 1986b). To best demonstrate this polymorphism, human genomic DNA, prepared from peripheral white blood cells according to a modification of the method of Jeffreys and Flavell (1977), was digested with endonucleases *Bst*EII and *Pst*I. Although the mutation site is detected by *Bst*EII, double enzyme digestion was performed because restriction map analysis of the α1AT genomic DNA sequence (Long et al., 1984) demonstrated that the expected relevant *Bst*EII fragments would be 7.8 and 7.6 kb with and without the polymorphism, respectively, a difference that is difficult to demonstrate with the usual agarose gel electrophoresis. A second enzyme (*Pst*I) was used in conjunction with *Bst*EII so that the fragment size difference could be clearly visualized (see Figure 3C for a schematic of the strategy used and for the location of the *Bst*EII and *Pst*I sites). To further aid in the visualization of the polymorphism, a specific M1(Ala^{213}) gene genomic probe of 0.95 kb (*Pst*I–*Bst*EII) encompassing the entirety of exon III (see schematic, Figure 3C) was used rather than the full-length cDNA probe.

Briefly, the analysis was performed with 20 μg of human genomic DNA digested (37 °C, 16 h) with *Pst*I (Bethesda Research Laboratories) followed by digestion (60 °C, 3 h) with *Bst*EII (New England Biolab) according to the manufacturer's specification. After digestion, the DNA was extracted with phenol–chloroform, ethanol precipitated, and resuspended in 10 mM tris(hydroxymethyl)aminomethane hydrochloride (Tris-HCl), pH 7.5, and 1 mM ethylenediaminetetraacetic acid (EDTA) at a concentration of 1 mg/mL. The resulting DNA (5 μg) was electrophoresed (25 V, 16 h) on a 1% agarose gel and analyzed by Southern blotting (Southern, 1975) with a ^{32}P-labeled M1(Ala^{213}) 0.95-kb gene genomic DNA probe from the region of the M1(Ala^{213}) gene as indicated in the schematic in Figure 3C.

Quantification of the Association Rate Constant of M1(Ala^{213}) and M1(Val^{213}) for Neutrophil Elastase. The association rate constant (K_a) of the α1AT preparations for neutrophil elastase was determined with α1AT purified from plasma of individuals homozygous for M1(Ala^{213}) or M1(Val^{213}) by using the criteria described above. Venous blood was collected in heparinized glass tubes and immediately centrifuged. The α1AT was purified (>95%) from plasma by positive selection affinity chromatography followed by molecular sieving and then negative selection affinity chromatography. The association rate constant (K_a) of these preparations was measured by the method of Beatty et al. (1980) with minor modifications described by Straus et al. (1985).

Comparison of the M1(Ala^{213}) and M1(Val^{213}) Sequences to Known α1AT Sequences. As this is the first study to correlate normal α1AT serum haplotypes with the respective genotype and because M1(Ala^{213}) turned out to be so common (see Results), the sequences of M1(Ala^{213}) and M1(Val^{213}) were compared to all of the published sequences of human α1AT cDNAs (Costanzo et al., 1983; Colau et al., 1984; Rosenberg et al., 1983; Coutelle et al., 1985) and coding exons (II–V) of genomic DNA (Long et al., 1984; Leicht et al., 1982; Nukiwa et al., 1986b). We have also included the sequence of the full-length human α1AT cDNA isolated by Courtney et al. (1984) sequenced by using the dideoxy method and oligonucleotide primers described above. To put the genomic sequences of M1(Ala^{213}) and M1(Val^{213}) in the context of the known α1AT sequences derived from amino acid sequence analysis, the gene sequences were also compared to all of the published data for the α1AT protein sequence. In addition, all of these sequences were compared to the baboon α1AT

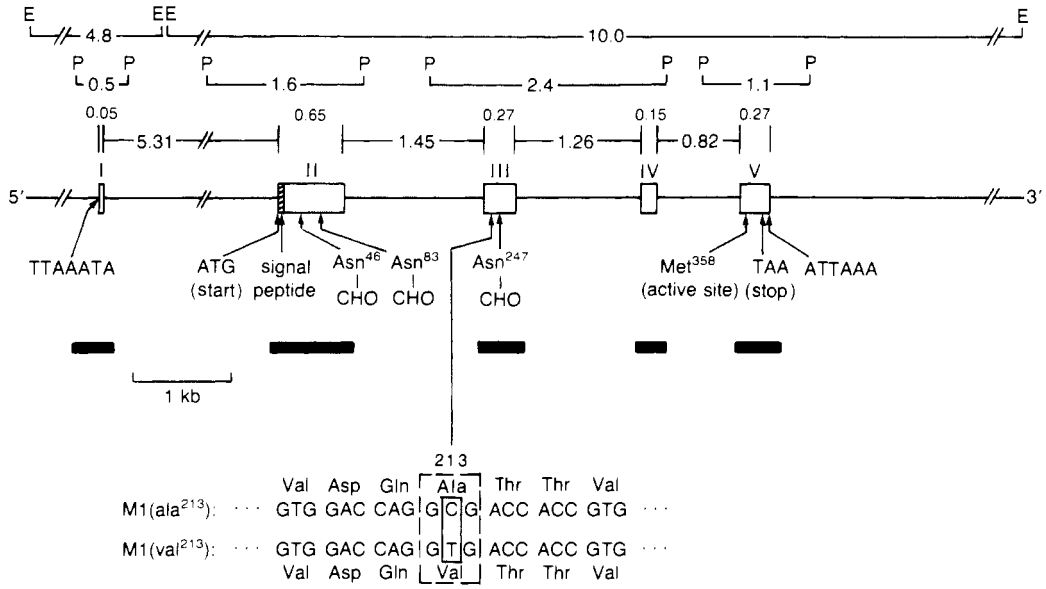

FIGURE 2: Structure of the α1-antitrypsin (α1AT) gene and the sequence difference between the α1AT M1(Ala^{213}) gene and the α1AT M1(Val^{213}) genes. Shown are the five coding exons (open rectangles, I–V) and four introns (connecting lines) of the α1AT gene. Sizes are indicated in kilobases. Also indicated are the putative RNA polymerase binding site (TTAAATA) 5′ to exon I, the start codon (ATG) in exon II, the signal peptide region (striped region in exon II), asparaginyl (Asn) residues in exons II and III that are the carbohydrate (CHO) attachment sites, the inhibitory active site (Met^{358}) in exon V, the stop codon (TAA) in exon V, and the polyadenylation site (ATTAAA) at the 3′ end of exon V. Indicated below the gene structure are the regions sequenced (solid bars); all sequencing was carried out in subcloned fragments prepared from the cloned gene with *Eco*RI (E) and *Pst*I (P). The sequence of M1(Ala^{213}) gene is identical with that of the M1(Val^{213}) gene except for a single base in exon III coding for the amino acid residue 213. The nucleotide T to C mutation in the box causes neutral amino acid change of Val^{213} to Ala^{213}.

cDNA sequence of Kurachi et al. (1981). Comparisons of the DNA and amino acids sequences were carried out by using an IBM-XT computer with DNASIS software (Hitachi America Ltd.).

RESULTS

Sequence of the Protein Coding Region of α1AT Haplotypes M1(Ala^{213}) and M1(Val^{213}). Our prior studies of the Z type α1AT gene led to the observation that some individuals conventionally defined by IEF criteria as having the M1 haplotype had a sequence of six amino acids around residue 213 similar to the Z gene [i.e., the putative M1(Ala^{213}) gene] while others differed at residue 213 where GTG (Val^{213}) substituted for GCG (Ala^{213}) [i.e., the putative M1(Val^{213}) gene] (Nukiwa et al., 1986b). In this study, direct cloning and sequencing of the promoter region, all exons, and all exon–intron junctions of the M1(Ala^{213}) and M1(Val^{213}) genes demonstrated that they indeed did differ at residue 213, but were otherwise identical (Figure 2). Thus, so far as the coding region is concerned, a single mutational event (GCG to GTG) separates the M1(Ala^{213}) from the M1(Val^{213}) gene, and the neutral amino acid change of alanine to valine is the only change in the primary structure of the α1AT types M1(Ala^{213}) and M1(Val^{213}).

Identification of the M1(Ala^{213}) and M1(Val^{213}) Phenotypes Using Endonuclease Mapping of the α1-Antitrypsin Gene. Although conventional IEF of serum can detect the differences among the M1, M2, and M3 α1AT proteins, it cannot detect the neutral amino acid difference of Ala to Val between the M1(Ala^{213}) and M1(Val^{213}) forms of α1AT, even when immobilins are used to achieve maximum separation of the M1 bands (Figure 3A,B). However, the knowledge that the DNA sequence difference between the M1(Ala^{213}) and M1(Val^{213}) genes generates a difference in a *Bst*EII recognition site permits the rapid identification of these two genes at the level of genomic DNA. In this context, double restriction endonuclease digestion (*Bst*EII and *Pst*I) clearly detected fragment-length polymorphism among individuals who had the M1 phenotype by IEF (Figure 3C). In some individuals, when genomic DNA with IEF pattern M1 was digested with *Bst*EII and *Pst*I, only a 0.95-kb fragment was detected (for example, individual "a", lanes 1, 6, 7, and 12) (Figure 3). From the sequence data (Figure 2), together with the IEF and Southern blotting data, it is clear that this individual was homozygous for the M1(Ala^{213}) gene. Other individuals with the IEF pattern M1 showed both 0.95- and 0.72-kb fragments, indicating the heterozygous presence of the M1(Ala^{213}) and M1(Val^{213}) genes (for example, individual "b", lanes 2, 8, and 13). In contrast, in other individuals with the IEF pattern M1, only a 0.72-kb fragment was detected, demonstrating the homozygous presence of the M1(Val^{213}) gene sequence (for example, individual "c"; lanes 3, 9, 14).

In contrast to the *Bst*EII-detected heterogeneity among IEF M1 phenotypes, the other major α1AT M-family haplotypes, M2 and M3, were both *Bst*EII positive; i.e., they demonstrated a 0.72-kb fragment indicating that these haplotypes are identical at amino acid 213 with M1(Val^{213}) (for example, for M2, individual "d", lanes 4, 10, and 15; for M3, individual "e", lanes 5, 11, and 16). Furthermore, the relatively common α1AT S haplotype was also *Bst*EII positive (not shown), consistent with the known sequence data showing that amino acid 213 is valine (Long et al., 1984), while the Z gene was *Bst*EII negative; i.e., amino acid residue 213 is alanine (Nukiwa et al., 1986b). Thus, while the α1AT M1(Ala^{213}) is genotypically unique among major M-family haplotypes, Ala^{213} is observed in some, but not all, common non-M haplotypes.

Transmission of the M1(Ala^{213}) Haplotype. On the basis of IEF patterns, it is thought that α1AT haplotypes are expressed in the autosomal codominant fashion (Gadek & Crystal, 1982; Fagerhol & Cox, 1981). Evaluation of a family of two generations with a combination of analysis of genomic DNA by *Bst*EII mapping, conventional serum isoelectric focusing, and measurement of the serum α1AT level by radial

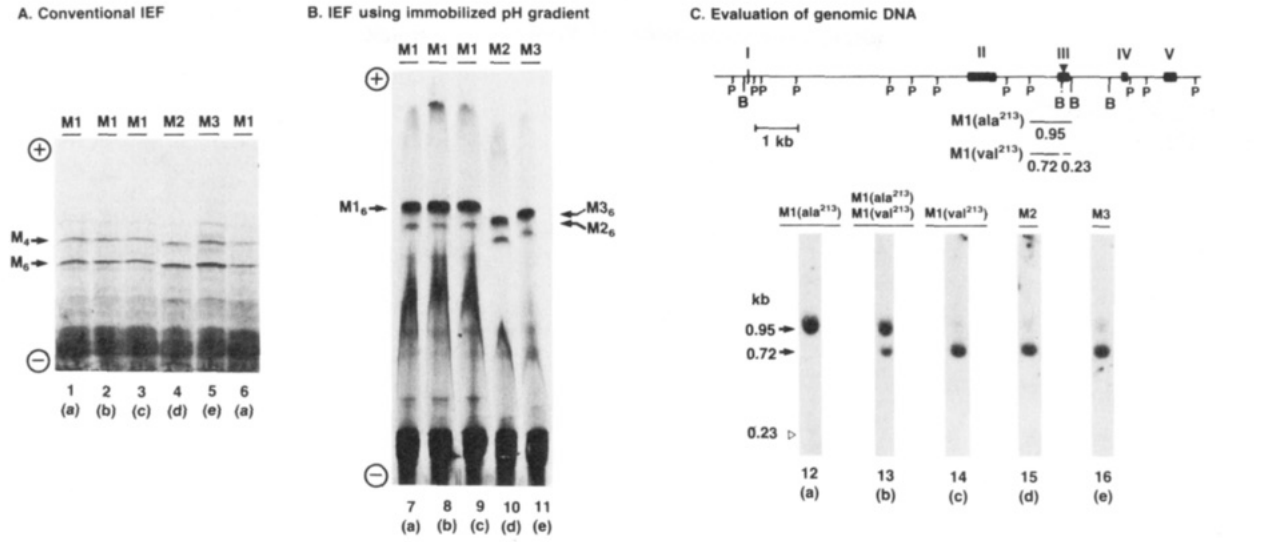

FIGURE 3: Use of the *Bst*EII fragment-length polymorphism to distinguish M1(Ala^{213}) from M1(Val^{213}) among individuals with the same serum M-type isoelectric focusing (IEF) pattern. Serum and genomic DNA samples were evaluated from five different individuals: individual "a" = M1(Ala^{213}) homozygote; individual "b" = M1(Ala^{213})M1(Val^{213}) heterozygote; individual "c" = M1(Val^{213}) homozygote; individual "d" = M2 homozygote; and individual "e" = M3 homozygote. (A) Conventional isoelectric focusing (IEF) gel patterns. The gels were focused at pH 4–5; the anoide (+) is at the top and the cathode (–) at the bottom. Indicated are the major IEF bands for the M-type α1AT haplotypes [referred to as the M_4 and M_6 bands; see Vaughan et al. (1982) and Jeppsson et al. (1985)]. The IEF pattern is at the top of each lane, and the individual evaluated is indicated in brackets below each lane number. Lane 1, individual a, M1(Ala^{213}) homozygote, IEF pattern M1; lane 2, b, M1(Ala^{213})M1(Val^{213}), heterozygote, IEF pattern M1; lane 3, c, M1(Val^{213}) homozygote, IEF pattern M1; lane 4, d, M2 homozygote, IEF pattern M2; lane 5, e, M3 homozygote, IEF pattern M3; lane 6, identical with lane 1. (B) Immobilized pH gradient isoelectric focusing gel patterns. The narrow pH gradient (pH 4.45–4.90) evaluates the M6 bands only, resulting in a wide separation of the M1, M2, and M3 proteins. Lanes 7, 8, 9, 10, and 11 evaluate individuals a–e and are parallel to lanes 1, 2, 3, 4, 5, and 6 of panel A, respectively. Note that even with this magnitude of separation there is no difference among $M1_6$ bands of the M1(Ala^{213}) and M1(Val^{213}) proteins. (C) Identification of M1(Ala^{213}) and M1(Val^{213}) by using the *Bst*EII fragment-length polymorphism. Shown at the top of the panel is a schematic of the α1AT gene with its five exons (solid boxes I–V) and four introns, the site of the exon III (Val^{213} to Ala^{213}) mutation (▼), and the restriction sites for the enzymes *Bst*EII (B) and *Pst*I (P). By use of a 0.95-kb α1AT M1(Ala^{213}) gene exon III probe (see Materials and Methods), the predicted fragments that are generated are 0.95 kb from M1(Ala^{213}) and 0.72 and 0.23 kb from M1(Val^{213}). The sequence recognized by *Bst*EII is G/GTNACC. Below the schematic are shown autoradiograms generated with a ^{32}P-labeled 0.95-kb α1AT M1(Ala^{213}) gene exon III probe by using Southern analysis of the *Bst*EII restriction fragments of genomic DNA of the M-family haplotypes shown in panels A and B. Lane 12, individual a, M1(Ala^{213}) homozygote, IEF pattern M1; lane 13, b, M1(Ala^{213})M1(Val^{213}) heterozygote, IEF pattern M1; lane 14, c, M1(Val^{213}) homozygote, IEF pattern M1; lane 15, d, M2 homozygote, IEF pattern M2 [in the 213 region, the M2 gene has a sequence identical with that of M1(Val^{213})]; lane 16, e, M3 homozygote, IEF pattern M3 [in the 213 region, the M3 gene as a sequence identical with that of M1(Val^{213})]. For the genomic DNA pattern for M1(Val^{213}), the expected 0.23-kb band (indicated by ▷) does not appear because this small fragment does not bind to the nitrocellulose filter efficiently.

immunodiffusion demonstrated that this is also true for the M1(Ala^{213}) haplotype (Figure 1).

Prevalence of the M1(Ala^{213}) and M1(Val^{213}) Haplotypes among Individuals with the IEF Pattern M1. Evaluation of a group of Caucasian individuals with the α1AT serum IEF pattern M1 demonstrated that approximately one-third of the M1 haplotypes were M1(Ala^{213}) with two-thirds M1(Val^{213}) (Table I). In this regard, of the 39 individuals studied, 4 demonstrated a homozygote 0.95-kb *Bst*EII pattern, 17 had a heterozygote 0.95-kb + 0.72-kb pattern, and 18 showed a homozygote 0.72-kb pattern. Assuming these frequencies hold for the Caucasian population as a whole, the haplotype frequency for M1(Ala^{213}) is 32% and for M1(Val^{213}) is 68%. These data suggest that the M1(Ala^{213}) haplotype is at least as frequent as the α1AT M2 haplotype and more frequent than the M3 haplotype (Dykes et al., 1984; Kueppers & Christopherson, 1978).

Consequences of Inheritance of the M1(Ala^{213}) or the M1(Val^{213}) Haplotype. Evaluation of the serum α1AT levels of individuals classified as M1(Ala^{213}) homozygotes, M1(Ala^{213})M1(Val^{213}) heterozygotes, and M1(Val^{213}) homozygotes revealed they were similar ($p > 0.1$; Figure 4A). Futhermore, both the M1(Ala^{213}) and M1(Val^{213}) α1AT molecules were equipotent as inhibitors of neutrophil elastase (Figure 4B). In this regard, the average association rate constant (K_a) of individuals homozygous for α1AT M1(Ala^{213}) was $(9.1 \pm 0.6) \times 10^6$ M^{-1} s^{-1}, comparable with that of individuals with α1AT M1(Val^{213}) [$(9.3 \pm 0.8) \times 10^6$ M^{-1} s^{-1}; $p > 0.5$). Thus, despite the fact that there is a single amino acid difference between the two proteins, there was no significant difference in the function of these common normal forms of α1AT.

Table I: Haplotype Frequency of α1AT Types M1(Ala^{213}) and M1(Val^{213}) among Individuals with Serum α1AT Isoelectric Focusing Pattern M1

phenotypes[a]	*n*	haplotype	*n*	haplotype frequency
M1(Ala^{213})M1(Ala^{213})	4	M1(Val^{213})	53	0.68
M1(Ala^{213})M1(Val^{213})	17	M1(Ala^{213})	25	0.32
M1(Val^{213})M1(Val^{213})	18			
total	39		78	

[a] Phenotypes were determined with a combination of *Bst*EII fragment analysis of genomic DNA, isoelectric focusing of serum, and serum α1AT level (see Materials and Methods).

Codon and Amino Acid Differences among Known α1AT Genes and Proteins. Among the human α1AT gene sequences that are known, only three [M1(Ala^{213}), M1(Val^{213}), and Z] were derived from DNA from individuals in whom the α1AT phenotypes were known (Table II). In all other cases, the α1AT protein phenotype of the source of the DNA was unknown as was the clinical status (in reference to α1AT-related disorders) of the donor. However, by comparing these sequences to that in this study, it is apparent that the cDNA

Table II: Known α1-Antitrypsin Sequences

Material sequenced	Region sequenced	Serum phenotype	exon II[a] –24	101	129	192	exon III 192	213	219	255	264	282	exon IV 282	302	331	exon V 332	342	358	363	376	394	Correct haplotype	Ref
Genomic DNA	All coding regions	M1		Arg	Lys			Ala	Pro	Pro	Glu			Val			Glu	Met	Glu	Glu		M1(Ala^{213})	c
				CGT	AAG			GCG	CCT	CCT	GAA			GTC			GAG	ATG	GAG	GAA			
Genomic DNA	All coding exons	M1		Arg	Lys			Val	Pro	Pro	Glu			Val			Glu	Met	Glu	Glu		M1(Val^{213})	c
				CGT	AAG			GTG	CCT	CCT	GAA			GTC			GAG	ATG	GAG	GAA			
cDNA	Entire coding region	Unknown		Arg	Lys			Val	Pro	Pro	Glu			Val			Glu	Met	Glu	Glu		M1(Val^{213})	d
				CGT	AAG			GTG	CCT	CCT	GAA			GTC			GAG	ATG	GAG	GAA			
cDNA	Entire coding region	Unknown		Arg	Lys			Val	Pro	Pro	Glu			Val			Glu	Met	Glu	Glu		M1(Val^{213})	e
				CGT	AAG			GTG	CCT	CCT	GAA			GTC			GAG	ATG	GAG	GAA			
cDNA	Entire coding region	Unknown		Arg	Lys			Val	Pro	Pro	Glu			Val			Glu	Met	Glu	Glu		M1(Val^{213})	c,f
				CGT	AAG			GTG	CCT	CCT	GAA			GTC			GAG	ATG	GAG	GAA			
cDNA	Entire coding region	Unknown		His	Lys			Val	Pro	Pro	Glu			Ile			Glu	Met	Glu	Glu		?	g
				CAT	AAG			GTG	CCT	CCT	GAA			ATC			GAG	ATG	GAG	GAA			
cDNA	Coding region amino acids –21 to 43, 172-231, 363-394	Unknown						Val	Pro										Glu	Glu		?	h
								GTG	CCC										GAG	GAA			
cDNA	Coding region amino acids 238-283	Unknown								Pro	Glu											?	i
										CCA	GAA												
cDNA	Coding region amino acids 363-394	Unknown																	Glu	Asp		M2 or M3^{b}	j,k
																			GAG	GAC			
Genomic DNA	Entire 12-kb gene	Unknown		Arg	Lys			Val	Pro	Pro	Val			Val			Glu	Met	Glu	Glu		S	e
				CGT	AAG			GTG	CCT	CCT	GTA			GTC			GAG	ATG	GAG	GAA			
Genomic DNA	All coding exons	M_3Z		Arg	Lys			Ala	Pro	Pro	Glu			Val			Lys	Met	Glu	Glu		Z	l
				CGT	AAA			GCG	CCT	CCT	GAA			GTC			AAG	ATG	GAG	GAA			
Protein (plasma)	Amino acids 1-394	Unknown		Arg	Lys			Val	Pro	Pro	Glu			Val			Glu	Met	Glu	Glu		M1(Val^{213})	m
Protein (pooled plasma)	Amino acids 243-350	Unknown								Pro	Glu			Val			Glu					?	n
Protein (plasma)	Amino acids 353-377	Unknown																Met	Glu	Asp		M2 or M3^{b}	o
Protein (plasma)	Amino acids 260-266	S									Val											S^{b}	p
Protein (plasma)	Amino acids 359-365	$X_{Christchurch}$																	Lys			$X_{Christchurch}$[b]	q
Protein (plasma)	Amino acids 336-343	Z															Lys					Z^{b}	r
Protein (liver)	Amino acids 1-33	Z																				Z^{b}	s
Protein (plasma)	Amino acids 344-358	Pittsburgh																Arg				Pittsburgh^{b}	t

[a]Region sequences indicated by dashes. Categorized by coding exons II, amino acids –24 to 192; exon III, 192-282; exon IV, 285-331; exon V, 332-394. [b]Probable haplotypes although entire sequence unknown. [c]This study. [d]Colan et al., (1984). [e]Long et al., (1984). [f]Courtney et al., (1984). [g]Rosenberg et al., (1984). [h]Leicht et al., (1982). [i]Constanzo et al., (1983). [j]Rogers et al., (1983). [k]Coutelle et al., (1985). [l]Nukiwa et al., (1986b). [m]Carrell et al., (1981, 1982). [n]Shochat et al., (1978). [o]Johnson et al., (1978). [p]Owen et al., (1976). [q]Brennan et al., (1986). [r]Owen et al., (1977). [s]Jeppsson et al., (1985). [t]Owen et al., (1983).

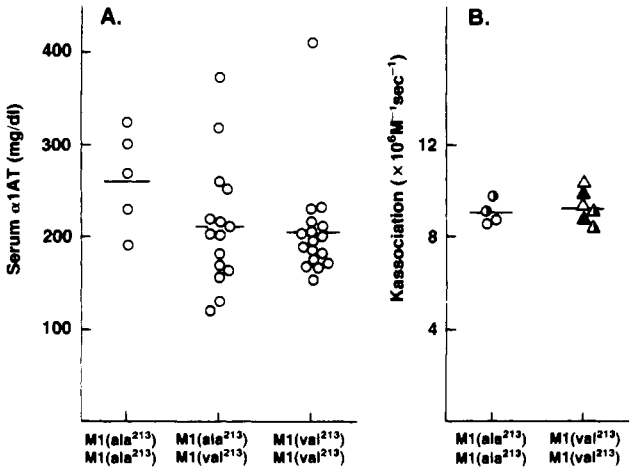

FIGURE 4: Consequences of inheritance of the M1(Ala^{213}) and M1(Val^{213}) α1AT genes. (A) Serum α1AT levels among individuals with the phenotypes M1(Ala^{213})M1(Ala^{213}) (n = 5), M1(Ala^{213})M1(Val^{213}) (n = 15), and M1(Val^{213})M1(Val^{213}) (n = 17). (B) Comparisons of the association rate constant (K_a of the M1(Ala^{213}) and M1(Val^{213}) types of α1AT for neutrophil elastase. The M1(Ala^{213}) and M1(Val^{213}) proteins were purified from plasma of individuals homozygous for these α1AT haplotypes, and the K_a was quantified as described under Materials and Methods. Each symbol (○, ●, △, ▲, ▲) represents a separate individual; each determination was carried out in duplicate. For panels A and B, horizontal lines indicate mean values.

sequence described by Colau et al. (1984) was from an individual with the M1(Val^{213}) haplotype, as were the cDNAs of Long et al. (1984) and Courtney et al. (1984). In contrast, although the cDNA of Rosenberg et al. (1984) includes Val^{213}, it also includes His^{101} (instead of Arg^{101}) and Ile^{302} (instead of Val^{302}) and thus cannot be matched with any known haplotype. The partial sequence data for Leicht et al. (1982) includes Val^{213}, and thus it is possible that it was derived from the M1(Val^{213}) haplotype; however, since the S haplotype also contains identical sequences in this region, a definitive haplotype cannot be assigned. Likewise the partial sequence of Costanzo et al. (1983) fits several haplotypes [e.g., M1(Ala^{213}), M1(Val^{213}), and Z], and thus its identity cannot be finalized. Interestingly, the partial cDNA sequence of Rogers et al. (1983) and Coutelle et al. (1985) contains Asp^{376} (instead of Glu^{376}) and thus clearly does not represent M1(Ala^{213}), M1(Val^{213}), S, or Z. Preliminary sequence data from our laboratory suggests both the M2 and M3 genes have GAC (Asp^{376}), and thus it is possible the Rogers et al. (1983) cDNA was derived from one of these normal haplotypes.

There have been two α1AT gene sequences published that are derived from α1AT haplotypes coding for the so-called "deficient" α1AT proteins, the S sequence described by Long et al. (1984) and the Z sequence described by Nukiwa et al. (1986b). The correlation of the Long et al. (1984) sequence with the S haplotype came when Long et al. (1984) cloned and sequenced one of the α1AT haplotypes from the Lawn et al. (1978) genomic DNA library and recognized that it coded for Val^{264} (instead of Glu^{264}), similar to the known short amino acid sequences of the S protein in this region (Owen et al., 1976). Comparison of the remainder of the Long et al. (1984) S sequence to the M1(Ala^{213}) and M1(Val^{213}) sequences demonstrates the S gene is identical with the M1(Val^{213}) gene except at amino acid 264. In contrast, the Z gene has an alanine at residue 213 and thus is identical with the M1(Ala^{213}) sequence except for the mutation at 342 [M1(Ala^{213}) is Glu^{342}; Z is Lys^{342}].

The only complete sequence of α1AT at the protein level is that of Carrell et al. (1982), carried out in α1AT purified from serum. This sequence is a composite of sequences from many "normals" (phenotype undescribed). Although the serum phenotypes were unknown, this sequence is identical with that of M1(Val^{213}) of this study. This observation is consistent with what would be expected for pooled normal α1AT on a statistical basis since M1(Val^{213}) is the most frequent (by at least 2-fold) α1AT haplotype associated with normal α1AT serum levels. Of the other known α1AT protein sequences, the partial sequence of Shochat et al. (1978) of a normal α1AT is too limited to deduce the likely correct haplotype. The same is true for that of Johnson et al. (1978), but it does contain Asp^{376}, suggesting it might represent M2 or M3. The only other α1AT protein sequences available are the short S sequence of Owen et al. (1976) that identified the S mutation as Val^{264}, the short Z sequence that confirmed the Z sequence differed from the M haplotypes at residue 342 (Owen et al., 1977), and the short N-terminal Z sequence of Jeppsson and Eriksson (1985) that is identical with that of all of the known α1AT haplotypes.

When all of the known human α1AT cDNA, and protein sequence data are compared to the sequence of the α1AT cDNA derived from the baboon (Kurachi et al., 1981), it is apparent that the closest match is that with the M1(Ala^{213}) gene. In this regard, M1(Ala^{213}) and the baboon sequence demonstrated identity at 95.9% (1203 of 1254) nucleotides in the coding exons and identity at 92.8% (388 of 418) amino acids. All of the other known complete human α1AT sequences showed less identity with the baboon sequence than the M1(Ala^{213}) sequence (Table II). Importantly, the amino acid 213 of the baboon is alanine (GCG), i.e., the same codon for amino acid 213 as in M1(Ala^{213}). In contrast, of the known complete α1AT coding exon sequences other than those of the M1(Ala^{213}) and Z gene, amino acid 213 is always valine (GTG). Furthermore, while the Z gene has GCG at this position, the Z gene differs from the baboon at another two nucleotides (amino acids 129 and 342) (Kurachi et al., 1981; Nukiwa et al., 1986b). Together, these observations strongly suggest that of all of the known human α1AT sequences M1(Ala^{213}) is the "oldest" in an evolutionary sense.

Discussion

Among Caucasians, the common, normal α1AT M-family haplotypes include M1, M2, and M3 plus the relatively rare haplotypes M4 and M5 (Gadek & Crystal, 1982; Kueppers, 1978; Fagerhol & Cox, 1981; Constans et al., 1980; Klasen et al., 1982; Weidinger et al., 1985). M1 is the most common (haplotype frequency 68–76%), followed by M2 (14–20%) and M3 (10–12%) (Dykes et al., 1984; Kueppers & Christopherson, 1978). The serum α1AT levels associated with these haplotypes are similar as is the function of the respective molecules as inhibitors of neutrophil elastase (Beckman & Beckman, 1980; Oakeshott et al., 1985).

By use of molecular biologic methods, this study demonstrates that the common M1 haplotype actually includes two haplotypes. Since these haplotypes comigrate on conventional IEF gels as well as on IEF gels with immobilized pH gradient, and differ only at amino acid residue 213, we have named them M1(Val^{213}) and M1(Ala^{213}). Because the difference among these haplotypes only involves a substitution of one neutral amino acid for another, it is not surprising that the differences were not previously detected with IEF analysis of serum (Gadek & Crystal, 1982; Kueppers, 1978; Fagerhol & Cox, 1981; Constans et al., 1980; Klasen et al., 1982; Weidinger et al., 1985). Of these two M1 haplotypes, M1(Val^{213}) is the more common, representing approximately two-thirds of the M1 haplotypes among a population of USA Caucasians. Interestingly, this means that the newly recognized M1(Ala^{213})

haplotype is likely the second most common form of α1AT among USA Caucasians. Consistent with this observation, the "normal" sequence of α1AT described by Carrell et al. (1982) is actually M1(Val^{213}). Since the Carrell et al. (1982) sequence was determined over several years from α1AT purified from pooled serum of normal individuals, it is understandable that, by chance, the M1(Val^{213}) was the sequence observed. Also consistent with the fact that M1(Val^{213}) is the most common M-type haplotype, of all of the human α1AT cDNAs and genomic DNAs that have been randomly isolated and sequenced, for those in which sufficient sequence information is available to definitively assign a haplotype, all are clearly the M1(Val^{213}) haplotype. Interestingly, in their summary of the normal α1AT sequence, Carrell et al. (1982) made note of an unpublished α1AT cDNA sequence that coded for an amino acid sequence identical with the Carrell et al. (1982) normal α1AT protein sequence except for the difference of Ala^{213} instead of Val^{213}, i.e., presumably that unpublished cDNA represented M1(Ala^{213}).

Quantification of the serum levels associated with M1(Ala^{213}) and M1(Val^{213}) demonstrated they were similar and the two molecules functioned in an identical manner as inhibitors of neutrophil elastase. It appears, therefore, that an Ala–Val substitution at residue 213 has no impact on the metabolism or function of the α1AT molecule. These observations are consistent with crystallographic analysis of the α1AT molecule in which amino acid residue 213 is located at the turn of segment 202–223, which forms a strongly twisted, double-stranded antiparallel ladder (Loebermann et al., 1984). While this places residue 213 at the surface of the molecule reasonably close to carbohydrate attachment site Asn^{247}, the IEF isoform patterns (depending primarily on differences in carbohydrate side chains) (Vaughan et al., 1982; Jeppsson et al., 1985) of M1(Val^{213}) and M1(Ala^{213}) are similar, suggesting they likely have similar carbohydrate side chains. Furthermore, the 213 residue in the three-dimensional structure is far from the active site at 358, and the 213 residue is not associated with any critical intramolecular bridging.

Because the difference between M1(Val^{213}) and M1(Ala^{213}) involves a substitution of a neutral amino acid, the two haplotypes cannot be distinguished at the protein level except by sequence analysis. However, the two haplotypes can be easily distinguished at the level of genomic DNA by using two methods. First, by use of Southern blotting, the codons for Val and Ala in the 213 region of exon III can be easily distinguished by using the combined restriction endonucleases *Pst*I and *Bst*EII (Nukiwa et al., 1986b) or the isoschizomer *Mae*III (Hejtmancik et al., 1986). Second, the two haplotypes can be distinguished by using 19-mer oligonucletides centered around the 213 region (Nukiwa et al., 1986b). However, because the Val^{213}/Ala^{213} substitution is also observed in the Z haplotype (Nukiwa et al., 1986b), either method must be combined with serum α1AT levels and IEF analysis to ensure that the 213 substitution is associated with normal levels and is on the M1 background.

With the increasing number of the available gene sequences correlated with the serum phenotype, the haplotypes of α1AT genes can be conceptualized in terms of evolution of the α1AT gene. While the coding sequences of the M1(Ala^{213}) and M1(Val^{213}) only differ at residue 213, some haplotypes share alanine at amino acid 213 [M1(Ala^{213}) and Z], while others share valine at 213 [M1(Val^{213}), M2, M3, and S], indicating that the point mutation at amino acid 213 may be a key point mutation in the evolutionary divergence of this gene. In this context, comparison of all the human α1AT sequences to the baboon α1AT cDNA sequence reveals that the human α1AT gene sharing the closest sequence homology with the baboon gene is M1(Ala^{213}). This fact suggests that it is probable that M1(Ala^{213}) is the oldest common human α1AT gene.

References

Beatty, K., Bieth, J., & Travis, J. (1980) *J. Biol. Chem. 255*, 3931–3934.

Beckman, G., & Beckman, L. (1980) *Hum. Hered. 30*, 81–83.

Bieth, J. G. (1986) in *Regulation of Matrix Accumulation* (Mecham, R., Ed.) pp 217–320, Academic, New York.

Brennan, S. O., & Carrell, R. W. (1986) *Biochim. Biophys. Acta 873*, 13–19.

Carrell, R. W. (1986) *J. Clin. Invest. 78*, 1427–1431.

Carrell, R. W., & Owen, M. C. (1979) *Essays Med. Biochem. 4*, 83–119.

Carrell, R. W., Jeppsson, J.-O., Vaughan, L., Brennan, S. O., Owen, M. C., & Boswell, D. R. (1981) *FEBS Lett. 135*, 301–303.

Carrell, R. W., Jeppsson, J.-O., Laurell, C.-B., Brennan, S. O., Owen, M. C., Vaughan, L., & Boswell, D. R. (1982) *Nature (London) 298*, 329–334.

Chen, E. Y., & Seeburg, P. H. (1985) *DNA 4*, 165–170.

Colau, B., Chuchana, P., & Bollen, A. (1984) *DNA 3*, 327–330.

Constans, J., Viau, M., & Gouaillard, C. (1980) *Hum. Genet. 55*, 119–121.

Costanzo, F., Castagnoli, L., Dente, L., Arcari, P., Smith, M., Costanzo, P., Raugei, G., Izzo, P., Pietropaolo, T. C., Bougueleret, L., Cimino, F., Salvatore, F., & Cortese, R. (1983) *EMBO J. 2*, 57–61.

Courtney, M., Buchwalder, A., Tessier, L.-H., Jaye. M., Benavente, A., Balland, A., Kohli, V., Lathe, R., Tolstoshev, P., & LeCocq, J. (1984) *Proc. Natl. Acad. Sci. U.S.A. 81*, 669–673.

Coutelle, C., Speer, A., Rogers, J., Kalsheker, N., Humphries, S., & Williamson, R. (1985) *Biomed. Biochim. Acta 44*, 421–431.

Cox, D. W., Johnson, A. M., & Fagerhol, M. K. (1980) *Hum. Genet. 53*, 429–433.

Dykes, D. D., Miller, S. A., & Polesky, H. F. (1984) *Hum. Hered. 34*, 308–310.

Fagerhol, M. K., & Cox, D. W. (1981) *Adv. Hum. Genet. 11*, 1–62.

Gadek, J. E., & Crystal, R. G. (1982) in *The Metabolic Basis of Inherited Disease* (Stanbury, J. B., Wyngaarden, J. B., Fredrickson, D. S., Goldstein, J. L., & Brown, M. S., Eds.) pp 1450–1467, McGraw-Hill, New York.

Gadek, J. E., Fells, G. A., Zimmerman, R. L., Rennard, S. I., & Crystal, R. G. (1981) *J. Clin. Invest. 68*, 889–898.

Görg, A., Postel, W., Weser, J., Patutschnick, W., & Cleve, H. (1985) *Am. J. Hum. Genet. 37*, 922–930.

Hejtmancik, J. F., Ward, P. A., Mansfield, T., Sifers, R. N., Harris, S., & Cox, D. W. (1986) *Lancet*, 767–769.

Jeffreys, A. J., & Flavell, R. A. (1977) *Cell (Cambridge, Mass.) 12*, 429–439.

Jeppsson, J.-O., & Eriksson, S. (1985) *Biochim. Biophys. Acta 831*, 30–33.

Jeppsson, J.-O., Laurell, C.-B., & Fagerhol, M. (1978) *Eur. J. Biochem. 83*, 143–153.

Jeppsson, J.-O., Lilja, H., & Johansson, M. (1985) *J. Chromatogr. 327*, 173–177.

Johnson, D., & Travis, J. (1978) *J. Biol. Chem. 253*, 7142–7144.

Klasen, E. C., Bos, A., & Simmelink, H. D. (1982) *Hum. Genet. 62*, 139–141.

Kueppers, F. (1978) in *Genetic Determinants of Pulmonary Disease* (Litwin, S. D., Ed.) pp 23–74, Marcel Dekker, New York.
Kueppers, F., & Christopherson, M. J. (1978) *Am. J. Hum. Genet. 30*, 359–365.
Kurachi, K., Chandra, T., Degen, S. J. F., White, T. T., Marchioro, T. L., Woo, S. L. C., & Davie, E. W. (1981) *Proc. Natl. Acad. Sci. U.S.A. 78*, 6826–6830.
Laurell, C.-B., & Eriksson, S. (1963) *Scand. J. Clin. Lab. Invest. 15*, 132–140.
Lawn, R. M., Fritsch, E. F., Parker, R. C., Blake, G., & Maniatis, T. (1978) *Cell (Cambridge, Mass.) 15*, 1157–1174.
Leicht, M., Long, G. L., Chandra, T., Kurachi, K., Kidd, V. J., Mace, M., Jr., Davie, E. W., & Woo, S. L. C. (1982) *Nature (London) 297*, 655–659.
Loebermann, H., Tokuoka, R., Deisenhofer, J., & Huber, R. (1984) *J. Mol. Biol. 177*, 531–556.
Long, G. L., Chandra, T., Woo, S. L. C., Davie, E. W., & Kurachi, K. (1984) *Biochemistry 23*, 4828–4837.
Mega, T., Lujan, E., & Yoshida, A. (1980) *J. Biol. Chem. 255*, 4053–4056.
Nukiwa, T., Brantly, M., Garver, R., Paul, L., Courtney, M., LeCocq, J.-P., & Crystal, R. G. (1986a) *J. Clin. Invest. 77*, 528–537.
Nukiwa, T., Satoh, K., Brantly, M. L., Ogushi, F., Fells, G. A., Courtney, M., & Crystal, R. G. (1986b) *J. Biol. Chem. 261*, 15989–15994.
Oakeshott, J. G., Muir, A., Clark, P., Martin, N. G., Wilson, S. R., & Whitfield, J. B. (1985) *Ann. Hum. Biol. 12*, 149–160.
Owen, M. C., & Carrell, R. W. (1977) *FEBS Lett. 79*, 245–247.
Owen, M. C., Carrell, R. W., & Brennan, S. O. (1976) *Biochim. Biophys. Acta 453*, 257–261.
Owen, M. C., Brennan, S. O., Lewis, J. H., & Carrell, R. W. (1983) *N. Engl. J. Med. 309*, 694–698.
Pannell, R., Johnson, D., & Travis, J. (1974) *Biochemistry 13*, 5439–5445.
Rogers, J., Kalsheker, N., Wallis, S., Speer, A., Coutelle, C. H., Woods, D., & Humphries, S. E. (1983) *Biochem. Biophys. Res. Commun. 116*, 375–382.
Rosenberg, S., Barr, P. J., Najarian, R. C., & Hallewell, R. A. (1984) *Nature (London) 312*, 77–80.
Sanger, F., Nicklen, S., & Coulson, A. R. (1977) *Proc. Natl. Acad. Sci. U.S.A. 74*, 5463–5467.
Schroeder, W. T., Miller, M. F., Woo, S. L. C., & Saunders, G. F. (1984) *Am. J. Hum. Genet. 37*, 868–872.
Shochat, D., Staples, S., Hargrove, K., Kozel, J. S., & Chan, S. K. (1978) *J. Biol. Chem. 253*, 5630–5634.
Southern, E. M. (1975) *J. Mol. Biol. 98*, 503–517.
Straus, S. D., Fells, G. A., Wewers, M. D., Courtney, M., Tessier, L.-H., Tolstoshev, P., LeCocq, J.-P., & Crystal, R. G. (1985) *Biochem. Biophys. Res. Commun. 130*, 1177–1184.
Travis, J., & Salvesen, G. S. (1983) *Annu. Rev. Biochem. 52*, 655–709.
Vaughan, L., Lorier, M. A., & Carrell, R. W. (1982) *Biochim. Biophys. Acta 701*, 339–345.
Weidinger, S., Jahn, W., Cujnik, F., & Schwarzfischer, F. (1985) *Hum. Genet. 71*, 27–29.

Expression of Human Factor IX and Its Subfragments in *Escherichia coli* and Generation of Antibodies to the Subfragments†

Shu-Wha Lin,‡ John J. Dunn,§ F. William Studier,§ and Darrel W. Stafford*,‡
Department of Biology, The University of North Carolina at Chapel Hill, Chapel Hill, North Carolina 27514, and Biology Department, Brookhaven National Laboratory, Upton, New York 11974
Received January 14, 1987; Revised Manuscript Received March 27, 1987

ABSTRACT: A cDNA clone encoding the entire human blood clotting factor IX (amino acids −3 to 415) has been placed under control of transcription and translation signals from bacteriophage T7 and expressed in *Escherichia coli*. The full-length cDNA and 13 different subfragments (which together cover the entire coding sequence of mature factor IX plus amino acids −40 to −19 of the prepro leader sequence) have each been joined to the coding sequence for the major capsid protein of T7 after the 326th codon and expressed as fusion proteins. All of the fusion proteins were insoluble, which facilitated their purification. A goat polyclonal antiserum against human factor IX reacted to different extents with the different fusion proteins, and rabbit polyclonal antibodies raised against the purified fusion proteins recognize the factor IX molecule, as demonstrated by immunoblotting techniques. Antibodies against at least one of the fusion proteins can also inhibit the biological activity of purified factor IX in a one-stage partial thromboplastin time bioassay. We expect these fusion proteins and the antibodies against them to be useful in studying the structure and function of factor IX.

Factor IX (Christmas factor) is a plasma glycoprotein essential for normal hemostasis. A defect in, or absence of, the factor IX molecule results in hemophilia B, an X-linked hereditary bleeding disorder. Purified human factor IX has a molecular weight of about 57 000 and contains approximately 17% carbohydrate (Fujikawa et al., 1973; Osterud & Flengsrud, 1975; DiScipio et al., 1978). The first 12 NH_2-terminal glutamic acid residues of the protein are posttranslationally modified to γ-carboxyglutamic acids (Gla) in a vitamin K dependent reaction (Fryklund et al., 1976; Olsen

† Supported by NIH Grant HL06350-C to S.-W.L. and D.W.S. and by grants from the Office of Health and Environmental Research of the U.S. Department of Energy to J.J.D. and F.W.S.
* Correspondence should be addressed to this author.
‡ The University of North Carolina at Chapel Hill.
§ Brookhaven National Laboratory.

0006-2960/87/0426-5267$01.50/0