Am. J. Hum. Genet. 50:208–221, 1992

Two Amino Acid Substitutions in Apolipoprotein B Are in Complete Allelic Association with the Antigen Group (x/y) Polymorphism: Evidence for Little Recombination in the 3′ End of the Human Gene

Alison M. Dunning,*,[1] Helmut-H. Renges,*,[1] Chun-Fang Xu,*,[1] Rachel Peacock,*,[1] Robert Brasseur,‡ Gerald Laxer,† Matti J. Tikkanen,§ Réné Bütler,‖ N. Saha,# Anders Hamsten,** Maryvonne Rosseneu,†† Philippa Talmud,*,[1] and Steve E. Humphries*,[1]

*Charing Cross Sunley Research Centre; and †Department of Crystallography, Bickbeck College, London; ‡Laboratory of Macromolecules at Interfaces, Free University of Brussels, Brussels; §Third Department of Medicine, University of Helsinki, Helsinki; ‖Central Laboratory Blood Transfusion Service, Swiss Red Cross, Berne; #Department of Pediatrics, National University of Singapore, Singapore; **King Gustav V Research Institute, Karolinska Institute and Department of Internal Medicine, Karolinska Hospital, Stockholm; and ††Department of Clinical Biochemistry, A.Z. St-Jan, Brugge, Belgium

Summary

We report the identification of an A-to-G base change, in exon 29 of the apolipoprotein B (apo B) gene, that results in the substitution of serine for asparagine at residue 4311 of mature apo B100. In a recent publication, Huang et al. have reported a C-to-T base change in exon 26 that causes the substitution of leucine for proline at residue 2712 of apo B. We have found complete linkage disequilibrium between the alleles at both these sites and an immunochemical polymorphism of LDL designated antigen group (x/y) (Ag(x/y)) in a sample of 118 Finnish individuals. This implies that either one of these substitutions—or both of them combined—could be the molecular basis of the Ag(x/y) antigenic determinants, with the allele encoding $serine_{4311}$ plus $leucine_{2712}$ representing the Ag(x) epitope, and that encoding $asparagine_{4311}$ plus $proline_{2712}$ the Ag(y) epitope. In a sample of 90 healthy Swedish individuals the Leu_{2712}/Ser_{4311} allele is associated both with reduced serum levels of LDL-cholesterol and apo B and with raised levels of HDL. However, these differences are of smaller effect than those associated with the *Xba*I RFLP of the apo B gene in this sample. We have also genotyped 523 individuals from European, Asian, Chinese, and Afro-Caribbean populations and have found complete association between the sites encoding residues 2712 and 4311 in all of these samples, although there are large allele frequency differences between these populations. In addition, there is strong linkage disequilibrium with allelic association between the alleles of these sites and those of the *Xba*I RFLP in all the populations examined. Taken together, these data suggest that, since the divergence of the major ethnic groups, there has been little or no recombination in the 3′ end of the human apo B gene.

Received February 28, 1991; final revision received August 19, 1991.

Address for correspondence and reprints: Alison M. Dunning, Centre for the Genetics of Cardiovascular Disorders, Department of Medicine, The Rayne Institute, University Street, London WC1E 6JJ, United Kingdom.

1. Present address: Centre for Genetics of Cardiovascular Disorders, Department of Medicine, The Rayne Institute, University Street, London WC1E 6JJ, United Kingdom.

0002-9297/92/5001-0023$02.00

Introduction

Apolipoprotein B100 (apo B) is a 550-kD protein secreted by the liver as a constituent of VLDL. During the metabolism of VLDL, other apoproteins and triglycerides are removed, leaving apo B as the sole protein component of LDL. As such, it serves both to maintain the integrity of the particle (Yang et al. 1986, 1989), and, as the ligand for the LDL receptor, it mediates the clearance of LDL-cholesterol from the

plasma. Since elevated plasma levels of LDL-cholesterol are one of the recognized risk factors for the development of coronary artery disease (CAD) (Whayne et al. 1981; Durrington et al. 1986), the identification of the factors that determine plasma LDL-cholesterol levels are of major public health importance. Twin studies and path analysis have demonstrated that genetic variation has a significant impact on both plasma levels of apo B and LDL-cholesterol, with a heritability of .5–.6 (Hamsten et al. 1986*a*; Berg 1987). Family studies using complex segregation analysis have found evidence for the plasma levels of apo B being determined by a major gene with contributions from environmental factors and genes of small or intermediate effect (Hasstedt et al. 1987; Pairitz et al. 1988).

Protein polymorphisms of apo B were first noted as the antigen-group (Ag) system of variants, detected using antisera from multiply transfused patients (Allison and Blumberg 1961). There are five reported pairs of epitopes, each pair representing alleles of the apo B gene; these are Ag(a_1/d), Ag(c/g), Ag(h/i), Ag(t/z), and Ag(x/y) (reviewed in Breguet et al. 1990). Berg et al. (1976), using the combined data from 10 different populations, reported that one of the epitopes of this system, the Ag(x−) (equivalent to Ag(y)), is associated with raised serum cholesterol and triglyceride levels. Since the cloning of the human apo B gene (Blackhart et al. 1986), nucleotide substitutions have been reported as candidates for the molecular bases of all the Ag epitopes (Ma et al. 1987, 1989; Dunning et al. 1988; Wang et al. 1988; Young and Hubl 1989; Xu et al. 1989; Wu et al. 1991).

An RFLP of the apo B gene, detected by the enzyme *Xba*I, has been documented by numerous groups to be associated with differences in serum cholesterol and apo B levels, with the presence of the cutting site (the X+ allele) being associated with raised levels (Berg et al. 1986; Law et al. 1986; Talmud et al. 1987; Aalto-Setälä et al. 1989; Paulweber et al. 1990). In addition, the same allele has been found to be associated with reduced fractional catabolic rate (FCR) of LDL (Demant et al. 1988; Houlston et al. 1988). However, the *Xba*I RFLP cannot be the direct cause of the effects seen, since the base change that creates the variant *Xba*I site is in the third base (wobble position) of codon 2488 and so does not alter the threonine residue at this position (Carlsson et al. 1986). The alleles of this RFLP must be in linkage disequilibrium with another base change that does have a functional effect leading to the reported differences in lipid parameters. The present study set out to identify common base changes, in linkage disequilibrium with the alleles of the *Xba*I site, that alter apo B structure and thus may be the cause of the associations seen.

Subjects and Methods

1. *Subjects*

a. British.—These subjects were two groups of unrelated hyperlipidemic individuals on whom FCR of LDL studies have been performed. FCR results and the apo B genotyping of the 17 individuals from Glasgow and of 22 individuals from Sheffield have been described elsewhere (Demant et al. 1988; Houlston et al. 1988).

b. North Karelia, Finnish.—This group comprised 89 unrelated men and women, with no history of hyperlipidemia, from a rural community in North Karelia, Finland. This population has been very stable for several hundred years, with no influence of Swedish-speaking individuals. They have been described in detail elsewhere (Ehnholm et al. 1982, 1984; Kuusi et al. 1985). Ag phenotyping and apo B RFLP genotyping have also been described elsewhere (Tikkanen et al. 1989; Xu et al. 1989).

c. Other Finnish.—A second group of Ag phenotyped, unrelated Finnish individuals was also examined. Phenotyping and apo B genotyping have been described by Dunning et al. (1988).

d. Swedish.—This group comprised 95 randomly selected, healthy male residents of Stockholm county. Data on lipid, lipoprotein, apoprotein, hemostatic, and metabolic variables have been presented elsewhere (Hamsten et al. 1986*b*, 1987). Individuals with serum cholesterol levels >9.5 mM and/or triglyceride levels >3.0 mM were excluded from the lipid and lipoprotein analyses. Age and body mass index (BMI) accounted, respectively, for 11.5% and 7.6% of the variance in serum HDL-cholesterol and triglyceride levels in this group.

e. Chinese.—This group comprised 83 healthy Chinese males, mean age 36 years, living in Singapore. A detailed description of this sample is presented elsewhere (Saha et al., submitted).

f. Afro-Caribbeans.—This group comprised 47 individuals aged 45–74 years (mean age 55 years), with at least three grandparents of Afro-Caribbean origin, drawn from family-practitioner registers in two Northwest London health centers. This sample has been described in detail elsewhere (Miller 1989).

g. Asians.—This group was a random subset of 152 men drawn from a sample of 1,433 south Asian men aged 40–65 years and living in the west of London who have been investigated for CHD risk factors. The original sample was assembled from four factories and from lists of general practitioners in west London. Of these men, 89% were Pujabi Sikhs, and the rest were evenly divided between Gujarati and Punjabi Hindus (McKeigue et al. 1991).

Apes.—DNA from four unrelated chimpanzees, four unrelated gorillas, and two orangutans was also investigated.

2. *DNA Isolation and Southern Blot Analysis*

These procedures were performed using standard techniques described elsewhere (Talmud et al. 1987).

3. *PCR*

Regions of the apo B gene were amplified by PCR (Saiki et al. 1988). The sequences of the oligonucleotide amplimers used, as well as their arrangements, are shown in figures 1 and 3. Oligonucleotides were obtained from Severn Biotech Ltd., Kidderminster, UK. The PCR reactions were performed in an automated thermal cycler (Cambio, Cambridge) by using *Thermus aquaticus* DNA polymerase (Perkin Elmer Cetus, Norwalk, CT) in the manufacturer's recommended buffer. After both initial denaturing of the DNA at 95°C for 5 min and annealing of the primers (55°C, 3 min), the program used was 50 cycles of 72°C for 3 min, 95°C for 1 min, and 55°C for 1 min. DNA generated by the PCR reactions was purified by Geneclean (Bio101, La Jolla, CA) according to the manufacturer's instructions and was eluted into TE buffer to give a DNA concentration of 50–100 ng/μl.

4. *Chemical Mismatch Cleavage Analysis*

Base differences in target DNA were initially detected by a modification of the Cotton et al. (1988) technique, described by Montandon et al. (1989). The exact procedure followed is described by Dunning et al. (1991).

5. *Direct Sequencing*

PCR was performed, and the products were purified as described above. Approximately 100 ng of purified product was annealed to a 30-fold molar excess of oligonucleotide primer, internal to the PCR primers (Severn Biotech, Kinderminster, UK) and was sequenced as described by King-Underwood et al. (1991).

6. *Blotting and Hybridization to Allele-specific Oligonucleotides (ASOs)*

PCRs were performed as described above but were not purified. Oligonucleotides (Severn Biotech) used as PCR primers, along with ASOs, are listed in figures 1 and 3. ASOs were labeled, at the 5′ end, with T4 polynucleotide kinase (BRL, Paisley, UK) and with [γ-^{32}P] ATP (Amersham, Amersham, UK) to a specific activity of approximately 0.1 μCi/pmol (Maniatis et al. 1982). Initially, slot blotting was performed by applying one-fiftieth of each PCR reaction to nylon filters (Hybond-N; Amersham, Amersham, UK) by using a slot-blotting manifold (Schleicher & Schuell, Dossel, Germany), but later one-tenth of the PCR was run on 1.2% agarose gels, was denatured in 1.5 M NaCl/0.5 M NaOH, and then was double-blotted between two pieces of nylon membrane (Biodyne A, Pall Processing Ltd, Portsmouth, UK), by using the denaturing solution as the transfer buffer. DNA was bound to the filters by baking at 80°C for 2 h. Filters were hybridized for 3 h in 5 × SSPE/0.5% SDS/5 × Denhardt's solution, with one or other of the ASOs, at 33°C and then was washed in 5 × SSPE/0.1% SDS for 10 min at the melting temperature of the ASO. Autoradiography was for 2–16 h.

7. *Computer Estimation of Effects of Amino Acid Substitutions*

The protein secondary structure was modeled using the Predict suite of 10 secondary structure prediction programs (Department of Biophysics, University of Leeds). Secondary structure predicted by this program is approximately 60% reliable. Segments of amino acid sequence extending 250 residues either side of the polymorphic sites were analyzed. Hydrophobicity profiles were generated for the same segments by using the Pc-Prot+ (Protein Analysis) Program (1990 version) and by calculating the mean hydrophobic moment (μ_H) and mean hydrophobicity (H_i), using a seven-residue window, according to the procedure of Eisenberg (1984). μ_H was then plotted against H_i for segments extending 15 amino acid residues either side of the relevant substitution sites (Rosseneu et al. 1990).

8. *Statistical Methods*

The gene-counting method was used for the estimation of gene frequencies. χ^2 Analysis was used for the estimation of Hardy-Weinberg equilibrium. Between-sample differences in apo B genotype distribu-

tion were determined by χ^2 analysis. Association between genotypes at the polymorphic loci was also estimated by χ^2 analysis. We considered statistical significance to be at the .05 level. The strength of associations was estimated by the correlation coefficient Δ (Chakravarti et al. 1984). For each lipid trait, multiple linear regression was used to adjust for age, gender, and BMI. The skewness of triglyceride distribution was adjusted by $\log_e$ transformation. For each adjusted variable, a one-way analysis of variance was performed to test the null hypothesis that phenotypic variation is not associated with genetic variation at the apo B locus. The percentage of variance ($R^2 \times 100$) explained by apo B genotypes was estimated by nonlinear regression.

Results

The chemical cleavage mismatch analysis technique was used to screen amplified DNA at bases 12326–13379 in exon 29 of the apo B gene from nine hyperlipidemic British patients. Seven of these patients (1–7 in fig. 1) are homozygous for the presence of the apo B *Xba*I cutting site and have been shown to have reduced FCR of LDL when compared with a similar group of patients homozygous for the absence of the *Xba*I cutting site (Demant et al. 1988; Houlston et al. 1988; Dunning et al. 1991). Patient 8 has been demonstrated to have LDL with a reduced affinity for a normal LDL receptor in an in vitro cell-binding assay (Frostegård et al. 1990; Dunning et al. 1991), and he is heterozygous for the *Xba*I RFLP. The sequence of the DNA amplified from these eight patients was compared with that from the probe DNA amplified from a subject homozygous for the absence of the *Xba*I site and with increased FCR of LDL as demonstrated in the study by Demant et al. (1988). The aim of the present experiment was to find a base change common to the X+ alleles—but not present in the X− alleles—that may be the cause of the differences, in both LDL-cholesterol levels and FCR of LDL, seen associated with the presence or absence of the *Xba*I cutting site. The exon 29 region examined by chemical cleavage mismatch analysis was covered in two overlapping PCR fragments (fig. 1, *top*).

In the analysis of the second PCR fragment, a novel 230-bp band was seen in the hydroxylamine tracks of all the samples (fig. 1, *bottom*). However, this band was present at reduced intensity in the probe sample and in the DNA from patient 8, suggesting that these two subjects were heterozygous for the base change while the others were homozygous. Direct sequencing was performed on the amplified DNA from these patients, and it was determined that the base change occurring is an A-to-G at nucleotide 13141 (fig. 2). This results in a serine-for-asparagine substitution at residue 4311 of the mature apo B100 protein.

The presence of this A-to-G base change in two of nine subjects suggested that this was a common variant, and thus a sample of unrelated individuals were genotyped to estimate the frequency of this polymorphism. The North Karelia and Finnish samples combined (together totaling 119 unrelated Finnish subjects) were genotyped for this base change by using ASOs (fig. 3, *top*) These subjects had previously been phenotyped for each of the Ag types (Dunning et al. 1988; Tikkanen et al. 1989; Xu et al. 1989). In this sample, the frequency of Asn_{4311} was found to be .79 (table 1), and the distribution of genotypes was in Hardy-Weinberg equilibrium. In addition, when genotype and phenotype data were compared, complete association was found between and $Asn_{4311} \rightarrow$ Ag(x/y) Ser, with the asparagine-encoding allele corresponding to Ag(y) and the serine-encoding allele to Ag(x).

Huang et al. (1990) have reported another polymorphism, in exon 26 of the apo B gene, which generates a substitution of leucine for proline at residue 2712 of the mature protein. Since this amino acid change may have a functional effect on apo B, we examined this polymorphism in the same sample of Finnish individuals (fig. 3, *bottom*). Complete association was also found between this polymorphism and Ag(x/y), with Ag(y) corresponding to Pro_{2712} and Asn_{4311} and Ag(x) with Leu_{2712} and Ser_{4311}. This unexpected finding indicates that either one or both of these loci may form the epitope detected as Ag(x/y).

Since there was complete allelic association between these three polymorphisms in this sample of Finnish individuals, it was decided to investigate this further by determining the genotypes at these sites in samples from other ethnic groups. These included individuals of Swedish, British, Asian, Chinese, and Afro-Caribbean origin (tables 1 and 2). Complete association was found between the alleles encoding these two sites in all of the individuals investigated (523 total). However, the allele frequency of Pro_{2712}/Asn_{4311}/Ag(y) varied from .89 in the Afro-Caribbean sample to .24 in the Chinese (table 1). Our estimations of the allele frequencies at residues 2712 and 4311 are in agreement with those observed for the Ag(x/y) polymorphism in similar ethnic groups (Breguet et al. 1990). Subsequently, Rapacz et al. (1991) have reported that

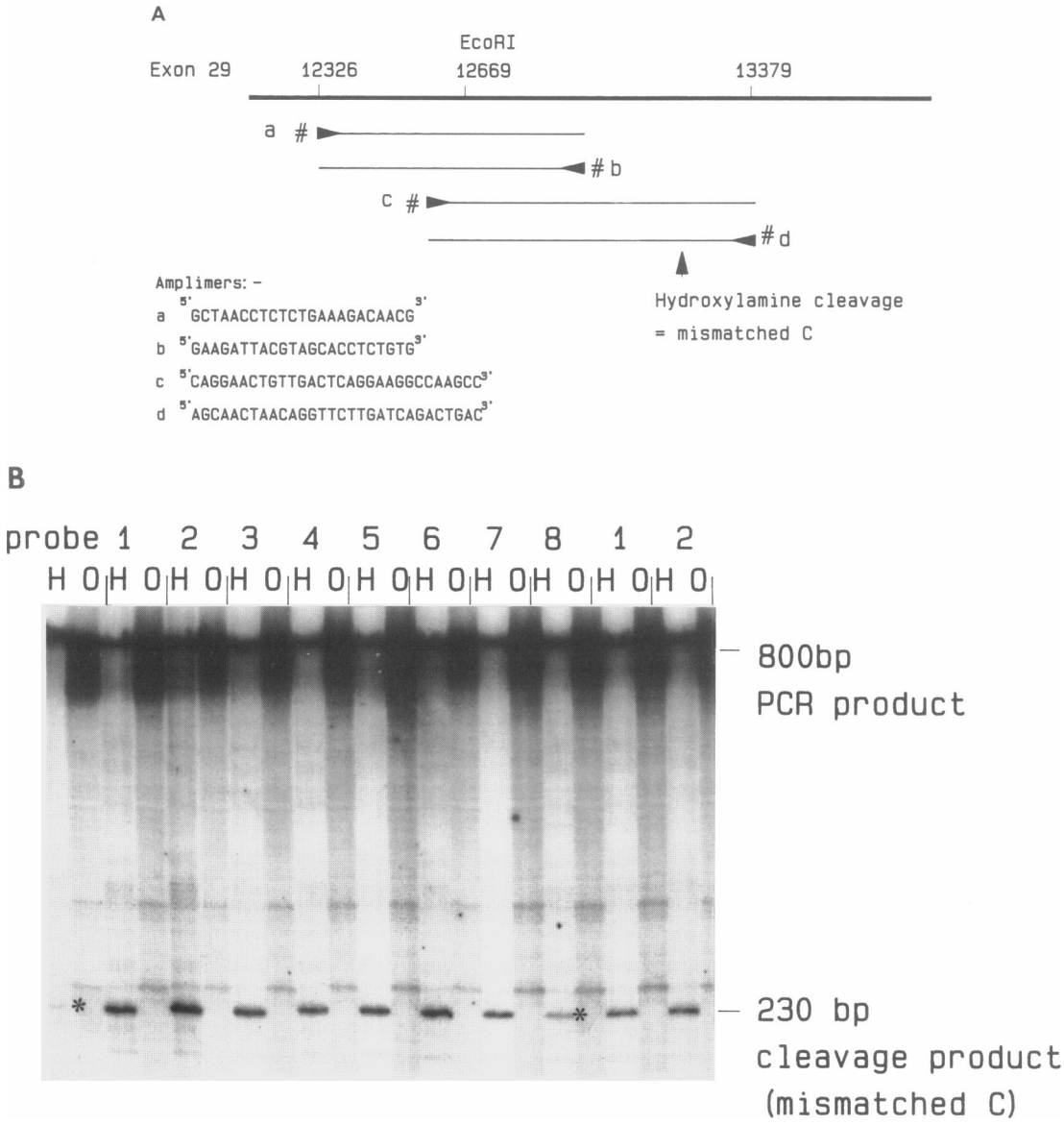

Figure 1 Chemical cleavage mismatch analysis in exon 29 of Apo B Gene. *Top,* Partial map showing gene region covered. Nucleotide numbers are given along the top. Overlapping PCR products used in the analysis are depicted below. Positions of amplimers are represented by arrows, and their sequences are shown as a, b, c, and d. Positions of 5′ end-labeling with ^{32}P are denoted by pound symbols (#). *Bottom,* Autoradiograph showing results of analysis. DNA samples 1–8 and probe (the "probe" DNA sample was hybridized against itself as a negative control) are presented (samples 1 and 2 are shown twice). In each case the hydroxylamine reaction track (H) is on the left, and the osmium tetroxide reaction (O) is on the right. The 800-bp uncleaved PCR product is visible at the top of the photograph, with the 230-bp hydroxylamine cleavage product beneath. The samples yielding reduced intensity bands (i.e., heterozygous for the base change) are marked by asterisks (*).

the Ag(x/y) protein polymorphism was not apparent in 20 chimpanzees and eight gorillas that they investigated, all of which were homozygous for the Ag(y) epitope. Using the identical system that we developed for the human apo B gene, we have genotyped the DNA from four chimpanzees, four gorillas, and two orangutans. Each of these animals was homozygous for proline and asparagine at the apo B residues corre-

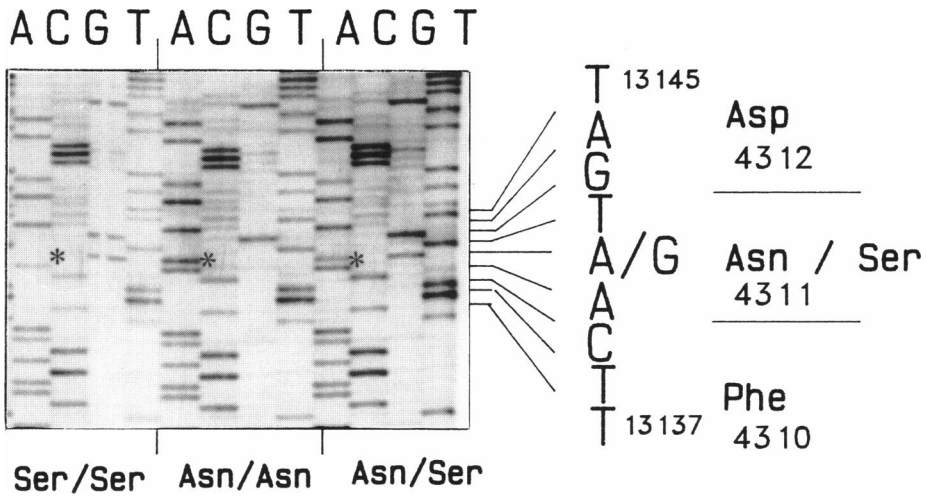

Figure 2 Direct sequencing of DNA amplified across mismatch. Nucleotides 13137–13145 (encoding amino acids 4310–4312) are shown for four DNA samples. The first four tracks show sequenced DNA from an individual homozygous for G at nucleotide 13141 encoding Ser. The second four tracks are from an individual homozygous for A encoding Asn, and the final tracks show DNA from an individual heterozygous for the base difference.

sponding to human residues 2712 and 4311, respectively; this would be compatible with these animals also being homozygous for Ag(y).

In order to try to elucidate what, if any, functional effect the Pro_{2712}→Leu and Asn_{4311}→Ser substitution may have, an attempt was made to determine the secondary structures of the polymorphisms of apo B by using the amino acid sequence data of the relevant portions of the protein. This was done by using both the output of the Predict secondary-structure prediction program and hydrophobicity profiles. The Pro_{2712}→Leu substitution appears to be in a very short helical section within a hydrophobic region of the protein generally comprising β-sheet and β-turn (fig. 4), and it is predicted that the presence of proline at this site would disturb the helical structure but would not significantly affect the hydrophobicity of the region. The Asn_{4311}→Ser change appears to occur in a short joining region between two long amphipathic helices (fig. 4), indicating that it is likely to be exposed on the surface of the LDL particle; however, it would not seem to have a major effect on the secondary structure.

In order to determine whether the polymorphisms at residues 2712 and 4311 are the cause of the differences in lipid and lipoprotein levels seen in association with the *Xba*I RFLP of apo B, we investigated the association between these loci and the *Xba*I site in the Finnish, Swedish, British, Asian, and Chinese samples and found strong linkage disequilibrium ($\Delta = .57$, $P < .001$). Moreover, all the unambiguously determined Ag(x)/Leu_{2712}/Ser_{4311} alleles are associated with the X− allele, although the Ag(y)/Pro_{2712}/Asn_{4311} al-

Table 1

Genotype Distribution and Relative Allele Frequencies in Population Samples

Population (*N*)	Pro_{2712}→Leu			Asn_{4311}→Ser			Frequency of Pro_{2712}/Asn_{4311}
	PP	PL	LL	NN	NS	SS	
Finnish (118)	65	44	9	65	44	9	.74
British (37)	21	13	3	21	13	3	.74
Swedish Controls (186)	58	23	5	58	23	5	.81
Asians (152)	44	70	38	44	70	38	.52
Chinese (82)	7	34	41	7	34	41	.29
Afro-Caribbeans (47)	37	10	0	37	10	0	.89

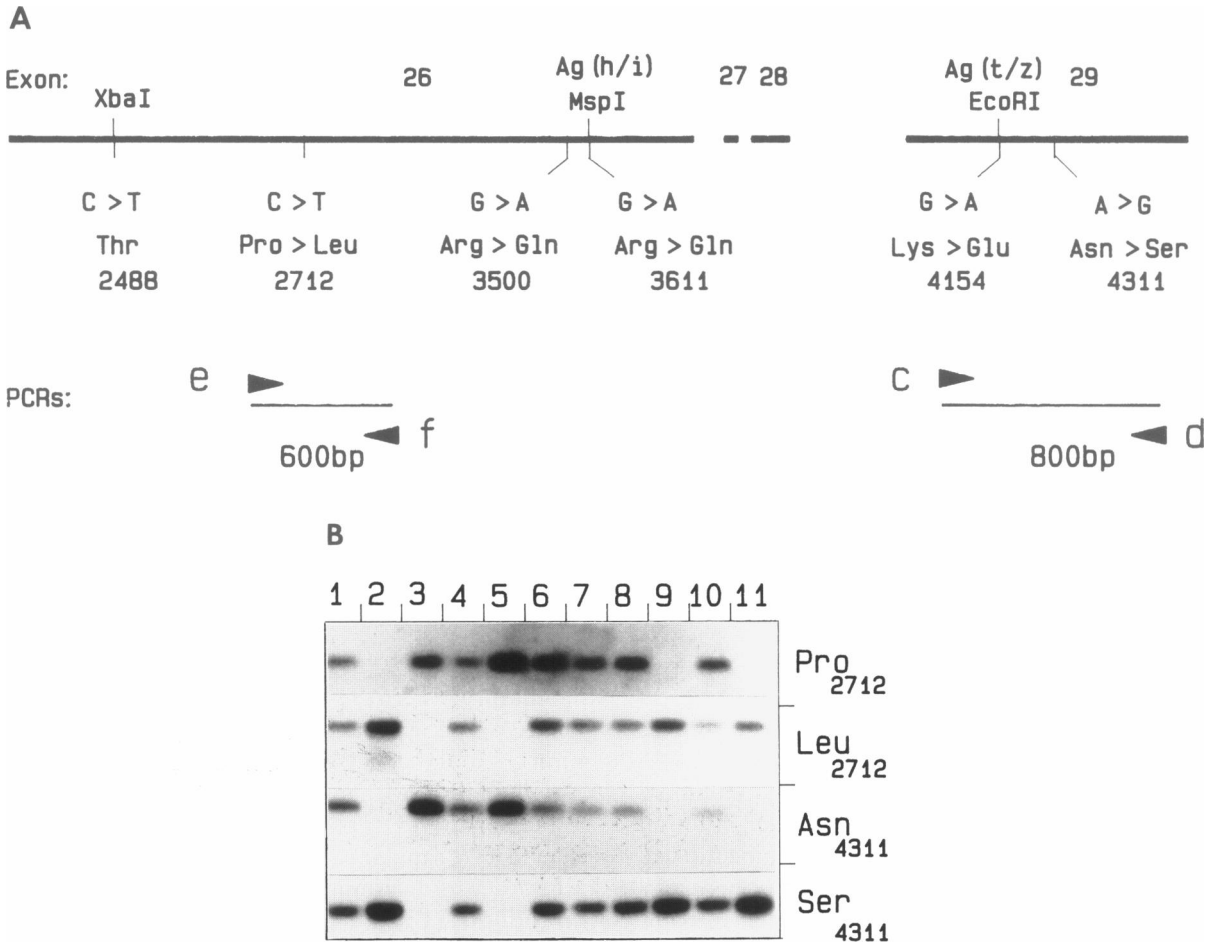

Figure 3 Allele-specific oligonucleotide melting to detect Pro_{2712}→Leu and Asn_{4311}→Ser. *Top,* Map of exons 26–29, showing positions of single base changes and amino acid substitutions that they create. Where the changes also generate RFLPs and antigen group (Ag) polymorphisms, these are shown along the top. The positions of the PCRs used to detect the base changes are shown below. The sequences of amplimers c and d are given in fig. 1a. The sequences, 5′ to 3′, of the other amplimers are as follows: e—ATCATCAGAACCATTGACCAGATGCTGAAAC; and f—TGACAATCACTCCATTACTAAGCTCCAGTG. The sequences, 5′ to 3′, of the ASOs are as follows: Pro—CACATACCAGAATTC; Leu—GAATTCTAGTATGTG; Asn—AATCTTCAATGATTATA; and Ser—TATAATCACTGAAGATT. *Bottom,* Autoradiograph showing example of results obtained. Quadruplicate blots of amplified DNA from 11 individuals are shown. These are hybridized to the ASOs recognizing the alleles for (from top to bottom) Pro_{2712}, Leu_{2712}, Asn_{4311}, Ser_{4311}. Sample 1 is an illustration of DNA from an individual heterozygous at both sites; sample 2 is from one homozygous for Leu_{2712}/Ser_{4311}; and sample 3 is from one homozygous for Pro_{2712}/Asn_{4311}.

leles are found unequivocally on both the X+ and X− alleles (table 3).

The effects that these different alleles have on the lipid, lipoprotein, and apolipoprotein parameters were compared in a sample of healthy Swedish individuals (tables 4 and 5). The *Xba*I RFLP accounts for as much as 9.1% of the phenotypic variance in serum LDL-cholesterol levels ($P < .05$; table 4) and for 8.9% of the variance in serum apo B levels, although this value does not reach statistical significance. In each case, the X− allele is associated with reduced mean levels of both LDL-cholesterol and apo B (table 5). By contrast, the Pro_{2712}→Leu/Asn_{4311}→Ser polymorphisms explain less of the variance seen in LDL-cholesterol and apo B levels—i.e., only 1.1% and 2.8%, respectively (table 4)—although in each case there is also a trend for the Leu_{2712}/Ser_{4311} allele to be associated with reduced mean levels (table 5). The polymorphisms at residues 2712 and 4311 do, however, explain a significant percentage of the variance

Table 2

Complete Association between Pro_{2712}→Leu, Asn_{4311}→Ser, and Ag(x/y)

Sample	Ag(x/y) xx	xy	yy	
Individuals from Finland:				
Pro_{2712}→Leu:				
LL	9			
LP		44		
PP			66	
Total				119
Asn^{4311}→Ser:				
SS	9			
SN		44		
NN			66	
Total				119
	Asn_{4311}→Ser			
	SS	SN	NN	
Combined ethnic groups:				
Pro_{2712}→Leu:				
LL	96			
PL		194		
PP			233	
Total				523

Table 3

Linkage Disequilibrium with Allelic Association between *Xba*I RFLP, Pro_{2712}→Leu, and Asn_{4311}→Ser in 420 Individuals from Different Populations

	*Xba*I		
	X−X−	X−X+	X+X+
(SL)	65	0	0
(SP)	86	76	0
(NP)	26	93	74

NOTE.—Δ = .57; $P < .001$.

(9.0%, $P < .05$; table 4) seen in serum HDL-cholesterol levels. In this instance the Leu_{2712}/Ser_{4311} allele is associated with raised mean HDL-cholesterol levels (table 5).

Discussion

Our findings demonstrate complete allelic association between Ag(x/y) and polymorphism at sites encoding apo B residues 2712 and 4311 in the sample

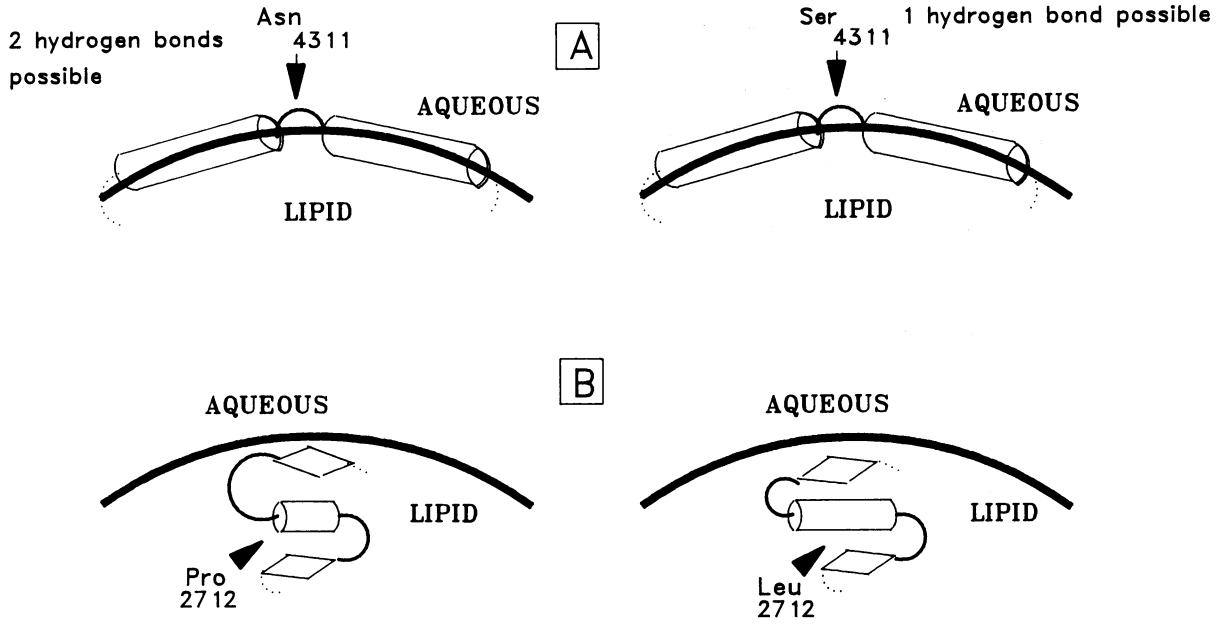

Figure 4 Cartoon representations of models of possible apo B structure in regions of amino acid substitutions. The phospholipid membrane of the LDL particle is represented by the thicker-lined arcs. Potential helical regions are denoted by cylinders, and likely β-sheet/β-turn regions are denoted by parallelograms. (Cartoons are not drawn to scale.) Ⓐ, Residue 4311. This residue probably lies in a short hydrophillic region between two longer amphipathic helices. The Asn→Ser change does not appear to alter the secondary structure; however, Asn could partake in the formation of two hydrogen bonds, and Ser could partake in only one during the formation of any tertiary structure. Ⓑ, Residue 2712. This residue is in a hydrophobic region generally consisting of potential β-sheet and β-turn; however, residue 2712 lies within a short helix which appears to be broken by the presence of Pro at this site but which is sustained by Leu.

Table 4

Percentage of Variance (R^2 × 100) Explained by *Xba*I and Pro_{2712}→Leu/Asn_{4311}→Ser Genotypes in 86 Swedish Individuals

	*Xba*I	Pro_{2712}→Leu and Asn_{4311}→Ser
Total cholesterol	5.4	1.5
Total triglyceride	1.5	1.1
LDL cholesterol	9.1*	1.1
HDL cholesterol	6.3	9.0*
HDL/LDL cholesterol	8.6*	7.8*
Apo B	8.9	2.8
Apo AI	2.2	1.8

* $P < .05$.

from Finland. Thus the epitopes recognized as Ag(x/y) could be generated by a single amino acid substitution at either residue 2712 or residue 4311. Alternatively, if both these residues are involved in the formation of the tertiary structure in the apo B protein, they may interact to form one epitope, with Ag(x) consisting of Leu_{2712} plus Ser_{4311} and Ag(y) of Pro_{2712} plus Asn_{4311}. Since there is little, if any, evidence of recombination in this region of the apo B gene, it is also possible that neither of these residues forms the epitope but that another, as yet undefined amino acid substitution, in complete linkage disequilibrium with these alleles, may create the Ag(x/y) epitopes. Recently, Wu et al. (1991) have published an association study, using 18 individuals protein phenotyped for the Ag polymorphisms, in which they conclude that the Pro_{2712}→Leu substitution is "responsible for the Ag(x/y) polymorphism." While our studies confirm that the alleles encoding residue 2712 are in complete allelic association with those encoding the Ag(x/y) epitopes, our findings demonstrate that such a conclusion may be unwarranted because the correlation seen between residue 2712 and Ag(x/y) is not proof of causation. The same caution must also be applied to the molecular bases of the other Ag polymorphisms, which have also been deduced from association studies. Investigations using in vitro expression of constructs with combinations of the amino acid variants together with antibodies which detect the different Ag epitopes are required to test which apo B residues actually form the Ag epitopes.

The secondary-structure protein modeling presented is inherently speculative but does give some indication as to which of the amino acid substitutions described here could be the molecular basis of the Ag(x/y) antigenic determinants. Pro_{2712} is within one of the hydrophobic "proline-rich" consensus sequences postulated to be able to penetrate the lipid of the LDL particle (DeLoof et al. 1987*b*) and is thus unlikely to be exposed on the surface of the LDL particle. DeLoof et al. (1987*b*) place Pro_{2712} as the ninth residue of a 25-residue consensus motif that occurs eight times within apo B100 (given in one-letter code, this consensus motif is FQVPDLHIPEFQLPHISHTIEVPTF). It is notable that in six of these eight motifs proline is conserved as the ninth residue and that in the other two it is replaced once by asparagine and once also by leucine, which indicates that the Pro→Leu change may be conservative within such a lipid-binding region. The Asn_{4311}→Ser change is also not predicted to have a major effect on apo B conformation, unless this residue is involved in hydrogen bonding, in which case Asn could participate in the formation of two bonds and Ser in only one. Residue 4311 occurs within a short joining region between two long amphipathic helices (fig. 4), and such helices are thought to have

Table 5

Comparison of Mean ± SE Serum Levels of Cholesterol and apo B by Genotype for *Xba*I RFLP and Pro_{2712}→Leu/Asn_{4311}→Ser Polymorphism in 83 Swedish Individuals

	*Xba*I			Pro_{2712}→Leu AND Asn_{4311}→Ser		
	X−X−	X−X+	X+X+	(LS)	(LN)	(PN)
No. of individuals	17	39	27	4	23	56
LDL cholesterol (mmol/liter)	3.57* ± .65	3.88 ± .83	4.25 ± .90	3.77 ± .89	3.85 ± .61	3.98 ± .93
HDL cholesterol (mmol/liter)	1.53 ± .38	1.33 ± .29	1.48 ± .29	1.85* ± .45	1.40 ± .35	1.40 ± .28
Apo B (mmol/liter)	99 ± 13	106 ± 18	111 ± 16	94 ± 14	106 ± 15	107 ± 18

* $P < .05$.

"surface-seeking" properties within the LDL particle (DeLoof et al. 1987*a*). This indicates that residue 4311 may be exposed on the surface of the particle and would therefore be available to act as an epitope for an antibody. Another possibility is that residues 2712 and 4311 are closely related in space on the surface of the lipoprotein particle, because of the folding of the protein. If this is so, then they could jointly form the Ag(x/y) epitopes, Asn_{4311} interacting with Pro_{2712} or Ser_{4311} with Leu_{2712} as the only two functional combinations of the protein. Although it is not yet possible to use computer modeling to explore the possibility that residues 2712 and 4311 may interact in the tertiary structure, such interaction cannot be excluded.

These data demonstrate strong linkage disequilibrium in this region of the apo B gene, with apparent complete allelic association between the Ag(x)/Leu_{2712}/Ser_{4311} and *Xba*I X− alleles in the 688 chromosomes where phase could be determined unambiguously. Berg et al. (1986) also found strong linkage disequilibrium between the Ag(x) and *Xba*I X− alleles in a sample of 75 Norwegian individuals, but they did not find complete allelic association; the reason for this difference remains unclear. Both sets of data indicate that the mutations creating the Ag(x)/Leu_{2712}/Ser_{4311} allele may have arisen on a chromosome carrying the X− allele, and our own data support the theory that they have not since been separated by recombination. Thus, our original aim of searching for common potentially functional variants of apo B, in linkage disequilibrium with the alleles of the *Xba*I RFLP, has been successful, although association studies alone cannot exclude or confirm that the Ag(x)/Leu_{2712}/Ser_{4311} allele is having a functional effect on apo B. The results of our initial analysis of the Swedish sample imply that the Ag(x)/Leu_{2712}/Ser_{4311} allele is associated with some, but not all, of the differences in both serum LDL-cholesterol and apo B levels seen in association with the *Xba*I X− allele. In addition, the same allele is associated with raised HDL-cholesterol levels. This pattern is consistent with the fact that the Leu_{2712}/Ser_{4311} alleles are a subset of the *Xba*I X− alleles. However, our data also predict that there are other, as yet undetected, functional variants of apo B that will account for the rest of the differences seen in association with the *Xba*I RFLP. Further detailed analyses in a much larger sample are required to address this question more fully.

Recently, Rapacz et al. (1991), using protein phenotyping, have reported that there is only one Ag haplotype—Ag(g,d,y,i,t)—in 20 chimpanzees and eight gorillas that they have studied. They thus suggest that this is the ancestral haplotype for human apo B. This also implies that the polymorphism, seen at each of the Ag loci in *Homo sapiens,* has evolved since the divergence of the ancestors of man and higher apes, estimated to be 5 million years ago. We genotyped DNA from four chimpanzees, four gorillas, and two orangutans and found them all to be homozygous for both proline and asparagine at the apo B residues corresponding to human residues 2712 and 4311, respectively. Thus these results support those of Rapacz et al., and it can be predicted that these apes may also have the Ag(y) epitope. However, since the complete association of Pro_{2712}/Asn_{4311} is also present in these apes, it is still not possible to determine whether either, both, or neither of these residues form the Ag(y) epitope.

Since Ag(x/y) and Ag(a1/d) are the only Ag loci that are polymorphic in every ethnic group worldwide, Breguet et al. (1990) suggest that these loci must be the oldest of the Ag variant sites and that they probably evolved before the postulated migration of *H. sapiens* out of Africa, estimated to have occurred approximately 100,000 years ago. If this is the case, then it indicates that there has been no recombination between the sites encoding Ag(x/y) and residues 2712 and 4311, a distance of approximately 7 kB of DNA, in the history of modern man. Analysis of a larger region of the apo B gene lends further credence to this. When our own data are used in conjunction with those of Breguet et al. (1990), it is possible to draw an evolutionary tree for the 3′ end of the apo B gene (fig. 5). The two haplotypes, designed A_0 and B_0 (fig. 5), differ by only the C-to-T base change that creates the *Xba*I polymorphic site. It is not possible to identify which of these chromosomes arose first, since both are present in all major ethnic groups; however, the frequencies of these two forms vary widely between populations. It is possible to deduce that all the other mutations arose on one or other of these chromosomes in a sequential order, and no subsequent recombination is required to explain the observed haplotypes. Haplotype A_1 can be postulated to have been created by a G-to-A base change on A_0, giving rise both to the *Eco*RI RFLP, at the DNA level, and to the Ag(t/z) polymorphism, at the protein level. The G-to-A mutation that generated Arg_{3500}→Gln (Innerarity 1990) (designated A_2 in fig. 5) appears to have arisen on an A_1 chromosome, since all the patients so far identified as having this disease have inherited this haplotype

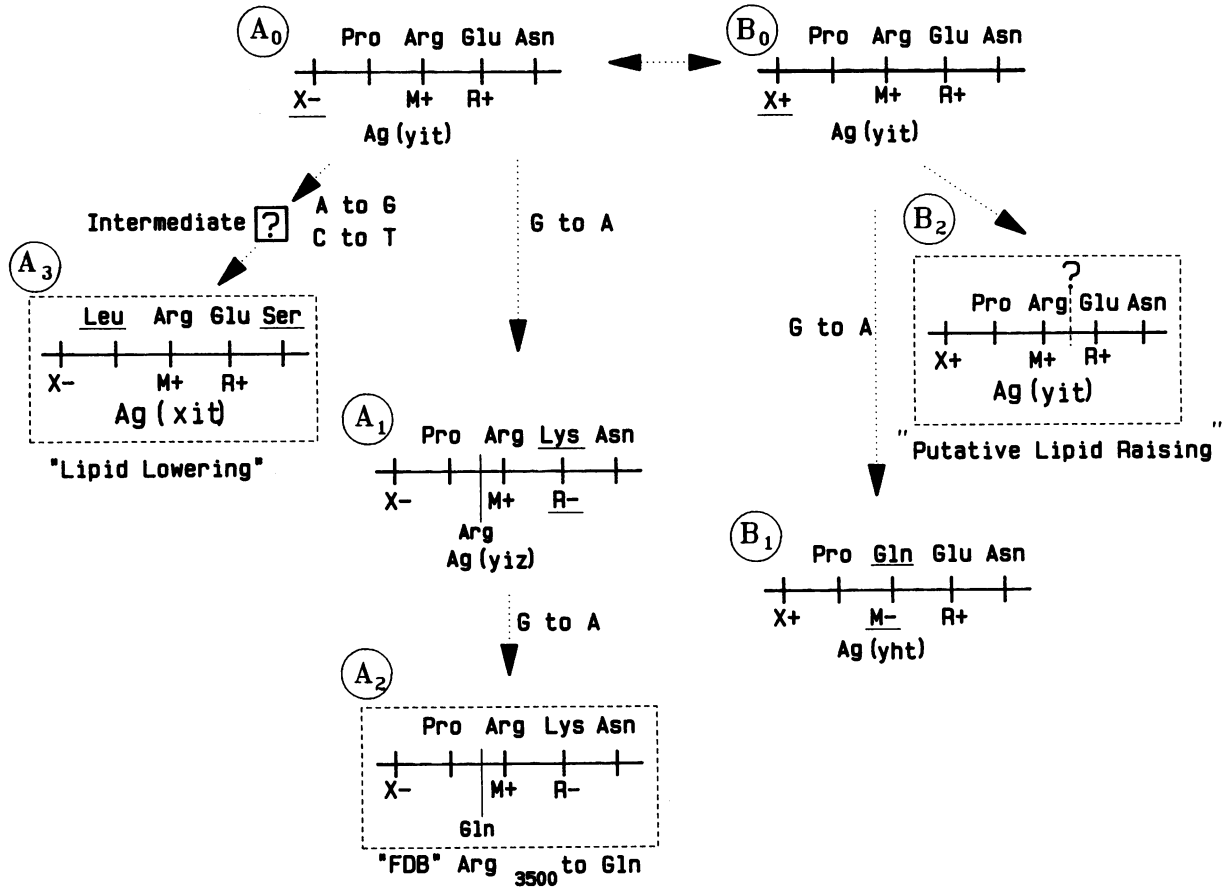

Figure 5 Postulated evolutionary tree for the 3′ end of apo B Gene. Line representations of the apo B haplotypes show amino acids above the line and show both the presence (+) or absence (−) of restriction-enzyme sites (X = *Xba*I; M = *Msp*I; and E = *Eco*RI) and partial Ag haplotypes beneath. Probable progenitor haplotypes (A_0 and B_0) are shown, and all sequential haplotypes have been given a number. Base changes creating the new haplotypes are shown next to the arrows. Full discussion is given in the text.

(Ludwig and McCarthy 1990; Myant et al. 1991). The only exception to the pattern that each haplotype differs from the previous one by a single base change is Ag(xit) (designated A_3 in fig. 5), which differs from A_0 at the two sites: Pro_{2712}→Leu and Asn_{4311}→Ser. Our studies have not revealed an intermediate haplotype (table 2), implying that these two base changes may have arisen within a very short time interval. Our data and those of Berg et al. (1976) suggest that this haplotype has a small lipid-lowering effect. We postulate that the Ag(h/i) polymorphism, the *Msp*I RFLP at the DNA level, has been created by a G-to-A base change on haplotype B_0 (results not shown), thereby giving rise to B_1. We also postulate that there is an additional functionally distinct haplotype (designated B_2) that has been created by an as yet unidentified sequence change occurring on the B_0 haplotype. This sequence change would be in allelic association with the *Xba*I X + site and has the effect of raising serum cholesterol levels (Berg et al. 1986; Law et al. 1986; Talmud et al. 1987; Aalto-Setälä et al. 1989; Paulweber et al. 1990).

There is some evidence—e.g., the extremely rare haplotype Ag(xiz)—that recombination events may have occurred in this gene region. (Breguet et al. 1990; table 4). However, in general, the evidence indicates that novel mutation, rather than recombination, is the driving force for change in the 3′ end of the apo B gene. This unexpected result poses the question of whether this lack of recombination is due to chance or selection. In order for selection to be the cause, it would require that the forms of apo B created by recombina-

tion be dominant and lethal; these are not attributes that have been generally proposed for variants of the apo B gene. However, since introns have been postulated to be "spare" DNA available for mutation and recombination (Gilbert 1985), the presence of huge exons and only two small introns in the 3′ end of the apo B gene may indicate that recombination in the apo B is disadvantageous.

Acknowledgments

This work was supported by the Sunley Research Trust and by grants from the British Heart Foundation (grant RG5), the Sigoid Juselius Foundation, the Academy of Finland, the University of Helsinki, the Swedish Medical Research Council (grants 08691 and 08152), the Swedish Heart-Lung Foundation, the Torsten and Ragnar Söderberg Foundation, and the Singapore National Heart Association. H.-H.R. is supported by a grant from the Deutsche Forschungsgemeinschaft; R.P. is supported by the Medical Research Council; and C.-F.X. is supported by the British Council. The ape DNA samples were a gift from Dr. Ian Gray and Professor Alec Jeffreys. We would also like to thank Drs. Simon Butler and Alan Mills for helpful discussions and Elaine Osman for help with preparation of the manuscript.

References

Aalto-Setälä K, Gylling H, Helve E, Kovanen P, Miettinen TA, Turtola H, Kontula K (1989) Genetic polymorphism of the apolipoprotein B gene locus influences serum LDL cholesterol level in familial hypercholesterolaemia. Hum Genet 82:305–307

Allison AC, Blumberg BS (1961) An isoprecipitation reaction distinguishing human serum protein types. Lancet 1: 634–637

Berg K (1987) Genetics of coronary heart disease and its risk factors. In: Bock G, Collins GM (eds) Molecular approaches to human polygenic disease. Ciba Foundation Symposium, 130, 1st ed. Wiley & Sons, Chichester, pp 14–28

Berg K, Hames C, Dahlen G, Frick MH, Krishan I (1976) Genetic variation in serum low density lipoproteins and lipid levels in man. Proc Natl Acad Sci USA 73:937–940

Berg K, Powell LM, Wallis SC, Pease R, Knott TJ, Scott J (1986) Genetic linkage between the antigenic group (Ag) variation and the apolipoprotein B gene: assignment of the Ag locus. Proc Natl Acad Sci USA 83:7367–7370

Blackhart BD, Ludwig EH, Pieratti VR, Caiati L, Onasch MA, Wallis SC, Powell L, et al (1986) Structure of the human apolipoprotein B gene. J Biol Chem 261:15364–15367

Breguet G, Bütler R, Bütler-Brenner E, Sanchez-Mazas A (1990) A worldwide population study of the Ag-system haplotypes, a genetic polymorphism of human low-density lipoprotein. Am J Hum Genet 46:502–517

Carlsson P, Darnfors C, Olofsson SO, Bjursell G (1986) Analysis of the human apolipoprotein B gene: complete structure of the B-74 region. Gene 49:29–51

Chakravati A, Buetow KH, Antonarakis SE, Waber PG, Boehm CD, Kazazian HH (1984) Nonuniform recombination within the human β-globin gene cluster. Am J Hum Genet 36:1239–1258

Cotton RG, Rodrigues NR, Campbell RD (1988) Reactivity of cytosine and thymine in single-base-pair mismatches with hydroxylamine and osmium tetroxide and its application to the study of mutations. Proc Natl Acad Sci USA 85:4397–4401

De Loof H, Rosseneu M, Brasseur R, Ruysschaert JM (1987*a*) Functional differentiation of amphiphilic helices of the apolipoproteins by hydrophobic moment analysis. Biochim Biophys Acta 911:45–52

De Loof H, Rosseneu M, Yang C-Y, Li W-H, Gotto AM, Chan L (1987*b*) Human apolipoprotein B: analysis of internal repeats and homology with other apolipoproteins. J Lipid Res 28:1455–1465

Demant T, Houlston RS, Caslake MJ, Series JJ, Shepherd J, Packard CJ, Humphries SE (1988) Catabolic rate of low density lipoprotein is influenced by variation in the apolipoprotein B gene. J Clin Invest 82:797–802

Dunning AM, Houlston R, Frostegard J, Revill J, Nilsson J, Hamsten A, Talmud P, et al (1991) Genetic evidence that the putative receptor binding domain of apolipoprotein B (residues 3130 to 3630) is not the only region of the protein involved in interaction with the low density lipoprotein receptor. Biochim Biophys Acta 1096:231–237

Dunning AM, Tikkanen MJ, Enholm C, Butler R, Humphries SE (1988) Relationships between DNA and protein polymorphisms of apolipoprotein B. Hum Genet 78:325–329

Durrington PN, Hunt L, Ishola M, Kane J, Stephens WP (1986) Serum apolipoproteins AI and B and lipoproteins in middle aged men with and without previous myocardial infarction. Br Heart J 56:206–212

Ehnholm C, Huttunen JK, Retinen P, Leine LI, Mutanen M, Kostiainen E, Iacono JM, et al (1984) Effect of a diet low in saturated fatty acids on plasma lipid, lipoprotein and HDL subfraction. Arteriosclerosis 4:265–269

Ehnholm C, Huttunen JK, Rietinen P, Leino M, Mutanen M, Kostiainen E, Pikkarainem J, et al (1982) Effect of a diet on serum lipoproteins in a population with a high risk of coronary disease. N Engl J Med 307:850–855

Eisenberg D (1984). Three dimensional structure of membrane and surface proteins. Annu Rev Biochem 53:595–623

Frostegård J, Hamsten A, Gudland M, Nilsson J (1990) Low density lipoprotein–induced growth of U937 cells—a

novel method to determine the receptor binding of low density lipoprotein. J Lipid Res 31:33–34

Gilbert W (1985) Genes-in-pieces revisited. Science 228: 823–828

Hamsten A, Eféndic S, Walldius G, Szamosi A, de Faire U (1987) Glucose tolerance and insulin response to glucose in nondiabetic young male survivors of myocardial infarction. Am Heart J 113:917–927

Hamsten A, Iselius L, Dahlén G, de Faire U (1986*a*) Genetic and cultural inheritance of serum lipids, low and high density lipoprotein cholesterol and serum apolipoproteins A-I, A-II and B. Atherosclerosis 60:199–208

Hamsten A, Walldius G, Dahlén G, Johansson B, de Faire U (1986*b*) Serum lipoproteins and apolipoproteins in young male survivors of myocardial infarction. Atherosclerosis 59:223–235

Hasstedt SJ, Wu L, Williams RR (1987) Major locus inheritance of apolipoprotein B in Utah pedigrees. Genet Epidemiol 4:67–76

Houlston RS, Turner PR, Revill J, Lewis B, Humphries SE (1988) The fractional catabolic rate of low density lipoprotein in normal individuals is influenced by variation in the apolipoprotein B gene. Atherosclerosis 71:81–85

Huang LS, Gavish D, Breslow JL (1990) Sequence polymorphism in the human apo B gene at position 8344. Nucleic Acids Res 18:5922

Innerarity TL (1990) Familial hypobetalipoproteinemia and familial defective apolipoprotein B100: genetic disorders associated with apolipoprotein B. Curr Opinions Lipidol 1:104–109

King-Underwood L, Gudnason V, Humphries S, Seed M, Patel D, Knight B, Soutar A (1991) Identification of the 664 proline to leucine mutation in the low density lipoprotein receptor in four unrelated patients with familial hypercholesterolaemia in the UK. Clin Genet 40:17–28

Kuusi T, Ehnholm C, Huttunen JK, Kostiainen E, Pietinen P, Leino U, Uusitalo U, et al (1985) Concentration and composition of serum lipoproteins during a low-fat diet at two levels of polyunsaturated fat. J Lipid Res 26:360–367

Law A, Wallis SC, Powell LM, Pease RJ, Brunt H, Priestley LM, Knott TJ, et al (1986) Common DNA polymorphism within coding sequence of apolipoprotein B gene associated with altered lipid levels. Lancet 1:1301–1303

Ludwig EH, McCarthy BJ (1990) Haplotype analysis of the human apolipoprotein B mutation associated with familial defective apolipoprotein B100. Am J Hum Genet 47: 712–720

Ma Y, Schumaker V, Butler R, Sparkes RS (1987) Two DNA restriction fragment length polymorphisms (RFLPs) associated with Ag (t/z) and Ag (c/g) antigenic sites of human apolipoprotein B. Arteriosclerosis 3:301–305

Ma Y, Wang X, Butler R, Schumaker VN (1989) Bsp 12861 restriction fragment length polymorphism detects AG (c/g) locus of human apolipoprotein B in all 17 persons studied. Arteriosclerosis 9:242–246

McKeigue PM, Shah B, Marmot MG (1991) Central obesity and insulin resistance underlie high diabetes prevalence and cardiovascular risk in South Asians. Lancet 337:382–386

Maniatis T, Fritsch EF, Sambrook J (eds) (1982) Molecular cloning: a laboratory manual, 1st ed. Cold Spring Harbor Laboratory, Cold Spring Harbor, NY

Miller GJ (1989) Ethnicity, lipoproteins and haemostatic factors. In: Cruickshank JK, Beevers DG (eds) Ethnic factors in health and disease, 1st ed. Wright, London, pp 280–288

Montandon AJ, Green PM, Giannelli F, Bentley DR (1989) Direct detection of point mutations by mismatch analysis: application to haemophilia B. Nucleic Acids Res 17: 3347–3358

Myant NB, Gallagher JJ, Knight BL, McCarthy SN, Frostegård J, Nilsson J, Hamsten A, et al (1991) Clinical signs of familial hypercholesterolaemia in patients with familial defective apolipoprotein B-100 and normal low-density lipoprotein receptor function. Atherosclerosis Thromb 11:691–703

Pairitz G, Davignon J, Mailloux H, Sing CF (1988) Sources of interindividual variation in the quantitative levels of apolipoprotein B in pedigrees ascertained through a lipid clinic. Am J Hum Genet 43:311–321

Paulweber B, Friedl W, Krempler F, Humphries SE, Sandhofer F (1990) Association of DNA polymorphism at the apolipoprotein B gene locus with coronary artery disease and serum very low density lipoprotein levels. Arteriosclerosis 10:17–24

Rapacz J, Chen L, Butler-Brunner E, Wu M-J, Hasler-Rapacz J, Butler R, Schumaker V (1991) Identification of the ancestral haplotype for apolipoprotein B suggests an African origin of *Homo sapiens sapiens* and traces their subsequent migration to Europe and the Pacific. Proc Natl Acad Sci USA 88:1403–1406

Rosseneu M, De Loof H, De Meutter J, Ruysschaert JM, Vanloo B, Brasseur R (1990) Identification and computer modelling of functional domains in plasma apolipoproteins. In: Brasseur R (ed) Molecular description of biological membrane components by computer aided conformational analysis, vol. 2, 1st ed. CRC, New York, pp 173–190

Saha N, Tay JSH, Humphries SE. Apolipoprotein B gene DNA polymorphisms (XbaI and EcoRI) serum lipids and apolipoprotein B in healthy Chinese (submitted)

Saiki RK, Gelfand DH, Stoffel S, Scharf SJ, Higuchi R, Horn GT, Mullis KB, et al (1988) Primer-directed enzymatic amplification of DNA with a thermostable DNA polymerase. Science 239:487–491

Talmud PJ, Barni N, Kessling AM, Carlsson P, Darnfors C, Bjursell G, Galton D, et al (1987) Apolipoprotein B gene variants are involved in the determination of serum cholesterol levels: a study in normo- and hyperlipidaemic individuals. Atherosclerosis 67:81–89

Tikkanen MJ, Xu C-F, Hamalainen T, Talmud P, Sarna S,

Huttunen JK, Pietinen P, et al (1989) XbaI polymorphism of the apolipoprotein B gene influences plasma lipid response to diet intervention. Clin Genet 37:327–334

Wang X, Schlapser P, Ma Y, Butler R, Elovson J, Schumaker VN (1988) The ag(a1/d) immunogenetic polymorphism coincides with a T to C substitution at nucleotide 1981, creating an AluI restriction site. Arteriosclerosis 8:429–435

Whayne TF, Alaupovic P, Curry MD, Lee ET, Anderson PS, Schechter E (1981) Plasma apolipoprotein B and VLDL-, LDL- and HDL-cholesterol as risk factors in the development of coronary artery disease in male patients examined by angiography. Atherosclerosis 39:411–424

Wu M-J, Butler E, Butler R, Schumaker VN (1991) Identification of the base substitution responsible for the Ag(x/y) polymorphism of apolipoprotein B-100. Arteriosclerosis Thromb 11:379–384

Xu CF, Nanjee N, Tikkanen MJ, Huttunen JK, Pietinen P, Butler R, Angelico F, et al (1989) Apolipoprotein B amino acid 3611 substitution from arginine to glutamine creates the Ag (h/i) epitope: the polymorphism is not associated with differences in serum cholesterol and apolipoprotein B levels. Hum Genet 82:322–326

Yang CY, Chen SH, Gianturco SH, Bradley WA, Sparrow JT, Tanimura M, Li WH, et al (1986) Sequence, structure, receptor-binding domains and internal repeats of human apolipoprotein B-100. Nature 323:738–742

Yang C-Y, Gu Z-W, Weng S-A, Kim TW, Chen S-H, Pownall HJ, Sharp PM, et al (1989) Structure of apolipoprotein B-100 of human low density lipoproteins. Arteriosclerosis 9:96–108

Young SG, Hubl ST (1989) An ApaLI restriction site polymorphism is associated with the MB19 polymorphism in apolipoprotein B. J Lipid Res 30:443–449