The genetics of atherosclerosis

By DAVID J. GALTON, *Medical Professorial Unit, St Bartholomew's Hospital, London EC1*

Coronary atherosclerosis—one of the major health problems of our time—arises from a complex interaction between environmental and genetic factors. Evidence for genetic factors comes from many studies reporting a strong familial aggregation of the disease. A heritability (that is the proportion of phenotypic variance that can be attributed to additive genetic factors) of 0·63 for coronary artery disease of early onset in family studies has been reported (Nora *et al.* 1980). Twin studies have revealed a higher concordance rate of the disease in monozygotic than in dizygotic pairs. In one representative study of coronary artery disease occurring before the age of 60 years, concordance rates were 0·83 and 0·22 for mono- and dizygotic twin pairs respectively, the disease end-point being angina pectoris or myocardial infarction, or both (Berg, 1983).

There is now a mass of clinical, epidemiological and experimental information to implicate altered metabolism of lipoproteins in the aetiology of atherosclerosis. In rare instances a simple monogenic disorder such as a defect in the cell-surface receptor for endocytosis of low-density lipoproteins (LDL) can strongly predispose to the development of premature atherosclerosis (Brown & Goldstein, 1976). However, in the majority of cases it is not possible to identify single genetic determinants and it is likely that several major genes may contribute to the manifestation of the disease. It will be of great interest to map which of the major genes segregate with the trait. This may allow identification from birth of individuals who are predisposed to develop premature atherosclerosis.

The most likely candidates for genetic variants predisposing to atherosclerosis are those coding for the lipid transport proteins; and the next section summarizes recent information on the apolipoprotein gene family.

The apolipoprotein genes

Use of recombinant DNA techniques has identified the chromosomal locations of all the major apolipoprotein genes and permitted a detailed study of their genomic structure. The apolipoprotein genes are dispersed amongst four chromosomes (Fig. 1). There is a cluster of three on the long arm of chromosome 11 (apo A1, C3 and A4). Another cluster occurs on chromosome 19 (apo C1, C2 and E; together with the LDL receptor). Two others are located on chromosome 1 (apo A2) and on the short arm of chromosome 2 (apo B).

There are considerable nucleotide sequence homologies amongst the apolipoprotein genes, and it is possible they arose from a common ancestral gene by duplications or conversion events. Thus their genomic structures reveal the presence of three introns in strikingly similar positions. One is within the 5′ untranslated region of the gene; the second within the leader sequence; and the third intron within the coding part of the gene. These divide the gene into four exons; and the fourth exon is variably expanded by internal repeat units of approximately sixty-six base pairs (Fig. 2). These may code for amino acid sequences that form the amphipathic helix and be responsible for the lipid-binding properties of these apolipoproteins. A possible scheme for the evolution of this multigene family is presented in Fig. 3. The common ancestral gene may have been

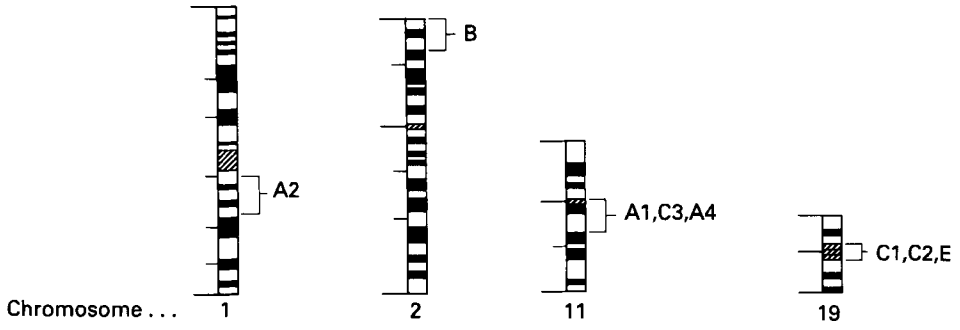

Fig. 1. Chromosomal localization of the apolipoprotein multigene family. A map of human chromosomes 1, 2, 11 and 19 showing approximate assignments of the apolipoprotein genes (see Breslow, 1985; Luo *et al.* 1986). Apolipoproteins A1, A2, C1, etc. are corresponding gene loci.

very similar to the present day apolipoprotein C1 in structure and length. Then by a series of gene duplications the other genes were formed and diversified by variable duplication of stretches of codons within their fourth exons.

Genetic polymorphisms

Mutations are a fundamental way of altering the genome and in some instances can produce a deleterious change in gene function. Such mutations may gradually become eliminated from the gene pool of the population under the influence of natural selection on the resulting phenotype. However, occasionally a mutation will arise that may have some selective advantage over the common allele and may gradually increase in frequency to the point of permanent fixation in the population (Galton, 1985). When the frequency of the uncommon variant occurs to an appreciable extent that cannot be accounted for by new mutations (arbitrarily taken as more than 1%) it is called a polymorphic variant. A genetic polymorphism is therefore the occurrence in the same population of two or more alleles at one locus, each occurring at an appreciable frequency and not accounted for by new mutations. The DNA polymorphism may not

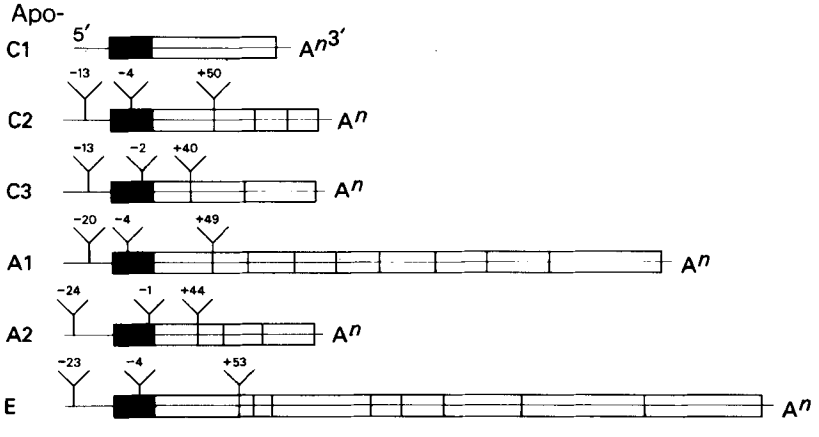

Fig. 2. Genomic structures of the apolipoprotein multigene family. The intron–exon organization of members of the apolipoprotein gene family. Values derived from Breslow (1985) and Luo *et al.* (1986). ■, Leader sequence; □, exons; Y, site of introns that are spliced out; A^n, polyadenine tail; C1, A1, etc. are corresponding gene loci. Numbers above intron sites are base-pairs from start codon.

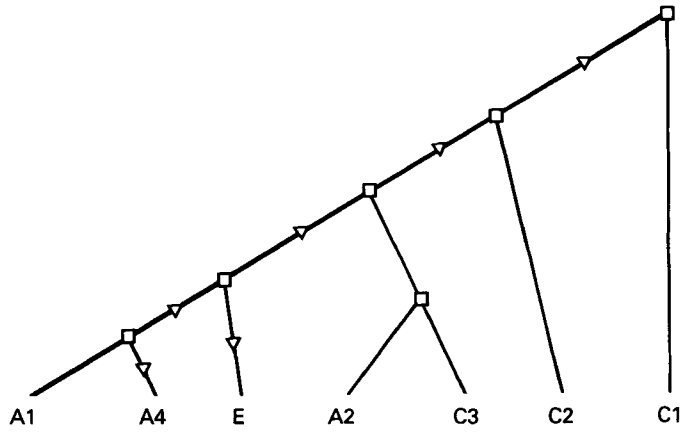

Fig. 3. A hypothetical scheme for the evolution of the apolipoprotein genes. The ancestral gene was probably similar to apo C1. Gene duplication (□) led to apo C2. Duplication of 22 codons in exon 3 and exon 4 (▽) diversified the lengths. Modified from Luo *et al.* (1986).

occur within structural genes. It has been calculated (Jeffreys, 1979) that one in every 100 nucleotides may be variable in flanking regions of the gene or in spacer regions (i.e. intergenic DNA). This makes the extent of DNA polymorphisms much greater than that observed for protein polymorphisms.

Genetic polymorphisms constitute a basic component of hereditary variation in natural populations on which evolutionary forces operate. A polymorphic variant that is of minor selective disadvantage, and may be declining in frequency may become a selective advantage if the prevailing environmental conditions alter slightly. This therefore provides a means whereby populations can evolve in different directions. In general the greater the genetic variation inherent in a natural population the greater the opportunities for evolutionary change.

From a preliminary analysis (human genomic probes have only been available since the mid-1970s) there is clearly a large amount of genetic polymorphism in human populations. Some of this polymorphism may be impermanent representing a transient phase of molecular evolution. The rarer of the variant alleles could be in the process of increasing or decreasing in frequency, under the influence of natural selection or by random genetic drift, or both. Alternatively, the polymorphism may be permanent, i.e. in equilibrium with the common form.

Detection of polymorphisms

DNA polymorphisms can be detected if the mutation directly affects the recognition site of a restriction endonuclease, either creating a new site or abolishing an old one. In either case a different sized DNA fragment will be produced (Fig. 4). More than 100 restriction enzymes are available each with different nucleotide sequence specificities and each enzyme is of potential use for detecting new polymorphisms. This is currently providing a wealth of new genetic markers for clinical studies. Alternatively, a DNA polymorphism can be produced by variable insertions of repetitive sequences within a short stretch of the genome. This leads to DNA fragments produced by digestion with restriction enzymes of highly variable length (Fig. 4).

The most useful method to detect DNA polymorphisms is by 'Southern blotting' (Southern, 1975). Usually leucocyte DNA is digested with various restriction enzymes

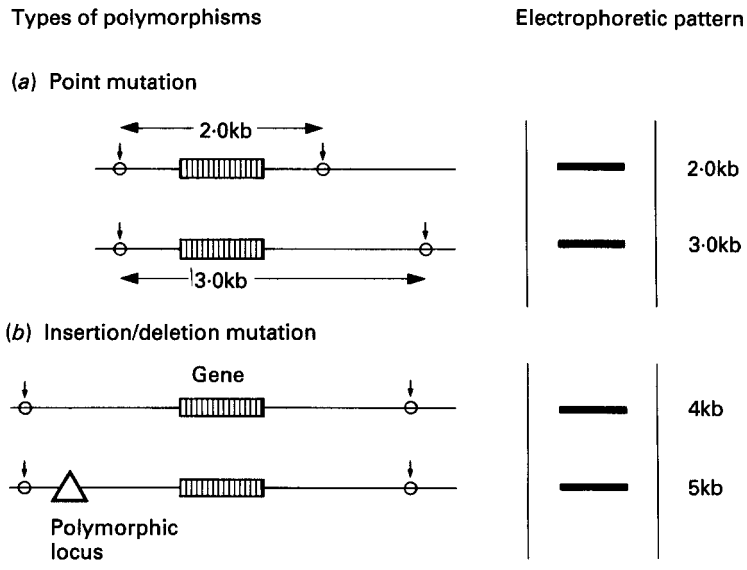

Fig. 4. Types of DNA polymorphisms revealed by 'Southern blotting'. This scheme illustrates a hypothetical gene (▬) digested with the restriction enzyme (○) together with the blotting patterns after gel electrophoresis of the DNA fragments. One type of polymorphism is due to nucleotide substitutions (*a*), the other is due to variable DNA insertions at the polymorphic locus (△) as in (*b*). kb, 1000 base pairs.

and fractionated according to size on agarose or polyacrylamide gel. The DNA on the gel is denatured with alkali, neutralized with Tris buffer and transferred by capillary flow to a nitrocellulose or other filter—the blotting process. The filter now carries a 'print' of the DNA from the gel that can be hybridized to a labelled gene probe and then visualized by autoradiography. This probe can be a cloned DNA sequence or a synthetic oligonucleotide. Using different restriction enzymes this method can generate far more allelic variants than is possible by a study of protein polymorphisms. It is now possible to take a gene which has no protein variants associated with it and in a relatively short time discover several DNA variants at that locus. Such DNA variants may be due to mutations in introns, exons, flanking sequences or spacer DNA. The latter can be in linkage disequilibrium with nucleotide variation within the coding region of structural genes and can be of diagnostic value.

Genetic polymorphisms related to atherosclerosis

An early locus to be studied for population associations with coronary artery disease was the apo A1–C3–A4 gene cluster on the long arm of chromosome 11. This region appears to be relatively prone to new mutations and DNA rearrangements. The apo A1 and A4 genes (that are closely related) have been separated by the interposition of the apo C3 gene in the opposite orientation to both, and intra-strand recombination has been described between the A1 and C3 genes. There are also multiple restriction site polymorphisms along the genome over a distance of approximately 14×10^3 base pairs (Breslow, 1985). These occur in exons, introns and flanking sequences as described in Fig. 5. These sites reveal a complex pattern of linkage associations. Linkage analysis has been done using haplotypes of pairs of allelic sites. The degree of linkage disequilibrium is not a simple function of the relative physical distance between sites over a region of 14×10^3 base pairs. For example, the Sst 1 and Pvu II restriction site polymorphisms are

https://doi.org/10.1079/PNS19870047 Published online by Cambridge University Press

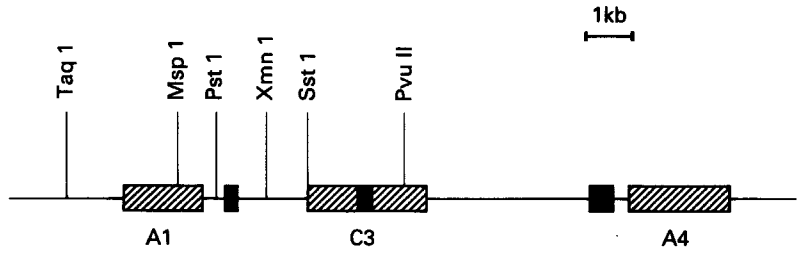

Fig. 5. Restriction map of the A1–C3–A4 genes showing multiple polymorphic sites. ▨, Exons–introns; —, spacer DNA; ■, repetitive elements; kb, 1000 base pairs. Taq 1, Msp 1, Pst 1, Xmn 1, Sst 1 and Pvu II are restriction endonucleases.

close together yet in linkage equilibrium; the Sst 1 and Msp 1 sites are further apart yet show linkage disequilibrium. The simplest explanation is that the processes of mutation and genetic drift are primarily responsible for the observed associations. The relative role of differential selection on haplotypes is difficult to assess. The effects of recombination over such short distances are potentially less significant than mutation and drift/selection in the formation of new haplotypes (Barker *et al.* 1984).

Apo C3 polymorphisms

Exon 4. A DNA polymorphism for the restriction enzyme Sst 1 occurs in the fourth exon, thirty-nine base pairs from the C-terminal codon (Rees *et al.* 1983). It is due to a cytosine to guanine transversion and can be used to identify two alleles, S1 and S2, by Southern blotting. The frequency distribution of this polymorphism has been extensively studied in different ethnic populations as well as in disease groups (Rees *et al.* 1985*a*). In a study of coronary artery disease defined by coronary angiography (Rees *et al.* 1985*b*) the frequency of the rare allele (S2) rose from 6% (*n* 68) in the controls to 22% (*n* 61) in severely affected subjects (Fig. 6). A similar study has been performed with young survivors of myocardial infarction (Ferns *et al.* 1985). The variant (S2) allele was found in 4% (*n* 47) of healthy controls compared with 21% (*n* 48) of infarction subjects (Fig. 6).

Intron 1. There is a Pvu II site polymorphism in the first intron of the apo C3 gene giving rise to two alleles of respective frequencies 0·79 and 0·21 in a West German population (Frossard *et al.* 1987). The frequency of the rare allele is increased in subjects (*n* 155) with coronary artery disease as defined by angiography compared with healthy controls with a relative incidence of 3·5.

The apo A1 gene

This gene is situated $2{\cdot}6 \times 10^3$ base pairs upstream from the apo C3 gene and is transcribed in the opposite orientation. There is a Pst 1 site polymorphism 314 base-pairs from the 3′ end of the apo A1 gene. The frequency of the rare (P2) allele in randomly selected controls was 4% (*n* 123) and was 3·3% in thirty subjects with no evidence of coronary disease by angiography. By contrast, in eighty-eight patients with coronary disease before the age of 60 years (defined by angiography), the P2 allele occurred in 31·8% ($P<0{\cdot}0001$) of subjects (Ordovas *et al.* 1986). This at present is the strongest genetic marker reported for coronary artery disease.

The next stage in these studies is to examine haplotype combinations of polymorphic restriction sites to see if they constitute stronger genetic markers for atherosclerosis

https://doi.org/10.1079/PNS19870047 Published online by Cambridge University Press

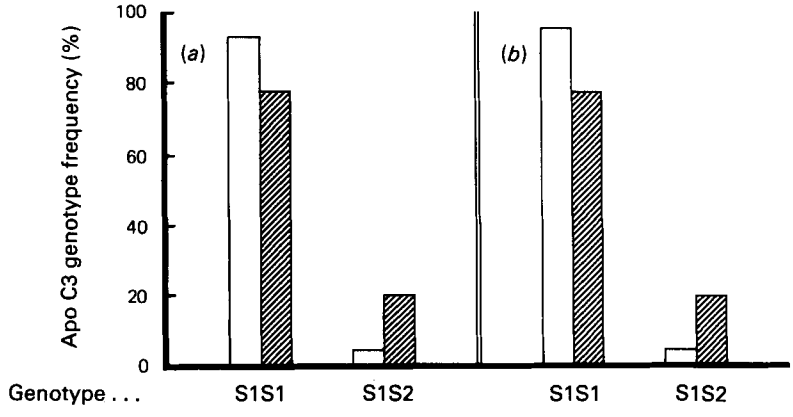

Fig. 6. Distribution of a Sst 1 polymorphism in the apolipoprotein C3 gene in early-onset coronary artery disease. Genotype frequencies are presented in controls (□), and patient groups (▨) with either severe or minimal coronary artery disease defined by (*a*) angiography (see Rees *et al.* 1985*b*) or (*b*) in survivors of myocardial infarction (see Ferns *et al.* 1985). (For details, see p.341.)

(Rees *et al.* 1986). The A1–C3–A4 cluster with multiple restriction site polymorphisms can provide valuable genetic markers since polymorphic sites that are physically close together can show a number of haplotype combinations. These can act as multi-allelic markers and may provide better identification of any putative atherogenic loci in their vicinity.

A possible interpretation of the above studies is that these DNA polymorphisms at the A1–C3 locus are harmless mutations but are acting as linkage markers for a deleterious atherogenic allele in this region. The reasons why the linkage markers vary amongst three populations (West Germany, Britain and North America) could be that (1) the atherogenic allele has mutated several times during evolution and hence associates with different background polymorphisms in different populations, or (2) there have been recombinational events between the atherogenic allele and the different polymorphic restriction sites. The mutation therefore can spread to a different background of chromosomal polymorphisms. The latter is somewhat unlikely in view of the short physical distances between the various linkage markers.

There are also other examples in the apolipoprotein gene family where multiple mutations have probably occurred. For example, the apo C2 gene mutation producing familial apo C2 deficiency is linked to one allele of a Taq 1 polymorphism in a North Italian pedigree, but to the other allele in a Dutch pedigree (Humphries *et al.* 1984). The apo C2 gene has probably mutated twice to account for these different allelic associations. A similar situation may have arisen with the putative atherogenic allele on the long arm of chromosome 11.

Conclusion

These studies represent first attempts to identify major alleles that segregate with the inheritance of premature coronary artery disease. Identification of these atherogenes may allow prediction from an early age of individuals who are predisposed to develop premature atherosclerosis. Thereafter treatment could be instituted to modify environmental factors that interact with such high risk genes to produce disease. One of the most important applications of this area of genetic research is a more precise appreciation of

https://doi.org/10.1079/PNS19870047 Published online by Cambridge University Press

the environmental factors that interact with the genome to predispose to atherosclerosis. Finally information on the major atherogenic loci may lead to newer forms of therapy directed against possible deleterious gene products.

Financial support for these studies have come from the Medical Research Council, Wellcome Trust and British Diabetic Association which is gratefully acknowledged. Financial support from the Fritz-Thysson Foundation is also gratefully acknowledged.

REFERENCES

Barker, D., Holm, T. & White, R. (1984). *American Journal of Human Genetics* **36,** 1159–1171.

Berg, K. (1983). In *Progress in Medical Genetics*, vol. V, pp. 36–90 [A. G. Steinberg, A. Bearn, A. G. Motulsky and B. Childs, editors]. Philadelphia: W. B. Saunders.

Breslow, J. L. (1985). *Annual Review of Biochemistry* **54,** 699–727.

Brown, M. S. & Goldstein, J. L. (1976). *Science* **191,** 150–154.

Ferns, G. A. A., Stocks, J., Ritchie, C. & Galton, D. J. (1985). *Lancet* **i,** 300–303.

Frossard, P. M., Coleman, R., Funke, H. & Assmann, G. (1987). In *Recent Advances in Arteriosclerosis Research*, pp. 53–62 [W. H. Hauss, R. W. Wissler and J. Grunwald, editors]. Dusseldorf: Westdeutscher Verlag.

Galton, D. J. (1985). *Molecular Genetics of Common Metabolic Disease.* London: Edward Arnold.

Humphries, S. E., Williams, L. G., Stalenhoef, A. F., Baggio, G., Crepaldi, G., Galton, D. J. & Williamson, R. (1984). *Human Genetics* **67,** 151–156.

Jeffreys, A. J. (1979). *Cell* **18,** 1–10.

Luo, C. C., Li, W. H., Moore, M. N. & Chan, L. (1986). *Journal of Molecular Biology* **187,** 325–340.

Nora, J. J., Lortscher, R. H., Spangler, R. D., Nora, A. H. & Kimberley, W. J. (1980). *Circulation* **61,** 503–508.

Ordovas, J. M., Schaefer, E. J. & Salem, D. (1986). *New England Journal of Medicine* **314,** 671–677.

Rees, A., Sharpe, C., Stocks, J., Shoulders, C. C., Baralle, F. E. & Galton, D. J. (1985*a*). *Journal of Clinical Investigation* **76,** 1090–1095.

Rees, A., Shoulders, C. C., Stocks, J., Galton, D. J. & Baralle, F. E. (1983). *Lancet* **ii,** 444–447.

Rees, A., Stocks, J., Paul, H., Ohuchi, Y. & Galton, D. J. (1986). *Human Genetics* **72,** 168–171.

Rees, A., Stocks, J., Williams, L. G., Caplin, J. L., Jowett, N. I., Camm, A. J. & Galton, D. J. (1985*b*). *Atherosclerosis* **58,** 269–275.

Southern, E. M. (1975). *Journal of Molecular Biology* **98,** 503–514.

Printed in Great Britain