Molecular genetics of lipoprotein disorders

Jan L. Breslow, M.D.

Genetics, lipoproteins, and atherosclerosis

The results of demographic studies have taught us that factors such as diet, weight, exercise and smoking habits are very important in determining susceptibility to cardiovascular disease. However, complex epidemiologic data and clinical observations have shown that these variables do not explain susceptibility to cardiovascular disease among all members of the population. We all know people who live to their 80s or 90s in spite of unhealthy life-styles, the so-called Winston Churchill syndrome. On the other hand, physicians often see people who maintain a reasonable diet, are not overweight, are active, and do not smoke but who suffer myocardial infarctions in their 20s, 30s, and 40s. It is clear that there must be a fundamental difference between these two groups of people. Investigators have theorized that the first group has intrinsic or genetic factors that prevent heart attacks, whereas the second group has genetic factors that result in premature heart attacks. Epidemiologic studies have shown that the levels of the cholesterol-carrying complexes in the blood, plasma lipoproteins, are strong predictors of which individuals in the population are likely to have premature heart attacks. Two of the plasma lipoproteins have been shown to play particularly strong but opposite roles. In individuals under 50 years of age elevated levels of low-density lipoprotein (LDL) and decreased levels of high-density lipoprotein (HDL) are strong predictors of heart disease. In individuals over 50 years of age LDL is less predictive, perhaps because people with extreme elevations of LDL have already died by that age, and a low HDL concentration is an even stronger predictor of heart disease. In fact, for an asymptomatic person 50 years of age an HDL measurement is the best single laboratory test to predict not only the occurrence of coronary artery disease but survival itself.

Lipoprotein metabolism and apolipoproteins

Lipoproteins are complexes of lipids and proteins (table 1) that are synthesized mainly by the liver and intestine and catabolized by hepatic and extrahepatic tissues. Their physiologic role is to transport dietary and/or endogenous synthesized lipids (cholesterol, triglyceride, and phospholipids) from one organ to another. In addition to LDL and HDL, there are two other lipoprotein classes called chylomicrons and very low-density lipoproteins (VLDL). Each lipoprotein class is metabolically heterogeneous and several subfractions have been described. Plasma lipoproteins are spherical particles with cores of nonpolar neutral lipid consisting of cholesteryl esters and triglycerides and coats of relatively polar materials consisting of phospholipid, free cholesterol, and proteins. The protein components of lipoproteins are called apolipoproteins and have been designated apo A-I, apo A-II, apo A-IV, apo B, apo C-I, apo CII, apo CIII, apo D, and apo E (table 2). The pathways of lipoprotein formation, interconversion, and catabolism in the body are complex and will not be reviewed here. However, research over the last 10 to 15 years has indicated that the apolipoproteins play an important role in each of these processes. For example, the inability to synthesize apo B in the disorder abetalipoproteinemia prevents secretion of several lipoprotein fractions, including chylomicrons, VLDL, and LDL. In addition, apo A-I has been shown to be a cofactor for the lecithin cholesterol acyl transferase enzyme, which is responsible for synthesizing most of the cholesteryl esters in plasma, and apo CII is a cofactor for the lipoprotein lipase enzyme, which is responsible for hydrolyzing triglycerides. Finally, high-affinity lipoprotein receptors that recognize apolipoproteins B and E have been discovered. Recent research has attempted to define at the DNA level those genetically determined factors that control apolipoprotein synthesis and structure. In addition, human genetic lesions affecting apolipoprotein synthesis and structure that result in plasma lipoprotein abnormalities and in some cases premature atherosclerosis are now being identified.

Modern molecular genetics

The human genome is quite complex, consisting of approximately $3 \times 10^9$ base pairs. If one assumes that the average gene is approximately $3 \times 10^3$ base pairs
PERSPECTIVE

TABLE 1
Properties and composition of human plasma lipoproteins

<table>
<thead>
<tr>
<th>Lipoprotein class</th>
<th>Size range (Å)</th>
<th>Density range (g/ml)</th>
<th>Electrophoretic mobility</th>
<th>Triglycerides (% wt)</th>
<th>Phospholipids (% wt)</th>
<th>Free cholesterol (% wt)</th>
<th>Esterified cholesterol (% wt)</th>
<th>Proteins (% wt)</th>
<th>Major apoproteins</th>
<th>Minor apoproteins</th>
</tr>
</thead>
<tbody>
<tr>
<td>Chylomicrons</td>
<td>0.94</td>
<td>750-12,000</td>
<td>Origin (cad)</td>
<td>80-95</td>
<td>3-6</td>
<td>1-3</td>
<td>2-4</td>
<td>1-2</td>
<td>A-I, A-IV, B, Cl, CII, E</td>
<td>A-II, CII</td>
</tr>
<tr>
<td>VLDL</td>
<td>0.94</td>
<td>300-700</td>
<td>Prebeta</td>
<td>45-65</td>
<td>15-20</td>
<td>4-8</td>
<td>16-22</td>
<td>6-10</td>
<td>B, E, Cl, CII, CII</td>
<td>A-I, A-IV, CII, E</td>
</tr>
<tr>
<td>LDL</td>
<td>1.063-1.21</td>
<td>180-300</td>
<td>Beta</td>
<td>4-8</td>
<td>18-24</td>
<td>6-8</td>
<td>45-50</td>
<td>18-22</td>
<td>B</td>
<td>Cl, CII, CII, E</td>
</tr>
<tr>
<td>HDL</td>
<td>1.063-1.21</td>
<td>50-120</td>
<td>Alpha</td>
<td>2-7</td>
<td>26-32</td>
<td>3-5</td>
<td>15-20</td>
<td>45-55</td>
<td>A-I, A-II, E</td>
<td>Cl, CII, CIII, D, F</td>
</tr>
</tbody>
</table>

long, this means that the human genome contains approximately $10^6$ genes. Some DNA in the human genome is repetitive, occurring with frequencies of up to thousands of copies per genome. However, the genes coding for proteins, such as the apolipoproteins, are in most cases single copy. The challenge is to pick the single gene for a given apolipoprotein out of the 1 million genes in the human genome. In addition, one would want to have enough of this material to be able to study its structure and function. As recently as 10 years ago this was an impossible task. However, with the introduction of molecular cloning, this procedure has become routine in many laboratories.

The first technologic advance that allowed this to

TABLE 2
Apolipoproteins and their association with human diseases

<table>
<thead>
<tr>
<th>Apolipoprotein</th>
<th>Plasma concentration (mg/ml)</th>
<th>Isoelectric point (PI)</th>
<th>Mol. wt. (K)</th>
<th>Primary amino acid sequence of mature proteins</th>
<th>Function</th>
<th>Association with clinical disorders</th>
</tr>
</thead>
<tbody>
<tr>
<td>A-I</td>
<td>1.0-1.2</td>
<td>5.40-5.85</td>
<td>28</td>
<td>243A.A.</td>
<td>Activates LCAT</td>
<td>Tangier disease</td>
</tr>
<tr>
<td>A-II</td>
<td>0.3-0.5</td>
<td>5.0</td>
<td>8.5</td>
<td>77A.A.</td>
<td>—</td>
<td>Apo A-I apo C-III deficiency</td>
</tr>
<tr>
<td>A-IV</td>
<td>0.16</td>
<td>5.45</td>
<td>46</td>
<td>Unknown</td>
<td>—</td>
<td>—</td>
</tr>
<tr>
<td>B-100</td>
<td>0.7-1.0</td>
<td>—</td>
<td>549</td>
<td>Unknown</td>
<td>Receptor-mediated catabolism of LDL</td>
<td>Abetalipoproteinemia, normotriglyceridemic abetalipoproteinemia (B-100 deficiency)</td>
</tr>
<tr>
<td>B-48</td>
<td>—</td>
<td>—</td>
<td>246</td>
<td>Unknown</td>
<td>Chylomicron formation</td>
<td>—</td>
</tr>
<tr>
<td>Cl</td>
<td>0.04-0.06</td>
<td>7.5</td>
<td>6.5</td>
<td>57A.A.</td>
<td>Activates (moderately) LCAT</td>
<td>—</td>
</tr>
<tr>
<td>CII</td>
<td>0.03-0.05</td>
<td>4.9</td>
<td>9</td>
<td>79A.A.</td>
<td>Activates lipoprotein lipase</td>
<td>Familial type I hyperlipoproteinemia</td>
</tr>
<tr>
<td>CIII</td>
<td>0.12-0.14</td>
<td>4.7-5.0</td>
<td>9</td>
<td>79A.A.</td>
<td>Inhibits catabolism of Apo E-containing lipoproteins</td>
<td>Apo A-I apo C-III deficiency</td>
</tr>
<tr>
<td>D</td>
<td>0.06-0.07</td>
<td>5.0-5.2</td>
<td>32.5</td>
<td>Unknown</td>
<td>—</td>
<td>—</td>
</tr>
<tr>
<td>E</td>
<td>0.025-0.050</td>
<td>5.7-6.0</td>
<td>34.2</td>
<td>299A.A.</td>
<td>Receptor-mediated catabolism</td>
<td>Familial type III hyperlipoproteinemia</td>
</tr>
</tbody>
</table>
Thus ensured with relatively few clones. The fragments generated from the human genome, a hexanucleotide recognition enzyme could cut the genome into approximately 10⁶ restriction fragments.

The second indispensable part of this technology has been the advent of molecular cloning. This breakthrough was the result of the development of cloning vectors that replicate autonomously in Escherichia coli even when joined to foreign DNA. The restriction fragments generated from the human genome can be used to replace regions of vector DNA not essential for propagation in bacteria. The foreign DNA inserted into these regions is then replicated as if it were a normal component of the vector. The cloning vectors in common use today are plasmids, bacteriophage lambda, cosmids, and bacteriophage M13. It is possible to grow large quantities of these vectors containing inserted foreign DNA so that one can then isolate sufficient material for study.

The third concept important to molecular genetics is library construction. A DNA library contains cloned DNA fragments representing all of the source DNA sequences. There are two types of libraries in common use. The first is a library of genomic DNA. DNA is partially digested with a restriction enzyme and cloned into a bacteriophage lambda vector. To achieve a 99% probability of having a given DNA sequence represented in a library of, on the average, 17 kilobase fragments of human DNA requires 8 × 10⁶ recombinant clones. The second type of library in common use is a cDNA or copy DNA library. A cDNA library is made from the polyA-containing mRNA of a tissue synthesizing the desired protein, and this cDNA is usually cloned in a plasmid vector. In a tissue making relatively large amounts of a given protein, the corresponding mRNA is enriched in the sequences of interest. Thus representation in a cDNA library in this situation is ensured with only 10³ to 10⁴ recombinant clones.

A fourth component of this technology is the ability to select the clone with the DNA sequences of interest from the particular library under study. There are three general methods in use. The first one, hybridization selection, relies on the ability to accomplish cell-free synthesis of the protein of interest with a preparation of mRNA. Clones from the DNA library are then used to select by hybridization the particular mRNA of interest from the other mRNAs. The particular clones capable of doing this presumably have DNA sequences corresponding to this mRNA. The second method is the use of oligonucleotide probes to directly screen genetic libraries. If the protein sequence is known, it is possible to envision all of the particular mRNA sequences that could code for a particular segment of the protein. One can synthesize such a mixture of oligonucleotides, radiolabel them, and use them as probes to select clones containing DNA sequences corresponding to the particular protein of interest. The final method is the use of expression vectors. In these vectors, bacterial clones actually produce the foreign protein corresponding to the cloned DNA. One can then use an antibody to detect which bacterial clone contains the DNA of interest.

A fifth important technology is that of DNA sequencing either by the chemical technique of Maxam and Gilbert or the enzymatic technique of Sanger. With these techniques it is possible to derive the sequence of the foreign DNA in the clone of interest and determine whether this corresponds to the known primary amino acid sequence.

Cloned DNA can be used in many ways. cDNA clones can be used to determine mRNA structure from which the amino acid sequence of the primary translation product of a protein of interest can be deduced. Genomic clones can be used to identify aspects of gene structure such as promoters and introns. cDNA and genomic clones can be used to determine tissue levels of specific mRNAs. They can also be used to assess chromosomal localization, linkage to other genes, and disease-associated polymorphisms. Finally, they can be genetically engineered to produce altered gene products that can be used to study protein structure-function relationships.

**Strategy for isolating and characterizing the apolipoprotein genes**

In our laboratory, we have isolated and are in the process of characterizing cDNA and genomic clones for four of the apolipoproteins: apo A-I, apo CII, apo CIII, and apo E. Our strategy has been to identify a
region of the apolipoprotein primary amino acid sequence specified by relatively unambiguous codons. We have then synthesized cDNA oligomers corresponding to all possible mRNAs that could code for these amino acids. This mixture of oligomers is radiolabeled and used as a probe to screen a library of cDNA made from adult human liver mRNA and cloned in a plasmid vector. A cDNA clone is then radiolabeled and used as a probe to screen a library of human genomic DNA cloned in bacteriophage lambda. The selected cDNA and genomic clones are characterized by restriction endonuclease mapping, subcloning, and DNA sequencing. Chromosomal localization of the apolipoprotein genes is accomplished through the use of cDNA probes and somatic cell hybrids. In the latter technique, human and rodent cells are fused and then propagated in culture. During this process, the hybrids tend to lose human chromosomes. After a period of time, the hybrid cells contain the rodent genome and only a few human chromosomes. At that time the hybrids are characterized as to which human chromosomes are still remaining, and DNA is prepared and probed for the apoprotein gene of interest. After looking at many hybrids, one observes which human chromosome is present whenever the apoprotein gene is present and which human chromosomes are absent whenever the apoprotein gene is absent. In this way it is possible to deduce which human chromosome contains a particular apoprotein gene. Through these studies we have been able to determine that the apo A-I and apo CIII genes reside on human chromosome 11. Another group has shown through family studies of the inheritance of protein polymorphisms that the apo A-IV and apo A-I genes are linked. Thus three of the apolipoprotein genes are on human chromosome 11. In addition, we have also shown that the genes for apo E and apo CII reside on human chromosome 19. I will summarize briefly our knowledge of the structure and mutations at each of these apoprotein chromosomal loci (table 3).

### Table 3: Apolipoprotein Genes

<table>
<thead>
<tr>
<th>Chromosome 11</th>
<th>Chromosome 19</th>
</tr>
</thead>
<tbody>
<tr>
<td>A-IV</td>
<td>E^a</td>
</tr>
<tr>
<td>A-I</td>
<td>CII</td>
</tr>
<tr>
<td>CIII</td>
<td></td>
</tr>
</tbody>
</table>

**Disorder**
- Apo A-I, apo CIII deficiency
- Hypertriglyceridemia
- Type III hyperlipoproteinemia

**Gene lesion**
- DNA insertion in apo A-I gene
- Altered apo CIII allele
- Apo E structural gene mutations

^Apoprotein A-1 and apo E genes structurally similar (intron location and intragenic duplications).

The apo A-I/apo C III gene locus

The gene for apo A-I has been characterized; it is approximately 2200 base pairs long and is interrupted by three intervening sequences or introns. In addition, the gene for apo CIII is very close to that for apo A-I. Studies indicate that apo A-I and apo CIII are coded for by opposite DNA strands. This is a very unusual configuration and may have some functional significance. In clinical studies we have examined DNA from individuals described by Norum et al. who have very low plasma HDL levels, absence of apo A-I and apo CIII, and very severe premature atherosclerosis. We have shown that these individuals have an alteration in their apo A-I gene caused by an insertion of DNA in the region that codes for the apo A-I protein. This is the first demonstration of a DNA insertion causing disease in humans and is also the first demonstration at the DNA level of a genetic lesion causing atherosclerosis. In another clinical study, Rees et al. have found that a specific DNA alteration near the apo A-I gene was present in 50% of individuals with elevated plasma triglyceride levels and in only 5% of normal individuals. Our investigations have shown that the DNA alteration associated with the apo A-I gene is actually an alteration in the apo CIII gene. Our findings raise the question of whether primary abnormalities in the apo CIII gene can cause elevated triglyceride levels in humans and by this mechanism be associated with human atherosclerosis.

The apo E gene locus

We have also performed extensive studies of the apo E gene. Apo E in normolipidemic humans is distributed equally between HDL and VLDL. This apolipoprotein is involved in receptor-mediated lipoprotein recognition by cells. In addition, apo E appears to be important in the metabolism of dietary cholesterol, since it has been observed in both animals and humans that the ingestion of dietary cholesterol leads to the appearance of unusual lipoproteins that are rich in apo
E. In the past we have demonstrated that apo E occurs in different individuals in one of six phenotypes. Through familial studies we showed that the apo E phenotypes were the result of three alleles at a single genetic locus and proposed that this variation in apo E was caused by structural gene mutations. A work has been confirmed and the actual amino acid residues at which genetic variation takes place have been identified. In other studies, we and others have shown that homozygosity for one of the apo E alleles, ε2, resulting in the apo E phenotype E2/2, is the underlying biochemical defect in the disorder type III hyperlipoproteinemia. This disorder is characterized by elevated plasma cholesterol and triglyceride levels, xanthomas, and premature coronary and peripheral vascular disease. An analysis of apo E phenotypes in the general population indicates that 1% to 2% are homozygous and 25% heterozygous for the ε2 apo E allele. As a consequence, the apo E genetic locus can be considered a common atherosclerosis gene. Further studies have indicated that the actual expression of hyperlipidemia and atherosclerosis in this condition is dependent on other environmental, endocrine, or genetic factors. In recent work, we have isolated the gene for apo E; it is approximately 4000 base pairs long and is interrupted by three intervening sequences. Our studies at the DNA level have indicated common features between the structures of the apo E and the apo A-I genes. Both genes have three introns in similar locations, and in each case the fourth exon consists largely of tandemly repeated DNA sequences that presumably arose by intragenic duplications. These similarities imply a close phylogenetic relationship between these genes and indicate that they may have arisen from a common evolutionary precursor. In addition, we have identified the particular DNA base alterations responsible for the different apo E structural gene alleles (table 4). Through this work it is now possible to construct DNA probes that recognize the normal apo E allele as well as each of the mutant apo E alleles. Therefore it is now possible to screen for mutant apo E alleles in the population with DNA probes.

### Summary

These types of studies lay the foundation for understanding the human apolipoproteins at the genetic level. In addition, through a few specific examples, we have already shown evidence at the DNA level that lesions in apolipoprotein genes can play a role in human atherosclerosis. This work is in its infancy, but we hope it will have a major impact on our ability to understand genetic susceptibility to atherosclerosis in the general population.

### References

2. McPherson J, Bruns GAP, Karathanasis SK, Breslow JL: Isolation, characterization and mapping to chromosome 19 of the human apolipoprotein E gene. (Submitted for publication)
Molecular genetics of lipoprotein disorders.
J L Breslow

Circulation. 1984;69:1190-1194
doi: 10.1161/01.CIR.69.6.1190

Circulation is published by the American Heart Association, 7272 Greenville Avenue, Dallas, TX 75231
Copyright © 1984 American Heart Association, Inc. All rights reserved.
Print ISSN: 0009-7322. Online ISSN: 1524-4539

The online version of this article, along with updated information and services, is located on the World Wide Web at:
http://circ.ahajournals.org/content/69/6/1190