Identification of functional domains in the plasma apolipoproteins by analysis of inter-species sequence variability

Richard B. Weinberg

Department of Internal Medicine, Section of Gastroenterology, The Bowman Gray School of Medicine, Winston-Salem, NC 27157

Abstract Molecular evolution theory posits that sequence motifs essential for protein function are constrained by selective pressure from changing over long stretches of evolutionary time. Thus, analysis of inter-species amino acid sequence variability, by identifying highly conserved intervals, should predict the location of domains critical for protein function. We have analyzed the amino acid sequences of the mammalian apolipoproteins A-I, A-IV, C-I, C-II, C-III, D, and E with a computer algorithm that calculates numerical residue variability scores. The application of a median sieve filter to the data facilitated identification of the exact boundaries of highly conserved domains, which coincided with the location of known structural features and functional domains in this family of proteins. The analysis also identified highly conserved intervals in every apolipoprotein whose function is unknown at present, but which are candidates for regions with specific functional roles. **—Weinberg, R. B.** Identification of functional domains in the plasma apolipoproteins by analysis of inter-species amino acid sequence variability. *J. Lipid Res.* 1994. **35:** 2212–2222.

Supplementary key words structure-function relationships • apolipoproteins • computer analysis • protein evolution • median filter

Inter-species comparison of amino acid and nucleotide sequences is a powerful tool in the study of protein evolution. Based on the principle that random mutations occur at a fixed frequency (1), analysis of amino acid and/or nucleotide sequence variability provides a molecular clock that can date the divergence of species (2), and disclose convergent, divergent, or parallel paths of evolution (3, 4). Moreover, sequence comparisons within protein families can reveal the molecular mechanisms by which proteins of distinctly new function have evolved from one another (5). A corollary to this principle is that sequence motifs essential for a protein function are constrained by selective pressure from changing over long stretches of evolutionary time. Thus, highly conserved sequences should identify domains that are critical for function. Conversely, sequences in nonfunctional domains, unaffected by selective constraints, change at a fixed rate. The greater the variability of a sequence over a short evolutionary span (e.g., within the same class or order), the lower the probability that it effects a specific function.

This relationship between sequence conservation and protein function suggests that inter-species sequence comparison should be a useful approach to the prediction of structure–function relationships in the plasma apolipoproteins, a family of lipid-binding proteins that regulate lipoprotein metabolism (6). We have therefore analyzed amino acid variability in the sequences of the exchangeable mammalian apolipoproteins with a computer algorithm that detects highly conserved regions. Here we report that this analysis successfully identified the location of known structural features and functional domains in these proteins, and also identified highly conserved intervals in every apolipoprotein whose function is unknown at present, but which are candidates for domains with specific functional roles.

MATERIALS AND METHODS

Apolipoprotein sequence acquisition

The sequences for the mammalian plasma apolipoproteins A-I, A-IV, C-I, C-II, C-III, E, and D; human retinol protein; tobacco hornworm (*Manduca sexta*) insecticyanin; and *Pierus brassicae* bilin binding protein were downloaded as single letter code print files from the Swiss Protein and GenBank data bases. The sequence of baboon apoA-IV was taken from Hixson et al. (7). Sequences were aligned with the sequence analysis program PILEUP (Genetics

Abbreviations: HDL, high density lipoprotein; VLDL, very low density lipoprotein; LDL, low density lipoprotein; LCAT, lecithin:cholesterol acyltransferase.

Computer Group, Madison, WI), which performs a progressive series of pairwise alignments between sequences to create a final multiple sequence alignment.

Sequence variability analysis

Analysis of interspecies sequence variability was performed using Lotus 123 software (version 2.01, Microsoft Corporation). A substitution matrix was used to assign numerical values to the degree of conservation for each possible amino acid interchange based on a combination of amino acid charge, hydrophobicity, and steric properties (8). A conservation score was calculated for each residue of an aligned sequence as the sum of the substitution scores for all possible pairwise comparisons at that position (**Appendix 1**). To facilitate comparisons among apolipoproteins, the conservation scores were converted to normalized variability scores by the formula: $V_x = (C_{max} - C_x)/C_{max}$, where V_x is the variability score at position x, C_x is the conservation score at position x, and C_{max} is the maximum score corresponding to a "perfect" match for all the compared sequences in that analysis. Variability scores were plotted as bar graphs versus sequence number.

The substitution matrix analysis generated high frequency noise from random point mutations that obscured the exact boundaries of highly conserved domains (**Fig. 1A**). A common method to eliminate such noise is to smooth the data with a "sliding window" mean, which moves a variable size window along the sequence and assigns the numerical average of all the values within the window to the residue in the center. However, this technique seriously blurred the edges between regions with disparate scores, and was also overly sensitive to outlying values within otherwise homogeneous regions (Fig. 1B). Median filtering provided better resolution. This technique also uses a moving window, but assigns the interval median value to the center of the window. The sequential application of median filters of increasing window size (cascade median filtering or median sieving) further increases the power of this technique to identify discrete protein domains (9). The Lotus 123 numerical sorting and macro features were used to execute the sieving process (**Appendix 2**). Three passes through the sequence using progressively wider window widths of 3, 5, and 7 amino acids were performed. Sieving at larger window widths did not yield further improvement in resolution.

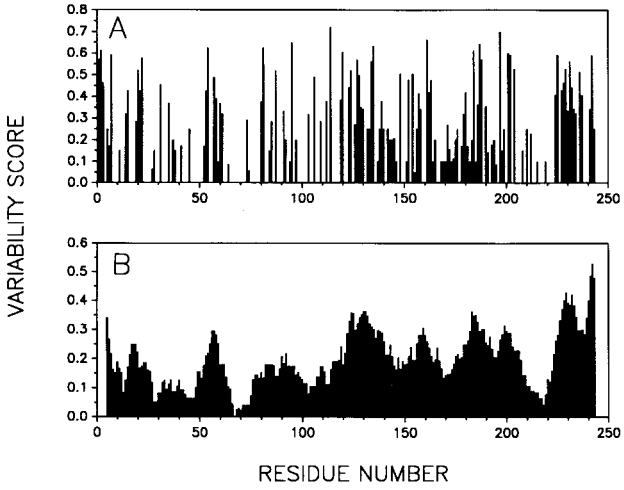

Fig. 1. Substitution scoring matrix analysis of apolipoprotein A-I. A) raw comparison scores; B) scores smoothed by a sliding-window mean filter, 11 amino acids in width.

RESULTS AND DISCUSSION

Properties of amphipathic helices in the plasma apolipoproteins

In the following discussion it will be useful to consider the properties of three classes of amphipathic alpha helices in the plasma apolipoproteins that can be distinguished on the basis of the distribution of charged residues on the polar face of the helix, the mean hydrophobicity of the residues on the nonpolar face, and the hydrophobic moment of the helix, a measure of the directionality of the hydrophobic face (10). Class A helices, the predominant type in most of the apolipoproteins, are characterized by a distinctive clustering of positively charged residues at the polar–nonpolar junction with negatively charged residues in the middle of the polar face, high hydrophobicity, and a large hydrophobic moment. Class A helices are very surface-active, and are thought to be the primary mediators of lipid binding. Class Y helices are present only in apoA-IV, and to a lesser extent apoA-I, and are characterized by the presence of two discrete clusters of negative residues on the polar face, low hydrophobicity, and a smaller hydrophobic moment than class A helices. Class Y helices are relatively hydrophilic and may not be able to penetrate deeply into lipid surfaces; consequently their binding to lipid may be easily disrupted. Class G* helices are similar to the alpha helices found in globular proteins in that they have a random distribution of charged residues in their polar faces and moderate hydrophobicity; however, they have a large hydrophobic moment. These properties may enable class G* helices to mediate intra- and intermolecular protein-protein interactions which stabilize tertiary conformation, self-association, or apolipoprotein-enzyme binding.

Apolipoprotein A-I

Apolipoprotein A-I (apoA-I) is a 28,100-dalton protein, 243 amino acids in length, that is synthesized in the intestine and liver (6). It is the major protein component of high density lipoproteins (HDL), and plasma levels are

inversely correlated with the risk of coronary atherosclerosis (11). ApoA-I is the primary activator of lecithin:cholesterol acyltransferase (LCAT) (12), a plasma enzyme that esterifies cholesterol at the HDL surface. ApoA-I may also be a ligand that mediates the interaction of high density lipoproteins and cell surfaces (13).

The matrix/median filter analysis for apoA-I was performed on sequences from eight mammalian species: human, baboon, macac, dog, pig, cow, rat, and rabbit (**Fig. 2**). The analysis revealed that the amino terminal half of apoA-I is more conserved than the carboxyl terminal half. The region of highest variability between amino acids 120 to 200 coincides with the location of four class A helical lipid binding domains (10), which suggests that the amphipathic helices can tolerate considerable substitution without disruption of their function. The short stretches of variability near the amino terminus correspond well with the boundaries between 11-mer repeats (10). It is of interest that the majority of naturally occurring apoA-I mutations are located within these variable regions (14).

Three regions of highly conserved sequence were identified by the analysis: *1)* residues 10 to 50, with a short variable gap at residues 20 to 23; *2)* residues 61 to 120, with two short variable gaps at residues 72 to 77 and 81 to 86; and *3)* residues 202 to 224, near the carboxyl terminus. The conserved region at the amino terminus coincides with the location of the single class G* helix in apoA-I, at residues 8–33. Although no specific function for this region is currently recognized, the observations of Marcel et al. (15) that this region shares epitopes with central regions of the apoA-I molecule suggests that this G* helix could mediate intramolecular protein–protein interactions critical for apoA-I folding.

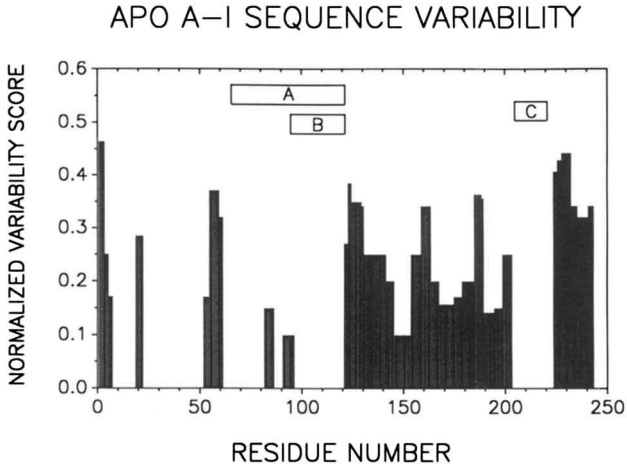

Fig. 2. Matrix/median sieve analysis of apolipoprotein A-I. Data from Fig. 1A were filtered with a median sieve. "A" marks the LCAT activation sequence determined by synthetic peptides (15); "B" marks the LCAT activation/hinge domain determined by monoclonal antibodies (16); "C" marks the high density lipoprotein receptor binding domain (17).

However, the other two conserved regions coincide extremely well with domains identified by experimental approaches as important for apoA-I function. The interval from residue 60 to 120 is identical to the core sequence (marked "A" in Fig. 2) determined using model peptides as essential for LCAT activation (16), and also encompasses the region (marked "B") identified by monoclonal antibodies as a hinged domain involved in the LCAT reaction (17). The interval from residue 202 to 224 coincides with a domain (marked "C") found to mediate the binding of apoA-I to the putative cellular high density lipoprotein receptor (18). This region also coincides with the region of maximum amphipathic moment identified by Nolte and Atkinson (19) using power spectrum analysis, which suggests that it could also serve to anchor the carboxy terminus to the lipoprotein surface. Both naturally occurring (20) and recombinant (21) mutations in these two conserved domains significantly disrupt protein function.

Apolipoprotein A-IV

Apolipoprotein A-IV (apoA-IV) is a 46,000-dalton protein, 376 amino acids in length, that is synthesized in the intestine and the liver (22). ApoA-IV is the most hydrophilic apolipoprotein (23); consequently it displays labile affinity for lipid surfaces (24, 25) and circulates primarily unassociated with plasma lipoproteins (26). ApoA-IV binds with high affinity to cell surfaces (27, 28), and facilitates phospholipid and cholesterol efflux from cells (29, 30). ApoA-IV activates LCAT (31), and accelerates the rate of HDL speciation catalyzed by cholesteryl ester transfer protein (32). These observations, taken together, suggest that apoA-IV may play a role in the earliest stages of peripheral HDL assembly, and may participate in the process of reverse cholesterol transport.

The matrix/median filter analysis of the human, baboon, rat, and mouse apoA-IV sequences revealed a distinctive pattern of repeating variable regions cyclically punctuated by short conserved intervals (**Fig. 3**). The variable regions all correspond to 11- and 22-mer alpha helical repeats (10). With few exceptions, the highly conserved intervals coincide exactly with the location of proline residues and/or beta turns, which punctuate the alpha helical domains in apoA-IV (33). The conservation of these sequential beta turns suggests that they may be important for the function of apoA-IV. For example, recent studies by Jonas et al. (34) have shown that apoA-IV forms the largest observed apolipoprotein–phospholipid micellar complexes. The multiple beta turns may confer flexibility and expandability to apoA-IV which help stabilize these complexes as they acquire cellular lipid and are transformed into spherical particles by the action of LCAT (35). This conjecture is supported by the recent ob-

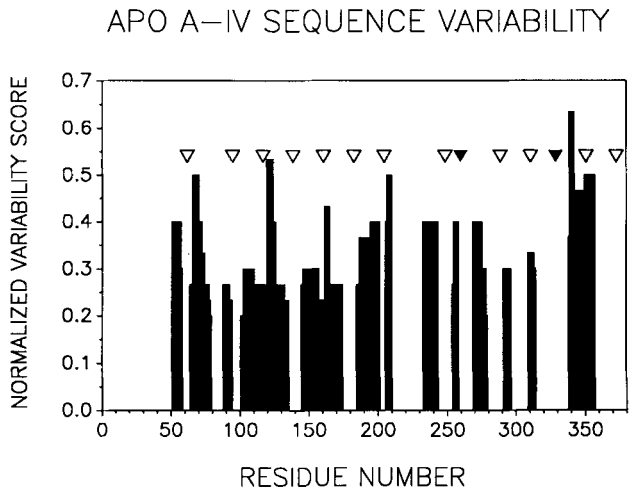

Fig. 3. Matrix/median sieve analysis of apolipoprotein A-IV; ▽ marks proline-containing beta turns; ▼ marks non-proline-containing beta turns.

servation that deletion of residues 118–161 abolishes 90% of the LCAT activation by apoA-IV (36).

The highly conserved 50 amino acids at the amino terminus encompass the only a class G* helical domain in apoA-IV, at residues 7 to 31 (10). As with apoA-I, no specific function for this region is currently recognized, but the presence of a G* helix makes it a candidate domain to mediate the distinctive high affinity self-association of apoA-IV (37). The conserved interval at residues 314–338 coincides with the location of a class Y helix at residues 311–332 (10). The conservation of class Y helices in apoA-IV may be responsible for the unusually low surface exclusion pressure of this apolipoprotein (25). The highly conserved interval between residues 354 to 370 at the carboxyl terminus is the location of a repeated series of four tetrapeptides with the sequence GLU-GLN-(GLN/ALA/VAL)-GLN. This unique motif, which is not present in any other apolipoprotein, is the site of several murine (38), baboon (7), and human (39, 40) mutations that affect apoA-IV structure (41), serum lipid levels (42, 43), and dietary responsiveness (7, 42, 44).

Apolipoprotein C-I

Apolipoprotein C-I (apoC-I) is a 6,613-dalton protein, 57 amino acids in length, that associates with very low density lipoproteins (VLDL) and HDL, and exchanges freely between their surfaces in circulation (6). It is the smallest apolipoprotein, and analysis of the molecular evolution of the apolipoproteins suggests that apoC-I may be the descendant of the primordial ancestral gene of the apolipoprotein family (5). Little is known about the specific function of apoC-I in human lipoprotein metabolism, but it has been found to activate LCAT (45), and inhibit hepatic lipase (46). It may also modulate the interaction of apoE and β-VLDL and inhibit the binding of β-VLDL to low density lipoprotein receptor-related proteins (47).

The matrix/median filter analysis of the human, baboon, dog, and rat apoC-I sequences revealed three highly conserved intervals between residues 13–27, 43–48, and 53–58 (**Fig. 4**). The conserved region at the amino terminal overlaps with the part of the molecule important for LCAT activation (marked "A"). Studies with synthetic fragments have established that truncation of residues 17–32 reduces LCAT activation by 50%, and that further truncation of residues 33–39 results in complete loss of LCAT activation (48). The carboxyl terminal 39–57 residues of apoC-I neither bind lipid nor activate LCAT (48), but their conservation suggests they could be important for a specific function, such as the modulation of apoE binding to β-VLDL.

Apolipoprotein C-II

Apolipoprotein C-II (apoC-II) is a 8,824-dalton protein, 79 residues in length, that is synthesized in the liver. Like apoC-I, apoC-II is associated with very low density lipoproteins (VLDL) and HDL, and can exchange between the surfaces of these lipoprotein classes in circulation (6). The major function of apoC-II is the activation of lipoprotein lipase (49), an endothelial enzyme that mediates the intravascular hydrolysis of chylomicron and VLDL triglycerides.

The matrix/median filter analysis for apoC-II was performed on six mammalian sequences: human, macac, dog, cow, rat, and guinea pig. The analysis revealed that

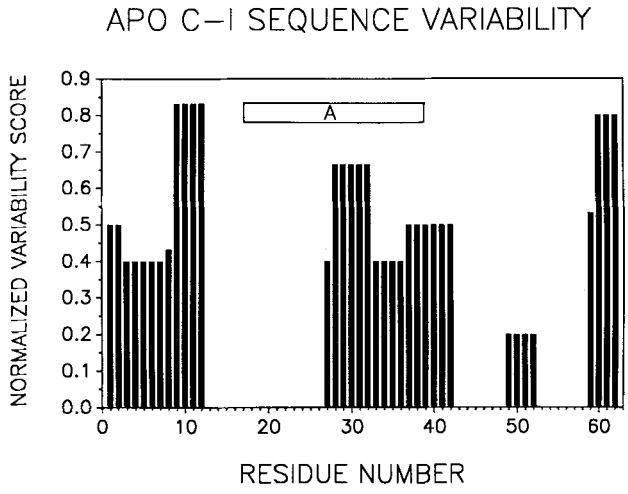

Fig. 4. Matrix/median sieve analysis of apolipoprotein C-I. "A" marks the LCAT activation sequence determined by synthetic peptides (45).

the amino and carboxyl terminal regions of this small protein are quite variable, but that a highly conserved region is located between residues 47 to 71 (**Fig. 5**). This region coincides perfectly with the domain (marked "A") previously identified by synthetic peptides (50) and cyanogen bromide fragments (51) as essential for the activation of lipoprotein lipase. This region also coincides with the location of the single class G* helix in apoC-II (between residues 60–76), which raises the possibility that this region may mediate the interaction between apoC-II and lipoprotein lipase (52).

Apolipoprotein C-III

Apolipoprotein C-III (apoC-III) is a 8,764-dalton protein, 79 amino acids in length, that is synthesized in the liver (6). ApoC-III comprises 50% of the total protein in VLDL and 2% of total protein in HDL (53). A carbohydrate side chain consisting of galactosamine, galactose, and 0–3 sialic acid residues is attached by an O-glycosidic linkage to threonine 74. Several lines of evidence implicate apoC-III in the metabolism of triglyceride-rich particles: apoC-III inhibits lipoprotein lipase (54) and hepatic triglyceride lipase in vitro (46) and in vivo (5, 56), and may inhibit hepatic chylomicron remnant uptake (57).

The matrix/median filter analysis for apoC-III was performed on five mammalian sequences: human, dog, cow, rat, and pig. The analyses revealed three short conserved intervals between residues 10–14, 39–41, and 49–54; otherwise, the protein appeared to be quite variable, particularly at the amino and carboxyl terminii (**Fig. 6**). The amino terminal half of apoC-III does not bind lipid (58), and no specific function for the interval between 10–14 is suggested in the literature, although it is located in the middle of the sole class G* helix in this protein. The conserved interval at 39–41 coincides with a thrombin cleavage site (58), which suggests that aqueous exposure of this region may be critical for some function. The conserved interval at residues 49–54 is in the middle of an amphipathic alpha helix (marked "A") that is critical for the lipid binding of the C-terminal half of the molecule (48), and is thus a candidate for the domain mediating the inhibition of lipoprotein lipase (59). The variability of the carboxyl terminus suggests that it may not be critical for function. In this regard, a $Thr_{74} \rightarrow Ala$ mutation, which deletes the O-linked carbohydrate attachment site, is not associated with dyslipidemia (60); similarly, the carbohydrate moiety is not necessary for in vitro synthesis, secretion, or lipid binding (61).

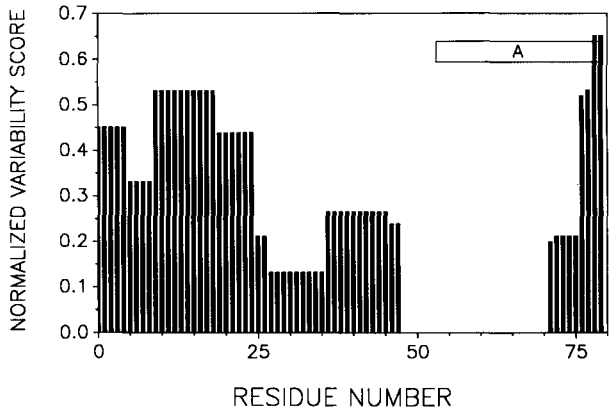

Fig. 5. Matrix/median sieve analysis of apolipoprotein C-II. "A" marks the lipoprotein lipase activating domain determined by peptide fragments (47) and synthetic peptides (48).

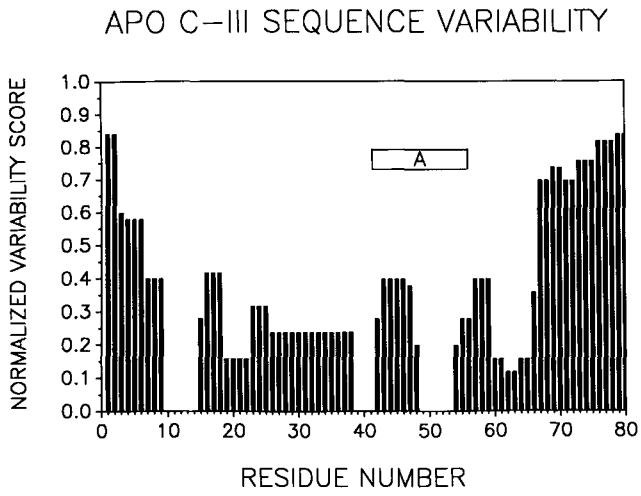

Fig. 6. Matrix/median sieve analysis of apolipoprotein C-III. "A" marks the critical lipid association sequence determined by synthetic peptides (45).

Apolipoprotein D

Apolipoprotein D (apoD) is a 19,300-dalton glycoprotein, 169 amino acids in length. Although it was first identified and purified from HDL subfractions (62), it is now apparent that it is distributed in a wide variety of tissues (63). The structure of apoD does not resemble the other exchangeable apolipoproteins (64), and it has a different evolutionary pedigree. ApoD is a member of the alpha-2 microglobulin family of hydrophobic ligand carrier proteins, which includes retinol-binding protein, lactoglobulin, uteroglobulin, insecticyanin, and bilin-binding protein (65). The common structural feature of this family is a "beta barrel" consisting of two orthogonally oriented sets of four anti-parallel beta strands that form a hydrophobic ligand binding pocket. The crystal structure of several proteins in this family has been described, and the residues proximate to the ligand within the pocket have been identified (66–68).

As the human, rabbit, mouse, and rat apoD sequences are practically identical, we instead compared the sequence of human apoD with three other members of the alpha-2 microglobulin family to determine whether the matrix/median filter method would correctly locate the ligand contact residues within conserved regions. The sequences of human apoD, human retinol-binding protein, tobacco hornworm (*Manduca sexta*) insecticyanin, and *Pierus brassicae* bilin-binding protein were aligned and gaps in the sequences were assigned the non-identity marker X. Note that as these four proteins differ in length, the residue numbers in **Fig. 7** do not correspond to the actual residue numbers of apoD.

The analysis revealed considerable sequence variability, which may, however, be an artifact of the alignment process, as gaps were scored as regions of non-identity. Two highly conserved regions were identified between residues 114 to 125, and 145 to 152 (Fig. 7). The former has previously been identified as the most highly conserved sequence interval in this protein family (68). In general, amino acid residues identified as ligand contact zones in human serum retinol-binding protein (68), insecticyanin (66), and bilin binding protein (67) were located within intervals of lower sequence variability.

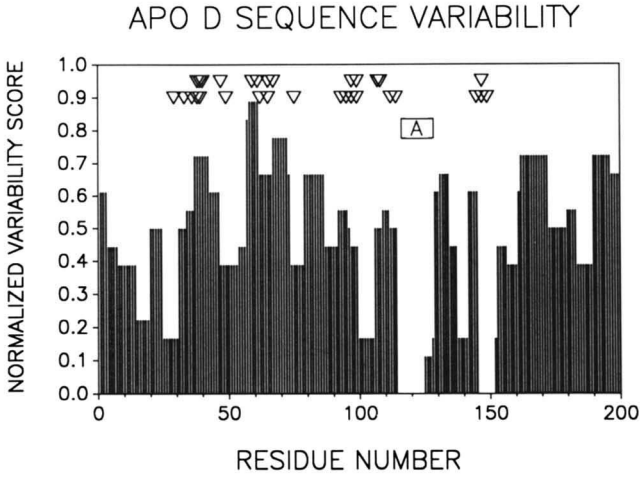

Fig. 7. Matrix/median sieve analysis of apolipoprotein D; ▽ marks ligand contact residues determined by X-ray diffraction of retinol binding protein (upper tier), and bilin binding protein and insecticyanin (lower tier). "A" marks the most conserved sequence in the alpha-2 microglobulin family (61).

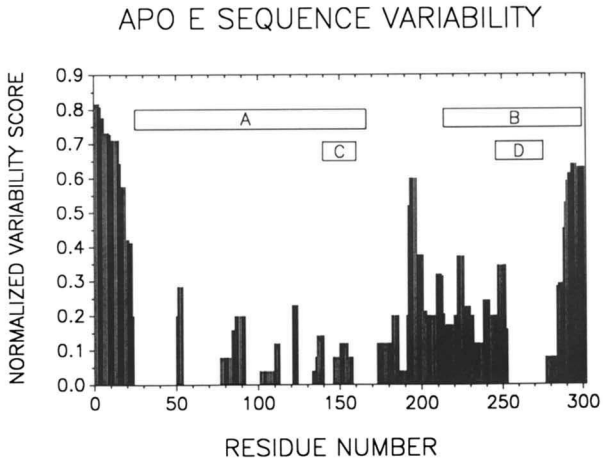

Fig. 8. Matrix/median sieve analysis of apolipoprotein E. "A" marks the four helical bundle domain identified by X-ray diffraction (70); "B" marks the lipid associating domain; "C" marks the LDL receptor binding domain (71); "D" marks the heparin binding domain (73).

Apolipoprotein E

Apolipoprotein E (apoE) is a 34,200-dalton glycoprotein, 299 amino acids in length, that is synthesized in the liver, endocrine tissues, the central nervous system, and macrophages (69). ApoE plays a major role in cholesterol metabolism: it serves as a ligand for both the LDL B/E receptor and a distinct hepatic chylomicron remnant receptor (70), and promotes efflux of cholesterol and lipids from the cell surface to HDL (71). ApoE also binds avidly to heparin (72), a property that may regulate its interaction with the vascular endothelium or interstitial cell matrix. Thrombin cleaves apoE into two fragments at amino acid 189 (73); the amino terminal fragment has been crystallized, and its structure has been determined by X-ray diffraction (74).

The matrix comparison/median filter analysis for apoE was performed on ten mammalian sequences: human, baboon, macac, dog, rat, mouse, guinea pig, rabbit, cow, and sea lion (**Fig. 8**). The analyses revealed the highest variability scores at the amino terminus from residue 1 to 24, at the carboxyl terminus from residues from 280 to 299, and between residues 182 to 205. This latter interval corresponds to the random coil linker region between the amino and carboxyl terminal domains (75) and includes the thrombin cleavage site. The region marked "B" has been identified as an alpha helical lipid binding domain (75), and its relative variability again suggests that amphipathic alpha helices can tolerate substitutions without loss of function.

The most notable finding was a long, highly conserved region in the amino terminal two-thirds of the protein between residues 22 and 182, interrupted by four short variable intervals at residues 80–89, 110–113, 121–124, and 133–140. This region coincides with a globular domain comprised of four clustered helical bundles (marked "A"), which has been identified by X-ray diffraction (74), as well as the location of four class G* helices (at residues 25–51,

52–83, 91–116, and 135–160) which may stabilize the helical bundle array. It also includes the sequence motif (marked "C") that mediates the binding of apoE to the LDL receptor (76). The short variable interval between residues 80–89 marks both the junction between two 11-mer repeats and a short loop at the junction of helices 2 and 3; similarly, the short variable interval between residues 120–123 marks both an 11-mer junction and a short loop joining helices 3 and 4 (10). Another region of highly conserved sequence was identified at the carboxyl terminus between residues 252 to 279; this corresponds to the heparin binding domain (marked "D"), identified by Weisgraber et al. (77), as well as the strongest amphipathic domain in apoE, identified by Nolte and Atkinson using power spectrum analysis (19).

Conclusions

Analyses of inter-species amino acid variability in the plasma apolipoproteins identified highly conserved intervals that correspond to known structural features and functional domains in these proteins. These include: the LCAT activating and receptor binding domains in apoA-I; a distinctive pattern of sequential beta turns possibly related to reverse cholesterol transport in apoA-IV; the LCAT activating domain in apoC-I; the lipoprotein lipase activating domain in apoC-II; a possible lipoprotein lipase inhibition domain in apoC-III; and the alpha helical bundle, LDL receptor binding, and heparin binding domains in apoE as highly conserved intervals. Comparison of apoD with three related proteins of the alpha-2 microglobulin family located ligand contact residues within relatively conserved intervals. The analyses also identified other highly conserved domains in every apolipoprotein whose function is not known at present, but whose conservation predicts a role in mediating new and as yet unrecognized metabolic functions.

The author thanks Mary Sorci-Thomas for inspiration and encouragement, and Mark Lively for searching the Swiss Protein and GenBank data bases. This research was supported by Grant R01-HL30897 from the National Heart, Lung, and Blood Institute of the National Institutes of Health.

Manuscript received 18 February 1994 and in revised form 27 June 1994.

APPENDIX 1

Example spreadsheet set up for substitution matrix conservation scoring of aligned protein sequences.

(Sequence Listing and Scoring)

	A	B	C	D	E	F	G	H
1	DOG[a]		PIG		RAT		SCORE	
2	L	1[c]	L	1	L	1	6[d]	
3	V	2	V	2	L	1	4	
4	L	1	L	1	D	3	2	
5	D	3	L	1	V	2	1	
6	V	2	X	4	V	2	2	
7	D	3	D	3	D	3	6	

(Substitution Matrix)

	I	J	K	L	M	N
		L[b]	V	D	X	#
L	2	1	0	0	1	
V	1	2	0	0	2	
D	0	0	2	0	3	
X	0	0	0	0	4	

[a]Amino acid sequences for each species are listed vertically in alternate columns in single letter code.

[b]The substitution matrix is placed at any free location on the spreadsheet. The numerical values denote the degree of conservation for amino acid exchanges. X identifies gaps in a sequence. The last column (**N**) contains a unique numerical identifier for each amino acid.

[c]The first amino acid is assigned a numerical identifier from the matrix using the formula: @VLOOKUP(A2,I1..N5,5), which directs the program to the location of the matrix and the column number containing the numerical amino acid identifiers. This formula is then copied down beside each sequence.

[d]A conservation score is calculated for each position in the sequence as the sum of all possible pairwise comparisons. For the first position in this example the formula is: @VLOOKUP(A2,I1..N5,D2) + @VLOOKUP(A2,I1..N5,F2) + @VLOOKUP(C2,I1..N5,F2). This formula is then copied down beside each sequence. The number of combinations required for n species is given by: $\sum_{x=1}^{n-1} (n - x)$

	A	B	C	D	E	F	G
1	{FOR COUNT,1,SEQ,1,\A}					SEQ =	376
2	/		/		/	COUNT =	7
3	C		C		C		
4	{UP}		{UP}		{UP}		
5	.		{UP}		{UP}		
6	.		.		{UP}		
7	{DOWN}		.		.		
8	~		{DOWN}		.		
9	{RIGHT}		{DOWN}		{DOWN}		
10	{RIGHT}		~		{DOWN}		
11	{UP}		{RIGHT}		{DOWN}		
12	~		{RIGHT}		~		
13	{RIGHT}		{UP}		{RIGHT}		
14	{RIGHT}		{UP}		{RIGHT}		
15	{UP}		~		{UP}		
16	/		{RIGHT}		{UP}		
17	D		{RIGHT}		{UP}		
18	S		{UP}		~		
19	R		{UP}		{RIGHT}		
20	D		/		{RIGHT}		
21	.		D		{UP}		
22	{END}		S		{UP}		
23	{DOWN}		R		{UP}		
24	~		D		/		
25	P		.		D		
26	~		{END}		S		
27	~		{DOWN}		R		
28	G		~		D		
29	{DOWN}		P		.		
30	/		~		{END}		
31	C		~		{DOWN}		
32	~		G		~		
33	{LEFT}		{DOWN}		P		
34	~		{DOWN}		~		
35	{UP}		/		~		
36	/		C		G		
37	R		~		{DOWN}		
38	E		{LEFT}		{DOWN}		
39	{END}		~		{DOWN}		
40	{DOWN}		{UP}		/		
41	~		{UP}		C		
42	{DOWN}		/		~		
43	{DOWN}		R		{LEFT}		
44	{LEFT}		E		~		
45	{LEFT}		{END}		{UP}		
46			{DOWN}		{UP}		
47			~		{UP}		
48			{DOWN}		/		
49			{DOWN}		R		
50			{DOWN}		E		
51			{LEFT}		{END}		
52			{LEFT}		{DOWN}		
53					~		
54					{DOWN}		
55					{DOWN}		
56					{DOWN}		
57					{DOWN}		
58					{LEFT}		
59					{LEFT}		

```
MEDIAN SIEVE PROGRAM FOR LOTUS 123


ENTER RAW SCORE VALUES IN AN EMPTY COLUMN

ADD TAIL OF TEN "0"s TO END OF SCORE RANGE

ENTER SEQUENCE LENGTH IN CELL G1

EDIT MACRO {FOR} STATEMENT TO SELECT MESH 1:
      MESH 1: {FOR COUNT,1,SEQ,1,\A}
      MESH 2: {FOR COUNT,1,SEQ,1,\B}
      MESH 3: {FOR COUNT,1,SEQ,1,\C}

POSITION INDICATOR AT SECOND VALUE IN SCORE RANGE

TO START SIEVING: PRESS ALT-X

WHEN FIRST MESH IS COMPLETE, REPEAT WITH NEXT MESH SIZE:

ADD TAIL OF TEN "0"s TO END OF NEWLY CREATED MESH SCORE RANGE
EDIT MACRO {FOR} STATEMENT TO SELECT NEXT LARGEST MESH SIZE
FOR MESH 2, POSITION INDICATOR AT THIRD VALUE IN MESH 1 SCORE RANGE
FOR MESH 3, POSITION INDICATOR AT FORTH VALUE IN MESH 2 SCORE RANGE

TO STOP SIEVING AT ANY TIME: PRESS CTRL-BREAK, THEN ESC
```

REFERENCES

1. Kimura, M. 1968. Evolutionary rate at the molecular level. *Nature.* **217:** 624–626.
2. Kimura, M. 1980. A simple method for estimating evolutionary rates of base substitutions through comparative studies of nucleotide sequences. *J. Mol. Evol.* **16:** 111–120.
3. Kimura, M. 1987. Molecular evolutionary clock and the neutral theory. *J. Mol. Evol.* **26:** 24–33.
4. Ohta, T. 1991. Multigene families and the evolution of complexity. *J. Mol. Evol.* **33:** 34–41.
5. Luo, D-C., W-H. Li, M. N. Moore, and L. Chan. 1986. Structure and evolution of the apolipoprotein multigene family. *J. Mol. Biol.* **187:** 325–340.
6. Mahley, R. W., T. L. Innerarity, S. C. Rall, and K. H. Weisgraber. 1984. Plasma lipoproteins: apolipoprotein structure and function. *J. Lipid Res.* **25:** 1277–1294.
7. Hixson, J. E., C. M. Kammerer, G. E. Mott, M. L. Britten, S. Birnbaum, P. K. Powers, and J. L. VandeBerg. 1993. Baboon apolipoprotein A-IV. Identification of Lys(76)→Glu that distinguishes two common isoforms and detection of length polymorphisms at the carboxyl terminus. *J. Biol. Chem.* **268:** 15667–15673.
8. MacLachlan, A. D. 1972. Repeating sequences and gene duplication in proteins. *J. Mol. Biol.* **64:** 417–437.
9. Bangham, J. A. 1988. Data-sieving hydrophobicity plots. *Anal. Biochem.* **174:** 142–145.
10. Segrest, J. P., M. K. Jones, H. De Loof, C. G. Brouillette, Y. V. Venkatachalapathi, and G. M. Anatharamaiah. 1992. The amphipathic helix in the exchangeable apolipoproteins: a review of secondary structure and function. *J. Lipid Res.* **33:** 141–166.
11. Miller, G. J. 1984. Epidemiological and clinical aspects of high density lipoproteins. *In* Clinical and Metabolic Aspects of High Density Lipoproteins, Miller, N. E., and G. J. Miller, editors. Elsevier, Amsterdam. 47–74.
12. Fielding, C. J., V. G. Shore, and P. E. Fielding. 1972. A protein cofactor of lecithin cholesterol acyltransferase. *Biochem. Biophys. Res. Commun.* **46:** 1493–1498.
13. Graham, D. L., and J. F. Oram. 1987. Identification and characterization of a high density lipoprotein-binding protein in a cell membrane by ligand blotting. *J. Biol. Chem.* **262:** 7439–7442.
14. Von Eckardstein, A., A. H. Funke, M. Walter, K. Altland, A. Benninghoben, and G. Assman. 1990. Structural analysis of human apolipoprotein A-I variants: amino acid substitutions are nonrandomly distributed throughout the apolipoprotein A-I primary structure. *J. Biol. Chem.* **265:** 8610–8617.
15. Marcel, Y. L., P. R. Provost, H. Koa, E. Raffai, N. V. Dac, J-C. Fruchart, and E. Rassart. 1991. The epitopes of apolipoprotein A-I define distinct structural domains including a mobile middle region. *J. Biol. Chem.* **266:** 3644–3653.
16. Anantharamaiah, G. M., Y. V. Venkatachalapathi, C. G. Brouillete, and J. P. Segrest. 1990. Use of synthetic peptide analogues to localize lecithin:cholesterol acyltransferase activating domain in apolipoprotein A-I. *Arteriosclerosis.* **10:** 95–105.
17. Banka, C. L., D. J. Bonnet, A. S. Black, R. S. Smith, and L. K. Curtiss. 1991. Localization of an apolipoprotein A-I epitope critical for activation of lecithin:cholesterol acyltransferase. *J. Biol. Chem.* **266:** 23886–23892.
18. Allan, C. M., N. H. Fidge, and J. Kanellos. 1992. Antibodies to the carboxyl terminus of human apolipoprotein A-I. The putative cellular binding domain of high density lipoprotein 3 and carboxyl terminal structural homology between apolipoproteins A-I and A-II. *J. Biol. Chem.* **267:** 13257–13261.
19. Nolte, R. T., and D. Atkinson. 1992. Conformational analysis of apolipoprotein A-I and E-3 based on primary sequence and circular dichroism. *Biophys. J.* **63:** 1221–1239.
20. Rall, S. C., K. H. Weisgraber, R. W. Mahley, Y. Ogawa, C. J. Fielding, G. Utermann, J. Haas, A. Steinmetz, H. J. Menzel, and G. Assmann. 1984. Abnormal lecithin:cholesterol acyltransferase activation by a human apolipoprotein apoA-I variant in which a single lysine residue is deleted. *J. Biol. Chem.* **259:** 10063–10070.
21. Minnich, A., X. Collet, A. Roghani, C. Cladara, R. L. Hamilton, C. J. Fielding, and V. I. Zannis. 1992. Site-directed mutagenesis and structure–function analysis of the human apolipoprotein A-I. *J. Biol. Chem.* **267:** 16553–16560.
22. Lefevre, M., and P. S. Roheim. 1984. Metabolism of apolipoprotein A-IV. *J. Lipid Res.* **25:** 1603–1610.
23. Weinberg, R. B. 19787. Differences in the hydrophobic properties of discrete alpha helical domains of rat and human apolipoprotein A-IV. *Biochim. Biophys. Acta.* **918:** 299–303.
24. Weinberg, R. B., and M. S. Spector. 1985. Human apolipoprotein A-IV: displacement from the surface of triglyceride-rich particles by HDL_2-associated C-apoproteins. *J. Lipid Res.* **26:** 26–37.
25. Weinberg, R. B., J. A. Ibdah, and M. C. Phillips. 1992. Adsorption of apolipoprotein A-IV to phospholipid monolayers spread at the air/water interface. *J. Biol. Chem.* **267:** 8977–8983.
26. Green, P. H., R. M. Glickman, J. W. Riley, and E. Quinet. 1980. Human apolipoprotein A-IV: intestinal origin and distribution in plasma. *J. Clin. Invest.* **65:** 911–919.
27. Savion, N., and A. Gamliel. 1988. Binding of apolipoprotein A-IV and apolipoprotein A-IV to cultured bovine aortic endothelial cells. *Arteriosclerosis.* **8:** 178–186.
28. Weinberg, R. B., and C. Patton. 1990. Binding of human apolipoprotein A-IV to human hepatocellular plasma membranes. *Biochim. Biophys. Acta.* **1044:** 255–261.
29. Steinmetz, A., R. Barbaras, N. Ghalim, V. Clavey, J. C. Fruchart, and G. Ailhaud. 1990. Human apolipoprotein A-IV binds to apolipoprotein A-I/A-II receptor sites and promotes cholesterol efflux from adipose cells. *J. Biol. Chem.* **265:** 7859–7863.
30. Bielicki, J. K., W. J. Johnson, R. B. Weinberg, J. M. Glick, and G. H. Rothblat. 1992. Efflux of lipid from fibroblasts to apolipoproteins: dependence on elevated levels of cellular unesterified cholesterol. *J. Lipid Res.* **33:** 1699–1709.
31. Steinmetz, A., and G. Utermann. 1985. Activation of lecithin:cholesterol acyltransferase by human apolipoprotein A-IV. *J. Biol. Chem.* **260:** 2258–2264.
32. Rye, K. A., K. H. Garrety, and P. J. Barter. 1992. Changes in the size of reconstituted high density lipoproteins during incubation with cholesteryl ester transfer protein: the role of apolipoproteins. *J. Lipid Res.* **33:** 215–224.
33. Boguski, M. S., M. Freeman, N. A. Elshourbagy, J. M. Taylor, and J. I. Gordon. 1986. On computer-assisted analysis of biological sequences: proline punctuation, consensus sequences, and apolipoprotein repeats. *J. Lipid Res.* **27:** 1011–1034.
34. Jonas, A., A. Steinmetz, and L. Churgay. 1993. The number of amphipathic alpha-helical segments of apolipoproteins A-I, E, and A-IV determines the size and functional properties of their reconstituted lipoprotein particles. *J. Biol. Chem.* **268:** 1596–1602.

35. Jonas, A., J. H. Wald, K. L. Toohill, E. S. Krul, and K. E. Kezdy. 1990. Apolipoprotein A-I structure and lipid properties in homogeneous, reconstituted spherical and discoidal high density lipoproteins. *J. Biol. Chem.* **265:** 22123-22129.
36. Emmanuel, F., A. Steinmetz, N. Theret, N. Gosselet, F. Attenot, S. Seguret, M. Latta, J. C. Fruchart, and P. P. Denefle. 1993. Studies on the structure-function relationships of human apolipoprotein A-IV: involvement of a specific helicoidal domain in LCAT activation. *Circulation.* **88** (4 part 2): I462.
37. Weinberg, R. B., and M. S. Spector. 1985. The self-association of human apolipoprotein A-IV: evidence for an in vivo circulating dimeric form. *J. Biol. Chem.* **260:** 14279-14286.
38. Reue, K., and T. Leete. 1991. Genetic variation in mouse apolipoprotein A-IV due to insertion and deletion in a region of tandem repeats. *J. Biol. Chem.* **266:** 12715-12721.
39. Lohse, P., M. R. Kindt, D. J. Rader, and H. B. Brewer. 1990. Genetic polymorphism of human plasma apolipoprotein A-IV is due to nucleotide substitutions in the apolipoprotein A-IV gene. *J. Biol. Chem.* **265:** 10061-10064.
40. Lohse, P., M. R. Kindt, D. J. Rader, and H. B. Brewer. 1990. Human plasma apolipoproteins A-IV-0 and A-IV-3: molecular basis for two rare variants of apolipoprotein A-IV-1. *J. Biol. Chem.* **265:** 12734-12739.
41. Weinberg, R. B., M. Jordan, and A. Steinmetz. 1990. Distinctive structure and function of human apolipoprotein variant, apoA-IV-2. *J. Biol. Chem.* **265:** 18372-18378.
42. Williams, S. C., S. G. Grant, K. Reue, B. Carrasquillo, A. J. Lusis, and A. J. Kinniburgh. 1989. *Cis*-acting determinants of basal and lipid-regulated apolipoprotein A-IV expression in mice. *J. Biol. Chem.* **264:** 19009-19016.
43. Menzel, H. J., E. Boerwinkle, S. Schrangl-Will, and G. Utermann. 1988. Human apolipoprotein A-IV polymorphism: frequency and effect on lipid and lipoprotein levels. *Hum. Genet.* **79:** 368-372.
44. McCombs, R. J., D. E. Marcadis, and R. B. Weinberg. 1993. Apolipoprotein A-IV-1/2 heterozygotes are hyporesponders to a high cholesterol diet. *Circulation.* **88** (4 part 2): I317.
45. Soutar, A. K., C. W. Garner, H. N. Baker, J. T. Sparrow, R. L. Jackson, A. M. Gotto, and L. C. Smith. 1975. Effect of the human plasma apolipoproteins and phosphatidylcholine acyl donor on the activity of lecithin:cholesterol acyltransferase. *Biochemistry.* **14:** 3057-3064.
46. Kinnunen, P. K., and C. Ehnholm. 1976. Effect of serum and C-apoproteins from very low density lipoproteins on human postheparin plasma hepatic lipase. *FEBS Lett.* **65:** 354-357.
47. Weisgraber, K. H., R. W. Mahley, R. C. Kowal, J. Herz, J. L. Goldstein, and M. S. Brown. 1990. Apolipoprotein C-I modulates the interaction of apolipoprotein E with beta-migrating very low density lipoproteins (beta-VLDL) and inhibits binding of beta-VLDL to low density lipoprotein receptor-related protein. *J. Biol. Chem.* **265:** 22453-22459.
48. Sparrow, J. T., and A. M. Gotto. 1980. Phospholipid binding studies with synthetic apolipoprotein fragments. *Ann. NY Acad. Sci.* **348:** 187-211.
49. LaRosa, J. C., R. I. Levy, R. Herbert, S. E. Lux, and D. S. Fredrickson. 1970. A specific apoprotein activator of lipoprotein lipase. *Biochem. Biophys. Res. Commun.* **41:** 57-62.
50. Kinnunen, W. C., R. L. Jackson, L. C. Smith, A. M. Gotto, and J. T. Sparrow. 1977. Activation of lipoprotein lipase by native and synthetic fragments of apoC-II. *Proc. Natl. Acad. Sci. USA.* **74:** 4348-4851.
51. Musliner, T. A., F. C. Church, P. N. Herbert, and N. J. Kingston. 1977. Lipoprotein lipase cofactor activity of a carboxyl terminal peptide of apolipoprotein C-II. *Biochim. Biophys. Acta.* **573:** 501-509.
52. Clarke, A. R., and J. J. Holbrooke. 1985. The mechanism of activation of lipoprotein lipase by apolipoprotein C-II. The formation of a protein-protein complex in free solution and at a triacylglycerol-water interface. *Biochim. Biophys. Acta.* **827:** 358-368.
53. Catapano, A. L. 1980. The distribution of apoC-II and apoC-III in very low density lipoproteins of normal and type IV subjects. *Atherosclerosis.* **35:** 419-424.
54. Wang, C. S., W. J. McConathy, H. U. Kloer, and P. Alaupovic. 1985. Modulation of lipoprotein lipase activity by apolipoproteins. Effect of apolipoprotein C-III. *J. Clin. Invest.* **75:** 384-390.
55. Ginsberg, H. N., N. A. Le, I. J. Goldberg, J. C. Gibson, A. Rubinstein, P. Wang-Iverson, R. Norum, and W. V. Brown. 1986. Apolipoprotein B metabolism in subjects with deficiency of apolipoproteins C-III and A-I. Evidence that apolipoprotein C-III inhibits catabolism of triglyceride-rich lipoproteins by lipoprotein lipase in vivo. *J. Clin. Invest.* **78:** 1287-1295.
56. Ito, Y., N. Azrolan, A. OConnell, A. Walsh, and J. L. Breslow. 1990. Hypertriglyceridemia as a result of human apoC-III gene expression in transgenic mice. *Science.* **249:** 790-793.
57. Windler, E., Y. Chao, and R. J. Havel. 1980. Regulation of the hepatic uptake of triglyceride-rich lipoproteins in the rat: opposing effects of homologous apolipoprotein E and the individual C apoproteins. *J. Biol. Chem.* **255:** 8303-8307.
58. Sparrow, J. T., H. J. Pownall, F. J. Hsu, L. D. Blumenthal, A. R. Culwell, and A. M. Gotto. 1977. Lipid binding by fragments of apolipoprotein C-III-1 obtained by thrombin cleavage. *Biochemistry.* **16:** 5427-5431.
59. Catapano, A. L. 1987. Activation of lipoprotein lipase by apolipoprotein C-II is modulated by the COOH terminal region of apolipoprotein C-III. *Chem. Phys. Lipids.* **45:** 39-47.
60. Maeda, H., R. K. Hashimoto, T. Ogura, S. Hiraga, and H. Uzawa. 1987. Molecular cloning of a human apoC-III variant: Thr 74 → Ala 74 mutation prevents O-glycosylation. *J. Lipid Res.* **28:** 1405-1409.
61. Roghani, A., and V. I. Zannis. 1988. Mutagenesis of the glycosylation site of human apoC-III. O-linked glycosylation is not required for apoC-III secretion and lipid binding. *J. Biol. Chem.* **263:** 17925-17932.
62. McConathy, W. J., and P. Alaupovic. 1976. Studies on the isolation and partial characterization of apolipoprotein D and lipoprotein D of human plasma. *Biochemistry.* **15:** 515-520.
63. Boyles, J. K., L. M. Notterpek, M. R. Wardell, and S. C. Rall. 1990. Identification, characterization, and tissue distribution of apolipoprotein D in the rat. *J. Lipid Res.* **31:** 2243-2256.
64. Drayna, D., C. Fielding, J. McLean, B. Baer, G. Castro, E. Chen, L. Comstock, W. Henzel, W. Kohr, K. Wion, and R. Lawn. 1986. Cloning and expression of human apolipoprotein D cDNA. *J. Biol. Chem.* **261:** 16535-16539.
65. Drayna, D. T., J. W. McLean, K. L. Wion, J. M. Trent, H. A. Drabkin, and R. M. Lawn. 1987. Human apolipoprotein D gene: gene sequence, chromosomal location, and

homology to the alpha-2μ-globulin superfamily. *DNA.* **6:** 199–204.
66. Holden, H. M., W. R. Rypniewski, J. H. Law, and I. Rayment. 1987. The molecular structure of insecticyanin from the tobacco hornworm *Manduca sexta L.* at 2.6 A resolution. *EMBO J.* **6:** 1565–1570.
67. Huber, R., M. Schneider, O. Epp, I. Mayr, A. Messerschmidt, J. Pfulgrath, and H. Kayser. 1987. Crystallization, crystal structure analysis, and preliminary molecular model of the bilin-binding protein from the insect *Pieris brassicae.* *J. Mol. Biol.* **195:** 423–434.
68. Cowan, S. W., M. E. Newcomer, and T. A. Jones. 1990. Crystallographic refinement of human serum retinol binding protein at 2 Å resolution. *Proteins.* **8:** 44–61.
69. Mahley, R. W. 1988. Apolipoprotein E: cholesterol transport protein with expanding role in cell biology. *Science.* **240:** 622–630.
70. Mahley, R. W., D. Y. Hui, T. L. Innerarity, and K. H. Weisgraber. 1981. Two independent lipoprotein receptors on hepatic membranes of the dog, swine, and man. *J. Clin. Invest.* **68:** 1197–1206.
71. Koo, C., T. L. Innerarity, and R. W. Mahley. 1985. Obligatory role of cholesterol and apolipoprotein E in the formation of large cholesterol-enriched and receptor-active high density lipoproteins. *J. Biol. Chem.* **260:** 11934–19943.
72. Mahley, R. W., K. H. Weisgraber, and T. L. Innerarity. 1979. Interactions of plasma lipoproteins containing B and E apolipoproteins with heparin and cell surface receptors. *Biochim. Biophys. Acta.* **575:** 81–89.
73. Innerarity, T. L., E. J. Friedlander, S. C. Rall, K. H. Weisgraber, and R. W. Mahley. 1983. The receptor-binding domain of human apolipoprotein E. *J. Biol. Chem.* **258:** 12341–12347.
74. Wilson, C., M. R. Wardell, K. H. Weisgraber, R. W. Mahley, and D. A. Agard. 1991. The three-dimensional structure of the LDL receptor binding domain of human apolipoprotein E. *Science.* **252:** 1817–1822.
75. Aggerbeck, L. P., J. R. Wetterau, K. H. Weisgraber, C. C. Wu, and F. T. Lindgren. 1988. Human apolipoprotein E3 in aqueous solution: evidence for two structural domains. *J. Biol. Chem.* **263:** 6249–6258.
76. Mahley, R. W., and T. L. Innerarity. 1983. Lipoprotein receptors and cholesterol homestasis. *Biochim. Biophys. Acta.* **737:** 197–222.
77. Weisgraber, K. H., S. C. Rall, R. W. Mahley, R. W. Milane, Y. L. Marcel, and J. T. Sparrow. 1986. Human apolipoprotein E: determination of the heparin binding sites of apolipoprotein E3. *J. Biol. Chem.* **261:** 2068–2076.