Apolipoprotein B-100: conservation of lipid-associating amphipathic secondary structural motifs in nine species of vertebrates

Jere P. Segrest,[1,*] Martin K. Jones,[*] Vinod K. Mishra,[*] Vincenzo Pierotti,[†] Stephen H. Young,[†] Jan Borén,[†] Thomas L. Innerarity,[†] and Nassrin Dashti[§]

Departments of Medicine and Biochemistry and the Atherosclerosis Research Unit,* UAB Medical Center, Birmingham, AL 35294-0012; The Gladstone Institute of Cardiovascular Disease,† San Francisco, CA 94141-9100; and Departments of Nutrition Sciences and Pediatrics,§ UAB Medical Center, Birmingham, AL 35294-0011

Abstract Development of a computer program called LOCATE allowed us to show that human apolipoprotein B-100 is composed of five domains, NH_2-α_1-β_1-α_2-β_2-α_3-COOH, enriched, alternately, in amphipathic α helixes and amphipathic β strands. Using updated versions of this program, here we compare the complete sequence of human apolipoprotein B-100 with partial sequences from eight additional species of vertebrates (chicken, frog, hamster, monkey, mouse, pig, rat, and rabbit). The lipid-associating amphipathic α helixes cluster in domains α_2 (between residues 2075 ± 25 and 2575 ± 25) and α_3 (between residues 4100 ± 100 and 4550 ± 50) in all species for which those regions have been sequenced but with little conservation of individual helixes. Lipid-associating amphipathic β strands cluster in domains β_1 (approximately residues 827-2000) and β_2 (approximately residue 2571 to residue 4000 ± 50) in all species for which these regions have been sequenced, with conservation of several individual amphipathic β strands. Hydrophobic segments are present in apolipoprotein B-100 sequences of all nine species but the frequency of occurrence is no greater than generally found in β sheet-containing proteins. We conclude that four alternating lipid-associating domains, -β_1-α_2-β_2-α_3-COOH, are common supramolecular features of apolipoprotein B-100 in nine vertebrate species.—**Segrest, J. P., M. K. Jones, V. K. Mishra, V. Pierotti, S. H. Young, J. Borén, T. L. Innerarity, and N. Dashti.** Apolipoprotein B-100: conservation of lipid-associating amphipathic secondary structural motifs in nine species of vertebrates. *J. Lipid Res.* 1998. **39:** 85–102.

Supplementary key words amphipathic α helixes • amphipathic β strands • hydrophobic amino acid sequences • plasma lipoproteins • computer analysis of amphipathic motifs • low density lipoprotein • protein homology • computer program LOCATE • amphipathic domains • protein–lipid interactions

Plasma apolipoproteins can be grouped into two general classes, the nonexchangeable apolipoproteins (apoB-100 and apoB-48), and the exchangeable apolipoproteins (apoA-I, apoA-II, apoA-IV, apoC-I, apoC-II, apoC-III, and apoE). ApoB-100 is highly insoluble in aqueous solution and is also one of the largest monomeric proteins known. Because of its size and insoluble nature, it has been difficult to deduce the structural motif(s) responsible for the lipid-associating properties of apoB-100. On the other hand, the exchangeable apolipoproteins are soluble in aqueous solutions, and the secondary structural motif (amphipathic α helix) responsible for their lipid association has been extensively studied (1, 2).

The amphipathic α helix is a common secondary structural motif in biologically active peptides and proteins. In a previous review article (3) we grouped amphipathic helixes into seven distinct classes, with class A representing the lipid-associating amphipathic helical domains of the exchangeable apolipoproteins. A variation on the class A theme is distinguished by a radial clustering of positive and negative residues into a pattern different from that of class A. Because of the presence of positively charged residues at the polar–nonpolar interface, as well as at the center of the polar face, we term this the class Y motif (4).

An additional class of amphipathic helix, termed G*, was also found in the exchangeable apolipoproteins. This class is distinguished by a random radial arrangement of positively and negatively charged residues.

Abbreviations: apo, apolipoprotein; Λ, calculated affinity; α_1, α_2, α_3, amphipathic α helical domains of apolipoprotein B-100; β_1, β_2, amphipathic β strand domains of apolipoprotein B-100.

[1]To whom correspondence should be addressed.

These amphipathic helixes are similar but not identical to the class G amphipathic helixes found in α helical globular proteins. Class G* amphipathic helixes differ from those of class G in having both a greater hydrophobic moment and a greater nonpolar face hydrophobicity. The class G* amphipathic helixes are postulated to prefer protein–protein interactions over protein–lipid interactions (4). The best characterized example of this class is the N-terminal 4-helix bundle globular domain of apoE (5) that consists of four class G* amphipathic helixes (4) and, depending upon local conditions, fold as a 4-helix bundle protein or associate with lipid (5).

To identify lipid-associating domains in apoB-100, we developed a computer program called LOCATE that searches amino acid sequences to identify potential amphipathic structural motifs using sets of rules for α helix and β strand termination (6). Two dense clusters of putative lipid-associating amphipathic helixes were located precisely in the middle and the C-terminal end of apoB-100; a cluster of class G* helixes was located at the N-terminus. The two regions between the three amphipathic helix clusters were found to be highly enriched in putative amphipathic β strands, while the three amphipathic helical domains were largely devoid of this putative lipid-associating motif. We proposed, therefore, that human apoB-100 has a pentapartite structure, NH_2-α_1-β_1-α_2-β_2-α_3-COOH, with α_1 representing a globular domain (6).

Recent enhancements to the program LOCATE have been used to compare the complete sequence of human apoB-100 with partial sequences from eight species of vertebrates (chicken, frog, hamster, monkey, mouse, pig, rat, and rabbit). Based on these analyses, we demonstrate in this report the ubiquitous presence of the pentapartite structure in apoB-100 from phylogenetically diverse vertebrate species.

METHODS

Programs for displaying amphipathic α helixes

The programs WHEEL, COMBO, and CONSENSUS are described in detail elsewhere (7). WHEEL creates a helical wheel diagram of a given sequence of amino acids arranged as an ideal α helix seen down the long axis from the amino terminal end. COMBO superimposes and averages a set of rotationally defined helical wheels. Selected residues of the set are projected onto two wheel diagrams. In the default option, the left-hand and right-hand wheels display the counts of all positively and negatively charged residues, respectively. CONSENSUS superimposes a set of helical wheels in the same fashion as COMBO and a single figure classifies the amino acid residues into five physical–chemical groups: positive, negative, polar, neutral, and hydrophobic.

Programs for locating amphipathic α helixes, amphipathic β strands, or hydrophobic segments in protein sequences

LOCATE finds and displays potentially lipid-associating motifs in an amino acid sequence where the amino acid file contains a sequence of one letter amino acids and the optional free energy of transfer file contains default settings for amino acid residues. Amino Acid Residue File: This file contains a sequence of one letter amino acid residues in the standard encoding. Free Energy of Transfer File: This file is a modification of the GES hydrophobicity scale (8) and contains a free energy of transfer, hydrocarbon to water, in kcal/mol for each amino acid residue used in the amino acid residue file: Arg −12.3; Asp −9.2; Lys −8.8; Glu −8.2; Asn −4.8; Gln −4.1; His −3.0; Pro −0.6; Ser −0.2; Ala 0.0; Gly 0.1; Thr 0.2; Tyr 0.4; Trp 1.9; Cys 2.0; Val 2.6; Leu 2.8; Ile 3.1; Met 3.4; Phe 3.7.

Display options. /DISPLAY_PS automatically runs LOCATE in X-Windows mode and allows a postscript display of selected single amphipathic α helixes, amphipathic β strands, or hydrophobic segments or a postscript display of selected sets of amphipathic α helixes as COMBO or CONSENSUS diagrams. /LINEAR eliminates the hydrophobicity scale and plots the helixes along the same vertical line.

Motif specific options. LOCATE_ALPHA identifies potential amphipathic α helixes within a given amino acid sequence using termination rules described elsewhere (7). /CLASS=list specifies the class of helixes that will be reported. A: helix is determined to be class A; Y: helix is determined to be class Y; G*: helix is anything else. The algorithm for identification of class A and class Y amphipathic helixes is diagrammatically illustrated in **Fig. 1**. /LAMBDA_HYB shows the Λ lipid affinity of each amphipathic helix on the x axis as calculated by the Λ function. /LAMBDA_HYB_CUTOFF specifies the minimum value that the Λ lipid affinity of the nonpolar face of the candidate helix can be and still be selected.

The Λ function for calculating lipid affinity was derived as follows. The snorkeling of basic residues allows for greater penetration of class A amphipathic helixes into the hydrophobic interior of phospholipid monolayers than would otherwise be possible; the greater the angle of the snorkel wedge (largely equivalent to the angle of the hydrophobic face), the greater the lipid penetration (9). Using neutron diffraction, Jacobs and White (10) have measured the gradient that H_2O forms

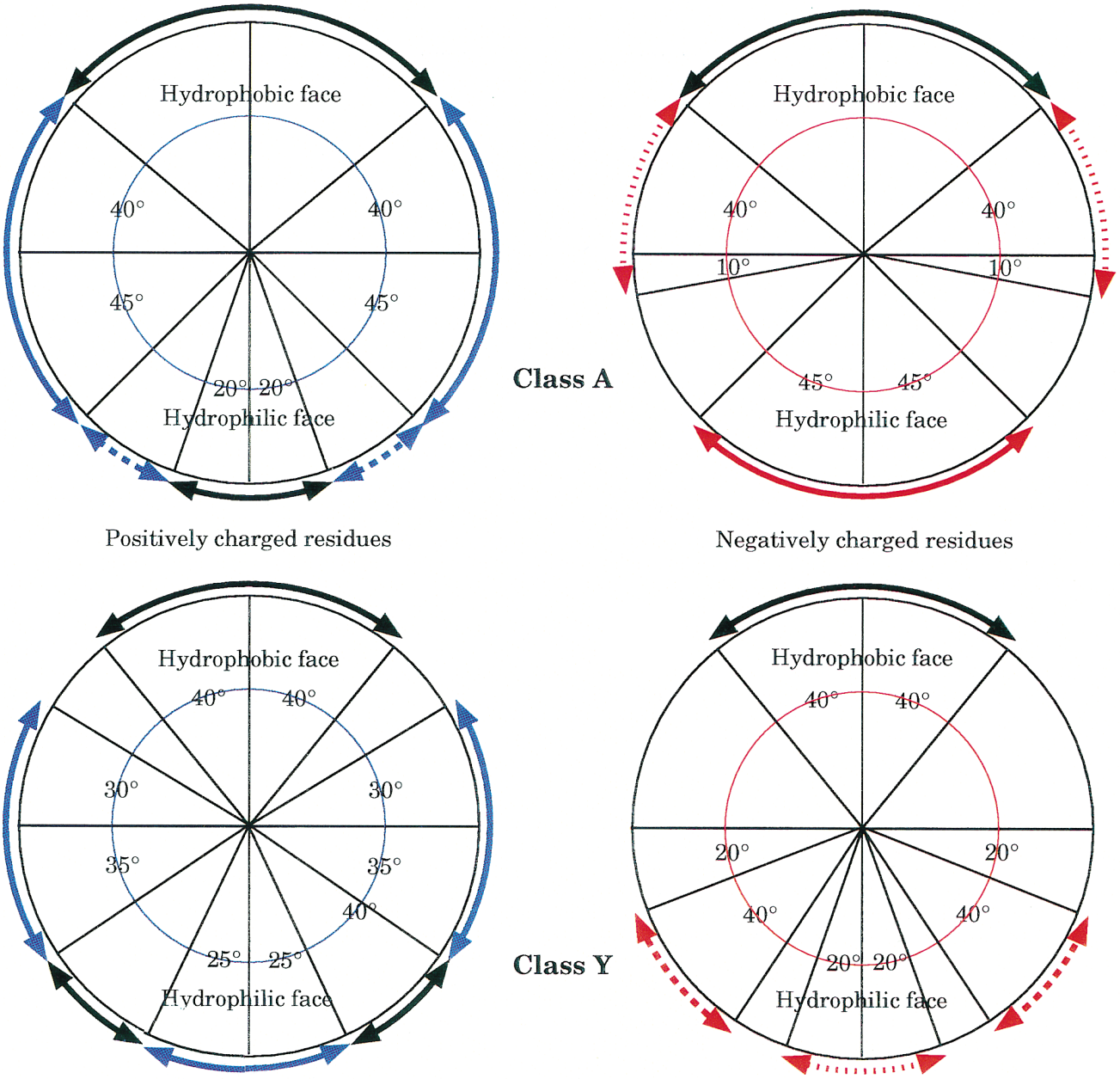


Fig. 1. LOCATE_ALPHA/CLASS algorithm for identifying class A and class Y amphipathic α helixes. Amphipathic helixes are identified using one of two options for radial orientation of the hydrophobic face on a helical wheel projection: the hydrophobic moment or the snorkel algorithm (4, 6). LOCATE_ALPHA/CLASS then uses rules for the allowed radial arrangement of positively charged (left half of the figure) and negatively charged residues (right half of the figure) to identify class A and class Y helixes. In the figure, the hydrophobic face is up and the hydrophilic face is down. The black double-headed arrows indicate the radial positions at which charged residues, either positive or negative, are disallowed. For positively charged residues, the solid blue double-headed arrows indicate the radial positions at which one or more positively charged residues are required and the dashed blue double-headed arrows indicate the radial positions at which a maximum of one positively charged residue is allowed for the left and right sides together. For negatively charged residues, the solid red double-headed arrow indicates the radial position at which two or more negatively charged residues are required, the dashed red double-headed arrows indicate the radial positions at which one or more positively charged residues are required for the left and right sides together, and the dotted red double-headed arrows the radial positions at which a maximum of one negatively charged residue is allowed. Radial positions for which no charged residue rules exist are left blank. A: Class A amphipathic α helixes (upper half). B: Class Y amphipathic α helixes (lower half).

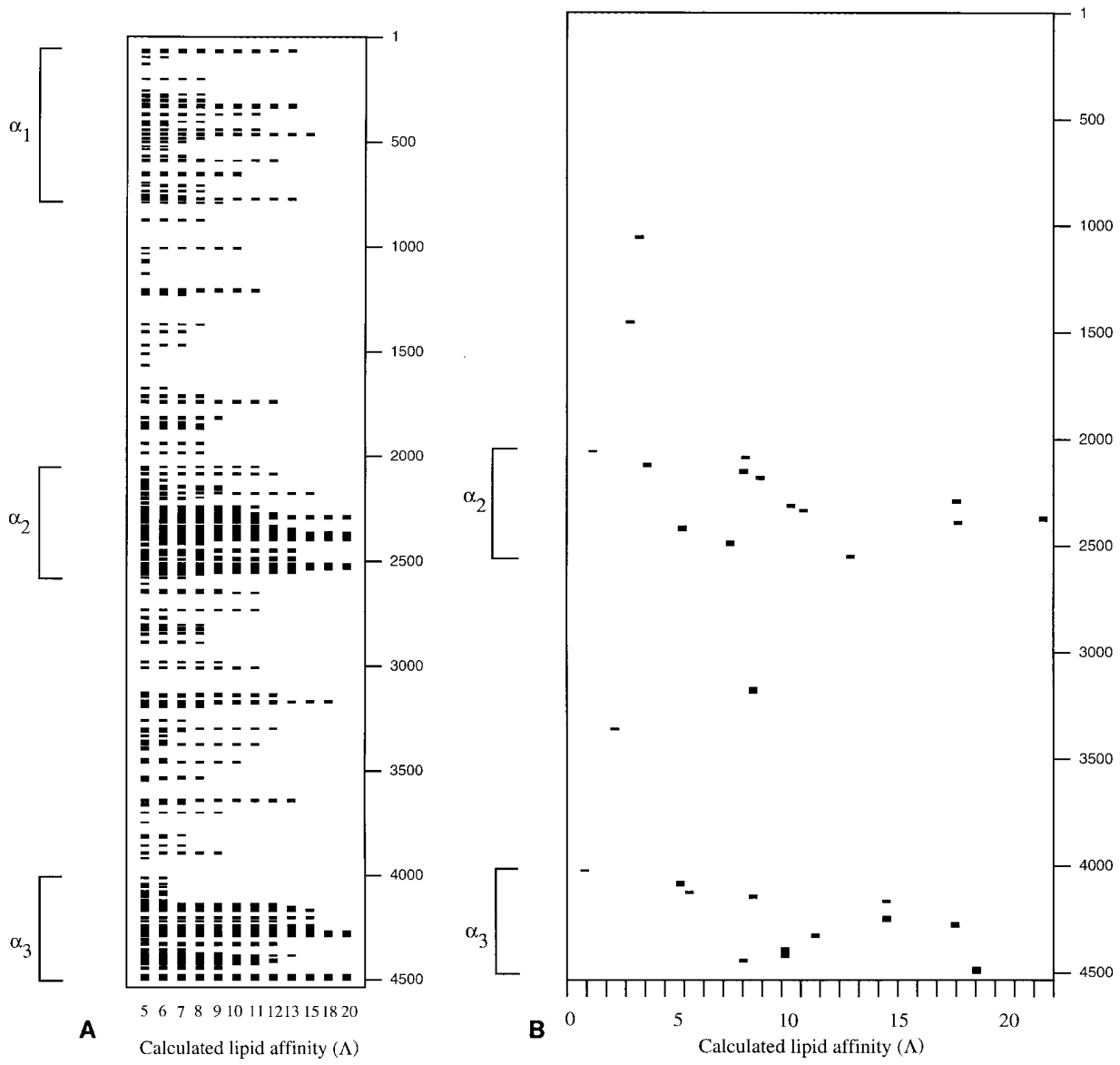


Fig. 2. LOCATE analyses of human apoB-100. A: LOCATE_ALPHA analysis. Plot of all amphipathic α helixes (length ≥10 residues) as the cutoff for Λ (calculated lipid affinity) increases on the X-axis from 5 to 20 kcals/mol. B: LOCATE_ALPHA/CLASS analysis. Plot of all class A and Y amphipathic α helixes (length ≥10 residues) identified by LOCATE (using default parameters). Λ for each helix identified is plotted on the X-axis. C: LOCATE_BETA analysis. Plot of all amphipathic β strands (length ≥8 residues, hydrophobic moment ≥2) as the cutoff for total hydrophobicity of the nonpolar face increases on the X-axis from 5 to 20 kcals/mol. D: LOCATE_HYB analysis. Plot of all hydrophobic segments (length ≥10 residues) as the cutoff for total hydrophobicity increases on the X-axis from 5 to 20 kcals/mol.

from the outside to the inside of a phospholipid monolayer. The hydrocarbon core starts at a depth of approximately 7 Å beneath the center of the phosphatidylcholine head group; at this depth, H_2O has a molar concentration approximately 15% of that at the level of the phosphatidylcholine head groups. As the free energy of the hydrophobic effect decreases with a decrease in concentration of H_2O, the deeper the penetration of an amphipathic helix into the interior of a phospholipid monolayer, the more effective the hydrophobicity (i.e., the lower the free energy) of its nonpolar face. Therefore, the overall lipid affinity of an amphipathic helix will partially depend upon its depth of lipid penetration (9).

Combining the water gradient determined by Jacobs and White (10) with the free energy of transfer of an amino acid residue from H_2O to varying H_2O:organic solvent mixtures, we have derived a free energy gradi-

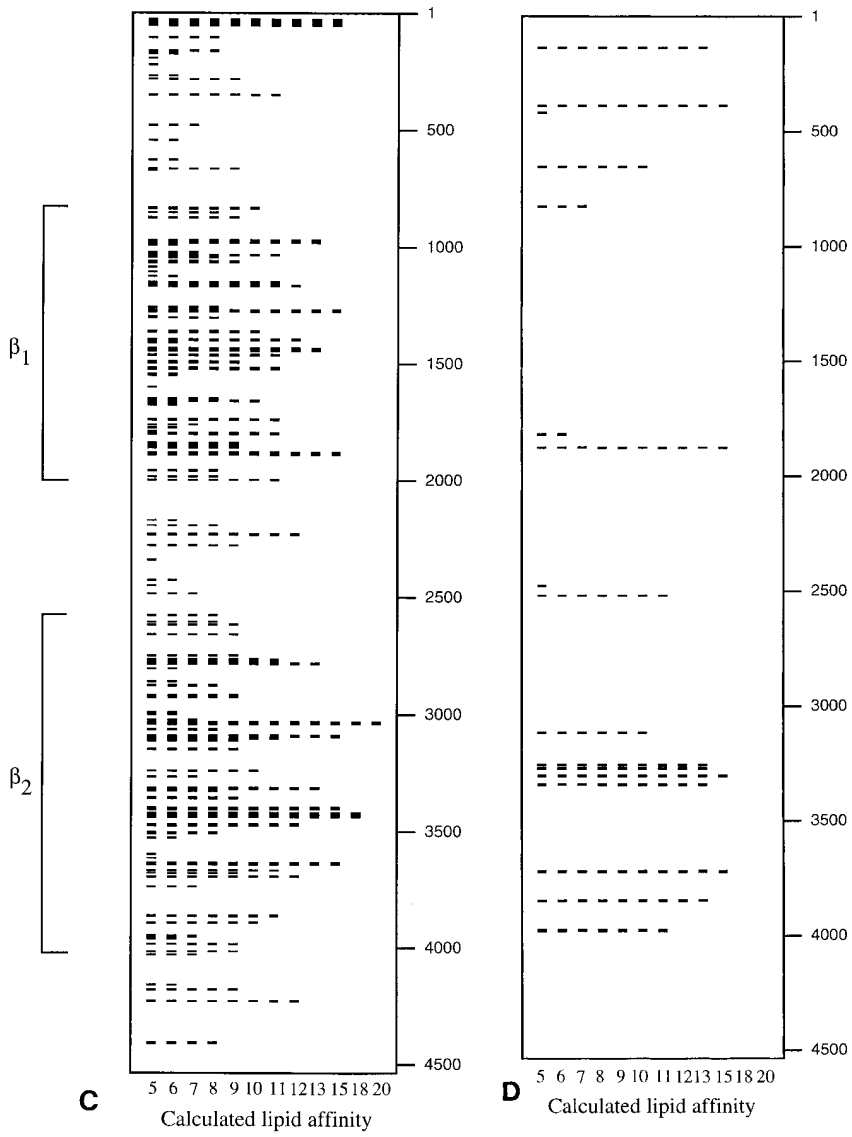


Fig. 2.

ent, δ_i(Å), that is a function of depth of penetration in Å, d_i(Å). Therefore, $\Lambda = \Sigma(\Delta G_{\Psi i} \times \delta_i(Å))$. The method of calculating Λ described in Palgunachari et al. (9) assumes that the lipid surface is always parallel to the helix axis.

A recently developed method of calculating lipid affinity (Λ_3) is based upon a three-dimensional helical cylinder analysis. The program, using the assumptions of the Λ algorithm applied to a three-dimensional model, considers each amphipathic helix to be a cylinder 14 Å in diameter with a pitch of 1.5 Å per residue. The Λ_3 algorithm then determines the orientation of the amphipathic helix relative to the plane of the hydrated phospholipids that produces the maximum lipid affinity. From this orientation, four properties are determined: *a*) angle of rotation of the amphipathic helix relative to the hydrated phospholipid plane, *b*) long axis tilt of the amphipathic helix relative to the hydrated phospholipid plane, *c*) depth of penetration of each residue into the hydrated phospholipid in three dimensions, and *d*) the calculated lipid affinity, Λ_3.

Briefly, the Λ_3 algorithm is based upon the rule that three points determine a plane; here the points are determined by charged residues. For any such plane, all charged residues must lie on the plane or be on one side of the plane, the polar side. Because the maximum value of Λ_3 may not occur for a plane that passes

through three charged residues, the plane is allowed to pivot through two charged residues and even one: the plane orientation is adjusted at 1° increments giving multiple rotations and tilts. As in the previous algorithm, Lys, but not Arg, is allowed to snorkel. The present algorithm provides the following output: *a*) a calculated Λ_3, *b*) a calculated angle of tilt, *c*) a diagrammatic representation of the plane of the hydrated phospholipid relative to the N- and C-terminal end of the amphipathic helix (displayed on the helical wheel output), *d*) depth of penetration (in Å) of each residue, and *e*) highlighting of all residues buried in the hydrated phospholipid (displayed on both the helical wheel and helical net outputs), and *f*) a depth of penetration contour map plotted on the helical net output.

LOCATE_BETA identifies potential amphipathic β strands within a given amino acid sequence using termination rules also described elsewhere (6). LOCATE_HYB identifies segments of hydrophobic residues within a given amino acid sequence using charged and/or polar residues for termination.

General options. These programs create either a linear or a two-dimensional plot. The two-dimensional plot displays the location of each sequence with the desired motif on the Y-axis versus a measure of its hydrophobic moment, hydrophobicity of the hydrophobic face (per residue or total), total hydrophobicity, or calculated lipid affinity (Λ) on the X-axis. Options allow: *1*) selections of amphipathic α helixes (negatively or positively charged) or amphipathic β strands (negatively or positively charged) and *2*) minimum allowable cutoff values for any given parameter (e.g., sequence length, hydrophobicity, lipid affinity, etc.).

/MINIMUM_LENGTH specifies the minimum value that the length of the candidate motif can be and still be selected. /NONPOLAR_HYB_CUTOFF specifies the minimum value that the total hydrophobicity of the nonpolar face of the candidate amphipathic motif can be and still be selected. /TOTAL_HYB shows the total (summed) hydrophobicity of each motif selected on the x axis. /TOTAL_HYB_CUTOFF specifies the minimum value that the total hydrophobicity of the candidate motif can be and still be selected.

Sequence alignments and determination of sequence and structural homologies

Partial sequences of apoB-100 from six species of vertebrates, monkey, pig, rat, rabbit, hamster, and chicken, were downloaded in the FASTA format using the ENTREZ Browser (http://atlas.nlm.nih.gov). Partial sequences of apoB-100 from two additional species, mouse and frog, were determined in the laboratories of co-authors, Stephen H. Young and Thomas L. Innerarity, respectively.

Sequence alignment and homology analyses of the nine apoB-100 sequences were performed on a Macintosh using the program MACAW (11, 12) downloaded from the Internet (ftp://ncbi.nlm.nih.gov/pub/schuler/macaw). To analyze structural (amphipathic and hydrophobic motif) homology, aligned sequences were subjected to LOCATE analysis using cutoff parameters described in the figure legends.

RESULTS

LOCATE analysis of human apoB-100

Before beginning the structural comparisons of apoB-100 from nine vertebrate species, we examined the structural parameters of human apoB-100 predicted by using the most recent version of LOCATE. To aid in the visual identification of clusters of amphipathic motifs with high calculated lipid affinity, each probable amphipathic motif is displayed in the form of a bar graph created by increasing the cutoff incrementally on the X-axis from 5 to 20 kcals/mol for *a*) calculated lipid affinity (Λ) of amphipathic α helixes (see Methods), *b*) the total hydrophobicity of the nonpolar face of amphipathic β strands, and *c*) total hydrophobicity of hydrophobic segments.

LOCATE_ALPHA analyses of amphipathic α helixes as a function of calculated lipid affinity are shown in **Fig. 2A**. Three dense clusters of amphipathic α helixes with high calculated lipid affinity, residues 58–795 (α_1), 2250–2584 (α_2), and 4129–4505 (α_3), are clearly identified. These same general regions were identified as the three amphipathic α helical domains in our original structural analysis of apoB-100 (6). The region 2045–2249 could be considered a portion of α_2 containing a lower density of amphipathic α helixes with weaker lipid affinity.

Figure 2B is a plot of all class A and Y amphipathic helixes identified by LOCATE_ALPHA/CLASS in human apoB-100; Λ for each helix is plotted on the x-axis. Using this option to LOCATE_ALPHA, class A and Y amphipathic α helixes are shown to cluster almost exclusively between residues 2052–2560 (α_2) and 4017–4515 (α_3); no class A or Y amphipathic helixes are found in the α_1 domain. Combining Figs. 2A and 1B, we estimate that in human apoB-100 the α_2 domain is located between residues 2045–2587 and the α_3 domain between residues 4017–4515.

Figure 3 shows COMBO analyses, using LOCATE_ALPHA/CLASS and LOCATE_ALPHA analyses in the X-Windows mode, of the amphipathic α helixes from the α_1, α_2, and α_3 domains of human apoB-100. Plots of

the distribution of positively and negatively charged residues and Leu and Phe residues, respectively, in the class A amphipathic α helixes in the α_2 and α_3 domains (Fig. 2B) are shown in Figs. 3A and B, respectively. These plots illustrate the features common to the class A amphipathic α helixes found in the exchangeable apolipoproteins (2, 4): clustering of the positively charged and negatively charged residues at the polar–nonpolar interfaces and the center of the polar face, respectively (Fig. 3A), a well-defined clustering of hydrophobic residues in the middle of the nonpolar face (Fig. 3B), a high nonpolar face mean hydrophobicity per residue of 2.05 (automatically calculated by both programs) and a mean length of 22.0 ± 9.2 amino acid residues, identical to the length of the tandem repeats found in the exchangeable apolipoproteins.

Analyses of amphipathic α helixes selected on the basis of a calculated high lipid affinity (defined as $\Lambda \geq 7$, third column from left, Fig. 2A), and thus selected independently of amphipathic α helix class, are shown in Figs. 3C–F. Figs. 3C and D are COMBO plots of the distribution of positively and negatively charged residues and Leu and Phe residues, respectively, in the amphipathic α helixes of amphipathic α helixes found in the α_2 plus α_3 domains. These amphipathic α helixes show well-defined clustering of the negatively charged residues in the center of the polar face (Fig. 3C), a well-defined clustering of the negatively charged residues in the center of the polar face (Fig. 3C), a well-defined clustering of hydrophobic residues in the middle of the nonpolar face (Fig. 3D), a high nonpolar face mean hydrophobicity per residue (2.23), and a mean length of 19.5 ± 9.0 amino acid residues; the clustering of the positively charged residues (Fig. 3C) occurs both at the polar–nonpolar interface and in the center of the polar face, presumably because these amphipathic helixes represent a mixture of class A and class Y (2, 4).

Figures 3E and F are COMBO plots of the distribution of positively and negatively charged residues and Leu and Phe residues, respectively, of amphipathic α helixes found in the α_1 domain; the amphipathic α helixes of this domain are distinctly different from those of the α_2 and α_3 domains. The most striking difference is that the positions of the positive and negative charge clusters in the amphipathic α helixes of the α_1 domain are random or perhaps reversed compared to class A amphipathic α helixes (Fig. 3E). Further, mean length is shorter and less variable (15.5 ± 4.4), and the clustering of hydrophobic residues in the middle of the nonpolar face is poorly defined (Fig. 3F), compared to the amphipathic helixes in the α_2 and α_3 domains.

Figure 2C is a LOCATE_BETA bar graph analysis of amphipathic β strands in human apoB-100. Two dense clusters of amphipathic β strands with high calculated lipid affinity can be identified: β_1:827–2001 and β_2:2571–4032. These two domains were noted in our previous LOCATE analysis of apoB-100 (6). In addition, there is a lower density cluster of amphipathic β strands not noted in our previous analysis, residues 24–354 (termed β_N), associated with the N-terminal half of the α_1 domain. Even the C-terminal half of this domain, as well as the α_2 and α_3 domains, contain a few scattered, relatively short amphipathic β strands.

Finally, it has been suggested that unique hydrophobic segments occur in apoB-100 (13). Fig. 2D is a LOCATE_HYB bar graph analysis of hydrophobic segments in human apoB-100 that are ≥10 residues in length with a total hydrophobicity ≥5 to ≤20; a total of 17 sequence intervals with a total hydrophobicity ≥5 were identified. LOCATE_HYB analysis (**Table 1**) of the test chimeric protein sequences, used as controls in our previous LOCATE analysis of apoB-100 (6), indicates that apoB-100, in general, is not unique in possessing short hydrophobic segments when compared to the motif B (β sheet) test protein; proteins rich in β sheets are also relatively rich in short hydrophobic segments. The cluster of four hydrophobic sequences between residues 3250–3347, however, conceivably could be unique. It will be useful, therefore, to see whether this cluster is conserved in the eight apoB-100 fragments from other vertebrate species.

Alignment of the nine apoB-100 sequences

Figure 4 shows the results of the sequence alignment produced by the sequence comparison program MACAW. These results provide two useful pieces of information: sequence alignment for structural homology analysis using the program LOCATE and information concerning sequence identity in aligned regions.

All nine apoB-100 molecules have been sequenced between residues 3139–3409 (i.e., overlap in this region), a sequence stretch that covers a portion of the putative LDL receptor-binding domain in β_2 (13). Six sequences overlap in the region of β_2 between residues 3500–3700. Five sequences overlap in the C-terminal α_3 domain (residues 4176 to the C-terminus). Five sequences overlap in the region of β_2 between residues 2713–3409. Four sequences overlap the region between residues 2436–2713, representing a small portion of the N-terminal end of β_2 and a small portion of the C-terminal end of α_2. Three sequences overlap the entire α_2 domain and a small portion of the C-terminal end of β_1. Only two sequences overlap the region of β_1 between residues 1380–2000. Three sequences overlap residues 400–1000, representing the C-terminal half of the α_1 domain and a very small portion of β_1. Finally, two sequences overlap the first 100 residues of the N-terminal end of the α_1 domain and none between residues 1000–1380 for β_1.

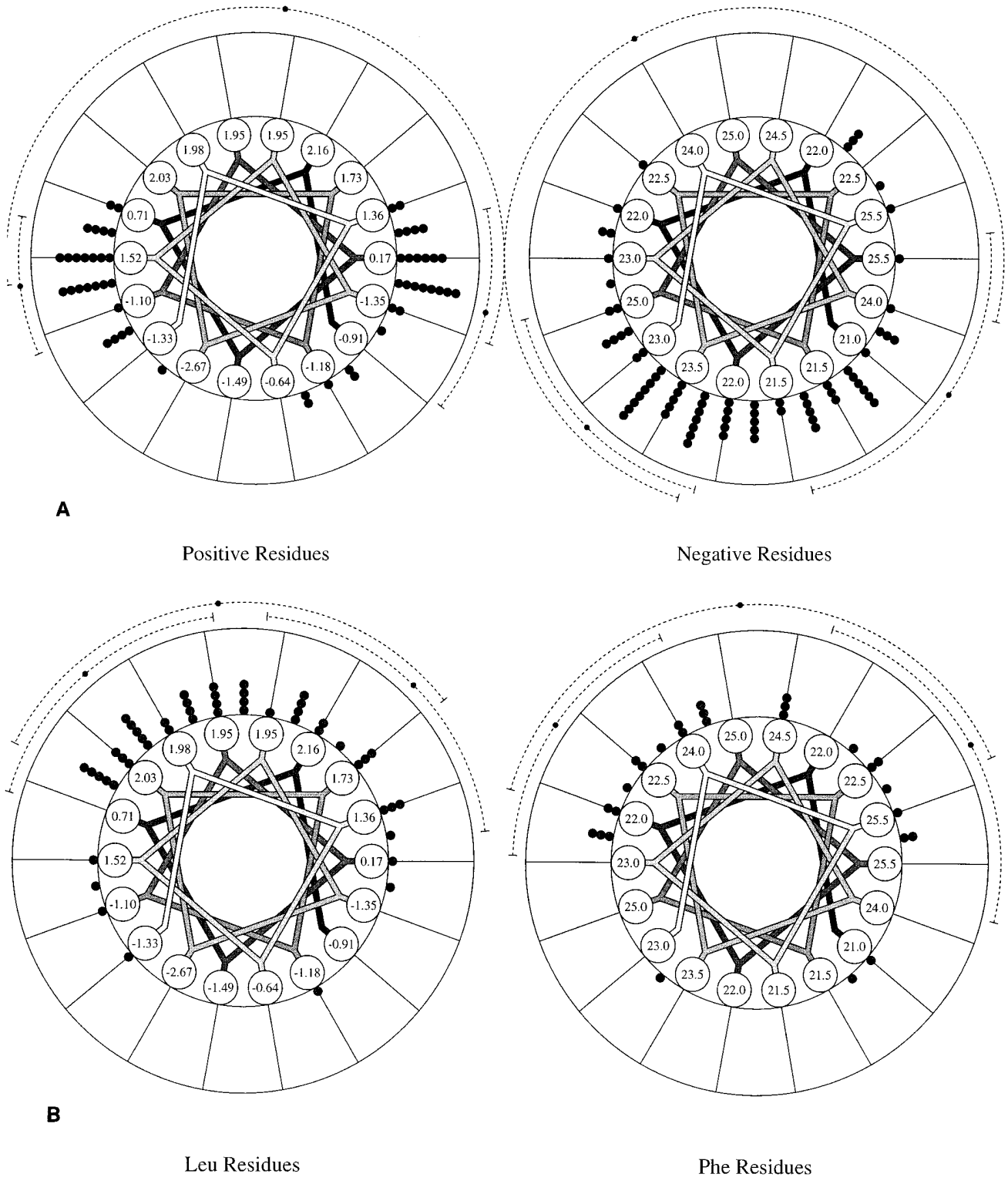


Fig. 3. COMBO analyses of the amphipathic α helixes from the α_1, α_2, and α_3 domains of human apoB-100. COMBO analyses showing the radial distribution of: (A) positively and negatively charged residues and **(B)** Leu and Phe residues in the class A amphipathic α helixes (located uniquely in the α_2 and α_3 domains, Fig. 2B); (C) positively and negatively charged residues and **(D)** Leu and Phe residues of amphipathic α helixes with high calculated high lipid affinity ($\Lambda \geq 7$) located in the α_2 and α_3 domains; **(E)** positively and negatively charged residues and **(F)** Leu and Phe residues of amphipathic α helixes with high calculated high lipid affinity ($\Lambda \geq 7$) located in the α_1 domain. COMBO superimposes and averages a set of wheels. Before the wheels are superimposed, each wheel is rotated so that the hy-

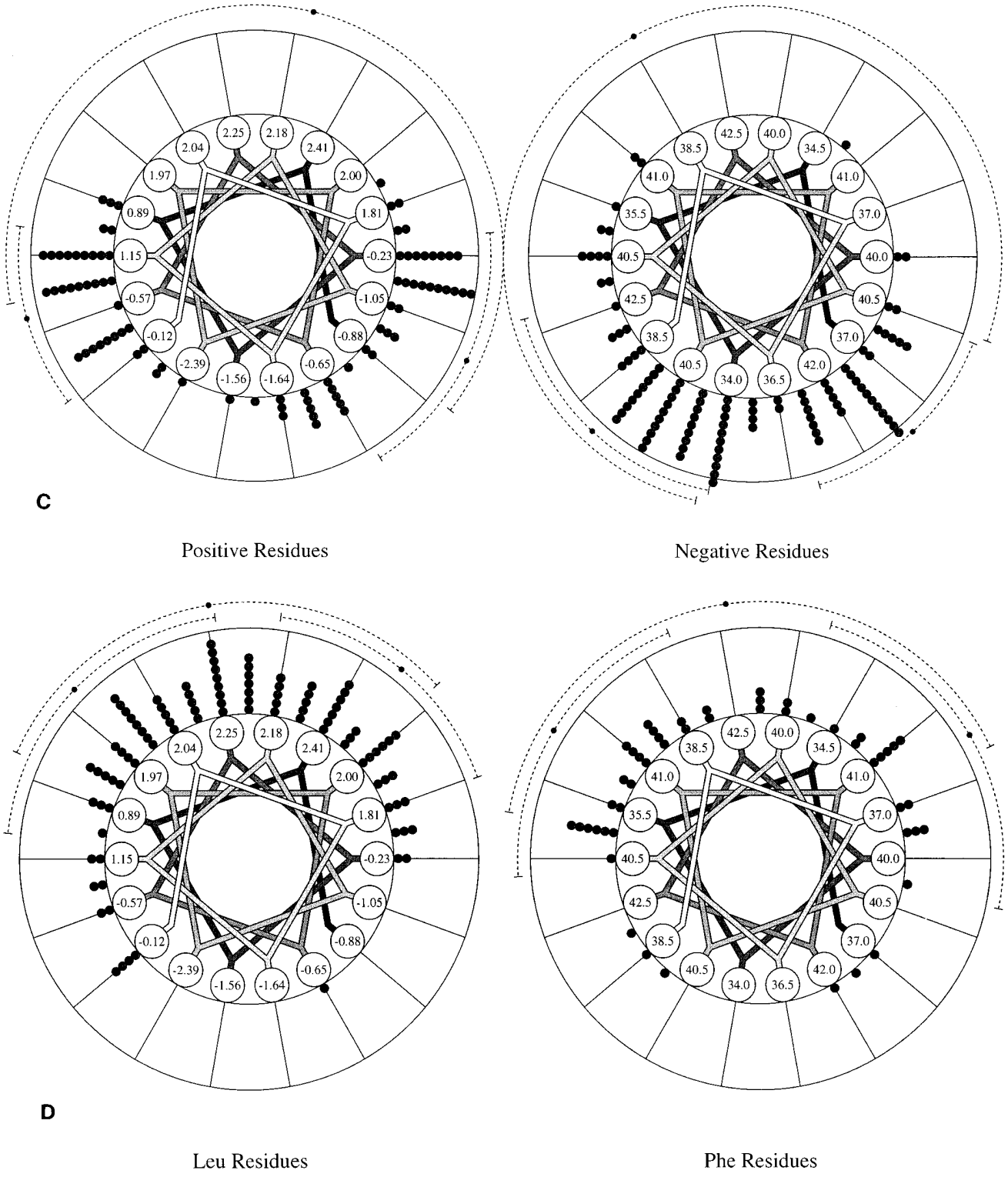
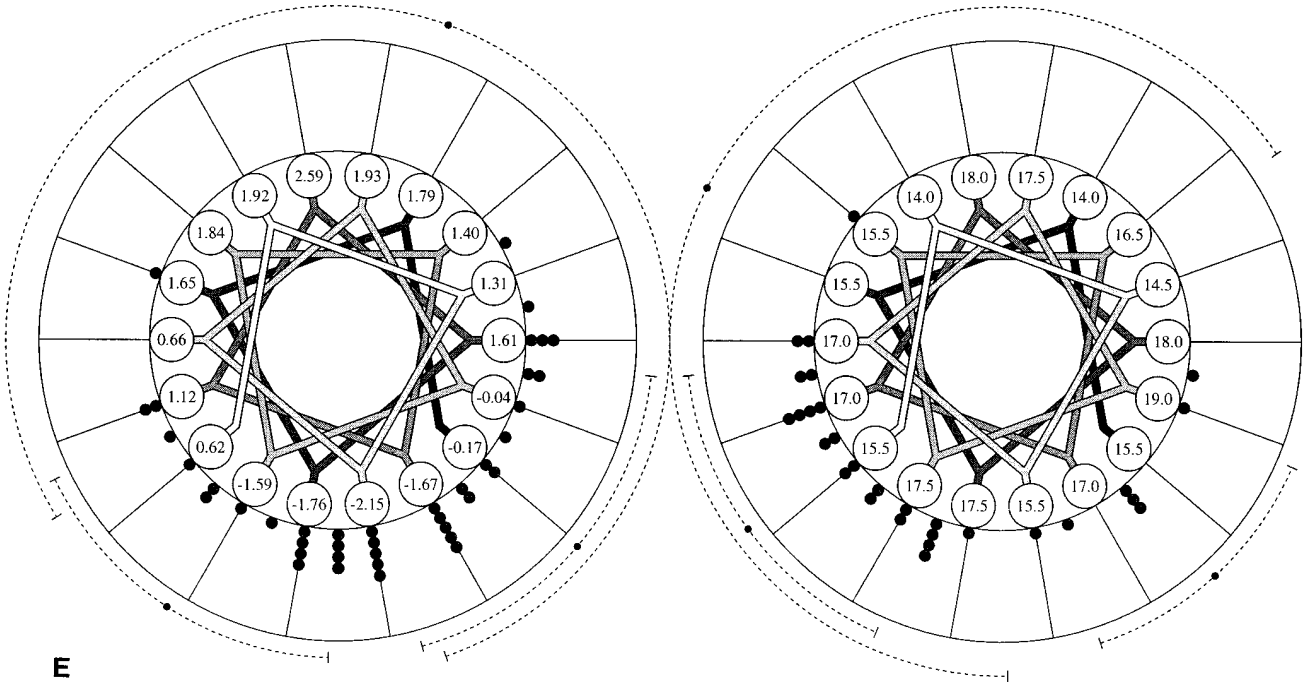


Fig. 3.

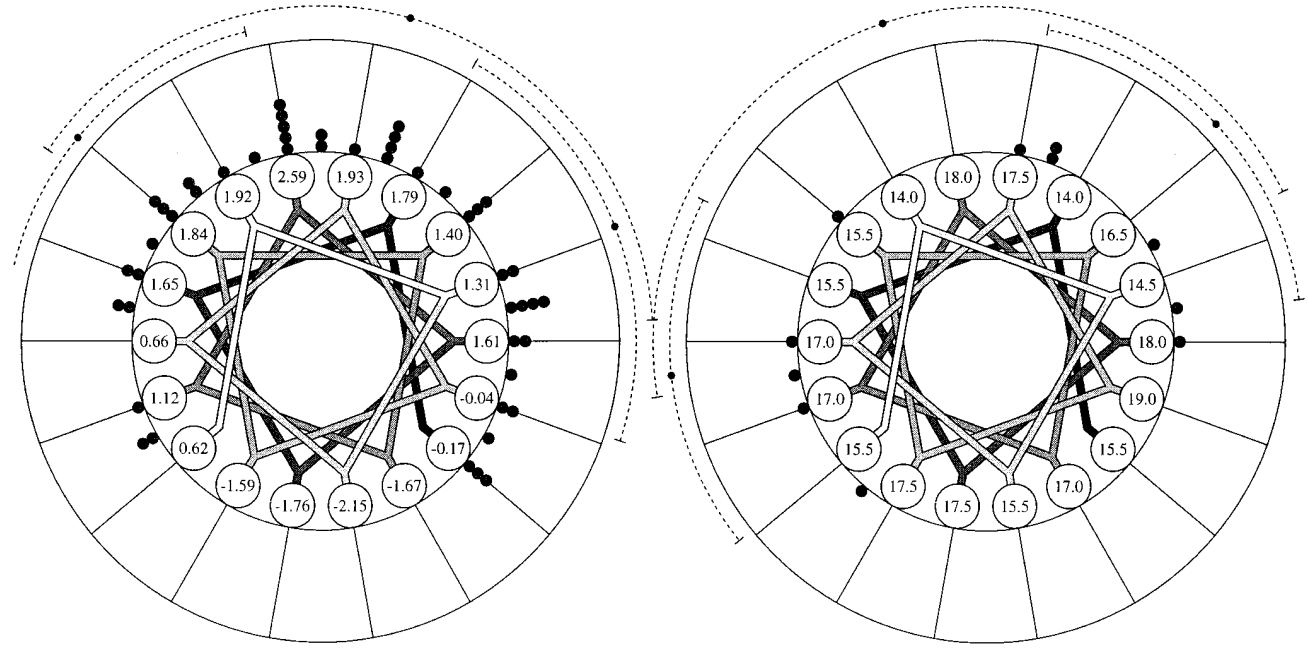
drophobic face points towards the top of the page. Selected residues of the set are projected onto two wheel diagrams. In default mode, the left-hand wheel displays the counts of all positively charged residues and the right-hand wheel displays the counts of all negatively charged residues or a pair of individual amino acid residues can be selected. The wheels display at every 10 degree position around the helix the count (small filled circles) of positively charged residues that fall into that position. The average hydrophobicities of the uncharged residues (left hand wheel) and the total number of residues for the set of sequences (right hand wheel) are displayed at every 20 degree position within the open circles of the superimposed helical wheel.



E

Positive Residues

Negative Residues



F

Leu Residues

Phe Residues

Fig. 3.

TABLE 1. Analysis of test proteins by the program LOCATE_HYB

	Motif A Dataset	Motif G Dataset	Motif C Dataset	Motif B Dataset	ApoB-100
	hydrophobic segments/100 residues				
Total hydrophobicity ≥10 kcal/mol	0.08	0.17	0.14	0.83	0.29
Segment length ≥10 residues					β_1 0.33 β_2 0.57

The four test datasets, described in detail by Segrest et al. (6), are composed of amino acid sequence chimeras of proteins enriched in the lipid-associating class A (apolipoproteins) amphipathic α helixes (motif A), the non-lipid-associating class G (globular proteins) amphipathic α helixes (motif G), and the non-lipid-associating class C (coiled-coil proteins) amphipathic α helixes (motif C), and non-lipid-associating proteins containing predominantly β strand secondary structure (motif B).

Structural homologies between the nine apoB-100 sequences

Figure 5 represents LOCATE_ALPHA analyses of the nine apoB-100 sequences aligned as shown in Fig. 4 and indicates that the α_2 and α_3 domains are generally conserved. Each of the four partial sequences that overlap in the C-terminal α_3 domain (pig, mouse, rat, and chicken) and the two sequences that overlap the entire α_2 domain (pig and mouse) show a well-defined cluster of presumed lipid-associating amphipathic α helixes ($\Lambda \geq 7$) that correspond closely to the α_3 domain and the α_2 domain of human apoB-100, respectively (Fig. 5A). The additional sequence (frog) that overlaps a small portion of the C-terminal end of α_2 also shows the beginning of amphipathic α helixes in that portion of α_2. In all five species in which the α_3 domain has been sequenced, an amphipathic α helix more than 50 residues long with high calculated lipid affinity is located in the neighborhood of residue 4480 (**Fig. 6A**).

Figure 5B is a LOCATE_ALPHA/CLASS analysis of all class A and Y amphipathic α helixes identified in the nine apoB-100 sequences. The distribution of these amphipathic α helixes corresponds closely to that of the presumed lipid-associating amphipathic α helixes shown in Fig. 5A; however, clustering of class A and Y amphipathic motifs is a more definitive marker of the α_2 and α_3 domains than calculated high lipid affinity.

The LOCATE_ALPHA analysis is Fig. 5A, in addition to marking the α_2 and α_3 domains, also marks the α_1 domain. Although a number of the amphipathic α helixes in the α_1 domain, unlike the α_2 and α_3 domains, may be conserved (arrows), more sequences that overlap the α_1 domain of human apoB-100 are needed to confirm this impression.

Figure 7 is a LOCATE_BETA analysis of all amphipathic β strands identified in the nine apoB-100 sequences that have a nonpolar face hydrophobicity ≥7. It is clear from this figure that amphipathic β strands cluster in domains β_1 (approximately residues 827–2000) and β_2 (approximately residues 2571–4000) in all species sequenced in these regions. In contrast to the presumed lipid-associating amphipathic α helixes, a number of the amphipathic β strands of human apoB-100 are conserved (Fig. 7, arrowheads); this is particularly true in the putative LDL receptor-binding domain (13) located between residues 3100–3500 in β_2. Another feature of the amphipathic β strands that appears to be conserved are the several gap regions that occur in the β_1 and β_2 domains of human apoB-100 (e.g., between residues 900–1000 and 3200–3300). It is worthy of note that a long amphipathic β strand is found within the first 60 N-terminal residues of both human and mouse apoB-100 (Fig. 6B).

Finally, **Fig. 8** is a LOCATE_HYB analysis of all hydrophobic segments identified in the nine apoB-100 sequences that have a total hydrophobicity ≥7.5. The important point to make from this figure is that there is little conservation of the cluster of hydrophobic segments noted earlier in the β_2 domain of human apoB-100 between residues 3250–3347; these residues are represented in all 9 sequences.

DISCUSSION

Conservation of amphipathic motifs

In this study we have used the computer programs LOCATE_ALPHA, LOCATE_BETA and LOCATE_HYB to compare the distribution of lipid-associating motifs (amphipathic α helixes, amphipathic β strands, and hydrophobic segments, respectively) in human apoB-100 to their distribution in partial sequences from eight additional species of vertebrates (monkey, pig, mouse, rat, hamster, rabbit, chicken, and frog). **Figure 9** is schematic map of the general distribution in the apoB-100 molecule of amphipathic α helixes and amphipathic β strands as determined here.

The LOCATE_ALPHA/CLASS algorithm that produced Figs. 2B and 5B used the radial position of charged residues only to identify class A and Y amphipathic α helixes; hydrophobicity of the hydrophobic face was not a consideration. It therefore seems remarkable that: *a*) the hydrophobic residues of the class A and Y amphipathic α show clustering of hydrophobic residues that is as well defined as that found in amphipathic α helixes selected on the basis of high calculated lipid affinity (Fig. 3B vs. Fig. 3D); and *b*) class A and Y amphipathic α helixes are confined almost ex-

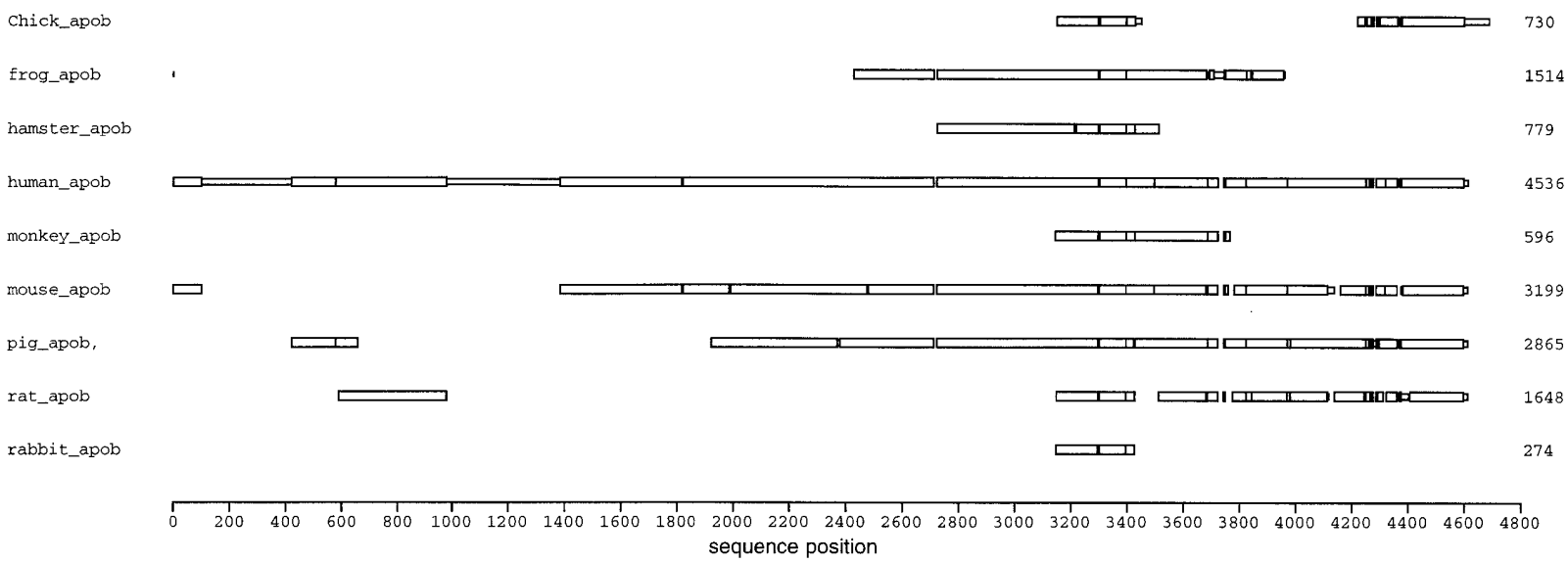


Fig. 4. Alignment of the complete human apoB-100 sequence and determination of sequence homology with partial apoB-100 sequences from chicken, frog, hamster, monkey, mouse, pig, rat, and rabbit. Alignment and sequence homology analysis were performed with the Macintosh version of the MACAW (Multiple Alignment Construction and Analysis Workbench) multiple sequence alignment program (11, 12) as described in Methods. Partial sequences of apoB-100 from six species of vertebrates: chicken, hamster, monkey, pig, rat, and rabbit, were obtained in the FASTA format as described in Methods. Partial sequences of apoB-100 from two additional species, mouse and frog, were determined in the laboratories of co-authors Stephen J. Young and Thomas L. Innerarity, respectively. This figure is a schematic of the results of the MACAW analysis. Sequences available for the apoB-100 of each vertebrate species are denoted by bars. The narrower bars represent sequences in which there is no overlap or homology. The thicker bars denote regions of linked homology blocks identified using the default settings for the program. Gaps represent regions of sequence in which insertions have occurred in other sequences.

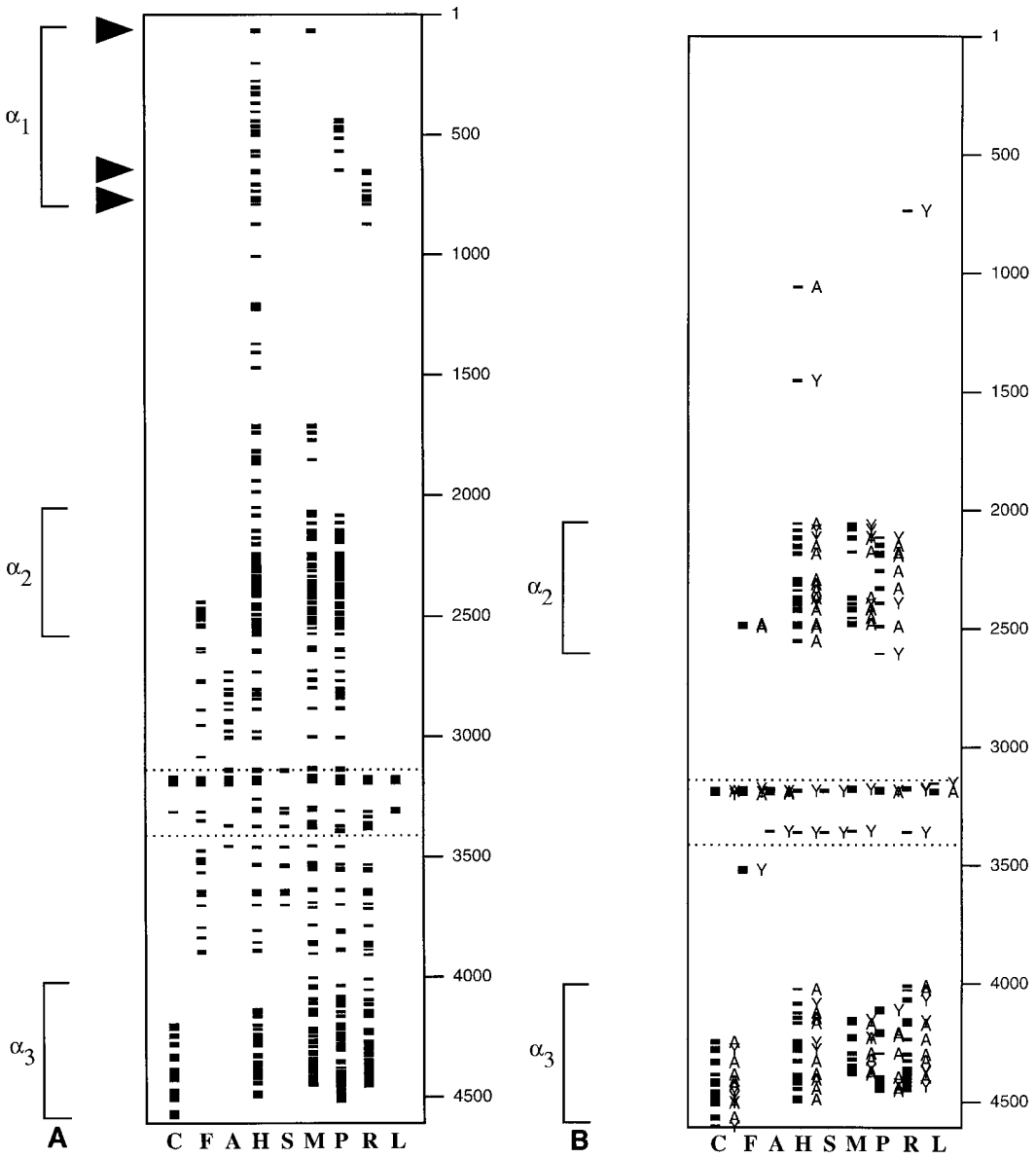


Fig. 5. Phylogenetic comparison of the distribution of amphipathic α helixes in apoB-100 from nine species of vertebrates: C (chicken), F (frog), A (hamster), H (human), S (monkey), M (mouse), P (pig), R (rat), and L (rabbit). Sequence alignment was as in Fig. 4. A: All amphipathic α helixes (length ≥10 residues) with $\Lambda \geq 7$ kcal/mol. Arrowheads indicate regions of possible conservation of amphipathic α helixes. B: All class A and class Y amphipathic α helixes with length ≥10 residues (using default parameters).

clusively to the α_2 and α_3 domains. These results suggest that class A and Y motifs are confined almost exclusively to amphipathic α helixes possessing high lipid affinity.

Lipid-associating amphipathic β strands cluster in two domains in all species for which these regions have been sequenced (Fig. 7). Unlike the lipid-associating amphipathic α helixes, several individual amphipathic β strands within these domains, indicated by arrows in Fig. 7, are conserved. The β domains show considerably fewer regions of "homology" to the exchangeable apolipoproteins than do the α domains (MACAW analyses, data not shown).

The β_1 domain is represented only partially in the sequences of pig, mouse, and rat. In these species, the amphipathic β strands of this domain are distributed, on average, between residues 827–2001. There are a number of obvious gaps in amphipathic β strand distribution in the β_1 domain and several of these gaps may be conserved (e.g., residues 878–966); however, the

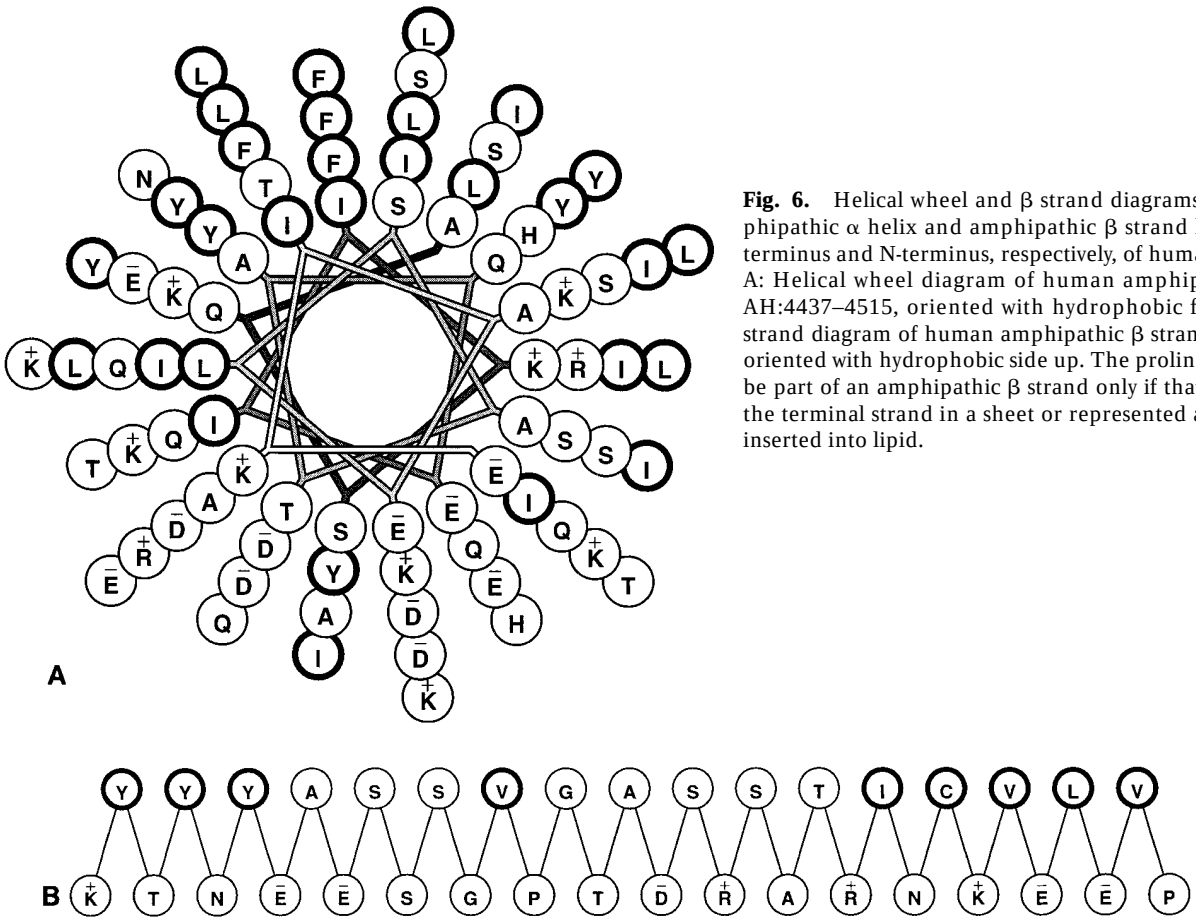


Fig. 6. Helical wheel and β strand diagrams of long amphipathic α helix and amphipathic β strand located at C-terminus and N-terminus, respectively, of human apoB-100. A: Helical wheel diagram of human amphipathic helix, AH:4437–4515, oriented with hydrophobic face up. B: β strand diagram of human amphipathic β strand, Aβ:24–58, oriented with hydrophobic side up. The proline at 38 would be part of an amphipathic β strand only if that strand were the terminal strand in a sheet or represented a lone strand inserted into lipid.

paucity of overlapping sequences in this domain precludes a firm conclusion.

The amphipathic β strands of the β_2 domain are distributed, on average, between residue 2571 to approximately residue 4000. There is a variability of 100 residues or so in the C-terminal end of this domain in the four sequences in which this region is present (human, pig, mouse, rat); there appears to be more conservation of individual amphipathic β strands in the N-terminal two-thirds of the β_2 domain than in the C-terminal one-third (Fig. 7). As is true for the β_1 domain, a number of obvious gaps in amphipathic β strand distribution are noted in the β_2 domain. The gap in the region of residue 3200 is conserved across all 9 sequences. The gap at residue 2700 is conserved in the three sequences, human, pig, and mouse, that overlap this region; there may also be a conserved gap of variable length just beyond residue 3500.

Finally, our analysis indicates that hydrophobic segments are present in apoB-100 but at a frequency no greater than in other β sheet-containing proteins (Table 1). The cluster of hydrophobic segments located in the neighborhood of residue 3300 in human apoB-100 is poorly conserved (Fig. 8).

Amphipathic α helical domains

The α_2 domain in human apoB-100 is "homologous" to certain amphipathic helix-containing regions of six of the exchangeable apolipoproteins A-I, A-II, A-IV, C-I, C-III, and E (MACAW analyses, data not shown). This domain is represented in its entirety in the sequences of human, pig, and mouse and partially represented in the frog sequence. In these species, the amphipathic α helixes of the α_2 domain cluster between residues 2075 ± 25 and 2575 ± 25 (Fig. 5); the highest concentration of amphipathic α helixes occurs in the C-terminal half of this domain.

The α_3 domain in human apoB-100 is "homologous" to specific amphipathic helix-containing regions of all seven of the exchangeable apolipoproteins A-I, A-II, A-IV, C-I, C-II, C-III, and E; this homology is especially striking with the C-terminal two-thirds of human apoA-I (MACAW analyses, data not shown). The α_3 domain is represented in its entirety in the sequences of human,

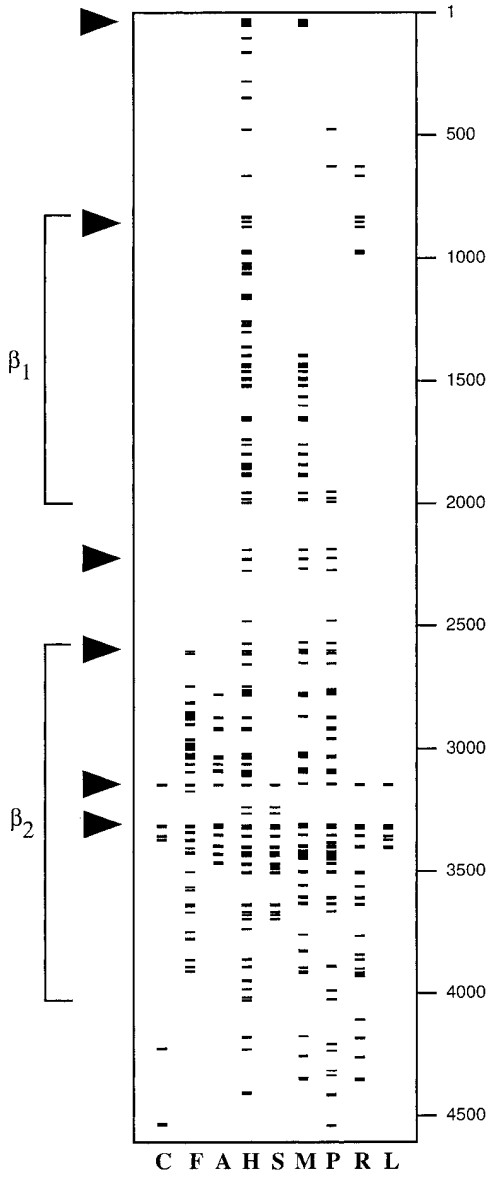


Fig. 7. Phylogenetic comparison of the distribution of amphipathic β strands in apoB-100 from nine species of vertebrates: C (chicken), F (frog), A (hamster), H (human), S (monkey), M (mouse), P (pig), R (rat), and L (rabbit). Sequence alignment was as in Fig. 4. The plot is of all amphipathic β strands (length ≥8 residues, hydrophobic moment ≥2 kcal/mol) with total hydrophobicity of the nonpolar face ≥7 kcal/mol (compare with Fig. 2C). Arrowheads indicate regions of apparent conservation of amphipathic β strands.

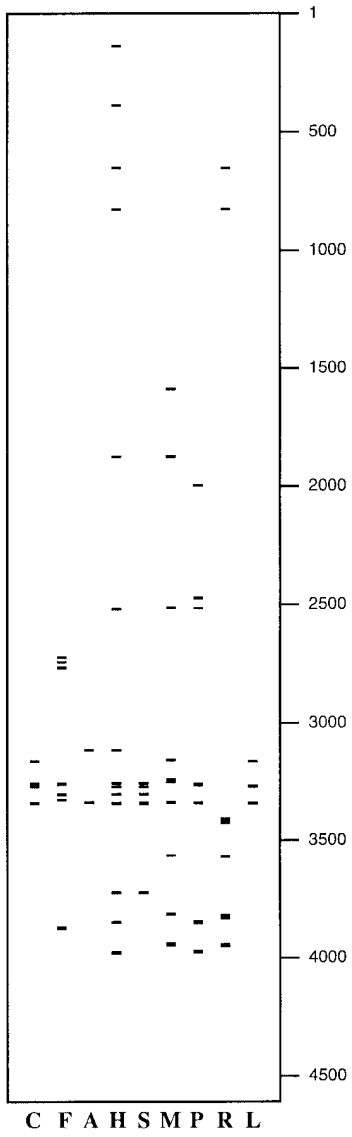


Fig. 8. Phylogenetic comparison of the distribution of hydrophobic segments in apoB-100 from nine species of vertebrates: C (chicken), F (frog), A (hamster), H (human), S (monkey), M (mouse). P (pig), R (rat), and L (rabbit). Sequence alignment was as in Fig. 4. The plot is of all hydrophobic segments (length ≥10 residues) with total hydrophobicity ≥7.5 kcal/mol (compare with Fig. 2D).

pig, mouse, rat, and chicken. In these species, the amphipathic α helixes of the α_3 domain cluster between residues 4100 ± 100 and 4550 ± 50; they are distributed more evenly than the amphipathic α helixes of the α_2 domain (Fig. 5).

The α_2 and α_3 domains of apoB-100 correspond closely to the two major apoB-100 lipid-associating domains reported by Yang, et al. (13) at residues 2100–2700 and 4100–4500 using the principle of nonreleasability of tryptic peptides from trypsin-treated intact LDL. Others have also suggested the presence of amphipathic α helixes in these regions of apoB-100 (14, 15).

What might be the function of the lipid-associating amphipathic α helical domains (α_2 and α_3) of apoB-100? We have proposed that these domains represent

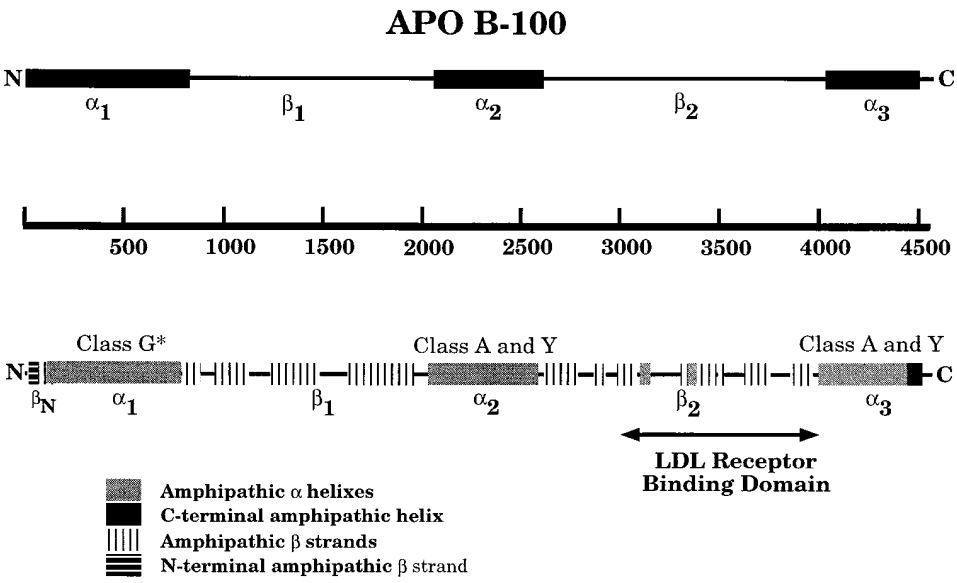


Fig. 9. Schematic diagram of the distribution of amphipathic α helixes and amphipathic β strands in apolipoprotein B-100.

the flexible, hinged domains of apoB-100, perhaps playing a role in LDL particle size variability (6).

It is unfortunate that only limited amino acid sequence data on the α_1 domain of apoB-100 from species other than human is currently available, as comparative analyses of α_1 sequence data from multiple species may be useful in the elucidation of mechanisms regulating the early steps in assembly of apoB-containing lipoproteins (see discussion below).

Amphipathic β strand domains

The presence of β structure in apoB-100 has been suggested in the past using spectroscopic methods (16–20). The first suggestion of the presence of amphipathic β strands in apoB-100 was made by Osterman et al. (21). While the amphipathic β strand is now accepted by many investigators as a major lipid-associating motif in apoB (15), there is little published direct evidence for the existence of this motif and little is known about its mode of association with phospholipid monolayers; even less is known about its biological functions.

There are a number of unanswered questions concerning amphipathic β strands in apoB-100. *a*) Do amphipathic β strands associate with lipid as individual strands or, as suggested recently by Don Small (personal communication), as antiparallel sheets or both? We favor the possibility that both strands and sheets can interact depending upon the local conditions. *b*) What is the lipid affinity of amphipathic β strands compared to class A amphipathic α helixes from apoB-100 as well as from exchangeable apolipoproteins? *c*) Assuming they exist, what is the lipid affinity of antiparallel amphipathic β sheets compared to amphipathic β strands? We hypothesize that large antiparallel amphipathic β sheets have what amounts to an infinitely high lipid affinity. *d*) What is the biological function of the amphipathic β strands in apoB beyond determining lipid affinity? *e*) What is the function of the gap regions between amphipathic β strands of the β_1 and β_2 domains?

Relationship of pentapartite model to truncated apoB-100 constructs

What experimental evidence is there to support the NH_2-α_1-β_1-α_2-β_2-α_3-COOH pentapartite model for apoB-100 shown in Fig. 9? Several studies have shown that the diameters of secreted lipoprotein particles are proportional to the length of C-terminal truncated apoB-100 fragments until the N-terminal end of the β_1 domain is approached (residues 1100–1300), at which point the linearity begins to fall off (22–24).

In another study, expression of C-terminal truncated forms of apoB-100 in HepG2 cells demonstrated that, with stepwise truncation, progressively more apoB-100 was recovered free of lipid in the cell culture medium (25); recovery of truncated apoB-100 was 100% at residue 1860, 48% at residue 1315, 20% at residue 1040, 18% at residue 770, and 0% at residue 590. Thus, beyond the N-terminal end of the β_1 domain (approximately residue 827, Fig. 9), the lipid affinity of truncated and secreted apoB-100 approaches zero.

In a separate study using the McA-Rh777 cell line,

apoB-100 truncated at residues 1040 and 815 were recovered in the lipid-free fraction, whereas apoB-100 truncated at or beyond residue 1270 formed discrete lipoprotein particles (23). Again the ability of apoB-100 to form lipoprotein particles disappears as the N-terminal end of the β_1 domain is approached.

Finally, when apoB-100 was truncated at residue 771 (apoB-17) and expressed in murine C127 cells, 60% was secreted as soluble, lipid-poor protein that formed discoidal complexes with dimyristoyl phosphatidylcholine (26). Circular dichroic spectra of this complex suggested lipid-associated apoB-17 was 39% α helix and 36% β sheet. LOCATE analyses of the α_1 "globular" domain of human apoB-100 (residues 1–800) predicts that 36% of this domain is lipid-associating amphipathic α helix ($\Lambda \geq 7$) and 26% is lipid-associating amphipathic β strand (total nonpolar face hydrophobicity ≥5).

We conclude that four alternating lipid-associating domains, -β_1-α_2-β_2-α_3-COOH, enriched in amphipathic β strands and amphipathic α helixes, respectively, are ubiquitous to the structure of apoB-100 from vertebrate species. While the α_1 "globular" domain of apoB-100 seems almost certain to be conserved in apoB-100 from all species of vertebrates, the degree of this conservation must await more overlapping sequences with this domain in human apoB-100. We speculate that, because of the apparent role of this domain in the assembly of apoB-containing lipoprotein particle, the α_1 "globular" domain will prove to be the most conserved of the α and β domains in apoB-100.

This work was supported in part by NIH grants HL-34343 to JPS, HL-41633 to SHY and HL-41633 to TLI. SHY would like to thank Dr. Brian J. McCarthy for advice and assistance during the sequencing of mouse apoB-100.

Manuscript received 23 July 1997 and in revised form 17 September 1997.

REFERENCES

1. Segrest, J. P., R. L. Jackson, J. D. Morrisett, and A. M. Gotto. 1974. A molecular theory of lipid–protein interactions in the plasma lipoproteins. *FEBS Lett.* **38:** 247–253.
2. Segrest, J. P., D. W. Garber, C. G. Brouillette, S. C. Harvey, and G. M. Anantharamaiah. 1994. The amphipathic α helix: a multifunctional structural motif in plasma apolipoproteins. *Adv. Protein Chem.* **45:** 303–369.
3. Segrest, J. P., H. De Loof, J. G. Dohlman, C. G. Brouillette, and G. M. Anantharamaiah. 1990. The amphipathic helix motif: classes and properties. *Proteins: Struct. Funct. Genet.* **7:** 1–15.
4. Segrest, J. P., M. K. Jones, H. De Loof, C. G. Brouillette, Y. V. Venkatachalapathi, and G. M. Anantharamaiah. 1992. The amphipathic helix in the exchangeable apolipoproteins: a review. *J. Lipid Res.* **33:** 141–166.
5. 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.
6. Segrest, J. P., M. K. Jones, V. K. Mishra, G. M. Anantharamaiah, and D. W. Garber. 1994. Apolipoprotein B-100 has a pentapartite structure composed of three amphipathic α helical domains alternating with two amphipathic β strand domains: detection by the computer program LOCATE. *Arterioscler. Thromb.* **14:** 1674–1685.
7. Jones, M. K., G. M. Anantharamaiah, and J. P. Segrest. 1992. Computer programs to identify and classify amphipathic α helical domains. *J. Lipid Res.* **33:** 287–296.
8. Engelman, D. M., T. A., Steitz, and A. Goldman. 1986. Identifying transbilayer helices in amino acid sequences of membrane proteins. *Annu. Rev. Biophys. Biophys.Chem.* **15:** 321–353.
9. Palgunachari, M. N., V. K. Mishra, S. Lund–Katz, M. C. Phillips, S. O. Adeyeye, S. Alluri, G. M. Anantharamaiah, and J. P. Segrest. 1996. Only the two end helixes of eight tandem amphipathic helical domains of human apoA-I have significant lipid affinity: implications for HDL assembly. *Arterioscler. Thromb. Vasc. Biol.* **16:** 328–338.
10. Jacobs, R. E., and S. H. White. 1989. The nature of hydrophobic binding of small peptides at the bilayer interface: implications for the insertion of transbilayer helices. *Biochemistry.* **28:** 3421–3437.
11. Schuler, G. D., S. F. Altschul, and D. J. Lipman. 1991. A workbench for multiple alignment construction and analysis. *Proteins Struct. Funct. Genet.* **9:** 180–190.
12. Lawrence, C. E., S. F. Altschul, M. S. Boguski, J. S. Liu, A. F. Neuwald, and J. C. Wootton. 1993. Detecting subtle sequence signals: a Gibbs sampling strategy for multiple alignment. *Science.* **262:** 208–214.
13. Yang, C-Y., Z-W. Gu, S. A. Weng, T. W. Kim, S-H. Chen, H. J. Pownall, P. M. Sharp, S. W. Liu, W. H. Li, A. M. Gotto, Jr., and L. Chan. 1989. Structure of apolipoprotein B-100 of human low density lipoproteins. *Arterioscler. Thromb.* **9:** 96–108.
14. De Loof, H., M. Rosseneu, C. Y. Yang, W. H. Li, A. M. Gotto, and L. Chan. 1987. Human apolipoprotein B: analysis of internal repeats and homology with other apolipoproteins. *J. Lipid Res.* **28:** 1455–1465.
15. Nolte, R. T. 1994. Structural analysis of the human apolipoproteins: an integrated approach utilizing physical and computational methods. Thesis Dissertation. Boston University.
16. Gotto, A. M., R. I. Levy, and D. S. Fredrickson. 1968. Observations on the conformation of human beta lipoprotein: evidence for the occurrence of beta structure. *Biochemistry.* **60:** 1436–1441.
17. Scanu, A., and R. Hirz. 1968. Human serum low-density lipoprotein protein: its conformation studied by circular dichroism. *Nature.* **218:** 200–201.
18. Chen, G., and J. P. Kane. 1979. Secondary structure in very low density and intermediate density lipoproteins of human serum. *J. Lipid Res.* **20:** 481–488.
19. Walsh, M. T., and D. Atkinson. 1983. Solubilization of low-density lipoprotein with sodium deoxycholate and recombination of apoprotein B with dimyristoylphosphatidylcholine. *Biochemistry.* **22:** 3170–3178.
20. Goormaghtigh, E., V. Cabiaux, J. De Meutter, M. Rosseneu, and J. M. Ruysschaert. 1993. Secondary structure of the particle associating domain of apolipoprotein B-100

in low-density lipoprotein by attenuated total reflection infrared spectroscopy. *Biochemistry.* **32:** 6104–6110.
21. Osterman, D., R. Mora, F. J. Kézdy, E. T. Kaiser, and S. C. Meredith. 1984. A synthetic amphiphilic β-strand tridecapeptide: a model for apolipoprotein B. *J. Am. Chem. Soc.* **106:** 6845–6847.
22. Young, S. D., S. T. Hubi, R. S. Smith, S. M. Snyder, and J. F. Terdiman. 1990. Familial dysbetalipoproteinemia caused by a mutation in the apolipoprotein B gene that results in a truncated species of apolipoprotein B (B–31): a unique mutation that helps to define the portion of the apolipoprotein B molecule required for the formation of buoyant, triglyceride-rich lipoproteins. *J. Clin. Invest.* **85:** 933–942.
23. Yao, Z., B. D. Blackhart, M. F. Linton, S. M. Taylor, S. G. Young, and B. J. McCarthy. 1991. Expression of carboxyl-terminally truncated forms of human apolipoprotein B in rat hepatoma cells. *J. Biol. Chem.* **266:** 3300–3308.
24. Schumaker, V. N., M. L. Phillips, and J. E. Chatterton. 1994. Apolipoprotein B and low-density lipoprotein structure: implications for biosynthesis of triglyceride-rich lipoproteins. *Adv. Protein Chem.* **45:** 205–248.
25. Graham, D., K. Lesley, J. J. Timothy, C. P. Tina, J. Richard, C. R. Pullinger, and J. Scott. 1991. Carboxyl-terminal truncation of apolipoprotein B results in gradual loss of the ability to form buoyant lipoproteins in cultured human and rat liver cell lines. *Biochemistry.* **30:** 5616–5621.
26. Herscovitz, H., M. Hadzopoulou-Cladaras, M. T. Walsh, C. Cladaras, V. I. Zannis, and D. M. Small. 1991. Expression, secretion and lipid-binding characterization of the N-terminal 17% of the apolipoprotein B. *Proc. Natl. Acad. Sci. USA* **88:** 7313–7317.