

N-terminal domain of apolipoprotein B has structural homology to lipovitellin and microsomal triglyceride transfer protein: a “lipid pocket” model for self-assembly of apoB-containing lipoprotein particles

Jere P. Segrest,^{1,*} Martin K. Jones,* and Nassrin Dashti[†]

Departments of Medicine and Biochemistry and the Atherosclerosis Research Unit,* University of Alabama Medical Center, Birmingham, AL 35294-0012, and Departments of Nutrition Sciences and the Atherosclerosis Research Unit,[†] University of Alabama Medical Center, Birmingham, AL 35294-0011

Abstract The process of assembly of apolipoprotein (apo) B-containing lipoprotein particles occurs co-translationally after disulfide-dependent folding of the N-terminal domain of apoB but the mechanism is not understood. During a recent database search for protein sequences that contained similar amphipathic β strands to apoB-100, four vitellogenins, the precursor form of lipovitellin, an egg yolk lipoprotein, from chicken, frog, lamprey, and *C. elegans* appeared on the list of candidate proteins. The X-ray crystal structure of lamprey lipovitellin is known to contain a “lipid pocket” lined by antiparallel amphipathic β sheets. Here we report that the first 1000 residues of human apoB-100 (the α_1 domain plus the first 200 residues of the β_1 domain) have sequence and amphipathic motif homologies to the lipid-binding pocket of lamprey lipovitellin. We also show that most of the α_1 domain of human apoB-100 has sequence and amphipathic motif homologies to human microsomal triglyceride transfer protein (MTP), a protein required for assembly of apoB-containing lipoproteins. Based upon these results, we suggest that an LV-like “proteolipid” intermediate containing a “lipid pocket” is formed by the N-terminal portion of apoB alone or, more likely, as a complex with MTP. This intermediate produces a lipid nidus required for assembly of apoB-containing lipoprotein particles; pocket expansion through the addition of amphipathic β strands from the β_1 domain of apoB results in the formation of a progressively larger high density lipoprotein (HDL)-like, then very low density lipoprotein (VLDL)-like, spheroidal lipoprotein particle.—Segrest, J. P., M. K. Jones, and N. Dashti. N-terminal domain of apolipoprotein B has structural homology to lipovitellin and microsomal triglyceride transfer protein: a “lipid pocket” model for self-assembly of apoB-containing lipoprotein particles. *J. Lipid Res.* 1999. 40: 1401–1416.

Supplementary key words plasma apolipoproteins • LOCATE • microsomal triglyceride transfer protein (MTP) • amphipathic β sheets • amphipathic α helices • co-translational • vitellogenin • lipovitellin precursor • α_1 domain • β_1 domain

Apolipoprotein (apo) B exists in two distinct forms, apoB-100 and apoB-48. In humans, apoB-100, containing

4536 amino acid residues, is synthesized in the liver, is a constituent of VLDL, intermediate density lipoprotein (IDL), and low density lipoprotein (LDL), and is a ligand for the LDL receptor (1). In humans, apoB-48 comprises residues 1–2152 of apoB-100, is synthesized in the intestine, and is a constituent of chylomicrons (1).

The mechanism whereby each apoB molecule physically assembles lipid into a triacylglycerol-rich lipoprotein is not well understood. It is known that lipoprotein assembly occurs co-translationally in HepG2 cells; i.e., while the C-terminal portion is still being synthesized on the ribosome of the endoplasmic reticulum, the N-terminal portion has assembled a small lipoprotein particle (1). It has also been shown that disruption of one or two of the six disulfide bonds in the N-terminal domain of apoB blocks MTP-dependent lipid transfer to nascent apoB and subse-

Abbreviations: apo, apolipoprotein; VLDL, very low density lipoprotein; IDL, intermediate density lipoprotein; LDL, low density lipoprotein; HDL, high density lipoprotein; kDa, kilodalton; hAPOB, human apoB-100; hB, hAPOB; LV, lamprey vitellin; cLV, chicken LV; fLV, frog LV; vLV, vertebrate LV (LV, cLV, and fLV); LV precursor; lamprey vitellogenin; MTP, microsomal triglyceride transfer protein; hMTP, human MTP; LV-1n, LV-1c, and LV-2, processed polypeptide chains of lamprey LV in order from N-terminus; β B:LV, β A:LV, and β D:LV, amphipathic β sheets lining triangular lamprey LV lipid pocket; β C:LV, additional amphipathic β sheet (barrel) in lamprey LV; I-V β :LV, five defined clusters of amphipathic β strands in lamprey LV; β C:B, β A:B, and β B:B: defined clusters of amphipathic β strands in hAPOB homologous to β B:LV, β A:LV, and β D:LV of lamprey LV; I-II β :MTP, two defined clusters of amphipathic β strands in hMTP; α :LV, α helical domain of lamprey LV; α :B, a helical domain of hAPOB homologous to the C-terminal half of the α helical domain of lamprey LV; α :MTP, α helical domain of hMTP homologous to the N-terminal half of the α helical domain of lamprey LV; I-III α :LV, three defined clusters of positively charged amphipathic α helices in lamprey LV; HD-vLV-I-V, five homologue domains of vLV; HD-vLV:B-I-IV, four homologue domains of vLV and hAPOB; HD-vLV:MTP-I-III, three homologue domains of vertebrate lipovitellins and hMTP; HD-MTP:B-I-IV, four homologue domains of hMTP and hAPOB.

¹ To whom correspondence should be addressed.

quent secretion of apoB-containing lipoproteins (2–5). Thus proper folding of the N-terminal domain of apoB appears to be required for assembly of apoB-containing lipoproteins.

We developed a computer program (LOCATE) to help identify lipid-associating domains in apoB-100. Initial LOCATE analysis indicated the presence of two domains of putative amphipathic β strands alternating with three domains of putative amphipathic α helices in a pentameric structure: $\text{NH}_2\text{-}\alpha_1\text{-}\beta_1\text{-}\alpha_2\text{-}\beta_2\text{-}\alpha_3\text{-COOH}$ (6). In a later study, we compared 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). We concluded that four alternating lipid-associating domains, $\text{-}\beta_1\text{-}\alpha_2\text{-}\beta_2\text{-}\alpha_3\text{-COOH}$, are common supramolecular features of apolipoprotein B-100 in nine vertebrate species and that class A and Y amphipathic α helices (see legend to Fig. 1) are confined almost exclusively to the α_2 and α_3 domains (7).

During a LOCATE database search (/DATABASE_SEARCH, see Methods) for protein sequences that contained similar amphipathic β strands to apoB-100, four vitellogenins, the precursor form of lipovitellin (LV), an egg yolk lipoprotein, from chicken, frog, lamprey, and *C. elegans* appeared on a list of 192 candidate proteins. Lipovitellin is a lipoprotein found in the egg yolk of oocytes of many egg-laying animals (8). Vitellogenin, the precursor of LV (to be referred to as LV precursor), is synthesized in the liver (9) and is then secreted into the plasma, where it is taken up by the ovary via the LDL receptor (10). Lamprey LV contains three polypeptide chains, LV-1n ($M_r = 66.8$ kDa), LV-1c ($M_r = 40.75$ kDa), and LV-2 ($M_r = 35.2$ kDa) (8). Lamprey LV is 15% lipid by weight (thus is a proteolipid rather than a classical lipoprotein) and serves as a source of amino acids and lipid during embryogenesis. The lipid is non-covalently associated and consists of phospholipid (approx. 27 molecules per LV monomer) triacylglycerol (approx. 11 molecules per monomer), and cholesterol (8, 11). The X-ray crystal structure for lamprey LV has been determined at 2.8 Å resolution (11).

In the present report, we demonstrate that LV from three vertebrate species and human microsomal triglyceride transfer protein (hMTP) have both sequence and amphipathic motif homology to the first 1000 and 900 residues, respectively, of human apoB-100 (hAPOB). We propose a model for the assembly of apoB-containing lipoprotein particles in which initial lipidation occurs in an LV-like lipid-binding pocket confined largely to the α_1 domain of apoB alone or, more likely, formed in combination with MTP.

METHODS

Programs for locating amphipathic α helices or amphipathic β strands in protein sequences

LOCATE finds and displays potentially lipid-associating motifs in an amino acid sequence where the amino acid residue file contains a sequence of one letter amino acids and the optional

free energy of transfer file contains default settings for amino acid residues. The free energy of transfer file, representing a modification of the GES scale hydrophobicity (12), contains a free energy of transfer, hydrocarbon to water, in kcal/mol for each amino acid residue used in the amino acid residue file and is as follows: 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; Gly 0.1; Ala 0.0; 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. /DATABASE_SEARCH allows the amino acid residue file to be read as a series of amino acid sequences for database search.

LOCATE_ALPHA identifies potential amphipathic α helices within a given amino acid sequence using termination rules described elsewhere (6). /NOCHECK_HYDROPHOBIC_RESIDUES: By default, the rules for extending the lengths of helices incorporate checks on the location of hydrophobic residues. If this option is selected, these checks are not performed. /CLASS=list specifies the classes of helices that will be reported. Helices determined to be of a certain type are reported if the list contains the respective character. A: helix is determined to be class A; Y: helix is determined to be class Y; G*: helix is any other amphipathic α helix. The algorithm for identification of class A and class Y amphipathic helices is described elsewhere (7). /LAMBDA_HYB_CUTOFF specifies the minimum value that the Δ lipid affinity of the candidate helix can be and still be selected.

The function for calculating lipid affinity (Δ) of amphipathic α helices (6) was derived as follows. The snorkeling of basic residues allows for greater penetration of class A amphipathic helices 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 (6). Using neutron diffraction, Jacobs and White (13) have measured the gradient that H_2O forms 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 non-polar face. Therefore, the overall lipid affinity of an amphipathic helix will partially depend upon its depth of lipid penetration (6). Combining the water gradient determined by Jacobs and White (13) 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 gradient, $\delta_i(\text{Å})$, that is a function of depth of penetration in Å, $\delta_i(\text{Å})$ (14). Two studies of the interactions of amphipathic helical peptides with lipid bilayers show that Δ lipid affinity calculations provide significantly better correlations with measured lipid affinity than other properties, such as hydrophobic moment, total hydrophobicity of the hydrophobic face, or total hydrophobicity (14, 15).

LOCATE_BETA identifies potential amphipathic β strands within a given amino acid sequence using termination rules also described elsewhere (6). LOCATE_MARK identifies the location of selected amino acid residues within a given amino acid sequence.

/MINIMUM_LENGTH specifies a minimum length of the candidate motif for selection. /CHARGE. If + or - is used, a number can follow the sign within the double quotes indicating the minimum number of those kind of charges to be on an amphipathic motif for the amphipathic motif to be reported. The allowable characters are as follows: "+," number of positive charges is greater than or equal to twice the number of negative charges; "-", number of negative charges is greater than or equal to twice

the number of positive charges. /NONPOLAR__HYB shows the average hydrophobicity of the nonpolar face of each amphipathic motif selected. /NONPOLAR__HYB__CUTOFF specifies a minimum average hydrophobicity of the nonpolar face of the candidate amphipathic motif for selection. /HYB__MOMENT__CUTOFF specifies a minimum hydrophobic moment of the candidate amphipathic motif for selection.

Protein sequence similarity database searches, dot matrix sequence alignments, local alignments of sequence blocks, and determination of structural homologies

Sequences of hAPOB, LV precursors (lamprey, frog, chicken, and *C. elegans*) and hMTP were downloaded in the FASTA format using the ENTREZ Browser (<http://atlas.nlm.nih.gov>). Protein sequence similarity searches against 84,000 sequences in the NBRF PIR protein sequence database (ftp.bchs.uh.edu/pub/gene-server/pir/pir_re/49/vms) were performed on an SGI Elan 4000 workstation with the Unix version of FASTA3 (16) (<ftp.virginia.edu:pub/fasta>) using ktup 1 and default parameters. Local sequence alignments by dot matrix analysis were performed with PLALIGN using default parameters on an SGI Elan 4000 workstation from the Unix version of FASTA20u6.shar (16) (<ftp.virginia.edu:pub/fasta>). Local alignments of sequence blocks were performed on an Alpha workstation with the Windows NT version of MACAW, Multiple Alignment Construction and Analysis Workbench (17, 18) (<ftp://ncbi.nlm.nih.gov/pub/schuler/macaw>) searching by segment pair overlap with a pairwise score cutoff of 35. To analyze similarities in amphipathic structural motifs, aligned sequences were subjected to LOCATE analysis on a VAXstation 4000/90 using cutoff parameters described in the figure legends.

Mapping of predicted amphipathic motifs and homologue domains onto the three dimensional structure of lamprey lipovitellin

The coordinates of the X-ray crystal structure for lamprey lipovitellin (11) were obtained from the Protein Databank (PDB). The amphipathic motifs in lamprey LV found by LOCATE and the vertebrate LV homologue domains identified by MACAW were mapped onto the three-dimensional structure of lamprey LV using TRIPOS SYBYL6.5 run on an SGI Elan 4000 workstation. For comparison of local sequence similarity to the position of amphipathic structural motifs, we created homologue domains from one or more sequence blocks. Homologue domains are defined as: *i*) widely spaced single sequence blocks or *ii*) a grouping of multiple closely clustered sequence blocks.

RESULTS

Database search for proteins containing the amphipathic β strand motif

Table 1 is a summary of the results of a LOCATE_BETA Swiss Prot database search for proteins containing >7 amphipathic β strands with hydrophobicity of the nonpolar face >10 kcal/mol, hydrophobic moment >1 and length >10 . A total of 192 proteins were selected: these included hAPOB (containing 27 amphipathic β strands, almost double that possessed by the protein with the next highest number) and all four LV precursors completely sequenced to date (*C. elegans*, lamprey, frog, and chicken); these five proteins are highlighted with bold type. Many of the selected proteins also appear to be membrane proteins (underlined) and viral coat proteins.

Identification of amphipathic motifs in lamprey LV

Based upon the results of Table 1 and because lamprey LV is known to contain a "lipid pocket" lined by antiparallel (most likely amphipathic) β sheets and surrounded by positively charged α helices (8), both features common to apoB, we investigated the possible structural relationship of lamprey LV to hAPOB. A number of investigators have observed that chicken and frog LV precursors have significant amino acid sequence similarity to hAPOB (9, 10, 19).

As noted in the discussion, the X-ray crystal structure for lamprey LV has been determined at 2.8 Å resolution (11). In the center of the monomer is a triangular-shaped cavity that apparently contains approximately 38 molecules of lipid. This "lipid pocket" is lined on its two major sides by antiparallel β sheets, β B:LV, and β A:LV, consisting of 8 and 7 strands, respectively. The bottom of the cavity is capped by a third antiparallel β sheet, β D:LV. An additional feature of the lamprey LV structure is a 17-helix antiparallel bundle that connects β A:LV to B:LV at the top of the triangular lipid cavity; this region is known to contain several positively charged α helices (20). Further, the central helices of the cluster form the wall of the lipid pocket in the gap between β A:LV and β B:LV. A fourth antiparallel β sheet folded as a β barrel, β C:LV, is attached to the top of β B:LV, and acts both as a putative hinge for

TABLE 1. Database search^a for proteins containing ≥ 7 amphipathic β sheets

Proteins	Number of Amphipathic β Sheets
Cell surface antigen: rickettsia	9
Sevenless protein: fruit fly	8
α -2-Macroglobulin precursor: rat	7
Endothelial actin-binding protein: human	11
ACV synthetase: 5 species	10–13
ApoB-100: human	27
Adenomatous <i>Polyposis coli</i> protein: human	8
3 Proteins: 3 species	7–9
<u>Bactericidal permeability increasing protein: human</u>	7
2 Proteins: 2 species	9, 7
<u>Brain calcium channel protein (2): rabbit</u>	7, 7
<u>Complement C3: 3 species</u>	7–9
<u>Complement C4: 2 species</u>	7, 8
5 Proteins: 5 species	7–11
<u>Adenylate cyclase: 3 species</u>	7–8
Desmosomal glycoprotein: 2 species	7, 8
Dynein: 3 species	8–10
17 Proteins: 17 species	7–15
<u>Fibronectin receptor α subunit: 2 species</u>	7, 9
<u>Membrane glycoproteins (11): 4 species</u>	7–13
Mycocerosic acid synthase: mycobacterium	8
<u>Cation-independent M-6-P receptor: 2 species</u>	8, 8
4 Proteins: 4 species	7–10
Viral coat proteins and polyproteins: 52 species	7–14
24 Proteins (including coat proteins): 20 species	7–14
Vitellogenin: chicken	10
Vitellogenin: <i>C. elegans</i>	7
Vitellogenin: African clawed frog	8
Vitellogenin: lamprey	8
13 Proteins: 13 species	7–14
Sequences within bounds	192

^a LOCATE_BETA with hydrophobicity of the nonpolar face ≥ 10 kcal/mol, hydrophobic moment ≥ 1 , minimal length ≥ 10 , proline break.

expansion of the "lipid pocket" and as the binding site for the dimeric unit by interacting with several of the α helices in the helical domain of the other subunit.

Lamprey LV precursor has been sequenced by Sharrock et al. (11). In order to simplify sequence comparisons among proteins, our position assignments for amino acid residues in lamprey LV omit the 14 residue-long signal sequence. For example, in the numbering scheme used by Sharrock et al. (11), LV-1n extends from residue 15 to residue 707, while in our scheme, LV-1n extends from residues 1–693, LV-1c from residues 694–1060, and LV-2 from residues 1292–1610. The segment 1061–1291 from lamprey vitellogenin, a heavily phosphorylated polyserine domain called phosvitin in higher vertebrates, is not part of the LV structure.

The amino acid sequence of lamprey LV was examined by the program LOCATE and the result compared to the X-ray structure. **Fig. 1, column LV** (highlighted in green) contains the results of this analysis: the LOCATE_BETA and LOCATE_ALPHA/CHARGE programs were used to identify probable amphipathic β strands (Fig. 1A) and positively charged amphipathic α helices (Fig. 1B), respectively. Six well-defined clusters of amphipathic β strands are found within the three LV polypeptide chains (the latter denoted by green brackets): I β :LV:8–123, II β :LV:584–686, III β :LV:800–936, IV β :LV:1001–1059, V β :LV:1350–1469, and VI β :LV:1533–1608. These amphipathic β strand clusters were mapped onto the three-dimensional structure of lamprey LV using molecular graphics, as noted in Methods.

Examination of the X-ray structure shows that: 1) the β B:LV antiparallel β sheet is made up of two separate domains (red brackets) located in polypeptide LV-1c between residues 774–932 and 989–1057; these two domains correspond closely to the location of the amphipathic β strand clusters III β :LV:800–936 and IV β :LV:1001–1059, respectively (see Fig. 1A). 2) The β A:LV antiparallel β sheet also is made up of two separate domains (red brackets) but they are located in two separate polypeptides, at the C-terminal end of LV-1n between residues 584–672, and at the N-terminal end of LV-1c between residues 723–742. The amphipathic β strand cluster II β :LV:584–686 corresponds nicely to the larger of the two domains of the β A:LV antiparallel β sheet, while a single amphipathic β strand marks the smaller domain (see Fig. 1A). 3) The β D:LV antiparallel β sheet is made up of a single domain (red bracket) located in polypeptide LV-2 between residues 1342–1513 that corresponds closely to the amphipathic β strand cluster V β :LV:1350–1469 (see Fig. 1A). 4) Finally, the β C:LV antiparallel β sheet also is made up of a single domain (red bracket) located at the N-terminal end of polypeptide LV-1n between residues 1–183 that corresponds to the amphipathic β strand cluster I β :LV:8–123 (see Fig. 1A).

Three clusters of positively charged amphipathic α helices are found in lamprey LV, the largest being a cluster of 7 helices, II α :LV:313–560. Examination of the X-ray structure of lamprey LV shows that the 17 helices of the α helical domain, α :LV, are located in polypeptide LV-1n be-

tween residues 284–602 (red bracket in Fig. 1B), corresponding closely to II α :LV:313–560.

Thus, although not all of the β strands of the four antiparallel sheets nor all of the α helices of the single α helical domain are amphipathic, many are. As each of the five structural motifs within the lamprey LV sequence appears as a well-defined cluster in Fig. 1, these results provide validation for LOCATE as a tool for defining the amphipathic secondary structural features of proteins homologous to lamprey LV, a result important for interpretation of the analyses that follow in this report.

Protein sequence similarity database searches with hAPOB, lamprey LV precursor, and hMTP

Table 2 is a summary of a FASTA3 protein sequence similarity database search using hAPOB as the query sequence. Any z-score (probability) <0.1 is considered strong evidence of homology (16); all homologous proteins share regions of similar tertiary structure (16). The 16 proteins or fragments shown with the highest z-score in Table 2 are all apoB sequences. Of the next six sequences, all with z-scores $<4.3 \times 10^{-5}$, three are apoB and three are LV precursors (underlined): the latter have z-scores of 2×10^{-11} (frog), 4.2×10^{-11} (lamprey), and 4.3×10^{-5} (chicken) and thus are clearly homologous to hAPOB. LV precursors from *C. elegans* (nematode) have z-scores varying from 0.13 to 0.76 (underlined), while hMTP appears last on the list with a z-score of 1.6 (underlined); phylogenetically distant homologous proteins can have z-scores up to 10 (16).

The right half of Table 2 delineates the region of optimal alignment in apoB-100 for each match; as a guide, a schematic of the hAPOB pentapartite structure has been placed above the alignments. The LV precursors align with the N-terminal α_1 domain and 100–200 contiguous residues of the β_1 domain of hAPOB (frog: residues 1–1023; lamprey: residues 15–929; chicken: residues 19–935). hMTP aligns with residues 98–676 of the N-terminal α_1 domain of hAPOB.

We then performed a FASTA3 protein sequence similarity database search using lamprey LV as the query sequence (data not shown). Eight of the ten sequences with the highest z-score are all LV precursors. Lamprey, chicken, and frog have z-scores of 0 and a trout fragment has a z-score of 2×10^{-34} ; full-length *C. elegans* LV precursors have z-scores that vary from 1.5×10^{-18} to 2.2×10^{-14} , strongly suggesting that nematode LV is homologous and has a tertiary structure similar to the vertebrate lipovitellins (16). Human apoB appears as the 11th protein on the list with a z-score of 8.6×10^{-10} , while hMTP appears last on the list with a z-score of 4.3×10^{-3} ; both proteins, thus, are homologous to lamprey LV. hAPOB has its optimal alignment with residues 1–931 of lamprey LV, while hMTP aligns with residues 80–683 of lamprey LV.

We also performed a FASTA3 protein sequence similarity database search using hMTP as the query sequence (data not shown). Lipovitellins are the first proteins, other than hMTP, to appear on the list and have z-scores strongly suggesting homology to hMTP; hAPOB has a

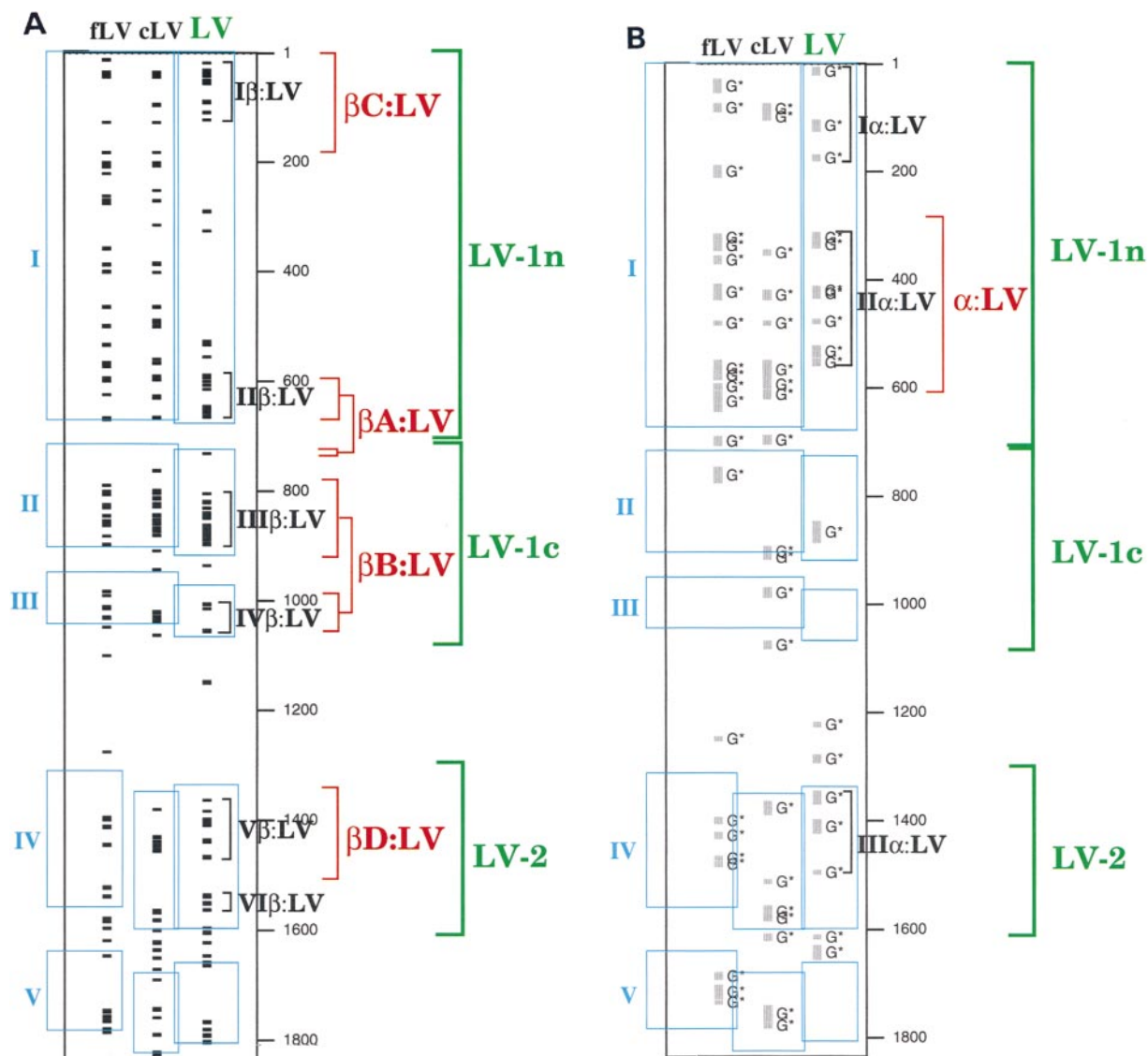


Fig. 1. Mapping of homologue domains, amphipathic motifs and known lamprey LV structural elements to the amino acid sequences of vertebrate (lamprey, chicken, frog) LV precursors. Amphipathic motifs identified by the program composite LOCATE in the amino acid sequences of the vertebrate LV precursors (fLV, frog; cLV, chicken; and LV, lamprey) are plotted within the black box. Amphipathic β strands are marked by black bars (A) and positively charged amphipathic α helices by gray bars (B). The amino acid sequence position for the motifs within the three LV precursors is indicated along the right hand edge of the black box. Color code: Green: (top) LV, lamprey LV precursor sequence (reference sequence); (right) brackets denote the location within the precursor of each processed polypeptide chain of lamprey LV: LV-1n, LV-1c, and LV-2. Red: (right) brackets map the location of each structural domain, defined by X-ray crystallography, within the sequence of lamprey LV (four β :LV β sheets; single α :LV α helical domain). Blue boxes: homologue domains of vertebrate (lamprey, chicken, and frog) LV precursors. Homologue domains were created from one or more sequence blocks as defined in the Methods. Five homologue domains were identified for the vertebrate LV (HD-vLV) and are mapped as blue boxes onto the amphipathic motifs: I (residues 1–680), II (residues 730–920), III (residues 970–1070), IV (residues 1320–1610), and V (residues 1640–1800). A: Composite LOCATE_BETA analysis of vertebrate LV precursors. Plot of all amphipathic β strands (length > 6 residues, hydrophobic moment > 1 , proline termination) having a total hydrophobicity of the nonpolar face > 5 kcal/mol. The location of the four X-ray-determined antiparallel β sheets of lamprey lipovitellin, β C:LV, β A:LV, β B:LV, and β D:LV, are indicated as red brackets to the right; five clusters of amphipathic β sheets in lamprey LV, I-V β :LV, are indicated as black brackets to the right. B: Composite LOCATE_ALPHA/CHARGE analysis of vertebrate LV precursors. Plot of all positively charged amphipathic α helices (length > 10 residues) having a Δ (calculated lipid affinity) > 2 kcal/mol selected using helix termination rules that ignore the presence of hydrophobic residues on the polar face. Class A represents the lipid-associating amphipathic helical domains of apolipoproteins (26). The most distinctive feature of class A is the unique clustering of positively charged residues at the polar–nonpolar interface and negatively charged amino acid residues at the center of the polar face. Two variations on the class A theme, termed Y and G*, have been described (26). The G* class is distinguished by a random radial arrangement of positively and negatively charged residues. Class G* amphipathic helices differ from class G amphipathic helices found in α helical globular proteins in having both a greater hydrophobic moment and a greater nonpolar face hydrophobicity and are postulated to prefer protein–protein interactions over protein–lipid interactions. The capital letter next to each sequence indicates that all of the amphipathic α helices identified are class G*. The location of the X-ray-determined α helical domain of lamprey lipovitellin, α :LV, is indicated as a red bracket to the right; three clusters of positively charged amphipathic α helices in lamprey LV, I-III α :LV, are indicated as black brackets to the right.

TABLE 2. Protein sequence similarity database search with apoB-100

APO B-100 (HUMAN)	z-sc	identity
The best scores		
apolipoprotein B-100 precursor - hum	0	100.0%:-----:;
apolipoprotein B - pig (fragments)	0	68.4%:-----:;
apolipoprotein B - mouse (fragment)	0	71.7%:-----:;
apolipoprotein B - mouse (fragment)	0	71.7%:-----:;
apolipoprotein B - crab-eating maca	0	92.6%:-----:;
apolipoprotein B - golden hamster	0	71.3%:-----:;
apolipoprotein B - rat (fragment)	0	83.6%:-----:;
apolipoprotein B-100 - rabbit (frag)	0	69.6%:-----:;
apolipoprotein B-100 - mouse (fragm	0	66.4%:-----:;
apolipoprotein B-100 - rat (fragmen	0	63.5%:-----:;
apolipoprotein B-100 - chicken (fra	0	62.2%:-----:;
apolipoprotein B-100 - golden hamst	0	65.3%:-----:;
apolipoprotein B - rabbit (fragment	1.6e-37	85.4%:-----:;
apolipoprotein B - chicken (fragmen	6.8e-32	33.6%:-----:;
apolipoprotein B-L - rat (fragment)	3.4e-24	63.8%:-----:;
apolipoprotein B - pig (fragment)	1.7e-18	71.3%:-----:;
<i>vitellogenin</i> A2 precursor - African	2e-15	21.8%:-----:;
apolipoprotein B - rat (fragment)	4e-13	65.0%:-----:;
truncated apolipoprotein B-48 - hum	9.9e-13	100.0%:-----:;
<i>vitellogenin</i> precursor - silver lam	4.2e-11	22.8%:-----:;
apolipoprotein B PI - rat (fragment	1.1e-08	59.7%:-----:;
<i>vitellogenin</i> II precursor - chicken	4.3e-05	21.6%:-----:;
hypothetical protein MG218 - Mycopl	0.0014	21.2%:-----:;
myosin heavy chain beta, cardiac mu	0.0015	21.1%:-----:;
myosin heavy chain - chicken (fragm	0.0021	21.1%:-----:;
myosin heavy chain, smooth muscle,	0.0026	21.2%:-----:;
-----14-----		
<i>vitellogenin</i> 6 precursor - Caenorha	0.13	19.1%:-----:;
-----7-----		
<i>vitellogenin</i> 2 precursor - Caenorha	0.67	20.9%:-----:;
-----2-----		
<i>vitellogenin</i> 5 precursor - Caenorhab	0.76	20.2%:-----:;
-----5-----		
<i>microsomal triglyceride transfer pr</i>	1.6	18.8%:-----:;

Columns from the left: 1) Proteins with best global alignment scores compared to human apoB-100; 2) global alignment z-score (probability), the smaller score, the more similar the alignment; 3) percent sequence identity of aligned region; 4) alignment: globally aligned portion is indicated by a dotted line and the C-terminus of the aligned protein is indicated by a colon.

ApoB-100 schematic in the upper right corner: The apoB-100 α and β domains are aligned with the human apoB-100 sequence (first row). Upper schematic: The α domains are indicated by black rectangles and the β domains by thin lines. Middle schematic: This is an amino acid residue scale that indicates the corresponding apoB residue position. Lower schematic: The amphipathic α helix-rich regions are indicated by dark gray rectangles and the amphipathic β strand-rich regions by light gray rectangles. Regions containing positively charged amphipathic α helices are indicated by + signs.

z-score of 1.1. Lamprey LV precursor has its optimal alignment with residues 80–653 of hMTP and frog LV precursor with residues 81–524, while hAPOB aligns with residues 1–641 of hMTP.

Finally, we performed a FASTA3 protein sequence similarity database search using the first 800 residues of hAPOB (the α_1 domain) as the query sequence (data not shown). The relevant result from this analysis is that the z-score for hMTP is improved to 0.063 from 1.1 and 1.6, supporting homology to the α_1 domain of hAPOB.

Dot matrix local sequence similarity comparisons of hAPOB, lamprey LV precursor, and hMTP

As a second method for local sequence similarity comparisons, we performed dot matrix analyses comparing the

N-terminal 1400 residues of hAPOB, lamprey LV precursor, and hMTP; the results are shown in Fig. 2. The locations of all lines of similarity given below are approximate.

Figure 2A compares hAPOB to lamprey LV precursor. Three lines of similarity appear on the diagonal aligning residues 30–1040 (hAPOB) with residues 1–1100 (lamprey LV). The first diagonal line, with a raw similarity score >200, aligns residues 30–725 (hAPOB) with residues 1–715 (lamprey LV). The other two diagonal lines, with raw similarity scores >50, align residues 730–960 with 715–925 and 960–1040 with 1020–1100, hAPOB versus lamprey LV, respectively.

Figure 2B compares lamprey LV precursor to hMTP. Two lines of similarity appear on the diagonal aligning residues 100–540 (hMTP) with residues 90–530 (lamprey

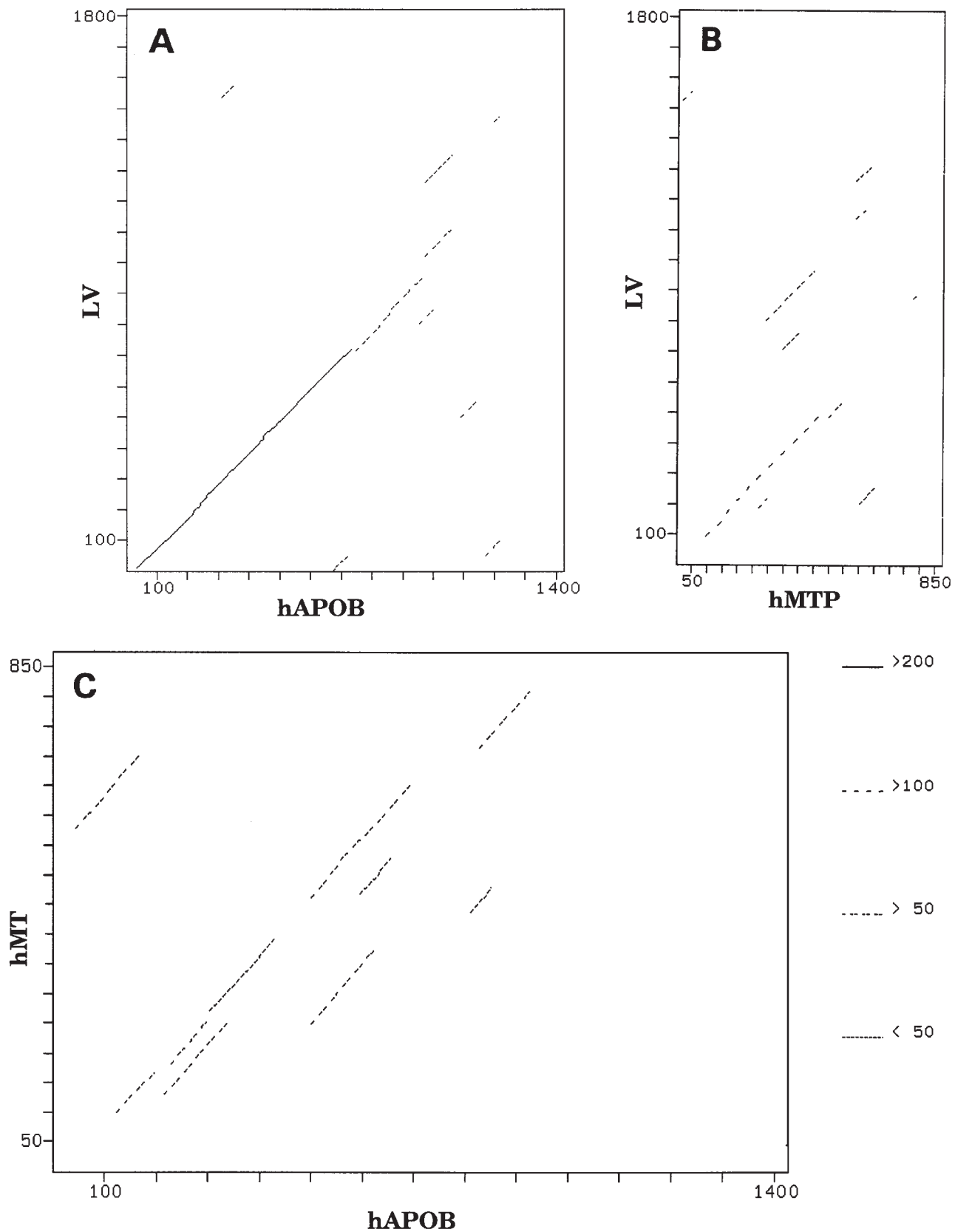


Fig. 2. Dot matrix comparisons of A: lamprey LV (Y-axis) versus hAPOB (residues 1–1400); B: lamprey LV (Y-axis) versus hMTP; C: hMTP (Y-axis) versus hAPOB (residues 1–1400). Plots were made using PLALIGN from the FASTA package (16).

LV). The first diagonal line, with a raw similarity score >100 , matches residues 100–480 for both hMTP and lamprey LV (see the results of Fig. 4). The second diagonal line, with a raw similarity score >50 , matches residues 500–540 (hMTP) with 490–530 (lamprey LV).

Finally, Fig. 2C compares hAPOB to hMTP. Five lines of similarity appear on or near the diagonal aligning residues 120–920 (apoB-100) to residues 100–820 (MTP); two major gaps between the five lines are seen at residues 430–500 (see the results of Fig. 5) and 690–810 of apoB-100. Three of the diagonal lines, with raw similarity scores >50 , correspond to residues 120–200, 230–300, and 500–690. A fourth, slightly off-axis, diagonal line, also with a raw similarity score >50 , corresponds to residues 720–820 and 820–920 of hMTP and hAPOB, respectively; this region of homology will play a prominent role in later discussions (see Fig. 5). The fifth diagonal line, with a raw similarity score <50 , corresponds to residues 300–430 of apoB-100.

Mapping of the structure of lamprey LV onto LV precursors from other species

Based upon the FASTA3 analysis of lamprey LV precursor, frog and chicken LV precursors have very high similarity scores to lamprey LV precursor; the similarity score of the invertebrate LV precursor from the nematode, *C. elegans*, to lamprey LV precursor is lower but still suggests homology. In order to increase the power of our later mapping of the structure of LV onto the amphipathic structures of hAPOB and hMTP, we performed a multiple local alignment of sequence blocks for LV precursors using the program MACAW (results not shown). For comparison of local sequence similarity to the position of amphipathic structural motifs, we created homologue domains from one or more sequence blocks as described in Methods.

Five homologue domains (see blue boxes in Fig. 1) were defined for the three vertebrate lipovitellins: HD-vLV-I:1–680, HD-vLV-II:730–920, HD-vLV-III:970–1070, HD-vLV-IV:1320–1610, and HD-vLV-V:1640–1800. These five homologue domains are related to the three polypeptide chains of lamprey LV in several ways: *i*) the juncture of the processed peptides LV-1n and LV-1c (green brackets in Fig. 1) is between residues 693 and 694 in lamprey, corresponding to the juncture between HD-vLV-I:1–680 and HD-vLV-II:730–920; *ii*) the C-terminus of HD-vLV-III:970–1070 corresponds to the C-terminus of LV-1c (residue 1074); *iii*) HD-vLV-IV:1320–1610 corresponds closely to LV-2 (residues 1292–1610). Finally, the gap between HD-vLV-III:970–1070 and HD-vLV-IV:1320–1610 represents the polyserine domain. The nature and function of HD-vLV-V:1640–1800, a sequence not part of the processed LV structure, is unclear.

Multiple local alignments of sequence blocks for all four LV precursors were performed but the percentage of the sequences covered by homologue domains for all four precursors is reduced significantly from that seen for the three vertebrate precursors, reflecting the phylogenetic distance between vertebrates and invertebrates (data not shown).

We then used the results of the multiple alignment to

map the known secondary structure of lamprey LV onto the amphipathic secondary structure predicted for the other LV precursors by composite LOCATE analysis (Fig. 1). For each composite analysis, the five vLV homologue domains defined for the three vertebrate LV precursors are enclosed by blue boxes (blue roman numerals).

Figure 1A is a composite LOCATE_BETA analysis of the vLV precursors: 1) HD-vLV-I:1–680 contains within it the N-terminal and the bulk of the C-terminal extremes of the β C:LV and β A:LV structural domains, respectively, denoted by red brackets. The I β :LV and II β :LV clusters of amphipathic β strands that map to these two structural domains are denoted by black brackets; corresponding, but less well defined, clusters of amphipathic β strands are present in the chicken and frog. 2) HD-vLV-II:730–920 contains the N-terminal two-thirds of the β B:LV structural domain, denoted by a red bracket. The III β :LV cluster of amphipathic β strands that maps to this structural domain is denoted by a black bracket; corresponding, equally well defined, clusters of amphipathic β strands are present in the chicken and frog. This homologue domain also contains the small C-terminal portion of β A:LV. 3) HD-vLV-III:970–1070 contains the C-terminal one-third of the β B:LV structural domain, denoted by a red bracket. The IV β :LV cluster of amphipathic β strands that maps to this structural domain is denoted by a black bracket; corresponding, equally well defined, clusters of amphipathic β strands are present in the chicken and frog. 4) HD-vLV-IV:1320–1610 contains the β D:LV structural domain, denoted by a red bracket. The V β :LV cluster of amphipathic β strands that maps to this structural domain is denoted by a black bracket; corresponding clusters of amphipathic β strands are present in the chicken and frog.

Figure 1B is a composite LOCATE_ALPHA/CHARGE analysis of the vLV precursors. The first vLV:homologue domain, HD-vLV-I:1–680, contains within it the 17 helices of the α :LV:284–602 structural domain of lamprey LV, denoted by the red bracket. The large positively charged amphipathic α helix cluster, II α :LV:313–560, that maps to this structural domain is denoted by a black bracket; equally well defined clusters of positively charged amphipathic α helices are present in the same region of chicken and frog.

As six disulfide bonds are found in the α_1 domain (residues 1–800) of apoB-100, we determined the location of all cysteine residues with a composite LOCATE_MARK analysis of the vLV precursors (data not shown). These analyses indicate that: *i*) six cysteines between residues 150–550 are conserved in the three vLV precursors, and *ii*) three of the six, located between residues 150–200, are conserved in all four precursors (see Discussion).

Mapping of the structure of lamprey LV onto hAPOB

Next, multiple local alignments of sequence blocks between apoB-100 and LV precursors were performed using the program MACAW (results not shown). Four homologue domains (see blue boxes in Fig. 3) were defined: HD-vLV:B-I:1–180, HD-vLV:B-II:450–680, HD-vLV:B-III:710–860 and HD-vLV:B-IV:980–1020 (vLV), HD-vLV:B-IV:925–986 (hAPOB).

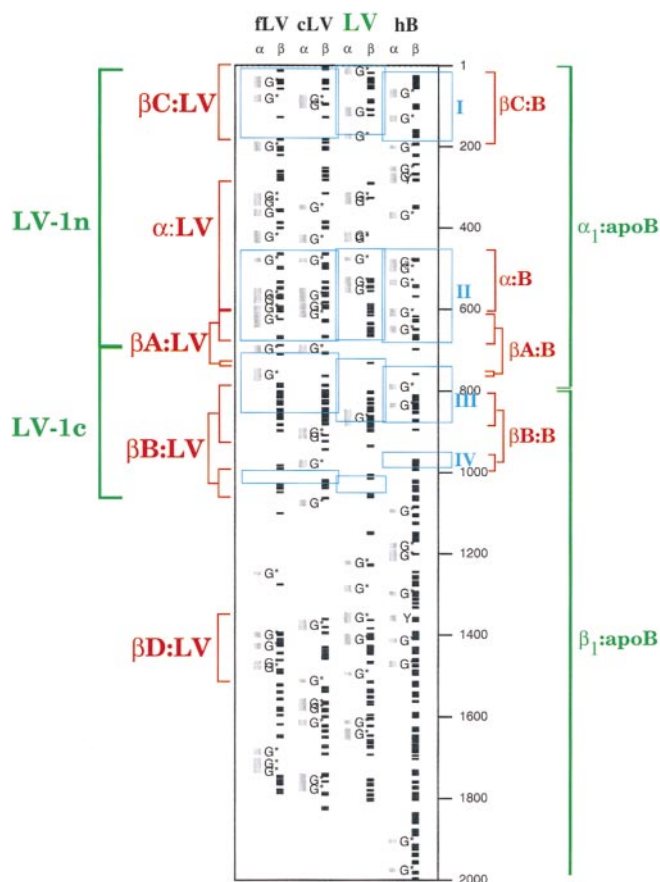


Fig. 3. Mapping of homologue domains, amphipathic motifs, and known lamprey LV structural elements to the amino acid sequences of vertebrate (lamprey, chicken, frog) LV precursors and the first 2000 residues of hAPOB. Amphipathic motifs identified by the program composite LOCATE in the amino acid sequences of the vertebrate LV precursors (fLV, frog; cLV, chicken, and LV, lamprey) are plotted within the black box. Positively charged amphipathic α helices are marked by gray bars (labeled α) and amphipathic β strands by black bars (labeled β). The amino acid sequence positions for the motifs within the three LV precursors and hAPOB are indicated along the right hand edge of the black box. Color code: Green: (top) LV, lamprey LV precursor sequence (reference sequence); (left) brackets denote the location within the precursor of the processed polypeptide chains of lamprey LV, LV-1n, and LV-1c; and (right) brackets denote the location of the α_1 and β_1 domains of hAPOB. Red: (left) brackets map the location of each structural domain, defined by X-ray crystallography, within the sequence of lamprey LV (four β :LV β sheets; single α :LV α helical domain); and (right) brackets map the corresponding homologues of these elements mapped onto hAPOB (β :B, β sheet domains; α :B, α helical domain). Blue boxes: Homologue domains of vertebrate LV precursors and hAPOB. Homologue domains were created as described in the Methods. Four homologue domains were identified for the vertebrate LV and hAPOB (HD-vLV:B) and are mapped as blue boxes onto the amphipathic motifs: I (residues 1–180), II (residues 450–680), III (residues 710–860) and IV (LV precursors, residues 980–1020; hAPOB, residues 925–986). For each protein, the left hand column, labeled α , represents a plot (gray bars) of all positively charged amphipathic α helices (length > 10 residues) having a Δ (calculated lipid affinity) > 2 kcal/mol selected using helix termination rules that ignore the presence of hydrophobic residues on the polar face. The capital letter next to each identified amphipathic helix indicates its amphipathic α helix class (all but two, in hAPOB, are G*). For each protein, the right hand column, labeled β , represents a plot (black bars) of amphipathic β strands (length > 6 residues, hydrophobic moment > 1, proline termination) having a total hydrophobicity of the nonpolar face > 5 kcal/mol.

Then, multiple local alignments of sequence blocks between all four LV precursors and apoB-100 were performed but the extent of the sequences representing homologue domains between hAPOB and all four LV precursors is reduced significantly from that seen between hAPOB and the three vertebrate precursors, again reflecting the phylogenetic distance between vertebrates and invertebrates (data not shown).

We then used the results of the multiple alignment to map the known secondary structure of lamprey LV onto the amphipathic secondary structure predicted for both the vertebrate LV precursors and hAPOB by composite LOCATE analysis. Figure 3 represents composite LOCATE_ALPHA/CHARGE and LOCATE_BETA analyses of the three vertebrate LV precursors and the first 2000 residues of hAPOB, searching for probable positively charged amphipathic α helices (α) and amphipathic β strands (β), respectively. The four vLV:B homologue domains defined for the three vertebrate LV precursors and hAPOB are enclosed by blue boxes (blue roman numerals). Red brackets map structural domains: (left) the location of each structural domain defined by X-ray crystallography in the sequence of lamprey LV and (right) the corresponding homologues of these domains in hAPOB. Three of the four antiparallel β sheet domains, and the N-terminal half

of the α helical domain, of lamprey LV map through the four homologue domains onto apoB-100.

HD I:1–180 maps to: *i*) the β C:LV antiparallel β sheet region of lamprey LV, residues 1–183, and *ii*) a corresponding cluster of amphipathic β strands in hAPOB (residues 16–186, β C:B). Of the six disulfide bonds found in the α_1 domain (residues 1–800) of apoB-100, both cysteines of C159–C185, are conserved not only in all three vLV precursors but in *C. elegans* as well (data not shown); interestingly, this disulfide bond is not among those required for MTP-dependent lipid transfer to nascent apoB and subsequent secretion of apoB-containing lipoproteins (2–5).

HD II:450–680, maps in its C-terminal portion to: *i*) the first of the two separate sequences that make up the β A:LV antiparallel β sheet domain of lamprey LV, residues 584–672, and *ii*) a corresponding well defined cluster of amphipathic β strands in hAPOB (residues 615–676, β A:B). Further, HD II:450–680 maps at its N-terminal end to: *i*) the C-terminal half of the α :LV α helical domain of lamprey LV, residues 284–602, and *ii*) a corresponding cluster of positively charged amphipathic α helices in hAPOB (residues 477–618, α :B).

HD III:710–860, maps to: *i*) the first of the two separate sequences that make up the β B:LV antiparallel β sheet do-

main of lamprey LV, residues 774–932 and *ii*) a corresponding, very well defined, cluster of amphipathic β strands in hAPOB (residues 808–897, β B:B). This homologue domain also maps to the second of the two separate sequences that make up the β A:LV antiparallel β sheet domain of lamprey LV, residues 723–742.

Finally, HD IV:980–1020 maps to: *j*) the second of the two separate sequences that make up the β B:LV antiparallel β sheet domain of lamprey LV, residues 989–1057 and *ii*) a corresponding cluster of amphipathic β strands in hAPOB (residues 967–996, β B:B).

The antiparallel β sheet, β D:LV, residues 1342–1513, forms the bottom of the triangular “lipid pocket” in the crystal structure of lamprey LV (10, 11). Although hAPOB possesses almost continuous tandem amphipathic β strands between residues 1000–2000, multiple local alignment of the three vertebrate LV precursors with hAPOB (Fig. 3) fails to demonstrate homology to the β D:LV amphipathic β strand cluster. However, local alignment analysis of hAPOB with individual LV precursors suggests modest sequence similarity between the sequences near residue 1150 of hAPOB and residue 1400 of lamprey or chicken (but not frog) LV precursors, respectively (data not shown).

Mapping of the structure of lamprey LV onto hMTP

We then used similar techniques to explore further the suggestion by FASTA3 analysis that hAPOB and hMTP may be homologous. Given the greater sequence similarity between lipovitellins and hMTP (data not shown) than between hAPOB and hMTP (Table 2), we began by mapping the secondary structure of lamprey LV onto the amphipathic structure of hMTP. To increase the significance, we used multiple alignment and composite LOCATE analyses to map the known secondary structure of lamprey LV onto the amphipathic secondary structures predicted for both the other vertebrate LV precursors and hMTP.

MTP has been reported to have sequence homology to the α clusters of chicken and frog lipovitellins (17). To examine the implications of this possibility, multiple alignments of sequence blocks for the three vertebrate LV precursors and hMTP were performed (data not shown). Three homologue domains (see blue boxes in Fig. 4) were defined: HD-vLV:MTP-I:80–155 (LV precursors), HD-vLV:MTP-I:98–166 (hMTP), HD-vLV:MTP-II:–340–480, and HD-LV:MTP-III:1030–1050 (LV precursors), HD-vLV:MTP-III:722–741 (hMTP).

Multiple alignments of sequence blocks for all four LV precursors and hMTP result in two homologue domains of similar size and location to those found for the three vertebrate LV precursors and hMTP (results not shown).

In Fig. 4, the three vLV:MTP homologue domains defined by this analysis are indicated by blue boxes (blue roman numerals) superimposed upon the composite LOCATE analyses of probable positively charged amphipathic α helices (α) and amphipathic β strands (β). Note that hMTP contains two large and well-defined clusters of probable amphipathic β strands, I β :MTP:16–243 and II β :MTP:722–802 (black brackets). Red brackets indicate structural domains: (left) the location of each structural

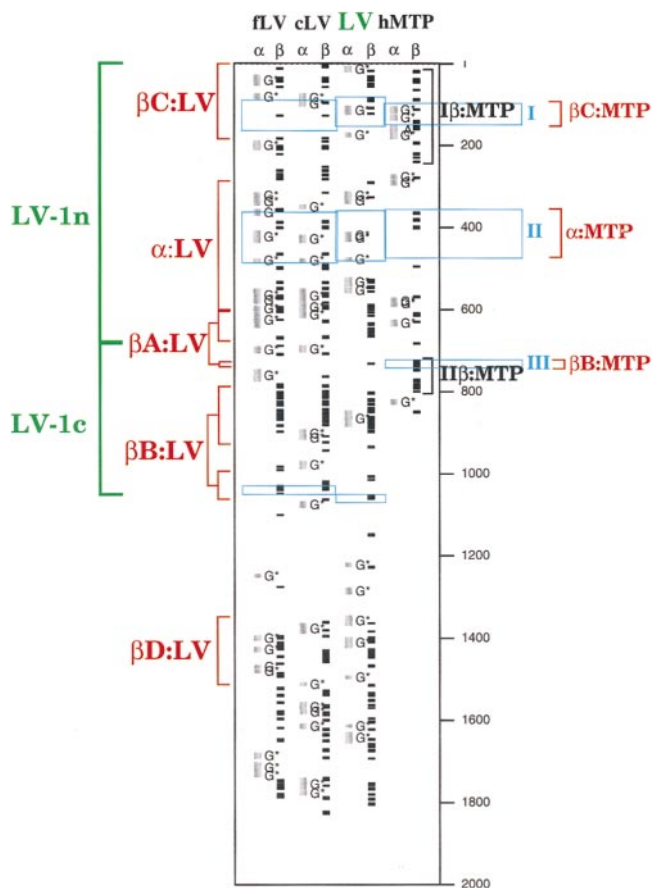


Fig. 4. Mapping of homologue domains, amphipathic motifs, and known lamprey LV structural elements to the amino acid sequences of vertebrate (lamprey, chicken, frog) LV precursors and hMTP. Amphipathic motifs identified by the program composite LOCATE in the amino acid sequences of the vertebrate LV precursors and hMTP are plotted within the black box. Positively charged amphipathic α helices are marked by gray bars (labeled α) and amphipathic β strands by black bars (labeled β). The amino acid sequence positions for the motifs within the three LV precursors and hMTP are indicated along the right hand edge of the black box. Color code: Green: (top) LV, lamprey LV precursor sequence (reference sequence); (left) brackets denote the location within the precursor of the processed polypeptide chains of lamprey LV, LV-1a, and LV-1c. Red: (left) brackets map the location of each structural domain, defined by X-ray crystallography, within the sequence of lamprey LV (four β :LV β sheets; single α :LV α helical domain); and (right) brackets map the corresponding homologues of these elements mapped onto hMTP (β :MTP, β sheet domains; α :MTP, α helical domain). Blue boxes: homologue domains of vertebrate LV precursors and hMTP. Homologue domains were created as described in Methods. Three homologue domains were identified for the vertebrate LV and hMTP (HD-vLV:MTP) and are mapped as blue boxes onto the amphipathic motifs: the largest, II, is located between residues 340–480. For each protein, the left hand column, labeled α , represent a plot (gray bars) of all positively charged amphipathic α helices (length > 10 residues) having a Δ (calculated lipid affinity) > 2 kcal/mol selected using helix termination rules that ignore the presence of hydrophobic residues on the polar face. The capital letter next to each identified amphipathic helix indicates its amphipathic α helix class (all are G*). For each protein, the right hand column, labeled β , represents a plot (black bars) of amphipathic β strands (length > 6 residues, hydrophobic moment > 1, proline termination) having a total hydrophobicity of the non-polar face > 5 kcal/mol.

domain defined by X-ray crystallography is mapped onto the sequence of lamprey LV; (right) the corresponding homologues of these domains are mapped onto hMTP.

Fragments of two of the four antiparallel β sheet domains and approximately one-third of the α helical domain of lamprey LV map through the three homologue domains onto hMTP: 1) HD I:98–166 maps to: *i)* the C-terminal end of the β C:LV antiparallel β sheet region of lamprey LV, and *ii)* the central section of amphipathic β strand cluster β B:MTP:16–243 (residues 98–166, β C:MTP). 2) HD II:–340–480 maps to: *i)* the upper central region of the α :LV α helical domain of lamprey LV, and *ii)* a region of hMTP (residues 340–480, α :MTP) that contains no identifiable positively charged amphipathic α helices but does contain a well-defined cluster of negatively charged amphipathic α helices (see Fig. 5). 3) Finally, HD III:1030–1050 maps to: *i)* the C-terminal end of the second of the two separate sequences that make up the β B:LV antiparallel β sheet domain of lamprey LV, residues

989–1057, and *ii)* the N-terminal end of II β :MTP:722–802 (residues 722–741, β B:MTP).

Analysis of the structural similarities of hAPOB and hMTP

Finally, a local alignment of sequence blocks for hMTP versus hAPOB was performed (data not shown). Four widely separated homologue domains (see blue boxes in Fig. 5) are defined: HD-MTP:B-I:21–43, HD-MTP:B-II:106–135, HD-MTP:B-III:–530–590 and HD-MTP:B-IV: 802–899 (hAPOB), HD-MTP:B-III:715–811 (hMTP). We then mapped the predicted amphipathic structure of hAPOB onto the predicted amphipathic structure of hMTP using local alignment and composite LOCATE analyses. In Fig. 5 the four MTP:B homologue domains defined by this analysis are indicated by blue boxes (blue roman numerals) superimposed upon the composite LOCATE analyses of positively charged amphipathic α helices (+ α), negatively charged amphipathic α helices (– α), amphipathic β strands (β), and cysteine residues (C). The results of

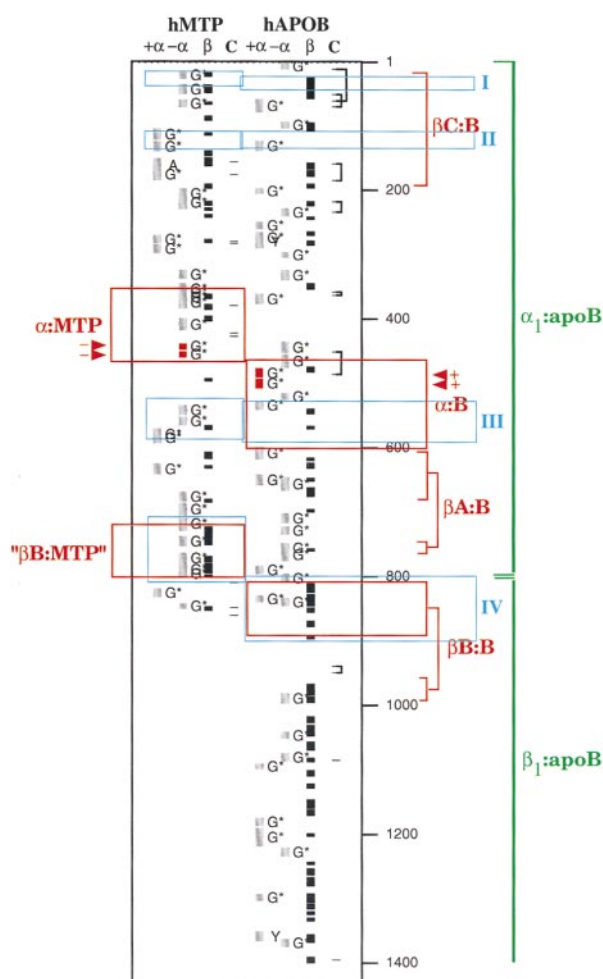


Fig. 5. Mapping of homologue domains, amphipathic motifs, cysteine residues, and hAPOB homologues of known lamprey LV structural elements to the amino acid sequences of hMTP and hAPOB. Amphipathic motifs and the position of cysteine residues identified by the program composite LOCATE in the amino acid sequences of hMTP and hAPOB are plotted within the black box. The locations of positively charged amphipathic α helices are represented by gray bars labeled + α (left column), negatively charged amphipathic α helices by gray bars, labeled – α (left middle column), amphipathic β strands by black bars, labeled β (right middle column), and cysteine residues by black lines with known disulfide bonds in hAPOB connected by brackets, labeled C (right column). The amino acid sequence positions for the motifs within hMTP and hAPOB are indicated along the right hand edge of the black box. Color code: Green: (right) brackets denote the location of the α_1 and β_1 domains of hAPOB. Red: the location of the corresponding homologues of each defined structural domain of lamprey LV are indicated by brackets or, in two cases, by boxes. Blue boxes: homologue domains of hMTP and hAPOB. Homologue domains were created as described in Methods. Four homologue domains were identified for hMTP and hAPOB (HD-MTP:B) and are mapped as blue boxes onto the amphipathic motifs. For each protein, the column labeled + α represents a plot of all positively charged amphipathic α helices (length > 10 residues) having a Δ (calculated lipid affinity) > 2 kcal/mol selected using helix termination rules that ignore the presence of hydrophobic residues on the polar face. For each protein, the column labeled – α represents a plot (gray bars) of all negatively charged amphipathic α helices (length > 10 residues) having a Δ (calculated lipid affinity) > 4 kcal/mol selected using helix termination rules that ignore the presence of hydrophobic residues on the polar face. For each amphipathic α helix plot, the capital letter next to each identified amphipathic helix indicates its amphipathic α helix class. For each protein, the column labeled β represents a plot (black bars) of amphipathic β strands (length > 6 residues, hydrophobic moment > 1, proline termination) having a total hydrophobicity of the nonpolar face > 5 kcal/mol. Finally, for each protein, the column labeled C represents cysteine residues and, for hAPOB, indicates disulfide bonds. The α helical domains of hMTP (α :MTP) and of hAPOB (α :B) are denoted by red boxes. Also denoted by red boxes are: 1) the first of the two separate sequences that make up the cluster of amphipathic β strands in hAPOB

(residues 808–897, β B:B) that are homologous to the β B:LV antiparallel β sheet domain of lamprey LV, and 2) the corresponding (homologous) cluster of amphipathic β strands in hMTP, “ β B:MTP.” All other domains in hAPOB that are homologues to the structural domains defined by X-ray crystallography in the sequence of lamprey LV are denoted by red brackets. The red arrowheads on the right marked with + signs denote two positively charged amphipathic α helices (red bars) at the C-terminal end of α :B; the two red arrowheads on the left marked with – signs denote two negatively charged amphipathic α helices (red bars) at the N-terminal end of α :MTP.

these analyses are: 1) HD I:21–43 and HD II:106–135 map the amphipathic β strand cluster of hAPOB, β C:B, that is homologous to β C:LV of lamprey LV to the N-terminal and central regions of the extensive N-terminal amphipathic β strand cluster of hMTP, I β :MTP (see Fig. 4). Interestingly, conservation of the region in the vicinity of residue 120 is shared between lamprey LV, hAPOB, and hMTP. 2) HD III:530–590 maps the C-terminal half of the α :B region of hAPOB onto hMTP. 3) Most strikingly, HD IV precisely maps the first of the two separate sequences that make up the amphipathic β strand cluster of hAPOB, β B:B, that is homologous to β B:LV of lamprey LV, residues 802–899 (red box), to the II β :715–811 amphipathic β strand cluster of hMTP (see Fig. 4), designated “ β B:MTP” (red box) in Fig. 5. This off-axis region of homology was also identified by the dot matrix local sequence similarity analysis of Fig. 2C.

In addition to these analyses, the upper pair of red boxes in Fig. 5 indicate the locations of the two domains homologous to the α helical domain of lamprey LV, α :MTP of hMTP and α :B of hAPOB. It is worthy of note that: 1) α :MTP, devoid of positively charged amphipathic α helices, contains numerous negatively charged amphipathic α helices, and 2) two positively charged amphipathic α helices (residues 477–491 and 492–508, red bars) at the N-terminal end of the α :B domain, indicated by red arrow-

heads on the right, are in apposition to two negatively charged amphipathic α helices (residues 438–447 and 450–559, red bars) at the C-terminal end of the α :MTP domain, indicated by red arrowheads on the left.

DISCUSSION

Experimental validation

In this report, we suggest that several functionally important domains in the first 1000 residues of lamprey LV are structurally homologous to corresponding domains in hAPOB. This conclusion is based to a significant extent upon the demonstration that domains of local sequence similarity contain common amphipathic structural motifs identified using LOCATE analysis. Thus it is important to compare the NH_2 - α_1 - β_1 - α_2 - β_2 - α_3 -COOH pentapartite model derived for apoB-100 using these analytical tools (6, 7) to known experimental data.

Figure 6 represents a composite LOCATE analysis of the N-terminal 2000 residues of hAPOB that graphically illustrates the distribution of amphipathic motifs within the α_1 and β_1 domains of hAPOB, residues 1–800 and 801–2000, respectively (6, 7). The first three columns labeled α , $+\alpha$, and C define three features characteristic of the α domain: a) this domain contains a dense cluster of poten-

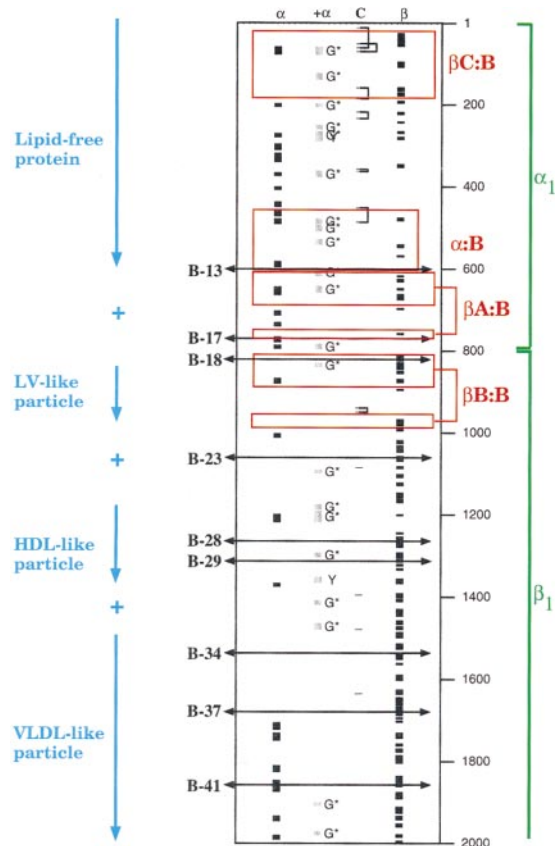


Fig. 6. Correlation of apoB-containing lipoprotein particle biosynthesis with the distribution of amphipathic motifs along the N-terminal 2000 residues of hAPOB and hAPOB homologues of known lamprey LV structural elements. Amphipathic motifs and the position of cysteine residues identified by the program composite LOCATE in the amino acid sequence of the first 2000 residues of hAPOB are plotted within the black box. Amphipathic α helices with high lipid affinity are denoted by black bars labeled α (left column); positively charged amphipathic α helices are denoted by gray bars labeled $+\alpha$ (left middle column), amphipathic β strands are denoted by black bars labeled β (right middle column), and cysteine residues are denoted by black lines labeled C, with known disulfide bonds in hAPOB connected by brackets (right column). The amino acid sequence positions for the motifs within the hAPOB sequence are indicated along the right hand edge of the black box. The horizontal double-headed arrows indicate the position of C-terminal truncations of nine apoB mutants, B-13, B-17, B-18, B-23, B-28, B-29, B-34, B-37, and B-41 (21–25). Color code: Green: (right) brackets denote the location of the α_1 and β_1 domains of hAPOB. Red: boxes map the location of the corresponding homologues of each defined structural domain of lamprey LV mapped onto hAPOB (β :B, β sheet domains; α :B, α helical domain). Blue: the vertical blue arrows indicate the approximate, progressively longer, intact portions of the N-terminal domain of the hAPOB sequence required for lipid-free protein and lipovitellin-like (proteolipoprotein-like), HDL-like and VLDL-like particle formation (21–25). LOCATE_ALPHA analysis (α): Plot of all amphipathic α helices (length > 10 residues) having a Δ (calculated lipid affinity) > 8 kcal/mol. LOCATE_ALPHA/CHARGE analysis ($+\alpha$): Plot of all positively charged amphipathic α helices (length > 10 residues) having a Δ (calculated lipid affinity) > 2 kcal/mol selected using helix termination rules that ignore the presence of hydrophobic residues on the polar face. The positively charged amphipathic α helix cluster in hAPOB, α :B, structurally homologous to the antiparallel α helical cluster in lamprey lipovitellin, is marked by a red box labeled on the right. The capital letter next to each sequence indicates the amphipathic α helix class (all but 2, of a total of 21, are G*). LOCATE_AA analysis (C): Plot of all cysteine residues in apoB-100 indicating the locations of the disulfide bonds. LOCATE_BETA analysis (β): Plot of all amphipathic β strands (length > 6 residues, hydrophobic moment > 1, proline termination) having a total hydrophobicity of the nonpolar face > 5 kcal/mol. The amphipathic β strands in hAPOB, β C:B, β A:B and β B:B, structurally homologous to the first three antiparallel β sheets of lamprey lipovitellin, are marked by red boxes and labeled on the right.

sulfide bonds. LOCATE_BETA analysis (β): Plot of all amphipathic β strands (length > 6 residues, hydrophobic moment > 1, proline termination) having a total hydrophobicity of the nonpolar face > 5 kcal/mol. The amphipathic β strands in hAPOB, β C:B, β A:B and β B:B, structurally homologous to the first three antiparallel β sheets of lamprey lipovitellin, are marked by red boxes and labeled on the right.

lipid-associating G* amphipathic α helices (column α), *b*) many of these G* amphipathic α helices are positively charged (column $+\alpha$), and *c*) the domain is rich in disulfide bonds suggesting a folded structure (column C). The last column labeled β defines the distinguishing feature of the β_1 domain, an almost continuous run of amphipathic β strands; amphipathic β strands are also present in the α_1 domain but do not dominate its structure to the extent they do in the β_1 domain (6, 7).

Expression of progressively smaller C-terminal truncated forms of apoB-100 in cell culture (indicated by horizontal double-headed arrows in Fig. 6) have demonstrated that near the N-terminal end of the β_1 domain the percentage of the truncated apoB-100 associated with lipid approaches zero. Expression of C-terminally truncated forms of apoB in HepG2 and McA-RH 7777 cells demonstrated that, with stepwise truncation, progressively more apoB was recovered in the $d > 1.25$ g/ml fraction of cell culture medium (21). These studies showed that 100% of apoB-41, 48% of apoB-29, 20% of apoB-23, and 18% of apoB-17 were recovered in the $d < 1.25$ g/ml fraction of culture medium (Fig. 6). On the other hand, no apoB-13 was detected in the lipoprotein fraction but was recovered in the $d > 1.25$ g/ml fraction (21). In a separate study using the McA-RH7777 cell line, apoB-18 and apoB-23 were recovered in $d > 1.23$ g/ml, whereas particles containing apoB-28, apoB-31, apoB-37, apoB-48, and apoB-53 formed discrete lipoprotein particles (22). In HepG2 cells transfected with minigenes coding for apoB-41, apoB-29, and apoB-23, these proteins were secreted on HDL-like particles (23) (Fig. 6). In yet another study, apoB-17 was recovered from the media free of lipid but when combined with phospholipid formed discoidal complexes, suggesting the presence of amphipathic motifs within the apoB-17 sequence (24).

The above studies indicate that there is a direct relationship between the size of the nascent polypeptide and the size of the particle that is being assembled. These studies also suggest that there is a threshold in apoB size (somewhere between apoB-23 and apoB-28) under which the polypeptide cannot form a lipoprotein particle (Fig. 6). Further, McLeod et al. (25) have shown that apoB-29 forms an HDL-like particle when expressed in McA-RH7777, but VLDL particle formation requires inclusion of multiple amphipathic β strands from the β_1 domain beyond apoB-29 (Fig. 6). Other studies have shown that the diameters of secreted lipoprotein particles are linearly proportional to the length of C-terminal truncated apoB-100 fragments until a minimal apoB fragment length is reached (1, 22); the linearity falls off near the end of the β_1 domain, between residues 1100–1300 (Fig. 6). Taken together, truncation studies suggest a rather abrupt change in the nature of the interactions between apoB and bulk lipid at the α_1 – β_1 boundary. They support our suggestions (6, 7) that: *a*) a major portion of the α_1 domain is globular and not an integral part of the bulk lipid of the apoB-containing lipoprotein particles; *b*) the β_1 domain is extended and is integrated into the bulk lipid of these same particles.

Thus we are confident that similarities of the α_1 domain of hAPOB to lamprey LV and hMTP observed by LOCATE analysis indeed reflect local regions of common structural homology between these diverse proteins. This confidence is increased by the ability of LOCATE to predict rather accurately the locations of the β sheets and the 17-helix bundle in the three-dimensional structure of lamprey LV (Fig. 1).

“Lipid pocket” hypothesis

The most widely quoted mechanism for the physical assembly of lipid particles containing apoB is the budding oil droplet model (1). In this model, the N-terminal portion of apoB is believed to become embedded in the inner monolayer of the endoplasmic reticulum (ER) membrane, where it nucleates an oil droplet from the supersaturated rough RE membranes. Upon completion of apoB synthesis, this oil-droplet emulsion is detached from the bilayer to form the nascent lipoprotein particle. One weakness of this model is that many investigators have searched extensively by electron microscopy for inner rough ER membrane blebs in liver microsomal preparations but have never found them (1).

Further, thermodynamic considerations make it unlikely that lipoproteins assemble through the wholesale remodeling or dismantling of plasma membrane bilayers as suggested by the budding oil droplet model for assembly of apoB-containing lipoprotein particles (1). Rather, it seems more reasonable that apolipoproteins may accrete the lipid of their corresponding lipoprotein particles gradually. Based upon the structural and sequence homologies between hAPOB and lamprey LV described in this report, we propose the following general hypothesis: a LV-like structural intermediate containing a lipid-binding pocket is formed by the N-terminal 22–26% of apoB (residues 1000–1200) and functions as a lipid nucleation site for the initiation of assembly of apoB-containing lipoprotein particles.

The structure of the lamprey LV “lipid pocket” is schematically represented in Figs. 7A and B. As there is a spacing of 4.7 Å between each β strand in an antiparallel β sheet, $\beta A:LV$ and $\beta B:LV$ contain 7- and 8-stranded antiparallel β sheets with widths of 33 and 38 Å, respectively; these are approximately the dimensions of the hydrophobic core of a phospholipid bilayer. Extrapolating from the known structure of plasma lipoprotein bilayer discs (26), we suggest that the lamprey LV lipid pocket most likely encloses a small phospholipid bilayer (Figs. 7A and B) split by a central triacylglycerol phase or “lens” (Fig. 7B).

The following observations are relevant for models of the assembly of apoB-containing particles. *i*) The first 1000 amino acid residues of hAPOB and LV are highly similar; thus the two proteins are homologous, meaning they share extensive regions of common local tertiary structure. *ii*) Clusters of amphipathic motifs of lamprey LV are mapped cleanly onto other vertebrate LV. *iii*) Amphipathic β clusters, $\beta A:LV$, $\beta B:LV$, and $\beta C:LV$, have well-defined homologues in apoB, while the bottom amphipathic β sheet of the LV “lipid pocket,” $\beta D:LV$, has no clear homologue. *iv*) The largest of the two sections of the $\beta B:B$ amphipathic β strand cluster of hAPOB maps pre-

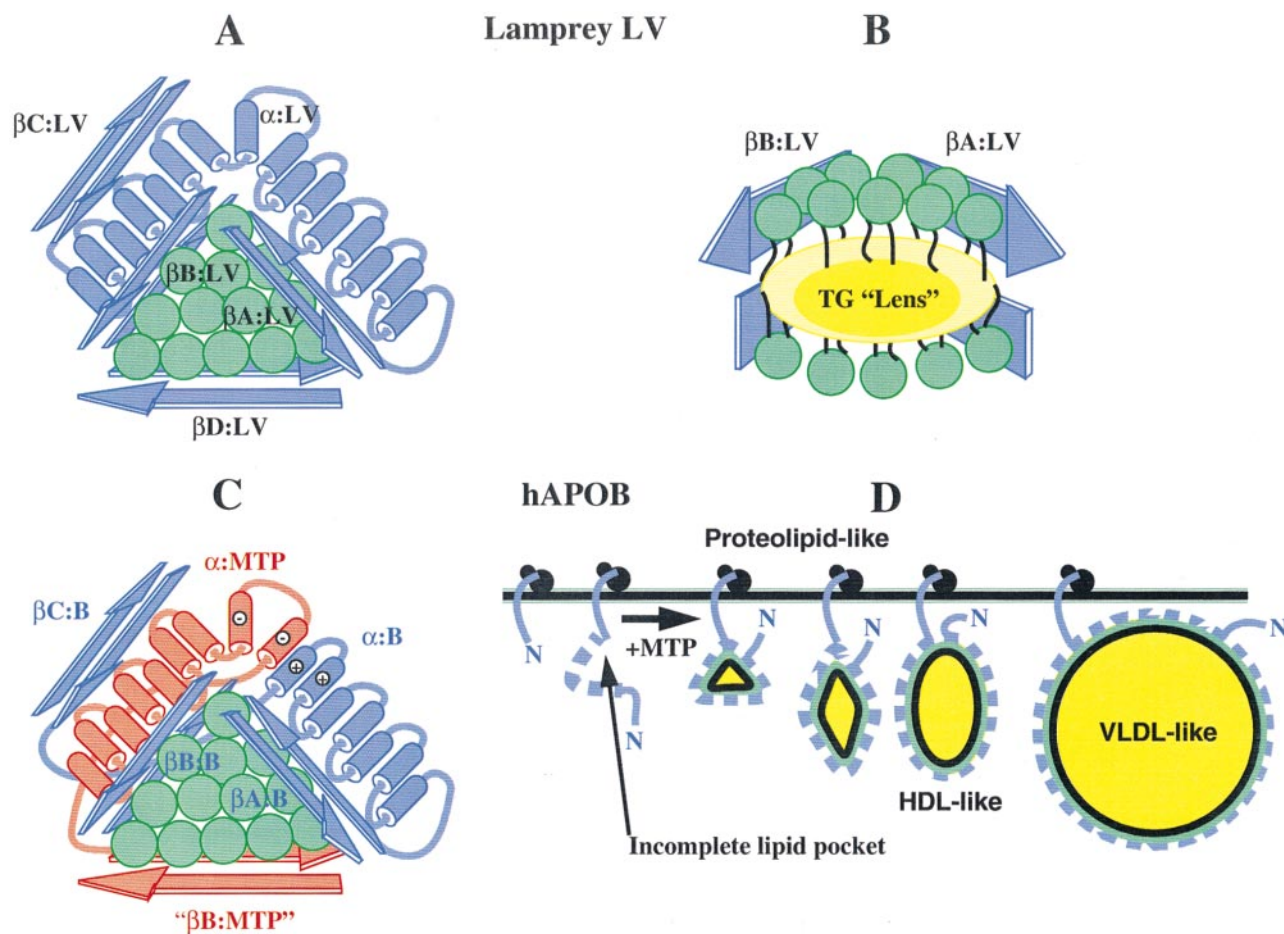


Fig. 7. “Lipid pocket” model for assembly of apoB-containing lipoprotein particles. **A:** Model of the structure of the lamprey LV “lipid pocket” viewed from above. Green circles represent phospholipid headgroups. The protein structural schematics in blue, cylinders for α helices, and antiparallel arrows for antiparallel β sheets, indicate the relative positions of the β sheets, β C:LV, β A:LV, β B:LV, and β D:LV, and the α helical cluster, α :LV. For simplicity, only two of the β strands in each β sheet are shown. The α :LV cluster consists of two layers of α helices that line the outside of the β sheets, β A:LV and β B:LV, arranged so that α helices in a given layer are parallel to each other and antiparallel to the α helices of the other layer. Further, the central helices of the cluster form the wall of the lipid pocket in the gap between β A:LV and β B:LV. **B:** Model of the structure of the lamprey LV “lipid pocket” viewed from the side. The green circles with black tails represent phospholipid molecules in the form of an expanded bilayer containing a central triacylglycerol “lens” (yellow). The relative positions of the β sheets, β A:LV and β B:LV, are indicated by protein structural schematics as in **A**. **C:** Model of the proposed hAPOB “lipid pocket” viewed from above in which β A:B and β B:B require the amphipathic β strand cluster, “ β B:MTP” and the α helical cluster LV homologue fragment, α :MTP from hMTP, to complete the pocket (compare with **A**). The protein structural schematics are color coded, blue for hAPOB and red for hMTP. Truncation of α :B occurs precisely at the edge of the gap between β A:LV and β B:LV. The α :LV domain is continued with α :MTP, starting with the central helices that form the wall of the lipid pocket in the gap between β A:LV and β B:LV. We hypothesize that the portion of α :MTP that lines the lipid pocket (negatively charged) represents the domain of hMTP that binds to the positively charged N-terminal portion of the α :B domain of hAPOB (30–32). **D:** An incomplete “lipid pocket” is formed by amphipathic β strands (blue dashes) located in the α_1 domain of hAPOB. MTP is required for this pocket to fill with lipid (yellow, neutral lipid; green, phospholipid head groups; black, fatty acyl chains), perhaps acting as a co-structural element to complete the pocket (see **C**). Once the pocket is filled, additional amphipathic β strands from the β_1 domain of hAPOB are co-translationally added, allowing the “lipid pocket” to expand until defined lipoprotein particles with HDL, then VLDL density result.

cisely onto the amphipathic β strand cluster, II β :722–802 (β B:MTP), of hMTP. ν) The C-terminal half of the α helical domain of LV (α :LV) maps onto a cluster of positively charged amphipathic α helices in hAPOB, residues 477–618, while the N-terminal portion of this α helical domain maps onto a cluster of negatively charged amphipathic α helices in hMTP, residues 340–480 (Fig. 5).

The autosomal recessive disorder, abetalipoproteinemia, produces a virtual absence of apoB-containing lipoproteins; in this disorder, microsomal triglyceride transfer protein (MTP) is not detectable (27). This observation

has led to the suggestion that the presence of MTP is necessary for the formation of apoB-containing, TAG-rich lipoproteins (29). Recent studies in cell lines that do not express apoB and MTP, i.e., HeLa and COS-1 cells, have clearly demonstrated that co-transfection of these cells with MTP and large C-terminally truncated apoB resulted in secretion of apoB-containing lipoproteins (28, 29). Expression of MTP was required for efficient secretion of apoB from these cells (28, 29).

The apparent absolute requirement of hMTP for assembly of apoB-containing lipoproteins could simply mean

that hMTP plays a purely non-structural role as a shuttle to fill the "lipid pocket" of the proposed LV-like apoB intermediate. The dense cluster of amphipathic β strands at the N-terminal end of hMTP, I β :MTP:16–243 (Fig. 4), seems likely to be a part of that protein's binding site for shuttling monomeric triacylglycerol to apoB.

However, as hMTP binds specifically to the unlipidated α_1 domain of hAPOB (30–32), MTP also may play a more central structural role in assembly of apoB-containing lipoproteins. We hypothesize that, in the absence of the MTP, the LV-like apoB "lipid pocket" intermediate is incomplete and no apoB-containing lipoprotein can be assembled. Particularly relevant to development of a model from this hypothesis is a recent study by Hussain et al. (32). They have shown that positively charged amino acid residues, presumably on amphipathic α helices located between residues 430–570 of hAPOB, are critical for MTP binding to apoB. Further, they find that 40% and 70% of this binding activity is abolished by truncation of the B:430–570 construct to residues 509 and 502, respectively. LOCATE identifies three positively charged amphipathic α helices between residues 477–491 and 492–508 (indicated in Fig. 5 by two red bars marked by red arrowheads with + signs) and 527–541 in the N-terminus of the α :LV homologue domain of hAPOB, α B:477–618; the locations of these three amphipathic helices, particularly the latter two, correlate extremely well with the Hussain et al. data (32).

In another recently published paper, Mann et al. (33) use molecular modeling to suggest a homology of β C:LV of lamprey LV to the N-terminal portions of both hAPOB and hMTP. Based upon the results of site-directed mutagenesis studies, these authors propose that initial hAPOB binding to hMTP occurs via their respective β C homologue domains (33), a possibility compatible with our homology and structural analyses and the "lipid pocket" model. These results, together with our suggestion that the dense cluster of amphipathic β strands at the N-terminal end of hMTP, I β :MTP:16–243 (see Fig. 4) represents that protein's binding site for shuttling monomeric triacylglycerol to apoB, led us to suggest the following. The β barrel structure proposed by Mann et al. (33) for β C:B may be the triacylglycerol-binding domain of hAPOB that accepts "shuttled" monomeric triacylglycerol from the β barrel structure proposed by Mann et al. (33) for the β C:MTP domain of hMTP.

Based upon our analyses and the literature, we propose the following "lipid pocket" model (Fig. 7C). Truncation of α :B occurs precisely at the edge of the gap between β A:LV and β B:LV. The α :LV domain is continued with α :MTP, starting with the central helices that form the wall of the lipid pocket in the gap between β A:LV and β B:LV. The negatively charged α :LV homologue domain of hMTP, residues 340–480 (α :MTP in red), binds, possibly through the helices indicated by red arrowheads with – signs in Fig. 5, to the positively charged α :LV homologue domain of hAPOB, residues 477–618 (α :B in blue). The result is the creation of a functional α :LV homologue domain by bringing β A:B and β B:B together; a third amphipathic β strand cluster, residues 722–802 (" β B:MTP") from hMTP, completes the construction of a competent

"lipid pocket". One attractive feature of this model is that it provides a defined mechanism for regulation of assembly of apoB-containing lipoprotein particles.

Alternate models can be envisioned. Although the similarity of the amphipathic β strands located near residues 1150 of apoB ("bD:B") to β D:LV is weak, another possibility for an apoB "lipid pocket" is one in which " β D:B" combines with β A:B and β B:B to complete the "lipid pocket." A variation of this model is one in which no "lipid pocket," i.e., no protein-rich, lipid-poor intermediate, is formed; rather β A:B and β B:B form the initial units in an expanding series of amphipathic β sheets that attach cotranslationally by some unknown mechanism to a relatively large lipid particle. In a slightly less simple possibility, β C:B combines with β A:B and β B:B to complete the "lipid pocket." Although the latter model requires a rearrangement of β C:B relative to β B:B from that found in the known lamprey LV structure, Raag et al. (8) and Banaszak, Sharrock, and Timmins (9) have suggested that β C:LV might be able to fold into the lamprey LV "lipid pocket" to accommodate increasing lipid; perhaps the truncation of α :B to half the length of α :LV forces β C:B into the pocket.

CONCLUSIONS

We conclude that the structural similarity of the N-terminal region of hAPOB to LV and hMTP harbors important implications for mechanisms involved in the assembly of apoB-containing lipoprotein particles. In particular, we suggest that a LV-like structural intermediate containing a "lipid pocket" is formed by the N-terminal portion of apoB in combination with MTP (or perhaps alone). This intermediate represents a self-assembled lipid nidus that initiates assembly of apoB-containing lipoprotein particles; pocket expansion through the addition of amphipathic β strands from the β_1 domain of apoB results in the formation of a spheroidal lipoprotein particle (Fig. 7D). Confirmation of this hypothesis awaits further structural and molecular biological studies of the N-terminal domain of apoB and the α :MTP and " β B:MTP" domains of hMTP. ■

We would like to thank Dr. Leonard Banaszak and his graduate student, Tim Anderson, for early helpful discussions of their unpublished assignment of the amino acid sequence of lamprey LV to its three-dimensional structure and for later making the PDB coordinates available to us. We would also like to thank Dr. William Pearson for help in making protein sequence comparisons. Our work was supported in part by NIH grant HL-34343.

Manuscript received 20 July 1998 and in revised form 26 April 1999.

REFERENCES

1. 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.

2. Shelness, G. S., and J. T. Thornburg. 1996. Role of intramolecular disulfide bond formation in the assembly and secretion of apolipoprotein B-100-containing lipoproteins. *J. Lipid Res.* **37**: 408–419.
3. Ingram, M. F., and G. S. Shelness. 1997. Folding of the amino-terminal domain of apolipoprotein B initiates microsomal triglyceride transfer protein-dependent lipid transfer to nascent very low density lipoprotein. *J. Biol. Chem.* **272**: 10279–10286.
4. Huang, X. F., and G. S. Shelness. 1997. Identification of cysteine pairs within the amino-terminal 5% of apolipoprotein B essential for hepatic lipoprotein assembly and secretion. *J. Biol. Chem.* **272**: 31872–31876.
5. Tran, K., J. Boren, J. Macri, Y. Wang, R. McLeod, R. K. Avramoglu, K. Adeli, and Z. Yao. 1998. Functional analysis of disulfide linkages clustered within the amino terminus of human apolipoprotein B. *J. Biol. Chem.* **273**: 7244–7251.
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. Segrest, J. P., M. K. Jones, V. K. Mishra, V. Pierotti, S. G. Young, J. Borén, T. L. Innerarity, and N. Dashti. 1998. Apolipoprotein B-100: conservation of lipid-associating amphipathic secondary structural motifs in nine species of vertebrates. *J. Lipid Res.* **39**: 85–102.
8. Raag, R., K. Appelt, N-H. Xuong, and L. Banaszak. 1988. Structure of the lamprey yolk lipid-protein complex: lipovitellin-phosvitin at 2.8 Å resolution. *J. Mol. Biol.* **200**: 553–569.
9. Banaszak, L., W. Sharrock, and P. Timmins. 1991. Structure and function of a lipoprotein: lipovitellin. *Annu. Rev. Biophys. Biophys. Chem.* **20**: 221–246.
10. Perez, L. E., M. J. Fenton, and I. P. Callard. 1991. Vitellogenin—homologs of mammalian apolipoproteins? *Comp. Biochem. Physiol.* **100B**: 821–826.
11. Sharrock, W. J., T. A. Rosenwasser, J. Gould, J. Knott, D. Hussey, J. I. Gordon, and L. Banaszak. 1992. Sequence of lamprey vitellogenin: implications for the lipovitellin crystal structure. *J. Mol. Biol.* **226**: 903–907.
12. Engelman, D. M., T. A. Steitz, and A. Goldman. 1985. Identifying transbilayer helices in amino acid sequences of membrane proteins. *Annu. Rev. Biophys. Biophys. Chem.* **15**: 321–353.
13. Jacobs, R. E., and S. H. White. 1989. The nature of the hydrophobic binding of small peptides at the bilayer interface: implications for the insertion of transbilayer helices. *Biochemistry.* **28**: 3421–3437.
14. 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 helices 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.
15. Mishra, V. K., and M. N. Palgunachari. 1996. Interaction of model class A1, A2, and class Y amphipathic helical peptides with membranes. *Biochemistry.* **35**: 11210–11220.
16. Pearson, W. R., and D. J. Lipman. 1988. Improved tools for biological sequence comparison. *Proc. Natl. Acad. Sci. USA.* **85**: 2444–2448.
17. 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.
18. Lawrence, C. E., S. F. Altschul, M. S. Boguski, J. S. Liu, A. F. Newwald, and J. C. Wootton. 1993. Detecting subtle sequence signals: a Gibbs sampling strategy for multiple alignment. *Science.* **262**: 208–214.
19. Stifani, S., R. George, and W. J. Schneider. 1991. Vitellogenin motifs conserved in nematodes and vertebrates. *J. Mol. Evol.* **32**: 429–438.
20. Shoulders, C. C., M. E. Narcisi, J. Read, S. A. Chester, D. Brett, J. Scott, T. A. Anderson, D. G. Levitt, and L. J. Banaszak. 1994. The abetalipoproteinemia gene is a member of the vitellogenin family and encodes an α helical domain. *Nature Struct. Biol.* **1**: 285–286.
21. Graham, D., K. Lesley, J. Timothy, T. C. Jones, R. J. Pease, 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.
22. 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: evidence that the length of apolipoprotein B has a major effect on the buoyant density of the secreted lipoproteins. *J. Biol. Chem.* **266**: 3300–3308.
23. Boren, J., L. Graham, M. Wettsten, J. Scott, A. White, and S-O. Olofsson. 1991. The assembly and secretion of apoB-100-containing lipoproteins in HepG2 cells. ApoB-100 is cotranslationally integrated into lipoproteins. *J. Biol. Chem.* **267**: 9858–9867.
24. 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 apolipoprotein B. *Proc. Natl. Acad. Sci. USA.* **88**: 7313–7317.
25. McLeod, R. S., Y. Wang, S. Wang, A. Rusinol, P. Links, and Z. Yao. 1996. Apolipoprotein B sequence requirements for hepatic very low density lipoprotein assembly. *J. Biol. Chem.* **271**: 18445–18455.
26. 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.
27. Wetterau, J. R., L. P. Aggerbeck, M-E. Bouma, C. Eisenberg, A. Munck, M. Hermier, J. Schmitz, G. Gay, D. J. Rader, and R. E. Gregg. 1992. Absence of microsomal triglyceride transfer protein in individuals with abetalipoproteinemia. *Science.* **258**: 999–1001.
28. Gordon, D. A., H. Jamil, D. Sharp, D. Mullaney, Z. Yao, R. E. Gregg, and J. Wetterau. 1994. Inhibition of the microsomal triglyceride transfer protein blocks the first step of apolipoprotein B assembly but not the addition of bulk core lipids in the second step. *Proc. Natl. Acad. Sci. USA.* **91**: 7628–7632.
29. Leiper, J., J. D. Bayliss, R. J. Pease, D. J. Brett, J. Scott, and C. C. Shoulders. 1994. Microsomal triglyceride transfer protein, the abetalipoproteinemia gene product, mediates the secretion of apolipoprotein B-containing lipoproteins and decreases cotranslational degradation of apolipoprotein B in transfected COS-7 cells. *J. Biol. Chem.* **269**: 21951–21954.
30. Wu, X., M. Zhou, L-S. Huang, J. Wetterau, and H. N. Ginsberg. 1996. Demonstration of a physical interaction between microsomal triglyceride transfer protein and apolipoprotein B during the assembly of apoB-containing lipoproteins. *J. Biol. Chem.* **271**: 10277–10281.
31. Hussain, M. M., A. Bakillah, and H. Jamil. 1997. Apolipoprotein B binding to microsomal triglyceride transfer protein decreases with increases in length and lipidation: implications in lipoprotein biosynthesis. *Biochemistry.* **36**: 13060–13067.
32. Hussain, M. M., A. Bakillah, N. Nayak, and G. S. Shelness. 1999. Amino acids 430–570 in apolipoprotein B are critical for its binding to microsomal triglyceride transfer protein. *J. Biol. Chem.* In press.
33. Mann, C. J., T. A. Anderson, J. Read, S. A. Chester, G. B. Harrison, S. Kochl, P. J. Ritchie, P. Bradbury, F. S. Hussain, J. Amey, B. Vanloo, M. Rosseneu, R. Infante, J. M. Hancock, D. G. Levitt, L. J. Banaszak, J. Scott, and C. C. Shoulders. 1999. The structure of vitellogenin provides a molecular model for the assembly and secretion of atherogenic lipoproteins. *J. Mol. Biol.* **285**: 391–408.