Limited Proteolysis and Biophysical Characterization of the Lipovitellin Homology Region in Apolipoprotein B[†]

Zhenghui Gordon Jiang, Margaretha Carraway, and C. James McKnight*

Department of Physiology and Biophysics, Boston University School of Medicine, 715 Albany Street, Boston, Massachusetts 02118

Received August 9, 2004; Revised Manuscript Received November 8, 2004

ABSTRACT: Apolipoprotein B (apoB) is the essential nonexchangeable protein in chylomicrons and very low-density lipoprotein-derived lipoprotein particles, including low-density lipoprotein (LDL). ApoB has been a key target for cardiovascular research because of its essential role in the assembly, secretion, delivery, and receptor binding of LDL. The three-dimensional structure of apoB has not been determined. However, the N-terminal region of apoB is homologous to the lipid storage protein lipovitellin, which allows the modeling of this region based on the X-ray structure of lipovitellin. The model of the N-terminal 17% of apoB (B17) suggests that, like lipovitellin, B17 consists of an N-terminal β-barrel domain, a helical domain, and a β-sheet domain (C-sheet). Here we test the validity of this model by limited proteolysis of B17 and the characterization of individual domains expressed in *Escherichia coli* and insect cell systems that are consistent with the model and proteolysis data. Circular dichroism studies of the individual domains indicate that they are folded and their secondary structures are in agreement with the model. We find that the helical domain and C-sheet of apoB interact with each other *in vitro*, suggesting a strong interaction between these two domains, even without a covalent peptide bond linkage. Our data suggest that the three lipovitellin-like domains exist in B17. Furthermore, the domains fold independently with secondary structures and stabilities like those of intact B17.

Apolipoproteins belong to a family of proteins found in plasma lipoprotein particles, where they pack, stabilize, and transport lipid emulsions in the circulation (*1*, *2*). Human apolipoprotein B (apoB)[1] is a nonexchangeable apolipoprotein in most lipoprotein particles, but not a component of high-density lipoproteins (HDLs). Two forms of apoB are present, apoB100 on very low-density lipoprotein (VLDL) produced in the liver and apoB48 on chylomicrons from the small intestine (*3*). In both VLDL and chylomicrons, apoB plays a primarily structural role in the assembly and secretion of lipoprotein particles. For low-density lipoproteins (LDLs), apoB is also a target for the receptor recognition that mediates the removal of LDLs from the bloodstream (*4*). ApoB is directly implicated in many cardiovascular diseases, including atherosclerosis, the leading cause of death in Western societies. The interaction between the aberrant apoB and subendothelium proteoglycans induces the retention of atherogenic lipoproteins on the artery wall, which is a major early event in the formation of atherosclerotic plaques (*5*).

ApoB100 is a secretory glycoprotein with 4536 amino acids and 16 N-linked oligosaccharides with a molecular mass of 550 kDa (*6*, *7*). It readily precipitates out of aqueous solution once it is deprived of the associated lipids (*8*). Secondary structure predictions of apoB based on the primary sequence suggest an NH_2-$\beta\alpha1$-$\beta1$-$\alpha2$-$\beta2$-$\alpha3$-COOH pentapartite domain arrangement (*9*−*11*). α and β represent domains composed of internally repeated amphipathic α-helices and β-sheets, respectively, which confer extremely strong lipid binding affinity. In contrast to the rest of the protein, the N-terminal $\beta\alpha1$ domain does not have an extensively repeated amphipathic structural motif, although many of its predicted helices and sheets display amphipathic properties. Cryo-electron microscopy studies of human LDL particles indicate that the N-terminal region of apoB projects away from the core of the particle like a knob on a ball (*12*, *13*). This observation is consistent with the data published in several labs which show that the N-terminal domain itself can be expressed in a lipid poor form (*14*, *15*). A protein longer than B19.5 is required for initiation of lipoprotein particle assembly in the presence of microsomal triglyceride transfer protein (MTP) in COS cells (*16*).

The major function of the N-terminal domain resides in its role in the initiation of protein folding and lipoprotein assembly. The transcription of apoB appears to be constitutive. The regulation of apoB production is controlled primarily by the balance between secretion and degradation (*17*−*19*). One hypothesis suggests the existence of a transient translocation pause after the synthesis of the N-terminal 85 kDa of this protein (*20*). The fate of the protein is largely

[†] This work was supported by grants from the National Institutes of Health (Grant HL-26335).

* To whom correspondence should be addressed. Telephone: (617) 638-4042. Fax: (617) 638-4041. E-mail: cjmck@bu.edu.

[1] Abbreviations: apoB, apolipoprotein B; CD, circular dichroism; ER, endoplasmic reticulum; GA, glutaraldehyde; GuHCl, guanidine hydrochloride; HDL, high-density lipoprotein; HPLC, high-performance liquid chromatography; IDL, intermediate-density lipoprotein; LDL, low-density lipoprotein; MBP, maltose-binding protein; MTP, microsomal triglyceride transfer protein; SDS−PAGE, sodium dodecyl sulfate−polyacrylamide gel electrophoresis; TPCK, L-1-tosylamido-2-phenylethyl chloromethyl ketone; Tris, tris(hydroxymethyl)aminomethane; VLDL, very low-density lipoprotein.

10.1021/bi048286y CCC: $30.25
Published on Web 12/31/2004

dependent on the folding of this N-terminal region in the endoplasmic reticulum (ER) and the availability of lipids. The N-terminal 11% of apoB contains six of eight disulfide bonds in the intact protein. The formation of correct disulfide bond pairs in the N-terminal region is essential for the folding and secretion of apoB (*21*−*23*). Furthermore, an N-linked oligosaccharide at position Asn158, when mutated, leads to the rapid degradation of the N-terminal 17% of apoB (B17) in rat hepatoma McA-RH7777 cells (*24*). The folding of the N-terminal region in the ER is required for the interaction with a critical chaperone during VLDL maturation, MTP. Lipoprotein secretion is blocked in the absence of functional MTP or the presence of MTP inhibitors (*25*−*27*). Two MTP binding regions have been identified on apoB: residues 1−300 and residues 430−570 (*28*, *29*). Segrest and co-workers hypothesized that MTP and apoB form a triangular pocket which precedes the transfer of lipids via MTP as the pocket expands through the addition of amphipathic β-strands from the β1 domain to eventually form a lipoprotein particle (*30*, *31*). An alternative model of the initiation is based on the observation that B17 is able to bind dimyristoylphosphatidylcholine multilamellar vesicles and convert the vesicles into discoidal-shaped particles *in vitro* (*14*). Therefore, the N-terminal domain of apoB could, in theory, initiate phospholipid binding itself and serve as a nucleation site for further lipid incorporation via MTP. However, our current understanding of the events happening in the ER with regard to apoB, MTP, and lipids remains limited by the lack of knowledge of the structure and biophysical properties of apoB.

The structural analysis of apoB has been difficult, mainly because of its large size and complications from its bound lipid. However, lipovitellin, a putative ancestor protein of apoB found in egg yolk, has been crystallized (*32*, *33*). Lipovitellin is an ancient lipid transport and storage protein that delivers lipids into the oocyte. This primitive delivery pathway existed before the circulating lipoprotein system evolved (*34*). The primary sequence of lipovitellin is homologous to the N-terminal 20.5% of apoB, with 20.8% identity as calculated with BLAST (*35*). On the basis of this homology, several labs have built models of apoB based on the three-dimensional structure of lipovitellin (*11*, *29*, *36*). According to these models, B17 consists of an N-terminal β-barrel domain, followed by a largely helical domain and then a β-sheet domain comparable to the C-sheet in lipovitellin (Figure 1a). A convincing feature of this model is that all the cysteine residues localize in proximity to form the correct disulfide bond linkages, even in the calculations lacking disulfide information. However, lipovitellin is found in a much more primitive metabolic pathway in a much simpler organism, though both are involved in lipid binding. The huge distance on the evolutionary tree, as well as their modest homology, casts doubt on the reliability of this model. Furthermore, even if the structures of the three major domains in apoB and lipovitellin are conserved, their relative orientation may be significantly different. In this paper, we use limited proteolysis, peptide mapping, and biophysical methods to experimentally test the validity of the lipovitellin-based model of the N-terminal region of apoB.

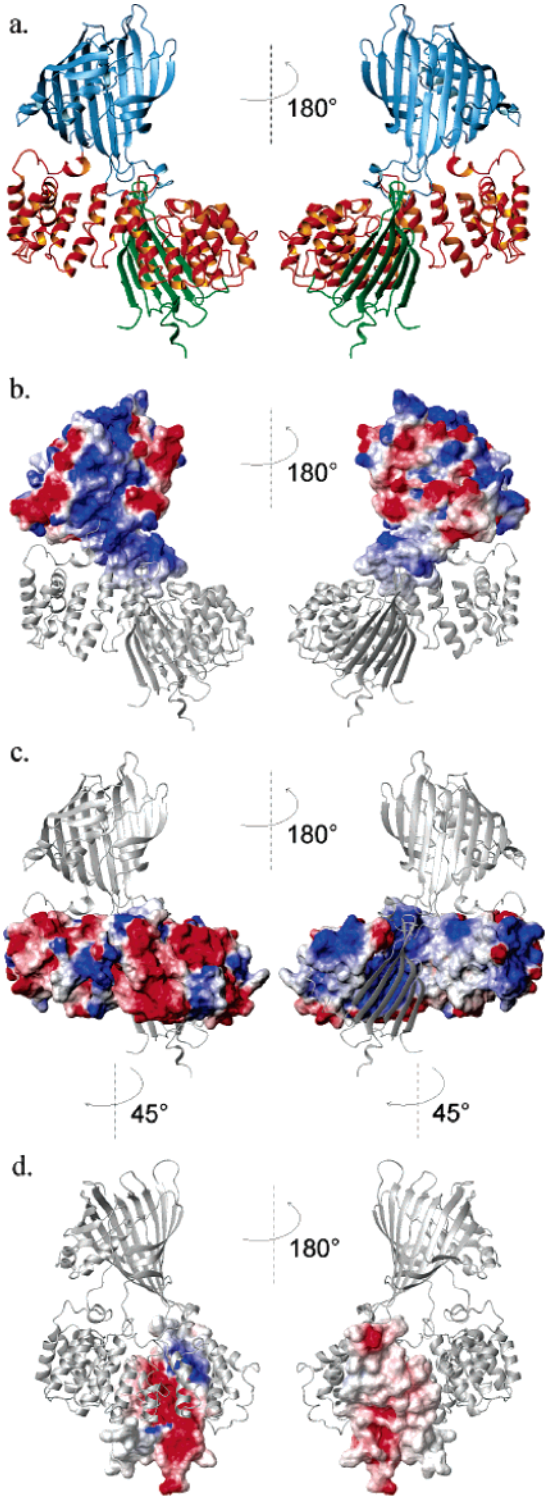

FIGURE 1: Three-dimensional structure model of B17 based on lipovitellin. The model was built based on the structure of lipovitellin (PDB entry 1LSH) with MODELLER 4 (*37*). Three segments, residues 1−18, 676−747, and 777−802, were not included in the model, because of the lack of homologous coordinates in the lipovitellin PDB file. The ribbon and surface potential representations were generated by MOLMOL (*51*). (a) Ribbon representation of B17: blue for the β-barrel domain, red for the helical domain, and green for the C-sheet domain. (b−d) Surface potential representations for individual domains in B17: blue for positive charges and red for negative charges. (b) Surface potential for the β-barrel domain. (c) Helical domain. (d) C-Sheet domain. The right-hand view is a 180° rotation about the *y*-axis. Panel d is a view after a 45° rotation along the *y*-axis from panel c, for a better view of the C-sheet domain.

EXPERIMENTAL PROCEDURES

Modeling of the N-Terminal Region of ApoB. The apoB protein sequence was aligned with the lipovitellin sequence through BLAST 2 SEQUENCE (*35*). The structure of lipovitellin [Protein Data Bank (PDB) entry 1LSH], including two chains (Lv-1N and Lv-1C), was used as a template for the modeling. An initial model of B17 was generated using MODELLER 4 with all the cysteine residues in free form (*37*). Disulfide bond information was added later based on the initial structure and previous biochemical information (*7*). The secondary structure contents were calculated from the model with DSSP (*38*).

Cloning of B17 and Its Domain Constructs. PCR products encoding B5.9 (residues 1−264), B13 (residues 1−611), and B17 (residues 1−782) were cloned into pDONR201 (Invitrogen) with the N-terminal signal sequence of apoB. B5.9 and B13 contained a six-His tag at their carboxyl termini. The target genes were transferred from pDONR201 into the pDEST8 vector (Invitrogen) through site-specific recombination reactions. The product was transformed into DH10Bac competent cells (Invitrogen) to generate the bacmid DNA. Bacmid DNA was isolated and transfected into Sf9 cells to generate infectious recombinant baculovirus particles (Invitrogen). The titer of each virus stock was measured with a plaque assay (Invitrogen).

For expression in *Escherichia coli*, B6.4-13 (residues 292−593), B6.4-15 (residues 292−680), B6.4-17 (residues 292−782), B13-15 (residues 611−680), and B13-17 (residues 611−782) were cloned into the pET24a vector (Novagen) using standard protocols (*39*, *40*). All constructs contained a six-His tag at their carboxyl termini. For B13-15 and B13-17, an additional Pro residue was introduced after the N-terminal Met residue to maximize the level of expression. The DNA sequences of all constructs were confirmed by DNA sequence analysis at the Boston University Molecular Genetics Core Facility.

Protein Expression and Purification. B5.9 (with the His tag and without the His tag), B13, and B17 were expressed in Sf9 cells (Invitrogen). The Sf9 cells were infected with recombinant baculovirus at a multiplicity of infection of 1−10 and cell densities between 1.5×10^6 and 2×10^6 cells/mL. Three days post-infection, media were harvested and cleared by centrifugation at 500*g* for 10 min. Supernatants were loaded onto a column prepacked with Affi-Gel heparin (Bio-Rad) and eluted with a step gradient of sodium chloride (from 0 to 1 M) in 20 mM sodium phosphate (pH 7.5). B5.9 (with the His tag) and B13 were further purified on Ni−NTA Sepharose (Qiagen) and eluted with increasing imidazole concentrations. B5.9 (without the His tag) and B17 after heparin purification were used for limited proteolysis.

For spectroscopic studies, B5.9 and B17 were further purified by high-performance liquid chromatography (HPLC) on a C4 reverse phase column (Vydac). Proteins were eluted with a water/acetonitrile gradient containing 0.1% trifluoroacetic acid. Fractions with the purified protein were supplemented with 8 M urea to 1.5 times the fraction volume (final urea concentration of 2.3 M). The volume was reduced by $^2/_3$ via rotary evaporation (Brinkmann). The denatured protein solution in 8 M urea then underwent stepwise dialysis against 20 mM sodium phosphate and 150 mM sodium chloride (pH 7.5) for refolding.

The constructs lacking the N-terminal B5.9 domain, B6.4-13, B6.4-15, B6.4-17, B13-15, and B13-17, were expressed as inclusion bodies in *E. coli* BL21(DE3) cells. Protein expression was induced with 1 mM IPTG at an optical density of 0.6−0.8 at 600 nm. Cell pellets were collected 3 h after induction by centrifugation for 30 min at 3000*g*. The cells were lysed with 1 mg/mL lysozyme in 10 mM Tris and 100 mM sodium phosphate (pH 8.0) for 30 min and then sonicated three times for 15 min, followed by centrifugation at 90000*g* for 30 min. Pellets were washed with 1% Triton X-100 and then 1 M urea. The inclusion bodies were solubilized in 8 M urea. The denatured proteins were loaded onto a Ni−NTA Sepharose column and purified with a gradient of imidazole in 6 M guanidine hydrochloride (GuHCl) according to the Qiagen Ni−NTA purification protocol. Unfolded proteins in GuHCl were slowly dripped into a refolding buffer [50 mM Tris, 0.8 M arginine, 5 mM EDTA, 10 mM reduced glutathione, and 2 mM oxidized glutathione (pH 8.0)] and gently stirred for 2 h at room temperature. Glutathione was omitted for the B13-15 and B13-17 constructs that lack disulfide bonds. The refolding reaction mixtures were then dialyzed against 10 mM sodium phosphate and 150 mM sodium chloride (pH 7.5). All the proteins were concentrated with an Amicon Ultra concentrating apparatus (Millipore). Protein concentrations were calculated from the UV absorbance at 280 nm (*41*).

Size Exclusion Chromatography. Protein samples were concentrated to 1−5 mg/mL in 10 mM sodium phosphate and 150 mM sodium chloride (pH 7.5), then injected into a 200 μL sample loop, and loaded on a Superdex GL 200 column (Amersham Biosciences). The protein was eluted with a flow rate of 0.5 mL/min in the same buffer at 4 °C. The elution was monitored by UV absorbance at 280 nm. Column fractions were analyzed by SDS−PAGE.

Limited Proteolysis. TPCK-treated trypsin from bovine pancreas (Sigma) was freshly prepared at 1 mg/mL in 1 mM HCl. The trypsin stock solution was added to 250 μL of 1 mg/mL B5.9 or B17 to produce the final protein:trypsin ratios of 10:1, 100:1, and 1000:1 (w/w). To maintain the solubility of B5.9, which is sensitive to ionic strength, 500 mM sodium chloride was included in all reaction mixtures. The mixtures were incubated at 37 °C, and 50 μL aliquots were taken at 5, 30, 60, and 120 min. The proteolysis in each aliquot was terminated by the addition of 16 μL of gel loading buffer containing SDS and β-mercaptoethanol for electrophoresis, or by the addition of 5 μL of 50% acetic acid and flash-freezing in liquid nitrogen. Digestion products were separated on Tricine gels (*42*), transferred onto polyvinylidene fluoride membranes, and submitted for N-terminal sequencing (Tufts University Core Facility, Tufts Medical School, Boston, MA). Large-scale digestion products were purified by HPLC as described above. Fractions containing purified fragments were confirmed on Tricine gels and then submitted for mass spectral analysis (Molecular Biology Core Facilities, Dana Farber Cancer Institute, Boston, MA).

Cross-Linking. Samples (100 μL) contained 0.2 mg/mL protein in 10 mM sodium phosphate, 150 mM sodium chloride, and 10% glycerol (pH 7.5). After equilibration at room temperature for 1 h, glutaraldehyde (EM grade, American Bioanalytical) was added to final concentrations of 0, 0.001, 0.002, 0.005, 0.01, 0.02, 0.05, 0.1, 0.5, 1, and 10% (w/v). The mixtures were incubated at room temperature

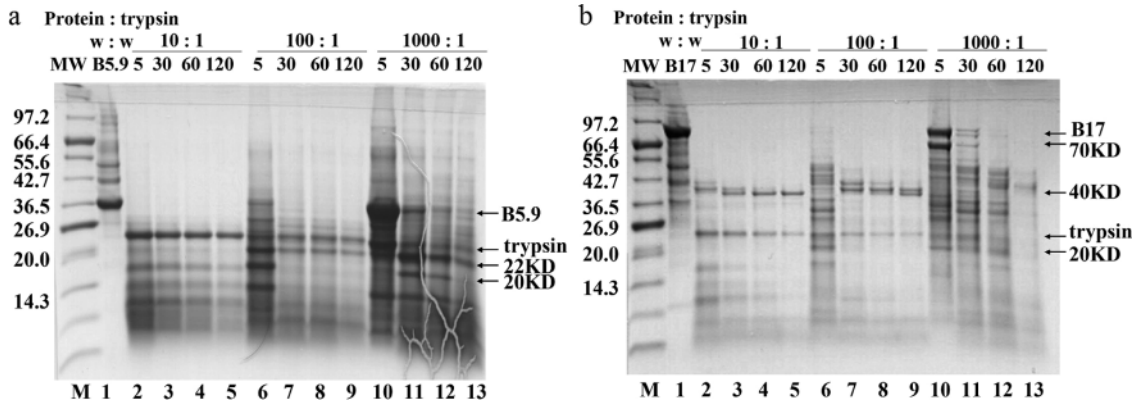

FIGURE 2: Limited trypsin digestion of B5.9 and B17. B5.9 and B17 at 1 mg/mL in 50 mM sodium phosphate and 500 mM sodium chloride (pH 7.5) were supplemented 10:1, 100:1, and 1000:1 (w/w) with trypsin. Samples were taken after 5, 30, 60, and 120 min and separated on a Tricine−SDS gel. (a) Proteolysis of B5.9: lane M, molecular mass markers; lane 1, B5.9 before digestion; lanes 2−5, 10:1 protein:trypsin ratio; lanes 6−9, 100:1 protein:trypsin ratio; and lanes 10−13, 1000:1 protein:trypsin ratio. (b) Proteolysis for B17. Same as panel a, except that B5.9 is replaced with B17.

for 5 h, with interspersed vortexing every hour. The cross-linking reactions were terminated by the addition of 15 μL of 0.5 M glycine. Reaction products were analyzed by 12.5% SDS−PAGE. For most reactions, 0.005% glutaraldehyde was sufficient for the assay. Higher concentrations of glutaraldehyde resulted in the formation of higher-molecular mass oligomers for all the proteins, including the MBP control. Unless otherwise specified, the 0.005% glutaraldehyde reactions are shown.

Circular Dichroism of ApoB Constructs. CD data were collected on AVIV 62 DS and AVIV 215 instruments. Protein solutions (1−2 μM) were made in 10 mM potassium phosphate (pH 7.5), except for those of B5.9 and B13, which were supplemented with 150 mM potassium fluoride to increase the solubility of the protein. The exact protein concentration was determined by UV absorbance at 280 nm immediately before the scan of each sample. Wavelength scans are the average of four scans taken in a 1 mm cuvette with a 5 s averaging time at every nanometer. The scans were corrected for the signal of the cuvette with buffer only. The secondary structure contents were calculated with CDPro (*43*). The averages of results from the three programs in CDPro, CONTIN, SELCON3, and CDSSTR, are reported (*44*−*46*).

The 0.1−0.2 μM protein samples in 10 mM potassium phosphate at pH 7.5 were used for chemical unfolding. The unfolding was performed using the AVIV titration accessory by the addition of a protein solution at the same concentration, prepared in 7 M GuHCl (pH 7.5), in 0.1 M steps to the 2.4 mL native protein solution in a 1 cm cuvette, with a constant volume throughout the titration. The unfolding of B5.9 was monitored at 217 nm, which was the closest we could measure to the wavelength of its maximum signal, without a loss of signal due to the absorbance of GuHCl. The other samples were monitored at 222 nm. All the unfolding experiments were performed at 25 °C. After each denaturant injection, the sample was stirred for 1 min and equilibrated for 5 s, and then the data were collected with an averaging time of 20 s.

RESULTS

Expression and Purification of B5.9 and B17. Although B5.9 is expressed at high levels in *E. coli*, the protein forms inclusion bodies that are extremely difficult to refold, possibly because of the presence of the eight cysteine residues. Therefore, we use insect cells to express apoB constructs that include the N-terminal β-barrel domain. In Sf9 cells, proteins with an N-terminal signal sequence fold in the ER and are secreted into the culture media (*40*). After a single-step heparin column purification, 90% of the impurities are removed. We assume that the proteins that are secreted after passing the ER quality controls are folded into their native conformation.

Limited Trypsin Digestion of B5.9 and B17. To test the validity of the lipovitellin homology model of the N-terminal region of apoB, we use trypsin to cleave both B5.9 and B17. Well-folded regions within proteins usually resist protease digestion. Therefore, the protected fragments after trypsin digestion represent stably folded domains. After trypsinolysis at different protein:trypsin ratios and varying incubation times, the digestion products are analyzed by SDS−PAGE and transferred onto polyvinylidene fluoride membranes for N-terminal sequencing. Accurate molecular masses of some fragments are then determined by matrix-assisted laser desorption ionization time-of-flight mass spectrometry. On the basis of the N-terminal sequence and the size of each fragment, we map the cleavage site on the protein based on the specific cleavage of trypsin at the carboxyl termini of lysine or arginine residues.

In the digestion of B5.9 expressed in Sf9 cells, protected fragments are only observed at a low trypsin concentration (1000:1 protein:trypsin weight ratio) (Figure 2a). Two major fragments with molecular masses of 20 and 22 kDa are observed. N-Terminal sequencing reveals that the 20 kDa fragment starts at residue 21 (HLRKY...), while the 22 kDa fragment begins right from the N-terminus (EEEML...) (Table 1). The cleavage near the N-terminus is consistent with the missing electron density of the N-terminal region of the lipovitellin crystal structure, indicating that it may be unfolded. Furthermore, mass spectral analysis shows that the 22 kDa fragment is actually a mixture of two species (Table 1). These two cleavage sites map to positions within the conserved hydrophobic loop that extends from the β-barrel domain and interacts with the helical and C-sheet domains. The masses of the two fragments are closest to those expected for cleavage at Lys189 and Arg192 within the

Table 1: Summary of Limited Proteolysis on B5.9 and B17

	fragment size (kDa)	N-terminal sequence	experimental MW	calculated MW[b]	predicted fragment	calculated vs experimental MW error
B5.9	20	HLRKY	N/A[a]	20065[c]/20432[c] (19779[c]/20065[c])	21−189/21−192 (21−187/21−189)	N/A
	22	EEEML	22548/22798	22386[c]/22752[c] (22110[c]/22386[c])	1−189/1−192 (1−187/1−189)	−0.7%/−0.2% (−1.9%/−1.8%)
B17	20	SVSLP	N/A[a]	18296	615−782 (C-terminus)	N/A
	40	GLSDE	39426	39363 (40483)	330−680 (330−690)	−0.2% (2.7%)
	70	EEEML	70611	70415[c] (69605[c])	1−614 (1−608)	−0.3% (−1.4%)

[a] No significant peak was observed in the mass spectrum of the 20 kDa fragment from B5.9 or the 20 kDa fragment from B17. [b] Values in parentheses correspond to the values for the next closest possible fragment. [c] B5.9 and B17 have an N-linked oligosaccharide at Asn158. The size of the oligosaccharide in B5.9 was 1041 Da as measured in a separate experiment. The size of the sugar is included in calculating the theoretical molecular weights of these fragments that contain Asn158 and its N-linked oligosaccharide.

^{187}RF^{189}KPI^{192}R sequence (Table 1). Cleavage at Arg187 is just within the error of the mass spectral data and cannot be rigorously ruled out. However, in addition to the larger difference in mass, Arg187 is predicted to be at the very beginning of the hydrophobic loop and is less likely to be accessible. Therefore, we refer to the 22 kDa fragments as ending at residues Lys189 and Arg192. Although the 20 kDa fragment did not yield a peak during mass spectral analysis, the size of the fragment on SDS−PAGE is consistent with cleavage at the these same sites in the hydrophobic loop (i.e., Lys189 and Arg192).

The digestion of B17 expressed in Sf9 cells shows a different pattern at different trypsin concentrations. At the 1000:1 protein:trypsin weight ratio, ∼50% of the 85 kDa protein is degraded to a 70 kDa fragment within 5 min and completely degraded in 2 h (Figure 2b). A 20 kDa fragment is also produced, and N-terminal sequencing indicates that it starts at the beginning of the C-sheet domain (Table 1 and Figure 2b). This 20 kDa fragment most likely ends at or near the C-terminal residue of the B17 construct (residue 782).

At 100:1 and 10:1 protein:trypsin ratios, B17 is degraded to a 40 kDa fragment in three consecutive steps (Figure 2b). The intensity of this fragment is not significantly diminished after 2 h, even at a 10:1 protein:trypsin weight ratio, suggesting the fragment is stable and tightly folded. N-Terminal sequencing and mass spectral data indicate that this fragment begins after a loop between the second and third helices of the helical domain and extends to the middle of the C-sheet domain (Table 1). It is noteworthy that this fragment also ends within a region missing in the electron density of the lipovitellin structure.

Protein Expression, Purification, and Refolding. On the basis of the limited proteolysis data, we made a series of constructs encoding different domains within B17. Two expression systems were adopted to express those domains. Constructs containing the N-terminal β-barrel domain, i.e., B5.9, B13, and B17, were cloned for expression in Sf9 insect cells, and the rest of the constructs were expressed in *E. coli* (Figure 3).

In insect cells, proteins expressed and secreted into the media are first purified with a heparin column. The eluant from the heparin purification is then loaded onto nickel resin and eluted with imidazole. B5.9 and B13 expressed in Sf9 cells are glycosylated at Asn158, with a sugar of 1041 Da, as determined by electrospray ionization mass spectrometry (data not shown). Removal of this sugar by mutating Asn158 to Ala completely blocked the secretion of B5.9 and partially blocked the secretion of B13, which is consistent with the report on the N158Q mutation in B17 (ref *24* and unpublished data).

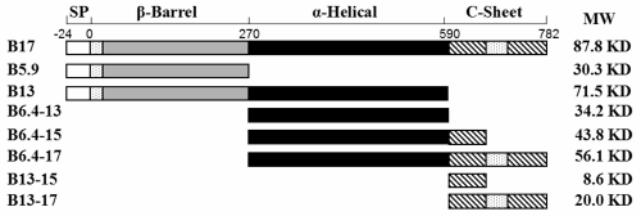

FIGURE 3: Schematic map of B17 and its domain constructs. B17, B5.9, and B13 were expressed in Sf9 cells. B6.4-13, B6.4-15, B6.4-17, B13-15, and B13-17 were expressed in *E. coli*. B5.9 was cloned in two forms, one with a six-His tag at the carboxyl terminal and one without. Except for B17, all other constructs had six-His tags at their carboxyl termini: white for the signal peptide, gray for the β-barrel domain, and black for the helical domain. Lines represent the C-sheet domain, and stippled regions represent sequences that are homologous to the regions missing in the lipovitellin crystal structure. The residue numbering is marked above B17.

In contrast to the four pairs of disulfide bonds that do not efficiently re-form in the *in vitro* refolding of the constructs containing the β-barrel, the two pairs of disulfide bonds in the helical domain constructs re-form in an appropriate redox buffer. Proteins expressed from *E. coli* are purified on nickel resins under denaturing conditions. Highly concentrated proteins containing the helical domain (B6.4-13, B6.4-15, and B6.4-17) in 6 M GuHCl are dripped into the refolding buffer with an optimized redox potential. The refolded proteins are then dialyzed into suitable oxidizing buffers for further biochemical analysis. Size exclusion chromatography indicates that the entire helical domain, B6.4-13, is a homogeneous species after refolding, with an apparent molecular mass close to that of a dimer (Figure 4). B6.4-15 exhibits a concentration-dependent monomer−dimer equilibrium. More dimer is observed at higher concentrations, suggesting that the beginning of the C-sheet may compete with the site for dimerization of the helical domain (data not shown). B6.4-17 has a molecular mass close to its monomer size as determined by size exclusion chromatography (Figure 4).

Constructs containing the C-sheet, B13-15 and B13-17, are expressed and purified in a manner similar to that of the helical domain constructs. Refolding is done in the same refolding buffer but without redox reagents, since they contain no cysteine residues. Both proteins are stable at low concentrations but aggregate readily at concentrations above 1 mg/mL, particularly B13-17. Both proteins elute from a size exclusion column before the volume predicted for a monomer, suggesting dimer formation (Figure 4). Some other

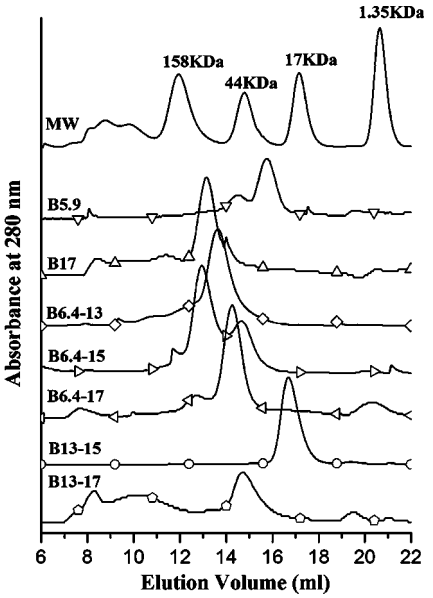

FIGURE 4: Size exclusion chromatography for the B17 domain constructs. Approximately 200 μL of a 1 mg/mL protein sample was prepared in 10 mM sodium phosphate and 150 mM sodium chloride (pH 7.5) and loaded onto a Superdex GL 200 size exclusion column. The protein was eluted with a flow rate of 0.5 mL/min at 4 °C. Protein elution was monitored by UV absorbance at 280 nm. The data for the molecular standard are depicted with a solid line. B5.9 (inverted triangles) was a monomer/dimer mixture. B17 (triangles) appeared as an almost homogeneous monomer. B6.4-13 (diamonds) eluted at its dimeric size. B6.4-15 (right-pointing triangles) eluted as a mixture of a monomer and dimer. B6.4-17 (left-pointing triangles) was mainly a monomer. B13-15 (circles) eluted as a homogeneous dimer, but aggregated over time. B13-17 (pentagons) was a mixture of dimers and other oligomers.

higher-molecular mass species are observed for B13-17, suggesting the formation of oligomers or small aggregates.

Chemical Cross-Linking of Individual Domains from B17. The molecular mass determined from size exclusion chromatography can be influenced by the molecular shape. The results are also biased by the high protein concentration necessary for detection. To observe the unbiased aggregation state in a more dilute solution, we use glutaraldehyde to cross-link the proteins at very low concentrations.

B5.9 appears to form some dimer in solution as determined by size exclusion chromatography (Figure 4). The cross-linking data clearly indicate the presence of a cross-linked dimer band at ~60 kDa (Figure 5). The helical domain, B6.4-13, after refolding is a homogeneous species, eluting at an apparent dimer molecular mass on the size exclusion column (Figure 4). Cross-linking confirms that B6.4-13 exists as a dimer in solution, with a band at 66 kDa formed (Figure 5). Little monomeric protein remains after cross-linking. In contrast, B6.4-17 is predominantly monomeric in solution. The 50 kDa band remains essentially unchanged after cross-linking, except for the presence of a smear toward the larger size, possibly from additional glutaraldehyde molecules on the protein (Figure 5).

The refolded B13-17, corresponding to the C-sheet, tends to aggregate in solution. Size exclusion chromatography of B13-17 indicates the formation of both aggregates and dimers (Figure 4). The cross-linking of B13-17 produces large-scale aggregation, as most of the proteins remain in the well after cross-linking. Still, a band at 40 kDa, corresponding to the dimer, can be observed (Figure 5).

Cross-linking of a negative control, maltose-binding protein (MBP), demonstrates that the cross-linking we observe is specific (Figure 5). MBP is a monomeric protein with many lysines on the surface. It remains monomeric after cross-linking under the same conditions.

We have also examined the interaction between different domains by mixing neighboring domains in the presence of glutaraldehyde (Figure 5). The cross-linking between B5.9 and B6.4-13, which would reconstitute most of B17 except for the C-sheet, is complicated by their similar molecular

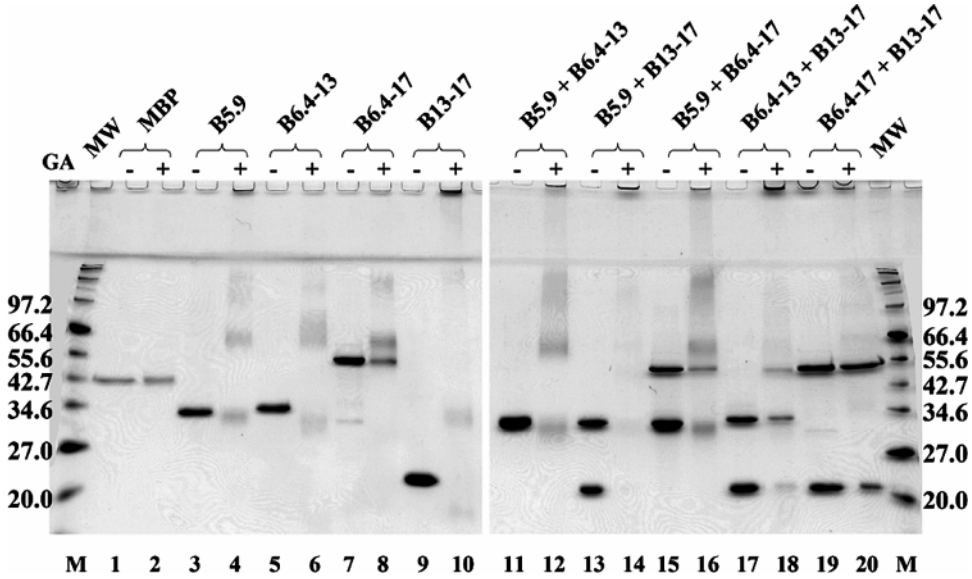

FIGURE 5: Cross-linking of B17 domain constructs. B17 domain constructs (0.02 mg/mL) were mixed in 10 mM sodium phosphate, 150 mM sodium chloride, and 10% glycerol at pH 7.5 for 1 h. Glutaraldehyde (GA) was added to a concentration of 0.005%, except for B6.4-13 with B13-17 and B6.4-17 with B13-17, where 0.002% glutaraldehyde was used. Samples were analyzed on 12.5% SDS−PAGE gels and stained with Coomassie blue: MW, molecular mass marker; lanes 1 and 2, maltose binding protein MBP); lanes 3 and 4, B5.9; lanes 5 and 6, B6.4-13; lanes 7 and 8, B6.4-17; lanes 9 and 10, B13-17; lanes 11 and 12, B5.9 and B6.4-13; lanes 13 and 14, B5.9 and B13-17; lanes 15 and 16, B5.9 and B6.4-17; lanes 17 and 18, B6.4-13 and B13-17; and lanes 19 and 20, B6.4-17 and B13-17.

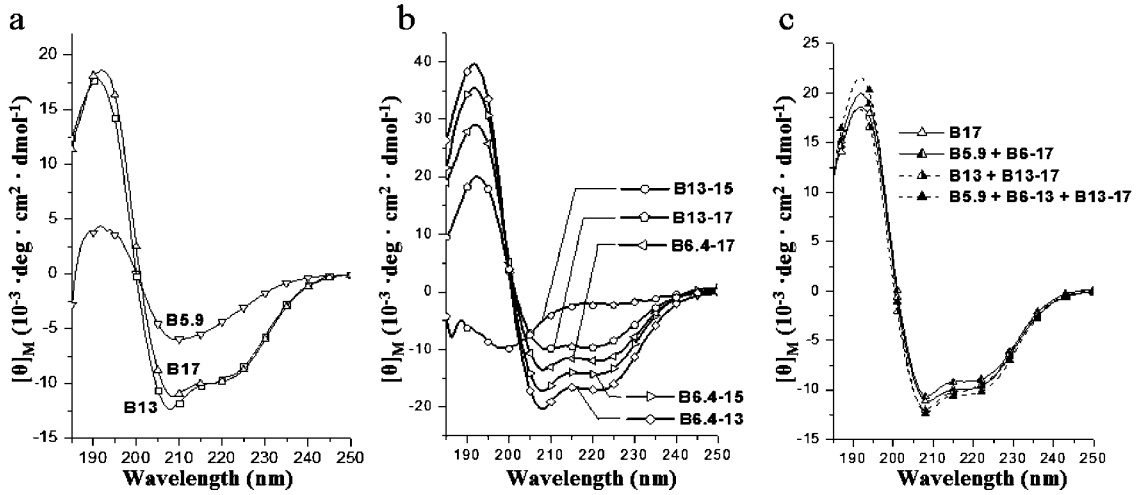

FIGURE 6: Far-UV CD spectra of B17 domain constructs. Each sample contained ~2 μM (exact molar concentration determined by UV absorbance) protein in 10 mM potassium phosphate (pH 7.5). B5.9 and B13 were supplemented with 150 mM potassium fluoride to maintain their solubility. Experiments were performed in a 1 mm cuvette at 25 °C. Each spectrum is the average of four scans with an averaging time of 5 s at every 1 nm wavelength. (a) Constructs containing the β-barrel domain: B5.9 (▽), B13 (□), and B17 (△). (b) Constructs lacking the β-barrel domain: B6.4-13 (◇), B6.4-15 (right-pointing triangles), B6.4-17 (left-pointing triangles), B13-15 (○), and B13-17 (⬠). (c) Comparison of B17 spectra to the sum of the individual domain construct spectra. The summation spectra were produced by adding the mean residue ellipticity of the individual domains weighted by their relative contributions to B17 (i.e., the ratio of residue numbers in the fragment to the 782 residues in B17): B17 (△), B5.9 and B6-17 (▲), B13 and B13-17 (▲), and B5.9, B6-13, and B13-17 (▲). For clarity, only every eighth data point is shown.

masses. Although a dimer band is clearly formed, it is difficult to differentiate whether it is from the two possible homodimers, the B5.9−B6.4-13 heterodimer, or a combination of all three species (Figure 5). B5.9 and B13-17 do not form specific cross-links with each other, which is consistent with the model, as they do not have a large interacting surface in the B17 model based on lipovitellin (Figure 5). B5.9 does not cross-link to B6.4-17 either (Figure 5), which makes the ambiguous dimer band in the B5.9 and B6.4-13 cross-linking reaction more likely to be homodimers. B6.4-13 and B13-17 form specific cross-links to each other, producing a band with the same size as B6.4-17 (Figure 5), indicating B6.4-13 and B13-17 interact with each other in a 1:1 manner. Given that B6.4-13 is a dimer by itself and B13-17 is a dimer/oligomer mixture, this heterodimeric interaction between B6.4-13 and B13-17 must have a lower K_d that can outcompete the homodimer interactions. The fact that B6.4-17 and B13-17 do not cross-link to each other suggests that B13-17 interacts with B6.4-13 in a specific manner (Figure 5). In addition to B6.4-13, B13-17 also cross-links to B13, forming a single heterodimer band on a Tricine−SDS gel (Figure S1 of the Supporting Information). Therefore, it is likely that the B13-17 and B6.4-13 interaction *in vitro* is similar to the intrinsic interaction that occurs in B6.4-17 and in intact B17.

Secondary Structure of the N-Terminal ApoB Domains. To examine the secondary structure of the different N-terminal apoB constructs, we have used circular dichroism (CD) spectroscopy (Figure 6). The far-UV CD spectra of B5.9 expressed in Sf9 cells indicate a β-sheet-dominated structure (39% β-sheet) with a minimum at 210 nm (Figure 6a and Table 2). The spectra of B13 and B17 indicate a mixture of α- and β-structures. The mean residue ellipticities of B13 and B17 are very close to each other, except that B13 has a stronger signal at 207 nm, which comes from the larger contribution of the β-barrel domain (Figure 6a). The calculated secondary structure contents are also very close to each other. The constructs containing the helical domain, B6.4-13, B6.4-15, and B6.4-17, are easily identified as helical proteins with characteristic minima at 222 and 209 nm. The helical content of the three constructs decreases from B6.4-13 to B6.4-15 to B6.4-17 as expected (60, 53, and 46% helical, respectively), due to the increase in size of the C-sheet (Figure 6b and Table 2). B13-15 is largely unfolded with a minimum at 200 nm in the CD spectrum (Figure 6b). This observation is consistent with the results of our limited proteolysis and modeling where B13-15 is only expected to fold when associated with the helical domain (B6.4-15) or as part of the C-sheet (B13-17). Surprisingly, B13-17 gives a helical spectrum with 31% α-helix and 19% β-sheet (Figure 6b and Table 2). A possible explanation is that the region not visible in the crystal structure of lipovitellin adopts some helical secondary structure. In fact, on the basis of the primary sequence, a large part of the missing region of the C-sheet is predicted to be helical. However, since the truncated C-sheet is not as highly soluble in solution, the helical content could be induced by domain truncation and/or protein self-association. It is also possible that some segments of B13-17 adopt helical structures, while more β-sheets are formed in those segments in the context of B6.4-17.

To examine whether the three domains, the β-barrel (B5.9), the helical domain (B6.4-13), and the C-sheet (B13-17), have the same conformation in isolation as in B17, we compared the B17 spectra with the sum of the spectra of the individual domains. The three summation spectra (B5.9 + B6.4-17, B13 + B13-17, and B5.9 + B6.4-13 + B13-17) overlap with the B17 curve to a close approximation. This suggests that the three domains can fold independently and no large secondary structure change occurs when the three domains are assembled as B17.

We have also examined the stability of the individual domains by monitoring their thermal and chemical unfolding by CD. The thermal unfolding of B5.9 results in an

Table 2: Summary of the Secondary Structure Contents (%) Predicted by CD

		CD[a]	model[b]	δ[c]
B5.9	α-helix	7.1	8.9	1.8
	β-strand	38.6	44.2	5.6
	others	54.6	40.1	−7.8
	not calcd[d]		6.7	
B13	α-helix	32.9	34.4	1.5
	β-strand	17.1	19.3	2.2
	others	49.7	42.5	−3.3
	not calcd[d]		3.9	
B17	α-helix	33.0	27.5	−5.5
	β-strand	16.9	21.7	4.8
	others	49.7	38.5	1.1
	not calcd[d]		12.3	
B6.4-13	α-helix	59.7	59.5	−0.2
	β-strand	5.1	0	−5.1
	others	34.9	38.2	5.6
	not calcd[d]		12.3	
B6.4-15	α-helix	52.5	47.2	−5.3
	β-strand	8.5	9.9	1.4
	others	38.6	39.9	4.3
	not calcd[d]		3.0	
B6.4-17	α-helix	46.3	38.2	−8.1
	β-strand	12.0	10.2	−1.8
	others	41.4	34.5	10.1
	not calcd[d]		17.0	
B13-15	α-helix	4.9	0	−4.9
	β-strand	30.7	50.0	19.3
	others	63.4	33.3	−13.4
	not calcd[d]		16.7	
B13-17	α-helix	30.7	16.7	−14.0
	β-strand	19.2	28.3	9.1
	others	49.8	20.0	18.2
	not calcd[d]		48.0	

[a] The reported data are the average of secondary structure contents calculated on the basis of the CD wavelength scan from 250 to 185 nm by CDPro, which contains three prediction tools: SELCON3, CONTIN, and CDSSTR (*43*−*46*). [b] The secondary structure annotation in the model was created by DSSP (*38*). [c] δ is the difference in secondary structure between the predictions from CD and the model. [d] Three types of sequences were not in the model and thus could not be annotated: the sequence not modeled due to the lack of PDB coordinates in lipovitellin, the N-terminal Met, and the C-terminal six-His tag.

irreversible aggregation at 80 °C (data not shown). In contrast, the thermal unfolding of the helical domain constructs, B6.4-13 and B6.4-15, is not complete even at 98 °C (data not shown). To compare the stability of different domains under the same conditions, we use GuHCl to chemically unfold the protein. Of the three tested domains, B5.9, B6.4-13, and B13-17, B5.9 unfolds first at 1 M GuHCl (Figure 7). The B5.9 unfolding curve indicates a cooperative unfolding process, suggesting the β-barrel domain is a single folding unit. The C-sheet construct, B13-17, starts to unfold at a higher denaturant concentration than B5.9 and loses 90% of its signal at 222 nm at 2 M GuHCl (Figure 7). The unfolding also appears to be a cooperative event. The unfolding transition of helical domain construct B6.4-13 covers a wide GuHCl concentration range from 1 to 4 M (Figure 7). It is a much less cooperative process and may involve several intermediate states. This indicates the helical region does not contain a single folding unit but, instead, many local folding units exist that unfold independently. The unfolding curve of B17 reflects the combined effect of the three independent domains (Figure 7). At low GuHCl concentrations, most of the signal loss is due to the unfolding of the C-sheet, B13-17, and to a lesser extent the β-barrel

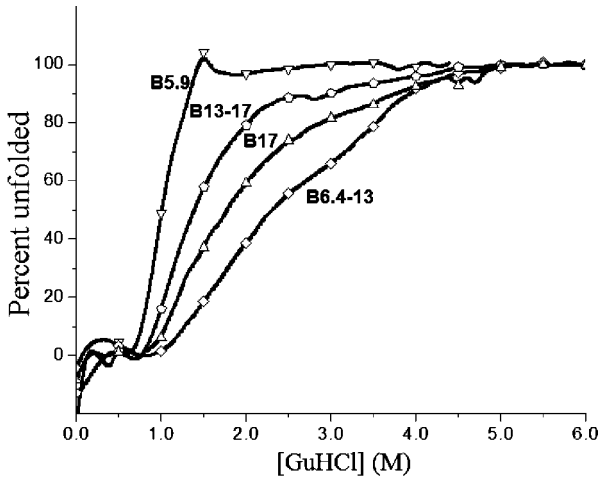

FIGURE 7: Chemical unfolding of B17 domain constructs. Each sample contained ~0.2 μM protein in 10 mM potassium phosphate (pH 7.5) in a 1 cm cuvette. The unfolding was performed under the constant volume mode, at 25 °C, by adding 7 M GuHCl (pH 7.5) at 0.1 M steps with stirring for 1 min followed by a 5 s delay. Data were collected with a 20 s averaging time at 222 nm, except for B5.9 that was monitored at 217 nm. The raw unfolding curves were converted to percent unfolded: B5.9 (▽), B6.4-13 (◇), B13-17 (⬠), and B17 (△). For clarity, only every fifth data point is shown.

domain, B5.9, at 222 nm. The loss of signal at higher GuHCl concentrations primarily results from the unfolding of the helical domain. Thus, the three domains within B17 are different in their chemical stability and unfolding cooperativity. The folding and assembly of the three domains into B17 do not provide a new more stable hydrophobic core. Instead, the intrinsic forces involved in the folding of the three individual domains appear to be independent of the domain−domain interactions in B17.

DISCUSSION

ApoB is a key protein in the metabolism of lipoproteins. A high-resolution structure of apoB has long been awaited, but has still not been determined. The N-terminal region of apoB is unique and critical for the assembly and secretion of VLDL particles (*47*). On the basis of its homology to lipovitellin, a putative ancestor of apoB, three-dimensional models have been constructed in several labs (*11*, *29*, *36*). In this paper, we combine biochemical and biophysical approaches to test the validity of this model experimentally.

One validation of the model is the model itself. The localization of all the cysteine residues in sites favoring the formation of disulfide bonds provides strong support for the model, given that the cysteine positions in apoB and lipovitellin differ. The tertiary packing of the helical domain in the model is similar to that of lipovitellin not only in the shape of the two layer helices but also in the distribution of hydrophilic and hydrophobic residues. Like the helical domain in lipovitellin, the helical domain in apoB also consists of two layers of helices with a large hydrophobic space between them, tethering the two layers together. The C-sheet domains are amphipathic in both proteins, with one side negatively charged, interacting with the helical domain, and the other side hydrophobic, possibly making a lipid-binding surface (Figure 1d).

Our limited proteolysis data also support the lipovitellin-based model. The proteins tested are B5.9 and B17 expressed

in Sf9 cells. The secretion of B5.9 and B17 through the quality control systems of the ER in Sf9 cells and the addition of a carbohydrate at the expected position are strong indications of correct folding. The advantage of using trypsin as the protease is its specific cleavage at lysine and arginine residues, making it simpler to map the cleavage sites. Conversely, it cannot identify unfolded regions lacking those residues. Additionally, we have not identified every fragment observed in the digestion; instead, we have focused on the stable, long-lived fragments. The observation of a cleavage indicates that the lysine or arginine is in an unstructured region. However, it is possible that the fragments are still associated with each other, with nicks in the loop regions.

The two fragments from the digestion of B5.9 are consistent with the lipovitellin-based model. The lipovitellin crystal structure is missing the 18 N-terminal residues, suggesting that this region may not be well folded. The cleavage site at position 20 indicates that the N-terminus of apoB adopts an unfolded state similar to lipovitellin after the signal peptide is cleaved. The cleavage sites at positions 189 and 192 can also be predicted according to the model, since a long loop from residue 188 to 203 exists in the lipovitellin-based model. This loop connects the 10th and 11th strands and interacts with the helical domain and C-sheet domain. From residue 21 to 187, there are 22 lysine or arginine residues, but no major loops are found in the lipovitellin-based model, consistent with the protease protection we observed with the 20 kDa fragment.

The cleavage patterns observed in the B17 digestion are more complicated. A complete C-sheet fragment of 20 kDa is observed at low trypsin concentrations. However, a 40 kDa fragment starting from the third helix in the helical domain to the middle of the C-sheet is observed at high trypsin concentrations. This indicates that the 40 kDa fragment adopts a more stable conformation than the C-sheet, which can be observed only at low trypsin concentrations. The C-sheet observed at the low trypsin concentration might be an intermediate product during the proteolysis, which will be further degraded, given more reaction time or protease. Additionally, the C-sheet fragment is observed together with many other unidentified fragments, suggesting that the C-sheet can still be in association with the helical domain, although it is nicked at the connecting region. This explanation is supported by the observed cross-linking between the helical domain and the C-sheet domain *in vitro*.

Trypsin cleavage occurs at accessible residues. Proteolysis, such as the cleavage of the C-sheet, may result in further unfolding and complete degradation of the protein, which could explain why only a small ratio of B17 is cleaved to the 40 kDa fragment at high trypsin concentrations. The protection of the 40 kDa fragment is the result of a specific sequential cleavage event. Since we did not identify each fragment at each reaction time point, we are not able to draw conclusions about the kinetics of the trypsin proteolysis. Instead, we focus on the major protected fragments in understanding why they are not susceptible to proteolysis.

The 40 kDa fragment from the B17 digest, containing most of the helical domain and part of the C-sheet, must be tightly folded within B17, as it protects 22 lysine and 10 arginine residues. The helical domain is cut after the first two helices at R329, in the loop connecting helix 2 and helix 3. However, there are many other sites in similar turns, such as R400, R412, K427, K480, K485, K493, K506, K510, R604, K605, and R608, within the 40 kDa fragment that are protected from digestion. Thus, the loop between the second and third helix may be longer than the model predicts or unstructured, allowing for cleavage. K480, K485, and K491 in the long loop from residue 469 to 493 in the model are not cleaved in the 40 kDa fragment, suggesting that this region folds into a stable structure. In the alignment between apoB and lipovitellin, a large part of this loop region lacks a homologous sequence in lipovitellin. This is a demonstration that the quality of homologous modeling is highly dependent on the accuracy of the alignment. More interestingly, the loop from residue 596 to 609, connecting the helical domain and the C-sheet, is also protected in this fragment, as R604, K605, and R608 are not cleaved. It indicates that this loop can adopt a folded conformation in B17. Finally, the 40 kDa fragment ends in a region missing in the lipovitellin structure, suggesting the C-sheet of apoB may contain an unstructured region like lipovitellin.

The lipovitellin-based model is also supported by the independent refolding of the three domains within B17 *in vitro*. The fact that B13-15 does not fold on its own is in agreement with our limited proteolysis data, which show that this region is within the protected fragment of either B6.4-15 or B13-17. Size exclusion chromatography and cross-linking experiments indicate that almost all the domains exhibit intermolecular interaction to some extent *in vitro*. As full-length apoB is monomeric, the homodimer interaction we observe is the result of the artificial domain truncation and the exposure of hydrophobic patches. It is noteworthy that B5.9 contains a largely hydrophilic surface in the model (Figure 1b). Interestingly, it is suggested that the β-barrel domain, B5.9, is one of the docking sites for MTP, when the latter is shuttling lipids to apoB during the assembly of VLDL in the ER (*29*). Given that the N-terminal region of MTP is also homologous to lipovitellin, one possibility is that this B5.9 homodimer mimics the dimer interaction between B5.9 and the N-terminal region of MTP.

In the model, a long loop, connecting the 10th and 11th strand in B5.9, sits in the groove between the helical domain and the C-sheet. This loop, which is composed of mainly hydrophobic residues with two basic residues on each side, is conserved in lipovitellin. In the B5.9 construct, this hydrophobic loop is cleaved by trypsin, but the loop is inaccessible in B17. This suggests that the B5.9 hydrophobic loop is buried in B17 and interacting with the helical and C-sheet domain as it does in lipovitellin. In our cross-linking experiments, B5.9 does not cross-link to B6.4-13, B13-17, or B6.4-17, suggesting that the loop is probably not sufficient for locking B5.9 in the right position or it must be inserted cotranslationally. Actually, in the lipovitellin structure, the other major component stabilizing the β-barrel domain is the A-sheet sitting at the bottom of the barrel, suggesting the B17-20 region could be interacting with B5.9 in apoB (*48*, *49*).

In contrast to B5.9, B13-17 cross-links to B6.4-13 specifically. The negatively charged face of B13-17 likely interacts with the positively charged half of the inside layer from the helical domain, generating a continuous hydrophobic surface, including the hydrophobic side of B13-17 and the hydrophobic half of the inside layer from the helical domain (Figure 1c,d). This interaction is so strong that it outcompetes

the homodimer interaction of both B6.4-13 and B13-17. Notably, in lipovitellin, the C-sheet uses this charge−charge interaction to associate itself with the helical domain and forms a hydrophobic pocket together with the A-sheet, creating an entirely hydrophobic surface for lipid binding (*49*). This structural scheme may be inherited by apoB for initiation of lipid incorporation and nucleation of the VLDL assembly.

The three domains within B17 expressed separately contain secondary structure similar to that predicted by the model. B6.4-13 has essentially the same helical content as predicted in the model (Table 2). The β-sheet content of B5.9 is 5.6% lower than the model predicts, which is reasonable, considering the less accurate β-sheet prediction by CD (Table. 2). Surprisingly, the CD spectrum of B13-17 exhibits a significant helical signal, with 30.7% helix, 19.2% β-sheet, and 49.8% random coil (Table. 2). This fragment contains a 40-residue region that is missing from the electron density of the lipovitellin crystal structure. The helical CD spectrum indicates that a large amount of the unmodeled regions in B13-17 could be in a helical form. However, it is also possible that higher than expected helicity is an artifact of domain truncation since the B13-17 construct is not highly soluble. No significant difference is observed between the measured B17 CD curve and those produced by summing the spectra of the different domains, suggesting that no large secondary structure changes occur when the three domains fold into B17. Furthermore, the B17 chemical denaturation curve is close to the sum of the denaturation curves of the three individual domains, indicating the three domains are likely to fold independently in B17. In contrast to the cooperative chemical unfolding of B5.9 and B13-17, the unfolding of B6.4-13, corresponding to the helical domain, is essentially linear, suggesting multistep, noncooperative unfolding. This observation is somewhat surprising, given the prediction based on the structure of lipovitellin where the helical domain contains a large global hydrophobic core between the two layers of helices (*48*, *50*). However, the flexible and adaptable property of this domain may be important in the assembly and secretion of VLDL.

In conclusion, our biochemical and biophysical data provide direct experimental support for the current B17 model based on lipovitellin. Most of our limited proteolysis data can be explained by this model. The three domains within B17 are independent folding units and contain the secondary structures predicted by the model. The clarification of the domain packing and the identification of loop regions in the N-terminal region of apoB should provide new strategies for the structural characterization of this protein.

ACKNOWLEDGMENT

We are grateful to Dr. Gonzalo Hernandez and Dr. Jonathan Lee for their help with mass spectral analysis. Amit Gulati and Dr. Wei Qiu helped us with the insect cell culture. We thank Dr. Donald M. Small, Dr. David Atkinson, and Dr. Haya Herscovitz for their insightful discussions of this work.

SUPPORTING INFORMATION AVAILABLE

A figure demonstrating specific cross-linking between B13 and B13-17. This material is available free of charge via the Internet at http://pubs.acs.org.

REFERENCES

1. Smith, L. C., Pownall, H. J., and Gotto, A. M., Jr. (1978) The plasma lipoproteins: Structure and metabolism, *Annu. Rev. Biochem. 47*, 751−757.
2. Mahley, R. W., Innerarity, T. L., Rall, S. C., Jr., and Weisgraber, K. H. (1984) Plasma lipoproteins: Apolipoprotein structure and function, *J. Lipid Res. 25*, 1277−1294.
3. Kane, J. P. (1983) Apolipoprotein B: Structural and metabolic heterogeneity, *Annu. Rev. Physiol. 45*, 637−650.
4. Brown, M. S., Kovanen, P. T., and Goldstein, J. L. (1981) Regulation of plasma cholesterol by lipoprotein receptors, *Science 212*, 628−635.
5. Skalen, K., Gustafsson, M., Rydberg, E. K., Hulten, L. M., Wiklund, O., et al. (2002) Subendothelial retention of atherogenic lipoproteins in early atherosclerosis, *Nature 417*, 750−754.
6. Cladaras, C., Hadzopoulou-Cladaras, M., Nolte, R. T., Atkinson, D., and Zannis, V. I. (1986) The complete sequence and structural analysis of human apolipoprotein B-100: Relationship between apoB-100 and apoB-48 forms, *EMBO J. 5*, 3495−3507.
7. Yang, C. Y., Chen, S. H., Gianturco, S. H., Bradley, W. A., Sparrow, J. T., et al. (1986) Sequence, structure, receptor-binding domains and internal repeats of human apolipoprotein B-100, *Nature 323*, 738−742.
8. Walsh, M. T., and Atkinson, D. (1983) Solubilization of low-density lipoprotein with sodium deoxycholate and recombination of apoprotein B with dimyristoylphosphatidylcholine, *Biochemistry 22*, 3170−3178.
9. Yang, C. Y., Gu, Z. W., Weng, S. A., Kim, T. W., Chen, S. H., et al. (1989) Structure of apolipoprotein B-100 of human low-density lipoproteins, *Arteriosclerosis 9*, 96−108.
10. Nolte, R. T. (1994) Structural analysis of the human apolipoproteins: An integrated approach utilizing physical and computational methods, Ph.D. Thesis, Boston University, School of Medicine, Boston.
11. Segrest, J. P., Jones, M. K., De Loof, H., and Dashti, N. (2001) Structure of apolipoprotein B-100 in low-density lipoproteins, *J. Lipid Res. 42*, 1346−1367.
12. Spin, J. M., and Atkinson, D. (1995) Cryoelectron microscopy of low-density lipoprotein in vitreous ice, *Biophys. J. 68*, 2115−2123.
13. Orlova, E. V., Sherman, M. B., Chiu, W., Mowri, H., Smith, L. C., et al. (1999) Three-dimensional structure of low-density lipoproteins by electron cryomicroscopy, *Proc. Natl. Acad. Sci. U.S.A. 96*, 8420−8425.
14. Herscovitz, H., Hadzopoulou-Cladaras, M., Walsh, M. T., Cladaras, C., Zannis, V. I., et al. (1991) Expression, secretion, and lipid-binding characterization of the N-terminal 17% of apolipoprotein B, *Proc. Natl. Acad. Sci. U.S.A. 88*, 7313−7317.
15. Carraway, M., Herscovitz, H., Zannis, V., and Small, D. M. (2000) Specificity of lipid incorporation is determined by sequences in the N-terminal 37 of apoB, *Biochemistry 39*, 9737−9745.
16. Shelness, G. S., Hou, L., Ledford, A. S., Parks, J. S., and Weinberg, R. B. (2003) Identification of the lipoprotein initiating domain of apolipoprotein B, *J. Biol. Chem. 278*, 44702−44707.
17. Borchardt, R. A., and Davis, R. A. (1987) Intrahepatic assembly of very low-density lipoproteins. Rate of transport out of the endoplasmic reticulum determines rate of secretion, *J. Biol. Chem. 262*, 16394−16402.
18. Dixon, J. L., Furukawa, S., and Ginsberg, H. N. (1991) Oleate stimulates secretion of apolipoprotein B-containing lipoproteins from Hep G2 cells by inhibiting early intracellular degradation of apolipoprotein B, *J. Biol. Chem. 266*, 5080−5086.
19. Sato, R., Imanaka, T., Takatsuki, A., and Takano, T. (1990) Degradation of newly synthesized apolipoprotein B-100 in a pre-Golgi compartment, *J. Biol. Chem. 265*, 11880−11884.
20. Du, E. Z., Kurth, J., Wang, S. L., Humiston, P., and Davis, R. A. (1994) Proteolysis-coupled secretion of the N terminus of apolipoprotein B. Characterization of a transient, translocation arrested intermediate, *J. Biol. Chem. 269*, 24169−24176.
21. Burch, W. L., and Herscovitz, H. (2000) Disulfide bonds are required for folding and secretion of apolipoprotein B regardless of its lipidation state, *J. Biol. Chem. 275*, 16267−16274.
22. Huang, X. F., and Shelness, G. S. (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.
23. Tran, K., Boren, J., Macri, J., Wang, Y., McLeod, R., et al. (1998) Functional analysis of disulfide linkages clustered within the amino terminus of human apolipoprotein B, *J. Biol. Chem. 273*, 7244−7251.
24. Vukmirica, J., Nishimaki-Mogami, T., Tran, K., Shan, J., McLeod, R. S., et al. (2002) The N-linked oligosaccharides at the amino terminus of human apoB are important for the assembly and secretion of VLDL, *J. Lipid Res. 43*, 1496−1507.
25. Jamil, H., Chu, C. H., Dickson, J. K., Jr., Chen, Y., Yan, M., et al. (1998) Evidence that microsomal triglyceride transfer protein is limiting in the production of apolipoprotein B-containing lipoproteins in hepatic cells, *J. Lipid Res. 39*, 1448−1454.
26. Wetterau, J. R., Gregg, R. E., Harrity, T. W., Arbeeny, C., Cap, M., et al. (1998) An MTP inhibitor that normalizes atherogenic lipoprotein levels in WHHL rabbits, *Science 282*, 751−754.
27. Bakillah, A., Nayak, N., Saxena, U., Medford, R. M., and Hussain, M. M. (2000) Decreased secretion of ApoB follows inhibition of ApoB-MTP binding by a novel antagonist, *Biochemistry 39*, 4892−4899.
28. Hussain, M. M., Bakillah, A., Nayak, N., and Shelness, G. S. (1998) Amino acids 430−570 in apolipoprotein B are critical for its binding to microsomal triglyceride transfer protein, *J. Biol. Chem. 273*, 25612−25615.
29. Mann, C. J., Anderson, T. A., Read, J., Chester, S. A., Harrison, G. B., et al. (1999) The structure of vitellogenin provides a molecular model for the assembly and secretion of atherogenic lipoproteins, *J. Mol. Biol. 285*, 391−408.
30. Segrest, J. P., Jones, M. K., and Dashti, N. (1999) 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. 40*, 1401−1416.
31. Dashti, N., Gandhi, M., Liu, X., Lin, X., and Segrest, J. P. (2002) The N-terminal 1000 residues of apolipoprotein B associate with microsomal triglyceride transfer protein to create a lipid transfer pocket required for lipoprotein assembly, *Biochemistry 41*, 6978−6987.
32. Baker, M. E. (1988) Is vitellogenin an ancestor of apolipoprotein B-100 of human low-density lipoprotein and human lipoprotein lipase? *Biochem. J. 255*, 1057−1060.
33. Meininger, T., Raag, R., Roderick, S., and Banaszak, L. J. (1984) Preparation of single crystals of a yolk lipoprotein, *J. Mol. Biol. 179*, 759−764.
34. Byrne, B. M., Gruber, M., and Ab, G. (1989) The evolution of egg yolk proteins, *Prog. Biophys. Mol. Biol. 53*, 33−69.
35. Tatusova, T. A., and Madden, T. L. (1999) BLAST 2 Sequences, a new tool for comparing protein and nucleotide sequences, *FEMS Microbiol. Lett. 174*, 247−250.
36. Herscovitz, H., Derksen, A., Walsh, M. T., McKnight, C. J., Gantz, D. L., et al. (2001) The N-terminal 17% of apoB binds tightly and irreversibly to emulsions modeling nascent very low-density lipoproteins, *J. Lipid Res. 42*, 51−59.
37. Sali, A., and Blundell, T. L. (1993) Comparative protein modelling by satisfaction of spatial restraints, *J. Mol. Biol. 234*, 779−815.
38. Kabsch, W., and Sander, C. (1983) Dictionary of protein secondary structure: Pattern recognition of hydrogen-bonded and geometrical features, *Biopolymers 22*, 2577−2637.
39. Sambrook, J., and Russell, D. W. (2000) *Molecular Cloning: A Laboratory Manual*, 3rd ed., Cold Spring Harbor Laboratory Press, Plainview, NY.
40. Khachfe, H. M. (2002) Spectroscopic and calorimetric studies of the 17% N-terminal domain of apolipoprotein B-100, Ph.D. Thesis, Boston University, School of Medicine, Boston.
41. Edelhoch, H. (1967) Spectroscopic determination of tryptophan and tyrosine in proteins, *Biochemistry 6*, 1948−1954.
42. Schagger, H., and von Jagow, G. (1987) Tricine-sodium dodecyl sulfate-polyacrylamide gel electrophoresis for the separation of proteins in the range from 1 to 100 kDa, *Anal. Biochem. 166*, 368−379.
43. Sreerama, N., and Woody, R. W. (2000) Estimation of protein secondary structure from circular dichroism spectra: Comparison of CONTIN, SELCON, and CDSSTR methods with an expanded reference set, *Anal. Biochem. 287*, 252−260.
44. Provencher, S. W. (1982) CONTIN: A general purpose constrained regularization program for inverting noisy linear algebraic and integral equations, *Comput. Phys. Commun. 27*.
45. Sreerama, N., and Woody, R. W. (1994) Protein secondary structure from circular dichroism spectroscopy. Combining variable selection principle and cluster analysis with neural network, ridge regression and self-consistent methods, *J. Mol. Biol. 242*, 497−507.
46. Manavalan, P., and Johnson, W. C., Jr. (1987) Variable selection method improves the prediction of protein secondary structure from circular dichroism spectra, *Anal. Biochem. 167*, 76−85.
47. Shelness, G. S., and Sellers, J. A. (2001) Very-low-density lipoprotein assembly and secretion, *Curr. Opin. Lipidol. 12*, 151−157.
48. Anderson, T. A., Levitt, D. G., and Banaszak, L. J. (1998) The structural basis of lipid interactions in lipovitellin, a soluble lipoprotein, *Structure 6*, 895−909.
49. Thompson, J. R., and Banaszak, L. J. (2002) Lipid−protein interactions in lipovitellin, *Biochemistry 41*, 9398−9409.
50. Ohlendorf, D. H., Wrenn, R. F., and Banaszak, L. J. (1978) Three-dimensional structure of the lipovitellin-phosvitin complex from amphibian oocytes, *Nature 272*, 28−32.
51. Koradi, R., Billeter, M., and Wuthrich, K. (1996) MOLMOL: A program for display and analysis of macromolecular structures, *J. Mol. Graphics 14*, 51−55, 29−32.

BI048286Y