Modular Structure of Solubilized Human Apolipoprotein B-100
LOW RESOLUTION MODEL REVEALED BY SMALL ANGLE NEUTRON SCATTERING

Alexander Johs‡, Michal Hammel‡, Ines Waldner‡, Roland P. May‡, Peter Laggner‡, and Ruth Prassl‡

From the ‡Institute of Biophysics and X-ray Structure Research, Austrian Academy of Sciences, Schmiedlstrasse 6, A-8042 Graz, Austria, the §Division of Cell Biology and Biophysics, School of Biological Sciences, University of Missouri, Kansas City, Missouri 64110, and the ¶Institut Laue-Langevin, 6, rue Jules Horowitz, BP 156, 38042 Grenoble Cedex 9, France

Being intimately involved in cholesterol transport and lipid metabolism human low density lipoprotein (LDL) plays a prominent role in atherogenesis and cardiovascular diseases. The receptor-mediated uptake of LDL is triggered by apolipoprotein B-100 (apoB-100), which represents the single protein moiety of LDL. Due to the size and hydrophobic nature of apoB-100, its structure is not well characterized. Here we present a low resolution structure of solubilized apoB-100. We have used small angle neutron scattering in combination with advanced shape reconstruction algorithms to generate a three-dimensional model of lipid-free apoB-100. Our model clearly reveals that apoB-100 is composed of distinct domains connected by flexible regions. The apoB-100 molecule adopts a curved shape with a central cavity. In comparison to LDL-associated apoB-100, the lipid-free protein is expanded, whereas according to spectroscopic data the secondary structure is widely preserved. Finally, the low resolution model was used as a template to reconstruct a hypothetical domain organization of apoB-100 on LDL, including information derived from a secondary structure prediction.

Apolipoprotein B-100 (apoB-100) is the sole protein component of low density lipoprotein (LDL) and encompasses a variety of functions in lipid metabolism. Like all plasma lipoproteins, LDL takes the shape of a globular particle consisting of a non-polar lipid core surrounded by an amphipathic coating of protein, phospholipid, and cholesterol (1, 2). LDL is the principal plasma cholesterol carrier and serves as a source of cholesterol for most tissues of the body through receptor-mediated recognition of apoB-100. From the medical point of view, LDL may become involved in a variety of metabolic disorders, such as hypercholesterolemia, hyperlipidemia, or atherogenesis, which are often linked to a structurally determined dysfunction of apoB-100.

Human apoB-100 is a glycoprotein with a molecular mass of about 550 kDa, consisting of 4536 amino acid residues. It is a single chain protein that is associated with hydrophobic molecules in a noncovalent fashion to facilitate their transport and targeting in a hydrophilic environment. The apolipoprotein is synthesized in hepatocytes, where the assembly of the lipoprotein particle takes place, which is mediated by the first 884 amino acid residues of apoB-100 in the presence of a microsomal triglyceride transfer protein (3). Furthermore, apoB-100 is a carrier of cell targeting signals that facilitate the uptake of lipoprotein particles by receptor-mediated endocytosis (4). A domain close to the COOH-terminal end, which is substantially enriched in basic amino acids (residues 3345–3381), was identified as the receptor binding domain (5–7).

A detailed analysis of the secondary structure by computational methods resulted in the so-called pentapartite model (8) proposing five consecutive domains with alternating α-helix and β-strand regions. According to this model the predicted domain organization for apoB-100 is NH₂–β₂α₁β₁–α₂β₂–α₃–COOH. The amphipathic nature of these secondary structural elements determines the extent of association with the lipid core. This information was gained by exploring the relative accessibility for trypsin digestion (9). The NH₂-terminal domain including residues 1–1000 (β₆₈) was found to be highly trypsin-releasable suggesting that it is not associated with the lipid core in LDL. In addition, this domain contains more than half of the total 25 cysteine residues of apoB-100, all of which are involved in disulfide bonds with other cysteine residues in the same domain. This is a strong indication for a very compact folding of this region, most likely in the form of an independent globular domain that was shown to be highly homologous to the structure of lipovitellin (10). The remaining domains (β₁–α₂β₂–α₃) are less characterized and were found to be more closely associated to the lipid core of LDL.

Over time several strategies were followed to elucidate the structure and morphology of LDL and apoB-100 on LDL by different biophysical techniques (11–14). The mapping of apoB-100 on the surface of LDL by triangulation techniques with monoclonal antibodies generated the generally accepted idea that the protein is a belt wrapped once around the LDL particle (15, 16). Studies using cryoelectron microscopy (cryo-EM) in vitreous ice resulted in a three-dimensional low resolu-
tion model of the LDL particle (17). Nevertheless, despite the successful crystallization of LDL (18–20), a high resolution structure of apoB-100 in LDL is not available yet and awaits further clarification.

Other approaches were applied to investigate the morphology of apoB-100 after extraction from its native lipid environment. Several detergents were used to solubilize apoB-100, and the resulting protein-detergent complex was subjected to negative staining EM and cryo-EM in vitreous ice (21–25). To a certain extent, these results differ depending on preparation procedure and imaging technique used. Curved shapes and horseshoe or thread-like morphologies were found. Flexible strings up to 1000 Å long, with width varying from 20 to 70 Å, were imaged. However, a common feature in all studies is that lipid-free apoB-100 adopts an elongated conformation.

As outlined above, the most conclusive experimental results on the structure of apoB-100 are reported from EM imaging, but they lack three-dimensional information. To address this issue, we have applied small angle neutron scattering (SANS) combined with the contrast matching technique to elucidate three-dimensional characteristics of lipid-free apoB-100 and to assess its structural flexibility at the same time. We have restored a low resolution model of apoB-100 while eliminating contributions of the solubilizing detergent. Finally, based on our low resolution structure model in combination with a secondary structure prediction, a hypothetical arrangement of distinct modules of apoB-100 within LDL is proposed.

**EXPERIMENTAL PROCEDURES**

**Preparation of LDL from Human Plasma**—Blood was taken from normolipidemic, single donors. To prevent clotting, oxycare, microbial degradation, and proteolytic cleavage, 1 mg/ml Na₂EDTA, 0.05% gentamycin sulfate (Serva, Heidelberg, Germany) and a protease inhibitor mixture (Roche Diagnostics, Basel, Switzerland) were added. Whole blood was centrifuged at 4500 rpm, 10 min, 5 °C, and the supernatant was carefully separated from the pellet and stored in aliquots at −70 °C after addition of 10% (w/v) sucrose as a cryoprotectant.

LDL (density range 1.02–1.063) was prepared by ultracentrifugation and dialyzed at 4 °C against standard buffer (10 mM sodium phosphate, 1 mg/ml Na₂EDTA, 0.05% gentamycin sulfate, pH 7.4). LDL samples were assessed for protein concentration by the BCA method (BCA protein assay kit, Pierce) Purity and homogeneity of solubilized apoB-100 were determined by SDS-PAGE. The protein concentration was determined by the BCA method, and purity and homogeneity of solubilized apoB-100 were assessed by SDS-PAGE. The concentration of Nonidet P-40 was determined by absorbance measurements at 280 nm. The absorption coefficient for Nonidet P-40 was calculated from linear regression of a standard dilution series in deionized water. The theoretical absorption from the protein was calculated from the amino acid composition (28) and subtracted from absorption values of samples containing both protein and Nonidet P-40 to give accurate values for detergent concentration.

**Circular Dichroism Spectroscopy**—Far-UV/CD spectra were obtained in a wavelength range from 190 to 250 nm with a Jasco J-715 spectropolarimeter (Jasco Inc., Easton, MD) equipped with a personal computer. Each spectrum was recorded as an average of three scans taken with the following parameters: step resolution, 0.5 nm; scan speed, 50 nm/min; response, 1 s; and bandwidth, 1 nm. The spectra were base line-corrected by subtracting the signal of detergent in standard buffer. The results were expressed in terms of mean residue ellipticity [θ] (degree cm² dmol⁻¹) defined as follows.

\[
[\theta] = \frac{\theta_{obs}}{10 \cdot c \cdot l}
\]  
(Eq. 1)

θ_{obs} is the measured ellipticity in degrees, c is the protein residue concentration in mol/liter, and l is the length of the light path in cm. The ellipticity data were analyzed for secondary structure by a non constrained least squares analysis using the CDSSTR algorithm based on reconstruction of experimental data (29, 30).

**SANS Experiments**—Neutron scattering data were obtained at beam lines D11 and D22 using the high flux reactor at the Institut Laue-Langevin (Grenoble, France). To determine the contrast match point of the detergent contrast variation experiments were performed on instrument D11. Scattering profiles were recorded of 28 mM Nonidet P-40 dissolved in buffer at several D₂O/H₂O ratios (0, 10, 20, 30, 40, and 100% D₂O). This detergent concentration, which is about 10 times the critical micellar concentration, was chosen to be the same as in the apoB-100-Nonidet P-40 complex used for further SANS experiments.

Sample-detector distances of 1.2, 3, and 12 m (collimations 5.5, 5.5, and 13.5 m), circular cells with a thickness of 2 mm, and a neutron wavelength of 6 Å (Δλ/λ of 10%) were used.
Low Resolution Structure of Solubilized ApoB-100

The contrast match point is given by the D$_2$O concentration at which the intensity of forward scattering approaches zero, \( I(0) = 0 \).

\( I(0) \) values were determined by extrapolation of the scattering profiles to zero angle using the Guinier approximation (31). The normalized intensities \( (I(0)/c)^{1/2} \) were plotted as a function of D$_2$O content (data not shown). The intercept of this linear dependence on the abscissa indicates the contrast match point of the detergent. In case of Nonidet P-40, the scattering intensity was matched at 18% D$_2$O.

The apoB-100-Nonidet P-40 complex was measured at the match point concentration of 18% D$_2$O/H$_2$O on instrument D22 at a protein concentration of 6.4 mg/ml and a detergent concentration of 28 mM using a rectangular cell with a thickness of 2 mm positioned in a thermostatted sample rack at 8 °C. The typical acquisition time was 30 min. The exposure of the sample was repeated 12 times to verify sample stability. Sample-detector/collimation distances of 14.0/14.4 m and 2.8/2.8 m, a wavelength \( \lambda \) of 6 Å \((\Delta \lambda/\lambda \text{ of 10%})\), and a rectangular beam aperture of 16 × 16 mm were used. To correct for background scattering the Nonidet P-40/buffer curve measured at 18% D$_2$O/H$_2$O was subtracted.

Data reduction was based on standard ILL software using the programs DETEC, RNILS, SPOILLY, RGUI, and RPILOT (32). The data acquired at both sample-detector distances were merged resulting in a total \( q \)-range of 0.004–0.32 Å$^{-1}$ \((q = 4\pi\sin(\theta)/\lambda, 2\theta \text{ is the scattering angle})\) used for further calculations.

Analysis of Reduced Scattering Data—Radius of gyration \( (R_g) \) of apoB-100 was determined by the Guinier approximation from the low \( q \)-regions \((qR_g \leq 1.3)\) of the scattering profiles (31) according to the following equation.

\[
\ln(I(q)) = \ln(I(0)) - \frac{R_g^2 \cdot q^2}{3} \quad \text{(Eq. 2)}
\]

At small angles when plotted as \( \log(I(q)) \) versus \( q^2 \) the scattering profile should be linear although this relationship is rigorously valid only for \( q \leq R_g^{-1} \) (33).

The pair-distance distribution function \( p(r) \),

\[
p(r) = \frac{1}{2\pi} \int_0^\infty r \cdot q \cdot I(q) \cdot \sin(r \cdot q) \cdot dq \quad \text{(Eq. 3)}
\]

was obtained by indirect Fourier transformation of the scattered intensities \( I(q) \) using the program GNOM (34). The \( p(r) \) corresponds to the distribution of distances \( r \) between any two volume elements within one particle weighted by the product of their scattering length density relative to that of the solvent. This offers an alternative calculation of \( R_g \) that is based on the full scattering curve and gives the maximum dimension of the macromolecule \( D_{\text{max}} \), as the distance where the \( p(r) \) function approaches zero (35).

Ab Initio Modeling—The low resolution shape of apoB-100 was restored from experimental data using the program DAMMIN (36). The scattering curve up to \( q_{\text{max}} = 0.25 \text{ Å}^{-1} \) was used for fitting corresponding to a resolution of ~25 Å \((2\pi/q_{\text{max}})\). DAMMIN calculates the scattering intensities from a multiphase model of a particle constructed of a finite number of dummy beads. The beads are characterized by a configuration vector \( (X) \) assigning them to a phase or to the solvent. The program starts searching for a model in a volume filled by hexagonally packed uniform beads. The dimensions were selected according to the resolution of the scattering profile while the overall dimension of the initial search object is taking into account. DAMMIN searches for the best model configuration minimizing the discrepancy function \( f(X) = \chi^2 + \alpha P(X) \) using the simulated annealing method (37), where \( \chi \) is the discrepancy between the calculated and experimental curve, and \( \alpha P(X) \) is a looseness penalty with positive weight \( \alpha > 0 \). The aim of this method is to randomly modify the coordinates of the beads while always approaching the configurations that decrease the energy \( f(X) \).

The bead radius used for the ab initio modeling of solubilized apoB-100 was set to 14 Å. This value was chosen due to experimental results \((D_{\text{max}} = 600 \text{ Å, resolution } \sim 25 \text{ Å})\) and certain constraints within the program DAMMIN. Different independent models with similar goodness of fit were obtained. Ten of these models were used for automated averaging by the program DAMAVER (38). This program computes the values for the normalized spatial discrepancy (NSD) between each pair in the set. The average NSD value for each model was calculated with respect to the rest of the set and the model with the lowest NSD was selected as reference model. All other models were superimposed onto the reference model using the program SUPCOMB (39). Finally, the entire assembly of beads was remapped onto a densely packed grid of beads where each grid point was characterized by its occupancy factor. The portion of points with higher non-zero occupancy was selected to yield the volume equal to the average excluded volume of the models.

Secondary Structure Prediction—The complete amino acid sequence of apoB-100 was subjected to the automated secondary structure prediction service SSpro 2.0 using bidirectional recurrent neural networks (40). The results of this prediction were visualized using POLYVIEW (41).

RESULTS

Circular Dichroism Spectroscopy—To assess the effects of delipidation on protein folding, circular dichroism spectra of both solubilized apoB-100 and native LDL were recorded and evaluated. Secondary structure contents were calculated from the normalized spectra (Fig. 1) and expressed as percent secondary structure as shown in Table 1.

These results are in agreement with a secondary structure prediction, and the values obtained for native LDL are consistent with data reported previously (42). The content of \( \alpha \)-helix and \( \beta \)-strands of detergent-solubilized apoB-100 appears slightly higher than that of native LDL, while the contents of turns and coil structures seem to be slightly lower. However, within the experimental error margin, the overall secondary structure of Nonidet P-40 solubilized apoB-100 is well conserved. Similar results were reported by Watt and Reynolds (26).

3 For further details the reader is referred to the web site www.embl-hamburg.de/ExternalInfo/Research/Sax/dammin.html showing an animation of this fitting approach.
for apoB-100 solubilized by n-dodecyl octaethylene glycol monoether, a detergent that is closely related to Nonidet P-40, although due to the choice of a different evaluation method, the absolute values in the results vary.

**SANS Data**—The final scattering curve was obtained by merging the scattering profiles of apoB-100 solubilized in Nonidet P-40 collected at two sample-detector distances (Fig. 2A).

The $R_G$ value derived from the Guinier approximation was $150 \pm 100 \text{ Å}$ (Fig. 2A, inset). The linearity of the fit suggests that the data accurately reflect the non-aggregated state of the protein in solution. To verify this, the molecular weight of the protein was calculated.

In a two-component system, the molecular weight of a particle in solution is proportional to $I(0)$ according to the following equations (43, 44),

\[
I(0) = \frac{c \cdot N_A}{M} \left[ \sum_i (b_i - \rho^0 \cdot \nu_i) \right]^2 \tag{Eq. 4}
\]

whereas

\[
\sum_i (b_i - \rho^0 \cdot \nu_i) = \frac{M}{N_A} (B - \rho^0 \cdot \bar{\nu}). \tag{Eq. 5}
\]

$M$ is the molecular mass, $c$ stands for the protein concentration, $b_i$ is the scattering length contained in a small volume, $\nu_i$, $\rho^0$ is the scattering length density of the solvent, $B$ is the scattering

**TABLE 1**

Secondary structure calculated from circular dichroic spectra by the CDSSTR method

<table>
<thead>
<tr>
<th>Sample</th>
<th>$\alpha$-Helix</th>
<th>$\beta$-Strand</th>
<th>Turns</th>
<th>Coil</th>
<th>NRMSD$^a$</th>
</tr>
</thead>
<tbody>
<tr>
<td>LDL</td>
<td>37%</td>
<td>17%</td>
<td>20%</td>
<td>25%</td>
<td>0.010</td>
</tr>
<tr>
<td>ApoB-100</td>
<td>40%</td>
<td>21%</td>
<td>19%</td>
<td>16%</td>
<td>0.007</td>
</tr>
<tr>
<td>Predicted$^b$</td>
<td>36%</td>
<td>23%</td>
<td>41%</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

$^a$ The NRMSD value indicates the root mean square deviation of fits to measured spectra.

$^b$ Results are from the secondary structure prediction. This algorithm is not capable of discerning turns and coil structures, thus a sum of both is given.

FIGURE 1. Far-UV circular dichroism spectra of delipidated apoB-100 in comparison with native LDL. Circular dichroic spectra of delipidated apoB-100 (dotted line) and native LDL (solid line) were recorded in a 0.02-cm cell thermostatted to 20 °C at protein concentrations of $5.5 \times 10^{-6}$ M and $2.5 \times 10^{-6}$ M, respectively. Raw spectra of apoB-100 and LDL were corrected for contribution of either 2.5 mM Nonidet P-40 or buffer and transformed to mean residue ellipticities.

**TABLE 1**

Secondary structure calculated from circular dichroic spectra by the CDSSTR method

<table>
<thead>
<tr>
<th>Sample</th>
<th>$\alpha$-Helix</th>
<th>$\beta$-Strand</th>
<th>Turns</th>
<th>Coil</th>
<th>NRMSD$^a$</th>
</tr>
</thead>
<tbody>
<tr>
<td>LDL</td>
<td>37%</td>
<td>17%</td>
<td>20%</td>
<td>25%</td>
<td>0.010</td>
</tr>
<tr>
<td>ApoB-100</td>
<td>40%</td>
<td>21%</td>
<td>19%</td>
<td>16%</td>
<td>0.007</td>
</tr>
<tr>
<td>Predicted$^b$</td>
<td>36%</td>
<td>23%</td>
<td>41%</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

$^a$ The NRMSD value indicates the root mean square deviation of fits to measured spectra.

$^b$ Results are from the secondary structure prediction. This algorithm is not capable of discerning turns and coil structures, thus a sum of both is given.

**FIGURE 2. Neutron scattering curves.** A, experimental SANS curve (circles) and the scattering profile computed from the ab initio model (line). The ranges corresponding to those used for the calculations of $R_G$, $R_{w_1}$, and $R_{w_2}$ are indicated. Inset: Guinier plot of the SANS data. The square indicates the $I(q)$ data used to obtain $R_G$. The straight line corresponds to the best fit through these points and extrapolation to $q = 0$ yields $I(0)$. $B$, cross-sectional Guinier fit within a $q$ range of $0.0075-0.016 \text{ Å}^{-1}$ for $R_{w_1}$ and $0.017-0.04 \text{ Å}^{-1}$ for $R_{w_2}$. C, distance distribution function computed from the experimental SANS data. The maximum dimension of the $p(r)$ is denoted by $D_{\text{max}}$. 

**FIGURE 2. Neutron scattering curves.** A, experimental SANS curve (circles) and the scattering profile computed from the ab initio model (line). The ranges corresponding to those used for the calculations of $R_G$, $R_{w_1}$, and $R_{w_2}$ are indicated. Inset: Guinier plot of the SANS data. The square indicates the $I(q)$ data used to obtain $R_G$. The straight line corresponds to the best fit through these points and extrapolation to $q = 0$ yields $I(0)$. $B$, cross-sectional Guinier fit within a $q$ range of $0.0075–0.016 \text{ Å}^{-1}$ for $R_{w_1}$ and $0.017–0.04 \text{ Å}^{-1}$ for $R_{w_2}$. C, distance distribution function computed from the experimental SANS data. The maximum dimension of the $p(r)$ is denoted by $D_{\text{max}}$. 

JULY 14, 2006 • VOLUME 281 • NUMBER 28 JOURNAL OF BIOLOGICAL CHEMISTRY 19735
**Low Resolution Structure of Solubilized ApoB-100**

length per unit mass, and \( v \) represents the partial specific volume of the macromolecule. The amount of exchangeable protons was assumed to be 80% resulting in a \( B \) value of \( 2.55 \times 10^{-14} \) cm. The scattered intensity at zero angle \( I(0) \) was obtained by extrapolation of the Guinier plot to \( q^2 \to 0 \). \( I(0) \) was found to be 0.61 cm\(^{-1}\).

The molecular mass calculated from the zero angle intensity was 547,000 Da, somewhat higher than 512,816 Da as calculated from the amino acid sequence. The discrepancy between the values can be explained by the contribution of glycosylated residues accounting for up to 7% of the molecular mass and a certain error in the protein concentration. The protein concentration was 547,000 Da, somewhat higher than 512,816 Da as calculated from the amino acid sequence. The discrepancy between the values was 547,000 Da, somewhat higher than 512,816 Da as calculated from the amino acid sequence.

The calculation of the gyration of the cross-section, the mean radius of gyration of the cross-section \( R_{xs} \), and the mean cross-sectional intensity at zero angle \( I(q) \) were identified in separate regions. Applying this formalism to our results, two linear regions with similar qualities of fit \( \chi \sim 0.6 \) display elongated bowed shapes and a central cavity. An average shape was calculated by superposition and alignment of 10 separate models (shown in Fig. 3A). The probability map with a defined cutoff results in an average shape that has a more compact character compared with the single models. Nevertheless, the averaged low resolution model confirms the results obtained for the single models. One can identify several distinct modules and a bent overall shape with a cavity located in the center of the particle (Fig. 3B). However, it has to be emphasized that the final refined model cannot be considered as a single unique protein conformation (51).

**DISCUSSION**

Using small angle neutron scattering combined with contrast matching, we have succeeded in constructing a three-dimensional model of delipidated apoB-100 as a constituent of a protein-detergent complex. *Ab initio* shape reconstruction from SANS data revealed models with similar goodness of fit. All of them showed typical characteristics that are conserved in the final averaged model. The most striking features of this model are its arcuate nature and the presence of a pronounced cavity in the center of the molecule. In particular, one can distinguish independent modules seen as alternating wide and narrow sections or protrusions in the model that are most likely connected by linkers. We assume that these linker regions confer a certain structural flexibility to the molecule, which is an important point concerning the spatial expansion of apoB-100 on lipoproteins.
Evidence for a domain arrangement of apoB-100 within LDL was already proposed in the literature (52). Our data strongly support these findings and additionally indicate that apoB-100 solubilized from native LDL by a non-ionic detergent retains this modular configuration. Similarly, results of a differential scanning calorimetry study describing the melting transition of both native and solubilized apoB-100 as a non-cooperative process of independent domains (13) are confirmed. In this respect, our results are also consistent with previous investigations by EM, which describe solubilized apoB-100 as a long flexible beaded thread (23).

Upon consideration of the conformations of solubilized and native apoB-100 several aspects become evident. First and most important for the interpretation of our data, the secondary structure as estimated in terms of \( \alpha \)-helical and \( \beta \)-sheet content is retained during delipidation. This is a strong indication that the characteristic folding of particular domains is not affected by delipidation. Second, the tertiary structure of solubilized apoB-100 is clearly different from that of apoB-100 wrapped around the surface of LDL particles. Native lipid-associated apoB-100 is confined to the geometry of LDL and surrounds the particle with the NH2- and COOH-terminal regions located close to each other (16). Moreover, apoB-100 is a highly amphiphilic protein that repeatedly penetrates the phospholipid surface monolayer or even the hydrophobic lipid interior of LDL. These specific features result in a tight integration with lipids and a highly condensed arrangement of apoB-100 in LDL. During the solubilization process apoB-100 is released from the lipidic environment, and the NH2- and COOH-terminal ends become separated due to the inherent flexibility of the molecule. Taking into account, that the hydrophobic regions in apoB-100 are distributed throughout the macromolecule (Ref. 53; in contrast to integral membrane proteins, where lipids are located within a narrow region of the protein), the surfactant molecules are spread all over the hydrophobic parts of the protein surface. This pronounced shielding effect of surfactants might be the driving force for the rearrangement of domains as well as for the conservation of the arcuate overall shape of apoB-100.

To relate the modular arrangement of our model to the primary sequence of apoB-100, we performed a secondary structure prediction. First, we correlated the results of our secondary structure prediction to the pentapartite model proposed by Segrest et al. (8). This model describes the molecule as an assembly of distinct domains with a defined sequence of amphipathic \( \alpha \)-helix- and \( \beta \)-strand-rich clusters as NH2-\( \beta \alpha_1-\beta_1-\alpha_2-\beta_2-\alpha_3-\)COOH. As shown in Fig. 4A, a high level of consistency was achieved. Second, a recently published homology model of the NH2-terminal domain of apoB-100 (residues 1–1000) has confirmed the compact globular arrangement of the \( \beta \alpha_1 \)-domain (54). This view is supported by cryo-EM studies, which supply evidence that the NH2-terminal domain protrudes as a knob or pointed end out of the LDL particle (17, 55). The size of this protrusion matches the dimensions of the homologous lipovitellin-like domain with a diameter of \( \sim 110 \) Å (54, 56). In regard to our ab initio model, the size and shape of this domain fits to one of the two ends of our model, although an accurate superposition with the homology model is not feasible due to the limited resolution of the reconstructed model. Accordingly, we have assigned the NH2 terminus as origin for mapping the results of the secondary structure prediction to our model.

So far, no adequate templates are available for structural modeling of apoB-100 domains except for the NH2-terminal domain. Thus, any statements on detailed structural characteristics of domains still remain speculative. Nevertheless, two \( \beta \)-strand regions, \( \beta_1 \) and \( \beta_2 \), encompassing residues 1000–2000
and 2600–4000 could be identified by secondary structure prediction. These \( \beta \)-strands (9) serve as irreversibly lipid-associated domains in LDL and anchor the protein to the lipid core. We assume that delipidation has preserved the secondary structure; however, the amphipathic \( \alpha \)-helices are released from the lipidic environment and adopt a sterically more favored overall arrangement as revealed in the low resolution model. A reversible lipid-associated \( \alpha \)-helical domain, located precisely in the middle of the molecule, approxi-
mately located between residues 2100 and 2600 (corresponding to 44–56% of the amino acid sequence shown in Fig. 4A). If assigned to the three-dimensional model, this domain becomes located in the cavity region (Fig. 4B).

A loop was proposed near the carboxy-terminal end of apoB-100 (16) that is stabilized by the binding of Arg-3500 to Trp-4369 (57) and modulates the binding of the LDL receptor to the LDL receptor binding site (residues 3359–3369). Approximately 20% of the total 4536 amino acids in apoB-100 is part of this loop. Based on the geometry of our model we can identify a pronounced bulky region preceding the COOH-terminus endorsing the presence of a loop (Fig. 4B). Nevertheless, the low resolution of the reconstructed model does not allow the course of the protein chain to be followed. Hydrophilic sites, in particular the receptor binding site, which is characterized by clusters of polar basic residues (7, 9), as well as glycosylation sites (10 of 16 N-glycosylation sites of apoB-100 being located between residues 2752 and 4210), should be assigned along the outer perimeter very close to this loop region and may additionally account for the bulkiness of this domain.

The COOH-terminal domain, which corresponds to amino acid residues 4100 through 4500, has been suggested to form a dense cluster of amphipathic \( \alpha \)-helices, which are reversibly lipid-associated (9). In our prediction this \( \alpha \)-domain comprises 11% of the amino acid sequence and when mapped onto the three-dimensional model nicely fits to the small, turned back end region in the model.

Finally, in an attempt to reconstruct an LDL particle we have accommodated the \textit{ab initio} model on a sphere. The simplest way to mimic the situation on LDL was to apply an arbitrary movement of the COOH-terminal domain toward the NH\(_2\)-terminal domain. This single modification resulted in a ring-like shape with the COOH-terminal end close to the NH\(_2\)-terminal end (Fig. 4C). Indeed, the diameter of this ring-shaped structure roughly corres-
dponded to the diameter of an intact LDL-particle (~250 Å) as depicted schematically in Fig. 4D. This hypothetical model of apoB-100 on LDL also reflects the protrusion seen by EM that is formed by the compact globular structure of the NH\(_2\)-terminal domain (17, 55). Hence, we speculate that the averaged low resolution model of solubilized apoB-100 presented in this study could equally be representative for the conformation and morphology of native apoB-100 in LDL.

Taken together, the present results open new perspectives on the structure of apoB-100 as they confirm and extend current knowledge from EM. Our data clearly emphasize that apoB-100 is partly flexible in nature. Similarly, the protein is subject to structural constraints and thus retains its secondary structure and in large parts its inherent curvature even after delipidation. An assignment of secondary structure information to the low resolu-

---

**FIGURE 4. Secondary structure prediction and theoretical model of apoB on LDL.** A, prediction of secondary structure based on the complete amino acid sequence of apoB-100; \( \alpha \)-helical structures are shown in red, \( \beta \)-sheet structures are shown in blue, and random coil and parts with low prediction confidence are shown in gray. Predictions were classified according to the pentapartite structure proposed by Segrest et al. (8). B, mapping of secondary structure information onto the low resolution model. The position of the NH\(_2\)-terminal domain was selected with respect to the dimensions of the homology model (54). Secondary structure elements were allocated proportionally to the fraction of their occurrence in the prediction. C, hypothetical model of the spatial arrangement of apoB in LDL after performing an arbitrary movement of the COOH-terminal domain toward the NH\(_2\)-terminal domain. D, model of the LDL particle after superposition of a 250 Å sphere onto the structure model of apoB-100 representing the lipid components.
tion model gave insights into the spatial organization of secondary structural elements within the protein. Thus, the present work provides fundamentals toward future endeavors to reveal a more detailed structure of lipid-free apoB-100 at higher resolution, i.e. by x-ray crystallography. In doing so, certain difficulties imposed by the structural flexibility and amphipathic properties of this protein will have to be taken into account.

Acknowledgments—We thank Veronika Sattler and Bernadette Zanner for technical and Paul Debagge for editorial assistance. We acknowledge—We thank Veronika Sattler and Bernadette Zanner for technical and Paul Debagge for editorial assistance. We thank Veronika Sattler and Bernadette Zanner for technical and Paul Debagge for editorial assistance. We thank Veronika Sattler and Bernadette Zanner for technical and Paul Debagge for editorial assistance.
Modular Structure of Solubilized Human Apolipoprotein B-100: LOW RESOLUTION MODEL REVEALED BY SMALL ANGLE NEUTRON SCATTERING
Alexander Johs, Michal Hammel, Ines Waldner, Roland P. May, Peter Laggner and Ruth Prassl

doi: 10.1074/jbc.M601688200 originally published online May 16, 2006

Access the most updated version of this article at doi: 10.1074/jbc.M601688200

Alerts:
• When this article is cited
• When a correction for this article is posted

Click here to choose from all of JBC's e-mail alerts

This article cites 52 references, 17 of which can be accessed free at http://www.jbc.org/content/281/28/19732.full.html#ref-list-1