Structure of Apolipoprotein B-100 of Human Low Density Lipoproteins

Chao-yuh Yang, Zi-Wei Gu, Shi-ai Weng, Tae W. Kim, San-Hwan Chen, Henry J. Pownall, Paul M. Sharp, Shyan-Weoi Liu, Wen-Hsiung Li, Antonio M. Gotto, Jr., and Lawrence Chan

We have analyzed low density lipoproteins (LDL) apolipoprotein (apo) B structure by direct sequence analysis of LDL apo B-100 tryptic peptides. Native LDL were digested with trypsin, and the products were fractionated on a Sephadex G-50 column. The partially digested apo B-100 still associated with lipids was recovered in the void volume (designated trypsin-nonreleasable, TN, peptides). The released peptides (designated trypsin-releasable, TR, peptides) in subsequent peaks were repurified on two successive high-performance liquid chromatography (HPLC) columns. The TN peak was delipidated and redigested with trypsin, and the resulting peptides were purified on two successive HPLC columns. Using this approach, we sequenced over 88% of LDL apo B-100, extending and refining our previous study (Nature 1986;323:738-742) which covered 52% of the protein. TN peptides made up 31%, and the TR peptides, 34% of the apo B-100 sequence; 23.7% were found under both TN and TR categories. Based on its differential trypsin releasability, apo B-100 can be divided into five domains: 1) residues 1—1000, largely TR; 2) residues 1001—1700, alternating TR and TN; 3) residues 1701—3070, largely TN; 4) residues 3071—4100, mainly TR and mixed; and 5) residues 4101—4536, almost exclusively TN. Domain 1 contained 14 of the 25 Cys residues in apo B. Domain 4 encompassed seven N-glycosylation sites, and contained the putative receptor binding domains. All 19 potential N-glycosylation sites were directly sequenced: 16 were found to be glycosylated and three were not. Three pairs of disulfide bridges were also mapped. Finally, a combination of cDNA sequencing, direct mRNA sequencing, and comparison of published apo B-100 sequences allowed us to identify specific amino acid residues within apo B-100 that seem to represent bona fide allelic variations. Our study provides information on LDL apo B-100 structure that will be important to our understanding of its conformation and metabolism. (Arteriosclerosis 9:96-108, January/February 1989)

Low density lipoproteins (LDL) are the major carriers of cholesterol in plasma, and their increased concentration in circulation is correlated with the development of atherosclerosis. In normal individuals, very low density lipoproteins (VLDL) are the major precursor lipoproteins of plasma LDL. LDL are removed from the circulation by both high-affinity receptor-mediated and receptor-independent pathways, the liver being the major organ responsible for LDL clearance. Apoprotein (apo) B-100 is the major protein constituent in LDL and is the ligand recognized by the LDL receptor.

The complete sequence of apo B-100 has been recently deduced from its cDNA sequence. It is one of the largest monomeric proteins known, with a calculated molecular weight of 513 kDaltons (kD). Apo B-100 shares some properties with other smaller apolipoproteins; both apo B-100 and apo E bind to the LDL receptor and interact with a variety of glycosaminoglycans including heparin. Further, like other apolipoproteins, apo B-100 contains sequences that spontaneously associate with lipids. However, apo B-100 is unique among the plasma apolipoproteins in the following respects: it is insoluble in aqueous media after delipidation, its conformation in association with lipids is quite different, LDL apo B-100 has a high β-sheet content, and it does not exchange among lipoprotein particles. Even though the complete primary structure of apo B-100 has been determined, its tertiary structure and conformation on LDL is still poorly understood. We have approached this problem by direct sequence analysis of LDL apo B-100 tryptic peptides, thereby inferring the relative accessibility of different regions of the protein to tryptic digestion and their subsequent release from the lipoprotein particle. In a previous communication, we directly sequenced 52% (2366 residues) of apo B-100 protein using this approach. The present analysis is an extension and refinement of our previous study. We have now obtained direct sequence information on 88% (4021 residues) of LDL apo B-100.
were pooled according to the separation profile (see column at a flow rate of 60 ml/h). Four ml fractions were monitored by ultraviolet absorbance at 220 nm. They was then applied to a Sephadex G-50 (2.6x200 cm) temperature for 24 hours, and the tryptic hydrolysate for 60 minutes and then alkylated with iodoacetic acid 0.1 M ammonium bicarbonate (pH 8.0) were reduced by a 200-fold excess of dithiothreitol (DTT) at room temperature for 3 hours. The reaction mixture according to the standard procedure by ultracentrifugation in KBr solutions. The protein was isolated from single donors according to the standard procedure by ultracentrifugation in KBr solutions. LDL in normal fasting donors. Each LDL preparation was isolated as an ammonium acetate buffer system. (A: 0.005 M KH2PO4/K2HPO4, pH 6.0; B: 10% A and 90% acetonitrile). The fractions from the Vydac column were reconstituted in 200 to 500 μl of Buffer A or a 1:1 mixture of Buffer A and B; the latter solution was used to ensure that peptide mixtures eluting at a high percentage of B could be dissolved and transferred from the sample tube to the column. The purity of the collected fractions was checked by manual sequencing. The pure peptides were completely sequenced and their amino acid compositions were determined by the Pico-Tag method. The fractions that still contained mixed peptides were further chromatographed on the same column using an ammonium acetate buffer system. (A: 0.025 M ammonium acetate, pH 6.5; B: 30% 0.05 M ammonium acetate, pH 6.5, 70% acetonitrile, vol/vol).

Preparation of Low Density Lipoproteins

Human plasma was obtained by plasmapheresis from normal fasting donors. Each LDL preparation was isolated as an ammonium acetate buffer system. (A: 0.005 M KH2PO4/K2HPO4, pH 6.0; B: 10% A and 90% acetonitrile). The fractions from the Vydac column were reconstituted in 200 to 500 μl of Buffer A or a 1:1 mixture of Buffer A and B; the latter solution was used to ensure that peptide mixtures eluting at a high percentage of B could be dissolved and transferred from the sample tube to the column. The purity of the collected fractions was checked by manual sequencing. The pure peptides were completely sequenced and their amino acid compositions were determined by the Pico-Tag method. The fractions that still contained mixed peptides were further chromatographed on the same column using an ammonium acetate buffer system. (A: 0.025 M ammonium acetate, pH 6.5; B: 30% 0.05 M ammonium acetate, pH 6.5, 70% acetonitrile, vol/vol).

Carboxymethylation

Reduction and carboxymethylation were performed according to the method of Anfinsen and Haber. 20 LDL in 0.1 M ammonium bicarbonate (pH 8.0) were reduced by a 200-fold excess of dithiothreitol (DTT) at room temperature for 60 minutes and then alkylated with iodoacetic acid (10-fold excess over DTT) for 3 hours. The reaction mixture was dialyzed against 0.1 M ammonium bicarbonate (pH 8.0) to remove the excess reagent and nonvolatile salt.

Trypsin Digestion and Sephadex G-50 Chromatography

Tryptic digestion of the reduced and alkylated LDL (200 mg LDL in 20 ml of 0.1 M NH4HCO3, pH 8.0) was performed with TPC and treated trypsin (1 mg). The enzyme/substrate ratio was 1:40. The mixture was left at room temperature for 24 hours, and the trypic hydrolysate was then applied to a Sephadex G-50 (2.6x200 cm) column at a flow rate of 60 ml/h. Four ml fractions were monitored by ultraviolet absorbance at 220 nm. They were pooled according to the separation profile (see below) and lyophilized.

Peptide Purification

Apo B-100 tryptic peptides were fractionated on a Waters or Beckman HPLC system equipped with a variable wavelength detector. The column temperature was set at 50°C. Peptides were monitored at 220 nm. For primary fractionation, we used a Vydac C18 reverse-phase column (1.0x25 cm) with a trifluoroacetic acid (TFA) buffer system (Buffer A: 0.1% TFA in water, vol/vol; Buffer B: 0.08% TFA on 95% acetonitrile and 5% water, vol/vol) at a flow rate of 4.0 ml/min. A linear gradient of Buffer B at 1% B per minute was applied. The eluent under each peak was collected manually in glass tubes. The identical peaks from different runs were pooled and dried on a Speedvac (Savant Instruments, NY). For further purification, we used a Hypersil ODS reverse-phase column (4.6x250 mm) with a phosphate buffer system (A: 0.005 M KH2PO4/K2HPO4, pH 6.0; B: 10% A and 90% acetonitrile). The fractions from the Vydac column were reconstituted in 200 to 500 μl of Buffer A or a 1:1 mixture of Buffer A and B; the latter solution was used to ensure that peptide mixtures eluting at a high percentage of B could be dissolved and transferred from the sample tube to the column. The purity of the collected fractions was checked by manual sequencing. The pure peptides were completely sequenced and their amino acid compositions were determined by the Pico-Tag method. The fractions that still contained mixed peptides were further chromatographed on the same column using an ammonium acetate buffer system. (A: 0.025 M ammonium acetate, pH 6.5; B: 30% 0.05 M ammonium acetate, pH 6.5, 70% acetonitrile, vol/vol).

Amino Acid Analysis

Amino acid composition was analyzed with a Waters Pico-Tag amino acid analyzer, using PITC derivatization. The pure peptides were first hydrolyzed in 150 μl of 6 N HCl containing 20 μl solution of 0.125% phenol and incubated for 1 hour at 150°C. The hydrolyzed samples were dried and redried by adding 20 μl of ethanolic solution (ethanol/water/TEA, 1:1:1) to ensure that a trace amount of ammonia was left. For derivatization, the samples were coupled with 20 μl of PITC solution (ethanol/water/TEA/PITC, 7:1:2:1) for 10 minutes, were dried again on a Speedvac, and were reconstituted in sample buffer (0.5 M sodium phosphate buffer, pH 7.4, and 5% acetonitrile) for analysis.

Peptide Sequencing

Sequences of short peptides of less than 20 residues were determined according to the method of Chang et al. using a modified Edman reagent, dimethylaminoazobenzeneisothiocyanate (DABITC) (Fluka, Switzerland). Most of the dimethylaminoazobenzenehydantoin (DABTH) amino acids were identified by thin-layer chromatography (TLC). HPLC methodology was used to identify DABTH- and DABTH-Leu, or residues which contained less than 100 pm of material. A Beckman HPLC system and a Brownlee PTH-C18 column (5 μm, 220x4.6 mm) were used as a modified system for the separation of DABTH-amino acids at 435 nm. The chromatographic conditions were: column temperature, 50°C; flow rate, 1.3 ml/min;
Buffer A: 0.03 M sodium acetate, pH 5.0; Buffer B: acetonitrile with 2% ethylene dihydrochloride; after injection of the sample, mobile phase isoelectric at 37%B for 2 minutes; then linear gradient from 37%B to 59%B in 23 minutes; then isocratic at 59%B for 5 minutes followed by 95%B for 1 minute; maintenance at 95%B isocratic flow for 2 minutes, followed by return of the system to 3%B for 1.5 minutes. The sequences of all the larger peptides were determined with a gas phase automatic sequencer (Applied Biosystems) equipped with a model 120A PTH Analyzer. The PTH amino acid derivatives were identified by HPLC using a Brownlee narrow bore column (2×220 mm).

**Carboxyl-Terminus Determination**

The determination of the carboxyl-terminus of the protein or the peptides was performed by carboxypeptidase-Y digestion and analyzed by Pico-Tag precolumn derivatization.\(^{27}\)

**Determination of N-Glycosylation Sites**

Purified tryptic peptides were sequenced by the manual modified Edman degradation procedure or the automated gas phase sequencer method. Evidence for N-glycosylation was obtained when a step revealed a negative result corresponding to an asparagine residue in the consensus sequence Asn-X-Thr/Ser where X represents any amino acid residue except proline.\(^{28}\) In all cases, amino acid analysis of the peptide confirmed the presence of an asparatic acid residue that was derived from the missing asparagine residue.

**Nucleotide Sequence of Cloned Apoprotein B-100 cDNAs**

All cDNA sequencing was performed by the dideoxy-nucleotide termination technique of Sanger et al.\(^{29}\) using synthetic oligonucleotide primers. For cDNA sequencing, both strands were sequenced in entirety for all apo B-100 cDNA clones.

**Direct Sequencing of HepG2 mRNA**

PolyA RNA was isolated from HepG2 cells by the guanidine isothiocyanate technique,\(^{30}\) followed by two passages over an oligo-dT-cellulose column. Sequencing was performed on 2 μg of the RNA by the technique of Gellebter,\(^{31}\) by using synthetic oligonucleotide primers.

**Statistical Analysis of DNA Sequence Data**

In making pairwise nucleotide sequence comparisons, we used the method of LJ et al.,\(^{32}\) in which nucleotide changes are classified as synonymous (causing no amino acid change) or nonsynonymous and each position in a codon is counted as a synonymous site, a nonsynonymous site, or one-third synonymous and two-thirds nonsynonymous, depending on the consequences of the changes possible at that position. For example, the second position of the codon UUA (Leu) is a nonsynonymous site and the first and third positions are each counted as one-third synonymous and two-thirds nonsynonymous. The method gives the number of (synonymous) changes per synonymous site and the number of (nonsynonymous) changes per nonsynonymous site.

**Results**

**General Strategy of Sequence Analysis**

Primary structure analysis of apo B-100 has been greatly hampered by its high molecular weight (>500 kD)\(^{5-6,17,33}\) and its insolubility in aqueous buffers. These problems have been overcome by use of LDL as the starting material and by use of a newly developed HPLC technique.\(^{21}\) LDL are soluble in aqueous buffers and contain 90% lipid and 20% protein.\(^{37-38}\) Most of the protein is apo B-100. Some regions of apo B-100 on LDL are tightly associated with the lipid and some other regions are more exposed to the aqueous environment. Trypsinization of LDL apo B-100 would allow cleavage into small peptides [trypsin-releasable (TR) peptides] of regions which contain Lys or Arg close to the surface of the LDL particle, while the lipid-associated regions [trypsin-nonreleasable (TN) regions], which are somehow buried or conformationally unfavorable and protected from tryptic digestion, should remain intact. Some peptides that are cleaved by trypsin, but stay associated with the partially digested LDL, will also be grouped with the TN peptides. The TR and TN peptides of apo B-100 are readily separated on a Sephadex G-50 column.\(^{38-39}\) The scheme used for analysis of apo B-100 structure is shown in Figure 1.

The separation profile of the tryptic peptides on a Sephadex G-50 column was presented previously (Figure 2A in Yang et al.\(^{7}\)). Seven peaks, designated T1 to T7, were identified. By definition, the void volume peak (T1) contained the TN peptides still associated with lipids. Peak T2 generally contained a very small amount of the RNA by the technique of Gellebter,\(^{31}\) by using synthetic oligonucleotide primers.
Figure 2. Chromatogram of the trypsin-releasable peptides of apo B-100, T3 obtained by reverse-phase high-performance liquid chromatography trifluoroacetic acid (TFA) buffer system (A: 0.1% TFA in H2O; B: 0.08% TFA in 95% acetonitrile and 5% H2O) was used to separate the peptide mixture, and the gradient was run from 5% B to 55% B in 75 minutes. The flow rate was at 4.5 ml per minute. The column temperature was 50°C.

Figure 3. Rechromatography of peaks 27 (A) and 42 (B) from Figure 2. The separation was performed on a Hypersil ODS 5μ reverse-phase column (4.6x250 mm) with a sodium phosphate buffer system (A: 5 mm NaH2PO4/Na2HPO4 in water; B: 10% A and 90% CH3CN).

Trypsin-Releasable and Trypsin-Nonreleasable Peptides

The same peptides were released by trypsin digestion whether the LDL was reduced and alkylated before or after the enzyme treatment. We identified 4021 amino
acid residues of trypsin peptides of apo B-100, including 1404 amino acid residues of TN peptides, 1544 amino acid residues of TR peptides, and 1073 amino acid residues of mixed peptides (MX), which were present in both fractions. MX peptides probably represent peptides from regions of apo B-100 that were partially digested by trypsin. They may also include peptides that were cleaved, but remained partially associated with the lipid on the Sephadex G-50 column because of their lipid-binding properties, their self-aggregability, or their association with other peptides in the void volume. Such peptides constitute about 24% of the apo B-100 sequence (Table 1).

In all, we have sequenced over 88% of the 4536 amino acid residues in mature apo B-100. The distribution of TN and TR peptides in apo B-100 is nonrandom, as we showed before. We previously proposed that the apo B-100 sequence can be divided into five domains based on the trypsin releasability of each region. Data from the present analysis indicates that the same five hypothetical domains can be identified with slight modification. The amino acid residues of TN, TR, MX, and not found (NF) peptides in each domain are depicted in Figure 4 and listed in Table 1. A schematic representation of apo B-100 on an LDL particle is shown in Figure 5.

**Location of N-Linked Carbohydrates**

Nineteen potential N-glycosylation sites have been identified in the apo B-100 sequence deduced from the cDNA sequence. By direct sequence analysis, 16 sites were found to be glycosylated and three sites were not. Two of the glycosylated sites were on one peptide, the rest were on 14 other peptides. The peptides are designated CHP1 to CHP16, matching the order of N-glycosylated asparagines from the N-terminal to C-terminal region of apo B-100; peptide CHP3+4 contains both the third and fourth N-glycosylated asparagines. The carbohydrate moieties were linked to asparagine residues at the following 16 positions: 158, 956, 1341, 1350, 1496, 2752, 2955, 3074, 3197, 3309, 3331, 3364, 3438, 3868, 4210, and 4404, respectively. The length of the glycopeptides ranges from 6 to 42 residues. Their respective masses were increased by the attached carbohydrate, and, on the Sephadex column, they were eluted inside fractions T1, T3, and T4 which generally contained other nonglycosylated peptides of higher molecular mass. For example, CHP9 contained six amino acid residues and eluted from fraction T3 which generally contains peptides of 20 to 30 residues. Similarly, on the Vydac C18 column, this hexapeptide also eluted early (8% buffer B of the TFA system) because of its increased hydrophilicity due to the attached carbohydrate. On the Vydac C18 column, the amount of buffer B needed in the phosphate buffer system to elute the N-glycosylated peptides is related to various properties (e.g., size, charge) of the peptides. On this column, carbohydrate also increases the hydrophilicity of the peptides and allows the glycosylated peptide to be eluted early in this system.

**Nucleotide Sequence of Apoprotein B-100 cDNA and mRNA**

We previously presented the complete sequence of overlapping cloned apo B-100 cDNAs. We have now completely sequenced all the independent cDNA clones that we identified in two cDNA libraries. Twenty-seven of the clones were derived from normal liver cDNAs, and four were from HepG2 cell cDNAs. Furthermore, we directly sequenced a human hepatoma cell line (HepG2) mRNA in the region corresponding to the COOH-terminal of 500 residues of apo B-100. Our results indicate that there is sequence heterogeneity among apo B-100 mRNA sequences from different sources. The comparison of our apo B-100 mRNA and cDNA sequences with other published sequences is presented in the Discussion section.

**Discussion**

Apo B-100 is a most difficult protein for structural analysis because of its huge size and its insolubility in aqueous buffer after delipidation. Recently, our laboratory presented a preliminary mapping of LDL-apo B-100 tryptic peptides. The basic approach was the painstaking separation of the LDL apo B peptides on successive HPLC columns. We found that there are two types of tryptic peptides, those that are readily released from LDL particles and those that are released only upon delipidation and retrypsinization. Operationally, in this paper, we have labeled the first group of peptides trypsin-releasable, and the second group, trypsin-nonreleasable. Therefore, the term "trypsin-releasable" refers to peptides that were labeled "trypsin-accessible," and "trypsin non-releasable" to those labeled "trypsin-inaccessible" in our previous study. We have made the change because the old terminology was easily mistaken as mechanistic, whereas the new one is purely descriptive and operational in nature. The present study presents the complete sequence of essentially all the trypptic peptides of apo B-100 obtained.
**Figure 4.** Distribution of trypsin-releasable versus trypsin-nonreleasable peptides in apo B-100. This is an updated, refined version of Figure 1A in our previous communication. It contains information on 1655 additional residues not previously analyzed. The red peptides are those that stayed associated with the lipoprotein particle during the initial tryptic digestion. They were released from the first peak (T1) on the Sephadex G-50 column (Figure 2A in Yang et al.). The blue peptides are the trypsin-releasable peptides (T3→T7, Figure 2A in Yang et al.). The purple peptides are those found in both peaks T1 and T3→T7. The black peptides are those that were not found by direct sequencing but predicted from cDNA sequences. The orange Cs are the Cys residues. The green Ns are Asn residues that are actually glycosylated by direct sequence analysis. For a key to the single-letter codes, see the legend to Table 2.
Figure 5. Schematic representation of apo B-100 structure on low density lipoprotein (LDL). This is a revised and updated version of Figure 2D in Yang et al.7 For simplicity, the trypsin-releasable regions are shown on the outside, and the trypsin-nonreleasable regions, inside the core of the LDL particle. Relative lengths of peptides are not drawn to scale, and the surface-core orientation is presented to allow easy visualization of trypsin releasability. It does not imply that the different parts of apo B-100 are physically on the surface or buried inside the lipoprotein particle. The five hypothetical domains are separated by dashed lines. The thrombin cleavage sites at residues 1297 and 3249 are marked. The locations of the 16 identified N-glycosylated carbohydrates are indicated by (O), cysteine residues are indicated by (•), and disulfide bridges are indicated by (=).

by such an approach. Since we are sequencing many peptides repeatedly, we believe that additional sequencing will no longer be fruitful. About two-thirds of the peptides can be assigned to exclusively TR or TN regions. However, a significant proportion (23.7%, labeled as mixed in Table 1) was found to be present in both the TR and the TN group of peptides.

The method for probing LDL apo B structure by limited proteolysis has been used by various investigators in the past.36-43 Since most previous studies using such an approach did not present any sequence information, comparison with the present analysis is difficult. The general conclusions from such studies are: 1) peptides that are released from LDL apo B by limited proteolysis are different, as a whole, from those that are not released;36,38,40 2) there are repeated sequences in apo B-100;36,41,42 and 3) some monoclonal antibody epitopes are destroyed, whereas others are unaltered by limited proteolysis.42 The observations in this study bear directly on and support the first conclusion. In separate studies, we also have evidence supporting the second5,7,14 and the last44 conclusions. In one previous report where LDL apo B-100 peptides released by limited trypsinization were sequenced and localized to eight regions of apo B-100,43 seven regions were found to match sequences that are TR or MX in our study; one (the last peptide in Table 1 of the study by Forgez et al.43) that was not found in the present study was juxtaposed between two exclusively TR peptides by our analysis. Nonidentical experimental conditions notwithstanding, the consistency of the results from two different laboratories indicates that different regions of apo B do assume specific conformations on LDL.

Trypsin-Releasable and Trypsin-Nonreleasable Domains of LDL Apoprotein B-100

In our preliminary analysis published previously,7 the LDL apo B-100 sequence could be divided into five broad domains based on the susceptibility of the region to trypsin digestion and release. Here, after analyzing over 88% of apo B-100 peptides, we find that the pattern of distribution of TR versus TN peptides is similar to that in the earlier analysis, and the same five domains can be identified with only minor modification (Table 1, and Figure 6): domain 1, an NH2-terminal region (residues 1 to 1000) that consists predominantly of TR peptides; domain 2, a region next to it from residues 1001 to 1700 that appears to be transitional between 1 and 3 with alternating TR and TN peptides, as well as a significant proportion of MX peptides; domain 3, a largely TN region spanning residues 1701 to 3070, with occasional TR peptides interspersed (TR peptides are noted especially close to the NH2- and COOH-terminal boundaries of this region); domain 4, a domain from residues 3071 to 4100 of mainly TR and MX peptides; and domain 5, a COOH-terminal domain from residue 4100 to the COOH-terminus of apo B-100 that consists almost exclusively of TN peptides.

The average hydrophobicities of the first four domains (−0.901, −0.861, −0.927, and −0.908 kcal per residue, respectively) differ only slightly from the average value of the entire apo B-100 sequence (−0.916 kcal per residue). The COOH-terminal domain is significantly more hydrophobic (−1.020 kcal per residue) than the rest. Its hydrophobic nature may be a factor that renders the region relatively TN upon trypsin digestion.

We examined the distribution of hydrophobic and polar, as well as "indifferent," amino acid residues in each of the five domains (Figure 7). There seems to be an inverse correlation between the abundance of hydrophobic residues and trypsin releasability. For example, domains 5 and 3 are, in the main, relatively TN; domain 5 especially and domain 3, to a lesser extent, seem to contain fairly high proportions of hydrophobic residues (compare especially panel T to panels D, H, and P). Similarly, within each domain, peptides that are exclusively TN seem to be overrepresented in hydrophobic residues (e.g., compare panels B to A, J to I, R to Q in Figure 7). However, exceptions do exist (e.g., compare panel F to E; comparison of panels N and M may be misleading since few peptides in domain 4 are TN). Thus, a careful analysis of the relative abundance of hydrophobic versus polar residues indicates that the charge characteristics of the amino acids present in each domain do not entirely account for the relative releasability of the region after trypsin cleavage. The sensitivity of a protein to trypsin digestion is poorly understood. Apart from the presence of arginyl and lysyl residues, it also depends on the conformation of the region and on the charge environment of the neighboring residues.45 While the factors that contribute to the trypsin releasability of LDL apo B are incompletely understood, the fact that the same apo B-100 peptides are reproducibly released when LDL are exposed to trypsin indicates...
that the protein maintains a unique conformation on LDL, and the pattern of trypsin releasability provides a useful map for comparison with results based on other techniques, such as monoclonal antibody binding, used for probing LDL apo B-100 structure.

**Disulfide Bridges and N-Linked Glycosylation**

The distribution of the 25 cysteine residues in apo B-100 is highly asymmetric (Figures 5 and 6). Fourteen of the 25 are found in the N-terminal domain 1. We also found that at least six of these residues are linked to their immediate neighbor Cys residues, i.e., Cys218 to Cys219, Cys544 to Cys545, and Cys556 to Cys557. Such a concentration of disulfide bridges in domain 1 indicates that this part of apo B-100 is globular in nature and may contribute to the relative trypsin-accessibility of this region. Knott et al. mapped another disulfide peptide in domain 4, which also consists of neighboring Cys residues, namely, Cys317 and Cys327. We are currently attempting to map the rest of the disulfide bridges, which will be the subject of a future communication.

In a previous paper, we indicated that there were 20 potential N-glycosylation sites. However, on reexamining the sites, we have excluded Asn-429 as a potential site because it is followed by a Pro, which is not allowed according to the Marshall rule. We have directly sequenced all the 19 potential N-glycosylation sites in apo B-100 (Figures 4, 5, and 6). Three of these sites were nonglycosylated. Of the 16 glycosylated peptides, seven are found in domain 4. The latter domain, which consists mainly of TR and MX peptides, appears to include a multifunctional region of apo B-100. It encompasses the
Figure 7. Distribution of amino acid residues of different charge properties in apo B-100. The five domains are those defined by trypsin releasability as detailed in Figures 5 and 6. TR = trypsin-releasable, TN = trypsin-nonreleasable; MX = mixed. The hydrophobic residues include Met, Val, Leu, Ile, Phe, Tyr, and Trp; acidic residues, Asp and Glu; basic residues, Arg and Lys; indifferent residues, Gly, Ala, Ser, Thr, Asn, Gin, His, and Cys. The exact numbers of individual residues are computed from Figure 4.

Comparison of Trypsin Releasability to Other Parameters of LDL Apoprotein B Conformation

Most physical measurements of LDL apo B conformation give a picture of the "average" structure of the entire polypeptide chain. For example, circular dichroism measurements indicate that LDL apo B contains a relatively high $\beta$-sheet content of 21%. However, this is only an average value, and it is not clear from such physical measurements which sequences within the entire apo B-100 sequence assume $\beta$-sheet conformation on LDL. There are only a limited number of experimental approaches that can probe LDL apo B conformation at the primary sequence level. In addition to the susceptibility of different regions of apo B-100 to proteolytic digestion and release, an approach taken in the present study, another approach taken by many different laboratories is the mapping of immunoreactivity of apo B epitopes with monoclonal antibodies. When denatured apo B-100 is used as an immunogen, the resulting monoclonal antibodies show differential reactivity to native LDL apo B. The difference in immunoreactivity between denatured and native LDL apo B is significant.

Figure 8. Variations in DNA-derived and peptide-derived amino acid sequences from different laboratories. The only peptide-derived amino acid sequence is that determined in the current study and is labeled AA seq. The data of Chen et al. has been updated by completion of the complete sequence analysis of all 31 cDNA clones as well as mRNA sequence of the 3' terminal region covering the COOH-terminal 500 residues.
### HUMAN APOLIPOPROTEIN B-100 STRUCTURE

**Yang et al.**

<table>
<thead>
<tr>
<th>Codon</th>
<th>AA seq</th>
<th>Nucleotide and deduced amino acid sequences</th>
</tr>
</thead>
</table>

- **1** to **4**
- **5** to **8**
- **9** to **12**
- **13** to **16**
- **17** to **20**
- **21** to **24**
- **25** to **28**
- **29** to **32**
- **33** to **36**
- **37** to **40**
- **41** to **44**
- **45** to **48**
- **49** to **52**

**X:** Unknown nucleotide or amino acid.

- **# codon at which non synonymous differences are observed.**
- **$+$ codon at which a variant has been observed in more than one laboratory.
apo B-100 and native LDL apo B can be explained by at least two factors: 1) the epitope which might be "exposed" in denatured apo B-100 may be "buried" in LDL apo B, and 2) the monoclonal antibody may be conformation-dependent and may not recognize native apo B. It will be interesting to compare trypsin releasability of specific sequences in apo B-100 to their monoclonal antibody accessibility. Unfortunately, in most instances, this is not possible at present, because the epitopes for specific antibodies have been mapped to extremely broad regions of apo B-100.44,47,48 In a recent study, we mapped the epitopes for several monoclonal antibodies to relatively short sequences.44 A comparison of trypsin releasability to epitope accessibility among these antibodies is presented in Table 2. In this table, LDL accessibility to each monoclonal antibody is defined as the amount of LDL that can be precipitated by the antibody in the presence of Staphylococcus aureus protein A. Monoclonal antibodies 7 and 22 recognize residues 4517 to 4536 on apo B-100. In this region, residues 4517 to 4526 are TN, and 4527 to 4536 are TR (see Figure 4) by trypsin digestion. The amount of LDL that is immunoprecipitable by these antibodies increases from ~40% of fresh LDL to >80% LDL kept in storage for a month.44 The epitope for monoclonal antibody 16 is generally not releasable by trypsin treatment even though it is quite accessible to the antibody. Conversely, even though the epitope for monoclonal antibody 20 encompasses largely releasable peptides, it is not accessible to the antibody in its native conformation. We note that in this case, adsorption of native LDL to a solid phase induces sufficient conformational change in this epitope to allow its recognition by the antibody.44,45

In summary, in the three short sequences inside the COOH-terminal region of apo B-100 that we studied, there is little correlation between the two methods. This is not entirely unexpected, for the reasons given above. Hahm et al.42 have found that limited proteolysis of LDL apo B destroys some epitopes, sparing others. The relative lack of correlation between the two methods underscores the importance of studying LDL apo B conformation by multiple simultaneous approaches and of being cautious in the interpretation of experimental data.

**Variations among Apoprotein B-100 Sequences from Different Laboratories**

By direct peptide sequencing, we have identified 4021 of the 4536 amino acid residues for the apo B-100 sequence. We have compared our data with one genomic sequence, four complete cDNA sequences, three partial cDNA sequences, and a partial mRNA sequence determined in the present study (Figure 8). Among all the amino acids we identified by direct peptide sequencing, only one did not match any of the published sequences, i.e., at residue 725, we observed D (aspartic acid), whereas it is N (asparagine) in all the published sequences. This difference can be explained by post-translational modification of the protein sequence by deamidation.

The compilation shown in Figure 8 allows us to evaluate the extent of DNA and protein sequence variation that exists among the various apo B-100 alleles in the human population. Our compilation reveals 75 differences among all the sequences compared: 1 three-codon deletion, 53 nonsynonymous changes, and 21 synonymous changes. This extends the compilation by Ludwig et al.49 who found 61 differences among the DNA sequences available then. As noted by these researchers, in some cases, the published sequence represented isolates from more than one cDNA or genomic library and therefore does not represent the haplotype of a given allele.

Some of the differences shown in Figure 8 appear to represent bona fide variations. In our direct peptide sequencing work, we observed both V and A at residue 3849, while the amino acid deduced from our cDNA sequence is A, and that deduced from the other four published DNA sequences is V. In our cDNA sequencing work,6 we observed in some clones the same three-codon deletion in the signal peptide region (codons -16 to -14) as that observed by Cladaras et al.,9 and we observed both ATT (I) and CTT (L) at codon 618 (refers to the codon number in the mature peptide region) and both CCT (P) and GCT (A) at codon 892. Some of the sequence variations have also been detected as restriction fragment polymorphisms in population studies. For example, the variation at codon 2488 gives rise to an Xba I polymorphism50 and at codon 4154, an AAA codon to a GAA codon change results in an EcoR I site polymorphism.51

We also note from Figure 8 that 12 of the sequence differences have been reported by more than one laboratory; two (codons 2488 and 4145) have been noted above, and the others are codons 71, 591, 2298, 3400, 3405, 3705, 3849, 3922, and 3937. As the chance of making the same sequencing error at a nucleotide position in two or more laboratories is small, these sequence differences probably represent bona fide variations.

In cases other than those mentioned above, it is difficult to distinguish sequencing errors from bona fide variations. However, in some cases, there is circumstantial evidence that suggests sequencing errors. At codon 766, Law et al.9 observed CGA, whereas all other authors observed CAG; at codon 1391, Cladaras et al.9 observed TCT, whereas all others observed TTC; at codon 2298, Hardman et al.50 observed TAT, whereas all others observed ATT; at codon 2906, Law et al.9 observed TCG, whereas all others observed TGC; at codon 3264, Carlson et al.51 observed CTG, whereas all others observed CGT; at codon 3404, Law et al.9 observed CCG, whereas all others observed GCC; and at codon 3797, Cladaras et al.9 observed CGA, whereas all others observed CAG. In each of these cases, the differences involve the rearrangement of two or three nucleotides, a situation that is unlikely to have arisen by point mutation. In our cDNA sequencing work,6 we observed GAC, CAT, GAA, GAG, GGA, and GGG at codons 4083, 4084, 4095, 4101, and 4109, respectively, whereas, in our mRNA sequencing work (present study), we observed AAC, AAU,CAA,UGU, GCA, and GCC at these codon positions. As the latter match the sequences from other laboratories, the former are probably not bona fide variations.

Table 3 shows pairwise comparisons of the nucleotide sequences cited in Figure 8. In each comparison, the number of changes per nonsynonymous site and the number of changes per synonymous site are given above.
and below the diagonal, respectively (see Methods for the definition of these quantities). The sequence by Ludwig et al. is, on the average, most similar to the sequence by Chen et al. is most similar to the sequence by Ludwig et al.

Acknowledgments

We thank Beate Beuermann, Billy Touchstone, Mickias Taffere, and Hui-Xih Yang for excellent technical assistance, Susan Kelly for the art work, and Suzanne Mascola for typing the manuscript.

References

1. Gordon T, Kannel WB. Predisposition to atherosclerosis in the head, heart, and legs. The Framingham Study. JAMA 1972;221:661-666
8. Law SW, Grant SM, Hilguchi K, et al. Human liver apolipoprotein B-100 cDNA: complete nucleic acid and derived amino acid sequence. Proc Natl Acad Sci USA 1986;83:8142-8146
31. Gelebter J. Dideoxynucleotide sequencing of RNA and uncloned cDNA. BRL Focus 1987;5:17
33. Elovson J, Jacobs JC, Schumaker VN, Pupponle DL. Molecular weights of apoprotein B obtained from human low density lipoprotein (apolipoprotein B-100) and from rat very low density lipoprotein (apolipoprotein B-PII). Biochemistry 1985;24:1569-1578

Index Terms: apolipoprotein B • peptide sequencing • low density lipoproteins • cDNA cloning • trypsin releasability • messenger RNA sequencing • polymorphism • N-glycosylation • protein conformation • internal repeats
Structure of apolipoprotein B-100 of human low density lipoproteins.
C Y Yang, Z W Gu, S A Weng, T W Kim, S H Chen, H J Pownall, P M Sharp, S W Liu, W H Li
and A M Gotto, Jr

doi: 10.1161/01.ATV.9.1.96

The online version of this article, along with updated information and services, is located on the
World Wide Web at:
http://atvb.ahajournals.org/content/9/1/96