Proc. Nat. Acad. Sci. USA
Vol. 71, No. 9, pp. 3631–3634, September 1974

The Primary Structure of High Density Apolipoprotein-Glutamine-I

(primary amino-acid sequence/lipid–protein interaction/α-helical structure)

H. NORDEAN BAKER, THOMAS DELAHUNTY, ANTONIO M. GOTTO, JR., AND RICHARD L. JACKSON

Division of Atherosclerosis and Lipoprotein Research, Department of Medicine, Baylor College of Medicine and The Methodist Hospital, Houston, Texas 77025

Communicated by Donald S. Fredrickson, July 8, 1974

ABSTRACT **The major protein constituent of human plasma high density lipoproteins has been isolated and its complete amino-acid sequence determined. The protein, designated apolipoprotein-glutamine-I by the presence of carboxyl-terminal glutamine, is a single polypeptide chain of 245 amino-acid residues, including three residues of methionine. The protein is devoid of cysteine, cystine, and isoleucine. Cleavage of apolipoprotein-glutamine-I with cyanogen bromide yields four fragments with 94, 90, 36, and 25 amino acids. The amino-acid sequence of each fragment was determined by conventional methods, with proteolytic digestion with trypsin, chymotrypsin, and thermolysin. The alignment of the cyanogen bromide fragments was determined by the isolation of the methionine-containing tryptic peptides from apolipoprotein-glutamine-I.**

Inspection of the sequence of apolipoprotein-glutamine-I suggests an interesting distribution of amino acids that may account for its helical structure and its ability to bind and transport lipid.

Human plasma high density lipoproteins (HDL) have been shown to contain two major apoprotein components that account for 90% of the total protein of the family (1, 2). These two proteins, designated apolipoprotein-glutamine-I (apoLP-Gln-I) and apolipoprotein-glutamine-II (apoLP-Gln-II), have been extensively characterized in several laboratories, and their physiochemical and lipid-binding properties have been recently reviewed (3–6). The amino-acid sequence of apoLP-Gln-II is known (7–9). It contains two identical monomeric units of 77 amino acids each; the units are linked by a single disulfide bond.

In the present communication, we describe the complete amino-acid sequence of the other major HDL apoprotein, apoLP-Gln-I. The protein contains 245 amino acids with a calculated molecular weight of 28,331. This new sequence information, coupled with that of apoLP-Gln-II and of apoLP-Ala (10) and apoLP-Ser (11, 12) from the very low density lipoproteins (VLDL), permits speculation as to the structural features of the plasma apoproteins that might account for their ability to bind and transport lipid. A theory to explain this ability has been recently presented (13).

MATERIALS AND METHODS

Isolation of ApoLP-Gln-I. Human plasma HDL was isolated by ultracentrifugal flotation between densities 1.063 and 1.210 g/ml, as described (14). Lipid-free HDL (apoHDL) was prepared by delipidation of the lipoprotein with diethyl ether–ethanol (3:1). ApoLP-Gln-I was isolated from apoHDL by a combination of gel filtration and DEAE–cellulose chromatography, as described (15).

Sequence Strategy. The complete amino-acid sequence of apoLP-Gln-I was established from the four cyanogen bromide fragments. The sequences were determined by both direct identification of the phenylthiohydantoins from Edman degradations (16) and by quantitative subtractive Edman degradations (17) of smaller peptides obtained from proteolytic digestion. The alignment of the fragments in the intact polypeptide chain was determined by the isolation of the overlap or methionine-containing peptides from a tryptic digest of maleylated apoLP-Gln-I.

Cyanogen Bromide Cleavage. ApoLP-Gln-I was treated with cyanogen bromide, and the four fragments were isolated by chromatography of the digest on Bio-Gel P-30 (BioRad Laboratories) in 25% formic acid (15). The sequences of these peptides were determined by studying the products of enzymatic digestion of the fragments with trypsin, chymotrypsin, and thermolysin. The sequence of each proteolytic peptide was determined by the subtractive Edman procedure (17). Amides were determined by amino-acid analysis after enzymatic digestion of the appropriate peptide with aminopeptidase M (Rohm and Haas) and/or carboxypeptidases A and B (Worthington). Residues 102 and 123 were identified as glutamine and glutamic acid, respectively, by thin-layer chromatography of the phenylthiohydantoin derivatives from T-15 and T-19 (18). As a check on the alignment of the tryptic peptides, the sequence of each fragment was determined automatically with a Beckman Sequencer, model 890B. The phenylthiohydantoin amino acids were determined by gas chromatography.

Isolation of Overlap Peptides. ApoLP-Gln-I was treated with [^{14}C]maleic anhydride (Schwarz/Mann) by the procedure of Butler and Hartley (19). From the specific activity of the isolated maleylated apoLP-Gln-I, it was determined that 97% of the predicted residues were acylated. The maleylated protein was then treated with trypsin, and the tryptic peptides were fractionated by gel-filtration chromatography. The methionine-containing peptides were pooled, demaleylated (19), and treated with trypsin a second time. The peptides were then purified by ion-exchange chromatography.

RESULTS

Characterization of the Cyanogen Bromide Peptides. The initial step in determining the amino-acid sequence of apoLP-

Abbreviations: VLDL, very low density lipoproteins; HDL, high density lipoproteins; apoLP-Gln-I (A-I) and apoLP-Gln-II (A-II), the two major apoproteins of human HDL, each with carboxyl-terminal glutamine; and apoLP-Ala (C-III) and apoLP-Ser (C-I), two apoproteins from human VLDL, with carboxyl-terminal alanine and serine, respectively.

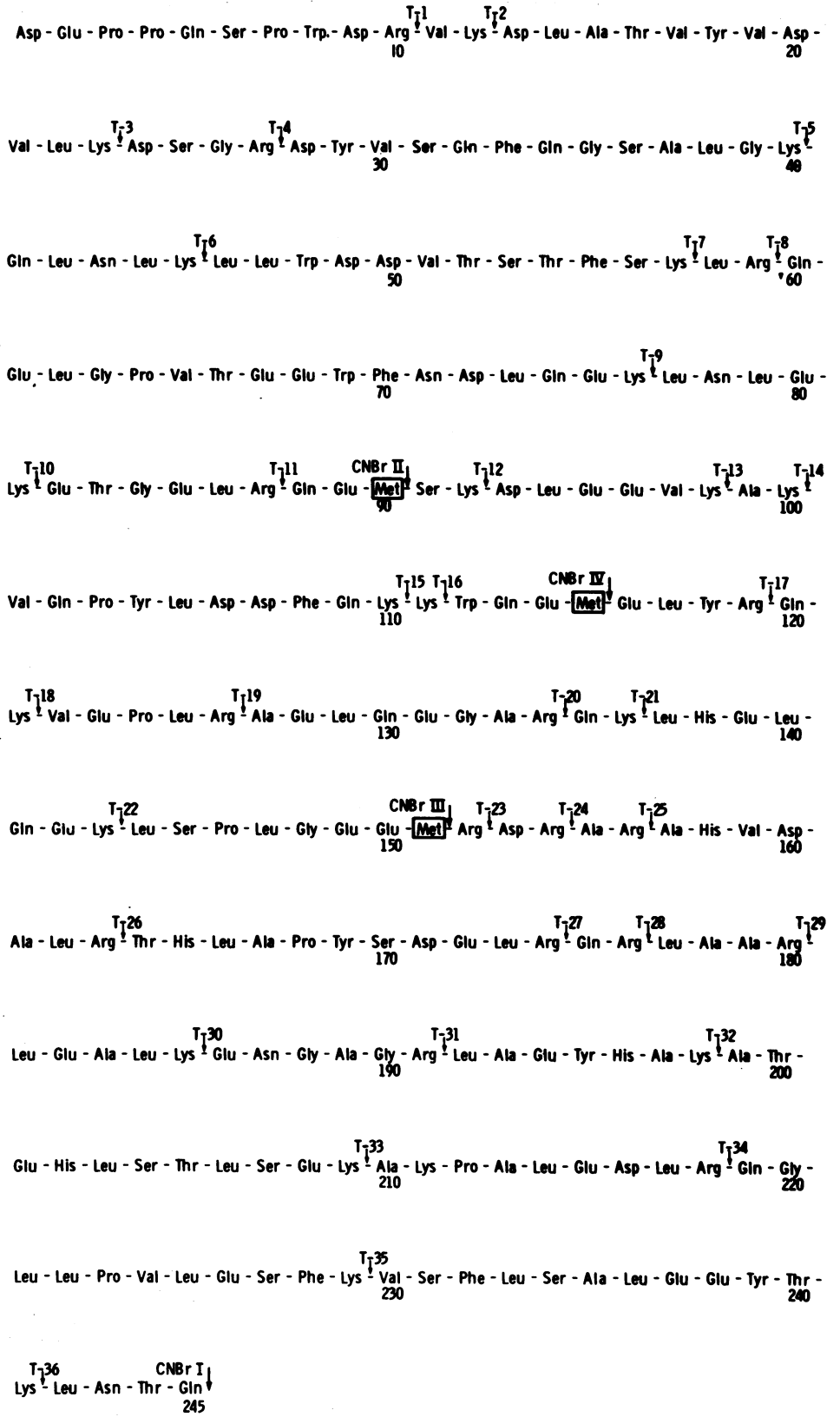

FIG. 1. The amino-acid sequence of human plasma, high density apolipoprotein-glutamine-I. Peptides obtained on cyanogen bromide cleavage (CNBr) and on tryptic cleavage (T) are shown.

Gln-I was cleavage of the intact polypeptide with cyanogen bromide. Since the protein contains three residues of methionine, four fragments were expected. The properties of these fragments have been described (15). Their amino-acid compositions are shown in Table 1. The sequences of the fragments, designated CNBr I, II, III, and IV and containing

TABLE 1. *Amino-acid composition of the cyanogen bromide fragments of apoLP-Gln-I**

Amino acid	CNBr I	CNBr II	CNBr III	CNBr IV	Total
Tryptophan	— —	2.6 (3)	— —	0.8 (1)	3.5 (4)
Lysine	6.1 (6)	6.0 (7)	2.9 (3)	4.3 (5)	19.1 (21)
Histidine	3.3 (4)	— —	0.9 (1)	— —	4.7 (5)
Arginine	8.0 (9)	3.9 (4)	3.0 (3)	— —	14.9 (16)
Aspartic acid	5.8 (6)	10.3 (12)	— —	3.1 (3)	20.3 (21)
Threonine	4.4 (5)	4.3 (5)	— —	— —	9.8 (10)
Serine	5.4 (6)	6.1 (6)	1.0 (1)	1.0 (1)	14.1 (14)
Glutamic acid	14.9 (13)	17.2 (16)	12.2 (12)	7.1 (6)	48.4 (47)
Proline	2.8 (3)	3.9 (4)	1.8 (2)	1.1 (1)	9.2 (10)
Glycine	3.0 (3)	4.6 (5)	2.0 (2)	— —	9.6 (10)
Alanine	11.9 (14)	2.7 (2)	2.0 (2)	1.0 (1)	18.3 (19)
Valine	2.7 (3)	6.1 (7)	1.0 (1)	1.8 (2)	12.3 (13)
Methionine	— —	0.4 (1)†	0.5 (1)†	0.6 (1)†	2.8 (3)
Leucine	15.7 (17)	12.1 (13)	7.2 (7)	2.2 (2)	38.1 (39)
Tyrosine	2.6 (3)	2.0 (2)	0.8 (1)	0.9 (1)	6.4 (7)
Phenylalanine	1.7 (2)	2.7 (3)	— —	0.9 (1)	6.0 (6)
Total residues	94	90	36	25	245

* The two columns of numbers represent the values obtained from amino-acid analysis and from the sequence (in parentheses).
—, not detected.
† As homoserine.

94, 90, 36, and 25 residues, respectively, were determined by conventional methods (manuscript submitted). The sequence of each fragment is shown in Fig. 1. All of the tryptic peptides from the cyanogen bromide fragments were soluble between pH 3 and 9, with the exception of T-7. This peptide precipitated during the digestion of CNBr II with trypsin. Peptide T-7 was removed by centrifugation and was solubilized in 0.2 M pyridine, pH 9.5. The amino-acid sequence of CNBr II was further complicated by the presence of three tryptic peptides (T-6, T-9, and T-12) that contained NH_2-terminal glutamine. These peptides had blocked NH_2-terminal residues and, thus, their sequences could not be determined by the Edman procedure. These "blocked" peptides were treated with thermolysin and/or chymotrypsin, and the smaller fragments were isolated and their sequences were determined. The amino-acid sequence of CNBr III was complicated by the presence of a dipeptide occurring twice in the sequence (Gln-Lys: T-18 and T-21); the dipeptide was isolated in excess of 100%. The identification of Gln-Lys at T-18 was obtained from the sequence of a chymotryptic peptide overlapping T-17, T-18, and T-19 (residues 119–125). The identification of the second dipeptide of Gln-Lys was determined from the isolation of a tryptic peptide from maleylated CNBr III. The peptide (corresponding to residues 135–151) was isolated by gel chromatography and contained a blocked amino-terminus. This peptide was demaleylated and treated a second time with trypsin; peptide T-21 was isolated from the digest. The only problem encountered in the sequence of CNBr I was the alignment of tryptic peptide T-28 (Gln-Arg). This problem was clarified by the isolation of the chymotryptic peptide from CNBr I that contained a sequence of Arg-Gln-Arg-Leu (residues 174–177).

Ordering of the Cyanogen Bromide Peptides. CNBr II had NH_2-terminal aspartic acid, and, since apoLP-Gln-I also has NH_2-terminal aspartic acid, this established CNBr II as the NH_2-terminal peptide of apoLP-Gln-I. Fragment CNBr I was devoid of homoserine and had COOH-terminal glutamine and, thus, corresponded to the COOH-terminal peptide of apoLP-Gln-I. To establish the arrangement of CNBr III and CNBr IV, we isolated two of the three overlap or methionine-containing peptides from a tryptic digest of demaleylated tryptic peptides obtained initially from maleylated apoLP-Gln-I. These overlap peptides [T-12 (residues 88–92) and T-17 (residues 112–119)] yielded the necessary information to order the cyanogen bromide fragments as II–IV–III–I (manuscript submitted).

DISCUSSION

The complete amino-acid sequence of human plasma high density apolipoprotein-glutamine-I is presented in Fig. 1. The calculated molecular weight for the 245 amino-acid residues is 28,331, which is slightly higher than that determined by equilibrium ultracentrifugation (2).

The sequence of apoLP-Gln-I contains several interesting features. All the residues of tryptophan, 15 of 21 aspartic acid residues, and 9 of 13 valine residues are located in the NH_2-terminal half of the sequence. In contrast, all residues of histidine, 13 of 16 alanine residues, and 24 of 39 leucine residues are in the COOH-terminal half. There is also a total absence of threonine in the middle of apoLP-Gln-I. No threonine occurs between residue position 83 in CNBr II and residue position 164 in CNBr I, which covers one-third of the entire length of the sequence. The absence of serine from this same region is also evident.

The amino acid sequence of apoLP-Gln-I has several structural features that may account for its high helical content and lipid-binding properties (4). The protein contains an abundance of amino acids known to be helical-forming residues (Glu, Ala, and Leu). In addition, there are several regions that contain presumed amphipathic helices, as has been suggested for apoLP-Gln-II, apoLP-Ala, and apoLP-Ser (13). These amphipathic helices are characterized as having two faces: a polar face and an apolar or hydrophobic face. For example, four such regions occur in segments 125–145, 147–167, 169–189, and 191–211. We have constructed space-filling models of these four regions and have placed them into an α-helical structure. Inspection of these regions shows the same interesting distribution of the charged residues on the polar face as previously noted (13). The negatively charged residues,

glutamic acid and aspartic acid, invariably occur in a narrow strip along the center of the polar face, while the positively charged residues, lysine and arginine, are located on the lateral edges. This arrangement of basic and acidic residues permits close orientation of the polar head group of phospholipid with the ion pairs of the protein (13). The greatest number of topographically close ion-pairs occurs between residue positions 74 and 211. This observation suggests that a major portion of the phospholipid-binding properties of the apoprotein may reside in this area. The other areas of the molecule may be important in protein–protein interaction.

We are indebted to Miss Esther Chou, Mr. Paul Kizer, Mrs. Vicki Liaw, Mrs. Alice Lin, Mr. Simon Mao, Mr. Kent McGregor, Mrs. Judy Mumford, and Mr. Albert Sanborn for their technical assistance and to Drs. Joel Morrisett, Henry Pownall, Louis Smith, James Sparrow, and David Taunton for their recommendations and comments on this work. We are indebted to Miss Debbie Mason for the preparation of the manuscript. This work was supported in part by Health, Education, and Welfare Research Grant HL 14194, by a grant from the John A. Hartford Foundation, Inc., and by the American Heart Association, Texas Affiliate. R.L.J. is an Established Investigator of the American Heart Association.

1. Shore, B. & Shore, V. (1969) *Biochemistry* **8**, 4510–4516.
2. Scanu, A., Toth, J., Edelstein, C., Koga, S. & Stiller, E. (1969) *Biochemistry* **8**, 3309–3316.
3. Scanu, A. M. & Wisdom, C. (1972) *Annu. Rev. Biochem.* **41**, 703–730.
4. Scanu, A. M. (1972) *Biochim. Biophys. Acta* **265**, 471–508.
5. Shore, V. & Shore, B. (1972) in *Blood Lipids and Lipoproteins: Quantitation, Composition and Metabolism*, ed. Nelson, G. J. (Wiley Interscience, New York), pp. 789–824.
6. Fredrickson, D. S., Lux, S. E. & Herbert, P. N. (1972) *Advan. Exp. Biol. Med.* **26**, 25–56.
7. Brewer, H. B., Lux, S. E., Ronan, R. & John, K. M. (1972) *Proc. Nat. Acad. Sci. USA* **69**, 1304–1308.
8. Lux, S. E., John, K. M. & Brewer, H. B. (1972) *J. Biol. Chem.* **247**, 7510–7518.
9. Lux, S. E., John, K. M., Ronan, R. & Brewer, H. B. (1972) *J. Biol. Chem.* **247**, 7519–7527.
10. Brewer, H. B., Shulman, R., Herbert, P., Ronan, R. & Wehrly, K. (1972) *Advan. Exp. Biol. Med.* **26**, 280.
11. Shulman, R., Herbert, P., Wehrly, K., Chesebro, B., Levy, R. I. & Fredrickson, D. S. (1972) *Circulation* **45**, II-246.
12. Jackson, R. L., Sparrow, J. T., Baker, H. N., Morrisett, J. D., Taunton, O. D. & Gotto, A. M. (1974) *J. Biol. Chem.*, in press.
13. Segrest, J. P., Jackson, R. L., Morrisett, J. D. & Gotto, A. M. (1974) *FEBS Lett.* **38**, 247–253.
14. Jackson, R. L. & Gotto, A. M. (1972) *Biochim. Biophys. Acta* **285**, 36–47.
15. Baker, H. N., Jackson, R. L. & Gotto, A. M. (1973) *Biochemistry* **12**, 3866–3871.
16. Pisano, J. J., Bronzert, T. & Brewer, H. B. (1972) *Anal. Biochem.* **45**, 43–59.
17. Konigsberg, W. (1972) in *Methods in Enzymology*, eds. Hirs, C. H. W. & Timasheff, S. N. (Academic Press, New York), Vol. 25, pp. 326–332.
18. Inagami, T. & Murakami, K. (1972) *Anal. Biochem.* **47**, 501–504.
19. Butler, P. J. G. & Hartley, B. S. (1972) in *Methods in Enzymology*, eds. Hirs, C. H. W. & Timasheff, S. N. (Academic Press, New York), Vol. 25, pp. 191–199.