Proc. Natl. Acad. Sci. USA
Vol. 82, pp. 1906–1910, April 1985
Biochemistry

Periodicity of leucine and tandem repetition of a 24-amino acid segment in the primary structure of leucine-rich α_2-glycoprotein of human serum

(protein structure/amphiphilic structure/gene duplication/membrane protein/galactosamine and glucosamine oligosaccharides)

NOBUHIRO TAKAHASHI, YOKO TAKAHASHI, AND FRANK W. PUTNAM*

Department of Biology, Indiana University, Bloomington, IN 47405

Contributed by Frank W. Putnam, November 15, 1984

ABSTRACT **The complete primary structure of the 3.1S leucine-rich α_2-glycoprotein (LRG) present in human plasma has been determined. This protein (M_r ≈45,000) consists of a single polypeptide chain with one galactosamine and four glucosamine oligosaccharides attached. The polypeptide has two intrachain disulfide bonds and contains 312 amino acid residues of which 66 are leucine. The amino acid sequence can be exactly divided into 13 segments of 24 residues each, eight of which exhibit a periodic pattern in the occurrence of leucine, proline, and asparagine. The consensus sequence for the repeating tetracosapeptide unit is Pro-Xaa-Xaa-Leu-Leu-Xaa-Xaa-Xaa-Xaa-Xaa-Leu-Xaa-Xaa-Leu-Xaa-Leu-Xaa-Xaa-Asn-Xaa-Leu-Xaa-Xaa-Leu. This periodicity suggests that the unique structure of LRG arose from a series of unequal crossovers of a precursor oligonucleotide sequence that encoded a building block rich in leucine. Overall, the amino acid sequence of LRG is not significantly homologous to the continuous sequence of any protein in the current data base. However, the consensus tetracosapeptide sequence shows strong homology to segments of many mitochondrial proteins, viral envelope proteins, and oncogene proteins that have a high leucine content and transmembrane domains. Tandem repetition of similar segments also occurs in apolipoproteins that have amphipathic helical potential. Prediction of the secondary structure by the Chou–Fasman rules and calculation of the hydrophilic/hydrophobic profile by several methods confirm the tandem repetition of largely hydrophobic structural units; these begin with a β-turn that leads into an organized structure with α-helical or β-sheet potential. These structural characteristics and the homology to mitochondrial proteins and apolipoproteins suggest that LRG is a membrane-derived or membrane-associated protein containing a series of domains capable of bipolar surface orientation.**

Although complete amino acid sequences have been reported for about half of the nearly 100 proteins that have been isolated from human plasma (1), little is known about the structure of a series of well-characterized α- and β-glycoproteins of uncertain or unknown function (2). We have begun study of some of these as part of a program for investigation of the structure and function of human plasma proteins (3, 4). We report here the complete amino acid sequence of a trace component of unknown function that was isolated from human serum in 1977 by Haupt and Baudner (5). They called it 3.1S leucine-rich α_2-glycoprotein (LRG) because of its sedimentation coefficient, its unusually high content of leucine, and its electrophoretic mobility. The protein was reported to consist of a single polypeptide chain ($M_r \approx 50{,}000$) and to contain about 23% carbohydrate by weight (5). To our knowledge, no other structural data have since been published.

LRG is a trace protein of human plasma. The average content of LRG in adult serum is 2.1 mg/100 ml (5). Although neither the site of synthesis nor the function are known and no relationship to disease has been reported, LRG is of interest because of its unusually high content of leucine, which is about 17% by weight. Thus, about every fifth residue is leucine. Whether the leucine is randomly distributed or occurs in a regular structural pattern could only be determined by amino acid sequencing, which we undertook and completed. We found that the sequence exhibits a novel periodic pattern unlike that previously reported for soluble globular proteins except apolipoproteins (6–8). The single polypeptide chain is composed of 312 amino acid residues, and the sequence can be exactly divided into 13 blocks of 24 residues; eight of these tetracosapeptide sequences exhibit a periodic pattern in the distribution of leucine and certain other residues. The periodic pattern is highly significant statistically; the segment comparison score in SD units for the top 100 scores is 18.06. This suggests that LRG arose by a series of unequal crossovers of an oligonucleotide sequence that encoded a prototype leucine-rich building block of about 24 amino acid residues.

The periodicity in the primary structure of LRG is due to a recurring pattern of leucine residues that produces a segment similar in hydrophobicity and length to that found in the transmembrane segments of amphiphilic (amphipathic) peptides and proteins (9, 10). Thus, we suggest that LRG may be a membrane-associated or membrane-derived protein. This hypothesis is supported by predictions of the secondary structure and the hydrophilic/hydrophobic character (hydropathy profile) of the polypeptide chain.

MATERIALS AND METHODS

Materials. Purified LRG prepared from human serum by the method of Haupt and Baudner (5) and antiserum specific for LRG were obtained from Behringwerke AG (Marburg/Lahn, F.R.G.). The protein migrated as a single band in Na-DodSO$_4$/polyacrylamide gel electrophoresis both in the presence and absence of 2-mercaptoethanol.

Methods. Methods used for the determination of the primary structure of proteins in our laboratory have been described (3, 4, 11). The purified LRG was reduced and carboxymethylated before sequence analysis. The carboxymethylated protein (5–15 mg) was subjected to separate digestions with L-1-tosylamido-2-phenylethyl chloromethyl ketone-treated trypsin, *Staphylococcus aureus* V8 protease, and endoproteinase Lys-C. The intact protein was also

The publication costs of this article were defrayed in part by page charge payment. This article must therefore be hereby marked "*advertisement*" in accordance with 18 U.S.C. §1734 solely to indicate this fact.

Abbreviation: LRG, leucine-rich α_2-glycoprotein.

*To whom reprint requests should be addressed.

cleaved with dilute acid at 108°C for 4 hr (12). Each enzymatic digest was separated by a combination of gel filtration and high-performance liquid chromatography (HPLC) on an Ultrasphere ODS column (Altex, Berkeley, CA) or a Synchropak RP-P column (SynChrom, Linden, IN) with a programmed gradient of 1-propanol containing 0.1% trifluoroacetic acid or 0.1% heptafluorobutyric acid (4). The digest with dilute acid was separated by an automated two-step HPLC system; this consists of a combination of anion-exchange chromatography with a Spherogel-TSK IEX-540 DEAE column (Altex) as the first step and chromatography with a reversed-phase Ultrasphere ODS column as the second. The purified peptides were analyzed with the amino acid analyzer, and their sequences were determined by automatic Edman degradation (3, 4, 11). Hexosamine analysis was also done with the amino acid analyzer, after acid hydrolysis (3, 11).

Computer Analysis of Sequence Data. The sequence data base of the *Atlas of Protein Sequence and Structure* updated to August 1984 and the programs SEARCH, ALIGN, RELATE, and PRPLOT were provided by the National Biomedical Research Foundation.† The programs SEARCH, RELATE, and ALIGN (13) were used either with the unitary matrix or with the mutation data matrix; all gave a score for statistical significance in standard deviations (SD) of the real score above a score of 100 random runs. We used PRPLOT to plot the hydrophilic/hydrophobic profile of the molecule by utilizing different scales including those of Kyte and Doolittle (14), Wolfenden *et al.* (15), Argos *et al.* (16), and Eisenberg (10). PRPLOT was also used to estimate the secondary structure predicted by the procedure of Chou and Fasman (17) based on their values for the tendency of individual amino acids to appear in α-helix, β-sheet, and β-turn structures.

RESULTS AND DISCUSSION

Amino Acid Composition. The amino acid composition calculated from the sequence analysis (Table 1) corresponds well with that obtained by amino acid analysis of the protein (5). The very high content of leucine (66 of 312 residues) is an unusual characteristic of this protein. By amino acid analysis the leucine content of the glycoprotein is 16.62% by weight (5). The value for the polypeptide portion calculated from the complete sequence is 21.75% by weight or 21.12 mole %. A search of the protein sequence data base† showed that the leucine content of LRG is exceeded only by those of a vespid venom peptide, a few hypothetical bacteriophage and mitochondrial proteins corresponding to gene sequences, and the hormone secretin. Most other sequences listed that had a high leucine content (15–20 mole %) also were for hypothetical mitochondrial proteins or were for known mitochondrial cytochromes and enzymes, viral coat proteins, or biologically active peptides. About one-third of the amino acid residues of LRG are strongly hydrophobic. Other notable features are the low content of tyrosine (3 residues), isoleucine (4 residues), and cysteine (4 residues). The latter are linked in two disulfide bonds; one is close to the amino terminus and is between cysteine-8 and cysteine-21, and the other is near the carboxyl terminus and is between cysteine-268 and cysteine-294.

Polypeptide Structure and Molecular Weight. Human LRG consists of a single polypeptide chain containing 312 amino acid residues, one GalN oligosaccharide, and four GlcN oligosaccharides (Fig. 1). The M_r calculated for the unglycosylated polypeptide chain is 34,346. An M_r of 35,970 was obtained from electrophoresis in $NaDodSO_4$/polyacrylamide gels when the results estimated in gels ranging from 5% to 20% gel concentration were extrapolated by use of a double-reciprocal plot. This value, which presumably reflects only the polypeptide chain, is much less than the M_r of 49,600 ± 4,000 determined by equilibrium sedimentation (5). However, the carbohydrate content of LRG (about 23% by weight) contributes significantly to the M_r and also makes it difficult to determine the M_r accurately by physicochemical methods. We estimate that the M_r of LRG is approximately 45,000, based on the value of 34,346 for the polypeptide portion and an estimate of 600 for the GalN oligosaccharide and of 2500 for each of the four GlcN oligosaccharides.

Table 1. Amino acid composition of human LRG based on the complete amino acid sequence determination

Amino acid	No. of residues	Amino acid	No. of residues
Aspartic acid	19	Valine	14
Asparagine	17	Methionine	3
Threonine	12	Isoleucine	4
Serine	21	Leucine	66
Glutamic acid	14	Tyrosine	3
Glutamine	20	Phenylalanine	10
Proline	22	Lysine	12
Glycine	23	Histidine	7
Alanine	21	Arginine	15
Half-cystine	4	Tryptophan	5

M_r of unmodified polypeptide chain is 34,346; number of residues is 312. Thr-2 is linked to a GalN oligosaccharide. Asn-44, Asn-151, Asn-234, and Asn-290 are linked to GlcN oligosaccharides.

Number and Location of Oligosaccharides. Fig. 1 gives the complete amino acid sequence of LRG and shows the sites of attachment of the five oligosaccharides. The GalN is O-linked to the second residue (Thr-2). Both in its proximity to the amino terminus and to a nearby proline, this linkage is similar to that of the GalN attached to the amino-terminal residue (threonine) in human hemopexin (11, 18). However, no obligatory signal sequence for the attachment of GalN is known, whereas GlcN is always N-linked to asparagine in the tripeptide sequence Asn-Xaa-(Thr/Ser). In LRG there are four GlcN oligosaccharides; these are linked to Asn-44, Asn-151, Asn-234, and Asn-290. However, the asparagine is not glycosylated in the putative acceptor sequence Asn-Leu-Ser at positions 271–273. Nothing has been published about the carbohydrate structure of the oligosaccharides of LRG; however, a complex dibranched GlcN oligosaccharide is the most common carbohydrate in human plasma glycoproteins (19).

Internal Homology. Although many plasma proteins exhibit strong evidence of internal duplication in amino acid sequence (20, 21), at first appearance LRG does not. Despite the frequency of leucine residues, there are only five instances of a Leu-Leu sequence and only one case of the triplet sequence Leu-Leu-Leu, compared to 53 instances of single leucine residues. This lack of a statistical distribution puzzled us. However, first by visual inspection and then by computer analysis with the program RELATE, we found that the amino acid sequence of LRG fits a 13-cycle periodic pattern dominated by a 24-residue peptide (tetracosapeptide) which has the consensus sequence Pro-Xaa-Xaa-Leu-Leu-Xaa-Xaa-Xaa-Xaa-Xaa-Xaa-Leu-Xaa-Xaa-Leu-Xaa-Leu-Xaa-Xaa-Asn-Xaa-Leu-Xaa-Xaa-Leu. All of the amino acid residues identified in this sequence occur at least six times in the alignment shown in Fig. 2. In the computer listing of the 100 segments having the highest scores for intrasequence homology (the top 100 of 39,903 comparisons), the lengths of all

†Barker, W. C., Hunt, L. T., Orcutt, B. C., George, D. G., Yeh, L. S., Chen, H. R., Blomquist, M. C., Johnson, G. C. & Dayhoff, M. O. (Aug. 1, 1984) *Atlas of Protein Sequence and Structure*, Protein Sequence Database, (Natl. Biomed. Res. Found., Washington, D.C.).

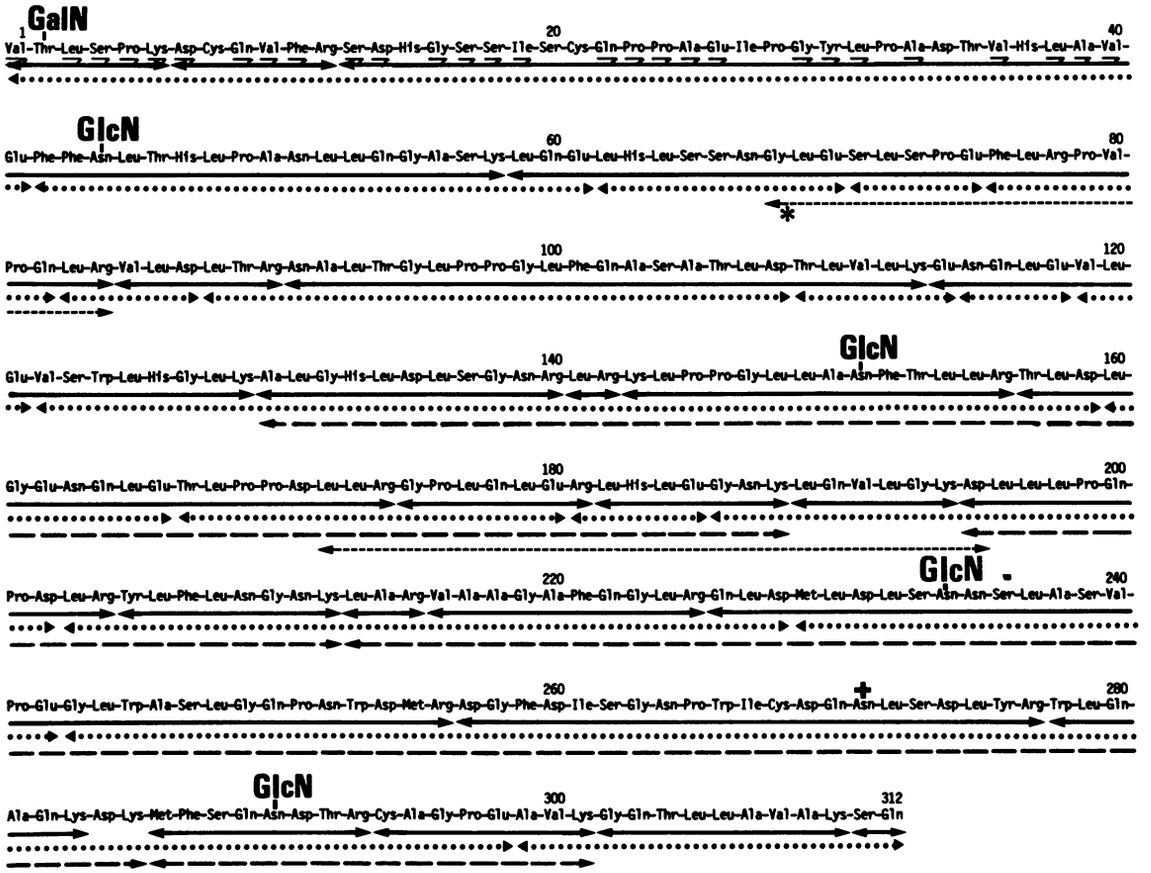

FIG. 1. Summary of the complete amino acid sequence of human LRG. The sequence is shown along with all peptides necessary for the proof of sequence. The peptides obtained from different digestions are as follows: ———, tryptic peptides; ·····, *S. aureus* V8 peptides; – – –, endoproteinase Lys-C peptides; ---------, dilute acid peptides. The asterisk indicates a dilute acid peptide of a tryptic peptide. GalN and GlcN indicate the attachment sites for the galactosamine oligosaccharide and the glucosamine oligosaccharides, respectively. A potential glycosylation site (Asn-271) is shown by a plus sign. No evidence for polymorphism in the sequence of LRG was found except for a possible ambiguity at position 215, where tyrosine may replace arginine in some molecules.

segments listed were divisible by 24, and all the segments fit exactly into the periodic pattern of Fig. 2. In this pattern, the complete sequence of 312 residues is exactly divided into an array of 13 tetracosapeptide segments with no gaps inserted. The segment comparison score for the tetracosapeptide segments is 18.06 SD units. A score of 3.0 SD units is considered significant (13).

The unique feature of the 13-cycle alignment is that it is based mainly on a periodic distribution of a single amino acid rather than on a recurrent sequence of several different amino acids. More than 50 of the 66 leucine residues fit the pattern, including almost all the leucines in the 200 successive residues (nos. 38–237) in which the most homologous tetracosapeptide segments are located. Four more leucines are aligned in column 8 of Fig. 2. Since leucine is one of the most hydrophobic amino acids and is usually present in an or-

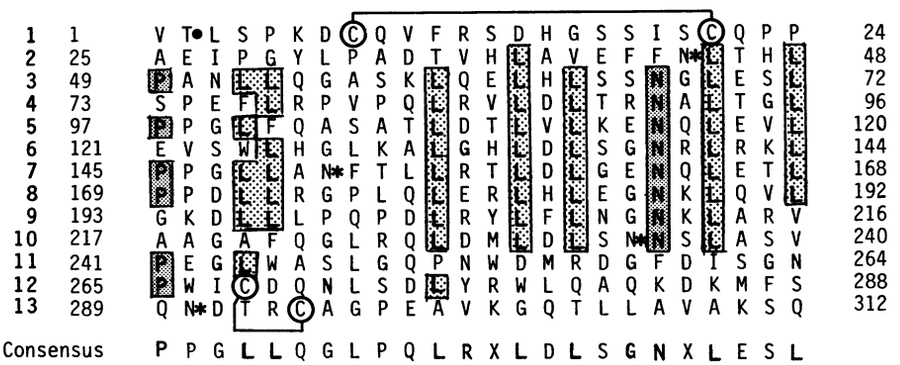

FIG. 2. Periodicity of leucine residues and tandem repetition of a 24-amino acid segment in the primary structure of LRG. The sequence of Fig. 1 with position numbers is shown in the one-letter code for amino acids (13) and is aligned exactly into 13 segments (numbered in bold type). Amino acids that occur at least six times in a vertical column are shown in bold type and are enclosed in shaded boxes. The consensus sequence is given at the bottom, with the most frequent residues in bold type. Others that occur three or four times are also given in the consensus sequence, and X is used where there is no clear preference. The cysteine (C) residues are circled and the disulfide bonds are represented schematically. An asterisk denotes each asparagine (N) to which a GlcN oligosaccharide is linked, and a solid circle at threonine-2 indicates that a GalN oligosaccharide is attached.

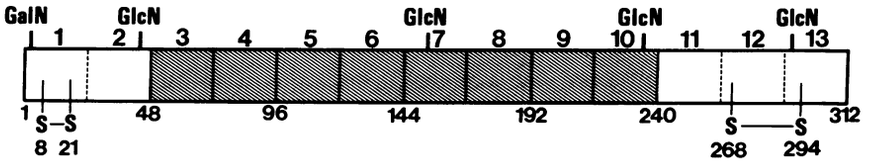

FIG. 3. Schematic representation of the tandem repetition of 24-amino acid segments in LRG. Each numbered block represents a segment (see Fig. 2). Shaded blocks represent homologous segments that exhibit a periodic pattern of leucine and other amino acids. The numbers below the diagram give the sequence positions. The locations of the GalN and GlcN oligosaccharides and of the two disulfide bonds are shown.

dered structure (α-helix or β-pleated sheet), the periodicity of leucine residues probably reflects a unique structural characteristic of LRG that has been attained by a series of unequal crossovers of an oligonucleotide sequence that coded for a prototype building block that was rich in leucine. Although the entire polypeptide sequence is exactly divisible into 13 tetracosapeptide segments, only the central 8 segments exhibit strong inter-segment homology. These comprise a central core structure, illustrated schematically in Fig. 3. The amino-terminal segment has a GalN oligosaccharide and an intra-segment disulfide bond. The periodic pattern of homology ends after segment 10. The last two segments are joined by a disulfide bond, and the last has a GlcN oligosaccharide.

Sequence Homology to Other Proteins. To examine the possible occurrence in other proteins of sequences homologous to the periodic tetracosapeptide of LRG, we used as a probe a consensus sequence that represents the amino acid of highest frequency in each vertical column of Fig. 2. The sequence used was Pro-Pro-Gly-Leu-Leu-Gln-Gly-Leu-Pro-Gln-Leu-Arg-Xaa-Leu-Asp-Leu-Ser-Gly-Asn-Xaa-Leu-Glu-Ser-Leu, in which Xaa indicates that no amino acid occurred with sufficient frequency.

By use of the program SEARCH and both the unitary matrix and the mutation data matrix, we found that LRG has a unique amino acid sequence that does not exhibit significant homology over a series of contiguous segments to any protein in the current sequence data base.† However, most of the repeating tetracosapeptide segments of LRG do exhibit striking homology of their basic pattern with segments of a number of polypeptides, most of which have a high leucine content. These polypeptides include many viral envelope proteins, oncogene products, and a number of known mitochondrial proteins, such as various cytochromes as well as hypothetical mitochondrial proteins for which the sequences have been deduced by cloning techniques. For example, when the unitary matrix was used, the highest score for the prototype tetracosapeptide sequence was given by a 24-residue segment of hypothetical protein 4 of the human mitochondrion, which has a leucine content of 21 mole % (22). Other segments of protein 4 and other mitochondrial proteins scored high, but the matching segments were not contiguous in the sequences of the proteins. When the mutation data matrix was used, the highest score was for the hemagglutinin precursor of various strains of human and duck influenza virus (residues 441–464). This segment corresponds to residues 78–101 in the second (HA_2) chain of the processed virus protein and exists as a long α-helix with a 36-Å hydrophobic exterior surface (23).

We do not suggest that the homology identified signifies a genetic or evolutionary relationship. Rather, it probably reflects the fact that a high leucine content may result in a periodicity in sequence similar to that of LRG. In an organized structure such as the α-helix or β-pleated sheet (for both of which leucine has a high potential), the hydrophobic residues tend to cluster in a similar orientation (9, 10). Such hydrophobic patches must either be in the interior of protein molecules or, if on the exterior, will tend to interact with hydrophobic membrane surfaces.

The high solubility of LRG and its resistance to heat coagulation are probably due to the combination of a high carbohydrate content, which greatly increases the hydrophilicity of the protein surface, and to the clustering of the hydrophobic leucine residues, which are probably internalized in the protein.

Tandem Repetitions of Protein Sequences. Although fibrous proteins such as collagen have frequent short (tripeptide) repeating units, tandem repetitions of the length and type found in LRG are rare. However, some apolipoproteins contain a series of 11-amino acid or 22-amino acid segments similar to although not closely homologous with the repeating sequence in LRG (6–8). Fitch (6) identified a 13-cycle repeat of an 11-amino acid (or 22-amino acid) sequence in human apolipoprotein A-I and proposed that it resulted from a series of intragenic unequal crossovers. Barker and Dayhoff (7) identified a similar periodicity in human apolipoproteins A-II, C-II, and C-III and suggested that the four lipoproteins arose from a single ancestral gene. Recently Boguski *et al.* (8) reported that rat apolipoprotein A-IV contains 13 tandem repetitions of a 22-amino acid segment with amphipathic helical potential. The consensus sequence of the apolipoproteins is similar to that of LRG in length, in the number of repetitions, and in the frequency of leucine and other hydrophobic residues. However, the sequence homology of LRG and apolipoproteins is not pronounced; hence, they probably have a functional rather than a genetic relationship. The apolipoprotein repeating sequence can be drawn as an α-helical wheel with amphiphilic character; that is, the nonpolar residues are segregated on one section of the circumference, and hydrophilic and charged residues are aligned on the rest. The Chou–Fasman rules (17) lead to the prediction of an alternating secondary structure, which is in accord with a plot of the hydropathic index. The result is a helical amphiphilic structure to which the function of lipid binding is attributed.

Hydrophilic/Hydrophobic Profile. When the computer program PRPLOT was used with different scales (10, 14–16) to evaluate the hydrophilic/hydrophobic profile of LRG, the results were in general agreement. Computation of the hydropathic index of Kyte and Doolittle (14) (Fig. 4) shows that the eight conserved segments of the central core are largely hydrophobic, whereas the amino-terminal segment and the last two segments are mainly hydrophilic. Each of the central eight segments has a predominantly hydrophobic pattern with a short hydrophilic section in the middle or near the carboxyl terminus. The patterns for the eight tandem segments are similar but not identical. The carboxyl-terminal segment appears to end with a hydrophobic segment; this could be buried inside the molecule, or it could be a transmembrane segment. In view of the high leucine content, the hydrophobicity of LRG is not unexpected. However, using the method of Eisenberg (10), we were unable to identify a series of potential membrane segments of 21–24 residues in length, such as those he described in some membrane proteins. Also, the periodic leucine residues did not all cluster together in an α-helical wheel as the hydrophobic residues tend to do in apolipoproteins and amphiphilic peptides (8–10).

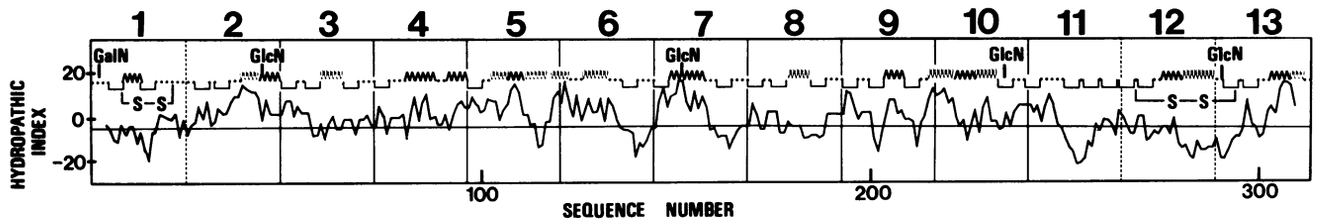

FIG. 4. Hydropathy profile (14) and secondary structure predictions (17) for LRG calculated by use of the PRPLOT program (see text footnote †). In the secondary structure prediction, the residues are represented by dots and are shown in α-helical (wwww), β-sheet (wwww), and β-turn (MM) conformational states. Locations are given for the oligosaccharides (GalN and GlcN) and for the disulfide bonds. The vertical lines divide the structure into the 13 numbered tandem segments for which the amino acid sequences are given in Fig. 2.

Secondary Structure. The secondary structure predicted by the Chou–Fasman rules (17) shows some repetition of structural elements in the eight tandem conserved segments of the central core of the molecule (shaded section of Fig. 3). In fact, the predicted structure is nearly identical for segments 3 and 8, which have 11 residues in common. Although the predicted secondary structure of the core segments is not identical, the general pattern is similar. Most of the conserved segments begin with a β-turn, followed by an organized structure (either α-helical or β-sheet) in the middle of the segment, and end with a β-turn. The two segments at the amino terminus and the two at the carboxyl terminus are each the loci of a disulfide bond and of carbohydrate, both of which greatly affect the local structure in a way that is hard to predict by the Chou–Fasman rules. These structural characteristics and the homology to mitochondrial proteins and apolipoproteins suggest that LRG is a membrane-derived or membrane-associated protein containing a series of domains capable of bipolar surface orientation.

We thank P. H. Davidson, S. A. Dorwin, J. Madison, and K. Huss for excellent technical assistance. This work was supported by Grant AM19221 from the National Institutes of Health.

1. Putnam, F. W., ed. (1984) *The Plasma Proteins* (Academic, Orlando, FL), 2nd Ed., Vol. 4.
2. Schwick, H. G. & Haupt, H. (1984) in *The Plasma Proteins*, ed. Putnam, F. W. (Academic, Orlando, FL), 2nd Ed., Vol. 4, pp. 167–220.
3. Lozier, J., Takahashi, N. & Putnam, F. W. (1984) *Proc. Natl. Acad. Sci. USA* **81**, 3640–3644.
4. Takahashi, N., Takahashi, Y., Ortel, T. L., Lozier, J., Ishioka, N. & Putnam, F. W. (1984) *J. Chromatogr.*, **317**, 11–25.
5. Haupt, H. & Baudner, S. (1977) *Hoppe-Seyler's Z. Physiol. Chem.* **358**, 639–646.
6. Fitch, W. M. (1977) *Genetics* **86**, 623–644.
7. Barker, W. C. & Dayhoff, M. O. (1977) *Comp. Biochem. Physiol. B* **57**, 309–315.
8. Boguski, M. S., Elshourbagy, N., Taylor, J. M. & Gordon, J. I. (1984) *Proc. Natl. Acad. Sci. USA* **81**, 5021–5025.
9. Kaiser, E. T. & Kézdy, F. J. (1983) *Proc. Natl. Acad. Sci. USA* **80**, 1137–1143.
10. Eisenberg, D. (1984) *Annu. Rev. Biochem.* **53**, 595–623.
11. Takahashi, N., Takahashi, Y. & Putnam, F. W. (1984) *Proc. Natl. Acad. Sci. USA* **81**, 2021–2025.
12. Inglis, A. S. (1983) *Methods Enzymol.* **91**, 324–332.
13. Barker, W. C., Ketcham, L. K. & Dayhoff, M. O. (1978) in *Atlas of Protein Sequence and Structure*, Dayhoff, M. O. ed. (Natl. Biomed. Res. Found., Washington, DC), Vol. 5, Suppl. 3, pp. 359–362.
14. Kyte, J. & Doolittle, R. F. (1982) *J. Mol. Biol.* **157**, 105–132.
15. Wolfenden, R., Anderson, L., Cullis, P. M. & Southgate, C. C. B. (1981) *Biochemistry* **20**, 849–855.
16. Argos, P., Rao, J. K. M. & Hargrave, P. A. (1982) *Eur. J. Biochem.* **128**, 565–575.
17. Chou, P. Y. & Fasman, G. O. (1978) *Annu. Rev. Biochem.* **47**, 251–276.
18. Takahashi, N., Takahashi, Y. & Putnam, F. W. (1985) *Proc. Natl. Acad. Sci. USA* **82**, 73–77.
19. Baenziger, J. (1984) in *The Plasma Proteins*, ed. Putnam, F. W. (Academic, Orlando, FL), 2nd Ed., Vol. 4, pp. 271–315.
20. Doolittle, R. F. (1984) in *The Plasma Proteins*, ed. Putnam, F. W. (Academic, Orlando, FL), 2nd Ed., Vol. 4, pp. 317–359.
21. Ortel, T. L., Takahashi, N. & Putnam, F. W. (1984) *Proc. Natl. Acad. Sci. USA* **81**, 4761–4765.
22. Anderson, S., Bankier, A. T., Barrell, B. G., de Bruijn, M. H. L., Coulson, A. R., Drouin, J., Eperon, I. C., Nierlich, D. P., Roe, B. A., Sanger, F., Schreier, P. H., Smith, A. J. H., Staden, R. & Young, I. G. (1981) *Nature (London)* **290**, 457–465.
23. Wilson, I. A., Skehel, J. J. & Wiley, D. C. (1981) *Nature (London)* **289**, 366–373.