Proc. Natl. Acad. Sci. USA
Vol. 83, pp. 2295–2299, April 1986
Biochemistry

Use of hydrophobicity profiles to predict receptor binding domains on apolipoprotein E and the low density lipoprotein apolipoprotein B–E receptor

(sequence/structure prediction/apolipoprotein E mutation)

H. DE LOOF*, M. ROSSENEU*, R. BRASSEUR†, AND J. M. RUYSSCHAERT†

*Department of Clinical Biochemistry, AZ St.-Jan, Brugge 8000, Belgium; and †Laboratory of Macromolecules at Interface, Brussels Free University, ULB, Brussels 1050, Belgium

Communicated by I. Prigogine, August 13, 1985

ABSTRACT **We have used mean hydrophobicity and hydrophobic moment calculations to predict the receptor binding domains in apolipoprotein E and in the low density lipoprotein apolipoprotein B–E receptor. In apolipoprotein E, two receptor binding domains, residues 136–160 and 214–236, having a high hydrophilicity and a high mean helical hydrophobic moment, were identified. The first domain has been located experimentally and mutations influencing the hydrophobicity parameters of the binding site have been shown to affect the receptor binding. The second domain is probably, either separately or in combination with the first domain, involved in receptor binding or in heparin binding. In the low density lipoprotein apolipoprotein B–E receptor, six protein domains were identified. In the first domain (residues 1–371), eight hydrophilic maxima, organized in pairs through disulfide bonds, form the four experimentally observed receptor binding sites. These sites consist of repeats of 26 amino acids but differ from those reported by others [Yamamoto, T., Davis, C. G., Brown, M. S., Schneider, W. J., Casey, M. L., Goldstein, J. W. & Russell, D. W. (1984) *Cell* 39, 27–38]. The second, more hydrophobic, domain (residues 372–640) forms the core of the receptor, explaining its homology with the precursor of mouse epidermal growth factor, while the cysteine residues in the third domain (residues 641–699), interacting with those of the first domain, further stabilize the molecule. Beyond the fourth hydrophilic domain (residues 700–767), to which carbohydrates are linked, a very hydrophobic membrane spanning region (residues 768–789) could be detected easily. The last domain (residues 790–839), situated in the cytoplasma, contains hydrophilic maxima, as this region might interact with clathrin-related proteins. These data suggest that hydrophobicity analysis can detect and predict protein domains: hydrophilic receptor sites as well as hydrophobic core-forming and membrane-spanning regions.**

The apolipoproteins are soluble plasma proteins whose major function is the transport, in the form of soluble complexes, of cholesterol, triglycerides, and phospholipids in plasma and the delivery of these complexes, called lipoproteins, to various cells. The interaction of lipoproteins with cells involves the interaction with cellular receptors (1, 2). Recently one of these, the low density lipoprotein (LDL) apolipoprotein B–E [apo(B–E)] receptor, has been isolated (3), characterized, and sequenced (4, 5). Among the known apolipoproteins, apoE and apoB, the major protein component of LDL, are known to interact with the LDL apo(B–E) receptor (2, 3). The amino acid sequence of apoB is still under investigation, while the sequences of apoE and some mutants have been elucidated (6). Residue modifications of apoB and apoE have shown the involvement of positive amino acids (6). Studies using monoclonal antibodies (7), protein cleavage (8), and apoE mutants (6) have further located the receptor binding site of apoE.

In apolipoproteins, amphipathic helical structure is thought to be involved in lipid binding and other physiological functions (9). The amphiphilicity can be expressed in terms of the mean helical hydrophobic moment $\langle\mu_H\rangle$ and the mean hydrophobicity $\langle H_i\rangle$, as proposed by Eisenberg *et al.* (10). Following this method transmembrane helices, helices from globular proteins, and "surface-seeking helices" cluster in different regions of a plot of $\langle\mu_H\rangle$ against $\langle H_i\rangle$ (10, 11). We extended these hydrophobicity analyses to predict other functional domains in apolipoproteins and their receptor protein.

METHODS

The mean α-helical hydrophobic moment $\langle\mu_H\rangle$ and the mean hydrophobicity $\langle H_i\rangle$ were calculated according to the procedure of Eisenberg *et al.* (10, 11). Then, a segment of 11 amino acids was moved through the protein sequence and the mean hydrophobicity and mean hydrophobic moment per segment were calculated. $\langle\mu_H\rangle$ was plotted versus $\langle H_i\rangle$ for all possible segments except those containing a proline residue, since this residue has a strong helix-breaking capacity (12). These two parameters were also plotted as a function of the midpoint of the amino acid segment along the sequence.

To predict the protein secondary structure, we used the method developed by Chou and Fasman (12–14). In most calculations we used the normalized "consensus" hydrophobicity scale (11). For calculation of the LDL apo(B–E) receptor we used the normalized scale of Argos, as described by Eisenberg (11), because it is especially suitable for membrane-related proteins.

RESULTS

Prediction of the Receptor Binding Sites in apoE. Several helices having a relatively high hydrophobicity ($-0.2 < H_i < 0.4$) and a high hydrophobic moment ($0.3 < \mu_H < 0.65$) can be detected in the $\langle\mu_H\rangle$ versus $\langle H_i\rangle$ plot of apoE (Figs. 1 and 2). Such helices can also be found in other apolipoproteins and are suggested to be involved in lipid binding (15). The most striking feature of these plots (Figs. 1 and 2) is, however, the appearance of two clusters of residues having a very low hydrophobicity ($H_i < -0.55$) together with a very high ($\mu_H > 0.52$, cluster A) to intermediate ($0.29 < \mu_H < 0.46$, cluster B) hydrophobic moment. Using Fig. 1, these residues can easily be located as belonging to regions 136–160 and 214–236 and having sequences Arg-Leu-Ala-Ser-His-Leu-

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.

Abbreviations: apo-, apolipoprotein; LDL, low density lipoprotein.

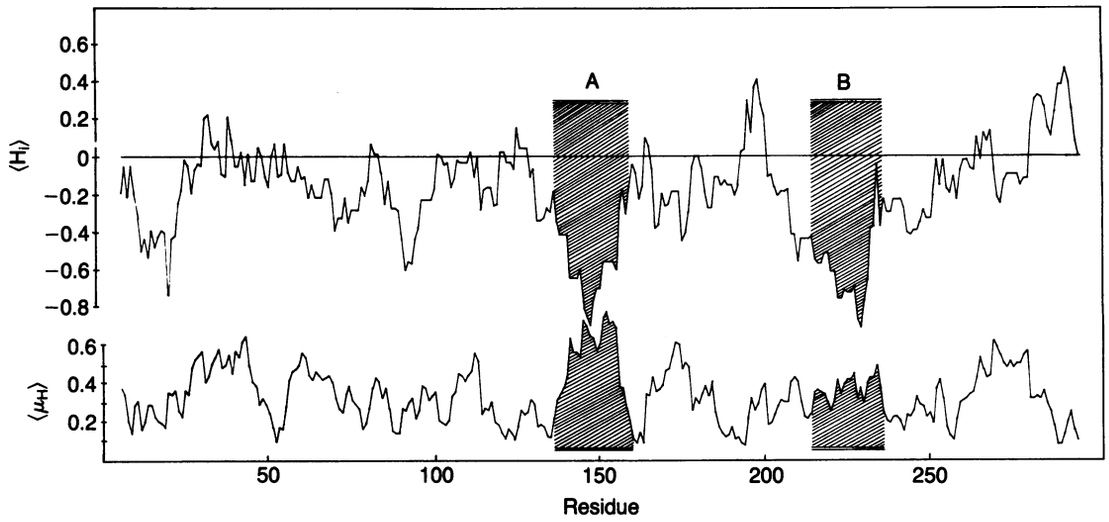

FIG. 1. Hydrophobicity profile of apoE. A plot of $\langle\mu_H\rangle$ and $\langle H_i\rangle$ as a function of the midpoints of the 11-residue-long segments used for the calculations is shown. The shaded areas between residues 136–160 and 214–236 correspond to the two clusters (A and B) described in the text. The amino acid sequence of apoE was taken from ref. 7.

Arg-Lys-Leu-Arg-Lys-Arg-Leu-Leu-Arg-Asp-Ala-Asp-Asp-Leu-Gln-Lys-Arg-Leu-Ala for the first cluster and Leu-Arg-Ala-Arg-Met-Glu-Glu-Met-Gly-Ser-Arg-Thr-Arg-Asp-Arg-Leu-Asp-Glu-Val-Lys-Glu-Gln-Val for the second cluster. This first domain corresponds to the receptor binding domain of apoE as proposed by Mahley *et al.* (6). To obtain further confirmation of the involvement of residues 136–160 in receptor binding we carried out the same calculations for all apoE mutants described so far (6). All of the substitutions affecting the receptor binding activity affected the plot of apoE in domain A. A close-up of this region for normal apoE and three mutations is shown in Fig. 3. The plot for the $Arg^{158}\rightarrow$Cys mutation (Fig. 3*B*), with receptor binding activity decreased to <2% (6), shows that the mean hydrophobicity of the 11-residue segments with midpoints at 153–155 increases from about −0.57 to −0.32 while the mean hydrophobic moments decrease from about 0.72 to 0.63. The $Cys^{112}\rightarrow$Arg, $Arg^{142}\rightarrow$Cys mutation (Fig. 3*C*), with <20% receptor binding activity, also results in extensive changes in the plot of $\langle\mu_H\rangle$ against $\langle H_i\rangle$ as both the hydrophilicity and the mean helical hydrophobic moment decrease significantly.

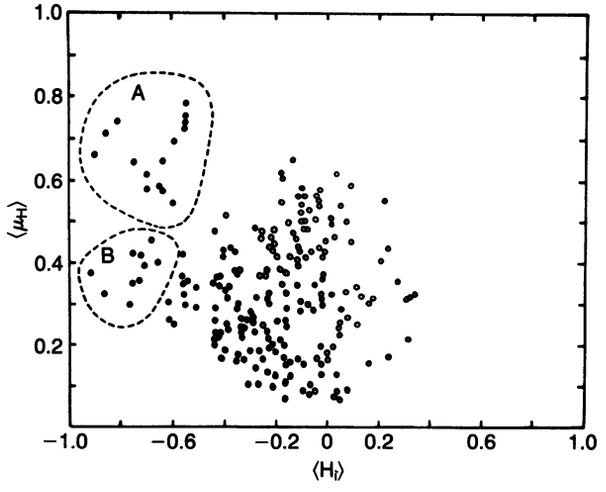

FIG. 2. Plot of $\langle H_i\rangle$ against $\langle\mu_H\rangle$ for all 11-residue-long segments of apoE except for those containing a proline residue, since this residue has strong helix breaking capacity (12). The two clusters of residues involved in receptor binding (A and B) are circled.

The modifications in the plot for a third mutation ($Lys^{146}\rightarrow$Gln) are less extensive (Fig. 3*D*). The mean hydrophobicities of this site do not exceed −0.5 and experimental data showed that this mutation retains 40% receptor binding activity.

We could detect a similar cluster of hydrophilic residues at positions 128–152 in the rat apoE sequence (16). This domain is exactly aligned to residues 136–160 of human apoE.

According to Chou and Fasman analysis (12–14), these two domains consist of two helical segments interrupted by a β-turn whereas Innerarity *et al.* (17) proposed a secondary structure comprised of a helix and a β-sheet separated by a β-turn for the first receptor binding site. The appearance of β-turns at the hydrophilic maxima is in accordance with the observations of Rose (18). This second domain has not yet been directly implicated in receptor binding, although the existence of a second binding region interacting with the β-very low density lipoprotein receptor has been proposed (19). An additional possibility might be combination of the A and B domains to form the receptor binding site of apoE. This would explain why protein cleavage at position 199 (17, 20) has a profound effect on receptor binding. Heparin binding (6) is also very likely to occur in this region.

Prediction of Functional Domains in the LDL apo(B–E) Receptor. A similar approach was used to predict the functional domains in the LDL apo(B–E) receptor, especially in terms of the apoB–E binding sites. Based on calculations according to Chou and Fasman (12–14), the receptor protein has a heterogeneous secondary structure consisting of 30% each β-sheet and β-turns and 20% helical segments. The calculation of a mean helical hydrophobic moment would therefore be meaningless and only the mean hydrophobicity of the 11-residue-long segments was plotted (Fig. 4). We identified a first structural domain consisting of the 371 NH_2-terminal amino acids, which includes 54 cysteine residues (15% of the residues). The mean hydrophobicity showed an alternation of hydrophobic and hydrophilic regions with well-characterized minima (Fig. 4). The most probable location of the receptor sites in this domain corresponds to the hydrophilic zones having negatively charged residues, able to interact with the positively charged residues of the apoE binding sites. Such zones could be localized around residues 26–38, 70–83, 108–119, 148–159, 196–208, 238–248, 279–289, and 330–342. Each of these regions, which contains at least four acidic residues and not more than two basic amino acids, has a net negative charge. Yamamoto *et al.* (4) in a search for

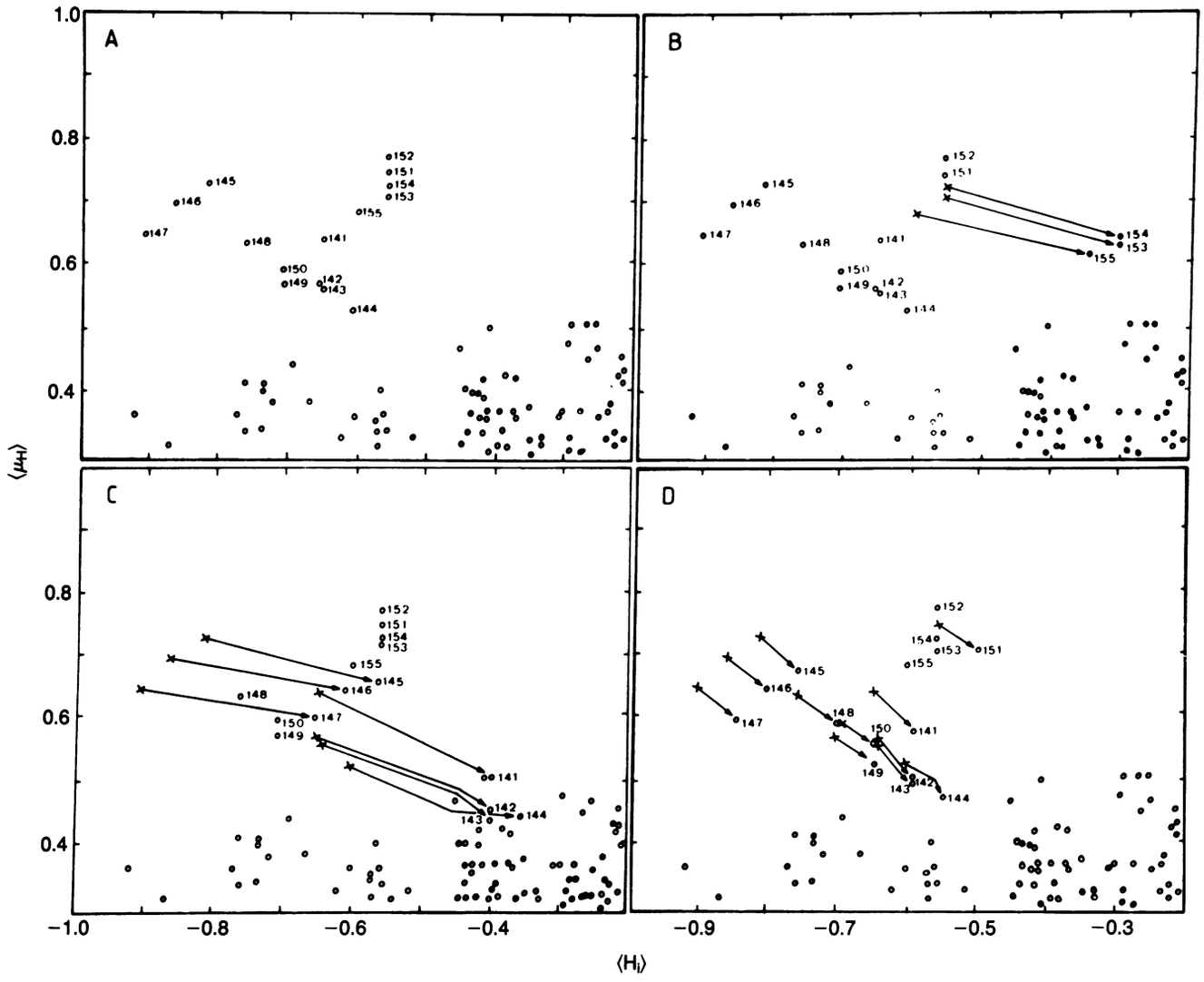

FIG. 3. Representation of the receptor binding domain of apoE, around cluster A, on an expanded scale (x axis, $\langle H_i \rangle$ from −1 to −0.2; y axis, $\langle \mu_H \rangle$ from 0.3 to 1). (*A*) Normal apoE. (*B–D*) The $Arg^{158} \rightarrow Cys$; $Cys^{112} \rightarrow Arg$, $Arg^{142} \rightarrow Cys$; and $Lys^{146} \rightarrow Gln$ mutants. Arrows indicate displacements of some residues compared to their positions in the plot of normal apoE.

repetitive sequences using the computer programs ALIGN and RELATE identified eight repetitive sequences, seven of which corresponded to the above mentioned regions, but did not include residues 330–342. Using a similar approach we could easily align the residue 319–343 segment with the C-terminal fragments of the seven first segments so that all repetitive sequences contained a cluster of acidic residues at the C-terminal end (Table 1). This residue 319–343 segment contains seven acidic, two basic, and 4 cysteine residues, and its secondary structure consists mainly of β-turns and β-sheets (88%). This is analogous to the seven other segments. In contrast, the last binding zone proposed by Yamamoto *et al.* (4) and by Goldstein and Brown (5), is more hydrophobic, contains fewer acidic residues, and has fewer than half the number of residues predicted to be in a β-turn conformation than the other sites. The secondary structure of all sites (Table 1) contains two or three β-turns together with 4 cysteine residues, which are likely to associate to form disulfide bridges within the length of these segments, thereby creating loops. In this process, negatively charged residues could come close to each other and create a two-sided cavity, where interaction with the closely packed positive charges of apoE would become possible (Fig. 5). The high degree of symmetry present in this first domain of the LDL apo(B–E) receptor and the existence of an extra segment between the fourth and the fifth of these sites suggest a folding center around residues 170–180. Through folding, extra cysteine bonds could form between residues belonging to the two halves of the first domain. Such a pathway, following after the formation of short loops, would be a relatively easy, rapid, and therefore highly probable protein folding process (ref. 21; Fig. 5).

The second domain identified in the LDL apo(B–E) receptor consists of the more hydrophobic zone between residues 372 and 640. This zone has a predominant β-sheet structure (45%) and contains only 16% β-turns compared to 23 and 44%, respectively, in the first domain. Moreover, this second domain contains but one cysteine and has a higher hydrophobicity (0.04) than that of the first 371 residues (−0.21). This domain is probably located in the inside of the receptor protein where it would create a hydrophobic core. This structural function could also explain the high homology with the precursor of the mouse epidermal growth factor (4). The third domain (residues 640–699) has a high hydrophobicity (0.14) and contains six cysteine residues. This domain resembles the first one in the high content of cysteine residues (10%). We postulate that the six remaining cysteines (at positions 646, 656, 660, 675, 677, and 690) interact with some

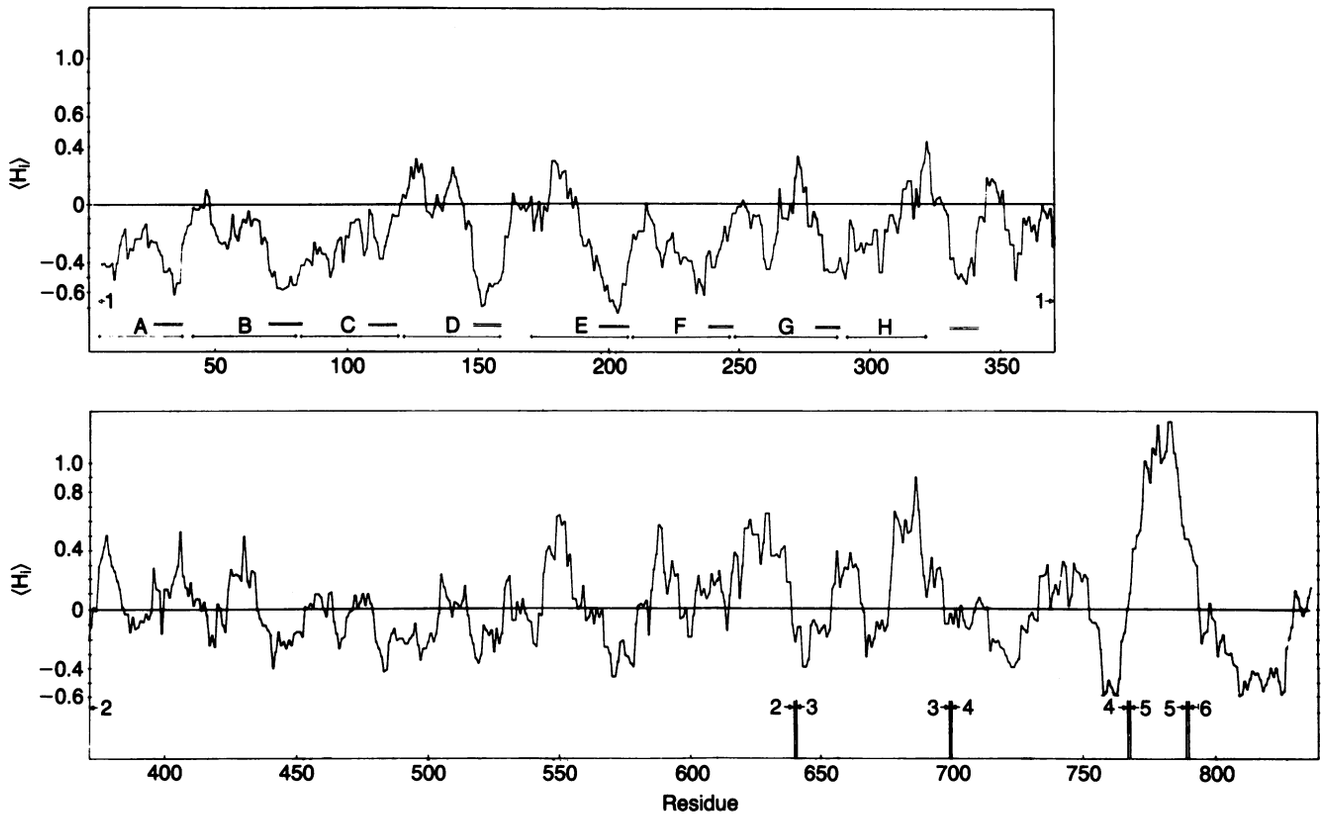

FIG. 4. Hydrophobicity profile of the LDL apo(B–E) receptor along the amino acid sequence, using 11-residue-long segments. The different protein domains, as described in the text, are indicated by numbers. The eight repeats in the first protein domain (A–H), as reported by Yamamoto *et al.* (4), are indicated by arrows. The eight hydrophilic maxima containing several negatively charged residues, responsible for receptor interaction, are indicated by double underlining. These fragments correspond to the C terminus of the repeats given in Table 1.

of the cysteines located between residues 1 and 371 to further stabilize the protein-binding domain of the receptor (Fig. 5). Residues 675 and 677, two cysteines very close to each other in the primary structure, might for example interact with two different segments of the first domain, as has been described for other extracellular proteins (21). Such a conformation is further supported by the facts that all cysteine residues appear to be involved in disulfide bonds and that the receptor protein is extremely stable, as long as the disulfide bonds are not reduced (5).

The fourth domain (residues 700–767) has a higher hydrophilicity (−0.10) because it contains several serine and threonine residues to which carbohydrate residues are linked (4). The fifth, very hydrophobic (1.04), domain, residues 768–789, corresponds to the *trans*-membrane region of 22 residues and can be identified easily on the hydrophobicity profile (Fig. 4). The last domain, residues 790–839, represents the cytoplasmic region, has a higher hydrophilicity (−0.26), and may be able to interact with the clathrin-related proteins.

Based on these considerations, we therefore propose a model (Fig. 5) for the LDL apo(B–E) receptor in which eight segments of about 25 residues each (Table 1) located between residues 1 and 371, with a loop shape stabilized by disulfide bonds, are organized in pairs belonging to the two symmetrical parts of the first protein domain to form the four binding sites. Such a model is supported by the experimental data showing that either four LDL particles containing one apoB molecule each or four apoE molecules on one high density lipoprotein particle interact with one receptor molecule (2).

Table 1. Optimal alignment of the eight segments hypothetically responsible for interaction of the LDL apo(B–E) receptor with its ligands

13–38	C Q *D* – – G K C I S Y K W V C *D* G S A *E* C Q *D* G S *D* *E* S
54–81	C G G R V N R C I P Q F W R C *D* G Q V *D* C *D* N G S *D* *E* Q
95–120	C H *D* – – G K C I S R Q F V C *D* S *D* R *D* C L *D* G S *D* *E* A
134–159	C N S – – S T C I P Q L W A C *D* N *D* P *D* C *E* *D* G S *D* *E* W
173–208	C L S – – G *E* C I H S S W R C *D* G G P *D* C K *D* K S *D* *E* *E*
222–247	C S *D* – – G N C I H G S R Q C *D* R *E* Y *D* C K *D* M S *D* E V
263–288	C H S – – G *E* C I T L *D* K V C N M A R *D* C R *D* W S *D* *E* P
319–343	C P *D* – – G F Q L V A Q R R C *E* *D* I *D* *E* C Q *D* P – *D* T C
Consensus sequence	C G C I C *D* *D* C *D* S *D* *E*
308–322	C S – – – H V C – – N D L K I G Y *E* C L C P *D* G – – – –

Segments 13–38 to 263–288 correspond to the C terminus of the repetitive sequences reported by Yamamoto *et al.* (4). Segment 319–343 is better correlated with the previous ones than is the 308–322 segment proposed by Yamamoto *et al.* Amino acids are indicated by the single-letter code; A, Ala; C, Cys; D, Asp; E, Glu; F, Phe; G, Gly; H, His; I, Ile; K, Lys; L, Leu; M, Met; N, Asn; P, Pro; Q, Gln; R, Arg; S, Ser; T, Thr; V, Val; W, Trp; Y, Tyr. Negatively charged residues are set in italic type.

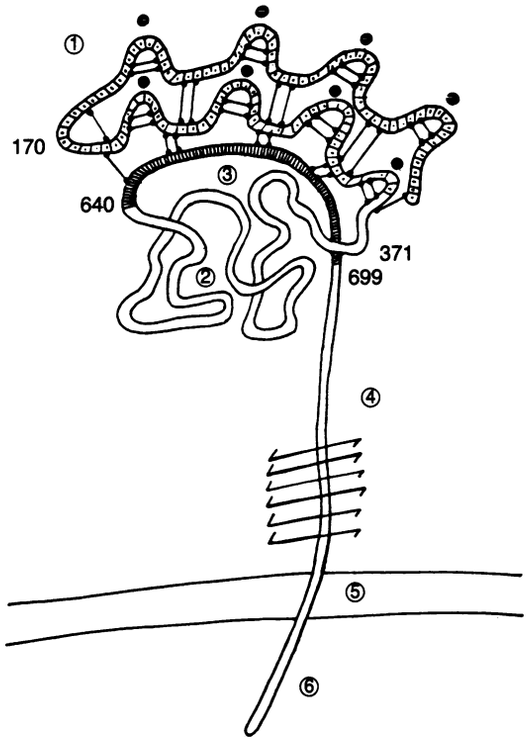

FIG. 5. Schematic model of the six domains of the LDL apo(B–E) receptor.

These sites are further stabilized through disulfide bridges with the cysteine residues between residues 640 and 699. The hydrophobic region at residues 372–640 constitutes the core of the molecule. The domain having the maximal hydrophobicity, residues 767–789, spans the lipid bilayer of the cellular membrane, while the intracellular region (residues 789–839) is more hydrophilic.

CONCLUSION

We have analyzed hydrophobicity profiles to predict and identify receptor binding domains in both a ligand (apoE) and its receptor [the LDL apo(B–E) receptor] protein. This type of analysis has led to the identification of "receptor-seeking peptides," analogous to the surface-seeking peptides proposed by Eisenberg *et al.* (10). Such residues appear to be specific for receptor binding, since they could not be identified by the same method in other apolipoproteins that have been sequenced or in other highly helical proteins such as hemoglobin, bacteriorhodopsin, or mellitin. The receptor binding site predicted by our method is in agreement with the domain identified experimentally by Mahley *et al.* (6) in normal apoE. The decreased receptor binding activity observed for some of the mutants in experimental receptor binding studies can also be predicted from the changes in $\langle \mu_H \rangle$ and $\langle H_i \rangle$ obtained in our calculations, thus supporting this theoretical approach. Using the same techniques, we detected a second binding domain (residues 214–236) that could interact either with the LDL apo(B–E) receptor together with the first domain or with other lipoprotein receptors and could be important in heparin binding.

The analysis of the hydrophobicity profile of the LDL apo(B–E) receptor has enabled the identification of the corresponding low hydrophobicity ligand binding regions in this protein. These regions coincide only in part with those previously proposed on the basis of repetitive amino acid sequences (4, 5).

This type of analysis can therefore be used to predict functional domains in proteins and to identify not only hydrophobic domains within the core of a protein or in a transmembrane region but also hydrophilic zones such as receptor binding sites on both the receptor and its ligand protein.

We are grateful to R. Lamote for assistance with various aspects of programming. H.D.L. is a recipient of an Instituut voor Wetenschappelyk Onderzoek in Nyverheid en Landbouw Fellowship.

1. Goldstein, J. L. & Brown, M. S. (1977) *Annu. Rev. Biochem.* **46**, 897–930.
2. Mahley, R. W. & Innerarity, T. L. (1983) *Biochim. Biophys. Acta* **737**, 197–222.
3. Schneider, W. J., Beisiegel, U., Goldstein, J. L. & Brown, M. S. (1982) *J. Biol. Chem.* **257**, 2664–2673.
4. Yamamoto, T., Davis, C. G., Brown, M. S., Schneider, W. J., Casey, M. L., Goldstein, J. L. & Russel, D. W. (1984) *Cell* **9**, 27–38.
5. Goldstein, J. L. & Brown, M. S. (1984) *J. Lipid Res.* **25**, 1450–1461.
6. Mahley, R. W., Innerarity, T. L., Rall, C. R. & Weisgraber, K. H. (1984) *J. Lipid Res.* **25**, 1277–1294.
7. Weisgraber, K. H., Innerarity, T. L., Harder, K. J., Mahley, R. W., Milne, R. W., Marcel, Y. L. & Sparrow, J. T. (1983) *J. Biol. Chem.* **258**, 12348–12354.
8. Innerarity, T. L., Friedlander E. J., Rall, S. C., Weisgraber, K. W. & Mahley, R. W. (1983) *J. Biol. Chem.* **258**, 12341–12347.
9. Kaiser, E. T. & Kezdy, F. J. (1984) *Science* **223**, 249–255.
10. Eisenberg, D., Weiss, R. M. & Terwilliger, T. C. (1982) *Nature (London)* **299**, 371–374.
11. Eisenberg, D. (1984) *Annu. Rev. Biochem.* **53**, 595–623.
12. Chou, P. Y. & Fasman, G. D. (1974) *Biochemistry* **13**, 211–222.
13. Chou, P. Y. & Fasman, G. D. (1974) *Biochemistry* **13**, 222–245.
14. Chou, P. Y. & Fasman, G. D. (1977) *J. Mol. Biol.* **115**, 135–175.
15. Krebs, K. E. & Phillips, M. C. (1984) *FEBS Lett.* **175**, 263–266.
16. McLean, J. W., Fukazawa, C. & Taylor, J. M. (1983) *J. Biol. Chem.* **258**, 8993–8999.
17. Innerarity, T. L., Weisgraber, K. H., Arnold, K. S., Rall, S. C. & Mahley, R. W. (1984) *J. Biol. Chem.* **259**, 7261–7267.
18. Rose, G. D. (1978) *Nature (London)* **272**, 586–590.
19. Wang-Iverson, P., Ginsberg, H. N., Peteanu, L. A., Le, N. A. & Brown, V. (1985) *Biochem. Biophys. Res. Commun.* **123**, 578–586.
20. Bradley, W. A., Hwang, S. C., Karlin, J. B., Lin, A. H. Y., Prasad, S. C., Gotto, A. M., Jr., & Gianturco, S. H. (1984) *J. Biol. Chem.* **259**, 14728–14735.
21. Schulz, G. E. & Schirmer, R. H. (1984) *Principles of Protein Structure* (Springer, New York).