# Disulfide Bridges of a Cysteine-Rich Repeat of the LDL Receptor Ligand-Binding Domain†

Stephan Bieri, Julianne T. Djordjevic, Norelle L. Daly, Ross Smith, and Paulus A. Kroon*

*Centre for Protein Structure, Function and Engineering, Department of Biochemistry, University of Queensland, Brisbane, Queensland 4072, Australia*

*Received March 3, 1995; Revised Manuscript Received July 21, 1995*⊗

ABSTRACT: The low density lipoprotein (LDL) receptor is the prototype of a family of structurally related cell surface receptors that mediate the endocytosis of multiple ligands in mammalian cells. Its ligand-binding domain consists of seven cysteine-rich ligand-binding repeats, each ~40 amino acid residues long. Ligand-binding repeats occur in other members of the LDL receptor (LDLR) gene family and in a number of functionally unrelated proteins. As a first step toward an understanding of the structure and function of LB repeats, we have expressed the amino-terminal ligand-binding repeat (LB1) of the human LDLR as a recombinant peptide (rLB1) and have determined its disulfide-pairing scheme. Oxidative folding of rLB1 yielded a single isomer which contained three disulfide bonds. This isomer reacted with a conformation-specific monoclonal antibody (IgG-C7) made to LB1 in the native LDLR, suggesting that rLB1 was correctly folded. rLB1 was resistant to digestion with trypsin, chymotrypsin, and V8 protease, consistent with a tightly folded structure. Disulfide bond connections were established using two separate approaches. Digestion with the nonspecific proteolytic enzyme proteinase K yielded an 8 amino acid peptide with a single disulfide bond which connected Cys(IV) and Cys(VI). In the second approach, disulfide bonds were sequentially reduced with tris(2-carboxyethyl)phosphine and the resulting cysteine residues alkylated with iodoacetamide. An analysis of peptides which contained two cysteinylacetamide residues, derived from a single reduced disulfide bond, showed that Cys(I) and Cys(III) were disulfide-bonded and confirmed the presence of a disulfide bond between Cys(IV) and Cys(VI). We infer that the remaining disulfide bond bridges Cys(II) and Cys(V). These disulfide bonds result in a cluster of negatively-charged residues, including the conserved Ser-Asp-Glu sequence, in a single loop, in place for interactions with positively charged residues on apoB-100 and apoE. The disulfide bond connectivity of rLB1 serves as a paradigm for other members of the LB family.

The low density lipoprotein (LDL)[1] receptor is the prototype of a family of structurally related cell surface receptors which mediate the endocytosis of multiple ligands in mammalian cells. Other members of this family include the very low density lipoprotein (VLDL) receptor (Takahashi et al., 1992), the LDL receptor-related protein/α2-macroglobulin protein (Herz et al., 1988), and renal glycoprotein gp330 (Raychowdhury et al., 1989). These receptors share a group of common structural elements, which include cysteine-rich ligand-binding (LB) or complement repeats (Hobbs et al., 1990) and cysteine-rich EGF-like repeats (Goldstein et al., 1985; Hobbs et al., 1990). Differences in the number and arrangement of these repeated sequences are thought to be responsible for the diversity of ligands which bind to this family of receptors (Hobbs et al., 1990). The LDL receptor (LDLR) has a binding domain which consists of seven contiguous LB repeats, each approximately 40 amino acids long (Südhof et al., 1985) with a repeating pattern of six cysteine residues (Cys(I)−Cys(VI)).[2] This combination of repeats allows the LDLR to bind plasma lipoproteins that contain two structurally unrelated ligands: apolipoprotein (apo) B-100, a 550 kDa glycoprotein which is present as a single copy on lipoproteins (Brown & Goldstein, 1986), and apoE, a 34 kDa glycoprotein which is present as multiple copies on lipoprotein particles (Mahley, 1988). LB repeats contain a conserved negatively-charged sequence (Ser-Asp-Glu), which is required for binding of apoB-100 and apoE (Esser et al., 1988; Russell et al., 1989). It is thought that these sequences bind regions of concentrated positive charge on apoB-100 and apoE (Mahley, 1988).

The LB motif has been found in an increasing number of proteins which are functionally unrelated to the LDLR family, including clotting factors (Haefliger et al., 1987), the human heparin sulfate proteoglycan called perlecan (Noonan et al., 1991), the linker chain of earthworm hemoglobin (Suzuki et al., 1993), and a receptor for subgroup A Rous sarcoma virus (Bates et al., 1993). In each of these proteins, the LB repeat is thought to function as a protein-binding domain which interacts with Lys and Arg residues, resembling the positively charged receptor-binding regions of apoE and apoB-100.

Chemical modification studies have shown that most of the cysteines in the LDLR form disulfide bridges (Yamamoto

† This work was supported by a research grant from the National Health and Medical Research Council of Australia to P.A.K. and R.S.

* Address correspondence to this author, at the Department of Biochemistry, University of Queensland, Brisbane, Queensland 4072, Australia. Telephone: Intl 61-7-3365-4876; FAX: Intl 61-7-3365-4699; EMAIL: pkroon@biosci.uq.oz.au.

⊗ Abstract published in *Advance ACS Abstracts,* September 1, 1995.

[1] Abbreviations: LDL, low density lipoprotein; VLDL, very low density lipoprotein; LDLR, LDL receptor; LB repeat, ligand-binding repeat; LB1, amino-terminal LB repeat; rLB1, recombinant LB1; apo, apolipoprotein; RP, reverse-phase; TCEP, tris(2-carboxyethyl)phosphine; AM, acetamide; PE, pyridylethyl.

[2] Cysteine residues identified with Roman numerals (Cys(I) through Cys(VI)) refer to the order of cysteine residues in rLB1, and not their absolute positions.

0006-2960/95/0434-13059$09.00/0 

et al., 1984; Goldstein et al., 1985). In view of the conserved nature of LB repeats over more than 600 million years (Suzuki et al., 1993), it is likely that their overall structures are similar. A model consistent with current data is one in which the six cysteine residues within a repeat form intrarepeat disulfide bonds, leading to a stable compact structure.

As a first step toward understanding the structure and function of LB repeats, we have expressed the amino-terminal repeat (LB1) of the LDLR and determined its disulfide connectivity. Although LB1 is not absolutely required for the binding of ligands (Esser et al., 1988; Russell et al., 1989), it is the only repeat for which a conformation-dependent monoclonal antibody is available. IgG-C7 binds to LB1 of the native LDLR, but only in the absence of reducing agents, and in the presence of $Ca^{2+}$ (van Driel et al., 1987). Following equilibration of disulfide bonds, recombinant LB1 (rLB1) is converted to a protease-resistant peptide with three unique disulfide bonds, connecting Cys(I)−Cys(III), Cys(II)−Cys(V), and Cys(IV)−Cys(VI). This disulfide-bonding pattern results in a cluster of negatively-charged residues, including the conserved Ser-Asp-Glu sequence, in a single loop, in place for interactions with positively charged residues on apoB-100 and apoE.

## MATERIALS AND METHODS

*Materials.* Vinylpyridine, bovine thrombin, iodoacetamide, glutathione−agarose, α2-chymotrypsin, and sequencing-grade trypsin were obtained from Sigma. Proteinase K was obtained from Boehringer Mannheim (Castle Hill, NSW, Australia). Tris(2-carboxyethyl)phosphine (TCEP) was kindly provided by Alun Jones, Centre for Drug Design and Development, University of Queensland, Brisbane, Australia. IgG-C7-producing hybridoma CRL 1691 was obtained from the American Type Culture Collection (Rockville, MD). Monoclonal antibody IgG-3D1 and nonstructural protein (NS1) from dengue virus were provided by Dr. Paul Young, University of Queensland. Reagents for the gas-phase sequencer were from Applied Biosystems (Foster City, LA). Other chemicals used were the highest reagent grade available.

*Expression and Folding of rLB1.* The cDNA encoding the first cysteine-rich repeat of the ligand-binding domain of the human LDLR was PCR-amplified, digested with appropriate restriction enzymes, and cloned into the expression vector pGEX-2T (Djordjevic, unpublished experiments). Recombinant human LB1 was expressed as a thrombin-cleavable GST fusion protein in *Escherichia coli* using a 15 L Chemap fermentor (Chemap AG, Männedorf, Switzerland). Following glutathione−agarose affinity chromatography (Smith & Johnson, 1988), the glutathione−agarose-bound GST fusion protein was reduced by incubation with 1 mM DTT for 1 h at 37 °C and cleaved with bovine thrombin (5 units/mg of fusion protein) to release rLB1. We have previously shown that reduction with DTT prior to thrombin cleavage increases the efficiency with which thrombin cleaves the fusion protein (Djordjevic, unpublished experiments). rLB1 (250−400 μg/mL) was incubated overnight in the presence of 3 mM GSH/0.3 mM GSSG in 50 mM Tris-HCl, 150 mM NaCl, and 2.5 mM $CaCl_2$, pH 8.5, at 4 °C, to allow thiol−disulfide exchange. Refolded rLB1 was purified by reverse-phase (RP) HPLC as described below.

*HPLC and FPLC Analyses.* Refolded rLB1 was purified using an Econosil C18 reverse-phase column (10-μm particle size, 60-Å pore, 10 × 250 mm) obtained from Alltech (Deerfield, IL). HPLC analyses were performed on a Beckman HPLC system. The detector was coupled to an IBM-compatible XT PC using a Delta Data Systems (Digital Solutions Pty. Ltd, Brisbane, Australia) interface. Data analysis was performed using software from Delta Data Systems. A trifluoroacetic acid/acetonitrile solvent system was used (solvent A: 0.1% trifluoroacetic acid in water; solvent B: 0.1% trifluoroacetic acid in acetonitrile) for the purification, typically with a flow rate of 2.5 mL/min.

Proteinase K fragments were isolated using RP-HPLC on a Waters Radial-Pak column (10-μm particle size, 8 × 100 mm, Milford, MA) under isocratic conditions using the trifluoroacetic acid/acetonitrile solvent system described above.

FPLC analyses were performed using a Smart System (AMRAD-Pharmacia, Australia). A μRPC C2/C18 reverse-phase column (3-μm particle size, 120-Å pore, 2.1 × 100 mm) and a trifluoroacetic acid/acetonitrile buffer system (see above) were used for all FPLC analyses.

*Immunoblot Analysis.* Mouse monoclonal antibody, IgG-C7, was prepared by injecting 6−11 week old, pristane-primed Balb-C mice with IgG-C7-producing hybridoma CRL 1691 (Ausubel et al., 1993). Cells were removed from ascites fluid by centrifugation (900*g*, 15 min at room temperature).

For dot blot analysis of rLB1, the peptide was applied directly to 0.2 μm nitrocellulose membrane (Schleicher and Schuell, West Germany) using a vacuum blotting apparatus. To obtain reproducible signals, membranes were baked in a vacuum oven at 80 °C, for a minimum of 8 h, because rLB1 bound poorly to the membranes. Nitrocellulose membranes were blocked at room temperature for at least 1 h with 5% nonfat skim milk or 5% BSA, in TBS (20 mM Tris-HCl, pH 7.5, 500 mM NaCl). Subsequent antibody incubations and washes were performed with 1% nonfat skim milk or 1% BSA, in TBS which contained 0.05% Tween-20. Dot blots were incubated with either IgG-C7 or an unrelated antibody IgG-3D1, for 1 h. Prior to detection, dot blots were incubated with rabbit anti-mouse IgG, for 1 h. To visualize rLB1, dot blots were incubated with horseradish peroxidase-conjugated goat anti-rabbit IgG for 30 min. Following three washes (15 min and 2 × 5 min), nitrocellulose membranes were incubated with enhanced chemiluminescence (ECL) substrate (Amersham International, U.K.) according to the manufacturer's instructions and exposed to X-ray film (AGFA, Curix) for time periods ranging from 5 s to 10 min.

*Proteinase K Digestion.* rLB1 (0.5 mg/mL) was digested with proteinase K for 19−24 h, at 24 or 37 °C, in 25 mM ammonium carbonate buffer (pH 7.5) and 0.5% SDS with an enzyme:peptide ratio of 1:14.

*Reduction and Alkylation with Iodoacetamide.* To alkylate each of the six cysteine residues, rLB1 was reduced with 45 mM DTT in the presence of 6 M GdnHCl, for 2 h at 50 °C, and incubated with supersaturated iodoacetamide as described below. Partial reduction and alkylation of rLB1 were performed essentially as described by Gray (1993). Briefly, HPLC-purified rLB1 (2 μg in 15−25 μL of elution buffer) was mixed with an equal volume of 20 mM tris(2-carboxyethyl)phosphine (TCEP) in 20 mM ammonium carbonate, pH 3, and incubated for 7−12 min at 65 °C. The partially

reduced peptide was immediately alkylated with supersaturated iodoacetamide (approximately 2.2 M). This solution was prepared by heating 50 mg of iodoacetamide in 100 $\mu$L of 0.5 M tris acetate and 2 mM $Na_2EDTA$, pH 8, at 65 °C, until it was clear, and was used within 5 min of preparation. Partially reduced rLB1 (30−50 $\mu$L) was rapidly injected into the iodoacetamide solution with a 250 $\mu$L Hamilton syringe (22 ga PT#3). During the injection the iodoacetamide solution was vortexed vigorously. Following the injection (approximately 25 s later) the reaction was quenched with 200 $\mu$L of 0.5 M citric acid. The sample was immediately injected onto a C2/C18 RP-FPLC column on a Smart System. The column was washed with 0.1% trifluoroacetic acid/10% acetonitrile until the effluent absorbance at 214 nm dropped below 0.05. Different alkylated forms of rLB1, $AM_n$-rLB1, where $n$ corresponds to the number of acetamide (AM) groups incorporated, were separated using an 18−30% acetonitrile gradient over 24 min. The separated forms were analyzed by electrospray mass spectrometry, and 0.1−1 nmol of protein was subjected to N-terminal amino acid sequence analysis.

*Reduction and Alkylation with Vinylpyridine.* $AM_2$-rLB1 was completely reduced with DTT, and the resulting free cysteine residues were labeled with 4-vinylpyridine to yield pyridylethyl (PE) derivatives (Gray, 1993). Briefly, $AM_2$-rLB1 (0.5 $\mu$g in 50−100 $\mu$L of FPLC elution buffer) was incubated with 100 $\mu$L of 0.5 M Tris acetate, pH 8, 2 mM $Na_2EDTA$, and 20 mM DTT, at 65 °C for 15 min. Following addition of 5 $\mu$L of 4-vinylpyridine, the reaction mixture was incubated in the dark, at room temperature, for 20 min. The reaction was stopped by adding 100 $\mu$L of 0.5 M citric acid. This resulted in the formation of rLB1, which was labeled with both AM and PE groups. The peptide ($AM_2/PE_4$-rLB1) was purified by RP-FPLC using an 18−30% acetonitrile gradient over 24 min. The presence of $AM_2/PE_4$-rLB1 was confirmed by electrospray mass spectrometry.

*Tryptic Digestion of $AM_2/PE_4$-rLB1.* $AM_2/PE_4$-rLB1 was taken to dryness in a Savant SpeedVac SC100 and then dissolved in 100 $\mu$L of 50 mM $NH_4CO_3$, pH 8.5, containing 2 mM $CaCl_2$. The peptide (2−5 $\mu$g) was cleaved with trypsin using a 1:10 (w/w) ratio of enzyme to protein, for 16 h at 37 °C. Cleavage products were separated by RP-FPLC using a 1−30% acetonitrile gradient over 60 min. Individual peaks were collected using the Smart System fraction collector and analyzed by electrospray mass spectrometry.

*Mass Spectrometry.* Peptides were analyzed using a triple quadrupole mass spectrometer (SCIEX API III, Thornhill, Canada), equipped with an ionspray interface and operated in the positive detection mode. Mass spectra were analyzed using MacSpec version 3.2 software (SCIEX).

*Amino Acid Sequence Analysis.* Peptides were sequenced using automated Edman degradation on an Applied Biosystems 470A protein gas-phase sequencer with on-line PTH analysis.

## RESULTS

*Expression and Refolding of the rLB1 Domain.* LB1 was expressed as a thrombin-cleavable GST fusion protein in *E. coli*, using the expression vector pGST-LB1 (Djordjevic, unpublished experiments). This vector encodes GST, followed by the first 46 amino acids of the mature human LDL receptor, comprising the first LB repeat and the first four amino acids of the second LB repeat. These last four amino acids serve as a linker sequence between Cys(VI) of the first LB repeat and Cys(I) of the second LB repeat of the LDLR. Cleavage with thrombin releases a 48 amino acid peptide: GSAVGDRCERNEFQCQDGKCISYKWVCDG-SAECQDGSDESQETCLSVT. This recombinant peptide comprises the LDL receptor fragment with an N-terminal Gly-Ser extension derived from the thrombin cleavage site (Smith & Johnson, 1988). The identity of the peptide was confirmed by Edman sequencing. RP-HPLC analysis of the peptide gave a variable number of peaks, depending on its storage history (data not shown). Oxidative refolding of rLB1 in the presence of a redox buffer (GSH/GSSG) resulted in a single major RP-HPLC peak under isocratic conditions (Figure 1). This form of rLB1 was stable when stored at 4 °C. The conformation-specific monoclonal antibody, IgG-C7, reacted with RP-HPLC-purified rLB1 in the presence of $Ca^{2+}$ (inset in Figure 1). In contrast, no signal was observed in the absence of $Ca^{2+}$. An unrelated monoclonal antibody, IgG-3D1, made against the nonstructural glycoprotein of dengue virus, failed to react with rLB1 in the presence or absence of $Ca^{2+}$, confirming the specificity of the recognition of rLB1 by IgG-C7. Thiol−disulfide exchange therefor leads to the formation of a disulfide-bonded form for rLB1 which is immunologically indistinguishable from LB1 in the LDLR, suggesting that rLB1 is correctly folded.

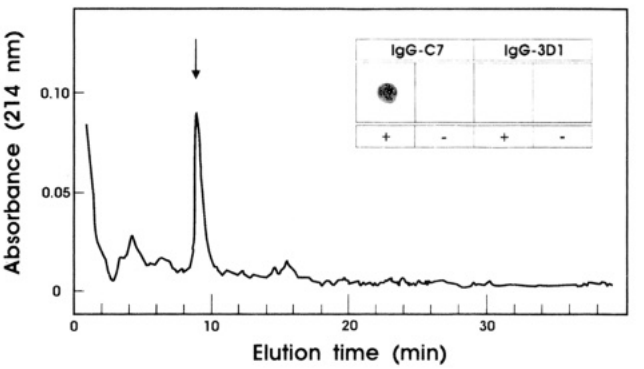



FIGURE 1: rLB1 was subjected to thiol−disulfide exchange and analyzed by RP-HPLC as described in Materials and Methods. The peptide eluted as a single peak (arrow). Inset: Dot blot analysis of RP-HPLC-purified rLB1. HPLC-purified rLB1 was dotted (4 $\mu$g/dot) onto a 0.2 $\mu$m nitrocellulose membrane. The membrane was baked overnight in a vacuum oven, at 80 °C, and blocked for 1 h in TBS, which contained 5% nonfat skim milk. Membranes were incubated with IgG-C7, or with IgG-3D1, in the presence (+) or in the absence (−) of 2 mM $Ca^{2+}$. Incubations in the absence of $Ca^{2+}$ also contained 30 mM EDTA. Signals were detected by enhanced chemiluminescence followed by exposure to X-ray film.

*Analysis of the Refolded rLB1 Domain.* Mass spectral analysis of rLB1 gave a molecular mass of 5220 Da (variation of ±1 Da), consistent with the presence of three disulfide bridges. To confirm the complete absence of free cysteine thiol groups, rLB1 was incubated in the presence of 6 M GdnHCl for 2 h at 50 °C, to expose thiol groups which might be buried, followed by incubation with supersaturated iodoacetamide to alkylate these groups. The data in Figure 2A show that the molecular mass of rLB1 remained 5220 Da, with no evidence for species with alkylated cysteine residues. In contrast, when rLB1 was first incubated with DTT, the molecular mass increased to 5569 Da (Figure 2B), consistent with the presence of six alkylated cysteine

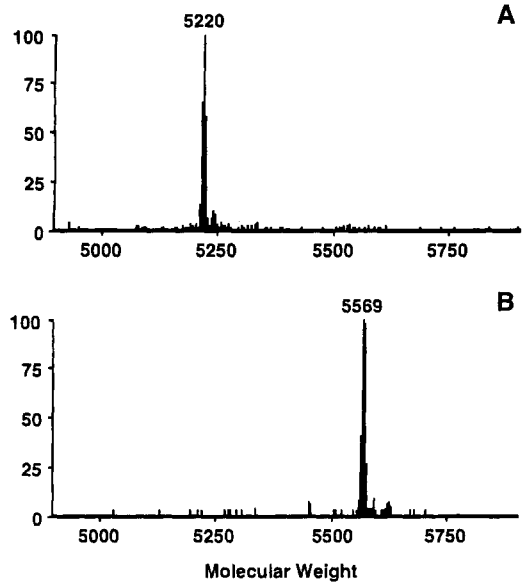



FIGURE 2: Mass spectral analysis of rLB1 for the presence of alkylated thiol groups. rLB1 was incubated in 6 M GdnHCl for 2 h at 50 °C in the absence (A) and presence (B) of 45 mM DTT and injected into supersaturated iodoacetamide as described in Materials and Methods. The treated peptides were immediately purified by RP-FPLC and analyzed by mass spectrometry.

residues. We conclude that the refolded rLB1 domain is completely oxidized and contains no free cysteine thiol groups.

*Arrangement of Disulfide Bridges of the rLB1 Domain. (A) Proteolytic Digestion.* The traditional strategy for analyzing disulfide connectivities is to use proteolytic enzymes to cleave the peptide chain and to isolate bridged fragments. However, rLB1 in its oxidized form was resistant to cleavage by most commonly used sequence-specific proteases. Trypsin and chymotrypsin (24 h, at 37 °C) at a 1:10 enzyme:rLB1 ratio, in the presence of 0.5 M urea or 0.1% SDS, did not cleave rLB1. Proteinase K, a relatively non-specific protease, did cleave rLB1. Incubation with proteinase K for 19−24 h, at 24 or 37 °C, in 25 mM ammonium carbonate buffer (pH 7.5) and 0.5% SDS, with an enzyme:peptide ratio of 1:14, resulted in significant fragmentation. One fragment isolated under isocratic conditions (18% acetonitrile) had a molecular mass of 898 Da. Sequence analysis showed that it contained two amino termini, corresponding to fragments Trp25-Gly29 and Cys44-Ser46. These data demonstrate that Cys(IV) and Cys(VI) in rLB1 are disulfide-bridged. No other proteinase K-generated fragments were informative.

*(B) Partial Reduction with TCEP.* An alternate approach for analyzing disulfide connectivities is to reduce individual disulfide bonds at acidic pH with reagents such as TCEP (Gray, 1993), leaving the remaining chain intact. The schematic in Figure 3 summarizes this process and shows each of the possible species formed. Reaction of rLB1 with TCEP at 65 °C resulted in a time-dependent reduction of disulfide bonds, which were then rapidly alkylated with iodoacetamide. A total of eight partially reduced and AM-labeled species were identified by RP-FPLC (peaks 1−8 in Figure 4) and mass spectral analysis. Peptides could not be detected in any of the other peaks. Peptides 1−8 (derived from peaks 1−8) contained 0, 2, 4, or 6 AM groups (Table 1), consistent with precursor peptides in which 0, 1, 2, or 3

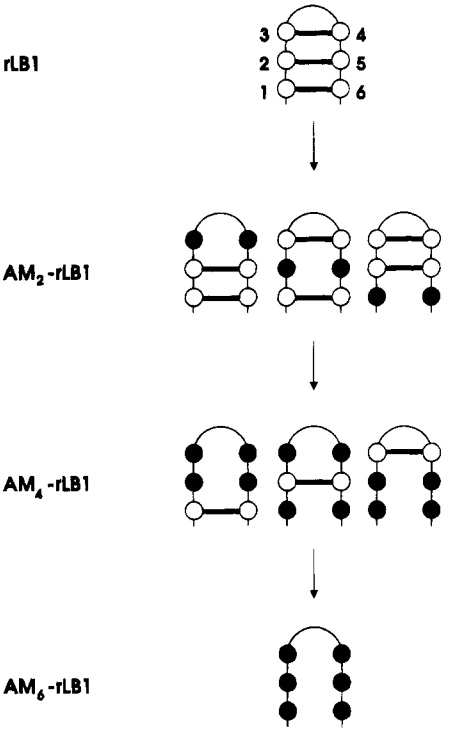



FIGURE 3: Products formed during the sequential reduction and alkylation of rLB1. (O) Disulfide-bonded cysteine residues; (●) alkylated cysteine residues formed by the reduction of a disulfide bond and alkylation of the resulting free thiol groups. The nomenclature for the partially reduced and alkylated forms of rLB1 is described in the text.

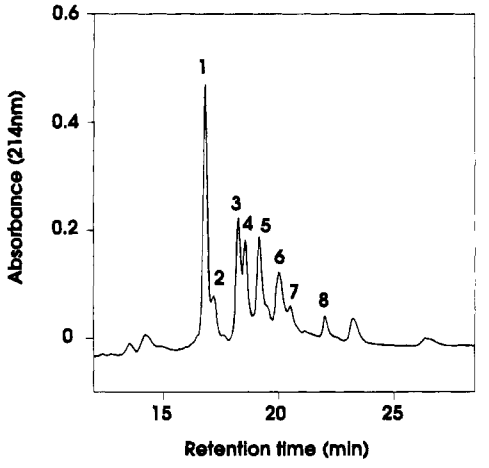



FIGURE 4: RP-FPLC separation of partially reduced and alkylated rLB1. Disulfide bonds were partially reduced with TCEP (10 min, 65 °C) and the resulting free thiols alkylated by injection into 2.2 M iodoacetamide as described in Materials and Methods. The resulting peptides were separated on a C2/C18 RP-FPLC column, using a linear (18−30%) gradient of acetonitrile.

disulfide bonds were reduced, respectively. These represent the full complement of possible species that can be formed (Figure 3).

Peaks 1, 2, and 8 contained peptides with 2 AM groups each. Sequence analysis of peptide 1 through to residue 21 showed that the first three cysteine residues were not alkylated (data not shown). In contrast, sequence analysis of peptide 8 showed that both Cys(I) and Cys(III) were alkylated, providing direct evidence for a Cys(I)−Cys(III) disulfide bridge. Because of the proximity of peak 2 to peak 1, and because of its low abundance, we did not attempt to

Table 1: Analysis of Peptides Formed from Partial Reduction and Alkylation of rLB1[a]

| peptide[b] | $M_r$ | $N^c$ | peptide[b] | $M_r$ | $N^c$ |
|---|---|---|---|---|---|
| 1 | 5337 | 2 | 5 | 5568 | 6 |
| 2 | 5337 | 2 | 6 | 5453 | 4 |
| 3 | 5220 | 0 | 7 | 5452 | 4 |
| 4 | 5453 | 4 | 8 | 5335 | 2 |

[a] Peptides were produced by partial reduction and alkylation of rLB1, purified by RP-FPLC as described in Figure 4, and analyzed by mass spectrometry. [b] Peptide numbers correspond to peaks in Figure 4. Peptide peaks were combined from several RP-FPLC runs. [c] The number of alkylated cysteine residues ($N$) in each of the isolated components was calculated using $M_r$ = 5220 for the completely oxidized peptide (peptide 3), and a mass of 58 Da for the AM group.

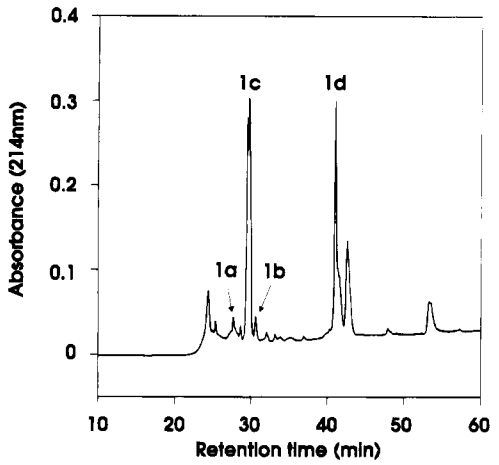



FIGURE 5: RP-FPLC separation of a tryptic digest of PE-labeled peptide 1. Peptide 1 was isolated as described in the legend for Figure 4 and was reduced with DTT and alkylated with vinylpyridine to yield the doubly labeled peptide $AM_2/PE_4$-rLB1. This peptide was digested with trypsin (1:10 w/w) for 16 h at 37 °C. Cleavage products were separated by RP-FPLC using a 1−30% acetonitrile gradient over 60 min.

isolate and sequence peptide 2.

The fact that none of the cysteine residues in the first 21 amino acids of peptide 1 were alkylated was consistent with a Cys(IV)−Cys(VI) disulfide bond. To confirm this, peptide 1 was completely reduced in the presence of DTT and reacted with vinylpyridine to label each of the four cysteine thiol groups formed with a PE group. The resulting peptide ($AM_2/PE_4$-rLB1) was cleaved with trypsin and analyzed by RP-FPLC (Figure 5). Mass spectral analysis of each of the peaks showed that only peaks 1a−1d contained peptides within the size range predicted. Mass spectra of peptides 1a−1d are shown in Figure 6. Although peak 1c is a doublet in the RP-FPLC profile, only a single fragment was detected by mass spectral analysis within the predicted size range. The location of tryptic sites relative to each of the cysteine residues of rLB1 is illustrated schematically in Figure 7, together with the calculated masses for unmodified tryptic peptides, and for peptides modified with all possible combinations of AM and PE groups. From a comparison of the molecular masses of peptides 1a−1c with the predicted masses of AM and PE-labeled fragments, we conclude that Cys(I), Cys(II), and Cys(III) are contained within peptides 1a, 1b, and 1c, respectively, and that each of these is labeled with PE. Furthermore, we conclude that peptide 1d contains cysteine residues Cys(IV), Cys(V), and Cys(VI), of which two are labeled with AM groups and one with a PE group. Amino acid sequence analysis of the first 10 residues of this peptide (W25−Q34) showed that Cys(IV) was labeled with AM and that Cys(V) was labeled with PE groups. Since peptide 1d contained two AM-labeled cysteine residues, we infer that both Cys(IV) and Cys(VI) were AM-labeled and that these cysteine residues were disulfide-bonded in the oxidized peptide. This is in agreement with the Cys(IV)−Cys(VI) disulfide bond determined by proteinase K digestion. In view of the Cys(I)−Cys(III) and Cys(IV)−Cys(VI) connectivities, we infer that the remaining disulfide bond bridges Cys(II) and Cys(V). The resulting overlapping disulfide bonds in rLB1 are illustrated in Figure 8A.

## DISCUSSION

In this paper we report the disulfide-bonding pattern for the amino-terminal LB repeat of the human LDLR. This repeat was expressed as a GST fusion protein and cleaved with thrombin to release rLB1. Oxidative refolding of rLB1 in the presence of a redox buffer yielded a unique state, the

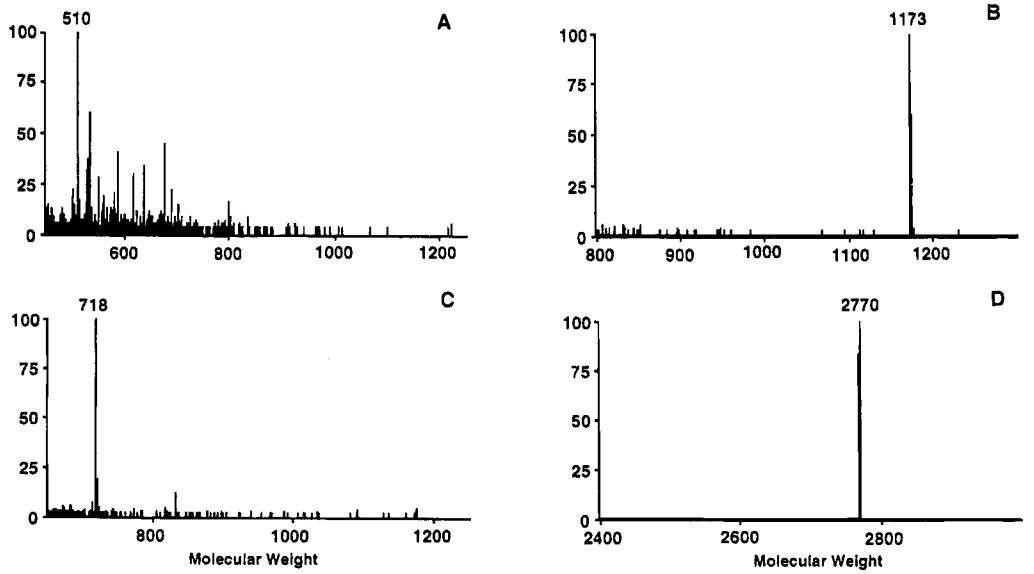
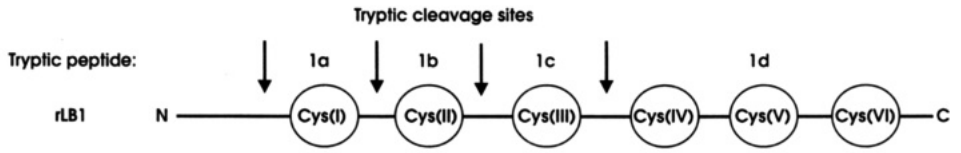



FIGURE 6: Mass spectral analysis of peptides formed by tryptic digestion of $AM_2/PE_4$-rLB1. Tryptic peptides were isolated as described in the legend for Figure 5 and analyzed by mass spectrometry. Results for peptides 1a, 1b, 1c, and 1d are shown in panels A to D, respectively.




| $M_r$ | 660 | 406 | 1069 | 612 |
|---|---|---|---|---|
| + AM | | 464 | 1127 | 670 |
| + PE | | 511 | 1174 | 717 |

| AM/PE | $M_r$ |
|---|---|
| 0/0 | 2550 |
| 3/0 | 2724 |
| 2/1 | 2771 |
| 1/2 | 2818 |
| 0/3 | 2865 |

FIGURE 7: Identification of peptides formed by tryptic digestion of $AM_2/PE_4$-rLB1. A schematic representation of rLB1, its tryptic cleavage sites (arrows), the location of cysteine residues relative to these sites, and peptides formed as a result of cleavage (1a−1d) are shown at the top of the figure. The molecular masses of each of the unmodified tryptic peptides, as well as the predicted masses for all combinations of AM/PE-modified fragments, are shown below this representation of the tryptic peptides. Shaded boxes contain molecular masses which agree with the experimentally determined values from Figure 6.

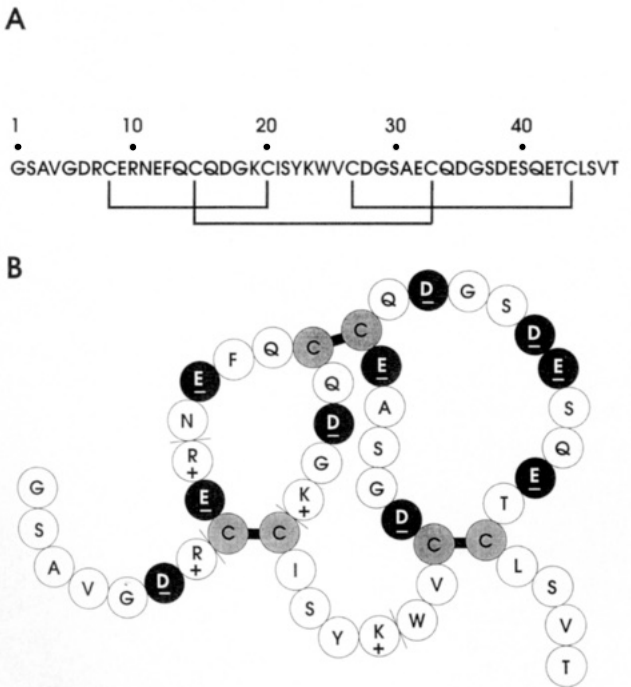



FIGURE 8: Structural models for rLB1. (A) A linear representation of the three overlapping disulfide bridges; (B) a model which illustrates the tricyclic structure of the disulfide-bridged peptide, and the concentration of acidic residues in the last of the loops. (−) indicates the location of tryptic cleavage sites.

presumed thermodynamic ground state, in which each of the cysteine residues were disulfide-bonded. In this state, it reacted with IgG-C7, a conformation-specific monoclonal antibody which binds LB1 of the native LDLR in the presence of $Ca^{2+}$ (van Driel et al., 1987). In the absence of any knowledge about the structure of LB1, recognition by IgG-C7 is currently the only way in which the conformation of the expressed repeat can be related to that in the LDLR. The fact that IgG-C7 reacts with rLB1 in the presence and not in the absence of $Ca^{2+}$ suggests that rLB1 is correctly folded.

We used a strategy based on the sequential reduction of disulfide bonds in conjunction with proteolytic enzymes to show that Cys(I)−Cys(III), Cys(II)−Cys(V), and Cys(IV)−Cys(VI) are connected by disulfide bonds in the folded peptide. In a separate study we have determined the three-dimensional structure of rLB1 by NMR spectroscopy (Daly et al., 1995). The resulting structure is consistent with the presence of these disulfide bonds. The overlapping disulfide bridges (Figure 8A) result in a constrained structure which is resistant to proteolytic digestion. The presence of disulfide bridges in LB1 and each of the other LB repeats is likely to be responsible for the extreme stability of the LDL receptor, which can be boiled in SDS and retain its ability to bind lipoproteins as long as its disulfide bonds remain intact (Brown & Goldstein, 1986). To our knowledge, this is the first description of the disulfide bridging pattern for an LB repeat. It is distinct from the disulfide bond connectivity for EGF-like domains (Cooke et al., 1987), insulin-like growth factor (Narhi et al., 1993), wheat germ agglutinin (Wright et al., 1984), neuronal venoms in several cone shells (Olivera et al., 1990), and charybdotoxin, a neurotoxin from the venom of buthid scorpions (Massefski et al., 1990). The occurrence of LB motifs in the LDL receptor of man and *Xenopus laevis* (Mehta et al., 1991), the VLDL receptor (Takahashi et al., 1992), the LDL receptor-related protein/α2-macroglobulin receptor (Herz et al., 1988), renal glycoprotein gp330 (Raychowdhury et al., 1989), complement factors (Haefliger et al., 1987), the proteoglycan perlecan (Noonan et al., 1991), a receptor for subgroup A Rous sarcoma virus (Bates et al., 1993), and the linker chain of the hemoglobin of *Lumbricus terrestris* (Suzuki et al., 1993) suggests a common origin for these motifs, dating back at least 600 million years (Suzuki et al., 1993). Südhof et al. (1985) have attributed the reuse of LB motifs in different proteins to the process of exon shuffling. In view of the highly conserved nature of this ancient family of motifs, the disulfide bond connectivity of rLB1 serves as a paradigm for other members of the LB family.

Figure 8B shows a structural model for the disulfide-bonded form of rLB1, which illustrates the concentration of acidic residues, including the highly conserved sequence Ser-Asp-Glu, in the last loop of the tricyclic structure. These charged residues have been identified as central to the function of LB repeats in the binding of ligands (Esser et al., 1988; Russell et al., 1989). However, despite their highly conserved sequences, individual LB repeats have distinct

functions. This has been demonstrated most clearly for the LDL receptor, whose ligands, apoE and apoB-100, require different combinations of the seven LB modules of the binding domain to achieve high affinity binding (Esser et al., 1988; Russell et al., 1989). Functional diversity of LB repeats is likely to be provided by nonconserved residues, which may lead to differences in the arrangement and accessibility of acidic residues which mediate ligand binding. We note that while studies of the function of LB repeats of the LDLR have focused primarily on intermolecular interactions between the LDLR and its lipoprotein ligands, other roles may involve intramolecular protein−protein interactions. The release of receptor-bound lipoproteins within the acidic environment of the endosome is one step which is likely to involve such interactions. This step is crucial for recycling receptors like the LDLR and is known to require segments of the EGF homology domain (Davis et al., 1987).

Because rLB1 was resistant to proteolytic digestion by most common specific proteases, a combination of different approaches was needed to delineate the disulfide-bonding pattern. Digestion with the nonspecific enzyme proteinase K resulted in the formation of one fragment which contained a single disulfide bond. Amino acid sequence analysis of the purified peptide was used to identify the Cys(IV)−Cys(VI) disulfide bridge. The second approach we used was based on a novel method described by Gray (1993), in which disulfide bonds are sequentially reduced with an organic phosphine, TCEP. The success of this approach can be attributed to the low pH of the reaction mixture (pH 3) which suppresses scrambling of disulfide bonds in partially reduced peptides (Gray, 1993). At this pH, cysteine thiol groups largely exist in the protonated (RSH) form, and not as thiolate anions ($RS^-$), the form required for disulfide exchange (Ryle & Sanger, 1955). However, the basic pH (pH 8) during the subsequent alkylation of cysteine residues does allow such exchange to occur. This was minimized by injecting the partially reduced peptide directly into a supersaturated solution of iodoacetamide (Gray, 1993), thus ensuring that the peptide was only exposed to basic pH in the presence of high concentrations of the reactive alkylating agent iodoacetamide. A diagnostic test for scrambling of disulfide bond is the appearance of more partially reduced and alkylated species than can be accounted for theoretically. In the current study, no more than the eight theoretically possible forms (Figure 3) were identified (peptides 1−8), suggesting that scrambling had not occurred.

We used the sequential reduction approach to identify two disulfide bonds, to ensure that this method was consistent with the more classical approach used for the first disulfide bond. We analyzed two peptides formed by the alkylation of cysteine residues derived from a single reduced disulfide bond (Table 1). For the first (peptide 8), direct sequence analysis showed that both Cys(I) and Cys(III) were alkylated with AM, consistent with a Cys(I)−Cys(III) disulfide bond. The absence of alkylated cysteine residues in the first 21 amino acids of peptide 1 was consistent with a Cys(IV)−Cys(VI) bond. The presence of this bond was demonstrated directly using a more complex strategy in which peptide 1 was completely reduced and alkylated with vinylpyridine to label the four free thiol groups. The resulting peptide was digested with trypsin, leading to five fragments, one of which contained two AM-labeled cysteine residues and one PE-labeled cysteine residue (Figure 7). Sequence analysis of this peptide showed that both Cys(IV) and Cys(VI) were alkylated with AM, confirming the presence of the Cys(IV)−Cys(VI) disulfide bond. The independent identification of the Cys(IV)−Cys(VI) disulfide bond in a proteinase K fragment and in a fragment produced by partial reduction with TCEP supports our earlier supposition that scrambling of disulfide bonds was not significant under the experimental conditions used.

In conclusion, this study has provided a model for disulfide bonding of LB-type repeats and is a first step toward a detailed delineation of their structure, and of their role(s) in mediating intermolecular protein−protein interactions between lipoprotein receptors and their ligands.

## REFERENCES


Ausubel, F. M., Brent, R., Kingston, R. E., Moore, D. D., Seidman, J. G., Smith, J. A., & Struhl, K. (1993) in *Current Protocols in Molecular Biology*, John Wiley and Sons, New York.

Bates, P., Young, J. A. T., & Varmus, H. E. (1993) *Cell 74*, 1043−1051.

Brown, M. S., & Goldstein, J. L. (1986) *Science 232*, 34−47.

Cooke, R. M., Wilkinson, A. J., Baron, M., Pastore, A., Tappin, M. J., Campbell, I. D., Gregory, H., & Sheard, B. (1987) *Nature 327*, 339−341.

Daly, N. L., Scanlon, M. J., Djordjevic, J. T., Kroor, P. A., & Smith, R. (1995) *Proc. Natl. Acad. Sci. U.S.A. 92*, 6334−6338.

Davis, C. G., Goldstein, J. L., Südhof, T. C., Anderson, R. G. W., Russell, D. W., & Brown, M. S. (1987) *Nature 326*, 760−765.

Esser, V., Limbird, L. E., Brown, M. S., Goldstein, J. L., & Russell, D. W. (1988) *J. Biol. Chem. 263*, 13282−13290.

Goldstein, J. L., Brown, M. S., Anderson, R. G. W., Russell, D. W., & Schneider, W. J. (1985) *Annu. Rev. Cell Biol. 1*, 1−39.

Gray, W. R. (1993) *Protein Sci. 2*, 1732−1748.

Haefliger, J. A., Tschopp, J., Nardelli, D., Wahli, W., Kocher, H. P., Tosi, M., & Stanley, K. K. (1987) *Biochemistry 26*, 3551−3556.

Herz, J., Hamann, U., Rogne, S., Myklebost, O., Gausepohl, H., & Stanley, K. K. (1988) *EMBO J. 7*, 4119−4127.

Hobbs, H. H., Russell, D. W., Brown, M. S., & Goldstein, J. L. (1990) *Annu. Rev. Genet. 24*, 133−170.

Mahley, R. W. (1988) *Science 240*, 622−630.

Massefski, W., Jr., Redfield, A. G., Hare, D. D., & Miller, C. (1990) *Science 249*, 521−524.

Mehta, K. D., Chen, W. J., Goldstein, J. L., & Brown, M. S. (1991) *J. Biol. Chem. 266*, 10406−10414.

Narhi, L. O., Hua, Q., Arakawa, T., Fox, G. M., Tsai, L., Rosenfeld, R., Holst, P., Miller, J. A., & Weiss, M. A. (1993) *Biochemistry 32*, 5214−5221.

Noonan, D. M., Fulle, A., Valente, P., Cai, S., Horigan, E., Sasaki, M., Yamada, Y., & Hassell, J. R. (1991) *J. Biol. Chem. 266*, 22939−22947.

Olivera, B. M., Rivier, J., Clark, C., Ramilo, C. A., Corpuz, G. P., Abogadie, F. C., Mena, E., Woodward, S. R., Hillyard, D. R., & Cruz, L. J. (1990) *Science 249*, 257−263.

Raychowdhury, R., Niles, J. L., McCluskey, R. T., & Smith, J. A. (1989) *Science 244*, 1163−1165.

Russell, D. W., Brown, M. S., & Goldstein, J. L. (1989) *J. Biol. Chem. 264*, 21682−21688.

Ryle, A. P., & Sanger, F. (1955) *Biochem. J. 60*, 535−540.

Smith, D. B., & Johnson, K. S. (1988) *Gene 67*, 31−40.

Südhof, T. C., Goldstein, J. L., Brown, M. S., & Russell, D. W. (1985) *Science 228*, 815−822.

Suzuki, T., & Riggs, A. F. (1993) *J. Biol. Chem. 268*, 13548−13555.

Takahashi, S., Kawarabayasi, Y., Nakai, T., Sakai, J., & Yamamoto, T. (1992) *Proc. Natl. Acad. Sci. U.S.A. 89*, 9252−9256.

van Driel, I. R., Goldstein, J. L., Südhof, T. C., & Brown, M. S. (1987) *J. Biol. Chem. 262*, 17443−17449.

Wright, C. S., Gavilanes, F., & Peterson, D. L. (1984) *Biochemistry 23*, 280−287.

Yamamoto, T., Davis, C. G., Brown, M. S., Schneider, W. J., Casey, M. L., Goldstein, J. L., & Russell, D. W. (1984) *Cell 39*, 27−38.


BI9504973