β-Sheet secondary structure of an LDL receptor domain from complement factor I by consensus structure predictions and spectroscopy

Christopher G. Ullman[a], Parvez I. Haris[a], Kathryn F. Smith[a], Robert B. Sim[b], Vincent C. Emery[c], Stephen J. Perkins[a,*]

[a]*Department of Biochemistry and Molecular Biology, University of Oxford, South Parks Road, Oxford OX1 3QU, UK*
[b]*MRC Immunochemistry Unit, Department of Biochemistry, University of Oxford, South Parks Road, Oxford OX1 3QU, UK*
[c]*Department of Virology, Royal Free Hospital School of Medicine, Rowland Hill Street, London NW3 2PF, UK*

Received 11 July 1995; revised version received 25 July 1995

Abstract Low density lipoprotein receptor domains (LDLrs) represent a large cell surface receptor superfamily of consensus length 39 residues. Alignment of 194 sequences indicated highly conserved Cys and Asp/Glu residues, and a consensus secondary structure with three β-strands was predicted. Sequence threading against known protein folds indicated consistency with small β-sheet proteins. Complement factor I contains two LDLrs, and the second of these was successfully expressed using a bacterial pGEX system. FT-IR spectroscopy on this indicated a small amount of β-sheet together with turns and loops. LDLr is proposed to have a β-sheet structure in which the five biologically important Asp/Glu residues are located on an exposed loop.

Key words: LDLr domain; FT-IR spectroscopy; Secondary structure prediction; Protein fold recognition; β-Sheet protein

1. Introduction

The low density lipoprotein receptor domain (LDLr) is a cysteine-rich repeat of approximately 40 amino acids. This was first identified in the low density lipoprotein cell surface receptor protein [1,2]. LDLr is also found in factor I [3,4] (Fig. 1), C6, C7, C8 and C9 of complement, and in members of the α_2M receptor family protein [2]. LDLr occurs as groups of repeats in the cell surface receptors, but as single or double repeats in the complement proteins. The domain is encoded by individual exons. Mutations in the LDLr allele are associated with familial hypercholesteremia, a disease which leads to premature atherosclerosis.

LDLr is characterised by six Cys residues and a conserved cluster of acidic residues. The disulphide bridge pairings and the atomic structure are unknown. The complement LDLrs have an unknown function, although in C9 they may aid formation of the membrane attack complex. C9 may be both a ligand and a receptor during its polymerisation, as a region within C9 is similar to the apoprotein E binding site for the LDL receptor [5]. In the cell surface receptors, the domain is involved in ligand binding. The conserved acidic residues interact electrostatically with basic regions in apoproteins B and E of VLDL and LDL, and are thought to bind to regions rich in basic amino acids in other ligands [2]. LDLr-3 to LDLr-7 in the LDL receptor are essential for apoprotein B binding, and LDLr-5 is essential for apoprotein E binding. In the α_2M receptor, a fragment with eight LDLr domains binds to α_2M.

A powerful strategy to analyse protein structures has been developed, based on the joint use of (a) averaged secondary structure predictions to define structural elements, (b) sequence threading against known protein folds to identify related protein folds, and (c) spectroscopy to identify the secondary structure [6,7]. By this approach, we show that LDLr contains turns and loops with a small amount of β-strand, and propose that the biologically important Asp/Glu residues are located on a surface loop.

2. Materials and methods

2.1. Alignments and structure predictions

A total of 194 LDLr sequences were extracted from the ENTREZ CD-ROM database (Release 14, December 1994). These were initially aligned using the program MULTAL [8], and adjusted manually to maximise the alignment of conserved or chemically similar residues [9]. These residues were defined as follows: tiny, G, A, S; aliphatic, I, L, V, M; aromatic, H, F, Y, W; positive, R, H, K; negative, D, E; hydroxyl, S, T; amide, N, Q [10]. Averaged structure predictions were calculated from the sequence alignments using the Chou-Fasman and β-biased GOR I and GOR III statistical methods [11–14], and the PHD neural network method [15]. Hydropathy analyses were based on summation of the total of hydrophobic or hydrophilic residues at each position [16], defined on the basis of consensus values [17]. Solvent accessibilities were predicted by the PHD neural network method [18]. The LDLr sequences were subjected to summation of pairwise energy and solvation terms to score their similarity with known protein folds using the program THREADER [19].

2.2. Cloning of the LDLr2 coding sequence from factor I

The LDLr2 coding sequence of factor I (residues 240–276 [3]) was amplified from a cDNA clone [3] by PCR at an annealing temperature of 42°C with the following primers: 5′ oligonucleotide GCA TGC CAA GGC AAA GGC TTC, 3′ oligonucleotide CGG ATC CTA TGC ACA GCC AAC TTC ATC TTC. The underlined region in the 3′ oligonucleotide corresponds to a *Bam*HI restriction site used to identify positive clones and the region shown in bold corresponds to a translation termination codon. The PCR product was ligated into the fusion protein expression vector pGEX-3X (Pharmacia), which had been digested with *Sma*I, and transformed into the *E. coli* strain JM109. Recombinant colonies were confirmed by PCR, digestion of plasmid DNA with *Bam*HI and DNA sequencing using the USB Sequenase enzyme (Version 2.0) and the PCR primers.

An overnight bacterial culture was diluted 1:10 in fresh LB-medium and protein expression was induced with 0.1 mM IPTG when the OD_{600} of the culture was 0.6. The culture was incubated for a further three hours and pelletted. The cells were resuspended in 1/100 of the starting volume in lysis buffer (PBS containing lysozyme 0.2 mg/ml, DNAse 1 μg/ml, Triton X-100 0.1% (v/v), PEFABLOC-SC (Pentapharm, Basel)

*Corresponding author. Fax: (44) (171) 794-9645.
E-mail: steve@rfhsm.ac.uk

Abbreviations: LDLr2, second low density lipoprotein receptor domain of factor I; FT-IR, Fourier transform infrared; CD, circular dichroism; NMR, nuclear magnetic resonance.

0014-5793/95/$9.50 © 1995 Federation of European Biochemical Societies. All rights reserved.
SSDI 0014-5793(95)00916-7

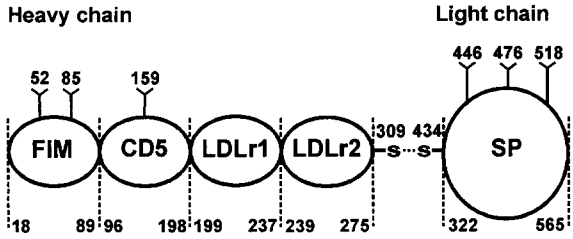


Fig. 1. Domain structure of human factor I. The heavy chain contains the factor I, C6, C7-specific (FIM) domain, the CD5 domain, and two low density lipoprotein (LDLr) domains, numbered as in [5], and identified in [6]. The light chain contains the serine protease (SP) domain. Putative carbohydrate chains and the disulphide link between the light and heavy chains are also indicated.

1 mM, benzamidine 50 mM, EDTA 5 mM, PBS is NaCl 150 mM, Na_2HPO_4 54 mM, NaH_2PO_4 16 mM, pH 7.2) for 1 h followed by sonication 3× 20 s on ice using an MSE sonicator. The lysate was clarified at 32,500 × *g* at 4°C for 20 min. The supernatant was passed through a 2.5 × 10 cm column containing 50 ml glutathione agarose (Sigma), equilibrated in PBS. The column was washed with 250 ml Dulbecco A PBS (Oxoid) containing 1 mM EDTA pH 8.0 and the fusion protein was eluted with 150 ml Dulbecco A PBS containing 10 mM reduced glutathione and 1 mM EDTA pH 8.0. The protein was dialysed against 50 mM Tris-HCl, 100 mM NaCl, pH 8.2 and cleaved with human Factor Xa (Boerhinger Mannheim) (1% w/w fusion protein) for 6 h at 37°C. Trifluoroacetic acid (0.1% v/v) was added to the cleaved GST:LDLr2 fusion protein and the proteins were separated by reversed phase chromatography on a 3 ml Resource RPC column (Pharmacia) using a linear 0–99.9% acetonitrile gradient in 0.1% TFA. LDLr2 eluted at approximately 25% acetonitrile. The domain was lyophilised and resuspended in 1 ml PBS (Sigma D5773) and further purified through a 1.6 × 100 cm Superdex 30 prep grade (Pharmacia) column. The domain was concentrated by ultrafiltration using a 10 ml Amicon stirred cell and a YM3 membrane (Amicon).

A polyclonal antiserum was raised to GST:LDLr2 in rabbits, and immunoglobulin was purified by triple 16% w/v sodium sulphate precipitation [20]. ELISAs using GST:LDLr2 antibodies were performed with factor I (prepared as in [21]) or an irrelevant GST construct bound to microtitre plates. The GST:LDLr2 antibodies were also tested against reduced and unreduced factor I in Western blot analysis.

2.3. Spectroscopic studies of LDLr2 and factor I

Fourier transform infrared (FT-IR) spectra were recorded and analysed on a Perkin-Elmer 1750 FT-IR spectrometer as described elsewhere [6]. Fluorescence emission spectra were recorded with 50 μg/ml LDLr2 in 10 mM potassium phosphate buffer, pH 7.0 and 0–6 M guanidine-HCl at 20°C, using a SPEX Fluoromax spectrofluorometer with 1 cm pathlength quartz microcells. The excitation frequency was set at 275 nm for maximal excitation of the Tyr residue in LDLr2. Emission spectra were recorded in the range 290–350 nm for each guanidine concentration, together with a buffer baseline to correct for the water Raman peak and background scattering. Circular dichroism (CD) spectroscopy was performed with 0.4 mg/ml LDLr2 in 10 mM potassium phosphate, pH 7.0 between 5–85°C using a Jobin-Yvon CD6 spectropolarimeter with 0.5 mm pathlength quartz cells. The instrument was routinely calibrated with an aqueous solution of recrystallised D10-camphosulphonic acid ($\theta^{0.1\%,1\ cm}$ = 0.308 at 290 nm).

3. Results and discussion

3.1. Sequence alignment and secondary structure prediction of the LDLr superfamily

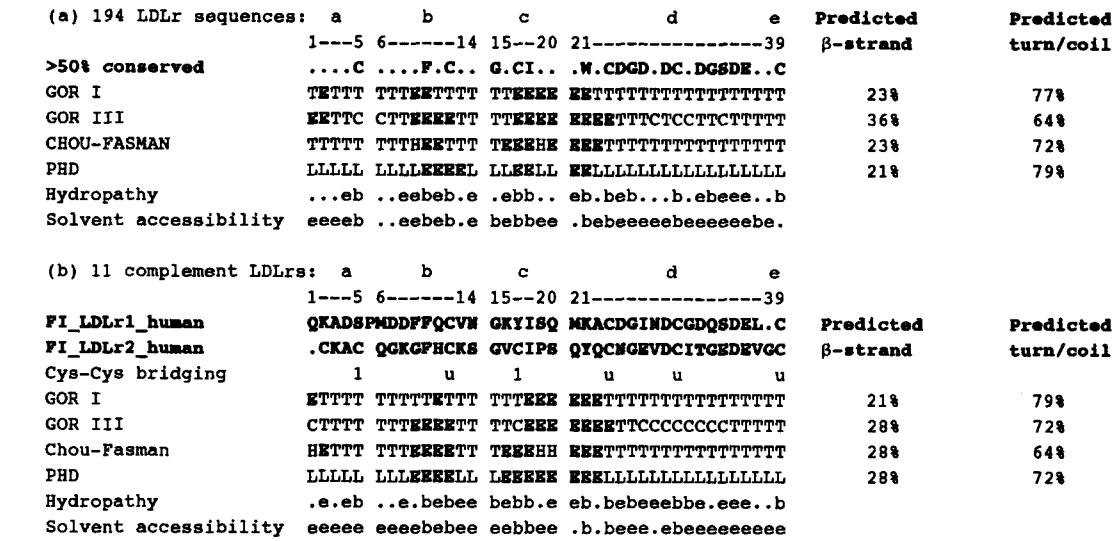
A total of 194 LDLr sequences originate from five protein groups, i.e. the complement LDLrs (7 proteins), together with LDL receptors (7 proteins), LDL receptor-related proteins (3 proteins), VLDL receptors (2 proteins), and other cell surface receptor proteins (4 proteins). The consensus alignment of Fig. 2 showed that LDLr is typically 39 residues long [9]. It is readily separated into five segments a to e, between which there are insertions that are most likely to be involved with loop structures. Three insertions between a–b, b–c and within e are more variable in length than those between c–d and d–e, which are caused by insertions in only 2 and 3 sequences out of 194. The alignment showed that 19 positions are conserved to better than 50% of the 194 sequences and 11 are conserved to better than 90% [9]. Cys-5, -12, -17, -24, -30 and -39 are conserved in 98–100% of the sequences, with the notable exception of LDLr1 of factor I in which Cys-5 and Cys-17 are omitted (and therefore likely to be disulphide linked). High conservation is also seen for Asp/Glu-36 (100% conserved), Asp/Glu-35 (98% conserved), Asp/Glu-29 (93% conserved), Asp/Glu-32 (88% conserved), Asp/Glu-25 (81% conserved) and Asp/Glu-27 (55% conserved). The residues ^{29}DCxDGSDE^{36} are implicated in ligand binding [22], and in the LDL receptor such residues interact electrostatically with the positively charged apoprotein B and apoprotein E [2].

Averaged secondary structure predictions were calculated for all 194 sequences (Fig. 3) and for each of the five LDLr groups to show very similar outcomes. Three β-strands each of length 2–4 residues were well-predicted by all four predictive methods, with one in each of the sequence segments b, c and d (Fig. 2). It is notable that each predicted β-strand is flanked by a sequence insertion (Fig. 3), as desired, and that a conserved hydrophobic residue occurs in each of the first and second β-strands at positions 10 (aromatic: 90% conserved) and 18 (aliphatic: 93% conserved), respectively. The GOR I/III, Chou-Fasman and PHD methods predict 21–36% β-strand and 64–79% turn/coil structure. The Chou-Fasman method occasionally predicted single or paired α-helical residues. As four such residues are required for a complete turn of an α-helix, these were disregarded. The predictions show that the five highly-conserved Asp/Glu residues all occur in a C-terminal region of turn or loop. The structure of this loop region would be stabilised by disulphide bridges formed by Cys-30 and Cys-39 and their partners. Conservation is high within this loop, as Asp rarely replaces Glu-36, and Glu rarely replaces Asp-25, -29, -32 and -35. Calculation of averaged hydropathies or solvent accessibilities gave similar results and show that this loop region has a high exposure to solvent (Fig. 3).

→

Fig. 2. Alignment of 194 LDLr domain sequences. For reason of brevity, mostly human sequences are shown. The alignment was based on 11 sequences from the complement proteins (FI_, C6_, C7_, C8a_ and C8b_, C9_), 49 from LDL receptors (ldlr_), 96 from LDL receptor-related proteins (lrp_), 16 from very low density LDL receptors (vldlr_), 13 from gp330 (gp330_), 8 from heparan sulphate proteoglycan (PERLEC_) and 1 from the Japanese quail LRP (RSV_REC_). Further sequences were taken from the following organisms: *C. elegans* (_celeg), Japanese quail, hamster, mouse, rabbit, rat (_rat) and *X. laevis*. Consensus structure predictions were based on residue positions occupied in over 50% of the LDLr sequences.

```
>90% conservation          *             * *           **      *    **      **       *
>50% conservation          *             * *         * **    * **** ** ***   **       *
Segment:                [-a-]         [---b---]       [-c--] [-----d------]  [-e-     ]
Consensus numbering:    1---5         6------14       15--20 21----------34  35--     39
FI_1_human__  DFADVVCYTQKADS.....PMDDFFQCVN.....GKYISQ.MKACDGINDCGDQS..DEL.....C
FI_2_human__  ..........CKAC.....QGKGFHCKS.....GVCIPS.QYQCNGEVDCITGE..DEVG.....C
C6_human____  CIPSKLCKIEEADC.....KNKFRCDS.....GRCIAR.KLECNGENDCGDNS..DERD.....C
C7_human____  SCEPTRGCPTEEGC.....GERFRCFS.....GQCISK.SLVCNGDSDCDEDS.ADEDR.....C
C8a_human___  SCSSSTTCVRQAQC.....GQDFQCKE.....TGRCLKR.HLVCNGDQDCLDGS..DEDD.....C
C8b_human___  DCVTNRPCGSQVRC.....EGFVCAQ.....TGRCVNR.RLLCNGDNDCGDQS..DEAN.....C
C9_human____  CVPTEPCEDAEDDC.....GNDFQCST.....GRCIKM.RLRCNGDNDCGDFS..DEDD.....C
ldlr_1_human  LLLAAAGTAVGDRC.....ERNEFQCQD.....GKCISY.KWVCDGSAECQDGS..DESQET...C
ldlr_2_human  .........LSVTC.....KSGDFSCGG.....RVNRCIPQ.FWRCDGQVDCDNGS..DEQG.....C
ldlr_3_human  .........PPKTC.....SQDEFRCHD.....GKCISR.QFVCDSDRDCLDGS..DEAS.....C
ldlr_4_human  .........PVLTC.....GPASFQCNS.....STCIPQ.LWACDNDPDCEDGS..DEWPQR...C
ldlr_5_human  .RGLYVFQGDSSPC.....SAFEFHCLS.....GECIHS.SWRCDGGPDCKDKS..DEEN.....C
ldlr_6_human  .........AVATC.....RPDEFQCSD.....GNCIHG.SRQCDREYDCKDMS..DEVG.....C
ldlr_7_human  .........VNVTLC.....EGPNKFKCHS.....GECITL.DKVCNMARDCRDWS..DEPIKE...C
lrp_1_human_  SALVAAAIDAPKTC.....SPKQFACRD.....QITCISK.GWRCDGERDCPDGS..DEAPEI...C
lrp_2_human_  ......PQSKAQRC.....QPNEHNCLG.....TELCVPM.SRLCNGVQDCMDGS..DEGPH....C
lrp_3_human_  TCLANPSYVPPPQC.....QPGEFACAN.....SRCIQE.RWKCDGDNDCLDNS..DEAPAL...C
lrp_4_human_  .........HQHTC.....PSDRFKCEN.....NRCIPN.RWLCDGDNDCGNSE..DESNAT...C
lrp_5_human_  .........SARTC.....PPNQFSCAS.....GRCIPI.SWTCDLDDDCGDRS..DESAS....C
lrp_6_human_  .........AYPTC.....FPLTQFTCNN.....GRCINI.NWRCDNDNDCGDNS..DEAG.....C
lrp_7_human_  ..........SHSC.....SSTQFKCNS.....GRCIPE.HWTCDGDNDCGDYS..DETHAN...C
lrp_8_human_  ...TNQATRPPGGC.....HTDEFQCRL.....DGLCIPL.RWRCDGDTDCMDSS..DEKS.....C
lrp_9_human_  .......EGVTHVC.....DPSVKFGCKD.....SARCISK.AWVCDGDNDCEDNS..DEEN.....C
lrp_10_human  .........ESLAC.....RPPSHPCAN.....NTSVCLPP.DKLCDGNDDCGDGS..DEGEL....C
lrp_11_human  LQDDLTCRAVNSSC.....RAQDEFECAN.....GECINF.SLTCDGVPHCKDKS..DEKPSY...C
lrp_12_human  .........NSRRC.....KKTFRQCSN.....GRCVSN.MLWCNGADDCGDGS..DEIP.....C
lrp_13_human  .........NKTAC.....GVGEFRCRD.....GTCIGN.SSRCNQFVDCEDAS..DEMN.....C
lrp_14_human  .........SATDCSSYFRLGVKGVLFQPCER.....TSLCYAP.SWVCDGANDCGDYS..DERD.....C
lrp_15_human  ......PGVKRPRC.....PLNYFACPS.....GRCIPM.SWTCDKEDDCEHGE..DETH.....C
lrp_16_human  ..........NKFC.....SEAQFECQN.....HRCISK.QWLCDGSDDCGDGS..DEAAH....C
lrp_17_human  .........EGKTC.....GPSSFSCPG.....THVCVPE.RWLCDGDKDCADGA..DESIAAG..C
lrp_18_human  ........LYNSTC.....DDREFMCQN.....RQCIPK.HFVCDHDRDCADGS..DESPE....C
lrp_19_human  .........EYPTC.....GPSEFRCAN.....GRCLSSRQWECDGENDCHDQS..DEAPKNPH.C
lrp_20_human  .......TSPEHKC.....NASSQFLCSS.....GRCVAE.ALLCNGQDDCGDSS..DERG.....C
lrp_21_human  FYLGSDGRTCVSNC.....TASQFVCKN.....DKCIPF.WWKCDTEDDCGDHS..DEPPD....C
lrp_22_human  .........PEFKC.....RPGQFQCST.....GICTNP.AFICDGDNDCQDNS..DEAN.....C
lrp_23_human  .........DIHVC.....LPSQFKCTN.....TNRCIPG.IFRCNGQDNCGDGE..DERD.....C
lrp_24_human  .........PEVTC.....APNQFQCSI.....TKRCIPR.VWVCDRDNDCVDGS..DEPAN....C
lrp_25_human  .........TQMTC.....GVDEFRCKD.....SGRCIPA.RWKCDGEDDCGDGS..DEPKEE...C
lrp_26_human  .........DERTC.....EPYQFRCKN.....NRCVPG.RWQCDYDNDCGDNS..DEES.....C
lrp_27_human  .........TPRPC.....SESEFSCAN.....GRCIAG.RWKCDGDHDCADGS..DEKD.....C
lrp_28_human  ..........TPRC.....DMDQFQCKS.....GHCIPL.RWRCDADADCMDGS..DEEA.....C
lrp_29_human  .......GTGVRTC.....PLDEFQCNN.....TLCKPL.AWKCDGEDDCGDNS..DENPEE...C
lrp_30_human  .........ARFVC.....PPNRPFRCKN.....DRVCLWI.GRQCDGTDNCGDGT..DEED.....C
lrp_31_human  ....EPPTAHTTHC.....KDKKEFLCRN.....QRCLSS.SLRCNMFDDCGDGS..DEED.....C
lrp_17_celeg  ........PSAVAC.....AEGTFPCSN.....GHCINQ.TKVCDGHNDCHDEQVSDESLAT...C
lrp_21_celeg  .........GNRTC.....TDQEFHCTSNAKLAQPKYECIPR.AWLCDGDVTCAGGE..DESTEL...C
lrp_34_celeg  .........RSAQC.....DPPLRFRCAH.....SRLCLNI.LQLCNGFNDCGPNDFSDEHLSM...C
vldlr_1_hum_  RESGATGTGRKAKC.....EPSQFQCTN.....GRCITL.LWKCDGDEDCVDGS..DEKN.....C
vldlr_2_hum_  .........VKKTC.....AESDFVCNN.....GQCVPS.RWKCDGDPDCEDGS..DESPEQ...C
vldlr_3_hum_  .........HMRTC.....RIHEISCGA.....HSTQCIPV.SWRCDGENDCDSGE..DEEN.....C
vldlr_4_hum_  .........GNITC.....SPDEFTCSS.....GRCISR.NFVCNGQDDCSDGS..DELD.....C
vldlr_5_hum_  .........APPTC.....GAHEFQCST.....SSCIPI.SWVCDDDADCSDQS..DESLEQ...C
vldlr_6_hum_  ....GRQPVIHTKC.....PASEIQCGS.....GECIHK.KWRCDGDPDCKDGS..DEVN.....C
vldlr_7_hum_  .........PSRTC.....RPDQFECED.....GSCIHG.SRQCNGIRDCVDGS..DEVN.....C
vldlr_8_hum_  ........KNVNQC.....LGPGKFKCRS.....GECIDI.SKVCNQEQDCRDWS..DEPLKE...C
PERLEC_1_hum  FRRLGTVPQFPRAC.....TEAEFACHS.....YNECVAL.EYRCDRRPDCRDMS..DELN.....C
PERLEC_2_hum  PQPLLPGSVRPLPC.....GPQEAACRN.....GHCIPR.DYLCDGQEDCEDGS..DELD.....C
PERLEC_3_hum  ........GPPPPC.....EPNEFPCGN.....GHCALK.LWRCDGDFDCEDRT..DEAN.....C
PERLEC_4_hum  .....PTKRPEEVC.....GPTQFRCVS.....TNMCIPA.SFHCDEESDCPDRS..DEFG.....C
GP330_1_rat_  PYGMKLQRDQMTCC.....GSLSFPCNN.....GKCVPS.FFRCDGVDDCHDNS..DEXQ.....C
GP330_2_rat_  .......GVFNNTC.....SPSAFACVR.....GGQCIPG.QWHCDRQNDCLDGS..DEXK.....C
GP330_3_rat_  .....PTHATSSTC.....PSTSFTCDN.....HVCIPK.DWVCDTDNDCSDGS..DEXN.....C
GP330_4_rat_  ........QASGTC.....QPTQFRCPD.....HRCISP.LYVCDGDKDCADGS..DEXG.....C
```



(a) 194 LDLr sequences:	a	b	c	d	e	Predicted	Predicted
	1---5	6------14	15--20	21----------39		β-strand	turn/coil
>50% conserved	C	....F.C..	G.CI..	.W.CDGD.DC.DGSDE..C			
GOR I	TETTT	TTTEETTTT	TTEEEE	EETTTTTTTTTTTTTTTTT		23%	77%
GOR III	EETTC	CTTEEEETT	TTEEEE	EEEETTTCTCCTTCTTTTT		36%	64%
CHOU-FASMAN	TTTTT	TTTHEETTT	TEEEHE	EEETTTTTTTTTTTTTTTT		23%	72%
PHD	LLLLL	LLLLEEEEL	LLEELL	EELLLLLLLLLLLLLLLLL		21%	79%
Hydropathy	...eb	..eebeb.e	.ebb..	eb.beb...b.ebeee..b			
Solvent accessibility	eeeeb	..eebeb.e	bebbee	.bebeeeeebeeeeeebe.			

(b) 11 complement LDLrs:	a	b	c	d	e		
	1---5	6------14	15--20	21----------39			
FI_LDLr1_human	QKADSPMDDFFQCVN		GKYISQ	MKACDGINDCGDQSDEL.C		Predicted	Predicted
FI_LDLr2_human	.CKAC	QGKGFHCKS	GVCIPS	QYQCNGEVDCITGEDEVGC		β-strand	turn/coil
Cys-Cys bridging	1	u	1	u	u	u	
GOR I	ETTTT	TTTTTETTT	TTTEEE	EEETTTTTTTTTTTTTTTT		21%	79%
GOR III	CTTTT	TTTEEEETT	TTCEEE	EEEETTCCCCCCCCTTTTT		28%	72%
Chou-Fasman	HETTT	TTTEEEETT	TEEEHH	EEETTTTTTTTTTTTTTTT		28%	64%
PHD	LLLLL	LLLEEEELL	LEEEEE	EEELLLLLLLLLLLLLLLL		28%	72%
Hydropathy	.e.eb	..e.bebee	bebb.e	eb.bebeeebbe.eee..b			
Solvent accessibility	eeeee	eeeebebee	eebbee	.b.beee.ebeeeeeeeee			

Fig. 3. Consensus secondary structure analyses for the LDLr superfamily. The residue numbering and alignment is from Fig. 2. Disulphide bridges are indicated by 1 or u (unknown). Average predictions from (a) 194 LDLr sequences and (b) 11 complement LDLr sequences are shown (E, β-strand; H, α-helix; T, turn; C or L, coil). The total β-strand and turn/coil structure from each prediction is summarised on the right. The hydropathy analyses are denoted by e for >67% hydrophilic residues and b for >67% hydrophobic residues. The PHD accessibility analyses are denoted by e (exposed) for >36% solvent accessibility and b (buried) for <9% solvent accessible.

3.2. Comparison with small β-sheet protein folds

The 194 LDLr sequences were scored for compatibility with 254 known protein folds. Averaging the 194 Z-scores gave the best values of −0.1 to −1.3 for eight folds, starting from a baseline of +10.0. A Z-score of less than −2.7 corresponds to high structural identity. Five of the 8 best-scoring folds were of size 46–53 residues and maintained by Cys–Cys bridges with short β-strands, while three had α-helix contents and could be discarded. These five folds were epidermal growth factor (Brookhaven code 1EGF), tissue-type plasminogen activator (1TPM), omega-aga-IVB toxin (1OMA), transcriptional elongation factor (1TFI), and sea anemone toxin (1ATX). The 1TFI fold was discarded as this is a zinc binding domain with reduced thiol groups. This left four folds which scored consistently within the top 6–14 (± 10) of the 254. These results suggest that the LDLr structure has no close structural homology with previously determined folds, but that parts of it resembles small β-sheet proteins.

3.3. Expression and spectroscopic characterization of LDLr2

In factor I, LDLr2 (Fig. 1) was expressed for structural studies (see section 2) for reason of its greater sequence similarity with the LDLr consensus (Fig. 3). The expressed LDLr2:GST protein migrated at approximately 31 kDa by SDS-PAGE. This was purified by glutathione agarose affinity chromatography to yield a single product. This was used to raise a polyclonal antiserum, which was found to recognise reduced factor I heavy chain and non-reduced factor I in Western blots, and to react in an ELISA assay with factor I and GST. Cleavage of LDLr2:GST in solution was achieved using Factor Xa. LDLr2 was separated by reversed phase chromatography and purified as a single peak by gel filtration on Superdex 30. The N-terminal 15 amino acids were sequenced and determined to be the correct sequence of GIPACQGKGFHCKSG... (where the residues GIP were derived from the linker between GST and LDLr2). Mass spectroscopy of LDLr2 at pH 3.5 showed a single peak at a molecular mass of 4099 ± 1. This mass agreed precisely with that calculated from a 40-residue LDLr2 sequence $GIPA^{240}CQ...DEVGCA^{276}$, in which the 5 Asp/Glu residues and the N-terminus are protonated, the Cys residues are disulphide bridged, and the C-terminus carboxyl is ionized. Minor peaks observed at masses of 4138 and 4176 may correspond to the binding of one or two K^+ cations. The use of Ellman's reagent showed that no free thiol groups were present in non-denatured LDLr2. Therefore we concluded that LDLr2 has been correctly expressed. The existence of a folded LDLr2 structure was supported by fluorescence data. Fluorescence spectra on LDLr2 were recorded using 0–6 M guanidine-HCl solutions. The emissions at 301.0 to 302.5 nm were plotted against guanidine concentration to show a sigmoidal curve, in which unfolding started at 1 M guanidine and ended at 4 M guanidine, after which no further changes were observed.

FT-IR and CD data supported the LDLr structure predictions of Fig. 3. FT-IR spectra of LDLr2 and factor I in ^2H_2O buffers (Figs. 4a and 4c) showed a broad absorbance band at 1645 cm^{-1} and 1638 cm^{-1}, respectively. Within this band, the second derivative spectra showed a single intense component at 1641 cm^{-1} in LDLr2 and 1637 cm^{-1} in factor I. β-Sheet bands commonly occur at 1634–1638 cm^{-1}, with an anti-parallel β-sheet component in the region of 1670–1680 cm^{-1} [23]. For LDLr2, the band at 1641 cm^{-1} was consistent with β-sheet, while for factor I the bands at 1682 and 1637 cm^{-1} agreed with this assignment. The absence of significant absorbance in the region 1650–1655 cm^{-1} indicated the lack of major α-helix structures in LDLr2. A minor band seen at 1649 cm^{-1} for factor I could be attributed to the single solvent-exposed α-helix of the SP domain (Fig. 1). β-Turns are characterised by bands close to 1670 cm^{-1}, and loop regions by bands close to 1640 cm^{-1} [23]. For LDLr2, strong bands were seen at 1673 cm^{-1} and 1641 cm^{-1}

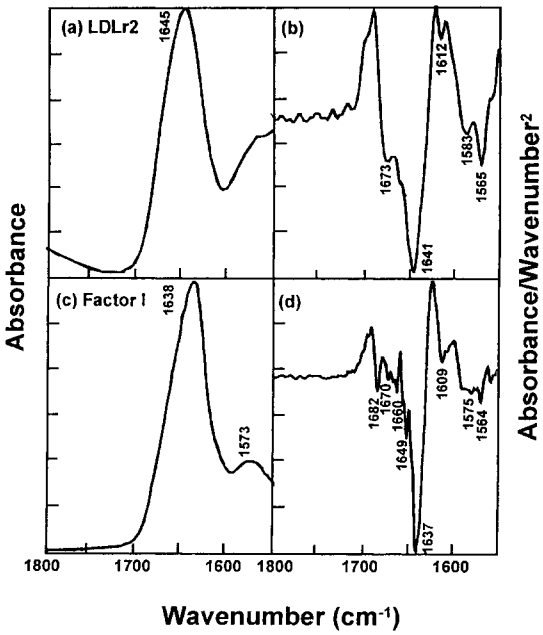


Fig. 4. FT-IR spectra at 20°C of LDLr2 and human factor I in ^2H_2O buffers. Sample concentrations were 1.6 and 2.4 mg/ml, respectively. Factor I was prepared as in [24]. The (a) and (c) absorbance and (b) and (d) second derivative spectra are shown for each sample.

(where the latter overlaps with the β-sheet band), and a weak one at 1670 cm^{-1} was seen in factor I. In summary, the factor I spectrum (Fig. 4c and 4d) corresponded to the summation of a 14% contribution from two LDLrs (Figs. 4a and 4b), a 43% contribution from the SP domain with a known β-sheet FT-IR spectrum [24], and the absence of α-helix in the FIM and CD5 domains. CD spectra were recorded to show that LDLr2 has a large amount of loop or random structure and a small amount of β-strand structure. The interpretation of CD spectra can however be affected by a large proportion of Cys residues [14], of which there are 6 in LDLr2.

4. Conclusions

The LDLr superfamily has evolved to be a key component of complement proteins and cell surface receptors. Despite many insertions into the consensus (or primordial) sequence, the basic fold of this domain has remained unchanged. The three central segments b, c and d are predicted to define the core structure of LDLr in terms of a β-sheet stabilised by three disulphide bridges and hydrophobic residues. The disulphide pairing is presently unknown, but will determine the conformation adopted by the β-strands. It is possible to conclude from Fig. 3 that the biologically important Asp/Glu residues are not related to any β-sheet features, and are located on an exposed surface loop of LDLr for ligand binding. This type of structure is analogous to other complement protein domains that possess conserved Cys residues. These structures, which include epidermal growth factor [25], the SP domain [26], the short consensus/complement repeat [14], and the C1q globular heads [27], are all predominantly composed of β-sheet structures.

Acknowledgements: We thank the Wellcome Trust, Royal Society and Royal Free Peter Samuel Trust Fund for financial support. We thank also Mr. A.C. Willis for N-terminal peptide sequencing, Dr R.T. Aplin for mass spectroscopy measurements, and Mr A.J. Beavil for access to fluorescence and CD spectroscopy facilities. Earlier structural analyses of LDLr were reported in [9].

References

[1] Schneider, W.J. (1989) Biochim. Biophys. Acta 988, 303–317.
[2] Moestrup, S.K. (1994) Biochim. Biophys. Acta 1197, 197–213.
[3] Catterall, C.F., Lyons, A., Sim, R.B., Day, A.J. and Harris, T.J. (1987) Biochem. J. 242, 849–856.
[4] Kunnath-Muglia, L.M., Chang, G.H., Sim, R.B., Day, A.J. and Ezekowitz, R.A. (1993) Mol. Immunol. 30, 1249–1256.
[5] Stanley, K.K. and Herz, J. (1987) EMBO J. 6, 1951–1957.
[6] Perkins, S.J., Smith, K.F., Williams, S.C., Haris, P.I., Chapman, D. and Sim, R.B. (1994) J. Mol. Biol. 238, 104–119.
[7] Edwards, Y.J.K. and Perkins, S.J. (1995) FEBS Lett. 358, 283–286.
[8] Taylor, W.R. (1988) J. Mol. Evol. 28, 161–169.
[9] Smith, K.F. (1992) PhD Thesis, University of London.
[10] Taylor, W.R. (1986) J. Mol. Biol. 188, 233–258.
[11] Chou, P.Y. and Fasman, G.D. (1978) Adv. Enzymol. 47, 45–148.
[12] Garnier, J., Osguthorpe, D.J. and Robson, B. (1978) J. Mol. Biol. 120, 97–120.
[13] Gibrat, J.-F., Garnier, J. and Robson, B. (1987) J. Mol. Biol. 198, 425–443.
[14] Perkins, S.J., Haris, P.I., Sim, R.B. and Chapman, D. (1988) Biochemistry 27, 4004–4012.
[15] Rost, B. and Sander, C. (1993) J. Mol. Biol. 232, 584–599.
[16] Perkins, S.J. (1986) Eur. J. Biochem. 157, 169–180.
[17] Eisenberg, D. (1984) Annu. Rev. Biochem. 53, 595–623.
[18] Rost, B. and Sander, C. (1994) Proteins: Structure, Function and Genetics 20, 216–226.
[19] Jones, D.T., Taylor, W.R. and Thornton, J.M. (1992) Nature 358, 86–89.
[20] Johnstone, A. and Thorpe, R. (1982) Immunochemistry in practise, Blackwell, Oxford.
[21] Perkins, S.J., Smith, K.F. and Sim, R.B. (1993) Biochem. J. 295, 101–108.
[22] Yamamoto, T., Davis, C.G., Brown, M.S., Schneider, W.J., Casey, M.L., Goldstein, J.L. and Russell, D.W. (1984) Cell 39, 27–38.
[23] Lee, D.C., Haris, P.I., Chapman, D. and Mitchell, R.C. (1990) Biochemistry 29, 9185–9193.
[24] Prestrelski, S.J., Byler, D.M. and Liebman, M.N. (1991) Biochemistry 30, 133–143.
[25] Cooke, R.M., Wilkinson, A.J, Baron, M., Pastore, A., Tappin, M.J., Campbell, I.D., Gregory, H. and Sheard, B. (1987) Nature 327, 339–341.
[26] Perkins, S.J. and Smith, K.F. (1993) Biochem. J. 295, 109–114.
[27] Smith, K.F., Haris, P.I., Chapman, D., Reid, K.B.M. and Perkins, S.J. (1994) Biochem. J. 301, 249–256.