Research article

The relationship between the L1 and L2 domains of the insulin and epidermal growth factor receptors and leucine-rich repeat modules

Colin W Ward*[1,3] and Thomas PJ Garrett[2,3]

Address: [1]CSIRO Health Sciences and Nutrition, 343 Royal Parade Parkville, Victoria 3052, Australia;, [2]Walter and Elisa Hall Institute, Royal Melbourne Hospital, Parkville, Victoria 3052, Australia; and [3]Cooperative Research Centre for Cellular Growth Factors, Royal Melbourne Hospital, Parkville, Victoria, 3050, Australia

E-mail: Colin W Ward* - Colin.Ward@hsn.csiro.au; Thomas PJ Garrett - Tom.Garrett@hsn.csiro.au

*Corresponding author

Published: 27 July 2001

BMC Bioinformatics 2001, **2**:4

Received: 6 July 2001
Accepted: 27 July 2001

This article is available from: http://www.biomedcentral.com/1471-2105/2/4

Abstract

Background: Leucine-rich repeats are one of the more common modules found in proteins. The leucine-rich repeat consensus motif is LxxLxLxxNxLxxLxxLxxLxx- where the first 11–12 residues are highly conserved and the remainder of the repeat can vary in size Leucine-rich repeat proteins have been subdivided into seven subfamilies, none of which include members of the epidermal growth factor receptor or insulin receptor families despite the similarity between the 3D structure of the L domains of the type I insulin-like growth factor receptor and some leucine-rich repeat proteins.

Results: Here we have used profile searches and multiple sequence alignments to identify the repeat motif lxx-Lxlxx-Nx-Lxx-Lxx-Lxx-Lxx- in the L1 and L2 domains of the insulin receptor and epidermal growth factor receptors. These analyses were aided by reference to the known three dimensional structures of the insulin-like growth factor type I receptor L domains and two members of the leucine rich repeat family, porcine ribonuclease inhibitor and internalin 1B. Pectate lyase, another beta helix protein, can also be seen to contain the sequence motif and much of the structural features characteristic of leucine-rich repeat proteins, despite the existence of major insertions in some of its repeats.

Conclusion: Multiple sequence alignments and comparisons of the 3D structures has shown that right-handed beta helix proteins such as pectate lyase and the L domains of members of the insulin receptor and epidermal growth factor receptor families, are members of the leucine-rich repeat superfamily.

Background

Many proteins have a modular architecture and are composed of a number of different, sometimes repeated structural units [1,2]. The four most common modules found in the extracellular regions of proteins are immunoglobulin (Ig) domains, epidermal growth factor (EGF)-like repeats, fibronectin type 3 (Fn3) modules and leucine-rich repeats [2]. Two of these, Fn3 modules [3–6] and EGF-like repeats [7–10], have been identified in members of the insulin receptor (IR) family.

There is some evidence to suggest that the L domains of the IR and EGFR families are leucine-rich repeats. At 10–16%, leucine is the most common residue in these do-

mains. Furthermore, the 3D structure of the L1/cys-rich/L2 fragment of the IGF-1R showed that the L domains were single-stranded right-handed β-helices [8] with structural similarities to pectate lyase, a right-handed beta helix protein [11,12] and the ribonuclease inhibitor, a right-handed beta-alpha superhelix protein [13]. Ribonuclease inhibitor (RI) is recognised as a member of the superfamily of leucine-rich repeat proteins [14–16] while pectate lyase is not, although similarities in the sequence patterns and 3D structures of pectate lyase and RI have been noted [15–17]. The IGF-1R is listed as a leucine-rich repeat protein in the SCOP database [http://scop.mrc-lmb.cam.ac.uk/scop/] but not in any of the annotated sequence databases such as SwissProt [http://srs.ebi.ac.uk/] or SMART [http://smart.embl-heidelberg.de/index.shtml] [18]). Similarly, none of the other L-domain containing proteins from the IR or EGFR families are listed as leucine-rich repeats in these data bases or in a recent summary of the complete protein tyrosine kinase family present in the human genome [19].

The superfamily of leucine-rich repeat proteins has been subdivided into six subfamilies termed: typical, RI-like, CC (cysteine-containing), PS (plant specific), SD22-like and bacterial [16]. These subfamilies are characterised by different lengths and consensus sequences of the repeats (Fig. 1). The bulk of the LRRs have repeats of 22–25 amino acid residues while RI, with its alternating repeats of 28 to 29 residues, is considered somewhat atypical [20]. The family has been expanded further to include the small proteoglycans, which were shown to consist of different combinations of two types of LRRs of 21 (S-type) and 26 (T-type) amino acid residues [21]. The LRR consensus sequence is LxxLxLxxNx-Lxx-Lxx-Lxx-Lxx- (Fig. 1) where the first 11–12 residues are highly conserved and the remainder of the repeat can vary in size [16,17,21]. Some repeats have C instead of N at the 4^{th} highly conserved position and I, V, M, F, Y, A or C at the positions denoted by L in the above consensus (Fig. 1).

In view of this variation in sequence motifs among LRR proteins, the sequences of the L1 and L2 domains of members of the IR and EGFR families were re-examined. The LRR motif is difficult to detect when examining a single sequence, but becomes more readily recognisable when multiple sequence alignments are analysed. The identification of such conserved sequence motifs was greatly aided by the availability of the 3D structures of the IGF-1R L1 and L2 domains [8], pectate lyase [11,12] and the known LRR proteins RI [13] and internalin 1B [22]. The data indicate that pectate lyase and the L domains of members of the IR and EGFR families should be included in the expanding family of LRR proteins

Results

Preliminary analyses of the SwissProt data base using profiles based on alignments of single (Prf1) or double LRR repeats (Prf2) failed to score members of the IR or EGFR families in the first 500 ranked scores. However searches with Prf-4 (based on four tandem repeats), ranked IR family members at positions 234 (INSR_RAT), 245 (INSR_MOUSE), 262 (IG1R_RAT) and 478 (IG1R_HUMAN) and EGFR family members at 146 (LT23_CAEEL), 174 (EGFR_MOUSE), 448 (EGFR_HUMAN) and 457 (EGFR_CHICK). The alignments with the mouse, human and chick EGFRs were to residues equivalent to 363–452 in the L2 domain of EGFR_human. In contrast, the alignments with the other sequences were to regions in the cytoplasmic domain. When the analyses were restricted to the receptor ectodomains, the alignments were to sequences in the L1 or L2 domains. Examples are: Prf-1 with residues 1–28, Prf-2 with residues 10–78 and Prf-4 with residues 46–153 in the human insulin receptor. An alternative approach, using a profile generated from the alignment of the L1 domains of 13 members of the IR family (see Fig. 2) was also encouraging as it ranked four known LRR proteins in the next 11 hits after the 26 known members of the IR or EGFR families. These were repeats 8–14 (residues 295–450) of ESA8_TRYEQ (ranked 28^{th}) and ESA8_TRYBB (ranked 30^{th}), and repeats 1–5 (residues 28–177) of TSHR_MOUSE (ranked 35^{th}) and TSHR_RAT (ranked 37^{th}).

However, in both approaches, the repeating units in the L1 and L2 domains, known from the 3D structure of IGF-1R [8], did not align exactly with the repeating units in the LRR proteins [20,23] or the LRR protein profiles (see Methods). The presence of insertions in some regions of the IR or EGFR proteins tended to confound the profile alignments. Consequently, the nature of the repeating sequence motifs in the L1 and L2 domains of the IR and EGFR families were examined manually using the 3D structure of the L1 and L2 domains of IGF-1R as a guide to adjust the alignments obtained from the profile analyses.

The consensus sequences for the six LRR protein subfamilies [16] are shown in Fig. 1. The sizes of the repeating units vary from 22–31 with eight highly conserved positions, numbered 1 to 8 in Fig 1. Conserved position 4 usually contains N or C but sometimes S, T, P, Q or A. The other 7 highly conserved positions are generally L but frequently contain alternative non-polar residues such as I, V, M, C, A, F, or Y, particularly in some LRR subfamilies. Also included are the type S and type T LRRs from the small proteoglycans [21]. The repeating motifs have been arranged in Fig. 1 so that they all begin with the 11-residue stretch LxxLxLxxNxL which contains

conserved position	1	2	3	4	5	6	7	8	
LRR_Typical	Lxx	LxLxxNx			Lxx/	LPxx/φFxx		Lxx/	(22-31)
LRR_RI type	Lxx	LxLxxNx			Lxxxxφxx		Lxxx	Lxxxxx/	(28-29)
				C					
LRR_CC type	Lxx	LxLxxCxxITDxxφxx					Lxxx/	Cxx/	(25-27)
								L	
LRR_SD22-like	Lxx	LxLxxNx/Ixx				Ixx	LEx	Lxx/	(21-23)
LRR_PS type	Lxx	LxLxxNx			LxGx	IPxx	LGx	Lxx/	(23-25)
LRR_bacterial type	Lxx	LxVxxNx			Lxx	LPx	LPxx/		(20-22)
LRR_PGC-S	Lxx	LxLxxNx			Ixx	φPxx	LPxx/		(21-23)
LRR_PGC-T	Lxx	LxLxxNx			Lxx/	φxxxxFxx		Lxx/	(26-28)
IR family	Ixx/LxIxxNx/Lxx					Lxx	Lxx	Lxx-	(23-42)
EGFR family	Ixx/LxIxxNx/Lxx					Lxx	Lxx	Lxx--	(22-46)
Pectate Lyase	xxx	φxφxxNx			φxx	φxx/	φx	φxx	(23-39)

φ = hydrophobic (A,V,I,L,M,F,Y,W)

Figure 1
Comparison of repeat motifs in IR and EGFR family L domains with the sequence motifs of LRR subfamilies. The motifs for LRR_Typical, LRR_PS type, LRR_bacterial type, LRR_SD22-like, LRR_CC type, LRR_R1 type-a, LRR_R1 type-b are based on [16], the motifs for the S and T type repeats of small proteoglycans (LRR_PGC) are from [21]. The size of the repeat can vary by insertions at the positions denoted by the dashes [16]. Also shown is a consensus motif for the pectate lyase superfamily based on the alignments in Fig. 5 and [26].

the highly conserved positions 1 to 5. This motif corresponds to a linear central sequence with two half turns at either end [20]. Deletions and/or insertions, particularly in the C-terminal portion of the repeats (after conserved positions 5, 6, 7 and 8) account for the size differences seen between different LRR protein members.

The repeat sequences in members of the IR family that correspond to the six turns of the parallel beta helix of the L1 and L2 domains of IGF-1R [8] are shown in Fig. 2. The corresponding predicted repeats for members of the EGFR family are shown in Fig. 3. LRR-like motifs are most readily identified in the 4^{th} and 5^{th} repeats in the IR family L1 and EGFR family L2 domains and in the 3^{rd}, 4^{th} and 5^{th} repeats of the IR family L2 and EGFR family L1 domains. The size of these repeats (22–30 residues) is similar to the range found in other LRRs (Fig. 1). The 4^{th} repeat in the L1 and L2 domains of the IR family contain an insert of 7–8 residues between conserved positions 1 and 2 (Fig. 2), which appears as an extra loop in the 3D structure [8]. The 4^{th} repeats of the majority of the EGFR L1 and L2 domains show similar inserts of 6–7 residues in this region, with the EGFR from *Schistosoma mansoni* (SMEGFRA) and *Drosophila melanogaster* (TOP_DROME) having inserts of 11 and 13 residues respectively (Fig. 3). Variable sized insertions are also seen in some (or all) of the 1^{st} and 2^{nd} repeats of the IR and EGFR L1 domains and the 1^{st}, 2^{nd} and 6^{th} repeats of the IR and EGFR family L2 domains.

The nature of the residues in the eight highly conserved positions of the LRR-like repeats in the L domains of IR and EGFR are summarised in Table 1. The most commonly found residues in the L domains of the IR and EGFR families are: isoleucine at positions 1 and 3; asparagine at position 4; and leucine at positions 2, 5, 6, 7 and 8. However other non-polar residues, particularly valine, phenylalanine and cysteine, occur in positions 1–3 and

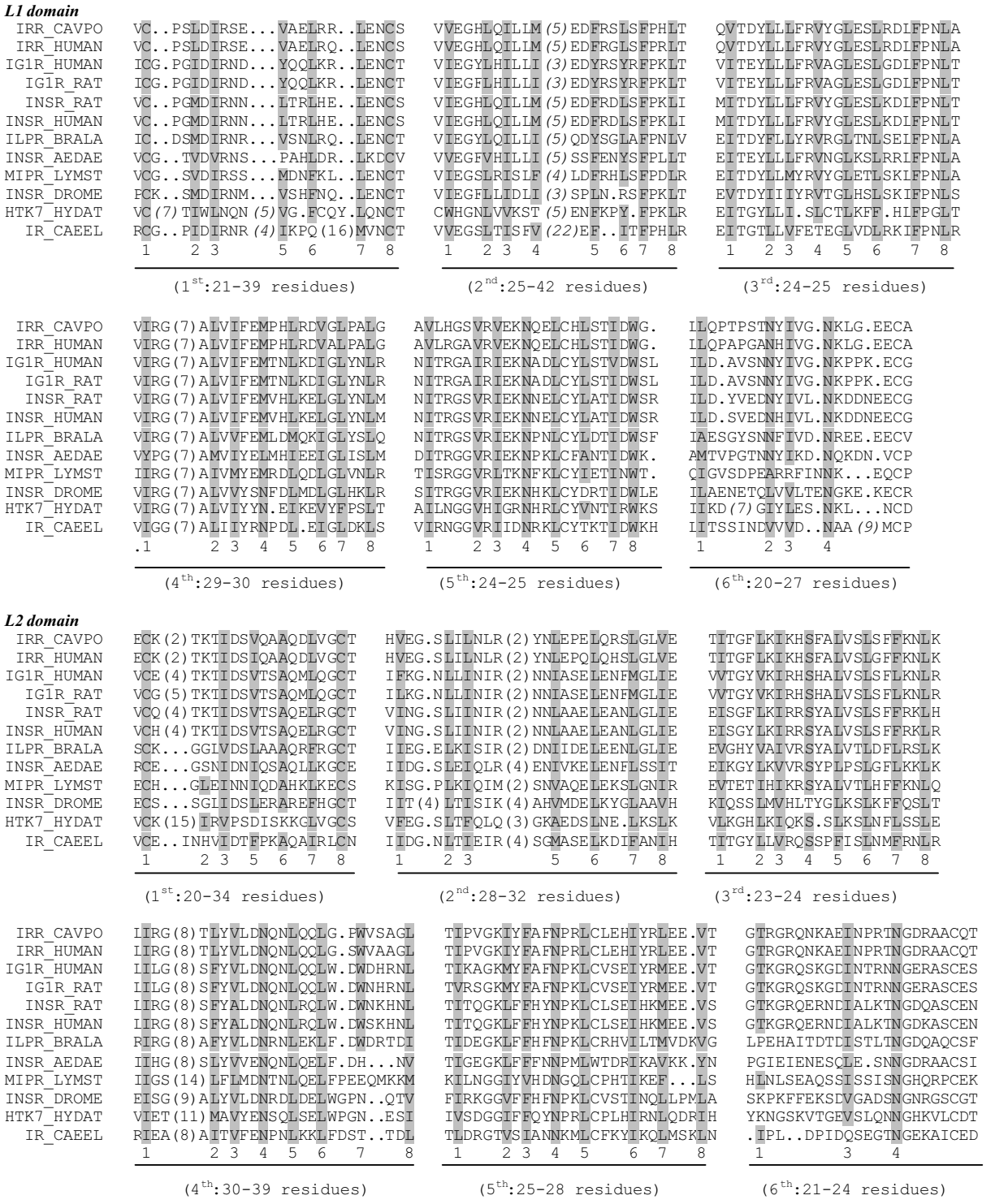

Figure 2

Leucine-rich repeats in the L1 and L2 domain of members of the insulin receptor family. The sequences were sourced from SwissProt except for IR_CAEEL which is GENBANK:AF012437. Residues equivalent to conserved positions 1 to 8 in the sequence motif LxxLxLxxNxLxxLxxLxxLxx are shaded. The INSR_MOUSE is not shown as it differs from INSR_RAT at only S139Y in the 6th repeat of the L1 domain and S415N in the 4th repeat of the L2 domain, positions that are not conserved in the LRR motif.

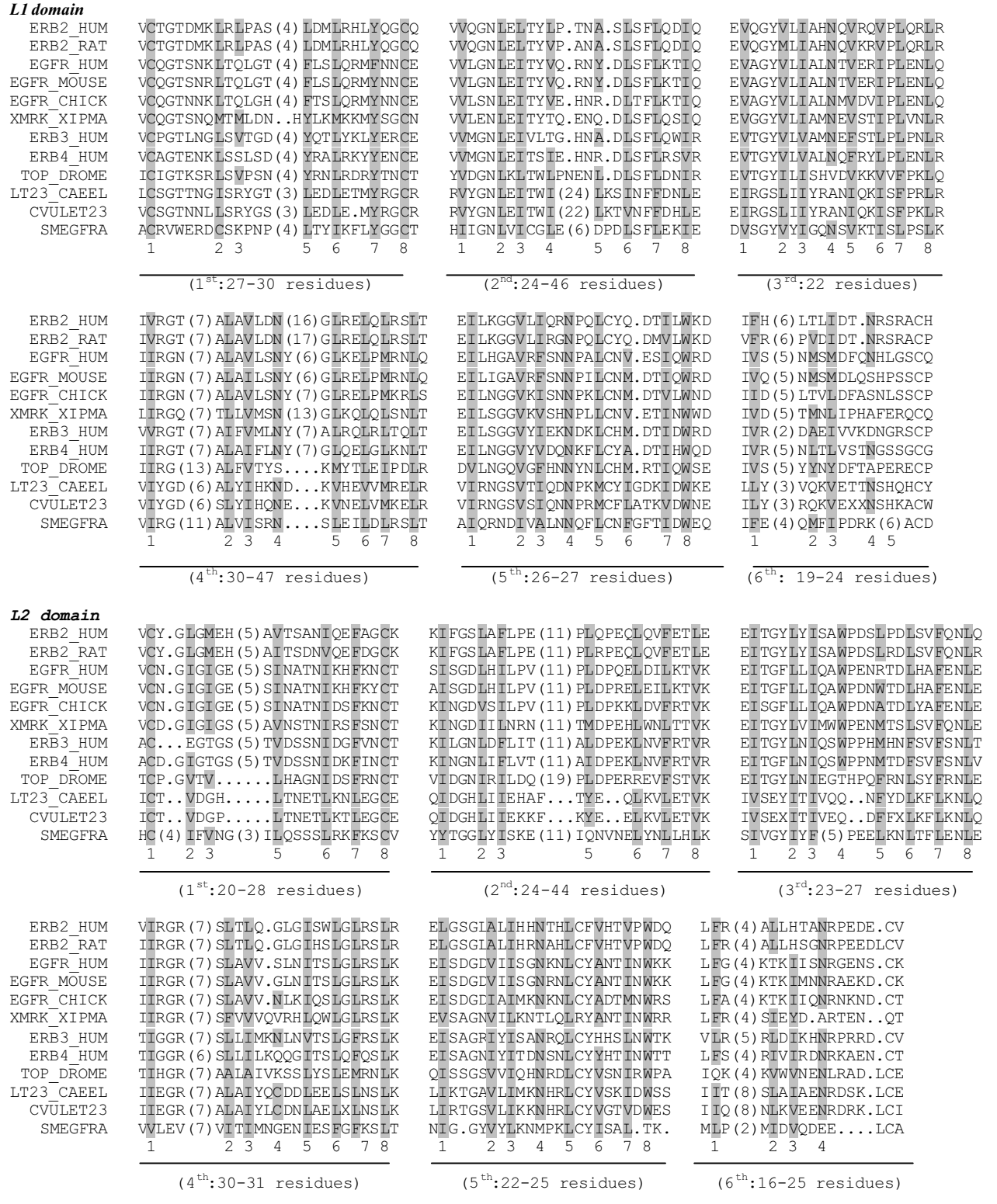

Figure 3
Leucine-rich repeats in the L1 and L2 domains of members of the epidermal growth factor receptor family. The sequences were sourced from SwissProt except for SMEGFRA which is from EMBL and CVULET23 which is from GENBANK. Residues equivalent to conserved positions 1 to 8 in the sequence motif LxxLxLxxNxLxxLxxLxxLxx are shaded.

Table 1: Frequency of occurrence of amino acid residues in the eight conserved positions of the LRR-like motifs in the L domains of the IR and EGFR families

Residue	IR Family (L1 + L2)								EGFR Family (L1 + L2)							
	1st	2nd	3rd	4th	5th	6th	7th	8th	1st	2nd	3rd	4th	5th	6th	7th	8th
H	-	1	-	-	-	-	1	-	-	-	2	-	-	1	-	-
K	2	7	-	-	-	-	-	-	-	-	1	2	-	-	-	-
R	-	-	-	-	-	1	-	-	-	-	2	-	3	1	-	-
D	-	-	-	-	1	1	-	-	-	-	-	1	1	-	-	-
E	-	-	-	-	-	-	1	-	-	1	-	-	-	-	-	-
N	-	9	-	**68**	-	-	2	-	-	-	-	**61**	-	-	-	-
Q	-	-	1	-	-	2	-	-	1	2	3	3	2	2	-	-
S	-	1	-	12	-	-	-	-	-	-	1	4	-	-	-	-
T	7	-	-	3	-	1	1	-	-	5	2	4	-	-	-	1
P	1	-	-	-	1	-	-	-	-	-	-	-	-	-	-	-
G	1	2	-	-	-	-	-	-	-	-	2	4	-	-	-	-
A	-	1	3	-	1	13	-	-	-	4	2	5	4	5	-	-
V	17	20	26	12	13	4	2	13	35	35	24	6	15	13	6	10
I	**91**	16	**95**	5	10	20	13	12	**57**	18	**78**	3	14	19	16	9
L	6	**70**	15	2	**85**	**80**	**51**	**64**	6	**70**	13	3	**57**	**69**	**46**	**53**
M	1	9	2	14	3	-	9	1	-	5	5	1	7	5	13	-
C	26	-	-	-	-	-		26	24	1	-	2	-	-	-	24
F	2	7	13	1	11	4	39	-	10	1	8	-	8	4	28	-
Y	1	-	1	-	5	4		1	1	2	1	-	6	1	11	-
W	1	-	-	-	-	-		13	-	-	-	8	1	-		23

Calculated from conserved positions indicated in Figs 2 & 3.

5–8 in some sequences (Figs 2 & 3). The most common alternatives to asparagine at the 4th conserved position are valine, methionine, serine and tryptophan (Table 1).

The locations of the eight conserved residues in the 3D structure of IGF-1R L1 domain have been compared with the location of the equivalent residues in the 3D structures of the known LRR proteins porcine RI and internalin 1B. As shown in Fig. 4, the 3D structures of the first part of the repeat (LxxLxLxxNx) are very similar in each of these proteins (β-strand flanked by two half turns), while the structures of the remainder of the repeats are more variable. This is despite the size variability of the sequence patterns seen in some of the repeats in the L domains of the IGF-1R and the more frequent occurrence of non-L residues at highly conserved positions 1–3 and 5–8.

The pectate lyase family [24,25] shows even greater variability, with large insertions in some repeats corresponding to loops overlaying the parallel beta helix core (see Fig. 4). As for other LRRs, the most conserved residues include positions 2–4, being two aliphatic side chains and the Asn ladder (Fig. 5). However, at each of the 8 highly conserved positions, isoleucine or valine is preferred over leucine [26]. In positions 2 and 3 the frequency of I or V is greater than 60%. The features which set apart the pectate lyase repeat from that of other LRR families are an aromatic residue in position 5, extending the third β-strand and forming a bulge (*top left* in Figure 4). Residues at positions 7 and 8 slide over to spatially overlap with positions 6 and 7 in other LRRs.

Discussion

There is considerable interest in the structure of the L domains of the IR and EGFR families and their relationship with other proteins, because of their importance in ligand binding (see [8,10,27,28]). In this paper, evidence is presented to show that these L domains contain all of the features of leucine-rich repeats. Multiple sequence alignments, coupled with the 3 D structure of the L1 and L2 domains of the IGF-1R [8], enabled the residues equivalent to the conserved residues in known LRR motifs, to be identified as summarised in Figs 2,3,4.

A variant of the motif LxxLxLxxNx-Lxx-Lxx-Lxx-Lxx- found in LRR proteins can be identified in the L1 and L2 domains of the IR and EGFR families, where I rather

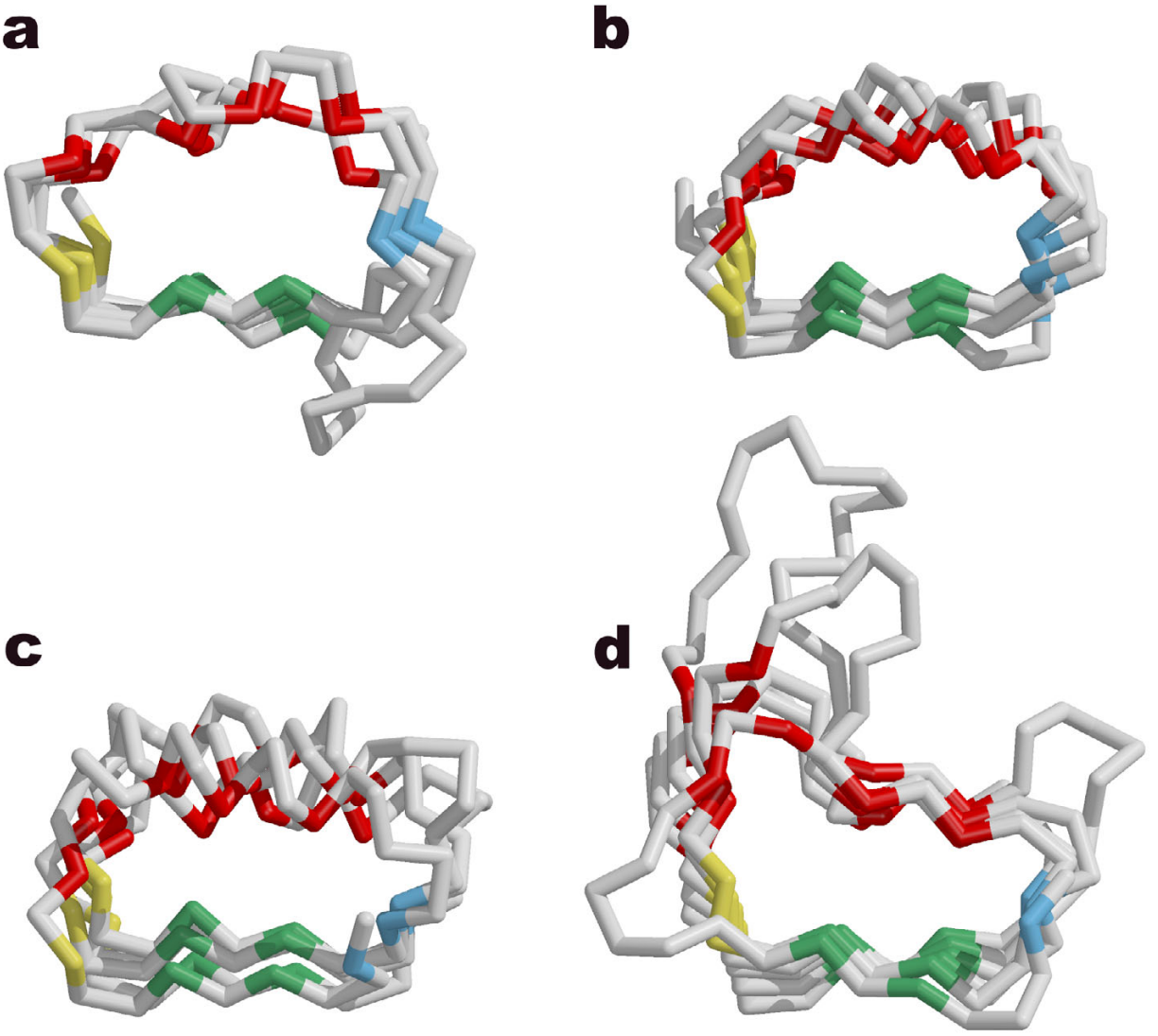

Figure 4
Comparison of the location of conserved motif residues in the 3D structures of IGF-1R L1 domain and other LRR and β-helix proteins. The 3D structures are *a* human, IG1R fragment (PDB: 1IGR, residues 50–143), *b,* internalin 1B from *Listeria monocytogenes* (PDB: 1D0B, residues 143–241),*c,* porcine ribonuclease inhibitor (PDB:2BNH, residues 24–148) and *d,* pectate lyase C from *Erwinia chrysanthemi* (PDB:1PEC, residues 86–245). Residues equivalent to conserved positions 1 to 8 in the sequence motif LxxLxLxxNxLxxLxxLxxLxx are colour coded as follows: position 1: blue; positions 2 & 3: green; position 4: yellow; positions 5 – 8: red.

than L is the most common residue at positions 1 and 3. Other non-polar amino acids frequently occur in some positions (Table 1, Figs 2 & 3), as found with other LRR proteins [16,17,21]. The L domains of the IR and EGFR families contain five full repeats with the 6^{th} partially truncated. Some sequences have insertions between conserved positions 1 and 2 and 4 and 5 (Figs 2 & 3), which complicate the analysis. However the combination of ex-

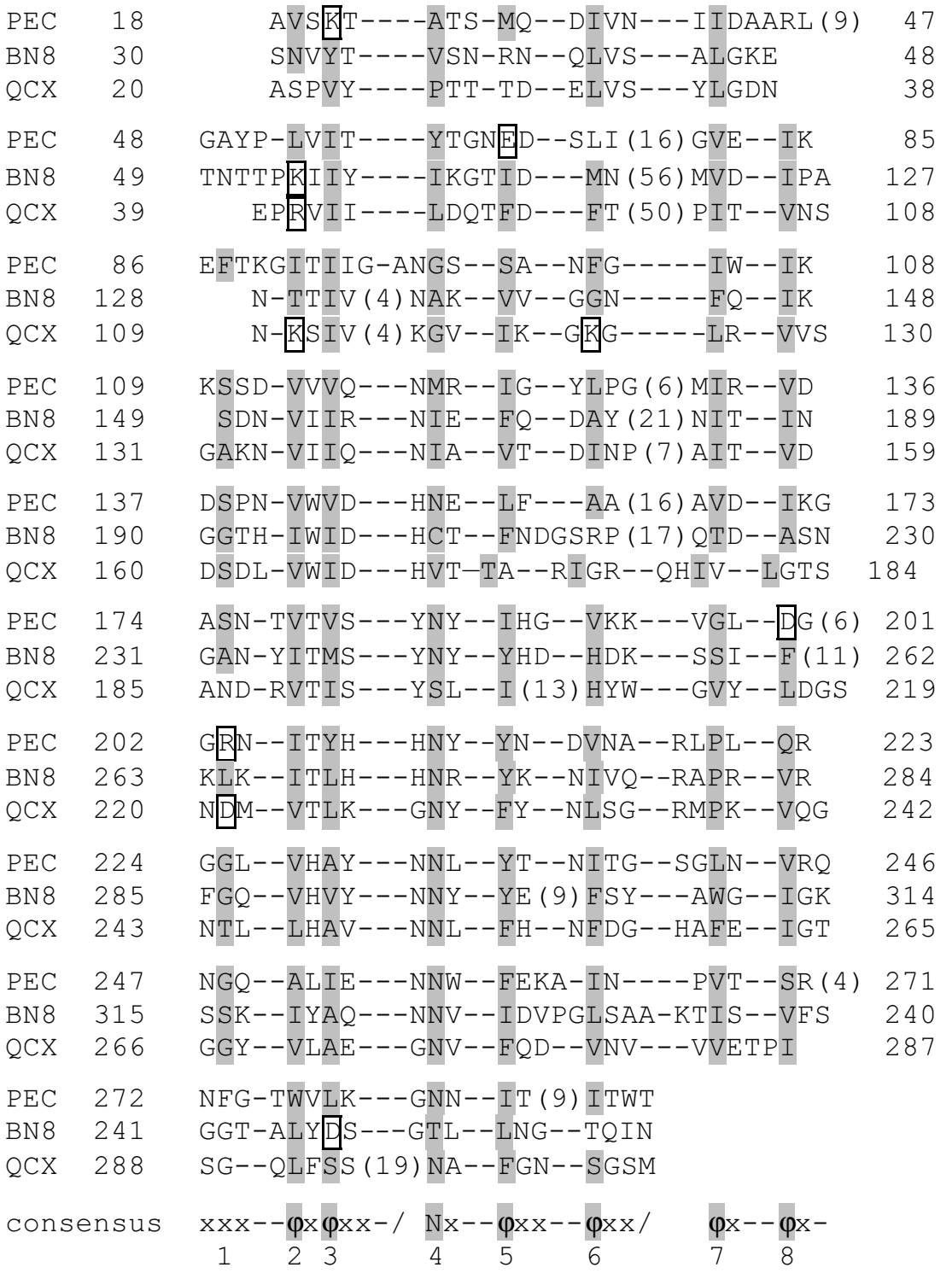

Figure 5
Structural repeats in the sequences of *Erwinia chrysanthemi* pectate lyase C (PEC, [11,12]), *Bacillus Subtilis* pectate lyase (BN8, [33]) and *Aspergillus niger* pectin lyase B (QCX, [34]). Alignment is based on a superposition of the three-dimentional structures, with residues in the sequence motif shaded. Charged residues (boxed) also form part of the motif but have either a compensating buried charge or are solvent exposed.

amining multiple sequence alignments with the known 3D structure of the IGF-1R allowed the sequence motifs to be established.

Leucine rich repeat proteins are members of a broader class of proteins termed solenoid proteins, where the repeating structural units in the polypeptide chain form a continuous superhelix [17]. Solenoid proteins, including LRRs, show the simplest relationship between sequence and structure, compared to the more complicated folds of globular proteins. Thus recognition of such motifs can provide valuable insights into the predicted structure of such protein domains [17]. The most conserved structural feature of LRRs is the LxxLxLxxNx region, while the remainder of the repeat can differ dramatically [15,17] as illustrated in Fig. 5. The central sequence xLxLx in this region of the repeat forms a β-strand with successive repeats of this β-strand forming a parallel β-sheet on one side of the LRR module. This corresponds to the second β-sheet in the L domains of the IGF-1R, which is the structural counterpart of the β-sheet that forms the inner face of RI and internalin (Fig. 4). This face is involved in protein-protein interactions in RI [15], U2 [29] and the IR family (see [8,10]).

As shown in Fig 4, the first region of the repeat LxxLxLxxNx in RI, internalin, the IGF-1R L domains and pectate lyase all adopt similar structural folds, while the remainder of the repeats are highly variable. Most of the repeats in the IR and EGFR L domains are between 21 and 30 residues, within the range commonly found in other LRRs. The inserts are accommodated as loops which do not, or are unlikely to, perturb the core structure [8]. Three of the repeats in pectate lyase have large inserts of 11–17 residues between the 6^{th} and 7^{th} conserved position (Fig. 5). The major difference between the 3D structures of RI and internalin versus the L domains of IGF-1R is the absence of the repetitive helix on the opposite face to the canonical β-sheet in the IGF-1R L domains and thus a lack of curvature although the L domains of IGF-1R are capped by α-helices at the N-terminal end and to a lesser extent at the C-terminus [8].

The existence of repeats in the L domains of IR and EGFR was first reported by Bajaj et al. [30] who described five repeats in the region equivalent to 1–119 in human IR. The subsequent 3D structure determination of the two L domains in the IGF-1R showed that they each contain five full and one partial repeat [8].

Conclusion

Here we have shown, using a combination of sequence analyses and 3D structure comparisons, that variations of the repeating motif typical of LRRs is present in the L domains of members of the IR and EGFR subfamilies and in β-helix proteins. This motif is not obvious, is difficult to detect with sequence analysis programs and has not been described previously. Comparison of the 3D structure of these domains with other protein structures showed that L domains matched equally well to the pectate lyase family and LRRs such as porcine ribonuclease inhibitor. We conclude that these three groups should be considered part of the same LRR superfamily. In the IR and EGFR subfamilies, isoleucine (or valine) is preferred over leucine at some positions of the repeat while in β-helix proteins isoleucine or valine (or occasionally phenylalanine) are always preferred over leucine.

Methods

Multiple Sequence Analysis

The sequence analysis programs used were from the sequence analysis software package of the Genetics Computer Group of the University of Wisconsin, Biotechnology Centre, Madison, Wisconsin, USA. Files of individual proteins were edited using the Seqed program and aligned using Pileup. Final adjustment to these alignments were made manually as required using Lineup. Profiles [31,32] were generated from these aligned sequences using ProfileMake. The SwissProt database was probed using ProfileSearch and the alignments displayed with ProfileSegments. Specific sequences were analysed using Gap or ProfileGap. Gap weight and Length weight penalties used were 3.0 and 0.3 respectively unless stated otherwise.

The proteins used to generate these profiles, were the chondroadherin precursor (CHAD_BOVIN, 10 LRRs), platelet glycoprotein 1B alpha chain precursor (GPBA_MOUSE, 8 LRRs), the bone proteoglycan 2 precursor (PGS2_RABIT, 10 LRRs) and the putative receptor protein tyrosine kinase from *Arabidopsis* (TMK1_ARATH, 8 LRRs). Three profiles were generated: Prf-1 based on the alignment of the 38 single LRRs in these four proteins; Prf-2 based on the alignment of the 19 tandem repeats of two LRRs from these four proteins and Prf-4 which was generated from the alignment of eight sequences, each containing four LRRs in tandem. The fragments from CHAD_BOVIN and PGS2_RABIT used in Prf-4 corresponded to repeats 1–4 and 7–10.

Abbreviations

3D, three dimensional; EGF, epidermal growth factor, EGFR, epidermal growth factor receptor, Fn3, fibronectin type 3; IR, insulin receptor; IGF-1, insulin-like growth factor 1; IGF-1R, the type I insulin-like growth factor receptor; IRR, insulin receptor related receptor; LRR, leucine-rich repeat; RI, ribonuclease inhibitor; SCOP, structural comparison of proteins; SMART, simple modular architecture research tool.

References

1. Bork P: **Mobile modules and motifs.** *Curr Opin Struct Biol* 1992, **2**:413-421
2. Bork P, Downing AK, Kieffer B, Campbell ID: **Structure and distribution of modules in extracellular proteins.** *Quart Rev Biophys* 1996, **29**:119-167
3. O'Bryan JP, Frye RA, Cogswell PC, Neubauer Z, Kitch B, Prokop C, Espinosa III R, Le Beau MM, Earp HS, Liu ET: **axl, a transforming gene isolated from primary human myeloid leukemia cells, encodes a novel receptor tyrosine kinase.** *Mol Cell Biol* 1991, **11**:5016-5031
4. Mulhern TD, Booker GW, Cosgrove L: **A third fibronectin type-3 domain in the insulin-family receptors.** *Trends Biochem Sci* 1998, **23**:465-466
5. Marino-Buslje C, Mizuguchi K, Siddle K, Blundell TL: **A third fibronectin type 3 domain in the extracellular region of the insulin receptor family.** *FEBS Lett* 1998, **441**:331-336
6. Ward CW: **Members of the insulin receptor family contain three fibronectin type 3 domains.** *Growth Factors* 1999, **16**:315-322
7. Ward CW, Hoyne PA, Flegg RH: **Insulin and epidermal growth factor receptors contain the cysteine repeat motif found in the tumour necrosis factor receptor.** *Proteins: Struct Funct Genet* 1995, **22**:141-153
8. Garrett TPJ, McKern NM, Lou M, Frenkel MJ, Bentley JD, Lovrecz GL, Elleman TC, Cosgrove L, Ward CW: **The structure of the first three domains of the type I insulin-like growth factor receptor.** *Nature* 1998, **394**:395-399
9. Ward CW, Garrett TPJ, McKern NM, Lawrence LJ: **Structure of the insulin receptor family: unexpected relationships with other proteins.** *Today's Life Sciences* 1999, **11**:26-32
10. Adams TE, Epa VC, Garrett TJ, Ward CW: **Structure and function of the type I insulin-like growth factor receptor.** *Cell Molec Life Sci.* 2000, **57**:1050-1093
11. Yoder MD, Lietzke SE, Jurnak F: **Unusual structural features in the parallel beta-helix in pectate lyases.** *Structure* 1993, **1**:241-251
12. Yoder MD, Keen NT, Jurnak F: **New domain motif: pectate lyase C, a secreted plant virulence factor.** *Science* 1993, **260**:1503-1507
13. Kobe B, Deisenhofer J: **Crystal structure of porcine ribonuclease inhibitor, a protein with leucine repeats.** *Nature* 1993, **366**:751-756
14. Kobe B, Deisenhofer J: **The leucine-rich repeat: a versatile binding motif.** *Trends Biochem Sci* 1994, **19**:415-421
15. Kobe B, Deisenhofer J: **Proteins with leucine-rich repeats.** *Curr Opin Struct Biol* 1995, **5**:409-416
16. Kajava AV: **Structural diversity of leucine-rich repeat proteins.** *J Mol Biol* 1998, **277**:519-527
17. Kobe B, Kajava AV: **When protein folding is simplified to protein coiling: the continuum of solenoid protein structures.** *Trends Biochem Sci* 2000, **25**:509-515
18. Schultz J, Copley RR, Doerks T, Ponting CP, Bork P: **SMART: a web-based tool for the study of genetically mobile domains.** *Nucleic Acids Res* 2000, **28**:231-234
19. Robinson DR, Wu Y-M, Lin S-F: **The protein tyrosine kinase family of the human genome.** *Oncogene* 2001, **19**:5548-5557
20. Kajava AV, Vassart G, Wodak SJ: **Modeling of the three-dimensional structure of proteins with typical leucine-rich repeats.** *Structure* 1995, **3**:867-877
21. Matsushima N, Ohyanagi T, Tanaka T, Kretsinger RH: **Super-motifs and evolution of tandem leucine-rich repeats within the small proteoglycans-biglycan, decorin, lumican, fibromodulin, PRELP, keratocan, osteoadherin, epiphycan, and osteoglycin.** *Proteins: Struct Funct Genet* 2000, **38**:210-225
22. Marino M, Braun L, Cossart P, Ghosh P: **Structure of the InlB leucine-rich repeats, a domain that triggers host cell invasion by the bacterial pathogen L-monocytogenes.** *Molecular Cell* 1999, **4**:1063-1072
23. Smiley BL, Stadnyk AW, Myler PJ, Stuart K: **The trypanosome leucine repeat gene in the variant surface glycoprotein expression site encodes a putative metal-binding domain and a region resembling protein-binding domains of yeast, Drosophila, and mammalian proteins.** *Mol Cell Biol* 1990, **10**:6436-6444
24. Henrissat B, Heffron SE, Yoder MD, Lietzky SE, Jurnak F: **Functional implications for structure-based sequence alignment of proteins in the extracellular pectate lyase superfamily.** *Plant Physiol.* 1995, **107**:963-976
25. Yoder MD, Jurnak F: **The parallel β helix and other coiled folds.** *FASEB J* 1995, **9**:335-342
26. Heffron S, Moe GG, Sieber V, Mengaud J, Cossart P, Vitali J, Jurnak F: **Sequence profile of a parallel b helix in the pectate lyase superfamily.** *J. Struct. Biol.* 1998, **122**:223-235
27. Lax I, Bellot F, Howk R, Ullrich A, Givol D, Schlessinger J: **Functional analysis of the ligand binding site of EGF-receptor utilizing chimeric chicken/human receptor molecules.** *EMBO J* 1989, **8**:421-427
28. Kohda D, Odaka M, Lax I, Kawasaki H, Suzuki K, Ullrich A, Schlessinger J, Inagaki F: **A 40-kDa epidermal growth factor/transforming growth factor alpha-binding domain produced by limited proteolysis of the extracellular domain of the epidermal growth factor receptor.** *J Biol Chem* 1993, **268**:1976-1981
29. Price SR, Evans PR, Nagai K: **Crystal structure of the spliceosomal U2B"-U2A' protein complex bound to a fragment of U2 small nuclear RNA.** *Nature* 1998, **394**:645-650
30. Bajaj M, Waterfield MD, Schlessinger J, Taylor WR, Blundell T: **On the tertiary structure of the extracellular domains of the epidermal growth factor and insulin receptors.** *Biochim.Biophys. Acta* 1987, **916**:220-226
31. Devereux J, Haeberli P, Smithies O: **A comprehensive set of sequence analysis programs for the VAX.** *Nucleic Acids Res* 1984, **12**:387-395
32. Gribskov M, Luthy R, Eisenberg D: **Profile analysis.** *Methods Enzymol* 1990, **183**:146-159
33. Pickersgill R, Jenkins J, Harris G, Nasser W, Robert-Baudouy J: **The structure of Bacillus subtilis pectate lyase in complex with calcium.** *Nat Struct Biol* 1994, **1**:717-723
34. Vitali J, Schick B, Kester HCM, Visser J, Jurnak F: **The three-dimensional structure of Aspergillus niger pectin lyase B at 1.7 Å resolution.** *Plant Physiol.* 1998, **116**:69-80