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

ABSTRACT

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.

INTRODUCTION

Background Many proteins have a modular architecture and are composed of a number of different, sometimes repeated structural units. 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. Two of these, Fn3 modules and EGF-like repeats, 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 domains. 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 with structural similarities to pectate lyase, a right-handed beta helix protein and the ribonuclease inhibitor, a right-handed beta-alpha superhelix protein. Ribonuclease inhibitor (RI) is recognised as a member of the superfamily of leucine-rich repeat proteins while pectate lyase is not, although similarities in the sequence patterns and 3D structures of pectate lyase and RI have been noted. The IGF-1R is listed as a leucine-rich repeat protein in the SCOP database but not in any of the annotated sequence databases such as SwissProt or SMART). 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. 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. 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. 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. 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. Some repeats have C instead of N at the 4th 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, pectate lyase and the known LRR proteins RI and internalin 1B. 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

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.