Review

Structural analysis of leucine-rich-repeat variants in proteins associated with human diseases

N. Matsushima⁠*, N. Tachi⁠b, Y. Kuroki⁠c, P. Enkhbayar⁠a,⁠d,⁠e, M. Osaki⁠f, M. Kamiya⁠f and R. H. Kretsinger⁠g

a Division of Biophysics, School of Health Sciences, Sapporo Medical University, Sapporo, Hokkaido 060-8556 (Japan), Fax: +81 11 612 3617, e-mail: matusima@sapmed.ac.jp
b Department of Occupational Therapy, School of Health Sciences, Sapporo Medical University, Sapporo, Hokkaido 060-8556 (Japan)
c Department of Biochemistry, School of Medicine, Sapporo Medical University, Sapporo, Hokkaido 060-8556 (Japan)
d Department of Biophysics, Faculty of Biology, National University of Mongolia, Ulaanbaatar 210646/377 (Mongolia)
e Division of Agricultural Science, Graduate School of Agriculture, Hokkaido University, Sapporo, Hokkaido 060-0810 (Japan)
f Division of Biological Science, Graduate School of Science, Hokkaido University, Sapporo, Hokkaido 060-0810 (Japan)
g Department of Biology, University of Virginia, Charlottesville, Virginia 22904 (USA)

Received 4 May 2005; received after revision 18 August 2005; accepted 1 September 2005
Online First 18 October 2005

Abstract. A number of human diseases have been shown to be associated with mutation in the genes encoding leucine-rich-repeat (LRR)-containing proteins. They include 16 different LRR proteins. Mutations of these proteins are associated with 19 human diseases. The mutations occur frequently within the LRR domains as well as their neighboring domains, including cysteine clusters. Here, based on the sequence analysis of the LRR domains and the known structure of LRR proteins, we describe some features of different sequence variants and discuss their adverse effects. The mutations in the cysteine clusters, which preclude the formation of sulfide bridges or lead to a wrong paring of cysteines in extracellular proteins or extracellular domains, occur with high frequency. In contrast, missense mutations at some specific positions in LRRs are very rare or are not observed at all.

Key words. Leucine-rich repeats; human diseases; cysteine clusters; typical LRR motif; RI-like LRR motif; Cryopyrin/Nalp3/PYPAF1; CARD15.

Introduction

Leucine-rich-repeat (LRR)-containing domains are present in 4748 proteins in the PFAM database (10 November 2004) (reviewed in [1–4]). LRR proteins have been identified in viruses (25), bacteria (403), archae (1) and eukaryotes (4319). Many LRR proteins are involved in protein-ligand interactions; these include plant immune response and the mammalian innate immune response (reviewed in [5]). Most LRRs are 20–30 amino acids long and the repeat number ranges from 2 to 52. The LRR proteins have been divided into seven classes [3]. One group of LRR proteins that includes small LRR proteoglycans (SLRP) has LRRs

* Corresponding author.
LRR proteins in disease

from more than one of the seven classes [4, 6]. All known LRRs adopt an arc shape (figs. 1, 2) [4]. A number of human diseases have been shown to be associated with mutation in the genes encoding LRR proteins (reviewed in [7–9]). They include NgR [10], LGI1 [11], Trk-A [12–14], polycystin 1/PKD1 [15, 16], nyctalopin [17, 18], FSHr [19–21], LH/CGr [22–24], TSHr [24, 25–27], keratocan [28, 29], GPIbα [30–32], GPIbβ [33–35], GPIX [36], LRRK2 [37], CIAS1 [38–42], CIITA [43, 44], and Nod2 [45, 46] (table 1). Mutations of these proteins are associated with schizophrenia [47], ADLTE/ADPEAF [48–58], CSNB1/XLCSNB [17, 18], CIPA [59, 60], ADPKD [61, 62], ODG1 [63–68], LCH [69], Graves disease [70], thyrotropin resistance [71–78], FGH [79], papillary cancer [80], hyperthyroidism [81–85], CNA2 [86, 87], BSS [88–135], PT- vWD [136–139], Parkinson’s disease [37, 140–144], CINCA/NOMID [39, 41, 42, 145–149], BLS II [43, 150–156] and Crohn’s disease [45, 46, 157, 158] (table 1). The mutations occur frequently within the LRR domains as well as in their neighboring domains, including cysteine clusters at the N- and C-termini.

Here we focus on understanding the adverse effects of different sequence variants based on the sequence analysis of the LRR domains and the known structure of LRR proteins.

**Structural principles and features of LRR proteins**

All LRR repeats can be divided into a highly conserved segment and a variable segment. The highly conserved segment consists of an 11-residue stretch, LxxLxLxxNxLxxLpxxoFxzxLxx, in which ‘L’ is Leu, Ile, Val, or Phe, ‘N’ is Asn, Thr, Ser, or Cys, and ‘C’ is Cys or Ser [159]. Seven classes of LRR motifs have been proposed, characterized by different lengths and consensus sequences of the variable segments of repeats [159, 160]. Comparative analysis of the known structures of 17 different LRR proteins has revealed the following features [4].

1) Three residues at positions 3–5 in the highly conserved segment form a short β-strand. The β-strands stack parallel and then form an arc (fig. 2). The concave face consists of a parallel β-sheet, and the convex face is made of a variety of secondary structures such as α-helix, 310-helix, polyproline II helix, and an extended structure or a tandem arrangement of β-turns. In most LRR proteins the β-strands on the concave surface and (mostly) helical elements on the convex surface are connected by short loops or β-turns.

2) The seven classes of LRR motifs adopt a variety of structural units. Typical LRRs have the consensus sequence of LxxLxLxxNxLxxLpxxoFxzxLxx, form a short β-strand. The β-strands stack parallel and then form an arc (fig. 2). The concave face consists of a parallel β-sheet, and the convex face is made of a variety of secondary structures such as α-helix, 310-helix, polyproline II helix, and an extended structure or a tandem arrangement of β-turns. In most LRR proteins the β-strands on the concave surface and (mostly) helical elements on the convex surface are connected by short loops or β-turns.

3) The seven classes of LRR motifs have the consensus sequence of LxxLxLxxNxLxxLpxxoFxzxLxx, form a short β-strand. The β-strands stack parallel and then form an arc (fig. 2). The concave face consists of a parallel β-sheet, and the convex face is made of a variety of secondary structures such as α-helix, 310-helix, polyproline II helix, and an extended structure or a tandem arrangement of β-turns. In most LRR proteins the β-strands on the concave surface and (mostly) helical elements on the convex surface are connected by short loops or β-turns.