A framework for interpreting the leucine-rich repeats of the *Listeria* internalins

Michael Marino*, Laurence Braun†, Pascale Cossart‡, and Partho Ghosh*‡

*Department of Chemistry and Biochemistry, University of California, San Diego, 9500 Gilman Drive, La Jolla, CA 92093-0314; and †Institut Pasteur, Unité des Interactions Bactéries-Cellules, 28 rue du Dr. Roux, 75015 Paris, France

The surface protein InlB of the bacterial pathogen *Listeria monocytogenes* is required for inducing phagocytosis in various non-phagocytic mammalian cell types *in vitro*. InlB causes tyrosine phosphorylation of host cell adaptor proteins, activation of phosphoinositide 3-kinase, and rearrangements of the actin cytoskeleton. These events lead to phagocytic uptake of the bacterium by the host cell. InlB belongs to the internalin family of *Listeria* proteins, which also includes InlA, another surface protein involved in host cell invasion. The internalins are the largest class of bacterial proteins containing leucine-rich repeats (LRR), a motif associated with protein–protein interactions. The LRR motif is found in a functionally diverse array of proteins, including those involved in the plant immune system and in the mammalian innate immune response. Structural and functional interpretations of the sequences of internalin family members are presented in light of the recently determined x-ray crystal structure of the InlB LRR domain.

A number of intracellular pathogens have the ability to invade mammalian cells that are normally non-phagocytic by inducing phagocytic behavior (1). The cellular machinery required for phagocytosis appears to be present in many types of non-phagocytic cells but in a constitutively unassembled or inactive state. Certain pathogens possess means by which to induce assembly of the phagocytic machinery necessary to bring about their own uptake. Among these pathogens is the facultative intracellular bacterium *Listeria monocytogenes*, a cause of meningitis and abortion in humans (2). *L. monocytogenes* induces its own phagocytosis in a large number of non-phagocytic cell types *in vitro* through the actions of the bacterial surface proteins InlA and InlB (also called internalin and internalin B, respectively) (3).

The related proteins InlA and InlB act on different sets of cell types. Although InlA is required for inducing phagocytosis in the enterocyte-like epithelial cell line Caco-2 (4), InlB is required for inducing phagocytosis in some hepatocyte, endothelial, epithelial, or fibroblast-like cell lines (e.g., Vero, HeLa, HEp-2, Chinese hamster ovary) (5–10). The different cell type specificities likely reflect the fact that InlA and InlB have different receptor specificities. The receptor for InlA is E-cadherin (11), and whereas the receptor for InlB has not been identified, it is clear that it is not E-cadherin. Nevertheless, InlB has been shown to have potent effects on host cell signaling events. InlB promotes activation of host cell phosphoinositide (PI) 3-kinase and tyrosine phosphorylation of certain adaptor proteins (Gab1, Shc, and Cbl) implicated in the membrane localization of phosphoinositide 3-kinase (PI 3-kinase) (12, 13). PI 3-kinase is known to have effects on actin polymerization (14), a process that is required for *L. monocytogenes* invasion of host cells (12). Reorganization of the actin cytoskeleton presumably leads to changes in the membrane that cause phagocytosis of the bacterium.

**Internalin Family**

InlA and InlB belong to the internalin family of *Listeria* proteins, which includes at least seven additional members in *L. monocytogenes* (InlC, InlC2, InlD, InlE, InlF, InlG, and InlH) (4, 15–18). Internalin proteins are also found in *Listeria ivanovii*, a pathogen of ruminants (19, 20), and also in nonpathogenic species of the genus *Listeria*. Except for InlC, which is a soluble and secreted protein, the *L. monocytogenes* internalins have sequence motifs that are consistent with bacterial cell surface localization. In contrast to InlB, which has been shown to be important to proliferation in hepatocytes *in vivo* (21, 22), the role of InlA in virulence has not been clearly demonstrated (23, 24). The biochemical functions and exact virulence properties of the other *L. monocytogenes* internalins remain to be determined.

Proteins of the internalin family share several sequence motifs, the most distinctive of these being leucine-rich repeats (LRR). The LRR motif is associated with protein–protein interactions and is found in both bacterial and eukaryotic proteins (25–27). The majority of bacterial LRR proteins are found in *Listeria* and belong to the internalin family, but LRR proteins have also been identified in *Yersinia* spp. (28), *Shigella flexneri* (29), *Salmonella typhimurium* (30), and *Burkholderia* (Pseudomonas) *solanacearum* (31). Among eukaryotes, the LRR motif is found in proteins encoded by the disease resistance (R) genes of the plant immune system (32–34) and by the toll and toll-like genes of *Drosophila* and mammals, which are involved in innate immune responses (35, 36). The LRR motif is not limited to virulence- or immunity-related proteins but is instead found in a large group of proteins with highly diverse functions and cellular localizations (25).

The LRR regions of InlA and InlB play critical roles in interactions with mammalian cells. The LRR region of InlA is necessary to promote invasion of permissive cells *in vitro* (37). Likewise, the LRR region of InlB is necessary for the activation of phosphoinositide 3-kinase, rearrangement of the actin cytoskeleton, and invasion of permissive cells *in vitro* (38). The recently determined x-ray crystal structure of the InlB LRR domain (39) provides a framework for the interpretation of sequences of internalin family members.

**Internalin LRR**

The internalin LRR regions are highly regular, with almost all repeats being 22 residues in length (Fig. 1). The LRRs are tandemly repeated from 6 times in InlG to 16 times in InlA. The 7.5 LRRs present in InlB give it its elongated and curved shape (Fig. 2). The repeats are similar to those observed in the LRRs of porcine and human ribonuclease inhibitors (RI) (40–42), the

This paper was presented at the National Academy of Sciences colloquium “Virulence and Defense in Host-Pathogen Interactions: Common Features Between Plants and Animals,” held December 9–11, 1999, at the Arnold and Mabel Beckman Center in Irvine, CA.

Abbreviations: LRR, leucine-rich repeat; RI, ribonuclease inhibitor.

*To whom reprint requests should be addressed. E-mail: pghosh@ucsd.edu.

8784–8788 | PNAS | August 1, 2000 | vol. 97 | no. 16
The U2LRR fragment of U2 small nuclear ribonucleoprotein (43), and the Ran GTP-ase activating protein rna1p (44). As in these structures, each repeat alternates between a short β-strand and an opposing antiparallel helical segment; the β-strands and helices are connected to each other by coils. The main chain wraps around in a right-handed sense, with the β-strands forming a parallel β-sheet and the helices stacking on each other, giving rise to the elongated shape of the LRR.

Although the β-strands are a highly conserved structural feature of LRR proteins, the helical segments are more divergent. The β-strands are almost always composed of three residues that are in precise register. On the other hand, the helical segments are variable in length, register, and type. In fact, one repeat of U2LRR contains a second β-strand instead of a helix. In InlB, the helical segments are short (three to five residues) and are exclusively composed of 310-helices. U2LRR possesses 23- and 25-residue repeats that also form 310-helices. This contrasts with RI and rna1p, which have long (10–14 residues) α-helices and also have longer repeat lengths (28 residues or greater). Apparently the more tightly wound 310-helix accommodates formation of short repeats more favorably than does the α-helix.

The curvature of the LRR region arises from the β-strands of adjacent repeats being packed more closely together than opposing helices. The β-strands are within interrepeat hydrogen bonding distance, whereas the helices form intrarepeat hydrogen bonds. The β-strands form the concave face and the helices the convex face.
convex face of the molecule (Fig. 2). This curvature would appear to limit the possible number of tandem LRRs, because a closed circle would be formed. However, the InlB β-strands are twisted and give the entire structure a right-handed superhelical twist, thereby placing no obvious limits on the possible number of tandem repeats (39). The β-strands of RI, U2LRR, and rna1p are not as extensively twisted and indeed place a limit on the number of possible tandem repeats. It is not clear at present what gives rise to this difference.

The coil regions connecting β-strands and 3_{10}-helices constitute the great majority (14–17 residues) of the 22-residue InlB LRR. Even though these regions lack conventional secondary structure, they form highly regular structures (rms deviation of 0.32 Å in Cα positions). This regularity may in part be caused by the presence of water molecules that act as extensions of secondary structure. These waters are intercalated between repeats and form bridging hydrogen bonds between main-chain atoms of adjacent repeats. Structural determination of InlB to high resolution (1.86 Å) allowed visualization of three distinct spines of water molecules that run the length of the repeat region (39). Some of these hydrogen bond-mediating waters are also observed in the 2.0-Å resolution structure of human RI (42). A regular pattern of hydrogen bond-mediating waters is expected to be present in other LRR proteins and to be observed in high-resolution structures of other LRR proteins.

LRR Flanking Sequences

The LRR regions are flanked by sequences that are conserved in the internalin family and that most likely play a role in providing stability. At the N terminus of the LRR region is a segment of ~40 residues termed the “N-terminal cap” in InlB, because it provides a hydrophilic cap on the hydrophobic core of the first LRR (39). The hydrophobic core would be exposed to solvent at either end of the LRR were it not for other portions of the protein. The N-terminal cap forms numerous polar and hydrophobic interactions with the first LRR. The most prominent of these is a hydrophobic interaction with Tyr at position 20 of the first LRR, a highly conserved residue in the internalin family (Fig. 1).

In addition to its structural role, the N-terminal cap has been suggested to be involved in function based on the observation of two highly solvent-exposed calcium ions that bind to this region (39). The calciums are not required for formation of protein structure and may instead act as metal ion bridges between InlB and mammalian cell surface receptors or binding proteins. The N-terminal cap and LRR regions of InlB are sufficient to elicit mammalian cell effects and to induce phagocytosis (38).

At the C terminus of the LRR region is a conserved sequence of ~100 residues termed the interrepeat (IR) region (17). The IR region is dispensable for the mammalian cell signaling activities of InlB but is required in InlA for inducing phagocytosis (37). The IR region has not yet been visualized, but crystals of intact InlB have been obtained that should elucidate the role of this as well as other regions of internalins (M.M. and P.G., unpublished data). At least a portion of the IR region is likely to provide a hydrophilic cap on the last LRR.

LRR Pattern: Structural Residues

A total of 10 positions of the 22-residue internalin LRR are highly conserved and serve structural rather than functional roles. Seven positions (2, 5, 7, 12, 15, 18, and 21) of the repeat contain mainly Leus or Iles that point inward and compose the hydrophobic core of the protein (Fig. 3). Other hydrophobic residues are accommodated at these positions, including Val, Phe, Met, and Ala (Fig. 1). Position 10 also faces inward and usually contains Asns or Glns, which form the so-called “Asn ladder” (45). The Asn hydrogen bonds to main-chain atoms in its own repeat and to those in the preceding repeat. This explains why position 10 in the first repeat is not constrained to be an Asn or Gln.

The seven hydrophobic positions along with the Asn-ladder position define the internalin LRR sequence motif (Fig. 4). This motif corresponds almost identically to the 22-residue repeat
LRR Pattern: Functional Residues

More than half the residues of the internalin LRR face outward, are variable, and can serve to define protein–protein interaction surfaces specific to each internalin. As with other LRR proteins, the interaction surfaces of the internalin are likely to be formed by the β-strand regions that constitute the concave faces of these molecules. The concave face has been observed to form the major binding surface in complexes of R1 and U2LRR with their target proteins (41, 43) and is inferred to form the binding surface for rnalp (44). The sides of R1 adjacent to the concave face are also involved in binding. In InIB, the concave face contains separate patches of hydrophobic and negatively charged residues that could constitute protein–protein interaction surfaces (39).

Sequence characteristics of the internalin also point to the concave face as being important to protein–protein interactions. Three-quarters of the hydrophobic residues predicted to be surface exposed in the L. monocytogenes internalin are located on the concave face (positions 3, 4, 6, 8, and 9) (Fig. 1). Likewise, greater than three-quarters of the negatively charged residues predicted to be surface exposed are also located on the concave face (Fig. 1). In general, the concave faces of the internalin possess the hydrophobic and negative charge characteristics observed for InIB. The sides adjacent to the concave face (positions 11 and 13 on one side, and positions 21 and 1 on the other) are highly enriched in positively charged residues, as also observed in InIB. In contrast to the distinctive pattern of residues found on the concave face and sides, the convex face (positions 16, 19, and 20) of the internalin appears to be relatively featureless (Fig. 1). The exception to this is InIC, for which the convex face is predicted to contain a small hydrophobic patch. Although the high sequence conservation of the internalin LRR allows strong predictions about protein–protein interaction sites based on the InIB structure, it does not lead as easily to predictions about target specificity.

The challenge for the future will be in identifying binding partners for individual internalin and in determining how the LRR structural motif is used to generate a variety of specific functional interactions.

M.M. was supported by National Institutes of Health Training Grant GM07240-24. This work was in part supported by a W. M. Keck Distinguished Young Scholar award and an American Heart Association grant (9930135N) to P.G.