

## Hypothesis

## A capping domain for LRR protein interaction modules

Hugo Ceulemans<sup>a</sup>, Marc De Maeyer<sup>b</sup>, Willy Stalmans<sup>a</sup>, Mathieu Bollen<sup>a,\*</sup><sup>a</sup>*Afdeling Biochemie, Faculteit Geneeskunde, Katholieke Universiteit Leuven, Leuven, Belgium*<sup>b</sup>*Centrum voor Transgene Technologie en Gentherapie (VIB), Herestraat 49, B-3000 Leuven, Belgium*

Received 7 June 1999; received in revised form 5 July 1999

**Abstract** Leucine-rich repeats (LRR) are protein interaction modules which are present in a large number of proteins with diverse functions. We describe here a novel motif (16–19 residues) downstream of the last, incomplete, LRR in a subfamily of LRR proteins. In the U2A' spliceosomal protein, this motif is folded into a cap that shields the hydrophobic core of the LRRs from the solvent. Modelling of the LRR-cap in the imidazoline-1 candidate receptor, using the known structure of U2A' as template, showed a conservation of the basic structural features.

© 1999 Federation of European Biochemical Societies.

*Key words:* Leucine-rich repeat; sds22

## 1. Introduction

Leucine-rich repeats (LRRs) are protein interaction modules of 20–29 amino acids [1–3]. They are present in proteins with diverse functions and occur in tandem arrays of up to 40 LRRs. Each repeat consists of an 11 residue fragment with the consensus sequence LxxLxLxxN/CxL, followed by a variable stretch of 9–18 residues. The crystal structure of two LRR-containing proteins, i.e. a ribonuclease inhibitor [4] and the U2A' spliceosomal protein [5], has been solved. In these proteins, the invariable 11 residue fragment of the LRRs comprises a short  $\beta$ -strand and a turn at the N/C position, while the variable LRR fragments are largely folded into a helicoidal conformation (Fig. 1). The LRRs are arranged so that all the  $\beta$ -strands and helicoidal structures are parallel to a common axis, resulting in a curved superhelix with the  $\beta$ -strands and helicoidal structures at the concave and convex sides, respectively. The curvature of the superhelix is believed to be determined by the length and pitch of the helicoidal structures.

## 2. Identification of an LRR-associated motif

Sds22, a putative mitotic regulator of protein phosphatase 1 [6], consists largely of a tandem array of 12 LRRs. The last LRR is incomplete, however, and only includes the invariant  $\beta$ -strand. We noticed that some residues in the short sequence following the last LRR of sds22 are conserved between yeast and man (Fig. 2). To explore whether this motif is also present in other LRR proteins, we have used residues 298–360 of human sds22, comprising the last two LRRs and the carboxy-terminal 21 amino acids, as a query to search the

NCBI non-redundant protein sequence database with the PSI-BLAST program [7]. The *E*-value cut-off was set as low as  $10^{-4}$  in order to reduce the influence of the LRR motif. After 10 iterations, convergence was reached. About 30 distinct protein matches with an *E*-value of  $< 10^{-6}$  were identified, originating from all major eukaryotic lineages. Based on homology, they could be ordered in nine clusters (Fig. 2). In addition to proteins with an unknown function, these clusters included polypeptides with a role in RNA processing (U2A' spliceosomal protein [5]), in RNA nuclear export (TAP [8]) or as regulator of protein phosphatases (sds22 [6] and PHAPIa/b [9]). These clusters define a subfamily of LRR proteins that contains a motif of 16–19 residues, 5–9 amino acids downstream of the last, incomplete, LRR and has the consensus sequence YRxx $\phi$ xxx $\phi$ Px $\phi$ xxLD, where  $\phi$  represents any hydrophobic residue. This subfamily of LRR proteins is also characterized by a last, incomplete, LRR that often contains an atypical proline and by a penultimate LRR of medium length (24–27 residues) with some typically spaced hydrophobic residues in the helicoidal fragment. The vast majority of the LRRs in this subfamily can be categorized as either 'sds22-like' or 'typical' in the classification of LRRs that was introduced by Kajava [3].

## 3. Structural and functional features of the LRR-cap

In the U2A' spliceosomal protein, the LRR-associated motif identified above assumes an amphipathic  $\alpha$ -helix, that covers the hydrophobic core of the LRR superhelix, followed by a loop and an extended conformation that is aligned with the  $\beta$ -strand of the last LRR (Fig. 1A). The first (Y) and last residue (D) of the motif are hydrogen-bonded. Since this LRR-associated domain seems to be involved in the shielding of the hydrophobic core of the LRR superhelix from solvent, we propose to designate this domain as the LRR-cap. The least-conserved LRR-capping domain was found in the imidazoline-1 candidate receptor (Fig. 2). However, using the U2A' spliceosomal protein structure as template, we modelled the LRR-cap of the imidazoline-1 candidate receptor with the Swiss-pdb-Viewer program [10]. We found that the basic features of the LRR-cap were conserved and that the variability was mainly situated in the loop segment (Fig. 1B).

Deletion of the sequence corresponding to the LRR-cap resulted in a nuclear exclusion of sds22 in yeast [6]. Combined with observations that all other LRR-cap-containing proteins investigated thus far (U2A' spliceosomal protein, TAP, PHAPIa/b) are nuclear, this suggests that the LRR-capping domain may be required for the nuclear targeting. It remains to be explored whether this putative nuclear targeting function of the LRR-cap involves the shielding of the hydrophobic

\*Corresponding author. Fax: (32) (16) 345995.  
E-mail: mathieu.bollen@med.kuleuven.ac.be

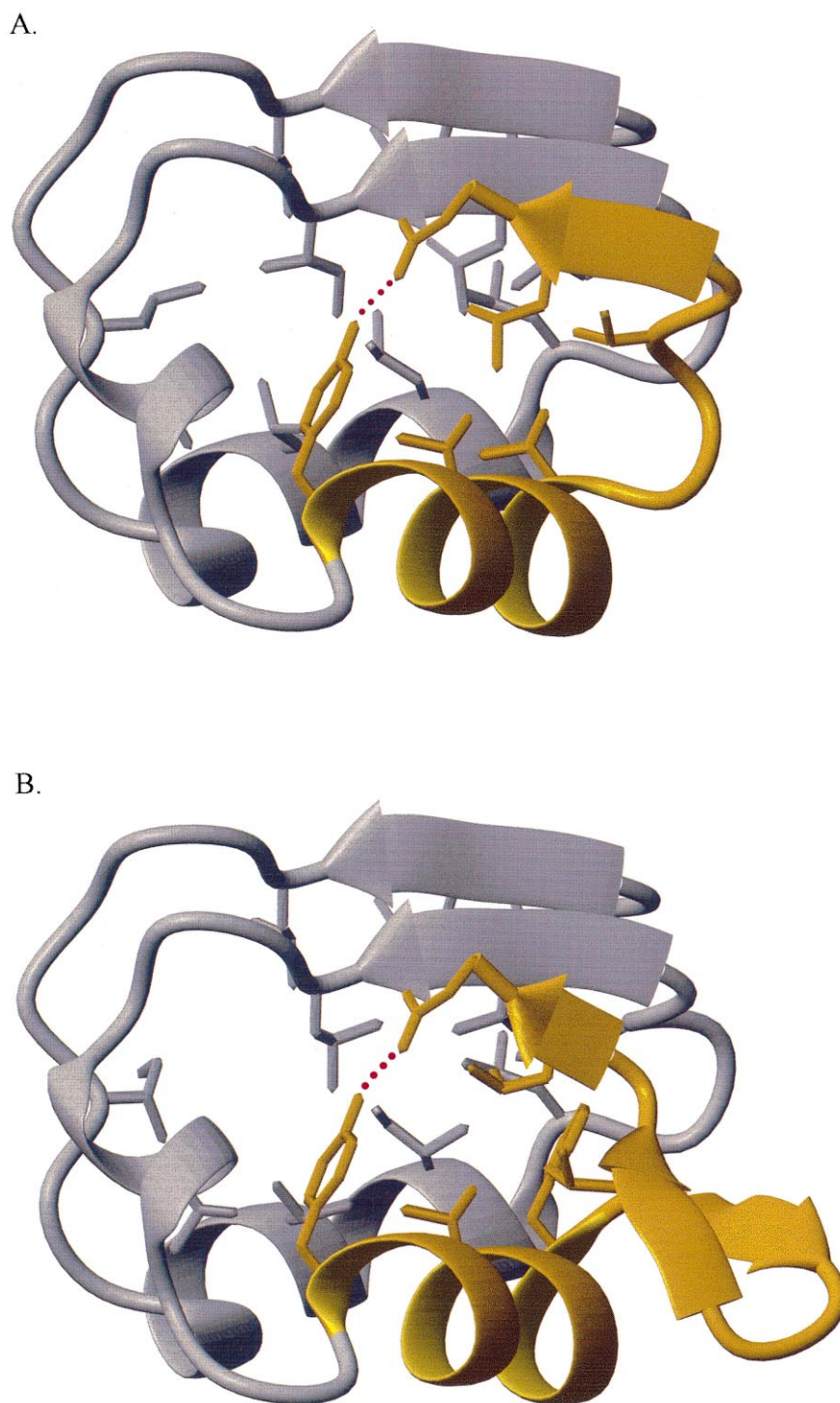


Fig. 1. (A) Backbone of the carboxy-terminus of the U2A' spliceosomal protein. Shown are the last two LRRs (gray) plus the LRR-cap (yellow). (B) Backbone of the last two LRRs (gray) plus the LRR-cap (yellow) of the imidazoline-1 candidate receptor. Arrows correspond to  $\beta$ -strands and ribbons to helices, as calculated with the Kabsch and Sander algorithm. Also shown are the side chains of the conserved hydrophobic residues and of the hydrogen-bonded (...) tyrosine and the C-terminal aspartic acid in the LRR-cap. This figure was produced using the Molmol program [11].

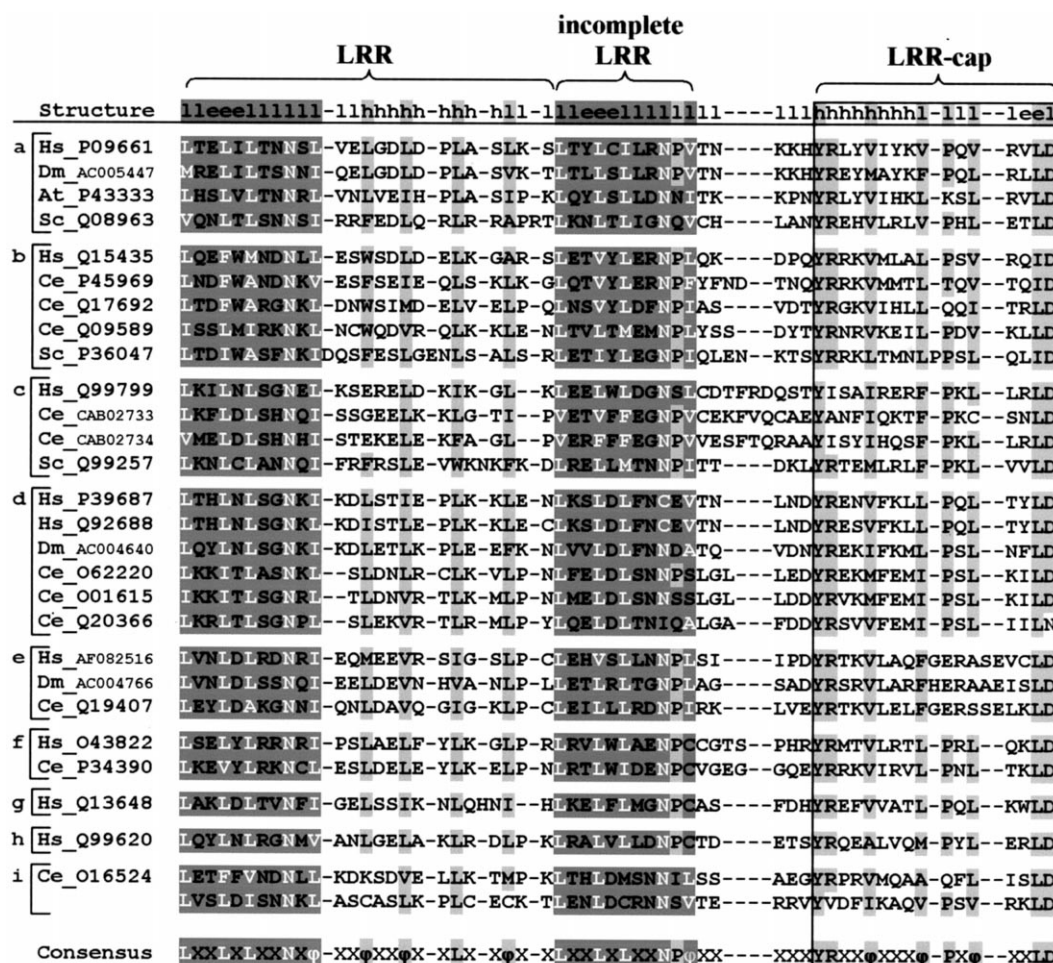


Fig. 2. Multiple alignment of protein clusters containing a LRR-cap. The alignment was constructed using the PSI-BLAST search results and the outcome of homology searches against translated DNA sequences. The figure shows only sequences from *Homo sapiens* (Hs), *Drosophila melanogaster* (Dm), *Caenorhabditis elegans* (Ce), *Arabidopsis thaliana* (At) and *Saccharomyces cerevisiae* (Sc). The sequences are identified with their SwissPROT/TrEMBL, EMBL or GenBank accession numbers. Capitals represent amino acids, X stands for any amino acid and φ for any hydrophobic residue. The top shows the secondary structure (h, helical; e, β-extended; l, loop conformation), as derived from the crystal structure of the U2A' spliceosomal protein (pdb # 1A9N). The LRR β-strands are depicted in dark gray boxes, with the conserved residues in white. The LRR-caps are framed. Conserved residues typical of the LRR-cap subfamily are depicted in light gray boxes. The square brackets delineate clusters of homologous proteins: a, U2A' spliceosomal protein; b, sds22; c, TAP; d, PHAPIa/b; e, imidazoline-1 candidate receptor; f, YF5/A2; g, TSLRP; h, B7; i, unknown (the latter protein contains two LRR domains and two LRR-caps).

core of the LRR superhelix or depends on its interaction with other protein(s).

## References

- [1] Kobe, B. and Deisenhofer, J. (1994) Trends Biochem. Sci. 19, 415–421.
- [2] Buchanan, S.G.St.C. and Gay, N.J. (1996) Prog. Biophys. Mol. Biol. 65, 1–44.
- [3] Kajava, A.V. (1998) J. Mol. Biol. 277, 519–527.
- [4] Kobe, B. and Deisenhofer, J. (1995) Nature 374, 183–186.
- [5] Price, S.R., Evans, P.R. and Nagai, K. (1998) Nature 394, 645–650.
- [6] Stone, E.M., Yamano, H., Kinoshita, N. and Yanagida, M. (1993) Curr. Biol. 3, 13–26.
- [7] Altschul, S.F., Madden, T.L., Schäffer, A.A., Zhang, J., Zhang, Z., Miller, W. and Lipman, D.J. (1997) Nucleic Acids Res. 25, 3389–3402.
- [8] Gruter, P., Tabernero, C., von Kobbe, C., Schmitt, C., Saavedra, C., Bach, A., Wilm, M., Felber, B.K. and Izaurralde, E. (1998) Mol. Cell 1, 649–659.
- [9] Li, M., Makinje, A. and Damuni, Z. (1996) Biochemistry 35, 6998–7002.
- [10] Guex, N. and Peitsch, M.C. (1997) Electrophoresis 18, 2714–2723.
- [11] Koradi, R., Billeter, M. and Wuthrich, K. (1996) J. Mol. Graph. 14, 29–32.