23 March, 2012

Mini-review, Biochemical Society Transactions, 2012

LRRK2: function and dysfunction

Analysis of LRRK2 accessory repeat domains: prediction of repeat length, number, and sites of Parkinson's disease mutations

Ryan D Mills[1,2], Terrence D Mulhern[1], Heung-Chin Cheng[1], Janetta G Culvenor[2]

[1]Department of Biochemistry and Molecular Biology, Bio21 Molecular Science and Biotechnology Institute, and [2]Department of Pathology, The University of Melbourne, Victoria 3010, Australia.

Corresponding author:

Janetta G. Culvenor, j.culvenor@unimelb.edu.au

Department of Pathology, The University of Melbourne, Victoria 3010, Australia.

Phone: 61 3 8344 3991

Fax: 61 3 8344 4004

Abstract

Various investigators have identified the major domain organisation of LRRK2, which includes a GTPase Roc domain (ras of complex proteins) followed by a COR domain (C-terminal of Roc) and a protein kinase domain. In addition, there are four domains composed of structural repeat motifs likely to be involved in regulation and localization of this complex protein. Here, we report our bioinformatic analyses of the human LRRK2 amino acid sequence to predict the repeat size, number and likely boundaries for the armadillo repeat, ankyrin repeat, the leucine-rich repeat, and WD-40 repeat regions of LRRK2. Homology modelling using known protein structures with similar domains was used to predict structures, exposed residues and location of mutations for these repeat regions. We predict that the armadillo repeats, ankyrin repeats, and leucine-rich repeats together form an extended N-terminal flexible "solenoid"-like structure composed of tandem repeat modules likely to be important in anchoring to the membrane and cytoskeletal structures as well as binding to other protein ligands. Near the C-terminus of LRRK2 the WD-40 repeat region is predicted to form a closed propeller structure important for protein complex formation.

Key words: LRRK2, armadillo, ankyrin, leucine-rich repeat, WD-40, GTPase, protein kinase, Parkinson's disease

Abbreviations used: LRRK2, leucine-rich repeat kinase 2; PSI-BLAST, Position-Specific Iterative Basic Local Alignment Search Tool; PDB, Protein Data Bank.

Introduction

In 2004 two landmark reports first identified mutations in the large multi-domain leucine-rich repeat kinase two (LRRK2) as causing late-onset autosomal-dominant Parkinson's disease [1, 2]. Various investigations have also indicated the presence of armadillo, ankyrin and WD-40 repeat regions in addition to the two catalytic domains, a Roc GTPase (ras of complex proteins) and a protein kinase domain, which are separated by a COR (C-terminal of Roc) domain [3]. However, a detailed analysis of all the repeat modules in LRRK2 is lacking. We have undertaken varied protein sequence bioinformatic analyses to predict the likely repeat lengths, number of repeat segments and likely boundaries for these modular domains. We then undertook homology modelling to predict the expected molecular shape and sites of several pathogenic mutations for the repeat domains.

Fig.1 summarizes the predicted arrangement for the LRRK2 domains with approximate repeat lengths and number.

Armadillo repeats form an extended super helix

Marin (2006) characterised a block of LRRK2-specific repeats at the N-terminus [4]. Our analysis suggests these repeats are variants of the armadillo repeat structure. Following alignment of LRRK2 orthologs to select a LRRK2-consensus repeat sequence, we identified 13 putative armadillo-type repeats after application of the Position-Specific Iterative Basic Local Alignment Search Tool (PSI-BLAST). For LRRK2 we predict 13 tandem multiples of about 42 residues, similar to the well characterized β-catenin 42 amino acid repeats [5] (Fig. 2A). These repeats typically contain three alpha-helix segments as shown in Fig.2, a short helix1 followed by two longer helices. In β-catenin helix 2 and helix 3 pack against each other in an anti-parallel mode. These structurally-conserved, sequence-degenerate repeats are predicted to make up the largest LRRK2 domain of about 600 amino acids. Homology modelling indicates an extended super helix with a hydrophobic core (Fig 3B); based on elements of delta-catenin, PDB code 3L6X and importin-alpha structures PDB code 1EE4 [6, 7]. The Parkinson's E334K mutation occurs on a predicted extended loop between armadillo repeats 6 and 7 (Fig.3B). Like β-catenin the LRRK2 armadillo domain is likely to have a shallow groove lined with positively charged residues, such as histidines from helix 2 and 3 of LRRK2 armadillo repeats.

Ankyrin repeats form gently curved extended structures

Similar to cytoskeletal erythrocyte ankyrin, we predict the LRRK2 protein also contains a set of 33 amino acid structural tandem repeats. Seven repeats were predicted with a profile hidden Markov model and PSI-BLAST using a set of known ankyrin domains (Fig. 2B). The LRRK2 consensus repeat is compared with human ankyrinR consensus sequence described by Michaely et al., 2002 (PDB code 1N11) [8]. Ankyrin domains are characterised by two anti-parallel alpha-helices followed by a loop to form a gently curved structure. The predicted structure shown in Fig. 3B for LRRK2 ankyrin domain is modelled after the ankyrin repeat domain of ribonuclease L (PDB 1WDY) [9]. In this predicted structure, the Parkinson disease-associated mutation R793M is exposed on the surface of this domain on helix 2 of the putative ankyrin repeat 4. Thus, we predict that this mutation alters the surface charge of this domain. It is unclear whether this mutation affects the structural integrity and/or the protein-protein interaction ability of the putative ankyrin repeats domain of LRRK2.

Leucine-rich repeats form horse-shoe curved structures

Leucine-rich repeats (LRR) commonly occur as tandem repeats about 25 amino acids in length in multiples of 2-30 to form a solenoid/horse-shoe shaped structure as first shown

for ribonuclease inhibitor [10]. For LRRK2, 14 leucine-rich repeats of about 24 amino acids were predicted by conformity with the consensus sequence shown at the bottom of Fig. 2C, and by profile analysis using a pool of 743 structurally verified LRR repeats (REP V1.1, [11]). A common feature is an 11-residue sequence (LxxLxLxxNxL) in which the leucine and asparagine residues may be substituted by other hydrophobic residues. Fig. 3B depicts a predicted structure of the LRRK2 LRR domain modelled after parts of the known structures with similar repeat lengths (PDB codes: 1OGQ, 1ZIW, 2FT3). In this structure, the LRRK2 LRR domain contains 14 LRRs, which are grouped into two modules. Module 1 and Module 2 formed by LRR1-LRR7 and LRR9-LRR14, respectively, are linked together by a shorter LRR8. The putative N-terminal extended capping LRR 1 is shown at the top of Fig. 2C. In known structures each LRR subunit contributes a β-strand at the concave face to form a curved β-sheet, while the convex face consists of a variety of secondary structures with alpha-helices and turns and connecting loops [12]. LRRK2 ortholog-conserved leucine repeat "position 8" basic residues in LRR 4 (R1067) and LRR 7 (K1138) at the concave face are likely to be involved in functional ligand interactions. The LRR domain mutation R1067Q mapped to the putative LRR 4 is predicted to alter the surface charge of the concave face in the predicted structure (Fig. 3B).

A recent study of this LRR domain by Vancraenenbroeck *et al.* also predicted 14 LRR repeats for this LRRK2 region with a similar structural-repeat alignment and similar homology structural prediction [13].

WD-40 repeat modules form closed β-sheet propeller-like structures

WD-40 repeats are so-named after their common length of 40 residues and for containing an unusual tryptophan-aspartic acid motif. They were first identified in LRRK2 by Paizan-Ruiz and Zimprich and their co-workers [1]. Sequence alignment of LRRK2 with F-box/WD repeat-containing protein 7, facilitated prediction of seven WD-40 repeats in LRRK2 (Fig 3 A). A WD-40 consensus sequence from Li and Roberts, 2001 [14] is included below Fig. 3A. These repeats of the standard 40 residue length are expected to adopt a closed formation of β-sheet structures arranged radially to comprise a propeller-like shape (Fig. 3B). Each WD-40 repeat generally forms four β-strands named from A-D (Fig. 3B). WD-40 repeats generally occur as multiples of 6-7 due to packing constraints of the closed structure. Several inserts were identified occurring between predicted β-strands, including an extended insert after WD-40 repeat 5, which may form an exposed surface for protein interactions. Structural modelling of the LRRK2 WD-40 repeat domain utilising alignment with the F-box/WD40 repeat protein 7 (PDB 2OVP) [15] is shown in Fig. 3B. A similar structure was predicted by Jorgensen et al., for this domain which indicated a positively charged central cleft [16].

The Parkinson's disease mutations G2385R and T2356I were mapped to elements of WD-40 β-sheet 5. These mutations may alter LRRK2 interaction with other proteins, ligands, LRRK2 domains, modify LRRK2 substrate recruitment or dimerisation.

In conclusion:

Detailed analysis of LRRK2 predicted repeat domains indicate that the protein is aptly named as "leucine-rich" with the N-terminal half composed of armadillo, ankyrin and leucine-rich repeat modules all containing many hydrophobic leucine residues. These N-terminal repeats are anticipated to form an extended flexible structure with a hydrophobic core exhibiting likely membrane and cytoskeletal anchoring and regulatory functions. The WD-40 repeat domain C-terminal to the catalytic domains may contribute to LRRK2 binding to other cellular proteins to form protein complexes and regulation of the catalytic activities

of LRRK2. Homology modelling of repeat domains indicated that several pathogenic Parkinson's mutations may interfere with charge distribution on interaction surfaces. Testing of sequence-structure predictions described to date, awaits experimentally determined structures of the LRRK2 domains to further improve understanding of LRRK2 molecular function and interactions.

Acknowledgements:

We thank the Australian National Health and Medical Research Council [project grant number 566743] and Australian Brain Foundation for financial support. Ryan Mills was supported by an Australian Postgraduate Research Award, a Dowd Neuroscience fellowship and an Australian Institute of Nuclear Science and Engineering award.

References:

1 Paisán-Ruíz, C., Jain, S., Evans, E. W., Gilks, W. P., Simón, J., van der Brug, M., López de Munain, A., Aparicio, S., Gil, A. M., Khan, N., Johnson, J., Martinez, J. R., Nicholl, D., Carrera, I. M., Pena, A. S., de Silva, R., Lees, A., Martí-Massó, J. F., Pérez-Tur, J., Wood, N. W. and Singleton, A. B. (2004) Cloning of the gene containing mutations that cause PARK8-linked Parkinson's disease. Neuron. **44**, 595-600

2 Zimprich, A., Biskup, S., Leitner, P., Lichtner, P., Farrer, M., Lincoln, S., Kachergus, J., Hulihan, M., Uitti, R. J., Calne, D. B., Stoessl, A. J., Pfeiffer, R. F., Patenge, N., Carbajal, I. C., Vieregge, P., Asmus, F., Müller-Myhsok, B., Dickson, D. W., Meitinger, T., Strom, T. M., Wszolek, Z. K. and Gasser, T. (2004) Mutations in LRRK2 cause autosomal-dominant parkinsonism with pleomorphic pathology. Neuron. **44**, 601-607

3 Mata, I. F., Wedemeyer, W. J., Farrer, M. J., Taylor, J. P. and Gallo, K. A. (2006) LRRK2 in Parkinson's disease: protein domains and functional insights. Trends Neurosci. **29**, 286-293

4 Marín, I. (2006) The Parkinson disease gene LRRK2: evolutionary and structural insights. Mol Biol Evol. **23**, 2423-2433

5 Huber, A. H., Nelson, W. J. and Weis, W. I. (1997) Three-dimensional structure of the armadillo repeat region of beta-catenin. Cell. **90**, 871-882

6 Ishiyama, N., Lee, S. H., Liu, S., Li, G. Y., Smith, M. J., Reichardt, L. F. and Ikura, M. (2010) Dynamic and static interactions between p120 catenin and E-cadherin regulate the stability of cell-cell adhesion. Cell. **141**, 117-128

7 Conti, E. and Kuriyan, J. (2000) Crystallographic analysis of the specific yet versatile recognition of distinct nuclear localization signals by karyopherin alpha. Structure. **8**, 329-338

8 Michaely, P., Tomchick, D. R., Machius, M. and Anderson, R. G. (2002) Crystal structure of a 12 ANK repeat stack from human ankyrinR. Embo J. **21**, 6387-6396

9 Tanaka, N., Nakanishi, M., Kusakabe, Y., Goto, Y., Kitade, Y. and Nakamura, K. T. (2004) Structural basis for recognition of 2',5'-linked oligoadenylates by human ribonuclease L. EMBO J. **23**, 3929-3938

10 Kobe, B. and Deisenhofer, J. (1993) Crystallization and preliminary X-ray analysis of porcine ribonuclease inhibitor, a protein with leucine-rich repeats. J Mol Biol. **231**, 137-140

11 Andrade, M. A., Ponting, C. P., Gibson, T. J. and Bork, P. (2000) Homology-based method for identification of protein repeats using statistical significance estimates. J Mol Biol. **298**, 521-537

12 Bella, J., Hindle, K. L., McEwan, P. A. and Lovell, S. C. (2008) The leucine-rich repeat structure. Cell Mol Life Sci. **65**, 2307-2333

13 Vancraenenbroeck, R., Lobbestael, E., Weeks, S. D., Strelkov, S. V., Baekelandt, V., Taymans, J. M. and De Maeyer, M. (2012) Expression, purification and preliminary biochemical and structural characterization of the leucine rich repeat namesake domain of leucine rich repeat kinase 2. Biochim Biophys Acta. **1824**, 450-460

14 Li, D. and Roberts, R. (2001) WD-repeat proteins: structure characteristics, biological function, and their involvement in human diseases. Cell Mol Life Sci. **58**, 2085-2097

15 Hao, B., Oehlmann, S., Sowa, M. E., Harper, J. W. and Pavletich, N. P. (2007) Structure of a Fbw7-Skp1-cyclin E complex: multisite-phosphorylated substrate recognition by SCF ubiquitin ligases. Mol Cell. **26**, 131-143

16 Jorgensen, N. D., Peng, Y., Ho, C. C., Rideout, H. J., Petrey, D., Liu, P. and Dauer, W. T. (2009) The WD40 domain is required for LRRK2 neurotoxicity. PLoS One. **4**, e8463

17 Andrade, M. A., Petosa, C., O'Donoghue, S. I., Müller, C. W. and Bork, P. (2001) Comparison of ARM and HEAT protein repeats. J Mol Biol. **309**, 1-18

Figure Legends:

Fig. 1 Arrangement of LRRK2 domains. Predicted size and number of repeat subunits for repeat domains are summarized to scale.

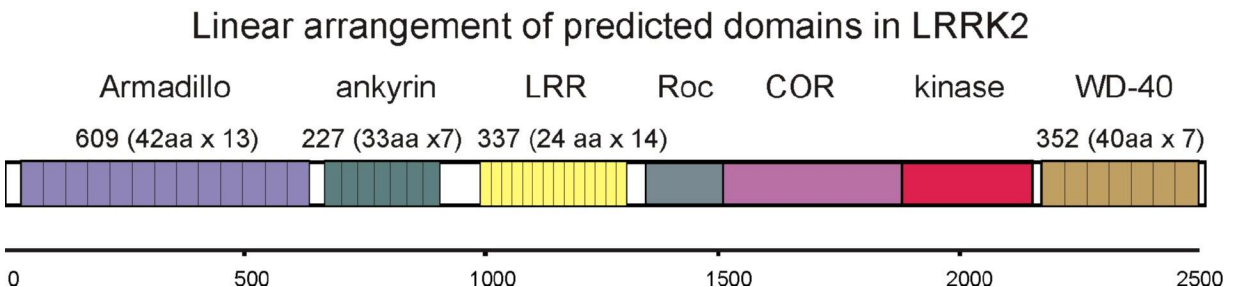
Fig. 2 Structural sequence alignments for putative repeats of LRRK2 armadillo, ankyrin, and leucine-rich repeat domains. Helix regions are shaded green, β-strand region is shaded yellow, sites of Parkinson's mutations boxed and in red text. A: LRRK2 consensus is compared with the armadillo consensus [17]. B: A LRRK2 ankyrin consensus is compared with consensus from ankyrinR [8]. C: The LRR consensus is shown below the figure[12]; LRR 1 is longer than standard, and repeat 8 shorter. The β-strand region is shaded yellow.

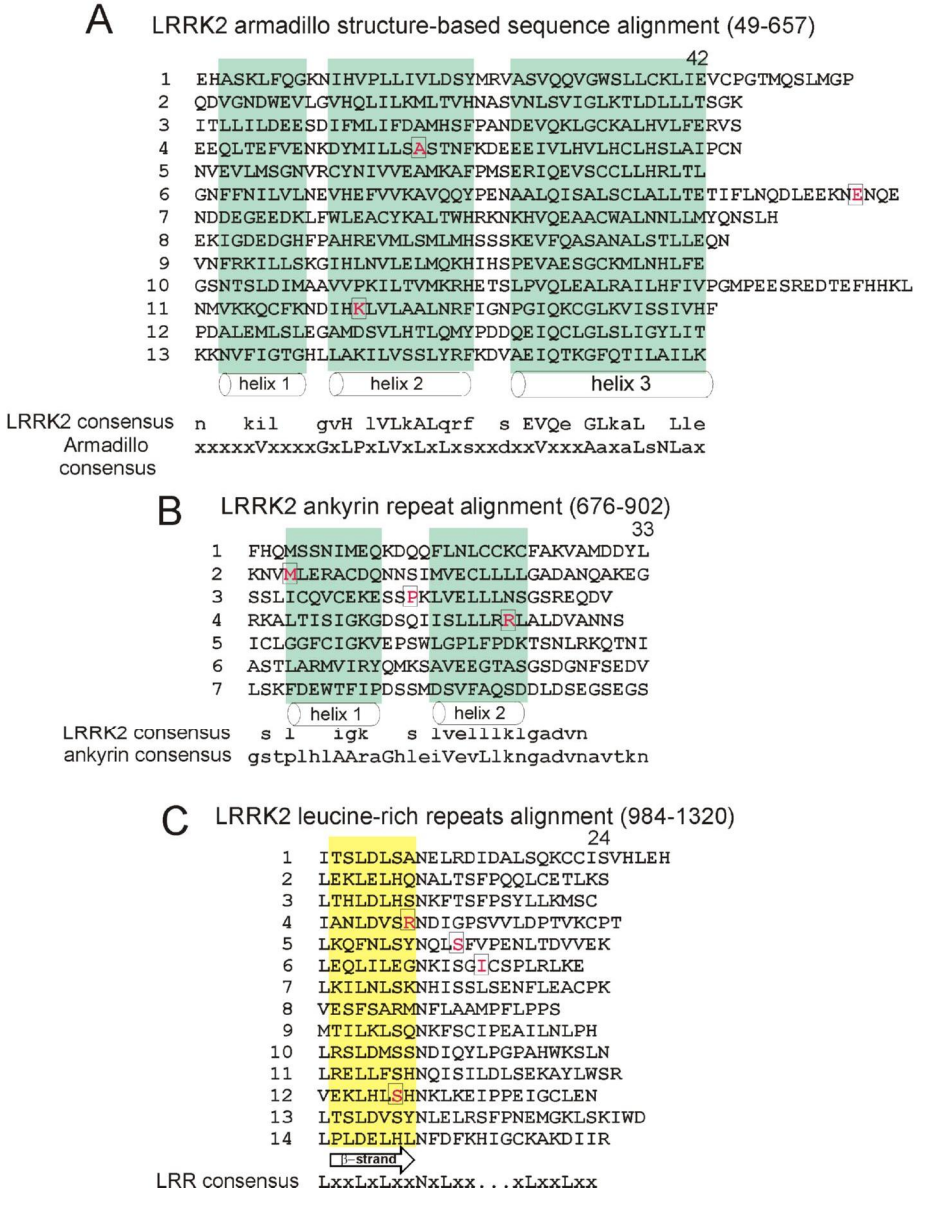
Fig. 3 A: LRRK2 WD-40 structural-based sequence alignment with β-strands shaded yellow.

Several predicted inserts are in pink shading. Sites of Parkinson's mutations are boxed. A consensus for WD-40 repeats is given below [14].

B: Homology models for LRRK2 repeat domains. Domains were modelled using known structures with similar repeat lengths, as described in text. Sites of some Parkinson's mutations are indicated.

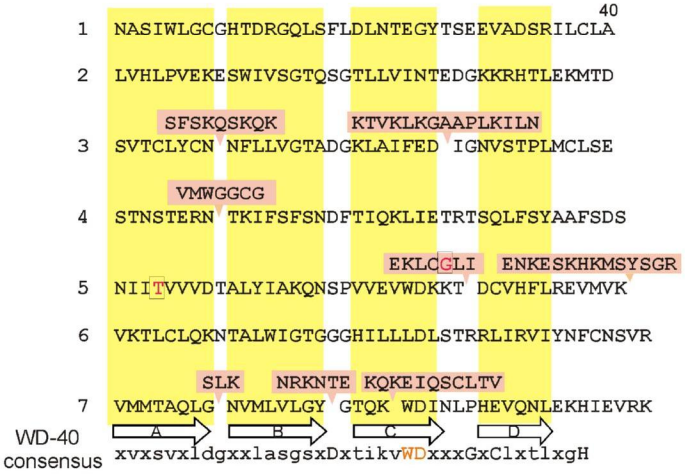
Fig 1 Mills et al, 2012




A
LRRK2 armadillo structure-based sequence alignment (49-657)
42
1 EHASKLFQGKNIHVPLLIVLDSYMRVASVQQVGWSLLCKLIEVCPGTMQSLMGP
2 QDVGNDWEVLGVHQLILKMLTVHNASVNLSVIGLKTLDLLLTSGK
3 ITLLILDEESDIFMLIFDAMHSFPANDEVQKLGCKALHVLFERVS
4 EEQLTEFVENKDYMILLSASTNFKDEEEIVLHVLHCLHSLAIPCN
5 NVEVLMSGNVRCYNIVVEAMKAFPMSERIQEVSCCLLHRLTL
6 GNFFNILVLNEVHEFVVKAVQQYPENAALQISALSCLALLTETIFLNQDLEEKNENQE
7 NDDEGEEDKLFWLEACYKALTWHRKNKHVQEAACWALNNLLMYQNSLH
8 EKIGDEDGHFPAHREVMLSMLMHSSSKEVFQASANALSTLLEQN
9 VNFRKILLSKGIHLNVLELMQKHIHSPEVAESGCKMLNHLFE
10 GSNTSLDIMAAVVPKILTVMKRHETSLPVQLEALRAILHFIVPGMPEESREDTEFHHKL
11 NMVKKQCFKNDIHKLVLAALNRFIGNPGIQKCGLKVISSIVHF
12 PDALEMLSLEGAMDSVLHTLQMYPDDQEIQCLGLSLIGYLIT
13 KKNVFIGTGHLLAKILVSSLYRFKDVAEIQTKGFQTILAILK
helix 1
helix 2
helix 3
LRRK2 consensus n kil gvH lVLkALqrf s EVQe GLkaL Lle
Armadillo consensus xxxxxVxxxxGxLPxLVxLxLxsxxdxxVxxxAaxaLsNLax
B
LRRK2 ankyrin repeat alignment (676-902)
33
1 FHQMSSNIMEQKDQQFLNLCCKCFAKVAMDDYL
2 KNVMLERACDQNNSIMVECLLLLGADANQAKEG
3 SSLICQVCEKESSPKLVELLLNSGSREQDV
4 RKALTISIGKGDSQIISLLLRRLALDVANNS
5 ICLGGFCIGKVEPSWLGPLFPDKTSNLRKQTNI
6 ASTLARMVIRYQMKSAVEEGTASGSDGNFSEDV
7 LSKFDEWTFIPDSSMDSVFAQSDDLDSEGSEGS
helix 1
helix 2
LRRK2 consensus s l igk s lvelllklgadvn
ankyrin consensus gstplhlAAraGhleiVevLlkngadvnavtkn
C
LRRK2 leucine-rich repeats alignment (984-1320)
24
1 ITSLDLSANELRDIDALSQKCCISVHLEH
2 LEKLELHQNALTSFPQQLCETLKS
3 LTHLDLHSNKFTSFPSYLLKMSC
4 IANLDVSRNDIGPSVVLDPTVKCPT
5 LKQFNLSYNQLSFVPENLTDVVEK
6 LEQLILEGNKISGICSPLRLKE
7 LKILNLSKNHISSLSENFLEACPK
8 VESFSARMNFLAAMPFLPPS
9 MTILKLSQNKFSCIPEAILNLPH
10 LRSLDMSSNDIQYLPGPAHWKSLN
11 LRELLFSHNQISILDLSEKAYLWSR
12 VEKLHLSHNKLKEIPPEIGCLEN
13 LTSLDVSYNLELRSFPNEMGKLSKIWD
14 LPLDELHLNFDFKHIGCKAKDIIR
β-strand
LRR consensus LxxLxLxxNxLxx...xLxxLxx

A

LRRK2 WD-40 structure-based sequence alignment (2164-2515)



B Homology models of predicted repeat domains in LRRK2