Prediction of the repeat domain structures and impact of parkinsonism-associated variations on structure and function of all functional domains of Leucine rich repeat kinase 2 (LRRK2)

Ryan D Mills¹,², Terrence D Mulhern¹, Fei Liu³, Janetta G Culvenor² and Heung-Chin Cheng¹

¹Department of Biochemistry and Molecular Biology, Bio21 Molecular Science and Biotechnology Institute, and ²Department of Pathology, University of Melbourne, Parkville, Victoria 3010, Australia. ³Department of Chemistry & Biomolecular Sciences, Macquarie University, NSW 2109, Australia.

Address all correspondence to:
Heung-Chin Cheng
Department of Biochemistry & Molecular Biology, Bio21 Institute
University of Melbourne
Parkville, Victoria 3010, Australia
Phone: 61 3 83442254
E-mail: heung@unimelb.edu.au

Key words: Protein kinase, protein interactions, phosphorylation, neurotoxicity, Parkinson disease

Grant support: National Health and Medical Research Council of Australia (NHMRC Project Grant No.566743 to JGC and HCC)
Abstract

Genetic variations of leucine-rich repeat kinase 2 (LRRK2) are the major cause of dominantly inherited Parkinson's disease (PD). LRRK2 protein contains seven predicted domains: a tandem Ras-like GTPase (ROC) domain and C-terminal of Roc (COR) domain, a protein kinase domain and four repeat domains. PD-causative variations arise in all domains, suggesting that aberrant functioning of any domain can contribute to neurotoxic mechanisms of LRRK2. Determination of the three-dimensional structure of LRRK2 is one of the best avenues to decipher its neurotoxic mechanism. However, with the exception of the Roc domain, the three-dimensional structures of the functional domains of LRRK2 have yet to be determined. Based upon the known three-dimensional structures of repeat domains of other proteins, the tandem Roc-COR domains of the C. tepidum Rab family protein, and the kinase domain of the D. discoideum Roco4 protein, we predicted (i) the motifs essential for protein-protein interactions in all domains, (ii) the motifs critical for catalysis and substrate recognition in the tandem Roc-COR and kinase domains, and (iii) the effects of some PD-associated missense variations on the neurotoxic action of LRRK2. Results of our analysis provide a conceptual framework for future investigation into the regulation and the neurotoxic mechanism of LRRK2.

(200 words)
INTRODUCTION

Since the initial discovery of the involvement of leucine-rich repeat kinase 2 (LRRK2; OMIM accession number: 609007) in familial Parkinsonism, the challenge has been to identify how this protein contributes to pathogenesis. Research is spurred by the possibility that LRRK2 dysfunction may underlie not only the familial Parkinsonism, but also common, sporadic Parkinson's disease (PD). The ultimate goal is to understand LRRK2 sufficiently well to allow intervention strategies either targeted to LRRK2 itself or associated cellular signaling components. Progress made to date is summarized well by previous reports (Cookson, 2010; Daniëls, et al., 2011a), and a good overview of the genetic data in the context of parkinsonism-associated genes is provided by Nuytemans et al. (Nuytemans, et al., 2010).

The first aim of this manuscript is to provide a detailed assessment of the possible functions of each of the seven putative domains of LRRK2, and how their structure and function may be affected by parkinsonism-associated variations. This is supported by structural predictions of armadillo, ankyrin, leucine-rich repeat (LRR) and WD-40 repeat domains with procedures detailed in Supplemental Information Part 1 (Figures 1A, 1C, 2A and 5A). Although the three-dimensional structures of the Roc-COR and kinase domains have not been solved, we are able to use the recently determined three-dimensional structures of the LRRK2 Roc domain (Deng, et al., 2008a), the tandem Roc and COR domains of Chlorobium tepidum Roco protein (Gotthardt, et al., 2008a) and the Dictyostelium discoideum Roco4 kinase domain (Gilsbach, et al., 2012) as the models to predict the regulatory and catalytic properties of the LRRK2 Roc-COR and kinase domains (Figures 3 and 4). To construct the models of the four repeat domains of LRRK2 shown in Figures 1B, 1D, 2B and 5B, we searched through structural databases for homologous structures containing repeat lengths similar to those predicted in LRRK2 and structures containing insert regions of similar length to those predicted in LRRK2 (Supp. Table S3-6 in Supplemental Information Part 2). Some features of the models have been published previously in a conference proceeding (Mills, et al., 2012). The full details of how the models were generated are described in Supplemental Information Part 1. The coordinates of the amino acid residues in the models are provided for use by readers to examine the models using viewing programs for the three-dimensional structures of proteins (Supplemental Information Part 3).
Genetic and biochemical studies revealed that PD-associated variations in LRRK2 promote neuronal death through a mechanism dependent upon both the kinase activity (Greggio, 2012; Greggio, et al., 2006; Lee, et al., 2010a; Ramsden, et al., 2011; Smith, et al., 2006c; West, et al., 2005) and guanine nucleotide-binding ability (Xiong, et al., 2012a). Furthermore, results of in vitro studies suggest that the guanine nucleotide-binding ability plays a significant role in regulation of LRRK2 kinase activity (Biosa, et al., 2013), suggesting functional interactions between the kinase domain and Roc-COR domains is crucial to the neurotoxic action of LRRK2. In addition to these interactions, interactions between the kinase domain and other functional domains are potentially important for the neurotoxic action of LRRK2 (Greggio, et al., 2006; Sheng, et al., 2012; Smith, et al., 2006c; West, et al., 2007). Of note, autophosphorylation of Ser-1292 located in the LRR domain plays a critical role in modulating the neurotoxicity caused by some PD-associated variations of LRRK2 (Supp. Figure S1, Supp. Tables S1 and S2) (Sheng, et al., 2012). For example, autophosphorylation at Ser-1292 is significantly elevated in the LRRK2 mutant carrying the neurotoxic PD-associated G2019S variation in the kinase domain. These findings have two implications. First, the LRR domain is critical to the neurotoxic action of the mutated LRRK2. Second, by autophosphorylating Ser-1292, LRRK2 kinase domain may modulate the protein-protein interaction activity of the LRR domain of LRRK2. In light of these findings and implications, the second aim of this manuscript is to use the models of the functional domains we present in this manuscript to predict the molecular basis of functional interactions among the various domains in LRRK2 and assess how phosphorylation and PD-associated variations aberrantly modulate these interactions to confer neurotoxicity to the LRRK2 mutants.

Because of the presence of Roc and kinase domains, LRRK2 is both a GTPase and a protein kinase. As a GTPase, LRRK2 is expected to exert its physiological function by interacting with down-stream effector proteins via the Roc domain in a GTP-dependent fashion. Likewise, as a protein kinase, LRRK2 is expected to perform its physiological function by phosphorylating specific cellular proteins. Furthermore, the clearly pathogenic variations including I1371V, R1441H/C/G, Y1669C, G2019S and I2020T found in familial PD patients are all mapped to the Roc domain, the Roc-domain-associated COR domain and the kinase domain (Supp. Figure S1 and Supp. Table S1). Tong et al. generated homozygous [R1441C]LRRK2 knock-in mice and found that they exhibited impairment of the nigrostriatal dopaminergic pathway (Tong, et al., 2009). As R1441 is located in the Roc domain, the findings indicate that aberrant function and/or regulation of the GTPase activity
of LRRK2 can contribute to neurotoxicity. In addition, the kinase activity is indispensable to the neurotoxic action of mutant LRRK2 in neuronal cells, and small-molecule inhibitors targeting the kinase domain of LRRK2 were found to attenuate the neurotoxic action of the recombinant [G2019S]LRRK2 mutant expressed in neurons of a mouse model (Greggio, et al., 2006; Lee, et al., 2010a; Smith, et al., 2006c). All these findings establish the functional importance of dys-regulation of the GTPase and kinase domains in the neurotoxic action of LRRK2 mutants. We therefore use the models of the Roc and COR domains and the kinase domain presented in this manuscript to assess how the pathogenic variations of LRRK2 impact on the functions of the Roc and COR domains and the kinase domain. The final aim of this manuscript is to assess the design of small-molecule inhibitors of the kinase and GTPase domains for use as therapeutics to reduce and/or slow down brain damage in PD patients.

ARMADILLO REPEAT DOMAIN

A structural model of the LRRK2 armadillo repeat domain

Armadillo repeats are units of around 42 amino acids, with each repeat forming three α-helices (named H1 to H3). The repeats stack against each other to form a superhelix, with sides consisting of parallel H2 and H3 helices. Ligands commonly occupy a cleft lined by H3 helices from consecutive armadillo repeats, as observed in structures of armadillo repeat domains of β-catenin and Adenomatous polyposis coli (APC) (Morishita, et al., 2011). In the LRRK2 N-terminal region, 13 repeat units of around 42 amino acids in length were identified (Figure 1). The repeats exhibit spacing of hydrophobic residues similar to characterised armadillo proteins, but also bear features unique to LRRK2 [for additional analysis, see Marín et al. (Marín, 2006)]. A model is shown in Figure 1 including the 13 predicted repeats, three insert regions between the repeats, as well as the locations of two variations found in PD patients.

Armadillo repeat domain: interactions

Cellular proteins that specifically bind to the LRRK2 armadillo repeat domain have yet to be identified. However, some well characterised examples of armadillo repeat domain interactions with Ras-like GTPases exist. A notable example is importin α of which the armadillo repeat domain mediates importin α interactions with specific cellular proteins containing the nuclear localization signals (NLS). The intermolecular armadillo repeat domain/NLS interactions allow importin α to direct nuclear import of the cellular proteins. In
addition to intermolecular interactions, armadillo repeat domain of importin α can also bind intramolecularly with the N-terminal importin β-binding (IBB) domain [(Kobe, 1999) and reviewed in (Lott and Cingolani, 2011)]. This intramolecular interaction controls shuttling of importin α from nucleus to cytosol. Once exported to the cytosol, this intramolecular interaction is relieved when the NLS of a cargo protein displaces the IBB domain from the armadillo repeat domain, allowing the armadillo repeat domain of importin α to load the cargo protein for nuclear translocation (Lott and Cingolani, 2011).

The second example of armadillo repeat domain function relates to regulation of actin polymerization in cells by the armadillo repeat domain-containing protein mDia1. mDia1 induces nucleation and polymerisation of actin monomers to form unbranched actin filaments. In the inactive state, five armadillo repeats near the N-terminus of mDia1 intramolecularly binds to the Diaphanous-autoregulatory domain (DAD domain) at the C-terminus (Alberts, 2001). The binding auto-inhibits its ability to promote nucleation and polymerization of actin monomers. The Ras-like GTPase Rho can disrupt this intramolecular autoinhibition of mDia1 by binding to its armadillo repeat domain in a GTP-dependent manner, allowing the DAD domain of mDia1 to promote formation of unbranched actin filaments (Otomo, et al., 2005).

For LRRK2, its armadillo domain could similarly act as an autoinhibitory domain suppressing its kinase activity, by analogy with mDia1. Assuming a similar mode of binding, crucial residues for interaction with GTPases would occur near the C-terminal end of H3 helices in the LRRK2 armadillo domain. As LRRK2 contains the Roc-COR GTPase tandem domains, it is worthwhile to investigate if the armadillo repeat domain can interact with and modulate the function of the tandem domains.

In addition to intramolecular interactions with other functional domains, the segment encompassing residues 210-310 was predicted by Pandey et al. as a region with a high propensity to undergo aggregation (Pandey, et al., 2012). This segment forms parts of repeats 4 and 6 and the entire repeat 5 in our model of the armadillo repeat domain of LRRK2 (Figure 1A). The recombinant LRRK2 N-terminal segment containing the armadillo repeat and ankyrin repeat domains, formed aggregates when expressed in primary neurons. Intriguingly, its expression in SH-SY5Y cells could attenuate cell death induced by 6-hydroxydopamine treatment (Pandey, et al., 2012). The cytoprotective mechanism of this N-terminal segment of LRRK2 remains to be determined. One possible mechanism is that it acts as a dominant negative mutant which prevents binding of neurotoxic interactor proteins to
intact LRRK2. Nonetheless, the findings highlight the significance of identifying cellular proteins that specifically interact with LRRK2 armadillo repeat domain.

**Armadillo repeat domain: PD-associated variations**

Two potentially pathogenic variations E334K and A419V have been found in the LRRK2 armadillo region. The E334K variation was found in two North American siblings with PD (Supp. Figure S1) (Nichols, et al., 2007; Xiromerisiou, et al., 2007). The E334K variation is predicted to lie within a 16-residue insert region separating repeats 6 and 7 (Figure 1A). A cytoplasmic component of adherens junctions, known as plakophilin-1, contains three insert regions, one of which enforces a substantial bend in its armadillo superhelix via interaction with the flanking armadillo repeats (Choi and Weis, 2005). If the 6-7 insert is similarly important for the overall shape of the LRRK2 armadillo domain, charge reversal due to the E334K variation within its predicted site of interaction with armadillo repeats 6 and 7 could have far-reaching consequences. This could be explored by comparison of E334K mutant LRRK2 with wild-type and constructs bearing the engineered point variations, to show whether disruption of the armadillo structure at the 6-7 insert can affect LRRK2 function.

The A419V variation was found in a number of studies in Asian individuals. Significant associations of this variation with development of PD were observed (Ross, et al., 2011). A419 is predicted to be located in the H3 helix of the armadillo repeat 8. Its change to the hydrophobic valine may perturb interactions between the H3 helix of repeat 8 with the H3 helix of repeat 9.

**ANKYRIN REPEAT DOMAIN**

**A structural model of the LRRK2 ankyrin repeat domain**

Ankyrin domains are made up of arrays of repeating units, each commonly 33 residues long and comprising two α-helices (termed H1 and H2) followed by a loop region, altogether forming a gently curved superstructure [reviewed in Li et al. (Li, et al., 2006)]. Our prediction reveals that LRRK2 ankyrin repeats 3 and 4 are 30 and 31 residues, respectively, shorter than the five other predicted 33-residue LRRK2 ankyrin repeats (Figure 1C). Based on a 30-residue ankyrin repeat at the C-terminus of the Bcl-3 ankyrin domain, the missing residues in ankyrin repeat 3 of LRRK2 may similarly enforce an atypical structure, with a loop section replacing most of H1 and connecting directly to H2.
Ankyrin repeat domain: interactions

One example of an ankyrin domain interaction potentially relevant to LRRK2 is the ability of ankyrin repeats to bind and inhibit protein kinase domains. In the crystal structure of the heterodimeric complex of the cyclin-dependent kinase 4 inhibitor (INK4) with cyclin-dependent kinase 6 (Cdk6), p16-INK4a employs its concave surface and H1-H2 linkers to contact both the N- and C-terminal lobes of the kinase domain of Cdk6, enforcing a 15º rotation between the lobes to form an inactive conformation (Russo, et al., 1998). A reported LRRK2 ankyrin domain-containing N-terminal fragment expressed in E. coli will be useful for future experiments exploring whether the ankyrin domain can inhibit the kinase domain (Lu, et al., 2010). Generally, negative regulation of LRRK2 kinase activity by the N-terminal region is suggested by the finding that LRRK2, in which the LRRs and all further N-terminal residues were deleted, autophosphorylated in vitro to a greater extent than wild-type full-length LRRK2 (Greggio, et al., 2008b). To narrow down the inhibitory regions in the LRRK2 N-terminal domains, the H1-H2 linkers of the ankyrin domain may be a good place to start, by analogy with INK4 inhibition of Cdk6.

Ankyrin repeat domain: PD-associated variations

M712V, a PD-associated variation within the LRRK2 ankyrin domain, was found in one North American PD patient with unknown family history of the disease (Paisán-Ruíz, et al., 2008). A second variation, R793M, was identified in three German patients and an unaffected control (Berg, et al., 2005). Met-712 is in H1 of repeat 2, while Arg-793 is in H2 of repeat 4 and may project from the convex surface of the domain. Both Met-712 and Arg-793 are evolutionarily conserved residues in LRRK2 orthologues, suggesting that their variations may perturb their ankyrin repeat domain structure and/or interaction of the ankyrin repeat domain with other cellular proteins.

LEUCINE-RICH REPEAT DOMAIN

Leucine-rich repeat domain: structure

LRRs immediately N-terminal to the Ras-like Roc domain are a common characteristic of members of the Roco family of multi-domain proteins (Bosgraaf and Van Haastert, 2003). A LRR unit contains 20-30 amino acid residues with the consensus motif LxxLxLxx\textsuperscript{N/C}xL, where L can be Leu, Val, or Ile, N can be Asn, Thr or Ser and x stands for any amino acid. This motif forms the conserved β-strand and the consecutive loop regions. Adjacent LRR repeat units are assembled to form parallel β-strands, each linked by an arched
segment that can be composed of various secondary structures, such as \( \alpha \)-helix, \( 3_{10} \) helix or polyproline II conformation [reviewed in Bella et al. (Bella, et al., 2008) and (Enkhbayar, et al., 2004)].

Our prediction reveals that the LRRs within LRRK2 range from 20 to 29 amino acids in length. A predicted LRR at the N-terminal end of the domain contains 29 residues, and could potentially form an \( \alpha \)-helix on its convex side (residues 984 to 1012) (Figure 2). The LRRK2 LRR region lacks the distinct pattern of Cys residues normally present in the capping structure found at the N-terminal end of many LRR domains, and it is possible that the longer N-terminal LRR in LRRK2 fulfills the function of the cap, masking the hydrophobic residues in the neighboring LRRs. Vancraenenbroeck et. al recently reported a similar LRR structural-sequence alignment for LRRK2 and generation of a similar homology model for this domain (Vancraenenbroeck, et al., 2012).

**Leucine-rich repeat domain: interactions**

Tandem LRRs form a stack with a tighter curvature compared with ankyrin domains, and interact with client proteins via a narrow, flat face formed by the loops linking the \( \beta \)-strands with the helical regions of each repeat. Residues projecting from position 8 of individual LRRs dominate this interaction; for example, Asn-131 of Ran GAP, which stabilizes the Gln residue of the GTPase Ran (Seewald, et al., 2002). Potential client protein-interacting residues in the LRRK2 LRRs are Arg-1067 and Lys-1138, which occupy position 8 in their respective LRRs, and are conserved throughout vertebrates (Figure 3A). These residues are candidates for future experiments employing targeted mutagenesis to disrupt the client protein-interacting face of the LRRK2 LRR domain. Interestingly, variation of the conserved residue Arg-1067 to an uncharged Gln residue has been found in a Singaporean PD patient, suggesting a link to parkinsonism (Skipper, et al., 2005).

**Leucine-rich repeat domain: PD-associated variations**

In screening of 53 German families affected by PD, LRRK2 S1096C and S1228T variations were identified in patients from two separate families (Berg, et al., 2005). Another variation, I1122V, was detected in two siblings with PD (Zimprich, et al., 2004), and one patient heterozygous for the variation R1067Q was found in a group of Singaporean PD patients (Skipper, et al., 2005). All but the S1228T variation lie in the N-terminal half of the domain (Figure 2B). Among them, Arg-1067, Ser-1096 and Ser-1228 are predicted to lie on the surface of the concave face of the domain. Thus, they are likely solvent-exposed and can
potentially be involved in protein-protein interactions (Figure 2B). Ile-1122, however, lies in the variable length region on the convex face, so it is difficult to predict whether it is solvent-exposed or buried (Figure 2B).

Some of the variations could potentially affect some interaction of LRRs with other LRRK2 domains, or with protein interactors of LRRK2. A potential interactor with the LRR domain is Rab5b, which was identified in a yeast two-hybrid screen of a human brain cDNA library using the LRRK2 leucine-rich repeat domain (LRR, residues 967-1360) (Shin, et al., 2008). Interestingly, Ser-1124 has been identified as a phosphorylation site, likely by LRRK2 autophosphorylation (Gloeckner, et al., 2010). Thus, variation of the nearby Ile-1122 may interfere with recognition of this site by the LRRK2 kinase domain.

Autophosphorylation of the LRR domain suggests the LRR and kinase domains come into close proximity in cells, and this process may be further explored using a reported LRRK2 construct encoding residues from the LRRs to the C-terminus expressed in insect cells (Anand, et al., 2009). In addition to autophosphorylation of Ser-1124, Ser-1292 was recently identified by Sheng et al. as an autophosphorylation site (Sheng, et al., 2012). More importantly, its phosphorylation level correlates with the increase in LRRK2 kinase activity induced by several well documented PD-associated LRRK2 missense variations including R1441G/C, N1437H, S2019S and I2020T variations mapped to the COR and kinase domains. Intriguingly, even though replacement of Ser-1292 by alanine has no impact on LRRK2 kinase activity, it ameliorates the defects in neurite outgrowth caused by these five PD-associated LRRK2 variations. These data suggest that autophosphorylation of Ser-1292 plays a significant role in mediating functional interactions of the LRR domain with the COR and kinase domains. In our model shown in Figure 3, Ser-1292 is mapped to the putative arched segment linking the putative repeats 13 and 14.

ROC AND COR TANDEM DOMAINS

Roc and COR tandem domain: structure

C-terminal to the LRRs, LRRK2 contains a Ras of complex (Roc) GTPase region, followed by an intervening stretch of amino-acids (the COR, C-terminal of Roc) which together form the central core shared by a family of multi-domain proteins called the Roco family proteins (Bosgraaf and Van Haastert, 2003). Since the Roc and COR domains are arranged in tandem with Roc located at the N-terminal side of COR, the Roc and COR domains are referred to as Roc-COR tandem domains (Gotthardt, et al., 2008b). In LRRK2, the Roc and COR domains correspond to the segments of Pro-1331 to Phe-1511 and Lys-
1512 to Leu-1840, respectively. The crystal structure of recombinant LRRK2 Roc domain complexed with Mg$^{2+}$-GDP reveals a domain-swapped dimer with the segment of Pro-1331 to Arg-1412 of one monomer forming extensive interaction with the segment of Ala-1413 to Lys-1512 of another monomer (Deng, et al., 2008a). The active site resides at the interface of the two monomers (Figure 3A). Although the structure suggests that the domain-swapped dimer can bind guanine nucleotide and the conserved active site residues can potentially align appropriately for catalysis (lower panel of Figure 3A), this structure has been challenged as a potential artifact formed in the process of crystallization (Gotthardt, et al., 2008b; Wittinghofer and Vetter, 2011).

Nonetheless, the putative functional units formed by of the swapped domain dimer of LRRK2 Roc domain structurally resemble Ras and other small GTPases (Figure 3A and Figure S2). It contains five α-helices and six β-strands linked together by loops. Many of them contain the conserved residues essential for specific binding of Mg$^{2+}$-guanine nucleotides and catalysing GTP hydrolysis [reviewed in (Wittinghofer and Vetter, 2011)]. The first functionally important loop is the β1/α1 loop connecting the β1 strand and α2 helix contains the segment (\textsuperscript{1341}GNTGSGKT\textsuperscript{1348}) corresponding to the conserved phosphate-binding motif (GxxxxGKS/T) for binding the phosphate groups in guanine nucleotides. Thus, the PD-associated T1348N variation mapped to this motif is expected to perturb binding of GTP and GDP. Indeed, Xiong et al. reported that the T1348N variation confers neurotoxicity to recombinant LRRK2 (Xiong, et al., 2010). The second functionally important loop is the α1/β2 loop containing the conserved Switch 1 motif (Lys-1356 to Val-1369) which undergoes extensive conformational changes when GTPases transit from the GTP-bound to the GDP-bound states. Thr-1368 in this motif corresponds to the conserved Thr residue essential for binding to the γ-phosphate of GTP (Supp. Figure S2). The third important significant loop is the β3/α2 loop which forms part of the conserved Switch 2 motif (Asp-1394 to Gln-1441) which is expected to exhibit significant conformational differences between the GTP- and GDP-bound states. Asp-1304 and Gly-1397 correspond to the conserved aspartate and glycine residues in the consensus DxxG segment commonly found in GTPases. Asp-1304 is predicted to bind the Mg$^{2+}$ ion complexed with the guanine nucleotide, while Gly-1397 binds the γ-phosphate of GTP. In most GTPases, a conserved glutamine in the β3/α2 loop plays a critical role in catalysing GTP hydrolysis. This conserved glutamine is replaced by Arg-1398 in LRRK2 (Supp. Figure S2). Thus, the PD-associated R1398H variation is expected to perturb GTP hydrolysis. Both Switch 1 and Switch 2 contain structural features that mediate specific binding of GTPases to their effector proteins to
modulate their activities. The PD-associated R1398H variation may perturb binding of Switch 2 of LRRK2 Roc domain to its effector proteins. The fourth functionally important loop is the β5/α4-loop containing the conserved N/TKxD motif (1452-THLD1455 where His-1453 substitutes the conserved lysine residue), and Asp-1455 in this motif corresponds to the conserved Asp that governs specific binding to the guanine moiety in GTP/GDP (Supp. Figure S2).

Gotthardt, et al. determined the crystal structure of the tandem Roc and COR domains from the C. tepidum Roco protein (Gotthardt, et al., 2008b) (Figure 3). Since this is the only known structure of tandem Roc-COR domains, it has been used as a model to predict the structure and function of the tandem Roc and COR domains of LRRK2 and assess how PD-associated variations impact the normal function of these two domains of LRRK2. In the next two sections, we use the structures of Ras, LRRK2 Roc domain and C. tepidum Roco protein as the models to discuss the structural basis of regulatory properties, dimerization of LRRK2, interactions of the LRRK2 tandem Roc and COR domains with its upstream regulatory proteins and downstream effector proteins.

**Roc and COR domains: regulation**

In a member of the Roco family from D. discoideum, known as Gbpc, there exists a Ras GEF-like domain C-terminal to the Roc/COR-kinase region, which can specifically promote release of GDP from the purified Gbpc Roc and COR fragment (van Egmond, et al., 2008). Since LRRK2 does not have a Ras GEF-like domain, it likely requires the assistance of an exogenous guanine nucleotide exchange factor to promote the release of GDP from its Roc domain for the guanine nucleotide exchange. LRRK2 forms protein complexes with the Ras-like GTPase cdc42 and one of its guanine nucleotide exchange factors (GEF) called ARHGEF7 in cells (Habig, et al., 2013; Haebig, et al., 2010). Using an assay that measures the binding of LRRK2 to GTP-agarose, Haebig et al. provided preliminary data supporting ARHGEF7 as a GEF of LRRK2. Future investigation should focus on defining the biochemical and structural basis of regulation of the tandem Roc and COR domains of LRRK2 by ARHGEF7. Most Ras-like GTPases are regulated by multiple GEFs. In addition to ARHGEF7, LRRK2 is likely to be regulated by other GEFs.

Similar to other Ras-like small GTPases, LRRK2 exhibits weak GTPase activity (Liu, et al., 2010; Liu, et al., 2012; Stafa, et al., 2012). Efficient hydrolysis of GTP requires the assistance of a GTPase-activator protein (GAP). The ArfGAP1 which facilitates the GTP hydrolysis by the ADP-ribosylation factor 1 (ARF) was recently found by two groups of
researchers to directly bind LRRK2 and enhance GTP hydrolysis (Stafa, et al., 2012; Xiong, et al., 2012b). However, both groups provided conflicting data on the motifs in LRRK2 that mediate its interaction with ArfGAP1. Stafa et al. mapped the Arf-GAP1-binding motif to a segment encompassing the armadillo repeat and ankyrin repeat domains of LRRK2 while Xiong et al. mapped it to the WD-40 and the kinase domain near the C-terminus of LRRK2 (Stafa, et al., 2012; Xiong, et al., 2012b). Future investigation should focus on defining the ArfGAP1-binding motif(s) in LRRK2 and understanding whether and how PD-associated variations in the armadillo-repeat, ankyrin-repeat, WD-40 repeat and/or kinase domains affect LRRK2 interaction with ArfGAP1.

The Roc and COR domains are implicated in auto-regulation of LRRK2. Several pieces of evidence suggest that activity of the kinase domain is under the control of the GTPase domain. First, autophosphorylation of LRRK2 can be influenced by addition of non-hydrolysable GTP analogues or GDP (Guo, et al., 2007; Smith, et al., 2006a; West, et al., 2007) or introduction of the T1348N variation designed to abolish guanine nucleotide binding (Ito, et al., 2007). Second, variation of Thr-1343 to glycine and Arg-1398 to glutamine in the Roc domain also influenced autophosphorylation of LRRK2 (Ito, et al., 2007; Xiong, et al., 2010). In Ras, Gly-13 and Gln-61 are located at the homologous positions of Thr-1343 and Arg-1398 of LRRK2 Roc domain (Deng, et al., 2008a) (Supp. Figure S2). Furthermore, Gly-13 and Gln-61 enhance GTP hydrolysis of Ras by interacting with the γ-phosphate of GTP and by stabilising the nucleophilic water molecule in the hydrolysis reaction, respectively (Maegley, et al., 1996; Milburn, et al., 1990). Thus, the artificial T1343G and R1398Q variations are expected to augment the LRRK2 GTPase activity (Deng, et al., 2008b; Ito, et al., 2007; Maegley, et al., 1996; Milburn, et al., 1990; Xiong, et al., 2010). Relevant to these findings, Xiong et al. reported that ArfGAP1 inhibits the cytotoxic action of LRRK2 when expressed in cells and in D. melanogaster (Xiong, et al., 2012b). Presumably, ArfGAP1 exerts its neurotoxicity-attenuating effect by enhancing the GTPase activity of the LRRK2 Roc domain and this in turn suppresses LRRK2 autophosphorylation (Xiong, et al., 2012b). Future investigation should focus on defining the structural and biochemical basis of cross-talk between the Roc and COR domains and the kinase domains of LRRK2 and how this may regulate the neurotoxicity of LRRK2.

**Roc and COR domains: interactions**

*Interactions mediated by the Roc domain*
A major determinant region for effector binding incorporates part of the guanine-nucleotide-dependent loop, known as Switch I, and the \( \beta \)-strand immediately C-terminal to switch I. Residues in this region vary between the subfamilies of Ras-like GTPases, contributing to their preference for subfamily-specific effectors and variation of these residues can alter effector selectivity (Bauer, et al., 1999; White, et al., 1995). In LRRK2, the central portion of this region is the \(^{1367}\text{ATVGIDV}\) segment, which is most similar to the corresponding region in the Rab subfamily. In most Rab members, the sequences of this effector-binding region conform to the consensus [A/S]TIG[I/V]DF. The LRRK2 sequence matches the Rab subfamily consensus except for a phenylalanine residue in Rab GTPases, which forms part of a triad of conserved hydrophobic residues that pack against parallel \( \alpha \)-helices in the Rab effectors Rab6IP1 and GCC185 (Fernandes, et al., 2009). The Rab-like effector region of the LRRK2 Roc domain may be sufficient to recruit Rab effectors, providing a further link to Rab-mediated signaling, in addition to the reported LRRK2 interactions with Rab5b and co-localisation with Rab7b (Higashi, et al., 2009; Shin, et al., 2008).

Besides the conventional Switch 1 and Switch 2 motifs used by GTPases to interact with effector proteins, other regions of the LRRK2 Roc domain are also involved in mediating binding of LRRK2 with other cellular proteins. A notable example is the long loop that links the \( \beta 2 \) and \( \beta 3 \) strands (Supp. Figure S2). This loop (referred to as the \( \beta 2/\beta 3 \)-loop) corresponding to residues 1379-1387 is not visible in the crystal structure of the Roc domain determined by Deng et al. (Deng, et al., 2008b) (Figure 3A), suggesting that it is flexible. Orenstein et al. recently reported that LRRK2 in neurons is degraded in lysosomes by chaperone-mediated autophagy and the \( \beta 2/\beta 3 \)-loop contains a putative pentapeptide motif critical for the translocation of LRRK2 to lysosomes (Orenstein, et al., 2013a). Similar to other protein substrates degraded by this process, LRRK2 binds to heat shock cognate protein hsc70 which then target it to the lysosome-associated membrane protein type 2A (LAMP-2A). Upon binding to LAMP2A, LRRK2 is translocated into lysosomes for degradation. LRRK2 contains multiple putative pentapeptide motifs which can direct its interaction with hsc70 (Orenstein, et al., 2013a). Among these pentapeptide motifs, the one \( ^{1379}\text{QIRDK} \) located in the \( \beta 2/\beta 3 \)-loop was found to be the most important hsc70-binding site (Supp. Figure S2). Intriguingly, the LRRK2 mutant carrying the most prevalent PD-associated G2019S variation was degraded by chaperone-mediated autophagy with a much lower efficiency. Since G2019S variation significantly increases the LRRK2 kinase activity and autophosphorylation at multiple sites in the Roc and COR domains (Greggio, et al., 2009), we
predict that enhanced autophosphorylation of the Roc and COR domain reduces the accessibility of the $^{1379}$QIRDK motif to hsc70 and this in turn may attenuate the efficiency of the degradation of LRRK2 by chaperone-mediated autophagy.

**Intramolecular interactions between the Roc and COR domains**

The tandem Roc and COR domains of the *C. tepidum* Roco protein form homodimers (Gotthardt, et al., 2008c) with extensive hydrophobic interactions among residues at the intramolecular Roc/COR interface (Figure 3B). Superimposition of the LRRK2 Roc domain with the Roc domain in the monomer A of this structure generates a model of the LRRK2 Roc/COR interface (Figure 3B). The model suggests that the interface contains Ile-1371, Arg-1441 and Tyr-1669 found mutated in some PD patients. Thus, variations of these residues can potentially affect intramolecular cross-talks between the Roc and COR domains in LRRK2.

**Intermolecular interactions between the COR domains of two neighboring monomers of tandem Roc and COR domains**

The crystal structure of the Roc-COR domains of the *C. tepidum* Roco protein reveals a dimer formed by two monomers (monomers A and B) of the tandem Roc and COR domains interacting at the COR domains of the two monomers (Gotthardt, et al., 2008c). Supp. Figure S3 reveals that the dimer is formed by interactions between determinants of the COR domains of the neighboring monomers. In particular, electrostatic interactions between Glu-906 of one monomer with Lys-922 of the neighboring monomer likely play a major role in stabilising the dimer (Supp. Figure S3, left panel). Since the COR domains of LRRK2 and *C. tepidum* Roco protein exhibit significant sequence variations, we are unable to predict if the COR domain of LRRK2 can form extensive electrostatic interactions similar to those shown in Figure S3 to stabilise the LRRK2 dimer.

**Intermolecular interactions between the Roc domains of two neighboring monomers of tandem Roc and COR domains**

A major portion of the Roc domain is missing in the monomer B of the three-dimensional structure of the dimer of Roc-COR domains of *C. tepidum* Roco protein. To create a model of the intact dimer, we superimposed the Roc domain of monomer A to the structure of monomer B. In this model, Arg-543 of one monomer is located in the active site of neighboring monomer (Figure S3, right panel). Gotthardt *et al.* suggested that this Arg-543 may mimic the arginine finger of GTPase activator protein (GAP) to facilitate GTPase hydrolysis in the Roc active site (Gotthardt, et al., 2008b). Indeed, variation of Arg-543
completely abolishes the GTPase activity of the *C. tepidum* Roc-COR domain dimer, suggesting its significance in catalysing GTP hydrolysis. Docking of the Roc domain structure to the monomer A of the dimer of *C. tepidum* tandem Roc and COR domains reveals Lys-1423 of LRRK2 as the residue located in the homologous position in LRRK2 Roc domain (Supp. Figure S2). In light of this, it will be worthwhile to examine the effect of variation of Lys-1423 on the GTPase activity of LRRK2.

**Roc and COR domains: PD-associated variations**

Three variations, R1441G, R1441C and R1441H, were found in families from the Basque region of Northern Spain, North America and Taiwan, respectively (Mata, et al., 2005; Paisán-Ruiz, et al., 2004). The Y1699C and R1628P variations in the COR domain of LRRK2 were detected in a German-Canadian family and an international cohort, respectively (Mata, et al., 2005; Zimprich, et al., 2004). The residue predicted as equivalent to LRRK2-R1628 in the *C. tepidum* Roco structure, Gln-730, is located in a solvent-exposed interface between α-helices 12 and 13 in subdomain 1 of the COR domain (Figure 3) (Gotthardt, et al., 2008a). *C. tepidum* Roc-COR domain residues predicted as equivalent to R1441 and Y1699 are in close proximity, near the interface between the Roc and COR domains, suggesting variation of these sites may affect Roc-COR interaction (Figure 3). This prediction is supported by an increased interaction of both LRRK2 COR-only or COR-kinase constructs with GST-tagged Roc domain resulting from the Y1699C variation (Daniëls, et al., 2011b).

**KINASE DOMAIN**

**Kinase domain: structure**

*Lessons from the structures of wild type and mutant Roco4 kinase domain*

LRRK2 contains a protein kinase domain, located C-terminal to the COR domain, in a similar arrangement to other members of the Roco family of multi-domain proteins. The *D. discoideum* Roco4 is a member of the Roco family proteins with the LRR, Roc-COR, kinase and WD domains arranged in a similar order as LRRK2. The kinase domains of *D. discoideum* Roco4 and human LRRK2 share 29.9 % sequence identity and 34.4 % sequence similarity (as aligned in Supp. Figure S4, with percentage similarity calculated by normalising each residue pair to maximum and minimum scores from the BLOSUM62 substitution matrix, and excluding gaps). Results of the targeted gene disruption study revealed that Roco4 governs cellulose biosynthesis on *Dictyostelium* cells (Gilsbach, et al., 2012). More importantly, the phenotypic defects of the Roco4-deficient *Dictyostelium* cells
can be rescued by expression of a recombinant Roco4/LRRK2 chimera in which the Roco4 kinase domain is replaced by that of LRRK2. These data suggest that the kinase domains of Roco4 and LRRK2 can phosphorylate similar protein substrates to rescue the phenotypic defects in the *Roco4*-null *Dictyostelium* cells. Thus, Gilsbach *et al.* suggested that the crystal structures of wild type and mutant Roco4 kinase domains are suitable models of LRRK2 kinase domain (Gilsbach, *et al.*, 2012).

The catalytic domain of all protein kinases contains two lobes: the minor lobe at the N-terminal portion and the C-terminal major lobe. The cleft between the two lobes contains the catalytically critical residues required for binding of substrates (ATP and protein substrates) and for catalysis. Inspection of the structures of many protein kinases reveals that when a protein kinase adopts the active conformation, the kinase domain is connected by two non-contiguous hydrophobic spines termed regulatory and catalytic spines anchored to the αF-helix (Kornev and Taylor, 2010; Kornev, *et al.*, 2008). Once assembled, the residues essential for substrate binding and catalysis are appropriately oriented for effective phosphorylation of the protein substrates. Figures 4 and S5 show the superimposed structures of wild type and mutant Roco4 kinase domains. In all these structures, both the catalytic and regulatory spines are properly assembled, suggesting that the catalytically critical residues are appropriately aligned for catalysis. The regulatory spine of Roco4 kinase domain is formed by His-1152 in the conserved HRDLRSPN motif of the catalytic loop, Phe-1178 in the DFG motif of the activation loop, Met-1082 of the αC-helix and Leu-1093 of β4 strand. Once assembled in the regulatory spine, the catalytic base Asp-1154 of the HRDLRSPN motif is appropriately aligned for catalysis. The side chain of Asp-1177 in the DFG motif chelates a magnesium ion that binds to β- and γ-phosphates of ATP. The assembly of Phe-1178 in the regulatory spine likely facilitates appropriate positioning of Asp-1177 for ATP binding (Supp. Figure S5B). Since Met-1082 is contiguous to Glu-1078 in the αC-helix, its assembly in the regulatory spine allows proper positioning of the conserved Glu-1078 to form a salt bridge with the conserved Lys-1055 in the AIK motif of the β3-strand (Supp. Figure S5B). Formation of this salt-bridge allows the lysine residue to bind β-phosphate of ATP.

In addition to the electrostatic interactions of the β- and γ-phosphates of ATP with the β3 strand Lys-1055 and the magnesium ion bound to Asp-1177 in the DFG-motif of the activation loop, binding of ATP to the active site is also controlled by the catalytic spine which forms hydrophobic interaction with the adenine ring of ATP. As shown in Supp. Figure S5B, Val-1040 and Ala-1053 in the N-lobe of the kinase domain form the cap of the
adenine ring while the side chain of Phe-1161 in the β7-strand in the major lobe of the kinase domain forms the base. Two contiguous residues Ile-1160 and Leu-1162 form hydrophobic interactions with Leu-1113 in the αD-helix. Interactions of Leu-1113 with Ile-1222 and Ile-1226 anchor the catalytic spine to αF-helix. Results of studies of the conformational dynamics of PKA in catalysis suggest that opening and closing of the catalytic spine at the points of contact with the adenine ring of ATP and ADP governs binding of ATP, transfer of phosphate and release of ADP (Kornev and Taylor, 2010; Masterson, et al., 2012; Masterson, et al., 2010). LRRK2 is activated by the PD-associated G2019S variation, future investigation can focus on understanding how the mutated Ser-2019 residue interplays with residues forming the catalytic and regulatory spines in controlling LRRK2 kinase activity (see below for a more detailed discussion).

**The role of activation loop in regulation of the kinase activity of LRRK2 and Roco4**

Recombinant Roco4 kinase domain readily undergoes autophosphorylation *in vitro* to a stoichiometry of 2 molecules of phosphate per molecule of enzyme. Results of biochemical analysis suggest Ser-1187 and Ser-1189 of the activation loop as the sites of phosphorylation. Furthermore, replacement of these two serine residues by alanine abolishes the kinase activity and the ability to rescue the phenotypic defects Roco4-deficient *Dictyostelium* cells, suggesting that autophosphorylation at these two sites of the activation loop is crucial for Roco4 kinase activity. Of relevance, Thr-2031, Ser-2032 and Thr-2035 in the activation loop of LRRK2 were identified as autophosphorylation sites (Greggio, et al., 2008a; Kamikawaji, et al., 2013; Li, et al., 2010a) (Supp. Figure S2). Unlike the essential role played by autophosphorylation Ser-1187 and Ser-1189 in the kinase activity and biological function of Roco4, autophosphorylation of Ser-2032 and Thr-2035 either has a modest effect on the ability of LRRK2 to undergo autophosphorylation at other sites *in vitro* and cytotoxicity or has no effect on the ability of LRRK2 to phosphorylate exogenous substrates (Greggio, et al., 2008a; Li, et al., 2010a). Thus, despite the high degree of sequence similarity between the kinase domains of Roco4 and LRRK2, the kinase activity of the two enzymes exhibit different degrees of reliance on the phosphorylation states of the activation loop autophosphorylation sites.

**Variations in the protein kinase domain: the G2019S and I2020T variation**
The G2019S variation was originally identified in two siblings from a diverse group of North American and European families with history of PD (Kachergus, et al., 2005) and has been detected in 35.7% of the tested Arab sporadic PD cases in North Africa (Ishihara, et al., 2007), and in European populations the variation accounts for approximately 5 to 6% of familial parkinsonism (Di Fonzo, et al., 2005; Nichols, et al., 2007; Nichols, et al., 2005). Laboratory models also suggest a potential role in disease, with reports of overexpression of G2019S mutant LRRK2 in animal models variously resulting in reduction in striatal dopamine and dopaminergic neuronal loss, and causing decreased neurite outgrowth in cultured SH-SY5Y neuroblastoma cells (Chen, et al., 2012; Lee, et al., 2010c; Li, et al., 2010b; Liu, et al., 2008; Plowey, et al., 2008). Using a quantitative phosphoproteomic approach to identify proteins exhibiting elevated phosphorylation level in substantia nigra of [G2019S]LRRK2 transgenic mice, Chen et al. found that MAPK kinase 4 (M KK4) phosphorylation at Ser-257 was upregulated, reminiscent of the previous findings by Hsu et al. that LRRK2 can interact with multiple MKKs to activate the MKK/JNK signaling pathways (Hsu, et al., 2010a; Hsu, et al., 2010b). They demonstrated up-regulation of the MKK4/JNK/Jun cell death signaling pathway, which in turn directs activation of caspases and expression of Bim and FasL in substantia nigra of the [G2019S]LRRK2 transgenic mice. Their data suggest that [G2019S]LRRK2 causes striatal dopaminergic neuronal loss by activating both the intrinsic and extrinsic apoptotic signaling pathways. Biochemical analysis reveals that the G2019S variation increased both autophosphorylation and phosphorylation of protein and peptide substrates above wild-type levels, in full-length recombinant LRRK2 purified from HEK293T cells (Covy and Giasson, 2009; Matta, et al., 2012; West, et al., 2005), in the (Δ1-969)LRRK2 deletion mutant expressed in baculovirus-infected insect cells (Anand, et al., 2009) and in COR-kinase fragment expressed in E. coli (Luzón-Toro, et al., 2007). In addition to inducing neuronal loss, G2019S variation also affects synaptic vesicle turnover. Matt, et al. discovered Enodophilin A (EndoA) crucial for synaptic vesicle endocytosis as a physiological substrate of LRRK2, and identified Ser-75 in Endo A as the phosphorylation site (Matta, et al., 2012). The G2019S variation results in a faster and a higher stoichiometry of phosphorylation of Endo A in vitro (Matta, et al., 2012).

The G2019S variation lies at the N-terminal end of the activation loop (Figure S4), the conformation of which is important in many kinases for regulating their protein kinase activity. Variations in the corresponding glycine residue have been reported to alter the function of human PTEN-induced kinase (PINK1), Cyclin-dependent kinase (Cdk1) in Drosophila melanogaster and Calcium and calmodulin-dependent serine/threonine-protein
kinase (CCaMK) in the garden pea *Pisum sativum* (Clegg, et al., 1993; Lévy, et al., 2004; Sim, et al., 2006). The [G1179S]Roco4 kinase domain with the homologous Gly-1179 in the activation loop mutated to serine was generated by Gilsbach et al. to mimic the G2019S variation in LRRK2. The crystal structure of this mutant Roco4 kinase domain reveals that Ser-1179 is engaged in hydrogen bonding with Arg-1077 in the αC-helix (Figure 4B), implying that Ser-2019 of [G2019S]LRRK2 is also engaged in hydrogen bonding with Gln-1919 at the homologous position in the αC-helix (Figure 4B). Indeed, the kinase activity of [G2019S/Q1919A]LRRK2 is much lower than that of [G2019S]LRRK2, supporting the notion that the active conformation of LRRK2 kinase domain is stabilised by hydrogen bonds formed between the side chains of Ser-2019 and Gln-1919.

The activation state of a number of protein kinases is governed by movement of the activation loop to adopt two configurations – the DFG-in and DFG-out configurations. The movement is caused by exchange in positions of the phenylalaine (corresponding to Tyr-2018 in LRRK2) and the glycine (corresponding to Gly-2019 of LRRK2). The adoption of the DFG-in configuration is essential for catalysis. A group of small molecule inhibitors, referred to as type II inhibitors (Liu, et al., 2013), suppress the activity of protein kinases by selectively binding and stabilising them when they adopt the conformation with the DFG-out configuration. Simulation of the dynamics of the loop movement reveals that the G2019S variation can potentially stabilise LRRK2 kinase domain in the DFG-in configuration (Liu, et al., 2013).

In contrast to the activating effect of the G2019S variation on LRRK2 kinase domain, the PD-associated I2020T variation induces a modest increase or has no effect on the ability of LRRK2 kinase domain to phosphorylate exogenous protein/peptide substrates (Funayama, et al., 2005; Nichols, et al., 2010; Ohta, et al., 2007). Furthermore, I2020T variation significantly reduces phosphorylation of recombinant LRRK2 at Ser-910 and Ser-935 in transfected HEK293 cells (Nichols, et al., 2010). Since phosphorylated Ser-910 and Ser-935 are docking sites of the 14-3-3 protein, interaction of recombinant LRRK2 with 14-3-3 protein in cells was significantly diminished. Elucidation of the structural basis of how G2019S and I2020T variations impact on LRRK2 kinase activity awaits determination of the crystal structure of LRRK2 kinase domain.

**Protein kinase domain: interactions**

*The kinase domain as a scaffold – lessons from the MAP3K*
Instances exist of MLK-family kinases with protein-protein interactions mediated by their kinase domain. For example, the catalytic region of human MEKK1 (residues 1199–1496) is sufficient for recruitment to a complex with JNKK1 that is involved in transduction of signals to the downstream kinase JNK (Xia, et al., 1998). Also, testing truncations of the MLK-family kinase TAK1 for interaction with JIP1 showed the kinase domain of TAK1 is sufficient for binding (Blanco, et al., 2007). Similar interactions for the LRRK2 kinase domain are suggested by its broad sequence similarity to these MLK-family kinases; however, the specific regions involved have not been mapped.

Interactions with potential protein substrates

A number of cellular proteins including α-synuclein, ArfGAP1, ARHGEF7 Akt1, 4EBP1, moesin, Futsch and endophilin A (EndoA) have been identified as candidate substrates of LRRK2 kinase domain (Haebig, et al., 2010; Jaleel, et al., 2007; Lee, et al., 2010d; Matta, et al., 2012; Ohta, et al., 2011a; Parisiadou, et al., 2009; Qing, et al., 2009; Stafa, et al., 2012; Xiong, et al., 2012b) (Supp. Table S7). A protein kinase recognises its protein substrates by two types of interactions: (i) interactions between the active site of the kinase with the consensus phosphorylation sequence in the protein substrates and (ii) distal interactions between the kinase and the protein substrate mediated by binding of docking motifs spatially separated from the phosphorylation site in the substrate and interaction motif or functional domain located distally from the active site of the kinase (Cheng, et al., 2011; Turk, 2008; Ubersax and Ferrell, 2007). With these two interactions, protein kinases are able to recognise their protein substrates with exquisite specificity. Using the positional scanning peptide library approach developed by Hutti et al. (Hutti, et al., 2004), two groups of researchers have defined the optimal phosphorylation sequence recognized by the active site of LRRK2 kinase domain (Nichols, et al., 2009; Pungaliya, et al., 2010). A peptide, referred to as Nictide derived from the sequence has been used as an in vitro peptide substrate to monitor the kinase activity of LRRK2 (Nichols, et al., 2009). Supplemental Table S7 lists the Nictide sequence and the phosphorylation sites of a number of protein substrates of LRRK2. Inspection of these sequences revealed two features. First, threonine appears at a much higher frequency than serine as the site of phosphorylation in the protein substrates. The preference for threonine as the phosphorylation site was further confirmed by Nichol et al. who demonstrated that LRRK2 failed to phosphorylate the Nictide derivative in which the target threonine is replaced by serine (Nichols, et al., 2009). Second, the phosphorylation sequences
in the protein substrates deviate significantly from the consensus phosphorylation sequence for LRRK2 (the underlined portion in the Nictide sequence, Supp. Table S7) defined by the peptide library screening. This finding suggests that the active site of LRRK2 kinase domain plays only a minor role in recognition of these proteins as LRRK2 substrates. Presumably, these protein substrates employ docking motifs spatially separated from the phosphorylation site to bind to interaction motifs located distally from the active site of LRRK2. The binding allows LRRK2 to specifically recognize these proteins as its preferred substrates. These interaction motifs for substrate recognition may reside in the kinase domain and or other functional domains of LRRK2. For example, ArfGAP1 binds to motifs outside the kinase domain for its efficient phosphorylation by LRRK2 (Stafa, et al., 2012; Xiong, et al., 2012b). Thus, further investigation to identify protein substrates of LRRK2 should also focus on defining how motifs located distally from the active site of LRRK2 kinase domain participate in directing LRRK2 phosphorylation of these substrates. Relevant to this, Lee et al. reported that LRRK2 phosphorylates distinct clusters of the protein substrates in the pre-synaptic and postsynaptic compartment of Drosophila neuromuscular junction (Lee, et al., 2010d).

LRRK2 preferentially phosphorylates 4E-BP, an ortholog of the protein synthesis inhibitor 4E-BP1 in the postsynaptic compartment but the microtubule-binding protein Futsch in the presynaptic compartment. Presumably, the distal interactions between the docking sites in 4E-BP and Futsch with interaction motifs in one or more of the functional domains governing protein-protein interactions are critical in directing their preferential phosphorylation in the different compartments of the neuromuscular junction.

**WD-40 REPEAT DOMAIN**

A number of PD-associated variations and phosphorylation sites map to the WD-40 repeat domain (Supp. Figure S1 and Supp. Tables S1 and S2), suggesting that this domain participates in the neurotoxic action and regulation of LRRK2. Indeed, deletion of the WD-40 repeat domain of zLRRK2 (homologous of human LRRK2 expressed in zebra fishes) induces neuronal loss and locomotive defects in zebra fish (Sheng, et al., 2010). Furthermore, Jorgensen, et al. reported that WD-40 is critical to the neurotoxic action of some LRRK2 mutants (Jorgensen, et al., 2009). In recent studies, ArfGAP1 was reported to bind to LRRK2 and activate its GTPase activity (Stafa, et al., 2012; Xiong, et al., 2012b); in one report, the ArfGAP1-binding sites in LRRK2 map to the kinase domain and WD-40 repeat domain (Xiong, et al., 2012b).
A structural model of the LRRK2 WD-40 repeat domain

Structures of WD-40 repeat domains have been extensively characterized, with over 50 published structures to date (reviewed in Li & Roberts, (Li and Roberts, 2001). Each WD-40 motif forms four β-strands (commonly named A to D from innermost to outermost), which lie flat side-by-side to make a single β-sheet, running parallel to the axis of the domain. A WD-40 repeat domain comprises at least six such β-sheets formed from WD-40 motifs, which radiate out from the centre of the domain, forming an overall bladed barrel shape known as a β-propeller. The WD-40 domain of LRRK2 appears to comprise seven WD-40 repeats (Figure 5), beginning immediately C-terminal to the predicted I helix of the protein kinase domain and terminating close to the LRRK2 C-terminus (Supp. Figure S1).

WD-40 domain: interactions

A recent report by Xiong et al. demonstrated that the WD-40 repeat domain mediates LRRK2 interaction with its upstream regulator and protein substrate ArfGAP1 (Xiong, et al., 2012b). The LRRK2-binding motif was mapped to a region encompassing the N-terminal sixty five residues of ArfGAP1 (Stafa, et al., 2012; Xiong, et al., 2012b). Future investigation to define the structural basis of ArfGAP1 interaction with LRRK2 will provide insights into how WD-40 repeat domains selectively bind to their protein ligands.

Inspection of the structures of the complexes formed by the WD-40 repeat domains of other cellular proteins with their cognate protein partners can shed light on the role of the WD-40 repeat domain of LRRK2 in protein-protein interactions. F-box and WD repeat domain-containing 7 (Fbw7), a subunit of the SCF ubiquitin protein ligase complex, is responsible for substrate recognition for the enzyme complex. It contains seven WD-40 repeats. The region of greatest similarity between Fbw7 and LRRK2 WD-40 sequences spans WD-40 repeat 2 and the A-strand of repeat 3, a region which includes part of the Fbw7 phospho-peptide interacting region (mainly the D-A and B-C linkers of WD-40 repeats 3 and 5). This region is crucial for Fbw7 function. Regulatory phosphorylation sites known as Cdc4-phosphodegron (CPD) sites are recognized by Cdc4 and its orthologs, such as Fbw7, which target their client proteins for ubiquitination and degradation by the Cdc34-SCF ubiquitin protein ligase complex (Skowyra, et al., 1997). It is possible that the LRRK2 WD-40 region binds phospho-peptides in a manner similar to Fbw7, using a surface formed predominantly by repeats 2 and 3. A cluster of basic residues identified by Jorgensen et al, 2009 is preserved in the LRRK2 WD-40 model presented here (Figure 6) (Jorgensen, et al.,
2009), in which the previously identified basic residues occur close to the N-terminal end of the A strands of blades 6 and 7, and near the C-terminal end of the B strands of blades 5 and 7.

**WD-40 domain: PD-associated variations**

The T2356I variation was found in two patients of British and Dutch backgrounds (Khan, et al., 2005; Macedo, et al., 2009) while the G2385R variation was originally identified in a Taiwanese family affected by PD (Mata, et al., 2005), detected in 9% of a Taiwanese cohort of sporadic PD cases, and 7.3% in a Chinese PD cohort in Singapore [(Di Fonzo, et al., 2006; Fung, et al., 2006; Tan and Skipper, 2007). Overexpression of G2385R mutant LRRK2 results in neuronal loss in *Drosophila melanogaster* (Ng, et al., 2009), and increases susceptibility of mouse primary neurons and HEK293T cells to toxicity induced by hydrogen peroxide (Tan and Skipper, 2007; West, et al., 2007). The predicted locations of both G2385R and T2356I variations on an edge of β-sheet 5 formed from its A-B and C-D linkers suggest that both could disrupt the same region of the WD-40 domain (Figure 5). A previous analysis appears to locate the G2385R variation on the C-terminal end of strand D of β-sheet 5, although details of the analysis are not given (Mata, et al., 2006). Here, the G2385R variation is predicted to lie on an extended loop projecting from stand C and stand D of β-sheet 5, and this loop may function to increase surface area for interaction with binding partners, as has been observed for extended loops in some structurally-characterised WD-40 domains (Ren, et al., 2010; Song, et al., 2008). One potential binding partner of the LRRK2 WD-40 repeat domain is the LRRK2 kinase domain. Rudenko et al. reported that the G2385R variation results in a reduction in the kinase activity of LRRK2. When the G2385R and G2019S variations were introduced together into LRRK2, the G2385R variation abolishes the activating effect of the G2019S variation on the kinase domain (Rudenko, et al., 2012). Together with the predicted location of Gly-2385 in our model (Figure 5B), these data suggest that the G2365R variation may perturb interactions between the WD-40 domain with other cellular proteins or with other functional domains of LRRK2. The predicted location of Thr-2356 in the middle of strand A suggests its role in maintaining the structural integrity of the LRRK2 WD-40 domain. Thus, the T2356I variation can potentially perturb the structural integrity of the WD-40 domain. Future investigation of the LRRK2 WD-40 domain will likely focus on the region harboring these two variations, predicted here as β-sheet 5 of our model (Fig. 5), and address whether this region mediates some crucial function of the domain. Possible WD-40 functions are interaction with LRRK2 binding partners, binding to
the kinase domain or dimerisation, as supported by the inability of LRRK2 lacking the WD-40 domain to form dimers in HEK293T cell lysates separated by native gel electrophoresis and size-exclusion chromatography (Jorgensen, et al., 2009).

INTER-DOMAIN INTERACTIONS – THE ROLE IN REGULATION OF CATALYTIC ACTIVITIES AND NEUROTOXIC ACTION OF LRRK2

Dimerisation of LRRK2

As discussed in the section of Roc and COR domains, the *C. tepidum* tandem Roc and COR domains forms homodimers. If the tandem Roc and COR domains of LRRK2 also form homodimers, interactions between the two neighboring monomers likely direct dimerisation of LRRK2 (Gotthardt, et al., 2008b; Gotthardt, et al., 2008c). Using size exclusion column chromatography and native gel analysis, Greggio *et al.* and Sen *et al.* demonstrated that recombinant [G2019S]LRRK2 expressed in HEK293FT cells exists as monomer, dimer and higher order oligomers (Greggio, et al., 2008a; Sen, et al., 2009). Among them, only the dimeric form of LRRK2 exhibited significant kinase activity in phosphorylating an exogenous protein substrate and autophosphorylation. The existence of LRRK2 dimer in cells was also confirmed by several other groups of researchers (Berger, et al., 2010; Biosa, et al., 2013; Civiero, et al., 2012; James, et al., 2012; Jorgensen, et al., 2009). Ito *et al.* recently re-examined the dimerisation status of LRRK2 in cells using blue native gel electrophoresis and glycerol gradient centrifugation (Ito and Iwatsubo, 2012). From their data, they suggested that the LRRK2 monomer migrated as a dimer in blue native gel due to its anomalous gel mobility. Furthermore, in contrast to the findings by other researchers, they found that LRRK2 monomer also exhibited protein kinase activity. One approach to address the discrepancy of these results is determination of the molecular mass of purified recombinant full-length LRRK2 in solution by analytical ultracentrifugation. In addition, determination of the crystal structure of full-length LRRK2 may reveal if LRRK2 forms homodimers and how dimerisation affects LRRK2 kinase activity.

Interactions between the Roc/COR domains with the kinase domain

Since small GTPases exert their biological functions by binding to effector protein molecules in a GTP-dependent manner, we predict that the Roc/COR domain binds to the
kinase domain in a GTP-dependent fashion and in turn modulates the activity of the kinase domain. Since LRRK2 undergoes autophosphorylation at multiple sites in the Roc and COR domains, we predict that the GTPase activity and/or the effector-binding activity of Roc and COR domains are modulated by LRRK2 kinase domain. In light of this, a number of studies that aim to decipher the molecular basis of interplay between these catalytic functional domains have been conducted (Biosa, et al., 2013; Greggio, et al., 2006; Smith, et al., 2006c; Taymans, et al., 2011; West, et al., 2007). Notably, Taymans, et al. and Biosa et al. demonstrated that variation of Thr-1348 that abolishes of the guanine nucleotide-binding ability of the Roc domain induces a significant decrease in the kinase activity of LRRK2 and [G2019S]LRRK2. Furthermore, incubation of both the wild type and [G2019S]LRRK2 with GTP or the non-hydrolysable GTP analog significantly increased their kinase activity (Biosa, et al., 2013; Taymans, et al., 2011). All these findings indicate that the Roc/COR domains, upon binding to GTP, modulate the activity of the kinase domain.

**Phosphorylation sites and variations mapped to the linking segments connecting the various functional domains**

In addition to phosphorylation sites mapped to the various functional domains, LRRK2 also undergoes phosphorylation at sites in the segment connecting the ankyrin and LRR domains and that connecting the COR and kinase domains (Thr-1849 and Ser-z1853) (Supp. Table S2). Seven residues including Ser-910, Ser-912, Ser-926, Ser-935, Ser-955, Ser-973 and Ser-976 located in the segment linking the ankyrin and LRR domains were identified as phosphorylation sites in both the endogenous and recombinant LRRK2 expressed in Swiss 3T3 cells, mouse brain and HEK293T cells (Supp. Table S2) [(Dzamko, et al., 2010; Gloeckner, et al., 2010; Ito, et al., 2007; Li, et al., 2011; Nichols, et al., 2010). It is unclear if they are autophosphorylation sites or sites of phosphorylation by other upstream regulatory kinases.

The exact function of these phosphorylation sites remains unclear. Among those sites in the ankyrin-LRR linker, Ser-910 and Ser-935 are docking sites of several 14-3-3 family proteins (Dzamko, et al., 2012; Li, et al., 2011). Variations that disrupt their phosphorylation abolish 14-3-3 binding, reduce kinase activity and induce LRRK2 translocation to cytoplasmic structures that resemble inclusion bodies (Dzamko, et al., 2010; Nichols, et al., 2010). How these events contribute to the neurotoxic action and physiological function of LRRK2 remains unclear. Intriguingly, introduction of the kinase-dead K1906M and/or D2017A variations only reduces S910/S935 phosphorylation of LRRK2 in cells, suggesting
that they are not autophosphorylated in cells. Instead, they are likely phosphorylated by one or more upstream protein kinases (Dzamko, et al., 2010; Li, et al., 2011; Nichols, et al., 2010). To further ascertain if phosphorylation at these sites is an autophosphorylation event, Dzamko et al. first treated cells with small-molecule inhibitors of LRRK2 kinase activity to suppress LRRK2 autophosphorylation in cells (Dzamko, et al., 2010). They then lysed the cells and isolated LRRK2 from the lysate by immunoprecipitation. After extensive washing to remove the inhibitors that bound to LRRK2 kinase domain, this LRRK2 preparation was incubated with radioactive ATP to allow autophosphorylation. Their results indicate that while the LRRK2 preparation undergoes autophosphorylation at other sites in vitro, no increase in phosphorylation of Ser-910 and Ser-935 was detected. All these results strongly suggest that LRRK2 does not autophosphorylate at Ser-910 and Ser-935 to modulate its binding with 14-3-3 proteins [reviewed in (Lobbestael, et al., 2012)].

There are two reports demonstrating phosphorylation of LRRK2 at Ser-910 and Ser-935 by upstream regulatory kinases. Li et al. provided data that implicate PKA as an upstream regulatory kinase phosphorylating Ser-935 (Li, et al., 2011). They demonstrated that recombinant LRRK2 is phosphorylated by PKA at Ser-935 in vitro. Furthermore, activation of PKA by forskolin induces enhanced phosphorylation of the endogenous LRRK2 at Ser-935. These data suggest that PKA is an upstream kinase phosphorylating LRRK2 at Ser-935. Intriguingly, the sequences around Ser-910 (KKSNSSI) and Ser-935 (RHSNSSL) do not fully conform with the consensus phosphorylation sequence of PKA (R/K-R/K-x-S/T-φ, where x stands for any amino acid and φ represents a hydrophobic residue), suggesting that in addition to determinants located near Ser-910 and Ser-935, determinants located distally from these two residues also participate in recognition of LRRK2 by PKA as a substrate. These distal determinants may mediate binding of LRRK2 to PKA prior to phosphorylation. Of relevance, the regulatory subunit of PKA contains determinants that specifically direct binding of PKA to its substrates. For example, the regulatory subunit of PKA is critical in directing phosphorylation of the NMDA receptor by PKA. Future investigation to define the molecular basis of binding of LRRK2 to PKA will shed light on how LRRK2 is regulated. In a recent study by Dzamko et al., members of the IkappaB kinase family including IKKa, IKKβ, IKKe and TBK1 can phosphorylate LRRK2 at S910 and S935 in vitro and in macrophage cell lines and in bone marrow-derived macrophages (Dzamko, et al., 2012). It is unclear if LRRK2 is also phosphorylated by the IkappaB kinases in neurons.

The physiological phosphatases targeting the various phosphorylation sites in LRRK2 have not been identified. A recent study by Lobbestael, et al. reported that the Protein
Phosphatase 1 (PP1) forms complexes with LRRK2 in cells (Lobbestael, et al., 2013). Furthermore, treatment of cells with PP1 inhibitors suppresses dephosphorylation of Ser-910, Ser-935, Ser-955 and Ser-973 of LRRK2. The results suggest that PP1 is an upstream phosphatase dephosphorylating LRRK2 at these sites.

**INTERACTIONS OF LRRK2 CATALYTIC DOMAINS WITH SMALL-MOLECULE INHIBITORS**

It is generally accepted that in the long term LRRK2 is likely to be a valid drug target (Lee, et al., 2010b; Smith, et al., 2006b). Most of the pathogenic variations of LRRK2 (I1371V, R1441C/G/H, Y1699C/G, G2019S, and I2020T) occur in or near the enzymatic domains, pointing to the therapeutic potential of regulating the GTPase or kinase activity of LRRK2 (Supp. Figure S1 and Supp. Table S1). The G2019S mutant, with up-regulated kinase activity, in particular is found as the most common variation in familial and sporadic PD patients (Smith, et al., 2006c). In the short term, however, further insight into the complex regulatory mechanisms of LRRK2 enzymatic activity is needed before efficacious therapeutic design can be realized (Lee, et al., 2010a). We first present an overview of existing strategies in inhibiting the LRRK2 Ser/Thr kinase activity. This is driven by recent evidence that the LRRK2 inhibitors are neuro-protective in animal models, and that LRRK2 pathogenesis requires its kinase activity, and that kinase inhibitors are effective in treating other chronic conditions such as cancer and diabetes. The potential of targeting the Roc or GTPase domain is discussed secondly as a complementary, albeit less explored, approach to LRRK2 activity regulation.

**LRRK2 kinase inhibitors**

The DFG- motif at the N-terminal end of the activation loop of protein kinases contains the conserved aspartate residue that chelates the magnesium or manganese ions complexed with ATP. The DFG-motif can adopt two distinct configurations: the DFG-in and DFG-out configurations. The Asp and Phe residues in the DFG motif swap positions in two configurations [reviewed in (Dar and Shokat, 2011)]. When the activation loop of protein kinases adopt the DFG-in configuration, the aspartate is appropriately aligned to chelate Mg\(^{2+}\)-ATP or Mn\(^{2+}\)-ATP for phosphotransfer reaction in catalysis. Inspection of the three-dimensional structures of many protein kinases complexed with small-molecule inhibitors reveals two major classes of kinase inhibitors: type I inhibitors that preferentially bind to the ATP-binding site when the activation loop adopts the DFG-in configuration and type II
inhibitors which bind to a pocket adjacent to the ATP-binding site when the activation loop adopts the DFG-out configuration (Zuccotto, et al., 2010). A full description of the designs of the type I and type II LRRK2 kinase inhibitors is presented in Supplemental Information I. G2019 of LRRK2 corresponds to the glycine in the DFG motif. How its variation to serine (as in the [G2019S]LRRK2 mutant) affects interactions of these inhibitors with residues in the active site of LRRK2 is discussed in Supplemental Information Part 1

**LRRK2 GTPase inhibitors**

Currently there are no compounds reported for regulating LRRK2 GTPase activity. It is known that the LRRK2 mutants have reduced GTP-hydrolysis level, and this GTP-bound conformation of the Roc domain enhances the kinase activity (West, et al., 2007). It is therefore reasonable to explore the strategy of inhibiting GTP-binding to the Roc domain or trapping the Roc domain in a non-GTP-binding conformation to reduce the kinase activity of LRRK2. This could prove a complementary approach to finding direct kinase inhibitors of LRRK2. The current lack of GTP-site targeting inhibitors is due to the general difficulty in developing a potent and specific inhibitor for the GTP-site when GTP itself is a high affinity ligand. However, a recent report disclosed the first GTP-site targeting compound, CID1067700 (GTPase inhibitors, Figure S6), specifically for the GTP-conformer of the GTPase Rab7 with nanomolar IC\textsubscript{50} values (Agola, et al., 2012). Additional structure activity relationship studies have identified a few key interaction motifs for its activity. Biochemical analysis and modeling study revealed that CID1067700 binds to the nucleotide-binding site in the active site of Rab7. It can specifically bind to the GTP-conformer of Rab7 with nanomolar IC\textsubscript{50} values. Intriguingly, even though the nucleotide-binding pockets of Rab7 and the other Ras-like GTPases are structurally very similar, CID1067700 binds to all other Ras-like GTPases examined by the authors with affinities much lower than that of Rab7, suggesting that CID1067700 binds to unique determinants in the nucleotide binding site of Rab7. This provides an interesting lead for developing potentially specific inhibitors for the GTP-conformer of LRRK2. The cross interactions of the GTPase domain and the kinase domain are complex, possibly mediated via the COR domain. For example, evidence suggests that phosphorylation of Thr1410 in the COR domain of LRRK2 alters the GTP hydrolysis rate but does not change GTP-binding (Pungaliya, et al., 2010). It is therefore essential to understand the GTP-site interactions at the Roc domain in the presence of the COR or the kinase domain and their ligands.
Challenges and opportunities in targeting LRRK2 activities as therapeutic strategies for PD treatment

The difficulty in developing effective therapeutics for most neurodegenerative diseases stems from the complex etiology of these conditions, as exemplified by the case of LRRK2 (Cuny, 2012). Recent efforts in elucidating the physiological and pathological role of LRRK2 will likely reveal new avenues of PD therapeutic design and point to combinations of therapeutic approaches. For example, Akt was recently found to be a possible physiological substrate of LRRK2, and the phosphorylation of Akt was reduced with LRRK2 mutants (R1441C, G2019S, and I2020T) leading to impaired survival pathways (Ohta, et al., 2011b). It is possible that therapeutics that compensate for this deficiency will be required in addition to LRRK2 activity regulation. Another physiological function of LRRK2 was suggested to be in synaptic vesicle storage and mobilization (Piccoli, et al., 2011). The G2385R variation in the WD-40 domain may disrupt the normal synaptic vesicle control and reduce the release of neurotransmitters. Targeting mutant LRRK2 activity to correct for such vesicle disruption could also be therapeutically relevant.

Most recently, the role of LRRK2 in lysosome targeting and chaperone-mediated autophagy has been investigated (Orenstein, et al., 2013b). WT LRRK2 and [G2019S]LRRK2 both bind to lysosomes for degradation. Increased binding to the lysosome of either the WT or the G2019S mutant can interfere with lysosomal activity and lead to toxicity. The G2019S mutant is degraded more poorly and also interferes with degradation of other proteins such as α-synuclein. Adding GTP or GAPDH does not change the WT LRRK2 or the [G2019S]LRRK2 mutant's binding to lysosomes. However, the addition of a kinase inhibitor (SU6656, a type I inhibitor, Figure S6) to the cells expressing WT LRRK2 increases LRRK2 binding to the lysosome and reduces LRRK2 degradation. Furthermore, introduction of the kinase-dead variation to LRRK2 and its mutants abrogates these effects, indicating the requirement of kinase activity for the defective degradation and higher toxicity. For the G2019S mutant, however, SU6656 does not affect its lysosome binding alone but enhances the binding when another chaperone-mediated autophagy substrate GAPDH is added. This may suggest that the G2019S mutant is already in a conformation that binds the lysosome in a defective way and type I kinase inhibitors may not be effective in reducing this aberrant interaction and could even enhance the defective lysosome binding in the presence of GAPDH and increase toxicity. On the other hand, this may provide a new opportunity in inhibitor design, because if a kinase inhibitor can lock the G2019S mutant in a conformation...
that prevents its binding to the lysosome or GAPDH then this may specifically reduce toxicity of the G2019S mutant. Such in-depth investigations on the mechanistic aspect of LRRK2 mutant toxicity will likely help developing more efficacious therapeutic strategies, and caution should be taken such that the LRRK2 kinase activity is not targeted without the appropriate molecular context.

CONCLUSIONS AND OUTSTANDING QUESTIONS

Knowledge of how the domains of LRRK2 function together, and how upstream regulatory influences are integrated by the domains, will be useful for any intervention strategy specifically directed at LRRK2. Not only will detailed information on the function of each domain suggest new sites to be targeted, but it could also reveal interactions between domains that influence the accessibility of standard target sites, such as the ATP-binding cleft of the kinase domain. Furthermore, characterisation of ligands for the repeat domains will help develop a more comprehensive picture of how LRRK2 is integrated into cell signaling networks. Of particular use would be experimental verification of LRRK2-ligand interactions required for normal and pathogenic signaling.

In addition to full-length LRRK2, two splice variants of LRRK2 mRNA were detected in mice. One variant lacks exon 5 which encodes a putative armadillo repeat [repeat 3 of our prediction (Figure 1A)] (Giesert, et al., 2013). Thus, the putative LRRK2 isoform will have one of the repeats missing in its armadillo repeat domain. Another variant only contains 19 exons and it encodes a putative truncated LRRK2 isoform lacking the armadillo repeat domain and a part of the LRR repeat domain at the N-terminal end as well as half of the kinase domain and all the WD-40 repeat domain at the C-terminus. To establish the physiological significance of the findings, future investigation to confirm endogenous protein expression of these two putative LRRK2 isoforms encoded by the two LRRK2 variants is needed.

Finally, determination of the three dimensional structure of full-length LRRK2 will provide us with the structural detail of interactions among the various functional domains of LRRK2. The report of expression and purification recombinant LRRK2 expressed in HEK293T cells indicates that it is feasible to perform crystallographic analysis and biophysical analyses to determine the three-dimensional structure of LRRK2 (Civiero, et al., 2012). Given the scarcity of information of the three-dimensional structures of the various functional domains of LRRK2, it is hoped that the sequence analysis described in this
manuscript, particularly predictions of potentially important ligand-interacting regions on the repeat domains, may help achieve these aims.
Acknowledgements:
RDM was supported by an Australian Postgraduate Award, a Dowd Foundation Neuroscience Research Scholarship and an AINSE (Australian Institute of Nuclear Science and Engineering) Postgraduate Research Award. TDM, JGC, FL and HCC acknowledge the support of the Australian National Health and Medical Research Council (NHMRC Project Grant No.566743 to JGC and HCC) and the Australian Brain Foundation.

Author contributions:
RDM, HCC, FL and TDM performed bioinformatic analyses, modeling and writing. HCC and JGC undertook project planning and editing.
Figure 1  Models that resemble the predicted repeat patterns of the armadillo repeat domain (amino acids 49-657) and ankyrin repeat domain (amino acids 676-902) of LRRK2.

A. Predicted armadillo repeat domain of LRRK2. The sequences corresponding to the three putative α-helices Helix1, Helix2 and Helix3 are shaded in green. The residues that are varied in familial parkinsonism are in red fonts. LRRK2 consensus corresponds to the LRRK2-specific armadillo-like consensus sequence. ARM consensus is the armadillo repeat consensus sequence derived from cellular proteins with bona fide armadillo repeats (Andrade, et al., 2001). For the LRRK2-specific consensus sequence, residues (or those with similar functional groups) occurring in ≥30% but ≤70% of the repeats is considered as a less conserved residues while those occurring in ≥70% is classified as highly conserved residues. In the two consensus sequences, the highly conserved and the less conserved residues are presented as capital and lower case letters, respectively. A LRRK2 splice variant which lacks the exon 5 of the LRRK2 gene, is expressed at a high level in astrocytes in mouse brain (Giesert, et al., 2013). This splice variant encodes a putative truncated LRRK2 isoform that lacks the segment encompassing residues 146-190 (underlined in red). This segment corresponds to repeat 3 of our predicted model. B. A model resembling the predicted repeat pattern of the LRRK2 armadillo domain was derived from δ-catenin [PDB code 3L6X, (Ishiyama, et al., 2010)] and yeast karyopherin (importin) alpha [PDB code 1EE4, (Conti and Kuriyan, 2000)] to reflect repeat lengths and insert regions (Supp. Table S3). Helix 1, Helix 2 and Helix 3 are represented as yellow, green and red ribbons, respectively. The repeat numbers (numbers in circles) are placed near Helix 1 of each repeat. C. Ankyrin-like repeats in LRRK2 were identified using a LRRK2-specific ankyrin repeat consensus sequence HMM. The sequences corresponding to the putative α-helices H1 and H2 are in green shade. LRRK2 consensus is the LRRK2-specific ankyrin-like consensus sequence. Ankyrin consensus is the consensus sequence from a structurally-characterised ankyrin protein most similar to the LRRK2-specific ankyrin-like consensus sequence, the cytoskeletal component ankyrinR [PDB code 1N11, (Michaely, et al., 2002)]. For the LRRK2-specific consensus sequence, residues (or those with similar functional groups) occurring in ≥30% but ≤70% of the repeats is considered as a less conserved residues while those occurring in ≥70% is classified as highly conserved residues. In the two consensus sequences, the highly conserved and the less conserved residues are presented as capital and lower case letters, respectively.
D. A model resembling the predicted repeat pattern of the LRRK2 ankyrin domain based on the IκB protein Bcl-3 [PDB code 1K1A, (Michel, et al., 2001)] and cyclin-dependent kinase 4 inhibitor (INK4) [PDB code 1BD8, (Jeffrey, et al., 2000)] and ribonuclease L [PDB code 1WDY, (Tanaka, et al., 2000; Tanaka, et al., 2004; Tanaka, et al., 2005)]. Helix 1 and Helix 2 of the repeats are presented as yellow and red ribbons, respectively. The repeat numbers (numbers in circles) are placed near Helix 2 of each repeat. Boxes in panels B and D show portions of alignments for the parkinsonism-associated variations, indicating conservation between Homo sapiens (Hs), chicken Gallus gallus (Gg) and puffer fish Takifugu rubripes (Tr), and sites of variation are shaded in yellow.

Figure 2  A model that resembles the predicted repeat pattern of the leucine-rich repeat domain of LRRK2 (amino acids 984-1320)

A. Predicted leucine-rich peats (LRRs) in LRRK2. Repeat 1 is longer than standard LRRs, while repeat 8 is shorter. The sequence corresponding to the putative conserved β-strand in each repeat is in green shade. The residues that are varied in parkinsonism are shown in red font. The serine residue in red circle is Ser-1292, which undergoes autophosphorylation. The consensus sequence for LRRs derived from proteins with bona fide LRRs is shown at the bottom of panel A (Bella, et al., 2008). Curves represent the two loops that connect the concave side to the convex side, while dots indicate the variable-length region at the convex face, which can be helical in longer repeats (~28 residues). B. A model resembling the predicted repeat pattern of the LRRK2 leucine-rich repeat domain. The various predicted repeat lengths in LRRK2 LRR domain were modeled using repeats from polygalacturonase-inhibiting protein (PGIP) [PDB code 1OGQ, (Di Matteo, et al., 2003)] and the Toll-like receptor TLR3 [PDB code 1ZIW, (Choe, et al., 2005)], combined with a 21-residue repeat from biglycan [PDB code 2FT3, (Scott, et al., 2006)] to represent a potential 20-residue repeat at the centre of the LRRK2 LRR domain (Tables S5, Supplemental Information Part 2). Boxes in panels B and D show portions of alignments for the parkinsonism-associated variations, indicating conservation between Homo sapiens (Hs), chicken Gallus gallus (Gg) and puffer fish Takifugu rubripes (Tr), and sites of variation are shaded in yellow.
Figure 3  Comparison of the structures of H-Ras, Roc domain of LRRK2 and Roco domain of C. tepidum

A. Structures of the Roc domain of LRRK2 and H-Ras in the GTP-bound state (PDB ID: 2ZEJ for LRRK2 Roc domain and 5P21 for H-Ras). The structure of Roc domain of LRRK2 was superimposed onto the structure of H-Ras in the GTP bound state (green and red: H-Ras; Olive green: LRRK2 Roc domain). The locations of the α-helices and β-strands in both structures are indicated. Lower panel: a close-up view of the active sites of LRRK2 Roc domain and H-Ras to show the conserved residues located in Switch 1 (Sw1) and Switch 2 (Sw2) participating in binding of effectors and/or catalysis of GTP hydrolysis in Ras (Thr-35 and Gln-61) and the homologous residues in LRRK2 (Thr-1348 and Arg-1398, respectively). Both Sw1 and Sw2 in the H-Ras structure are shown as red ribbons. Lys-1347 and Thr-1348 of LRRK2, which are predicted to bind guanine nucleotides (Deng, et al., 2008a) are shown. Ile-1371 and Arg-1441 are varied in familial parkinsonism. Two segments corresponding to residues 1357-1368 and residues 1379-1387 that are missing in the LRRK2 Roc domain structures are indicated by dotted lines and labeled as ∆LRR2(1357-1368) and ∆LRR2(1379-1387), respectively. B. Structure of monomer A of the Roc-COR tandem domains of the C. tepidum Roco protein (PDB ID: 3DPU) superimposed on a model of the Roc domain of LRRK2 (Grey ribbons: COR domain of C. tepidum Roco protein; Red ribbons: Roc domain of C. tepidum Roco protein; Olive green ribbons: LRRK2 Roc domain). Lower panel: a close-up view of the interface of Roc and COR domains of C. tepidum Roco protein (monomer A) that shows the residues involved at the interface of the Roc and COR domains of C. tepidum Roco protein. Ile-1371 and Arg-1441 of the Roc domain and Tyr-1699 of COR domain of LRRK2 are mutated in some familial PD patients. Ile-1371 and Arg-1441 of LRRK2 and the homologous residues Leu-487 and Tyr-804, respectively in C. tepidum Roco protein are shown. The three-dimensional structure of LRRK2 COR domain has not been solved. Thus, the location of Tyr-1699 in relation to the Roc domain remains unclear. Since Tyr-804 of C. tepidum Roco protein homologous to Tyr-1699 of LRRK2 is located at the Roc/COR interface, it is logical to predict that Tyr-1699 is also located at the Roc/COR interface of LRRK2.

Figure 4  A model accounting for the activation caused by the Parkinsonism-associated G2019S variation
A. Sequences of the segments containing Gly-2019 and Ile-2020 in human LRRK2 and its orthologues. B. Structure of the [G1179S]Roco4 kinase domain (PDB: 4F1O). Right panel: an enlarged view showing the hydrogen bond formed between the side chains of the mutated residue (Ser-1179) in the activation loop and Arg-1077 in the αC-helix. C. Structure of the [L1180T]Roco4 kinase domain (PDB:4F1O). Right panel: an enlarged view showing Thr-1180 located in close proximity to the residues forming the hydrophobic regulatory spine (R-spine). The corresponding residues mutated in LRRK2 are shown in brackets.

Figure 5  A model that resembles the predicted repeat pattern of WD-40 repeat domain of LRRK2 (amino acids 2164-2515)

A. Arrangement of the predicted WD-40 repeats in the LRRK2 C-terminal region. Location of repeats was predicted by alignment with F-box WD-40 repeat protein 7 (Fbw7), the WD-40 repeat-containing protein most similar to the LRRK2 region with known atomic resolution structure [(Hao, et al., 2007); PDB code 2OVP]. The sequences corresponding to the predicted β-strands A, B, C and D in each repeat are in green shade. The predicted extended loop regions are in pink-shaded boxes. The consensus WD-40 sequence derived from proteins with bona fide WD-40 repeats is shown at the bottom of the panel (Li and Roberts, 2001). B. A model resembling the predicted repeat pattern of the LRRK2 WD-40 repeat domain. To build the model, extended loop regions from WD-40 proteins of Damaged-DNA binding protein 1 (DDB1) [PDB code 3EI4, (Scrima, et al., 2008)] and Chromatin assembly factor 1 (Caf1 [PDB code 3C9C, (Song, et al., 2008)]) were combined with Ribonucleic acid export 1 (Rae1) [PDB code 3MMY, (Ren, et al., 2010)] (Supplemental Information Part 2, Supp. Tables S6). β-strands A, B, C and D of each repeat are shown as green, blue, red and yellow ribbons, respectively. The repeat number of each predicted repeat is shown besides the β-strand A of the repeat. C. G2385R and T2356I are parkinsonism-associated variations. Boxes show portions of alignments for each parkinsonism-associated variation, indicating conservation between Homo sapiens (Hs), chicken Gallus gallus (Gg) and puffer fish Takifugu rubripes (Tr), and sites of variations are shaded in yellow.
References


Berger Z, Smith KA, Lavoie MJ. 2010. Membrane localization of LRRK2 is associated with increased formation of the highly active LRRK2 dimer and changes in its phosphorylation. Biochemistry 49(26):5511-23.


Fung HC, Chen CM, Hardy J, Singleton AB, Wu YR. 2006. A common genetic factor for Parkinson disease in ethnic Chinese population in Taiwan. BMC Neurol. 6:47.


of leucine-rich repeat kinase 2 to the endosomal-lysosomal compartment in lewy body disease. J Neuropathol Exp Neurol. 68(9):994-1005.


Pandey N, Fahey MT, Jong YJ, O'Malley KL. 2012. Sequences located within the N-terminus of the PD-linked LRRK2 lead to increased aggregation and attenuation of 6-hydroxydopamine-induced cell death. PLoS One 7(9):e45149.


LRRK2 link enhanced GTP-binding and kinase activities to neuronal toxicity. Hum Mol Genet. 16(2):223-232.


Figure 1
Figure 2
Figure 3
Figure 4
Figure 5
Author/s:
Mills, RD; Mulhern, TD; Liu, F; Culvenor, JG; Cheng, H-C

Title:
Prediction of the Repeat Domain Structures and Impact of Parkinsonism-Associated Variations on Structure and Function of all Functional Domains of Leucine-Rich Repeat Kinase 2 (LRRK2)

Date:
2014-04-01

Citation:

Persistent Link:
http://hdl.handle.net/11343/49796