

## Structural and functional *in silico* analysis of LRRK2 missense substitutions

Fernando Cardona<sup>1,2,\*</sup>, Marta Tormos-Pérez<sup>1,3,#</sup>, and Jordi Pérez-Tur<sup>1,2,\*</sup>

<sup>1</sup>Unitat de Genètica Molecular. Institut de Biomedicina de València (C.S.I.C.). Valencia, Spain.

<sup>2</sup>CiberNed. Spain

<sup>3</sup>Departamento de Biotecnología. Escuela Técnica Superior de Ingenieros Agrónomos. Universitat Politècnica de València. Valencia, Spain.

# Current address: Instituto de Biología Molecular y Celular del Cáncer, Consejo Superior de Investigaciones Científicas (CSIC), Salamanca University, Salamanca, Spain

\* Corresponding authors: [fcardona@ibv.csic.es](mailto:fcardona@ibv.csic.es) and [jpereztur@ibv.csic.es](mailto:jpereztur@ibv.csic.es)

Institut de Biomedicina de València, Consejo Superior de Investigaciones Científicas (IBV-CSIC). Jaume Roig 11, València 46010, Spain.

Contract grant sponsor: The Ministerio de Ciencia e Innovación, Spain; Contract grant number:

SAF2009-10434; Contract grant sponsor: Generalitat Valenciana; Contract grant number:

ACOMP/2010/096.

## **Abstract**

The *LRRK2* gene (*Leucine-Rich Repeat Kinase 2*, PARK8) is mutated in a significant number of cases of autosomal dominant Parkinson's disease (PD) and in some sporadic cases of late-onset PD. LRRK2 is a large, complex protein that comprises several interaction domains: armadillo, ankyrin, leucine-rich repeats and WD40 domains; two catalytic domains: ROC-GTPase and serine/threonine kinase; and a COR domain (unknown function). Pathogenic mutations are scattered all over the domains of LRRK2, although the prevalence of mutations in some domains is higher (ROC-GTPase, COR and kinase). In this work, we model the structure of each domain to predict and explore the effects of described missense mutations and polymorphisms. The results allow us to postulate the possible effects of pathogenic mutations in the function of the protein, and hypothesize the importance of some polymorphisms that have not been linked directly to PD, but act as risk factors for the disease. In our analysis, we also study the effects of PD-related mutations in the kinase domain structure and in the phosphorylation of the activation loop to determine effects on kinase activity.

**Key words:** function, kinase, LRRK2, Parkinson's disease, stability, structure

## Introduction

Parkinson's disease (PD) is a neurodegenerative disorder caused by genetic and environmental factors. Clinic features are resting tremor, bradykinesia, rigidity and postural instability, caused by the progressive loss of dopaminergic neurons in the *substantia nigra* of the central nervous system (CNS).

To date, 15 *loci* associated with the appearance of several variants of the disease have been identified [1]. In 2004, our group identified mutations in the *LRRK2* gene (*Leucine-Rich Repeat Kinase 2*, *PARK8*, MIM# 607060) as being responsible for some forms of PD [2]. Its gene product (dardarin or LRRK2) is a multidomain protein of 2527 amino acids, with an approximate molecular weight of 280 kDa. The function of this protein is not fully known, although several studies have indicated a role in vesicle traffic, cytoskeleton dynamics, signal transduction, translation regulation, apoptosis, oxidative stress response and mitochondrial function. Furthermore, there is evidence to suggest a common pathway with  $\alpha$ -synuclein (MIM#163890), parkin (MIM# 602544) and TAU (MIM# 157140) in PD [1]. Today it is considered that some specific mutations (e.g., p.R1441G/C or p.G2019S) are a major cause of the disease since they are responsible for approximately 2-4% of cases in Caucasian populations [3]. The phenotype associated with these mutations is late-onset PD, commonly indistinguishable from the idiopathic disease. Moreover, the neuropathology associated with LRRK2 mutations is variable, suggesting that this protein is involved in several pathways. Dardarin-associated PD causes loss of dopaminergic neurons and sometimes abnormal accumulation of ubiquitinated proteins, mainly  $\alpha$ -synuclein and occasionally hyperphosphorylated tau [4].

LRRK2 comprises several independent domains, some of which are typically involved in protein-protein interaction: armadillo (ARM), ankyrin (ANK), leucine-rich repeats (LRR) and WD40 domains. The N-terminal region of LRRK2 contains seven ankyrin-type repeats that form the ANK domain. In addition, 13 leucine-rich repeats (LRR) have also been identified in this region. This domain participates in interactions with other proteins through the extensive surface with accessibility to the solvent. There have also been predicted C-terminal repeats that form a WD40 domain [5]. The presence or absence of the armadillo domain in the N-terminal has been a controversial matter, but recent studies conclude that LRRK2 contains armadillo-type repeats (ARM-like) [6]. Moreover, LRRK2 encodes two enzymatic activities: kinase and GTPase. The GTPase domain belongs to the ROCO family, with the GTPase typically in tandem with a COR domain. The ROC-COR (ROCO) module is conserved throughout evolution, suggesting the functional interdependence of the two domains [6]. Activation of the

kinase domain (KIN) probably occurs by phosphorylation in a sequence of 25-30 amino acids loop that is flanked by conserved motifs DF/YG and APE, whose conformation changes to allow access to the substrate [5]. These conformation changes may be mediated by molecular interactions such as hydrogen bonds [7].

Some authors have proposed that LRRK2 predominantly forms a dimer *in vivo* that autoregulates itself by phosphorylation, and it seems that dimerization takes place through several LRRK2 regions, but the exact parts involved in this phenomenon remain unclear [8]. Some authors postulate that the ROCO domain is critical, but not exclusive, to dimerization [9]. Although other proteins of the ROCO family form dimers through COR-COR interactions [10], these interactions were not observed in LRRK2, and there is evidence for ROC-ROC interactions obtained from three-dimensional structures [11]. This model suggests that residues R1441 and I1371, associated with PD, are located on the dimerization surface and stabilize the dimer.

At least 20 mutations linked to autosomal dominant PD, and approximately the same number for sporadic PD, have been described. Mutations in LRRK2 are the most common genetic cause of autosomal dominant PD, especially in certain populations [12]. The p.G2019S mutation has been detected in 5-6% of familial PD, and in 1-2% of sporadic cases. This prevalence can be higher in specific populations such as Ashkenazi Jews and Arabs from North Africa. Dardarin mutations are found mostly at the C-terminal half, suggesting that it is more important at a functional level. However, some are also found in the N-terminal region along the ARM, ANK and LRR domains, and some authors have postulated that mutations in these domains can interfere, similarly to those in WD40, with the binding to other proteins or molecules [5]. The most studied mutations are those found in catalytic domains. p.G2019S gives rise to a hyperactive kinase, increasing autophosphorylation and phosphorylation of generic substrates [13]. Other mutations commonly found in patients with PD, such as p.R1441G/C, which affects the GTPase domain, appear to reduce this activity, but enhance kinase activity [4], whereas the p.R1441H mutation inhibits this activity. In the WD40 domain, the p.G2385R mutation has also been described as attenuating kinase activity [14]. There have also been reports that pathogenic mutations p.R1441C, p.Y1699C or p.G2019S cause neuronal degeneration, protein aggregates or intracytoplasmic inclusions in cultured neural cells [15], as well as increased apoptosis [16]. In general, the overexpression of mutant LRRK2 causes loss of viability, whereas inactivation of the kinase reduces this toxicity [15]. However, the influence of different amino acid substitutions in protein kinase activity is controversial, and the mechanisms by which dardarin

mutations cause PD appear to be complex and cannot be explained only by kinase activity. In this work, we intend to provide new evidence for and an appropriate framework to determine the pathogenicity of the genetic variants in dardarin by performing an *in silico* analysis of the variants described as pathogenic or polymorphic.

## Methods

The tertiary structure models were performed by sequence homology using EasyModeller [17] with default settings for the LRR, COR and KIN domains, for which a suitable template was found. The rest of the domains were modeled by threading for the LOMETS meta-server (<http://zhanglab.ccmb.med.umich.edu/LOMETS/>) with default settings, and by the best model obtained, based on the z-score and the absence of unfolded areas. The alignments of the templates and the LRRK2 sequences, as well as the percentage of similarity and threading parameters, are shown in **Suppl. Fig. 1**. A Mutational analysis was performed, as previously described [18], for the missense substitutions described in these domains according to the Human Gene Mutation Database (<http://www.hgmd.cf.ac.uk>). To view and edit the structures, we used the PyMOL software (DeLano Scientific Inc). The structures obtained for each wild-type domain were deposited in the Protein Model Database (<http://mi.caspur.it/PMDB/>). For the phosphorylation motif analysis, we used GPS 2.1 (<http://gps.biocuckoo.org/online.php>) and NetPhos/NetPhosK (<http://www.cbs.dtu.dk/services/NetPhos/>; <http://www.cbs.dtu.dk/services/NetPhosK/>). For the analysis of stability and function/pathogenicity alterations, several methods were used [19-27] and the results are provided in **Suppl. Table 1** and **Suppl. Table 2**. Those results that were coincident in at least 75% of the analyses were accepted.

## Results

To predict the structural effects of the PD-related mutations, the structure of each LRRK2 domain and the possible alteration caused by each missense mutation were analyzed *in silico*. Previously, as some domain definition, and even the presence or absence of the ARM domain, is controversial [5-6], we also carried out a redefinition of the domain boundaries based on the predicted, or previously resolved, tertiary structures. The results obtained (**Fig. 1**) indicate the presence of the ARM domain, and also some differences from other works [5] in a given domain definition. So the COR and KIN domains were 54 and 8 amino acids smaller than expected, respectively, whereas the LRR and ROC domains were respectively 5 amino acids bigger than previously described.

Afterward, we predicted the effect of the missense mutations related to PD (**Fig. 2**) in secondary and tertiary structures, on the electrostatic surface, in polar contacts and charge distribution, as well as stability, function alteration and possible pathogenicity.

### **ARM domain**

According to the model obtained (**Suppl. Fig. 2a**), this domain is formed by 21  $\alpha$ -helices grouped into threes to form supercoils and are linked by loops to form a curved structure where loops were more exposed to the solvent. The tertiary structure prediction was made by threading, and the best model obtained was selected by using the  $\alpha$ -importin of *Saccharomyces cerevisiae* as a template (PDB ID 1wa5B). This domain contains three described amino acid substitutions: p.A211V, p.E334K and p.N363S. Whereas no significant change was observed for the p.A211V substitution, the results show that the p.E334K mutation, described as PD-associated, affected the electrostatic surface and charge distribution in this domain (**Suppl. Fig. 2a**). The p.N363S mutation determined the loss of two hydrogen bonds with H365 and V366, which were in an  $\alpha$ -helix near a loop (**Suppl. Fig. 2a**), and which may affect its flexibility and stability (**Table 1**). This amino acid substitution also generated a target motif for PRKG1 (**Fig. 3a**), which performs important functions in nervous system development and regulation, and has been related with axon branching and neurite elongation [28].

### **ANK domain**

This domain is formed by 11  $\alpha$ -helices linked by loops that are more exposed to the solvent (**Suppl. Fig. 2b**). The three-dimensional model of this domain was run by threading using the human gankyrin (PDB ID 1uohA) as a template. In this domain, four substitutions, p.M712V, p.P755L, p.R793M and p.I810V, are described. Of those, p.M712V and p.I810V were not clearly associated with PD [29]. Nevertheless, and as seen in **Table 1** and **Suppl. Fig. 2b**, the p.M712V model showed that this mutation would affect both protein stability and the electrostatic surface. Moreover, p.P755L and p.R793M were described as being PD-related. The former affects protein stability (**Table 1**), whereas the latter alters the function, charge distribution and electrostatic surface (**Table 1** and **Suppl. Fig. 2b**). These effects noted on the electrostatic surface have been previously predicted by other authors [30].

### **LRR domain**

The tertiary structure of this domain (**Suppl. Fig. 3**) was modeled by homology using Lingo1 as a template (PDB ID 2id5C). The model shows that this domain is composed of 13 parallel  $\beta$ -sheets linked by large loops to form a curved structure that is highly exposed to the solvent. Similar results have been

previously described for this domain [31]. Eight amino acids substitutions were described along this domain, four of which were considered to be pathogenic: p.Q1111H, p.I1122V, p.I1192V and p.S1228T. The p.R1067Q polymorphism caused the loss of a positive charge, the alteration of the electronic surface and it altered domain stability (**Suppl. Fig. 3**). The effects of this mutation on the surface charge have been previously hinted at by other authors [30-31]. Otherwise, p.Q1111H, p.I1122V and p.S1228T affected the stability of the domain, p.L1165P altered the function of the protein (**Table 1**) and p.A1151T caused the appearance of a hydrogen bond with p.K1129 (**Suppl. Fig. 3**), belonging to a near loop, which could imply increased rigidity between these loops. p.R1067Q, p.Q1111H, p.I1122V, p.I1192V and p.S1228T have been previously predicted as being localized at the hydrophobic core of the domain, and could affect interaction with other proteins, but without affecting the structure [31].

Pathogenic mutation p.S1228T introduced a motif for MAP3K7 (**Fig. 3a**), which is involved in neurite outgrowth and neuronal death [32]. This variant also affected the distribution of the hydrogen bonds (**Suppl. Fig. 3**), so p.S1228 formed hydrogen bonds with S1203, N1230 and S1250, while p.T1228 did so only with S1203 and N1230, which could weaken that with the loop, which contains S1250. Substitution p.H1216R implied a gain of a positive charge on the domain surface (**Table 1** and **Suppl. Fig. 3**). No significant changes were detected for p.S1096C and p.I1192V.

Polymorphism p.A1151T also introduced a phosphorylation motif for several kinases (**Fig. 3a**), as GSK3, which phosphorylates Tau to produce amyloid-beta peptides in Alzheimer's disease, is necessary for axon formation [33].

Although p.I1192 has been described to be pathogenic by Nichols and colleagues [34], no clear evidences for this association has yet been provided because diagnosis in some cases is unclear, and only two of the four affected members of the family who could be tested present this change. Besides, major changes were caused by polymorphisms p.R1067Q and p.L1165P (**Table 1** and **Suppl. Fig. 3**), so perhaps these mutations should be studied further.

### **ROC domain**

The structure of this domain was solved by X-ray crystallography in 2008 (PDB ID 2zej) [11], and comprises six  $\beta$ -sheets and five  $\alpha$ -helices. We modeled and analyzed the different substitutions described for ROC by homology using this structure as a template. This domain (**Fig. 4a**) is formed by five  $\alpha$ -helices and six  $\beta$ -sheets that are linked by loops exposed to the solvent. Eight amino acid substitutions are described in this domain, and six are associated with PD; of those, three affect the same

residue, p.R1441. As shown in **Fig. 4a**, the three changes at this position implied loss of a positive charge and affected the domain surface. These mutations are on the dimerization surface and probably hinder this process. This amino acid is probably involved in forming hydrogen bonds between two monomers of LRRK2 to stabilize the dimer (**Suppl. Fig. 4**), as previously proposed [11]. Loss of p.R1441 implied the loss of these bonds with p.F1401 and p.T1404 at the other monomer. According to our results, the three mutations on R1441 had an effect on protein stability, whereas only p.R1441C/G affected the function and were predicted as pathogenic. However, the p.R1441H mutation had a drastic effect on the tertiary structure as it introduced torsion into this domain, and changed the inner and outer angles, as well as total domain length (**Fig. 4b**), and probably difficult dimerization, similarly to p.R1483Q, which was not described as pathogenic. p.I1371V and p.R1514Q had no apparent structural effects, but were predicted to cause stability changes (**Table 1**). According to our results, p.K1468E was predicted as functionally altered/pathogenic, and it changed the charge distribution (**Table 1**) and the electrostatic surface of this area (**Fig. 4b**). No significant changes were detected for p.A1442P.

It is remarkable that almost all the substitutions described for this domain were associated with PD and that almost all of them were predicted to produce marked effects on the domain structure, stability or function, which were especially obvious for the mutations at position 1441. This demonstrates the importance of this domain in the function of this protein, if it affects its activity and/or binding of substrates, or if it affects the dimerization of the protein.

### **COR domain**

The structural model of this domain was performed by homology using the previously solved structure (PDB ID 3dpuA) as a template for the homologous domain in prokaryote *C. tepidum* [10]. According to this model, the domain is composed of nine  $\alpha$ -helices and six  $\beta$ -sheets that are linked by loops exposed to the solvent (**Fig. 4c**). The N-terminal region contains six  $\alpha$ -helices and two small  $\beta$ -sheets, while the C-terminal region has three  $\alpha$ -helices and four consecutive antiparallel  $\beta$ -sheets. The prediction for human COR differed primarily from the prokaryote homolog in its C-terminal half, an area responsible for dardarin dimerization in *C. tepidum*, thus supporting that in human dardarin, dimerization occurred through its ROC domain, as previously proposed [11]. For this domain, there were two PD-related mutations, p.R1628P and p.Y1699C, and other six coding substitutions, two of them at position p.R1728. The models show that pathogenic mutation p.R1628P was predicted as such (**Table 1**), and that it drastically altered the secondary structure of the domain by abolishing the formation of two  $\beta$ -sheets



with, in addition, loss of a positive charge and changes in the electrostatic surface (**Table 1** and **Fig. 4c**). Mutation p.Y1699C was also predicted as being functionally altered (**Table 1**), with a slightly altered electrostatic surface (**Fig. 4c**). Furthermore, p.R1725Q and p.R1728H affected stability, and the latter also affected the protein function. For these two changes, as well as for p.R1728L, the models reveal loss of a positive charge together with the alteration to the electrostatic surface (**Table 1** and **Fig. 4c**). No significant changes were detected for p.V1613A and p.L1795F.

As in the ROC domain, almost all the substitutions described in the COR domain affected stability, function or structure (**Table 1** and **Fig. 4c**). The reason why only p.R1628P and p.Y1699C were associated with PD could be that they are at the interface with the ROC domain, and could therefore affect the signal transduction between the GTPase and kinase domains [10]. This seems especially obvious for p.R1628P, which abolished the formation of two  $\beta$ -sheets and removed a positive charge in this area, which probably made the interactions and the domain structure especially difficult.

One mutation that has been recently described for this domain (p.S1761R) [35] presents a homolog residue in *C. tepidum* (S852). Previous results obtained with our model suggest that this mutation can affect dimerization through this domain [35].

### **KIN domain**

According to the model obtained (**Suppl. Fig. 5a** and **Fig. 5**), this domain is formed by seven  $\alpha$ -helices and six  $\beta$ -sheets linked by loops, which are exposed to the solvent. The structure model was performed by homology using the Roco4 kinase domain (PDB IDs 4f0f, 4f0g, 4f1m and 4f1o), a protist homologous protein whose kinase domain was recently solved, [36] as a template. Six mutations have been broadly described in this catalytic domain: p.R1941H, p.Y2006H, p.I2012T, p.G2019S, p.I2020T and p.T2031S.

The effect of amino acid substitution in the activation loop is particularly interesting as this loop regulates kinase activity. In fact, the most frequent mutation in PD, p.G2019S, is found in this loop, and has been demonstrated to give rise to enhanced kinase activity [37,14,38]. There are two plausible explanations for this; introducing a phosphorylation site (most of the mutations introduce an S or a T) or stabilizing active or inactive forms by hydrogen bonds [36]. Note that these possibilities are not exclusive, so the phosphate group could also alter the hydrogen bond distribution. In this work, we analyzed both possibilities.

The p.G2019 equivalent amino acid in the *Dictyostelium discoideum* Roco4 kinase domain has been proposed to introduce a hydrogen bond with p.R1918, which stabilizes the active form. This interaction cannot occur in the inactive form as amino acids are located too far away from each other [36]. We modeled the human domain using this structure as a template (**Fig. 5**) and also found an extra hydrogen bond in the active form (**Fig. 5a**), but with p.Q1919. The p.S2019 variant introduced an extra bond between itself and p.I2020, which could strengthen the polar contacts. p.T2031S substitution, which has been described to increase the active kinase form [38], also introduced an extra hydrogen bond, which could stabilize this form (**Fig. 5a**). In turn, p.I2020T formed an extra hydrogen bond, which could stabilize the inactive form (**Fig.5b**). Some authors have described this form as an equal [38] or an overactivated kinase [37], whereas other studies have described greater ATP-affinity and kinase activity in this variant [39].

The p.I2012T variant, near the catalytic loop, showed an extra hydrogen bond in both the active and inactive forms, but with different amino acids (**Fig. 5c** and **5d**), which seemed to contribute to maintain conformation. In the active conformation, it bound to an amino acid of one helix involved in substrate recognition (**Fig. 5c**), while it bound to the same loop in inactive conformation (**Fig. 5d**).

p.I2012T, p.G2019S and p.I2020T are suggested as being pathogenic and are in the activation loop of the kinase domain (**Suppl. Fig. 5a**), and they introduce or change phosphorylation motifs (**Fig. 3b**). Thus, the p.I2012T mutation removed a putative phosphorylation motif for EGFR/Alk tyrosine kinases, but introduced a new motif for the PRKG1/MAP2K2 serine/threonine kinases. Otherwise, p.G2019S introduced new motifs for JAK3/BLK tyrosine kinase and PKA serine/threonine kinase. Finally, p.I2020T created target motifs for BLK tyrosine kinase and TAK1 serine/threonine kinase, without affecting the EGFR/Alk motif. The fact that the different mutations altered several recognition targets is particularly interesting if we consider that the activity of protein kinases is usually regulated by the phosphorylation of the activation loop, and that some of these mutations have been described as possessing increased kinase activity. Besides, p.T2031S, which is also in this loop, forms part of an autophosphorylation target motif (**Fig. 3b**) that is also related with kinase activity regulation [40]. In fact, although T2031, S2032 and T2035 have been described as autophosphorylation motifs, and Ser-2032 and Thr-2035 are important sites that regulate LRRK2 kinase activity [40], our results suggest that they can be recognized by other important kinases (**Fig. 3b**). p.T2031S substitution changed a MAP3K motif for a MAP3K11 one, and removed a CK1/GRK1 phosphorylation site (S2032). p.Y2006H substitution also

altered the activation loop phosphorylation pattern by removing a tyrosine target of Fes kinase (**Fig. 3b**). Moreover in p.R1941H and p.Y2006H, the electrostatic surface was also altered (**Suppl. Fig. 5a**).

Some structural and stability effects were observed in p.R1941H and p.Y2006H polymorphisms (**Suppl. Fig. 5a** and **Table 1**), which support the idea that they can be PD risk factors.

### **WD40 domain**

According to the model obtained (**Suppl Fig. 5b**), this domain is formed by 30  $\beta$ -sheets linked by loops exposed to the solvent to form a  $\beta$ -propeller. The model was run by threading and using mouse holoenzyme paf-ah (PDB ID 1vyhC) as a template. For this domain seven amino acid changes have been described, two of which have been suggested as being pathogenic (p.T2356I and p.G2385R), and the results indicate major changes for these mutations (**Table 1**). In addition, these mutations are located three-dimensionally near each other, indicating that their position can be especially important for the function as it alters the interaction with other proteins or ligands, as proposed elsewhere [30]. Our results predict stability and electrostatic surface alterations for p.R2143H (**Table 1** and **Suppl. Fig. 5b**), negative charge loss and electrostatic surface alteration for p.D2175H, and protein stability and polar contacts distribution alteration for p.T2356I; in the last case, stability affectations were probably due to loss of hydrogen bonds (**Table 1** and **Suppl. Fig. 5b**). p. G2385R implied a gain of positive charge and electrostatic surface alterations due to the larger lateral chain size (**Table 1** and **Suppl. Fig. 5b**). p.V2390M affected stability, whereas p.L2466H affected stability and function, also including a gain of a positive charge (**Table 1**). Finally, p.Y2189C and p.T2356I removed the Tyk2/JakA and MAPK2K/LKB motifs, respectively (**Fig. 3a**).

## **Discussion**

One of the major challenges that the interpretation of genetic variants faces is translating the finding of a genetic substitution for a biologically sound mechanism that can link it with a given disorder. This is of particular interest when a gene is prone to show genetic variability that is not apparently associated with a disease caused by mutations in that same gene. This is the case of dardarin, which is responsible for familial forms of PD [2], but with a high degree of coding substitutions [41]. In line with this, we analyzed 20 amino acid substitutions that were reported to be pathogenic and 25 other coding substitutions that are allegedly polymorphisms. Usually, the pathogenicity of each new identified variant is predicted by taking into account its segregation in a particular pedigree, when it is found to be the result of genetic screening, when the phylogenetic conservation of the residue is affected and with the alteration

of functionally important residues when the protein function is known or can be inferred. We suggest that the molecular modeling of dardarin is a useful tool to predict the protein structure and the effect of missense substitutions to, therefore, contribute to assign a role in the pathogenicity of the molecule. Most of the mutations involved in PD, as well as many of the described polymorphisms, variably affect the secondary or tertiary structure, the electrostatic surface, hydrogen bonds, charge distribution, stability, the function or the phosphorylation sites involved in its regulation. The results of our analysis can also help to establish working hypotheses for future experiments.

First, we define the domain boundaries based on predicted or previously resolved structures. The results obtained (Fig. 1) show the presence of an ARM domain (which has been controversial) and some differences with other works [5] in some domain definitions. Hence the COR and KIN domains are 54 and 8 amino acids smaller than expected, respectively, whereas the LRR and ROC domains are respectively 5 and 2 amino acids bigger than previously described. These differences may well be important because they change the domain where polymorphism p.R1514Q is sited, which was usually sited in the COR domain and the ROC domain in our model (**Fig. 2**).

Our results indicate different mechanisms of action for the mutations analyzed herein. First, through the alteration of protein-protein interaction domains, including dimerization; second, by the modification of phosphorylation motifs; third by structural alterations, including changes in the secondary or tertiary structure, polar contacts distribution, and modification of torsion and angles.

Some substitutions produce major changes in the surface of the affected domains, which may be critical to determine the molecules with which dardarin can interact. This can explain, at least partially, the pathogenicity of mutations p.E334K (ARM), p.R793M (ANK), p.Q1111H (LRR) and p.G2385R (WD40) (**Suppl. Fig. 2, Suppl. Fig. 3 and Suppl. Fig. 5b**). The electrostatic potential of the surface, which is also important for the efficiency and stability of the interaction, also significantly altered in these mutants, except for p.G2385R. This alteration is not exclusive of pathogenic mutations as some polymorphisms, for example p.R1067Q and p.H1216R in the LRR domain or p.D2157H and p.R2143H in WD40, also produce an alteration on the electrostatic surface of the protein. This alteration would suggest that these polymorphisms act as genetic risk factors for the disease.

With the ROC and COR domains, some major changes on the electrostatic surface were also detected for some mutations (p.R1441G/C/H and p.R1628P) and polymorphisms (p.K1468E and p.R1728H/L) (**Fig. 4a and Fig. 4b**). In these cases, pathogenicity can be explained by the role of the ROC

domain in the dimerization of dardarin (p.R1441) [11,10] or by the interaction with other molecules, such as GTP, for the ROC-GTPase domain.

In addition, the phosphorylation pattern can be affected by amino acid changes, which thereby affect the functional regulation of the protein. Note that this protein is also related to the ERK/MKK pathways [1]. In some cases, the interaction (STE20 family) or the direct phosphorylation (PKC) by these kinases has been demonstrated [42].

No drastic structural changes were observed for the pathogenic mutations of the kinase domain (**Table 1**), although stability and function were predicted as being affected in most cases. It is noteworthy that these substitutions are in the activation loop, and that they introduce phosphorylatable residues, which alter the phosphorylation pattern of this loop (**Table 1** and **Fig. 3b**), and whose phosphorylation is important for regulating kinase activity [43]. p.G2019S is in the conserved region DFG/DYG, so changes in the side chain of these positions can affect its activation, even though they do not affect the structure. In fact it has been postulated [44] that p.G2019S and p.I2020T provoke a conformational change that mimics the activated state, especially for p.G2019S, by abolishing the flexible nature of this conserved region by replacing glycine. We propose that this flexibility is affected by changes in the pattern of the polar bonds, so the p.S2019 variant establishes an extra hydrogen bond with p.I2020 (**Fig. 5a**). p.I2020T also forms an extra hydrogen bond, which can stabilize the inactive form (**Fig. 5b**), which could explain the decrease in kinase activity described previously [37]. The generation of an additional autophosphorylation site in the activation domain could also reinforce this increased activity [43]. In turn, p.I2012T introduces a phosphorylation target for PRKG1 and MAP2K2, and probably contributes to maintain the (active/inactive) conformation by polar contacts (**Fig. 5c** and **Fig. 5d**). In addition, p.G2019S removes a motif for EGFR/Alk, and introduces two new motifs for JAK3/BLK and PRKAR1A, whereas p.I2020T also introduces motifs for BLK and TAK1, but does not remove any EGFR/Alk site. Some of these kinases play key roles in the nervous system function and development. EGFR is involved in axon regeneration and neurite outgrowth [45], Alk in neuroblastoma development [46], Fes in axon guidance [47] and CDK1 in neuron degeneration [48]. p.T2031, p.S2032 and p.T2035 are auto-phosphorylation motifs in the activation loop [40], and they present phosphorylatable sequences by other kinases (**Fig. 3b**). p.T2035 lies in a putative recognition site for several kinases that are important for the nervous system, such as GSK3 and CDK1 and DYRK1A, and is involved in neuronal differentiation in response to nerve growth factor [49], which phosphorylates  $\alpha$ -synuclein to facilitate its aggregation [50].

Therefore, the non physiologic phosphorylation of mutant forms of dardarin by serin/threonin kinases, such as MAP2K2, PRKAR1A or MAP3K7, as well as changes in the autophosphorylation pattern (**Fig. 3**), can explain the increased (p.G2019S and p.T2031S) or decreased (p.I2012T and p.I2020T) kinase activity observed in these mutant forms [15,37,14,38].

Some of the tested substitutions affect the number or distribution of polar contacts, which is the case of polymorphisms p.N363S (in the ARM domain, **Suppl. Fig. 2a**) and p.A1151T (LRR, **Suppl. Fig. 3**), or pathogenic mutations p.S1228T (**LRR, Suppl. Fig. 5**) and p.T2356I (WD40, **Suppl. Fig. 5b**). An alteration to these polar contacts can affect the stability or function of the protein (affected in most of these substitutions; **Table 1**), as well as the interaction with other domains or proteins. In fact at the ROC-GTPase domain, p.R1441 mutations can also hinder the dimerization of LRRK2 in this way. As shown in **Suppl. Fig. 4**, p.R1441 forms hydrogen bonds with p.F1401 and p.T1404 of the other monomer, which probably does not occur in the mutants. In fact, the p.R1441H mutation probably involves the gain of a hydrogen bond within the monomer [43]. At the ROC domain, some dardarin auto-phosphorylation sites have also been identified [51], one in p.T1404, which postulates that phosphorylation at this position is required for phosphorylation of the p.T1491 residue involved in GTP binding and, therefore, in kinase activation. From this information, it can be hypothesized that position p.T1404 should be more accessible to phosphorylation through loss of the interaction with p.R1441. Therefore the substitutions at position p.R1441 can increase kinase activity through this mechanism.

At the ROC domain some LRRK2 autophosphorylation sites have also been identified [51], one in p.T1404, which postulates that phosphorylation at this position is required for the phosphorylation of the p.T1491 residue involved in GTP binding and, therefore, in kinase activation. Based on this information, it can be hypothesized that position p.T1404 should be more accessible to phosphorylation through loss of the interaction with p.R1441, therefore the substitutions at position R1441 can enhance kinase activity through this mechanism.

Liu and colleagues [52] used molecular modeling and docking, and proposed that the increased kinase activity seen for p.G2019S might be due to electrostatic alterations on the protein surface, which would result in substrate recognition alterations, but in a substrate-dependent way. Nevertheless for this mutation, it was not possible for us to detect any significant changes in electrostatic surface potential. This discrepancy might be related to specific effects given the substrate used by Liu and colleagues rather than to the effect that the mutation had. In a recent study [53], the same group described the effect of

p.G2019S substitution in kinase stabilization. Although these authors did not find any differences between the models they obtained with both templates, their results differ from ours in that the amino acid established the hydrogen bond with the p.S2019 variant. Along these lines, the differences we observed can be related to local discrepancies which might be lost when globally comparing the modeling obtained with the two templates used by these authors. In their model, the bond with p.G2019 was established with p.E1920, and with p.E1920 and p.Y2018 for p.S2019. In our model it was established with p.I2020, a residue that is also involved in PD, and with p.Q1919. These discrepancies can be due to the differences found in the templates used; whereas Liu and colleagues used B-raf kinase, we utilized Roco4 kinase. It is important to note that in the *D. discoideum* Roco4 kinase structure, p.R1077 (the equivalent to p.R1918 in human LRRK2) interacts only with the equivalent mutant form p.S1179 (p.S2019 in human), but not with the wild-type form p.G1179. For this reason, the role of this substitution in the stabilization of the active form in LRRK2 remains unclear.

Moreover, although several polymorphisms have been described along LRRK2 with no association with PD, this analysis shows that many of them may affect the structure, stability or function. This can be explained if these polymorphisms are found in the protein regions not involved in those functions related to the binding of other substrates, catalytic activity or dimerization. Another possible reason is that the effect of these polymorphisms was compensated by the non mutated copy of the protein. Nonetheless, we must also consider the possibility that these mutations are involved in PD, but they exert a mild enough effect to not result in a significant change in the genetic risk for the disease. Association studies of polymorphisms in dardarin have proven the existence of significant differences in the penetrance of pathogenic mutations according to race, or even to the population analyzed [54]. This raised the possibility that some mutations with a pronounced effect on the protein, such as p.N363S (ARM), p.R793M (ANK), p.R1067Q (LRR) or p.R1728H/L (COR), may be related to PD in some populations, and perhaps in combination with other population-specific environmental and genetic factor.

As expected, our analysis shows how the mutations that affect the enzymatic and the COR domains result in increased alterations on the structure, the stability or the function of these domains (**Table 1**, **Suppl. Table 1** and **Suppl. Table 2**), thus confirming the key role of those domains in the function of the protein.

It is also remarkable that many of the described mutations involve the loss or gain of an R residue (approximately one third); that is, the loss or gain of a positive charge. Besides the known importance of

the p.R1441 residue, the presence of conserved R residues on the surface of the LRR and WD40 domains, possibly involved in interactions with other proteins [5], can alter the specificity of binding to different proteins.

Finally, it is noteworthy that some analyzed mutations can affect the interaction or relative position of the different domains among them. However, the size and complexity of this protein did not allow us to run a model for the entire protein, or even combinations of two or three domains, with acceptable results. For the same reason, we were unable to analyze the effect that a particular mutation might have on not only the domain where it appears, but also on adjacent domains, which could also prove important in the structure or for dimerization.

The results obtained in this study reveal the usefulness of structural modeling and bioinformatics tools to predict the effects that mutations have on a protein. However, it is important to always take into account that these results are predictions based on mathematics and homology, and can help decide which are more suitable to, for example, generate experimental models in order to further investigate the etiology or treatment for a particular disorder. The determination of the real structural and functional effects of these mutations remains a goal to be reached in order to elucidate their role in the pathogenesis of PD.

## Acknowledgements

The authors wish to thank the different departments at the IBV for their help and assistance in developing this project. The authors also wish to state that have no conflict of interest to declare.

## References

1. Seol W (2010) Biochemical and molecular features of LRRK2 and its pathophysiological roles in Parkinson's disease. *BMB Rep* 43 (4):233-244
2. Paisan-Ruiz C, Jain S, Evans EW, Gilks WP, Simon J, van der Brug M, Lopez de Munain A, Aparicio S, Gil AM, Khan N, Johnson J, Martinez JR, Nicholl D, Carrera IM, Pena AS, de Silva R, Lees A, Marti-Masso JF, Perez-Tur J, Wood NW, Singleton AB (2004) Cloning of the gene containing mutations that cause PARK8-linked Parkinson's disease. *Neuron* 44 (4):595-600
3. Huang Y, Cheung L, Rowe D, Halliday G (2004) Genetic contributions to Parkinson's disease. *Brain Res Brain Res Rev* 46 (1):44-70
4. Santpere G, Ferrer I (2009) LRRK2 and neurodegeneration. *Acta Neuropathol* 117 (3):227-246
5. Mata IF, Wedemeyer WJ, Farrer MJ, Taylor JP, Gallo KA (2006) LRRK2 in Parkinson's disease: protein domains and functional insights. *Trends Neurosci* 29 (5):286-293
6. Marin I (2008) Ancient origin of the Parkinson disease gene LRRK2. *J Mol Evol* 67 (1):41-50
7. Kornev AP, Haste NM, Taylor SS, Eyck LF (2006) Surface comparison of active and inactive protein kinases identifies a conserved activation mechanism. *Proceedings of the National Academy of Sciences of the United States of America* 103 (47):17783-17788. doi:0607656103 [pii]



10.1073/pnas.0607656103

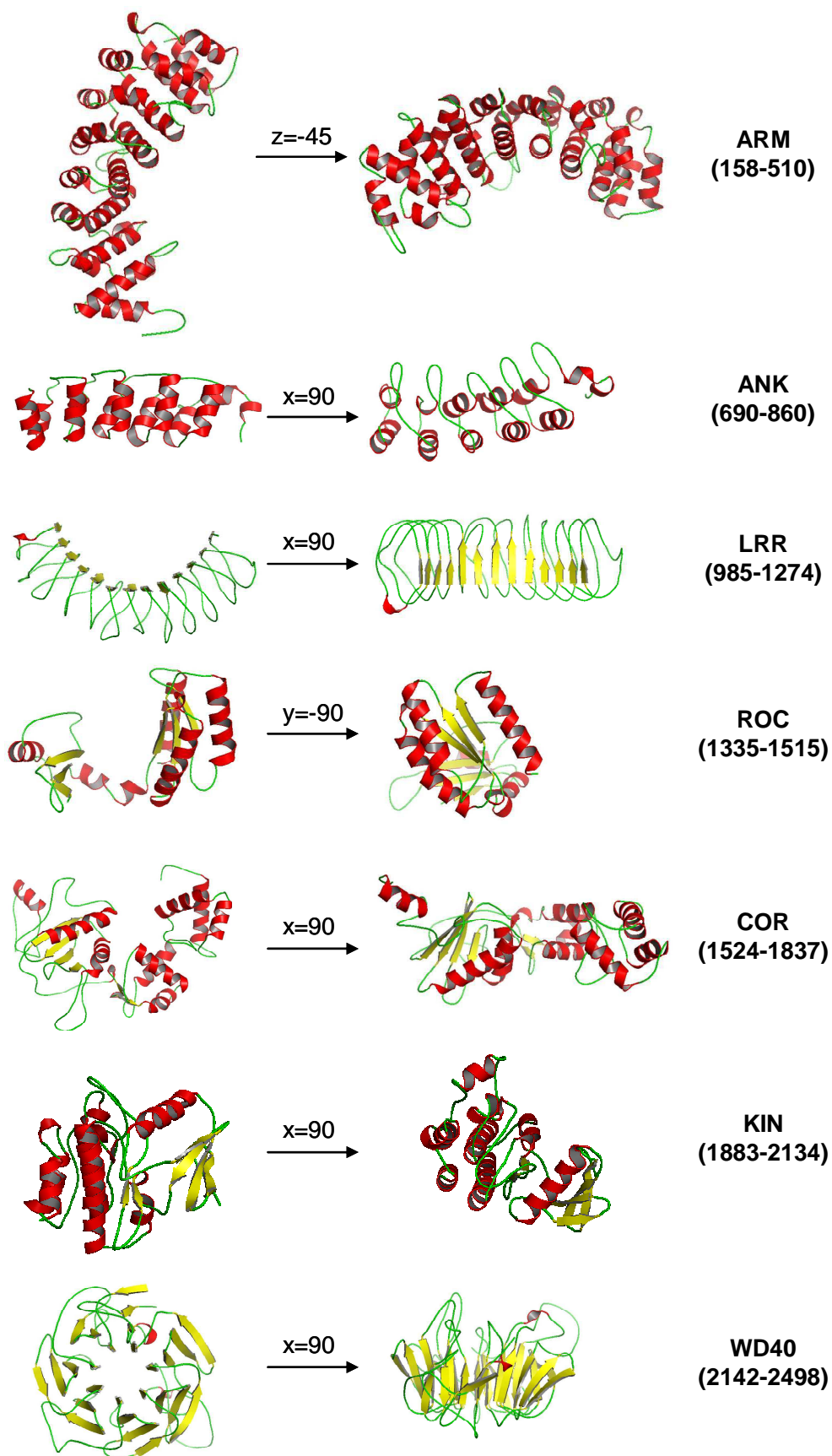
8. Greggio E, Zambrano I, Kaganovich A, Beilina A, Taymans JM, Daniels V, Lewis P, Jain S, Ding J, Syed A, Thomas KJ, Baekelandt V, Cookson MR (2008) The Parkinson disease-associated leucine-rich repeat kinase 2 (LRRK2) is a dimer that undergoes intramolecular autophosphorylation. *J Biol Chem* 283 (24):16906-16914
9. Klein CL, Rovelli G, Springer W, Schall C, Gasser T, Kahle PJ (2009) Homo- and heterodimerization of ROCO kinases: LRRK2 kinase inhibition by the LRRK2 ROCO fragment. *J Neurochem* 111 (3):703-715
10. Gotthardt K, Weyand M, Kortholt A, Van Haastert PJ, Wittinghofer A (2008) Structure of the Roc-COR domain tandem of *C. tepidum*, a prokaryotic homologue of the human LRRK2 Parkinson kinase. *Embo J* 27 (16):2239-2249
11. Deng J, Lewis PA, Greggio E, Sluch E, Beilina A, Cookson MR (2008) Structure of the ROC domain from the Parkinson's disease-associated leucine-rich repeat kinase 2 reveals a dimeric GTPase. *Proceedings of the National Academy of Sciences of the United States of America* 105 (5):1499-1504
12. Lesage S, Brice A (2009) Parkinson's disease: from monogenic forms to genetic susceptibility factors. *Hum Mol Genet* 18 (R1):R48-59
13. Luzon-Toro B, Rubio de la Torre E, Delgado A, Perez-Tur J, Hilfiker S (2007) Mechanistic insight into the dominant mode of the Parkinson's disease-associated G2019S LRRK2 mutation. *Hum Mol Genet* 16 (17):2031-2039
14. Jaleel M, Nichols RJ, Deak M, Campbell DG, Gillardon F, Knebel A, Alessi DR (2007) LRRK2 phosphorylates moesin at threonine-558: characterization of how Parkinson's disease mutants affect kinase activity. *Biochem J* 405 (2):307-317
15. West AB, Moore DJ, Biskup S, Bugayenko A, Smith WW, Ross CA, Dawson VL, Dawson TM (2005) Parkinson's disease-associated mutations in leucine-rich repeat kinase 2 augment kinase activity. *Proceedings of the National Academy of Sciences of the United States of America* 102 (46):16842-16847
16. Webber PJ, West AB (2009) LRRK2 in Parkinson's disease: function in cells and neurodegeneration. *Febs J* 276 (22):6436-6444
17. Kuntal BK, Aparoy P, Reddanna P (2010) EasyModeller: A graphical interface to MODELLER. *BMC research notes* 3:226
18. Cardona F, Sanchez-Mut JV, Dopazo H, Perez-Tur J (2011) Phylogenetic and in silico structural analysis of the Parkinson disease-related kinase PINK1. *Hum Mutat* 32 (4):369-378
19. Capriotti E, Fariselli P, Casadio R (2005) I-Mutant2.0: predicting stability changes upon mutation from the protein sequence or structure. *Nucleic acids research* 33 (Web Server issue):W306-310. doi:33/suppl\_2/W306 [pii]  
10.1093/nar/gki375
20. Cheng J, Randall A, Baldi P (2006) Prediction of protein stability changes for single-site mutations using support vector machines. *Proteins* 62 (4):1125-1132. doi:10.1002/prot.20810
21. Yin S, Ding F, Dokholyan NV (2007) Eris: an automated estimator of protein stability. *Nature methods* 4 (6):466-467. doi:nmeth0607-466 [pii]  
10.1038/nmeth0607-466
22. Teng S, Srivastava AK, Wang L (2010) Sequence feature-based prediction of protein stability changes upon amino acid substitutions. *BMC genomics* 11 Suppl 2:S5. doi:1471-2164-11-S2-S5 [pii]  
10.1186/1471-2164-11-S2-S5
23. Majewski J, Ott J (2003) Amino acid substitutions in the human genome: evolutionary implications of single nucleotide polymorphisms. *Gene* 305 (2):167-173. doi:S0378111903003792 [pii]
24. Kumar P, Henikoff S, Ng PC (2009) Predicting the effects of coding non-synonymous variants on protein function using the SIFT algorithm. *Nature protocols* 4 (7):1073-1081. doi:nprot.2009.86 [pii]  
10.1038/nprot.2009.86
25. Ramensky V, Bork P, Sunyaev S (2002) Human non-synonymous SNPs: server and survey. *Nucleic acids research* 30 (17):3894-3900
26. Ferrer-Costa C, Gelpi JL, Zamakola L, Parraga I, de la Cruz X, Orozco M (2005) PMUT: a web-based tool for the annotation of pathological mutations on proteins. *Bioinformatics (Oxford, England)* 21 (14):3176-3178. doi:bti486 [pii]  
10.1093/bioinformatics/bti486
27. Capriotti E, Calabrese R, Casadio R (2006) Predicting the insurgence of human genetic diseases associated to single point protein mutations with support vector machines and evolutionary information. *Bioinformatics (Oxford, England)* 22 (22):2729-2734. doi:btl423 [pii]  
10.1093/bioinformatics/btl423

28. Zhao Z, Wang Z, Gu Y, Feil R, Hofmann F, Ma L (2009) Regulate axon branching by the cyclic GMP pathway via inhibition of glycogen synthase kinase 3 in dorsal root ganglion sensory neurons. *J Neurosci* 29 (5):1350-1360
29. Lesage S, Condroyer C, Lannuzel A, Lohmann E, Troiano A, Tison F, Damier P, Thobois S, Ouvrard-Hernandez AM, Rivaud-Pechoux S, Brefel-Courbon C, Destee A, Tranchant C, Romana M, Leclere L, Durr A, Brice A (2009) Molecular analyses of the LRRK2 gene in European and North African autosomal dominant Parkinson's disease. *J Med Genet* 46 (7):458-464
30. Mills RD, Mulhern TD, Cheng HC, Culvenor JG (2012) Analysis of LRRK2 accessory repeat domains: prediction of repeat length, number and sites of Parkinson's disease mutations. *Biochemical Society transactions* 40 (5):1086-1089
31. Vancraenenbroeck R, Lobbstaël E, Weeks SD, Strelkov SV, Baekelandt V, Taymans JM, De Maeyer M (2012) Expression, purification and preliminary biochemical and structural characterization of the leucine rich repeat namesake domain of leucine rich repeat kinase 2. *Biochimica et biophysica acta* 1824 (3):450-460
32. Neubert M, Ridder DA, Bargiotas P, Akira S, Schwaninger M (2011) Acute inhibition of TAK1 protects against neuronal death in cerebral ischemia. *Cell Death Differ* 18 (9):1521-1530
33. Garrido JJ, Simon D, Varea O, Wandosell F (2007) GSK3 alpha and GSK3 beta are necessary for axon formation. *FEBS Lett* 581 (8):1579-1586
34. Nichols WC, Elsaesser VE, Pankratz N, Pauculo MW, Marek DK, Halter CA, Rudolph A, Shults CW, Foroud T (2007) LRRK2 mutation analysis in Parkinson disease families with evidence of linkage to PARK8. *Neurology* 69 (18):1737-1744
35. Lorenzo-Betancor O, Samaranch L, Ezquerro M, Tolosa E, Lorenzo E, Irigoyen J, Gaig C, Pastor MA, Soto-Ortolaza AI, Ross OA, Rodriguez-Oroz MC, Valdeoriola F, Marti MJ, Luquin MR, Perez-Tur J, Burguera JA, Obeso JA, Pastor P (2012) LRRK2 haplotype-sharing analysis in Parkinson's disease reveals a novel p.S1761R mutation. *Mov Disord* 27 (1):146-151
36. Gilsbach BK, Ho FY, Vetter IR, van Haastert PJ, Wittinghofer A, Kortholt A (2012) Roco kinase structures give insights into the mechanism of Parkinson disease-related leucine-rich-repeat kinase 2 mutations. *Proceedings of the National Academy of Sciences of the United States of America* 109 (26):10322-10327
37. Gloeckner CJ, Kinkl N, Schumacher A, Braun RJ, O'Neill E, Meitinger T, Kolch W, Prokisch H, Ueffing M (2006) The Parkinson disease causing LRRK2 mutation I2020T is associated with increased kinase activity. *Hum Mol Genet* 15 (2):223-232
38. Nichols RJ, Dzamko N, Morrice NA, Campbell DG, Deak M, Ordureau A, Macartney T, Tong Y, Shen J, Prescott AR, Alessi DR (2010) 14-3-3 binding to LRRK2 is disrupted by multiple Parkinson's disease-associated mutations and regulates cytoplasmic localization. *Biochem J* 430 (3):393-404
39. Reichling LJ, Riddle SM (2009) Leucine-rich repeat kinase 2 mutants I2020T and G2019S exhibit altered kinase inhibitor sensitivity. *Biochem Biophys Res Commun* 384 (2):255-258. doi:S0006-291X(09)00818-3 [pii] 10.1016/j.bbrc.2009.04.098
40. Li X, Moore DJ, Xiong Y, Dawson TM, Dawson VL (2011) Reevaluation of phosphorylation sites in the Parkinson disease-associated leucine-rich repeat kinase 2. *J Biol Chem* 285 (38):29569-29576
41. Mata IF, Wilhoite GJ, Yearout D, Bacon JA, Cornejo-Olivas M, Mazzetti P, Marca V, Ortega O, Acosta O, Cosentino C, Torres L, Medina AC, Perez-Pastene C, Diaz-Grez F, Vilarino-Guell C, Venegas P, Miranda M, Trujillo-Godoy O, Layson L, Avello R, Dieguez E, Raggio V, Micheli F, Perandones C, Alvarez V, Segura-Aguilar J, Farrer MJ, Zabetian CP, Ross OA (2011) Lrrk2 p.Q1111H substitution and Parkinson's disease in Latin America. *Parkinsonism Relat Disord* 17 (8):629-631
42. Zach S, Felk S, Gillardon F (2010) Signal transduction protein array analysis links LRRK2 to Ste20 kinases and PKC zeta that modulate neuronal plasticity. *PLoS ONE* 5 (10): e13191. doi:10.1371/journal.pone.0013191
43. Nolen B, Taylor S, Ghosh G (2004) Regulation of protein kinases; controlling activity through activation segment conformation. *Mol Cell* 15 (5):661-675
44. Anand VS, Reichling LJ, Lipinski K, Stochaj W, Duan W, Kelleher K, Pungaliya P, Brown EL, Reinhart PH, Somberg R, Hirst WD, Riddle SM, Braithwaite SP (2009) Investigation of leucine-rich repeat kinase 2 : enzymological properties and novel assays. *Febs J* 276 (2):466-478
45. Koprivica V, Cho KS, Park JB, Yiu G, Atwal J, Gore B, Kim JA, Lin E, Tessier-Lavigne M, Chen DF, He Z (2005) EGFR activation mediates inhibition of axon regeneration by myelin and chondroitin sulfate proteoglycans. *Science* 310 (5745):106-110
46. Mosse YP, Laudenslager M, Longo L, Cole KA, Wood A, Attiyeh EF, Laquaglia MJ, Sennett R, Lynch JE, Perri P, Laureys G, Speleman F, Kim C, Hou C, Hakonarson H, Torkamani A, Schork NJ, Brodeur GM, Tonini GP, Rappaport E, Devoto M, Maris JM (2008) Identification of ALK as a major familial neuroblastoma predisposition gene. *Nature* 455 (7215):930-935
47. Fiore R, Puschel AW (2003) The function of semaphorins during nervous system development. *Front Biosci* 8:s484-499

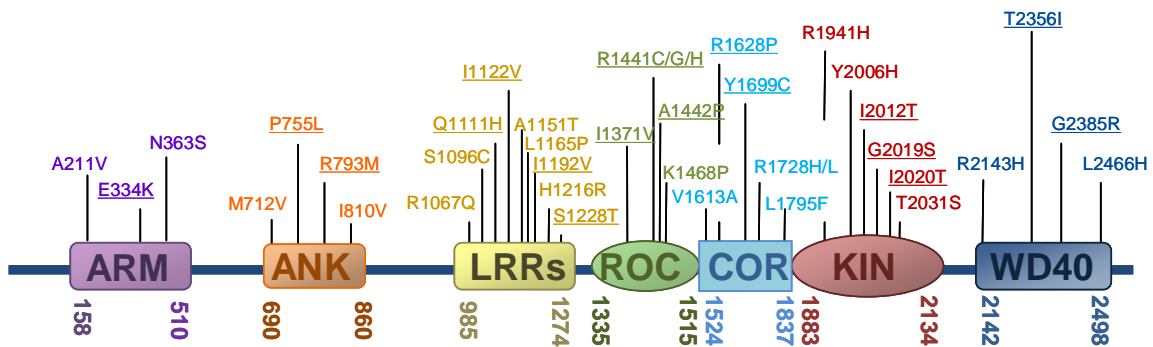
48. Yuan Z, Becker EB, Merlo P, Yamada T, DiBacco S, Konishi Y, Schaefer EM, Bonni A (2008) Activation of FOXO1 by Cdk1 in cycling cells and postmitotic neurons. *Science* 319 (5870):1665-1668
49. Kelly PA, Rahmani Z (2005) DYRK1A enhances the mitogen-activated protein kinase cascade in PC12 cells by forming a complex with Ras, B-Raf, and MEK1. *Mol Biol Cell* 16 (8):3562-3573
50. Kim EJ, Sung JY, Lee HJ, Rhim H, Hasegawa M, Iwatsubo T, Min do S, Kim J, Paik SR, Chung KC (2006) Dyrk1A phosphorylates alpha-synuclein and enhances intracellular inclusion formation. *J Biol Chem* 281 (44):33250-33257
51. Kamikawaji S, Ito G, Iwatsubo T (2009) Identification of the autophosphorylation sites of LRRK2. *Biochemistry* 48 (46):10963-10975
52. Liu M, Kang S, Ray S, Jackson J, Zaitsev AD, Gerber SA, Cuny GD, Glicksman MA (2011) Kinetic, Mechanistic, and Structural Modeling Studies of Truncated Wild-Type Leucine-Rich Repeat Kinase 2 and the G2019S Mutant. *Biochemistry*
53. Liu M, Bender SA, Cuny GD, Sherman W, Glicksman M, Ray SS (2013) Type II Kinase Inhibitors Show an Unexpected Inhibition Mode against Parkinson's Disease-Linked LRRK2 Mutant G2019S. *Biochemistry*
54. Kumari U, Tan EK (2009) LRRK2 in Parkinson's disease: genetic and clinical studies from patients. *Febs J* 276 (22):6455-6463

Table 1. Summary of the effects of PD-related mutations in stability, function or structure of LRRK2 domains. Mutations described as pathogenic are marked by an asterisk. The columns “stability” and “pathogenicity/function” are the result of considering as good the predictions coincident of the 75% of the methods in Tables S1 y S2, respectively. The column “other alteration” is the result of the structural alterations or putative phosphorylation alterations detected.

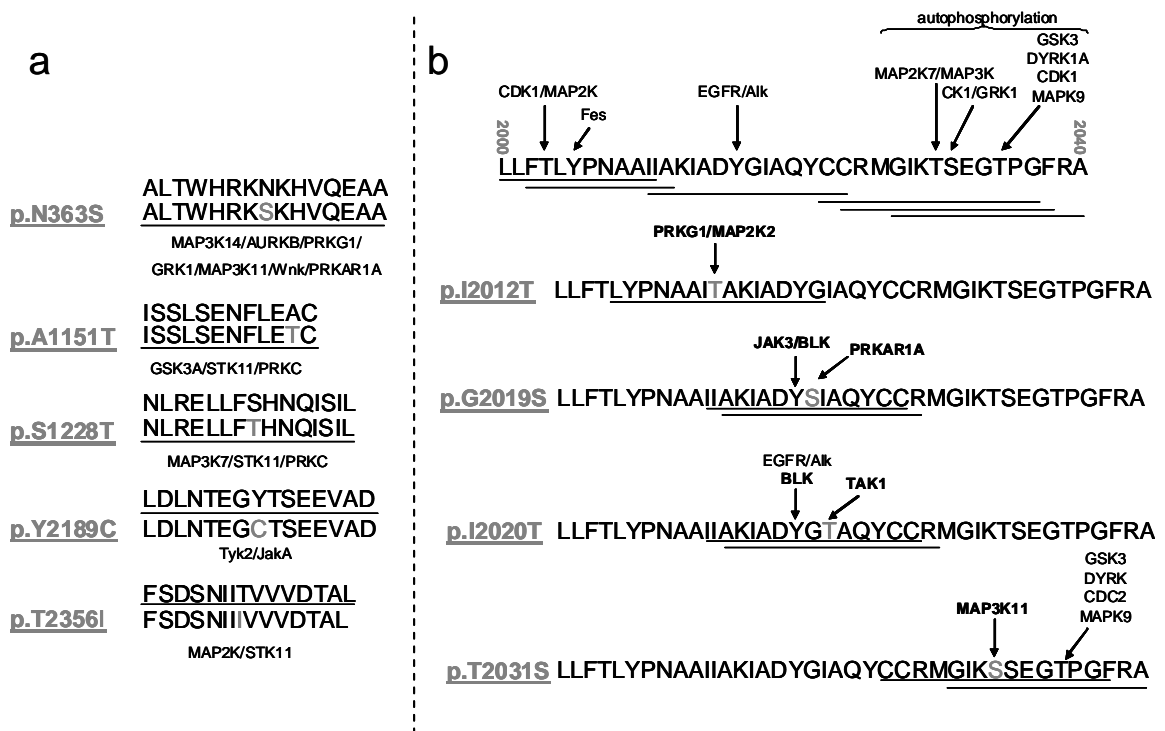
Domain	Mutation	Stability	Path/Func	Other alterations
ARM	p.A211V			
	p.E334K*			charge (-/+); electrostatic surface
	p.N363S	affected		phosphorylation; polar contacts
ANK	p.M712V	affected		electrostatic surface
	p.P755L*		affected	
	p.R793M*			charge (+ <sup>o</sup> ), electrostatic surface
	p.I810V			electrostatic surface
LRR	p.R1067Q	affected		charge (+/-), electrostatic surface
	p.S1096C			
	p.Q1111H*	affected		electrostatic surface
	p.I1122V*	affected		
	p.A1151T			phosphorylation; polar contacts
	p.L1165P	affected	affected	
	p.I1192V*			
	p.H1216R			charge ( <sup>o</sup> /+), electrostatic surface
p.S1228T*	affected		phosphorylation; polar contacts	
ROC	p.I1371V*	affected		
	p.R1441C*	affected	affected	dimerization; charge (+ <sup>o</sup> ), electrostatic surface
	p.R1441G*	affected	affected	dimerization; charge (+ <sup>o</sup> );electrostatic surface
	p.R1441H*	affected		tertiary structure; dimerization; charge (+ <sup>o</sup> );electrostatic surface
	p.A1442P*			
	p.K1468E		affected	charge (+/-), electrostatic surface
	p.R1483Q	affected		tertiary structure; dimerization
p.R1514Q*	affected			
COR	p.V1613A			
	p.R1628P*		affected	secondary structure; charge (+ <sup>o</sup> ), electrostatic surface
	p.Y1699C*		affected	electrostatic surface
	p.R1725Q	affected		charge (+ <sup>o</sup> ), electrostatic surface
	p.R1728H	affected	affected	electrostatic surface
	p.R1728L			charge (+ <sup>o</sup> ), electrostatic surface
	p.L1795F			
p.Q1823K			charge ( <sup>o</sup> /+), electrostatic surface	
KIN	p.R1941H		affected	electrostatic surface
	p.Y2006H	affected	affected	phosphorylation; electrostatic surface
	p.I2012T*	affected	affected	phosphorylation
	p.G2019S*		affected	phosphorylation
	p.I2020T*	affected	affected	phosphorylation
	p.T2031S			phosphorylation
WD40	p.R2143H	affected		electrostatic surface
	p.D2175H			charge (- <sup>o</sup> ), electrostatic surface
	p.Y2189C			phosphorylation
	p.T2356I*	affected		phosphorylation; polar contacts
	p.G2385R*		affected	charge ( <sup>o</sup> /+), electrostatic surface
	p.V2390M	affected		
	p.L2466H	affected	affected	



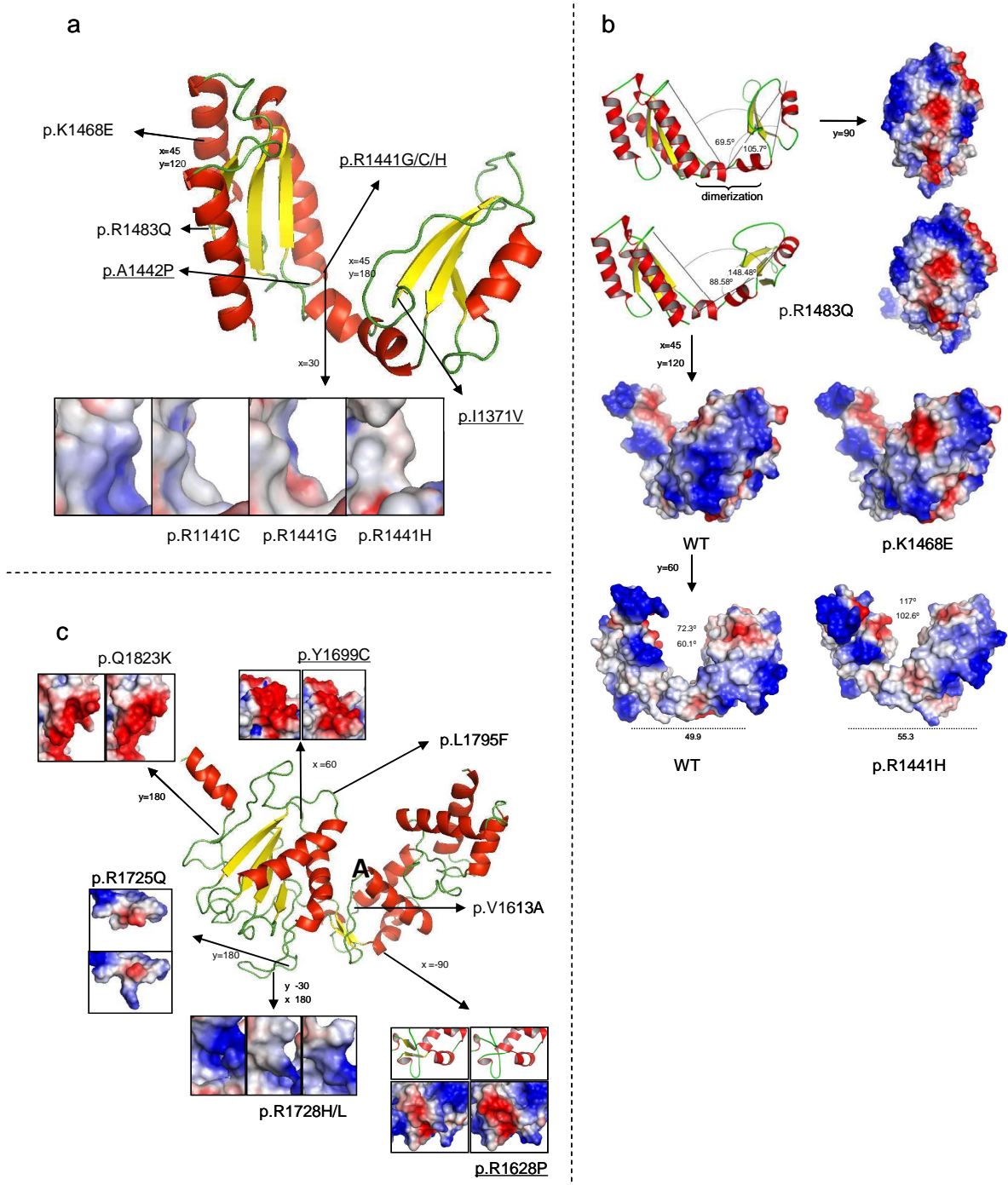
**Fig. 1.** Models and limit definition of the LRRK2 domains. The number of the first and last amino acid of the domain is shown in parentheses. Two views of the domain are shown, turned as indicated over the arrows.



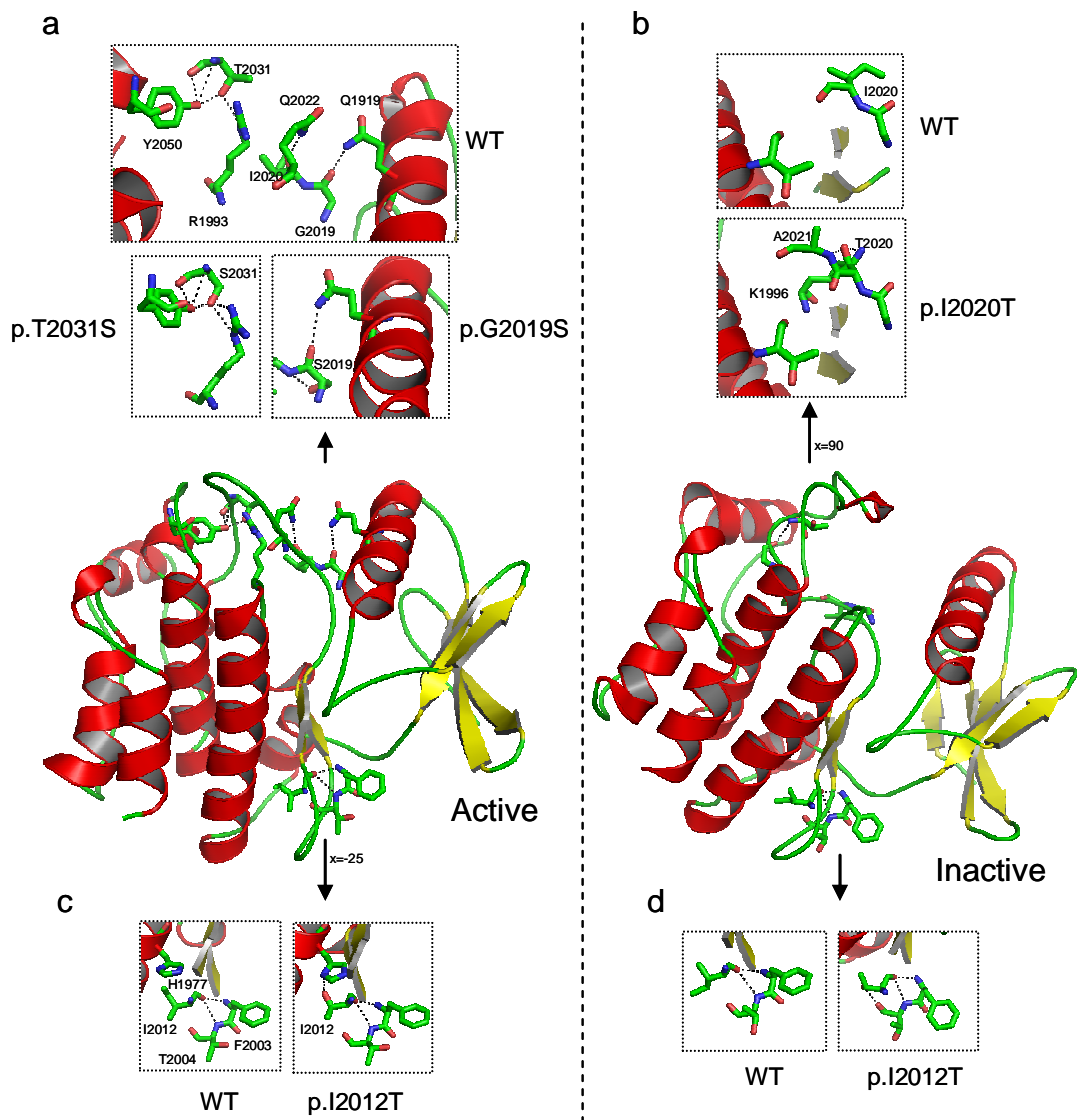
**Fig. 2.** Distribution of the PD-related substitutions in relation to the functional domains of LRRK2. The mutations described as pathogenic are underlined.



**Fig. 3.** Effects of the PD-related mutations on the LRRK2 putative sequences for kinase recognition. a) Alteration of kinase target sequences outside the kinase domain. b) Alteration of the kinase target sequence in the activation loop of the kinase domain. The sequences described as a substrate of auto-phosphorylation are grouped by a bracket. Target sequences are underlined in order of appearance of kinase name (from left to right). The mutations described as pathogenic are underlined. The mutated amino acid is shown in gray. Sometimes the same sequence is a putative target of several kinases.



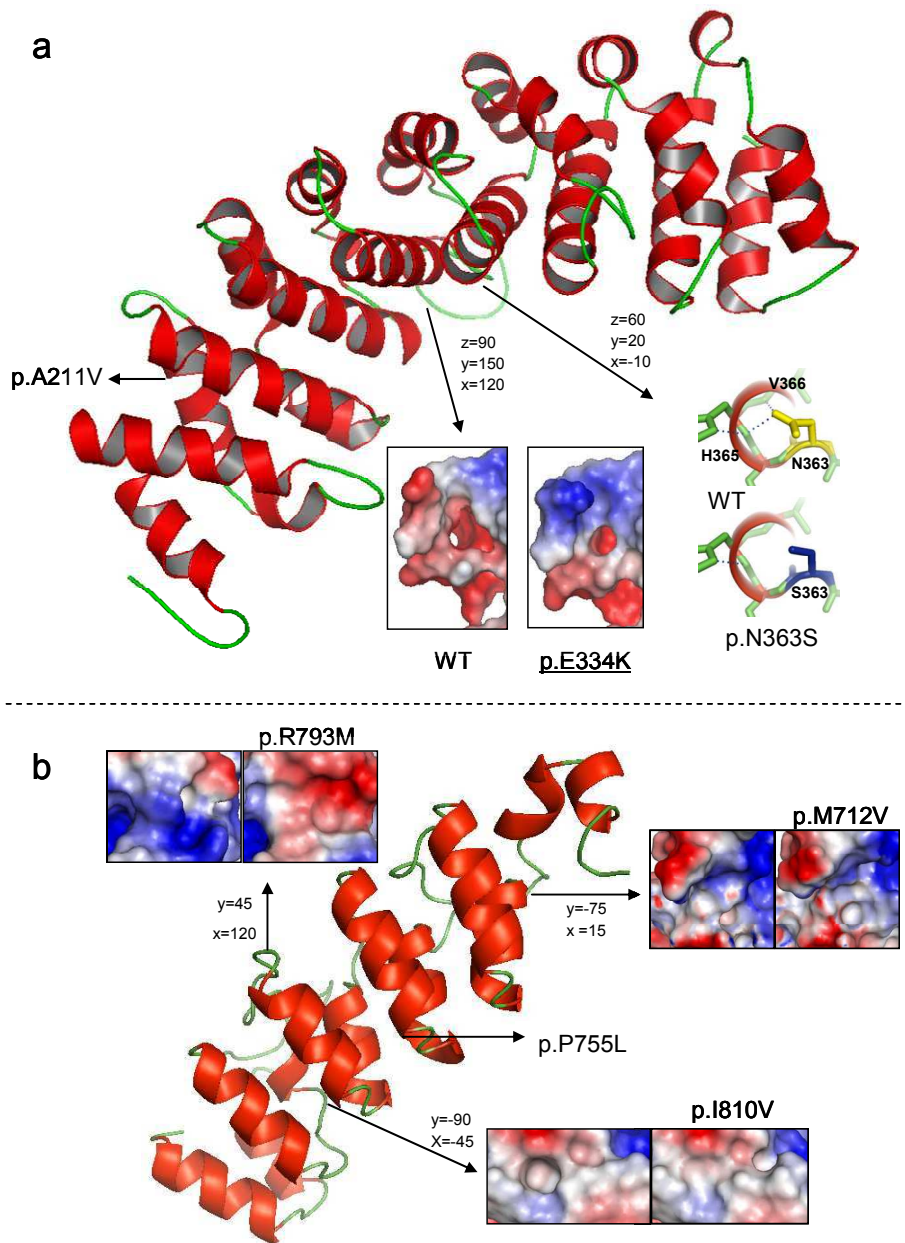
**Fig. 4.** Effects of PD-related mutations in the LRRK2 structural models. a) ROC domain structure and effects of p.R1441 substitutions on the electrostatic surface. b) Effects of p.R1441H, K1468E and p.R1483Q in electrostatic surface angles and distances of the ROC domain. c) The COR structure and effects of p.R1628P, affecting the secondary structure and the electrostatic surface; p.R1728H/L, affecting the electrostatic surface; and p.Y1699C, p.R1725Q and p.Q1823K, affecting the domain surface. The numbers over the arrow indicate the rotation of each axis used to obtain the figures. The numbers over the angles indicate the degrees ( $^\circ$ ), while the numbers under the lines indicate the distance ( $\text{\AA}$ ).



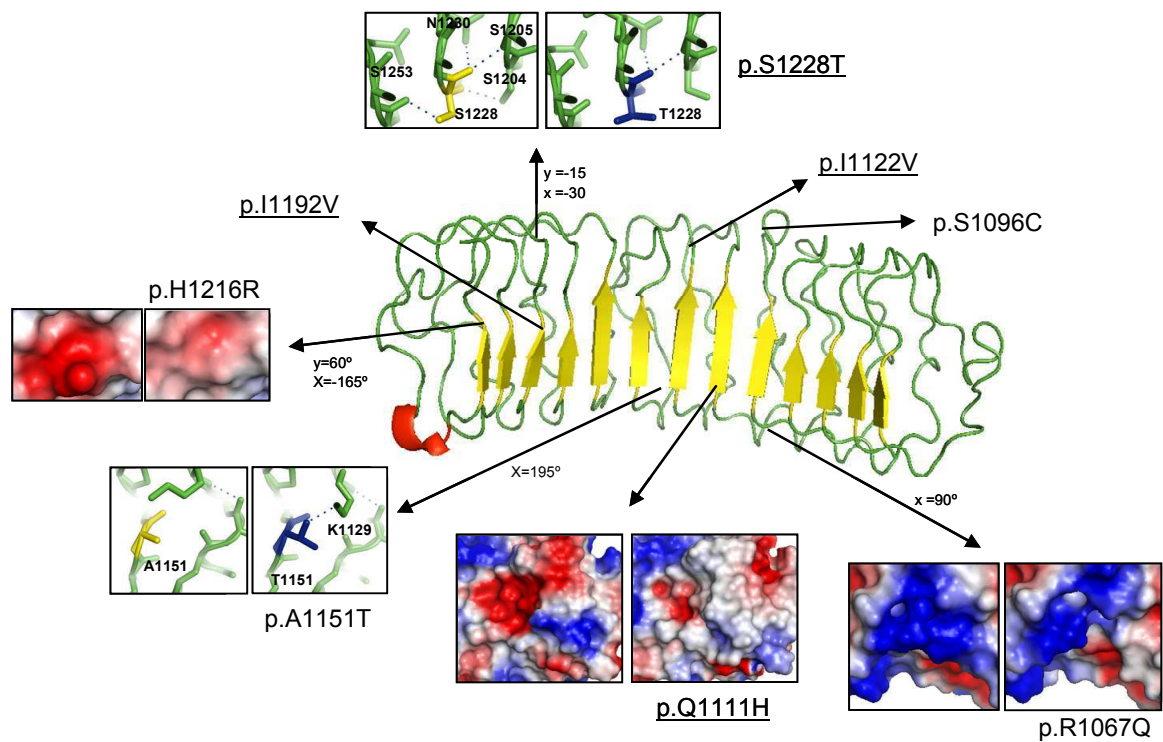
**Fig. 5.** Effects of the PD-related mutations on structural models of the LRRK2 kinase domain. a) Effect of p.G2019S, p.T2031S and p.I2012T on the polar contact distribution of the active form. b) Effect of p.I2020T and p.I2012T on the polar contact distribution of the inactive form. c) Effect of p.I2012T on the polar contact distribution of the active form. d) Effect of p.I2012T on the polar contact distribution of the inactive form. The numbers over the arrow indicate the rotation of each axis used to obtain the figures. Hydrogen bonds are illustrated with dotted lines.



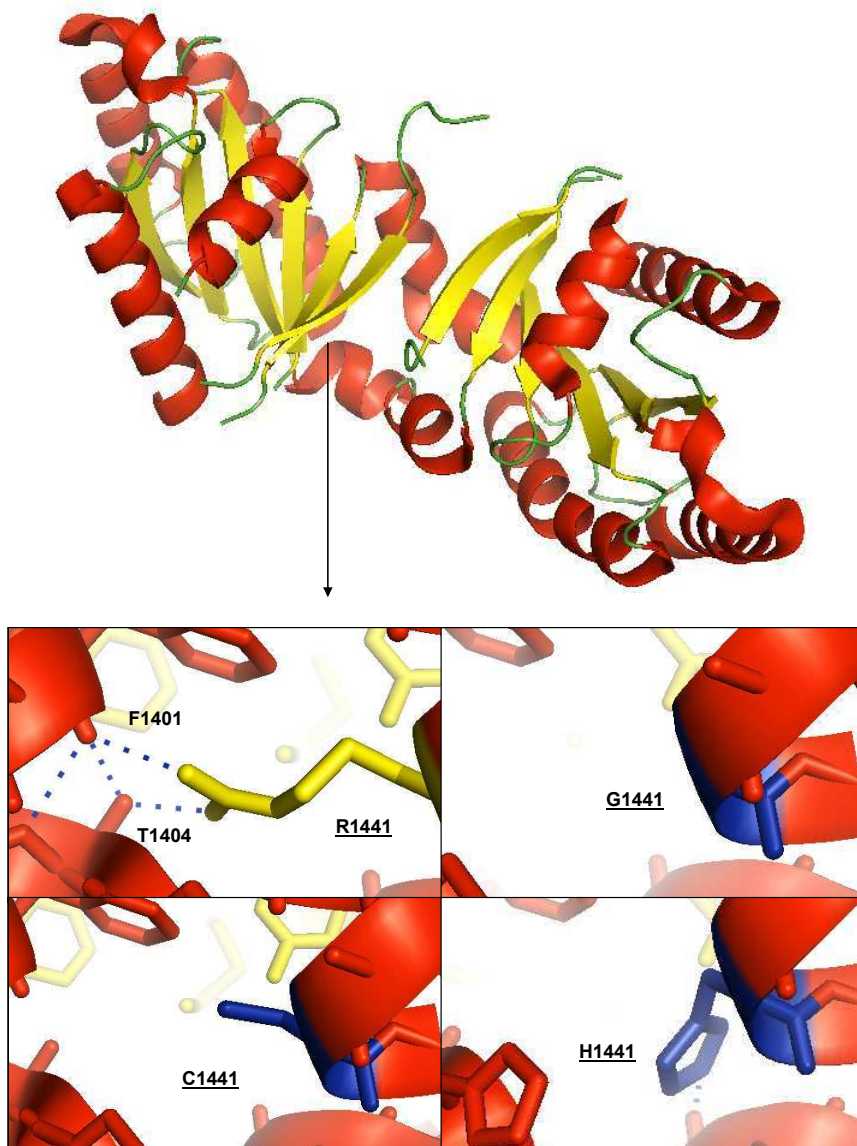




Suppl. Fig. 2. Effects of PD-related mutations in LRRK2 structural models. a) ARM structure and effects of p.E334K and p.N363S in electrostatic surface and polar contacts distribution, respectively. b) ANK domain and effects of p.R793M in electrostatic surface and effects of p.M712V and p.I810V in the surface of the domain. The numbers over the arrow indicate the rotation of each axis used to show the figures. Mutations described as pathogenic are underlined.



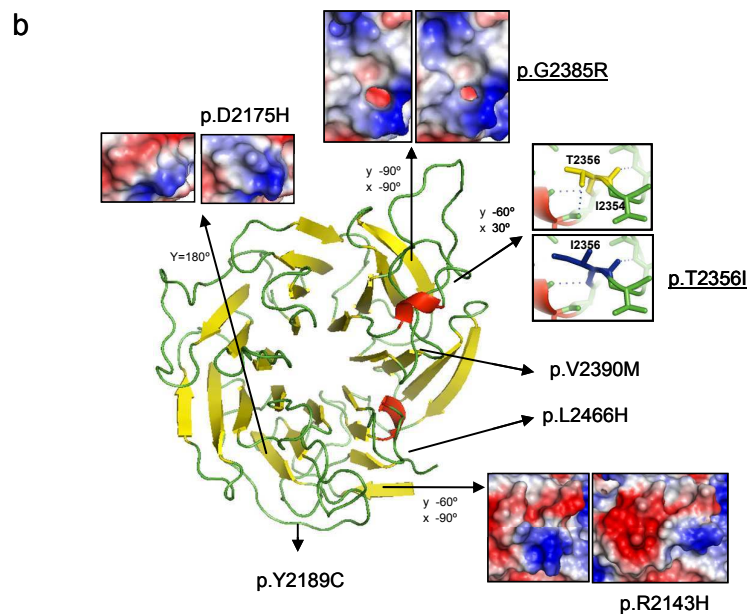
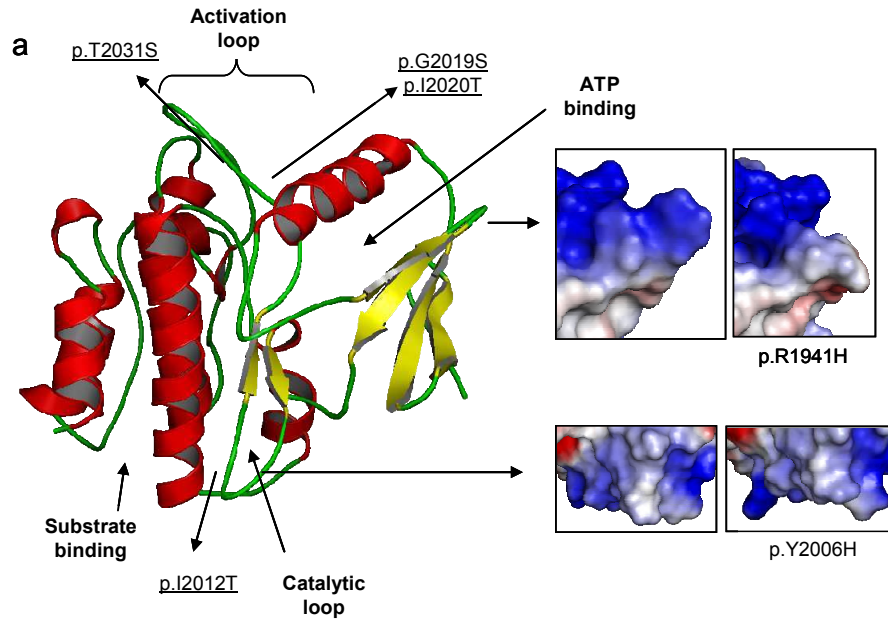
Suppl. Fig. 3. Effects of PD-related mutations in LRR domain structural model. The figure shows the effects p.R1067Q, p.Q1111H and p.H1216R in the electrostatic surface and of p.A1151T and p.S1228T in polar contacts distribution. The numbers over the arrow indicate the rotation of each axis used to show the figures. Mutations described as pathogenic are underlined.



Suppl. Fig. 4. Effects of substitution of R1441 on polar contacts involved in LRRK2 dimers stabilization.

Substitution of R1441 affect to the polar contacts with F1401, T1404 of the other monomer of LRRK2.

This figure was obtained using the structure previously solved.



Suppl. Fig. 5. Effects of PD-related mutations in LRRK2 structural models. a) KIN domain structure and effects of p.R1941H in electrostatic surface and p.Y2006H in the surface of the domain. Subdomains important for the kinase activity are indicated in the figure. b) WD40 structure and effects of p.R2143H and p.T2356I, affecting electrostatic surface and polar contact distribution, respectively; and effects of p.D2175H and p.G2385R in the surface of the domain. The numbers over the arrow indicate the rotation of each axis used to show the figures. Mutations described as pathogenic are underlined.

Supplementary Table 1. Prediction of effects of amino acid substitution in protein stability. Stability of was predicted using four methods, and parameters obtained are showed. Pathogenic mutations are marked by \*. For the elaboration of Table 1, only results coincident in at least 75% of the methods were taken into account (in bold).

Mutation	I-Mutant			MuPro			Eris	MuStab	
	DDG	RI	Stability	CS1	CS2	Stability	DDG	Pred	Conf
p.A211V	-0.27	1		0.17	-0.63		<b>-6,04</b>	Increase	24.6
p.E334K*	-0.66	4	Decrease	-0.48	<b>-0.81</b>	Decrease	<b>-3,51</b>	Increase	31.8
<b>p.N363S</b>	0.22	1		<b>-1</b>	<b>-0.88</b>	<b>Decrease</b>	<b>-2,27</b>	<b>Decrease</b>	<b>93</b>
<b>p.M712V</b>	<b>-1.04</b>	<b>9</b>	<b>Decrease</b>	-0.31	<b>-0.85</b>	Decrease	<b>-7,85</b>	<b>Decrease</b>	<b>85.9</b>
p.P755L*	<b>-0.72</b>	4	Decrease	-0.48	-0.66	Decrease	-1,55	<b>Decrease</b>	<b>85.5</b>
p.R793M*	<b>-0.95</b>	4	Decrease	0.46	0.67	Increase	<b>-5,45</b>	<b>Decrease</b>	<b>79.6</b>
p.I810V	-0.66	8	Decrease	<b>-1</b>	<b>-0.98</b>	<b>Decrease</b>	-1,27	<b>Decrease</b>	<b>90.2</b>
<b>p.R1067Q</b>	<b>-2.16</b>	<b>9</b>	<b>Decrease</b>	-0.69	-0.61	Decrease	<b>1,92</b>	<b>Decrease</b>	<b>90.7</b>
p.S1096C	-0.70	2	Decrease	-0.29	-0.69	Decrease	<b>6,38</b>	<b>Decrease</b>	<b>92.8</b>
<b>p.Q1111H*</b>	<b>-1.09</b>	7	<b>Decrease</b>	<b>-1</b>	<b>-0.72</b>	<b>Decrease</b>	<b>-2,75</b>	<b>Decrease</b>	<b>79.8</b>
<b>p.I1122V*</b>	-0.06	6	Decrease	<b>-1</b>	<b>-0.91</b>	<b>Decrease</b>	<b>5,49</b>	<b>Decrease</b>	<b>93.7</b>
p.A1151T	0.61	7	Decrease	-0.51	-0.58	Decrease	0,35	<b>Decrease</b>	<b>88</b>
<b>p.L1165P</b>	<b>-0.78</b>	7	<b>Decrease</b>	<b>-1</b>		<b>Decrease</b>	<b>2,31</b>	<b>Decrease</b>	<b>90.7</b>
p.I1192V*	-0.03	6	Decrease	<b>-1</b>	<b>-0.89</b>	<b>Decrease</b>	-0,53	<b>Decrease</b>	<b>83.7</b>
p.H1216R	-0.06	4		-0.08	<b>-0.87</b>	Decrease	-0,14	<b>Decrease</b>	<b>82.1</b>
<b>p.S1228T*</b>	<b>-1.08</b>	2	Decrease	-0.27	<b>-0.75</b>	Decrease	<b>5,14</b>	<b>Decrease</b>	<b>93.7</b>
<b>p.I1371V*</b>	<b>-1.48</b>	<b>9</b>	<b>Decrease</b>	<b>-1</b>	<b>-0.91</b>	<b>Decrease</b>	-0,99	<b>Decrease</b>	<b>92.7</b>
<b>p.R1441C*</b>	<b>-2.02</b>	<b>9</b>	<b>Decrease</b>	-0.64	<b>-0.85</b>	<b>Decrease</b>	-0,68	<b>Decrease</b>	<b>82.1</b>
<b>p.R1441C*</b>	<b>-1.30</b>	6	<b>Decrease</b>	<b>-1</b>	<b>-0.99</b>	<b>Decrease</b>	0,08	<b>Decrease</b>	<b>92.9</b>
<b>p.R1441H*</b>	<b>-1.97</b>	<b>9</b>	<b>Decrease</b>	<b>-1</b>	<b>-0.99</b>	<b>Decrease</b>	0,59	<b>Decrease</b>	<b>81</b>
p.A1442P*	-0.34	3	Decrease	-0.06	-0.52	Decrease	1,99	Increase	23.2
p.K1468E	0.18	0	Decrease	-0.43		Decrease	-1,56	<b>Decrease</b>	<b>81.2</b>
<b>p.R1483Q</b>	<b>-1.38</b>	<b>9</b>	<b>Decrease</b>	<b>-0.92</b>		<b>Decrease</b>	-0,04	<b>Decrease</b>	<b>89.6</b>
<b>p.R1514Q*</b>	<b>-1.35</b>	<b>9</b>	<b>Decrease</b>	<b>-1</b>	<b>0.71</b>	Decrease	<b>-2,42</b>	<b>Decrease</b>	<b>92.5</b>
p.V1613A	<b>-0.93</b>	<b>9</b>	<b>Decrease</b>	<b>-1</b>	<b>-0.99</b>	<b>Decrease</b>	<b>-8,82</b>	<b>Decrease</b>	<b>92.3</b>
p.R1628P*	<b>-2.24</b>	7	<b>Decrease</b>	-0.23	-0.55	Decrease	<b>7,99</b>	<b>Decrease</b>	<b>83</b>
p.Y1699C*	<b>-0.79</b>	1	Decrease	<b>-0.78</b>	-0.59	Decrease	<b>-12,5</b>	<b>Decrease</b>	<b>91.8</b>
<b>p.R1725Q</b>	<b>-1.26</b>	<b>8</b>	<b>Decrease</b>	-0.04			<b>-2,42</b>	<b>Decrease</b>	<b>81</b>
<b>p.R1728H</b>	-0.41	<b>8</b>	Decrease	<b>-0.81</b>	<b>-0.89</b>	<b>Decrease</b>	<b>-9,74</b>	<b>Decrease</b>	<b>81.2</b>
p.R1728L	0	5	Decrease	0.67	-0.63	Decrease	-0,37	<b>Decrease</b>	<b>82.7</b>
p.L1795F	-0.01	<b>8</b>	Decrease	<b>-1</b>	<b>-0.98</b>	<b>Decrease</b>	-1,53	<b>Decrease</b>	<b>87.9</b>
p.Q1823K	-0.32	1	Increase	0.03			<b>-7,85</b>	<b>Decrease</b>	<b>80.5</b>
p.R1941H	<b>-1.21</b>	<b>8</b>	<b>Decrease</b>	-0.18	<b>-0.89</b>	Decrease	-1,28	<b>Decrease</b>	<b>78.6</b>
<b>p.Y2006H</b>	-0.39	5	Decrease	<b>-1</b>	<b>-0.99</b>	<b>Decrease</b>	<b>-3,34</b>	<b>Decrease</b>	<b>81.6</b>
<b>p.I2012T*</b>	<b>-1.91</b>	<b>9</b>	<b>Decrease</b>	<b>-1</b>	<b>-0.90</b>	<b>Decrease</b>	<b>-2,78</b>	<b>Decrease</b>	<b>94.1</b>
p.G2019S*	<b>-0.98</b>	<b>8</b>	<b>Decrease</b>	<b>-1</b>	<b>-0.92</b>	<b>Decrease</b>	-1,83	Increase	26.2
<b>p.I2020T*</b>	<b>-2.31</b>	<b>9</b>	<b>Decrease</b>	<b>-1</b>	<b>-0.91</b>	<b>Decrease</b>	<b>2,76</b>	<b>Decrease</b>	<b>94.8</b>
p.T2031S	0.02	4		-0.31	-0.68	Decrease	<b>-4,27</b>	<b>Decrease</b>	<b>93.4</b>
<b>p.R2143H</b>	<b>-1.58</b>	<b>9</b>	<b>Decrease</b>	-0.07	<b>-0.72</b>	Decrease	<b>-2,41</b>	<b>Decrease</b>	<b>81.6</b>
p.D2175H	-0.2	3	Decrease	-0.61			>10	Decrease	28.6
p.Y2189C	<b>-0.72</b>	2	Decrease	-0.51			>10	<b>Decrease</b>	<b>87.3</b>
<b>p.T2356I*</b>	-0.4	1		<b>-1</b>	<b>-0.86</b>	<b>Decrease</b>	<b>-5,23</b>	<b>Decrease</b>	<b>79.8</b>
p.G2385R*	<b>-1.61</b>	<b>8</b>	<b>Decrease</b>	-0.08	-0.5	Decrease	>10	Increase	22.8
<b>p.V2390M</b>	-0.61	<b>8</b>	Decrease	<b>-1</b>		<b>Decrease</b>	<b>6,90</b>	<b>Decrease</b>	<b>92.5</b>
<b>p.L2466H</b>	<b>-2.33</b>	<b>8</b>	<b>Decrease</b>	<b>-0.96</b>	<b>-0.98</b>	<b>Decrease</b>	<b>7,47</b>	<b>Decrease</b>	<b>89.6</b>



Supplementary Table 2. Prediction of pathogenicity of amino acid substitution. Pathogenicity was predicted using five methods, and parameters obtained are showed. Pathogenic mutations are marked by \*. For the elaboration of Table 1, only results coincident in at least 75% of the methods were taken into account (in bold). DI=dissimilarity index

Mutation	DI	SIFT		PolyPhen		Pmut			PhD-SNP	
		Score	Funct	Prediction	PSIC Score	NN	R	Prediction	Effect	R
p.A211V	<b>1.22</b>	0.03	affect		<b>1.61</b>	<b>0.16</b>	<b>6</b>			<b>8</b>
p.E334K*	1.99	0.00	affect		<b>1.26</b>	<b>0.33</b>	<b>3</b>			<b>6</b>
p.N363S	<b>0.37</b>			poss damag	<b>1.51</b>	<b>0.13</b>	7			7
p.M712V	<b>0.81</b>	0.00	affect	poss damag	<b>1.51</b>	<b>0.29</b>	<b>4</b>			<b>6</b>
p.P755L*	1.94	0.01	affect	prob damag	<b>2.55</b>	<b>0.49</b>	<b>0</b>			<b>4</b>
p.R793M*	1.80	0.05	affect	poss damag	<b>1.58</b>	<b>0.34</b>	<b>3</b>			7
p.I810V	<b>0.56</b>	0.01	affect		<b>0.9</b>	<b>0.04</b>	9			7
p.R1067Q	2.03	0.00	affect		<b>1.36</b>	<b>0.15</b>	<b>6</b>			7
p.S1096C	<b>0.55</b>	0.01	affect		<b>1.46</b>	<b>0.30</b>	<b>3</b>		disease	<b>6</b>
p.Q1111H*	<b>0.43</b>	0.03	affect	poss damag	<b>1.7</b>	<b>0.05</b>	9			<b>2</b>
p.I1122V*	<b>0.56</b>				<b>0.92</b>	<b>0.04</b>	9			8
p.A1151T	<b>1.12</b>				<b>1.17</b>	<b>0.09</b>	8			<b>3</b>
p.L1165P	1.94			poss damag	<b>1.93</b>	<b>0.71</b>	<b>4</b>	Patholog	disease	8
p.I1192V*	<b>0.56</b>				<b>0.71</b>	<b>0.04</b>	9			<b>6</b>
p.H1216R	1.79			prob damag	<b>2.30</b>	<b>0.34</b>	<b>3</b>			<b>5</b>
p.S1228T*	<b>0.13</b>	0.03	affect		<b>1.20</b>	<b>0.04</b>	9			7
p.I1371V*	<b>0.56</b>	0.02	affect		<b>0.71</b>	<b>0.05</b>	8			<b>4</b>
p.R1441C*	3.39			poss damag	<b>1.89</b>	<b>0.83</b>	<b>6</b>	Patholog		<b>4</b>
p.R1441G*	1.50			poss damag	<b>1.78</b>	<b>0.63</b>	<b>2</b>	Patholog		<b>0</b>
p.R1441H*	1.79			poss damag	<b>1.69</b>	<b>0.31</b>	<b>3</b>			<b>0</b>
p.A1442P*	<b>0.56</b>	0.04	affect		<b>1.45</b>	<b>0.45</b>	<b>0</b>		disease	<b>2</b>
p.K1468E	1.99	0.00	affect	poss damag struct effect	<b>1.27</b>	<b>0.27</b>	<b>4</b>			<b>6</b>
p.R1483Q	2.03				<b>1.35</b>	<b>0.14</b>	7			7
p.R1514Q*	2.03				<b>0.19</b>	<b>0.14</b>	7			9
p.V1613A	<b>1.22</b>	0.01	affect		<b>1.37</b>	<b>0.23</b>	<b>5</b>			9
p.R1628P*	<b>0.76</b>	0.03	affect	prob damag	<b>2.03</b>	<b>0.75</b>	<b>5</b>	Patholog	disease	8
p.Y1699C*	<b>1.25</b>	0.01	affect	prob damag	<b>2.40</b>	<b>0.77</b>	<b>5</b>	Patholog		<b>0</b>
p.R1725Q	2.03	0.04	affect		<b>1.38</b>	<b>0.14</b>	7			<b>0</b>
p.R1728H	1.79	0.02	affect	poss damag	<b>1.58</b>	<b>0.31</b>	<b>3</b>			<b>5</b>
p.R1728L	<b>0.73</b>			prob damag	<b>2.03</b>	<b>0.63</b>	<b>2</b>	Patholog		8
p.L1795F	<b>0.74</b>	0.02	affect		<b>1.24</b>	<b>0.20</b>	<b>6</b>			7
p.Q1823K	<b>0.69</b>			poss damag	<b>1.52</b>	<b>0.14</b>	7			7
p.R1941H	1.79			poss damag	<b>1.69</b>	<b>0.31</b>	<b>3</b>		disease	7
p.Y2006H	<b>0.93</b>	0.00	affect	poss damag	<b>1.61</b>	<b>0.09</b>	8		disease	<b>1</b>
p.I2012T*	<b>0.49</b>	0.02	affect	poss damag	<b>1.61</b>	<b>0.42</b>	<b>1</b>		disease	<b>1</b>
p.G2019S*	1.53	0.00	affect	poss damag	<b>1.58</b>	<b>0.28</b>	<b>4</b>		disease	7
p.I2020T*	<b>0.49</b>	0.01	affect	poss damag	<b>1.61</b>	<b>0.42</b>	<b>1</b>		disease	8
p.T2031S	<b>0.13</b>	0.00	affect		<b>0.36</b>	<b>0.05</b>	9			7
p.R2143H	1.79				<b>0.65</b>	<b>0.31</b>	<b>3</b>			<b>0</b>
p.D2175H	<b>1.05</b>				<b>1.12</b>	<b>0.61</b>	<b>2</b>	Patholog		
p.Y2189C	<b>1.25</b>				<b>1.94</b>	<b>0.77</b>	<b>5</b>	Patholog		
p.T2356I*	<b>0.49</b>	0.05	affect		<b>0.45</b>	<b>0.42</b>	<b>1</b>			<b>1</b>
p.G2385R*	1.50	0.00	affect		<b>0.10</b>	<b>0.60</b>	<b>1</b>	Patholog		<b>2</b>
p.V2390M	<b>0.81</b>	0.02	affect		<b>1.19</b>	<b>0.13</b>	7			<b>6</b>
p.L2466H	<b>0.63</b>	0.01	affect	poss damag	<b>1.72</b>	<b>0.75</b>	<b>4</b>	Patholog		<b>3</b>