

Structural model of the dimeric Parkinson's protein LRRK2 reveals a compact architecture involving distant interdomain contacts

Giambattista Guaitoli^{a,b,1}, Francesco Raimondi^{c,d,1}, Bernd K. Gilsbach^{a,e,f,1}, Yacob Gómez-Llorente^{g,1}, Egon Deyaert^{h,i,1}, Fabiana Renzi^g, Xianting Li^j, Adam Schaffner^{g,j}, Pravin Kumar Ankush Jagtap^{k,l}, Karsten Boldt^b, Felix von Zweydford^{a,b}, Katja Gotthardt^f, Donald D. Lorimer^m, Zhenyu Yue^j, Alex Burginⁿ, Nebojsa Janjic^o, Michael Sattler^{k,l}, Wim Versées^{h,i}, Marius Ueffing^{b,2}, Iban Ubarretxena-Belandia^{g,2,3}, Arjan Kortholt^{e,2,3}, and Christian Johannes Gloeckner^{a,b,2,3}

^aGerman Center for Neurodegenerative Diseases, 72076 Tübingen, Germany; ^bCenter for Ophthalmology, Institute for Ophthalmic Research, Eberhard Karls University, 72076 Tübingen, Germany; ^cDepartment of Life Sciences, University of Modena and Reggio Emilia, 41125 Modena, Italy; ^dCell Networks, University of Heidelberg, 69120 Heidelberg, Germany; ^eDepartment of Cell Biochemistry, University of Groningen, Groningen 9747 AG, The Netherlands; ^fStructural Biology Group, Max Planck Institute for Molecular Physiology, 44227 Dortmund, Germany; ^gDepartment of Structural and Chemical Biology, Icahn School of Medicine at Mount Sinai, New York, NY 10029; ^hStructural Biology Brussels, Vrije Universiteit Brussel, 1050 Brussels, Belgium; ⁱVlaams Instituut voor Biotechnologie, Structural Biology Research Center, Vrije Universiteit Brussel, 1050 Brussels, Belgium; ^jDepartments of Neurology and Neuroscience, Friedman Brain Institute, Icahn School of Medicine at Mount Sinai, New York, NY 10029; ^kCenter for Integrated Protein Science Munich at Department of Chemistry, Technische Universität München, 85747 Garching, Germany; ^lInstitute of Structural Biology, Helmholtz Zentrum München, 85764 Munich, Germany; ^mBeryllium Discovery Corporation, Bainbridge Island, WA 98110; ⁿBroad Institute, Cambridge, MA 02142; and ^oSomaLogic, Boulder, CO 80301

Edited by Quyen Q. Hoang, Indiana University School of Medicine, Indianapolis, IN, and accepted by Editorial Board Member Gregory A. Petsko May 25, 2016 (received for review December 24, 2015)

Leucine-rich repeat kinase 2 (LRRK2) is a large, multidomain protein containing two catalytic domains: a Ras of complex proteins (Roc) G-domain and a kinase domain. Mutations associated with familial and sporadic Parkinson's disease (PD) have been identified in both catalytic domains, as well as in several of its multiple putative regulatory domains. Several of these mutations have been linked to increased kinase activity. Despite the role of LRRK2 in the pathogenesis of PD, little is known about its overall architecture and how PD-linked mutations alter its function and enzymatic activities. Here, we have modeled the 3D structure of dimeric, full-length LRRK2 by combining domain-based homology models with multiple experimental constraints provided by chemical cross-linking combined with mass spectrometry, negative-stain EM, and small-angle X-ray scattering. Our model reveals dimeric LRRK2 has a compact overall architecture with a tight, multidomain organization. Close contacts between the N-terminal ankyrin and C-terminal WD40 domains, and their proximity—together with the LRR domain—to the kinase domain suggest an intramolecular mechanism for LRRK2 kinase activity regulation. Overall, our studies provide, to our knowledge, the first structural framework for understanding the role of the different domains of full-length LRRK2 in the pathogenesis of PD.

LRRK2 | Parkinson's disease | structural modeling | EM | CL-MS

Parkinson's disease (PD) is the second most common age-related neurodegenerative disease, and it is clinically characterized by movement impairment, bradykinesia, rigidity, and resting tremors. Pathologically, it is characterized by the progressive loss of dopaminergic neurons in the substantia nigra and the formation of Lewy bodies (1). Although the majority of cases are sporadic, mutations in the leucine-rich repeat kinase 2 (LRRK2) gene [*PARK8*; Online Mendelian Inheritance in Man (OMIM) no. 609007] have been unequivocally linked to late-onset autosomal dominant PD (2, 3).

Mutations in the *LRRK2* gene are found in 5–15% of families with autosomal dominant PD, making them the most common cause of Mendelian PD identified so far (4). Unfortunately, the biological function of LRRK2 and the mechanism by which it contributes to PD pathogenesis are not well understood. LRRK2 is a multidomain, 286-kDa protein exhibiting both GTPase and kinase activities (5–7) that belongs to the recently identified Roco protein family of G proteins. Members of this family have a Ras of complex proteins (Roc) G-domain and an adjacent conserved C-terminal of Roc (COR) dimerization domain in

common (8, 9). Besides the enzymatic core region, LRRK2 contains four predicted solenoid domains commonly involved in protein–protein interactions (10). These domains include the N-terminal ankyrin, armadillo, and namesake leucine-rich repeat (LRR) domains, along with a C-terminal WD40 domain. Various PD-associated mutations in LRRK2 have been shown to augment kinase activity (6, 7, 11, 12) and increase the phosphorylation of Rab proteins, recently identified as likely physiological substrates (12). In addition, LRRK2 mutations within the Roc domain have been associated with decreased LRRK2 GTPase activity (13, 14).

Significance

Leucine-rich repeat kinase 2 (LRRK2) represents a promising drug target for treatment and prevention of Parkinson's disease (PD), because mutations in LRRK2 are the most common cause of Mendelian forms of the disease. PD-associated LRRK2 variants show decreased GTPase and increased kinase activity. By integrating multiple experimental inputs provided by chemical cross-linking, small-angle X-ray scattering, and a negative-stain EM map, we present, to our knowledge, the first structural model of the full-length LRRK2 dimer. The model reveals a compact folding of the LRRK2 dimer with multiple domain–domain interactions that might be involved in the regulation of LRRK2 enzymatic properties.

Author contributions: G.G., F. Raimondi, B.K.G., Y.G.-L., E.D., F. Renzi, Z.Y., A.B., M.S., W.V., M.U., I.U.-B., A.K., and C.J.G. designed research; G.G., F. Raimondi, B.K.G., Y.G.-L., E.D., F. Renzi, X.L., A.S., P.K.A.J., F.v.Z., K.G., D.D.L., and W.V. performed research; N.J. contributed new reagents/analytic tools; G.G., F. Raimondi, B.K.G., Y.G.-L., E.D., F. Renzi, X.L., A.S., P.K.A.J., K.B., F.v.Z., K.G., D.D.L., A.B., M.S., W.V., I.U.-B., A.K., and C.J.G. analyzed data; and G.G., F. Raimondi, B.K.G., E.D., A.B., M.S., W.V., M.U., I.U.-B., A.K., and C.J.G. wrote the paper.

The authors declare no conflict of interest.

This article is a PNAS Direct Submission. Q.Q.H. is a guest editor invited by the Editorial Board.

Freely available online through the PNAS open access option.

Data deposition: The structure of the leucine-rich repeat domain of the *Chlorobium tepidum* Roco protein has been deposited in the Protein Data Bank, www.pdb.org (PDB ID code 5IL7).

¹G.G., F. Raimondi, B.K.G., Y.G.-L., and E.D. contributed equally to this work.

²M.U., I.U.-B., A.K., and C.J.G. contributed equally to this work.

³To whom correspondence may be addressed. Email: iban.ubarretxena@mssm.edu, a.kortholt@rug.nl, or johannes.gloeckner@dzne.de.

This article contains supporting information online at www.pnas.org/lookup/suppl/doi:10.1073/pnas.1523708113/-DCSupplemental.

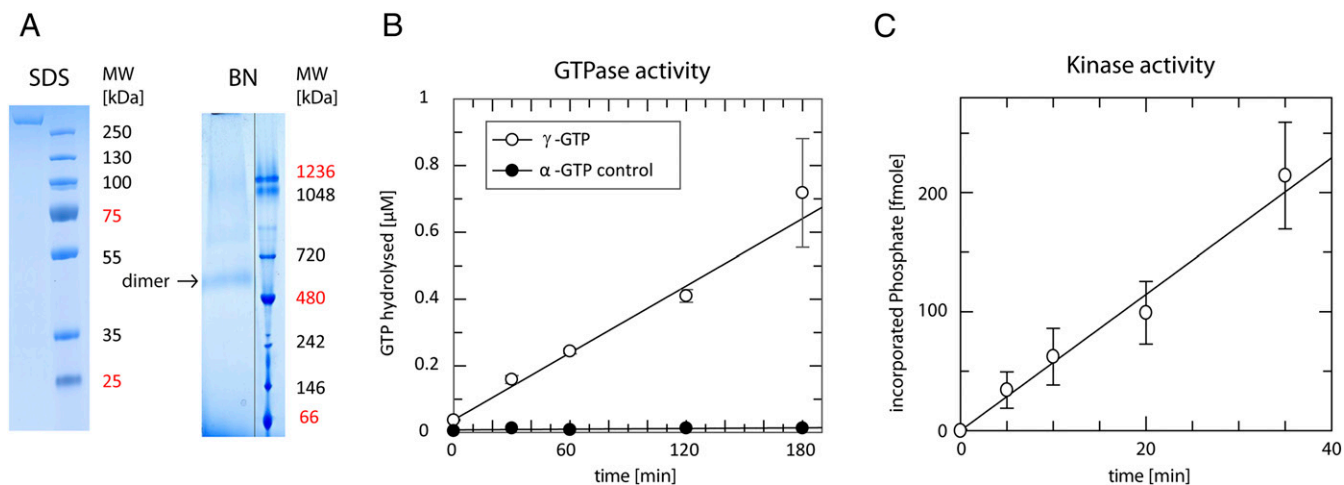


Fig. 1. (A, Left) SDS/PAGE to demonstrate the purity of LRRK2. Five percent of a typical batch (corresponding to $\sim 600 \text{ cm}^2$ of HEK293T culture) was separated on a NuPage 4–12% gradient gel and stained with colloidal Coomassie. (Right) BN-PAGE (4–12% NativePage) analysis reveals that LRRK2 is predominantly dimeric after purification (colloidal Coomassie stain). MW, molecular weight. (B) Purified LRRK2 exhibits GTPase activity. The assay was performed with $0.1 \mu\text{M}$ LRRK2 and either $10 \mu\text{M}$ γ -GTP or $10 \mu\text{M}$ α -GTP (specificity control) at 30°C . The observed GTP hydrolysis rate is 0.03 min^{-1} (± 0.0014) ($n = 3$ biological replicates with 2 technical replicates for each). (C) Kinase activity toward the synthetic LRRK2 substrate LRRKtide in the presence of GDP ($n = 3$ biological replicates with 2 technical replicates for each).

These results have prompted the development of highly potent and specific kinase inhibitors with K_i values in the low nanomolar range (15). However, the value of this therapeutic approach has been challenged by the observation that treatment with LRRK2 kinase inhibitors results in significant lung toxicity in primates (16). In addition, LRRK2 knockout rodent models display severe kidney phenotypes (17, 18). These results suggest that effective therapeutic intervention may require more subtle allosteric modulation of LRRK2 kinase activity, and highlight the critical importance of understanding the regulation and function of this enzyme. Here, we present a structural model of full-length, dimeric LRRK2 derived from an integrative structural modeling approach based on multiple experimental constraints provided by chemical cross-linking small-angle X-ray scattering (SAXS), as well as a 3D map of LRRK2 generated by negative-stain EM. The resulting model of LRRK2 reveals a compact architecture with close contacts of the C-terminal helix of the WD40 domain with the N-terminal ankyrin domain. These two domains, as well as the LRR domain, are in close proximity to the kinase domain, suggesting that they could play a potential role in modulating LRRK2 kinase activity.

Results

Purified LRRK2 Is an Active Dimer in Solution. To improve the yield and purity of LRRK2 necessary for structural studies, the expression and purification procedure for recombinant full-length LRRK2 was optimized from previously published protocols (19). Following transient transfection of HEK293T cells, N-terminal Strep/FLAG (SF)-tagged LRRK2 was purified via the tandem Strep-tag II moiety. The optimized protocol allowed a yield of ~ 100 – $150 \mu\text{g}$ of LRRK2 at a purity $>98\%$ from confluent cells covering $\sim 600 \text{ cm}^2$ (4×14 -cm dishes, 5 – 6×10^8 cells). The purity of the material is shown in Fig. 1A. In addition, the secondary structure of the protein was assessed by circular dichroism (CD) measurements in the far-UV region (250–190 nm). The CD spectra shown in Fig. S1 indicate that the protein is folded with a large fraction (52%) of α -helical secondary structure and just 9% of β -strand content. These data are in good agreement with estimations based on the final structural models of the LRRK2 dimer predicting 34–36% α -helical and 12–14% β -strand content. The purified LRRK2 was analyzed by dynamic light scattering (DLS) and Blue Native polyacrylamide gel electrophoresis (BN-PAGE) to evaluate purity and oligomeric state. DLS analysis revealed that LRRK2 is predominantly

dimeric in solution with an experimentally determined molecular mass of 581 kDa, which is in excellent agreement with the calculated mass of the SF-tagged LRRK2 dimer (native protein, 572 kDa; SF-tag fusion protein, 584 kDa) (Fig. S2). The BN-PAGE shows a prominent band with an apparent molecular mass corresponding to dimeric LRRK2 (Fig. 1A). Consistent with previous reports showing that LRRK2 is able to homodimerize (6, 20–23), these findings demonstrate that LRRK2 is a dimer in solution with no significant presence of higher oligomeric species under these conditions. In addition, no changes in the oligomeric state were visible when LRRK2 was purified in the presence of different G-nucleotides (Fig. S2), and biochemical analysis confirmed that the purified LRRK2 displays both GTPase and kinase activities (Fig. 1B and C).

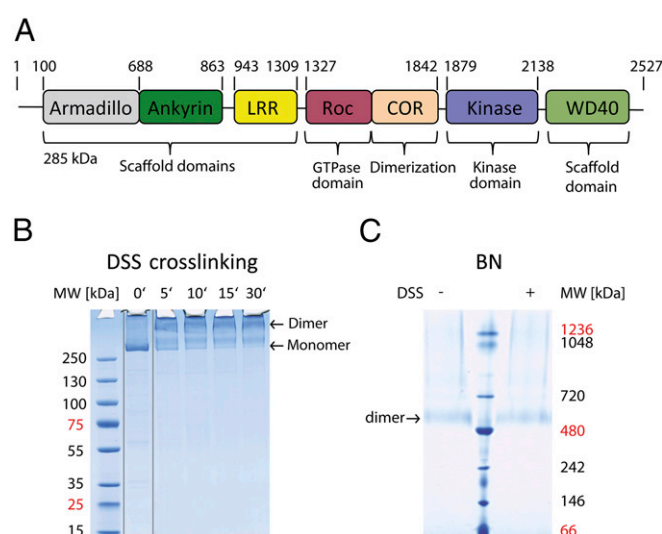


Fig. 2. (A) Schematic overview of the LRRK2 domain structure. (B) Time course for DSS cross-linking at room temperature (RT). Samples were separated on a NuPage 4–12% gradient gel and stained with colloidal Coomassie. (C) BN-PAGE analysis before and after 30 min of DSS cross-linking at RT (4–12% NativePage; colloidal Coomassie stain).

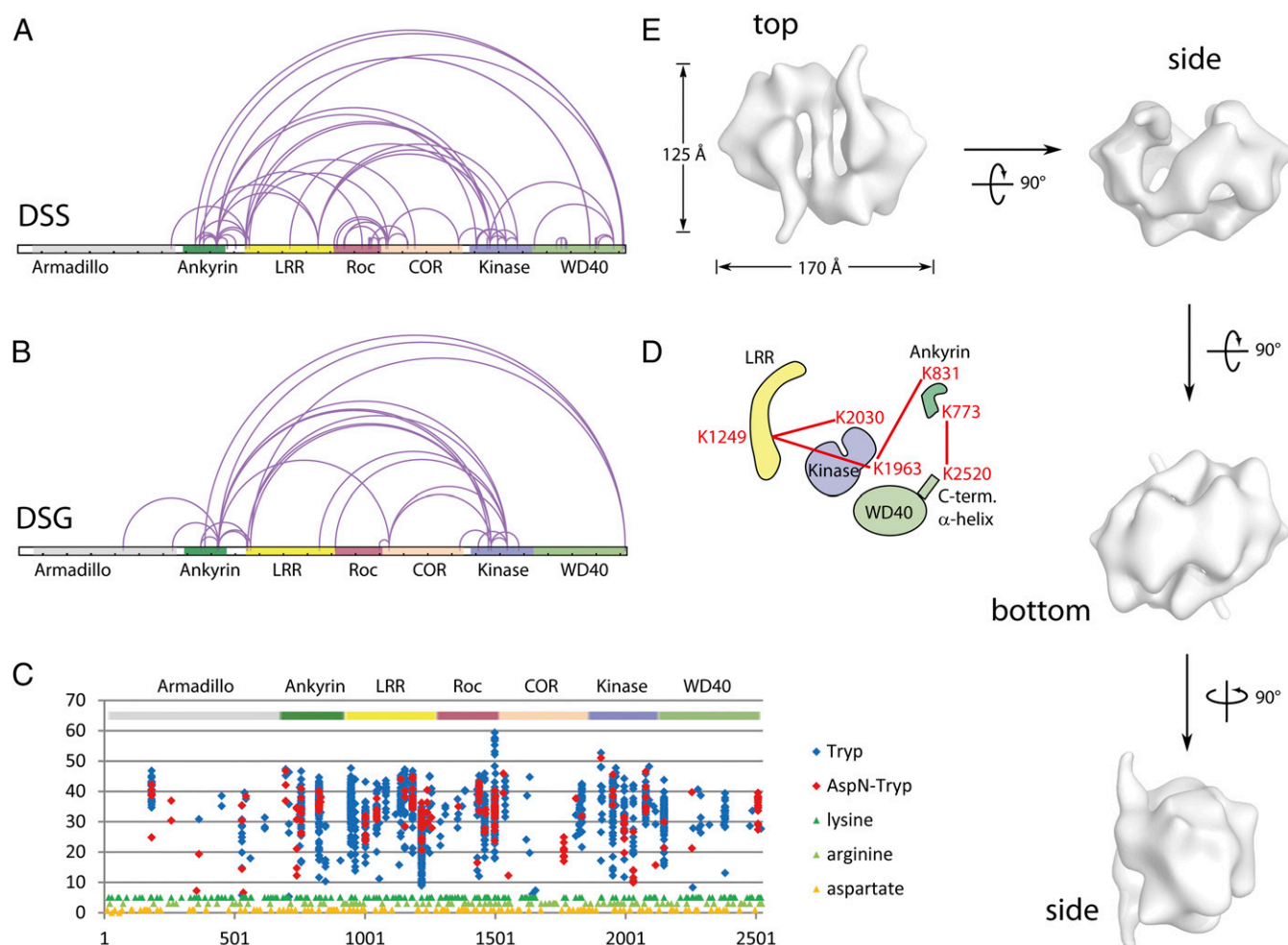


Fig. 3. (A and B) DSS and DSG cross-links mapped onto the LRRK2 sequence with domains indicated. The plots were generated with xiNET (65). (C) DSS-derivatized lysine residues (monolinks) plotted over the residue number in LRRK2 for two different proteolytic enzyme combinations: trypsin-only (Tryp) and AspN-trypsin (ApsN-Tryp). The distribution of theoretical cleavage sites (trypsin: lysine, arginine; AspN: aspartate) is represented by triangles at the bottom of the plot. (D) Representative cross-links between N- and C-terminal domains (N-terminal: ankyrin, LRR; C-terminal: WD40) and the kinase domain demonstrating a compact folding of LRRK2. (E) Three-dimensional reconstruction of LRRK2. Surface views of the symmetrical EM map of LRRK2 (rendered at a threshold corresponding to a molecular mass of 600 kDa).

Chemical Cross-Linking Coupled with Mass Spectrometry Reveals Intradomain Contacts Consistent with a Compact Architecture of LRRK2.

Intramolecular chemical cross-linking is a powerful tool to analyze domain-domain interactions in complex proteins (24). We conducted a systematic cross-linking approach, combined with proteolytic cleavage and the subsequent identification of the cross-linked peptides by mass spectrometry (MS) to provide information about domain interactions in full-length LRRK2. We used the amine-specific/lysine-selective *N*-hydroxysuccinimide (NHS) esters disuccinimidyl suberate (DSS) and disuccinimidyl glutarate (DSG) with spacer arm lengths of 11.4 and 7.7 Å, respectively. As a first step, the conditions for selective cross-linking were optimized to favor cross-links over saturation of the accessible lysine side chains with monolinks of each cross-linking reagent. For the ensuing datasets, cross-linker/protein molar ratios of 50:1 and 25:1 were used, although the ratio of 25:1 gave the best results in terms of the formation of cross-links over monolinks; Fig. 2B shows a time course for DSS cross-linking. Using the optimized conditions, efficient cross-linking of the LRRK2 dimer without the appearance of higher aggregates was observed by BN-PAGE analysis (Fig. 2C). Notably, the cross-linking had no impact on the migration or intensity of LRRK2. After chemical cross-linking and before MS, the protein was subjected to proteolysis by different proteolytic enzymes

and their combinations, including AspN and trypsin, as well as LysC and GluC. The peptides were subsequently separated by size exclusion chromatography (SEC) to obtain fractions enriched in cross-linked peptide species. An overview of the chromatographic separation and fractions is shown in Fig. S3 A and B. Extracted tandem mass spectra (MS/MS) of the SEC fractions analyzed with xQuest/xProphet (25) demonstrate good enrichment of cross-linked species in the early fractions, in agreement with published observations (26). Based on these initial tests, we focused our subsequent analyses on fractions A2 to A5.

In its primary sequence, LRRK2 exhibits ~7% frequency of lysine residues. Lysine residues that were found to be modified by the reagent but without forming cross-links (monolinks) were also taken into account as an indicator of solvent accessibility. The frequencies represented by individual ID scores of monolink peptides were plotted versus LRRK2 residue number (Fig. 3C). The plot shows various hotspots of modification, indicating that these residues are highly accessible, and thus likely exposed to the solvent. Other regions of the protein, such as the COR domain, display reduced accessibility, in agreement with previous work showing that the COR domain of Roco proteins is involved in dimerization and is also likely surrounded by other domains in the full-length protein (27). Additional lysine residues with high accessibility to the cross-linking

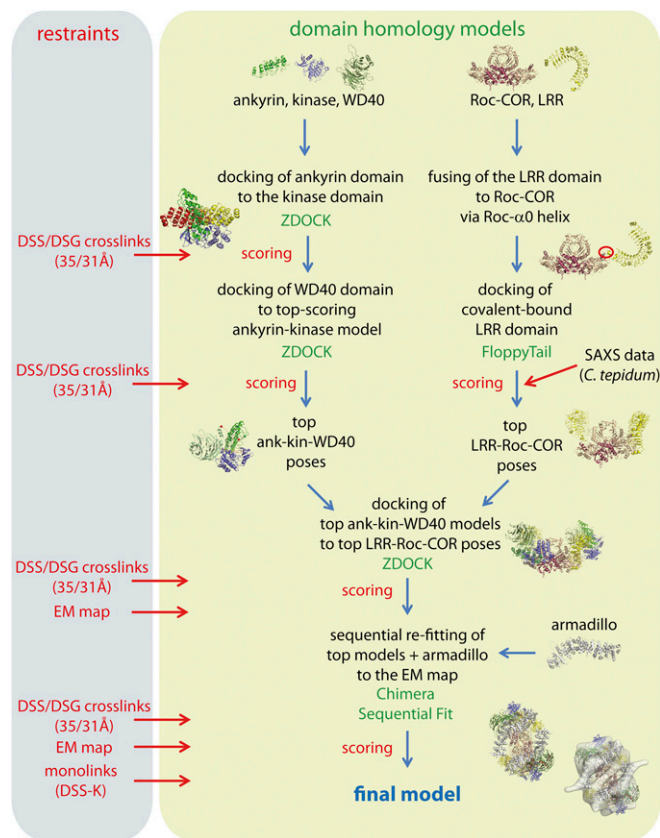


Fig. 4. Integrative modeling workflow.

reagent were observed in the neighborhood of known phosphorylation sites (19, 28). In contrast, the N terminus exhibits a low coverage with monolink peptides. This observation may be explained by the lower frequency of lysine residues in this part of the sequence, which leads to rather large peptide fragments that can remain undetected in our assay.

For the DSS cross-linking, the combined final dataset contains 64 different cross-links with a cut-off ID score of 28, a delta score (ΔS) < 0.95, and a false discovery rate (FDR) < 0.05 (Dataset S1). For DSG cross-links, 30 different cross-links matched these criteria (Dataset S2). A schematic overview of the found cross-links is given in Fig. 3A and B. Interestingly, the dataset revealed several long-range interdomain cross-links (by primary sequence), such as between the ankyrin domain and the outer C-terminal helix of the WD40 domain (i.e., between residues K773 and K2520). In addition, several contacts between the N-terminal ankyrin, as well as LRRs and the kinase domain (e.g., between residues K831 and K1963, as well as residues K1249 and K2030), were found, suggesting that these domains are in close proximity to each other (Fig. 3D).

The identification of various long-range interdomain cross-links and the interactions of the LRRK2 N-terminal part with the C terminus are consistent with a compact overall architecture of the dimeric LRRK2 holoenzyme.

Architecture of Dimeric LRRK2 Revealed by Negative-Stain Electron Microscopy. We analyzed LRRK2 by EM to generate a low-resolution 3D density map of the dimeric LRRK2 protein as input for an integrative structural modeling approach. We first imaged LRRK2 purified from brain tissue of BAC transgenic mice expressing wild-type 3 \times FLAG-tagged LRRK2 (29) embedded in negative stain to ensure sufficient contrast (30). Fig. S4B displays a representative field view of negatively stained LRRK2 showing

abundant elongated particles with a few protein aggregates. To ascertain the identity of the recorded particles and the oligomeric state they represent, we used gold markers in combination with antibody labeling (31). LRRK2 was first labeled with either monoclonal anti-FLAG or polyclonal anti-LRR antibodies, followed by the addition of 5 nm of gold-conjugated anti-IgG secondary antibody, and revealed particles marked with two gold moieties when imaged (Fig. S4C). Due to the fact that immunogold labeling relies on the concerted action of two antibodies, and is therefore relatively inefficient, a large proportion of particles were either labeled with just one nanogold moiety or unlabeled. The particles labeled with two nanogold moieties corresponded to LRRK2 dimers, in agreement with our DLS data, and had approximate dimensions of 125 \times 170 Å. We note that Civiero et al. (22) recently used a comparable labeling and imaging strategy on LRRK2 purified from HEK293 cells with a similar outcome.

Initially, we manually selected LRRK2 particles whose shape and dimensions resembled the dimeric particles identified by immunogold labeling, and obtained a dataset of \sim 2,600 particles that was subjected to reference-free alignment, classification, and 3D reconstruction by the cross-common lines method using EMAN (32). After the first refinement cycle, a clear twofold symmetry axis emerged, even though no geometrical constraints were imposed. A subsequent refinement imposing a twofold symmetry yielded an initial model at a resolution of \sim 33 Å. To improve the resolution of the map, \sim 127,000 single images of LRRK2 particles from an independent purification were automatically selected using projection matching with EMAN2 (33) and then subjected to iterative rounds of reference-free classification and heterogeneity analysis using 2D maximum likelihood analysis in Fourier space (2DMLF) (34). The dataset (51,000 particle images) corresponding to LRRK2 dimers was used for further refinement of the initial model. The reprojections of the final volume showed good agreement with experimental class averages, and the data displayed a comprehensive coverage of the Euler angular space. As calculated from the 0.5 value of the Fourier ring correlation (FRC) curve, the final resolution of the reconstruction was 22 Å (Fig. 3E and Fig. S4D). A comparison of EM maps reconstructed with and without imposing a twofold symmetry is shown in Fig. S5. The EM negative-stain map of dimeric LRRK2 suggests a compact globular architecture, where dimerization occurs via a single twofold rotation axis.

Determination of the LRRK2 Domain Assembly by Integrative Modeling. Seven unique domains have been identified from primary sequence of LRRK2, including three N-terminal repeat domains (armadillo, ankyrin, and LRR), followed by the RocCOR G-domain/dimerization interface, a kinase domain, and a C-terminal WD40 repeat (9). Single-domain crystal structures of the LRR (reported here), RocCOR, and kinase domains from LRRK2 homologs (e.g., a Roco protein from *Chlorobium tepidum*, Roco4 from *Dictyostelium discoideum*) are available and can be used in integrative modeling studies (35, 36). In addition, a structure of the human LRRK2 Roc domain has recently been published (37). Suitable templates to model the armadillo, ankyrin, and WD40 domains have been obtained from homologous proteins through the HHpred (38) homology detection and structure prediction server. The detailed analysis and refinement of the LRRK2 domain homology models are described in SI Materials and Methods. The domain borders, as well as the alignments used for the homology modeling of the single domains, are shown in Fig. 24 and Figs. S6–S9. The resulting domain borders are in agreement with previous work (39). For the determination of the quaternary domain assembly of the LRRK2 holoenzyme, a hierarchical and multiple-step integrative modeling approach was used. This modeling strategy was optimized by considering constraints provided by the cross-linking as well as SAXS data from

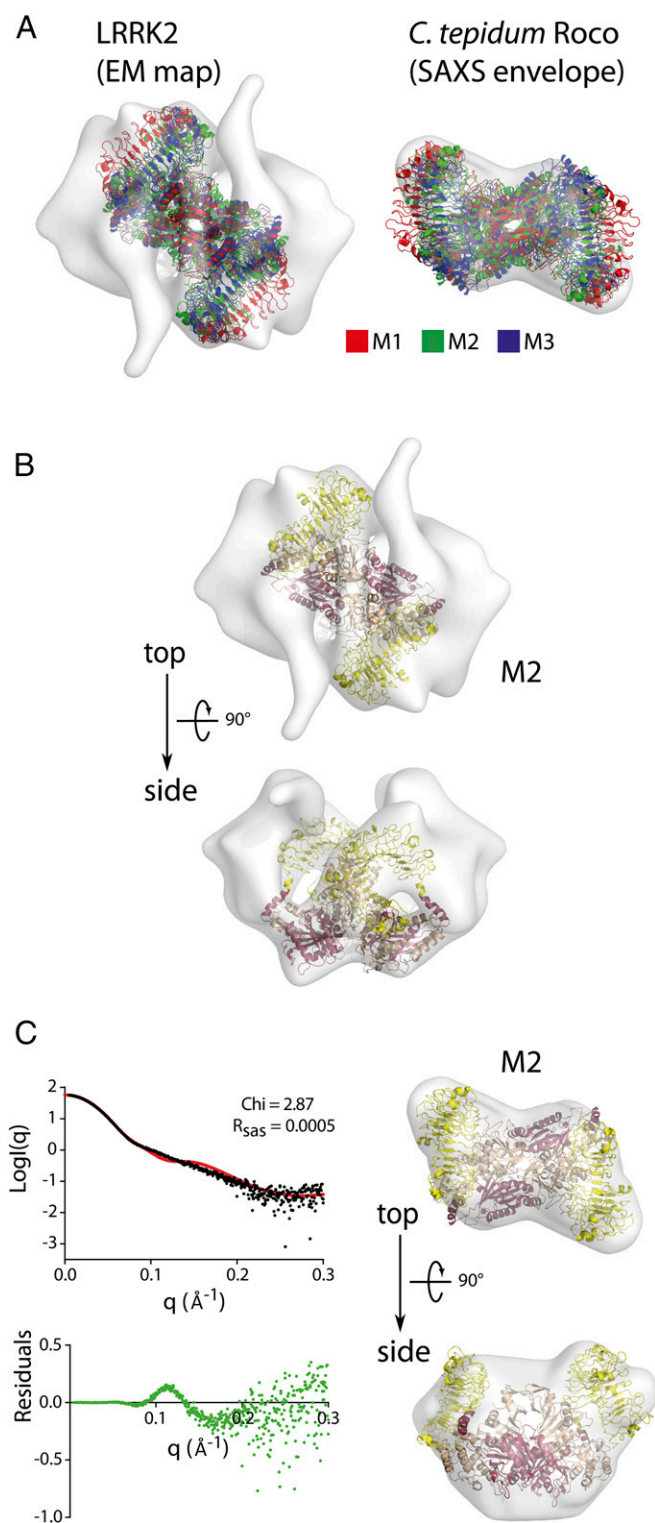


Fig. 5. (A) Superimposed top-scoring LRR-RocCOR conformers M1–M3 fitted to the EM map (Left) as well as to the ab initio averaged bead model envelope calculated from the SAXS data for the *C. tepidum* Roco protein (Right). (B) Docking of the LRR-RocCOR model M2 to the EM map. (C, Upper Left) Fit of the experimental SAXS data of the *C. tepidum* LRR-RocCOR protein (black) on the theoretical scattering curve (red line) corresponding to LRR-RocCOR conformer M2 using the program CRYSOLO. (C, Lower Left) Residuals of the fit are shown below (green). (C, Right) Docking of the LRR-RocCOR model M2 to the ab initio averaged bead model envelope calculated from the SAXS data for the *C. tepidum* Roco protein. The LRR-RocCOR conformer M2 gave the best-scoring final models (Fig. 7). q , scattering vector.

a *C. tepidum* Roco protein and the EM density map of dimeric LRRK2 (Fig. 4).

Hybrid Model for Core of LRRK2 by Docking LRR to RocCOR and Comparison with SAXS Data for the Homologous *C. tepidum* Roco protein. Modeling of LRRK2 quaternary structure was initiated by assembling the LRR domain on the RocCOR domain homology model structures. The recently reported structure of a bacterial Roco protein (*C. tepidum*) was used as a template (36) for the RocCOR dimer. The anchoring of the LRR domain relies mainly on the position of the α -helix (amino acids 1,309–1,328) of the Roc G-domain. The structure of the LRR domain of the *C. tepidum* Roco protein (PDB ID code 5IL7), which served as a template to model the LRRK2 LRR domain, also contains the Roc α -helix (36) at its C terminus, allowing the alignment of the LRR and RocCOR models at this position. Although bacterial Roco proteins lack the kinase domain, they can be used as a model for the LRR-RocCOR region (9). Two different alignments for the LRRK2 and *C. tepidum* Roco LRR domains were tested for threading calculations (*SI Materials and Methods*), leading to two alternative LRR model structures.

The Rosetta Floppy Tail (40) approach was subsequently adapted to sample in an exhaustive manner the possible rigid-body-like orientations of the LRR domain with respect to RocCOR, while keeping them covalently attached through the α -helix. The N-terminally preceding linker (amino acids 1,308–1,315) was sampled with full flexibility to allow local backbone conformational changes to be propagated downstream through the Rosetta internal coordinate system (i.e., foldtree), resulting in rigid-body rototranslations of the RocCOR domain relative to LRR. The best docking solutions (*SI Materials and Methods*) were selected based on cross-linking constraints provided by cross-linking coupled with MS (CL-MS), as well as by fitting to the EM map of dimeric LRRK2. Several good cross-linked fit (XL^{fit}) scoring models were obtained.

To validate the robustness of the top-scoring LRR-RocCOR conformers further, their theoretical scattering curves were compared with the experimental SAXS data of a *C. tepidum* LRR-RocCOR construct (Fig. S10) using CRYSOLO, which resulted in a broad range of χ values ranging from 2.08 to 21. The two LRR-RocCOR conformers that score best based on XL^{fit} and EM map docking, referred to as models M1 and M2, have an χ value of 6.70 and 2.87, respectively. However, both models display an overall conformation that is close to the conformation of the model that shows the best agreement with the experimental SAXS data ($\chi = 2.08$), referred to as model M3 (Fig. S11). An overlay of the resulting LRR-RocCOR models M1–M3 is shown in Fig. S12B. Fig. 5A shows the models docked to the LRRK2 EM map as well as the SAXS envelope of the bacterial Roco protein. All three conformers are compact [experimental radius of gyration (R_g^{exp}) = 45.0 Å, $R_{g, M1}$ = 46.4 Å, $R_{g, M2}$ = 44.8 Å, $R_{g, M3}$ = 44.4 Å], with an LRR domain that folds back toward the C-terminal end/top of the RocCOR module, and show an EM correlation score of >0.84 (Fig. S12A). This general conformation is also in excellent agreement with the ab initio averaged bead model envelope calculated based on the SAXS data (Fig. 5B and C and Fig. S11B). Although models M2 and M3 show good agreement with the SAXS data (especially considering the size of the protein), while clearly fitting the data better than model M1, deviations could also be due to inherent sequence variations between the proteins (human versus *C. tepidum* Roco), where the bacterial protein lacks a kinase, ankyrin, and WD40 domain. All three models, LRR-RocCOR M1 to M3, were then subjected to subsequent docking steps for the remaining domains.

Generation of a Hybrid Model for Dimeric LRRK2 Holoenzyme by Hierarchical Docking. An ensemble-based hierarchical docking approach was used to assemble the ankyrin, kinase, and WD40 domains on the LRR-RocCOR top-scoring conformers. Histograms showing the distributions of either XL^{fit} or EM fit (EM^{fit} , or EM

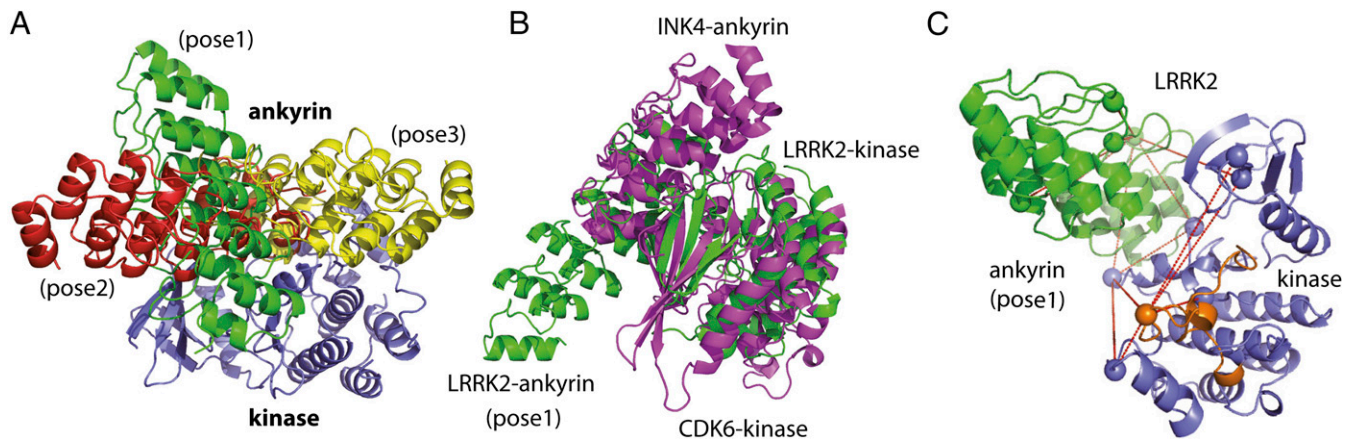


Fig. 6. Ankyrin-domain docked to the kinase domain mode via ZDOCK. (A) Three top poses obtained from the ankyrin domain docking to the kinase domain selected by experimental determined cross-linking constraints. (B) Superposition of one of the most favorable docking solutions to a crystal structure (PDB ID code 1G3N) of the inhibitory complex of the INK4 ankyrin domain with the CDK6 kinase domain (magenta) gives confidence to the model (green). (C) Experimentally determined cross-links (red lines) on the best-scoring ankyrin-kinase docking solution (pose1). The LRRK2 activation loop is highlighted in orange. Cross-linked lysine residues are shown as spheres.

map correlation value) for the three alternative LRR-RocCOR conformers are illustrated in Fig. S12C. The cross-linking data revealed several cross-links of the kinase domain to the ankyrin repeats, suggesting that the ankyrin domain is involved in the regulation of the kinase activity. Therefore, the ankyrin domain was separately docked onto the kinase domain, and the best docking solutions were filtered out through the XL^{fit} scoring approach. The model structure of the kinase in the nucleotide-free (apo) state showed a higher number of satisfied cross-link-related distances (Fig. S13), and was therefore used for the subsequent docking steps. Interestingly, the top-scoring ankyrin-kinase poses (Fig. 6A) are similar to existing crystal structures of a kinase domain of CDK6 with the inhibitory ankyrin domain of INK4 (41), suggesting a common mode of regulation of kinases by these domains (Fig. 6B). The top-scoring model for the LRRK2 ankyrin-kinase module with highest similarity to the latter crystal structure is shown in Fig. 6C, with highlighted cross-links satisfying the defined Euclidean distance limits.

The three best ankyrin-kinase models (Fig. 6A) were subsequently docked to the WD40 model. The nine best representative ankyrin-kinase-WD40 complexes were rigid-body-docked to the top-scoring three LRR-RocCOR conformers by preserving noncrystallographic (point) symmetry (*SI Materials and Methods*) imposed on the LRR-RocCOR model. The best solutions, selected based on cross-link constraints, as well as the 22-Å EM density map, were chosen for subsequent assembling of the remaining armadillo domain. Because no cross-linking constraints satisfying the confidence criteria were determined by the CL-MS method, mainly because of the distribution of lysine residues in this area (Fig. 2D) resulting in rather large peptides, the positioning of the armadillo domain relied primarily on fitting it into the remaining space of the EM density map. The resulting dimeric LRRK2 quaternary structure models were refined on the EM map through 200 cycles of sequential refitting within the University of California, San Francisco (UCSF) Chimera (42) fitmap algorithm. The sequential refitting was carried out by keeping each domain block internally rigid through the following order: LRR-RocCOR (chains A and B), ankyrin-kinase-WD40 (chain A), ankyrin-kinase-WD40 (chain B), armadillo (chain A), and armadillo (chain B).

The resulting top-scoring model for the dimeric LRRK2 holoenzyme shown in Fig. 7 is in good agreement with both the experimental CL-MS data and the EM map. The scoring results for models based on three different LRR-RocCOR conformers (M1–M3) are summarized in *Datasets S3–S5*. For the top pose (model 1)

(*Dataset S4*) based on LRR-RocCOR conformer M2 ($XL^{fit} = 83.4$) the cross-linking distance limits (*Materials and Methods*) were satisfied for 42 of 64 DSS (66%) and 15 of 30 DSG (50%) cross-links, respectively. In addition to the total number of satisfied cross-links, the XL^{fit} metric considers the deviation of the violated cross-links from the distance threshold as well as the distances of consecutive domain terminals. The models showing the best agreement with the CL-MS data had EM^{fit} correlation values greater than 0.90 (M1/M2 = 0.912). The top-scoring model (Fig. 7) had an accumulated terminal distance of 173 Å. An alternative model (model 14) was identified for the conformer M2, satisfying even a higher number of cross-linking constraints (60/ $XL^{fit} = 80.2$) at a similar EM^{fit} (0.9063) but at the expense of a higher accumulated terminal distance of adjacent repeat domains (266 Å). A comparison of both models is shown in Fig. S14A and B.

Interestingly, the final models strongly suggest that there are close contacts of the kinase domain with the N-terminal ankyrin and LRR repeat domains. In the final model, the kinase-ankyrin module localizes in a position close to the LRR domain and the α 0-helix connecting the LRR and the Roc G-domain (Fig. S14C).

Discussion

Understanding the structure of LRRK2 and how mutations alter that structure to disrupt normal function, and ultimately contribute to the pathogenesis of PD, is essential for developing potential therapeutics. Although important progress toward this goal has been made by the determination of the crystal structure of the dimeric RocCOR interface of an orthologous Roco protein from *C. tepidum* (36), as well as the kinase domain of the LRRK2 orthologous *D. discoideum* protein Roco4 (35), the 3D arrangement of the various domains in LRRK2 remains largely unknown. Our results confirm that LRRK2, similar to other Roco protein family members and in contrast to classical small G proteins, forms constitutive dimers (27). To obtain a structural model of the dimeric LRRK2 holoenzyme, we have used an integrative modeling pipeline combining domain homology models with molecular docking and experimental constraints provided by chemical cross-linking, a low-resolution LRRK2 density map generated by EM negative staining, and SAXS data for the homologous *C. tepidum* Roco protein. Both chemical cross-linking of recombinant human LRRK2 and the EM density map from LRRK2 purified from transgenic mice concordantly reveal a compact folding of the dimeric enzyme. Furthermore, the cross-linking approach allowed the identification of various interdomain contacts, including interactions between

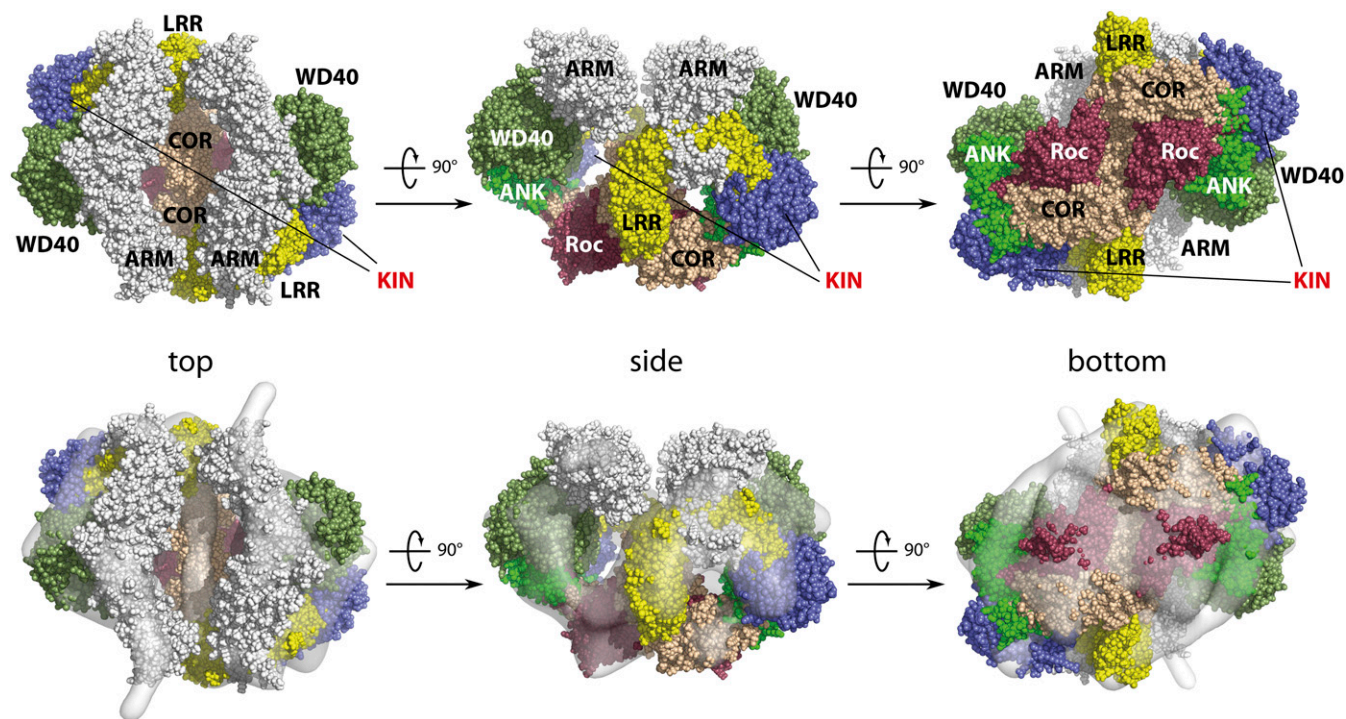


Fig. 7. Final LRRK2 models fitted to the EM map. (*Upper*) Top, side, and bottom views of the resulting top-scoring model (model 1/LRR-RocCOR conformer M2). (*Lower*) Superimposition with the EM negative-stain map. The seven LRRK2 domains are diagrammed: ARM (armadillo repeats), ANK (ankyrin repeats), LRR, Roc (G-domain Roc), COR (COR dimerization domain), KIN (kinase domain), and WD40 (WD40 repeat domain). The scoring results are provided in [Dataset S4](#) and [S5](#).

distant domains in agreement with previous reports (23). Given the presence of a Roc G-domain and kinase domain in one protein and the similarities of these modules to MAP kinase signaling cascades, it has been hypothesized that an intramolecular regulation mechanism involving these domains must exist for LRRK2 (5, 43). On the other hand, by testing direct binding between the G-domain Roc and the kinase domain, previous work demonstrated that these two domains do not bind to each other at high affinity (43). Interestingly, our model reveals spatial proximity of the kinase domain with the ankyrin domain and the LRR domain by chemical cross-linking. Furthermore, the top-scoring docking solutions for the ankyrin domain to the kinase domain resemble the structure determined for a complex from the ankyrin-repeat protein INK4 and the kinase domain of CDK6 (41). INK4 proteins are well-characterized tumor suppressors that inhibit cyclin kinases (44). In addition, based on the LRRK2 primary protein sequence, the LRR domain is directly coupled to both the C terminus of the ankyrin domain and the N terminus of the Roc G-domain. Although mechanistically not well understood, recent studies did show that the N terminus of human LRRK2 and related Roco proteins is essential for function *in vivo* (45). Our data suggest that the LRRK2 ankyrin-LRR domains may play a direct role in the regulation of the LRRK2 kinase activity, guiding future in-depth functional characterization of LRRK2 at a molecular level.

The final structural model, integrating multiple experimental data, supports findings demonstrating that the RocCOR domains form the primary dimerization interface and that the terminal domains are not likely to contribute much to the dimerization of LRRK2. However, although chemical cross-linking provides robust information for physical contacts of different functional domains within a protein, this type of data cannot distinguish intra- versus intermolecular cross-links (i.e., cross-links can be formed *in cis*, within a monomer, or *in trans* between two monomers). Because there is no straightforward analytical strategy to discriminate between *cis* and

trans cross-links, we considered this problem by computational analysis. All *cis* and *trans* distances were considered, with the shortest selected for scoring ([Datasets S6–S8](#)). The resulting models satisfy the highest number of cross-links and show good agreement with the EM density map of dimeric LRRK2. This finding suggests that most of the cross-links occur *in cis*. However, the LRRK2 COR-kinase linker regions (amino acids 1,849–1,878) are in close spatial proximity in the *C. tepidum* RocCOR template structure (36). In consequence, the positions of the C-terminal domains in the final models are also in agreement with an intertwined homodimer as found for other dimeric proteins (46). Therefore, an arrangement where the C-terminal kinase and WD40 domain would interact with the N-terminal ankyrin and LRR domains of the other monomer would still be in agreement with our experimental constraints. For this reason, we did not assign the domains to a specific monomer/chain in the final models.

It also has to be considered that cross-linking data are affected by the presence of multiple conformations. Consistently, we observed heterogeneity in the shape of the EM particles. For a large and dynamic multidomain protein, such as LRRK2, the presence of multiple simultaneous conformations in solution is not unlikely, which could also give a reasonable explanation for the observation that no model was derived that could satisfy all cross-links and EM restraints at the same time.

In conclusion, by combining chemical cross-linking, EM analysis, integrative modeling, and biochemical experiments, we present, for the first time to our knowledge, a compact architecture of the LRRK2 dimer with a domain assembly where distant domains engage in multiple contacts. Our model also suggests that the N-terminal ankyrin and LRR repeats are involved in the intramolecular regulation of the biological activity of LRRK2. The model presented will be useful for future structure-driven functional studies and, furthermore, will support future rational drug design studies (i.e., by

suggesting promising mechanisms for allosteric inhibitors targeting domain–domain contacts).

Materials and Methods

Generation of LRRK2 Expression Constructs. The generation of N-terminal, SF-tagged (NSF), full-length LRRK2 expression constructs has been previously described (19). Similar methods were used to create an N-terminal deletion construct, 1,280-end (Δ 1,279-LRRK2) using the pDEST-NSF-tandem affinity purification (47) Gateway cloning vector.

Cell Culture and Transfection. HEK293T cells (CRL-11268; American Type Culture Collection) were cultured in 14-cm dishes in DMEM (PAA) supplemented with 10% (vol/vol) FCS (PAA) and appropriate antibiotics. Cells were transfected at a confluence between 50% and 70% with 8 μ g of plasmid DNA per dish using polyethyleneimine (Polysciences) solution as described previously (48). After transfection, cells were cultured for 48 h.

Protein Purification. Purification of SF-tagged LRRK2 purified from 4 \times 14-cm dishes (600 cm²) of confluent HEK293T cells was performed as described previously (19, 49). Briefly, after removal of the medium, cells were lysed in 1 mL of lysis buffer and washing buffer [50 mM Hepes (pH 8.0), 100 mM NaCl, 1 mM DTT, 5 mM MgCl₂, 0.5 mM EDTA, 10% (vol/vol) glycerol, 0.01% Triton X-100], supplemented with 0.55% Nonidet P-40 and Complete Protease Inhibitors (Roche), per 14-cm dish. Incubation with lysis buffer was then performed for 40 min at 4 °C on a shaker. Cell debris and nuclei were removed by centrifugation of the lysates at 10,000 \times g for 10 min. Cleared lysates were incubated with 100 μ L of settled Strep-Tactin Superflow resin (IBA) for 1 h at 4 °C. Beads were transferred to microspin columns (GE Healthcare) and washed five times with 500 mL of washing buffer prior to final elution using 400 μ L of elution buffer (washing buffer supplemented with 200 mM D-threitol; IBA).

DLS. DLS was measured using a DynaPro NanoStar instrument with a 50- μ L volume in a disposable cuvette (Eppendorf UVette) using a protein concentration of 0.1 mg/mL. Ten measurements of 10 s each were collected. Data were analyzed using Dynamics v.7.1.9 software.

Cross-Linking Reagents. The 1:1 mixtures of DSS H12/D12 and DSG H4/D4 were purchased from Creative Molecules.

Chemical Cross-Linking. For NHS-ester–based chemical cross-linking, 2 μ L of DSS-H12/D12 solution (12.5 mM in DMSO) or DSG-H6/D6 solution (12.5 mM in DMSO) was added to 100 μ g of purified LRRK2 in a total volume of 0.4 mL of elution buffer. The reactions were carried out under constant shaking for up to 45 min. The reaction was then stopped by adding Tris-HCl (pH 7.5) solution to a final concentration of 100 mM. The mix was incubated for 15 min at room temperature under constant shaking.

Proteolytic Cleavage and SEC Prefractionation. After chemical cross-linking, protein samples were precipitated with chloroform and methanol (49). Proteolysis was performed as described earlier (49), and the resulting peptide mixtures were separated by SEC following the protocols described by Leitner et al. (25). Additional information is provided in *SI Materials and Methods*.

MS Analysis. SpeedVac-dried SEC fractions were redissolved in 0.5% TFA and analyzed by LC-MS/MS using a nanoflow HPLC system (Ultimate 3000 RSLC; Thermo Fisher) coupled to an Orbitrap Velos (Thermo Fisher) tandem mass spectrometer. For identification, cross-linked peptides were separated on the nano-HPLC by 120-min gradients and analyzed by a data-dependent approach acquiring CID MS/MS spectra of the 10 most intense peaks, excluding single- and double-charged ions.

Data Analysis Using xQuest/xProphet. Before analysis via xQuest/xProphet (v2.1.1) (25), MS/MS spectra were extracted from the RAW files using ReAdW (v4.3.1). In the case of Q-Exacte data, the ProteoWizard/msconvert (3.0.6002) suite was used with the parameters described by Leitner et al. (25). The identification of monolinks, looplinks, and cross-links was done based on the identification of DSS H12/D12 or DSG H6/D6 pairs following protocols published by Leitner et al. (25). Briefly, for database search, a database containing 46 proteins was prepared containing common contaminants and interactors of LRRK2. Additionally, a decoy database containing the inverted sequences was generated. For xQuest/xProphet, standard parameters according Leitner et al. (25) were used with carbamidomethyl as a fixed modification and methionine oxidation as a variable modification. For DSS-D12 and DSG-D6, isotope differences of 12.075321 Da and 6.03705 Da, respectively, were used. For CID spectra, the parent ion

tolerance was set to 10 ppm (MS) and the fragment ion tolerance to 0.3 Da (MS/MS). For higher-energy collisional dissociation (HCD) spectra, the fragment ion tolerance was set to 0.02 Da. Only those cross-linked peptides that fulfilled specific minimal criteria (ID score > 28, Δ S < 0.95, FDR < 0.05) were considered for modeling. Additionally, the MS/MS spectra were evaluated by manual inspection to ensure a good representation of the fragment series of both cross-linked peptides.

EM Analysis and 3D Map Reconstruction. A 3- μ L aliquot of LRRK2 [40 ng/ μ L protein, 20 mM Tris-HCl (pH 8.0), 150 mM NaCl] was adsorbed for 45 s onto glow-discharged holey grids (Quantifoil R2/4 on Cu/Rh 300 mesh) coated with thin carbon. The specimen was stained with 2% (wt/vol) uranyl acetate, blotted three times, and finally air-dried. The initial dataset of images used to generate the initial model was selected by imaging the specimens with a TF20 (FEI) field emission gun transmission electron microscope (FEG) at 200 kV under low-dose conditions, using a 4,000 \times 4,000-pixel CCD camera at the equivalent calibrated magnification of 88,249 \times and -1.5 - μ m defocus. A second dataset was taken on a J2100F (JEOL) FEG, using a 2,000 \times 2,000-pixel CCD camera at 63,450 \times and -1.5 - μ m defocus.

CCD images were evaluated for drift and astigmatism in Fourier space, and the contrast transfer function (CTF) was estimated using CTFIND3 (50). An initial dataset of 3,500 dimer particles was selected using EMAN (32) and then normalized and CTF-corrected using Xmipp software (51). The particles were subject to 2DMLF analysis as implemented in Xmipp (52). In this process, particle images with an irregular background, close neighbors, overlapping particles, and aggregates were discarded to yield a dataset of 2,600 particles. The 2DMLF-screened particle images were classified by reference-free alignment into 24 class averages by EMAN. The 24 class averages were then used to reconstruct an initial 3D model by the cross-common lines technique and refined through 16 cycles by EMAN without imposing any symmetry constraint. A second 16-iteration refinement cycle was operated imposing a twofold symmetry to produce a definitive initial model. The second dataset of 65,000 particles was selected, CTF-corrected, and 2DMLF-screened in the same way to yield a total of 51,000 particles. This dataset was used to refine the initial model following the multi-resolution refinement protocol described by Scheres et al. (52) in steps 37–47. The reconstructions of the final volume and the class averages were visualized in EMAN. The Euler angular coverage was calculated by Xmipp (52). The resolution was estimated according to the 0.5 value of the FRC curve, and the maps were low-pass-filtered at this spatial frequency.

SAXS Analysis of the *C. tepidum* LRR-RocCOR Domain Construct. A *C. tepidum* Roco protein fragment corresponding to the LRR-RocCOR domain (amino acids 1–946) was cloned in a pProEX vector and expressed and purified in a nucleotide-free form, as described earlier (36). The SAXS experiment was performed on the BM29 beamline at the European Synchrotron Radiation Facility in Grenoble, France, with an online HPLC set-up. Fifty microliters of an 8-mg/mL sample was injected on a Bio SEC-3 HPLC column (Agilent) equilibrated with 20 mM Hepes (pH 7.5), 150 mM NaCl, 5 mM MgCl₂, 5% glycerol, and 1 mM DTT. Individual scatter curves corresponding to the elution peak were buffer-subtracted and averaged. PRIMUS was used for determination of the R_g using Guinier approximation (53) and GNOM for the calculation of the pair distance distribution function P(r) (54). All models of the human LRR-RocCOR module were compared with the experimental data using CRY SOL (55) and the R_{SAS} metric was calculated with ScÅtter (56). Ab initio envelopes were calculated using DAMMIF (average of 20 runs) (57), and docking into the envelope was performed with Supcomb (58). The Situs pdb2vol module was used to convert the ab initio models into volumetric maps (59). Experimental and modeling parameters for the SAXS analysis are provided in [Dataset S9](#).

Integrative Modeling. A detailed description of the integrative modeling procedure is provided in *SI Materials and Methods* and graphically summarized in Fig. 4 (workflow). Briefly, structures of the single domains were obtained by homology modeling and assembled through an integrative modeling procedure consisting of hierarchical, rigid-body docking calculations carried out by means of the ZDOCK algorithm (60) using established protocols (61–63). Structural models obtained from global docking approaches were scored and filtered by applying cross-link–derived distance filters, using 35-Å and 31-Å thresholds for DSS and DSG, respectively (64). The overall fit of a given docking pose to the cross-link dataset was measured through the XL^{fit} score (*SI Materials and Methods*). Homology models for the ankyrin and kinase domains were used for subsequent docking of the WD40 domain. The best scoring ankyrin–kinase–WD40 ternary complexes were finally used for rigid-body docking on selected LRR-RocCOR conformers; these were obtained by manually placing the LRR domain on the RocCOR structural model by superimposing the common α -helix and by exhaustively sampling the orientation of the covalently attached LRR domain with respect to the RocCOR unit through an adaptation of the Floppy Tail approach (40). The obtained LRR-RocCOR conformers were additionally cross-validated via

comparison with the experimental SAXS data of the *C. tepidum* LRR-RocCOR construct. Final models for dimeric full-length LRRK2 were selected from initial docking solutions combining cross-link constraints and the negative-stain EM map based on XL^{fit} and EM^{fit} scores. See Table S1 for statistics for the crystallographic analysis of the LRR domain of the *C. tepidum* Roco protein.

SDS and BN Gel Electrophoresis. SDS/PAGE was performed using 4–12% NuPage (Life Technologies) gradient gels. BN gel electrophoresis has been performed on 4–12% NativePage gels (Life Technologies) according to the vendor's protocols. The gels have been stained with colloidal Coomassie, as described elsewhere (49).

Kinase Assay. LRRK2 kinase activity was measured by incorporation of 32P into the LRRKtide peptide in 30 °C kinase buffer [consisting of 20 mM Tris (pH 7.5), 10 mM MgCl₂, 1 mM EGTA, 1 mM sodium orthovanadate, 1 mM NaF, 5 mM β-glycerolphosphate, 0.02% Triton X-100, 2 mM DTT] in the presence of 500 μM GDP. To start the reaction, 0.03-mg/mL full-length LRRK2 was mixed with 75 μM LRRKtide and 25 μM ATPγ32P (2 Ci/mmol) and was stopped by adding 100 mM ice-cold EDTA. Samples were spotted on nitrocellulose filters, washed with 50 mM phosphoric acid, and dried before scintillation counting (PerkinElmer).

GTPase Assay (Charcoal). LRRK2 GTPase activity was measured by adding 10 μM ATPγ32P to 0.1 μM recombinant LRRK2 at 30 °C. Aliquots of 10 μL

were taken at certain time points and mixed with 400 μL of charcoal solution (50 g/L charcoal in 20 mM phosphoric acid) to stop the reaction. The charcoal was pelleted, and the amount of free 32Pi in the supernatant was determined by scintillation counting.

Statistical Analysis. Linear regression curves were calculated using GraFit v5 (Erithacus Software) weighted by individually determined errors for each data point.

ACKNOWLEDGMENTS. We thank Profs. Alfred Wittinghofer, Francesca Fanelli, and Rob Russell for helpful discussions; the staff of the Medical Proteome Center Tubingen for technical assistance; and the staff of the BM29 beamline at the European Synchrotron Radiation Facility. Computational resources and staff expertise were provided by the Department of Scientific Computing at the Icahn School of Medicine at Mount Sinai. This work has been supported in part by the Linked Efforts to Accelerate Parkinson's Solutions initiative of The Michael J. Fox Foundation for Parkinson's Research; an Alexander Von Humboldt postdoctoral fellowship (to F. Raimondi); a Boehringer Ingelheim Fonds doctoral fellowship (to P.K.A.J.); a Netherlands Organisation for Scientific Research NWO-VIDI grant (to A.K.); VUB SRP financing, Fonds Wetenschappelijk Onderzoek Vlaanderen, and Hercules Stichting (W.V.); NIH Instrumentation Award 1S10RR026473-01A1 (to I.U.-B.); and the EM resources at the New York Structural Biology Center.

- Fahn S (2003) Description of Parkinson's disease as a clinical syndrome. *Ann N Y Acad Sci* 991:1–14.
- Zimprich A, et al. (2004) Mutations in LRRK2 cause autosomal-dominant parkinsonism with pleomorphic pathology. *Neuron* 44(4):601–607.
- Paisán-Ruiz C, et al. (2004) Cloning of the gene containing mutations that cause PARK8-linked Parkinson's disease. *Neuron* 44(4):595–600.
- Gasser T (2009) Molecular pathogenesis of Parkinson disease: Insights from genetic studies. *Expert Rev Mol Med* 11:e22.
- Biosa A, et al. (2013) GTPase activity regulates kinase activity and cellular phenotypes of Parkinson's disease-associated LRRK2. *Hum Mol Genet* 22(6):1140–1156.
- Gloeckner CJ, et al. (2006) The Parkinson disease causing LRRK2 mutation I2020T is associated with increased kinase activity. *Hum Mol Genet* 15(2):223–232.
- West AB, et al. (2005) Parkinson's disease-associated mutations in leucine-rich repeat kinase 2 augment kinase activity. *Proc Natl Acad Sci USA* 102(46):16842–16847.
- Marin I, van Egmond WN, van Haastert PJ (2008) The Roco protein family: A functional perspective. *FASEB J* 22(9):3103–3110.
- Giltsbach BK, Kortholt A (2014) Structural biology of the LRRK2 GTPase and kinase domains: Implications for regulation. *Front Mol Neurosci* 7:32.
- Andrade MA, Perez-Iratxeta C, Ponting CP (2001) Protein repeats: Structures, functions, and evolution. *J Struct Biol* 134(2–3):117–131.
- Jaleel M, et al. (2007) LRRK2 phosphorylates moesin at threonine-558: Characterization of how Parkinson's disease mutants affect kinase activity. *Biochem J* 405(2):307–317.
- Steger M, et al. (2016) Phosphoproteomics reveals that Parkinson's disease kinase LRRK2 regulates a subset of Rab GTPases. *eLife* 5:5.
- Lewis PA, et al. (2007) The R1441C mutation of LRRK2 disrupts GTP hydrolysis. *Biochem Biophys Res Commun* 357(3):668–671.
- Tsika E, Moore DJ (2013) Contribution of GTPase activity to LRRK2-associated Parkinson disease. *Small GTPases* 4(3):164–170.
- Cohen P, Alessi DR (2013) Kinase drug discovery—What's next in the field? *ACS Chem Biol* 8(1):96–104.
- Fuji RN, et al. (2015) Effect of selective LRRK2 kinase inhibition on nonhuman primate lung. *Sci Transl Med* 7(273):273ra15.
- Baptista MA, et al. (2013) Loss of leucine-rich repeat kinase 2 (LRRK2) in rats leads to predictive abnormal phenotypes in peripheral organs. *PLoS One* 8(11):e80705.
- Herzig MC, et al. (2011) LRRK2 protein levels are determined by kinase function and are crucial for kidney and lung homeostasis in mice. *Hum Mol Genet* 20(21):4209–4223.
- Gloeckner CJ, et al. (2010) Phosphopeptide analysis reveals two discrete clusters of phosphorylation in the N-terminus and the Roc domain of the Parkinson-disease associated protein kinase LRRK2. *J Proteome Res* 9(4):1738–1745.
- 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):5511–5523.
- Sen S, Webber PJ, West AB (2009) Dependence of leucine-rich repeat kinase 2 (LRRK2) kinase activity on dimerization. *J Biol Chem* 284(52):36346–36356.
- Civiero L, et al. (2012) Biochemical characterization of highly purified leucine-rich repeat kinases 1 and 2 demonstrates formation of homodimers. *PLoS One* 7(8):e43472.
- Greggio E, et al. (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.
- Leitner A, et al. (2010) Probing native protein structures by chemical cross-linking, mass spectrometry, and bioinformatics. *Mol Cell Proteomics* 9(8):1634–1649.
- Leitner A, Walzthoeni T, Aebersold R (2014) Lysine-specific chemical cross-linking of protein complexes and identification of cross-linking sites using LC-MS/MS and the xQuest/xProphet software pipeline. *Nat Protoc* 9(1):120–137.
- Leitner A, et al. (2012) Expanding the chemical cross-linking toolbox by the use of multiple proteases and enrichment by size exclusion chromatography. *Mol Cell Proteomics* 11(3):M111.014126.
- Terheyden S, Ho FY, Gilsbach BK, Wittinghofer A, Kortholt A (2015) Revisiting the Roco G-protein cycle. *Biochem J* 465(1):139–147.
- Webber PJ, et al. (2011) Autophosphorylation in the leucine-rich repeat kinase 2 (LRRK2) GTPase domain modifies kinase and GTP-binding activities. *J Mol Biol* 412(1):94–110.
- Li X, et al. (2010) Enhanced striatal dopamine transmission and motor performance with LRRK2 overexpression in mice is eliminated by familial Parkinson's disease mutation G2019S. *J Neurosci* 30(5):1788–1797.
- Ohi M, Li Y, Cheng Y, Walz T (2004) Negative staining and image classification—Powerful tools in modern electron microscopy. *Biol Proced Online* 6:23–34.
- Hainfeldt JF, Powell RD (2000) New frontiers in gold labeling. *J Histochem Cytochem* 48(4):471–480.
- Ludtke SJ, Baldwin PR, Chiu W (1999) EMAN: Semiautomated software for high-resolution single-particle reconstructions. *J Struct Biol* 128(1):82–97.
- Tang G, et al. (2007) EMAN2: An extensible image processing suite for electron microscopy. *J Struct Biol* 157(1):38–46.
- Scheres SH, et al. (2005) Maximum-likelihood multi-reference refinement for electron microscopy images. *J Mol Biol* 348(1):139–149.
- Giltsbach BK, et al. (2012) Roco kinase structures give insights into the mechanism of Parkinson disease-related leucine-rich-repeat kinase 2 mutations. *Proc Natl Acad Sci USA* 109(26):10322–10327.
- 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.
- Deng J, et al. (2008) Structure of the ROC domain from the Parkinson's disease-associated leucine-rich repeat kinase 2 reveals a dimeric GTPase. *Proc Natl Acad Sci USA* 105(5):1499–1504.
- Soding J, Biegert A, Lupas AN (2005) The HHpred interactive server for protein homology detection and structure prediction. *Nucleic Acids Res* 33(Web Server issue):W244–W248.
- 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. *Biochem Soc Trans* 40(5):1086–1089.
- Kleiger G, Saha A, Lewis S, Kuhlman B, Deshaies RJ (2009) Rapid E2-E3 assembly and disassembly enable processive ubiquitylation of cullin-RING ubiquitin ligase substrates. *Cell* 139(5):957–968.
- Jeffrey PD, Tong L, Pavletich NP (2000) Structural basis of inhibition of CDK-cyclin complexes by INK4 inhibitors. *Genes Dev* 14(24):3115–3125.
- Petersen EF, et al. (2004) UCSF Chimera—A visualization system for exploratory research and analysis. *J Comput Chem* 25(13):1605–1612.
- Taymans JM (2012) The GTPase function of LRRK2. *Biochem Soc Trans* 40(5):1063–1069.
- Cánepa ET, et al. (2007) INK4 proteins, a family of mammalian CDK inhibitors with novel biological functions. *IUBMB Life* 59(7):419–426.
- van Egmond WN, van Haastert PJ (2010) Characterization of the Roco protein family in *Dictyostelium discoideum*. *Eukaryot Cell* 9(5):751–761.
- MacKinnon SS, Wodak SJ (2015) Landscape of intertwined associations in multi-domain homo-oligomeric proteins. *J Mol Biol* 427(2):350–370.
- Gloeckner CJ, Boldt K, Schumacher A, Roepman R, Ueffing M (2007) A novel tandem affinity purification strategy for the efficient isolation and characterisation of native protein complexes. *Proteomics* 7(23):4228–4234.
- Ben-Zimra M, Koler M, Orly J (2002) Transcription of cholesterol side-chain cleavage cytochrome P450 in the placenta: Activating protein-2 assumes the role of steriodogenic factor-1 by binding to an overlapping promoter element. *Mol Endocrinol* 16(8):1864–1880.

49. Gloeckner CJ, Boldt K, Ueffing M (2009) Strep/FLAG tandem affinity purification (SF-TAP) to study protein interactions. *Curr Protoc Protein Sci* Chapter 19:Unit19.20.
50. Mindell JA, Grigorieff N (2003) Accurate determination of local defocus and specimen tilt in electron microscopy. *J Struct Biol* 142(3):334–347.
51. Marabini R, et al. (1996) Xmipp: An image processing package for electron microscopy. *J Struct Biol* 116(1):237–240.
52. Scheres SH, Núñez-Ramírez R, Sorzano CO, Carazo JM, Marabini R (2008) Image processing for electron microscopy single-particle analysis using XMIPP. *Nat Protoc* 3(6):977–990.
53. Konarev PV, Volkov VV, Sokolova AV, Koch MHJ, Svergun DI (2003) PRIMUS—A Windows-PC based system for small-angle scattering data analysis. *J Appl Crystallogr* 36:1277–1282.
54. Svergun DI (1992) Determination of the regularization parameter in indirect-transform methods using perceptual criteria. *J Appl Crystallogr* 25:495–503.
55. Svergun DI, Barberato C, Koch MHJ (1995) CRYSOLO—A program to evaluate X-ray solution scattering of biological macromolecules from atomic coordinates. *J Appl Crystallogr* 28:768–773.
56. Rambo RP, Tainer JA (2013) Accurate assessment of mass, models and resolution by small-angle scattering. *Nature* 496(7446):477–481.
57. Franke D, Svergun DI (2009) DAMMIF, a program for rapid ab-initio shape determination in small-angle scattering. *J Appl Crystallogr* 42:342–346.
58. Kozin M, Svergun DI (2001) Automated matching of high- and low-resolution structural models. *J Appl Crystallogr* 34:33–41.
59. Wriggers W (2010) Using Situs for the integration of multi-resolution structures. *Biophys Rev* 2(1):21–27.
60. Mintseris J, et al. (2007) Integrating statistical pair potentials into protein complex prediction. *Proteins* 69(3):511–520.
61. Ambrosio M, et al. (2010) Superactive mutants of thromboxane prostanoid receptor: Functional and computational analysis of an active form alternative to constitutively active mutants. *Cell Mol Life Sci* 67(17):2979–2989.
62. Fanelli F, et al. (2011) Light on the structure of thromboxane A₂ receptor heterodimers. *Cell Mol Life Sci* 68(18):3109–3120.
63. Mignogna ML, et al. (2015) The intellectual disability protein RAB39B selectively regulates GluA2 trafficking to determine synaptic AMPAR composition. *Nat Commun* 6:6504.
64. Erzberger JP, et al. (2014) Molecular architecture of the 40S-eIF1-eIF3 translation initiation complex. *Cell* 158(5):1123–1135.
65. Combe CW, Fischer L, Rappsilber J (2015) xiNET: Cross-link network maps with residue resolution. *Mol Cell Proteomics* 14(4):1137–1147.
66. Sali A, Blundell TL (1993) Comparative protein modelling by satisfaction of spatial restraints. *J Mol Biol* 234(3):779–815.
67. Sievers F, et al. (2011) Fast, scalable generation of high-quality protein multiple sequence alignments using Clustal Omega. *Mol Syst Biol* 7:539.
68. Cole C, Barber JD, Barton GJ (2008) The Jpred 3 secondary structure prediction server. *Nucleic Acid Res* 36(Web Server issue):W197–W201.
69. Jones DT (1999) Protein secondary structure prediction based on position-specific scoring matrices. *J Mol Biol* 292(2):195–202.
70. Shen MY, Sali A (2006) Statistical potential for assessment and prediction of protein structures. *Protein Sci* 15(11):2507–2524.
71. Chen VB, et al. (2010) MolProbity: All-atom structure validation for macromolecular crystallography. *Acta Crystallogr D Biol Crystallogr* 66(Pt 1):12–21.
72. Adams PD, et al. (2010) PHENIX: A comprehensive Python-based system for macromolecular structure solution. *Acta Crystallogr D Biol Crystallogr* 66(Pt 2):213–221.
73. Tyka MD, et al. (2011) Alternate states of proteins revealed by detailed energy landscape mapping. *J Mol Biol* 405(2):607–618.
74. Leaver-Fay A, et al. (2011) ROSETTA3: An object-oriented software suite for the simulation and design of macromolecules. *Methods Enzymol* 487:545–574.
75. Stein A, Kortemme T (2013) Improvements to robotics-inspired conformational sampling in rosetta. *PLoS One* 8(5):e63090.
76. Seeber M, et al. (2011) Wordom: A user-friendly program for the analysis of molecular structures, trajectories, and free energy surfaces. *J Comput Chem* 32(6):1183–1194.
77. Kahraman A, et al. (2013) Cross-link guided molecular modeling with ROSETTA. *PLoS One* 8(9):e73411.
78. Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ (1990) Basic local alignment search tool. *J Mol Biol* 215(3):403–410.
79. André I, Bradley P, Wang C, Baker D (2007) Prediction of the structure of symmetrical protein assemblies. *Proc Natl Acad Sci USA* 104(45):17656–17661.
80. Raman S, et al. (2009) Structure prediction for CASP8 with all-atom refinement using Rosetta. *Proteins* 77(Suppl 9):89–99.
81. McWilliam H, et al. (2013) Analysis tool web services from the EMBL-EBI. *Nucleic Acid Res* 41(Web Server issue):W597–W600.
82. Söding J (2005) Protein homology detection by HMM-HMM comparison. *Bioinformatics* 21(7):951–960.
83. Wang C, Bradley P, Baker D (2007) Protein-protein docking with backbone flexibility. *J Mol Biol* 373(2):503–519.
84. Kim DE, Chivian D, Baker D (2004) Protein structure prediction and analysis using the Robetta server. *Nucleic Acid Res* 32(Web Server issue):W526–W531.
85. Merkley ED, et al. (2014) Distance restraints from crosslinking mass spectrometry: Mining a molecular dynamics simulation database to evaluate lysine-lysine distances. *Protein Sci* 23(6):747–759.
86. Cock PJ, et al. (2009) Biopython: Freely available Python tools for computational molecular biology and bioinformatics. *Bioinformatics* 25(11):1422–1423.
87. Louis-Jeune C, Andrade-Navarro MA, Perez-Iratxeta C (2012) Prediction of protein secondary structure from circular dichroism using theoretically derived spectra. *Proteins* 80(2):374–381.

Supporting Information

Guaitoli et al. 10.1073/pnas.1523708113

SI Materials and Methods

Proteolytic Cleavage and SEC Prefractionation. After chemical cross-linking, protein samples were precipitated by chloroform methanol (49). Proteolysis was performed as described earlier (49). Briefly, the protein precipitates were dissolved in 50 mM ammonium bicarbonate containing 0.2% RapiGest (Waters) reduced with DTT and alkylated with iodoacetamide, and proteolysis was performed by adding either 2 μg of trypsin (Promega) or a combination of 2 μg of trypsin and 0.2 μg of AspN (Promega). Samples were incubated at 37 °C overnight. After proteolysis, the RapiGest surfactant was hydrolyzed by acidification and removed by centrifugation. The samples were further cleaned up via C18-StageTips (Thermo Fisher) following standard protocols. Sample volumes were reduced to ~ 10 μL in a SpeedVac. Subsequently, the peptide mixtures were separated by SEC following protocols as described by Leitner et al. (25). Briefly, SpeedVac-dried samples were redissolved in 30 μL of SEC buffer [30% (vol/vol) acetonitrile, 0.1% TFA]. Samples were loaded in two fractions (15 μL) and separated on a Superdex Peptide PC 3.2/30 column (GE-Healthcare) at a flow rate of 50 μL on an Ettan LC system (GE-Healthcare). One hundred-microliter fractions were collected. UV traces were recorded at 214 and 260 nm.

Gold Labeling. Aliquots of wild-type LRRK2 (40 ng/ μL) were incubated with either monoclonal anti-Flag M2 or an in-house-generated polyclonal anti-LRR antibody on ice for 8 h. Five nanometers of gold-labeled rabbit anti-IgG secondary antibody (Sigma) was added to this mixture at a final concentration of 40 μM and incubated on ice overnight. The samples were adsorbed onto carbon-coated grids and stained with 2% uranyl acetate following the same procedure used for the nonlabeled particles.

Homology Modeling of LRRK2 Domains Based on X-Ray Structures of Homologous Roco Proteins.

Kinase domain.

Active-like conformation. The structural model of the kinase domain of human LRRK2 protein (UniProt ID code Q5S007, residues 1,879–2,138) in its active-like conformation was obtained by homology modeling through MODELER (66). The adenosine-5'-[(beta,gamma)-methylene]triphosphate (AppCp)-bound structure of the Roco4 kinase domain from *Dictyostelium discoideum* (PDB ID code 4F0F) was used as a template for comparative modeling calculations (35). Several templates, different in the length of deletions in several loop regions (characterized by lower sequence homology), were tested. Alignments were generated by means of ClustalOmega (67) (Fig. S6). The final template structure giving the most reliable models held deletions of the following residue stretches: 1,057–1,071 and 1,260–1,263. The α -helical constraints were imposed to residues 2,105–2,106 according to secondary structure predictions by means of JPRED (68) and PsiPRED (69). For each template, 100 models were generated by randomizing all of the Cartesian coordinates of standard residues in the initial model. The 20 models with the lowest number of stereochemical constraint violations were selected and ranked according to the discrete optimized protein energy (DOPE) score (70). The best 10 models were validated for their stereochemical quality through the Molprobit tool (71) available in the Phenix software package (72). The top-quality model was finally subjected to 15 cycles of backbone-restrained relaxation through the Rosetta FastRelax protocol (73) available in Rosetta3.5 (74).

Apo-like conformation. To model the structure of the LRRK2 kinase domain in its apo-like conformation, the apo structure of the Roco4 kinase domain from *D. discoideum* (PDB ID code 4F0G) was used (35).

The same alignment, as well as modeling and refinement protocols used for the active-like structure predictions, was used. Because the activation loop is missing in the 4F0G structure, we ran loop prediction calculations of the corresponding LRRK2 kinase domain residues (i.e., 2,020–2,034) on the top-quality structure obtained from homology modeling. One thousand models were generated using the next-generation, robotics-based kinematic closure method (75) available in Rosetta3.5 (74), utilizing default settings.

To select the best representative loop conformation, clustering of the best 10% conformers (according to Rosetta score) has been performed through the quality threshold (QT) algorithm implemented in Wordom software (76). To assess structure similarity, C α -atoms root mean square deviation (C α -RMSD) with a cut-off of 3 Å was used. Best model reliability was finally probed by mapping cross-link-related distances through Xwalk (77). The model from the first largest cluster, with the highest Rosetta score and stereochemical quality, holding at the same time the highest number of cross-links mapped, was retained as a representative structure of the apo-like state for further calculations.

RocCOR domain. The structural model of the monomeric RocCOR of human LRRK2 protein (UniProt ID code Q5S007, residues 1,309–1,842) was obtained by homology modeling through MODELER (66). The structure of the monomeric RocCOR domain of the *C. tepidum* Roco protein (PDB ID code 3DPU, chain A), published recently, was used as an initial template for comparative modeling calculations (36). For optimal modeling of LRRK2-specific sequence insertions compared with the RocCOR template structure, 16 different templates were tested, differing in length of deletions in several loop regions of both the Roc and COR domains, characterized by lower homology (final alignment is shown in Fig. S6). Alignments were generated by means of ClustalOmega (67) and refined according to BLAST-derived alignment (78). The template giving the most reliable models was characterized by deletions of the following residue stretches: 441–447, 496–499, 541–545, 570–580, 625–627, 647–649, 686–687, 711–712, 805–807, 820–822, 840–841, and 858–863. LRRK2 RocCOR residue stretches 1,546–1,549 and 1,702–1,706 were subjected to α -helical constraints (according to secondary structure prediction). The crystallographic structure of the human LRRK2 Roc domain (PDB ID code 2ZEJ, chain B) was used as an additional template to model the Roc domain (37). Because the dimeric form of human LRRK2 Roc from the 2ZEJ crystal might be an artifact caused by domain swapping (36), only residues 1,335–1,356 and 1,412–1,512 (i.e., those parts not involved in the dimer interface) have been taken into account for model building.

The Roc domain is missing from chain B of *Chlorobium tepidum* RocCOR (PDB ID code 3DPU) structure (36). Hence, to build a fully dimeric LRRK2 RocCOR domain, the monomeric template derived from chain A of the dimeric *C. tepidum* Roco protein was fitted on chain B by superimposing the C-terminal COR domains (amino acids 790–938). The use of a dimeric RocCOR template was indeed instrumental to model the RocCOR dimer interface properly.

For each template, 200 models were generated by randomizing all of the Cartesian coordinates of standard residues in the initial model. The 20 models with the lowest number of stereochemical

constraint violations have been selected and ranked according to the DOPE score (70). The best 10 models were validated for their stereochemical quality through the Molprobtity tool (71) available in the Phenix software package (72). The top-quality model was finally subjected to 15 cycles of backbone-restrained symmetrical relaxation (79) through the Rosetta FastRelax protocol (73) available in Rosetta3.5 (74). Noncrystallographic (point) symmetry definitions were generated starting from the modeled dimer and taking chain A as the input for symmetrical relaxation.

LRR domain. The structural model of the LRR domain of human LRRK2 protein (UniProt ID code Q5S007, residues 943–1,327) was obtained by homology modeling through the threading protocol (80) available in Rosetta3.5 (74). The structure of the LRR domain of the *C. tepidum* Roco protein (PDB ID code 5IL7, chain A), published along with this paper (crystallographic details are provided in Table S1), was used as a template for comparative modeling calculations. Two alternative alignments of the LRRK2 and *C. tepidum* LRR sequences were generated: a global one, obtained through EMBOSS Stretcher (81) using a BLOSUM30 matrix and a local one obtained from BLAST (78). The former alignment is characterized by 28.05% of identity, whereas the latter is characterized by 29.61% of identity. A search for suitable templates carried out with HHpred (82) showed no template with sequence identity greater than 26% (PDB ID code 3RGZ, as of August 2014), confirming that the *C. tepidum* LRR crystallographic structure is the best template to date to model LRRK2's LRR structure. This finding is further supported by the observation that none of the templates identified by HHpred contained a C-terminal helix suitable to model the characteristic α -helix (which, in LRRK2, is predicted to reside on amino acids 1,315–1,326) and its N-terminally preceding linker. In the *C. tepidum* LRR template structure, the last C-terminal LRR motif is peculiar in that it is characterized by an additional β -strand insertion, creating a β -strand-turn- β -strand-turn motif that precedes the C-terminal α -helix region. The parallel β -strands of this unit have the typical orientation of a parallel β -sheet, which is different from the orientation of the preceding LRR β -sheet.

Both starting alignments indicate divergence between LRRK2 and *C. tepidum* LRR sequences. In particular, BLAST alignment suggests that the C-terminal β -strand-turn- β -strand-turn motif might be absent in the LRRK2 LRR domain and that only one β -strand is likely present.

For each alignment, 10,000 starting models were generated by threading the target sequence onto the template. The gaps in unaligned regions were closed by loop modeling through the cyclic coordinate descent method (83), using three and nine amino acid fragments obtained from the Robetta web server (84). Finally, all-atom refinement was carried out for each model through Rosetta FastRelax (72).

To select the best representative loop conformation, clustering of the best 10% conformers was performed through the QT algorithm from Wordom (76), assessing structure similarity through $C\alpha$ -RMSD with a 3-Å cut-off. The model from the largest cluster, with the highest Rosetta score and stereochemical quality assessed through Molprobtity (71), was retained as the best representative structure.

Homology Modeling of LRRK2 Armadillo, Ankyrin, and WD40 Domains. Suitable templates for the armadillo (62–655), ankyrin (688–863), and WD40 (2,139–2,527) domains were identified through the HHpred web server (82).

To model the LRRK2 armadillo domain (amino acids 62–655), the following templates were used: PDB ID codes 4EV8 (chain A) and 2Z6G (chain A) (alignment is shown in Fig. S6). The following parts were removed from the templates for proper modeling of sequence insertions: PDB ID code 4EV8: 395–397

and PDB ID code 2Z6G: 394–395. Residues 108–124, 327–336, 339–344, 382–390, and 525–539 were subjected to α -helical constraints (according to secondary structure predictions).

The LRRK2 ankyrin domain (amino acids 688–863) was modeled using the structure PDB ID code 4UUC (chain A; final alignment is shown in Fig. S6) as a template. The following parts were removed from the templates for proper modeling of sequence insertions: 142–143 and 223–224.

The LRRK2 WD40 (amino acids 2,139–2,527) was modeled using the following template structures: PDB ID code 1VYH (chain C) deleted on residues 118–120, 160–162, 181–200, 264–266, and 274–277 and PDB ID code 3ZWL (chain B) deleted on residues 30–32, 72–74, 113–115, 141–143, 236–238, 246–248, and 310–312.

For each domain, 100 models were generated by randomizing all of the Cartesian coordinates of standard residues in the initial model. The 20 models with the lowest number of stereochemical constraint violations were selected and ranked according to the DOPE score (70). The best 10 models were validated for their stereochemical quality through the Molprobtity tool (71) available in the Phenix software package (72). The top-quality model was finally subjected to 15 cycles of backbone restrained relaxation (73).

Integrative Modeling.

Filtering of structural models based on cross-links fit. To assess the reliability of each structure/complex model, a simple metric was devised that ranks a given structure based on cross-link-related distances. To this end, a score (s_{ij}) is assigned to each cross-link pair between sites i and j through the following criteria:

$$s_{ij} = \begin{cases} 1 & \text{if } d_{ij} \leq \text{cutoff} \\ \text{else} & \\ 1 - \alpha_{ij} & \end{cases}, \quad [S1]$$

where d_{ij} is the $C\alpha$ - $C\alpha$ distance, computed through Wordom (76); cut-off stands for the distance-cut-off set to 35 Å and 31 Å for DSS and DSG, respectively; and α_{ij} is the normalized deviation ($d_{ij} - \text{cut-off}$) of the corresponding cross-linked residues i and j . As a result, sites within the cross-link distance thresholds are maximally rewarded, whereas those sites exceeding the cut-off are penalized proportionally to the distance.

A given model/complex is classified by the total score (XL^{fit}) given by the sum of the individual score terms of each cross-link.

Sequence connectivity of the various domains was taken into account by adding a scoring term inversely proportional to terminal distances. Given the ambiguity of the cross-link data and the dimeric nature of LRRK2, all of the possible distance combinations on the two dimer chains have been measured and, for each pair, only the smallest distance (or deviation) was considered in the final score calculation.

Modeling of the LRR-RocCOR assembly. The integrative modeling of the LRRK2 quaternary structure was started by joining the top structural models of the LRR domain to the dimeric RocCOR model. The α -helices, present in both domain models, were instrumental to dock and preorient the LRR domain manually with respect to RocCOR. In detail, a junction point between the two domains was created by optimally superimposing the backbone atoms of residues 1,319–1,320 on both RocCOR dimer chains. Noncrystallographic (point) symmetry definitions were generated starting from the manually modeled LRR-RocCOR dimer and taking chain A as the input for 15 cycles of symmetrical FastRelax (73, 79). Multiple conformers of the dimeric, symmetrical LRR-RocCOR complex were obtained by running 1,000 independent FastRelax calculations, both unrestrained as well as distance-restrained (for a total of 2,000 structures). In the latter case, cross-link-derived distance constraints were modeled

by using Rosetta flat harmonic potentials defined by the following distance (d)-dependent function $f(d)$:

$$f(d) = \begin{cases} 0 & \text{if } d \leq t + x_0 \\ \text{else} & \\ \left(\frac{d - x_0 - t}{\sigma} \right)^2 & \end{cases}, \quad [\text{S2}]$$

where x_0 is the equilibrium distance (15 Å and 13 Å for DSS and DSG, respectively), t is the tolerance (15 Å and 13 Å for DSS and DSG, respectively), and σ is an empirical weighting factor given by:

$$\sigma = \left(\frac{\max(\text{Id score})}{\text{Id score}} \right) \times \left(\frac{1}{\text{Freq}} \right), \quad [\text{S3}]$$

where *Id score* is the top score found for a given cross-link pair and *Freq* is the number of times a given cross-link is found on the samples. The use of cross-link *Id scores* and recurrences was first proposed for the implementation of Bayesian scoring functions in recent integrative modeling efforts (64). The maximum value of 30 Å as a cut-off to model DSS-related distances agrees with the maximal value suggested by thorough analysis of cross-link-related distances from large molecular dynamics trajectories datasets (85).

The top 50% of solutions, ranked according to the Rosetta full-atom scoring function, were reranked according to the XL^{fit} score.

The first three top conformations were subjected to the Rosetta Floppy Tail (40) approach, which was adapted to sample the possible rigid-body-like orientations of the LRR domain exhaustively with respect to the RocCOR one, while keeping them covalently attached through the $\alpha 0$ -helix. To this end, the conformation of the linker region that N-terminally precedes the $\alpha 0$ -helix (i.e., residues 1,308–1,315) was sampled through small, shear, and fragment insertion moves in centroid mode, followed by small and shear perturbations and then by side-chain repacking to refine final conformations. As a result, backbone conformational changes of the linker region are propagated downstream through the internal Rosetta coordinate system (i.e., foldtree), resulting in rigid-body rototranslations of the RocCOR domain relative to the LRR one. Additionally, side-chain repackings were performed for residues surrounding the linker region, as well as for newly formed interfaces resulting from rigid-body rototranslational movements of the RocCOR relative to the LRR. For each of the starting LRR-RocCOR conformers, 1,000 unrestrained and 1,000 cross-link distance-restrained Floppy Tail calculations were carried out by using default parameters for the production run and three amino acid fragments obtained from the Robetta web server (84). Predictions were performed only on chain A while keeping the symmetrical RocCOR chain B in the input structure to avoid the LRR domain sampling unrealistic conformations with respect to RocCOR. Symmetrical LRR-RocCOR conformers were generated by fitting a copy of chain A on the RocCOR chain B.

For each run, the top 50% solutions, ranked according to the Rosetta full-atom scoring function, were reranked according to

the XL^{fit} score. The top 50% scoring models were finally retained for clustering with the QT algorithm from Wordom (76), measuring similarity through C α -RMSD with a 7.5-Å cut-off.

The cluster representatives that fitted both the cross-link datasets and the negative-staining EM map well were selected and subjected to final refinement through Rosetta FastRelax (73).

Fitting in the negative-staining EM map was carried out through UCSF Chimera (42) software, using 50 cycles of the map in the map global search algorithm.

Ankyrin, kinase, and WD40 domains assembling through integrative rigid body docking and cross-link-derived distance filters. Quaternary structure predictions of the ankyrin, kinase, and WD40 domains relative to the LRR-RocCOR assembly were carried out through ZDOCK rigid-body docking calculations (60). Hierarchical docking was used to determine realistic complexes for the ankyrin and kinase domains, which were, in turn, used for subsequent docking of the WD40 domain. The final top ankyrin-kinase-WD40 ternary complexes were finally used for rigid-body docking on the previously selected top LRR-RocCOR conformers.

For each stage, global docking calculations were carried out in a dense sampling mode (i.e., using a rotational sampling of 6°). For each run, the top 50% of the best 4,000 solutions were retained and reranked according to the XL^{fit} score. The top 50% scoring models were finally clustered with Wordom (76) using a C α -RMSD cut-off of 7.5 Å.

In each stage, the representative complexes for a cluster that best satisfied cross-link-derived distance filters were retained and used for subsequent docking stages. By this strategy, a total of 12 ankyrin-kinase-WD40 complexes were selected and used for docking to alternative dimeric LRR-RocCOR conformers. The same docking and filtering procedure as above was applied to select a pool of plausible ankyrin-LRR-RocCOR-kinase-WD40 conformers. To recover symmetry for each pose, a copy of docked ankyrin-kinase-WD40 ternary domain complex was fitted by chain swapping of the symmetrical LRR-RocCOR dimer, using ad hoc BioPython (86) scripts. Only those cluster representatives with more than 50% of satisfied cross-link distances and with an EM negative-stain map correlation value greater than 0.8 were retained for further inspection.

Armadillo domain assembling and LRRK2 quaternary structure refinement. Given the scarcity of available cross-links for the armadillo domain (only two DSSs and one DSG found between the armadillo and ankyrin domains), its relative orientation was determined by independently fitting two copies of the armadillo model structure into the EM map and selecting those poses that best fitted the density in correspondence of the N-terminal region (i.e., the one that could be assigned to the armadillo domain).

The prefitted armadillo domains were integrated to the top ankyrin-LRR-RocCOR-kinase-WD40 complexes determined in the integrative docking stage. The whole dimeric LRRK2 quaternary structure models were refined on the EM map through 200 cycles of sequential refitting within the UCSF Chimera (42) fitmap algorithm. The sequential refitting was carried out by keeping each block internally rigid through the following order: LRR-RocCOR (chains A and B), ankyrin-kinase-WD40 (chain A), ankyrin-kinase-WD40 (chain B), armadillo (chain A), and armadillo (chain B).

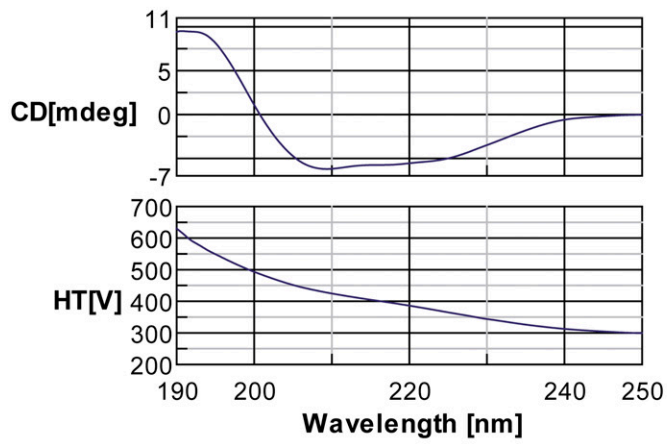
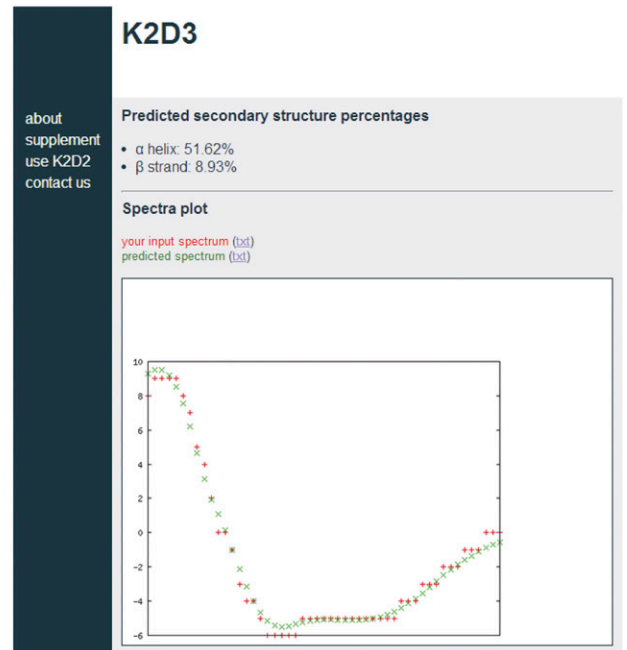
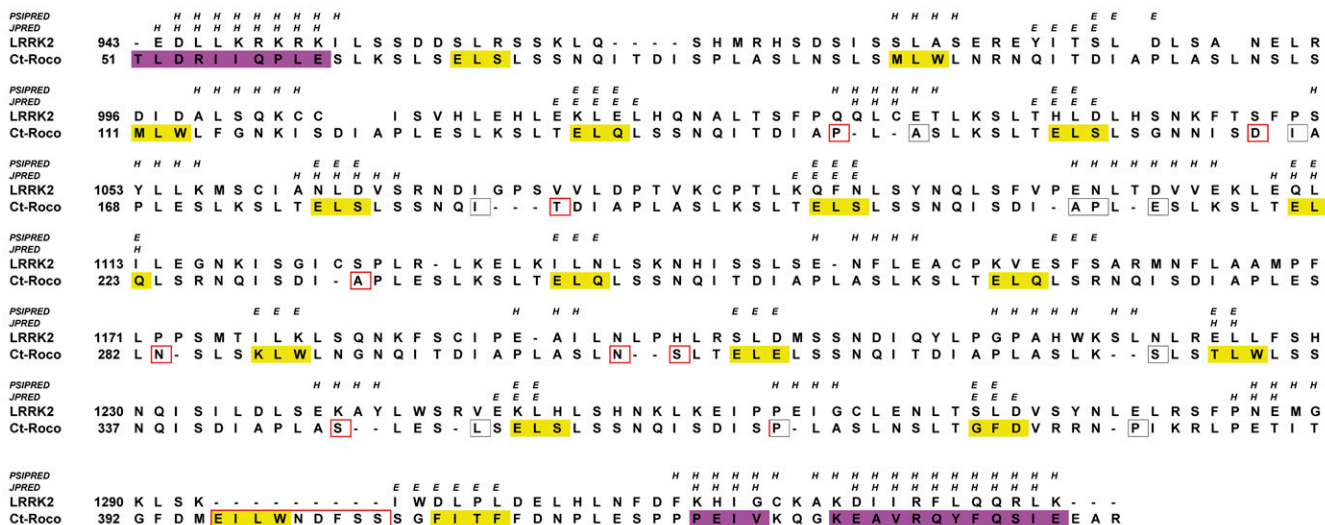
A**B**

Fig. 51. (A) LRRK2 CD far-UV spectrum acquired using a 1-mm path length quartz cuvette. HT[V], high tension [volts]; mdeg, millidegree. (B) LRRK2 α -helix and β -sheet content estimation using the K2D3 web server (87). The wavelength range used was 240–190 nm.

A

LRR domain: alignment 1 (a.a. 943-1327) EMBOSS Stretcher BLOSUM30 /BLAST C-ter mixed

Identity: 124/459 (28.0%)



B

LRR domain: alignment 2 (a.a. 943-1327) EMBOSS Stretcher BLOSUM30/BLAST N-ter/C-Ter/mixed

Identity: 124/459 (30%)

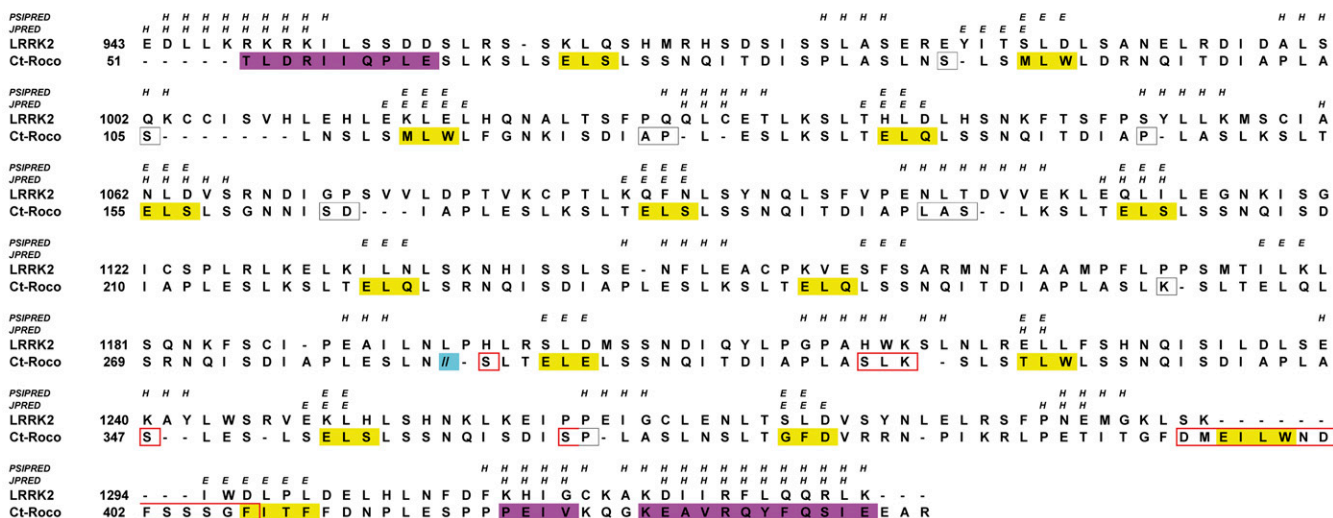


Fig. S7. Alignments used for homology modeling. (A) Alignment of the LRRK2 LRR domain (amino acids 943–1,327) to the *C. tepidum* Roco LRR domain template structure PDB ID code 5IL7 (alternative 1). (B) Alignment of the LRRK2 LRR domain (amino acids 943–1,327) to the *C. tepidum* Roco LRR domain template structure PDB ID code 5IL7 (alternative 2). A secondary structure-specific coloring scheme is used for the template sequences, where α -helices and β -strands are colored in violet and yellow, respectively. Amino acids deleted from the template are highlighted in red boxes. For the LRRK2 sequences, the PsiPRED and JPRED secondary structure predictions are also reported. BLOSUM30, Blocks Substitution Matrix 30; EMBOSS, The European Molecular Biology Open Software Suite.

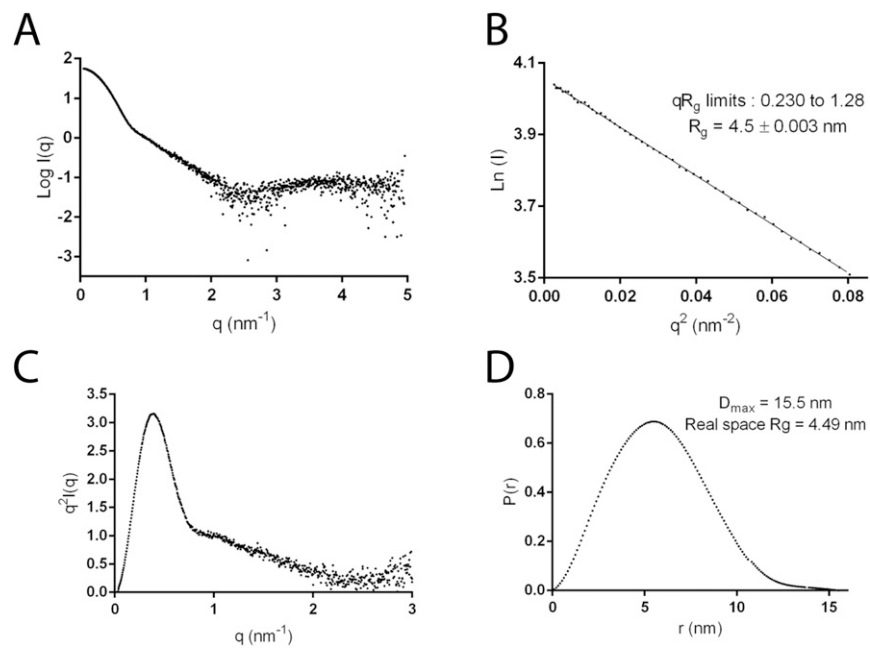


Fig. S10. SAXS analysis of the *C. tepidum* LRR-Roc-COR construct. (A) Experimental scattering curve. (B) Guinier analysis. (C) Kratky plot. (D) Pair-distance distribution function $P(r)$. D_{max} , maximum particle dimension; I , scatter intensity; q , scattering vector.

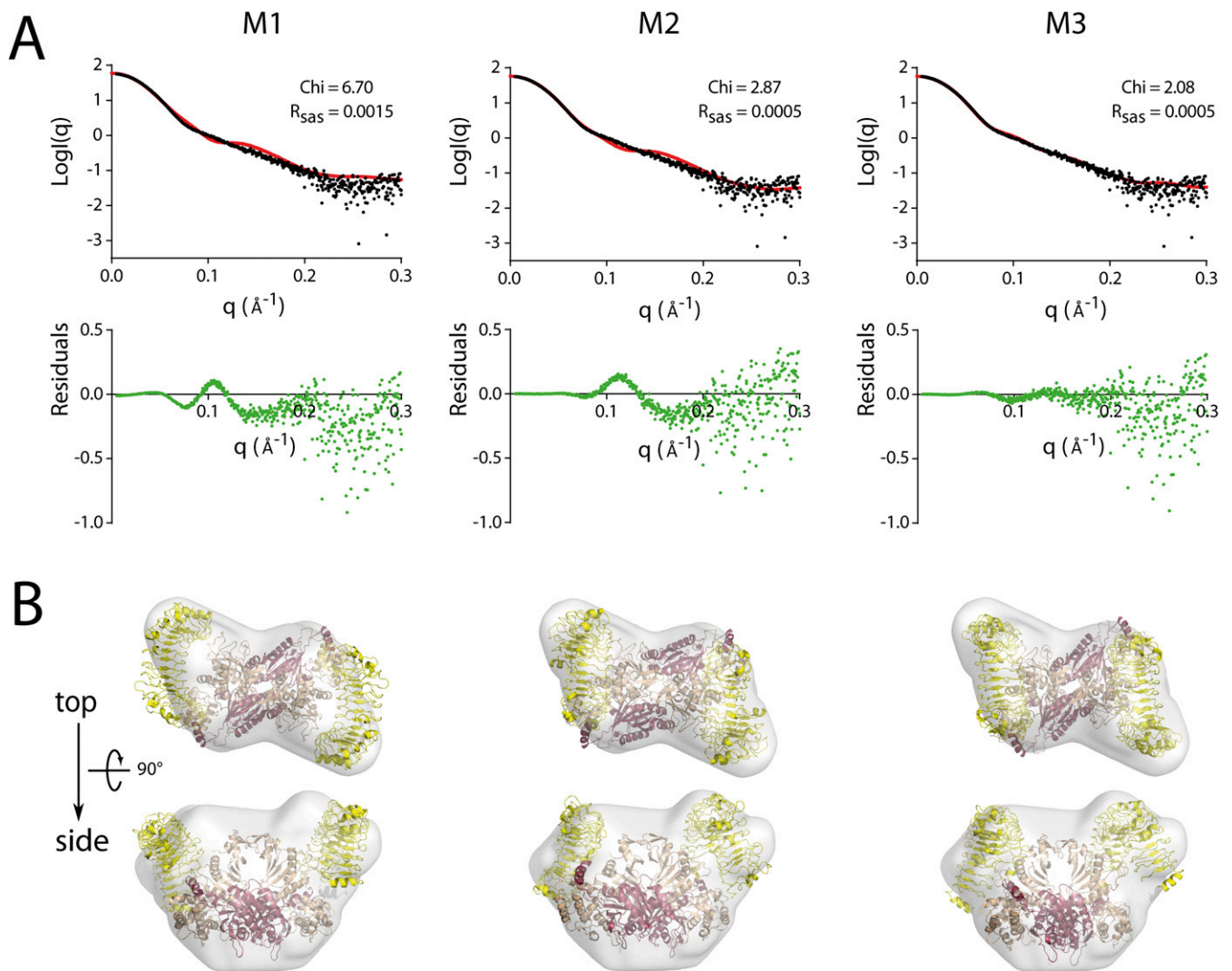


Fig. S11. (A) Fit of the experimental SAXS data of the *C. tepidum* LRR-Roc-COR protein (in black) on the theoretical scatter curves (red lines) corresponding to the LRR-Roc-COR conformers M1, M2, and M3 using the program CRYSOLOG. The residuals of the fit are shown below (green). (B) Docking of the LRR-Roc-COR models M1–M3 in the ab initio averaged bead model envelope calculated from the SAXS data with DAMMIF (average of 20 models).

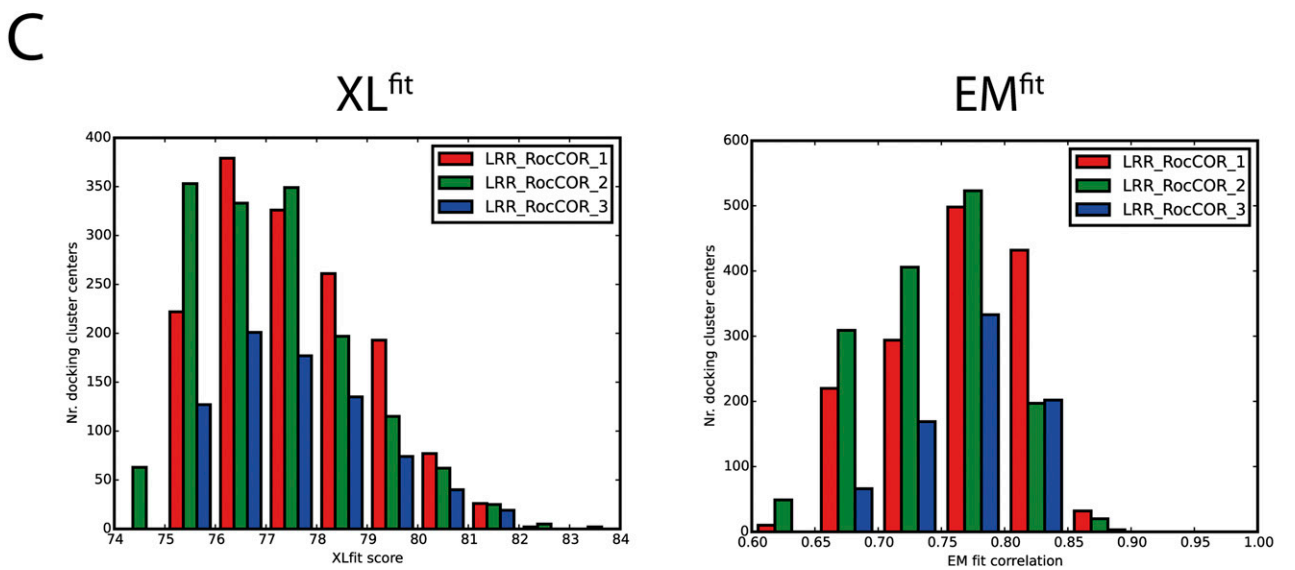
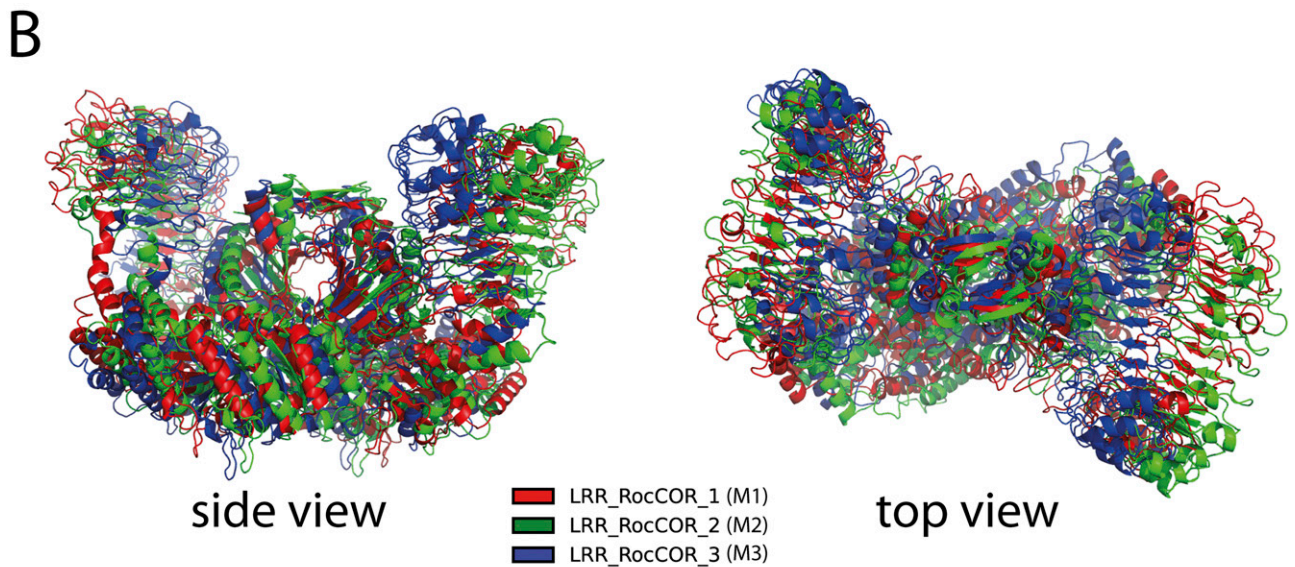
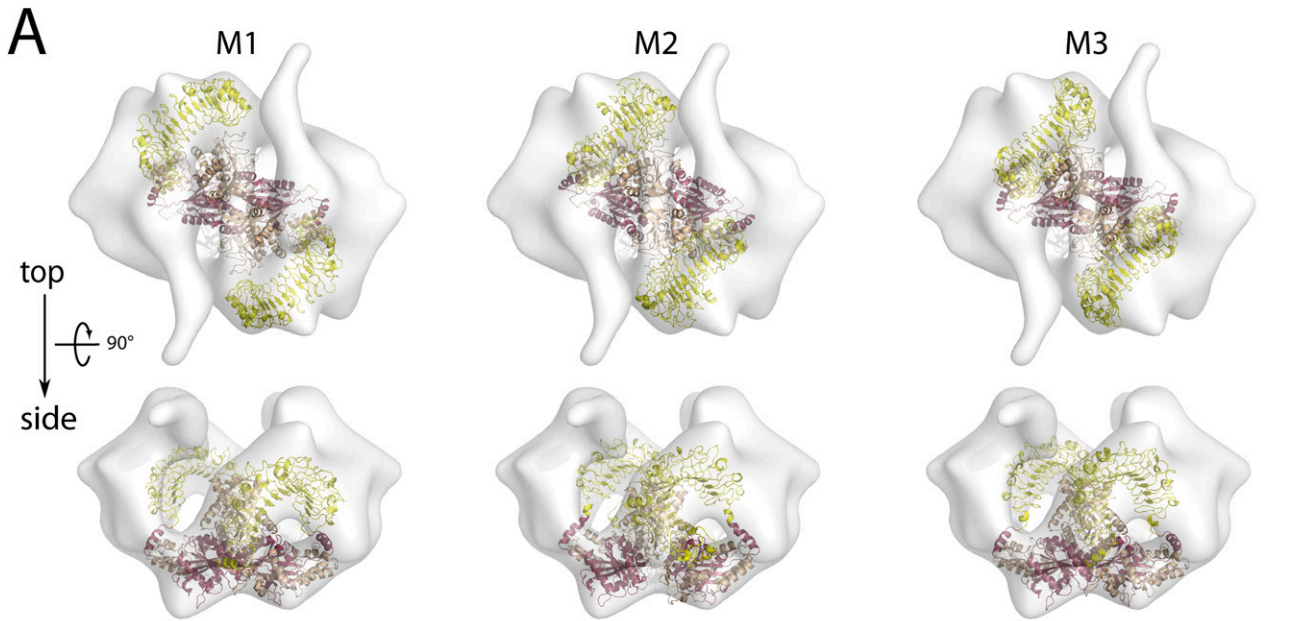


Fig. S12. (A) Docking of the top-scoring LRR-RocCOR conformers to the EM map (M1: $EM^{fit} = 0.864$, M2: $EM^{fit} = 0.856$, and M3: $EM^{fit} = 0.841$). (B) Superimposition of alternative LRR-Roc-COR docking solutions. (C) XL^{fit} score (Left) and EM^{fit} (Right) histograms for three alternative LRR-Roc-COR docking solutions that have been selected for subsequent modeling steps.

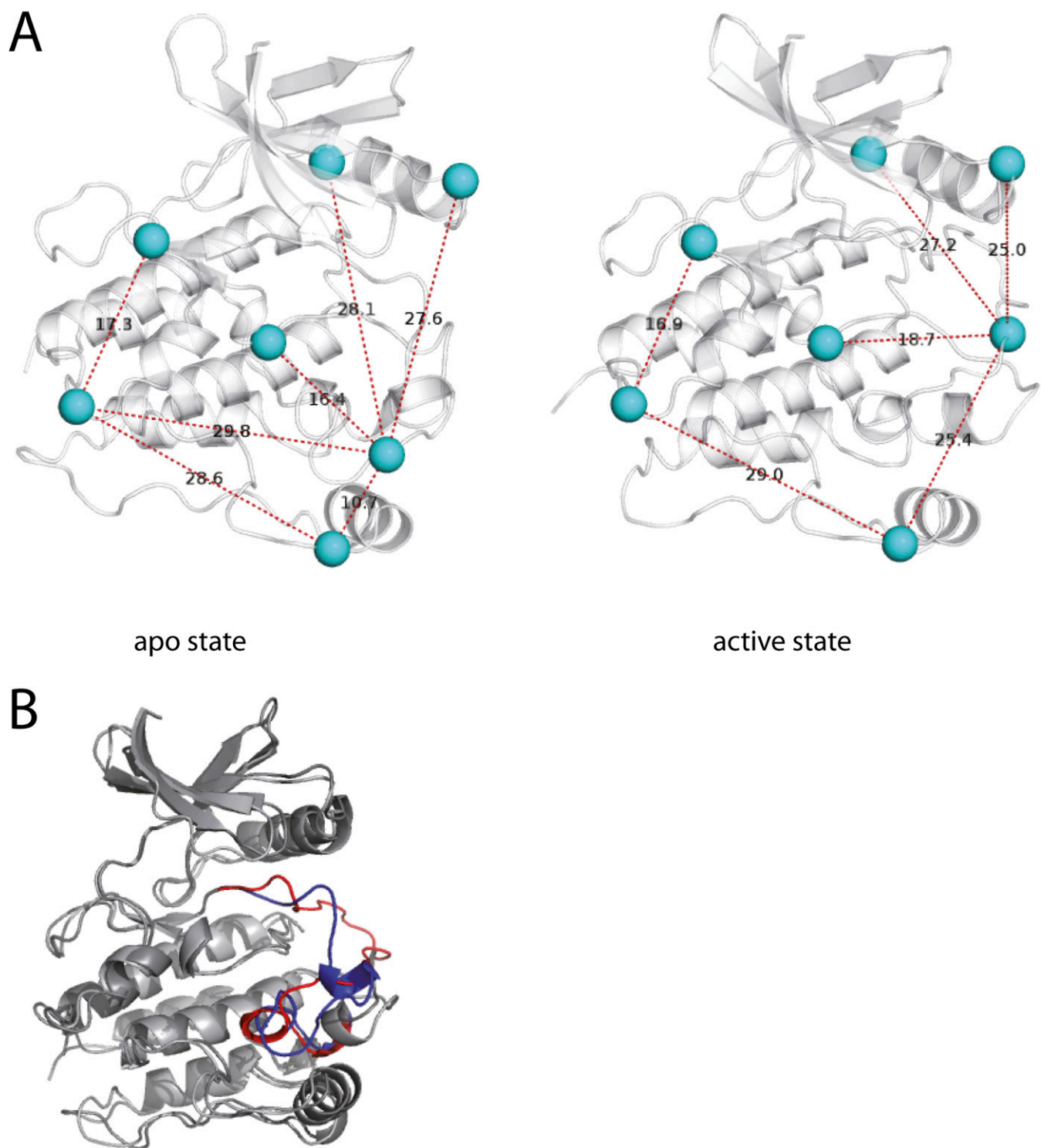


Fig. S13. (A) Kinase domain models for the apo conformation as well as the active conformation. (B) Superimposed models. The different conformations of activation segments are highlighted: apo state (blue) and active state (red).

Table S1. Statistics for the crystallographic analysis of the LRR domain of the *C. tepidum* Roco protein (PDB ID code 5IL7)

Variable	Value
Space group	P 21, 21, 21
Cell constants (Å)	53.22, 98.94, 155.40
a, b, c, α , β , γ	90°, 90°, 90°
Resolution, Å	47.14–2.30
Data completeness (in resolution range), %	99.8 (47.14–2.30)
$\langle I/\sigma(I) \rangle$	4.12 (at 2.29 Å)
R , R_{free}	0.21, 0.26
Wilson B-factor	25.6
Outliers	0 of 37,250 reflections
Total no. of atoms	6,871
Average B-factor	30.0

Dataset S1. List of DSS cross-links

[Dataset S1](#)

The dataset reports the XQuest/XProphet analysis according to Leitner et al. (25).

Dataset S2. List of DSG cross-links

[Dataset S2](#)

The dataset reports the XQuest/XProphet analysis according to Leitner et al. (25).

Dataset S3. Top-scoring models for LRR-Roc-COR M1

[Dataset S3](#)

The dataset reports the name of the rank, model (PDB file), EM^{fit} score (EM map correlation value), total number of satisfied cross-links, number of satisfied DSS cross-links, number of satisfied DSG cross-links, accumulated terminal distance value (terminal distance) of consecutive domain terminals (in angstroms), cross-link fit score (XL^{fit} score), and surface accessibility score (SASA) considering identified monolinks.

Dataset S4. Top-scoring models for LRR-Roc-COR M2

[Dataset S4](#)

The dataset reports the name of the rank, model (PDB file), EM^{fit} score (EM map correlation value), total number of satisfied cross-links, number of satisfied DSS cross-links, number of satisfied DSG cross-links, accumulated terminal distance value (terminal distance) of consecutive domain terminals (in angstroms), XL^{fit} score, and SASA considering identified monolinks.

Dataset S5. Top-scoring models for LRR-Roc-COR M3 best satisfying the SAXS data for the *C. tepidum* LRR-Roc-COR protein

[Dataset S5](#)

The dataset reports the name of the rank, model (PDB file), EM^{fit} score (EM map correlation value), total number of satisfied cross-links, number of satisfied DSS cross-links, number of satisfied DSG cross-links, accumulated terminal distance value (terminal distance) of consecutive domain terminals (in angstroms), XL^{fit} score, and SASA considering identified monolinks.

Dataset S6. Distance data used for scoring of the models based on conformer M1

[Dataset S6](#)

The dataset reports the distances between cross-linked lysine residues (for either DSS or DSG) and distances of consecutive domains (terminal distances). The shortest distances have been considered for scoring, and those distances below the cross-linked (either DSS or DSG)-specific distance cut-off are highlighted in red.

Dataset S7. Distance data used for scoring of the models based on conformer M2

[Dataset S7](#)

The dataset reports the distances between cross-linked lysine residues (for either DSS or DSG) and distances of consecutive domains (terminal distances). The shortest distances have been considered for scoring and those distances below the cross-linked (either DSS or DSG)-specific distance cut-off are highlighted in red.

Dataset S8. Distance data used for scoring of the models based on conformer M3

[Dataset S8](#)

The dataset reports the distances between cross-linked lysine residues (for either DSS or DSG) and distances of consecutive domains (terminal distances). The shortest distances have been considered for scoring and those distances below the cross-linked (either DSS or DSG)-specific distance cut-off are highlighted in red.

Dataset S9. Experimental and modeling parameters for the SAXS analysis of the *C. tepidum* Roco protein

[Dataset S9](#)

Data-collection parameters:	C.t. LRR-RocCOR NF
Beamline:	BM29, ESRF, Grenoble
Wavelength, (Å):	0.99
s range, (Å ⁻¹):	0.032 - 0.498
Concentration (mg/ml):	8
Temperature, (K):	298.15
Structural parameters:	
I(0) (cm ⁻¹) [from p(r)]:	57.70 ± 0.02
R _g (Å) [from p(r)]:	44.90 ± 0.03
I(0) (cm ⁻¹) (from Guinier):	57.73 ± 0.02
R _g (Å) (from Guinier):	45.00 ± 0.03
D _{max} (Å):	155
Porod volume estimate, V _p (Å ³):	381650
Molecular-mass determination:	
Molecular mass M _r (kDa) from Porod volume (V _p /1.7):	225
Calculated M _r (kDa) from sequence (monomer):	110
Modeling parameters:	
Shape reconstruction	DAMMIF
Symmetry	P2
# of models averaged/total	20
DAMAVER NSD (var)	1.182 ± 0.159
Software employed:	
Data evaluation:	PRIMUS, GNOM
Computation of model intensities:	CRY SOL
Three-dimensional graphics representations:	UCSF Chimera