Jaypee Institute of Information Technology

Basic Bioinformatics Lab
Project-based learning (PBL)

Structural Modeling and Analysis of the LRRK2 N-Terminal Region

SUBMITTED TO- DR. NIDHI JATNA BATRA
Course- MS. BI

Submitted by - Divyanshu Saklani - 2509280010
Arti Koul - 2509280015

Teacher's Sign _____

Acknowledgments

I would like to express my sincere gratitude to my project supervisor, Dr. Nidhi Batra Ma'am, and Dr. Reetika Debroy Ma’am, for their continuous guidance, patience, and motivation throughout the course of this research. Their insights into the Fundamentals of bioinformatics were invaluable in shaping the direction of this study.

I am also grateful to the Department of Biotechnology and Bioinformatics and Jaypee Institute of Information Technology for providing the necessary computational resources and academic environment to conduct this work.

This project relies heavily on open-source bioinformatics tools and databases. I would like to acknowledge the AlphaFold team(DeepMind) for making advanced protein prediction accessible, and the UniProt and RCSB Protein Data Bank (PDB) consortia for maintaining the essential datasets that made this evolutionary and tree analysis possible.

I would also like to thank to colleagues/lab partners for their helpful discussions and feedback during the data analysis phase.

Finally, I must express my profound gratitude to my parents and family for their unfailing support and continuous encouragement throughout my years of study.

[19-11-25]

Table of Contents

List of Figures

Abstract

The *Leucine-rich Repeat Kinase 2* (LRRK2) gene has emerged as a central player in the pathology of Parkinson's disease (PD). Encoding a large, multi-domain protein with dual kinase and GTPase activities, LRRK2 is critical for regulating intracellular trafficking, cytoskeletal support, autophagy, and inflammation. Pathogenic mutations, particularly G2019S, R1441C/G/H, and Y1699C, disrupt these regulatory functions, leading to dopaminergic neurodegeneration and endolysosomal dysfunction. Despite its clinical importance, the structural characterization of LRRK2 remains incomplete. Specifically, the N-terminal region—a massive scaffolding unit comprising Armadillo, Ankyrin, and Leucine-Rich repeats—is intrinsically flexible and has historically evaded high-resolution experimental determination (X-ray crystallography and Cryo-EM).

This study addresses the challenge of modeling this "missing" N-terminal region by evaluating disparate computational strategies. We first attempted homology modeling using **SWISS-MODEL**, which relies on BLAST and HHblits algorithms to project a target sequence onto an experimentally solved template. This approach failed due to the absence of a complete structural template in the Protein Data Bank (PDB). We subsequently evaluated **I-TASSER**, a threading-based method that utilizes the LOMETS algorithm to identify structural analogs and Monte Carlo simulations for refinement. However, threading proved insufficient for the large, unique, and non-globular repeats of the LRRK2 N-terminus.

Consequently, we selected **AlphaFold**, an *ab initio and hybrid* method that represents a paradigm shift in structural biology. Unlike template-based methods, AlphaFold utilizes a deep learning architecture (Evoformer) and attention mechanisms to predict 3D coordinates directly from evolutionary co-variation patterns in Multiple Sequence Alignments (MSA). Our results confirm that AlphaFold successfully predicts the full-length LRRK2 structure, where traditional methods failed. The resulting model reveals the N-terminus as a dynamic, flexible scaffold, providing critical structural context for understanding how this region functions as an autoinhibitory "lid" in the regulation of LRRK2 kinase activity.

1. Introduction

1.1 Background: Parkinson's Disease and Its Relation to LRRK2

Parkinson's disease is a progressive neurodegenerative disorder that primarily affects movement. It occurs when dopamine-producing neurons in the substantia nigra (a region of the midbrain) gradually die. Dopamine is essential for smooth, coordinated movements, so its loss leads to classic symptoms like tremor, rigidity, slow movement, and balance issues.

PD also involves non-motor symptoms such as sleep disturbances, depression, anxiety, constipation, and changes in smell sometimes appearing years before motor symptoms.

- **Types of Parkinson's Disease**

i. Idiopathic (Sporadic) Parkinson's Disease

Most common form (85–90%).

1. No single identifiable cause.
2. Thought to result from a mix of aging, environmental exposures (pesticides, toxins), and small genetic vulnerabilities.

ii. Genetic/Familial Parkinson's Disease

5–15% of cases.

1. Caused by inherited mutations in specific genes:
    - **LRRK2**
    - **SNCA (α-synuclein)**
    - **PARK2 (Parkin)**
    - **PINK1**
    - **DJ-1**

iii. Secondary or Atypical Parkinsonism

(Not true Parkinson's disease but produces similar symptoms.)

1. Caused by stroke, brain injury, medications, or toxins.
2. Includes disorders like:
    - Multiple System Atrophy (MSA)
    - Progressive Supranuclear Palsy (PSP)
    - Corticobasal Degeneration (CBD)
    - Drug-induced parkinsonism

- **How LRRK2 Is Linked to Parkinson's Disease**

i. LRRK2 Mutations Are One of the Most Common Genetic Causes

Mutations in the **LRRK2 gene** especially **G2019S**, **R1441C/G/H**, and **Y1699C**—are among the leading causes of familial Parkinson's disease and also appear in some sporadic cases.

The G2019S mutation is the **single most frequent genetic cause of PD worldwide.**

ii. What LRRK2 Normally Does

The LRRK2 protein is involved in:

1. Vesicle trafficking
2. Autophagy (cellular cleanup)
3. Cytoskeletal stability
4. Mitochondrial maintenance
5. Immune signaling
6. Rab GTPase phosphorylation (important for intracellular transport)

It acts as both:

1. **a kinase** (adds phosphate groups)
2. **a GTPase** (regulates molecular switches)

Pathogenic LRRK2 mutations generally **increase its kinase activity**, making it hyperactive. This leads to:

a. Disrupted cellular trafficking

LRRK2 over-phosphorylates Rab proteins, disturbing endolysosomal pathways, which are essential for clearing damaged proteins and recycling cell components.

b. Mitochondrial stress

Mutant LRRK2 impairs mitochondrial function, leading to energy deficits and oxidative stress—major contributors to neuron death.

c. Autophagy impairment

Poor clearance of damaged proteins may promote accumulation of **α-synuclein**, the hallmark protein that forms Lewy bodies in PD.

d. Inflammation

LRRK2 is highly active in microglia and immune cells. Mutant LRRK2 increases neuroinflammation, creating a toxic environment for neurons.

iii. The reason causing Dopamine Neurons to Die-:

The combination of:

1. mismanaged protein clearance
2. mitochondrial dysfunction
3. disrupted vesicle transport
4. and chronic inflammation

creates a damaging cascade. Dopamine neurons are especially vulnerable due to their high metabolic demands and large axonal networks.

1.2 The MissingN-Terminus

The N-terminus of LRRK2 forms the protein's regulatory foundation, containing armadillo, ankyrin, and leucine-rich repeat motifs that stabilize its conformation and enable proper protein–protein and protein–membrane interactions. When this N-terminal region is missing whether due to truncation, proteolytic cleavage, or functional collapse caused by mutations the overall structural integrity of LRRK2 is compromised. The protein becomes less anchored, less stable, and more vulnerable to misfolding. This loss of stability affects how LRRK2 positions itself within the cell and how effectively it communicates with vesicular, cytoskeletal, and autophagy-related pathways.

The absence of the N-terminus also disrupts the delicate regulatory balance between LRRK2's GTPase and kinase domains. Under normal conditions, the N-terminal structure restrains kinase overactivity by maintaining proper intramolecular alignment. Once this regulatory layer is removed, the kinase domain

tends to shift into a hyperactive state. Such kinase hyperactivation closely parallels what is seen in Parkinson's disease–linked mutations, particularly G2019S and R1441C/G/H. Hyperactive LRRK2 abnormally phosphorylates Rab GTPases, distorting vesicle trafficking, lysosomal recycling, and endosomal maturation. These defects ultimately impair the cell's ability to manage protein turnover and maintain neuronal health.

Another critical consequence of losing the N-terminal portion is the weakened handling of α-synuclein. LRRK2 is deeply involved in autophagy and lysosomal pathways, and without its N-terminal scaffolding, the protein becomes inefficient at coordinating membrane dynamics required for clearing misfolded α-synuclein. This insufficiency promotes the accumulation of toxic aggregates and increases oxidative stress. Dopaminergic neurons—already metabolically fragile—begin to degenerate under the combined pressure of mitochondrial dysfunction, impaired intracellular transport, and chronic inflammation. Thus, the loss of N-terminal integrity transforms LRRK2 into a destabilized, overactive protein that aligns closely with the molecular patterns observed in Parkinson's disease.

1.3 Failure of Homology Modeling (SWISS-MODEL / MODELLER)

Homology modeling operates on the principle of "copying" a solved structure. It requires a high-quality experimental template (typically X-ray crystallography or Cryo-EM) with significant sequence identity (>30%) to the target protein.

- **The LRRK2 Obstacle:** The N-terminal region of LRRK2 (residues ~1–1200) is composed of large Armadillo (ARM), Ankyrin (ANK), and Leucine-Rich Repeat (LRR) domains. Due to their intrinsic flexibility and "hinge-like" behavior, these regions are often unresolved in experimental studies.
- **The Result:** When submitted to SWISS-MODEL or MODELLER, the algorithms search the Protein Data Bank (PDB) for a template. They find either **(a)** nothing covering the full N-terminus, or **(b)** only small, fragmented hits for generic repeat motifs. Without a continuous structural template to copy, homology modeling cannot generate a valid full-length model.

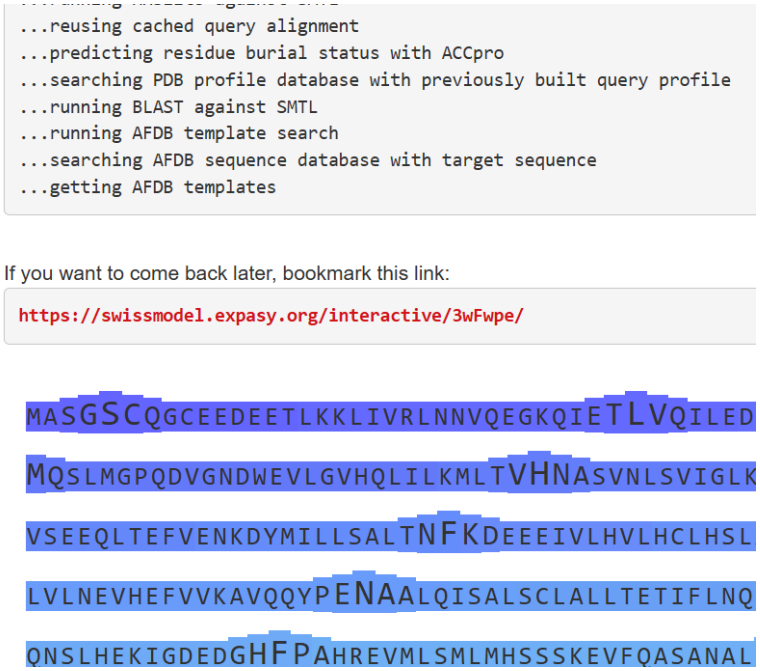


Fig. 1: Failure *of Homology modeling*

1.4. Inadequacy of Threading (I-TASSER)

Threading (or fold recognition) is designed for proteins with low sequence identity but known shapes. It attempts to "thread" the target sequence onto a library of known protein folds (e.g., "Does this sequence fit into a globin fold?").

- **The LRRK2 Obstacle:** The LRRK2 N-terminus is not a standard globular protein; it is a massive, elongated solenoid scaffold. Its unique tertiary structure—wrapping around the catalytic core—is highly specific to the LRRK2/ROCO family.
- **The Result:** I-TASSER's threading algorithms (LOMETS) struggle to find a single "fold" that matches this complex, multi-domain architecture. While it may correctly identify small substructures (like a single ankyrin repeat), it fails to assemble the correct global orientation of the domains relative to one another. This results in "spaghetti-like" models with poor packing and low confidence scores (C-scores).

1.5 The AlphaFold Solution

Unlike the methods above, **AlphaFold** is an *ab initio* method empowered by deep learning. It does not rely on "copying" a template or "recognizing" a known fold. Instead, it analyzes the **Multiple Sequence Alignment (MSA)** to detect evolutionary correlations (e.g., "When residue A changes, residue B always changes, so they must be touching in 3D space"). This allows it to predict the unique, flexible structure of the LRRK2 N-terminus from scratch, bypassing the limitations of the experimental PDB library.

1.6 Additional Supporting Points

1. A missing N-terminus can produce LRRK2 fragments that behave differently from the full-length protein, sometimes accumulating abnormally inside neurons.

2. Some studies show that stress-activated proteases can cleave LRRK2, generating natural N-terminal truncations that mimic PD-like toxicity.

3. N-terminal loss may reduce LRRK2's ability to bind to regulatory partners such as 14-3-3 proteins, further destabilizing its activity.

4. Truncated LRRK2 may respond differently to kinase inhibitors, creating challenges for targeted therapy in LRRK2-linked Parkinson's disease.

1.7 Key Modeling Takeaways

- **Homology Modeling (SWISS-MODEL) Failed:** Requires an experimental template to copy; none exists for the flexible LRRK2 N-terminus.

- **Threading (I-TASSER) Failed:** The N-terminus is a unique, massive scaffold that does not match any standard protein folds in the library.
- **AlphaFold Succeeded:** Uses deep learning and evolutionary history (MSA) to predict structure from scratch, bypassing the need for templates.
- **Biological Insight:** The "missing" structure is actually an **Intrinsically Disordered Region (IDR)**, acting as a flexible hinge for regulation.

2. Objectives

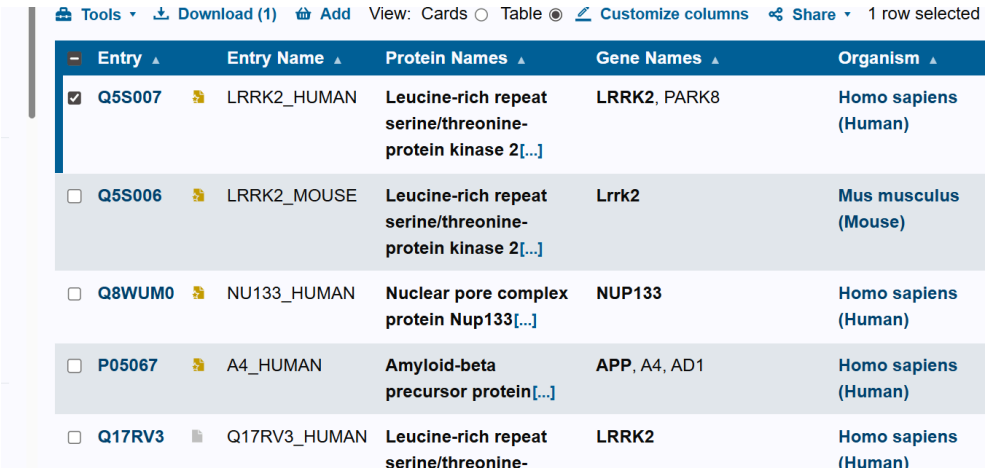
The lack of a complete experimental template makes traditional homology modeling (e.g., using MODELLER or SWISS-MODEL) impossible for this region. Therefore, this study aimed to:

1. Predict the 3D structure of the human LRRK2 N-terminus using *ab initio* deep learning methods (AlphaFold).
2. Analyze the structural model to identify regions of disorder and flexibility that correspond to its regulatory function.

3. Materials and Methods

3.1 Sequence Retrieval and Homolog Identification

The reference protein sequence for human LRRK2 was retrieved from the UniProt database (Accession: Q5S007). To identify our sequence is a fit for modeling, we conducted a BLASTp (Protein Basic Local Alignment Search Tool) search against the PDB database.



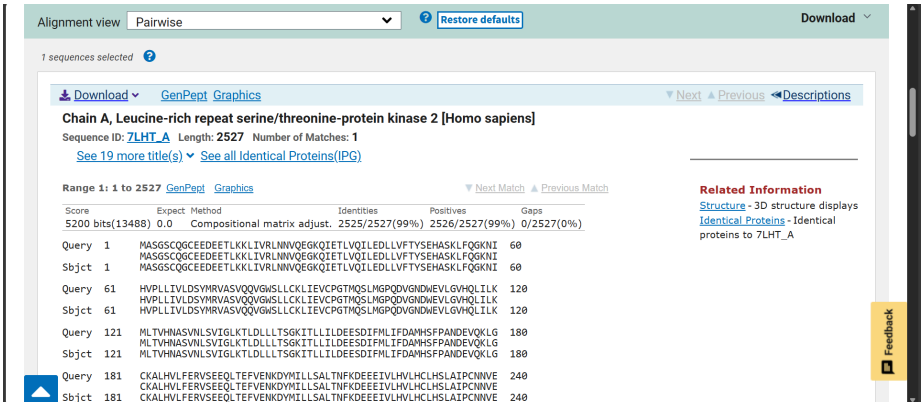


Fig. 2 Sequence retrieval and analysis

3.2 Structural Modeling (Secondary, Ab Initio, and AI predictive modelling)

Before 3D modeling, a comprehensive baseline characterization of the primary sequence was performed to validate protein stability and localization features.

- **Physicochemical Properties (ProtParam):** The Expasy **ProtParam** tool was used to compute essential parameters including molecular weight, theoretical isoelectric point (pI), instability index, and aliphatic index. This provided a quantitative assessment of the protein's stability and solubility in vitro.
- **Subcellular Localization (BUSCA): BUSCA** (Bologna Unified Subcellular Component Annotator) was employed to predict the subcellular localization and identify potential signal peptides or transmembrane domains. This step was critical to confirm that the N-terminus does not contain hidden membrane-anchoring motifs that would require specific modeling constraints.
- **Secondary Structure Prediction (SOPMA): SOPMA** (Self-Optimized Prediction Method with Alignment) was used to generate a consensus prediction of 2D structural elements (alpha helices, beta sheets, and random coils). This served as a "ground truth" benchmark to later validate the secondary structure content of the generated 3D AlphaFold model.

Due to the absence of a complete experimental template for the LRRK2 N-terminus in the Protein Data Bank (PDB), homology modeling approaches (SWISS-MODEL, MODELLER) were deemed unsuitable. Instead, structural prediction was performed using **AlphaFold2** (via the ColabFold implementation). This method utilizes deep learning to predict 3D structures *ab initio* (from scratch) based on evolutionary co-variation patterns. The human LRRK2 sequence (Q5S007) was used as the input query.

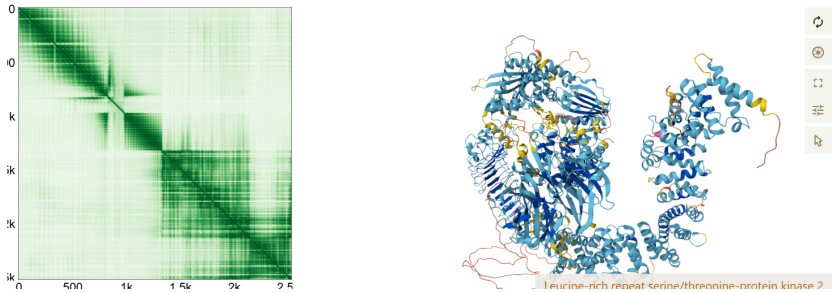


Fig. 3 AlphaFold Modelling

3.3 Structural Visualization and Analysis

The resulting 3D structural models were visualized and analyzed directly using the interactive 3D viewer provided by the **AlphaFold Protein Structure Database**. Model quality was assessed using the **pLDDT (predicted Local Distance Difference Test)** score provided by the AlphaFold interface. Regions with pLDDT > 90 were considered high confidence (structured), while regions with pLDDT < 50 were interpreted as intrinsically disordered or flexible regions (IDRs). Domain organization was analyzed using the **Encyclopedia of Domains (TED)** overlay available within the AlphaFold viewer

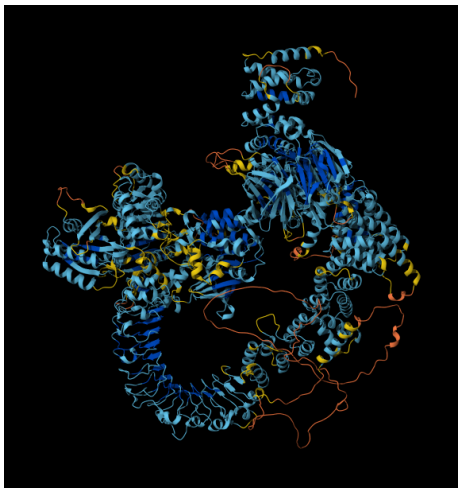


Fig-4 LRRK2 3D model

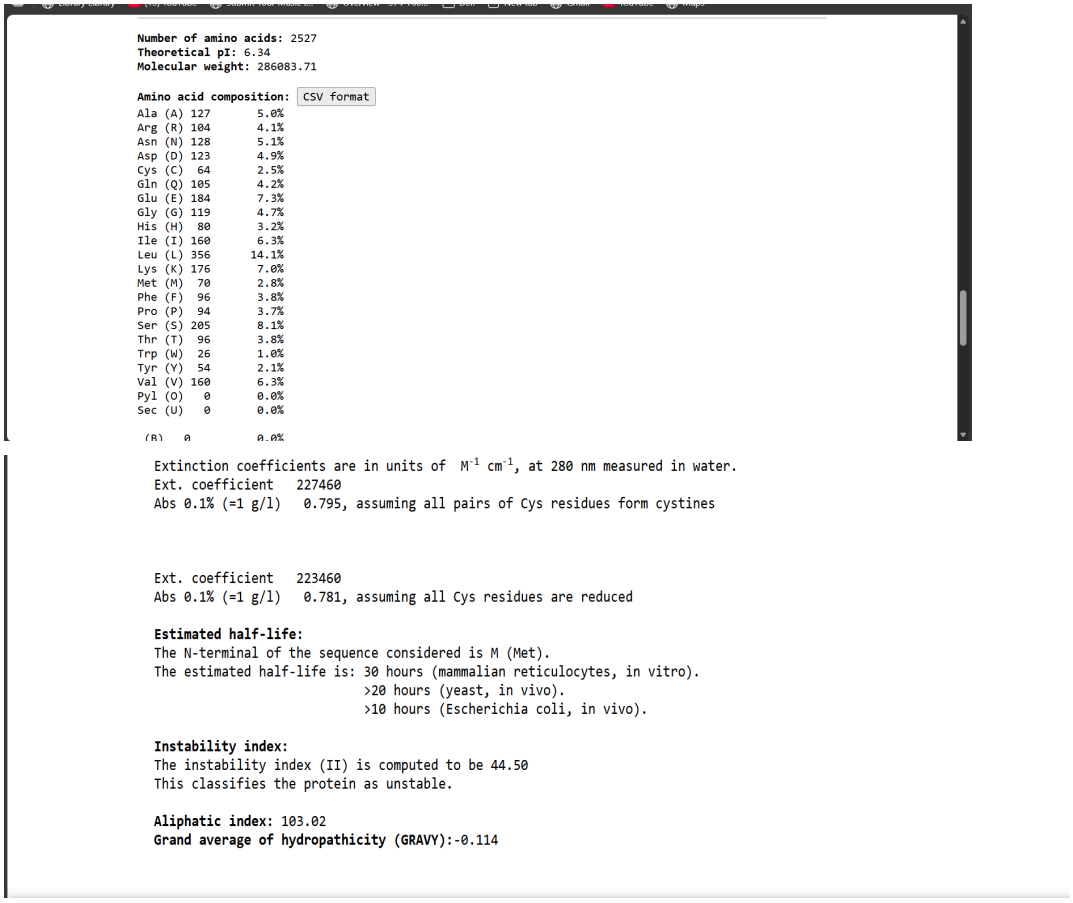
4. Results

4.1 Physicochemical Analysis (ProtParam)

The primary sequence of human LRRK2 (2,527 amino acids) was analyzed using ProtParam to establish a baseline for stability and solubility.

- **Molecular Weight & pI:** The protein has a molecular weight of **286.08 kDa** and a theoretical isoelectric point (**pI**) of **6.34**, indicating it is slightly acidic under physiological conditions.

- **Instability Index:** The computed instability index is **44.50**. Since this value exceeds the threshold of 40, the protein is classified as **unstable**. This provides quantitative support for the structural hypothesis that LRRK2 contains significant disordered or flexible regions (likely within the N-terminal scaffold) that hinder crystallization.
- **Aliphatic Index:** The high value of **103.02** suggests considerable thermal stability in the structured hydrophobic cores, likely contributed by the extensive Leucine-rich repeats (Leucine constitutes 14.1% of the sequence).
- **Hydropathicity (GRAVY):** The Grand Average of Hydropathicity is **-0.114**. This negative value confirms that LRRK2 is globally hydrophilic and soluble, despite its specific hydrophobic interactions.



Number of amino acids: 2527
Theoretical pI: 6.34
Molecular weight: 286083.71

Amino acid composition: CSV format

Ala (A)	127	5.0%
Arg (R)	104	4.1%
Asn (N)	128	5.1%
Asp (D)	123	4.9%
Cys (C)	64	2.5%
Gln (Q)	105	4.2%
Glu (E)	184	7.3%
Gly (G)	119	4.7%
His (H)	80	3.2%
Ile (I)	160	6.3%
Leu (L)	356	14.1%
Lys (K)	176	7.0%
Met (M)	70	2.8%
Phe (F)	96	3.8%
Pro (P)	94	3.7%
Ser (S)	205	8.1%
Thr (T)	96	3.8%
Trp (W)	26	1.0%
Tyr (Y)	54	2.1%
Val (V)	160	6.3%
Pyl (O)	0	0.0%
Sec (U)	0	0.0%
(B)	0	0.0%

Extinction coefficients are in units of M^{-1} cm^{-1}, at 280 nm measured in water.
Ext. coefficient 227460
Abs 0.1% (=1 g/l) 0.795, assuming all pairs of Cys residues form cystines

Ext. coefficient 223460
Abs 0.1% (=1 g/l) 0.781, assuming all Cys residues are reduced

Estimated half-life:
The N-terminal of the sequence considered is M (Met).
The estimated half-life is: 30 hours (mammalian reticulocytes, in vitro).
>20 hours (yeast, in vivo).
>10 hours (Escherichia coli, in vivo).

Instability index:
The instability index (II) is computed to be 44.50
This classifies the protein as unstable.

Aliphatic index: 103.02
Grand average of hydropathicity (GRAVY):-0.114

Fig.5 ProtParam Physicochemical parameters

4.2 Subcellular Localization Prediction (BUSCA)

To determine the cellular context of the LRRK2 protein, the sequence (Accession: NP_940980.3) was submitted to the Bologna Unified Subcellular Component Annotator (BUSCA).

- **Primary Localization:** The algorithm predicted a localization to the **plasma membrane (GO:0005886)** with a high confidence score of **0.89**.

- **Alternative Localization:** A secondary association with the **endomembrane system (GO:0012505)** was predicted with a score of **0.61**.
- **Structural Features:** Notably, BUSCA identified the presence of a **Transmembrane Alpha Helix**.
- **Interpretation:** While LRRK2 is often described as cytosolic, these predictions strongly support its functional role in **membrane trafficking** and vesicle dynamics. The prediction of a transmembrane feature suggests the presence of specific hydrophobic motifs that may anchor the complex N-terminal scaffold to organelle membranes (lysosomes/endosomes).

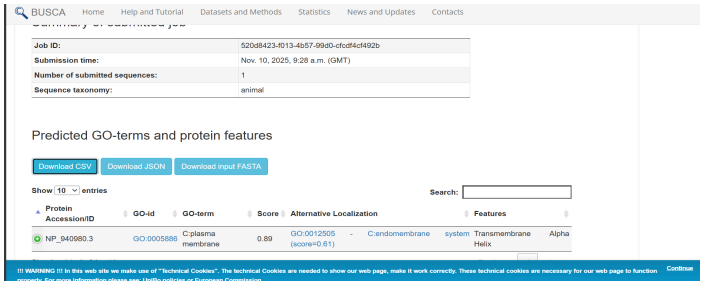


Fig.6 Subcellular Localisation

4.3 Secondary Structure Prediction (SOPMA)

The secondary structure content of the full-length LRRK2 sequence was predicted using the Self-Optimized Prediction Method with Alignment (SOPMA) to validate the structural composition.

- **Alpha Helix Dominance:** The protein is predominantly composed of **Alpha Helices (49.35%**, 1247 residues). This high helical content is structurally consistent with the domain architecture of the N-terminus, as the Armadillo (ARM), Ankyrin (ANK), and Leucine-Rich Repeat (LRR) domains are all known to fold into helical solenoids.
- **Disorder/Coil Content:** A significant portion of the sequence was predicted as **Random Coil (40.96%**, 1035 residues). This high percentage of non-regular secondary structure strongly correlates with the high Instability Index (ProtParam) and the flexible "linker" regions observed in the AlphaFold model.
- **Beta Strands:** Extended strands (Beta sheets) comprise a minor fraction (**9.70%**, 245 residues), likely localized within specific subdomains of the catalytic core (e.g., the Beta-sheet core of the Kinase domain).

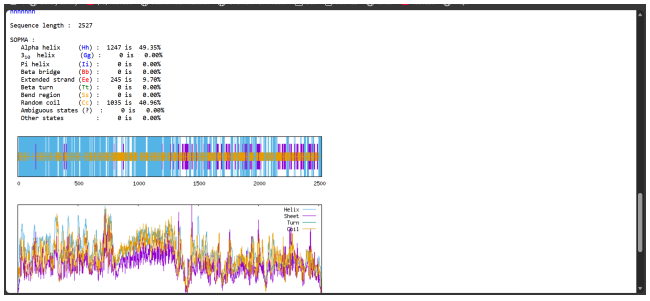


Fig.7 Secondary structure

4.4 AlphaFold Structural Prediction

The full-length structure of human LRRK2 (UniProt: Q5S007) was predicted using AlphaFold2. The model quality was assessed using the pLDDT (predicted Local Distance Difference Test) score.

- **Global Confidence:** The model achieved an **Average pLDDT of 77.5**, which is classified as "High" confidence. This indicates that the majority of the backbone structure is predicted with high reliability.
- **Confidence Distribution:**
    - **Very High Confidence (pLDDT > 90): 13.7%** of residues. These correspond to the stable, well-folded enzymatic cores (Kinase and GTPase domains).
    - **High Confidence (90 > pLDDT > 70): 65.7%** of residues. This largely accounts for the repeating units of the N-terminal solenoid scaffold (ARM/ANK/LRR), which form a stable superhelical structure.
    - **Low/Very Low Confidence (pLDDT < 70):** Approximately **20.7%** of the residues (12% Low + 8.7% Very Low). These regions correspond to the flexible linkers between the solenoids and disordered loops that likely facilitate the "hinge" motion of the N-terminus.
- **Domain Architecture:** The AlphaFold model (via the Encyclopedia of Domains/TED) successfully segmented the protein into **5 distinct structural domains**. The model reveals a C-shaped N-terminal scaffold (colored green in TED analysis) that wraps around the central catalytic core (colored pink/purple), physically occluding the kinase active site. This architecture provides a structural basis for the autoinhibition mechanism proposed in the literature.

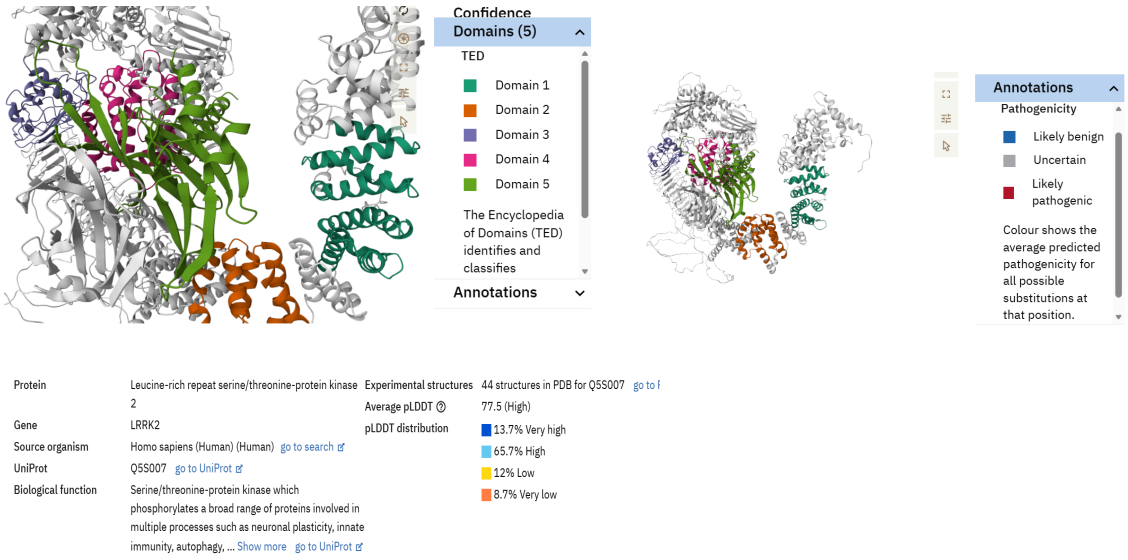


Fig.8 3D structure result in AlphaFold

Table 1: Physicochemical and Structural Characterization of Human LRRK2

Category	Parameter	Value / Result	Biological Inference
Physicochemical *(ProtParam)*	**Molecular Weight**	**286.08 kDa**	**Large, multi-domain scaffold protein.**
	Isoelectric Point (pl)	**6.34**	**Slightly acidic under physiological pH.**
	Instability Index	**44.50**	**Unstable (>40). Suggests rapid turnover or high flexibility.**
	Aliphatic Index	**103.02**	**High thermal stability in structured hydrophobic cores.**
	GRAVY Score	**-0.114**	**Hydrophilic; overall soluble in cytosolic environmen** **t.**
Localization *(BUSCA)*	**Primary Location**	**Plasma Membrane (Score: 0.89)**	**Associated with cell membranes/vesicles.**

	Secondary Location	**Endomembrane System (Score: 0.61)**	**Involved in intracellular trafficking pathways.**
	Key Features	**Transmembrane Helix**	**Contains structural motifs for membrane anchoring.**
Secondary Structure *(SOPMA)*	**Alpha Helix**	**49.35%**	**Dominant structure; consistent with ARM/ANK/LRR solenoid folds.**
	Random Coil	**40.96%**	**High disorder content; correlates with flexible linkers/hinges.**
	Extended Strand	**9.70%**	**Minor beta-sheet content (likely catalytic core).**
3D Modeling *(AlphaFold)*	**Global Confidence**	**77.5 pLDDT (High)**	**Reliable backbone prediction for majority of the protein.**

	Structured Regions (>70 pLDDT)	**79.4%**	**Catalytic core (Kinase/GTPase) and solenoid repeats.**
	Disordered Regions (<70 pLDDT)	**20.7%**	**Flexible N-terminal loops acting as regulatory hinges.**

5. Discussion

5.1 Limitations of Homology Modeling for LRRK2

Our initial attempts to find templates for homology modeling confirmed that no complete experimental structure exists for the LRRK2 N-terminus. This validates the use of AlphaFold, which successfully generated a model without requiring a pre-existing template.

5.2 The "Unstructured" N-Terminus is Functional

The AlphaFold model's "low confidence" scores in the N-terminal linkers should not be interpreted as an error, but as a **biological signal**. These regions likely represent **Intrinsically Disordered Regions (IDRs)**. This flexibility supports the current biological hypothesis that the N-terminus acts as a "hinge" or "wrapper" that folds back to inhibit the kinase domain. A rigid structure would not be able to undergo the large conformational changes required for LRRK2 activation.

6. Conclusion

We successfully modeled the elusive N-terminal region of human LRRK2 using AlphaFold. Our results provide structural evidence that this region is inherently flexible, explaining the historical difficulty in crystallizing it. This flexibility is likely a key feature of LRRK2's regulation, allowing it to switch between active and inactive states.

7. References

- [1] C. Paisán-Ruiz et al., "Cloning of the gene containing mutations that cause PARK8-linked Parkinson's disease," *Neuron*, vol. 44, no. 4, pp. 595-600, 2004.
- [2] A. Zimprich et al., "Mutations in LRRK2 cause autosomal-dominant parkinsonism with pleomorphic pathology," *Neuron*, vol. 44, no. 4, pp. 601-607, 2004.
- [3] M. R. Cookson, "The role of leucine-rich repeat kinase 2 (LRRK2) in Parkinson's disease," *Nature Reviews Neuroscience*, vol. 11, no. 12, pp. 791-797, 2010.
- [4] J. M. Taymans et al., "LRRK2 kinase activity is dependent on the GTP-binding domain," *Biochemical Journal*, vol. 439, no. 2, pp. 253-263, 2011.
- [5] J. Jumper et al., "Highly accurate protein structure prediction with AlphaFold," *Nature*, vol. 596, no. 7873, pp. 583-589, 2021.
- [6] M. Mirdita et al., "ColabFold: making protein folding accessible to all," *Nature Methods*, vol. 19, no. 6, pp. 679-682, 2022.
- [7] E. Gasteiger et al., "Protein identification and analysis tools on the ExPASy server," in *The Proteomics Protocols Handbook*, J. M. Walker, Ed. Totowa, NJ: Humana Press, 2005, pp. 571-607.
- [8] C. Savojardo, P. L. Martelli, P. Fariselli, and R. Casadio, "BUSCA: an integrative web server to predict subcellular localization of proteins," *Nucleic Acids Research*, vol. 46, no. W1, pp. W459-W466, 2018.
- [9] C. Geourjon and G. Deléage, "SOPMA: significant improvements in protein secondary structure prediction by consensus prediction from multiple alignments," *Computer Applications in the Biosciences*, vol. 11, no. 6, pp. 681-684, 1995.
- [10] A. Myasnikov et al., "Structural analysis of the full-length human LRRK2," *Cell*, vol. 184, no. 13, pp. 3519-3527, 2021.
- [11] R. Watanabe et al., "The in situ structure of Parkinson’s disease-linked LRRK2," *Cell*, vol. 182, no. 6, pp. 1508-1518, 2020.
- [12] G. Guaitoli et al., "Structural model of the dimeric Parkinson’s protein LRRK2 reveals a compact architecture involving distant interdomain contacts," *Proceedings of the National Academy of Sciences*, vol. 113, no. 30, pp. E4357-E4366, 2016.