JETIR.ORG ISSN: 2349-5162 | ESTD Year : 2014 | Monthly Issue
JOURNAL OF EMERGING TECHNOLOGIES AND INNOVATIVE RESEARCH (JETIR)
An International Scholarly Open Access, Peer-reviewed, Refereed Journal

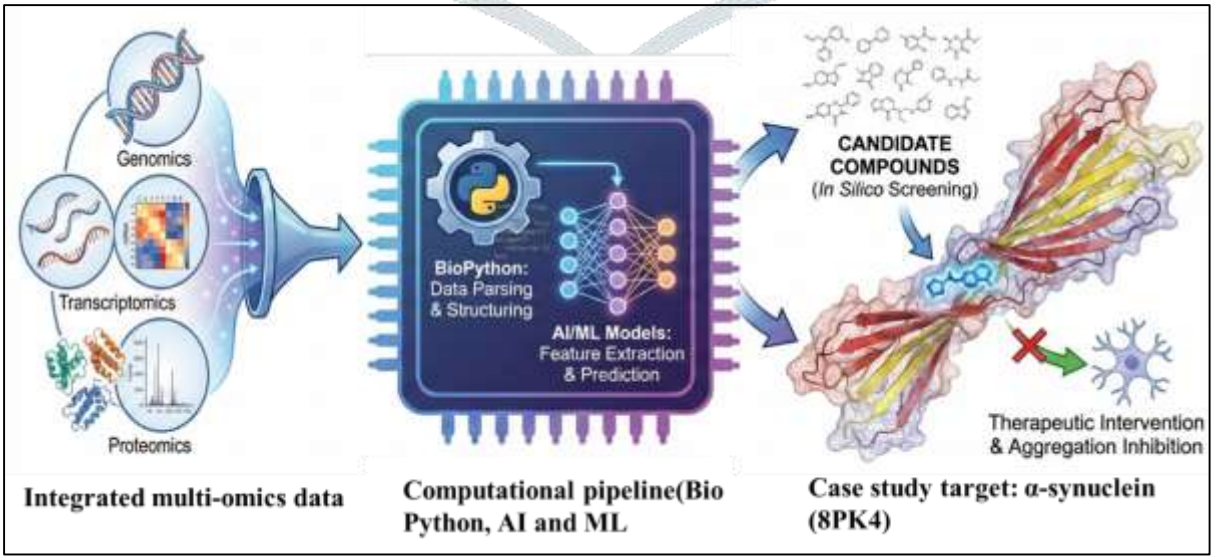
Mult-omics-based drug discovery pipelines incorporating machine learning, AI, and BioPython in Parkinson's disease: A case study of 5A polymorph of alpha-synuclein

Neha Singh[1] Uma Kumari[1]

[1]Project Trainee at Bioinformatics Project and Research Institute, Noida - 201301, India
[1]Senior Bioinformatics Scientist, Bioinformatics Project and Research Institute, Noida - 201301, India
Corresponding Author: Uma Kumari(uma27910@gmail.com)

Abstract:

Parkinson's Disease (PD) is a progressive neurodegenerative disorder driven by complex molecular dysfunctions in the brain over many years. A central hallmark is the accumulation of α-Synuclein, which forms distinct protein aggregates known as polymorphs. These polymorphic variants differ in toxicity, propagation, and responsiveness to therapies, underscoring the need for advanced, systems-level approaches in drug development. Recent breakthroughs in multi-omics technologies, artificial intelligence (AI), and computational biology offer powerful tools to unravel PD pathogenesis and accelerate therapeutic discovery. Yet, integrating these innovations into reproducible and unified pipelines remains a major challenge. Multi-omics disciplines—including genomics, transcriptomics, proteomics, metabolomics, and lipidomics—enable identification of novel drug targets, patient stratification based on pathology, and prioritization of compounds with higher therapeutic potential. A notable example is the 5A polymorph (PDB ID: 8PK4), which illustrates how structural variations in α-Synuclein can inform rational, structure-based drug design. Building on this, we propose a conceptual pipeline that integrates multi-omics data with AI, BioPython, and machine learning to design drugs targeting specific polymorphs. This framework combines disease modules, virtual screening, and predictive modeling. Finally, we address current limitations, ethical considerations, and highlight future opportunities such as digital twin modeling to transform PD drug discovery.



Keywords: Parkinson's Disease, α-Synuclein polymorphs,Multi-omics, Genomics, Proteomics, Metabolomics, Artificial Intelligence (AI), Machine Learning (ML), Structure-based drug design, Digital twin modeling

1. Introduction

Parkinson's disease (PD) is the second most common neurodegenerative disorder after Alzheimer's, characterised by both motor symptoms such as tremors, stiffness, and slowed movement, and non-motor symptoms including memory loss, sleep disturbances, and autonomic dysfunction. A central feature of PD is the misfolding and aggregation of alpha-synuclein, a presynaptic protein that plays a critical role in disease progression. **[1]** Despite decades of research, current therapies remain largely symptomatic, and the variability in disease course across patients continues to challenge understanding. Genome-wide association studies (GWAS) have identified genetic regions, such as SNCA, associated with PD, but these findings alone do not fully account for the disease's complex mechanisms. **[2]** The integration of multi-omics approaches spanning genomics, transcriptomics, proteomics, metabolomics, and lipidomics has revealed coordinated dysregulation across pathways including mitochondrial function, vesicle trafficking, lipid metabolism, neuroinflammation, and synaptic signalling. This systems-level perspective enables the identification of disease modules and molecular signatures that single datasets cannot capture. However, the vast scale of omics data requires advanced computational tools to extract meaningful biological insights**. [3]** Artificial intelligence (AI) and machine learning (ML) techniques, such as deep neural networks, ensemble models, and graph-based methods, have shown promise in subtyping neurodegenerative diseases, discovering biomarkers, and repurposing drugs. By integrating multi-omics data, AI models can identify disease-related molecular networks and compounds that influence alpha-synuclein aggregation and toxicity**.[4]** Yet, many studies overlook the structural diversity of alpha-synuclein fibrils, which may hold the key to targeted therapies. Bio Python offers a powerful framework to unify omics data, structural biology, molecular docking, and ML outputs into cohesive pipelines for drug discovery. This review outlines a conceptual framework for studying the α-synuclein 5A polymorph, presenting a universal programming architecture that converges structural polymorphism, systems-level omics, and AI-driven modelling. Such integration provides a blueprint for future multiphasic drug development in PD and related synucleinopathies. **[5]**

2. Alpha-synuclein in PD (biology & structural polymorphism)

Alpha-synuclein is an intrinsically disordered protein of 140 residues containing a synuclein-specific sequence motif in its primary structure**.[6]** The protein is divided into three separate structures: (1) N-terminal segment (residues 1 60), (2) an outer non-amyloid-B domain (NAC) (residues 61- 95), and (3) an outer domain (residues 96 -140). The N-terminal region contains the most conserved repeats of 11 residues, with the KTKEGV consensus motif pivotal to amphipathic 1a-helial structure and lipid membrane association. The most central part of the NAC, which is also responsible due to the presence of many hydrophobic amino acids, is the main factor in the amyloidogenic aggregation potential of the protein. **[7,8]**To date, there are 6 well-characterised missense SNCA mutations in the gene encoding α-synuclein, with strong evidence of causality, plus a few newer variants under study. Six mutations are as follows

Mutation	Amino acid change	References
A53T (Alanine to Threonine at position 53)	p.Ala53Thr	[16]
A30P (Alanine to Proline at 30)	p.Ala30Pro	[17]
E46K (Glutamate to Lysine at 46)	p.Glu46Lys	[17]
H50Q (Histidine to Glutamine at 50)	p.His50Gln	[18]
G51D (Glycine to Aspartate at 51)	p.Gly51Asp	[19]
A53E (Alanine to Glutamate at 53)	p.Ala53Glu	[20]

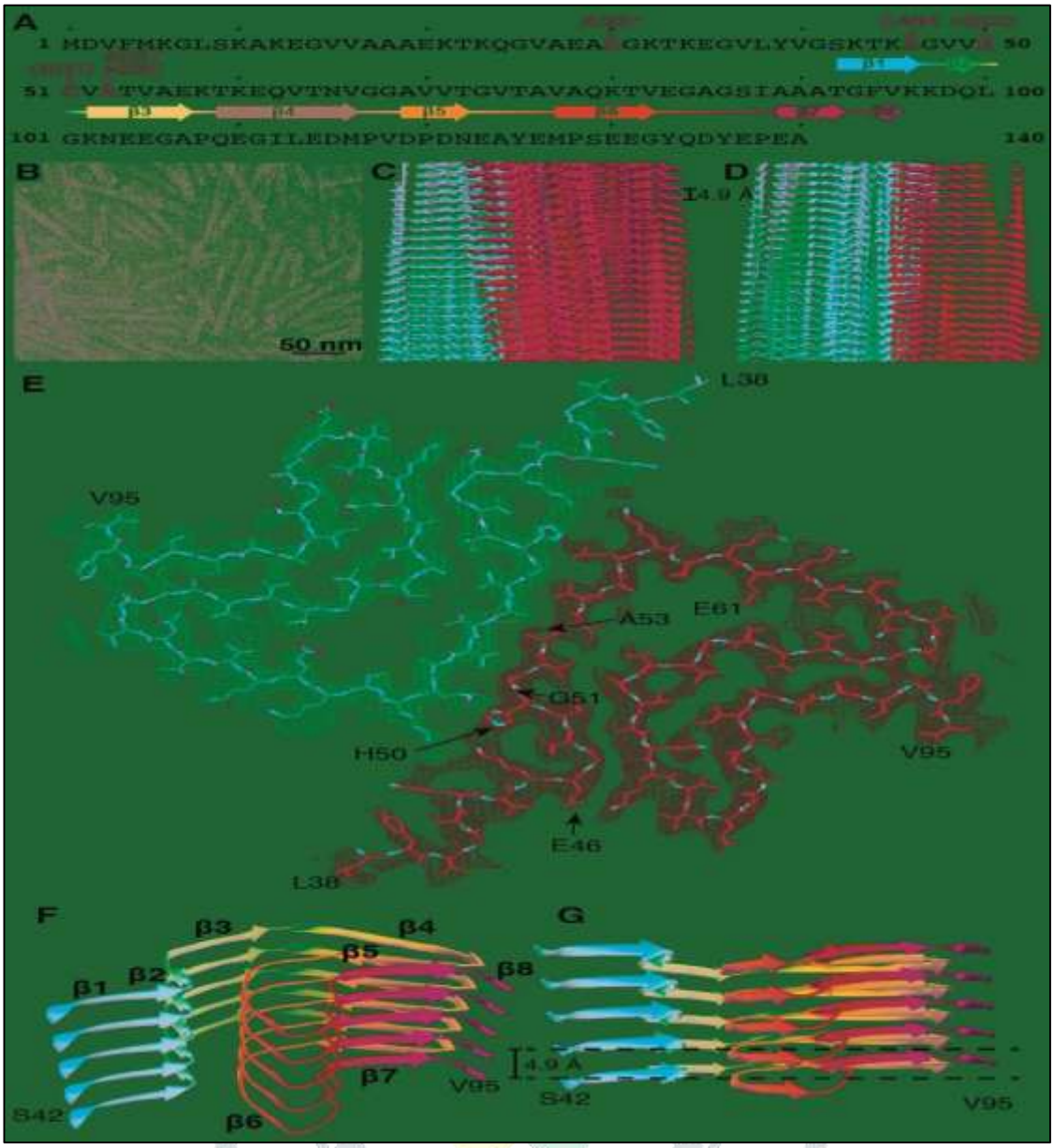


Figure 2. A schematic representation of the sequence of human alpha-synuclein (alpha-Syn). There are labelled familial mutations. The beta strand regions are indicated with arrows going from blue (Low) to red (High). A cryo-electron microscopy (Cryo-EM) image of the distribution and general look of alpha-Syn fibrils. A Cryo-EM reconstruction of the alpha-Syn (1-121) fibrils, showing two protofilaments (Orange and Blue). A cross-section of (C), illustrating the separation of beta-strands. A cross-section of a fibril along the axis, illustrating the arrangement of the two alpha-Syn protofilaments and atomic model fit. Labels of the initial (L38) and the last (V95) residues fitting. Arrows labelled the positions of four of the five mutations associated with familial Parkinson's disease. The distribution of beta-strands of a single protofilament of the alpha-Syn fibril, which corresponds to residues 42 to 95. The colour scheme is the same as in (A). The same as in (F), but from a perpendicular view of the fibril axis, showing some differences in the heights of different areas of a single protofilament**. [9]**

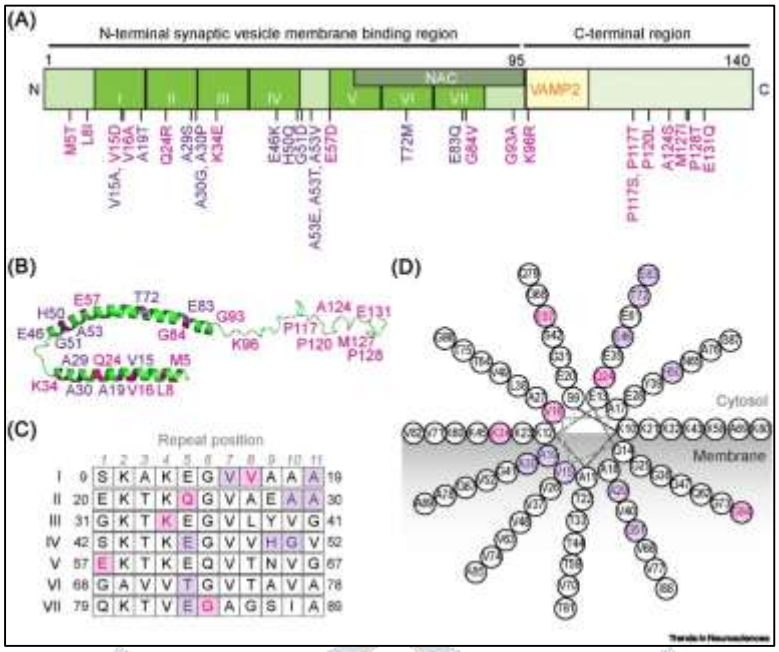


Figure.3 With respect to the α-synuclein protein, there are several key features. The N-terminal section allows for the formation of an amphipathic alpha-helix, allowing the protein to associate with synaptic vesicles. The C-terminal region facilitates binding to its partner protein, VAMP2. In addition, the N-terminal section contains 11 imperfect KTKEGV sequences, as well as the NAC domain, which has been shown to promote aggregation of this protein. Mutations within the α-synuclein protein have been identified to cause a variety of diseases (purple) or may have unknown effects (pink). Analysis of the locations of the mutations reveals that most mutations cluster at positions 5, 10, and 11 of the KTKEGV repeating sequences. Further analysis indicates that the mutation clusters at positions 10 and 11 are directed toward the membrane; however, the mutation cluster at position 5 is directed towards the cytosol. NAC, non-amyloid-β component**.[10]**

Localisation:

Alpha-synuclein shows a highly localised mode of localisation as it is found to be mainly localised to the presynaptic ends of the neurons, where it is strongly linked to synaptic vesicles. On the contrary, it is limited in its presence in neuronal somata, dendritic arbours, or extrasynaptic space. . Alpha-synuclein occurs mainly as a cytosolic protein endogenously, but low levels of exogenously produced, extracellular alpha-synuclein are found in the fluid of the body, like cerebrospinal fluid (CSF) and plasma**.[11.12]**

In case of stress or dysfunction of the mitochondria in cells, the endogenous release of alpha-synuclein is significantly increased in the neurons. Neuronal activity regulates the exocytosis of alpha-synuclein in glutamatergic neurotransmission. Glutamate-stimulated postsynaptic receptors promote the enhanced exocytosis of alpha-synuclein through synaptic vesicle exocytosis**. [13]**

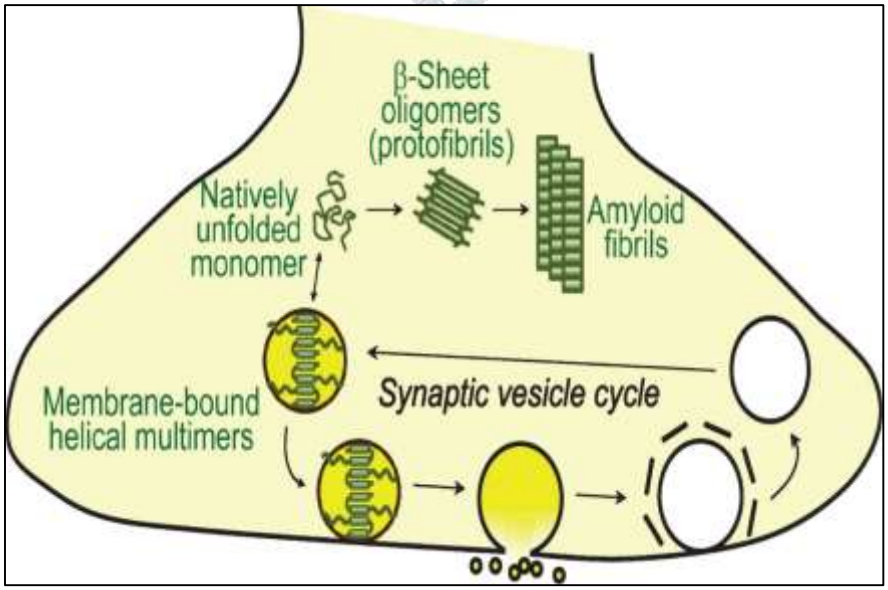


Figure.4 The figure illustrates the different conformations of α-synuclein, both physiologically and pathologically, at the synapse. It has been shown that the cytosolic form of α-synuclein is natively unstructured and typically exists as a monomeric protein. However, when α-synuclein binds to synaptic vesicles (SVs) through its membrane-binding domain (MBD), it adopts an α-helical conformation, due to the interaction of its N-terminal residues with the lipid bilayer of SVs, resulting in membrane-bound conformations. Furthermore, upon interaction with membranes, α-synuclein undergoes multimerisation and is subsequently required for its viable physiological role at the synapse. In terms of pathology, cytosolic α-synuclein in the unstructured form can misfold to form pathological oligomers, specifically those that contain β-sheets (aka protofibrils), and over time can aggregate and develop into amyloid fibrils. **[14]**

Function: In healthy individuals, the function of alpha synuclein is not well understood; however, a hypothesis has been proposed regarding its role in vesicular transport between brain cells. Its role in the regulation of synaptic neurotransmission was confirmed by the downregulation of alpha synuclein in knockout mice and primary neurons.

In the context of disease, alpha-synuclein is a major component of Lewy bodies and Lewy neurites and has been implicated in the degeneration of dopamine-producing neurons, which underlie the motor dysfunction observed in Parkinson's disease.**[15]**

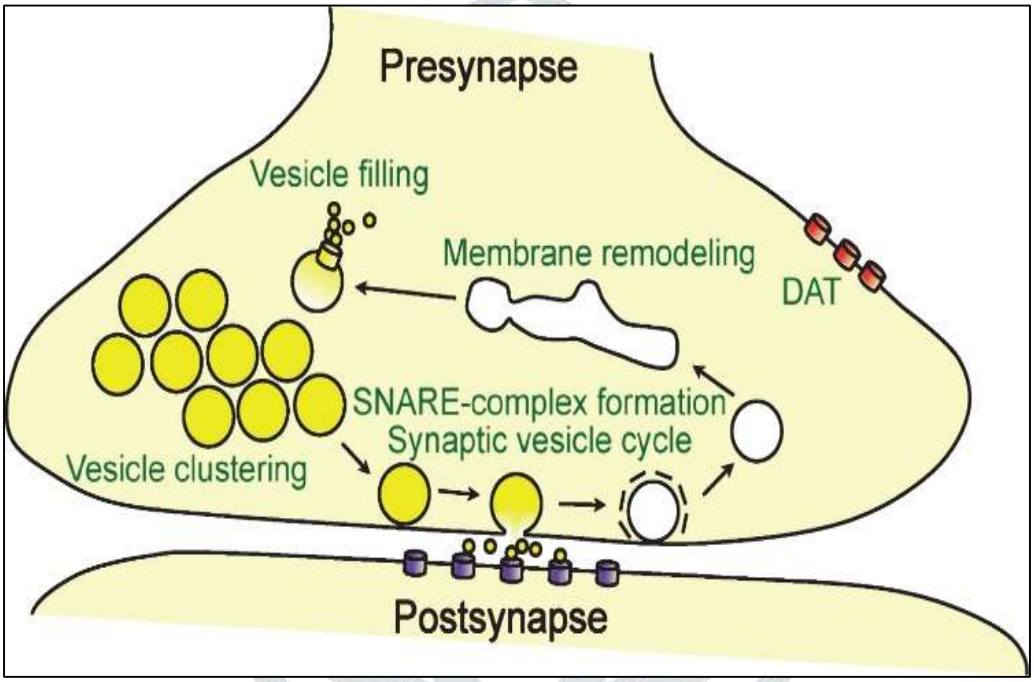


Figure. 5 Synaptic processes influenced by synaptic α-synuclein function as depicted in the figure; includes elements of membrane remodelling and modulations of dopamine transporter (DAT) and vesicular monoamine transporter (VMAT2) clustering and maintaining pools of synaptic vesicles, enabling/promoting independent SNARE-complex assembly during vesicular exocytosis, as well as modulation of synaptic vesicular release cycles.**[16]**

Proposed Stages of α-Synuclein Spread

Spreading of α-Synuclein is divided into different stages. In some cases, Lewy pathology at stage 1 involves the olfactory bulb, and a new finding is that brain involvement can be limited to the olfactory bulb or to anterior olfactory structures only (glomeruli, olfactory mucosa), suggesting a new hypothesis for studying possible routes of spread in PDs.[28] A recent axonal tracing study revealed the existence of an anterograde pathway between the substantia nigra and the olfactory bulb in rats.The idea that a neurotropic virus uses a "key-lock" mechanism to infiltrate unprotected nerve cell fibres may help to explain why sPD(spreading Parkinson's disease) only affects humans among a wide variety of vertebrates. For α-synuclein to undergo a stable (albeit pathological) conformation, the pathogen would have to deactivate endogenous chaperones and cause α-synuclein to form a stable conformation before prion-like propagation.

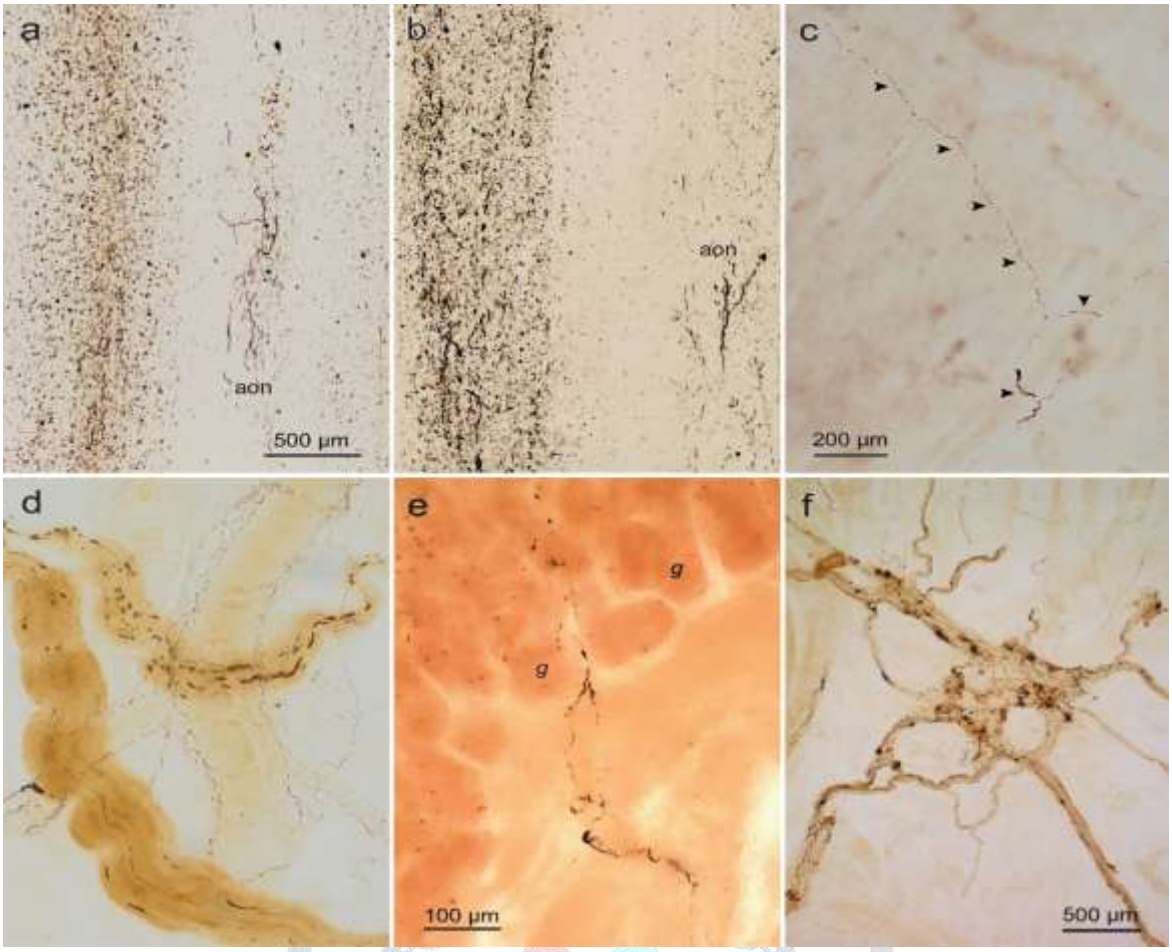


Figure.6 (a-f): Detection of Lewy Body Pathology in the olfactory bulb and gastric Auerbach plexus via immunohistochemical methods for α-synuclein (100 µm sections from polyethene glycol-embedded material). (a) Olfactory bulb and anterior olfactory nucleus (50 y/o male, stage 2). Lewy neurites and Lewy bodies were found in both the dorsal motor nucleus of the vagal nerve and the intermediate reticular zone of the medulla as well. It would seem likely that this non-demented individual ultimately would have developed Parkinson's Disease (PD) if he had lived longer. (b) Olfactory bulb and anterior olfactory nucleus of a 63 y/o female (stage 2). There was significantly less severe pathology compared to the case illustrated in A, but there were still some Lewy neurites and Lewy bodies found in the dorsal motor nucleus of the vagal nerve, intermediate reticular zone, nucleus raphes magnus, and locus coeruleus. (c) A tangentially cut section through the gastric cardia illustrates Lewy neurites (indicated by arrowheads) in the same case as in B. (d) A section of a gastric cardia that is cut tangentially to its surface are shown with intramural Lewy Pathology along with a large, branching blood vessel visible in the background surrounded by an extensive network of thread-like sympathetic nerve fibres, which illustrate the presence of both Lewy neurites and Lewy bodies in the dorsal motor nucleus of the vagal nerve, intermediate reticular zone and nucleus raphes magnus Micrographs (e) and (f) demonstrate details of Lewy neurites found in PD penetrating the muscularis mucosa and extending upward into the lamina propria of the stomach where they wrap around the gastric glands (g) as shown in a section at right angles to the mucosa (a) (section through the left). (f) A section through the gastric cardia of a patient with PD and a disease duration of 11 years. Scale bars: (a) applies to (b), (c) applies to (d); stage numbers in parentheses refer to the neuropathological stages of sporadic PD (1-6). Reproduction of micrographs (e) and (f) with permission from.**[17]**

Structural Polymorphs:

Cryo-EM has been extensively used for studying the structural basis of α-synuclein aggregation, resulting in the identification of multiple fibril polymorphs with distinct molecular arrangements. Understanding molecular arrangements is crucial for understanding the heterogeneity of synucleinopathies, including Parkinson's disease (PD), multiple system atrophy (MSA), and dementia with Lewy bodies (DLB).

The first high-resolution cryo-EM structures of full-length α-synuclein fibrils were reported by (Guerrero-Ferreira et al., 2019).They solved two novel polymorphs (2a and 2b). Their study showed that even α-synuclein have an identical sequence, then also it can adopt structurally distinct folds, stabilised by different protofilament interfaces and salt-bridge networks (e.g., K45–E57 in polymorph 2a). Importantly, hydrophobic

amino acids such as Valine and Alanine are present in the hydrophobic NAC region (residues 61-95, which was shown to constitute the fibril core, highlighting its critical role in aggregation.

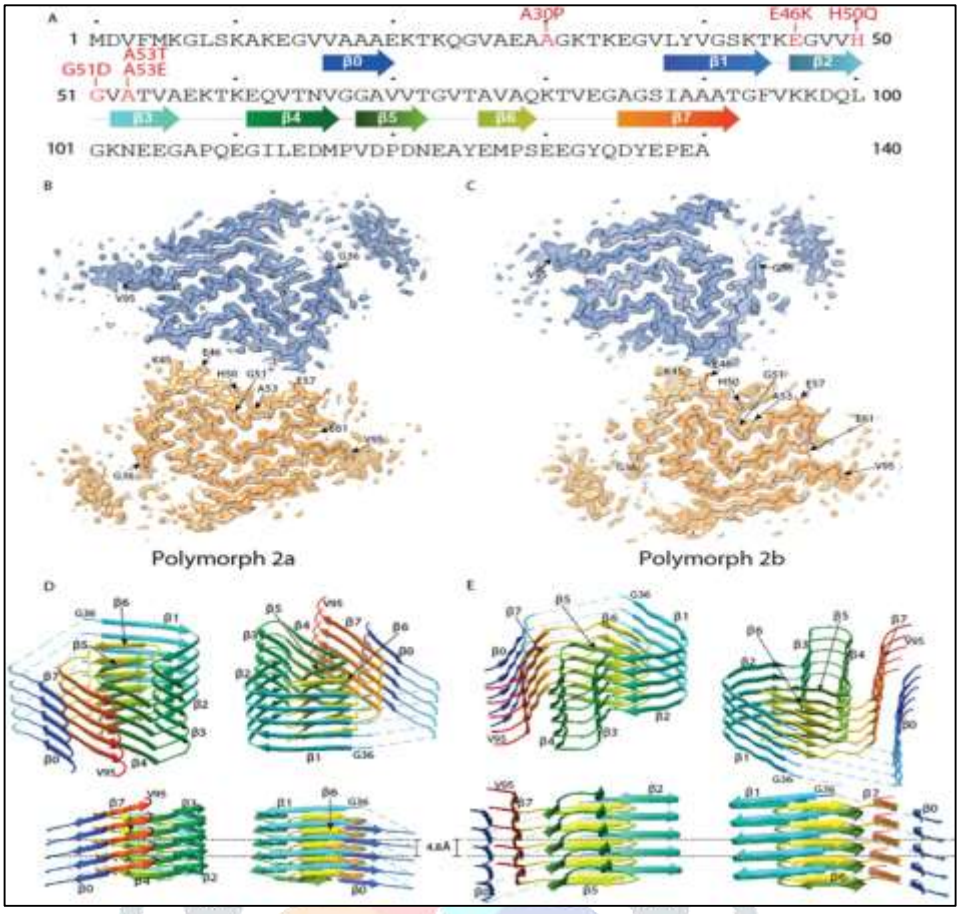


Figure.7 The figure shows the cryo-EM structures and β-strand organisation of α-synuclein fibril polymorphs 2a and 2b, highlighting their distinct protofilament interfaces and the positions of familial Parkinson's disease mutations within the fibril core.**[18]**

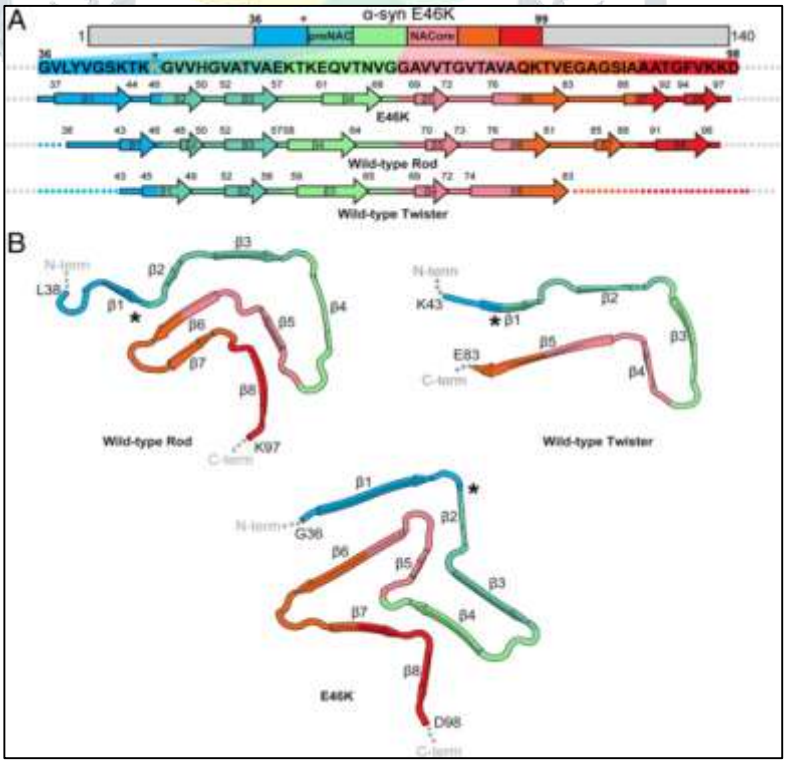


Figure.8 (A) A structural model of the secondary structure organisation of α-synuclein fibrils depicting the N-terminal, preNAC, NAC core and C-terminal regions, with β-strands and loops of the E46K mutant fibril compared to those of the wild-type rod and wild-type twister polymorphs. The position of the E46K mutation within the fibril core is shown and demonstrates its close proximity to important β-strands interfaces and salt-bridge networks important for stabilising and packing the fibrils.(B) A backbone trace representation of the protofilament folds, based on our structural data, of the wild-type rod, wild-type twister, and E46K α-synuclein fibrils. Each of the β-strands (β1–β8) is colour-coded to highlight differences in conformation between the different polymorphs. The E46K mutation appears to affect the way the inter-strand interactions are organised, resulting in a different protofilament topology and an increase in the stability of the fibrils compared to wild-type structures.**[19]**

Overall, abundant complementary ligand binding studies have shown that small molecules bind to fibril polymorphs in a structure-dependent form. As an example, the (Liu et al., 2024), study revealed that a common chemical scaffold can occupy different sites on wild-type and mutant α-synuclein fibrils, which illustrates polymorph-specific occupancy. This information is crucial in rational therapeutic design because drug candidates should be able to take cognisance of the conformational differences of α-synuclein fibrils.[60]Collectively, all these structural developments define that the α-synuclein fibril polymorphism is at the centre of both the molecular pathogenesis of PD and the development of specific therapeutics. Here, the analysis of the polymorph is timely and will offer insight into the fibril core enrichment with hydrophobic amino acids like valine and alanine, which have been proposed to play an active role in aggregation and stability.The structural variety of the α-synuclein fibrils is a challenge and an opportunity for therapy development. One polymorph can be strongly bound by small molecules or antibodies but weakly bound by another. Equally, diagnostic ligands in Single Photon Emission Computed (SPECT) or Positron Emission Tomography (PET) imaging cannot identify some fibril conformations in cases where there are differences in binding pockets. Therefore, the rational design of therapeutic agents as well as imaging biomarkers can only be achieved through a detailed insight into polymorph-specific structures.[42]

3. Molecular Landscape of Parkinson's Disease

Parkinson's disease has a complex molecular landscape that ranges from genetic, epigenetic and environmental influences. Genome-wide studies revealed more than 90 risk loci and single-nucleotide polymorphisms (SNPs) associated with PD; these pathways related to synaptic function, mitochondrial homeostasis and proteostasis. The disease susceptibility is increased by rare high-penetrance mutations in *SNCA*, *LRRK2*, *PINK1*, *DJ-1*, and *GBA.* Altered methylation and histone signatures in dopaminergic neurons also modulate disease onset and progression, as revealed by epigenomic analyses.[15]

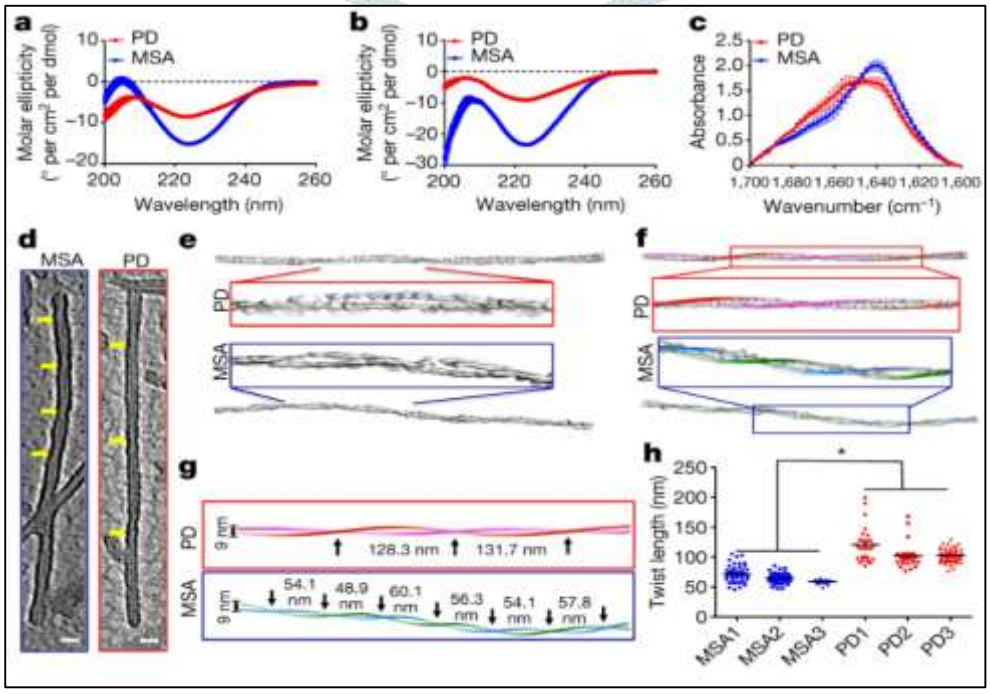


Figure.9 a, Circular dichroism spectra of α-syn aggregates from the CSF of patients with PD (red) or patients with MSA (blue), amplified by two rounds of α-syn-PMCA. Spectra were recorded from 35 μM suspensions of α-syn aggregates. Measurements were taken for all of the PD (n = 43) and MSA (n = 43) samples analysed and data (molar ellipticity) are mean ± s.e.m. **b**, A similar experiment was performed for α-syn aggregates that were amplified from the brain of patients with PD (n = 3) or patients with MSA (n = 3). **c**, FTIR spectra of α-syn aggregates that were obtained after two rounds of seeding and amplification of samples of CSF from patients with PD (n = 10) or patients with MSA (n = 10). The solution of aggregated proteins (5 μl; 5 mg ml^{-1}) was analysed with an FTIR-4100 spectrometer (JASCO). **d**, Cryo-ET was performed to evaluate structural differences between fibrils from patients with PD and fibrils from patients with MSA. Central slices of representative subtomograms of PD-associated fibrils and MSA-associated fibrils are shown. The negative-stained fibrils were imaged with a 300-kV electron microscope . Yellow arrows indicate twists in the filaments. Scale bar, 20 nm. **e**, Three-dimensional density maps segmented from the original tomograms. Boxed densities are magnified views. **f**, Three-dimensional helical models were built that overlapped with the corresponding densities of PD- and MSA-associated fibrils, including a magnification of the central region. **g**, Helical models showing the periodicity of twisting of PD- or MSA-associated fibrils. Black arrows indicate the twist in the 3D model of the filament. **h**, Quantification of the periodic spacing (in nm) in many different fibrils derived from samples from patients with PD (n = 3) or patients with MSA (n = 3). Each dot corresponds to a different fibril and data are mean ± s.e.m. *P < 0.05 by one-way ANOVA followed by Tukey's multiple comparison test. **[20]** Genetic predisposition, multi-level biochemical disruptions, and structurally heterogeneous α-synuclein structural forms require need for integration of multi-omics pipelines. A combination of genomics, proteomics, metabolomics, and structural biology with machine learning provides a powerful approach to look into the molecular landscape of PD and identify therapeutic targets.

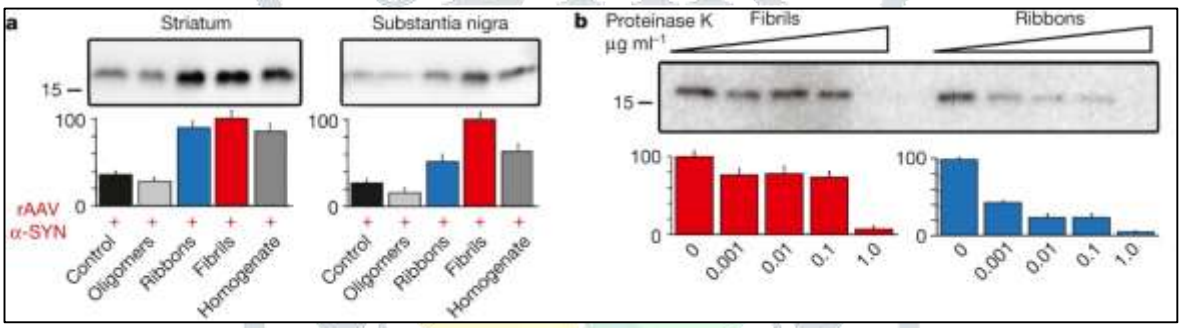


Figure.10 a, Sarkosyl-insoluble α-SYN is detected in striatum and substantia nigra after striatal inoculation of brain homogenate, oligomers, ribbons and fibrils in combination with rAAV-driven α-SYN overexpression. **b**, Degradation of sarkosyl-insoluble α-SYN in brain extracts from fibrils and ribbons inoculated animals with rAAV-driven α-SYN overexpression by increasing proteinase K concentrations. The results reveal that each strain imprints its intrinsic structure to endogenous α-SYN. Western blot analysis was performed using the monoclonal anti-α-SYN antibody clone 42. The data correspond to the mean and associated standard error calculated from 3 independent measurements from 2 biological replicates. **[21]**

4. Multi-omics integration for drug discovery

In genetic association studies, epigenomic profiling is a valuable tool for understanding the structure of disease-risk variants in a cell-type-specific manner and has enabled researchers to build a more refined model of disease risk. Several studies have examined the relationship between chromatin accessibility and GWAS data in single cells, demonstrating that many loci associated with an increased risk of PD are regulated by populations of neurons and glia. For example, researchers have used a similar approach using single single-nucleus ATAC-seq-based approach to identify and link inherited PD risk variants to specific regulatory element categories and have been able to use these links to prioritise functional investigations of non-coding variants likely responsible for altering the program of gene expression in cells involved with ND (neurodegeneration). These findings illustrate how integrating genomics and single-cell epigenomic analyses improves the ability to model disease mechanisms by linking genetic risk to specific cellular environments, thereby strengthening downstream network models and therapies identified using multi-omic frameworks and sequencing pipelines.

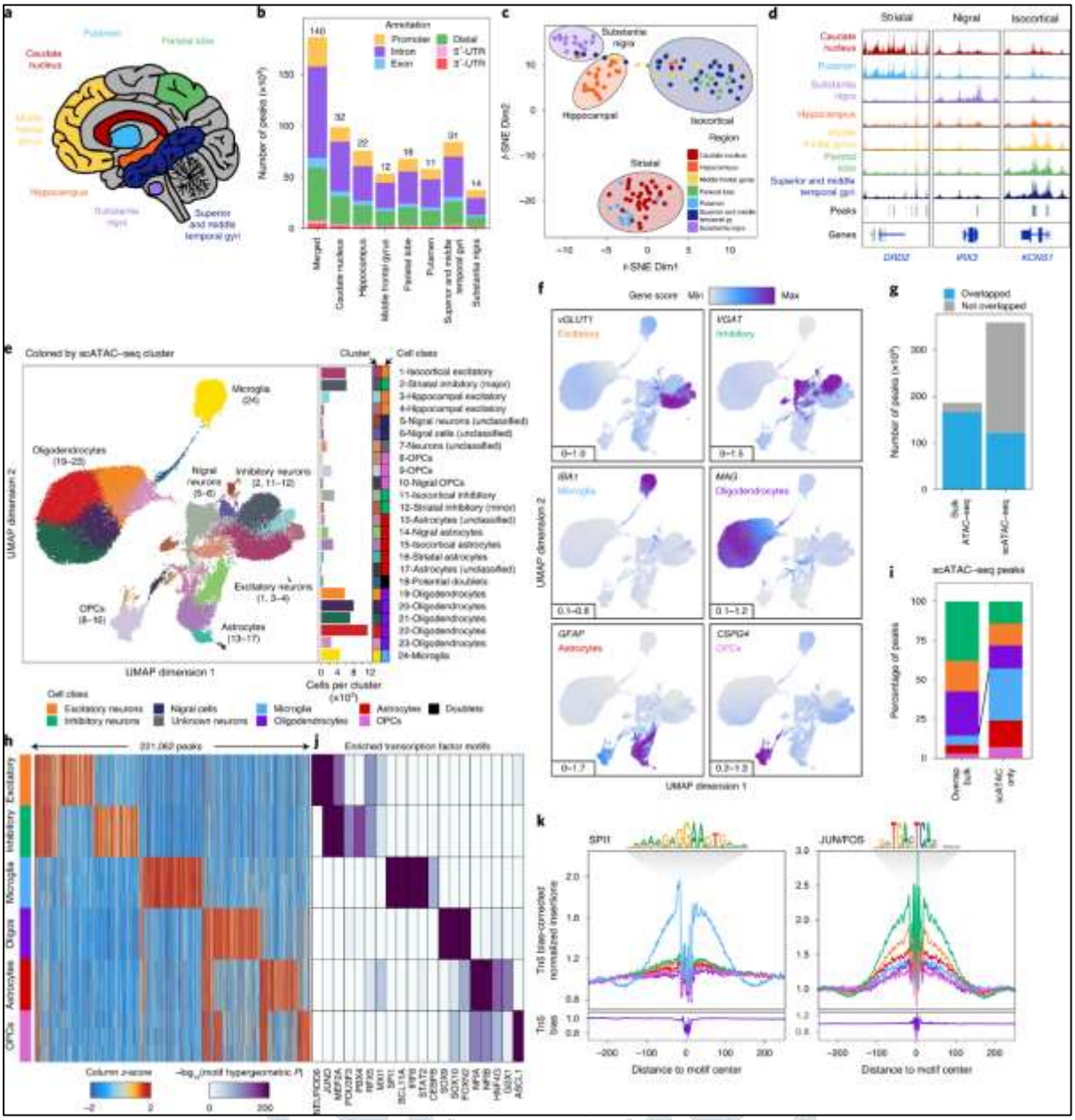


Figure 11.(A) The study profiles various areas within the brain. (B) A bar graph displays the number of reproducible peaks obtained from each distinct brain area's biological samples. The 'Merged' bar displays all the merged biologically reproducible peaks analysed up to this point. The top of each graph displays the total count of biologically reproducible brain sample(s) analysed by area (C). The first chart displays a representation of t-SNE (t-distributed stochastic neighbour embedding). All points on the charts represent a piece of tissue, with samples combined if appropriate (D), These are sequencing results for area-specific ATAC-seq (assay for transposase-accessible chromatin using sequencing) peaks. From left to right, the peaks are for DRD2 (striatum; chr11:113,367,951–113,538,919), IRX3 (substantia nigra; chr16:54,276,577–54,291,319) and KCNS1 (isocortex; chr20:45,086,706-45,107,665). The tracks have been equalised according to the total reads recorded from TSS regions (TSS = Transcription Start Site) (E). The chart has been colored according to the assigned cluster using UMAP, a uniform manifold approximation and projection after iterative LSI (Latent Semantic Indexing) of the scATAC-seq (single cell ATAC) data derived from 10 different biological samples; each dot represents one of the cells (n = 70,631); the colours correspond with the cluster of cells (F). The chart on the left displays the same information as E, with the addition of the cell's gene activity score for the lineage-defining gene. The scores around the bottom left of each panel indicate the lowest and highest scores. The overlap between the bulk ATAC-seq and single-cell ATAC-seq peak calls is illustrated by a bar graph (g). The "bulk ATAC-seq" refers to the peaks obtained from the merged peak sets of bulk ATAC-seq, which fall within the peaks of the scATAC-seq, while the "scATAC-seq" refers to peaks representing the merged peaks of scATAC-seq falling within those from the bulk ATAC-seq. An overlap is defined as any base(s) in common between both merged peak sets. A heat map (h) shows the chromatin accessibility of all 221,062 binarised peaks from the scATAC-seq; rows represent individual pseudobulk replicates (3) per cell type, and columns represent peaks; oligodendrogliocytes (Oligos) are represented here. A bar chart (i)

summarises how many of the 221,062 peaks from the scATAC-seq binarised peak set are also seen in bulk ATAC-seq ("Overlap bulk") versus those seen only in the scATAC-seq data ("scATAC only") for peaks unique to one cell type only (n = 172,111). The colours of the bars match those in Fig. 1h. A total of 221,062 binarised peaks from the scATAC-seq data were assessed for motif enrichment (j). Figure 1h displays the analysis of peak motifs of binarised type (motifs that matched the binarisation region). To overcome motif redundancy, prediction of transcription factor drivers was based on the average gene expression pattern across the GTEx brain tissues, and the accessibility of cis-regulatory regions (promoter regions) of transcription factors, as determined by groupings of cellular classes within a scATAC-seq dataset. Figure 1k displays a footprinting analysis of SPI1 (CIS-BP M6484_1.02) and JUN/FOS (CIS-BP M4625_1.02) as indicated by the six major cell classes.**[22]**

Integrating multi-omics data requires computational frameworks that recombine the inherent heterogeneous nature of genomics, transcriptomics, proteomics, metabolomics, and lipidomics datasets. Careful methodological design plays a crucial role in combining data with varying scales, dimensions, noise profiles, and missing values. Broadly, multi-omics integration strategies are divided into three main categories: early, intermediate, and late integration. In early integration, all omics data merge into a single feature matrix after suitable normalisation, batch correction, and scaling. This "data level fusion" strategy allows machine learning models to uncover cross-layer relationships.[19] Early integration excels in matched datasets, e.g., Hashemi et al. used concatenated genomics-proteomics matrices with random forests to prioritise PD synucleinopathy targets.[77]

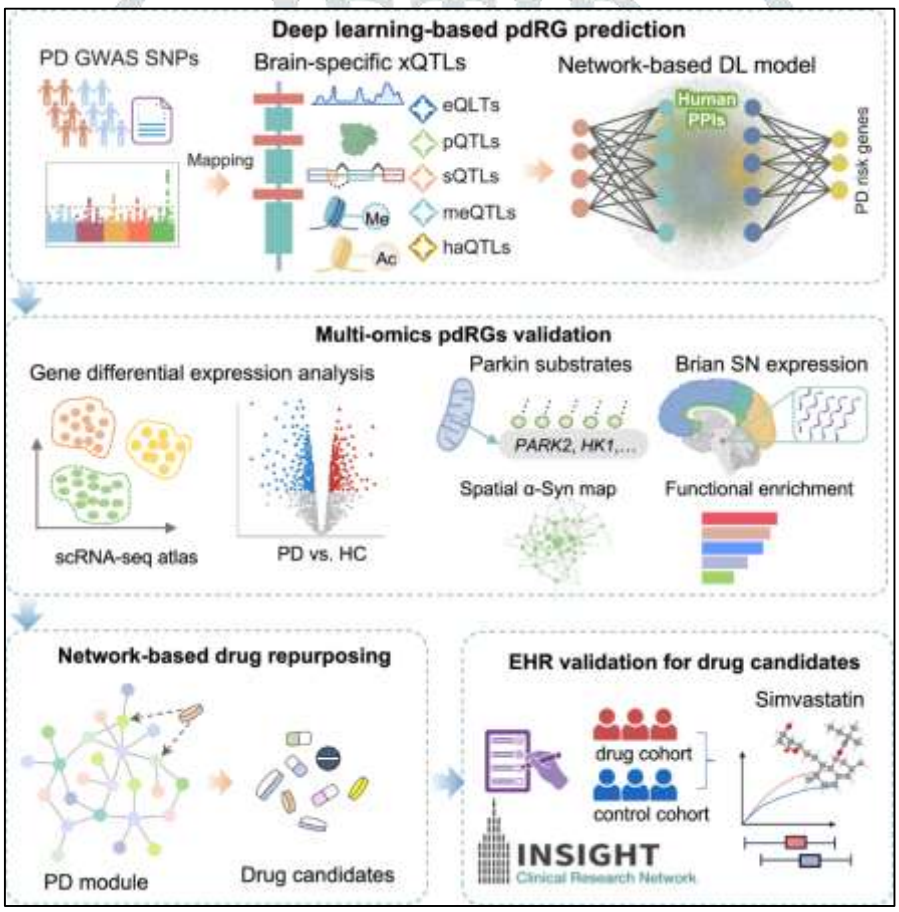


Figure.12 This process consists of four major phases: (1) Prediction of PD Risk Genes Using Deep Learning: First, the research team mapped out known PD-related SNPs to five types of xQTLs (gene expression, protein expression, splicing, methylation, histone acetylation). The next step was to include these genes in the set of candidate PD risk genes and apply the network-based DL approach to predict which of the genes in the Human Protein Architecture Network (HPA) are likely to be involved with PD; (2) Validation of Candidate PD Risk Genes Using Multi-Omics Knowledge; (3) Repurposing Drugs Using the Candidate PD Risk Gene Network; and (4) Confirmatory Analysis of Candidate PD Risk Gene Network-Repurposed Drugs Using EHR Data. Images created via BioRender.com . **[23]**

Intermediate integration involves multi-layer biological networks, such as (gene to protein to metabolite to lipid interactions). To connect different omics layers, it identifies molecular "bridge nodes". This includes: Weighted Gene Co-expression Network Analysis (WGCNA), Probabilistic or Bayesian graphical models,

Pathway-based integration (KEGG, Reactome), Knowledge graph-based integration, and Multi-omics factor analysis (MOFA).

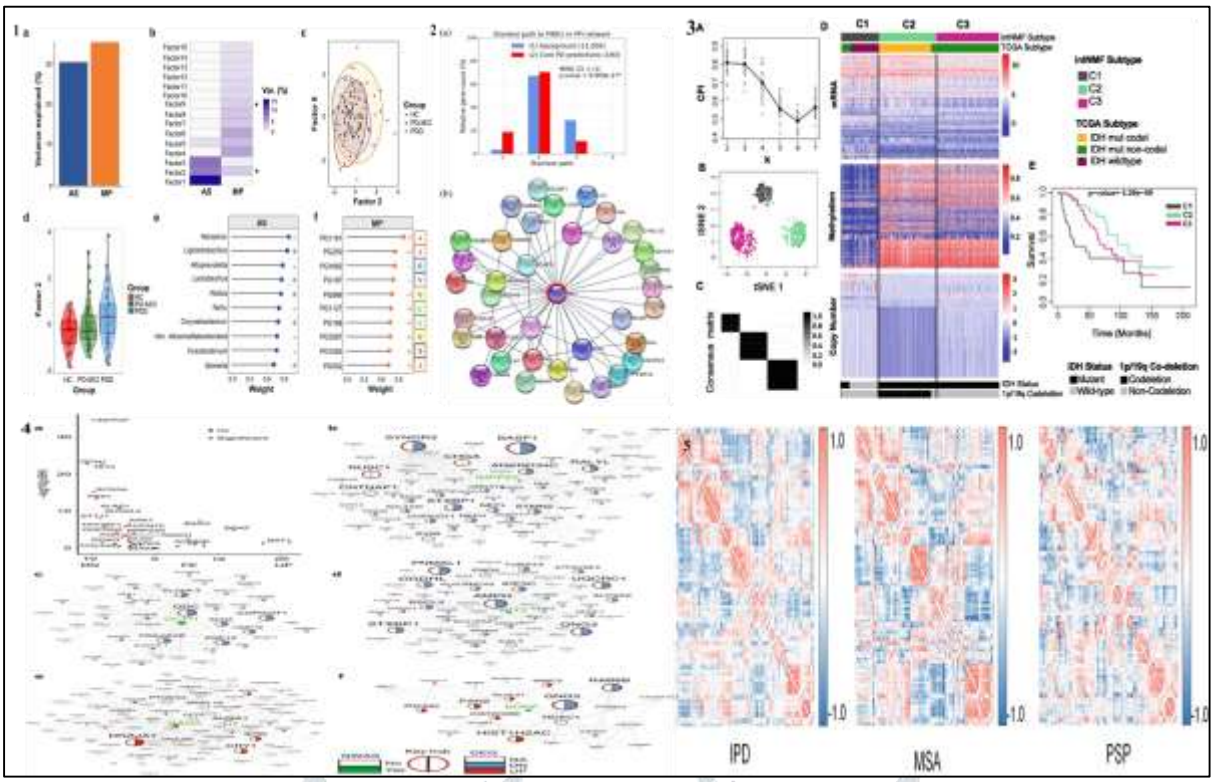


Figure.13 1(a) The measurement type (view) accounts for a percentage of the overall variation and (b) The percentage of the overall variation attributable to latent factors (LFs) 1-15. Asterisk (*) symbols indicate a statistically significant relationship between a factor and study groups. (c) Latent Factor 2 on the x-axis plotted against Latent Factor 9 on the y-axis, where each dot and ellipse is color coded based on group membership. (d) Latent Factor 2 data presented as box plots, categorised and colour-coded based on CI status. (e) Lollipop plot illustrating the most highly ranked genera in Latent Factor 2 based on per cent rank. (f) Lollipop plot illustrating the highest-ranking PGs from saliva as ranked by percent rank for Latent Factor 2; human-protein group data are colored orange, while bacterial-protein group data are colored blue. 2.Two sets of gene distributions were plotted in the PPI network based on the shortest path length of the Core PD gene set (in red) and the background gene set (in blue). The parentheses next to each set represent the number of available genes in that set. To determine if there is a statistically significant difference between the two distributions, we performed a one-tailed MWU test at an alpha level of (0.05) using the MWU value of (2). Figure 2B shows the PPI network of PINK1 and its 36 immediate neighbours that were predicted using the Core PD set of genes. Cytoscape 3.10.050 was used to visualise the gene network. 3. The three dimensional data of DNA methylation, RNA Expression, and Copy number Variations (CNV) used to make the Integrative Clustering of lower grade gliomas will (a) Show a Cluster prediction index based on the Cluster number and identify the best number of Cluster(s), (b) Provide a visual representation of the Subgroups with t-SNE, (c) Present a Sample Block Structure of the samples in the Consensus Matrix, which indicates the strength of the Consensus Matrix and supports stopping the clustering process and moving to analysis, (d) Provide Heat Maps of the mRNA, DNA methylation, and Copy Number Variations for Clustering while also showing the biological relevance of the IDH mutations and Codeletions, and (e) Provide Estimates of Survival Probabilities based on Kaplan Meier curves and Cox Proportional Hazards Log-Rank Test comparisons to allow evaluation of the difference in estimated survival probabilities between sites that do or do not have IDH mutations. 4. The three dimensional data of DNA methylation, RNA Expression, and Copy number Variations (CNV) used to make the Integrative Clustering of lower grade gliomas will (a) Show a Cluster prediction index based on the Cluster number and identify the best number of Cluster(s), (b) Provide a visual representation of the Subgroups with t-SNE, (c) Present a Sample Block Structure of the samples in the Consensus Matrix, which indicates the strength of the Consensus Matrix and supports stopping the clustering process and moving to analysis, (d) Provide Heat Maps of the mRNA, DNA methylation, and Copy Number Variations for Clustering while also showing the biological relevance of the IDH mutations and Codeletions, and (e) Provide Estimates of Survival Probabilities based on Kaplan Meier curves and Cox Proportional Hazards Log-Rank Test comparisons to allow evaluation of the difference in estimated survival probabilities between sites that do or do not have IDH

mutations.5. The metabolic network of different parkinsonian disorders based on Regional Radiomics Similarity Network (R2SN). **[24]**

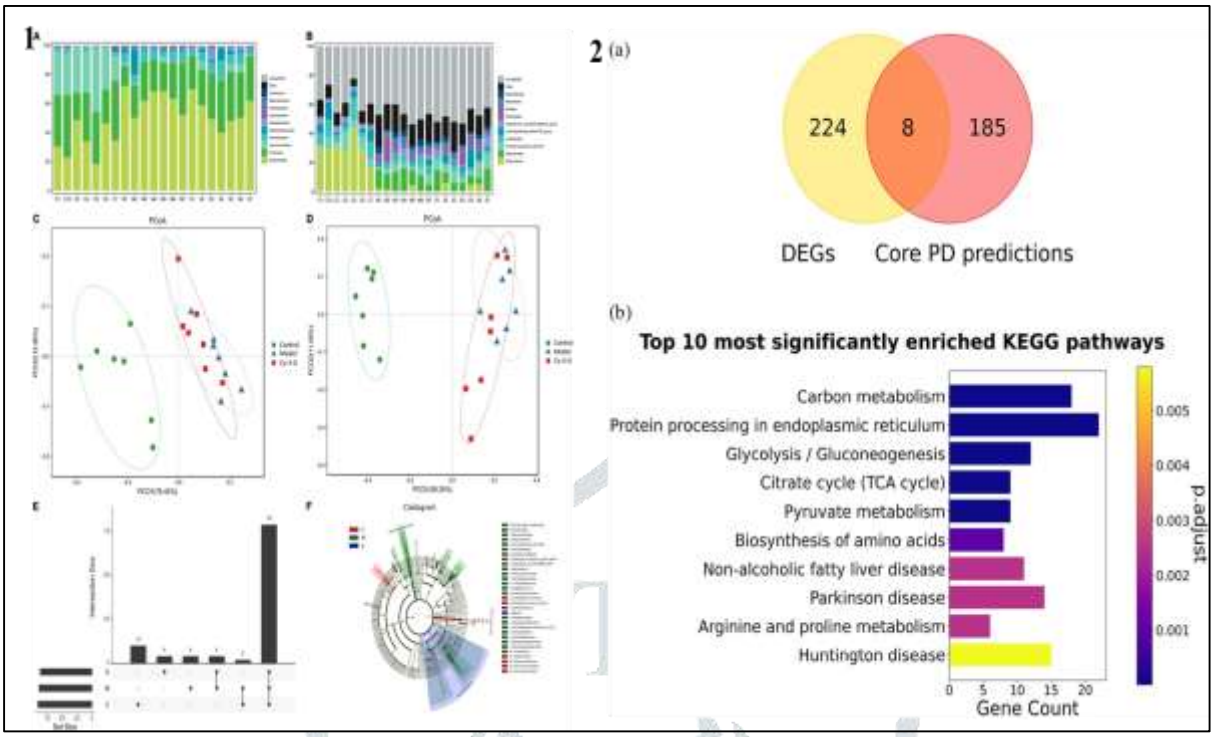


Figure.14 1. Using 16S rRNA sequencing, a histogram of species distributions was produced at the phylum level (A) and the genus level (B) for all three groups: Control, Model, and Cy-3-G, showing different colours representing different bacteria at the phylum or genus levels. We also generated PCoA plots depicting the differences between the Control, Model, and Cy-3-G Groups at the phylum (C) and genus (D) levels. We created an Upset plot (E) representing differences in microbiota composition among the three groups and produced a Cladogram (F) to show differences in the distribution of taxa among the three groups. 2.a) Venn diagram showing the overlap between DEGs from Novak et al.7 and Core PD predictions. b) The 10 KPs that were significantly enriched in the Core PD predictions, according to p.adj values (adjusted p-values), were calculated using a multiple hypothesis testing method developed by Benjamini and Hochberg. Gene Count indicates how many Core PD predictions were involved in a given KEGG pathway (KP).**[25]**

Additionally, network-based approaches include Similarity Network Fusion (SNF), Weighted Gene Co-expression Network Analysis (WGCNA), and probabilistic Bayesian network integration to build a multi-layer molecular network that represents common molecules linking genomic, transcriptional, and metabolic features. Further, Multi-omic Factor Regression (MFR), auto-encoder deep learning, graph neural networks (GNN), canonical correlation analyses (CCA), and multiple-kernel learning (MKL)-based regression and/or machine-learning techniques improve the ability of researchers to identify nonlinear relations, predict biomarkers and understand the pathology of disease through direct use of omics profiles.

5. Artificial Intelligence and Machine Learning in Drug Discovery

Artificial Intelligence (AI) and Machine Learning (ML) are transforming drug discovery by analyzing vast biological datasets, including genomics, transcriptomics, proteomics, metabolomics, and structural biology. These frameworks uncover hidden molecular relationships, predict drug-target interactions, and identify therapeutic leads. Neurodegenerative diseases like Parkinson's, driven by protein misfolding and network dysfunction, benefit from ML's ability to model complex, non-linear biological mechanisms. AI/ML approaches include supervised, unsupervised, semi-supervised, and deep learning models, with algorithms such as SVMs, Random Forests, and Gradient Boosting applied to diagnosis, biomarker prediction, and drug-response modeling. ML-based drug-target interaction frameworks integrate chemical descriptors, protein sequences, and structural data, enhancing predictive accuracy.

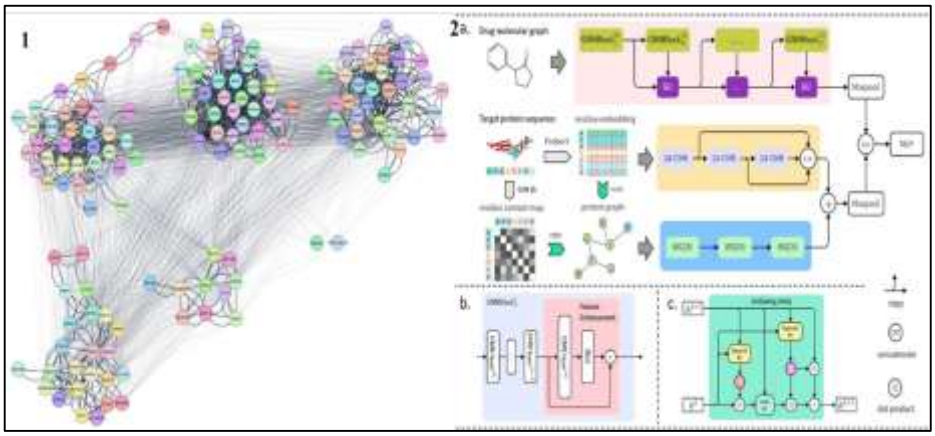


Figure.15 The structure of Mutationally Armoured Patterns for Parkinson's Disease (MAP4PD) gives a framework of MAP4PD clustered networks with variations. The lighter edges represent intercluster connections while the thicker edges represent intracluster connections. Utilising Graph Neural Networks for Drug Target Interaction (DPI) as outlined in Figure 15. (A) GNNBlocks DPI overall structure. In GNNBlocks, drug graphs and molecules are encoded through the modules, while in Target Sequences, embedding is performed through prot-BERT and through multi-scale convolutional neural networks. In terms of ligands, Target graphs, created with prot-BERT and ESM-1b, are embedded via weighted graph convolutional networks (GCN). (B) GNNBlock integrates feature enhancement from the last layer of GNN. (C) GNNBlock features that contain gating units (GU) placed between two consecutive GNNBlocks can be modified to increase feature representation. **[25]**

6. Deep Learning Applications to Structure-Based Drug design

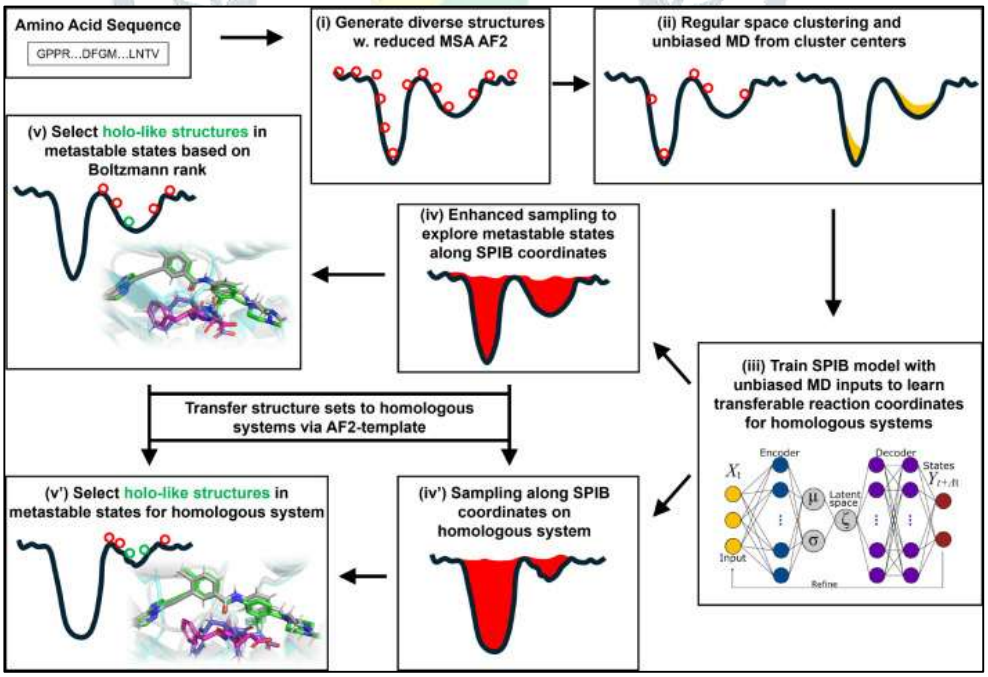


Figure.16 Flowchart of the AF2RAVE-Glide process: (i) reduced MSA AF2 provides way to create decoy structures through simulation of folding processes; (ii) combination of affinity clustering in space of decoys creates clusters; an initial MD simulation using CLUSTER will provide an unbiased start to perform subsequent simulations for structures with ligand affinity; (iii) the process of forming a predictive information bottleneck (P) is being used to determine reaction coordinates of the structures generated in (ii). These reactions provide a basis for the models of chemical free energy utilised to determine the free energy landscape. (iv) and (v): Activation or high activity states of holo-like-proteins that fall into the Boltzmann distribution

upon simulation of the ligand with holo-like proteins may provide a pathway for ligand design or induced-fit dockings/structures of ligand-bound proteins that fall into the Boltzmann distribution. For (iv') and (v'), the use of the MSA decoi sets and the P reaction coordinates of these using a similar MSA will produce results identical to those obtained from homologous proteins that are in the same MSA structural family. **[25]**

7. Case Study: 5A Polymorph of α-Synuclein

7.1 Structural and Biochemical Hallmarks of the 5A α-Synuclein Polymorph

The research team used cryo-electron microscopy to resolve a new structural variant of the human synuclein protein, which they named the 5A polymorph. The 5A polymorph was characterised by an entirely new protofilament fold and differs from other α-synuclein fibrils in terms of its unique β-sheet stacking pattern and inter-strand packing arrangement. As compared to other α-synuclein fibrils, the 5A polymorph has distinct surface topologies, differences in electrostatic potential distribution, as well as different exposure patterns of hydrophobic and charged residues along the fibrillar axis, and these variations are critical to determining how the 5A α-syn will form intermolecular interactions, stability as a fibril and ability to template seeds. Biochemically, the 5A polymorph possesses greater conformational stability and resistance to proteolytic degradation than previously identified strains of α-synuclein due to its ability to remain stable over long time periods within neurons, thus giving it the ability to alter the pathogenic effects of other strains of α-synuclein. Another key observation regarding the 5A polymorph is the presence of specific grooves, cavities, and interfaces on the fibrillar surface that are potential ligand-binding sites. The structures associated with these sites do not exist or are not available to be bound in other strains of α-synuclein. Given the above observations, it appears to be reasonable to conclude that targeting therapeutics using structured polymorphs of α-synuclein is preferable to general aggregation inhibition.[66,111]

7.2 In Silico Workflow for Polymorph-Specific Therapeutic Targeting

To systematically investigate the druggability of the 5A polymorph, we propose a modular in silico approach using BioPython as the primary computational tool. Initially, the structure is analysed using Bio.PDB for biologically relevant fibril sections, by removing any crystallography-related artefacts, and also calculating residue-level descriptors like solvent accessibility, secondary structure content, backbone dihedral distributions, and inter-residue contact networks. By using the methods outlined in the previous section, we will be able to identify regions that have a stable structure and specific ways for these structures to interact with ligands. The next step is to perform structure-based virtual screening of these candidate-binding sites using molecular docking software and evaluate them against small-molecule libraries containing FDA-approved drugs, BBB-permeable compounds, and other neuroactive compound libraries.[112] The output generated from this docking software will be placed into Python workflows to allow for the automatic retrieval of the corresponding binding affinity predictions, stability parameters of each docking pose, and the ligand interaction fingerprint for each of the ligands evaluated. The results of the molecular docking will be filtered using several criteria, including consensus scoring, reproducibility of predicted interactions across fibril regions, and physicochemical properties indicative of potential for CNS drug development. Using this multi-layered validation approach, we will determine whether the selected ligands are worthy of further investigation and whether they exhibit reliable, biochemically accurate interactions with the selected binding site.[113]

7.3 The Importance of Multiomics in Machine Learning-Based Compound Prioritisation for Therapy Development

Structural predictions alone cannot capture the biological complexity of Parkinson's disease (PD). Integrated docking with multi-omics approaches generates significant insights into drug-related pathways, linking features to PD-associated proteins involved in mitochondrial function, vesicle trafficking, lipid metabolism, proteostasis, and neuroinflammation. Metabolomic and lipidomic profiles further highlight the roles of membrane composition and energy homeostasis in the propagation of α-synuclein. Machine learning models combine 3D structural descriptors, ligand interactions, and pathway activities into ensemble frameworks. These models rank compounds by their specificity for polymorphs and their relevance to PD, enabling a holistic drug-prioritisation strategy that extends beyond affinity and aligns structural binding with disease-driving molecular networks.

7.4 Drug Repurposing and Identification of New Compounds

Advances in drug design, including virtual screening and fragment-based discovery, enable the identification of therapeutic options using high-resolution 3D structural data of proteins and binding partners. This approach is particularly valuable for drug repurposing, where compounds with established safety and pharmacokinetics can be tested against the 5A α-synuclein polymorph and dysregulated pathways in Parkinson's disease. Structural insights into the 5A fibril—its stability, seeding interfaces, and membrane interactions—support library-based screening for novel compounds. By addressing polymorphic fibril structures, this strategy overcomes protein aggregation challenges. Integrating multi-omics, AI, and BioPython pipelines emphasizes strain-specific targeting for precision drug discovery in synucleinopathies.

8. Current Challenges and Directions for Future Work

The integration of artificial intelligence (AI) with multi-omics data in Parkinson's disease research faces significant methodological and biological challenges. Omics datasets vary in scale, noise, and sample size, with genomics being relatively stable compared to the sparse and condition-dependent transcriptomic, proteomic, metabolomic, and lipidomic data. This imbalance complicates cross-layer biological interpretation. Most datasets rely on bulk tissue, limiting cell-type specificity and failing to capture spatial and temporal dynamics of α-synuclein pathology. AI models often use static snapshots, reducing generalizability across disease stages. Reproducibility, interpretability, and ethical concerns—such as privacy, consent, bias, and equitable access—remain critical barriers to precision medicine.

9. Conclusion and Future Outlook

Multi-omics data integrated with artificial intelligence (AI), machine learning (ML), and computational biology is transforming drug discovery for Parkinson's Disease (PD). Genomics, transcriptomics, proteomics, metabolomics, and lipidomics together provide a systems-level view of disease mechanisms that extends beyond the scope of any single dataset. AI-driven pipelines, built with tools like BioPython, enable the identification of molecular networks, biomarkers, and structure-based therapeutic targets. Central to this review is evidence that α-synuclein exists in multiple polymorphic forms, each with distinct toxicity and progression profiles. The 5A polymorph exemplifies how structural diversity informs precision medicine, advancing polymorph-specific drug discovery through integrated computational frameworks.

References

1. Henderson, M. X.; Trojanowski, J. Q.; Lee, V. M.-Y. α-Synuclein Pathology in Parkinson's Disease and Related α-Synucleinopathies. Neurosci. Lett. **2019**, 709, 134316. https://doi.org/10.1016/j.neulet.2019.134316

2. Bandbe, T., Johri, V., Kumari, U. (2025). Structure-guided Genome-wide Association Analysis of ALK Variants with GWAS Data Using R. *Computational Biology and Bioinformatics*, *13*(2), 72-85. https://doi.org/10.11648/j.cbb.20251302.13

3. The genetic architecture of Parkinson's disease - The Lancet Neurology. https://www.thelancet.com/journals/laneur/article/PIIS1474-4422(19)30287-X/abstract (accessed 2026-01-17).

4. Vineeta Johri, Uma Kumari "NGS-based polyglutamine expansion pattern detection using machine learning models to predict the conformation states of huntingtin" IOSR Journal of Pharmacy and Biological Sciences" Vol.20,Issue 5, 2025,10.9790/ 3008-2005031927

5. Uma kumari, Meenakshi Pradhan, Saptarshi Mukherjee, Sreyashi Chakrabarti, Ngs Analysis approach for neurodegenerative disease with Biopython",2023 ,volume 10,issue 9 , http://doi.one/10.1729/Journal.36043

6. Role of α-Synuclein Carboxy-Terminus on Fibril Formation in Vitro | Biochemistry. https://pubs.acs.org/doi/10.1021/bi027363r (accessed 2026-01-13).

7. Dehay, B.; Bourdenx, M.; Gorry, P.; Przedborski, S.; Vila, M.; Hunot, S.; Singleton, A.; Olanow, C. W.; Merchant, K. M.; Bezard, E.; Petsko, G. A.; Meissner, W. G. Targeting α-Synuclein for Treatment of Parkinson's Disease: Mechanistic and Therapeutic Considerations. Lancet Neurol. **2015**, 14 (8), 855–866. https://doi.org/10.1016/S1474-4422(15)00006-X.

8. Guardia-Laguarta, C.; Area-Gomez, E.; Rüb, C.; Liu, Y.; Magrané, J.; Becker, D.; Voos, W.; Schon, E. A.; Przedborski, S. α-Synuclein Is Localized to Mitochondria-Associated ER Membranes. J. Neurosci. **2014**, 34 (1), 249–259. https://doi.org/10.1523/JNEUROSCI.2507-13.2014.

9. Guerrero-Ferreira, R.; Taylor, N. M.; Mona, D.; Ringler, P.; Lauer, M. E.; Riek, R.; Britschgi, M.; Stahlberg, H. Cryo-EM Structure of Alpha-Synuclein Fibrils. eLife **2018**, 7, e36402. https://doi.org/10.7554/eLife.36402.

10. Sharma, M.; Burré, J. α-Synuclein in Synaptic Function and Dysfunction. Trends Neurosci. **2023**, 46 (2), 153–166. https://doi.org/10.1016/j.tins.2022.11.007.

11. Bendor, J.; Logan, T.; Edwards, R. H. The Function of α-Synuclein. Neuron **2013**, 79 (6), 10.1016/j.neuron.2013.09.004. https://doi.org/10.1016/j.neuron.2013.09.004.

12. Lee, H.-J.; Bae, E.-J.; Jang, A.; Ho, D.-H.; Cho, E.-D.; Suk, J.-E.; Yun, Y.-M.; Lee, S.-J. Enzyme-Linked Immunosorbent Assays for α-Synuclein with Species and Multimeric State Specificities. J. Neurosci. Methods **2011**, 199 (2), 249–257. https://doi.org/10.1016/j.jneumeth.2011.05.020.

13. Yamada, K.; Iwatsubo, T. Extracellular α-Synuclein Levels Are Regulated by Neuronal Activity. Mol. Neurodegener. **2018**, 13 (1), 9. https://doi.org/10.1186/s13024-018-0241-0.

14. Barnett, E. M.; Cassell, M. D.; Perlman, S. Two Neurotropic Viruses, Herpes Simplex Virus Type 1 and Mouse Hepatitis Virus, Spread along Different Neural Pathways from the Main Olfactory Bulb. Neuroscience **1993**, 57 (4), 1007–1025. https://doi.org/10.1016/0306-4522(93)90045-H.

15. Braak, H.; Tredici, K. D.; Rüb, U.; de Vos, R. A. I.; Jansen Steur, E. N. H.; Braak, E. Staging of Brain Pathology Related to Sporadic Parkinson's Disease. Neurobiol. Aging **2003**, 24 (2), 197–211. https://doi.org/10.1016/S0197-4580(02)00065-9.

16. Burré, J. The Synaptic Function of α-Synuclein. J. Park. Dis. **2015**, 5 (4), 699–713. https://doi.org/10.3233/JPD-150642.

17. The α-synuclein hereditary mutation E46K unlocks a more stable, pathogenic fibril structure | PNAS. https://www.pnas.org/doi/10.1073/pnas.1917914117 (accessed 2026-01-10).

18. Li, Y.; Zhao, C.; Luo, F.; Liu, Z.; Gui, X.; Luo, Z.; Zhang, X.; Li, D.; Liu, C.; Li, X. Amyloid Fibril Structure of α-Synuclein Determined by Cryo-Electron Microscopy. Cell Res. **2018**, 28 (9), 897–903. https://doi.org/10.1038/s41422-018-0075-x.

19. Boyer, D. R.; Li, B.; Sun, C.; Fan, W.; Zhou, K.; Hughes, M. P.; Sawaya, M. R.; Jiang, L.; Eisenberg, D. S. The α-Synuclein Hereditary Mutation E46K Unlocks a More Stable, Pathogenic Fibril Structure. Proc. Natl. Acad. Sci. **2020**, 117 (7), 3592–3602. https://doi.org/10.1073/pnas.1917914117.

20. Peelaerts, W.; Bousset, L.; Van der Perren, A.; Moskalyuk, A.; Pulizzi, R.; Giugliano, M.; Van den Haute, C.; Melki, R.; Baekelandt, V. α-Synuclein Strains Cause Distinct Synucleinopathies after Local and Systemic Administration. Nature **2015**, 522 (7556), 340–344. https://doi.org/10.1038/nature14547.

21. Corces, M. R.; Shcherbina, A.; Kundu, S.; Gloudemans, M. J.; Frésard, L.; Granja, J. M.; Louie, B. H.; Eulalio, T.; Shams, S.; Bagdatli, S. T.; Mumbach, M. R.; Liu, B.; Montine, K. S.; Greenleaf, W. J.; Kundaje, A.; Montgomery, S. B.; Chang, H. Y.; Montine, T. J. Single-Cell Epigenomic Analyses Implicate Candidate Causal Variants at Inherited Risk Loci for Alzheimer's and Parkinson's Diseases. Nat. Genet. **2020**, 52 (11), 1158–1168. https://doi.org/10.1038/s41588-020-00721-x.

22. Dou, L.; Xu, Z.; Xu, J.; Zang, C.; Su, C.; Pieper, A. A.; Leverenz, J. B.; Wang, F.; Zhu, X.; Cummings, J.; Cheng, F. A Network-Based Systems Genetics Framework Identifies Pathobiology and Drug Repurposing in Parkinson's Disease. Npj Park. Dis. **2025**, 11 (1), 22. https://doi.org/10.1038/s41531-025-00870-y.

23. Wang, W.; Zhu, G.; Wang, Y.; Li, W.; Yi, S.; Wang, K.; Fan, L.; Tang, J.; Chen, R. Multi-Omics Integration in Mice With Parkinson's Disease and the Intervention Effect of Cyanidin-3-O-Glucoside. Front. Aging Neurosci. **2022**, 14. https://doi.org/10.3389/fnagi.2022.877078.

24. Chen, J.; Fan, N.; Lu, Y.; Yang, J.; Song, W.; Sheng, H.; Yang, Y.; Chen, S.; Wang, J. Intelligence on the Graph: Graph Neural Networks for Mechanistic Drug Target Discovery. J. Pharm. Anal. **2025**, 101508. https://doi.org/10.1016/j.jpha.2025.101508.

25. Deng, G.; Shi, C.; Ge, R.; Hu, R.; Wang, C.; Qin, F.; Pan, C.; Mao, H.; Yang, Q. Efficient Substructure Feature Encoding Based on Graph Neural Network Blocks for Drug-Target Interaction Prediction. Front. Pharmacol. **2025**, 16. https://doi.org/10.3389/fphar.2025.1553743.