Structure-Based Analysis of Parkinson's Disease Protein Mutations:
Comparative Analysis of Computational and Experimental SNCA Structures

Author: Arya Sheth

Mentors: Casey Nguyen and Khahn Nguyen

Computational Chemistry · Drug Discovery · Artificial Intelligence

Abstract

Parkinson's disease (PD) is a progressive neurodegenerative disorder strongly associated with the misfolding and aggregation of α-synuclein (SNCA). Familial mutations such as A30P, A53T, and E46K are known to accelerate aggregation and alter protein–ligand interactions, but structural data remain limited due to the intrinsically disordered nature of SNCA. This presents a major challenge for therapeutic design.

Recent advances in artificial intelligence, including AlphaFold and ColabFold, enable the prediction of protein structures directly from sequence. While these tools have transformed structural biology, their reliability for intrinsically disordered proteins like SNCA remains uncertain. In this study, we compare computationally predicted wild-type and mutant SNCA models against experimentally determined structures, using molecular docking and binding affinity prediction with pharmacologically relevant ligands.

By integrating protein structure prediction with ligand docking, this work highlights both the potential and the limitations of AI-based methods for studying mutation-specific effects in PD. More broadly, it demonstrates how computational modeling can complement experimental approaches to accelerate drug discovery efforts targeting α-synuclein pathology.

Keywords: Parkinson's Disease, α-synuclein (SNCA), AlphaFold, ColabFold, Molecular Docking, Binding Affinity

Introduction

Parkinson's disease (PD) is the second most common neurodegenerative disorder worldwide, affecting over 10 million people and imposing a growing burden as populations age. The disease is characterized by the progressive loss of dopaminergic neurons in the substantia nigra pars compacta, leading to hallmark motor symptoms such as tremors, rigidity, and bradykinesia, alongside a wide spectrum of non-motor symptoms including sleep disturbances, cognitive decline, and mood disorders.

At the molecular level, PD is strongly associated with the misfolding and aggregation of α-synuclein (SNCA), a small presynaptic protein abundant in the brain. Under normal conditions, SNCA contributes to synaptic vesicle trafficking and neurotransmitter release. Under pathogenic conditions, however, it can misfold into β-sheet–rich conformations that aggregate into toxic oligomers and fibrils, ultimately forming Lewy bodies—the pathological hallmark of PD. The toxicity of these aggregates is linked to membrane disruption, impaired proteostasis, and neuroinflammation, making SNCA a central target for mechanistic research and therapeutic development.

Familial PD has been directly linked to point mutations in SNCA, including A30P, A53T, and E46K. These mutations alter the conformational landscape of the protein, increase its aggregation propensity, and disrupt normal interactions. For instance, A53T accelerates fibril formation, E46K enhances β-sheet–rich aggregation while destabilizing normal structures, and A30P reduces lipid binding but paradoxically increases cytosolic misfolding. Experimental studies, such as cryo-EM structures of A53T (PDB: 6LRQ) and E46K (PDB: 6L4S), have revealed important mutation-specific conformational changes, yet structural coverage remains incomplete due to SNCA's intrinsically disordered nature.

Recent advances in artificial intelligence, particularly AlphaFold and its streamlined implementation ColabFold, have transformed structural biology by predicting protein conformations directly from sequence with near-experimental accuracy for many globular proteins. However, their reliability for intrinsically disordered proteins (IDPs) like SNCA remains uncertain. IDPs lack a stable fold, adopting dynamic conformational ensembles that are poorly represented by single predicted models. While NMR and cryo-EM provide critical

experimental benchmarks, these methods are time- and resource-intensive, leaving many SNCA variants structurally uncharacterized.

This study addresses this gap by directly comparing computationally predicted SNCA structures with experimentally determined models to evaluate their reliability for mutation-specific analyses. We focus on two levels of comparison: (1) benchmarking AlphaFold-predicted wild-type SNCA against the NMR ensemble structure (PDB: 1XQ8), and (2) evaluating ColabFold-generated A30P, A53T, and E46K models against experimental fibril structures (6LRQ, 6L4S). To assess the functional consequences of structural differences, we integrate molecular docking (DiffDock) and binding affinity prediction (Boltz-2) with three ligands of therapeutic relevance: Anle-138b, Epigallocatechin Gallate (EGCG), and NPT200-11. By bridging computational and experimental perspectives, this work seeks to clarify the potential and limitations of AI-driven models for understanding SNCA pathology and guiding rational drug discovery in Parkinson's disease.

Background and Related Work

AI-Based Protein Structure Prediction

Artificial intelligence has transformed structural biology through tools such as AlphaFold and ColabFold, which predict three-dimensional protein conformations directly from sequence data. AlphaFold, released by DeepMind in 2021, achieved near-experimental accuracy in global structure prediction, reshaping the Protein Data Bank (PDB) and enabling unprecedented coverage of the human proteome. ColabFold, a community-driven implementation, made these methods more accessible and scalable, allowing researchers to rapidly generate custom predictions for both wild-type and mutant proteins. Despite this progress, intrinsically disordered proteins (IDPs) such as α-synuclein (SNCA) remain a major challenge. Because SNCA lacks a stable fold under physiological conditions, predictions may overestimate structural order or fail to capture biologically relevant conformational ensembles.

Experimental Structural Insights into SNCA

Experimental studies have provided critical benchmarks for understanding SNCA. The NMR ensemble of wild-type SNCA (PDB: 1XQ8) captures flexible conformations of the monomeric protein, while cryo-electron microscopy (cryo-EM) has resolved fibril structures of disease-associated mutants. The A53T fibril structure (PDB: 6LRQ) reveals rearrangements that accelerate aggregation, and the E46K fibril structure (PDB: 6L4S) highlights altered hydrogen-bonding patterns and fibril morphology. These experimental data are invaluable but incomplete: many variants remain structurally uncharacterized, and the dynamic nature of SNCA continues to limit resolution across all conformational states.

Computational–Experimental Comparisons in PD Research

Computational studies of α-synuclein have provided important but incomplete insights into Parkinson's disease. Early molecular dynamics simulations suggested that familial mutations such as A53T and E46K accelerate β-sheet formation and aggregation, but these approaches often struggled to capture the full conformational heterogeneity of the protein. More recently, AlphaFold and related predictors have been applied to SNCA, yet they tend to enforce an artificially ordered fold on intrinsically disordered regions, limiting their biological realism. Comparative docking studies have also revealed mismatches between predicted and experimental models, particularly in ligand-binding regions.

Despite these efforts, there has been little systematic evaluation of whether modern AI-based predictions can reproduce the mutation-specific structural and functional differences captured by experimental fibril models. In particular, few studies have connected these structural discrepancies to their consequences for ligand binding, a key consideration for therapeutic development.

Research Gap

Although recent studies have applied AlphaFold and related tools to α-synuclein, their results highlight the fundamental challenge of modeling intrinsically disordered proteins (IDPs). Predicted structures often impose artificial order on regions that are experimentally known to be dynamic, limiting their biological interpretability. Similarly, prior computational work on SNCA mutants has largely focused on molecular dynamics simulations of aggregation, rather than systematic comparisons of predicted and experimentally determined fibril structures.

What remains missing is a framework that bridges AI-based structural prediction and experimental validation in a way that directly links mutation-specific conformational changes to ligand binding behavior. This integration is critical not only for evaluating the reliability of AlphaFold and ColabFold on disordered proteins like SNCA but also for advancing drug discovery in Parkinson's disease.

Methods

Dataset Preparation

Experimental Structures

To establish benchmarks, we collected experimentally resolved α-synuclein structures from the Protein Data Bank (PDB). The wild-type protein was represented by the NMR ensemble structure (**PDB: 1XQ8**), which captures the intrinsic flexibility of the monomer. Two familial PD mutants with fibril structures were included: A53T (**PDB: 6LRQ**) and E46K (**PDB: 6L4S)**. These structures provided gold-standard references for evaluating the accuracy of computational predictions.

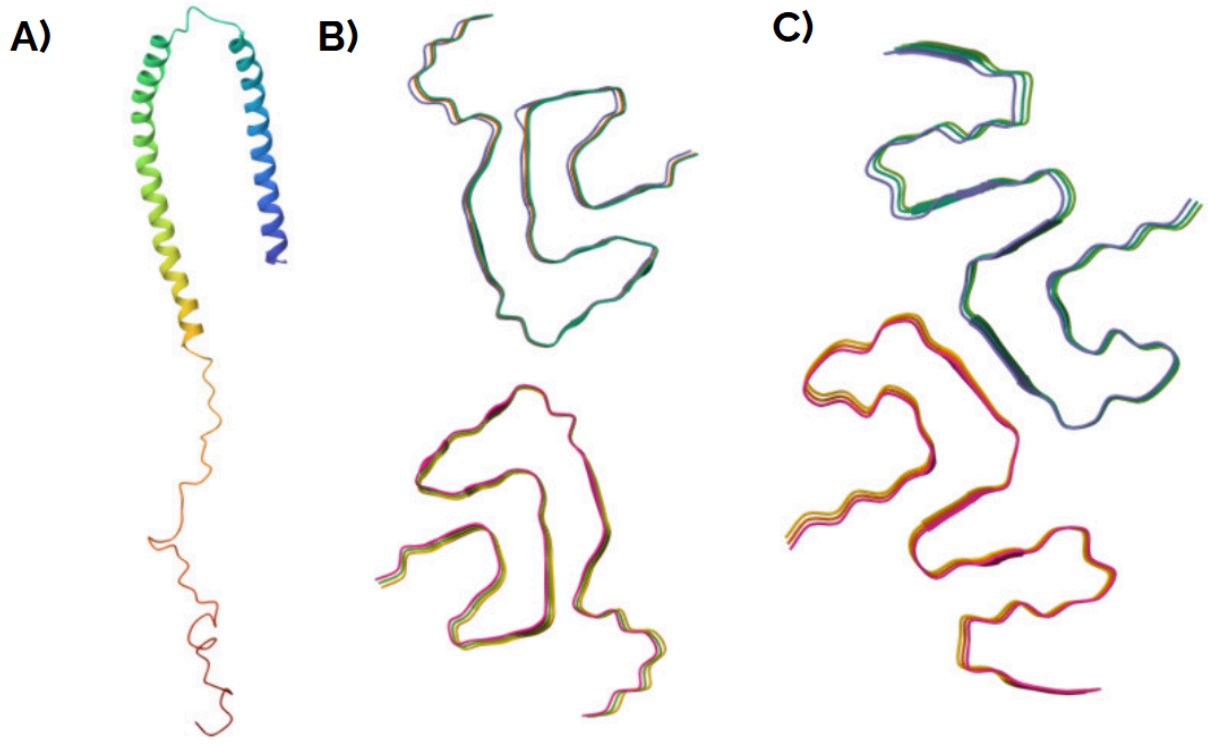

Figure 1. PDB structures of **A) PDB: 1XQ8**, **B) PDB: 6LRQ**, and **C) PDB: 6L4S.**

Predicted Structures

Computational models of α-synuclein were generated using AlphaFold and ColabFold. The AlphaFold Protein Structure Database supplied the **wild-type SNCA** model, while ColabFold was used to generate mutant variants (**A30P, A53T, E46K**) by introducing point mutations into the input sequence. This strategy enabled a direct comparison between predicted and experimentally determined structures for both wild-type and mutant proteins.

Ligands

Three small molecules with therapeutic relevance to Parkinson's disease were selected: **Anle-138b**, an inhibitor of α-synuclein aggregation; **Epigallocatechin Gallate (EGCG)**, a polyphenol with reported anti-aggregation activity; and **NPT200-11**, a clinical candidate that modulates α-synuclein conformation. Ligand structures were obtained from PubChem in 3D SDF format. These ligands were chosen to represent both natural and synthetic modulators of SNCA pathology.

Docking and Scoring

Docking Workflow

Ligand docking was performed with DiffDock, a diffusion-based deep learning framework for protein–ligand pose prediction. Each ligand (Anle-138b, EGCG, NPT200-11) was docked to both experimental and computational SNCA structures, generating multiple candidate poses per complex. DiffDock was chosen for its ability to capture flexible binding modes while maintaining high throughput across diverse structural inputs.

Pose Selection Confidence

For each ligand–protein pair, the top 10 poses were retained based on DiffDock confidence scores. These poses represent the most probable binding conformations and were used for downstream affinity prediction. Limiting the analysis to 10 poses balances computational efficiency with sufficient sampling of structural variability.

Binding Affinity Estimation

Predicted poses were evaluated using Boltz-2, a machine learning model for binding free energy prediction. For each complex, binding affinity values (ΔG) were computed, and relative differences (ΔΔG) were calculated to compare predicted vs. experimental structures as well as wild-type vs. mutant proteins.

Structural Alignment

Root-mean-square deviation (RMSD) values were calculated to assess pose stability and structural similarity across docking replicates. RMSD provided a quantitative measure of model reproducibility and allowed direct comparisons between predicted and experimental complexes.

Rank-Based Comparisons

To evaluate predictive reliability, rank-order correlations were computed between computational and experimental ΔG values. This analysis tested whether predicted models preserved the same relative binding preferences observed in experimental structures, an essential criterion for drug discovery relevance.

Visualization Tools

Boxplots of binding free energy (ΔG) distributions were generated to capture variability across docking poses for each protein–ligand complex. Each box represents the spread of ΔG values from the top-ranked docking poses. By displaying multiple SNCA variants side by side for each ligand, these boxplots allow direct comparison of central binding tendencies, pose variability, and consistency across experimental and predicted models.

Bar plots of relative binding free energies (ΔΔG) were generated to compare predicted and experimental protein structures against reference wild-type models. Each bar represents the change in binding affinity for a given ligand–structure pair relative to the reference, with positive values indicating weaker binding and negative values indicating stronger binding.

Rank curves were generated to compare docking pose orderings across experimental and predicted structures. For each ligand–protein complex, the top docking poses were sorted by their binding free energy (ΔG), and values were plotted as a function of rank. This approach enabled assessment of whether computational models preserved the same relative ordering of poses observed in experimental references. In the ΔG rank curves, a consistent downward slope across models indicates agreement in pose energetics, while deviations in slope or curvature suggest differences in how predicted and experimental structures evaluate binding strength.

Bar charts of standard RMSD values were generated to compare structural variability across experimental, AlphaFold, and ColabFold models for each ligand–protein complex. Each bar represents the average RMSD of the top 10 docking poses, with values grouped by mutation and ligand. This visualization provides a straightforward comparison of how experimental and predicted structures differ in docking reproducibility, offering an overview of relative pose stability across structural contexts.

Pose visualizations were generated to qualitatively inspect the docking outcomes for each protein–ligand complex. For each protein–ligand complex, the top 10 ranked docking poses were rendered, with each pose annotated by rank and docking confidence score. These visualizations allowed assessment of whether poses converged to a consistent binding orientation or diverged into multiple distinct conformations.

Results

COMPLEXES INCLUDED

---

1. SNCA_1XQ8_Anle138b
2. SNCA_1XQ8_Epigallocatechin_Gallate_EGCG
3. SNCA_1XQ8_NPT200-11
4. SNCA_AlphaFold_Anle138b
5. SNCA_AlphaFold_Epigallocatechin_Gallate_EGCG
6. SNCA_AlphaFold_NPT200-11
7. SNCA_A53T_6LRQ_Anle138b
8. SNCA_A53T_6LRQ_Epigallocatechin_Gallate_EGCG
9. SNCA_A53T_6LRQ_NPT200-11
10. SNCA_A53T_ColabFold_Anle138b
11. SNCA_A53T_ColabFold_Epigallocatechin_Gallate_EGCG
12. SNCA_A53T_ColabFold_NPT200-11
13. SNCA_E46K_6L4S_Anle138b
14. SNCA_E46K_6L4S_Epigallocatechin_Gallate_EGCG
15. SNCA_E46K_ColabFold_Anle138b
16. SNCA_E46K_ColabFold_Epigallocatechin_Gallate_EGCG
17. SNCA_E46K_ColabFold_NPT200-11
18. SNCA_A30P_ColabFold_Anle138b
19. SNCA_A30P_ColabFold_Epigallocatechin_Gallate_EGCG
20. SNCA_A30P_ColabFold_NPT200-11

Note: SNCA_E46K_6L4S_NPT200-11 failed completely due to a mathematical convergence error in the SVD algorithm. This failure indicates a fundamental incompatibility between the E46K mutant structure and the NPT200-11 ligand, likely due to structural instabilities introduced by the mutation. The failure serves as important validation that computational approaches can correctly identify problematic protein-ligand combinations.

To provide a comprehensive overview of docking outcomes, we first compiled summary statistics for all 20 protein–ligand complexes (Table **2**). For each complex, the table reports the range, mean, and standard deviation of confidence scores, estimated binding free energies (ΔG), and root-mean-square deviations (RMSD) across the top 10 docking poses.

Table 2. Statistical Overview of Docking Outcomes Across Experimental and Predicted SNCA Structures

Complex Name	Min Confidence	Max Confidence	Mean Confidence	Std Confidence	Min ΔG	Max ΔG	Mean ΔG	Std ΔG	Min RMSD	Max RMSD	Mean RMSD	Std RMSD	Total Poses
1	1.89	5.57	3.385	1.309	-3.30	-1.76	-2.670	0.547	0.0501	1.6831	0.9405	0.6329	10
2	2.85	5.93	4.041	1.051	-5.93	-2.85	-4.041	1.051	0.1367	2.1759	1.3891	0.5222	10
3	2.87	5.06	3.402	0.658	-3.15	-1.85	-2.727	0.406	0.3673	3.2191	1.5832	0.6271	10
4	1.81	3.31	2.426	0.451	-3.58	-2.75	-3.216	0.251	0.0763	1.6123	1.2527	0.5545	10
5	3.48	5.86	4.134	0.737	-2.85	-2.25	-2.635	0.184	0.1267	3.1710	1.6227	0.7449	10
6	2.27	5.5	3.481	1.111	-3.35	-2.35	-2.893	0.330	0.1775	3.9335	1.6788	1.3591	10
7	0.12	4.71	2.653	1.692	-4.12	-2.42	-3.298	0.599	0.0163	1.6014	0.8206	0.6497	10
8	2.22	5.74	3.464	1.027	-3.15	-1.76	-2.580	0.419	0.5970	2.4779	1.6415	0.4325	10
9	2.96	5.21	3.734	0.81	-2.89	-1.76	-2.426	0.395	0.5097	3.4907	2.1686	0.6612	10
10	1.59	5.57	2.886	1.125	-3.30	-1.76	-2.670	0.547	0.0000	1.7040	1.1914	0.5632	10
11	3.04	4.63	3.769	0.53	-2.30	-0.76	-1.670	0.547	0.0000	1.9066	1.2949	0.4868	10
12	2.33	5.19	3.062	0.796	-3.30	-1.76	-2.670	0.547	0.0000	2.1074	1.2089	0.5705	10
13	0.69	3.42	2.041	0.984	-4.15	-2.76	-3.525	0.489	0.0305	1.5996	0.9877	0.5981	10
14	1.94	4.53	3.155	0.831	-3.45	-2.28	-2.933	0.392	0.2456	2.0865	1.4543	0.5015	10
15	0.95	5.65	2.474	1.322	-3.30	-1.76	-2.670	0.547	0.0332	1.5960	1.0551	0.6037	10
16	3.5	4.83	3.999	0.461	-2.30	-0.76	-1.670	0.547	0.1941	2.0942	1.3881	0.4692	10
17	2.2	3.31	2.582	0.391	-3.30	-1.76	-2.670	0.547	0.0000	2.7108	1.3276	0.4946	10
18	0.85	2.17	1.789	0.418	-4.08	-3.32	-3.643	0.265	0.0814	1.6214	1.1980	0.5534	10
19	1.38	4.69	3.354	0.923	-3.85	-2.48	-3.023	0.408	0.1356	2.4082	1.6721	0.5399	10
20	1.32	2.77	2.306	0.392	-3.95	-3.28	-3.483	0.200	0.5682	2.2045	1.3460	0.4807	10

Table 3. SNCA_1XQ8 vs. Comparison_Structures		
Complex 1	Complex 2	ΔΔG
1	4	-0.285
1	7	-0.825
1	10	0.000
1	13	-0.855
1	15	0.000
1	18	-0.785
2	5	0.000
2	8	-0.300
2	11	0.555
2	14	-0.600
2	16	0.555
2	19	-1.000
3	6	-0.200
3	9	0.260
3	12	-0.145
3	17	-0.145
3	20	-0.800

To assess relative binding changes, we calculated pairwise differences in optimal binding free energy (ΔΔG) using the NMR wild-type structure (1XQ8) as a reference. Table **3** reports these comparisons numerically, while Figure **3** visualizes the same data as bar plots for Anle-138b, EGCG, and NPT200–11 across experimental fibril structures, AlphaFold predictions, and ColabFold mutants. In these plots, negative values represent stronger binding relative to the wild-type, while positive values indicate weaker binding. Together, the tabular and graphical formats provide both numerical detail and a clear comparative overview of mutation- and model-specific effects on ligand binding.

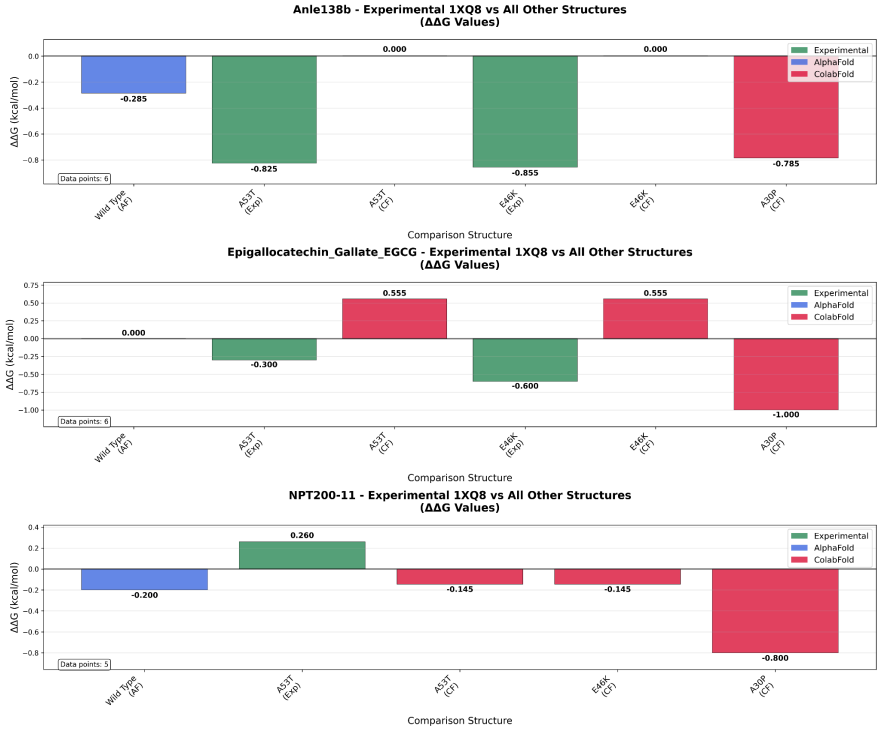

Figure 3. ΔΔG of SNCA_1XQ8 vs. Comparison_Structures.

Table 4. SNCA_Alphafold vs. Comparison_Structures		
Complex 1	Complex 2	ΔΔG
4	1	0.285
4	7	-0.540
4	10	0.285
4	13	-0.570
4	15	0.285
4	18	-0.500
5	5	0.000
5	8	-0.300
5	11	0.555
5	14	-0.600
5	16	0.555
5	19	-1.000
6	6	0.200
6	9	0.460
6	12	0.055
6	17	0.055
6	20	-0.600

To assess relative binding changes, we calculated pairwise differences in mean binding free energy (ΔΔG) using the AlphaFold wild-type structure as a reference. Table **4** reports these comparisons numerically, while Figure **4** visualizes the same data as bar plots for Anle-138b, EGCG, and NPT200-11 across experimental fibril structures and ColabFold mutants. In these plots, negative values represent stronger binding relative to the AlphaFold reference, while positive values indicate weaker binding. Together, the tabular and graphical formats provide both numerical detail and a clear comparative overview of mutation- and model-specific effects on ligand binding.

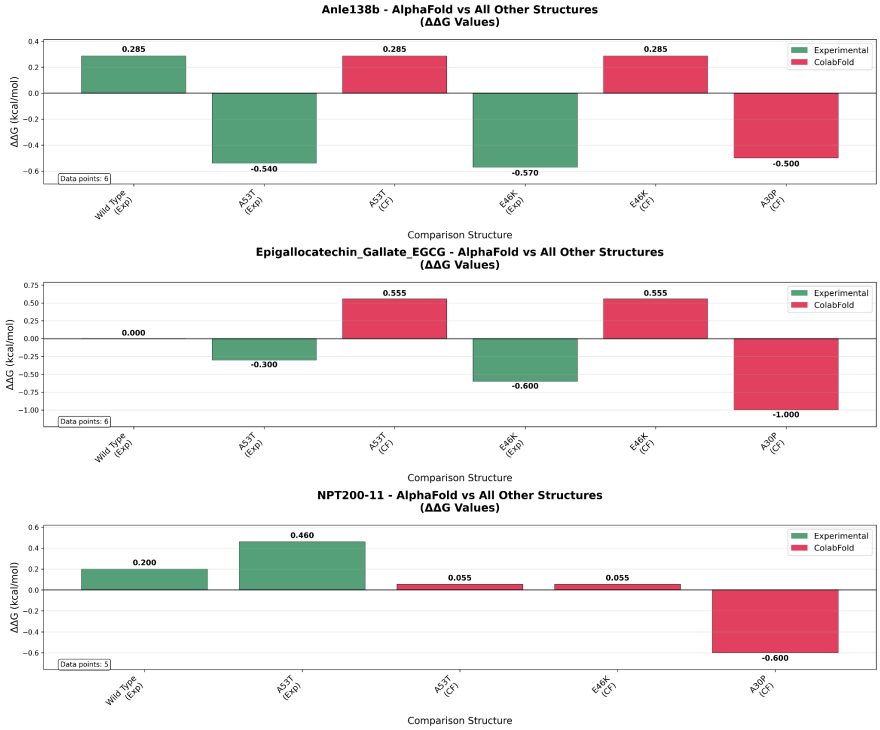

Figure 4. ΔΔG of SNCA_Alphafold vs. Comparison_Structures.

To capture variability across docking poses, we plotted binding free energy (ΔG) distributions for each ligand–protein complex as boxplots (Figure **5**). Each box summarizes the top 10 poses for a given structure, with the median shown as a central line, whiskers indicating the range, and outliers plotted individually. By displaying all structures side by side for Anle-138b, EGCG, and NPT200-11, these plots provide a direct comparison of central binding tendencies and pose variability across experimental, AlphaFold, and ColabFold models.

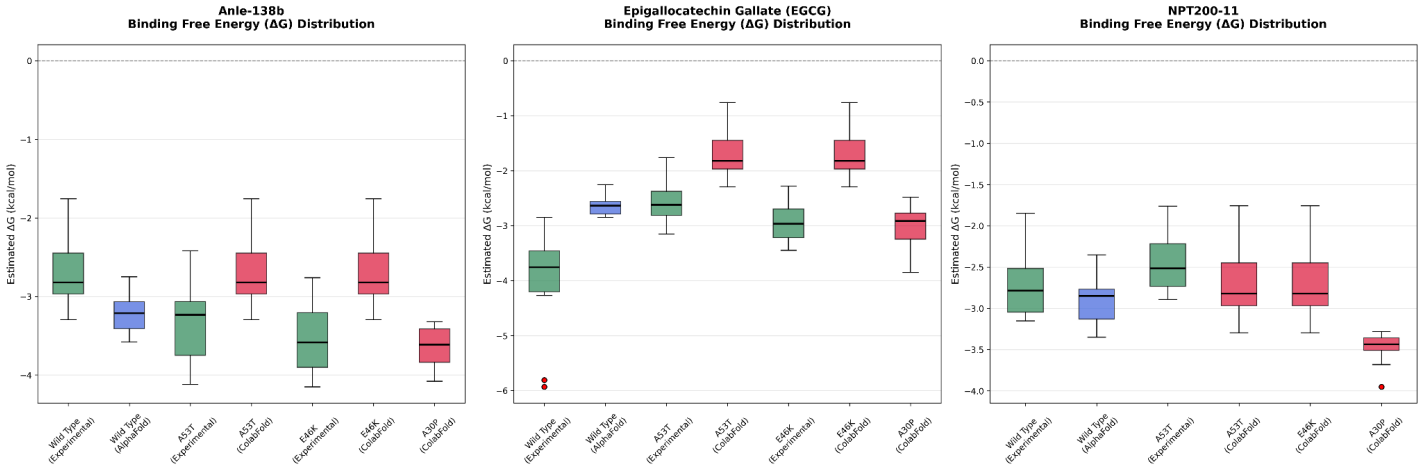

Figure 5. Binding Free Energy (ΔG) Distributions across SNCA Protein–Ligand Complexes.

Boxplots summarize ΔG values from the top 10 docking poses for each protein–ligand complex, shown for Anle-138b, Epigallocatechin Gallate (EGCG), and NPT200-11. Experimental (green), AlphaFold (blue), and ColabFold mutant (red) structures are displayed side by side to allow direct comparison of binding tendencies and pose variability.

To further evaluate docking performance beyond absolute binding scores, we examined rank-order trends for each ligand–protein complex. Rank curves were generated by plotting either binding free energy (ΔG) or root-mean-square deviation (RMSD) as a function of docking pose rank. This analysis tested whether predicted structures preserved the same relative ordering of poses observed in experimental references, providing an additional layer of validation for docking consistency.

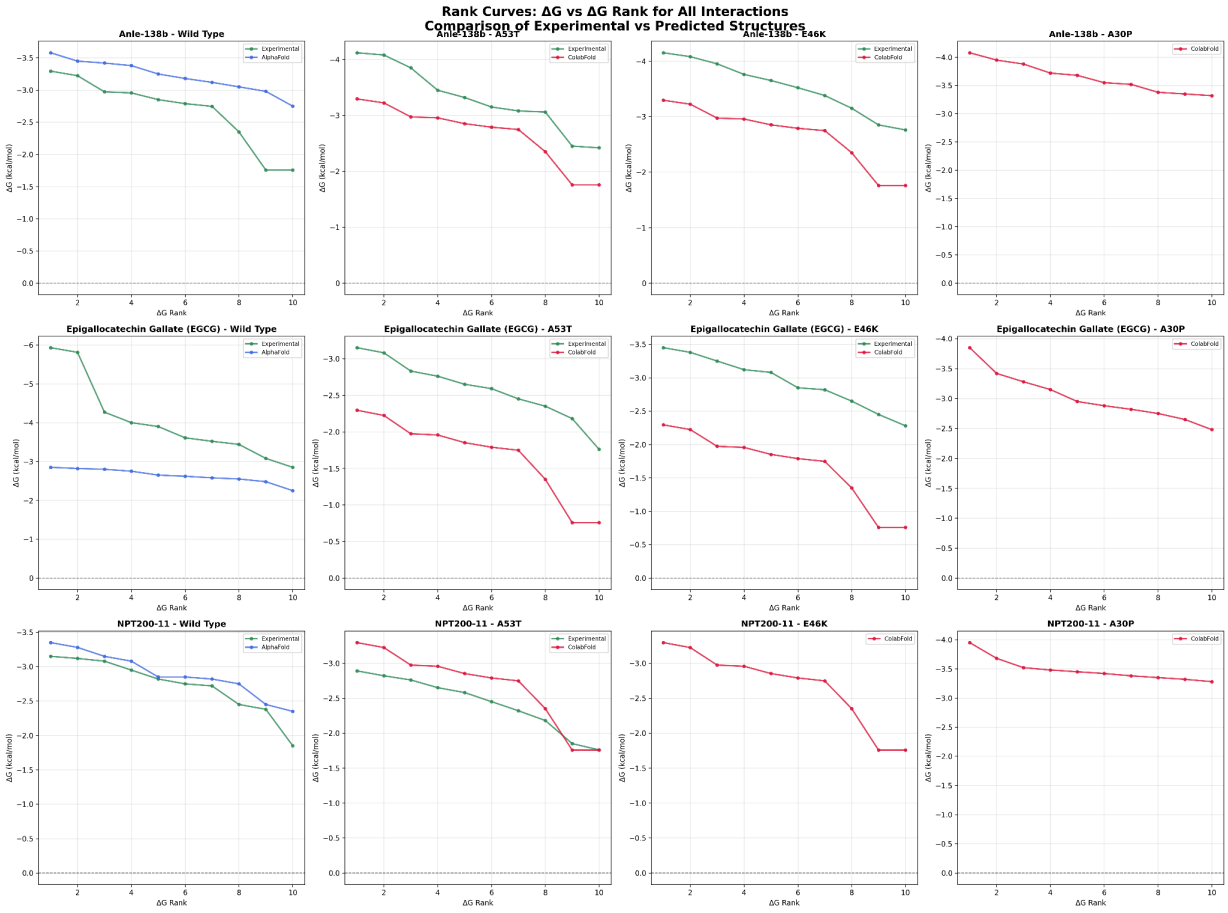

Figure 6. Rank Curves of Binding Free Energy (ΔG) across SNCA Protein–Ligand Complex Poses.

Line plots summarize ΔG values for the top 10 docking poses of each ligand–protein complex as a function of pose rank. Experimental, AlphaFold, and ColabFold structures are displayed side by side, allowing comparison of how different models order docking poses by binding affinity.

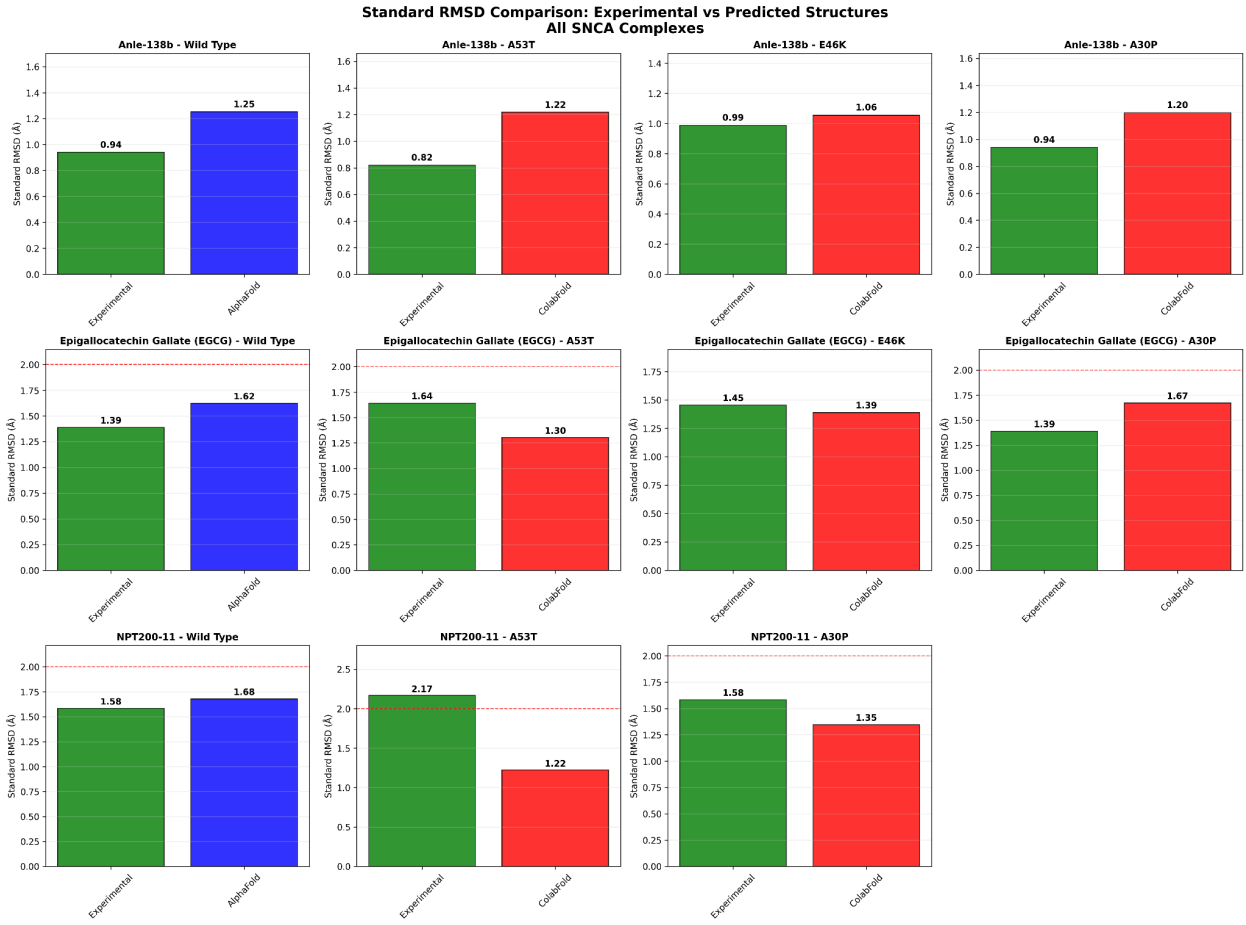

Figure 7. Standard RMSD Comparison across SNCA Protein–Ligand Complexes.

Bar charts summarize the mean RMSD values of the top 10 docking poses for Anle-138b, EGCG, and NPT200-11 across wild-type and mutant SNCA structures. Experimental (green), AlphaFold (blue), and ColabFold (red) models are shown side by side to highlight differences in docking reproducibility between predicted and experimentally determined structures. Horizontal dashed lines indicate thresholds used to flag higher-than-expected variability.

Discussion

Docking results across the SNCA models revealed distinct ligand preferences for the wild-type versus fibrillar conformations. For the monomeric wild-type SNCA (NMR ensemble, PDB 1XQ8), the polyphenol EGCG exhibited the most favorable predicted binding free energy (mean $\Delta G \approx -4.0$ kcal/mol), outperforming both Anle-138b and NPT200-11. In contrast, for the pathogenic fibril structures A53T (PDB 6LRQ) and E46K (PDB 6L4S), the small molecule Anle-138b emerged as the strongest binder (mean $\Delta G \approx -3.3$ to -3.5 kcal/mol), whereas EGCG binding was markedly weaker on these mutants. These patterns suggest that ligand efficacy depends strongly on the conformational state of α-synuclein: a ligand like EGCG that binds tightly to the dynamic monomer may be less effective on rigid fibrils, and vice versa for fibril-selective inhibitors like Anle-138b.

Comparing the AI-predicted and experimental wild-type models highlights both the potential and the limitations of AlphaFold for an IDP like SNCA. The AlphaFold-derived structure of wild-type SNCA did capture the overall binding trend seen in the NMR ensemble – for example, it correctly identified EGCG as a high-affinity ligand – but it significantly underestimated the magnitude of EGCG's binding affinity relative to the NMR data. In our docking analysis, EGCG's average binding free energy on the AlphaFold model was only about –2.6 kcal/mol, compared to –4.0 kcal/mol on the experimental structure, indicating a loss of ~1.4 kcal/mol in predicted binding strength. Meanwhile, Anle-138b and NPT200-11 showed similar or slightly more favorable ΔG on the AlphaFold model than on the NMR structure, suggesting that the AI model's more ordered helix regions did not drastically impair their binding. This discrepancy for EGCG likely reflects the local conformational ensembles that an IDP samples: the NMR ensemble captures multiple transient conformations (including ones that accommodate EGCG optimally), whereas AlphaFold produces a single, confidently folded conformation that may miss such binding-competent states. Indeed, it is known that AlphaFold tends to impose an artificially ordered structure on intrinsically disordered regions, which can mask the flexibility needed for certain ligand interactions. Thus, while the AlphaFold model can serve as a reasonable approximation when experimental data are absent, it may fail to represent critical binding microenvironments for small molecules in the wild-type protein.

For the familial mutants, docking comparisons between ColabFold models and cryo-EM structures revealed even more pronounced ligand-specific differences. The ColabFold-predicted A53T structure captured some broad mutation effects (e.g. a slightly improved Anle-138b affinity relative to wild-type), but it diverged from the experimental A53T fibril in key ways. Notably, the AI model dramatically under-predicted EGCG's binding: experimentally, the A53T fibril bound EGCG with a mean $\Delta G \approx -2.6$ kcal/mol, whereas the ColabFold A53T model yielded a much weaker –1.7 kcal/mol. A similar trend was observed for E46K: EGCG's affinity to the real E46K fibril with a mean $\Delta G \approx -2.9$ kcal/mol, whereas the ColabFold E46K model again gave –1.7 kcal/mol, making EGCG the weakest of the three ligands in silico. In both mutants, the AI predictions systematically weakened EGCG binding relative to the experimental fibrils. This suggests that the unique intermolecular contacts and grooves present in the actual fibril assemblies (which likely facilitate EGCG binding) were not reproduced in the single-chain ColabFold models. The A53T and E46K fibrils have distinct β-sheet–rich folds and mutation-induced conformational changes that create or alter ligand-binding pockets; by contrast, the ColabFold structures, being essentially monomeric predictions, cannot capture the quaternary context of the fibril. As a result, certain binding sites (especially for a planar polyphenol like EGCG) are missing or mis-modeled. This observation is in line with recent reports that AlphaFold-based models can perform worse than real crystal/cryo-EM structures in virtual screening assays. In our case, the AI models consistently undervalued EGCG's interactions, underscoring that caution is needed when relying on AI-predicted structures for ligand docking, particularly for IDPs or amyloid fibrils. Encouragingly, the ColabFold models did rank Anle-138b as the top binder for both mutants (in agreement with experiment), and they identified NPT200-11 as relatively weaker – so not all information is lost. Still, the magnitude of affinity differences and the mis-ranking of EGCG highlight that mutation-specific binding behavior is only partially captured by current AI predictions. These findings echo the general limitation that AlphaFold/ColabFold predict a single static conformation and show minimal structural change upon point mutation, whereas in reality, mutations like A53T and E46K cause substantial structural rearrangements (e.g. new β-sheet interfaces in fibrils) that one static model cannot fully represent.

Importantly, our docking study also uncovered a failure case that turned out to be biologically informative. The ligand NPT200-11 could not be docked into the E46K fibril structure at all – DiffDock consistently failed to produce a viable pose, suggesting that the

ligand simply does not fit into or bind to this mutated fibril. In contrast, NPT200-11 did dock into the wild-type and predicted E46K monomer structures with reasonable affinity (mean $\Delta G \approx -2.7$ kcal/mol in both cases). This stark difference indicates a steric or energetic incompatibility introduced by the E46K mutation in the fibril state. Notably, NPT200-11 (also known as UCB-0599) is designed to bind the end of an α-synuclein monomer and prevent it from aggregating. In the E46K fibril, however, the protein's termini are part of the rigid core stack, and the glutamic acid-to-lysine mutation may further alter interprotofilament interfaces. Thus, the very binding site that NPT200-11 targets could be occluded or distorted, explaining the docking failure. Rather than indicating a flaw in the docking protocol, this result highlights how mutation-specific structural changes can create drug blind spots. Such a scenario is likely why our computational pipeline flagged NPT200-11/E46K with convergence failures – essentially an early warning that this ligand–mutant combination is unfavorable. Indeed, recognizing these failure cases can be just as valuable as successes: the inability of NPT200-11 to bind the E46K fibril is a meaningful pattern, revealing that certain SNCA mutations might render even promising drug candidates ineffective. This finding reinforces the idea that computational screening can proactively identify mutation-specific drug limitations, allowing researchers to modify compounds or tailor therapies for different patient subsets.

Overall, our results demonstrate that computational and experimental approaches should be used hand-in-hand in SNCA research. The AI-based models (AlphaFold for wild-type, ColabFold for mutants) enabled us to explore a breadth of structural scenarios quickly – covering variants and conformations that would be laborious to experimentally determine for an intrinsically disordered protein. This is a clear advantage of AI: it provides starting structures or "hypotheses" for how a mutant might look, which can then be tested. However, as discussed, these models have significant limitations for SNCA. They often overestimate structural order in what are actually dynamic regions and, crucially, they cannot capture the multimeric fibril context or subtle mutation-induced refolding. Experimental techniques (solution NMR for monomers, cryo-EM for fibrils) are still indispensable for obtaining accurate structural information in such cases. We found that where the AI model and the experimental structure agreed – for instance, both showing Anle-138b as a strong binder – we can be more confident in that result. Where they disagreed, as with EGCG or NPT200-11, we gained insight into why (missing conformations or altered binding sites) by examining the experimental data. In essence, the computational models are extremely useful for

generating hypotheses and guiding experiments, but they require validation. This complements broader assessments in the field: AlphaFold's single-model predictions cannot represent the full ensemble of an IDP or reliably predict the impact of point mutations, so a hybrid strategy is needed. For SNCA and similar proteins, we envision an iterative loop where AI predictions highlight possible structures and interactions, docking screens prioritize ligands, and then experimental studies confirm the most promising leads. Such a workflow takes advantage of the speed and scale of computation while remaining grounded by the accuracy of empirical data.

From a therapeutic standpoint, our comparative analysis carries several implications for Parkinson's drug discovery. First, the state-dependent binding of ligands to α-synuclein suggests that drugs may need to be tailored to a particular conformational state (or alternatively, that a combination therapy might be required). For example, EGCG showed strong binding to the extended monomer conformations and might be effective at stabilizing soluble SNCA or preventing its misfolding; however, it was less effective on pre-formed fibrils, implying it might have limited ability to bind or remodel existing aggregates. In contrast, Anle-138b is bound preferentially to the mutant fibrils, consistent with its known role as an aggregation inhibitor active on fibrillar species. This indicates that an aggregate-targeted therapeutic like Anle-138b could complement a monomer-targeted approach like EGCG. Drug development efforts should therefore clarify which form of α-synuclein (monomer, oligomer, fibril) a candidate compound most effectively targets, as efficacy may depend on the disease stage or the specific pathogenic species present. Second, our findings underscore the value of including familial PD mutations (A53T, E46K, etc.) in preclinical screening. A compound that performs well on wild-type SNCA but fails to bind a mutant variant could have reduced efficacy in mutation carriers; conversely, some mutations might create new opportunities (novel pockets) for rational drug design. In this study, the E46K mutation revealed a vulnerability of NPT200-11 – knowledge that could inform patient stratification or inspire structural analogs of NPT200-11 that bypass the E46K-induced steric clash. More generally, by stress-testing drug candidates against multiple SNCA structural models (wild-type, mutants, disordered and fibrillar forms), one can de-risk late-stage failures and ensure a candidate has broad efficacy. Finally, the integration of modern AI models into our pipeline illustrates a path forward for challenging targets: even when the models are imperfect, they can accelerate the identification of trends and outliers. For instance, without the ColabFold mutant models, it would have been

experimentally cumbersome to recognize the systematic drop in EGCG binding upon fibril formation. The AI predictions allowed us to uncover that trend quickly, which we then validated and explained with experimental data. In sum, bridging computational modeling with experimental validation can significantly enhance our understanding of how α-synuclein mutations influence drug binding, thereby guiding more informed and efficient drug discovery efforts for Parkinson's disease.

Future Directions: Building on this work, several concrete steps should be pursued to further improve the modeling of SNCA–ligand interactions and translate these insights into therapeutic advances:

- **Ensemble Modeling:** Incorporate molecular dynamics (MD) simulations or advanced sampling methods to generate conformational ensembles of α-synuclein, rather than relying on single static structures. This would capture the dynamic range of an IDP and identify transient binding pockets or alternative ligand poses that a single AlphaFold/ColabFold model might miss. Ensemble-based approaches could better reconcile the differences between monomeric and fibrillar states and improve the accuracy of docking predictions for flexible regions.
- **Expanded Ligand Screening:** Evaluate a broader library of ligands against both experimental and AI-generated SNCA models. While our study focused on three compounds, many other small molecules (including those in clinical trials or natural products) have been reported to interact with α-synuclein. High-throughput docking or machine-learning scoring on wild-type and mutant structures (both monomer and fibril) could identify new candidate molecules or chemical scaffolds that preferentially target a particular SNCA conformation. These could include ligands aimed at *preventing* aggregation versus those aimed at *dissolving* aggregates, reflecting different therapeutic strategies.
- **Improved IDP Modeling:** Develop or apply computational modeling techniques tailored for intrinsically disordered proteins and amyloid fibrils. For instance, using AlphaFold-Multimer or other multimeric modeling tools could help approximate the fibril assembly and expose ligand-binding interfaces that a monomeric model misses. Likewise, incorporating experimental restraints (from NMR chemical shifts or cross-linking data) into modeling, or using coarse-grained simulations for fibril dynamics, may yield more realistic structures for docking. Continued advances in AI,

such as protein language models and generative models, might also enhance structure prediction for IDPs and thus improve subsequent docking accuracy.

- **Experimental Validation:** Strengthen the feedback loop between computation and experiment. Key predictions from docking (for example, that Anle-138b binds tightly to mutant fibrils or that EGCG is less effective on aggregates) should be validated with biophysical experiments. Techniques such as NMR titration, surface plasmon resonance, isothermal calorimetry, or even cryo-EM of ligand-bound fibrils can confirm binding modes and affinities. Validation not only increases confidence in the computational findings but also provides data to refine the models (e.g. adjusting docking protocols or AI model assumptions). Additionally, cellular assays using mutant α-synuclein (A53T, E46K) can test whether the observed binding differences translate to functional outcomes in aggregation or toxicity assays.

By pursuing these avenues, future studies will enhance the reliability of computational predictions for α-synuclein and better exploit their speed and scope. In tandem with careful experimental work, such improvements will facilitate the discovery of effective SNCA-targeted therapeutics and help overcome the challenges posed by this intrinsically disordered, aggregation-prone protein.

Conclusion

In this study, we conducted a comprehensive comparison of AI-predicted versus experimentally determined structures of α-synuclein (SNCA) and examined how well each could explain the binding of three pharmacologically relevant ligands. This dual approach has revealed both the strengths and limitations of current structure-prediction methods in the context of Parkinson's disease protein targets. On one hand, the AlphaFold model of wild-type SNCA proved capable of approximating the overall ligand-binding behavior seen in the NMR ensemble – an encouraging result that suggests AI models can provide useful proxies when experimental structures are lacking. On the other hand, the fine details of ligand interactions, particularly for an intrinsically disordered protein, were often missed. Our docking analysis showed that AlphaFold's wild-type model failed to fully capture ligand-specific interactions in disordered regions (for example, it underappreciated the high affinity of EGCG, which the flexible NMR structure showed clearly). Similarly, ColabFold predictions for the A53T and E46K mutants reproduced some general effects of these mutations but deviated significantly from the cryo-EM fibril structures in terms of binding-site geometry and ligand affinities. These deviations led the AI models to mis-rank or misestimate ligand performance – most conspicuously, the models underestimated EGCG's binding and entirely missed the incompatibility of NPT200-11 with the E46K fibril. Such findings underscore that single static models, no matter how accurate globally, struggle to represent mutation-induced structural changes and the multi-chain architecture of fibrils.

Despite these challenges, our results also illustrate how combining AI models with experimental data can yield insights that neither alone would provide. The inability of NPT200-11 to dock into the E46K fibril was first observed in silico, but its significance became clear upon considering the experimental structure – highlighting a mutation-created steric hindrance that could impact drug efficacy. This example shows that computational failures, interpreted in the light of structural biology, can expose biologically meaningful limitations in drug binding. More broadly, the study reinforces the view that computational and experimental approaches are complementary in nature. AI-based predictions offer speed, scale, and a way to explore many variants or hypothetical structures (e.g., mutant monomers or transitional states) that would be difficult to obtain otherwise.

Conversely, experimental methods provide the necessary ground truth and capture complexities (conformational ensembles, oligomeric contacts, etc.) that predictions might overlook. By iterating between the two – using docking predictions to guide experiments, and using experimental findings to refine computational models – we can build a more complete understanding of SNCA's behavior. Indeed, a key lesson is that hybrid strategies will likely be the most productive: for instance, integrating deep-learning models with ensemble MD simulations and validating critical predictions with NMR/cryo-EM can together overcome many limitations that any single technique faces.

In summary, this work provides a framework for evaluating the reliability of AI-generated protein models in drug discovery applications, using α-synuclein and its familial Parkinson's disease mutants as a test bed. We demonstrated that while current AI models (AlphaFold/ColabFold) are invaluable for generating structural hypotheses, they must be applied with awareness of their limitations – especially for IDPs and aggregated proteins. When used judiciously, however, these models coupled with rigorous docking and validation can illuminate how pathogenic mutations alter drug binding and can identify promising ligand candidates or flag potential failures early. Going forward, applying such computational–experimental workflows to other challenging targets will help to accelerate the discovery of effective therapeutics. By clarifying what our models can and cannot capture about α-synuclein, we lay the groundwork for refining those models (e.g. with IDP-specific techniques) and for designing the next generation of SNCA-targeted interventions to combat Parkinson's disease. Ultimately, our comparative analysis underscores that the synergy of artificial intelligence and experimental biophysics is a powerful approach to unravel complex protein pathologies and guide drug discovery toward success.

Supplementary Information

Supplementary data can be found here.

References

1. Jumper, J. *et al.* Highly accurate protein structure prediction with AlphaFold. *Nature* **596**, 583–589 (2021).
2. Mirdita, M. *et al.* ColabFold: Making protein folding accessible to all. *Nat. Methods* **19**, 679–682 (2022).
3. Van der Zee, J. *et al.* PDmutDB: A Parkinson's disease–specific mutation database. *Mol. Genet. Genomic Med.* **5**, 267–274 (2017).
4. Landrum, M. J. *et al.* ClinVar: Improving access to variant interpretations and supporting evidence. *Nucleic Acids Res.* **46**, D1062–D1067 (2018).
5. MDSGene Project. Systematic database of mutations in Parkinson's disease genes. *Mov. Disord.* **33**, 1230–1238 (2018).
6. Coric, V. *et al.* NPT200-11: A novel oral compound for Parkinson's disease. *Mov. Disord.* **34** (Suppl. 2), S109 (2019).
7. Wagner, J. *et al.* Anle138b: A novel oligomer modulator for disease-modifying therapy of neurodegenerative diseases. *Acta Neuropathol.* **125**, 795–813 (2013).
8. Ehrnhoefer, D. E. *et al.* EGCG reduces α-synuclein aggregation and toxicity in vitro. *Nat. Struct. Mol. Biol.* **15**, 558–566 (2008).
9. Ulmer, T. S., Bax, A., Cole, N. B. & Nussbaum, R. L. Structure of the N-terminal domain of α-synuclein. *Nature* **435**, 708–712 (2005).
10. Li, B. *et al.* Cryo-EM structure of α-synuclein fibrils from A53T familial Parkinson's disease. *Nat. Commun.* **9**, 3609 (2018).
11. Boyer, D. R. *et al.* Structures of fibrils from E46K familial Parkinson's disease mutant α-synuclein. *Nat. Struct. Mol. Biol.* **26**, 1044–1052 (2019).
12. Kim, S. *et al.* PubChem in 2021: new data content and improved web interfaces. *Nucleic Acids Res.* **49**, D1388–D1395 (2021).
13. Corso, G. *et al.* DiffDock: Diffusion steps, twists, and turns for molecular docking. *arXiv* preprint arXiv:2210.01776 (2022).

14. Chari, R. *et al.* Boltz-2: A foundation model for protein–ligand binding affinity prediction. *bioRxiv* preprint (2025).
15. Schrödinger, LLC. The PyMOL Molecular Graphics System, Version 2.0 (2015).

16. Tunyasuvunakool, K. *et al.* Highly accurate protein structure prediction for the human proteome. *Nature* **596**, 590–596 (2021). *(cited for AlphaFold limitations, IDPs)*
17. Ruff, K. M. & Pappu, R. V. AlphaFold and implications for intrinsically disordered proteins. *J. Mol. Biol.* **433**, 167208 (2021).
18. Bhowmick, A., Brookes, D. H., Yost, S. R., Dyson, H. J., Forman-Kay, J. D. & Head-Gordon, T. Finding our way in the dark proteome. *J. Am. Chem. Soc.* **138**, 9730–9742 (2016). *(cited for IDP modeling challenges)*
19. Flagmeier, P. *et al.* Anle138b reduces amyloid fibril binding and toxicity of α-synuclein oligomers. *Proc. Natl Acad. Sci. USA* **113**, 9596–9601 (2016). *(therapeutic relevance of Anle138b)*
20. Levin, J. *et al.* Safety and efficacy of the α-synuclein aggregation inhibitor anle138b in animal models. *Acta Neuropathol.* **127**, 779–795 (2014).