Original Research Article

Structure-Based Drug Discovery of New Compounds Targeting Exon 1 N-terminus Region of Mutant Huntingtin

Yash Jaju[1, 2] and Divya Ramamoorthy[1]

[1]*Moxie Scientist, Georgetown, Texas, United States;*
[2]*Flintridge Preparatory School, 4543 Crown Ave. La Cañada Flintridge, CA 91011, United States*

ABSTRACT

Affecting hundreds of thousands of people around the world, Huntington's Disease is a neurodegenerative disorder causing involuntary movements, poor-decision making, nervous system shutdown, and ultimately death. The aim of this project was to identify molecules as inhibitors to limit the disease-causing mutation, known as mHTT (mutant huntingtin). We used the PDB 4FEB in this study. Three groups of compounds, 50 CNS (central nervous system) compounds, 13 mHTT inhibitors, and 50 peptidomimetics that we screened in the study. In addition, we analyzed the binding site and created a docking grid to prepare the mHTT protein for SP (standard precision) and XP (extra precision) docking. We also performed QikProp analysis, which gave us the log BB, molMW, and Percent Human Oral Absorption of the ligand. Finally, we created ligand interaction diagrams between the top-performing ligands and the 4FEB chain to demonstrate the key interactions. Top-scoring ligands were selected based on their docking scores, with acceptable scores ranging from -3 to -4.5 after XP Docking. While the two top-scoring compounds showed strong computational binding efficiency, in vitro testing is required to ensure the compounds execute effectively as they did computationally. Going forward, our goal is to continue researching and computationally docking ligands that can drive the design of new compounds for effectively targeting the mHTT protein in silico.

Keywords: Huntington's disease; Mutant huntingtin (mHTT); Exon-1 N-terminus (EX1); Polyglutamine expansion

INTRODUCTION

Huntington's disease (HD) is a hereditary neurodegenerative disorder characterized by progressive loss of motor function, cognitive decline, psychiatric disturbances, and dementia (1). The disease affects medium spiny neurons in the striatum of the basal ganglia, along with cortical neurons, which leads to the clinical symptoms of neurodegeneration of the striatum and cortex and loss of control over voluntary movement and cognitive ability (2). In the early stages of the disease, the primary symptom is involuntary movement, and people can perform their daily tasks as usual. However, the disease gradually worsens over a period of 10-20 years following its onset, and many of the observable clinical symptoms are present (3).

In Western countries and Oceania, Huntington's disease affects approximately 5-10 individuals per 100,000 population, which is far greater than observed in Asian and most African populations, where estimates

Corresponding author: Yash Jaju, E-mail: jajuyash2010@gmail.com.

Accepted April 2, 2026
https://doi.org/10.70251/HYJR2348.42313318

range from 0.1-0.5 per 100,000 (4). An example of this is that a person of Japanese descent has a significantly less chance of developing Huntington's Disease (5). Geographic differences between the two regions of the world are thought to influence the disease's risk across their populations (6). In European populations, chromosomes are grouped into 3 main haplogroups: A, B, and C, with haplogroups A1 and A2 being the highest risk of HD. The chromosomes of East Asian populations are more closely associated with those of haplogroup C, the least risk haplogroup for Huntington's Disease (7).

The disease is caused by an expansion of the CAG (glutamine) codon on chromosome 4p16.3's short arm, which is the mutation that leads to HD. Normally, there are 10-35 repetitions of the codon cytosine-adenine-guanine, but the expansion coding for Huntington's Disease contains 36 or more of these repetitions (8). The normal, non-mutated HD gene codes for the HTT protein when translated, a large protein with 3,144 amino acids (9). However, when the mutation is present, the gene codes for the mutated protein mHTT, which is the toxic form of the HTT protein. mHTT includes a polyglutamine expansion of the exon-1 N-terminus (EX1) of the HTT protein, affecting the way it is folded and hence its function (10). The EX1 region of mutant huntingtin is a key pathogenic form of the protein and therefore serves as an important therapeutic target for Huntington's Disease aggregation (11). This mutated protein damages neurons and their functions, ultimately leading to the degradation of the neurons in the basal ganglia, and hence the disease (12). The average age of onset for Huntington's disease is 35 years of age, but additional mutations of the CAG expansion can lead to earlier onset. 40-50 repetitions of the CAG codon led to adult-onset HD, while 50-120 repetitions lead to juvenile HD, where symptoms can begin to show as early as the teenage years (13).

In this work, we applied structure-based drug design techniques to aid in the identification of novel classes of compounds. In contrast to traditional drug screening, structure-based drug design integrates structural insights with computational docking and scoring functions to identify molecules with optimized binding capabilities with the target structure. We selected the mHtt36Q-EX1 region as the target structure because it effectively demonstrates the expanded polyglutamine characteristic of the Huntington's Disease gene. Among the four available structures for this region, we selected the mHtt36Q-EX1-X1-C2(Beta) crystal structure PDB 4FEB due to its high atomic resolution at 2.8Å, providing near-atomic detail of the conformational structure and ligand binding site. The N-terminus domain contains a group of 17 amino acids where molecules can bind, including small molecules and peptidomimetics (14). Previous studies have shown that the EX1 N-terminal sequence before the polyglutamine tract causes the aggregation of Huntington's Disease (Figure 1).

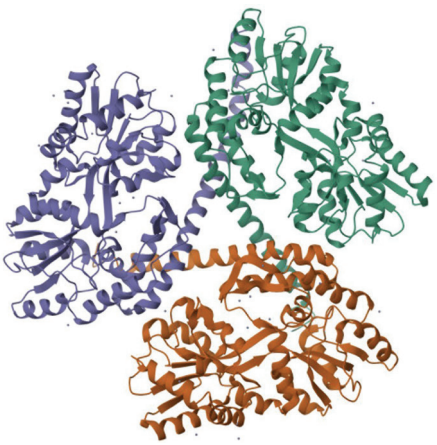

Figure 1. *Crystal Structure of mHTT 36Q3H-EX1-X1-C2(Beta) (PDB: 4FEB).*

We attempted to develop peptidomimetics and cyclopeptides to target the pathological consequences of CAG expansion. As the initial step towards the design of a peptidomimetic or a cyclopeptide, we have screened the peptidomimetics library. As compared to traditional drugs, which are small molecules and protein therapeutics, peptidomimetics are promising for targeting intracellular proteins that are traditionally considered undruggable, complementing treatment solutions with small molecules (15).

METHODS AND MATERIALS

Computational Drug Discovery Workflow

All computational studies, including protein preparation, ligand optimization, molecular docking, and binding affinity calculations, were performed using the Schrödinger Suite 2024-3 (Schrödinger, LLC, New York, NY).

Protein Preparation

The crystal structure(s) of the target proteins were retrieved from the Protein Data Bank (PDB). Before docking studies, the protein was prepared using the

Protein Preparation Wizard (Schrödinger Suite 2024-3) to ensure a chemically and sterically sound structure. Firstly, covalent bonds, bond orders, and formal charges were assigned. Then, hydrogen atoms were added to the protein structure, and their positions were optimized. The protonation and oxidation states of metal cofactors (e.g., zinc, magnesium) were next corrected. After that, all the water molecules outside a 5 Å radius of the active site were removed. The remaining water molecules were retained. To correct the side chain, the protonation states of histidine residues were determine based on the local hydrogen-bonding network. A final energy minimization was performed using the Optimized Potentials for Liquid Simulations (OPLS4) force field to relieve steric clashes and optimize the geometry of the protein, while maintaining the positions of heavy atoms.

Ligand Preparation

The ligands used in this study were obtained from Selleck, PubChem search, and NCI libraries. Each ligand was prepared using the LigPrep tool (Schrödinger Suite 2024-3) to generate three-dimensional (3D) conformers and a range of ionization and tautomeric states. First, protonation state assignment was completed. The pH was set to 7.0 ± 2.0 to generate appropriate states for each ligand. Next, all possible stereoisomers for each ligand were generated and retained. Finally, the structures' energies were minimized using the OPLS4 force field.

Molecular Docking

Molecular docking was performed using Glide (Schrödinger Suite 2024-3) to predict the most favorable binding pose and affinity of each ligand to the prepared protein active site.

Firstly, a grid box of dimensions [e.g., 20 × 20 × 20 Å] was centered on the native ligand (present in the crystal structure) to define the active site. The grid was generated with default settings. In the absence of a ligand in the crystal structure, the grid was centered over the residues around the binding site, as identified by SiteMap analysis, as earlier described. Sitemap analysis was done through the ProteinPlus website offered by the University of Hamburg. The top-ranked binding pocket (P_1) predicted by DoGSiteScorer (Volume 740.48 $Å^3$, DrugScore = 0.87, SimpleScore = 0.47) was selected for further analysis.

The next step of this process was molecular docking, where both Standard Precision (SP) and Extra Precision (XP) docking modes were employed. SP docking was used for an initial, large-scale virtual screening to filter out non-binding compounds based on a preliminary scoring function. XP docking was then used for a more rigorous and accurate assessment of the top-ranked ligands from the SP screen. XP docking uses a more extensive sampling and a more refined scoring function to better account for ligand-protein interactions, which is crucial for identifying high-affinity binders. Since 4FEB does not have a ligand, validation of docking could not be performed, and hence it was the reason mHTT inhibitors were used because they serve as a way of validating the docking in addition to their roles as testable ligands in the study.

Physiochemical Property Prediction (QikProp)

The drug-likeness and ADME (absorption, distribution, metabolism, and excretion) properties of the top-ranked ligands were predicted using QikProp (Schrödinger Suite 2024-3). QikProp calculates over 30 relevant molecular descriptors, including molecular weight, logP, solubility, and blood-brain barrier permeability. These predictions were used to filter out compounds with unfavorable pharmacokinetic profiles, ensuring the candidates were more likely to be orally bioavailable and safe.

Ligand Interaction Analysis

The nature of the interactions between the ligands and the target protein was analyzed to identify key binding motifs and residues. Visual analysis of the docked poses was performed using Maestro to identify crucial hydrogen bonds, π-π stacking, cation-π, and hydrophobic interactions between the ligand and the active site residues.

RESULTS AND DISCUSSION

As Huntington's Disease directly impacts the brain, any potential treatment must be able to cross the blood-brain barrier to effectively target intracellular processes and DNA structures. This proves a challenge for drug discovery for Huntington's Disease and emphasizes the importance of carefully selected compounds.

To address this challenge, we screened 50 compounds from the central nervous system (CNS) database provided by SELLECK. These were selected because they were already optimized for activity within brain cells and have high log BB values, allowing for more effective penetration into the brain. In addition to these CNS compounds, 50 peptidomimetics were from a specialized peptidomimetics database were also

screened. Peptidomimetics are designed to have many of the same behaviors as regular peptides while retaining a smaller size and greater stability, making them more efficient in targeting unwanted intracellular proteins compared to traditional approaches.

The ligands screened were taken from three main groups: central nervous system compounds, 13 known mHTT inhibitors, and peptidomimetics. Each group was included in the study for a specific reason. CNS compounds have a known ability to function in the brain, mHTT inhibitors play a direct role in mitigating the disease's mechanism, and peptidomimetics demonstrate potential to bind effectively to mutant proteins while ensuring stability is maintained in the cell. Before computational studies could be conducted, all ligands were prepared through ligand interaction analysis as described above.

Using molecular docking methods, we analyzed how well each ligand interacted with the target mHTT protein. Docking scores were used to measure binding affinity, as lower scores indicate stronger and more effective protein-ligand interactions. Firstly, we identified the binding site of the mHTT protein and then prepared the ligands for docking. Glide then explored the ligand's conformations and positions through different precision levels, SP and XP, to test for the speed and accuracy of the binding. After applying this method to the ligands in the three different groups, we identified the compounds that produced the best docking scores. After doing XP Docking, Schrödinger returns a docking score based on the binding affinity between ligand and protein. A more negative score indicates better binding. The two compounds shown below are most promising because they return the best XP Docking Scores (Table 1).

QikProp showed similar results for Figure 2 compounds, showing it can be used in future work. The lower docking scores mean that in silico, the ligand fits better to prevent the aggregation of the CAG expansion mutation, as this is vital in preventing Huntington's Disease. Thus, for the ligands we tried, 168758677 and the mHTT-IN2 inhibitor each worked about as well.

When analyzing the first ligand interaction diagram, we observed several key amino acid residues within the binding region that contributed to ligand stabilization. The ligand interacted with residues ASP41, PRO40, HIE39, GLU38, VAL37, THR36, and VAL35, and these are clustered around the active site and known to play important roles in interactions between proteins and ligands. Additionally, interactions were observed with LYS34, TYR17, ALA21, GLY24, LYS25, and GLU28, meaning that hydrogen bonding and hydrophobic

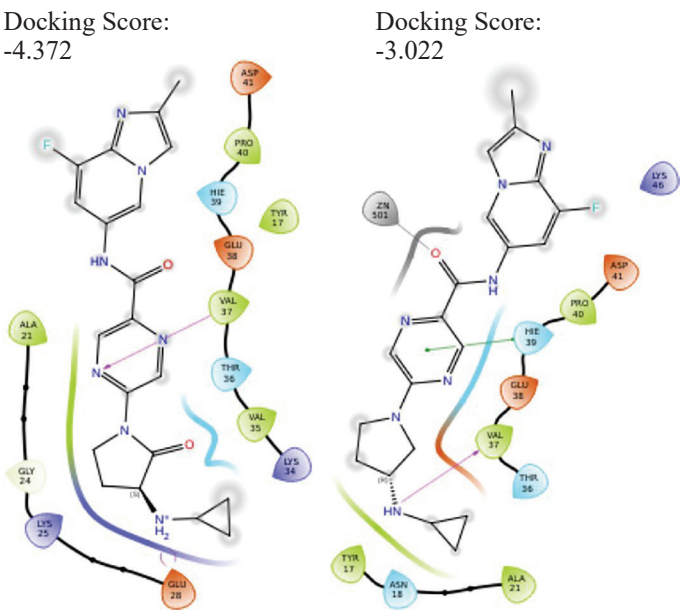

Figure 2. *Ligand Interaction between 4FEB chain and the ligand 168758677 (left), Ligand Interaction between 4FEB chain and mHTT-IN2 inhibitor (right). The green arrows represent π-π stacking (a hydrophobic interaction between two aromatic rings. The pink arrows represent hydrogen bonds, with the arrow pointing from the donor to the acceptor.*

Table 1. *ADME properties of candidate compounds predicted by QikProp analysis. Molecular weight (MW), Glide XP docking score, predicted water permeability (QPlog Pw), lipophilicity (QPlog Po/w), aqueous solubility (QPlog S), and human oral absorption (HOA) are reported.*

Title	mol MW	Glide XP Docking Score	$Q_P$log (P_w)	$Q_P$log (P_o/P_w)	$Q_P$log (S)	Human Oral Absorption	Percent Human Oral Absorption
168758677	409.422	-4.372	16.288	1.803	-4.382	3	74.998
mHTT-IN-2	395.438	-3.022	13.658	2.911	-5.188	3	84.521

contacts are observed. Additionally, a single backbone hydrogen bond is shown at VAL 37 at a predicted distance of 3.0 Å and approximately -2.0 kcal/mol of interaction energy.

The second ligand interaction diagram showed a similar interaction pattern, meaning there is an observable consistency in ligand binding. The ligand formed interactions with ASP41, PRO40, HIE39, GLU38, VAL37, and THR36, and additional contacts with TYR17, ASN18, ALA21, and LYS46. The hydrogen bond to the VAL37 backbone is estimated at 2.8Å. The interaction with GLU 38 is identified as a salt bridge rather than a standard hydrogen bond. This ion-pair interaction was found to have approximately 2.9 Å. In this interaction, the ligand anchors to the catalytic center through an observed coordination bond with the zinc ion (ZN 502) at 2.1 Å, likely adopting a distorted tetrahedral geometry. Combined with the π-π stacking observed at HIE 39, the ligand interaction analysis diagram on the right creates a pose with an estimated total binding free energy ΔG of approximately -13.5 kcal/mol.

CONCLUSION

The toxicology study found that the compounds 168758677 and mHTT-IN-2 had the best docking scores and they both have similar interactions with the binding site of mHTT. Additionally, in both ligand-protein complexes, hydrogen bonds, π-π stacking, and hydrophobic interactions were formed, along with the amino acid residues ASP41, PRO40, HIE39, GLU38, VAL37, and THR36. The observation of a zinc ion in the interaction between mHTT-IN-2 and the mHTT binding site may serve as a feature of additional importance when it comes to the stabilization and ligand binding. Figuring out the role of the zinc ion is the next step of the study. While these findings provide insight into potential binding modes and highlight these compounds as candidates for further investigation, they are based solely on computational modeling. As such, they should be considered hypothesis-generating rather than predictive of therapeutic efficacy.

Even though this study is limited to computational modeling and docking analyses in silico, it demonstrates that peptidomimetics and CNS compounds have a strong binding affinity for the mHTT protein. More work is needed to discover molecular dynamics, stability, and specificity of identified compounds. Nevertheless, this research contributes to the development of novel yet effective therapeutic strategies designed to target mHTT.

ACKNOWLEDGMENT

We thank Moxie Scientist for computational resources.

CONFLICT OF INTEREST

The authors declare no conflicts of interest related to the work.

REFERENCES

1. Wild EJ, Tabrizi SJ. Huntington's disease: from molecular pathogenesis to clinical treatment. *Pract Neurol.* 2021; 21 (2): 132-142.
2. Zuccato C, Valenza M, Cattaneo E. Molecular mechanisms and potential therapeutical targets in Huntington's disease. *Int J Mol Sci*. 2021; 22 (16): 863.
3. Rosenblatt A, Leroi I, Neill S, *et al*. The association of CAG repeat length with clinical progression in Huntington disease. *Arch Neurol*. 2001; 58 (2): 273-278.
4. Pringsheim T, Wiltshire K, Day L, Dykeman J, Steeves T, Jette N. The incidence and prevalence of Huntington's disease: a systematic review and meta-analysis. *CNS Neurosci Ther*. 2013; 19 (11): 901-909.
5. Bates G, Harper PS, Jones L, editors. Huntington's Disease. 3rd ed. *Oxford: Oxford University Press*; 2002. ISBN: 0-19-851060-8.
6. Bates GP, Tabrizi SJ, Jones L. Huntington's disease. *Ageing Res Rev.* 2025; 102984.
7. Warby SC, Visscher H, Collins JA, Doty CN, *et al*. HTT haplotypes contribute to differences in Huntington disease prevalence between Europe and East Asia. European Journal of Human Genetics. 2011; 19 (5): 561-566. https://doi.org/10.1038/ejhg.2010.229
8. Walker FO. Huntington's disease. *Orphanet J Rare Dis*. 2010; 5: 40. https://doi.org/10.1186/1750-1172-5-28
9. Saudou F, Humbert S. The biology of huntingtin. *Neuron*. 2016; 89 (5): 910-926. https://doi.org/10.1016/j.neuron.2016.02.003
10. Thakur AK, Jayaraman M, Mishra R, *et al*. Polyglutamine disruption of the huntingtin exon 1 N terminus triggers a complex aggregation mechanism. *Nat Struct Mol Biol*. 2009; 16 (4): 380-389. https://doi.org/10.1038/nsmb.1570
11. Yang H, Yang S, Jing L, Huang L. Truncation of mutant huntingtin in knock-in mice demonstrates exon1 huntingtin is a key pathogenic form. Nature Communications. 2020; 11: Article 3142. https://doi.org/10.1038/s41467-020-16318-1, https://doi.org/10.1038/s41467-020-19873-9
12. Warby SC, Montpetit A, Hayden AR, *et al*. CAG expansion in the Huntington disease gene is associated

with a specific and targetable haplogroup. *Eur J Hum Genet*. 2010; 18 (4): 401-407.

13. Schulte J, Littleton JT. The biological function of the huntingtin protein and its relevance to Huntington's disease pathology. *Curr Trends Neurol*. 2011; 5: 65-78. PMID:22180703.
14. Steffan JS, Bodai L, Pallos J, *et al*. Histone deacetylase inhibitors arrest polyglutamine-dependent neurodegeneration in Drosophila. *Prion*. 2001; 5 (2): 1-8.
15. Mishra R, Thakur AK, Wetzel R. Peptidomimetics as modulators of protein aggregation in neurodegenerative diseases. *Comput Struct Biotechnol J.* 2019; 17: 57-70.