IJHSR
ijhighschoolresearch.org
RESEARCH ARTICLE

ML-Driven Approach to Discovery of Peptide-Based BACE1 Inhibitors

Nicholas Chen

ACG Parnell College, 2 Titoki Street, Parnell, Auckland 1052, New Zealand; nicholas.chen1216@gmail.com
Mentor: Vladimir Akhmetov

ABSTRACT: Alzheimer's Disease (AD) is a progressive neurodegenerative disorder currently affecting over 50 million individuals globally. The amyloid-beta (Aβ) hypothesis remains one of the most understood models for AD pathogenesis, prompting extensive research efforts in this area. The recently FDA-approved monoclonal antibody Lecanemab suppresses formation of Aβ plaques but is accompanied by significant side effects including immunogenicity. Peptides can achieve comparable affinity with fewer adverse effects and offer cost-effective and flexible development processes. Recent advancements in Deep Reinforcement Learning (DRL) algorithms, including the Nobel Prize recognized AlphaFold3 (AF3) and RoseTTAFold (RF) diffusion, have enabled the rapid generation of potent binders, revolutionizing the field of peptide drug discovery. Here, we apply open-source codes to generate *de novo* catalytic-site targeted binder of β-secretase (BACE1), one of the key enzymes in the production of Aβ plaques in the brain.

KEYWORDS: Translational Medical Sciences, Drug Identification and Testing, Machine Learning, Alzheimer's Disease, BACE1.

Introduction

AD stands as one of the predominant causes of dementia worldwide, with approximately 50 million individuals affected presently, a figure projected to double every five years. By 2050, it is estimated that the number of AD patients will rise to 152 million.[1] The burden of AD extends beyond the individual affected, impacting their families and exerting substantial economic strain, costing economies globally 1.3 trillion US dollars per year.[2] Hence, the development of novel drugs to prevent and treat AD is of utmost importance. AD is defined by the presence of extracellular amyloid plaques and intracellular neurofibrillary tangles, along with neuroinflammation and a reduction in neurotransmitters in patients' brains.[3] Aberrant neural network activity, dysfunction, and loss of synapses, as well as degeneration of specific neuronal populations, are the main underlying causes of cognitive decline in AD.

Amyloid-β (Aβ) and the Amyloid Hypothesis:

Despite tremendous effort, the exact pathology of AD onset remains unclear.[4] The predominantly understood mechanism suggests accumulation of amyloid-β (Aβ) plaques in the brain, triggering a cascade of events that lead to neuronal degradation and cognitive impairment.[5] These senile plaques are characterized by a central Aβ core surrounded by dystrophic neurites, typically arising from the aberrant aggregation of soluble Aβ42 and Aβ40 among nerve cells.[6] During this event of amyloidosis of the brain, Aβ peptides and soluble N-terminal peptides APPα and APPβ (sAPPα, sAPPβ) are cleaved from the amyloid precursor protein (APP) by secretases, forming clusters of soluble toxic oligomeric Aβ.[7] These oligomers bind to zinc (Zn(II)) to initiate the formation of hydrophobic deposits of Aβ plaques.[8] The plaques interfere with neuronal interactions, disrupt neurotransmission, and lead to synaptic loss, which is currently regarded as the main symptom of AD.[9,10]

Another principal pathological hallmark of AD is neurofibrillary tangles (NFT).[11] NFT consists of abnormal fibers of misfolded and hyperphosphorylated Tau (pTau) within neuronal cells.[12] Notably, hyperphosphorylation of Tau results in the formation of tangles that disrupt microtubule stability, leading to neuronal cytoskeleton deterioration and impaired axonal transport. Evidence suggests that Aβ is upstream of Tau in AD pathogenesis, triggering the conversion of Tau from a normal to a toxic state.[13] Thus, preventing Aβ formation has become a strategic focus in AD drug discovery.

β-secretases are shown as amyloidogenic proteases and therefore considered promising targets to reduce Aβ plaques in the brain. Preventing the onset and progression of Alzheimer's disease (AD). However, developing effective inhibitors for these targets is challenging due to the vastness of the chemical space. The synthetically accessible chemical space is estimated to contain between 10^{33} and 10^{60} molecules, of which only a small fraction has been explored.[14,15] This vast chemical space poses significant challenges for current drug design and inhibitor development methodologies. Traditional approaches, such as Analog-Based Drug Design (ABDD), high-throughput screening (HTS), and fragment-based drug discovery (FBDD), rely heavily on extensive trial and error, making them both time-consuming and expensive.[16] On average, it takes approximately 10-15 years and $2.8 billion to discover a viable chemical inhibitor and bring it to market.[17,18] These conventional methods depend largely on human knowledge and intuition, limiting the exploration to a small segment of the chemical space and often resulting in the discovery of molecules with structures and properties similar to known

DOI: 10.36838/v7i4.29

compounds.[16] Consequently, about 98% of drug candidates fail before reaching the final stage of the drug development process.[19] ML-assisted drug discovery may help navigate vast chemical space and generate structures that could be missed by humans due to inherent biases in molecular design.

Generative artificial intelligence:

The latest breakthroughs in artificial intelligence (AI) have enabled new approaches to de novo drug discovery that address the limitations of human knowledge and the trial-and-error process. Deep Reinforcement Learning (DRL) approaches, in particular, can understand complex patterns in datasets that are often incomprehensible to humans.[20] These approaches generate millions of new compounds and inhibitors with bioactive and pharmacological properties similar to existing inhibitors. As a result, DRL allows for a more efficient exploration of chemical binding space, drastically outperforming human-driven methods.[21] Many of these generated inhibitors have displayed promising efficacy during subsequent steps of the drug design process, such as *in silico* and *in vitro* testing, demonstrating the effectiveness of this new drug discovery paradigm.[22]

Denoising diffusion probabilistic models (DDPMs), a powerful class of DRL models recently demonstrated to generate new photorealistic images from text prompts, have several properties well-suited for protein design.[20] These models generate highly diverse outputs by denoising data corrupted with Gaussian noise, enabling the creation of new samples with properties similar to the initial training dataset. They are guided by reinforcement learning at each step of the iterative generation process toward design objectives by providing conditioning information, allowing the generation of inhibitors with specific properties.[23]

DDPMs such as RoseTTAFold (RF) diffusion and AlphaFold 3 (AF3) can generate novel protein structures with high precision to inhibit target sites of the protein.[23,24] They also rank the suitability of individual proteins as potential drug targets by evaluating factors such as confidence level, size and accessibility of binding pockets, and the uniqueness of the predicted protein fold. A combination of these models can be employed together, linked by Message-passing neural networks (MPNN).[25] In this approach, the generated protein backbone model by RF diffusion is analysed by MPNN for molecular library screening, predicting a peptide sequence from the protein backbone. AF3 is then used to predict protein structures from the peptide sequence, generating a rank of the best inhibitors binding to targeted sites. Finally, the ease of inhibition is analysed by comparing the depth of the catalytic site and the position of attachment of the inhibitor.

Our proposal utilizes DRL methodology, combining RF diffusion, protein MPNN, and AF3 to generate peptide-based inhibitors of BACE1. These inhibitors are then analysed for structural similarity and chemical interactions while also proposing peptide stapling positions and groups to enhance affinity. We hope that our strategy will provide insights for further AD studies and contribute to the discovery of novel candidate drugs for AD.

The role of Secretases and their inhibition strategies:

APP is metabolized through two distinct routes: the amyloidogenic and non-amyloidogenic pathways. In the amyloidogenic pathway, APP is initially cleaved by β-site APP-cleaving enzyme (BACE1) at the Asp1 site, generating sAPPβ and a 99-amino acid membrane-bound C-terminal fragment (CTF) C99.[26] Subsequently, γ-secretase cleaves C99 to release Aβ and CTFγ, with both β and γ-secretase required for the formation of Aβ plaques.[27] Meanwhile, in the non-amyloidogenic pathway, APP is primarily cleaved first by α-secretase at the Aβ Leu17 site, producing sAPPα and C83. C83 is further cleaved by γ-secretase to produce p3 peptide and the APP intracellular domain (AICD) in the region containing the Aβ peptide, thereby preventing Aβ production.[28]

Since the non-amyloidogenic pathway does not produce Aβ, a preference for the amyloidogenic over the non-amyloidogenic APP processing pathway predisposes to AD pathology. Thus, targeting BACE1 is a rational approach to inhibit Aβ formation by inhibiting the amyloidogenic pathway while allowing the non-amyloidogenic pathway to proceed. Its cleavage of APP is the rate-limiting step in Aβ production, with an extended catalytic site that confers substrate specificity and has formed the basis for developing inhibitors.[29] However, the design and development of the BACE1 inhibitors have proven to be very challenging due to the large catalytic site and structural flexibility of BACE1.[30] The current FDA-approved BACE1 inhibitor, Lecanemab, is a monoclonal antibody (mAb) that offers high selectivity and is effective at low nanomolar concentrations *in vitro*. However, mAbs lack blood–brain barrier (BBB) permeability due to their extended size (~150 kDa), which is tailored to occupy the large BACE1 catalytic site.[31] In addition, there is the potential for Fc receptor-mediated immunogenicity, mediated by microglia, which can cause vasogenic edema and cerebral microhemorrhage.[32] Such amyloid-related imaging abnormalities occurred in 12.6% of trial participants, compared to 1.7% in the placebo group, and among those affected, 17.3% experienced additional brain bleeding, compared to 9% in the placebo group.[33] The benefits of Lecanemab do not seem to outweigh the risks of serious side effects, leading to its rejection by the European Medicines Agency (EMA).[34]

Other approaches involving small-molecule BACE1 inhibitors have high BBB permeability due to their small size (~0.5kDa) but offer low selectivity to the target and thus increase the likelihood of drug-drug interactions (DDIs).[35] Current small-molecule BACE1 inhibitors, such as LY2886721 and E2609, have been discontinued due to a lack of cognitive improvement in patients and adverse effects in the liver caused by its lack of specificity and affinity to the target (Table 1).[36,37] Peptide-based drugs, on the other hand, offer higher selectivity than small-molecule drugs and greater permeability across the blood-brain barrier (BBB) than monoclonal antibodies (mAbs), making them promising candidates for rational drug design and inhibition.[38]

Table 1: Small molecule BACE1 inhibitors in clinical trials.[39] *Terminated due to abnormal liver biochemistry, **Removed from pipeline. This table is reproduced with permission under Creative Commons Attribution 4.0 international. From "BACE1 inhibitor drugs in clinical trials for Alzheimer's disease" by Robert Vassar. 2014.

Company	Drug	Phase
AstraZeneca/Lilly	AZD3293	Phase 2/3
CoMentis	CTS-21166	Phase 1
Eisai/Biogen Idec	E2609	Phase 2
High Point	HPP854	Phase 1
Janssen/Shionogi	------	Phase 1
Lilly	LY2886721	Phase 2*
Merck	MK-8931	Phase 2/3
Novartis	------	Phase 1
Pfizer	PF-05297909	Phase 1
Roche	RG7129	Phase 1**
Takeda	TAK-070	Phase 1
Vitae/Boehringer Ingelheim	VTP-37948	Phase 1

Peptides, Small molecules and Monoclonal antibodies:

The development of drugs targeting protein-protein interactions (PPIs) poses significant challenges for small molecules.[40,41] Small molecules generally struggle to bind with high affinity to the large and flat binding sites typical of PPIs, unlike mAbs. However, mAbs, while capable of high-affinity binding, are typically limited to extracellular targets due to their poor cell permeability. Therefore, even if they bound to the target with a high affinity *in vitro*, it is difficult to inhibit the target *in vivo* within the human brain. Specifically, peptide drugs offer several advantages over both small molecule drugs and monoclonal antibodies for targeting PPIs. Their molecular size and tuneable properties make them well-suited for this purpose.[42] Peptides are generally larger than conventional small molecules, granting them an antibody-like affinity for binding to flat PPI interfaces while being smaller in molecular weight compared to antibodies.[43]

Thus, with their simple conformational structure and a molecular size between those of small molecules and large protein mAbs, peptides are expected to have higher target affinity and less toxicity compared to small molecules. With their high molecular structural diversity, peptide and peptidomimetic frameworks offer significant potential for structural variations through simple modifications or replacements of amino acid fragments (e.g., N-methylation). The synthesis of peptides and peptidomimetics is well-documented and allows for the generation of many molecules at low costs, making them highly promising for drug development.[44]

Thermodynamics and Entropy penalty:

The restriction of a small molecule's motion upon binding to a protein causes a loss of configurational entropy, resulting in a penalty to its binding affinity. This phenomenon, known as Enthalpy-Entropy Compensation (EEC), presents a significant challenge in drug design. Overcoming EEC is essential to maximize the binding interaction of a candidate drug with its target protein, thereby improving its affinity.[45]

The strength of this interaction is evaluated through the binding free energy (ΔG_b), which is the primary parameter to optimize. The drug design process is iterative and has a strong structural component: crystallographic structures of the target in complex with a known ligand allow for the identification of binding sites and a static representation of ligand–target binding interactions. Based on this information, initial virtual screening of a large library of compounds is usually performed to identify those capable of forming a stable complex with the target, indicated by a large negative value of ΔG_b.[46] ΔG_b for a ligand–target complex at a given temperature can be expressed as $\Delta G_b = \Delta H_b - T\Delta S_b$, where ΔG_b is the sum of an enthalpic (ΔH_b) and an entropic ($T\Delta S_b$) term. This dictates the binding affinity (K_a) through the equation $K_a = e^{\frac{-\Delta Gb}{RT}}$, so thus extremely high affinity is only achieved when both ΔH_b and ΔS_b contribute favorably to binding.[47]

In recent years, new strategies have emerged for interpreting thermodynamic data, becoming key drivers for lead optimization programs in drug development. Enthalpy and entropy are made favorable through molecular attractions between the binder and the target, including hydrophobic, hydrophilic, and electrostatic interactions. Conformational flexibility is entropically unfavorable and is determined by the number of rotatable bonds. Consequently, efforts to maximize favorable enthalpic and entropic contributions through intermolecular forces and restricting the number of rotatable bonds are strategies to induce the greatest decrease in ΔG_b.[46,48] This includes macrocyclization techniques for peptide to effectively reduce conformational freedom through linkers.[49]

By applying these strategies using computational chemistry, developers can estimate the entropy penalty ($-T\Delta Sb$) of their designed drug candidates and methods for optimizing it. Given that EEC may be expected to some degree during the optimization process, a combination of biophysically determined thermodynamic data and high-quality structural information is highly useful in overcoming the compensatory tendency. For example, directing hydrogen bonds to already structured regions of the protein or utilizing multiple hydrogen bonding interactions from a single group, where the entropic penalty has already been paid, can help achieve this goal.[50]

Stapled helical peptides:

Peptides with various conformations have drawbacks, including poor stability against circulating proteolytic enzymes and an intrinsic inability to penetrate cell membranes. Despite these challenges, peptide-based therapeutics have seen significant development and marketing over the past two decades, supported by organic synthetic techniques that address the undruggable characteristics of peptides, such as poor stability in circulation against proteolytic enzymes, poor cell membrane penetration, and the tendency to lose native conformation in aqueous solution.[51] Among the various peptidomimetics developed as PPI inhibitors, stapled helical peptides have gained particular attention for their increased target affinity, enhanced cell penetration, and protection against proteolytic degradation compared to linear bioactive peptides. As referred to by Xiang *et al.*, peptide stapling involves cross-linking either side

DOI: 10.36838/v7i4.29

chain to side chain or side chain to the terminus of anchoring amino acids, forming preorganized stable helical conformations with reduced entropic penalty.[51] This method usually involves the use of olefin through olefin metastasis to achieve peptide cyclisation.

To increase our inhibitor's permeability and affinity to BACE1, these generated peptides can be converted into stapled helical peptides by adopting a preformed, stable α-helical conformation using Grubbs metathesis catalysts to tether the side chains of amino acids.[51] Common stapled amino acids include Lys and either Asp or Glu, using orthogonal protecting groups for staple formation on the same face of the helix at positions *i*, *i+4*, *i+7*, and *i+11*.[52] This approach would reduce flexibility and intrinsic disorder of the inhibitor in their unbound free state, thereby incurring less entropy penalty upon binding to the target.[53]

Methods

The crystal structure of BACE1 at pH 4.5 (2ZHT in the Protein Data Bank), representing the enzyme's most active conformation, is selected as the protein model. Previous studies have revealed that the BACE1 active site contains two aspartic acid residues, Asp32 and Asp228, that play crucial roles in substrate binding.[54,55] Therefore, the coordinates of these residues are used to define the binding site and therefore peptide's backbone design space using RF Diffusion. The contacts with these residues are used to confirm a valid protein structure using Message-passing neural networks (MPNN) and then AlphaFold3 (AF3), utilizing information from the molecular library of amino acids.

Using the open-source code for RF diffusion, we specified continuous chains for contigs as A1-385:30-40 for generating a backbone for a 30-mer peptide and A1-385:10-20 for a 16-mer peptide, utilizing the PDB file 2ZHT.[56] The process was run for 200 iterations, focusing on catalytic residues Asp32 and Asp228 as primary hotspots. Additional hydrophobic residues of BACE1, including Phe47, Phe109, and Ile110, were included to enhance binding interactions, ensuring the generated binder not only targets the catalytic site but also forms stable interactions with other key residues to secure binding to BACE1.

Next, we applied ProteinMPNN to generate amino acid sequences based on the generated backbone, with a sampling temperature of 0.5 to diversify sequence sampling. Following this, AF3 was used for sequence validation and 3D protein structure prediction. Parameters for AF3 included enabling an initial guess with desired coordinates, setting the recycle count to 3, and using AF3 v3 parameters to optimize prediction accuracy. Finally, the generated peptides were screened for interaction strength with the target, eliminating those with minimal or no interaction to ensure specificity and efficacy of the binder.

The algorithm returns the optimal binding conformation of each peptide, yielding the lowest binding energy along with the smallest values of predicted aligned error (PAE) and root mean square deviation (RMSD). Since lower binding energies correspond to more stable binding interactions, peptides with lower binding energies bind more favorably to BACE1. Consequently, these peptide molecules act as stronger inhibitors, preventing BACE1 from binding to APP and producing Aβ. Thus, among the generated candidate compounds and existing BACE1 drugs, the molecules with the lowest binding energies are determined to have stronger binding interactions with, and more potent inhibition of, BACE1.

Result and Discussion

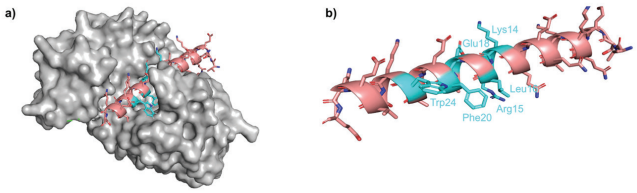

Figure 1: Molecular design of 30-mer: a) BACE1 in complex with the generated 30-mer peptide, b) Generated 30-mer peptide. The residues marked in blue have weak interactions with BACE1.

Among several promising outputs, one generated structure has attracted particular attention owing to the presence of multiple noteworthy peptide-protein interactions. Moreover, peptides containing more than twenty amino acid residues are likely to maintain secondary structure elements. Thus, the structure is likely to show experimental binding due to a smaller entropy penalty (Figure 1). Below, we analyze interactions such as hydrogen bonding, electrostatic attractions, and van der Waals forces between the generated peptides and the target molecule.

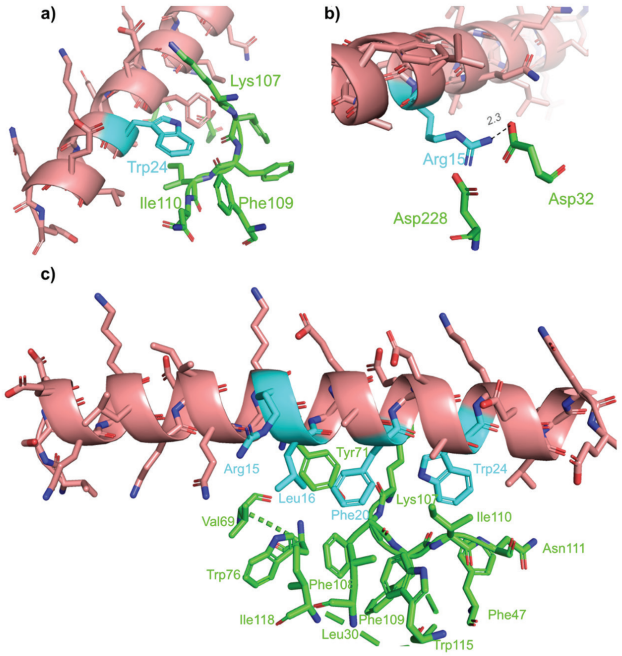

Figure 2: Residue interactions of 30-mer with BACE1: a) Hydrophobic interactions around 5Å of Trp24, b) Electrostatic interactions and Hydrogen bonding of Arg15 to catalytic Asp32 and Asp228, c) Overall depiction of charge interactions of Arg15 and hydrophobic interactions of Leu16, Phe20, and Trp20 with the hydrophobic surface of BACE1, composed of hydrophobic residues Leu30, Phe47, Val69, Tyr71, Trp76, Phe108, Phe109, Ile110, Trp115, and Ile118.

In the generated 30-mer peptide sequence VSEEQKKLQEIQGKRLGEAFEEAWKDVEQE, numerous residues exhibit chemical interactions with the target, BACE1. Electrostatic interactions are facilitated by charged residues Lys6, Lys14, Arg15, Glu18, Glu21, Glu22, and Asp26 with BACE1. Notably, the positively charged Arg15 positions itself within 5Å of the negatively charged catalytic aspartic acids Asp32 and Asp228, forming electrostatic interactions and hydrogen bonds that enhance catalytic site attachment and inhibition. Furthermore, the hydrophobic residues Leu16, Phe20, and Trp24 are aligned on the same face of the helix and positioned in a groove formed by the hydrophobic surface of BACE1, exemplifying the lock-and-key mechanism for interaction. The collective interactions of these three hydrophobic residues create a strong driving force that anchors the peptide to the surface of BACE1 (Figure 2). In addition, the hydrophobic interactions, along with the charge interactions between Arg15 and Asp32/Asp228, increase the selectivity of the peptide, further enhancing its inhibitory potential.

Thus, these interactions provide good attachment of the binder with high favorable enthalpy. The entropy penalty present with the binder can be minimized using macrocyclization approaches. Lys6 and Glu10 residues are charged and aligned on the same face of the helix at positions *i* and *i+4*, making them good candidates for peptide stapling through macrocyclization. Connecting an olefin to these residues mimics a natural salt bridge interaction through charge interactions. They are also both solvent-exposed, meaning that stapling on these residues does not influence the direct intermolecular attraction between the peptide and the target or the dissolution effect involved. However, the hydrophobicity of the olefin bridge would tend to stick to the protein and orient away from the solvent to minimize enthalpy, presenting a design challenge that could be overcome using hydrophilic molecules like triazoles. Alternatively, olefin attachment could be made to the hydrophobic residues Leu16 and Pro20, which are also aligned on the same face of the helix at positions *i* and *i+4*. These residues are positioned to face the hydrophobic surface of BACE1 and can interact with the olefin through hydrophobic interactions. However, there is a chance of undesirable interactions with BACE1. Further investigations and testing should be carried out to determine the optimal stapling positions to use.

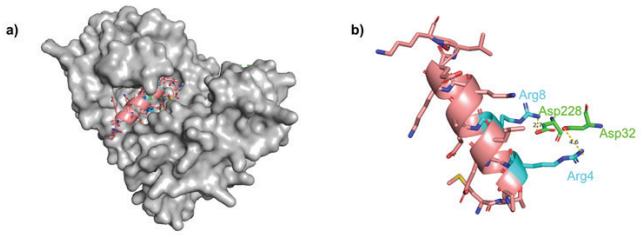

Figure 3: Molecular design of 16-mer: a) Crystal model of BACE1 in complex with the generated 16-mer peptide, b) Generated structure of 16-mer and its residue interaction with catalytic Asp32 and Asp228.

One of the generated 16-mer peptides also caught our attention as an alternative (Figure 3). Arg4 and Arg8 are both within 5 Å of the catalytic aspartic acids Asp32 and Asp228. Its smaller chain size enhances permeability through the blood-brain barrier (BBB), making it a promising drug candidate. However, this smaller size also results in a higher entropy penalty compared to the larger 30-mer peptide, leading to reduced affinity for the target. This drawback, along with fewer interactions with BACE1, positions it as a second choice compared to the 30-mer.

Compared to existing inhibitors like Lecanemab, our generated peptide binders are smaller, potentially allowing for greater permeability across the blood-brain barrier (BBB) and reduced side effects. Leveraging DRL in the development and refinement of our inhibitor may also lower production costs relative to Lecanemab; however, further research and clinical studies are essential before any definitive claims can be made. Future research could test these candidate BACE1 inhibitors through later stages of the drug design pipeline, including assessment, validation, and molecular dynamic simulations. Further iterative diffusion and optimizations based on this binder could yield more desirable pharmacological and bioactive properties, leading to experimental validation. Notably, the algorithm has limitations, as binders with longer chains (>50-mer) are often generated to exhibit weaker or no interactions with the target. Additionally, this approach focused solely on the Aβ hallmark; incorporating a multi-target strategy to address both Aβ and Tau could potentially yield a more effective therapeutic outcome. With over 50 million people worldwide currently affected by Alzheimer's Disease (AD), the urgency to develop novel treatments is paramount. The recent emergence of AlphaFold3 and diffusion algorithms has revolutionized the generation of peptide binders, allowing researchers to produce viable candidates in minutes with a high likelihood of experimental validation success. This paper showcases the capabilities of generative AI in creating structurally and biochemically plausible peptide binders for their targets. Through iterative processes, we generated and analysed binders of 16 and 30-mer, showing significant potential for advancing to drug testing stages. While some binders initially showed minimal interaction with the target, these issues were resolved through further iterations. Structural deformities were also addressed similarly. Our work demonstrates the feasibility of Deep Reinforcement Learning algorithms in rapidly generating binders, which can accelerate drug discovery for targeting diseases.

■ Conclusion

Focusing on BACE1, we employed AF3 and RF diffusion for the *de novo* generation of peptide-based inhibitors. The peptide backbone is generated by RF diffusion, then optimized and folded by Message-Passing Neural Networks (MPNN) and AF3 to create candidate inhibitors targeting the active site. Through this method, we identified promising 30-mer and 16-mer peptides as drug candidates, demonstrating significant chemical interactions with BACE1 and positioned for advancement into preclinical research. We also proposed peptide stapling methods to enhance peptide affinity, with optimal stapling parameters and affinity to be determined through laboratory testing. Overall, this approach provides a simple, robust, and accurate method for discovering BACE1

DOI: 10.36838/v7i4.29

inhibitors and potential AD drugs, facilitating efficient exploration of the chemical space and accelerating the development of novel AD treatments. We anticipate that these methods can also be applied to efficiently generate and optimize millions of candidate compounds for other major therapeutic targets, revolutionizing *de novo* drug discovery.

Data Availability

The open-source code to RF diffusion and AlphaFold3 can be found in this Google Collab https://colab.research.google.com/github/sokrypton/ColabDesign/blob/main/rf/examples/diffusion.ipynb#scrollTo=TuRUfQJZ4vkM or under GitHub https://github.com/RosettaCommons/RFdiffusion?tab=License-1-ov-file#readme.

Copyright (c) 2023 University of Washington. Developed at the Institute for Protein Design by Joseph Watson, David Juergens, Nathaniel Bennett, Brian Trippe, and Jason Yim.

The generated candidate drugs associated with this work are publicly available at https://github.com/nnicholas-c/Drug-Discovery-for-BACE1.

Acknowledgments

I attest that this paper's ideas, graphics, and writing are entirely my own. I would like to thank Dr Vladimir Akhmetov from Massachusetts Institute of Technology for his mentorship, guidance, and assistance in drafting this paper. I also wish to thank Indigo Research for their continuous support and for providing this opportunity. Lastly, I extend my thanks to the Institute for Protein Design at the University of Washington for their open-source RF diffusion code, which was instrumental in the synthesis of these drugs.

References

1. Breijyeh, Z., & Karaman, R. (2020). Comprehensive Review on Alzheimer's Disease: Causes and Treatment. *Molecules (Basel, Switzerland)*, *25*(24), 5789. https://doi.org/10.3390/molecules25245789
2. World Health Organisation (2023). Dementia fact sheet https://www.who.int/news-room/fact-sheets/detail/dementia#:~:text=Alzheimer%20disease%20is%20the%20most%20common%20form%20and%20may%20contribute,frontal%20lobe%20of%20the%20brain).
3. Efthymiou, A. G., & Goate, A. M. (2017). Late onset Alzheimer's disease genetics implicates microglial pathways in disease risk. *Molecular neurodegeneration*, *12*(1), 43. https://doi.org/10.1186/s13024-017-0184-x
4. Mathys, H., Boix, C.A., Akay, L.A. *et al*. Single-cell multiregion dissection of Alzheimer's disease. *Nature* (2024). https://doi.org/10.1038/s41586-024-07606-7
5. Das, B., & Yan, R. (2019). A Close Look at BACE1 Inhibitors for Alzheimer's Disease Treatment. *CNS drugs*, *33*(3), 251–263. https://doi.org/10.1007/s40263-019-00613-7
6. Tschanz, J.T. (2011). Senile Plaques. In: Kreutzer, J.S., DeLuca, J., Caplan, B. (eds) Encyclopedia of Clinical Neuropsychology. Springer, New York, NY. https://doi.org/10.1007/978-0-387-79948-3_491
7. Chasseigneaux, S., & Allinquant, B. (2012). Functions of Aβ, sAPPα and sAPPβ: similarities and differences. *Journal of neurochemistry*, *120 Suppl 1*, 99–108. https://doi.org/10.1111/j.1471-4159.2011.07584.x
8. Rezaei-Ghaleh, N., Giller, K., Becker, S., & Zweckstetter, M. (2011). Effect of zinc binding on β-amyloid structure and dynamics: implications for Aβ aggregation. *Biophysical journal*, *101*(5), 1202–1211. https://doi.org/10.1016/j.bpj.2011.06.062
9. Zhang, Y., Chen, H., Li, R., Sterling, K., & Song, W. Amyloid β-based therapy for Alzheimer's disease: challenges, successes and future. *Sig Transduct Target Ther* 8, 248 (2023). https://doi.org/10.1038/s41392-023-01484-7
10. Hardy, J., & Selkoe, D. J. (2002). The amyloid hypothesis of Alzheimer's disease: progress and problems on the road to therapeutics. *Science (New York, N.Y.)*, *297*(5580), 353–356. https://doi.org/10.1126/science.1072994
11. Mohandas, E.; Rajmohan, V.; Raghunath, B (2009).1. Neurobiology of Alzheimer's disease. Indian Journal of Psychiatry 51(1):p 55-61. | DOI: 10.4103/0019-5545.44908
12. Rawat, P., Sehar, U., Bisht, J., Selman, A., Culberson, J., & Reddy, P. H. (2022). Phosphorylated Tau in Alzheimer's Disease and Other Tauopathies. *International journal of molecular sciences*, *23*(21), 12841. https://doi.org/10.3390/ijms232112841
13. Bloom G. S. (2014). Amyloid-β and tau: the trigger and bullet in Alzheimer disease pathogenesis. *JAMA neurology*, *71*(4), 505–508. https://doi.org/10.1001/jamaneurol.2013.5847
14. Polishchuk, P. G.; Madzhidov, T. I.; Varnek, A. Estimation of the size of drug-like Chemical Space based on GDB-17 data. Journal of Computer-Aided Molecular Design (2013), 27, 675–679.
15. Mouchlis, V. D., Afantitis, A., Serra, A., Fratello, M., Papadiamantis, A. G., Aidinis, V., Lynch, I., Greco, D., & Melagraki, G. (2021). Advances in de Novo Drug Design: From Conventional to Machine Learning Methods. *International Journal of Molecular Sciences*, *22*(4), 1676. https://doi.org/10.3390/ijms22041676
16. Sousa, T.; Correia, J.; Pereira, V.; Rocha, M. (2021). Generative deep learning for targeted compound design. Journal of Chemical Information and Modeling, 61, 5343–5361.
17. DiMasi, J. A.; Grabowski, H. G.; Hansen, R. W. (2016). Innovation in the pharmaceutical industry: New estimates of R&D costs. Journal of Health Economic, 47, 20–33.
18. Rashid, M. B. (2021). Artificial intelligence effecting a paradigm shift in drug development. SLAS Technology, 26, 3–15.
19. Xue, H.; Li, J.; Xie, H.; Wang, Y. (2018). Review of Drug Repositioning approaches and resources. International Journal of Biological Sciences, 14, 1232–1244.
20. Tong, X., Liu, X., Tan, X., Li, X., Jiang, J., Xiong, Z., Xu, T., Jiang, H., Qiao, N., & Zheng, M. (2021). Generative Models for De Novo Drug Design. *Journal of medicinal chemistry*, *64*(19), 14011–14027. https://doi.org/10.1021/acs.jmedchem.1c00927
21. Vanhaelen, Q.; Lin, Y.-C.; Zhavoronkov, A. (2020). The advent of Generative Chemistry. ACS Medicinal Chemistry Letters, 11, 1496–1505 https://doi.org/10.1021/acsmedchemlett.0c00088
22. Popova, M., Isayev, O., & Tropsha, A. (2018). Deep reinforcement learning for de novo drug design. *Science advances*, *4*(7), eaap7885. https://doi.org/10.1126/sciadv.aap7885
23. Watson, J.L., Juergens, D., Bennett, N.R. *et al*. De novo design of protein structure and function with RFdiffusion. *Nature* 620, 1089–1100 (2023). https://doi.org/10.1038/s41586-023-06415-8
24. Jumper, J., Evans, R., Pritzel, A. *et al*. Highly accurate protein structure prediction with AlphaFold. *Nature* 596, 583–589 (2021). https://doi.org/10.1038/s41586-021-03819-2
25. Tang, M., Li, B., & Chen, H. (2023). Application of message passing neural networks for molecular property prediction. *Current Opinion in Structural Biology*, *81*, 102616.
26. Selkoe, D. J. Alzheimer's disease: genes, proteins, and therapy. *Physiol. Rev*. 81, 741–766 (2001). https://journals.physiology.org/doi/full/10.1152/physrev.2001.81.2.741

27. Silva Hecimovic., Jun Wang., Georgia Dolios., Maribel Martinez., Rong Wang., & Alison M. Goate. (2004). Mutations in APP have independent effects on Aβ and CTFγ generation. Neurobiology of Disease, Volume 17, Issue 2, ISSN 0969-9961. https://doi.org/10.1016/j.nbd.2004.04.018.
28. Zhang, Y., Chen, H., Li, R., Sterling, K., & Song, W. Amyloid β-based therapy for Alzheimer's disease: challenges, successes and future. *Sig Transduct Target Ther* 8, 248 (2023). https://doi.org/10.1038/s41392-023-01484-7
29. Evin, G., & Hince, C. (2013). BACE1 as a therapeutic target in Alzheimer's disease: rationale and status. *Drugs & Aging*, *30*(10), 755–764. https://doi.org/10.1007/s40266-013-0099-3
30. Hall, A., & Gijsen, H. J. M. (2017). Targeting β-secretase (BACE) for the treatment of Alzheimer's disease. Doi: 10.1016/B978-0-12-409547-2.13809-0.
31. Yan, R., Han, P., Miao, H., Greengard, P., & Xu, H. (2001). The transmembrane domain of the Alzheimer's beta-secretase (BACE1) determines its late Golgi localization and access to beta-amyloid precursor protein (APP) substrate. *The Journal of biological chemistry*, *276*(39), 36788–36796. https://doi.org/10.1074/jbc.M104350200
32. Panza, F., Solfrizzi, V., Imbimbo, B. P., & Logroscino, G. (2014). Amyloid-directed monoclonal antibodies for the treatment of Alzheimer's disease: the point of no return?. *Expert opinion on biological therapy*, *14*(10), 1465-1476.
33. Van Dyck, C. H., Swanson, C. J., Aisen, P., Bateman, R. J., Chen, C., Gee, M., Kanekiyo, M., Li, D., Reyderman, L., Cohen, S., Froelich, L., Katayama, S., Sabbagh, M., Vellas, B., Watson, D., Dhadda, S., Irizarry, M., Kramer, L. D., & Iwatsubo, T. (2023). Lecanemab in Early Alzheimer's Disease. *The New England Journal of Medicine*, *388*(1), 9–21. https://doi.org/10.1056/NEJMoa2212948
34. European Medicines Agency 26 July 2024 https://www.ema.europa.eu/en/documents/smop-initial/questions-answers-refusal-marketing-authorisation-leqembi-lecanemab_en.pdf
35. Imai, K., Takaoka, A. Comparing antibody and small-molecule therapies for cancer. *Nat Rev Cancer* 6, 714–727 (2006). https://doi.org/10.1038/nrc1913
36. Bazzari FH, Bazzari AH. BACE1 Inhibitors for Alzheimer's Disease: The Past, Present and Any Future? *Molecules*. 2022; 27(24):8823. https://doi.org/10.3390/molecules27248823
37. May, P. C., Willis, B. A. *et al.* (2015). The potent BACE1 inhibitor LY2886721 elicits robust central Aβ pharmacodynamic responses in mice, dogs, and humans. *The Journal of neuroscience: the official journal of the Society for Neuroscience*, *35*(3), 1199–1210. https://doi.org/10.1523/JNEUROSCI.4129-14.2015
38. Wang, L., Wang, N., Zhang, W., Cheng, X., Yan, Z., Shao, G., Wang, X., Wang, R., & Fu, C. Therapeutic peptides: current applications and future directions. *Sig Transduct Target Ther* 7, 48 (2022). https://doi.org/10.1038/s41392-022-00904-4
39. Vassar, R. BACE1 inhibitor drugs in clinical trials for Alzheimer's disease. *Alz Res Therapy* 6, 89 (2014). https://doi.org/10.1186/s13195-014-0089-7
40. Wells J.A., McClendon C.L. Reaching for high-hanging fruit in drug discovery at protein–protein interfaces. *Nature.* (2007); 450:1001–1009. doi: 10.1038/nature06526.
41. Buchwald P. Small-molecule protein–protein interaction inhibitors: Therapeutic potential in light of molecular size, chemical space, and ligand binding efficiency considerations. *IUBMB Life*. (2010); 62:724–731. doi: 10.1002/iub.383.
42. Qian Z., Dougherty P.G., Pei D. Targeting intracellular protein-protein interactions with cell-permeable cyclic peptides. *Curr. Opin. Chem. Biol.* (2017); 38:80–86. doi: 10.1016/j.cbpa.2017.03.011.
43. Ran X., Gestwicki J.E. Inhibitors of protein–protein interactions (PPIs): An analysis of scaffold choices and buried surface area. *Curr. Opin. Chem. Biol*. (2018); 44:75–86. doi: 10.1016/j.cbpa.2018.06.004.
44. Monti, A., Vitagliano, L., Caporale, A., Ruvo, M., & Doti, N. (2023). Targeting Protein-Protein Interfaces with Peptides: The Contribution of Chemical Combinatorial Peptide Library Approaches. *International journal of molecular sciences*, *24*(9), 7842. https://doi.org/10.3390/ijms24097842
45. Reynolds, C. H., & Holloway, M. K. (2011). Thermodynamics of ligand binding and efficiency. *ACS medicinal chemistry letters*, *2*(6), 433-437.
46. Chodera, John D., and David L. Mobley. "Entropy-enthalpy compensation: role and ramifications in biomolecular ligand recognition and design." *Annual Review of Biophysics* 42 (2013): 121.
47. Ohtaka, H., & Freire, E. (2005). Adaptive inhibitors of the HIV-1 protease. *Progress in biophysics and molecular biology*, *88*(2), 193-208.
48. Freire, Ernesto. "A thermodynamic approach to the affinity optimization of drug candidates." *Chemical Biology & Drug Design* 74.5 (2009): 468–472.
49. Bechtler, C., & Lamers, C. (2021). Macrocyclization strategies for cyclic peptides and peptidomimetics. *RSC medicinal chemistry*, *12*(8), 1325–1351. https://doi.org/10.1039/d1md00083g
50. Ward, Walter H. J., and Geoffrey A. Holdgate. "Isothermal titration calorimetry in drug discovery." P*rogress in Medicinal Chemistry* 38 (2001): 309–376.
51. Xiang, L., Si, C., Wei-Dong, Z., & Hong-Gang, H. (2020). Stapled Helical Peptides Bearing Different Anchoring Residues Chemical Reviews 2020 120 (18), 10079-10144 DOI: 10.1021/acs.chemrev.0c00532
52. Bernal, F., & Katz, S. G. (2014). Synthesis of stabilized alpha-helical peptides. *Methods in molecular biology (Clifton, N.J.)*, *1176*, 107–114. https://doi.org/10.1007/978-1-4939-0992-6_9
53. Ke, C., Liuliu, H., Bin Shen. (2019). Rational cyclization-based minimization of entropy penalty upon the binding of Nrf2-derived linear peptides to Keap1: A new strategy to improve therapeutic peptide activity against sepsis. *Biophysical Chemistry* https://doi.org/10.1016/j.bpc.2018.11.002.
54. Hu, H., Chen, Z., Xu, X., & Xu, Y. (2019). Structure-Based Survey of the Binding Modes of BACE1 Inhibitors. *ACS chemical neuroscience*, *10*(2), 880–889. https://doi.org/10.1021/acschemneuro.8b00420
55. Shimizu, H., Tosaki, A., Kaneko, K., Hisano, T., Sakurai, T., & Nukina, N. (2008). Crystal Structure of an Active Form of BACE1, an Enzyme Responsible for Amyloid β Protein Production. *Molecular and Cellular Biology*, *28*(11), 3663–3671. https://doi.org/10.1128/MCB.02185-07
56. University of Washington (2023). Institute for Protein Design. Joseph, W., David J., Nathaniel, B., Brian T., & Jason, Y. https://colab.research.google.com/github/sokrypton/ColabDesign/blob/main/rf/examples/diffusion.ipynb#scrollTo=TuRUfQJZ4vkM

Author

Nicholas Chen is currently a senior at ACG Parnell College in New Zealand. He is interested in pharmacology, specifically leveraging artificial intelligence for drug discovery and development, and hopes to major in Biochemistry in college.

DOI: 10.36838/v7i4.29