Optimized Pathway for Designing Small Molecules with a Hybrid Quantum-Classical Generative Adversarial Network

Leena Adwankar[1,2], Nitya Pisolkar[1,3], Meera Minocha[1,4], Vamika Tawari[1,5], Akhil Muthyala[1,6], Sakshi Buregoni[1,5*], Victoria Ho[1,7*] & Larry McMahan[1]

Received May 18, 2026
Accepted September 27, 2026
Electronic access November 15, 2026

The modern process of drug discovery requires large financial investment and time, which is why researchers are utilizing new computational chemistry methods like machine learning to accelerate molecular synthesis. Alzheimer's disease is a neurodegenerative disease associated with the aggregation of many proteins, including Butyrylcholinesterase (BChE). The treatment of its symptoms requires molecules that can penetrate the highly selective blood-brain barrier (BBB) and inhibit BChE. Our model architecture utilizes a Wasserstein generative adversarial network (GAN) with a quantum generator, classical discriminator, and a gradient penalty. It was trained selectively on BChE inhibitors from the IPAD-DB database of Alzheimer's protein inhibitors. Post-processing algorithms checked size, lipophilicity, planarity, aromatic rings, and other functional groups critical for effective molecules. Out of 100 generated molecules, 100% were small, 89% had amine groups, 30% passed BBB permeability checkers, and 93% were predicted to have low-aggregation potential. From molecular docking, 100% had docking scores of the generated BCHE inhibitors ranging from -8 to -4, indicating moderate binding capability. Out of the 100 initially generated molecules, 5 molecules showed promising BBB permeability as well as moderate docking scores to BChE through molecular docking, proving to be preliminary computational leads that, with extensive experimental validation, could show potential as real Alzheimer's symptomatic treatment drugs. This framework demonstrates the benefits of a hypothesis-driven, hybrid quantum-classical approach to accelerating early-stage identification of BBB-penetrant BChE inhibitors while reducing experimental time.

Keywords: Drug Discovery, Computational Chemistry, Machine Learning, Generative Adversarial Network, Alzheimer's Disease

Introduction

Alzheimer's disease is one of many progressive neurodegenerative diseases. It is the most common cause of dementia, affecting millions of individuals worldwide and becoming a large public health risk[1]. Symptoms of this disease include memory loss, trouble concentrating, mood swings, aggression, and a decline in motor skills in late stages; currently, there is no cure. However, drug discovery is a long and rigorous process for this disease, often taking five to ten years and costing billions of dollars[2]. One solution to speed up early-stage progress is using computational methods, to come up with potential candidates that could be utilized in further wet lab testing, enabling the efficient screening of high-volume chemical spaces while reducing experimental cost and time[3].

Classical generative adversarial networks offer great potential but face significant challenges in handling the high-dimensional and highly nonlinear datasets associated with Alzheimer's disease, such as neuroimaging[4]. These challenges come from the complexity of biological systems and the high growth of feature interactions in molecular data. A GAN is a type of deep learning model consisting of a generator that creates synthetic data and a discriminator that evaluates its authenticity. Recently, the concept of GANs has surfaced with a paradigm that fully utilizes quantum entanglement to describe complex probability distributions and thereby detect complex interactions among biomarkers for disease. Quantum-based models use principles such as superposition to encode and process high-dimensional and high volumes of information more efficiently than classical counterparts[5]. To provide further detail and background, a Generative Adversarial Network is composed of two neural networks that work in opposition: a generator and a discriminator. Both networks are trained together in a competitive process: the discriminator

[1] *Department of Computer Science and Engineering, Aspiring Scholars Directed Research Program, Fremont, California*
[2] *Irvington High School*
[3] *Archbishop Mitty High School*
[4] *Saratoga High School*
[5] *Sakshi Buregoni, American High School*
[6] *Emerald High School*
[7] *Harker Upper School*
* *These authors contributed equally to this work*

© The National High School Journal of Science 2026

improves its ability to distinguish real from generated samples, while the generator improves its ability to produce outputs that can fool the discriminator[6]. Through repeated feedback and weight updates, the networks progressively refine their outputs, ultimately allowing the generator to produce highly realistic and diverse samples that resemble the original dataset[7]. Quantum computing uses the principles of quantum mechanics to process information with better speed and accuracy than comparable classical methods. However, it can require specific quantum hardware with large memory, so it is best implemented by simulation[8]. The Python package pennylane has functions to create the quantum simulation with the model itself[9]. It takes the normal input for a generator, but instead runs it through a parameterized quantum circuit[10]. The hybrid GAN format keeps the discriminator classical, reducing unnecessary training time and resources otherwise needed[11].

Quantum computing primarily utilizes the principles of superposition and entanglement as ways to improve the speed and diversity of the model[12]. In a classical neural network, each unit processes a single numerical state at a time, meaning the model must evaluate possible molecules sequentially. Because the quantum circuit also encodes complex probability distributions through entangled qubit states, it can represent high-dimensional molecular relationships with fewer parameters[13]. This means the model can more effectively capture subtle interactions between molecular substructures that influence properties such as docking scores and BBB permeability. Consequently, the quantum generator may produce a broader and more chemically diverse set of candidate molecules compared to a classical GAN.

However, the use and application of Quantum GANs (QGANs) for neurodegenerative diseases is still uncharted, with limited research being carried out for generating potential drug candidates[14]. At the moment, much of this is due to data homogeneity and the lack of the extensive amount of data and computational resources needed to be able to conduct training models for this purpose[15].

This work looks at the potential of hybrid quantum–classical generative architectures for molecular generation and evaluation (Figure 1).

We hypothesized that incorporating a quantum generator within a GAN framework could more efficiently explore chemical space than a classical architecture (Figure 1). Using a curated dataset of BChE inhibitors, the model generated a diverse set of chemically feasible molecules while maintaining features critical for potential BChE therapeutics and inhibition.

Methods

This study follows a retrospective, cross-sectional research design utilizing in-silico predictive modeling. No human subjects or personal data were used, and the results are not meant to be directly used in clinical trials, as our research works on accelerating the very preliminary stages of drug discovery research. General data processing, numerical computation, and visualization were carried out using NumPy[16], Pandas[17], and Matplotlib[18] in Python, and cheminformatics operations (substructure matching, sanitization, and descriptor calculation) were performed using RDKit[19].

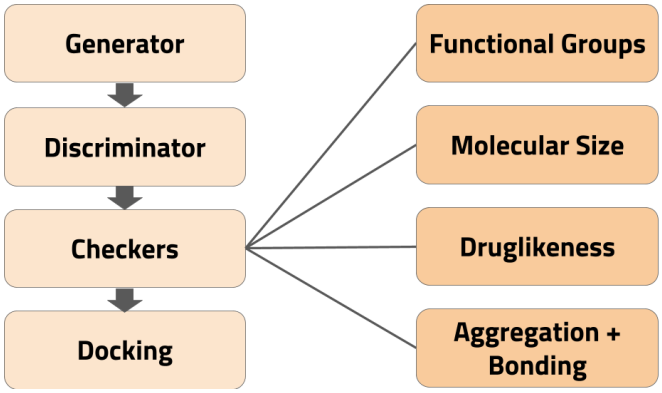


Figure. 1 Four-stage pipeline for generating and validating candidate BChE inhibitors. A quantum generator produces molecular structures, which are evaluated by a classical discriminator, screened using drug-relevant property checkers, and finally assessed through molecular docking against BChE. Training used a curated dataset of fewer than 300 experimentally verified BChE inhibitors.

Phase I: Molecular Preprocessing and Filtering

This study was conducted in three phases. The first phase involved evaluating the properties of the different molecules in order to identify molecules that could have the ability to inhibit BChE and identify the qualities needed for potential blood-brain barrier (BBB) permeability. Successfully crossing the BBB remains one of the primary pharmacokinetic obstacles in developing effective central nervous system therapeutics[20]. Because no single computational method fully captures BBB permeability, this phase combined two independent checkers.

In order to check for BBB permeability, we applied the Python implementation of the Brain Or IntestinaL EstimateD permeation method (BOILED-Egg method), which uses lipophilicity and polarity to graph Log of Partition Coefficient (WLogP) vs Topological Polar Surface Area (TPSA). Molecules with properties lying within the yellow region of the graph are likely to be BBB-permeable. WLogP (lipophilicity) values were calculated using the Wildman-Crippen algorithm to estimate the logarithm of the octanol-water partition coefficient. TPSA (topological polar surface area) values were calculated by summing the contributions of

© The National High School Journal of Science 2026

polar atoms to the molecular surface area, following standard computational methods. These parameters were then used as inputs to the BOILED-Egg model.

As a second check on BBB permeability, we also implemented the BBB Score algorithm of Gupta et al.[21], which sums five weighted molecular descriptors, aromatic ring count, heavy atom count, a hydrogen-bond count, TPSA, and pKa, into a single score ranging from 0 to 6, with scores of 4 or above classified as CNS-active. Because RDKit does not provide a pKa predictor, the pKa term was held at the algorithm's own disclosed dataset-mean default of 8.81 for every molecule. Pairing this method with BOILED-Egg provided two independent permeability estimates rather than relying on a single model.

Next, we designed algorithms ("checkers") to check our future generated molecules for the characteristics often present in strong inhibitors. One such characteristic checked was molecular weight. The ideal range is between 400 and 600 Daltons, which is calculated by summing up individual atomic masses using standard (average) atomic weights.

In addition, we applied Lipinski's Rule of Five[22], a general druglikeness filter. This rule evaluates four criteria: molecular weight, calculated octanol-water partition coefficient, hydrogen bond donor count, and hydrogen bond acceptor count. A molecule was classified as passing when it violated none of these four criteria.

We also searched for functional groups within the molecules, including amine, carbonyl, and hydroxyl groups and aromatic rings. The purpose of this was to check that the molecules were developing real molecular structures, not irregular fragments. Each of the rings has varying purposes in bioactive molecules, but are no measure of their ability to inhibit BChE. The aromatic ring may enhance target binding through hydrophobic and pi-pi interactions[23]. The amine group can influence BBB permeation depending on its ionization state at physiological pH[24]. The carbonyl group may contribute to molecular stability and binding via hydrogen-bonding interactions. The hydroxyl group increases polarity and can improve aqueous solubility. Functional group detection was done using substructure matching based on pre-built RDKit library functions, with each molecule being searched for the presence of these identifiers.

Other checkers ensure that the molecule interacts correctly with the protein and body. These checkers look for rotatable bonds and self-aggregation potential. Rotatable bonds ensure flexibility to attach to BChE, while self-aggregation prevention prevents extra build-up of the molecule. These flexibility and polarity requirements were also formalized using the Veber criteria[25], which classify a molecule as bioavailable when it has 10 or fewer rotatable bonds and a TPSA of 140 $\AA^2$ or less.

To confirm that the molecules have high bioavailability, we looked at the generated molecules' self-aggregation potential. The lower the self-aggregation potential, the greater the likelihood there is that the inhibitor molecule would be able to maintain its stability. The self-aggregation potential metric, which helps approximate the tendency of molecules to form intermolecular clusters, was estimated via the aforementioned metrics; specifically, WLogP, TPSA, and the number of hydrogen bond acceptors and donors.

Additionally, to assess whether or not generated molecules represented reproductions of the training data, each generated molecule's Tanimoto similarity to its nearest neighbor in the IPAD-DB reference library was calculated using Morgan fingerprints, with lower similarity indicating greater novelty. To see whether the generator produced a structurally varied set of molecules, we additionally calculated internal diversity, defined as the mean pairwise Tanimoto distance of all molecules in a batch, alongside validity and uniqueness.

Using a list of successes and failures of the checkers, we constructed two databases: desirable and undesirable molecules. These datasets functioned as the foundation for the second phase of molecular generation, providing negative and positive examples for the generator and discriminator.

Phase II: Molecular Generation

The second phase involves the generation of molecules. The model used a WGAN-GP architecture with a quantum generator, classical discriminator, and gradient penalty. We trained the two neural networks for 150 epochs with a learning rate of 2×10^{-5}. The generator outputs 100 SMILES strings based on a sample of 127 molecules from IPAD-DB[26], 114 for training and 13 for testing with a batch size of 1 graph per gradient step, specifically the set of BChE inhibitors from their open-access natural compounds, synthetic compounds, and drug collections. SMILES (Simplified Molecular Input Line Entry System) strings are a way to represent chemical structures using short ASCII text strings, encoding atoms, bonds, and connectivity in a linear text format (Figure 2).

The quantum element was implemented into the generator with the Python package PennyLane[27], creating a parameterized quantum circuit. Quantum hardware is not necessary for computation, and these results were attained only run on CPU.

The QuantumDenseLayer, utilizing the quantum principles of superposition and entanglement, amplitude-embeds its input into n qubits (n=6 in the generator, n=7 in the critic), applies a single StronglyEntanglingLayers operation (depth 1), and measures the Pauli-Z expectation on each wire. The QLSTM cell comprises four sub-circuits (forget, input, update, and output gates), each applying a linear projection followed by angle embedding, one BasicEntanglerLayers operation (depth 1) of parameterized rotations, and per-wire Pauli-Z measurement, recombined through standard sigmoid/tanh gat-

© The National High School Journal of Science 2026

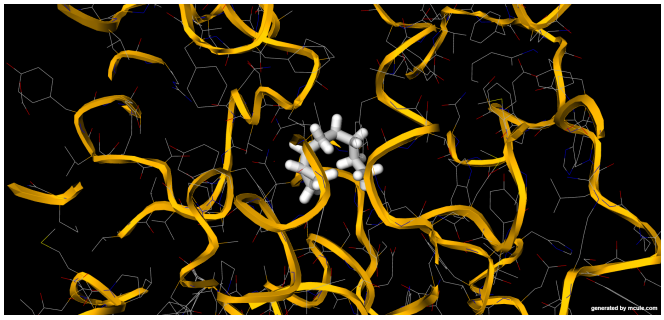


Figure. 2 Representative Molecule 1-Click Docking rendering using AutoDock Vina, showing one of 100 generated molecules docked in human BChE (PDB: 1P0I). The docking grid was centered on Ser198 (X = 137.508, Y = 120.311, Z = 41.198), 22 Å per side, exhaustiveness = 2. AutoDockTools prepared the receptor (Gasteiger charges assigned, crystallographic waters removed); Mcule's stereoisomer generator resolved ligand stereocenters. The top-ranked pose by Vina score was retained per ligand.

ing.

All circuits run on the analytic default.qubit statevector simulator through the PyTorch[28] interface. This is to ensure measurements are shot-free and gradients are obtained by backpropagation rather than the parameter-shift rule.

The discriminator is a Wasserstein critic (input projection, quantum dense layer, quantum LSTM, adaptive pooling, a linear layer with ReLU and dropout, and a single unbounded output), while the generator combines MFConv and GCNConv graph layers with the quantum layers and forms the adjacency matrix as the Gram product of its output. The training uses WGAN with gradient penalty over graphs of up to 64 atoms for 150 epochs, updating on one graph at a time. Hydrogens are excluded, and molecules exceeding the atom limit are omitted rather than truncated. The critic-to-generator update ratio was 5:1 for epochs 1–30, 2:1 for 31–60, and 1:1 thereafter; the gradient-penalty coefficient was 10; optimization used Adam (β_1 = 0.5, β_2 = 0.9) at a learning rate of 2×10^{-5} with cosine annealing and warm restarts; the latent dimension was 64. Checkpoints were selected by the smallest absolute estimated Wasserstein distance among epochs whose gradient-penalty term fell below 0.5, so that a critic that had ceased to discriminate could not be selected.

Generation proceeds on a graph scaffold sampled independently of any training molecule. Atom types are assigned subject to a valence budget determined by each node's graph degree, and bond orders (single, double, triple) are read from the magnitudes of the generator's adjacency output, again constrained by remaining valence. Candidate five- and six-membered rings are tested for aromaticity and retained where RDKit sanitization succeeds. Only the largest connected fragment is kept. Stereochemistry is not recovered. RDKit sanitization is applied to every candidate during generation, and structures that fail are discarded and the attempt repeated, with all attempts counted so that reported validity reflects the full generation budget.

To test whether any performance advantages, such as training time and accuracy, are attributable to the quantum layers rather than to architecture or training settings, the hybrid GAN was run twice, and a classical version was trained under identical hardware, software, data splits, and hyperparameters. For an n-qubit circuit, the quantum layer utilizes O(n) trainable weights, whereas the equivalent classical dense layer operating on a 2n-dimensional input vector requires O(n*2n) weights. This allows the architecture to operate in high-dimensional feature spaces while keeping total trainable quantum parameters to 91 across the entire model. Additionally, to benchmark against existing models, results were compared to the open-source model TenGAN. Examples of metrics weighed include: validity, uniqueness, novelty, and internal diversity.

Finally, we evaluated the model using solubility, synthesizability, druglikeness, and the checkers. Using the checkers previously designed in Phase 1, we assessed the generated molecules based on their size, presence of functional groups, rotatable bonds, self-aggregation potential, flat structure, BBB analysis, Lipinski and Veber criteria, and Tanimoto similarity to the existing dataset in order to gain a better understanding of the chemical properties present in our molecules. We also examined the novelty, validity, and internal diversity of the molecules.

Phase III: Model Optimization and Docking

The third phase focused on improving the existing model. The principal hyperparameter varied was the learning rate for the generator and critic, tested at 1×10^{-4} and 2×10^{-4} before settling on 2×10^{-5}, the value used for all reported runs. Otherwise, the checker set was fixed after Phase II, and no changes were made to the layer architecture. Candidate configurations were assessed informally by inspecting the chemical validity of the generated molecules and their online docking scores against BChE, rather than through a controlled sweep varying one factor at a time with repeated runs per configuration.

We performed molecular docking using Mcule's 1-Click Docking[29], which uses the AutoDock Vina scoring algorithm, pairing each of the 100 SMILES strings with the target BChE protein. Mcule's default database did not contain BChE, so we downloaded the 1P0I crystal structure of human BChE and used its biological assembly PDB file. For each SMILES string, we also isolated the main compound, if there were multiple, since only the main structure could be used for the docking. Full grid placement, exhaustiveness, and receptor/ligand preparation parameters are detailed in the Figure 2 caption[30].

© The National High School Journal of Science 2026

To validate that the docking pipeline was producing meaningful results, two compounds were docked under identical conditions: tacrine, a known BChE inhibitor, as a positive control, and acetic acid, a molecule with no known BChE activity, as a negative control.

To identify the most promising generated candidates, a scoring function was constructed that combines molecular qualities described in the caption into a single weighted 0–1 ranking score (Equation 1, Table 1). Each component was normalized to a 0–1 scale prior to weighting to ensure comparability across all metrics.

$$\text{Composite} = 0.35D + 0.2(B) + 0.05\left(\frac{G}{6}\right) + 0.2L + 0.1(1-T) + 0.06A + 0.02H + 0.02C \quad (1)$$

Equation 1: A weighted 0–1 ranking score combining docking score, both independent BBB permeability checks, druglikeness, structural novelty, and functional-group presence, with each term normalized to 0–1 prior to weighting.

To test the reproducibility of our generation, the full model was retrained and re-evaluated across two runs under identical hyperparameters and training data. Checker-based descriptor metrics remained consistent across runs, supporting the stability of the underlying generative process. Because molecular docking was performed manually and not repeated for every run, all docking scores, composite molecule-scoring results, and associated figures reported in this manuscript reflect that same single representative run.

Results

To ensure the viability of the hybrid-quantum classical architecture, the model was trained for 150 epochs using a sample of 127 molecules from the Inhibitors of Proteins associated with Alzheimer's Disease Database (IPAD-DB), an experimentally validated database.

Training time per epoch fell in two step-like drops, from a mean of 83.5 s over epochs 1–30 to 38.3 s over epochs 31–60 and 23.9 s thereafter, a 71.4% reduction between the first and last phases (Figure 4). These drops coincide exactly with the scheduled reductions in the critic-to-generator update ratio at epochs 30 and 60 (5:1, then 2:1, then 1:1) and scale approximately with it (observed ratios of 2.18 and 1.60 against 2.50 and 2.00 predicted by the schedule). The parameter-matched classical WGAN-GP, which contains no quantum layers, showed the same profile under the same schedule (1.28 s, 0.63 s, and 0.45 s for the corresponding phases; a 64.6% reduction). Within-run epoch timing therefore reflects the update schedule rather than any property of the quantum architecture, and we do not interpret it as evidence of quantum efficiency.

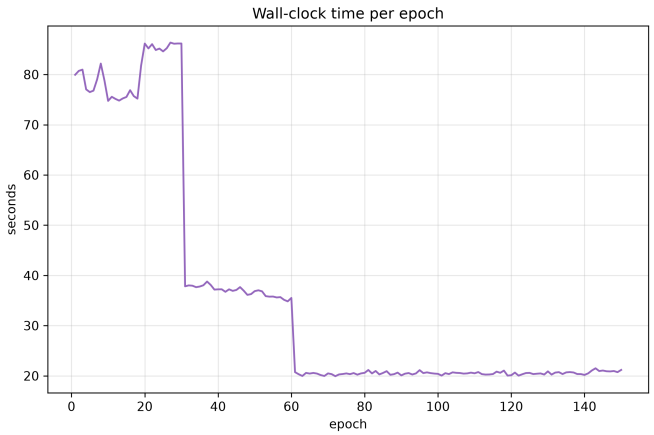


Figure. 3 Wall-clock time per Epoch. Line graph of wall-clock training time (seconds), recorded continuously for each of the 150 training epochs where training time per epoch began above 80 seconds and declined sharply in two step-like drops before stabilizing around 20 seconds by the end of training, matching the critic update ratio design.

Under matched conditions, the classical baseline trained substantially faster: 98.5 s in total (0.66 s per epoch) against 5,800.8 s (38.7 s per epoch) for the hybrid GAN, a 58.9-fold difference, with 40.7% of hybrid GAN epoch time spent inside the variational circuits. Both models satisfied the Lipschitz constraint early, the gradient-penalty term falling below 0.5 at epoch 4 (hybrid GAN) and epoch 3 (classical) and ending at 0.067 and 0.016, respectively. The estimated Wasserstein distance did not converge toward zero in either model; it rose monotonically from approximately zero to 4.58 (hybrid GAN) and 2.76 (classical) over 150 epochs, indicating that in both cases the critic continued to separate generated from training graphs rather than the generator closing the gap.

The variational circuits did receive substantial gradient signal. Mean gradient norms on the quantum parameters were 0.30 times (generator) and 0.29 times (critic) those on the classical parameters, with total quantum-weight movement of 2.17 over training, indicating that the circuits were actively optimized rather than passively carried.

Following training, the model generated 100 candidate molecules to determine whether the architecture could produce chemically feasible structures with druglike properties. All 100 were classified as small molecules; 89% contained amine functional groups; 93% exhibited low self-aggregation potential, indicating minimal risk of molecule clustering that could impair solubility; 37.2% had docking scores $\leqq -6.5$; and 100% had docking scores $\leqq -4$.

As a benchmark, the known BChE inhibitor tacrine docked at -8.2 kcal/mol, while the non-inhibitor acetic acid docked at -2.9 kcal/mol, confirming that the docking pipeline distin-

© The National High School Journal of Science 2026

Table 1 Molecule Scorer Weights: Assigned to each normalized component and the reasoning behind its relative contribution to the composite score.

Symbol	Component	Weight	Purpose
D	Docking score	0.35	Computational evidence of predicted BChE binding
B	BOILED-Egg permeability	0.2	Primary permeability signal
G	Gupta BBB Score	0.05	Cross-check against second permeability model
L	druglikeness	0.2	General developability (Lipinski, Veber, logP, MW)
T	Novelty (1-T)	0.1	Distinctness from existing inhibitors
A, H, C	Functional groups (amine, hydroxyl, carbonyl)	0.1	Structural complexity indicators



Figure. 4 WGAN-GP losses graph. Line graph tracking the generator and critic loss recorded continuously across all 150 training epochs. The critic loss dropped from 10 to -4, showing successful detection of generation flaws. The generator loss increased from 0 to 3, showing generation improvement.

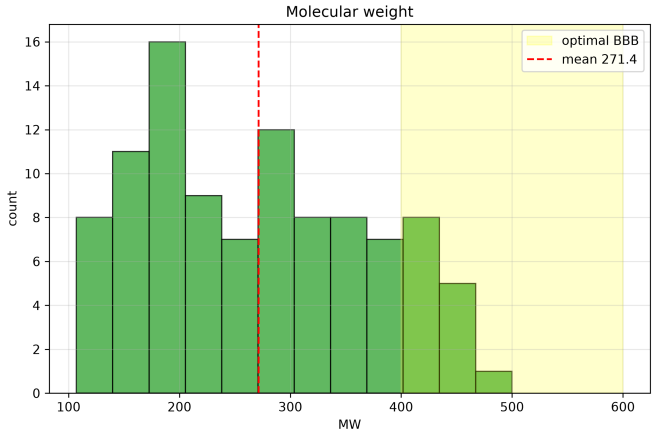


Figure. 5 Molecular Weight Distribution. Bar graph showing the frequency distribution of molecular weights across the 100 generated candidate molecules, assessed post-generation using the checker algorithms. Weights ranged from roughly 100-500 Da, generally below the optimal blood-brain barrier (BBB) permeability range of 400-600 Da (highlighted yellow area), indicating the slight heavy-atom limit imposed during the graph-to-SMILES generation step to optimize for BBB permeability. The distribution mean was 271.4 Da, indicating that while the generated molecules were consistently small enough to avoid steric exclusion, most fell short of the weight range typically associated with strong BChE-inhibitor binding.

guishes a strong reference binder from an inert control.

The generated molecules reached a mean molecular weight below the optimal BBB permeability range, though within the capabilities of small-molecule generation (Figure 6).

Notably, 34% of molecules fell within the optimal LogP range required for BBB permeability, despite the entire library remaining below the 400 Dalton threshold associated with many other CNS compounds (Figure 7).

While the average druglikeness scores remained low overall, a subset of the generated molecules exhibited moderate druglikeness (Figure 7).

Additionally, ∼37% of molecules also had hydroxyl functional groups, ∼89% had amine functional groups, and 99% showed signs of low aggregation potential, all components that aid drug performance as well. To provide an independent cross-check on BBB permeability, the same 100 candidate molecules were also scored using the BBB algorithm of Gupta et al.21, a composite metric (aromatic ring count, heavy atom count, MWHBN, TPSA, and pKa) scored from 0 to 6, with a score of 4 or above classified as CNS-active.

Gupta BBB scores ranged from 1.30 to 5.35, with a distribution mean of 3.28. 34% of molecules met the permeability threshold, including 19% scoring above 4.5, while 47% scored below 3. Cross-referencing this metric against the BOILED-Egg classification for the same 100 molecules showed agreement on permeability status for 83% of the library.

To assess whether the generated candidates represented genuinely novel chemical matter, each candidate was compared against the IPAD-DB reference library using Tanimoto similarity over Morgan fingerprints, with the single highest-similarity IPAD-DB compound retained as the nearest neighbor. Nearest-neighbor similarity scores ranged from 0.070 to

© The National High School Journal of Science 2026

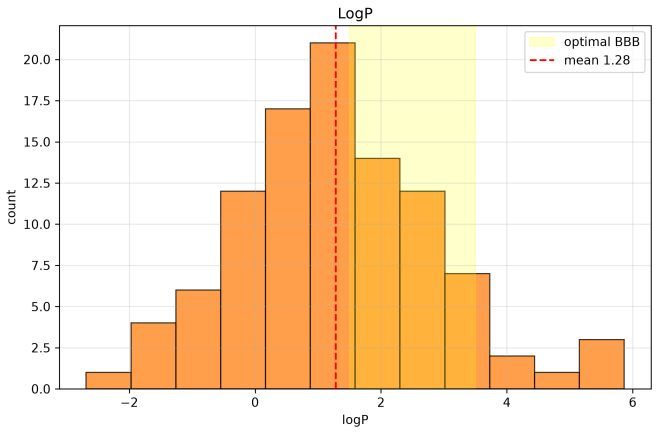


Figure. 6 LogP Distribution. Bar graph showing the frequency distribution of lipophilicity (LogP) values across the 100 generated candidate molecules, assessed post-generation using the checker algorithms, with the optimal blood-brain-barrier (BBB) permeability range of LogP 1-3 highlighted in yellow. Despite much of the entire library falling below the 400 Da molecular weight threshold typically associated with CNS-active compounds, 34% of molecules fell within the optimal LogP range. The distribution mean was 1.28, close to the lower bound of the optimal range, which suggests that lipophilicity was more readily optimized by the model than molecular weight.

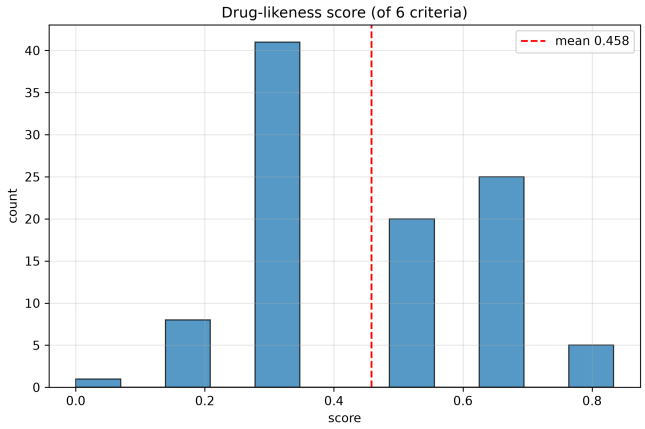


Figure. 7 Distribution of Druglikeness Scores. Bar graph showing the frequency distribution of composite druglikeness scores (0-0.8 scale) across the 100 generated candidate molecules, assessed post-generation using the checker algorithms. The majority of molecules scored low, but a subset scored in the moderate 0.5-0.7 range, with an overall distribution mean of 0.458, indicating that while most generated molecules had limited druglikeness, a good fraction showed promising druglike character.

0.214, with a distribution mean of 0.127 (SD = 0.026). Using a novelty threshold of Tanimoto similarity < 0.4, all 100 molecules were classified as structurally novel to the IPAD-DB library.

Two established rule-based druglikeness filters, Lipinski's Rule of Five and the Veber criteria25, were also applied to the 100 generated candidates. Eighty-seven percent of molecules passed all four Lipinski criteria with zero violations, with individual criterion pass rates of 99% for molecular weight, 97% for LogP, 96% for hydrogen bond acceptors, and 93% for hydrogen bond donors. Ninety-three percent of molecules passed both Veber criteria, with 97% meeting the rotatable-bond threshold ($\leq$10) and 95% meeting the TPSA threshold ($\leq$140 $\AA^2$); the mean rotatable bond count was 4.25, and mean TPSA was 68.5 $\AA^2$.

Across all 100 candidates, composite scores ranged from 0.304 to 0.743, with a mean of 0.5124 (median 0.5034).

Generation quality was assessed across two independent training runs of the hybrid GAN, differing only in random seed, and one run of the parameter-matched classical WGAN-GP. Due to the two runs producing nearly identical results, the data was assessed only against the first quantum run. The hybrid GAN produced 94.3% validity while the classical was 93.5%, with 100% uniqueness and all 100 candidates classified as novel in every run. Internal structural diversity was effectively identical, 0.909 compared to 0.912, as was nearest-neighbour Tanimoto similarity to the training library, 0.127 against 0.118. Rule-based filters gave Lipinski pass rates of 87% for the hybrid GAN and 84% for the classical, and Veber pass rates of 93% compared to 78% for the classical; blood-brain-barrier permeability was 39% and 38%; mean druglikeness on the six-criterion composite was 0.458 and 0.425.

Because the two hybrid GAN seeds bound run-to-run variation, architectural differences can be assessed against it. For validity, Lipinski and Veber pass rates, BBB permeability, aromatic-ring and carbonyl frequency, high-potential fraction, and mean druglikeness, the hybrid–classical difference did not exceed the spread between hybrid GAN seeds, and we do not attribute these differences to architecture. The exception was ring count: the hybrid GAN produced 1.68 and 1.36 rings per molecule against 0.95 for the classical baseline, a difference (0.57) larger than the seed spread (0.32). Nearest-neighbour similarity also differed by more than the seed spread, but by only 0.008 Tanimoto units, with both architectures more than ten standard deviations below the 0.4 novelty threshold. With two seeds, these observations remain suggestive rather than established.

To situate performance against an established molecular generation framework, TenGAN, a transformer-based SMILES generative adversarial network, was evaluated on the same task. TenGAN produced 50.0% validity, 41.7% uniqueness among valid outputs, 30.0% novelty among unique outputs, and internal diversity of 0.71; 54.2% of its molecules satisfied Lipinski's Rule of Five, 45.8% the Veber criteria, and 2.1% were predicted blood-brain-barrier permeable. Both

© The National High School Journal of Science 2026

architectures presented here, classical and hybrid quantum-classical GAN, exceeded these figures on every metric.

Furthermore, TenGAN generates SMILES token by token, so an invalid string is simply an invalid output. Our pipeline constructs molecules on a graph scaffold with valence-aware atom assignment and applies RDKit sanitization during generation, discarding failed attempts and retrying; validity is therefore partly enforced by construction rather than learned, and an attempt-based validity rate is not directly comparable to a token-level one. The uniqueness and diversity gaps more plausibly reflect genuine differences in chemical-space coverage.

Applying the molecular scoring function to candidates identified the following top three molecules (Table 2).

While the quantum generator did successfully produce small and structurally distinct druglike molecules with the optimal functional groups and lipophilicity, many did, in fact, fall below the optimal molecular weight range for BChE inhibition. This suggests that quantum superposition did help in chemical space; further refinement is needed to consistently generate molecules with stronger binding interactions and BBB permeability, finding the balance between our low heavy atom limit and maintaining BBB permeability.

Docking scores were compared across architectures on all successfully docked candidates. The hybrid GAN achieved a mean of –6.25 kcal/mol (SD 0.92, n = 86), with 37.2% scoring $\leq$ –6.5 and 100% $\leq$ –4, against –5.49 kcal/mol (SD 0.86, n = 91) for the parameter-matched classical baseline, with 17.6% and 86.8% respectively. The 0.76 kcal/mol difference was significant (Welch t = –5.74, $p = 4.2 \times 10^{-8}$; Cohen's d = 0.87), as was the difference in the fraction scoring $\leq$ –6.5 (Fisher p = 0.004). Unlike the descriptor-based metrics, where architectural differences fell within the spread between hybrid GAN seeds, the docking advantage exceeded it, and docking contributes the largest single weight (0.35) to the composite ranking score.

Discussion

This study examines whether a hybrid quantum-classical GAN can serve as a practical framework for early-stage candidate generation against butyrylcholinesterase (BChE), a symptomatic target in Alzheimer's disease. The motivation for a quantum generator is that parameterized circuits can encode high-dimensional distributions with comparatively few trainable weights, a property suitable when available training data is limited. The model successfully generated chemically feasible small molecules, with a high proportion exhibiting desirable druglike properties such as low aggregation potential and relevant functional groups.

Under matched conditions, the hybrid GAN architecture did not demonstrate a general efficiency advantage over classical controls. The quantum layers were actively optimized rather than passively carried: mean gradient norms on the quantum parameters reached 0.30 (generator) and 0.29 (critic) of those on the classical parameters, with cumulative quantum-weight movement of 2.17 across training. Their distinguishing property in this architecture is parameter economy rather than speed, since each variational circuit carries O(n) trainable weights for n qubits, whereas a classical layer reading the same 2^n amplitude inputs requires at least $2^n \times n$. This economy did not translate into faster convergence or better generated molecules under matched conditions, and simulating the circuits on classical hardware accounted for 40.7% of training time. We therefore make no claim of a quantum speed or convergence advantage.

Several data-related factors may have affected the outcomes of this study. The primary limitation lies in the size and specificity of the training dataset, as the model was trained on a relatively limited collection of known BChE inhibitors from IPAD-DB, which may have constrained chemical diversity and reduced the likelihood of generating highly optimized candidates. As a result, the generated molecules may contain inherent biases reflected from the training data and may not capture every aspect of chemically viable BChE inhibitors. Nearest-neighbor Tanimoto similarity to the training library was low across all 100 candidates (range 0.070–0.214, mean 0.127, SD 0.026), indicating that the model did not simply reproduce training structures. However, this similarity could be attributed to model underfitting since the estimated Wasserstein distance rose throughout training rather than converging, and no candidate greatly resembled a known BChE inhibitor.

Some generated molecules that satisfied checker-based criteria showed weaker docking performance. This result could likely be due to the arrangement of the amine groups or the need for more high-energy bonds. Additionally, molecular docking provides a simplified and static estimate of docking scores and does not account for protein flexibility or dynamic interactions. Furthermore, the relatively low molecular weights observed among generated candidates (average of 271.4 Da) reflect the heavy-atom limit created during the graph-to-SMILES step of the generation pipeline, which restricted the candidates to a small number of heavy atoms per molecule. Future architectures should relax this condition generation directly within the BBB-optimal size range, optimizing the potential BBB permeability against the optimal weight range.

Another limitation comes from the accuracy of our various checker algorithms. The BOILED-Egg method is used to determine BBB permeability, but despite its usefulness as an initial screening tool, it does not have complete accuracy in real life. Because the method relies on a fixed physicochemical property space, it does not account for biological mechanisms such as active transport or enzymatic interactions that play a critical role in BBB penetration in vivo. As a result, certain

© The National High School Journal of Science 2026

Table 2 Top three candidate molecules by composite score. SMILES, composite score, and supporting metrics for the three highest-ranked molecules, as identified by the Molecule Scorer (Equation 1).

SMILES	Composite	Docking (kcal/mol)	Gupta BBB	BOILED-Egg	druglikeness	Tanimoto Novelty
CC1=CC(C)C(C2=CC(O)=C(C)C2N)C1(C)C	0.743	-7.8	4.57	Permeable	0.667	0.867
CCC(C)C(=C1CCc2ccc(S)n2CN1COC)N(C)OC(C)O	0.737	-7.3	4.16	Permeable	0.833	0.862
C=C(C)C(CC)C1C=C=C(CC)C2CN1C(O)(NO)C2(CN(C)C)C(C)(C)C=O	0.711	-7.2	3.01	Permeable	0.833	0.837

molecules classified as BBB-permeable may fail under real conditions, while others that fall outside the defined region may still be permeable when tested experimentally. This limitation suggests that BOILED-Egg predictions should be interpreted as probabilistic, not definitive, and highlights the need for further experimental validation in the lab.

A similar issue applies to the Gupta BBB Score used as our second permeability check. It also remains a structure-only model that cannot capture biological mechanisms. Furthermore, because RDKit has no built-in pKa predictor, the pKa term was held at the algorithm's own disclosed dataset-mean fallback (8.81), which likely compresses discriminating power toward the population average. The two methods agreed on classification for 83%, but the remaining disagreements underscored that neither test should be treated as definitive without experimental validation.

Notably, the high pass rates for Lipinski's Rule of Five (87%) and the Veber criteria (93%) also stand in contrast to the much lower BBB-permeability classification rates (34–39% across our two independent checkers), indicating that general bioavailability rules alone are insufficient and reinforcing the need for the BBB-specific scoring methods used here.

Additionally, since our checker algorithms are rough heuristics, a possible future enhancement to this project would be increasing their accuracy. A possible list of additional factors for the self-aggregation checker, for example, could include the logarithm of the partition coefficient, aromatic ring count, hydrogen bonding, solubility, fraction of sp2 carbons, planarity, and charge or ionization awareness (e.g., presence of ionizable groups). In addition, numerical thresholds used in these checkers could be turned into a weighted approach.

To gain greater accuracy on the molecules generated to bypass the blood-brain barrier and inhibit BChE, our next aim is to refine our model architecture. The current WGAN-GP framework improves training stability by enforcing a Lipschitz constraint on the discriminator. To enhance the model even further and prevent mode collapse from becoming a potential issue, we will implement minibatch discrimination. By incorporating minibatch distribution, the discriminator will be able to run through numerous samples at once rather than individually, to make detection of non-diverse samples easier and thus encourage the generator to respond with more diverse results while adhering to biological constraints.

A potential future direction for improving molecule generation also comes from transitioning to a conditional GAN (cGAN) architecture. In a cGAN, molecule generation can be explicitly guided by predefined constraints, such as BBB permeability, molecular weight, functional groups, and self-aggregation. By conditioning the generator on checker-based requirements, molecules that fail to meet essential criteria could be filtered out during generation.

External comparison here is limited to a single framework. TenGAN was selected as a published, openly available generative model that could be run on the same task, but a complete benchmark would also include VAE, GraphVAE, MolGAN, and diffusion-based architectures. Each would need to be retrained on the same 127-molecule BChE subset for the comparison to be fair, rather than compared against figures reported on large general-purpose chemical datasets, which was beyond the scope of this study and is left to future work.

Apart from adding to the architecture, we will add more checkers and conduct a full sweep over learning rate, critic-to-generator update ratio, and gradient-penalty coefficient, with multiple seeds per configuration and confidence intervals.

However, the potential candidates are only early-stage computational leads and require extensive further testing and wet-lab validation to show any real promise as Alzheimer's drugs. Structural features such as functional-group presence, such as the 89% amine incidence reported above, indicate compatibility with known BChE-inhibitor chemotypes but do not themselves demonstrate active-site complementarity. Their purpose was structural validation, confirming that the generative model produces chemically complex molecules rather than fragments. Amine, carbonyl, and hydroxyl therefore speak to the generative process's ability to construct real, viable small molecules. Ultimately, target engagement can only be confirmed through purified-enzyme BChE inhibition assays with concentration-response curves yielding IC50 or Ki values. Currently, the inhibitory potential of our candidate molecules should be interpreted as an untested computational hypothesis.

The clearest architectural difference appeared in docking scores. The hybrid GAN advantage is correlated with a higher count of rings per molecule, with the complexity improving the shape complementarity in the BChE active site.

This model's distinguishing property is the molecular

© The National High School Journal of Science 2026

checker outputs and docking scores. Furthermore, as quantum hardware matures, architectures of this kind may become practical where circuit simulation currently dominates the cost, but that remains to be demonstrated rather than assumed, and further experimental testing is necessary regardless.

Acknowledgements

We thank the Aspiring Scholars Directed Research Program for providing us with the opportunity to conduct research. We give thanks to Dr. Edward Njoo for his guidance on the chemistry aspect of the project. We appreciate the work done by previous contributors Linda Chang, Sahil Vijay, Sriaditya Pendyala, and Lakshmi Narasimhan Prasanna.

References

1 S. Abdul Manap, A. Almadodi, S. Sultana, A. Sebastian, K. Kavani, N. Lyenouq, S. Shankar, Alzheimer's disease: a review on the current trends of the effective diagnosis and therapeutics. Frontiers in Aging Neuroscience. Vol. 16, 2024., https://doi.org/10.3389/fnagi.2024.1429211.

2 J. Cummings, G. Lee, A. Ritter, K. Zhong, The costs of developing treatments for Alzheimer's disease: a retrospective exploration. Alzheimer's & Dementia. Vol. 18, pg. 469–477, 2021., https://doi.org/10.1002/alz.12450.

3 E. Lin, C.-H. Lin, H.-Y. Lane, Relevant applications of generative adversarial networks in drug design and discovery: molecular de novo design, dimensionality reduction, and de novo peptide and protein design. Molecules. Vol. 25, pg. 3250, 2020., https://doi.org/10.3390/molecules25143250.

4 Q. Jiao, Y. Zhang, Z. Wang, H. Liu, X. Li, MediQ-GAN: quantum-inspired GAN for high resolution medical image generation. arXiv, 2025., https://arxiv.org/html/2506.21015v2.

5 J. Amin, M. Sharif, T. Saba, M. Raza, A. Rehman, Quantum AI for psychiatric diagnosis: enhancing dementia classification with quantum machine learning. Frontiers in Psychiatry. Vol. 16, 2025., https://doi.org/10.3389/fpsyt.2025.1648060.

6 I. Goodfellow, J. Pouget-Abadie, M. Mirza, B. Xu, D. Warde-Farley, S. Ozair, A. Courville, Y. Bengio, Generative adversarial nets. Advances in Neural Information Processing Systems. Vol. 27, 2014., https://papers.nips.cc/paper_files/paper/2014/file/5ca3e9b122f61f8f06494c97b1afccf3-Paper.pdf.

7 A. Creswell, T. White, V. Dumoulin, K. Arulkumaran, B. Sengupta, A. A. Bharath, Generative adversarial networks: an overview. IEEE Signal Processing Magazine. Vol. 35, pg. 53–65, 2018., https://doi.org/10.1109/MSP.2017.2765202.

8 M. Schuld, F. Petruccione, Supervised learning with quantum computers. Springer, 2018., https://doi.org/10.1007/978-3-319-96424-9.

9 V. Bergholm, J. Izaac, M. Schuld, C. Gogolin, C. Blank, K. McKiernan, N. Killoran, J. Carrasquilla, PennyLane: automatic differentiation of hybrid quantum-classical computations. arXiv, 2018., https://arxiv.org/abs/1811.04968.

10 PennyLane, Quantum GANs. PennyLane Demos., https://pennylane.ai/qml/demos/tutorial_quantum_gans.

11 J. Li, R. Topaloglu, S. Ghosh, Quantum generative models for small molecule drug discovery. arXiv, 2021., https://arxiv.org/abs/2101.03438.

12 M. A. Nielsen, I. L. Chuang, Quantum computation and quantum information. Cambridge University Press, 2010.

13 D. Sikdar, M. Cui, A. Chau, A. Bhamra, S. Prasanna, L. McMahan, Hybrid quantum-classical generative adversarial network for synthesizing chemically feasible molecules. Journal of Emerging Investigators, 2023., https://doi.org/10.59720/22-143.

14 L. Lan, H. You, Z. Zhang, Z. Fan, W. Zhao, N. Zeng, Y. Chen, X. Zhou, Generative adversarial networks and its applications in biomedical informatics. Frontiers in Public Health. Vol. 8, 2020., https://doi.org/10.3389/fpubh.2020.00164.

15 M. Seiler, K. Ritter, Pioneering new paths: the role of generative modelling in neurological disease research. Pflugers Archiv – European Journal of Physiology. Vol. 477, pg. 571–589, 2024., https://doi.org/10.1007/s00424-024-03016-w.

16 NumPy documentation, NumPy., https://numpy.org/doc/.

17 Pandas documentation, Pandas., https://pandas.pydata.org/docs/#.

18 Matplotlib: visualization with Python, Matplotlib., https://matplotlib.org/.

19 RDKit: open-source cheminformatics software, RDKit., https://www.rdkit.org/.

20 W. M. Pardridge, Drug transport across the blood-brain barrier. Journal of Cerebral Blood Flow & Metabolism. Vol. 32, pg. 1959–1972, 2012., https://doi.org/10.1038/jcbfm.2012.126.

21 M. Gupta, H. J. Lee, C. J. Barden, D. F. Weaver, The blood–brain barrier (BBB) score. Journal of Medicinal Chemistry. Vol. 62, pg. 9824–9836, 2019., https://doi.org/10.1021/acs.jmedchem.9b01220.

22 C. A. Lipinski, F. Lombardo, B. W. Dominy, P. J. Feeney, Experimental and computational approaches to estimate solubility and permeability in drug discovery and development settings. Advanced Drug Delivery Reviews. Vol. 23, pg. 3–25, 1997., https://doi.org/10.1016/S0169-409X(96)00423-1.

23 Y. Zhu, X. Wang, Y. Wang, Y. Wang, J. Wu, Y. Zhang, Aromatic rings as molecular determinants for the molecular recognition of protein kinase inhibitors. Molecules. Vol. 26, pg. 1776, 2021., https://doi.org/10.3390/molecules26061776.

24 H. Pajouhesh, G. R. Lenz, Medicinal chemical properties of successful central nervous system drugs. NeuroRx. Vol. 2, pg. 541–553, 2005., https://doi.org/10.1602/neurorx.2.4.541.

25 D. F. Veber, S. R. Johnson, H.-Y. Cheng, B. R. Smith, K. W. Ward, K. D. Kopple, Molecular properties that influence the oral bioavailability of drug candidates. Journal of Medicinal Chemistry. Vol. 45, pg. 2615–2623, 2002., https://doi.org/10.1021/jm020017n.

26 C. Peng, Y. Zhang, J. Liu, X. Li, H. Wang, IPAD-DB: a manually curated database for experimentally verified inhibitors of proteins associated with Alzheimer's disease. The Journal of Biological Databases and Curation. Vol. 2024, pg. baae048, 2024., https://doi.org/10.1093/database/baae048.

27 PennyLane, PennyLane., https://pennylane.ai/.

28 PyTorch, PyTorch., https://pytorch.org/.

29 Mcule.com, 1-Click Docking, 2026., https://mcule.com/apps/1-click-docking/.

30 O. Trott, A. J. Olson, AutoDock Vina: improving the speed and accuracy of docking with a new scoring function, efficient optimization, and multithreading. Journal of Computational Chemistry. Vol. 31, pg. 455–461, 2010., https://doi.org/10.1002/jcc.21334.

© The National High School Journal of Science 2026