# CNN-Based Scoring Functions Enhance Binding Energy Discrimination: A Comparative Analysis of GNINA 1.0 and AutoDock 4.2 for Acetylcholinesterase Inhibitors

Napa Boonma[1,*], Sanhajutha Puangmala[1], and Prasan Tangyuenyongwatana[2]

[1]College of Oriental Medicine, Rangsit University, Pathum Thani 12000, Thailand
[2]Department of Pharmaceutical Industry, School of Pharmacy, Eastern Asia University, Pathum Thani 12110, Thailand
*Corresponding author, E-mail: napa.b@rsu.ac.th

## Abstract

Artificial intelligence, particularly convolutional neural networks (CNNs), has revolutionized molecular docking by enhancing pose prediction and ligand discrimination capabilities. This study demonstrates the superior performance of GNINA 1.0 compared to traditional docking methods through comprehensive redocking and binding affinity analyses of acetylcholinesterase inhibitors. GNINA achieved exceptional redocking accuracy, with an average RMSD of 0.472 ± 0.01 Å (n = 3), which is well below the 2.0 Å acceptance threshold, compared to PyRx (AutoDock Vina-based), which had an average RMSD of 1.269 ± 0.25 Å. Critically, GNINA's CNN-based scoring provided superior ligand discrimination, with a broad binding affinity range of 4.11 kcal/mol (-4.83 to -8.94 kcal/mol), whereas AutoDock 4.2 exhibited limited energy differentiation, spanning only 0.7 kcal/mol (-8.7 to -9.4 kcal/mol). Linear regression analysis confirmed a strong correlation ($r^2$ = 0.7988) between GNINA's binding affinity and $IC_{50}$ values, suggesting that both measurements are effective. These results, consistent with docking's typical 2–3 kcal/mol standard error, demonstrate GNINA's ability to generate differentiated binding energies and precise poses without requiring predefined binding sites, making it significantly more effective for ligand ranking and virtual screening than empirical scoring methods.

***Keywords***: *GNINA 1.0, convolutional neural networks, molecular docking, AutoDock 4.2, acetylcholinesterase inhibitors, rhodanine derivatives*

## 1. Introduction

Traditional molecular mechanics docking methods, such as AutoDock 4.2, are widely used computational tools for predicting the binding conformations and affinities of small molecules (ligands) to macromolecular targets (proteins) with known structures (Meng et al., 2011; Leelananda et al., 2016). These methods typically treat the protein as a rigid structure, allowing only the ligand to be flexible during the docking process (Morris et al., 2009). The docking workflow involves two main components: a search algorithm that explores different ligand conformations and orientations within the protein's binding site, and a scoring function that estimates the binding affinity by evaluating intermolecular interactions using molecular mechanics force fields or empirically derived terms (Li et al., 2019). While AutoDock 4.2 uses grid-based energy evaluation and simulated annealing algorithms (Rizvi et al., 2013), AutoDock Vina introduces a more efficient and accurate scoring algorithm, achieving significant speed improvements through multithreading and improved binding mode predictions (Trott & Olson, 2010). However, these methods still have significant limitations, primarily due to inaccuracies in the scoring functions. These functions often fail to accurately predict the true binding affinity, leading to a high rate of false positives and false negatives in virtual screening campaigns (Shirali et al., 2025). The challenge of developing more precise scoring functions remains a critical area of research in the field.

The development of molecular docking software from AutoDock Vina to its successors, SMINA (Lans et al., 2020), and GNINA 1.0 represents a significant evolution in computational drug discovery, with each iteration bringing notable improvements (McNutt et al., 2021). The most significant leap in this lineage is the development of GNINA 1.0, a fork of SMINA that integrates deep learning into the docking workflow. GNINA 1.0 employs a hybrid docking pipeline that integrates traditional sampling with deep learning-based scoring to achieve superior pose prediction accuracy, and it utilizes an ensemble of convolutional neural

networks (CNNs) as a scoring function, a departure from the traditional empirical scoring functions used by its predecessors.

The methodology of the docking process begins with Monte Carlo sampling via SMINA (a fork of AutoDock Vina), which explores the ligand's conformational space using multiple parallel Markov Chain Monte Carlo (MCMC) chains controlled by the exhaustiveness parameter (default value = 8). This sampling generates candidate ligand poses within a defined binding pocket or across the entire protein surface for blind docking, with default pose retention parameters provided by the Neurosnap.ai platform determining how many poses per chain are retained for further evaluation. Following initial empirical scoring with the Vina force field, GNINA applies an ensemble of convolutional neural networks (CNNs) (Wang et al., 2020; Jin et al., 2021; Riggs & Sakidja, 2023) as the final rescoring function (McNutt et al., 2025). The 3D convolutional neural network (CNN) is used to evaluate protein-ligand interactions by treating molecular docking as a 3D image recognition task. Rather than relying solely on explicit physical equations, the software discretizes the binding site into a 3D grid of voxels (0.5 Å resolution), where different atom types such as aromatic carbons, nitrogens, and oxygens are encoded into separate density channels (Ragoza et al., 2017). This 4D tensor representation allows the CNN to learn complex, non-linear spatial features of molecular recognition directly from structural data. The network is trained to perform two simultaneous tasks: (1) classifying pose quality by predicting the probability that a ligand is within 2.0 Å RMSD of the native structure, and (2) predicting the binding affinity score of the complex. These CNNs, trained on thousands of protein-ligand crystal structures, process 3D voxelized representations of receptor-ligand complexes to predict pose quality using multiple metrics: binary pose classification assesses the geometric accuracy of the docking pose, CNN affinity estimates the binding strength, and CNN_VS ranks active compounds against non-binders during virtual screening.

Shameema et al. (2025) reported acetylcholinesterase (AChE) inhibitory activity and molecular docking of 3-α-carboxyethyl- and 3-benzamidoacetic acid rhodanine derivatives, using rivastigmine as a reference compound. The AChE inhibition activity ($IC_{50}$ values) showed distinct variations, while the docking binding affinity (kcal/mol) obtained from AutoDock 4.2 produced very similar docking scores (0.1-0.5 kcal/mol). However, in Autodock 4, the typical error of docking scores is in the range of 2.0-3.0 kcal/mol (Morris et al., 2009). Therefore, this research aims to determine whether CNN-based pose scores generated by GNINA 1.0 demonstrate stronger correlations with experimental binding data than traditional docking scores for a curated set of compounds.

## 2. Objectives

The primary objective of this study was to perform redocking of the co-crystallized ligand of AChE (PDB ID: 1EVE) under blind docking conditions. This approach was chosen to compare the performance of GNINA 1.0 (which uses SMINA) with that of PyRx 0.8 (which implements the AutoDock Vina search engine). Specifically, the blind docking condition was employed to evaluate the ability of both programs to accurately identify the correct binding pocket without prior spatial knowledge, a task at which GNINA 1.0 is designed to be highly proficient. The second objective was to evaluate the binding affinities of rhodanine derivatives reported by Shameema et al. (2025) using GNINA 1.0 and AutoDock 4.2, and to compare the linear correlations with the $IC_{50}$ values.

## 3. Materials and Methods

### *3.1 Computational Hardware*

All GNINA docking computations were performed using the Neurosnap.ai web service, a cloud-based molecular modeling platform. The server-grade hardware features NVIDIA RTX 4090 GPUs with 24 GB of GDDR6X VRAM. The designated CPU and system RAM are managed by the web service and have sufficient capacity to support GPU-accelerated docking computations without inducing computational bottlenecks.



*Proceedings of RSU International Research Conference (RSUCON-2026)*

### *3.2 Software and Docking Parameters*

Molecular docking analyses were conducted with GNINA 1.0 (https://github.com/gnina/gnina). The acetylcholinesterase receptor structure (PDB: 1EVE) was obtained from the Protein Data Bank (https://www.rcsb.org/), prepared using PDBFixer (https://github.com/openmm/pdbfixer), and ligand structures were obtained from Shameema and coworkers' report.

The GNINA 1.0 program was run with default parameters for comprehensive protein docking, leveraging its deep-learning-based scoring functions. The essential parameters include the default CNN (convolutional neural network) scoring function, which employs the primary score for affinity ranking known as CNN affinity (in kcal/mol), and the CNN pose score, a value ranging from 0 to 1 that indicates the model's prediction of the likelihood of a specific binding pose and evaluates the model's confidence in the predicted binding pose. The exhaustiveness was set to 8, the number of output poses to 10, and the random seed to 0. Redocking of the 1EVE cognate ligand was performed with GNINA 1.0 (n = 3 independent runs) using the crystallographic ligand to define the docking box under blind docking conditions, with default CNN rescoring. Vina scoring was used during Monte Carlo sampling, followed by an ensemble of 5 CNN models for final pose rescoring and ranking, which means there are 5 different neural network models (trained slightly differently) that all look at the pose at the same time. Their scores are then averaged together to get a more accurate result. SMINA-based docking (Koes et al., 2013) was used as the search algorithm with exhaustiveness set to 8, which controls the number of Monte Carlo chains and provides sufficient sampling of ligand conformational space, and 20 output poses per run were generated. The top-ranked pose from each run was compared with the crystallographic ligand, and heavy-atom RMSD values were computed using VEGA-ZZ.

To evaluate the programs' ability to independently identify the binding pocket, a blind docking approach was used to redock the cognate ligand. In this context, 'blind docking' is defined as a global search in which the grid box is configured to encompass the entire surface of the AChE protein (center: X= 16.1822, Y= 55.2552, Z= 127.3038; dimensions: X= 51.1127, Y= 59.7531, Z= 47.2471 Å), rather than being locally centered on the known catalytic site. This approach was deliberately chosen over traditional focused docking; while focused docking would likely yield similar pose predictions for both programs, a global search rigorously tests the efficiency of GNINA 1.0's CNN scoring and AutoDock Vina's search engine in locating the correct binding site without any prior spatial knowledge.

### *3.3 Docking Samples*

Docking samples were obtained from Shameema et al. (2025). The structures were drawn in 2D with ChemSketch as shown in Figure 1, then converted to 3D structures using energy minimization with the MMFF94 force field in Avogadro and saved in SDF format. A targeted subset of six compounds, alongside the reference standard rivastigmine, was selected for docking analysis based on previously published data. These compounds were chosen because they exhibit a broad range of experimental $IC_{50}$ values. This selection criterion was established to evaluate whether the advanced scoring functions of GNINA 1.0 could better differentiate binding affinities and yield stronger correlations with experimental in vitro data than traditional AutoDock 4.2 scoring, which previously showed narrow differences in binding affinities for these structures.

### *3.4 Data Analysis*

Data analysis was performed in GraphPad Prism 10. RMSD values from the three independent GNINA 1.0 redocking runs were reported as mean ± SD. Linear regression between predicted binding affinity (Y) and experimental $IC_{50}$ values (X) was conducted using simple linear regression (XY columns), with 95% confidence intervals for the regression line and parameters (slope and intercept).

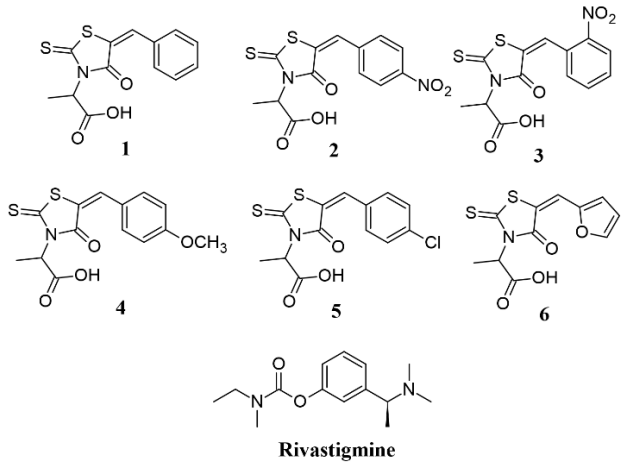


**Figure 1** Structures of rhodanine derivatives and rivastigmine

## 4. Results and Discussion

### *4.1 Docking Validation: Redock Accuracy Evaluates*

The redocking accuracy test was performed using GNINA 1.0 by conducting three independent redocking runs, and the resulting pose is shown in Figure 2A, with an average RMSD of 0.472 ± 0.01 Å. The protocol also included blind redocking using PyRx 0.8 as a reference, yielding a higher RMSD of 1.269 ± 0.25 Å. The commonly accepted criterion for a successful redocking procedure is an RMSD below 2.0 angstrom; therefore, GNINA 1.0 demonstrates higher redocking accuracy, while the PyRx redocking results are also below 2.0 angstrom. PyRx uses AutoDock Vina as a docking engine, which is the predecessor of SMINA. The poses of these two dockings are shown in Figure 2.

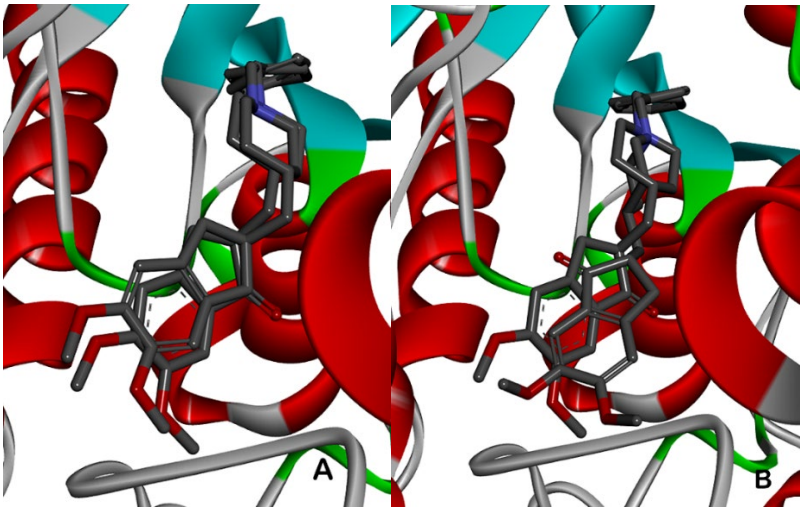


**Figure 2** Redocking of the X-ray ligand into 1EVE using GNINA 1.0 (A) and PyRx 0.8 (B)

### *4.2 Docking Samples*

Six compounds were docked into 1EVE protein using GNINA 1.0 and obtained the binding affinity (kcal/mol), and CNN pose scores, which are shown in Table 1 with the docking scores from AutoDock 4.2 (Shameema et al., 2025).

**Table 1** The $IC_{50}$ and binding energy (kcal/mol) from GNINA 1.0 and AutoDock 4.2

| Compound | $IC_{50}$ (μM) | log$IC_{50}$ (μM) | GNINA Binding affinity (kcal/mol) | CNN pose scores | AutoDock 4.2 Binding affinity (kcal/mol) |
|---|---|---|---|---|---|
| Compound 1 | 39.76 ± 0.78 | 1.60 ± 0.01 | -8.94 | 0.9265 | -9.0 |
| Compound 2 | 27.29 ± 1.08 | 1.44 ± 0.02 | -8.79 | 0.7233 | -9.3 |
| Compound 3 | 56.44 ± 3.45 | 1.75 ±0.03 | -6.87 | 0.4560 | -8.9 |
| Compound 4 | 40.44 ± 2.45 | 1.61 ± 0.03 | -8.07 | 0.8424 | -9.0 |
| Compound 5 | 34.44 ± 1.45 | 1.54 ± 0.02 | -8.28 | 0.4340 | -9.4 |
| Compound 6 | 112.58 ± 5.33 | 2.05 ± 0.02 | -4.83 | 0.5938 | -8.7 |
| Rivastigmine | 54.37 ± 4.52 | 1.74 ± 0.04 | -5.97 | 0.764 | NA |

The data support the argument that GNINA provides better ligand discrimination than AutoDock 4.2, and the results are consistent within the standard error. GNINA shows a binding affinity range of ~4.11 kcal/mol (-4.83 to -8.94 kcal/mol), whereas AutoDock 4.2 exhibits a range of ~0.7 kcal/mol (-8.7 to -9.4 kcal/mol). Figure 3(A) shows that compound 2 overlays closely with the X-ray ligand of 1EVE, while Figure 3(B) demonstrates compound 6, which has a high $IC_{50}$ and a low GNINA binding affinity. The latter structure does not overlay well with the X-ray ligand of 1EVE. For CNN pose scores, the data were compared with $IC_{50}$ values and GNINA binding affinity. The results indicate that higher CNN scores tend to be associated with stronger (more negative) binding affinities, with a Pearson correlation coefficient (r) of -0.60. In contrast, there is no strong linear relationship between the CNN scores and biological activity ($IC_{50}$), with a correlation coefficient (r) of -0.26. Therefore, a high CNN score does not reliably predict a low $IC_{50}$ (i.e., high potency).

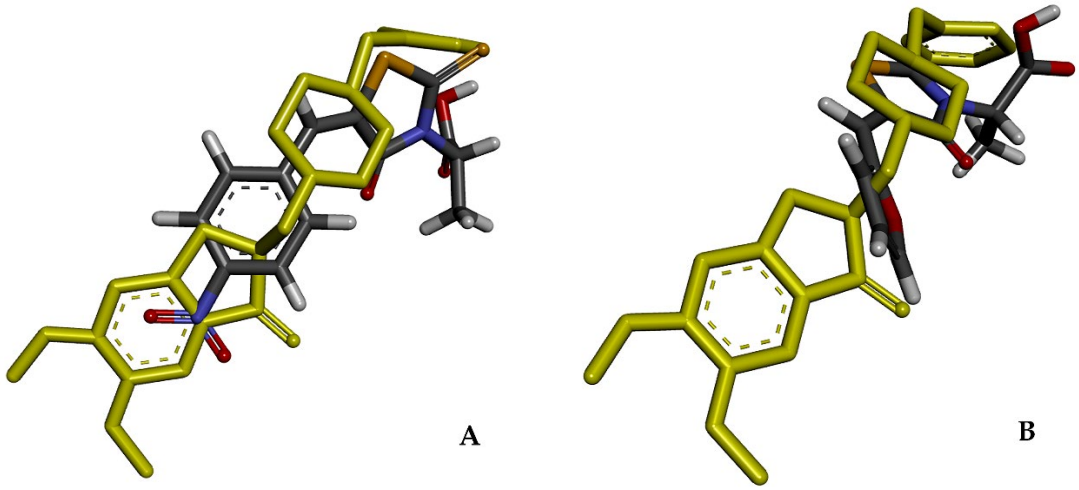


**Figure 3** (A) Compound 2 overlaid closely with the X-ray ligand of 1EVE (shown in yellow). (B) Compound 6, which has a high IC50 and low binding affinity predicted by GNINA, overlaid with the X-ray ligand (yellow)

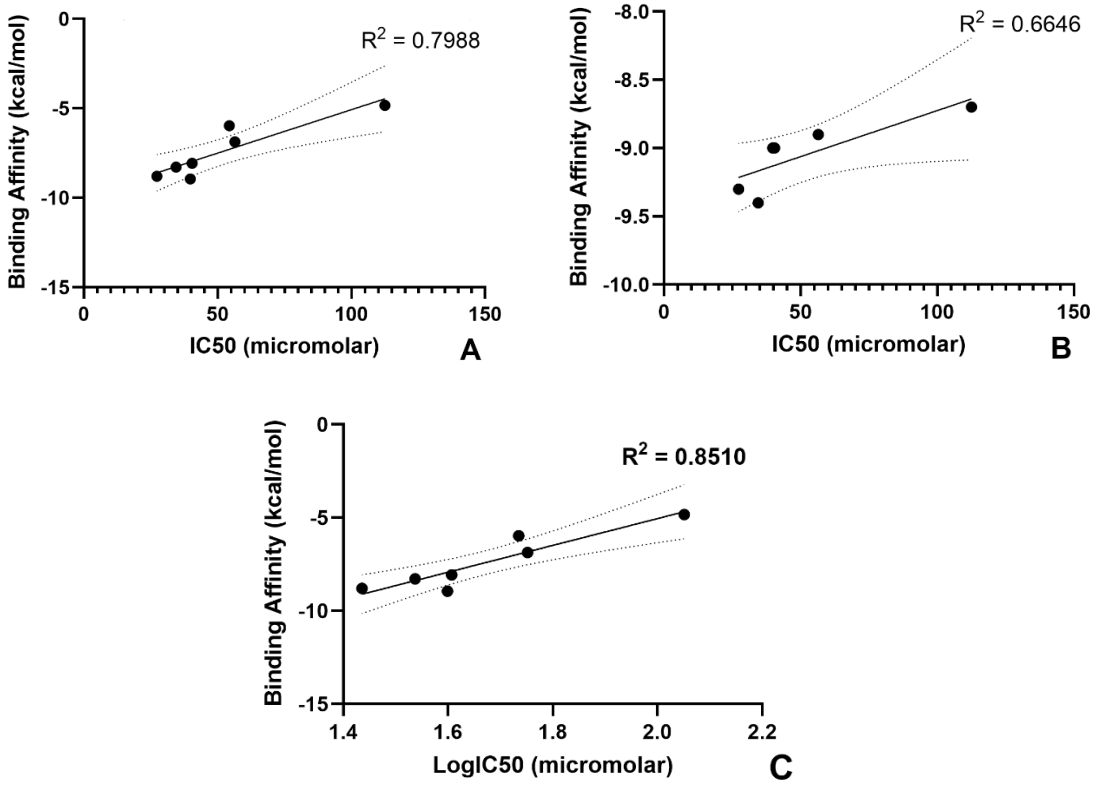


**Figure 4** Correlation graphs between binding affinity (kcal/mol) and $IC_{50}$ (µM) predicted by GNINA 1.0 (A), AutoDock 4.2 (B), and the correlation between binding affinity (kcal/mol) and $LogIC_{50}$ (µM) predicted by GNINA 1.0 (C). The Y-axes are scaled independently to clearly display the correlation trends within each dataset. Notably, a significant range compression is evident in the AutoDock 4.2 results (Graph B) compared to the wider energy range predicted by GNINA 1.0 (Graphs A and C)

From the GNINA 1.0 docking results, two parameters, binding affinity and CNN pose scores, were considered. In addition, scientific judgment is an important factor in determining which pose or ranking score is most reliable. The linear regression analysis (Figure 4A) demonstrates that GNINA docking yields a strong and significant correlation ($R^2 = 0.7988$, $p < 0.0067$, 95% CI of the slope: 0.0205 to 0.0765). However, according to thermodynamic principles, the binding free energy is linearly related to the logarithm of the binding affinity rather than the raw concentration. When the experimental data were transformed into $LogIC_{50}$, the linear correlation improved substantially ($R^2 = 0.8510$, $p < 0.0031$, 95% CI of the slope: 3.735 to 10.66). This tighter linear fit and increased statistical significance suggest that GNINA's scoring function more accurately reflects the underlying thermodynamic binding interactions of the ligands.

For binding affinity, the SMINA scoring function implemented in GNINA provides a good energy spread, ranging from -4.83 to -8.94 kcal/mol, while AutoDock 4.2 shows a compressed binding energy range, meaning that the predicted binding energies of the compounds are very close to each other. The typical error range of binding energy predictions is approximately 2.0-3.0 kcal/mol. Notably, while GNINA 1.0 demonstrated a wide range of binding energies, allowing clear differentiation between the compounds, AutoDock 4.2 exhibited significant range compression, scoring all compounds within a very narrow energy window. This narrow range limits AutoDock's ability to distinguish between compounds with distinct experimental $IC_{50}$ values.

When interpreting GNINA docking results, false positives can occur where a compound, such as compound 5, exhibits excellent binding affinity (e.g., -8.5 kcal/mol) and high CNN scores in its top-ranked poses, yet these poses lie entirely outside the binding pocket. While a ligand may have the perfect size and shape to achieve a high van der Waals score, it might lack critical hydrogen bonds or incur desolvation penalties that the scoring function does not fully capture. In the case of compound 5, although the first four ranks show strong metrics outside the binding site, the fifth rank is located within the binding pocket with a good binding energy(-8.28 kcal/mol); however, its low CNN pose score (0.4340) and inconsistent 3D interactions with the binding site compared to X-ray structure indicate that the neural network has low confidence in this predicted pose.The findings of this study have direct real-world applications in early-stage drug discovery and lead optimization. In a practical setting, the inability of traditional empirical scoring functions (like AutoDock 4.2) to distinguish between compounds with varying biological activities can lead to the costly synthesis and testing of false positives. By demonstrating that GNINA 1.0 provides a broader energy range and stronger correlation with experimental $IC_{50}$ values, this study supports the integration of deep learning-based scoring functions into standard virtual screening pipelines to more accurately prioritize lead compounds. Furthermore, as the pharmaceutical industry rapidly advances toward fully AI-driven structural predictions and docking models (e.g., DiffDock, DynamicBind), validating the reliability of CNN-based scoring functions like GNINA serves as a crucial benchmark for researchers transitioning from traditional molecular modeling to next-generation computational drug design approaches.

## 5. Conclusion

GNINA demonstrates superior performance in redocking studies, achieving a top-1 pose RMSD of $0.472 \pm 0.01$ Å. This improved accuracy is associated with GNINA's CNN-based scoring architecture, which provides superior ligand discrimination by spanning a broader, more differentiated spectrum of binding energies. In contrast, AutoDock 4.2's empirical scoring function exhibits limited energy-range differentiation, producing predictions that cluster within narrow bands and consequently lack sufficient discriminatory power to distinguish compounds with different binding affinities. These findings demonstrate that deep learning-based rescoring functions, such as those in GNINA, fundamentally enhance pose prediction accuracy and ligand ranking capability. However, the paradigm is shifting away from traditional sampling-and-scoring methods toward end-to-end generative AI docking frameworks such as DynamicBind (Lu et al., 2024) and DiffDock (Corso et al., 2023). Unlike conventional approaches that treat the protein as a rigid receptor, these newer models leverage diffusion processes and geometric deep learning to account for protein flexibility and complex conformational changes during ligand binding. By treating docking as a generative modeling problem, they offer the potential to explore a vastly larger chemical space more quickly and accurately.

## 6. Acknowledgements

The authors would like to thank the College of Oriental Medicine and the School of Pharmacy at Eastern University for their support of this research.

## 7. References

Corso, G., Stärk, H., Jing, B., Barzilay, R., & Jaakkola, T. (2023). DiffDock: Diffusion steps, twists, and turns for molecular docking. *International Conference on Learning Representations*, 1-33. https://doi.org/10.48550/arXiv.2210.01776

Jin, B., Song, Y., & Kihara, D. (2021). Guiding conventional protein–ligand docking software with convolutional neural networks. *Journal of Chemical Information and Modeling*, *61*(3), 1065–1076. https://doi.org/10.1021/acs.jcim.0c00542

Koes, D.R., Baumgartner, M.P., & Camacho, C.J. (2013). Lessons learned in empirical scoring with Smina from the CSAR 2011 benchmarking exercise. *Journal of Chemical Information and Modeling*, *53*(8), 1893-1904. https://doi.org/10.1021/ci300604z

Lans, I., Anoz-Carbonell, E., Palacio-Rodríguez, K., Aínsa, J. A., Medina, M., & Cossio, P. (2020). In silico discovery and biological validation of ligands of FAD synthase, a promising new

antimicrobial target. *PLoS Computational Biology*, *16*(8), Article e1007898. https://doi.org/10.1371/journal.pcbi.1007898

Leelananda, S. P., & Lindert, S. (2016**).** Computational methods in drug discovery. *Beilstein Journal of Organic Chemistry*, *12*(1), 2694-2718. https://doi.org/10.3762/bjoc.12.267

Li, J., Fu, A. & Zhang, L. (2019). An overview of scoring functions used for protein–ligand interactions in molecular docking. *Interdisciplinary Science Computational Life Science*, *11*(2), 320–328 https://doi.org/10.1007/s12539-019-00327-w

Lu, W., Zhang, J., Huang, W., Zhang, Z., Jia, X., Wang, Z., Shi, L., Li, C., Wolynes, P.G., Zheng, S. (2024). DynamicBind: Predicting ligand-specific protein-ligand complex structure with a deep equivariant generative model. *Nature Communication*, *15*(1), Article 1071. https://doi.org/10.1038/s41467-024-45461-2

McNutt, A. T., Francoeur, P., Aggarwal, R., Masuda, T., Meli, R., Ragoza, M., Sunseri, J., & Koes, D. R. (2021). GNINA 1.0: Molecular docking with deep learning. *Journal of Cheminformatics*, *13*(1), Article 43. https://doi.org/10.1186/s13321-021-00522-2

McNutt, A.T., Li, Y., Meli, R., Aggarwal, R., Koes, D.R. (2025). GNINA 1.3: The next increment in molecular docking with deep learning. *Journal of Cheminformatics*, *17*(1), Article 28. https://doi.org/10.1186/s13321-025-00973-x

Meng, X.Y., Zhang, H.X., Mezei, M., & Cui, M. (2011). Molecular docking: A powerful approach for structure-based drug discovery. *Current Computer-Aided Drug Design*, *7*(2), 146-157. https://doi.org/10.2174/157340911795677602

Morris, G.M., Huey, R., Lindstrom, W., Sanner, M.F., Belew, R.K., Goodsell, D.S., & Olson, A.J. (2009). AutoDock4 and AutoDockTools4: Automated docking with selective receptor flexibility. *Journal of Computational Chemistry*, *30*(16), 2785-2791. https://doi.org/10.1002/jcc.21256

Ragoza, M., Hochuli, J., Idrobo, E., Sunseri, J., & Koes, D. R. (2017). Protein-ligand scoring with convolutional neural networks. *Journal of Chemical Information and Modeling*, *57*(4), 942–957. https://doi.org/10.1021/acs.jcim.6b00740

Riggs, G., & Sakidja, R. (2023). Convolutional neural network analysis of molecular docking for drug discoveries [Poster presentation]. *2023 Fall Meeting of the APS Prairie Section*, University of Missouri, Columbia, Missouri, United States.

Rizvi, S.M., Shakil, S., & Haneef, M.A. (2013). A simple click by click protocol to perform docking: AutoDock 4.2 made easy for non-bioinformaticians. *EXCLI Journal*, *12*, 831-857.

Shameema, S., Dhanapal, S., Hariharan, M., Nandha Kumar, C. S., Siva Muthu, A., Venkatachalapathi, S., & Sundaram, K. (2025). Acetylcholinesterase inhibition activity and molecular docking studies of 3-α-carboxy ethyl/3-benzamidoacetic acid rhodanine derivatives. *Letters in Applied NanoBioScience*, *14*(1), Article 32. https://doi.org/10.33263/LIANBS141.032

Shirali, A., Stebliankin, V., Karki, U., Shi, J., Chapagain, P.P., & Narasimhan, G. (2025). A comprehensive survey of scoring functions for protein docking models. *BMC Bioinformatics*, *26*(1), Article 25. https://doi.org/10.1186/s12859-024-05991-4

Trott, O., & Olson, A.J. (2010). AutoDock Vina: Improving the speed and accuracy of docking with a new scoring function, efficient optimization, and multithreading. *Journal of Computational Chemistry*, *31*(2), 455-461. https://doi.org/10.1002/jcc.21334

Wang, X., Terashi, G., Christoffer, C.W., Zhu, M., & Kihara, D. (2020). Protein docking model evaluation by 3D deep convolutional neural networks. *Bioinformatics*, *36*(7), 2113–2118, https://doi.org/10.1093/bioinformatics/btz870