ACS OMEGA

Open Access

http://pubs.acs.org/journal/acsodf

Article

Discovering Influenza Virus Neuraminidase Inhibitors via Computational and Experimental Studies

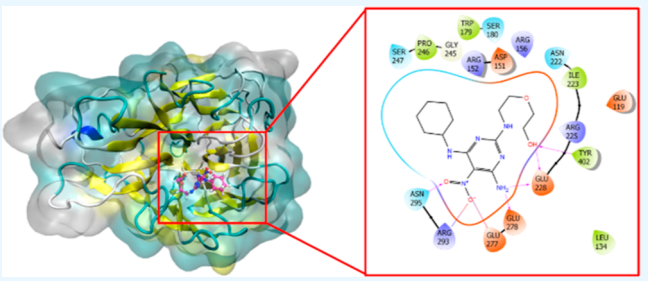
Trung Hai Nguyen,[¶] Ngoc Quynh Anh Pham,[¶] Quynh Mai Thai, Van V. Vu, Son Tung Ngo,* and Jim-Tong Horng*

Cite This: *ACS Omega* 2024, 9, 48505−48511

Read Online

ACCESS | Metrics & More | Article Recommendations | Supporting Information

ABSTRACT: Influenza A and B viruses spread out worldwide, causing several global concerns. Discovering neuraminidase inhibitors to prevent influenza A and B viruses is thus of great interest. In this work, a machine learning model was trained and tested to evaluate the ligand-binding affinity to neuraminidase. The model was then used to predict the binding affinity of compounds from the CHEMBL database, which is a manually curated database of bioactive molecules with drug-like properties. The physical insights into the binding process of ligands to neuraminidase were clarified via molecular docking and molecular dynamics simulations. Experimental investigation on enzymatic activity validated our computational results and suggested that 2 compounds were potential inhibitors of neuraminidase of the influenza A and B viruses.



INTRODUCTION

Influenza A and B viruses have caused major influenza outbreaks or pandemics that have affected millions of people worldwide.[1−3] The viruses are divided into subtypes based on two proteins on their surface: hemagglutinin (H) and neuraminidase (N). These subtypes have been responsible for major pandemics throughout the 20th and 21st centuries such as the H1N1 pandemic in 1918, H2N2 pandemic in 1957, and H3N2 pandemic in 1968[1,4] and outbreaks of H5N1,[2,5] H1N1,[6,7] H5N8,[8] and H7N9[9] in recent years. New strains of the viruses continue to emerge, and the risk of drug resistance has sparked great interest into finding potential antiviral compounds.[10−16]

Neuraminidase is a key surface glycoprotein that plays an important role in viral replication and infection. It is a proven target for developing drugs against influenza A and B viruses.[17,18] Several drugs recommended for treating influenza virus, such as oseltamivir, zanamivir, and peramivir, are neuraminidase inhibitors. Existing neuraminidase inhibitors were rationally developed many years ago since they mimic the structure of sialic acid, the natural ligand of neuraminidase. However, these drugs suffer from serious limitations such as the potential emergence of oseltamivir-resistant strains[19,20] and the poor oral bioavailability of zanamivir.[21] On the other hand, more virulent variants such as H5N1 and H7N9 have emerged. Therefore, researching novel inhibitors capable of effectively inhibiting neuraminidase and overcoming limitations of existing inhibitors continues to be a topic of interest.

Computer-aided drug design (CADD) is a powerful tool for rapid and accurate screening of several million compounds for potential inhibitors of enzymes.[22−24] The adoption of CADD methods has rapidly increased due to their potential to significantly reduce the cost and time of a new drug development.[25] CADD can be used for both searching for new inhibitors and repurposing existing drugs.[26−29] CADD has contributed to the discovery of severally available drugs such as *dorzolamide*,[24,30] *saquinavir*, *ritonavir*, and *indinavir*.[22]

The machine learning (ML) method has been widely used to predict inhibitors for neuraminidase. Classification models based on support vector machine, naïve Bayes, and random forest (RF)[31−33] have been applied to classify active versus inactive compounds for inhibiting neuraminidase. Training data include a few hundred active and inactive compounds, and standard molecular descriptors such as MACCS and ECFP4 have been used as feature input into the models. To the best of our knowledge, no regression models have been trained for neuraminidase inhibitors.

In this work, we aim to use a combination of computational and experimental approaches to find potential inhibitors for inhibiting neuraminidase. Validation of the computational approach against experiments helps enhance its reliability for future applications. In particular, the trained ML regression model was employed to predict the ligand-binding free energy

Received: August 5, 2024
Revised: November 6, 2024
Accepted: November 15, 2024
Published: November 25, 2024

of approximately 2 million compounds of the ChEMBL database to neuraminidase. The experimental studies were then carried out to validate the ML outcomes. The shortlist of potential candidates was obtained. The experiment would then validate the ML outcomes. Besides, molecular docking and MD simulations were executed to clarify the physical insights into the binding process of these compounds to neuraminidase.

MATERIALS AND METHODS

Data Set. A set of 1154 compounds with SMILES and their corresponding association constants K_i was collected from BindingDB. The binding free energy was computed from K_i as $\Delta G = RT \ln K_i$, where R is the molar gas constant and $T = 298$ K is the absolute temperature. The experimental binding free energy ΔG was employed as a label for training ML regression models. The set was randomly divided into a train set consisting of 989 compounds and a test set consisting of 165 compounds. The train set was used to train ML models, and the test set was used for performance evaluation. The ΔG distribution of train and test sets is shown Figure S1 in the Supporting Information. The experimental binding free energy in the training data ranges in a wide range from strongly active (−14.2 kcal/mol) to essentially inactive (5.1 kcal/mol) compounds. About 40% of compounds in the training data have a binding free energy of −7 kcal/mol (equivalent to IC50 of about 10 μM) or weaker. The best ML model was chosen to perform prediction of ligand-binding free energy to neuraminidase for the CHEMBL database,[34] consisting of nearly 2 million compounds. Compounds that have already been in the train and test sets were excluded from the CHEMBL set. The top 100 compounds having strongest binding affinity were selected for further investigations using molecular dynamics (MD) simulations, enzymatic activity, and cytotoxicity assays.

Training ML Models. We trained four regression models including linear regression (LR), RF, extreme gradient boosting (XGBoost),[35] and convolutional networks on graphs (GraphConv).[36] LR served as a baseline model due to its simplicity and it being less prone to overfitting. RF and XGBoost are both ensemble methods. They differ in that in RF, regression tree learners are fit independently based on bootstrapping and random subspace of the train sample, while in XGBoost, the learners are instructed sequentially, with each student attempting to rectify the errors made by their predecessors. Furthermore, in RF, prediction is made by averaging over the predictions of all the trees in the ensemble, while in XGBoost, the weighted sum of predictions from all learners is used as a final prediction. The features for LR, RF, and XGBoost are physicochemical descriptors computed using the RDKitDescriptors toolkit included within DeepChem.[37] RDKitDescriptors calculated 200 physicochemical descriptors that were finally reduced to 104 features after removing the ones having a mostly zero value and highly correlated features. In the case of LR and RF, the median was used for imputing missing values. Conversely, XGBoost does not require imputation as it is capable of managing missing values automatically. For LR, the features were standardized to have a zero mean and a standard deviation of one. The deep learning method GraphConv can learn features on the fly and, therefore, does not require manual feature extraction. Input into the model is a molecular graph, which is passed to convolutional layers. The convolutional layers will acquire a fixed-length embedding vector that is subsequently fed into a densely connected layer.

The hyperparameters of LR, RF, and XGBoost were optimized by minimizing the mean square error (MSE) calculated from the training set through the application of the 10-fold cross-validation technique. The Hyperopt library[38] was utilized to identify the optimal set of hyperparameters. In our exploration of the GraphConv model, we experimented with various configurations, including the number of units in the graph_cov layers and dense layers as well as adjustments to the learning rate and dropout rates. The performance seemed to be more sensitive to the network size than to the learning and dropout rates. The Python library Scikit-Learn[39] was utilized to train LR and RF models, while the XGBoost library was employed for XGBoost models. The GraphConv model was trained using the DeepChem[37] library. Setting information on ML models is provided in the Supporting Information.

Molecular Docking. AutoDock Vina[40] was used to dock CHEMBL ligands into the binding pocket of neuraminidase, whose 3D structure in complex with zanamivir was obtained from the protein data bank with PDB ID 4B7Q,[41] representing the neuraminidase of the H1N1 2009 influenza virus with a high resolution of 2.73 Å. The docking empirical parameters were modified to improve docking accuracy according to our previous study.[42] The force field parameters from AutoDockTools were used to prepare the protein and ligands for docking. The chemicalize Web server, a tool of ChemAxon, was utilized to predict the ligand protonation states.[43] The center of the docking grid was chosen as the center of mass of zanamivir and the size of the grid was 24 × 24 × 24 Å^3. The docking poses with the lowest docking energy were selected for subsequent MD simulations.

Molecular Dynamics Simulations. MD simulations were performed to sample the conformational space of complexes between neuraminidase and ligands in aqueous solution. The Amber99SB-iLDN force field[44] was employed to parametrize interatomic interactions of the protein and counterions. For water molecules, the TIP3P water model[45] was used. The general Amber force field[46] was employed for Lennard-Jones and bonded interactions for the ligand. AmberTools18[47] and ACPYPE packages[48] were applied to fit the point charges of the ligand using the restrained electrostatic potential method.[46] The fitting procedure required as an input the electrostatic potential grid, which was calculated by density functional theory based on the double hybrid functional Mp2, basis set 6-31G(d,p), and implicit solvent (ε = 78.4). The neuraminidase−ligand complexes were inserted into a water box with dodecahedral periodic boundary conditions. The box size was chosen such that there was a minimum distance of 16.0 Å between the protein−ligand complex and the box edge. The box had a volume of 569.75 nm^3 and contained a total of 56,000 atoms. The water box for MD simulations of free ligands had a volume of 56.16 nm^3 and a total number of atoms of 5500.

Energy minimization with the steepest descent algorithm was first performed to remove steric clashes and drive the conformation to a local minimum. Then, short MD simulations of 100 ps under *NVT* and *NPT* conditions were performed to equilibrate the system. During this equilibration step, Cα atoms were restrained by applying a weak harmonic restraining potential. Finally, 50 ns MD simulations were performed to generate data for structural and energetic analyses. To improve statistical sampling, MD simulation for

https://doi.org/10.1021/acsomega.4c07194
ACS Omega 2024, 9, 48505−48511

the complex was repeated 2 times using different random number seeds. We used GROMACS version 2019.6[49] to perform MD simulations.

Neuraminidase Enzymatic Activity Assays. The NA-Fluor Influenza Neuraminidase Assay Kit (Applied Biosystems, Foster City, CA) using MUNANA as the substrate was utilized to assess the effectiveness of compounds.[50] The virus stock was titrated using the NA activity assay, and the optimal virus dilution (1:64 dilution) for the neuraminidase inhibition assay was selected. Compounds were tested for NA inhibitory activity at 100 μM. Fluorescence was measured by an ELISA Reader (Molecular Devices; Lmax II384) at an emission of 460 nm and an excitation of 360 nm. A zanamivir control (0.02 μM, MedChemExpress, Monmouth Junction, NJ, USA) was included for comparison.

Analysis Tools. The statistical errors of correlation coefficient and root-mean-square error (RMSE) were estimated using 1000 rounds of the bootstrapping method.[51] The intermolecular side chain (SC) contact between the ligand and the residual neuraminidase was counted when the spacing between non-hydrogen atoms of them was $\leq$4.5 Å. The intermolecular hydrogen bond (HB) between the residual neuraminidase and ligand was counted when the angle $\angle$ of the acceptor−hydrogen−donor was $\geq$135° and the distance between the acceptor and donor was $\leq$3.5 Å.

The experimental data are depicted as mean values accompanied by their standard error of the mean (SEM, derived from no fewer than two independent experiments. Statistical analyses were conducted using a two-tailed, unpaired Student's t test with Prism software (version 8.0, GraphPad Software, San Diego, CA, USA). Significance levels were set at $p < 0.05$ (*), $p < 0.01$ (**), and $p < 0.001$ (***), with "ns" denoting nonsignificant differences between the indicated settings.

RESULTS AND DISCUSSION

ML Calculations. The predictive performance of trained ML models was assessed by using three performance metrics, namely, RMSE, Pearson's R, and Spearman's ρ correlation coefficients. Table 1 shows a performance comparison of the four ML models for the test set consisting of 165 compounds. The baseline model LR gave the poorest performance with a large RMSE (RMSE = 2.80 ± 0.32 kcal/mol) and a low correlation (Pearson's R = 0.46 ± 0.07; Spearman's ρ = 0.58 ± 0.05) with respect to the experiment. This is not unexpected since LR is a linear model, which is unable to capture nonlinear relationships between the input features and the target label of binding free energy. Using more powerful nonlinear methods such as RF, XGBoost, and GraphConv significantly improves the predictive performance, as shown in Table 1. The GraphConv model gave the best performance with the lowest test RMSE (RMSE = 1.86 ± 0.22) and highest test correlation (Pearson's R = 0.80 ± 0.04, Spearman's ρ = 0.84 ± 0.03). Figure 1 shows a comparison between predicted and experimental binding free energies. However, GraphConv's performance does not differ significantly from that of the XGBoost model, which is the second best. Assuming that compounds having a binding free energy of −7 kcal/mol or lower are binder and those with a higher value than that are nonbinder, we can calculate the confusion matrix, which is shown in Table S1 in the Supporting Information. The accuracy in predicting binder vs nonbinder is about 82%.

Table 1. Comparison of ML Models' Performance in Predicting Binding Free Energy ΔG for 165 Test Compounds to Neuraminidase

model	RMSE (kcal/mol)	Pearson's R	Spearman's ρ
LR	2.80 ± 0.32	0.46 ± 0.07	0.58 ± 0.05
RF	2.03 ± 0.23	0.76 ± 0.03	0.83 ± 0.03
XGBoost	1.94 ± 0.22	0.78 ± 0.03	0.83 ± 0.03
graphConv	1.86 ± 0.22	0.80 ± 0.04	0.84 ± 0.03

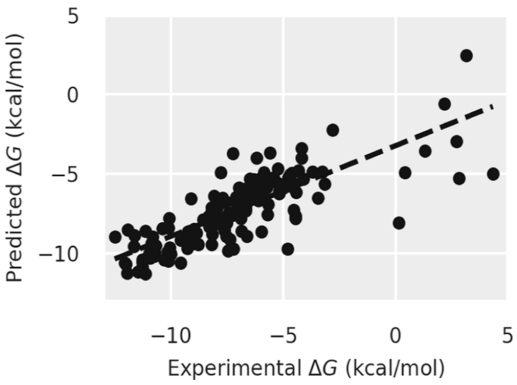


Figure 1. Comparative analysis of binding free energy derived from experimental data versus predictions generated by the GraphConv model for a set of 165 test compounds.

The GraphConv model was chosen to make predictions of binding free energies for nearly 2 million compounds in the CHEMBL data set. The distribution of predicted binding free energies is shown in Figure S2 in the Supporting Information. The mean and standard deviation of the distribution are −6.84 and 0.97 kcal/mol, respectively. A short list of 400 compounds having the strongest binding free energy to neuraminidase (ranging from −12.4 to −10.10 kcal/mol) were chosen for further investigations. Among these 400 compounds, 184 of them were previously tested and therefore removed from the list. These 184 compounds are mostly active with the experimental binding free energy ranging from −14.9 to −7.0 kcal/mol, and 95% of them have an experimental binding free energy more negative than −8.5 kcal/mol. It should be noted that these compounds were not included in our training data, and the ML predicted binding free energies for them are highly accurate with RMSE = 1.57 kcal/mol. Then, 11 compounds (see Table S2 in the Supporting Information for the list of compounds) were randomly selected for experimental investigations to assess their enzymatic inhibitory activity. Moreover, we also performed molecular docking and MD simulations to study their binding conformation to the enzyme.

Enzymatic Assays. The inhibition of neuraminidase activity was then evaluated with the Influenza Neuraminidase Assay Kit, using zanamivir as a positive control (Figure 2). Two of the 11 compounds (compounds 1 and 7) exhibited potential inhibition of NA activity. However, four compounds (compounds 4−6 and 11) yielded higher values of readings than the virus-only control. We suspect that these compounds may exhibit unexpected stimulatory effects on the neuraminidase enzyme, either directly or indirectly, resulting in higher enzymatic activity levels than that in the virus-alone condition. However, we cannot exclude the autofluorescence exhibited by

https://doi.org/10.1021/acsomega.4c07194
ACS Omega 2024, 9, 48505−48511

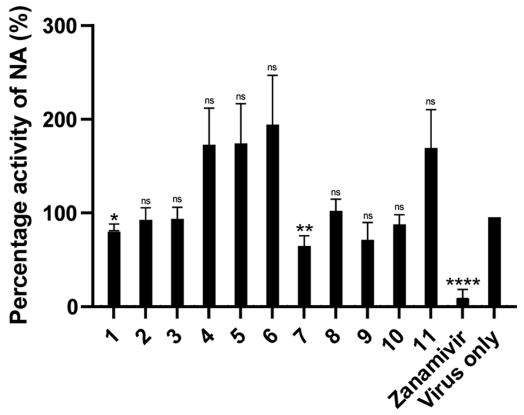


Figure 2. Validation of the compound by NA enzymatic activity assay. Eleven compounds were evaluated at a concentration of 100 μM, with zanamivir as the control at 0.02 μM. 1:64 dilution of virus was selected for the assay. Data are normalized to the virus-only group, which is arbitrarily set to 100%. The graph summarizes n = 3 independent experiments. Error bars show mean ± SEM (unpaired two-tailed t test). *p < 0.1, **p < 0.01, ****p < 0.0001, and ns = not significant. Comparisons between the virus-only group and compound 1, compound 7, and zanamivir yielded p-values of 0.0146, 0.0079, and <0.0001, respectively.

these compounds. Moreover, compounds could potentially stabilize the neuraminidase enzyme, prolonging its activity and enhancing the signal generated by substrate cleavage. Additionally, these compounds might inadvertently interact with assay components or viral particles in a manner that artificially amplifies the fluorescence signal generated by the cleavage of the MUNANA substrate.

Structural Insights from Molecular Docking and MD Simulations. In order to gain physical insight into the binding process of the top CHEMBL compounds to neuraminidase, 11 selected compounds were docked into the binding site of neuraminidase. Table S2 in the Supporting Information shows docking energy, which ranges from −12 to −9 kcal/mol. Docking poses of 11 compounds binding to neuraminidase are shown in Table S3 and the Supporting Information. For most of these poses, at least two HBs were formed between the ligand and receptor's residues.

A well-known limitation of docking methods is that they ignore the dynamics and treat the protein conformation essentially as rigid. Therefore, in the next step, we performed MD simulations to refine the docking structure of CHEMBL1430043 and neuraminidase. Figure S3 (in the Supporting Information) shows the RMSD of the two independent MD trajectories where the complex conformation is stabilized after about 100 ns.

Figure 3 shows the binding pose of CHEMBL1430043 taken as a representative structure in MD trajectory 1. In this binding pose, the compound makes several HB contacts with neuraminidase's residues in the binding pocket. A very similar binding pose was also observed during another independent MD simulation (see Figure S5 in the Supporting Information), which indicates that the structure was well equilibrated.

To study the nature of the interactions between CHEMBL1430043 and neuraminidase, we calculated the probability of the compound making HBs and hydrophobic contacts with the protein, and the result is shown in Figure 4. The important residues that make significant contacts with the compound include Arg152, Ser180, Glu228, Ser247, Glu277,

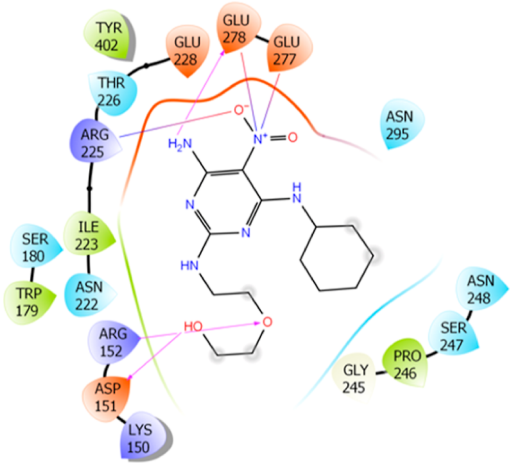


Figure 3. Binding pose of CHEMBL1430043 to neuraminidase in the representative structure using the clustering method with a non-hydrogen atom RMSD cutoff of 0.2 nm over an interval of 100−200 ns of MD trajectory 1.

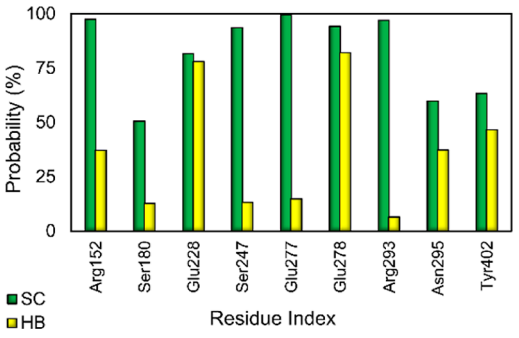


Figure 4. Probability of forming SC and HB contacts between CHEMBL1430043 and neuraminidase of essential residues. The calculations were performed over the time interval of 100−200 ns during the MD simulation. See Figure S6 in the Supporting Information for a plot of the probability of forming SC and HB contacts across all residues.

Glu278, Agr293, Asn295, and Tyr402 and are expected to determine the binding process.

Secondary Structure Dynamics of Complexes. To further gain a comprehensive understanding of the changes induced by CHEMBL1430043 within the complex, the secondary structure was also investigated via the define secondary structure of proteins protocol (Figure 5). In detail, the β-content had average values of 46.5 ± 0.9% and 46.4 ± 0.9%. for trajectory 1 and trajectory 2, respectively. The α-content of both trajectories fluctuates between 0% and 5% with an average of 3.3 ± 0.6% and 2.3 ± 0.7% for trajectory 1 and trajectory 2, respectively. The average values of coil-content and turn-content of trajectory 1 were 39.7 ± 1.1% and 10.5 ± 0.7% respectively. These values remained relatively consistent when compared to those of trajectory 2, which were 40.1 ± 1.0% and 11.2 ± 0.9%, respectively.

Free Energy Landscapes. Free energy landscape (FEL) of the CHEMBL1430043−neuraminidase complex over the equilibrium region 100−200 ns of MD simulations was obtained using the GROMACS tool "gmx sham".[52,53] The FEL was obtained and is shown in Figure 6. A minima structure was found, which has coordinates (PC1, PC2) at (−1.4, −0.25), of which the ligand-binding pose to

https://doi.org/10.1021/acsomega.4c07194
ACS Omega 2024, 9, 48505−48511

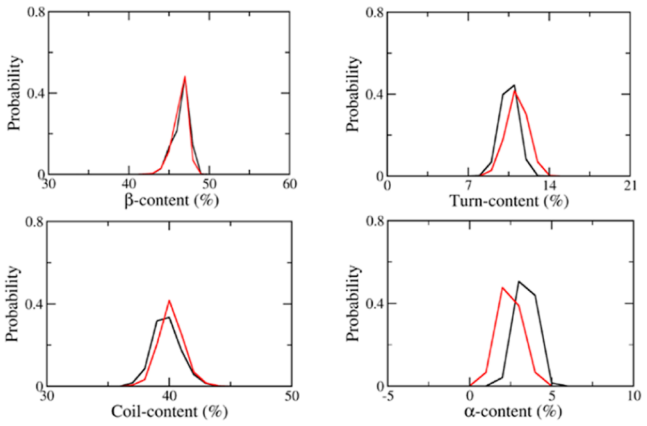


Figure 5. Secondary structure of the neuraminidase in complex with the CHEMBL1430043 inhibitor.

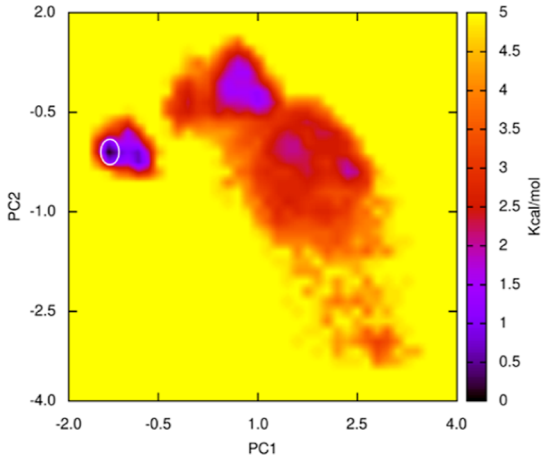


Figure 6. FEL of the simulated neuraminidase-CHEMBL1430043 systems based on the principal component analysis. In particular, the first and second principal components were utilized as two reaction coordinates. The systems were analyzed over an interval of 100−200 ns of equilibrium of both simulated MD trajectories.

neuraminidase is described in Figure 7. The population distribution of this minima is 15%.

CONCLUSIONS

We employed ML approaches to virtually screen the large CHEMBL compound database consisting of nearly 2 million compounds. The GraphConv model shows the best performance on the test set with a Pearson's R correlation of 0.8 and an RMSE of 1.86 kcal/mol. Molecular docking and MD simulations were employed to understand structural insights into the binding process between the top compounds and neuraminidase. The MD simulations shed light into SC and HB contacts between the top compound and neuraminidase and indicate important residues that stabilize the protein−ligand interaction. Experimental investigations on the enzymatic inhibition of 11 compounds randomly selected from the top 400 compounds indicated that compounds 1 and 7 might be able to inhibit neuraminidase. This shows that virtual screening helps significantly increase the hit rate, which is normally less than 1% for conventional high-throughput screening methods. Such findings underscore the immense value of virtual screening in expediting the identification of potential drug candidates with IAV-inhibitory properties.

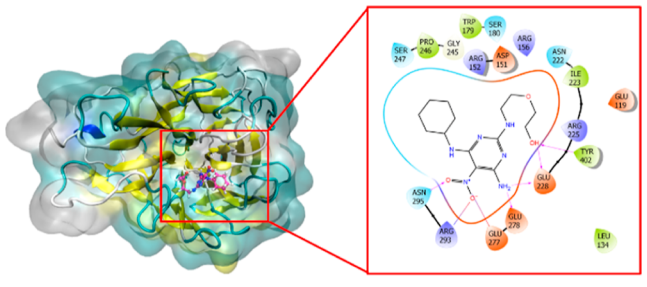


Figure 7. Representative structures of the neuraminidase + CHEMBL1430043, which are located in the corresponding minima of the FEL.

ASSOCIATED CONTENT

Data Availability Statement

All relevant data necessary to reproduce all results in the paper are within the main text, Supporting Information file, and the GitHub repository (https://github.com/nguyentrunghai/Neuraminidase/releases/tag/v1.0). Python code for training ML models, training data set, MD input files, parameter files, and topology files can be found in this Github repository.

Supporting Information

The Supporting Information is available free of charge at https://pubs.acs.org/doi/10.1021/acsomega.4c07194.

Setting information on ML models; confusion matrix for the test compounds; ML predicted binding free energy and docking energy for 11 compounds; docked binding pose of inhibitors with residues in N1 protein; distribution of experimental binding free energies for the train and test sets; distribution of predicted binding free energies for ChEMBL compounds; RMSD and RMSF during MD simulations for the neuraminidase in complex with CHEMBL1430043; radius of gyration and HBs during MD simulations for the neuraminidase in complex with CHEMBL1430043; binding pose of CHEMBL1430043 to neuraminidase; and the probability of forming SC and HB contacts between CHEMBL1430043 and neuraminidase (PDF)

The training and testing data (XLSX)

AUTHOR INFORMATION

Corresponding Authors

Son Tung Ngo − *Laboratory of Biophysics, Institute for Advanced Study in Technology, Ton Duc Thang University, Ho Chi Minh City 72915, Vietnam; Faculty of Pharmacy, Ton Duc Thang University, Ho Chi Minh City 72915, Vietnam;* orcid.org/0000-0003-1034-1768; Email: ngosontung@tdtu.edu.vn

Jim-Tong Horng − *Department of Biochemistry and Molecular Biology, College of Medicine, Molecular Infectious Disease Research Center, Chang Gung Memorial Hospital, and Research Center for Emerging Viral Infections, College of Medicine, Chang Gung University, Taoyuan 333, Taiwan;* Present Address: Research Center for Industry of Human Ecology and Research Center for Chinese Herbal Medicine, Graduate Institute of Health Industry Technology, Chang Gung University of Science and Technology, Taoyuan 333, Taiwan; Email: jimtong@cgu.edu.tw

https://doi.org/10.1021/acsomega.4c07194
ACS Omega 2024, 9, 48505−48511

Authors

Trung Hai Nguyen − *Laboratory of Biophysics, Institute for Advanced Study in Technology, Ton Duc Thang University, Ho Chi Minh City 72915, Vietnam; Faculty of Pharmacy, Ton Duc Thang University, Ho Chi Minh City 72915, Vietnam;* orcid.org/0000-0003-1848-3963

Ngoc Quynh Anh Pham − *Department of Biochemistry and Molecular Biology, College of Medicine, Chang Gung University, Taoyuan 333, Taiwan*

Quynh Mai Thai − *Laboratory of Biophysics, Institute for Advanced Study in Technology, Ton Duc Thang University, Ho Chi Minh City 72915, Vietnam; Faculty of Pharmacy, Ton Duc Thang University, Ho Chi Minh City 72915, Vietnam;* orcid.org/0000-0003-2149-5690

Van V. Vu − *NTT Hi-Tech Institute, Nguyen Tat Thanh University, Ho Chi Minh City 72820, Vietnam*

Complete contact information is available at: https://pubs.acs.org/10.1021/acsomega.4c07194

Author Contributions

[¶]T.H.N. (nguyentrunghai@tdtu.edu.vn) and N.Q.A.P. contributed equally to this work. All authors designed the studies, collected and analyzed the data, and wrote the manuscript. THN trained and tested ML models. THN predicted potential inhibitors for NA via ML calculations. NQAP collected the database of available inhibitors vs NA from bindingdb.org. QMT performed molecular docking and MD simulations. VVV helped in discussing the results. STN and JTH provided concept, supervise, writing, editing, etc.

Notes

The authors declare no competing financial interest.

■ ACKNOWLEDGMENTS

This work was supported by the Ho Chi Minh City Foundation for Science and Technology Development under grant number 115/QĐ-SKHCN. This research was made possible through the generous financial support of Chang Gung Memorial Hospital, Taoyuan, Taiwan, under the grants BMRP416 and CMRPD1M0881-3. We would like to express our appreciation for the additional support received from the National Science and Technology Council, which granted support under the reference numbers 113-2321-B-182-003, 112-2321-B-182-003, 112-2320-B-182-034-MY3, and 111-2321-B-182-001. Furthermore, our research benefited from funding from the Research Center for Emerging Viral Infections within the Featured Areas Research Center Program, part of the Higher Education Sprout Project supported by the Ministry of Education and the National Science and Technology Council, under the grants 110-2634-F-182-001 and 111-2634-F-182-001.

■ REFERENCES

(1) Palese, P. Influenza: old and new threats. *Nat. Med.* **2004**, *10* (S12), S82−S87.

(2) Yen, H.-L.; Webster, R. Pandemic Influenza as a Current Threat. In *In Vaccines for Pandemic Influenza*; Compans, R. W., Orenstein, W. A., Eds.; Springer Berlin Heidelberg, 2009; Vol. *333*, pp 3−24.

(3) Matsuzaki, Y.; Sugawara, K.; Takashita, E.; Muraki, Y.; Hongo, S.; Katsushima, N.; Mizuta, K.; Nishimura, H. Genetic diversity of influenza B virus: The frequent reassortment and cocirculation of the genetically distinct reassortant viruses in a community. *J. Med. Virol.* **2004**, *74* (1), 132−140.

(4) Hsieh, Y. C.; Wu, T. Z.; Liu, D. P.; Shao, P. L.; Chang, L. Y.; Lu, C. Y.; Lee, C. Y.; Huang, F. Y.; Huang, L. M. Influenza pandemics: past, present and future. *J. Formos. Med. Assoc.* **2006**, *105* (1), 1−6.

(5) Ferguson, N. M.; Fraser, C.; Donnelly, C. A.; Ghani, A. C.; Anderson, R. M. Public Health Risk from the Avian H5N1 Influenza Epidemic. *Science* **2004**, *304* (5673), 968−969.

(6) Patel, M; Dennis, A; Flutter, C; Khan, Z Pandemic (H1N1) 2009 influenza. *Br. J. Anaesth.* **2010**, *104*, 128−142.

(7) Neumann, G.; Noda, T.; Kawaoka, Y. Emergence and pandemic potential of swine-origin H1N1 influenza virus. *Nature* **2009**, *459* (7249), 931−939.

(8) Li, M.; Liu, H.; Bi, Y.; Sun, J.; Wong, G.; Liu, D.; Li, L.; Liu, J.; Chen, Q.; Wang, H.; He, Y.; Shi, W.; Gao, G. F.; Chen, J. Highly Pathogenic Avian Influenza A(H5N8) Virus in Wild Migratory Birds, Qinghai Lake, China. *Emerg. Infect. Dis.* **2017**, *23* (4), 637−641.

(9) Wu, Y.; Bi, Y.; Vavricka, C. J.; Sun, X.; Zhang, Y.; Gao, F.; Zhao, M.; Xiao, H.; Qin, C.; He, J.; Liu, W.; Yan, J.; Qi, J.; Gao, G. F. Characterization of two distinct neuraminidases from avian-origin human-infecting H7N9 influenza viruses. *Cell Res.* **2013**, *23* (12), 1347−1355.

(10) Perrier, A.; Eluard, M.; Petitjean, M.; Vanet, A. In Silico Design of New Inhibitors Against Hemagglutinin of Influenza. *J. Phys. Chem. B* **2019**, *123* (3), 582−592.

(11) Choi, W.-S.; Jeong, J. H.; Kwon, J. J.; Ahn, S. J.; Lloren, K. K. S.; Kwon, H.-I.; Chae, H. B.; Hwang, J.; Kim, M. H.; Kim, C.-J.; Webby, R. J.; Govorkova, E. A.; Choi, Y. K.; Baek, Y. H.; Song, M.-S. Screening for Neuraminidase Inhibitor Resistance Markers among Avian Influenza Viruses of the N4, N5, N6, and N8 Neuraminidase Subtypes. *J. Virol.* **2018**, *92* (1), 10.

(12) Albohy, A.; Zhang, Y.; Smutova, V.; Pshezhetsky, A. V.; Cairo, C. W. Identification of Selective Nanomolar Inhibitors of the Human Neuraminidase, NEU4. *ACS Med. Chem. Lett.* **2013**, *4* (6), 532−537.

(13) Zhang, G.-Q.; Chang, H.; Gao, Z.; Deng, Y.-p.; Zeng, S.; Shang, L.; Ding, D.; Liu, Q. Neuraminidase-Activatable NIR Fluorescent Probe for Influenza Virus Ratiometric Imaging in Living Cells and Colorimetric Detection on Cotton Swabs. *ACS Mater. Lett.* **2023**, *5* (3), 722−729.

(14) Tam, N. M.; Nguyen, M. T.; Ngo, S. T. Evaluation of the Absolute Affinity of Neuraminidase Inhibitor using Steered Molecular Dynamics Simulations. *J. Mol. Graph. Model.* **2017**, *77*, 137−142.

(15) Nagao, M.; Yamaguchi, A.; Matsubara, T.; Hoshino, Y.; Sato, T.; Miura, Y. De Novo Design of Star-Shaped Glycoligands with Synthetic Polymer Structures toward an Influenza Hemagglutinin Inhibitor. *Biomacromolecules* **2022**, *23* (3), 1232−1241.

(16) Waldmann, M.; Jirmann, R.; Hoelscher, K.; Wienke, M.; Niemeyer, F. C.; Rehders, D.; Meyer, B. A Nanomolar Multivalent Ligand as Entry Inhibitor of the Hemagglutinin of Avian Influenza. *J. Am. Chem. Soc.* **2014**, *136* (2), 783−788.

(17) Das, K.; Aramini, J. M.; Ma, L.-C.; Krug, R. M.; Arnold, E. Structures of influenza A proteins and insights into antiviral drug targets. *Nat. Struct. Mol. Biol.* **2010**, *17* (5), 530−538.

(18) Stohr, K. Preventing and treating influenza. *Br. Med. J.* **2003**, *326* (7401), 1223−1224.

(19) Hurt, A. C. The epidemiology and spread of drug resistant human influenza viruses. *Curr. Opin. Virol.* **2014**, *8*, 22−29.

(20) Bloom, J. D.; Gong, L. I.; Baltimore, D. Permissive Secondary Mutations Enable the Evolution of Influenza Oseltamivir Resistance. *Science* **2010**, *328* (5983), 1272−1275.

(21) Cass, L. M. R.; Efthymiopoulos, C.; Bye, A. Pharmacokinetics of Zanamivir After Intravenous, Oral, Inhaled or Intranasal Administration to Healthy Volunteers. *Clin. Pharmacokinet.* **1999**, *36* (Supplement 1), 1−11.

(22) Van Drie, J. H. Computer-aided drug design: the next 20 years. *J. Comput. Aided Mol. Des.* **2007**, *21* (10−11), 591−601.

(23) Yu, W.; MacKerell, A. D. Computer-Aided Drug Design Methods. In *Antibiotics: Methods and Protocols*; Sass, P., Ed.; Springer: New York: New York, NY, 2017; pp 85−106.

(24) Sliwoski, G.; Kothiwale, S.; Meiler, J.; Lowe, E. W. Computational Methods in Drug Discovery. *Pharmacol. Rev.* **2014**, *66* (1), 334−395.
(25) Marshall, G. R. Computer-Aided Drug Design. *Annu. Rev. Pharmacol. Toxicol.* **1987**, *27*, 193−213.
(26) Ngo, S. T.; Minh, N. H.; Thuy, H. L. T.; Minh, Q. P.; Vi Khanh, T.; Nguyen Thanh, T.; Van, V. Assessing Potential Inhibitors for SARS-CoV-2 Main Protease from Available Drugs using Free Energy Perturbation Simulations. *RSC Adv.* **2020**, *10*, 40284−40290.
(27) Ngo, S. T.; Fang, S.-T.; Huang, S.-H.; Chou, C.-L.; Huy, P. D. Q.; Li, M. S.; Chen, Y.-C. Anti-Arrhythmic Medication Propafenone a Potential Drug for Alzheimer's Disease Inhibiting Aggregation of Aβ: In Silico and In Vitro Studies. *J. Chem. Inf. Model.* **2016**, *56* (7), 1344−1356.
(28) Tam, N. M.; Pham, M. Q.; Ha, N. X.; Nam, P. C.; Phung, H. T. T. Computational Estimation of Potential Inhibitors from Known Drugs Against the Main Protease of SARS-CoV-2. *RSC Adv.* **2021**, *11* (28), 17478−17486.
(29) Thai, Q. M.; Phung, H. T. T.; Pham, N. Q. A.; Horng, J.-T.; Tran, P.-T.; Tung, N. T.; Ngo, S. T. Natural compounds inhibit Monkeypox virus methyltransferase VP39 in silico studies. *J. Biomol. Struct. Dyn.* **2024**, 1−9.
(30) Vijayakrishnan, R. Structure-based drug design and modern medicine. *J. Postgrad. Med.* **2009**, *55* (4), 301−304.
(31) Li, Y.; Kong, Y.; Zhang, M.; Yan, A.; Liu, Z. Using Support Vector Machine (SVM) for Classification of Selectivity of H1N1 Neuraminidase Inhibitors. *Mol. Inf.* **2016**, *35* (3−4), 116−124.
(32) Zhang, L.; Ai, H.; Zhao, Q.; Zhu, J.; Chen, W.; Wu, X.; Huang, L.; Yin, Z.; Zhao, J.; Liu, H. Computational Prediction of Influenza Neuraminidase Inhibitors Using Machine Learning Algorithms and Recursive Feature Elimination Method. In *Bioinformatics Research and Applications*; Cai, Z., Daescu, O., Li, M., Eds.; Springer International Publishing: Cham, 2017; pp 344−349.
(33) Lian, W.; Fang, J.; Li, C.; Pang, X.; Liu, A.-L.; Du, G.-H. Discovery of Influenza A virus neuraminidase inhibitors using support vector machine and Naïve Bayesian models. *Mol. Diversity* **2016**, *20* (2), 439−451.
(34) Mendez, D.; Gaulton, A.; Bento, A. P.; Chambers, J.; De Veij, M.; Félix, E.; Magariños, M.; Mosquera, J. F.; Mutowo, P.; Nowotka, M.; Gordillo-Marañón, M.; Hunter, F.; Junco, L.; Mugumbate, G.; Rodriguez-Lopez, M.; Atkinson, F.; Bosc, N.; Radoux, C. J.; Segura-Cabrera, A.; Hersey, A.; Leach, A. R. ChEMBL: towards direct deposition of bioassay data. *Nucleic Acids Res.* **2018**, *47* (D1), D930−D940.
(35) Chen, T.; Guestrin, C. XGBoost: A Scalable Tree Boosting System. *KDD '16: Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*; ACM Digital Library, 2016, pp 785−794.
(36) Duvenaud, D. K.; Maclaurin, D.; Iparraguirre, J.; Bombarell, R.; Hirzel, T.; Aspuru-Guzik, A.; Adams, R. P. Convolutional Networks on Graphs for Learning Molecular Fingerprints. In *Advances in Neural Information Processing Systems*; Cortes, C., Lawrence, N., Lee, D., Sugiyama, M., Garnett, R., Eds.; Curran Associates, Inc., 2015; Vol. *28*.
(37) Ramsundar, B.; Eastman, P.; Walters, P.; Pande, V.; Leswing, K.; Wu, Z. Deep Learning for the Life Sciences: Applying Deep Learning to Genomics, Microscopy. In *Drug Discovery, and More*; O'Reilly Media, 2019.
(38) Bergstra, J.; Yamins, D.; Cox, D. Making a Science of Model Search: Hyperparameter Optimization in Hundreds of Dimensions for Vision Architectures. *Proceedings of the 30th International Conference on Machine Learning*; PMLR, 2013; Vol. *28*, pp 115−123.
(39) Pedregosa, F.; Varoquaux, G.; Gramfort, A.; Michel, V.; Thirion, B.; Grisel, O.; Blondel, M.; Prettenhofer, P.; Weiss, R.; Dubourg, V.; Vanderplas, J.; Passos, A.; Cournapeau, D.; Brucher, M.; Perrot, M.; Duchesnay, E. Scikit-learn: Machine Learning in Python. *J. Mach. Learn. Res.* **2011**, *12*, 2825−2830.
(40) Trott, O.; Olson, A. J. AutoDock Vina: Improving the speed and accuracy of docking with a new scoring function, efficient optimization, and multithreading. *J. Comput. Chem.* **2010**, *31* (2), 455−461.
(41) van der Vries, E.; Collins, P. J.; Vachieri, S. G.; Xiong, X.; Liu, J.; Walker, P. A.; Haire, L. F.; Hay, A. J.; Schutten, M.; Osterhaus, A. D. M. E.; Martin, S. R.; Boucher, C. A. B.; Skehel, J. J.; Gamblin, S. J. H1N1 2009 Pandemic Influenza Virus: Resistance of the I223R Neuraminidase Mutant Explained by Kinetic and Structural Analysis. *PLoS Pathog.* **2012**, *8* (9), No. e1002914.
(42) Pham, T. N. H.; Nguyen, T. H.; Tam, N. M.; Vu, T.; Pham, N. T.; Huy, N. T.; Mai, B. K.; Tung, N. T.; Pham, M. Q.; Vu, V.; Ngo, S. T. Improving ligand-ranking of AutoDock Vina by changing the empirical parameters. *J. Comput. Chem.* **2022**, *43* (3), 160−169.
(43) Chemicalize was Used for Prediction of Chemical Properties. https://chemicalize.com/,developedbyChemAxon (accessed May 2023).
(44) Aliev, A. E.; Kulke, M.; Khaneja, H. S.; Chudasama, V.; Sheppard, T. D.; Lanigan, R. M. Motional Timescale Predictions by Molecular Dynamics Simulations: Case Study using Proline and Hydroxyproline Sidechain Dynamics. *Proteins: Struct., Funct., Bioinf.* **2014**, *82* (2), 195−215.
(45) Jorgensen, W. L.; Chandrasekhar, J.; Madura, J. D.; Impey, R. W.; Klein, M. L. Comparison of simple potential functions for simulating liquid water. *J. Chem. Phys.* **1983**, *79* (2), 926−935.
(46) Wang, J.; Wolf, R. M.; Caldwell, J. W.; Kollman, P. A.; Case, D. A. Development and testing of a general amber force field. *J. Comput. Chem.* **2004**, *25* (9), 1157−1174.
(47) Case, D.; Cerutti, D.; Cheatham, T.; Darden, T.; Duke, R.; Giese, T.; Gohlke, H.; Goetz, A.; Greene, D.; Homeyer, N. *Amber18*; University of San Francisco, 2017.
(48) Sousa da Silva, A. W.; Vranken, W. F. ACPYPE - AnteChamber PYthon Parser interfacE. *BMC Res. Notes* **2012**, *5* (1), 367.
(49) Abraham, M. J.; Murtola, T.; Schulz, R.; Páll, S.; Smith, J. C.; Hess, B.; Lindahl, E. GROMACS: High performance molecular simulations through multi-level parallelism from laptops to supercomputers. *SoftwareX* **2015**, *1−2*, 19−25.
(50) Sethy, B.; Hsieh, C.-F.; Lin, T.-J.; Hu, P.-Y.; Chen, Y.-L.; Lin, C.-Y.; Tseng, S.-N.; Horng, J.-T.; Hsieh, P.-W. Design, Synthesis, and Biological Evaluation of Itaconic Acid Derivatives as Potential Anti-Influenza Agents. *J. Med. Chem.* **2019**, *62* (5), 2390−2403.
(51) Efron, B. Bootstrap Methods: Another Look at the Jackknife. *Ann. Stat.* **1979**, *7*, 1−26.
(52) Mu, Y.; Nguyen, P. H.; Stock, G. Energy Landscape of a Small Peptide Revealed by Dihedral Angle Principal Component Analysis. *Proteins: Struct., Funct., Bioinf.* **2005**, *58* (1), 45−52.
(53) Papaleo, E.; Mereghetti, P.; Fantucci, P.; Grandori, R.; De Gioia, L. Free-Energy Landscape, Principal Component Analysis, and Structural Clustering to Identify Representative Conformations from Molecular Dynamics Simulations: The Myoglobin Case. *J. Mol. Graph. Model.* **2009**, *27* (8), 889−899.

https://doi.org/10.1021/acsomega.4c07194
ACS Omega 2024, 9, 48505−48511