Discovering Neuraminidase Inhibitors via Computational and Experimental Studies

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

[a]Laboratory of Biophysics, Institute for Advanced Study in Technology, Ton Duc Thang University, Ho Chi Minh City 72915, Vietnam

[b]Faculty of Pharmacy, Ton Duc Thang University, Ho Chi Minh City 72915, Vietnam

[c]Department of Biochemistry and Molecular Biology, College of Medicine, Chang Gung University, Kweishan, Taoyuan 333, Taiwan

[d]NTT Hi-Tech Institute, Nguyen Tat Thanh University, Ho Chi Minh City 72820, Vietnam

ABSTRACT: Influenza A viruses spread out worldwide causing several global concerns. Discovering neuraminidase inhibitors to prevent the influenza A virus is thus of great interests. 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 possibility of compounds from the CHEMBL database, which is 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 studies on enzymatic and antiviral activity as well as cytotoxicity have validated our computational results and suggested that 2 compounds were potential inhibitors of neuraminidase of the influenza A virus.

INTRODUCTION

Influenza A viruses have caused major influenza outbreak or pandemic that affected millions of people worldwide. 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.[1-2] and H5N1,[3-4], H1N1,[5-6] H5N8,[7] and H7N9[8] in recent years. New strains of the viruses continue to emerge and the risk of drug resistance have sparked great interest into finding potential anti-viral compounds.[9-15]

Neuraminidase is a key surface glycoprotein which plays an important role in viral replication and infection. It is a proven target for developing drugs against influenza A viruses[16-17]. Several drugs recommended for treating influenza virus such as oseltamivir, zanamivir, and peramivir are neuraminidase inhibitors. However, these drugs suffer from serious limitations such as the emergence of oseltamivir-resistant strains[18-19], the poor oral bioavailability of zanamivir [20]. On the other hand, more virulent variants, such as H5N1 and H7N9 has emerged. Therefore, researching novel inhibitors capable of effectively inhibiting neuraminidase continue to be a topic of interest.

Computer-aided drug design (CADD) plays as a powerful tool for rapidly and accurately screening of several million compounds for potential inhibitors of enzymes[21]. The adoption of CADD methods is rapidly increased due to their potential to significantly reduce the cost and time of a new drug development.[22] CADD can be used in both purposes including searching for new inhibitors and repurposing for existing drugs.[23-25] CADD has been donating to the discovery of severally available drugs such as *dorzolamide*,[26-27] *saquinavir*, *ritonavir*, and *indinavir*.[21]

In this work, we aim to use a combination of computational and experimental approaches to find potential inhibitors for inhibiting neuraminidase. In particular, the trained machine learning (ML) model was employed to predict the ligand-binding affinity of ca. 2 million compounds of ChEMBL database to neuraminidase. The experimental studies were then carried out to validate the ML outcomes. The shortlist of potential candidates were obtained. The experiment would be then validated the ML outcomes. Besides, molecular docking and MD simulations were used to clarify the physical insights into the binding process of these compounds to neuraminidase.

MATERIALS & 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 calculated from K_i as $\Delta G = RTlnK_i$, where R is the molar gas constant, T = 298 K is the absolute temperature. The experimental binding free energy ΔG was used 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 machine learning (ML) models and the test set was used for performance evaluation. The ΔG distribution of train and test sets is shown in **Figure 1**. The best ML model was selected to make prediction of binding free energy to Neuraminidase for the CHEMBL data set[28] consisting of nearly 2 million compounds. Compounds which 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 simulations, enzymatic activity, antiviral activity and cytotoxicity assays.

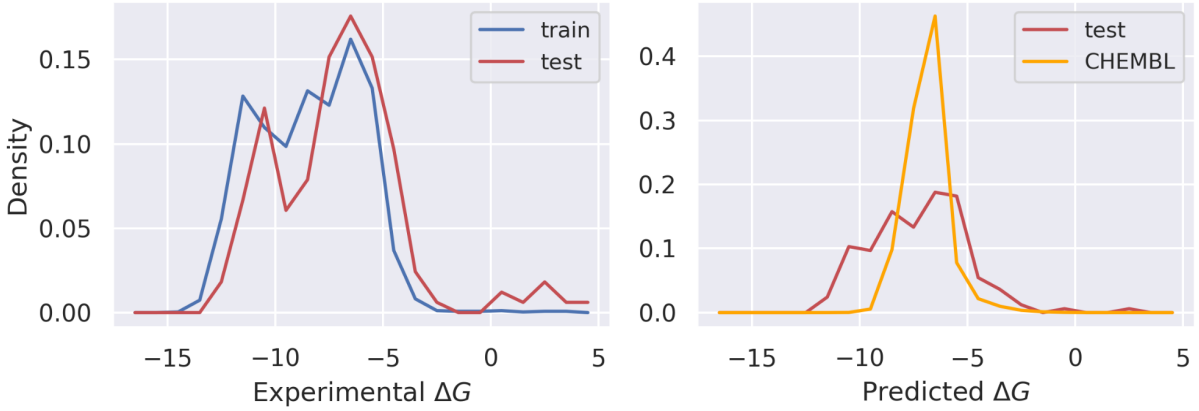

Figure 1. Distribution of experimental binding free energies for the train and test sets (left) and of the binding free energies predicted by GraphConv model for the test and CHEMBL sets (right).

We trained four regression models including linear regression (LR), random forest (RF), extreme gradient boosting (XGBoost)[29] and convolutional networks on graphs (GraphConv).[30] LR served as a baseline model due to its simplicity and 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 sequentially trained such that each learner tries to fix the mistake made by previous ones. Furthermore, in RF prediction is made by averaging over the predictions of all trees in the ensemble, while in XGBoost, weighted sum of predictions from all learners is used as a final prediction. Features for LR, RF and XGBoost are physicochemical descriptors which were calculated using the RDKitDescriptors tool kit implemented in DeepChem.[31] RDKitDescriptors calculated 200 physicochemical descriptors which were finally reduced to 104 features after removing the ones having mostly zero value and highly correlated features. For LR and RF, missing values were imputed with the median, while for XGBoost, imputation is not required because it can automatically handle missing values. 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 learn a fixed-length embedding vector called molecular fingerprint which is then input into a densely connected layer.

Hyperparameters of LR, RF and XGBoost were tuned by minimizing the mean square error (MSE) estimated from the train set using the 10-fold cross validation method. The Hyperot library[32] was used to search for the optimal set of hyperparameters. For the GraphConv model, we tried different numbers of units in the graph_cov layers and dense layers, learning rate, and dropout rates. The performance seemed to be more sensitive to the network size than the learning and dropout rates. We used the Python library Scikit-Learn[33] to train LR and RF models and the XGBoost library for XGBoost models. We used the library DeepChem[31] to train the GraphConv model.

Molecular Docking

AutoDock Vina[34] was used to dock CHEMBL ligands into binding pocket of neuraminidase whose 3D structure in complex with Zanamivir was obtained from the protein data bank with PDB ID 4B7Q. The docking empirical parameters were modified to improve docking accuracy according to our previous study[35]. The force field parameters from AutoDockTools were used to prepare the protein and ligands for docking. The chemicalize webserver, a tool of ChemAxon, was utilized to predict the ligand protonation states.[44] The center of docking grid was chosen as the center of mass of Zanamivir and the size of the grid as $24 \times 24 \times 24$ Å^3. The docking poses with lowest docking energy were selected for subsequent MD simulations.

Molecular Dynamics Simulations

MD simulations were performed to sample conformational space of complexes between neuraminidase and ligands in aqueous solution. Amber99SB-iLDN force field[36] was employed to parameterized inter-atomic interactions of the protein and counter ions. For water molecules, TIP3P water model[37] was used. The general Amber force field (GAFF)[38] was employed for Lennard Jones and bonded interactions for the ligand. AmberTools18[39] and ACPYPE packages[40] was applied to fit the point charges of the ligand using the restrained electrostatic potential (RESP) method[38]. The fitting procedure required as an input the

electrostatic potential grid which was calculated by DFT based on the double hybrid functional Mp2, basis set 6-31G(d,p), and implicit solvent ($\varepsilon = 78.4$). 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 in total 56000 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 the complex was repeated for 2 times using different random number seeds. We used the software GROMACS version 2019.6[41] to perform MD simulations.

Cell culture and viral amplification

MDCK cells were cultured in E10 medium comprising Dulbecco's modified Eagle medium (DMEM, Invitrogen, Carlsbad, CA) supplemented with 10% fetal bovine serum (FBS; JRH Biosciences, Lenexa, KS, USA), 100 U/mL penicillin, 100 μg/mL streptomycin (Sigma–Aldrich, St Louis, MO, USA), 2 mM L-glutamine (Gibco BRL, Gaithersburg, MD, USA), and 0.1 mM nonessential amino acid mixture (Gibco). The cells were maintained in a 37 °C incubator with 5% CO_2. The influenza virus A/WSN/33 (WSN) obtained from the American Type Culture Collection was propagated in MDCK cells. Cells were incubated with the influenza virus at a multiplicity of infection (MOI) of 0.5 for 1 h at room temperature and cultured for 24 or 48 h at 37 °C in fresh DMEM. Viral titers in supernatants were determined using MDCK cells by the 50% tissue culture infectious dose (TCID50) assay[42].

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[43]. 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 excitation of 360 nm. A zanamivir control (0.02 μM, MedChemExpress, Monmouth Junction, NJ, USA) was included for comparison.

Cell-based antiviral activity and cytotoxicity assays

The antiviral efficacy of compounds was assessed in MDCK cells through quantification of their ability to mitigate virus-induced cell death. MDCK cells were cultured in a 96-well plate (2×10^4 cells per well) in E10 medium and incubated at 37°C under 5% CO_2 overnight. The wells were then washed once with phosphate-buffered saline (PBS). Compounds were added at a concentration of 50 μM, along with 9 x TCID50 of the virus. The cells were subjected to individual compound incubation as a means to evaluate its cytotoxic effects. The surviving cells were measured with the colorimetric 3-(4,5-dimethylthiazol-2-yl)-2,5-diphenyl tetrazolium bromide (MTT) assay. The plate was placed in a 5% CO_2, 37°C incubator for 72 h. After incubation, the cell wells were rewashed with PBS, and 50 μL of MTT solution was added to each well, followed by incubation for 3 h at 37°C with 5% CO_2. The medium was removed, and 200 μL of dimethyl sulfoxide was added to each well to dissolve the formazan crystals. The absorbance of each well was measured at 570 nm on an ELISA Reader (Molecular Devices; Lmax II384). A control group of drug ITA19 with a concentration of 5 μM was employed[43].

Statistical analysis

The statistical errors of correlation coefficient and RMSE were estimated using 1000 rounds of the bootstrapping method.[45] The intermolecular sidechain contact (SC) between the ligand to the residual neuraminidase was counted when the spacing between non-hydrogen atoms of them is ≤4.5 Å. The intermolecular hydrogen bond (HB) between the residual neuraminidase and ligand was counted when the angle ∠ between acceptor-hydrogen-donor is ≥135° and the distance between acceptor and donor is ≤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 non-significant differences between the indicated settings.

RESULTS AND DISCUSSION

Machine Learning Calculations

The predictive performance of trained ML models was assessed by using three performance metrics, namely, root-mean-square error (RMSE), Pearson's R and Spearman's ρ correlation coefficients. Table 1 shows performance comparison of the four ML models for the test set consisting of 165 compounds. The baseline model LR gave the poorest performance with large RMSE (RMSE = 2.80 ± 0.32 kcal/mol) and low correlation (Pearson's R = 0.46 ± 0.07; Spearman's ρ = 0.58 ± 0.05) with respect to 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 2 shows comparison between predicted and experimental binding free energies. However, GraphConv's performace does not differ significantly from XGBoost model which is the second best.

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

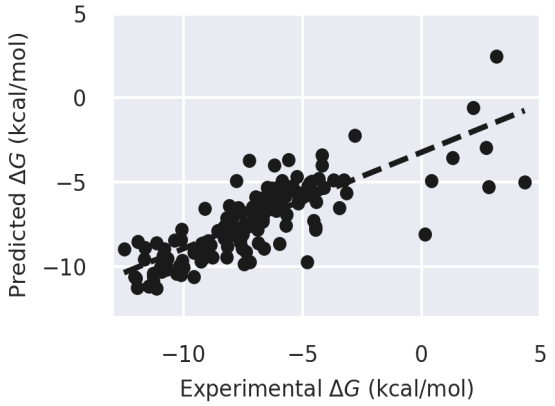

Figure 2. Comparison of binding free energy between experiment and prediction made by the GraphConv model for 165 test compounds.

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 Fig 1. The mean and standard deviation of the distribution are -6.84 kcal/mol and 0.97 kcal/mol, respectively. A short list of 400 compounds having strongest binding free energy to neuraminidase (ranging from -12.4 kcal/mol to -10.10 kcal/mol) was chosen for further investigations. Among these 400 compounds, 184 of them were previously tested and therefore, removed from the list. Then 11 compounds (see Table S1 in SI for the list of compounds) were randomly selected (see table S1 in SI) for experimental investigations to assess their enzymatic inhibitory and anti-viral activities and cytotoxicity. Moreover, we also performed molecular docking and MD simulations to study their binding conformation to the enzyme.

Enzymatic and antiviral activities and cytotoxicity assays

Our research, guided by virtual screening, revealed a notable breakthrough in the quest for effective influenza A virus (IAV) inhibitors. The inhibition of neuraminidase activity was then evaluated with the Influenza Neuraminidase Assay Kit, using zanamivir as a positive control (Figure 3). Two of the eleven compounds (compounds 1 and 7) selected through this innovative approach exhibited potential inhibition of NA activity. However, four compounds (compounds 4-6, and 11) yield 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 compared to the virus-alone condition. However, we can't exclude the autofluorescence exhibited by these compounds. Moreover, compounds could potentially stabilize the neuraminidase enzyme, prolonging its activity and enhancing the signal generated by the substrate cleavage. Additionally, these compounds might inadvertently interact with assay components or viral particles in a manner that artificially amplifies the fluorescent signal generated by the cleavage of the MUNANA substrate. This may also explain why compound 9 did not exhibit clear inhibition of NA activity but demonstrated the ability to inhibit viruses in a cell-based assay with about 55% protection (Figure 4). Compound 6 also showed a significant protection of about 30%. Nevertheless, it's important to consider that compounds 6 and 9, with no inhibition in the NA enzymatic activity, have off-target effects in the cell-based antiviral activity assay. The ineffectiveness of compounds 1 and 7 in the antiviral assay might be attributed to their membrane penetration ability.

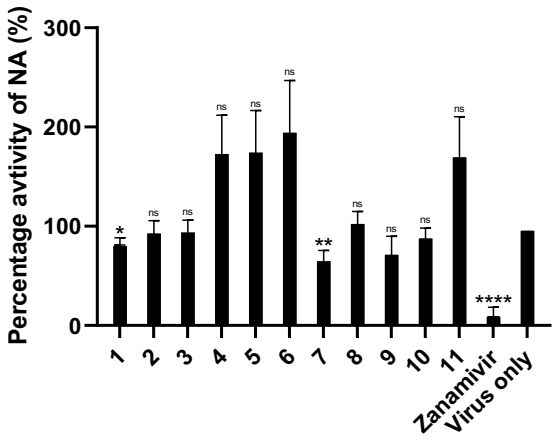

Figure 3. Validation of the compound by NA enzymatic activity assay. Eleven compounds were evaluated at a concentration of 100 µM, with Zanamivir as the control at 0.02 µM. 1:64 dilution of virus was selected for the assay. Data is normalized to the virus-only group, which is arbitrarily set to 100%. 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, 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.

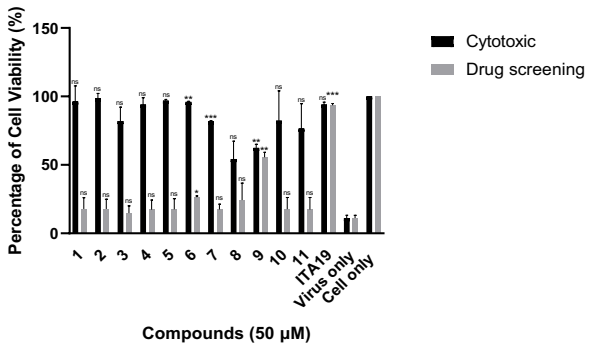

Figure 4. A cell-based antiviral assay for compounds screening. Eleven compounds were assessed at a concentration of 50 μM, with ITA19 used as the control at 5 μM. Data were normalized to the cell-only group, which was set at 100%. For cytotoxicity assessment, comparisons between the cell-only group and compounds 6, 7, and 9 yielded p-values of 0.0087, 0.0003, and 0.0047, respectively. For drug screening, comparisons between the virus-only group and compounds 6, 9, and ITA19 yielded p-values of 0.0199, 0.0084, and 0.0008, respectively. The graph represents n = 2 independent experiments. Error bars indicate mean ± SEM (unpaired two-tailed t-test). *p<0.1, **p < 0.01, ****p < 0.0001, ns = not significant.

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 S1 in SI shows docking energy which ranges from -12 to -9 kcal/mol. Docking poses of 11 compounds binding to neuraminidase are showed in Table S2 in SI. For most of these poses, at least two hydrogen bonds 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. Fig S1 (in SI) shows the RMSD of the two independent MD trajectories where the complex conformation is stabilized after about 25 ns.

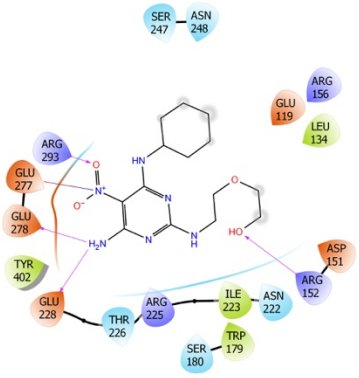

Figure 5. Binding pose of CHEMBL1430043 to neuraminidase in the representative structure of MD trajectory 1.

Fig. 5 shows biding pose of CHEMBL1430043 taken as a representative structure in MD trajectory 1. In this binding pose, the compound makes several hydrogen bond contacts with neuraminidase's residues in the binding pocket. A very similar binding pose was also observed during another independent MD simulation (see Fig S2 in SI) which indicates the structure was well equilibrated.

To study the nature of the interactions between CHEMBL1430043 and neuraminidase we calculated the probability of the compound making hydrogen bonds and hydrophobic contacts with the protein and the result is showed in Fig 6. The important residues which make significant contacts with the compound include Glu119, Asp151, Arg152, Trp179, Glu228, Glu287 and are expected to determine the binding process.

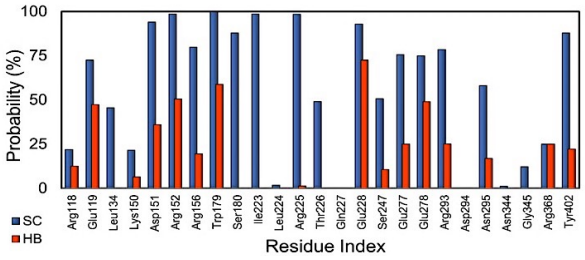

Figure 6. Probability of forming side chain (SC) and hydrogen bond (HB) contacts between CHEMBL1430043 and neuraminidase. The probability was calculated as the fraction of time the contacts were formed over MD trajectory 1. See Fig. S3 in SI for the similar plot for MD trajectory 2.

CONCLUSIONS

We have employed machine learning approaches to virtually screen the large CHEMBL compound database consisting of nearly 2 million compounds. The convolutional networks on graphs (GraphConv) model show the best performance on the test set with Pearson's R correlation of 0.8 and RMSE of 1.86 kcal/mol. Molecular docking and MD simulations were employed to understand structural insights into the binding process between top compounds and neuraminidase. The MD simulations shed light into side chain and hydrogen bond contacts between the top compound and neuraminidase and indicate important residues which stabilize the protein-ligand interaction. Experimental investigations on the enzymatic inhibition and antiviral activities as well as cytotoxicity of 11 compounds randomly selected from the top 400 compounds indicated that the drug screening hit rate was elevated to an impressive range of 9.1% to 18.2% when informed by virtual screening predictions. This is starkly contrasting the less than 1% hit rates achieved by 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. However, we acknowledge that further refinements are essential in our ongoing pursuit. While these compounds exhibit promise in NA inhibition, their efficacy in the cell-based assay was not good compared to the control group. This observation suggests the virtual screening process may only encompass some relevant factors on the virus's surface. As we chart our research path forward, we recognize the need for strategic compound modifications to enhance their

NA inhibition capacity. Future investigations will be dedicated to fine-tuning these compounds to elevate their antiviral potential and advance the fight against IAV.

ASSOCIATED CONTENT

Data and Software Availability

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

AUTHOR INFORMATION

Corresponding Author

Son Tung Ngo - Laboratory of Biophysics, Institute for Advanced Study in Technology, Ton Duc Thang University, Ho Chi Minh City, Vietnam; Faculty of Pharmacy , Ton Duc Thang University, Ho Chi Minh City, Vietnam; Email: ngosontung@tdtu.edu.vn

Jim-Tong Horng - Department of Biochemistry and Molecular Biology, College of Medicine, Chang Gung University, Kweishan, Taoyuan, Taiwan; Email: jimtong@cgu.edu.tw

ACKNOWLEDGMENT

This work was supported by Ho Chi Minh City Foundation for Science and Technology Development under grant number 115/QĐ-SKHCN.

REFERENCES

1. Palese, P., Influenza: old and new threats. *Nat Med* **2004,** *10* (12 Suppl), S82-7.
2. 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.
3. Ferguson, N. M.; Fraser, C.; Donnelly, C. A.; Ghani, A. C.; Anderson, R. M., Public health. Public health risk from the avian H5N1 influenza epidemic. *Science* **2004,** *304* (5673), 968-9.
4. Yen, H.-L.; Webster, R., Pandemic Influenza as a Current Threat. In *Vaccines for Pandemic Influenza*, Compans, R. W.; Orenstein, W. A., Eds. Springer Berlin Heidelberg: 2009; Vol. 333, pp 3-24.
5. WHO Pandemic (h1n1) 2009 briefing note 4. http://www.who.int/csr/disease/swineflu/notes/h1n1_situation_20090724/en/.
6. Neumann, G.; Noda, T.; Kawaoka, Y., Emergence and pandemic potential of swine-origin H1N1 influenza virus. *Nature* **2009,** *459* (7249), 931-9.
7. Mingxin, L.; Haizhou, L.; Yuhai, B.; Jianqing, S.; Gary, W.; Di, L.; Laixing, L.; Juxiang, L.; Quanjiao, C.; Hanzhong, W.; Yubang, H.; Weifeng, S.; George, F. G.; Jianjun, C., Highly Pathogenic Avian Influenza A(H5N8) Virus in Wild Migratory Birds, Qinghai Lake, China. *Emerg Infect Dis* **2017,** *23* (4), 637.
8. 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.
9. 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.
10. 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).
11. 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.
12. 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 Materials Lett* **2023,** *5* (3), 722-729.
13. 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.
14. 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.
15. 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.
16. Das, K.; Aramini, J. M.; Ma, L.-C.; Krug, R. M.; Arnold, E., Structures of influenza A proteins and insights into antiviral drug targets. *Nature Structural & Molecular Biology* **2010,** *17* (5), 530-538.
17. Klaus, S., Preventing and treating influenza. *BMJ* **2003,** *326* (7401), 1223.
18. Hurt, A. C., The epidemiology and spread of drug resistant human influenza viruses. *Current Opinion in Virology* **2014,** *8*, 22-29.
19. Bloom, J. D.; Gong, L. I.; Baltimore, D., Permissive Secondary Mutations Enable the Evolution of Influenza Oseltamivir Resistance. *Science* **2010,** *328* (5983), 1272-1275.
20. Cass, L. M. R.; Efthymiopoulos, C.; Bye, A., Pharmacokinetics of Zanamivir After Intravenous, Oral, Inhaled or Intranasal Administration to Healthy Volunteers. *Clinical Pharmacokinetics* **1999,** *36* (1), 1-11.
21. Van Drie, J. H., Computer-aided drug design: the next 20 years. *J Comput Aided Mol Des* **2007,** *21* (10), 591-601.
22. Marshall, G. R., Computer-Aided Drug Design. *Ann. Rev. Pharmacol. Toxicol.* **1987,** *27*, 193-213.
23. Ngo, S. T.; Hung Minh, N.; Le Thi Thuy, H.; Pham Minh, Q.; 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.
24. 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.
25. 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.
26. Vijayakrishnan, R., Structure-based drug design and modern medicine. *J. Postgrad. Med.* **2009,** *55* (4), 301-304.
27. Sliwoski, G.; Kothiwale, S.; Meiler, J.; Lowe, E. W., Computational Methods in Drug Discovery. *Pharmacol. Rev.* **2014,** *66* (1), 334-395.
28. Mendez, D.; Gaulton, A.; Bento, A. P.; Chambers, J.; De Veij, M.; Félix, E.; Magariños, María P.; Mosquera, Juan F.; Mutowo, P.; Nowotka, M.; Gordillo-Marañón, M.; Hunter, F.; Junco, L.; Mugumbate, G.; Rodriguez-Lopez, M.; Atkinson, F.; Bosc, N.; Radoux, Chris J.; Segura-Cabrera, A.; Hersey, A.; Leach, Andrew R., ChEMBL: towards direct deposition of bioassay data. *Nucleic Acids Res* **2018,** *47* (D1), D930-D940.
29. 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* **2016**, 785–794.
30. 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.
31. Ramsundar, B.; Eastman, P.; Walters, P.; Pande, V.; Leswing, K.; Wu, Z., *Deep Learning for the Life Sciences: Applying Deep Learning to Genomics, Microscopy, Drug Discovery, and More*. O'Reilly Media: 2019.
32. 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* **2013,** *28*, 115-123.
33. 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.
34. Trott, O.; Olson, A. J., AutoDock Vina: Improving the speed and accuracy of docking with a new scoring function, efficient optimization, and multithreading. *Journal of Computational Chemistry* **2010,** *31* (2), 455-461.
35. Pham, T. N. H.; Nguyen, T. H.; Tam, N. M.; Y. Vu, T.; Pham, N. T.; Huy, N. T.; Mai, B. K.; Tung, N. T.; Pham, M. Q.; V. Vu, V.; Ngo, S. T., Improving ligand-ranking of AutoDock Vina by changing the empirical parameters. *Journal of Computational Chemistry* **2022,** *43* (3), 160-169.
36. 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.
37. Jorgensen, W. L.; Chandrasekhar, J.; Madura, J. D.; Impey, R. W.; Klein, M. L., Comparison of simple potential functions for simulating liquid water. *The Journal of Chemical Physics* **1983,** *79* (2), 926-935.
38. Wang, J.; Wolf, R. M.; Caldwell, J. W.; Kollman, P. A.; Case, D. A., Development and testing of a general amber force field. *Journal of Computational Chemistry* **2004,** *25* (9), 1157-1174.
39. 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**.
40. Sousa da Silva, A. W.; Vranken, W. F., ACPYPE - AnteChamber PYthon Parser interfacE. *BMC Research Notes* **2012,** *5* (1), 367.
41. 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.
42. Rosmarinic acid interferes with influenza virus A entry and replication by decreasing GSK3β and phosphorylated AKT expression levels. *Journal of Microbiology, Immunology and Infection* **2022/08/01,** *55* (4).
43. 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. *Journal of Medicinal Chemistry* **February 12, 2019,** *62* (5).
44. Chemicalize was Used for Prediction of Chemical Properties. https://chemicalize.com/, developed by ChemAxon.
45. Efron, B., Bootstrap Methods: Another Kook at the Jackknife. *Ann. Stat.* **1979,** *7*, 1-26.

Table of Contents Only
