MedChemExpress compounds prevent neuraminidase N1 *via* physics- and knowledge-based methods†

Quynh Mai Thai, ‡[ab] Trung Hai Nguyen,‡[ab] Huong Thi Thu Phung, [c] Minh Quan Pham, [de] Nguyen Kim Tuyen Pham,[f] Jim-Tong Horng[g] and Son Tung Ngo ★[ab]

Influenza A viruses spread out worldwide, causing several global concerns. Hence, discovering neuraminidase inhibitors to prevent the influenza A virus is of great interest. In this work, a machine learning model was employed to evaluate the ligand-binding affinity of *ca.* 10 000 compounds from the MedChemExpress (MCE) database for inhibiting neuraminidase. Atomistic simulations, including molecular docking and molecular dynamics simulations, then confirmed the ligand-binding affinity. Furthermore, we clarified the physical insights into the binding process of ligands to neuraminidase. It was found that five compounds, including micronomicin, didesmethyl cariprazine, argatroban, Kgp-IN-1, and AY 9944, are able to inhibit neuraminidase N1 of the influenza A virus. Ten residues, including Glu119, Asp151, Arg152, Trp179, Gln228, Glu277, Glu278, Arg293, Asn295, and Tyr402, may be very important in controlling the ligand-binding process to N1.

Introduction

The influenza A virus pandemics killed several millions of people in the last century, including the H1N1 pandemic in 1918, H2N2 pandemic in 1957, and H3N2 pandemic in 1968.[1,2] In this century, the influenza A virus, including H5N1,[3,4] H1N1,[5,6] H5N8,[7] and H7N9 (ref. 8) caused flu for a large number of people worldwide. Numerous studies were carried out to find a potential inhibitor for treating the disease.[9–15] In particular,

aLaboratory of Biophysics, Institute for Advanced Study in Technology, Ton Duc Thang University, Ho Chi Minh City, Vietnam. E-mail: ngosontung@tdtu.edu.vn

bFaculty of Pharmacy, Ton Duc Thang University, Ho Chi Minh City, Vietnam

cNTT Hi-Tech Institute, Nguyen Tat Thanh University, Ho Chi Minh City, Vietnam

dInstitute of Natural Products Chemistry, Vietnam Academy of Science and Technology, Hanoi, Vietnam

eGraduate University of Science and Technology, Vietnam Academy of Science and Technology, Hanoi, Vietnam

fFaculty of Environment, Sai Gon University, 273 An Duong Vuong, Ward 3, District 5, Ho Chi Minh City, Vietnam

gDepartment of Biochemistry and Molecular Biology, College of Medicine, Chang Gung University, Kweishan, Taoyuan, Taiwan

† Electronic supplementary information (ESI) available: Binding pose of N1 with the experimental ligand (green) and docked ligand (pink) at active site; interaction of target protein, N1 with inhibitors *via* docking simulation and MD; RMSD of inhibitor + protein systems and free inhibitor in solution; calculated results for available N1 complexes compared with the experimental values; schematic description of binding free energy calculation *via* the LIE approach; and formation of SC and HB with N1 of all residues. See DOI: https://doi.org/10.1039/d4ra02661f

‡ Contribution equally to the work.

neuraminidase is the most popular drug target since it is an important element in the delivery of viral progeny to human cells.[16] Oseltamivir, zanamivir, and peramivir are among the drugs approved for inhibiting the target. However, numerous resistances still persist.[17–19] Therefore, studying new inhibitors that are able to effectively inhibit neuraminidase remains an interesting issue.

Computer-aided drug design (CADD) is a powerful tool for rapidly and accurately screening several million compounds for potential enzyme inhibitors. It has been initially reported since October 5, 1981, when an article entitled "*Next Industrial Revolution: Designing Drugs by Computer at Merck*" was published by Fortune magazine.[20] CADD's influence is rapidly increasing due to a significant decrease in the cost and time of new drug development.[21] CADD can be used for both purposes, including searching for new inhibitors and repurposing existing drugs.[22–24] CADD has been contributing to the discovery of severally available drugs such as dorzolamide,[25,26] saquinavir, ritonavir, and indinavir.[20] In CADD, the computational approach is frequently used to probe potential inhibitors that could bind well to a protein target. Thus, determining of ligand-binding free energy is one of the most critical factors in CADD.[27] Then, researchers developed numerous schemes,[28] including physics- and knowledge-based methods, to solve this problem.[29–39] The combination of these approaches may enhance CADD.

In this work, we aim to use a combination of knowledge- and physics-based methods to search for potential inhibitors from the MedChemExpress (MCE) database for inhibiting

neuraminidase. In particular, the trained machine learning (ML) model was utilized to rapidly and accurately probe the ligand-binding affinity of *ca.* 10 000 compounds in the MCE database for neuraminidase. We then performed atomistic simulations to confirm the ML results and gain physical insights into the protein–ligand binding process. It should be noted that both docking and LIE calculations were initially validated over eight complexes including 4B7Q,[40] 4B7N,[40] 4B7J,[40] 5NZE,[41] 5NZ4,[41] 5NWE,[41] 5NZF,[41] and 5NZN.[41] The critical residues controlling the ligand-binding process of N1 target were probed to clarify the binding mechanism. The structural change of N1 active site was also investigated to understand how ligands effect the enzymic target. Finally, a shortlist of potential candidates emerged. The outcome probably boosts the flu therapy.

Materials and methods

Convolutional networks on graphs calculations

The trained ML model used convolutional networks on graphs (GraphConv)[39] was utilized to search for potential inhibitors for preventing the biological activity of neuraminidase according to unpublished work.[42] The model was published online at the GitHub URL **https://github.com/nguyentrunghai/Neuraminidase/tree/main/ML/code**. In particular, the molecular features can be learned on the fly by the deep learning method GraphConv. A molecular graph is distributed to convolutional layers that will learn molecular fingerprints.

Structure of receptor and ligand

The X-ray diffraction structure of H1N1 influenza virus neuraminidase was downloaded from the Protein Data Bank with the identification of 4B7Q.[40] The protein structure is complex with zanamivir, illustrating a sophisticated molecular interaction at the active site of the protein. Besides, the ligands were obtained from the MedChemExpress (MCE) database, a comprehensive repository containing information on approximately 10 000 chemical compounds, was utilized as the foundational resource for performing calculations.

Molecular docking simulations

The modified AutoDock Vina (mVina), which used a set of modified empirical parameters for improving ligand-ranking, was employed to dock inhibitors to the neuraminidase N1 binding site. The docking grid size was set to dimensions of $24 \times 24 \times 24$ Å. We chose exhaustiveness as the default value based on the previous benchmark.[43,44] The docking modes differ from one another by an amount energy of 7 kcal mol^{-1}.

Molecular dynamics simulations

MD simulations were performed to refine the molecular docking outcomes of MCE compounds to neuraminidase *via* GROMACS version 2019.6.[45] Particularly, neuraminidase and neutralized ions were presented by using the Amber99SB-iLDN force field.[46] The TIP3P water model was instantaneously used to parameterize water molecules.[47] Consequently, MCE compound was parameterized using general Amber force field[48] with the assistance of AmberTools18 (ref. 49) and ACPYPE packages.[50] It should be denoted that the quantum chemical calculation using B3LYP hybrid functional at the 6-31G(d,p) level of theory was carried out to gain chemical information about the ligand. During which, the restrained electrostatic potential (RESP) approach was employed to allocate the atomic charges over quantum simulations with implicit solvent ($\varepsilon = 78.4$).[48] Moreover, the neuraminidase + ligand complex was placed into a dodecahedron periodic boundary condition (dPBC) box with a volume of *ca.* 570 nm^3 (see Fig. 1A). The solvated complex comprises 56 000 atoms totally. Besides, the free MCE compound was also inserted into a dPBC box with a volume of *ca.* 66 nm^3 (see Fig. 1B), which system thus consists of 6500 atoms totally.

The steepest descent method was used to find the lowest energy state for both solvated complexes and free ligands in solution systems. The systems were then relaxed over canonical
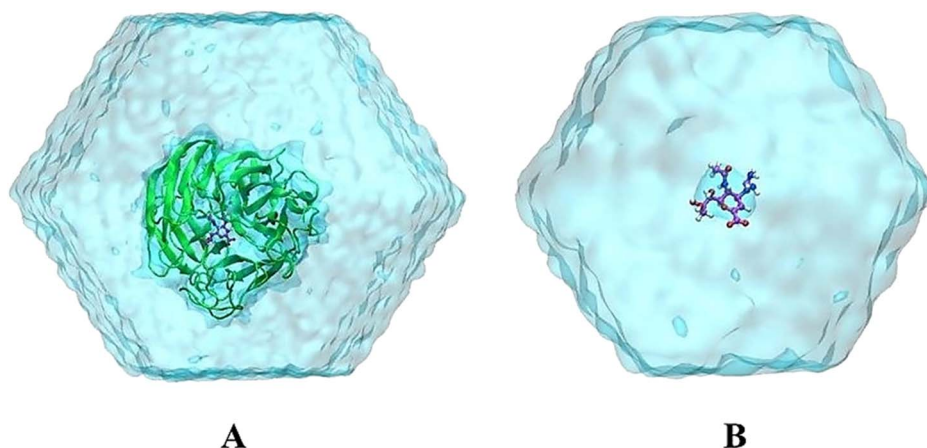

Fig. 1 The outset configurations of MD simulations consist of (A) N1 + zanamivir (4B7Q) in solution and (B) free zanamivir in solution. VMD 1.9.3 (ref. 51) was used to provide the figure.

and isothermal–isobaric ensemble simulations with a length of 100 ps each. The N1 + ligand and free ligand in solution were then simulated over 100.0 and 5.0 ns of MD simulations for each trajectory, respectively. Particularly, the systems were simulated at 310 K and 1 atm.

Linear interaction energy

Linear interaction energy (LIE) method can be utilized to probe the ligand-binding free energy.[52] The thermodynamic diagram of the approach is expressed in Fig. S1 of the ESI file.† The ligand-binding free via LIE approach, ΔG_{LIE}, is thus calculated as the average of the van der Waals (vdW) and electrostatic (cou) interaction energy differences of a ligand with their surrounding atoms upon association. In particular, the ligand in the bound state, bound to N1, is annotated as a subscription of b. Besides, the state of free ligand in solution is denoted as a subscription of f. The formula for estimating the ligand-binding free energy via LIE method, ΔG_{LIE}, can be expressed as follow.

$$\Delta G_{LIE} = \alpha(\langle V_{l-s}^{vdW}\rangle_b - \langle V_{l-s}^{vdW}{}_f\rangle) + \beta(\langle V_{l-s}^{cou}\rangle_b - \langle V_{l-s}^{cou}{}_f\rangle) + \gamma \quad (1)$$

whereas the empirical parameters, α, β, and γ can be obtained from the previous works.[53,54] Among these, the parameters α and β represent nonpolar and polar interactions, respectively.[53] The coefficient γ is associated with the alteration of the binding cleft hydrophobic mechanism corresponding to various inhibitors.

Analysis tools

The chemicalize webserver, a tool of ChemAxon, was utilized to predict the ligand protonation states.[55] The correlation error was calculated using 1000 rounds of the bootstrapping method.[56] The intermolecular sidechain contact (SC) between the ligand and the residual neuraminidase was counted when the spacing between their 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.

Results and discussion

The trained ML model using the GraphConv technique was used to rapidly and accurately probe the ligand-binding affinity of MCE compounds to neuraminidase N1. It should be noted that the model adopted the correlation coefficient from the corresponding experiments of $R = 0.80 \pm 0.04$ according to the unpublished work.[42] Moreover, the root-mean-square error (RMSE) between ML outcomes and experimental data is of RMSE = 1.86 ± 0.22 referring to the unpublished work.[42] The ML predicted ligand-binding affinity dropped in the range from -0.06 to -11.12 kcal mol^{-1}, with a median of -6.46 kcal mol^{-1}. The distribution of the ML predicted ligand-binding free energy of MCE databases is described in Fig. S1 of the ESI file.† Five compounds form an appropriate value of the ligand-binding

Table 1 Top five compounds suggested by ML model[a]

No.	Compound	ΔG_{ML}
1	Micronomicin	-10.01
2	Didesmethyl cariprazine	-9.86
3	Argatroban	-9.79
4	Kgp-IN-1	-9.62
5	AY 9944	-9.52

[a] The unit of energy is kcal mol^{-1}.

affinity of smaller than -9.50 kcal mol^{-1} (see Table 1). The atomistic simulations were then performed to simultaneously confirm and explain the ML outcomes.

Molecular docking simulations were usually performed to preliminary probe the binding affinities and poses of the ligands to receptors.[57,58] In this work, AutoDock Vina with the modified empirical parameters[43] was executed to explore the ligand-binding pose and affinity to N1. Initially, a benchmark was carried out to assess the performance of the docking protocol. The obtained results are described in Table 2. Interestingly, the correlation coefficient between docking and experimental data is of $R_{mVina} = 0.72 \pm 0.20$ (cf. Fig. 2). Besides,

Table 2 The data from docking simulations and experiments of known inhibitors against N1 neuraminidase[a]

No.	PDB ID	RMSD	ΔG_{mVina}	ΔG_{EXP}
1	4B7Q[40]	1.5	-10.6	-13.38
2	4B7N[40]	1.1	-11.4	-12.05
3	4B7J[40]	1.3	-9.9	-10.92
4	5NZE[41]	1.5	-10.2	-8.71
5	5NZ4 (ref. 41)	1.6	-10.1	-8.45
6	5NWE[41]	1.2	-9.9	-6.27
7	5NZF[41]	0.8	-9.4	-5.53
8	5NZN[41]	0.8	-10.0	-5.02

[a] The units of RMSD, K_i and energy are Å, nM and kcal mol^{-1} respectively.
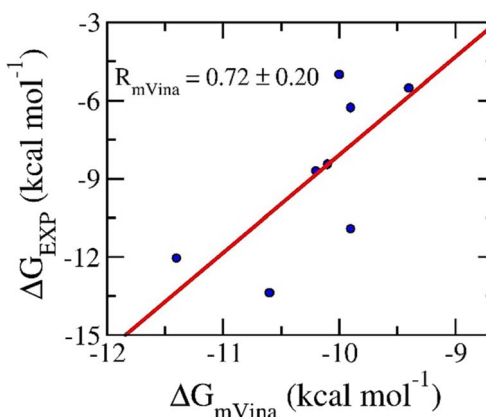

Fig. 2 Correlation between docking and experimental data. Docking results were obtained via mVina application.

the successful-docking rate, $\hat{p}$, is of 100% since the average of RMSD between docking and experimental poses of RMSD = 1.1 ± 0.1 angstrom (*cf.* Table S1 of the ESI file†). It should be noted that the ligand-binding pose is counted as a successfully docked structure if the RMSD between the docked and experimental shapes is less than 2 angstrom. Therefore, it may be argued that the protocol is an appropriate one to search for ligand-binding pose and affinity for the N1 target.

AutoDock Vina with the altered empirical parameters was thus executed to find the ligand-binding pose of five top-lead compounds to N1. The outcomes of molecular docking simulations were mentioned in Table 3. The ligand-binding free energy adopts in the range from −11.5 to −13.5 kcal mol^{-1}, with a median of −12.3 kcal mol^{-1}. Consequently, the binding poses of five ligands to N1 are shown in Fig. 3 and Table S2 of the ESI file.† The residues, including Glu119, Asp151, Glu228, Glu277, Glu278, Arg293, Asn295 form rigid contacts with five ligands implying that these residues may play an imperative role in the binding process of ligands to N1.

As mentioned above, the molecular docking simulations provided an appropriate outcome. However, because the protocol uses many constrains resulting in a decreasing accuracy of the outcome, MD simulations were thus required to be performed to refine the results. Each MD simulation with a length of 100 ns was carried out to turn the N1 + inhibitor complex into relaxed states, in which the docked conformations were employed as the initial structure of MD simulations. All

systems reach the state of stability after 10 ns of MD simulations. Therefore, the structures of the complexes over stable intervals, 50–100 ns (*cf.* Table S3 of the ESI file†), were then used for estimating the free energy difference of binding *via* LIE approach.[59] Besides, the free ligands in solution were also mimicked over 5.0 ns of MD simulations. The snapshots extracted over an interval 2.5–5.0 ns, which is a stable domain, were utilized for calculating the ligand-binding free energy *via* LIE protocol.[59]

The ligand-binding free energy between ligands and N1 can be investigated *via* the LIE approach.[52] In particular, the different free energy between two states involving *bound* and *unbound* states (Fig. S1 of the ESI file†) can be calculated *via* MD simulations by using eqn (1). In conventional, the empirical parameters $\alpha = 0.18$, $\beta = 0.50$, and $\gamma = 0.00$ were normally used.[60,61] Unfortunately, the computational values do not form any correlation to the respective experiments with a value of $R = -0.74$ (*cf.* Table S4 of the ESI file†). Using the different set of empirical parameters involving $\alpha = 0.288$, $\beta = -0.049$, and $\gamma = -5.880$,[54] the LIE outcomes provide an appropriate with the correlation coefficient of $R_{LIE} = 0.77 \pm 0.16$ (*cf.* Fig. 4 and Table S4 of the ESI file†). The obtained results imply that the physical insights into the binding process of ligands to N1 are possibility similar to Aβ systems,[54] Severe acute respiratory syndrome corona virus 2 (SARS-CoV-2) main protease,[62,63] and Monkeypox virus (MPVX) methyltransferase VP39.[57] The average of LIE results is of $\Delta G_{LIE} = -9.78 \pm 0.25$ kcal mol^{-1} that is slightly overestimated for experiments, $\Delta G_{EXP} = -9.26 \pm 0.20$ kcal mol^{-1}. Overall, one could argue that the LIE approach serves as a suitable protocol for estimating the ligand-binding free energy of N1.

The ligand-docking pose of five top-lead compounds to N1 was used as the starting structure of MD simulations. The complex was relaxed over 100 ns of MD simulations, which were also repeated two times. Based on the MD simulation results, the binding mechanism of N1 + inhibitors can be estimated by analyzing the intermolecular HB and SC contacts between N1 residues and their ligands. In this context, the intermolecular HB and SC contacts between N1 and their ligands were calculated over the relaxed intervals of MD simulations. The outcome

Table 3 Docking energy of the top five compounds to N1[a]

No.	Compound	ΔG_{mVin}[a]
1	Micronomicin	−11.6
2	Didesmethyl cariprazine	−11.5
3	Argatroban	−13.5
4	Kgp-IN-1	−12.8
5	AY 9944	−12.1

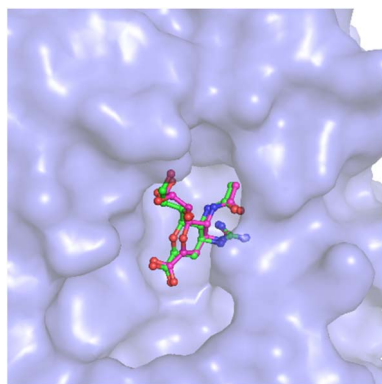[a] The unit of energy is kcal mol^{-1}.

Fig. 3 Supper position of docking and experimental pose of the complex 4B7Q. In particular, redocked conformation of zanamivir to N1 was provided by mVina approach.
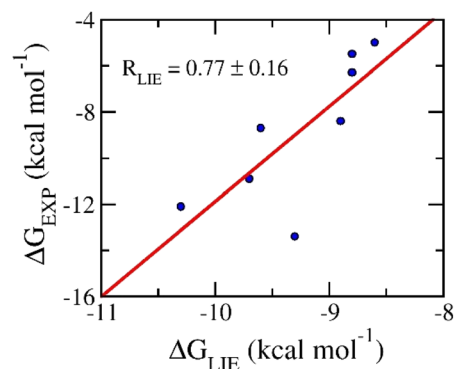

Fig. 4 Correlation between LIE and experimental data. In particular, LIE results were obtained over MD simulations.
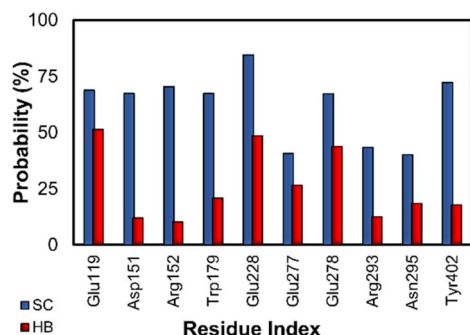
Fig. 5 Essential residues forming SC and HB for the inhibitors. In particular, the probability of SC and HB contacts were calculated over an interval 50–100 ns of MD simulations.

of the analysis was fully reported in Fig. S2 of the ESI file.† The N1 essential residues, which rigidly form strong HB and SC contacts with inhibitors, are described in Fig. 5. These residues are Glu119, Asp151, Arg152, Trp179, Gln228, Glu277, Glu278, Arg293, Asn295, and Tyr402. On average, the HB and SC contacts between these residues and ligands occupied more than 57 and 29%, respectively. One could argue that possible

mutations at these residues can significantly alter the ligand-binding affinity.

The structural change in the N1 binding site under the effects of their ligands was also probed over the equilibrium conformations of the solvated complex *via* MD simulations. The conformational change of ten essential residues including Glu119, Asp151, Arg152, Trp179, Gln228, Glu277, Glu278, Arg293, Asn295, and Tyr402 was probed *via* the clustering calculation with a non-hydrogen RMSD cutoff 0.12 nm. Fig. 6 represents the superposition between MD-refined conformational and starting shapes, which are noted with colorful and gray colors, respectively. In particular, the green arrows mention the structural change of the corresponding residues. The alteration of the N1 active site possibly implies that the neuraminidase biological activity would be inhibited.

The equilibrium snapshots of the complex over an interval 50–100 ns were used for computing the cou and vdW free interaction energies between ligands and surrounding atoms. Moreover, the free ligand in solution was also simulated to estimate the ligand-binding free energy. The difference in free interaction energies between bound and unbound states can be used as a major factor for computing the ligand-binding free energy *via* LIE scheme. The outcome was presented in Table 4.
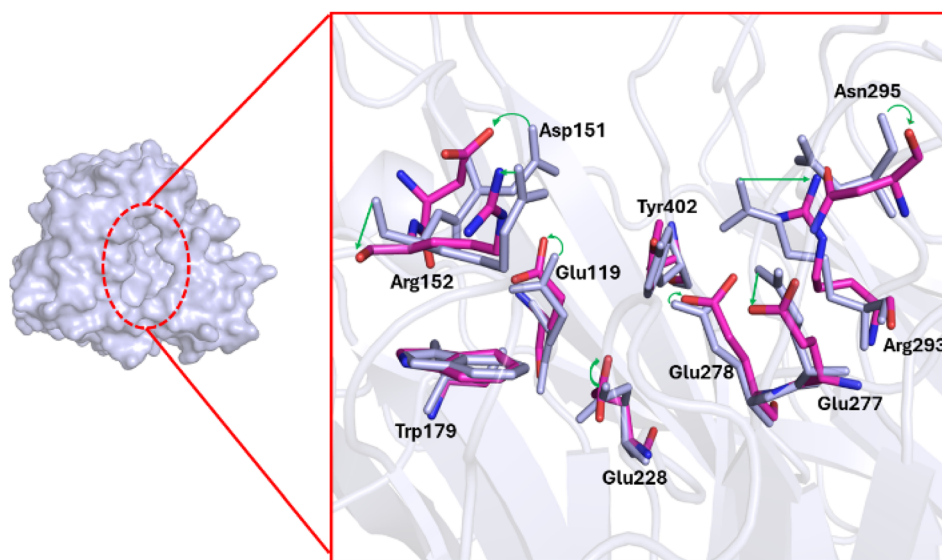

Fig. 6 The representative conformation of 10 crucial residues was determined through non-hydrogen RMSD clustering analysis with a cutoff of 1.2 Å. The colorful residues indicate the MD refined conformation compared with the initial structure, which is painted in gray. Changes in the residues are marked by green arrows.

Table 4 LIE data of five top-lead MCE compounds to neuraminidase[a]

No.	Name	$\langle V_{\text{l–s}}^{\text{cou}} \rangle_{\text{b}} - \langle V_{\text{l–s}}^{\text{cou}} \rangle_{\text{u}}$	$\langle V_{\text{l–s}}^{\text{vdW}} \rangle_{\text{b}} - \langle V_{\text{l–s}}^{\text{vdW}} \rangle_{\text{u}}$	ΔG_{LIE} [a]	$\langle \Delta G \rangle$	Predicted K_i range
1	Micronomicin	−9.05	−12.42	−9.01 ± 0.46	−9.51	Sub-micromolar
2	Didesmethyl cariprazine	−0.52	−13.17	−9.65 ± 0.19	−9.76	High-nanomolar
3	Argatroban	−23.37	−17.07	−9.65 ± 0.63	−9.72	High-nanomolar
4	Kgp-IN-1	−7.47	−18.23	−10.76 ± 0.92	−10.19	High-nanomolar
5	AY 9944	−4.55	−14.40	−9.80 ± 0.87	−9.66	High-nanomolar

[a] The computed error of LIE approach is the standard error of the mean. The unit of energy is of kcal mol^{-1}.

In particular, the LIE approach is in good agreement with the ML outcome due to the RMSE between ML and LIE results is of 0.70 kcal mol^{-1}. On average two computational approaches involving ML and ML, the predicted inhibition constant, ΔG, of five top-lead MCE compounds thus range from sub-micromolar to high-nanomolar (*cf.* Table 4). It may thus be argued that top-lead MCE compounds can inhibit the biological activity of N1.

Conclusions

In this context, physics- and knowledge-based approaches were employed to find potential inhibitors for prohibiting neuraminidase N1 from the MCE database. In particular, the ML model was tested that it formed a good correlation to experiment, $R = 0.80 \pm 0.04$, according to the unpublished work.[42] Moreover, the molecular docking and LIE calculations were indicated to be in good agreement with the corresponding experiments with correlation coefficients of $R_{\text{mVina}} = 0.72 \pm 0.20$ and $R_{\text{LIE}} = 0.77 \pm 0.16$, respectively. In particular, the physical insights into the binding process of ligands to N1 are possible similar to Aβ systems,[54] Severe acute respiratory syndrome corona virus 2 (SARS-CoV-2) main protease,[62,63] and Monkeypox virus (MPVX) methyltransferase VP39,[57] because they use the same empirical parameters of LIE approach. In particular, one can be explained since the parameters γ associated with the hydrophobic mechanism of the enzymic binding site, while, the parameters α and β related with nonpolar and polar terms.[53]

Five MCE compounds including micronomicin, didesmethyl cariprazine, argatroban, Kgp-IN-1, and AY 9944 were simultaneously suggested by both physics- and knowledge-based approaches that can inhibit the N1 with predicted inhibition constants ranging from sub-micromolar to high-nanomolar. Furthermore, ten residues including Glu119, Asp151, Arg152, Trp179, Gln228, Glu277, Glu278, Arg293, Asn295, and Tyr402 may play important elements regulating the ligand-binding process to N1. In particular, these residues rigidly form HB and SC contacts with ligands.

Conflicts of interest

There are no conflicts to declare.

Acknowledgements

References

1 P. Palese, *Nat. Med.*, 2004, **10**, S82–S87.

2 Y. C. Hsieh, T. Z. Wu, D. P. Liu, P. L. Shao, L. Y. Chang, C. Y. Lu, C. Y. Lee, F. Y. Huang and L. M. Huang, *J. Formosan Med. Assoc.*, 2006, **105**, 1–6.

3 N. M. Ferguson, C. Fraser, C. A. Donnelly, A. C. Ghani and R. M. Anderson, *Science*, 2004, **304**, 968–969.

4 H.-L. Yen and R. Webster, in *Vaccines for Pandemic Influenza*, ed. R. W. Compans and W. A. Orenstein, Springer, Berlin Heidelberg, 2009, vol. 333, ch. 1, pp. 3–24.

5 WHO, Pandemic (h1n1) 2009 briefing note 4, http://www.who.int/csr/disease/swineflu/notes/h1n1_situation_20090724/en/.

6 G. Neumann, T. Noda and Y. Kawaoka, *Nature*, 2009, **459**, 931–939.

7 L. Mingxin, L. Haizhou, B. Yuhai, S. Jianqing, W. Gary, L. Di, L. Laixing, L. Juxiang, C. Quanjiao, W. Hanzhong, H. Yubang, S. Weifeng, F. G. George and C. Jianjun, *Emerging Infect. Dis.*, 2017, **23**, 637.

8 Y. Wu, Y. Bi, C. J. Vavricka, X. Sun, Y. Zhang, F. Gao, M. Zhao, H. Xiao, C. Qin, J. He, W. Liu, J. Yan, J. Qi and G. F. Gao, *Cell Res.*, 2013, **23**, 1347–1355.

9 A. Perrier, M. Eluard, M. Petitjean and A. Vanet, *J. Phys. Chem. B*, 2019, **123**, 582–592.

10 W.-S. Choi, J. H. Jeong, J. J. Kwon, S. J. Ahn, K. K. S. Lloren, H.-I. Kwon, H. B. Chae, J. Hwang, M. H. Kim, C.-J. Kim, R. J. Webby, E. A. Govorkova, Y. K. Choi, Y. H. Baek and M.-S. Song, *J. Virol.*, 2018, **92**, e01580.

11 A. Albohy, Y. Zhang, V. Smutova, A. V. Pshezhetsky and C. W. Cairo, *ACS Med. Chem. Lett.*, 2013, **4**, 532–537.

12 G.-Q. Zhang, H. Chang, Z. Gao, Y.-p. Deng, S. Zeng, L. Shang, D. Ding and Q. Liu, *ACS Mater. Lett.*, 2023, **5**, 722–729.

13 N. M. Tam, M. T. Nguyen and S. T. Ngo, *J. Mol. Graphics Modell.*, 2017, **77**, 137–142.

14 M. Nagao, A. Yamaguchi, T. Matsubara, Y. Hoshino, T. Sato and Y. Miura, *Biomacromolecules*, 2022, **23**, 1232–1241.

15 M. Waldmann, R. Jirmann, K. Hoelscher, M. Wienke, F. C. Niemeyer, D. Rehders and B. Meyer, *J. Am. Chem. Soc.*, 2014, **136**, 783–788.

16 D. Kobasa, S. Kodihalli, M. Luo, M. R. Castrucci, I. Donatelli, Y. Suzuki, T. Suzuki and Y. Kawaoka, *J. Virol.*, 1999, **73**, 6743–6751.

17 A. J. Hay and F. G. Hayden, *The Lancet*, 2013, **381**, 2230–2232.

18 X. Liu, T. Li, Y. Zheng, K.-W. Wong, S. Lu and H. Lu, *Emerging Microbes Infect.*, 2013, **2**, e27.

19 R. Trebbien, S. Pedersen, K. Vorborg, K. Franck and T. Fischer, *Eurosurveillance*, 2017, **22**, 1–8.

20 J. H. Van Drie, *J. Comput.-Aided Mol. Des.*, 2007, **21**, 591–601.

21 G. R. Marshall, *Annu. Rev. Pharmacol. Toxicol.*, 1987, **27**, 193–213.

22 S. T. Ngo, N. Hung Minh, H. Le Thi Thuy, Q. Pham Minh, T. Vi Khanh, T. Nguyen Thanh and V. Van, *RSC Adv.*, 2020, **10**, 40284–40290.

23 S. T. Ngo, S.-T. Fang, S.-H. Huang, C.-L. Chou, P. D. Q. Huy, M. S. Li and Y.-C. Chen, *J. Chem. Inf. Model.*, 2016, **56**, 1344–1356.

24 N. M. Tam, M. Q. Pham, N. X. Ha, P. C. Nam and H. T. T. Phung, *RSC Adv.*, 2021, **11**, 17478–17486.

25 R. Vijayakrishnan, *J. Postgrad. Med.*, 2009, **55**, 301–304.

26 G. Sliwoski, S. Kothiwale, J. Meiler and E. W. Lowe, *Pharmacol. Rev.*, 2014, **66**, 334–395.

27 W. Yu and A. D. MacKerell, in *Antibiotics: Methods and Protocols*, ed. P. Sass, Springer New York, New York, NY, 2017, pp. 85–106, DOI: 10.1007/978-1-4939-6634-9_5.

28 U. Ryde and P. Soderhjelm, *Chem. Rev.*, 2016, **116**, 5520–5566.

29 S. T. Ngo, H. M. Hung and M. T. Nguyen, *J. Comput. Chem.*, 2016, **37**, 2734–2742.

30 R. W. Zwanzig, *J. Chem. Phys.*, 1954, **22**, 1420–1426.

31 D. L. Beveridge and F. M. DiCapua, *Annu. Rev. Biophys. Biophys. Chem.*, 1989, **18**, 431–492.

32 S. T. Ngo, T. H. Nguyen, N. T. Tung, P. C. Nam, K. B. Vu and V. V. Vu, *J. Comput. Chem.*, 2019, **41**, 611–618.

33 W. Jiang and B. Roux, *J. Chem. Theory Comput.*, 2010, **6**, 2559–2565.

34 Y. Meng, D. Sabri Dashti and A. E. Roitberg, *J. Chem. Theory Comput.*, 2011, **7**, 2721–2727.

35 W. Jiang, J. Thirman, S. Jo and B. Roux, *J. Phys. Chem. B*, 2018, **122**, 9435–9442.

36 J. C. Miles and A. J. Walker, *IEE Proc. Intell. Transp. Syst.*, 2006, **153**, 183–198.

37 DeepChem Home Page, **https://github.com/deepchem/deepchem**.

38 T. Chen and C. Guestrin, *KDD '16: Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, 2016, pp. 785–794, DOI: **10.1145/2939672.2939785**.

39 D. K. Duvenaud, D. Maclaurin, J. Iparraguirre, R. Bombarell, T. Hirzel, A. Aspuru-Guzik and R. P. Adams, in *Advances in Neural Information Processing Systems*, ed. C. Cortes, N. Lawrence, D. Lee, M. Sugiyama and R. Garnett, Curran Associates, Inc., 2015, vol. 28.

40 E. van der Vries, P. J. Collins, S. G. Vachieri, X. Xiong, J. Liu, P. A. Walker, L. F. Haire, A. J. Hay, M. Schutten, A. D. M. E. Osterhaus, S. R. Martin, C. A. B. Boucher, J. J. Skehel and S. J. Gamblin, *PLoS Pathog.*, 2012, **8**, e1002914.

41 J. Pokorná, P. Pachl, E. Karlukova, J. Hejdánek, P. Řezáčová, A. Machara, J. Hudlický, J. Konvalinka and M. Kožíšek, *Viruses*, 2018, **10**, 339.

42 T. H. Nguyen, N. Q. A. Pham, Q. M. Thai, V. V. Vu, S. T. Ngo and J.-T. Horng, *chemrxiv*, 2024, preprint, DOI: **10.26434/chemrxiv-2024-89cw5**.

43 T. N. H. Pham, T. H. Nguyen, N. M. Tam, T. Y. Vu, N. T. Pham, N. T. Huy, B. K. Mai, N. T. Tung, M. Q. Pham, V. V. Vu and S. T. Ngo, *J. Comput. Chem.*, 2021, **43**, 160–169.

44 N. T. Nguyen, T. H. Nguyen, T. N. H. Pham, N. T. Huy, M. V. Bay, M. Q. Pham, P. C. Nam, V. V. Vu and S. T. Ngo, *J. Chem. Inf. Model.*, 2020, **60**, 204–211.

45 M. J. Abraham, T. Murtola, R. Schulz, S. Páll, J. C. Smith, B. Hess and E. Lindahl, *SoftwareX*, 2015, **1–2**, 19–25.

46 A. E. Aliev, M. Kulke, H. S. Khaneja, V. Chudasama, T. D. Sheppard and R. M. Lanigan, *Proteins: Struct., Funct., Bioinf.*, 2014, **82**, 195–215.

47 W. L. Jorgensen, J. Chandrasekhar, J. D. Madura, R. W. Impey and M. L. Klein, *J. Chem. Phys.*, 1983, **79**, 926–935.

48 J. Wang, R. M. Wolf, J. W. Caldwell, P. A. Kollman and D. A. Case, *J. Comput. Chem.*, 2004, **25**, 1157–1174.

49 D. A. Case, I. Y. Ben-Shalom, S. R. Brozell, D. S. Cerutti, T. E. Cheatham III, V. W. D. Cruzeiro, T. A. Darden, R. E. Duke, D. Ghoreishi, M. K. Gilson, H. Gohlke, A. W. Goetz, D. Greene, R. Harris, N. Homeyer, Y. Huang, S. Izadi, A. Kovalenko, T. Kurtzman, T. S. Lee, S. LeGrand, P. Li, C. Lin, J. Liu, T. Luchko, R. Luo, D. J. Mermelstein, K. M. Merz, Y. Miao, G. Monard, C. Nguyen, H. Nguyen, I. Omelyan, A. Onufriev, F. Pan, R. Qi, D. R. Roe, A. Roitberg, C. Sagui, S. Schott-Verdugo, J. Shen, C. L. Simmerling, J. Smith, R. SalomonFerrer, J. Swails, R. C. Walker, J. Wang, H. Wei, R. M. Wolf, X. Wu, L. Xiao, D. M. York and P. A. Kollman, *AMBER 18*, University of California, San Francisco, 2018.

50 A. W. Sousa da Silva and W. F. Vranken, *BMC Res. Notes*, 2012, **5**, 1–8.

51 W. Humphrey, A. Dalke and K. Schulten, *J. Mol. Graphics*, 1996, **14**, 33–38.

52 H. Gutiérrez-de-Terán and J. Åqvist, in *Computational Drug Discovery and Design*, ed. R. Baron, Springer, New York, 2012, vol. 819, ch. 20, pp. 305–323.

53 M. Almlöf, B. O. Brandsdal and J. Åqvist, *J. Comput. Chem.*, 2004, **25**, 1242–1254.

54 S. T. Ngo, B. K. Mai, P. Derreumaux and V. V. Vu, *RSC Adv.*, 2019, **9**, 12455–12461.

55 Chemicalize was Used for Prediction of Chemical Properties, developed by ChemAxon, **https://chemicalize.com/**.

56 B. Efron, *Annu. Stat.*, 1979, **7**, 1–26.

57 Q. M. Thai, H. T. T. Phung, N. Q. A. Pham, J.-T. Horng, P.-T. Tran, N. T. Tung and S. T. Ngo, *J. Biomol. Struct. Dyn.*, 2024, 1–9, DOI: **10.1080/07391102.2024.2321509**.

58 T. H. Nguyen, N. M. Tam, M. V. Tuan, P. Zhan, V. V. Vu, D. T. Quang and S. T. Ngo, *Chem. Phys.*, 2023, **564**, 111709.

59 J. Aqvist, C. Medina and J.-E. Samuelsson, *Protein Eng.*, 1994, **7**, 385–391.

60 S. Bjelic, M. Nervall, H. Gutiérrez-de-Terán, K. Ersmark, A. Hallberg and J. Åqvist, *Cell. Mol. Life Sci.*, 2007, **64**, 2285–2305.

61 T. Hansson, J. Marelius and J. Åqvist, *J. Comput.-Aided Mol. Des.*, 1998, **12**, 27–35.

62 S. T. Ngo, N. M. Tam, M. Q. Pham and T. H. Nguyen, *J. Chem. Inf. Model.*, 2021, **61**, 2302–2312.

63 S. T. Ngo, T. H. Nguyen, N. T. Tung, V. V. Vu, M. Q. Pham and B. K. Mai, *Phys. Chem. Chem. Phys.*, 2022, **24**, 29266–29278.