

Prediction of the Inhibition of Influenza Virus Neuraminidase Various Strains by Means of a Generalized Model Constructed Using the Data on the Position of Known Ligands

A. V. Mikurova^{a, *}, A. V. Rybina^a, and V. S. Skvortsov^a

^a Institute of Biomedical Chemistry, ul. Pogodinskaya 10, Moscow, 119121 Russia

*e-mail: a.mikurova@ibmc.msk.ru

Received October 30, 2020; revised November 20, 2020; accepted November 27, 2020

Abstract—Several variants of models for predicting the IC_{50} values of inhibitors of influenza virus neuraminidase are presented for both individual strains and also for combinations of data for neuraminidases of several strains. They are based on the use of calculated energy contributions to the amount of change in the free energy of enzyme-inhibitor complexes. In contrast to previous works, aimed at the complex modeling, we added a procedure of comparison of the docking variants with one of the neuraminidase inhibitors, for which the structure of the complexes was determined experimentally. Selection of reference molecules for the comparison of structure similarity was made using the Tanimoto metrics and the limit of the RMSD value for a similar part of the structure was no more than 2 Å. Using this limitation and filtering datasets for a particular strain by the Q^2 value obtained in the leave-one-out control procedure it was possible to construct equations for predicting the IC_{50} value with a Q^2 value close to the minimum confidence threshold (0.57 in this work). Taking into consideration that in this version of the prediction models, a minimum set of energy contributions is used, which does not employ expensive calculations of entropy contributions, the result obtained supports the correctness of using a generalized model based on the data on the position of known ligands to predict the inhibition of neuraminidase of the influenza virus of various strains.

Keywords: influenza virus neuraminidase, inhibitors, computational methods, QSAR

DOI: 10.1134/S1990750821020086

INTRODUCTION

This work represents a logic continuation of our earlier studies aimed at the development of models for predicting the IC_{50} value (concentration causing 50% inhibition) for influenza virus neuraminidase inhibitors [1–3]. The search for new anti-influenza drugs continues. The need for them is dictated by the high variability of the influenza virus, which leads to the rapid emergence of viral drug resistance development [4]. Currently, the inhibitors of neuraminidase of influenza A and B viruses oseltamivir, zanamivir, peramivir, and laninamivir are registered as second generation drugs. The emergence of resistance has been repeatedly shown in relation to these drugs [4], so that the transition to the third and subsequent generations is inevitable. Despite the fact that the coronavirus has noticeably reduced public interest in the influenza virus, the problem of combating drug resistance and the creation of universal drugs effective against neuraminidase of various strains of the virus group still requires its solution.

This work is based on the hypothesis of minimizing the probability of (drug) resistance development in the case when an inhibitor does not have a narrow speci-

ficity for a particular serotype, but binds with approximately equally affinity to neuraminidases of all (or most) strains, albeit with a lower affinity. In this case, the most general calculated parameters, independent of the choice of a specific neuraminidase variant, should serve as a basis for creating predictive equations.

Earlier, we have shown that it is possible to construct predictive equations based on a set of experimentally determined (by X-ray structural analysis) and modeled variants of the structure of neuraminidases of various strains [2]. In addition, it was shown [3] that the use of the componentwise enthalpy contributions to the change in the free energy of the complex, calculated by the MMPBSA method [5], was sufficient to create models that allow dividing the set of ligands into weak, medium, and strong inhibitors. This work considers the possibility of using a priori information about the position of already known ligands to improve the predictive power of the model. In the process of modeling the complexes, we use the molecular docking procedure. Its main problem is the selection of an adequate solution. Frequently, the closest solution to the observed solution does not

match the best according to the scoring function. As a rule, if we analyze the best 100 variants selected according to the evaluation function, then there is one close to the experimentally observed variant among them. There is no other way to select from the list, except by the magnitude of the evaluation function or by the previously known position of individual structural groups. For example, for neuraminidase inhibitors, this is the position of the COO⁻ group or the localization of a positive charge on the nitrogen atom. The Protein Data Bank (PDB) contains a sufficient amount of data on the position of individual ligands at the binding site of various influenza virus neuraminidases [6]. After analyzing the pairwise similarity of one of these molecules with the modelled ligand, it is possible to find common or similar parts of the structure and estimate how close to the known ligands they are located in the ligand/enzyme complex. This information can be used to select the final version of the complex because the complex structure is subsequently optimized using molecular dynamics in which small deviations do not play a significant role.

MATERIALS AND METHODS

The following data on the structure and inhibitory activity (the IC₅₀ value) of chemical compounds have been used in this study: (1) a set of various low molecular weight compounds active towards neuraminidases from 5 strains A/Tokyo/3/67, A/tern/Australia/G70C/75, A/PR/8/34, A/Aichi/2/68 and B/Lee/40 (221 potential complexes) [1]; (2) a set of three known drugs (oseltamivir, zanamivir, peramivir) active towards 30 neuraminidase variants: A/teal/Hong Kong/W312/97, A/duck/Alberta/35/76, A/duck/Singapore/3/97, A/duck/Germany/1215/73, A/turkey/Ontario/6118/68, A/shearwater/Australia/1/72, A/duck/Czechoslovakia/56, A/quail/Italy/1117/65, A/duck/Memphis/546/74, A/turkey/Minnesota/916/1980, A/duck/Memphis/546/74, A/turkey/Minnesota/916/1980, as well as 8 variants of mutations, A/duck/Memphis/546/74 plus 9 variants of mutations (90 complexes) [2].

The general plan on modeling the complexes was described earlier [2, 3]. Development of predictive models was carried out using the parameters of the complexes calculated by the MM-PBSA (MM-GBSA) method. Data preparation, docking, and molecular dynamics were performed using the Dock 6.9 [7] and Amber 18.0 [8] software (force fields FF14SB and GAFF2). During each preparatory step the simulation time for molecular dynamics was 0.5 ns. The final simulation was performed for a 1 ns interval and 25 variants of the complexes were recorded every 40 ps; using these data the averaged values of the energy contributions to the change in the binding energy calculated by the MM-PBSA (MM-GBSA) method were obtained. The following set of independent variables was used in the equations:

molecular weight of the inhibitor; change in the magnitude of electrostatic interaction (ELE); change in the value of van der Waals interactions (VDW), as well as hydrophobic (PBSUR) and solvation (PBCAL) contributions to the change in free energy calculated by the Poisson-Boltzmann (PB) method, similar contributions calculated by the generalized Born method (GB) (GBSUR and GB). Linear regression equations were estimated by the value of Q² in a cross-check procedure using the leave-one-out method.

An important difference of this work from the previous ones was the use of an additional criterion during selection of a variant of the complex after docking. The procedure included a number of sequential steps (including preparatory ones):

(1) Selection of complexes of influenza virus neuraminidases from PDB with unique ligand structures. In total, 12 structures were used in this study: (PDBID: 2QWB, 1F8E, 1F8D, 3K37, 4KS4, 4KS1, 2QWF, 2QWD, 3CL0, 2QWE, 2SIM, 1INF). These complexes included: sialic acid, aromatic inhibitors, oseltamivir, zanamivir and their derivatives (Fig. 1).

(2) Spatial alignment of all complexes with known ligands, as well as neuraminidase variants used as a binding site during docking (Fig. 2).

(3) Each of the available molecular structures (a total of 185 original compounds) was compared pairwise with each of the 12 structures selected at step 1. Using the Instant JChem package (ChemAxon, Hungary) [9], the common part of the molecular structure was isolated and the degree of structure similarity was determined (using the Tanimoto metrics).

(4) After the docking procedure, up to 100 variants of the ligand position were saved. By calculating the root mean square deviation (RMSD) between the atoms of the coinciding part of the structure of the candidate molecule and that of the 12 molecules with which the candidate had the maximum similarity (or identity), the best variant was selected. During the selection process, the following parameters could be varied: the cutoff value of the evaluation function (in our work, the "Grid Score" value was no more than -20), the minimum value of the Tanimoto coefficient (0.5), and the maximum allowable RMSD value (2 Å).

Calculations were performed using a hybrid high-performance computing complex of the Federal Research Center "Computer Science and Control" of the Russian Academy of Sciences (FRC IU RAS) [10] IBM based on CPU Power9 and graphics accelerators Nvidia Tesla V100.

RESULTS AND DISCUSSION

The scheme for the selection of complexes used in this work assumes that some of them will be discarded. For example, Fig. 3 shows distribution of the maximum possible Tanimoto coefficient obtained by comparing 185 compounds from the total set with 12 inhib-

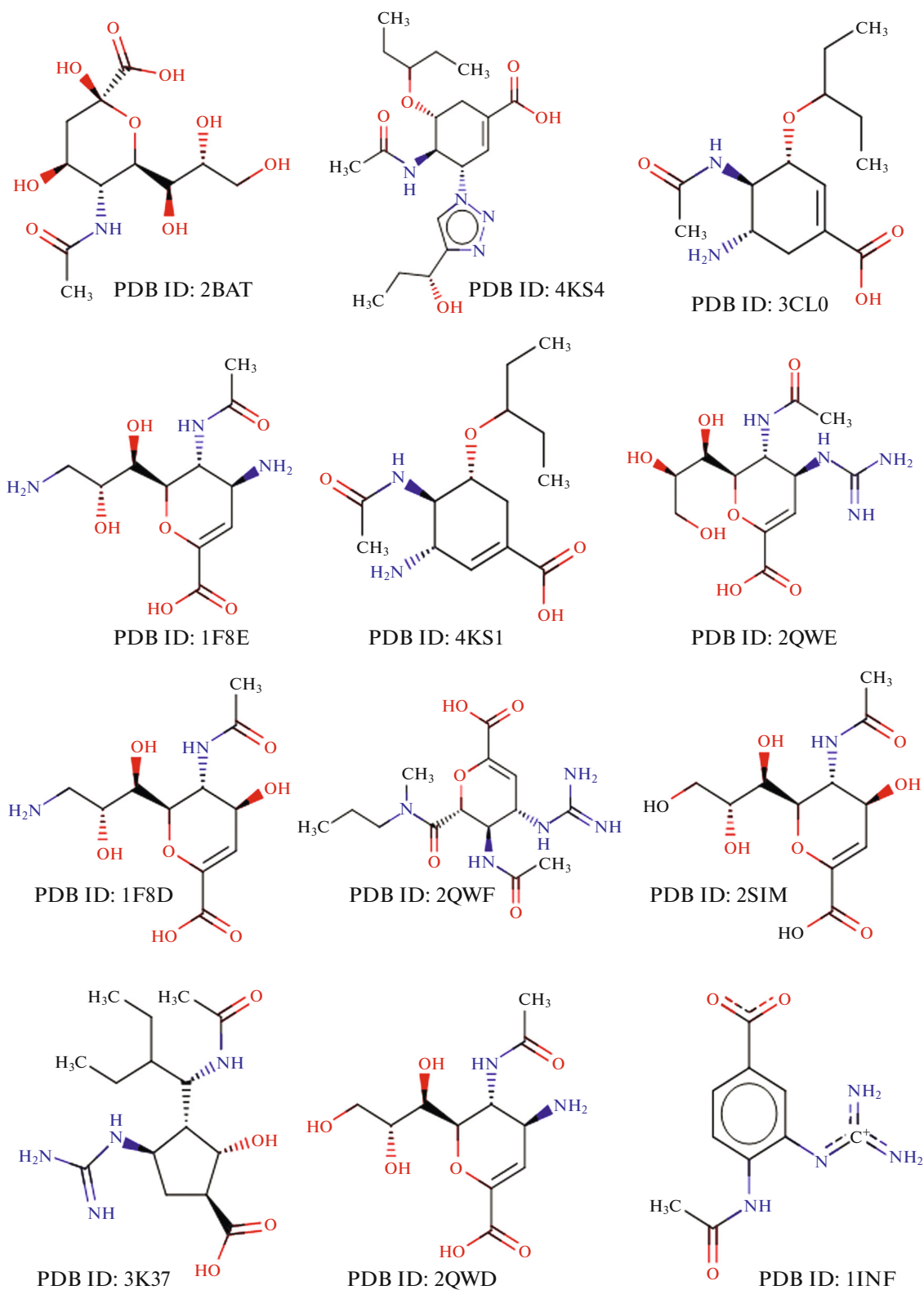


Fig. 1. Twelve structures of influenza virus neuraminidase inhibitors with known positions the complexes with neuraminidases deposited in the PDB.

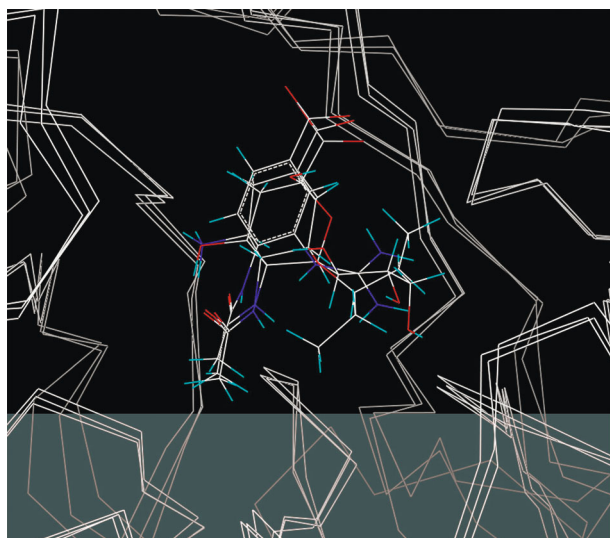


Fig. 2. An example of the spatial alignment of known neuraminidase complexes and the resultant superimposition of ligands.

itors taken from crystal structures (for demonstration, a comparison of the same compounds with one single molecule, oseltamivir, is also given). It can be seen that 20 compounds should be discarded already at this stage, since they do not satisfy the condition that the Tanimoto coefficient should be greater than 0.5. In addition, docking may not find a solution at all, or the cutoff by the given parameters, cannot yield any solution (RMSD is not more than 2 Å). The selection procedure resulted in selection of 252 out of 311 potentially possible (167 for the first set and 85 for the second one).

The main parameters of the equations for predicting the IC₅₀ value, or more precisely the value $\text{pIC}_{50} = -\log(\text{IC}_{50})$, built for various combinations of data subsets, are presented in Table 1. High learning R^2 values for set 1 can be the result of overlearning, although it gives a very good result in the leave-one-out cross validation test (in contrast to sets 2 and 4). An equation with 7 independent variables and a constant was always used as a starting point. In other words, the number of observations should be at least 40, which is well demonstrated by such a parameter as R_{rm}^2 —averaged over 10 training procedures, when the pIC_{50} values were mixed randomly. Simple pooling of sets did not provide a significant improvement. However, the result became significant when only sets with Q^2 higher greater than 0.4 were pooled (set 1 + 3 + 5 + 7 + 8, $Q^2 = 0.57$). Unfortunately, the quality of the data collected from the literature cannot be reliably confirmed, especially since the datasets themselves are often compiled from different sources published by different research groups and at different times. The sets 6, 7, and 8 represent another variant.

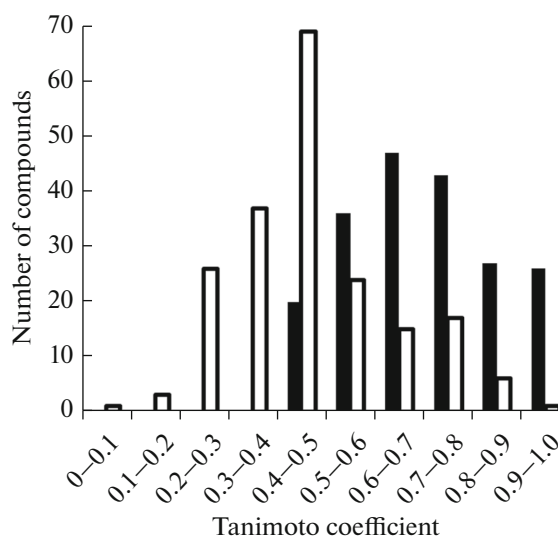


Fig. 3. Distribution of 185 compounds by the value of the Tanimoto coefficient obtained by pairwise comparison with ligands with a known position in the crystals of influenza virus neuraminidase. Black fill shows the maximum possible values obtained during comparison with all 12 structures, the unfilled shows comparison with oseltamivir.

The data were obtained by two groups of researchers, but the problem with these samples consisted in the narrow range of IC₅₀ values.

Thus, we can conclude that the use of an additional restriction in the selection of variants for the ligand position after the docking procedure makes it possible to obtain an adequate set of calculated parameters; using these parameters it is possible to construct an equation combining data on the inhibition of neuraminidases from various influenza virus strains. Inclusion of data on inhibition of individual strains into a total dataset is reasonable in the case of those datasets, which demonstrate minimum acceptable result in the leave-one-out test (e.g. they have Q^2 value higher than 0.4).

FUNDING

This work was performed within the framework of the Program for Basic Research of State Academies of Sciences for 2013–2020. The software porting on hybrid Power9 cluster was supported by the Russian Foundation for Basic Research (project 18-29-03100).

COMPLIANCE WITH ETHICAL STANDARDS

This article does not contain any research involving humans or the use of animals as objects.

Table 1. Parameters of pIC₅₀ prediction linear regression equations obtained during training and the test results

Training set	R_L^2	MSE _L	n	R_{rm}^2	Q ²	MSE _{LOO}	N	pIC ₅₀ range
1	0.82	0.82	20	0.45	0.66	1.63	4	2.0 : 8.8
2	0.55	0.45	20	0.34	0.1	1.1	5	3.2 : 9.3
3	0.62	0.72	43	0.2	0.51	0.93	5	4.6 : 10.1
4	0.39	1.66	61	0.16	0.32	1.86	2	3.27 : 9.85
5	0.54	1.59	23	0.3	0.47	1.86	3	2.1 : 8.5
6	0.38	0.92	28	0.28	0.15	1.36	2	4.56 : 9.88
7	0.49	0.27	29	0.21	0.39	0.35	3	6.58 : 9.5
8	0.55	0.31	28	0.28	0.44	0.39	3	6.94 : 10.22
from 1 to 5	0.35	2.3	167	0.155	0.24	2.7	4	2.0 : 10.1
from 6 to 8	0.42	0.45	85	0.2	0.04	0.7	1	4.56 : 10.22
from 1 to 8	0.45	2.4	252	0.1	0.22	1.7	6	2.0 : 10.22
1 + 3 + 5	0.47	2.2	86	0.07	0.38	2.5	3	2.0 : 10.1
7 + 8	0.37	0.5	58	0.1	0.17	0.6	6	6.58 : 10.22
1 + 3 + 5 + 7 + 8	0.59	1.75	143	0.01	0.57	1.84	4	2.0 : 10.22

R_L^2 is R^2 of learning; MSE_L is mean standard error of learning; n —is the number of observation during learning (selected complexes); R_{rm}^2 is averaged R^2 in the test of random mixing of the dependent value; Q² is Q² of the model; MSE_{LOO}—is mean standard error for the leave-one-out method; N is an optimal number of variables in the leave-one-out control (maximal 7 + constant). The set composition: 1—data on inhibition of the A/Tokyo/3/67 strain; 2—inhibition of the A/tern/Australia/G70C/75 strain; 3—inhibition of the A/PR/8/34 strain; 4—inhibition of the A/Aichi/2/68 strain; 5—inhibition of the B/Lee/40 strain; 6—inhibition of 30 variants of influenza virus A neuraminidase by oseltamivir; 7—inhibition of 30 variants of influenza virus A neuraminidase by zanamivir; 8—inhibition of 30 variants of influenza virus A neuraminidase by peramivir.

CONFLICT OF INTEREST

The authors declare that they have no conflicts of interest.

REFERENCES

- Mikurova, A.V., Rybina, A.V., and Skvortsov, V.S., *Biomeditsinskaya Khimiya*, 2016, vol. 62, no. 6, pp. 691–703. <https://doi.org/10.18097/PBMC20166206691>
- Mikurova, A.V. and Skvortsov, V.S., *Biomeditsinskaya Khimiya*, 2018, vol. 64, no. 3, pp. 247–252. <https://doi.org/10.18097/PBMC20186403247>
- Mikurova, A.V. and Skvortsov, V.S., *Biomeditsinskaya Khimiya*, 2019, vol. 65, no. 6, pp. 520–525. <https://doi.org/10.18097/PBMC20196506520>
- Breslav, N.V., Shevtchenko, E.S., Abramov, D.D., Priplipov, A.G., Zhuravleva, M.M., Oskerko, T.A., Kolobukhina, L.V., Merkulova, L.N., Shchelkanov, M.Yu., Burtzeva, E.I., and Lvov, D.K., *Voprosy Virusologii*, 2013, vol. 58, no. 1, pp. 28–32.
- Kollman, P.A., Massova, I., Reyes, C., Kuhn, B., Huo, S., Chong, L., Lee, M., Lee, T., Duan, Y., Wang, W., Donini, O., Cieplak, P., Srinivasan, J., Case, D.A., Cheatham, T.E., 3rd, *Acc. Chem. Res.*, 2000, vol. 33, 889–897.
- <https://www.rcsb.org> (application date October 1, 2020).
- Allen, W.J., Balius, T.E., Mukherjee, S., Brozell, S.R., Moustakas, D.T., Lang, P.T., Case, D.A., Kuntz, I.D., and Rizzo, R.C., *J. Comput. Chem.*, 2015, vol. 36, no. 15, pp. 1132–1156.
- Salomon-Ferrer, R., Case, D.A., and Walker, R.C., *WIREs Comput. Mol. Sci.*, 2013, vol. 3, pp. 198–210.
- <https://www.chemaxon.com> (application date October 1, 2020).
- Federal Research Center Computer Science and Control of Russian Academy of Sciences [Electronic resource]: site. Moscow: FRC CS RAS. URL: <http://hh-pcc.frcsc.ru> (application date November 9, 2020).

Translated by A. Medvedev