2D-SAR, Topomer CoMFA and molecular docking studies on avian influenza neuraminidase inhibitors

Bing Niu a,⁎,1, Yi Lu a,1, Jianying Wang a, Yan Hu a, Jiahui Chen a, Qin Chen a,⁎, Guangwu He c,⁎, Linfeng Zheng b,c,⁎⁎

a School of Life Sciences, Shanghai University, Shanghai 200444, China
b Department of Radiology, Shanghai General Hospital, Shanghai Jiao Tong University, Shanghai 200080, China
c Department of Radiology, Shanghai First People’s Hospital, Baoshan Branch, Shanghai 200940, China

A B S T R A C T

Avian influenza is a serious zoonotic infectious disease with huge negative impacts on local poultry farming, human health and social stability. Therefore, the design of new compounds against avian influenza has been the focus in this field. In this study, computational methods were applied to investigate the compounds with neuraminidase inhibitory activity. First, 2D-SAR model was built to recognize neuraminidase inhibitors (NAIs). As a result, the accuracy of 10 cross-validation and independent tests is 96.84% and 98.97%, respectively. Then, the Topomer CoMFA model was constructed to predict the inhibitory activity and analyses molecular fields. Two models were obtained by changing the cutting methods. The second model is employed to predict the activity (r² = 0.784 and r = 0.982). Molecular docking was also used to further analyze the binding sites between NAIs and neuraminidase from human and avian virus. As a result, it is found that same binding Total Score has some differences, but the binding sites are basically the same. At last, some potential NAIs were screened and some optimal opinions were taken. It is expected that our study can assist to study and develop new types of NAIs.

© 2018 The Authors. Published by Elsevier B.V. on behalf of Research Network of Computational and Structural Biotechnology. This is an open access article under the CC BY-NC-ND license (http://creativecommons.org/licenses/by-nc-nd/4.0/).

1. Introduction

The avian influenza (AI) virus is a negatively-stranded RNA virus that belongs to influenza virus A of the Orthomyxoviridae in the virus classification. The main host of this virus is birds, and humans and other mammals can also be infected. Influenza viruses can be divided into different subtypes based on the antigenicity differences of hemagglutinin (HA) and neuraminidase (NA). Currently, 18 subtypes of HA (H1–H18) and 11 NA (N1–N11) have been found in the world [1,2]. Among them, H5, H7 and H9 are the most harmful subtypes of birds, and some strains of H5 and H7 subtypes can cause high incidence and mortality of birds, which is called highly pathogenic avian influenza (HPAI) virus.

The avian influenza outbreaks caused serious economic losses to the poultry industry due to the death of birds, mass slaughter and restrictions on national and international trade [3]. 2014–2015, H5N2 HPAI broke out in the United States. By the end of November 2015, the outbreak has led to about 47 million poultry euthanized or killed. The economic losses were up to 3.3 billion U.S. dollars, and 18 trading partners banning the import of U.S. poultry [4,5]. The outbreak of AI in Miyazaki in Japan in 2010 forced public officials to ban tourists from entering the disaster-stricken areas, then slaughter all the infected birds, causing a loss of about 8.1 billion yen [6]. On January 27, 2007, a commercial turkey farm reported an H5N1 outbreak in the United Kingdom, leading to a mass culling of poultry in the farm [7].

NA, also known as sialidase, is a mushroom like homologous four dimer glycoprotein on the surface of influenza viruses A and B [8], and plays a key role in viral infection, replication, maturation and release [9,10]. The studies of the crystal structure of NA show that there is an active site on each subunit that binds to the inhibitor, and the residues are highly conserved. The conserved residues in the active pocket of influenza virus A and B almost identical [11]. Therefore, neuraminidase becomes an important target for the design of influenza drugs. Since 2010, neuraminidase inhibitors have become the only group of antiviral drugs recommended by the World Health Organization for the treatment and prevention of influenza A and B in humans [12].

Since computer-aided drug design can shorten the cycle of drug development and reduce the risk, in recent years, the structure-based
relationship (SAR) / quantitative structure activity relationship (QSAR) model based on machine learning algorithms and molecular docking technologies have been widely used in the field of chemical informatics [13–16]. Xue et al. [17] used heuristic algorithm (HM) and support vector machine (SVM) to construct prediction models to predict the binding affinity of 94 compounds to human serum albumin, leading to a good correlation coefficient \( r^2 \) of 0.86 and 0.94 and root-mean-square errors of 0.212 and 0.134 albumin drug binding affinity units, respectively. Sun et al. [18] constructed four QSAR models (PLS, HQSAR, CoMSIA and Almond model) by using 32 N-substituted oseltamivir derivatives NAIs. The \( r^2 \) and \( q^2 \) of the optimal model are 0.950 and 0.846, respectively. Li et al. [19] used 35 resveratrol derivatives with NA inhibitory activity by NA activity assay to build both CoMFA and CoMSIA model. In the CoMFA model, the \( q^2 \) is 0.62 with a standard error of estimate of 0.093 and \( r^2 \) is 0.973 on the training set. All these researches show a well predicted effect of the compounds, but they are generally focus on only one or two types of derivatives, such as sialic acid analogs and resveratrol derivatives, and the currently known NAIs also include cyclohexene derivatives, cyclopendane derivatives, benzoic acid derivatives, pyrrolidine derivatives, flavonoid analogs, and caffeic acid derivatives [20–22]. This will lead to a result that the model is only suitable for these derivatives, and it is difficult to get a good prediction effect once the structure of test compounds are different from the training set for modelling.

In this study, for this problem we collected more molecules (197 NAIs and 185 non-inhibitors) and more classification of inhibitors to broaden the application of the QSAR model. Six different machine learning algorithms were taken to construct 2D-SAR models to identify whether a compound is a NAI or not. Then partial least squares (PLS) method was used to build Topomer CoMFA model by all the inhibitors to predict the NA inhibitory activity. Molecular docking was also applied to simulate and analysis the interaction between NA and inhibitors. Finally, some potential NAIs were designed and screened, thus some advice for design new NAIs can be obtained. We hope that all these results will help pharmacologists to develop new drugs with higher NA inhibitory activity and cheaper price.

2. Materials and Methods

2.1. Data Preparation

In this work, 197 NAIs and 185 non-inhibitors were downloaded from the DUD database (http://dud.docking.org). The inhibitor molecules include sialic acid analogs, cyclopentane derivatives, benzoic acid derivatives, pyrrolidine derivatives, and flavonoid analogues.

45 molecular descriptors calculated for each molecule were used to construct a 2D-SAR model. First, three-dimensional structures of the molecules were optimized by molecular mechanics as implemented using the MM2 force field with the Polak–Ribiere algorithm until the root-mean-square gradient became <0.1 kcal/mol. MM2 force field [30] is one of the important force fields based on molecular mechanics for the optimization of small organic molecules. It is designed to reproduce equilibrium geometry of small organic molecules very precisely. This force field ignores the electronic motions in the molecular system and calculates energy as a function of position of atoms. It can optimize small molecules considering intra and intermolecular interaction energies considering of stretching of bonds, bending of angles, rotation around single bond, steric and electrostatic interactions between pairs of non-bonded atoms. And then, the descriptors were obtained for the most stable conformation of each molecule by using the AM1 semiempirical method at the restricted Hartree–Fock level with no configuration interaction. The AM1 method [31] was selected because it is a simple geometrical optimization that requires no complex mathematical calculation. Structural optimization using the AM1 method is rapid, and electronic structures are generated easily [32]. The molecular descriptors of all the compounds (382, including non-inhibitors) are shown in the Supplement 1 and the whole data set is randomly divided into training set (285) and test set (97).

In the Topomer CoMFA model, the pIC_{50} is used to represent the biological activity of the inhibitors. For molecules represented by IC_{50} were transformed according to the formula pIC_{50} = -log IC_{50}. 197 inhibitor molecules were also divided into training set (149) and test set (48).

2.2. 2D-SAR Model

2.2.1. CfsSubsetEval and Best First Algorithm

For a data set containing \( n \) vectors, there are \( 2^n \) possible combinations of feature subset. The best way to find an optimal subset is to try all the possible feature combinations. However, because of the large amount of computation, this strategy is difficult to implement. The CfsSubsetEval (CFS) method combined with Best-first (BF) search was employed to search the optimal feature subset in this study. CFS [33,34] is a heuristic feature-selection algorithm for evaluating the worth of a subset of attributes by considering the individual predictive ability of each feature along with the degree of redundancy between them. Therefore, feature-class and feature-feature correlations of training set were first calculated by CFS and the merit was calculated according to function (1):

\[
S_n = \frac{R^2}{\sqrt{1 + n(1-R^2)}}
\]

where \( S_n \) is the heuristic “merit” of a feature subset \( S \), \( R^2 \) is the mean feature-class correlation, and \( R^2 \) is the average feature-feature inter-correlation.

Eq. (1) forms the core of CFS and imposes a ranking on feature subsets in the search space of all possible feature subsets. Then, Best-first search was applied to search the feature subset space. BF is a general heuristic search algorithm which explores a graph by expanding the most promising node chosen according to an evaluation function [35,36]. BF search starts using the following steps:

1) Use greedy hill-climbing to enhance the backtracking facility to search for the space of attribute subset.

2) Set the number of consecutive non-improved nodes that allow control the completion level of the backtracking done.

3) Best first may start with the empty set of attributes and search forward, or start with the full set of attributes and search backward, or start at any point and search in both directions.

The search will terminate if five consecutive fully expanded subsets do not improve “merit” on the current best subset in order to avoid exploring the entire feature subset.

2.2.2. Modelling Methods

Various different machine learning methods, such as Adaboost [37], Bagging [38], J48, and some other methods were used for the training set data to construct the classification prediction model by the 10-fold cross-validation test. The test set is used to evaluate the prediction ability of the model.

2.2.3. Prediction Measurement

For measuring the success rates in this kind of binary classifications, a set of four metrics are usually used in this part. They are: (1) overall accuracy (Acc), (2) stability or Mathew’s correlation coefficient (MCC), (3) sensitivity (Sn), and (4) specificity (Sp). The SN, SP, ACC, and MCC...
can be represented as:

\[
\begin{align*}
\text{Sn} &= 1 - \frac{N^+}{N^-} \quad 0 \leq \text{Sn} \leq 1 \\
\text{Sp} &= 1 - \frac{N^-}{N^+} \quad 0 \leq \text{Sp} \leq 1 \\
\text{Acc} &= 1 - \frac{N^{+} + N^{-}}{N^{+} + N^{-}} \quad 0 \leq \text{Acc} \leq 1 \\
\text{MCC} &= \sqrt{\frac{1 + \frac{N^{+} - N^{-} \cdot N^{-} - N^{+}}{N^{+} + N^{-}}}{1 + N^{+} - N^{-} \cdot N^{-} - N^{+} \cdot N^{-}} } \quad -1 \leq \text{MCC} \leq 1
\end{align*}
\]

where \( N^+ \) is the total number of the positive events investigated while \( N^- \) the number of positive events incorrectly predicted as the negative ones; \( N^- \) the total number of the negative events investigated while \( N^- \) is the number of the negative events incorrectly predicted as the positive ones.

### 2.3. Topomer CoMFA

Topomer CoMFA is a rapid fragment-based three-dimensional quantitative structure-activity relationship (3D-QSAR) method [39,40]. Unlike traditional CoMFA, the Topomer CoMFA does not require subjective alignment of 3D ligand conformers and uses automatic alignment rules, so analysis is faster. The steps of the Topomer CoMFA are as follows:

1. Split the 3D molecular structure into fragments containing common features, open valence bonds or linkages.
2. Align each segment based on overlapping part to provide an absolute orientation of any segment.
3. Calculate the steric and electrostatic fields of the top-aligned segments.
4. Use PLS regression to build the model and the jackknife test to evaluate the model.

The \( r^2 \) and \( q^2 \) were used to evaluate the Topomer CoMFA models [41]. Cutoff values of \( r^2 \) and \( q^2 \) are 0.8 and 0.5, respectively. The optimum model was determined by the highest \( q^2 \), and the validity of the model depends on the \( r^2 \) value [42].

### 2.4. Molecule Docking

Two different NA crystal structures were downloaded from the Protein Data Bank (PDB) database: N2 protein of H3N2 (A/Tanzania/205/2010) (PDB ID: 4GZP) [43], H5N2 (A/Northern pintail/Washington/40964/2014) (PDB ID: 5HUK) [44]. Proteins were prepared with protein structure preparation module of the SYBYL X-2.0. All ligands and water molecules were removed, and hydrogen atoms were added. In addition, charges were added to N- and C-terminal regions of the NAs which became NH\(^+\) and COO\(^-\). The Meanwhile, during ligand converted, the two-dimensional (2D) representations of NAs were converted in three-dimensional (3D) ones and were minimized at physiological pH 7.0 with hydrogen atoms and charge by using Powell energy gradient method and the Gasteiger-Huckel system. At last, the Surfex-Dock module was used for molecule docking [45].

### 3. Results and Discussion

#### 3.1. 2D-SAR Model

For an unknown compound, it should first be determined whether it is an inhibitor of neuraminidase by 2D-SAR prediction model. After removing the factors with strong correlation, the remaining 22 parameters were used for variable screening, and 7 variables were finally selected (see Table 2). The model was built using 6 common machine learning methods, and the results are shown in Table 1. The ACC is 96.84% for cross-validation model and 98.97% for the independent test set by using the k-nearest neighbors algorithm (KNN), indicating that the model prediction effect was good. Through literature review, there is little classification model used in the prediction of NAIs. Li et al. [46] use SVM to establish four classification models to predict whether collected compounds were active or weakly active. The ACC and the MCC of the optimal model on the test set are 89.71% and 0.81 respectively. There may be some difference between the structures of compounds with active and weakly active, the difference between the collected inhibitor and non-inhibitor molecules is much larger, so the model has higher ACC and MCC values in this study.

The single factor modelling method was chosen to analyze the correlation of the ACC and molecular descriptors. According to Table 1, the IB1 method was selected to perform single factor modelling on the 7 molecular descriptors used for modelling. The results are shown in Table 2 and the correlation between selected descriptors are also calculated in Table 3. The independent predictive value of any single factor is low, but accumulating factors results in increased ACC. From the correlation matrix, the correlation between most selected descriptors is low, the correlation between TIndex and Rad is a little higher but not high enough to delete one of them [13]. TIndex was obtained via mathematical operations from the corresponding molecular graphs of compounds [47]. Though it is difficulty to encode stereo-chemical information, it can be easily and rapidly computed for any constitutional formula yielding good correlation abilities. The ACC and MCC of TIndex is the highest among 7 single factor model indicate that this factor might be the most important variable of the 2D-SAR model.

Sensitivity analysis was also performed on the selected 7 molecular descriptors (see Fig. 1) to further analyze the relationship between molecular descriptors and inhibitory activity. The change trends of NRBo and Ovality are similar to that between dipole length (DPLL) and inhibitory activity, activity increasing first and then decreasing as DPLL value increases. The TIndex, Rad, TVCon and the molecular activities showed a trend of increasing first, then decreasing and then increasing. Sol is closely related to the chemical structure of the molecule. The presence of hydrophilic groups such as hydroxyl, carboxyl, and amino groups can greatly increase the water solubility of the molecule. At the same time, the hydrophilic group easily interacts with the active pocket of the NA protein through hydrogen bonding and inhibits NA activity, so

### Table 1

| Classifier  | Training set | Test set |
|-------------|--------------|----------|
| SN (%)      | SP (%)       | ACC (%)  | MCC |
| Naive Bayes | 79.59, 97.83 | 88.42, 0.78 | 92.00, 97.87, 89.69, 0.81 |
| SVM         | 80.95, 97.83 | 89.12, 0.80 | 82.00, 93.62, 87.63, 0.76 |
| KNN         | 96.60, 97.10 | 96.84, 0.94 | 100.00, 97.87, 98.97, 0.98 |
| AdaBoost    | 89.80, 97.10 | 93.33, 0.87 | 72.00, 95.74, 83.51, 0.69 |
| Bagging     | 92.52, 95.65 | 94.04, 0.88 | 94.00, 89.36, 91.75, 0.84 |
| C4.5        | 93.20, 94.20 | 93.68, 0.87 | 96.00, 95.74, 95.88, 0.92 |

The significance of bold shows the best predicted result of the model.

### Table 2

| Molecular descriptors | Description               | SN (%) | SP (%) | ACC (%) | MCC |
|-----------------------|---------------------------|--------|--------|---------|-----|
| DPLL                  | Dipole length             | 74.15  | 69.57  | 71.93   | 0.44|
| TIndex                | Molecular topological index| 72.79  | 76.09  | 74.39   | 0.49|
| NRBo                  | Number of rotatable bonds | 65.99  | 60.14  | 63.16   | 0.26|
| Ovality               | Ovality                   | 65.99  | 77.54  | 71.58   | 0.44|
| Rad                   | Radius                    | 78.91  | 76.94  | 72.63   | 0.45|
| TVCon                 | Total valence connectivity| 69.39  | 78.09  | 72.63   | 0.46|
| Sol                   | Water solubility          | 62.59  | 80.43  | 71.23   | 0.44|
the inhibitory activity of compounds increases with the increasing of Sol.

3.2. Topomer CoMFA

According to the 2D-SAR model, a compound can be quickly determined whether it is a NAI or not. To further predict its inhibitory activity, the 3D-QSAR model should be used. Topomer CoMFA model was selected for quantitative analysis in various 3D-QSAR models. This model has been widely used in the auxiliary design of avian influenza, HIV, central nervous system diseases and other tumor-targeted therapeutic drugs [48–50]. In the Topomer CoMFA model, the activity of the inhibitor molecules are related to the segmentation methods [51]. In the modelling process, once the segmentation is completed, the input structure will be standardized and generate Topomers with the same substructure. As more identical substructures are identified in the test set, the predictive power of the model is better. In this study, the compound was divided into two segments, R1 and R2. The two Topomer CoMFA model was obtained by changing the segmentation methods by using training set. The q² and r² of the two model are shown in the Table 4. For a reliable predictive model, the q² should be >0.5 [52], the Model 2 was statistically significant (q² = 0.748 and r² = 0.982). Li et. al [19]. developed both CoMFA and CoMISA model by 35 NAIs in their research, the q² and r² of CoMFA model is 0.722 and 0.996. Obviously, the r² of our model is lower than them, but our q² is higher, and please note that the NAIs in our model is 197. It means that our model not only has a good predictive effect but also a wide range of applications.

Table 3
Correlation matrix of the selected descriptors.

|        | DPLL | TIndx | NRBo  | Ovality | Rad   | TVCon | Sol   |
|--------|------|-------|-------|---------|-------|-------|-------|
| DPLL   | 1    |       |       |         |       |       |       |
| TIndx  | 0.194| 1     |       |         |       |       |       |
| NRBo   | 0.152| −0.115| 1     |         |       |       |       |
| Ovality| −0.095| −0.412| −0.199| 1       |       |       |       |
| Rad    | 0.220| 0.883 | −0.167| −0.297 | 1     |       |       |
| TVCon  | −0.133| −0.430| −0.077| 0.167   | −0.371| 1     |       |
| Sol    | −0.322| −0.671| 0.098 | 0.225   | −0.607| 0.314| 1     |

Table 4
The results of two Topomer CoMFA model.

|        | Model 1 | Model 2 |
|--------|---------|---------|
| Segmentation methods |         |         |
| q²     | 0.686   | 0.748   |
| r²     | 0.854   | 0.982   |

Fig. 1. Sensitivity analysis results of selected molecular descriptor. A: DPLL; B: TIndex; C: NRBo; D: Ovality; E: Red; F: TVCon; G: Sol.
But the model was less predictive of certain compounds (shown in Fig. 2), such as Compound 80 and Compound 68, by using the independent test set. As we all know, the verification of 3D-QSAR analysis strongly depends on the selected training data set [26,53]. Compounds 80 and 68 are both benzoic acid derivatives (the molecular structure is shown in the Fig. 3). In this experiment, there are 19 benzoic acid derivatives, 13 in the training set and 6 in the test sets. Compared with other compounds, the number is slightly less, and the structure difference between the collected benzoic acid derivatives is relatively large, which may lead in poor prediction ability of the model for such kind of compounds. The plot of experimental pIC50 and predicted pIC50 for all the training set and independent test set molecules are show in Fig. 1 and the value are show in Supplement 2.

The CoMFA model also provides recommendations for the redesign NAIs of high selectivity, low toxicity, and high activity. In the electrostatic contour maps, the blue contours represent the positively charged areas of the molecule that favored an increase in activity, while the red contours stand for the negatively charged areas of molecule that favored an increase in the activity. And in the steric field contours, the green contours mean a bulky group here would be favorable for higher inhibitory activity, while the yellow color means oppositely [54]. We chose Compound 37 and 06 as a sample, the molecular structures were shown in Fig. 3. Through the analysis of steric and electrostatic field of CoMFA model, the yellow regions near the cutting place of R1 fragment in Compound 37 (see Fig. 4-A and B) indicate that the introduction of small groups can improve the inhibitory activity of the compounds. The R2 fragment of Compound 06 is obvious smaller than Compound 37, which could explain why the inhibitory activity of Compound 06 is much higher (the pIC50 value of Compound 06 and Compound 37 is 8.52 and 6.74, respectively). For the other compounds, adding appropriate groups in the suitable position can improve the biological activity of compounds.

3.3. Molecule Docking

To further analyze the interaction of a protein receptor with its ligand and revealing their binding mechanism, all the NA inhibitor molecules were simulated docking with N2 proteins of two different strains. The 3D structure of NA was measured using the single-crystal X-ray diffraction method by Yang et al. [55] and Zhu et al. [56]. Since NA is a tetramer composed by four identical polypeptides [57], the active pocket on only one of the subunits are selected for molecular docking in this study (the active pocket are shown in Fig. 5). All the docking results of the inhibitors were summarized in Table 5 (for detailed results, see Supplement 3). The docking score between NAIs and 4GZP were better than SHUK. We speculate that this may be related to the drug were designed based on human and the gene fragments of A/ Northern pintail/Washington/40,964/2014 strain were recombined [58,59]. For some compounds, the pIC50 values are relatively close, but Total Score has great difference, especially when docking with 5HUK. Taking Compounds 06 and 09 as an example, the pIC50 value of Compound 6 and 9 are all equal to 8.52, but the Total Score of Compound 6 is 6.1 and Compound 9 is just 4.13 (see Supplement 3). When using SYBYL for molecular docking, only hydrogen bonds between drugs and proteins can be identified. Therefore, we cannot exclude there are other interactions between some compounds and NA, such as covalent bonds and electrostatic interactions.

Compounds 132 and 120 were taken as an example, the interaction sites with 2 NA and molecular structure were shown in Fig. 6. Although the binding sites of the two compounds are different from the protein, but the Total Score is very close, for Compounds 132 the Total Score...
with 4GZP and 5HUK are 8.25 and 8.29, respectively. For Compounds 120, the action sites of 4GZP are Arg292, Arg371 and Tyr406, and Compounds 120 can bind to 5HUK with Asp151, Asp152, Arg292 and Arg371, the difference in one compound binding to 5HUK and 4GZP is also very small. Summery all the binding site of inhibitors to NA proteins, the interaction sites were mainly Arg292, Arg371, Asp151, Glu227, and Glu277, these sites were identical to those reported in many research [43–45]. Hydrophilic groups including carboxyl and amino groups tend to form hydrogen bonds with NA. The negative group (such as –COOH) easily binds to the Arg292 and Arg371 site,
and the positive group facilitates (such as -NH₂, -C≡N₃H₄) binding to the Asp₁₅₁ site [46,47]. This result is basically consistent with sensitivity analysis. The larger the Sol, the stronger the NAI activity is. Through molecular docking, it can provide some ideas for the future use of NAIs especially for poultry and promote the development of new veterinary drugs.

| Type                      | Number | 4GZP Max ST | Mean ST | The NO. of ST ≥ 5 | 5HUK Max ST | Mean ST | The NO. of ST ≥ 5 |
|---------------------------|--------|-------------|---------|------------------|-------------|---------|------------------|
| Cyclopentane derivatives  | 32     | 11.56       | 8.74    | 32               | 9.11        | 6.70    | 30               |
| Benzoic acid derivatives  | 19     | 8.88        | 6.73    | 14               | 8.14        | 5.42    | 11               |
| Sialic acid analogues     | 74     | 8.8         | 6.23    | 57               | 7.49        | 5.32    | 44               |
| Pyrrolidine derivatives   | 37     | 8.61        | 6.82    | 37               | 7.98        | 5.48    | 25               |
| Sialic acid analogues     | 35     | 7.79        | 6.53    | 35               | 7.57        | 5.59    | 26               |
| Total                     | 197    | 11.56       | 6.70    | 175              | 9.11        | 5.63    | 136              |

a ST represent Total Score.

Fig. 6. Docking results of NAIs (Compound 132 and 120) and 2 different NA. A and D: the structure of Compound 132 and 120; B and E: Views of the binding site of Compound 132 and 120 with 4GZP; C and F: Views of the binding site of Compound 132 and 120 with 5HUK. The yellow broken line indicates a hydrogen bond.
3.4. Design of Potential NAIs

The use of herbal medicine has been accepted in many countries, including regions with improved healthcare systems [60, 61]. Most herbal medicinal herbs are inexpensive and have been shown an anti-influenza activity in long-term practice [62, 63]. Herbs are rich in biologically active compounds, including phenols, flavonoids, flavonoids, and flavanols. Some flavonoids were designed based on the results of CoMFA model, and 2D-SAR model were applied to judge whether they were NAIs. All the compounds are shown in Supplement 4. Then the NA inhibitory activity of the compounds tested as NAIs were predicted by Topmer CoMFA model. The predicted pIC50 of these potential NAIs are shown in Table 6. Thirty compounds show a positive result though 2D-SAR model and some of them seems to have a good NA inhibitory activity (if the pIC50 > 5, means IC50 may lower than 10^{-5} mol/L). The pIC50 of Compounds M13 is 5.47 shows that this compound may have a good NA inhibitory activity. Through molecule docking, the interaction sites of M13 and 5HUK were shown in Fig. 7, the docking sites are almost the same as known NAIs. The hydroxyl of Compound M13 can interact with Trp178 residues, and the oxygen atoms on the ring can also interact with Arg118 and Arg371 through hydrogen bonds (green line in Fig. 7). The aromatic nucleus on the ligand can also interact with

| NO. | Pred pIC50 | NO. | Pred pIC50 | NO. | Pred pIC50 |
|-----|------------|-----|------------|-----|------------|
| M1  | 5.35       | M11 | 4.45       | M21 | 4.36       |
| M2  | 5.39       | M12 | 4.25       | M22 | 3.86       |
| M3  | 5.09       | M13 | 5.47       | M23 | 5.44       |
| M4  | 4.34       | M14 | 3.42       | M24 | 4.68       |
| M5  | 4.78       | M15 | 4.25       | M25 | 4.88       |
| M6  | 4.55       | M16 | 4.41       | M26 | 4.13       |
| M7  | 3.35       | M17 | 4.25       | M27 | 4.24       |
| M8  | 4.25       | M18 | 4.71       | M28 | 4.53       |
| M9  | 4.39       | M19 | 4.95       | M29 | 4.11       |
| M10 | 4.17       | M20 | 4.92       | M30 | 4.35       |

The significance of bold shows the best NA inhibitory activity in the selected ones.

The predicted pIC50 of selected compounds.

![Fig. 7. The hydrophobicity surface of 5HUK with Compound M13 (Left) and interaction site Compound M13 and 5HUK (Right).](image-url)
the positively (Arg151 and Arg292) and negatively charged amino acids (Glu277) residues by electrostatic interaction (orange line in Fig. 7). The CoMFA Contour map of Compound M13 (shown in Supplement 5) can also provide some recommendations to redesign the compound with higher inhibitory activity. We believe that these results will help pharmacologists to develop and design new drugs that are much cheaper and with highly effective against influenza. This will make a positive contribution to more patients to overcome this disease, even in less developed areas.

4. Conclusions

In this study, 2D-SAR and 3D-QSAR prediction models were constructed using the collected inhibitor molecules (n = 197) and non-inhibitor molecules (n = 185). First, NAI and non-inhibitors were classified by establishing a 2D-SAR model. Ten cross-validation tests have an accuracy of 96.84%, and independent tests have an accuracy of 98.97%. The Topomer CoMFA model was then built using only NAI s. Two models were obtained by changing the segmentation methods. Model 2 is selected with higher q² and r² values, the q² is 0.784 and the r² is 0.982. Molecular docking was also used to further analyze the binding sites between the NAI s with NA from two different hosts. The results showed that the Total Score had some differences between human and avian viruses, but the binding sites are basically the same. At last, some potential NAI s were screened by 2D-SAR model, and the NA inhibitory activity were predicted by Topomer CoMFA model. Compound M13 shows a good NA inhibitory activity and the predicted pIC50 is 5.47. In conclusion, we hope that our work will help to study drug and drug activity against avian influenza.

Conflicting Interests

The authors declare no conflicting interests.

Acknowledgement

The present study was supported by the National Key Research and Development Program of China (grant no. 2016YFD0501101), Shanghai Pujiang Program (17P1408000), Natural Science Foundation of Shanghai (17ZR1422500), Interdisciplinary Program of Shanghai Jiao Tong University (project number:YG2016MS26) and High Performance Computing Center Program of Shanghai University.

Appendix A. Supplementary Data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.jsb.2018.11.007.

References

[1] Ying W, et al. Bat-derived influenza-like viruses H17N10 and H18N11. Trends Microbiol 2014;22(4):183–91.
[2] Tong S, et al. New World bats harbor diverse influenza viruses. Virology J 2013;10(2):45–56.
[3] Kapczynski DR, et al. Protection of commercial turkeys following inactivated or recombinant influenza A(H5N1) by 3D-QSAR CoMFA and CoMSIA modeling. Med Chem Commun 2011;2(4):378–86.
[4] Zhang J, et al. Neuraminidase pharmacophore model derived from diverse classes of NAIs. J Integr Med 2015;13(5):531–42.
[5] Zheng M, et al. QSAR analyses on avian influenza virus neuraminidase inhibitors. Biomed Research Int 2014;2014:1–11.
[6] Okuyama T. Analysis of optimal timing of tourism demand recovery policies from tourists in Japan. Int J Tourism Res 2013;15(1):107–18.
[7] Guo H, et al. Molecular characterization of surface proteins hemagglutinin and neuraminidase from recent H5N1 avian influenza viruses. Mol Divers 2012;16(3):513–24.
[8] Stoll V, et al. Influenza neuraminidase inhibitors: structure-based design of a novel inhibitor series. Biochemistry 2003;42(3):718–27.
[9] Cheng CK, et al. From neuraminidase inhibitors to conjugates: a step towards better anti-influenza drugs? Future Med Chem 2014;6(7):757.
[10] Pizzorno A, Abed Y, Boivin G. Influenza drug resistance. Semin Respir Crit Care Med 2011;32(4):409–22.
[11] Caballero I, et al. Docking and quantitative structure-activity relationship studies for 3-fluoro-4-[(pyrrolidin-2-yl)carbonyl]benzoquinone and 4-(pyridin-2-yl)benzoquinone. J Med Chem 2012;55(4):1350–61.
[12] Guo H, et al. Three-dimensional quantitative structure-activity relationship studies on novel series of benzoazinone containing compounds acting as SIR inhibitors using CoMFA and CoMSIA. Bioorg Med Chem 2006;14(5):2369–79.
Ghose AK, et al. Conformational searching methods for small molecules I: study of the SYBYL SEARCH method. J Comput Chem 1993;14(9):1050–65.

Li Y, et al. Using support Vector Machine (SVM) for Classification of Selectivity of H1N1 Neuraminidase Inhibitors. Molecular Informatics 2016;35(3–4):116–24.

da Silva Junkes B, Amboni RDdMC, Augusto R. Semiempirical topological index: A novel molecular descriptor for quantitative structure–retention relationship studies. Internet Electron J Mol Design 2003;2:33–40.

Kumar S, Tiwari M. Topomer-CoMFA-based predictive modelling on 2,3-diaryl substituted-1,3-thiazolidin-4-ones as non-nucleoside reverse transcriptase inhibitors. Med Chem Res 2015;24(1):245–57.

Tian Y, et al. Design some New Type-I c-met Inhibitors based on Molecular Docking and Topomer CoMFA Research. Mol Inform 2014;33(8):536.

Gary T, José-Maria C, Trabanco AA. QSAR design of triazolopyridine mGlu2 receptor positive allosteric modulators. J Mol Graph Model 2014;53:82–91.

Xiang Y, Song J, Zhang Z. Topomer CoMFA and virtual screening studies of azaindole class renin inhibitors. Comb Chem High Throughput Screen 2014;17(5):458.

Golbraikh A, Tropsha A. Beware of q(2)! J Mol Graph Model 2002;20(4):269–76.

Zhang Q, et al. Rational questing for potential novel inhibitors of FabK from Streptococcus pneumoniae by combining FMO calculation, CoMFA 3D-QSAR modeling and virtual screening. J Mol Model 2011;17(6):1483.

Liu XL, et al. Microwave Assistant one Pot Synthesis, Crystal Structure, Antifungal Activities and 3D-QSAR of Novel 1,2,4-Triazolo[4,3-a]pyridines. Chem Biol Drug Des 2014;84(1):342–7.