

QSAR analyses on avian influenza virus neuraminidase inhibitors using CoMFA, CoMSIA, and HQSAR

Mingyue Zheng · Kunqian Yu · Hong Liu ·
Xiaomin Luo · Kaixian Chen · Weiliang Zhu ·
Hualiang Jiang

Received: 21 April 2006 / Accepted: 17 September 2006 / Published online: 11 November 2006
© Springer Science+Business Media B.V. 2006

Abstract The recent wide spreading of the H5N1 avian influenza virus (AIV) in Asia, Europe and Africa and its ability to cause fatal infections in human has raised serious concerns about a pending global flu pandemic. Neuraminidase (NA) inhibitors are currently the only option for treatment or prophylaxis in humans infected with this strain. However, drugs currently on the market often meet with rapidly emerging resistant mutants and only have limited application as inadequate supply of synthetic material. To dig out helpful information for designing potent inhibitors with novel structures against the NA, we used automated docking, CoMFA, CoMSIA, and HQSAR methods to investigate the quantitative structure–activity relationship for 126 NA inhibitors (NIs) with great structural diversities and wide range of bioactivities against influenza A virus. Based on the binding conformations discovered via molecular docking into the crystal structure of NA, CoMFA and CoMSIA models were successfully built with the cross-validated q^2 of 0.813 and 0.771, respectively. HQSAR was also carried out as a complementary study in that HQSAR technique does not require 3D information of these compounds and could provide a detailed molecular fragment contribution to the inhibitory activity. These models also show clearly how steric, electrostatic, hydrophobicity, and individual fragments affect

the potency of NA inhibitors. In addition, CoMFA and CoMSIA field distributions are found to be in well agreement with the structural characteristics of the corresponding binding sites. Therefore, the final 3D-QSAR models and the information of the inhibitor–enzyme interaction should be useful in developing novel potent NA inhibitors.

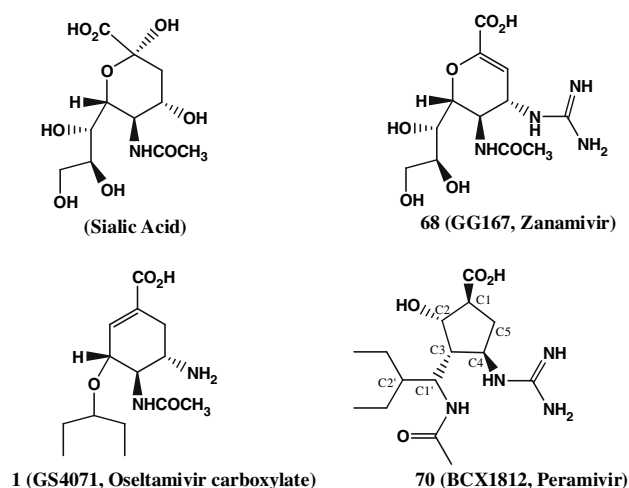
Keywords Avian influenza · Neuraminidase inhibitor · Docking · 3D-QSAR · CoMFA · CoMSIA · HQSAR

Introduction

The high case-fatality rate of the present H5N1 avian influenza epidemic in Asia, Europe and Africa has raised serious concerns about a global flu pandemic, which could become a massive threat to human being if the virus mutates for human-to-human transmission [1, 2]. In the early phases of a pandemic, the only treatment recommended would be neuraminidase inhibitors (NIs) [3, 4]. Neuraminidase (NA) [5] is a surface glycoprotein of influenza viruses that cleaves terminal sialic acids from carbohydrates, which is critical for viral release from infected cells and facilitates viral spread in the respiratory tract. The highly conserved catalytic active site of NA and crystal structures of inhibitor/NA complexes makes it an excellent target for anti-influenza drug design. Indeed, several potent inhibitors [6–9] (**1**, **68**, and **70** with bioactivities listed in Table 3) have been studied extensively, and two NI drugs are commercially available at present: zanamivir (**68**, from GlaxoSmithKline and Biota) and oseltamivir

M. Zheng · K. Yu · H. Liu · X. Luo · K. Chen ·
W. Zhu (✉) · H. Jiang
Drug Discovery and Design Center, State Key Laboratory
of Drug Research, Shanghai Institute of Materia Medica,
Chinese Academy of Sciences, 555 Zu Chong Zhi Road,
Zhangjiang Hi-Tech Park, Shanghai 201203, P.R. China
e-mail: wlzhu@mail.shcnc.ac.cn

(the ethyl ester prodrug of GS4071, **1**, from Hoffman La Roche and Gilead Sciences) [9, 10]. On an influenza pandemic, any strategy involving NI use would require stockpiles of these drugs. However, in most countries, the current levels of NI use are low and synthesis material sources of these drugs are limited. It is estimated that storage covering 20–25% of the population would be adequate during such an influenza pandemic [2, 3]. Furthermore, H5N1 virus evolves rapidly and strains resistant to the drug oseltamivir are emerging [11]. Therefore, there is an urgent need to design more potent NIs with novel structures.



Scheme. 1

Apparently, the interaction mechanism of the NA with its inhibitors would be greatly helpful in providing clues for designing small molecule inhibitors. Rational drug design utilizing available X-ray crystal structures of NA–inhibitor complexes has led to the discovery of some six-membered ring NIs including those two marketed drugs mentioned above [9, 10]. Recently, a series of pyrrolidine and cyclopentane-based NIs were also developed by employing an iterative structure-based methods [12, 13], some of which have shown equal or better efficacies on comparison with oseltamivir carboxylate and zanamivir [14]. Correlating the physicochemical properties of compounds or the computed binding energies with inhibitory activities is believed to gain an insight into the interaction mechanism of NA to inhibitors, and recently Verma et al. reported 17 quantitative structure–activity relationships (QSAR) for different sets of NIs [15]. However, most of these relationships were confined to compounds with a same scaffold; hence the predictability was very limited. To take the structural diversity into account, Yi et al. [16] applied the comparative molec-

ular similarity analysis (CoMSIA [17]) on 32 NIs from five different chemical classes, and derived a QSAR models with the q^2 around 0.70 at 10–14 optimal principle components (OPCs). This study might account for the key residues at binding site and the substituent feature requirements of NIs, whereas the OPC number of 10–14 might also indicate a problem of dimensionality and over-fitting. Indeed, for a five compounds test set, the predicted standard error (s) raises one or two orders of magnitude comparing with that from training set. Besides, this model contains few or no oseltamivir analogues (in fact, only the GS4071 was included but with an incorrect scaffold as depicted), which shows both clinical and pharmacological perspective on NA H5N1 influenza treatment [18]. In recent years, novel NIs kept emerging, with both different inhibitory profiles and chemically more accessible scaffolds [12, 19]. Thereby, it is high time to construct new QSAR models to get further insight into the relationship between molecular structures and bioactivities, in order to provide guidelines for the development of potent NIs with novel structures.

Currently, we report our new QSAR models, including Comparative Molecular Field Analysis (CoMFA) [20], CoMSIA, and HQSAR [21]. These models were first developed for a training set with 108 compounds and further validated using a test set with 18 compounds—together 126 known NI structures covering benzoic acids, carbocyclic derivatives, cyclopentanes, pyrrolidines, and miscellaneous chemical categories. The protocol adopted is described as follows: Those 126 NIs were initially docked to the binding site of NA (the crystal structure of H1N9 subtype) to identify their binding conformations and positions. With the binding information, 3D-QSAR models were constructed by CoMFA and CoMSIA and then mapped back to the topology of the active site of NA to reveal the accordance of contour maps with the distribution of the key residues in the binding site. Next, HQSAR was carried out as a complementary study to our CoMFA and CoMSIA analyses, in that HQSAR technique does not require 3D information of these compounds. In addition to deriving highly predictive models, the objective of the CoMFA and CoMSIA is to study the effect of different molecular field contributions on the biological activity; HQSAR, on the other hand, is to explore the detailed molecular fragment contribution. Therefore, we believe these 3D- and 2D-QSAR models would provide new useful information for structural modification of current inhibitors and the design of novel NIs with chemically more accessible structures.

Computational details

Materials

In total, 126 known NIs were prepared in this study, and their bioactivities against influenza A virus were mainly taken from the literature [12, 16, 22–26] (Table 1). IC_{50} values are available for most inhibitors except for a series of pyrrolidine NIs [12], for which IC_{50} values were calculated from K_i values using the Cheng–Prusoff equation [27]. The NA inhibitory activity data from different sources are comparable and we took the IC_{50} values from ref. [28] as the standard to normalize the experimental data, which was also adopted by Yi et al. [16] in their 3D-QSAR study against NIs. Inhibitory properties for the complete set of compounds are listed in Table 3, in which IC_{50} values were converted to the corresponding pIC_{50} ($-\log IC_{50}$). From this data set, 108 compounds (unasterisked, Table 3) were selected as members of the training set for model construction, and the other 18 compounds (asterisked, Table 3) as members of the test set for external model validation. Training and test sets were selected such that structurally diverse molecules possessing activities of a wide range were included in both sets. Thus, the data set is suitable for the followed QSAR model development.

Molecular modeling

The crystal structure of a wild-type NA complex with inhibitor GS4071 (**1**) was retrieved from the Brookhaven Protein Data Bank (PDB entry: 2QWK [29]). All ligands and water molecules were removed at first and the polar hydrogen atoms were added. Then, the potential of the 3D structures of NA was assigned according to the Amber 4.0 force field with Kollman-united-atom [30] charges encoded in **SYBYL 6.8** molecular simulation package [31]. The molecular geometries of all NIs in Table 1 were also modeled using the **SYBYL**. The initial structures were first minimized using molecular mechanics with the Tripos force field [32] with Gasteiger–Marsilli charges [33–35], and the energy convergence gradient value was set to 0.005 kcal/(mol Å). Then those structures were fully optimized, and Mulliken charges were calculated based on the semiempirical AM1 method [36], available in the **SYBYL** distribution of **MOPAC**. To find out the binding mode of the NIs to the NA, the docking program **AUTODOCK 3.0.3** [37, 38] was used to automatically dock the ligands to the enzyme. The Lamarckian genetic algorithm (LGA) [38] was applied to deal with the protein–inhibitor interactions. A Solis

and Wets local search [39] is performed for the energy minimization on a user-specified proportion of the population. The docked structures of the inhibitors were generated after a reasonable number of evaluations. The whole docking operation in this study could be stated as follows. First, the NA structure was checked for polar hydrogen and assigned for partial atomic charges, a PDBQ file was then created. Atomic solvation parameters and fragmental volumes were assigned to the NA using the **ADDSOL** module of the **AUTODOCK** program. Meanwhile, some of the torsion angles of the inhibitors were defined, allowing the conformation search for the ligands during the docking process. It should be noted that all amide bonds of NIs were treated as trans state, and would not be explored during molecular docking stage. Secondly, the grid map with $70 \times 70 \times 70$ points and a spacing of 0.375 Å was calculated using the **AUTOGRID** program to evaluate the binding energies between the inhibitors and the protease. The grid center was set at the active site position (27.162 18.86 62.012) and the affinity and electrostatic potential grids were calculated for each type of atom in the inhibitors. The energetic configuration of a particular ligand was found by tri-linear interpolation of affinity values and electrostatic interaction of the eight grid points around each atom of the ligand. Thirdly, some important parameters for LGA calculations were reasonably setup according to requirements of the Amber force field and our problem. The starting position and conformation was set as random, and the initial number of individuals in population is 300. The step size was set to 0.2 Å for translation and 5° for orientation and torsion. The maximum number of generations, energy evaluations, and docking runs were set to 5.0×10^5 , 2.5×10^6 , and 20, respectively. The elitism value is 1, which automatically survives into next generation. The mutation rate is 0.02, which is a probability that a gene would undergo a random change. The crossover rate, the probability of proportional selection, is 0.80. The pseudo-Solis and Wets local search method with a maximum of 300 iterations per local search was used. The probability of performing local search on an individual in the population is 0.06. The maximum number of consecutive successes before doubling or halving the local search step size is 4, and the same as failures. The lower bound on F , the termination criterion for the local search, is 0.01. Twenty docking simulations were performed for each inhibitor and conformations of multiple runs were then clustered with the root-mean square deviation (RMSD) tolerance of 0.5 Å. Finally, the conformation and position with the highest predicted binding affinity in the top-

Table 1 Structures of 126 NIs

| Cpd | R ₁ | R ₂ | R ₃ |
|-----|---|--|--------------------|
| 1 | CH ₂ Et ₂ | NH ₃ ⁺ | COMe |
| 2 | C ₃ H ₇ | NH ₃ ⁺ | COMe |
| 3 | CH ₂ OMe | NH ₃ ⁺ | COMe |
| 4 | CH ₂ CH ₂ CF ₃ | NH ₃ ⁺ | COMe |
| 5 | CH ₂ CH=CH ₂ | NH ₃ ⁺ | COMe |
| 6 | Cyclopentyl | NH ₃ ⁺ | COMe |
| 7 | Cyclohexyl | NH ₃ ⁺ | COMe |
| 8 | Phenyl | NH ₃ ⁺ | COMe |
| 9 | H | NH ₃ ⁺ | COMe |
| 10 | Me | NH ₃ ⁺ | COMe |
| 11 | C ₂ H ₅ | NH ₃ ⁺ | COMe |
| 12 | C ₄ H ₉ | NH ₃ ⁺ | COMe |
| 13 | C ₅ H ₁₁ | NH ₃ ⁺ | COMe |
| 14 | C ₆ H ₁₃ | NH ₃ ⁺ | COMe |
| 15 | C ₇ H ₁₅ | NH ₃ ⁺ | COMe |
| 16 | C ₈ H ₁₇ | NH ₃ ⁺ | COMe |
| 17 | C ₉ H ₁₉ | NH ₃ ⁺ | COMe |
| 18 | C ₁₀ H ₂₁ | NH ₃ ⁺ | COMe |
| 19 | CH ₂ CHMe ₂ | NH ₃ ⁺ | COMe |
| 20 | CH(Me)Et(R) | NH ₃ ⁺ | COMe |
| 21 | CH(Me)Et (S) | NH ₃ ⁺ | COMe |
| 22 | CH(Et)CH ₂ CH=CH ₂ (R) | NH ₃ ⁺ | COMe |
| 23 | CH(Et)CH ₂ CH=CH ₂ (S) | NH ₃ ⁺ | COMe |
| 24 | CH(Et)C ₇ H ₁₅ | NH ₃ ⁺ | COMe |
| 25 | CH(Et)CH ₂ -c-C ₆ H ₁₁ | NH ₃ ⁺ | COMe |
| 26 | CH(Et)CH ₂ CH ₂ -c-C ₆ H ₁₁ | NH ₃ ⁺ | COMe |
| 27 | CH ₂ C ₆ H ₅ | NH ₃ ⁺ | COMe |
| 28 | CH(Et)CH ₂ CH ₂ C ₆ H ₅ (R) | NH ₃ ⁺ | COMe |
| 29 | CH(Et)CH ₂ CH ₂ C ₆ H ₅ (S) | NH ₃ ⁺ | COMe |
| 30 | CH ₂ CH ₂ CH ₂ -C ₆ H ₄ (4-C ₆ H ₅) | NH ₃ ⁺ | COMe |
| 31 | CH(C ₃ H ₇) ₂ | NH ₃ ⁺ | COMe |
| 32 | H | NHC(=NH ₂ ⁺)NH ₂ | COMe |
| 33 | C ₃ H ₇ | NHC(=NH ₂ ⁺)NH ₂ | COMe |
| 34 | C ₄ H ₉ | NHC(=NH ₂ ⁺)NH ₂ | COMe |
| 35 | CH(Me)Et | NHC(=NH ₂ ⁺)NH ₂ | COMe |
| 36 | CH ₂ Et ₂ | NHC(=NH ₂ ⁺)NH ₂ | COCF ₃ |
| 37 | C ₃ H ₇ | NH ₃ ⁺ | COCF ₃ |
| 38 | C ₃ H ₇ | NH ₃ ⁺ | COEt |
| 39 | C ₃ H ₇ | NH ₃ ⁺ | SO ₂ Me |
| 40 | C ₃ H ₇ | NHCH=NH ₂ ⁺ | COMe |
| 41 | C ₃ H ₇ | NHC(=NMe ⁺)NH ₂ | COMe |

Table 1 continued

| | | | | |
|-----------|-------------------------------|---|---|--|
| | Cpd | R₁ | R₂ | |
| | 42 | Me | H | |
| | 43 | F | H | |
| | 44 | H | Me | |
| | Cpd | R₁ | R₂ | |
| | 45 | Me | C ₃ H ₇ | |
| | 46 | Me | C ₄ H ₉ | |
| | 47 | Me | CHEt ₂ | |
| | 48 | Me | CH ₂ CH ₂ C ₆ H ₅ | |
| | 49 | Me | Cyclohexyl | |
| | 50 | Et | C ₃ H ₇ | |
| | 51 | Et | C ₄ H ₉ | |
| | 52 | C ₃ H ₇ | C ₃ H ₇ | |
| | 53 | H | C ₄ H ₉ | |
| | 54 | H | CHEt ₂ | |
| | 55 | H | COEt | |
| | 56 | H | COCHMe ₂ | |
| | 57 | H | COCHEt ₂ | |
| | Cpd | R₁ | R₂ | R₃ |
| | 58 | Me | Me | NH ₃ ⁺ |
| | 59 | C ₃ H ₇ | Me | NH ₃ ⁺ |
| | 60 | C ₃ H ₇ | C ₃ H ₇ | NH ₃ ⁺ |
| | 61 | Et | Et | NH ₃ ⁺ |
| | 62 | CH ₂ CH ₂ C ₆ H ₅ | Et | NH ₃ ⁺ |
| | 63 | -CH(Me)CH=CHCH(Me)- | | NH ₃ ⁺ |
| | 64 | CH ₂ CH ₂ NH ₃ ⁺ | H | NHC(=NH ₂ ⁺)NH ₂ |
| | 65 | C ₃ H ₇ | H | NHC(=NH ₂ ⁺)NH ₂ |
| | 66 | Me | Me | NHC(=NH ₂ ⁺)NH ₂ |
| 67 | C ₃ H ₇ | Me | NHC(=NH ₂ ⁺)NH ₂ | |
| | Cpd | | | Cpd |
| | 68 | | | 72 |
| | Cpd | | | Cpd |
| | 69 | | | 73 |

Table 1 continued

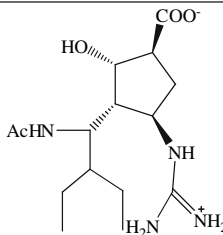
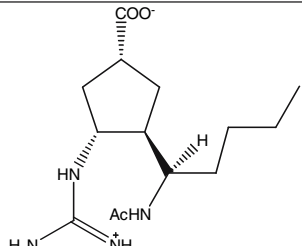
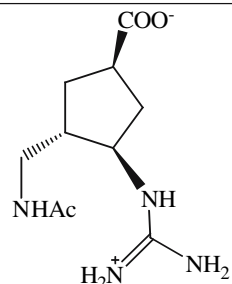
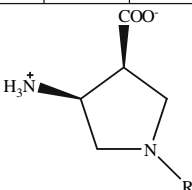
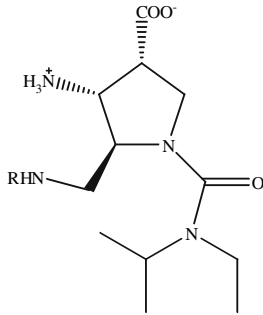
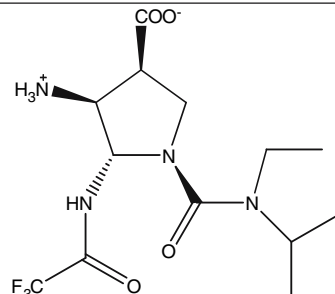
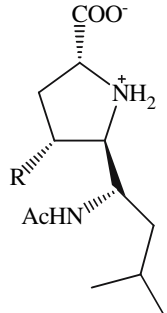
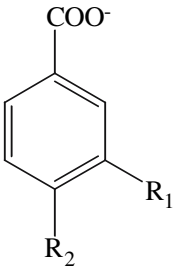
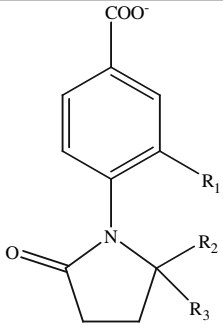
| | | | | |
|---|---|--|-------------------------|---|
|  | Cpd 70 |  | Cpd 74 | |
|  | Cpd 71 |  | Cpd | R |
| | | | 75 | OCHet ₂ |
| | | | 76 | CONet ₂ |
| | | | 77 | CON(C ₃ H ₇) ₂ |
| | | | 78 | CON(Et)CH(Me) ₂ |
| | | | 79 | CON[CH(Me) ₂] ₂ |
| | | | 80 | CON(CH ₂ CH ₂ OH)CH(Me) ₂ |
| | | | 81 | CON[(CH ₂) ₃ OH]CH(Me) ₂ |
| | | | 82 | CON[(CH ₂) ₅ OH]CH(Me) ₂ |
| | | | 83 | CON[(CH ₂) ₃ COO ⁻]CH(Me) ₂ |
| 84 | CON[(CH ₂) ₄ COO ⁻]CH(Me) ₂ | | | |
| 85 | CON[(CH ₂) ₃ NH ₃ ⁺]CH(Me) ₂ | | | |
| 86 | CON[(CH ₂) ₂ -2-Pyridyl]CH(Me) ₂ | | | |
|  | | | Cpd | R |
| | | | 87 | COMe |
| | | | 88 | COEt |
| | | | 89 | COCH=CH ₂ |
| | | | 90 | COCF ₃ |
| | | | 91 | SO ₂ Me |
|  | | | Cpd | R |
| | | | 92 | CF ₃ |
| | | | 93 | Me |
| | | | 94 | CH=CH ₂ |
|  | | | Cpd | R |
| | | | 95 | 4-imidazolyl |
| | | | 96 | 4-thiazolyl |
| | | | 97 | 2-thiazolyl |
| | | | 98 | 2-imidazolyl |
| | | | 99 | 3-pyrazolyl |
| | | | 100 | 3-isoxazolyl |
| | | | 101 | 5-isoxazolyl |
| | | | 102 | COOMe |
| | | | 103 | CH=CHCH ₃ (Z) |
| | | | 104 | CH=CH ₂ |
| | | | 105 | CH=CHCH ₃ (E) |

Table 1 continued

| | | | | |
|---|------------|--|---|---------------------------------|
|  | 106 | CH=C(Me) ₂ | | |
| | 107 | CH=CF ₂ | | |
| | 108 | CH=CHCl(<i>E</i>) | | |
| | 109 | CH=CHCl(<i>Z</i>) | | |
| | Cpd | R₁ | R₂ | |
| | 110 | NHC(=NH ₂ ⁺)NH ₂ | NHCOMe | |
| | 111 | CH(=NH ⁺)OH | NHCOMe | |
| | 112 | CH ₂ CH ₂ NO ₂ | NHCOMe | |
| | 113 | NHC(=NH ₂ ⁺)NH ₂ | NH ₂ SO ₂ CH ₃ | |
| | 114 | CH(=NH ⁺)OH | NH ₂ SO ₂ CH ₃ | |
|  | 115 | NHC(=NH ₂ ⁺)NH ₂ | NHCOCH(Me) ₂ | |
| | 116 | NHC(=NH ₂ ⁺)NH ₂ | CONHCH ₃ | |
| | 117 | NHC(=NH ₂ ⁺)NH ₂ | SO ₂ CH ₃ | |
| | 118 | NHC(=NH ₂ ⁺)NH ₂ | CH ₂ SOCH ₃ | |
| | 119 | NHC(=NH ₂ ⁺)NH ₂ | CH ₂ SO ₂ CH ₃ | |
| | Cpd | R₁ | R₂ | R₃ |
| | 120 | NHC(=NH ₂ ⁺)NH ₂ | H | H |
| | 121 | NHC(=NH ₂ ⁺)NH ₂ | H | CH ₂ OH |
| | 122 | NHC(=NH ₂ ⁺)NH ₂ | H | CH ₂ NH ₂ |
| | 123 | H | CH ₂ OH | CH ₂ OH |
| | 124 | NHC(=NH ₂ ⁺)NH ₂ | CH ₂ OH | CH ₂ OH |
| | 125 | NHCH ₂ Et ₂ | H | H |
| | 126 | NHCH ₂ Et ₂ | CH ₂ OH | CH ₂ OH |

ranked cluster was extracted from the optimized NI–NA complex. The geometrical optimizations of docked complexes were carried out using molecular mechanics method encoded in **SYBYL** with following parameters: the Amber-all-atom charge of the Amber 4.0 force field [40] and Powell method with the root-mean square (RMS) energy gradient of 0.05 kcal/(mol Å). The whole system was minimized to convergence. Although the solvation energies could not be explicitly considered during the minimization, the energy calculations were performed with a distance-dependent dielectric constant of 5 to mimic the solvation effect of the inhibitors in the protein environment [41]. These conformations and positions were aligned together inside the binding pocket of NA and used for further CoMFA and CoMSIA analyses, to explore the specific contributions of electrostatic, steric, and hydrophobic effects to the molecular bioactivities.

CoMFA

Steric and electrostatic interactions were calculated using a sp³ carbon atom as steric probe and a +1 charge as electrostatic probe with Tripos force field. The grid spacing was 2.0 Å in the *x*, *y*, and *z* directions, and

the grid region was automatically generated by the CoMFA routine to encompass all molecules with an extension of 4.0 Å in each direction. The default value of 30 kcal/mol was set to the maximum steric and electrostatic energy cutoffs. With standard options for variable scaling and SAMPLE [42] method, leave-one-out (LOO) cross-validations were carried out to determine the OPC number; the final CoMFA model was obtained with this optimum number of components and non-cross-validated conventional correlations.

CoMSIA

Five physicochemical properties, viz. steric, electrostatic, hydrophobic fields, and hydrogen bond donor and acceptor were evaluated. The steric contribution was reflected by the third power of the atomic radii of the atoms. The electrostatic properties have been introduced as Gasteiger–Masili charges, and an atom based hydrophobicity is estimated according to the parameterization developed by Ghose and co-workers [43, 44]. The lattice dimension was selected with a sufficiently large margin (4 Å) to enclose all aligned molecules as in CoMFA. Any singularities were avoided at atomic positions in CoMSIA fields because a

Gaussian type distance dependence of the physico-chemical properties was adopted; thus, no arbitrary cutoffs were required. Similarity indices were computed using a probe with a charge of +1, a radius of 1 Å, a hydrophobicity of +1, and 0.3 as attenuation factor α for the Gaussian-type distance. The statistical evaluation for the CoMSIA analyses were performed in the same way as described for CoMFA.

HQSAR

By transforming the chemical representation of a molecule into its corresponding molecular hologram, HQSAR requires no explicit 3D information for the ligands [45]. In this study, the method with its default parameters compiled in **SYBYL 6.8** was performed by the procedure described as follows: At first, a predefined set of rules was used to hash a molecule into a molecular fingerprint that encoded the frequency of occurrence of various molecular fragment types including: atoms, bonds, connections, chirality, and donor and acceptor. Next, the molecular fingerprint was cut into strings at a fixed interval as specified by a hologram length (HL) parameter. Then, all of the generated strings were hashed into a fixed length array. The **SYBYL** Line Notation (SLN) for each string was mapped to a unique integer in the range of 0–2³¹ using a cyclic redundancy check algorithm. This numerical representation of molecules was exploited by a subsequent correlation analysis with PLS LOO method. Finally, the optimal HQSAR model was constructed by screening the 12 default HL values ranging from 53 to 401.

Results and discussion

Molecular modeling

In this work, we selected the automated docking to align molecules for CoMFA and CoMSIA—a key procedure in such 3D QSAR studies. As we known, QSAR analyses based on some rough alignment such as atomic RMS fit and rigid body field fit techniques can not provide detailed information about how ligands bind to receptor, and superimposed conformations derived from such simple alignment rules are usually not even close to the bioactive conformations. However, the automated docking provides us a way to predict the postulated bioactive conformations of ligands, which can not only be used to build more reliable 3D QSAR models, but also eliminate the

difficulty in aligning compounds that covers more than one structure types.

To date, a number of crystal structures of NA-inhibitor complexes have been released, in which different kinds of NIs bound to the active site in a similar pattern [7, 12, 13, 46–48]. Investigation on these structures also showed that there is virtually no change in the orientation of the side chain amino acids of the NA active site, except for some minor conformational changes in Glu276 and Arg224. Consequently, in the present study we only took one crystal structure, the NA-GS4071 complex as the template for docking procedure. As depicted in the Fig. 1, the conformation of GS4071 from the X-ray model and that from the docked complex could be superimposed with a RMSD of 0.58 Å, which suggests that the parameter set for docking is capable of reproducing the X-ray structure. Furthermore, docked results of the complexes for zanamivir and peramivir, two other important NIs, were also examined. The superimposition of the docked structures with the corresponding crystal structures (PDB entry code 1NNC [48] for zanamivir and 1L7F [49] for peramivir), as shown in Fig. 2 revealed that the RMSD between the calculated and observed positions were 0.35 Å for zanamivir and 0.49 Å for peramivir. These results indicated that **AUTODOCK** was competent in reproducing the experimentally found binding position and conformation of NIs, even without consideration of the minor rearrangement of a few side chains caused by induce-fit.

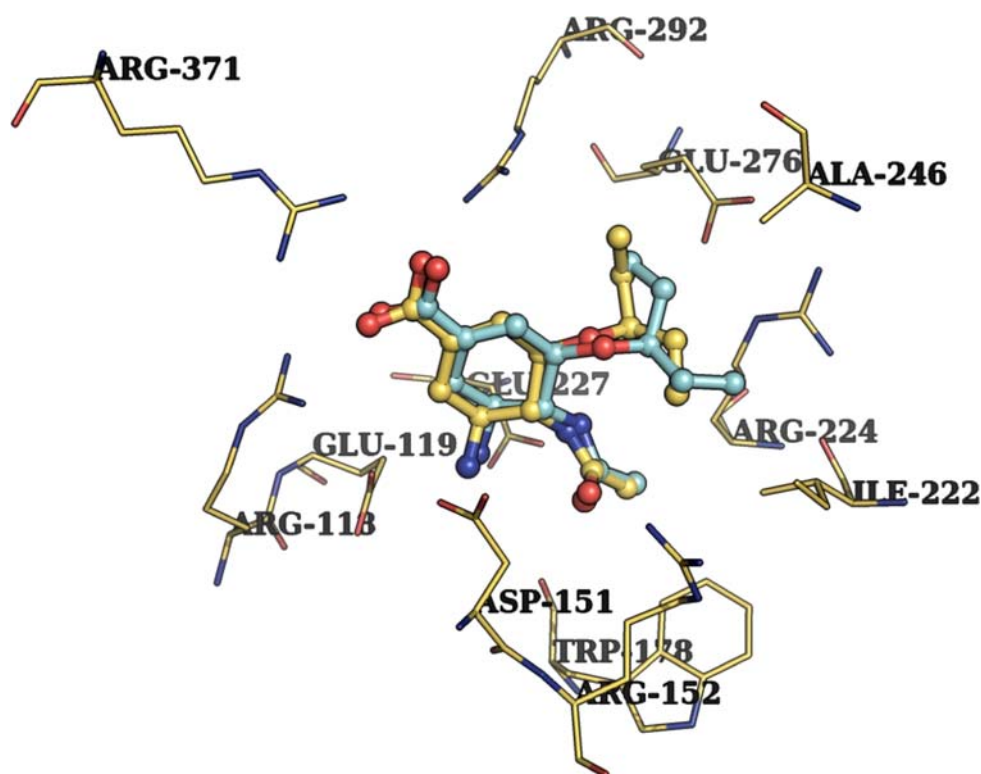
The predicted binding free energies (ΔG s) for all the NIs to NA are listed in Table 3. A linear regression analysis reveals a fair correlation between the experimental inhibitory potencies ($-\log IC_{50}$) of the NIs and the predicted ΔG s to the enzyme, with the r^2 value 0.526 (Eq. 1 and Fig. 3). These correlations demonstrate that the binding conformations and binding models of the NIs to the NA are reasonable:

$$-\log IC_{50} = -7.702 - 0.750 \times \Delta G \quad (n = 126, r^2 = 0.526, F = 137.332, s = 1.027) \quad (1)$$

CoMFA and CoMSIA

It is essential to understand enzyme-inhibitor interactions for designing novel synthetic candidates, however, those interactions are difficult to describe. Structure-based design is focused on the elucidation of enzyme-inhibitor interactions but does not always lead to predictive models. On the other hand, 3D-QSAR models do not necessarily reflect topological features of the enzyme structure. These models are generally

Fig. 1 Comparison of the position of GS4071 calculated by the **AUTODOCK 3.0.3** program (light blue) and that observed in the crystal structure (yellow), for which 12 residues important to interaction are represented in capped-sticks model



constructed using alignment rules, which are not always consistent with the characteristics of the binding conformations. In this study, these two approaches were successfully combined: the complex 3D model of NIs with NA was obtained by **AUTODOCK** and predictive 3D-QSAR models were derived from the alignment conformations extracted directly from the 3D models of the docked complexes.

Table 2 lists all statistical indexes from CoMFA and CoMSIA analysis. For CoMFA, the model was constructed with cross-validated q^2 of 0.813 at five optimum components, the conventional r^2 is 0.942, F value is 373.577 and s is 0.495; For CoMSIA, a six-component model was obtained with q^2 , r^2 , F and s of 0.771, 0.943, 313.213, and 0.493, respectively. The high values of the conventional r^2 relating to different

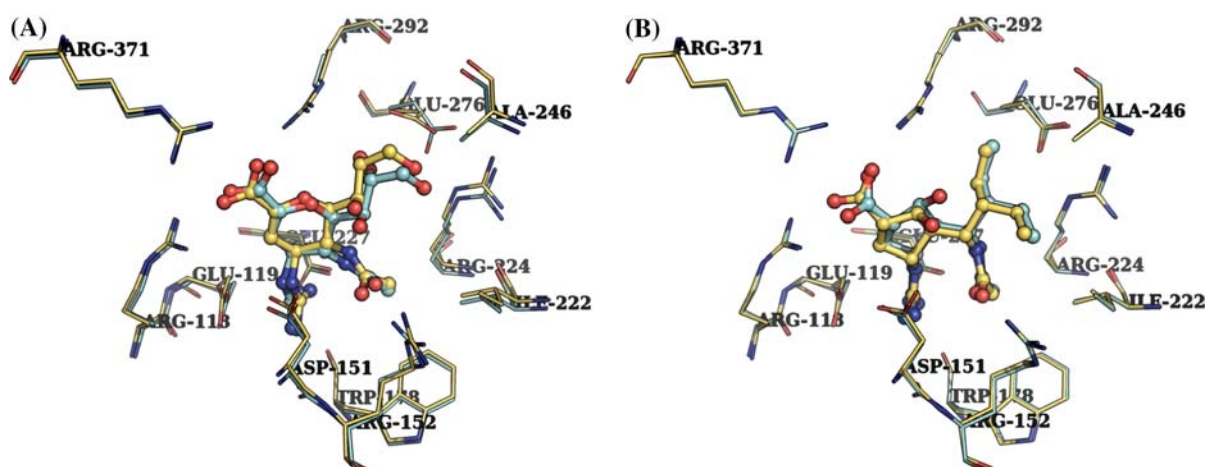
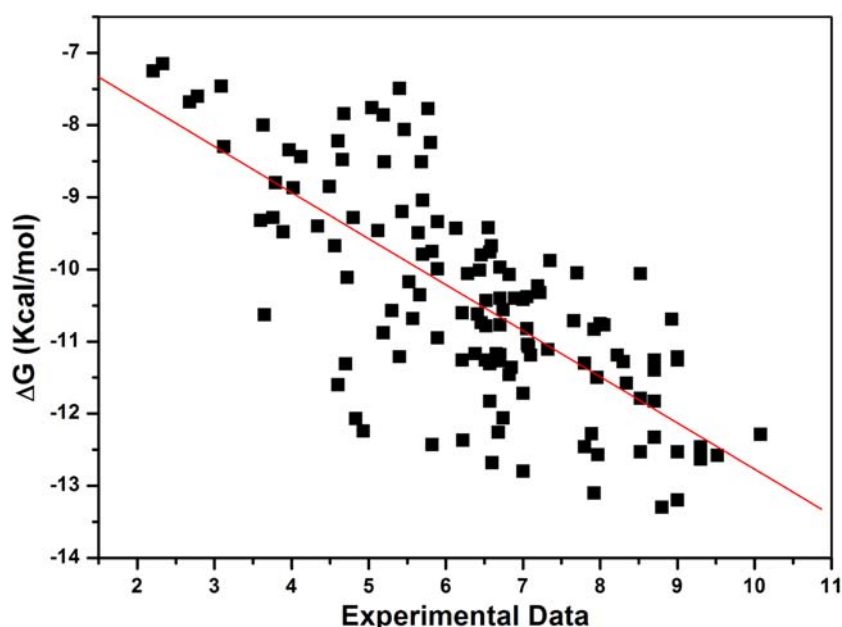


Fig. 2 Overlay of the complex (light blue, determined by docking) with its corresponding X-ray model (yellow): **(A)** Docked complex of Zanamivir-NA (the structure in 2QWK) is aligned with the X-ray complex structure (PDB entry code

1NNC); **(B)** Docked complex of Peramivir-NA (the structure in 2QWK) is aligned with the corresponding X-ray model (PDB entry code 1NNC)

Fig. 3 Correlation of the predicted binding free energies (ΔG , kcal/mol) of NIs with NA to the experimental activities ($-\log IC_{50}$, EA)



descriptor variables (steric, electrostatic, hydrophobic, hydrogen bond donor and acceptor) illustrates that these variables are appropriate to describe the interaction mode of various NIs with NA, as well as the field properties around the inhibitors. With further investigation on the field contribution, we may find that the sum of steric and hydrophobic fields in the CoMSIA model is approximately equal to the steric contribution in the corresponding CoMFA model, which indicates those two models are in consistent with each other.

The binding site of NA has been characterized as highly polar by virtue of the ten charged and only four hydrophobic constitutional amino acids (Fig. 4) [12]. In our results, the distribution of electrostatic field in

CoMFA and the sum of electrostatic and hydrogen bond fields account for around 60% of the total, indicating that polar interaction is indeed essential to the binding of NIs with NA. Table 3 gives the predicted activities of the 108 compounds versus their experimental activities. Taking into account both the statistical indexes and the observed correlation in Fig. 5, we may conclude that the CoMFA and CoMSIA models are highly predictive. To validate the reliability of the constructed models further, the inhibitory activities of the 18 NIs in test set are predicted with these two models, which were also indicated in Fig. 5 with blue triangle. The testing results with correlation coefficients, r^2_{test} , of 0.758 and 0.754 for CoMFA and CoMSIA models, respectively, indicate that the

Table 2 Statistical results of CoMFA, CoMSIA and HQSAR models

| PLS statistics | Model | | | | |
|---------------------|---------------|-------|---------------|-------|-------|
| | CoMFA | | CoMSIA | | HQSAR |
| q^2 | 0.813 | | 0.771 | | 0.768 |
| OPC number | 5 | | 6 | | 6 |
| r^2 | 0.942 | | 0.943 | | 0.811 |
| s | 0.495 | | 0.493 | | 0.725 |
| F | 373.577 | | 313.213 | | |
| Contribution | Steric | 0.411 | Steric | 0.149 | |
| | | | Electrostatic | 0.188 | |
| | Electrostatic | 0.589 | Hydrophobic | 0.274 | |
| | | | Donor | 0.240 | |
| | | | Acceptor | 0.149 | |
| r^2_{test} | 0.908 | | 0.896 | | 0.702 |
| s_{test} | 0.521 | | 0.563 | | 0.907 |

CoMFA and CoMSIA models would be reliable, and should be useful in designing NIs with novel structures.

The visualization of the results of the best CoMFA and CoMSIA models have been performed using the StDev*Coeff mapping option contoured by contribution. Favored and disfavored levels fixed at 80% and 20%, respectively. The CoMFA contour map is shown in Fig. 6, with peramivir and NA residues within 4 Å of the NIs also displayed in aid of visualization. In terms of electrostatics in this model, two highly polar regions are worth noting. There is a large electronegative favored region at the *C1*-substituent of the inhibitor. This area corresponds to the acidic pocket where the carboxylic acid of peramivir has strong ionic interactions with a triad of arginine residues in the NA pocket: Arg118, Arg292 and Arg371. And around the *C4*-guanidino group, the model tends to have prominent electropositive favored contours, which can meet the electrostatic requirement of a negatively charged environment provided by three acidic residues, Glu119, Asp151 and Glu227. Obviously, electrostatic and hydrogen-bonding interactions involved in those two highly polar regions are indispensable to the binding affinity of a potent NI, hence the acidic carboxylate and basic amino or guanidino groups are usually kept constant in the NI structural optimization. Those binding features could also be noted from the alignments of NI structures in the binding site of NA (see Fig. 5), where the structural complementarity with the receptor surface physicochemical properties is rendered clearer.

Besides those two highly polar regions mentioned above, as shown in the CoMFA contour map, blue areas could also be observed at the *C2*- and *C3*-groups, which imply that the NI activity might be improved by placing substituents with positive charges in those places. For example, peramivir is much more potent than its analogue, **74**, in which the interaction with the positive charged hydrogen of *C2*-hydroxyl is absent. Whereas, to keep this interaction and have 2-CH₂ replaced by a secondary amino, Maring et al. found a series of novel pyrrolidines, **95–109**, as potent NIs [12]. But two red areas present near the *C2'*-ethylbutyl group in the contour map could not be explained with the displayed peramivir. Actually, this region corresponds to the polar glycerol side chain of another highly potent NI, zanamivir, hence the model indeed reflects that extra binding affinity can be gained through the electrostatic interactions at this region.

With regard to the steric interaction, the CoMFA model shows a large spacial hindrance region present beside the *C1*-carboxylate group, indicating that bulky substituent at this position would be detrimental to the NI potency. This could explain why the replacement of hydrogen to methyl at R2 group of GS4071 results in a drastic loss of inhibitory activity (**44**, Table 3). Two additional yellow contours beneath the *C1'*-acetamido moiety indicate that subsite containing this group is sterically restricted, which accords well with the properties of a small hydrophobic pocket formed by the surrounding residues Arg152, Trp178 and Ile222.

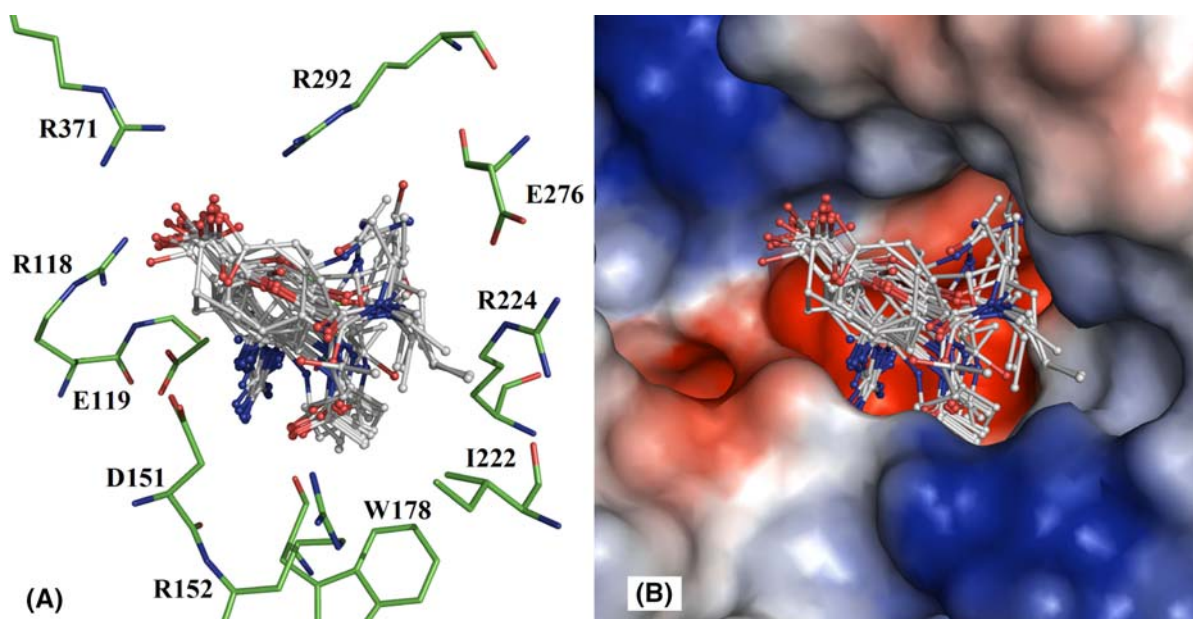


Fig. 4 The alignment of NIs in the binding site of NA. (Note that only compounds mentioned in the discussion section are included). (A) Key residues of the binding site are presented in green sticks. (B) The NA surface was rendered with electrostatic potential

Table 3 In vitro inhibitory activities of 126 NIs and their predicted activities by CoMFA, CoMSIA and HQSAR

| ID | pIC ₅₀ | References | Inhibitor charge (e) | CoMFA | CoMSIA | HQSAR | ΔG (kcal/mol) |
|-----|-------------------|------------|----------------------|-------|--------|-------|---------------|
| 1 | 9 | [25] | 0 | 8.41 | 8.295 | 7.23 | −11.21 |
| 2 | 6.89 | [25] | 0 | 6.247 | 6.547 | 6.87 | −10.4 |
| 3 | 6.7 | [25] | 0 | 7.691 | 6.692 | 7.04 | −9.97 |
| 4 | 6.65 | [25] | 0 | 6.378 | 6.148 | 6.61 | −11.17 |
| 5 | 5.66 | [25] | 0 | 6.258 | 6.203 | 6.54 | −10.35 |
| 6 | 7.66 | [25] | 0 | 7.36 | 7.487 | 7.55 | −10.71 |
| 7 | 7.22 | [25] | 0 | 6.818 | 7.049 | 7.88 | −10.32 |
| 8 | 6.28 | [25] | 0 | 5.957 | 6.562 | 5.88 | −10.06 |
| 9 | 5.2 | [23] | 0 | 5.06 | 5.267 | 6.28 | −8.51 |
| 10 | 5.43 | [23] | 0 | 5.685 | 5.682 | 6.43 | −9.2 |
| 11 | 5.7 | [23] | 0 | 5.982 | 6.092 | 6.83 | −9.79 |
| 12 | 6.52 | [23] | 0 | 6.405 | 6.703 | 6.93 | −10.78 |
| 13 | 6.7 | [23] | 0 | 6.106 | 6.798 | 6.96 | −11.27 |
| 14 | 6.82 | [23] | 0 | 7.104 | 7.073 | 7.02 | −11.46 |
| 15 | 6.57 | [23] | 0 | 6.894 | 7.576 | 7.08 | −11.83 |
| 16 | 6.74 | [23] | 0 | 6.646 | 6.907 | 7.15 | −12.06 |
| 17 | 6.68 | [23] | 0 | 6.861 | 6.48 | 7.21 | −12.26 |
| 18 | 6.22 | [23] | 0 | 6.981 | 6.777 | 7.28 | −12.37 |
| 19 | 6.7 | [23] | 0 | 6.76 | 6.688 | 6.85 | −10.77 |
| 20 | 8 | [23] | 0 | 7.63 | 7.585 | 7.45 | −10.75 |
| 21 | 8.05 | [23] | 0 | 7.751 | 7.531 | 7.32 | −10.77 |
| 22 | 9 | [23] | 0 | 7.891 | 7.841 | 7.35 | −11.26 |
| 23 | 8.52 | [23] | 0 | 8.628 | 8.205 | 7.56 | −11.79 |
| 24 | 9 | [23] | 0 | 8.121 | 8.44 | 7.82 | −12.53 |
| 25 | 7.8 | [23] | 0 | 8.329 | 8.392 | 8.71 | −12.46 |
| 26 | 9 | [23] | 0 | 8.676 | 8.2 | 8.7 | −13.2 |
| 27 | 6.21 | [23] | 0 | 5.604 | 5.502 | 6.18 | −10.6 |
| 28 | 9.52 | [23] | 0 | 9.22 | 8.658 | 7.62 | −12.58 |
| 29 | 7.92 | [23] | 0 | 7.714 | 7.533 | 7.64 | −13.1 |
| 30 | 7.05 | [23] | 0 | 6.767 | 6.884 | 6.47 | −10.38 |
| 31 | 7.8 | [10] | 0 | 8.284 | 8.544 | 7.64 | −11.3 |
| 32 | 7 | [12] | 0 | 6.355 | 7.044 | 7.95 | −10.42 |
| 33* | 8.7 | [12] | 0 | 8.3 | 8.506 | 8.39 | −12.33 |
| 34 | 8.52 | [25] | 0 | 8.094 | 8.541 | 8.45 | −12.53 |
| 35 | 9.3 | [25] | 0 | 8.809 | 8.979 | 8.95 | −12.63 |
| 36 | 9.3 | [25] | 0 | 9.538 | 9.481 | 8.73 | −12.46 |
| 37 | 7 | [25] | 0 | 7.834 | 8.006 | 7.32 | −12.8 |
| 38* | 5.82 | [25] | 0 | 5.873 | 5.107 | 7.02 | −12.43 |
| 39 | 4.6 | [25] | 0 | 4.623 | 4.857 | 4.96 | −11.6 |
| 40 | 6.85 | [25] | 0 | 7.409 | 6.489 | 6.55 | −11.36 |
| 41 | 8.34 | [25] | 0 | 8.553 | 8.369 | 6.95 | −11.58 |
| 42 | 5.64 | [25] | 0 | 5.634 | 6.075 | 5.74 | −9.49 |
| 43* | 8.52 | [25] | 0 | 8.999 | 8.157 | 6.39 | −10.06 |
| 44 | 5.82 | [25] | 0 | 5.384 | 5.583 | 5.97 | −9.75 |
| 45 | 7.19 | [24] | 0 | 7.392 | 7.142 | 6.78 | −10.23 |
| 46* | 6.74 | [24] | 0 | 6.93 | 7.339 | 6.85 | −10.56 |
| 47 | 8.22 | [24] | 0 | 7.657 | 7.441 | 7.1 | −11.19 |
| 48 | 7 | [24] | 0 | 6.904 | 7.613 | 7.11 | −11.72 |
| 49 | 6.7 | [24] | 0 | 6.628 | 6.378 | 7.79 | −11.18 |
| 50 | 7.05 | [24] | 0 | 7.682 | 7.516 | 7.28 | −10.82 |
| 51 | 7.07 | [24] | 0 | 7.908 | 7.872 | 7.3 | −11.07 |
| 52 | 7.92 | [24] | 0 | 8.344 | 7.825 | 7.29 | −10.83 |
| 53 | 6.7 | [24] | 0 | 6.939 | 6.948 | 6.16 | −10.4 |
| 54 | 7.96 | [24] | 0 | 8.162 | 8.421 | 6.71 | −11.5 |
| 55 | 5.57 | [24] | 0 | 4.359 | 4.222 | 5.33 | −10.68 |
| 56* | 5.19 | [24] | 0 | 5.168 | 5.376 | 5.25 | −10.88 |
| 57 | 5.4 | [24] | 0 | 5.891 | 5.425 | 5.8 | −11.21 |
| 58 | 5.89 | [16] | 0 | 6.248 | 6.272 | 6.85 | −9.99 |
| 59 | 7.06 | [16] | 0 | 7.096 | 7.213 | 7.36 | −11.04 |
| 60 | 8.3 | [16] | 0 | 8.706 | 8.29 | 8.25 | −11.28 |
| 61* | 8.93 | [16] | 0 | 8.276 | 7.89 | 7.86 | −10.69 |

Table 3 continued

| ID | pIC ₅₀ | References | Inhibitor charge (e) | CoMFA | CoMSIA | HQSAR | ΔG (kcal/mol) |
|------|-------------------|------------|----------------------|-------|--------|-------|---------------|
| 62 | 8.7 | [16] | 0 | 7.857 | 7.735 | 8.56 | −11.26 |
| 63 | 8.7 | [16] | 0 | 7.816 | 7.774 | 8.21 | −11.4 |
| 64* | 4.93 | [16] | +1 | 4.428 | 4.903 | 5.15 | −12.24 |
| 65 | 6.6 | [16] | 0 | 7.545 | 7.058 | 8.36 | −12.68 |
| 66* | 7.97 | [16] | 0 | 7.088 | 6.662 | 7.95 | −12.57 |
| 67 | 8.8 | [16] | 0 | 8.547 | 8.324 | 8.63 | −13.3 |
| 68 | 8.7 | [16] | 0 | 8.462 | 8.427 | 8.39 | −11.83 |
| 69 | 5.04 | [16] | −1 | 5.26 | 4.982 | 4.84 | −7.76 |
| 70 | 10.08 | [16] | 0 | 9.22 | 9.549 | 9.94 | −12.29 |
| 71 | 4.56 | [16] | 0 | 4.778 | 4.856 | 5.37 | −9.67 |
| 72* | 6.44 | [16] | 0 | 7.161 | 6.738 | 6.48 | −10.01 |
| 73 | 4.83 | [16] | 0 | 4.582 | 4.373 | 5.94 | −12.07 |
| 74 | 7.89 | [16] | 0 | 8.257 | 8.374 | 6.78 | −12.28 |
| 75 | 4.66 | [26] | 0 | 4.537 | 3.838 | 5.11 | −8.48 |
| 76 | 4.6 | [26] | 0 | 4.369 | 4.401 | 4.91 | −8.22 |
| 77 | 4.49 | [26] | 0 | 4.761 | 4.827 | 5.12 | −8.85 |
| 78 | 5.8 | [26] | 0 | 5.452 | 4.726 | 5.03 | −8.24 |
| 79 | 5.4 | [26] | 0 | 5.381 | 4.947 | 5.12 | −7.49 |
| 80* | 4.68 | [26] | 0 | 4.881 | 4.739 | 5.15 | −7.84 |
| 81 | 5.68 | [26] | 0 | 5.206 | 5.552 | 5.31 | −8.51 |
| 82 | 5.7 | [26] | 0 | 5.987 | 6.007 | 5.38 | −9.04 |
| 83 | 4.72 | [26] | 0 | 4.73 | 4.906 | 5.26 | −10.11 |
| 84 | 5.89 | [26] | 0 | 6.053 | 6.907 | 5.54 | −10.95 |
| 85 | 4.34 | [26] | 0 | 4.992 | 3.923 | 5.21 | −9.4 |
| 86 | 5.89 | [26] | 0 | 6.201 | 5.972 | 5.46 | −9.34 |
| 87 | 5.12 | [26] | 0 | 5.313 | 5.354 | 4.82 | −9.46 |
| 88 | 4.8 | [26] | 0 | 4.587 | 4.966 | 4.73 | −9.28 |
| 89* | 4.02 | [26] | 0 | 4.621 | 4.438 | 4.33 | −8.87 |
| 90* | 6.55 | [26] | 0 | 6.181 | 5.78 | 4.9 | −9.42 |
| 91 | 3.89 | [26] | 0 | 3.835 | 4.208 | 4.68 | −9.48 |
| 92 | 6.13 | [16] | 0 | 5.121 | 4.86 | 4.71 | −9.43 |
| 93 | 3.79 | [16] | 0 | 4.917 | 4.871 | 4.05 | −8.8 |
| 94 | 3.76 | [16] | 0 | 4.773 | 4.853 | 4.67 | −9.28 |
| 95 | 6.21 | [12] | 0 | 5.675 | 5.809 | 6.42 | −11.26 |
| 96 | 6.46 | [12] | 0 | 6.298 | 6.241 | 6.95 | −9.8 |
| 97 | 6.51 | [12] | 0 | 6.842 | 6.642 | 6.84 | −11.26 |
| 98 | 6.46 | [12] | 0 | 6.437 | 6.639 | 7.41 | −10.74 |
| 99* | 7.35 | [12] | 0 | 7.187 | 7.377 | 7.43 | −9.88 |
| 100 | 6.57 | [12] | 0 | 6.204 | 5.873 | 6.78 | −11.31 |
| 101 | 6.57 | [12] | 0 | 6.88 | 6.847 | 6.68 | −9.76 |
| 102* | 6.59 | [12] | 0 | 7.554 | 7.34 | 5.79 | −9.67 |
| 103 | 6.82 | [12] | 0 | 6.63 | 6.769 | 6.41 | −10.07 |
| 104 | 6.52 | [12] | 0 | 6.242 | 6.536 | 6.42 | −10.43 |
| 105 | 7.7 | [12] | 0 | 7.57 | 7.474 | 6.41 | −10.05 |
| 106 | 6.41 | [12] | 0 | 6.563 | 6.821 | 6.91 | −10.62 |
| 107 | 6.38 | [12] | 0 | 6.239 | 6.419 | 6.63 | −11.17 |
| 108* | 7.1 | [12] | 0 | 6.358 | 6.711 | 6.32 | −11.19 |
| 109 | 5.52 | [12] | 0 | 6.757 | 6.501 | 6.32 | −10.17 |
| 110* | 5.77 | [16] | 0 | 5.205 | 5.857 | 4.58 | −7.77 |
| 111 | 2.33 | [16] | 0 | 2.018 | 2.322 | 1.55 | −7.15 |
| 112 | 2.68 | [16] | −1 | 2.9 | 2.804 | 2.68 | −7.68 |
| 113 | 4.12 | [16] | 0 | 3.296 | 4.02 | 4.71 | −8.44 |
| 114 | 2.78 | [16] | −1 | 3.327 | 2.916 | 1.89 | −7.6 |
| 115* | 3.09 | [16] | 0 | 3.011 | 3.5 | 4.7 | −7.46 |
| 116 | 3.97 | [16] | 0 | 5.041 | 4.453 | 4.91 | −8.34 |
| 117 | 5.46 | [16] | 0 | 5.674 | 4.965 | 5.6 | −8.06 |
| 118 | 5.19 | [16] | 0 | 4.553 | 4.467 | 4.86 | −7.86 |
| 119 | 3.63 | [16] | 0 | 4.02 | 4.338 | 3.56 | −8 |
| 120 | 3.6 | [22] | 0 | 3.387 | 3.462 | 3.13 | −9.32 |
| 121 | 4.7 | [22] | 0 | 4.381 | 4.577 | 3.56 | −11.31 |

Table 3 continued

| ID | pIC ₅₀ | References | Inhibitor charge (e) | CoMFA | CoMSIA | HQSAR | ΔG (kcal/mol) |
|------|-------------------|------------|----------------------|-------|--------|-------|---------------|
| 122 | 2.2 | [22] | 0 | 3.398 | 2.974 | 3.61 | −7.25 |
| 123 | 3.12 | [22] | −1 | 2.53 | 3.124 | 3.99 | −8.3 |
| 124 | 5.3 | [22] | 0 | 5.151 | 5.664 | 5.55 | −10.57 |
| 125* | 3.65 | [22] | −1 | 3.657 | 4.148 | 4.17 | −10.63 |
| 126 | 7.32 | [22] | −1 | 6.486 | 7.355 | 6.22 | −11.11 |

* Inhibitors in testing set

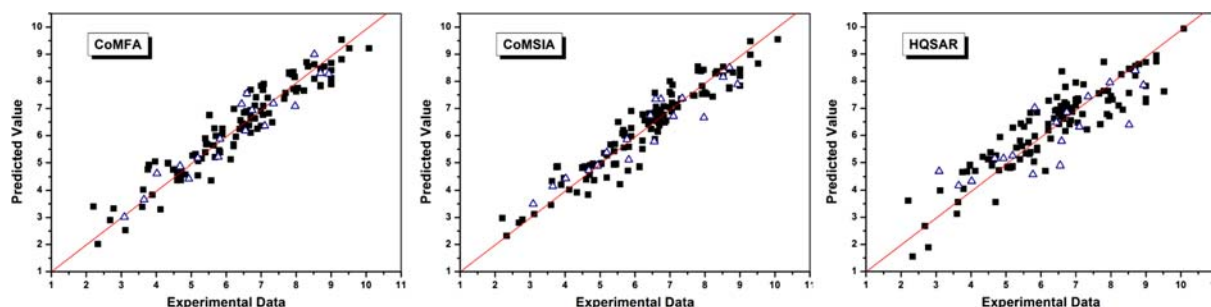


Fig. 5 Plot of the observed activity (pIC₅₀) versus the predicted activity (the data for the training set are marked as •, and testing set as Δ) derived from CoMFA, CoMSIA and HQSAR models, respectively

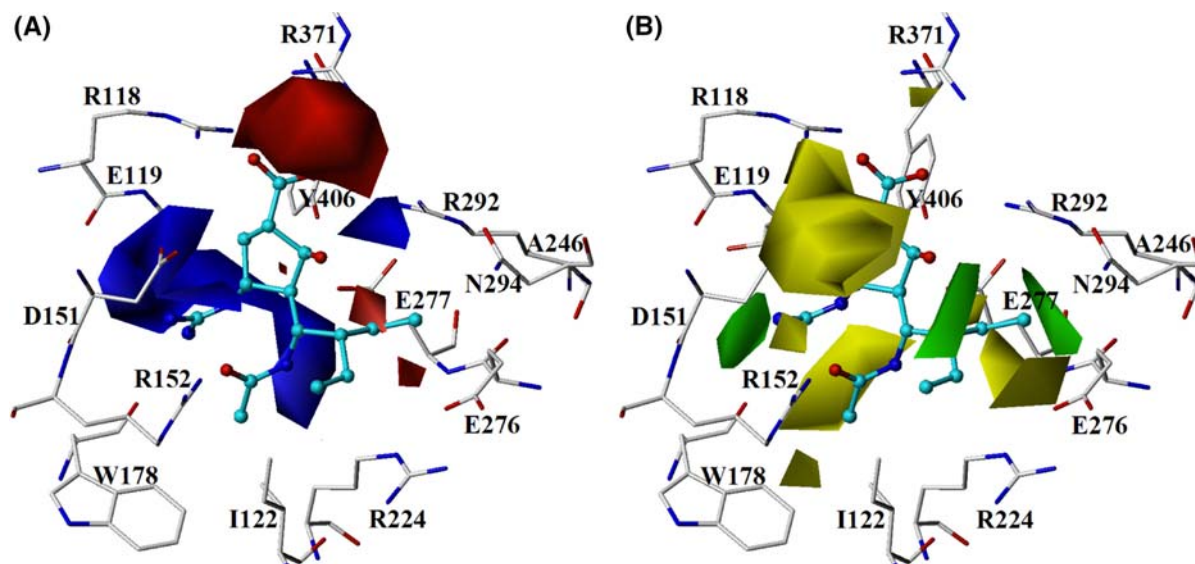


Fig. 6 CoMFA (A) electrostatics and (B) steric contour maps as compared with the topology of the binding site of NA. Residues within 4 Å of inhibitor are shown. The residues are represented as sticks, and inhibitors are shown in ball-and-stick. Sterical

favored areas are in green; sterical unfavored areas are in yellow. Positive charge-favored areas are in blue; positive charge unfavored areas are in red

Therefore, bulkier group replacement at this or equivalent position would lead to a drop of NI activity. As we may find from Table 3, **38** is less potent than its acetamido counterpart **33**. Similarly, **115** is weaker as compared to **110**.

The CoMFA model also shows two long narrow regions of steric favored contours near the two iden-

tical branched ethyls of the C2'-group of peramivir. In the CoMSIA model, these two regions are predicted to be hydrophobic favorable (Fig. 7A). Both of these two models suggest that reducing the length of alkyl chains at this or equivalent position will decrease the bioactivity of NIs, which make account for the potency order of **60** > **59** > **58**. In addition, considering the size and

shape of these two steric contours, we might also speculate that ethyl would be optimal for binding, which can be verified by the fact that **61** is more potent than **60**, though two ethyl substituents are smaller in space than the two propyls. It is interesting to find a steric hindrance contour between the two branched ethyls while this region is predicted as hydrophobic unfavorable in the CoMSIA model (Fig. 7A), contrasting clearly with side regions. These contours explain why **10–18**, with non-branching alkyl chain at the C2' position, have relatively weak inhibitory activities. Hydrophobic unfavorable contours are also seen overlapping with the C1-carboxylate and C1'-acetamido moieties, which further stress polar functional groups at these sites are necessary to NIs binding to the active site of NA.

For the hydrogen bond donor and acceptor maps of CoMSIA, as shown in Fig. 7B, the cyan and orange contours around the binding site indicate that hydrogen bond donors and acceptors are advantageous in these areas, respectively. As expected, this is in agreement with the inhibitor–enzyme binding model. In the NA–NI complexes, Glu119 and Asp151 form hydrogen bonds as hydrogen bond acceptors with the guanidino or amino moiety of the NI structure, while Arg118, Arg292, and Arg371 form hydrogen bonds with the carboxylate moiety as hydrogen bond donors. The consistency between the CoMSIA field distributions and the structural topological properties of the NI binding site demonstrates the reasonability of the CoMSIA result. Steric and electrostatic contour map

derived from CoMSIA shows general similarity as compared to that from CoMFA, which is not presented for brevity.

The binding mode can be readily observed by combining the CoMFA and CoMSIA contour maps with the 3D structural topology of the NI binding site. Not only the field property coincide perfectly with the environmental characteristics of the binding pocket but also some indications for further design and structural modification of NIs could be found. More important, the colored polyhedrons of CoMFA and CoMSIA located in the cavity of the binding pocket are direct indexes for the chemical feature and spacial magnitude of the substituents in the process of new compound synthesis.

HQSAR

The statistic results of the HQSAR analysis on the training set are listed in Table 2. The cross-validated PLS analysis yielded a q^2 of 0.768 with six optimal components, and conventional PLS analysis gave a r^2 value of 0.811 with a s of 0.725. The hologram that gives the lowest standard error has a length of 61. The correlation between the experimental data and HQSAR predicted values are plotted in Fig. 5. All those results suggest that the HQSAR model was successfully built, and this model should be good enough to rank molecular activities for further drug discovery.

The HQSAR module implemented in SYBYL uses a color code to discriminate the main atomic contri-

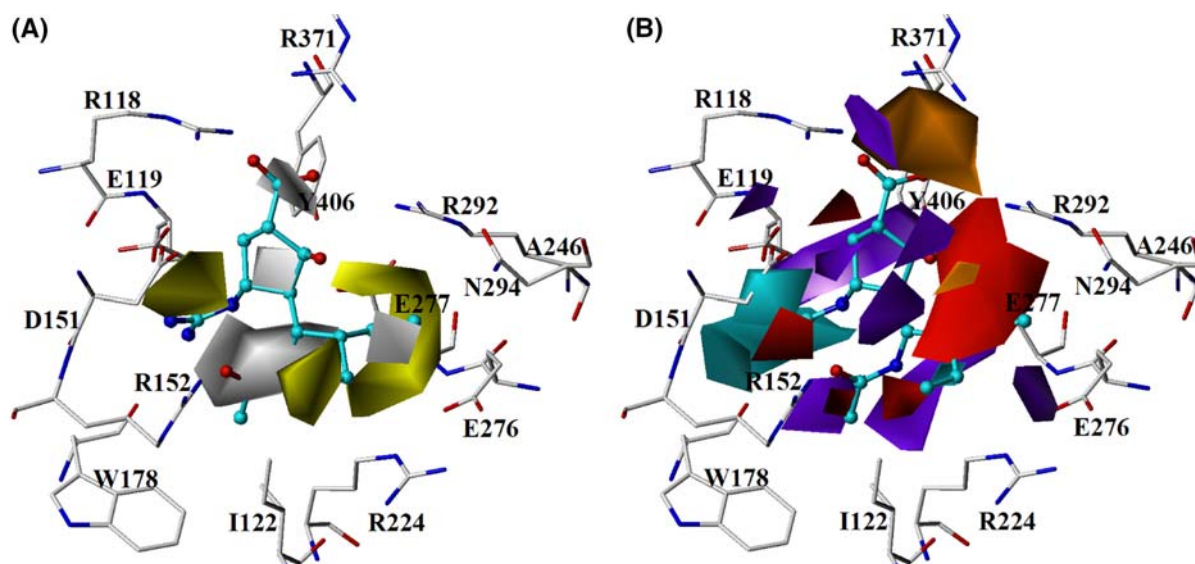


Fig. 7 CoMSIA contour maps for: (A) hydrophobicity; (B) hydrogen bond donor and acceptor. Hydrophobic favored areas are colored in yellow and hydrophobic unfavored areas are in gray. The cyan-colored contours show areas where a hydrogen

bond donor is favored, and purple-colored contours indicate the opposite. Hydrogen bond acceptors are favored in areas indicated by orange contours, whereas red contours are areas where hydrogen bond acceptors are disfavored

butions to activity. Figure 8 depicts the individual atomic contributions to the activity of peramivir (**70**). For comparison, other four compounds (**71**, **72**, **73**, **74**) with the cyclopentane scaffold were also included in this figure. The conformations presented here were the same as those used in the CoMFA and CoMSIA analyses. The colors at the red end of the spectrum (red, red orange, and orange) reflect poor (or negative) contributions, while colors at the green end (yellow, green blue, and green) indicate favorable (positive) contributions. Atoms with intermediate contributions are colored in white. As shown, the 5-ring core—(*1S*, *2S*, *3R*, *4R*)-cyclopentane of **70** is suggested to be strongly related to the biological activity of this compound; cyclopentane rings with different chirality: i.e. (*1R*, *3S*, *4R*) for that of **71**, (*1S*, *3R*, *4R*) for **72** and **73**, and (*1R*, *3R*, *4R*) for **74**, are partially yellow-colored, indicating their less contribution to the inhibitory potency. Therefore, we may infer that **70** should present higher affinity to NA as compared to other cyclopentane NIs, and this could be confirmed by the experimental data listed in Table 3.

As we know, the success of 3D-QSAR model such as CoMFA and CoMSIA is strongly dependent on the conformation of the molecule used in the analysis, and on the orientation of the molecule relative to the other molecules in 3D space. From Fig. 8 and the docking simulation we may notice that cyclopentane ring of NI with other chirality than (*1S*, *2S*, *3R*, *4R*) generally causes these compounds to take different conformation and orientation in the binding site of NA. The distinct differences in structure may result in difficulty of aligning these compounds with **70** in the most prevalent way, thus the modification clues derived from the CoMFA and CoMSIA analyses can not be applied to such structures. Whereas, HQSAR technique does not require the 3D information of these compounds, so it could serve as a complementary study to our 3D QSAR analyses. For example, from the molecular hologram presentations shown in Fig. 8, we may clearly find that (*1S*, *2S*, *3R*, *4R*)-

cyclopentane is a favorable configuration as for NI structural optimization, which, especially the importance of the chirality at *C1* and *C2* atoms, can not be straightforwardly interpreted by CoMFA and CoMSIA analyses.

Validation of the 3D-QSAR models

QSAR models tend to perform better on data on which the model was constructed than on new data. For CoMFA and CoMSIA models, there is even more optimism because the apparent performance is substantially limited by the molecule structure type of the training set and the alignment rule. Since one of our aims is to develop more universal and robust models for designing novel compounds in inhibiting NA, we carried out more stringent validation on an external test set with 18 compounds covering observable structural diversity, to determine the generalizability of current models. The predicted bioactivities for these NIs based on the above models were depicted in Fig. 5 (labeled as triangle). The PLS analysis on the testing set released r^2_{test} values of 0.908, 0.896 and 0.702; and s_{test} values of 0.521, 0.563, and 0.907 for CoMFA, CoMSIA, and HQSAR, respectively, demonstrating again that the CoMFA, CoMSIA, and HQSAR models are fairly reliable and have rather strong predictive ability, and should be useful in designing new NIs.

Except for the concern of the generalizability, the high internal validation performance of our QSAR models might be a result of chance correlation. To address this problem, these three models were validated by applying the Y-randomization of response test (in this work, the experimental activity values) [50, 51]. It consists of repeating the calculation procedure several times after shuffling the Y vector randomly. If all models obtained by the Y-randomization test have relatively high values for both q^2 and r^2 statistics, this is due to a chance correlation and implies that the current modeling method cannot lead to an

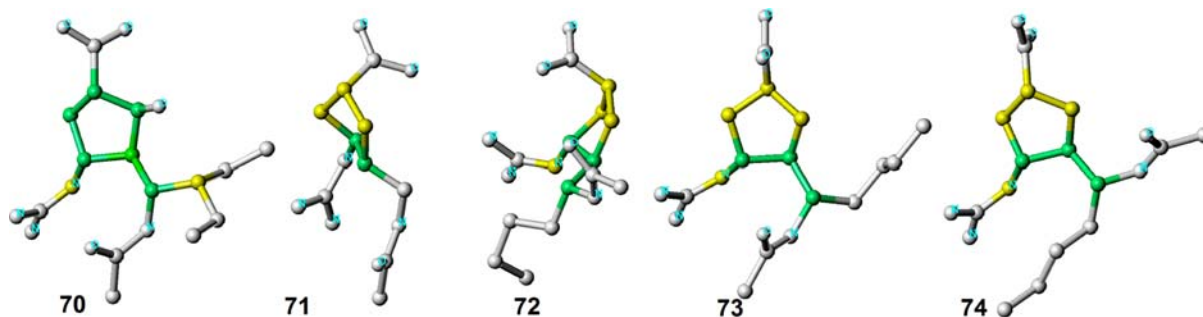


Fig. 8 The individual atomic contributions to the activity of some NIs (**70–74**, Table 1). The colors at red end of the spectrum reflect poor contributions, while colors at green end indicate favorable contributions

Table 4 Y-Randomization results of CoMFA, CoMSIA and HQSAR models

| Iteration | CoMFA | | CoMSIA | | HQSAR | |
|-----------|---------|---------|---------|---------|---------|---------|
| | r^2 | q^2 | r^2 | q^2 | r^2 | q^2 |
| 1 | 0.06087 | 0.02613 | 0.03359 | 0.00700 | 0.00004 | 0.27470 |
| 2 | 0.00050 | 0.06217 | 0.00011 | 0.05587 | 0.00675 | 0.01233 |
| 3 | 0.00287 | 0.07728 | 0.01106 | 0.00452 | 0.00673 | 0.01450 |
| 4 | 0.00309 | 0.06925 | 0.01152 | 0.00495 | 0.03866 | 0.00979 |
| 5 | 0.02495 | 0.00113 | 0.00021 | 0.37390 | 0.00426 | 0.04193 |
| 6 | 0.00080 | 0.27630 | 0.00003 | 0.43000 | 0.00200 | 0.15350 |
| 7 | 0.00424 | 0.04414 | 0.00728 | 0.01734 | 0.03321 | 0.00651 |
| 8 | 0.02441 | 0.00040 | 0.01375 | 0.00040 | 0.02045 | 0.00008 |
| 9 | 0.00199 | 0.09985 | 0.01244 | 0.00248 | 0.03296 | 0.00430 |
| 10 | 0.00795 | 0.01594 | 0.00122 | 0.2014 | 0.01232 | 0.00541 |

acceptable model using the available data set. This was not the case for the data set and methodology used in this work. Ten random shuffles of the Y vector were performed and the results are shown in Table 4. The low q^2 and r^2 values show that the good results in our original models are not due to a chance correlation or structural dependency of the training set.

Conclusions

In this study, 108 NIs were chosen to build highly predictive CoMFA, CoMSIA and HQSAR models. The statistics results on both the training and an external test set of 18 NIs suggest these models are well constructed; CoMFA/CoMSIA field maps and HQSAR fragment contribution plot provide us useful information for the NI structure optimization. Therefore, these 3D and 2D QSAR models are expected to be fast and effective tools to speed up the development anti-AIV drugs with novel structures, to reduce the impact of the likely rapid global spread of influenza pandemics.

Acknowledgements We thank Professor Arthur J. Olson for his kindness in offering us the **AUTODOCK 3.0.3** program. We gratefully acknowledge financial support from National Natural Science Foundation of China (Grants 20102007, 29725203, and 20072042), the State Key Program of Basic Research of China (Grant 2004CB518901), and Shanghai Science and Technology Commission (03DZ19228, 05JC14092 and 05QMX1464).

References

- de Jong MD, Hien TT (2006) *J Clin Virol* 35:2
- Ferguson NM, Cummings DA, Cauchemez S, Fraser C, Riley S, Meeyai A, Iamsirithaworn S, Burke DS (2005) *Nature* 437:209
- Laver G (2005) *Nature* 434:821
- Gani R, Hughes H, Fleming D, Griffin T, Medlock J, Leach S (2005) *Emerg Infect Dis*. Available from <http://www.cdc.gov/ncidod/EID/vol11no09/04-1344.htm>
- Colman PM (1994) *Protein Sci* 3:1687
- Young D, Fowler C, Bush K (2001) *Philos Trans R Soc Lond B Biol Sci* 356:1905
- Lew W, Chen X, Kim CU (2000) *Curr Med Chem* 7:663
- Kim CU, Chen X, Mendel DB (1999) *Antivir Chem Chemother* 10:141
- von Itzstein M, Wu WY, Kok GB, Pegg MS, Dyason JC, Jin B, Van Phan T, Smythe ML, White HF, Oliver SW, et al (1993) *Nature* 363:418
- Kim CU, Lew W, Williams MA, Liu H, Zhang L, Swaminathan S, Bischofberger N, Chen MS, Mendel DB, Tai CY, Laver WG, Stevens RC (1997) *J Am Chem Soc* 119:681
- Le QM, Kiso M, Someya K, Sakai YT, Nguyen TH, Nguyen KH, Pham ND, Nguyen HH, Yamada S, Muramoto Y, Horimoto T, Takada A, Goto H, Suzuki T, Suzuki Y, Kawaoka Y (2005) *Nature* 437:1108
- Maring CJ, Stoll VS, Zhao C, Sun M, Krueger AC, Stewart KD, Madigan DL, Kati WM, Xu Y, Carrick RJ, Montgomery DA, Kempf-Grote A, Marsh KC, Molla A, Steffy KR, Sham HL, Laver WG, Gu YG, Kempf DJ, Kohlbrenner WE (2005) *J Med Chem* 48:3980
- Stoll V, Stewart KD, Maring CJ, Muchmore S, Giranda V, Gu YG, Wang G, Chen Y, Sun M, Zhao C, Kennedy AL, Madigan DL, Xu Y, Saldivar A, Kati W, Laver G, Sowin T, Sham HL, Greer J, Kempf D (2003) *Biochemistry* 42:718
- Chand P, Bantia S, Kotian PL, El-Kattan Y, Lin TH, Babu YS (2005) *Bioorg Med Chem* 13:4071
- Verma RP, Hansch C (2006) *Bioorg Med Chem* 14:982
- Yi X, Guo Z, Chu FM (2003) *Bioorg Med Chem* 11:1465
- Klebe G, Abraham U, Mietzner T (1994) *J Med Chem* 37:4130
- Doucette KE, Aoki FY (2001) *Expert Opin Pharmacother* 2:1671
- Chand P, Babu YS, Bantia S, Rowland S, Dehghani A, Kotian PL, Hutchison TL, Ali S, Brouillette W, El-Kattan Y, Lin TH (2004) *J Med Chem* 47:1919
- Cramer RD, David EP, Jeffrey DB (1988) *J Am Chem Soc* 110:5959
- Tong W, Lowis DR, Perkins R, Chen Y, Welsh WJ, Goddette DW, Heritage TW, Sheehan DM (1998) *J Chem Inf Comput Sci* 38:669
- Atigadda VR, Brouillette WJ, Duarte F, Ali SM, Babu YS, Bantia S, Chand P, Chu N, Montgomery JA, Walsh DA,

- Sudbeck EA, Finley J, Luo M, Air GM, Laver GW (1999) *J Med Chem* 42:2332
23. Kim CU, Lew W, Williams MA, Wu H, Zhang L, Chen X, Escarpe PA, Mendel DB, Laver WG, Stevens RC (1998) *J Med Chem* 41:2451
24. Lew W, Wu H, Mendel DB, Escarpe PA, Chen X, Laver WG, Graves BJ, Kim CU (1998) *Bioorg Med Chem Lett* 8:3321
25. Williams MA, Lew W, Mendel DB, Tai CY, Escarpe PA, Laver WG, Stevens RC, Kim CU (1997) *Bioorganic Med Chem Lett* 7:1837
26. Wang GT, Chen Y, Wang S, Gentles R, Sowin T, Kati W, Muchmore S, Giranda V, Stewart K, Sham H, Kempf D, Laver WG (2001) *J Med Chem* 44:1192
27. Cheng Y, Prusoff WH (1973) *Biochem Pharmacol* 22:3099
28. Smith PW, Sollis SL, Howes PD, Cherry PC, Starkey ID, Copley KN, Weston H, Scicinski J, Merritt A, Whittington A, Wyatt P, Taylor N, Green D, Bethell R, Madar S, Fenton RJ, Morley PJ, Pateman T, Beresford A (1998) *J Med Chem* 41:787
29. Varghese JN, Smith PW, Sollis SL, Blick TJ, Sahasrabudhe A, McKimm-Breschkin JL, Colman PM (1998) *Structure* 6:735
30. Weiner SJ, Kollman PA, Case DA, Singh C, Ghio G, Alagona S, Profeta P, Weiner P (1984) *J Am Chem Soc* 106:765
31. SYBYL (2000) Version68; Tripos Associates; St Louis, MO
32. Clark M, Cramer R, Vannodenbosch N (1989) Validation of the general-purpose tripos 5.2 force-field. *J Comput Chem* 10:982
33. Gasteiger J, Marsili M (1981) *Org Magn Reson* 15:353
34. Marsili M, Gasteiger J (1980) *Croat Chem Acta* 53:601
35. Gasteiger J, Marsili M (1980) *Tetrahedron* 36:3219
36. Dewar MJS, Zoebisch EG, Healy EF, Stewart JJP (1985) *J Am Chem Soc* 107:3902
37. Morris GM, Goodsell DS, Huey R, Hart WE, Halliday S, Belew R, Olson AJ (1999) AUTODOCK Version 3.0.3. The Scripps Research Institute, Molecular Graphics Laboratory, Department of Molecular Biology
38. Morris GM, Goodsell DS, Halliday RS, Huey R, Hart WE, Belew RK, Olson AJ (1998) *J Comput Chem* 19:1639
39. Solis FJ, Wets RJ-B (1981) *Math Operat Res* 6:19
40. Weiner SJ, Kollman PA, Nguyen DT, Case DA (1986) *J Comput Chem* 7:230
41. Mehler EL, Solmajer T (1991) *Protein Eng* 4:903
42. Bush BL, Nachbar RB Jr (1993) *J Comput Aided Mol Des* 7:587
43. Viswanadhan VN, Ghose AK, Revankar GR, Robins RK (1989) *J Chem Inf Comput Sci* 29:163
44. Ghose AK, Crippen GM (1986) *J Comput Chem* 7:565
45. Tong W, Lewis DR, Perkins R, Chen Y, Welsh WJ, Goddette DW, Heritage TW, Sheehan DM (1998) *J Chem Inf Model* 38:669
46. Babu YS, Chand P, Bantia S, Kotian P, Dehghani A, El-Kattan Y, Lin TH, Hutchison TL, Elliott AJ, Parker CD, Ananth SL, Horn LL, Laver GW, Montgomery JA (2000) *J Med Chem* 43:3482
47. Finley JB, Atigadda VR, Duarte F, Zhao JJ, Brouillette WJ, Air GM, Luo M (1999) *J Mol Biol* 293:1107
48. Varghese JN, Epa VC, Colman PM (1995) *Protein Sci* 4:1081
49. Smith BJ, McKimm-Breshkin JL, McDonald M, Fernley RT, Varghese JN, Colman PM (2002) Structural studies of the resistance of influenza virus neuraminidase to inhibitors, vol 45, pp 2207–2212
50. Tropsha A, Gramatica P, Gombar VK (2003) *Quant Struct-act Relatsh* 22:1
51. Wold S, Eriksson L (1995) In: van de Waterbeemd H (ed) *Chemometrics methods in molecular design*. VCH, Weinheim, pp 309–318