

# Docking and 3D QSAR study of thiourea analogs as potent inhibitors of influenza virus neuraminidase

Jiaying Sun · Shaoxi Cai · Hu Mei · Jian Li · Ning Yan · Yuanqiang Wang

Received: 10 June 2009 / Accepted: 21 January 2010 / Published online: 7 March 2010  
© Springer-Verlag 2010

**Abstract** Surflex-Dock was applied to study interactions between 30 thiourea analogs and neuraminidase (NA). The docking results showed that hydrogen bonding and electrostatic interactions were highly correlated with the activities of neuraminidase inhibitors (NIs), followed by hydrophobic and steric factors. Moreover, there was a strong correlation between the predicted binding affinity (total score) and experimental  $pIC_{50}$  (correlation coefficient  $r=0.870$ ;  $P<0.0001$ ). A three dimensional holographic vector of atomic interaction field (3D-HoVAIF) was employed to construct a QSAR model. The  $r^2$ ,  $q^2$  and  $r^2_{test}$  values of the optimal QSAR model were 0.849, 0.724 and 0.689, respectively. From the QSAR model, it could be seen that electrostatic, hydrophobic and steric interactions were closely related to inhibitory activity, which was consistent with the docking results. Based on the docking and QSAR results, five new compounds with high predicted activities were designed.

**Keywords** Thiourea analogs · Neuraminidase inhibitors · Docking · QSAR · HoVAIF

## Introduction

Neuraminidase (NA) is one of two glycoproteins on the surface of influenza virus [1]. NA is responsible for viral release from infected cells and viral transport through the mucus in the respiratory tract [1]. NA has been recognized as a potential target for the of control influenza virus [2]. Neuraminidase inhibitors (NIs) form key components of pandemic preparedness plans as treatment and prophylaxis could reduce virus transmission [3]. Sialic acid analogs were the first NIs reported. Based on the structure of sialic acid, different series of NIs were prepared, such as cyclohexenes [4], benzoic acids [5], pyrrolidine derivates [6, 7] and so on. During the past decade, thiourea derivatives have been reported as being effective against HIV and to have bactericidal action [8]. However, few studies have evaluated substituted acyl(thio)ureas and 2H-1,2,4-thiadiazolo [2,3- $\alpha$ ] pyrimidines for their antiviral activities [8]. In 2006, a new class of substituted acyl (thio)urea and 2H-1,2,4-thiadiazolo [2,3- $\alpha$ ] pyrimidine derivatives with highly specific anti-influenza virus activities in cell culture were prepared by Sun et al. [8]. Their in vitro inhibitory activities against influenza neuraminidase (H1N1) were also investigated and found to correlate well with their antiviral efficacy in cell culture.

In order to further develop potent antiviral agents, many researchers have carried out quantitative structure activity relationship (QSAR) of NIs. For example, in 2003 Yi et al. [9] established CoMSIA (comparative molecular similarity indices analysis) models of 37 NIs including cyclohexene, cyclopentane, pyrrolidine and benzoic acid derivatives. Conformation of 37 NIs were derived from AUTODOCK 3.0. The  $q^2$  values of the four QSAR models obtained were 0.701 (including steric and electrostatic), 0.562 (consisting of steric, electrostatic and hydrophobic), 0.704 (including

J. Sun (✉) · S. Cai · H. Mei (✉) · J. Li · N. Yan  
College of Bioengineering, Chongqing University,  
Chongqing 400044, China  
e-mail: wy7472@126.com  
e-mail: meihu@cqu.edu.cn

Y. Wang  
Colleges of Chemistry and Bioengineering,  
Chongqing Institute of Technology,  
Chongqing 400050, China

steric, electrostatic and hydrogen bond) and 0.651 (including steric, electrostatic, hydrophobic and hydrogen bond), respectively. In 2006, Rajeshwar et al. [2] investigated QSAR of benzoic acid, carbocyclic ring, cyclopentane, isoquinoline derivatives using physicochemical and electronic parameters, and obtained 17 QSAR models with good statistical results. QSAR of 46 NIs including cyclohexene, cyclopentane, pyrrolidine and benzoic acid derivatives was researched by heuristic method (HM) and radial basis function network (RBFNN) was reported by Lü et al. [10]. The linear HM model indicated that hydrophobic and hydrogen bonding interactions between molecules played important roles in regulating the activities of NIs, while at the same time the nonlinear RBFNN models proved to have higher predictive ability than that of the linear model. Nair et al. [11] studied QSAR of 40 thiourea analogs using spatial, topological, electronic, thermodynamic and E-state indices. The genetic algorithm based genetic function approximation method of variable selection was used to generate QSAR models. The results indicated that the atom type log P and shadow indices made enormous contributions to NI activities.

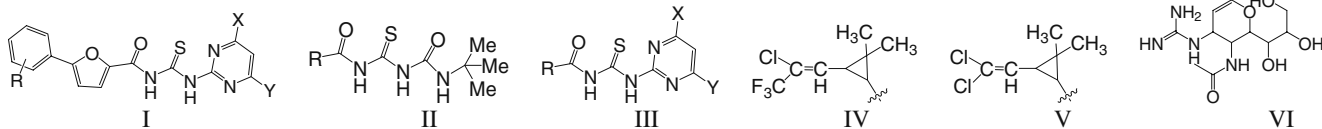
In the current paper, the aim was to construct QSAR models of NIs and to study NI–NA interactions, with a view to providing information and reference data for drug design and optimization of lead compounds. To this end, Surflex-Dock was applied to study the interactions between 40 thiourea derivatives [11] and NA. Then, QSARs were performed using a three dimensional holographic vector of atomic interaction field (3D-HoVAIF), which was proposed based on a 2D structural descriptor developed by Liu et al. [12] in our laboratory. Based on docking and QSAR results, new compounds with high inhibitory activities were designed.

## Methods and materials

### Docking

Surflex-Dock was applied to study molecular docking. The crystal structure of NA with zanamivir GG167 (Scheme 1: series VI) was retrieved from the RCSB Protein Data Bank (PDB entry code: 1a4g) [13]. Surflex-Dock uses an

empirical scoring function and a patented search engine to dock ligands into a protein's binding site [14]. A Protomol, which was used to guide molecular docking, is a computational representation of the intended binding site to which putative ligands are aligned. Protomols can be produced by one of three routes [15]: (1) automatic: Surflex-Dock finds the largest cavity in the receptor protein; (2) ligand-based: a ligand in the same coordinate space as the receptor; (3) residue-based: specified residues in the receptor. Thus, a Protomol can be generated automatically or defined based on a cognate ligand or known active site. In the current paper, a Protomol was generated automatically. Two parameters determining the extent of the Protomol—a threshold parameter of 0.31 and a bloat parameter of 1 Å—were established. All the water molecules in NA 1a4g (receptor) were deleted, and hydrogen atoms were added to 1a4g [16, 17]. The protein structure was utilized in subsequent docking experiments without energy minimization. In addition, treatment of docking small molecules (ligands) was as follows: energy minimization method: Powell; force field: tripos; charge: MMFF94; max iterations: 1,000; termination: 0.001 kcal/(mol\*Å); root mean square (RMS) displacement: 0.001 Å; other parameters: treated by default. Surflex-Dock scores (total scores) are expressed in  $-\log_{10}(K_d)$  units to represent binding affinities. In the docking procedure, ten binding poses per ligand were obtained, and the binding pose with the highest total score was taken into consideration for ligand–receptor interactions. Docking results were validated by finding the correlation coefficient between experimental  $pIC_{50}$  and total scores as well as the RMSD (root mean squared deviation) between the docking position calculated for zanamivir and that observed in the crystal structure. The strengths of individual scoring functions combine to produce a consensus that is more robust and accurate than any single function for evaluating ligand–receptor interactions. Thus, the CScore (consensus score) [18] was used for ranking the affinity of ligands bound to the active site of a receptor. CScore integrates a number of popular scoring functions and provides several functions: D\_Score [19], PMF (potential of mean force)\_Score [20], G\_Score [21] and CHEM Score [22]. CScore was automatically computed from the six scores (0, 1, 2, 3, 4 and 5). The best CScore is 5. Structures with scores of 3 or 4 merit further consideration. Structures with a CScore of



**Scheme 1** Molecular skeletons of series I–VI. *I* Polysubstituted pyrimidinyl acyl(thio)urea analogs (1–8); *II* tert-butylaminocarbonyl acyl(thio)urea analog (9–17); *III* aryl and chrysanthemoyl Q groups

(18–30); *IV* : 3-(2-chloro-3,3,3-trifluoropropenyl)-2,2-dimethyl cyclopropyl (CFPC); *V* 3-(2,2-dichloro ethenyl)-2,2-dimethyl cyclopropyl (DCPC), *VI* zanamivir, GG167 (31)

0 are consistently considered bad by all scoring functions and should be dropped.

### QSAR analysis

Ordinary atoms of organic molecules of pharmaceutical interest include H, C, N, P, O, S, Cl, Br and I, which are partitioned into five types in the Periodic Table of the elements. According to their hybridization state, the atoms were further divided into ten subtypes. Thus, there were 55 potential interactions in each molecule (Table 1). In this paper, three kinds of potential energy fields—electrostatic, steric and hydrophobic—were employed in the representation of different interactions, producing  $3 \times 55 = 165$  interaction items for organic molecules of various drugs.

There are three atomic interaction potential energies:

$$E_{mn}(E) = \sum_{i \in m, j \in n} \frac{e^2}{4\pi\epsilon_0} \cdot \frac{Z_i \cdot Z_j}{r_{ij}} \quad (1 \leq m \leq n \leq 10) \quad (1)$$

$$E_{mn}(S) = \sum_{i \in m, j \in n} \epsilon_{ij} \cdot D \cdot \left[ \left( \frac{R_{ij}^*}{r_{ij}} \right)^{12} - 2 \cdot \left( \frac{R_{ij}^*}{r_{ij}} \right)^6 \right] \quad (2)$$

$(1 \leq m \leq n \leq 10)$

$$E_{mn}(H) = \sum_{i \in m, j \in n} S_i \cdot a_i \cdot S_j \cdot a_j \cdot e^{-r_{ij}} \cdot T_{ij} \quad (1 \leq m \leq n \leq 10) \quad (3)$$

Electrostatic interaction is an important non-bonded interaction obeying Coulomb's law. In Eq. 1,  $r_{ij}$  denotes interatomic Euclid distance, with the unit of meter (m);  $e$  is the elementary charge ( $1.6021892 \times 10^{-19}$  C);  $\epsilon_0$  represents the dielectric constant  $8.85418782 \times 10^{-12}$  C<sup>2</sup>/(J · m) in vacuum;  $Z$  is the amount of net electric charge;  $m$  and  $n$  are atomic types. The electrostatic interactions among all atoms included in a molecule could be entered into this equation, and then accumulated together into each of the 55 interaction items according to their atom-pair attributes.

Steric interaction is described by the Lennard-Jones formula (Eq. 2), where  $\epsilon_{ij} = (\epsilon_{ii} \cdot \epsilon_{jj})^{1/2}$  is the potential well of atomic pairs, with its value taken from reference [23, 24];  $R_{ij}^3 = (C_h \cdot R_{ii}^3 + C_h \cdot R_{jj}^3)/2$  [24, 25], is the van der Waals' radius of the modified atom-pair, with a correction factors  $C_h$  of 1.00, 0.95 and 0.90 in the case of  $sp^3$ ,  $sp^2$  and  $sp$  hybridization, respectively [25–27].

Hydrophobic interaction is defined as interatomic hydrophobic interaction through force field in “hint” proposed by Kellogg et al. [27], where  $S$  is the solvent accessible surface area (SASA) for atoms [26–28], giving information on surface area using a spherical water-molecule probe at the atomic surface;  $a$  is atomic hydrophobic constant, value taken from reference [29, 30];  $T$  is a sign function, indicating entropy change resulting from different types of atomic interaction [30–33].

The 3D molecular structures of 30 compounds were generated automatically using the software Chemoffice 8.0, then the semi-empirical quantum chemistry software MOPAC6.0 contained in Chem3D was used to obtain final optimized molecular structures at the AM1 level (cut-off value of 0.001 kJ mol<sup>-1</sup>). Simultaneously, atomic partial charges were calculated by the Mulliken method in the form of single points. The atomic charges and the spatial positions of all atoms in a molecule were entered into the program C Super-3D.EXE, giving rise to HoVAIFA descriptors by taking forms of Cartesian coordinates and partial charges, respectively. For any molecules containing ten atom subtypes, all 165 descriptors were obtained, namely  $V_1$ – $V_{55}$ ,  $V_{56}$ – $V_{110}$  and  $V_{111}$ – $V_{165}$  correspond to electrostatic, steric and hydrophobic interactions, respectively.

Genetic algorithms (GA) were implemented by matlab software (version 7.0). GA variable screening parameters were established as follows: 200 initial populations, 100 genetic generations, crossover probability was 0.5, mutation probability was 0.01, and the evaluation function was cross-validation correlation coefficient  $q^2$ . A statistical model was obtained by partial least squares (PLS) regression. PLS model parameter establishment was as follows: 95% confidence level; seven

**Table 1** All ten subtypes of atoms and the resulting 55 types of interactions in holographic vector of atomic interaction field (HoVAIFA)

No	Atom types	1	2	3	4	5	6	7	8	9	10
1	H	1-1	1-2	1-3	1-4	1-5	1-6	1-7	1-8	1-9	1-10
2	C(sp <sup>3</sup> )		2-2	2-3	2-4	2-5	2-6	2-7	2-8	2-9	2-10
3	C(sp <sup>2</sup> )			3-3	3-4	3-5	3-6	3-7	3-8	3-9	3-10
4	C(sp)				4-4	4-5	4-6	4-7	4-8	4-9	4-10
5	N(sp <sup>3</sup> )					5-5	5-6	5-7	5-8	5-9	5-10
6	N(sp <sup>2</sup> )						6-6	6-7	6-8	6-9	6-10
7	N(sp)							7-7	7-8	7-9	7-10
8	O(sp <sup>3</sup> ), S(sp <sup>3</sup> )								8-8	8-9	8-10
9	O(sp <sup>2</sup> ), S(sp <sup>2</sup> )									9-9	9-10
10	F, Cl, Br, I										10-10

cross validations. The model was valuated by cross-validation, Y random permutation tests, and external validation.

## Dataset

Structures and  $pIC_{50}$  of 40 NIs (Scheme 1, Table 2) were derived from reference [11].  $IC_{50}$  values were measured spectrofluorometrically using 20-(4-methylumbelliferyl)- $\alpha$ -D-acetylneuraminic acid as substrate for neuraminidase to yield a fluorescent product that could be quantified. In addition, samples with  $IC_{50} > 20 \mu M$  were discarded; the

remaining 30 samples were divided randomly into a training set (24 samples) and test set (6 samples) from reference [11].

## Models and discussion

### Molecular docking

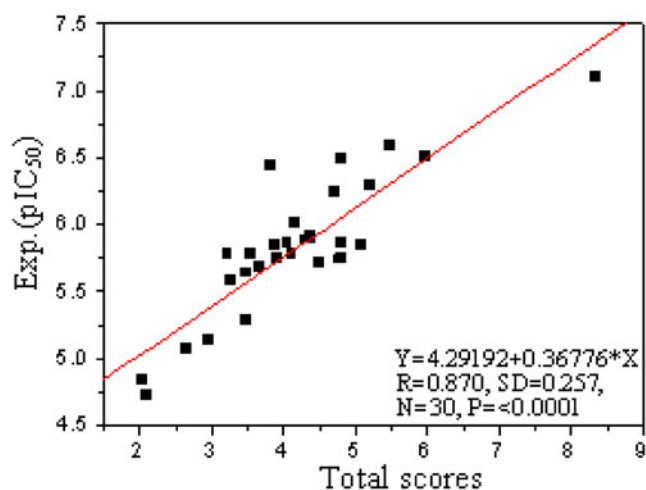
There was a strong correlation between total scores (the Surflex-Dock scores) and experimental  $pIC_{50}$  (correlation coefficient  $r=0.870$ ;  $SD=0.257$ ;  $P<0.0001$ ; Fig. 1). RMSD

**Table 2** Structures, activities and docking scores of 30 neuraminidase inhibitors (NIs). *CScore* Consensus score

ID	R	X	Y	$IC_{50}$ ( $\mu M$ )	Experimental $pIC_{50}$ (M)	Predicted $pIC_{50}$ (M)		Residual (M)		Hydrogen bond number	Total scores <sup>b</sup>	CScore
						PLS	Ref. [11]	PLS	Ref. [11]			
1 <sup>a</sup>	2-Cl	OEt	Me	1.65	5.78	6.25	6.69	-0.47	-0.91	6	4.11	3
2	2-Cl	OEt	OEt	0.08	7.10	7.03	6.96	0.07	0.14	9	8.33	4
3	2-Cl	OH	Me	0.32	6.49	6.04	6.08	0.45	0.41	6	4.80	3
4	2-Cl	OMe	OMe	1.77	5.75	6.12	5.85	-0.37	-0.10	6	4.81	3
5 <sup>a</sup>	2-Cl	Cl	Cl	14.5	4.84	5.90	6.20	-1.06	-1.36	5	2.04	4
6	4-NO <sub>2</sub>	Cl	Cl	1.66	5.78	5.69	5.85	0.09	-0.07	3	3.22	4
7 <sup>a</sup>	4-NO <sub>2</sub>	OEt	Me	2.30	5.64	6.09	6.41	-0.45	-0.77	4	3.49	4
8	4-NO <sub>2</sub>	OH	Me	0.36	6.44	6.47	5.87	-0.03	0.57	6	3.82	3
9	5-(2-Cl-Ph)-2-furyl	-	-	1.42	5.85	6.01	6.25	-0.16	-0.40	7	5.09	5
10	5-(4-NO <sub>2</sub> -Ph)-2-furyl	-	-	1.30	5.89	5.84	5.87	0.05	0.02	7	4.32	3
11 <sup>a</sup>	Ph	-	-	1.79	5.75	5.91	5.50	0.41	0.25	5	3.92	4
12	OMe	-	-	1.83	5.74	5.69	5.53	0.05	0.21	7	4.78	3
13	(2,4-Cl <sub>2</sub> -Ph)-OCH <sub>2</sub>	-	-	1.67	5.78	5.74	5.82	0.04	-0.04	7	3.55	3
14	2,6-F <sub>2</sub> -Ph	-	-	1.43	5.84	5.67	5.91	0.17	-0.07	7	3.89	5
15	S-(+)-2-Me-1-(4-Cl-Ph)-Pr	-	-	1.35	5.87	6.06	5.94	-0.19	-0.07	6	4.80	3
16	cis-(-)-CFPC	-	-	0.51	6.29	6.21	6.18	0.08	0.11	8	5.21	5
17 <sup>a</sup>	trans-(-)-DCPC	-	-	0.26	6.59	6.50	6.46	0.09	0.13	7	5.47	4
18	5-(4-NO <sub>2</sub> -Ph)-2-furyl	OMe	Me	1.22	5.91	5.96	5.81	-0.05	0.10	7	4.37	5
19	5-(2-Cl-Ph)-2-furyl	OMe	Cl	1.29	5.89	5.92	5.91	-0.03	-0.02	6	4.38	3
20	6-Cl-3-pyridyl	Me	OH	8.58	5.07	5.17	5.42	-0.10	-0.35	5	2.66	4
21	2-Cl-3-pyridyl	Me	Me	7.19	5.14	5.49	5.19	-0.35	-0.05	5	2.95	4
22	2-Cl-3-pyridyl	OMe	Cl	2.59	5.59	5.22	5.35	0.37	0.24	7	3.27	3
23	5,6-Cl <sub>2</sub> -3-pyridyl	OMe	OMe	18.5	4.73	4.67	4.88	0.06	-0.15	6	2.10	4
24 <sup>a</sup>	Ph	Me	Me	2.10	5.68	6.01	5.95	-0.33	-0.27	4	3.66	3
25	2-Me-1-(4-Cl-Ph)-Pr	OEt	OEt	0.31	6.51	6.27	6.69	0.24	-0.18	5	5.98	4
26	CFPC	OMe	OMe	0.97	6.01	6.14	5.93	-0.13	0.08	5	4.16	4
27	CFPC	Me	Me	0.58	6.24	6.19	6.44	0.05	-0.20	5	4.72	5
28	2-F-4-Cl-Ph	Me	Me	1.36	5.87	5.83	5.97	0.04	-0.10	4	4.05	3
29	2-F-4-Cl-Ph	OMe	Cl	5.10	5.29	5.57	5.32	-0.28	-0.03	4	3.49	5
30	(2,4-Cl <sub>2</sub> -Ph)-OCH <sub>2</sub>	OMe	OMe	1.89	5.72	5.77	5.81	-0.05	-0.09	3	4.49	3
31	-	-	-	-	-	5.36	-	-	-	9	7.73	5

<sup>a</sup> Samples in test set

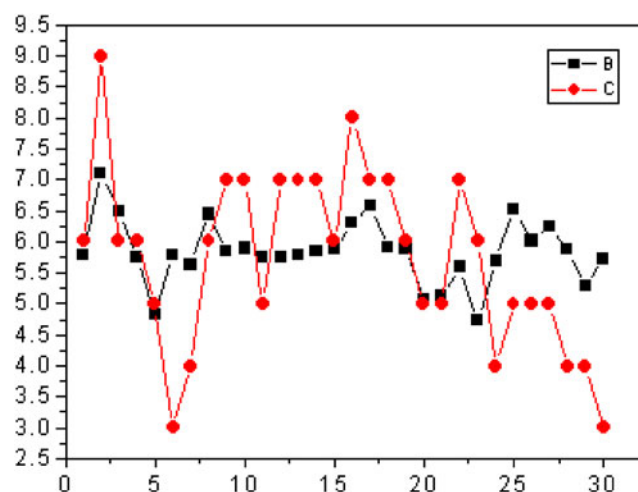
<sup>b</sup> Total scores (the Surflex-Dock scores) are expressed in  $-\log_{10}(K_d)$  units to represent binding affinities



**Fig. 1** Correlation between total scores and experimental activities ( $pIC_{50}$ ) of neuraminidase inhibitors (NIs)

of 1.16 Å and similarity value (a measure of similarity between solution coordinates and reference coordinates) of 0.75 were obtained. Moreover, the CScores showed that docking results were reasonable (CScore of zanamivir = 5; CScores of all 36 samples given in Table 2).

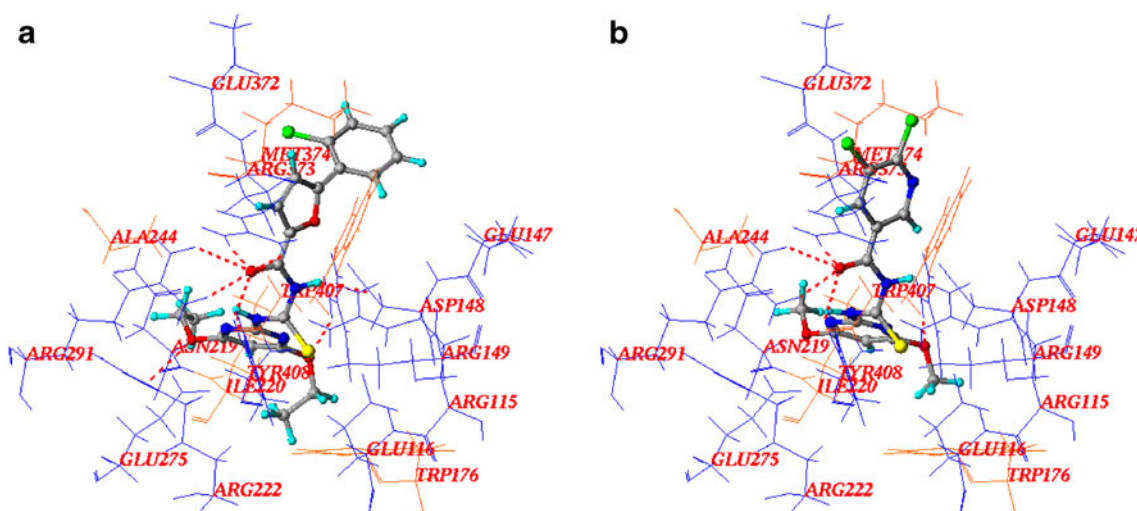
Figure 2a illustrates hydrogen bonding (dashed lines) interactions between ID 2, with the highest activity, and key residues (including neutral residues Asp148 and Tyr408, basic residues Arg149, Arg222, Arg291 and Arg373) in the active site. A total of nine hydrogen bonds of four types ( $-N \cdots H-O-$ ,  $-O \cdots H-N$ ,  $=O \cdots H-N$  and  $=O \cdots O-H$ ) were formed. Figure 2b shows hydrogen bonding (dashed lines) interactions between ID 23, with the lowest activity, and key amino acid residues (neutral residue Asp148, acid residue Glu 116, basic residues Arg149, Arg291 and Arg373) in the active site. A total of six hydrogen bonds of three types ( $-O \cdots H-N$ ,  $=O \cdots H-N$  and  $=O \cdots O-H$ ) were



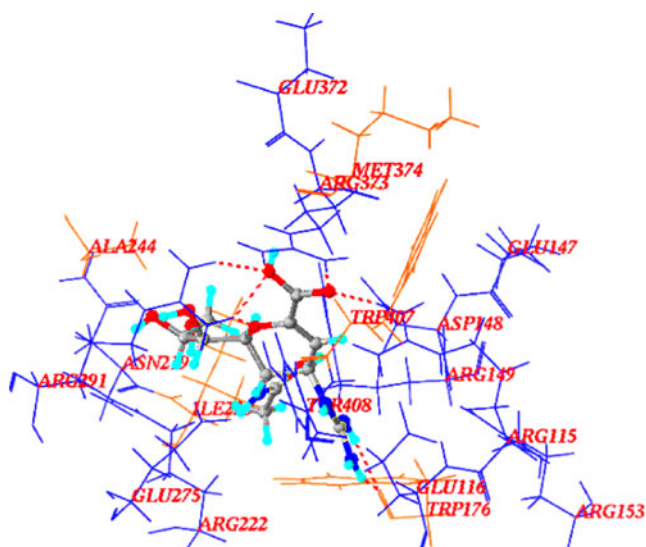
**Fig. 3** Scatter plots of hydrogen bond number ( $x$ -axis) for 30 NIs against experimental activity ( $y$ -axis;  $pIC_{50}$ )

formed. Hydrogen bond numbers of the other samples are given in Table 2. From Fig. 3, it can be seen that the number of hydrogen bonds was correlated with NI activity. Moreover, the correlation coefficient  $r$  was 0.410, with  $SD=1.334$  and  $P=0.024$ .

Figure 2 shows the presence of hydrophobic interactions between alkyl groups in Y substituent groups and residues Ile220 and Trp176), between  $=CH-$  in pyrimidine and Ile220. For example, the activity of ID 2 was higher than that of ID 1, because the ethyl group in the Y substituent group of ID 2 was more hydrophobic than the methyl group in ID 1. The same conclusion could be drawn from ID 4 and 19, ID 20 and 21. In addition, there were hydrophobic interactions between R substituent groups of NIs and residues Met374 and Trp407, for instance, ID 23 and 26, ID 26 and 30, ID 27 and 28. Thus, hydrophobic interactions correlated with activity.

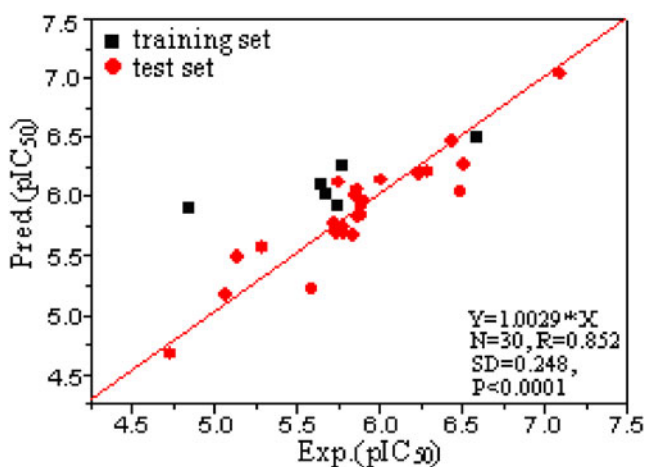


**Fig. 2** Hydrogen bonding interactions (dashed lines) between ID 2 (a), and 23 (b) and key amino acids of the active site

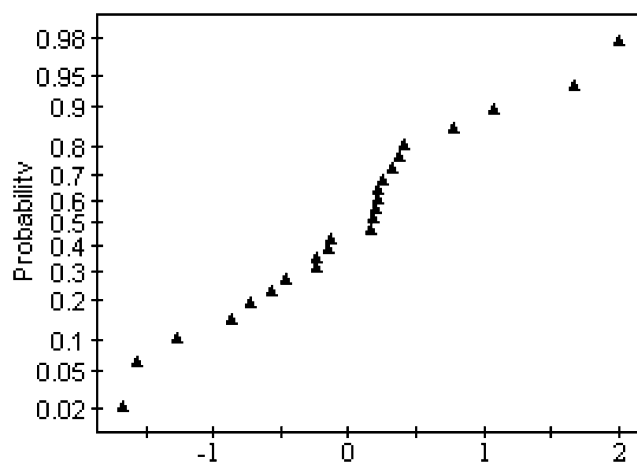


**Fig. 4** Hydrogen bonding interactions (*dashed lines*) between the positions calculated from the GG167 crystal structure of zanamivir and key amino acids of the active site

Besides hydrogen bonding and hydrophobic interactions, electrostatic interactions could be produced between NIs and residues in the active pocket. For example, R substituent groups with electronegativity were related to activity, e.g., the activity of ID 14 was higher than that of ID 11. The same result was obtained with ID 24 and 27, and ID 24 and 28. There were electrostatic interactions between oxygen atoms in X substituent groups and residues Ile220 and Arg291 (amino of amide group in Ile220, guanidyl in Arg291), between the carbonyl of the amide group in the sample and guanidyl in Arg 373, between the chlorine atom in the R substituting group and the amino of the amide group in Glu 372, and between the oxygen atom in the Y substituent group and the amino of the amide group in Tyr408. Thus, polar X substituent groups were



**Fig. 5** Plot of experimental vs predicted  $pIC_{50}$  of 30 NIs

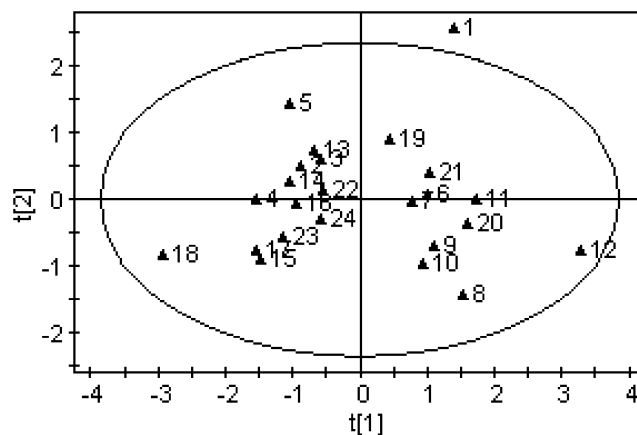


**Fig. 6** Cumulative normal probability plot of residuals

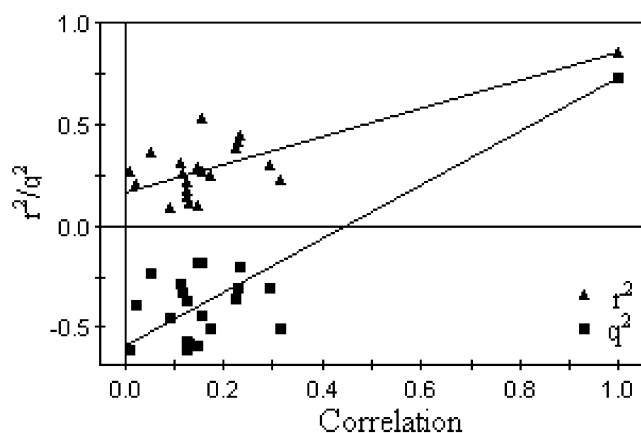
favorable to enhanced activity. For example, the activity of ID 3 was higher than that of ID 1 because the polarity of OH (X substituent group) was stronger than that of OEt in ID 3. The same conclusions were obtained from ID 5 and 19, ID 7 and 8, ID 8 and 18, and so on.

Steric interaction also had an effect on activity. For instance, the Y substituent group of ID 2 is ethoxyl, but that of ID 1 is methyl. Ethoxyl has a larger volume in comparison with methyl, therefore the activity of ID 2 was higher than that of ID 1. The same conclusion could be drawn from ID 20 and 21. In addition, the activity of ID 18 was higher than ID 7, because the volume of the X substituent group of ID 18 was smaller than that of ID 7.

The above docking results show that hydrogen bonding, and hydrophobic, electrostatic and steric interactions affected activity. Moreover, these interactions were correlated mainly with X, Y and R substituent groups together with the common acyl group structure of NIs. Thus, X, Y and R substituent groups could be modified to design new compounds with high inhibitory activity.



**Fig. 7** Scatter plot of  $t_1$  vs  $t_2$

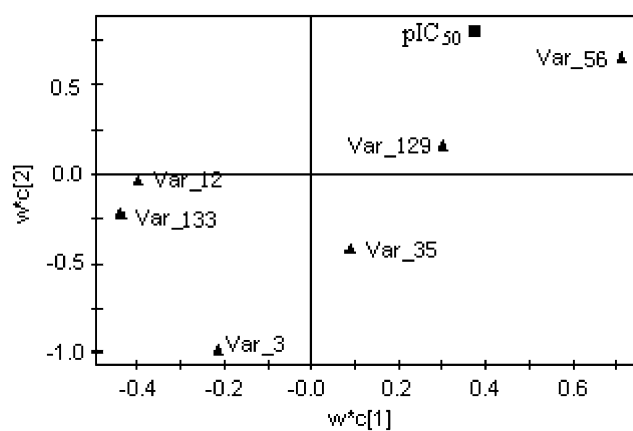
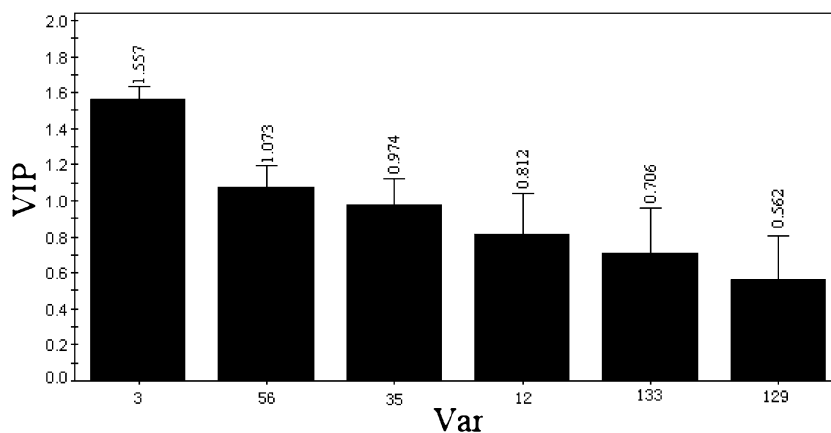


**Fig. 8** Results of  $Y$  random permutations test

Figure 4 illustrates a total of nine hydrogen bonds of two types ( $-O\cdots H-N-$  and  $=O\cdots H-N-$ ) were formed between the position calculated for the zanamivir (GG167) crystal structure and key amino acid residues (Arg115, Arg149, Trp176, Arg291 and Arg373) in the active site. Hydrogen bonds were distributed between the carboxy group, acyl group, guanidine and oxygen of cyclohexene in zanamivir and residues in the active pocket. Hydrophobic interactions could be observed between methyl in zanamivir and residues Ile 220 and Trp176, and between methylene in zanamivir and Ile220. Electrostatic interactions could be formed between hydroxyl in zanamivir and guanidino in Arg291, between the carboxyl group in zanamivir and guanidino in Arg373, between the carboxyl group of residue Glu116 and the carbonyl in Tyr408 and guanidino in zanamivir, and between the amino acyl group in zanamivir and the carbonyl in Trp407.

From Figs. 2 and 4, it can be seen that the same hydrogen bonding (Arg149, ARG291 and Arg373), hydrophobic interaction (Ile220 and Trp176) and electrostatic interaction (Arg291, Arg373 and Tyr408) residues occur in the active pocket, indicating that they might share some similarity in binding mode.

**Fig. 9** Variable importance in projection (VIP) values



**Fig. 10** Loading scatter plot

### QSAR analysis

In total, 6 descriptors ( $V_3$ ,  $V_{12}$ ,  $V_{35}$ ,  $V_{56}$ ,  $V_{129}$  and  $V_{133}$ ) were obtained by GA variable screening.  $r^2$ ,  $q^2$  and RMSEEs (root mean square error of estimations) of the optimal PLS model with five components were 0.849, 0.724 and 0.223, respectively. The QSAR model is given below:

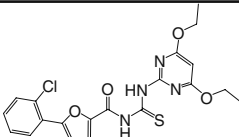
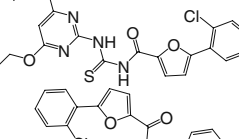
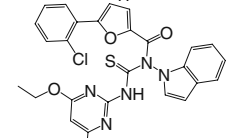
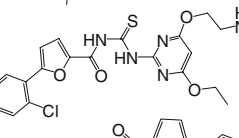
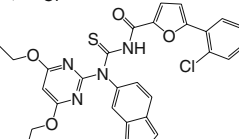
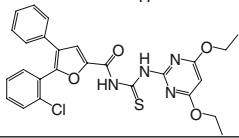
$$Y = 11.5576 - 0.888*V_3 - 0.729*V_{12} - 0.922*V_{35} + 0.501*V_{56} + 0.005*V_{129} - 0.784*V_{133} \quad (4)$$

From the PLS model, it was clear that electrostatic interaction was a dominant factor affecting activity, followed by hydrophobic and steric interactions.

Generally speaking,  $r^2$  and  $q^2$  of a reliable QSAR model should be larger than 0.8 and 0.5, respectively [34, 35]. So, the present model was indeed excellent. Moreover, an  $r^2_{\text{test}}$  value of 0.689 with  $SD=0.421$  and  $P<0.0001$  were obtained. The linear regression equation through origin for the test set was as follows:

$$Y = 1.06376*X (r = 0.830, \quad SD = 0.421, \quad N = 6, \quad P0.0001) \quad (5)$$

**Table 3** Structures, predicted activities and docking scores of five newly designed compounds

ID	Structures	Experimental pIC <sub>50</sub>	Predicted pIC <sub>50</sub>	Total Scores	CScore
2		7.10	7.03	8.33	4
2a		-	7.90	9.29	4
2b		-	7.24	8.91	3
2c		-	7.10	8.56	5
2d		-	7.00	8.80	4
2e		-	6.99	9.10	4

Here,  $Y$  and  $X$  were predictive pIC<sub>50</sub> values and experimental pIC<sub>50</sub> values, respectively.

Figure 5 shows a plot of predicted against experimental pIC<sub>50</sub> values of the 30 samples. It is obvious that all samples were distributed uniformly around the diagonal except ID 5. The reason might be the lower activity of ID 5 in comparison with that of other samples. The linear regression equation through the origin is given below:

$$Y = 1.0029 * X (r = 0.852, SD = 0.248, N = 30, P0.0001) \quad (6)$$

The predicted activity, and the error between the experimental and predicted values, of all NIs are given in Table 2. From Table 2, the predicted activities of the test set were better comparison with previously published estimates [11], which suggested that our QSAR model had good predictive capability.

In Eq. 4,  $V_3$ ,  $V_{12}$  and  $V_{35}$  represent electrostatic interactions between the first type of atoms ( $H^{s1}$ ) and the third type of atoms ( $C^{sp2}$ ), between the second type of atoms ( $C^{sp3}$ ) and the third type of atoms ( $C^{sp2}$ ), and between the fifth type of atoms ( $N^{sp3}$ ) and the fifth type of atoms ( $N^{sp3}$ ), respectively.  $V_{56}$  represents steric interac-

tion between the first type of atoms ( $H^{s1}$ ) and the first type of atoms ( $H^{s1}$ ).  $V_{129}$  and  $V_{133}$  represent hydrophobic interactions between the second type of atoms ( $C^{sp3}$ ) and the tenth type of atoms (halogen atoms, X), and between the third type of atoms ( $C^{sp2}$ ) and the sixth type of atoms ( $N^{sp2}$ ), respectively.

Comparing docking with QSAR results, hydrophobic interactions were found to occur between the second type of atoms ( $C^{sp3}$ ) and the tenth type of atoms (halogen atoms, X), in other words,  $V_{129}$  was consistent with hydrophobic interaction between NIs and residue Met374 in docking analysis (Fig. 2). Hydrophobic interaction between the third type of atoms ( $C^{sp2}$ ) and the sixth type of atoms ( $N^{sp2}$ ), namely  $V_{133}$ , reflected hydrophobic interaction between NIs and residues Trp407 and Ile220 (Fig. 2). Information on the electrostatic and steric interactions of the variables selected was also consistent with the docking results.

The above QSAR results indicated that 3D-HoVAIF descriptors could appropriately characterize structural feature of 30 NIs. The optimal PLS model was robust and had good predictive capability. It was clear that QSAR results were consistent with those of docking.



Cumulative normal probability plots of residuals display the residuals standardized on a double log scale. The standardized residual is the raw residual divided by residual standard deviation (RSD). In Fig. 6, the points of the normal probability plot of the residuals for the training lay almost on a straight line between  $-2$  and  $+2$ , and the points on the probability plot followed close to a straight line, which indicated an approximately normal distribution of residuals with no outliers.

Figure 7 displays a scatter plot of  $t_1$  (the first component) versus  $t_2$  (the second component). With the exception of outlier ID 1, the samples were located in the tolerance ellipse based on Hotelling's T2 with 95% confidence.

The model was further validated by the Y random permutations test, and the order of Y was randomly permuted a number of times (20 by default). Figure 8 displays a plot of the correlation coefficient between the original Y and the permuted Y versus the cumulative  $r^2$  and  $q^2$ , with regression lines drawn. Moreover, the intercepts of the regression line for  $r^2$  and  $q^2$  were  $-0.012$  and  $-0.681$ , respectively. These results indicated that high  $r^2$  and  $q^2$  were not caused by casual factors.

Variable importance in projection (VIP) values of PLS reflect the importance of variables in fitting both the X- and Y-scores in the model. VIP is normalized, and the average squared VIP value is 1. Thus variables in the model with a  $VIP > 1$  are more important. VIP values of  $V_3$  and  $V_{56}$  were larger than 1 (Fig. 9), therefore  $V_3$  and  $V_{56}$  were more important in explaining the inhibitory activities of NIs.

The PLS loading plot (Fig. 10) revealed a positive correlation between  $V_{56}$ ,  $V_{129}$  and  $pIC_{50}$ , and a negative correlation between  $V_3$ ,  $V_{12}$ ,  $V_{35}$ ,  $V_{133}$  and  $pIC_{50}$ .  $V_{56}$ , which dominated the first component, was highly positively related to  $pIC_{50}$ ; however,  $V_{133}$ , which also dominated the first component, was negatively correlated with  $pIC_{50}$ .  $V_{12}$  was not closely correlated with  $pIC_{50}$  in the first component.  $V_3$ , governing the second component, was highly negatively correlated with  $pIC_{50}$ , but  $V_{56}$  governing the second component was highly positively related to  $pIC_{50}$ . Thus, it could be inferred that  $V_{56}$  and  $V_3$  were the dominant variables affecting activity, which was consistent with the VIP analysis.

### Molecular design

Based on docking and QSAR results, ID 2 with the highest activity was taken as a template to design new compounds. Five new compounds with high predicted activity were designed and assessed (Table 3). The best predicted  $pIC_{50}$  value was 12.4% higher than that of the template molecule. The activities of the designed compounds improved along with increases in hydrophobic, steric, hydrogen bonding and electrosteric properties. For

example, the activity of compound 2a was enhanced along with hydrophobic and steric properties of the Y substituent group. Moreover, the total scores of these five compounds were higher than that of the template molecule, with a good score consistency.

### Conclusions

Docking results of 30 thiourea analogs indicated that hydrogen bonding and electrostatic interactions were important factors affecting inhibitory activity, followed by hydrophobic and steric factors. Moreover, there was a strong significant correlation between docking total scores and  $pIC_{50}$  ( $r=0.870$  and  $P<0.0001$ ). In addition, the QSAR model representing classic electrostatic, steric and hydrophobic interactions was robust and had good predictive ability ( $r^2=0.849$ ,  $q^2=0.724$ ,  $RMSEE=0.223$ ). Furthermore, QSAR results were in good agreement with docking results.

Based on docking and QSAR results, five new compounds with high predicted activity were designed. The predicted  $pIC_{50}$  value of the best compound was 12.4% higher than that of template molecule ID 2.

The results showed that, as a method of structural description, HoVAIFA can characterize the complex interactions between drug molecules and related biomolecules. HoVAIFA parameters with clear physicochemical meaning are easy to interpret. QSAR studies based on HoVAIFA can be calculated only in the knowledge of molecular 2D structures, and do not consider conformation. In addition, HoVAIFA calculations are convenient and less time consuming than other methods. Therefore, HoVAIFA warrants further study and is expected to be widely used in molecular structure and function studies in the future.

**Acknowledgments** The authors thank Professor Gang Chen of the National Technology College of Singapore for offering us docking results of the Sybyl program and help. This work was supported by the National High Technology Research and Development Program of China (Grant No. 2006AA02Z312).

### References

1. McKimm-Breschkin JL (2000) Resistance of influenza viruses to neuraminidase inhibitors—a review. *Antiviral Res* 47:1–17
2. Verma RP, Hansch C (2006) A QSAR study on influenza neuraminidase inhibitors. *Bioorg Med Chem* 14:982–996
3. Stephenson I, Clark TW, Pareek M (2008) Antiviral treatment and prevention of seasonal influenza: a comparative review of recommendations in the European Union. *J Clin Virol* 42:244–248
4. Kim CU, Lew W, Williams MA, Wu H, Zhang LJ, Chen XW, Escarpe PA, Mendel DB, Graeme Laver W, Stevens RC (1998) Structure–activity relationship studies of novel carbocyclic influenza neuraminidase inhibitors. *J Med Chem* 41:2451–2460
5. Brouillette WJ, Atigadda VR, Luo M, Air GM, Babu YS, Bantia S (1999) Design of benzoic acid inhibitors of influenza neuraminidase

- containing a cyclic substituituon for the N-acetyl grouping. *Bioorg Med Chem Lett* 9:1901–1906
- Wang GT, Chen YW, Wang S, Gentles R, Sowin T, Kati W, Muchmore S, Giranda V, Stewart K, Sham H, Kempf D, Laver WG (2001) Design, synthesis, and structural analysis of influenza neuraminidase inhibitors containing pyrrolidine cores. *J Med Chem* 44:1192–1201
  - Krueger AC, Xu YB, Kati WM, Kempf DJ, Maring CJ, McDaniel KF, Molla A, Montgomery D, Kohlbrenner WE (2008) Synthesis of potent pyrrolidine influenza neuraminidase inhibitors. *Bioorg Med Chem Lett* 18:1692–1695
  - Sun CW, Zhang XD, Huang H, Zhou P (2006) Synthesis and evaluation of a new series of substituted acyl(thio)urea and thiadiazolo [2, 3-a] pyrimidine derivatives as potent inhibitors of influenza virus neuraminidase. *Bioorg Med Chem* 14:8574–8581
  - Yi X, Guo Z, Chu FM (2003) Study on molecular mechanism and 3D-QSAR of influenza neuraminidase inhibitors. *Bioorg Med Chem* 11:1465–1474
  - Lü WJ, Chen YL, Ma WP, Zhang XY, Luan F, Liu MC, Chen XG, Hu ZD (2008) QSAR study of neuraminidase inhibitors based on heuristic method and radial basis function network. *Eur J Med Chem* 43:569–576
  - Nair PC, Sobhia ME (2008) Quantitative structure activity relationship studies on thiourea analogues as influenza virus neuraminidase inhibitors. *Eur J Med Chem* 43:293–299
  - Liu SS, Cai SC, Cao CZ, Li ZS (2000) Molecular electronegative distance vector (MEDV) related to 15 properties of alkanes. *J Chem Inf Comput Sci* 40:1337–1348
  - Taylor NR, Cleasby A, Singh O, Skarzynski T, Wonacott AJ, Smith PW, Sollis SL, Howes PD, Cherry PC, Bethell R, Colman P, Varghese J (1998) Dihydropyranocarboxamides related to zanamivir: a new series of inhibitors of influenza virus sialidases. 2. Crystallographic and molecular modeling study of complexes of 4-amino-4H-pyran-6-carboxamides and sialidase from influenza virus types A and B. *J Med Chem* 41:798–807
  - Jain AN (2003) Surflex: fully automatic flexible molecular docking using a molecular similarity-based search engine. *J Med Chem* 46:499–511
  - Ruppert J, Welch W, Jain AN (1997) Automatic identification and representation of protein binding sites for molecular docking. *Protein Sci* 6:524–533
  - Muthas D, Sabnis YA, Lundborg M, Karlén A (2008) Is it possible to increase hit rates in structure-based virtual screening by pharmacophore filtering? An investigation of the advantages and pitfalls of post-filtering. *J Mol Graph Model* 26:1237–1251
  - Clark RD (2008) A ligand's-eye view of protein binding. *J Comput Aided Mol Des* 22:507–521
  - Clark RD, Strizhev A, Leonard JM, Blake JF, Matthew JB (2002) Consensus scoring for ligand/protein interactions. *J Mol Graph Model* 20:281–295
  - Jones G, Willet P, Glen RC, Leach AR, Taylor R (1997) Development and validation of a genetic algorithm for flexible docking. *J Mol Biol* 267:727–748
  - Muegge I, Martin YC (1999) A general and fast scoring function for protein-ligand interactions: a simplified potential approach. *J Med Chem* 42:791–804
  - Kuntz ID, Blaney JM, Oatley SJ, Langridge R, Ferrin TE (1982) A geometric approach to macromolecule-ligand interactions. *J Mol Biol* 161:269–288
  - Eldridge MD, Murray CW, Auton TR, Paolini GV, Mee RP (1997) Empirical scoring functions: I. The development of a fast empirical scoring function to estimate the binding affinity of ligands in receptor complexes. *J Comput Aided Mol Des* 11:425–445
  - Levitt M (1983) Protein folding by restrained energy minimization and molecular dynamics. *J Mol Biol* 170:723–764
  - Levitt M, Perutz MF (1988) Aromatic rings act as hydrogen bond acceptors. *J Mol Biol* 201:751–754
  - Kellogg GE, Burnett JC, Abraham DJ (2001) Very empirical treatment of solvation and entropy; a force field derived from Log Po/w. *J Comput Aided Mol Des* 15:381–393
  - Kellogg GE, Semus SF, Abraham DJ (1991) HINT—a new method of empirical hydrophobic field calculation for CoMFA. *J Comput Aided Mol Des* 5:545–552
  - Hahn M (1995) Receptor surface models. 1. Definition and construction. *Med Chem* 38:2080–2090
  - Abraham DJ, Kellogg GE, Holt JM, Ackers GK (1997) Hydropathic analysis of the non-covalent interactions between molecular subunits of structurally characterized hemoglobins. *Mol Biol* 272:613–632
  - Gunnarsson GT, Umesh RD (2004) Hydropathic interaction analyses of small organic activators binding to antithrombin. *Bioorg Med Chem* 12:633–640
  - Pei J, Wang Q, Zhou J, Lai L (2004) Estimating protein-ligand binding free energy; atomic solvation parameters for partition coefficient and solvation free energy calculation. *Proteins* 57:651–664
  - Kellogg GE, Joshi GS, Abraham DJ (1992) New tools for modeling and understanding hydrophobicity and hydrophobic interactions. *Med Chem Res* 1:444–453
  - Kellogg GE, Abraham DJ (1992) Complementary hydrophobicity map predictions of drug structure from a known receptor/receptor structure from known drugs. *J Mol Graph* 10:212–217
  - Hasel W, Hendrikson TF, Still WC (1988) A rapid approximation to the solvent accessible surface areas of atoms. *Tetrahedron Comput Methodol* 1:103–116
  - Clark M, Crame RD (1993) The probability of chance correlation using partial least squares (PLS). *Quant Struct Act Relat* 2:137–145
  - Yao SW, Lopes VHC, Fernández F, García-Mera X, Morales M, Rodríguez-Borges JE, Cordeiro MNDS (2003) Synthesis and QSAR study of the anticancer activity of some novel indane carbocyclic nucleosides. *Bioorg Med Chem* 11:4999–5006