European Journal of Medicinal Chemistry 45 (2010) 1008–1014

Contents lists available at ScienceDirect

European Journal of Medicinal Chemistry

journal homepage: http://www.elsevier.com/locate/ejmech

Original article

# Docking and 3D-QSAR studies of influenza neuraminidase inhibitors using three-dimensional holographic vector of atomic interaction field analysis

Jiaying Sun [a,b,*], Shaoxi Cai [a,b], Ning Yan [a], Hu Mei [a,b,*]

[a] College of Bioengineering, Chongqing University, Chongqing 400044, China
[b] Key Laboratory of biorheological science and technology of Educational Ministry, Chongqing 400044, China

ARTICLE INFO

*Article history:*
Received 27 June 2009
Received in revised form
12 November 2009
Accepted 20 November 2009
Available online 27 November 2009

*Keywords:*
Neuraminidase inhibitors (NIs)
Docking
Three-dimensional holographic vector of atomic interaction field analysis (HoVAIFA)
3D-QSAR

ABSTRACT

Surflex-Dock is employed to investigate interactions between neuraminidase inhibitors (NIs) and neuraminidase (NA), which illuminate that carboxyl group, amino (guanidino) group, amide group, hydroxy group are crucial. Hydrogen bonds and hydrophobic interactions impact on activities of NIs. There is a strong correlation between binding affinity and $pIC_{50}$, with $r = 0.813$. We have developed three-dimensional holographic vector of atomic interaction field analysis (HoVAIFA) as a new method of 3D-QSAR to understand chemical–biological interactions. Good results, $R^2 = 0.789$ and $R^2cv = 0.732$, show that HoVAIFA can be applicable to molecular structural characterization and bioactivity prediction. Electrostatic, steric and hydrophobic interactions affect activities of NIs. HoVAIFA and docking results are corresponding, which illustrates that HoVAIFA is an effective methodology for characterization of complex interactions of drug molecules.



## 1. Introduction

Influenza is a serious danger to human health acute toxicity of a fast-spreaded respiratory infection, which is one of the main causes of death [1]. Given the lack of effective drugs for prevention and treatment of influenza, the development of a novel anti-influenza virus drug is of great significance. Influenza is an RNA virus that contains two major surface glycoproteins, namely, neuraminidase (NA) and hemagglutinin (HA). NA can cleave the a-ketosidic connections of sialic acid and nearby residues of sugar [2]. It also destroys HA on the virus surface allowing the emergence of progeny virus units from infected cells. So, NA is a potential target to control influenza virus. Chemicals that inhibit NA can protect the host from viral infection. Based on the NA crystal structures elucidated, many high selective neuraminidase inhibitors (NIs) are reasonably designed. At present, both zanamivir and oseltamivir are effective inhibitors for both A and B forms of neuraminidase [3,4]. Zanamivir is administered by oral inhalation due to high polar compounds, and oseltamivir is a prodrug that is converted after oral intake to its active form, the carboxylic acid (GS 4071).

Quantitative structure activity relationship (QSAR) is an important method for designing drug, so, construction of quantitative correlation between the molecular structure and biological activity for these compounds has an important significance to research and development of high efficiency anti-influenza drug. For example, 17 QSAR models for different series of compounds including benzoic acids [2,5–7], carbocyclic derivatives [8,9], cyclopentanes [9,10], isoquinolines [11–14], and pyrrolidines [15] were developed using MLR (multiple linear regressions) to understand chemical–biological interactions governing activities of NIs, by reporting Verma et al. [16].

However, an efficient approach for investigating protein–ligand interactions, molecular docking plays a key role in rational drug design [17]. So, protein–ligand interactions were investigated using Surflex-Dock in the present paper. QSAR studies of above-mentioned compounds were carried out utilizing three-dimensional holographic vector of atomic interaction field (3D-HoVAIF), and the influence of molecular structure on neuraminidase inhibiting activities were also discussed in detail. 3D-HoVAIF is proposed based upon a 2D structural descriptor developed by Liu et al. [18] in our laboratory. Proceeding from two spatial invariants, namely atom relative distance and atomic properties on the bases of three common non-bonded (electrostatic, van der Waals and hydrophobic) interactions which are directly associated with bioactivities, 3D-HoVAIF method derives multidimensional vectors to represent molecular steric structural characteristics.

* Corresponding authors. College of Bioengineering, Chongqing University, Chongqing 400044, China. Tel./fax: +86 23 65112677.
*E-mail addresses:* wy7472@126.com (J. Sun), meihu@cqu.edu.cn (H. Mei).

0223-5234/$ – see front matter 
doi:10.1016/j.ejmech.2009.11.043

**Table 1**
All 10 subtypes of atoms and the resulting 55 types of interactions in HoVAIFA.

| No | Atomic 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($sp^3$) | | 2–2 | 2–3 | 2–4 | 2–5 | 2–6 | 2–7 | 2–8 | 2–9 | 2–10 |
| 3 | C($sp^2$) | | | 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($sp^3$) | | | | | 5–5 | 5–6 | 5–7 | 5–8 | 5–9 | 5–10 |
| 6 | N($sp^2$) | | | | | | 6–6 | 6–7 | 6–8 | 6–9 | 6–10 |
| 7 | N(sp) | | | | | | | 7–7 | 7–8 | 7–9 | 7–10 |
| 8 | O($sp^3$), S($sp^3$) | | | | | | | | 8–8 | 8–9 | 8–10 |
| 9 | O($sp^2$), S($sp^2$) | | | | | | | | | 9–9 | 9–10 |
| 10 | F, Cl, Br, I | | | | | | | | | | 10–10 |

## 2. Experiment

### 2.1. Methods and materials

#### 2.1.1. Docking

Surflex-Dock was applied to study molecular docking. Crystal structure of NA was retrieved from RCSB Protein Data Bank (PDB entry code: 2ht7) [19]. This is a particular structure with oseltamivir (GS4701). Surflex-Dock uses an empirical scoring function and a patented search engine to dock ligands into a protein's binding site [20]. Protomol is used to guide molecular docking. Protomol is a computational representation of the intended binding site to which putative ligands are aligned. Production of protomol supplies three manners [21]: (1) Automatic: Surflex-Dock finds the largest cavity in the receptor protein; (2) Ligand: A ligand in the same coordinate space as the receptor; (3) Residues: Specified residues in the receptor. Surflex-Dock scores are expressed in $-\log_{10}(K_d)$ units to represent binding affinities. Surflex-Dock scores are evaluated by CScore (Consensus Score) [22]. CScore integrates a number of popular scoring functions for ranking the affinity of ligands bound to the active site of a receptor. 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. CScore provides several functions: D Score [23], PMF (Potential of Mean Force) Score [24], G Score [25] and CHEM Score [26]. The consensus scores range from 1 to 5. The best CScore is 5. The protein structure was utilized in subsequent docking experiments without energy minimization. All ligands and water molecules have been removed at first and the polar hydrogen atoms were added. Automatic docking is employed. Other parameters are established by default in software. All NIs are minimized using default parameter.

#### 2.1.2. QSAR

Ordinary atoms of organic molecules with pharmaceutical interests including H, C, N, P, O, S, Cl, Br, I, which are partitioned into 5 types in the Element Periodic Table. According to their hybridization state, the atoms are furthermore divided into 10 subtypes. Thus, there are 55 interactions in a molecule (Table 1). In this paper, three kinds of potential energy fields, electrostatic, steric and hydrophobic, are 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:

$$E_{mn}(E) = \sum_{i\in m, j\in n} \frac{e^2}{4\pi\varepsilon_0}\cdot\frac{Z_i\cdot Z_j}{r_{ij}} (1 \le m \le n \le 10) \quad (1)$$

$$E_{mn}(S) = \sum_{i\in m, j\in n} \varepsilon_{ij}\cdot D\cdot\left[\left(\frac{R^*_{ij}}{r_{ij}}\right)^{12} - 2\cdot\left(\frac{R^*_{ij}}{r_{ij}}\right)^{6}\right] (1 \le m \le n \le 10) \quad (2)$$

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

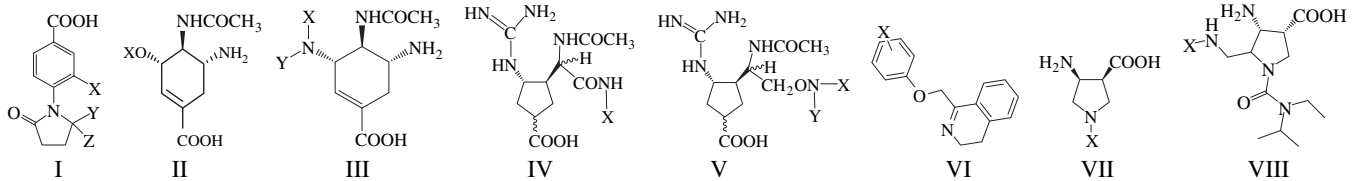
Electrostatic interaction is an important non-bonded interaction obeying Coulomb's law. In Eq. (1), $r_{ij}$ denotes interatomic Euclid distance, with the unit of meter (m); $e$ is the elementary charge ($1.602\ 189\ 2 \times 10^{-19}$ C); $\varepsilon_0$ represents dielectric constant $8.854\ 187\ 82 \times 10^{-12}$ $C^2/(J\cdot m)$ in vacuum; $Z$ is the amounts of net electric charges; $m$ and $n$ are atomic types. Electrostatic interactions among all atoms included in a molecule could be given out by this equation, and then accumulating them together into each of the 55 interaction items according to their atom-pair attributes.

Steric interaction is described by Lennard–Jones formula (Eq. (2)). Amongst, $\varepsilon_{ij} = (\varepsilon_{ii}\cdot\varepsilon_{jj})^{1/2}$ is potential well of atomic pairs, with its value taken from reference [27,28]; $R^3_{ij} = (C_h\cdot R^3_{ii} + C_h\cdot R^3_{jj})/2$ [29,30], is van der Waals' radius for modified atom-pair, with corrected factor $C_h$ of 1.00 in case of $sp^3$ hybridization, 0.95 $sp^2$ hybridization and 0.90 sp hybridization [31].

Hydrophobic interaction is defined as interatomic hydrophobic interaction through force field in "hint" proposed by Kellogg et al. [32]. Amongst, $S$ is the solvent accessible surface area (SASA) for atoms [33], indicating information on surface area when water-molecule probe roiling its sphere at the atomic surface; $A$ is atomic hydrophobic constant, taken the value from reference [34]; $T$ is sign function, indicating entropy change resulting from different types of atomic interaction [35–38].

Three-dimensional molecular structures of the 124 compounds are automatically generated by software Chemoffice8.0, and then semi-empirical quantum chemistry software MOPAC6.0 contained in Chem3D is used to obtain final optimized molecular structures at AM1 levels (cut-off value of 0.001 kJ/mol). Simultaneously, atomic partial charges are calculated by Mulliken Method in the form of single-point. Spatial positions for all atoms in a molecule and the atomic charges are put into C program Super-3D.EXE, giving rise to HoVAIFA descriptors by taking forms of Cartesian coordinates and partial charges, respectively. For any molecule contain 10 subtypes of atoms, all 165 descriptors are obtained, namely $V_1 \sim V_{55}$, $V_{56} \sim V_{110}$ and $V_{111} \sim V_{165}$ correspond to electrostatic, steric and hydrophobic interactions, respectively.

Genetic algorithm (GA) is implemented by matlab software (version 7.0). GA variable screening parameter establishment:




**Scheme 1.** Various Molecular Skeletons: Series I ~ VIII. I (Benzoic acid derivatives 1–7); II (Cyclohexene derivatives 8–41, 55–64); III (Cyclohexene derivatives 42–54); IV (Cyclopentane derivatives 65–78); V (Cyclopentane derivatives 79–91); VI (Isoquinoline derivatives 92–107); VII (Pyrrolidine Derivatives 108–119); VIII (Pyrrolidine Derivatives 120–124).

**Table 2**
Experiment, calculation, errors values of $pEC_{50}$ and total scores for all samples.

| ID | X | Y | Z | Exp. (M) | $Cal_{EST.}$ (M) | $Cal_{LOO}$ (M) | $Err_{EST.}$ (M) | $Err_{LOO}$ (M) | Total Scores | CScore |
|---|---|---|---|---|---|---|---|---|---|---|
| 1 | NHC(=NH)$NH_2$ | H | H | 3.60 | 4.094 | 4.329 | 0.494 | 0.729 | 4.68 | 3 |
| 2 | NHC(=NH)$NH_2$ | H | $CH_2OH$ | 4.70 | 3.915 | 3.592 | −0.785 | −1.108 | 5.96 | 4 |
| [a]3 | NHC(=NH)$NH_2$ | H | $CH_2NH$ | 2.59 | 3.529 | – | 0.939 | – | 3.86 | 4 |
| 4 | H | $CH_2OH$ | $CH_2OH$ | 3.12 | 5.508 | 5.629 | 2.388 | 2.509 | 4.73 | 3 |
| 5 | NHC(=NH)$NH_2$ | $CH_2OH$ | $CH_2OH$ | 5.30 | 4.389 | 4.027 | −0.911 | −1.273 | 5.99 | 4 |
| 6 | NHCH$(C_2H_5)_2$ | H | H | 3.65 | 4.078 | 4.141 | 0.428 | 0.491 | 4.64 | 4 |
| 7 | NHCH$(C_2H_5)_2$ | $CH_2OH$ | $CH_2OH$ | 7.32 | 5.737 | 5.490 | −1.583 | −1.830 | 6.83 | 4 |
| 8 | CH$(C_2H_5)C_2H_5$ | – | – | 9.00 | 6.967 | 6.891 | −2.033 | −2.109 | 7.67 | 5 |
| 9 | $CH_2CH_2CH_3$ | – | – | 6.89 | 6.624 | 6.612 | −0.266 | −0.278 | 7.44 | 5 |
| 10 | $CH_2OCH_3$ | – | – | 5.70 | 6.085 | 6.115 | 0.385 | 0.415 | 5.92 | 3 |
| 11 | $CH_2CH_2CF_3$ | – | – | 6.65 | 6.314 | 6.294 | −0.336 | −0.356 | 6.39 | 4 |
| 12 | $CH_2CH{=}CH_2$ | – | – | 5.66 | 6.476 | 6.550 | 0.816 | 0.890 | 5.79 | 3 |
| 13 | Cyclopentyl | – | – | 7.66 | 6.797 | 6.753 | −0.863 | −0.907 | 7.51 | 4 |
| [a]14 | Cyclohexyl | – | – | 7.22 | 7.015 | – | −0.205 | – | 6.68 | 4 |
| [a]15 | Phenyl | – | – | 6.28 | 7.476 | – | 1.196 | – | 7.02 | 5 |
| 16 | H | – | – | 5.20 | 6.068 | 6.097 | 0.868 | 0.897 | 4.40 | 5 |
| 17 | Me | – | – | 5.43 | 6.333 | 6.356 | 0.903 | 0.926 | 5.02 | 4 |
| 18 | $C_2H_5$ | – | – | 5.70 | 6.562 | 6.581 | 0.862 | 0.881 | 5.85 | 3 |
| [a]19 | $C_3H_7$ | – | – | 6.74 | 6.726 | – | −0.014 | – | 7.44 | 5 |
| 20 | $C_4H_9$ | – | – | 6.52 | 6.863 | 6.871 | 0.343 | 0.351 | 6.15 | 5 |
| 21 | $C_5H_{11}$ | – | – | 6.70 | 6.979 | 6.986 | 0.279 | 0.286 | 7.67 | 3 |
| 22 | $C_6H_{13}$ | – | – | 6.82 | 7.081 | 7.088 | 0.261 | 0.268 | 7.81 | 4 |
| 23 | $C_7H_{15}$ | – | – | 6.57 | 7.170 | 7.189 | 0.600 | 0.619 | 7.77 | 4 |
| 24 | $C_8H_{17}$ | – | – | 6.74 | 7.250 | 7.269 | 0.510 | 0.529 | 7.02 | 3 |
| 25 | $C_9H_{17}$ | – | – | 6.68 | 7.323 | 7.349 | 0.643 | 0.669 | 7.8 | 5 |
| 26 | $C_{10}H_{21}$ | – | – | 6.22 | 7.390 | 7.442 | 1.170 | 1.222 | 5.93 | 5 |
| [a]27 | $CH_2CHMe_2$ | – | – | 6.70 | 7.011 | – | 0.311 | – | 6.08 | 3 |
| 28 | CH(Me)$C_2H_5$(R) | – | – | 8.00 | 7.077 | 7.052 | −0.923 | −0.948 | 6.81 | 3 |
| 29 | CH(Me)$C_2H_5$(S) | – | – | 8.05 | 7.108 | 7.081 | −0.942 | −0.969 | 6.81 | 3 |
| 30 | CH$(C_2H_5)_2$ | – | – | 9.00 | 7.315 | 7.254 | −1.685 | −1.746 | 7.67 | 4 |
| 31 | CH$(C_2H_5)CH_2CH{=}CH_2$(R) | – | – | 9.00 | 7.356 | 7.298 | −1.644 | −1.702 | 7.27 | 4 |
| 32 | CH$(C_2H_5)CH_2CH{=}CH_2$(S) | – | – | 8.52 | 7.356 | 7.315 | −1.164 | −1.205 | 7.27 | 4 |
| 33 | CH$(C_2H_5)C_7H_{15}$ | – | – | 9.00 | 7.935 | 7.845 | −1.065 | −1.155 | 8.25 | 4 |
| [a]34 | Cy-$C_6H_{11}$ | – | – | 7.22 | 7.233 | – | 0.013 | – | 6.68 | 5 |
| [a]35 | CH$(C_2H_5)CH_2$-Cy-$C_6H_{11}$ | – | – | 7.80 | 7.972 | – | 0.172 | – | 7.61 | 5 |
| 36 | CH$(C_2H_5)CH_2CH_2$- Cy-$C_6H_{11}$ | – | – | 9.00 | 8.047 | 7.955 | −0.953 | −1.045 | 7.67 | 3 |
| 37 | $C_6H_5$ | – | – | 6.28 | 5.629 | 5.606 | −0.651 | −0.674 | 7.02 | 3 |
| 38 | $CH_2C_6H_5$ | – | – | 6.21 | 5.629 | 5.609 | −0.581 | −0.601 | 7.19 | 5 |
| [a]39 | CH$(C_2H_5)CH_2CH_2C_6H_5$(R) | – | – | 9.52 | 7.790 | – | −1.730 | – | 7.84 | 3 |
| [a]40 | CH$(C_2H_5)CH_2CH_2C_6H_5$(S) | – | – | 7.92 | 7.512 | – | −0.408 | – | 7.84 | 3 |
| 41 | $CH_2CH_2CH_2$-$C_6H_4$(4-$C_6H_5$) | – | – | 7.05 | 7.181 | 7.187 | 0.131 | 0.137 | 7.22 | 5 |
| 42 | Me | $(CH_2)_2CH_3$ | – | 7.19 | 6.365 | 6.333 | −0.825 | −0.857 | 6.71 | 4 |
| 43 | Me | $(CH_2)_3CH_3$ | – | 6.74 | 7.241 | 7.260 | 0.501 | 0.520 | 6.73 | 3 |
| [a]44 | Me | $CH(CH_2CH_3)_2$ | – | 8.22 | 6.873 | – | −1.347 | – | 5.78 | 5 |
| 45 | Me | $(CH_2)_2$-$C_6H_5$ | – | 7.00 | 6.578 | 6.557 | −0.422 | −0.443 | 7.50 | 5 |
| 46 | Me | Cy-$C_6H_5$ | – | 6.70 | 6.980 | 6.989 | 0.280 | 0.289 | 7.18 | 4 |
| 47 | $CH_2CH_3$ | $(CH_2)_2CH_3$ | – | 7.05 | 6.980 | 6.978 | −0.070 | −0.072 | 5.62 | 5 |
| 48 | $CH_2CH_3$ | $(CH_2)_3CH_3$ | – | 7.07 | 7.249 | 7.261 | 0.179 | 0.191 | 6.12 | 3 |
| 49 | $(CH_2)_2CH_3$ | $(CH_2)_2CH_3$ | – | 7.92 | 7.117 | 7.090 | −0.803 | −0.830 | 6.86 | 4 |
| [a]50 | H | $(CH_2)_3CH_3$ | – | 6.70 | 7.049 | – | 0.349 | – | 7.03 | 3 |
| [a]51 | H | $CH(CH_2CH_3)_2$ | – | 7.96 | 6.640 | – | −1.320 | – | 8.39 | 3 |
| 52 | H | $COCH_2CH_3$ | – | 5.57 | 5.745 | 5.753 | 0.175 | 0.183 | 5.29 | 4 |
| 53 | H | $COCHMe_2$ | – | 5.19 | 5.907 | 5.935 | 0.717 | 0.745 | 5.75 | 4 |
| 54 | H | $COCH(CH_2CH_3)_2$ | – | 5.40 | 6.168 | 6.193 | 0.768 | 0.793 | 6.04 | 5 |
| 55 | H | – | – | 5.20 | 6.068 | 6.097 | 0.868 | 0.897 | 4.4 | 5 |
| 56 | Me | – | – | 5.43 | 6.333 | 6.356 | 0.903 | 0.926 | 5.02 | 3 |
| 57 | $C_2H_5$ | – | – | 5.70 | 6.562 | 6.581 | 0.862 | 0.881 | 5.85 | 5 |
| 58 | $C_3H_7$ | – | – | 6.74 | 6.803 | 6.804 | 0.063 | 0.064 | 7.44 | 4 |
| 59 | $C_4H_9$ | – | – | 6.52 | 6.962 | 6.973 | 0.442 | 0.453 | 6.15 | 3 |
| 60 | $CH_2CHMe_2$ | – | – | 6.70 | 6.952 | 6.958 | 0.252 | 0.258 | 6.08 | 4 |
| 61 | CH(Me)$CH_2CH_3$(R) | – | – | 8.00 | 7.077 | 7.052 | −0.923 | −0.948 | 6.81 | 4 |
| 62 | CH(Me)$CH_2CH_3$(S) | – | – | 8.05 | 7.077 | 7.051 | −0.973 | −0.999 | 6.81 | 3 |
| [a]63 | CH$(C_2H_5)_2$ | – | – | 9.00 | 7.749 | – | −1.251 | – | 7.67 | 3 |
| 64 | CH$(C_3H_7)_2$ | – | – | 7.80 | 7.749 | 7.745 | −0.051 | −0.055 | 7.31 | 4 |
| 65 | $CH_2CH_3$ | – | – | 6.48 | 5.408 | 5.348 | −1.072 | −1.132 | 6.71 | 5 |
| 66 | CH$(CH_3)_2$ | – | – | 5.87 | 5.556 | 5.540 | −0.314 | −0.330 | 7.19 | 4 |
| 67 | $CH_2CH{=}CH_2$ | – | – | 6.14 | 5.277 | 5.229 | −0.863 | −0.911 | 5.97 | 3 |
| [a]68 | $(CH_2)_3CH_3$ | – | – | 5.00 | 5.877 | – | 0.877 | – | 6.94 | 5 |
| [a]69 | $CH(CH_3)CH_2CH_3$ | – | – | 6.39 | 5.620 | – | −0.770 | – | 7.43 | 5 |
| [a]70 | $CH(CH_2CH_3)_2$ | – | – | 7.10 | 5.728 | – | −1.372 | – | 6.78 | 4 |
| 71 | $CH(CH_3)(CH_2)_2CH_3$ | – | – | 5.49 | 5.718 | 5.747 | 0.228 | 0.238 | 5.40 | 4 |
| [a]72 | $CH(CH_3)CH_2CH(CH_3)_2$ | – | – | 5.08 | 5.733 | – | 0.653 | – | 5.77 | 4 |
| 73 | $CH(CH_3)(CH_2)_3CH_3$ | – | – | 5.06 | 5.718 | 5.850 | 0.658 | 0.687 | 6.82 | 5 |
| 74 | $CH(CH_2CH_3)(CH_2)_3CH_3$ | – | – | 5.74 | 5.846 | 6.062 | 0.106 | 0.110 | 6.03 | 4 |

**Table 2** (*continued*)

| ID | X | Y | Z | Exp. (M) | $Cal_{EST.}$ (M) | $Cal_{LOO}$ (M) | $Err_{EST.}$ (M) | $Err_{LOO}$ (M) | Total Scores | CScore |
|---|---|---|---|---|---|---|---|---|---|---|
| [a]75 | $(CH_2)_2C_6H_5$ | – | – | 5.19 | 5.741 | – | 0.551 | – | 5.43 | 4 |
| 76 | $CH(CH_3)(CH_2)_2C_6H_5$ | – | – | 6.64 | 6.081 | 6.118 | −0.559 | −0.578 | 5.82 | 4 |
| 77 | $CH(CH_3)CH_2OCH_3$ | – | – | 5.36 | 6.085 | 6.817 | 0.725 | 0.758 | 5.48 | 5 |
| [a]78 | $CH(CH_2CH_3)CH_2OCH_3$ | – | – | 5.66 | 6.284 | – | 0.624 | – | 6.06 | 4 |
| 79 | $CH_3$ | $(CH_2)_2CH_3$ | – | 6.03 | 6.752 | 6.500 | 0.722 | 0.787 | 6.88 | 4 |
| [a]80 | $CH_3$ | $CH(CH_3)_2$ | – | 5.49 | 6.516 | – | 1.026 | – | 6.48 | 5 |
| 81 | $CH_3$ | $CH_2CH{=}CH_2$ | – | 6.19 | 6.463 | 6.771 | 0.273 | 0.310 | 7.03 | 3 |
| 82 | $CH_3$ | $(CH_2)_3CH_3$ | – | 5.62 | 6.696 | 6.964 | 1.076 | 1.151 | 6.14 | 4 |
| 83 | $CH_3$ | $(CH_2)_5CH_3$ | – | 4.91 | 6.817 | 6.138 | 1.907 | 2.054 | 5.68 | 5 |
| 84 | $CH_3$ | $(CH_2)_2C_6H_5$ | – | 5.10 | 5.831 | 6.194 | 0.731 | 1.038 | 6.16 | 5 |
| [a]85 | $CH_2CH_3$ | $CH_2CH_3$ | – | 7.82 | 5.983 | – | −1.837 | – | 8.61 | 4 |
| 86 | $CH_2CH_3$ | $(CH_2)_2CH_3$ | – | 6.89 | 6.228 | 6.317 | −0.662 | −0.696 | 7.20 | 3 |
| 87 | $CH_2CH_3$ | $(CH_2)_3CH_3$ | – | 6.37 | 6.319 | 5.092 | −0.051 | −0.053 | 6.55 | 5 |
| 88 | $CH_2CH_3$ | $CH_2C_6H_5$ | – | 6.17 | 5.285 | 7.363 | −0.885 | −1.078 | 6.86 | 5 |
| [a]89 | $CH_2CH_3$ | $CH_2CH_2OH$ | – | 6.14 | 8.486 | – | 2.346 | – | 6.25 | 4 |
| 90 | $CH_2CH_2CH_3$ | $CH_2CH(CH_3)_2$ | – | 6.70 | 7.296 | 6.876 | 0.596 | 0.663 | 6.74 | 4 |
| 91 | $CH_2CH_2CH_3$ | $CH_2CH_2CH_3$ | – | 7.22 | 6.900 | 2.915 | −0.320 | −0.344 | 8.47 | 3 |
| 92 | $4\text{-}NO_2$ | – | – | 2.90 | 2.914 | 2.325 | 0.014 | 0.015 | 5.93 | 3 |
| [a]93 | 4-Br | – | – | 2.77 | 2.362 | – | −0.408 | – | 4.44 | 4 |
| [a]94 | 4-CN | – | – | 2.84 | 2.363 | – | -0.477 | – | 4.76 | 3 |
| 95 | 4-Cl | – | – | 2.81 | 2.361 | 2.340 | −0.449 | −0.485 | 4.54 | 5 |
| 96 | 4-F | – | – | 2.63 | 2.362 | 2.344 | −0.268 | −0.290 | 4.54 | 5 |
| 97 | H | – | – | 2.58 | 2.362 | 2.791 | −0.218 | −0.236 | 4.32 | 4 |
| 98 | $4\text{-}CH_3$ | – | – | 2.68 | 2.782 | 2.836 | 0.102 | 0.111 | 4.69 | 4 |
| 99 | $4\text{-}OCH_3$ | – | – | 2.62 | 2.820 | 2.480 | 0.200 | 0.216 | 4.96 | 3 |
| 100 | 4-OH | – | – | 2.24 | 2.462 | 3.052 | 0.222 | 0.240 | 4.84 | 3 |
| 101 | $4\text{-}OC_2H_5$ | – | – | 2.65 | 3.021 | 3.175 | 0.371 | 0.402 | 5.08 | 4 |
| 102 | $4\text{-}OC_3H_7$ | – | – | 2.79 | 3.145 | 3.149 | 0.355 | 0.385 | 5.11 | 5 |
| 103 | $4\text{-}OC_4H_9$ | – | – | 2.78 | 3.119 | 3.430 | 0.339 | 0.369 | 5.50 | 4 |
| 104 | $4\text{-}C(CH_3)_3$ | – | – | 3.15 | 3.402 | 2.766 | 0.252 | 0.280 | 5.20 | 5 |
| 105 | $3\text{-}CH_3$ | – | – | 2.78 | 2.767 | 2.336 | −0.013 | −0.014 | 4.50 | 5 |
| 106 | 3-F | – | – | 2.67 | 2.361 | 2.324 | −0.309 | −0.334 | 4.45 | 3 |
| 107 | 3-Cl | – | – | 2.82 | 2.361 | 4.832 | −0.459 | −0.496 | 4.12 | 4 |
| 108 | $OCH(Et)_2$ | – | – | 4.66 | 4.768 | 4.742 | 0.108 | 0.172 | 5.78 | 4 |
| 109 | $CON(Et)_2$ | – | – | 4.60 | 4.733 | 2.340 | 0.133 | 0.142 | 5.65 | 3 |
| [a]110 | $CON(n\text{-}Pr)_2$ | – | – | 4.49 | 4.843 | – | 0.353 | – | 4.96 | 4 |
| [a]111 | $CON(Et)CH(Me)_2$ | – | – | 5.80 | 4.838 | – | −0.962 | – | 4.41 | 4 |
| [a]112 | $CON[CH(Me)_2]_2$ | – | – | 5.40 | 5.131 | – | −0.269 | – | 5.64 | 4 |
| 113 | $CON(CH_2CH_2OH)CH(Me)_2$ | – | – | 4.68 | 3.979 | 3.664 | −0.701 | −1.016 | 5.77 | 5 |
| 114 | $CON[(CH_2)_3OH]CH(Me)_2$ | – | – | 5.68 | 5.558 | 5.553 | −0.122 | −0.127 | 6.02 | 3 |
| 115 | $CON[(CH_2)_5OH]CH(Me)_2$ | – | – | 5.70 | 6.241 | 6.279 | 0.541 | 0.579 | 6.67 | 4 |
| 116 | $CON[(CH_2)_3COOH]CH(Me)_2$ | – | – | 4.72 | 5.861 | 6.753 | 1.141 | 2.033 | 6.18 | 5 |
| 117 | $CON[(CH_2)_4COOH]CH(Me)_2$ | – | – | 5.89 | 6.242 | 6.253 | 0.352 | 0.363 | 6.20 | 3 |
| 118 | $CON[(CH_2)_3NH_2]CH(CH_3)_2$ | – | – | 4.34 | 5.007 | 5.038 | 0.667 | 0.698 | 5.51 | 5 |
| 119 | $CON[(CH_2)_2$-2-Pyridyl$]CH(Me)_2$ | – | – | 5.89 | 5.758 | 3.929 | −0.132 | −1.961 | 7.21 | 4 |
| [a]120 | $COCH_3$ | – | – | 5.12 | 4.552 | – | −0.568 | – | 4.82 | 4 |
| 121 | $COCH_2CH_3$ | – | – | 4.80 | 4.722 | 4.715 | −0.078 | −0.085 | 4.72 | 3 |
| 122 | $COCH{=}CH_2$ | – | – | 4.02 | 4.065 | 4.069 | 0.045 | 0.049 | 4.75 | 4 |
| 123 | $COCF_3$ | – | – | 6.55 | 4.365 | 4.196 | −2.185 | −2.354 | 6.30 | 4 |
| 124 | $SO_2CH_3$ | – | – | 3.89 | 5.101 | 5.158 | 1.211 | 1.268 | 5.36 | 3 |

[a] Samples in test set. Total Sores: Surflex-Dock scores are expressed in $-\log_{10}(K_d)$ units to represent binding affinities.

initial populations are 200, genetic generations are 100, crossover probability is 0.5 and mutation probability is 0.01, the evaluation function is cross-validation correlation coefficient $Q^2$. Statistical model was obtained by multiple linear regressions (MLR). External predictive ability of model was evaluated by $Q^2_{ext}$ [39].

$$Q^2_{ext} = 1 - \sum_{i=1}^{test}\left(y_i - \hat{y}_i\right)^2 \Big/ \sum_{i=1}^{test}\left(y_i - \hat{y}_{tr}\right)^2 \quad (4)$$

### 2.2. Dataset and structural characterization

Here 124 NIs are studied (Scheme 1 and Table 1 for their skeletons and structures), whose molecular structures and biological values ($pEC_{50}$, effective concentration of the compound required to achieve 50% protection of MT-4 cells against the cytopathic effect of virus, listed in Table 2, taken from reference [16]). With the exclusion of various outliers, some different 2D-QSAR models are already obtained by respective use of molecular molar volume, instruction variables, calculated molar refractive index, calculated substituent hydrophobic parameters, octanol and water partitioning coefficient. However, the exclusion of the outliers must be taken carefully. So taking all samples in Tables 1–10 [16], into full consideration, we use the developed descriptors (HoVAIFA) to obtain 3D-QSAR models with totally good prediction results which are quite comparative to that obtained in paper [16]. 124 samples were randomly divided into training set (including 96 samples) and test set (comprised 28 samples which were outliers deleted in reference [16]).

## 3. Results and discussions

### 3.1. Docking

Fig. 1 illuminated hydrogen bonding interactions between amino acid residues (consisting of basic residue ARG118, ARG152, ARG292, acidic residue GLU119 and neutral residue TYR406) and ID

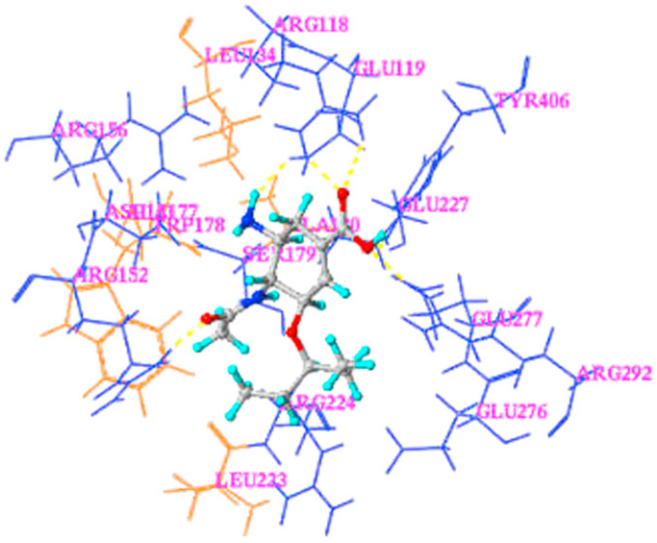



**Fig. 1.** Hydrogen bonding interactions between GS4701 (ID 8) and key amino acid residues.

8 (GS4071). 6 hydrogen bonds (dashed line) are produced. Moreover, types of hydrogen bonds included C═O···H–N, H–N···H–N, C–O···H–N, H–O···H–N, and C═O···H–O. From Fig. 1, hydrophobic interactions can form between alkyl groups, carbocyclic ring in NIs (ID 8) and hydrophobic residues including LEU134, ALA180, ALA177, TRP178, and LEU223. It is thus evident that hydrogen bonding interactions and hydrophobic interactions affect activity of NIs. Furthermore, CScore is 5. It is thus evident that docking was reasonable. Total scores (the predicted binding affinities) and CScore of all NIs are provided in Table 2. CScores of all samples are good.

Fig. 2 elucidated comparison of ID 8 (GS4701) and that observed in the crystal structure. RMSD (root mean squared deviation) was 0.66 Å, and similarity (a measure of similarity between solution coordinates and reference coordinates) was 0.72.

Moreover, A linear regression analysis reveals a fair correlation between experimental $pEC_{50}$ and binding affinities, with correlation coefficient $r = 0.813$ Eq. (5) and Fig. 3.

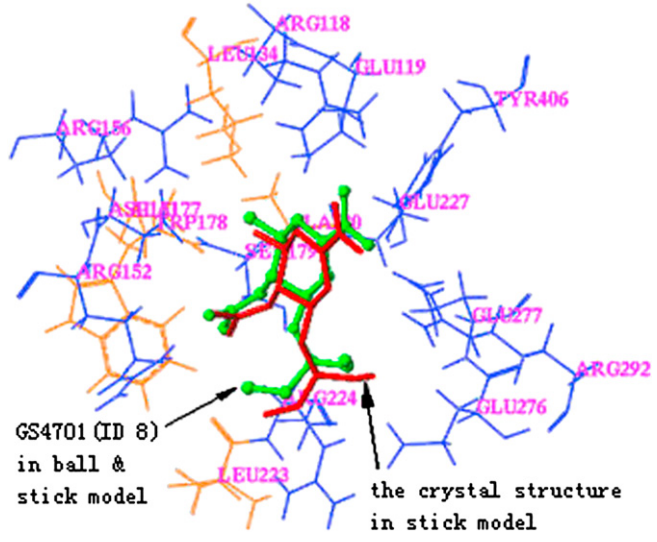



**Fig. 2.** Comparison of the position of GS4701 (ID 8) and that observed in the crystal structure.

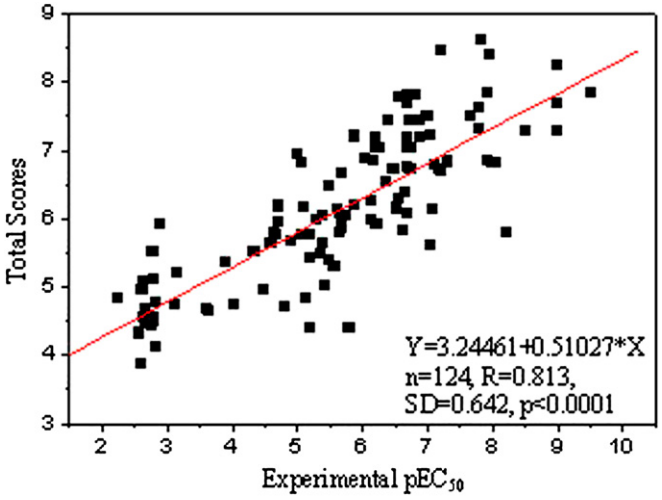



**Fig. 3.** Correlation of the predicted binding affinities (total scores) of 124 NIs with NA to experimental $pEC_{50}$.

$$Y = 3.24461(\text{Error}: 0.20129) + 0.51027(\text{Error}: 0.03305)^{*}X \quad (n = 124,\ R = 0.813,\ SD = 0.642,\ p < 0.0001) \tag{5}$$

### 3.2. QSAR

7 descriptors (including $V_{17}$, $V_{56}$, $V_{70}$, $V_{78}$, $V_{122}$, $V_{132}$ and $V_{148}$) were obtained by GA variable screening. The 3D-QSAR regression model with 7 variables has good estimation capacity ($R^2 = 0.789$, $SD = 0.832$) and the best predictive ability ($R^2_{CV} = 0.732$, $SD_{CV} = 0.936$), which is given below:

$$Y = 5.815 + 0.715^{*}V_{17} + 0.399^{*}V_{56} - 1.842^{*}V_{70} + 0.577^{*}V_{78} - 0.489^{*}V_{122} + 1.369^{*}V_{132} + 0.400^{*}V_{148}$$
$$n = 96,\ R^2 = 0.789,\ SD = 0.832,\ F = 46.901;\quad R^2_{CV} = 0.732,\ SD_{CV} = 0.936,\ F_{CV} = 34.338 \tag{6}$$

According to commonly recognition statistical standard, reliable model about QSAR is $q^2 \geq 0.5$ [40]. Therefore, the present model is indeed excellent with a predictive ability of 73.2%. Moreover, $Q^2_{ext}$ of the QSAR model were 0.705. Above results elucidated that the QSAR model was strong and had good external predictive ability.

In Eq. (6) $V_{17}$ is the electrostatic interaction between the second type of atoms ($C^{sp3}$) and the eighth type of atoms ($O^{sp3}$), $V_{56}$ is the steric interaction between first type of atoms ($H^{s1}$) and the first type of atoms ($H^{s1}$), $V_{70}$ indicates the steric interaction between the second type of atoms ($C^{sp3}$) and the sixth type of atoms ($N^{sp2}$), $V_{78}$ is the steric interaction between the third type of atoms ($C^{sp2}$) and the sixth type of atoms ($N^{sp2}$), $V_{122}$ is the hydrophobic interaction between the second type of atoms ($C^{sp3}$) and the third type of atoms ($C^{sp2}$), $V_{132}$ is the hydrophobic interaction between the third type of atoms ($C^{sp2}$) and the fifth type of atoms ($N^{sp3}$), $V_{148}$ is the hydrophobic interaction between the fifth type of atoms ($N^{sp3}$) and the eighth type of atoms ($O^{sp3}$). From QASR model, it can be concluded that the steric effect and hydrophobic interactions are the most important interactions, and the next is electrostatic interaction. Moreover, it can be seen that activity of NIs is high negatively correlated with the steric interaction between $sp^3$-hybridized C atoms ($C^{sp3}$) and $sp^2$-hybridized N atoms ($N^{sp2}$). Activity of NIs is high positively correlated with the hydrophobic

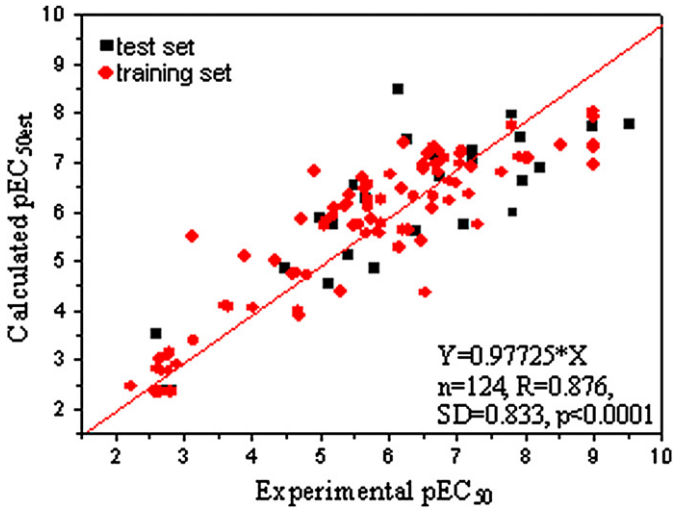



**Fig. 4.** Plot of calculated $pEC_{50est}$ vs. experimental $pEC_{50}$ of NIs in training set and test set.

interaction between the third type of atoms ($C^{sp2}$) and the fifth type of atoms ($N^{sp3}$). In addition, the addition of alkyl group X is helpful to improve activity, but alkyl chain should not be too long, because the steric interaction between unhybridized H atoms and unhybridized H atoms has a positive influence on the activity.

Fig. 4 showed plots of estimated activity value ($pEC_{50est}$) against observed $pEC_{50}$ values of samples in training set and test set. From Fig. 4, it can be seen that almost all samples are uniformly distributed around diagonal, not obviously exceptional point has selected. Furthermore, equation through origin was given below:

$$Y(\text{calculated } p\text{EC50}) = 0.97725(\text{error} : 0.01229)\,{}^{*}X(\text{experimental } p\text{EC50}) \quad (n = 124,\ R = 0.876,\ SD = 0.833,\ p < 0.0001) \tag{7}$$

It elucidates that 3D-HoVAIF can appropriately illustrate structural feature of compounds, and make the correct reflection in the statistical model. In addition, the estimated activity value ($pEC_{50-est}$), leave-one-out (LOO) predicted value ($pEC_{50loo}$) and respective error between the experimental and calculated values of 124 NIs are given in Table 2.

### 3.3. Comparison between QSAR and docking

From above results, it can be seen that the electrostatic interaction between the second type of atoms ($C^{sp3}$) and the eighth type of atoms ($O^{sp3}$) correspond to electrostatic interaction between ID 8 and residues ARG 224, ARG292 and TYR406. The hydrophobic interaction between the second type of atoms ($C^{sp3}$) and the third type of atoms ($C^{sp2}$) is conformity to hydrophobic interaction between ID 8 and residues LEU223 and ALA180; the hydrophobic interaction between the third type of atoms ($C^{sp2}$) and the fifth type of atoms ($N^{sp3}$) conforms to hydrophobic interaction between alkyl groups, carbocyclic ring in ID 8 and residue residues LEU134, ALA177, TRP178, and ALA180; the hydrophobic interaction between the fifth type of atoms ($N^{sp3}$) and the eighth type of atoms ($O^{sp3}$) consistent with hydrophobic interaction between alkyl groups in ID 8 and residue TRP178. In addition, from the docking, it can be concluded that the steric interactions including between first type of atoms ($H^{s1}$) and the first type of atoms ($H^{s1}$), between the second type of atoms ($C^{sp3}$) and the sixth type of atoms ($N^{sp2}$), between the third type of atoms ($C^{sp2}$) and the sixth type of atoms ($N^{sp2}$) can decide molecular shape. Therefore, QSAR and docking results illuminate that electrostatic, hydrophobic and steric interactions have an effect on activity of NIs in some degree. It is thus clear that QSAR results accord with docking results.

## 4. Conclusions

In this paper, the docking results elucidate that hydrogen bonds and hydrophobic interactions mainly affect bioactivity of NIs. Moreover, carboxyl group, amino (guanidino) group, amide group, hydroxy group are crucial to form hydrogen bonds interactions between NIs and key residues in active site. In addition, there is a strong correlation between binding affinity and experimental $pIC_{50}$ with significant correlation coefficient $r = 0.813$ and $p < 0.0001$.

The QSAR results elucidate that the steric effect, hydrophobic and electrostatic interactions affect activities of NIs. $R^2$ and $R^2_{CV}$ of the QSAR model are 0.789 and 0.732, respectively. Moreover, $Q^2_{ext}$ of 0.705 is obtained. The results show that the QSAR model is robust and has good predictive ability. The QSAR results coincide with docking results, which illustrates that HoVAIFA is an effective description methodology for characterization of the complex interactions of drug molecules. HoVAIFA parameters, having clear physicochemical meaning and not considering the superposition of conformation, are of easy interpretation and can be calculated directly with more advantages in modeling stability and predictive ability than traditional methods of molecular characterization. Therefore, HoVAIFA is worth further study and is expected to be widely used in the bioactivity prediction of various theoretical drugs and other diverse substances. QSAR and docking can supplement each other. Therefore, QSAR combination with docking is useful to drug design and optimation of lead compounds.

## Acknowledgements

The authors thank professor Gang Chen in National Technology College of Singapore for offering us docking results of the sybyl program and help. This work was supported by financial support from the national high technology research and development program of China (Grant No. 2006AA02Z312).

## Appendix. Supplementary data

Supplementary data associated with this article can be found in the online version, at doi:10.1016/j.ejmech.2009.11.043.

## References

[1] R.F. Service, Science 275 (1997) 756–757.
[2] M. Von Itzstein, W.Y. Wu, G.B. Kok, M.S. Pegg, J.C. Dyason, B. Jin, T.V. Phan, M.L. Smythe, H.F. White, S.W. Oliver, P.M. Colman, J.N. Varghese, D.M. Ryan, J.M. Woods, R.C. Bethell, V.J. Hotham, J.M. Cameron, C.R. Penn, Nature 363 (1993) 418–423.
[3] C.U. Kim, N. Bischofberger, M.A. Williams, W. Lew, J. Merson, C. Sweet, J. Antivir. Res. Soc. 34 (1997) A74.
[4] C.J. Dunn, K.L. Goa, Drugs 58 (1999) 761–784.
[5] P.W. Smith, I.D. Starkey, P.D. Howes, S.L. Sollis, S.P. Keeling, P.C. Cherry, M. von Itzstein, W.Y. Wu, B. Jin, J. Med. Chem 31 (1996) 143–150.
[6] R.A. Fromtling, Drugs Fut. 25 (2000) 430–433.
[7] S. Singh, M.J. Jedrzejas, G.M. Air, M. Luo, J. Med. Chem. 38 (1995) 3217–3225.
[8] P. Chand, Y.S. Babu, S. Bantia, N. Chu, L.B. Cole, P.L. Kotian, J. Med. Chem. 40 (1997) 4030–4052.
[9] P.D. Howes, A. Cleasby, D.N. Evans, H. Feilden, P.W. Smith, S.L. Sollis, N. Taylor, A.J. Wonacott, J. Med. Chem. 34 (1999) 225–234.
[10] M.A. Williams, W. Lew, D.B. Mendel, C.Y. Tai, P.A. Escarpe, W.G. Laver, R.C. Stevens, C.U. Kim, J. Am. Chem. Soc. 7 (1997) 1837–1842.
[11] D.B. Mendel, C.Y. Tai, P.A. Escarpe, W.X. Li, C.U. Kim, M.A. Williams, W. Lew, L. Zhang, N. Bischofberger, J.H. Huffman, J. Med. Chem. 41 (1998) 2451–2460.

[12] W.X. Li, P.A. Escarpe, E.J. Eisenberg, K.C. Cundy, C. Sweet, K.J. Jakeman, J. Merson, W. Lew, M. Williams, L.J. Zhang, C.U. Kim, N. Bischofberger, M.S. Chen, D.B. Mendel, Antimicrobial. Agents Chemother. 42 (1998) 647–653.
[13] J.A.L. Ives, J.A. Carr, D.B. Mendel, C.Y. Tai, R. Lambkin, L. Kelly, J.S. Oxford, F.G. Hayden, N.A. Roberts, J. Am. Med. Assoc. 7 (2002) 307–317.
[14] Z.Y. Liu, Y.G. Wu, Z.R. Li, Chin J Antibiot 27 (2002) 529–531.
[15] L.A. Sorbera, Drugs Fut 25 (2000) 249–251.
[16] R.P. Verma, C. Hansch, Bioorg. Med. Chem. 14 (2006) 982–996.
[17] J. Drews, Science 287 (2000) 1960–1964.
[18] S.S. Liu, S.C. Cai, C.Z. Cao, Z.S. Li, J. Chem. Inf. Comput. Sci. 40 (2000) 1337–1348.
[19] R.J. Russell, L.F. Haire, D.J. Stevens, P.J. Collins, Y.P. Lin, G.M. Blackburn, A.J. Hay, S.J. Gamblin, J.J. Skehel, Nature 443 (2006) 45–49.
[20] A.N. Jain, J. Med. Chem. 46 (2003) 499–511.
[21] J. Ruppert, W. Welch, A.N. Jain, Protein Sci. 6 (1997) 524–533.
[22] R.D. Clark, A. Strizhev, J.M. Leonard, J.F. Blake, J.B. Matthew, J. Mol. Graph. Model 20 (2002) 281.
[23] G. Jones, P. Willet, R.C. Glen, A.R. Leach, R. Taylor, J. Mol. Biol. 267 (1997) 727.
[24] I. Muegge, Y.C. Martin, J. Med. Chem. 42 (1999) 791–804.
[25] I.D. Kuntz, J.M. Blaney, S.J. Oatley, R. Langridge, T.E. Ferrin, J. Mol. Biol. 161 (1982) 269–288.
[26] M.D. Eldridge, C.W. Murray, T.R. Auton, G.V. Paolini, R.P. Mee, J. Comput. Aided. Mol. Des. 11 (1997) 425–445.
[27] C. Zhao, C. Maring, M. Sun, K. Stewart, V. Stoll, Y. Xu, Y. Gu, A. Krueger, T. Hcrrin, H. Sham, W.G. Laver, D. Madigan, A. Kennedy, W. Karl, D. Montgomery, A. Saldivar, D. Kempf, W. Kohlbrenner, J. Med. Chem. 44 (2001) 1192–1201.
[28] M. Levitt, J. Mol. Biol 170 (1983) 723–764.
[29] M. Levitt, M.F. Perutz, J. Mol. Biol. 201 (1988) 751–754.
[30] G.E. Kellogg, J.C. Burnett, D.J. Abraham, Comput. Aided Mol. Des. 15 (2001) 381–393.
[31] G.E. Kellogg, S.F. Semus, D.J. Abraham, Comput. Aided. Mol. Des. 5 (1991) 545–552.
[32] M. Hahn, Med. Chem. 38 (1995) 2080–2090.
[33] D.J. Abraham, G.E. Kellogg, J.M. Holt, G.K. Ackers, Mol. Biol. 272 (1997) 613–632.
[34] G.T. Gunnarsson, R.D. Umesh, Biorg. Med. Chem. 12 (2004) 633–640.
[35] G.E. Kellogg, G.S. Joshi, D.J. Abraham, Med. Chem. Res. 1 (1992) 444–453.
[36] G.E. Kellogg, D.J. Abraham, Mol. Graph 10 (1992) 212–217.
[37] W. Hasel, T.F. Hendrikson, W.C. Still, Tetrahed. Comp. Method 1 (1988) 103–116.
[38] J. Pei, Q. Wang, J. Zhou, L. Lai, Proteins 57 (2004) 651–664.
[39] A. Golbraikh, A. Tropsha, J. Mol. Graph. Model. 20 (2002) 269–276.
[40] M. Clark, R.D. Crame, The probability of chance correlation using partial least squares (PLS). Quant. Struct. Act. Relat. 2 (1993) 137–145.