Investigation into Adamantane-Based M2 Inhibitors with FB-QSAR

Hang Wei[1], Cheng-Hua Wang[2], Qi-Shi Du[3,5,*], Jianzong Meng[3,4] and Kuo-Chen Chou[5]

[1]*Editorial Board, Journal of Guangxi University, Guangxi University, Nanning, Guangxi, 530004, China;* [2]*College of Life Science and Pharmaceutical Engineering, Nanjing University of Technology, Nanjing, 210009, China;* [3]*College of Life Science and Technique, Guangxi University, Nanning, Guangxi, 530004, China;* [4]*Nanning Fermentation and Enzyme Engineering Research and Technique Center, Nanning , 530003, China;* [5]*Gordon Life Science Institute, San Diego, California 92130, USA*

Abstract: Because of their high resistance rate to the existing drugs, influenza A viruses have become a threat to human beings. It is known that the replication of influenza A viruses needs a pH-gated proton channel, the so-called M2 channel. Therefore, to develop effective drugs against influenza A, the most logic strategy is to inhibit the M2 channel. Recently, the atomic structure of the M2 channel was determined by NMR spectroscopy (Schnell, J.R. and Chou, J.J., Nature, 2008, 451,591-595). The high-resolution NMR structure has provided a solid basis for structure-based drug design approaches. In this study, a benchmark dataset has been constructed that contains 34 newly-developed adamantane-based M2 inhibitors and covers considerable structural diversities and wide range of bioactivities. Based on these compounds, an in-depth analysis was performed with the newly developed fragment-based quantitative structure–activity relationship (FB-QSAR) algorithm. The results thus obtained provide useful insights for dealing with the drug-resistant problem and designing effective adamantane-based antiflu drugs.

Key Words: FB-QSAR, influenza A, M2 protein, amantadine, rimantadine, drug design.

I. INTRODUCTION

The outbreak of H5N1 avian influenza virus (AIV) in Asia, Europe and Africa, with rapid spread and high mortality rate in human, has raised extensive concerns about a pending global pandemic [1-4]. Even more seriously is that the rapid mutation rate of the viruses will make it very difficult to contain the influenza outbreak, just like the Hispanic case in 1918 and the Asian cases in 1957 and 1968. In the early pandemic phase, the only treatments so far available are with the neuraminidase inhibitors (NIs), oseltamivir and zanamivir [3,5,6], or the M2 channel inhibitors (MIs), amantadine and rimantadine [7,8].

As an integral membrane protein in the viral lipid envelope, M2 is a pH-gated proton channel that consists of four tightly packed transmembrane (TM) helices [9]. The M2 channel plays a key role in release of viral nucleoproteins and preventing premature conformational rearrangement of newly synthesized haemagglutinin during transport to the cell surface. Blocking its proton conductance could effectively control the viral replication [10,11]. The adamantane-based drugs, amantadine and rimantadine [7,8], have been used as first-choice antiviral drugs against community outbreaks of influenza A for many years. But, recent reports show that the resistance to these adamantane-based drugs in humans, birds and pigs has reached more than 90% [12]. Therefore, we are challenged to develop new potent MIs (M2 channel inhibitors) against the terrible viruses.

*Address correspondence to this author at the College of Life Science and Technique, Guangxi University, Nanning, Guangxi, 530004, China; Tel/Fax: 086-771-327-3843/086-771-323-8107;
E-mail: qishi_du@yahoo.com.cn

Recently the long-sought 3D (three-dimensional) structure of the M2 proton channel was successfully determined by high-resolution (nuclear magnetic resonance) spectroscopy [9]. Such a milestone work has provided us with a solid foundation to deal with the drug-resistant problem and develop new potent drugs against the influenza viruses [13].

Many lines of evidences have indicated that computational approaches, such as structural bioinformatics [14-18], molecular docking [19-24], molecular packing [25,26], pharmacophore modeling [27,28], Mote Carlo simulated annealing approach [29], protein subcellular location prediction [30,31], identification of membrane proteins and their types [32], identification of enzymes and their functional classes [33], identification of GPCR and their types [34,35], identification of proteases and their types [36,37], protein cleavage site prediction [38-40], signal peptide prediction [41,42], and QSAR (quantitative structure–activity relationship) [43-51], can timely provide very useful information or insights for both basic research and drug development and hence are widely welcome by science community. The present study is attempted to apply the newly developed fragment-based QSAR, or FB-QSAR, to investigate the M2 channel in hope to provide useful insights for developing efficient antiflu drugs.

II. MATERIALS AND METHODS

In recent years, a series of new adamantane-based compounds were synthesized by an iterative structure-based methods; some of the compounds thus obtained have shown equal or better efficacies in comparison with the original adamantane-based compounds: amantadine and rimantadine [7,8,52]. Many useful insights can be gained about the M2-inhibiting mechanism by studying the correlation of the

1573-4064/09 $55.00+.00

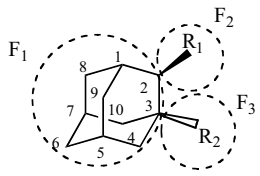
physicochemical properties of these compounds with their inhibition activities (see, e.g., [13]). However, to the best of our knowledge, no report has ever been seen in using QSAR to study these new adamantane-based compounds.

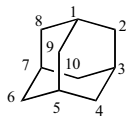
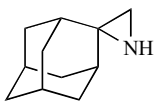
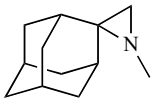
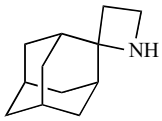
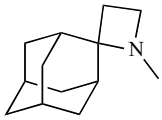
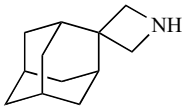
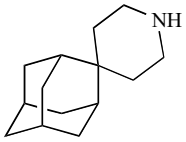
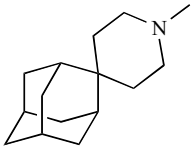
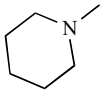
Listed in Table **1** are 34 adamantane-based M2 inhibitors against H3N2 influenza A virus that were collected from four different sources [7,8,52,53] in the last 10 years. Since the bioactivities from different sources were measured by using different methods and concentration units, it is necessary to normalize these values. In view of this, the following equation is introduced to normalize the bioactivity values:

$$\mathrm{pPIC50}_{\mathrm{sample}} = \frac{1}{\log(100 \times \mathrm{IC50}_{\mathrm{sample}}\eta)} \quad (1)$$

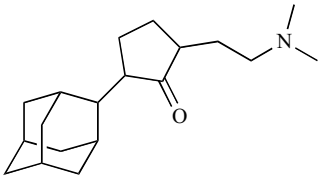
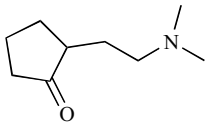
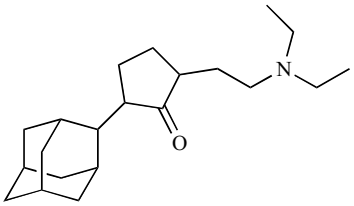
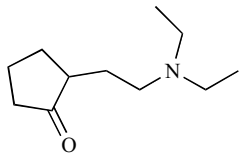
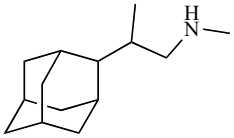
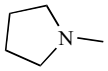
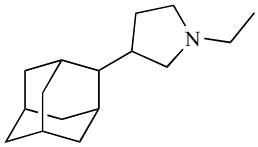
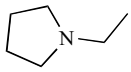
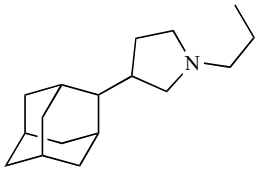
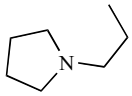
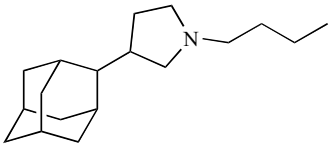
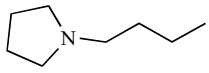
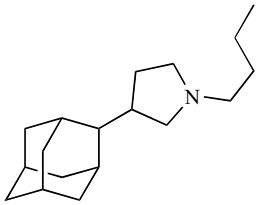
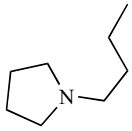
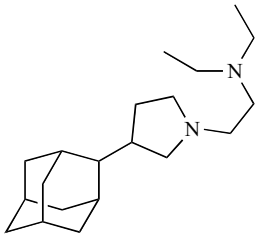
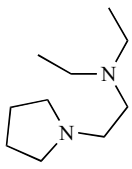
where $\mathrm{IC50}_{\mathrm{sample}}$ are the experimental bioactivities of MI (M2 channel inhibitor) samples, $\mathrm{pPIC50}_{\mathrm{sample}}$ the pseudo-pIC50 of the MI samples, η the scaling factor for different

Table 1. The Molecular Structures and Fragments of the 34 M2 Channel Inhibitors

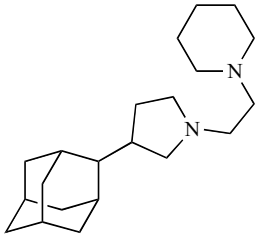
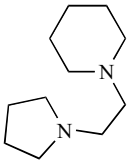
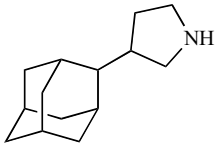
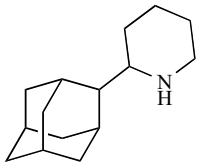
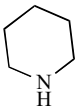
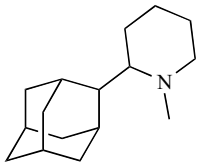
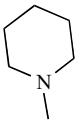
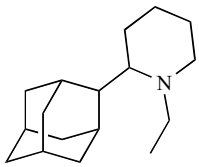
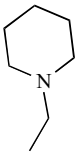
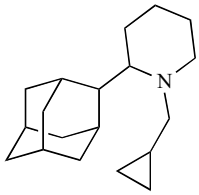
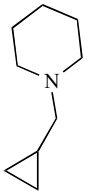
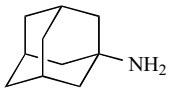
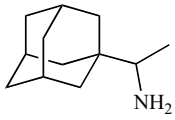
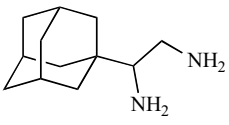
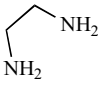


Mol	F_1	F_2	F_3	Expt pPIC50	code [1]
		(R_1 in F_2)	(R_2 in F_3)		
M01		R_1 : 	R_2 : H	0.4163	a
M02		R_1 : 	R_2 : H	0.4244	a
M03		R_1 : 	R_2 : H	0.7939	a
M04		R_1 : 	R_2 : H	0.6783	a
M05		R_1 : 	R_2 : H	0.8184	a
M06		R_1 : 	R_2 : H	1.1020	a
M07		R_1 : 	R_2 : H	0.4279	a

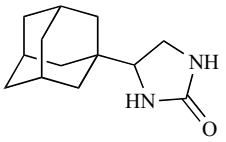
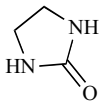
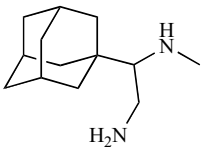
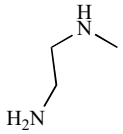
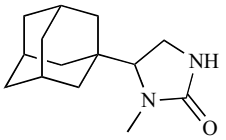
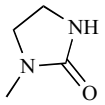
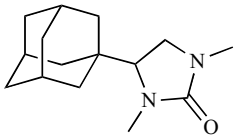
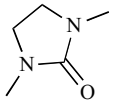
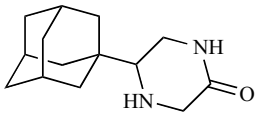
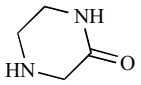
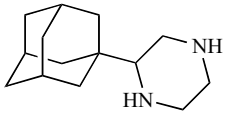
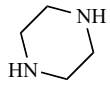
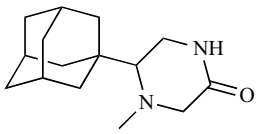
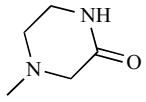
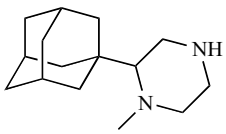
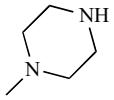
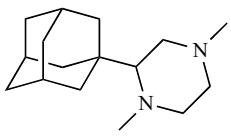
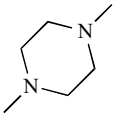
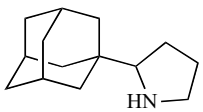
(Table 1. Contd….)

Mol	F_1	F_2	F_3	Expt pPIC50	code [(1)]
M08		R_1 : 	R_2 : H	0.4854	b
M09		R_1 : 	R_2 : H	0.5160	b
M10		R_1 : 	R_2 : H	0.4615	b
M11		R_1 : 	R_2 : H	0.4227	b
M12		R_1 : 	R_2 : H	0.4268	b
M13		R_1 : 	R_2 : H	0.4306	b
M14		R_1 : 	R_2 : H	0.5359	b
M15		R_1 : 	R_2 : H	0.4840	b

(Table 1. Contd….)

Mol	F_1	F_2	F_3	Expt pPIC50	code [1]
M16		R_1 : 	R_2 : H	2.2927	b
M17		R_1 : 	R_2 : H	0.6282	b
M18		R_1 : 	R_2 : H	0.3792	c
M19		R_1 : 	R_2 : H	0.5159	c
M20		R_1 : 	R_2 : H	0.8129	c
M21		R_1 : 	R_2 : H	0.5159	c
M22		R_1 : H_2	R_2 : NH_3	0.5000	a
M23		R_1 : H_2	R_2 : 	0.7924	a
M24		R_1 : H_2	R_2 : 	0.6364	d

(Table 1. Contd….)

Mol	F_1	F_2	F_3	Expt pPIC50	code [(1)]
M25		R_1 : H_2	R_2 : 	0.4609	d
M26		R_1 : H_2	R_2 	0.48467	d
M27		R_1 : H_2	R_2 : 	0.3651	d
M28		R_1 : H_2	R_2 : 	0.4277	d
M29		R_1 : H_2	R_2 : 	0.4402	d
M30		R_1 : H_2	R_2 : 	0.5914	d
M31		R_1 : H_2	R_2 : 	0.3236	d
M32		R_1 : H_2	R_2 : 	0.3841	d
M33		R_1 : H_2	R_2 : 	0.3504	d
M34		R_1 : H_2	R_2 : 	0.4953	d

(1) In Table **1**, "a" refers to [7]; "b" to [8]; "c" to [53]; and "d" to [52].

sources (Table **2**), and the coefficient 100 is for adjusting the numerical magnitudes to the region easier to handle.

In the four subsets of experimental data, the amantadine and rimantadine possess the key common components, and hence they should have the same bioactivity in the four sources. Therefore the scaling factor η were calculated based on the value of $IC50_{amantadine}$ or $IC50_{rimantadine}$. In the MI subset of [53], there is rimantadine but no amantadine. For consistency, the scaling factor η for the samples in the subset from [53] was adapted to:

$$\eta = \frac{1}{\mathrm{IC50}_{\mathrm{rimantadine}}} \times \frac{1}{3}\sum_{i=1}^{3}\left(\frac{\mathrm{IC50}_{\mathrm{rimantadine}}}{\mathrm{IC50}_{\mathrm{amantadine}}}\right) \approx \frac{1}{\mathrm{IC50}_{\mathrm{rimantadine}}} \times 0.248747 = \frac{1}{56} \times 0.248747 \approx 0.0044 \quad (2)$$

where the values of *i* (1, 2, and 3) corresponds to the index codes "a" [7], "b" [8], and "d" [52] of Table **1**, respectively. Actually the scaling factor of the subset from [53] was derived from the values of amantadine in the other three subsets.

As shown in Table **1**, each compound is divided into three fragments: F_1, F_2 and F_3. The F_2 and F_3 are the fragments containing substituent group R_1, and group R_2, respectively. The F_1 is the remaining part of the molecule after the fragments F_2 and F_3 are deleted. However, for computational convenience, we used the whole molecule for F_1. The substituents expand from hybrid rings (including nitrogen atoms or oxygen atoms) to alkyl groups, and hydrogen atoms.

In the current study, the newly-developed fragment-based QSAR algorithm [46], abbreviated as FB-QSAR, was adopted. FB-QSAR can provide more structural information for rational drug design and has been successfully used in building the predictive model of neuraminidase inhibitors for drug development against H5N1 influenza virus [43]. It involves the following four steps during practical applications: **(1)** choosing the fragments from the relevant molecular family to allocate the substituents; **(2)** calculating the descriptor (such as physicochemical property) for each of the selected fragments; **(3)** preparing input files for the training dataset and test dataset; and **(4)** developing the predictive model and using it to predict the bioactivity of the query sample.

It is essential for designing novel synthetic candidates to find a feasible formulation to reflect the target-inhibitor interactions [19,27,28]. In this regard, FB-QSAR has provided an effective model that depends on less descriptors but results in better prediction [43,46]. The less number of descriptors might also represent a more universal and robust model with less risk of over-fitting problem [54]. Accordingly, in using FB-QSAR the selection of molecular descriptors is important. The notations and physical meanings of the 20 descriptors in the current study are listed in Table **3**.

The fundamental equation of FB-QSAR is formulated as follows:

$$\sum_{k=1}^{M} b_k \left(\sum_{l=1}^{L} a_l p_{i,k,l}\right) = \mathrm{pK}_i \quad (i = 1,2,...,N) \quad (3)$$

where N is the number of molecule samples, M the number of fragments in the molecular family, L the number of physicochemical properties of fragments, a_l the sensitive coefficient of the l^{th} physicochemical property, b_k the weight coefficients of k^{th} fragment in the molecular family, $p_{i,k,l}$ the l-th property of fragment k in molecule i, and pK_i the bioactivity of i^{th} molecule. For the current study case where the number of molecular fragments is 3 and the number of molecular descriptors is 20, the FB-QSAR model in this study can be reduced to

$$\sum_{k=1}^{3} b_k \left(\sum_{l=1}^{20} a_l p_{i,k,l}\right) = \mathrm{pPIC50}_i^{\mathrm{train}} \quad (i = 1,2,\cdots,29) \quad (4)$$

where $\mathrm{pPIC50}_i^{\mathrm{train}}$ is the bioactivity of the i-th compound in the training dataset.

Beginning from an initial guess of the coefficients $\{b_k^{(0)}\}=1$, the coefficients $\{a_l\}$ and $\{b_k\}$ were solved using the double least square (IDLS) technique iteratively and alternately. The convergence criterion for the iterative procedure was set as follows [43,46]:

$$\left|Q^{(n+1)} - Q^{(n)}\right| = \left| \sqrt{\frac{1}{N}\sum_{i=1}^{N}(\mathrm{pPIC50}_i - \mathrm{pPIC50}_i^{(n+1)})^2} - \sqrt{\frac{1}{N}\sum_{i=1}^{N}(\mathrm{pPIC50}_i - \mathrm{pPIC50}_i^{(n)})^2} \right| \le \varepsilon \quad (\varepsilon = 1 \times 10^{-6}) \quad (5)$$

Table 2. The IC50s of Amantadine/ Rimantadine and Scaling Factors for Samples from Different Sources

Source[(1)]	IC50 Amantadine/Rimantadine	H Scaling Factor (1/IC50)	Scaling Degree IC50×η	Note
a	1.98	0.5051	1.0000	[7]
b	12.8	0.0781	0.9997	[8]
c	56[(2)]	0.0044[(3)]	0.2464	[53]
d	49.1	0.0204	1.0016	[52]

(1) See footnote (1) of Table **1**.

(2) 56 is the IC50 of rimantadine for code c, i.e. from [53].

(3) 0.0044 is the result of $1/IC50_{rimandatine}$ in [53] multiplied by 0.248747.

Table 3. The Symbols and Physic-Chemical Implications of the 20 Descriptors

Descriptor	Class Description	Physical Implication
SlogP	2D physical property	Log of the octanol/water partition coefficient (including implicit hydrogens).
E_sol	3D conformation	Solvation free energy.
E_ele	3D potential Energy	Electrostatic component of the potential energy.
E_vdw	3D potential Energy	van der Waals component of the potential energy.
Dipole	3D conformation dependent charge	Dipole moment calculated from the partial charges of the molecule.
ASA	3D surface area descriptor	Water accessible surface area.
Vdw_vol	2D physical property	Van der Waals volume.
Diameter	2D adjacency Descriptor	Largest vertex eccentricity in graph.
BalabanJ	2D adjacency Descriptor	Balaban averaged distance sum connectivity.
Apol	2D physical property	Sum of atomic polarizabilities.
FASA_H	3D conformation dependent charge	Fractional hydrophobic surface area.
FASA_P	3D conformation dependent charge	Fractional polar surface area.
$FCASA^+$	3D conformation dependent charge	Fractional charge-weighted positive surface area.
$FCASA^-$	3D conformation dependent charge	Fractional charge-weighted negative surface area.
WeinerPol	2D adjacency descriptor	Meiner polarity number.
A_count	2D physical property	Number of atoms.
A_nH	2D atom counts	Number of hydrogen atoms.
E_str	3D potential Energy	Bond stretch energy.
E_strain	3D potential Energy	E minus energy of local minimum.
logP(o/w)	2D physical property	Log octanol/water partition coefficient.

where $Q^{(n)}$ represents the square root of the summation of squared differences between the experimental bioactivities and the predicted bioactivities in the *n*-th step, and $Q^{(n+1)}$ is that in the (n+1)-th step. Then the coefficients $\{a_l\}$ and $\{b_k\}$, which are the solution of Eq.4, can be used to predict the bioactivities of query compounds *via* the following equation

$$\mathrm{pPIC50}_j^{\mathrm{pred}} = \sum_{k=1}^{3} b_k^{(n)} \Delta g_{j,k} = \sum_{l=1}^{20} a_l^{(n)} \Delta g_{j,l} = \sum_{k=1}^{3}\sum_{l=1}^{20} b_k^{(n)} a_l^{(n)} p_{j,k,l}$$

(j=1,2,…,5) (6)

where the term $b_k^{(n)}\Delta g_{j,k}$ is the contribution of fragment $F_{j,k}$ of the j-th molecule in the test dataset, $a_l^{(n)}\Delta g_{j,l}$ the contribution of physical property $P_{j,l}$, and $b_k^{(n)}a_l^{(n)}p_{j,k,l}$ the contribution of physical property $p_{j,k,j}$ of fragment k in molecule j, and n the iteration times.

To demonstrate a comparison of FB-QSAR with the other QSAR methods, the widely adopted 2D-QSAR [55] and Free-Wilson QSAR [56] were also applied for the same study. As pointed by Du *et al.* [43,46], the 2D-QSAR is a special case of FB-QSAR with the coefficients $\{b_k\}$ being fixed to 1, or that the whole molecule is treated as one fragment. Therefore the 2D-QSAR results can be conveniently obtained by just taking the first iterative results of FB-QSAR. In the current study, the 60 (3×20) adjustable variables were used for the Free-Wilson QSAR algorithm, while only 23 (20+3) variables used for the FB-QSAR algorithm.

RESULTS

All the results calculated using the methods described in the last section are listed in Table **4**, from which we can see that the results by the resubstitution test [57] on the training dataset with the Free-Wilson QSAR are R=1 and Q=0, indicating 100% correct. However, when it was operated on the test dataset, the Free-Wilson QSAR yielded the worst results of R=0.1093 and Q=2.7944. This means that the high correct resubstitution-rate by the Free-Wilson QSAR on the training

Table 4. The Calculation Results of Training Set and Test Set Using Three Methods: FB-QSAR, Free-Wilson QSAR, and Traditional 2D-QSAR

Training ID	Expt. $pPIC_{50}$	FB-QSAR	Free-Wilson QSAR	2D-QSAR
M01	0.4163	0.4236	0.4163	0.3779
M02	0.4244	0.4519	0.4244	0.5545
M03	0.7940	0.8161	0.7939	0.8216
M04	0.6783	0.6446	0.6783	0.6710
M06	1.1020	0.8139	1.1020	0.8682
M07	0.4279	0.4482	0.4279	0.3584
M08	0.4854	0.4635	0.4854	0.6529
M09	0.5160	0.5140	0.5160	0.5943
M10	0.4615	0.5239	0.4615	0.4154
M11	0.4227	0.4180	0.4227	0.5192
M13	0.4306	0.4005	0.4306	0.6317
M14	0.5359	0.5052	0.5359	0.3276
M15	0.4840	0.5272	0.4840	0.8317
M16	2.2927	2.2876	2.2927	1.8216
M17	0.6283	0.5771	0.6283	0.3649
M18	0.3792	0.6336	0.3792	0.5717
M19	0.5159	0.5535	0.5159	0.6699
M21	0.5159	0.5153	0.5159	0.5000
M23	0.7924	0.5781	0.7924	0.5203
M24	0.6364	0.6408	0.6364	0.6500
M25	0.4609	0.6367	0.4609	0.5985
M26	0.4847	0.6331	0.4847	0.3938
M27	0.3651	0.3769	0.3651	0.3522
M28	0.4277	0.3666	0.4277	0.3431
M29	0.4402	0.4009	0.4402	0.2509

(Table 4. Contd….)

Training ID	Expt. $pPIC_{50}$	FB-QSAR	Free-Wilson QSAR	2D-QSAR
M30	0.5914	0.6664	0.5914	0.9592
M32	0.3841	0.3356	0.3841	0.6302
M33	0.3504	0.3455	0.3504	0.0817
M34	0.4953	0.4387	0.4953	0.5784
R	---	0.9620	1.0000	0.8396
Q	---	0.0980	0.0000	0.1948
Test ID				
M05	0.8184	0.9508	-1.0536	1.0098
M12	0.4268	0.3166	-1.6125	0.3559
M20	0.8129	0.5720	4.1260	0.5795
M22	0.5000	0.6698	4.9406	3.8882
M31	0.3236	0.2106	1.1506	0.5673
R	---	0.8007	0.1093	0.6002
Q	---	0.1608	2.7944	1.5255

dataset is an over-fitting result with very poor cluster-tolerant capacity [58] and low predictive accuracy. This is because it contained 60 adjustable variables that were much more than the number of the samples involved. In contrast, for the same testing datasets, the FB-QSAR yielded the best prediction results of R=0.8007 and Q=0.1608.

The very high success rate achieved by the FB-QSAR approach is due to the iterative double least square (IDSL) technique, as demonstrated in (Fig. **1** and Fig. **2**). It can be seen from the two figures that the correlation coefficient R increases steadily with the iteration steps, while the trend of the predictive error Q is just the opposite.

FB-QSAR availed itself of the 20 descriptors to construct the successful model. The values of the 20 physicochemical descriptors are listed in Table **5**. The bioactivities of the query samples were computed using these values and Eq.6. The comparisons between FB-QSAR and the other two QSAR methods are shown in Fig. (**3**).

For facilitating analysis, the 34 compounds in Table **1** are classified into two groups. In group 1 the substituents R_1 are located in the position 2 (F_2 is in R_1), while in group 2 the substituents R_2 are in the position 3 (F_3 is in R_2). In order to identify the detailed fragment contributions, the values of fragment contributions of the seven compounds (M05, M06,

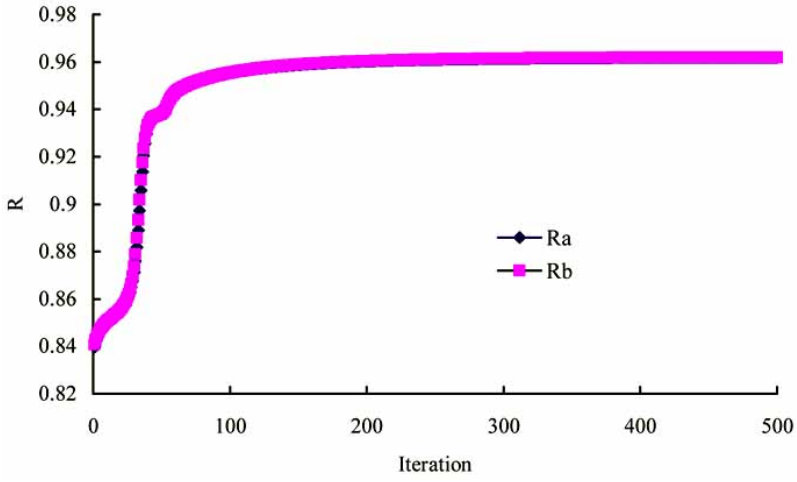


Fig. (1). The correlation coefficients between experimental and predicted bioactivities increase with the iterations. R_a is for $\{a_i^{(n)}\}$ iteration and R_b is for $\{b_j^{(n)}\}$ iteration. The R_a and R_b are getting close with the iterations. The first value $R_a^{(0)}$=0.839575 is the result of standard 2D-QSAR and the converged value $R_a^{(n)}$=0.961976 is the result of FB-QSAR. The remarkable improvement is from the better theoretical model and advanced mathematical technique of FB-QSAR.

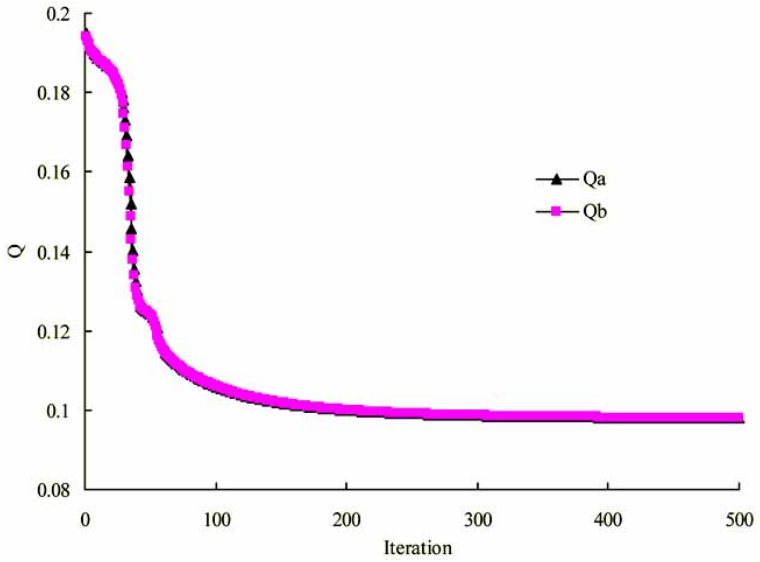


Fig. (2). The errors between predicted bioactivities and experimental bioactivities of NA inhibitors decrease with the iterations. The Q is the average square root of the summation of squared differences between predicted bioactivities and experimental bioactivities (cf. Eq.10). Q_a is for $\{a_l^{(n)}\}$ iteration and Q_b is for iteration. The R_a and R_b are getting close with the iterations. The first value $Q_a^{(0)}$=0.194843 is the result of standard 2D-QSAR and the converged value $Q_a^{(n)}$=0.09958 is the result of FB-QSAR. The remarkable improvement is from the better theoretical model and advanced mathematical technique of FB-QSAR.

M12, M16, M20, M23, and M31) are compiled in Table **6**. Among them the molecules M05, M06, M12, M16, and M20 belong to group 1, while the M23 and M31 belong to group 2.

It is instructive to point out that the contributions from fragments F_3 of the five molecules in group 1 are the same, and so are the contributions from fragments F_2 of the two molecules in group 2. This is because the fragments in each group have the same chemical structures.

For the seven compounds listed in Table **6**, the contributions of F_2 are generally larger than those of F_3. For example, the compound with the highest bioactivity, 2.2927, in group 1 is M16; while the corresponding compound in group 2 is M23 (rimantadine) whose bioactivity is only 0.7924. This suggests that the fragments in position 2 may be more sensitive than those in position 3. Rimantadine (M23) has a substituent ethylamine in position 3. If the ethylamine substituent is moved from position 3 to position 2, its bioactivity would increase to 4.2076 from the original value 0.7924 according to the current FB-QSAR model. Of course, such a predicted result needs further proofed by experiments.

Of the 34 compounds in Table **1**, M05, M12, M20, M22 and M31 were selected as the samples of the test dataset, and the remaining 29 as the samples of the training dataset. The criteria for selecting the above five test samples are: (1) they must be in two different groups; (2) their values must have a

Table 5. Coefficients of 20 Descriptors* in FB-QSAR Model

Descriptor	Coefficient	Descriptor	Coefficient
Diam	0.76899	FASA_P	0.07735
BalabJ	-0.02396	$FCASA^+$	-0.47807
E_ele	0.08539	$FCASA^-$	-0.02779
E_sol	-0.24497	Vdw_vol	-0.71044
E_vdw	-0.02085	WeinerPol	0.09033
Apol	0.04015	A_count	-0.21632
Dipole	0.03179	A_nH	0.03209
SlogP	0.09165	E_str	-0.05189
FASA_H	0.08508	E_strain	0.12689

* The physical meanings of the descriptors can be found in Table **2**.

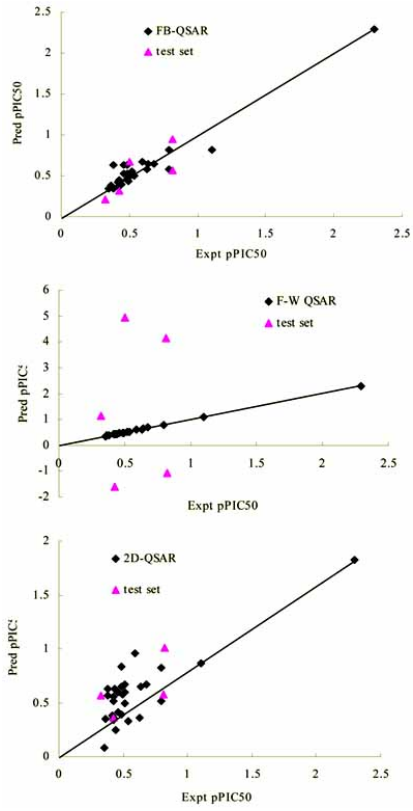


Fig. (3). Correlations of the experimental activity (pPIC50) versus the predicted activity (the data for the training set are marked as ◆, and test set as▲) derived from FB-QSAR, Free-Wilson QSAR and 2D-QSAR, respectively. For the Free-Wilson QSAR model, the predicted values of training set are exactly the same as the experimental data, but the predicted values in the test set nearly show no relationship with the experimental data. For the FB-QSAR model, all the predicted data concentrate on the diagonal line zone, shows the least deviation. Compared with the above two, 2D-QSAR model is better than Free-Wilson QSAR, but worse than the FB-QSAR.

larger span. The tested outcomes obtained by means of the three QSAR approaches are listed in Table **4**, from which we can see the FB-QSAR approach yields the best predictive results.

DISCUSSION

In the area of computer-aided drug design (CADD), QSAR is one of the most widely used approaches. However, chemists are often disappointed by the poor predicted results and the lack of physicochemical explanations for the models they used. The reason is that the structure-activity optimization surface is not as smooth as originally anticipated. The chemical space, described by QSAR, is filled of potential cliffs. To describe the potential cliffs in a chemical space from all aspects, we not only need better physicochemical descriptors, but also need more structural descriptors. In this sense, FB-QSAR provides an effective and feasible means to deal with this kind of problems. Moreover, the iterative double least square (IDLS) technique can help the FB-QSAR to

Table 6. Fragment Contributions of 7 Molecules (M05, M06, M12, M16, M20, M22, M23, M31 and M33)

	Molecule	F1	F2	F3	Pred.	Expt.
Group 1	M05	1.1054	2.0998	0.0437	0.9508	0.8184
	M06	1.2336	2.0912	0.0437	0.8139	1.1020
	M12	1.1959	1.5561	0.0437	0.3166	0.4268
	M16	1.0226	3.3538	0.0437	2.2876	2.2927
	M20	1.4612	2.0769	0.0437	0.5720	0.8129
Group 2	M23	1.1372	1.8429	0.1275	0.5781	0.7924
	M31	1.6145	1.8429	0.0178	0.2106	0.3236

avoid the "over-fitting" problem often encountered in using the other QSAR methods.

CONCLUSION

It was found through the current study that, among the existing QSAR methods, the FB-QSAR not only can make the best predicted results for the adamantane-based MIs in inhibiting influenza A virus, but also can provide detailed structural information for the rational drug design. The FB-QSAR is a powerful tool for the fragment-based drug design. As revealed by the above fragment analysis (cf. Table **1** and Table **6**), position 2 is more sensitive than position 3. Therefore, in conducting the adamantane-based antiflu drug design, our target should be focused on the fragment F_2 in position 2.

ACKNOWLEDGEMENTS

The authors wish to express their sincere gratitude to the three anonymous Reviewers, whose constructive comments were very helpful for strengthening the presentation of this paper. This work is financially supported by the International Science and Technology Cooperation Project 2008DFA30710 and by the Chinese National Science Foundation (NSFC).

REFERENCES

[1] Ferguson, N.M.; Cummings, D.A.; Cauchemez, S.; Fraser, C.; Riley, S.; Meeyai, A.; Iamsirithaworn, S.; Burke, D. S. Strategies for containing an emerging influenza pandemic in Southeast Asia. *Nature*, **2005**, *437*, 209-214.

[2] Wei, D.Q.; Du, Q.S.; Sun, H.; Chou, K. C. Insights from modeling the 3D structure of H5N1 influenza virus neuraminidase and its binding interactions with ligands. *Biochem. Biophys. Res. Comm.*, **2006**, *344*, 1048-1055.

[3] Wang, S.Q.; Du, Q.S.; Chou, K. C. Study of drug resistance of chicken influenza A virus (H5N1) from homology-modeled 3D structures of neuraminidases. *Biochem. Biophys. Res. Comm.*, **2007**, *354*, 634-640.

[4] Du, Q.S.; Wang, S.Q.; Chou, K. C. Analogue inhibitors by modifying oseltamivir based on the crystal neuraminidase structure for treating drug-resistant H5N1 virus. *Biochem. Biophys. Res. Comm.*, **2007**, *362*, 525-531.

[5] Lew, W.; Chen, X.; Kim, C. U. Discovery and development of GS 4104 (oseltamivir): an orally active influenza neuraminidase inhibitor. *Curr. Med. Chem.*, **2000**, *7*, 663-672.

[6] Dunn, C.J.; Goa, K. L. Zanamivir: a review of its use in influenza. *Drugs*, **1999**, *58*, 761-784.

[7] Kolocouris, N.; Zoidis, G.; Foscolos, G.B.; Fytas, G.; Prathalingham, S.R.; Kelly, J.M.; Naesens, L.; De Clercq, E. Design and synthesis of bioactive adamantane spiro heterocycles. *Bioorg. Med. Chem. Lett.*, **2007**, *17*, 4358-4362.

[8] Stamatiou, G.; Kolocouris, A.; Kolocouris, N.; Fytas, G.; Foscolos, G.B.; Neyts, J.; De Clercq, E. Novel 3-(2-adamantyl)pyrrolidines with potent activity against influenza A virus-identification of aminoadamantane derivatives bearing two pharmacophoric amine groups. *Bioorg. Med. Chem. Lett.*, **2001**, *11*, 2137-2142.

[9] Schnell, J.R.; Chou, J. J. Structure and mechanism of the M2 proton channel of influenza A virus. *Nature*, **2008**, *451*, 591-595.

[10] Hay, A.J.; Wolstenholme, A.J.; Skehel, J.J.; Smith, M. H. The molecular basis of the specific anti-influenza action of amantadine. *EMBO J.*, **1985**, *4*, 3021-3024.

[11] Wang, C.; Takeuchi, K.; Pinto, L.H.; Lamb, R. A. Ion channel activity of influenza A virus M2 protein: characterization of the amantadine block. *J. Virol.*, **1993**, *67*, 5585-5594.

[12] Deyde, V.M.; Xu, X.; Bright, R.A.; Shaw, M.; Smith, C.B.; Zhang, Y.; Shu, Y.; Gubareva, L.V.; Cox, N.J.; Klimov, A. I. Surveillance of resistance to adamantanes among influenza A(H3N2) and A(H1N1) viruses isolated worldwide. *J. Infect. Dis.*, **2007**, *196*, 249-257.

[13] Huang, R.B.; Du, Q.S.; Wang, C.H.; Chou, K. C. An in-depth analysis of the biological functional studies based on the NMR M2 channel structure of influenza A virus. *Biochem. Biophys. Res. Comm.*, **2008**, *377*, 1243-1247.

[14] Chou, K. C. Review: structural bioinformatics and its impact to biomedical science. *Curr. Med. Chem.*, **2004**, *11*, 2105-2134.

[15] Chou, K. C. Modelling extracellular domains of GABA-A receptors: subtypes 1, 2, 3, and 5. *Biochem. Biophys. Res. Commun.*, **2004**, *316*, 636-642.

[16] Chou, K. C. Molecular therapeutic target for type-2 diabetes. *J. Proteome Res.*, **2004**, *3*, 1284-1288.

[17] Chou, K. C. Insights from modelling the 3D structure of the extracellular domain of alpha7 nicotinic acetylcholine receptor. *Biochem. Biophys. Res. Commun.*, **2004**, *319*, 433-438.

[18] Chou, K. C. Coupling interaction between thromboxane A2 receptor and alpha-13 subunit of guanine nucleotide-binding protein. *J. Proteome Res.*, **2005**, *4*, 1681-1686.

[19] Chou, K.C.; Wei, D.Q.; Zhong, W. Z. Binding mechanism of coronavirus main proteinase with ligands and its implication to drug design against SARS. (Erratum: ibid., 2003, Vol.310, 675). *Biochem. Biophys. Res. Comm.*, **2003**, *308*, 148-151.

[20] Li, Y.; Wei, D.Q.; Gao, W.N.; Gao, H.; Liu, B.N.; Huang, C.J.; Xu, W.R.; Liu, D.K.; Chen, H.F.; Chou, K. C. Computational approach to drug design for oxazolidinones as antibacterial agents. *Med. Chem.*, **2007**, *3*, 576-582.

[21] Wang, J.F.; Wei, D.Q.; Chen, C.; Li, Y.; Chou, K. C. Molecular modeling of two CYP2C19 SNPs and its implications for personalized drug design. *Protein Pept. Lett.*, **2008**, *15*, 27-32.

[22] Zhang, R.; Wei, D.Q.; Du, Q.S.; Chou, K. C. Molecular modeling studies of peptide drug candidates against SARS. *Med. Chem.*, **2006**, *2*, 309-314.

[23] Gao, W.N.; Wei, D.Q.; Li, Y.; Gao, H.; Xu, W.R.; Li, A.X.; Chou, K. C. Agaritine and its derivatives are potential inhibitors against HIV proteases. *Med. Chem.*, **2007**, *3*, 221-226.

[24] Zheng, H.; Wei, D.Q.; Zhang, R.; Wang, C.; Wei, H.; Chou, K. C. Screening for New Agonists against Alzheimer's Disease. *Med. Chem.*, **2007**, *3*, 488-493.

[25] Chou, K.C.; Nemethy, G.; Scheraga, H. A. Energetic approach to packing of a-helices: 2. General treatment of nonequivalent and nonregular helices. *J. Am. Chem. Soc.*, **1984**, *106*, 3161-3170.

[26] Chou, K.C.; Maggiora, G.M.; Nemethy, G.; Scheraga, H. A. Energetics of the structure of the four-alpha-helix bundle in proteins. *Proc. Nat. Acad. Sci. USA*, **1988**, *85*, 4295-4299.

[27] Sirois, S.; Wei, D.Q.; Du, Q.S.; Chou, K. C. Virtual Screening for SARS-CoV Protease Based on KZ7088 Pharmacophore Points. *J. Chem. Inf. Comput. Sci.*, **2004**, *44*, 1111-1122.

[28] Chou, K.C.; Wei, D.Q.; Du, Q.S.; Sirois, S.; Zhong, W. Z. Review: Progress in computational approach to drug development against SARS. *Curr. Med. Chem.*, **2006**, *13*, 3263-3270.

[29] Chou, K. C. Energy-optimized structure of antifreeze protein and its binding mechanism. *J. Mol. Biol.*, **1992**, *223*, 509-517.

[30] Chou, K.C.; Shen, H. B. Cell-PLoc: A package of web-servers for predicting subcellular localization of proteins in various organisms. *Nat. Protoc.*, **2008**, *3*, 153-162.

[31] Chou, K.C.; Shen, H. B. Review: Recent progresses in protein subcellular location prediction. *Anal. Biochem.*, **2007**, *370*, 1-16.

[32] Chou, K.C.; Shen, H. B. MemType-2L: A Web server for predicting membrane proteins and their types by incorporating evolution information through Pse-PSSM. *Biochem. Biophys. Res. Commun.*, **2007**, *360*, 339-345.

[33] Shen, H.B.; Chou, K. C. EzyPred: A top-down approach for predicting enzyme functional classes and subclasses. *Biochem. Biophys. Res. Commun.*, **2007**, *364*, 53-59.

[34] Chou, K. C. Prediction of G-protein-coupled receptor classes. *J. Proteome Res.*, **2005**, *4*, 1413-1418.

[35] Xiao, X.; Wang, P.; Chou, K. C. GPCR-CA: A cellular automaton image approach for predicting G-protein-coupled receptor functional classes. *J. Comput. Chem.*, **2008**, DOI 10.1002/jcc.21163.

[36] Chou, K.C.; Shen, H. B. ProtIdent: a web server for identifying proteases and their types by fusing functional domain and sequential evolution information. *Biochem. Biophys. Res. Comm.*, **2008**, *376*, 321-325.

[37] Shen, H.B.; Chou, K. C. Identification of proteases and their types. *Anal. Biochem.*, **2009**, *385*, 153-160.

[38] Chou, K. C. A vectorized sequence-coupling model for predicting HIV protease cleavage sites in proteins. *J. Biol. Chem.*, **1993**, *268*, 16938-16948.

[39] Chou, K. C. Review: Prediction of HIV protease cleavage sites in proteins. *Anal. Biochem.*, **1996**, *233*, 1-14.

[40] Shen, H.B.; Chou, K. C. HIVcleave: a web-server for predicting HIV protease cleavage sites in proteins. *Anal. Biochem.*, **2008**, *375*, 388-390.

[41] Chou, K.C.; Shen, H. B. Signal-CF: a subsite-coupled and window-fusing approach for predicting signal peptides. *Biochem. Biophys. Res. Commun.*, **2007**, *357*, 633-640.

[42] Shen, H.B.; Chou, K. C. Signal-3L: a 3-layer approach for predicting signal peptide. *Biochem. Biophys. Res. Commun.*, **2007**, *363*, 297-303.

[43] Du, Q.S.; Huang, R.B.; Chou, K. C. Review: recent advances in QSAR and their applications in predicting the activities of chemical molecules, peptides and proteins for drug design. *Curr. Protein Pept. Sci.*, **2008**, *9*, 248-259.

[44] Du, Q.S.; Mezey, P.G.; Chou, K. C. Heuristic Molecular Lipophilicity Potential (HMLP): A 2D-QSAR Study to LADH of Molecular Family Pyrazole and Derivatives. *J. Comput. Chem.*, **2005**, *26*, 461-470.

[45] Du, Q.S.; Huang, R.B.; Wei, Y.T.; Du, L.Q.; Chou, K. C. Multiple Field Three Dimensional Quantitative Structure-Activity Relationship (MF-3D-QSAR). *J. Comput. Chem.*, **2008**, *29*, 211-219.

[46] Du, Q.S.; Huang, R.B.; Wei, Y.T.; Pang, Z.W.; Du, L.Q.; Chou, K. C. Fragment-Based Quantitative Structure-Activity Relationship (FB-QSAR) for Fragment-Based Drug Design. *J. Comput. Chem.*, **2009**, *30*, 295-304.

[47] Prado-Prado, F.J.; Gonzalez-Diaz, H.; de la Vega, O.M.; Ubeira, F.M.; Chou, K. C. Unified QSAR approach to antimicrobials. Part 3: First multi-tasking QSAR model for Input-Coded prediction, structural back-projection, and complex networks clustering of antiprotozoal compounds. *Bioorg. Med. Chem.*, **2008**, *16*, 5871-5880.

[48] Dea-Ayuela, M.A.; Perez-Castillo, Y.; Meneses-Marcel, A.; Ubeira, F.M.; Bolas-Fernandez, F.; Chou, K.C.; Gonzalez-Diaz, H. HP-Lattice QSAR for dynein proteins: Experimental proteomics (2D-electrophoresis, mass spectrometry) and theoretic study of a Leishmania infantum sequence. *Bioorg. Med. Chem.*, **2008**, *16*, 7770-7776.

[49] Cruz-Monteagudo, M.; Munteanu, C.R.; Borges, F.; Natália, M.; Cordeiro, D.S.; Uriarte, E.; Chou, K.C.; Gonzalez-Diaz, H. Stochastic molecular descriptors for polymers. 4. Study of complex mixtures with topological indices of mass spectra spiral and star networks: the blood proteome case. *Polymer*, **2008**, *49*, 5575-5587.

[50] Aguero-Chapin, G.; Antunes, A.; Ubeira, F.M.; Chou, K.C.; Gonzalez-Diaz, H. Comparative study of topological indices of macro/supra-molecular RNA complex networks. *J. Chem. Inf. Mod.*, **2008**, *48*, 2265-2277.

[51] Gonzalez-Diaz, H.; Gonzalez-Diaz, Y.; Santana, L.; Ubeira, F.M.; Uriarte, E. Proteomics, networks, and connectivity indices. *Proteomics*, **2008**, *8*, 750-778.

[52] Tataridis, D.; Fytas, G.; Kolocouris, A.; Fytas, C.; Kolocouris, N.; Foscolos, G.B.; Padalko, E.; Neyts, J.; De Clercq, E. Influence of an additional 2-amino substituent of the 1-aminoethyl pharmacophore group on the potency of rimantadine against influenza virus A. *Bioorg. Med. Chem. Lett.*, **2007**, *17*, 692-696.

[53] Kolocouris, A.; Tataridis, D.; Fytas, G.; Mavromoustakos, T.; Foscolos, G.B.; Kolocouris, N.; De Clercq, E. Synthesis of 2-(2-adamantyl)piperidines and structure anti-influenza virus A activity relationship study using a combination of NMR spectroscopy and molecular modeling. *Bioorg. Med. Chem. Lett.*, **1999**, *9*, 3465-3470.

[54] Wang, T.; Yang, J.; Shen, H.B.; Chou, K. C. Predicting membrane protein types by the LLDA algorithm. *Protein Pept. Lett.*, **2008**, *15*, 915-921.

[55] Hansch, C. M.R.M.; Fujita, T.; Maloney, P.P.; Geiger, F.; Streich, M. The correlation of biological activity of plant growth regulators and chloromycetin derivatives with Hammett constants and partition coefficients. *J. Am. Chem. Soc.*, **1963**, *85*, 2817-2824.

[56] Free, S. M., Jr.; Wilson, J. W. A mathematical contribution to structure-activity studies. *J. Med. Chem.*, **1964**, *7*, 395-399.

[57] Chou, K.C.; Zhang, C. T. Review: Prediction of protein structural classes. *Crit. Rev. Biochem. Mol. Biol.*, **1995**, *30*, 275-349.

[58] Chou, K. C. A key driving force in determination of protein structural classes. *Biochem. Biophys. Res. Commun.*, **1999**, *264*, 216-224.

Received: 20 January, 2009 **Revised: 18 March, 2009** **Accepted: 19 March, 2009**