Available online at www.sciencedirect.com
ScienceDirect
European Journal of Medicinal Chemistry 44 (2009) 625–631
EUROPEAN JOURNAL OF MEDICINAL CHEMISTRY
http://www.elsevier.com/locate/ejmech

Original article

QSAR studies for diarylpyrimidines against HIV-1 reverse transcriptase wild-type and mutant strains

Yong-Hong Liang, Fen-Er Chen*

Department of Chemistry, Fudan University, Shanghai 200433, PR China

Received 3 December 2007; received in revised form 19 February 2008; accepted 20 March 2008
Available online 8 April 2008

Abstract

For the activity values against HIV-1 RT wild-type and mutant strains (L100I, Y181C and Y188N) of 34 diarylpyrimidines (DAPYs) taken from literature, multiple linear regression analysis and the crossvalidation of Leave-One-Out method are carried out against the activity values with the hydrophobicity index log P, the modified steric parameter L_Y, the Muliken charge of nitrogen atom on the right wing N_C and the indicator index I for the substituents R_3' and R_1 on the left wing, and four good QSAR models are established: $R = 0.8211 (N = 34)$, $R = 0.8599$ $(N = 33)$, $R = 0.8711 (N = 30)$ and $R = 0.9079$ $(N = 29)$ for the cases of wild type, and mutant strain forms L100I, Y181C and Y188N, respectively. Additionally, the results indicate that log P plays a vital role in the prediction of the activity of DAPYs, and the cyano group on the left wing is important for inhibiting the mutant forms L100I and Y188N.

Keywords: HIV-1 RT; NNRTIs; DAPYs; Hydrophobicity index; 2D-QSAR

1. Introduction

Non-nucleoside reverse transcriptase inhibitors (NNRTIs), one of the two kinds of inhibitors against the HIV-1 reverse transcriptase (HIV-1 RT), have attracted more attention due to their high specificity and low toxicity [1]. However, the rapid emergence of resistant HIV viral strains carrying mutation at residues that surround the NNRTIs' binding pocket limits the usefulness of NNRTIs. Thus, the design and development of new and more potent mutation-resistant inhibitors is still an arduous task for the treatment of the HIV-1 infected patients [2].

Recently, a novel class of NNRTI compounds diarylpyrimidines (DAPYs, Fig. 1) [3–5], which demonstrate the beneficial effect on HIV-infected patients with NNRTI-resistant viruses, have led to the hope for treatment of the AIDS patients. Etravirine (TMC125), an early DAPY compound, is currently in phase IIIB clinical trials in the United States. Based on the anti-HIV data (IC_{50} is 50% protection of MT-4 cells form HIV-1 cytopathicity as determined by the MTT method) reported by Kukla et al. [3], Thakur and his coworkers [6] have constructed two 2D-quantitative structure–activity relationship (2D-QSAR) models with the hydrophobicity index log P, a topological Balaban branching index (J), molar volume (MV), and two indicator parameters I_1 and I_2, and concluded that the hydrophobicity index log P plays no role in the prediction of inhibition activity of DAPYs. However, Hansch et al. [7] proposed that the hydrophobic index log P plays an important role in the prediction of inhibition activity of NNRTIs. As pointed out above, the design of anti-AIDS drugs should be prefaced the development of drug resistant, but the structure–activity relationships for NNRTIs against HIV-1 RT mutant forms are still less studied [8].

In continuation with our early work in designing mutation-resistant inhibitors NNRTIs [9], 2D-QSAR for the wild-type, mutant forms (L100I, Y181C and 188L) of HIV-1 RT has been investigated by multiple linear regression analysis and the crossvalidation of Leave-One-Out (LOO) [10] method in the present work. The purpose of this work is to study the

* Corresponding author. Fax: +86 21 65643811.
E-mail address: rfchen@fudan.edu.cn (F.-E. Chen).

0223-5234/$ - see front matter © 2008 Elsevier Masson SAS. All rights reserved.
doi:10.1016/j.ejmech.2008.03.021

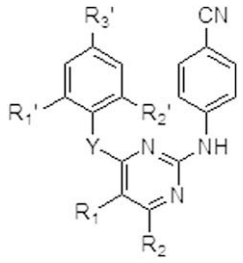

Fig. 1. The parent structure of DAPY derivatives.

following two aspects: (1) whether or not the hydrophobic index log *P* plays a role in the prediction of the activity of DAPYs against HIV-1 RT wild type; (2) to establish 2D-QSAR models for DAPYs against HIV-1 RT mutant forms.

2. Materials and methods

2.1. Experimental data

The experimental data comprising 34 IC_{50} values for wild type, 33 IC_{50} for L100I and Y181C, and 31 IC_{50} for Y188L are collected from the literature [3,11,12] and are converted to log 1/*C* (that is, $-\log IC_{50}$). The structures of compounds used in the study and their experimental activities (log 1/*C*) are given in Tables 1, 2 and 4, respectively.

2.2. Calculation of molecular descriptors

In order to establish the 2D-QSAR model, compounds should be represented as theoretical molecular descriptors [13,14]. In this work, the structures of 34 compounds are built and preoptimized by means of the molecular mechanics force field (MM+) in Hyperchem demo version 7.5 [15], and the hydrophobicity index log *P* is obtained by Hyperchem. Then, the initial structures are optimized by the semi-empirical method PM3 [16,17] in the Gaussian 03 program [18]. Previous theoretically studied results by ONIOM2 (B3LYP/6-31G(d): PM3) calculation [19] indicated that two main interactions exist between DAPY and non-nucleoside binding pocket (NNIBP)—hydrogen bonding interaction with Lys101 and π−π stacking or H···π interaction with Tyr181 and Tyr188. Thus, three physicochemical parameters (the Muliken charge of nitrogen for the right NH group N_C, an indicator index *I* and the modified steric index L_Y) are considered for these interactions. The calculated parameters are listed in Table 1.

Table 1
Physicochemical and indicator parameters of DAPY derivatives

No.	R_1', R_2', R_3'	R_1	R_2	Y	log *P*	*A* log *P*	L_Y	N_C	*I*
1	2,4,6-TriMe	–	–	N	1.64	5.549	3.00	0.27	0
2	2,6-DiMe-4-CN	–	–	O	1.39	4.952	2.81	0.25	0
3	2,6-DiMe-4-CN	–	–	N	1.21	4.942	3.00	0.25	0
4	2,6-DiMe-4-Br	–	–	O	1.72	5.821	2.81	0.27	0
5	2,6-diMe-4-CN	–	–	S	2.06	6.380	2.56	0.27	0
6	2,6-DiMe-4-(HCC)	–	–	O	1.53	6.200	2.81	0.27	0
7	2,4,6-TriMe	–	–	S	2.16	6.117	2.56	0.27	0
8	2,4,6-TriMe	–	–	O	1.82	5.559	2.81	0.27	0
9	2,6-DiBr-4F	–	–	N	0.68	5.793	3.83	0.24	0
10	2,4,6-TriCl	–	–	N	0.51	6.084	3.52	0.24	0
11	2,6-DiMe	–	–	N	1.49	5.063	3.00	0.27	0
12	2,4-DiCl-6-Me	–	–	N	0.89	5.906	3.26	0.24	0
13	2,6-DiMe-4-Cl	–	–	N	1.27	5.727	3.00	0.24	0
14	2,6-DiBr-4-Me	–	–	N	1.44	6.074	3.83	0.24	0
15	2,6-DiMe-4-Br	–	–	N	1.54	5.811	3.00	0.24	0
16	2,6-DiMe-4-CN	Br	–	N	1.39	5.690	3.00	0.25	1
17	2,6-DiMe-4-CN	Br	–	N	1.56	5.700	2.81	0.25	1
18	2,4,6-TriMe	Br	–	O	1.82	6.290	3.00	0.27	0
19	2,4,6-TriMe	HCC	–	N	1.63	6.676	3.00	0.27	0
20	2,4,6-TriMe	Vinyl	–	N	2.1	6.101	3.00	0.27	0
21	2,4,6-TriMe	Ph	–	N	3.14	–	3.00	0.27	0
22	2,6-DiMe-4-CN	CN	–	N	1.06	4.821	3.00	0.27	0
23	2,4,6-TriMe	CN	–	N	1.49	5.428	3.00	0.29	0
24	2,6-DiMe-4-CN	Cl	–	N	1.11	5.606	3.00	0.25	1
25	2,6-DiMe-4-CN	Cl	–	O	1.29	5.616	2.81	0.25	1
26	2,4,6-TriMe	Cl	–	N	1.72	6.224	3.00	0.24	0
27	2,6-DiMe-4-CN	Me	–	N	1.49	5.428	3.00	0.24	0
28	2,4,6-TriMe	Me	–	N	1.92	6.035	3.00	0.26	0
29	2,6-DiMe-4-CN	NO_2	–	O	0.70	4.846	2.81	0.34	0
30	2,6-DiMe-4-CN	NH_2	–	O	−0.21	4.205	2.81	0.24	0
31	2,6-DiMe-4-CN	NHAc	–	O	−0.38	4.072	2.81	0.27	0
32	2,6-DiMe-4-CN	Br	NH_2	O	1.54	5.493	2.81	0.28	1
33	2,6-DiMe-4-CN	Br	CH_2OH	N	0.81	5.524	3.00	0.25	1
34	2,6-DiMe-4-CN	Br	NHOH	N	2.18	5.486	3.00	0.28	1

Table 2
Experimental and predicted values of log 1/*C* for the test set of 34 DAPY derivatives against HIV-1 RT wild type

No.	Obsd	Eq. (3)		Eq. (4)		Eq. (5)	
		Calcd	Δ	Calcd	Δ	Calcd	Δ
1	9.00	8.69	0.31	8.67	0.33	8.76	0.24
2	8.96	8.88	0.08	8.81	0.15	8.97	−0.01
3	9.40	8.98	0.42	8.89	0.51	9.06	0.34
4	8.54	8.58	−0.05	8.60	−0.06	8.58	−0.04
5	8.24	8.41	−0.16	8.43	−0.19	8.20	0.04
6	8.26	8.66	−0.41	8.64	−0.38	8.41	−0.15
7	8.44	8.35	0.09	8.39	0.05	8.34	0.10
8	8.54	8.56	−0.02	8.57	−0.03	8.68	−0.14
9	9.22	9.28	−0.06	9.25	−0.03	9.28	−0.06
10	9.15	9.19	−0.03	9.17	−0.02	9.02	0.13
11	9.15	8.74	0.41	8.71	0.44	8.88	0.27
12	9.00	9.14	−0.14	9.07	−0.07	8.99	0.01
13	8.68	9.03	−0.36	8.95	−0.27	8.94	−0.26
14	9.15	9.18	−0.02	9.15	0.00	9.16	−0.01
15	8.64	8.96	−0.33	8.90	−0.26	8.91	−0.27
16	9.40	8.92	0.47	8.86	0.54	8.88	0.52
17	8.85	8.83	0.03	8.77	0.08	8.79	0.06
18	8.26	8.59	−0.33	8.62	−0.36	8.44	−0.18
19	8.38	8.68	−0.30	8.67	−0.29	8.20	0.18
20	8.60	8.50	0.10	8.54	0.06	8.54	0.06
21	7.62	7.71	−0.09	7.58	0.04	–	–
22	9.30	8.66	0.64	9.27	0.03	9.33	−0.03
23	9.00	8.60	0.40	9.06	−0.06	9.06	−0.06
24	8.92	9.00	−0.08	8.91	0.01	8.91	0.01
25	8.85	8.86	−0.01	8.83	0.02	8.82	0.03
26	8.57	8.84	−0.27	8.85	−0.28	8.72	−0.15
27	9.10	8.92	0.17	8.91	0.19	9.04	0.06
28	8.77	8.62	0.15	8.66	0.11	8.65	0.12
29	7.74	8.35	−0.61	7.80	−0.06	7.83	−0.09
30	9.00	8.87	0.13	8.93	0.07	9.05	−0.05
31	8.72	8.58	0.14	8.70	0.02	8.79	−0.07
32	8.85	8.63	0.23	8.57	0.28	8.62	0.23
33	8.40	8.97	−0.57	8.94	−0.54	8.93	−0.53
34	8.40	8.35	0.05	8.44	−0.04	8.70	−0.30

For the hydrogen bonding interaction, the Muliken charge of N for the right NH group is calculated by the PM3 method. The other main interaction is the π−π stacking with Tyr181, and Sherill's calculated results [20] indicated that the π−π interaction for sandwich configurations of benzene−benzonitrile is the strongest among the benzene dimers, benzene−phenol, benzene−toluene, benzene−fluorobenzene, and benzene−benzonitrile complexes. Thus, an indicator index *I* is used to account for this interaction, where, $I = 1$ for R_3', the CN group and R_1 is halogen. Once two substituents R_1' and R_2' are presented, the conformation favoring the left phenyl ring π−π stacking with Tyr181 is partially due to limiting the rotational freedom by increasing the steric effect between the left phenyl ring and the right phenyl ring. In order to consider this effect, we put forward a modified steric index L_Y as follows:

$$L_Y = k\left(\frac{L_{R_1'} + L_{R_2'}}{2}\right)$$

In this equation, *k* is the calibration constant. Once Y is N, the same atom as that of the right wing, *k* is equal to 1. On the other hand, once Y is another atom, the steric effect between the left phenyl ring and the right phenyl ring will be diminished as the bond length C−Y is shorter or longer than that of C−N. Thus, if Y is the sulfur atom, $k = L_{NH_2}/L_{SH}$, and if Y is the oxygen atom, $k = L_{OH}/L_{NH_2}$. And $L_{R_1'}$ and $L_{R_2'}$ are the STERIMOL lengths for R_1' and R_2'.

3. Results and discussion

3.1. HIV-1 RT wild-type study

Firstly, the multiple linear regression (MLR) method with four physicochemical parameters is applied to model 34 DAPYs against HIV-1 RT wild type on Microsoft Office Excel 2003, and a common 2D-QSAR model Eq. (1) with the correlation coefficient *R* equal to 0.6343 is established as shown in Table 4. In this full model Eq. (1), log *P* is the hydrophobicity index calculated by Hyperchem [12], L_Y is the modified steric parameter for two substituents R_1' and R_2' on the left phenyl ring, N_C is the muliken charge of nitrogen atom on the left wing, and *I* is the indicator index for R_3' and R_1. In contrast with the 2D-QSAR models for two other NNRTIs, TIBO and HEPT [7], the activities of DAPY derivatives are in the inverse ratio with their hydrophobicity index log *P* in this model.

Previous studies [7] have shown that the activity of a drug usually correlates with the hydrophobicity index log *P* in the inverse parabola relationship. Thus, based on the model Eq. (1), the square log *P* is considered, and leads to a better 2D-QSAR model Eq. (2). This parabolic model Eq. (2) gives an optimum value of log *P* equal to 0.68, and the correlation coefficient *R* is higher than 0.72. In order to evaluate the descriptors in the 2D-QSAR model Eq. (2), we rank the descriptors according to their effect on increasing the value of *S* when removed from the model. And we find that the indictor index *I* plays little role in the prediction of the activity of DAPYs against HIV-1 RT wild type. The data presented in Table 3 (correlation matrix) also show that the hydrophibicity index log *P*, L_Y and N_C play a dominant role in the prediction of log 1/*C*. Once the indictor index *I* is removed, a model Eq. (3) with the correlation coefficient *R* equal to 0.7255 is obtained as depicted in Table 5. Table 3 shows the correlation matrix among four molecular descriptors and the correlation of log $1/C_{LAI}$ with each single variable. It reveals that the three optimal descriptors are not seriously intercorrelated, thus justifying the inclusion of all the variables in the relationship.

In order to improve the 2D-QSAR model's prediction, we analyze the outliers that the model Eq. (3) underestimated

Table 3
Correlation matrix for the molecular description and their correlations with the activity of DAPYs against HIV-1 RT wild type

	log $1/C_{LAI}$	log *P*	L_Y	N_C	*I*
log $1/C_{LAI}$	1.0000	0.4192	0.4114	0.4852	0.0877
log *P*	0.4192	1.0000	0.1978	0.1447	0.0064
L_Y	0.4114	0.1978	1.0000	0.4060	0.1360
N_C	0.4852	0.1447	0.4060	1.0000	0.0672
I	0.0877	0.0064	0.1360	0.0672	1.0000

Table 4
Experimental and predicted values of log 1/*C* for the test set of 34 DAPY derivatives against HIV-1 RT mutant forms L100I, Y181C and Y188L

No.	log $1/C_{L100I}$ Eq. (6)			log $1/C_{Y181C}$ Eq. (7)			log $1/C_{Y188L}$ Eq. (9)		
	Obsd	Calcd	Δ	Obsd	Calcd	Δ	Obsd	Calcd	Δ
1	7.74	7.36	0.38	8.12[a]	7.46	0.66	7.32	7.27	0.05
2	7.14	6.98	0.16	7.43	7.49	−0.06	7.72	7.35	0.37
3	7.47	7.16	0.31	8.15[a]	7.60	0.55	8.11[a]	7.32	0.78
4	6.95	7.18	−0.23	7.23	7.33	−0.10	7.10	7.27	−0.17
5	6.76	7.18	−0.43	7.17	7.39	−0.22	7.36	7.60	−0.25
6	6.73	7.05	−0.32	6.77	7.28	−0.52	7.30	7.18	0.12
7	7.27	7.30	−0.03	7.60	7.45	0.16	7.70	7.71	−0.01
8	7.66	7.26	0.40	7.42	7.39	0.03	7.26	7.36	−0.10
9	7.82	8.18	−0.36	7.82	8.18	−0.35	7.19	7.43	−0.24
10	7.85	7.85	0.01	8.00	8.00	0.00	7.00	7.48	−0.47
11	6.67	7.27	−0.60	7.21	7.42	−0.20	6.59	7.19	−0.60
12	7.22	7.44	−0.22	7.74	7.82	−0.08	7.07	7.39	−0.33
13	6.49	7.17	−0.69	7.72	7.70	0.02	6.94	7.43	−0.49
14	8.52	8.28	0.24	8.47	8.23	0.24	–	–	–
15	7.12	7.30	−0.18	7.77	7.77	0.00	7.16	7.53	−0.37
16	8.15	8.43	−0.27	8.02	8.05	−0.03	8.48	8.38	0.10
17	8.18	8.27	−0.09	7.66	7.97	−0.31	8.23	8.46	−0.23
18	7.96	7.50	0.46	7.44	7.48	−0.04	7.57	7.34	0.23
19	7.55	7.36	0.20	7.37	7.43	−0.07	7.38	7.24	0.14
20	7.46	7.78	−0.32	7.44	7.69	−0.25	7.54	7.66	−0.12
21	–	–	–	–	–	–	–	–	–
22	7.04	7.13	−0.09	7.12	7.38	−0.26	8.07[a]	7.12	0.95
23	7.20	7.27	−0.07	7.40	7.21	0.19	7.11	6.99	0.12
24	8.28	8.34	−0.05	7.85	8.02	−0.17	8.33	8.34	−0.01
25	8.12	8.14	−0.02	7.47	7.85	−0.38	8.32	8.29	0.02
26	7.77	7.42	0.35	7.49	7.75	−0.25	7.47	7.56	−0.09
27	7.80	7.27	0.53	7.92	7.68	0.25	7.80	7.43	0.36
28	7.92	7.59	0.33	7.92	7.64	0.28	7.74	7.53	0.22
29	6.92	6.90	0.02	6.52	6.58	−0.06	6.11	6.49	−0.39
30	7.80	7.48	0.32	8.03	7.84	0.19	8.24	8.07	0.18
31	7.51	7.68	−0.17	7.52	7.63	−0.10	8.08	7.98	0.10
32	8.48	8.26	0.22	8.15[a]	7.68	0.48	8.34	8.17	0.17
33	8.70	8.33	0.37	8.22	7.82	0.40	8.22	8.28	−0.05
34	8.92	9.07	−0.15	7.92	7.91	0.01	–	–	–

[a] The compounds are removed in the improved 2D-QSAR models Eq. (7) and Eq. (9).

the activity for the compounds **22** and **23** and overestimated the activity for the outlier **29** because the model Eq. (3) lacks a parameter for R_1. Once the substituent R_1 is Ph or NO_2, the hydrogen bonding between the inhibitor and NNIBP is reduced by the large steric effect [10,19,23], and while the substituent is the cyano group, the π−H interaction will be increased [19,20]. Thus, an indicator for this substituent is considered, $R_1 = CN$, $I' = 1$; $R_1 = Ph$, NO_2, $I' = -1$, otherwise, $I' = 0$. Combined with the above three parameters, a better model Eq. (4) is established with $R = 0.8211$ and $S = 0.26$ as depicted in Table 5. To examine this improvement, the crossvalidation of LOO method was carried out in SYBYL 7.2 [21] for testing the predictive ability of Eq. (4). The results of LOO method with $R_{CV} = 0.7413$ and $S_{CV} = 0.28$ show that Eq. (4) is more stable and better than the model Eq. (3) with $R_{CV} = 0.5272$ and $S_{CV} = 0.36$. Additionally, the linear relationship between the experimental and calculated values for this model is given in Table 2 and Fig. 2A, and shows that the model Eq. (4) can be good to predict the activity of DAPYs. Recently, Thakur et al. [6] constructed a 2D-QSAR model (constructed by MLR with four parameters *J*, MV, I_1 and I_2 against IC_{50}) for the front 31 compounds in Table 1 as the training dataset with the correlation coefficient *R* equal to 0.5749, and six outliers exist in their model. Thus, the 2D-QSAR model Eq. (4) is better than that reported by Thakur et al. [6]. However, we find that the hydrophobicity index log *P* is not equal to those of Thakur. Thus, we calculated the hydrophobicity index *A* log *P* [22] by the software Dragon Evaluation version 5.4 [24]. As shown in Table 2, the 2D-QSAR model Eq. (5) also indicated that the hydrophobicity index log *P* plays a vital role in the prediction of the activity of DAPYs.

3.2. Study of the HIV-1 RT mutant forms

DAPYs are the new generation NNRTIs that retain high activity against mutant strains associated with resistance to NNRTIs [25]. Thus, to construct 2D-QSAR for DAPYs against HIV-1 RT mutant forms play a vital role to develop and design mutation-resistant NNRTIs. In this section, we will establish 2D-QSAR models for the activity of DAPYs against mutant forms L100I, Y181C and Y188L, and reveal

Table 5
2D-QSAR models for DAPYs against HIV-1 RT wild-type, mutant forms L100I, Y181C and Y188L[a]

Type	Model		N	R	S	F	R_{CV}	S_{CV}
LAI	$\log 1/C = 9.72(\pm 1.32) - 0.21(\pm 0.09)\log P + 0.34(\pm 0.25)L_Y - 6.66(\pm 3.10)N_C + 0.10(\pm 0.15)I$	(1)	34	0.6342	0.35	4.88	0.4221	0.39
LAI	$\log 1/C = 9.87(\pm 1.20) - 0.20(\pm 0.07)(\log P)^2 + 0.27(\pm 0.20)\log P + 0.24(\pm 0.23)L_Y - 6.76(\pm 2.80)N_C + 0.02(\pm 0.14)I$	(2)	34	0.7259	0.32	6.24	0.4963	0.37
LAI	$\log 1/C = 9.92(\pm 1.15) - 0.20(\pm 0.07)(\log P)^2 + 0.28(\pm 0.19)\log P + 0.23(\pm 0.22)L_Y - 6.82(\pm 2.73)N_C$	(3)	34	0.7255	0.31	8.05	0.5272	0.36
LAI	$\log 1/C = 9.79(\pm 1.03) - 0.10(\pm 0.04)(\log P)^2 + 0.09(\pm 0.17)\log P + 0.28(\pm 0.19)L_Y - 6.73(\pm 2.53)N_C + 0.49(\pm 0.15)I'$	(4)	34	0.8211	0.26	11.59	0.7413	0.28
LAI	$\log 1/C = 6.37(\pm 1.03) - 0.16(\pm 0.08)(A\log P)^2 + 1.49(\pm 0.91)A\log P + 0.43(\pm 0.16)L_Y - 8.08(\pm 2.11)N_C + 0.42(\pm 0.13)I'$	(5)	33	0.8504	0.22	14.11	–	–
L100I	$\log 1/C = 3.78(\pm 0.75) + 0.47(\pm 0.13)(\log P)^2 - 0.87(\pm 0.27)\log P + 1.25(\pm 0.24)L_Y + 1.20(\pm 0.15)I$	(6)	33	0.8599	0.22	15.10	0.8070	0.36
Y181C	$\log 1/C = 8.35(\pm 1.21) + 0.24(\pm 0.11)(\log P)^2 - 0.45(\pm 0.25)\log P + 0.60(\pm 0.24)L_Y - 9.69(\pm 0.14)N_C + 0.44(\pm 2.80)I$	(7)	33	0.7637	0.30	7.55	0.6524	0.32
Y181C	$\log 1/C = 8.27(\pm 0.90) + 0.30(\pm 0.09)(\log P)^2 - 0.57(\pm 0.18)\log P + 0.73(\pm 0.17)L_Y - 10.25(\pm 0.11)N_C + 0.38(\pm 2.12)I$	(8)	30	0.8711	0.22	15.10	0.7873	0.26
Y188L	$\log 1/C = 10.05(\pm 0.85) + 0.47(\pm 0.14)(\log P)^2 - 0.95(\pm 0.30)\log P - 9.24(\pm 3.22)N_C + 1.07(\pm 0.17)I$	(9)	31	0.8063	0.37	12.08	0.6776	0.42
Y188L	$\log 1/C = 9.92(\pm 0.61) + 0.63(\pm 0.11)(\log P)^2 - 1.23(\pm 0.22)\log P - 8.99(\pm 2.33)N_C + 1.18(\pm 0.13)I$	(10)	29	0.9079	0.26	28.17	0.8766	0.27

[a] Statistical parameters: R = correlation coefficient, S = standard error, F = the Fisher criterion, R_{CV} = predictive correlation coefficient, S_{CV} = predictive standard error.

the role of hydrophobic, steric and molecular properties for DAPYs against mutant strain forms.

The mutant form L100I, combined with three molecular descriptors log P, L_Y, and I to model the activity of 33 DAPYs, leads to a good 2D-QSAR model Eq. (5) with $R = 0.8599$ and $S = 0.22$ as depicted in Table 5. In this 2D-QSAR model Eq. (5), the Muliken charge of N for the right NH group plays no role in the prediction of DAPYs against mutant form L100I because this mutant form decreases the hydrogen bond interaction between the inhibitor and Lys101. Additionally, the coefficients of L_Y and I are positive, so increasing the steric index and the presence of cyano group on the left phenyl ring favor DAPYs against mutant form L100I. Thus, design mutation-resistant inhibitors DAPYs for L100I should be increased the interaction between the inhibitor and Tyr181, Tyr188 and Trp229. And our recently designed DAPYs (2-phenylamino-4-naphanoxypyrimidine) and Gullemont's TMC278 (rilpivirine) analogues which retain the cyano group on the left phenyl ring exhibit high activity against mutant form L100I [4,5,23].

In the 2D-QSAR model Eq. (6) for mutant form Y181C, the indicator index I plays little role in the prediction of DAPYs against HIV-1 RT mutant form Y181C as no $\pi-\pi$ stacking interaction exists in this mutant form, and the coefficient is only 0.44. However, as shown in Table 4, there are only three outliers that the model Eq. (6) is underestimated the activity for the compounds **1**, **3** and **32** because the inhibitors in multiple conformations bind to HIV-1 RT [2]. Once the three outliers are removed [6,7,26], a better 2D-QSAR model Eq. (7) is obtained with the correlation coefficient R equal to 0.8711. The crossvalidation of LOO method is performed to test the prediction of Eq. (7) as above for Eq. (4) from Eq. (3), the results with the predictive correlation coefficient $R_{CV} = 0.7873$ and the predictive standard error $S_{CV} = 0.26$ indicate that this outlier is reasonable and Eq. (7) is stable and has good predicted accuracy.

For the mutant form Y188L, the multiparametric model based on three indexes log P, N_C and I gives excellent results as Eq. (8) in Table 5. In this 2D-QSAR model Eq. (8), the coefficient of the indicator index I is positive. Therefore, the cyano group plays a vital role in DAPYs against the mutant form Y188L as in the mutant form L100I. Table 4 lists the predicted values and the residual errors for the model Eq. (8), and shows that this linear model can be used to correlate the molecular structure with log $1/C$ for the mutant form Y188L except for two underestimated outliers **3** and **22**. To improve the prediction of 2D-QSAR model Eq. (8), two outliers have also been deleted. And a wonderful

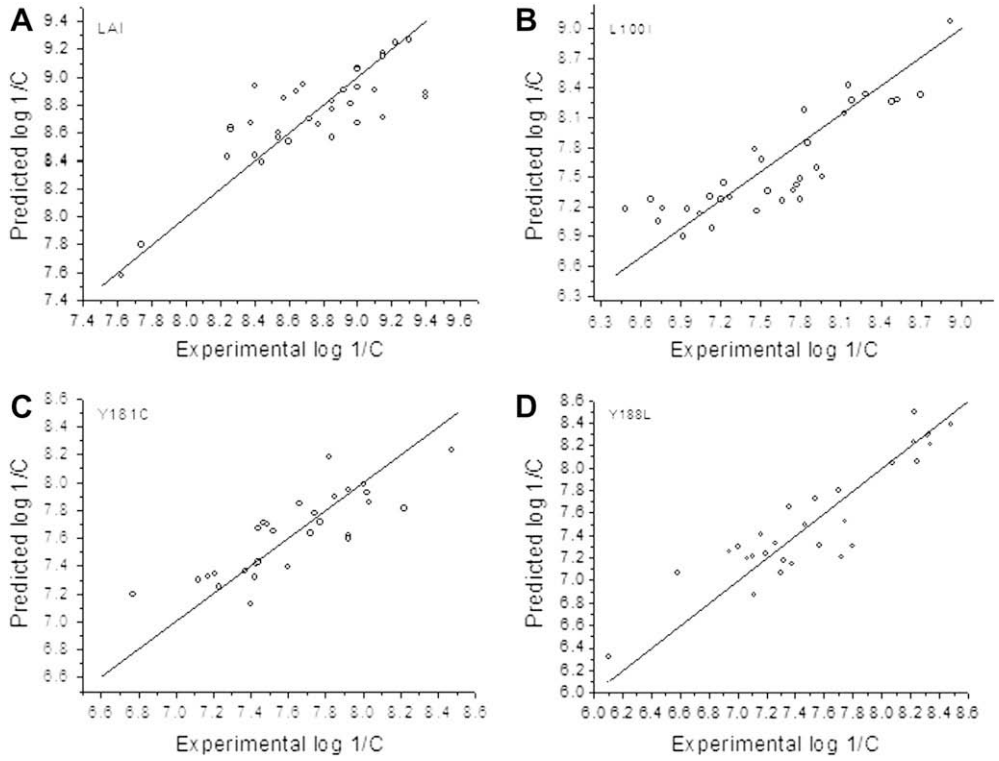

Fig. 2. Plot of predicted vs. experimental activity for DAPYs according to four 2D-QSAR models: (A) based on Eq. (4), (B) based on Eq. (5), (C) based on Eq. (7), and (D) based on Eq. (9).

2D-QSAR model Eq. (9) with the correlation coefficient R equal to 0.9079 is obtained. On comparison of the results of LOO method for the improved 2D-QSAR model Eq. (9) with those for the model Eq. (8), Eq. (9) is more stable and has better predicted accuracy with $R_{CV} = 0.8766$ and $S_{CV} = 0.27$ than Eq. (8). As shown in Fig. 2D, this model exhibits good prediction for DAPYs against the mutant form Y188L.

4. Conclusion

In summary, four 2D-QSAR models for DAPYs against HIV-1 RT wild-type and mutant strains are able to predict those compounds that exhibit favorable activity values as potent, and those compounds in agreement with experimental values as inactive compounds for the activity of DAPYs. In contrast with Thakur's conclusion [6], the four 2D-QSAR models indicate that the hydrophibicity index log P plays a vital role in the prediction of the activity of DAPYs against wild-type, and mutant forms L100I, Y181C and Y188L. And the cyano group on the left phenyl ring plays an important role in DAPYs against L100I and Y188L as shown in the 2D-QSAR models for the mutant forms L100I and Y188L. These studied results will aid the future design of better mutation-resistant inhibitors.

Acknowledgment

This work was supported by the grants from the National Natural Science Foundation of China (no.: 30672536).

References

[1] S.G. Sarafianos, K. Das, S.H. Hughes, E. Arnold, Curr. Opin. Struct. Biol. 4 (2004) 716–730.
[2] K. Das, P.J. Lewi, S.H. Hughes, E. Arnold, Prog. Biophys. Mol. Biol. 88 (2005) 209–231.
[3] D.W. Ludovici, B.L. De Corte, M.J. Kukla, H. Ye, C.Y. Ho, M.A. Lichenstein, R.W. Kavash, K. Andries, M.P. De Bethune, H. Azijn, R. Pauwels, P.J. Lewi, J. Heeres, L.M.H. Koymans, M.R. De Jonge, K.J.A. Van Aken, F.F.D. Daeyaert, K. Das, E. Arnold, P.A.J. Janssen, Bioorg. Med. Chem. Lett. 11 (2001) 2235–2239.
[4] J. Guillemont, E. Pasquier, P. Palandjian, D. Vernier, S. Gaurrand, P.J. Lewi, J. Heeres, M.R. De Jonge, L.M.H. Koymans, F.F.D. Daeyaert, M.H. Vinkers, E. Arnold, K. Das, R. Pauwels, K. Andries, M.-P. de Bethune, E. Bettens, K. Hertogs, P. Wigerinck, P. Timmerman, P.A.J. Janssen, J. Med. Chem. 48 (2005) 2072–2079.
[5] C. Mordant, B. Schmitt, E. Pasquier, C. Demestre, L. Queguiner, C. Masungi, Peeters, A.L. Smeulders, E. Bettern, K. Hertogs, J. Heeres, P. Lewi, J. Guillemont, Eur. J. Med. Chem. 42 (2007) 567–579.
[6] A. Thakur, M. Thakur, A. Bharadwaj, S. Thakur, Eur, J. Med. Chem. 43 (2008) 471–477.
[7] R. Garg, S.P. Gupta, H. Gao, M.B. Babu, A.K. Debnath, C. Hansch, Chem. Rev. 99 (1999) 3525–3601.
[8] P.R. Duchowicz, M. Fernández, J. Caballero, E.A. Castro, F.M. Fernández, Bioorg. Med. Chem. 14 (2006) 5876–5889.
[9] Y.Z. Xiong, F.E. Chen, J. Balzarini, E. De Clercq, C. Pannecouque, Eur. J. Med. Chem., doi:10.1016/j.ejmech.2007.08.001, in press.
[10] M. Stone, J. R. Stat. Soc., Ser. B 36 (1974) 111–147.
[11] K. Das, A.D. Clark Jr., P.J. Lewi, J. Heeres, M.R. De Jonge, L.M.H. Koymans, M. Vinkers, F. Daeyaert, D.W. Ludovici, M.J. Kukla, B. De Corte, R.W. Kavash, C.Y. Ho, H. Ye, M.A. Lichtenstein, K. Andries, R. Pauwels, M.-P. De Bethune, P.L. Boyer, P. Clark, S.H. Hughes, P.A.J. Janssen, E. Arnold, J. Med. Chem. 47 (2004) 2550–2560.
[12] H.A. De Kock, P. Wigerinck, PCT Int. Appl., WO2006094930, 2006.

[13] B. Hemmateenejad, R. Miri, M. Jafarpour, M. Tabarzad, M. Shamsipur, QSAR Comb. Sci. 26 (2007) 1065–1075.
[14] S.D. Chen, H.X. Liu, Z.Y. Wang, QSAR Comb. Sci. 26 (2007) 889–896.
[15] HyperChem. Release 7.5 for Windows. Molecular Modeling System. Evaluation Copy, Hypercube, Inc., Gainesville, FL, USA, 2002.
[16] J.J.P. Stewart, J. Comput. Chem. 10 (1989) 209–220.
[17] J.J.P. Stewart, J. Comput. Chem. 10 (1989) 221–264.
[18] M.J. Frisch, G.W. Trucks, H.B. Schlegel, et al., Gaussian 03 Revision B.03, Gaussian, Inc., Pittsburgh, PA, 2003.
[19] Y.H. Liang, F.E. Chen, Drug Discov. Ther. 1 (2007) 57–60. Available from: <http://www.ddtjournal.com/getabstract.php?id=25>.
[20] M.O. Sinnokrot, C.D. Shrill, J. Am. Chem. Soc. 126 (2004) 7690–7697.
[21] SYBYL 7.2, Tripos International, 1699 South Hanley Rd., St. Louis, Missouri 63144, USA, 2004.
[22] Dragon Evaluation Version 5.4, 2006, <http://www.disat.unimib.it/chm>.
[23] Y.H. Liang, F.E. Chen, J. Balzarini, E. De Clercq, C. Pannecouque, unpublished results.
[24] V.N. Viswanadhan, A.K. Ghose, G.N. Revankar, R.K. Robins, J. Chem. Inf. Comput. Sci. 29 (1989) 163–172.
[25] P. Kirkpatrick, Nat. Rev. Drug Discov. 4 (2005) 105.
[26] K. Roy, J.T. Leonard, Bioorg. Med. Chem. 12 (2004) 745–754.