




Cite this: *RSC Adv.*, 2018, 8, 40529

In silico studies of diarylpyridine derivatives as novel HIV-1 NNRTIs using docking-based 3D-QSAR, molecular dynamics, and pharmacophore modeling approaches

Youlan Wan,† Yafeng Tian,† Wenjie Wang, Shuangxi Gu, Xiulian Ju and Genyan Liu *

A series of novel diarylpyridine derivatives has recently been identified as HIV-1 non-nucleoside reverse transcriptase inhibitors (NNRTIs), and most of them exhibited potent activities against HIV-1 strains, with EC_{50} values in the low nanomolar range. However, the three-dimensional quantitative structure–activity relationships (3D-QSARs) and pharmacophore characteristics of these compounds remain to be studied. In the present study, forty-two diarylpyridine derivatives were firstly docked into HIV-1 reverse transcriptase, and molecular dynamics (10 ns) simulations were further performed to validate the reliability of the docking results, which indicated that residues Lys101, Tyr181, Tyr188, Phe227, and Trp229 might play important roles in binding with these diarylpyridines. The “U”-shaped docking conformations of all compounds were then used to construct 3D-QSAR and pharmacophore models. The satisfactory statistical parameters of CoMFA ($q_{\text{loo}}^2 = 0.665$, $r_{\text{ncv}}^2 = 0.989$, $r_{\text{pred}}^2 = 0.962$, etc.) and CoMSIA ($q_{\text{loo}}^2 = 0.727$, $r_{\text{ncv}}^2 = 0.988$, $r_{\text{pred}}^2 = 0.912$, etc.) models demonstrated that both constructed models had excellent predictability, and their contour maps gave insights into the structural requirements of the diarylpyridines for the anti-HIV-1 activity. A docking-conformation-based pharmacophore model, containing three hydrophobic centers, three hydrogen-bond acceptors, and three hydrogen-bond donors, was also established. The observations in this study might provide important information for the rational design and development of novel HIV-1 NNRTIs.

Received 1st August 2018
Accepted 26th November 2018

DOI: 10.1039/c8ra06475j

rsc.li/rsc-advances

Introduction

Human immunodeficiency virus (HIV) reverse transcriptase (RT) is an asymmetric heterodimer which plays a pivotal role in the viral replication stage for catalyzing the replication of single-stranded viral RNA to produce double-stranded DNA before the viral genome is integrated into the DNA of the host.^{1–4} This characteristic makes the HIV RT an attractive target for the development of anti-HIV drugs.⁵ To date, many HIV RT inhibitors have been developed, which could be mainly classified into nucleoside RT inhibitors (NRTIs) and non-nucleoside RT inhibitors (NNRTIs).^{6,7} NRTIs are analogues of the normal dNTP substrates.⁸ They can compete with endogenous nucleotides at the polymerase active site which could availablely terminate the viral DNA synthesis.⁹ NNRTIs bind to an allosteric site that is an approximately 10 Å distance from the polymerase active site.¹⁰ In fact, the structure studies of HIV RT indicate that the NNRTI-

binding pocket (NNIBP) does not exist in the absence of NNRTIs.^{11–14}

NNRTIs is an important component of highly active anti-retroviral therapy, which is a standard treatment for acquired immune deficiency syndrome (AIDS) patients at present.¹⁵ Chemical structures of NNRTIs are diverse.^{16,17} Different from the first-generation NNRTIs that have almost rigid butterfly-like conformation, the second-generation NNRTIs are more flexible and always adopt “U” (or “horseshoe”) conformation to interact with the HIV type 1 (HIV-1) RT (Fig. 1).^{14,18} The obvious attribution of the second-generation NNRTIs is that they can bind to the RT through torsional flexibility (“wiggling”) and local repositioning (“jiggling”).¹⁹ Thus, the characteristic of conformational flexibility makes the second-generation NNRTIs easily adapt to the steric interference caused by different mutations at the allosteric binding site.²⁰ Diarylpyrimidines (DAPYs) is one class of the efficient and promising second-generation NNRTIs with strong resilience to a number of HIV-1 mutants. Etravirine and rilpivirine, as the second-generation NNRTIs of DAPY derivatives, have been approved by U.S. Food and Drug Administration in 2008 and 2011, respectively.²¹ They have been proved to have highly potent against both WT HIV-1 and many clinically prevalent mutants with low nanomolar IC_{50} values.²²

Key Laboratory for Green Chemical Process of Ministry of Education, School of Chemical Engineering and Pharmacy, Wuhan Institute of Technology, Wuhan 430205, P. R. China. E-mail: liugenyan@yahoo.com; Tel: +86 27 87194980

† Youlan Wan and Yafeng Tian contributed equally to this work.



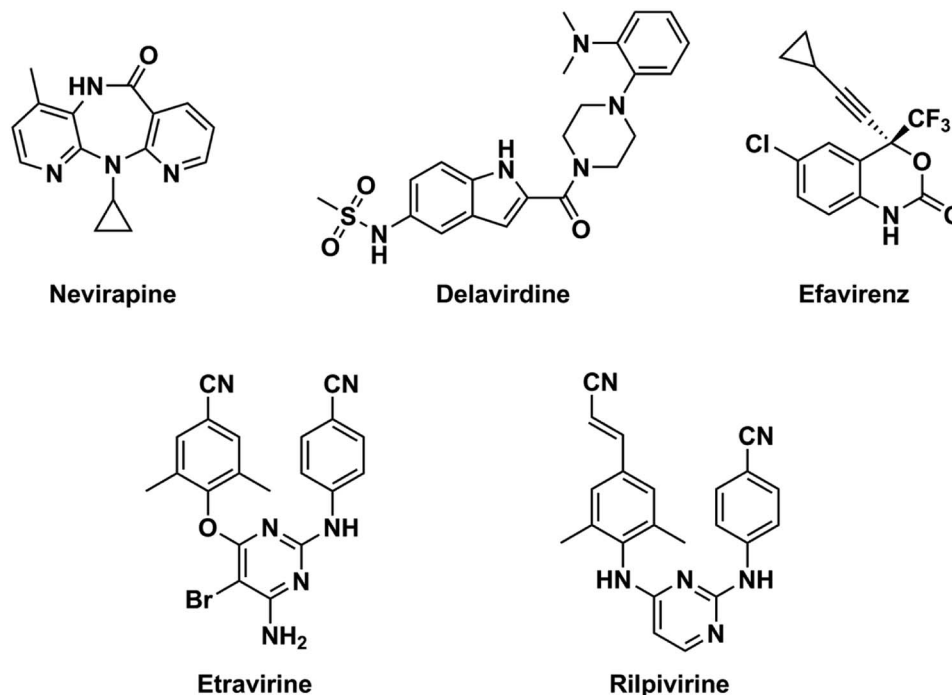


Fig. 1 Chemical structures of the first-generation (nevirapine, delavirdine, and efavirenz) and the second-generation (etravirine and rilpivirine) NNRTIs.

Although it has been reported that second-generation NNRTIs have remarkable potencies against HIV-1 RT mutants, the drug resistances and severe side effects of NNRTIs were appeared to HIV-infected patients for their long-term usage and remained the major obstacles for prolonged antiretroviral therapy.²³ Thus, seeking for the novel and safe NNRTIs with broad spectrum anti-HIV-1 activity is necessary.²⁴ Computer-aided drug design methods, such as molecular docking and pharmacophore models, are often used as effective tools for anticipating the discovery of novel hits by reducing time and costs.

Recently, some novel diarylpyridine derivatives as NNRTIs were designed by replacing the pyrimidine ring of DAPYs with a pyridine ring (Table 1).^{25,26} Most of them exhibited good antiviral activity against wild-type (WT) HIV-1 strains ($EC_{50} = 0.035\text{--}4.41\ \mu\text{M}$). To understand the three-dimensional quantitative structure–activity relationships (3D-QSARs) and pharmacophore characteristics of these novel diarylpyridines, in the present study, the docking-based 3D-QSAR and pharmacophore modeling studies were performed on the novel series of diarylpyridines. Molecular docking was first performed to investigate the interactions between these diarylpyridines and the HIV-1 RT and to obtain the possible drug-like conformations of them. To verify the rationality of the docking results and analyze the dynamics behavior of the diarylpyridines, molecular dynamics (MD) simulations were carried out on HIV-1-RT-diarylpyridine complexes. The docking conformations of the diarylpyridines were then selected to construct 3D-QSAR models, of which the contour maps could figuratively elucidate the relationships between diarylpyridine structures and their inhibitory activities. Moreover, a pharmacophore model based on the docking poses was constructed and also provided an insight into the key

structural features of the diarylpyridines for the activity. This combined ligand- and acceptor-based modeling study might be helpful for designing and developing novel diarylpyridine candidates as HIV-1 NNRTIs.

Materials and methods

Molecular construction and structure optimization

Forty-two novel diarylpyridine derivatives were selected from the published literatures^{25,26} for the molecular simulation. Their structures and anti-HIV-1 activities ($pEC_{50} = -\log EC_{50}$) were listed in Table 1. These molecules were then constructed by SYBYL-X 2.1 running on a windows 7 workstation. All the molecules were minimized with Gasteiger–Hückel charges by the default Powell method which used conjugate-gradient minimization.²⁷ The termination and max iterations were set to $0.005\ \text{kcal}\ (\text{mol}^{-1}\ \text{\AA}^{-1})$ and 1000, respectively. Other parameters were set as default values.

Molecular docking

The crystal structure of the HIV-1 RT for molecular docking was available in the Protein Data Bank (PDB ID: 3MEC). The PDB database includes 72 crystal structures of the WT HIV-1 RT.²⁸ 3MEC was chosen as a docking receptor because of its high resolution ($2.3\ \text{\AA}$) and its co-crystallized ligand etravirine (TMC125), which adopt a “U”-shaped conformation²⁹ at the allosteric site.

All co-crystallized water molecules and sulfate ions were removed before docking. The binding pocket was generated by the automatic mode using the extracted ligand TMC125.³⁰ In order to verify the reliability of molecular docking, the cognate

Table 1 Chemical structures, experimental and predicted pEC₅₀ values of diarylpyridine derivatives acting as HIV-1 NNRTIs

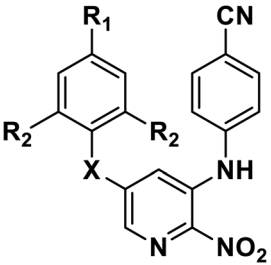
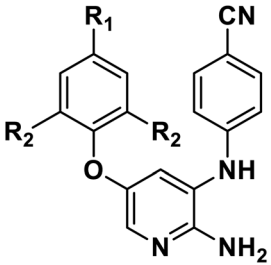
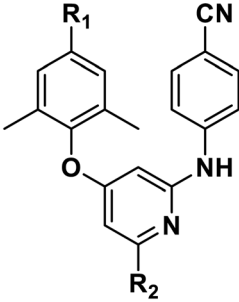
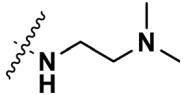
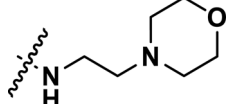
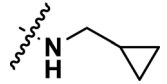
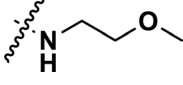
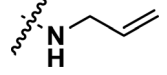
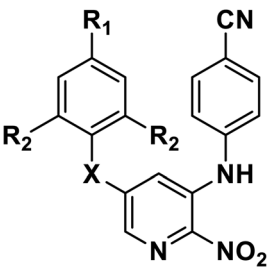
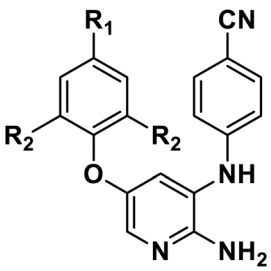
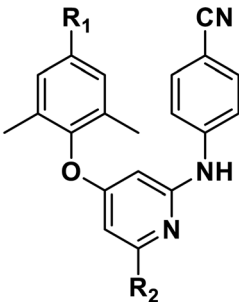
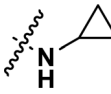
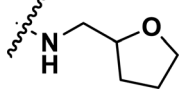
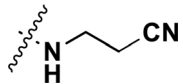
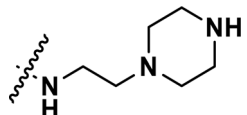
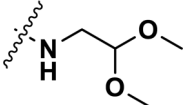
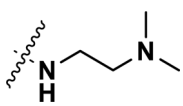
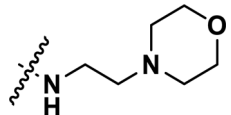
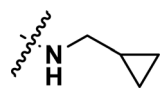
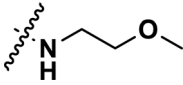
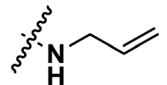
No.	R ₁	R ₂	EC ₅₀ (μM)	Actual pEC ₅₀	CoMFA		CoMSIA	
					Predicted pEC ₅₀	Residual	Predicted pEC ₅₀	Residual
<div style="display: flex; justify-content: space-around; align-items: center;"> <div style="text-align: center;">  <p>compounds 1-9 X = O compounds 10-11 X = NH</p> </div> <div style="text-align: center;">  <p>compounds 12-20</p> </div> <div style="text-align: center;">  <p>compounds 21-42</p> </div> </div>								
1	CN	CH ₃	4.12	5.385	5.374	0.011	5.418	-0.033
2 ^a	CH ₃	CH ₃	0.29	6.538	6.652	-0.114	6.607	-0.069
3 ^b	CH ₃	Br	0.05	7.301	7.174	0.127	7.098	0.203
4	H	CH ₃	2.03	5.693	5.680	0.013	5.707	-0.014
5	Cl	CH ₃	1.41	5.851	5.788	0.063	5.717	0.134
6	Br	CH ₃	4.41	5.356	5.377	-0.021	5.385	-0.029
7	Cl	Cl	0.31	6.509	6.528	-0.019	6.490	0.019
8	Br	Br	0.79	6.102	6.321	-0.219	6.334	-0.232
9	F	F	1.01	5.996	5.858	0.138	6.020	-0.024
10	H	CH ₃	1.06	5.975	6.061	-0.086	6.000	-0.025
11 ^{a,b}	CH ₃	CH ₃	0.04	7.398	7.409	-0.011	6.831	0.567
12 ^b	CN	CH ₃	0.07	7.155	7.088	0.067	7.176	-0.021
13 ^a	CH ₃	Cl	0.14	6.854	6.978	-0.124	6.950	-0.096
14 ^b	CH ₃	Br	0.07	7.155	7.249	-0.094	7.191	-0.036
15	H	CH ₃	0.60	6.222	6.231	-0.009	6.218	0.004
16	H	OCH ₃	35.73	4.447	4.444	0.003	4.479	-0.032
17	Cl	CH ₃	0.37	6.432	6.310	0.122	6.336	0.096
18	Br	CH ₃	0.84	6.076	6.206	-0.130	6.153	-0.077
19 ^a	Cl	Cl	0.26	6.585	6.522	0.063	6.362	0.223
20	Br	Br	0.50	6.301	6.245	0.056	6.228	0.073
21 ^b	CH ₃	Cl	0.37	6.432	6.410	0.022	6.419	0.013
22 ^b	CN	Cl	0.84	6.076	6.125	-0.049	6.108	-0.032
23	CH ₃		0.043	7.367	7.340	0.027	7.471	-0.104
24	CH ₃		0.072	7.143	7.149	-0.006	7.070	0.073
25	CH ₃		1.99	5.701	5.648	0.053	5.689	0.012
26 ^a	CH ₃		0.87	6.060	5.962	0.098	5.855	0.205
27	CH ₃		0.70	6.155	6.117	0.038	6.229	-0.074

Table 1 (Contd.)

No.	R ₁	R ₂	EC ₅₀ (μM)	Actual pEC ₅₀	CoMFA		CoMSIA	
					Predicted pEC ₅₀	Residual	Predicted pEC ₅₀	Residual
 <p>compounds 1-9 X = O compounds 10-11 X = NH</p>			 <p>compounds 12-20</p>			 <p>compounds 21-42</p>		
28 ^b	CH ₃		0.035	7.456	7.489	-0.033	7.486	-0.030
29	CH ₃		1.79	5.747	5.793	-0.046	5.702	0.045
30 ^a	CH ₃		0.24	6.620	6.780	-0.160	6.572	0.048
31	CH ₃		0.71	6.149	6.125	0.024	6.136	0.013
32	CH ₃		1.63	5.788	5.793	-0.005	5.789	-0.001
33 ^b	CN		0.041	7.387	7.321	0.066	7.390	-0.003
34 ^b	CN		0.082	7.086	7.121	-0.035	7.138	-0.052
35	CN		1.52	5.818	5.870	-0.052	5.800	0.018
36 ^a	CN		0.25	6.602	6.665	-0.063	6.658	-0.056
37	CN		0.40	6.398	6.386	0.012	6.262	0.136

The particle-mesh Ewald summation technique was used to compute long-ranged electrostatic interactions.³² The coulomb and van der Waal's cut-offs were set to 1.0 and 1.4 nm, respectively. During the MD simulation, the time step was defined as 2 fs. The coordinate trajectories were written at intervals of 10 ps. Potential energy (PE), hydrogen bond number, root-mean-square deviation (RMSD), root-mean-square fluctuation (RMSF), radius of gyration (R_g), and solvent accessible surface area (SASA) were calculated by GROMACS inbuilt commands, *g_energy*, *g_hbond*, *g_rmsd*, *g_rmsf*, *g_gyration*, and *g_sasa*, respectively. The free binding energies were calculated using a GROMACS tool for high-throughput MM-PBSA calculations (*g_mmpbsa*).

3D-QSAR

QSAR is a method to establish computational or mathematical models, which attempt to find a statistically significant correlation between structure and function by a chemometric technique.³³ Nowadays, multiple methodologies are maturing for 3D-QSAR studies, such as molecular shape analysis (MSA), hypothetical active site lattice (HASL), comparative molecular field analysis (CoMFA), comparative molecular similarity indices analysis (CoMSIA), and genetically evolved receptor models (GERM). In this study, CoMFA and CoMSIA were performed on 42 diarylpyridine derivatives. The statistical technique for building the models was partial least square (PLS) analysis.

Various statistical parameters were taken into consideration to validate the quality of the constructed CoMFA and CoMSIA models. Cross-validated correlation coefficient with leave-one-out method (q_{loo}^2), non-cross-validated correlation coefficient (r_{ncv}^2), F-statistic values (F), and standard error of estimate (SEE) were calculated to measure the internal prediction of the models. However, these statistical parameters could only judge how well a model may recur the response for the training set molecules³⁴ and could not prove the prediction ability of a model for external molecules. Thus, other metrics, for example, r_{test}^2 , r_0^2 , $r_0'^2$, r_m^2 , $r_m'^2$, k , k' , r_{pred}^2 , root mean square error (RMSE), and mean absolute error (MAE), were considered for external validation of the models.³⁵ r_{test}^2 is the regression line coefficient of correlation for the test set. r_0^2 (predicted vs. observed activities) and $r_0'^2$ (observed vs. predicted activities) are the regression lines coefficients of correlation with the intercept set to zero. k (predicted vs. observed activities) and k' (observed vs. predicted activities) are the slopes of regression lines. As for a good model, it should satisfy the following conditions: r_0^2 or $r_0'^2$ is close to r^2 ; $[(r^2 - r_0^2)/r^2] < 0.1$ or $[(r^2 - r_0'^2)/r^2] < 0.1$; $0.85 \leq k \leq 1.15$ or $0.85 \leq k' \leq 1.15$.³⁶ A novel validation parameter r_m^2 is calculated as the following equation:

$$r_m^2 = r^2 \times \left(1 - \sqrt{r^2 - r_0^2}\right) \quad (1)$$

Δr_m^2 and $\overline{r_m^2}$ are also important to prove the external prediction of a model. Δr_m^2 is the difference between r_m^2 (predicted vs. observed activities) and $r_m'^2$ (observed vs. predicted activities). $\overline{r_m^2}$ is the average value of r_m^2 and $r_m'^2$. The values of $\overline{r_m^2}$ for the training set ($\overline{r_{m(\text{training})}^2}$) and the test set ($\overline{r_{m(\text{test})}^2}$) should be more than 0.5. In the case that the previous condition is met, the $\Delta r_{m(\text{training})}^2$ and $\Delta r_{m(\text{test})}^2$ should be less than 0.2.³⁷ The r_{pred}^2 , RMSE, and MAE were calculated by the following equations,^{37,38} respectively.

$$r_{\text{pred}}^2 = 1 - \frac{\sum (Y_{\text{pred}(\text{test})} - Y_{\text{test}})^2}{\sum (Y_{\text{test}} - \overline{Y_{\text{training}}})^2} \quad (2)$$

$$\text{RMSE} = \sqrt{\frac{1}{n} \sum (Y_{\text{obs}} - Y_{\text{pred}})^2} \quad (3)$$

$$\text{MAE} = \frac{1}{n} \times \sum |Y_{\text{obs}} - Y_{\text{pred}}| \quad (4)$$

The r_{pred}^2 should be more than 0.5 and the MAE value should satisfy the following requirement: $\text{MAE} \leq 0.1 \times \text{training set range}$.

Furthermore, sensitivity, specificity, accuracy, positive prediction value (PPV), negative prediction value (NPV) and Matthew's correlation coefficient (MCC) were calculated to assess the quality of 3D-QSAR models. Based on a comparison between the predicted and observed values, the compounds can be classified into four categories: true positives (TP), false negatives (FN), false positives (FP), true negatives (TN).³⁹ The above statistical parameters were defined respectively in the following equations:

$$\text{Sensitivity} = \frac{n(\text{TP})}{n(\text{TP}) + n(\text{FN})} \quad (5)$$

$$\text{Specificity} = \frac{n(\text{TP})}{n(\text{TP}) + n(\text{FN})} \quad (6)$$

$$\text{Accuracy} = \frac{n(\text{TP}) + n(\text{TN})}{n(\text{TP}) + n(\text{FP}) + n(\text{TN}) + n(\text{FN})} \quad (7)$$

$$\text{PPV} = \frac{n(\text{TP})}{n(\text{TP}) + n(\text{FP})} \quad (8)$$

$$\text{NPV} = \frac{n(\text{TN})}{n(\text{TN}) + n(\text{FN})} \quad (9)$$

$$\text{MCC} = \frac{n(\text{TP}) \times n(\text{TN}) - n(\text{FP}) \times n(\text{FN})}{\sqrt{(n(\text{TF}) + n(\text{FP})) \times (n(\text{TP}) + n(\text{FN})) \times (n(\text{TN}) + n(\text{FP})) \times (n(\text{TN}) + n(\text{FN}))}} \quad (10)$$

Pharmacophore model

Genetic Algorithm with Linear Assignment of Hypermolecular Alignment of Datasets (GALAHAD) is a novel approach for pharmacophore screening module. The GALAHAD alignment is from a set of ligand molecules that bind at a common target site by Tripos' proprietary technology to generate pharmacophore hypotheses.⁴⁰ GALAHAD runs in two main stages, including that the ligands are flexibly aligned to each other in internal coordinate space, and then the produced conformations are aligned in Cartesian space. However, these molecules are considered as rigid bodies in the second step. In this study, ten compounds (3, 11, 12, 14, 21, 22, 28, 33, 34, and 40) with high activities and diverse structures were selected to build pharmacophore models. A virtual suggestion according to the experimental activity data is automatically performed to set the parameters of population size, max generations, and mols required to hit for GALAHAD. All conformations of selected molecules were frozen in the setting of advanced parameters.

Results and discussion

Molecular docking

In this study, we firstly applied molecular dockings to investigate the binding mode of the diarylpyridine derivatives in the HIV-1 RT (PDB ID: 3MEC), and the docked conformations of these diarylpyridines were then used to construct the subsequent 3D-QSAR and pharmacophore models. The cognate ligand TMC125 was extracted and then re-docked into the binding site of 3MEC to verify whether the used docking method and parameters were appropriate and reliable. The re-docked conformation was compared with the original conformation of the co-crystal TMC125 (Fig. 2), and the difference between two conformations was very slight with a RMSD of 0.24 Å (<2.0 Å)⁴¹ for all atoms of TMC125. The re-docked results showed that TMC125 formed hydrogen bonds with Lys101 (Lys101-C=O...H-, 1.78 Å; Lys101-N-H...N-, 2.57 Å) and Glu138 (Glu138-C-O...H-, 2.39 Å). The bromine substituent of the pyrimidine ring pointed to an opening cavity between Lys101 and Glu138. The left phenyl substituent located at a hydrophobic pocket that formed by Tyr181, Tyr188, Phe227, and Trp229, and interacted with these aromatic residues by π - π stacking. The twist tendency of the re-docked TMC125 and its interaction with the enzyme were similar with the results of a previous study reported by Lansdon *et al.*²⁹ suggesting that the method of docking was reliable and feasible.

All diarylpyridines were then docked into the NNIBP of the HIV-1 RT using the same docking pattern. All compounds adopted a similar "U" -shape with the original ligand TMC125 in the NNIBP. The similar hydrogen bonds, π - π stacking and hydrophobic interactions were observed in the docking results. The left benzene rings of all compounds interacted with Tyr181, Tyr188, Phe227, and Trp229 by π - π stacking and hydrophobic interactions. Hydrogen bonds were found between Lys101 and the imino group of the right linker, the amino/nitro substituent neighbor to the right linker, or the nitrogen atom of the pyridine ring. The binding patterns of two most active compounds

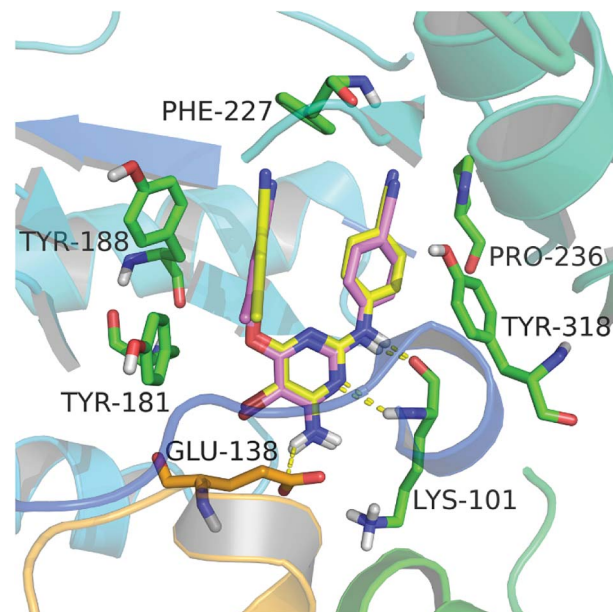


Fig. 2 The re-docked results of etravirine in the binding pocket of the HIV-1 RT (PDB ID: 3MEC). The re-docked etravirine (pink) almost completely superimposes with the co-crystallized etravirine (yellow). The residues from chains A and B are displayed in green and orange, respectively. Hydrogen bonds are shown as yellow dash lines.

11 and 28 were shown in Fig. 3a and b, respectively. As for compound 11, the nitro group out of the pyridine ring and the imino group of the right linker hydrogen-bonded with Lys101 (Lys101-N-H...O = , 2.02 Å; Lys101-C=O...H-, 1.89 Å). The nitrogen atoms of the pyridine rings located at the entrance of the opening between Lys101 and Glu138. As shown in Fig. 3b, compound 28 formed two hydrogen bonds with Lys101 (Lys101-N-H...N-, 2.27 Å; Lys101-C=O...H-, 1.65 Å). The imino group of the R₂ substituents interacted with Glu138 *via* a hydrogen bond (Glu138 = O...H-, 2.01 Å). In this study, Glu138 was found to be the only amino acid residue of the chain B, which could interact with these diarylpyridine derivatives. This result is in agreement with a previous report by Das *et al.*⁴² All found interactions are also consistent with the results from the reported literatures.^{21,22} It seems that Lys101 is a key residue which could form at least one hydrogen bond with diarylpyridine derivatives.

MD simulations

To further validate the docking results and understand the dynamic interactions between ligands and the HIV-1 RT, 10 ns MD simulations were carried on two complexes, 3MEC-11 and 3MEC-28. Fig. 4a showed the average PE values of 3MEC-11 and 3MEC-28 during the 10 ns MD simulations, ranging from -4.812×10^6 to -4.793×10^6 kJ mol⁻¹, which confirmed the stability of two complexes. The total numbers of hydrogen bond interactions of two complexes during the 10 ns MD simulations were shown in Fig. 4b. As shown in Fig. 4c and d, the RMSD values of both protein backbones had a sharply rise during the first 0.3 ns and then remained steady (Fig. 4c); the RMSD values

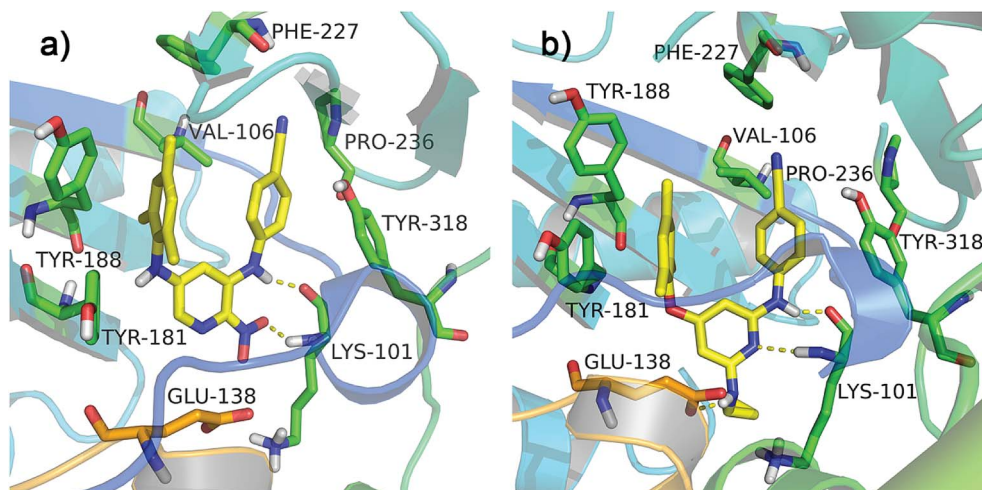


Fig. 3 The docking results of compounds **11** (a) and **28** (b) in the binding pocket of the HIV-1 RT (PDB ID: 3MEC). The residues from chains A and B are displayed in green and orange, respectively. Hydrogen bonds are shown as yellow dash lines.

of the total atoms of ligands **11** and **28** maintained at 0.10 and 0.12 nm after 4.0 and 7.5 ns, respectively. These data demonstrated that two systems were stable during the 10 ns MD simulations.

The RMSF values of all residues of chains A and B were lower than 0.40 nm (Fig. 4e and f). Low RMSF values of some key residues in the NNIBP illustrated these residues had good stability in a dynamic environment. The R_g and SASA values of two complexes during the 10 ns MD simulations were shown in Fig. 4g and h, respectively. It could be observed that the R_g and SASA values were equilibrated after a short-time increase at the beginning of the simulations. The binding energies of complexes 3MEC-**11** and 3MEC-**28** were -37.77 and -36.58 kcal mol $^{-1}$, respectively. Moreover, the binding modes of complexes 3MEC-**11** and 3MEC-**28** at 5 ns and 10 ns were shown in Fig. 5. Compounds **11** and **28** maintained a “U”-shape and had the same interactions with the residues of the NNIBP during the MD processes. It should be noted that the hydrogen bonds between compound **28** and Lys101 (chain A) were observed, but the hydrogen bond with Glu138 (chain B) was disappeared. This might be because the distance between the compound **28** and Glu138 was too far at the selected moment for the MD simulation, and this observation also indicated that the Lys101 was a key residue on the other hand. In general, all MD results indicated that complexes 3MEC-**11** and 3MEC-**28** were structurally stable and the docking results were reliable.

Statistical results of CoMFA and CoMSIA

The docking conformations of 42 diarylpyridines were then used for building 3D-QSAR models (Fig. 6). The compounds were divided into a training set (35 compounds) and a test set (7 compounds) randomly. The PLS statistical results were listed in Table 2. All the permutation and combination of steric (S), electrostatic (E), hydrophobic (H), hydrogen-bond donor (D), and hydrogen-bond acceptor (A) fields of CoMSIA were calculated by PLS, and CoMSIA models with low values of q_{100}^2 (<0.5) and r_{ncv}^2 (<0.9) were not listed in Table 2. The S, E, H, D, and A

contributions of CoMSIA (S + E + H + D + A) were 7.5%, 25.4%, 21.1%, 19.5%, and 26.6%, respectively. It could be observed that E and A fields played more significant roles. The CoMFA model (S + E) ($q_{100}^2 = 0.665$, SEE = 0.084, $r_{ncv}^2 = 0.989$) and the CoMSIA model (S + E + H + D + A) ($q_{100}^2 = 0.727$, SEE = 0.093, $r_{ncv}^2 = 0.988$) were used for further analysis. As shown in Table 3, the MAE_{training} and MAE_{test} of CoMFA and CoMSIA were both less than $0.1 \times$ training set range. The values of the slope k and k' of test set were close to 1. Regarding the values of r_m^2 metrics, the $\Delta r_{m(\text{training})}^2$ and $\Delta r_{m(\text{test})}^2$ of CoMFA and CoMSIA were lower than 0.2; the $r_{m(\text{training})}^2$ and $r_{m(\text{test})}^2$ of CoMFA and CoMSIA were higher than 0.5. The correlation between the actual and predicted activities of 35 compounds in the training set and 7 compounds in the test set were shown in Fig. 7. The sensitivity, specificity, accuracy, PPV, NPV, and MCC of CoMFA and CoMSIA were listed in Table 4. The sensitivity values of both models were 96.7%, indicating that 96.7% active compounds were correctly predicted as actives. The specificity of CoMFA and CoMSIA were 91.7% and 83.3%, which manifested the percentages of correctly predicted inactive compounds out of total inactives. The MCC is a correlation coefficient between the observed and predicted binary classifications, and is ranging from -1 to 1 , in which the value of 1 signifies a perfect prediction of model. The MCC values of both CoMFA (0.8833) and CoMSIA (0.8220) models were close to 1 . All these statistical results of CoMFA and CoMSIA were similar and well within the acceptable limit, which manifested that the developed CoMFA and CoMSIA models with good predictive ability were reliable and reasonable.

CoMFA and CoMSIA contour maps

The CoMFA and CoMSIA contour maps were further analyzed to investigate the QSARs of the diarylpyridine derivatives. Compound **28** as a reference molecule was in combination with the contour maps. Fig. 8a and b showed the steric contour maps of CoMFA and CoMSIA models, respectively. The green contours represent that bulky groups are favored at this

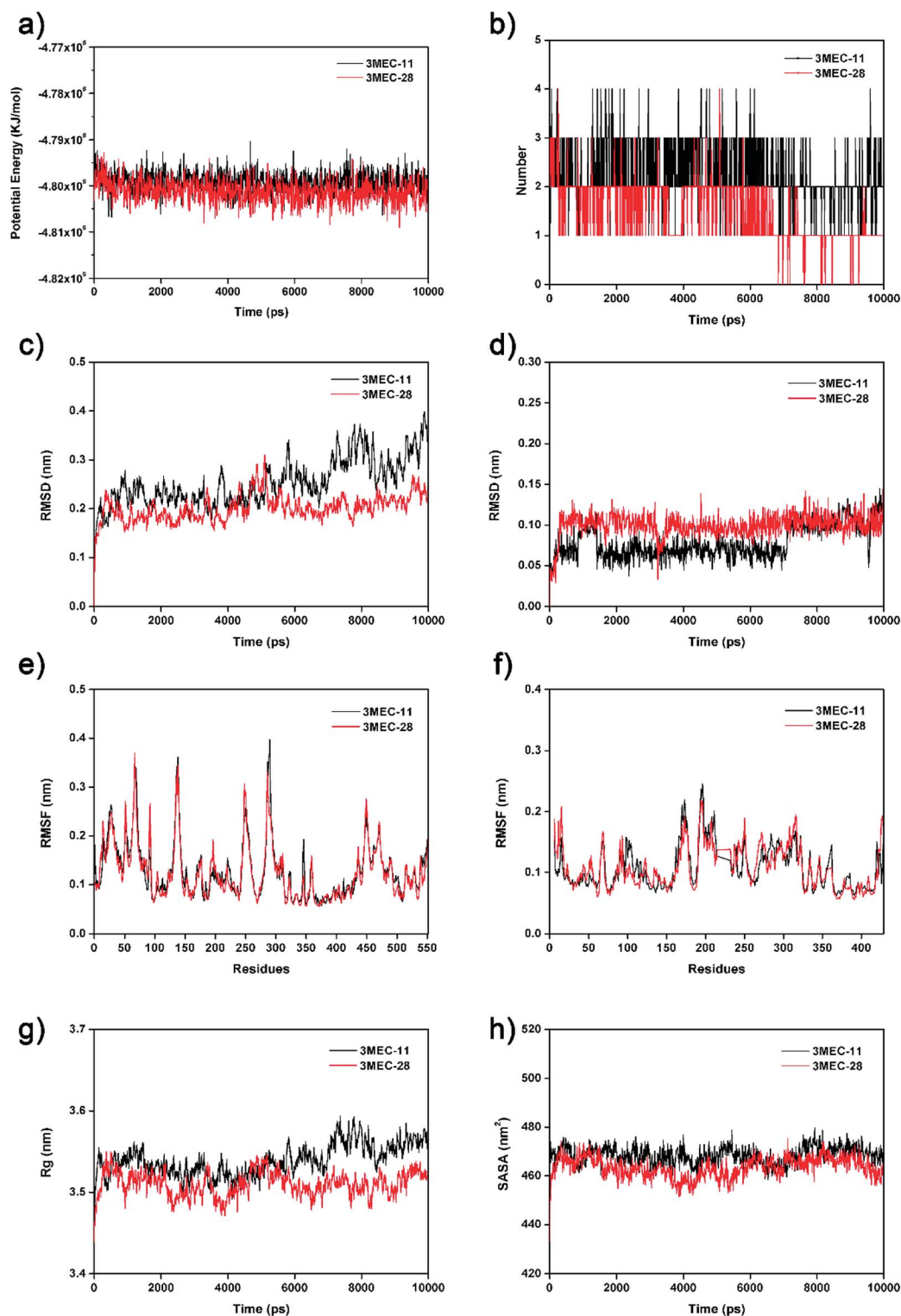


Fig. 4 The 10 ns MD results of two protein–ligand complexes (3MEC-11 and 3MEC-28). (a) Potential energies. (b) Total number of hydrogen bonds between the ligands and the HIV-1 RT. (c) RMSD values of backbone atoms. (d) RMSD values of the ligands. (e) RMSF values of chain A. (f) RMSF values of chain B. (g) R_g of backbone atoms. (h) SASA of the ligands.

position and the yellow contours mean that bulky groups are disfavored here. The green contour near the *para*-position of the left benzene ring indicated that the compound with bulky substituent at this position might be beneficial for the activity.

This was consistent with the experimental results that the activities of compounds with a methyl group at the *para*-position were higher than those of compounds with a hydrogen or a bromine at the same position, for example, **11** (*para*-CH₃) > **10**

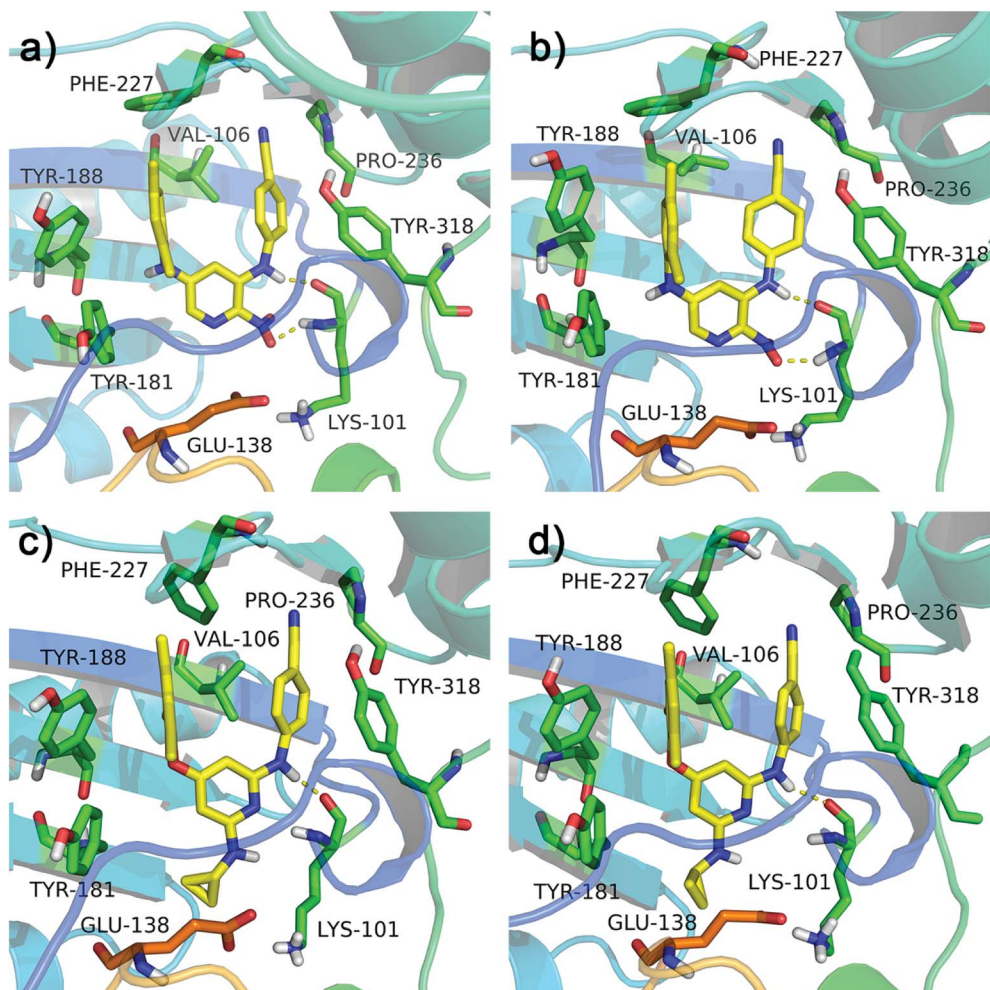


Fig. 5 Binding modes of the protein–ligand complexes 3MEC-11 at the moment of (a) 5 ns and (b) 10 ns, and 3MEC-28 at the moment of (c) 5 ns and (d) 10 ns during the MD simulations.

(*para*-H), 3 (*para*-CH₃) > 8 (*para*-Br), and 14 (*para*-CH₃) > 20 (*para*-Br). It could be observed that three small yellow contours located at the *ortho*-position of the left benzene ring in Fig. 8a

and a middle yellow contour situated at the same place in Fig. 8b, indicating that bulky substituents at the *ortho*-position were not beneficial for the activity. This visualization might

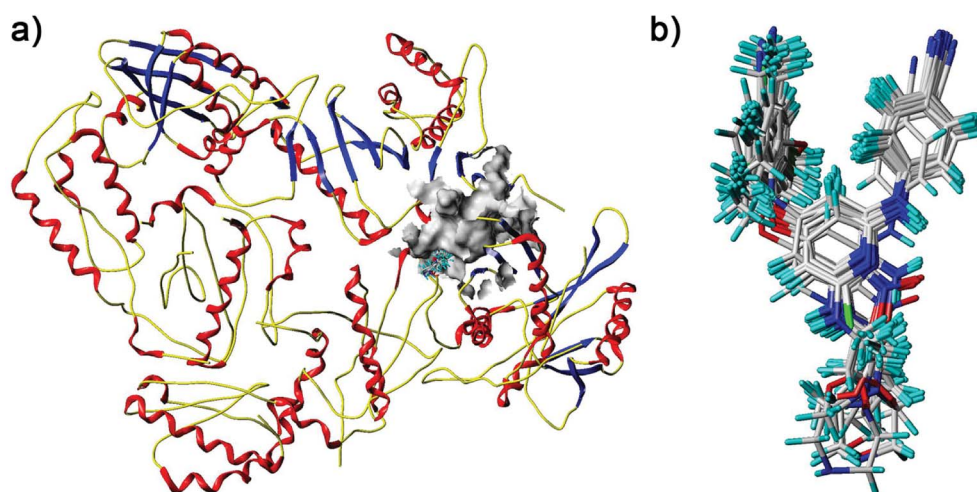


Fig. 6 Docked diarylpyridine derivatives in the nonnucleoside binding site of the HIV-1 RT (PDB ID: 3MEC). (a) The surface of the residues which are 5 Å from ligands are shown in gray color. (b) The docking conformations of all the compounds.

Table 2 PLS statistical results of the CoMFA and CoMSIA models

Model ^a	q_{loo}^2	N	SEE	r_{ncv}^2	F	Field contribution (%)					
						S	E	H	D	A	
CoMFA	S + E	0.665	10	0.084	0.989	221.653	0.396	0.604			
	S + E + H + D + A	0.727	12	0.093	0.988	150.555	0.075	0.254	0.211	0.195	0.266
	S + E	0.613	15	0.110	0.985	85.333	0.229	0.771			
	E + D	0.525	14	0.168	0.964	38.423		0.700		0.300	
	S + E + D	0.679	12	0.119	0.980	90.300	0.142	0.547		0.311	
	S + E + A	0.602	11	0.146	0.969	64.793	0.128	0.473			0.399
	S + H + D	0.543	8	0.194	0.938	48.987	0.192		0.383	0.424	
	S + H + A	0.651	8	0.163	0.956	70.442	0.154		0.376		0.471
	S + D + A	0.607	15	0.112	0.985	81.432	0.205			0.355	0.440
	CoMSIA	E + H + D	0.622	14	0.081	0.992	169.485		0.407	0.309	0.284
E + H + A		0.601	8	0.184	0.944	54.568		0.359	0.264		0.377
E + D + A		0.526	12	0.152	0.968	54.940		0.482		0.188	0.330
H + D + A		0.698	11	0.113	0.981	109.043			0.361	0.274	0.365
S + E + H + D		0.686	13	0.074	0.993	221.377	0.105	0.379	0.264	0.253	
S + E + H + A		0.640	8	0.169	0.953	65.648	0.099	0.319	0.236		0.346
S + E + D + A		0.696	13	0.115	0.982	90.245	0.110	0.387		0.205	0.299
S + H + D + A		0.748	10	0.116	0.979	114.403	0.091		0.310	0.249	0.349
S + E + D + A		0.670	12	0.106	0.984	113.716		0.279	0.232	0.211	0.278

^a S, E, H, D, and A mean steric, electrostatic, hydrophobic, hydrogen-bond donor, and hydrogen-bond acceptor fields, respectively.

Table 3 Statistical validation parameters of the CoMFA and CoMSIA (S + E + H + D + A) models using training and test set

Validation parameters	CoMFA	CoMSIA
r_{pred}^2	0.962	0.912
RMSEC ^a	0.069	0.073
MAE _{training}	0.051	0.049
$r_{\text{o(training)}}^2$	0.989	0.988
$r'_{\text{o(training)}}^2$	0.989	0.988
$r_{\text{m(training)}}^2$	0.989	0.988
$r'_{\text{m(training)}}^2$	0.979	0.974
$\Delta r_{\text{m(training)}}^2$	0.010	0.014
$r_{\text{m(training)}}^2$	0.984	0.981
RMSEP ^b	0.101	0.249
MAE _{test}	0.110	0.181
r^2_{test}	0.955	0.659
$r_{\text{o(test)}}^2$	0.942	0.649
$r'_{\text{o(test)}}^2$	0.951	0.577
$\frac{r^2 - r'_{\text{o(test)}}^2}{r^2}$	0.004	0.015
k	0.993	1.018
k'	1.007	0.982
$r_{\text{m(test)}}^2$	0.948	0.594
$r'_{\text{m(test)}}^2$	0.896	0.470
$\Delta r_{\text{m(test)}}^2$	0.052	0.124
$r_{\text{m(test)}}^2$	0.922	0.532
0.1 × training set range	0.301	0.301

^a Root mean square error of calibration. ^b Root-mean-square error of prediction.

explain why the compound **16** with a methoxy group at the *ortho*-position has low pEC₅₀ value. It was also in agreement with the order of activity data: **3** (*ortho*-Br) > **2** (*ortho*-CH₃), **7** (*ortho*-Cl) > **5** (*ortho*-CH₃), **14** (*ortho*-Br) > **13** (*ortho*-CH₃), **19** (*ortho*-Cl) > **17** (*ortho*-CH₃). A yellow contour covered a part of

the R₂ substituents (compounds **23–42**) and was close to the pyridine ring, indicating that the reducing the R₂ length might increase the antiviral activity, for example, **28** > **25** and **38** > **35**. This result is consistent with a previous conclusion reported by Yang *et al.*²⁵ Moreover, a green contour neighbor to the terminus of the R₂ suggested that the existing of a bulky group at the R₂ terminal might enhance the activity.

The electrostatic fields are presented as blue contours where electropositive substituents are favorable for the activity and as red contours where electronegative substituents are beneficial (Fig. 9). One blue contour was near the nitro and amino groups which is out of the pyridine ring of compounds **1–20**. Although both groups could form hydrogen bonds with the HIV-1 RT, the electrostatic field forces might make their activity different. These results suggested that the amino group was better than the nitro group for the antiviral activity in compounds **1–20**. This was supported by the experimental activities as observed in the following compounds, **12** > **1**, **13** > **2**, **15** > **4**, **14** > **5**, **18** > **6**, **19** > **7**, **20** > **9**. A red contour located close to the chlorine substituent and the R₂ substituents of compounds **21** and **22**, revealing the importance of electronegative groups at this position. A blue contour near the linker between the left benzene ring and the pyridine ring was observed in the CoMFA contour maps (Fig. 9a). This indicated that using an electropositive atom as the linker was beneficial for the activity. Compounds **10** (X = NH, pEC₅₀ = 5.975) and **11** (X = NH, pEC₅₀ = 7.398) with a nitrogen atom as the linker exhibited better activities than the corresponding compounds **4** (X = O, pEC₅₀ = 5.693) and **2** (X = O, pEC₅₀ = 6.538) with an oxygen atom as the linker, respectively.

The hydrophobic, hydrogen-bond donor, and hydrogen-bond acceptor contours of the CoMSIA model were shown in Fig. 10. As shown in Fig. 10a, the yellow and white contours mean the hydrophobic substituents are favorable and

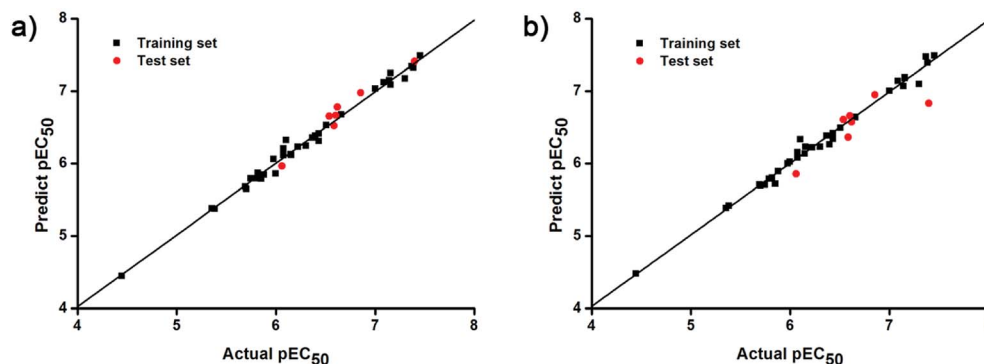


Fig. 7 Plots of actual versus predicted pEC_{50} values for training (black squares) and test (red dots) set compounds on the basis of (a) the CoMFA and (b) the best CoMSIA models.

Table 4 Sensitivity, specificity, accuracy, PPV, NPV, and MCC values of the CoMFA and CoMSIA (S + E + H + D + A) models

Model	Sensitivity	Specificity	Accuracy	PPV	NPV	MCC
CoMFA	0.9667	0.9167	0.9524	0.9667	0.9167	0.8833
CoMSIA	0.9667	0.8333	0.9286	0.9355	0.9091	0.8220

unfavorable for the activity, respectively. There were two yellow contours that located at the *ortho*-position of the left benzene ring, indicating that the introduction of hydrophobic groups in this region might enhance the activity. The compounds with the relative hydrophobic substituents ($-CH_3$, $-F$, $-Cl$ and $-Br$) at the *ortho*-position of the left benzene ring exhibited good antiviral activity. A white contour situated at the end of the R_2 substituents (compounds 23–42), which pointed to the entrance channel of the binding pocket. Combined with the steric contour maps, this indicated that hydrophilic bulky groups at the R_2 terminus were favorable for the activity. As shown in Fig. 10b, a cyan contour near the imino group that connected

the pyridine ring with the right benzene ring indicated that a hydrogen-bond donor at this position was advantage of the activity. The previously docked results were also demonstrated this structure was important for the diarylpyridines to form hydrogen bonds with the residue Lys101 of the HIV-1 RT. A purple contour was near the imino group of the R_2 substituents, indicating a hydrogen-bond donor was not improper here. As shown in Fig. 10c, the magenta contour indicated hydrogen-bond acceptors were favored and the red contour represented hydrogen-bond acceptors were disfavored. Two magenta contours were near the *para*-positions of the right and left benzene rings, respectively. However, in the previous docking study, hydrogen bonds were not observed at these positions.

Pharmacophore elucidation

GALAHAD was applied to study the pharmacophore characteristics of the diarylpyridine derivatives. A possible pharmacophore model was constructed using ten diarylpyridines with high activities and diverse structures (Fig. 11). The different

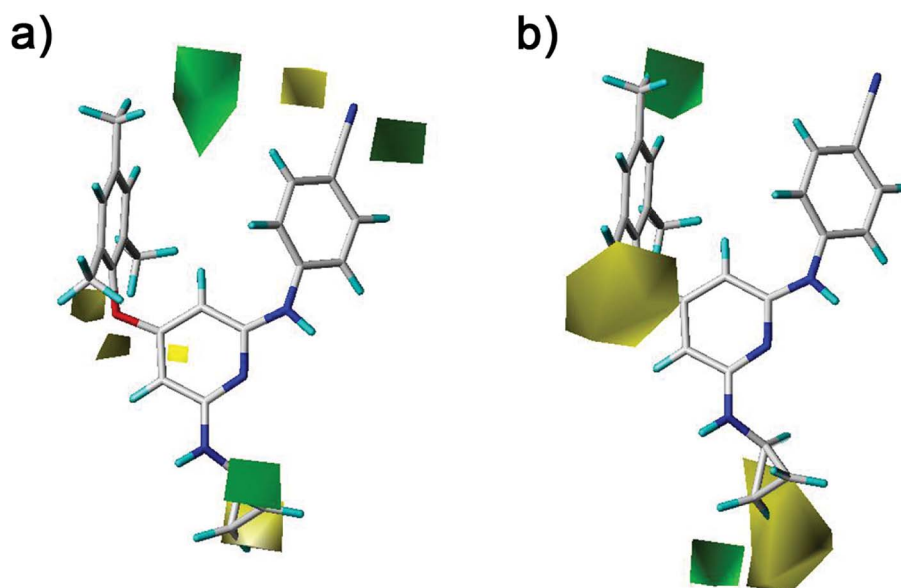


Fig. 8 Contour maps of steric fields derived from (a) the CoMFA and (b) the best CoMSIA models using compound 28 as a reference molecule.

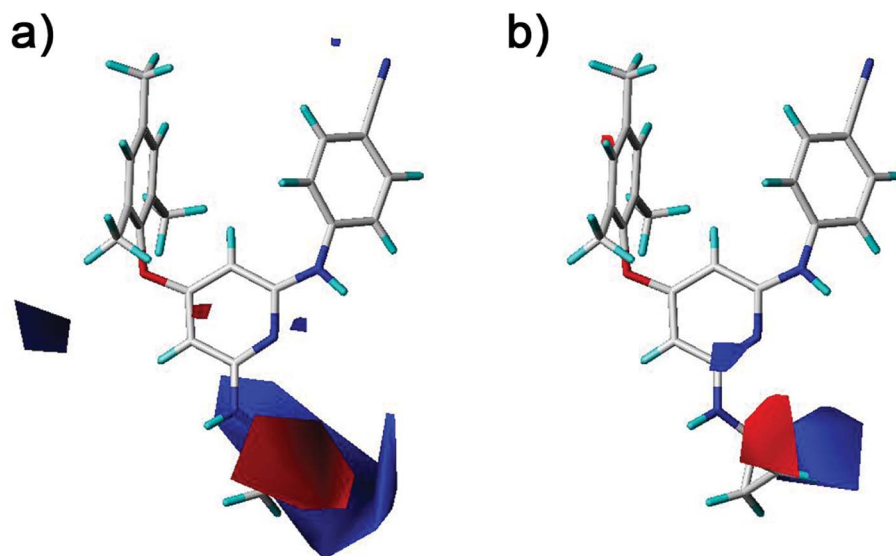


Fig. 9 Contour maps of electrostatic fields derived from (a) the CoMFA and (b) the best CoMSIA models using compound **28** as a reference molecule.

colors represent different pharmacophore features. As shown in Fig. 11, the built model contained three hydrophobe (HY, cyan) features, three hydrogen-bond acceptors (HAS, green), and three hydrogen-bond donors (HDs, magenta). Three HY features located at the center of the three aromatic rings, respectively. This indicated that maintaining hydrophobic structures in common scaffolds of NNRTI diarylpyridines were necessary to interact with hydrophobic residues in the NNIBP of the HIV-1 RT. The HA3 feature suggested that the atom of the left linker should be a hydrogen-bond acceptor. A HA feature at the cyano group of the right benzene ring indicated that the nitrogen atom of the cyano group might be a hydrogen-bond acceptor

candidate. It could be observed that there was a HD/HA feature at the same nitrogen atom that linked the pyridine ring and the right benzene ring. The HD3 feature was a big sphere which covered the nitrogen atom of the R₂ substituent of compounds **28**, **33**, **34**, and **40**. This indicated that a hydrogen-bond donor at this space had relatively flexible distance to the pyridine ring. In addition, the nitrogen atoms of the nitro groups were out of the sphere HD2 whereas the nitrogen atoms of the amino groups were included in the sphere. This result was consistent with the 3D-QSAR contour maps and experimental activities, indicating an amino group (hydrogen-bond donor) was better than a nitro group (hydrogen-bond acceptor) for improving the activity.

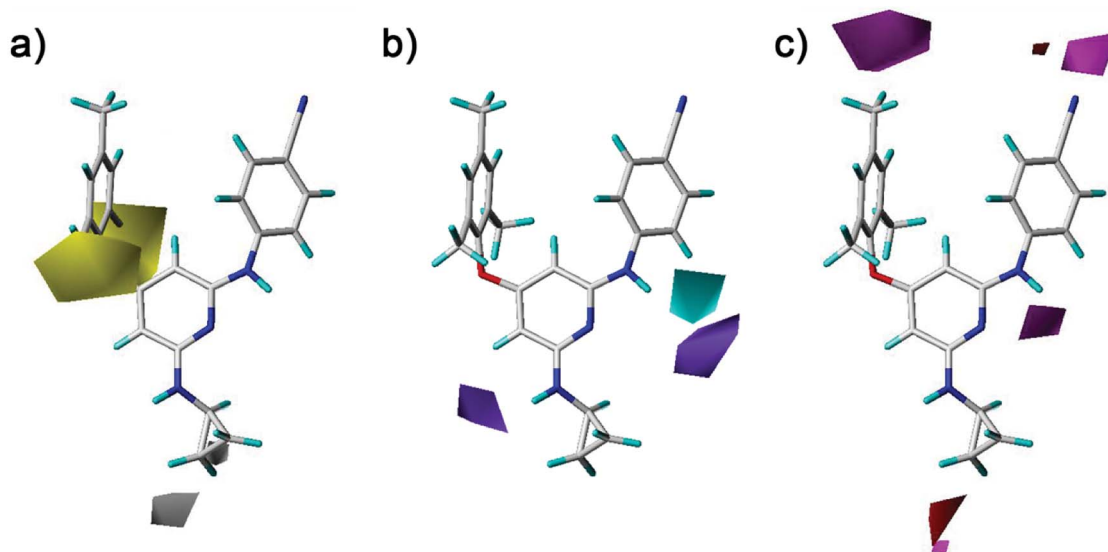


Fig. 10 Contour maps of (a) hydrophobic, (b) hydrogen-bond donor, and (c) hydrogen-bond acceptor fields of the best CoMSIA model using compound **28** as a reference molecule.

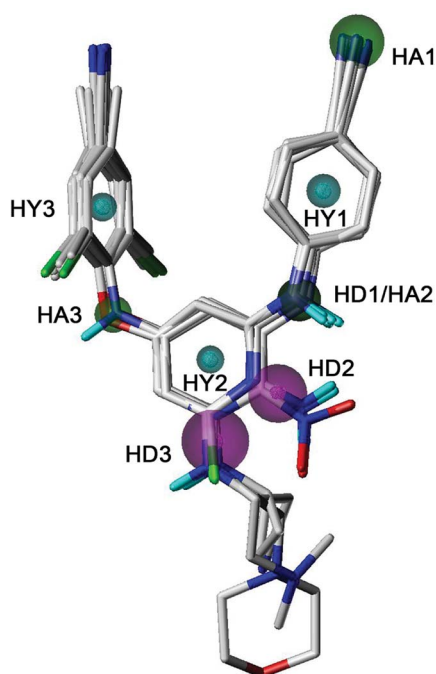


Fig. 11 The best pharmacophore model aligned with ten diarylpyridine derivatives of HIV-1 RT NNRTIs. The hydrophobic, hydrogen-bond acceptor, and hydrogen-bond donor features are colored in cyan, green, and magenta, respectively.

Conclusions

In the present study, a novel series of diarylpyridine derivatives were docked into the HIV-1 RT and the 10 ns MD simulations were performed on complexes of the HIV-1 RT with two diarylpyridine derivatives (compounds **11** and **28**). The CoMFA, CoMSIA, and GALAHAD were performed on these docked conformations of the diarylpyridine derivatives to understand their 3D-QSAR relationships and pharmacophore characteristics. The docking and pharmacophore results suggested that diarylpyridine derivatives as NNRTIs should contain hydrophobic parts because the NNIBP of the HIV-1 RT was consisted of hydrophobic residues. The maintenance of the left aromatic ring is necessary to interact with Tyr181, Tyr188, Phe227, and Trp229 by π - π stacking and hydrophobic interactions in the NNIBP. The results also manifested that the group of the right linker should be a hydrogen-bond acceptor/donor to interact with the key residue Lys101 by hydrogen bonds. The statistical results of the 3D-QSAR models demonstrated that the constructed CoMFA and CoMSIA models had preferable predictive ability for novel diarylpyridines or DAPYs. Based on the comprehensive consideration of the docking, 3D-QSAR, and pharmacophore results, the following substituents that are introduced into corresponding positions might be beneficial for improving the anti-HIV-1 activity of diarylpyridines: (1) small hydrophobic substituents at the *ortho*-positions of the left aromatic ring; (2) bulky groups at the *para*-position of the left aromatic ring; (3) a short chain as R_2 with a bulky hydrophilic terminus; (4) a electropositive substituent as the left linker and

a hydrogen-bond donor as the right linker. In conclusion, the molecular modeling results in this paper might provide significant information for the design and development of novel diarylpyridines targeting to the HIV-1 RT.

Conflicts of interest

There are no conflicts to declare.

Acknowledgements

This work was supported by National Natural Science Foundation of China (No. 21807082, 21877087), Hubei Provincial Natural Science Foundation of China (No. 2017CFB121), Hubei Provincial Department of Education of China (No. Q20171503), and Wuhan International Scientific and Technological Cooperation Project (No. 2017030209020257).

References

- 1 Y. Hsiou, J. Ding, K. Das, J. A. D. Clark, S. H. Hughes and E. Arnold, *Structure*, 1996, **4**, 853–860.
- 2 T. Ilina, K. LaBarge, S. G. Sarafianos, R. Ishima and M. A. Parniak, *Biology*, 2012, **1**, 521–541.
- 3 E. D. Clercq, *J. Med. Chem.*, 1995, **38**, 2491–2517.
- 4 H. Mitsuya, R. Yarchoan and S. Broder, *Science*, 1990, **249**, 1533–1544.
- 5 H. Jonckbeere, J. Anne and E. D. Clercq, *Med. Res. Rev.*, 2000, **20**, 129–154.
- 6 D. D. Richman, *Nature*, 2001, **410**, 995–1001.
- 7 L. R. Boone, *Curr. Opin. Invest. Drugs*, 2006, **7**, 128–135.
- 8 N. Sluis-Cremer, D. Arion and M. A. Parniak, *Cell. Mol. Life Sci.*, 2000, **57**, 1408–1422.
- 9 R. S. Goody, B. Müller and T. Restle, *FEBS Lett.*, 1991, **291**, 1–5.
- 10 E. D. Clercq, *Antiviral Res.*, 1998, **38**, 153–179.
- 11 C. Tantillo, J. Ding, A. Jacobo-molina, R. G. Nanni, P. L. Boyer, S. H. Hughes, R. Pauwels, K. Andries, P. A. Janssen and E. Arnold, *J. Mol. Biol.*, 1994, **243**, 369–387.
- 12 J. Ding, K. Das, C. Tantillo, W. Zhang, A. D. Clark, S. Jessen, X. Lu, Y. Hsiou, A. Jacobo-Molina and K. Andries, *Structure*, 1995, **3**, 365–379.
- 13 J. S. Ren, E. Robert, G. Elspeth, S. Donald, R. Carl, K. Ian, K. James, D. Graham, J. Yvonne, S. David and S. David, *Nat. Struct. Biol.*, 1995, **2**, 293–302.
- 14 J. Ding, K. Das, H. Moereels, L. Koymans, K. Andries, P. A. Janssen, S. H. Hughes and E. Arnold, *Nat. Struct. Biol.*, 1995, **2**, 407–415.
- 15 E. L. Asahchop, M. A. Wainberg, R. D. Sloan and C. L. Tremblay, *Antimicrob. Agents Chemother.*, 2012, **56**, 5000–5008.
- 16 P. Zhan, X. Chen, D. Li, Z. Fang, E. D. Clercq and X. Liu, *Med. Res. Rev.*, 2013, **33**, E1–E72.
- 17 E. D. Clercq, *Chem. Biodiversity*, 2004, **1**, 44–64.
- 18 P. Zhan, X. Liu, Z. Li, C. Pannecouque and E. D. Clercq, *Curr. Med. Chem.*, 2009, **16**, 3903–3917.

- 19 G. Liu, W. Wang, Y. Wan, X. Ju and S. Gu, *Int. J. Mol. Sci.*, 2018, **19**, 1436.
- 20 G. Liu, Y. Wan, F. Sai, S. Gu and X. Ju, *Mol. Diversity*, 2018, DOI: 10.1007/s11030-018-9860-1.
- 21 R. Schrijvers, *Expert Opin. Pharmacother.*, 2013, **14**, 1087–1096.
- 22 D. Li, P. Zhan, E. D. Clercq and X. Liu, *J. Med. Chem.*, 2012, **55**, 3595–3613.
- 23 W. Wang, Y. Tian, Y. Wan, S. Gu, X. Ju, X. Luo and G. Liu, *Struct. Chem.*, 2018, DOI: 10.1007/s11224-018-1204-3.
- 24 Y. Song, Z. Fang, P. Zhan and X. Liu, *Curr. Med. Chem.*, 2014, **21**, 329–355.
- 25 J. Yang, W. Chen, D. Kang, X. Lu, X. Li, Z. Liu, B. Huang, D. Daelemans, C. Pannecouque and E. D. Clercq, *Eur. J. Med. Chem.*, 2016, **109**, 294–304.
- 26 Z. Liu, Y. Tian, J. Liu, B. Huang, D. Kang, E. D. Clercq, D. Daelemans, C. Pannecouque, P. Zhan and X. Liu, *Eur. J. Med. Chem.*, 2017, **140**, 383–391.
- 27 M. J. D. Powell, *Math. Program.*, 1977, **12**, 241–254.
- 28 A. Paneth, W. Płonka and P. Paneth, *J. Mol. Model.*, 2017, **23**, 317–325.
- 29 E. B. Lansdon, K. M. Brendza, M. Hung, R. Wang, S. Mukund, D. Jin, G. Birkus, N. Kutty and X. H. Liu, *J. Med. Chem.*, 2010, **53**, 4295–4299.
- 30 J. Ruppert, W. Welch and A. N. Jain, *Protein Sci.*, 1997, **6**, 524–533.
- 31 B. Hess, H. Bekker, H. J. C. Berendsen and J. G. E. M. Fraaije, *J. Chem. Theory Comput.*, 2008, **4**, 1463–1472.
- 32 D. M. York, T. A. Darden and L. G. Pedersen, *J. Chem. Phys.*, 1993, **99**, 8345–8348.
- 33 J. Verma, V. M. Khedkar and E. C. Coutinho, *Curr. Top. Med. Chem.*, 2010, **10**, 95–115.
- 34 A. Tropsha, P. Gramatica and V. K. Gombar, *QSAR Comb. Sci.*, 2003, **22**, 69–77.
- 35 P. P. Roy and K. Roy, *QSAR Comb. Sci.*, 2010, **27**, 302–313.
- 36 A. Golbraikh and A. Tropsha, *J. Mol. Graphics Modell.*, 2002, **20**, 269–276.
- 37 P. K. Ojha, I. Mitra, R. N. Das and K. Roy, *Chemom. Intell. Lab. Syst.*, 2011, **107**, 194–205.
- 38 K. Roy, R. N. Das, P. Ambure and R. B. Aher, *Chemom. Intell. Lab. Syst.*, 2016, **152**, 18–33.
- 39 V. P. Zambre, R. Giridhar and M. R. Yadav, *Med. Chem. Res.*, 2013, **22**, 4685–4699.
- 40 N. J. Richmond, C. A. Abrams, P. R. Wolohan, E. Abrahamian, P. Willett and R. D. Clark, *J. Comput.-Aided Mol. Des.*, 2006, **20**, 567–587.
- 41 D. Yusuf, A. M. Davis, G. J. Kleywegt and S. Schmitt, *J. Chem. Inf. Model.*, 2008, **48**, 1411–1422.
- 42 K. Das, P. J. Lewi, S. H. Hughes and E. Arnold, *Prog. Biophys. Mol. Biol.*, 2005, **88**, 209–231.