Journal of Computer-Aided Molecular Design (2005) 19: 243–258
DOI 10.1007/s10822-005-4790-2

Computational studies and drug design for HIV-1 reverse transcriptase inhibitors of 3′,4′-di-O-(S)-camphanoyl-(+)-cis-Khellactone (DCK) analogs

Hai-feng Chen[a,c,f], Bo-tao Fan[a,*], Chen-yang Zhao[a], Lan Xie[c], Chun-hong Zhao[b], Ting Zhou[d], Kuo-Hsiung Lee[d] & Graham Allaway[e]

[a]*Department of Chemistry, University Paris 7-Denis Diderot, 1 rue Guy de la Brosse 75005, Paris, France;* [b]*Beijing Institute of Pharmacology and Toxicology, 100850, Beijing China;* [c]*Laboratory of Computer Chemistry, Shanghai Institute of Organic Chemistry, 200032 Shanghai, China;* [d]*School of Pharmacy, University of North Carolina at Chapel Hill, 27599 Chapel Hill, NC, USA;* [e]*Panacos Pharmaceuticals Inc, 209 Perry Parkway, 20877 Gaithersburg, MD , USA;* [f]*Present address: Department of Molecular Biology and Biochemistry, University of California, Irvine, CA, 92697-3900, USA*

Received 6 December 2004; accepted 25 March 2005

Key words: DCK, drug design, molecular dynamics simulation

Summary

Molecular docking and molecular dynamics simulation were applied to study the binding mode of 3′,4′-di-*O*-(*S*)-camphanoyl-(+)-*cis*-khellactone (DCK) analogs anti-HIV inhibitors with HIV-1 RT. The results suggest that there is a strong hydrogen bond between DCK O16 and NH of Lys101, and that DCK analogues might act similarly as other types of HIV-1 RT inhibitors. The investigation about drug resistance for DCK shows no remarkable influence on the most frequently observed mutation K103N of HIV-1 RT. Based on the proposed mechanism, some new structures were designed and predicted by a SVM model. All compounds exhibited potent inhibitory activities against HIV replication in H9 lymphocytes with EC_{50} values lower than 1.95 μM. The rationality of the method was validated by experimental results.

Introduction

HIV-1 reverse transcriptase (HIV-1 RT) is one of the important targets for the rational drug development of anti-AIDS pharmaceuticals [1]. In general, the inhibitors of HIV-1 RT are classified into two main categories: nucleoside inhibitors (NRTIs) such as AZT [2], ddl [3], ddc [4] and 3TC [5], and non-nucleoside inhibitors (NNRTIs), like HEPT, TIBO, Nevirapine and BHAP [6]. Because NNRTIs have the advantage of high potency, low toxicity and exquisite selectivity [1], the investigations of new inhibitors of this category are always in the headlines. The developed drug-resistance of HIV against the current drugs decreases rapidly the efficiency of these compounds. The discovery of new NNRTIs is always important. 3′,4′-di-*O*-(*S*)-camphanoyl-(+)-*cis*-khellactone (DCK, M01) was identified as a potent anti-HIV NNRT agent (EC_{50} = 4.90×10^{-2} μM) with a therapeutic index (TI) of 1.37×10^5 in H9 lymphocytes [7], which is more efficient than AZT (EC_{50} = 4.6×10^{-2} μM and TI = 10,000) in the same assay. But the poor solubility and bioavailability limited the further development of this compound toward clinical drugs.

HEPT, Nevirapine [8–10], and TIBO [11, 12] were investigated for their inhibition mechanism on HIV-1 RT. However, there is no related

*To whom correspondence should be addressed. Fax: +33-1-44276814; E-mail: fan@paris7.jussieu.fr

report about the binding mode between DCK analogues and HIV-1 RT. In this paper, we report the work about the applications of molecular docking, molecular dynamics simulation and Support Vector Machine (SVM) to study the possible inhibitory mechanism of DCK analogues on HIV-1 RT. Based on these investigations, some new compounds were designed and synthesized. All these synthesized compounds exhibit inhibitory activity. One of them shows high activity ($EC_{50} = 2.0 \times 10^{-4}$ μM). This theoretical study can help us to well understand the interaction mechanism between inhibitors DCK analogues and HIV-1 RT, and to make quantitative prediction of their inhibitory activities.

Computational methods and materials

Dataset for Analysis

The structures of DCK analogues and activities EC_{50} (concentration in μM of 50% bounding to the H9 cells) are listed in Table 1.

Support Vector Regression Method

SVM is a new and very promising classification and regression method developed by Vapnik et al. [13]. A detailed description of SVM theory can be found in several excellent books and tutorials [14–17]. SVM was originally developed for classification problems, but they can also be extended to solve non-linear regression problems by the introduction of ϵ-insensitive loss function.

SVM approach has been proposed to minimize the structural risk rather than the empirical risk; that is to preserve a good generalization ability rather than optimizing the agreement with a given (limited) training set, and so constitutes a trade-off between the complexity of the model and its capability to reproduce experimental observations.

The regression performance depends on a good setting of parameters: C, ϵ and the kernel type and corresponding kernel parameters. Parameter C is a regularization constant which determines the trade off between the model complexity and the degree to which deviations larger than ϵ are tolerated in optimization formulation. The selection of the kernel function and corresponding parameters is

Table 1. Structures of DCK and analogs, bioactivities (pEC_{50}) and log *P*.

No.	Structure	pEC_{50}	log *P*
M01	DCK(R3 = R4 = R5 = R6 = H)	1.31	4.94
M02	3-Me-DCK	1.56	5.22
M03	4-Me-DCK	2.23	5.09
M04	5-Me-DCK	2.22	5.40
M05	4,5-di-Me-DCK	0.92	5.56
M06	4,6-di-Me-DCK	1.24	5.56
M07	3-OMe-DCK	1.98	4.07
M08	4-OMe-DCK	1.29	3.94
M09	5-OMe-DCK	1.36	4.68
M10	3-Cl-4-Me-DCK	3.12	4.99
M11	3-CH_2Br-DCK	1.23	5.42
M12	3-CH_2OAc-DCK	1.17	4.34
M13	3-CH_2OH-DCK	1.46	4.21
M14	3-CH_2NH_2-DCK	0.17	3.87
M15	3-CH_2NEt_2-DCK	−0.57	5.32
M16	3-CH_2OAc-4-Me-DCK	1.55	4.50
M17	6-CH_2OH-4-Me-DCK	0.95	4.56
M18	3-CH_2OH-4-Me-DCK	1.79	4.37
M19	3-$CHBr_2$-DCK	0.02	5.35
M20	6-CH_2Br-DCK	1.61	5.61
M21	3-CH_2Br-4-Me-DCK	2.13	5.58

very important because they define the distribution of the training set of samples in the high dimensional feature space.

All SVM models in our present study were implemented using the shareware program LibSVM developed by Chih-Chung Chang and Chih-Jen Lin [18]. The radial basis function (RBF) was used as kernel function in this work. For RBF kernel, the most important parameter is the width (γ) of the radial basis function. All calculation programs implementing SVM were written in M-file based on MATLAB script.

Molecular Docking

The complex of HIV-1 RT and ligand was extracted from PDB Databank (PDB code: 1C1B). The ligand compound GCA186 which is complexed with HIV-1 RT in the X-ray crystal structure is shown in Figure 1. Three-dimensional structures building and modeling were performed using the SYBYL 6.9 [19] program package. Energy minimization was performed using Tripos force field [20] with a distance-dependent dielectric and the Powell conjugate gradient algorithm with a convergence criterion of 0.01 kcal/(mol Å). All calculations were performed on a SGI Octane2 and Origin 300 workstation.

The molecule M03 (see Figure 2) was selected as reference compound in bioactivity tests. Because the HIV-1 RT inhibitors investigated and GCA186 have six-member heterocycles (Figure 2), we need only to align their heterocycles (atom number: 11,2,3,4,5 and 6 for M03, atom number: 1,2,3,4,5 and 6 for GCA186) together. The active site was determined based on structure of the co-crystallized GCA186-protein complex. Firstly, the ligand GCA186 was extracted from the HIV-1 RT complex, and then the remaining receptor structure was completed by adding polar hydrogen atoms and missing residues (Glu89 and Val90). This structure was optimized with constrained molecular dynamics available within the SYBYL 6.9 program. Partial atom charges of HIV-1 RT were calculated with Kollman-all-atom approximation [21]. Gasteiger–Hückel charges were calculated for all inhibitor molecules, as recommended in the AutoDock 3.0 package, which was used for performing automated docking of inhibitors to receptor. The development and the principle of this software have been described elsewhere [22,23]. During the docking process, a series of the docking parameters were set on. The number of generations, energy evaluations, and docking runs were set to 370,000, 1,500,000 and 50, respectively.

Molecular Dynamics (MD) Simulations

Molecular dynamics simulations were performed using the GROMACS 3.2 package [24,25]. For all atoms, the standard parameters of the GROMACS force field were used. The initial coordinates of the complex for the reference (active) molecule M03 and HIV-1 RT, M03 and mutated HIV-1 RT (K103N) were extracted from the energy minimized molecular docking conformation. The complex to be computed was centered in a periodic, rectangular simulation box with a

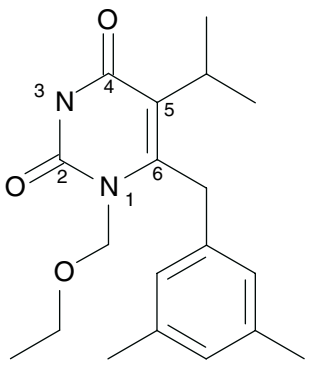

Figure 1. The structure of ligand GCA186 complexed with HIV-1 RT in X-ray crystal (PDB code :1C1B).

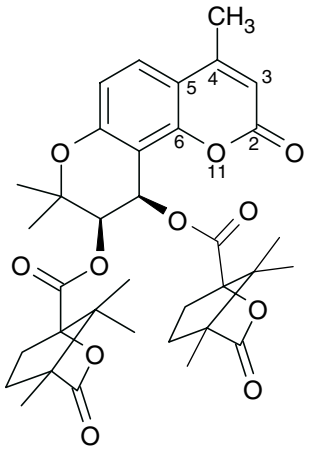

Figure 2. The structure of M03.

minimum distance of 0.5 nm between the molecule and the box wall. The box was filled with SPC water molecules. To relax the configuration of the solvent and the local strain (due to generation of hydrogen positions), and to remove bad Van der Waals contacts (particles that are too close) a steepest descent minimization of 10,000 steps was adopted using GROMACS. The optimized structures were then heated stepwise, to a finite temperature of 300 K. The time step used in all calculations was 1.0 fs. After 20 ps equilibration, MD simulation was employed to record time trajectory. The complex was simulated for 1 ns at 300 K. Coordinates were saved every 0.02 ps for the purpose of subsequent analysis. This process was applied to all complexes formed between the studied compounds and HIV-1 RT.

Calculation of Descriptors

Twenty-seven Charged Partial Surface Area (CPSA) [26] and 56 VolSurf descriptors [27] were calculated with default methods implemented in SYBYL 6.9. But only three descriptors were selected after a number of tests with training set compounds. They are partial positive surface area (PPSA1), hydrophobic regions at sixth energy level (D6) and polarizability (POL). Log *P*, HOMO and LUMO were calculated with HYPERCHEM 7.0.

Results and Discussion

Stability of structure

For M03-HIV-1 RT complex obtained from molecular docking, 1 ns molecular dynamics simulation was performed. To examine the variations in the intramolecular conformations of HIV-1 RT, the root-mean-squared deviation (rmsd) with respect to the starting structure was calculated. Simulation time vs. rmsd of the backbone of the protein during the full simulation (1000 ps of molecular dynamics) with M03 is presented in Figure 3. The rmsd variation of HIV-1 RT was about 0.50 nm. The backbone of HIV-1 RT demonstrated a relatively small variation after 600 ps. This shows that the complex becomes stable after 600 ps of simulation. The total energy of the system and solvent accessible surface for

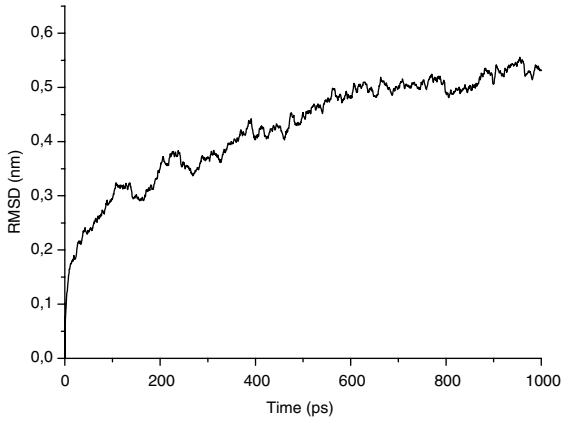

Figure 3. Simulation time vs. rmsd of the backbones of HIV-1 RT.

HIV-1 RT are illustrated in Figures 4 and 5. The average energy of complex is about −1,131,000 kcal/mol. The average solvent accessible surface [28] for HIV-1 RT is about 2830 nm^2.

The principal hydrophobic and hydrogen bonding interactions are represented in Figure 6. There is one important hydrogen bond between oxygen (O16) atom of carbonyl in MO3 and the nitrogen atom in residue Lys101 (O16...HN), with a distance of 2.83 Å and an angle (O···H-N) 152°. Figure 7 illustrates the variations of distance between O16 of DCK and HN of Lys101, O11 of DCK and HN of Lys101. The average distance of O16...H-N (Lys101) is about 2.2 Å. The variation for the angle of O16...H-N is presented in Figure 8. The average angle is about 140°. This suggests that there is a strong hydrogen bond between O16 and HN (Lys101). The average distance of O11...H-N (Lys101) is about 3.5 Å. The hydrogen bond between O11 and HN (Lys101) is therefore very weak. During 1 ns simulation, this hydrogen bond just exists a few ps. For both NNRTs TIBO and HEPT, there is also one hydrogen bond between inhibitor and the residue Lys101 [30]. This illustrates that DCK analogues might have similar interaction mechanism to the other types of NNRTs. There are also some important hydrophobic interactions. The camphanoyl substituent at position 3′ has hydrophobic interactions with the residues Ile94, Pro95, His96 and Trp229, and electrostatic interactions with the residues of Tyr181, Tyr183 and Tyr188. For another camphanoyl substituent at position 4′, there are electrostatic interactions with the residues

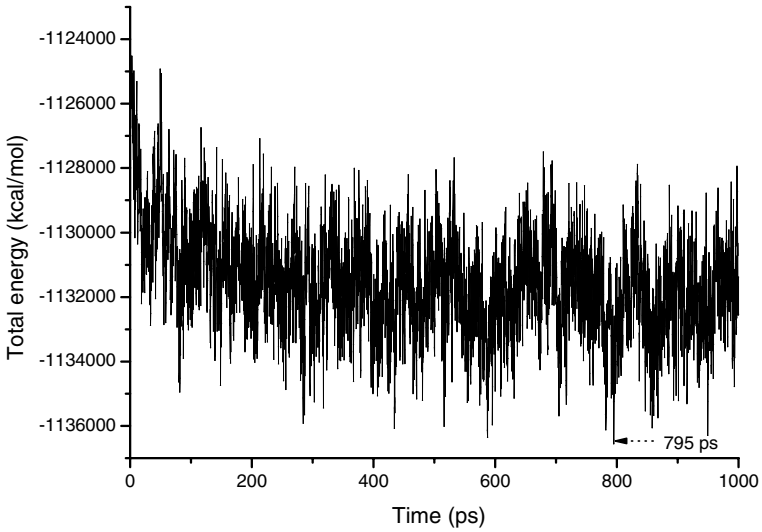

Figure 4. Total energy changes of M03-HIV-1 RT complex as the function of time. Dotted arrow indicates the local energy-minima conformation.

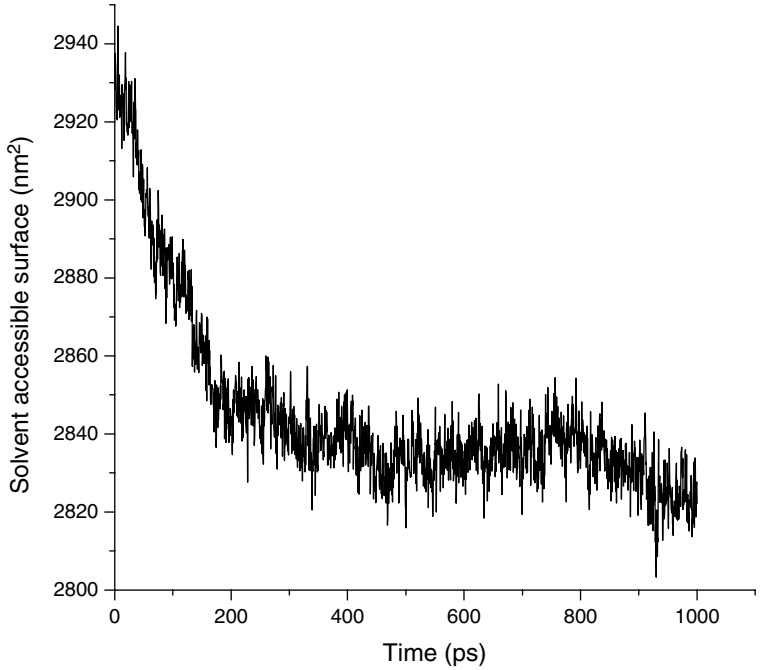

Figure 5. The variation of solvent accessible surface for HIV-1 RT during the simulation.

of His235, Phe227 and Tyr318, and hydrophobic interactions with the residues of Val106, Pro225, Leu234 and Pro236. These hydrogen bond and interactions contribute significantly to the stabilization of complex.

Because DCK analogues have different substituents at positions 3, 4, 5 and 6 (Table 1), we selected these positions to investigate the interaction between M03 and HIV-1 RT. Figure 9 illustrates lipophilic potential (a) [31] and electrostatic potential (b) of M03 and residues within 6 Å around the substituents of positions 3 and 4 (see Figure 2). The substituents at positions 3 and 4 of khellactone are near the polar residues Asn136

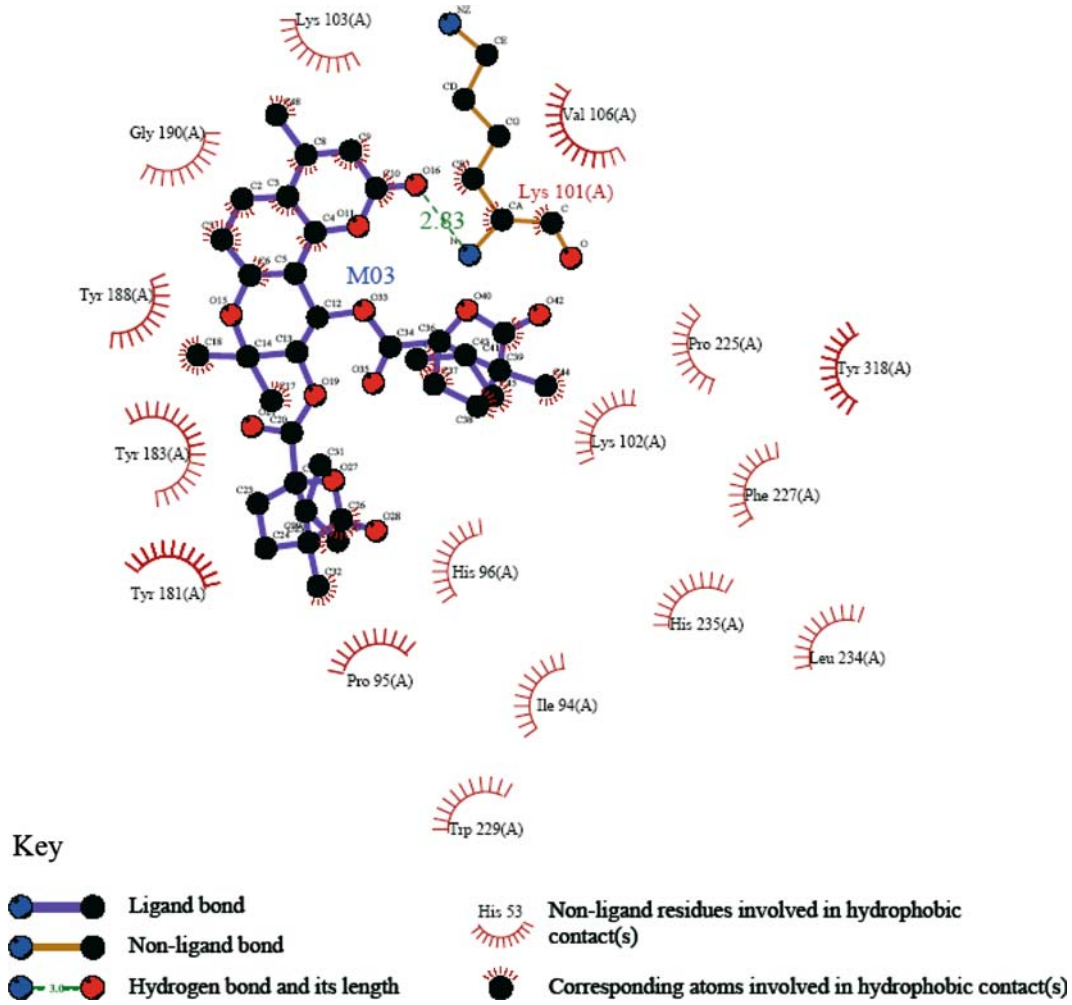

Figure 6. Two-dimensional representation for the interacting mode of compound M03 with HIV-1 RT, drawn by LIGPLOT program [29].

(B chain), Asn137 (B chain), Glu138(B chain) and Lys101 (A chain). The blue regions of lipophilic potential near the positions 3 and 4 suggest that hydrophilic substituents of inhibitor could be favorable to anti-HIV. Red regions of electrostatic potential near the positions 3 and 4 show that electronegative groups are favorable to activity. Among the substituents Cl, CH_2Br and CH_2OH at position 3, their Hammett constants σ_m are 0.37, 0.12 and 0.0, respectively [32]. Therefore, the capacity of accepting electron is in the following sequence: M10(3-Cl) > M21(3-CH_2Br) > M18 (3-CH_2OH), in agreement with experimental observation. The distance between these residues and the substituents of positions 3 and 4 is only about 4 Å, therefore bulk groups might be unfavorable to activity. This explains that compound M15 (3-CH_2NEt_2) has low activity.

The lipophilic and electrostatic potentials of M03 and residues within 6 Å around substituents of positions 5 and 6 are presented in Figure 10. The residues Val179, Val189, Gly190 and Ser191 have interactions with the substituents of positions 5 and 6. Brown regions of lipophilic potential near the substituent of position 5 and 6 show that hydrophobic groups could be favorable to the interaction between inhibitor and HIV-1 RT. Red and cyan regions of electrostatic potential near the substituents of positions 5 and 6 suggest that electronegative groups could increase the activity. This could explain that the activity of M20 (6-CH_2Br) is higher than that of M17 (6-CH_2OH).

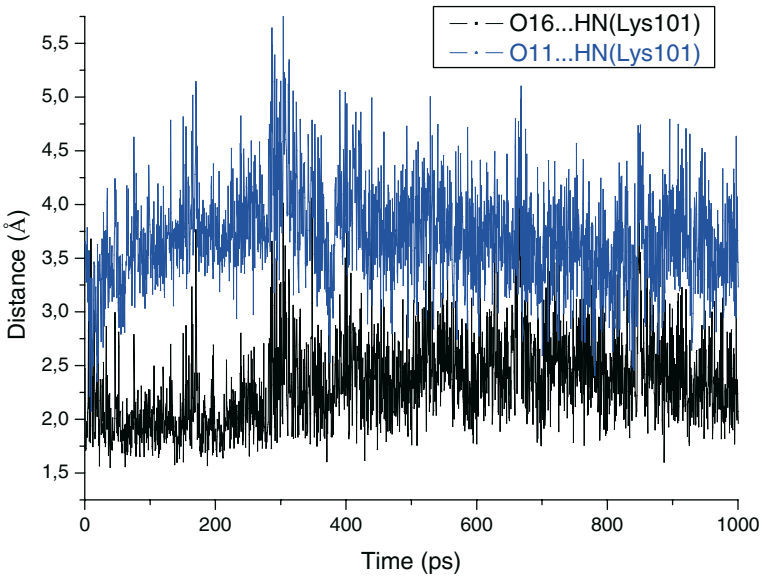

Figure 7. Simulation time vs. distances of O16 HN(Lys101) and O11...HN(Lys101).

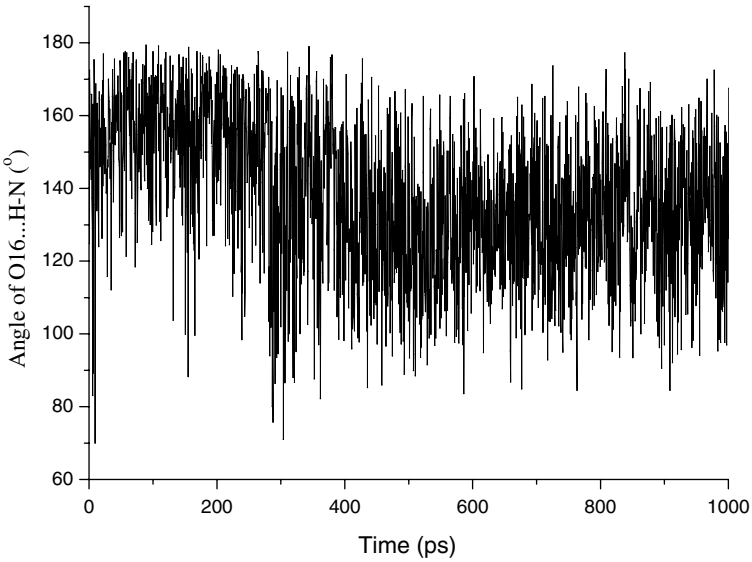

Figure 8. Simulation time vs. angle O16...H-N.

These investigations could help us to better understand the interaction between inhibitor and HIV-1 RT.

Figure 11 illustrates the superposition results of M03 and **4b** (3-CH_2CN-4-Me-DCK), compound proposed to be synthesized based on docking study (see Table 4). Their three fused rings could be aligned together. This type of compounds has same interaction mechanism with HIV-1 RT.

Analysis of drug resistance

The efficiency of NNRTIs is seriously compromised by the emergence of mutant viral strains [33]. Some mutations, most notably K103N, have been selected both *in vitro* and *in vivo* by most currently available NNRTIs [34]. K103N is also the most frequently observed mutation among patients failing highly active

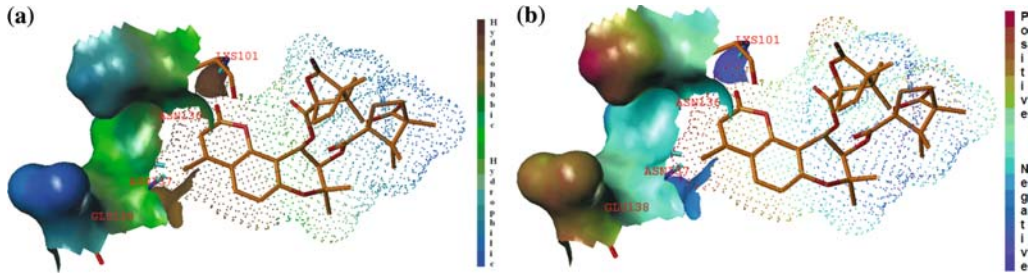

Figure 9. Lipophilic potential (a) and electrostatic potential (b) of M03 and residues within 6 Å around substituents of positions 3 and 4. The structure in dot sphere represents M03, the other ones are the residues of HIV-1 RT. Brown represents highest lipophilic area of the molecule, blue highest hydrophilic area, red positive charge, purple negative charge.

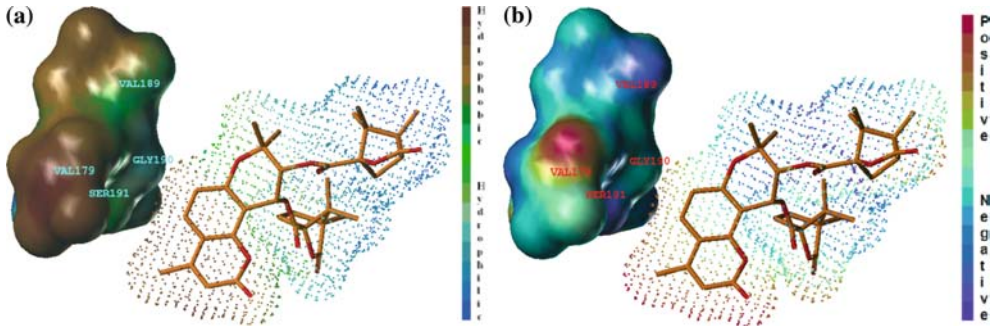

Figure 10. Lipophilic potential (a) and electrostatic potential (b) of M03 and residues within 6 Å around substituents of positions 5 and 6. The structure in dot sphere represents M03, the other ones are the residues of HIV-1 RT. Brown represents the highest lipophilic area of the molecule, blue the highest hydrophilic area, Red the positive charge, purple the negative charge.

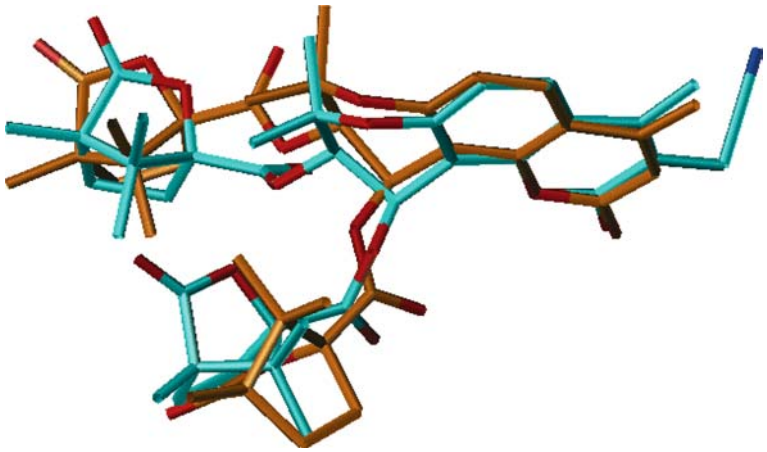

Figure 11. Superposition of compounds M03 and **4b** (orange: M03; cyan: **4b**).

antiretroviral therapy (HAART) because it confers resistance to all of the clinically approved NNRTIs [35]. DCK analogues as one type of the NNRTIs, the investigation of drug resistance for these compounds could contribute to the design of more effective therapeutic agents.

The mutated structure of HIV-1 RT (K103N) was optimized using Tripos force field with SYBYL6.9. Then the DCK analogues were docked into mutated HIV-1 RT. The histogram of binding free energy for these compounds with wild and mutated HIV-1 RT is presented in Figure 12. The binding

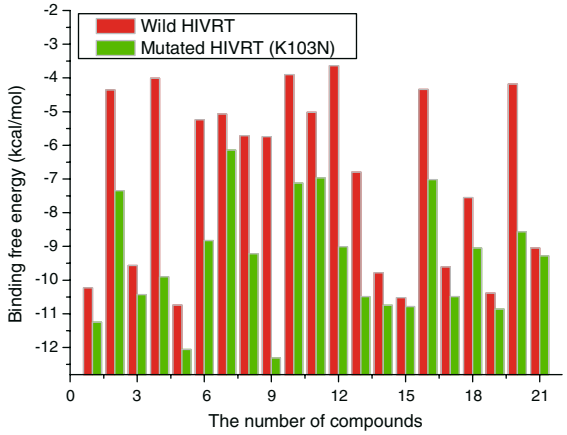

Figure 12. Histogram of binding free energy of DCK analogues with wild and mutated HIV-1 RT.

free energy between each compound and mutated HIV-1 RT is lower than that of wild receptor.

To examine the variations in the intramolecular conformations of the mutated HIV-1 RT and ligand within the solvent environment, 1.0 ns molecular dynamics simulation was performed. Simulation time vs. rmsd of the backbone of mutated HIV-1 RT with M03 is presented in Figure 13. The rmsd difference of HIV-1 RT is about 0.40 nm. The result shows that the complex became stable after 700 ps of simulation time. Total energy vs. simulation time is presented in Figure 14. The average energy of complex is about −1,132,000 kcal/mol. The average solvent accessible surface of mutated HIV-1 RT is about 2820 nm^2 (Figure 15). The values for

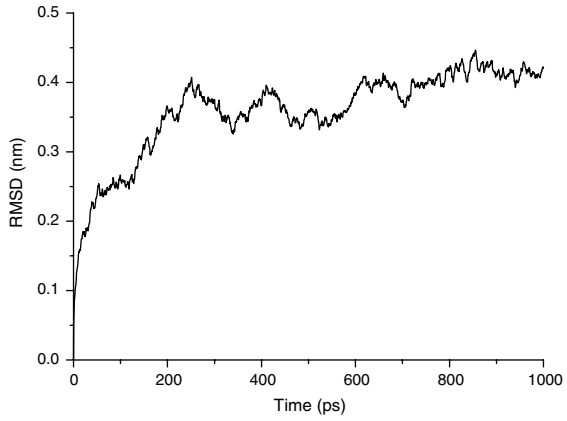

Figure 13. Simulation time vs. rmsd of the backbones of mutated HIV-1 RT.

total energy and solvent accessible surface are slightly lower than those of wild HIV-1 RT. This shows that the complex of M03 with mutated HIV-1 RT is more stable than that with wild HIV-1 RT. The variation of distance for three possible hydrogen bonds with simulation time is shown in Figure 16. The variation of distance O16...HN (Lys101) is not significant. During 1.0 ns simulation, the distance is always kept about 2.0 Å. This suggests that this hydrogen bond is always very strong after the mutation of K103N. The distance of O11...HN (Lys101) is about 3.5 Å. There is no hydrogen bond interaction between these groups, similar to wild HIV-1 RT. The distance O35...HN (Asn103) (full 1.0 ns simulated time) was counted and the average value is about 5.5 Å. This result indicates that the hydrogen bond interaction could be excluded, except in some regions corresponding perhaps to its lowest energy conformations. Compared with the wild HIV-1 RT, the mutation of K103N does not weaken the hydrogen bond between O16 and Lys101. From these results, we could conclude that the DCK analogues could be a potent agent against the drug resistance due to the mutation of K103N.

The principal hydrophobic and hydrogen bonding interactions between M03 and mutated HIV-1 RT are illustrated in Figure 17. We found that O16 and O42 are linked by hydrogen bond to residues Lys101 and Asn103. The length of hydrogen bond is 2.66 Å (O16...N of Lys101) and 2.70 Å (O42...N of Asn103), respectively. Beside these hydrogen bond interactions, there are also hydrophobic effects around the camphanoyl substituent at position 3′ with the residues Pro95, Leu100, Tyr181, Tyr183 and Tyr188, the camphanoyl substituent at position 4′ with the residues Val106, Tyr318 and Pro225. These hydrophobic interactions seem to be favorable to the stabilization of complex.

The energy-minimum conformers for M03-HIV-1 RT and M03-HIV-1 RT(K103N) at simulation time 795 and 984 ps are shown in Figure 18a, b There is only one hydrogen bond between O16 and HN of residue Lys101 (see Figure 18(A)). Two hydrogen bonds were found between M03 and residues Lys101 and Asn103 (see Figure 18(B)). This also shows that K103N mutation might not have influence on the activity of DCK analogues.

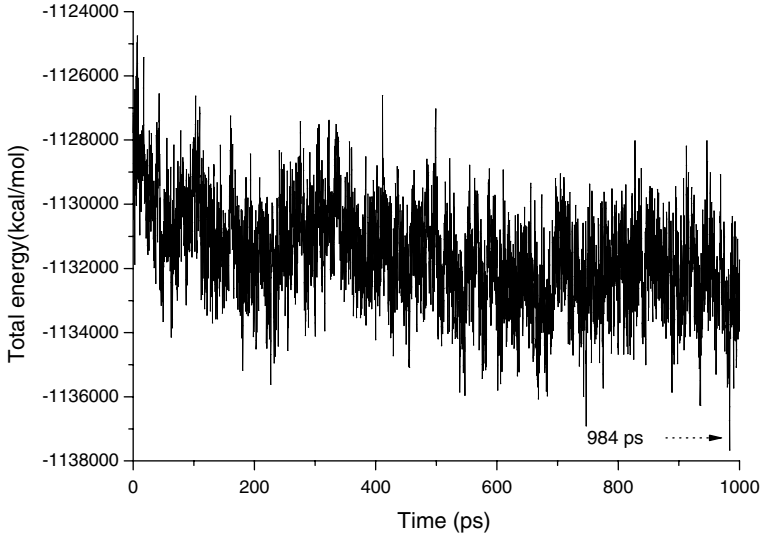

Figure 14. Total energy changes of M03 and HIV-1 RT (K103N) as the function of time. Dotted arrow indicates the local energy-minima conformation.

SVM Model

As mentioned in Introduction section, some DCK analogues exhibit potent activity against HIV-1 RT, but their further development was limited by some drawbacks, such as poor solubility and low bioavailability. In another hand, the results obtained from docking study and molecular dynamics simulation showed that the DCK analogues could act with mutated HIV-1 RT without significant changes in their activity. These results lead us to develop some models in order to design new DCK analogue molecules, which can keep their high inhibitory activity for HIV-1 RT with better solubility and bioavailability. The SVM was shown to be a powerful non-linear method applied to drug design [36]. In this work, the new technique was used to construct QSAR model for DCK analogues. In SVM model, to select the suitable additional descriptors, LOO cross-validation method was used. The parameters ϵ, γ and C were determined as 0.30, 1.70 and 500, respectively using training set. The experimental activities (EA), selected descriptors and predicted activities (PA) are listed in Table 2.

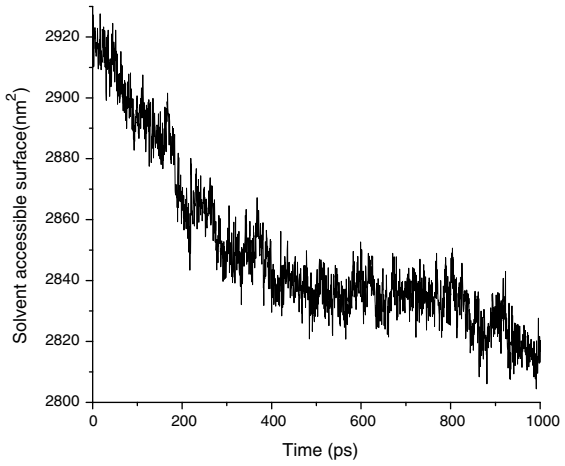

Figure 15. The variation of solvent accessible surface for HIV-1 RT (K103N) during the simulation.

Examination of SVM model

The five descriptors used in SVM model are gathered in Table 2. They are partial positive surface area (PPSA1), hydrophobic regions at sixth energy level (D6), polarizability (POL), HOMO and LUMO. The correlation matrix (Table 3) for these five descriptors shows no high correlations between them. PPSA1 is the parameter-related with the positive charge. It represents the electrostatic interaction between the residues of receptor and ligand. D6 may reflect the hydrophobic property of a molecule. The HOMO and LUMO are localized on aromatic moiety of the

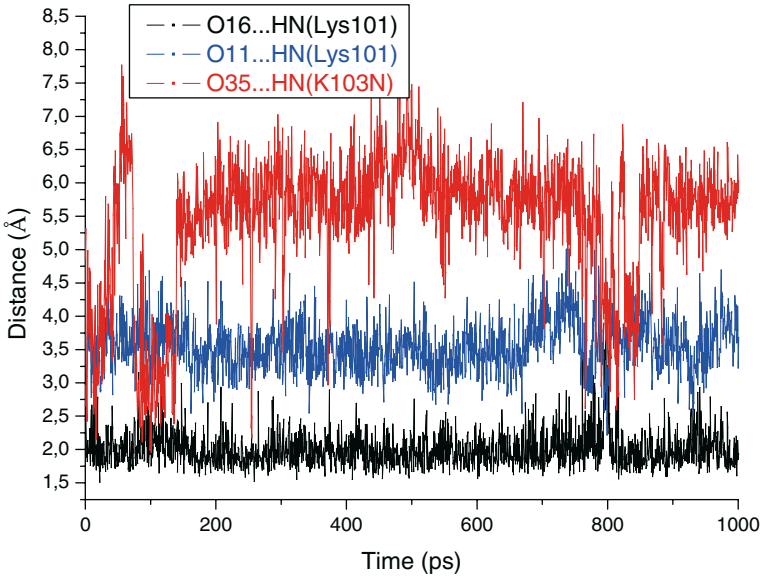

Figure 16. Simulation time vs. distances of atom pair.

compounds. Their energies describe the ability of the aromatic ring to make π–π interactions, respectively, with aromatic or basic residues of receptor [37]. These descriptors could be correlated with the results of molecular dynamics simulation, and selecting them is reasonable. It can be remarked from Table 3 that there are some correlations between the parameters PPSA1, HOMO and LUMO. The distribution of charges and energies of highest occupied and lowest unoccupied orbitals are all electron dependent properties. But their mutual dependence is not very significant. This cannot influence the construction of QSAR model with these parameters. Figure 19 illustrates the correlation between predicted (PA) and experimental (EA) activities. The correlation coefficient r^2 between EA and PA is 0.918 for training set, with standard error (SE) equal to 0.234. This model could well reflect the relations between EA and PA.

Design of new compounds

Since the high hydrophobicity might be one important cause of the failure for DCK analogues in their drug development, the following modification is then focused on improvement of molecular water-solubility, potentially enhancing molecular absorption and oral bioavailability. Guided by the investigation about binding mode and interaction mechanism discussed above, our structural modification will center on the 3,4-substituents. Because the electronegativity at position 3 have potent relationship with anti-HIV activity, we design a series of mono-, di-, and tri-substituted DCK analogues by taking into account of this factor, such as 3-nitromethyl (CH_2NO_2), 3-cyanomethyl (CH_2CN), 3-methylcarbamate (CH_2OCONH_2), 3-hydroxymethyl (CH_2OH) and halides (X). All designed new compounds, generated using previously developed structure generation system [38], were then predicted with SVM models for their activity against HIV replication, as shown in Table 4. The log *P*, HOMO and LUMO values were equally calculated with HYPERCHEM 7.0. From these data, we can see that the modification of substituents on position 3 results in many new compounds with log *P* values ranging from 0.8 to 5.58. Most of these compounds have log *P* values less than 5, which can meet the ''rule of 5' for determining the drug-like compounds [39]. Interestingly, the predicted activities of 3-cyanomethyl-DCK (**4c**) and 3-nitromethyl-DCK (**3c**) reach to 2.93 and 2.34 log unit. They should be very potent against HIV replication. The synthesis and biological evaluation of these compounds may give us more meaningful structural information and further rational design.

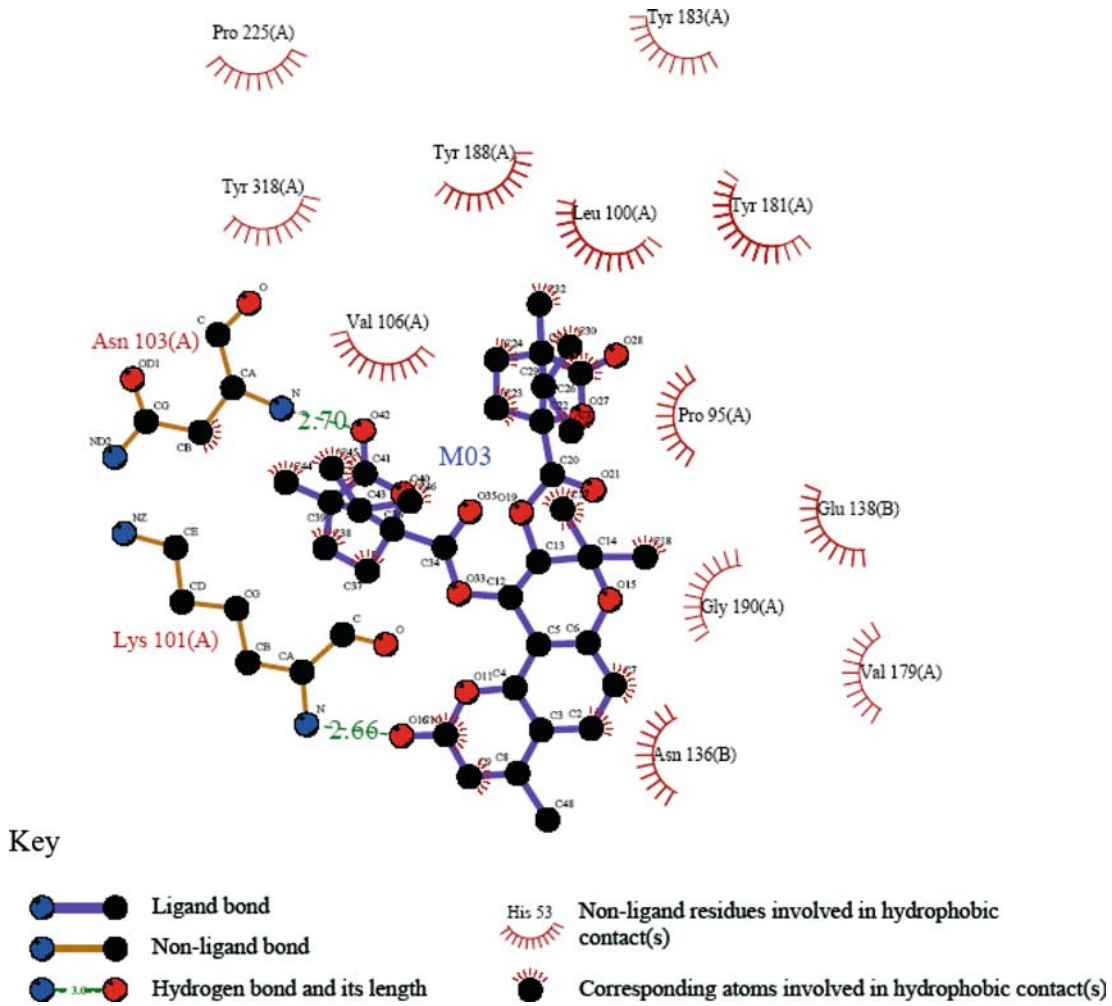

Figure 17. Two-dimensional representation for the interaction mode of compound M03 with mutated HIV-1 RT(K103N), drawn by LIGPLOT program [29].

Synthesis and validation

Fourteen newly synthesized DCKs were tested against HIV-1 replication in acutely infected H9 lymphocytes, to evaluate their anti-HIV activity [40]. AZT was used as control test. Bioassay data are shown in Table 5. Actually, all designed DCKs exhibited inhibitory activities against HIV replication in H9 lymphocytes with EC_{50} values lower than 1.95 μM. The most promising compounds are mono- and di-substituted 3-cyanomethylated-DCK analogues **4c** and **4b** with EC_{50} values of 0.0002 and 0.0024 μM in H9 lymphocytes. These two compounds were both more potent than AZT in the same assay. The mono- and di-substituted 3-nitromethyl-DCKs (**3b** and **3c**) exhibited also potent anti-HIV activities with EC_{50} values of 0.023 and 0.028 μM, respectively. Another interesting compound is 3-fluoromethyl-5-methoxy-4-methyl-DCK (**9a**) with an EC_{50} value of 0.018 μM, much more potent than its calculated activity. In general, bioassay confirmed the anti-HIV activities of novel DCK analogues and showed similar active patterns.

The correlation between predicted and experimental activities for newly synthesized compounds is shown in Figure 20. The correlation coefficient r^2 between EA and PA is 0.690, with standard error (SE) equal to 0.634. The data were classed into three types: low, moderate and high.

We observed a decrease for r^2 with respect to constructed model using training set. It should be

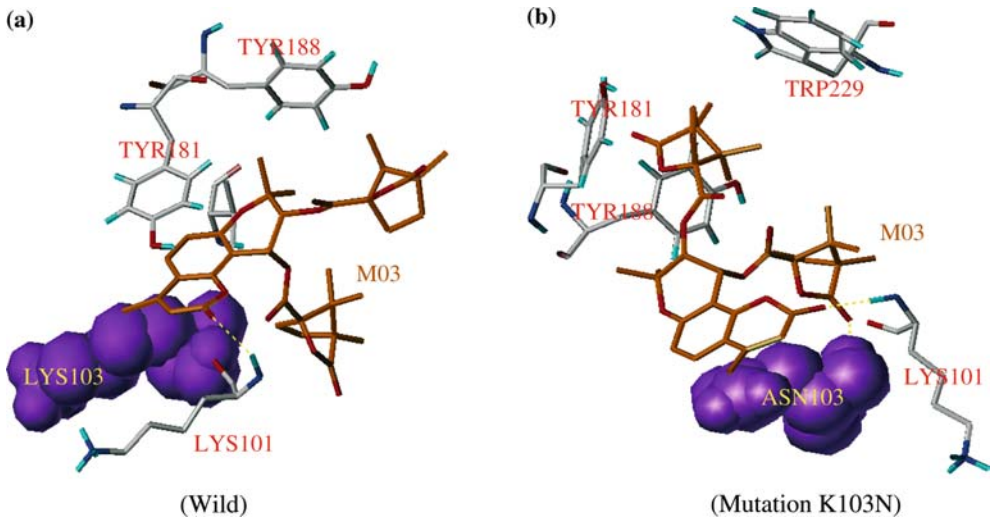

Figure 18. Energy-minimum conformations for M03 with wild and mutated HIV-1 RT (For clear shown the hydrogen bond of mutation K103N, the complex of mutation was rotated for some degrees).

pointed out that we have only 14 newly synthesized compounds. In fact, a number of structures has been virtually generated. But the synthesis is difficult and very cost. In this case, we have to choose the new molecules with the best predicted values for synthesis. Furthermore, our goal is not only to study theoretically the QSAR, we hope also to obtain some molecules which have the potential to be developed as drugs. This diminution in number leads to the decrease of performance. Anyhow, the tendency is good, and the prediction is relatively respected. For example, the

Table 2. The experimental and predicted activity of training set.

No.	pEC_{50}	PPSA1	D6	POL	HM	LM	PA
M01	1.31	650.301	−1.13	−1.662	−9.4102	−0.8685	1.61
M02	1.56	670.653	0.209	−0.896	−9.399	−0.8867	1.77
M03	2.23	659.221	0.638	−0.896	−9.4987	−0.9004	2.09
M04	2.22	659.341	0.799	−0.896	−9.5093	−0.9079	2.20
M05	0.92	673.929	−0.648	−0.129	−9.4073	−1.1006	0.86
M06	1.24	731.349	0.048	−0.129	−9.1141	−0.6099	1.48
M07	1.98	673.225	−0.541	−0.63	−9.0742	−0.8018	1.68
M08	1.29	673.98	−1.077	−0.63	−9.459	−0.8519	1.25
M09	1.36	704.464	−0.916	−0.63	−9.3657	−0.713	1.18
M10	3.12	650.464	1.924	−0.091	−9.5507	−1.1224	2.82
M11	1.23	638.118	0.102	0.201	−9.5563	−1.0942	1.22
M12	1.17	730.236	1.227	0.938	−9.5606	−1.0147	1.47
M13	1.46	671.047	−0.434	−0.63	−9.4097	−0.886	1.38
M14	0.17	731.794	−1.023	−0.332	−9.5905	−1.0507	0.47
M15	−0.57	842.811	−1.238	2.733	−8.6964	−0.7195	−0.27
M16	1.55	744.91	1.12	1.705	−9.5449	−1.0324	1.25
M17	0.95	696.383	−0.97	0.136	−9.6363	−1.0515	0.65
M18	1.79	697.088	1.12	0.136	−9.5191	−0.9856	1.77
M19	0.02	650.565	−0.005	1.297	−9.6006	−1.2444	0.32
M20	1.61	624.975	−0.862	0.201	−9.754	−1.1825	1.31
M21	2.13	633.894	1.656	0.201	−9.5224	−1.1201	2.43

Table 3. The analysis of correlation between descriptors (r).

	PPSA1	D6	POL	HOMO	LUMO
PPSA1	1.0000	−0.2078	0.6155	0.6385	0.4856
D6	−0.2078	1.0000	0.1193	−0.3043	−0.3174
POL	0.6155	0.1193	1.0000	0.2494	−0.2040
HOMO	0.6385	−0.3043	0.2494	1.0000	0.7511
LUMO	0.4856	−0.3174	−0.2040	0.7511	1.0000

compounds **3c**, **4b** and **4c** are the most active according to the predicted values using our model. The biological assay shows that these three compounds possess good activities. Except **3c** whose value is little far from expected one, the **4b** and **4c** are in agreement with the predicted values. Moreover, the **4c** is the best one, in agreement with calculated results. Current bioassay data verified the reliability of the SVM model described above. Most experimental biological data matched well with predictive potency, which validated the SVM model for DCK analogues.

Conclusion

Molecular docking and molecular dynamics simulation were applied to study the binding mode between DCK analogues and HIV-1 RT. MD simulation suggests that the hydrogen bond between O16 and NH of Lys101 is very strong. While the hydrogen bond between O11 and NH of Lys101 only exists during a few ps simulation time (full 1.0 ns simulation time). This study illustrates that DCK analogues might have similar interaction mechanism to other type of HIV-1 RT inhibitors, such as HEPT, TIBO, etc. Lipophilic and electrostatic potential analysis shows us that the substituents at positions 3 and 4 of khellactone are near the hydrophilic and polar residues of

Table 4. The substituent and prediction activity of new compounds.

Code	Substituent	PA	log *P*
1a	3-Me-4-Me-5-OMe-DCK	0.61	5.12
2a	3-CH_2Br-4-Me-5-OMe-DCK	0.23	5.32
3a	3-CH_2NO_2-4-Me-5-OMe-DCK	0.15	0.80
4a	3-CH_2CN-4-Me-5-OMe-DCK	1.52	4.94
5a	3-CH_2OAc-4-Me-5-OMe-DCK	0.09	4.24
7b	3-CH_2OCONH_2-4-Me-DCK	1.65	3.95
7c	3-CH_2OCONH_2-DCK	0.81	3.80
9a	3-F-4-Me-5-OMe-DCK	1.02	4.36
8a	3-Cl-4-Me-5-OMe-DCK	0.60	4.74
6a	3-CH_2OH-4-Me-5-OMe-DCK	0.98	4.11
4b	3-CH_2CN-4-Me-DCK	2.34	5.19
3b	3-CH_2NO_2-4-Me-DCK	1.29	1.06
3c	3-CH_2NO_2-DCK	2.34	0.90
4c	3-CH_2CN-DCK	2.93	5.04

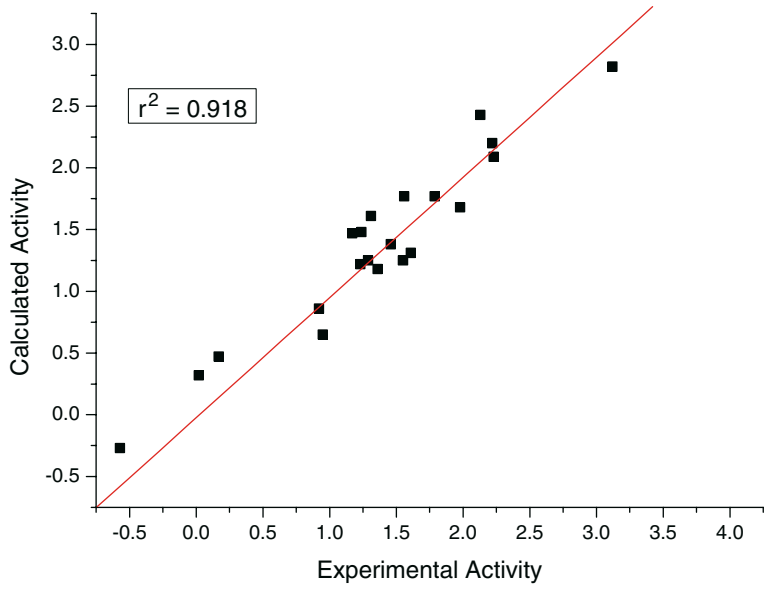

Figure 19. Correlations between experimental and predicted activity.

Table 5. Anti-HIV activity of new DCK analogues in the H9 cell lines.

Code	EC_{50} (μm)[a]	pEC_{50}
1a	1.95	−0.29
2a	1.62	−0.21
3a	0.27	0.57
4a	0.11	0.96
5a	1.50	−0.18
7b	0.353	0.45
7c	0.0777	1.11
9a	0.018	1.74
8a	1.80	−0.26
6a	0.0522	1.28
3b	0.0230	1.64
3c	0.0288	1.54
4b	0.0024	2.62
4c	0.0002	3.70
AZT	0.0062	2.21

[a]Concentration that inhibits replication of virus by 50%. Assays in H9 lymphocytes were performed by Panacos Pharmaceuticals, Inc., Gaithersburg, MD, USA.

Glu138, Asn136 (B chain) and Lys101 (A chain). This suggests that hydrophilic and electronegative substituents at positions 3 and 4 are favorable to anti-HIV activity.

Because the efficiency of NNRTIs is seriously compromised by the emergence of mutant viral strains, the influence of drug resistance of mutated HIV-1 RT (K103N) on DCK analogues' activity was investigated. The results from docking and MD simulation show that DCK analogues could also inhibit the mutated HIV-1 RT (K103N).

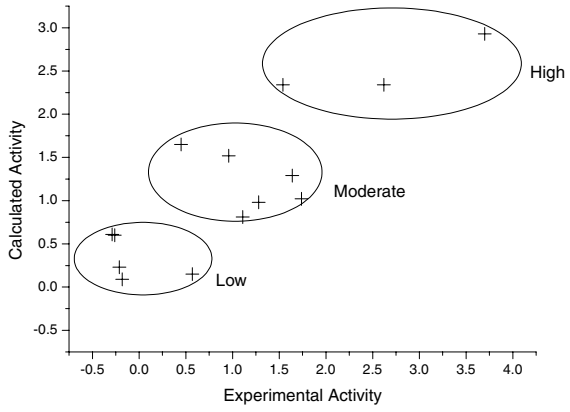

Figure 20. Correlations between experimental and predicted activity of SVM Model for newly compounds.

Based on these investigations, a set of compounds with modification of position 3 and or position 4 substituents was designed. SVM was used to construct the relationship model between DCK structural descriptors and anti-HIV activity. This model is stable and robust, and it can be used to predict anti-HIV activity. Fourteen newly designed compounds by means of SVM model were synthesized, and most of them have potent anti-HIV activity. Their further evaluations are now being processed.

Acknowledgements

The authors thank Prof. Arthur J. Olson for his kindness in offering us the AutoDock 3.0.3 program. We gratefully acknowledge financial support from the Chinese Academy of Sciences – National Center of Scientific Research in France Cooperation Program (CNRS/CAS No. 14916) and Embassy of France in China. The authors also thank Dr. Chih-Jen Lin in offering us the LibSVM software in the present study.

References

1. Hopkins, A.L., Ren, J.S., Esnouf, R.M., Willcox, B.E., Jones, E.Y., Ross, C., Miyasaka, T., Walker, R.T., Tanaka, H., Stammers, D.K. and Stuart, D.I., J. Med. Chem., 39 (1996) 1589.
2. Mitsuya, H., Weinhold, K.J., Furman, P.A., St. Clair, M.H., Nusinoff, L.S., Gallo Bolognesi, R.C. D., Barry, D.W. and Broder, S., Proc. Natl. Acad. Sci. USA, 82 (1985) 7096.
3. Mitsuya, H. and Broder, S., Proc. Natl. Acad. Sci. USA, 83 (1986) 1911.
4. Yarchoan, R., Mitsuya, H., Thomas, R.V., Pluda, J.M., Hartman, N.R., Perno, C.F., Marczyk, K.S., Allain, J.P., Johns, D.G. and Broder, S., Science, 245 (1989) 412.
5. Coates, J.A.V., Cammack, N., Jenkinson, H.J., Mutton, I.M., Pearson, B.A., Storer, R., Cameron, J.M. and Penn, C.R., Antimicrob. Agents Chemother., 36 (1992) 202.
6. De Clercq, E., Med. Res. Rev., 13 (1993) 229.
7. Huang, L., Kashiwada, Y., Cosentino, L.M., Fan, S. and Lee, K.H., Bioorg. Med. Chem. Lett., 4 (1994) 593.
8. Kireev, D.B., Chretien, J.R., Grierson, D.S. and Monneret, C., J. Med. Chem., 40 (1997) 4257.
9. HeravI, M.J. and Parastar, F., J. Chem. Inf. Comput. Sci., 40 (2000) 147.
10. Rizzo, R., Rives, J.T. and Jorgenson, W.L., J. Med. Chem., 44 (2001) 145.
11. Eriksson, M.A.L., Pitera, J. and Kollman, P.A., J. Med. Chem., 42 (1999) 868.
12. Huuskonen, J., J. Chem. Inf. Comput. Sci., 41 (2001) 425.

13. Vapnik, V., Statistical Learning Theory. Wiley, New York, 1998.
14. Cristianini, N. and Shawe-Taylor, J., An Introduction to Support Vector Machines. Cambridge University Press, Cambridge, UK, 2000.
15. Joachims, T., Learning to Classify Text Using Support Vector Machines: Methods, Theory, and Algorithms. Kluwer, 2002.
16. Schölkopf, B. and Smola, A., Learning with Kernels. MIT Press, Cambridge, MA, 2002.
17. Herbrich, R., Learning Kernel Classifiers. MIT Press, Cambridge, MA, 2002.
18. Hsu, C.W. and Lin, C.J., IEEE Trans. Neural Networks., 13 (2002) 415.
19. SYBYL. Tripos Associates Inc,Version 6.9, St. Louis, MO.
20. Clark, M., Cramer, III R.D. and Opdenbosh, N.Van , J. Comput. Chem., 10 (1989) 982.
21. Weiner, S.J.P., Kollman, A., Case, D.A., Singh, U.C., Ghio, C., Alagona, G., Profeta, S.J. and Weiner, P., J. Am. Chem. Soc., 106 (1984) 765.
22. Morris, G.M., Goodsell, D.S., Halliday, R.S., Huey, R., Hart, W.E., Belew, R.K. and Olson, A.J., J. Comput. Chem., 19 (1998) 1639.
23. Morris, G.M., Goodsell, D.S., Huey, R. and Olson, A.J., J. Comput. – Aided Mol. Des., 10 (1996) 293.
24. GROMACS, version 1.6; BIOSON Research Institute, Department of Biophysical Chemistry, University of Groningen, The Netherlands.
25. Berendsen H., van der Spoel D. and van Drunen R., (1995) Comput. Phys. Commun. 91 43.
26. Stanton, D.T. and Jurs, P.C., Anal. Chem., 62 (1990) 2323.
27. Clementi, S., Cruciani, G., Fifi, P., Riganelli, D., Valigi, R. and Musumarra, G., Quant. Struct. – Act. Relat., 15 (1995) 108.
28. Eisenberg, D. and McLachlan, A.D., Nature, 319 (1986) 199.
29. Wallace, A.C., Laskowski, R.A. and Thornton, J.M., Protein Eng., 8 (1995) 127.
30. Chen, H.F., Yao, X.J., Li, Q., Yuan, S.G., Panaye, A., Doucet, J.P. and Fan, B.T., SAR and QSAR Environ. Res., 14 (2003) 455.
31. Viswanadhan, V.N., Ghose, A.K., Revankar, G.R. and Robins, R.K., J. Chem. Inf. Comput. Sci., 29 (1989) 163.
32. Hansch, C., Leo, A. and Taft, R.W., Chem. Rev., 91 (1991) 165.
33. Clercq, E.De , Biochim. Biophys. Acta, 1587 (2002) 258.
34. Bacheler, L.T., Drug Resist. Updates, 2 (1999) 56.
35. Blagović, M. U., Rives, J. T. and Jorgensen, W. L., J. Med. Chem., 47 (2004) 2389.
36. Chen, H.F., Yao, X.J., Petitjean, M., Xia, H.R., Yao, J.H., Panaye, A., Doucet, J.P. and Fan, B.T., QSAR Comb. Sci., 23 (2004) 603.
37. Macchiarulo, A., De Luca, L., Costantino, G., Barreca, M.L., Gitto, R., Pellicciari, R. and Chimirri, A., J. Med. Chem., 47 (2004) 1860.
38. Chen, H.F., Dong, X.C., Zeng, B.S., Gao, K., Yuan, S.G., Fan, B.T., Zheng, C.Z., Panaye, A. and Doucet, J.P., SAR and QSAR Environ. Res., 14 (2003) 251.
39. Lipinski, C.A., Lombardo, F., Dominy, B.W. and Feeney, P.J., Adv. Drug Del. Rev., 23 (1997) 3.
40. Xie, L., Zhao, C.H., Zhou, T., Chen, H.F., Fan, B.T., Ma, J.Z., Li, J.Y., Bao, Z.Y., Lee, K.H., Allaway, G. and Lo, Z.W. Biorg. Med. Chem., (2005) in press.