Journal of Biomolecular Structure and Dynamics

ISSN: 0739-1102 (Print) 1538-0254 (Online) Journal homepage: https://www.tandfonline.com/loi/tbsd20

3D-QSAR, molecular docking, and molecular dynamics simulation of a novel thieno[3,4-d]pyrimidine inhibitor targeting human immunodeficiency virus type 1 reverse transcriptase

Han Chu, Qing-xiu He, Jun-wei Wang, Ya-ting Deng, Juan Wang, Yong Hu, Yuan-qiang Wang & Zhi-hua Lin

To cite this article: Han Chu, Qing-xiu He, Jun-wei Wang, Ya-ting Deng, Juan Wang, Yong Hu, Yuan-qiang Wang & Zhi-hua Lin (2020) 3D-QSAR, molecular docking, and molecular dynamics simulation of a novel thieno[3,4-d]pyrimidine inhibitor targeting human immunodeficiency virus type 1 reverse transcriptase, Journal of Biomolecular Structure and Dynamics, 38:15, 4567-4578, DOI: 10.1080/07391102.2019.1697366

To link to this article: https://doi.org/10.1080/07391102.2019.1697366

JOURNAL OF BIOMOLECULAR STRUCTURE AND DYNAMICS
2020, VOL. 38, NO. 15, 4567–4578
https://doi.org/10.1080/07391102.2019.1697366

3D-QSAR, molecular docking, and molecular dynamics simulation of a novel thieno[3,4-d]pyrimidine inhibitor targeting human immunodeficiency virus type 1 reverse transcriptase

Han Chu[a,b], Qing-xiu He[a,b], Jun-wei Wang[a,b], Ya-ting Deng[a,b], Juan Wang[a,b], Yong Hu[a,b], Yuan-qiang Wang[a,b] and Zhi-hua Lin[a,b,c]

[a]Department of Pharmacy and Bioengineering, Chongqing University of Technology, Chongqing, P. R. China; [b]Key Laboratory of Screening and Activity Evaluation of Targeted Drugs, Chongqing, P. R. China; [c]School of Chemistry and Chemical Engineering, Chongqing University, Chongqing, P. R. China

Communicated by Ramaswamy H. Sarma

ABSTRACT

Human immunodeficiency virus type 1 reverse transcriptase (HIV-1 RT) is one of the most attractive drug targets for the treatment of AIDS. In this study, 67 thieno[3,4-d]pyrimidine derivatives were selected as novel HIV-1 RT inhibitors to combat viral resistance, and were subjected to 3 D-QSAR studies using CoMFA, CoMSIA, and T-CoMFA. In the 3 D-QSAR study, two methods of ligand-based alignment and pharmacophore-based alignment were used. The results showed that CoMFA ($n = 8$; $q^2 = 0.594$; $r^2 = 0.974$) and CoMSIA ($n = 7$; $q^2 = 0.528$; $r^2 = 0.965$) have good stability and predictability. The molecular docking study showed that the hydrogen bonding and van der Waals interactions of key residues such as Leu100, Lys101, Val106, Phe227 and Pro236 play an important role in ligand-receptor binding. Based on these results, 12 new thieno[3,4-d]pyrimidines were designed and their activities were predicted; the results indicated that these compounds have good predictive activity and reasonably good ADME/T profiles. MD simulation analysis of 50 ns showed that compound 23j formed four hydrogen bonds with the residues (Lys101, Lys104, Val106 and Thr318), and binds more closely to HIV-1 RT than compound 23j. Furthermore, the group at the R_1 position and the horseshoe-like conformation of these compounds are critical for the inhibitory activity and stability. These results provide useful insights for the discovery and design of a new generation of HIV-1 RT inhibitors.

ARTICLE HISTORY
Received 10 May 2019
Accepted 22 October 2019

KEYWORDS
Novel HIV-1 RT inhibitors; thieno[3,4-d]pyrimidine derivatives; 3D-QSAR; molecular docking; molecular dynamics simulation

1. Introduction

Acquired immunodeficiency syndrome (AIDS) caused by human immunodeficiency virus type 1 (HIV-1) has been one of the major problems endangering human life since it was reported in 1981 (Hymes et al., 1981). HIV-1 reverse transcriptase (HIV-1 RT), one of the enzymes necessary for HIV replication, can catalyze the conversion of single-stranded RNA into double-stranded pre-HIV DNA (Anthony, 2004; Sarafianos et al., 2009), and is one of the important targets for the development of anti-HIV-1 drugs (Tarasova, Poroikov, & Veselovsky, 2018). The non-nucleoside reverse transcriptase inhibitors (NNRTIs) targeting HIV-1 RT have good antiviral activity, selectivity, and pharmacokinetic properties (Esté & Cihlar, 2010; Shattock, Warren, McCormack, & Hankins, 2011), and are of great significance in the combination therapy of HIV infection and AIDS (Martins, Ramos, & Fernandes, 2008). At present, the first generation of NNRTIs used in the clinic mainly includes Efavirenz, Nevirapine, and Delavirdine (Kertesz et al., 2010), and these compounds act in the vicinity of the HIV-1 RT catalytic region, which alters the protein geometry, eventually leading to enzyme failure (Buzón et al., 2010). However, evidence suggests that HIV-1 has become resistant to these drugs, and that the first generation NNRTIs are not effective for patients with mutant HIV-1, and so, the new generation of NNRTIs needs to break through in terms of drug resistance (Mayer & Venkatesh, 2010; Zhan, Pannecouque, De Clercq, & Liu, 2016). Currently, common mutations in HIV-1 RT include K103N, Y108C, and Y188L (de Béthune, 2010; Johnson et al., 2011). The new generation of NNRTI designs can be considered from some aspects such as: a) enhancing compound conformational flexibility and positional adaptability (Das, Lewi, Hughes, & Arnold, 2005); b) boarding backbone hydrogen bonding (Ghosh, Chapsal, Weber, & Mitsuya, 2008); c) targeting highly conserved residues in the target (Zhan, Liu, Li, Pannecouque, & De Clercq, 2009). In-depth study of new compounds of this type of mutant HIV-1 RT is imminent.

In recent years, the thieno[3,4-d]pyrimidine derivatives reported by Kang et al. showed higher inhibitory activity and better performance in drug resistance than the older generation NNRTIs, and show great potential for further research. In this study, 67 compounds were selected and the relationship between the molecular structure of thieno[3,4-d]pyrimidine

CONTACT Yuan-qiang Wang wangyqnn@cqut.edu.cn; Zhi-hua Lin zhlin@cqut.edu.cn Department of Pharmacy and Bioengineering, Chongqing University of Technology, Chongqing, 400050 P. R. China

Supplemental data for this article is available online at https://doi.org/10.1080/07391102.2019.1697366.

derivatives and its inhibitory activity (EC_{50}) was studied by comparative molecular force field analysis (CoMFA) (Cramer, Patterson, & Bunce, 1988), comparative molecular similarity index analysis (CoMSIA) (Klebe, Abraham, & Mietzner, 1994), and Topomer CoMFA (T-CoMFA) (Cramer, 2003). Then, by analyzing the contour maps of CoMFA and CoMSIA, 12 new compounds were designed, and the activity of these new compounds was predicted by the reliable model obtained above. In addition, this study also evaluates the binding mode of compounds to proteins using molecular docking and molecular dynamics simulation. This study can avoid the blindness of drug design and provide theoretical and practical value for the design, discovery, and synthesis of new and efficient NNRTIs.

2. Materials and methods

2.1. Preparation of data set and structural optimization

The 67 thieno[3,4-d]pyrimidines were collected from literature (Kang et al., 2017; Kang et al., 2016), and their structures and activity data are shown in Table S1. The bioactivity values of these compounds *in vitro* were reported as EC_{50}, which were converted to the corresponding pEC_{50} ($pEC_{50} = \log EC_{50}$). The compounds were randomly divided into two sets considering both the distribution of bioactivity values and structural diversities. A training set of 54 compounds (80%) was used to construct the models, whereas the remaining 8 compounds (20%) were added to the test set to evaluate the models.

The 3 D-molecular models of these compounds were sketched using SYBYL-X 2.1 software and minimized using the Tripos molecular mechanics force field in the Powell method with a convergence criterion of 0.005 kcal• (mol • Å)$^{-1}$ and 10000 maximum iterations. Partial atomic charges were assigned by the Gasteiger-Huckel method. Other parameters were set to the default value. The minimized structures were prepared for 3 D-QSAR studies as the primal conformations.

2.2. CoMFA and CoMSIA analysis

In SYBYL-X 2.1, the field properties of the models were calculated using 3 D cubic lattices with a spaced grid of 2.0 Å. For the CoMFA method, it was assumed that ligand–receptor binding was non-covalent. An sp^3 hybridized carbon atom with +1 charge as the probe was used to detect the size and distribution of the steric and electrostatic fields (Cheng, Wang, Yu, Li, & Huang, 2018). In the CoMSIA model, besides the steric and electrostatic field, the hydrophobic, H-bond acceptor and donor fields were also introduced. Further, similarity indices were calculated between a probe and each atom of the molecule based on a Gaussian distance function in the CoMSIA method.

Molecular alignment is one of the important parameters for generating QSAR models. The statistical values of CoMFA and CoMSIA depend on the molecular alignment methods. Therefore, this study used ligand-based alignment (Alignment I) and pharmacophore-based alignment (Alignment II) to find the best 3 D-QSAR model. Alignment I is shown in Figure 1(A). A common skeleton thienopyrimidine ring (as shown in the thick part) was identified in all 67 molecules, and the most active compound 23 was used as the template molecule, and all other molecules were aligned using the Align Database module of SYBYL-X 2.1 (Patel, Patel, Kaushik-Basu, & Talele, 2008), as shown in Figure 1(C). Alignment II is shown in Figure 1(D). The pharmacophore models were generated using the Genetic Algorithm with Linear Assignment of Hypermolecular Alignment of Datasets (GALAHAD) module of SYBYL-X 2.1 (Richmond et al., 2006). Fifty-four compounds were selected to construct pharmacophore models in consideration of their relatively high activity and diverse structures. The Population Size and the Max Generations was set according to an automatically performed virtual suggestion based on the experimental activity data. The model with the highest SPECIFICITY, N_HITS, STERICS, MOL_QRY, HBOND, and the lowest ENERGY value was applied to compound alignments. The CoMFA and CoMSIA models established by the above two alignment methods will then be analyzed further.

2.3. Topomer CoMFA analysis

Topomer CoMFA was the second-generation CoMFA method. Unlike the traditional CoMFA method, Topomer CoMFA automated the generation of models, which enhanced the reproducibility of CoMFA and greatly improved the computational efficiency. The steric and electrostatic fields were calculated using the carbon sp^3 probe and partial least-squares (PLS) regression was used for generating the Topomer CoMFA model. In this study, each molecule was broken into three fragments and the segmentation is shown in Figure 1(C).

2.4. Partial least square (PLS) analysis and model validation

Partial least squares (PLS) method, an extension of multiple regression analysis, was applied to construct the relationship between the structure and biological activities using SAMPLS. The optimum number of components (n) and the highest cross-validation correlation coefficient (q^2) for the correlation models were calculated by the leave-one-out (LOO) cross-validation procedure (Bush & Nachbar, 1993). The conventional multiple correlation coefficients (r^2), standard error of estimate (*SEE*), and the Fisher test (F) values was obtained by non-cross-validated analysis (Li, Wei, Zhao, & Du, 2014).

The values of q^2 and r^2 determined by Partial least square analyses, evaluated the internal predictive ability and robustness of the models, whereas the external validation method as the most valuable validation method, was employed to estimate the external prediction capability of QSAR models (Li et al., 2017). Biological activities of compounds in the test set were predicted by the 3 D-QSAR models derived from the training set.

2.5. Molecular docking

Molecular docking is one of the drug design methods used to predict the optimal binding mode between a small molecule ligand and a receptor protein. In this study, molecular

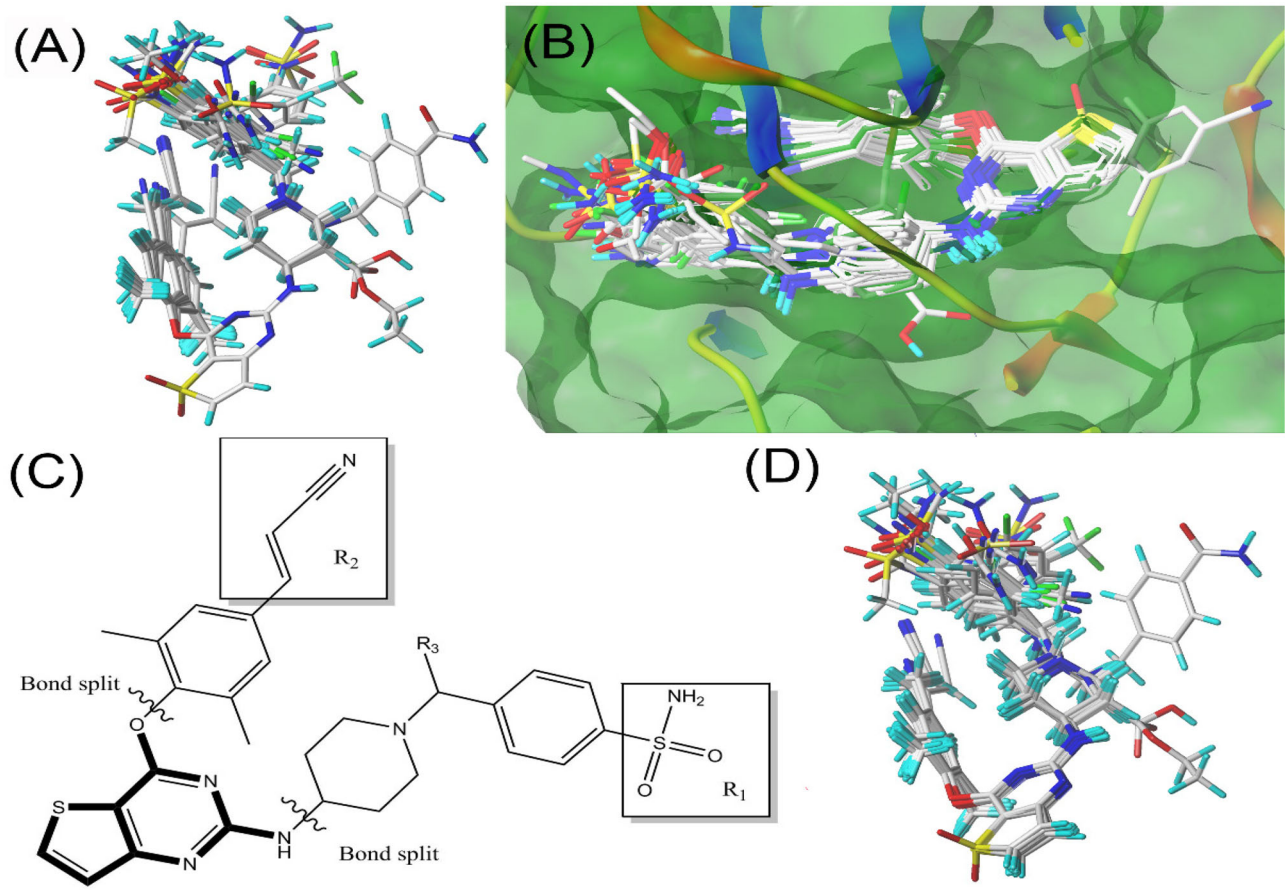

Figure 1. A: Alignment I of all compounds; B: The docking conformation of HIV-1 RT protein with all compounds; C: The Structure of compound 23; D: Alignment II of all compounds.

docking was performed using the SurFlex-Dock module of Sybyl, and the crystal structure of HIV-1 RT (PDB: 3M8Q) was obtained from Protein Data Bank (http://www.rcsb.org/pdb/). The binding pocket was generated using the Protomol generation technique of SYBYL. The crystal structure of HIV-1 RT was hydrogenated and charged using SYBYL software. At the same time, the original ligand molecules (DJZ562) in the crystal structure were removed for docking calculations. The docking pocket was formed with amino acid residues at 0.5 Å in the range of the original ligand molecule, and the remaining parameters were default.

The cognate ligand DJZ562 (Figure 2(A)) was redocked into the binding pocket to verify the reliability of the docking method. After that, 54 compounds were docked into the binding pocket using the same parameters. Taking both the total score and CScore into consideration, the best conformation of each molecule was selected for further analysis. The binding visualization was performed using PyMOL software.

2.6. Molecular dynamics (MD) simulation and free energy analysis

The MD simulations were carried out using the AMBER16 package. Before simulation, the molecular mechanics method is usually adopted to optimize the biological macromolecular system. The general AMBER force field (GAFF) was used for the inhibitors, and the ff99SB force field was employed for the protein (Berendsen, Postma, van Gunsteren, DiNola, & Haak, 1984).

All protein inhibitor complex systems were immersed in a box of the TIP3P water model (Jorgensen, Chandrasekhar, Madura, Impey, & Klein, 1983). The systems were then neutralized by the addition of Na^+ or Cl^- counter ions. First, several minimization steps were performed for the systems to avoid possible steric crashes. Then, each system was gradually heated from 0 K to 300 K during the heating stage and kept at 300 K during the following equilibrium and production stages. A time step of 2 fs was used for the heating stage, equilibrium stage, and the entire production stage. A periodic boundary condition was employed to maintain constant temperature and the pressure (NPT) ensembles. The pressure was set at 1 atm and was controlled by an anisotropic (x-, y-, z-) pressure scaling protocol with a pressure relaxation time of 1 ps. The temperature was regulated using Langevin dynamics with a collision frequency of 2 ps^{-1}. The Particle Mesh Ewald (PME) method was adopted to handle long-range electrostatics and a 1 nm cutoff was set to treat real-space interactions. All covalent bonds involving hydrogen atoms were constrained using the SHAKE algorithm. Each system underwent 50 ns MD simulation and the trajectory of simulated systems was saved every 100 ps.

The binding free energy is an important indicator reflecting the degree of binding compactness. The positive and negative binding free energies reflect the possibility of ligand

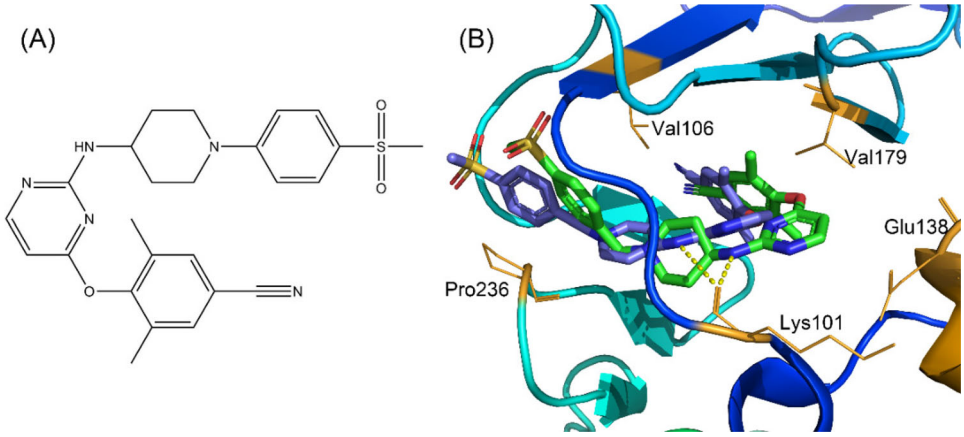

Figure 2. A: The structure of DJZ562. B: The re-docked result of the HIV-1RT co-crystal ligand DJZ562 in its binding site. The co-crystal conformation is shown as green sticks and the re-docked conformation is shown as blue sticks. The hydrogen-bond interactions are illuminated by dashed lines.

binding to the receptor. In this study, MM/PBSA (Molecular mechanics/Poisson Boltzmann Surface Area) (Kollman et al., 2000) and MM/GBSA (Molecular Mechanics/Generalized Born Surface Area) (Homeyer & Gohlke, 2012) were used to calculate the binding free energy for the saved trajectories of MD simulations. The corresponding formulas were as follows:

$$\begin{aligned}\Delta G_{bind} &= G_{complex} - G_{protein} - G_{ligand} \\ &= \Delta H - T\Delta S \approx \Delta G_{gas} + \Delta G_{sol} - T\Delta S \\ \Delta G_{gas} &= \Delta E_{ele} + \Delta G_{vdw};\ \Delta G_{sol} = \Delta G_{PB/GB} + \Delta G_{SA}\end{aligned}$$

In the formulas, ΔG_{bind} was the final binding free energy, and $G_{complex}$, $G_{protein}$, and G_{ligand} were the free energies of the complex, HIV-1 RT protein, and ligand, respectively. ΔG_{gas} was the gas-phase interaction energy between the protein and the ligand, which consisted of ΔE_{ele} (electrostatic energy) and ΔE_{vdw} (van der Waals energy). ΔG_{sol} was the solvation free energy, which was the sum of the electrostatic solvation energy $\Delta G_{PB/GB}$ (polar contribution) and non-electrostatic solvation energy ΔG_{SA} (nonpolar contribution). In the actual calculation, conformational entropy ($T\Delta S$) was time-consuming and had little influence on the total free energy; hence, $T\Delta S$ could be ignored. In this study, the molecular dynamics simulation was carried out by selecting compound 23 with the highest activity and compound 23j with the highest predicted activity. First, this could verify the results of 3 D-QSAR and molecular docking. Second, this further revealed the binding mode of this series of compounds to HIV-1 RT. Simultaneously, this could be found in key amino acids, suggesting a rational design for a new generation of HIV-1 RT inhibitors.

2.7. ADME/T properties prediction and synthetic accessibility prediction

In this study, the ADME/T properties of the designed inhibitor for HIV-1 RT were predicted using the pkCSM online server (http://biosig.unimelb.edu.au/pkcsm/prediction). For a given compound, the webserver predicted the percentage that will be absorbed through the human intestine. A molecule with an absorbance of less than 30% was considered poorly absorbed. The higher the VD was, the more of a drug was distributed in tissues rather than plasma. The VDss was considered low if below 0.71 L•kg^{-1} (logVDss < -0.15) and high if above 2.81 L•kg^{-1} (logVDss > 0.45). The predictors evaluated the designed molecules to determine whether it was likely to be metabolized by either P450 or would be a cytochrome P450 inhibitor. The dosing rates to achieve steady-state concentrations were determined by Total Clearance (log mL•(min•kg)$^{-1}$) which was related to bioavailability. There was likely to be a toxic character for the designed novel molecules assessed by the prediction of AMES toxicity/Hepatotoxicity/Skin Sensitization.

The synthetic accessibility of the novel derivatives was evaluated by applying the online tool SwissADME (http://www.swissadme.ch/index.php) whose scale revealed that a score closer to 10 had a high structural complexity, and was therefore difficult to synthesize, whereas a score of 1 indicated a relatively easy synthetic route.

3. Results and analysis

3.1. 3 D-QSAR model analysis

Using 54 thienopyrimidines derivatives as the training set, the results of modeling of CoMFA, CoMSIA, and T-CoMFA are shown in Table 1 (31 possible combinations of COMSIA molecular fields are shown in Table S2).

In general, for internal validation, the robust and reliable ability of the CoMFA/CoMSIA model should meet $q^2 > 0.5$, $r^2 > 0.9$. The Topomer CoMFA model satisfies $q^2 > 0.2$ for internal validation. The favorable model should have a low value of *SEE* and high values of q^2, r^2 and *F*. As can be seen from Table 1, the statistical results of the CoMFA/CoMSIA model using the Alignment I method, which led to the lowest *SEE* and the highest q^2, r^2 and *F*, were superior to those of the Alignment II. Hence, the Alignment I method was more conducive to further 3 D-QSAR research. In Alignment II, the pharmacophores do not stack well due to the rigid alignment of the GALAHAD. Therefore, the established 3 D-QSAR model is not good. Comprehensive analysis of all model parameters in the table showed that Alignment I has an excellent effect. Therefore, the CoMFA and CoMSIA models established by Alignment I were selected as the superior

Table 1. Summary of 3 D-QSAR results.

	Alignment I			Alignment II	
PLS statistics	CoMFA	CoMSIA	T-CoMFA	CoMFA	CoMSIA
n	8	7	15	7	10
q^2	0.594	0.528	0.692	0.489	0.520
r^2	0.974	0.965	0.989	0.939	0.981
SEE	0.139	0.157		0.208	0.121
F	207	184		102	221
Field Contribution[a]					
S	0.444	0.044			0.041
E	0.556	0.266			0.279
H		0.131			0.160
D		0.266			0.228
A		0.292			0.292

[a]Field Contribution: CoMFA and CoMSIA with different field contribution such as S (steric); E (electrostatic); H (hydro-phobic); D (H-bond donor); A (H-bond acceptor).

3 D-QSAR model to guide the modification of thienopyrimidine HIV-1 RT inhibitors. The linear correlation analysis between the training set and the test set of the 3 D-QSAR model is shown in Figure 3. It can be seen from the figure that most of the compounds in the training set and test set of CoMFA and CoMSIA are located near or close to the trend line, which proves that the actual activity value of the compound fits well with the predicted activity value. 3 D contour map analysis was performed using compound 23 with the highest actual activity as a template. The contour map of the electrostatic field and the steric field of CoMFA is shown in Figure 4(A,B). CoMSIA-SEHDA only shows contour maps of the electrostatic field, hydrogen bond donor field, and hydrogen bond acceptor field that contributes a large amount. Figure 4(A) is an electrostatic field contour map of CoMFA. The blue region indicates that increasing the positive charge here is beneficial to increase the activity of the compounds, and the red region indicates that increasing the electronegativity here is beneficial to the increase the activity. This combination indicates that the presence of a negatively charged group at the R_2 position favors activity, the cyano substitution of compound 63 (EC_{50} value of 8.82) at the R_2 position carries a greater negative charge and is more active than the methyl substituted compound 74 (EC_{50} value of 7.43). Figure 4(B) is a steric field contour map of CoMFA. The combination of the green (80% contribution) and yellow (20% contribution) colored regions on the R_1 substitution indicates that a larger group should be present at this position.

These were confirmed by some compounds in the system. In case of compounds 57 and 63, compound 63 substituted with an amide group at the R_1 position can be better filled in the binding pocket, making the binding more stable, and its activity is significantly higher than that of compound 57 (EC_{50} value of 8.12). However, not all the substitutions of large groups are beneficial to the activity. Compound 58 (EC_{50} value of 8.12) was substituted with a piperidine ring at the R_1 position, and its activity was significantly reduced due to the inflexibility and large size of the piperidine ring. Figure 4(C) is the electrostatic contour map of CoMSIA. The blue region represents an increase in the positive charge to facilitate improved activity of the compound, and the red portion represents an increase in the electronegativity of the compound which is advantageous for its activity. This is consistent with the electrostatic field contour map of CoMFA. The contour plots of the hydrogen bond acceptor (HBA) and hydrogen bond donor (HBD) of CoMSIA are shown in Figure 4(D,E), respectively. Combining the two contour maps, the R_1 position favors both hydrogen bond donor atoms and hydrogen bond acceptor atoms, indicating that hydrogen bonds are easily formed at this point. The contour maps of HBA and HBD can be verified by compounds 28, 21, and 39. The substituent groups at the R_1 position of compound 28 (EC_{50} value of 7.48) and compound 39 (EC_{50} value of 7.80) provide HBA and HBD atoms (-F and -CH_3, respectively), respectively. Bromine substitution of compound 31 (EC_{50} value of 8.20) neither satisfies the requirement for hydrogen bond formation, nor does it satisfy the CoMFA stereoscopic field contour map, and its activity is significantly reduced. In addition, the HBD contour map at R_3 indicates that this is not conducive to hydrogen bond donor atoms.

3.2. Molecular docking analysis

The Surflex-Dock module of SYBYL software was used for molecular docking. Before docking all the compounds, the co-crystallized ligand DJZ562 of 3M8Q was re-docked into the binding pocket to validate whether the used method and parameters were reliable and feasible. As shown in Figure 2(B), the re-docked conformation of DJZ562 was almost identical to its co-crystallized conformation. The NH linker connecting the central pyrimidine ring and the right pyridine ring forms a hydrogen bond with the main chain of Lys101. These interactions were in accordance with a previous study (Kertesz et al., 2010), and indicated that the generated docking-protocol was reliable and could be used for subsequent investigation.

All 54 compounds were then docked into the binding site using the same docking parameters (Figure 1(B)). The docking results showed that almost all compounds entered the HIV-1 RT binding pocket in a similar horseshoe-like conformation. In addition, the NH linker connecting the central pyrimidine ring and the right pyridine ring forms a hydrogen bond with the main chain of Lys101, which is almost a characteristic marker of this series of compounds. As shown in Figure 5(A), Compound 23 is encapsulated within the HIV-1 RT active region and the horseshoe-liked conformation of the compound allows it to better fill the entire pocket. Consistent with HBA and HBD contour maps of CoMSIA, the sulfonamide group at the R_2 position of compound 23 provides a hydrogen bond donor and a hydrogen bond acceptor to form hydrogen bonds with the residues Lys104 and Val106, respectively (yellow dotted line represents Hydrogen bond). The combination of these two hydrogen bonds greatly enhances the stability of binding. In addition, the open region formed by the Pro236 hairpin ring can accommodate larger volume groups, which validates the stereoscopic contour map of CoMFA. Compounds with large substituents such as compound 23 make good use of this open region. This makes the binding of the compound to the protein more stable. Notably, the open region and the nearby hydrophobic amino acids such as Val106, Pro236,

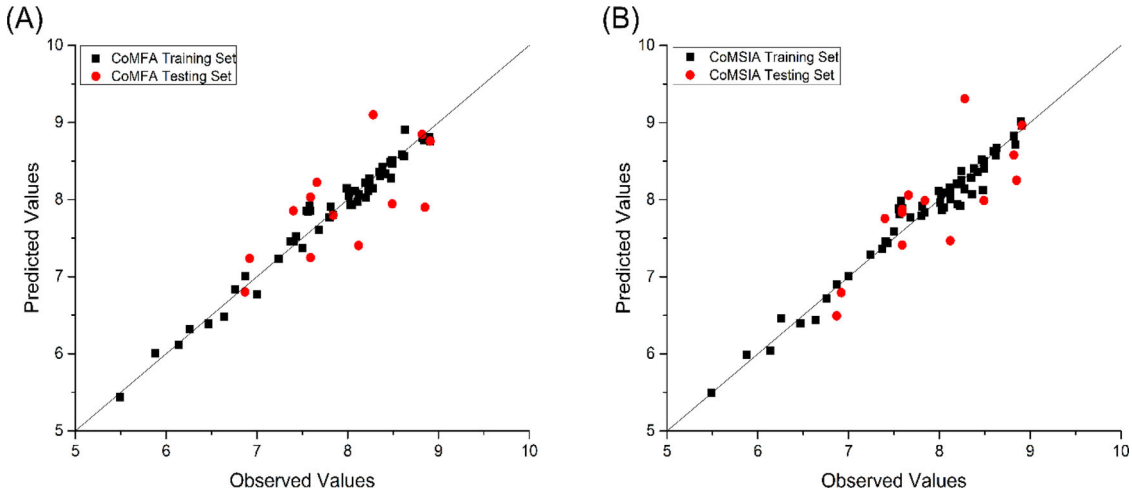

Figure 3. Correlation between training set and test set observed value data and predicted value data. A: COMFA model results; B: COMSIA model results.

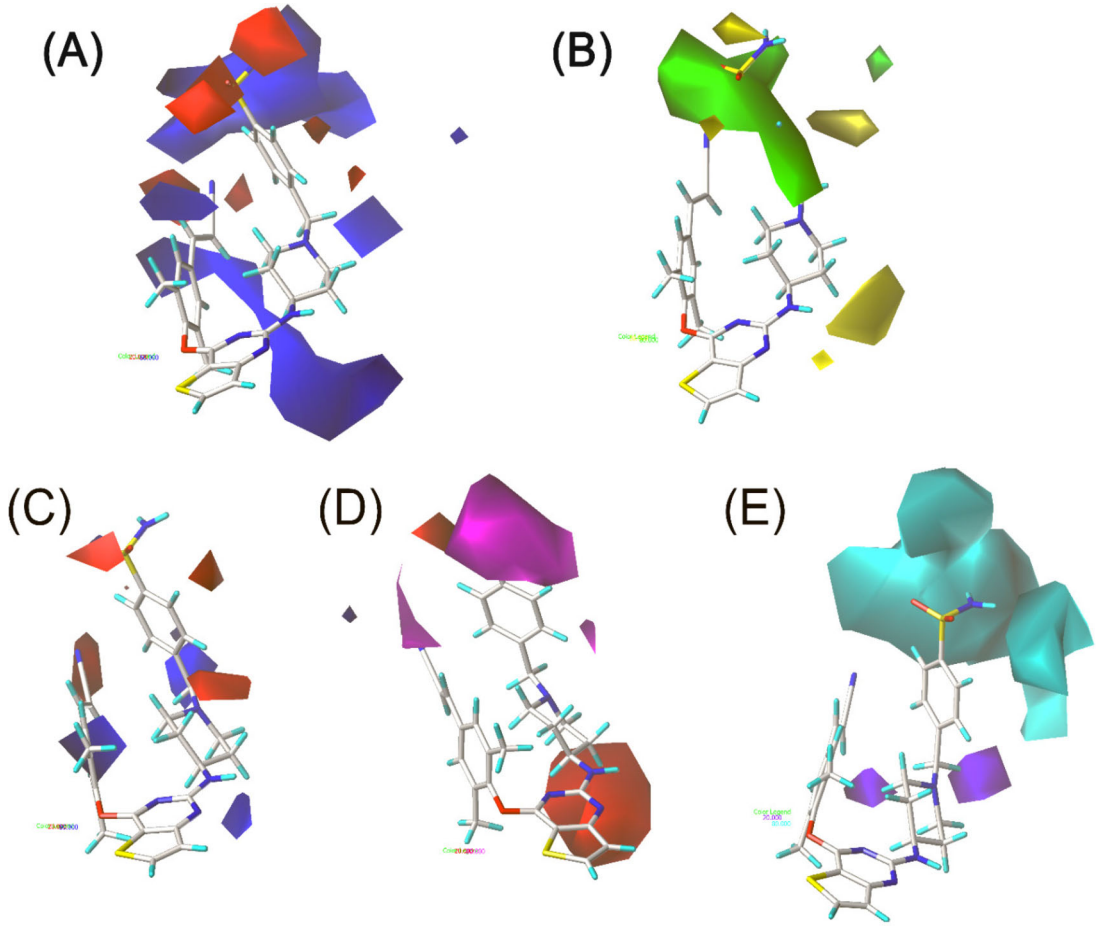

Figure 4. The contour map of CoMFA (CoMSIA) for HIV-1 RT. A: CoMFA electrostatic field; B: CoMFA stereoscopic field; C: CoMSIA electrostatic field; D: CoMSIA hydrogen bond acceptor field; E: CoMSIA hydrogen bond donor field.

Leu234, LYS104 and Phe227 can form a hydrophobic interaction with the compounds. The open region filled with the thiophene ring, often acting as a channel entrance, is surrounded by several relatively inflexible residues such as Leu100, Lys101, Glu138 and Val179. This may be a great inspiration for research on the new generation of NNRTIs. In addition, the O-benzene ring on the left side of most compounds and its substituents can form a hydrophobic interaction with the hydrophobic residues Tyr181, Val179 and Trp229, which ensures the stability of the binding. A 2 D model of Compound 23 docked with HIV-1 RT is shown in Figure 5(B). The central pyrimidine ring of compound 23 forms a σ-π hyperconjugate effect with the side chain of residue Leu100, and the C-phenyl group on the right side

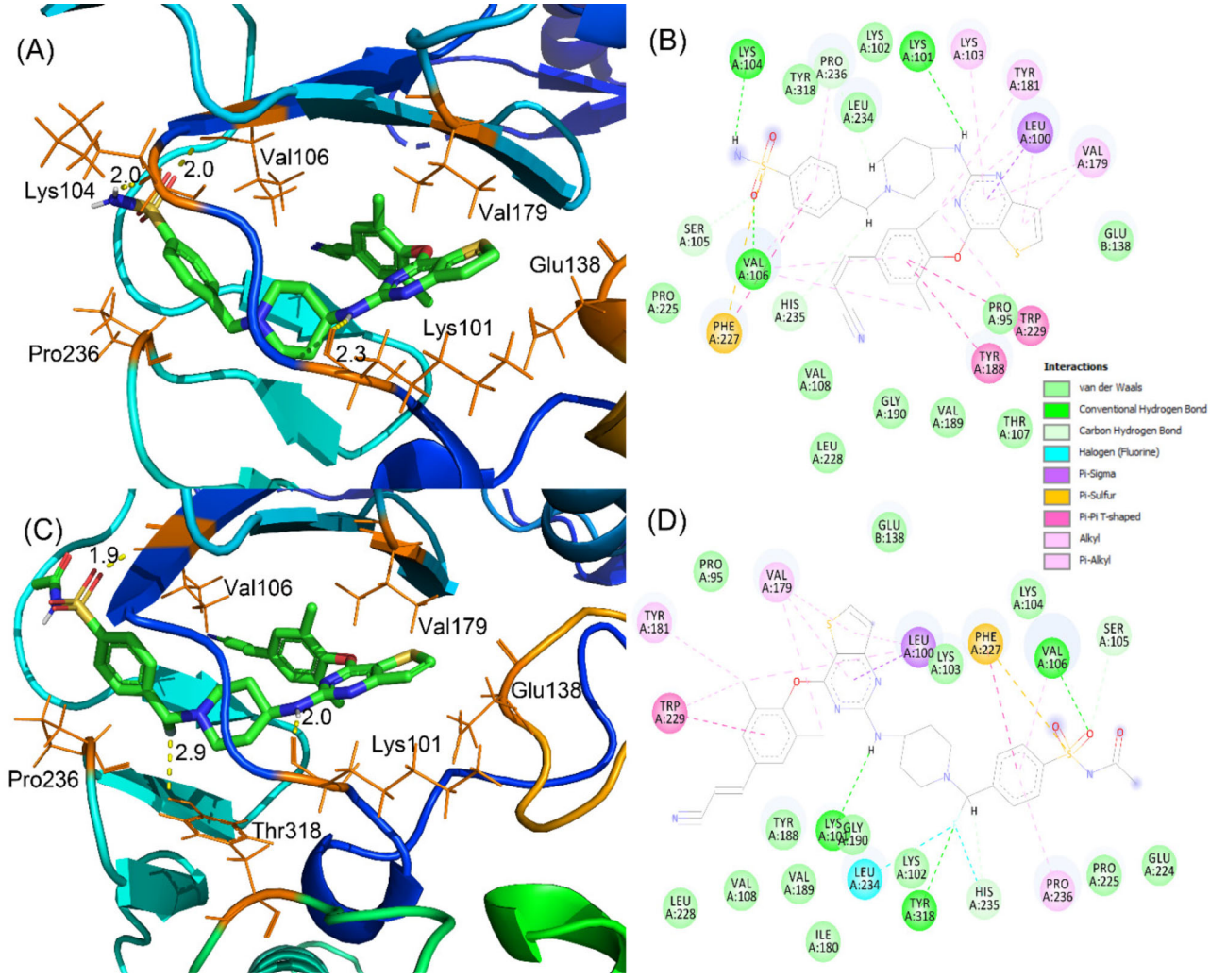

Figure 5. A: Docking model of HIV-1 RT with compound 23; B: Docking model of HIV-1 RT with compound 23(2D); C: Docking model of HIV-1 RT with compound 23j; D: Docking model of HIV-1 RT with compound 23j (2D).

also forms a σ-π hyperconjugate effect with the side chain of residue Pro236. This greatly enhances the stability of the entire system. The π-π stack formed by the O-phenyl group and the benzene ring of Trp229 is beneficial to binding. In the figure, the lavender linkage indicates a hydrophobic relationship between the amino acid residue and the corresponding ligand, and the light green amino acid residue without attachment indicates van der Waals interaction between the residue and ligand. The blue ring outside the partial amino acid residue indicates the surface of the receptor active pocket that has solubility, and that the diameter of the ring is proportional to the size of the dissolved surface.

3.3. Generation of new leads

By analyzing the results of 3 D-QSAR and molecular docking, compound 23 was used as the lead compound, and the dominant modified region determined by the above results was modified. The structure and predicted activity values of the newly designed compounds are shown in Table 2.

Molecular docking of all new compounds showed that all of them had the same docking pattern as compound 23 and formed "characteristic" hydrogen bonds with the backbone of Lys101. Activity prediction was performed using the established 3 D-QSAR model, and the prediction results are listed in Table 2. The predicted values of the CoMFA and CoMSIA models indicate that the new HIV-1 RT inhibitors are active. The R_1 group was modified by increasing the group volume or introducing groups to increase the number of hydrogen bond donors and acceptors. The results showed that the activity of the compound was improved when $-SO_2NHCOCH_3$ was introduced here. Based on the modification of R_1, when the R_3 group was introduced with -F as a hydrogen bond acceptor, the obtained compound 23j showed remarkably improved inhibitory activity. Finally, a series of halogen-substituted compounds (23k, 23l) was also designed to have a significant increase in HIV-1 RT inhibitory activity. Taking compound 23j as an example, as shown in Figure 5(C,D), the conformation of compound 23j into the binding pocket was substantially identical to that of compound 23, but the introduction of $-SO_2NHCOCH_3$ further filled the open region at Pro236. In addition, the -F group introduced at the R_3 position also forms a hydrogen bond with Tyr318. These all guarantee the stability of the complex and fully prove the reliability of the 3 D-QSAR model.

Table 2. Structures and predicted Synthetic accessibility of the novel compounds.

No.	R_1	R_2	R_3	Predicted values (pEC_{50})		Synthetic accessibility
				CoMFA	CoMSIA-SEHDA	
23a	-SOOH		H	7.33	7.42	5.07
23b			H	8.35	8.89	5.06
23c			H	8.08	8.27	5.00
23d	$-SO_2NH_2$	-COOH	H	7.82	8.52	4.65
23e	$-SO_2NH_2$		H	7.69	7.83	4.99
23f			H	8.05	8.35	4.94
23g	$-SO_2NH_2$		H	8.07	8.65	5.05
23h		$-CH_3$	H	7.97	8.85	4.83
23i			H	8.35	8.87	5.06
23j			F	9.20	9.36	5.35
23k			Cl	9.38	9.40	5.34
23l			Br	9.13	9.34	5.43

3.4. Molecular dynamics simulations

The 50 ns MD simulations were performed on compounds 23 and 23j to further understand the detailed dynamic binding mode. The system overall convergence and stability of MD simulations were monitored by the root-mean-square deviation (RMSD) of backbone atoms (C, Ca, N and O) with respect to the initial docking structure (Figure 6(A,B)). The fluctuations of 23 and 23j were maintained at an average of 2 Å and 2.5 Å, suggesting the stability of the complex conformation. By comparing the MD simulated representative snapshots of the complexes, it was found that they all acquired a horseshoe-like conformation similar to HIV-1 RT and maintained the characteristic H-bonding interactions with the backbone of Lys101. These were basically similar with the initial docking structures, thereby proving the docking reliability. Thus, subsequent discussions on MD simulations would be reasonable to analyze the conformations extracted from the MD trajectories.

Moreover, to further analyze the residue contribution of HIV-1 RT, the root-mean-square fluctuations (RMSFs) versus the residue number of all the compounds are shown in Figure 6(C,D). The two complexes had similar dynamic fluctuation trends and RMSF distributions, indicating that these inhibitors had a binding mode similar to HIV-1 RT. The key residues include Leu100, Lys101, Val106, Val179, Phe227 and Pro236, exhibiting rigid behavior with low RMSF values (around 1Ang), indicating a stable compound and complex properties. These results support the reliability of MD simulation. Residue fluctuations in complex 23-HIV-1 RT are generally higher than the residue fluctuations in 23j-HIV-1 RT. These results indicate that 23j has good binding affinity with HIV-1 RT.

Obviously, the final simulation structure and the initial docked structure are all located in the original binding pocket, but both Compound 23 and Compound 23j are shifted to the right to varying degrees (Figure 7(A,B)). First, the conformation of Compound 23 did not change significantly. It also moves back a distance on the basis of the rightward shift, which directly leads to an increase in the distance between the sulfonamide group at the R_1 position and the residues Lys104 and Val106 (>4 Å). This breaks the hydrogen bonds in the original conformation (Figure 7(C)). Second, the left portion of compound 23 moves upwards by a large amount. There is a strong van der Waals interaction between the residues Val108, Tyr188, and compound 23, which may be one of the factors influencing this phenomenon. However, in general, the hydrogen bonding between Lys101 and the NH linker is stable, and there is still a strong van der Waals interaction between Leu100, Phe227 and Compound 23. In addition, the sulfonamide group is filled in the open region formed by Pro236. These factors indicate that compound 23 still has good activity.

Compound 23j shifts to the right front with respect to the initial conformation in a horseshoe-like conformation, but this does not affect the extremely strong van der Waals interaction between the amino acid residues Leu100 and Phe227 (Figure 7(D)). In addition, this also enhances the electrostatic interaction of Trp229 and Ser105 with compound 23j and significantly enhances the contribution of these residues to the binding free energy. Further, the substituent of compound 23j at the position of R_1 undergoes a leftward rotation after the MD simulations were balanced, and forms a hydrogen bond with Lys104 and Val106 (2.9 Å, 2.5 Å, respectively). Compound 23j also forms a hydrogen bond with Lys101 and Thr318, with a total of four hydrogen bonds. These favorable interactions greatly enhance the tightness of binding and may be the reason why the predicted activity of compound 23j is higher than that of the compound 23, which also proves the reliability of the 3 D-QSAR model.

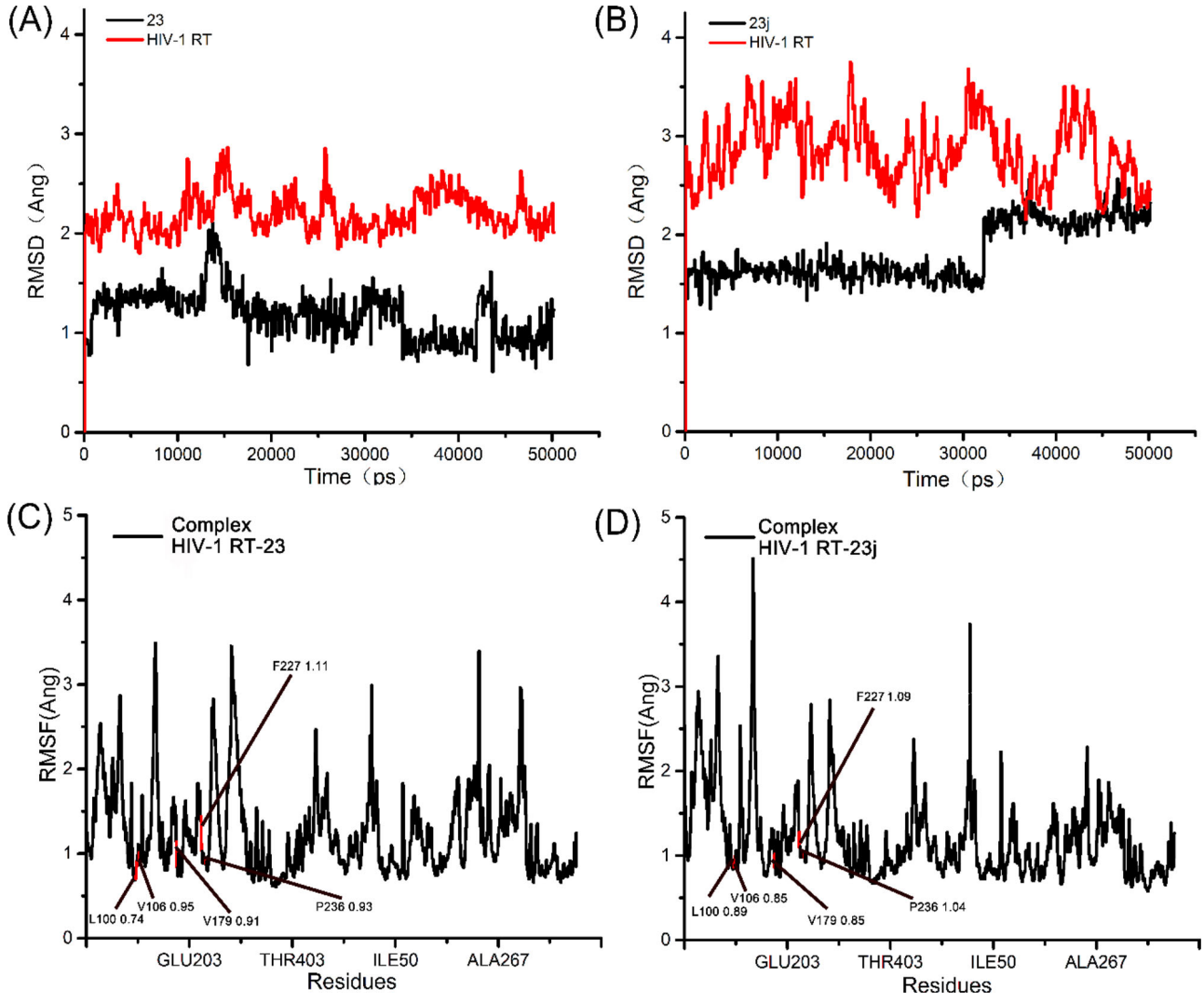

Figure 6. A, B: RMSD of the protein, compound 23 and compound 23j; C, D: The RMSF of the complexes of 23 and 23j.

Based on the MD simulation, binding free energies were calculated using the MM-PBSA and MM-GBSA methods to assess the energetic aspects of the two representative complexes. The binding free energy terms of complexes are listed in Table 3. It can be observed that the van der Waals energy (ΔE_{vdw}) is the main factor affecting the total binding energy. In the MM/GBSA model and the MM/PBSA model, total gas phase free energy consists of van der Waals energy and electrostatic energy. However, the total solvation free energy in the MM/GBSA model is composed of EGB (Polar solvation energy) and ESURF (Non-polar solvation energy), whereas in the MM/PBSA model, total solvation free energy is determined by EPB (Polar solvation energy), ENPOLAR (Non-polar solvation energy), and EDISPER (Edispersion).

Edispersion is calculated by numerical determination of the solvent accessible surface area. Finally, the ΔG_{gas} of complex 23-HIV-1 RT and 23j-HIV-1 RT was $-104.5\,kcal{\cdot}mol^{-1}$ and $-118.54\,kcal{\cdot}mol^{-1}$, respectively. In the MM/GBSA model, the ΔG_{solv} of complex 23-HIV-1 RT and 23j-HIV-1 RT was $28.70\,kcal{\cdot}mol^{-1}$ and $35.77\,kcal{\cdot}mol^{-1}$, respectively. In the MM/PBSA model, the ΔG_{solv} of complex 23-HIV-1 RT and 23j-HIV-1 RT was $95.11\,kcal{\cdot}mol^{-1}$ and $105.17\,kcal{\cdot}mol^{-1}$, respectively. The results showed that the ΔG binding of compounds 23 and 23j by MM/PBSA was -75.84 and $-82.77\,kcal{\cdot}mol^{-1}$, respectively, and that by MM/GBSA was -9.43 and $-13.37\,kcal{\cdot}mol^{-1}$, respectively. Furthermore, the ΔG binding of HIV-1 RT with 23j was more negative than that with 23, indicating that 23j bonded with HIV-1 RT more tightly than 23.

3.5. Binding free energy analysis

To determine the key residues that contribute to the activity of the inhibitor, and to understand the interaction between the inhibitor and the residues surrounding the HIV-1 RT binding pocket, the energy contribution of each residue in the HIV-1 RT protein was calculated by MM/PBSA.py. As seen from Figure 8, the interaction of the residue with the inhibitor fluctuated between 2.84 and $-3.15\,kcal{\cdot}mol^{-1}$.

Modification of compound 23j at the R_1 position resulted in more interaction with residues such as Leu234 and Trp229 in the open region generated by Pro236, which results in tighter binding. More residues in Complex 23j-HIV-1 RT are better involved in energy contribution, showing a better binding pattern. Residues that contribute more to the binding energy such as Leu100, Val106, Pro236 and Val179,

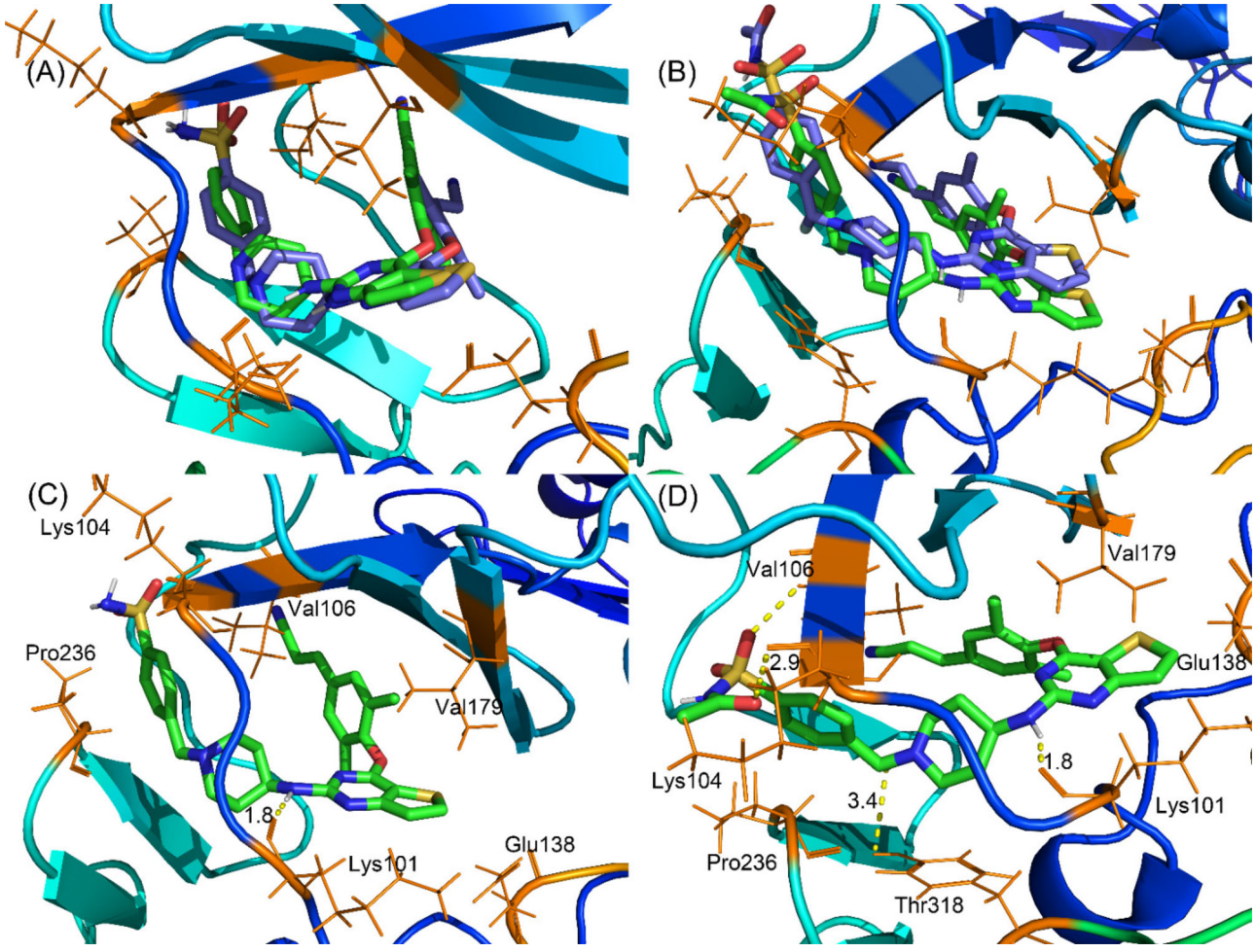

Figure 7. Structural comparison between initial (blue) and representative snapshots from the MD simulation (green) of 23-HIV-1 RT (A) and 23j-HIV-1 RT (B); The binding mode of 23-HIV-1 RT (C) and 23j-HIV-1 RT (D) after MD ++simulation.

Table 3. The binding free energies and individual energy components (kcal·mol^{-1}).

	Compound	ΔVDWAALS	ΔEEL	ΔEPB	ΔENPOLAR	ΔEDISPER	ΔG bind
MM/PBSA	23	−81.54	−23.00	57.69	−53.04	90.46	−9.43
	23j	−84.28	−34.26	65.29	−56.35	96.22	−13.37

EPB/EGB: Polar solvation energy; ENPOLAR/ESURF: Non-polar solvation energy; VDWAALS: van der Waals energy; EEL: Electrostatic energy; EDISPER: Edispersion.

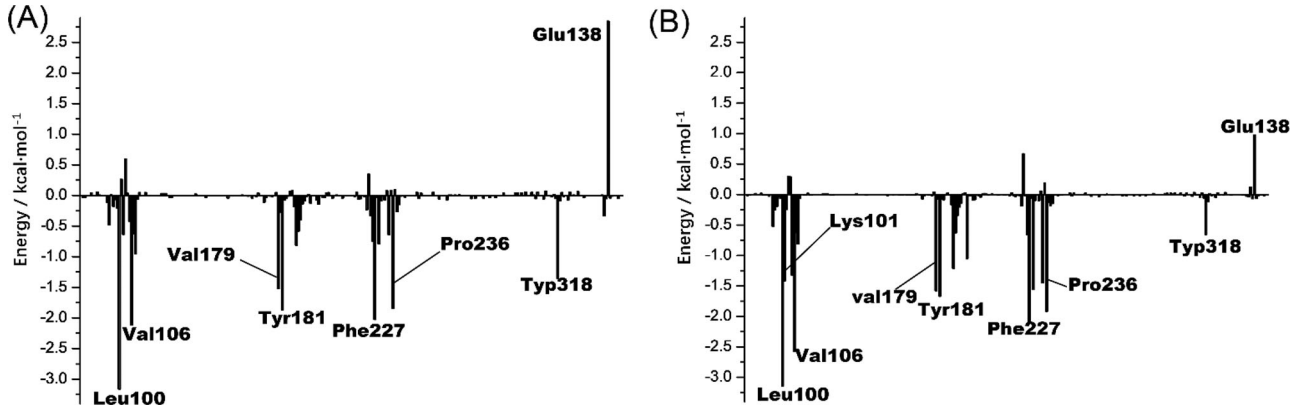

Figure 8. Comparison of per-residue energy decomposition of 23-HIV-1 RT (A), 23j-HIV-1 RT (B).

interact more with the R_1, and R_3, and the common skeleton of the compounds, so that the R_2 position can be further optimized. In the open region formed by amino acid residues such as Leu100, Lys101, Glu138 and Val179, the negative contribution of Glu138 to binding energy should be considered. Optimizing the thiophene ring could be a good

Table 4. ADMET properties of novel designed molecules.

No	Absorption	Distribution	Metabolism CYP							Excretion	Toxicity		
			2D6	3A4	1A2	2C19	2C9	2D6	3A4				
	Intestinal absorption (human) (%absorbed)	VDss (human) (logL/Kg)	Substrate		Inhibitor					Total clearance (log mL/min/kg)	AMES	Hepatotoxicity	Skin Sensitization
23	71.135	−0.607	No	Yes	No	No	Yes	No	No	0.762	No	Yes	No
23a	83.659	−0.581	No	Yes	No	Yes	Yes	No	Yes	0.408	No	Yes	No
23b	49.024	0.036	No	Yes	No	No	No	No	No	0.543	No	Yes	No
23c	61.037	0.351	No	Yes	No	No	No	No	No	0.136	No	Yes	No
23d	90.320	−0.596	No	Yes	No	Yes	Yes	No	Yes	0.472	No	Yes	No
23e	56.201	0.074	No	Yes	No	No	No	No	No	0.645	No	Yes	No
23f	61.256	−1.277	No	Yes	No	No	No	No	No	0.208	No	Yes	No
23g	59.692	−0.557	No	Yes	No	No	No	No	No	0.244	No	Yes	No
23h	83.659	−0.581	No	Yes	No	Yes	Yes	No	Yes	0.408	No	Yes	No
23i	91.520	−0.598	No	Yes	No	No	Yes	No	Yes	0.236	No	Yes	No
23j	85.384	−0.416	No	Yes	No	Yes	No	No	Yes	0.368	No	Yes	No
23k	85.329	−0.403	No	Yes	No	Yes	No	No	Yes	0.216	No	Yes	No
23l	71.135	−0.607	No	Yes	No	No	Yes	No	No	0.762	No	Yes	No

direction. Among this series of compounds, the key residues Leu100, Phe227, Val106 and Pro236 contributed significantly to the binding free energy, whereas the contributions of the easily mutated residues Lys103, Tyr108 and Tyr188 were greatly reduced. This may be the reason why this series of compounds was more active against mutant HIV-1 RT. At the same time, these key residues verify the reliability of the docking results. Energy decomposition is a good way to verify the binding of receptors to ligands. These results thus provide useful clues for the rationalization of drug design.

3.6. ADME/T profiles and synthetic accessibility

To further study the druggability of the virtually designed compounds, ADME/T properties were also predicted using the pkCSM online server (Table 4). The intestinal absorbance of the compounds ranged from 49.024% to 91.520%, indicating a high absorption characteristic, especially for compound 23i. The logVDss of most compounds was less than −0.15, which meant that they were distributed more in the plasma rather than in tissues. For metabolism, all compounds were predicted as substrates for the CYP450 3A4 subtype, which meant they might be metabolized by CYP 3A4. Moreover, compound 23b, 23c, 23e, 23f and 23g could not inhibit the CYP450 family, whereas the other compounds might inhibit the CYP450 2C19/2C9/2D6/3A4 subtype. Based on the predicted total clearance, all the compounds could be cleared by hepatic and renal tissue in a combinatory manner. The predicted toxicity indicated all the compounds might be harmful for the liver, and that all compounds had no skin sensitization and mutagenic potency. All of the designed molecules with synthetic accessibility scores ranging between 4.65 and 5.43, demonstrated low complexity of structure thus indicating synthetic feasibility (Table 3). Computational pharmacokinetic and toxicological studies and synthetic accessibility indicated that the virtually designed compounds may serve as lead compounds for further development.

4. Conclusion

This study used a combination of 3 D-QSAR, molecular docking, and MD simulation to analyze the decisive inhibition mechanism of a new generation of HIV-1 RT inhibitors. The 3 D-QSAR model established in this study is robust and has high predictive power, revealing the key structural factors affecting HIV-1 RT inhibitory activity. The docking results are consistent with other studies. Based on these results, 12 new thieno[3,4-d]pyrimidines were designed and predicted for activity, and the results indicate that these compounds have good predictive activity and a reasonably good ADME/T profile. Molecular docking, MD simulation, and free energy decomposition results indicate that the group at the R_1 position of the compound and the horseshoe-like conformation are critical for the activity and stability of the inhibitor. Hydrogen bonding and van der Waals interactions of key residues such as Leu100, Lys101, Val106, Phe227 and Pro236 play a key role in ligand-receptor binding. These results thus provide a useful reference for the rational design of a new generation of HIV-1 RT inhibitors.

Acknowledgements

The authors are grateful to the national natural science foundation of China (81171508, 31400667); Key project of Chongqing natural science foundation (cstc2018jcyjAX0683, cstc2015jcyjBX0080); Scientific and technological research project of Chongqing municipal education commission (KJZD-K201801102, KJ1600908, KJQN201801132); Opening Foundation of State Key Laboratory of Silkworm Genome Biology (sklsgb1819-2). Computational support from the Supported by the scientific research startup fund of Chongqing university of technology (2017ZD42).

Disclosure statement

The authors declare no conflict of interest, financial or otherwise.

References

Anthony, N. J. (2004). HIV-1 integrase: A target for new AIDS chemotherapeutics. *Current Topics in Medicinal Chemistry*, *4*(9), 979–990. doi:10.2174/1568026043388448

Berendsen, H. J., Postma, J. V., van Gunsteren, W. F., DiNola, A., & Haak, J. R. (1984). Molecular dynamics with coupling to an external bath. *The Journal of Chemical Physics*, *81*(8), 3684–3690. doi:10.1063/1.448118

Bush, B. L., & Nachbar, R. B. (1993). Sample-distance partial least squares: PLS optimized for many variables, with application to CoMFA. *Journal*

of Computer-Aided Molecular Design, *7*(5), 587–619. doi:10.1007/BF00124364

Buzón, M. J., Massanella, M., Llibre, J. M., Esteve, A., Dahl, V., Puertas, M. C., … Sharkey, M. (2010). HIV-1 replication and immune dynamics are affected by raltegravir intensification of HAART-suppressed subjects. *Nature Medicine*, *16*(4), 460. doi:10.1038/nm.2111

Cheng, L. P., Wang, T. C., Yu, R., Li, M., & Huang, J. W. (2018). Design, synthesis and biological evaluation of novel zanamivir derivatives as potent neuraminidase inhibitors. *Bioorganic & Medicinal Chemistry Letters*, *28*(23-24), 3622–3629. doi:10.1016/j.bmcl.2018.10.040

Cramer, R. D. (2003). Topomer CoMFA: A design methodology for rapid lead optimization. *Journal of Medicinal Chemistry*, *46*(3), 374–388. doi:10.1021/jm020194o

Cramer, R. D., Patterson, D. E., & Bunce, J. D. (1988). Comparative molecular field analysis (CoMFA). 1. Effect of shape on binding of steroids to carrier proteins. *Journal of the American Chemical Society*, *110*(18), 5959–5967. doi:10.1021/ja00226a005

Das, K., Lewi, P. J., Hughes, S. H., & Arnold, E. (2005). Crystallography and the design of anti-AIDS drugs: Conformational flexibility and positional adaptability are important in the design of non-nucleoside HIV-1 reverse transcriptase inhibitors. *Progress in Biophysics and Molecular Biology*, *88*(2), 209–231. doi:10.1016/j.pbiomolbio.2004.07.001

de Béthune, M.-P. (2010). Non-nucleoside reverse transcriptase inhibitors (NNRTIs), their discovery, development, and use in the treatment of HIV-1 infection: A review of the last 20 years (1989–2009). *Antiviral Research*, *85*(1), 75–90. doi:10.1016/j.antiviral.2009.09.008

Esté, J. A., & Cihlar, T. (2010). Current status and challenges of antiretroviral research and therapy. *Antiviral Research*, *85*(1), 25–33. doi:10.1016/j.antiviral.2009.10.007

Ghosh, A. K., Chapsal, B. D., Weber, I. T., & Mitsuya, H. (2008). Design of HIV protease inhibitors targeting protein backbone: An effective strategy for combating drug resistance. *Accounts of Chemical Research*, *41*(1), 78–86. doi:10.1021/ar7001232

Homeyer, N., & Gohlke, H. (2012). Free energy calculations by the molecular mechanics Poisson − Boltzmann surface area method. *Molecular Informatics*, *31*(2), 114–122. doi:10.1002/minf.201100135

Hymes, K. B., Greene, J. B., Marcus, A., William, D. C., Cheung, T., Prose, N. S., … Laubenstein, L. J. (1981). Kaposi's sarcoma in homosexual men—A report of eight cases. *The Lancet*, *318*(8247), 598–600. doi:10.1016/S0140-6736(81)92740-9

Johnson, V. A., Calvez, V., Günthard, H. F., Paredes, R., Pillay, D., Shafer, R., … Richman, D. D. (2011). 2011 update of the drug resistance mutations in HIV-1. *Topics in Antiviral Medicine*, *19*(4), 156. doi:10.1002/9780470773680.ch5

Jorgensen, W. L., Chandrasekhar, J., Madura, J. D., Impey, R. W., & Klein, M. L. (1983). Comparison of simple potential functions for simulating liquid water. *The Journal of Chemical Physics*, *79*(2), 926–935. doi:10.1063/1.445869

Kang, D., Fang, Z., Huang, B., Lu, X., Zhang, H., Xu, H., … Liu, X. (2017). Structure-based optimization of thiophene [3, 2-d] pyrimidine derivatives as potent HIV-1 non-nucleoside reverse transcriptase inhibitors with improved potency against resistance-associated variants. *Journal of Medicinal Chemistry*, *60*(10), 4424–4443. doi:10.1021/acs.jmedchem.7b00332

Kang, D., Fang, Z., Li, Z., Huang, B., Zhang, H., Lu, X., … Liu, X. (2016). Design, synthesis, and evaluation of thiophene [3, 2-d] pyrimidine derivatives as HIV-1 non-nucleoside reverse transcriptase inhibitors with significantly improved drug resistance profiles. *Journal of Medicinal Chemistry*, *59*(17), 7991–8007. doi:10.1021/acs.jmedchem.6b00738

Kertesz, D. J., Brotherton-Pleiss, C., Yang, M., Wang, Z., Lin, X., Qiu, Z., … Dunten, P. W. (2010). Discovery of piperidin-4-yl-aminopyrimidines as HIV-1 reverse transcriptase inhibitors. N-benzyl derivatives with broad potency against resistant mutant viruses. *Bioorganic & Medicinal Chemistry Letters*, *20*(14), 4215–4218. doi:10.1016/j.bmcl.2010.05.040

Klebe, G., Abraham, U., & Mietzner, T. (1994). Molecular similarity indices in a comparative analysis (CoMSIA) of drug molecules to correlate and predict their biological activity. *Journal of Medicinal Chemistry*, *37*(24), 4130–4146. doi:10.1021/jm00050a010

Kollman, P. A., Massova, I., Reyes, C., Kuhn, B., Huo, S., Chong, L., … Cheatham, T. E. (2000). Calculating structures and free energies of complex molecules: Combining molecular mechanics and continuum models. *Accounts of Chemical Research*, *33*(12), 889–897. doi:10.1021/ar000033j

Li, M., Wei, D., Zhao, H., & Du, Y. (2014). Genotoxicity of quinolones: Substituents contribution and transformation products QSAR evaluation using 2D and 3D models. *Chemosphere*, *95*(1), 220–226. doi:10.1016/j.chemosphere.2013.09.002

Li, S., Fan, J., Peng, C., Chang, Y., Guo, L., Hou, J., … Sun, P. (2017). New molecular insights into the tyrosyl-tRNA synthase inhibitors: CoMFA, CoMSIA analyses and molecular docking studies. *Scientific Reports*, *7*(1), 11525. doi:10.1038/s41598-017-10618-1

Martins, S., Ramos, M. J., & Fernandes, P. A. (2008). The current status of the NNRTI family of antiretrovirals used in the HAART regime against HIV infection. *Current Medicinal Chemistry*, *15*(11), 1083–1095. doi:10.2174/092986708784221467

Mayer, K. H., & Venkatesh, K. K. (2010). Antiretroviral therapy as HIV prevention: Status and prospects. *American Journal of Public Health*, *100*(10), 1867–1876. doi:10.2105/AJPH.2009.184796

Patel, P. D., Patel, M. R., Kaushik-Basu, N., & Talele, T. T. (2008). 3D QSAR and molecular docking studies of benzimidazole derivatives as hepatitis C virus NS5B polymerase inhibitors. *Journal of Chemical Information and Modeling*, *48*(1), 42–55. doi:10.1021/ci700266z

Richmond, N. J., Abrams, C. A., Wolohan, P. R., Abrahamian, E., Willett, P., & Clark, R. D. (2006). GALAHAD: 1. Pharmacophore identification by hypermolecular alignment of ligands in 3D. *Journal of Computer-Aided Molecular Design*, *20*(9), 567–587. doi:10.1007/s10822-006-9082-y

Sarafianos, S. G., Marchand, B., Das, K., Himmel, D. M., Parniak, M. A., Hughes, S. H., & Arnold, E. (2009). Structure and function of HIV-1 reverse transcriptase: Molecular mechanisms of polymerization and inhibition. *Journal of Molecular Biology*, *385*(3), 693–713. doi:10.1016/j.jmb.2008.10.071

Shattock, R. J., Warren, M., McCormack, S., & Hankins, C. A. (2011). Turning the tide against HIV. *Science*, *333*(6038), 42–43. doi:10.1126/science.1206399

Tarasova, O., Poroikov, V., & Veselovsky, A. (2018). Molecular docking studies of HIV-1 resistance to reverse transcriptase inhibitors: Mini-review. *Molecules*, *23*(5), 1233. doi:10.3390/molecules23051233

Zhan, P., Liu, X., Li, Z., Pannecouque, C., & De Clercq, E. (2009). Design strategies of novel NNRTIs to overcome drug resistance. *Current Medicinal Chemistry*, *16*(29), 3903–3917. doi:10.2174/092986709789178019

Zhan, P., Pannecouque, C., De Clercq, E., & Liu, X. (2016). Anti-HIV drug discovery and development: Current innovations and future trends: Miniperspective. *Journal of Medicinal Chemistry*, *59*(7), 2849–2878. doi:10.1021/acs.jmedchem.5b00497