



J. Serb. Chem. Soc. 84 (3) 303–316 (2019)
JSCS–5185

Journal of
the Serbian
Chemical Society

JSCS-info@shd.org.rs • www.shd.org.rs/JSCS

UDC 616.98:578.828+66.011:539.16:615.2
Original scientific paper

Molecular modeling studies of HIV-1 non-nucleoside reverse transcriptase inhibitors using 3D-QSAR, virtual screening and docking simulations

JIAN-BO TONG*, SHANG-SHANG QIN, SHAN LEI and YANG WANG

Shaanxi Key Laboratory of Chemical Additives for Industry, Shaanxi University of Science and Technology, Xi'an 710021, P. R. China

(Received 4 September, revised 15 October, accepted 15 November 2018)

Abstract: Acquired immunodeficiency syndrome (AIDS) is a significant human health threat around the world and therefore, the study of anti-HIV drug design has become an important task for today's society. In this paper, a three-dimensional quantitative structure–activity relationships study (3D-QSAR) was conducted on 72 HIV-1 non-nucleoside reverse transcriptase inhibitors (NNRTIs) using Topomer comparative molecular field analysis (Topomer CoMFA). The multiple correlation coefficients of fitting, cross-validation, and external validation were 0.899, 0.788 and 0.942, respectively. The results indicated that the obtained model had both a favorable estimation stability and a good prediction capability. Topomer Search was used to search appropriate R groups from the ZINC database, thereby, 14 new compounds were designed, and 12 of the new compounds were predicted to be more active than the template molecule. These results strongly suggest that the Topomer search was effective in screening and could be a useful guide in the design of new HIV-1 drugs. The ligands of the template molecule and the new designed compounds were used for molecular docking to study the interaction of these compounds with the protein receptor. The results showed that the ligands would generally form hydrogen-bonding interactions with the residues Ala28, Asp29, Gly49 and Ile50 of the protein receptor, thereby providing additional insights for the designing of even more effective drugs.

Keywords: 3D-QSAR; NNRTIs; Topomer CoMFA; Topomer search; new drug design; molecular docking.

INTRODUCTION

AIDS caused by human immunodeficiency virus type-1 (HIV-1) has been threatening human health and no drug could hitherto cure it completely.^{1–3} There are two types of HIV that infect humans: HIV-1 and HIV-2. HIV-1 is the more virulent form because it is more easily transmitted.³ There are millions of people

*Corresponding author E-mail: jianbotong@aliyun.com
<https://doi.org/10.2298/JSC180904098T>

living with HIV, and its rapid spread and high fatality rate resulted in large economic losses and social impacts.⁴ While over 20 antiretroviral drugs have been approved for use in HIV-infected patients,⁵ many of these drugs are becoming less effective because of resistance resulting from changes in the mutation-prone HIV. Therefore, the successful management of HIV infection requires new patient-friendly and effective therapies. The host proteins involved in the viral replication cycle have been used as drug targets when designing inhibitors to prevent the spread of infection. These targets include reverse transcriptase, protease, integrase, polymerase, glycoprotein, as well as the host cell receptor (CD4) and co-receptors (CCR5 and CXCR4⁶). NNRTIs are among the most potent and promising of the anti-AIDS agents that specifically target the HIV-1 reverse transcriptase (RT). The HIV-1 RT is an asymmetric heterodimer, comprised of a p66 subunit (560 amino acids) and a p51 subunit (440 amino acids)⁷ that has been used as an effective target for antiretroviral drugs. However, the effectiveness of NNRTIs^{8,9} drugs is hampered by the rapid emergence of drug-resistant viruses and severe side effects associated with long-term use. Therefore, there is a need to develop additional, highly potent NNRTIs with a broad spectrum of antiviral activity and improved pharmacokinetic properties. Consequently, more efficient strategies that facilitate and shorten the drug discovery process would be extremely beneficial.¹⁰

The availability of computational techniques based on QSARs provide a possible means of accelerating the drug design process, since a QSAR is a technique that attempts to summarize chemical and biological information in a form that allows the rapid generation of relationships between chemical structures and biological activity.¹¹ The success of a QSAR study depends on a proper selection of the variables and on a useful representation of the information. The best variables are those for which small changes in the value result in significant changes in activity. A 3D-QSAR model better reflects the interactions between a substrate and the receptor than does a 2D-QSAR model. Comparative molecular field analysis (CoMFA)¹² is a widely-used method of 3D-QSAR investigation. In this paper, Topomer CoMFA,^{13,14} a second generation of CoMFA, was employed to construct a 3D-QSAR model utilizing data for 72 HIV-1 NNRTIs to analyze the chemical–biological interactions governing their activities. The Topomer CoMFA model was combined with the Topomer search¹⁵ technology to conduct a ligand-based virtual screening of possible compounds in order to lay the foundation of improved drug design.

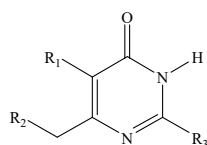
MATERIALS AND METHODS

Preparation of data set

A total of 72 dihydro(alkylthio)benzylxopyrimidines (S-DABOs), a class of highly potent NNRTIs, was selected from the published literature.¹⁶ The full data set was randomly divided into two parts: a training set of 57 compounds (compounds **1** to **57**) to build the 3D-QSAR

model and a test set of 15 compounds (compounds **58** to **72**) for use in the evaluation of the model. The biological activity was characterized by the half-maximal (50 %) inhibitory concentration (IC_{50}). The IC_{50} values in μM range were converted to the M range and then converted to a logarithmic scale ($pIC_{50} = -\log IC_{50}$).¹⁷ The chemical structures and their pIC_{50} values are listed in Table I.

TABLE I. The structures and activity values of the S-DABO derivatives; compounds **58–72** were chosen as the test set



No.	R1	R2	R3	pIC_{50}	No.	R1	R2	R3	pIC_{50}
1	Me	2-Naphthyl	S- <i>sec</i> -Bu	4.23	37	Me	2,6-Di-Cl-Ph	S- <i>tert</i> -Bu	5.96
2	H	1-Naphthyl	S-Cyclopentyl	4.31	38	H	2,6-Di-F-Ph	S-Me	6.10
3	Me	1-Naphthyl	S-Cyclopentyl	4.35	39	Me	2-Cl-Ph	S- <i>sec</i> -Bu	6.10
4	Me	4-F-Ph	S- <i>sec</i> -Bu	4.59	40	Me	2-F-Ph	S- <i>sec</i> -Bu	6.10
5	Me	4-Cl-Ph	S- <i>sec</i> -Bu	4.77	41	Me	3-NO ₂ -Ph	S- <i>sec</i> -Bu	6.10
6	H	1-Naphthyl	S- <i>sec</i> -Bu	4.79	42	H	2-F-Ph	S- <i>sec</i> -Bu	6.22
7	H	2-Naphthyl	S- <i>sec</i> -Bu	4.83	43	H	3-NO ₂ -Ph	S- <i>sec</i> -Bu	6.22
8	H	4-F-Ph	S- <i>sec</i> -Bu	4.83	44	H	2,6-Di-Cl-Ph	S- <i>tert</i> -Bu	6.22
9	H	4-Cl-Ph	S- <i>sec</i> -Bu	5.02	45	H	2,6-Di-Cl-Ph	S- <i>n</i> -Bu	6.30
10	H	Ph	S- <i>tert</i> -Bu	5.07	46	H	2,6-Di-Cl-Ph	S-Cyclopentyl	6.40
11	H	3-Me-Ph	S- <i>tert</i> -Bu	5.09	47	H	2,6-Di-F-Ph	S- <i>n</i> -Bu	6.70
12	Me	3-Me-Ph	S- <i>sec</i> -Bu	5.27	48	H	2,6-Di-F-Ph	S- <i>tert</i> -Bu	6.70
13	Me	2,6-Di-Cl-Ph	S-Cyclohexyl	5.31	49	H	2,6-Di-Cl-Ph	S- <i>sec</i> -Bu	6.70
14	Me	Ph	S-Me	5.31	50	Me	2,6-Di-Cl-Ph	S- <i>sec</i> -Bu	6.92
15	Me	Ph	S- <i>sec</i> -Bu	5.32	51	H	2,6-Di-F-Ph	S- <i>sec</i> -Bu	7.00
16	Me	3-Me-Ph	S- <i>tert</i> -Bu	5.34	52	Me	2,6-Di-F-Ph	S- <i>sec</i> -Bu	7.00
17	Me	Ph	S-Cyclohexyl	5.37	53	H	2,6-Di-F-Ph	S-Cyclohexyl	7.05
18	H	3-Cl-Ph	S- <i>sec</i> -Bu	5.42	54	Me	2,6-Di-F-Ph	S- <i>tert</i> -Bu	7.05
19	Me	4-NO ₂ -Ph	S- <i>sec</i> -Bu	5.44	55	H	2,6-Di-F-Ph	S-Cyclopentyl	7.10
20	Me	3-Me-Ph	S-Cyclopentyl	5.47	56	Me	2,6-Di-F-Ph	S-Cyclopentyl	7.10
21	H	2-Cl-Ph	S- <i>sec</i> -Bu	5.49	57	H	2,6-Di-F-Ph	S- <i>iso</i> -Pr	7.30
22	Me	3-F-Ph	S- <i>sec</i> -Bu	5.52	58	Me	1-Naphthyl	S- <i>sec</i> -Bu	4.35
23	H	2,6-Di-Cl-Ph	S-Me	5.52	59	H	2-Naphthyl	S-Cyclohexyl	4.48
24	H	Ph	S-Cyclohexyl	5.52	60	H	Ph	S- <i>sec</i> -Bu	5.27
25	H	3-Me-Ph	S- <i>iso</i> -Pr	5.54	61	Me	Ph	S-Cyclopentyl	5.47
26	H	Ph	S-Cyclopentyl	5.55	62	H	3-Me-Ph	S-Cyclopentyl	5.59
27	H	3-Me-Ph	S-Cyclohexyl	5.59	63	Me	Ph	S- <i>iso</i> -Pr	5.60
28	Me	3-Me-Ph	S-Me	5.60	64	H	3-Me-Ph	S- <i>sec</i> -Bu	5.62
29	Me	3-Me-Ph	S- <i>iso</i> -Pr	5.60	65	Me	3-Cl-Ph	S- <i>sec</i> -Bu	5.74
30	H	4-NO ₂ -Ph	S- <i>sec</i> -Bu	5.62	66	H	3-F-Ph	S- <i>sec</i> -Bu	5.92
31	Me	3-Me-Ph	S-Cyclohexyl	5.66	67	H	2-NO ₂ -Ph	S- <i>sec</i> -Bu	6.22
32	Me	Ph	S- <i>tert</i> -Bu	5.72	68	H	2,6-Di-Cl-Ph	S-Cyclohexyl	6.40
33	Me	2,6-Di-Cl-Ph	S-Cyclopentyl	5.80	69	Me	2,6-Di-F-Ph	S-Me	6.70
34	H	2,6-Di-Cl-Ph	S- <i>iso</i> -Pr	5.89	70	Me	2,6-Di-F-Ph	S- <i>n</i> -Bu	7.05

TABLE I. Continued

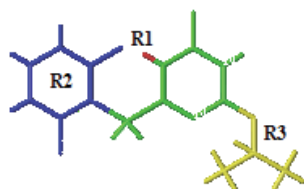
No.	R1	R2	R3	pIC ₅₀	No.	R1	R2	R3	pIC ₅₀
35	Me	2,6-Di-Cl-Ph	S- <i>iso</i> -Pr	5.94	71	Me	2,6-Di-F-Ph	S-Cyclohexyl	7.15
36	Me	2,6-Di-Cl-Ph	S- <i>n</i> -Bu	5.94	72	Me	2,6-Di-F-Ph	S- <i>iso</i> -Pr	7.30

Molecular structure construction

The 3D structures of these compounds were generated using Sketch Molecule of the SYBYL-X 2.0 package. All molecular structures used in both the training set and the test set were optimized using Tripos force field¹⁸ and Gasteiger–Hückel charges.¹⁹ The structural energy minimization was terminated using the Powell gradient algorithm with a convergence criterion of 0.005 kcal* mol⁻¹, reached after a maximum of 1000 iterations.²⁰ Generally, the lowest energy conformer of the most active compound was selected as a molecular template, which in this study was compound **57**.

Topomer CoMFA analysis

The Topomer CoMFA model was developed using SYBYL-X 2.0. The Topomer CoMFA is the second generation of CoMFA, and is a 3D-QSAR technique that automates the creation of models for predicting the biological activity or property of compounds.^{21,22} In Topomer CoMFA, the molecules are cut into two or more fragments, and all the fragments are aligned automatically,²³ the choice of splitting modes or the identification of the best R-group is the key step for the Topomer CoMFA. Steric and electrostatic interaction energies were calculated using the carbon sp³ probe. The Topomer CoMFA model was built by partial least square (PLS)²⁴ based on the optimal number of principal components (*N*) and evaluated by leave-one-out (LOO) cross-validation.²⁵ The bioactivity of the test set molecules were calculated by the model in order to evaluate the predictive ability of the model. In the process of Topomer CoMFA, the mode of cutting would affect the quality of the model. In this study, each of the training set structures was broken into three fragments shown as Ra (yellow), Rb (red) and Rc (blue) groups (Fig. 1).

Fig. 1. Cutting style of molecule **57**.

The partial least-squares (PLS) statistical method, which is an extension of multiple regression analysis, was used to build the Topomer CoMFA model. The leave-one-out (LOO) cross-validation approach was used to acquire the optimal number of components (*NC*), the correlation coefficient (*q*²), the non-cross-validation correlation coefficient (*r*²), the standard error of estimate (*SEE*), and the Fischer ratio value (*F*). The 3D contour maps are represented by field contribution maps using the StDev*Coeff field type. *Q*_{ext}² indicates external validation statistics of the model. *q*², *r*² and *Q*_{ext}² are calculated according to the formulas:²⁶

$$q^2 = 1 - \frac{\sum (Y_{\text{obs}} - Y_{\text{pred}})^2}{\sum (Y_{\text{obs}} - Y_{\text{mean}})^2}$$

* 1 kcal = 4184 J

$$r^2 = 1 - \frac{\sum (Y_{\text{obs}} - Y_{\text{CVpred}})^2}{\sum (Y_{\text{obs}} - Y_{\text{mean}})^2}$$

$$Q_{\text{ext}}^2 = 1 - \frac{\sum_{i=1}^{\text{test}} (Y_i - Y_{\text{test}})^2}{\sum_{i=1}^{\text{test}} (Y_i - Y_{\text{trave}})^2}$$

where Y_{mean} is the average activity value of the entire data set, while Y_{obs} , Y_{pred} , Y_{CVpred} , Y_{test} and Y_{trave} represent observed values, predicted values, cross-validated activity values, test predicted activity values and training predicted activity values, respectively. Often, high q^2 , Q_{ext}^2 and r^2 values are considered as strong evidence of high predictive ability of the model.

Molecular virtual screening

Topomer Search is a fast 3D ligand-based virtual screening tool that can search large libraries of compounds for fragments that are similar to the chemical structures of known lead compounds. In addition to pharmacophoric properties, Topomer Search uses topomeric fields to compare molecules, and it allows screening for whole molecules, R groups, or scaffolds, using topomer-based similarity.²⁷ The Topomer distance was used to estimate the similarity between the query fragment and molecular fragments being screened after 3-D reassembling, with lower values demonstrating greater similarity. This principle is explained as follows: The molecules in the database are incised into fragments, which are compared with the topomer similarity of the R groups of the training molecules. Then, the Topomer CoMFA model is used to predict their contributions to activity. Finally, a series of R groups is obtained. In this paper, compound **57** with the highest activity was chosen as the basic scaffold, and R1, R2 and R3 groups acted as the query to search in the ZINC (2012) database (130,000 compounds) for similar fragments. Topomer distance (TOPDIST) was set as 185²⁸ to evaluate the binding degree, and other parameters were defaulted by SYBYL-X 2.0.

Molecular docking

Research into the mechanism of the interaction between an enzyme and ligands is properly meaningful for the design of new drugs. Molecular docking technology can effectively explore the interaction between an enzyme and ligands.²⁹ Docking and scoring technology is applied to drug discovery for predicting the biological activity. Generally, van der Waals forces, hydrogen bonds, hydrophobic interactions, and electrostatic interactions are regarded as the primary factors in a docking study. The molecular docking studies in this work were performed using Surflex-Dock of SYBYL-X 2.0. Docking and scoring technology was applied to drug discovery for predicting the biological activity. The crystal structure (PDB code: 1QBT)³⁰ was taken from the RCSB Protein Data Bank.³¹ The 1QBT was prepared by adding hydrogen and charges, treating the terminal residues and extracting the ligand. Next, the 3D molecular model of the biomacromolecule was generated. All the ligands were prepared in accordance with the method used for the training molecules. The number of the maximum output poses was set as 20, and other parameters (additional starting conformations per molecule, angstroms to expand search grid and maximum number of rotatable bonds per molecule) were set, and the SYBYL-X 2.0 default values were considered for calculation. The output poses were evaluated by scoring functions, including total score, G-score, D-score, Chem-score, PMF-score and C-score³² (consensus score) which reflect the scoring consistency of the five other scores. Generally, the higher the total score value is, the greater is the selectivity of the output pose.

RESULTS AND DISCUSSION

Topomer CoMFA statistical results

The statistical results obtained from the Topomer CoMFA models are summarized in Table II. The multiple correlation coefficient of fitting (r^2), cross-validation (q^2) and external validation (Q_{ext}^2) are 0.899, 0.788 and 0.942, respectively. Meanwhile, the standard estimated error (SEE) is 0.271 and the F value is 66.584. When q^2 is larger than 0.5 and r^2 is greater than 0.6, it is generally believed that the model has statistical significance.³³ The biological activity predicted by the Topomer CoMFA model for all the molecules are shown in Table III. The linear regression between the experimental pIC_{50} and the predicted pIC_{50} , for both the training set and the test set is shown in Fig. 2, where the values are distributed near the 45° diagonal. This is further evidence of the strength of the results stemming from the Topomer CoMFA model.

TABLE II. The statistical results of Topomer CoMFA; N – optimal number of components; q^2 – the multiple correlation coefficient of cross validation; r^2 – the multiple correlation coefficient of fitting; SEE – standard estimated error; F – F test value; Q_{ext}^2 – the multiple correlation coefficient of external validation

Field	Parameter					
	N	q^2	r^2	SEE	F	Q_{ext}^2
SE	6	0.788	0.899	0.271	66.584	0.942

TABLE III. Experimental and predicted pIC_{50} values and corresponding residual values for all compounds

No.	pIC_{50} Exp, M	pIC_{50} Pred, M	Residual, M	No.	pIC_{50} Exp, M	pIC_{50} Pred, M	Residual, M
1	4.23	4.49	0.26	37	5.96	5.97	0.01
2	4.31	4.43	0.12	38	6.10	6.59	0.49
3	4.35	4.41	0.06	39	6.10	5.85	-0.25
4	4.59	4.53	-0.06	40	6.10	5.96	-0.14
5	4.77	5.03	0.26	41	6.10	6.13	0.03
6	4.79	4.61	-0.15	42	6.22	6.20	-0.02
7	4.83	4.51	-0.32	43	6.22	6.16	-0.06
8	4.83	4.77	-0.06	44	6.22	5.99	-0.23
9	5.02	5.05	0.03	45	6.30	6.01	-0.29
10	5.07	5.36	0.29	46	6.40	6.15	-0.25
11	5.09	5.47	0.38	47	6.70	6.85	0.15
12	5.27	5.78	0.51	48	6.70	6.83	0.13
13	5.31	5.98	0.67	49	6.70	6.33	-0.37
14	5.31	5.10	-0.21	50	6.92	6.31	-0.61
15	5.32	5.68	0.36	51	7.00	7.17	0.17
16	5.34	5.44	0.10	52	7.00	7.14	0.14
17	5.37	5.35	-0.02	53	7.05	6.84	-0.19
18	5.42	5.59	0.17	54	7.05	6.80	-0.25
19	5.44	5.34	-0.10	55	7.10	6.99	-0.11

TABLE III. Continued

No.	pIC_{50} Exp, M	pIC_{50} Pred, M	Residual, M	No.	pIC_{50} Exp, M	pIC_{50} Pred, M	Residual, M
20	5.47	5.60	0.13	56	7.10	6.96	-0.14
21	5.49	5.88	0.31	57	7.30	6.94	-0.36
22	5.52	5.48	-0.04	58	4.35	4.59	0.24
23	5.52	5.75	0.23	59	4.48	4.19	-0.29
24	5.52	5.38	-0.14	60	5.27	5.70	0.43
25	5.54	5.58	0.04	61	5.47	5.50	0.03
26	5.55	5.52	-0.03	62	5.59	5.62	0.03
27	5.59	5.48	-0.11	63	5.60	5.45	-0.15
28	5.60	5.20	-0.4	64	5.62	5.81	0.19
29	5.60	5.55	-0.05	65	5.74	5.56	-0.18
30	5.62	5.36	-0.26	66	5.92	5.51	-0.41
31	5.66	5.45	-0.21	67	6.22	5.74	-0.48
32	5.72	5.34	-0.38	68	6.40	6.01	-0.39
33	5.80	6.13	0.33	69	6.70	6.56	-0.14
34	5.89	6.10	0.21	70	7.05	6.82	-0.23
35	5.94	6.08	0.14	71	7.15	6.82	-0.33
36	5.94	5.99	0.05	72	7.30	6.91	-0.39

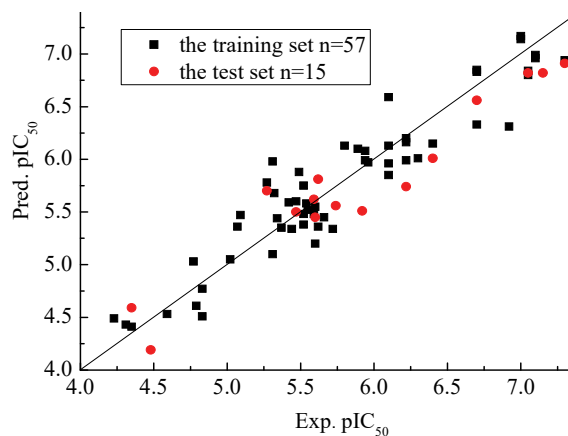


Fig. 2. Linear regression between the experimental and predicted pIC_{50} values for the 72 inhibitors.

Contour maps of the Topomer CoMFA model

The 3D-QSAR contour maps of the Topomer CoMFA model are shown in Fig. 3a–f with compound **57** as the template structure. The Topomer CoMFA results were graphically interpreted by field contribution maps using the “Compound Filed*Coeff” type of field. Through different colors, the contour maps display information about various factors that increase or decrease the molecular bioactivity generated by steric and electrostatic fields effects. The steric interactions of the R1, R2 and R3 groups are represented by green and yellow contours

in Fig. 3a–c, while the electrostatic interactions of the R1, R2 and R3 groups are denoted by red and blue contours in Fig. 3d–f. The green contours represent regions where a large or bulky substituent results in higher activity. The contrary is true for the yellow contours. The red isopleths indicate regions where a negatively charged substituent results in higher activity and the blue isopleths indicate regions where a positively charged substituent results in higher activity.

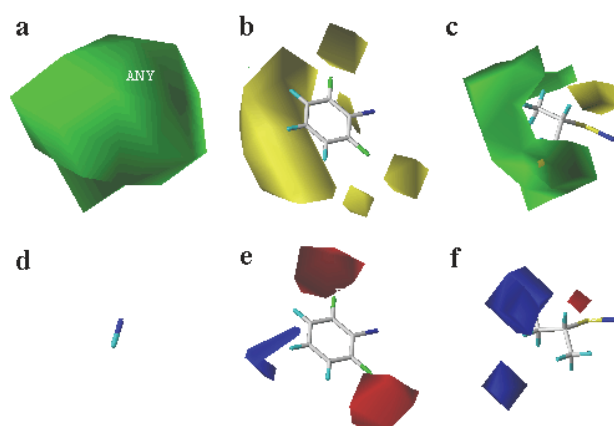


Fig. 3. 3D contours generated by the Topomer CoMFA model: a) steric field map of R1; b) steric field map of R2; c) steric field map of R3; d) electrostatic field map of R1; e) electrostatic field map of R2; f) electrostatic field map of R3 (Green and yellow contours represent steric favorable and unfavorable regions, respectively. Blue and red contours represent regions that favor electropositive and electronegative groups, respectively).

As shown in Fig. 3a and d, the big green contour at the R1 position indicates that a bulky substituent at this position would increase the activity. In contrast, the electrostatic fields at the R1 position has no effect on activity, which could be seen from the activities of compounds **16** (R1=Me) > **11** (R1=H), **39** (R1=Me) > **21** (R1=H) and **65** (R1=Me) > **18** (R1=H). In Fig. 3b, there is a large yellow contour around the benzene ring of R2 group, indicating that an R2 substituent of greater volume decreases the activity. For example, the inhibitor activity of **15** (R2 = Ph) > **1** (R2 = 2-naphthyl), **20** (R2 = 3-Me-Ph) > **3** (R2 = 1-naphthyl), and **68** (R2 = 2,6-di-Cl-Ph) > **59** (R2 = 2-naphthyl). In Fig. 3e, the large red contour and a small blue contour at the R2 position indicate that introduction of electropositive substitutions at this position decreases the activity. For example, the biological activity of **22** (R2 = 3-F-Ph) > **15** (R2 = Ph), **49** (R2 = 2,6-di-Cl-Ph) > **21** (R2 = 2-Cl-Ph) and **72** (R2 = 2,6-di-F-Ph) > **35** (R2 = 2,6-di-Cl-Ph). In Fig. 3c and f, a large green contour, a small yellow contour, a big blue contour and a small red contour at the R3 position, indicate that a large electropositive substituent at this position could increase the activity. For example, the biological

activity of **31** (R3 = *S*-cyclohexyl) > **20** (R3 = *S*-cyclopentyl), **50** (R3 = *S*-*sec*-Bu) > **36** (R3 = *S*-*n*-Bu) and **16** (R3 = *S*-*tert*-Bu) > **12** (R3 = *S*-*sec*-Bu).

Molecular screening and molecular design

The results of molecular screening using the Topomer Search technology were further evaluated through the use of Topomer distance (TOPDIST) and the contribution values of R-groups (TOPCOMFA_R). In general, an R group with higher priority contribution value was used to replace the R group of the template molecule in the same limit of the TOPDIST. In this study, the most active compound, **57**, was chosen as the basic scaffold. Five thousand R1, R2 and R3 groups were screened from the drug-like molecules in ZINC database using similar fragments. Finally, two R2 and seven R3 groups were found that had higher TOPCOMFA_R contributions than the R-groups of the template molecule.

In this work, the two R2 groups and seven R3 groups with higher contribution values were employed to substitute alternately for the R2 and R3 of the molecular template. Then 14 new molecules were designed. All of the compounds were optimized using the method that had been applied to the training molecules. The Topomer CoMFA model was then used to predict the activities of these new structures. The 14 new compounds and their predicted activities are given in Table IV. It can be seen from the data that there are 12 new compounds with activities that are higher than that of the template molecule. Close examination of the data in Table IV shows that the 12 new compounds have higher activities primarily because of the small volume and electronegativity of substituent at the R2 position of the molecules. Moreover, a bulky substituent in R3 contributes to the activity of the 14 new compounds. This observation is consistent with the analysis of the 3D contour of the Topomer CoMFA model. The result that the activities of compounds **2** and **4** are lower than that of the template molecule **57** may be due to one or both of the following causes: a) the mismatch between the R2 and R3 groups and b) unfavorable interactions in the molecule.

TABLE IV. Structures and predicted pIC₅₀ values of the newly designed molecules

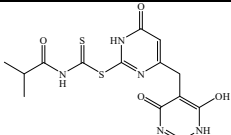
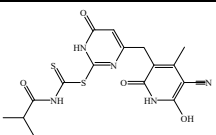
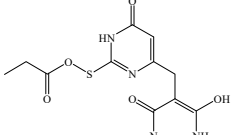
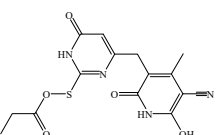
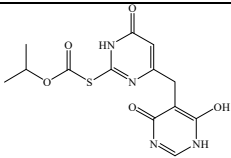
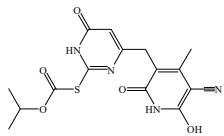
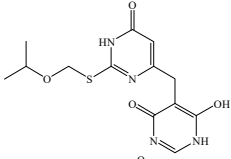
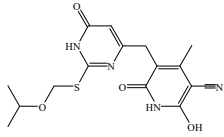
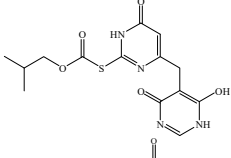
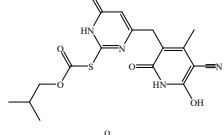
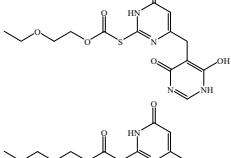
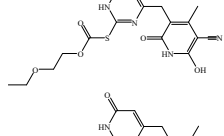
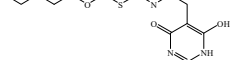

No.	Structure	Predicted pIC ₅₀	No.	Structure	Predicted pIC ₅₀
1-1^a		7.36	2-1^a		7.48
1-2		7.26	2-2^a		7.39

TABLE IV. Continued

No.	Structure	Predicted pIC ₅₀	No.	Structure	Predicted pIC ₅₀
1-3 ^a		7.35	2-3 ^a		7.48
1-4		7.21	2-4 ^a		7.33
1-5 ^a		7.34	2-5 ^a		7.47
1-6 ^a		7.32	2-6 ^a		7.45
1-7 ^a		7.35	2-7 ^a		7.47

^aCompounds with higher activity than that of the template molecule

Docking study

To validate the docking reliability, the crystal structure of protein (PDB code: 1QBT) with the cognate ligand was redocked. As the reference ligand, the cognate ligand was taken out of its protein–ligand complex (1QBT) and redocked in its binding site. As can be seen from Fig. 4a, the redocked ligand and the reference ligand can be almost completely superimposed. Their rotational positions are very similar. The result shows that the docking method is rational and reliable. The prototype molecule generated in this docking study is displayed in Fig. 4b. In Fig. 5, the ligand is represented by sticks, the amino acid residues by lines, and the hydrogen bonds by red dotted lines, and it reveals that the key residues Asp25, Asp30, Gly48, and Ile50 in chain A and Asp25, Asp30, Gly48 and Ile50 in chain B interact with the inhibitor through hydrogen bonds.

Considering the location of the cognate ligand as the binding site, the training set molecules and new designed molecules were docked into the receptor. The higher the total score, the better the selectivity of the output pose in the same C-score. Therefore, the best output pose of each molecule with full marks in the C-score was selected. The pose of template molecule **57** was chosen to explain

the binding mode between the protein receptor and inhibitors, and the docking results are shown in Fig. 6a. As can be seen from Fig. 6a, the template molecule forms hydrogen-bonding interactions with Ala28, Asp29, Gly49 and Ile50 in chain A of the protein receptor and with Gly49 and Ile50 in the chain B of the protein receptor. The total score is 7.0064, the crash score is -0.6330 and the polar score is 1.8215.

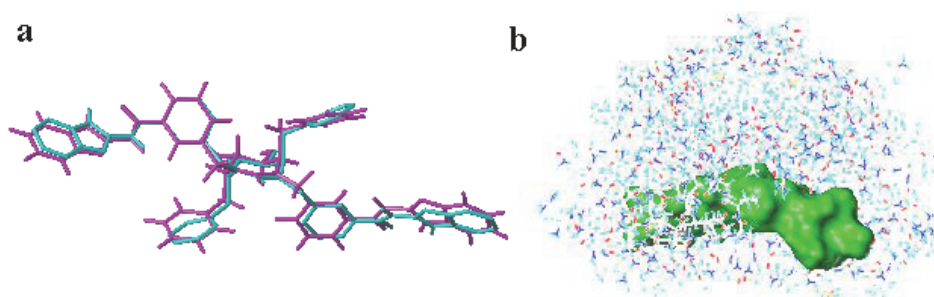


Fig. 4. a) Superimposition of the redocked reference ligand (the blue stick model) and the reference ligand (the purple stick model); b) the protomol (the green region represents the prototype molecule).

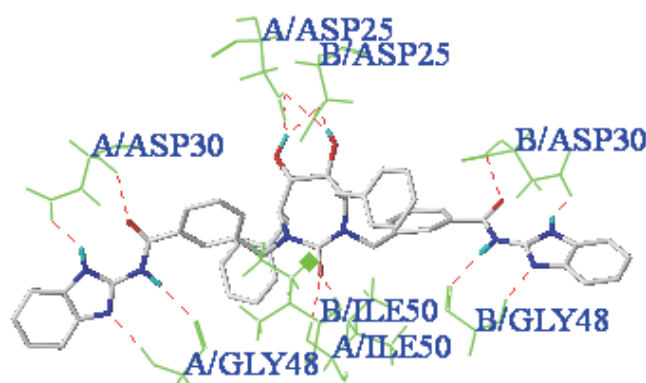


Fig. 5. The hydrogen-bond interactions (the ligand is represented by sticks, the amino acid residues by lines and the hydrogen bonds dotted lines).

In order to illustrate further the binding relationship between the ligands and the protein receptor, the new designed molecules were also subjected to the docking study. The total scores are given in Table V. The docking results for molecules **1-4**, **1-6** and **2-6**, as examples for the newly designed inhibitors, are analyzed in Fig. 6b–d. In Fig. 6b, compound **1-4** has hydrogen-bonding interactions with Asp30, Thr31, Gly49 and Ile50 in chain A, and Arg8, Gly49 and Ile50 in chain B. The total score, crash and polar are 8.1039, -0.8966 , 4.0317, respectively. Seven active sites (Asp25 in chain A, and Asp25, Ala28, Asp29, Ile47, Gly48

and Gly49 in chain B) of the protein receptor have hydrogen-bonding interactions with the ligand (compound **1-6**) in Fig. 6c. The total score, crash and polar are 8.5314, -1.1183, 6.5322, respectively. As shown in Fig. 6d, hydrogen-bonding interactions between the ligand (compound **2-6**) and the protein receptor are generated between Asp25 in chain A, and Asp25, Ala28, Asp29 and Asp30 in chain B. The total score, crash and polar are 8.8812, -1.3272, 4.9789, respectively.

TABLE V. Total-score and predicted pIC_{50} of newly designed molecules

No.	Pred.	Total score	No.	Pred.	Total score
1-1	7.36	6.8876	2-1	7.48	6.1404
1-2	7.26	7.1161	2-2	7.39	7.5404
1-3	7.35	7.6069	2-3	7.48	6.7080
1-4	7.21	8.1039	2-4	7.33	6.7009
1-5	7.34	7.8423	2-5	7.47	8.0498
1-6	7.32	8.5314	2-6	7.45	8.8812
1-7	7.35	7.9667	2-7	7.47	8.1133

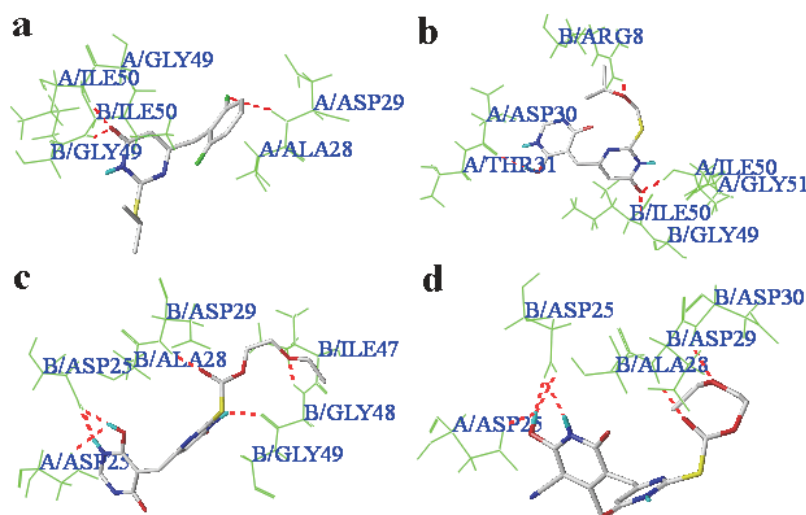


Fig. 6. a) The hydrogen-bond interaction between 1QBT and compound: a) **57**, b) **1-4**, c) **1-6** and d) **2-6** (the red dotted lines represent hydrogen bonding).

The results indicate that generally ligands would form hydrogen-bonding interactions with Ala28, Asp29, Gly49 and Ile50 of the protein receptor.

CONCLUSIONS

A series of S-DABO derivatives as non-nucleoside reverse transcriptase inhibitors were studied by computer-aided drug design processes. The Topomer CoMFA 3D-QSAR method generated a model with good internal and external prediction capability from a training set of 57 S-DABO derivatives. A test set of

15 molecules was employed to validate the external predictive ability of the model. The models can be extrapolated to predict novel and more potent inhibitors, and the contour maps obtained from Topomer CoMFA analyses can provide a useful insight for structure-based design for designing new chemical entities with high HIV-1 inhibitory activity. Then, 14 new HIV-1 inhibitors were designed adopting the Topomer search technology, of which 12 had higher activities than that of the template molecule. For a better understanding of the binding modes of inhibitors at the active site of HIV-1 protein, molecular docking analyses of representative compounds were performed. The docking results indicate that the ligands would generally form hydrogen-bonding interactions with the residues Ala28, Asp29, Gly49 and Ile50 of the protein receptor. This study could serve as a basis for the development of HIV-1 NNRTIs, and provide a theoretical reference for the synthesis of new drugs.

Acknowledgements. This work was supported by the National Natural Science Funds of China [21475081], the Natural Science Foundation of Shaanxi Province of China [2015JM2057] and the Graduate Innovation Fund of Shaanxi University of Science and Technology. We are thankful to Dr Donald G. Barnes for carefully reviewing this manuscript.

ИЗВОД

СТУДИЈА МОЛЕКУЛСКИМ МОДЕЛОВАЊЕМ ИНХИБИТОРА НЕНУКЛЕОЗИДНЕ РЕВЕРСНЕ ТРАНСКРИПТАЗЕ ЗА HIV-1 КОРИСТЕЊИ 3D-QSAR, ВИРТУЕЛНИ СКРИНИНГ И СИМУЛАЦИЈЕ ДОКИНГОМ

JIAN-BO TONG, SHANG-SHANG QIN, SHAN LEI и YANG WANG

Shaanxi Key Laboratory of Chemical Additives for Industry, Shaanxi University of Science and Technology, Xi'an 710021, P. R. China

Синдром стечене имунодефицијенције (сида) је значајна претња здрављу људи у свету, зато је студија дизајна анти-HIV лекова постала значајан задатак савременог друштва. У овом раду је проведена студија тродимензионалног односа структуре и активности (3D-QSAR) на 72 HIV-1 инхибитора нуклеозидне реверсне транскриптазе (NNRTIs) користећи Торомер упоредну анализу молекулског поља (Topomer CoMFA). Коефицијенти фитовања вишеструке корелације, унакрсне и спољашње евалуације су били 0,899, 0,788, односно 0,942. Резултати указују да добијени модел има и повољну стабилност у процењивању и добру способност предвиђања. Topomer Search је употребљен за тражење одговарајућих R група из ZINC базе података, одакле је дизајнирано 14, а за 12 од нових једињења је предсказано да би била активнија од темплатног молекула. Ови резултати снажно сугеришу да је Topomer Search ефикасан у скринингу и може бити користан водич дизајну нових HIV-1 лекова. Лиганди темплатног молекула и новосинте- тисана једињења су искоришћени за молекулски докинг у циљу проучавања интеракција ових једињења са протеинским рецептором. Резултати показују да ће генерално лиганди формирати водонично-везивне интеракције са остацима Ala28, Asp29, Gly49 и Ile50 протеина, чиме се добија додатни увид за дизајн још ефикаснијих лекова.

(Примљено 4. септембра, ревидирано 15. октобра, прихваћено 15. новембра 2018)

REFERENCES

1. F. Sterpone, P. Derreumaux, S. Melchionna S, *J. Phys. Chem. B* **122** (2018) 2544

2. Y. Y. Sun, J. F. Li, F. Q. Zhou, J. L. Li, B. Yin, *Phys. Chem. Chem. Phys.* **18** (2016) 12964
3. J. D. Reeves, R. W. Doms, *J. Gen. Virol.* **83** (2002) 1253
4. M. E. S. Soliman, *Drug. Dev. Res.* **74** (2013) 283
5. T. A. Ayele, A. Worku, Y. Kebede, K. Alemu, A. Kasim, Z. Shkedy, *Syst. Rev.* **6** (2017) 173
6. J. P. Moore, S. G. Kitchen, P. Pugach, J. A. Zack, *AIDS Res. Hum. Retroviruses* **20** (2004) 111
7. L. A. Kohlstaedt, J. Wang, J. M. Friedman, P. A. Rice, T. A. Steitz, *Science* **256** (1992) 1783
8. D. Jayaweera, P. Dilanchian, *Expert Opin. Pharmacother.* **13** (2012) 2601
9. H. H. Lu, P. Xue, Y. Y. Zhu, X. L. Ju, X. J. Zheng, *Bioorg. Med. Chem.* **25** (2017) 2491
10. P. Zhan, X. Chen, D. Li, Z. Fang, E. D. Clercq, *Med. Res. Rev.* **33** (2013) 7
11. R. P. Bhole, K. P. Bhusari, *Arch. Pharm.* **344** (2011) 119
12. P. G. Baraldi, P. A. Borea, M. Bergonzoni, B. Cacciari, E. Ongini, M. Racanatini, G. Spalluto, *Drug Dev. Res.* **46** (2015) 126
13. S. Yu, J. Yuan, J. Shi, X. Ruan, T. Zhang, Y. Wang, W. Du, *Chemom. Intell. Lab. Syst.* **146** (2015) 34
14. J. B. Tong, P. Zhan, X. S. Wang, Y. J. Wu, *J. Chemom.* **31** (2017) 2934
15. J. B. Tong, M. Bai, X. Zhao, *Med. Chem. Res.* **25** (2016) 2619
16. M. A. D. Brito, C. R. Rodriguez, J. J. Cirino, R. B. D. Alencastro, H. C. Castro, *J. Chem. Inf. Model.* **48** (2008) 1706
17. A. Gangjee, O. Adair, S. F. Queener, *J. Med. Chem.* **42** (1999) 2447
18. C. W. Swalina, R. J. Zauhar, M. J. Degrazia, G. Moyna, *J. Biomol. NMR* **21** (2001) 49
19. J. Liu, F. Wang, Z. Ma, X. Wang, Y. Wang, *Int. J. Mol. Sci.* **12** (2011) 946
20. L. Wu, Y. Wang, Y. Liu, S. Yu, H. Xie, *Oncotarget* **5** (2014) 7677
21. R. D. Cramer, P. Cruz, G. Stahl, W. C. Curtiss, B. Campbell, B. B. Masek, F. Soltanshahi, *J. Chem. Inf. Model.* **48** (2009) 2180
22. R. J. Jilek, R. D. Cramer, *J. Chem. Inf. Comput. Sci.* **44** (2004) 1221
23. A. Golbraikh, A. Tropsha, *J. Mol. Graphics Modell.* **20** (2002) 269
24. R. D. Cramer, R. D. Clark, D. E. Patterson, A. M. Ferguson, *J. Med. Chem.* **39** (1996) 3060
25. Z. Shao, J. E. Meng, *Neurocomputing* **173** (2016) 778
26. J. B. Bhonsle, V. Divakaramenon, D. P. Huddler, A. J. Magill, R. P. Hicks, *J. Med. Chem.* **50** (2008) 6545
27. J. B. Tong, M. Bai, X. Zhao, *J. Serb. Chem. Soc.* **81** (2016) 3
28. Y. Xiang, J. Song, Z. Zhang, *Comb. Chem. High Throughput Screening* **17** (2014) 458
29. U. Singh, R. P. Gangwal, G. V. Dhoke, R. Prajapati, M. Damre, *Arabian J. Chem.* **10** (2012) S617
30. P. K. Jadhav, P. Ala, F. J. Woerner, C. H. Chang, S. S. Garber, E. D. Anton, *J. Med. Chem.* **40** (1997) 181
31. P. W. Rose, C. Bi, W. F. Bluhm, C. H. Christie, D. Dimitropoulos, S. Dutta, R. K. Green, D. S. Goodsell, A. Prlic, M. Quesada, G. B. Quinn, A. G. Ramos, J. D. Westbrook, J. Young, C. Zardecki, H. M. Berman, P. E. Bourne, *Nucleic Acids Res.* **41** (2013) D475
32. R. D. Clark, A. Strizhev, *J. Mol. Graphics Modell.* **20** (2002) 281
33. J. Tong, P. Zhan, M. Bai, T. Yao, *J. Chemom.* **30** (2016) 523.