3D-QSAR and docking studies of HIV-1 Integrase inhibitors using R-group search and Surflex-dock

JIAN-BO TONG*, MIN BAI and XIANG ZHAO

College of Chemistry and Chemical Engineering, Shaanxi University of Science & Technology, Xi'an 710021, PR China

*Correspondence author: E-mail (jianbotong@aliyun.com)

***Abstract*:** In this paper, a three-dimensional quantitative structure-activity relationship (3D-QSAR) study for 62 HIV-1 integrase inhibitors was established using Topomer CoMFA. The multiple correlation coefficient of fitting, cross validation and external validation were 0.942, 0.670 and 0.748, respectively. The results indicated that the Topomer CoMFA model obtained has both favorable estimation stability and good prediction capability. Topomer Search was used to search R group from ZINC database. As the result, a series of R groups with relatively high activity contribution was obtained. By No.42 molecule filtering, 1 Ra groups and 21 Rb groups were selected. We employed the 1 Ra groups and 21 Rb groups to alternately substitutes for the Ra and Rb of sample 42. Finally, we designed 21 new compounds and further predicted their activities using the Topomer CoMFA model and there were 10 new compounds with higher activity than that of the template molecule. The results suggested the Topomer Search technology could be effectively used to screen and design new HIV-1 integrase inhibitors and has good predictive capability to guide the design of new HIV/AIDS drugs. Molecular docking elucidated the conformations of the compounds and key amino acid residues at the docking pocket of IN protein.

Keywords: quantitative structure-activity relationship(QSAR), integrase inhibitors, Topomer CoMFA, Topomer Search, molecular docking, new drug design

RUNNING TITLE: 3D-QSAR and docking studies

INTRODUCTION

Acquired immunodeficiency syndrome (AIDS) caused by the human

*Correspondence author. Fax: 86-29-86168312; Tel: 86-29-86168315;
E-mail address: jianbotong@aliyun.com(J.B. Tong)

immunodeficiency virus (HIV) has resulted in the deaths of about 30 million people since it was first reported in 1981.[1] Anti-HIV drug development is one of the leading tasks in the drug discovery area due to the improving rate of sufferers with HIV and related infections.[2] The host proteins involved in viral replication cycle have been used as drug targets to design inhibitors to prevent the spread of infection, such as reverse transcriptase, protease, integrase, polymerase, Glycoprotein(gp41 and gp120), as well as the host cell receptor(CD4) and coreceptor (CCR5 and CXCR4).[3] IN plays a pivotal role in the integration of the viral genome into the host genome enabling HIV to efficiently propagate in human CD4+ cells.[4] And it is an essential enzyme for the viral replication and has no mammalian counterparts, so IN is an attractive target for the development of anti-AIDS drugs.[5, 6] Human immunodeficiency virus-1 (HIV-1) is characterized by reverse transcription of the viral RNA genome to cDNA and its integration into the host cell genome. Then, the integrated proviral DNA with a long terminal repeat (LTR) at each end is transcribed, leading to synthesis of viral proteins and completion of the viral replication cycle. Drugs blocking HIV integration not only inhibit virus replication, but also enhance T cell survival . HIV-1 integrase (IN), a viral gene-encoded enzyme, catalyzes the integration, which proceeds by two spatially and temporally distinct steps, 30 processing and DNA strand transfer, in the context of the retroviral preintegration complex.[7, 8] HIV-1 IN is a 32 kDa polynucleotidyl transferase comprising three domains: the N-terminal domain, the C-terminal domain, and the catalytic domain. The catalytic domain contains a DDE motif (D64, D116, and E152) that forms metal chelating interactions with one or two divalent metal ions, such as Mn2+ and Mg2+. IN catalyzes the insertion of reverse transcribed viral DNA into the host cell's chromosomes in two steps: (a) 30-processing, the excision of two terminal nucleotides leaving 30-hydroxyl ends of the viral DNA, and (b) strand transfer, insertion of the 30-hydroxyl ends onto the host DNA by a nucleophilic addition. Currently, two other IN inhibitors (Elvitegravir and Dolutegravir)have been approved for clinical use.[9]

The availability of computational techniques on quantitative structure activity relationships (QSARs) might provide a potential direction for accelerating the drug

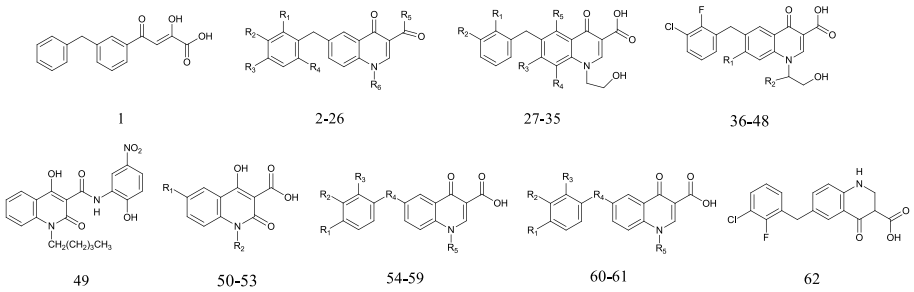
design process. In fact, QSAR can be viewed as a technique attempting to summarize chemical and biological information in a form that allows one to generate relationships between chemical structure and biological activity.[10] As is well known, the success of a QSAR study depends also on the selection of variables (molecular descriptors) and on the representation of the information. Variables should give the maximum of information in the activity variations. 3D-QSAR model would better reflect the interactions between the substrate and receptor compared to 2D-QSAR. Comparative molecular field analysis (CoMFA)[11] is the method used widely of 3D-QSAR. In this paper, Topomer CoMFA, [12, 13] the second generation of CoMFA was employed to construct the 3D-QSAR model for 62 HIV-1 integrase inhibitors to analyze the chemical-biological interactions governing their activities toward HIV-1 PR. The Topomer CoMFA model would be also applied to conduct ligand-based virtual screening combining the Topomer Search[14] technology to lay the foundation of new drug design.

PRINCIPLES AND METHODS

Data set

In this study, the structures and experimental data of the 62 HIV-1 integrase inhibitors obtained from the literature[15] are shown in Table 1. The dataset was systematically divided into the training set (45 compounds) and the test set (17 compounds). The number of test set compounds was approximately 30% that of the training set compounds, which was considered as a proper ratio. [16] The training set was applied to build the 3D-QSAR model and, for the test set, was used to verify the predictive ability of the model. The bioactivities of inhibitors were presented in pIC_{50}($-lgIC_{50}$). IC_{50} is the drug concentration inhibiting 50% of the cellular growth followed by 1 h of drug exposure.

Table 1 Structures and bioactivities of 62 integrase inhibitors



NO.	R1	R2	R3	R4	R5	R6	IC_{50}(μ M)	PIC_{50}	Pred.
1	__						0.05	7.3010	7.4712
2*	H	H	H	H	H	H	1.63	5.7878	5.7911
3	H	H	H	H	H	CH_3	2.30	5.6383	5.6374
4	H	Cl	H	H	OH	H	0.80	6.0969	6.2438
5	Cl	H	H	H	OH	H	0.41	6.3872	6.0375
6*	F	H	H	H	OH	H	0.50	6.3010	6.3172
7	Me	H	H	H	OH	H	1.08	5.9666	5.9750
8	OMe	H	H	H	OH	H	1.17	5.9318	5.8369
9	CF_3	H	H	H	OH	H	0.72	6.1427	5.8379
10	Cl	H	H	Cl	OH	H	0.37	6.4318	6.4454
11*	H	Cl	H	Cl	OH	H	0.25	6.6021	6.2876
12	Cl	Cl	H	H	OH	H	0.07	7.1549	7.0814
13	Cl	Cl	H	H	OH	Me	0.083	7.0809	7.1392
14*	Cl	Cl	H	H	OH	Et	0.031	7.5086	7.3118
15	Cl	Cl	H	H	OH	Pr	0.055	7.2596	7.3060
16	Cl	Cl	H	H	OH	iPr	0.026	7.5850	7.4435
17	Cl	Cl	H	H	OH	Bu	0.065	7.1871	7.0327
18*	Cl	Cl	H	H	OH	CH_2CO_2H	0.032	7.4949	7.4264
19	Cl	Cl	H	H	OH	$(CH_2)_2CO_2H$	0.038	7.4202	7.3982
20	Cl	Cl	H	H	OH	CH_2CONH_2	0.035	7.4559	7.4534
21*	Cl	Cl	H	H	OH	$(CH_2)_2CONH_2$	0.116	6.9355	7.2167
22	Cl	Cl	H	H	OH	$(CH_2)_2NH_2$	0.215	6.6676	7.2085
23	Cl	Cl	H	H	OH	$(CH_2)_2OH$	0.021	7.6778	7.4673
24	Cl	Cl	H	H	OH	$(CH_2)_3OH$	0.077	7.1135	7.2954
25*	Cl	F	H	H	OH	$(CH_2)_2OH$	0.044	7.3565	6.6186
26	F	Cl	H	H	OH	$(CH_2)_2OH$	0.024	7.6198	7.8180
27	Cl	Cl	H	H	F	H	0.084	7.0757	7.0352
28	Cl	Cl	F	H	H	__	0.025	7.6021	7.6280
29	Cl	Cl	H	F	H	__	0.034	7.4685	7.4825
30	Cl	Cl	OMe	H	H	__	0.012	7.9208	7.6462
31*	Cl	Cl	Cl	H	H	__	0.043	7.3665	7.4009
32	Cl	Cl	Me	H	H	__	0.041	7.3872	7.5342
33*	Cl	Cl	CF_3	H	H	__	0.674	6.1713	6.9872

34	Cl	Cl	CN	H	H	__	0.050	7.3101	7.4231
35	F	Cl	OMe	H	H	__	0.009	8.0458	7.9970
36	H	(S)-Me	__	__	__	__	0.0148	7.8297	7.9084
37	H	(R)-Me	__	__	__	__	0.0383	7.4168	7.9076
38*	H	(S)-Et	__	__	__	__	0.009	8.0458	7.9384
39	H	(S)-Pr	__	__	__	__	0.0082	8.0862	7.7193
40*	H	(S)-iPr	__	__	__	__	0.0082	8.0862	7.9061
41	H	(S)-tBu	__	__	__	__	0.006	8.2218	7.9751
42	H	(S)-cyclohexyl	__	__	__		0.0056	8.2518	7.9954
43	H	(S)-Ph	__	__	__	__	0.0098	8.0088	8.0302
44*	OMe	(S)-Pr	__	__	__	__	0.0058	8.2366	7.8725
45	OMe	(S)-iPr	__	__	__	__	0.0072	8.1427	8.0813
46	OMe	(R)-iPr	__	__	__	__	0.0144	7.8416	7.6121
47	OMe	(S)-tBu	__	__	__	__	0.0058	8.2366	8.1655
48*	OMe	(S)-cyclohexyl	__	__	__		0.0067	8.1739	8.1191
49	__						9	5.0458	5.1364
50	Bn	CH_3	__	__	__	__	6	5.2218	5.2892
51*	4-F-Bn	CH_3	__	__	__	__	0.9	6.0458	5.2451
52	OPh	CH_3	__	__	__	__	14	4.8539	4.7144
53*	4-F-Bn	$(CH_2)_4CH_4$	__	__	__	__	5	5.3010	4.8765
54	H	H	H	S	$(CH_2)_2OH$	__	18.5	4.7328	4.7834
55	Cl	H	Cl	CH_2	$(CH_2)_2OH$	__	0.2	6.6990	5.8825
56	Cl	H	Cl	CH_2	$(CH_2)_3OH$	__	1.3	5.8861	5.7093
57*	Cl	H	Cl	CH_2	$(CH_2)_4OH$	__	0.6	6.2218	5.9280
58	Cl	H	Cl	CH_2	$(CH_2)_2N$-$(CH_3)_2$		24.1	4.6180	5.1304
59	Cl	H	Cl	CH_2	$(CH_2)_2O$-CH_3	__	16.5	4.7825	5.4701
60	F	Cl	NH	__	__	__	2.1	5.6778	5.5582
61*	H	H	S	__	__	__	1.6	5.7959	6.7275
62	__						0.0435	7.3615	7.4321

*Chosen as the test set

Molecular structure construction

The 3D structures of 62 HIV-1 integrase inhibitors were constructed using the sketch molecule of Sybyl 2.0-X package. All molecules were optimized using tripos force field and gradient descent method with an energy charge of 0.005 kcal/mol. Partial charges for all the molecules were added using the Gasteiger-Hückel method. The maximum iteration coefficient was 1000. Other parameters were defaulted by Sybyl 2.0-X.

Topomer CoMFA modeling

Topomer CoMFA is a rapid fragment-based 3D-QSAR method to predict significant R-group of molecules.The Topomer CoMFA method identifies bioactivity values with the help of a compound library as a source with automated rules.[11] The process of standard Topomer CoMFA is completed by the following two steps: the first step is generating the Topomer 3D models for each fragment of the molecule. Topomer CoMFA divides one compound into two or more fragments. By confirming how to break compounds' structures, the Topomer CoMFA can identify the fragments' features and charges automatically.[17]The second step consists of performing CoMFA with partial least squares (PLS) of leave-one-out (LOO) cross-validation in order to form a predictive model. [18] During the process of building the model, the CoMFA method is used to deal with the large amounts of data. By objective measures and automatic matching to analyse compounds' characters, Topomer CoMFA is more efficient in forming predictive models.

In the process of Topomer CoMFA, the measure of fracture would affect the quality of the model. In This study, each of the training set structure was broken into two sets of fragments shown as Ra (blue) and Rb (red) groups as shown in Fig. 1. Initially, as molecule 42 had the highest activity, it was selected as the template molecule. Based on compound 42, the cutting style was confirmed. The molecule was cut to obtain the Ra group and Rb group. Other training molecules were identified automatically and cut in this style. The molecules not identified need to be cut manually. Then the steric and electrostatic field energy between the molecules was calculated. The descriptors obtained were considered as the independent variables and the pIC_{50} values were regarded as the dependent variables in partial least square (PLS)[19] to build the Topomer CoMFA model. The model was evaluated by leave-one-out-cross validation (LOO-CV) approach. The test molecules were predicted by the Topomer CoMFA model to verify the predictive ability of the model obtained.

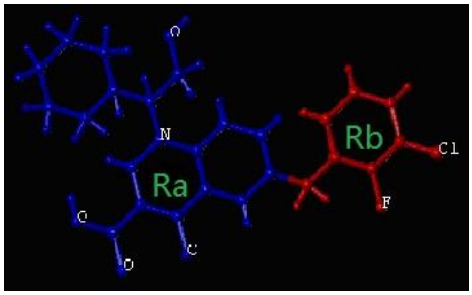


Fig. 1 Cutting style of sample 42

Molecular screening

Molecular screening was carried out using the Topomer Search technology, a fast 3D ligand-based virtual screening tool. The principle is explained as following: the molecules in the database are cut into fragments, which are compared with the Topomer similarity of R groups of training molecules. Then the Topomer CoMFA model are used to predict their contributions to activity. Finally, a series of R groups will be obtained. In this study, Topomer Search was employed to search R groups with relatively high activity contribution from drug-like in ZINC (2012) database (130,000 compounds). Topomer distance was set as 185 to evaluate the binding degree, and other parameters were defaulted by Sybyl2.0-X.

Molecular docking

Molecular docking studies were performed using Surflex-dock of Sybyl 2.0-X. Surflex-dock uses an empirical scoring function and a patented search engine to dock ligands into a protein's binding site. Surflex-dock is particularly successful at eliminating false positive results and can, therefore, be used to narrow down the screening pool significantly, while still retaining a large number of active compounds.

In this study, the protein-ligand complex with crystal structure (PDB ID: 3NF7) [20] of HIV-1 protease was taken from RCSB Protein Data Bank. 3NF7 was prepared by adding hydrogen, adding charges, treating the terminal residues and extracting the ligand. Then the prototype molecule was generated. All the ligands were prepared in accordance with the method used to training molecules. The number of the maximum output poses was set as 20 and other set parameters were defaulted by Sybyl 2.0-X. The output poses were evaluated by scoring functions including Total score, G-score, D-score, Chem-score, PMF-score and C-score (consensus score) which reflects the

scoring consistency of other five scores. Generally, the higher the C-score, the better the selectivity of the output pose.

RESULTS AND DISCUSSION

Topomer CoMFA modeling results and evaluation

To generate statistically significant 3D-QSAR models, we used the ligand-based alignment rule. In this study, the regression analysis was carried out using the partial least squares (PLS) method,[21, 22] some statistical parameters were used to analysis the stand or fall of these models, including the cross-validated coefficient (q^2), the standard deviation of error prediction (r^2), standard error of estimate (SEE) and F-statistic values, a high q^2 and r^2 value ($q^2 > 0.5$, $r^2 > 0.6$) is considered as a proof of high predictive ability of the model.[23] The statistical results of model in this study are displayed in Table 2. As can be seen from the table, the q^2 value of 0.751, an optimized component of 6 and r^2 value of 0.942, which suggested the model also has predictive ability ($q^2 > 0.2$). The pIC_{50} value of test set was predicted with the q_{pred}^2 value of 0.748. The linear regression between experimental pIC_{50} and predicted pIC_{50} for training set and test set are shown in Fig. 2. Table 1 shows the predicted bioactivities (pIC_{50}) for training set and test set. The results indicate that the model has both favorable estimation stability and good prediction capabilities.

Table 2 The statistical results of Topomer CoMFA

Statistical parameters[a]	*N*	r^2	q^2	$q_{pred}{}^2$	*SEE*	*SD*	SD_{CV}	*F*
Topomer CoMFA	6	0.942	0.670	0.748	0.277	0.28	0.67	103.344

[a]*N:* optimal components, r^2: The multiple correlation coefficient of fitting, q^2: The multiple correlation coefficient of cross validation, q^2_{pred}: The multiple correlation coefficient of external validation, *SEE*: standard estimated error, *SD*: fitting standard deviation, SD_{cv}: cross validation tandard deviation, F: Fisher value

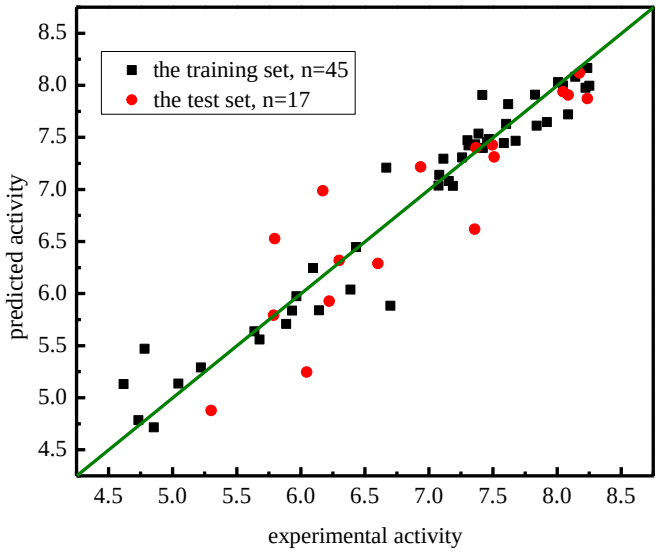


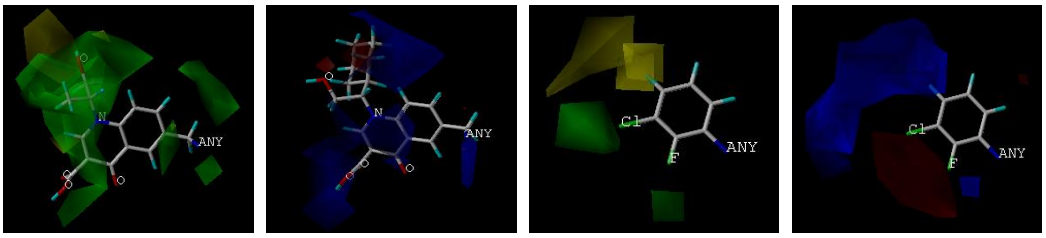
Fig. 2 Linear regression between experimental and predicted pIC_{50} of 62 inhibitors

3D contour plots of Topomer CoMFA model

The three-dimensional contour plots of the Topomer CoMFA model are shown in Fig. 3(a-d) with the sample 42 as the reference structure. The contour maps provide information on factors affecting the activities of the molecules. This is particularly important when increasing or reducing the activity of a compound by changing its molecular structural. The steric interaction of the Ra and Rb groups is represented by green and yellow contours in Fig. 3(a) and Fig. 3(c), respectively. While the electrostatic interaction of the Ra and Rb groups is denoted by red and blue contours in Fig. 3(b) and Fig. 3(d), respectively. The green contours represent regions where the large or bulky substituent is favorable for the activity. The opposite is true for the yellow contours. The red isopleths indicate regions where the negative charged substituent is favorable for the activity and the blue isopleths indicate regions where an increase of the positive charged substituent enhances the activity.

As shown in Fig. 3(a), a green contour covering the cyclohexyl group links to R_2 indicates the presence of a bulky group for good biological activity. This is in agreement with the experimental data: 38(-Et)>37(-Me), 41(-tBu)>40(-iPr). The molecule 42 has the highest activity because of the bulky substituent (-cyclohexyl) at R_2-position. Besides, a green contour near the R_2-position of the molecule 42

indicates the bulky substituent in this position may be favorable for the activity. For example, molecule 45(-OMe) has higher activitiy than molecule 45(-H). From Fig. 3(b), there is a large blue contour around cyclohexyl(R_2), which suggest that the positive charged substituent at R_2-position may favor the activity. This is in agreement with the experimental data: 39(-Pr), 40(-iPr), 41(-tBu), 42(-cyclohexyl). In Fig. 3(c) and Fig. 3(d), a yellow and a large blue contour at 4-position of the phenyl ring indicate that the small and positive charged substituent is preferred in this region. A red contour at 2, 3-position of the phenyl ring in Fig. 3(d), suggesting introduction of the electronegtive substituent into this position will be benefit for inhibitory activity. It can show the fact that the -Cl and -F have been introduced in this position.



(a) (b) (c) (d)

Fig. 3 3D contour of Topomer CoMFA model

(a)steric field map of Ra; (b)electrostatic field map of Ra;

(c)steric field map of Rb; (d)electrostatic field map of Rb

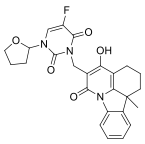
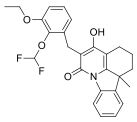
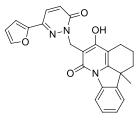
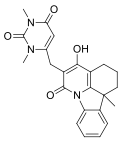
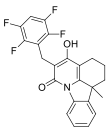
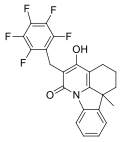
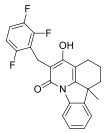
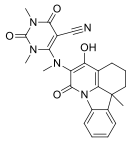
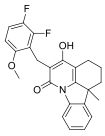
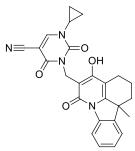
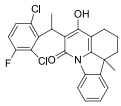
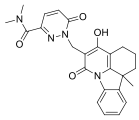
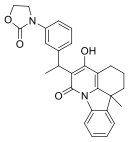
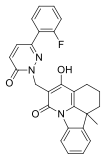
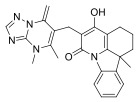
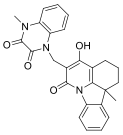
Molecular screening and molecular design

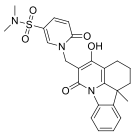
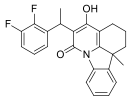
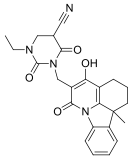
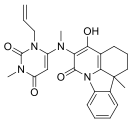
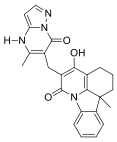
The results of molecular screening using Topomer Search technology are evaluated by the Topomer distance (TOPDIST) and the contribution values of R-groups(TOPCOMFA_R). Under normal circumstances, we give priority to TOPCOMFA_R in the same limit of the TOPDIST. In this study, 5000 Ra groups and 1000 Rb groups were screened from Drug-like in ZINC (2012) database. Eventually, 1 Ra groups and 21 Rb groups with higher TOPCOMFA_R than that of template molecule were selected.

We employed the 1 Ra group and 21 Rb groups to alternately substitutes for the Ra and Rb of sample 42 and designed 21 new molecules. All molecules were optimized using the method applied to the training molecules and further predicted their activities using the Topomer CoMFA model obtained. The structures and

predicted activities of 21 new compounds are displayed in Table 3. It can be seen from Table 3 that there are 10 new compounds with higher activity than that of the template molecule. And as revealed from Table 3, 10 new compounds have higher activities because of the introduction of the electronegtive substituent into 2, 3-position of the phenyl ring of Rb. Moreover, the bulky substituent in Ra make contributions for the activity of 10 new compounds . This is consistent with the analysis of the 3D contour of Topomer CoMFA model.

Table 3 Structures and predicted pIC_{50} of new designed molecules

NO.	structure	Pred.	NO.	structure	Pred.
1*		8.2652	12*		8.2761
2		8.0578	13		8.0433
3*		8.6310	14*		8.6773
4*		8.3685	15		8.1185
5*		8.3283	16		8.1452
6*		8.3148	17*		8.8139
7		7.8841	18		8.0917
8		7.7815	19*		8.4089

9*		8.4612	20		7.9982
10		8.1987	21		8.0043
11		8.1189			

* compounds with higher activity than that of the template molecule

Docking results

To validate the 3D-QSAR results, docking simulation was performed to study the binding environment. Here, the Surflex program (Sybyl2.0-X) was used to explore the probable binding conformation. In this study, the training inhibitors and the new-designed inhibitors were applied to perform the docking study with the IN receptor, respectively.

Fig. 4(a) and Fig. 4(b) Depict the docking results of molecule 42 and 24 in the training set. As can be seen from Fig. 4(a), compound 42 was docked into the binding cavity with the carboxyl directing towards the hydrophobic group of His78, Val79, Ser81, Asn144, Val150 and Tyr194. The molecule forms hydrogen bonding interactions with Gly82, His183, Lys188 and Arg199. The hydrogen bond distances observed are 1.9 Å (Gly82-O ··· H-O-), 2.5 Å (His183-N ··· H-O-), 1.9 Å (Lys188-HN-H···O-), and 2.5 Å (Arg199-HN-H···O-), respectively. As shown in Fig. 4(b), compound 24 was docked into the binding cavity with the carboxyl directing towards the hydrophobic group of His78, Val79, Ser81, Tyr83, Asn144, Val150 and Ser153. 5 hydrogen bonds were formed between compound 24 and IN receptor. The hydrogen bond distances observed are 2.0 Å (Gly82-O···H-O-), 2.2 Å (His183-N··· H-O-), 1.9 Å (Lys188-HN-H ··· O-), 2.5 Å (Arg199-HN-H ··· O-) and 2.5 Å (Arg199-HN-H···O-).

The docking results between new-designed molecule 17 and 14 and IN receptor

are displayed in Fig. 4(c) and Fig. 6(c). From Fig. 4(c), we can find compound 17 was docked into the binding cavity with the carboxyl directing towards the hydrophobic group of Val79, Ala80, Ser81, Gln146, Asn155 and Lys186. 7 hydrogen bonds were formed between compound 17 and IN receptor. The hydrogen bond distances observed are 2.0 Å (Gly82-O ··· H-O-), 2.3 Å (His183-N ··· H-O-) 2.4 Å (Lys188-HN-H···O-), 2.4 Å (Lys188-HN-H···O-), 2.4 Å (Lys188-HN-H···N-), 2.8 Å (Lys188-HN-H···N-) and 2.7 Å (Arg199-HN-H···O-). As shown in Fig. 4(d), compound 14 was docked into the binding cavity with the carboxyl directing towards the hydrophobic group of His78, Ala80, Ser81, Asn144, Gln146 and Lys188. 6 hydrogen bonds were formed between compound 14 and IN receptor. The hydrogen bond distances observed are 2.6 Å (His78-O···H-O-), 2.3 Å (Val79-O···H-O-), 2.4 Å (Lys188-HN-H···F-), 2.5 Å (Lys188-HN-H···F-), 2.6 Å (Lys188-HN-H···F-), 2.5 Å (Arg199-HN-H···F-) and 2.5 Å (Arg199-HN-H···F-).

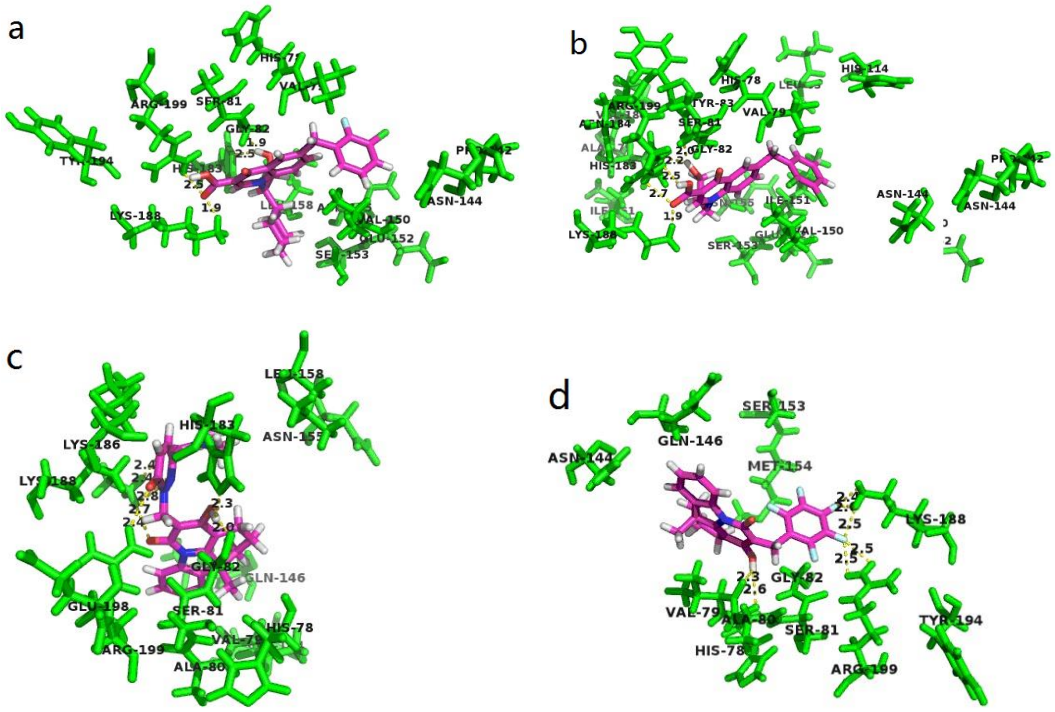


Fig. 4 Docking result of the compound 42 (a), 24 (b) 17(c) and 14 (d) into the binding site of HIV-1 integrase protease. Hydrogen bonds are shown as yellow lines, with distance unit of Å. The inhibitor and the important residues are shown as stick model.

CONCLUSIONS

In the present work, 62 HIV-1 integrase inhibitors were studied by computer-aided drug design processes, such as 3D-QSAR/Topomer CoMFA studies

and molecular docking simulations. The built models are favored by internal and external predictions and the stastics are convincing and comparable. The models can not only be extrapolated to predict novel and more potent inhibitors, but the contour maps obtained from Topomer CoMFA analyses provid a useful insight for structure-based design for designing new chemical entities with high HIV-1 inhibitory activity. For a better understanding of the binding modes of inhibitors at the active site of HIV-1 protein, molecular docking analyses of the representative compounds were performed. Some key residues such as His78, Val79, Ala80, Ser81, Val150, hydrophobic interactions, as well as hydrogen bonds (Gly82101, His183, Lys188, Arg199) between inhibitors and the active site were observed. This study could serve as a basis for the development of HIV-1 inhibitors.

***Acknowledgments*:** We gratefully acknowledge supports of this research by the Science and Technology Project of Science and Technology Department of Shaanxi(2011K07-13), the Specific Research Projects of Science and Technology Department of Shaanxi(12JK0629, 11JK0602), the Graduate Innovation Fund of Shaanxi University of Science & Technology.

REFERENCES

1. H. Sharma, T. W. Sanchez , N. Neamati , M. Detorio , R. F. Schinazi , X. L. Cheng , J. K. Buolamwini, *Bioorganic & Medicinal Chemistry Letters*. **23** (2013) 6146.

2. R. V. Patel, Y. S. Keum, S. W. Park. *European Journal of Medicinal Chemistry.* **xxx** (2014) 1.

3. J. P. Moore, S. G. Kitchen, P. Pugach, J. A. Zack, *AIDS. Res. Hum. Retroviruses.* **20**(2004) 111.

4. D. G. Zhang , B. Debnath , S. H. Yu , T. W. Sanchez , F. Christ , Y. Liu , Z. Debyser , N. Neamati , G. S Zhao, *Bioorganic & Medicinal Chemistry*. **22** (2014) 5446.

5. B. W. Li, F. H. Zhang, E. Serrao, H. Chen, T. W. Sanchez, L. M. Yang, N. Neamati, Y.T. Zheng, H. Wang, Y. Q. Long, *Bioorganic & Medicinal Chemistry*. **22** (2014) 3146.

6. S. Ray, Z. Fatima, A. Saxena, *Med. Chem*. **10** (2010)147.

7. D.W. Zhang, H. Q. He, S.X. Guo, *Analytical Biochemistry*. **460**(2014) 36.

8.D. J. Hazuda, *Braz. J. Infect. Dis*. **14** (2010) 513.

9. W. G. Powderly, *J. Antimicrob. Chemother*. **65** (2010) 2485.

10. R. P. Bhole, K. P. Bhusari, *Archiv. Der. Pharmazie*. **344**(2011) 119.

11. Y. J. Ji, M. Shu, Y. Lin, Y. Q. Wang, R. Wang, Y. Hu, Z. H. Lin, *J. Mol. Struct*. **1045**(2013) 35.

12. R. D. Cramer, *J. Med. Chem*. **46**(2003) 374.

13. R. D. Cramer, P. Cruz, G. Stahl, W. C. Curtiss, B. Campbell, B. B. Masek, F. Soltanshahi, *J. Chem. Inf. Model*. **48**(2008) 2180.

14. X. Miao, Molecular Design Targeting on Tau Protein for Alzheimer's Disease. Chongqing, Chongqing University, 2013.

15. Z. J. Cheng , Y. Zhang, W. Z. Fu, *European Journal of Medicinal Chemistry*. **45**(2010) 3970.

16. E. Cichero, S. Cesarini, L. Mosti, P. Fossa, *J. Mol. Model*. **16** (2010) 1481.

17. A. Golbraikh, A. Tropsha, *J. Mol. Graph. Model*. **20** (2002) 269.

18. M. Le Bret, J. Polanski, F. Zouhiri, L. Jeanson, D. Desmaele, J. d'Angelo, J. F. Mouscadet, R. Gieleciak, J. Gasteiger, *J. Med. Chem*. **45** (2002) 4647.

19. L. Xu, X. G. Shao, Methods of Chemometrics. Science Press, Beijing, 2004, pp. 166-169.

20. D. I. Rhodes, T. S. Peat, N. Vandegraaff, D. Jeevarajah, G. Le1, E. D. Jones, J. A. Smith, J. AV Coates, L. J. Winfield, N. Thienthong, J. Newman, D. Lucent, J. H. Ryan, G. P.l Savage, C. L. Francis, J. J. Deadman, *Antiviral Chemistry & Chemotherapy*. **21**(2011) 155.

21. L. Ståhle. S. Wold, *J. Chemom*. **1**(1987) 185.

22. S. Wold, C. Albano. W. J. Dunn, et al., *Chemometrics*, **138**(1984)17.

23. R. M. Esnouf, J. Ren, A. L. Hopkins. C. K. Ross, et al., *Proc. Natl. Acad. Sci. U. S. A*. 94(1997) 3984.

Table Caption

Table 1 Structures and bioactivities of 62 integrase inhibitors

Table 2 The statistical results of Topomer CoMFA

Table 3 Structures and predicted pIC_{50} of new designed molecules

Figure Caption

Fig. 1 Cutting style of sample 42

Fig. 2 Linear regression between experimental and predicted pIC_{50} of 62 inhibitors

Fig. 3 3D contour of Topomer CoMFA model

(a)steric field map of Ra; (b)electrostatic field map of Ra；

(c)steric field map of Rb; (d)electrostatic field map of Rb

Fig. 4 Docking result of the compound 42 (a), 24 (b) 17(c) and 14 (d) into the binding site of HIV-1 integrase protease.