

Understanding the structure-activity relationship of betulinic acid derivatives as anti-HIV-1 agents by using 3D-QSAR and docking

Ping Lan · Wan-Na Chen · Zhi-Jian Huang ·
Ping-Hua Sun · Wei-Min Chen

Received: 5 August 2010 / Accepted: 5 October 2010 / Published online: 27 October 2010
© Springer-Verlag 2010

Abstract Novel anti-HIV-1 agents derived from betulinic acid have been greatly concerned. 3D-QSAR and molecular docking studies were applied to rationalize the structural requirements responsible for the anti-HIV activity of these compounds. The CoMFA and CoMSIA models resulted from 28 molecules gave r_{cv}^2 values of 0.599 and 0.630, r^2 values of 0.994 and 0.958, respectively. To estimate the predictive ability of the 3D-QSAR model, an external validation was employed. Based on the contour maps generated from both CoMFA and CoMSIA, we have identified some key features in the betulinic acid derivatives that are responsible for the anti-HIV activity. Molecular docking was used to explore the binding mode between these derivatives and HIV gp120. We have therefore designed a series of novel betulinic acid derivatives by utilizing the SAR results revealed in the present study, which were predicted with excellent potencies in the developed models. The results provide a valuable method to design new betulinic acid derivatives as anti-HIV-1 agents.

Keywords Anti-HIV-1 activity · Betulinic acid · Docking · gp120 · 3D-QSAR

Introduction

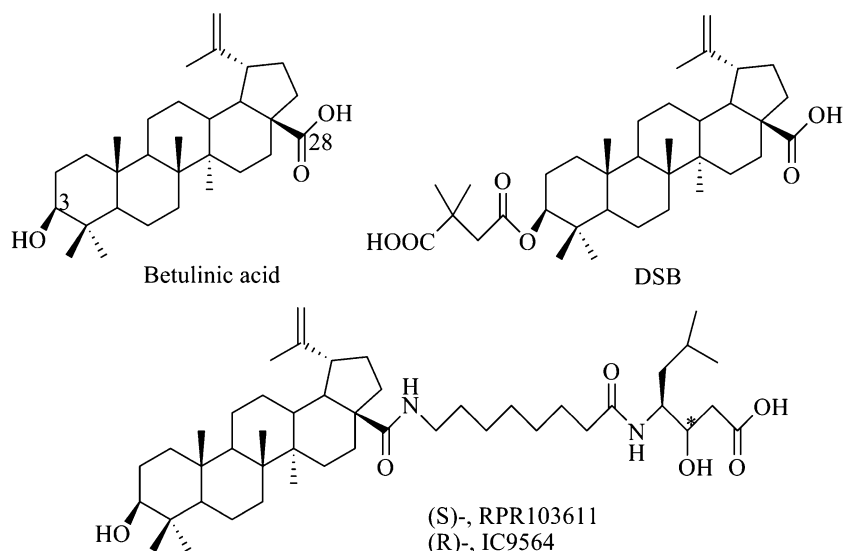
Human immunodeficiency virus (HIV) infection caused acquired immunodeficiency syndrome (AIDS) has become a

serious life-threatening problem to human kind [1–3]. Many strategies, such as highly active antiretroviral therapy (HAART) which combines several anti-HIV drugs including anti-HIV reverse transcriptase (RT) and protease agents, have been investigated in today's clinic treatments [1]. However, these strategies suffer from serious limitations such as the rapid development of multi-drug resistant (MDR) HIV viral strains and associated side effects of these inhibitors. Therefore, new classes of anti-HIV agents with novel structures or mechanisms of action are urgently needed to overcome the limitations of the current approaches [4–7].

Betulinic acid (Fig. 1) is a lupine-type pentacyclic triterpenoid saponin commonly found in some plants of *Betulaceae* family. The potent anti-HIV activity of betulinic acid has been greatly concerned. Some investigations on the structural modification of betulinic acid were carried out, and many derivatives with excellent anti-HIV activities have been obtained. 3-*O*-(3', 3'-dimethylsuccinyl)-betulinic acid (DSB, bevirimat, Fig. 1), one of the most famous derivatives of betulinic acid, which had passed the phase II clinical trials in 2007 [8, 9], had an EC_{50} of 0.00035 μ M and a therapeutic index (TI) of 20,000 in H9 lymphocytes [1]. Other famous derivatives such as IC9564 and RPR103611 (Fig. 1) which possessed side chain modification at the C-28 position also exhibited excellent anti-HIV potencies though they have not been under clinical evaluation yet. Previous studies demonstrated that the molecular mechanisms of action for betulinic acid derivatives were quite unique in comparison with currently known anti-HIV drugs that target HIV reverse transcriptase or protease. It was revealed in previous researches that chemical modification at C-3 position resulted in potent anti-HIV-1 derivatives that inhibit HIV-1 maturation, while C-28 position produced derivatives that inhibit HIV-1 entry by targeting HIV-1 gp120 [8].

P. Lan · W.-N. Chen · Z.-J. Huang · P.-H. Sun · W.-M. Chen (✉)
Guangdong Province Key Laboratory of Pharmacodynamic
Constituents of TCM and New Drugs Research, College of
Pharmacy, Jinan University,
Guangzhou 510632, People's Republic of China
e-mail: twmchen@jnu.edu.cn

Fig. 1 Structures of betulinic acid, DSB, IC9564 and RPR103611



A series of betulinic acid derivatives modified at C-3 and C-28 positions has been reported as bi-functional anti-HIV-1 agents to inhibit both HIV-1 maturation and entry [5, 6]. To explore the structure-activity relationship and the molecular mechanisms of action of these derivatives, the 3D-QSAR methods including comparative molecular field analysis (CoMFA) and comparative molecular similarity indices analysis (CoMSIA) methods as well as molecular docking were applied. The CoMFA and CoMSIA methods were performed to predict the activities of these derivatives and provided the regions in space where interactive fields may influence the activity. Molecular docking was employed to explore the binding modes of these compounds and the HIV-1 gp120. Based on the excellent performances of the 3D-QSAR and molecular docking studies, a set of novel betulinic acid derivatives have been designed and displayed excellent inhibitory potencies in the developed models. The results can not only help in understanding the structure-activity relationship of these compounds, but also be served as a useful guide for the design of new derivatives with better activities.

Materials and methods

Data sets for analysis

The *in vitro* anti-HIV-1 activity data reported as IC_{50} (concentration that inhibits HIV-1 NL4-3 replication by 50%) in MT4 cells by the betulinic acid derivatives were used for the current study. All the compounds and associated inhibitory activity data were obtained from literature [5, 6] reported by the same research group. Although these derivatives were reported over a period of

three years, compounds 1 and 17 displayed almost the same bioactivity data in both the two literatures, therefore, it was confirmed that all the reported bioactivity data were in the same experimental error. IC_{50} was converted to pIC_{50} by taking $\text{Log}(1/IC_{50})$. The pIC_{50} values were used as the dependent variables in all models subsequently developed. All 34 derivatives were divided into a training set of 28 molecules and a test set of six molecules, the ratio of the training and the test set was nearly 4.6 to 1. The selection of the test set was completed randomly with the consideration of ranging the distribution of the data sets in normal one. The structures and associated biological activities were shown in Table 1 and Table 2.

Alignment procedure

All the molecular modeling studies, CoMFA and CoMSIA reported herein were performed using SYBYL 8.1 program package of Tripos, Inc. [10]. The structural energy minimization was performed using the SYBYL energy minimizer (Tripos Force Field) and Gasteiger-Hückel charge, with a 0.05 kcal/(Å mol) energy gradient convergence criterion [11]. All of the structures were aligned into a lattice box by fitting with pentacyclic ring group as a common structure using compound 14 as a template, which was the most active derivative. The aligned molecules were shown in Fig. 2.

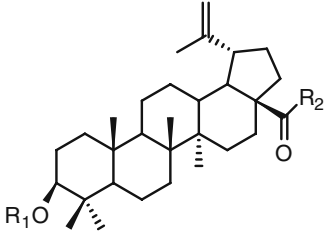
CoMFA and CoMSIA procedures

The standard CoMFA and CoMSIA procedures, as implemented in SYBYL 8.1, were performed. Both the CoMFA and CoMSIA descriptor fields were calculated at each lattice with grid spacing of 1 Å and extending to 4 Å units

in all three dimensions within defined region. The van der Waals potentials and coulombic terms, which represented steric and electrostatic fields, respectively, were calculated using Tripos force field [11]. In CoMFA method a sp^3 hybridized carbon atom with a charge of 1e was used as a probe atom, the steric and electrostatic fields were truncated at 30 kcal mol^{-1} [12]. The steric, electrostatic,

hydrophobic, hydrogen bond donor and hydrogen bond acceptor potential fields in CoMSIA method were calculated at each lattice intersection of a regularly spaced grid of 1 \AA and extending to 4 \AA using a probe atom with radius 1.0 \AA , +1.0 charge, and hydrophobic and hydrogen bond properties of +1. The attenuation factor was set to the default value of 0.3 [13].

Table 1 The structures of the training and test set molecules



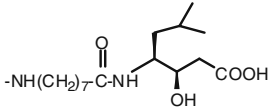
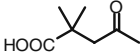
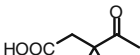
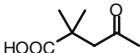
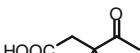
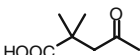
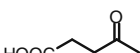
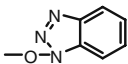

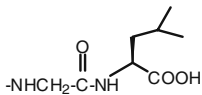

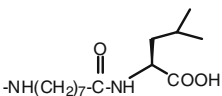

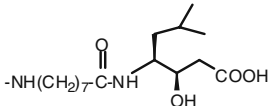
Compd. No.	Substituent	
	R ₁	R ₂
1	H	
2		OH
3		OH
4		-NH-(CH ₂) ₁₀ -COOH
5		-NH-(CH ₂) ₁₀ -COOH
6		-NH-(CH ₂) ₇ -COOH
7		
8		
9		
10		

Table 1 (continued)

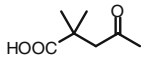
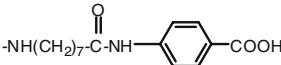
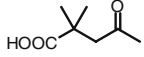
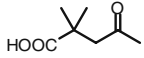
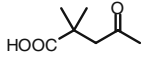
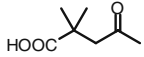
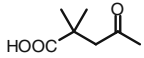
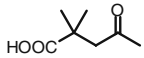
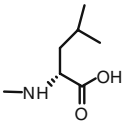
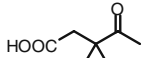
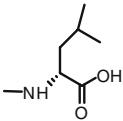
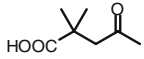
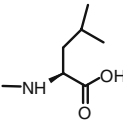
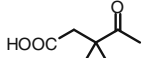
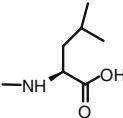
11		
12		-NH-(CH ₂) ₅ -NH-COCH ₃
13		-NH-(CH ₂) ₆ -NH-COCH ₃
14		-NH-(CH ₂) ₇ -NH-COCH ₃
15		-NH-(CH ₂) ₈ -NH-COCH ₃
16		-NH-(CH ₂) ₉ -NH-COCH ₃
17	H	-NH-(CH ₂) ₇ -NH-COCH ₃
18	H	-NH-(CH ₂) ₆ -NH-COCH ₃
19	H	-NH-(CH ₂) ₈ -NH-COCH ₃
20	H	-NH-(CH ₂) ₆ -NH-COCH ₂ COOH
21	H	-NH-(CH ₂) ₄ -COOH
22	H	-NH-(CH ₂) ₅ -COOH
23	H	-NH-(CH ₂) ₆ -COOH
24	H	-NH-(CH ₂) ₇ -COOH
25	H	-NH-(CH ₂) ₅ -CO-Gly
26	H	-NH-(CH ₂) ₆ -CO-Gly
27	H	-NH-(CH ₂) ₇ -CO-Gly
28	H	-NH-(CH ₂) ₅ -CO-Ala
29	H	-NH-(CH ₂) ₅ -CO-Pro
30	H	-NH-(CH ₂) ₅ -CO-Gly-Gly
31		
32		
33		
34		

Table 2 The actual pIC₅₀s, predicted pIC₅₀s (Pred.) and their residuals (Res.) of the training and test set molecules

Compd. no.	pIC ₅₀ Actual	CoMFA		CoMSIA	
		Pred.	Res.	Pred.	Res.
1	7.276	7.244	0.032	7.343	-0.067
2	7.125	6.923	0.202	6.550	0.575
3	6.347	6.541	-0.194	6.533	-0.186
4	7.921	7.789	0.132	7.737	0.184
5	7.456	7.449	0.007	7.507	-0.051
6	7.114	7.145	-0.031	6.864	0.250
7	7.456	7.493	-0.037	7.326	0.130
8*	7.796	7.912	-0.116	7.268	0.528
9	7.061	7.135	-0.074	6.737	0.324
10	7.018	7.027	-0.009	7.147	-0.129
11	8.155	8.101	0.054	8.086	0.069
12*	7.921	7.795	0.126	7.838	0.083
13	8.155	8.085	0.070	8.134	0.021
14	8.585	8.849	-0.264	9.048	-0.463
15	8.444	8.366	0.078	8.554	-0.110
16	7.921	7.806	0.115	8.228	-0.307
17	7.328	7.092	0.236	7.337	-0.009
18	7.398	7.493	-0.095	7.215	0.183
19	6.796	6.921	-0.125	6.723	0.073
20	6.337	6.412	-0.075	6.323	0.014
21	4.406	4.471	-0.065	5.200	-0.794
22*	5.585	5.738	-0.153	5.830	-0.245
23	4.680	4.665	0.015	4.922	-0.242
24	5.867	5.894	-0.027	5.602	0.265
25	4.365	4.348	0.016	4.199	0.166
26	5.208	5.343	-0.135	5.714	-0.506
27	6.796	6.694	0.102	6.469	0.327
28	4.106	4.088	0.018	4.030	0.076
29	4.091	4.003	0.088	4.114	-0.023
30*	4.237	4.993	-0.756	4.455	-0.218
31*	7.824	7.492	0.332	7.525	0.299
32	8.097	8.094	0.003	8.041	0.056
33	7.699	7.733	-0.034	7.520	0.179
34*	8.000	8.302	-0.302	7.714	0.286

* Test set molecules.

Partial least squares (PLS) analysis

To form the basis for a statistically significant model, the partial least squares (PLS) was used to linearly correlate the CoMFA and CoMSIA fields to the pIC₅₀ values. The cross-validation analysis was performed using the leave-one-out (LOO) method in which one molecule was removed from the data set and its activity was predicted using the model derived from the rest of the data set. PLS was conjunct with the cross-validation option to determine the optimum

number of components (ONC) which were then used in deriving the final CoMFA and CoMSIA model without cross-validation. The ONC was the number of components resulted in highest cross-validated correlated correlation coefficient (r_{cv}^2), which was defined as follows:

$$r_{cv}^2 = 1 - \frac{\sum (Y_{obs} - Y_{pre})^2}{\sum (Y_{obs} - Y_{mean})^2} \quad (1)$$

Where Y_{pre} , Y_{obs} and Y_{mean} are predicted, observed, and mean values of the target property (pIC₅₀), respectively [14–16]. Column filtering was used at the default value of 2.0 kcal mol⁻¹ in the cross-validation part. The final models were developed with ONC by using non-cross-validated analysis equal yielded the highest correlation coefficient (r^2).

Predictive correlation coefficient (r_{pred}^2)

The predictive abilities of each model were determined from a test set of six compounds that were not included in the training set. These molecules were aligned to the template and their pIC₅₀ values were predicted. The predictive correlation coefficient (r_{pred}^2), based on the molecules of test set, was defined as follows: $r_{pred}^2 = (SD-PRESS)/SD$

SD is the sum of the squared deviations between the predicted biological activities of the test set and mean activities of the training molecules and PRESS is the sum of squared deviations between predicted and actual activity values for each molecule in the test set [17–19].

External validation

It was indicated in the previous researches that a high cross-validated correlation coefficient (r_{cv}^2) is the necessary condition for a 3D-QSAR model to have a high predictive power; nevertheless, it is not a sufficient condition. In fact, the low values of r_{cv}^2 and r_{pred}^2 can serve as an indicator of a low predictive ability of a model, however, the opposite is not necessarily true [20]. In many cases, a model with high r_{cv}^2 and r_{pred}^2 values can be proved to be inaccurate. Even though a model may exhibit a good predictive ability based on the statistics for the test set, it is not always sure that the model will perform well on a new set of data [21]. The only way to estimate the true predictive power of a model is to test it on an external validation. To evaluate the true predictive abilities of the established models, both the CoMFA and CoMSIA models were subjected to rigorous external validation process. Several statistics such as r_m^2 , r_0^2 , R and k were employed according to literature [20–25].

Let \tilde{y}_i and y_i be the predicted and observed pIC₅₀ values of the test set molecules, respectively. If we plot y versus \tilde{y} for the ideal QSAR model, the regression line will bisect the angle formed by positive directions of the orthogonal

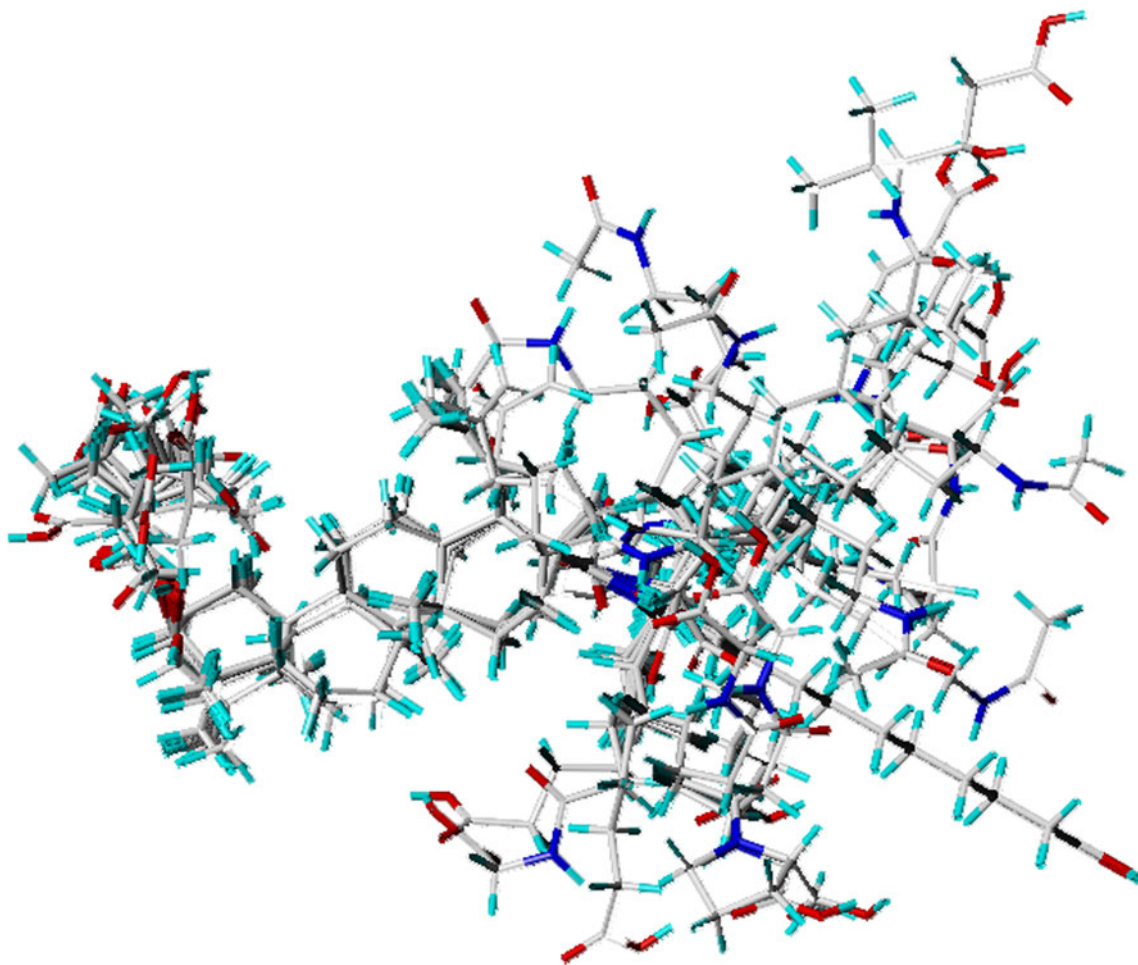


Fig. 2 Alignment of the compounds used in the training set

axes \tilde{y} and y . The regression line can be described by expression $y^r = a\tilde{y} + b$, where

$$a = \frac{\sum (y_i - \bar{y}_o)(\tilde{y}_i - \bar{y}_p)}{\sum (\tilde{y}_i - \bar{y}_p)^2} \quad (2)$$

and

$$b = \bar{y}_o - a\bar{y}_p. \quad (3)$$

In these equations, \bar{y}_o and \bar{y}_p are the average values of the observed and predicted pIC_{50} values of the test set molecules. For the ideal model, the slope a is equal to 1, intercept b is equal to 0, and correlation coefficient:

$$R = \frac{\sum (y_i - \bar{y}_o)(\tilde{y}_i - \bar{y}_p)}{\sqrt{\sum (y_i - \bar{y}_o) \sum (\tilde{y}_i - \bar{y}_p)^2}}. \quad (4)$$

For the regression is equal to 1. A 3D-QSAR model may have a high predictive power, if it is close to the ideal one. Meanwhile, the regression of y against \tilde{y} through the origin:

$y_i^{r_0} = k\tilde{y}_i$, should be characterized by k close to 1. Slope k is calculated as follow:

$$k = \frac{\sum y_i \tilde{y}_i}{\sum \tilde{y}_i^2}. \quad (5)$$

A good value of squared correlation coefficient (r^2) between observed and predicted values of the test set compounds does not always necessarily mean that the predicted values are very close to corresponding observed activity. Therefore, for better external predictive potential of the model, a modified r^2 [$r_m^2(\text{test})$] was introduced by the following equation:

$$r_m^2(\text{test}) = r^2(1 - \sqrt{|r^2 - r_0^2|}). \quad (6)$$

Where, the r^2 was the non-cross-validated correlation coefficient obtained from the PLS process, and the r_0^2 was calculated as follows:

$$r_0^2 = 1 - \frac{\sum (\tilde{y}_i - y_i^{r_0})^2}{\sum (\tilde{y}_i - \bar{y}_p)^2}. \quad (7)$$

Where, the $y_i^{r_0}$ was obtained by this formula:

$$y_i^{r_0} = k \tilde{y}_i. \quad (8)$$

According to literature [20–25], 3D-QSAR models were considered acceptable if they satisfy all of the following conditions:

$$r_{cv}^2 > 0.5, r^2 > 0.6, [(r^2 - r_0^2)/r^2] 0.1, 0.85 \leq k \leq 1.15 \text{ and } r_{m(\text{test})}^2 > 0.5. \quad (9)$$

Molecular docking

Most of the molecules involved in the present study were betulinic acid derivatives modified at C-28 position, which inhibited HIV-1 entry by targeting the HIV gp120. To investigate the protein-ligand interactions, some derivatives were docked into the binding site of HIV gp120. The Surflex-Dock uses an empirical scoring function and a patented search engine to dock ligands into a protein's binding site was applied to study molecular docking [10]. Crystal structure of HIV gp120 was retrieved from RCSB Protein Data Bank (PDB entry code: 1G9N). The gp120 structure was utilized in subsequent docking experiments without energy minimization. The ligands were docked into corresponding protein's binding site by an empirical scoring function and a patented search engine in Surflex-Dock [10]. All ligands and water molecules have been removed and the polar hydrogen atoms were added. Protomol, an idealized representation of a ligand that makes every potential interaction with the binding site, was employed to guide molecular docking. Establishment of protomol supplies three manners: (a) Automatic: Surflex-Dock finds the largest cavity in the receptor protein; (b) Ligand: A ligand in the same coordinate space as the receptor; (c) Residues: Specified residues in the receptor [10]. In this paper, the automatic docking was applied. To visualize the binding mode between the protein and ligand, the molecular computer aided design (MOLCAD) program was applied. MOLCAD calculates and displays the surfaces of channels and cavities, as well as the separating surface between protein subunits [10]. MOLCAD program provides several types to create a molecular surface. The Robbin surfaces which illustrated the secondary structure elements of the binding structure was applied, meanwhile the MOLCAD Robbin and Multi-Channel surfaces program displayed with several potentials were established. Other parameters were established by default in software.

Results and discussion

Analysis of CoMFA model

The statistical parameters associated with CoMFA model were listed in Table 3. The CoMFA model of betulinic acid derivatives gave a cross-validated correlation coefficient (r_{cv}^2) of 0.599 (>0.5) with an optimized component of 5, which suggested that the model should be considerably reliable to predict the IC₅₀ values. A high non-cross-validated correlation coefficient (r^2) of 0.994 with a standard error estimate (SEE) of 0.123, excellent *F* value of 675.283 and predictive correlation coefficient (r_{pred}^2) of 0.913 was obtained. Contributions of steric and electrostatic fields were 0.518 and 0.482, respectively. The actual and predicted pIC₅₀ values of the training set and test set by the model were given in Table 2, and the graph of actual activity versus predicted pIC₅₀ of the training set and test set was illustrated in Fig. 3.

Analysis of CoMSIA model

The CoMSIA model using steric, electrostatic, hydrophobic, hydrogen bond donor and hydrogen bond acceptor fields gave a good cross-validated correlation coefficient (r_{cv}^2) of 0.630 (>0.5) with an optimized component of 3. A high non-cross-validated correlation coefficient (r^2) of 0.958 with a standard error estimate (SEE) of 0.300, *F* value of 181.291 and predictive correlation coefficient (r_{pred}^2) of 0.938 were also obtained. Contributions of steric, electrostatic, hydrophobic, hydrogen bond donor and hydrogen bond acceptor fields were 0.177, 0.245, 0.114,

Table 3 Partial least squares (PLS) results summary of CoMFA and CoMSIA models

PLS statistics	CoMFA	CoMSIA
r_{cv}^2 ^a	0.599	0.630
r^2 ^b	0.994	0.958
ONC ^c	5	3
SEE ^d	0.123	0.300
<i>F</i> value	675.283	181.291
r_{pred}^2 ^e	0.913	0.938
Field contribution		
Steric	0.518	0.177
Electrostatic	0.482	0.245
Hydrophobic	–	0.114
H-bond Donor	–	0.240
H-bond Acceptor	–	0.224

^a cross-validated correlation coefficient; ^b non-cross-validated coefficient; ^c optimal number of components; ^d standard error of estimate; ^e predictive correlation coefficient.

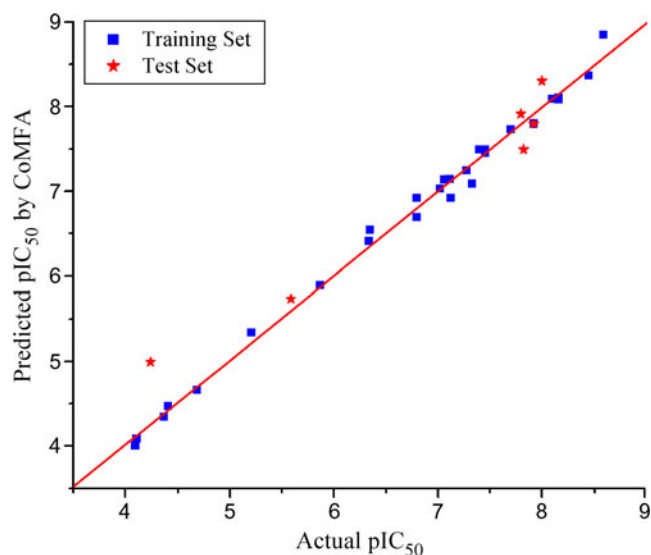


Fig. 3 Graph of actual versus predicted pIC_{50} of the training set and the test set using CoMFA

0.240 and 0.224, respectively. The actual and predicted pIC_{50} values and residual values for training set and test set compounds were given in Table 2. The relationship between actual activity and predicted pIC_{50} of the training set and test set compounds was illustrated in Fig. 4.

Analysis of external validation

The results of the external validation were shown in Table 4. The CoMFA and CoMSIA models using six derivatives in the test set, gave slope a values of 1.167 and 1.174 (close to 1), intercept b values of -1.321 and -1.056 (close to 0). The excellent r_m^2 (test) values of 0.843 and 0.862 (>0.5) as well as

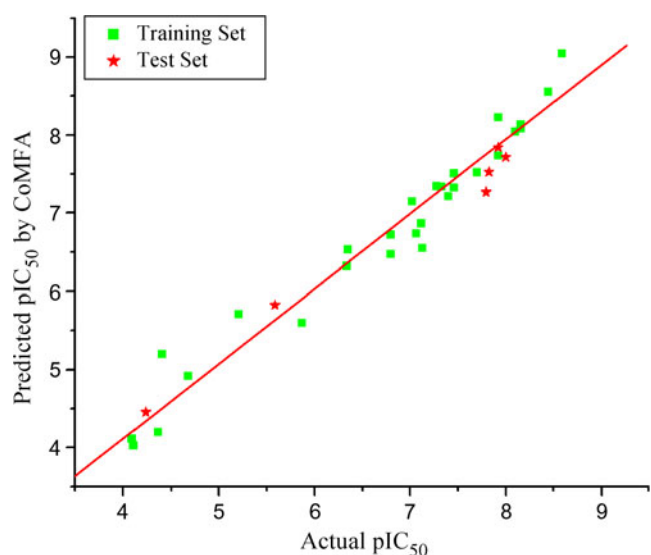


Fig. 4 Graph of actual versus predicted pIC_{50} of the training set and the test set using CoMSIA

Table 4 Results of the external validation for CoMFA and CoMSIA models

Parameters	CoMFA	CoMSIA
Slope a	1.167	1.174
Intercept b	-1.321	-1.056
Correlation coefficient R	0.982	0.992
Slope k	0.971	0.968
r_0^2	0.971	0.968
r_m^2 (test)	0.843	0.862
$[(r^2 - r_0^2) / r^2]$	0.023	-0.010

high slope of regression lines through the origin (k) values of 0.971 and 0.968 ($0.85 \leq k \leq 1.15$), and correlation coefficient (R) values of 0.982 and 0.992 (close to 1), the calculated $[(r^2 - r_0^2) / r^2]$ values of 0.023 and -0.010 (<0.1), respectively, were also obtained. The results of the external validation demonstrated that the CoMFA and CoMSIA models established by present study possessed a high accommodating capacity, and indicated that they may be reliable for being used to predict the activities of new derivatives.

CoMFA versus CoMSIA

When the CoMFA and CoMSIA models were applied to the six test compounds, both models gave satisfactory results. Compared with CoMFA, the CoMSIA model displayed better r^2_{cv} , r^2 and r_{pred}^2 values, which indicated that the CoMSIA model possessed higher predictive power than CoMFA. In the external validation analysis, although both the CoMFA and CoMSIA showed excellent parameters, the CoMSIA model was found to be slightly more reliable, it had better R and r_m^2 (test) values than the CoMFA model.

Graphical interpretation of CoMFA

To visualize the information content of the derived 3D-QSAR model, CoMFA contour maps were generated by plotting the coefficients from CoMFA model, which indicated regions in 3D space around the molecules where changes in the steric and electrostatic fields were predicted to increase or decrease the activity [9]. The CoMFA steric and electrostatic contour maps were showed in Fig. 5a and b using compound 14 as a reference structure. In Fig. 5a, green contours (80% contribution) refer to sterically favored regions while yellow contours (20% contribution) refer to sterically unfavored regions. In Fig. 5b blue contours (80% contribution) indicate regions where electron-donating groups are favored; red contours (20% contribution) reveal regions where electron-withdrawing groups are favored.

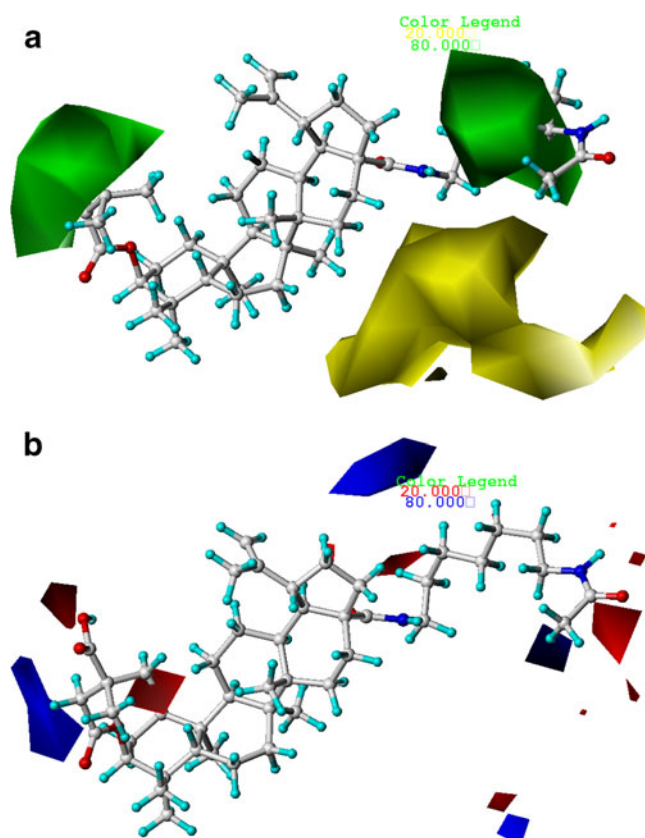


Fig. 5 Std* coeff contour maps of CoMFA analysis with 2 Å grid spacing in combination with compound 14. **(a)** Steric fields: green and yellow contours indicate regions where bulky groups increase or decrease the activity, and **(b)** Electrostatic fields: blue and red contours indicate regions where electron-donating or electron-withdrawing groups increase the potency

As shown in Fig. 5a, the steric favorable green contour on C-3 position indicated that a bulky substituent at this site would increase potency. This was consistent with the fact that C-3 modified derivatives (2-16, 31-34) exhibited much better IC_{50} values than that of C-3 hydroxyl derivatives (17-30). In fact, the C-3 modified derivatives such as DSB inhibit HIV-1 maturation, while C-28 modified derivatives (IC9564 and RPR103611) inhibit HIV-1 entry by targeting HIV-1 gp120. It can be inferred that compounds (2-16 and 31-34) modified at both C-3 and C-28 positions acted as bi-functional anti-HIV-1 agents, exhibited better activities than compounds 17-30 which could be regarded as mono-functional agents. Compared compound 1 with 10, as well as 14 and 17, it can easily be found that their activity discrepancies could also be explained by this green contour. The bulky groups at C-3 position were important for their potency, especially the 3',3'-dimethylsuccinyl group (11, 13, 14 and 15) and its analogues such as 2',2'-dimethylsuccinyl group (32). The steric favorable green contour around the middle of the C-28 side chain indicated that bulky groups in this position would benefit the activity.

Most of the derivatives revolved in this paper possessed a $-NH-(CH_2)_nNH-COR$ chain, the green contour demonstrated the impotence of the long alkyl side chain at the C-28 site. This observation can be confirmed by the fact that compound 14, which bear a $-NH-(CH_2)_7NH-COCH_3$ chain at C-28, was found to be at least 20-fold more potent than compound 2 which only possessed a carboxyl group. The big yellow contour near the terminal of C-28 side chain revealed that bulky groups at this position would decrease the activity. By checking up all the C-28 modified compounds, it was found that derivatives 25, 28 and 29 have the activity order of 25 ($R_2 = -NH-(CH_2)_5-CO-Gly$) > 28 ($R_2 = -NH-(CH_2)_5-CO-Ala$) > 29 ($R_2 = -NH-(CH_2)_5-CO-Pro$).

The electrostatic contour of CoMFA (Fig. 5b) showed red contours enclosing the C-3 position and the terminal of its side chain where electron-withdrawing groups were expected to benefit the potency. The blue contour at the middle of C-3 side chain indicated that electron-donating groups would be favorable. Hence, compounds 2-16 and 31-34 possessed both electron-donating group- methyl group at the middle of C-3 side chain, and the electron-withdrawing group-carboxyl group at the terminal of C-3 position showed excellent activities. A big blue contour appeared at the middle of the C-28 side chain, another smaller one existed at the end of C-28 side chain, indicated that the C-28 pharmacophore $-NH-(CH_2)_nNH-COCH_3$ was important for the activity, they also revealed that the terminal carboxylic group at C-28 was not a requirement for the activity. Compounds 13-17 which had an electron-donating methyl group at the terminal of C-28 side chain were the most active derivatives, whereas compounds 4-6 and 8-10 with a carboxyl group at the terminal were less active. The activity discrepancies of compounds 17-19 and compounds 20-24 can also be explained by this blue contour.

Graphical interpretation of CoMSIA

Figure 6a-e provided the CoMSIA steric, electrostatic, hydrophobic, hydrogen bond donor and acceptor contour plots for compound 14. The CoMSIA steric and electrostatic contour maps were almost the same to the corresponding CoMFA ones.

In Fig. 6c, white (20% contribution) and yellow (80% contribution) contours highlighted areas where hydrophilic and hydrophobic properties were favored. The white and yellow contours around the terminal and middle of the C-3 side chain suggested that a hydrophilic group at the terminal and a hydrophobic substituent at the middle of the C-3 side chain may benefit the potency. These contours demonstrated the extreme importance of 3',3'-dimethylsuccinyl group and its analogues such as 2',2'-dimethylsuccinyl group again. Hence, derivatives 2-16 and 31-34 possessed both hydrophobic methyl group at the middle of C-3 side chain, and the hydrophilic carboxyl group at the terminal of C-3 position

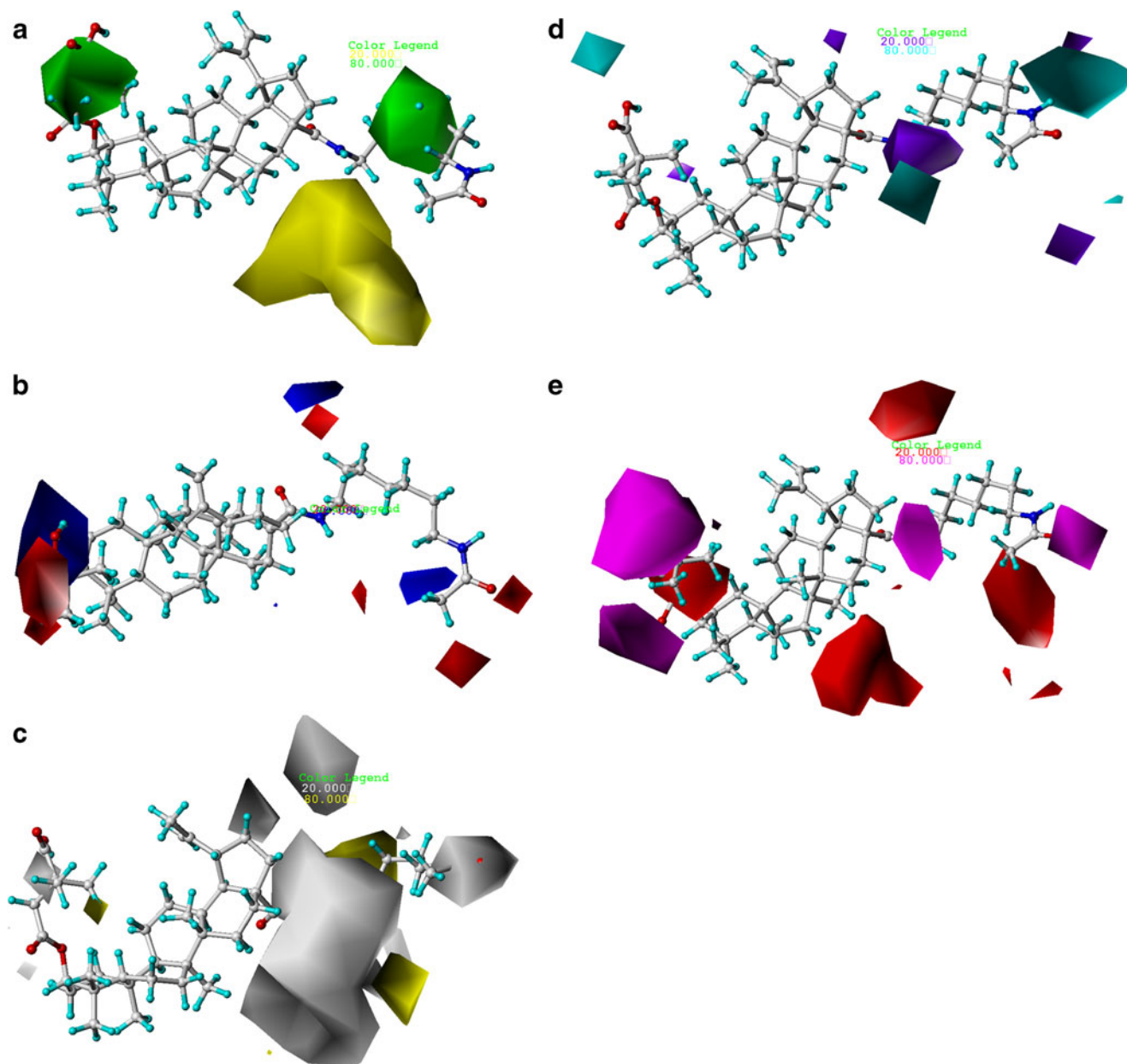


Fig. 6 Std* coeff contour maps of CoMSIA analysis with 2 Å grid spacing in combination with compound 14. **(a)** Steric contour map. Green and yellow contours refer to sterically favored and unfavored regions. **(b)** Electrostatic contour map. Blue and red contours refer to regions where electron-donating and electron withdrawing groups are favored. **(c)** Hydrophobic contour map. White and yellow contours

refer to regions where hydrophilic and hydrophobic substituent are favored. **(d)** Hydrogen bond donor contour map. The cyan and purple contours indicate favorable and unfavorable hydrogen bond donor groups. **(e)** Hydrogen bond acceptor contour map. The magenta and red contours demonstrated favorable and unfavorable hydrogen bond acceptor groups

exhibited excellent IC_{50} values. The huge white contour near the beginning of $-NH-(CH_2)_nNH-COCH_3$ chain indicated that the hydrophilic acylamino group at C-28 site was essential for the inhibitory potency. Another big yellow contour around the middle of the C-28 side chain revealed the importance of hydrophobic $-(CH_2)_n-$ group, this was consistent with the fact that derivative 5 ($R_2=-NH-(CH_2)_{10}-COOH$) displayed much better IC_{50} values than compounds 3 ($R_2=OH$) and 7 ($R_2=1H$ -benzotriazol-1-oxy-). The white

contour surround the carbamoyl group of $-NH-(CH_2)_nNH-COCH_3$ chain suggested that a hydrophilic group at this position may be favored. Derivatives 12-16 which possessed a hydrophilic carbamoyl group at this site were the most active compounds, and derivatives 4-7 which without a hydrophilic substituent at this position were less active. The yellow contour near the terminal of the C-28 side chain indicated that a hydrophobic group at this position may increase the activity. The most active compounds 12-16 all

possessed a hydrophobic methyl substituent at this site, meanwhile, derivatives which had a hydrophilic carboxyl group at the end of C-28 (e.g., 4-6, 6-10 and 17-34) showed decreased activities.

In Fig. 6d, the cyan (80% contribution) and purple (20% contribution) contours indicated favorable and unfavorable hydrogen bond donor groups. The cyan contour near the carboxyl group of 3',3'-dimethylsuccinyl group at C-3 position indicated that the carboxyl group may act as hydrogen bond donor, replaced it with hydrogen bond acceptor group could decrease the activity. In fact, the carboxyl group acted as H-bond donor and formed H-bond with the carbonyl group of the Gly42 residue. Derivatives 17-30 without 3',3'-dimethylsuccinyl group at C-3 position exhibited significantly decreased potencies. Another cyan contour near the second imino group of $-\text{NH}-(\text{CH}_2)_n\text{NH}-\text{COCH}_3$ chain suggested that a hydrogen bond donor substituent at this position may be favored. This may explain the reason why compounds 4-7 and 21-30 without an H-bond donor group at this site showed decreased activities.

In Fig. 6e, the magenta (80% contribution) and red (20% contribution) contours identified favorable and unfavorable positions for hydrogen bond acceptors. The purple contour near the carbonyl group of 3',3'-dimethylsuccinyl group at C-3 position indicated that the carbonyl group may act as hydrogen bond acceptor, removed it would decrease the potency. In fact, the carbonyl group at this site acted as H-bond acceptor by forming H-bond with the amino group of residue Arg144. The purple contour near the carbonyl of C-28 position indicated that it may act as H-bond acceptor. Actually, the carbonyl at C-28 position was an H-bond acceptor and formed H-bond with the imino group of residue Gly41. Another purple contour around the carbonyl

group of $-\text{NH}-(\text{CH}_2)_n\text{NH}-\text{COCH}_3$ chain also suggested that a hydrogen bond acceptor group at this position could benefit the activity. As a matter of fact, the carbonyl group was responsible for forming H-bond with the imino group of residue Leu123.

Molecular docking analysis

The docking studies of betulinic acid derivatives have not been reported up to now. In this paper, the most active derivative 14 and the most inactive compound 29 were selected for more detailed analysis. The MOLCAD Robbin surfaces structure of the HIV gp120 within the compound 14 and 29 were shown in Fig. 7. In Fig. 7a, the carbonyl group of the 3',3'-dimethylsuccinyl group at C-3 position acted as hydrogen bond acceptor by forming H-bond with the amino group of residue Arg144; while the carboxyl group acted as hydrogen bond donor and formed H-bond with the carbonyl group of the Gly42 residue. The carbonyl at C-28 position and $-\text{NH}-(\text{CH}_2)_n\text{NH}-\text{COCH}_3$ chain acted as hydrogen bond acceptors and formed H-bond with the imino groups of residues Gly41 and Leu123, respectively. The observations obtained from Fig. 7a were in agreement with that of CoMSIA hydrogen bond donor and acceptor contour maps. In Fig. 7b, the carbonyl at C-28 position of compound 29 formed H-bond with the imino groups of Gly41 residue as usual; the carboxyl group acted as hydrogen bond donor by forming H-bond with the carbonyl group of Ala 183 residue. Since the C-3 position of compound 29 remained unmodified, there were not any H-bonds found in the C-3 site. It can be inferred that this kind of binding mode resulted in weaker binding power between compound 29 with the pocket. This may also

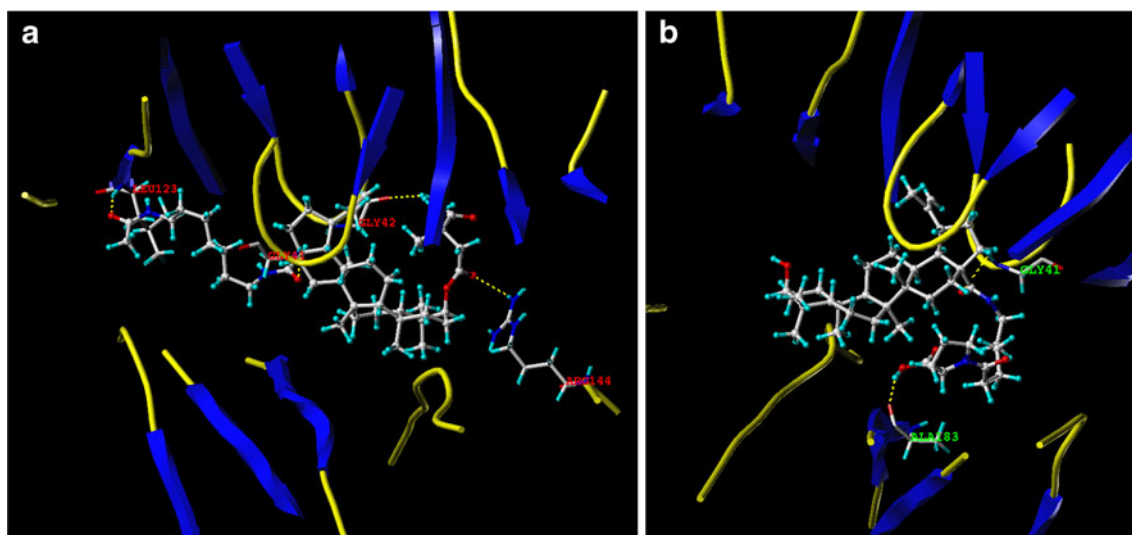


Fig. 7 MOLCAD Robbin surfaces structure of selected compound 14 (a) and 29 (b) in complex with the HIV gp120 (PDB code: 1G9N). Key residues and hydrogen bonds were labeled. The alpha helices

were shown as helices or cylinders, while beta sheets were shown as arrows and the loop regions as tubes

explain why derivatives (2-16, 31-34) modified at C-3 site exhibited much better IC_{50} values than that derivatives (17-30) with C-3 hydroxyl group.

The MOLCAD Robbin and Multi-Channel surfaces structure of the binding site of the HIV gp120 were also developed and displayed with cavity depth, electrostatic potential, lipophilic potential as well as hydrogen bonding sites to explore the ligand-receptor interactions, furthermore, to examine the 3D contour maps obtained by CoMFA and CoMSIA.

Figure 8 depicted the MOLCAD cavity depth potential surface of the binding site of HIV gp120 within compound 14, the cavity depth color ramp ranges from blue (low depth values = outside of the pocket) to light red (high depth values = cavities deep inside the pocket). As shown in Fig. 8, the whole molecule was in the yellow area which indicated that it was anchored to the deep inside HIV gp120.

Figure 9 demonstrated the MOLCAD electrostatic potential surface of the binding region, the color ramp for EP ranges from red (most positive) to purple (most negative). In Fig. 9, the carbonyl of the 3',3'-dimethylsuccinyl group was found in a red area, which indicated that electron-withdrawing properties were essential for the

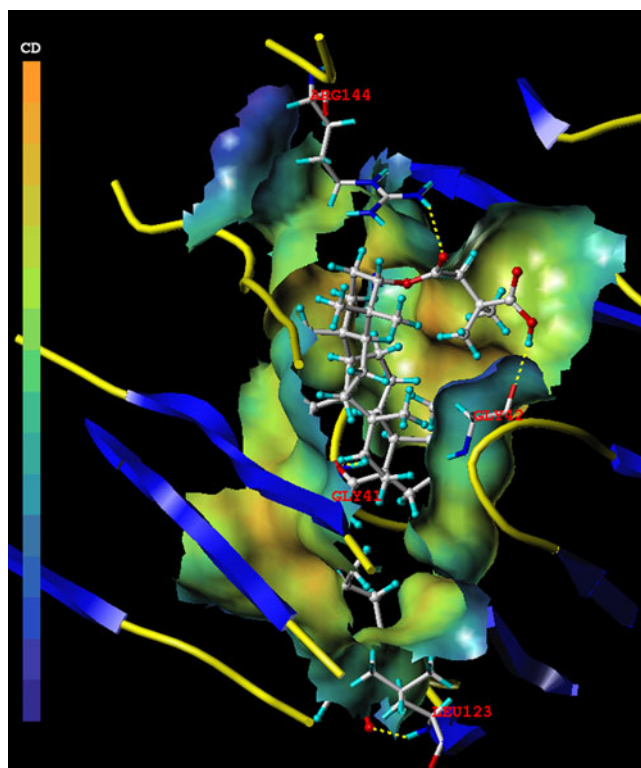


Fig. 8 The MOLCAD Robbin and Multi-Channel surfaces structure displayed with cavity depth potential of the HIV gp120 within the compound 14. Key residues and hydrogen bonds were labeled. The cavity depth color ramp ranges from blue (low depth values = outside of the pocket) to light red (high depth values = cavities deep inside the pocket)

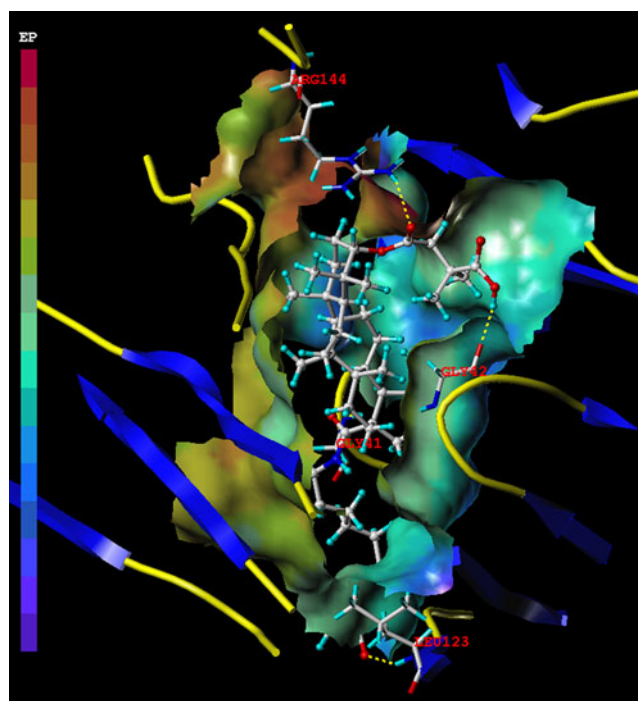


Fig. 9 The MOLCAD Robbin and Multi-Channel surfaces structure displayed with electrostatic potential of the HIV gp120 within the compound 14. Key residues and hydrogen bonds were labeled. The color ramp for EP ranges from red (most positive) to purple (most negative)

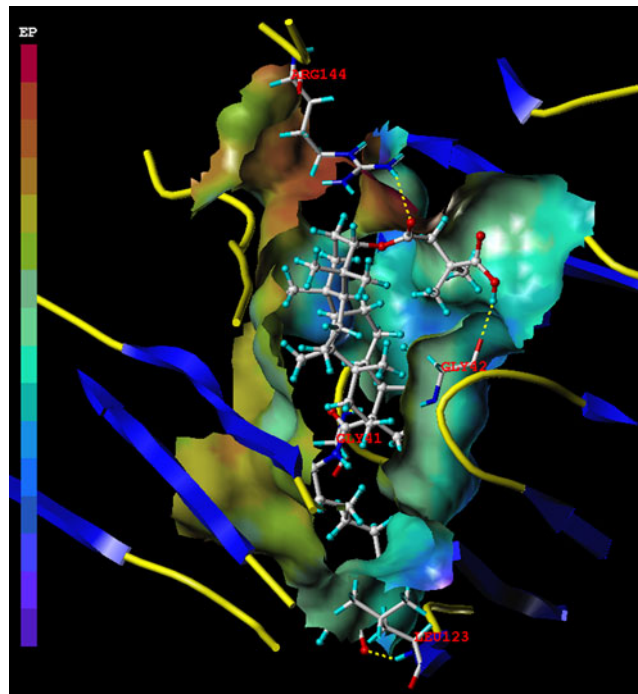


Fig. 10 The MOLCAD Robbin and Multi-Channel surfaces structure displayed with lipophilic potential of the HIV gp120 within the compound 14. Key residues and hydrogen bonds were labeled. The color ramp for LP ranges from brown (highest lipophilic area of the surface) to blue (highest hydrophilic area)

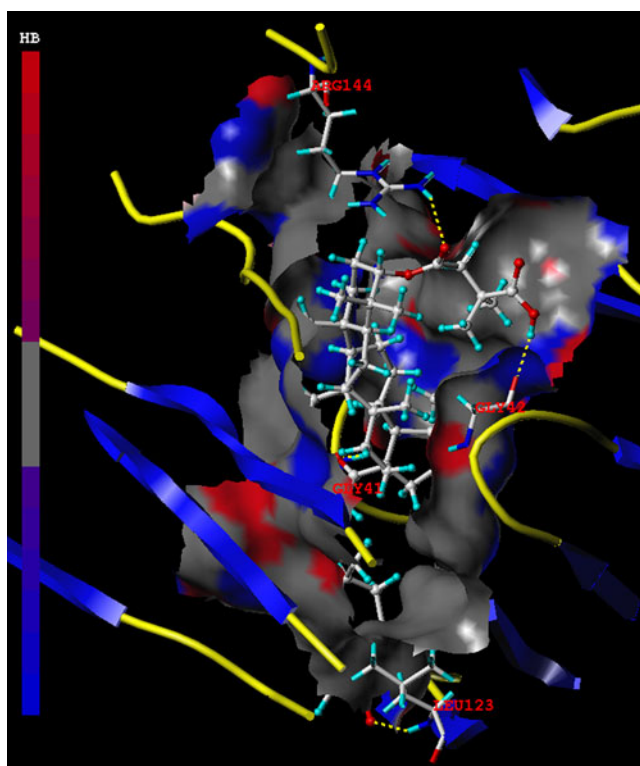


Fig. 11 The MOLCAD Robbin and Multi-Channel surfaces structure displayed with hydrogen bonding sites of the HIV gp120 within the compound 14. Key residues and hydrogen bonds were labeled. The color ramp for HP ranges from red (hydrogen bond donors) to blue (hydrogen bond acceptors)

potency; the two methyl substituents of the 3',3'-dimethylsuccinyl group were in a blue area, which suggested that electron-donating properties were favored. The acylamino group at C-28 site was found in a yellow area, indicating that an electron-withdrawing group would be favorable. The terminal of the C-28 side chain was in a blue area which suggested that an electron-donating substituent at this position would be favored. The observations obtained from this electrostatic potential surface satisfactorily matched the corresponding CoMFA and CoMSIA electrostatic contour maps.

Figure 10 showed the MOLCAD lipophilic potential surface of the binding area, the color ramp for LP ranges

from brown (highest lipophilic area of the surface) to blue (highest hydrophilic area). The carboxyl of the 3',3'-dimethylsuccinyl group at C-3 position was found in a blue area, suggesting that a hydrophilic substituent may benefit the potency; the two methyl groups were in a green area which indicated that hydrophobic substituent would be favorable. The terminal of the $-\text{NH}-(\text{CH}_2)_n\text{NH}-\text{COCH}_3$ side chain was found in a brown area which demonstrated that a hydrophobic substituent would increase the activity. These observations were in agreement with that obtained from CoMSIA hydrophobic contour map.

Figure 11 displayed the MOLCAD hydrogen bonding sites of the binding surfaces, ligands can be docked to proteins by matching the patterns displayed on the surface, the color ramp for HB ranges from red (hydrogen donors) to blue (hydrogen acceptors). In Fig. 11, the carbonyl of the 3',3'-dimethylsuccinyl group was oriented to a red area, which indicated that the surface of this region were hydrogen bond donors, and a hydrogen bond acceptor property may be favored; the carboxyl group was anchored into a blue area, which suggested that the surface of this site were hydrogen bond acceptors, and a hydrogen bond donor substituent would be favorable. The observations taken from this hydrogen bonding sites surfaces satisfactorily matched the corresponding CoMSIA hydrogen bond donor and acceptor contour maps.

Summary of structure-activity relationship

The structure-activity relationship revealed by 3D-QSAR and molecular docking studies were illustrated in Fig. 12. In general, the 3',3'-dimethylsuccinyl group and $-\text{NH}-(\text{CH}_2)_n\text{NH}-\text{COCH}_3$ chain or their analogues at C-3 and C-28 position were essential for excellent anti-HIV-1 potency. In detail, the electron-withdrawing, hydrophilic and hydrogen bond donor substituent in R_1 position would be favored; the electron-donating and hydrophobic groups in R_2 position would increase the activity; the minor, electron-donating and hydrophobic groups in R_3 position would benefit the potency; the carbonyl substituent of the acylamino and carbamoyl groups at C-28 side chain were responsible for binding to the HIV gp120.

Fig. 12 Structure-activity relationship revealed by 3D-QSAR and docking studies

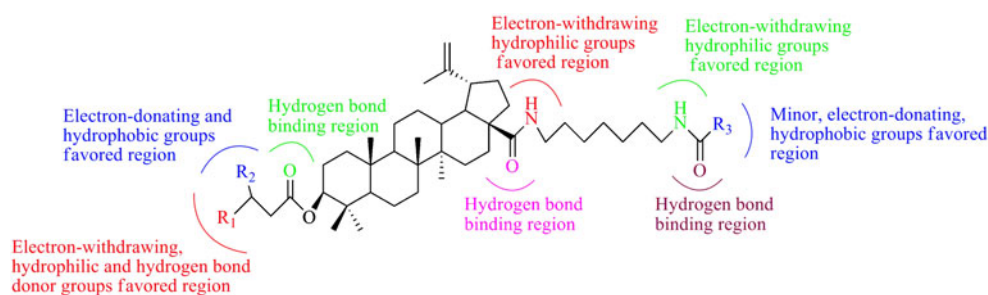
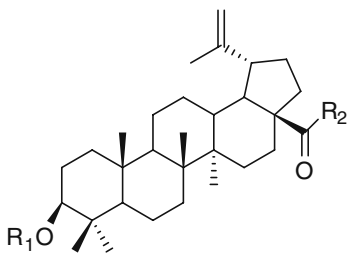


Table 5 Structures and predicted pIC₅₀ values of newly designed derivatives


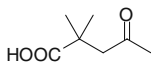
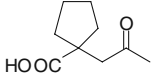
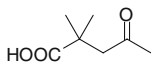
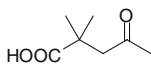
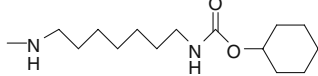
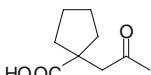
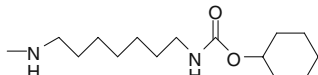
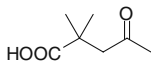
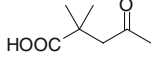
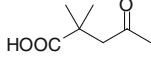
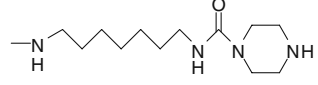
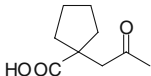
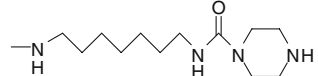
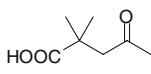
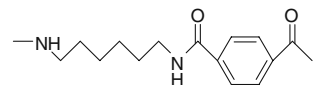
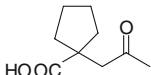
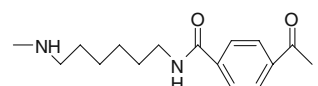
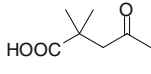
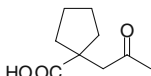
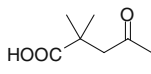
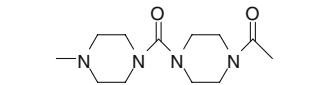
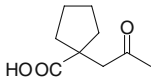
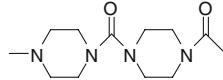
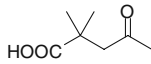
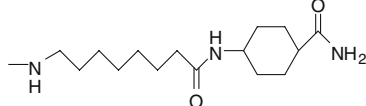
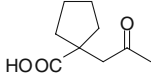
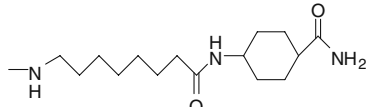
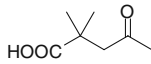
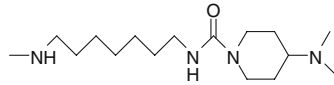
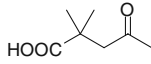
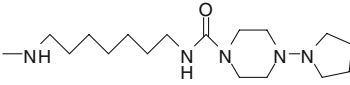
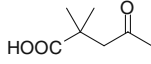
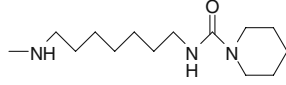
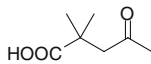
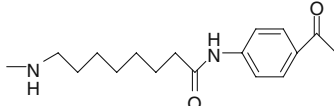
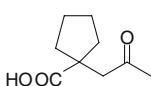
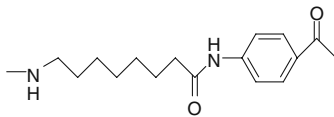
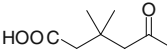
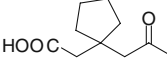
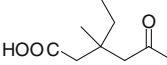
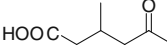
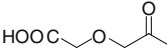
Compd. No.	Substituent		Predicted pIC ₅₀	
	R ₁	R ₂	CoMFA	CoMSIA
D1		-NH-(CH ₂) ₇ -NH-CONH ₂	8.814	8.865
D2		-NH-(CH ₂) ₇ -NH-CONH ₂	8.839	8.849
D3		-NH-(CH ₂) ₇ -NH-COO(CH ₃) ₃	8.389	8.828
D4			8.411	8.829
D5			8.127	8.540
D6		-NH-(CH ₂) ₇ -NH-CONH-(CH ₂) ₂ -NH-COCH ₃	8.690	8.832
D7		-NH-(CH ₂) ₇ -NH-CONH-CONH ₂	8.748	8.804
D8			8.344	8.415
D9			8.355	8.394
D10			8.551	8.451
D11			8.231	8.188
D12		-NH-(CH ₂) ₇ -NH-COOCH ₃	8.561	8.867
D13		-NH-(CH ₂) ₇ -NH-COOCH ₃	8.675	8.670
D14			8.288	8.185

Table 5 (continued)

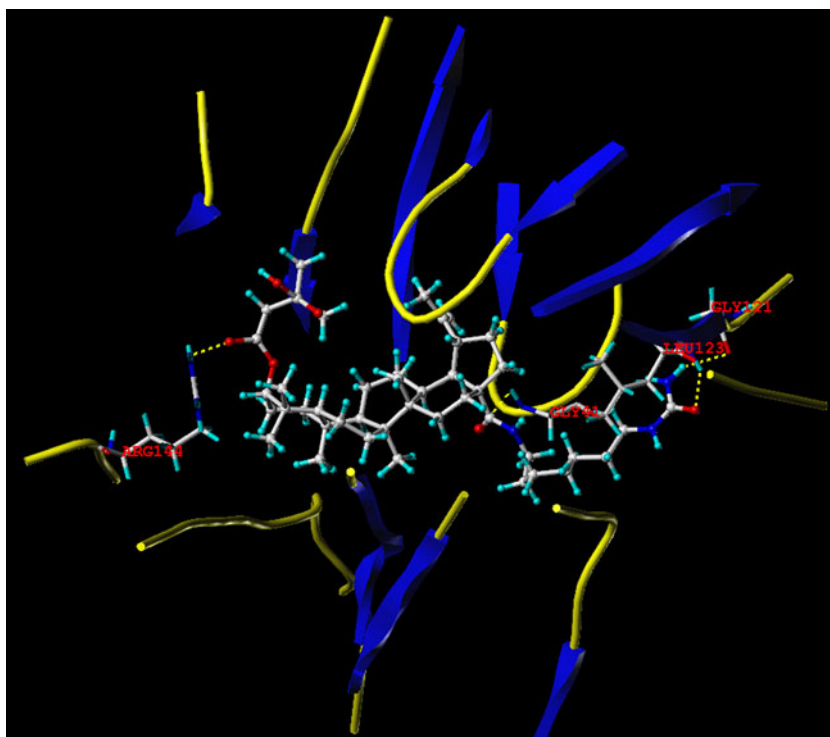
D15			8.290	8.203
D16			8.484	8.499
D17			8.124	8.255
D18			8.330	8.412
D19			8.306	8.440
D20			8.257	8.195
D21			8.104	8.048
D22			8.135	8.035
D23		-NH-(CH ₂) ₇ -NH-COCH ₃	8.209	8.345
D24		-NH-(CH ₂) ₇ -NH-COCH ₃	7.968	8.388
D25		-NH-(CH ₂) ₇ -NH-COCH ₃	8.181	8.438
D26		-NH-(CH ₂) ₇ -NH-COCH ₃	8.327	8.338
D27		-NH-(CH ₂) ₇ -NH-COCH ₃	8.230	8.363

Design for novel derivatives

By utilizing the structure-activity relationship revealed by this study, we have designed a series of novel derivatives based on the synthetic availability of these compounds. Firstly, the 3',3'-dimethylsuccinyl group and its analogues (e.g., 3',3'-cyclopentylsuccinyl, 3',3'-dimethylglutaryl, 3',3'-

cyclopentylglutaryl, 3'-ethyl-3'-methylglutaryl, 3'-methylglutaryl, 3',3'-dihydroxylsuccinyl, 3',3'-dioxolanyl succinyl and so on) which possessed a hydrogen bond acceptor substituent at the beginning, an electron-donating and hydrophobic substituent at the middle and an electron-withdrawing, hydrophilic and hydrogen bond donor substituent at the terminal were introduced in the C-3 position.

Fig. 13 MOLCAD Robbin surfaces structure of designed molecule D1 in complex with the HIV gp120 (PDB code: 1G9N). Key residues and hydrogen bonds were labeled. The alpha helices were shown as helices or cylinders, while beta sheets were shown as arrows and the loop regions as tubes



Besides, the satisfied $-\text{NH}-(\text{CH}_2)_n\text{NH}-\text{CO}-\text{R}$ pharmacophore with hydrogen bond acceptor, electron-withdrawing and hydrophilic groups at the beginning and middle sites, the minor, electron-donating and hydrophobic groups at the terminal were introduced in the C-28 position. These designed molecules were aligned to the database and their activities were predicted by the CoMFA and CoMSIA models previously established. The chemical structures and predicted pIC_{50} values of these compounds were shown in Table 5. As shown in Table 5, most of the designed compounds displayed excellent predicted pIC_{50} values in CoMFA and CoMSIA models. Compounds D1, D2, D3, D4, D6, D7, D12 and D13 which possessed electron-donating substituent at the terminal of the C-28 side chain exhibited best predicted pIC_{50} values. Compared with compound 14, the designed analogues D23-D31 showed slightly decreased activities. It can be found that the most potential substituent at C-3 position should be 3',3'-dimethylsuccinyl group. We have also performed a docking study on D1 which displayed highest predicted pIC_{50} values. The MOLCAD Robbin surfaces structure of the HIV gp120 within the designed D1 was shown in Fig. 13. The carbonyl group of the 3',3'-dimethylsuccinyl group at C-3 position acted as hydrogen bond acceptor by forming H-bond with the amino group of residue Arg144 as usual. The carbonyl at C-28 position and $-\text{NH}-(\text{CH}_2)_n\text{NH}-\text{COCH}_3$ chain acted as hydrogen bond acceptors and formed H-bond with the imino groups of residues Gly41 and Leu123, respectively. The amino group at the terminal of C-28 side

chain acted as hydrogen bond donor by forming H-bond with the carbonyl group of the Gly121 residue. The observations obtained from Fig. 13 were similar with that of Fig. 7a. The results validated the 3D-QSAR models established by present study.

Conclusions

Betulinic acid and its derivatives have long been regarded as a source of novel anti-HIV drug candidates. We employed 3D-QSAR and docking methods to explore the structure-activity relationship of a series of betulinic acid derivatives with anti-HIV activities. In 3D-QSAR studies, both the CoMFA and CoMSIA methods were satisfactory according to the statistical validation results as well as the contour map analysis. The excellent predictive ability of CoMFA and CoMSIA observed for the test set of compounds indicated that these models could be successfully used for predicting the IC_{50} values. The good performance of the external validation demonstrated that they may be reliable for predicting the activities of new derivatives. Furthermore, the 3D contour maps offered enough information to understand the structure-activity relationship and identified structural features influencing the activity. The structure-activity relationship was validated by newly designed derivatives. The correlation of the results obtained from 3D-QSAR and docking studies can be served as a useful guideline for the further modification of betulinic acid as anti-HIV agents.

Acknowledgments This research work is supported by the Natural Science Foundation of Guangdong Province (No. 9151063201000053), China.

References

1. Sun IC, Wang HK, Kashiwada Y, Shen JK, Cosentino LM, Chen CH, Yang LM, Lee KH (1998) Anti-AIDS Agents. 34. Synthesis and structure-activity relationships of betulin derivatives as anti-HIV agents. *J Med Chem* 41:4648–4657
2. Hashimoto F, Kashiwada Y, Cosentino LM, Chen CH, Garrett PE, Lee KH (1997) Anti-AIDS Agents-XXVII. Synthesis and anti-HIV activity of beulinic acid and dihydrobetulinic acid derivatives. *Bioorg Med Chem* 12:2133–2143
3. Gerrish D, Kim IC, Kumar DV, Austin H, Garrus JE, Baichwal V, Saunders M, Mckinnon RS, Anderson MB, Carlson R, Arranz-Plaza E, Yager KM (2008) Triterpene based compounds with potent anti-maturation activity against HIV-1. *Bioorg Med Chem Lett* 18:6377–6380
4. Evers M, Poujade C, Soler F, Ribeill Y, James C, Lelievre Y, Gueguen JC, Reisdorf D, Morize I, Pauwels R, Clercq ED, Henin Y, Bousseau A, Mayaux JF, Pecq JBL, Dereu N (1996) Betulinic acid derivatives: a new class of human immunodeficiency virus type 1 specific inhibitors with a new mode of action. *J Med Chem* 39:1056–1068
5. Huang L, Ho P, Lee KH, Chen CH (2006) Synthesis and anti-HIV activity of bi-functional betulinic acid derivatives. *Bioorg Med Chem* 14:2279–2289
6. Dang Z, Lai W, Qian K, Ho P, Lee KH, Chen CH, Huang L (2009) Betulinic acid derivatives as human immunodeficiency virus type 2 (HIV-2) inhibitors. *J Med Chem* 52:7887–7891
7. Soler F, Poujade C, Evers M, Carry JC, Henin Y, Bousseau A, Huet T, Pauwels R, Clercq ED, Mayaux JF, Pecq JBL, Dereu N (1996) Betulinic acid derivatives: a new class of specific inhibitors of human immunodeficiency virus type 1 entry. *J Med Chem* 39:1069–1083
8. Qian K, Yu D, Chen CH, Huang L, Morris-Natschke SL, Nitz TJ, Salzwedel K, Reddick M, Allaway GP, Lee KH (2009) Anti-AIDS Agents. 78. design, synthesis, metabolic stability assessment, and antiviral evaluation of novel betulinic acid derivatives as potent anti-human immunodeficiency virus (HIV) agents. *J Med Chem* 52:3248–3258
9. Pohjala L, Alakurtti S, Ahola T, Yli-Kauhaluoma J, Tammela P (2009) Betulin-derived compounds as inhibitors of alphavirus replication. *J Nat Prod* 72:1917–1926
10. Sybyl 8.1, Tripos Inc, St. Louis, US
11. Song QL, Sun PH, Chen WM (2010) Exploring 3D-QSAR for ketolide derivatives as antibacterial agents using CoMFA and CoMSIA. *Lett Drug Des Discov* 7:149–159
12. Lan P, Huang ZJ, Sun JR, Chen WM (2010) 3D-QSAR and molecular docking studies on fused pyrazoles as p38 α mitogen-activated protein kinase inhibitors. *Int J Mol Sci* 11:3357–3374
13. Cichero E, Cesarini S, Fossa P, Spallarossa A, Mosti L (2009) Thiocarbamates as non-nucleoside HIV-1 reverse transcriptase inhibitors: docking-based CoMFA and CoMSIA analyses. *Eur J Med Chem* 44:2059–2070
14. Puntambekar DS, Giridhar R, Yadav MR (2006) Understanding the antitumor activity of novel tricyclicpiperazinyl derivatives as farnesyltransferase inhibitors using CoMFA and CoMSIA. *Eur J Med Chem* 41:1279–1292
15. Cichero E, Cesarini S, Mosti L, Fossa P (2010) CoMFA and CoMSIA analyses on 4-oxo-1, 4-dihydroquinoline and 4-oxo-1, 4-dihydro-1, 5-, -1, 6- and -1, 8-naphthyridine derivatives as selective CB2 receptor agonists. *J Mol Model* 16:677–691
16. Murugesan V, Prabhakar YS, Katti SB (2009) CoMFA and CoMSIA studies on thiazolidin-4-one as anti-HIV-a agents. *J Mol Graph Model* 27:735–743
17. Choo HYP, Choi S, Jung SH, Koh HY, Pae AN (2003) The 3D-QSAR study of antitumor arylsulfonylimidazolidinone derivatives by CoMFA and CoMSIA. *Bioorg Med Chem* 11:4585–4589
18. Choo HYP, Lim JS, Kam Y, Kim SY, Lee J (2001) A comparative study of quantitative structure activity relationship methods based on antitumor diarylsulfonylureas. *Eur J Med Chem* 36:829–836
19. Tsukada T, Takahashi M, Takemoto T, Kanno O, Yamane T, Kawamura S, Nishi T (2010) Structure-based drug design of tricyclic 8H-indeno[1, 2-d][1, 3]thiazoles as potent FBPase inhibitors. *Bioorg Med Chem Lett* 20:1004–1007
20. Golbraikh A, Tropsha A (2002) Beware of q^2 . *J Mol Graph Model* 20:269–276
21. Roy PP, Roy K (2008) On some aspects of variable selection for partial least squares regression models. *QSAR Comb Sci* 27:302–313
22. Lu P, Wei X, Zhang R (2010) CoMFA and CoMSIA 3D-QSAR studies on quionolone carboxylic acid derivatives inhibitors of HIV-1 integrase. *Eur J Med Chem* 45:3413–3419
23. Basu A, Jasu K, Jayaprakash V, Mishra N, Ojha P, Bhattacharya S (2009) Development of CoMFA and CoMSIA models of cytotoxicity data of anti-HIV-1-phenylamino-1H-imidazole derivatives. *Eur J Med Chem* 44:2400–2407
24. Roy PP, Paul S, Mitra I, Roy K (2009) On two novel parameters for validation of predictive QSAR models. *Molecules* 14:1660–1701
25. Roy K, Roy PP (2009) Comparative chemometric modeling of cytochrome 3A4 inhibitory activity of structurally diverse compounds using stepwise MLR, FA-MLR, PLS, GFA, G/PLS and ANN techniques. *Eur J Med Chem* 44:2913–2922