QSAR Studies of 3,3′-(Substituted-benzylidene)-bis-4-hydroxycoumarin, Potential HIV-1 Integrase Inhibitor

Bo-Jian Li[a] (李柏堅), Chih-Chia Chiang[b] (姜治家) and Ling-Yih Hsu[a]* (徐令儀)
[a]*Department of Biological Science and Technology, China University of Science and Technology, Nangang, Taipei, Taiwan, R.O.C.*
[b]*Department of Chemistry, R.O.C. Military Academy, Fengshan, Kaohsiung 830, Taiwan, R.O.C.*

A series of biscoumarins against HIV-1 integrase was subjected to quantitative structure-activity relationship (QSAR) analysis. The result of the study showed that the inhibitory activity, as determined in the cell culture model, was highly correlated with the electronic (LUMO) and lipophilic effects exhibited by the substitution (R). The relationship can be expressed by the following regression equation: pIC50 = -1.903 (± 1.962) + 0.442 (± 0.066) logP – 4.504 (± 1.788) ε_{LUMO}.

Keywords: HIV-1 integrase; Inhibitory activity; Structure-activity relationship; Biscoumarin.

INTRODUCTION

Quantitative structure-activity relationship (QSAR) is a tool to rationalize the interaction of chemical compounds with living subject. The idea of QSAR methods is to draw conclusions by likeness assuming that similarity of drugs with respect to certain chemical properties will result in similar biological response. QSAR can be regarded as a computer-derived rule that quantitatively describes biological activity in terms of chemical descriptors. Once a QSAR is known, prediction or generation of new compounds with better activity is promising.

Human immunodeficiency virus (HIV) is a lent virus (a member of the retrovirus family) that can lead to acquired immunodeficiency syndrome (AIDS). Eventually most HIV-infected individuals develop AIDS. These individuals mostly die from opportunistic infections or malignancies associated with the progressive failure of the immune system. Anti-HIV (also called antiretroviral) medications are used to control the reproduction of the virus and to slow or halt the progression of HIV-related disease. Four classes of anti-HIV drugs including non-nucleoside reverse transcriptase inhibitors, nucleoside reverse transcriptase inhibitors, protease inhibitors and fusion inhibitors have been approved by the U.S. Food and Drug Administration (FDA) for the treatment of AIDS. When used in combinations, these medications are termed Highly Active Antiretroviral Therapy (HAART), sometimes referred to as a "cocktail". The successful use of HAART can dramatically suppress human immunodeficiency virus 1 (HIV-1) viral replication and effect significant immune reconstitution.[1-4]

However, despite full access to antiretroviral agents, the emergence of antiretroviral-resistant HIV-1 strains and/or drug toxicities can spoil effective treatment. Consequently, the need to develop antiretroviral agents with novel mechanisms of action persists for the treatment of HIV-1 infected patients. In October 2007, FDA approved the first drug in the integrase-inhibitor class for the treatment of HIV-1 as part of combination antiretroviral therapy, adding to the available chemotherapeutic agents for the effective treatment of HIV/AIDS.

HIV-1 integrase (HIV-1 IN) has emerged as an important therapeutic target for the design of anti-HIV agents because integration is essential for the replication of HIV and the enzyme responsible for integration appears to be absent in the mammalian host. Integrase catalyzes two reactions; 3′-end processing, in which two deoxynucleotides are removed from the 3′-ends of the viral DNA and the strand transfer reaction, in which the processed 3′-ends of the viral DNA are covalently ligated to the host chromosomal DNA.[5-7]

Integration of the proviral DNA is a key step in allowing viral DNA to become permanently inserted into the host genome. This step is essential for the subsequent transcrip-

* Corresponding author. E-mail: hlychit@cc.cust.edu.tw

tion of the viral genome which leads to production of new viral genomic RNA and viral proteins needed for the production of the next round of infectious virus. Furthermore, integrase uses a single active site to hold two different configurations of DNA substrates, which may restrict the ability of HIV to develop drug resistance to integrase inhibitors. Systematic screening of potential inhibitors has been undertaken using mostly purified integrase-based assays. From such screens several integrase inhibitor classes have now been identified.[8-12]

Nowadays, natural and synthetic coumarin derivatives represent a wide range of pharmacological properties. They have been reported to be used as anticoagulant,[13] antipsoriasis,[14] antibiotics,[14,15] fungicides,[16] antioxidant activity,[17] anti-inflammatory,[17] and antitumor agents.[18-21] More recently, coumarin derivatives have been evaluated in the treatment of human immunodeficiency virus, due to their ability to inhibit HIV-1 IN.[22,23]

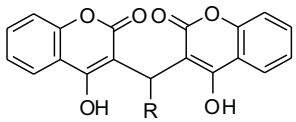
In continuation of our works on the discovery of novel HIV-1 IN inhibitor a quantitative structure-activity relationship (QSAR) study of some active 3,3′-(substituted-benzylidene)-bis-4-hydroxycoumarins is reported. In this paper, we tried to analyze the cause of difference of the activity of compounds against HIV-1 IN by using multivariable linear regression analysis and to obtain the information of mechanism of action at the molecular level. Physicochemical and quantum-chemical parameters taken into consideration in QSAR study are log P for hydrophobic effects, highest occupied molecular orbital energy (ε_{HOMO}), lowest unoccupied molecular orbital energy (ε_{LUMO}) as the electronic influences, MR (molar refractivity) for the steric interactions and NRB (number of rotatable bonds) for molecule flexibility.

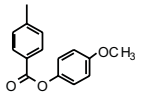
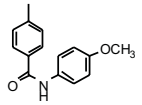
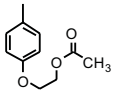
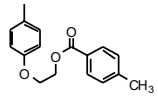
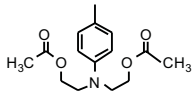
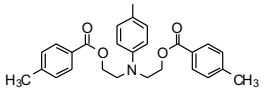
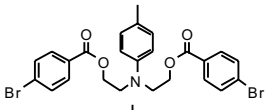
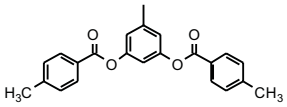
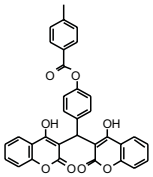
MATERIALS AND METHODS

The IC_{50} (50% inhibitory concentration in μM) of compounds **1-9** (Table 1) against HIV-1 IN were taken from our previous reported work.[22] The IC_{50} data were converted to pIC_{50} (negative logarithmic of IC_{50} value) for QSAR study. An increase in pIC_{50} value indicates the enhancement of the activity. The correlations were sought between inhibitory activity and various substituent constants at position R of molecule.

The molecular structures of the compounds in selected series were sketched using Chem Draw ultra 10.0 module of CS ChemOffice 2007 molecular modeling software. The sketched structures were then transferred to Chem3D module for generation of three dimensional structure (3D). The geometries of generated 3D structures were pre-optimized using MM2 force field as implemented in the Chem3D module of CS ChemOffice 2007. These molecular geometries were refined using the quantum chemical program package MOPAC 6.0 applying the PM3 pa-

Table 1. Structure and activity of coumarin analogues



Compd No	Substituent (R)	IC_{50}[a] (μM)	pIC_{50}[b]
1		3.1	5.509
2		23.1	4.636
3		96.0	4.018
4		23.1	4.636
5		102	3.991
6		1.8	5.745
7		0.96	6.018
8		0.58	6.237
9		0.49	6.310

[a] *in vitro* IC_{50} (50% inhibitory concentration in μM) against HIV-1 IN.

[b] negative logarithmic of IC_{50} value.

rameterization together with eigenvector following geometry optimization procedure. The gradient norm 0.001 kcal/Å was used to calculate electronic, geometric and energetic parameters for the isolated molecules. The descriptor values for all the molecules were calculated using "compute properties" module of program. Calculated thermodynamic descriptors included critical temperature (T), ideal gas thermal capacity (C), critical pressure (Pc), boiling point (BP), Henry's law constant (H), bend energy (Eb), heat of formation (Hf), total energy (TE), and logrhythm of partition coefficient (log P), the key physicochemical parameter describing lipophilicity. Steric descriptors derived were Connolly accessible area (CAA), Connolly molecular area (CMA), Connolly solvent excluded volume (CSEV), exact mass (EM), molecular weight (MW), principal moment of inertia-X component (PMI-X), principal moment of inertia-Y component (PMI-Y), principal moment of inertia-Z component (PMI-Z), number of rotatable bonds (NRB), molar refractivity (MR), and Ovality (OVAL). Electronic descriptors such as dipole (DPL), electronic energy (ElcE), highest occupied molecular orbital energy (ε_{HOMO}), lowest unoccupied molecular orbital energy (ε_{LUMO}), repulsion energy (RE), VDW-1,4-energy (E14), and Non-1,4-VDW energy (E). Correlation and regression analysis of the QSAR study were run on a PC computer using the Microsoft Excel program. Multiple regression analysis which involves finding the best fit of dependent variable (inhibitory activities against HIV-1 IN) to a linear combination of independent variables (descriptors) are used by the least squares method. In the equations, the figures in the parentheses are the standard errors of the regression coefficients, *n* is the number of compounds, R is the multiple correlation coefficient, R^2 is the determination coefficient, Q^2 is the leave-one-out(LOO) cross validated R^2, *F* is the significance test (*F*-test), p is the probability factor and SEE is the standard error of estimate. *F*-test values are for all equations statistically significant at the 1% level of probability.

RESULTS AND DISCUSSSION

In organic chemistry, a compound containing a carbonyl group (HC-C=O) is normally in rapid equilibrium with an enol tautomer (C=C-OH). In coumarin molecule the keto-enol tautomerism is possible. Comparison of the energy (by Hartree-Fock/6-31G* calculation) between the keto and enol structures of coumarin revealed that the enol form is more stable than the keto form by 21 kcal/mol (Fig. 1). The two hydrogen bonds between the hydrogen of enol and the oxygen of lactone in the *bis*-coumarin molecule may contribute to the stable conformation.

The values of the physicochemical parameters and frontier molecular orbital energy levels[23,24] for the compounds are given in Table 2. The inhibitory activities of the studied compounds against HIV-1 IN are reported by the pIC_{50} (log 1/C) values, where C is the molar concentration of the observed HIV-1 IN inhibitory activity.

Generally, the number of compounds should be four times greater than the number of descriptors to get a reliable model using multiple linear regression. In this study, we reduced down our pool of descriptors using principal components analysis for building multilinear QSAR models. The treatment started with the reduction in the number of molecular descriptors. If two descriptors intercorrelated highly with each other, then only one of them was selected; descriptors with insignificant variance for the data set treated were also rejected. This helps to speed up the descriptor selection and reduce the probability of including

Table 2. QSAR parameters and biological activities

Compd No.	IC_{50} (μM)	pIC_{50}	logP	ε_{Lumo}	ε_{Homo}	NRB[a]	MR[b]
1	3.1	5.509	4.990	-1.1	-8.91	7	151.162
2	23.1	4.636	4.312	-1.1	-8.52	7	153.111
3	96	4.018	3.219	-0.99	-9.09	8	135.3
4	23.1	4.636	5.604	-0.99	-9.06	9	160.514
5	102	3.991	3.343	-0.96	-8.73	12	158.709
6	1.8	5.745	8.113	-0.92	-8.30	14	209.137
7	0.96	6.018	8.796	-0.97	-8.38	14	214.3
8	0.58	6.237	7.575	-0.94	-9.29	9	186.086
9	0.49	6.310	6.716	-1.12	-9.25	9	232.034

[a] NRB = Number of Rotatable Bonds; [b] MR = Molar Refractivity

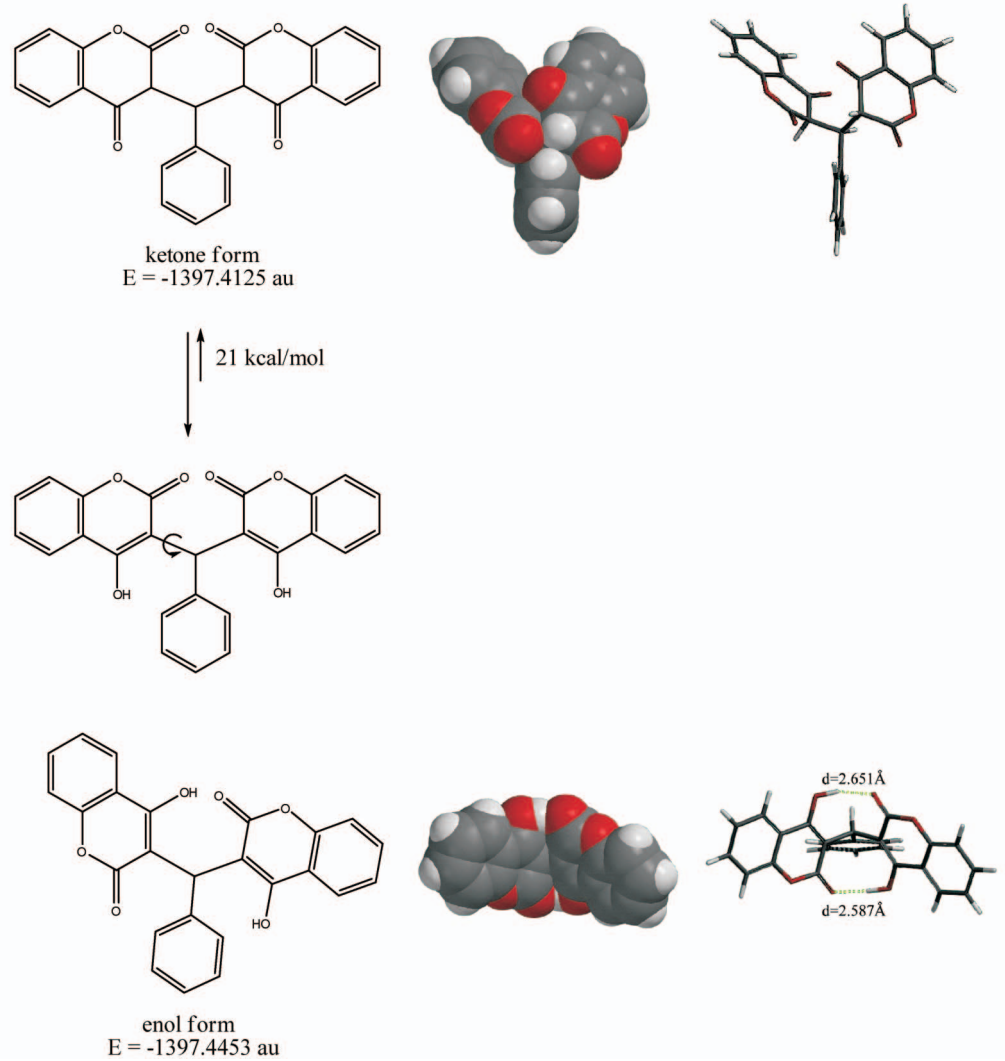


Fig. 1. Molecular models of the minimum-energy conformation of *bis*-coumarin.

unrelated descriptors by chance. Therefore, log P, lowest unoccupied molecular orbital energy (ε_{LUMO}), highest occupied molecular orbital energy (ε_{HOMO}), number of rotatable bonds (NRB) and molar refractivity (MR) were derived for the multivariable regression analysis. The descriptors derived for modeling the inhibitory activity of biscoumarins are summarized in Table 2. The correlation between the different physicochemical descriptors as independent variable and the pIC_{50} as dependent variable was determined. Some of the statistically significant models are given below and their statistical measures are listed in Table 3.

Model 1

Model 1 is a monoparametric equation modeled for HIV-1 IN inhibitory activity and has good correlation between biological activity and descriptor as indicated by R = 0.87 and explains 76% variance in inhibition. Low standard deviation of the model demonstrates accuracy of the model. The model showed overall significance level better than 99%, with the $F_{(1,7)} = 21.97$ against the values $F_{(0.01;1,7)} = 12.25$ from the F distribution. The p-value is 0.0022 (<0.01), which shows there is significant relationship between log P and biological activity. Leave-one-out (LOO) cross validated value ($Q^2 = 0.634$) reflects the good predictive power of the model. The plot of experimental *vs*. predicted property is shown in Fig. 2(a).

$$pIC_{50} = 2.942(\pm 0.515) + 0.391(\pm 0.084)\log P \quad (1)$$

Table 3. Statistical parameters of the 4 best scored equations (number of descriptor are not exceed 2)

Eq.	ND[a]	R[b]	R^2 [c]	Q^2 [d]	F[e]	p[f]	SEE[g]
(1)	1	0.871	0.758	0.634	$F_{(1,7)} = 21.97$	0.0022	0.486
(2)	2	0.886	0.786	0.151	$F_{(2,6)} = 10.99$	0.0098	0.495
(3)	2	0.931	0.868	0.690	$F_{(2,6)} = 19.65$	0.0023	0.388
(4)	2	0.939	0.883	0.768	$F_{(2,6)} = 22.56$	0.0016	0.366

[a] ND = the number of descriptors, [b] R = correlation coefficient, [c] R^2 = coefficient of determination, [d] Q^2 = leave-one-out (LOO) cross validated R^2, [e] F = Fisher-ratio (the significance of the equations), [f] p = probability factor, [g] SEE = standard error of estimate.

Model 2

Model 2 is a biparametric (log P and MR) equation and was considered to be a poor predictor of activity because of its poor validation ($Q^2 = 0.151$) although the coefficient of determination ($R^2 = 0.786$) is higher as compared with that of model 1. Further, the descriptors log P and MR are found to be highly correlated to each other as shown in the correlation matrix (Table 4). The correlation coefficient between log P and MR is 0.836 which is outside the acceptable range (< 0.8).

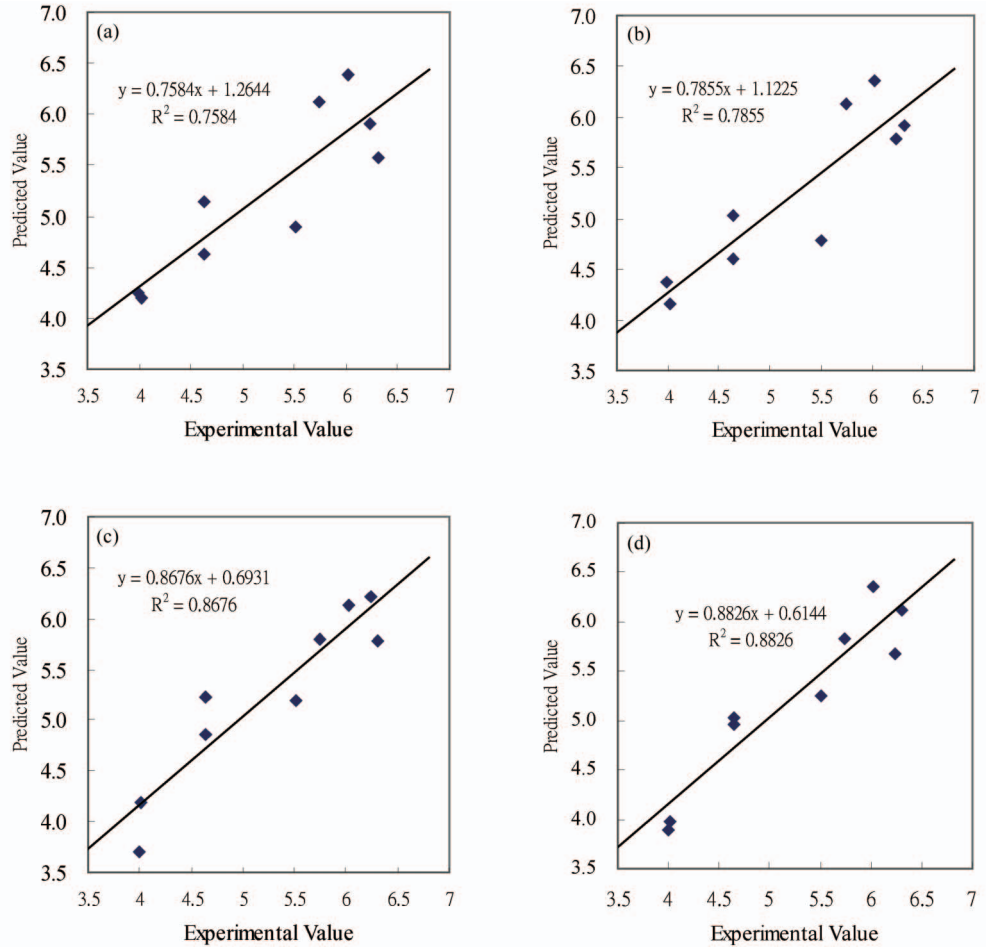


Fig. 2. (a) Correlation between the predicted pIC_{50} and the experimental pIC_{50} by Eq. (1). (b) Correlation between the predicted pIC_{50} and the experimental pIC_{50} by Eq. (2). (c) Correlation between the predicted pIC_{50} and the experimental pIC_{50} by Eq. (3). (d) Correlation between the predicted pIC_{50} and the experimental pIC_{50} by Eq. (4).

$$pIC_{50} = 2.139(\pm 1.060) + 0.279(\pm 0.155)\log P + 0.008(\pm .009)MR \quad (2)$$

Model 3

The replacement of MR by NRB drastically improves the quality of prediction, as indicated by much better statistical parameters in Table 3 ($R^2 = 0.868$, $Q^2 = 0.690$). A plot of experimental *vs.* predicted property is shown in Fig. 2(c). The negative coefficient of NRB revealed that the molecular flexibility of biscoumarins is detrimental to the activity.

$$pIC_{50} = 3.673(\pm 0.527) + 0.495(\pm 0.081)\log P - 0.135(\pm 0.061)NRB \quad (3)$$

Model 4

Model 4 is the best model since it shows the best correlation coefficient R = 0.94 and explains 88% variance in inhibition. Further, the smaller standard error of estimate, higher *F* and leave-one-out (LOO) cross validated value ($Q^2 = 0.768$) demonstrates satisfactory predictive ability of the model. The statistical parameters of Model 4 are shown in Table 3. This model indicates that the ε_{Lumo} is highly correlated to the activity. The negative sign of the coefficient of ε_{Lumo} indicates that inhibitory activity increases as ε_{Lumo} decreases. The distribution of the LUMO orbitals in the coumarin **8** is shown in Fig. 3. By coumarin molecule the maximum value of the LUMO is located on the enol carbon. Atoms with the highest density of this orbital could be the place of the highest reactivity of the substance and the compound could react there as a nucleophilic reagent according to frontier molecular orbital theory. The plot of experimental *vs*. predicted property is shown in Fig. 2(d).

Table 4. Correlation matrix for the selected descriptors

	pIC_{50}	LogP	ε_{Lumo}	ε_{Homo}	NRB	MR
pIC_{50}	1					
LogP	0.871	1				
ε_{Lumo}	-0.071	0.304	1			
ε_{Homo}	-0.061	0.237	0.240	1		
NRB	0.226	0.571	0.680	0.616	1	
MR	0.818	0.836	0.066	0.140	0.573	1

$$pIC_{50} = -1.903(\pm 1.962) + 0.442(\pm 0.066)\log P - 4.504(\pm 1.788)\varepsilon_{Lumo} \quad (4)$$

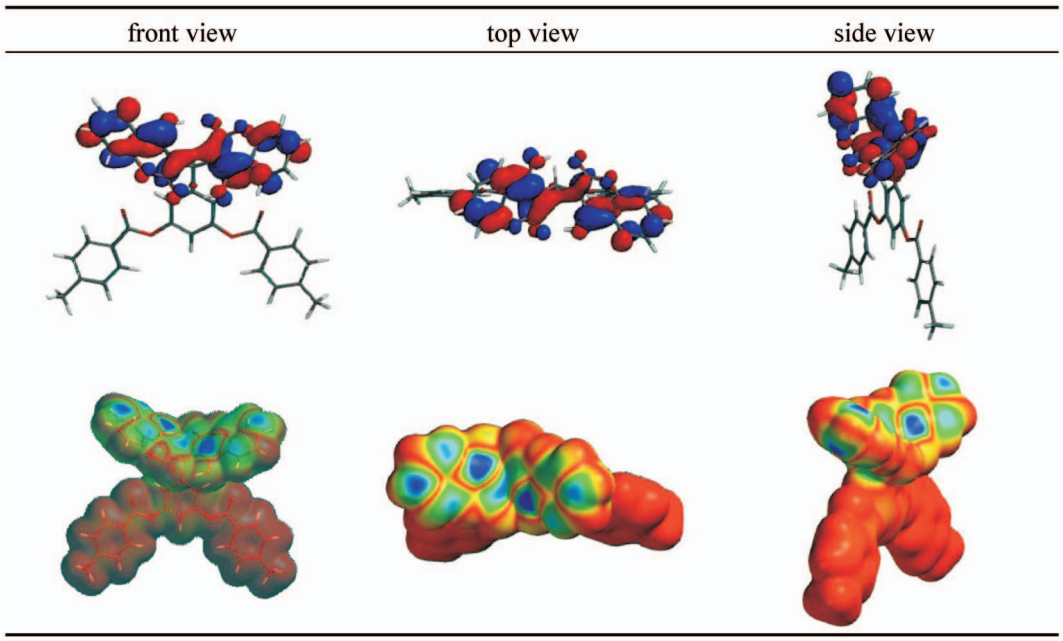


Fig. 3. LUMO topology distribution and electron density surface map for compound **8**. On the electron density surface map (lower figure), colors toward the red indicate small values of the LUMO, while colors toward the blue indicate large values of the LUMO. By coumarin molecule, the maximum value of the LUMO (the dark blue spot) is located on the enol carbon.

Table 5. Experimental and predicted pIC_{50} values

Compd. No.	pIC_{50} (Exp.)	pIC_{50} (Pred.)				Residual[a]			
		Eq(1)	Eq(2)	Eq(3)	Eq(4)	Eq(1)	Eq(2)	Eq(3)	Eq(4)
1	5.509	4.896	4.774	5.197	5.258	-0.613	-0.735	-0.312	-0.251
2	4.636	4.630	4.601	4.861	4.958	-0.006	-0.036	0.225	0.322
3	4.018	4.202	4.150	4.186	3.979	0.185	0.132	0.168	-0.038
4	4.636	5.136	5.022	5.231	5.034	0.500	0.385	0.594	0.397
5	3.991	4.251	4.377	3.707	3.899	0.260	0.385	-0.284	-0.092
6	5.745	6.118	6.121	5.797	5.827	0.374	0.376	0.052	0.082
7	6.018	6.386	6.354	6.135	6.354	0.368	0.336	0.117	0.337
8	6.237	5.908	5.782	6.206	5.679	-0.329	-0.455	-0.031	-0.557
9	6.310	5.572	5.920	5.781	6.111	-0.738	-0.390	-0.529	-0.199

[a] Residual = Pred. – Exp.

Table 6. Experimental and LOO predicted pIC_{50} values

Compd. No.	pIC_{50} (Exp.)	pIC_{50} (Pred.)				Residual[a]			
		Eq(1)	Eq(2)	Eq(3)	Eq(4)	Eq(1)	Eq(2)	Eq(3)	Eq(4)
1	5.509	4.800	4.532	5.092	5.157	-0.709	-0.977	-0.416	-0.352
2	4.636	4.629	4.549	4.934	5.101	-0.007	-0.087	0.297	0.464
3	4.018	4.289	4.248	4.265	3.957	0.272	0.230	0.247	-0.061
4	4.636	5.198	5.134	5.311	5.090	0.561	0.498	0.675	0.454
5	3.991	4.360	4.663	3.064	3.826	0.368	0.672	-0.927	-0.165
6	5.745	6.249	6.195	5.831	5.874	0.504	0.451	0.086	0.129
7	6.018	6.599	6.456	6.236	6.550	0.581	0.438	0.218	0.532
8	6.237	5.829	5.644	6.187	5.484	-0.407	-0.593	-0.049	-0.753
9	6.310	5.460	4.479	5.658	5.931	-0.850	-1.831	-0.652	-0.379

[a] Residual = LOO Pred. – Exp.

CONCLUSION

QSAR analysis was performed on a series of biscoumarins using molecular modeling program Chemoffice 2007. QSAR models were proposed for inhibitory activity of the biscoumarins using chemSAR descriptors employing sequential multiple regression analysis method. The selected models were checked for multicollinearity and correlation with *F*-test value. The predictive power of each model was validated with determination coefficient R^2 method and leave-one-out cross validation Q^2 method. The application of the QSAR paradigm may be useful in elucidating the mechanisms of chemical–biological interaction for HIV-1 IN. Our QSAR results show that electronic (LUMO) and lipophilic properties of the molecules are the two most important determinants of the activity. The finding of the study will be helpful in the design of potent inhibitory analogs of biscoumarin.

ACKNOWLEDGMENT

Financial support from National Science Council of the Republic of China is gratefully acknowledged.

SUPPLEMENTARY INFORMATION

Supplementary data associated with the experimental and predicted values of activity data are shown in Table 5 and 6, respectively.

Received April 20, 2010.

REFERENCES

1. Autran, B.; Carcelain, G.; Li, T. S.; Blanc, C.; Mathez, D.; Tubiana, R. *Science* **1997**, *277*, 112.
2. Pakker, N. G.; Notermans, D. W.; Boer, R. J.; Roos, M. T.; Wolf, F.; Hill, A. *Nat. Med.* **1998**, *4*, 208.
3. Wong, J. K.; Gunthard, H. F.; Havlir, D. V.; Zhang, Z. Q.; Haase, A. T.; Ignacio, C. C. *Proc. Natl. Acad. Sci.* **1997**, *94*, 12574.

4. Zhang, Z. Q.; Notermans, D. W.; Sedgewick, G.; Cavert, W.; Wietgrefe, S.; Zupancic, M. *Proc. Natl. Acad. Sci.* **1998**, *95*, 1154.
5. Katz, R. A.; Skalka, A. M. *Ann. Rev. Biochem.* **1994**, *63*, 133.
6. Rice, P.; Craigie, R.; Davies, D. R. *Curr. Opin. Struct. Biol.* **1996**, *6*, 76.
7. Farnet, C. M.; Wang, B.; Lipford, J. R.; Bushman, F. D. *Proc. Natl. Acad. Sci.* **1996**, *93*, 9742.
8. Fesen, M. R.; Pommier, Y.; Leteurtre, F.; Hiroguchi, S.; Yung, J.; Kohn, K. W. *Biochem. Pharmacol.* **1994**, *48*, 595.
9. LaFemina, R. L.; Graham, P. L.; LeGrow, K.; Hastings, J. C.; Wolfe, A.; Young, S. D.; Emini, E. A.; Hazuda, D. J. *Antimicrob. Agents Chemother.* **1995**, *39*, 320.
10. Fesen, M. R.; Kohn, K. W.; Leteurtre, F.; Pommier, Y. *Proc. Natl. Acad. Sci.* **1993**, *90*, 2399.
11. Dupont, R.; Jeanson, L.; Mouscadet, J. F.; Cotelle, P. *Bioorg. Med. Chem. Lett.* **2001**, *11*, 3175.
12. Hazuda, D. J.; Felock, P.; Witmer, M.; Wolfe, A.; Stillmock, K.; Grobler, J. A.; Espeseth, A.; Gabryelski, L.; Schleif, W.; Blau, C.; Miller, M. D. *Science* **2000**, *287*, 646.
13. Egan, D.; Kennedy, R. O.; Moran, E.; Cox, D.; Prosser, E.; Thornes, R. D. *Drug Metab. Rev.* **1990**, *22*, 503.
14. Maxwell, A. *Mol. Microbiol.* **1993**, *9*, 681.
15. Abd-Allah, O. A. *Farmaco* **2000**, *55*, 641.
16. El-Agrody, A. M.; El-Latif, M. S. A.; El-Hady, N. A.; Fakary, A. H.; Bedair, A. H. *Molecules* **2001**, *6*, 519.
17. Emmanuel-Giota, A. A.; Fylaktakidou, K. C.; Hadjipavlou-Litina, D. J.; Litinas, K. E.; Nicolaides, D. N. *J. Heterocycl. Chem.* **2001**, *38*, 717.
18. Dexeus, F. H.; Logothetis, C. J.; Sella, A.; Fitz, K.; Amato, R.; Reuben, J. M.; Dozier, N. J. *J. Clin. Oncol.* **1990**, *8*, 325.
19. Nofal, Z. M.; El-Zahar, M. I.; El-Karim, S. S. *Molecules* **2000**, *5*, 99.
20. Madari, H.; Panda, D.; Wilson, L.; Jacobs, R. *Cancer Res.* **2003**, *63*, 1214.
21. Payá, M.; Ferrándiz, M. L.; Miralles, F.; Montesinos, C.; Ubeda, A.; Alcaraz, M. J. *Arzneim.-Forsch.* **1993**, *43*, 655.
22. Chiang, C. C.; Mouscadet, J. F.; Tsai, H. J.; Liu, C. T.; Hsu, L.Y. *Chem. Pharm. Bull.* **2007**, *55*, 1740.
23. Ghose, A. K.; Crippen, G. M. *J. Chem. Inf. Comput. Sci.* **1987**, *27*, 21.
24. Viswanadhan, V. N.; Ghose, A. K.; Revankar Revankar, G. R.; Robins, R. K. *J. Chem. Inf. Comput. Sci.* **1989**, *29*, 163.