

Comparative molecular field analysis (CoMFA) and comparative molecular similarity indices analysis (CoMSIA) of thiazolone derivatives as hepatitis C virus NS5B polymerase allosteric inhibitors

Beilei Lei · Juan Du · Shuyan Li · Huanxiang Liu ·
Yueying Ren · Xiaojun Yao

Received: 15 January 2008 / Accepted: 13 July 2008 / Published online: 21 August 2008
© Springer Science+Business Media B.V. 2008

Abstract Three-dimensional quantitative structure-activity relationship (3D-QSAR) models for a series of thiazolone derivatives as novel inhibitors bound to the allosteric site of hepatitis C virus (HCV) NS5B polymerase were developed based on CoMFA and CoMSIA analyses. Two different conformations of the template molecule and the combinations of different CoMSIA field/fields were considered to build predictive CoMFA and CoMSIA models. The CoMFA and CoMSIA models with best predictive ability were obtained by the use of the template conformation from X-ray crystal structures. The best CoMFA and CoMSIA models gave q^2 values of 0.621 and 0.685, and r^2 values of 0.950 and 0.940, respectively for the 51 compounds in the training set. The predictive ability of the two models was also validated by using a test set of 16 compounds which gave r^2_{pred} values of 0.685 and 0.822, respectively. The information obtained from the CoMFA and CoMSIA 3D contour maps enables the interpretation of their structure-activity relationship and was also used to the design of several new inhibitors with improved activity.

Keywords 3D-QSAR · Hepatitis C virus NS5B polymerase · CoMFA · CoMSIA

Electronic supplementary material The online version of this article (doi:10.1007/s10822-008-9230-7) contains supplementary material, which is available to authorized users.

B. Lei · J. Du · S. Li · Y. Ren · X. Yao (✉)
Department of Chemistry, Lanzhou University,
Lanzhou 730000, China
e-mail: xjyao@lzu.edu.cn

H. Liu
School of Pharmacy, Lanzhou University, Lanzhou 730000,
China

Introduction

The hepatitis C virus (HCV) is a member of the Flaviviridae family [1] and was identified in 1989 as the leading pathogen for non-A, non-B viral hepatitis [2]. It has been a major public health problem that infects an estimated 170 million people worldwide [3], and often referred to as the silent killer. Patients can remain asymptomatic for decades before developing to liver cirrhosis and/or hepatocellular carcinoma [4]. Currently there is no vaccine on the market to prevent the disease for specific antiviral agent directed against HCV infection. The approved treatments for HCV infection are based upon the use of pegylated interferon- α (PEG-IFN) [5] in combination with the oral nucleoside drug ribavirin. However, there are serious side effects and poor response rates. So, novel, more efficacious and tolerable anti-HCV therapies are urgently needed.

HCV NS5B is a virus-encoded RNA-dependent-RNA polymerase (RdRp). It is one of the essential enzymes responsible for HCV viral replication and has attracted most attention as a potential therapeutic target [6]. In the past decade, there have been tremendous efforts dedicated to the development of novel, effective inhibitors of anti-HCV NS5B polymerase, and numerous classes of non-nucleoside inhibitors with different scaffolds have appeared in the literatures, such as benzimidazole [7–9], proline sulfonamide [10], indole [11–14], diketoacid [15–17], phenylalanine [18], thiophene [19], pyridazinone [20, 21] and acrylic acid [22, 23] and so on. Recently, a novel class of HCV NS5B polymerase inhibitors with thiazolone scaffold has been identified by the structure-based design method [24]. Based on this scaffold, a series of thiazolone derivatives were synthesized and they were found to be bound to the allosteric site of HCV NS5B.

Three-dimensional quantitative structure-activity relationship (3D-QSAR) approaches, such as comparative molecular

field analysis (CoMFA) [25] and comparative molecular similarity analysis (CoMSIA) [26] have been widely used in drug design [27, 28]. In this investigation, several 3D-QSAR models were built for a series of thiazolone derivatives [24, 29–31] to gain insight into the key structural factors affecting their activity. The developed models can then be used to the further modification of the molecules with desired activity. In order to obtain reasonable QSAR model and investigate the influence of the conformation of the template molecule on the quality of the model, two different conformations of the molecules template were considered. In addition, CoMSIA models based on different field/fields were performed to choose the best combination of fields. The results indicated that the template conformation extracted from X-ray crystallographic (XRC) complex performed better than that obtained from the lower energy conformation. The developed models can not only be used in rapid and accurate prediction of the activities of new designed inhibitors, but also provide some beneficial information for designing new inhibitors with improved inhibitory activity.

Materials and methods

Data sets and biological activity

A data set of 67 thiazolone derivatives for the 3D-QSAR studies which have been discovered to be novel inhibitors bound to the allosteric site of HCV NS5B polymerase was taken from literatures [24, 29–31]. The biological data obtained as IC_{50} (drug concentration, in μM) were converted into pIC_{50} ($-\log IC_{50}$) values and used as dependent variable in the CoMFA and CoMSIA QSAR analyses. Besides, X-ray crystallographic (XRC) complex structure of the High Throughput Screening hit with the HCV NS5B polymerase (PDB: 2HWH) [24] is available that provides template conformation for modeling the compounds in the present study. To validate the built model, the data set was divided into training and test sets. In the splitting the original data set into training and test sets, the similarity of the structural diversity and the range of biological activities in the two sets were considered to ensure the representative ability of the training set compounds. Then, 51 compounds were selected for the training set and 16 compounds for the test set. The structures and biological data of the compounds in the training and test sets are listed in Table 1.

Template selection, molecular modeling and alignment

In the development of 3D-QSAR models, the choice of the template molecule and the template conformation are of great importance. In this paper, compound 37 was selected

as the template molecule. This compound was chosen mainly because it is one of the most active compounds, and its binding mode in the allosteric site of HCV NS5B polymerase was available from the high-resolution X-ray crystal structure.

In this investigation, two different conformations of compound 37 were selected to assess their effects on the performance of the of 3D-QSAR models. The first one (Template Conformation 1) was the lowest energy conformation minimized by Tripos force field and partial atomic charges were calculated using Gasteiger–Hückel method. The conformation representing the global minimum of the ligand may not bind to the receptor and some degree of torsional freedom is required for the drug to adapt to the receptor binding site to yield a drug-receptor complex of lower energy [32, 33]. The conformation of compound 37 extracted from the XRC complex structure (PDB: 2HWH) was selected as the second template conformation (Template Conformation 2). The conformations of the Template 1 and the Template 2 are shown in Fig. 1b and c. Three-dimensional structure building and all computational studies were performed by SYBYL 6.9 molecular modeling package [34] on a Silicon Graphics O2 workstation running under the IRIX 6.5 operating system. Molecular structures were built with molecular sketch program, and then energy minimizations were performed using the Tripos force field with a distance-dependent dielectric function and Powell method with a convergence criterion of 0.01 kcal/mol. Partial atomic charges were calculated using Gasteiger–Hückel method. The conformational search was also carried out using the multisearch routine in SYBYL to get a reasonably low energy conformation for each compound. Compounds 3–7, 10–12, 14, 23 and 24 contain an stereogenic atom (shown as C* in the scaffold in Table 1), and they were built reasonably as *S*-enantiomeric structure in terms of the binding conformation of structurally similar compound 1 in the complex with HCV NS5B (PDB: 2HWI).

Molecules alignment was performed using database alignment in SYBYL with the common substructure asterisked in compound 37 (Fig. 1a). In this paper, the selection of atoms for molecules alignment was based on the common substructure of the studied compounds due to its effectiveness and easy implementation. The resulting alignment of the training set and the test set based on two different template conformations are shown in Fig. 2.

CoMFA and CoMSIA models

After the alignment of the molecules in the training set, molecular models were built. The CoMFA descriptors, steric and electrostatic field energies were calculated at each grid point using a sp^3 carbon probe atom with a

+1.0 charge and a van der Waals radius of 1.52 Å. using default settings in SYBYL. The steric and electrostatic interactions were calculated using the Tripos force field. All of the other parameters were as default settings. For the CoMSIA, five physicochemical properties, namely steric, electrostatic, hydrophobic, hydrogen bond donor and acceptor fields, were calculated using an sp^3 carbon probe atom with a +1.0 charge atom and a radius of 1.0 Å placed at regular grid spacing of 2 Å and a similar lattice box as used in CoMFA calculations. Here, steric indices were related to the third power of the atomic radii,

electrostatic descriptors were derived from atomic partial charges, hydrophobic fields were derived from atom-based parameters [35], and H-bond donor and acceptor indices were obtained by a rule-based method based on experimental results [36].

PLS regression analysis

The conventional CoMFA and CoMSIA descriptors derived above were used as independent variables, and pIC_{50} values were used as the dependent variable in partial

Table 1 The structures of the studied molecules and their experimental activity

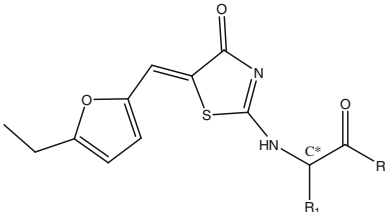
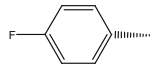
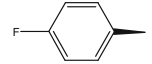
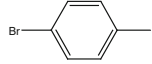
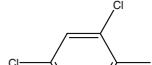
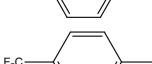
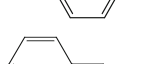
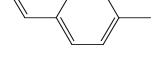
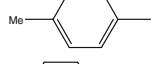
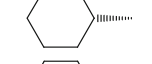
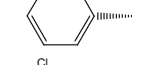
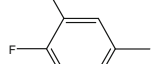
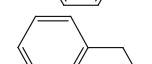
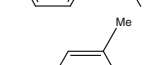
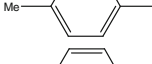
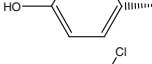
structure	Compd.	substituent		pIC_{50}
		R_1	R_2	
	1		-OH	5.52
	2		-OH	4.30
	3*		-OH	5.22
	4		-OH	5.12
	5*		-OH	4.68
	6		-OH	4.85
	7		-OH	4.96
	8		-OH	4.60
	9*		-OH	4.26
	10		-OH	4.62
	11		-OH	4.17
	12*		-OH	4.92
	13		-OH	4.37
	14		-NHOH	5.26
	15		-NH ₂	4.80

Table 1 continued

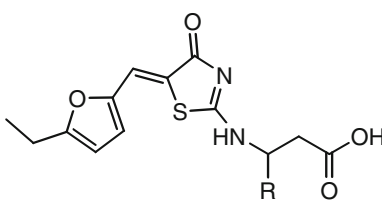
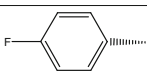
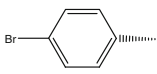
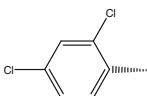
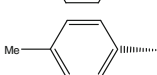
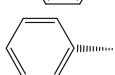
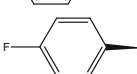
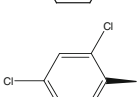
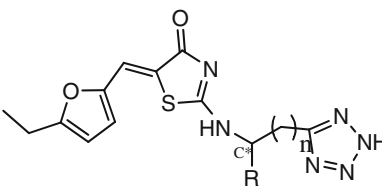
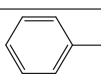
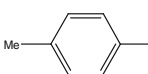
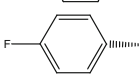
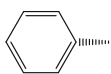
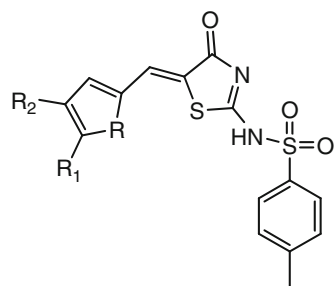
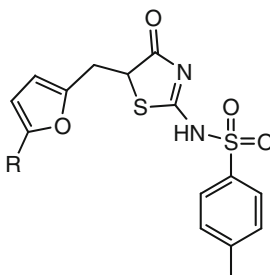
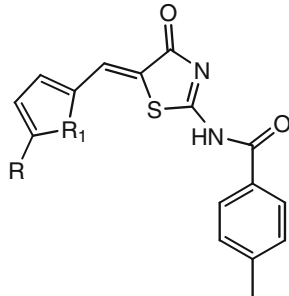
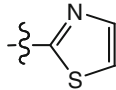
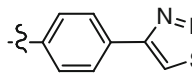
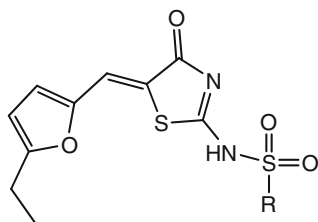
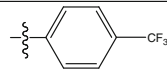
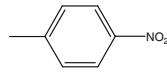
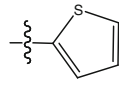
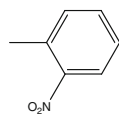
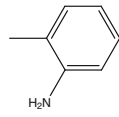
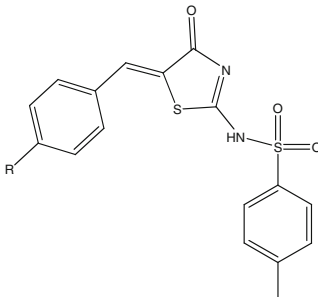
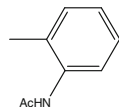
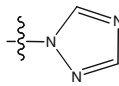
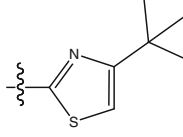
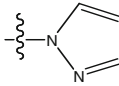
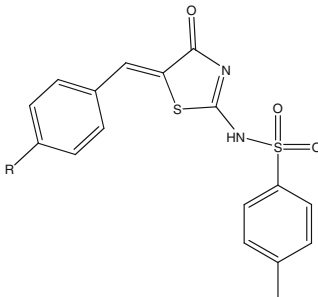
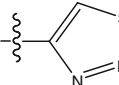
	16	<div>R</div> 		4.36	
	17			4.78	
	18			5.07	
	19*			4.57	
	20			4.23	
	21			4.89	
	22			5.05	
	23	<div>R</div> 		n	4.72
	24*			0	4.85
	25			0	5.01
	26			1	4.62
	27	R	R ₁	R ₂	4.96
	28	O	H	H	5.55
	29*	O	I	H	5.57
	30	O	Br	H	5.57
	31	O	Et	H	5.85
	32*	O	Ph	H	5.26
	33*	O	Me	Me	5.49
	34	S	H	H	4.92
	35	S	Cl	Me	5.00
	36	S	Me	H	4.85
	37	S	NO ₂	H	5.38
37	O	Me	H	5.70	
	38	<div>R</div>		H	4.13
	39			I	4.82
	40			Br	4.40
	41			Et	5.14

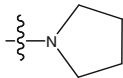
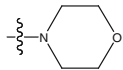
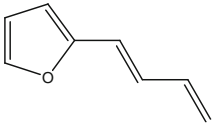
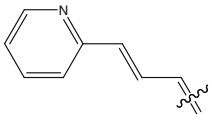
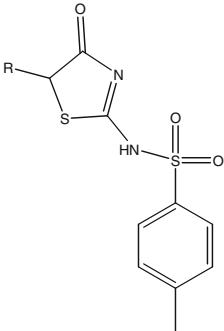
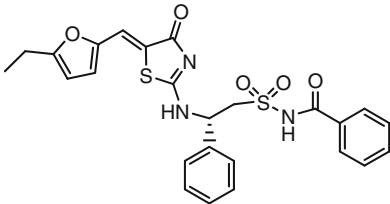
Table 1 continued

	R	R ₁	
	42	H	S 4.48
	43*	Me	O 4.80
	44	Et	O 4.89
	45		O 4.36
	46*		O 4.96
	47		5.49
	48		5.49
	49		5.10
	50		4.74
	51		4.96
	52*		4.47
	53		4.49
	54*		5.31
	55		5.00
	56		5.85

least squares (PLS) regression analyses to derive 3D-QSAR models using the standard implementation in the SYBYL package. The predictive value of the models was evaluated by leave-one-out (LOO) cross-validation. The

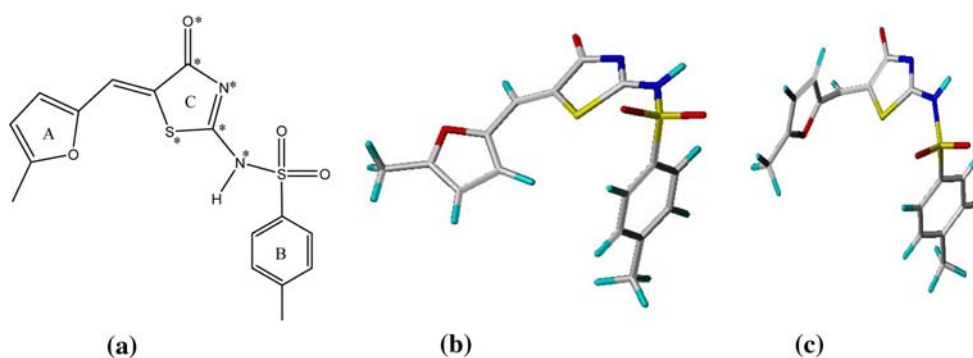
final model was constructed with the optimum number of components equal to that yielding the highest q^2 . The non-cross-validated correlation r^2 coefficient and standard error of estimate SEE (standard error of estimate) were

Table 1 continued

	57*		5.85																					
	58		4.92																					
<hr/>																								
		R																						
	59		5.30																					
	60		6.22																					
<hr/>																								
	61	<table><thead><tr><th>R₁₁</th><th>R₁₂</th><th>R₂</th></tr></thead><tbody><tr><td>F</td><td>H</td><td>4-Me</td></tr><tr><td>Cl</td><td>Cl</td><td>4-NO₂</td></tr><tr><td>F</td><td>H</td><td>3-NO₂</td></tr><tr><td>F</td><td>H</td><td>3-COOMe</td></tr><tr><td>F</td><td>H</td><td>3-COOH</td></tr><tr><td>F</td><td>H</td><td>4-NO₂</td></tr></tbody></table>	R ₁₁	R ₁₂	R ₂	F	H	4-Me	Cl	Cl	4-NO ₂	F	H	3-NO ₂	F	H	3-COOMe	F	H	3-COOH	F	H	4-NO ₂	5.03
R ₁₁	R ₁₂	R ₂																						
F	H	4-Me																						
Cl	Cl	4-NO ₂																						
F	H	3-NO ₂																						
F	H	3-COOMe																						
F	H	3-COOH																						
F	H	4-NO ₂																						
	62	Cl	Cl	4-NO ₂	5.18																			
	63*	F	H	3-NO ₂	4.82																			
	64	F	H	3-COOMe	4.85																			
	65*	F	H	3-COOH	4.92																			
	66	F	H	4-NO ₂	5.15																			
<hr/>																								
	67				4.80																			

*Test set

Fig. 1 (a) The structure of the template molecule, the asterisk (*) represents the atoms that are aligned in the model; (b) the lowest energy conformation as the Template 1; (c) the conformation of Template 2 extracted from the X-ray crystallographic (XRC) complex structure of the template molecule with the HCV NS5B polymerase

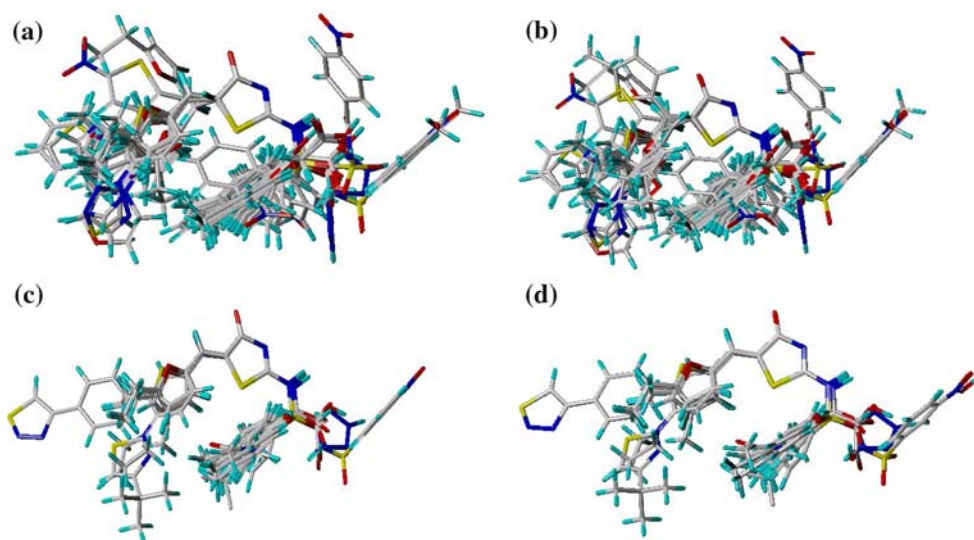


subsequently computed for the final PLS models. The bootstrapping procedure [37] was also used to validate each model. In this procedure, n random selections out of the training set of n objects are performed several times (100 times in this investigation) to simulate different

samplings from a larger set of objects. In each run some objects may not be included in the PLS analysis, and some might be included more than once.

To validate the built models, a test set of compounds with known activities but not used in model generation was used.

Fig. 2 Alignment of the data set: (a) Template 1 for the training set; (b) Template 2 for the training set; (c) Template 1 for the test set; (d) Template 2 for the test set



CoMFA and CoMSIA coefficient maps were generated by interpolation of the pairwise products between the PLS coefficients (coeff) and the standard deviations (stdev) of the corresponding CoMFA and CoMSIA descriptor values.

Results and discussion

CoMFA statistical results

The training data set of 51 compounds was aligned to derive both CoMFA and CoMSIA models. The models were generated with two different conformations of the same template molecule. A test set of 16 compounds was used to validate the reliabilities of the built models.

Two models were built based on CoMFA method. By use of the Template Conformation 1, a q^2 of 0.574 with six optimum number of components (ONC) based on both steric and electrostatic fields was observed. For the Template Conformation 2, a q^2 of 0.621 with five ONC based on both steric and electrostatic fields was obtained. The statistical parameters of the two models are shown in Tables 2 and 3. From these tables, it can be seen that the statistical parameters of Template 2 CoMFA model are obviously better than those of the Template 1 CoMFA model indicating that the former is more reliable and might have a better predictive ability than the latter. The better model gives a non-cross-validated correlation coefficient r^2 of 0.950, an F of 169.338, and a SEE_{tr} of 0.108, respectively. It also has an r^2_{boot} of 0.962 ± 0.015 and a SEE_{boot} of 0.091 ± 0.050 . The corresponding field contributions of steric and electrostatic are 56.1% and 43.9%. The detailed predicted results including the predicted pIC_{50} as well as the corresponding residuals for each compound of the training set are listed in Table 4.

We also analyzed the outliers in these models as shown in Table 4. Generally, if the residual value of a compound is greater than two times the standard error of the residuals, the compound is considered as an outlier [38]. In the Template 1 CoMFA model, there were five outliers (compounds 1, 30, 38, 39 and 59). While in the Template 2 CoMFA model, there were three outliers (compounds 30, 38 and 39). Compounds 30, 38 and 39 seem to be three outliers in both CoMFA and CoMSIA models. However, the residues of these three compounds are relatively small compared to some other similar works [39, 40] and less than the value of three times the standard error of the residuals.

To validate the predictive ability of the obtained models, the test set not included in the generation of the models was used. The predicted parameters (r^2_{pred} and SEE_{ts}) of the two models for the test set are also listed in Tables 2 and 3. The detailed predicted results are listed in Table 5. Obviously, the Template 2 CoMFA model gives a higher r^2_{pred} and a lower SEE_{ts} than Template 1 CoMFA model. It proves that Template 2 CoMFA model exhibits better performance in both the internal and external validation. The plot of the experimental versus the predicted pIC_{50} values for the training set and the test set by Template 2 CoMFA model is shown in Fig. 3.

CoMSIA statistical results

The leave-one-out (LOO) cross-validated correlation coefficient (q^2) values for the CoMSIA models based on different field/fields are available in Supporting Information. It is known that a good internal validation cannot guarantee a very good external validation [41]. For each Template conformation, the top six models (the parameters are shown in Tables 2 and 3) having good q^2 were selected to predict the test set so as to find out a model having good

Table 2 Statistical data for QSAR models with CoMFA and CoMSIA (Template 1)

	CoMFA			CoMSIA			
Fields ^a	S, E	S, E, A	A, S, H	H, D, A	S, E, H, A	H, D, A, S	S, E, H, D, A
q^2 ^b	0.574	0.621	0.648	0.631	0.644	0.651	0.640
ONC ^c	6	5	5	6	5	5	5
r^2 ^d	0.936	0.909	0.876	0.886	0.902	0.876	0.900
Outliers ^e	5	4					
SEE _{tr} ^f	0.122	0.145	0.169	0.164	0.15	0.169	0.151
F^g	108.019	89.948	63.483	56.946	82.712	63.795	81.258
r^2_{boot} ^h	0.964 ± 0.013	0.927 ± 0.023	0.903 ± 0.027	0.913 ± 0.029	0.926 ± 0.025	0.904 ± 0.024	0.947 ± 0.019
SEE _{boot} ⁱ	0.089 ± 0.054	0.128 ± 0.067	0.143 ± 0.069	0.140 ± 0.074	0.128 ± 0.065	0.144 ± 0.069	0.106 ± 0.059
r^2_{pred} ^j	0.534	0.787	0.611	0.516	0.717	0.632	0.705
SEE _{ts} ^k	0.297	0.201	0.272	0.303	0.231	0.264	0.237
<i>Field contribution</i>							
S	0.513	0.259	0.232		0.172	0.217	0.163
E	0.487	0.400			0.297		0.265
H			0.396	0.538	0.295	0.367	0.277
D				0.126		0.115	0.085
A		0.341	0.372	0.336	0.236	0.301	0.211

^aFields used: S = steric, E = electrostatic, H = hydrophobic, A = hydrogen bond acceptor, D = hydrogen bond donor

^bLeave-one-out (LOO) cross-validated correlation coefficient

^cOptimum number of components

^dNon-cross-validated correlation coefficient

^eResidual values greater than two times the standard error of the residuals are considered outliers

^fStandard error of estimate for the training set

^gRatio of r^2 to $1 - r^2$ (explained to unexplained)

^hAverage of correlation coefficient for 100 samplings using the bootstrapped procedure

ⁱAverage standard error of estimate for 100 samplings using the bootstrapped procedure

^jPredicted correlation coefficient for the test set

^kStandard error of estimate for the test set

performance in both internal and external validation. The predicted pIC_{50} values along with the corresponding residual values for each compound in the training set are listed in Table 4. By use of the Template 1, the best CoMSIA model gives a q^2 of 0.621 with five ONC based on steric, electrostatic and hydrogen bond acceptor fields. While for the CoMSIA models generated using the Template 2, the best one (bold in the Table 2) gives a q^2 of 0.685 with five ONC also based on steric, electrostatic and hydrogen bond acceptor fields. Compared with the results of CoMFA models, a similar trend can be concluded by comparison of the above CoMSIA models, i.e. the best Template 2 CoMSIA model is more reliable and seems to have better internal predictive ability than those based on the Template 1. The better model gives a non-cross-validated correlation coefficient r^2 of 0.940, an F of 142.136, and a SEE_{tr} of 0.117, respectively. It also has an r^2_{boot} of 0.944 ± 0.019 and a SEE_{boot} of 0.110 ± 0.058. The corresponding field contributions of steric, electrostatic and hydrogen bond acceptor are 28.2%, 39.1% and 32.7%,

respectively. The outliers in the models are also picked out to comparison (Table 4). In the best Template 1 CoMFA model, the outliers were 30, 37, 39 and 59. In the best Template 2 CoMSIA model, the outliers were 30, 38 and 39. Compounds 30, 38 and 39 seem to be three major outliers for almost all models.

The predicted parameters (r^2_{pred} and SEE_{ts}) of the models for the test set are also listed in Tables 2 and 3. It can be seen clearly that the best Template 2 CoMSIA model gives a higher r^2_{pred} and a lower SEE_{ts} and thus proves that this model exhibits the best performance considering both the internal and external validation. The detailed predicted pIC_{50} values and the corresponding residuals for each compound of the test set are listed in Table 5. The plot of the experimental pIC_{50} versus the predicted pIC_{50} values for the training set and the test set by best Template 2 CoMSIA model is shown in Fig. 4.

From the above discussion, it can be concluded that the CoMFA and CoMSIA models based on Template 2 are obviously superior to those based on Template 1, which is

Table 3 Statistical data for QSAR method with CoMFA and CoMSIA (Template 2)

	CoMFA			CoMSIA			
Fields ^a	S, E	S, E, A	A, S, H	S, E, H	S, E, H, A	H, D, A, S	S, E, H, D, A
q^2 ^b	0.621	0.685	0.689	0.677	0.697	0.671	0.682
ONC ^c	5	5	5	5	6	5	6
r^2 ^d	0.950	0.940	0.908	0.943	0.956	0.907	0.952
Outliers ^e	3	3					
SEE _{tr} ^f	0.108	0.117	0.146	0.115	0.102	0.146	0.106
F^g	169.338	142.136	88.37	148.684	159.67	87.744	145.227
r^2_{boot} ^h	0.962 ± 0.015	0.944 ± 0.019	0.936 ± 0.020	0.956 ± 0.018	0.964 ± 0.016	0.930 ± 0.022	0.961 ± 0.014
SEE _{boot} ⁱ	0.091 ± 0.050	0.110 ± 0.058	0.118 ± 0.064	0.097 ± 0.056	0.087 ± 0.053	0.122 ± 0.064	0.095 ± 0.052
r^2_{pred} ^j	0.685	0.822	0.634	0.774	0.766	0.675	0.76
SEE _{ts} ^k	0.244	0.184	0.263	0.207	0.211	0.248	0.213
<i>Field contribution</i>							
S	0.561	0.282	0.245	0.223	0.185	0.219	0.175
E	0.439	0.391		0.410	0.305		0.268
H			0.402	0.368	0.287	0.369	0.277
D						0.127	0.079
A		0.327	0.352		0.222	0.284	0.202

^aFields used: S = steric, E = electrostatic, H = hydrophobic, A = hydrogen bond acceptor, D = hydrogen bond donor

^bLeave-one-out (LOO) cross-validated correlation coefficient

^cOptimum number of components

^dNon-cross-validated correlation coefficient

^eResidual values greater than two times the standard error of the residuals are considered outliers

^fStandard error of estimate for the training set

^gRatio of r^2 to $1 - r^2$ (explained to unexplained)

^hAverage of correlation coefficient for 100 samplings using the bootstrapped procedure

ⁱAverage standard error of estimate for 100 samplings using the bootstrapped procedure

^jPredicted correlation coefficient for the test set

^kStandard error of estimate for the test set

reasonable and in agreement with our hypothesis. Our following discussions mainly concentrate on the models derived from Template 2.

CoMFA and CoMSIA contour maps

From the above discussion, by consideration of both the internal and external predictive power of the built models, the best CoMFA and CoMSIA selected to construct the stdev*coeff contour maps were CoMFA model and CoMSIA (SEA) models using Template Conformation 2. So our following discussion will focus on the 3D QSAR models derived using Template Conformation 2. The contour maps usually provide a detailed understanding of the key structural features required for the biological activity. For the CoMFA steric field, the green (sterically favorable) and yellow (sterically unfavorable) contours represent 80% and 20% level contributions, respectively. Similarly, the red (negative charge favorable) and blue (negative charge unfavorable) contours in the CoMFA electrostatic field

represent 80% and 20% level contributions, respectively. For the best CoMSIA model, only three fields are used and discussed in detail: steric, electrostatic and hydrogen bond acceptor. For the CoMSIA steric and electrostatic fields, the legends are the same as that in CoMFA. In the hydrogen bond receptor field, the magenta (favorable) and red (unfavorable) represent 85% and 15% level contributions, respectively. The CoMFA and CoMSIA contour maps are shown in Figs. 5 and 6, using Template conformation 2 as the reference structure.

The CoMFA steric contour map is shown in Fig. 5a. A large green region surrounding the 5-position of the A ring indicates that a bulky substituent is preferred in the position to produce higher inhibitory activity. This conclusion is consistent with the experiment results, such as the order for the activity of compounds 38–41: 38 (–H) < 40 (–Br) < 39 (–I) < 41 (–Et). This can also be shown by the activity of other compounds: 27 (–H) < 28 (–I) ≈ 29 (–Br) < 37 (–CH₃) < 30 (–Et), 43 (–CH₃) < 44 (–Et). However, the activity of compound 31 (–Ph) is lower than compound 37

Table 4 Experimental and predicted activities for the training set molecules

Compound	Exp pIC ₅₀	Template 1				Template 2			
		CoMFA		CoMSIA (SEA)		CoMFA		CoMSIA (SEA)	
		Pred.	Residual	Pred.	Residual	Pred.	Residual	Pred.	Residual
1	5.52	5.24	0.28*	5.27	0.25	5.60	−0.08	5.59	−0.07
2	4.30	4.26	0.04	4.24	0.06	4.31	−0.01	4.25	0.05
4	5.12	5.28	−0.16	5.25	−0.13	5.15	−0.03	5.16	−0.04
6	4.85	4.78	0.07	4.90	−0.05	4.68	0.17	4.80	0.05
7	4.96	5.17	−0.21	5.15	−0.19	4.99	−0.03	5.07	−0.11
8	4.60	4.62	−0.02	4.72	−0.12	4.57	0.03	4.68	−0.08
10	4.62	4.66	−0.04	4.55	0.07	4.57	0.05	4.55	0.07
11	4.17	4.24	−0.07	4.10	0.07	4.14	0.03	4.09	0.08
13	4.37	4.44	−0.07	4.38	−0.01	4.53	−0.16	4.38	−0.01
14	5.26	5.25	0.01	5.21	0.05	5.16	0.11	5.16	0.10
15	4.80	4.89	−0.09	4.83	−0.03	4.78	0.02	4.84	−0.04
16	4.36	4.38	−0.02	4.37	−0.01	4.42	−0.06	4.33	0.03
17	4.78	4.84	−0.06	4.85	−0.07	4.73	0.05	4.81	−0.03
18	5.07	5.06	0.01	5.04	0.03	4.98	0.09	5.01	0.06
20	4.23	4.17	0.06	4.24	−0.01	4.30	−0.06	4.25	−0.02
21	4.89	5.04	−0.15	4.94	−0.05	5.00	−0.11	4.99	−0.10
22	5.05	4.99	0.06	4.88	0.17	5.01	0.04	4.95	0.10
23	4.72	4.74	−0.02	4.85	−0.13	4.82	−0.10	4.85	−0.13
25	5.01	4.95	0.06	4.98	0.03	4.95	0.06	4.93	0.08
26	4.62	4.60	0.02	4.61	0.01	4.63	0.00	4.63	0.00
27	4.96	4.91	0.05	5.03	−0.07	4.95	0.01	5.01	−0.05
28	5.55	5.67	−0.12	5.37	0.18	5.63	−0.08	5.37	0.18
30	5.85	5.61	0.24*	5.57	0.28*	5.51	0.34*	5.56	0.29*
31	5.26	5.28	−0.02	5.38	−0.12	5.32	−0.06	5.43	−0.17
34	5.00	5.03	−0.03	5.09	−0.09	5.01	−0.01	5.11	−0.11
35	4.85	4.88	−0.03	5.12	−0.27	4.81	0.04	4.82	0.03
36	5.38	5.36	0.02	5.32	0.06	5.37	0.01	5.33	0.05
37	5.70	5.55	0.15*	5.40	0.30*	5.67	0.03	5.68	0.02
38	4.13	4.39	−0.26*	4.34	−0.21	4.37	−0.24*	4.39	−0.26*
39	4.82	4.53	0.29*	4.45	0.37*	4.49	0.33*	4.50	0.32*
40	4.40	4.45	−0.05	4.44	−0.04	4.43	−0.03	4.49	−0.09
41	5.14	5.08	0.06	5.00	0.14	5.12	0.02	4.97	0.17
42	4.48	4.56	−0.08	4.42	0.06	4.59	−0.11	4.46	0.02
44	4.89	4.84	0.05	4.80	0.09	4.82	0.07	4.78	0.11
45	4.36	4.31	0.05	4.42	−0.06	4.37	−0.01	4.41	−0.05
47	5.49	5.57	−0.08	5.62	−0.13	5.52	−0.03	5.60	−0.11
48	5.49	5.51	−0.02	5.43	0.06	5.48	0.01	5.40	0.09
49	5.10	5.11	−0.01	5.33	−0.23	5.13	−0.03	5.31	−0.21
50	4.74	4.70	0.04	4.76	−0.02	4.76	−0.02	4.74	0.00
51	4.96	4.82	0.14	4.96	0.00	4.81	0.16	4.97	−0.01
53	4.49	4.45	0.04	4.59	−0.10	4.51	−0.02	4.59	−0.10
55	5.00	4.95	0.05	4.96	0.04	5.19	−0.19	5.00	0.00
56	5.85	5.96	−0.11	5.76	0.09	5.90	−0.05	5.70	0.15
58	4.92	4.95	−0.03	5.00	−0.08	4.98	−0.06	5.01	−0.09
59	5.30	5.61	−0.31*	5.61	−0.31*	5.28	0.02	5.41	−0.11
60	6.22	6.08	0.14	6.18	0.04	6.26	−0.04	6.28	−0.06

Table 4 continued

Compound	Exp pIC ₅₀	Template 1				Template 2			
		CoMFA		CoMSIA (SEA)		CoMFA		CoMSIA (SEA)	
		Pred.	Residual	Pred.	Residual	Pred.	Residual	Pred.	Residual
61	5.03	5.06	−0.03	5.01	0.02	5.03	0.00	4.97	0.06
62	5.18	5.12	0.06	5.12	0.06	5.27	−0.09	5.16	0.02
64	4.85	4.84	0.01	4.83	0.02	4.92	−0.07	4.84	0.01
66	5.15	5.07	0.08	5.16	−0.01	5.09	0.06	5.16	−0.01
67	4.80	4.80	0.00	4.84	−0.04	4.78	0.02	4.88	−0.08

* outliers

Table 5 Experimental and predicted activities for the test set molecules

Compound	Exp pIC ₅₀	Template 1				Template 2			
		CoMFA		CoMSIA (SEA)		CoMFA		CoMSIA (SEA)	
		Pred.	Residual	Pred.	Residual	Pred.	Residual	Pred.	Residual
3	5.22	5.22	0.00	5.19	0.03	5.06	0.16	5.09	0.13
5	4.68	4.41	0.27	4.53	0.15	4.57	0.11	4.53	0.15
9	4.26	4.21	0.05	4.31	−0.05	4.36	−0.10	4.36	−0.10
12	4.92	5.20	−0.28	5.09	−0.17	5.09	−0.17	5.01	−0.09
19	4.57	4.42	0.15	4.24	0.33	4.40	0.17	4.26	0.32
24	4.85	4.85	0.00	4.86	−0.01	4.91	−0.06	4.86	−0.01
29	5.57	5.62	−0.05	5.36	0.21	5.54	0.03	5.36	0.21
32	5.49	5.26	0.23	5.32	0.17	5.16	0.33	5.28	0.21
33	4.92	4.91	0.01	5.00	−0.08	4.95	−0.03	5.04	−0.12
43	4.80	4.83	−0.03	4.63	0.17	4.70	0.10	4.60	0.20
46	4.96	4.37	0.59	4.53	0.43	4.40	0.56	4.56	0.40
52	4.47	4.55	−0.08	4.59	−0.12	4.55	−0.08	4.61	−0.14
54	5.31	4.69	0.62	5.11	0.20	4.97	0.34	5.15	0.16
57	5.85	5.08	0.77	5.46	0.39	5.28	0.57	5.65	0.20
63	4.82	5.01	−0.19	5.01	−0.19	5.02	−0.19	5.01	−0.19
65	4.92	4.90	0.02	4.90	0.02	4.95	−0.03	4.92	0.00

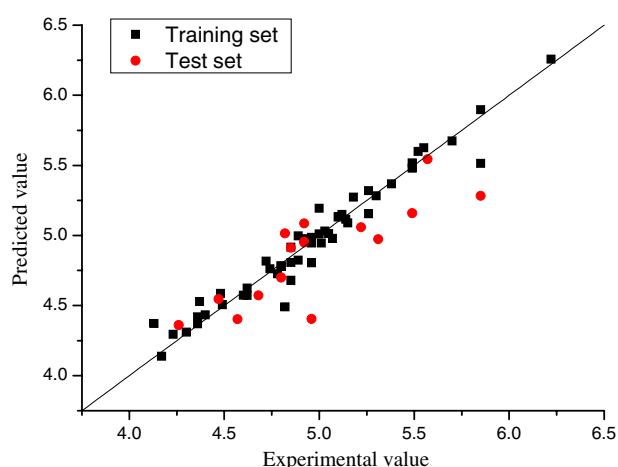
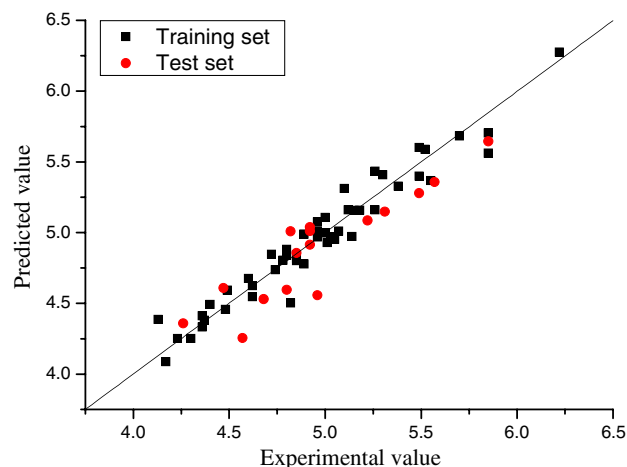
**Fig. 3** The experimental pIC₅₀ versus the predicted pIC₅₀ values for the training set and the test set based on Template 2 CoMFA model**Fig. 4** The experimental pIC₅₀ versus the predicted pIC₅₀ values for the training set and the test set based on Template 2 CoMSIA SEA model

Fig. 5 CoMFA contour maps shown in the presence of the Template 2. **(a)** Contour maps for steric field; **(b)** contour maps for electrostatic field

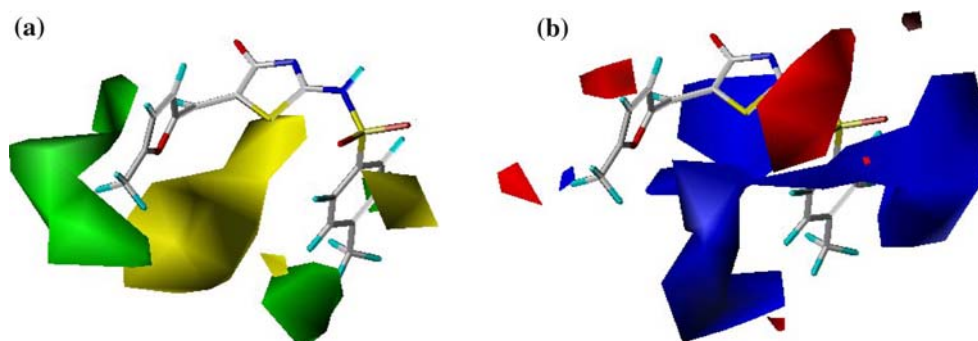
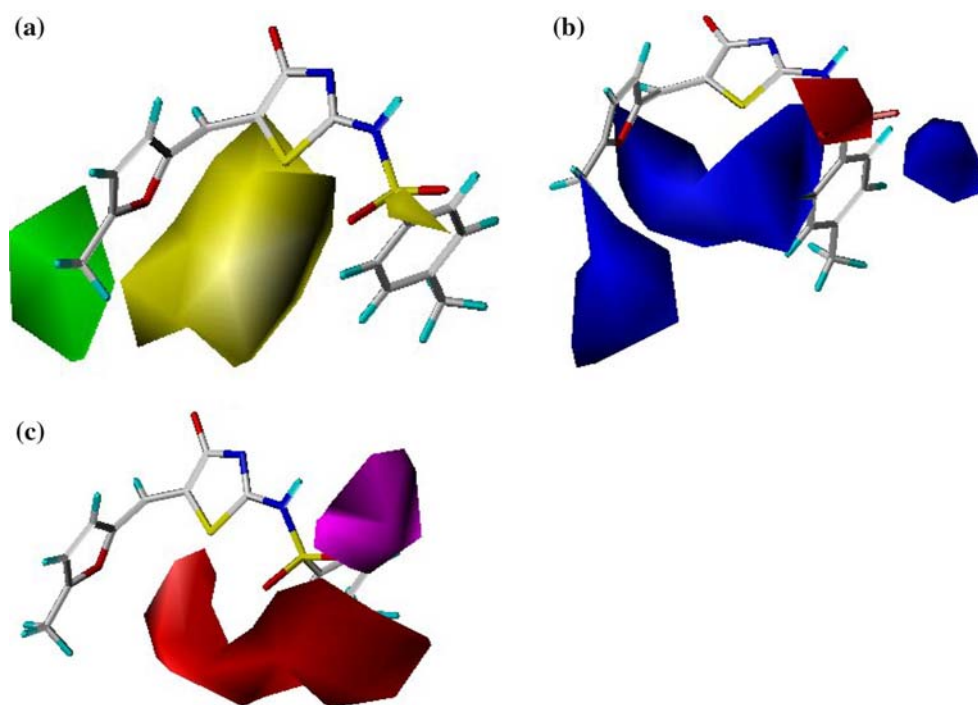


Fig. 6 CoMSIA contour maps shown in the presence of the Template 2. **(a)** Contour maps for steric field; **(b)** contour maps for electrostatic field; **(c)** contour maps for H-bond acceptor field



($-\text{CH}_3$), it is because the phenyl ring is closer to the large yellow contour. There is another small green region near the *para*-position of the B ring implying that a relatively bulky substituent is beneficial to the inhibitory activity. A large region of yellow contour close to the A ring and C ring suggest that bulky group would decrease the activity. In addition to compound 31 whose phenyl ring is closer to the large yellow contour leading to lower activity, there is another lower activity compound 11 whose benzyl substituent enters into the large yellow region. Compounds 38–41 are conformationally flexible molecules, due to the orientation of the flexible molecules, the furan ring of the four compounds enters into the large yellow contour map resulting in consistently low activity of compounds 38–41 compared to compounds 27–30. There is another small yellow region above the sulfonyl. In this region there is no substituent in all compounds. However, it is in agreement with the binding mode. The O atoms on the sulfonyl have

important H-bond interaction with the allosteric site, so a substituent on that position maybe impacts the interaction or decreases the binding affinity.

The CoMFA electrostatic contour map is shown in Fig. 5b. A favorable negative red contour region is found around the sulfonyl. This region indicates that electronegative groups could have a positive influence on the inhibitory activity. Most active compounds have the sulfonyl substituent. As the experimental data shown, compounds 1–26 and 42–46 which have a carbonyl on that position are much less active than other compounds with a sulfonyl on that position. Two large blue regions near the two *ortho*-position of B ring suggest that electropositive substituent would increase the inhibitory activity. This is in agreement with the fact that the binding affinities of compounds 47 and 48 with $-\text{H}$ group on the position are higher than that of the compounds 50 ($-\text{NO}_2$), 51 ($-\text{NH}_2$) and 52 ($-\text{NHAc}$). There is a third blue region near the terminal of the 2-position of the A

ring which means that an electropositive substituent on the position would enhance the activity.

The CoMSIA steric and electrostatic contour map shown in Fig. 6a and b are generally in accordance with the CoMFA map (Fig. 5) with few minor differences: The CoMFA steric contour map has two small green contours near the benzene ring while the CoMSIA steric contour map not. There has a small red contour behind the furyl ring in the CoMFA electrostatic contour map, but the CoMSIA electrostatic contour map has nothing at that position. Other locations of the contours are coincident although the shapes have some differences. These differences can be attributed to the difference of the description method of fields.

The CoMSIA hydrogen bond acceptor contour map is shown in Fig. 6c. The magenta region indicates that H-bond acceptor group enhances activity, which is supported by the importance of the existence of sulfonyl and carbonyl to the activity in the region. In fact, there are two hydrogen bonds between the sulfonyl and the residues Arg 501 according to the XRC complex structure (PDB: 2HWH). The interaction is shown in Fig. 8. In contrast, the red contour surrounding the *ortho*-position of B ring represent regions where H-bond acceptor group decreases activity. It can also be explained by the order of the activity for compound 50 ($-\text{NO}_2$) < 51 ($-\text{NH}_2$).

Yan et al. [30, 31] have discussed stereochemistry of the inhibitors in relation to the observed IC_{50} values and concluded that (*S*)-enantiomer binds preferentially to the allosteric pocket over (*R*)-enantiomer. This conclusion can be proved by the CoMSIA hydrogen bond acceptor contour map in this work. The compound 16 and 21 are a pair of enantiomers depicted in Fig. 7 in the presence of the hydrogen bond acceptor contour map of the CoMSIA model. The carboxyl of compound 16 ((*R*)-enantiomer) is close to the red contour which means the presence of the carboxyl decreases activity, yet the carboxyl group of compound 21 ((*S*)-enantiomer) is close to the magenta region indicating the carboxyl enhances activity. It can explain the fact that (*S*)-enantiomer shows better inhibiting activity than (*R*)-enantiomer.

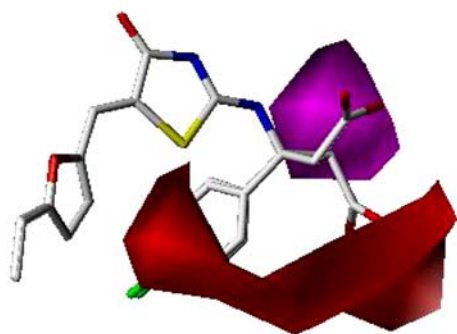
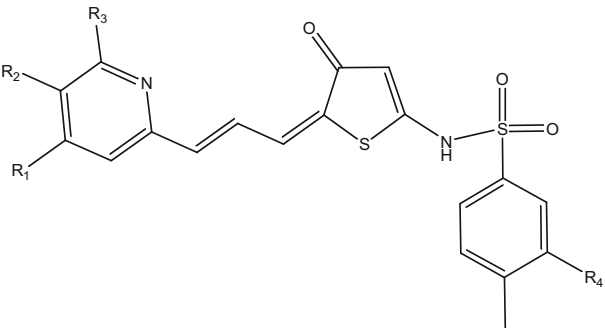


Fig. 7 H-bond acceptor contour map of CoMSIA model shown in the presence of compounds 16 and 21

Design of new molecules and prediction of their activity

CoMFA and CoMSIA methods have been widely used not only for investigating the structural factors related to the biological activity but also for designing and predicting new compounds with better activity. An accurate prediction of activity of newly designed compounds can be performed by the computation of binding free energies, but it is complicated and time-consuming. In this study, several analogues of the most potent molecule (compound 60) were designed based on the information of the best Template 2 CoMSIA model contour maps due to its high predictive power. Four important substituent positions were identified and shown in Table 6. Firstly, the favorable steric green contour and the electrostatic blue contour near the pyridine ring suggests that the substitution of electropositive groups at R_1 , R_2 and R_3 positions can increase activity. Secondly, there is a large electrostatic blue contour near the benzene ring, substituent of electron-withdrawing group at R_4 position can increase the electropositivity of 5 and 6 on the benzene ring. Thirdly, from the interaction between compound 37 and the HCV NS5B polymerase (Fig. 8), it can be seen that there exist two Arg (Arg 501 and Arg 422) close to the benzene ring, the

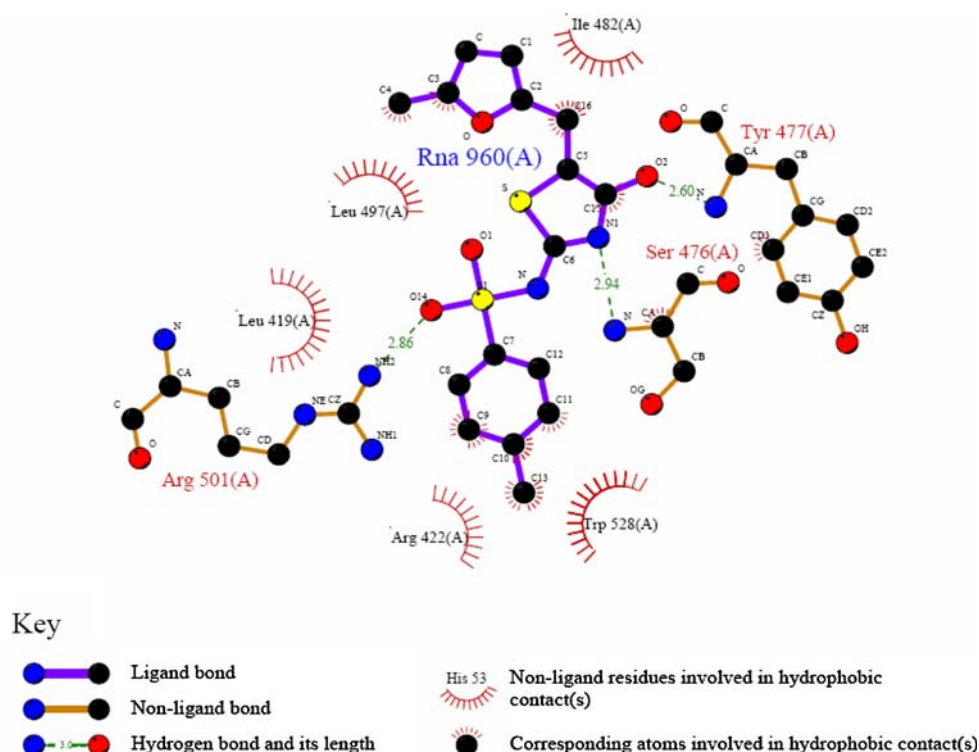
Table 6 Structures and predicted pIC_{50} values of the designed compounds (60a–60h)



Compound	R_1	R_2	R_3	R_4	pIC_{50}^a
60	H	H	H	H	6.28
60a	Me	H	H	H	6.42
60b	H	Et	H	H	6.38
60c	H	H	Me	H	6.36
60d	H	H	H	NO_2	6.50
60e	Me	H	H	NO_2	6.64
60f	H	Et	H	NO_2	6.65
60g	Me	Et	H	NO_2	6.73
60h	Me	Et	Me	NO_2	6.81

^a pIC_{50} values predicted by the best Template 2 CoMSIA model

Fig. 8 Schematic representation of interactions between compound 37 and the HCV NS5B polymerase according to the X-ray crystallographic (XRC) complex structure (PDB: 2HWH) produced using the Ligplot program developed by Wallace et al. [42]



polar H on the Arg could have H-bond interaction with the H-bond acceptor on the substituent of the benzene ring. The R_4 substituent containing H-bond acceptor atoms could increase activity. According to the above discussion, the most potent compound 60 was performed through substitution at the R_1 – R_4 positions. All the designed analogues have increased predicted pIC_{50} values compared to the reference compound 60. The predicted pIC_{50} values of the designed compounds are listed in Table 6.

Conclusion

In this study, CoMFA and CoMSIA models have been built to explain the structure-activity relationship of a series of thiazolone derivatives. Two different template conformations were used to study the impact on the model predictive ability. In addition, CoMSIA models based on different field/fields were performed to find the best combination of fields. The best CoMFA and CoMSIA models were obtained based on the template conformation extracted from X-ray crystallographic (XRC) complex and the best CoMSIA model was based on SEA fields. The internal predictive abilities of the both best models were satisfactory. The external predictive ability of the best CoMSIA model ($r^2_{pred} = 0.822$) was much better than the best CoMFA model due to the importance of the contribution of the hydrogen bond acceptor field. All of the new designed molecules based on the best CoMSIA model have shown improved pIC_{50} values.

The results of this study can provide further support to the design of thiazolone derivatives acting as potential inhibitors for hepatitis C virus NS5B polymerase.

Acknowledgement This work was supported by the Program for New Century Excellent Talents in University (No. NCET-07-0399).

References

- Matsuura Y (1995) *Virus* 45:105
- Choo QL, Kuo G, Weiner LK, Overby LR, Bradley D, Houghton M (1989) *Science* 244:359. doi:10.1126/science.2523562
- Alter MJ, Kruszon-Moran D, Nainan OV, McQuillan GM, Gao F, Moyer LA et al (1999) *N Engl J Med* 341:556. doi:10.1056/NEJM199908193410802
- Richard ETS (2006) *Nat Rev Drug Discov* 5:715. doi:10.1038/nrd2134
- Manns MP, McHutchison JG, Gordon SC, Rustgi VK, Shiffman M, Reindollar R et al (2001) *Lancet* 358:958. doi:10.1016/S0140-6736(01)06102-5
- Wu JZ, Yao NH, Walker M, Hong Z (2005) *Mini Rev Med Chem* 5:1103. doi:10.2174/138955705774933310
- Hirashima S, Suzuki T, Ishida T, Noji S, Yata S, Ando I et al (2006) *J Med Chem* 49:4721. doi:10.1021/jm060269e
- Beaulieu PL, Bousquet Y, Gauthier J, Gillard J, Marquis M, McKercher G et al (2004) *J Med Chem* 47:6884. doi:10.1021/jm040134d
- Patel PD, Patel MR, Kaushik-Basu N, Talele TT (2008) *J Chem Inf Model* 48:42. doi:10.1021/ci700266z
- Gopalsamy A, Chopra R, Lim K, Ciszewski G, Shi M, Curran KJ et al (2006) *J Med Chem* 49:3052. doi:10.1021/jm060168g
- Harper S, Avolio S, Pacini B, DiFilippo M, Altamura S, Tomei L et al (2005) *J Med Chem* 48:4547. doi:10.1021/jm050056±

12. Harper S, Pacini B, Avolio S, DiFilippo M, Migliaccio G, Laufer R et al (2005) *J Med Chem* 48:1314. doi:[10.1021/jm049122i](https://doi.org/10.1021/jm049122i)
13. Gopalsamy A, Lim K, Ciszewski G, Park K, Ellingboe JW, Bloom J et al (2004) *J Med Chem* 47:6603. doi:[10.1021/jm0401255](https://doi.org/10.1021/jm0401255)
14. Stansfield I, Pompei M, Conte I, Ercolani C, Migliaccio G, Jairaj M et al (2007) *Bioorg Med Chem Lett* 17:5143. doi:[10.1016/j.bmcl.2007.06.093](https://doi.org/10.1016/j.bmcl.2007.06.093)
15. Summa V, Petrocchi A, Pace P, Matassa VG, DeFrancesco R, Altamura S et al (2004) *J Med Chem* 47:14. doi:[10.1021/jm0342109](https://doi.org/10.1021/jm0342109)
16. DiSanto R, Fermeglia M, Ferrone M, Paneni MS, Costi R, Artico M et al (2005) *J Med Chem* 48:6304. doi:[10.1021/jm0504454](https://doi.org/10.1021/jm0504454)
17. Kim J, Han JH, Chong Y (2006) *Bull Korean Chem Soc* 27:1919
18. Chan L, Reddy TJ, Proulx M, Das SK, Pereira O, Wang W et al (2003) *J Med Chem* 46:1283. doi:[10.1021/jm0340400](https://doi.org/10.1021/jm0340400)
19. Louise-May S, Yang W, Nie X, Liu D, Deshpande MS, Phadke AS et al (2007) *Bioorg Med Chem Lett* 17:3905. doi:[10.1016/j.bmcl.2007.04.103](https://doi.org/10.1016/j.bmcl.2007.04.103)
20. Zhou Y, Li LS, Webber S, Ayida B, Bertolini T, Sun Z et al (2007) *Antiviral Res* 74:A51. doi:[10.1016/j.antiviral.2007.01.067](https://doi.org/10.1016/j.antiviral.2007.01.067)
21. Zhou Y, Li LS, Webber S, Dragovich P, Murphy D, Tran C et al (2007) *Antiviral Res* 74:A38. doi:[10.1016/j.antiviral.2007.01.035](https://doi.org/10.1016/j.antiviral.2007.01.035)
22. Pfefferkorn JA, Nugent R, Gross RJ, Greene M, Mitchell MA, Reding MT et al (2005) *Bioorg Med Chem Lett* 15:2812. doi:[10.1016/j.bmcl.2005.03.106](https://doi.org/10.1016/j.bmcl.2005.03.106)
23. Pfefferkorn JA, Greene ML, Nugent RA, Gross RJ, Mitchell MA, Finzel BC et al (2005) *Bioorg Med Chem Lett* 15:2481. doi:[10.1016/j.bmcl.2005.03.066](https://doi.org/10.1016/j.bmcl.2005.03.066)
24. Yan S, Appleby T, Larson G, Wu JZ, Hamatake R, Hong Z et al (2006) *Bioorg Med Chem Lett* 16:5888. doi:[10.1016/j.bmcl.2006.08.056](https://doi.org/10.1016/j.bmcl.2006.08.056)
25. Cramer RD, Patterson DE, Bunce JD (1988) *J Am Chem Soc* 110:5959. doi:[10.1021/ja00226a005](https://doi.org/10.1021/ja00226a005)
26. Klebe G, Abraham U, Mietzner T (1994) *J Med Chem* 37:4130. doi:[10.1021/jm00050a010](https://doi.org/10.1021/jm00050a010)
27. Zhang ZY, An LY, Hu WX, Xiang YH (2007) *J Comput Aided Mol Des* 21:145. doi:[10.1007/s10822-006-9090-y](https://doi.org/10.1007/s10822-006-9090-y)
28. Avery MA, Muraleedharan KM, Desai PV, Bandyopadhyaya AK, Furtado MM, Tekwani BL (2003) *J Med Chem* 46:4244. doi:[10.1021/jm030181q](https://doi.org/10.1021/jm030181q)
29. Ding Y, Smith KL, Varaprasad CVNS, Chang E, Alexander J, Yao N (2007) *Bioorg Med Chem Lett* 17:841. doi:[10.1016/j.bmcl.2006.08.104](https://doi.org/10.1016/j.bmcl.2006.08.104)
30. Yan S, Appleby T, Larson G, Wu JZ, Hamatake RK, Hong Z et al (2007) *Bioorg Med Chem Lett* 17:1991. doi:[10.1016/j.bmcl.2007.01.024](https://doi.org/10.1016/j.bmcl.2007.01.024)
31. Yan S, Larson G, Wu JZ, Appleby T, Ding Y, Hamatake R et al (2007) *Bioorg Med Chem Lett* 17:63. doi:[10.1016/j.bmcl.2006.09.095](https://doi.org/10.1016/j.bmcl.2006.09.095)
32. Jain AN, Koile K, Chapman D (1994) *J Med Chem* 37:2315. doi:[10.1021/jm00041a010](https://doi.org/10.1021/jm00041a010)
33. Labrie P, Maddaford SP, Fortin S, Rakhit S, Kotra LP, Gaudreault RC (2006) *J Med Chem* 49:7646. doi:[10.1021/jm060239b](https://doi.org/10.1021/jm060239b)
34. Sybyl 6.9 (1999) Tripos Associates, Inc., St. Louis
35. Viswanadhan VN, Ghose AK, Revankar GR, Robins RK (1989) *J Chem Inf Comput Sci* 29:163. doi:[10.1021/ci00063a006](https://doi.org/10.1021/ci00063a006)
36. Klebe G (1994) *J Mol Biol* 237:212. doi:[10.1006/jmbi.1994.1223](https://doi.org/10.1006/jmbi.1994.1223)
37. Vong R, Geladi P, Wold S, Esbensen K (1988) *J Chemometr* 2:281. doi:[10.1002/cem.1180020406](https://doi.org/10.1002/cem.1180020406)
38. Bang SJ, Cho SJ (2004) *Bull Korean Chem Soc* 25:1525
39. Assefa H, Kamath S, Buolamwini JK (2003) *J Comput Aided Mol Des* 17:475. doi:[10.1023/B:JCAM.0000004622.13865.4f](https://doi.org/10.1023/B:JCAM.0000004622.13865.4f)
40. Sekhar YN, Nayana MRS, Sivakumari N, Ravikumar M, Mahmood SK (2008) *J Mol Graph Model* 26:1338. doi:[10.1016/j.jmgm.2008.01.008](https://doi.org/10.1016/j.jmgm.2008.01.008)
41. Golbraikh A, Tropsha A (2002) *J Mol Graph Model* 20:269. doi:[10.1016/S1093-3263\(01\)00123-1](https://doi.org/10.1016/S1093-3263(01)00123-1)
42. Wallace AC, Laskowski RA, Thornton JM (1995) *Protein Eng* 8:127. doi:[10.1093/protein/8.2.127](https://doi.org/10.1093/protein/8.2.127)