DOI 10.5012/bkcs.2010.31.10.2761

Articles

Ligand Based CoMFA, CoMSIA and HQSAR Analysis of CCR5 Antagonists

Changdev G. Gadhe,[†] Sung Haeng Lee,[†,‡] Thirumurthy Madhavan,[†] Gugan Kothandan,[†] DuBok Choi,[§] and Seung Joo Cho[†,‡,*]

[†]*Department of Bio-New Drug Development, College of Medicine, Chosun University, Gwangju 501-759, Korea*
[*]*E-mail: chosj@chosun.ac.kr*
[‡]*Department of Cellular · Molecular Medicine and Research Center for Resistant Cells, College of Medicine, Chosun University, Gwangju 501-759, Korea*
[§]*Biotechnology Lab, BK Company R&D Center, Jeonbuk 573-400, Korea*
Received June 24, 2010, Accepted August 19, 2010

In this study, we have developed QSAR models for a series of 38 piperidine-4-carboxamide CCR5 antagonists using CoMFA, CoMSIA and HQSAR methods. Developed models showed good statistics in terms of q^2 and r^2 values. Best predictions obtained with standard CoMFA model ($r^2 = 0.888$, $q^2 = 0.651$) and combined CoMSIA model ($r^2 = 0.892$, $q^2 = 0.665$) with electrostatics and H-bond acceptor parameter. The validity of developed models was assessed by test set of 9 compounds, which showed good predictive correlation coefficient for CoMFA (0.804) and CoMSIA (0.844). Bootstrapped analysis showed statistically significant and robust CoMFA (0.968) and CoMSIA (0.936) models. Best HQSAR model was obtained with a q^2 of 0.662 and r^2 of 0.936 using atom, connection, hydrogen, donor and acceptor as parameters and fragment size (7-10) with optimum number of 6 components. Predictive power of developed HQSAR model was proved by test set and it was found to be 0.728.

Key Words: CCR5, QSAR, CoMFA, CoMSIA, HQSAR

Introduction

Human Immunodeficiency Virus type-1 (HIV-1) infection, which eventually leads to the acquired immunodeficiency syndrome (AIDS), was first discovered by Barrae *et al* in 1983.[1] At present, AIDS remains to be a lethal disease threatening human's health especially sub-Saharan Africa and Southeast Asia. The chemokine receptor CCR5 has proven to be an exciting target for pharmaceutical industry in the HIV-1 and inflammation therapeutic areas. This fact inspired a great amount of research in last decade to identify anti HIV-1 therapeutic by targeting the CCR5 mediated entry mechanism. This effort recently resulted in FDA approval for small molecular inhibitor, Maraviroc (Selzentry) for the treatment of HIV-1 infection.[2] HIV primarily infects vital cells in the human immune system such as helper T-cells (specifically CD4+ T-cells), macrophages and dendrite cells.[3-5] HIV enters into these cells by adsorption of viral glycoprotein to receptors on the target cell. This step is followed by the fusion of viral envelope with the cell membrane and release of the HIV capsid into the cell.[6-7] HIV infection leads to low levels of CD4+ T-cells through three main mechanisms: Firstly, direct viral killing of infected cells; secondly, increased rates of apoptosis in infected cells; and thirdly, killing of infected CD4+ T-cells by CD8 cytotoxic lymphocytes that recognize infected cells. CCR5 receptor is a member of G-Protein coupled receptors (GPCR) superfamily.[8,9] It has been identified as a primary co-receptor on CD4+ cells for entry of macrophage-tropic (M-tropic or R5) HIV-1 isolates.[10]

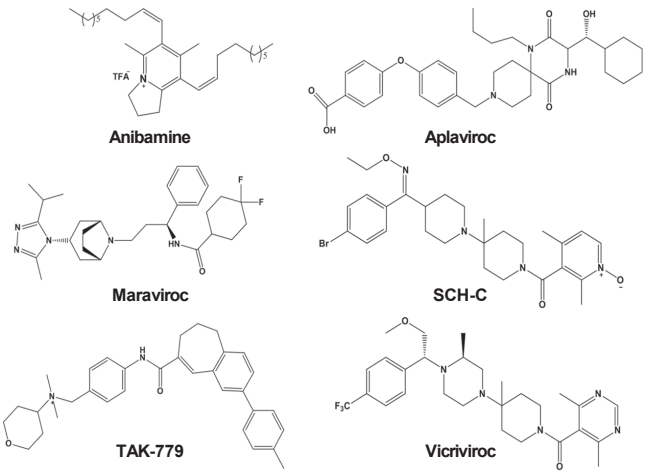


Figure 1. Chemical structures of several known CCR5 antagonists.

Few natural ligands (MIP-1a, MIP-1b and RANTES) have been identified for CCR5.[11] Currently, there is an only few small molecular inhibitors are being investigated as anti-HIV-1 agents in human clinical trials[12-17] and shown in Figure 1. To date, only one drug Maraviroc (Selzentry) was approved by FDA on August 6th 2007. Still, they raised concerns that it may be associated with serious complications such as liver damage, lymphoma, infections and heart attack.

Literature survey showed that previously 3D-QSAR and virtual screening studies of CCR5 antagonists 1-(3,3-diphenyl-

propyl)-piperidinyl amides have been performed by Aher *et al*. and Afantitis *et al*.[18,19] They shows that for high affinity binders key chemical and structural requirement can be identified using physicochemical parameter, topological property and 3D field such as steric, electrostatics, hydrophobic, hydrogen bond donor/acceptor around a set of aligned ligand molecules. 3D-QSAR models, CoMFA and CoMSIA on a series of piperidine-based CCR5 antagonists have been developed by Song *et al*.[20] Whereas, Y. Zhuo *et al* performed 3D-QSAR study for 1,3,4-trisubstituted pyrrolidine based CCR5 inhibitors.[21] An excellent work docking and 3D-QSAR study on 1-amino-2-phenyl-4-(piperidin-1-yl)-butanes derivatives was done by Xu *et al*.,[22] where they developed homology model for CCR5, and docking based conformations of all molecules were used for 3D-QSAR model generation.

There has been need for developing newer, selective and potent inhibitors of CCR5. A successful 3D-QSAR model not only helps in better understanding of structure activity relationship data for any class of compounds, but also ensure researcher an in depth analysis of lead compounds in further studies. Here, we have chosen a series of piperidine-4-carboxamide derivatives containing potent and selective CCR5 functional antagonists.[23] Hence, we started CoMFA,[24] CoMSIA[25] and HQSAR[26] studies as validation tools for QSAR analysis of CCR5 inhibitors. CoMFA and CoMSIA are the methods which are used in rational drug design. To develop novel antagonistic agents with high potency and selectivity, detailed information is required about structural features influencing biological activity towards CCR5 receptor. Since the crystal structure of CCR5 or its complex with any ligand is not available, the interactions between the CCR5 receptor and its antagonists are not known. Correlation between the structural properties of compounds with their biological activities is believed to gain an insight into the interaction mechanisms of CCR5 to antagonists. This could provide useful information for designing novel anti-HIV-1 drugs. CoMFA and CoMSIA analyses involve alignment of molecules in a structurally and pharmacologically reasonable manner on the basis of assumption that each molecule acts *via* common macromolecular target binding site. Since there is no experimental structure was reported, HQSAR analysis was also done which does not require-3D structure, bioactive conformation and molecular alignment. It generates QSAR equation by molecular fragments fingerprints as independent variable and biological activity as dependent variable.

Computational Details

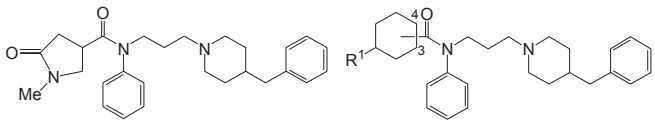
Dataset. Dataset used in this study comprised of piperidine-4-carboxamide derivatives reported as potent CCR5 inhibitors by Imamura *et al.*[23] The binding affinities of given antiretroviral concentration in nano-molar (nM) range were converted to the molar (M) concentration and then converted to logarithmic scale for further analyses on dataset, using the following formula.

$$pIC_{50} = -\log IC_{50}$$

Where, IC_{50} is concentration of antagonist to provide 50% inhibition of CCR5 receptor. Consequently we used 38 piperidine-4-carboxamide derivatives with their activities (Table 1-4) to establish QSAR models. Total dataset was divided randomly into a training set of 29 compounds and test set (wide range of activity) of 9 compounds.

Molecular modeling. All computational studies were performed using molecular modeling package SYBYL 8.1[27] installed on a Linux system. Since crystal structure of CCR5 with any ligand is not available, the most active molecule 5h (IC_{50} = 0.29 nM, pIC_{50} = 9.53) was considered as a template molecule and random search method was carried out to obtain least energy conformation. Least energy conformation was considered as a

Table 1. Structures and biological activity values of piperidine-4-carboxamide derivatives (Series-1)

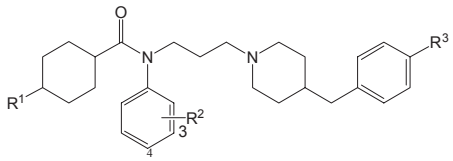


1

Compound	R^1	Position	pIC_{50}
1			6.31
5a*	Cbz	4	7.32
6a	H	4	5.37
7a*	Ac	4	7.78
7b	*i*-PrCO	4	7.41
7c	Bz	4	7.48
7d	Ms	4	7.40
5b	Cbz	3	5.92
6b	H	3	6.10
7e*	Ac	3	6.17
7f*	Ms	3	6.74

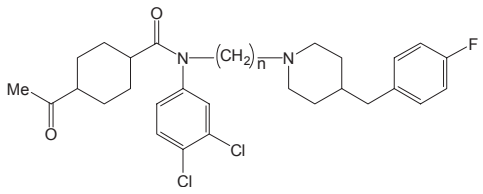
Asterisk (*) indicates test set compounds.

Table 2. Structures and biological activity values of piperidine-4-carboxamide derivatives (Series-2)



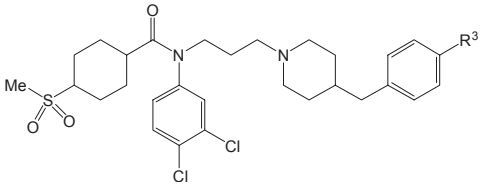
Compound	R^1	R^2	R^3	pIC_{50}
5c	Ac	3-Cl	F	8.52
5d	Ac	4-Me	F	8.55
5e	Ac	3,4-diCl	H	8.72
5f	Ac	3,4-diCl	F	8.92
5g	Ac	3-Cl, 4-Me	H	9.20
5h	Ac	3-Cl, 4-Me	F	9.54
5i	Ac	3-Cl, 4-*i*-Pr	F	7.74
5j	Ac	3-Cl, 4-MeO	F	7.52
5k	Ms	3,4-diCl	F	8.48
5l	Ms	3-Cl, 4-Me	F	9.30

Table 3. Structures and biological activity values of piperidine-4-carboxamide derivatives (Series-3)



Compound	n	pIC_{50}
12	4	7.28
18	2	6.31

Table 4. Structures and biological activity values of piperidine-4-carboxamide derivatives (Series-4)



Compound	R^3	pIC_{50}
11a	CF_3	8.06
11b*	NO_2	8.62
13*	NH_2	8.03
14a	NHAc	8.23
14b	NHMs	8.66
11c	Morpholino	8.55
11d	OMe	8.10
11e	SMe	8.51
11f	SO_2Me	8.66
11g*	SO_2Et	8.72
11h	SO_2i-Pr	8.82
11i*	SO_2NH_2	8.47
11j	SO_2NHMe	8.82
11k*	SO_2NMe_2	8.92
11l	SO_2 (Morpholino)	8.88

Asterisk (*) indicates test set compounds.

bioactive conformation and then minimized it by applying Tripos force field with Powell method and conjugate gradient minimizer. The minimization was terminated when the energy gradient convergence criterion of 0.05 kcal/mol was reached or when the 5000 step minimization cycle was exceeded. Rests of the molecules were built by changing required substitution on the template molecule and minimized similarly as template molecule by keeping core part constrained. Finally Gasteiger-Hückel charges were applied to all the molecules of dataset and it was subsequently used for QSAR studies.

Structural alignment. Molecular alignment is one of the most important and sensitive parameter in 3D-QSAR. Least energy conformation of template molecule (5h) was considered as a bioactive conformation and other molecules of the dataset were aligned on it. Atom fit method was used to align all molecules of dataset and shown in Figure 2.

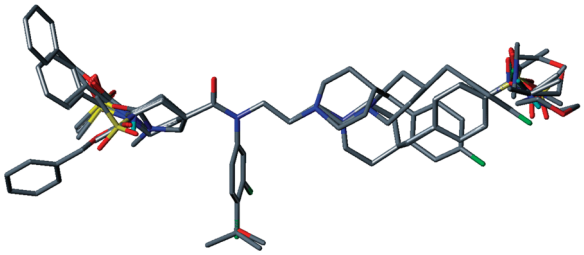


Figure 2. Molecular alignment of all molecules over template molecule (5h).

CoMFA and CoMSIA interaction energy. For the alignment set, the steric and electrostatic CoMFA field was calculated at each lattice intersection of a regularly spaced grid of 2.0 Å unit. The grid box dimensions were determined automatically in such a way that the region boundaries were extended beyond 4.0 Å in each direction from the co-ordinates of each molecule. The steric and electrostatic interactions between probe and remaining molecules were calculated. The generated steric and electrostatic fields were scaled by CoMFA_STD scaling method in SYBYL with default energy of +30 kcal/mol. The electrostatic interactions are modeled by Columbic potential and van der Waals interaction using Lennard-Jones potential respectively, and calculated by using standard Tripos force field. The distance dependent dielectric constant 1.00 was used. A sp^3 hybridized carbon atom having +1 charge serves as a probe for calculating steric and electrostatic field. Any singularities were avoided at atomic position in CoMSIA field because a Gaussian type distance dependence of physicochemical properties was adopted; thus no arbitrary cutoff was required. Similarly indices were calculated as 0.3 as attenuation factor (α) for Gaussian type distance.

RF CoMFA. Region focusing is an application of weights to the lattice point in a CoMFA region to enhance or to attenuate the contribution of points to subsequent analysis. When the weights are StdDev*Coefficient values, the process is exactly equivalent to image enhancement of the derived CoMFA maps. The sharpness of focusing is controlled by a user provided exponential factor.

HQSAR. Hologram QSAR (HQSAR) is a technique which employs fragment fingerprints as predictive variables of biological activity or other structural related data.[26] HQSAR doesn't require 3D structure of bioactive conformation or molecular alignments. HQSAR model generation deals with the 2D structure directed fragment fingerprints.[28] These molecular fingerprints are broken into strings at fixed intervals as specified by a hologram length (HL) parameter. The hologram length determines the number of bins in the hologram into which the fragments are hashed. Each corresponding fragment SYBYL line notation (SLN) is then mapped to a pseudo-random integer in the range (50 - 500) using cyclic redundancy check (CRC) algorithm. The integer generated by the CRC algorithm is unique and reproducible for each and every unique SLN string. The hashing then occurs by folding the pseudorandom integer for a particular SLN string into the bin range defined. In HQSAR, bins contain

information about the number of fragments hashed into each bin. The optimal HQSAR model was derived from screening through the 12 default HL values, which were a set of 12 prime numbers (53, 59, 61, 71, 83, 97, 151, 199, 257, 307, 353 and 401). The pattern of fragment counts from dataset molecules are then correlate to observed biological dataset using partial least squares analysis. The model development was performed using parameters *viz*., atom (A), bond (B), connection (C), chirality (Ch), hydrogen (H), and donor/acceptor (DA). The validity of the model depends on statistical parameters such as non-cross-validated r^2, cross-validated r^2_{cv} (q^2) by LOO, predictive r^2_{pred}, and standard error. The robustness of the model depends on the more challenging test set prediction reflected by its predictive r^2_{pred} value.

PLS analysis. The regression analysis was carried out by using partial least square (PLS) method[29-31] in CoMFA and CoMSIA. The cross-validation analysis was performed using leave-one-out (LOO) method, wherein one compound was removed from the dataset and its activity was predicted using the model derived from the rest of the molecules. To speed up analysis and to reduce the noise, the column with values (σ) below 2.0 kcal/mol was filtered off. Subsequently, PLS analyses were performed without cross validation, using optimum number of components (ONC) obtained from LOO method. The q^2 that resulted in optimum number of components and lowest standard error of prediction (SEP) were considered for further analyses and calculated using following formula:

$$q^2 = 1 - \frac{\sum_{\gamma} (\gamma_{pred} - \gamma_{actual})^2}{\sum_{\gamma} (\gamma_{actual} - \gamma_{mean})^2}$$

where, γ_{pred}, γ_{actual} and γ_{mean} are predicted, actual, and mean values of the target property (pIC_{50}) respectively. Equal weights for CoMFA were assigned to steric and electrostatic fields using CoMFA_STD scaling option. To derive 3D-QSAR models CoMFA and CoMSIA descriptors were used as an independent variable and pIC_{50} activity value as dependent variable. To further assess the robustness and statistical confidence of the models, bootstrapping analysis for 1000 runs was performed.

Predictive correlation coefficient (r^2_{pred}). The predictive ability of generated QSAR models were determined by using a test set of 9 compounds that were excluded during model generation. The energy minimization and geometry optimization of these 9 molecules is as same as the training set compounds described above, and their activity was predicted by using the model generated from training set. The predictive correlation coefficient (r^2_{pred}), based on test set molecules is computed by using following formula.

$$r^2_{pred} = \frac{(SD - PRESS)}{SD}$$

Where, SD is sum of squared deviation between the biological activity of test set molecules and mean activity of training set molecules. PRESS is predictive residual sum of square and is calculated by taking difference in predicted and actual activity of test set molecules. For all conventional analysis (non cross validation) the 'minimum sigma' standard deviation threshold was set to 2.0 kcal/mol.

Results and Discussions

CoMFA and CoMSIA analysis. We derived 3D-QSAR CoMFA and CoMSIA models for piperidine-4-carboxamide derivatives with good predictivity in terms of q^2 and r^2 values. The derived standard CoMFA model shows better statistics ($q^2 = 0.651$, $r^2 = 0.888$ with an ONC = 4), and combined CoMSIA model with electrostatics and hydrogen bond acceptor fields ($q^2 = 0.665$, $r^2 = 0.892$ with an ONC = 5) indicates good predictions. For both CoMFA and CoMSIA models the standard error of estimate (SEE) was found to be 0.396 and 0.397, respectively. After region focusing, a q^2 value for CoMFA model (0.753) was distinctly increased, because of the weight for those meaningful grid points are increased and noise is decreased. Further robustness of developed CoMFA and CoMSIA models was assessed by bootstraps analysis with standard deviation. For CoMFA model, BS-r^2 was found to be 0.968 with 0.017 SD and for CoMSIA model BS-r^2 was 0.936 with SD = 0.020. The statistical ANOVA test produced good result for CoMFA (47.54) and CoMSIA (38.00). The predictive ability of CoMFA and CoMSIA models depends upon the contribution of steric, electrostatic and H-bond acceptor field. The steric and electrostatic contribution for CoMFA model was found to be 50.7% and 49.3%. CoMSIA model was build with electrostatics and H-bond acceptor fields with contributions 66.9% and 33.1% respectively. The field contribution value for CoMFA indicates that both steric and electrostatic field contributes almost similar for the model development. But, in case of CoMSIA the electrostatic field has more impact than H-bond acceptor field and it is resulted in higher q^2 value. The regression analysis for both CoMFA and CoMSIA models were given in Table 5. Actual and predicted values along with residual values for CoMFA and CoMSIA training set and test set molecules were summarized in Table 6.

HQSAR analysis. HQSAR model generation was carried out using several parameters such as fragment distinction, fragment size and best hologram length (BHL). The HQSAR models were generated using default fragment size (4-7) combined with various hologram length and fragment types. Table 7 summarizes the results for different fragment type and hologram length. With the best fragment type parameters, PLS analyses were performed to investigate whether different fragment sizes could improve statistical results or not. The HQSAR results from different fragment sizes are summarized in Table 8. The best model ($q^2 = 0.662$, $r^2 = 0.936$) obtained with BHL = 97, NOC = 6 and with A/C/H/DA fragment types. Influence of various fragment sizes shows that larger fragment size was favored for improving statistical results in the form of q^2 value. However increment in q^2 is significant when fragment sizes changes from 4-7 to 8-11. The increment found in q^2 was 0.101. We chosen final model for HQSAR analyses with higher q^2 and lower SEE values as summarized in Table 8. The final model was built using A/C/H/DA key parameter, and 7-10 as the fragment size with 0.314 as SEE. Low residual values obtained for developed HQSAR model indicates its reliability, and can be

Table 5. PLS summary of standard CoMFA model and combined CoMSIA model with different parameters

Models	N	q^2	r^2	r^2_{pred}	SEE	F-value	BS-r^2	SD	Contributions (%)				
									S	E	A	H	D
CoMFA	**4**	**0.651**	**0.888**	**0.804**	**0.396**	**47.54**	**0.968**	**0.017**	**50.7**	**49.3**	**-**	**-**	**-**
RF	6	0.753	0.981	0.657	0.169	191.4	0.994	0.004	63.8	36.2	-	-	-
S	3	0.431	0.779	0.916	0.545	29.32	0.851	0.033	100	-	-	-	-
E	5	0.651	0.908	0.850	0.367	45.35	0.953	0.025	-	100	-	-	-
A	6	0.622	0.784	0.704	0.574	13.33	0.853	0.035	-	-	100	-	-
H	3	0.496	0.805	0.749	0.512	34.34	0.887	0.045	-	-	-	-	-
SE	3	0.606	0.825	0.950	0.485	39.32	0.896	0.062	28.2	71.8	-	-	-
EA	**5**	**0.665**	**0.892**	**0.844**	**0.397**	**38.00**	**0.936**	**0.020**	**-**	**66.9**	**33.1**	**-**	**-**
AH	4	0.580	0.858	0.632	0.447	36.12	0.911	0.024	-	-	50.0	50.0	-
EH	4	0.626	0.916	0.837	0.343	65.32	0.941	0.017	-	65.7	-	34.3	-
SEA	4	0.620	0.853	0.852	0.454	34.68	0.913	0.035	17.4	48.4	34.2	-	-
SEH	3	0.598	0.848	0.837	0.452	46.51	0.909	0.022	17.4	55.0	-	27.6	-
SED	4	0.507	0.840	0.866	0.474	31.40	0.892	0.049	20.7	55.5	-	-	23.9
EAH	5	0.640	0.925	0.912	0.332	56.46	0.955	0.021	-	44.0	29.1	26.9	-
AHD	5	0.561	0.857	0.533	0.457	27.54	0.912	0.040	-	-	36.3	48.0	15.7
SEAH	3	0.615	0.851	0.780	0.447	47.70	0.911	0.029	12.5	37.7	28.7	21.1	-
SEHD	4	0.516	0.874	0.852	0.420	41.59	0.913	0.037	17.3	43.4	-	21.4	17.9
SAHD	5	0.537	0.873	0.821	0.430	31.70	0.925	0.048	20.7	-	33.4	34.1	11.8
EAHD	5	0.605	0.895	0.717	0.392	39.07	0.943	0.029	-	39.2	22.9	22.6	15.3
SEAHD	5	0.581	0.899	0.849	0.385	40.73	0.944	0.021	11.9	34.0	20.2	17.9	16.0

N = optimum number of components as determined by the PLS leave one out cross validation study, q^2 = cross validated correlation coefficient by LOO, r^2 = conventional correlation coefficient, r^2_{pred} = predictive correlation coefficient, SEE = standard error of estimate, r^2-bs = correlation coefficient after 1000 runs of bootstrapping, SD-bs= standard deviation for 1000 runs of bootstrapping, S = steric, E = electrostatics, A = hydrogen bond acceptor, H = hydrophobic, D = hydrogen bond donor, Models used for CoMFA and CoMSIA analysis are shown in bold font.

Table 6. Actual and predicted pIC_{50} values of training set and test set with residual values by CoMFA and CoMSIA analysis

Compound	pIC_{50}	Predicted pIC_{50}		Residual Value		Compound	pIC_{50}	Predicted pIC_{50}		Residual Value	
		CoMFA	CoMSIA	CoMFA	CoMSIA			CoMFA	CoMSIA	CoMFA	CoMSIA
Training set						5h (temp)	9.53	8.89	8.94	0.64	0.59
1	6.31	6.25	6.63	0.06	−0.32	5i	7.74	8.48	8.68	−0.74	−0.94
11a	8.06	8.48	8.23	−0.42	−0.17	5j	7.52	8.33	8.09	−0.81	−0.57
11c	8.55	8.63	8.15	−0.08	0.40	5k	8.48	8.48	8.30	0.00	0.18
11d	8.10	8.39	8.15	−0.29	−0.05	5l	9.30	8.86	8.70	0.44	0.60
11e	8.51	8.49	8.44	0.02	0.07	6a	5.37	5.87	5.55	−0.50	−0.18
11f	8.66	8.59	8.56	−0.07	0.10	6b	6.10	5.87	5.64	0.23	0.46
11h	8.82	8.72	8.67	0.10	0.15	7b	7.41	7.54	7.31	−0.13	0.10
11j	8.82	8.70	9.02	0.12	−0.20	7c	7.48	7.71	7.73	−0.23	−0.25
11l	8.88	8.83	9.02	0.05	−0.14	7d	7.40	7.79	7.89	−0.39	−0.49
12	7.28	6.98	7.24	0.30	0.04	**Test set**					
14a	8.23	8.40	8.39	−0.17	−0.16	11b	8.62	8.28	8.78	0.34	−0.16
14b	8.66	8.75	8.79	−0.09	−0.13	11g	8.72	8.61	9.83	0.11	−0.41
18	6.31	5.79	6.20	0.52	0.11	11i	8.47	8.28	8.65	0.19	−0.18
5b	5.92	5.98	5.93	−0.06	−0.01	11k	8.92	8.46	9.00	0.46	−0.08
5c	8.52	8.34	8.90	0.18	−0.38	13	8.03	8.31	8.08	−0.28	−0.05
5d	8.55	8.74	8.55	−0.19	0.00	5a	7.32	7.32	7.75	0.00	−0.43
5e	8.72	8.30	8.34	0.42	0.38	7a	7.78	7.95	8.24	−0.17	−0.46
5f	8.92	8.52	8.60	0.40	0.32	7e	6.17	5.87	5.85	0.30	0.32
5g	9.20	8.66	8.69	0.54	0.51	7f	6.74	5.81	6.10	0.93	0.64

Table 7. Results of HQSAR analyses for various fragment distinctions on key statistical parameters using default fragment sizes (4-7)

Model	Fragment distinction	N	q^2	StdErr	r^2	SEE	BHL
1	A/B	4	0.511	0.827	0.898	0.378	199
2	A/C	4	0.517	0.822	0.935	0.302	257
3	B/C	3	0.469	0.845	0.831	0.476	307
4	A/B/C	2	0.420	0.865	0.681	0.642	257
5	A/B/Ch	4	0.544	0.790	0.910	0.354	353
7	C/H/Ch	3	0.544	0.783	0.810	0.506	353
8	A/C/DA	4	0.527	0.813	0.867	0.431	257
9	A/B/C/H	3	0.520	0.803	0.773	0.552	97
10	A/B/H/Ch	3	0.505	0.815	0.815	0.498	151
11	**A/C/H/DA**	**5**	**0.561**	**0.804**	**0.912**	**0.358**	**97**
12	B/C/Ch/DA	4	0.497	0.839	0.867	0.431	307
13	A/B/C/Ch/DA	2	0.469	0.829	0.645	0.677	257
14	A/B/C/H/Ch	3	0.526	0.798	0.774	0.551	97
15	A/B/C/H/DA	5	0.497	0.857	0.904	0.357	257
16	A/B/C/H/Ch/DA	5	0.520	0.837	0.895	0.391	199

A-atom; B-bond; C-connection; Ch-chirality; DA-donor/acceptor; Best model is highlighted in bold font.

Table 8. HQSAR analyses for the various fragment sizes on the key statistical parameters using best model (A/C/H/DA)

Atom Counts	N	q^2	StdErr	r^2	SEE	BHL
1-4	1	0.396	0.867	0.543	0.754	307
2-5	1	0.427	0.844	0.554	0.745	151
3-6	5	0.524	0.834	0.925	0.330	353
4-7	5	0.561	0.804	0.896	0.389	97
5-8	5	0.581	0.783	0.878	0.422	199
6-9	5	0.582	0.782	0.876	0.426	353
7-10	**6**	**0.662**	**0.719**	**0.936**	**0.314**	**97**
8-11	5	0.630	0.735	0.894	0.394	97

The final model chosen for HQSAR analysis is highlighted in bold font.

used to predict biological activity of novel compounds.

Predictive ability of CoMFA, CoMSIA and HQSAR models. The predictive ability of QSAR models was validated by a test set of 9 compounds, which were excluded during model generation. Both CoMFA and CoMSIA models showed good predictive power in terms of r^2_{pred} and, which is equal to 0.804 and 0.844, respectively. The graph of predicted *versus* actual activities for training set and test set of CoMFA and CoMSIA models are shown in Figure 3A and 3B. The developed HQSAR model shows good predictive power of 0.728. The detailed predicted *versus* actual activities along with residual values for training set and test set was depicted in Table 9 and plotted in Figure 4.

Interpretation of contour maps. The different density maps of the steric, electrostatic and H-bond acceptor parameters of CoMFA and CoMSIA models are represented as 3D coefficient contour maps. The contour maps surround all lattice points where the QSAR is found to be strongly associate with changes in the molecular field values (which basically mean changes in structure) as well as in binding affinity or any other measure of biological property. According to the standard SYBYL settings,

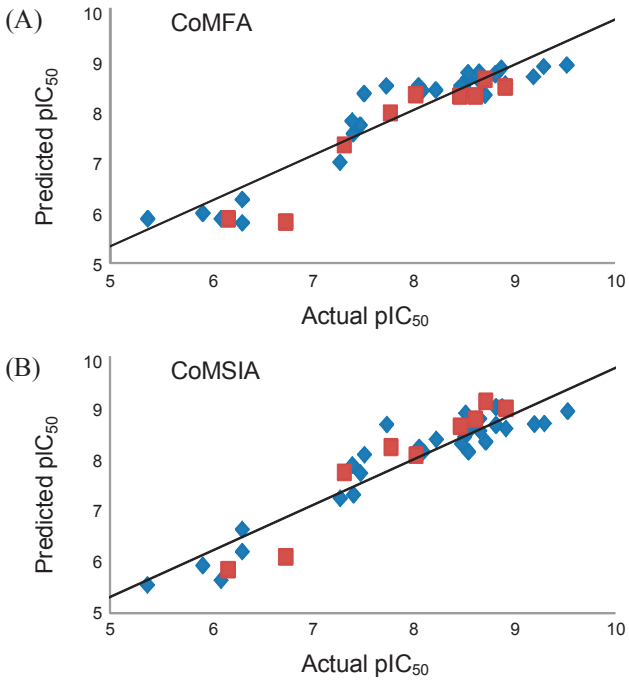


Figure 3. Scatter plot diagram of predicted *versus* actual activity of training set (blue diamond) and test set (red square) compounds by CoMFA (A) and CoMSIA (B) analysis.

steric interactions are represented by green and yellow colored contours while electrostatic interactions are represented as red and blue contours. Green contours stand for points where the Lennard-Jones potential has to be increased by appropriate groups to increase the biological activity whereas the yellow contours are used to underline the points where such a potential has to be decreased by suitable substituent's to correlate with increased binding affinity. The electrostatic red color shows the regions where the presence of a negative charge is expected

Table 9. HQSAR analyses for the various fragment sizes on the key statistical parameters using best model (A/C/H/DA)

Compound	Actual pIC_{50}	Predicted pIC_{50}	Residual Value
	Training set		
1	6.31	6.42	−0.11
11a	8.06	8.40	−0.34
11c	8.55	8.17	0.38
11d	8.10	8.35	−0.25
11e	8.51	8.41	0.10
11f	8.66	8.62	0.04
11h	8.82	8.94	−0.12
11j	8.82	8.80	0.02
11l	8.88	9.06	−0.18
12	7.28	7.09	0.19
14a	8.23	8.38	−0.15
14b	8.66	8.54	0.12
18	6.31	6.36	−0.05
5b	5.92	5.88	0.04
5c	8.52	8.47	0.05
5d	8.55	8.89	−0.34
5e	8.72	8.79	−0.07
5f	8.92	8.47	0.45
5g	9.20	9.07	0.13
5h (temp)	9.53	8.93	0.60
5i	7.74	8.02	−0.28
5j	7.52	8.29	−0.77
5k	8.48	8.41	0.07
5l	9.30	8.87	0.43
6a	5.37	5.36	0.01
6b	6.10	5.92	0.18
7b	7.41	7.61	−0.20
7c	7.48	7.37	0.11
7d	7.40	7.50	−0.10
	Test set		
11b	8.62	8.41	0.21
11g	8.72	8.64	0.08
11i	8.47	8.74	−0.27
11k	8.92	8.77	0.15
13	8.03	8.36	−0.33
5a	7.32	6.22	1.10
7a	7.78	7.16	0.62
7e	6.17	6.36	−0.19
7f	6.74	7.08	−0.34

to enhance the activity, whereas the blue color indicates regions where placing more positive charge is expected to correlate with increased binding affinity. In case of H-bond acceptor field magenta color denotes where the presence of acceptor atom is expected to increase the activity, whereas red color indicates decrease in inhibitory activity. In CoMFA contour map template molecule (5h) was shown in background. 3D-CoMFA contour map for best fitted model with steric field for template (IC_{50} = 0.29 nM, pIC_{50} = 9.53) molecule was shown in Figure 5. The

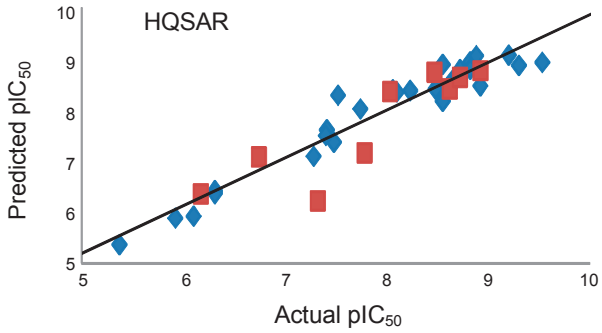


Figure 4. Scatter plot diagram of predicted *versus* actual activity of training set (blue diamond) and test set (red square) compounds by HQSAR analysis.

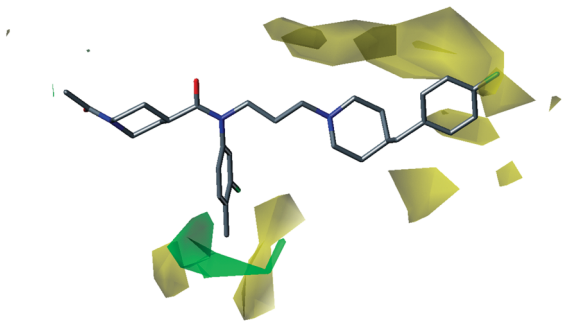


Figure 5. Steric stdev* coefficient contour map for highly active compound (5h) generated by CoMFA analysis, where sterically favored/disfavored areas are shown in green/yellow color.

presence of green colored contour on central phenyl ring near the meta and para position indicates that, this position is favorable for small bulky group like Cl and Me, and increases binding potency. In case of compound 5i (IC_{50} = 18 nM, pIC_{50} = 7.74) and 5j (IC_{50} = 30 nM, pIC_{50} = 7.52), the bulkier substitution at para position of central phenyl ring leads to decrease in binding affinity. The two yellow contours near para position indicate that, sterically bulkier groups (4-*i*-Pr and 4-MeO) are unfavorable for inhibitory activity. The alkyl linker plays a crucial role in demonstrating the inhibitory potency. It is clear that the increased or decreased in alkyl chain length in compound 14 (IC_{50} = 52 nM, pIC_{50} = 7.28) and in compound 18 (IC_{50} = 480 nM, pIC_{50} = 6.31) leads to reduced binding potency. Variation in alkyl chain length resulted in change in conformation of 4-benzylpiperidine moiety. It might be because of steric hindrance with receptor cavity which is indicated by the presence of yellow contours near 4-benzylpieridine moiety. At R^3 position introduction of -CF_3 group in compound 11a (IC_{50} = 8.7 nM, pIC_{50} = 8.01) resulted in decrease in CCR5 binding potency, but introduction of more polar group like NO_2 in compound 11b (IC_{50} = 2.4 nM, pIC_{50} = 8.62) retained the inhibitory potency. It suggests that polar groups are preferable at this position. The alkylsulfonyl (11f-11h) and aminosulfonyl (11i-11l) derivatives exhibit good activity, which indicate larger groups, had better effect on the activity (11h, k, and l).

CoMFA electrostatic contour map for template molecule (5h) is shown in Figure 6. Blue contour near para position of

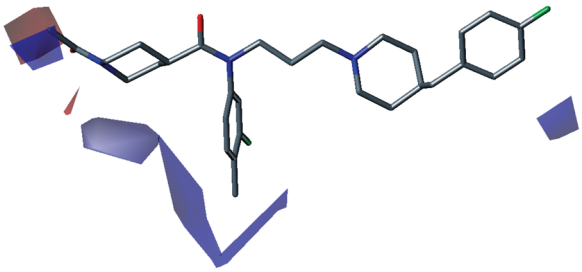


Figure 6. Electrostatics stdev* coefficient contour map for highly active compound (5h) generated by CoMFA analysis, where blue/red polyhedra depict the favorable site for positively/negatively charged groups.

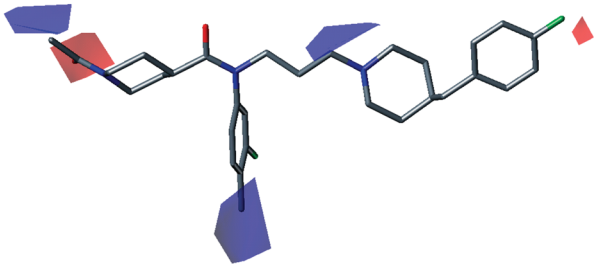


Figure 7. Electrostatics stdev* coefficient contour map for highly active compound (5h) generated by CoMSIA analysis, where blue/red polyhedra depict the favorable site for positively/negatively charged groups.

central phenyl ring, indicates that the small electropositive substituent are favorable for good inhibitory potency. The electropositive substituent's like 4-*i*-Pr in compound 5i (IC_{50} = 18 nM, pIC_{50} = 7.74) and polar 4-MeO in compound 5j (IC_{50} = 30 nM, pIC_{50} = 7.74) resulted in slightly loss of binding potency. It indicates that the central phenyl ring may interact with the sterically restricted and hydrophobic pocket of receptor, and show small blue polyhedron adjacent to these functional groups. The carbonyl group of 1-acetylpiperidin-4-yl substituent's of template molecule show red contour near the position of carbonyl group. The carbonyl group position and orientation is important for CCR5 binding. Polar groups are required for high binding potency. Compounds with 1-acylpiperidin-3-yl group substitution on 5b (IC_{50} = 1200 nM, pIC_{50} = 5.92), 7e (IC_{50} = 680 nM, pIC_{50} = 6.17) and 7f (IC_{50} = 180 nM, pIC_{50} = 6.74) shows blue contour near this substitution, indicates unfavorable for inhibitory potency.

CoMSIA electrostatics contour map (Figure 7) shows blue and red polyhedra around template molecule. At R^1 position red contour near carbonyl group indicates this group is favorable for inhibitory activity. According to experimental study it is clear that, amide moiety is essential for CCR5 binding. The 1-acetylpiperidin-4-yl substituent afforded compound 7a (IC_{50} = 16 nM, pIC_{50} = 7.78) showed increased binding potency over compound 1 (IC_{50} = 480 nM, pIC_{50} = 6.31), because of presence of acetylpiperidinyl moiety. Here, amide group orientation plays a major role in CCR5 inhibition. Bulkier substitutions on 'N' of piperidine ring, increases the inhibitory potency when compared it with compound 1. Replacement of acetyl group on piperidine moiety resulted in substantial reduction in binding affinity, indicates amide carbonyl group is essential for binding affinity. In case of compound 6a (IC_{50} = 4300 nM, pIC_{50} = 5.37), lack of substitution on piperidine N, resulted in loss of potency. It indicates that for better inhibitory potency, piperidine 'N' must be substituted. The presence of electronegative substituent (R^3 = F) on 4-benzylpiperidine moiety had slightly better effect on inhibitory potency (5e *vs*. 5f, 5g *vs*. 5h).

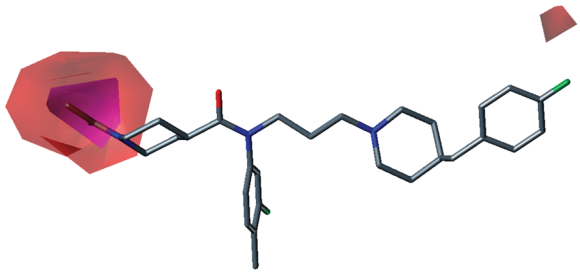


Figure 8. H-bond Acceptor stdev* coefficient contour map for highly active compound (5h) generated by CoMSIA analysis. Magenta contour indicates H-bond acceptor groups enhance binding affinity, whereas red contour decrease binding affinity

Figure 8 depict CoMSIA H-bond acceptor contour map. The template molecule was shown in background and surrounded by red and magenta colored contour. Small magenta contour occurs near R^1 (C = O) group. It indicates that carbonyl group acts as an H-bond acceptor. Big red colored contour occurs near methyl part of acetyl group indicates that position is unfavorable for H-bond acceptor group. Another small sized red contour occurs near R3 position indicates that corresponding position is unfavored for H-bond acceptor groups.

HQSAR atomic contribution map. HQSAR technique gives us straight forward clues about the individual atomic contribution to the biological activity through the different color codes. HQSAR color codes the individual atoms of molecules depending upon their contribution towards biological activity. Atomic contribution map for few molecules was depicted in Figure 9. Highly active compound 5h (IC_{50} = 0.29 nM, pIC_{50} = 9.53) color coded with yellow, green-blue and white color indicates contributes positively for inhibitory activity. The carbonyl group of R^1 substitution shows yellow color code indicates that this group is contributing positively for inhibitory activity. This is well supported by CoMFA and CoMSIA electrostatic red contour map. The 4-Me substitution on central phenyl ring shows yellow color code for hydrogen indicate that contribute positively for activity. It is supported by CoMFA and CoMSIA steric and electrostatics green and blue contour maps. Another highly active compound 5l (IC_{50} = 0.50 nM, pIC_{50} = 9.30) of the series shows similar color coding system as like compound 5h. Both compounds show blue and yellow color code on piperidine ring of 4-benzylpiperidine moiety, it indicates positive contribution for inhibitory activity. The least active molecule of dataset 6a (IC_{50} = 4300 nM, pIC_{50} = 5.37) shows red, red-orange and orange color code for atomic contribution. The color code red on piperidine ring signifies that, because of lack of substitution at R^1 position this compound is not able to make essential amide bond, which is necessary for potent bioactivity. Unsubstituted

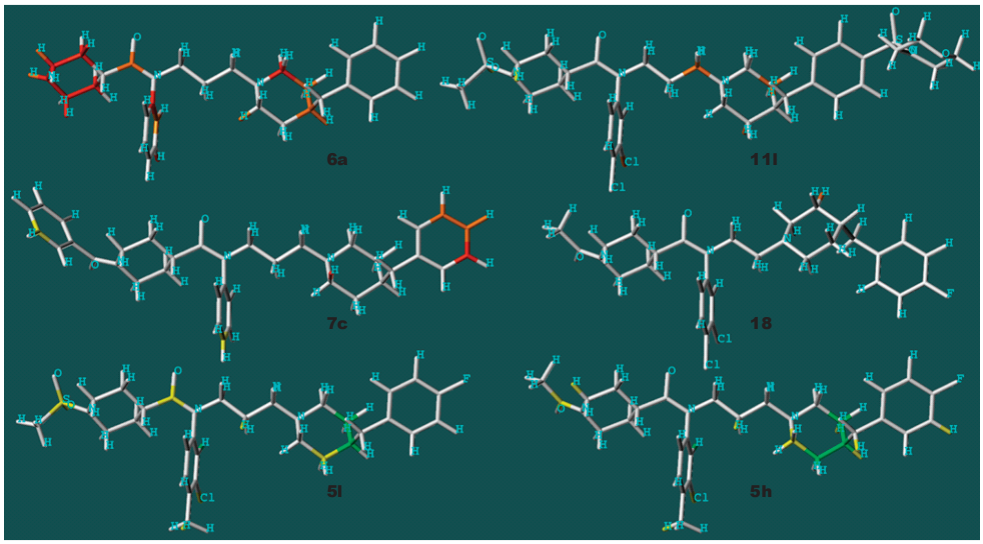


Figure 9. Positive and negative contribution map for few molecules obtained by HQSAR analysis. Color codes red, red orange, and orange show the unfavorable or negative contribution to the activity, the color codes yellow, green blue, and green denote favorable or positive contribution to the activity. The white color shows the intermediate contribution to the activity.

central phenyl ring also shown red color code. On the other hand piperidine moiety shows red and red-orange color code, this implies that it is unfavorable for activity. Compare to compound 7a (IC_{50} = 16 nM, pIC_{50} = 7.78) and 7c (IC_{50} = 33 nM, pIC_{50} = 7.48), compound 6a (IC_{50} = 0.29 nM, pIC_{50} = 5.37) does not have any substitution either at R^1, R^2 and R^3 position. But, compound 7a and compound 7c have substituted acetyl and benzoyl group at R^1 position, which made them highly active. It indicates that to exhibit good inhibitory potency, compound must be substituted at least in the R^1 position.

In summary, CoMFA and CoMSIA contour maps offer enough information for us to understand 3D-QSAR relationship between structures and their biological activities. CoMFA, CoMSIA and HQSAR result supports one another; indicate good agreement between the results.

Conclusion

This study was conducted to rationalize the CCR5 antagonistic activity of 38 piperidine-4-carboxamide derivatives by CoMFA, CoMSIA and HQSAR analyses. All these generated models showed good statistical correlation in terms of q^2 and r^2 values and well explained the variance between biological activities. Additionally higher bootstrapped r^2 values and lowest standard error of estimate indicates the stability of QSAR models. The statistical significance and robustness of developed QSAR models were judged by test set prediction. Comparing the CoMFA and CoMSIA analyses results, we can see that the CoMSIA model has slightly better predictive ability than CoMFA model. The spatial effect of steric, electrostatics and H-bond acceptor descriptors on biological activity of aligned molecules were given by contour map. We could suggest from contour map analysis that, incorporation of small electropositive group at para position of central phenyl ring could increase binding potency and it is supported by yellow color code in HQSAR contribution map. Contour map analysis revealed that compound with carbonyl or sulfonyl group at R^1 position is favorable for activity and supported by favorable yellow color code in HQSAR contribution map. At R^3 position polar substituents are favorable for retaining the inhibitory activity and larger substituent's had better effect on activity. The molecular modeling techniques are successfully employed in this study to correlate biological activities with structures. Good statistical correlation and satisfactory predictive power of developed models indicate that these models can serve as computational tools for rational design of novel CCR5 antagonists with enhanced activity and prediction prior to synthesis.

Acknowledgments. This work was supported by National Research Foundation of Korea (NRF) Grant funded by the Ministry of Education, Science and Technology (MEST) through the Research Center for Resistant Cells (R13-2003-009).

References

1. Chermann, J. C.; Barre-Sinoussi, F.; Dauguet, C.; Brun-Vezinet, F.; Rouzioux, C.; Rozenbaum, W.; Montagnier, L. *Antibiot. Chemother*. **1983**, *32*, 48.
2. Dorr, P.; Westby, M.; Dobbs, S.; Griffin, P.; Irvine, B.; Macartney, M.; Mori, J.; Rickett, G.; Smith-Burchnell, C.; Napier, C. *Antimicrob. Agents Chemother*. **2005**, *49*, 4721.
3. Furtado, M. R.; Callaway, D. S.; Phair, J. P.; Kunstman, K. J.; Stanton, J. L.; Macken, C. A.; Perelson, A. S.; Wolinsky, S. M. *New Engl. J. Med*. **1999**, *340*, 1614.
4. Coffin, J.; Haase, A.; Levy, J. A.; Montagnier, L.; Oroszlan, S.; Teich, N.; Temin, H.; Toyoshima, K.; Varmus, H.; Vogt, P. *Nature* **1986**, *321*, 10.
5. Koot, M.; van't Wout, A. B.; Kootstra, N. A.; Goede, R. E.; Tersmette, M.; Schuitemaker, H. *J. Infect. Dis*. **1996**, *173*, 349.
6. Fackler, O. T.; Peterlin, B. M. *Curr. Biol*. **2000**, *10*, 1005.
7. Piot, P.; Bartos, M.; Ghys, P. D.; Walker, N.; Schwartländer, B. *Nature* **2001**, *410*, 968.
8. Strader, C. D.; Fong, T. M.; Tota, M. R.; Underwood, D.; Dixon,

R. A. F. *Annu. Rev. Biochem*. **1994**, *63*, 101.
9. Cocchi, F.; DeVico, A. L.; Garzino-Demo, A.; Arya, S. K.; Gallo, R. C.; Lusso, P. *Science* **1995**, *270*, 1811.
10. Cocchi, F.; DeVico, A. L.; Garzino-Demo, A.; Lusso, P.; Gallo, R. C. *Science* **1996**, *274*, 1393.
11. Alkhatib, G.; Combadiere, C.; Broder, C. C.; Feng, Y.; Kennedy, P. E.; Murphy, P. M.; Berger, E. A. *Science* **1996**, *272*, 1955.
12. Li, G.; Haney, K. M.; Kellogg, G. E.; Zhang, Y. *J. Chem. Inf. Model.* **2009**, *49*, 120.
13. Strizki, J. M.; Xu, S.; Wagner, N. E.; Wojcik, L.; Liu, J.; Hou, Y.; Endres, M.; Palani, A.; Shapiro, S.; Clader, J. W. *Proc. Nat. Acad. Sci. USA* **2001**, *98*, 12718.
14. Tagat, J. R.; McCombie, S. W.; Nazareno, D.; Labroli, M. A.; Xiao, Y.; Steensma, R. W.; Strizki, J. M.; Baroudy, B. M.; Cox, K.; Lachowicz, J. *J. Med. Chem.* **2004**, *47*, 2405.
15. Strizki, J. M.; Tremblay, C.; Xu, S.; Wojcik, L.; Wagner, N.; Gonsiorek, W.; Hipkin, R. W.; Chou, C. C.; Pugliese-Sivo, C.; Xiao Y. *Antimicrob. Agents Chemother*. **2005**, *49*, 4911.
16. Maeda, K.; Nakata, H.; Koh, Y.; Miyakawa, T.; Ogata, H.; Takaoka, Y.; Shibayama, S.; Sagawa, K.; Fukushima, D.; Moravek, J. *J. Virol*. **2004**, *78*, 8654.
17. Nichols, W. G.; Steel, H. M.; Bonny, T.; Adkison, K.; Curtis, L.; Millard, J.; Kabeya, K.; Clumeck, N. *Antimicrob. Agents Chemother*. **2008**, *52*, 858.
18. Aher, Y. D.; Agrawal, A.; Bharatam, P. V.; Garg, P. *J. Mol. Model.* **2007**, *13*, 519.
19. Afantitis, A.; Melagraki, G.; Sarimveis, H.; Koutentis, P. A.; Markopoulos, J.; Igglessi-Markopoulou, O. *J. Comput.-Aided Mol. Des.* **2006**, *20*, 83.
20. Song, M.; Breneman, C. M.; Sukumar, N. *Bioorg. Med. Chem.* **2004**, *12*, 489.
21. Zhuo, Y.; Kong, R.; Cong, X.; Chen, W.; Wang, C. *Eur. J. Med. Chem.* **2008**, *43*, 2724.
22. Xu, Y.; Liu, H.; Niu, C.; Luo, C.; Luo, X.; Shen, J.; Chen, K.; Jiang, H. *Bioorg. Med. Chem.* **2004**, *12*, 6193.
23. Imamura, S.; Nishikawa, Y.; Ichikawa, T.; Hattori, T.; Matsushita, Y.; Hashiguchi, S.; Kanzaki, N.; Iizawa, Y.; Baba, M.; Sugihara, Y. *Bioorg. Med. Chem.* **2005**, *13*, 397.
24. Cramer, R. D.; Patterson, D. E.; Bunce, J. D. *J. Am. Chem. Soc.* **1988**, *110*, 5959.
25. Klebe, G.; Abraham, U.; Mietzner, T. *J. Med. Chem.* **1994**, *37*, 4130.
26. Hurst T, Heritage T. 213th ACS Natl. Meeting, San Francisco, CA, 1997, CINF 019.
27. S. H. R. SYBYL8.1; Tripos Inc., St. Louis, MO 63144 USA.
28. Ash, S.; Cline, M. A.; Homer, R. W.; Hurst, T.; Smith, G. B. *J. Chem. Inf. Comput. Sci.* **1997**, *37*, 71.
29. Dunn, W. J.; Wold, S.; Edlund, V.; Hellherg, S.; Gasteiger, J. *Quant. Struct.-Act. Relat.* **1984**, *3*, 131.
30. Wold, S.; Sjöström, M.; Eriksson, L. *Chemom. Intell. Lab. Syst.* **2001**, *58*, 109.
31. Cramer, R. D. *Perspect. Drug Discovery Des.* **1993**, *1*, 269.