ELSEVIER

Three-dimensional quantitative structure–activity relationship analyses of piperidine-based CCR5 receptor antagonists

Minghu Song, Curt M. Breneman* and N. Sukumar

Department of Chemistry, Rensselaer Polytechnic Institute, 110 8th Street, Troy, New York 12180, USA

Abstract—The CCR5 chemokine receptor has recently been found to play a crucial role in the viral entry stage of HIV infection and has therefore become an attractive potential target for anti-HIV therapeutics. On the other hand, the lack of CCR5 crystal structure data has impeded the development of structure-based CCR5 antagonist design. In this paper, we compare two three-dimensional Quantitative Structure–Activity Relationship (3D-QSAR) methods: Comparative Molecular Field Analysis (CoMFA) and Comparative Molecular Similarity Indices Analysis (CoMSIA) on a series of piperidine-based CCR5 antagonists as an alternative approach to investigate the interaction between CCR5 antagonists and their receptor. Superimposition of antagonist structures was performed using two alignment rules: atomic/centroid rms fit and rigid body field fit techniques. The 3D QSAR models were derived from a training set of 72 compounds, and were found to have predictive capability for a set of 19 holdout test compounds. The resulting contour maps produced by the best CoMFA and CoMSIA models were used to identify the structural features relevant to biological activity in this series of compounds. Further analyses of these interaction-field contour maps also showed a high level of internal consistency.
© 2003 Elsevier Ltd. All rights reserved.

1. Introduction

Acquired immunodeficiency syndrome (AIDS) is a fatal pandemic caused by infection with the human immunodeficiency virus type 1 (HIV-1). Initially, the HIV virus enters the cell of a macrophage or T-cell by fusion of the viral and cellular membranes. Once internalized, the viral capsid core is disassembled and the exposed viral RNA is reverse-transcribed into a double-stranded viral cDNA that is integrated into the host chromosome with the help of the retroviral integrase enzyme. The integrated DNA, called a provirus, serves as a template to form viral messenger RNA which subsequently yields the new functional proteins of the virus. In the last stage of the virus life cycle, the virus re-assembles into a new viral envelope, and then buds from the surface of cell membrane to become a new infective viral agent. As the consequence, the immune system of infected people is severely weakened and patients are potentially vulnerable to opportunistic infections.[1–3] Such an understanding of AIDS pathogenesis offers us a set of potential targets for the inhibition of HIV replication.

Most anti-AIDS medications available to date belong to two classes of inhibitors that target viral enzymes: HIV reverse transcriptase and HIV protease. The highly active antiretroviral therapy (HAART) provides an effective strategy to suppress viral replication and prolong the lives of infected individuals;[4] however, these therapeutics cannot eradicate the virus from these patients.[5] The effectiveness of these therapies has also been limited by other concerns, such as high cost, long-term side effects, complicated dosing regimens and, more seriously, the emergence of multidrug-resistant viral variants.[6] Therefore, these challenges prompt the search for new treatment agents beyond viral reverse transcriptase and protease inhibitors.[3,7–9]

Recent studies have shown that in addition to the CD4 receptor, a second set of seven-membrane-domain receptors on the cell surface called chemokine receptors also plays a critical role in the membrane-fusion process in HIV infection.[10–12] As a member of the chemokine receptor family, CCR5 is predominantly used by the most commonly transmitted M-tropic strains of HIV in the early stage of the replication cycle. It has been found that blocking the function of CCR5 strongly reduces HIV activity while exhibiting few side effects.[13–15] Furthermore, experiments on CCR5-deficient mice indicate a novel role of CCR5 in down-regulation of mammalian immune function.[16]

These studies reveal that CCR5 is an attractive potential target for anti-HIV treatments and therefore motivates the development of CCR5 antagonists as a new class of anti-HIV therapeutics. Monoclonal antibodies,[17] endogenous ligands of CCR5 receptor including RANTES, MIP-1α and MIP-1β[18], as well as their modified analogues[19,20] have been reported to exhibit anti-HIV activity. In addition, there is great interest in developing small molecule CCR5 antagonists to overcome any possible bioavailability problems associated with large molecules.[21–30] Some of these small molecular antagonist are currently in the early clinical development,[21–24] however, none of them are yet on the market. Hence, there is still a demand for new novel antagonist agents with high potency and selectivity, the development of which requires a more detailed investigation into how the structural features of these antagonists influence their biological activity toward the CCR5 receptor. Unfortunately, since the crystal structure of CCR5 or its complex with any ligands is not presently available, the underlying interaction mechanism between the CCR5 receptor and its antagonists remains a mystery. The structural features described by using 3D physicochemical property modeling may offer some practical guidelines for the future development of CCR5 antagonists.

Three-dimensional quantitative structure–activity relationship (3D-QSAR) methods such as Comparative Molecular Field Analysis (CoMFA),[31] have been widely applied to rational molecular design problems.[32,33] These methods not only provide a better understanding for the relationship between structural variations and an investigated biological activity, but can also be employed as a predictive model to virtually screen a library of potential candidate compounds. To our knowledge, there are no published QSAR models for CCR5 antagonists to date. In this paper, we describe the development of 3D-QSAR models to elucidate the structure–activity relationship (SAR) of small molecule CCR5 antagonists using CoMFA and its extension-Comparative Similarity Indices Analysis (CoMSIA).[34,35] These 3D QSAR models can be used to identify the physicochemical or structural features essential for enhancing their binding affinity and furthermore provide deeper insight into the mechanism of the CCR5 receptor–antagonist interaction. The information acquired will then be exploited in further modeling studies on CCR5 receptor antagonists.

2. Data set and methodology

2.1. Biological data

In this study, 91 piperidine-derived CCR5 antagonists were utilized to construct 3D QSAR models using biological data obtained from the literature.[25–28] In that work, the binding affinity was expressed as an IC_{50} value for, α binding to the CCR5-expressing Chinese hamster ovary (CHO) cell in a high-throughput assay against [^{125}I]-MIP-1α.[26] The negative logarithm of IC_{50} value [pIC_{50} or -log(IC_{50})] was adopted as a dependent variable in the QSAR analyses, with the IC_{50} values

expressed in the unit of nM. For those antagonists whose IC_{50} was evaluated using racemic mixtures, (S)-enantiomer configurations were adopted in the modeling. This selection was justified by the fact that (S)-enantiomer configurations have been found to be more potent than (R)-enantiomer configurations.[26] For those molecules for which IC_{50} values were not reported as specific values, the highest concentration level used in the testing was taken as the upper limit of their response value. Due to the relative paucity of data in this response range, inclusion of the less-well established data for these two low-affinity compounds (**38** and **45**) in the training dataset was justified by the need to represent a broader spectrum of responses in the QSAR model.

The structures of 91 compounds and their biological activities are shown in Table 1. The original collection of these compounds was partitioned into a training set with 72 compounds and a test set with 19 compounds. The test set was obtained by arbitrarily setting aside 19 compounds that approximately represent the distribution of biological data and the structural classes in the training set. This dataset was then used to assess the predictive power of the constructed QSAR models. The mean (standard deviation) of the biological activity pIC_{50} in the training set and test set were 1.39 (0.84) and 1.29 (0.82), respectively.

3. Molecular structure building and alignment
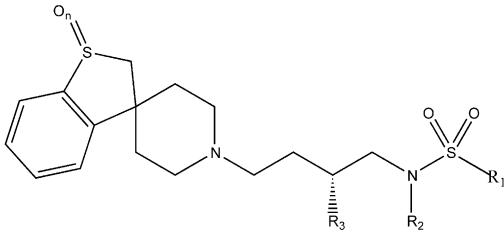
3.1. Molecular modeling

All computational studies were performed using the molecular modeling package SYBYL 6.8[36] on a Silicon Graphics Origin 2000 32-processor server. All molecules were considered in the neutral form and constructed using the SYBYL sketch program or assembled from the standard fragment database. The flexibility of the substituents was taken into account by performing a systematic conformational search. Partial atomic charges were assigned to each atom using the Gasteiger–Hückel method.[37–39] All structural optimizations were performed using a conjugate gradient minimizer and the standard Tripos force field[40] with a distance-dependent dielectric function. The minimization was terminated when the energy gradient convergence criterion of 0.001 kcal/mol was reached or when the 5000-step minimization cycle limit was exceeded. In order to explore the effect of different charges on the 3D QSAR analysis, AM1-ESP charges were also calculated for the same set of compounds using the MOPAC AM1 Hamiltonian implemented in SYBYL.

3.2. Structural alignment

Molecular alignment is considered one of the most sensitive parameters in 3D QSAR analyses. In the present study, two alignment rules were adopted: the atom/centroid RMS fit method and the rigid body field-fit procedure. We refer to these as Alignment I and Alignment II, respectively.

Table 1. Structures of piperidine-based CCR5 antagonists and their biological activities (pIC$_{50}$, nM)

(a)

(1a)

Compound	n	R$_1$	R$_2$	R$_3$	pIC$_{50}$[a]
1[d]	0	C$_6$H$_5$	CH$_3$	(S)-3,4-diCl–C$_6$H$_3$	−3
2	1	C$_6$H$_5$	CH$_3$	(S)-3,4-diCl–C$_6$H$_3$	−1.544
3	2	C$_6$H$_5$	CH$_3$	(S)-3,4-diCl–C$_6$H$_3$	−2
4	1	2-Thienyl	CH$_3$	(S)-3,4-diCl–C$_6$H$_3$	−1.778
5	2	2-Thienyl	CH$_3$	(S)-3,4-diCl–C$_6$H$_3$	−2.079
6	1	CH$_3$	CH$_3$	(S)-3,4-diCl–C$_6$H$_3$	−2.929
7	1	Cyclopentyl	CH$_3$	(S)-3,4-diCl–C$_6$H$_3$	−2
8	1	Cyclohexyl	CH$_3$	(S)-3,4-diCl–C$_6$H$_3$	−2
9	1	2-Cl–C$_6$H$_4$	CH$_3$	(S)-3,4-diCl–C$_6$H$_3$	−1.903
10	1	3-Cl–C$_6$H$_4$	CH$_3$	(S)-3,4-diCl–C$_6$H$_3$	−1.845
11[d]	1	4-Cl–C$_6$H$_4$	CH$_3$	(S)-3,4-diCl–C$_6$H$_3$	−1.6
12	2	3-NO$_2$–C$_6$H$_4$	CH$_3$	(S)-3,4-diCl–C$_6$H$_3$	−2.176
13	2	4-NO$_2$–C$_6$H$_4$	CH$_3$	(S)-3,4-diCl–C$_6$H$_3$	−1.778
14	1	4-MeO–C$_6$H$_4$	CH$_3$	(S)-3,4-diCl–C$_6$H$_3$	−1.602
15	1	4-Ph–C$_6$H$_4$	CH$_3$	(S)-3,4-diCl–C$_6$H$_3$	−1.602
16[d]	1	Naphth-2-yl	CH$_3$	(S)-3,4-diCl–C$_6$H$_3$	−1.778
17[d]	1	Indan-5-yl	CH$_3$	(S)-3,4-diCl–C$_6$H$_3$	−1.845
18	1	Pyridin-3-yl	CH$_3$	(S)-3,4-diCl–C$_6$H$_3$	−2
19	1	Quinolin-8-yl	CH$_3$	(S)-3,4-diCl–C$_6$H$_3$	−1.954
20	1	Quinolin-3-yl	CH$_3$	(S)-3,4-diCl–C$_6$H$_3$	−2.079
21	1	1-Me-Imidazol-4-yl	CH$_3$	(S)-3,4-diCl–C$_6$H$_3$	−2.531
22	0	C$_6$H$_5$	H	(S)-3,4-diCl–C$_6$H$_3$	−3.204
23	0	C$_6$H$_5$	H	(S)-3,4-diCl–C$_6$H$_3$	−2.362
24	2	C$_6$H$_5$	CH$_2$CH$_3$	(S)-3,4-diCl–C$_6$H$_3$	−2.114
25[d]	2	C$_6$H$_5$	CH$_2$CH$_3$	(S)-3,4-diCl–C$_6$H$_3$	−2.204
26[b]	0	C$_6$H$_5$	CH$_3$	C$_6$H$_5$	−2.653
27[d]	0	C$_6$H$_5$	CH$_3$	(S)-3-Cl–C$_6$H$_4$	−2.431
28[b]	1	C$_6$H$_5$	CH$_3$	C$_6$H$_5$	−1.544
29	1	C$_6$H$_5$	CH$_3$	(S)-3-Cl–C$_6$H$_4$	−1
30	1	C$_6$H$_5$	CH$_3$	(S)-4-Cl–C$_6$H$_4$	−2.431
31	1	C$_6$H$_5$	CH$_3$	(S)-4-F–C$_6$H$_4$	−2.756
32	1	C$_6$H$_5$	CH$_3$	(S)-3,4-OCH$_2$O–C$_6$H$_3$	−2.301
33[b,d]	2	C$_6$H$_5$	CH$_3$	C$_6$H$_5$	−1.477
34	2	C$_6$H$_5$	CH$_3$	(S)-3-Cl–C$_6$H$_4$	−1.176
35[d]	2	C$_6$H$_5$	CH$_3$	(S)-3,4-OCH$_2$O–C$_6$H$_3$	−1.903

(b)
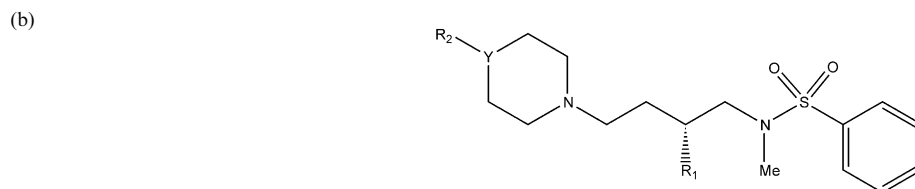
Compound	R$_2$	R$_1$	Y	pIC$_{50}$[b]
36[b]	C$_6$H$_5$	C$_6$H$_5$	CH	−2.079
37	C$_6$H$_5$	(S)-3-Cl–C$_6$H$_4$	CH	−1.477
38[c]	C$_6$H$_5$	(S)-4-F–C$_6$H$_4$	CH	−3
39	C$_6$H$_5$	(S)-3,4-OCH$_2$O-C$_6$H$_3$	CH	−2.477
40[b]	C$_6$H$_5$	3-F–C$_6$H$_4$	CH	−2
41[b]	C$_6$H$_5$	3-Me–C$_6$H$_4$	CH	−1.903
42[b,d]	C$_6$H$_5$	3-Et–C$_6$H$_4$	CH	−2.041
43[b]	C$_6$H$_5$	3–CF$_3$–C$_6$H$_4$	CH	−2.699
44[b,d]	C$_6$H$_5$	4-Me–C$_6$H$_4$	CH	−2.301
45[b,c]	C$_6$H$_5$	4-Ome–C$_6$H$_4$	CH	−3
46[b]	C$_6$H$_5$	3,5-diMe–C$_6$H$_3$	CH	−2.204
47[b,d]	C$_6$H$_5$	3,4-diF–C$_6$H$_3$	CH	−2.756
48[b]	C$_6$H$_5$	3,4-diMe–C$_6$H$_3$	CH	−1.778
49[b]	C$_6$H$_5$	3-Me,4-F–C$_6$H$_3$	CH	−2.255

Compound	R_2	R_1	Y	pIC_{50}[b]
50[b]	C_6H_5	3-F,4-Me–C_6H_3	CH	−2.041
51	C_6H_5	(S)-3-Cl–C_6H_4	N	−2.845
52	2-Me–C_6H_5	(S)-3-Cl–C_6H_4	N	−3.38
53	2-Me–C_6H_4	(S)-3-Cl–C_6H_4	CH	−2.602
54	2-MeO–C_6H_4	(S)-3-Cl–C_6H_4	CH	−1.845
55	3–CF_3–C_6H_4	(S)-3-Cl–C_6H_4	CH	−2.079
56	4-Cl–C_6H_4	(S)-3-Cl–C_6H_4	CH	−2.301
57	3-F–C_6H_4	(S)-3-Cl–C_6H_4	CH	−1.398

(c)
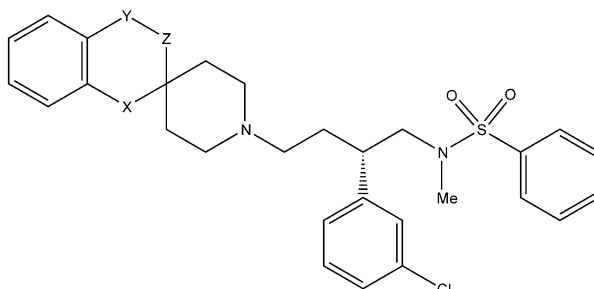
Compound	X	Y–Z	pIC_{50}[b]
58	—	CH_2CH_2	−2.255
59	—	$NHCH_2$	−1.699
60	—	$C(O)CH_2$	−0.699
61	—	C(O)NH	−1.653
62[d]	—	$C(O)N(CH_3)$	−2
63[d]	—	$C(O)NHCH_2$	−1.544
64	—	$NHC(O)CH_2$	−1.544
65	—	$CH(OH)CH_2$	−2
66	CH_2	O	−2.415

(d)
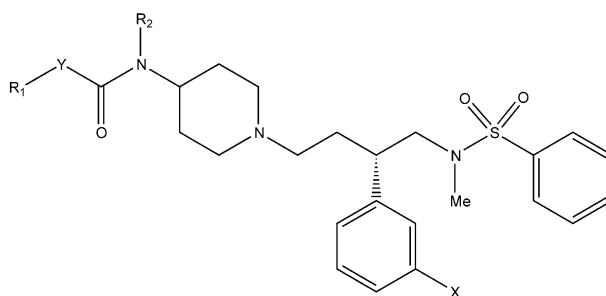
Compound	R_1	R_2	X	Y	pIC_{50}[b]
67[b]	t-Bu	Et	H	O	−1.398
68[b]	C_6H_5	Et	H	O	−1
69[b]	Bn	Et	H	O	−0.301
70	Bn	Et	Cl	O	−0.301
71[b]	Bn	CH_3	H	O	−0.699
72[b]	Bn	n-Pr	H	O	−0.301
73[d]	Bn	n-Pr	Cl	O	−0.602
74[b]	Bn	n-Bu	H	O	−0.699
75	Bn	Allyl	H	O	−0.176
76[b]	2-Me–C_6H_4–CH_2	n-Pr	H	O	−0.602
77[b,d]	3-Me–C_6H_4–CH_2	n-Pr	H	O	−0.477
78[b]	4-Me–C_6H_4–CH_2	n-Pr	H	O	−0.477
79[b,d]	4–CF_3–C_6H_4–CH_2	n-Pr	H	O	−0.778
80[d]	4-NO_2–C_6H_4–CH_2	n-Pr	H	O	−0.176
81	4-NO_2–C_6H_4–CH_2	Allyl	Cl	O	−0.301
82	3-$NH_2COC_6H_4$–CH_2	n-Pr	Cl	O	0.097
83[d]	4-$NH_2COC_6H_4$–CH_2	n-Pr	H	O	−0.301
84	4-$NH_2COC_6H_4$–CH_2	n-Pr	Cl	O	−0.477
85[b]	Bn	H	H	NH	−2
86[b]	Bn	n-Pr	H	NH	−0.398
87[b]	C_6H_5	n-Pr	H	NH	−0.602
88[b,d]	Bn	n-Pr	H	$N(CH_3)$	−1.301
89[b]	(R)-αα-Me–Bn	n-Pr	H	NH	−0.778
90[b]	(S)-αα-Me–Bn	n-Pr	H	NH	−1.875
91	4-NO_2–Bn	Allyl	H	NH	0.125

[a] The experimental IC_{50} values obtained from refs 19–22 were converted into pIC_{50} (the negative logarithm of IC_{50}, in units of nM).
[b] Activity was measured using the racemic mixture and the chiral (S)-enantiomer with higher potency was utilized in the modeling.
[c] The IC_{50} of these compounds were not reported as absolute values and the highest concentration in the test was adopted as the bioactivity in this study.
[d] The compounds held out as the test dataset.

Alignment I. In keeping with recognized procedure, the structure with the best biological activity (compound **91** in Table 1d) was utilized as the template. Other compounds were superimposed on the template using an atom/centroid root mean square (RMS) fitting procedure. The atoms used for alignment are marked with an asterisk (*) on the core structure illustrated in Figure 1.

Alignment II. The second alignment rule involved use of the field fit (rigid body) procedure implemented in SYBYL. First, the steric and electrostatic fields of the template (compound **91**) were calculated as reference fields, then the field fit procedure was used to change the geometry of all other molecules from their initial orientations in Alignment I such that their fields overlap the template fields as closely as possible. The results of Alignments I and II may be seen on the left and right sides of Figure 2, respectively.

4. CoMFA and CoMSIA analysis

In deriving the 3D-QSAR interaction fields, a 3D cubic lattice that extending at least 4 Å beyond the volumes of all investigated molecules on all axes was defined automatically with a grid spacing of 2.0 Å. In the CoMFA analysis, Lennard–Jones 6–12 and Coulomb potentials were employed to calculate the CoMFA steric and electrostatic interaction fields, respectively. An sp^3-hybridized carbon atom with a charge of $+1$ was used as the probe atom and the steric and electrostatic energy cutoff was 30

Figure 1. The core structure used for molecular modeling and atom-based alignment. The atoms for alignment are marked with an asterisk. The centroids of piperidine, 2-phenyl and sulfonamide phenyl rings are defined as dummy atoms for molecular alignment.

kcal/mol. In the CoMSIA analysis, the following standard parameters for the probe atom were used: a radius of 1 Å, a charge of $+1$, hydrophobicity of $+1$, hydrogen-bond donor or acceptor properties of $+1$. A default value 0.3 was adopted for the attenuation factor α.

In both studies, partial least squares (PLS) was carried out in conjugation with the leave-one-out cross-validation procedure to determine the optimum number of PLS components for the final non-cross-validated 3D-QSAR models. The optimum number of components produces the smallest root mean predictive sum of squared errors, which usually corresponds to the highest cross-validated coefficient $r_{cv}^2 = 1 - \frac{\sum (Y_{obs} - Y_{pre})^2}{\sum (Y_{obs} - Y_{mean})^2}$. The predictive power of the resulting 3D-QSAR models was assessed on the holdout test set using the r_{pred}^2 metric, which is defined as $r_{pred}^2 = \frac{SD-PRESS}{SD}$. SD is defined as the sum of squared deviations between the biological activities of the test set and the mean value of the training set responses, and PRESS is the sum of the squared deviation between the predicted and experimental bioactivities for the test compounds.

To further assess the statistical confidence of the derived models, a 100-cycle bootstrap analysis was performed using the optimized number of latent variables determined while building the PLS model. Finally, to qualitatively visualize the contribution from these interaction fields to the PLS model, contour maps were generated as scalar products of coefficients and the standard deviations (std*coeff) associated with CoMFA or CoMSIA values at each lattice point.

5. Results and discussion

5.1. CoMFA and CoMSIA statistical results

The statistical results of from CoMFA and CoMSIA modeling are summarized in Table 2. CoMFA analysis using the atom/centroid method of alignment (Alignment I) gave a cross-validated r_{cv}^2 of 0.726 with an optimal
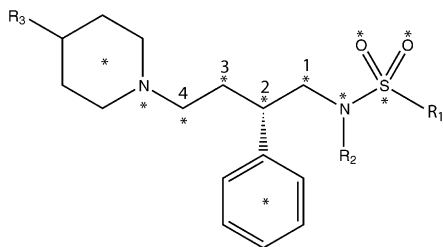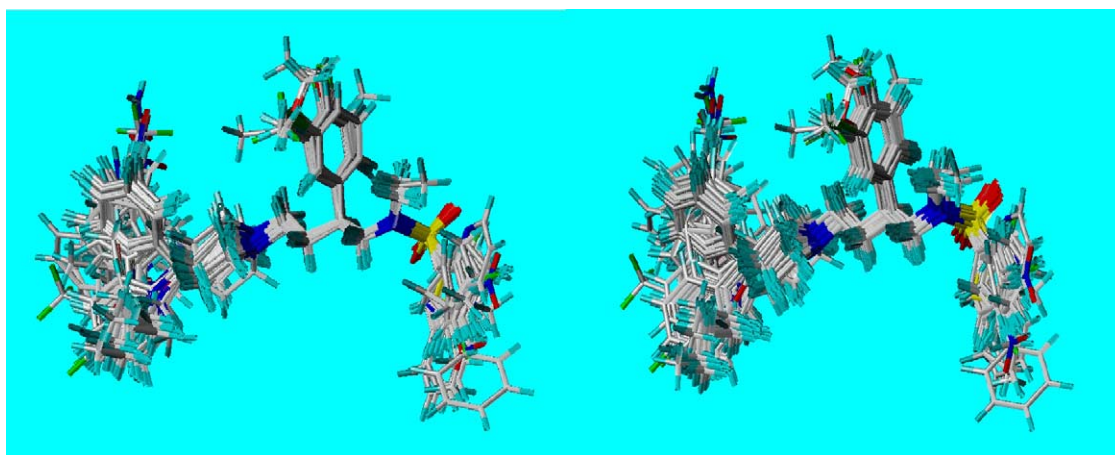

Figure 2. A stereoview of the aligned molecules based on Alignment I (left) and Alignment II (right).

number of latent variables (4) and a conventional r^2 of 0.891 for the non-cross-validated final model. Alternatively, the CoMFA model based on field-fit alignment (Alignment II) produced a statistically better model with the same number of latent variables. In this case, an r^2_{cv} of 0.756 and a conventional r^2 of 0.918 was observed. Reasonable predictions for the holdout test compounds were also achieved: an r^2 of 0.878 was seen for Alignment I and an r^2 of 0.837 was observed for Alignment II. These statistical results reveal that the resultant CoMFA model has good predictive capacity. The satisfactory quality of the best CoMFA model (using Alignment II) is represented in Figure 3, which shows a scatter plot of experimental pIC$_{50}$ values versus the predicted pIC$_{50}$ for the training set and test set.

In the CoMSIA studies, several models were generated using the two alignment methods and as well as different field combinations. The models developed using both steric and electrostatic fields produced a cross-validated r^2_{cv} of 0.645 (Alignment I) and 0.674 (Alignment II) with four latent variables. The corresponding conventional r^2 values of 0.839 (Alignment I) and 0.855 (Alignment II) were also determined. When the hydrophobic field was included, the cross-validated r^2_{cv} and conventional r^2 for both alignments were improved to 0.72 and 0.88, respectively. It was also observed that the further inclusion of hydrogen-bonding fields did not improve the quality of models. The above data indicate that a combination of steric, electrostatic and hydrophobic fields can provide an adequate description for the receptor-ligand interaction for this series of CCR5

Table 2. Summary of CoMFA and CoMSIA results for various models using Gasteiger-Hückel charges

	CoMFA		CoMSIA (ALL)		CoMSIA (SEHA)		CoMSIA (SEH)		CoMSIA (SE)	
	I[a]	II[b]	I	II	I	II	I	II	I	II
	4	4	5	5	4	5	4	4	4	4
Components										
r^2_{cv}	0.726	0.756	0.718	0.708	0.716	0.721	0.727	0.720	0.645	0.674
SEP[c]	0.455	0.430	0.466	0.474	0.478	0.464	0.454	0.461	0.518	0.497
$r^2_{conventional}$	0.891	0.918	0.881	0.885	0.871	0.899	0.880	0.879	0.839	0.855
SEE[d]	0.287	0.250	0.302	0.297	0.312	0.278	0.301	0.303	0.349	0.331
F[e]	136.585	186.341	98.112	101.809	113.328	118.069	123.373	121.606	86.989	95.589
r^2_{pred}	0.878	0.837	0.880	0.888	0.878	0.887	0.872	0.889	0.891	0.877
Fields contributions[f] (%)										
S	52.4	47.7	12.0	12.7	16.7	17.2	18.6	19.5	29.8	30.7
E	47.6	52.3	33.6	32.4	39.8	40.3	48.8	48.9	70.2	69.3
H			20.9	20.7	28.2	28.1	32.6	31.5		
A			13.4	12.9	15.3	14.5				
D			20.1	21.4						
r^2_{bs} [g]	0.925	0.941	0.904	0.918	0.943	0.933	0.907	0.906	0.878	0.898
SD$_{bs}$[g]	0.019	0.015	0.018	0.018	0.014	0.014	0.018	0.018	0.026	0.021

[a] Results using Alignment I.
[b] Results using Alignment II.
[c] SEP, standard error of prediction.
[d] SEE, standard error of estimation.
[e] F-test value.
[f] Field contributions: S, E, H, A, D, ALL denote the steric, electrostatic, hydrophobic, hydrogen-bond acceptor and donor, as well as the combination of the above five interaction fields, respectively.
[g] r^2_{bs} and SD$_{bs}$ represent the mean value of 100-run bootstrap predicted cross-validated r^2_{cv} and its standard deviation.
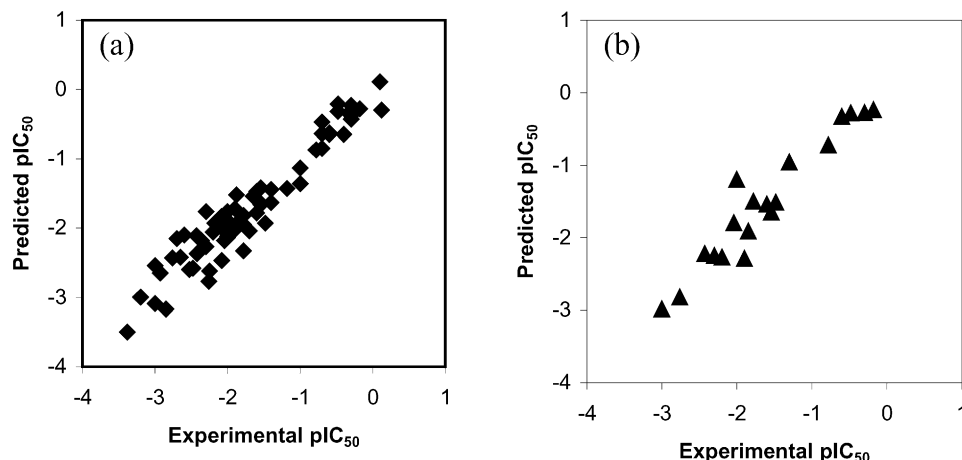

Figure 3. Predicted versus experimental activity (pIC$_{50}$, nM) of compounds in the training set (a) and the test set (b) for the CoMFA model produced using Alignment II.

antagonists. The relative insignificance of the hydrogen-bonding field may be due to its similarity with a combination of electrostatic, steric and hydrophobic fields. Since no significant difference was observed between the models developed using the two different alignment approaches, the CoMSIA model with the slightly higher cross-validated r_{cv}^2 (Alignment I) was chosen for the final analysis. The scatter plot of experimental versus predicted binding affinity pIC_{50} for the training set and the test set (Table 1) are presented in Figure 4 for the CoMSIA model with steric, electrostatic and hydrophobic fields using Alignment I. The detailed experimental and predicted IC_{50} values based on the selected CoMFA and CoMSIA models for the test compounds are shown in Table 4. It was observed that the predicted pIC_{50} residuals of these CoMFA and CoMSIA models were all less than one log unit. In CoMFA, 90% of the compounds were predicted to within 0.5 log units, while for CoMSIA, 84% of the compounds were predicted to

this level of accuracy and the remaining three compounds were found to have residuals of slightly larger than 0.5 log units. In addition, the performance of 100-fold bootstrap analyses, shown in Table 2, suggested that there is a good internal consistency within the underlying dataset.

Although the model created using semi-empirical AM1-ESP charges also gave favorable statistics (shown in Table 3), those results were not as good as those based on Gasteiger–Hückel charges. The Gasteiger–Hückel charge model was therefore selected for the work presented in the following section.

While the limited amount of chemical variation of the C-2 phenyl substituents within this dataset precludes a detailed interpretation of the CoMFA and CoMSIA data, it is possible to make certain qualitative observations from the results. For example, the contour plot of

Table 3. Summary of CoMFA and CoMSIA results for several models using AM1-ESP charges

	CoMFA		CoMSIA (ALL)		CoMSIA (SEHA)		CoMSIA (SEH)		CoMSIA (SE)	
	I[a]	II[b]	I	II	I	II	I	II	I	II
	6	7	4	6	6	6	7	6	6	6
Components										
r_{cv}^2	0.673	0.688	0.678	0.693	0.691	0.693	0.691	0.680	0.692	0.637
SEP[c]	0.505	0.497	0.493	0.489	0.490	0.489	0.495	0.500	0.490	0.532
$r_{conventional}^2$	0.936	0.945	0.806	0.882	0.890	0.913	0.904	0.907	0.890	0.897
SEE[d]	0.223	0.200	0.383	0.304	0.293	0.260	0.276	0.270	0.278	0.283
F[e]	158.610	199.604	69.777	80.754	87.813	113.843	86.213	105.107	87.549	94.573
r_{pred}^2	0.790	0.837	0.810	0.819	0.694	0.698	0.655	0.743	0.856	0.842
Fields contributions[f] (%)										
S	55.7	58.2	14.7	14.0	17.5	17.3	20.8	19.5	38.9	36.0
E	44.3	41.8	15.8	20.6	27.4	28.0	34.8	34.1	61.1	64.0
H			25.4	25.6	38.2	36.2	44.4	46.4		
A			19.0	14.9	16.9	18.5				
D			25.2	25.0						

[a] Results using Alignment I.
[b] Results using Alignment II.
[c] SEP, standard error of prediction.
[d] SEE, standard error of estimation.
[e] F-test value.
[f] Field contributions: S, E, H, A, D, ALL denote the steric, electrostatic, hydrophobic, hydrogen-bond acceptor and donor, as well as the combination of the above five interaction fields, respectively.
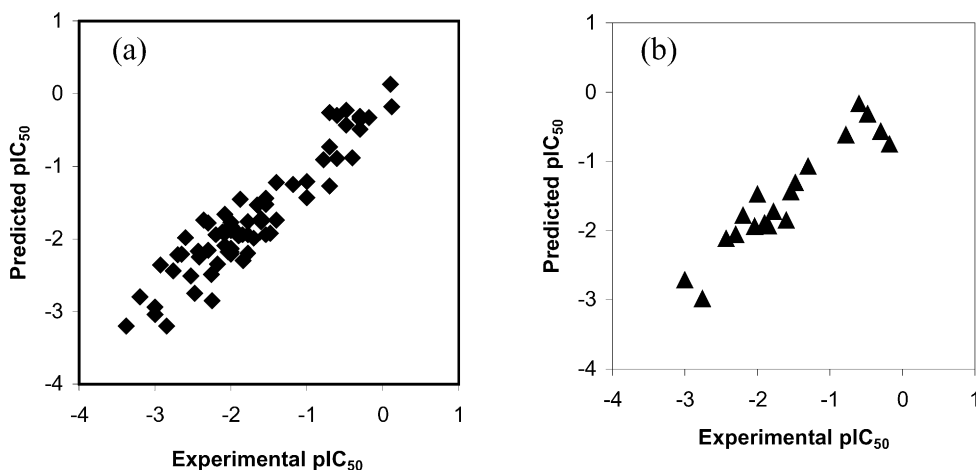

Figure 4. Predicted versus experimental activity (pIC_{50}, nM) of compounds in the training set (a) and the test set (b) for the CoMSIA model produced using Alignment I.

496

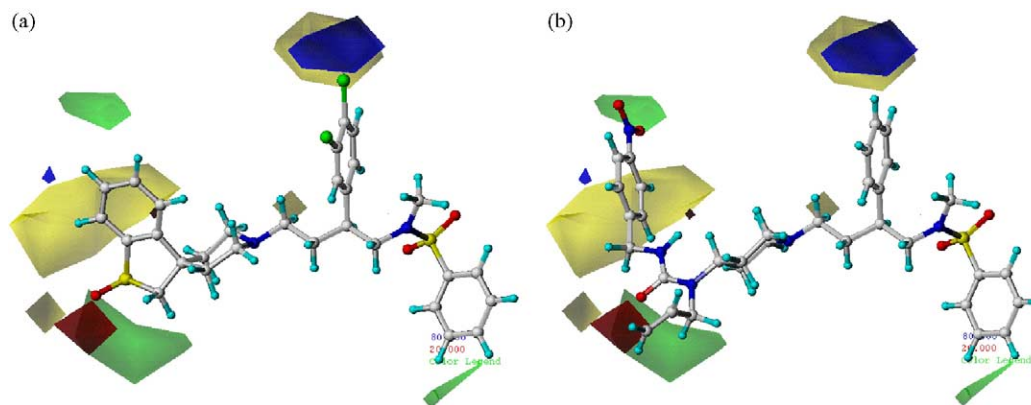*M. Song et al. / Bioorg. Med. Chem. 12 (2004) 489–499*

Figure 5. CoMFA stDev*Coeff contour plots with Alignment II for steric and electrostatic fields. Compounds **2** (a) and **91** (b) are displayed as references. Sterically favored/disfavored areas are shown in green/yellow, while the blue/red polyhedra depict the favorable site for positively/negatively charged groups. Favored and disfavored levels of these displayed interaction fields were fixed at 80 and 20%, respectively.

Table 4. Experimental pIC_{50} and the corresponding predicted values for the best CoMFA and CoMSIA models for the test compounds

Compd	pIC_{50}[a]	CoMFA[b]	CoMSIA[c]
1	−3	−2.986	−2.718
11	−1.6	−1.543	−1.856
16	−1.78	−1.501	−1.731
17	−1.85	−1.913	−1.935
25	−2.2	−2.275	−1.787
27	−2.43	−2.222	−2.122
33	−1.48	−1.517	−1.316
35	−1.9	−2.292	−1.897
42	−2.04	−1.797	−1.497
44	−2.3	−2.252	−2.059
47	−2.76	−2.823	−2.988
62	−2	−1.203	−1.482
63	−1.54	−1.653	−1.445
73	−0.6	−0.312	−0.172
77	−0.48	−0.283	−0.324
79	−0.78	−0.727	−0.625
80	−0.18	−0.24	−0.757
83	−0.3	−0.28	−0.571
88	−1.3	−0.694	−1.076

[a] Experimental pIC_{50} values.
[b] Predicted pIC_{50} of CoMFA model using Alignment II.
[c] Predicted pIC_{50} of CoMSIA using Alignment I and the combination of steric, electrostatic and hydrophobic fields.

the best CoMFA model based on Alignment II (field-fit) is presented in Figure 5. In this figure, the structures of Compound **2** (the initial benchmark structure of this series of CCR5 antagonists) and the most potent Compound **91** are shown superimposed on the important steric and electrostatic field regions. The green polyhedra characterize the regions where bulky substituents would increase the binding affinity, whereas yellow contours indicate regions where steric bulk would not be tolerated. The blue (red) polyhedra depict the favorable sites for positively (negatively) charged groups.

The large yellow polyhedron located at the left side of Figure 5 indicates that bulky substituents at the 4-position of the piperidine ring would not be favorable. This region of disfavorable steric interactions appears to account for the poorer activity of Compounds **58** ($IC_{50} = 180$ nM) and **27** ($IC_{50} = 270$ nM). The low binding affinity of these compounds may result from an improper orientation of a rigid aromatic ring system in

that area, compared with Compound **37** ($IC_{50} = 30$ nM) that has a more flexible 4-phenyl ring that can be oriented in a more favorable fashion. Although Compounds **2** ($IC_{50} = 35$ nM) and **3** ($IC_{50} = 100$ nM) have the same orientations of aromatic rings connected to the 4-position of the piperidine ring as in Compounds **1**, **27**, and **58** (Fig. 3a), this deficiency could be offset by favorable hydrogen-bond interactions between their sulfoxide or sulfone moieties and the receptor. This hypothesis is also consistent with the potent affinity observed for most of the compounds in Table 1d. This favorable interaction between the CCR5 receptor and the carbonyl groups (in Table 1d) or sulfoxide/sulfone groups in the spiro series (in Table 1a) was demonstrated by the appearance of a red contour (favorable negative charge region) close to the oxygen atom of the above groups. These oxygens could promote the formation of hydrogen-bonds with the corresponding receptor atoms. Above the yellow contour, there is a green polyhedron in the vicinity of the *para* position of the benzamide phenyl ring in Compound **91**, implying that the introduction of bulky groups in that position would enhance the binding affinity. Additionally, the medium-sized green polyhedron on the lower-left corner of the images suggests that this region would prefer substituents with moderate bulk, such as ethyl, propyl, or allyl groups over hydrogen. In the upper middle part of Figure 5a and b, the region favorable for positive charges (the blue polyhedron) juxtaposed with the forbidden steric region depicted by the yellow polyhedron indicates a strict requirement for *para* substitution on the C-2 phenyl ring. In strongly binding compounds, that region may be limited to small moieties bearing positive partial charges, an observation that is in agreement with the less potent activity of analogues substituted with electron-rich halogen atoms at that position.

An example of a CoMSIA model employing a combination of steric, electrostatic and hydrophobic fields and atom/centroid RMS fitting (Alignment I) is presented in Figure 6. The color scheme for electrostatic and steric fields is represented in the same manner as that used for CoMFA. In addition, the cyan and purple contours represent regions that are favorable to hydrophobic or
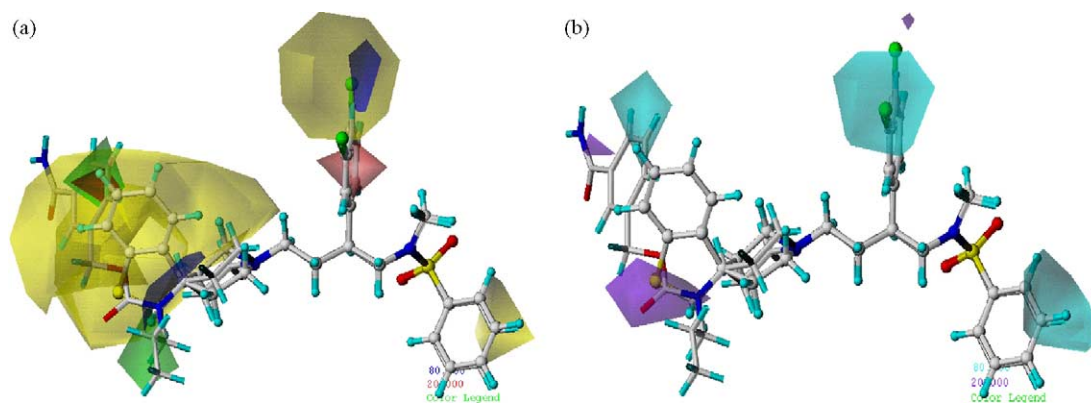
Figure 6. CoMSIA stDev*Coeff contour plots with Alignment I for steric and electrostatic fields (a) and hydrophobic field only (b). Compounds **2** and **91** are both embedded into the contour maps. Sterically favored/disfavored areas are shown in green/yellow, while the blue/red polyhedra depict favorable sites for positively/negatively charged groups. Cyan/purple contours indicate the regions where hydrophobic/hydrophilic groups would enhance the binding affinity. Favored and disfavored levels of these displayed interaction fields were fixed at 80 and 20%, respectively.

hydrophilic groups, respectively. In contrast to CoMFA contribution maps that highlight the 'pseudo-receptor' region where aligned molecules would interact favorably or unfavorably, the CoMSIA method provides a diagram that allows each field contribution to be mapped back directly to regions within molecular structures so that information regarding the contribution of particular functional groups in certain regions may be evaluated.[34,35]

The CoMSIA steric and electrostatic contour maps were found to be consistent with the CoMFA results. Besides the structural features already mentioned in the CoMFA analysis, there are three additional polyhedra that may represent important interactions between the ligands and the binding site. One red polyhedron located inside the C-2 phenyl ring suggests that an electron-rich ring would be favorable for binding. This is consistent with the electron withdrawing effects of 4-halogen substitutions, which are seen to decrease the binding affinity. Another blue polyhedron (region favorable to positive charges) that appears around the 4-position of the piperidine ring is consistent with the dramatic decrease in affinity that is observed when piperidine is replaced by piperazine. This effect may be seen by comparing the binding of Compound **37** ($pIC_{50} = 30$ nM) with that of Compound **51** ($pIC_{50} = 700$ nM), as well as the relative binding of Compound **53** ($pIC_{50} = 400$ nM) with Compound **52** ($pIC_{50} = 2400$ nM). Comparison of semi-empirical AM1-ESP partial atomic charges on two piperidine and piperazine model compounds showed that the piperidine nitrogen exhibits a larger negative charge (-0.492) than either nitrogen on piperazine (-0.475). This is consistent with the greater basicity of piperidine, and is a likely result of electrostatic field and inductive effects caused by the additional ring nitrogen. This is also consistent with the CoMSIA results, in which the region near the 4-position of piperidine ring is tagged as being favorable for positively charged groups, and is a poor place for negative electrostatic potential to exist. This observation is in keeping with the known importance of the piperidine nitrogen atom in a possible salt bridge interaction with the CCR5 receptor as tested by site-directed mutagenesis experiments.[41]

Inspection of the hydrophobic field contribution map in Figure 6b showed that it is consistent with the analysis of the standard fields (steric and electrostatic fields) for both CoMFA and CoMSIA. The red contours observed around the carbonyl or sulfoxide groups in the CoMFA map do not appear in the CoMSIA map. Instead, a purple polyhedron representing a zone of favorable hydrophilic interactions appears in that region. This result provides further support for the hypothesis that the oxygen atom of those groups is involved in a favorable hydrogen-bonding interaction with the CCR5 receptor. The hydrophilic purple polyhedron that appears near the upper-left corner of the CoMSIA map in the vicinity of the *para* position of the phenyl in the benzyloxy or benzamide series resides in a region that had earlier been seen to be sterically favorable in CoMFA. This is consistent with the observation that compounds with polar substituents, (compounds **80–84)** such as nitro and carboxamide, can exhibit potent affinity even in the absence of bulky groups in that region.

In the CoMSIA hydrophobic map, the cyan region around substituents at the 3-position of the C-2 phenyl group suggests that this phenyl ring is located in a small hydrophobic pocket that favors electron-rich rings. Moreover, there is another region favorable to hydrophobic groups near the *meta* and *ortho* positions of the sulfonamide-bearing phenyl ring, suggesting that a potential hydrophobic interaction with the receptor would be favorable to binding. As discussed previously, the 4-position of the phenylsulfonyl moiety was shown to tolerate a wide range of substituents, suggesting that this phenyl ring lies in an extended pocket within the receptor.

Although the electronic and hydrophilic features of the sulfonamide and piperidine groups are presumed to be key pharmacophore features for activity, these groups are conserved among most of the compounds investigated in this study. The resulting lack of structural variation of these key groups leads to a small variance in the interaction fields around these regions. Consequently, the significance of these key scaffolds is not represented in the contour maps described earlier. Inclusion of additional screening data from a more

diverse set of antagonists could provide sufficient structural variation within this key pharmacophore region, and serve to further demonstrate the importance of those relationships.

6. Conclusions

In this study, CoMFA and CoMSIA models have been created to explain the observed structure–activity relationship for a series of piperidine-based CCR5 antagonists. These 3D-QSAR models have demonstration a good predictive capability. High bootstrapped r^2 values and small standard deviations were observed in the modeling results, lending a level of confidence to these analyses. The models provide a useful tool for the virtual screening of proposed CCR5 antagonists for binding affinity within similar structural classes. It was also observed in this investigation that a combination of steric, electrostatic and hydrophobic fields provides an adequate description of the guest–host interactions relevant to this group of CCR5 antagonists. The CoMFA and CoMSIA model maps were found to display important molecular features for antagonist behavior. The structural requirements for the antagonists identified by these 3D-QSAR contour maps provides a rational outline for the development and optimization of new potentially active CCR5 antagonists.

Acknowledgements

References and notes

1. Moore, J. P.; Stevenson, M. *Nat. Rev. Mol. Cell Biol.* **2000**, *1*, 40.
2. Doms, R. W. In *HIV/AIDS Annual Update 2002, Proceedings of the 12th Annual Clinical Care Options for HIV Symposium*; iMedOptions, LLC: Key Biscayne, FL, 2002.
3. Mehanna, A. S. In *Burger's Medicinal Chemistry and Drug Discovery*; Abraham, D. J., Ed.; John Wiley & Sons: New York, 2003; Vol. 5.
4. De Clercq, E. *J. Med. Chem.* **1995**, *38*, 2491.
5. 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.
6. Richman, D. D. *Nature* **2001**, *410*, 995.
7. Blair, W. S.; Lin, P. F.; Meanwell, N. A.; Wallace, O. B. *Drug Discov. Today* **2000**, *5*, 183.
8. De Clercq, E. *Med. Res. Rev.* **2002**, *22*, 531.
9. Nair, V. *Rev. Med. Virol.* **2002**, *12*, 179.
10. Deng, H. K.; Liu, R.; Ellmeier, W.; Choe, S.; Unutmaz, D.; Burkhart, M.; DiMarzio, P.; Marmon, S.; Sutton, R. E.; Hill, C. M.; Davis, C. B.; Peiper, S. C.; Schall, T. J.; Littman, D. R.; Landau, N. R. *Nature* **1996**, *381*, 661.
11. Dragic, T.; Litwin, V.; Allaway, G. P.; Martin, S. R.; Huang, Y. X.; Nagashima, K. A.; Cayanan, C.; Maddon, P. J.; Koup, R. A.; Moore, J. P.; Paxton, W. A. *Nature* **1996**, *381*, 667.
12. Fauci, A. S. *Nature* **1996**, *384*, 529.
13. Dean, M.; Carrington, M.; Winkler, C.; Huttley, G. A.; Smith, M. W.; Allikmets, R.; Goedert, J. J.; Buchbinder, S. P.; Vittinghoff, E.; Gomperts, E.; Donfield, S.; Vlahov, D.; Kaslow, R.; Saah, A.; Rinaldo, C.; Detels, R.; O'Brien, S. J. *Science* **1996**, *273*, 1856.
14. Liu, R.; Paxton, W. A.; Choe, S.; Ceradini, D.; Martin, S. R.; Horuk, R.; MacDonald, M. E.; Stuhlmann, H.; Koup, R. A.; Landau, N. R. *Cell* **1996**, *86*, 367.
15. Samson, M.; Libert, F.; Doranz, B. J.; Rucker, J.; Liesnard, C.; Farber, C. M.; Saragosti, S.; Lapoumeroulie, C.; Cognaux, J.; Forceille, C.; Muyldermans, G.; Verhofstede, C.; Burtonboy, G.; Georges, M.; Imai, T.; Rana, S.; Yi, Y. J.; Smyth, R. J.; Collman, R. G.; Doms, R. W.; Vassart, G.; Parmentier, M. *Nature* **1996**, *382*, 722.
16. Zhou, Y. H.; Kurihara, T.; Ryseck, R. P.; Yang, Y.; Ryan, C.; Loy, J.; Warr, G.; Bravo, R. *J. Immunol.* **1998**, *160*, 4018.
17. Olson, W. C.; Rabut, G. E. E.; Nagashima, K. A.; Tran, D. N. H.; Anselma, D. J.; Monard, S. P.; Segal, J. P.; Thompson, D. A. D.; Kajumo, F.; Guo, Y.; Moore, J. P.; Maddon, P. J.; Dragic, T. *J. Virol.* **1999**, *73*, 4145.
18. Cocchi, F.; Devico, A. L.; Garzinodemo, A.; Arya, S. K.; Gallo, R. C.; Lusso, P. *Science* **1995**, *270*, 1811.
19. Simmons, G.; Clapham, P. R.; Picard, L.; Offord, R. E.; Rosenkilde, M. M.; Schwartz, T. W.; Buser, R.; Wells, T. N. C.; Proudfoot, A. E. I. *Science* **1997**, *276*, 276.
20. Heveker, N.; Montes, M.; Germeroth, L.; Amara, A.; Trautmann, A.; Alizon, M.; Schneider-Mergener, J. *Curr. Biol.* **1998**, *8*, 369.
21. Howard, O. M. Z.; Oppenheim, J. J.; Hollingshead, M. G.; Covey, J. M.; Bigelow, J.; McCormack, J. J.; Buckheit, R. W.; Clanton, D. J.; Turpin, J. A.; Rice, W. G. *J. Med. Chem.* **1998**, *41*, 2184.
22. Howard, O. M. Z.; Korte, T.; Tarasova, N. I.; Grimm, M.; Turpin, J. A.; Rice, W. G.; Michejda, C. J.; Blumenthal, R.; Oppenheim, J. J. *J. Leukocyte Biol.* **1998**, *64*, 6.
23. Baba, M.; Nishimura, O.; Kanzaki, N.; Okamoto, M.; Sawada, H.; Iizawa, Y.; Shiraishi, M.; Aramaki, Y.; Okonogi, K.; Ogawa, Y.; Meguro, K.; Fujino, M. *Proc. Natl. Acad. Sci. U.S.A.* **1999**, *96*, 5698.
24. Strizki, J. M.; Xu, S.; Wagner, N. E.; Wojcik, L.; Liu, J.; Hou, Y.; Endres, M.; Palani, A.; Shapiro, S.; Clader, J. W.; Greenlee, W. J.; Tagat, J. R.; McCombie, S.; Cox, K.; Fawzi, A. B.; Chou, C. C.; Pugliese-Sivo, C.; Davies, L.; Moreno, M. E.; Ho, D. D.; Trkola, A.; Stoddart, C. A.; Moore, J. P.; Reyes, G. R.; Baroudy, B. M. *Proc. Natl. Acad. Sci. U.S.A.* **2001**, *98*, 12718.
25. Finke, P. E.; Meurer, L. C.; Oates, B.; Mills, S. G.; MacCoss, M.; Malkowitz, L.; Springer, M. S.; Daugherty, B. L.; Gould, S. L.; DeMartino, J. A.; Siciliano, S. J.; Carella, A.; Carver, G.; Holmes, K.; Danzeisen, R.; Hazuda, D.; Kessler, J.; Lineberger, J.; Miller, M.; Schleif, W. A.; Emini, E. A. *Bioorg. Med. Chem. Lett.* **2001**, *11*, 265.
26. Dorn, C. P.; Finke, P. E.; Oates, B.; Budhu, R. J.; Mills, S. G.; MacCoss, M.; Malkowitz, L.; Springer, M. S.; Daugherty, B. L.; Gould, S. L.; DeMartino, J. A.; Siciliano, S. J.; Carella, A.; Carver, G.; Holmes, K.; Danzeisen, R.; Hazuda, D.; Kessler, J.; Lineberger, J.; Miller, M.; Schleif, W. A.; Emini, E. A. *Bioorg. Med. Chem. Lett.* **2001**, *11*, 259.
27. Finke, P. E.; Oates, B.; Mills, S. G.; MacCoss, M.; Malkowitz, L.; Springer, M. S.; Gould, S. L.; DeMartino, J. A.; Carella, A.; Carver, G.; Holmes, K.; Danzeisen, R.; Hazuda, D.; Kessler, J.; Lineberger, J.; Miller, M.; Schleif, W. A.; Emini, E. A. *Bioorg. Med. Chem. Lett.* **2001**, *11*, 2475.
28. Finke, P. E.; Meurer, L. C.; Oates, B.; Shah, S. K.; Loebach, J. L.; Mills, S. G.; MacCoss, M.; Castonguay, L.; Malkowitz, L.; Springer, M. S.; Gould, S. L.; DeMartino, J. A. *Bioorg. Med. Chem. Lett.* **2001**, *11*, 2469.

29. Palani, A.; Shapiro, S.; Clader, J. W.; Greenlee, W. J.; Cox, K.; Strizki, J.; Endres, M.; Baroudy, B. M. *J. Med. Chem.* **2001**, *44*, 3339.
30. Palani, A.; Shapiro, S.; Josien, H.; Bara, T.; Clader, J. W.; Greenlee, W. J.; Cox, K.; Strizki, J. M.; Baroudy, B. M. *J. Med. Chem.* **2002**, *45*, 3143.
31. Cramer, R. D.; Patterson, D. E.; Bunce, J. D. *J. Am. Chem. Soc.* **1988**, *110*, 5959.
32. Kubinyi, H. 3D Qsar in Drug Design: Theory Methods and Applications; Kluwer Academic Publishers: Dordrecht, 1993.
33. Greco, G.; Novellino, E.; Martin, Y. C. In *Designing Bioactive Molecules: Three-Dimensional Techniques and Applications*; Y. C. M., Willett, P., Ed.; American Chemical Society: Washington, D.C., 1998.
34. Klebe, G.; Abraham, U.; Mietzner, T. *J. Med. Chem.* **1994**, *37*, 4130.
35. Klebe, G. *Perspect. Drug Discov*. **1998**, *12*, 87.
36. *SYBYL*, version 6.8; Tripos, Inc.: St. Louis, MO 63144-2913, 2001.
37. Gasteiger, J.; Marsili, M. *Tetrahedron* **1980**, *36*, 3219.
38. Gasteiger, J.; Saller, H. *Angew. Chem. Int. Edit.* **1985**, *24*, 687.
39. Streitwieser, A. *Molecular Orbital Theory for Organic Chemists*; Wiley: New York, 1961.
40. Clark, M.; Cramer, R. D.; Vanopdenbosch, N. *J. Comput. Chem*. **1989**, *10*, 982.
41. Castonguay, L. A.; Weng, Y. M.; Adolfsen, W.; Di Salvo, J.; Kilburn, R.; Caldwell, C. G.; Daugherty, B. L.; Finke, P. E.; Hale, J. J.; Lynch, C. L.; Mills, S. G.; MacCoss, M.; Springer, M. S.; DeMartino, J. A. *Biochemistry* **2003**, *42*, 1544.