Discovery of Novel HIV Entry Inhibitors for the CXCR4 Receptor by Prospective Virtual Screening

Violeta I. Pérez-Nueno,*,[†] Sofia Pettersson,[†] David W. Ritchie,[‡] José I. Borrell,[†] and
Jordi Teixidó*,[†]

Grup d'Enginyeria Molecular, Institut Químic de Sarrià (IQS), Universitat Ramon Llull, Barcelona, Spain, and
INRIA Nancy Grant Est, Laboratoire Lorrain de Recherche en Informatique et ses Applications (LORIA),
UMR 7503, BP 239, 54506 Vandoeuvre-les-Nancy, France

The process of HIV entry begins with the binding of the viral envelope glycoprotein gp120 to both the CD4 receptor and one of CXCR4 or CCR5 chemokine coreceptors. There is currently considerable interest in developing novel ligands which can attach to these coreceptors and hence block virus-cell fusion. This article compares the application of structure-based (docking) and ligand-based (QSAR analyses, pharmacophore modeling, and shape matching) virtual screening tools to find new potential HIV entry inhibitors for the CXCR4 receptor. The comparison is based on retrospective virtual screening of a library containing different known CXCR4 inhibitors from the literature, a smaller set of active CXCR4 inhibitors selected from a large combinatorial virtual library and synthesized by us, and some druglike presumed inactive molecules as the reference set. The enrichment factors and diversity of the retrieved molecular scaffolds in the virtual hit lists was determined. Once the different virtual screening approaches had been validated and the best parameters had been selected, prospective virtual screening of our virtual library was applied to identify new anti-HIV compounds using the same protocol as in the retrospective virtual screening analysis. The compounds selected using these computational tools were subsequently synthesized and assayed and showed activity values ranging from 4 to 0.022 μg/mL.

INTRODUCTION

According to the World Health Organization, about 33 million people live with Acquired Immune Deficiency Syndrome (AIDS).[1] The entry of human immunodeficiency virus (HIV) into the host cell begins with binding of the viral envelope glycoprotein gp120 to both the CD4 cell surface receptor and one of CXCR4 or CCR5 chemokine coreceptors and leads to fusion of the viral capsid with the cell membrane. Current antiretroviral therapies (ARTs) against AIDS are generally based on reverse transcriptase inhibitors and protease inhibitors. Despite advances in the development of these potent agents which block HIV transcription and assembly, there remain problems regarding drug resistance, latent viral reservoirs, and drug induced toxic effects, which can all compromise effective control of the virus. Hence there is a need to develop new classes of anti-HIV drugs with different modes of action. Several researchers have recognized that knowledge of the mechanism of viral entry into the host cell provides further therapeutic targets against HIV infection.[2,3] To date, at least three subclasses of HIV viral entry/fusion inhibitors have emerged, namely the following: CD4 binding or attachment inhibitors, which target initial recognition and binding of the viral glycoprotein gp120 to the cell-surface CD4 antigen;[4] chemokine coreceptor binding inhibitors, which target binding of virus to the CCR5 or

CXCR4 coreceptor;[5] and cell fusion inhibitors, which target the gp41 viral glycoprotein.[6] Therefore, there is considerable interest in developing novel ligands which can modulate these receptors and block virus-cell fusion.[7−11]

To make progress toward this goal, we compiled a data set of CXCR4 antagonists from the literature comprising several AMD3100 derivatives, macrocycles, KRH1636 derivatives, dipicolil amine zinc(II) complexes, cyclic peptides, and tetrahydroquinolin-amine derivatives. Several of the AMD3100 derivatives are novel and have been synthesized in our group.[12] To this set was added some 4700 presumed inactive druglike compounds from the Maybridge Screening Collection[13] which have several 1D properties similar to those of the actives. The active molecules synthesized by us belong to a diverse but restricted set of compounds, selected using our PRALINS[14] program (Program for Rational Analysis of Libraries in Silico) from a large virtual combinatorial library. This library was designed to preserve the main features of AMD3100, i.e. polynitrogenated systems separated by a *p*-phenylene moiety, which is treated as an ideal reference CXCR4 antagonist. The compounds selected by PRALINS showed activities ranging from 20 to 0.008 μg/mL, and experimental binding assays confirmed that their mode of action was indeed to block the CXCR4 receptor.[12]

In order to find other active compounds without having to synthesize the whole of the combinatorial virtual library, ligand-based and structure-based virtual screening tools were used. For ligand-based virtual screening, QSAR analysis was performed with MOE,[15] and a good quantitative structure−activity relationship function was obtained. 3D pharmaco-

* Corresponding author phone: +34-93-267.20.00; fax: +34-93-2050.62.66; e-mail: j.teixido@iqs.url.es (J.T.); phone: +33 3 83 59 30 00; fax: +33 3 83 27 83 19; e-mail: violetaperezn@iqs.url.edu.
† Universitat Ramon Llull.
‡ INRIA Nancy Grant Est.

HIV Entry Inhibitors for the CXCR4 Receptor

J. Chem. Inf. Model., Vol. 49, No. 4, 2009 **811**

phore modeling using MOE and Discovery Studio[16] was applied in order to study the characteristic features of the actives necessary for interaction with the coreceptor. Shape matching using the PARAFIT,[17] ROCS,[18] and HEX[19] programs was also carried out to select molecules from the library with similar shapes to known actives. Because the 3D structure of CXCR4 has not yet been solved, a homology model of the protein built previously[20] using bovine rhodopsin[21] as the template was used for receptor-based analyses using AUTODOCK,[22] GOLD,[23] FRED,[24] and HEX.[25]

In order to validate the different virtual screening approaches and to set the best parameters for each one, a retrospective virtual screening analysis was performed on the compiled active and inactive data sets. Once the best approaches were selected, prospective analysis of the as yet unsynthesized compounds in our combinatorial virtual library was applied to establish a ranked list of new candidate CXCR4 inhibitors. A final virtual hit list was obtained from a consensus ranking of the different virtual screening approaches. Five molecules in the resulting hit list were synthesized and tested and were found to have activity values ranging from 4 to 0.022 μg/mL. The most active of these are monocyclams, as might be expected of AMD3100 derivatives,[26−28] and these coincided with the compounds in the first ranking positions of our hit list.

METHODS

Library Design. In this study, AMD3100, one of the earliest and still one of the most potent CXCR4 antagonists to be developed, was used as a reference ligand from which a combinatorial library was derived.[12,29] The compounds in this library were designed in such a way as to retain the main physicochemical features of this ligand, i.e. a central *p*-phenylene moiety with at least two nitrogen-containing substituents, one in the benzylic position and the other(s) in a heterocyclic system, and with similar distances between such nitrogens as those observed in cyclam. These considerations led us to design target compounds such as the diamines, **1,** as shown in Figure 1. A retrosynthetic analysis of those cases in which $R_1 = R_2$ and the number, n, of methyl linkers led to the selection of symmetrical diimines **2** as precursors, which can be extended with further methyls to give terephthalaldehyde (**3**) and two equivalents of the corresponding amine **4** where $n \geq 1$ (see Figure 2). When $R_1 = R_2$ and $n = 0$, compounds **2** are in fact symmetrical hydrazones which can be obtained by condensation of terephthalaldehyde and the corresponding hydrazine **4** ($n = 0$). These dihydrazones were also included in our library. In order to obtain nonsymmetric ($R_1 \neq R_2$) diamines **1** ($n \geq 1$) and dihydrazones **2** ($n = 0$), it was necessary to modify slightly our synthetic approach by using 4-(diethoxymethyl)benzaldehyde (**5**) as the core precursor. Thus, the intermediate hydrazono and aminobenzaldehydes **6** and **7** allowed such nonsymmetric compounds and other nonsymmetric aminohydrazones **8** to be included as further compounds in the combinatorial library (see Scheme 1). Overall, the virtual library consists of 66 amino/hydrazono-amine/hydrazone compounds (**1**, **2**, and **8**), 11 amino/hydrazono-aldehyde compounds (**6** and **7**), and 11 cyclam-amine/hydrazone compounds (**9** and **10**). Some representative examples of these structures are shown in Figure 3.
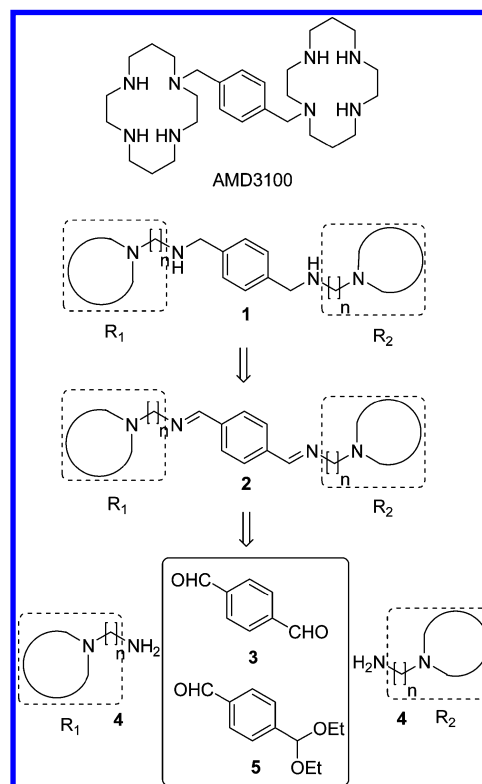

Figure 1. The AMD3100 reference antagonist for CXCR4 and schematic illustration of the target library construction. Top: the AMD3100 reference antagonist for CXCR4, with a *p*-phenyl linker and nitrogen-containing heterocyclic systems on each side of the linker. Bottom: a schematic illustration of the construction of the target library which preserves these features.
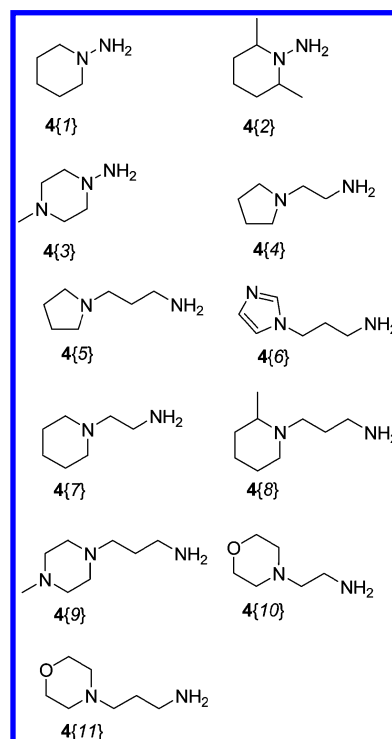
Figure 2. Amine and hydrazine building blocks used for the combinatorial virtual library.

Virtual Screening Data Sets. For the retrospective virtual screening analysis, a data set of 248 CXCR4 antagonists with activity values lower than 100 μM against CXCR4 was assembled from the literature. This set was used for receptor-
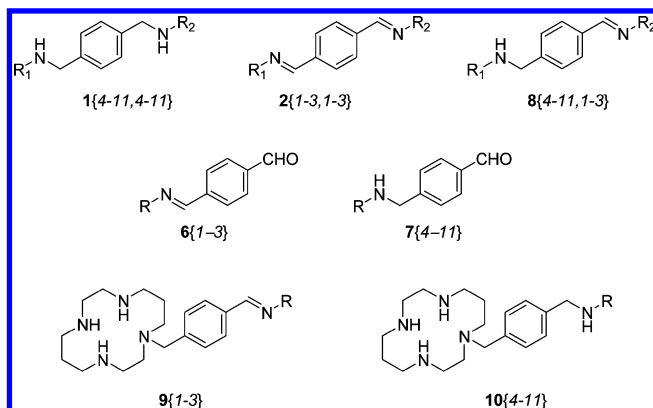
Scheme 1. Synthetic Scheme for the Symmetrical and Nonsymmetrical Diamines **1**, Dihydrazones **2,** and Aminohydrazones **8**
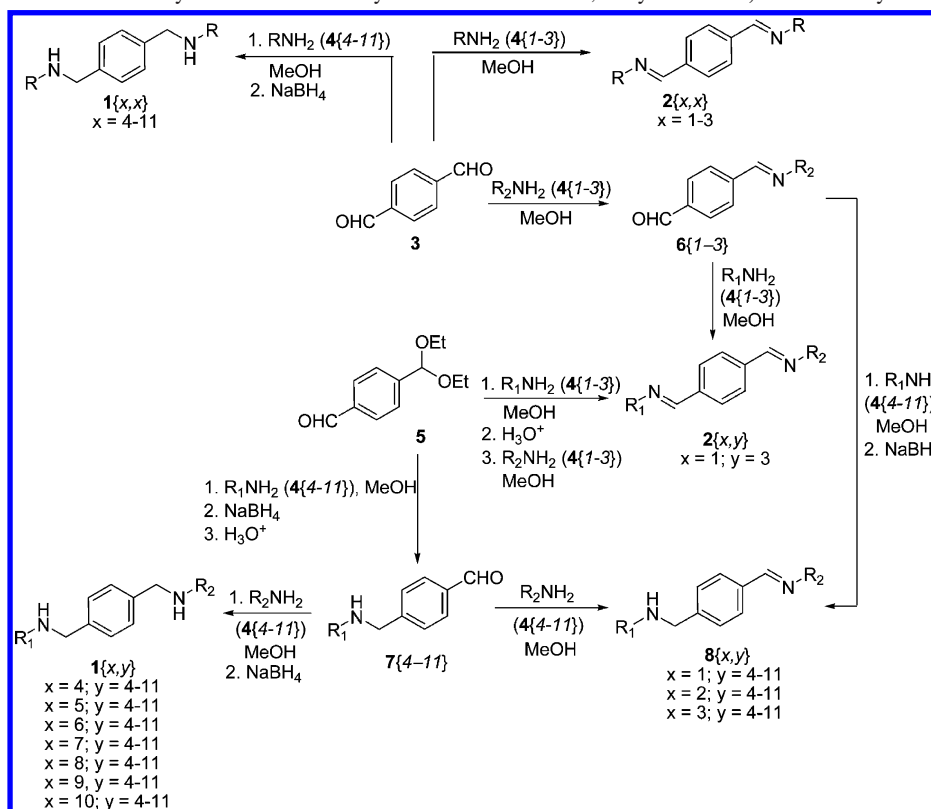
based docking and ligand-based screening analyses. A subset of the 103 most active compounds plus 48 compounds representative of other scaffold classes was then used for pharmacophore modeling. As summarized in Table 1, these compounds mainly belong to seven representative families, i.e., AMD3100 derivatives, macrocycles, KRH1636 derivatives, dipicolil amine zinc(II) complexes, tetrahydroquinolin-amine derivatives, cyclic peptides, and also the most active CXCR4 inhibitors from our combinatorial virtual library which had been synthesized by us.[12] Figure 4 shows some representative members of each family. These data sets were augmented with two further sets of druglike presumed inactive compounds from the Maybridge Screening Collec-

Figure 3. Representative examples of compounds in the combinatorial virtual library. Compounds **1** are symmetrical ($R_1 = R_2$) or nonsymmetrical ($R_1 \neq R_2$) diamines, compounds **2** are symmetrical ($R_1 = R_2$) or nonsymmetrical ($R_1 \neq R_2$) dihydrazones, compounds **8** are aminohydrazones, compounds **6** and **7** correspond to hydrazonobenzaldehydes and aminobenzaldehydes, respectively, and compounds **9** and **10** are hydrazone or amino substituted monocyclams.

tion (1462 for pharmacophore modeling and 4696 for the docking and shape matching approaches), selected in such a way that several of their 1D properties were similar to those of the actives (i.e., molecular weight, number of rotatable single bonds, numbers of hydrogen-bond donor and acceptor atoms, number of hydrophobic atoms, and octanol−water partition coefficient), as shown in Table 2.

For the prospective virtual screening analysis, the same presumed inactive compounds as in the retrospective analysis were used, and a subset of 34 hitherto unsynthesized compounds from the amino/hydrazono-amine/hydrazone (compounds **1**, **2**, and **8**), hydrazono/amino-aldehyde (compounds **6** and **7**), and cyclam-hydrazone/amine (**9** and **10**) families was selected from the virtual library for synthesis and testing. The 3D structures of all compounds were protonated at physiological pH, assigned Gasteiger partial charges, and geometry-optimized using the MMFF94 force field. All molecules were aligned with the MOE FlexAlign module[46] using as superposition template the AMD3100 conformation obtained previously from a CXCR4 docking study[20] (see Figure 5).

QSAR Analysis. QSAR analysis applies statistical methods to describe quantitative relationships between chemical structures and biological activities of a series of analogues. The process can be divided into three general steps: 1) data set selection, 2) data analysis, and 3) model validation. In the present QSAR study, a data set of 39 compounds with known EC_{50} activity values consisting of AMD3100 plus 38 further compounds synthesized by us (structures **1, 2, 6, 7,** and **8**) was used. This data set was divided into a training subset of 30 compounds and an external test set of 9 compounds, as described in Tables 3 and 4. A total of 194 descriptors were calculated with

Table 1. Summary of the CXCR4 Inhibitor Families Used in the Current Study

family	no. of compds	refs
CXCR4 Inhibitors for Retrospective Docking and Shape Based Virtual Screening		
tetrahydroquinolin-amine derivatives	123	11, 30−34
KRH1636 derivatives	23	11, 35−38
macrocycles	4	39
AMD3100 derivatives	94	11, 26, 27, 39−42
cyclic peptides	2	43
other	2	44
total	248	
CXCR4 Inhibitors for Retrospective Pharmacophore Model Based Virtual Screening		
KRH1636 derivatives	13	11, 35−38
dipicolil amine zinc(II) complexes	10	45
AMD3100 derivatives and macrocycles	90	11, 26, 27, 39−42
active molecules from the combinatorial virtual library (amino-amine, amino-aldehyde)	38	12
total	151	

MOE, including 2D and 3D descriptors. These descriptors were then pruned using correlation analysis and forward-selection and backward-elimination methods.

Partial Least Squares (PLS) regression was used to build the QSAR models using the above descriptors as independent variables and using the biological activities as the dependent variables. Model outliers were detected using the Grubbs test,

as implemented in MOE, by quantifying how far away the experimental biological activities are from the model by calculating the *Z-SCORE* ratio, defined as the difference between the experimental and model pEC_{50} values divided by the RMSE (root mean squared error) of the whole data set. Molecules with *Z-SCOREs* of 2.5 or higher were considered to be possible outliers. The model was then
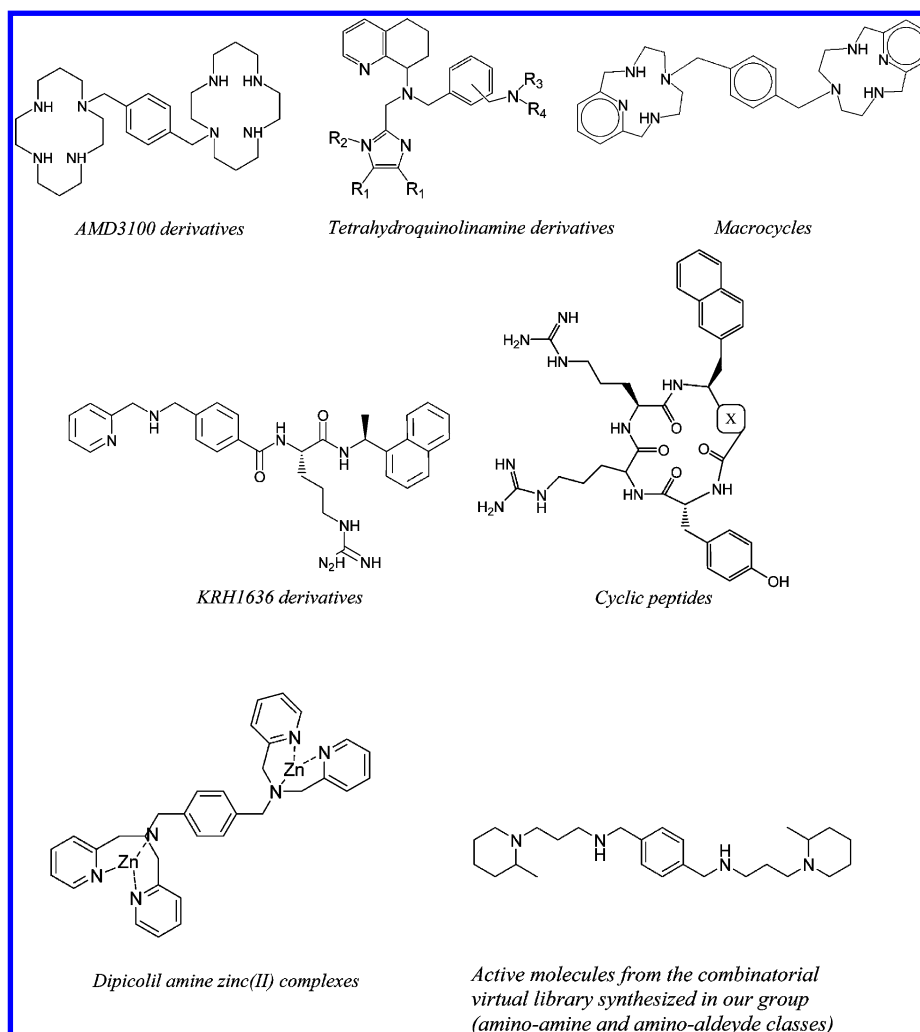

Figure 4. Representative structures of seven families of CXCR4 inhibitor.

Table 2. Summary of the 1D Physico-Chemical Properties of Active and Inactive Molecules in the Screening Databases Used in Pharmacophore Modelling, Docking, and Shape Matching Approaches[a]

comparison of data sets used in pharmacophore modeling	MW	b_1rotN	a_acc	a_don	a_hyd	SlogP
151 CXCR4 actives	485.2 (104.9)	6.9 (3.9)	3.5 (1.5)	1.5 (1.5)	26.3 (5.0)	−0.8 (2.5)
1462 inactives	381.4 (64.9)	5.1 (2.0)	4.0 (1.1)	1.2 (1.1)	16.6 (2.6)	2.6 (0.9)
comparison of data sets used in docking and shape matching approaches	MW	b_1rotN	a_acc	a_don	a_hyd	SlogP
248 CXCR4 actives	507.3(74.4)	9.2(4.9)	4.9(1.1)	1.7(1.3)	27.6(4.2)	4.3(3.0)
4696 inactives	497.4(45.6)	6.2(2.4)	3.6(1.6)	0.9(1.0)	21.8(4.1)	5.5(1.9)

[a] This table shows the average and standard deviation (in parentheses) of the following properties: MW (molecular weight); b_1rotN (number of rotatable single bonds); a_acc (number of hydrogen-bond acceptor atoms); a_don (number of hydrogen-bond donor atoms); a_hyd (number of hydrophopic atoms); S_logP (octanol-water partition coefficient).
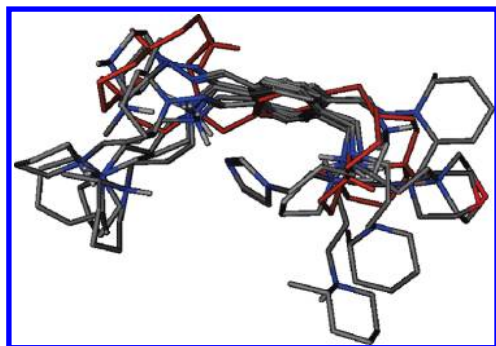
Figure 5. The MOE alignments of active database compounds with AMD3100 (shown in brown).

validated using leave-one-out (LOO) cross-validation and validation with an external test set (9 compounds). Several statistical parameters were used to evaluate the performance of the model:

• Correlation coefficient R2, cross-validated R2, and test set validation R2 against an external data set, where x is the experimental pEC50 and y is the model value

$$R = \frac{\sum (x - \bar{x}) \cdot (y - \bar{y})}{\sqrt{\sum (x - \bar{x})^2 \cdot \sum (y - \bar{y})^2}} \quad (1)$$

• Root mean squared error, *RMSE*, for the model, the cross-validation, and the external test set validation, where *PRESS* is the prediction error sum of squares and n the number of compounds

$$RMSE = \sqrt{\frac{PRESS}{n}} \quad (2)$$

PRESS is an important cross-validation parameter to measure the accuracy of a model. When *PRESS* is less than *SSY* (sum of the squared deviations for the experimental values from their mean), it indicates that the model is significant and predicts better than chance. Furthermore, a *PRESS/SSY* ratio of less than 0.4 indicates that the model is a reasonable QSAR model.[47]

• Cross-validated R2 has widely been used as a criterion of model robustness and predictive ability, with a threshold of 0.5 (0.6 for model R2).[48] Nevertheless, a high cross-validated R2 is considered a necessary condition for a model to have a high predictive power, but it is not a sufficient condition. Therefore, models are often evaluated with

external test sets to estimate their true predictive power. For example, Tropsha et al. consider a QSAR model to be predictive if the following conditions are satisfied[49]

$$\frac{R^2 - R_0^2}{R^2} < 0.1 \quad (3)$$

$$0.85 \leq k \leq 1.15 \quad (4)$$

where R_o is the correlation coefficient, and k is the value of the slope for the regression line through the origin (i.e., with the intercept set to 0).

• The Fisher test, or F-test, reflects the ratio of the variance explained by the model and the variance due to the error in the model. High values of the F-test indicate the reliability of the QSAR equation.

Ligand-Based Pharmacophore Modeling. Pharmacophore modeling studies were performed using the MOE and Discovery Studio software suites with four families of known actives from the above virtual screening data set, namely the following: AMD3100 derivatives, KRH1636 derivatives, dipicolil amine zinc(II) complexes, and the most active CXCR4 inhibitors from the combinatorial virtual library. 50 conformations and a maximum of 255 conformations of each compound were calculated in MOE (MMFF94 forcefield) and Discovery Studio (Catalyst Confirm algorithm), respectively. The training set consisted of the most active compound from each family of CXCR4 inhibitors. The pharmacophore queries were built on the alignment of these four structures with the FlexAlign module in MOE and using the Common Feature Pharmacophore Generation protocol in Discovery Studio. The pharmacophore scheme of PCH (polarity-charge-hydrophobicity) was applied throughout the MOE study. Chemical features and their tolerance radii were selected between those suggested by MOE to achieve better balance between sensitivity and specificity. Also, in Discovery Studio, hydrogen bond acceptor, hydrogen bond donor, hydrophobic, ionizable positive, and charged positive pharmacophore features were used. The maximum number of omitted features was set to one.

Ligand-Based Shape Matching Virtual Screening. Shape based virtual screening was performed using PARAFIT 08 Shape Tanimoto, ROCS 2.2 Combo Score and Shape Tanimoto, and HEX 4.8 Shape Tanimoto scores by superposing each database compound onto the docked AMD3100 query conformation. The PARAFIT and HEX superpositions were calculated using the conformation of each database

HIV Entry Inhibitors for the CXCR4 Receptor

J. Chem. Inf. Model., Vol. 49, No. 4, 2009 **815**

Table 3. Training Set Used for the QSAR Model Building Calculations[a]

	Compound	Name	pEC$_{50}$	predicted pEC$_{50}$	Residue		Compound	Name	pEC$_{50}$	predicted pEC$_{50}$	Residue
1		8{1,11}	4.512	4.864	-0.352	16		1{5,7}	5.327	5.465	-0.138
2		8{2,4}	4.367	5.303	-0.936	17		1{5,10}	5.177	5.070	0.107
3		8{2,5}	5.251	5.772	-0.521	18		1{6,7}	5.254	5.321	-0.067
4		8{2,9}	5.005	4.957	0.048	19		1{6,8}	1.106	outlier	outlier
5		8{3,5}	5.281	4.695	0.586	20		1{6,11}	4.305	5.038	-0.733
6		8{3,6}	4.483	4.578	-0.095	21		1{7,9}	5.190	5.520	-0.330
7		8{3,8}	5.424	5.135	0.289	22		1{7,11}	5.142	5.105	0.037
8		8{3,9}	4.503	4.238	0.265	23		1{8,8}	7.715	6.843	0.872
9		8{3,11}	4.647	4.061	0.586	24		1{8,10}	5.987	5.599	0.388
10		1{4,4}	4.511	4.917	-0.406	25		1{8,11}	5.906	5.890	0.016
11		1{4,5}	5.299	5.291	0.008	26		1{9,9}	4.642	5.047	-0.405
12		1{4,6}	4.852	4.962	-0.110	27		6{2}	4.432	4.724	-0.292
13		1{4,9}	4.659	4.913	-0.254	28		6{1}	4.235	4.169	0.066
14		1{5,5}	5.600	5.548	0.052	29		7{9}	4.233	3.811	0.422
15		1{5,6}	6.250	5.351	0.899	30		AMD3100	8.688	8.688	0

[a] pEC$_{50}$ (derived from EC$_{50}$ in μM) refers to the experimental activity values. The two last columns show the predicted and residual pEC$_{50}$ values obtained from QSAR model 1. Compound **19** gave a *Z-SCORE* > 2.5 and was therefore considered to be an outlier and was excluded from the training set.

compound that was calculated by MOE FlexAlign. However, as described previously,[20] the ROCS superpositions used ten further conformations of each molecule calculated by OMEGA.[50] Spherical harmonic consensus shape matching[51] was also performed using PARAFIT 08 by superposing each database compound onto a consensus shape query molecule calculated from three known CXCR4 actives from different scaffold families (an AMD derivative, a macrocycle derivative, and a KRH derivative). Database molecules were ranked according to their shape Tanimoto scores with respect to the query shape. The ROCS calculations also used the "color optimization" mode to maximize both the shape and chemical property overlays (e.g., proton donor/acceptor, cationic/anionic, and hydrophobicity/aromaticity).

Receptor-Based Virtual Screening. Receptor-based screening against CXCR4 was performed using AUTODOCK 3.0, GOLD 3.0.1, FRED 2.2.1, and HEX 4.8. In AUTODOCK and GOLD, ten independent LGA and GA runs were carried out, respectively, using the same protocol as described.[20] In GOLD,

the ligands were constrained to form a hydrogen bond with a carbonyl oxygen of either Glu288, Asp171, or Asp262 which had been identified previously as key binding residues by site-directed mutagenesis (SDM).[44,52–54] The ligand databases were ranked by AUTODOCK Docked Energy, Gold GoldScore and ChemScore, and a consensus score "Rank-by-Rank"[55] of these three scoring functions. In FRED, exhaustive rigid body optimization was carried out starting from the ligand conformations aligned to the docked AMD3100 conformation. PLP, Chemgauss3, Shapegauss, OEChemScore, ScreenScore, ChemScore scoring functions, and a consensus combination of these scores were used to rank the ligand databases. In HEX, docking and ranking was performed using a six-dimensional shape-only superposition correlation search with a translational distance range of 10 Å from the SDM-defined active site center and Hex Docked energy, respectively.

Analyzing Virtual Screening Hit Lists and Pharmacophores. Before virtual screening protocols and pharmacophoric models may be used prospectively, it is first

Table 4. External Test Set Used for QSAR Model Validation[a]

	Compound	Name	pEC$_{50}$	predicted pEC$_{50}$	Residue
1		**8**{1,5}	5.079	5.297	-0.218
2		**8**{2,11}	4.375	5.006	-0.631
3		**1**{5,8}	6.357	6.045	0.312
4		**1**{5,9}	5.889	5.336	0.553
5		**1**{5,11}	5.369	5.239	0.130
6		**1**{8,9}	7.142	5.783	1.359
7		**1**{9,11}	4.647	5.101	-0.454
8		**7**{5}	4.337	4.249	0.088
9		**7**{8}	5.183	4.962	0.221

[a] The column headings are described in Table 3.

Table 5. Correlation Analysis for the Descriptors Used in the QSAR Models

	pEC$_{50}$	VAdjEq	Q_VSA_HYD	dipoleY	SlogP_VSA8	SMR_VSA5	FASA+
pEC$_{50}$	100						
VAdjEq	−60	100					
Q_VSA_HYD	84	−68	100				
dipoleY	41	−14	20	100			
SlogP_VSA8	56	−26	41	42	100		
SMR_VSA5	54	−39	37	49	68	100	
FASA+	38	−16	59	−6	−34	−12	100

necessary to validate them by measuring their ability to retrieve actives from a database of compounds with known biological activities. Several formulas have been proposed to score quantitatively the quality of hit lists achieved in this way.[56] For example, for a database of *D* compounds containing *A* actives, and where H_t is the number of
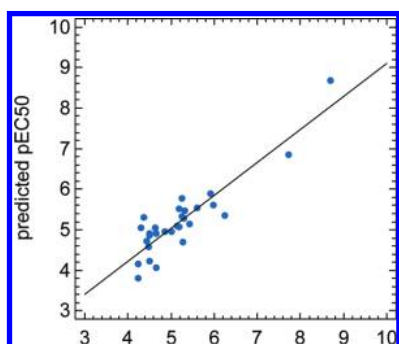
Figure 6. Correlation of experimental versus predicted pEC$_{50}$ for QSAR model 1.

compounds in a hit list, and H_a is the number of actives in that list, the following terms may be defined:[57]

Percent yield of actives:

$$Y(\%) = \frac{H_a}{H_t} \times 100 \quad (5)$$

Percent ratio of the actives in the hit list:

$$A(\%) = \frac{H_a}{A} \times 100 \quad (6)$$

Enrichment (enhancement):

$$EF = \frac{H_a/H_t}{A/D} = \frac{H_a \times D}{H_t \times A} \quad (7)$$

Goodness of Hit list:

$$GH = \left(\frac{H_a(3A + H_t)}{4H_tA}\right) \times \left(1 - \frac{H_t \times H_a}{D - A}\right) \quad (8)$$

False Negatives:

$$A - H_a \quad (9)$$

False Positives:

$$H_t - H_a \quad (10)$$

For each scoring method, the resulting hit lists were analyzed using the above terms. Following the pharmacoph-

HIV Entry Inhibitors for the CXCR4 Receptor

J. Chem. Inf. Model., Vol. 49, No. 4, 2009 **817**

Table 6. Prediction of Activity Values Using QSAR Model 1

Compound	Name	predicted EC$_{50}$ (μM)
	1{7,8}	0.66
	10{8}	0.87
	8{2,8}	1.58
	8{2,7}	1.88
	1{4,8}	2.16
	2{1,2}	2.24
	10{7}	2.70
	10{6}	2.87
	1{7,7}	2.93
	8{1,8}	3.70
	9{2}	3.85
	10{9}	4.61
	10{4}	4.75
	10{5}	5.23
	9{1}	5.90
	1{4,7}	5.94
	8{1,7}	5.98
	10{11}	7.30
	1{6,6}	7.53
	8{1,4}	9.28
	1{7,10}	9.95
	8{2,10}	10.13
	2{2,3}	10.52

Table 6. Continued
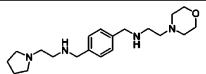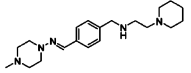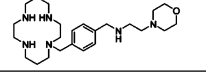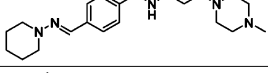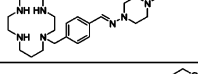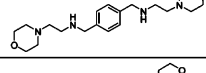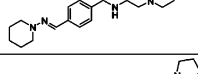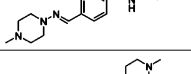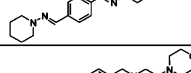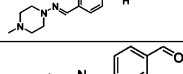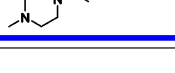
Compound	Name	predicted EC$_{50}$ (μM)
	1{*4,10*}	17.35
	8{*3,7*}	18.96
	10{*10*}	19.70
	8{*1,9*}	21.50
	9{*3*}	24.95
	1{*10,10*}	30.06
	8{*1,10*}	30.44
	8{*3,4*}	43.40
	2{*1,3*}	48.09
	8{*3,10*}	58.34
	6{*3*}	> 100

Table 7. Summary of the Results Obtained for the Retrospective Screening Analysis of the Generated Pharmacophore Models[a]

model	H_a	H_t	false +	false -	EF	Y(%)	A(%)	GH
1	142	142	0	9	10.68	100	94	0.99
2	139	140	1	12	10.61	99	92	0.97
3	122	157	35	29	8.30	78	81	0.77
4	133	186	53	18	7.64	72	88	0.73
5	123	168	45	28	7.82	73	81	0.73
6	132	132	0	19	10.68	100	87	0.97
7	96	96	0	55	10.68	100	64	0.91
8	106	107	1	45	10.58	99	70	0.92
9	92	92	0	59	10.68	100	61	0.90

[a] The quantities H_a, H_t, false +, false -, EF, Y(%), A(%), and GH are defined in eqs 5−10. Overall, QSAR model 1 can be seen to give the best statistics.
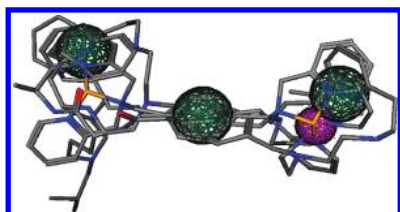
Figure 7. Alignment of the compounds in the training set and the pharmacophore model 1. Hydrophobic and aromatic features (Hyd|Aro) are shown in green. Cationic (Cat) features are shown in purple.

ore modeling, shape matching, and docking calculations, all compounds were sorted into ranked lists based upon their rmsd, shape matching scores, and docking energies, respec-

tively. These lists were then used to plot the percentage of known actives found versus the percentage of the ranked database screened and to calculate enrichment factors (EFs) at 1%, 5%, and 10% of the screened database.

RESULTS

PLS Analysis and Validation of QSAR Models. After descriptor pruning had been applied, five descriptors were selected to build the QSAR models, namely the following: VAdjEq, Q_VSA_HYD, dipoleY, SlogP_VSA8, and FASA+. Table 5 shows the correlation analysis for these descriptors. Three QSAR models were calculated as follows:

Model 1 (Figure 6):

$$pEC_{50} = 2.52586 + 0.00940 \cdot (Q\_VSA\_HYD) + 0.00507 \cdot (SlogP\_VSA8) + 0.10611 \cdot (dipoleY)$$

$N = 29$, $R^2 = 0.81$, RMSE = 0.42, $F = 36.45$, $R^2_{LOO} = 0.75$, RMSE$_{LOO}$ = 0.49, $R^2_{test} = 0.69$, RMSE$_{test}$ = 0.57, $n = 9$, $R_o^2 = 0.77$, $(R^2 - R_o^2)/R^2 = 0.049$, $k = 0.99$, PRESS = 5.20, SSY = 27.93, PRESS/SSY = 0.19

Model 2:

$$pEC_{50} = 2.52568 + 0.00940 \cdot (Q\_VSA\_HYD) + 0.10611 \cdot (dipoleY) + 0.00507 \cdot (SlogP\_VSA8) + 0.00130 \cdot (FASA+)$$

$N = 29$, $R^2 = 0.81$, RMSE = 0.42, $F = 26.24$, $R^2_{LOO} = 0.75$, RMSE$_{LOO}$ = 0.49, $R^2_{test} = 0.69$, RMSE$_{test}$ = 0.57, $n = 9$, $R_o^2 = 0.77$, $(R^2 - R_o^2)/R^2 = 0.049$, $k = 0.99$, PRESS = 5.20, SSY = 27.93, PRESS/SSY = 0.19

HIV Entry Inhibitors for the CXCR4 Receptor

J. Chem. Inf. Model., Vol. 49, No. 4, 2009 **819**

Table 8. Pharmacophore-Based Prospective Virtual Screening Results[a]

compound	model 1	model 2	model 3	model 4	model 5
2{1,2}	-	-	-	-	-
2{1,3}	0.9169	0.9086	-	0.9870	-
8{1,4}	0.7352	0.5261	0.5712	0.7153	1.4686
8{1,7}	-	0.5190	0.6307	0.7327	-
8{1,8}	0.6071	0.4723	0.6284	0.9790	-
8{1,9}	0.6714	0.3698	0.5487	0.7998	-
8{1,10}	-	0.5110	0.6307	0.7305	1.2161
2{2,3}	0.7664	0.9129	-	0.8400	-
8{2,7}	-	0.8506	0.6338	0.6678	1.3628
8{2,8}	0.4375	0.4742	0.6199	0.6860	-
8{2,10}	-	0.7040	0.6339	0.6660	0.8185
8{3,4}	0.4417	0.5261	0.5711	0.6690	0.9981
8{3,7}	0.5551	0.4368	0.6306	0.6317	0.8388
8{3,10}	0.5551	0.4289	0.6306	0.6291	0.8327
1{4,7}	0.7565	0.3962	0.5227	0.4716	0.8880
1{4,8}	0.4157	0.3923	0.4166	0.4902	0.8354
1{4,10}	0.7308	0.3284	0.5119	0.5366	0.7563
1{6,6}	0.6027	0.9477	0.4127	-	-
1{7,7}	0.6990	0.5077	0.5190	0.6907	1.0076
1{7,8}	0.4157	0.3625	0.4166	0.5070	0.8043
1{7,10}	0.7308	0.3284	0.5119	0.4728	0.7508
1{10,10}	1.1377	0.4894	0.5079	0.4643	0.7041
6{3}	-	0.9007	-	-	-
9{1}	0.4288	0.4014	0.4803	0.5879	1.1812
9{2}	0.3235	0.3832	0.4742	0.5254	0.8217
9{3}	0.4747	0.4012	0.4805	0.5449	0.5568
10{4}	0.3409	0.4270	0.4640	0.4845	0.6615
10{5}	0.3671	0.3329	0.4250	0.4800	0.7907
10{6}	0.3925	0.4079	0.4250	0.4800	1.1394
10{7}	0.2826	0.3789	0.4512	0.3739	0.5394
10{8}	0.3321	0.4755	0.4248	0.3568	0.5311
10{9}	0.2655	0.3462	0.4248	0.3320	0.5408
10{10}	0.2809	0.3786	0.4512	0.3727	0.5405
10{11}	0.3755	0.2856	0.4248	0.3483	0.5304

[a] This table lists the overall (RMSD) score obtained for each compound using pharmacophore models 1, 2, 3, 4, and 5. Hyphens denote compounds that do not match the pharmacophore model.

Model 3:

$$pEC_{50} = 2.52606 + 0.00940 \cdot (Q\_VSA\_HYD) + 0.00507 \cdot (SlogP\_VSA8) + 0.10611 \cdot (dipoleY) - 0.00040 \cdot (VAdjEq)$$

$N = 29$, $R^2 = 0.81$, RMSE $= 0.42$, $F = 26.24$, $R^2_{LOO} = 0.75$, RMSE$_{LOO} = 0.49$, $R^2_{test} = 0.69$, RMSE$_{test} = 0.57$, $n = 9$, $R_o^2 = 0.77$, $(R^2-R_o^2)/R^2 = 0.049$, $k = 0.99$, PRESS $= 5.20$, SSY $= 27.93$, PRESS/SSY $= 0.19$

One compound **1**{6,8} was deleted from the training set because it gave a *Z-SCORE* > 2.5, which indicated it is an outlier. This was confirmed by recalculating the models without it to obtain better overall statistics. All three resulting models showed R^2 values above 0.6 and R^2 for the cross-validation and external test set validation above 0.5. In all cases, the *PRESS/SSY* ratio was below 0.4, $(R^2-R_o^2)/R^2$ was less than 0.1, and k was between the above thresholds. Because the statistical results were broadly similar for all models, model 1 was selected as the most parsimonious because it used only three descriptors, whereas models 2 and 3 required four descriptors. The use of *dipoleY*, an external 3D descriptor, as independent variable in the three models enhanced the importance of a correct alignment of the molecules in order to obtain a reliable predicted activity value. Prediction of activity values for the training set and the external

test set using model 1 are shown in Tables 3 and 4. Predictions were made for compounds in the virtual library that had not yet been synthesized (**1, 2**, and **8**) and for monocyclams **9** and **10**. These results are shown in Table 6.

Pharmacophore Hypothesis Generation and Validation. Pharmacophore models were generated and retrospective analyses were performed to select models which achieved a good balance between sensitivity and specificity. Several models were proposed (Table 7), five using MOE (Models 1 to 5) and four using Discovery Studio (Models 6 to 9). Model 1 was built using the MOE Pharmacophore Elucidate module. Features in models 2 and 3 were selected from the consensus analysis performed with MOE Pharmacophore Query module. Models 4 and 5 were manually designed based on the description of the interactions of AMD3100 and CXCR4.[44,54,58−62] Finally, models 6 to 9 were built in Discovery Studio using the Hypogen and HipHop algorithms to generate hypotheses and to select the best common pharmacophore features produced. The retrospective analysis of the models showed that pharmacophore model 1 (Figure 7) was highly selective with our data set, giving no false positives and only nine false negatives. This model accurately classified and ranked all the known actives in the data set, except for the KRH1636 analogues which were positioned at the end of the hit list. Visual inspection of the hit lists in the retrospective analysis showed that the ranking of each compound depended on the model and type of compound. More reliable results were obtained using a consensus of the five MOE models in Table 7.

A prospective analysis using the consensus pharmacophore model was then applied to select new compounds for synthesis and testing. These molecules included hitherto unsynthesised compounds from the virtual combinatorial library (i.e., amino/hydrazono-amine/hydrazone compounds **1, 2,** and **8**, amino/hydrazono-aldehyde compounds **6** and **7** and cyclam-amine/hydrazone compounds **9** and **10**). All of these compounds can be seen to match the pharmacophore model equally well. The screened compounds selected by the consensus of pharmacophore models and their score values are shown in Table 8.

Docking Enrichments. In order to analyze the ability of the receptor model structure to discriminate active compounds from decoys, retrospective analysis of docking enrichment curves was performed as described previously.[20] Next, enrichment curves for the virtual combinatorial library compounds were calculated using the same protocol. Figure 8 shows the enrichment curves obtained. Inspection of these results shows that the enrichments obtained with the FRED consensus, Consensus scoring (AUTODOCK Docked Energy, GOLD GoldScore and ChemScore), and ChemScore scoring functions are the best, as was observed in the retrospective analysis. Looking at the first percentages of the ranked hit lists, the compounds selected by these three scoring functions can be seen to belong to **9, 10, 1, 2,** and **8**. The compounds found at the top 10% of the ranked hit list using these three scoring functions as well as AUTODOCK Docked Energy, HEX Docked Energy, and FRED Chemgauss3 are nearly the same. The screened compounds selected by these scoring functions and their score value are shown in Table 9.
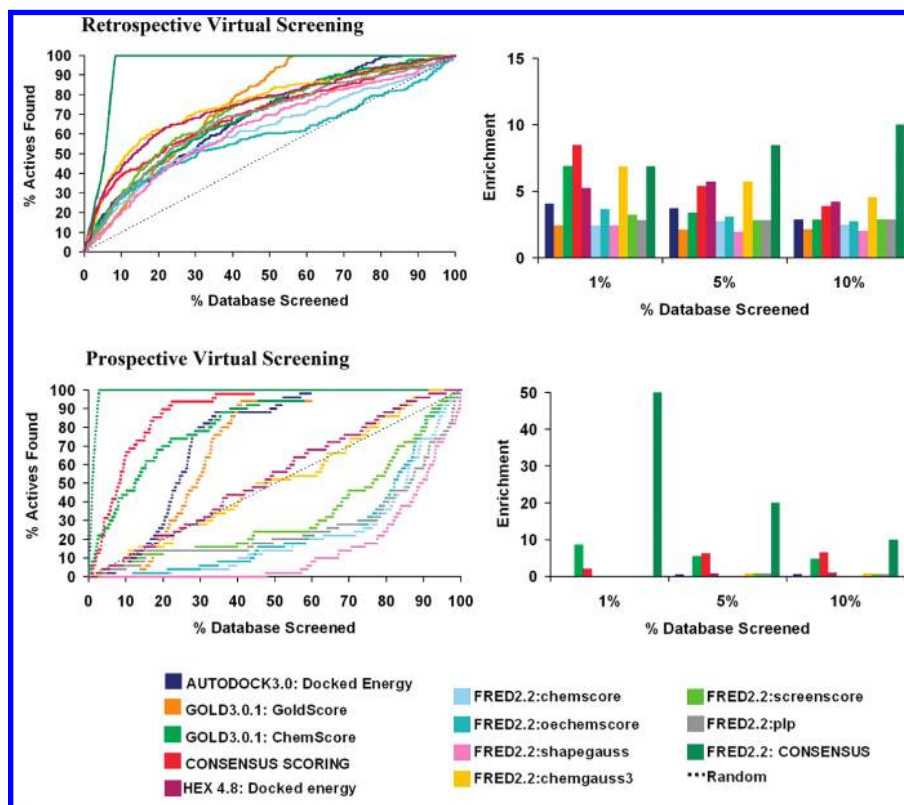
Figure 8. CXCR4 docking-based enrichment plots. On the left, enrichment results for several docking protocols for retrospective (top) and prospective (bottom) virtual screening analyses. The dotted black line represents the expected values if actives are selected at random. On the right, enrichment factors for actives found within the top-ranking 1%, 5%, and 10% of the screened inhibitor database (top) and screened virtual combinatorial library (bottom).

Table 9. Docking Scores of Hits from the Screened Combinatorial Library[a]

compound	consensus score	GOLD ChemScore	FRED consensus	AUTODOCK docked energy	HEX docked energy	FRED Chemgauss3
10{*8*}	512.70	27.20	446	−16.27	−345.80	−16.97
10{*9*}	455.30	25.88	408	−17.69	−373.90	−30.96
10{*11*}	1129.30	8.07	388	−15.76	−396.30	−31.67
10{*5*}	920.30	15.29	464	−16.59	−343.70	−14.77
9{*2*}	707	26.42	329	−14.78	−409.50	−29.98
8{*2,8*}	743.70	35.55	119	−14.59	−374.70	−47.45
1{*7,8*}	738	31.29	427	−15.14	−407.80	−62.70
8{*1,8*}	904.30	25.06	212	−14.69	−325.70	−49.30
1{*4,10*}	922.70	28.29	207	−14.03	−421.60	−46.24
8{*2,7*}	656	30.80	160	−14.46	−421.20	−57.50
1{*4,8*}	492.70	33.11	422	−14.92	−443.60	−42.65
1{*7,10*}	738.30	26.30	341	−14.10	−420.50	−40.59
1{*4,7*}	683	28.39	185	−15.90	−381.3	−69.14

[a] This table shows compounds found within the top 10% of the ranked hit list using Consensus score (AUTODOCK Docked Energy, GOLD GoldScore, and ChemScore), GOLD ChemScore, FRED consensus (PLP, Chemgauss3, Shapegauss, OEChemScore, ScreenScore, ChemScore), AUTODOCK Docked Energy, HEX Docked Energy, and Chemgauss3 scoring functions.

Shape Matching Enrichments. Because no crystallographic ligand conformation is available for the current system, the SDM-compatible conformation of AMD3100 found previously from computational docking was used as the database query. In order to study the performance of this query structure and the parameters used in the screening protocol, a retrospective analysis of shape matching enrichment curves was first performed.[20] Next, enrichment curves for the combinatorial virtual library compounds were calculated using the same protocol. Moreover, a consensus query was built from three different scaffold CXCR4 known actives (an AMD derivative, a macrocycle derivative, and a KRH derivative), and a retrospective analysis was performed.

Enrichment curves for the virtual combinatorial library compounds were also calculated showing similar results to the basic PARAFIT AMD3100 query shape Tanimoto score. Figure 9 shows that the ROCS Combo Score and PARAFIT Tanimoto Score and Consensus Shape Tanimoto give the best EFs, as in the retrospective analysis. HEX Shape Tanimoto and ROCS Shape Tanimoto also perform well. Overall, the ligand-based shape matching tools perform better than the docking tools used here. However, looking at the first percentages of the ranked hit lists obtained, the compounds selected by these shape matching methods belong to **9**, **10**, **1**, **2,** and **8**, as found with the docking tools. Molecules found at the top 10% hit ranking list are the same

HIV ENTRY INHIBITORS FOR THE CXCR4 RECEPTOR

J. Chem. Inf. Model., Vol. 49, No. 4, 2009 **821**
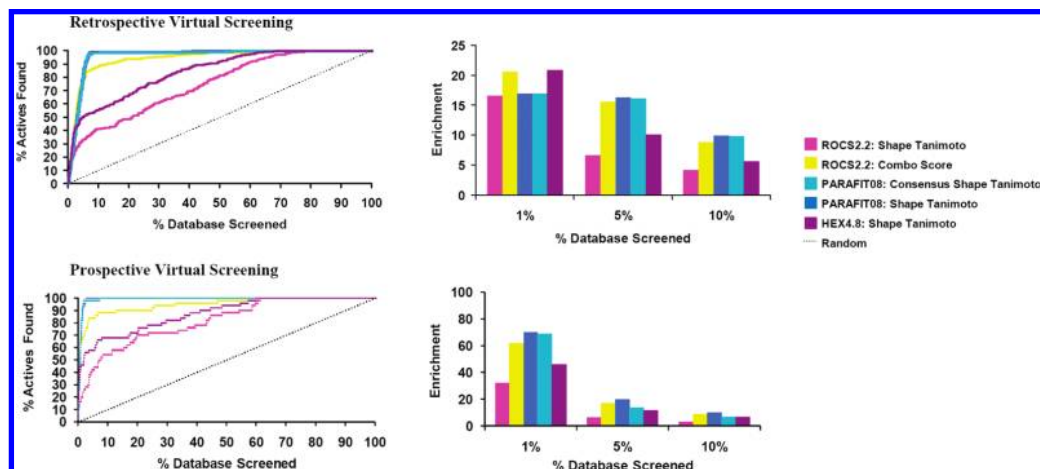
Figure 9. CXCR4 shape matching-based enrichments. On the left, enrichment curves obtained for various shape matching protocols on the known inhibitor database (top) and compounds from the virtual combinatorial library (bottom). The dotted line represents the expected enrichment if actives are selected at random. On the right, enrichment values for actives found within the top-ranking 1%, 5%, and 10% of the screened database (top) and screened virtual combinatorial library (bottom).

Table 10. Shape Matching Scores for Hits from the Screened Combinatorial Library[a]

compound	PARAFIT Shape Tanimoto	PARAFIT Consensus Shape	ROCS Shape Tanimoto	ROCS Combo Score	HEX Shape Tanimoto
10{*11*}	0.9725	0.9763	0.5970	0.9010	0.9109
10{*8*}	0.9492	0.9506	0.5070	0.6320	0.8479
10{*9*}	0.8996	0.9028	0.4820	0.8160	0.8422
10{*5*}	0.9193	0.9334	0.5280	0.8720	0.8490
9{*2*}	0.9570	0.9507	0.6330	0.6940	0.9145
8{*2,8*}	0.9651	0.966	0.5170	0.6840	0.8741
1{*7,8*}	0.9597	0.9471	0.5420	0.7320	0.8855
8{*1,8*}	0.9393	0.9416	0.5730	0.6070	0.8948
2{*1,2*}	0.9230	0.9307	0.5200	0.5580	0.8449
8{*2,7*}	0.9101	0.9053	0.4950	0.5480	0.8581
1{*7,10*}	0.9498	0.9547	0.4460	0.6240	0.8457
1{*4,7*}	0.9015	0.9139	0.4950	0.8100	0.8442

[a] This table shows compounds found within the top 10% of the ranked database using PARAFIT Shape Tanimoto and Consensus Shape Tanimoto, ROCS Combo Score and Shape Tanimoto, and HEX Shape Tanimoto scores.

using these different shape matching approaches. The screened compounds selected by these shape-based methods and their score value are shown in Table 10.

Hit Selection. A consensus "Rank-by-Vote"[55] of all the first hit ranking lists compounds found was performed, and five compounds were selected to be synthesized: **1**{*7,8*}, **8**{*2,8*}, **8**{*1,8*}, **10**{*11*}, and **10**{*8*}. Both of the cyclam-amine compounds (**10**) were classified in the top of the ranked list in the virtual screenings, but we selected the two best ranked and the three best classified amino/hydrazono-amines. Compound **8**{*1,8*} was toxic at a concentration of 4.1 μg/mL and showed no activity below this concentration (Table 11). However, compounds **1**{*7,8*} and **8**{*2,8*} showed anti-HIV activity values of 0.6 and 0.4 μg/mL, respectively, and the cyclam-amine compounds **10**{*11*} and **10**{*8*} showed the best anti-HIV activities of 0.058 μg/mL and 0.022 μg/mL, respectively.

Table 11. Summary of the Five VS-Selected Hits[a]

Compound	Name	EC$_{50}$ / μg/ml	CC$_{50}$ / μg/ml
	8{*1,8*}	> 4.1	4.1
	8{*2,8*}	0.6	14.6
	1{*7,8*}	0.4	> 25
	10{*8*}	0.022	> 25
	10{*11*}	0.058	> 25

[a] EC$_{50}$ denotes anti-HIV activity, and CC$_{50}$ is the cytotoxicity value (μg/ml).

DISCUSSION

A combination of ligand-based and receptor-based screening tools was used to select molecules from the virtual combinatorial library. The different approaches used generally select similar molecules at the first percentages of the ranked hit lists. Compounds selected by the various ligand-based virtual screening tools are practically the same, whereas

those selected by the structure-based docking tools also include some others. All shape-based and pharmacophore ligand-based approaches, and consensus scoring of AUTODOCK and GOLD scoring functions, FRED consensus and Chemgauss3, and the HEX Docked Energy approaches select nearly the same molecules at first percentages of database screened. However, although ligand-based searches

give better results than structure-based docking for both retrospective and prospective virtual screening analyses, the pharmacophore models and also AUTODOCK Docked Energy give the best correlation with experimental data. Of the five compounds selected by the Rank-by-Vote consensus, compound **8**{*1,8*} was toxic below 5 μg/mL, but **1**{*7,8*} and **8**{*2,8*} showed activity values below 1 μg/mL, and the remaining two, **10**{*11*} and **10**{*8*}, both of which are monocyclams, showed activity values below 0.06 μg/mL. Our proposed QSAR model agrees well with the experimental results, especially for the nonmonocyclam compounds, with predicted activities of 0.66, 1.58, 7.30, and 0.87 μM for **1**{*7,8*}, **8**{*2,8*}, **10**{*11*}, and **10**{*8*}, which differ by only 0.37, 0.02, 7.17, and 0.82 μM, respectively, from the experimental biological values.

Overall, our screening procedure selects the most active compounds from our combinatorial virtual library (i.e., **1**{*8,8*} 0.008 μg/mL, **1**{*8,9*} 0.03 μg/mL, **1**{*5,6*} 0.2 μg/mL, **1**{*8,10*} 0.4 μg/mL) in the first ranking positions of the final consensus list. Moreover, the first five unsynthesized compounds which were also predicted to be active were ranked in order of their known activities. Hence our screening procedure can be seen to perform rather well.

CONCLUSION

A database of CXCR4 inhibitors and similar presumed inactive compounds was compiled from the literature in order to perform retrospective virtual screening. This database was used to compare docking-based and ligand-based (i.e., pharmacophore modeling and shape matching) virtual screening approaches. Additionally, a large virtual combinatorial library of candidate CXCR4 antagonists was designed, and the above screening approaches were used to select five compounds for synthesis and testing. The actives identified in this way had activities in the range 20 to 0.008 μg/mL. Experimental binding assays of those compounds confirmed that their mode of action was to block the CXCR4 receptor. Activity values were used for the development of ligand-based QSAR models in order to use them to predict activity of hitherto unsynthesised molecules. Prospective virtual screening, using the same protocol as in retrospective screening analysis, was then used to guide the selection of other molecules from the virtual combinatorial library. Molecules found at the first positions of the consensus ranked hit list showed activity values in the range from 4 to 0.022 μg/mL.

ACKNOWLEDGMENT

REFERENCES AND NOTES

(1) UNAIDS. AIDS epidemic update: December 2007. http://www.un-aids.org/en/KnowledgeCentre/HIVData/EpiUpdate/EpiUpdArchive/2007/default.asp (accessed Nov 11, 2008).

(2) De Clercq, E. Emerging anti-HIV drugs. *Expert Opin. Emerging Drugs* **2005**, *10*, 241–274.

(3) De Clercq, E. Anti-HIV chemotherapy: current state of the art. *Med. Chem. Res.* **2004**, *13*, 439–478.

(4) Kadow, J.; Wang, H. G.; Lin, P. F. Small-molecule HIV-1 gp120 inhibitors to prevent HIV-1 entry: an emerging opportunity for drug development. *Curr. Opin. Invest. Drugs (Thomson Sci.)* **2006**, *7*, 721–726.

(5) Berger, E. A.; Murphy, P. M.; Farber, J. M. Chemokine receptors as HIV-1 coreceptors: Roles in viral entry, tropism, and disease. *Annu. Rev. Immunol.* **1999**, *17*, 657–700.

(6) Jiang, S.; Lin, K.; Strick, N.; Neurath, A. R. Inhibition of HIV-1 infection by a fusion domain binding peptide from the HIV-1 envelope glycoprotein GP41. *Biochem. Biophys. Res. Commun.* **1993**, *195*, 533–538.

(7) De Clercq, E. New antiviral agents in preclinical or clinical development. *Adv. Antiviral Drug Des.* **2004**, *4*, 1–62.

(8) De Clercq, E. New Anti-HIV Agents and Targets. *Med. Res. Rev.* **2002**, *22*, 531–565.

(9) Bean, P. New Drugs Targets for HIV. *Clin. Infect. Dis.* **2005**, *41*, 96–100.

(10) Markovic, I.; Clouse, K. A. Recent advances in understanding the molecular mechanisms of HIV-1 entry and fusion: revisiting current targets and considering new options for therapeutic intervention. *Curr. HIV Res.* **2004**, *2*, 223–34.

(11) Kazmierski, W. M.; Peckman, J. P.; Duan, M.; Kenakin, T. P.; Jenkinson, S.; Gudmundsson, K. S.; Piscitelli, S. C.; Feldman, P. L. Recent Progress in the Discovery of New CCR5 and CXCR4 Chemokine Receptor Antagonists as Inhibitors of HIV-1 Entry. Part 2. *Curr. Med. Chem. - Anti Infect. Agents* **2005**, *4*, 133–152.

(12) Pettersson, S.; Pérez-Nueno, V. I.; Ros-Blanco, L.; Puig de la Bellacasa, R.; Rabal, O.; Batllori, X.; Clotet, B.; Clotet-Codina, I.; Armand-Ugón, M.; Esté, J.; Borrell, J. I.; Teixidó, J. Discovery of novel non-cyclam polynitrogenated CXCR4 coreceptor inhibitors. *ChemMedChem* **2008**, *3*, 1549–1557.

(13) *Maybridge Bringing life to drug discovery*; Maybridge Databases Autumn 2005; Fisher Scientific International: England, 2005.

(14) Pascual, R.; Borrell, J. I.; Teixido, J. Analysis of selection methodologies for combinatorial library design. *Mol. Diversity* **2003**, *6*, 121–133.

(15) *MOE (Molecular Operating Environment), 2006.08 Release*; Chemical Computing Group, Inc.: Montreal, Canada, 2004.

(16) *Discovery Studio, version 2.0*; Accelrys Software Inc.: San Diego, 2007.

(17) Lin, J.; Clark, T. An analytical, variable resolution, complete description of static molecules and their intermolecular binding properties. *J. Chem. Inf. Model.* **2005**, *45*, 1010–1016.

(18) Grant, A. J.; Pickup, B. T. A fast method of molecular shape comparison: a simple application of a Gaussian description of molecular shape. *J. Comput. Chem.* **1996**, *17*, 1653–1659.

(19) Ritchie, D. W.; Kemp, G. J. L. Fast computation, rotation, and comparison of low resolution spherical harmonic molecular surfaces. *J. Comput. Chem.* **1999**, *20*, 383–395.

(20) Pérez-Nueno, V. I.; Ritchie, D. W.; Rabal, O.; Pascual, R.; Borrell, J. I.; Teixidó, J. Comparison of Ligand-Based and Receptor-Based Virtual Screening of HIV Entry Inhibitors for the CXCR4 and CCR5 Receptors Using 3D Ligand Shape-matching and Ligand-Receptor Docking. *J. Chem. Inf. Model.* **2008**, *48*, 509–533.

(21) Palczewski, K.; Kumasaka, T.; Hori, T.; Behnke, C. A.; Motoshima, H.; Fox, B. A.; Le Trong, I.; Teller, D. C.; Okada, T.; Stenkamp, R. E.; Yamamoto, M.; Miyano, M. Crystal structure of rhodopsin: A G-protein-coupled receptor. *Science* **2000**, *289*, 739–745.

(22) Morris, G. M.; Goodsell, D. S.; Halliday, R. S.; Hart, W.; Belew, R. K.; Olson, A. J. Automated Docking Using a Lamarckian Genetic Algorithm and Empirical Binding Free Energy Function. *J. Comput. Chem.* **1998**, *19*, 1639–1662.

(23) Verdonk, M. L.; Cole, J. C.; Hartshorn, M. J.; Murray, C. W.; Taylor, R. D. Improved Protein-Ligand Docking Using GOLD. *Proteins: Struct., Funct., Genet.* **2003**, *52*, 609–623.

(24) McGann, M. R.; Almond, H. R.; Nicholls, A.; Grant, J. A.; Brown, F. K. Gaussian docking functions. *Biopolymers* **2003**, *68*, 76–90.

(25) Ritchie, D. W.; Kemp, G. J. L. Protein docking using spherical polar Fourier correlations. *Proteins: Struct., Funct., Genet.* **2000**, *39*, 178–194.

(26) Hatse, S.; Princen, K.; De Clercq, E.; Rosenkilde, M. M.; Schwartzb, T. W.; Hernandez-Abad, P. E.; Skerlj, R. T.; Bridger, G. J.; Schols, D. AMD3465, a monomacrocyclic CXCR4 antagonist and potent HIV entry inhibitor. *Biochem. Pharmacol.* **2005**, *70*, 752–761.

(27) Princen, K.; Hatse, S.; Vermeire, K.; Aquaro, S.; De Clercq, E.; Gerlach, L.-O.; Rosenkilde, M.; Schwartz, T. W.; Skerlj, R.; Bridger, G.; Schols, D. Inhibition of Human Immunodeficiency Virus Replication by a Dual CCR5/CXCR4 Antagonist. *J. Virol.* **2004**, *78*, 12996–13006.

(28) Rosenkilde, M. M.; Gerlach, L.-O.; Hatse, S.; Skerlj, R. L.; Schols, D.; Bridger, G.; Schwartz, T. W. Molecular mechanism of action of monocyclam versus bicyclam non-peptide antagonist in the CXCR4 chemokine receptor. *J. Biol. Chem.* **2007**, *282*, 27354–27365.

(29) Teixidó, J.; Borrell, J. I.; Nonell, S.; Pettersson, S.; Ros, L.; Puig de la Bellacasa, R.; Rabal, M. O.; Pérez-Nueno, V. I.; Esté, J.; Clotet-Codina, I.; Armand-Ugón, M. Nuevos sistemas polinitrogenados como agentes anti-VIH. ES Patent ES200602764, 2006 (filing date: October 26, 2006).

(30) Bridger, G.; Skerlj, R.; Kaller, A.; Harwing, C.; Bogucki, D.; Wilson, T. R.; Crawford, J.; McEachern, E. J.; Atsma, B.; Nan, S.; Zhou, Y. World Patent WO 0022600, 2002.

(31) Bridger, G.; Skerlj, R.; Kaller, A.; Harwing, C.; Bogucki, D.; Wilson, T. R.; Crawford, J.; McEachern, E. J.; Atsma, B.; Nan, S.; Zhou, Y. World Patent WO 0022599, 2002.

(32) Bridger, G.; Skerlj, R.; Kaller, A.; Harwing, C.; Bogucki, D.; Wilson, T. R.; Crawford, J.; McEachern, E. J.; Atsma, B.; Nan, S.; Zhou, Y. World Patent WO 00234745, 2002.

(33) Bridger, G.; Skerlj, R.; Kaller, A.; Harwing, C.; Bogucki, D.; Wilson, T. R.; Crawford, J.; McEachern, E. J.; Atsma, B.; Nan, S.; Zhou, Y. World Patent WO 055876, 2003.

(34) Bridger, G.; Skerlj, R.; Kaller, A.; Harwing, C.; Bogucki, D.; Wilson, T. R.; Crawford, J.; McEachern, E. J.; Atsma, B.; Nan, S.; Zhou, Y.; Smith, C. D.; Di Fluir, R. M. U.S. Patent 0019058, 2004.

(35) Ichiyama, K.; Yokohama-Kumakura, S.; Tanaka, Y.; Tanaka, R.; Hirose, K.; Bannai, K.; Edamatsu, T.; Yanaka, M.; Niitani, Y.; Miyako-Kurosaki, N.; Takaku, H.; Koyanagi, Y.; Yamamoto, N. *Proc. Natl. Acad. Sci. U.S.A.* **2003**, *100*, 4185–4190.

(36) Murakami, T.; Yoshida, A.; Tanaka, R.; Mitsuhashi, S.; Hirose, K.; Yanaka, M.; Yamamato, N.; Tanaka, Y. KRH-2731: An Orally Bioavailable CXCR4 Antagonist Is a Potent Inhibitor of HIV-1 Infection. In *2004 Antivirals Pipeline Report*; Camp, R., Ed.;Proceedings of the 11th Conference on Retroviruses and Opportunistic Infections, San Francisco, CA, Feb 8−11, 2004; Treatment Action Group: San Francisco, CA, 2004; Abstract No. 541.

(37) Yamazaki, T.; Saitou, A.; Ono, M.; Yokohama, S.; Bannai, K.; Hiroswe, K.; Yanaka, M. World Patent WO 029218, 2003.

(38) Yamazaki, T.; Kikumoto, S.; Ono, M.; Saitou, A.; Takahashi, H.; Kumakura, S.; Hirose, K. World Patent WO 024697, 2004.

(39) Bridger, G. J.; Skerlj, R. T.; Padmanabhan, S.; Martellucci, S. A.; Henson, G. H.; Struyf, S.; Witvrouw, M.; Schols, D.; De Clercq, E. Synthesis and Structure-Activity Relationships of Phenylenebis(methylene)-Linked Bis-azamacrocycles That Inhibit HIV-1 and HIV-2 Replication by Antagonism of the Chemokine Receptor CXCR4. *J. Med. Chem.* **1999**, *42*, 3971–3981.

(40) De Clercq, E. Inhibition of HIV Infection by Bicyclams, Highly Potent and Specific CXCR4 Antagonists. *Mol. Pharmacol.* **2000**, *57*, 833–839.

(41) Esté, J. A.; Cabrera, C.; De Clercq, E.; Struyf, S.; Damme, J. V.; Bridger, G.; Skerlj, R. T.; Abrams, M. J.; Henson, G.; Gutierrez, A.; Clotet, B.; Schols, D. Activity of Different Bicyclam Derivatives against Human Immunodeficiency Virus Depends on Their Interaction with the CXCR4 Chemokine Receptor. *Mol. Pharmacol.* **1999**, *55*, 67–73.

(42) Egberink, H. F.; De Clercq, E.; Van Vliet, A. L. W.; Balzarini, J.; Bridger, G. J.; Henson, M. C.; Schols, D. Bicyclams, Selective Antagonists of the Human Chemokine Receptor CXCR4, Potently Inhibit Feline Immunodeficiency Virus Replication. *J. Virol.* **1999**, *73*, 6346–6352.

(43) Tamamura, H.; Araki, T.; Ueda, S.; Wang, Z.; Oishi, S.; Esaka, A.; Trent, J. O.; Nakashima, H.; Yamamoto, N.; Peiper, S. C.; Otaka, A.; Fujii, N. Identification of novel low molecular weight CXCR4 antagonists by structural tuning of cyclic tetrapeptide scaffolds. *J. Med. Chem.* **2005**, *48*, 3280–3289.

(44) Gerlach, L.-O.; Skerlj, R. T.; Bridger, G. J.; Schwartz, T. W. Molecular Interaction of Cyclam and Bicyclam Non-peptide Antagonists with the CXCR4 Chemokine Receptor. *J. Biol. Chem.* **2001**, *276*, 14154–14160.

(45) Tamamura, H.; Ojida, A.; Ogawa, T.; Tsutsumi, H.; Masuno, H.; Nakashima, H.; Yamamoto, N.; Hamachi, I.; Fujii, N. Identification of a new class of low molecular weight antagonists against the chemokine receptor CXCR4 having the dipicolylamine-zinc(II) complex structure. *J. Med. Chem.* **2006**, *49*, 3412–3415.

(46) Labute, P. Flexible Alignment of Small Molecules. Chemical Computing Group, Inc.: Montreal, Canada, 2004. (Available on the Internet at http://www.chemcomp.com/journal/malign.htm (accessed June 2, 2008)).

(47) Agrawal, V. K.; Singh, J.; Gupta, M.; Jaliwala, Y. A.; Khadikar, P. V.; Supuran, C. T. QSAR studies on benzopyran potassium channel activators. *Eur. J. Med. Chem.* **2006**, *41*, 360–366.

(48) Golbraikh, A.; Shen, M.; Xiao, Z.; Xiao, Y.-D.; Lee, K.-H.; Tropsha, A. Rational selection of training and test sets for the development of validated QSAR models. *J. Comput.-Aided Mol. Des.* **2003**, *17*, 241–253.

(49) Golbraikh, A.; Tropsha, A. Beware of q2! J. *Mol. Graphics Modell.* **2002**, *20*, 269–276.

(50) *OMEGA, version 2.1.0*; OpenEye Scientific Software Inc.: Santa Fe, NM., 2006.

(51) Pérez-Nueno, V. I.; Ritchie, D. W.; Borrell, J. I.; Teixidó, J. Clustering and classifying diverse HIV entry inhibitors using a novel consensus shape based virtual screening approach: Further evidence for multiple binding sites within the CCR5 extracellular pocket. *J. Chem. Inf. Model.* **2008**, *48*, 2146–2165.

(52) Hatse, S.; Princes, K.; Vermeire, K.; Gerlach, L.-O.; Rosenkilde, M. M.; Schwartz, T. W.; Bridger, G.; De Clercq, E.; Schols, D. Mutations at the CXCR4 interaction sites for AMD3100 influence anti-CXCR4 antibody binding and HIV-1 entry. *FEBS Lett.* **2003**, *546*, 300–306.

(53) Brelot, A.; Heveker, N.; Montes, M.; Alizon, M. Identification of Residues of CXCR4 Critical for Human Immunodeficiency Virus Coreceptor and Chemokine Receptor Activities. *J. Biol. Chem.* **2000**, *275*, 23736–23744.

(54) Hatse, S.; Princen, K.; Gerlach, L.-O.; Bridger, G.; Henson, G.; Clercq, E.; Schwartz, T. W.; Schols, D. Mutation of Asp171 and Asp262 of the chemokine receptor CXCR4 impairs its coreceptor function for human immunodeficiency virus-1 entry and abrogates the antagonistic activity of AMD3100. *Mol. Pharmacol.* **2001**, *60*, 164–173.

(55) Wang, R.; Wang, S. How does consensus scoring work for virtual library screening? An idealized computer experiment. *J. Chem. Inf. Comput. Sci.* **2001**, *41*, 1422–1426.

(56) Güner, O. F.; Henry, D. R. Metric for analyzing hit lists and pharmacophores. Chapter 11. In *Pharmacophore, perception, development and use in drug design*; Güner, O. F., Ed.; International University Line: La Jolla, CA, 2000; pp 195−212.

(57) Gürner, O. F.; Hoffman, R.; Li, H. Techniques and strategies in 3D data mining. In Report by Wendy A. Warr; 217th ASC National Metting and Exposition, Anaheim, California, March 12−25, 1999; Wendy Warr & Associates: London, 1999; pp 50−53.

(58) Hunter, T. M.; McNae, I. W.; Simpson, D. P.; Smith, A. M.; Moggach, S.; White, F.; Walkinshaw, M. D.; Parsons, S.; Sadler, P. J. Configurations of nickel-cyclam antiviral complexes and protein recognition. *Chem.--Eur. J.* **2007**, *13*, 40–50.

(59) Liang, X.; Parkinson, J. A.; Wishäupl, M.; Gould, R. O.; Paisey, S. J.; Park, H.; Hunter, T. M.; Blindauer, C. A.; Parsons, S.; Sadler, P. J. Structure and dynamics of metallomacrocycles: recognition of zinc xylyxl-bicyclam by an HIV coreceptor. *J. Am. Chem. Soc.* **2002**, *124*, 9105–9112.

(60) Hunter, T. M.; McNae, I. W.; Liang, X.; Bella, J.; Parsons, S.; Walkinshaw, M. D.; Sadler, P. J. Protein recognition of macrocycles: Binding of anti-HIV metallocyclams to lysozyme. *Proc. Natl. Acad. Sci. U.S.A.* **2005**, *102*, 2288–2292.

(61) Rosenkilde, M. M.; Gerlach, L.-O.; Jakobsern, J. S.; Skerlj, R. T.; Bridger, G. J.; Schwartz, T. W. Molecular mechanism of AMD3100 antagonism in the CXCR4 receptor. *J. Biol. Chem.* **2004**, *279*, 3033–3041.

(62) Valks, G. C.; McRobbie, G.; Lewis, E. A.; Hubin, T. J.; Hunter, T. M.; Sadler, P. J.; Pannecouque, C.; De Clercq, E.; Archibald, S. J. Configurationally restricted bismacrocyclic CXCR4 receptor antagonists. *J. Med. Chem.* **2006**, *49*, 6162–6165.