JOURNAL OF CHEMICAL INFORMATION AND MODELING

Article

pubs.acs.org/jcim

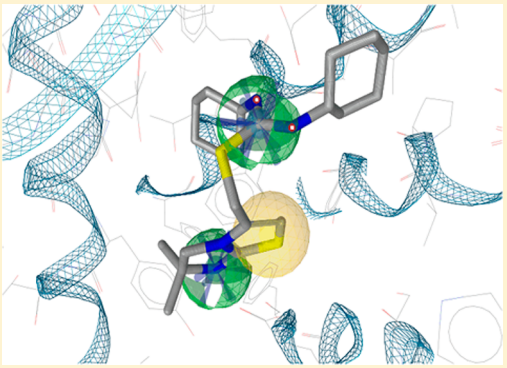
Highly Specific and Sensitive Pharmacophore Model for Identifying CXCR4 Antagonists. Comparison with Docking and Shape-Matching Virtual Screening Performance

Arnaud S. Karaboga,[†,§] Jesús M. Planesas,[‡,§] Florent Petronin,[†] Jordi Teixidó,[‡] Michel Souchet,*[,†] and Violeta I. Pérez-Nueno*[,†,‡]

[†]Harmonic Pharma, Espace Transfert, 615 rue du Jardin Botanique, 54600 Villers lès Nancy, France

[‡]Grup d'Enginyeria Molecular, Institut Químic de Sarrià (IQS), Universitat Ramon Llull, Barcelona, Spain

ABSTRACT: HIV infection is initiated by fusion of the virus with the target cell through binding of the viral gp120 protein with the CD4 cell surface receptor protein and the CXCR4 or CCR5 coreceptors. There is currently considerable interest in developing novel ligands that can modulate the conformations of these coreceptors and, hence, ultimately block virus−cell fusion. Herein, we present a highly specific and sensitive pharmacophore model for identifying CXCR4 antagonists that could potentially serve as HIV entry inhibitors. Its performance was compared with docking and shape-matching virtual screening approaches using 3OE6 CXCR4 crystal structure and high-affinity ligands as query molecules, respectively. The performance of these methods was compared by virtually screening a library assembled by us, consisting of 228 high affinity known CXCR4 inhibitors from 20 different chemotype families and 4696 similar presumed inactive molecules. The area under the ROC plot (AUC), enrichment factors, and diversity of the resulting virtual hit lists was analyzed. Results show that our pharmacophore model achieves the highest VS performance among all the docking and shape-based scoring functions used. Its high selectivity and sensitivity makes our pharmacophore a very good filter for identifying CXCR4 antagonists.



INTRODUCTION

HIV cell infection is a multistage and complex process. It starts with the virus entry[1] into the host membrane. The first stage of the entry process is the virus attachment to the cell membrane through receptor CD4. Second, viral proteins *gp120*[2] and *gp41* bind to the coreceptors CCR5 and CXCR4[3] to facilitate virus anchoring to the host cellular membrane. Finally, the viral envelope undergoes conformational changes that allows the insertion of the *gp41* fusion peptide into the host membrane.

One of the strategies to block the HIV entry process is inhibiting the binding of the viral protein *gp120* to the coreceptors CXCR4 or CCR5. The first interaction between *gp120* and these coreceptors is through their extracellular loops. CXCR4 and CCR5 are both chemokine receptors which belong to the GPCR superfamily[4] and are implicated in a wide range of human diseases.[5] This subfamily of GPCR proteins is responsible for transmitting signals from the extracellular environment as well as the recognition of extracellular ligands. All GPCRs have a common characteristic structure relatively hydrophobic with seven transmembrane domains. The N-terminus is located in the extracellular side and the transmembrane helices are connected by three extracellular loops whereas the C-terminus is located in the intracellular side facing the cytoplasm and three intracellular loops connect the interior domain of the transmembrane helices. The transmembrane helices are expanded in the plasmatic membrane as an anticlockwise bundle of alpha-helices. Therefore extracellular loops fulfill important roles in receptor activation and consequently in ligand binding.[6]

An important achievement was accomplished at the end of 2010 when Wu et al.[7] crystallized five independent CXCR4 structures with resolution of 2.5−3.2 Å. These crystal structures contain a small organic molecule derived from isothiourea[8] (IT1t) in Protein Data Bank structures (PDBs) 3ODU, 3OE6, 3OE8, and 3OE9 and a 16-residue cyclic peptide[9] (CVX15) in PDB 3OE0, as antagonists bound to the CXCR4 structure. As many other chemokine receptors, these CXCR4 structures show different monomer association.[10] PDBs 3OE0 and 3OE6 show monomeric structure. PDBs 3ODU and 3OE9 have been crystallized as homodimers, and PDB 3E8O is a homotrimeric protein. These processes of homo and mainly oligodimerization seem to play an important role in the allosterism process.[11]

In all the PDBs, we can observe that the ligand binding site is located in the second extracellular loop, ECL2, close to the N-terminus. Also, superposition of ligand binding pockets shows similar overlapping volume. The residues that induce common interactions in both ligands are Tyr 116, Asp 187, and Glu 288.

Received: January 17, 2013

   dx.doi.org/10.1021/ci400037y | *J. Chem. Inf. Model.* XXXX, XXX, XXX−XXX

Those residues are common in CXCR4 antagonist binding as it has been widely discussed in previous literature.[12,13] Although the overall orientation differs between ligands, hydrophobic interactions highly contribute in the binding affinity of the majority of CXCR4 antagonists. The residues Ala95, Leu91, Val112, His113, Tyr116, Phe292, Tyr255, and Ile259 play important roles in hydrophobic binding. Moreover, Tyr45 on TM1, Trp94 and Asp97 on TM2, Tyr116 on TM3, Asp171 on TM4, Asp262 on TM6, and His281 and Glu288 on TM7 are critical residues involved in stabilizing interactions with known CXCR4 antagonists.[14−16]

Nevertheless, in the case of the peptide ligand a ligand-induced conformational change in the extracellular side of the helix 5[7] is observed. Hence, the structure-based virtual screening (VS) of small peptides might be more dependent on the CXCR4 structure used for the screening.

Planesas et al.[17] recently compared and evaluated for the first time all CXCR4 structures published by Wu et al. together with their previously published CXCR4 rhodopsin-based homology model from the point of view of their receptor based virtual screening performance. In their work, they showed that the crystal structure 3OE6 retrieved the most accurate results in their docking-based VS study, especially regarding the early recovery (average enrichment factor (EF) of 11.7 at 1% of decoys screened) and the area under the curve (AUC) of the ROC plot (average AUC of 0.84), despite being the lowest resolution structure (3.2 Å). Moreover, scaffold retrieval analysis of PDB 3OE6 confirmed that all chemotype families used in their study were early recovered for all scoring functions. Hence, in the present work, we use PDB 3OE6 as the most suitable crystal structure to carry on the docking benchmarking of our CXCR4 database.

Hence, we used the PDB 3OE6 crystal structure to build a CXCR4 pharmacophore model by incorporating protein−ligand interaction information to define the pharmacophoric features. So far, this is the first CXCR4 pharmacophore published in this way. Here we present the convenience of using this pharmacophore model for identifying new CXCR4 antagonists given its high sensitivity and selectivity, and we compare its performance with classical structure-based and ligand-based methods.

METHODS

CXCR4 Database and Decoy Preparation. We compiled a set of 228 CXCR4 antagonist inhibitors from the literature, which mainly consists of 20 representative families of CXCR4 inhibitors, as listed in Table 1. In the compiled data set, all molecules have activity values of <59 μM against CXCR4. Figure 1 shows some representative members of each family.

To avoid potential bias of the virtual screening results due to large differences in basic properties such as molecular weight, number of rotatable single bonds, number of hydrogen-bond acceptor atoms, number of hydrogen-bond donor atoms, total polar surface area (TPSA), and number of hydrophobic atoms, it is important to work with focused screening libraries.[18] Hence, a set of 4696 presumed inactive compounds was assembled from the Maybridge Screening Collection database[19] in such a way that several physicochemical properties were similar to those of the active compounds. Table 2 shows that the average and standard deviations of these properties calculated by MOE 2009 are quite similar for the active and inactive pools. This decoy compilation follows the observations made by Krüger and Evers[20] who propose the use of a commercial database as an effective approach to represent the chemical space of a decoy set.

Table 1. Families of CXCR4 Inhibitors Compiled in the Database

family	number of compounds	ref
AMD/bicyclam derivatives	55	69−72, 44, 73
amine−amine/amine−aldehyde derivatives	41	74
AMD070 derivatives	29	75, 76
KRH derivatives/ peptidomimetic Kureha	18	77, 78, 79
monocyclam derivatives	17	44, 80−82, 43, 42, 12, 74
phenantroline derivatives	14	83
bis-azamacrocycles	13	84, 69
(DpaZn) dipicolilamine zinc derivatives	10	85
tetrahydroquinoline imidazopyridine derivatives	8	86, 87
macrocycles	4	84, 69, 88
isothiourea derivatives	3	7, 8, 79, 15
pyrimidine derivatives	3	89
chalcone derivatives	3	90
tetrahydroquinoline benzimidazole derivatives	2	91−93
bis-imidazole derivatives	2	94, 79
ethanediamine derivatives	2	79
chemocentrix	1	95, 15
Garland Marshall diazepine	1	79
spiro-containing compounds	1	96, 79
MSX-112	1	97

The 3D structures of all ligands were built using MOE. The structures were protonated at physiological pH (i.e., pH 7), Gasteiger partial charges were assigned, and the geometry was optimized using the MMFF94 force field.

Anti-HIV Data Set Preparation for Analyzing the Sensitivity and Specificity of the Pharmacophore Model. We assembled four data sets of anti-HIV active ligands comprising 928 integrase inhibitors, 2374 protease inhibitors, 2484 reverse transcriptase inhibitors, and 354 CCR5 antagonists, respectively. Ligand structures for the integrase, protease, and reverse transcriptase inhibitor data sets where extracted from Binding DB database[21] by searching for the "target Biomolecule" of interest and allowing Binding DB to retrieve also results linked to ChEMBL and PubChem databases. The CCR5 antagonists' data set was assembled from the work of Pérez-Nueno et al.[22] The 3D structures of all ligands were protonated at physiological pH (i.e., pH 7), Gasteiger partial charges were assigned, and the geometry was optimized using the MMFF94 force field in MOE.

Table 2 shows the average and standard deviations for several 1D physicochemical properties calculated by MOE 2009 for the integrase, protease, reverse transcriptase, and CCR5 databases, compared to the properties of the CXCR4 antagonists and CXCR4 compiled decoys.

These anti-HIV data sets were used to analyze both the sensitivity and specificity of the pharmacophore model. Specifically, sensitivity was measured by the ability of the model to identify the CXCR4 antagonists in the first ranking positions, while specificity was measured by the ability of the model to identify only the CXCR4 antagonists and nothing else. Hence, we performed two experiments; in the first one, we analyze the sensitivity of the pharmacophore model by

dx.doi.org/10.1021/ci400037y | J. Chem. Inf. Model. XXXX, XXX, XXX−XXX

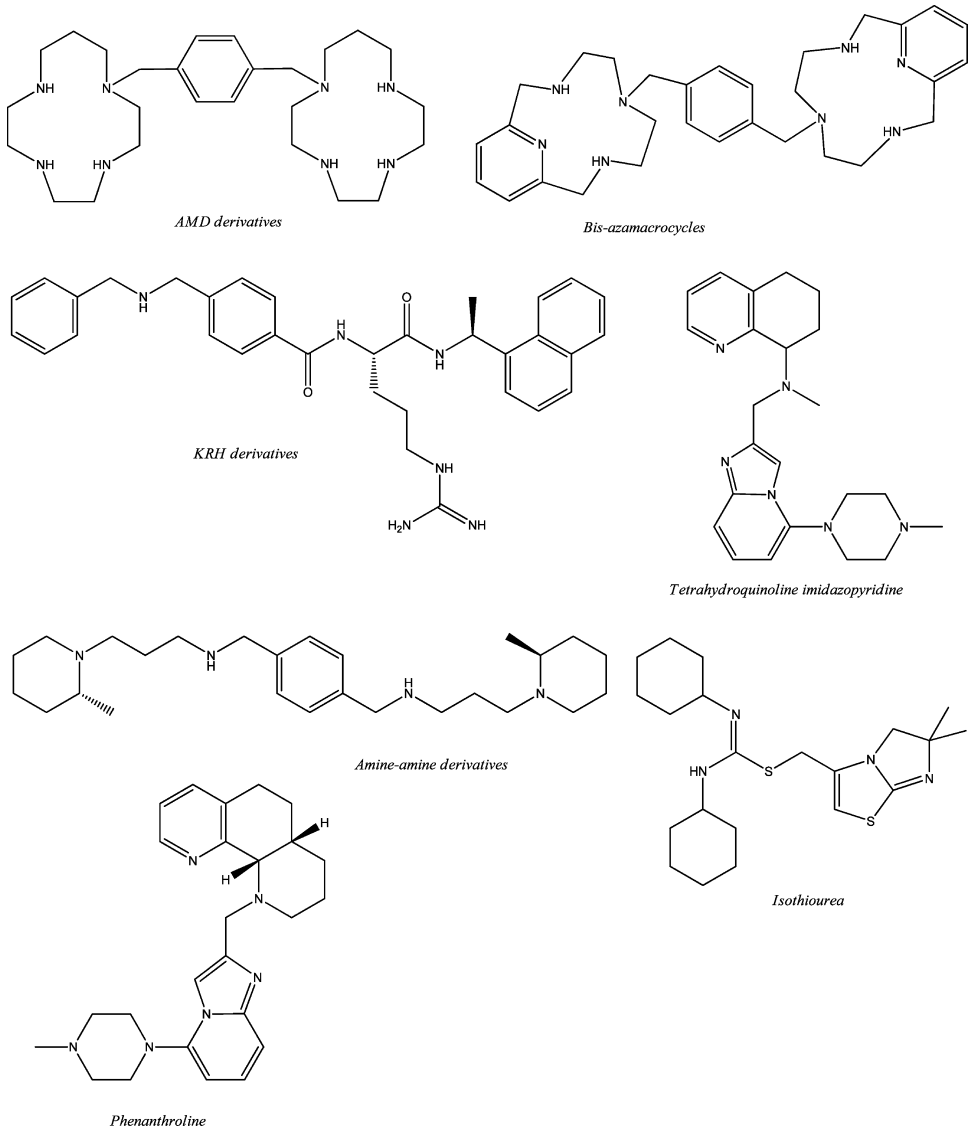


Figure 1. Representative structures of chemotype families in the CXCR4 database.

Table 2. Summary of the 1D Physicochemical Properties of Active and Inactive Molecules in the Screening Database[a]

molecule	molecular weight	number of rotatable single bounds b_1rotN	number of hydrogen-bond acceptor atoms, a_acc	number of hydrogen-bond donor atoms, a_don	number of hydrophobic atoms, ahyd	total polar surface area, TPSA
228 CXCR4 antagonists	455.2 (115.5)	7.0 (3.8)	2.4 (1.4)	0.9 (1.1)	24.9 (6.2)	73.5 (32.2)
4696 decoys	497.4 (45.5)	5.8 (2.2)	3.2 (1.6)	0.9 (0.9)	23.7 (4.1)	83.5 (33.4)
354 CCR5 antagonists	560.9 (86.6)	9.3 (3.6)	3.3 (1.6)	0.8 (0.8)	30.3 (4.7)	74.0 (31.5)
928 HIV integrase inh.	465.7 (299.0)	8.1 (10.8)	4.7 (3.4)	2.5 (3.9)	20.4 (3.9)	138.3 (131.1)
2374 HIV protease inh.	597.3 (115.1)	11.69 (4.1)	4.8 (2.1)	3.0 (1.7)	31.0 (5.9)	123.6 (39.8)
2484 HIV reverse transcriptase inh.	388.1 (217.0)	5.9 (6.9)	3.5 (2.2)	1.3 (1.8)	17.7 (10.2)	81.7 (65.0)

[a]Each column displays respective average number with standard deviation in parentheses.

considering the different anti-HIV compound databases as actives, and the same data set assembled from Maybridge as decoys; in the second one we analyze the specificity of the pharmacophore model by considering the CXCR4 antagonists as actives and all the other anti-HIV compound databases as decoys.

Sensitivity is generally favored over specificity to prevent the loss of active candidates. Conversely, if the pharmacophore is the ultimate decision manner in a VS cascade, a more specific selection threshold may be set. In any case, it is always needed to make a compromise between sensitivity and specificity of a

dx.doi.org/10.1021/ci400037y | J. Chem. Inf. Model. XXXX, XXX, XXX−XXX

model. Improvement of the specificity of a model is always at the cost of losing sensitivity, and vice versa.

Pharmacophore Modeling. A highly specific and sensitive pharmacophore model for identifying CXCR4 antagonists was built using protein−ligand interaction information. We used the automatic pharmacophore generation protocol in LigandScout 3.1 from Inte:Ligand[23] to generate the pharmacophore model directly from the receptor−ligand interactions as revealed in the 3OE6 CXCR4−isothiourea complex. In this first step, seven pharmacophore features of the ligand were identified: (i) two positively charged features located on the two nitrogen atoms of the thiourea group; (ii) two hydrogen-bond donor features respectively centered on the nitrogen of the thiourea and directed toward the backbone carbonyl of Cys286 and on the nitrogen of the aromatic ring to interact with the carboxyl moiety of residue Glu288; (iii) three hydrophobic features positioned on the two gemdimethyl carbons and on the aromatic ring near the thiomethyl linker. In a second step, we manually curate the previous model by keeping and adjusting the features that we considered relevant and mandatory for the protein−ligand interaction, and we removed those considered as unspecific. Eventually, we generated a five points pharmacophore as displayed in Figure 2.

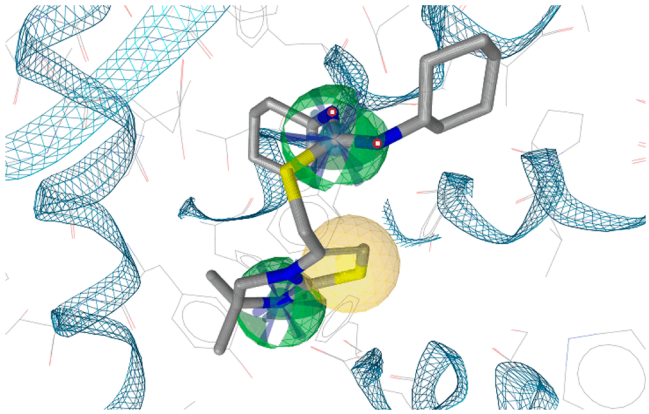


Figure 2. CXCR4 pharmacophore model with a high activity CXCR4 antagonist aligned. Five-featured manually refined final pharmacophore model. The pharmacophore hydrophobic features are shown in yellow. Positively charged features are shown in blue, and hydrogen bond donor features are shown in green.

Our compiled database of actives and decoys was docked onto the CXCR4 3OE6 PDB (see the Docking-Based Virtual Screening section). Given that DockScore scoring function performed the best, the DockScore pose obtained for every active and decoy was used as the starting conformation. These conformations were afterwards converted into LigandScout format using the Idbgen routine implemented in LigandScout 3.1, and 25 conformations per ligand were calculated using the FAST method of OMEGA.[24] The routine successfully converted 211 out of 228 actives and 4695 out of 4696 decoys. The remaining 17 CXCR4 inhibitors until the 228 total actives in the database were not considered in the analysis given that LigandScout default force-filed parameters failed to type their atoms (zinc-containing compounds).

Docking-Based Virtual Screening. In the last years GPCR and also chemokine receptors have begun to be determined successfully by protein crystallography. In the case of the chemokine coreceptor CXCR4, Wu et al. obtained five CXCR4 structures at the end of 2010. According to the work of Planesas et al.,[17] we performed the docking-based virtual screening using 3OE6 CXCR4 crystal structure obtained from the Protein Data Bank (PDB).[25] We used different representative force-filed based, empirical, and knowledge-based scoring functions:[26] GoldScore,[27] ChemScore,[27] and GoldScore rescored with ChemScore from Gold 4.1, Glide,[28] TotalScore provided in Surflex[29] from Sybyl X,[30] AutoDock Vina,[31] and Ligandfit scoring functions (LigScore1-Dreiding; LigScore2-Dreiding;[32] PLP1; PLP2;[33,34] Jain,[34] PMF, and Dock_Score[35]) implemented in Discovery Studio 2.5.

Protein structure 3OE6 is a monomeric protein chain with a crystal resolution of 3.2 Å. It does not contain any crystallized water molecules, and it has a large binding pocket in accordance with the observation that many CXCR4 inhibitors have peptidic or peptidomimetic character. The 3OE6 structure was protonated, histidines were kept in the same protonation state, and chemical molecules used as crystallization aiders were suppressed. This preparation procedure was carried out prior to docking, using the own protein preparation module implemented in each docking tool: the ADT program was used to prepare the corresponding PDBQT-protein file for AutoDock Vina, Protein Preparation Wizard was used for Glide, Protein Setup tool for Gold, Prepare Protein Structure for Sybyl, and Prepare Protein in the case of Discovery Studio. The force fields used to protonate and minimize the protein chain were the following: OPLS-2005,[36] TRIPOS,[37] and CHARMm[38] for Glide, Sybyl X, and Discovery Studio, respectively.

The CXCR4 binding pocket was defined by setting the volume, where the ligands could interact with the protein, delimited by residues Asp171, Asp262, and Glu288, that is in fact the region which comprises the bound ligand IT1t. Several mutagenic studies confirm interactions between Asp171, Asp262, and Glu288 residues with different ligands such as bicyclam AMD3100,[39−41] monocyclams AMD3465 and AMD3529,[42] or non cyclams like AMD070[42−44] in different ways. This triad of acidic residues also interacts with peptidic ligands like T-140, a 14-residue polypeptide[45,46] or cyclopentapeptides antagonists like FC131 and their derivates.[47]

Shape-Matching-Based Virtual Screening. Ligand-based virtual screening was performed using PARASURF/PARAFIT,[48,49] ROCS,[50] and HPCC.[51] Prior to the screening, OMEGA was used to calculate 25 conformations of each active and decoy according to the same parameters than those used to generate the compound conformations filtered by the pharmacophore model. All shape-based methods were assessed according to their ability to superpose each of the database compounds onto the query isothiourea extracted from the 3OE6 crystal structure. For each compound the conformation showing the best overlay with the query molecule was considered and scored.

PARASURF was used to calculate, from semiempirical quantum mechanics theory, the spherical harmonic (SH) molecular shape of all molecules. PARAFIT was then used to superpose every database compound onto the isothiourea query by exploiting the special rotational properties of the SH expansions. These superpositions used the SH Shape Tanimoto as the objective function, which was also used to rank the ligand database.

Superposition of atom-centered Gaussian functions[50] was then performed with ROCS to compute shape-based overlays of all conformers of every compound in the database onto the isothiourea query. Database molecules were then ranked

dx.doi.org/10.1021/ci400037y | J. Chem. Inf. Model. XXXX, XXX, XXX−XXX

according to their "color optimization" mode (Combo Score) to maximize both the shape and chemical property overlays.

HPCC was used to calculate spherical harmonic polynomial functions derived from atomic coordinates for all conformers of every compound in the database to represent their molecular shape as described in the work of Karaboga et al.[52] These harmonic based surfaces were discretized as triangle meshes. A chemotype-like property was then assigned to these triangles by using pharmacophoric features (hydrogen-bond donor, hydrogen-bond acceptor, hydrophobic, anion, cation, aromatic, and their combinations), in a way that the similarity calculation between molecules A and B was carried out by comparing the chemotypelike distribution associated to their respective shapes. HPCC was used in "shape plus chemistry" mode (Combo Score) to maximize both the shape and chemical property overlays.

Analyzing Virtual Screening and Pharmacophore Performance. All the screened compounds were ranked by their retrieved scores according to the pharmacophore, shape-matching, and docking approaches (LigandScout Pharmacophore Fit Score, Parafit Shape Tanimoto, ROCS Combo Score and HPCC Combo Score, and the aforementioned docking scoring functions, respectively), and the final ranking lists were represented in a ROC plot.[53] The area under the ROC plot (AUC)[53,54] was quantified to evaluate the VS performance. As high AUC values do not guarantee that the top ranked poses will correspond to active compounds,[55] we calculated the partial area under the curve[56,57] (pAUC) at 1%, 2%, and 5% of false positives. The corresponding values were normalized to partial area indexes.[58]

$$\mathrm{pAUC}_{x\%} = \frac{1}{2}\left[1 + \frac{\mathrm{AUC}_{x\%} - \mathrm{AUC}_{\text{random at } x\%}}{\mathrm{AUC}_{\text{perfect at } x\%} - \mathrm{AUC}_{\text{random at } x\%}}\right]$$

Furthermore, evaluation of early recovery of actives[59] was assessed by the enrichment factor (EF)[60] at 1%, 2%, and 5% of screened database and, additionally, by the percentage of actives retrieved at 1%, 2%, and 5% of decoys screened.

The comparison of results between the different anti-HIV data sets for analyzing the sensitivity and specificity of the pharmacophore model was performed using the pAUC and also a relative EF, as stated by Venkatraman et al.[61] Because it is not rigorous to compare EF between different data sets, given that EF values are easily influenced by the number of actives in the data set, von Korff et al.[62] advocate the use of a relative EF (EFrel), where the achieved enrichment for a given cutoff (say x%) is normalized by the maximum possible enrichment in that cutoff.

RESULTS AND DISCUSSION

Pharmacophore Performance. The performance of our five-featured CXCR4 pharmacophore model was validated by evaluating its ability to distinguish between the set of 228 known highly active CXCR4 antagonist inhibitors compiled from the literature and the 4696 compiled Maybridge decoy compounds. In an ideal pharmacophore virtual screening exercise, the ideal pharmacophore model should select all the highly active CXCR4 compounds and none of the decoy compounds. That is to say, it should be 100% sensitive (retrieving all highly active ligands) and 100% specific (retrieving zero false positives). Here we analyze the performance of the proposed pharmacophore model according to these parameters.

Our pharmacophore retrieved 298 hits (Figure 3). From those, 114 out of 4696 were decoys, and 184 out of 211 were

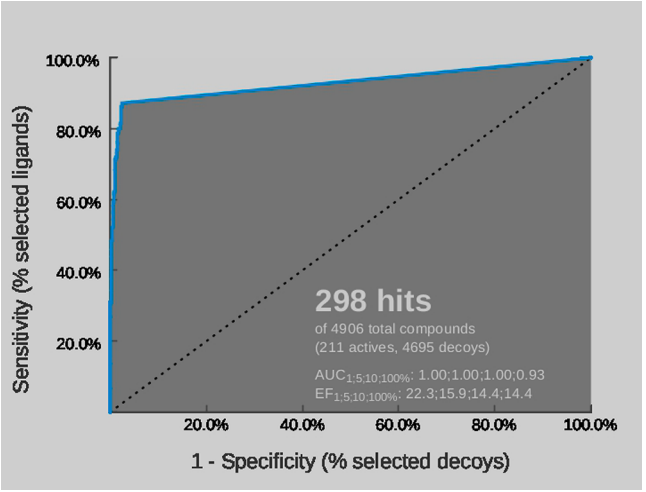


Figure 3. ROC plot validation of the pharmacophore model applied to CXCR4 antagonists. Values of area under the curve (AUC) and enrichment factor (EF) are displayed at 1, 5, 10, and 100% of screened database, respectively. These values highlight the high sensitivity and specificity of the designed pharmacophore model.

CXCR4 active compounds. Hence, the pharmacophore model retrieved the 2.4% of the total decoys and the 83.2% of the total active compounds considered. Therefore, it is very sensitive (retrieves nearly all active ligands) and very specific (retrieves very few false positives), approaching to an ideal pharmacophore. Figure 3 shows the ROC plot validation of the pharmacophore model applied to CXCR4 antagonists, where its high sensitivity and specificity can be appreciated by the early recovery values. At 1%, 2%, and 5% of the screened database, the pharmacophore model retrieves EF values of 22.3, 21.1, and 15.9, respectively, and pAUC values of 0.75, 0.82, and 0.88, respectively. The high specificity is reflected by a total AUC value of 93%. Hence, the designed five-featured pharmacophore model is able to discriminate very well between active and decoys. Nevertheless, it is worth mentioning that the pharmacophore model does not comprise any excluded volume features. Our analyses show that some ligand conformations matching the pharmacophore properly have an internal energy quite high which could lead to putative internal clashes, mainly for the case of big ligands for which a part of the ligand match well the five pharmacophoric features but the rest of the ligand could clash with the protein. Hence, adding excluded volumes to complete the protein−ligand context in the model might be a way to improve further the specificity of the present pharmacophore.

Analyzing the Sensitivity and Specificity of the Pharmacophore Model. The sensitivity and specificity of the pharmacophore model was analyzed by performing two types of experiments. The sensitivity was analyzed by considering the different anti-HIV compound databases as actives, and the same data set assembled from Maybridge as decoys. Results for this analysis are shown in Table 3a. The specificity was analyzed by considering our compiled database of CXCR4 antagonists as actives and all the other anti-HIV compound databases as decoys. Results for this analysis are shown in Table 3b.

dx.doi.org/10.1021/ci400037y | J. Chem. Inf. Model. XXXX, XXX, XXX−XXX

Table 3a. Comparison Values for EF_{rel} and pAUC Obtained When the Pharmacophore Is Applied to Different HIV Inhibitors Being Considered As Actives and the Same Data Set Assembled from Maybridge as Decoys

		CXCR4 antagonists + Maybridge decoys	CCR5 antagonists + Maybridge decoys	HIV integrase inh. + Maybridge decoys	HIV protease inh. + Maybridge decoys	HIV reverse transcriptase inh. + Maybridge decoys
EF_{rel}	1%	**95.9**	60.0	58.9	87.1	35.2
	2%	**90.8**	56.0	50.0	83.7	30.8
	5%	***68.6***	-[a]	-[a]	72.2	-[a]
pAUC	1%	***0.75***	0.55	0.51	0.52	0.50
	2%	***0.82***	0.58	0.52	0.53	0.50
	5%	***0.88***	-[a]	-[a]	0.54	-[a]

[a]Absence of hits retrieved by the pharmacophore at the percentage referred. The pharmacophore ROC curve does not arrive until 100% of screened database. LigandScout, as it is the case for all pharmacophore software, does not retrieve values for compounds not matching the pharmacophore over a defined threshold.

Table 3b. Comparison Values for the EF_{rel} and pAUC Obtained When the Pharmacophore Is Applied to Different HIV Inhibitors Being Considered As Decoys and the CXCR4 Antagonists As Actives

		CXCR4 antagonists + CCR5 antagonists	CXCR4 antagonists + HIV integrase inh.	CXCR4 antagonists + HIV protease inh.	CXCR4 antagonists + HIV reverse transcriptase inh.
	1%	100.0	100.0	96.0	100.0
EF_{rel}	2%	100.0	100.0	92.2	100.0
	5%	-[a]	-[a]	68.0	-[a]
	1%	0.55	0.69	0.66	0.76
pAUC	2%	0.58	0.69	0.67	0.83
	5%	-[a]	-[a]	0.70	-[a]

[a]Absence of hits retrieved by the pharmacophore at the percentage referred. The pharmacophore ROC curve does not arrive until 100% of screened database. LigandScout, as it is the case for all pharmacophore software, does not retrieve values for compounds not matching the pharmacophore over a defined threshold.

For the first kind of analysis, we can see in Table 3a that the pharmacophore model is more sensitive to the CXCR4 antagonists. The pharmacophore is able to identify the active CXCR4 ligands with the highest EF_{rel} and pAUC values (EF_{rel} of 95.9, 90.8, and 68.6, and pAUC of 0.75, 0.82, and 0.88 at 1%, 2%, and 5% of database screened, respectively), whereas all the other HIV inhibitors are retrieved with lower early recovery values (EF_{rel} of 60.0, 58.9, 87.1, and 35.2 and pAUC of 0.55, 0.51, 0.52, and 0.50 at 1% of screened database for CCR5 antagonists, HIV integrarse inhibitors, HIV protease inhibitors, and HIV reverse transcriptase inhibitors, respectively).

For the second kind of analysis, we can see in Table 3b that the pharmacophore is able to identify only CXCR4 antagonists with very high EF_{rel} and pAUC values and nothing else when using as decoys all the other anti-HIV compound databases. The pharmacophore is able to identify the active CXCR4 ligands in front of the other anti-HIV compounds with very high EF_{rel} and pAUC values (EF_{rel} of 100.0, 100.0, 96.0, and 100.0 at 1% of screened database and EF_{rel} of 100.0, 100.0, 92.2, and 100.0 at 2% of screened database, for CCR5 antagonists, HIV integrarse inhibitors, HIV protease inhibitors, and HIV reverse transcriptase inhibitors, respectively).

Overall, in both experiments the highest EF_{rel} and pAUC values are retrieved when screening the CXCR4 antagonists, only slightly approached by the values when screening HIV protease inhibitors. The pharmacophore model is able to discriminate well similar compounds, regarding the average of several 1D physicochemical properties, such as CXCR4 and CCR5 antagonists (see Table 2). Hence, the superior results for CXCR4 antagonists prove our pharmacophore model to be a filter of great sensitivity and specificity for identifying CXCR4 antagonists.

Comparison with Docking-Based and Shape-Matching Virtual Screening Performance. The CXCR4 compiled data set of actives and decoys was docked against 3OE6 CXCR4 PDB using the aforementioned docking functions. Moreover the isothiourea ligand in 3OE6 crystal structure was used to shape query our database using PARAFIT, ROCS, and HPCC. Figure 4 shows the ROC plot comparing the performance of all of these approaches together with the pharmacophore values obtained. Tables 4a and 4b show the early recovery performance for all approaches in terms of EF at 1%, 2%, and 5% of the screened database, and the percentage of true positives selected at 1%, 2%, and 5% of false positives, respectively. Tables 5a and 5b show the AUC and the pAUC at 1%, 2%, and 5% of decoys retrieved for all the approaches.

As it can be observed in Figure 4, the pharmacophore ROC curve (in magenta) does not arrive until 100% of screened database given that LigandScout, as it is the case for all pharmacophore software, does not retrieve values for compounds not matching the pharmacophore over a defined threshold. Hence, in order to better compare the pharmacophore model with the rest of scoring functions, we deeply analyze early recovery performance until 5% of total database screened.

In general, ROC plot analysis shows that the pharmacophore model obtains the highest AUC, and early recovery performance, expressed as EF, percentage of actives recovered at a percentage of decoys screened, and pAUC (values shown in bold in Tables 5a, 4a, 4b, and 5b, respectively), followed by LigandFit docking scoring functions, the rest of docking scoring functions, and shape-matching scoring functions. It is worth noting that the AUC value for the pharmacophore model is an estimated value (0.93), since its curve does not arrive until 100% of screened database, as mentioned before.

Regarding the EF (Table 4a), it can be observed that apart from the pharmacophore, the LigandFit docking scoring functions DockScore, LigScore2, PLP1, and PLP2 are also selectively choosing ligands from decoys from the compiled data set. At 1% of total database screened, EF values achieved for these LigandFit scoring functions were 22.6, 17.9, 15.5, and 15.1, respectively, and 20.2, 17.2, 14.1, and 13.4, at 2% of screened database, compared with the EF values for the pharmacophore model of 22.3 and 21.1 at 1% and 2% of database screened. Other scoring functions like PMF, LigScore1, Jain, Glide, AutoDock Vina, and Surflex returned lower

dx.doi.org/10.1021/ci400037y | J. Chem. Inf. Model. XXXX, XXX, XXX−XXX

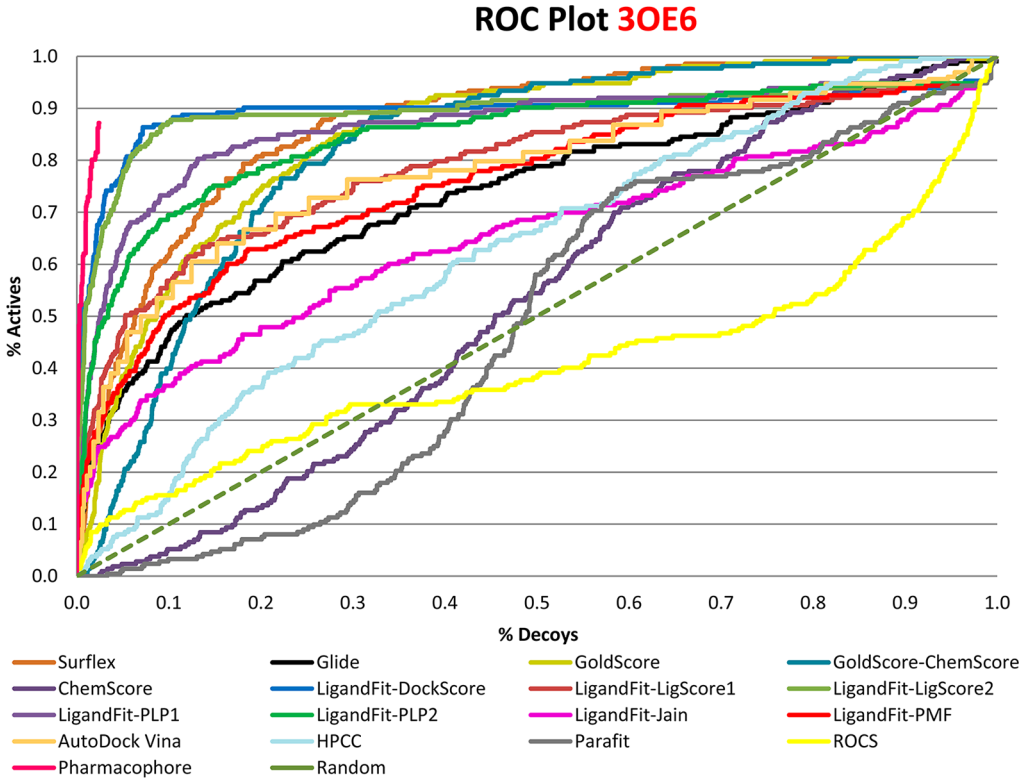


Figure 4. ROC plot validation of the pharmacophore model compared to various docking and shape-matching VS methods applied to CXCR4 antagonists. The dotted green line represents the expected enrichment if actives were selected at random. The pharmacophore ROC plot (curve in magenta) returned only 2.4% of the total decoys set.

but also quite selective EF values at the 1% cutoff (13.6, 13.2, 11.3, 10.4, 9.3, and 8.5, respectively) and also at the 2% cutoff (10.8, 11.8, 8.7, 9.6, 9.7%, and 8.9, respectively). The same trend can be observed for the percentage of actives at a percentage of decoys screened (Table 4b). The pharmacophore model obtains values of 71.1% and 80.6% of actives recovered at 1% and 2% of decoys screened, followed by the LigandFit docking scoring functions DockScore, LigScore2, PLP2, and PLP1, with values of 52.6% and 62.0%, 52.1% and 59.6%, 33.3% and 45.0%, and 31.9% and 45.5% of actives recovered at 1% and 2% of decoys screened, respectively.

Comparing the AUC (Table 5a) and pAUC values at 1%, 2%, and 5% of screened decoys (Table 5b), for all the approaches, it can be observed that the pharmacophore model obtains the highest AUC (0.93) and pAUC values (0.75 at 1%, 0.82 at 2%, and 0.88 at 5% of decoys screened, respectively) with high difference among all scoring functions, only comparable to LigandFit DockScore, with an AUC value of 0.88 and pAUC values of 0.72 at 1%, 0.75 at 2%, and 0.81 at 5% of decoys screened, respectively. LigScore2, PLP1, PLP2, PMF, LigScore1, Jain, Glide, AutoDock Vina, and Surflex show moderately good selectivity (Table 5b). From these scoring functions, LigScore2 achieved the highest pAUC values (0.65 at 1%, 0.71 at 2%, and 0.79 at 5% of screened decoys), followed by PLP1 (values of 0.59%, 0.64%, and 0.74%), PLP2 (values of 0.58%, 0.64%, and 0.70%), PMF (values of 0.57%, 0.59%, and 0.63%), LigScore1 (values of 0.56%, 0.60%, and 0.65%), Jain (values of 0.55%, 0.57%, and 0.60%), Glide (values of 0.54%, 0.57%, and 0.61%), AutoDock Vina (values of 0.54%, 0.57%, and 0.63%), and Surflex TotalScore (values of 0.53%, 0.56%, and 0.62%).

As mentioned above, the ROC plot analysis shows that shape-matching scoring functions perform the poorest in our study. Several papers have dealt with the question if docking performs better or not than shape-matching.[63,64] A known fact is that the performance of a method is highly dependent on the nature of the target of interest and the data set used to screen, which must be assembled properly as mentioned before. In our case study, several papers have shown that the CXCR4 target has multiple interaction subsites within its pocket,[65,12,14−17] and this can be easily understood when noticing the wide range of different active scaffold families existing for CXCR4 (Table 1). Hence, it is easy to think that a single shape query would not retrieve good results when performing shape-matching VS on a diverse set of CXCR4 inhibitors, which can bind in different ways (binding subsites) and hence have different active shapes. Therefore, we must take this into account when analyzing our results, which show shape-matching scoring functions performing the poorest (AUC values for shape-matching scoring functions approximately random), since we used the isothiourea ligand in the 3OE6 structure as a unique representative shape-matching query. Recent studies have shown that using a consensus query, which can average the essential features of several known high-affinity scaffold families to take into account multiple binding subsite shapes, can highly improve shape-matching VS performance.[13,17,66−68]

Scaffold Retrieval Analysis. The CXCR4 database that we have built contains 228 actives from twenty chemotype families (see Table 1). To evaluate the ability of the different approaches used to retrieve these diverse scaffold families, we analyzed the percentage of actives returned by the scoring functions attending to its intrinsic chemotype classification at 1%, 2%, and 5% of screened database. Figure 5 shows the

dx.doi.org/10.1021/ci400037y | J. Chem. Inf. Model. XXXX, XXX, XXX−XXX

Table 4a. Early Recovery Performance Expressed by the Enrichment Factor at 1%, 2%, and 5% of the Screened Database[a]

	% screened database	enrichment factors
	1%	5.6
GoldScore	2%	5.2
	5%	6.3
	1%	0.0
ChemScore	2%	0.0
	5%	0.5
	1%	0.5
GoldScore and	2%	1.9
ChemScore	5%	3.3
	1%	10.4
Glide	2%	9.6
	5%	6.3
	1%	8.5
Surflex	2%	8.9
TotalScore	5%	7.3
	1%	**22.6**
LigandFit	2%	20.2
DockScore	5%	13.4
	1%	13.2
LigandFit	2%	11.8
LigScore1	5%	8.1
	1%	17.9
LigandFit	2%	17.2
LigScore2	5%	12.6
	1%	15.5
LigandFit PLP1	2%	14.1
	5%	10.2
	1%	15.1
LigandFit PLP2	2%	13.4
	5%	9.9
	1%	11.3
LigandFit Jain	2%	8.7
	5%	5.4
	1%	13.6
LigandFit PMF	2%	10.8
	5%	6.7
	1%	9.3
AutoDock Vina	2%	9.7
	5%	7.3
	1%	0.0
Parafit	2%	0.0
	5%	0.3
	1%	4.3
ROCS	2%	4.3
	5%	2.4
	1%	1.4
HPCC	2%	1.9
	5%	1.6
	1%	**22.3**
Pharmacophore	2%	**21.1**
	5%	**15.9**

[a]Values in bold represent the best enrichment factor values obtained when comparing all scoring functions.

Table 4b. Early Recovery Performance Expressed As Percentage of Actives Recovered When 1%, 2%, and 5% of Decoys Were Respectively Selected[a]

	% false positive screened	% true positives
	1%	7.0%
GoldScore	2%	15.0%
	5%	38.5%
	1%	0.0%
ChemScore	2%	0.0%
	5%	2.3%
	1%	0.5%
GoldScore and	2%	3.8%
ChemScore	5%	18.3%
	1%	16.4%
Glide	2%	23.9%
	5%	35.7%
	1%	14.1%
Surflex	2%	23.5%
TotalScore	5%	43.7%
	1%	52.6%
LigandFit	2%	62.0%
DockScore	5%	77.9%
	1%	23.0%
LigandFit	2%	32.9%
LigScore1	5%	45.2%
	1%	52.1%
LigandFit	2%	59.6%
LigScore2	5%	77.0%
	1%	31.9%
LigandFit PLP1	2%	45.5%
	5%	64.3%
	1%	33.3%
LigandFit PLP2	2%	45.0%
	5%	57.3%
	1%	15.5%
LigandFit Jain	2%	22.1%
	5%	28.6%
	1%	19.7%
LigandFit PMF	2%	27.7%
	5%	37.1%
	1%	16.7%
AutoDock Vina	2%	25.9%
	5%	41.2%
	1%	0.0%
Parafit	2%	0.0%
	5%	1.3%
	1%	5.7%
ROCS	2%	8.5%
	5%	12.3%
	1%	2.0%
HPCC	2%	3.8%
	5%	8.0%
	1%	**71.1%**
Pharmacophore	2%	**80.6%**
	5%	*b*

[a]Values in bold represent the best percentages obtained when comparing all scoring functions. [b]Only 2.4% of total decoys are retrieved by the pharmacophore.

scaffold retrieval rates for the seven most representative chemotype families and the rest of chemotypes are grouped into a generic family named "others". Therefore, the AMD family includes AMD derivates, AMD070 derivatives, and monocyclam derivates (see Table 1). Macrocycle family includes also bis-azamacrocycles, and tetrahydroquinoline

dx.doi.org/10.1021/ci400037y | J. Chem. Inf. Model. XXXX, XXX, XXX−XXX

Table 5a. AUC Values Calculated from the Corresponding ROC Curves for All the Scoring Functions[a]

	AUC
GoldScore	0.85
ChemScore	0.53
GoldScore and ChemScore	0.81
Glide	0.74
Surflex TotalScore	0.87
LigandFit DockScore	0.88
LigandFit LigScore1	0.78
LigandFit LigScore2	0.88
LigandFit PLP1	0.86
LigandFit PLP2	0.84
LigandFit Jain	0.64
LigandFit PMF	0.75
AutoDock Vina	0.75
Parafit	0.44
ROCS	0.41
HPCC	0.63
Pharmacophore	**0.93**[b]

[a]Values in bold represent the best AUC values obtained when comparing all scoring functions. [b]AUC value is estimated based on a random classification of the actives and decoys not retrieved by the pharmacophore. The pharmacophore ROC curve does not arrive until 100% of screened database given that only values for 2.4% of the decoys set (6.1% of the total database) are retrieved.

family contains tetrahydroquinoline imidazopyridine and tetrahydroquinoline benzimidazole compounds.

Figure 5 shows that at 5% of screened database all the CXCR4 scaffold families were recovered only by the pharmacophore and DockScore function, with an average percentage recovery of 80% and 52%, respectively (average understood as the total percentage of compounds recovered at 5% of screened database divided by 8 scaffold families). At this percentage the pharmacophore retrieves 100% of the macrocycle derivatives and 100% of the tetrahydroquinoline compounds present in the database, while the DockScore scoring function retrieves only 100% of the macrocycle derivatives. Comparing the compounds retrieved for the same scoring function now at 2% of screened database, the unique families with no compounds retrieved were phenanthroline, tetrahydroquinoline, and the generic family "others" for the DockScore function, while the pharmacophore retrieved 30.8% of phenanthroline compounds, 30.0% of tetrahydroquinolines, and 28.6% of "others". Otherwise the pharmacophore model did not retrieve at 2% any isothiourea compound neither KRH derivatives, but DockScore retrieved 33.3% of the isothiourea chemofamily and 21.1% of the KRH structures. It is worth noting that in Figure 5 the isothiourea family is the shortest chemotype group containing only three structures.

Consequently we can affirm that the pharmacophore model selects actives independently from their chemotype nature in a very effective way, as well as the DockScore function, but also we observe that both scoring functions show indeed, a moderately different and complementary mode of retrieving the CXCR4 chemotype families at early percentages of database screened. This observation evidence the differences in the nature of both methods ranking the same binders.

Table 5b. Partial Area Indexes Calculated from the AUC Values at 1%, 2%, and 5% of Decoys Retrieved for All the Scoring Functions[a]

	% false positive screened	pAUC
GoldScore	1%	0.51
	2%	0.53
	5%	0.59
ChemScore	1%	0.50
	2%	0.50
	5%	0.49
GoldScore and ChemScore	1%	0.50
	2%	0.50
	5%	0.52
Glide	1%	0.54
	2%	0.57
	5%	0.61
Surflex TotalScore	1%	0.53
	2%	0.56
	5%	0.62
LigandFit DockScore	1%	0.72
	2%	0.75
	5%	0.81
LigandFit LigScore1	1%	0.56
	2%	0.60
	5%	0.65
LigandFit LigScore2	1%	0.65
	2%	0.71
	5%	0.79
LigandFit PLP1	1%	0.59
	2%	0.64
	5%	0.74
LigandFit PLP2	1%	0.58
	2%	0.64
	5%	0.70
LigandFit Jain	1%	0.55
	2%	0.57
	5%	0.60
LigandFit PMF	1%	0.57
	2%	0.59
	5%	0.63
AutoDock Vina	1%	0.54
	2%	0.57
	5%	0.63
Parafit	1%	0.50
	2%	0.50
	5%	0.49
ROCS	1%	0.51
	2%	0.52
	5%	0.53
HPCC	1%	0.50
	2%	0.51
	5%	0.51
Pharmacophore	1%	**0.75**
	2%	**0.82**
	5%	**0.88**

[a]Values in bold represent the best partial area indexes obtained when comparing all the scoring functions.

CONCLUSION

A large database of highly active CXCR4 antagonist inhibitors and similar presumed inactive compounds was compiled from the literature to perform retrospective virtual screening. The set

dx.doi.org/10.1021/ci400037y | J. Chem. Inf. Model. XXXX, XXX, XXX−XXX

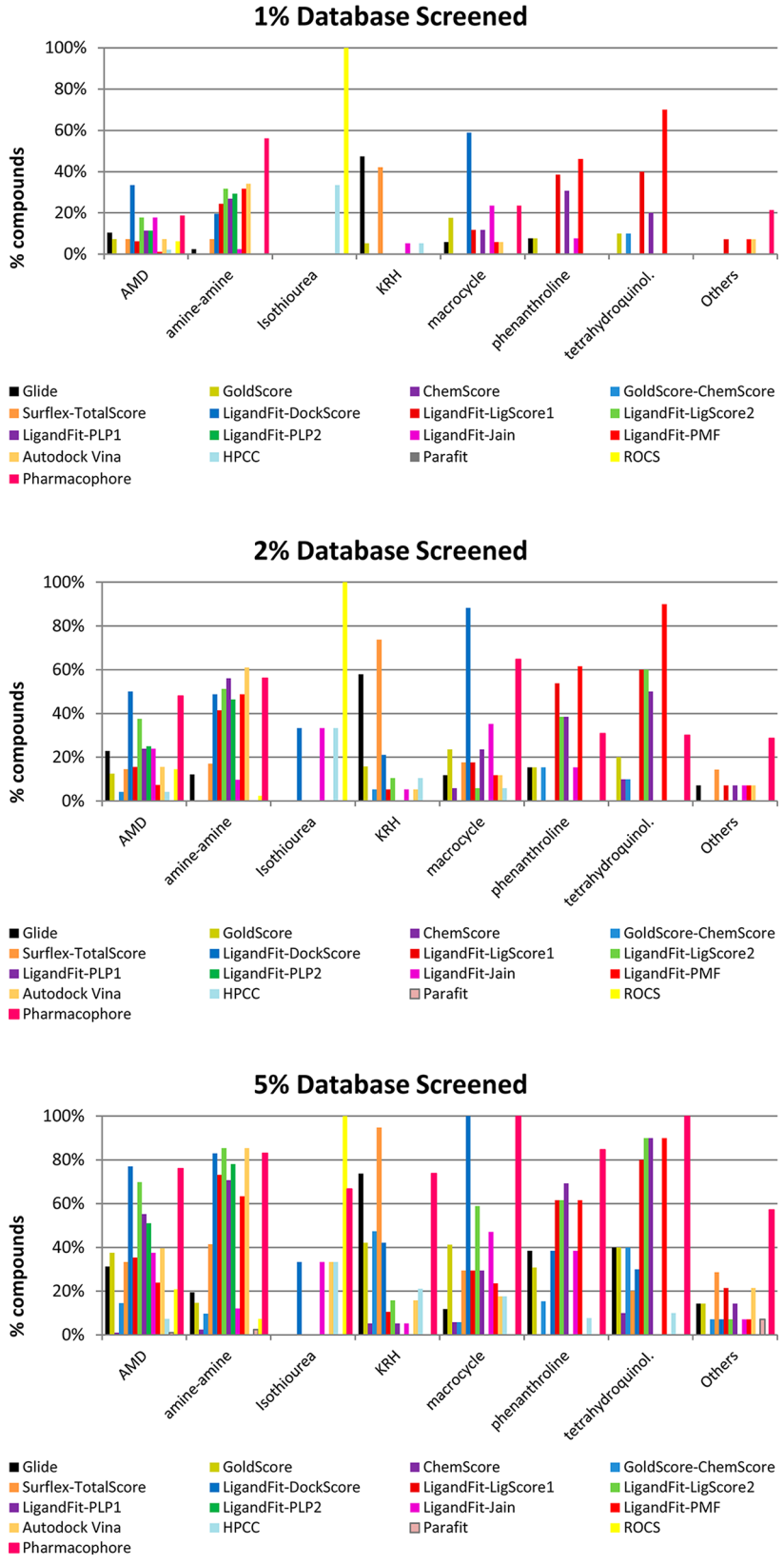


Figure 5. Scaffold diversity retrieval analysis for CXCR4 inhibitors, showing the percentage of actives by chemotype family retrieved using the pharmacophore, docking, and shape-matching scoring functions (colored bars), at 1%, 2%, and 5% of the screened database, respectively.

of active compounds compiled shows a wide scaffold diversity comprising more than 20 different chemotype families. Similarly, the set of decoys selected from the commercial vendor Maybridge possesses the drug-like features required to be similar to the active set to avoid bias in the screening results. This database was used to validate a new CXCR4 pharmacophore model built by using protein−ligand interaction information. So far, this is the first CXCR4

dx.doi.org/10.1021/ci400037y | J. Chem. Inf. Model. XXXX, XXX, XXX−XXX

pharmacophore published in this way. Furthermore, the compiled database was used to perform a benchmark comparison between the pharmacophore model and several docking-based and ligand shape-based virtual screening approaches. The area under de curve of the ROC plot, the early recovery, and the diversity results obtained show that the present pharmacophore model is highly superior to the docking-based and ligand shape-based approaches used in this study, especially the ligand shape-based approaches which were applied using only one shape-query to filter the compiled database. Consequently, the performance of the ligand shape-based approaches appeared to be the poorest. Analysis of these results, and according to other experiments previously carried on CXCR4 by other teams, shows that CXCR4 is a protein target with multiple interaction subsites within the binding pocket. As a matter of fact, the known antagonists distribute over the pocket in at least three different binding modes and that implies that any ligand shape-based screening strategy should be carried out using a multishape query (which takes into account the shape diversity of the several existent CXCR4 scaffold families), in order to be at the performance level of the docking-based approaches. The docking-based approaches and the pharmacophore model already cover in some way the CXCR4 multisite binding information. The docking-based protocols energetically explore the whole pocket region when computing the docking energy of the database compounds. The pharmacophore model takes into account the protein−ligand interaction information to define the essential chemical features needed for the active CXCR4 ligands to bind. These features have wide tolerance radii, therefore the pharmacophore model is also in certain extent representative of the binding of the different scaffold families.

Overall, the total area under de curve of the ROC plot and the early recovery results of the present pharmacophore model show that it is a highly specific and sensitive screening filter, which makes it very appropriate for identifying CXCR4 antagonists. Moreover, the scaffold retrieval analysis shows that the pharmacophore model is able to retrieve a diverse scaffold pool. Hence, using this pharmacophore model to prospectively screen a database of compounds might facilitate the identification of novel lead structures. Therefore, we state that the use of protein−ligand interaction information to define pharmacophoric features and a robust and divers compiled database of very active compounds are essential to built a good-performance model, especially when dealing with targets with multiple interaction sites, as in the case of CXCR4.

AUTHOR INFORMATION

Corresponding Author

*Tel: +33-354 958 604. Fax: +33-383 593 046. E-mail: souchet@harmonicpharma.com (M.S.), pereznueno@harmonicpharma.com (V.I.P.-N.).

Author Contributions

§A.S.K. and J.M.P.: These authors contributed equally to this work.

Notes

The authors declare no competing financial interest.

ACKNOWLEDGMENTS

The authors thank to the Servei de Disseny de Farmacs (SDF) in the Centre de Supercomputacio de Catalunya (CESCA) for providing access to some of the computational software used through. The authors thank OpenEye Scientific Software Inc. for providing a license for ROCS. The authors are also grateful to ChemAxon for the license for the Marvin and Jchem toolkits.

REFERENCES

(1) Tilton, J. C.; Doms, R. W. Entry inhibitors in the treatment of HIV-1 infection. *Antiviral Res.* **2010**, *1*, 91−100.

(2) Trushin, S. A.; Bren, G. D.; Badley, A. D. CXCR4 tropic HIV-1 gp120 inhibition of SDF-1α-induced chemotaxis requires Lck and is associated with cofilin phosphorylation. *Open Virol. J.* **2010**, 157.

(3) Harrigan, P. R.; Swenson, L.; McGovern, R. A.; Pollock, G. J. The Determinants and Consequences of HIV Coreceptor Switching. *HIV Med. Update* **2009**, *4*, 1−4.

(4) Haga, T.; Takeda, S. *G protein-coupled receptors: structure, function, and ligand screening*; CRC Press: Boca Raton, 2005; Vol. *6*.

(5) Choi, W. T.; An, J. Biology and clinical relevance of chemokines and chemokine receptors CXCR4 and CCR5 in human diseases. *Exp. Biol. Med* **2011**, *6*, 637−647.

(6) Peeters, M.; van Westen, G.; Li, Q.; IJzerman, A. Importance of the extracellular loops in G protein-coupled receptors for ligand recognition and receptor activation. *Trends Pharmacol. Sci.* **2011**, *1*, 35−42.

(7) Wu, B.; Chien, E. Y. T.; Mol, C. D.; Fenalti, G.; Liu, W.; Katritch, V.; Abagyan, R.; Brooun, A.; Wells, P.; Bi, F. C. Structures of the CXCR4 chemokine GPCR with small-molecule and cyclic peptide antagonists. *Science* **2010**, *6007*, 1066−1071.

(8) Thoma, G.; Streiff, M. B.; Kovarik, J.; Glickman, F.; Wagner, T.; Beerli, C.; Zerwes, H. G. Orally bioavailable isothioureas block function of the chemokine receptor CXCR4 in vitro and in vivo. *J. Med. Chem.* **2008**, *24*, 7915−7920.

(9) Moncunill, G.; Armand-Ugon, M.; Clotet-Codina, I.; Pauls, E.; Ballana, E.; Llano, A.; Romagnoli, B.; Vrijbloed, J. W.; Gombert, F. O.; Clotet, B. Anti-HIV activity and resistance profile of the CXC chemokine receptor 4 antagonist POL3026. *Mol. Pharmacol.* **2008**, *4*, 1264.

(10) Wang, J.; Norcross, M. Dimerization of chemokine receptors in living cells: key to receptor function and novel targets for therapy. *Drug Discov Today* **2008**, *13−14*, 625−632.

(11) Sohy, D.; Parmentier, M.; Springael, J. Y. Allosteric transinhibition by specific antagonists in CCR2/CXCR4 heterodimers. *J. Biol. Chem.* **2007**, *41*, 30062.

(12) Pettersson, S.; Pérez-Nueno, V. I.; Mena, M. P.; Clotet, B.; Esté, J. A.; Borrell, J. I.; Teixidó, J. Novel monocyclam derivatives as HIV entry inhibitors: design, synthesis, anti-HIV evaluation, and their interaction with the CXCR4 co-receptor. *ChemMedChem* **2010**, *8*, 1272−1281.

(13) Perez-Nueno, V. I.; Ritchie, D. W.; Borrell, J. I.; Teixido, 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**, *11*, 2146−2165.

(14) Kawatkar, S. P.; Yan, M.; Gevariya, H.; Lim, M. Y.; Eisold, S.; Zhu, X.; Huang, Z.; An, J. Computational analysis of the structural mechanism of inhibition of chemokine receptor CXCR4 by small molecule antagonists. *Exp. Biol. Med.* **2011**, *7*, 844−850.

(15) Neves, M. A. C.; Simões, S.; Sá e Melo, M. L. Ligand-guided optimization of CXCR4 homology models for virtual screening using a multiple chemotype approach. *J. Comput. Aided Mol. Des.* **2010**, *12*, 1023−1033.

(16) Lam, A. R.; Bhattacharya, S.; Patel, K.; Hall, S. E.; Mao, A.; Vaidehi, N. Importance of receptor flexibility in binding of cyclam compounds to the chemokine receptor CXCR4. *J. Chem. Inf. Model.* **2010**, *1*, 137−147.

(17) Planesas, J. M.; Pérez-Nueno, V. I.; Borrell, J. I.; Teixidó, J. Impact of the CXCR4 structure on docking-based virtual screening of HIV entry inhibitors. *J. Mol. Graph. Model.* **2012**, *38*, 123−136.

(18) Verdonk, M. L.; Berdini, V.; Hartshorn, M. J.; Mooij, W. T. M.; Murray, C. W.; Taylor, R. D.; Watson, P. Virtual screening using

protein-ligand docking: avoiding artificial enrichment. *J. Chem. Inf. Comput. Sci.* **2004**, *3*, 793−806.
(19) *Maybridge Bringing life to drug discovery, Maybridge Databases*; Fisher Scientific International: England, 2005.
(20) Krüger, D. M.; Evers, A. Comparison of Structure-and Ligand-Based Virtual Screening Protocols Considering Hit List Complementarity and Enrichment Factors. *ChemMedChem* **2009**, *1*, 148−158.
(21) Liu, T.; Lin, Y.; Wen, X.; Jorissen, R. N.; Gilson, M. K. BindingDB: a web-accessible database of experimentally determined protein−ligand binding affinities. *Nucleic Acids Res.* **2006**, *suppl 1*, D198.
(22) Pérez-Nueno, V. I.; Ritchie, D. W.; Rabal, O.; Pascual, R.; Borrell, J. I.; Teixido, 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**, *3*, 509−533.
(23) Wolber, G.; Langer, T. LigandScout: 3-D pharmacophores derived from protein-bound ligands and their use as virtual screening filters. *J. Chem. Inf. Model.* **2005**, *1*, 160−169.
(24) *OMEGA*, v. 2.1.0.; Open Eye Scientific Software: Santa Fe, NM, 2006.
(25) Berman, H.; Battistuz, T.; Bhat, T.; Bluhm, W.; Bourne, P.; Burkhardt, K.; Feng, Z.; Gilliland, G.; Iype, L.; Jain, S. The protein data bank. *Acta Crystallogr. D Biol. Crystallogr* **2002**, *6*, 899−907.
(26) Wang, R.; Lu, Y.; Fang, X.; Wang, S. An extensive test of 14 scoring functions using the PDBbind refined set of 800 protein-ligand complexes. *J. Chem. Inf. Comput. Sci.* **2004**, *6*, 2114−2125.
(27) Verdonk, M. L.; Cole, J. C.; Hartshorn, M. J.; Murray, C. W.; Taylor, R. D. Improved protein−ligand docking using GOLD. *Proteins: Structure, Function, Bioinf.* **2003**, *4*, 609−623.
(28) Friesner, R. A.; Banks, J. L.; Murphy, R. B.; Halgren, T. A.; Klicic, J. J.; Daniel, T.; Repasky, M. P.; Knoll, E. H.; Shelley, M.; Perry, J. K. Glide: a new approach for rapid, accurate docking and scoring. 1. Method and assessment of docking accuracy. *J. Med. Chem.* **2004**, *7*, 1739−1749.
(29) Pham, T. A.; Jain, A. N. Customizing scoring functions for docking. *J. Comput. Aided Mol. Des.* **2008**, *5*, 269−286.
(30) *SYBYL -X 1.1*; Tripos International, St. Louis, MO, 2010.
(31) Trott, O.; Olson, A. J. AutoDock Vina: improving the speed and accuracy of docking with a new scoring function, efficient optimization, and multithreading. *J. Comput. Chem.* **2010**, *2*, 455−461.
(32) Krammer, A.; Kirchhoff, P. D.; Jiang, X.; Venkatachalam, C. M.; Waldman, M. LigScore: a novel scoring function for predicting binding affinities. *J. Mol. Graph. Model.* **2005**, *5*, 395−407.
(33) Gehlhaar, D. K.; Verkhivker, G. M.; Rejto, P. A.; Sherman, C. J.; Fogel, D. R.; Fogel, L. J.; Freer, S. T. Molecular recognition of the inhibitor AG-1343 by HIV-1 protease: conformationally flexible docking by evolutionary programming. *Chem. Biol.* **1995**, *5*, 317−324.
(34) Bustanji, Y.; Taha, M. O.; Yousef, A. M.; G. Al-Bakri, A. Berberine potently inhibits protein tyrosine phosphatase 1B: Investigation by docking simulation and experimental validation. *J. Enzime Inhib. Med. Ch.* **2006**, *2*, 163−171.
(35) Feher, M. Consensus scoring for protein-ligand interactions. *Drug Discov. Today* **2006**, *9−10*, 421−428.
(36) Kaminski, G. A.; Friesner, R. A.; Tirado-Rives, J.; Jorgensen, W. L. Evaluation and reparametrization of the OPLS-AA force field for proteins via comparison with accurate quantum chemical calculations on peptides. *J. Chem. Inf. Model.* **2001**, *28*, 6474−6487.
(37) Clark, M.; Cramer, R. D., III; Van Opdenbosch, N. Validation of the general purpose Tripos 5.2 force field. *J. Comput. Chem.* **1989**, *8*, 982−1012.
(38) Brooks, B. R.; Bruccoleri, R. E.; Olafson, B. D.; Swaminathan, S.; Karplus, M. CHARMM: A program for macromolecular energy, minimization, and dynamics calculations. *J. Comput. Chem.* **1983**, *2*, 187−217.
(39) Hatse, S.; Princen, 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**, *2−3*, 300−306.
(40) Gerlach, L. O.; Skerlj, R. T.; Bridger, G. J.; Schwartz, T. W. Molecular interactions of cyclam and bicyclam non-peptide antagonists with the CXCR4 chemokine receptor. *J. Biol. Chem.* **2001**, *17*, 14153−14160.
(41) Rosenkilde, M. M.; Gerlach, L. O.; Jakobsen, J. S.; Skerlj, R. T.; Bridger, G. J.; Schwartz, T. W. Molecular mechanism of AMD3100 antagonism in the CXCR4 receptor. *J. Biol. Chem.* **2004**, *4*, 3033.
(42) Rosenkilde, M. M.; Gerlach, L. O.; Hatse, S.; Skerlj, R. T.; Schols, D.; Bridger, G. J.; Schwartz, T. W. Molecular mechanism of action of monocyclam versus bicyclam non-peptide antagonists in the CXCR4 chemokine receptor. *J. Biol. Chem.* **2007**, *37*, 27354−27365.
(43) Princen, K.; Hatse, S.; Vermeire, K.; Aquaro, S.; De Clercq, E.; Gerlach, L. O.; Rosenkilde, M.; Schwartz, T. W.; Skerlj, R.; Bridger, G. Inhibition of human immunodeficiency virus replication by a dual CCR5/CXCR4 antagonist. *J. Virol.* **2004**, *23*, 12996−13006.
(44) Hatse, S.; Princen, K.; Clercq, E. D.; Rosenkilde, M. M.; Schwartz, 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**, *5*, 752−761.
(45) Tamamura, H.; Xu, Y.; Hattori, T.; Zhang, X.; Arakaki, R.; Kanbara, K.; Omagari, A.; Otaka, A.; Ibuka, T.; Yamamoto, N. A low-molecular-weight inhibitor against the chemokine receptor CXCR4: a strong anti-HIV peptide T140. *Biochem. Biophys. Res. Commun.* **1998**, *3*, 877−882.
(46) Trent, J. O.; Wang, Z.; Murray, J. L.; Shao, W.; Tamamura, H.; Fujii, N.; Peiper, S. C. Lipid bilayer simulations of CXCR4 with inverse agonists and weak partial agonists. *J. Biol. Chem.* **2003**, *47*, 47136−47144.
(47) Våbenø, J.; Nikiforovich, G. V.; Marshall, G. R. Insight into the binding mode for cyclopentapeptide antagonists of the CXCR4 receptor. *Chem. Biol. Drug. Des.* **2006**, *5*, 346−354.
(48) Lin, J. H.; Clark, T. An analytical, variable resolution, complete description of static molecules and their intermolecular binding properties. *J. Chem. Inf. Model.* **2005**, *4*, 1010−1016.
(49) Ritchie, D. W.; Kemp, G. J. L. Protein docking using spherical polar Fourier correlations. *Proteins* **2000**, *2*, 178−194.
(50) Grant, J. A.; Gallardo, M.; Pickup, B. A fast method of molecular shape comparison: A simple application of a Gaussian description of molecular shape. *J. Comput. Chem.* **1998**, *14*, 1653−1666.
(51) Cai, W.; Xu, J.; Shao, X.; Leroux, V.; Beautrait, A.; Maigret, B. SHEF: a vHTS geometrical filter using coefficients of spherical harmonic molecular surfaces. *J. Mol. Model.* **2008**, *5*, 393−401.
(52) Karaboga, A. S.; Petronin, F.; Marchetti, G.; Souchet, M.; Maigret, B. Benchmarking of HPCC: a novel 3D molecular representation combining shape and pharmacophoric descriptors for efficient molecular similarity assessments. *J. Mol. Graph. Model.* **2012**, *41*, 20−30.
(53) Fawcett, T. An introduction to ROC analysis. *Pattern Recog. Lett.* **2006**, *8*, 861−874.
(54) Jiang, Y.; Metz, C. E.; Nishikawa, R. M. A receiver operating characteristic partial area index for highly sensitive diagnostic tests. *Radiology* **1996**, *3*, 745.
(55) Sonego, P.; Kocsor, A.; Pongor, S. ROC analysis: applications to the classification of biological sequences and 3D structures. *Brief. Bioinform.* **2008**, *3*, 198−209.
(56) Obuchowski, N. A. ROC analysis. *Am. J. Roentgenol.* **2005**, *2*, 364−372.
(57) Lasko, T. A.; Bhagwat, J. G.; Zou, K. H.; Ohno-Machado, L. The use of receiver operating characteristic curves in biomedical informatics. *J. Biomed. Inform.* **2005**, *5*, 404−415.
(58) McClish, D. K. Analyzing a portion of the ROC curve. *Med. Decision Making* **1989**, *3*, 190−195.
(59) Jain, A. N. Bias, reporting, and sharing: computational evaluations of docking methods. *J. Comput. Aided Mol. Des.* **2008**, *3−4*, 201−212.

(60) Mackey, M. D.; Melville, J. L. Better than Random? The Chemotype Enrichment Problem. *J. Chem. Inf. Model.* **2009**, *5*, 1154−1162.
(61) Venkatraman, V.; Pérez-Nueno, V. I.; Mavridis, L.; Ritchie, D. W. Comprehensive comparison of ligand-based virtual screening tools against the DUD data set reveals limitations of current 3D methods. *J. Chem. Inf. Model.* **2010**, *12*, 2079−2093.
(62) von Korff, M.; Freyss, J.; Sander, T. Comparison of ligand-and structure-based virtual screening on the DUD data set. *J. Chem. Inf. Model.* **2009**, *2*, 209−231.
(63) McGaughey, G. B.; Sheridan, R. P.; Bayly, C. I.; Culberson, J. C.; Kreatsoulas, C.; Lindsley, S.; Maiorov, V.; Truchon, J. F.; Cornell, W. D. Comparison of topological, shape, and docking methods in virtual screening. *J. Chem. Inf. Model.* **2007**, *4*, 1504−1519.
(64) Hawkins, P. C. D.; Skillman, A. G.; Nicholls, A. Comparison of shape-matching and docking as virtual screening tools. *J. Med. Chem.* **2007**, *1*, 74−82.
(65) Wong, R. S. Y.; Bodart, V.; Metz, M.; Labrecque, J.; Bridger, G.; Fricker, S. P. Comparison of the potential multiple binding modes of bicyclam, monocylam, and noncyclam small-molecule CXC chemokine receptor 4 inhibitors. *Mol. Pharmacol.* **2008**, *6*, 1485−1495.
(66) Perez-Nueno, V. I.; Ritchie, D. W. Using consensus-shape clustering to identify promiscuous ligands and protein targets and to choose the right query for shape-based virtual screening. *J. Chem. Inf. Model* **2011**, *6*, 1233−1248.
(67) Pérez-Nueno, V. I.; Ritchie, D. W. Applying in silico tools to the discovery of novel CXCR4 inhibitors. *Drug Dev. Res* **2011**, *1*, 95−111.
(68) Pérez-Nueno, V. I.; Ritchie, D. W. Identifying and characterizing promiscuous targets: Implications for virtual screening. *Expert Opin. Drug Discov.* **2012**, *1*, 1−17.
(69) Bridger, G. J.; Skerlj, R. T.; Padmanabhan, S.; Martellucci, S. A.; Henson, G. W.; 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**, *19*, 3971−3981.
(70) De Clercq, E. Inhibition of HIV infection by bicyclams, highly potent and specific CXCR4 antagonists. *Mol. Pharmacol.* **2000**, *5*, 833−839.
(71) Esté, J. A.; De Vreese, K.; Witvrouw, M.; Schmit, J. C.; Vandamme, A. M.; Anné, J.; Desmyter, J.; Henson, G. W.; Bridger, G.; De Clercq, E. Antiviral activity of the bicyclam derivative JM3100 against drug-resistant strains of human immunodeficiency virus type 1. *Antiviral Res.* **1996**, *2*, 297−307.
(72) Esté, J. A.; Cabrera, C.; De Clercq, E.; Struyf, S.; Van Damme, J.; Bridger, G.; Skerlj, R. T.; Abrams, M. J.; Henson, G.; Gutierrez, A. Activity of different bicyclam derivatives against human immunodeficiency virus depends on their interaction with the CXCR4 chemokine receptor. *Mol. Pharmacol.* **1999**, *1*, 67−73.
(73) Egberink, H. F.; De Clercq, E.; Van Vliet, A. L. W.; Balzarini, J.; Bridger, G. J.; Henson, G.; Horzinek, M. C.; Schols, D. Bicyclams, selective antagonists of the human chemokine receptor CXCR4, potently inhibit feline immunodeficiency virus replication. *J. Virol.* **1999**, *8*, 6346−6352.
(74) Pérez-Nueno, V. I.; Pettersson, S.; Ritchie, D. W.; Borrell, J. I.; Teixido, J. Discovery of novel HIV entry inhibitors for the CXCR4 receptor by prospective virtual screening. *J. Chem. Inf. Model.* **2009**, *4*, 810−823.
(75) Skerlj, R.; Bridger, G.; McEachern, E.; Harwig, C.; Smith, C.; Kaller, A.; Veale, D.; Yee, H.; Skupinska, K.; Wauthy, R. Design of novel CXCR4 antagonists that are potent inhibitors of T-tropic (X4) HIV-1 replication. *Bioorg. Med. Chem. Lett.* **2011**, 1414−1418.
(76) Bridger, G. J.; Skerlj, R.; Kaller, A.; Harwig, C.; Bogucki, D.; Wilson, T.; Crawford, J.; Mceachern, E.; Atsma, B. *Chemokine Receptor Binding Heterocyclic Compounds with Enhanced Efficacy*. United States Patent WO/2003/055876, 2003.
(77) Ichiyama, K.; Yokoyama-Kumakura, S.; Tanaka, Y.; Tanaka, R.; Hirose, K.; Bannai, K.; Edamatsu, T.; Yanaka, M.; Niitani, Y.; Miyano-Kurosaki, N. A duodenally absorbable CXC chemokine receptor 4 antagonist, KRH-1636, exhibits a potent and selective anti-HIV-1 activity. *Proc. Natl. Acad. Sci. USA* **2003**, *7*, 4185−4190.
(78) Yamazaki, T.; Saitou, A.; Ono, M.; Yokoyama, S.; Bannai, K.; Hirose, K.; Yanaka, M. *Novel Nitrogenous Compound and Use Thereof*. World Patent WO/2003/029218, 2003.
(79) Mosley, C. A.; Wilson, L. J.; Wiseman, J. M.; Skudlarek, J. W.; Liotta, D. C. Recent patents regarding the discovery of small molecule CXCR4 antagonists. *Expert Opin. Ther. Pat.* **2009**, *1*, 23−38.
(80) Bodart, V.; Anastassov, V.; Darkes, M. C.; Idzan, S. R.; Labrecque, J.; Lau, G.; Mosi, R. M.; Neff, K. S.; Nelson, K. L.; Ruzek, M. C. Pharmacology of AMD3465: a small molecule antagonist of the chemokine receptor CXCR4. *Biochem. Pharmacol.* **2009**, *8*, 993−1000.
(81) Bridger, G. J.; Boehringer, E. M.; Wang, Z.; Schols, D.; Skerlj, R. T.; Bogucki, D. E. *Methods to treat conditions mediated by chemokine receptors*. United States Patent 7,160,872 B2, 2008.
(82) Bridger, G. J.; Boehringer, E. M.; Wang, Z.; Schols, D.; Skerlj, R. T.; Bogucki, D. E. *Antiviral compounds*. United States Patent 6,506,770 B1, 2003.
(83) Catalano, J. G.; Gudmundsson, K. S.; Svolto, A.; Boggs, S. D.; Miller, J. F.; Spaltenstein, A.; Thomson, M.; Wheelan, P.; Minick, D. J.; Phelps, D. P. Synthesis of a novel tricyclic 1, 2, 3, 4, 4a, 5, 6, 10b-octahydro-1, 10-phenanthroline ring system and CXCR4 antagonists with potent activity against HIV-1. *Bioorg. Med. Chem. Lett.* **2010**, *7*, 2186−2190.
(84) Bridger, G. J.; Skerlj, R. T.; Thornton, D.; Padmanabhan, S.; Martellucci, S. A.; Henson, G. W.; Abrams, M. J.; Yamamoto, N.; Vreese, K. D. Synthesis and structure-activity relationships of phenylenebis (methylene)-linked bis-tetraazamacrocycles that inhibit HIV replication. Effects of macrocyclic ring size and substituents on the aromatic linker. *J. Med. Chem.* **1995**, *2*, 366−378.
(85) 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**, *11*, 3412−3415.
(86) Jenkinson, S.; Thomson, M.; McCoy, D.; Edelstein, M.; Danehower, S.; Lawrence, W.; Wheelan, P.; Spaltenstein, A.; Gudmundsson, K. Blockade of X4-Tropic HIV-1 cellular entry by GSK812397, a potent noncompetitive cxcr4 receptor antagonist. *Antimicrob. Agents Chemother.* **2010**, *2*, 817.
(87) Gudmundsson, K. S.; Boggs, S. D.; Catalano, J. G.; Svolto, A.; Spaltenstein, A.; Thomson, M.; Wheelan, P.; Jenkinson, S. Imidazopyridine-5, 6, 7, 8-tetrahydro-8-quinolinamine derivatives with potent activity against HIV-1. *Bioorg. Med. Chem. Lett.* **2009**, *22*, 6399−6403.
(88) Bridger, G. J.; Skerlj, R. T.; Hernandez-Abad, P. E.; Bogucki, D. E.; Wang, Z.; Zhou, Y.; Nan, S.; Boehringer, E. M.; Wilson, T.; Crawford, J. Synthesis and Structure− Activity Relationships of Azamacrocyclic CXC Chemokine Receptor 4 Antagonists: Analogues Containing a Single Azamacrocyclic Ring are Potent Inhibitors of T-Cell Tropic (X4) HIV-1 Replication. *J. Med. Chem.* **2009**, *3*, 1250−1260.
(89) Yen, C. F.; Hu, C. K.; Chou, M. C.; Tseng, C. T.; Wu, C. H.; Huang, Y. H.; Chen, S. J.; King, C. H. R. *Pyrimidine compounds*. United States Patent US 7,589,197 B2, 2009.
(90) Hachet-Haas, M.; Balabanian, K.; Rohmer, F.; Pons, F.; Franchet, C.; Lecat, S.; Chow, K. Y. C.; Dagher, R.; Gizzi, P.; Didier, B. Small neutralizing molecules to inhibit actions of the chemokine CXCL12. *J. Biol. Chem.* **2008**, *34*, 23189−23199.
(91) Miller, J. F.; Turner, E. M.; Gudmundsson, K. S.; Jenkinson, S.; Spaltenstein, A.; Thomson, M.; Wheelan, P. Novel N-substituted benzimidazole CXCR4 antagonists as potential anti-HIV agents. *Bioorg. Med. Chem. Lett.* **2010**, *7*, 2125−2128.
(92) Gudmundsson, K. S.; Sebahar, P. R.; Richardson, L. D.; Miller, J. F.; Turner, E. M.; Catalano, J. G.; Spaltenstein, A.; Lawrence, W.; Thomson, M.; Jenkinson, S. Amine substituted *N*-(1-*H*-benzimidazol-2ylmethyl)-5, 6, 7, 8-tetrahydro-8-quinolinamines as CXCR4 antagonists with potent activity against HIV-1. *Bioorg. Med. Chem. Lett.* **2009**, *17*, 5048−5052.

(93) Gudmundsson, K. *Pyranopyridine compounds. WO/2007/008539*, 2007.
(94) Murakami, T.; Kumakura, S.; Yamazaki, T.; Tanaka, R.; Hamatake, M.; Okuma, K.; Huang, W.; Toma, J.; Komano, J.; Yanaka, M. The novel CXCR4 antagonist KRH-3955 is an orally bioavailable and extremely potent inhibitor of human immunodeficiency virus type 1 infection: comparative studies with AMD3100. *Antimicrob. Agents Chemother.* **2009**, *7*, 2940−2948.
(95) Thomas, W. D.; Leleti, M. R.; Pennell, A. M. K. *CXCR4 modulators*. United States Patent 2007/0275965, 2007.
(96) Kokubo, M.; Tanaka, M.; Ochiai, H.; Takaoka, Y.; Shibayama, S. *Compound containing basic group and use thereof*. WO/2007/049771, 2007.
(97) Liang, Z.; Zhan, W.; Zhu, A.; Yoon, Y.; Lin, S.; Sasaki, M.; Klapproth, J. M. A.; Yang, H.; Grossniklaus, H. E.; Xu, J. Development of a unique small molecule modulator of CXCR4. *PloS One* **2012**, *4*, e34038−e34038.