

On the Value of Homology Models for Virtual Screening: Discovering hCXCR3 Antagonists by Pharmacophore-Based and Structure-Based Approaches

Dane Huang,[†] Qiong Gu,[†] Hu, Ge,[†] Jiming Ye,[‡] Noeris K. Salam,^{||} Arnie Hagler,^{§,||} Hongzhuan Chen,[⊥] and Jun Xu^{†,*}

[†]School of Pharmaceutical Sciences & Institute of Human Virology, Sun Yat-sen University, Guangzhou, China 510006

[‡]Health Innovations Research Institute and School of Health Sciences, Royal Melbourne Institute of Technology (RMIT) University, P.O. Box 71, Melbourne, VIC 3083, Australia

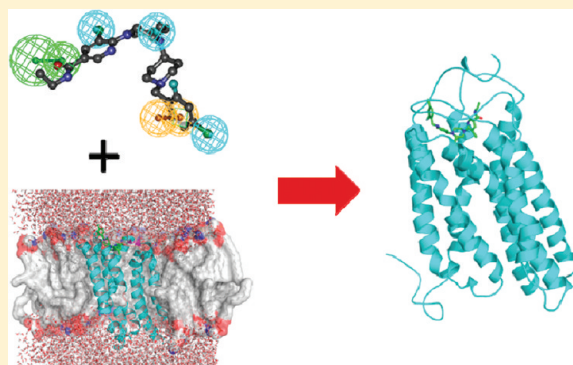
[§]Department of Chemistry, University of Massachusetts, 701 Lederle Graduate Research Tower, 710 North Pleasant Street, Amherst, Massachusetts 01003-9336, United States

^{||}Shifa Biomedical, 1 Great Valley Parkway, Suite 8, Malvern, Pennsylvania 19355, United States

[⊥]School of Medicine, Institute of Medical Sciences, Shanghai Jiao Tong University, Shanghai, China 200025

Supporting Information

ABSTRACT: Human chemokine receptor CXCR3 (hCXCR3) antagonists have potential therapeutic applications as antiviral, antitumor, and anti-inflammatory agents. A novel virtual screening protocol, which combines pharmacophore-based and structure-based approaches, was proposed. A three-dimensional QSAR pharmacophore model and a structure-based docking model were built to virtually screen for hCXCR3 antagonists. The hCXCR3 antagonist binding site was constructed by homology modeling and molecular dynamics (MD) simulation. By combining the structure-based and ligand-based screenings results, 95% of the compounds satisfied either pharmacophore or docking score criteria and would be chosen as hits if the union of the two searches was taken. The false negative rates were 15% for the pharmacophore model, 14% for the homology model, and 5% for the combined model. Therefore, the consistency of the pharmacophore model and the structural binding model is $219/273 = 80\%$. The hit rate for the virtual screening protocol is $273/286 = 95\%$. This work demonstrated that the quality of both the pharmacophore model and homology model can be measured by the consistency of the two models, and the false negatives in virtual screening can be reduced by combining two virtual screening approaches.



■ INTRODUCTION

Chemokines are small (8–10 kDa) peptides responsible among other things for leukocytes migration to sites of infection or injury. Chemokines exert their chemotactic functions by binding to chemokine receptors. Chemokine receptors are G protein-coupled receptors (GPCRs) with seven transmembrane domains. Abnormal expressions of chemokines and their receptors are implicated in the pathogenesis of several human diseases, including autoimmune, chronic inflammatory diseases, immunodeficiency, cancer, and viral infection.^{1–4}

Human chemokine receptor CXCR3 (hCXCR3) is a G_{αi} protein-coupled receptor in the CXC chemokine receptor superfamily. Binding of CXC chemokines CXCL9 (MIG), CXCL10, and CXCL11 (IP-10, I-TAC) to hCXCR3 is necessary for CXCR3 signal transduction and activation of the expression of Th1, NK cell, pDC as well as mast cells.^{5,6} Dysregulated expression of CXCR3 is involved in the development of chronic

hepatitis B, bronchial asthma, transplantation rejection, autoimmune diseases, and dermatosis.⁷

Many hCXCR3 antagonists have been discovered^{8–15} in the past few years. The first QSAR study for hCXCR3 antagonists was published in 2009.¹⁶ Another QSAR model for CXCR3 antagonists was reported in 2010. This model was based on multiple linear regressions (MLR) and least-squares support vector machine (LS-SVM) approaches.¹⁷ The binding model of hCXCR3 and its antagonists is unknown because of the lack of an hCXCR3 crystal structure. In order to improve the performance of virtual screening campaign for hCXCR3 antagonists, we propose a new virtual screening protocol that combines pharmacophore-based and structure-based approaches. First, an hCXCR3 homology model was built.

Received: February 5, 2012

Published: April 30, 2012

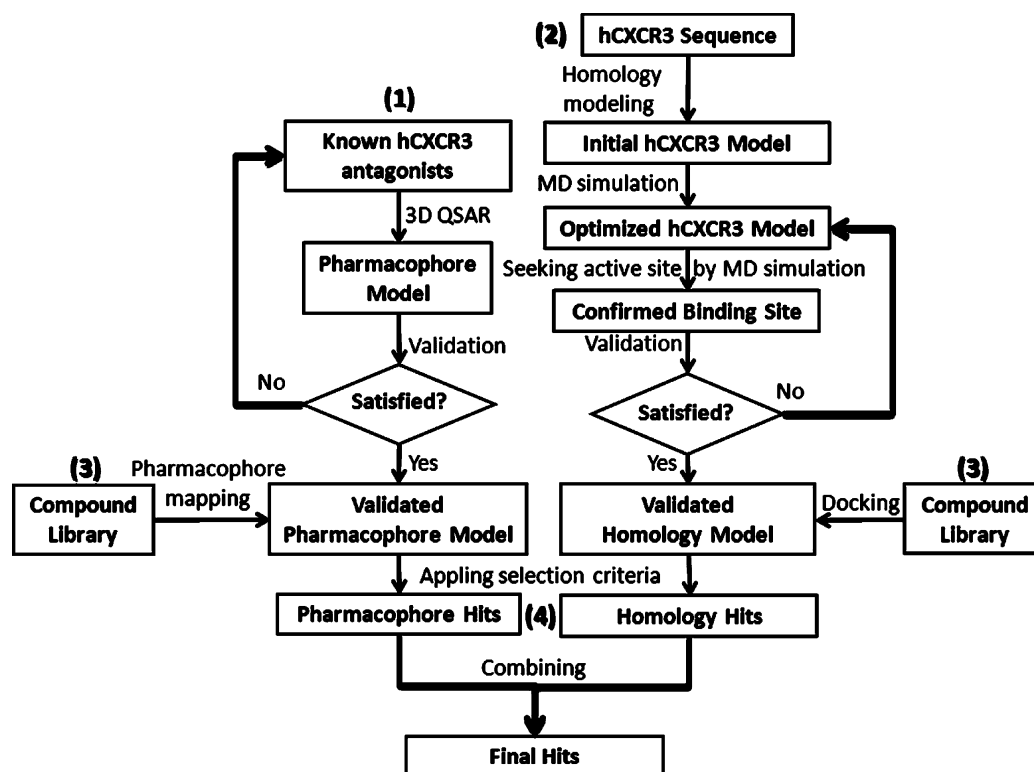


Figure 1. Virtual screening protocol that combines pharmacophore-based and structure-based approaches for hCXCR3 antagonist lead identification.

Then, a three-dimensional QSAR pharmacophore model was constructed. Finally, the two models were combined to predict hCXCR3 antagonist hits. The protocol is depicted in Figure 1.

The protocol consists of four steps: (1) building a 3D QSAR pharmacophore model from known inhibitors, (2) building an hCXCR3 homology model and refining the model by MD simulations, (3) virtual screening against pharmacophore and homology models with a compound library (containing 180,952 compounds from the NCI database and 286 known hCXCR3 antagonists), and (4) combining the two hit-sets from pharmacophore and homology models to produce the final virtual screening hits.

MATERIALS AND METHODS

Data as hCXCR3 Homology Model Templates. The sequence of hCXCR3 was taken from the NCBI protein database. The X-ray structure of hCXCR4 (PDB code, 3ODU) was selected as the template of hCXCR3. The sequence identity and similarity between hCXCR3 and hCXCR4 are 32.4% and 54.0%, respectively. The sequence alignment of hCXCR3 and hCXCR4 was performed using the Align Multiple Sequence module of Discovery Studio 2.1 (DS 2.1, Accelrys).¹⁸ Because the X-ray structure of hCXCR4 contains only 293 residues (27 to 319) from the whole sequence (352 residues), the residues for the N-terminal domain (containing the chemokine binding site) have no structural templates.^{19,20} To construct a model for the missing 27 N-terminal residues, a second structural template was taken from a bovine rhodopsin crystal structure (PDB code: 1F88). The sequence identity and similarity between hCXCR3 and the second structure template were 18.2% and 48.9%, respectively.

Method for MD Simulations. MD simulations were performed using Desmond v2.4 and the OPLS-AA force field.²¹ The simulations were divided into two steps. During the first

step, the energy minimized conformation of hCXCR3 obtained from the homology model was used as the initial structure for the MD simulations. The hCXCR3 was embedded in a hydrated lipid bilayer (palmitoyl oleo phosphatidyl ethanolamine, POPE), in which the lipid and water molecules were represented explicitly; and the bilayer's long axis was normalized to the membrane–water interface. The α -helical domains of the receptor were placed at the same level as the lipid bilayer. Overlapped lipid and water molecules were removed. Water was represented as a simple point charge (SPC) model. Seven chloride counterions were added to neutralize the system. The boundary conditions of the system were orthorhombic, measuring $71 \text{ \AA} \times 80 \text{ \AA} \times 112 \text{ \AA}$, and contained 148 membrane molecules and 11,872 water molecules.

During the second step, the most active hCXCR3 antagonist was docked to the averaged structure generated from the first step MD simulations. The antagonist binding site was predicted by homology to the binding site in the homologous hCXCR4.¹⁹ Then the ligand–receptor complex was simulated with MD using the same settings of the parameters as in the first step MD simulation.

The system was equilibrated in an isothermal–isobaric ensemble at 310 K and 1 bar using the Berendsen coupling scheme with $5 \text{ kcal mol}^{-1} \text{ \AA}^{-2}$ harmonic position restraints applied to all non-hydrogen atoms of the protein for 1.2 ns. The unrestrained system was then simulated for an additional 1.2 ns to equilibrate the aspect ratio of the simulation box further. During the equilibration process, van der Waals and short-range electrostatic interactions were cut off at 9 \AA . Long-range electrostatic interactions were computed using the particle mesh Ewald method.²² All bond lengths to hydrogen atoms were constrained using M-SHAKE.²³ A RESPA

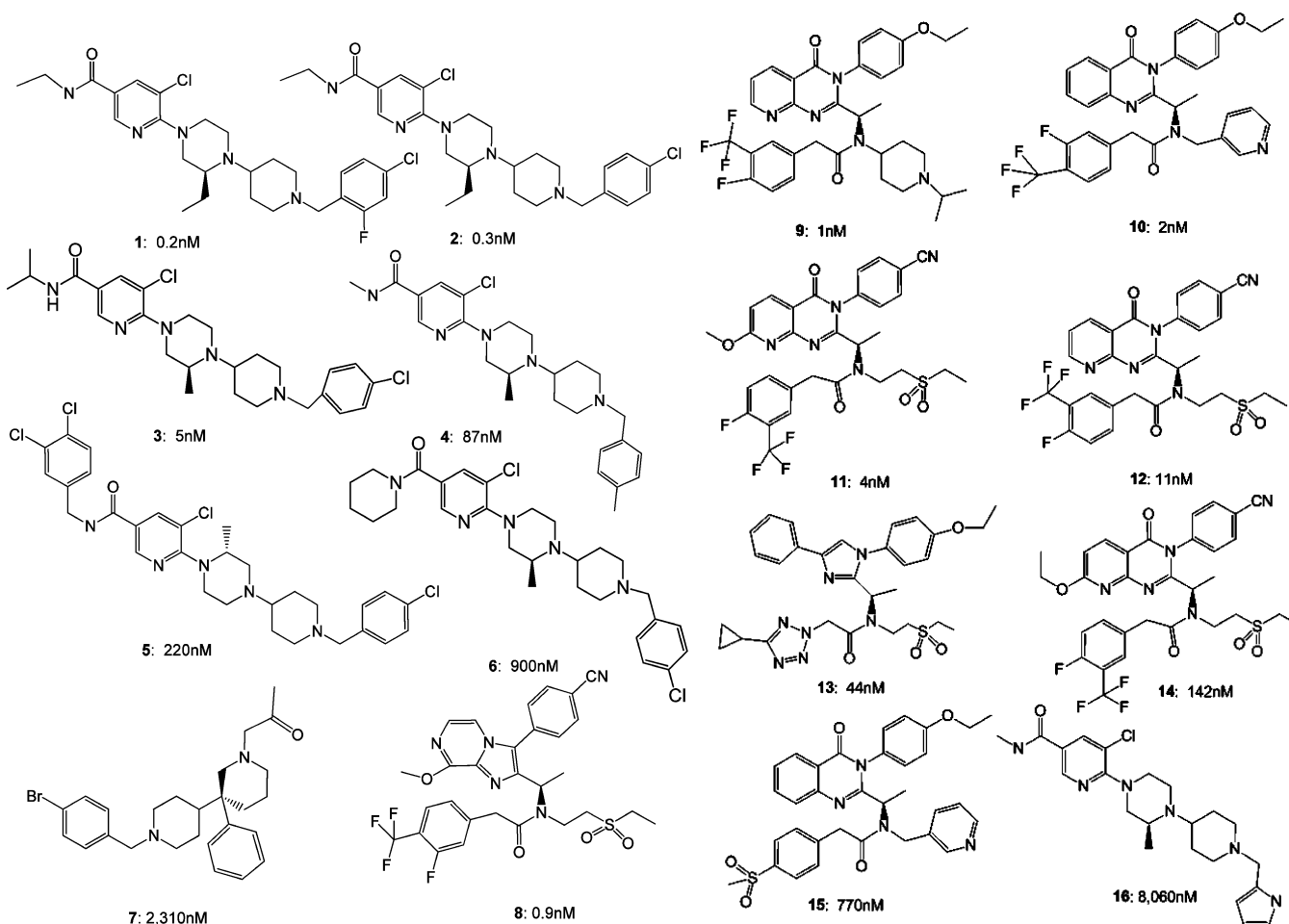


Figure 2. Structures and bioactivities of the 16 training compounds used to generate 3D pharmacophore model.

integrator²⁴ was used with a time step of 2 fs, and long-range electrostatics were computed every 6 fs.

During the second step, MD simulations were carried out for 12 ns starting with the averaged conformation from the last few ns after the equilibration. The optimized hCXCR3 homology model was acquired by averaging the conformation over the last few nanoseconds.

Data for Validating Homology Model. A training set comprising 16 compounds^{11,12,14,25,26} was selected on the basis of diversity of the chemical structures and bioactivity. The CXCR3 antagonist activities were measured in CXCR3 mediated chemotaxis assays for the responses to CXCL10 (IP-10). Their binding affinities (IC₅₀s) range over 5 orders magnitude from 0.2 nM to 8060 nM. The structures of the training set are depicted in Figure 2. The test set included 286 hCXCR3 antagonist structures derived from refs 9–11,14,15,25–31, and 180,982 NCI compounds were docked against the optimized homology model with the MOE docking algorithm to validate the binding mode. The binding site was minimized using the AMBER 99 force field in MOE with the default parameter. All the compounds were docked, employing Triangle Matcher as the placement method and the function London dG as the first scoring function. The refinement was set to force field (AMBER 99), and the docked poses were energy minimized in the receptor pocket.

Similarly, compounds were also docked independently with Glide 5.7³² (Glide v5.7, Schrödinger) utilizing the standard precision (SP) protocol, as well as LigandFit (DS 2.1,

Accelrys).¹⁸ The center of the Glide grid was defined by the antagonist bound to the homology model. Default settings were used for both the grid generation and docking. This included a scaling of the ligand radii by 0.8 to partially account for suboptimal fits that could be accommodated by minor receptor movements. The LigandFit docking process was performed with Monte Carlo simulations. All the dock parameters were set as default.

Data for Creating Pharmacophore Model. The training set for validating homology model was used to generate the pharmacophore model. A test set of 49 known antagonists^{11,12,26,28} outside the training set was used to test the pharmacophore model. Another larger compound library containing 180,952 compounds selected from the NCI database with 286 known hCXCR3 antagonists^{9–11,14,15,25–31} was used to validate the pharmacophore model as well.

All the compounds were converted to 3D structures, and energy was minimized in MOE.³³ The 3D conformers of all the compounds were generated using the Diverse Conformation Generation module in DS 2.1¹⁸ running with the Best conformations option. A total of 250 conformers were generated within an energy threshold of 20.0 kcal/mol above the global energy minimum for each compound.

Method for Creating the Pharmacophore Model. The Feature Mapping module of DS2.1¹⁸ was employed to determine the inherent chemical features of the compounds in the training set. The Feature Mapping protocol computes all possible pharmacophore feature mappings for the selected

ligands. Hydrophobic features were most prevalent, while charge-related pharmacophore features (such as pos_ionizable, neg_ionizable, pos_charge, neg_charge) were not observed. Therefore, the HBA (hydrogen bond acceptor), HBD (hydrogen bond donor), HY (hydrophobic), and RA (ring aromatic) were chosen to build the 3D pharmacophore models. The numbers of minimal and maximal features were set to 4 and 5, respectively, to generate the pharmacophore hypotheses. The fisher validation value was set to 95%. Ten hypotheses were generated by Catalyst (Accelrys).¹⁸ The final pharmacophore model was constructed by combining the 10 hypotheses using the consensus principle.

Method for Validating the Pharmacophore Model.

The pharmacophore model was validated by measuring correlation coefficient,^{18,34} total cost,^{18,34} null cost,^{18,34} root-mean-square deviation (rmsd),^{18,34} configuration cost,^{18,34} and fit value.^{18,34} Four different validation methods were used for valuation. They are as follows: predicting the relative activities of a test set of 49 antagonists, Fischer's randomization,^{18,35} calculating the enrichment factor³⁶(EF) obtained from a database of 100 K compounds, and the goodness of fit score³⁶ (GF). The confidence level (Fischer's randomization validation value) of the pharmacophore model was calculated using the following formula

$$S = 100[1 - (1 + X)/Y] \quad (1)$$

where S = the significance level or confidence level, X = the number of randomized hypotheses scored better than the selected pharmacophore model, and Y = the number of random hypotheses generated.

The enrichment factor³⁶(EF) calculated as fraction of actives in top hits, divided by fraction of actives in database, and goodness of fit score³⁶ (GF) were calculated using the following formulas

$$EF = (Ha \times D)/(Ht \times A) \quad (2)$$

$$GF = [(Ha \times (3A + Ht))/(4Ht \times A)] \times (1 - (Ht - Ha)/(D - A)) \quad (3)$$

where, Ht = total number of hits, Ha = the total number of active molecules in the Ht hits, D = total number of molecules in database, and A = the total number of actives in the database.

RESULTS AND DISCUSSION

Homology Modeling Results. The Ramachandran plot of the optimized homology model shows that 92% of the residues remain in "favorable" areas, while the remaining 8% of the residues are in allowed areas (Figure 3). The profile-3D score^{37,38} of the homology model was 147.9, which was close to the expected score (152.6).

The "Kyte-Doolittle hydrophobicity plot" is a plot of the amino acids in a protein against their hydrophobicity index and is used to predict the transmembrane region of protein.³⁹ The final homology model was further validated by the hydrophobicity distribution plot. The hydrophobicity of a residue is calculated as the average of the two residues before it, its own value, and the two residues after it (Figure 4). Peaks I–VII (Figure 4) represent seven transmembrane regions in both CXCR4 structure and the final homology model of hCXCR3. As shown in Figure 4, the trends of the two curves are similar, and the two structures are consistent giving some confidence to the proposed structure.

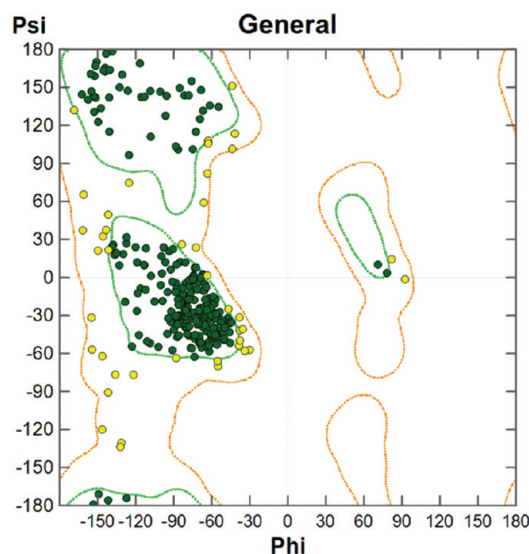


Figure 3. Ramachandran ψ – Φ distribution plot for the final homology model. A total of 92% of the residues remain in the favorable areas (filled circles), while the remaining 8% of the residues reside in allowed areas (open circles).

Pharmacophore Modeling Results. Quality of the Pharmacophore Model. Ten pharmacophore models (hypotheses) were created from the hCXCR3 antagonist training set, from which the pharmacophore features, HBA, HBD, RA, and HY, were derived. The 10 pharmacophore models were constructed with different combinations of these pharmacophore features. The pharmacophore models and their pharmacophore feature combinations are listed in Table 1.

In Table 1, the value in "Total cost" should be close to the fixed cost which is 65.20. The value in "Cost difference"^{18,34} should be close to the null cost which is 218.51. The values in "rmsd"^{18,34} represent the difference between the pharmacophore model and the training compounds. The quality of the hypothesis is inversely proportional to its rmsd. The "Correlation"^{18,34} measures the relation between predicted activity and observed activity. A perfect model would result in a correlation of 1. "Max. fit"^{18,34} is the maximum fit value for each hypothesis. This fit function measures how well the ligand features map to the pharmacophore features, specifically the distance that separates the feature on the molecule from the centroid of the hypothesis feature. The "Error cost"^{18,34} is a value that increases when the rmsd values between estimated and observed activities for the training set molecules increase.

Most hypotheses have five pharmacophore features, in which HY is the most prevalent. Hypothesis 1 has the most diversified pharmacophore features and covers most of the pharmacophore features derived from the training compounds. On the basis of the above analyses of all the parameters in Table 1, hypothesis 1 is ranked as the best. Its pharmacophore features and configuration are depicted in Figure 5.

In the training set, compound 1 is most active (IC_{50} = 0.2 nM), and compound 16 is least active (IC_{50} = 8060 nM). The two compounds are superimposed on the hypothesis 1 pharmacophore model in Figure 6. Compound 1 maps well to the pharmacophore model of the hypothesis 1, while compound 16 can only be partially mapped to the model, consistent with their relative activities, as would be expected because they are components of the training set.

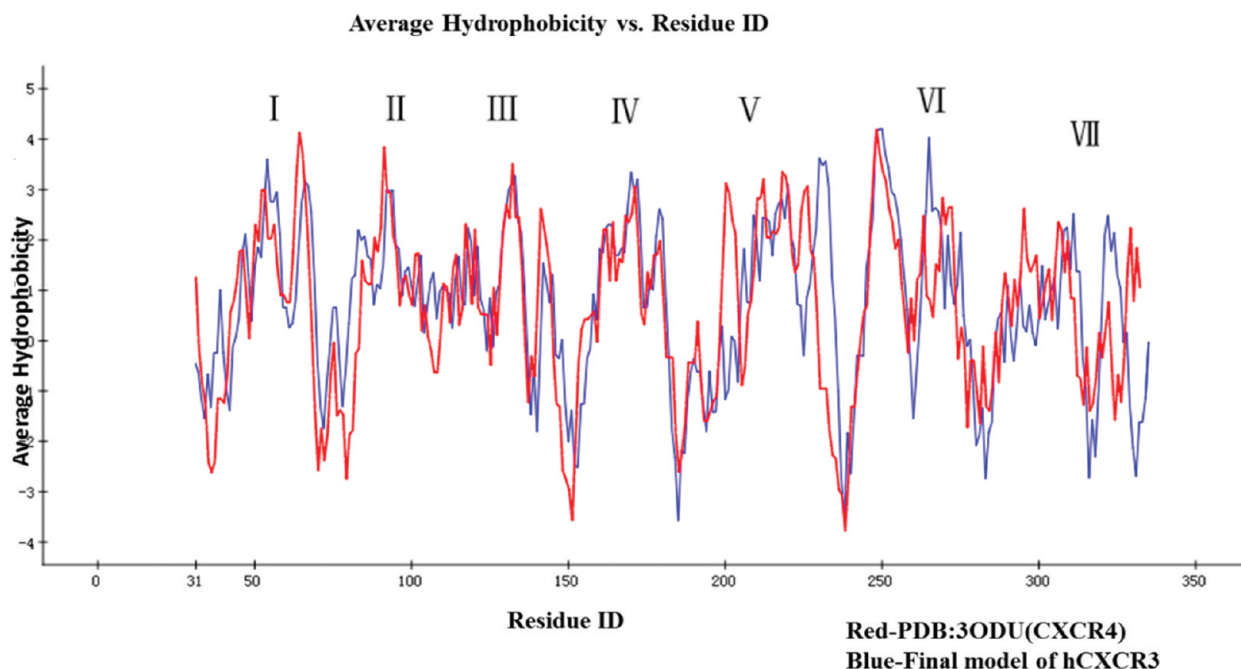


Figure 4. Superimposed hydrophobicity distribution plots of hCXCR4 crystal structure (red) and hCXCR3 homology model (blue).

Table 1. Pharmacophore Models and Their Pharmacophore Feature Combinations

No.	Total cost	Cost difference	rmsd	Correlation	Max. fit	Features	Error cost
1	79.38	139.13	1.23	0.96	13.01	HBA,HYP,HYP,HYP,RA	58.64
2	90.00	128.51	1.47	0.95	15.83	HYP,HYP,HYP,HYP,RA	63.80
3	94.43	124.08	1.90	0.91	11.23	HBD,HYP,HYP,HYP,RA	75.34
4	94.96	123.55	1.67	0.94	15.87	HBA,HYP,HYP,HYP,HYP	68.66
5	102.99	115.52	2.14	0.89	12.15	HYP,HYP,HYP,HYP,RA	83.22
6	103.54	114.97	2.16	0.89	12.20	HYP,HYP,HYP,HYP,RA	83.73
7	103.54	114.97	2.13	0.89	13.12	HYP,HYP,HYP,HYP,RA	82.67
8	103.62	114.89	2.16	0.89	12.35	HYP,HYP,HYP,RA,RA	83.65
9	104.80	113.71	1.90	0.92	13.59	HYP,HYP,HYP,HYP	75.34
10	105.29	113.22	2.20	0.88	12.64	HYP,HYP,HYP,HYP,HYP	85.01

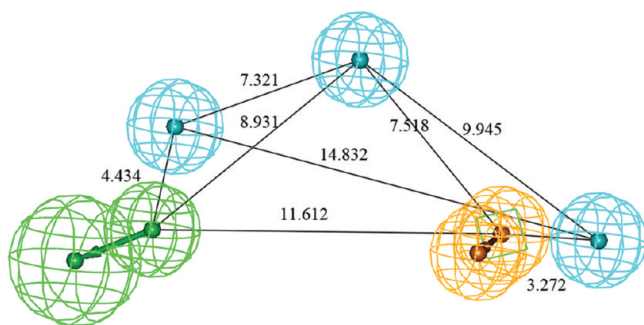


Figure 5. Best pharmacophore hypothesis for CXCR3 antagonists. 3D spatial arrangements and the distances between the centroids are represented in solid black lines. The pharmacophore features are color coded green (hydrogen bond donor), orange (aromatic ring), and cyan (hydrophobic group).

The 16 training compounds were divided into three activity groups: (1) Group I, most active ($IC_{50} \leq 100$ nM, +++); (2) Group II, active (100 nM $< IC_{50} \leq 1000$ nM, ++); and (3) Group III, less active ($IC_{50} > 1000$ nM, +), as shown in Table 2. In Table 2, "Fit value" is calculated to measure if ligand features are mapped to the generated pharmacophore features or not, and it also contains a distance term, which measures the

distance that separates the feature on the molecule from the centroid of the hypothesis feature.^{18,34} Fit value measures the ability of the model to account for the activity for the activity of the compound. The relative "Error" is defined as the ratio of the predicted activity and observed activities. It is given as $B_{(\text{predicted } IC_{50} \text{ value})}/A_{(\text{experiment } IC_{50} \text{ value})}$ (if $B > A$) and, $-A_{(\text{experiment } IC_{50} \text{ value})}/B_{(\text{predicted } IC_{50} \text{ value})}$ (if $A > B$). Note the negative value that the predicted IC_{50} value is lower than the experiment IC_{50} value. We see in Table 2 that the predicted activities are consistent with the observed activities for all compounds in the training set (see the last two columns in Table 2).

Testing the Preferred Pharmacophore Model. On the basis of the above analyses, hypothesis 1 is designated the "preferred pharmacophore model" for hCXCR3 antagonist activity prediction. As a test of this model, we predicted the activities of 49 antagonists^{11,12,26,28} outside the training set and compared then with the experimental values in Table 3. Although there are some quantitative differences, we see from the last two columns that the preferred pharmacophore model is able to accurately predict the relative activities of 47 (96%) of the compounds in the test set. The resulting correlation coefficient (r) between the experimental and predicted activities of the test

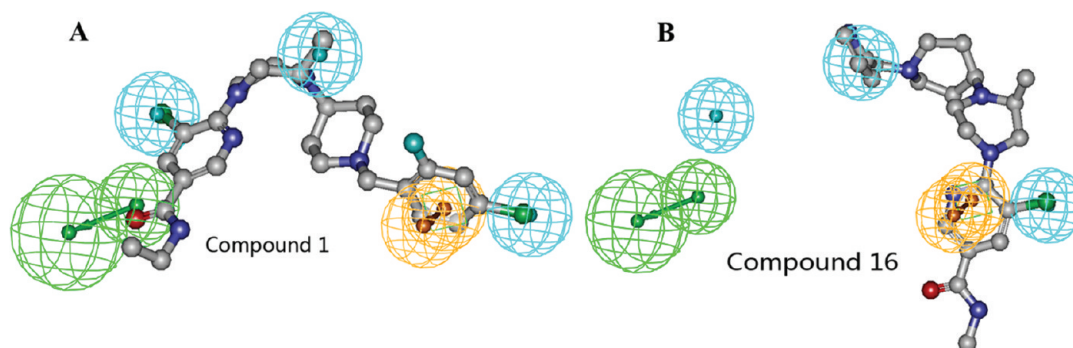


Figure 6. (A) Compound 1, the most potent compound in the training set, superposes well on the pharmacophore defined by hypothesis 1, while (B) compound 16, the least potent compound, is missing two pharmacophore features in the hypothesis.

Table 2. Ability of Hypothesis 1 To Account for the Observed Activities of the Training Set

No.	Fit value	Predicted IC ₅₀ (nM)	Observed IC ₅₀ (nM)	Error	Activity scale	Predicted scale
1	11.58	0.082	0.2	-2.5	+++	+++
2	11.28	0.16	0.3	-1.8	+++	+++
3	9.83	4.6	0.9	5.1	+++	+++
4	9.87	4.2	1	4.2	+++	+++
5	9.96	3.5	2	1.7	+++	+++
6	10.11	2.4	4	-1.7	+++	+++
7	9.62	7.5	5	1.5	+++	+++
8	9.86	4.3	11	-2.6	+++	+++
9	8.56	86	44	2.0	+++	+++
10	8.60	79	87	-1.1	+++	+++
11	8.39	126	142	-1.1	++	++
12	7.78	522	220	2.4	++	++
13	7.96	339	770	-2.3	++	++
14	7.55	873	900	-1.0	++	++
15	7.45	1102	2310	-2.1	+	+
16	7.23	1855	8060	-4.3	+	+

set was 0.88 (Figure 7). The two activities that were not accurately predicted had large observed IC₅₀s, and the compounds were not very active.

Fisher's randomization method^{18,35} was employed to measure the statistical significance of the preferred pharmacophore model in order to ensure that the chosen pharmacophore model was significantly better than would be obtained from the random combinations of the pharmacophore features. Nineteen random pharmacophore models were generated from the random combinations of the pharmacophore features (Figure 8) at the 95% confidence; they were compared against the preferred pharmacophore model. The fixed cost values of these 19 random models are plotted along with those of the preferred pharmacophore model. None of these 19 random models had lower cost or RMSDs than the preferred pharmacophore model (blue line in Figure 8). Thus, the Fisher randomization results indicate that the preferred pharmacophore model was not generated by chance.

The enrichment factor (EF)³⁶ and goodness of fit score (GF)³⁶ of the pharmacophore were calculated for hypothesis 1. A compound library containing 180,952 NCI compounds and 286 CXCR3 antagonists were screened against hypothesis 1. Hits were determined if their conformations could be mapped on to all five pharmacophore features. The screen produced 1201 compounds, all of which contained the five pharmacophore features. These were comprised of 243 (85%) of the known antagonists, and we also got 958 hits compounds from

the NCI database. Of course, there is no guarantee that the "inactives" from the database are not active, but for the purpose of calculating enrichment this is a conservative assumption. We also got 243 (85%) hits from 286 known CXCR3 antagonists. Thus, on the basis of this, the EF was 128 and the GF was 0.36 (data shown in Table 5). Thus, it is 128 times more probable to pick an active compound from the top ~1200 hits database than expected by chance.

Refined Homology Model. The initial hCXCR3 homology model was refined by MD simulations. In the first step of the MD simulation, the apo protein was equilibrated. In the second step, an optimized structure for a receptor–ligand complex was created and MD carried out in order to derive a structure to be used in virtual screening (Figure 9).

As shown in Figure 9, the MD simulation started with the homology model, structure A. The system stabilized after 1.2 ns, and the system energy decreased by 1823 kcal/mol. After the first phase, a refined structure, B, was obtained by extracting the configuration with the lowest potential energy from the trajectory. The binding pocket in structure B was identified by homology with the template, CXCR4,¹⁹ (3ODU). Then, compound 1 (most active compound in the training set) was docked into the binding pocket, and MD simulation of the ligand bound structure was carried out for 12 ns. Finally, the optimized hCXCR3 homology model, C, was obtained by averaging the last few ns of the MD trajectory.

Table 3. Experimental and Estimated IC₅₀ Values of a Test Set of hCXCR3 Antagonists Using the Pharmacophore Model of Hypothesis 1

No.	Fit value	Predicted IC ₅₀ (nM)	Observed IC ₅₀ (nM)	Error	Activity scale	Predicted scale
1	9.27	16	32	−1.93	+++	+++
2	8.03	293	231	+1.27	+++	+++
3	7.19	2023	3,160	−1.56	+	+
4	6.72	5945	1,260	+4.72	+	+
5	6.70	6213	1,260	+4.93	+	+
6	7.01	3016	4,500	−1.49	+	+
7	7.89	402	280	+1.44	++	++
8	8.04	284	80	+3.56	+++	++
9	9.63	7.3	39	−5.31	+++	+++
10	10.33	1.5	2.3	−1.56	+++	+++
11	7.15	2207	910	+2.43	+	++
12	6.76	5373	3,100	+1.73	+	+
13	6.99	3176	8,900	−2.80	+	+
14	7.45	1106	810	+1.37	++	+
15	9.03	28	18	+1.61	+++	+++
16	8.74	56	58	−1.03	+++	+++
17	9.39	12.7	6	+2.12	+++	+++
18	7.38	1296	2,700	−2.08	+	+
19	9.48	10.2	7	+1.46	+++	+++
20	9.63	7.3	2	+3.65	+++	+++
21	10.81	0.48	1	−2.07	+++	+++
22	10.48	1.03	5	−4.86	+++	+++
23	10.99	0.32	1	−3.10	+++	+++
24	9.04	28.7	80	−2.79	+++	+++
25	9.19	20.0	46	−2.28	+++	+++
26	9.64	7.1	11	−1.55	+++	+++
27	11.47	0.11	0.4	−3.81	+++	+++
28	10.38	1.31	2.3	−1.75	+++	+++
29	9.80	4.98	4	+1.24	+++	+++
30	10.17	2.11	6.6	−3.13	+++	+++
31	9.50	9.95	2.6	+3.83	+++	+++
32	10.66	0.69	2.5	−3.64	+++	+++
33	9.14	22.83	3.8	+6.01	+++	+++
34	9.74	5.67	3.2	+1.77	+++	+++
35	10.05	2.80	10	−3.58	+++	+++
36	10.06	2.71	8	−2.95	+++	+++
37	9.65	6.99	19	−2.72	+++	+++
38	8.90	39.65	15	+2.64	+++	+++
39	10.81	0.48	1.6	−3.31	+++	+++
40	10.60	0.78	1.1	−1.41	+++	+++
41	10.10	2.46	0.9	+2.73	+++	+++
42	10.19	1.99	1.4	+1.42	+++	+++
43	9.09	25.29	11	+2.30	+++	+++
44	8.81	48.41	44	+1.10	+++	+++
45	10.93	0.37	0.9	−2.45	+++	+++
46	10.67	0.66	1.7	−2.56	+++	+++
47	9.25	17.53	10	+1.75	+++	+++
48	9.97	3.34	4	−1.20	+++	+++
49	9.65	6.93	29	4.19	+++	+++

An overview of the protein sitting in the membrane–aqueous environment is given in Figure 10A, where it is shown that the end of the ligand inserts into a deep pocket of the protein, while the other end extends to the solvent exposed opening of the pocket. The ligand hydrogen bonds to three “hot residues” (Ser38, Gln194, and Arg202) in the active site; A histogram of the frequency of H-bond distances over the trajectory (Figure 10B) shows that these H-bonds are maintained over the entire trajectory with over 80% of the time spent at donor–acceptor

distances around 3 Å. A close up of the binding mode of compound 1 in the hCXCR3 homology model is depicted in Figure 10C. The ligand forms, and there is π – π interaction with Phe37, polar interaction with Ser38, and hydrogen bonds with Gln194 and Arg202. These binding interactions further rationalize the activity of compound 1. This optimized structure (Figure 9C) was then used as the basis for virtual screening of compounds for hCXCR3 antagonist lead identification.

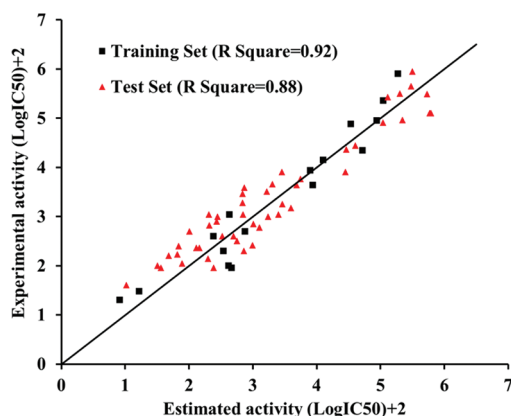


Figure 7. Correlation of the estimated and observed activities for the compounds listed in Tables 1 and 2.

Validating the Binding Mode with Known hCXCR3 Antagonists. In order to calibrate the docking score we exploited the 16 compounds in the training set. These were docked into the binding mode of the optimized hCXCR3 homology model, and the docking scores were compared to the observed IC_{50} values as listed in Table 4. A lower (more negative) docking score correlates with a stronger binding affinity. As shown in Table 4, although as is typically the case with virtual screening, there is quantitative differences between the relative docking scores and the IC_{50} s, in general a cutoff of about -21 can be used pragmatically to select actives. Thus, for example, compounds 15 and 16 are the least active compounds; and they have the highest docking score (-20.6 and -19.7 , respectively).

We then assessed the model and cutoff by docking a test set of 286 known antagonists and 180,952 NCI compounds. Hits were defined by a docking scores less than -21 kcal/mol on the basis of the results in Table 4. Out of the 181,238 compounds, 3580 had docking scores less than -21 kcal/mol. These were comprised of 246 (86%) hits from the 286 known antagonists and 3334 compounds from the NCI database. This corresponds to an enrichment factor, EF, of 44 and a GF of 0.26 (data shown in Table 5). Thus, it is 44 times more probable to pick

active compounds from the database than expected by chance, a necessary condition for the model to satisfy.

Merging Pharmacophore-Based and Structure-Based Approaches. By combining the two virtual screenings (structure-based and ligand-based screenings) results, 273 (95%) of the compounds satisfied either pharmacophore or docking score criteria and would be chosen as hits if we took the union of the two searches. The false negative rates were 15% for the pharmacophore model, 14% for the homology model, and 5% for the combined model. As shown in Table 5, the value of EF (332) and GF (0.58) were highly improved by intersection of the two virtual screenings. Thus, only 418 compounds satisfied both the docking and pharmacophore criteria, and out of these, 219 known actives were recovered. This constitutes a significant reduction (and concomitant cost savings) in the number of compounds that need to be bought or synthesized and assayed.

To further test the validity of the homology model, we carried out docking studies with an additional two docking algorithms/scoring functions to ensure there was no bias introduced by the calibration with MOE. Docking studies were carried out with Ligandfit and Glide. The results are presented in Table 6. As shown in this table, a significant enrichment is obtained with the homology model in all cases, albeit not quite as high as with MOE, which might be expected. This confirms the general validity of the homology model and demonstrates the ability of homology models to be used productively in virtual screening studies.

CONCLUSIONS

A new virtual screening protocol that combines a ligand-based pharmacophore model and a structure-based docking model has been built to virtually screen for hCXCR3 antagonists. The hCXCR3 antagonist binding site was identified by homology modeling and molecular dynamics (MD) refinement. Using 286 known hCXCR3 antagonists and 180,952 NCI compounds to test the protocol, the pharmacophore model produced 243 hits from the known antagonists and 958 hits from NCI compounds, while the homology model identified 246 hits from the known antagonists and 1996 hits from NCI compounds. For the screening of 286 known antagonists, the intersection of the two hit sets had 219 compounds, while the union of the two

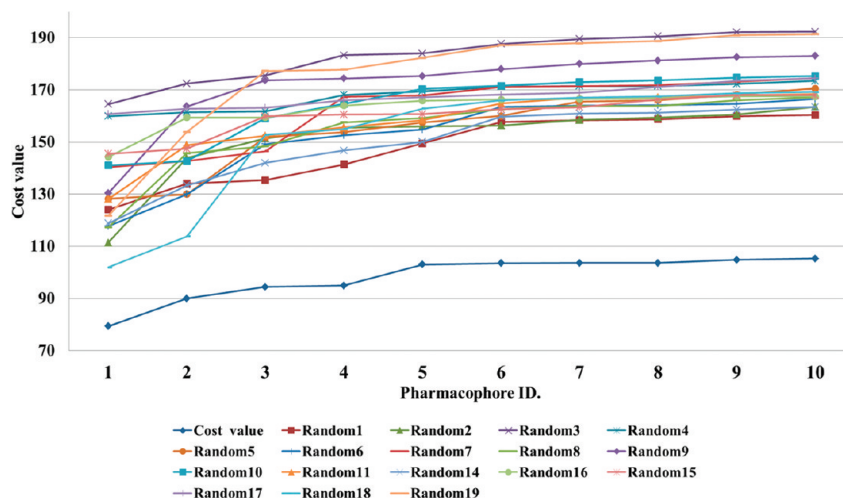


Figure 8. Fisher's randomization test results of pharmacophore generation. Difference cost values of 10 proposed pharmacophore models at the 95% confidence level. The lowest line is the preferred pharmacophore model (hypothesis 1) built as described in Materials and Methods.

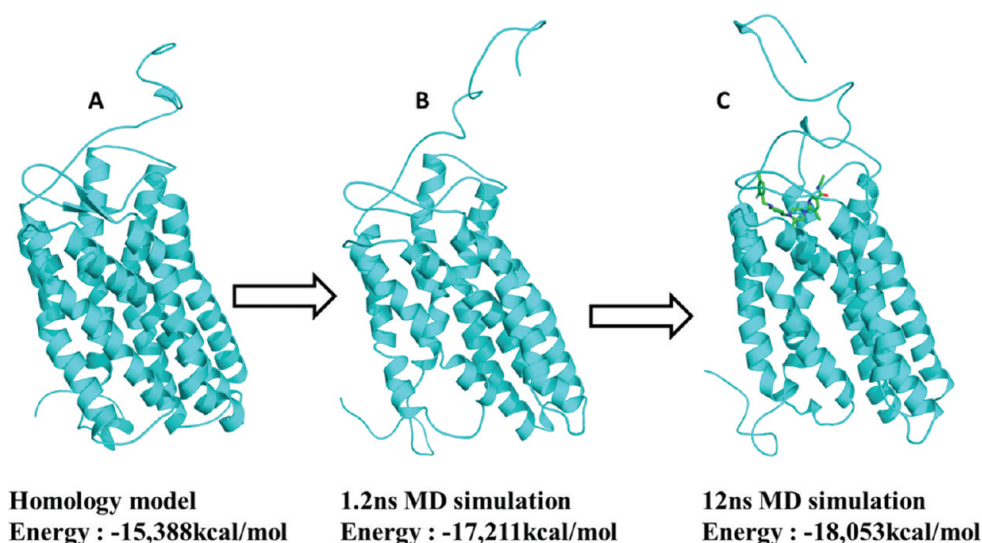


Figure 9. (A) Initial hCXCR3 homology model. (B) Equilibrated hCXCR3 homology model after 1.2 ns MD simulation. (C) hCXCR3–ligand complex (compound 1) after 12 ns.

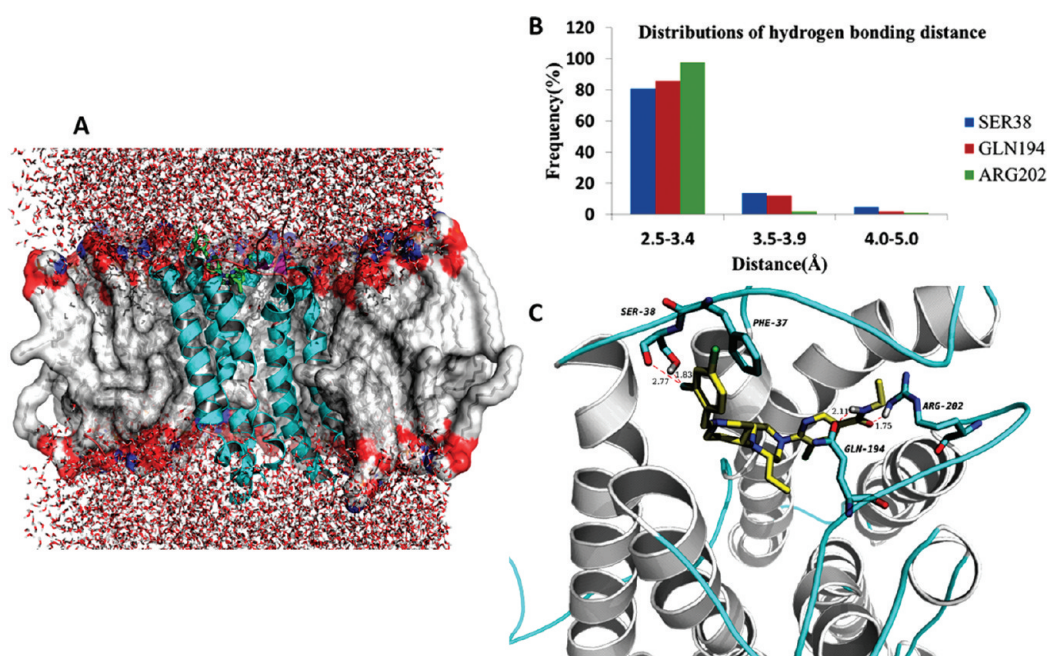


Figure 10. (A) Overview of the binding mode of compound 1 and hCXCR3 with the water and lipid environment. (B) Distributions of hydrogen bonding distance between compound 1 and residues SER38, GLN194, and ARG202 within 12 ns MD simulation. (C) Binding mode of compound 1 in the hCXCR3 homology model.

Table 4. MOE Docking Value of 16 Compounds in the Training Set

Compound No.	1	2	3	4	5	6	7	8
Score	−27.4	−27.5	−26.2	−26.7	−28.5	−24.12	−27.28	−25.5
IC ₅₀ (nM)	0.2	0.3	0.9	1	2	4	5	11
Compound No.	9	10	11	12	13	14	15	16
Score	−27.4	−26.7	−24.5	−30.6	−25.5	−28.3	−20.6	−19.6
IC ₅₀ (nM)	44	87	142	220	770	900	2,310	8,060

hit sets comprised 273 compounds. Therefore, the consistency of the pharmacophore model and the structural binding model is $219/273 = 80\%$. The hit rate for the virtual screening protocol is $273/286 = 95\%$, and the rate of false positives was reduced to 0.06%. By combining the structure-based and

ligand-based screenings results, 95% of the compounds satisfied either pharmacophore or docking score criteria and would be chosen as hits if the union of the two searches was taken. The false negative rates were 15% for pharmacophore model, 14% for homology model, and 5% for combined model.

Table 5. Statistical Parameters from Screening of 286 Known Antagonists and 180,952 NCI Compounds with MOE

Parameters	Pharmacophore model	Binding site model (MOE)	Union set of pharmacophore model + binding site model (MOE)	Intersection set of pharmacophore model + binding site model (MOE)
Total number of actives in database (A)	286	286	286	286
Total number of active molecules in hit list (Ha)	243	246	273	219
% Yield of actives [(Ha/Ht) × 100]	20.23	6.87	6.26	52.39
Total number of molecules in database (D)	181,238	181,238	181,238	181,238
Total number of hit molecules from the database (Ht)	1201	3580	4360	418
Total inactive molecules in hit list (Hi)	958	3334	4087	199
% Ratio of actives [(Ha/A) × 100]	84.97	86.0	95.5	76.6
Enrichment factor (EF)	128.16	43.5	39.7	332
% Ratio of false negatives [(A - Ha)/A × 100]	15.03	14.0	4.55	23.4
% Ratio of false positives [(Ht - Ha)/(D - A) × 100]	0.53	1.84	2.26	0.11
Goodness of fit score (GF)	0.36	0.26	0.28	0.58

Table 6. Confirmation of Homology Model by Docking with Glide and LigandFit

Parameters	Binding site model (Glide)	Binding site model (LigandFit)
Total number of actives in database (A)	286	286
Total number of active molecules in hit list (Ha)	75	62
% Yield of actives [(Ha/Ht) × 100]	2.09	1.73
Total number of molecules in database (D)	181,238	181,238
Total number of hit molecules from the database (Ht)	3580	3580
Total inactive molecules in hit list (Hi)	3505	3518
% Ratio of actives [(Ha/A) × 100]	26.2	21.7
Enrichment factor (EF)	13.3	11.0
% Ratio of false negatives [(A - Ha)/A × 100]	73.8	78.3
% Ratio of false positives [(Ht - Ha)/(D - A) × 100]	1.94	1.94
Goodness of fit score (GF)	0.08	0.07

This work demonstrated that the quality of both the pharmacophore model and homology model can be measured by the consistency of the two models, and false negatives in virtual screening can be reduced by combining two virtual screening approaches. More importantly perhaps, the intersection of the hits from the pharmacophore and docking methods identified 219 or 76% of the actives. This result enable us to reduce the number of compounds that must be submitted for expensive experimental assays from a virtual screen.

■ ASSOCIATED CONTENT

■ Supporting Information

Chemical structures used for model building and pdb file for the hCXCR3 homology model. This material is available free of charge via the Internet at <http://pubs.acs.org>.

■ AUTHOR INFORMATION

Corresponding Author

*E-mail: junxu@biochemomes.com; xujun9@mail.sysu.edu.cn.

Notes

The authors declare no competing financial interest.

■ ACKNOWLEDGMENTS

This work was funded in part of the National Science and Technology Major Project of China (2010ZX09102-305), the National High-tech R&D Program of China (863 Program) (2012AA020307), Guangdong Recruitment Program of Creative Research Groups, the National Natural Science Foundation of China (No. 81001372, 81173470).

■ REFERENCES

- (1) Gerard, C.; Rollins, B. J. Chemokines and disease. *Nat. Immunol.* **2001**, *2*, 108–115.
- (2) Charo, I. F.; Ransohoff, R. M. The many roles of chemokines and chemokine receptors in inflammation. *N. Engl. J. Med.* **2006**, *354*, 610–621.
- (3) Zeremski, M.; Dimova, R.; Brown, Q.; Jacobson, I. M.; Markatou, M.; Talal, A. H. Peripheral CXCR3-associated chemokines as biomarkers of fibrosis in chronic hepatitis C virus infection. *J. Infect. Dis.* **2009**, *200*, 1774–1780.
- (4) Wareing, M. D.; Lyon, A. B.; Lu, B.; Gerard, C.; Sarawar, S. R. Chemokine expression during the development and resolution of a pulmonary leukocyte response to influenza A virus infection in mice. *J. Leukocyte Biol.* **2004**, *76*, 886–895.
- (5) Clark-Lewis, I.; Mattioli, I.; Gong, J. H.; Loetscher, P. Structure–function relationship between the human chemokine receptor CXCR3 and its ligands. *J. Biol. Chem.* **2003**, *278*, 289–295.
- (6) Vola, A.; Luster, A. D. Chemokines and their receptors: Drug targets in immunity and inflammation. *Annu. Rev. Pharmacol. Toxicol.* **2008**, *48*, 171–197.
- (7) Huan, C.; Gui-Ying, Z. Advances in chemokine receptor CXCR3 research. *Pract. Prev. Med.* **2010**, *17*, 1466–1468.
- (8) Knight, R. L.; Allen, D. R.; Birch, H. L.; Chapman, G. A.; Galvin, F. C.; Jopling, L. A.; Lock, C. J.; Meissner, J. W.; Owen, D. A.; Raphy, G.; Watson, R. J.; Williams, S. C. Development of CXCR3 antagonists. Part 4: Discovery of 2-amino-(4-tropinyl)quinolines. *Bioorg. Med. Chem. Lett.* **2008**, *18*, 629–633.
- (9) Watson, R. J.; Allen, D. R.; Birch, H. L.; Chapman, G. A.; Galvin, F. C.; Jopling, L. A.; Knight, R. L.; Meier, D.; Oliver, K.; Meissner, J.

- W.; Owen, D. A.; Thomas, E. J.; Tremayne, N.; Williams, S. C. Development of CXCR3 antagonists. Part 3: Tropenyl and homotropenyl-piperidine urea derivatives. *Bioorg. Med. Chem. Lett.* **2008**, *18*, 147–151.
- (10) Watson, R. J.; Allen, D. R.; Birch, H. L.; Chapman, G. A.; Hannah, D. R.; Knight, R. L.; Meissner, J. W.; Owen, D. A.; Thomas, E. J. Development of CXCR3 antagonists. Part 2: Identification of 2-amino(4-piperidinyl)azoles as potent CXCR3 antagonists. *Bioorg. Med. Chem. Lett.* **2007**, *17*, 6806–6810.
- (11) Shao, Y.; Anilkumar, G. N.; Carroll, C. D.; Dong, G.; Hall, J. W., 3rd; Hobbs, D. W.; Jiang, Y.; Jenh, C. H.; Kim, S. H.; Kozlowski, J. A.; McGuinness, B. F.; Rosenblum, S. B.; Schulman, I.; Shih, N. Y.; Shu, Y.; Wong, M. K.; Yu, W.; Zawacki, L. G.; Zeng, Q., II. SAR studies of pyridyl-piperazinyl-piperidine derivatives as CXCR3 chemokine antagonists. *Bioorg. Med. Chem. Lett.* **2011**, *21*, 1527–1531.
- (12) Du, X. H.; Gustin, D. J.; Chen, X. Q.; Duquette, J.; Mcgee, L. R.; Wang, Z. L.; Ebsworth, K.; Henne, K.; Lemon, B.; Ma, J.; Miao, S.; Sabalan, E.; Sullivan, T. J.; Tonn, G.; Collins, T. L.; Medina, J. C. Imidazo-pyrazine derivatives as potent CXCR3 antagonists. *Bioorg. Med. Chem. Lett.* **2009**, *19*, S200–S204.
- (13) Wang, Y. H.; Busch-Petersen, J.; Wang, F.; Kiesow, T. J.; Graybill, T. L.; Jin, J.; Yang, Z.; Foley, J. J.; Hunsberger, G. E.; Schmidt, D. B.; Sarau, H. M.; Capper-Spudich, E. A.; Wu, Z. N.; Fisher, L. S.; McQueney, M. S.; Rivero, R. A.; Widdowson, K. L. Camphor sulfonamide derivatives as novel, potent and selective CXCR3 antagonists. *Bioorg. Med. Chem. Lett.* **2009**, *19*, 114–118.
- (14) Bongartz, J. P.; Buntinx, M.; Coesemans, E.; Hermans, B.; Lommen, G. V.; Wauwe, J. V. Synthesis and structure-activity relationship of benzetimide derivatives as human CXCR3 antagonists. *Bioorg. Med. Chem. Lett.* **2008**, *18*, 5819–5823.
- (15) Thoma, G.; Baenteli, R.; Lewis, I.; Wagner, T.; Oberer, L.; Blum, W.; Glickman, F.; Streiff, M. B.; Zerwes, H. G. Special ergolines are highly selective, potent antagonists of the chemokine receptor CXCR3: Discovery, characterization and preliminary SAR of a promising lead. *Bioorg. Med. Chem. Lett.* **2009**, *19*, 6185–6188.
- (16) Afantitis, A.; Melagraki, G.; Sarimveis, H.; Igglessi-Markopoulou, O.; Kollias, G. A novel QSAR model for predicting the inhibition of CXCR3 receptor by 4-N-aryl-[1,4] diazepane ureas. *Eur. J. Med. Chem.* **2009**, *44*, 877–884.
- (17) Afantitis, A.; Melagraki, G.; Sarimveis, H.; Koutentis, P. A.; Igglessi-Markopoulou, O.; Kollias, G. A combined LS-SVM & MLR QSAR workflow for predicting the inhibition of CXCR3 receptor by quinazolinone analogs. *Mol. Diversity* **2010**, *14*, 225–235.
- (18) *Discovery Studio*, version 2.1; Accelrys, Inc.: San Diego, CA, 2007.
- (19) Wu, B.; Chien, E. Y.; Mol, C. D.; Fenalti, G.; Liu, W.; Katritch, V.; Abagyan, R.; Brooun, A.; Wells, P.; Bi, F. C.; Hamel, D. J.; Kuhn, P.; Handel, T. M.; Cherezov, V.; Stevens, R. C. Structures of the CXCR4 chemokine GPCR with small-molecule and cyclic peptide antagonists. *Science* **2010**, *330*, 1066–1071.
- (20) Gao, J. M.; Xiang, R. L.; Jiang, L.; Li, W. H.; Feng, Q. P.; Guo, Z. J.; Sun, Q.; Zeng, Z. P.; Fang, F. D. Sulfated tyrosines 27 and 29 in the N-terminus of human CXCR3 participate in binding native IP-10. *Acta Pharmacol. Sin.* **2009**, *30*, 193–201.
- (21) Bowers, K. J.; Chow, E.; Xu, H.; Dror, R. O.; Eastwood, M. P.; Gregersen, B. A.; Klepeis, J. L.; Kolossvary, I.; Moraes, M. A.; Sacerdoti, F. D.; Salmon, J. K.; Shan, Y.; Shaw, D. E. *Scalable Algorithms for Molecular Dynamics Simulations on Commodity Clusters*. In Proceedings of the 2006 ACM/IEEE Conference on Supercomputing, ACM: Tampa, FL, 2006.
- (22) Darden, T.; York, D.; Pedersen, L. Particle mesh Ewald: An N [center-dot] log(N) method for Ewald sums in large systems. *J. Chem. Phys.* **1993**, *98*, 10089–10092.
- (23) Krautler, V.; Van Gunsteren, W. F.; Hunenberger, P. H. A fast SHAKE: Algorithm to solve distance constraint equations for small molecules in molecular dynamics simulations. *J. Comput. Chem.* **2001**, *22*, 501–508.
- (24) Tuckerman, M.; Berne, B. J.; Martyna, G. J. Reversible multiple time scale molecular dynamics. *J. Chem. Phys.* **1992**, *97*, 1990–2001.
- (25) Du, X.; Chen, X.; Mihalic, J. T.; Deignan, J.; Duquette, J.; Li, A. R.; Lemon, B.; Ma, J.; Miao, S.; Ebsworth, K.; Sullivan, T. J.; Tonn, G.; Collins, T. L.; Medina, J. C. Design and optimization of imidazole derivatives as potent CXCR3 antagonists. *Bioorg. Med. Chem. Lett.* **2008**, *18*, 608–613.
- (26) Liu, J.; Fu, Z.; Li, A. R.; Johnson, M.; Zhu, L.; Marcus, A.; Danao, J.; Sullivan, T.; Tonn, G.; Collins, T.; Medina, J. Optimization of a series of quinazolinone-derived antagonists of CXCR3. *Bioorg. Med. Chem. Lett.* **2009**, *19*, 5114–5118.
- (27) Crosignani, S.; Missotten, M.; Cleva, C.; Dondi, R.; Ratinaud, Y.; Humbert, Y.; Mandal, A. B.; Bombrun, A.; Power, C.; Chollet, A.; Proudfoot, A. Discovery of a novel series of CXCR3 antagonists. *Bioorg. Med. Chem. Lett.* **2010**, *20*, 3614–3617.
- (28) Johnson, M.; Li, A. R.; Liu, J.; Fu, Z.; Zhu, L.; Miao, S.; Wang, X.; Xu, Q.; Huang, A.; Marcus, A.; Xu, F.; Ebsworth, K.; Sablan, E.; Danao, J.; Kumer, J.; Dairaghi, D.; Lawrence, C.; Sullivan, T.; Tonn, G.; Schall, T.; Collins, T.; Medina, J. Discovery and optimization of a series of quinazolinone-derived antagonists of CXCR3. *Bioorg. Med. Chem. Lett.* **2007**, *17*, 3339–3343.
- (29) Cole, A. G.; Stroke, I. L.; Brescia, M. R.; Simhadri, S.; Zhang, J. J.; Hussain, Z.; Snider, M.; Haskell, C.; Ribeiro, S.; Appell, K. C.; Henderson, I.; Webb, M. L. Identification and initial evaluation of 4-N-aryl-[1,4]diazepane ureas as potent CXCR3 antagonists. *Bioorg. Med. Chem. Lett.* **2006**, *16*, 200–203.
- (30) Li, A. R.; Johnson, M. G.; Liu, J.; Chen, X.; Du, X.; Mihalic, J. T.; Deignan, J.; Gustin, D. J.; Duquette, J.; Fu, Z.; Zhu, L.; Marcus, A. P.; Bergeron, P.; McGee, L. R.; Danao, J.; Lemon, B.; Carabeo, T.; Sullivan, T.; Ma, J.; Tang, L.; Tonn, G.; Collins, T. L.; Medina, J. C. Optimization of the heterocyclic core of the quinazolinone-derived CXCR3 antagonists. *Bioorg. Med. Chem. Lett.* **2008**, *18*, 688–693.
- (31) McGuinness, B. F.; Carroll, C. D.; Zawacki, L. G.; Dong, G.; Yang, C.; Hobbs, D. W.; Jacob-Samuel, B.; Hall, J. W., 3rd; Jenh, C. H.; Kozlowski, J. A.; Anilkumar, G. N.; Rosenblum, S. B. Novel CXCR3 antagonists with a piperazinyl-piperidine core. *Bioorg. Med. Chem. Lett.* **2009**, *19*, S205–S208.
- (32) *Glide*, v5.7; Schrödinger, LLC: New York, 2011.
- (33) *Manual of Molecular Operating Environment (MOE)*, version 2010; Chemical Computing Group, Inc.: Montreal, Quebec, Canada, 2010.
- (34) Zhao, W.; Gu, Q.; Wang, L.; Ge, H.; Li, J.; Xu, J. Three-dimensional pharmacophore modeling of liver-X receptor agonists. *J. Chem. Inf. Model.* **2011**, *51*, 2147–2155.
- (35) John, S.; Thangapandian, S.; Sakikah, S.; Lee, K. W. Potent bace-1 inhibitor design using pharmacophore modeling, in silico screening and molecular docking studies. *BMC Bioinf.* **2011**, *12*.
- (36) Güner, O. F. *Pharmacophore Perception, Development, and Use in Drug Design*; International University Line: LaJolla, CA, 2000.
- (37) Luthy, R.; McLachlan, A. D.; Eisenberg, D. Secondary structure-based profiles: use of structure-conserving scoring tables in searching protein sequence databases for structural similarities. *Proteins* **1991**, *10*, 229–239.
- (38) Luthy, R.; Bowie, J. U.; Eisenberg, D. Assessment of protein models with three-dimensional profiles. *Nature* **1992**, *356*, 83–85.
- (39) Kyte, J.; Doolittle, R. F. A simple method for displaying the hydrophobic character of a protein. *J. Mol. Biol.* **1982**, *157*, 105–132.