Fingerprint Directed Scaffold Hopping for Identification of CCR2 Antagonists

Pramod C. Nair and M. Elizabeth Sobhia*

Centre for Pharmacoinformatics, National Institute of Pharmaceutical Education and Research (NIPER), Sector 67, SAS Nagar, Punjab-160062, India

Received May 7, 2008

Chemokine receptors have evolved as attractive targets for disease conditions which arise due to immunomodulation involving host-defense mechanisms. CCR2, a chemokine receptor, is targeted for diseases like arthritis, multiple sclerosis, vascular disease, obesity, and type 2 diabetes. This study provides a new strategy of a ligand based technique which exploits fingerprint led fragment features in conjunction with structure-guided design for identifying new scaffolds for CCR2. A fragment based mining (FBM) technique was employed on a chemical database to identify novel scaffold hops. The hits were subjected to 3-point pharmacophore fingerprint procedures with Tanimoto similarity metric to compare pharmacophoric fingerprints. The final 66 hits generated by these exercises were predicted by the validated HQSAR model, and the top predicted were suggested as probable scaffolds for CCR2 antagonism. The identified scaffolds were validated through molecular docking studies. The ligands were docked by providing receptor flexibility in the extra cellular domain (1 and 3), N terminal domain, and in the transmembrane (TM1 & TM7) helix region with IFD approach. Some of the scaffolds showed H-bonding potential which was not explored by the data set molecules. All identified scaffolds highlighted a key hydrogen bonding interaction with Thr292 as supported by mutational studies. The observed pi stacking interaction with Tyr188 in data set molecules was also produced by the new scaffolds. Taking the advantage of receptor flexibility the scaffolds explored the hydrophobic binding cleft between helix 1 and 7 occupied by residues Leu44, Leu45, Leu48 and Ile300, Ile303, Ile304, respectively. Two of the identified molecules have promising outcomes and can be considered as novel scaffolds for CCR2 binding.

1. INTRODUCTION

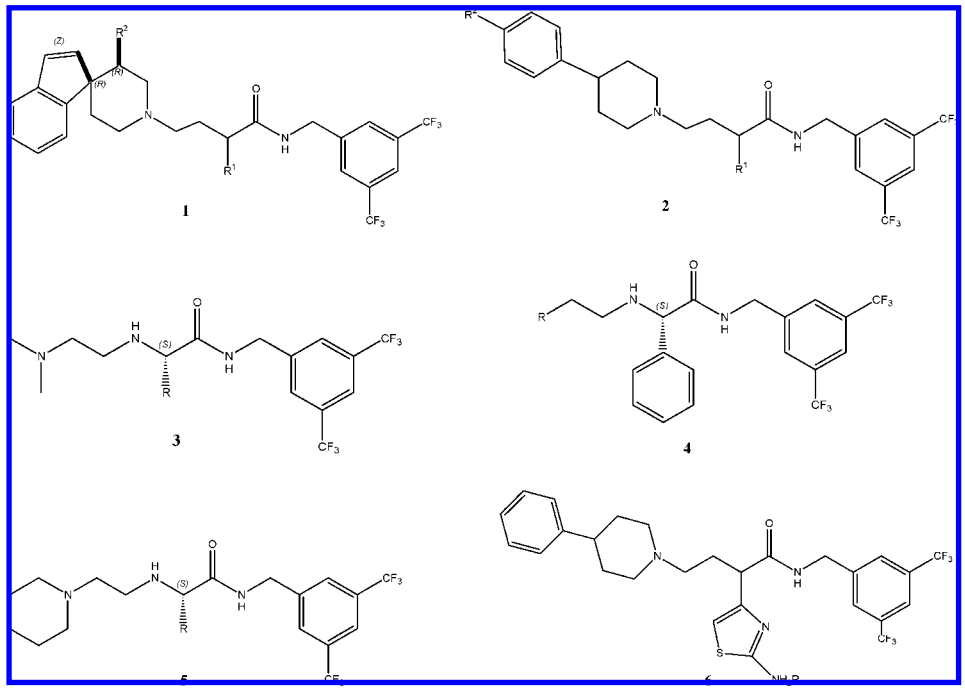
Chemokines play a key role in immunomodulation and host-defense mechanisms. These are small molecular weight water-soluble proteins which selectively recruit monocytes, neutrophils, and lymphocytes to sites of vascular injury and inflammation.[1−3] Human chemokines are classified into four families based on the differences in their structure and role. CC chemokines are the largest family of chemokines and are named so because the first two of four cysteine residues in these molecules lie one beside the other.[2] The mononuclear cells get attracted to sites of chronic inflammation by CC chemokines.[1−3] Monocyte chemoattractant protein-1 (MCP-1) is the most thoroughly distinguished CC chemokine and termed as chemokine ligand CCL2 in the regular nomenclature. Impairment of monocytic trafficking in inflammation models can be achieved to a significant extent by the loss of MCP-1 effector function. CCR2a and CCR2b are two isoforms of the monocyte chemoattractant protein-1 receptor formed by alternative splicing which differ in the C terminal tail.[4] CCR2b is a longer variant, and its targeting has proved to be helpful; several molecules are in phase 1 and 2 clinical trials for varied diseased conditions such as arthritis, multiple sclerosis, vascular disease, obesity, and type 2 diabetes.[2,4−6]

Ligand-based approaches are valuable when the target information is not available, and in such instances QSAR, pharmacophore mapping, and similarity search methods are proved to be useful. QSAR techniques are more often used as a tool for lead optimization within the congeneric domain of molecules; however, there are also reports found for its applicability in diverse chemical space to identify new leads. Success stories are reported for identifying lead hops using QSAR based chemical database mining approaches.[7−11] Shen et al. reported the applicability of predictive QSAR models for identification of new leads for functionalized amino acid (FAA) for their anticonvulsant effects, and such approaches can be considered as a rational drug discovery strategy.[8] Zhang and Muegge have reported scaffold hopping for 7 different biological targets. They have performed similarity based virtual screening using 2D and 3D descriptors.[10] Pharmacophore and similarity based methods are well suited when the target structure remains unsolved.[12−17] There is strong literature evidence suggesting that the lead identification can be achieved by pharmacophore based methods.[18−21] These methods involve identification of features from a set of compounds which act on the same target, whereas similarity based methods go with the fact that compounds with similar structural properties such as descriptors or fingerprints may own similar biological activity. Several pharmacophore feature generation methods for virtual screening are reported in the literature, like the 3D conformation based systematic search methods, clique detection using chemical graphs and distance constraints, maximum likelihood, and genetic algorithms approaches.[18−20] Molecular similarity methods also prove to be equally significant in lead identification using several approaches like chemical graphs, physicochemical properties, topological indices, 3D

* Corresponding author e-mail: mesophia@niper.ac.in.

10.1021/ci800157j CCC: $40.75
Published on Web 09/03/2008

Chart 1. Representative Chemical Scaffolds of a Data Set as CCR2 Antagonists



pharmacophore patterns, and molecular fields.[12−17] As far as the computational fingerprint-based similarity methods are concerned the Tanimoto coefficient remains to be the method of choice compared with other similarity methods.[13,21,22] Extensive virtual screening studies have been reported by several groups. Among them many have reported 2D fingerprinting approaches based on the Tanimoto coefficient as one of the efficient ways to identify similar molecules with resembling activities.[21] However there are groups who did report the biasness involving this approach.[23,24] Recent advances in similarity searches like the introduction of the group fusion technique such as turbo similarity searches have improved the scaffold-hopping strategies.[25] A report by Matter and others suggests that 2D fingerprints alone as a primary descriptor or in combination with other metrics are applicable for a general library design for the lead discovery encompassing the global diversity.[26] Also, these fingerprints are efficient in selecting representative subsets of bioactive compounds from a diverse set of compounds.[26] In our previous attempts we carried out fragment based HQSAR studies on a diverse class of CCR2 antagonists which provided critical information regarding fragment features required for CCR2 antagonism.[27,28] In the present study we have attempted to arrive at new scaffolds using a combinatorial approach in which fingerprint led fragment features are utilized in conjunction with structure guided analysis. The absence of X-ray structural information of the target prompted us to use a combinatorial approach which involves tools like HQSAR, fragment based mining, 3-point pharmacophoric fingerprinting, and similarity search. In addition we also validated the productivity of the obtained scaffolds using the modeled structure.

2. DATA SET

The data set of CCR2 antagonists reported by Merck Laboratories was used for the study. This includes a chemical series comprising 4-amino-2-alkyl-butyramides, 3,5-bis(trifluoromethyl)benzyl L-arylglycinamide, and α-aminothiazole-γ-aminobutanoic amides.[29−32]The CCR2 binding data on human monocytes or CHO cells expressing human CCR2b (hCCR2b) were considered for hologram QSAR modeling studies.[29−32] The data set used for analysis includes binding values (IC_{50}) ranging from 1000 nM to 0.43 nM. The data set compounds contain an amide and a bismethyl trifluorophenyl group in common. Percentage data of the racemates were ignored for QSAR analysis. The IC_{50} (μM) values were taken in molar (M) range and converted to pIC_{50} according to the formula

$$pIC_{50} = -\log IC_{50} \quad (1)$$

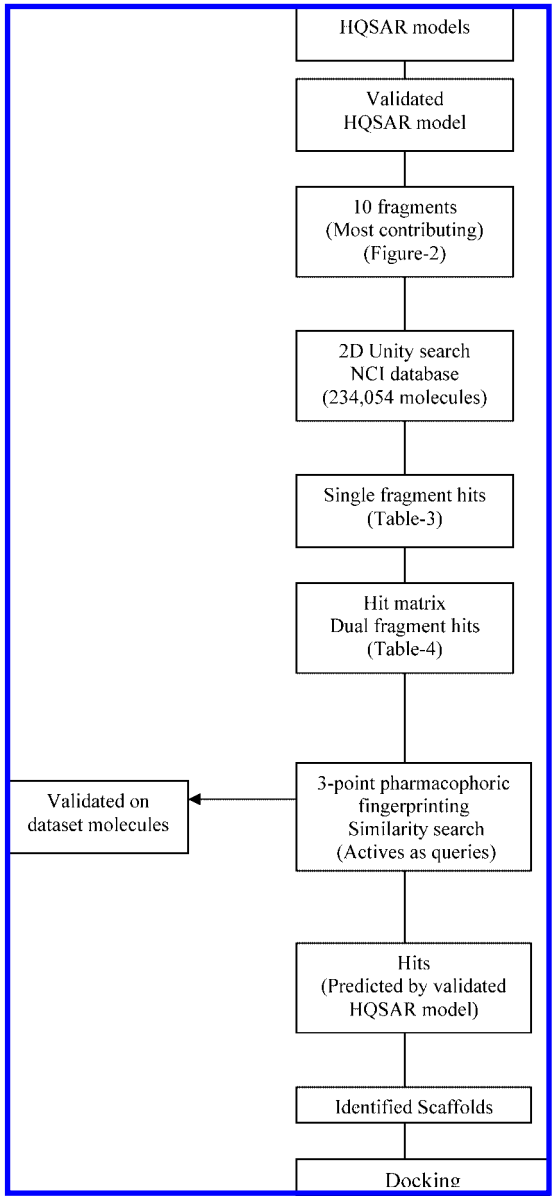
The data set with 36 molecules was randomly segregated into a training and a test set comprising 29 and 7 molecules, respectively (see Supporting Information S-Table 1a-1f). Chart 1 shows the representative scaffolds of the data set molecules used in the hologram analysis.

3. COMPUTATIONAL DETAILS

Scheme 1 shows various steps involved in HQSAR directed scaffold hopping. The first three steps of Scheme 1 show the HQSAR model development, validation, and the selection of the 10 most contributing fragments. The fourth step shows fragment based mining (FBM) using the NCI database. The fifth and the sixth steps describe the single and dual fragment hits obtained from the mining exercise. Step seven shows 3-point pharmacophore fingerprinting coupled with Tanimoto similarity performed on dual fragment hits. Steps eight and nine show screened scaffolds as predicted by the validated HQSAR model. The final step involves docking studies performed on identified hits.

3.1. HQSAR. Molecular modeling studies were performed using the molecular modeling package SYBYL7.1[33] installed on a Silicon Graphics Fuel Workstation. For QSAR analysis the structures were sketched and minimized individually by using Powell's conjugate gradient method.[34] Hologram

Scheme 1. Summary of the HQSAR Directed Scaffold Hopping Studies for Search of CCR2 Antagonists



QSAR (HQSAR) employs fragment fingerprints as predictive variables of biological activity or other structural related data. Fragments with different atom counts were generated, and fragments with 4 to 7 atoms were hashed into bins 1 to 85 of the fingerprint.[35] These molecular fingerprints were broken into strings at fixed interval as specified by a hologram length (HL) parameter. The hologram length determines the number of bins in the hologram into which the fragments are hashed. Each corresponding fragment SLN is mapped to a pseudo-random integer in the range of 0−2 using a cyclic redundancy check (CRC) algorithm.[33,35] The integer generated by the CRC algorithm is unique and reproducible for each and every unique SLN string. The hashing then occurs by folding the pseudorandom integer for a particular SLN string into the bin range defined. The bins contain the information about the number of fragments hashed into the bin. The optimal HQSAR model was derived from screening through the 12 default HL values, which were a set of 12 prime numbers ranging from 53 to 401. The model development was performed using parameters viz., atoms, bonds, connections, chirality, hydrogens, and donor/acceptor. The validity of the model depends on statistical parameters such as non-cross-validated r^2_{ncv}, cross-validated r^2_{cv} by Leave-One-Out (LOO), predictive r^2_{pred}, and standard error. The robustness of the model depends on the more challenging external test prediction reflected by its predictive r^2_{pred} value. Like r^2_{cv}, the predictive r^2_{pred} can assume a negative value reflecting a complete lack of predictive ability of the training set.[36] The predictive r^2_{pred} is defined as

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

where SD is the sum of the squared deviations between the biological activity of molecules in the test set and the mean biological activity of the training set molecules. The PRESS is the sum of the squared deviations between predicted and actual activity values for every molecule in the test set.

3.2. Database Mining. 2D search was employed for NCI database mining and for the dual fragment search within individual hit database.[37] Fragments are a group of atoms where there is a possibility to reach any atom from any other atom along a bonded path. In 2D searching, the query is defined by specific atoms and bonds. Parameters like isotope, charge, and stereochemistry were considered for the search. The database mining was employed using UNITY 2D search available with Sybyl7.1.[33]

3.3. 3-Point Pharmacophoric Fingerprinting and Similarity Search. Pharmacophore and property based fingerprinting are more valuable in lead hopping strategy compared with the simple structural feature fingerprinting like MACCS.[10,38] In the present study fingerprinting was achieved with the typed graph triangle (TGT) fingerprint approach, which is a 3-point pharmacophore fingerprint calculated from a 2D molecular graph. In 3-point pharmacophore fingerprinting each atom is given a type in the set {Donor, Acceptor, Polar, Anion, Cation, Hydrophobe}.[39] Then all triplets are coded as features using the three graph distances and as three atom types of each triplet (Typed Graph Triangle). The fingerprint is the set of all tuples of the form (*u*,*v*,*w*,*d*,*e*,*f*), where *u*, *v*, and *w* are atom types and *d*, *e*, and *f* are graph distances between the atoms.[39] The graph distance is defined as the number of bonds in the shortest path between the atoms in the chemical graph. The Tanimoto similarity metric based on the Tanimoto coefficient was used to compare fingerprints involving different queries to get hits with 85% similarity. The 3-point pharmacophoric fingerprinting and the similarity search results were obtained using the Molecular Operating Environment package.[39]

3.4. Molecular Docking. Flexible ligand docking methods like FlexX and Glide failed to rank the molecules as per the experimental binding activity.[40,41] We performed Induced Fit Docking (IFD) to address the issue of loop mobility. An Induced Fit Docking (IFD) technique reported recently combines rigid-receptor docking with protein refinement.[42,43] This technique considers protein flexibility while docking possible ligands. The protein 1KP1 was prepared according to the recommended protein preparation procedure using the Maestro software package. The data set antagonists and the identified scaffolds were optimized using LigPrep using the OPLS force field. Glide SP (standard precision) was used for all docking calculations.[41] The first stage of the IFD protocol performs an initial softened-potential docking of the ligands to the rigid receptor, with van der Waals radii scaling of 0.5 for both proteins and the ligands. Sampling of the

Table 1. Statistical Parameters Obtained for Different HQSAR Parameters[a]

	A	B	Cc	H	C	D&A	r^2_{ncv}	SE	BHL	r^2_{cv}	NOC
Model-1	x	x	x				0.661	0.494	97	0.467	3
Model-2	x	x	x	x			0.617	0.525	199	0.315	3
Model-3	**x**	**x**	**x**		**x**		**0.934**	**0.233**	**151**	**0.686**	**6**
Model-4	x	x	x			x	0.605	0.513	151	0.483	1
Model-5	x	x	x	x	x	x	0.976	0.142	353	0.765	9

[a] r^2_{ncv} = non-cross-validated correlation coefficient; SE = standard error; BHL = best hologram length; r^2_{cv} = cross-validated correlation coefficient; NOC = number of optimum components; *A* = atoms; *B* = bond; Cc = connections; *H* = hydrogen atoms; *C* = chirality; D&A = donor and acceptor. The model chosen for analysis is highlighted in bold font.

protein for each of the top 20 ligand poses (ranked by GlideScore) was performed using Prime. Residues within 5 Å of any ligand pose were refined; this consisted of a side-chain conformational search and optimization, followed by full minimization of the residues and the ligand. Complexes within 30.0 kcal/mol of the minimum energy structure were forwarded for redocking. The related ligand was redocked into each low-energy, induced-fit structure with default Glide settings (van der Waals radii scaling of 1.0 for protein and 0.8 for the ligand). The final ranking of the ligand is performed with Glide and E-model score (kcal/mol).

4. RESULTS AND DISCUSSION

4.1. HQSAR Model Development and Validation. The first criterion for model development was to distinguish molecular fragments based on their atoms, bonds, connections, chirality, hydrogens, and donor/acceptor. The atom parameters enable fragment determination based on elemental atom types, while the bonds and connections consider the bond orders and hybridization states within fragments, respectively. The default parameters namely atoms, bonds, and connections were tried first for the model development which gave a poor r^2_{ncv} of 0.661 and a standard error (SE) of 0.494 (Table 1). Inclusion of a chirality feature which distinguishes the molecule based on its atomic and bond stereochemistry showed a remarkable thrust toward the r^2_{ncv} with a value of 0.934, cross-validated r^2_{cv} of 0.686, and limiting SE to 0.233. Since the chirality parameter distinguishes the R-enantiomers from the S counterparts at the chiral center, it recounted well in the case of enantiopure molecules under the current study. Other parameters excluding chirality were also tried in combination with the three default parameters, but the statistical results were not significant suggesting the poor quality of models (Table 1). Thus, the model with chirality and three default parameters was considered for further analysis (Model-3). The second criterion was to check the statistical parameters of the models based on the fragment size. The fragment size refers to the minimum and maximum length of the fragment in a hologram fingerprint. The statistical parameters obtained for Model-3 with different fragment size is shown in Table 2. There was no significant difference noticed in the statistical value obtained for the atom count 4−7 and 5−8, with the hologram (151) and number of components (6). Further increase in the atom size reduced the overall statistical power of the model. So, the model developed with atom count of 4−7 was chosen for model development. The model which proceeded for further analysis contains essential parameters namely atoms, bonds, connections, chirality, and atom count 4−7. It is noteworthy that in the present study all 7 molecules of the test set were predicted well by the model with a notable predictive r^2_{pred} of 0.917. The plot of actual versus predicted pIC_{50} of the training and test set molecules is shown in Figure 1.

Table 2. Statistical Parameters Obtained for Model-3 with Different Atom Counts[a]

atom count	1 to 4	2 to 5	3 to 6	**4 to 7**	5 to 8	6 to 9
r^2_{cv}	0.678	0.662	0.646	**0.686**	0.703	0.708
r^2_{ncv}	0.947	0.943	0.923	**0.934**	0.940	0.888
SE	0.229	0.222	0.252	**0.233**	0.222	0.296
BHL	257	307	199	**151**	151	199
NOC	10	7	6	**6**	6	5

[a] r^2_{cv} = cross-validated correlation coefficient; r^2_{ncv} = non-cross-validated correlation coefficient; SE = standard error; BHL= best hologram length; NOC = number of optimum components. The best model is highlighted in bold font.

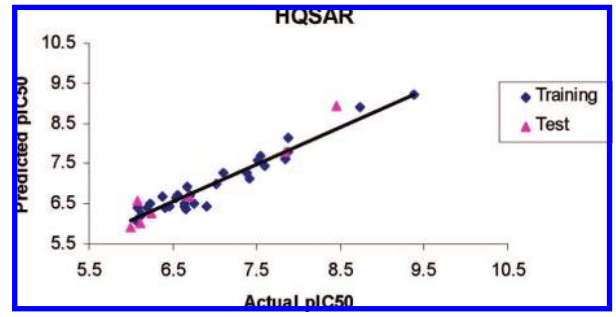


Figure 1. Plot of actual versus predicted pIC_{50} of the training and test set molecules by HQSAR analysis. The training set and test set molecules are shown in blue (diamond) and pink (triangle) spots, respectively.

4.2. Fragment Selection and Database Mining. The outline of the entire exercise performed in the HQSAR directed scaffold hopping study is shown in Scheme 1. The various steps performed in this exercise are elaborated in computational details (section 3). In HQSAR, the molecules are sliced into smaller fragments which provide critical information about the activity of the compounds. Such information may aid in hypothesizing the influence of fragments on CCR2 antagonistic activity; considering this, molecular fragments were employed for database mining. The top 10 fragments having positive coefficients reciprocating favorable contribution toward CCR2 antagonism were selected for database searching (Figure 2). The Frag 4, Frag 8, and Frag 9 reflect the presence of a Nam atom. It is noted in the data set that all the molecules contain an amide group, and presumably the presence of an amide group forms one of the important pharmacophoric features for CCR2 antagonism. Frag 3 represents the importance of 5 membered spiro ring systems noted in the case of the butyramides class. The importance of thiophene and thiazole containing ring systems are depicted in Frag 6, Frag7, and Frag 8 (Figure 2). The aminothiazole class of compounds being the most potent analogs with a thiazole moiety is likely to account for improved potency over other analogs. The National Cancer Institute (NCI) database consisting of 234,054 molecules was exploited for the mining purpose.[37] In the present study, molecules with two and more than two fragment properties

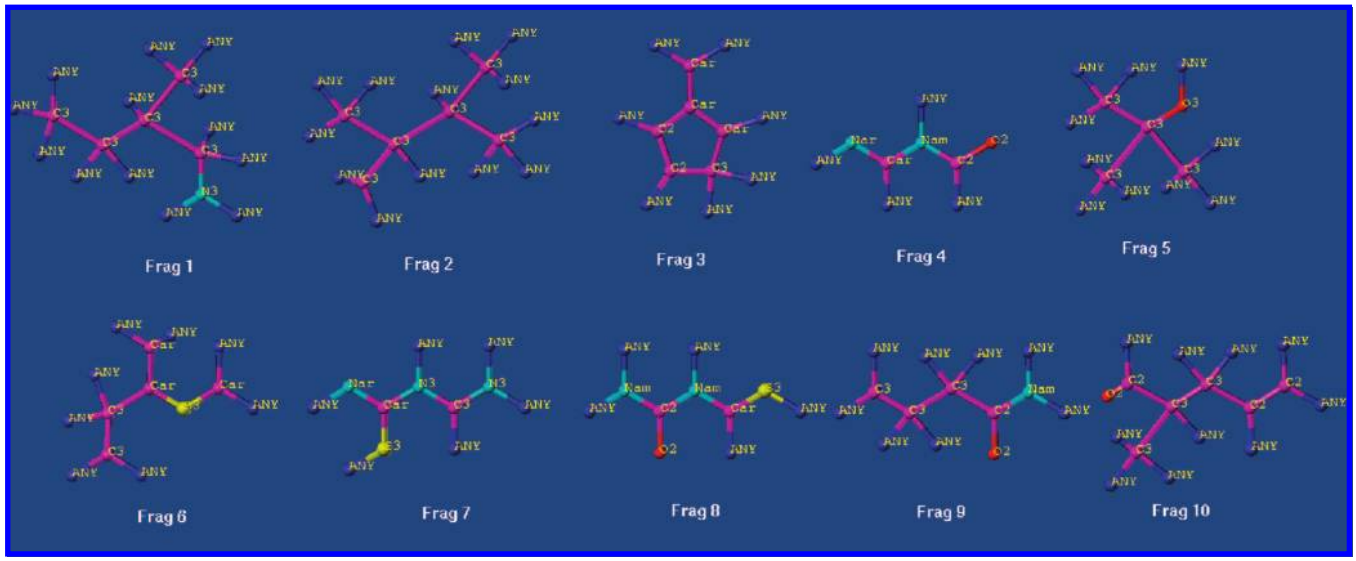


Figure 2. Top 10 fragments toward CCR2 antagonism for database mining. Magenta = carbon atoms, where C2, C3, and Car indicate sp^2, sp^3, and aromatic carbon, respectively. Red = oxygen atoms, where O2 and O3 indicate sp^2 and sp^3 oxygen, respectively. Cyan = nitrogen atom, where N3, Nar, and Nam indicate sp^3, aromatic, and amide nitrogen, respectively. Yellow = sulfur atom, where S3 is sp^3 sulfur. Blue = any atom, obtained during fragmentation.

Table 3. Dual Feature Hit Matrix

fragments	1	2	3	4	5	6	7	8	9	10
1	**4608**									
2	947	**9584**								
3	0	1	**148**							
4	308	1342	0	**5362**						
5	0	5	0	2	**94**					
6	6	0	0	0	0	**182**				
7	8	0	0	0	0	112	**142**			
8	106	267	0	40	0	0	0	**2599**		
9	15	5	0	6	0	118	112	33	**3349**	
10	108	935	0	123	0	0	0	638	12	**2344**

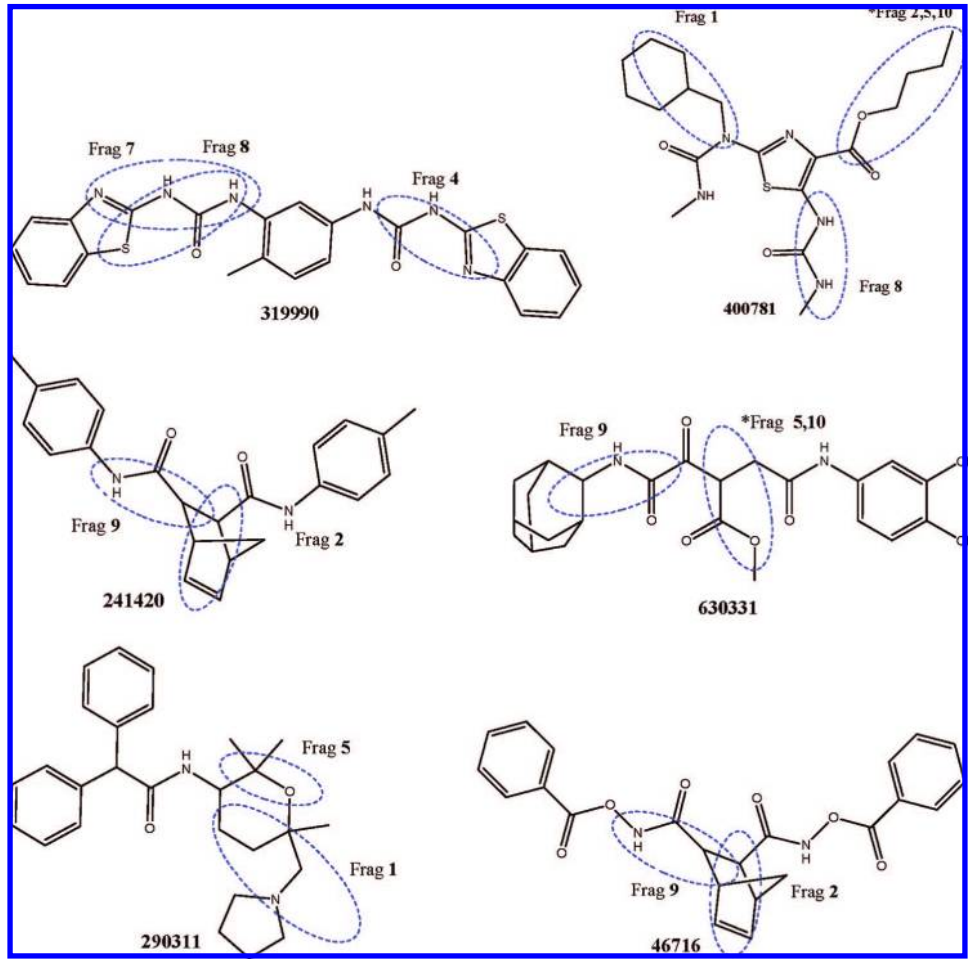
were obtained by searching molecules from a single fragment hits database. The single fragment hits database was generated for 10 individual fragments, and the hit matrix was developed on the basis of this. The total hits obtained after the search of dual featured fragments was 5249; on removal of repeat hits the number was reduced to 4211 (subdatabase), and these hits were subjected for further scrutiny. Table 3 shows a dual fragment featured hit matrix. Considering the hit matrix diagonally the off diagonal elements show the dual fragment hits, whereas the diagonal element shows the hits with single fragment (shown in bold).

4.3. 3-Point Pharmacophoric Fingerprinting and Similarity Search. Molecular fingerprints characterize a set of features derived from the structure of a molecule. The fingerprint based similarity search was performed based on a 3-point pharmacophore query determined from the three most actives **9**, **25**, and **35** from the data set of 36 representing different classes of molecules. The queries were first validated using the data set molecules. Query **9** picked up the butyramide class of molecules; additionally molecules from the aminothiazole series were also picked (**30**−**33**) (see Supporting Information S-Table 1a). This shows both classes of molecules have some common chemical features toward CCR2 activity. Query **25** retrieved only the aryl glycinamide class of compounds from the data set, while query **35** picked up **11, 12,** and **4** besides all the molecules of its class. Thus the ability to recover its own class of molecules using three different queries **9, 25**, and **35** validated the 3-point pharmacophoric fingerprint coupled with Tanimoto similarity metric. The 3-point pharmacophore based fingerprinting was performed using the subdatabase containing 4211 molecules. The query **9** gave 36 hits out of 4211, while **35** gave 30 hits. Surprisingly no hits were found with the **25** query; one of the probable reasons for this could be that this series of molecules is less active in comparison with the other series, thus not expected to be present in the list of 10 fragments chosen for the study. The total number of hits obtained by queries **9** and **35** was 66, and these were subjected to prediction by the validated HQSAR model. The hits with pIC50 > 8 were considered as probable scaffolds for CCR2 antagonism as these involve the most contributing fragments for CCR2 antagonism. Few scaffolds predicted by HQSAR are shown in Table 4 with their corresponding NCI codes. Since dual fragment features in the molecules were taken to retain an improved trait toward CCR2 antagonism, we tried to explore the features in the identified scaffolds. A few scaffolds with the fragment features highlighted are shown in Chart 2. **NSC 319990** shows 3 essential fragments: Frag4, Frag7, and Frag8. Aliphatic or alicyclic units represented by Frag 2 are seen in **NSC 241420** and **NSC 46716**. It is noteworthy that some scaffolds **NSC 630331** and **NSC 400781** showed an amalgamation of one or more fragments in a single fragment feature. The importance of fragment

Table 4. Few Scaffold Hops Predicted with a Validated HQSAR Model along with Docking Scores

molecules	HQSAR predicted pIC_{50}	Glide energy (kcal/mol)	docking score (kcal/mol)	Glide E-model (kcal/mol)
TAK-779	-	−50.29	−6.41	−85.24
35	-	−71.78	−9.40	−136.09
9	-	−54.45	−8.87	−96.75
25	-	−52.43	−8.31	−90.35
NSC 241420	8.50	−51.01	−7.94	−82.18
NSC 290311	8.29	−47.91	−6.81	−73.00
NSC **319990**	8.70	**−68.76**	−9.46	**−122.17**
NSC 400781	8.05	−55.26	−7.30	−84.84
NSC **46716**	8.10	**−68.39**	−8.47	**−113.04**
NSC 630331	10.14	−57.06	−7.37	−90.48

Chart 2. Identified NCI Scaffolds with Dual Fragment Features[a]



[a] The * indicates a combination of 2 or more fragments.

features for CCR2 binding is already discussed above, and its presence in NCI scaffolds is elaborated on in the docking section.

4.4. Molecular Docking. Incorporation of the intrinsic flexibility of protein binding sites is one of the problems in structure based approaches.[44,45] The flexibility of protein becomes crucial in cases when the protein structure is of low resolution failing to acknowledge loop mobility. Also problems are often seen when the target structure is a modeled structure or if the ligand shows an induced fit phenomenon leading to protein structure rearrangements.[45] In the case of CCR2, the absence of experimental structural coordinates is an existent predicament. Although structure-based studies involving homology modeling and docking for some CCR2 antagonists were performed by Berkhout et al., their results were debatable with the recent work reported by Marshall et al.[46,47] The theoretical model developed by Berkhout et al. is not publicly available. The binding sites revealed by both groups for the known CCR2 antagonist **TAK779** are in close proximity to each other. To gain insight into the exact binding regions, we analyzed the extra cellular loop region in the theoretical model developed by Shi et al.[48] (PDB 1KP1). A binding cleft is observed between the extra cellular loop region of first and second and the N terminal domain. Another subpocket is noticed between helix 1 and 2 of the extra cellular domain. It is quite possible that different loop arrangements in the extra cellular domain are responsible for docking the ligands at different sites as studied by two groups. So we decided to dock our ligands in both sites using IKP1. We first docked ligands by flexible ligand docking using FlexX and Glide.[40,41] FlexX failed to dock **TAK779** which we considered as the reference molecule. On the other hand Glide docked all the ligands but failed to score them as per the binding affinity data. To circumvent this problem we introduced flexibility in the loop region using IFD of Schrodinger.[42,43] The best predicted scaffolds along with **TAK779** were subjected to IFD using modeled structure 1KP1. The ligand sampling using IFD yielded several different poses for each ligand, and only those with the top scored conformations of the molecules were considered for further analysis. The scoring was assigned based on the Glide energy, docking score, and E-model energy. There is a good correlation noticed between scores obtained and binding affinity data of the molecules. Table 4 shows the different scores for the data set molecules along with the identified NCI hits. The side-chain χ dihedral angles represent protein flexibility suggesting the deviation from the initial angle. Rearrangements at side-chain His33, Val37, Gln39, Ile40, Gly41, and Ala42 were observed in most of the molecules except **25**. The probable reasons are that this class of molecules is smaller when compared to others, and they do not show the pi stacking interaction which is seen in the majority of the docked molecules. Due to this a vacant region is seen in the docked model of **25** which is either filled by phenyl or spiroindene groups in other classes showing higher potency over arylglycinamide.

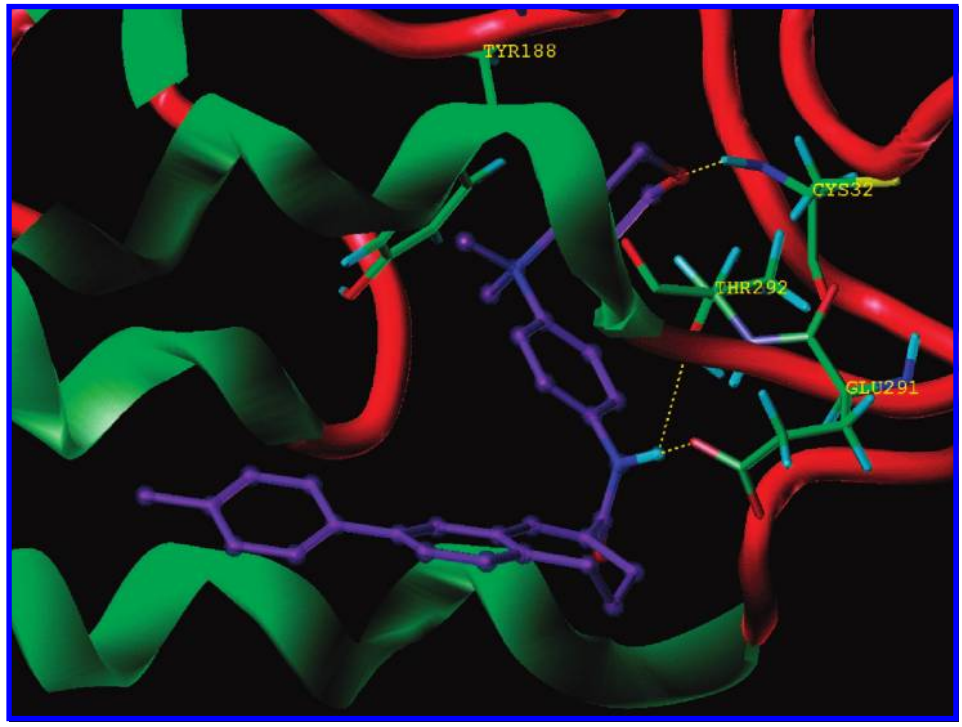


Figure 3. Docked structure of **TAK779** in CCR2 (1KP1). The residues hydrogen bonded to the **TAK779** are shown.

Docking Model of TAK779. The docked orientation of **TAK779** is shown in Figure 3. The tetrahydropyran ring is buried into the pocket with a favorable hydrophobic interaction with Tyr188. The O of the tetrahydropyran shows an H-bonding with backbone NH of Cys32. The ligand orientation of **TAK779** in CCR2 is much closer to the one reported by Marshall et al. differing in a few interactions near the biaryl and terminal phenyl portion of **TAK779**.[47] Our docking model shows H-bonding interaction of amide N of **TAK779** with Glu291 (Figure 3). Also, the early docking studies by Bekhout et al. suggested that Glu291 is one of the important residues in the binding pocket.[46] An H-bond interaction with Thr292 with amide NH of **TAK779** is noticed which is supported by mutational studies that Thr292 is important for CCR2 antagonism (Figure 3).[46] The phenyl ring attached to the biaryl forms hydrophobic interaction with Leu48 and Ile300. The central region of **TAK779** comprising of biaryl ring enters into the hydrophobic channel created by Leu44 and Leu293. The tail portion of **TAK779** fits in the hydrophobic region sandwiched between Leu48, His297, and Ile300. These observations indicate possibly the penetration of hydrophobic groups in a molecule was into the lipophilic groove bounded by residues Leu44, Leu48, Leu293, His297, and Ile300.

Docking Model of 35. The docked conformation of **35** is shown in Figure 4. The bistrifluro moiety of **35** is seen near the biaryl segment of **TAK779**. The phenyl ring attached to the piperidine moiety forms a Π-Π stacking and hydrophobic interaction with Tyr188 and Val37, respectively (Figure 4). The piperidine ring shows apolar contacts with Leu45 and Val37. The bistrifluro moiety enters the apolar pocket in a way similar to that of **TAK779** biaryl rings. On the other hand the acetamidothiazole shows strong hydrophobic contacts with Leu45 and Ile300 as seen in the tail portion of **TAK779**. As seen in Figure 4 there is an H-bond interaction between amide N and the carbonyl O of Thr292. The amino group of flexible acetamido substituted on the thiazole shows H-bonding with His297 (Figure 4).

Docking Model of 9. The docked superposed conformations of **9** and **35** show (Figure 5) the fine superposition of bismethyl triflurophenyl moiety. The cyclopropyl ring of **9** overlies on the thiazoles ring of **35**. The Π-Π stacking interaction in this compound is absent; instead the spiroindenepiperidine moiety shows weak hydrophobic contacts with residues namely His33, Val37, and Tyr188. This could be one of the reasons for reduced activity of this class of compounds over aminothiazoles. The cyclopropane ring orients in a fashion to make hydrophobic contacts with Val37 and Leu45. In comparison with **35** one of the methyl trifluro moieties shows apolar contacts with Leu293 and His297 while the other with Leu45, Leu48, and ILeu300. The amide nitrogen of butyramide shows strong H-bonding interactions with a pivotal residue Thr292 (Figure 6).

Docking Model of NSC 241420. The bicyclo heptene moiety of **NSC 241420** shows good superposition with the piperidine moiety of **35**. The 2 units of p-tolylformamide originating from the bicyclopiperidine stretch form a V-shaped orientation lying on the bismethyl triflurophenyl and acetamidothiazole moieties. One of the p-tolyl moieties shows apolar contacts with Ile300. Due to the shorter fragment length of p-tolyl moieties it forms fewer hydrophobic contacts in the lipophilic groove. Also the absence of Π-Π stacking with Tyr188 and only hydrogen bonding observed with Thr292 suggests the molecule to be a weak binder (see Supporting Information S-Figure 1a).

This suggests that the Π-Π stacking and the optimum size of the molecule is essential for embedding into the lipophilic groove to enhance the binding. Molecular optimization near the p-tolyl region and substitutions at bicycloheptene can further improve the binding of **NSC 241420.**

Docking Model of NSC 46716. Even in case of **NSC 46716** the bicycloheptene moiety superposes well with the piperi-

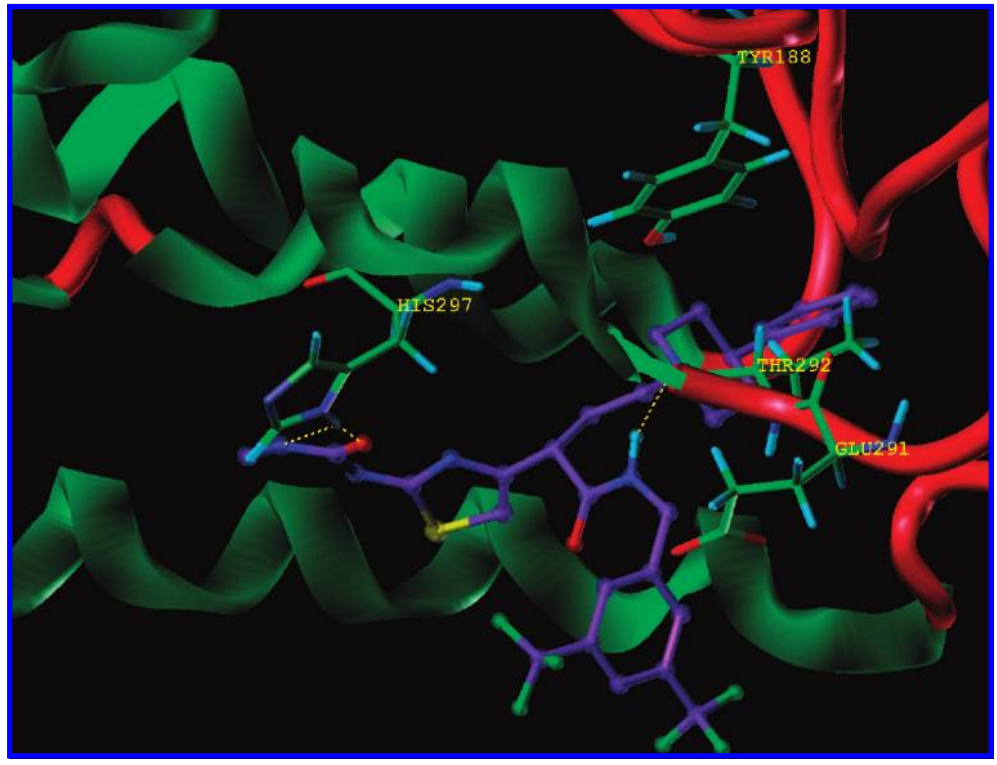


Figure 4. Docked structure of **35** in CCR2 (1KP1). The residues hydrogen bonded to **35** are shown.

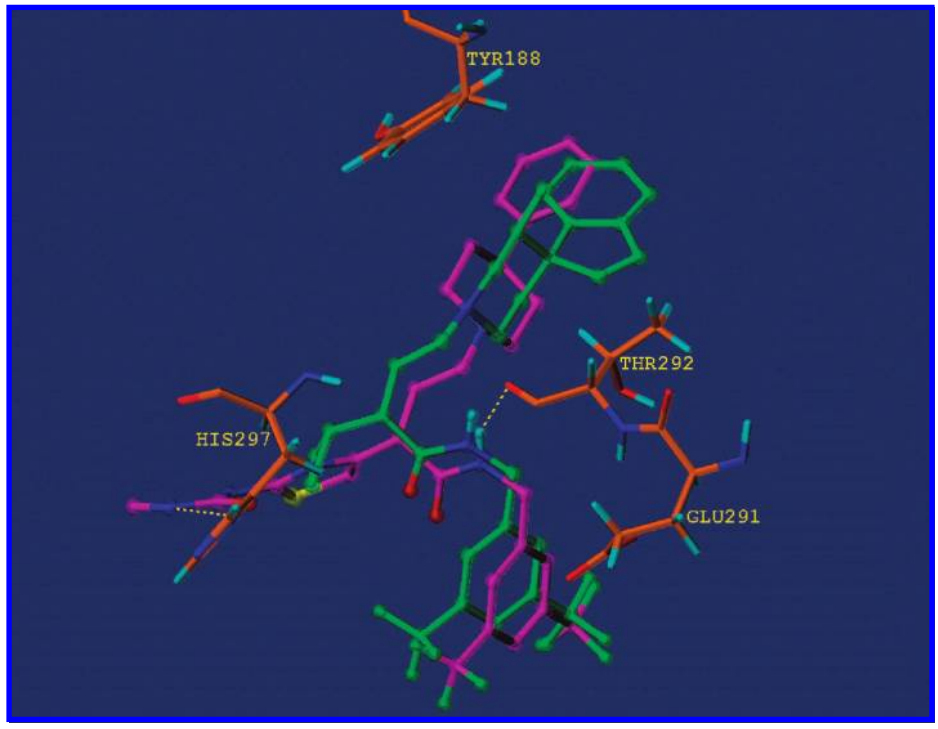


Figure 5. The docked overlay of **9** and **35** with residues hydrogen bonded to **35** are shown.

dine moiety of **35**. The benzoyloxy formamide units shows fine overlay with the bismethyl triflurophenyl and acetamidothiazole moieties. The former units essentially enters well in the lipophilic groove as observed in the case of **35**. Thus the strength of binding would be much higher as suggested by the docking results due to the larger benzoyloxy formamide units compared to the p-tolylformamide units of **NSC 241420**.

Also the hydrogen bonding interaction of two amide N and one carbonyl O of **NSC 46716** with Thr292 (see Supporting Information S-Figure 1b) increases the binding strength compared to **NSC 241420.** The presence of groups forming Π-Π interaction with Tyr188 would further improve the binding.

Docking Model of NSC 319990. The **NSC 319990** is presumably the best scaffold identified by the scaffold hopping exercise. The molecule seems to contain all the required pharmacophoric features for CCR2 binding. The molecule also has the highest score among the docked scaffolds (Table 4). The superposed conformations of **NSC**

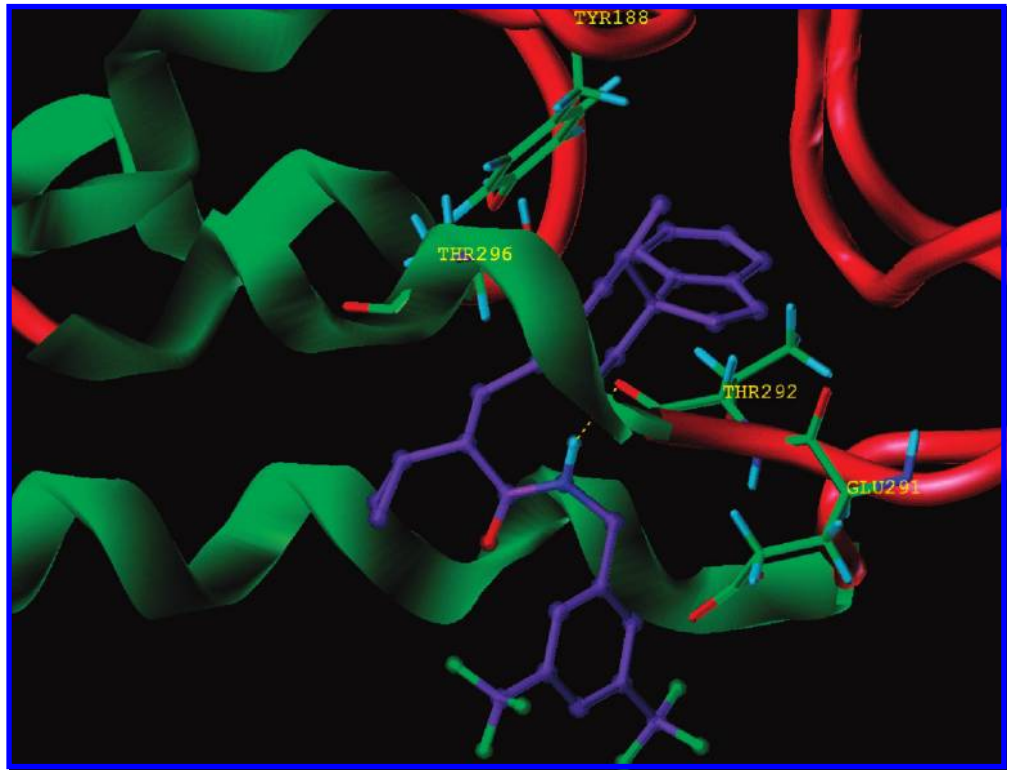


Figure 6. Docked structure of **9** in CCR2 (1KP1). The residues hydrogen bonded to the **9** are shown.

Figure 7. The docked overlay of **NSC 319990** and **35** with residues hydrogen bonded to **NSC 319990** are shown.

319990 with **35** show fine overlap covering the major pharmacophoric features (Figure 7). Overlay of **35**, **9**, and **NSC 319990** with their Connolly surfaces is shown in Figure 8.

First the phenyl Π-Π interaction of **35** with Tyr188 is maintained by one of the benzothiazole rings ortho to the central toluene. Second the most pivotal H bonding which is an interaction with Thr292 is retained (Figure 9). In comparison to **35** stronger interactions are observed in this molecule with 2 strong H bonds formed between amide nitrogens (ortho to toluene) and Thr292 indicating a probable binding with CCR2. The key lipophilic pocket explored by the aminothiazole class of compounds could be a reason for its superior potency over other data set molecules. The acetamidoaminothiazole substituent of **35** and **NSC 319990** explores the same pocket where acetamidoaminothiazole superposes with the benzothiazole moiety. An H-bonding interaction seen in **35** is preserved with carbonyl O (para to toluene) with His297. Also an additional H-bonding interaction of benzthiazole N with Asn301 is also observed (Figure 9). Figure 8 showing the Connolly surface shows the data set molecules take-up a distorted Y-shaped conformation,

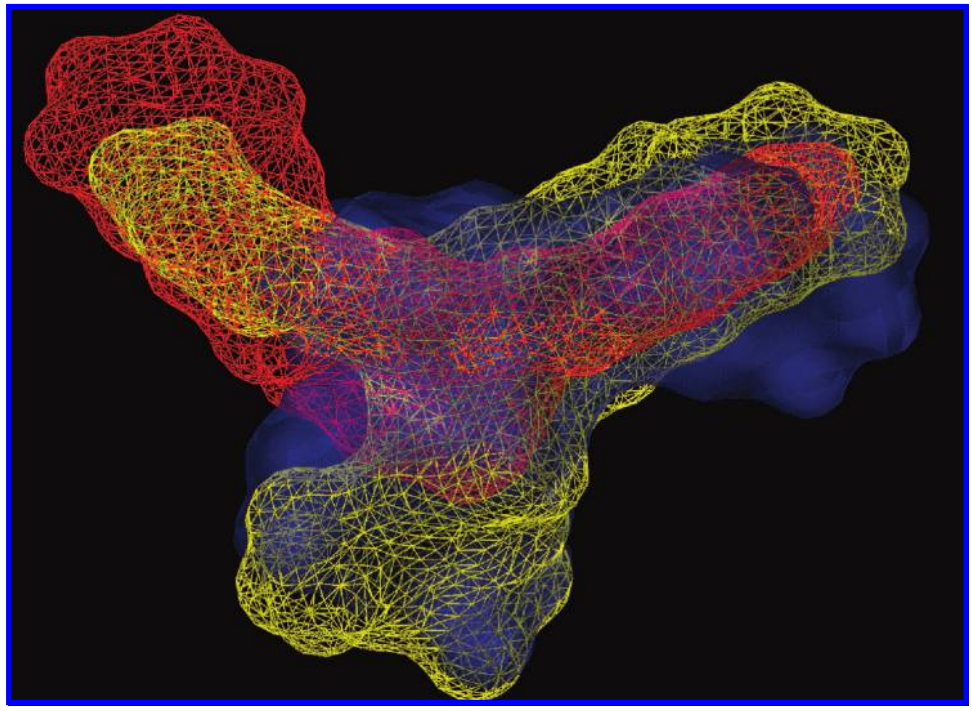


Figure 8. Connolly surface for **9, 35**, and **NSC 319990** for shape comparison. Red mesh view shows shape of **NSC 319990**. Yellow mesh view shows shape of **35**. Blue transparent view shows shape of **9**.

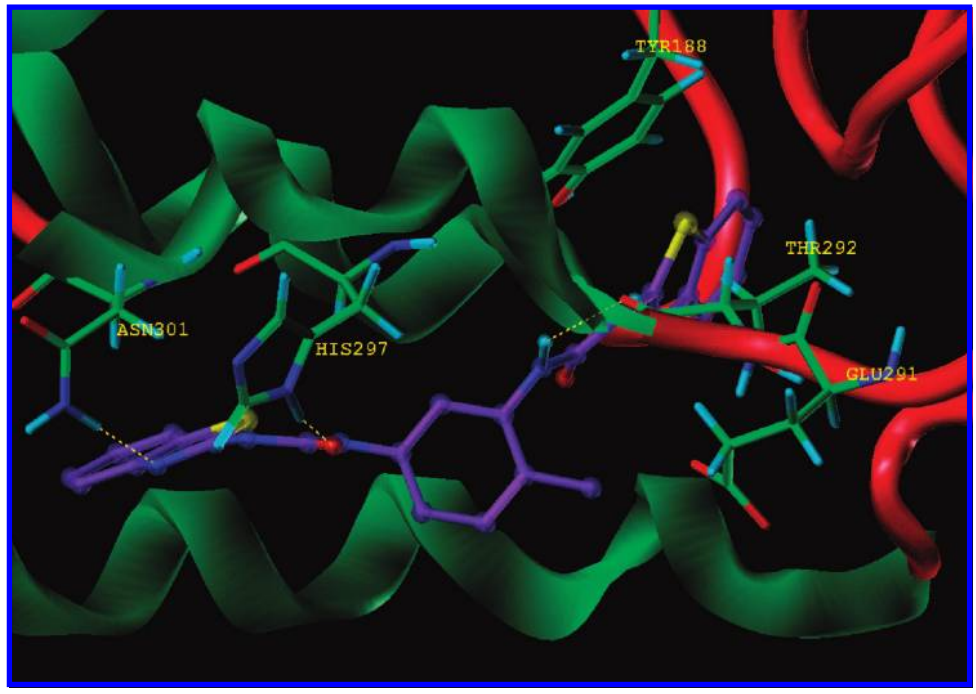


Figure 9. Docked structure of **NSC 319990** in CCR2 (1KP1). The residues hydrogen bonded to the **NSC 319990** are shown.

whereas the **NSC 319990** lacking one arm of the Y conformation still retains the essential pharmacophoric features. Further, an analysis was carried out on **NSC 319990** to check its druglike characteristics. It is interesting to note that this molecule has a mol wt of 474, the number of H-bond donors are 4, the number of H-bond acceptors are 8, AlogP = 3.772, and the number of rotatable bonds present are 4 in comparison with **35** having a mol wt of 627, the number of H-bond donors are 3, the number of H-bond acceptors are 13, AlogP = 5.064, and the number of rotatable bonds present are 8. Therefore we presume that this molecule can turn out to be a novel scaffold either equivalent to **35** or a better lead in terms of its potency and basic pharmacokinetic properties. This molecule can further be targeted by medicinal chemistry approaches to obtain new mimetics as CCR2 binders. Overall taking into consideration similarity, pharmacophoric characteristic, HQSAR prediction, analysis of single and dual fragment hits, docked pose, and druglike properties, it gives the speculation that **NSC 319990** probably could turn out to be a novel scaffold for CCR2 binding.

Though we carried out *in silico* validation for this molecule from different perspectives of computational analysis the fate of the molecule needs to be proved further experimentally.

5. COMPARISON WITH OTHER RELATED VIRTUAL SCREENING APPROACHES

Fingerprint based mining (FBM), an approach explored in this current study, offers a new strategy of ligand based mining where fragment features are derived from a validated HQSAR model. In this approach the fragments contributing to a particular biological end point are used for the mining purpose. For example, CCR2 antagonist activity has been used in the present case. The hits from the mining exercise are subjected to a 3-point pharmacophore fingerprinting, calculated from a 2D molecular graph. There are other virtual screening approaches that are reported in the literature. For example, Tropsha et al. adopted a QSAR based database mining approach wherein topological indices were calculated for a representative data set as well as the database to be screened.[8] In this approach, chemical similarity was calculated between the molecules in the data set with every molecule in the database using Euclidean distances based on the k Nearest Neighbor (kNN) method. The database was ranked, and the molecules closer to the active probes were selected based on similarity threshold. The selected molecules predicted by a majority of the QSAR were further subjected to docking or proved by biological activity.[8,9] An analysis of this method with ours shows that in our present study fragment features were used for database mining purposes, whereas topological indices were used for the same purpose by Tropsha et al. The major difference between the two methods lies in the fact that the present study makes use of fragments resulting from the validated QSAR model unlike Tropsha et al. which uses the descriptors calculated from the active data set and the database to be screened. Also, in the present study, we made use of Tanimoto derived scores for comparing the queries with the database molecules, whereas Tropsha et al. adopted Euclidean distances based on the k Nearest Neighbor (kNN) method. Both studies used statistically significant QSAR models to predict the activity of the hits obtained.[7−9]

Another ligand based similarity approach termed as the turbo similarity search developed by Hert et al. discusses the importance of the group fusion technique.[8,25] The steps of descriptor calculation and sorting of database are much similar to Tropsha et al. with an additional step of similarity calculation for every molecule in the database. The sorted databases are merged hence the name group fusion which convincingly proved to be better than the simple similarity search methods for the studied biological targets.[25] The method adopted by Hert et al. does not involve any QSAR development and prediction steps unlike Tropsha et al. and the present study.[8,25] However, their results conclude the importance of substructural analysis is persuasive enough for obtaining diverse scaffold hops.[25] This step has been taken into consideration in the present study also by making use of the fragments (substructure) features along with pharmacophore fingerprint emphasizing its efficient role in scaffold hopping.

6. CONCLUSIONS

In the present study we have utilized the HQSAR model as a virtual screening tool for identification of scaffolds as CCR2 antagonists. Different ligand based approaches were together exploited including QSAR, pharmacophoric fingerprinting, and similarity measure. The scaffold hopping strategy was achieved with fragments as descriptors generated by the validated HQSAR model for chemical database mining. Dual fragment features in the molecules were taken up for the study in order to retain an improved trait toward CCR2 antagonistic activity and arrive at a precise number of hits. Pharmacophore based similarity searching was done in order to retain the pharmacophoric features and screen out the remaining hits. The pharmacophore based similarity searching was validated with the data set molecules in turn to check its efficiency. The queries for the similarity search were considered from 3 actives from the different series of the data set. The resulting molecules were predicted by the HQSAR model, and the best predicted were suggested as probable scaffolds for CCR2 antagonism. Docking studies were performed on the identified scaffolds. The Frag 4, Frag 8, and Frag 9 reflect the presence of a Nam atom. Docking studies reveal the importance of amide N in H-bonding with Thr292. The data set molecules showed a crucial interaction with Thr292 which were retained by the identified scaffolds. In summary, this study provides a new strategy of a ligand based technique exploiting fragment features in combination with structure-guided identification of novel scaffolds. Two of the identified molecules **NSC 319990** and **NSC 46716** have promising outcomes and can be considered as novel scaffolds for CCR2 binding. This technique can help medicinal chemists in the identification of scaffolds surpassing the congeneric chemical space often observed in a regular QSAR methodology.

ACKNOWLEDGMENT

The authors thank the anonymous reviewers for their valuable comments on the manuscript. The authors thank the Department of Science and Technology (DST) and the Council of Scientific and Industrial Research (CSIR) for financial support.

Supporting Information Available: Additional details for structure of molecules with actual and predicted activity of different classes of compounds (S-Table1a-1f) and docking figures for **NSC 241420** and **NSC 46716** (S-Figure 1a and S-Figure 1b, respectively). This material is available free of charge via the Internet at http://pubs.acs.org.

REFERENCES AND NOTES

(1) Tsou, C. L.; Peters, W; Si, Y.; Slaymaker, S.; Aslanian, A. M.; Weisberg, S. P.; Mack, M.; Charo, I. F. Critical roles for CCR2 and MCP-3 in monocyte mobilization from bone marrow and recruitment to inflammatory sites. *J. Clin. Invest.* **2007**, *117*, 902–909.
(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) Lumeng, C. N.; Deyoung, S. M.; Bodzin, J. L.; Saltiel, A. R. Increased inflammatory properties of adipose tissue macrophages recruited during diet-induced obesity. *Diabetes* **2007**, *56*, 16–23.
(4) Sohy, D.; Parmentier, M.; Springael, J. Y. Allosteric transinhibition by specific antagonists in CCR2/CXCR4 heterodimers. *J. Biol. Chem.* **2007**, *282*, 30062–69.

(5) Szczucinski, A.; Losy, J. Chemokines and chemokine receptors in multiple sclerosis. Potential targets for new therapies. *Acta Neurol. Scand.* **2007**, *115*, 137–146.
(6) Mine, S.; Okada, Y.; Tanikawa, T.; Kawahara, C.; Tabata, T.; Tanaka, Y. Increased expression levels of monocyte CCR2 and monocyte chemoattractant protein-1 in patients with diabetes mellitus. *Biochem. Biophys. Res. Commun.* **2006**, *344*, 780–785.
(7) Oloff, S.; Mailman, R. B.; Tropsha, A. Application of validated QSAR models of D1 dopaminergic antagonists for database mining. *J. Med. Chem.* **2005**, *48*, 7322–7332.
(8) Shen, M.; Beguin, C.; Golbraikh, A.; Stables, J. P.; Kohn, H.; Tropsha, A. Application of predictive QSAR models to database mining: identification and experimental validation of novel anticonvulsant compounds. *J. Med. Chem.* **2004**, *47*, 2356–2364.
(9) Zhang, S.; Wei, L.; Bastow, K.; Zheng, W.; Brossi, A.; Lee, K. H.; Tropsha, A. J. Antitumor Agents 252. Application of validated QSAR models to database mining: discovery of novel tylophorine derivatives as potential anticancer agents. *J. Comput.-Aided Mol. Des.* **2007**, *21*, 97–112.
(10) Zhang, Q.; Muegge, I. Scaffold Hopping Through virtual screening using 2D and 3D similarity descriptors: Ranking, voting, and consensus scoring. *J. Med. Chem.* **2006**, *49*, 1536–1548.
(11) Esposito, E. X.; Hopfinger, A. J.; Madura, J. D. Methods for applying the quantitative structure-activity relationship paradigm. *Methods Mol. Biol.* **2004**, *275*, 131–214.
(12) Thorner, D. A.; Willett, P.; Wright, P. M.; Taylor, R. Similarity searching in files of three-dimensional chemical structures: representation and searching of molecular electrostatic potentials using field-graphs. *J. Comput.-Aided Mol. Des.* **1997**, *11*, 163–174.
(13) Willett, P. Similarity-based virtual screening using 2D fingerprints. *Drug Discovery Today* **2006**, *11*, 1046–1053.
(14) Mason, J. S.; Morize, I.; Menard, P. R.; Cheney, D. L.; Hulme, C.; Labaudiniere, R. F. New 4-point pharmacophore method for molecular similarity and diversity applications: Overview of the method and applications, including a novel approach to the design of combinatorial libraries containing privileged substructures. *J. Med. Chem.* **1999**, *42*, 3251–3264.
(15) Bender, A.; Glen, R. C. Molecular similarity: a key technique in molecular informatics. *Org. Biomol. Chem.* **2004**, *2*, 3204–3218.
(16) McGregor, M. J.; Muskal, S. M. Pharmacophore fingerprinting. 1. Application to QSAR and focused library design. *J. Chem. Inf. Comput. Sci.* **1999**, *39*, 569–574.
(17) Bajorath, J. Integration of virtual and high-throughput screening. *Nat. Rev. Drug Discovery* **2002**, *1*, 882–894.
(18) Martin, Y. C.; Bures, M. G.; Danaher, E. A.; DeLazzer, J.; Lico, I.; Pavlik, P. A. A fast new approach to pharmacophore mapping and its application to dopaminergic and benzodiazepine agonists. *J. Comput.-Aided Mol. Des.* **1993**, *7*, 83–102.
(19) Jones, G.; Willett, P.; Glen, R. C. A genetic algorithm for flexible molecular overlay and pharmacophore elucidation. *J. Comput.-Aided Mol. Des.* **1995**, *9*, 532–549.
(20) Patel, Y.; Gillet, V. J.; Bravi, G.; Leach, A. R. A comparison of the pharmacophore identification programs: Catalyst, DISCO and GASP. *J. Comput.-Aided Mol. Des.* **2002**, *16*, 653–681.
(21) Willett, P.; Winterman, V. A comparison of some measures of intermolecular structural similarity. *Quant. Struct.-Act. Relat.* **1986**, *5*, 18–25.
(22) Salim, N.; Holliday, J.; Willett, P. Combination of fingerprint-based similarity coefficients using data fusion. *J. Chem. Inf. Comput. Sci.* **2003**, *43*, 435–442.
(23) Martin, Y. C.; Kofron, J. L.; Traphagen, L. M. Do Structurally Similar Molecules Have Similar Biological Activity. *J. Med. Chem.* **2002**, *47*, 4350–4358.
(24) Flower, D. R. On the properties of bit string based measures of chemical similarity. *J. Chem. Inf. Comput. Sci.* **1988**, *38*, 379–386.
(25) Hert, J.; Willett, P.; Wilton, D. J.; Acklin, P.; Azzaoui, K.; Jacoby, E.; Schuffenhauer, A. New methods for ligand-based virtual screening: use of data fusion and machine learning to enhance the effectiveness of similarity searching. *J. Chem. Inf. Model.* **2006**, *46*, 462–470.
(26) Matter, H. Selecting optimally diverse compounds from structure databases: a validation study of two-dimensional and three-dimensional molecular descriptors. *J. Med. Chem.* **1997**, *40*, 1219–1229.
(27) Nair, P. C.; Srikanth, K.; Sobhia, M. E. QSAR studies on CCR2 antagonists with chiral sensitive hologram descriptors. *Bioorg. Med. Chem. Lett.* **2008**, *18*, 1323–1330.
(28) Srikanth, K.; Nair, P. C.; Sobhia, M. E. Probing the structural and topological requirements for CCR2 antagonism: holographic QSAR for indolopiperidine derivatives. *Bioorg. Med. Chem. Lett.* **2008**, *18*, 1450–1456.
(29) Zhou, C.; Guo, L.; Parsons, W. H.; Mills, S. G.; MacCoss, M.; Vicario, P. P.; Zweerink, H.; Cascieri, M. A.; Springer, M. S.; Yang, L. Alpha-aminothiazole-gamma-aminobutanoic amides as potent, small molecule CCR2 receptor antagonists. *Bioorg. Med. Chem. Lett.* **2007**, *17*, 309–314.
(30) Butora, G.; Morriello, G. J.; Kothandaraman, S.; Guiadeen, D.; Pasternak, A.; Parsons, W. H.; MacCoss, M.; Vicario, P. P.; Cascieri, M. A.; Yang, L. 4-Amino-2-alkyl-butyramides as small molecule CCR2 antagonists with favorable pharmacokinetic properties. *Bioorg. Med. Chem. Lett.* **2006**, *16*, 4715–4722.
(31) Yang, L.; Zhou, C.; Guo, L.; Morriello, G.; Butora, G.; Pasternak, A.; Parsons, W. H.; Mills, S. G.; MacCoss, M.; Vicario, P. P.; Zweerink, H.; Ayala, J. M.; Goyal, S.; Hanlon, W. A.; Cascieri, M. A.; Springer, M. S. Discovery of 3,5-bis(trifluoromethyl)benzyl L-arylglycinamide based potent CCR2 antagonists. *Bioorg. Med. Chem. Lett.* **2006**, *16*, 3735–3739.
(32) Butora, G.; Jiao, R.; Parsons, W. H.; Vicario, P. P.; Jin, H.; Ayala, J. M.; Cascieri, M. A.; Yang, L. 3-Amino-1-alkyl-cyclopentane carboxamides as small molecule antagonists of the human and murine CC chemokine receptor 2. *Bioorg. Med. Chem. Lett.* **2007**, *17*, 3636–3641.
(33) *SYBYL Molecular Modeling System*, *version 7.1*; Tripos Inc.: St. Louis, MO, 2007.
(34) Powell, M. J. D. Restart procedures for the conjugate gradient method. *Math. Program.* **1977**, *12*, 241–254.
(35) Lowis, D. R. *HQSAR, A New, Highly Predictive QSAR Technique*; Technical Notes; 1997/1; Tripos Inc.: St. Louis, MO, 1997.
(36) Cramer, R. D.; Bunce, J. D.; Patterson, D. E. Recent advances in comparative molecular field analysis (CoMFA). *Prog. Clin. Biol. Res.* **1989**, *29*, 161–165.
(37) *NCI Database Distributed with SYBYL, version 7.1*; Tripos Inc.: St. Louis, MO 63144-2913, 2007.
(38) Baber, J. C.; Shirley, W. A.; Gao, Y.; Feher, M. The use of consensus scoring in ligand-based virtual screening. *J. Chem. Inf. Comput. Sci.* **2006**, *46*, 277–288.
(39) *MOE (Molecular Operating Environment)*, *version 2006.08*; Chemical Computing Group Inc.: Montreal, Canada, 2006.
(40) Kramer, B.; Rarey, M.; Lengauer, T. Evaluation of the FLEXX incremental construction algorithm for protein-ligand docking. *Proteins: Struct., Funct., Genet.* **1999**, *37*, 228–241.
(41) Friesner, R. A.; Banks, J. L.; Murphy, R. B.; Halgren, T. A.; Klicic, J. J.; Mainz, D. T.; Repasky, M. P.; Knoll, E. H.; Shelley, M.; Perry, J. K.; Shaw, D. E.; Francis, P.; Shenkin, P. S. Glide: a new approach for rapid, accurate docking and scoring. 1. Method and assessment of docking accuracy. *J. Med. Chem.* **2004**, *47*, 1739–1749.
(42) Sherman, W.; Day, T.; Jacobson, M. P.; Friesner, R. A.; Farid, R. Novel procedure for modeling ligand/receptor induced fit effects. *J. Med. Chem.* **2006**, *49*, 534–553.
(43) *Induced Fit Docking protocol*; *Glide version 4.5, Prime version 1.6*; Schrödinger, LLC,:New York, NY, 2005.
(44) Teague, S. J. Implications of protein flexibility for drug discovery. *Nat. Rev. Drug Discovery.* **2003**, *2*, 527–541.
(45) Murray, C. W.; Baxter, C. A.; Frenkel, A. D. The sensitivity of the results of molecular docking to induced fit effects: application to thrombin, thermolysin and neuraminidase. *J. Comput.-Aided Mol. Des.* **1999**, *13*, 547–562.
(46) Berkhout, T. A.; Blaney, F. E.; Bridges, A. M.; Cooper, D. G.; Forbes, I. T.; Gribble, A. D.; Groot, P. H.; Hardy, A.; Ife, R. J.; Kaur, R.; Moores, K. E.; Shillito, H.; Willetts, J.; Witherington, J. CCR2: Characterization of the Antagonist Binding Site from a Combined Receptor Modeling/Mutagenesis Approach. *J. Med. Chem.* **2003**, *46*, 4070–4086.
(47) Marshall, T. G.; Lee, R. E.; Marshall, F. E. Common angiotensin receptor blockers may directly modulate the immune system via VDR, PPAR and CCR2b. *Theor. Biol. Med. Model.* **2006**, *3*, 1.
(48) Shi, X. F.; Liu, S.; Xiangyu, J.; Zhang, Y.; Huang, J.; Liu, S.; Liu, C. Q. Structural analysis of human CCR2b and primate CCR2b by molecular modeling and molecular dynamics simulation. *J. Mol. Model.* **2002**, *8*, 217–22.

CI800157J