

Clustering and Classifying Diverse HIV Entry Inhibitors Using a Novel Consensus Shape-Based Virtual Screening Approach: Further Evidence for Multiple Binding Sites within the CCR5 Extracellular Pocket

Violeta I. Pérez-Nuño,[†] David W. Ritchie,^{*,‡} Jose I. Borrell,[†] and Jordi Teixidó[†]

Grup d'Enginyeria Molecular, Institut Químic de Sarrià (IQS), Universitat Ramon Llull, Barcelona, Spain,
Department of Computing Science, King's College, University of Aberdeen, Aberdeen, U.K.

Received July 28, 2008

HIV entry inhibitors have emerged as a new generation of antiretroviral drugs that block viral fusion with the CXCR4 and CCR5 membrane coreceptors. Several small molecule antagonists for these coreceptors have been developed, some of which are currently in clinical trials. However, because no crystal structures for the coreceptor proteins are available, the binding modes of the known inhibitors within the coreceptor extracellular pockets need to be analyzed by means of site-directed mutagenesis and computational experiments. Previous studies have indicated that there is more than one binding site within the CCR5 extracellular pocket. This article investigates and develops this hypothesis using a novel spherical harmonic-based consensus shape clustering approach. The consensus shape approach is evaluated using retrospective virtual screening of CXCR4 and CCR5 inhibitors. Multiple combinations of CCR5 ligands in multiple trial superpositions are constructed to find consensus queries that give high virtual screening enrichments. Receiver–operator–characteristic performance analyses for both CXCR4 and CCR5 inhibitors show that the new consensus shape matching approach gives better virtual screening enrichments than existing shape matching and docking virtual screening techniques. The results obtained also provide strong evidence to support the notion that there are three main binding sites within the CCR5 extracellular cavity.

INTRODUCTION

Human immunodeficiency virus (HIV) entry inhibitors have emerged as a new generation of antiretroviral drugs which work by blocking interactions between the viral surface gp120 protein and the CXCR4 and CCR5 plasmatic membrane coreceptors of the host cell.^{1–4} A considerable number of small molecule antagonists for CXCR4 and CCR5 have been found to be effective for preventing viral entry, and some of them have been evaluated in clinical trials.^{5–9} However, no crystal structures of these coreceptors or their ligand-bound complexes are available. Consequently, several site-directed mutagenesis (SDM) and computational experiments have been carried out to identify the binding modes of the existing inhibitors. Analysis of the key CXCR4 SDM residues points to a well-defined localized binding cavity,¹⁰ but the CCR5 SDM residues are found to be spatially well-distributed around the pocket within the extracellular loops.¹¹ Moreover, the small-molecule inhibitors for CXCR4 are generally quite similar to each other, whereas CCR5 has many different inhibitors which derive from several diverse scaffold families. Several earlier computational binding experiments have indicated that different CCR5 ligands bind in fundamentally different ways within the CCR5 extracellular pocket.^{12–16} Furthermore, considering that (a) it is very difficult to superpose all the different families of CCR5 active compounds, (b) the results of retrospective virtual screening

(VS) enrichment studies are strongly dependent on the conformation of the query molecule, (c) SDM results suggest a large binding pocket within the extracellular loop region of the CCR5 structure, and (d) not all SDM mutations affect the binding of all ligands, there is good evidence to support a hypothesis that the known binders belong to two or more groups and that the members of each group bind to the same general region of the extracellular pocket. However, it is not clear a priori which actives might belong to which group. For example, different computational binding mode studies of ligands such as Atravirine,^{13,15} AD101,^{11,17} SCH-C,^{11,13–15,17} TAK-779,^{11,13,14,17–21} TAK-720,¹⁹ 2-aryl-4-(piperidin-1-yl)-butanamines,^{14,16} and 1,3,4-trisubstituted pyrrolidines piperidines,^{14,16} predict that they each bind in different ways within the CCR5 pocket. Hence it is difficult to obtain a clear picture of how these diverse ligands function.

Here we investigate the multisite binding hypothesis using a new consensus shape matching technique based on spherical harmonic (SH) representations of surface shapes²² to perform rapid and exhaustive comparison and clustering of multiple combinations of ligands in multiple trial superpositions. This novel SH-based shape-matching approach uses one or more “pseudomolecules”, obtained from the consensus shapes of the most active molecules, as VS queries against a database of known actives and decoys. The algorithm has been implemented in the ParaFit module of the ParaSurf suite of programs.²³ The new consensus shape-matching approach has been developed specifically to help analyze targets with

* To whom correspondence should be addressed. Tel: +44 1224 272282.
Fax: +44 1224 273422. E-mail: d.w.ritchie@abdn.ac.uk.

[†] Universitat Ramon Llull.

[‡] University of Aberdeen.

large ligand-binding pockets, although it may also be used in conventional VS studies where there are multiple known actives.

This article presents our new consensus shape matching VS approach and applies it to a large database of CXCR4 and CCR5 active inhibitors and comparable inactive decoys which was compiled previously.¹² Several trial CCR5 and CXCR4 consensus shape queries are constructed by clustering and superposing selected known actives, and the utility of each query is assessed using receiver-operator-characteristic (ROC)²⁴ plots. The area under the curve (AUC) of each ROC plot is used to provide an objective measure of the ability of each consensus shape to recognize known actives with similar shapes. To find the best clusters of binders with which to characterize the different groups of CCR5 antagonists, we conduct systematic experiments in which AUCs are compared for clusters formed in different ways. The best consensus queries thus found are then further analyzed in the context of the receptor pocket using rigid-body soft docking²⁵ onto the homology modeled CCR5 receptor.¹² Our virtual screening results show that the CCR5 inhibitors may be clustered into four superconsensus (SC) families, and our docking results show that these may be docked to three overlapping regions within the CCR5 extracellular pocket. These results provide strong evidence to support the notion that there are three main binding sites within the CCR5 extracellular pocket.

METHODS

Shape Representations. We use the ParaSurf and ParaFit modules of CEPOS InSilico Ltd.²³ to calculate and superpose molecular surfaces. ParaSurf calculates molecular shape and electronic properties from semiempirical quantum mechanics theory, and encodes these properties as SH expansions.²⁶ Surface shapes are represented as radial distance expansions of the molecular surface, $r(\theta, \varphi)$, with respect to a selected harmonic coordinate origin (CoH), which is normally set equal to the molecular center of gravity (CoG).²⁷ For example, the radial surface shape of molecule A is represented as

$$r_A(\theta, \varphi) = \sum_{l=0}^L \sum_{m=-l}^l a_{lm} y_{lm}(\theta, \varphi) \quad (1)$$

where θ and φ are the spherical coordinates, $y_{lm}(\theta, \varphi)$ are real spherical harmonics, a_{lm} are the expansion coefficients, and L is the order or highest polynomial power of the expansion. Here, $L = 6$ is used in all calculations. ParaFit calculates superpositions between pairs of molecules by exploiting the special rotational properties of the SH functions.²² For example, rotated SH expansion coefficients for molecule B may be calculated as

$$b'_{lm} = \sum_{m'=-l}^l R_{mm'}^{(l)}(\alpha, \beta, \gamma) b_{lm'} \quad (2)$$

where (α, β, γ) are zyz Euler rotation angles and $R_{mm'}^{(l)}(\alpha, \beta, \gamma)$ are real Wigner rotation matrix elements.²² To calculate a superposition between a pair of molecules, the CoH of molecule B is translated to that of the fixed reference molecule A, and a rotational search is then performed to find the rotation which minimizes the distance, D_{AB} , between the corresponding pairs of SH expansions

$$D_{AB} = \int (r_A(\theta, \varphi) - r'_B(\theta, \varphi))^2 d\Omega \quad (3)$$

Thanks to the orthogonality of the basis functions, this expression reduces to

$$D_{AB} = \sum_{l=0}^L \sum_{m=-l}^l a_{lm}^2 + b_{lm}^2 - 2a_{lm}b'_{lm} = |a|^2 + |b|^2 - 2ab' \quad (4)$$

Hence the distance function for any orientation may be calculated very rapidly from the original expansion coefficients. In VS, it is convenient to rearrange and normalize the basic distance expression (eq 4) to give a similarity score. Here, we use the Tanimoto score, S_{AB} , calculated as

$$S_{AB} = \frac{ab'}{|a|^2 + |b|^2 - ab'} \quad (5)$$

ParaSurf can calculate all necessary SH molecular properties in a matter of a few minutes. Once the surface shapes have been calculated, the ParaFit program can perform on the order of hundreds of molecular comparisons per second. Hence the overall approach is well-suited to tasks that require the calculation of multiple molecular comparisons such as high throughput VS.²⁷

SH Consensus Shape Matching. Using the SH representation, a “consensus shape”, $\bar{r}(\theta, \varphi)$, may be constructed as the average of N individual molecular shape expansion coefficient vectors, a_{lm}^k , for $k = 1, \dots, N$ as

$$\bar{r}(\theta, \varphi) = \frac{1}{N} \sum_{k=1}^N \sum_{l=0}^L \sum_{m=-l}^l a_{lm}^k y_{lm}(\theta, \varphi) \quad (6)$$

However, before computing the average, each molecule in the consensus must first be rotated to minimize the distance between it and the remaining $N - 1$ molecules. In practice, because these rotations are not known a priori, the consensus shape is constructed iteratively as follows. First, all-against-all rotational pairwise superpositions are calculated to find the two most similar surface shapes. Then, the average of these two shapes is taken as the initial seed shape for the consensus, and the remaining $N - 2$ SH shapes are rotated into superposition with the seed shape. The overall average of all SH coefficients is then computed to give the first estimate of the consensus shape. The consensus average is then refined by superposing the member molecular shapes back onto the average and by recalculating a new average shape. This procedure is repeated until convergence to optimal overlap is reached between each molecule and the consensus shape. This protocol is similar to techniques used for refining electron microscopy density images of similar molecules observed in different orientations.²⁸ In the present case, convergence is typically achieved in just three or four cycles. Hence calculating a consensus shape is a quick process. Figure 1 illustrates the overall procedure schematically.

ROC Plot Analyses. Here, all VS results are presented as ROC plots of true positive rate versus false positive rate or equivalently Sensitivity versus $1 - \text{Specificity}$. These quantities are calculated as

$$\text{True Positive Rate} = \text{TP}/(\text{TP} + \text{FN}) = \text{Sensitivity} \quad (7)$$

$$\text{False Positive Rate} = \text{FP}/(\text{TN} + \text{FP}) = 1 - \text{TN}/(\text{TN} + \text{FP}) = 1 - \text{Specificity} \quad (8)$$

where TP represents the number of correctly identified actives (true positives), FP represents the number of inactives

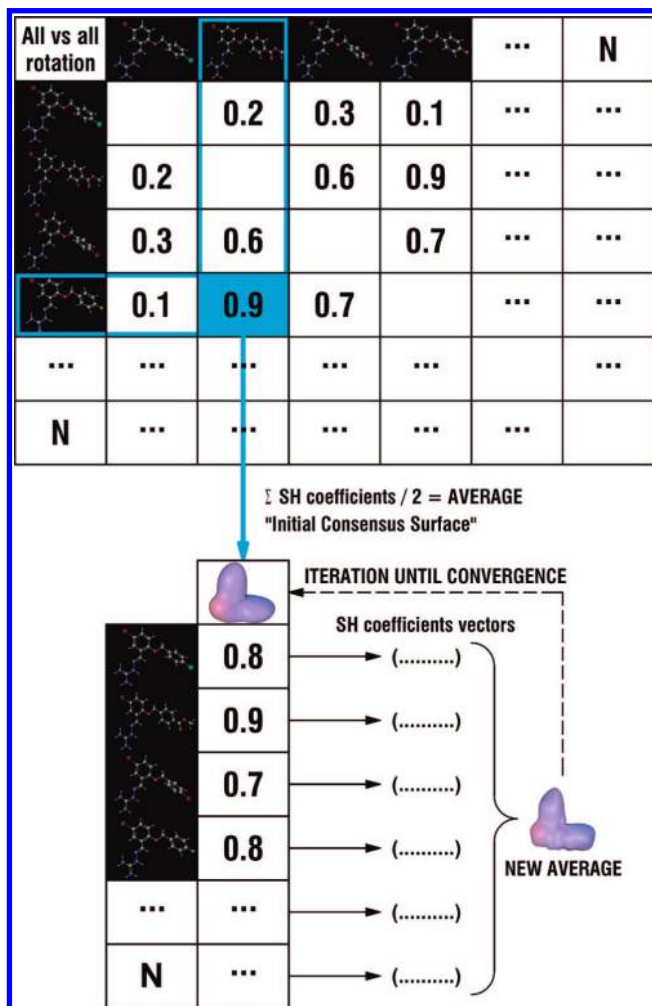


Figure 1. Flow diagram of the consensus shape calculation. First, ParaFit all-against-all rotational superpositions are calculated for the group of *N* molecules that will form the consensus. The two most similar SH shapes are selected and superposed to form a seed consensus shape. Then, all molecules are rotationally superposed onto the seed consensus. A new consensus shape is computed from the average SH coefficients of the superposed shapes. The consensus members are then superposed again onto the consensus average, and the process is iterated until convergence.

incorrectly predicted as active (false positives), TN represents the number of correctly identified inactives (true negatives), and FN represents the number of actives incorrectly predicted to be inactive (false negatives). Each ROC plot is calculated by ranking the database molecules by similarity with the query (or by docking energy with the protein target), and by summing the number of TPs, FPs, TNs, and FNs on either side of each rank position. ROC plots are particularly useful when comparing different VS queries with different numbers of actives and inactives because the AUC of a ROC plot gives an objective measure of query performance which is essentially independent of the actual number of positive and negative instances (i.e., ROC curves do not suffer from "class skew").²⁹

Consensus Shape-Based Virtual Screening. Obviously, using SH surfaces to compute average shapes will result in some smoothing and loss of detail compared to the original individual molecular shapes. However, this can be considered a desirable property because it provides an unbiased way to combine the most significant features of a related group of molecules. Nonetheless, it is important to select the member shapes carefully to achieve a good balance between capturing

Table 1. Families of CXCR4 and CCR5 Antagonists Used in the Current Study

family	number of compounds	ref
CXCR4 inhibitors		
tetrahydroquinolinamines	123	7, 30–34
KRH derivatives (Kureha Chemical Industries)	23	7, 35–38
macrocycles	4	39
AMD derivatives (AnorMED)	94	7, 39–44
cyclic peptides	2	45
other	2	46
total	248	
CCR5 inhibitors		
SCH derivatives (Schering-Plough)	120	17, 47, 48
diketopiperazines	9	49–53
anilide piperidine <i>N</i> -oxides	22	54
AMD derivatives (AnorMED)	3	44
4-piperidines	10	55, 56
4-aminopiperidine or tropanes	26	55, 57, 58
1,3,4-trisubstituted pyrrolidinepiperidines	9	59
phenylcyclohexylamines	9	60–65
TAK derivatives (Takeda)	66	66, 67
1-phenyl-1,3-propanodiamines	57	68–70
1,3,5-trisubstituted pentacyclics	9	71
<i>N,N'</i> -diphenylureas	4	72
5-oxopyrrolidine-3-carboxamides	5	73
guanyldihydrazone derivatives	33	74
4-hydroxypiperidine derivatives	36	75
other	6	76
total	424	

the most significant features for binding and smoothing away too much detail. In this work, this balance is achieved, and the overall approach is validated by monitoring the utility of various consensus shapes as VS queries against our database of known CCR5 and CXCR4 binders and decoys.¹² Since this database was first described, some newly published CCR5 inhibitors have been added to the set of actives. Table 1 lists the representative families of CXCR4 and CCR5 inhibitors in the updated database, and Figure 2 shows some representative members of each family. Consensus shape-based VS was applied to these families using query structures constructed from: (a) the consensus shape of the three most active compounds of different scaffolds families in the databases (i.e., an AMD derivative, a macrocycle derivative, and a KRH derivative for CXCR4, and a piperidine derivative, a SCH derivative, and a 1,3,4-trisubstituted pyrrolidine-piperidine derivative for CCR5), and (b) the consensus shape of *all* CXCR4 or CCR5 active inhibitors in the database.

The consensus shape-based approach was also used to investigate the CCR5 multiple binding site hypothesis. First, the CCR5 inhibitors were clustered using Ward's hierarchical clustering method,⁷⁷ as implemented in the JKlustor module of JChem,⁷⁸ using both chemical (topological) fingerprints and two-dimensional pharmacophore fingerprints. The optimal number of clusters to be selected was calculated using Kelley's method,⁷⁹ also implemented in JKlustor, and the consensus shapes of these fingerprint-defined clusters were calculated using ParaFit, as described above. Then, ParaFit was used again to compute all-against-all rotational superpositions of the fingerprint-defined consensus shapes. This produced a shape-based similarity matrix, which was reclustered to identify clusters of similar consensus shapes,⁸⁰ which were again superposed and averaged to compute a small

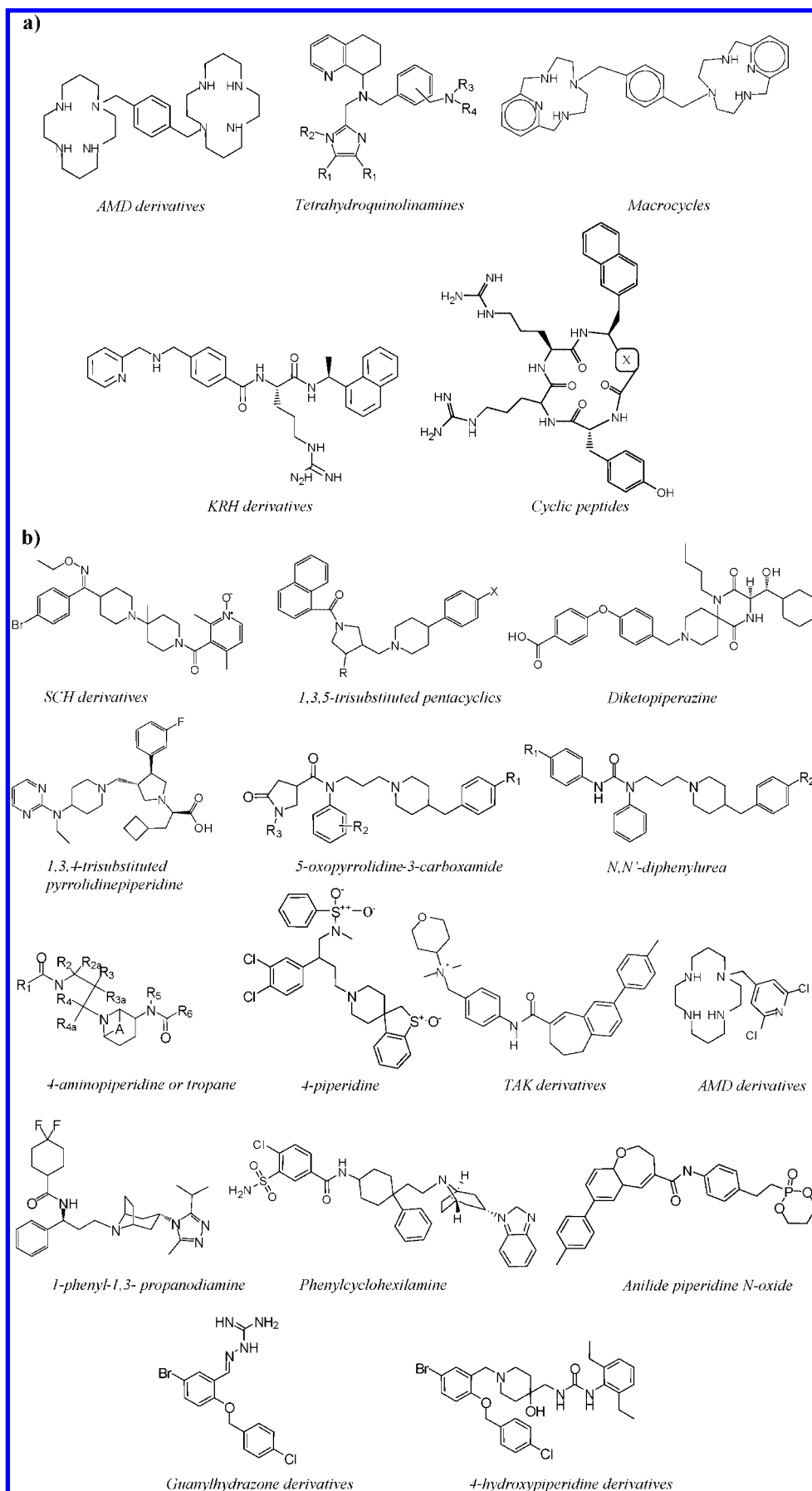


Figure 2. Representative structures of (a) five families of CXCR4 inhibitors and (b) fifteen families of CCR5 inhibitors.

number of SC shapes. In other words, the resulting SC SH shapes correspond to the shapes of pseudomolecules con-

structed from volumetric unions of fingerprint-based and shape-based subclusters of known actives.

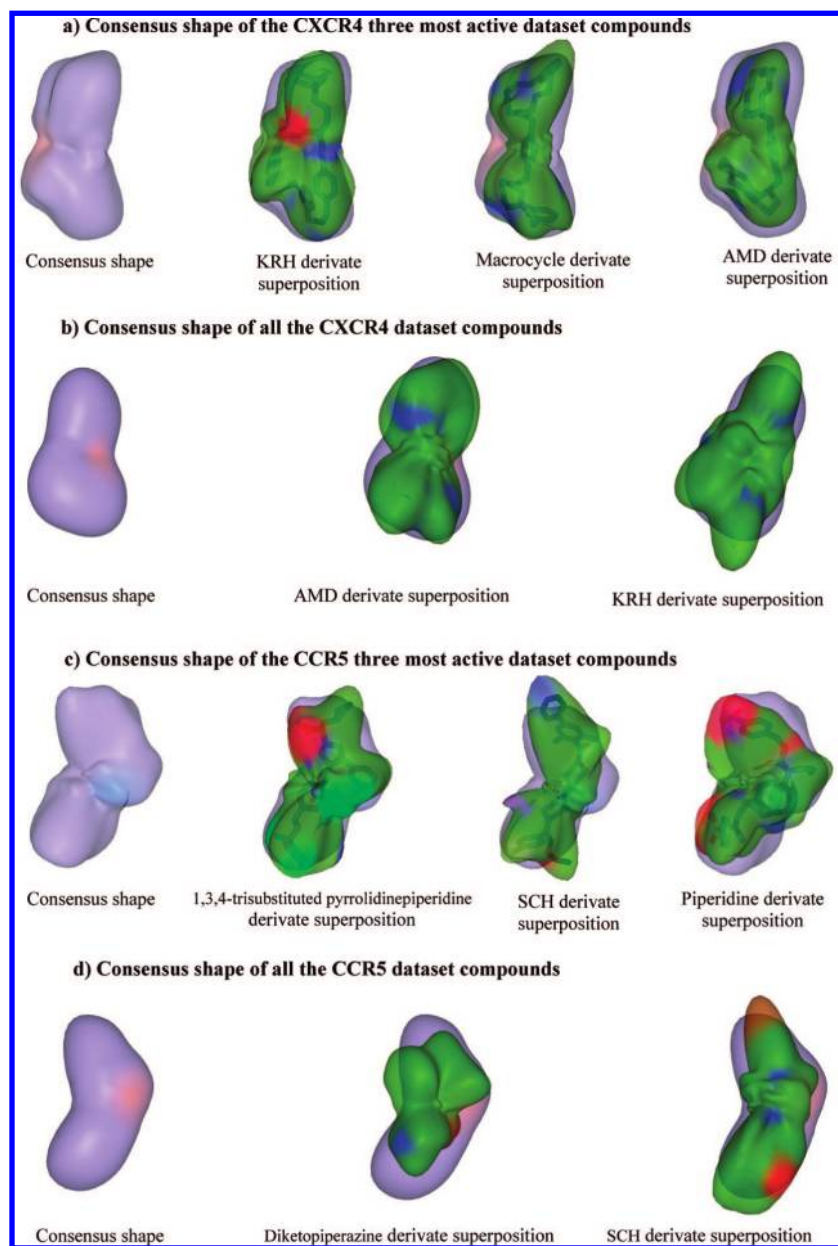


Figure 3. CXCR4 and CCR5 antagonist consensus shapes. (a) The image on the left shows the consensus shape calculated from the three most active compounds of different scaffold families in the CXCR4 inhibitor database: an AMD derivative, a macrocycle derivative, and a KRH derivative. The following three images show the superpositions of these compounds onto the consensus. (b) The consensus shape calculated from all CXCR4 database actives, and example superpositions onto the consensus of two randomly selected compounds (an AMD derivative and a KRH derivative). (c) On the left, the consensus shape calculated from the three most active compounds of different CCR5 database scaffold families: a 1,3,4-trisubstituted pyrrolidinepiperidine derivative, a SCH derivative, and a piperidine derivative. On the right, the superpositions of these compounds onto the consensus. (d) Consensus shape calculated from all CCR5 actives, along with example superpositions onto the consensus of two randomly selected actives (a diketopiperazine derivative and a SCH derivative).

To explore whether the computed SC pseudomolecules are sterically feasible in the context of the CCR5 extracellular pocket, each pseudomolecule was rigidly docked into our model-built CCR5 structure using blind Hex docking with default search parameters.²⁵ This structure was built by homology using bovine rhodopsin as template (PDB code 1HZX: 20% sequence identity and 35% similarity with respect to CCR5), as described previously.¹² To compare quantitatively the ability of the consensus shapes to identify known binders, VS was performed using our ligand database and the screening utility of each query shape was analyzed objectively using ROC analyses. Finally, VS results for CXCR4 and CCR5 consensus shape queries were compared to conventional ROCS 2.2,⁸¹ Hex 4.8,²⁵ and ParaFit 08²²

shape-matching VS, and to rigid-docking-based VS using Hex 4.8, AutoDock 3.0,⁸² GOLD 3.01,⁸³ and FRED 2.2.⁸⁴

RESULTS

CXCR4 and CCR5 Inhibitor Consensus Shapes. Figure 3a shows the consensus shape calculated from the three most active compounds of different scaffold families in the CXCR4 inhibitor database (an AMD derivative, a macrocycle derivative, and a KRH derivative). Figure 3b shows the consensus shape computed from all the CXCR4 inhibitors in our database. Visual inspection of these figures shows that the first consensus shape captures rather well the overall shape of the three selected inhibitors, whereas the all-

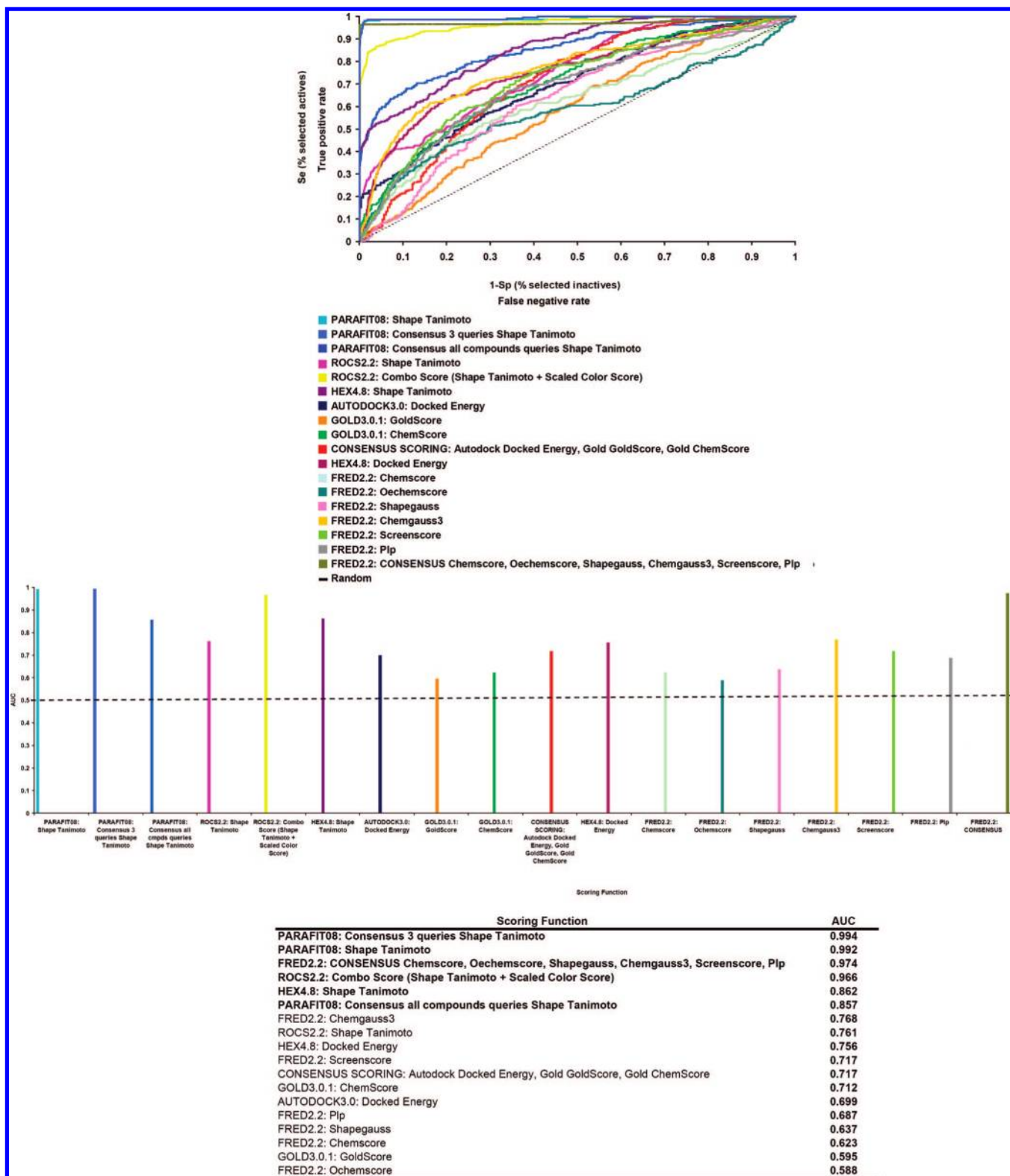












Figure 4. ROC plot validation of various shape-matching and docking VS methods compared to the consensus shape approach applied to CXCR4 antagonists. The dotted black line represents the expected enrichment if actives were selected at random. The lower bar chart and table report the AUC values obtained from the corresponding ROC curves. The scoring functions which give the best VS performance are shown in bold.

molecule consensus has much less local surface detail, yet still broadly retains the gross features of the member shapes. Figure 3c shows the consensus shape calculated for the three most active compounds of different scaffolds families in the CCR5 inhibitor database (a piperidine derivative, a SCH derivative, and a 1,3,4-trisubstituted pyrrolidinepiperidine derivative). Figure 3d shows the consensus shape of *all* the

CCR5 active inhibitors. In this case, it can be seen that using all database compounds to construct the consensus query causes a more spherical average shape than the CXCR4 inhibitors, because of the greater number and diversity of compounds in the CCR5 database.

CXCR4 Virtual Screening. Figure 4 shows the performance of the CXCR4 consensus shaped-based VS queries

Table 2. CCR5 Antagonist Clustering Results Using Ward's Clustering of Chemical Fingerprint Descriptors^a

CLUSTER	Compounds Found	Number of compounds	Consensus Shape
1	(8) <i>1,3,4-trisubstituted pyrrolidinepiperidines</i> (3) <i>1,3,5-trisubstituted pentacyclics</i> (5) <i>5-oxopyrrolidine-3-carboxamides</i> (4) <i>N,N'-diphenylureas</i> (2) TAK derivatives (1) 4-piperidines (1) others (MRK-1 CPMD 167)	24	
2	(1) <i>1,3,4-trisubstituted pyrrolidinepiperidines</i> (6) <i>1,3,5-trisubstituted pentacyclics</i> (13) 1-phenyl-1,3- propanodiamines (3) 4-piperidines (3) <i>AMD derivatives</i> (9) <i>Diketopiperazines</i> (1) SCH derivatives (2) Phenylcyclohexilamines (3) others (GSK, Merck2, Merck3)	41	
3	(22) <i>Anilide piperidine N-oxides</i> (1) TAK derivatives (1) others (1-benzazepine)	24	
4	(21) 1-phenyl-1,3- propanodiamines (5) Phenylcyclohexilamines	26	
5	(11) 1-phenyl-1,3- propanodiamines	11	
6	(12) 1-phenyl-1,3- propanodiamines	12	
7	(26) <i>4-aminopiperidine or tropanes</i> (6) 4-piperidines (2) Phenylcyclohexilamines (1) others (Merck1)	35	
8	(23) SCH derivatives	23	
9	(20) SCH derivatives	20	
10	(37) SCH derivatives	37	
11	(22) SCH derivatives	22	
12	(17) SCH derivatives	17	
13	(19) TAK derivatives	19	
14	(44) TAK derivatives	44	
15	(33) <i>Guanyldihydrazone derivatives</i>	33	
16	(36) <i>4-hydroxypiperidine derivatives</i>	36	

^a Kelley's method predicts 16 clusters as the optimal number. The number of compounds found for each family in each cluster is specified in parenthesis. The families marked in bold italics comprise the entire family in a unique cluster. The families marked in italics comprise the entire family between two clusters. The ten initial consensus shapes obtained after the grouping of clusters are also shown.

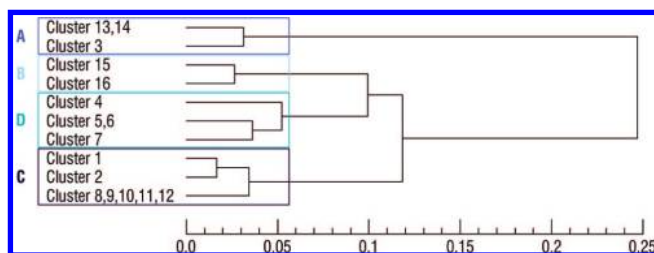
compared to docking-based and shape-based screening using a single high affinity ligand (AMD3100). This figure shows that the consensus shape queries give higher AUCs than the other approaches, although the single-ligand ParaFit query also performs well. As might be expected from consideration of Figure 1, the three-ligand consensus performs considerably better than the all-ligand consensus, due to the high degree of smoothing and loss of surface detail in the all-ligand shape. On the other hand, considering the very good performance of the single high affinity ligand and the marginally superior performance of the three-ligand query suggests that all three ligands share highly similar shapes (as confirmed by Figure 3a), which probably all bind in similar way within the CXCR4 pocket.

Regarding the shape matching approaches, ParaFit Shape Tanimoto, ROCS Combo Score, and Hex Shape Tanimoto all give comparable AUCs to the ParaFit consensus shape query. Of the docking tools, FRED Consensus gives the best AUC, followed by FRED Chemgauss3, Hex Docked Energy, and rank-by-rank docking Consensus Scoring, which gives a better enrichment than the individual Autodock Docked Energy, Gold GoldScore, and Gold ChemScore scoring functions. The ROCS Combo Score and FRED Chemgauss3 scoring functions both include descriptions of shape and molecular chemical properties. If the protein structures contain errors, as is likely with model-built structures, those docking functions that include terms that favor chemical complementarity might be expected to be more resilient to

Table 3. CCR5 Antagonist Clustering Results Using Ward's Clustering of 2D Pharmacophore Fingerprint Descriptors^a

cluster	compounds found	number of compounds
1	(3) <i>1,3,4-trisubstituted pyrrolidinepiperidines</i> (7) <i>1,3,5-trisubstituted pentacyclics</i>	48
2	(38) 1-phenyl-1,3- propanodiamines (6) <i>1,3,4-trisubstituted pyrrolidinepiperidines</i> (2) <i>1,3,5-trisubstituted pentacyclics</i> (14) 1-phenyl-1,3- propanodiamines (1) 4-aminopiperidine or tropanes (1) others (1-benzazepine)	24
3	(4) 1-phenyl-1,3-propanodiamines (8) 4-piperidines (5) 5-oxopyrrolidine-3-carboxamides (3) <i>Diketopiperazines</i> (5) anilide piperidine <i>N</i> -oxides (3) phenylcyclohexilamines (2) <i>N,N'</i> -diphenylureas (2) SCH derivatives (3) others (MRK-1 COMPD 167, Merck1, Merck2)	35
4	(17) 4-aminopiperidine or tropanes (1) phenylcyclohexilamines (11) SCH derivatives (1) others (Merck3)	30
5	(8) 4-aminopiperidine or tropanes (6) <i>Diketopiperazines</i> (2) anilide piperidine <i>N</i> -oxides	16
6	(15) anilide piperidine <i>N</i> -oxides (5) phenylcyclohexilamines	20
7	(18) SCH derivatives (2) <i>N,N'</i> -diphenylureas (1) 4-piperidines (1) others (GSK 108)	22
8	(59) SCH derivatives (2) TAK derivatives	61
9	(28) SCH derivatives	28
10	(2) SCH derivatives (22) TAK derivatives	24
11	(42) TAK derivatives	42
12	(1) <i>AMD derivatives</i> (1) 1-phenyl-1,3- propanodiamines (1) 4-piperidines (4) <i>guanyldihydrazone derivatives</i> (36) 4-hydroxypiperidine derivatives	43
13	(2) <i>AMD derivatives</i> (29) <i>guanyldihydrazone derivatives</i>	31

^a Kelley's method predicts 13 clusters as the optimal number. The number of compounds found for each family in each cluster is specified in parenthesis. The families marked in bold italics comprise the entire family in a unique cluster. The families marked in italics comprise the entire family between two clusters.

**Figure 5.** Dendrogram of the ten initial CCR5 antagonist groups clustered using Ward's clustering of spherical harmonic distances between the consensus surface shapes of each group. Four main SC groups, labeled A, B, C, and D, are recognized.

structural errors in the receptor. Therefore, it is perhaps not surprising that the FRED Chemgauss3 gives an AUC which is closer to that of the ligand-based scoring functions and

much better than the other docking scoring functions such as FRED Shapegauss, Chemscore, Oechemscore, Shapegauss, Chemgauss3, Screenscore, and Plp.

Overall, the enrichment results for CXCR4 show that ligand-based shape-matching approaches provide better VS performance than structure-based docking tools, except for FRED Consensus Scoring, which gives considerably better enrichments than the other FRED scoring functions. This indicates that inaccuracies probably exist in the homology-built structure of the receptor and that, consequently, the use of ligand-based techniques should provide a more reliable way to identify new inhibitors for this target. It is also worth mentioning that in all docking studies the protein structure was assumed to be rigid, which in reality is not true. The quality of any docking calculation will intrinsically depend on the conformation of the receptor and especially the conformations of the side chains lining the binding region. Hence keeping the protein rigid is potentially a large source of uncertainty that can further influence the performance of docking-based VS calculations.

Clustering Known CCR5 Inhibitor Families. Table 2 shows the result of clustering the CCR5 inhibitors using Ward's clustering of chemical (topological) fingerprints. In this case, Kelley's method gives the optimal number of clusters as 16. Table 3 gives the corresponding results for clustering using 2D pharmacophoric fingerprints, which gives just 13 more tightly grouped clusters according to Kelley's method. Inspection of these clusters shows that the pharmacophoric fingerprint clustering tends to distribute compounds from different chemical families into more different clusters, whereas clustering on chemical fingerprints tends to group the compounds more closely according to the known inhibitor families. For example, all members of two entire families are assigned to a single cluster in two cases (i.e., the 5-oxopyrrolidine-3-carboxamides and *N,N'*-diphenylureas are entirely assigned to cluster 1, and the AMD derivatives and diketopiperazines are entirely assigned to cluster 2), and the members of several other families are entirely assigned to separate clusters (i.e., the members of the anilide piperidine *N*-oxides, 4-aminopiperidine or tropanes, guanyldihydrazone derivatives and 4-hydroxypiperidine derivatives are all assigned to separate clusters). Furthermore, chemical fingerprint clustering nicely separates the 1-phenyl-1,3-propanodiamines and the SCH and TAK families into different clusters depending on their different R-groups. Hence chemical fingerprint-based clustering was selected as the most appropriate point from which to proceed. Further inspection of these clusters shows that by grouping clusters 5 and 6 and similarly grouping clusters 8, 9, 10, 11, and 12, and also clusters 13 and 14, a total of just ten clusters are obtained which correctly groups together all the compounds belonging to a given scaffold family. Hence a total of ten CCR5 inhibitor clusters were selected for further analysis using the consensus shape-based approach.

Calculating Consensus and Super-Consensus CCR5 Inhibitor Clusters. SH consensus surface shapes were calculated for each of the ten selected clusters, as described in Methods (eq 6). An all-against-all SH comparison of each consensus surface was calculated using ParaFit, and the resulting pairwise Tanimoto similarity coefficients were used to calculate consensus superclusters using a further round of Ward's hierarchical clustering. Figure 5 shows a dendro-

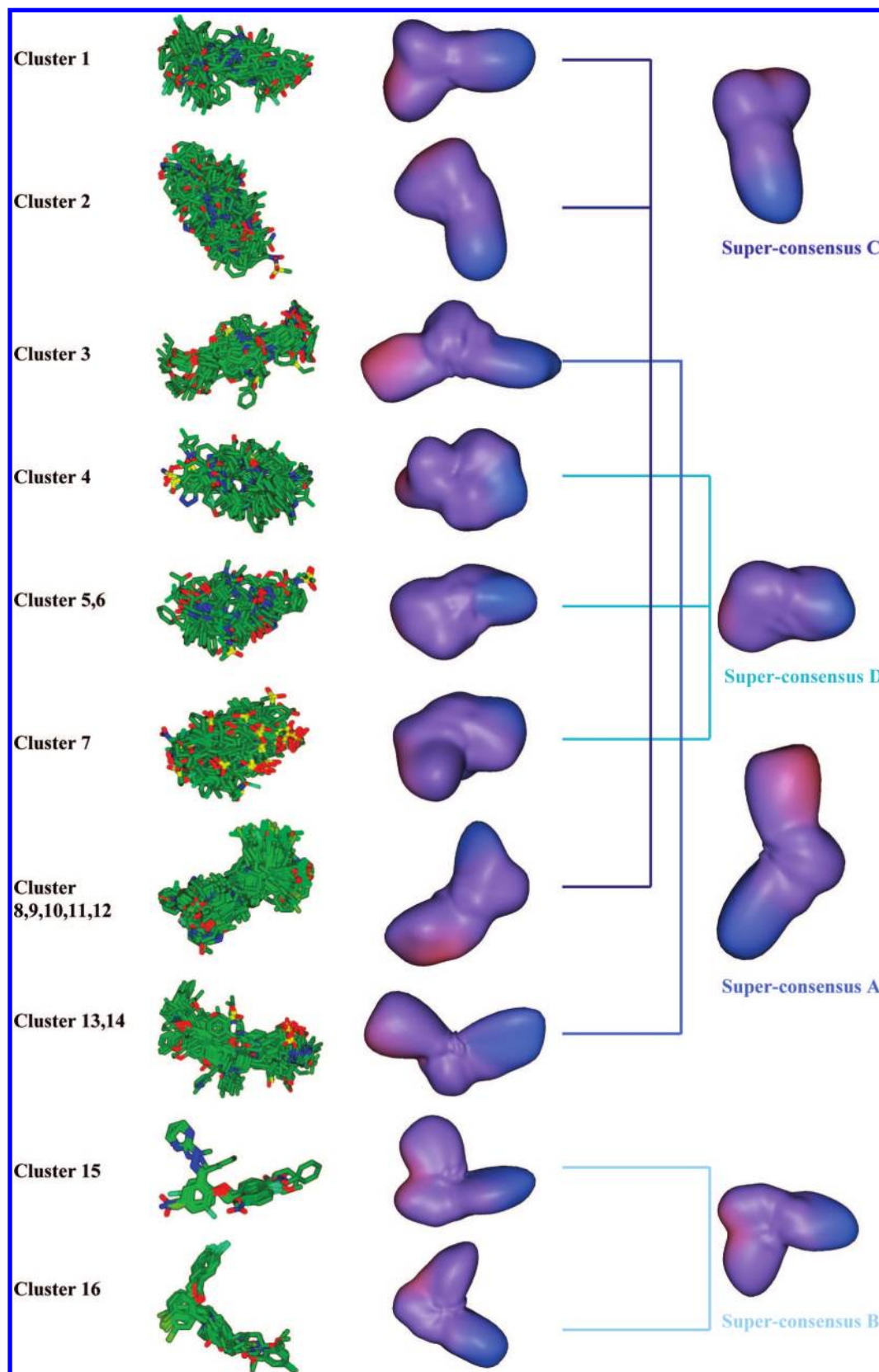


Figure 6. Molecular superpositions and consensus shapes of the ten Ward's clusters used to calculate the final SH SC shapes.

gram of the resulting SCs in which the initial ten consensus shape surfaces are clustered to give four main representative SC groups, A, B, C, and D. Figure 6 shows the 3D molecular overlays, the SH shapes of the 10 fingerprint clusters, and the SC shapes calculated from the clustered consensus surfaces. All molecular orientations shown in this figure were

derived directly from the SH consensus surface shape superposition calculations. If it is supposed that the calculated superclusters correspond to four fundamental families of inhibitors, it might further be hypothesized that these fundamental families bind within different regions of the CCR5 pocket. Figure 7 shows a schematic illustration of this

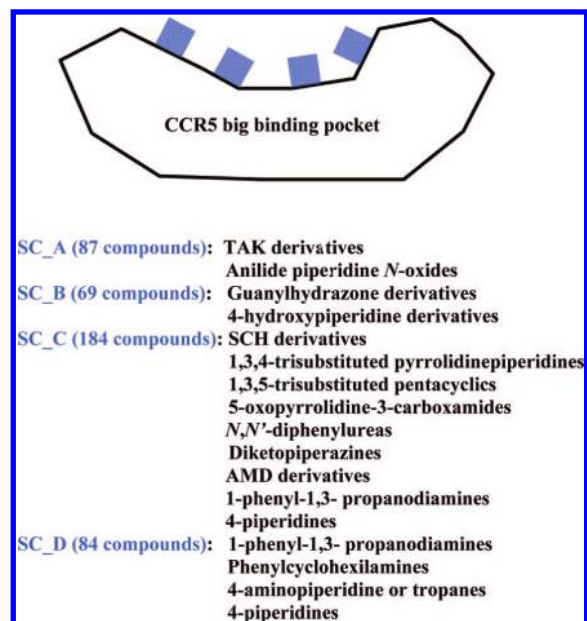


Figure 7. Schematic illustration of the hypothesized binding regions suggested by SC clustering. For each SC, the number of compounds used to construct the consensus and the family to which they belong are given.

hypothesis, along with the calculated scaffold family membership of each fundamental supercluster.

CCR5 Scaffold Family Virtual Screening. SH consensus surface shapes were also calculated for each of the 15 CCR5 scaffold families in the database. Results for the comparison of the VS performance of the SC scaffold queries are shown in Figure 8. It can be seen that the consensus query for each family performs quite well individually, except those for 4-hydroxypiperidine and guanylhydrazone derivatives. These two SC B family queries give poor AUCs because they have somewhat different molecular and consensus shapes than those of the other inhibitor families (see Figure 6).

As can be seen from the lower table in Figure 8, the scaffold family consensus AUCs correspond very well to the proposed SC clusters. More specifically, it can be seen that ordering the individual family consensus groups by AUC is almost sufficient to map each family member directly to its proposed SC. For example, the large group of high AUC scaffolds maps to SC C (namely, the SCH derivatives, 5-oxopyrrolidine-3-carboxamides, diketopiperazines, 1,3,4-trisubstituted pyrrolidinepiperidines, *N,N'*-diphenylureas, 1,3,5 trisubstituted pentacyclics, 4-piperidines, and AMD derivatives). The next group of good AUC scaffolds maps to SC A (i.e., the anilide piperidine *N*-oxides and the TAK derivatives). Similarly, the final low AUC scaffold families map to SC groups D (the 4-aminopiperidine and 1-phenyl-1,3-propanodiamines derivatives) and B (the 4-hydroxypiperidine and guanylhydrazone derivatives). The only exception to this AUC-based mapping is the phenylcyclohexilamine family which has a very high AUC, but the clustering of which places it in SC group D.

CCR5 Inhibitor Consensus Shape Virtual Screening. Figure 9 shows the VS results obtained using the four SC shapes as queries. It can be seen that SC C gives the best overall VS performance with an AUC of 0.91. It is perhaps not surprising that this SC query performs very well because it includes the three most active compounds in the database and also a large number of other actives (i.e., 184/424) with

similar shapes to the 4-piperidine derivatives, SCH derivatives, and 1,3,4-trisubstituted pyrrolidinepiperidine derivatives. The SC A query (87/424 actives) also performs rather well with an AUC of 0.79, and the SC D query (84/424 actives) performs reasonably well (AUC = 0.63). However, the ROC plot for SC B shows that this query exhibits good sensitivity and selectivity in the first percentages of the database screened, but the overall AUC is low (0.41) because the database contains relatively few members of the two SC B families (i.e., a total of only 69/424). However, if the members of clusters B and D are grouped together to form a single SC, as might be suggested by the dendrogram in Figure 5, the screening performance becomes essentially random (AUC = 0.51). Thus, despite the small populations of these two groups, their members have significantly different overall shapes, and they should be classified as two distinct structural groups for VS purposes. Performing a similar exercise with other combinations of SC clusters shows similar but less dramatic reductions in AUCs compared to the AUCs of the unmerged clusters. For example, merging A and B gives AUC = 0.65, merging A and D gives AUC = 0.65, and merging A and C gives AUC = 0.87 (compared to the original AUCs of A = 0.79, B = 0.41, C = 0.91, and D = 0.63). This behavior supports the chemical fingerprint clustering results which suggest that the CCR5 inhibitor families may be clustered into no fewer than four main groups. A similar SC cluster analysis was performed for the pharmacophore fingerprint clusters (details not shown), and this also indicated no less than four main SC families with AUCs of 0.79, 0.43, 0.94, and 0.74 for SC clusters A, B, C, and D (the cluster members of B, C, and D differed slightly from the chemical fingerprint analysis). Hence both clustering approaches ultimately indicate that the CCR5 antagonists may be grouped into four main SC families.

Although it is impractical to generate and test large numbers of different potential SC clusters, we wished to ensure that the VS performance of the selected four clusters was not being dominated by a small number of high affinity actives. Hence the AUCs for SC clusters A, B, C, and D were recalculated with the three most active CCR5 inhibitors removed from each cluster. For SC A, the recalculated AUC was unchanged (0.79). For SC C, the AUC was reduced only marginally from 0.91 to 0.90. For SCs B and D, the AUC was reduced from 0.41 to 0.38 and from 0.63 to 0.61, respectively. The different behavior of the A and C clusters compared to the B and D clusters seems to be because the remaining compounds in the B and D clusters have lower activities than those that were removed, whereas the reduced A and C clusters still contain several other molecules with high activities. Nonetheless, because the observed reduction in AUC is either small or modest in all cases, it may be concluded that the original four SC clusters seem to capture very well the general features of many high affinity binders.

Figure 10 shows the ROC plot analysis of the consensus shape-matching VS approach applied to the CCR5 inhibitor database. The consensus shape constructed from the three most active compounds gives the best VS performance (AUC = 0.99), followed by SC C, which comprises the families containing the greatest number of active compounds. As with the CXCR4 inhibitors, the consensus query constructed with the three most active compounds achieves higher perfor-

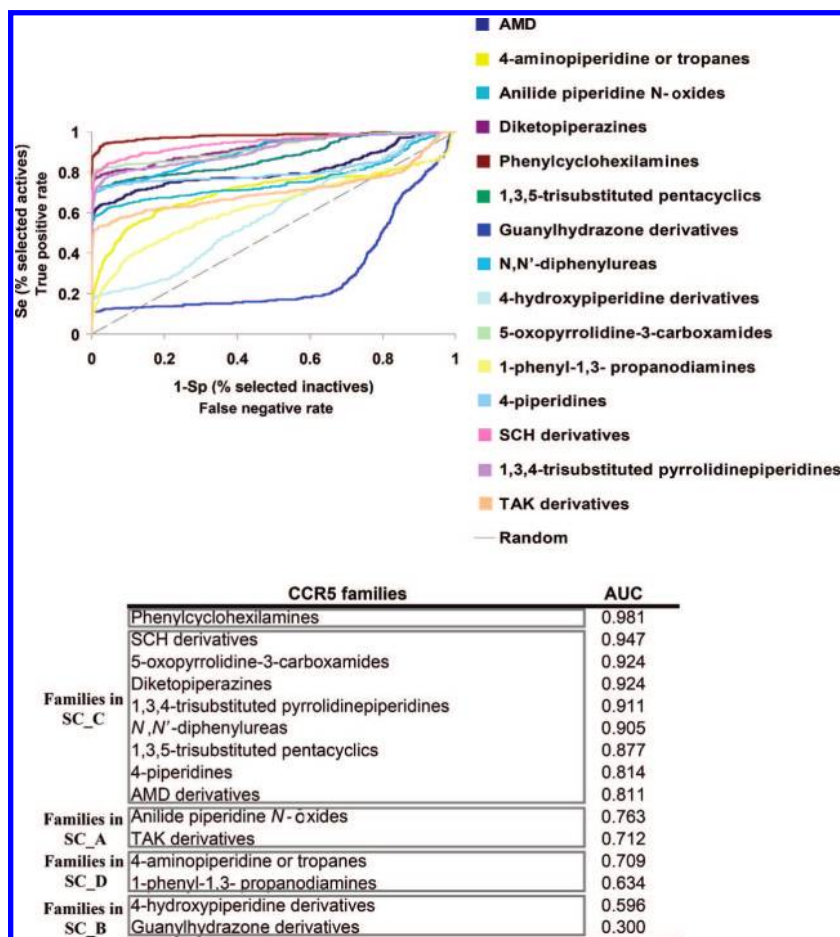


Figure 8. ROC plot evaluation of CCR5 scaffold family consensus shape-matching VS. The dotted black line represents the expected enrichment if actives were selected at random. The lower table reports the AUC values obtained for the consensus ROC curves for the different scaffold families, grouped according to their assigned SC clusters. This shows that the consensus families that give the highest AUC values belong to SC C, followed by those belonging to SC A, D, and B.

mance than the query constructed using all CCR5 inhibitors (AUC=0.87). This supports the notion that using too many molecules to make a consensus shape causes an undesirable degree of surface shape smoothing and the loss of important surface details. Figure 11 shows the ROC plots for all shape-matching VS approaches for the CCR5 inhibitor database. It can be seen that the consensus shape queries generally give larger AUCs than ROCS, Hex, and the single-query ParaFit queries.

Comparing Shape Based and Docking Based CCR5 Virtual Screening. Figure 12 shows the VS performance of selected CCR5 consensus shape queries and several conventional ligand based shape matching and receptor based docking approaches. It can be seen that the three-ligand consensus and SC C queries give the best overall screening performance, along with the FRED consensus scoring and Autodock docking methods, both of which also perform very well even though they do not take into account protein flexibility. As observed previously, the SC A query also performs well, but the SC B and D queries and combinations thereof give rather poor VS results. However, it is interesting to note that the Gaussian-based docking scoring functions in FRED generally give better VS performance than the Gaussian-based superposition scoring functions of ROCS. Comparing Figures 12 and 4 shows that the docking and single-query shape-matching results are generally better for CXCR4 than for CCR5. However, the AUCs for the CCR5

consensus-based query results are almost at the same high level as for CXCR4. Hence, despite CCR5 having a larger binding pocket and a much more diverse set of inhibitors than CXCR4, the use of consensus-based queries can be seen to find many CCR5 inhibitors remarkably well.

Blind Docking Super-Consensus Pseudomolecules. Figure 13 shows the results obtained for blind docking the SC pseudomolecules into the CCR5 extracellular pocket. It can be seen that the SC A pseudomolecule is docked onto one side of the CCR5 binding pocket (Site 1) near residues Ala29, Arg31, Leu33, Tyr37, Thr82, Trp86, Tyr108, and Glu283, delimited by transmembrane (TM) loops 1, 2, 3, and 7, whereas the SC C pseudomolecule is docked onto the opposite side of the pocket (Site 2) near residues Tyr108, Phe113, Ile198, Ile200, Asn252, Glu283, and Glu286, delimited by TM loops 3, 5, and 6. The SC B and SC D pseudomolecules are docked onto the central region of the binding pocket (Site 3) near residues Tyr108, and Glu283, delimited by TMs 3, 6, and 7, thus overlapping the predicted SC A and C binding sites. Figure 14 shows more detailed views of the calculated docking modes. Docking the four SCs derived from pharmacophore fingerprint clustering also gives similar binding modes (details not shown). Thus our docking calculations consistently suggest the existence of two or three main binding sites within the CCR5 pocket. These consensus-based docking predictions are consistent with experimental data.^{11,13,14,16}

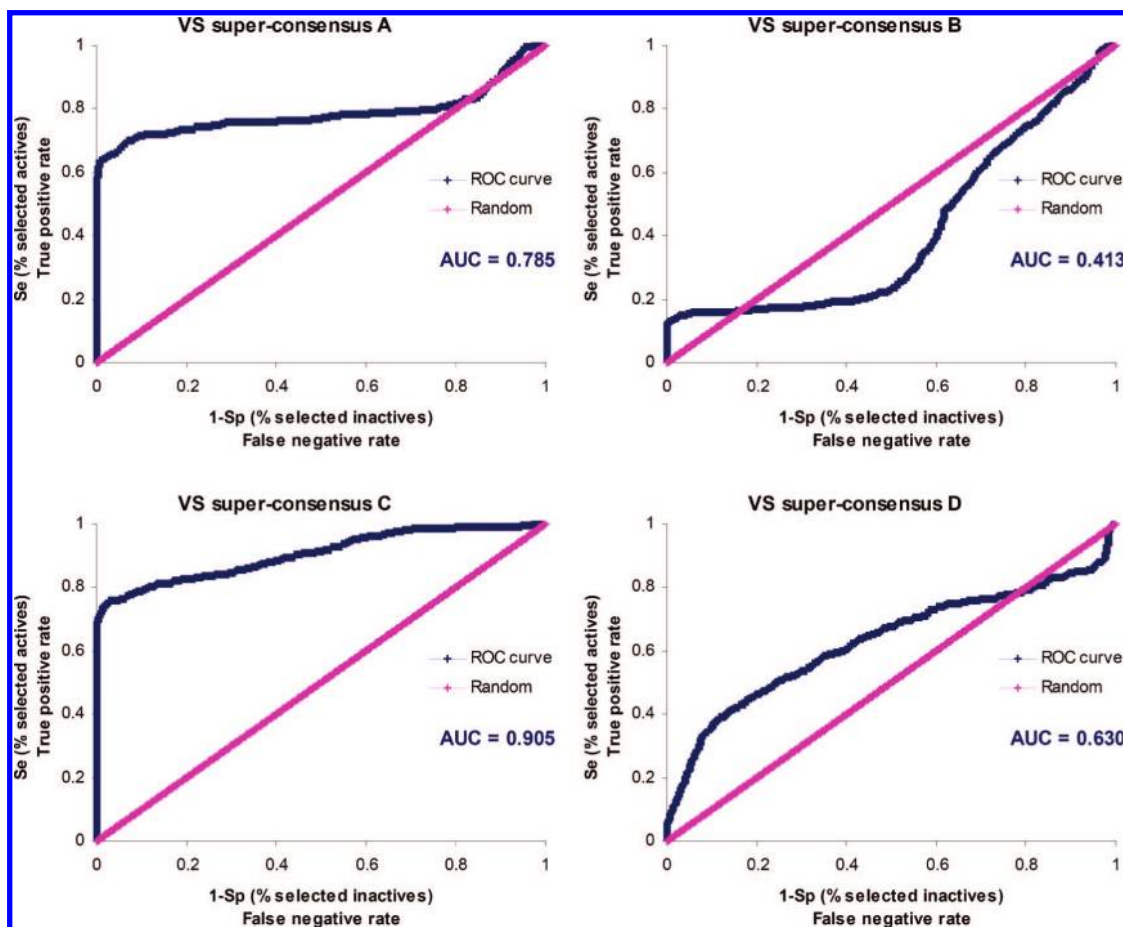


Figure 9. ROC plot validation of the CCR5 inhibitor SC pseudomolecules. The AUC is given for each SC, A, B, C, and D. The diagonal pink line represents the expected enrichment if actives were selected at random.

To confirm that the SC queries are properly matched with their predicted target sites, the three proposed binding sites were each treated as if they were separate targets for docking-based VS using rigid body docking of the corresponding SC pseudomolecules. In other words, when docking to Site 1, compounds belonging to SC A were treated as actives, and compounds belonging to SC B, C, and D were treated as inactives. In a similar manner, when docking to Site 2, compounds belonging to SC C were treated as actives, and compounds belonging to SC A, B, and D were treated as inactives. Similarly for Site 3, compounds belonging to SC B and D were treated as actives, and compounds belonging to SC A and C were treated as inactives. Figure 15 shows the docking VS performance for each of the three proposed CCR5 binding regions. Comparing Figures 15 and 9, it can be observed that docking VS onto Sites 1, 2, and 3 (AUC = 0.83, 0.96, and 0.85, respectively) improves the SC A, C and B/D shape matching AUCs (AUC = 0.79, 0.91, and 0.41/0.63, respectively). Given that SCs A and C already give good shape matching enrichments, reassigning the B and D members as inactives only marginally improves the corresponding AUCs. However, treating the large set of C and A members as inactives for Site 3 gives much higher AUCs for the SC B and D queries, which clearly supports the notion that the CCR5 antagonists bind to at least three main sites within the extracellular pocket.

DISCUSSION

The results of this study show that spherical harmonic consensus shapes can provide effective 3D query structures

for shape-based VS. For the CXCR4 and CCR5 ligands studied here, our results show that well-chosen consensus shape queries can give better (CXCR4) or significantly better (CCR5) virtual screening enrichments than conventional single-molecule VS queries. The CXCR4 results show that consensus shape based queries give higher AUCs (i.e., better enrichments) than conventional ligand-based and rigid-body receptor-based screening approaches. However, for CXCR4, these results are nonetheless broadly similar to the basic ParaFit one-molecule shape-matching approach because the inhibitors for this target share rather similar molecular shapes which individually match quite well the selected query shape. For CCR5, which has a much larger and more diverse set of inhibitor families, the SC family C and the SC all-family queries both give very good overall VS performance. However, this seems to be at least partly because a high proportion of all scaffold families cluster into the family C superconsensus grouping. Hence, by construction, the spherical harmonic consensus shape derived from these family members provides a single representative pseudomolecular shape which recognizes well many of the individual member structures.

Regarding the more challenging problem of understanding how so many diverse inhibitor families might bind within the CCR5 pocket, our consensus shape-based approach provides a straightforward way to identify clusters of inhibitor families from a large set of known actives which is broadly consistent with current experimental SDM data.^{11,13,14,16} More specifically, our clustering results indi-

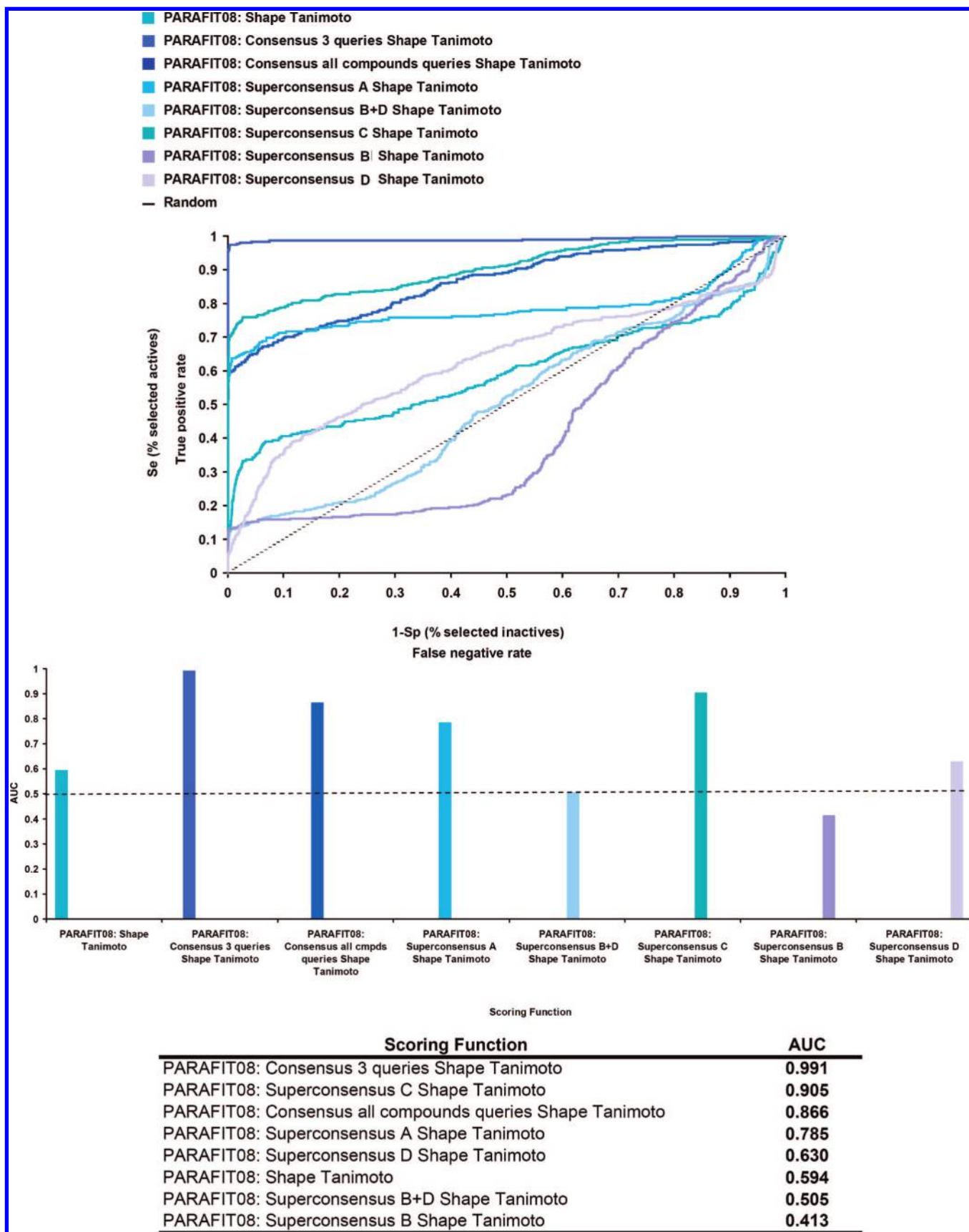


Figure 10. ROC plot evaluation of consensus shape-matching VS for the CCR5 antagonists. The dotted black line represents the expected enrichment if actives were selected at random. The lower bar chart and table report the AUC values of the corresponding VS ROC curves.

cate that the CCR5 inhibitors are in fact described very well by four main consensus families, of which SC family C is the most highly populated. Our docking results suggest that

the families of compounds belonging to SC A bind within Site 1, and this is consistent with SDM-based experimental results for TAK derivative binding.^{11,13,18–21,85} Furthermore,

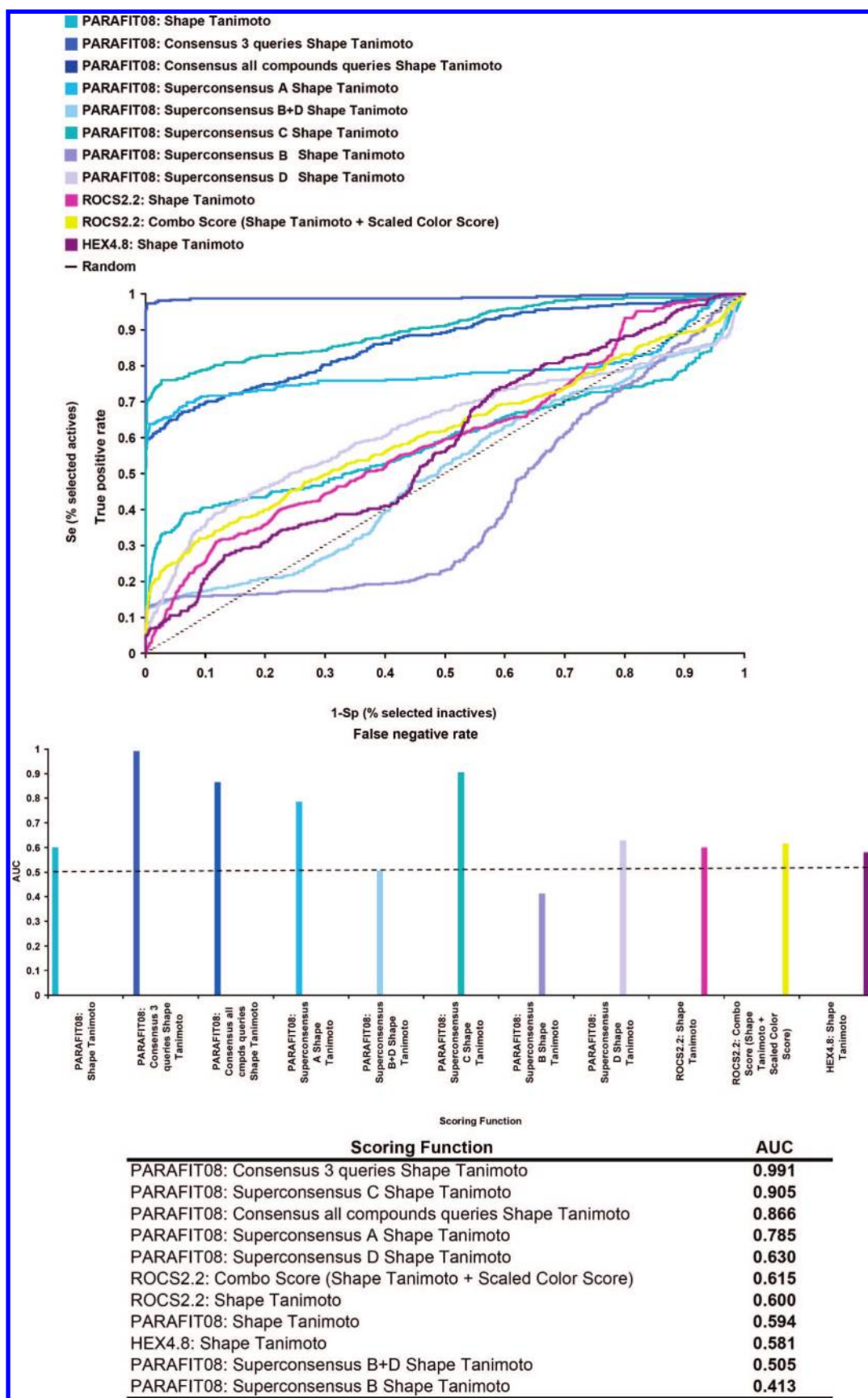


Figure 11. ROC plot comparison of various shape-matching VS methods for the CCR5 antagonists. The dotted black line represents the expected enrichment if actives were selected at random. The lower bar chart and table report the AUC values of the corresponding VS ROC curves.

our docking results suggest that SC C ligands bind within Site 2, and this is consistent with published experimental

results for CCR5 binding of certain SCH derivatives, 1,3,4-trisubstituted pyrrolidinepiperidine derivatives, and diketopi-

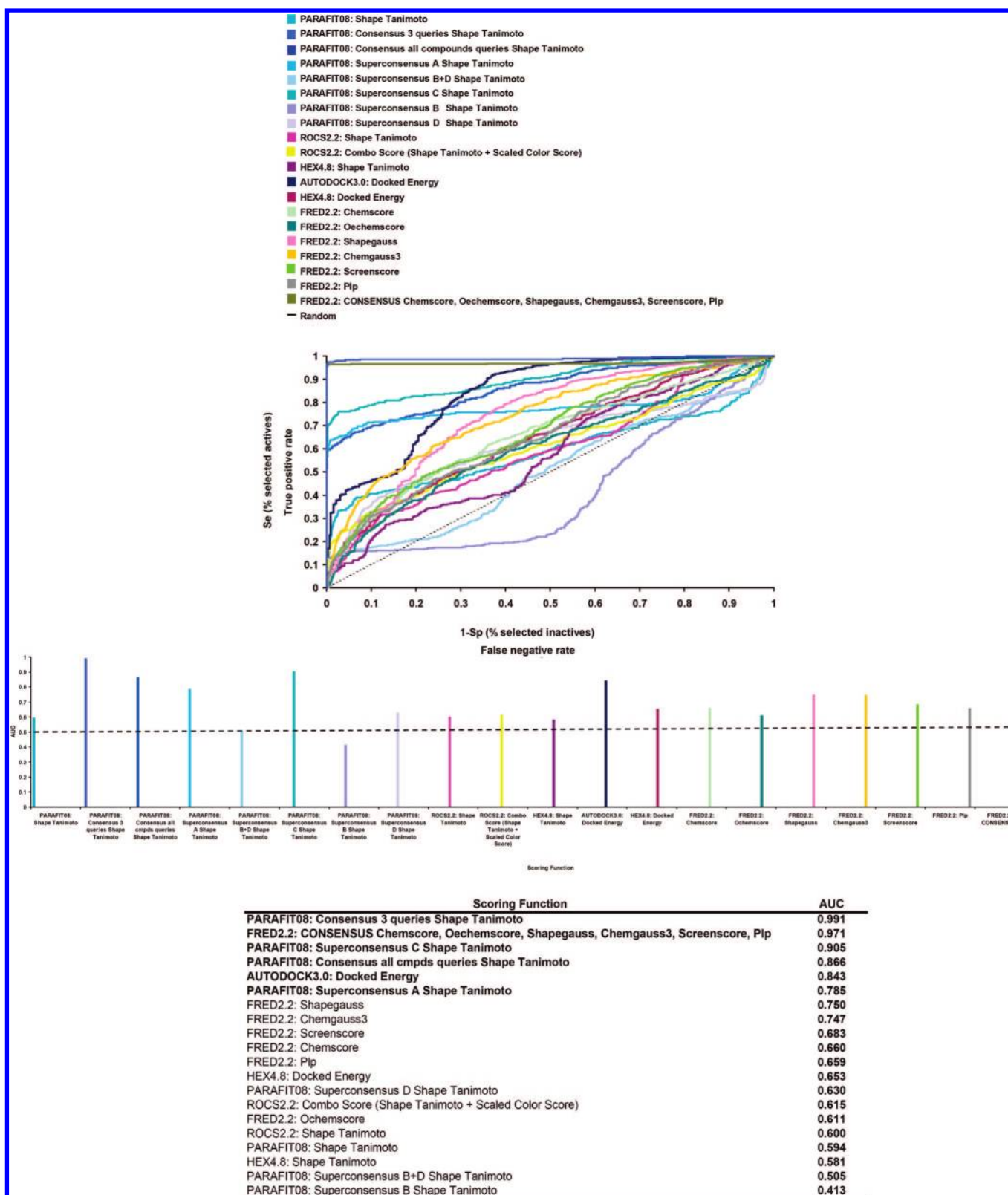


Figure 12. ROC plot validation of various shape-matching and docking VS methods compared to the consensus shape approaches for the CCR5 antagonists. The dotted black line represents the expected enrichment if actives were selected at random. The lower bar chart and table report the AUC values of the corresponding VS ROC curves. The scoring functions which give the best VS performance are shown in bold.

perazines.^{11,13,15–17,85} To our knowledge, there is not yet any experimental SDM evidence to relate compounds belonging to SC B and D to any specific binding site. However, previous docking predictions by Kellenberger et al. suggest a binding mode which includes both Site 1 and

Site 2,¹⁴ and this would be consistent with our prediction of Site 3, which is spatially located between Sites 1 and 2. Overall, the clusters and SC clusters found using our SH-based approach, and the direct correspondence of these with the spatial locations of the three binding sites predicted by

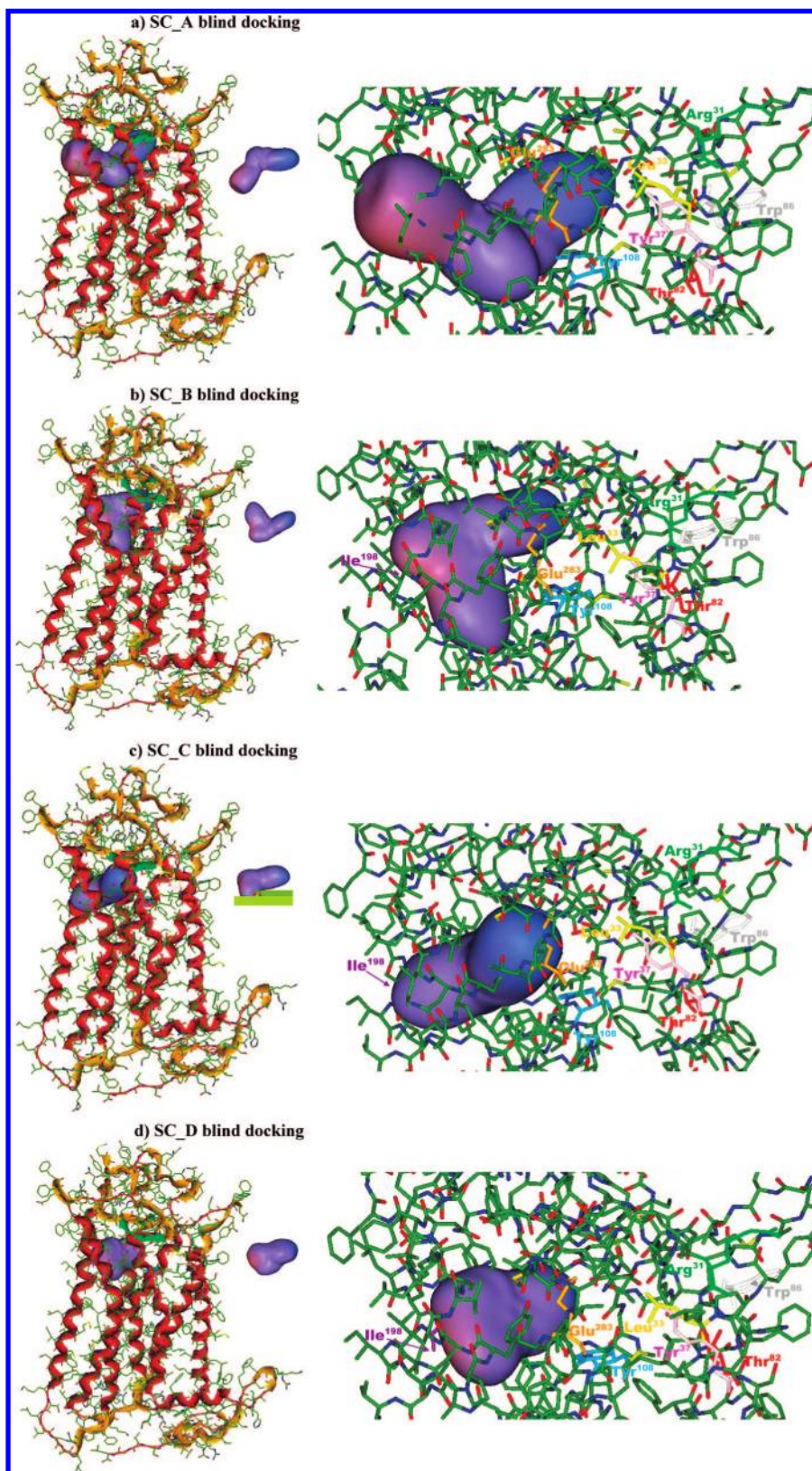


Figure 13. Hex blind docking results for the SC pseudomolecules. The images on the left show the final docked position of the SC pseudomolecules. The images on the right show close-up views of the docked conformations, annotated with the locations of known SDM binding site residues. In each case, the pseudomolecule was initially placed outside the CCR5 receptor pocket, as shown. (a) SC A blind docked onto one side of the CCR5 pocket. (b) SC B blind docked in the middle region of the pocket. (c) SC C blind docked onto the opposite side of the binding pocket. (d) SC D blind docked in the middle region of the pocket.

rigid-body pseudomolecule docking, are clearly consistent with and add weight to the previous computational predictions of Kellenberger et al. and also recently by Kondru et al.⁸⁵ In these earlier studies, different clinical drug candidates

were used to establish the nature of the binding pocket in CCR5. Although all CCR5 antagonists were predicted to bind to the same main hydrophobic pocket, in agreement with our results, both previous studies indicate that ligands may

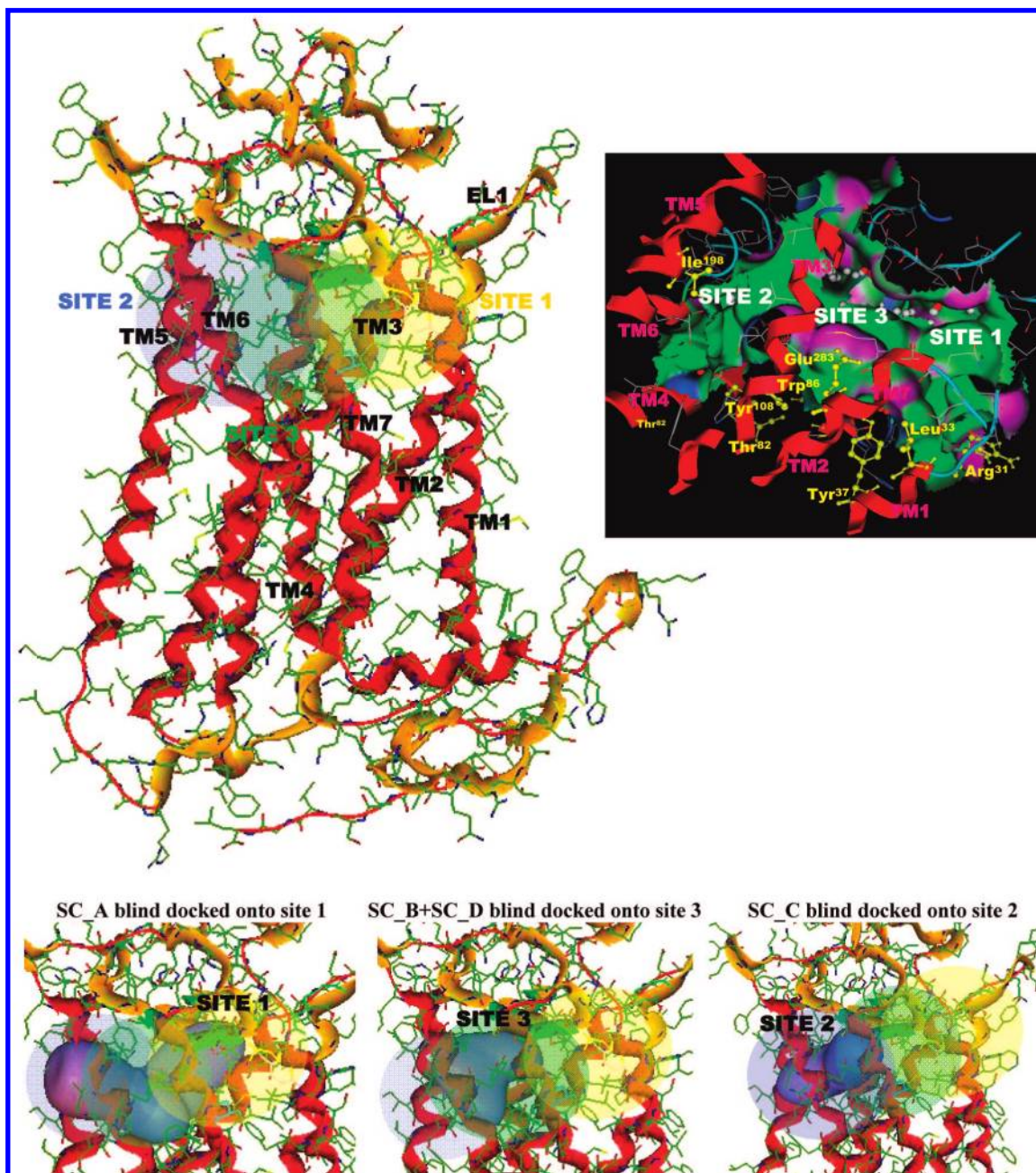


Figure 14. CCR5 binding pocket subsites proposed by the consensus VS and docking results. Here, the SC A pseudomolecule is docked onto the first subsite (Site 1), delimited by TMs 1, 2, 3, and 7. The SC C pseudomolecule is docked onto a second subsite (Site 2), delimited by TMs 3, 5, and 6. The SC B and SC D pseudomolecules are docked onto a third subsite (Site 3), delimited by TMs 3, 6, and 7, and which overlaps the SC A and SC C binding subsites. The top right image shows the CCR5 model with the proposed binding regions specified. On the top left, the van der Waals interaction surface of the CCR5 receptor cavity colored by H-Bonding (purple), hydrophobicity (green), and mild polar (blue) regions. TMs and important binding residues delimiting the three binding regions are shown in red and yellow, respectively. The bottom row of images show close-up views of the SC pseudomolecules in the three proposed subsites.

occupy different subcavities. This is clearly demonstrated by the different CCR5 mutant binding profiles obtained by Kondru et al., which is consistent with the significantly different electrostatic shapes and polarities of the CCR5 antagonists analyzed. Their docking predictions, which are based on SDM and CCR5 homology modeling data, suggest that the CCR5 receptor can accommodate structurally and electrostatically diverse antagonists by utilizing a unique set of interactions for every ligand, which is also consistent with our clustering results.

Nonetheless, the only completely reliable way to verify the validity of docking-based predictions is through comparison with a known crystallographic structure. Clearly, such

a gold standard reference is not available for CCR5. Hence any comparison with previous docking studies can, at best, serve only to add further support to the original prediction. On the other hand, for practical purposes, an unbiased and objective way to validate a structural prediction even when no crystal structure is available is to test its utility in the context of VS. The fact that the VS results obtained here using consensus and four SC shape-based similarity and docking queries give significantly enhanced screening enrichments compared to single-molecule based queries lends very strong support both to the initial validity of the notion of SC structures, and to the hypothesis that the members of

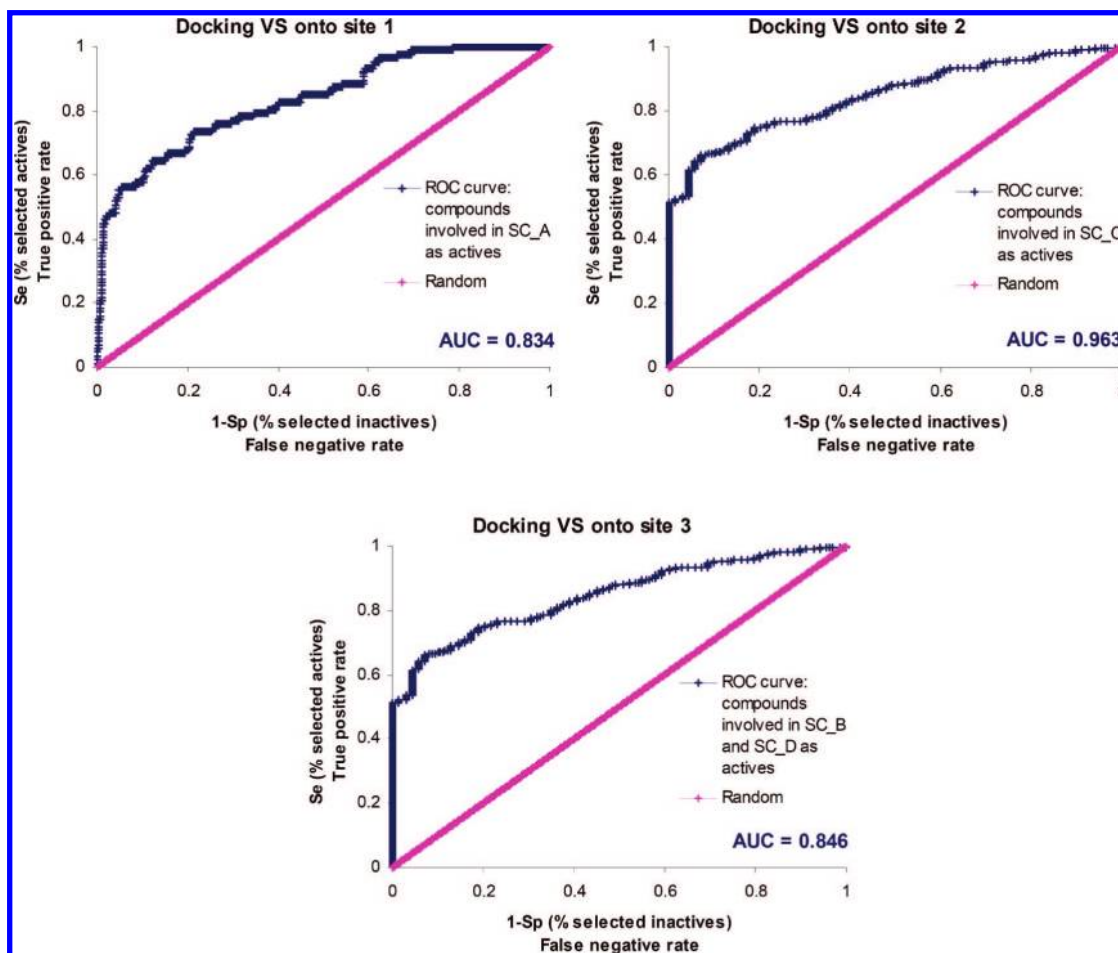


Figure 15. ROC plot validation of docking VS onto the three identified CCR5 subsites for the CCR5 antagonists. The AUC is given for each docking VS (onto Site 1, 2, and 3). The diagonal pink line represents the expected enrichment if actives were selected at random. For docking VS onto Site 1, compounds belonging to SC A are treated as actives, and compounds belonging to SC B, C, and D are treated as inactives. For docking VS onto Site 2, compounds belonging to SC C are treated as actives, and compounds belonging to SC A, B, and D are treated as inactives. For docking VS onto Site 3, compounds belonging to SC B and D are treated as actives, and compounds belonging to SC A and C are treated as inactives.

these SC clusters bind within at least three main sites in the CCR5 extracellular pocket.

CONCLUSION

This study has shown that using spherical harmonic consensus shapes as queries can be a useful strategy to improve hit enrichments in shape-based VS. We have developed a straightforward and fast method to construct consensus molecular shapes from SH surface envelopes. This consensus shape approach has been applied and validated by VS using a database of CXCR4 and CCR5 antagonists. For both receptor targets, ROC plot analyses show an improvement of VS results using the new approach. Moreover, the CCR5 multiple-binding-region hypothesis has been quantitatively explored by constructing different trial SH consensus query structures and by measuring their VS utility against our CCR5 inhibitor database. This study found four main SC clusters whose members are predicted to bind to three different but somewhat overlapping sites within the CCR5 pocket. The good VS results obtained with these virtual structures suggest they may profitably be used to search for novel inhibitors in prospective VS campaigns against other databases. Pseudomolecules corresponding to these SC clusters were docked into the CCR5 pocket, and

the locations of these positions were related to the locations predicted by previous docking studies. Several compounds within each consensus group have experimentally supported or computationally predicted binding modes which are consistent with the locations of the SC clusters docked here. Therefore, the SC structures calculated here provide strong supporting evidence for the CCR5 multiple-ligand-binding-site hypothesis, and help to give a better picture of how the CCR5 antagonists are probably distributed in the CCR5 receptor pocket.

ACKNOWLEDGMENT

The authors are grateful to OpenEye Scientific Software Inc., ChemAxon, and Cepos Insilico Ltd. for providing Academic Licences for ROCS, JChem, and ParaSurf, respectively. V.I.P.N. thanks the Generalitat de Catalunya—DURSI for a grant within the Formació de Personal Investigador (2008FI) program. This work was supported by The TV3 Marathon Foundation (AIDS-2001) promoted by the Catalan Radio and Television Corporation (Corporació Catalana de Ràdio i Televisió, CCRTV) and the Programa Nacional de Biomedicina (Ministerio de Educación y Ciencia, SAF2007-63622-C02-01).

REFERENCES AND NOTES

- Jiang, S.; Lin, K.; Strick, N.; Neurath, A. R. Inhibition of HIV-1 infection by a fusion domain binding peptide from the HIV-1 envelope glycoprotein GP41. *Biochem. Biophys. Res. Commun.* **1993**, *195*, 533–538.
- Kadow, J.; Wang, H. G.; Lin, P. F. Small-molecule HIV-1 gp120 inhibitors to prevent HIV-1 entry: An emerging opportunity for drug development. *Curr. Opin. Invest. Drugs* **2006**, *7*, 721–726.
- Berger, E. A.; Murphy, P. M.; Farber, J. M. Chemokine receptors as HIV-1 coreceptors: Roles in viral entry, tropism, and disease. *Annu. Rev. Immunol.* **1999**, *17*, 657–700.
- Markovic, I.; Clouse, K. A. Recent advances in understanding the molecular mechanisms of HIV-1 entry and fusion: revisiting current targets and considering new options for therapeutic intervention. *Curr. HIV Res.* **2004**, *2*, 223–234.
- De Clercq, E. New antiviral agents in preclinical or clinical development. *Adv. Antiviral Drug Des.* **2004**, *4*, 1–62.
- De Clercq, E. New Anti-HIV Agents and Targets. *Med. Res. Rev.* **2002**, *22*, 531–565.
- Kazmierski, W. M.; Peckman, J. P.; Duan, M.; Kenakin, T. P.; Jenkinson, S.; Gudmundsson, K. S.; Piscitelli, S. C.; Feldman, P. L. Recent progress in the discovery of new CCR5 and CXCR4 chemokine receptor antagonists as inhibitors of HIV-1 entry. Part 2. *Curr. Med. Chem.* **2005**, *4*, 133–152.
- Maeda, K.; Nakata, H.; Ogata, H.; Koh, Y.; Miyakawa, T.; Mitsuya, H. The current status of, and challenges in, the development of CCR5 inhibitors as therapeutics for HIV-1 infection. *Curr. Opin. Pharmacol.* **2004**, *4*, 447–452.
- Palani, A.; Tagat, J. R. Discovery and development of small-molecule chemokine coreceptor CCR5 antagonists. *J. Med. Chem.* **2006**, *49*, 2851–2857.
- Gerlach, L. O.; Skerlj, R. T.; Bridger, G. J.; Schwartz, T. W. Molecular Interaction of Cyclam and Bicyclam Non-peptide Antagonists with the CXCR4 Chemokine Receptor. *J. Biol. Chem.* **2001**, *276*, 14154–14160.
- Seibert, C.; Ying, W.; Gavrilov, S.; Tsamis, F.; Kuhmann, S. E.; Palani, A.; Tagat, J. R.; Clader, J. W.; McCombie, S. W.; Baroudy, B. M.; Smith, S. O.; Dragice, T.; Moore, J. P.; Sakmar, T. P. Interaction of small molecule inhibitors of HIV-1 entry with CCR5. *Virology* **2006**, *349*, 41–54.
- Pérez-Nueno, V. I.; Ritchie, D. W.; Rabal, O.; Pascual, R.; Borrell, J. I.; Teixidó, J. Comparison of ligand-based and receptor-based virtual screening of HIV entry inhibitors for the CXCR4 and CCR5 receptors using 3D ligand shape-matching and ligand-receptor docking. *J. Chem. Inf. Model.* **2008**, *48*, 509–533.
- Maeda, K.; Das, D.; Ogata-Aoki, H.; Nakata, H.; Miyakawa, T.; Tojo, Y. Structural and molecular interactions of CCR5 inhibitors with CCR5. *J. Biol. Chem.* **2006**, *281*, 12688–12698.
- Kellenberger, E.; Springael, J.-Y.; Parmentier, M.; Hachet-Haas, M.; Galzi, J.-L.; Rognan, D. Identification of nonpeptide CCR5 receptor agonists by structure-based virtual screening. *J. Med. Chem.* **2007**, *50*, 1294–1303.
- Wang, T.; Duan, Y. Binding modes of CCR5-targeting HIV entry inhibitors: Partial and full antagonists. *J. Mol. Graphics Modell.* **2008**, *26*, 1287–1295.
- Castonguay, L. A.; Weng, Y.; Adolfsen, W.; Di Salvo, J.; Kilburn, R.; Caldwell, C. G.; Daugherty, B. L.; Finke, P. E.; Hale, J. J.; Lynch, C. L.; Mills, S. G.; MacCoss, M.; Springer, M. S.; DeMartino, J. A. Binding of 2-aryl-4-(piperidin-1-yl)butanamines and 1,3,4-trisubstituted pyrrolidines to human CCR5: A molecular modeling guided mutagenesis study of the binding pocket. *Biochemistry* **2003**, *42*, 1544–1550.
- Tsamis, F.; Gavrilov, S.; Kajumo, F.; Seibert, C.; Kuhmann, S.; Ketas, T.; Trkola, A.; Palani, A.; Clader, J. W.; Tagat, J. R.; McCombie, S.; Baroudy, B.; Moore, J. P.; Sakmar, T. P.; Dragice, T. Analysis of the mechanism by which the small-molecule CCR5 antagonists SCH-351125 and SCH-350581 inhibit human immunodeficiency virus type 1 entry. *J. Virol.* **2003**, *77*, 5201–5208.
- Dragice, T.; Trkola, A.; Thompson, D. A.; Cormier, E. G.; Kajumo, F. A.; Maxwell, E.; Lin, S. W.; Ying, W.; Smith, S. O.; Sakmar, T. P.; Moore, J. P. A binding pocket for a small molecule inhibitor of HIV-1 entry within the transmembrane helices of CCR5. *Proc. Natl. Acad. Sci. U.S.A.* **2000**, *97*, 5639–5644.
- Nishikawa, M.; Takashima, K.; Nishi, T.; Furuta, R. A.; Kanzaki, N.; Yamamoto, Y.; Fujisawa, J.-I. Analysis of binding sites for the new small-molecule CCR5 antagonist TAK-220 on human CCR5. *Antimicrob. Agents Chemother.* **2005**, *49*, 4708–4715.
- Paterlini, M. G. Structure modeling of the chemokine receptor CCR5: Implications for ligand binding and selectivity. *Biophys. J.* **2002**, *83*, 3012–3031.
- Fano, A.; Ritchie, D. W.; Carrieri, A. Modelling the structural basis of human CCR5 chemokine receptor function: from homology model-building and molecular dynamics validation to agonist and antagonist docking. *J. Chem. Inf. Model.* **2006**, *46*, 1223–1235.
- Ritchie, D. W.; Kemp, G. J. L. Fast computation, rotation, and comparison of low resolution spherical harmonic molecular surfaces. *J. Comput. Chem.* **1999**, *20*, 383–395.
- ParaSurf; version08; CEPOS InSilico Ltd.: Erlangen, Germany, 2008; <http://www.ceposinsilico.de/Pages/Products.html> (accessed May 26, 2008).
- Triballeau, N.; Acher, F.; Brabet, I.; Pin, J.-P.; Bertrand, H.-O. Virtual screening workflow development guided by the “receiver operating characteristic” curve approach. application to high-throughput docking on metabotropic glutamate receptor subtype 4. *J. Med. Chem.* **2005**, *48*, 2534–2547.
- Ritchie, D. W.; Kemp, G. J. L. Protein docking using spherical polar Fourier correlations. *Proteins: Struct., Funct., Genet.* **2000**, *39*, 178–194.
- Lin, J.; Clark, T. An analytical, variable resolution, complete description of static molecules and their intermolecular binding properties. *J. Chem. Inf. Model.* **2005**, *45*, 1010–1016.
- Mavridis, L.; Hudson, B. D.; Ritchie, D. W. Toward high throughput 3D virtual screening using spherical harmonic surface representations. *J. Chem. Inf. Model.* **2007**, *47*, 1787–1796.
- Frank, J. Three-dimensional electron microscopy of macromolecular assemblies. Oxford University Press: Oxford, U.K., 2006.
- Fawcett, T. An introduction to ROC analysis. *Pattern Recognit. Lett.* **2006**, *27*, 861–874.
- Bridger, G.; Skerlj, R.; Kaller, A.; Harwing, C.; Bogucki, D.; Wilson, T. R.; Crawford, J.; McEachern, E. J.; Atsma, B.; Nan, S.; Zhou, Y. World Patent WO 0022600, 2002.
- Bridger, G.; Skerlj, R.; Kaller, A.; Harwing, C.; Bogucki, D.; Wilson, T. R.; Crawford, J.; McEachern, E. J.; Atsma, B.; Nan, S.; Zhou, Y. World Patent WO 0022599, 2002.
- Bridger, G.; Skerlj, R.; Kaller, A.; Harwing, C.; Bogucki, D.; Wilson, T. R.; Crawford, J.; McEachern, E. J.; Atsma, B.; Nan, S.; Zhou, Y. World Patent WO 00234745, 2002.
- Bridger, G.; Skerlj, R.; Kaller, A.; Harwing, C.; Bogucki, D.; Wilson, T. R.; Crawford, J.; McEachern, E. J.; Atsma, B.; Nan, S.; Zhou, Y. World Patent WO 055876, 2003.
- Bridger, G.; Skerlj, R.; Kaller, A.; Harwing, C.; Bogucki, D.; Wilson, T. R.; Crawford, J.; McEachern, E. J.; Atsma, B.; Nan, S.; Zhou, Y.; Smith, C. D.; Di Fluor, R. M. U.S. Patent 0019058, 2004.
- Ichiyama, K.; Yokohama-Kumakura, S.; Tanaka, Y.; Tanaka, R.; Hirose, K.; Bannai, K.; Edamatsu, T.; Yanaka, M.; Niitani, Y.; Miyako-Kurosaki, N.; Takaku, H.; Koyanagi, Y.; Yamamoto, N. A duodenally absorbable CXC chemokine receptor 4 antagonist, KRH-1636, exhibits a potent and selective anti-HIV-1 activity. *Proc. Natl. Acad. Sci. U.S.A.* **2003**, *100*, 4185–4190.
- Murakami, T.; Yoshida, A.; Tanaka, R.; Mitsuhashi, S.; Hirose, K.; Yanaka, M.; Yamamoto, N.; Tanaka, Y. KRH-2731: An Orally Bioavailable CXCR4 Antagonist Is a Potent Inhibitor of HIV-1 Infection. In *2004 Antivirals Pipeline Report*; Proceedings of the 11th Conference on Retroviruses and Opportunistic Infection, San Francisco, CA, Feb 8–11, 2004; Camp, R., Ed.; Treatment Action Group: San Francisco CA, 2004; Abstract 541.
- Yamazaki, T.; Saitou, A.; Ono, M.; Yokohama, S.; Bannai, K.; Hirose, K.; Yanaka, M. World Patent WO 029218, 2003.
- Yamazaki, T.; Kikumoto, S.; Ono, M.; Saitou, A.; Takahashi, H.; Kumakura, S.; Hirose, K. World Patent WO 024697, 2004.
- Bridger, G. J.; Skerlj, R. T.; Padmanabhan, S.; Martellucci, S. A.; Henson, G. W.; Struyf, S.; Witvrouw, M.; Schols, D.; De Clercq, E. Synthesis and structure–activity relationships of phenylenebis(methylene)-linked bis-azamacrocycles that inhibit HIV-1 and HIV-2 replication by antagonism of the chemokine receptor CXCR4. *J. Med. Chem.* **1999**, *42*, 3971–3981.
- De Clercq, E. Inhibition of HIV infection by bicyclams, highly potent and specific CXCR4 antagonists. *Mol. Pharmacol.* **2000**, *57*, 833–839.
- Esté, J. A.; Cabrera, C.; De Clercq, E.; Struyf, S.; Van Damme, J.; Bridger, G.; Skerlj, R. T.; Abrams, M. J.; Henson, G.; Gutierrez, A.; Clotet, B.; Schols, D. Activity of different bicyclam derivatives against human immunodeficiency virus depends on their interaction with the CXCR4 chemokine receptor. *Mol. Pharmacol.* **1999**, *55*, 67–73.
- Egberink, H. F.; De Clercq, E.; Van Vliet, A. L. W.; Balzarini, J.; Bridger, G. J.; Henson, G.; Horzinek, M. C.; Schols, D. Bicyclams, selective antagonists of the human chemokine receptor CXCR4, potently inhibit feline immunodeficiency virus replication. *J. Virol.* **1999**, *73*, 6346–6352.
- Hatse, S.; Princen, K.; De Clercq, E.; Rosenkilde, M. M.; Schwartz, T. W.; Hernandez-Abad, P. E.; Skerlj, R. T.; Bridger, G. J.; Schols, D. AMD3465, a monomacrocyclic CXCR4 antagonist and potent HIV entry inhibitor. *Biochem. Pharmacol.* **2005**, *70*, 752–761.

- (44) Princen, K.; Hatse, S.; Vermeire, K.; Aquaro, S.; De Clercq, E.; Gerlach, L.-O.; Rosenkilde, M.; Schwartz, T. W.; Skerlj, R.; Bridger, G.; Schols, D. Inhibition of human immunodeficiency virus replication by a dual CCR5/CXCR4 antagonist. *J. Virol.* **2004**, *78*, 12996–13006.
- (45) Tamamura, H.; Araki, T.; Ueda, S.; Wang, Z.; Oishi, S.; Esaka, A.; Trent, J. O.; Nakashima, H.; Yamamoto, N.; Peiper, S. C.; Otaka, A.; Fujii, N. Identification of novel low molecular weight CXCR4 antagonists by structural tuning of cyclic tetrapeptide scaffolds. *J. Med. Chem.* **2005**, *48*, 3280–3289.
- (46) Rosenkilde, M. M.; Gerlach, L. O.; Hatse, S.; Skerlj, R. L.; Schols, D.; Bridger, G.; Schwartz, T. W. Molecular mechanism of action of monociclam versus biciclam non-peptide antagonist in the CXCR4 chemokine receptor. *J. Biol. Chem.* **2007**, *282*, 27354–27365.
- (47) Palani, A.; Shapiro, S.; Clades, J. W.; Greenlee, W. J.; Blythin, D.; Cox, K.; Wagner, N. E.; Strizki, J.; Baroudy, B. M.; Dan, N. Biological evaluation and interconversion studies of rotamers of SCH 351125, an orally bioavailable CCR5 antagonist. *Bioorg. Med. Chem. Lett.* **2003**, *13*, 705–708.
- (48) Billick, E.; Seibert, C.; Pugach, P.; Ketas, T.; Trkola, A.; Endres, M. J.; Murgolo, N. J.; Coates, E.; Reyes, G. R.; Baroudy, B. M.; Sakmar, T. P.; Moore, J. P.; Kuhmann, S. E. The differential sensitivity of human and rhesus macaque CCR5 to small-molecule inhibitors of human immunodeficiency virus type 1 entry is explained by a single amino acid difference and suggests a mechanism of action for these inhibitors. *J. Virol.* **2004**, *78*, 4134–4144.
- (49) Maeda, K.; Yoshimura, K.; Shibayama, S.; Habashita, H.; Tada, H.; Sagawa, K.; Mikayawa, T.; Auki, M.; Fukushima, D.; Mitsuya, H. Novel low molecular weight spirodiketopiperazine derivatives potently inhibit R5 HIV-1 infection through their antagonistic effects on CCR5. *J. Biol. Chem.* **2001**, *276*, 35194–35200.
- (50) Shibayama, S.; Sagawa, K.; Watanabe, N.; Takeda, K.; Tada, H.; Fukushima, D. World Patent WO 2004054616, 2004.
- (51) Takaoka, Y.; Okamoto, M.; Genba, Y. World Patent WO 2004026874, 2004.
- (52) Takaoka, Y.; Nishizawa, R.; Shibayama, S.; Sagawa, K.; Matsuo, M. Y. World Patent WO 2002074770, 2002.
- (53) Imawaka, H.; Shibayama, S.; Takaoka, Y. World Patent WO 2003035074, 2003.
- (54) Cumming, J.; Tucker, H. World Patent WO 2003042177, 2003.
- (55) Cumming, J. World Patent WO 2003042178, 2003.
- (56) Cumming, J. World Patent WO 2003080574, 2003.
- (57) Cumming, J.; Winter, J. World Patent WO 2004018425, 2004.
- (58) Burrows, J.; Cumming, J. World Patent WO 2002076, 2002.
- (59) Willoughby, C. W.; Rosauer, K. G.; Hale, J. J.; Budhu, R. J.; Mills, S. G.; Chapman, K. T.; MacCoss, M.; Malkowitz, L.; Springer, M. S.; Gould, S. L.; DeMartino, J. A.; Siciliano, S. J.; Cascieri, M. A.; Carella, A.; Catver, G.; Colmes, K.; Schlieff, W. A.; Danzeisen, R.; Hazuda, D.; Kessler, J.; Lineberger, J.; Miller, M.; Emini, E. A. 1,3,4 Trisubstituted pyrrolidine CCR5 receptor antagonists bearing 4-aminoheterocycle substituted piperidine side chains. *Bioorg. Med. Chem. Lett.* **2003**, *13*, 427–431.
- (60) Kazmierski, W. M.; Aquino, C. J.; Bifulco, N.; Boros, E. E.; Chauder, B. A.; Chong, P. Y.; Duan, M.; Deanada, F. Jr.; Koble, C. S.; Malean, E. W.; Peckham, J. P.; Perkins, A. C.; Thompson, J. B.; Vanderwall, D. World Patent WO 2004054974, 2004.
- (61) Duan, M.; Kazmierski, W. M.; Aquino, C. J. World Patent WO 200405481, 2004.
- (62) Peckham, J. P.; Aquino, C. J.; Kazmierski, W. M. World Patent WO2004055010, 2004.
- (63) Aquino, C. J.; Chong, P. Y.; Duan, M.; Kazmierski, W. M. World Patent WO 2004055011, 2004.
- (64) Youngman, M.; Kazmierski, W. M.; Yang, H.; Aquino, C. J. World Patent WO 2004055012, 2004.
- (65) Yang, H.; Kazmierski, W. M.; Aquino, C. J. World Patent WO 2004055016, 2004.
- (66) Aramaki, Y.; Seto, M.; Okawa, T.; Oda, T.; Kanzaki, N.; Shiraishi, M. Synthesis of 1-benzothiepine and 1-benzazepine derivatives as orally active CCR5 antagonists. *Chem. Pharm. Bull.* **2004**, *52*, 254–258.
- (67) Seto, M.; Aramaki, Y.; Okawa, T.; Miyamoto, N.; Aikawa, K.; Kanzaki, N.; Shiraishi, M. Orally active antagonists as anti-HIV-1 agents: Synthesis and biological activity of 1-benzothiepine 1,1-dioxide and 1-benzazepine derivatives containing a tertiary amine moiety. *Chem. Pharm. Bull.* **2004**, *52*, 577–590.
- (68) Perros, M.; Price, D. A.; Stammen, B. L. C.; Wood, A. World Patent WO 2003084954, 2003.
- (69) Basford, P. A.; Stephenson, P. T.; Taylor, S. C. J.; Wood, A. World Patent WO 2003084954, 2003.
- (70) Armour, D. R.; Price, D. A.; Stammen, B. L. C.; Wood, A.; Perros, M.; Edwards, M. P. World Patent WO 2000038680, 2000.
- (71) Rusconi, S.; Scozzafava, A.; Mastrolorenzo, A.; Supuran, T. C. New advances in HIV entry inhibitors development. *Curr. Drug Targets Infect. Disord.* **2004**, *4*, 339–355.
- (72) Imamura, S.; Kurasawa, O.; Nara, Y.; Ichikawa, T.; Nishikawa, Y.; Iida, T.; Hashiguchi, S.; Kanzaki, N.; Lizawa, Y.; Baba, M.; Sugihara, Y. CCR5 antagonists as anti-HIV-1 agents. Part 2: Synthesis and biological evaluation of *N*-[3-(4-benzylpiperidin-1-yl)propyl]-*N,N'*-diphenylureas. *Bioorg. Med. Chem.* **2004**, *12*, 2295–2306.
- (73) Imamura, S.; Ishihara, Y.; Hattori, T.; Kurasawa, O.; Matsushita, Y.; Sugihara, Y.; Kanzaki, N.; Lizawa, Y.; Baba, M.; Hashiguchi, S. CCR5 antagonists as anti-HIV-1 agents. 1. Synthesis and biological evaluation of 5-oxopyrrolidine-3-carboxamide derivatives. *Chem. Pharm. Bull.* **2004**, *52*, 63–73.
- (74) Wei, R. G.; Arnaiz, D. O.; Chou, Y.-L.; Davey, D.; Dunning, L.; Lee, W.; Lu, S.-F.; Onuffer, J.; Ye, B.; Phillips, G. *Bioorg. Med. Chem. Lett.* **2007**, *17*, 231–234.
- (75) Lu, S.-F.; Chen, B.; Davey, D.; Dunning, L.; Jaroch, S.; May, K.; Onuffer, J.; Phillips, G.; Subramanyam, B.; Tseng, J.-L.; Wei, R. G.; Wei, M.; Ye, B. *Bioorg. Med. Chem. Lett.* **2007**, *17*, 1883–1887.
- (76) Debnath, A. K. Generation of predictive pharmacophore models for CCR5 antagonists: Study with piperidine- and piperazine-based compounds as a new class of HIV-1 entry inhibitors. *J. Med. Chem.* **2003**, *46*, 4501–4515.
- (77) Ward, J. H. Hierarchical grouping to optimize an objective function. *J. Am. Statist. Assoc.* **1963**, *58*, 236–244.
- (78) *JKlustor*, version 5.0.4; ChemAxon Ltd.: Budapest, Hungary, 2008; <http://www.chemaxon.com/jchem/doc/user/JKlustor.html> (accessed May 26, 2008).
- (79) Kelley, L. A.; Gardner, S. P.; Sutcliffe, M. J. An automated approach for clustering an ensemble of NMR-derived protein structures into conformationally-related subfamilies. *Protein Eng.* **1996**, *9*, 1063–1065.
- (80) Kleiweg, P. Data Clustering Software. <http://www.let.rug.nl/~kleiweg/index.html> (accessed May 26, 2008).
- (81) Grant, A. J.; Pickup, B. T. A fast method of molecular shape comparison: a simple application of a Gaussian description of molecular shape. *J. Comput. Chem.* **1996**, *17*, 1653–1659.
- (82) Morris, G. M.; Goodsell, D. S.; Halliday, R. S.; Hart, W.; Belew, R. K.; Olson, A. J. Automated docking using a Lamarckian genetic algorithm and empirical binding free energy function. *J. Comput. Chem.* **1998**, *19*, 1639–1662.
- (83) Verdonk, M. L.; Cole, J. C.; Hartshorn, M. J.; Murray, C. W.; Taylor, R. D. Improved protein–ligand docking using GOLD. *Proteins: Struct., Funct., Genet.* **2003**, *52*, 609–623.
- (84) McGann, M. R.; Almond, H. R.; Nicholls, A.; Grant, J. A.; Brown, F. K. Gaussian docking functions. *Biopolymers* **2003**, *68*, 76–90.
- (85) Kondru, R.; Zhang, J.; Ji, C.; Mirzadegan, T.; Rotstein, D.; Sankuratri, S.; Dioszegi, M. Molecular interactions of CCR5 with major classes of small-molecule anti-HIV CCR5 antagonists. *Mol. Pharmacol.* **2008**, *73*, 789–800.

CI800257X