ARTICLES

Unsupervised 3D Ring Template Searching as an Ideas Generator for Scaffold Hopping: Use of the LAMDA, RigFit, and Field-Based Similarity Search (FBSS) Methods

Martin Bohl* and Björn Loeprecht

Tripos GmbH, Martin-Kollar-Strasse 17, D-81829 Munich, Germany

Bernd Wendt

Tripos Discovery Research Ltd., Tripos Discovery Research Centre, Bude-Stratton Business Park, Bude, Cornwall EX23 8LY, United Kingdom

Trevor Heritage†

Tripos, Inc., 1699 South Hanley Road, St. Louis, Missouri 63144

Nicola J. Richmond‡ and Peter Willett

Krebs Institute for Biomolecular Research and Department of Information Studies, University of Sheffield, 211 Portobello Street, Sheffield S1 4DP, United Kingdom

Received November 11, 2004

Crystal structures taken from the Cambridge Structural Database were used to build a ring scaffold database containing 19 050 3D structures, with each such scaffold then being used to generate a centroid connecting path (CCP) representation. The CCP is a novel object that connects ring centroids, ring linker atoms, and other important points on the connection path between ring centroids. Unsupervised searching in the scaffold and CCP data sets was carried out using the atom-based LAMDA and RigFit search methods and the field-based similarity search method. The performance of these methods was tested with three different ring scaffold queries. These searches demonstrated that unsupervised 3D scaffold searching methods can find not only the types of ring systems that might be retrieved in carefully defined pharmacophore searches (supervised approach) but also additional, structurally diverse ring systems that could form the starting point for lead discovery programs or other scaffold-hopping applications. Not only are the methods effective but some are sufficiently rapid to permit scaffold searching in large chemical databases on a routine basis.

INTRODUCTION

Very subtle changes in the structures of bioactive compounds can sometimes lead to substantial differences in their biological activity (i.e., structural similarity is coupled with functional diversity). For example, moving from the aromatic ring A of an estrogen to a 4-ene structure can change an estrogenic steroid to a progestin, or moving from testosterone to 19-nortestosterone changes an androgenic to an anabolic steroid. It is also very common to find that inhibitors from different chemical classes show similar potency and the same mechanism of action. These observations lead to the concept of a pharmacophore, that is, those parts of the molecule that are essential for the desired receptor recognition, binding, and subsequent activity, whereas other parts may exhibit very considerable structural variation. The latter, nonpharmacophore moieties may contribute to activity by placing side chains and pharmacophoric groups in an appropriate position for receptor binding, or by being involved in nonspecific binding to protein surfaces. If such portions of the molecule can be modified without substantially changing the activity profile (i.e., structural diversity is coupled with functional similarity), then one could obtain a potentially valuable range of lead structures for drug discovery.

Existing computational approaches that can be used to generate or identify heterogeneous structures that may display similar potency include de novo design, virtual screening, flexible docking, pharmacophore searching, bioisosteric replacements, and topomer shape similarity applications inter alia (see, e.g., refs 1−4). The generation or selection of structural alternatives is most easily effected when the 3D structures of receptors or ligand−receptor complexes are available. If this is not the case, then pharmacophoric features or similarity to known active ligands can be taken into account when searching, with constraints being based on ligand features such as shape, size, topology, hydrophobicity, charge distribution, hydrogen-bonding features, 2D finger-

* Corresponding author e-mail: mbohl@tripos.com.

† Present address: Elsevier MDL, 14600 Catalina Street, San Leandro, CA 94577.

‡ Present address: GlaxoSmithKline, Cheminformatics, Gunnels Wood Road, Stevenage SG1 2NY, U. K.

10.1021/ci049657k CCC: $33.50
Published on Web 07/26/2006

prints, 3D structure descriptors, and so forth. In this paper, we report a similarity searching procedure that is based on a core ring system, which functions as a central template that can hold and position side chains, functional groups, and pharmacophoric features. Our procedure is in marked contrast to most similarity-searching procedures, which measure chemical similarity on the basis of the entire molecule, but has the great advantage that it may permit the identification of novel ring systems that are able to hold substituents in similar locations. There is much current interest in such scaffold-hopping methods.[5−21]

A common approach to scaffold hopping is based on 3D pharmacophore searching (using UNITY[22] or analogous substructure searching software), which selectively allows structural variation in some parts of a molecule while avoiding variation in the others. When defining the 3D query in such a supervised environment, the user is able to distinguish between the mandatory and nonmandatory parts, but this means that query formulation is necessarily subjective, and thus potentially problematic in the absence of detailed knowledge of the pharmacophore. Here, we consider the following methods for aligning ring systems and, hence, for carrying out scaffold hopping: a traditional rigid-body alignment method, the RigFit procedure that is implemented in the FlexS[23−26] software system; a novel method based on an image recognition algorithm, the LAMDA procedure described recently by Richmond et al.;[27] and the field-based similarity search (FBSS) program that is based on steric fields [28] and that was used in our previous study of scaffold hopping.[8] In addition to evaluating the effectiveness of these three methods, we also compare the results obtained with those obtained from static UNITY 3D searches for ring systems. Our approach to scaffold hopping has the following advantages: it is unsupervised, in that it does not require the user to define a specific substructural query; it is able to rank hits according to a similarity score; and it provides an alignment of the top-ranked structures for visual inspection of the results. In fact, we have applied the three alignment methods not just to conventional ring scaffold structures (as represented by all of their constituent atoms) but also to a reduced representation consisting of ring centroids and the centroid connecting path (CCP) between them. With the exception of the RigFit searches, only steric complementarity is taken into account here; a scaffold-hopping application might additionally take account of the matching of electronic features, hydrogen-bonding capabilities, hydrophobic regions, and appropriate side-chain directionality and so forth.

The successful replacement of a central ring system by a different cyclic system that holds substituents in appropriate locations permits the controlled introduction of region-focused structural variation in a molecule. This may, for example, prove important in the identification of new lead candidates, in generating valuable alternatives when modifying the core structure in combinatorial library design, as part of a de novo design strategy, within a structure-based ligand design application, in patent circumventions, in optimizing the absorption, distribution, metabolism, and excretion (ADME)/toxicity profile of a drug candidate, or for a "drug rescue" project. That said, it must be emphasized that the search procedures described below (and previously by other workers) can only generate structural ideas for scaffold hopping, with final inspection and selection requiring the skills of a creative chemist.

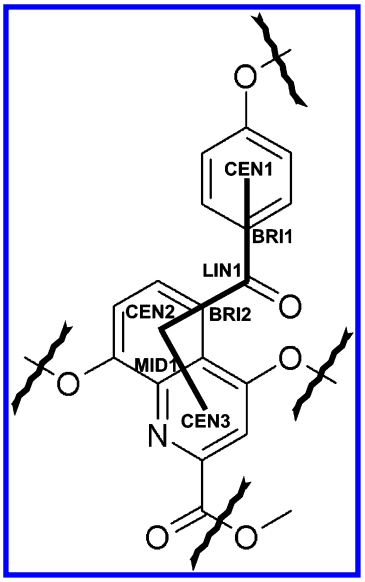

Figure 1. Conversion of an example CSD structure (YADDUC) into a ring template by removing side chains along the wavy lines and generation of the CCP by connecting the CEN (ring centroid), BRI (bridge), LIN (linker), and MID (ring-bond midpoint) points.

EXPERIMENTAL DETAILS

Construction of the Database of Ring Templates. A total of 89 686 high-resolution crystal structures ($R < 0.05$) were extracted from the Cambridge Structural Database (CSD) version 5.2.1 and ConQuest version 1.2,[29−31] subject to the structures not being disordered, not being polymeric, and not containing any obvious errors. These crystal structures were then processed as described below to yield the final database that was used in the searching experiments.

The generation of the ring scaffolds from those sets of 3D crystal coordinates that contained at least one ring involved the use of the ring, side chain, and linker concepts described by Bemis and Murcko[32] and by Xu.[10] A linker connects two or more rings in a given structure on a direct path, whereas a side chain is connected to just one ring system and contains neither ring nor linker atoms. In our work, all connected ring atoms and all conjugated systems directly attached to rings (in linkers or side chains) were kept together as a single ring fragment. A fully conjugated linker and the interconnected rings constituted a single ring template or scaffold (for example, biphenyl, phenyl−C═C−phenyl, and phenyl−C(═O)−phenyl but not phenyl−C−phenyl), with fused rings and spiro systems providing other examples of connected rings. In the cases of a nonconjugated side chain or a nonconjugated linker, only the atom directly attached to the ring system was retained (we refer to this as the *attachment atom*). The same was done if the side chain or linker was partially conjugated but the attachment atom was not part of the conjugation. Acyclic substituents attached to conjugated systems in linkers and in side chains were stripped. Hydrogen atoms directly attached to ring atoms were considered as part of the ring system, whereas no hydrogens were included when they were attached to linker or attachment atoms. The generation of a typical ring template is illustrated in Figure 1.

After removing all of the nonring atoms (with the exception of the attachment atoms, conjugated side chains,

and conjugated linkers), the next step filtered out ring templates containing atoms other than C, H, N, P, O, S, F, Cl, Br, and I and rings containing more than eight atoms. Finally, one randomly selected conformation was extracted for each connected 2D structure by applying a Tripos internal subroutine, this yielding a database containing 24 244 unique ring scaffolds.

In addition to comparing the 3D structures of ring templates themselves, we have developed a novel representation of ring structures for this study: the centroid connecting path. For each ring contained in a 3D ring template, the 3D coordinates of the geometric center of all atoms in the ring were determined: this is the *centroid*, or CEN. Individual rings were extracted from scaffolds by SYBYL line notation[22] atom mapping. Rings were subsequently detected while increasing the ring size up to eight-membered rings. Atoms in previously detected rings were kept in a matching atom list. Additional rings were only found if at least one atom of the investigated ring was not contained in the current matching atom list. Information about common atoms of directly connected rings was kept. The centroids of a given ring template were then connected by the shortest path between them. This path may go through *ring-bond midpoints* (MID) for fused rings or may involve ring atoms (*bridge atoms*, BRI), for example, in spiro compounds, when rings are directly connected through a bond (i.e., in biphenyl), or when rings are connected through *linker atoms* (LIN). In the last named case, the linker atoms that are needed to establish a connecting path between the centroids were also included in the shortest path; for example, the CCP of phenyl1−C(═O)−phenyl2 is of the form CEN1−BRI1−LIN−BRI2−CEN2. An example of a CCP containing CEN, BRI, MID, and LIN points is shown in Figure 1. Once CEN, BRI, MID, and LIN points had been defined, the nonexisting connections between them were established using the following distance thresholds: 1.62 Å for CEN−MID and 2.0 Å for CEN−BRI. In this study, we used an atom type of nitrogen for CEN and carbon for other CCP point types. Note that a branched CCP may result if three or more rings or linkers are attached or fused to a given central ring.

CCP structures are generally noncoplanar, but planar and linear CCP structures do occur. CCP representations do not contain any hydrogen atoms and are best analyzed in concert with their respective 3D ring scaffold structures, which maintain some information about outgoing bonds in the form of hydrogens directly attached to rings. In addition, one may choose to keep other information about orientations of potential substituents for rings, attachment atoms, and linkers. This, however, will not affect the results of CCP-based searches.

The CCP provides a novel way of representing the 3D structure of a ring system but at a lower level of complexity than the normal, all-atom representation; similar approaches have been used in other studies[33−36] where a 2D molecular fragment consisting of a subset of atoms is represented by a single center to reduce the complexity of a chemical structure. In our work, the generation of CCPs and the fitting of original 3D structures onto the aligned CCPs were performed by two SYBYL programming language[22] scripts: in all, the database of 24 244 distinct CSD ring scaffolds yielded a database of 19 050 unique CCPs after the elimination of trivial CCPs consisting of just a single CEN point. In what follows, we refer to these as the *24K database* and the *19K database*.

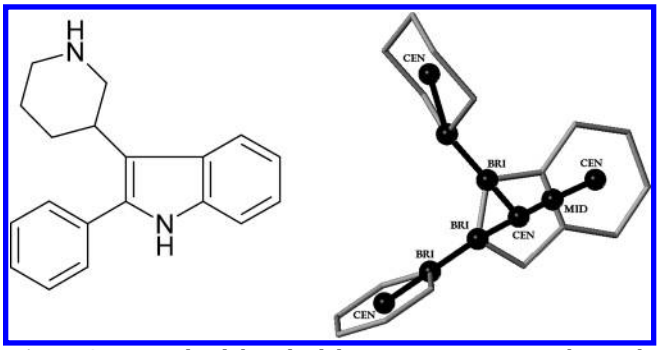

Figure 2. Piperidinylphenylindole query structure together with CCP representation (hydrogens in query not shown).

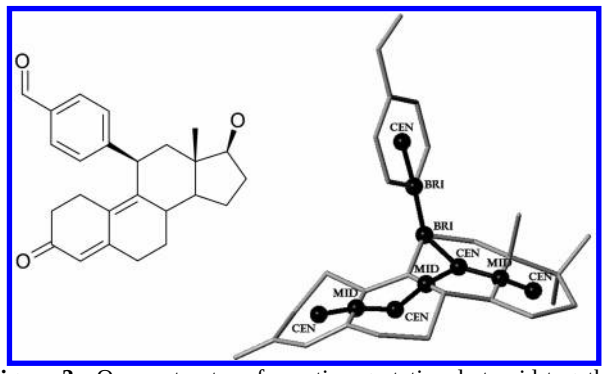

Figure 3. Query structure for antiprogestational steroid together with CCP representation (hydrogens in query not shown).

Inhibitor Query Scaffolds. Three ring-scaffold queries were used to evaluate the search performance of the various methods. These queries were built in the same way that the structures in the 24K and 19K databases had been created, as described above.

h5-HT$_{2A}$ Receptor Antagonists. h5-HT$_{2A}$ receptor antagonists are under investigation for the treatment of schizophrenia. The query structure used here is shown in Figure 2 and was taken from work by Rowley et al.[37,38] The 3D structure of this query was first sketched and then refined by the MMFF94s force field,[39] followed by semiempirical AM1[40] geometry optimization.

Antiprogestational Steroids. Progesterone antagonists can be used to terminate pregnancy by menses induction: further details and structure−activity patterns for progesterone derivatives have been reported by Bohl[41] and by Bursi and Groen.[42] The 3D query structure for the antiprogestational steroid was taken from the CSD crystal structure SESJUV[43] and is shown in Figure 3.

Phosphodiesterase PDE4 Inhibitors. PDE4 inhibitors are being considered for the treatment of chronic obstructive pulmonary disorder, Crohn's disease, multiple sclerosis, rheumatic arthritis, asthma, and other inflammatory diseases. The query structure is shown in Figure 4 and was taken from a pharmacophore model developed by van der Mey et al.;[44−47] the query is consistent with the results of a pharmacophore analysis of PDE4 inhibitors reported by Crespo et al.[48] The 3D structure for this query was processed using MMF94s and AM1 as for the piperidinylphenylindole query.

Methods for Superpositioning and Scaffold Searching. *UNITY Searching*. A rigid 3D search was carried out for each of the three query patterns above using the UNITY software, version 4.2.2, in connection with SYBYL software, version 6.7.2, under Irix, version 6.5.[22] The UNITY queries

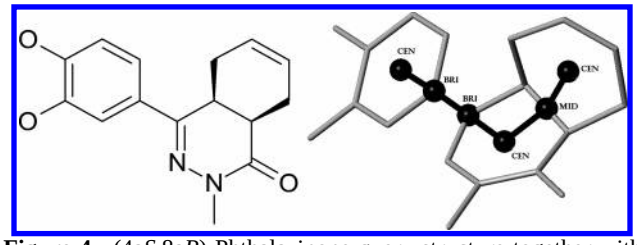

Figure 4. (4a*S*,8a*R*)-Phthalazinone query structure together with CCP representation (hydrogens in query not shown).

for the three target systems contained only the hydrophobic centroids for each ring, with the exception of the PDE4 inhibitor query, where four additional spatial points were used to enforce the (4aS,8aR) configuration. The tolerance for all spatial points was set to 0.75 Å, and Lipinski's "rule of 5" was not applied (Lipinski count of 0). The UNITY 3D search was carried out on the 24K database of ring scaffolds, yielding 34, 12, and 9 hit structures for the h5-HT_{2A} receptor antagonist, antiprogestin, and PDE4 inhibitor queries, respectively. These hits, together with a randomly selected set of nonhit structures, were used to build a file of 200 ring scaffolds, together with the corresponding CCP structures, that was then used for the detailed comparison of the three alignment methods described below.

Linear Assignment Method for Data Set Alignment (LAMDA). LAMDA[27] is a novel algorithm, based on image recognition[49] in two dimensions, for generating atom-based alignments of rigid 3D structures. The algorithm is sufficiently general to align any 3D object represented by a set of point coordinates, and we have adopted the basic procedure to align scaffolds represented by the CCP approach. Each CCP point is described by its type (CEN, BRI, MID, or LIN) and by its 3D coordinates, and then, the cost of matching a pair of CCP points, one from each scaffold, is computed for all possible pairs of points. A set of CCP point equivalences is then determined using an algorithm for solving the linear assignment problem. A postprocessing stage is used to reduce the chance of a poor scaffold alignment arising from mismatched CCP points, and an alignment transformation is then computed to optimally superimpose the CCP points comprising one scaffold onto their equivalent points in the other scaffold. A final refinement of this alignment is then carried out by calculating a new transformation based on a set of CCP point equivalences consisting of pairs that are approximately overlaid, irrespective of their point type. The various steps of the procedure are described in detail by Richmond et al.[27]

Mismatched points can arise when there is symmetry present in the CCP of a scaffold (as can also occur in the case of atom-based alignments). This could be remedied by identifying those points in the scaffold that have a mirror image and filtering them from the set of equivalences prior to calculating the alignment transformation. However, the condensed nature of the CCP representation means that there may be too few points available to calculate the appropriate transformation: at least three noncoplanar points are required. Instead, we adopted a simple, brute-force approach in which combinations of points were exhaustively substituted by their mirror images, the corresponding alignment computed, and then that alignment retained that had the lowest penalty score. When aligning data sets of 3D molecules, it is often undesirable to alter the stereochemistry of each structure. Hence, LAMDA checks for, and subsequently eliminates, any reflection present in the alignment transformation. However, when searching for novel scaffolds, one may want to retrieve the alternative absolute configuration of a molecular structure, and thus, reflections are allowed at the user's discretion.

The differences between CCP point types can be accounted for in the scoring function. For example, matching a CEN with a BRI point could carry a higher cost than matching a CEN with a CEN. Suppose Q is the query scaffold and P a candidate scaffold containing $|Q|$ and $|P|$ CCP points, respectively. Then, the scoring function S, which is a measure of the quality of the alignment, is defined as follows:

$$S = \sum_{i=1}^{|P|} w_i |q_i - T(p_i)| + 1.5(|Q| - |P|) \qquad \text{if } |Q| \geq |P|$$

or

$$S = \sum_{i=1}^{|Q|} w_i |q_i - T(p_i)| \qquad \text{otherwise}$$

where T is the alignment transformation from P onto Q, q_i in Q and p_i in P [$i = 1...$ min($|Q|$,$|P|$)] are pairs of CCP points such that $T(p_i)$ is the nearest CCP point to q_i, and w_i is a weight of matching different types of CCP points. In this study, all weights were set to 1.0 as the use of different weighting schemes was found to only have a minor influence on the results.

Field-Based Similarity Searching (FBSS). FBSS calculations[8,28] were carried out using only shape fields for rigid query and database structures, three GA iterations, and 10 000 operations per database structure.

RigFit included in FlexS. FlexS[23−26] version 1.9.2.L was used with the SUPERPOSITION_MODE of 1 (equal to rigid-body fitting or RigFit mode) and the VOLUME_OVERLAP_TH parameter of 0.0, instead of the default value of 0.60, to avoid prefiltering of the results by a selection criterion based on relative volumes. The total FlexS score was used as a measure of the quality of the alignment.

Computing Times. Searching the 200-scaffold structure data set on an SGI Origin 200 (R10 000, 180 MHz, and 128 MB of memory) took approximately 15 s, 210 min, 11 min, and 5 s for UNITY, FBSS, RigFit-FlexS, and LAMDA, respectively. RigFit-FlexS took 10−11 h on one CPU of this SGI workstation to process the 19K CCP data set, whereas searching this database with LAMDA on a Linux workstation took 3−45 min, depending on the complexity of the query and on the computer configuration (PC or workstation); for Linux PC searches, LAMDA-CCP can certainly be expected to process at least 1000 structures a minute, making it suitable for routine use with large chemical databases. A review of different alignment techniques and of their associated runtimes is provided by Lemmen and Lengauer.[50]

RESULTS AND DISCUSSION

Searches of the 200-Structure Database. The three different methods (LAMDA, RigFit, and FBSS) were used to search for similar chemical scaffolds in a database of 200 ring fragments and then to align the identified 3D structures with the input query structure. The full ring template structures were aligned by the RigFit and FBSS methods, and the 3D CCPs were aligned by the LAMDA and RigFit

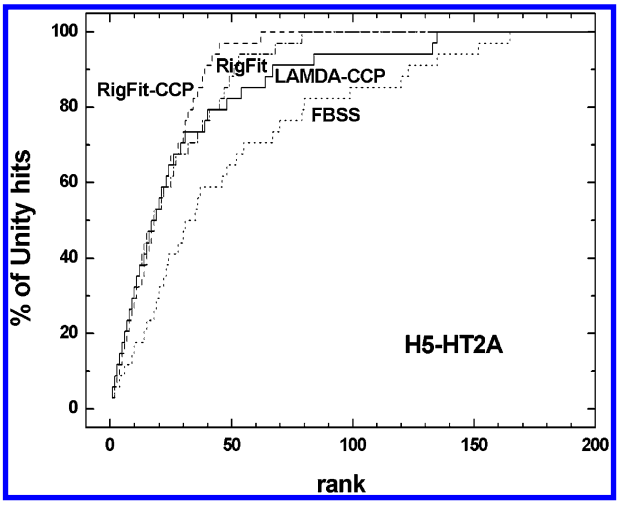

Figure 5. Enrichment of UNITY hits for h5-HT_{2A} receptor antagonists.

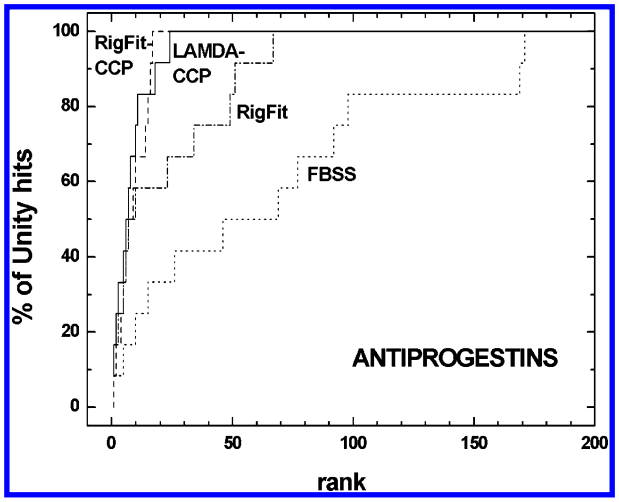

Figure 6. Enrichment of UNITY hits for antiprogestational steroids.

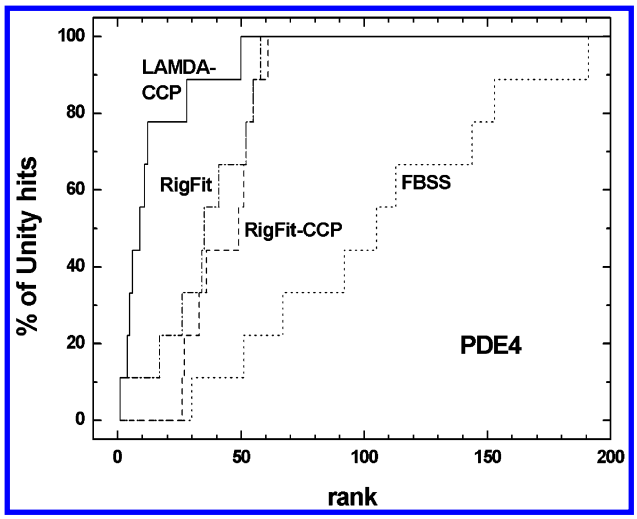

Figure 7. Enrichment of UNITY hits for PDE4 phosphodiesterase inhibitors.

methods. The success of the three unsupervised methods in identifying the hits found by the supervised UNITY search is shown in the cumulative recall plots shown in Figures 5−7. These plot the percentage of the actives that have been retrieved at some rank position against that rank position (so that the best performance is associated with curves that

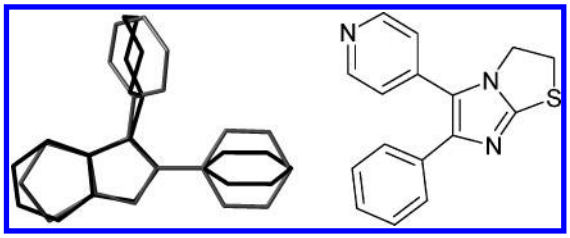

Figure 8. Ring template extracted from CSD structure SOSKUG (gray) fitted onto LAMDA-CCP alignment with h5-HT_{2A} query (black).

are toward the upper left-hand corner of each plot). In these figures, CCP structure representations were used for the solid-line LAMDA-CCP and dashed RigFit-CCP curves, and full scaffold structures were used for the dotted FBSS and dash-point RigFit curves.

Several conclusions can be drawn from the results shown in Figures 5−7. First, all of the methods are able to identify many of the UNITY hits despite the fact that they do not require the specification of a detailed pharmacophore pattern; of these methods, the FBSS approach, which we used in our initial studies of scaffold searching, is consistently the worst of those tested. Second, the absolute level of search performance is quite variable, with the best overall performance being obtained with the complex steroid example. Third, in all cases, the simple CCP representation enables searches to be carried out that are comparable, or superior, to (in terms of UNITY hits) the more detailed, full scaffold representation. Inspection of the alignments shows that the CCP method is better able to avoid suboptimal alignments, especially in those cases where the query scaffold and the database scaffold are of different sizes. This is quite a common occurrence, arising most frequently from those cases where the database scaffold contains more rings than the query template but where a substructure of this database scaffold is able to match the query well. The atom-based FBSS and RigFit methods generally produce noticeably suboptimal alignments in such cases as they are trying to superimpose the entire structures, whereas this appears to be much less of a problem with the simpler, CCP representation. Both of the two types of RigFit search can assign too high an alignment score when fitting very congested or bridged structures that concentrate many atomic spheres in a given volume segment.

The CCP representation does have limitations. Thus, its simplicity leads to some loss of information that can result in the production of reverse alignments (for example, head-to-tail mappings as in the alignment of CEN1−MID1−CEN2−MID2−CEN3 with CEN3−MID2−CEN2−MID1−CEN1); this problem is particularly noticeable if one or both of the scaffolds are highly symmetrical. Another CCP characteristic is that the orientation of ring planes is not taken into account (see, e.g., Figure 8), which can be a limitation if ring-plane positions play an important role in the placement of pharmacophoric groups. On the other hand, the simplicity of the representation, and the accordingly lower number of associated structural constraints, can be of benefit when the ring orientation is flexible to some extent or when the ring orientation cannot be deduced from modeling studies.

The atom-based LAMDA and RigFit methods are considerably more efficient than the field-based FBSS method considered by Bohl et al.,[8] and both of them are found to be

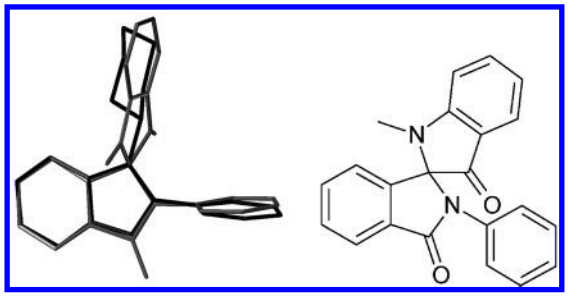

Figure 9. Ring template extracted from CSD structure VANBIV (gray) fitted onto LAMDA-CCP alignment with h5-HT_{2A} query (black).

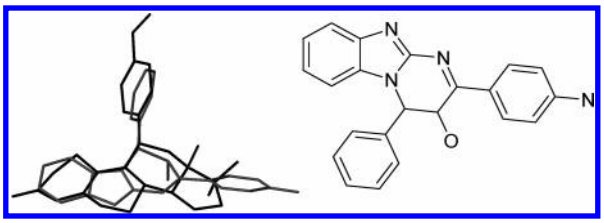

Figure 10. Ring template extracted from CSD structure YORYUZ (gray) fitted onto RigFit-CCP alignment with steroid query (black).

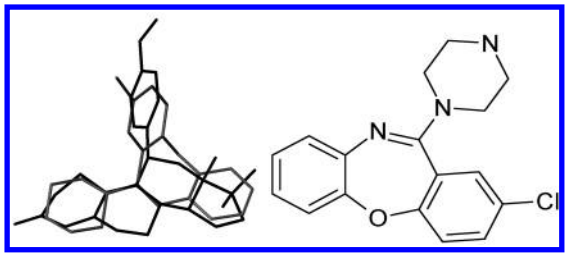

Figure 11. Ring template extracted from CSD structure AMOXAP (gray) aligned to steroid query (black) by RigFit using full ring-system structures.

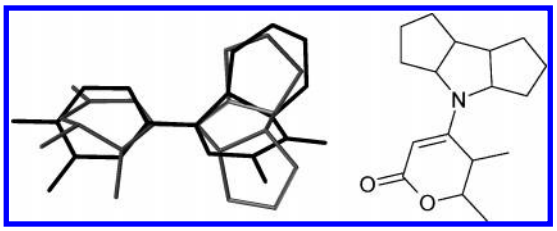

Figure 12. Ring template extracted from CSD structure SEZROE (gray) fitted onto LAMDA-CCP alignment with PDE4 query (black).

generally more effective than FBSS for identifying good ring template alignments at the top of the rankings. The CCP-based LAMDA and RigFit methods are able to find interesting scaffold alignments of approximately the same quality.

Searches of the 19K and 24K Databases. The searches thus far have been rather artificial in that if one has a good 3D pharmacophore model available then one should use that to probe a database, and not the unsupervised approaches considered here. However, any model will normally require substantial refinement (and may also require knowledge taken from the binding site or information from structure−activity relationships) before it is comprehensive, and several scaffolds were identified in RigFit and LAMDA searches of the 24K and 19K databases that had not been retrieved in the original UNITY searches. Examples of such scaffolds are shown in Figures 8−13, in which the hydrogen atoms and CCPs have been excluded for ease of comprehension (and similar comments apply to the other pairs of aligned

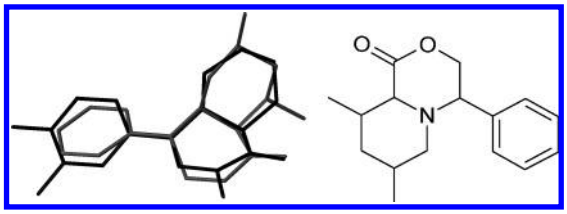

Figure 13. Ring template extracted from CSD structure TUHPAN (gray) fitted onto LAMDA-CCP alignment with PDE4 query (black).

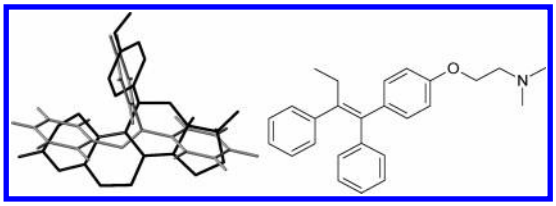

Figure 14. Ring template extracted from WD-1994-10847 (gray) fitted onto LAMDA-CCP alignment with steroid query (black).

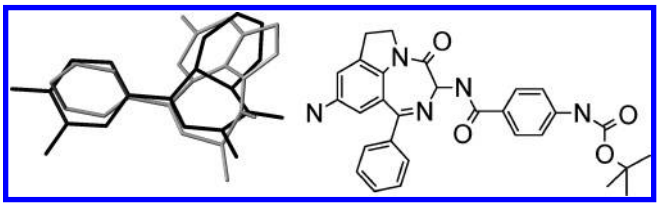

Figure 15. Ring template extracted from WD-1997-13346 (gray) fitted onto LAMDA-CCP alignment with PDE4 query (black).

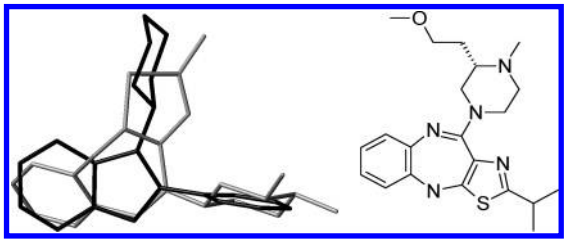

Figure 16. Ring template extracted from WD-2004-3636 (gray) fitted onto LAMDA-CCP alignment with h5-HT_{2A} query (black).

structures that are presented in Figures 14−16 and discussed later in this paper). Thus, the searching tools considered here can retrieve not only many of the ring structures identified by detailed pharmacophore searches but also additional and novel scaffolds, hence demonstrating their suitability for scaffold-hopping applications.

If a pharmacophore model is available, then it would be worth checking ideas generated by 3D scaffold alignments, such as those shown in Figures 8−13, to determine whether the proposed ring system is able to place substituents and pharmacophore points (hydrogen-bond donors and acceptors, charged groups, hydrophobic substituents, etc.) in an appropriate region relative to the core structure. Another helpful step to aid the efficient inspection of a large number of high-ranking hits would be to group them either by means of their CCPs or by means of a clustering procedure based on topological indices or 2D fingerprints. A few representatives from each cluster could then be investigated further if so desired.

New ideas typically do not come from scaffolds that display only minor structural variations when compared with the original query template. Of much more interest are nonredundant hits that have reasonable alignment scores but that are structurally different (for example, in the composition of the CCP). The extent to which this combination of

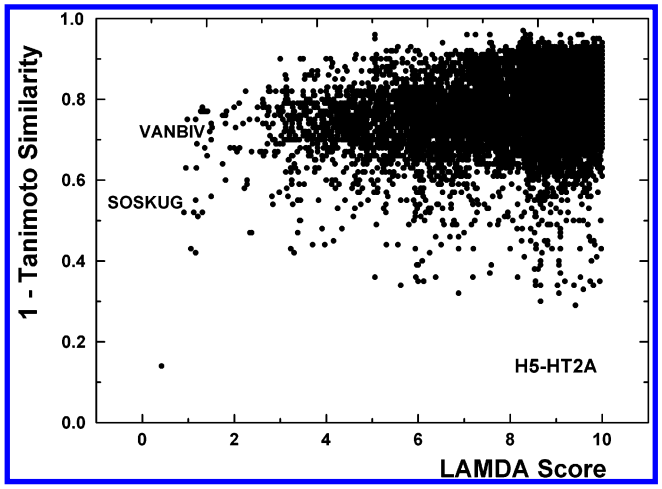

Figure 17. Plot of dissimilarity (1 − Tanimoto similarity) versus the LAMDA alignment score obtained by fitting the 19K CCP structures onto the CCP of the h5-HT_{2A} query; only 10 510 structures with LAMDA scores between 0 and 10 are shown. The superpositions for the SOSKUG and VANBIV structures are shown in Figures 8 and 9, respectively.

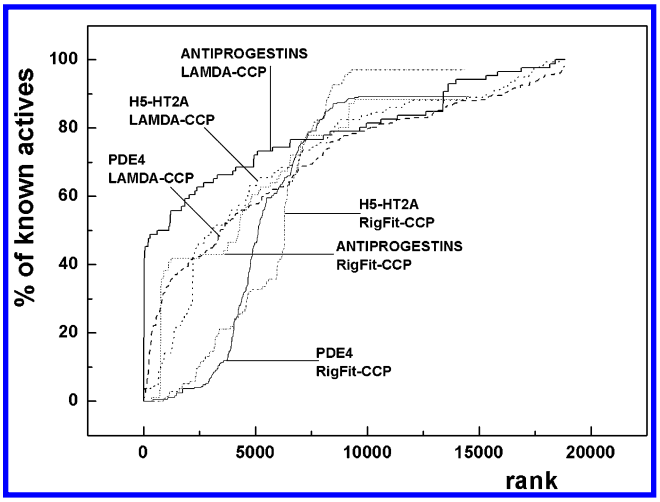

Figure 18. Enrichment of known WDA actives within the 19K CCP database using LAMDA and RigFit.

alignment and diversity occurs in practice is exemplified by Figure 17, which plots the diversity (as determined by the complement of the Tanimoto similarity between the query template and a database scaffold structure using UNITY 2D fingerprints[51−53]) on the *Y* axis against the alignment (as determined by the LAMDA-CCP score) on the *X* axis for searches of the 19K database using the h5-HT_{2A} query. Preferred hits are those that have a high diversity score and a low alignment score (the LAMDA score is a penalty function, and hence, a low score represents a good fit of the two structures that are being compared), that is, those toward the upper left-hand corner of the diversity−alignment plot. In this area of the plot, one will, however, also find scaffolds with similar ring connectivity but having heteroatoms in different positions of the ring system, making them sufficiently dissimilar in the fingerprints. In the opposite (lower right-hand) corner, structures can be found that contain the same individual rings but connected or fused differently together. These structures are highly similar in terms of 2D fingerprints, but their 3D representations will not align well.

A cutoff of 10 was applied to the LAMDA scores for the three searches of the 19K database. This resulted in totals of 4499, 10 510, and 18 495 hits for the steroid (high complexity), h5-HT_{2A} (medium complexity), and PDE4 (low complexity) templates, respectively. The h5-HT_{2A} hits are plotted in Figure 17, where it will be seen that there are a fair number of scaffolds that do appear to exhibit the desired combination of diversity and alignment characteristics; two of these scaffolds are marked in the plot and have been shown previously in Figures 8 and 9. However, it typically requires the pharmacological and structural experience of a scientist with knowledge in the particular therapeutic area to find the appropriate balance between chemical novelty and structural similarity when analyzing the ideas resulting from a scaffold-hopping experiment.

Searches for World Drug Alerts Structures in the 19K Database. The final experiments again involved the 19K database, this time using the LAMDA-CCP and RigFit-CCP methods to search for sets of structures that had been identified in the Derwent *World Drug Alerts* (WDA) database.[54] Specifically, keyword searches were carried out on the WDA for carbon-containing molecules that had the keywords "Progesterone-Antagonist", "Phosphodiesterase-Inhibitor-IV" or "Antiserotonin-2A". These correspond to the three query types discussed previously, and as before, we make the assumption that all of the actives in each case exhibit their potency through the same mode of action, same binding site, and a similar pose in the binding cavity. Totals of 86 antiprogestational scaffolds, 324 PDE4 inhibitor scaffolds, and 136 h5-HT_{2A} receptor antagonist scaffolds were identified in the WDA repository and added to the 19K database. The enrichment plots for the LAMDA-CCP and RigFit-CCP searches for the three different queries used in this study are shown in Figure 18. It will be seen that the RigFit-CCP searches are noticeably less effective than the LAMDA-CCP searches for all three activity classes, and the former method also had difficulties identifying appropriate base fragments in some cases. Therefore, only the LAMDA-CCP alignments are considered in the remainder of the discussion.

In the case of the antiprogestins, most of the known actives have a very similar steroid backbone. However, our method has been able to retrieve *cis*-tamoxifen at rank 615 within the 19K CCP data set. Tamoxifen, which is shown in Figure 14, has been filed with WD-1994-10847, and its structural similarity to antihormonal steroids has been noted previously.[55] In the case of the PDE4 query, the structure WD-1997-13346 was identified at rank 76: the corresponding superposition is shown in Figure 15. Finally, although the initial enrichment of known actives is not as good for h5-HT_{2A} antagonists as for the other two classes, LAMDA has been able to find an active compound (WD-2004-3636) with quite a different ring scaffold at rank 418; the alignment here is shown in Figure 16. When examining the enrichment results, it should be kept in mind that active compounds may use different modes of binding and that identifying new active lead structures should not be reduced to a replacement of ring scaffolds.

CONCLUSIONS

This study demonstrates that unsupervised 3D scaffold searching methods are able to find not only the types of ring systems that might be retrieved in carefully defined phar-

macophore searches but also additional, structurally diverse ring systems. These could form the starting point for the scaffold-hopping tasks that are often associated with a new lead discovery program or an attempt to rescue a project where the current drug candidate faces patent, ADME profile, or other problems. The methods have been shown to be effective in operation, and some are sufficiently rapid to permit routine scaffold searching in large chemical databases.

ACKNOWLEDGMENT

We wish to thank Tripos, Inc. for funding, the Cambridge Crystallographic Data Centre for providing the CSD database and the ConQuest software to generate the ring-template database, BioSolveIT GmbH for providing the FlexS/RigFit software, and Dr. D. Wilton for his work on the FBSS program. We are also grateful for valuable discussions with Dr. C. Lemmen (BioSolveIT GmbH) and Dr. R. D. Clark (Tripos, Inc.) regarding FlexS parameter settings and the LAMDA program, and the reviewers for their comments on an earlier draft of this paper.

Supporting Information Available: A list of CSD refcodes for the data set containing 200 structures. This material is available free of charge via the Internet at http://pubs.acs.org.

REFERENCES AND NOTES

(1) Andrews, K. M.; Cramer, R. D. Toward General Methods of Targeted Library Design: Topomer Shape Similarity Searching with Diverse Structures as Queries. *J. Med. Chem.* **2000**, *43*, 1723−1740.
(2) Schuffenhauer, A.; Gillet, V. J.; Willett, P. Similarity Searching in Files of Three-Dimensional Chemical Structures: Analysis of the BIOSTER Database Using Two-Dimensional Fingerprints and Molecular Field Descriptors. *J. Chem. Inf. Comput. Sci.* **2000**, *40*, 295−307.
(3) Sheridan, R. P. The Most Common Chemical Replacements in Drug-Like Compounds. *J. Chem. Inf. Comput. Sci.* **2002**, *42*, 103−108.
(4) Cramer, R. D.; Poss, M. A.; Hermsmeier, M. A.; Caulfield, T. J.; Kowala, M. C.; Valentine, M. T. Prospective Identification of Biologically Active Structures by Topomer Shape Similarity Searching. *J. Med. Chem.* **1999**, *42*, 3919−3933.
(5) Lauri, G.; Bartlett, P. A. CAVEAT: A Program to Facilitate the Design of Organic Molecules. *J. Comput.-Aided Mol. Des.* **1994**, *8*, 51−66.
(6) Schneider, G.; Neidhart, W.; Giller, T.; Schmid, G. Scaffold-Hopping by Topological Pharmacophore Search: A Contribution to Virtual Screening. *Angew. Chem., Int. Ed.* **1999**, *38*, 2894−2896.
(7) Olender, R.; Rosenfeld, R. A Fast Algorithm for Searching for Molecules Containing a Pharmacophore in Very Large Virtual Combinatorial Libraries. *J. Chem. Inf. Comput. Sci.* **2001**, *41*, 731−738.
(8) Bohl, M.; Dunbar, J.; Gifford, E. M.; Heritage, T.; Wild, D. J.; Willett, P.; Wilton, D. J. Scaffold Searching: Automated Identification of Similar Ring Systems for the Design of Combinatorial Libraries. *Quant. Struct.-Act. Relat.* **2002**, *21*, 590−597.
(9) Wagener, M. Scaffold Hopping by Searching a Database of Potential Scaffolds. Paper presented at the Euro-QSAR conference, Bornemouth, U. K., September 9−13, 2002.
(10) Xu, J. A New Approach to Finding Natural Chemical Structure Classes. *J. Med. Chem.* **2002**, *45*, 5311−5320.
(11) Lewell, X. Q.; Jones, A. C.; Bruce, C. L.; Harper, G.; Jones, M. M.; Mclay, I. M.; Bradshaw, J. Drug Rings Database with Web Interface. A Tool for Identifying Alternative Chemical Rings in Lead Discovery Programs. *J. Med. Chem.* **2003**, *46*, 3257−3274.
(12) Lloyd, D. G.; Buenemann, C. L.; Todorov, N. P.; Manallack, D. T.; Dean, P. M. Scaffold Hopping in *de novo* Design. Ligand Generation in the Absence of Receptor Information. *J. Med. Chem.* **2004**, *47*, 493−496.
(13) Bender, A.; Mussa, H. Y.; Gill, G. S.; Glen, R. C. Molecular Surface Point Environments for Virtual Screening and the Elucidation of Binding Patterns (MOLPRINT 3D). *J. Med. Chem.* **2004**, *47*, 6569−6583.
(14) Bender, A.; Mussa, H. Y.; Glen, R. C.; Reiling, S. Similarity Searching of Chemical Databases Using Atom Environment Descriptors (MOLPRINT 2D): Evaluation of Performance. *J. Chem. Inf. Comput. Sci.* **2004**, *44*, 1708−1718.
(15) Jenkins, J. L.; Glick, M.; Davies, J. W. A 3D Similarity Method for Scaffold Hopping from Known Drugs or Natural Ligands to New Chemotypes. *J. Med. Chem.* **2004**, *47*, 6144−6159.
(16) Kho, R.; Hodges, J. A.; Hansen, M. R.; Villar, H. O. Ring Systems in Mutagenicity Databases. *J. Med. Chem.* **2005**, *48*, 6671−6678.
(17) Ahlström, M. M.; Ridderström, M.; Luthman, K.; Zamora, I. Virtual Screening and Scaffold Hopping Based on GRID Molecular Interaction Fields. *J. Chem. Inf. Model.* **2005**, *45*, 1313−1323.
(18) Rush, T. S., III; Grant, J. A.; Mosyak, L.; Nicholls, A. A Shape-Based 3-D Scaffold Hopping Method and Its Application to a Bacterial Protein−Protein Interaction. *J. Med. Chem.* **2005**, *48*, 1489−1495.
(19) Wilkens, S. J.; Janes, J.; Su, A. I. HierS: Hierarchical Scaffold Clustering using Topological Chemical Graphs. *J. Med. Chem.* **2005**, *48*, 3182−3193.
(20) Stiefl, N.; Watson, I. A.; Baumann, K.; Zaliani, A. ErG: 2D Pharmacophore Descriptions for Scaffold Hopping. *J. Chem. Inf. Model.* **2006**, *46*, 208−220.
(21) Zhang, Q.; Muegge, I. Scaffold Hopping through Virtual Screening Using 2D and 3D Similarity Descriptors: Ranking, Voting, and Consensus Scoring. *J. Chem. Inf. Model.* **2006**, *49*, 1536−1548.
(22) SYBYL and UNITY are registered trademarks of Tripos, Inc., 1699 South Hanley Road, St. Louis, MO 63144. http://www.tripos.com (accessed June 2006).
(23) FlexS; BioSolveIT GmbH: Sankt Augustin, Germany. http://www.biosolveit.de (accessed June 2006).
(24) Lemmen, C.; Lengauer, T. Time-Efficient Flexible Superposition of Medium-Sized Molecules. *J. Comput.-Aided Mol. Des.* **1997**, *11*, 357−368.
(25) Lemmen, C.; Lengauer, T.; Klebe, G. FlexS: A Method for Fast Flexible Ligand Searching. *J. Med. Chem.* **1998**, *41*, 4502−4520.
(26) Lemmen, C.; Hiller, C.; Lengauer, T. RigFit: A New Approach to Superimpose Ligand Molecules. *J. Comput.-Aided Mol. Des.* **1998**, *12*, 491−502.
(27) Richmond, N. J.; Willett, P.; Clark, R. D. Alignment of Three-Dimensional Molecules Using an Image Recognition Algorithm. *J. Mol. Graphics Modell.* **2004**, *23*, 199−209.
(28) Wild, D. J.; Willett, P. Similarity Searching in Files of Three-Dimensional Chemical Structures: Alignment of Molecular Electrostatic Potential Fields with a Genetic Algorithm. *J. Chem. Inf. Comput. Sci.* **1996**, *36*, 159−167.
(29) Cambridge Crystallographic Data Centre, 12 Union Road, Cambridge, CB2 1EZ, U. K. is at URL http://www.ccdc.cam.ac.uk/ (accessed June 2006).
(30) Allen, F. H. The Cambridge Structural Database: A Quarter of a Million Crystal Structures and Rising. *Acta Crystallogr., Sect. B* **2002**, *58*, 380−388.
(31) Bruno, I. J.; Cole, J. C.; Edgington, P. R.; Kessler, M.; Macrae, C. F.; McCabe, P.; Pearson, J.; Taylor, R. New Software for Searching the Cambridge Structural Database and Visualizing Crystal Structures. *Acta Crystallogr., Sect. B* **2002**, *58*, 389−397.
(32) Bemis, G. W.; Murcko, M. A. The Properties of Known Drugs. 1. Molecular Frameworks. *J. Med. Chem.* **1996**, *39*, 2887−2893.
(33) Avidon, V. V.; Pomerantsev, I. A.; Golender, V. E.; Rozenblit, A. B. Structure−Activity Relationship Oriented Languages for Chemical Structure Representation. *J. Chem. Inf. Comput. Sci.* **1982**, *22*, 207−214.
(34) Rarey, M.; Dixon, J. S. Feature Trees: A New Molecular Similarity Measure Based on Tree Matching. *J. Comput.-Aided Mol. Des.* **1998**, *12*, 471−490.
(35) Rarey, M.; Stahl, M. Similarity Searching in Large Combinatorial Chemistry Spaces. *J. Comput.-Aided Mol. Des.* **2001**, *15*, 497−520.
(36) Gillet, V. J.; Willett, P.; Bradshaw, J. Similarity Searching Using Reduced Graphs. *J. Chem. Inf. Comput. Sci.* **2003**, *43*, 338−345.
(37) Rowley, M.; Bristow, L. J.; Hutson, P. H. Current and Novel Approaches to the Drug Treatment of Schizophrenia. *J. Med. Chem.* **2001**, *44*, 477−501.
(38) Rowley, M.; Hallett, D. J.; Goodacre, S.; Moyes, C.; Crawforth, J.; Sparey, T. J.; Patel, S.; Marwood, R.; Patel, S.; Thomas, S.; Hitzel, L.; O'Connor, D.; Szeto, N.; Castro, J. L.; Hutson, P. H.; MacLeod, A. M. 3-(4-Fluoropiperidin-3-yl)-2-Phenylindoles as High Aaffinity, Selective, and Orally Bioavailable h5-HT2A Receptor Antagonists. *J. Med. Chem.* **2001**, *44*, 1603−1614.
(39) Halgren, T. A. Characterization of MMFF94, MMFF94s, and Other Widely Available Force Fields for Conformational Energies and for Intermolecular Interaction Energies and Geometries. *J. Comput. Chem.* **1999**, *20*, 730−748.
(40) Dewar, M. J. S.; Zoebisch, E. G.; Healy, E. F.; Stewart, J. J. P. Development and Use of Quantum Mechanical Molecular Models. 76. AM1: a New General Purpose Quantum Mechanical Molecular Model. *J. Am. Chem. Soc.* **1985**, *107*, 3902−3909.
(41) Bohl, M. Theoretical Investigations on Steroid Structure and Quantitative Structure−Activity Relationships. In *Molecular Structure and*

Biological Activity of Steroids; Bohl, M., Duax, W. L., Eds.; CRC Press: Boca Raton, FL, 1992; pp 91−155.
(42) Bursi, R.; Groen, M. B. Application of (Quantitative) Structure−Activity Relationships to Progestagens: From Serendipity to Structure-Based Design. *Eur. J. Med. Chem.* **2000**, *35*, 787−796.
(43) Bohl, M.; Schubert, G.; Ponsold, K.; Reck, G.; Hoehne, E.; Simon, K. Molecular Mechanics and X-ray Crystal Structure Investigations on Conformations of 11β Substituted 4,9-Dien-3-One Steroids. *J. Mol. Graphics* **1989**, *7*, 122−153.
(44) Van der Mey, M.; Hatzelmann, A.; Van der Laan, I. J.; Sterk, G. J.; Thibaut, U.; Timmerman, H. Novel Selective PDE4 Inhibitors. 1. Synthesis, Structure−Activity Relationships, and Molecular Modeling of 4-(3,4-Dimethoxyphenyl)-2*H*-Phthalazin-1-Ones and Analogues. *J. Med. Chem.* **2001**, *44*, 2511−2522.
(45) Van der Mey, M.; Hatzelmann, A.; Van Klink, G. P. M.; Van der Laan, I. J.; Sterk, G. J.; Thibaut, U.; Ulrich, W. R.; Timmerman, H. Novel Selective PDE4 Inhibitors. 2. Synthesis, and Structure−Activity Relationships of 4-Aryl-Substituted *cis*-Tetra- and *cis*-Hexahydrophthalazinones. *J. Med. Chem.* **2001**, *44*, 2523−2535.
(46) Van der Mey, M.; Boss, H.; Hatzelmann, A.; Van der Laan, I. J.; Sterk, G. J.; Timmerman, H. Novel Selective PDE4 Inhibitors. 3. *In vivo* Antiinflammatory Activity of a New Series of N-Substituted *cis*-Tetra- and *cis*-Hexahydrophthalazinones. *J. Med. Chem.* **2002**, *45*, 2520−2525.
(47) Van der Mey, M.; Boss, H.; Couwenberg, D.; Hatzelmann, A.; Sterk, G. J.; Goubitz, K.; Schenk, H.; Timmerman, H. Novel Selective PDE4 Inhibitors. 4. Resolution, Absolute Configuration, and PDE4 Inhibitory Activity of *cis*-Tetra- and *cis*-Hexahydrophthalazinones. *J. Med. Chem.* **2002**, *45*, 2526−2533.
(48) Crespo, M. I.; Pagès, L.; Vega, A.; Segarra, V.; López, M.; Doménech, T.; Miralpeix, M.; Beleta, J.; Ryder, H.; Palacios, J. M. Design, Synthesis and Biological Activities of New Thieno[3, 2-*d*]pyrimidines as Selective Type 4 Phosphodiesterase Inhibitors. *J. Med. Chem.* **1998**, *41*, 4021−4035.
(49) Belongie, S.; Malik, J.; Puzicha, J. Matching Shapes and Object Recognition Using Shape Contexts. *IEEE Trans. Patt. Anal. Machine Intell.* **2002**, *24*, 509−522.
(50) Lemmen, C.; Lengauer, T. Computational Methods for the Structural Alignment of Molecules. *J. Comput.-Aided Mol. Des.* **2000**, *14*, 215−232.
(51) Brown, R. D.; Martin, Y. C. Use of Structure−Activity Data to Compare Structure-Based Clustering Methods and Descriptors for Use in Compound Selection. *J. Chem. Inf. Comput. Sci.* **1996**, *36*, 572−584.
(52) 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.
(53) Willett, P.; Barnard, J. M.; Downs, G. M. Chemical Similarity Searching. *J. Chem. Inf. Comput. Sci.* **1998**, *38*, 983−996.
(54) The World Drug Alert database is produced by Thomson/Derwent and can be found at URL http://thomsonderwent.com/products/pca/worlddrugalerts (accessed Sep 2005).
(55) Duax, W. L.; Griffin, J. F.; Weeks, C. M.; Wawrzak, Z. The Mechanism of Action of Steroid Antagonists: Insights from Crystallographic Studies. *J. Steroid Biochem.* **1988**, *31*, 481−492.

CI049657K