DOI: 10.1002/cmdc.200500005

# Scaffold-Hopping Potential of Ligand-Based Similarity Concepts

Steffen Renner and Gisbert Schneider*[a]

Manipulating living systems at the molecular level requires a profound knowledge of the variability of small-molecule effectors that provoke a particular cellular response. Medicinal chemistry relies on libraries of molecular probes that can be rationally designed to contain a desired degree of chemotype diversity. Despite great advances in the field of virtual screening and rational compound library design, "scaffold hopping" remains a challenging goal.[1] The concept of scaffold hopping is aimed at finding isofunctional but structurally dissimilar molecular entities.[2–5] The ideal screening methods that perform successful scaffold hops would not only find a maximum number but also a maximally diverse set of active compounds from a given chemical subspace. Only until recently has the focus in the development and evaluation of virtual screening methods often been purely on the retrieval of large numbers of "active" molecules, irrespective of the number of retrieved chemotypes. This has led to the impression that methods which employ a low level of abstraction from the molecular structure, such as substructure fingerprints, are among the most efficient ligand-based virtual screening methods.[6,7] In contrast to substructure-based molecular descriptors, pharmacophore models and physicochemical metrics represent a comparably high level of abstraction from chemical structure. Consequently, such methods have been employed for the design of screening libraries, relying on their scaffold-hopping potential.[2,3,8–12] In this study we compared the scaffold-hopping efficiency of topological, three-dimensional and molecular-surface-based pharmacophore pair descriptors with a popular substructure fingerprint method.

Two molecules are considered to have different scaffolds if they have different topologies.[4] This idea is based on the concept that druglike molecules are built up from a scaffold and side chains.[13] There are several reasons for seeking a set of diverse scaffolds. Different chemotypes offer a choice in terms of chemical accessibility and prospects for lead optimization. Multiple lead structures ("backup" compounds) lower the chance of attrition in drug development through undesirable ADMET (absorption, distribution, metabolism, excretion, and toxicity) properties.[5] Scaffold hopping can also be applied to move from natural substrates to more druglike chemotypes.[4,5,14] Furthermore, the creation of intellectual property is facilitated when multiple novel bioactive agents are available.

Different virtual-screening concepts have been proposed for scaffold hopping.[4] These include three-dimensional pharmacophore models,[9,15] pseudoreceptors,[16] protein-structure-based de novo design,[1,17] and ligand-based similarity searching.[18] Typically, rapid similarity searching is based on the comparison of descriptor vectors rather than on the explicit alignment of molecules to a reference and can thus be efficiently applied to screening large datasets.[18] Herein, we concentrated on such methods.

Similarity searching is founded on the similarity principle, which states that similar molecules exhibit similar biological effects.[19] A straightforward similarity-searching approach is to compare the connection tables to assess the similarity between two molecules. Such methods include substructure fingerprints like the MACCS keys,[20] which are based on exact chemical substructures. Substructure matching approaches were reported to be among the most successful for virtual screening.[6,7] The classification of intermolecular interactions into general pharmacophore types provides a way to obtain a more general description of the underlying chemotypes of molecules.[3,9] Three such descriptors were employed in the work reported herein: the topological CATS descriptor,[2,21] the three-dimensional CATS3D descriptor,[11] and the molecular-surface-based SURFCATS descriptor (Figure 1).

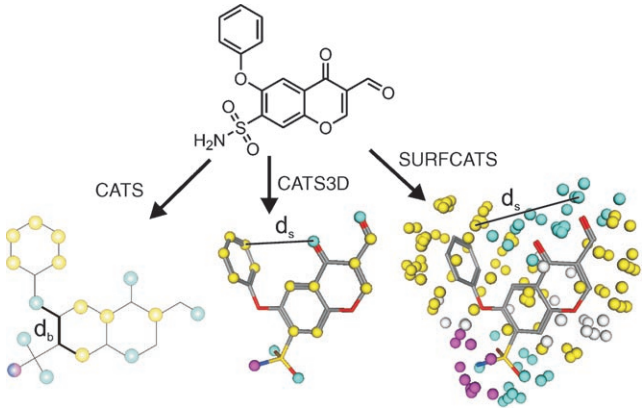


**Figure 1.** The CATS family of descriptors: CATS, CATS3D, and SURFCATS. All descriptors are based on a PPP (potential pharmacophore point)-type description of the underlying molecule. For each descriptor, pairs of PPPs are transformed into a correlation vector. CATS is calculated from the topological distances of atom-based PPP pairs. For CATS3D the spatial distances between atom-based PPPs are used instead. SURFACTS uses the spatial distances between PPPs on the contact surface of a molecule. Here the PPPs represent the atom types of the nearest atom to each surface point.

Molecular representations that are based on three-dimensional conformations like molecular surface-based descriptors are independent from the molecular connectivity and should have a favorable scaffold-hopping potential.[22,23] The three CATS descriptors describe a molecule in the form of a histogram that contains the normalized frequencies of all pairs of potential pharmacophore points (PPP) in a molecule. In our study, PPP pairs were further subdivided into PPP–PPP distan-

[a] *S. Renner, Prof. Dr. G. Schneider*
*Beilstein Endowed Chair for Cheminformatics*
*Institute of Organic Chemistry & Chemical Biology*
*Johann Wolfgang Goethe University*
*Siesmayerstraße 70, 60323 Frankfurt (Germany)*
*URL: www.modlab.de*
*Fax: (+49) 69-798-29826*
*E-mail: g.schneider@chemie.uni-frankfurt.de*

Supporting information for this article is available on the WWW under http://www.chemmedchem.org or from the author.

  

ces and different pharmacophore types. For CATS, pairs of PPPs with shortest topological distances of up to ten bonds were counted, matching at least one of the pharmacophore types: anion, cation, hydrogen-bond donor, hydrogen-bond acceptor, or hydrophobic.[21] For CATS3D and SURFCATS, pairs of PPPs were considered to fall into one of 20 equal-distance bins from 0–20 Å. For the latter two methods, one of the pharmacophore types anion, cation, hydrogen-bond donor, hydrogen-bond acceptor, polar (hydrogen-bond acceptor *and* hydrogen-bond donor), or hydrophobic were assigned with the ph4_aType function in the software suite MOE.[24] For SURFCATS, surface points were calculated with the Gauss–Connolly function in MOE with a spacing of 2 Å which were subsequently assigned to the pharmacophore type of the nearest atom. Finally, each bin of the three descriptors was scaled by the added occurrence of the respective PPPs.[11] For comparison with a conceptually different descriptor, the MACCS keys were used as implemented in MOE.

To assess the degree of scaffold hopping, one must define the term "scaffold". Herein, we followed the concept of Xu and Johnson employing the software suite Meqi,[25] which has recently been devised for the analysis of chemical libraries.[26,27] Two different definitions of a scaffold were applied: cyclic system ("Scaffold", Sc) and reduced cyclic system ("Reduced Scaffold", ReSc) (Figure 2). In Meqi, each molecular topology is

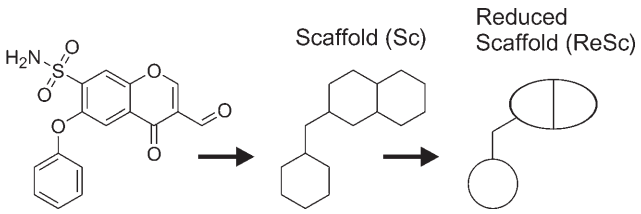


**Figure 2.** Definition of the cyclic system "Scaffold" (Sc) and the reduced cyclic system "Reduced Scaffold" (ReSc). Herein, we define the scaffold of a molecule as the side-chain-depleted molecular graph without annotation of atom types. A reduced scaffold is a more general representation which does not discriminate between rings consisting of different numbers of heavy atoms, but systems containing different numbers of rings are still not considered to be equal.

specified by a particular *m*olecular *eq*uivalence *i*ndex (meqi) which is used to distinguish between different scaffolds and reduced scaffolds.

Ligands from ten different target classes from the COBRA database[28] of annotated ligands (version 2.1, 4705 molecules) were used as reference for retrospective virtual screening: angiotensin-converting enzyme (ACE, 44 compounds, 28 scaffolds, 17 reduced scaffolds), cyclooxygenase 2 (COX2, 94, 27, 14), corticotropin-releasing factor (CRF antagonists, 63, 33, 23), dipeptidyl peptidase IV (DPP, 25, 13, 7), human immunodeficiency virus protease (HIVP, 58, 46, 31), matrix metalloproteinase (MMP, 77, 47, 19), neurokinin receptors (NK, 118, 65, 49), peroxisome proliferator-activated receptor (PPAR, 35, 29, 17), β-amyloid-converting enzyme (BACE, 44, 13, 12), and thrombin (THR, 188, 100, 36). According to the number of scaffolds and reduced scaffolds in relation to the number of molecules, the datasets range from sets with a low scaffold diversity (for example, COX2) to sets with a large relative scaffold diversity (such as PPAR and HIVP). The complete COBRA database contained 1628 different scaffolds and 637 distinct reduced scaffolds. For retrospective screening, each molecule from each target class was taken iteratively as the reference molecule for a virtual screening experiment, in which all other molecules were ranked according to their similarity to the reference molecule.

For quantification of "similarity" three similarity indices were employed: Manhattan distance [Eq. (1)], Euclidean distance [Eq. (2)], and Tanimoto similarity [Eq. (3)]:

$$D_{A,B} = \sum_{i=1}^{N} |x_{iA} - x_{iB}| \quad (1)$$

$$D_{A,B} = \sqrt{\sum_{i=1}^{N} (x_{iA} - x_{iB})^2} \quad (2)$$

$$S_{A,B} = \frac{\sum_{i=1}^{N} x_{iA}x_{iB}}{\sum_{i=1}^{N} (x_{iA})^2 + \sum_{i=1}^{N} (x_{iB})^2 - \sum_{i=1}^{N} x_{iA}x_{iB}} \quad (3)$$

in which $D_{A,B}$ denotes the distance and $S_{A,B}$ denotes the similarity between objects *A* and *B*. *N* is the total number of vector elements and $x_i$ is the value of the vector element *i*.

Virtual screening experiments were evaluated by the enrichment factor *ef,* which represents the ratio between the percentages of active molecules ("hits") in a top *x*% fraction of the ranked database to the expected percentage of active molecules. In summary, we employed ten different datasets, four descriptors (CATS, CATS3D, SURFCATS, MACCS), and three molecular representations (atomic, scaffold, and reduced scaffold) for the retrospective screening experiments.

The average relative performance of the four methods for the first 5% of the database over the ten activity classes is summarized in Figure 3. The relative performance of one particular method within one activity class was defined as the *ef* yielded with this method divided by the average *ef* of the four methods (using the same similarity index). The influence of different similarity indices on the overall enrichment was low, and for most parts was indistinguishable within the standard deviations. For all molecular representations, the ranking of the methods in terms of the enrichment factors for the top 5% of the hit lists was found to be MACCS > CATS > CATS3D ≈ SURFCATS for consideration of the average values only. With regard to the enrichment of scaffolds and reduced scaffolds, CATS, CATS3D, and SURFCATS slightly improved in comparison with the MACCS keys. An explanation for the high performance of the MACCS keys in scaffold enrichment might be that the connectivity between the substructures is not taken into account by this descriptor. This can lead to an effective retrieval of molecules with slightly different scaffolds but similar side-chain decoration.

  

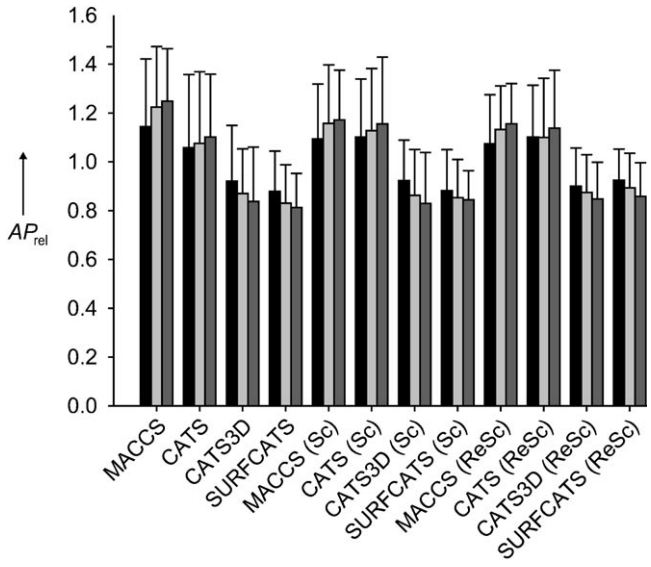


**Figure 3.** Average relative performance ($AP_{rel}$) for the first 5% over ten ligand classes from the COBRA database. Comparison of the performance of MACCS, CATS, CATS3D, and SURFCATS for molecule scaffolds (Sc) and reduced scaffolds (ReSc). Three similarity metrics were applied: the Tanimoto similarity (black bars), the Euclidean distance (light grey bars), and the Manhattan distance (dark grey bars).

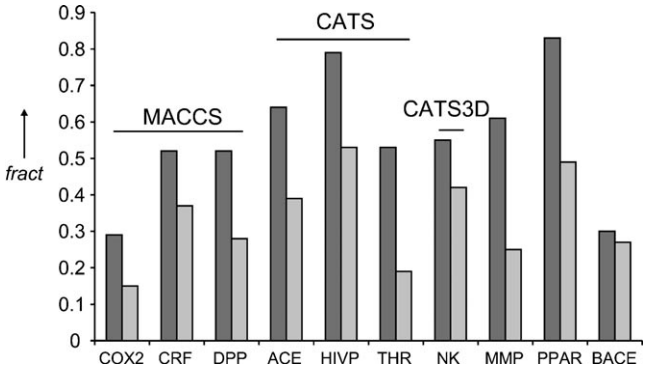


**Figure 4.** Scaffold diversity of the ligand classes. The diversity is given by the fraction (*fract*) of scaffolds (light gray) or reduced scaffolds (dark gray) relative to the number of molecules in a data set. With enrichment factors for the first 5%, MACCS performed best for the classes COX2, CRF, and DPP, CATS performed best for the classes ACE, HIVP, and THR, and CATS3D performed best for NK. For MMP, PPAR, and BACE, no method clearly dominated in performance.

Does this finding justify the conclusion that substructure fingerprints are best suited for the purpose of scaffold hopping? To find an answer to this question, a more detailed analysis was performed with observation of the enrichment of the individual ligand classes. We compared the Tanimoto *ef* values obtained by the four methods for the ten different activity classes (Supporting Information). None of the descriptors outperformed any other descriptors within the error. Judging from the average values only, MACCS performed best for COX2, CRF, and DPP for full molecules, scaffolds, and reduced scaffolds. CATS performed best for ACE, HIVP, and THR. CATS3D was best for NK in all molecular representations. SURFCATS was not found to be best for any one class. However, each descriptor of the CATS family was found to be better than the other descriptor family members for some ligand classes. This underscores the dependence of descriptor performance on the screening database. In other words, there is no *globally* best descriptor. We stress that this interpretation has limited relevance owing to large standard deviations, and therefore represents trends only. Further investigations with additional descriptors, metrics, and larger high-quality drug databases will be needed to scrutinize these findings.

Figure 4 shows the fraction of scaffolds and reduced scaffolds relative to the number of molecules found in the ten ligand classes. For the classes preferred by MACCS, the average fraction of scaffolds was 0.44 ($\pm$0.13), and the average fraction of reduced scaffolds was 0.27 ($\pm$0.11). For CATS, the fractions were 0.65 ($\pm$0.13) and 0.37 ($\pm$0.17), and for CATS3D, 0.55 and 0.42. One might speculate that MACCS performed best in classes with low numbers of different topologies, that is, low scaffold diversity, and that CATS and CATS3D performed best in classes with a high degree of scaffold diversity. We conclude that pharmacophore descriptors might be more suited for the design of diverse compound libraries than for substructure fingerprints. Still, one must be aware that these results are similar within the error margin.

In an earlier publication, we reported that different descriptors are often found to retrieve different molecules, despite having equal enrichment factors.[21] In the present study, we witness a similar situation: descriptors complement each other in the retrieval of different scaffolds and reduced scaffolds (Table 1).

**Table 1.** Overlap of the results for pairs of descriptors in the first 5% of the ranked hit list.[a]

| Descriptor | MACCS | CATS | CATS3D | SURFCATS |
|---|---|---|---|---|
| Scaffold Representations | | | | |
| MACCS | **13.8** | | | |
| CATS | 8.6 | **15.4** | | |
| CATS3D | 8.2 | 9.3 | **13.2** | |
| SURFCATS | 7.7 | 8.9 | 9.8 | **12.9** |
| Reduced Scaffold Representations | | | | |
| MACCS | **8.9** | | | |
| CATS | 6.1 | **9.8** | | |
| CATS3D | 5.8 | 6.5 | **8.7** | |
| SURFCATS | 5.3 | 6.1 | 6.5 | **8.1** |

[a] Average numbers over all ten classes of retrieved scaffold representations which were found by both methods. The numbers on the diagonal (in bold) are the average numbers of scaffolds found with the respective descriptor. The similarity index employed was the Tanimoto coefficient.

Two of the virtual hit lists were further investigated: the results for the COX2 inhibitors celecoxib (Figure 5) and rofecoxib (Supporting Information). For each scaffold class, the best-ranking hits were surveyed. Although the two reference molecules share a common reduced scaffold, different scaffold classes were retrieved at different ranking positions. Again, the four

  

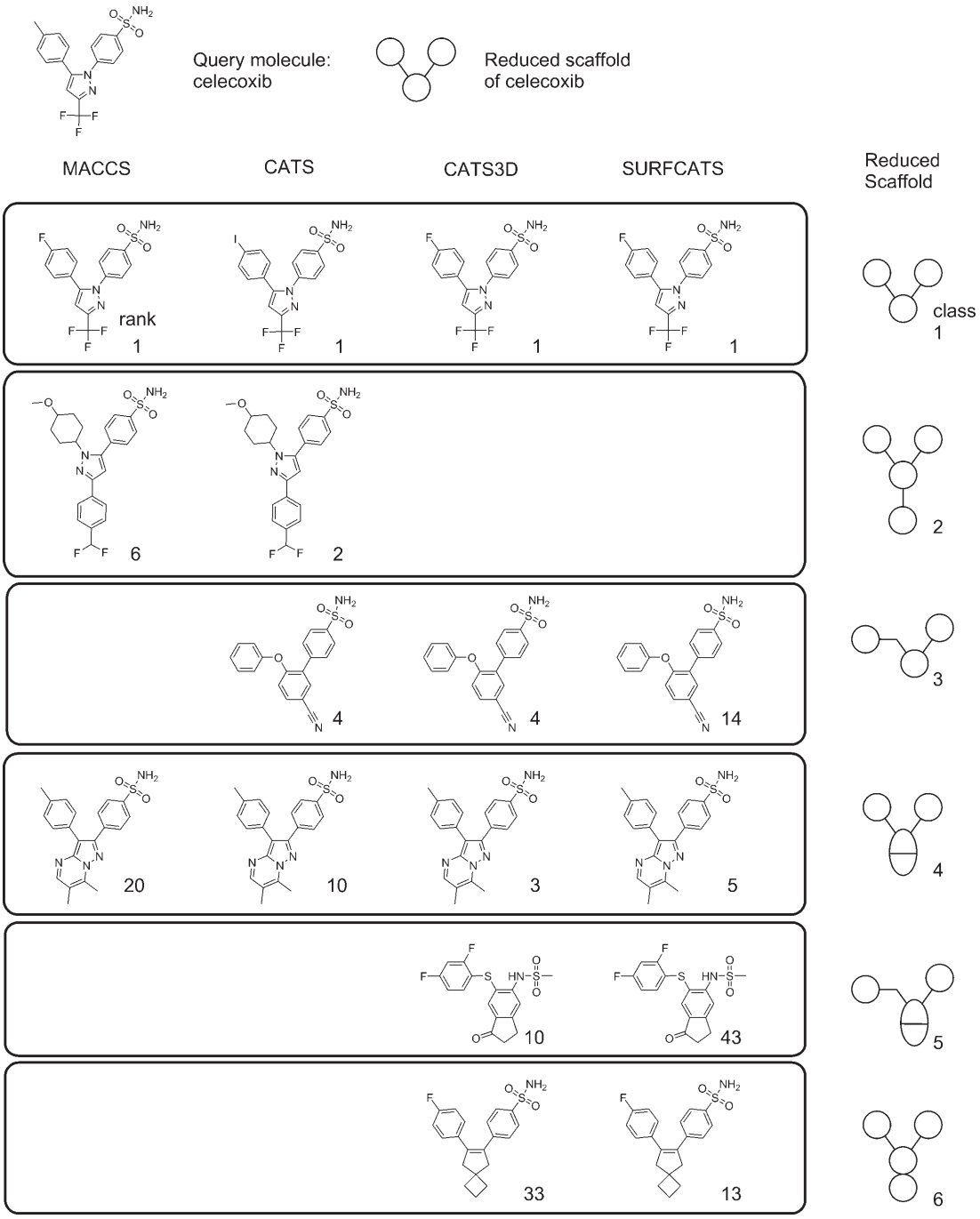


**Figure 5.** Best hits for each reduced scaffold obtained with celecoxib. For each descriptor, the best-scored molecule in each reduced-scaffold class that was retrieved in the first percent of the ranked database is shown.

similarity-searching methods differed in their ability to retrieve diverse scaffolds which results in a complementation of the methods. This outcome is remarkable, especially in light of the relatedness of the query structures.

The two reduced scaffolds that were found exclusively with the MACCS keys for rofecoxib (ReSc classes 6 and 7) reflect that MACCS keys include no direct information of the size of the retrieved molecules (Supporting Information). These molecules might have been rejected by the other methods as a result of their large size. Still large reduced scaffolds were also found with CATS for celecoxib (ReSc class 2), which might have resulted from the restriction of the descriptor to a maximal path length of 10 bonds in the present study. Such a cutoff might be inappropriate for a database with potentially long ligands and respective pharmacophores such as those annotated to HIVP, MMP, and PPAR—particularly in prospective screens.

In conclusion, we found that both substructure fingerprints (MACCS) and pharmacophore-pair descriptors (CATS) are suited for retrospective scaffold retrieval. For more diverse ligand

  

classes, the pharmacophore-based CATS descriptors slightly outperformed substructure (MACCS) keys as an average trend. The fact that structurally focused collections of pharmacologically active compounds are typically employed for retrospective screening studies might explain the often-found high performance of substructure keys or related descriptors.[6,7] Our results suggest that for the particular purpose of scaffold hopping, a reasonable strategy might be to use more generalized molecular representations like pharmacophore descriptors. The use of several complementing methods can be recommended for the purpose of scaffold hopping. We hope that our study will stimulate further investigations on this important topic of medicinal chemistry.

## *Acknowledgements*

*S.R. is grateful for a PhD fellowship from Merz Pharmaceuticals GmbH, Frankfurt am Main. This research was supported by the Beilstein-Institut zur Förderung der Chemischen Wissenschaften, Frankfurt am Main, and the Deutsche Forschungsgemeinschaft (DFG, SFB 579, project A11).*

**Keywords:** bioinformatics · drug design · molecular diversity · similarity searching · virtual screening

[1] G. Schneider, U. Fechner, *Nat. Rev. Drug Discovery* **2005**, *4*, 649 – 663.
[2] G. Schneider, W. Neidhart, T. Giller, G. Schmid, *Angew. Chem.* **1999**, *111*, 3068 – 3070; *Angew. Chem. Int. Ed.* **1999**, *38*, 2894 – 2896.
[3] G. Schneider, O. Clément-Chomienne, L. Hilfiger, P. Schneider, S. Kirsch, H.-J. Böhm, W. Neidhart, *Angew. Chem.* **2000**, *112*, 4305 – 4309; *Angew. Chem. Int. Ed.* **2000**, *39*, 4130 – 4133.
[4] H.-J. Böhm, A. Flohr, M. Stahl, *Drug Discovery Today Technol.* **2004**, *1*, 217 – 224.
[5] J. L. Jenkins, M. Glick, J. W. Davies, *J. Med. Chem.* **2004**, *47*, 6144 – 6159.
[6] R. D. Brown, Y. C. Martin, *J. Chem. Inf. Comput. Sci.* **1996**, *36*, 572 – 584.
[7] J. Hert, P. Willett, D. J. Wilton, P. Acklin, K. Azzaoui, E. Jacoby, A. Schuffenhauer, *Org. Biomol. Chem.* **2004**, *2*, 3256 – 3266.
[8] H. Matter, *J. Med. Chem.* **1997**, *40*, 1219 – 1229.
[9] J. S. Mason, A. C. Good, E. J. Martin, *Curr. Pharm. Des.* **2001**, *7*, 567 – 597.
[10] L. Naerum, L. Norskov-Lauritsen, P. H. Olesen, *Bioorg. Med. Chem. Lett.* **2002**, *12*, 1525 – 1528.
[11] S. Renner, T. Noeske, C. G. Parsons, P. Schneider, T. Weil, G. Schneider, *ChemBioChem* **2005**, *6*, 620 – 625.
[12] S. Renner, V. Ludwig, O. Boden, U. Scheffer, M. Göbel, G. Schneider, *ChemBioChem* **2005**, *6*, 1119 – 1125.
[13] G. W. Bemis, M. A. Murcko, *J. Med. Chem.* **1996**, *39*, 2887 – 2893.
[14] M.-L. Lee, G. Schneider, *J. Comb. Chem.* **2001**, *3*, 284 – 289.
[15] I. J. P. De Esch, J. E. Mills, T. D. Perkins, G. Romeo, M. Hoffmann, K. Wieland, R. Leurs, W. M. Menge, P. H. Nederkoorn, P. M. Dean, H. Timmermann, *J. Med. Chem.* **2001**, *44*, 1666 – 1674.
[16] D. G. Lloyd, C. L. Buenemann, N. P. Todorov, D. T. Manallak, P. M. Dean, *J. Med. Chem.* **2004**, *47*, 493 – 496.
[17] M. Stahl, N. P. Todorov, T. James, H. Mauser, H.-J. Böhm, P. M. Dean, *J. Comput.-Aided Mol. Des.* **2002**, *16*, 459 – 478.
[18] P. Willett, J. M. Barnard, G. M. Downs, *J. Chem. Inf. Comput. Sci.* **1998**, *38*, 983 – 996.
[19] M. A. Johnson, G. M. Maggiora, *Concepts and Applications of Molecular Similarity*, Wiley, New York, **1990**.
[20] MACCS keys, MDL Information Systems Inc., San Leandro, CA, USA (http://www.mdl.com).
[21] U. Fechner, L. Franke, S. Renner, P. Schneider, G. Schneider, *J. Comput.-Aided Mol. Des.* **2003**, *17*, 687 – 698.
[22] A. Bender, H. Y. Mussa, G. S. Gill, R. C. Glen, *J. Med. Chem.* **2004**, *47*, 6569 – 6583.
[23] T. Clark, *J. Mol. Graphics Modell.* **2004**, *22*, 519 – 525.
[24] MOE Molecular Operating Environment, Chemical Comuting Group Inc., Montréal, Canada (http://www.chemcomp.com).
[25] Pannanugget Consulting L.L.C., Kalamazoo, MI, USA (http://www.pannanugget.com).
[26] Y. J. Xu, M. Johnson, *J. Chem. Inf. Comput. Sci.* **2001**, *41*, 181 – 185.
[27] Y. J. Xu, M. Johnson, *J. Chem. Inf. Comput. Sci.* **2002**, *42*, 912 – 926.
[28] "COBRA: Collection of bioactive reference compounds for focused library design": P. Schneider, G. Schneider, *QSAR Comb. Sci.* **2003**, *22*, 713 – 718.

Received: June 28, 2005
Revised: October 14, 2005
Published online on December 27, 2005

  