# Impact of Conformational Flexibility on Three-Dimensional Similarity Searching Using Correlation Vectors

Steffen Renner,[†] Christof H. Schwab,[‡] Johann Gasteiger,[§] and Gisbert Schneider*[,†]

Institut für Organische Chemie und Chemische Biologie, Johann Wolfgang Goethe-Universität, Siesmayerstrasse 70, D-60323 Frankfurt am Main, Germany, Molecular Networks GmbH Computerchemie, Nägelsbachstrasse 25, D-91052 Erlangen, Germany, and Computer-Chemie-Centrum, Universität Erlangen-Nürnberg, Nägelsbachstrasse 25, D-91052 Erlangen, Germany

Received March 2, 2005

Many three-dimensional (3D) virtual screening concepts, like automated docking or pharmacophore searching, rely on the calculation of a "bioactive" or "receptor-relevant" conformation of a molecule to assess its biological activity. We investigated the dependence of the presence of conformations near the "bioactive" conformation on three-dimensional similarity searching with pharmacophore-based correlation vectors (CATS3D approach). Cocrystal structures of 11 target classes served as queries for virtual screening of a database of annotated ligands. Different numbers of conformations were calculated. Single 3D structures were obtained using the 3D structure generator CORINA and conformational ensembles by the conformation generation program ROTATE. This approach was able to reproduce conformations for high resolution cocrystal structures. For virtual screening we found that using only the CORINA-generated single conformation already resulted in a significant enrichment of isofunctional molecules having the same biological property profile. This observation was also made for ligand classes with many rotatable bonds. Although more similar conformations were considered to be more similar in the CATS3D description, the impact of using multiple conformations on the enrichment of actives was not as high as expected. CATS3D provides an alignment-free three-dimensional virtual screening approach that is less dependent on the presence of conformations which are close to the "bioactive" conformation of a molecule compared to methods that rely on an explicit three-dimensional alignment of molecules.

## INTRODUCTION

Understanding the principles of ligand−receptor interactions is a key prerequisite on the way toward an understanding of biological systems at the molecular level. Computational virtual screening is an often applied approach in pharmaceutical research that builds on such knowledge to efficiently propose novel ligand candidates for a given receptor.[1] Virtual screening for novel ligands corresponds to searching for molecules that can comprise the necessary arrangement of interacting groups which are essential for binding. A variety of computational methods is available for the estimation of ligand binding properties among which automated docking methods and pharmacophore models belong to the most widely used applications.[2] Both concepts are based on fitting a three-dimensional conformation to a query and thus fundamentally rely on the presence of the potentially "bioactive" conformation (a receptor-bound conformation) of the molecule under consideration. This conformation is not easy to find since it usually does not correspond to the global energy minimum conformation in an unbound state; for many ligands their bound conformation does not correspond to a minimum at all.[3] A recent publication has evaluated commercially available software for their ability to reproduce biological active conformations of molecules.[4] While this is a comparably easy task for small and rigid ligands with only a few rotatable bonds, all programs tested in this study often failed to reproduce the bioactive conformation of ligands with more than eight rotatable bonds. Although it is clear that in principle each bioactive conformation could be reproduced, there are still practical limits in the number of conformations that can be handled efficiently due to the exponential increase in the number of potential conformations with an increasing number of rotatable bonds.[5]

A strategy to avoid the computation of many conformations can be to use representations of the molecules that are less sensitive to the exact three-dimensional conformation of a molecule. One possibility is to compare molecules based on their topological graph-representation. One such approach is based on atom pairs,[6] e.g. topological pharmacophore pairs as given by the CATS descriptor:[7] A molecule is represented by a multidimensional histogram giving relative frequencies of pairs of pharmacophore points. This representation provides a molecular descriptor vector that can be directly used to measure the distance or similarity to other molecules. In addition, this representation is rotation- and translation-invariant and thus enables rapid pairwise comparisons of molecules without the need to explicitly align the molecules. The original idea of this kind of alignment-free molecular representation was described by Moreau and Broto[8] and further explored by Gasteiger and co-workers.[9] Recently, we

* Corresponding author phone: +49-69-798-24874; fax: +49-69-798-24880; e-mail: gisbert.schneider@modlab.de.
[†] Johann Wolfgang Goethe-Universität.
[‡] Molecular Networks GmbH Computerchemie.
[§] Universität Erlangen-Nürnberg.

10.1021/ci050075s CCC: $33.50 
Published on Web 08/08/2006

published CATS3D,[10,11] a three-dimensional (3D) extension of the CATS approach, that is conceptually similar to the 3D autocorrelation vector of Sheridan and co-workers[12] and that has already been shown to be successful in prospective screening, in particular for scaffold hopping,[11,13,14] or in combination with fuzzy pharmacophore models.[13,15] We compared CATS and CATS3D for their ability to retrieve isofunctional molecules from a database of known bioactive molecules for various targets in retrospective screening experiments.[10] Using only a single 3D conformation per molecule for the reference and database molecules, it was shown that even the incorporation of this possibly erroneous 3D information often led to an improvement in retrospective screening. While it is obvious that virtual screening methods that are based on explicit three-dimensional representations of the molecules (such as docking experiments or pharmacophore searches) strongly rely on the presence of conformations as close as possible to the "bioactive" conformation, there might be a weaker dependence for methods that are based on more abstract representations of the molecules such as the CATS3D approach.

In the present study, we examined the influence of the incorporation of different amounts of multiple conformations on the ability of the CATS3D approach to find isofunctional molecules in a retrospective screening experiment. For this study we used a large reference set of high-resolution cocrystal structures, the *PDBbind* database,[16] as our reference for ligand based retrospective screening in the COBRA database.[17] For this set of ligands we compared the effect of different numbers of predicted ligand conformations in a screening database (the COBRA database) on similarity searching with the aim to result in activity-enriched subsets of isofunctional molecules.

For the conformer calculation we used the rule- and data-based programs CORINA[18] (single conformations) and ROTATE[5,19] (multiple conformations). Since we were interested in the calculation of conformations for large data sets for virtual screening rather than an exhaustive exploration of the conformation space, we used a strategy to avoid both extensive calculation times and numbers of conformations. The conformation space was sampled by ROTATE only to a limited extent in order to guarantee a manageable amount of conformers per molecule and reasonable CPU times for both the generation of conformers and virtual screening experiments. Since ROTATE was not included in the study of ref 4, we first evaluated ROTATE using the same setup and data set as in ref 4 in order to get an idea of the performance of the program.

## MATERIALS AND METHODS

**Data Sets**. Two data sets of high-quality crystal structures of receptor-bound ligands were used as reference. The Boström data set[4] was taken from a published evaluation of conformation generation programs to evaluate the performance of ROTATE. The data set consisted of 32 cocrystal structures with a resolution of 2 Å and better, containing molecules with one to 10 rotatable bonds. The *PDBbind* database[16] (version 2002) served as a larger reference set of high-quality crystal structures of receptor-bound ligands for the virtual screening experiments. For retrospective virtual screening we used the COBRA database[17] (version 3.12) consisting of 5376 annotated ligands that were compiled from scientific literature. The ligands of the *PDBbind* database were grouped according to their target annotation. We removed all clusters containing less than five ligands. Clusters were also removed for which no ligands were found in the COBRA database with the same target annotation as in *PDBbind*. From multiple incidences of identical ligands all but the one with the best resolution were removed. The final set of reference ligands consisted of 11 groups ("activity classes") with a total number of 177 ligands. The corresponding set of "active" ligands in the COBRA database contained 674 molecules, which means that the COBRA database contained 4702 additional ligands that were not considered as "active" in either of the 11 activity classes. The final annotated activity classes and their abbreviations were as follows: acetylcholinesterase (ACHE, 6 compounds from *PDBbind*, 13 compounds from COBRA, overlap: 0), carbonic anhydrase II (CAII, 30, 25, 2), elastase (ELA, 8, 8, 0), factor Xa (FXA, 5, 226, 5), HIV-protease (HIVP, 58, 61, 8), neuraminidase (NEU, 8, 28, 1), protein tyrosine kinase c-src (PTK-CSRC, 7, 16, 0), protein tyrosine phosphatase 1b (PTP1B, 5, 36, 0), stromelysin 1 (STRO1, 7, 19, 0), thrombin (THR, 32, 194, 10), and urokinase type plasminogen activator (UTPA, 11, 48, 3). Since we were not interested in the absolute performance of the method, but in the relative performance using different degrees of conformational information, we did not remove ligands that were present in both databases ("overlap"). An overview over the average number of rotatable bonds and the average molecular weights of the activity classes is given in Table 1. Prior to further procession of the data all molecules were neutralized with a script written in the SVL-language of MOE. [20]

**Table 1.** Average Number of Rotatable Bonds and Molecular Weight of the Activity Classes[a]

| | PDBbind | | COBRA | |
|---|---|---|---|---|
| activity class | rotatable bonds | molecular weight | rotatable bonds | molecular weight |
| ACHE | 6.7 (5.8) | 334 (116) | 8.2 (4.5) | 253 (81) |
| CAII | 7.2 (3.5) | 321 (84) | 7.3 (3.0) | 366 (100) |
| ELA | 16.4 (3.2) | 545 (60) | 10.9 (4.2) | 431 (126) |
| FXA | 11.0 (5.3) | 435 (29) | 12.0 (5.7) | 489 (82) |
| HIVP | 21.7 (9.7) | 637 (116) | 19.3 (6.2) | 614 (116) |
| NEU | 12.9 (1.4) | 305 (20) | 12.6 (7.7) | 320 (130) |
| PTK-CSRC | 24.9 (3.4) | 557 (64) | 7.6 (3.2) | 444 (80) |
| PTP1B | 6.0 (0.0) | 277 (34) | 10.2 (6.4) | 464 (150) |
| STRO1 | 12.9 (6.4) | 487 (108) | 17.1 (5.6) | 489 (106) |
| THR | 10.7 (5.0) | 423 (125) | 15.4 (5.1) | 500 (107) |
| UTPA | 6.6 (1.9) | 294 (86) | 10.2 (5.8) | 165 (116) |

[a] Values in parentheses are standard deviations.

**Generation of 3D Structures and Ensembles of Conformations**. The 3D structure generator CORINA (version 2.64)[18] was used to calculate single 3D structures for each compound in the COBRA database. These single conformations were then used as input geometries for the conformer generator ROTATE (version 1.15).[19] For exploration of the conformation space, ROTATE applies a set of rules that resulted from a statistical analysis of the conformational preferences of the open-chain portions in small molecule crystal structures.[21] The rules and corresponding data are stored in the so-called Torsion Angle Library (TA library) that has been derived from the Cambridge Structural

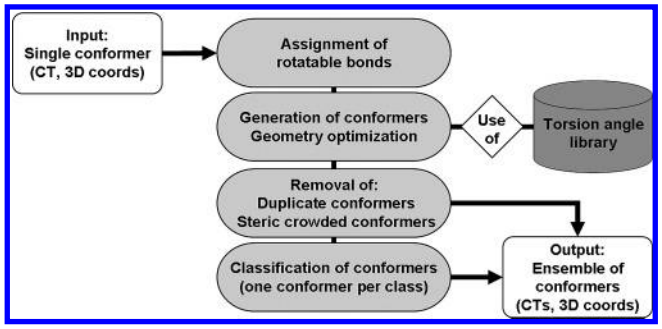


**Figure 1.** General principle of the ROTATE algorithm.

Database (CSD).[22] Figure 1 shows the general concept of the algorithm.

The connection table and a 3D model of the molecule (such as obtained from CORINA) are required as input to ROTATE. After assignment of rotatable bonds the rotamers are generated by using the rules and the data contained in the TA library. Currently, the TA library contains over 900 entries of torsion angle fragments (torsion patterns) consisting of four atoms with their adjacent neighbors and a single rotatable bond in the center. For each pattern the observed distributions of torsion angles are stored as histograms. These histograms contain the implicit information about the conformational preferences of molecules in a structured molecular environment with varying intermolecular directional forces and dielectric conditions in the different crystal packings. It was shown that the geometries adopted by small molecules in crystal structures are correlated to those of ligand conformations at the binding site of a biological receptor.[23] Thus, this approach generates observed conformations likely to be of biological relevance.

After assigning appropriate torsion angle histograms from the TA library to each rotatable bond the histograms were transformed into empirical potential energy functions.[21] These symbolic energy functions were used to select a set of preferred torsion angles for the rotatable bond under consideration (initial torsion angles), and all possible combinations of the initial torsion angles of all rotatable bonds were generated. After a new conformation was generated it was geometry optimized by applying the empirical energy function and was only accepted in the absence of steric hindrance. Duplicate conformations were rejected.

The number of conformations drastically increases with the number of rotatable bonds. To restrict the number of generated conformers, while retaining a maximum of structural diversity, similar geometries were combined into classes. The classification used in this study was based on the root-mean-square-deviation (RMSD) between the conformations in torsion angle space. The RMS threshold which decides whether two conformations belong to the same class is adjustable, and each class is finally represented by a single conformation. Besides the classification based on the RMSD between all torsion angles of two conformations, ROTATE also offers a comparison of rotamers in Cartesian coordinates space. Furthermore, energy thresholds with respect to the lowest-energy conformations can be defined, and the number of rotatable bonds to be processed can be specified and weighted by the topological position within the molecule (innermost rotatable bonds are rotated first). Thus, the conformational space of a molecule can be explored to a degree required by the user.

**Superimposition and Calculation of the RMSD Value**. The RMSD of atomic coordinates of a reference conformation and a ROTATE-generated conformation served as quality criterion defining pairwise conformational "similarity". Calculations of RMSD values were done with the program MATCH3D, a least-squares superimposition procedure in Cartesian coordinate space.[24] This software takes into account the symmetry of nondistinguishable but differently numbered groups (e.g. the two oxygen atoms in a carboxylate group) for the calculation and thereby avoids artificially introduced high RMSD values. Only nonhydrogen atoms were considered for the calculation.

**Calculation of the CATS3D Descriptor**. CATS3D encodes the conformation of a molecule in the form of a correlation vector (CV) that contains the scaled frequencies of all pairs of atoms of a molecule. Atom-pairs were subdivided into groups that were characterized by atom−atom distance ranges and six different pharmacophore types. Twenty equal distance bins from 0 to 20 Å were used. The large distance range up to 20 Å was used to incorporate almost all atom-pairs found in the ligands. Previous analysis indicated that only a few ligands exceeded a length of 20 Å (results not shown). Bin widths were applied based on previous experience from successful applications of the CATS3D descriptor. One of the pharmacophore types cation, anion, hydrogen-bond acceptor, hydrogen-bond donor, polar (hydrogen-bond acceptor AND hydrogen-bond donor), or hydrophobic was assigned to each atom using the ph4_aType function of MOE.[20] Using 20 distance bins for each of the 21 possible combinations of pairs of pharmacophore points resulted in a 420-dimensional descriptor. The value stored in each bin was divided by the added incidences of the two respective features. Each dimension of the CATS3D CV was calculated according to eq 1

$$\mathrm{CV}_d^T = \frac{1}{N_1 + N_2} \sum_{i=1} \sum_{j=1} \frac{1}{2} \delta_{ij,d}^T \qquad (1)$$

where $i$ and $j$ are atom indices, $d$ is a distance range, $T$ is the pair of pharmacophoric types of atoms $i$ and $j$, $N_1$ and $N_2$ are the total number of atoms of types of $i$ and $j$ present in a molecule, and $\delta_d^T$ (Kronecker delta) evaluates to 1 for all pairs of atoms of type $T$ within the distance range $d$. The factor of 0.5 in the sum avoids double counting of pairs. Pairs of atoms with themselves were not considered.

**Retrospective Virtual Screening**. For retrospective screening the general procedure was to rank a database according to increasing distance to a reference molecule.[10] The distance of a database entry to a reference molecule was calculated by the Manhattan distance (eq 2)

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

where $A$ and $B$ indicate the CATS3D vectors of two molecules, $x_i$ is the value of vector element $i$, and $N$ is the total number of vector elements ($N = 420$). For databases of multiple conformations only the best scoring conformation (the conformations with the smallest distance) for each molecule was retained for the final ranked list. Based on the ranked list an enrichment factor was calculated as a

performance measure for each retrospective screening experiment. The enrichment factor ef was defined as given by eq 3

$$\mathrm{ef} = \left(\frac{F_{\mathrm{act}}}{F_{\mathrm{all}}}\right) / \left(\frac{D_{\mathrm{act}}}{D_{\mathrm{all}}}\right) \quad (3)$$

where $F_{act}$ and $D_{act}$ are the numbers of molecules from the activity class of the reference that are found in the subset and are contained in the whole database, and $F_{all}$ and $D_{all}$ are the total numbers of molecules in the subset and in the whole database, respectively. An enrichment factor of 1 corresponds to a random distribution of active molecules in the ranked database, thus, an effective pharmacophore model should result in an ef > 1.

For each activity class the enrichment factors of all respective reference molecules were averaged to allow for statistical interpretation.

## RESULTS AND DISCUSSION

**Reproduction of "Bioactive" Conformations with the Program ROTATE.** As a first experiment we wanted to test the ability of the program ROTATE to reproduce "bioactive" conformations of small molecules in general. For this purpose we used a data set of ligand−receptor complexes that was used previously for an evaluation of several conformer generation programs,[4] the Boström data set with CORINA-generated 3D models as starting geometries for the conformer generation process. In the original study by Boström a restrictive RMSD threshold of 0.5 Å was used to define a "hit", that is, a successful reproduction of the crystal structure conformation. Considering the fuzzy molecular representation of molecules in the CATS3D descriptors using atom-pairs and distance bins of 1 Å, a less stringent thresholds might be sufficient for the purpose of our study. We reexamined the results of Boström and found that higher RMSD thresholds (e.g. 0.6 and 1.0 Å) led to higher hit rates for all programs and different runs. Applying higher RMSD thresholds takes into account that the resolution of the ligand atoms can be significantly lower (B-factors of up to 32) than the overall resolution of a protein−ligand complex.[4]

For comparison we used ROTATE to generate conformations for the ligands of the Boström data set. As input structure the CORINA-generated model was taken. A maximum of six preferred torsion angles per rotatable bond was used and only the five most central rotatable bonds in a molecule were processed. The latter was done to restrict the number of calculated conformations by processing only these rotatable bonds that should evoke the largest conformational changes. For final classification of all rotamers the comparison in torsion angle space was performed with different torsion angle thresholds: a threshold of RMSD ≤ 15° (resulting set of conformational ensembles further referred to as R4) in order to remove only very similar conformations (which resulted in the highest total number of output conformations) and thresholds of ≤ 120° (R1), ≤ 60° (R2), and ≤ 45° (R3). The results obtained with ROTATE are shown in Table 2.

In terms of reproducing biologically active conformations the best performing program in Boström's study was MacroModel[25] using the AMBER force field and a GB/SA solvent model that reproduced 22 of the 32 ligand conformations. This method was also the most time-consuming approach with almost 10.5 h of CPU time (SGI workstation MIPS R10000, 180 MHz). The fastest method in Boström's study, OMEGA,[26] only needed 18 s of CPU time but reproduced only 13 of the 32 ligand conformations. Therefore, with the reproduction of 19 ligand conformations in about 160 s (x86 Linux workstation, AMD Athlon XP 1800+) in the R4 ensemble, ROTATE performed with good quality and acceptable CPU time compared to the other conformer generator packages and methods tested.

In terms of average number of output conformations, ROTATE generated 122 conformers compared to 155 by MacroModel and 13 by OMEGA. One hundred twenty-two conformations per molecule still is a fairly large number to be handled, especially if the number of molecules drastically increases in data sets for virtual screening. As shown in Table 2, the average number of conformers decreases if the RMSD threshold to filter out similar geometries increases (R1 to R3 ensembles). With three to 39 conformations on average in the R1 to R3 ensemble, these numbers of conformations

**Table 2.** Best Fit, Defined by the Lowest RMS Deviation between the ROTATE-Generated Conformations Using Different RMS Thresholds in Torsion Angle Space for Classification and the Unmodified X-ray Structures of the 32 Ligands[b]

| PDB code | rotatable bonds | best RMSD (in [Å]) in R1 | R2 | R3 | R4 |
|---|---|---|---|---|---|
| 1a28 | 1 | *0.42* | *0.42* | *0.42* | *0.42* |
| 1tng | 1 | *0.09* | *0.09* | *0.09* | *0.09* |
| 1tnh | 1 | *0.22* | *0.22* | *0.22* | *0.22* |
| 1qft | 2 | *0.27* | *0.27* | *0.27* | *0.13* |
| 1ftm | 3 | *0.17* | 0.56 | *0.17* | *0.17* |
| 1phg | 3 | 0.70 | 0.70 | *0.47* | *0.38* |
| 3bto | 3 | 0.54 | 0.54 | 0.54 | 0.54 |
| 1d3g | 3 | 0.61 | *0.30* | *0.44* | *0.30* |
| 1c83 | 4 | *0.38* | *0.38* | *0.38* | *0.38* |
| 1ecv | 4 | *0.39* | *0.39* | *0.37* | *0.37* |
| 1fcz | 4 | *0.31* | *0.31* | *0.31* | *0.31* |
| 1gr2 | 4 | 0.64 | 0.64 | 0.64 | 0.60 |
| 1ian | 4 | 1.04 | *0.36* | *0.38* | *0.38* |
| 1bzs | 4 | 1.52 | 1.52 | 1.52 | 1.40 |
| 1frb | 4 | 1.14 | 0.61 | 0.59 | 0.59 |
| 1bju | 5 | 0.51 | *0.50* | *0.50* | *0.50* |
| 1dyr | 5 | 1.03 | 0.66 | *0.47* | *0.43* |
| 2izg | 5 | 0.94 | 0.69 | *0.41* | *0.41* |
| 1cbx | 5 | 0.55 | *0.29* | *0.27* | *0.27* |
| 5std | 5 | 0.92 | 0.51 | 0.69 | *0.33* |
| 6std | 5 | 0.90 | 0.90 | 0.53 | 0.53 |
| 7std | 5 | 1.06 | *0.45* | *0.45* | *0.45* |
| 1cbs | 5 | *0.46* | *0.46* | *0.46* | *0.46* |
| 1dam | 6 | 0.96 | 0.75 | 0.75 | 0.52 |
| 1ejn | 7 | 0.59 | 0.52 | 0.51 | *0.50* |
| 1caq | 8 | 1.62 | 1.44 | 1.44 | 1.39 |
| 1mtv | 8 | 1.13 | 1.10 | 1.10 | 1.07 |
| 1mtw | 8 | 1.53 | 1.27 | 1.27 | 1.27 |
| 1ppc | 8 | 2.03 | 1.35 | 1.35 | 1.35 |
| 1f0u | 10 | 1.77 | 1.07 | 0.96 | 0.94 |
| 1fkg | 11 | 1.84 | 1.83 | 1.68 | 1.68 |
| 1pph | 11 | 1.78 | 1.46 | 0.86 | 0.86 |
| av no. of conformers | | 3 | 22 | 39 | 122 |
| best mean RMSD [Å] | | 0.88 | 0.71 | 0.64 | 0.60 |
| hits (0.5 Å) | | *9* | *13* | *17* | *19* |
| hits (≤1.0 Å) | | 20 | 24 | 26 | 26 |
| CPU time [s][a] | | 138 | 141 | 145 | 163 |

[a] x86 Linux workstation, AMD Athlon XP 1800+. [b] Hits yielding an RMSD ≤ 0.5 Å are italicized.

**Table 3.** Average Number of Conformations per Molecule Calculated for 11 Activity Classes[a]

| | PDBbind | | | COBRA | | |
|---|---|---|---|---|---|---|
| activity class | R1 | R2 | R3 | R1 | R2 | R3 |
| ACHE | 1.7 | 10.7 | 19.2 | 1.5 | 10.7 | 20.8 |
| CAII | 3.5 | 18.9 | 31.3 | 3.1 | 22.0 | 38.9 |
| ELA | 3.8 | 34.9 | 51.5 | 2.8 | 27.0 | 54.9 |
| FXA | 3.8 | 38.4 | 69.0 | 3.9 | 32.8 | 59.4 |
| HIVP | 2.9 | 30.1 | 56.4 | 2.8 | 28.9 | 51.4 |
| NEU | 2.0 | 24.5 | 45.1 | 2.4 | 19.0 | 35.2 |
| PTK-CSRC | 2.6 | 26.7 | 48.0 | 4.0 | 21.9 | 41.4 |
| PTP1B | 4.0 | 14.6 | 25.6 | 3.2 | 25.4 | 46.9 |
| STRO1 | 3.4 | 35.4 | 64.9 | 3.2 | 31.0 | 55.6 |
| THR | 3.5 | 26.5 | 49.6 | 3.4 | 34.3 | 63.6 |
| UTPA | 3.3 | 11.0 | 19.2 | 3.3 | 22.8 | 48.0 |
| average | 3.1 | 24.7 | 43.6 | 3.1 | 25.1 | 46.9 |

[a] Multiple conformations were calculated with the following thresholds for the classification in torsion angle space: R1: 120°, R2: 60°, and R3: 45°.

per molecule seem to be more appropriate and manageable for our retrospective screening experiments where several thousands of structures had to be handled and analyzed. Although the number of hits, i.e., the number of reproduced ligand conformations decreased with the 0.5 Å criterion, using the less stringent hit criterion of 1 Å ROTATE still reproduced 20 (R1), 24 (R2), and 26 (R3) of the bioactive conformations. For comparison, all methods in the Boström study reproduced on average 25.7 of the 32 reference conformations by applying the hit criterion of ≤1 Å, in particular MacroModel (AMBER force field, GB/SA solvent model) with 28 and OMEGA with 26 conformations. With ROTATE, only the five most central (innermost) rotatable bonds of a molecule were taken into account for conformer generation. Therefore, ROTATE can be regarded to be successful in the generation of a conformation that is close to the bioactive one (≤1 Å) rather than exactly reproducing the receptor bound geometry (hit criterion of 0.5 Å).

The major goal of our study was to investigate the influence of using multiple conformations on the enrichment of actives in result sets of virtual screening experiments and not to perform an exhaustive exploration of the conformational space of a molecule. Therefore, we decided to use ROTATE with program parameter settings that resulted in the R1, R2, and R3 ensembles for any of our further screening experiments in order to ensure a manageable number of conformations per molecule and reasonable CPU times.

**Impact of Conformational Flexibility on Similarity Searching.** *Calculation of Conformations for the PDBbind Data Set and the COBRA Database.* Conformations were calculated for the selected reference molecules from *PDBbind* database and all molecules from the COBRA database. For each database single conformations were calculated with CORINA. These single conformations were then submitted to ROTATE applying the strategy as described in the previous section for the generation of multiple conformations.

Three databases differing in the number of multiple conformations were calculated using ROTATE. For the final classification we used torsion angle thresholds of 120° (resulting database further referred to as R1), 60° (R2), and 45° (R3). Table 3 gives an overview over the average number of conformations that were calculated per molecule for the different activity classes of both data sets. On average, approximately three conformations were generated for each molecule in the R1 data sets, roughly 25 conformations in the R2, and about 45 conformations per molecule in the R3 data sets. For some of the activity classes (e.g. PTP1B, UTPA) the number of conformations differed remarkably between the reference data set and the COBRA database. This indicates that the topological similarity between the entries of the two collections was low.

*Reproducing the Crystal-Structure Conformations of Reference Ligands.* To assess the reproduction of the receptor-bound conformations of the *PDBbind* reference ligands we calculated the RMSD value of all generated conformations to their corresponding experimentally determined geometry. The results of the calculation are shown in Table 4. Boström considered conformations with an RMSD of less than 0.5 Å to their reference structure as successfully reproduced conformations. According to this threshold, only for two activity classes (PTP1B, UTPA) the bioactive conformation could be reproduced, even with the R3 database containing the largest number of calculated conformations. Applying the less stringent RMSD criterion of 1 Å, the CORINA conformations already reproduced the bioactive conformation for three of the 11 activity classes (NEU, PTP1B, UTPA). For six activity classes the bioactive conformation could be reproduced in the R3 database. The best RMSD values were found for PTP1B and UTPA, the two classes with the minimum of rotatable bonds of 6 and 6.5 on average (Table 1), using the maximum of conformations. Only for two classes RMSD values higher than 1.3 Å were obtained: for HIVP and PTK-CSRC. This was anticipated since these ligands contained the largest number of rotatable bonds (21.7 for HIVP, 24.9 for PTK-CSRC), and we allowed only the five most central rotatable bonds of a molecule to be processed. The largest improvement using more conformations could be obtained for Factor Xa inhibitors (FXA) which contain 11 rotatable bonds on average (Table 1). For UTPA and PTP1B the second- and the third-best improvement were found. The smallest improvement was obtained for PTK-CSRC, which is probably caused by the fact that not all rotors were processed for the generation of multiple conformations, and these two classes had the most additional bonds that were not rotated.

An example of the Factor Xa inhibitor Fxv673 (PDB code: 1KSN) CORINA conformation (red) and the best R3 conformation (green) superimposed onto the reference is shown in Figure 2. The bound ligand conformation has a central kink that is not found in the geometrically more stretched CORINA conformation (RMSD = 2.3 Å). The best ROTATE conformation reproduced the kink which resulted in an improved RMSD of 1.1 Å. Summarizing, using more conformations resulted in a lower RMSD, and in most cases conformations were found close to the receptor-bound conformation.

*Relationship between Conformation RMSD and CATS3D Distance.* After it was demonstrated that the procedure for conformer calculation was successful in retrieving ligands near to the "bioactive" conformation, we investigated to which degree this behavior would be reflected in CATS3D. CATS3D descriptors were calculated for all conformations of the *PDBbind* subsets. The average of the best found Manhattan distances of the calculated conformations to the

**Table 4.** Average Best Cartesian RMSD of the Calculated Conformations to the Reference Conformations of the *PDBbind* Molecules[a]

| | RMSD in Å | | | | improvement over CORINA | | |
|---|---|---|---|---|---|---|---|
| activity class | CORINA | R1 | R2 | R3 | *I* (R1) | *I* (R2) | *I* (R3) |
| ACHE | 1.5 (1.8) | 1.1 (1.2) | 0.7 (0.7) | 0.7 (0.8) | 1.4 | 2.1 | 2.0 |
| CAII | 1.1 (0.4) | 0.9 (0.4) | 0.8 (0.4) | 0.8 (0.4) | 1.2 | 1.4 | 1.5 |
| ELA | 2.2 (0.4) | 1.6 (0.3) | 1.4 (0.4) | 1.3 (0.4) | 1.4 | 1.6 | 1.7 |
| FXA | 2.0 (0.4) | 1.7 (0.4) | 1.0 (0.2) | 0.8 (0.2) | 1.2 | 2.1 | 2.5 |
| HIVP | 3.1 (0.8) | 2.6 (0.9) | 2.2 (0.7) | 2.1 (0.7) | 1.2 | 1.4 | 1.5 |
| NEU | 1.0 (0.6) | 1.2 (0.4) | 0.8 (0.5) | 0.8 (0.5) | 0.9 | 1.3 | 1.3 |
| PTK-CSRC | 2.2 (0.5) | 2.2 (0.2) | 1.8 (0.1) | 1.8 (0.1) | 1.0 | 1.2 | 1.2 |
| PTP1B | 0.9 (0.2) | 0.5 (0.1) | 0.4 (0.1) | 0.4 (0.0) | 1.6 | 2.1 | 2.4 |
| STRO1 | 2.1 (0.8) | 1.5 (0.7) | 1.2 (0.6) | 1.2 (0.6) | 1.4 | 1.7 | 1.8 |
| THR | 1.9 (0.9) | 1.4 (0.8) | 1.2 (0.7) | 1.1 (0.7) | 1.4 | 1.6 | 1.7 |
| UTPA | 0.9 (0.5) | 0.5 (0.4) | 0.4 (0.3) | 0.4 (0.2) | 1.7 | 2.3 | 2.4 |
| Average | 1.7 (0.7) | 1.4 (0.6) | 1.1 (0.6) | 1.0 (0.5) | 1.3 (0.2) | 1.7 (0.4) | 1.8 (0.5) |

[a] Improvements are given for usage of multiple conformations in comparison to the RMSD obtained with a single CORINA conformation. The improvement *I* (Rx, $x = 1,2,3$) was calculated by RMSD (CORINA conformation)/RMSD (best Rx conformation). Values in parentheses are standard deviations.

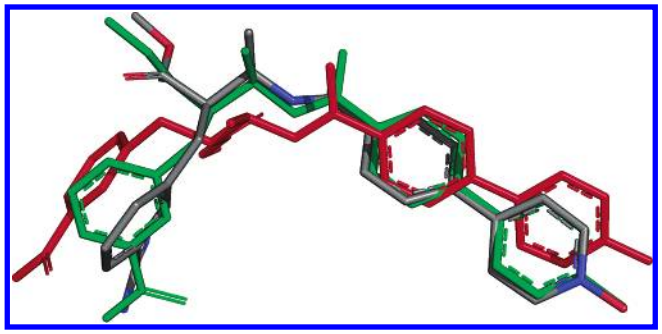

**Figure 2.** Superposition of the CORINA conformation (red) and the best R3 conformation (green) of the Factor Xa inhibitor Fxv673 (PDB code 1KSN) to the reference conformation from the crystal structure.

respective crystal structure conformations for each set of ligands is shown in Table 5. The relation between the RMSD values from molecular superposition and the CATS3D distances is shown in Figure 3. For most target classes a decrease in distance was observed using conformations with lower RMSD. This means that more similar conformations were considered to be more similar also by the CATS3D description.

Applying the CATS3D Manhattan distance criterion an average 1.5-fold improvement was obtained using more conformations. For comparison with the RMSD criterion using the R3 conformations resulted in an average 1.8-fold improvement over the single CORINA conformation. This discrepancy probably is an effect of binning atom-pairs into bins of 1 Å. Small conformational modifications changing the distances between pairs of atoms less than 1 Å do not drastically affect the CATS3D descriptor. Consequently, the increasing similarity of the calculated conformations to the crystal structure references from single CORINA conformations to R3 is reflected in the CATS3D molecule description, though not as strong as in the RMSD value obtained from explicit molecular superposition.

*Retrospective Screening.* To determine the impact of multiple conformations on 3D similarity searching the presumable "bioactive" conformations of the reference ligands selected from the *PDBbind* database were used to screen the COBRA database for ligands with similar biological activity. The results of the experiments are compiled in Table 6. Most reference classes were able to significantly enrich the first percent of the ranked database with molecules from the same activity class. Only for PTK-CSRC and PTP1B no active molecules at all were found in the top 1% of the ranked database.

For probing the impact of multiple conformations for similarity searching with CATS3D correlation vectors we were interested in the improvement of using multiple conformations over single conformations and not just in the overall performance of each class. Interestingly, while improvement was observed for some of the activity classes, no obvious improvement in the enrichment factor was found when multiple conformations were incorporated (average improvement = 1.1). Omitting the ef deterioration observed for ACHE, a slightly better average of 1.3 was found. The largest improvement was observed for FXA and THR yielding an improvement in the enrichment factor of 1.8 for R2 and R3, respectively. For the other activity classes much smaller improvements were detected. In all cases no large difference in the ef between R2 and R3 was observed.

Furthermore, no obvious correlation between the improvement of the RMSD from Table 5 and the improvement in similarity searching (Table 6) was found. Figure 4 shows the plots of the enrichment factors versus the best RMSD values to the receptor-bound (bioactive) conformation found in the various conformational ensembles (single CORINA conformation, R1, R2, and R3) for the different activity classes. For example for UTPA, for which the RMSD could be largely improved for the *PDBbind* data set, the usage of multiple conformations for COBRA led only to a small improvement for R3. On the other hand, FXA resulted in the largest improvement in both RMSD and enrichment. For THR even a larger ef-improvement (1.8-fold) was found compared to the CATS3D distance (1.4-fold improvement, Table 5). HIVP and STRO1, the two classes with the most rotatable bonds in the COBRA data set, showed nearly no improvement for R3. In R1 both classes even showed a small deterioration in the ef. This is likely to be due to the rotation of only the five innermost rotatable bonds in the molecules. This limitation seems to prevent the reproduction of the crystal structure, that is, the presumably "bioactive" conformation. This substantiates the observation in the previous experiment where the two classes with more than 20 rotatable bonds resulted in the two largest RMSD values (HIVP and PTK-CSRC in Table 4). Nevertheless, regarding the ef

**Table 5.** Average Best CAT3D Manhattan Distance of the Calculated Conformations to the Reference Conformations of the *PDBbind* Molecules[a]

| | CATS3D Manhattan distance | | | | improvement over CORINA | | |
|---|---|---|---|---|---|---|---|
| activity class | CORINA | R1 | R2 | R3 | *I* (R1) | *I* (R2) | *I* (R3) |
| ACHE | 3.3 (3.7) | 3.3 (3.7) | 3.0 (3.2) | 2.7 (2.8) | 1.0 | 1.1 | 1.2 |
| CAII | 3.3 (2.1) | 2.7 (2.1) | 2.5 (1.8) | 2.4 (1.8) | 1.2 | 1.3 | 1.4 |
| ELA | 7.0 (2.0) | 5.0 (1.5) | 4.5 (1.5) | 4.4 (1.4) | 1.4 | 1.6 | 1.6 |
| FXA | 7.5 (3.1) | 5.4 (2.0) | 3.8 (0.9) | 3.7 (1.0) | 1.4 | 2.0 | 2.0 |
| HIVP | 9.2 (3.2) | 8.9 (3.1) | 7.7 (2.9) | 7.4 (2.9) | 1.0 | 1.2 | 1.2 |
| NEU | 4.6 (1.4) | 3.6 (1.4) | 3.1 (1.5) | 3.0 (1.6) | 1.3 | 1.5 | 1.5 |
| PTK-CSRC | 10.6 (4.3) | 9.7 (3.4) | 9.1 (3.2) | 9.1 (3.2) | 1.1 | 1.2 | 1.2 |
| PTP1B | 2.7 (0.9) | 2.4 (0.7) | 2.1 (0.5) | 1.8 (0.5) | 1.1 | 1.3 | 1.5 |
| STRO1 | 6.0 (1.9) | 4.9 (1.6) | 4.3 (1.8) | 4.2 (1.8) | 1.2 | 1.4 | 1.4 |
| THR | 6.6 (4.0) | 5.4 (3.3) | 4.9 (2.9) | 4.8 (2.8) | 1.2 | 1.3 | 1.4 |
| UTPA | 2.5 (1.6) | 1.7 (1.0) | 1.6 (1.1) | 1.5 (0.9) | 1.5 | 1.6 | 1.7 |
| average | 5.8 (2.7) | 4.8 (2.6) | 4.2 (2.3) | 4.1 (2.3) | 1.2 (0.1) | 1.4 (0.2) | 1.5 (0.2) |

[a] Improvements are given for usage of multiple conformations in comparison to the distances obtained with a single CORINA conformation. The improvement *I* (Rx, $x = 1,2,3$) was calculated by distance (CORINA conformation)/distance (best Rx conformation). Values in parentheses are standard deviations.

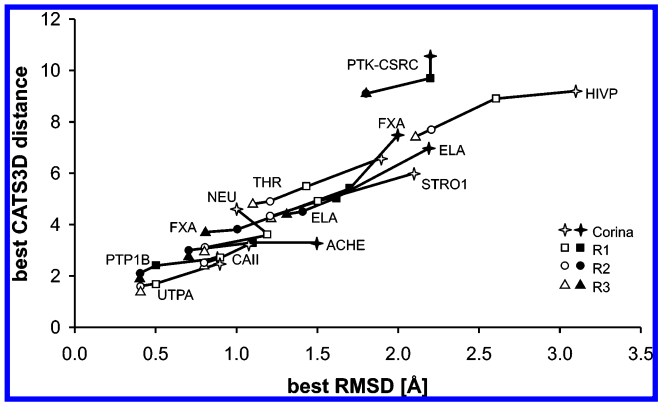


**Figure 3.** Best average CATS3D Manhattan distances (cf. Table 5), versus the best RMSD values (cf. Table 4), obtained with the four conformational ensembles (CORINA, R1, R2, and R3; cf. Table 5) for each of the activity classes. ACHE: acetylcholinesterase, CAII: carbonic anhydrase II, ELA: elastase, FXA: factor Xa, HIVP: HIV-protease, NEU: neuraminidase, PTK-CSRC: protein tyrosine kinase c-src, PTP1B: protein tyrosine phosphatase 1b, STRO: stromelysin 1, THR: thrombin, UTPA: urokinase type plasminogen activator.

values, both HIVP and STRO1 performed well, even with a single conformation. In contrast, THR, which ranked third in the number of rotatable bonds in the COBRA database, improved significantly with more conformations (Table 6, Figure 4). For ELA with an average number of 11.9 rotatable bonds in the COBRA database (Table 1), the enrichment did not increase although the RMSD to the receptor-bound conformation was lowered from 2.2 Å (CORINA single conformation) to 1.3 Å (R3 database).

To find an explanation for the low impact of multiple conformations on CATS3D similarity searching, we further investigated our data sets by the self-similarity and cross-similarity of the *PDBbind* and COBRA data sets. We calculated the average Manhattan distances between all molecules in the two data sets (self-similarity) and the average Manhattan distance between all molecules from the *PDBbind* data sets to all molecules from the respective sets from the COBRA database (cross-similarity). When more than a single conformation per molecules was present, the lowest distance was used. Furthermore, we recorded the average Manhattan distance from the *PDBbind* references to the best 10 inactives found in virtual screening. The results of this analysis are shown in Figure 5.

Comparing the self-similarities and cross-similarities of the two subsets for each target class the values were found to be similar within the standard deviation for CA, FXA, HIVP, NEU, STRO, and THR, which indicates similar distributions. Large cross-similarities relative to the self-similarity of at least one of the target class subsets from either *PDBbind* or COBRA were found for PTK-CSRC and PTP1B, which could explain the lack of retrieved hits in the retrospective screening experiments.

Using more conformations in the cross-similarity resulted in an average improvement of R3 over CORINA of 1.1 for all target classes except ACHE with 1.0. This is considerably lower than the average improvement obtained from conformations compared to the "bioactive" conformation of the same molecule using CATS3D (improvement = 1.5, Table 5) and the RMSD (improvement = 1.8, Table 4). These results indicate that for less similar molecules − though still binding to the same receptor − the impact of the particular conformation decreases in CATS3D. Since this effect was found for all target classes, it is unlikely that it results from different binding modes or binding to different binding sites on the receptor. In particular, since considerable enrichment was found for nine of the 11 target classes, the ligands within the activity classes were shown to be more similar than randomly selected molecules. The following consideration might present a reasonable explanation for the observed lower impact of conformations on molecular similarity for more distant molecules: (1) with increasing spatial distance between atom-pairs, the more likely their distance should fall into different distance bins considering different conformations. (2) Atom-pairs within a large distance should be less frequent than atom-pairs with small and intermediate distances. Actually, this observation was made by Zauhar et al. using ray-tracing techniques to compute distance frequencies.[27] As a consequence, it is more likely that two molecules are considered to be similar on the basis of small and medium atom-pair distances than on the basis of large distances.

Using more conformations also increased the similarity of false-positives: considering the 10 best-ranking inactives in the virtual hit lists, an average improvement of 1.1 was observed (1.1 for all target classes except for ELA with an

**Table 6.** Result of the Retrospective Screening of the COBRA Database with the *PDBbind* Reference Structures[a]

| | enrichment factor | | | | improvement over CORINA | | |
|---|---|---|---|---|---|---|---|
| activity class | CORINA | R1 | R2 | R3 | *I* (R1) | *I* (R2) | *I* (R3) |
| ACHE | 5.1 (4.0) | 2.5 (3.9) | 1.3 (3.1) | 1.3 (3.1) | 0.5 | 0.3 | 0.3 |
| CAII | 3.8 (4.0) | 4.5 (3.9) | 4.6 (4.0) | 4.6 (4.7) | 1.2 | 1.2 | 1.2 |
| ELA | 1.6 (4.4) | 1.6 (4.4) | 1.6 (4.4) | 1.6 (4.4) | 1.0 | 1.0 | 1.0 |
| FXA | 4.8 (2.4) | 7.0 (2.3) | 8.5 (2.3) | 8.7 (2.5) | 1.5 | 1.8 | 1.8 |
| HIVP | 12.2 (11.8) | 11.4 (11.2) | 13.2 (13.1) | 13.3 (13.3) | 0.9 | 1.1 | 1.1 |
| NEU | 22.2 (10.5) | 21.3 (11.0) | 23.5 (10.0) | 22.1 (10.3) | 1.0 | 1.1 | 1.0 |
| PTK-CSRC | 0.0 (0.0) | 0.0 (0.0) | 0.0 (0.0) | 0.0 (0.0) | | | |
| PTP1B | 0.0 (0.0) | 0.0 (0.0) | 0.0 (0.0) | 0.0 (0.0) | | | |
| STRO1 | 9.0 (8.9) | 6.7 (7.2) | 8.2 (9.0) | 9.0 (10.7) | 0.7 | 0.9 | 1.0 |
| THR | 2.9 (3.7) | 4.0 (4.8) | 5.2 (5.8) | 5.3 (5.8) | 1.4 | 1.8 | 1.8 |
| UTPA | 4.9 (5.8) | 6.6 (8.4) | 6.2 (8.9) | 6.0 (8.9) | 1.3 | 1.3 | 1.2 |
| average | 6.0 (6.5) | 6.0 (6.1) | 6.6 (6.9) | 6.5 (6.6) | 1.1 (0.3) | 1.1 (0.4) | 1.1 (0.4) |

[a] Enrichment factors were calculated for the first percent of the ranked databases. The improvement *I* (Rx, $x = 1,2,3$) was calculated by ef (best Rx conformation)/ef (CORINA conformation). For the calculation of the average improvement the improvement for PTK-CSRC and PTP1B were set to 1. Values in parentheses are standard deviations.

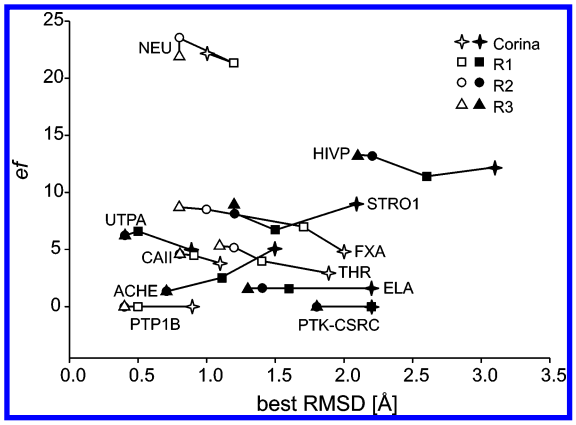


**Figure 4.** Enrichment factors, ef (cf. Table 6), versus the best RMSD values, obtained with the four conformational ensembles (CORINA, R1, R2, and R3; cf. Table 4) for each of the activity classes. If less than four data points are visible for an activity class, some of the data points have equal values. ACHE: acetylcholinesterase, CAII: carbonic anhydrase II, ELA: elastase, FXA: factor Xa, HIVP: HIV-protease, NEU: neuraminidase, PTK-CSRC: protein tyrosine kinase c-src, PTP1B: protein tyrosine phosphatase 1b, STRO: stromelysin 1, THR: thrombin, UTPA: urokinase type plasminogen activator.

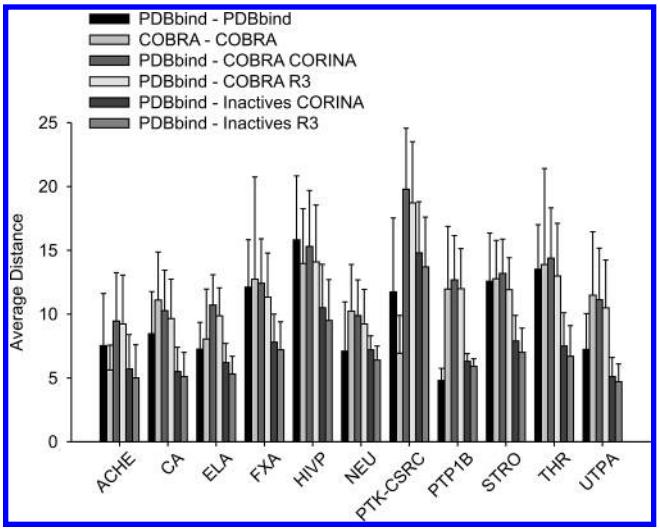


**Figure 5.** Comparison of self-similarity and cross-similarity of the *PDBbind* and COBRA data sets, using different numbers of conformations (CORINA and R3). Additionally, the average distances of the best 10 inactives obtained from the virtual screenings to the receptor bound queries were compared using different amounts of conformations. Error bars indicate the standard deviations.

improvement of 1.2). In our experiments this increase was found to be in the same range as the average over all active molecules. Since the ranks of the inactives are considered for calculation of the ef value, it is not unexpected that no clear correlation between ef and RMSD was found.

The low effect of using multiple conformations on our virtual screening seems to be based on a combination of several effects. The results and the improvement in virtual screening depended on the data set, in particular on the query structures. More distant isofunctional molecules were less sensitive to conformational differences according to CATS3D compared to topologically identical molecules. This could explain why most of the actives in the COBRA data set were insensitive to the actual conformation of the query molecule. Furthermore, false-positives were found on better ranks in the virtual hit lists with increasing numbers of conformations. The observed low sensitivity of CATS3D to the particular three-dimensional conformation explains the reported successful applications of this descriptor employing only a single conformation per molecule.[10,14] When a higher sensitivity to the actual conformation of a molecule is a desired goal, one might consider using smaller bin widths or adding weights for more distant atom-pairs.

## CONCLUSIONS

The program ROTATE was found to be successful in finding "bioactive" conformations, although the conformation space was only searched to a limited extent. Further investigations applying different strategies to explore the conformation space with ROTATE are ongoing. These search strategies apply flexible values for the number of processed rotors, the number of torsion angle values per rotor, and the RMSD threshold for the classification of conformers depending on the number of rotatable bonds in the molecule. Furthermore, the impact of the usage of TA libraries derived from other small molecule crystal structure databases on the performance of ROTATE is analyzed. Investigating the impact of multiple conformations on 3D similarity searching with CATS3D demonstrated that using only a single conformation per molecule, as the one obtained from CORINA, already resulted in an enrichment of actives. This

observation was also made for ligand classes with many rotatable bonds. On average these results only slightly improved when using multiple conformations. The particular enrichment and the corresponding improvement yielded using more conformations seemed to be dependent on the activity class, that is, the biological target under consideration. Multiple conformations resulted in the retrieval of more false positives. Furthermore, topologically different molecules seemed to be less sensitive to conformational changes using the CATS3D description than identical molecules. The combination of these effects led to the situation that no clear correlation between the improvement of the enrichment factor and the improvement of the RMSD to the bioactive conformation could be derived when screening with CATS3D correlation vectors. Since the calculation of multiple conformations is computationally expensive, we suggest the use of single conformations for large databases, e.g. virtual combinatorial libraries, for first-pass virtual screening. Single conformations can be computed efficiently with CORINA, even for large databases. In subsequent screening rounds, e.g. with smaller data sets or flexible ligands, the consideration of multiple conformations might lead to an improvement of the virtual screening results for the particular data set under investigation.

## ACKNOWLEDGMENT

We are grateful to Norbert Dichter for technical assistance. This research was supported by the Beilstein-Institut zur Förderung der Chemischen Wissenschaften.

## REFERENCES AND NOTES

(1) *Virtual Screening for Bioactive Molecules*; Böhm, H.-J., Schneider, G., Eds.; Wiley-VCH: Weinheim, New York, 2000.
(2) Schneider, G.; Böhm, H.-J. Virtual Screening and Fast Automated Docking Methods. *Drug Discovery Today* **2002**, *7*, 64−70.
(3) Perola, E.; Charifson, P. S. Conformational Analysis of Drug-Like Molecules Bound to Proteins: An Extensive Study of Ligand Reorganization upon Binding. *J. Med. Chem.* **2004**, *47*, 2499−2510.
(4) Boström, J. Reproducing the Conformation of Protein-Bound Ligands: A Critical Evaluation of Several Popular Conformational Searching Tools. *J. Comput.-Aided Mol. Des.* **2002**, *15*, 1137−1152.
(5) Schwab, C. H. Conformational Analysis and Searching. In *Handbook of Cheminformatics*; Gasteiger, J., Ed.; Wiley-VCH: Weinheim, New York, 2003; pp 262−301.
(6) Carhart, R. E.; Smith, D. H.; Venkataraghavan, R. Atom Pairs as Molecular Features in Structure−Activity Studies: Definition and Applications. *J. Chem. Inf. Comput. Sci.* **1985**, *25*, 64−73.
(7) 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.
(8) Moreau, G.; Broto, P. The Autocorrelation of a Topological Structure: A New Molecular Descriptor. *Nouv. J. Chim.* **1980**, *4*, 359−360.
(9) Wagener, M.; Sadowski, J.; Gasteiger, J. Autocorrelation of Molecular Surface Properties for Modeling Corticosteroid Binding Globulin and Cytosolic Ah Receptor Activity by Neural Networks. *J. Am. Chem. Soc.* **1995**, *117*, 7769−7775.
(10) Fechner, U.; Franke, L.; Renner, S.; Schneider, P.; Schneider, G. Comparison of Correlation Vector Methods for Ligand-Based Similarity Searching. *J. Comput.-Aided Mol. Des.* **2003**, *17*, 687−698.
(11) Renner, S.; Noeske, T.; Parsons, C. G.; Schneider, P.; Weil, T.; Schneider, G. New Allosteric Modulators of Metabotropic Glutamate Receptor 5 (mGluR5) Found by Ligand-Based Virtual Screening. *ChemBioChem* **2005**, *6*, 620−625.
(12) Sheridan, R. P.; Miller, M. D.; Underwood, D. J.; Kearsley, S. K. Chemical Similarity Using Geometric Atom Pair Descriptors. *J. Chem. Inf. Comput. Sci.* **1996**, *36*, 128−136.
(13) Renner, S.; Ludwig, V.; Boden, O.; Scheffer, U.; Göbel, M.; Schneider, G. New Inhibitors of the Tat-TAR RNA Interaction Found with a "Fuzzy" Pharmacophore Model. *ChemBioChem* **2005**, *6*, 1119−1125.
(14) Renner, S.; Schneider, G. Scaffold-Hopping Potential of Ligand-Based Similarity Concepts. *ChemMedChem* **2006**, *1*, 181−185.
(15) Renner, S.; Schneider, G. Fuzzy Pharmacophore Models from Molecular Alignments for Correlation-Vector Based Virtual Screening. *J. Med. Chem.* **2004**, *47*, 4653−4664.
(16) Wang, R.; Fang, X.; Lu, Y.; Wang, S. The PDBbind Database: Collection of Binding Affinities for Protein−Ligand Complexes with Known Three-Dimensional Structures. *J. Med. Chem.* **2004**, *47*, 2977−2980.
(17) Schneider, P.; Schneider, G. Collection of Bioactive Reference Compounds for Focused Library Design. *QSAR Comb. Sci.* **2003**, *22*, 713−718.
(18) Sadowski, J.; Gasteiger, J.; Klebe, G. Comparison of Automatic Three-Dimensional Model Builders Using 639 X-Ray Structures. *J. Chem. Inf. Comput. Sci.* **1994**, *34*, 1000−1008. (b) CORINA is available from Molecular Networks GmbH, Erlangen, Germany (http://www.mol-net.com).
(19) ROTATE is available from Molecular Networks GmbH, Erlangen, Germany (www.mol-net.com).
(20) MOE, Molecular Operating Environment. Distributor is available from Chemical Computing Group, Montreal, Canada (www.chemcomp.com).
(21) Klebe, G.; Mietzner, T. A Fast and Efficient Method to Generate Biologically Relevant Conformations. *J. Comput.-Aided Mol. Des.* **1994**, *8*, 583−606. (b) Klebe, G.; Mietzner, T.; Weber, F. Different Approaches Toward an Automatic Structural Alignment of Drug Molecules: Applications to Sterol Mimics, Thrombin and Thermolysin Inhibitors. *J. Comput.-Aided Mol. Des.* **1999**, *13*, 35−49.
(22) Allen, F. H.; Kennard, O.; Watson, D. G. Crystallographic Databases: Search and Retrieval of Information from the Cambridge Structural Database. In *Structure Correlation*; Bürgi, H.-B., Dunitz, J. D., Eds; Wiley-VCH: Weinheim, New York, 1994; Vol. 1, pp 71−110. (b) The Cambridge Structural Database (CSD) is available from Cambridge Crystallographic Data Centre, Cambridge, U.K. (www.ccdc.cam.ac.uk).
(23) Klebe, G. Structure Correlation and Ligand/Receptor Interactions. In *Structure Correlation;* Bürgi, H.-B., Dunitz, J. D., Eds.; Wiley-VCH: Weinheim, New York, 1994; Vol. 2, pp 543−603.
(24) Personal communications, Jens Sadowski, AstraZeneca R&D Mölndal, Sweden (MATCH3D was developed at the Computer-Chemie-Centrum, University of Erlangen-Nürnberg, Germany).
(25) Mohamadi, F.; Richards, N. G. J.; Guida, W. C.; Liskamp, R.; Lipton, M.; Caufield, C.; Chang, G.; Hendrickson, T.; Still, W. C. Macromodel − An Integrated Software System for Modeling Organic and Bioorganic Molecules Using Molecular Mechanics. *J. Comput. Chem.* **1990**, *11*, 440−467. (b) MacroModel is available from Schrödinger LLC, Portland, OR, U.S.A. (www.schrodinger.com).
(26) Boström, J.; Greenwood, J. R.; Gottfries, J. Assessing the Performance of OMEGA with Respect to Retrieving Bioactive Conformations. *J. Mol. Graphics Modell.* **2003**, *21*, 449−462. (b) OMEGA is available from OpenEye Scientific Software, Santa Fe, NM, U.S.A. (www.eyesopen.com).
(27) Zauhar, R. J.; Moyna, G.; Tian, L.; Li, Z.; Welsh, W. J. Shape Signatures: A New Approach to Computer-Aided Ligand- and Receptor-Based Drug Design. *J. Med. Chem.* **2003**, *46*, 5674−5690.

CI050075S