Articles

Fuzzy Pharmacophore Models from Molecular Alignments for Correlation-Vector-Based Virtual Screening

Steffen Renner and Gisbert Schneider*

Johann Wolfgang Goethe-Universität, Institut für Organische Chemie und Chemische Biologie, Marie-Curie-Strasse 11, D-60439 Frankfurt, Germany

Received December 15, 2003

A pharmacophore-based approach for compiling focused screening libraries is presented. It integrates information from three-dimensional molecular alignments into correlation vector-based database screening. The pharmacophore model is represented by a number of spheres of Gaussian-distributed feature densities. Different degrees of "fuzziness" can be introduced to influence the model's resolution. Transformation of this pharmacophore representation into a correlation vector results in a vector of feature probabilities which can be utilized for rapid virtual screening of compound databases or virtual libraries. The approach was validated by retrospective screening for cyclooxygenase 2 (COX-2) and thrombin ligands. A variety of models with different degrees of fuzziness were calculated and tested for both classes of molecules. Best performance was obtained with pharmacophore models reflecting an intermediate degree of fuzziness, yielding an enrichment factor of up to 39 for the first 1% of the ranked database. Appropriately weighted fuzzy pharmacophore models performed better in retrospective screening than similarity searching using only a single query molecule. The new pharmacophore method was shown to complement existing approaches.

Introduction

Virtual screening has proven to be a fast and efficient tool for hit discovery and optimization.[1] Virtual compound classification prior to experimental testing is a very fast and cost saving way of finding new screening candidates with desired biological activity. Classification methods range from simple comparison of physicochemical descriptors to complex estimations of binding constants by docking calculations or quantitative structure−activity relationship (QSAR) models.[2] Among a plethora of different approaches, pharmacophore-based applications are very popular and have shown to be successful in the capture of new active molecules and lead candidates.[3] Pharmacophore models represent the location of generalized receptor−ligand interaction sites in three-dimensional (3D) space which are considered to form a specific binding pattern. Two basically different approaches for the representation of the pharmacophore can be distinguished: 3D pharmacophore models and pharmacophore fingerprints.[4]

The traditional 3D-approach, implemented in program packages such as Catalyst,[5] DISCO,[6] GASP,[7] or MOE,[8] usually determines the most conserved features of a set of structurally aligned known active ligands. The spatial configuration of generalized interaction sites provides the basis for subsequent virtual screening for new molecules with the same biological activity. Those molecules which comprise all or a user-defined minimum number of the features are presumed to be active. A general drawback of this approach is the necessity to align a molecule to the pharmacophore query before it can be classified as potentially active or inactive.[9] This step can limit the screening of very large databases or virtual libraries. Another drawback can be the lack of information about less conserved regions. This is not easily possible to take into consideration since a minimum number of features have to be satisfied by a molecule to be classified as active.[4] "Excluded volumes" can compensate for a part of this problem by preventing the selection of molecules that are too large for the binding pocket.[4]

Pharmacophore fingerprints describe the spatial arrangement of pharmacophore features typically as a bitstring where each bit corresponds to a certain feature, or in the form of a correlation vector (CV).[10,11] The latter has the advantage that no explicit alignment is needed to estimate the activity of a given molecule, which allows for rapid screening of large compound databases and makes the method resistant to alignment errors. Still, the time-limiting factor is the calculation of 3D conformers. Usually, for each molecule a separate CV is stored. The common application is to take one or more known active query molecules and calculate the similarity of the CVs of each candidate molecule to the CV of the query (or "reference") molecule. The most similar compounds are then presumed to be active.

While there is information retained about all features present in a molecule, a limitation of many pharmacophore methods is that they usually do not contain

* To whom correspondence should be addressed. Phone: +49 69 798 29821. Fax: +49 69 798 29826. E-mail: gisbert.schneider@modlab.de.

10.1021/jm031139y CCC: $27.50
Published on Web 08/11/2004

information about the conservation and tolerance of individual pharmacophore features. First successful applications have been described that go into this direction,[12] but still information about conserved features should be more reliable. A recent attempt to introduce fuzziness into CV-based retrospective screening showed no significant improvement.[13] A shortcoming of this approach might have been that a uniform degree of fuzziness was used for all pairs of features of the molecules.

Another property of pharmacophore fingerprints that can either be considered as benefit or as drawback is that the descriptor of a molecule is not necessarily unique for that molecule. On one hand this can lead us to molecules that are structurally different from the reference molecules and do not exhibit any desired biological activity. On the other hand it can help discover compounds with similar biological activity but significantly different scaffolds, which is often a desired result and is sometimes referred to as "scaffold hopping".[11] Examples of pharmacophore fingerprints are our own CATS pharmacophore descriptors[11,14] and the well-established three- and four-point pharmacophore descriptors.[15,16]

The present study presents a pharmacophore approach combining some advantages of both methods, which we termed SQUID (*S*ophisticated *Qu*antification of *I*nteraction *D*istributions). The aim was to use the information about the conservation and tolerance of features from the alignment of known active ligands and to transform this information into a CV for fast database screening. The information about feature conservation is stored in the form of feature probabilities. A further aim of this work was to explore the effect of different degrees of structural generalization from the original molecular architecture. In SQUID pharmacophore models the distribution of features from the structural alignment is described by a number of Gaussian feature densities with different positions, conservation weights, and tolerances, which are referred to as potential pharmacophore points (PPPs) herein. Thus, the pharmacophore model may contain large numbers of small PPPs or small numbers of large PPPs. The second situation results in a more generalized or "fuzzy" representation of the alignment. Virtual screening with fuzzy pharmacophore representations might allow for the capture of more structural dissimilar molecules with similar biological activity.

Methods

Pharmacophore Model Perception. An overview of the calculation of a SQUID pharmacophore model and the derivation of a CV is given in Figure 1. The starting point of the calculation is a 3D alignment of reference molecules. This alignment is transformed into a three-dimensional distribution of generalized pharmacophoric interaction points (Figure 1a). For SQUID pharmacophore models this field of features is approximated by a set of Gaussian PPPs for a generalized description of the molecules. Dependent on a defined cluster radius, the local density maxima of the pharmacophore feature distribution are determined for the clustering of features into PPPs (Figure 1b). As a result, the distribution of features is approximated by Gaussian PPPs with appropriate tolerances, which are placed into the center of the clustered features. Then the PPPs are weighted according to the conservation of their underlying pharmacophore features among the reference molecules (Figure 1c). In the last step the spatial distribution of PPP densities is transformed into a CV of feature densities describing the distribution of pairs of PPPs within the CV. For this purpose the feature density distributions between all pairs of PPPs is summed up in distance bins of the CV (Figure 1d).

Alignments of reference compounds were obtained either with the flexible alignment tool or with the homology alignment tool of MOE.[8] Flexible alignments were generated with the features "hydrogen-bond donor", "hydrogen-bond acceptor", "aromaticity", "hydrophobe", and "volume" with default weights. Alignments of protein structures were generated with the homology alignment tool in MOE and default parameters. Transformation of the ligands was included in the transformation of protein structures.

For the assignment of SQUID pharmacophore features, six different generalized interaction types were considered: "cationic", "anionic", "polar", "hydrogen-bond donor", "hydrogen-bond acceptor", and "hydrophobic". One of these features was assigned to each of the atoms of the alignment of known active ligands with the ph4_aType function available in MOE.[8] Atoms for which no match was found were not further considered. In this way the 3D alignment of reference molecules was transferred into a three-dimensional distribution of pharmacophoric features, defined by generalized interaction types. This distribution could be considered as a field of the features of the molecules (Figure 1a).

For a more generalized or fuzzy approximation of this field by Gaussian PPPs, the single features had to be clustered into PPPs. Each PPP should represent a local maximum of the distribution of the pharmacophoric features. To define the degree of "fuzziness" of the pharmacophore model, the "cluster radius" was introduced as a variable, which defines the spatial area within which feature maxima were determined. The cluster radius determines the total number of PPPs that are present in a pharmacophore model and thus the degree of structural generalization from an atomistic view. To determine the local maxima of the feature distribution, a cluster radius dependent local feature density LFD was calculated for each position of an atom of the molecular ensemble. For an atom k of the pharmacophore feature-type T, the local feature density was calculated by

$$\mathrm{LFD}(\mathrm{atom_k}^{\mathrm{T}}) = \sum_{i=1}^{n} \max\left\{0,\ 1 - \frac{D_2(\mathrm{atom_k}^{\mathrm{T}}, \mathrm{atom_i}^{\mathrm{T}})}{r_c}\right\} \quad (1)$$

where n is the number of atoms of type T in the molecular ensemble, D_2 is the Euclidean distance between two atoms, and r_c is the cluster radius. Only atoms within r_c around an atom contribute to the LFD of this atom. The shorter the distance of an atom to the atom under consideration the higher is the impact of this atom on the LFD. A local maximum of the distribution of a pharmacophore feature T was defined as the

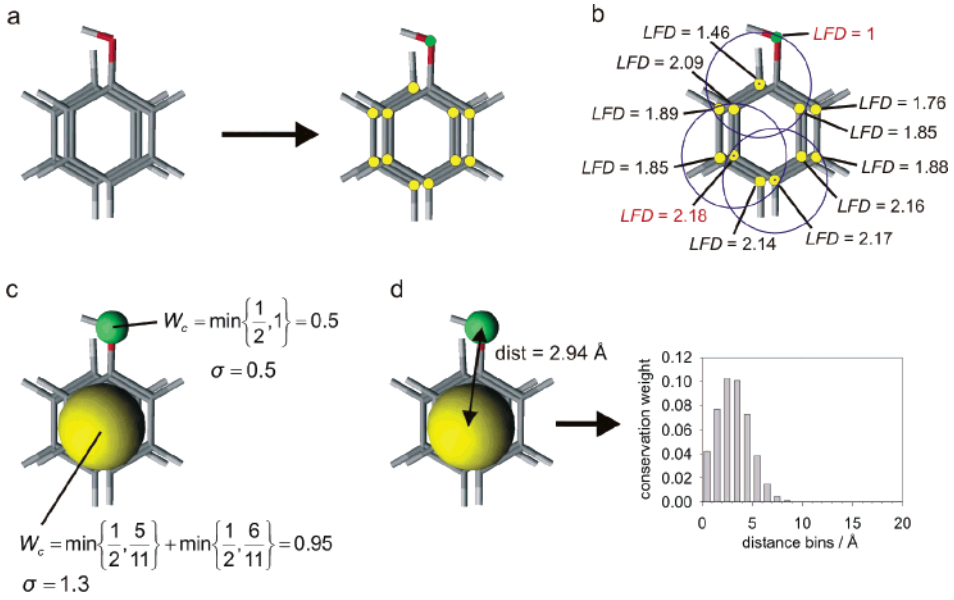


Figure 1. Schematic overview of the calculation of a SQUID pharmacophore model and a corresponding CV descriptor. (a) An alignment of active reference molecules is transformed into a 3D distribution of generalized pharmacophore interaction points. Yellow points represent hydrophobic interacting atoms, green points represent atoms which can interact as hydrogen bond acceptor and donor. (b) The interaction points are clustered to form PPPs based on the feature density distribution of the respective features and based on the defined resolution of the resulting model. The resolution of the model is affected by the cluster radius. Each PPP should represent a local maximum in the local feature density distribution. The local feature density LFD for each interaction point is calculated with eq 1. In this example a cluster radius of 1.5 Å was used. The circles exemplary show the cluster radii around the features which are marked by dots. The two maxima are highlighted in red. As can be seen in Figure 1b, the hydrophobic maximum is the only feature with five neighboring hydrophobic features within its cluster radius. All other hydrophobic features with lower LFD values have fewer hydrophobic features within their cluster radii. After determination of the maxima, all remaining interaction points are assigned to the nearest maximum of their respective feature-type. (c) PPPs are defined at the position of the geometric center of the atoms which are assigned to a local maximum. Two PPPs (yellow = hydrophobic, green = polar = hydrogen bond acceptor and donor) have been calculated in this example. The standard deviations σ of the PPPs are defined by the median distance of the atoms of the corresponding maximum to the position of the PPP, with a minimum value of 0.5 Å like for the polar PPP. The polar PPP has a conservation weight W_c of 0.5 since it consists only of atoms of one of the two molecules. The W_c of 0.96 for the hydrophobic PPP results from the slightly different number of hydrophobic interactions in the two molecules. (d) For virtual screening the pharmacophore model is transformed into a CV with eq 3. This CV represents a vector of probabilities for the presence of pairs of interactions within certain distances. A section of the resulting CV representing polar−hydrophobic pairs of PPPs within a distance range of 0 to 20 Å is shown. The highest probability of finding such a pair is such within a distance range of 2 to 4 Å. With more pairs of PPPs the probability distributions of these pairs of the bins are added. For virtual screening the whole CV is scaled to a maximum of 1.

position of an atom of feature-type T for which there was no other atom of type T present within r_c yielding a higher LFD value (Figure 1b). If there were two atoms with the same LFD value, one of the atoms was selected randomly.

For clustering of the atoms into PPPs, all atoms of the system were then assigned to the nearest maximum of their type. The geometric center of the atoms of a PPP was taken as the position of the PPP. The median distance of all atoms contributing to a PPP to the center of this PPP was taken as the value of the standard deviation σ of the PPP. For this value a minimum of 0.5 Å was defined (Figure 1c). The standard deviation of a PPP was assumed to be roughly proportional to the number of atoms, which are represented by a PPP.

A basic assumption of our approach is that conserved features of a set of similarly acting molecules should have a stronger impact on the biological function than features which are not very well conserved. The premise for this assumption is that the molecules have comparable activities on their receptor. Thus the PPPs should not be weighted by the number of atoms they represent; instead they were weighted by the conservation of the features they represent within the reference molecules. The information about feature conservation was expressed by the conservation weight:

$$W_C(\mathrm{PPP_k}) = \sum_{i=1}^{n} \min\left\{\frac{1}{n}, \frac{\text{no. atoms from molecule}_i \text{ of PPP}_k}{\text{no. atoms of PPP}_k}\right\} \quad (2)$$

where n is the number of molecules in the model. This function returns a maximum value of 1 for PPPs representing the same number of atoms from all molecules of the ensemble, and a minimum value of n^{-1} for PPPs which consist only of atoms of one molecule (Figure 1c). The conservation weight of a PPP of type T can be considered as the probability of finding a feature of type T at the position of the PPP. The resulting SQUID pharmacophore model can thus be considered as the three-dimensional probability distribution of the features found in a 3D molecular alignment.

Database Screening. For the sake of very rapid database screening, we did not explicitly align molecules to a 3D pharmacophore model. Instead the 3D model was transformed into a CV which can be utilized for efficient screening of databases containing CV-encoded

molecules. In the way the three-dimensional pharmacophore model accounts for the probabilities for the presence of features at positions in the explicit 3D space, the CV descriptor should account for the same probability distribution within the CV. For this purpose, the pharmacophore model and the molecules of a screening database were both constructed as a pairwise distance and feature-type dependent autocorrelation vector. The pharmacophore CV contains the probabilities for the presence of pairs of atoms within certain distances (Figure 1d) instead of the probabilities for the presence of single atoms within 3D space in the underlying pharmacophore model.

All entries in the screening database were encoded by the CATS3D descriptor,[14] which is a 3D extension of the topological CV-based CATS descriptor.[11] As described for the calculations of the pharmacophore model (vide supra), each atom of a molecule was assigned to one of the six generalized atom types of the ph4_aType function of MOE. For the 21 possible pairs of these generalized atom types {cationic−cationic, cationic−anionic, cationic-polar, etc., hydrogen-bond acceptor−hydrophobic, hydrophobic−hydrophobic}, the Euclidean distances of all atoms were measured, and the pair frequencies were partitioned into 20 equal distance bins in [0,20] Å, which resulted in a CV of 420 dimensions. The value stored in each bin was scaled by the added incidences of the two respective features. Thus each dimension of the CATS3D CV was calculated according to

$$\mathrm{CV}_d^{\mathrm{TP}} = \frac{1}{A+B}\sum_{i=1}^{A}\sum_{j=1}^{B}\frac{1}{2}\delta_{ij,d}^{\mathrm{TP}} \tag{3}$$

where i and j are atoms, d is a distance range, TP are the pharmacophoric feature-types T of the pair of atoms i and j, A and B are the number of atoms of the feature-types of atom i and atom j, respectively, and δ_d^{TP} (Kronecker delta) evaluates to 1 for all pairs of atoms of types TP 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. In contrast to the original CATS3D version used for similarity searching,[14] the CV was finally scaled to a maximum value of 1.

For comparison of database compounds with the SQUID pharmacophore model, both CV representations must have the same number of dimensions and the same arrangement of distance bins and pairs of generalized atoms. Using the conservation weights and the standard deviations of the PPPs, the Gaussian probability density for the presence of each of the PPP pairs within the distance ranges of each bin of the CV was calculated (eq 4). The resulting SQUID CV contains the normalized sum over the probability densities of all pairs of PPPs:

$$\mathrm{CV}_d^{\mathrm{TP}} = \frac{1}{\mathrm{no.pairs(TP)}}\sum_{i=1}^{A}\sum_{j=1}^{B}\frac{1}{2}\delta_{ij}^{\mathrm{TP}} \times \left(\frac{W_c(i)W_c(j)}{\sqrt{2\pi}(\sigma_i\sigma_j)}\exp\left(-\frac{1}{2}\frac{(D_2(i,j)-\mathrm{center}_d)^2}{(\sigma_i+\sigma_j)^2}\right)\right) \tag{4}$$

where i and j are PPPs, d is a distance range, TP are the feature-types T of the pair of PPPs i and j, A and B are the number of PPPs of the feature-types of i and j, W_c is the conservation weight of a PPP, σ is the standard deviation of a PPP, center_d is the center of the distance range d, and δ^{TP} evaluates to 1 for all pairs of PPPs of types TP. The factor of 0.5 in the sum avoids double counting of pairs. Pairs of PPPs with themselves were not considered. The values of each dimension were scaled by the total number of possible pairs of PPPs of the two features considered. Finally the CV was scaled to a maximum value of 1, which is a premise for the interpretation of the bin-values as probabilities. Equation 4 transforms the two Gaussian probability distributions of two PPPs found in a certain distance into one Gaussian probability distribution for the presence of the pair of PPPs within the distance ranges given by the CV (Figure 1d). The pairwise Gaussian distributions were summed over all possible pairs of PPPs of the system.

In addition to the conservation weights we introduced "feature-type weights" for the mutual weighting of the importance of the generalized feature-types within the CV, e.g. by supervised training using sets of reference molecules. These weights were introduced since we observed that the conservation weights alone were not always adequate to result in successful retrieval of active molecules by retrospective screening. Weight values were specified for each of the pharmacophore feature-types present in the model. The sums of single feature-type weights were used as feature-type weights for each pair of feature-types present in the CV. Subsequently the sum of the probabilities in the CV was scaled individually for each pair of feature-types over all distance bins. It was scaled to the value of the feature-type weight for the respective pair. Finally the CV was again scaled to a maximum of 1.

A score for the similarity between the SQUID model and an individual molecule was calculated according to eq 5.

$$S(a,b) = \frac{\sum_{i=1}^{n}(a_i b_i)}{1+\sum_{i=1}^{n}((1-a_i)b_i)} \tag{5}$$

where a_i is the value of the i-th element of the SQUID descriptor, b_i is the value of the i-th element of a molecule descriptor, and n is the total number of dimensions. The value a_i may be considered as the idealized probability of the presence of atom features in b_i. This results in high scores for molecules with many features in regions of the query descriptor which have a high probability. To penalize the presence of such atom pairs in regions with a low probability, the denominator weights the presence of atom pairs with the inverted probabilities of the descriptor of the pharmacophore model. A value of 1 was added to the denominator to avoid division by zero and high scores resulting from a very low value in the denominator of the term.

All methods for the calculation of the pharmacophore model, the calculation of the CVs of the pharmacophore

model and the calculation of the CATS3D descriptors were implemented in the SVL programming language in MOE.[8] For all calculations in MOE we used version 2003.02 of the software.

Retrospective Screening. The quality of a SQUID model was evaluated by its capability to produce high scoring values for molecules with the desired biological activity and low scoring values for molecules without this activity. This capability was expressed by the enrichment factor ef (eq 6). ef represents the ratio between the percentage of active molecules in a top x% fraction of the sorted database to the percentage of active molecules in the whole database.

$$\mathrm{ef} = \left(\frac{F_{\mathrm{act}}}{F_{\mathrm{all}}}\right) \Big/ \left(\frac{D_{\mathrm{act}}}{D_{\mathrm{all}}}\right) \tag{6}$$

where F_{act} and D_{act} are the numbers of known active molecules in the subset and the whole database, and F_{all} and D_{all} are the total numbers of molecules in the subset and 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 above 1.

In virtual screening, one is usually interested in the enrichment of the first few percentages of a database. In cases where the focus was a more general assessment of the enrichment capabilities of pharmacophore models over the whole set of reference molecules, we used the following enrichment value *ev* for the comparison of the pharmacophore models.

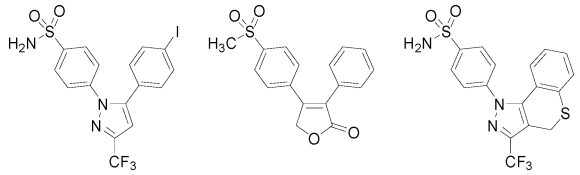
$$\mathrm{ev} = \sum_{i=1}^{100} (101 - i)\mathrm{ef}(i\%) \tag{7}$$

where ef(*i*%) is the enrichment factor for the first *i*% of the ranked database. This returns the weighted sum of the enrichment factors of the whole database. The smaller the fraction of the database, the higher is the weight for the ef.

Data Sets. The COBRA data collection was used for retrospective screening experiments.[17] The database version consisted of 4705 annotated ligands of biological targets compiled from scientific literature. All ligands were present in neutralized form. A database with up to 50 conformations per molecule was calculated with MOE using default parameter settings. Each resulting conformation was energy minimized with the MMFF94 force field with default parameters in MOE. For 73 molecules the algorithm failed to calculate conformations. For these molecules, single 3D conformations were calculated with CORINA.[18] The resulting database consisted of 124 910 conformations. For retrospective screening, we removed those molecules from the database, which were used for pharmacophore model generation. The resulting databases for retrospective screening consisted of 92 active molecules and 4611 inactive molecules for cyclooxygenase 2 (COX-2), and 188 actives and 4517 inactive compounds for thrombin.

For calculation of a COX-2 pharmacophore model, the crystal structures of COX-2 with the specific inhibitor SC-558 (1CX2) and the structures of COX-2 with the unspecific inhibitors flurbiprofen (3PGH) and indomethacin (4COX) were used to model a template alignment for the flexible alignment of the specific COX-2 inhibitors rofecoxib and molecule 5 (M5) from Palomer et al.[19] For calculation of the thrombin pharmacophore model we used the pdb structures 1C4V, 1D4P, 1D6W, 1D9I, 1DWD, 1FPC, and 1TOM.[20]

Scheme 1. 2D Structures of the Known Active COX-2 Inhibitors Used for the Calculation of the SQUID Pharmacophore Model



1 SC-558 **2** Rofecoxib **3** M5

Results and Discussion

The SQUID approach was tested for two classes of biological targets: COX-2 and thrombin. Both targets are well characterized in the literature, and crystal structures of the receptors with bound inhibitors are available. This was important since our method depends on a meaningful alignment of ligands. For both targets, pharmacophore models had already been developed by other groups,[19,20] and diverse sets of ligands are known, which provided a good starting point for retrospective screening. SQUID pharmacophore models were calculated using different PPP cluster radii and tested for performance by retrospective screening.

Pharmacophore Model of COX-2 Ligands. Palomer et al.[19] derived a pharmacophore model for COX-2 inhibitors on the basis of five specific inhibitors SC-558 (**1**), rofecoxib (**2**), DFU, celecoxib, and a molecule which they termed "molecule 5" (M5, **3**). For calculation of a 3D structural alignment of these ligands, they used a template alignment of all COX-2 ligands, for which there was a crystal structure of the ligand−receptor complex available. Crystal structures were at hand for SC-558 (1CX2) and the two unspecific inhibitors flurbiprofen (3PGH) and indomethacin (4COX). The alignment of these molecules was performed by superposition of their protein structures. The remaining ligands were aligned to the template alignment with the program Catalyst.[5] This approach was taken as a reference for the development of a pharmacophore model with our own program SQUID. The molecules DFU and celecoxib were not included in the SQUID pharmacophore model, because they are close analogues of rofecoxib and SC-558. The 2D structures of the remaining molecules are shown in Scheme 1. Crystal structures 1CX2, 3PGH, and 4COX were aligned with the homology alignment tool of MOE.[5] Rofecoxib and M5 were aligned to this template alignment with the flexible alignment tool of MOE. First, rofecoxib was aligned to the fixed template alignment. Then, M5 was aligned to the fixed alignment resulted from the previous step. For the final alignment, the unspecific inhibitors were removed. The resulting alignment of COX-2 inhibitors is shown in Figure 2. In accordance with the model of Palomer et al. the crucial pharmacophore features of these molecules are the sulfonyl group and the two aromatic six-membered rings.[19] The aromatic rings close to the sulfonyl group,

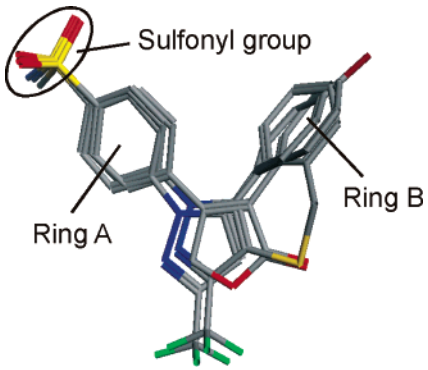


Figure 2. Three-dimensional alignment of the COX-2 inhibitors. Rofecoxib and M5 were aligned to the crystal structure conformation of SC-558 bound to COX-2. According to Palomer et al., essential interactions for specific COX-2 inhibitors are the aromatic rings A and B and the sulfonyl group.[19]

further referred to as "ring A", are nearly parallel to each other in the model. The angles between the planes of the distant aromatic rings, further referred to as "ring B", seem to be less constrained. The least conserved region of the model is the linker region between the two aromatic ring centers.

SQUID pharmacophore models were calculated with cluster radii from 0.5 Å to 3.5 Å in steps of 0.1 Å. A sample set of these pharmacophore models is shown in Figure 3. The models consisted of only three generalized interaction types: hydrogen-bond donors, hydrogen-bond acceptors, and hydrophobic interactions. The model resulting from 1 Å cluster radius is the most detailed one. Here atoms in close proximity are combined to PPPs, which results in a low abstraction from the chemical scaffolds. In contrast to all other models shown, the preferred angle between the two aromatic rings A and B are preserved in this model. The models resulting from 1.5 and 2.0 Å exhibit a higher degree of generalization from molecular structure. Many atoms, especially in the regions of the aromatic rings A and B, were combined to form large PPPs, covering several atoms from each of the molecules. Up to 2.0 Å only hydrophobic atoms were combined. The models from the cluster radii 2.5 Å and 3.0 Å still represent the overall shape of the molecular alignment with three hydrophobic PPPs, but in the 3.5 Å model the shape of the alignment is only marginally visible. In all models with a cluster radius up to 2.0 Å the sulfonyl group is represented by two highly conserved hydrogen-bond acceptor PPPs, one hydrogen-bond donor PPP, and one hydrophobic PPP. In the models resulting from cluster radii greater than 2.0 Å all oxygen atoms of the sulfonyl group are represented by a single large PPP. Moreover, the hydrophobic PPP vanished since the methyl group was assigned to the PPP of ring A.

Retrospective Screening for COX-2 Inhibitors. As the results of retrospective screening were very sensitive to the feature-type weights (data not shown), we decided to perform a restrained exhaustive search for the optimization of these weights. For every calculated model, each of the feature-type weights for features present in the pharmacophore model was varied from 0.1 to 0.5 in steps of 0.1, which resulted in 125 different weighting schemes for the COX-2 pharmacophore models. Each of the resulting descriptors was evaluated by retrospective screening. To obtain statistically more significant results, five different subsets of the COBRA database were created. For each of the subsets 50% of actives and 50% of inactives were randomly chosen from the original database for retrospective screening.

The results of the optimization procedure are shown in Figure 4. For each model calculated with a different cluster radius, the average enrichment factors for the first 1%, 5%, and 10% of the five ranked databases obtained with the best found weighting scheme are shown. The highest average enrichment factor of 39 for the first 1% of the database was obtained with the model calculated with a cluster radius of 1.4 Å and feature-

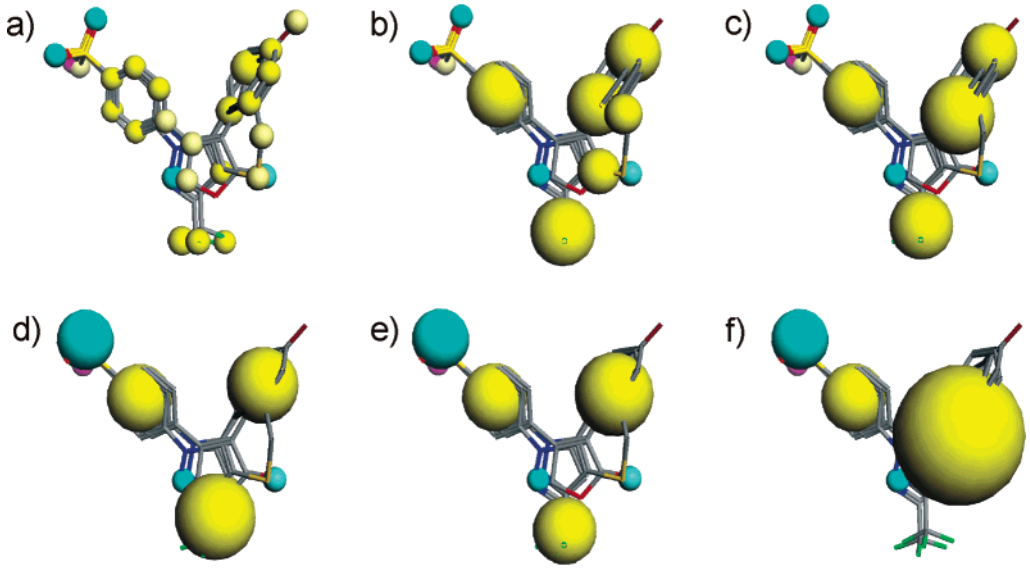


Figure 3. Pharmacophore models calculated with different cluster radii from the aligned COX-2 inhibitors. The cluster radii of the six shown models are (a) 1.0 Å, (b) 1.5 Å, (c) 2.0 Å, (d) 2.5 Å, (e) 3.0 Å, and (f) 3.5 Å. The colored spheres are representing the PPPs of the models. The radii of the spheres denote the standard deviations of the spatial distributions of the atoms of each PPP. Yellow PPPs represent hydrophobic interactions, magenta PPPs represent hydrogen-bond donors, and cyan PPPs represent hydrogen-bond acceptors. The intensity of the color of a PPP denotes the conservation of the PPP among the aligned molecules.

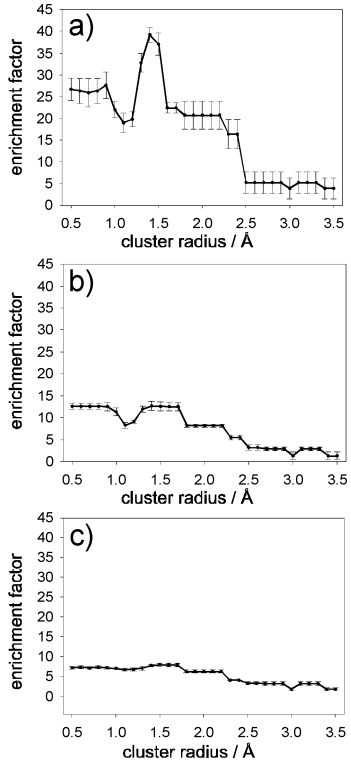


Figure 4. Results of the retrospective screenings for COX-2 inhibitors with the pharmacophore models calculated with cluster radii from 0.5 Å to 3.5 Å with the optimized feature-type weights of each model. The shown enrichment factors are average values from screening of five randomly selected subsets of the COBRA database. The enrichment factors are calculated from the first 1% in part a, for the first 5% in part b and for the first 10% in part c.

type weights of 0.1 for hydrogen-bond donors, 0.4 for hydrogen-bond acceptors, and 0.3 for hydrophobic interactions.

As it could be anticipated, the standard deviations of the enrichment factors were the highest for the first 1% and decreased for the first 5% and 10%. Nevertheless according to their standard deviations the enrichment factors for the first 1% of the database still seem to be appropriate for an evaluation of our pharmacophore models. All three curves exhibit the same general characteristics for different cluster radii, although the differences between the models vanish more and more considering the enrichment of the first 5% and 10% of the database (Figure 4).

Considering the performance of the models for the enrichment in the first 1% of the database, large enrichment factors could be obtained for all models with a cluster radius from 0.5 Å to 2.4 Å. As can be seen in Figure 3, these models only differ in the description of the hydrophobic interactions, while models with 2.5 Å and greater cluster radii differ from the other models in the description of the oxygen atoms of the sulfonyl group. The models with a large cluster radius use a single PPP for the description of these atoms while the models with small cluster radii use two PPPs. It seems that a single PPP for the description of these oxygen atoms is not sufficient for a reasonably performing pharmacophore model. The models from 0.5 Å to 2.4 Å can be divided into four groups. The pharmacophore models of the first group from 0.5 Å to 0.9 Å with enrichment factors of roughly 27 consist only of PPPs merging atoms from different molecules within close spatial proximity, e.g., all aromatic rings are described by six PPPs. From 1.0 Å to 1.2 Å, a minimum in the performance of the models was observed. In these models, ring A is represented by six PPPs, and ring B is represented by four or five PPPs, which might not be an adequate number for the description of an aromatic six-ring.

The three best performing models were obtained with cluster radii of 1.3 Å to 1.5 Å. Both models from 1.4 Å and 1.5 Å describe ring A with a single PPP and ring B with three and two PPPs, respectively. Like within the poorly performing models employing cluster radii from 1.0 Å to 1.2 Å, in the model obtained with a cluster radius of 1.3 Å, ring B is represented by four PPPs, but ring A is represented by three PPPs. The larger tolerances of the three PPPs of ring A might have compensated the unfavorable description of ring B. Within the models from 1.6 Å to 2.4 Å, the hydrophobic interactions are represented by a decreasing number of five to three hydrophobic PPPs.

For comparison, a pharmacophore model was calculated including the two additional COX-2 inhibitors DFU and celecoxib from the model of Palomer et al.[19] A slightly better ef for the first 1% of the database (ef = 40) was obtained with a model calculated with a cluster radius of 1.5 Å and feature-type weights of 0.2 for hydrogen-bond donors, 0.5 for hydrogen-bond acceptors, and 0.5 for hydrophobic interactions (data not shown).

To test if our approach for the optimization of feature-type weights is also valid in situations with significantly fewer reference molecules, we repeated the optimization procedure with only the molecules from the pharmacophore model as reference molecules for assessment of the enrichment capabilities of the SQUID models. For all models with cluster radii from 0.5 Å to 2.4 Å, several weighting schemes were found that ranked two of the three reference molecules into the first percent of the database. In no case were found all three molecules in the first percent. Ranking of all models according to eq 7 resulted in four similarly top-scoring 1.4 Å models with different weighting schemes. Among these models, the previously found best working model was found, with feature-type weights of 0.1 for hydrogen-bond donors, 0.4 for hydrogen-bond acceptors, and 0.3 for hydrophobic interactions. The worst of the other three models still resulted in an ef of 34, screening the database with the 92 COX-2 inhibitors.

To compare our method with another established method, we performed retrospective screenings with the molecules from which the pharmacophore models were calculated. For this approach, we encoded these molecules with the CATS3D descriptor, but without scaling the descriptor to a maximum of 1. The database molecules were scored by the Euclidean distance to the query molecule, and the database was sorted according to the calculated distances to the query molecule. A comparison of the results of the similarity search with the results obtained from the best SQUID model is shown in Figure 5. Rofecoxib performed best in comparison to the other two COX-2 inhibitors. This might be a consequence of its comparably small size. The

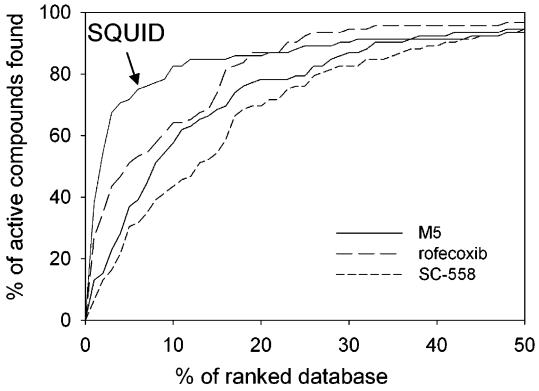
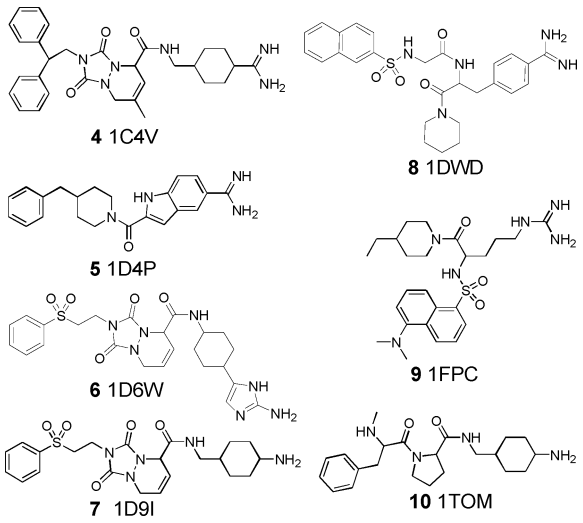


Figure 5. Comparison of the enrichment curves of the best found pharmacophore model resulting from 1.4 Å cluster radius with the results of the retrospective screening with the CATS3D descriptors of the COX-2 inhibitors used for the model calculation.

Scheme 2. 2D Structures of the Known Active Thrombin Inhibitors Used for the Calculation of the SQUID Pharmacophore Model[a]



[a] The names beneath the molecules are the pdb identifiers of the protein structures from which the conformations of these molecules were extracted.

pharmacophore model performed better than rofecoxib for the first 15% of the database. With the SQUID approach 75% of the active COX-2 inhibitors were ranked into the first 6% of the database. In comparison, rofecoxib retrieved 75% of the actives among the top 16% of the ranked database. Interestingly, the performance of the pharmacophore model decreased significantly for the last 25% of the active molecules in comparison to the COX-2 inhibitors.

Pharmacophore Model of Thrombin Ligands. A diverse set of seven noncovalent, nonpeptidic thrombin inhibitors was adopted from Patel et al.[20] The 2D structures of these molecules are shown in Scheme 2. All ligands were aligned by superposition of the protein structures with the homology alignment tool of MOE. The resulting alignment of the thrombin inhibitors is shown in Figure 6. According to Patel and co-workers the major interactions are B, H1, H2, and H3, where B is a basic interaction which interacts with the carboxylic group of Asp189. H1, H2, and H3 are hydrophobic interactions. Less conserved interactions are D1 and A1, where D1 is a hydrogen-bond donor and A1 is a

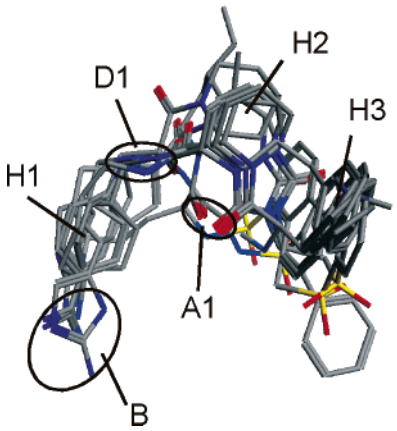


Figure 6. Three-dimensional alignment of the thrombin inhibitors. The molecules were aligned by superposition of their appropriate protein structures. All essential interactions with the receptor, according to Patel et al., are indicated.[20] B is a basic interaction, H1, H2, and H3 are hydrophobic interactions, A1 is a hydrogen-bond acceptor, and D1 is a hydrogen-bond donor.

hydrogen-bond acceptor. SQUID pharmacophore models were calculated from the 3D alignment with cluster radii from 0.5 Å to 3.5 Å within steps of 0.1 Å.

A sample set of the resulting models is shown in Figure 7. Four generalized interaction types were found in the ligands based on the ph4_aType function of MOE: hydrogen-bond acceptor, hydrogen-bond donor, polar, and hydrophobic. Since all ligands were presented in neutralized state, interaction B was not identified as cationic feature; instead it was represented by hydrogen-bond donor and polar interactions and an additional hydrogen-bond acceptor. In the 1.0 Å and 1.5 Å models the description of the three hydrophobic interactions H1, H2, and H3 is very detailed using a large number of PPPs. With a cluster radius of 2.0 Å, only four PPPs are left. In the models with cluster radii of 2.5 Å, 3.0 Å, and 3.5 Å, these hydrophobic interactions are represented by only three PPPs. Both A1 and D1 are structurally conserved features in the alignment. All appropriate atoms from the different molecules lie in near proximity to each other. A1 is represented by a small conserved PPP in all models except for the 3.5 Å model, where it is represented by a large PPP, including other hydrogen-bond acceptors. D1 is also represented by a small conserved PPP except for the models with 3.0 Å and 3.5 Å cluster radius.

Retrospective Screening for Thrombin Inhibitors. For retrospective screening with the SQUID pharmacophore models obtained from the alignment of thrombin inhibitors, the same procedure for feature-type weight optimization was applied as for the screening for COX-2 inhibitors. For the thrombin optimization, 625 weighting schemes had to be evaluated per model.

The results of the optimization procedure are shown in Figure 8. The best average enrichment factor of 18 for the first 1% of the database was obtained with the model calculated with a cluster radius of 2.0 Å and feature-type weights of 0.4 for polar, 0.5 for hydrogen-bond donors, 0.3 for hydrogen-bond acceptors, and 0.5 for hydrophobic interactions.

We detected two peaks for each of the enrichment factors, one for models with a high degree of generalization with cluster radii from 2.0 Å to 2.2 Å, and one for

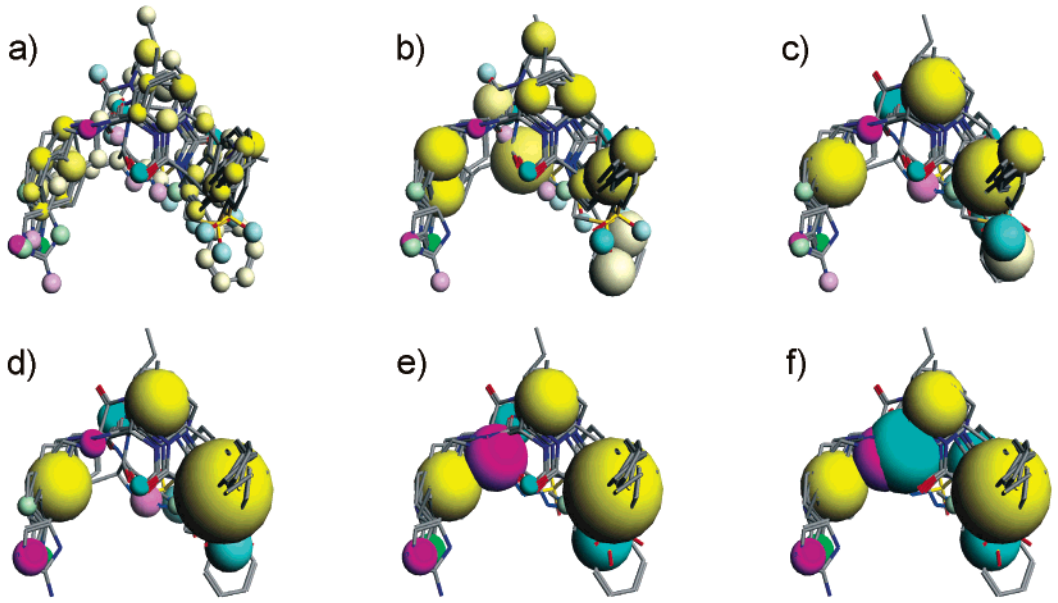


Figure 7. Pharmacophore models calculated with different cluster radii from the aligned thrombin inhibitors. The cluster radii of the six shown models are (a) 1.0 Å, (b) 1.5 Å, (c) 2.0 Å, (d) 2.5 Å, (e) 3.0 Å, and (f) 3.5 Å. Yellow PPPs represent hydrophobic interactions, magenta PPPs represent hydrogen-bond donors, cyan PPPs represent hydrogen-bond acceptors, and green PPPs represent polar interaction.

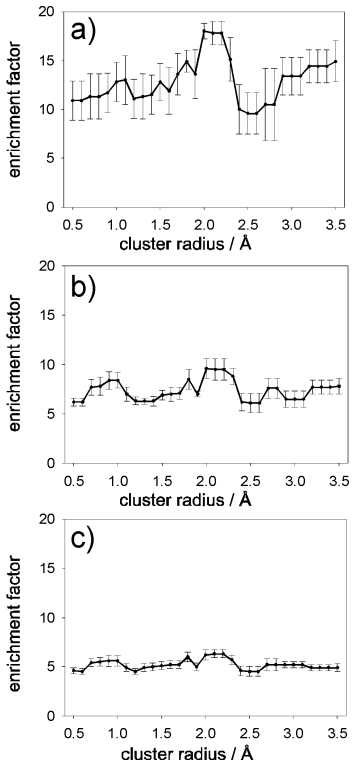


Figure 8. Results of the retrospective screenings for thrombin inhibitors with the pharmacophore models calculated with cluster radii from 0.5 Å to 3.5 Å with the optimized feature-type weights of each model. The shown enrichment factors are average values from screening of five randomly selected subsets of the COBRA database. The enrichment factors are calculated from the first 1% in a, for the first 5% in b and for the first 10% in c.

models with a lower degree of generalization with cluster radii of 1.0 Å and 1.1 Å. Interestingly, models with cluster radii greater than 2.8 Å performed very well, too. As can be seen in Figure 7, the model with a cluster radius of 1.0 Å from the first peak mainly clustered atoms within near proximity into PPPs, while already favoring conserved atoms. The model with a cluster radius of 2.0 Å from the second peak represents the features with a drastically diminished overall number of PPPs. In particular the three hydrophobic interactions are represented by four PPPs, in contrast to all other models with a smaller cluster radius. The models resulting from cluster radii larger than 2.8 Å consist mostly of PPPs with large tolerances, but unlike for COX-2, these PPPs represent the shape of the molecular alignment very well.

Like for COX-2, the optimization procedure was repeated with only the molecules from the pharmacophore model as reference molecules. For many models, weighting schemes were found which ranked two of the seven reference molecules into the first 1% of the database. In no case more molecules were found in the first 1%. Ranking of all models according to eq 7 resulted in the previously found best working 2.0 Å model with feature-type weights of 0.4 for polar interactions, 0.5 for hydrogen-bond donors, 0.4 for hydrogen-bond acceptors, and 0.5 for hydrophobic interactions.

The result of the 2.0 Å SQUID model was compared with results from retrospective screening with CATS3D descriptors calculated from the molecules used for the calculation of the pharmacophore model. Enrichment curves are shown in Figure 9. Major differences were observed in the performance of the individual thrombin inhibitors. The inhibitors from the crystal structures 1FPC and 1DWD performed best. The three inhibitors from structures 1D4P, 1D9I, and 1TOM performed even worse than a random distribution of active molecules within some regions of the ranked database. The SQUID pharmacophore model performed better than the most successful similarity search for the first 40% of the database. Fifty percent of the active molecules were ranked into the first 6% of the database by the phar-

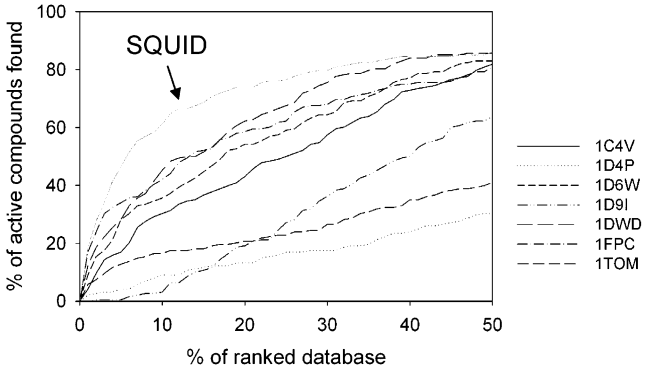


Figure 9. Comparison of the enrichment curves of the best found pharmacophore model resulting from 2.0 Å cluster radius with the results of the retrospective screening with the CATS3D descriptors of the thrombin inhibitors used for the model calculation.

macophore model in comparison to the best thrombin inhibitor from 1DWD, which ranked 50% of the active molecules into the first 13% of the database.

Method Performance. For an additional comparison of the SQUID pharmacophore model with an established method, we calculated pharmacophore models from the two alignments of COX-2 and thrombin reference compounds with the pharmacophore tool of MOE.[8] For both models, we used the atom type scheme PCH_ALL which consists of atom types for cationic, anionic, hydrogen-bond donor, hydrogen-bond acceptor, aromatic ring centers, and hydrophobic interactions. In contrast to SQUID, one PPP in MOE can describe multiple atom types, which can be combined by logic operators. As a starting point for the alignments, pharmacophore models were calculated automatically with the consensus pharmacophore function using MOE default parameters. This function clusters features into PPPs which are more conserved than a threshold value. For the threshold, 50% conservation was used. Retrospective screening with these first pharmacophore models was very slow, and the program failed to screen the whole database. As a consequence, we modified the models manually by removing PPPs which were not among the key features of the pharmacophore models published by Palomer et al.[19] or Patel et al.[20] respectively (Figure 2, Figure 6). For the thrombin model the radii and the positions of the PPPs for H1, H2, and D1 were manually adjusted for a more accurate representation of the underlying ring structures and the cluster of hydrogen-bond donors. Additional multiple features of the PPPs were also removed. The resulting MOE pharmacophore models are shown in Figure 10. Both models were evaluated by retrospective screening of the COBRA database.

With the MOE COX-2 model (Figure 10a), we retrieved 84 matching molecules among which we found 49 (58%) of the known COX-2 inhibitors. In comparison, the COX-2 SQUID model found 47 (56%) active molecules in the first 84 compounds from the ranked database. Reinsertion of a PPP from the first MOE model, which represents the central five-ring of the COX-2 inhibitors by an acceptor, aromatic, or hydrophobic interaction, resulted in 48 actives (91%) out of 53 matches. Within the first 53 molecules of the ranked database, the SQUID pharmacophore model retrieved

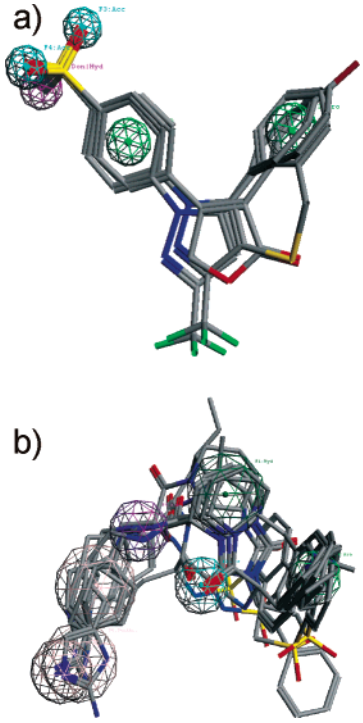


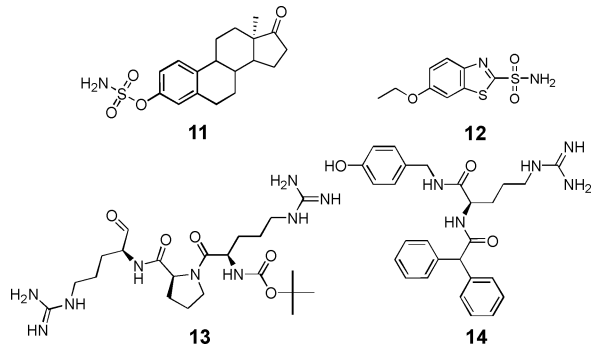
Figure 10. MOE Pharmacophore models for COX-2 (a) and thrombin (b). In the COX-2 pharmacophore model the rings A and B are represented by two aromatic ring center PPPs, and the sulfonyl group is represented by two acceptor PPPs and one PPP for a donor or hydrophobic interaction. In the thrombin pharmacophore model the hydrophobic interactions H1 and H2 are represented by hydrophobic PPPs while H3 is represented by an aromatic PPP. For A1 a hydrogen-bond acceptor PPP and for D1 a hydrogen-bond donor PPP was found. The basic interaction B was represented by two PPPs, one for hydrogen-bond acceptor or hydrogen-bond donor, and one for hydrogen-bond acceptor and hydrogen-bond donor.

only 38 (72%) active compounds. A comparison of the actives found by MOE and SQUID showed that the overlap was only 25 molecules, i.e., that both methods complement each other very well. SQUID retrieved an additional 13 actives which were missed by the refined MOE model.

With the MOE thrombin model (Figure 10b), we retrieved five actives (31%) among 16 matches, in comparison to the SQUID model which retrieved 13 actives (81%) among the first 16 molecules of the sorted database. Retrospective screening with the partial match option of the MOE pharmacophore search function requiring only six of the seven PPPs as matching criterion resulted in 489 matches including 87 (18%) thrombin inhibitors. With SQUID 119 actives (24%) were found among the first 489 molecules of the ranked database. The two sets of actives have 64 molecules in common. Again, we conclude that the two pharmacophore searching approaches complement each other.

To gain further confidence in our approach, we took a look at the two top-scoring nonactive molecules from each of the best pharmacophore models for COX-2 and thrombin (Scheme 3). Molecules **11**[21] and **12**[22] were found with the COX-2 pharmacophore model with 1.4 Å cluster radius, and molecules **13**[23] and **14**[24] were found with the thrombin pharmacophore model with 2.0 Å cluster radius. Ethoxzolamide (**12**) is an inhibitor of carbonic anhydrase. Also, it has been shown recently

Scheme 3. 2D Structures of the Best Scored Nonactive Molecules Found with the Best Pharmacophore Models[a]



[a] Compounds **11** and **12** were found with the COX-2 pharmacophore model with 1.4 Å cluster radius. Compounds **13** and **14** were found with the thrombin model derived from 2.0 Å cluster radius.

that celecoxib is a nanomolar inhibitor of carbonic anhydrase.[25] EMATE (**11**) is an inhibitor of estrone sulfatase, and a nanomolar inhibitory effect of EMATE on carbonic anhydrase activity has been reported.[26] This indicates that both "nonactive" molecules share common features with the COX-2 inhibitors from the pharmacophore model. According to our knowledge, no COX-2 activity has been reported for the molecules **11** and **12**, and we think it would be worthwhile testing them.

Molecule **13** (BOC-D-Arg-Pro-Arg) is an inhibitor of Factor Xa for which nanomolar inhibition of thrombin has been reported.[23] It thus represents a real hit. BIBP3226 (**14**) is an antagonist of the neuropeptide Y_1 receptor. To our knowledge, thrombin activity has not been tested for this molecule. Nevertheless obvious similarities to the thrombin inhibitors are present in BIBP3226.

Summarizing, we demonstrated that the SQUID approach can complement existing pharmacophore methods by retrieving additional active compounds. The SQUID technique contains many parameters that must be tuned appropriately, which certainly is a disadvantage of the method. In future implementations we intend to address this issue. In particular, the use of "feature-type weights" in addition to "conservation weights" adds complexity to the models which we intend to eliminate. Still, it should be kept in mind that the feature-type weights allow for an optimization of a pharmacophore model in situations where many reference molecules are at hand.

Conclusions

We developed a fuzzy pharmacophore model approach for the compilation of focused screening libraries ("SQUID"). The method is characterized by the following points:

- Information from molecular alignments of known active compounds is included in the pharmacophore models.
- Features of the aligned molecules are represented by a set of spheres of Gaussian densities, weighted by the conservation of the respective feature at the location of the sphere.
- The degree of fuzziness of the pharmacophore model can be varied, thereby affecting the number and size of PPP spheres.
- Transformation of the pharmacophore model into a correlation vector of probabilities allows for a rapid database screening that does not depend on aligning database molecules to the pharmacophore model. In this way we obtained a quantitative approach for the ranking of compound databases with respect to putative biological activity.

We challenged our approach using inhibitors of COX-2 and thrombin. For both classes of molecules, pharmacophore models were calculated with cluster radii from 0.5 Å to 3.5 Å and evaluated by retrospective screening. Although optimization of the relative weights of the feature-types in the correlation vectors was essential in some cases to obtain a satisfying enrichment, we showed that optimization on the basis of some few reference molecules can be sufficient. The best retrospective screening results for COX-2 were obtained with the model resulting from a cluster radius of 1.4 Å, yielding an enrichment factor of 39 for the first 1% of the ranked database. For thrombin, the best results for the enrichment in the first 1% of the database were obtained with the model resulting from a cluster radius of 2.0 Å, yielding ef = 18. For both targets, the best models outperformed retrospective screening by CATS3D similarity searching. This showed that, independent from the overall enrichment and thus independent of the explicit selection of active molecules, the pharmacophore model outperformed conventional similarity searching. In comparison to conventional pharmacophore searching with MOE, SQUID identified additional actives. We demonstrated that the SQUID pharmacophore model approach provides a potentially useful new method for virtual screening. The inherent fuzzy description of the molecules should support the goal of 'scaffold hopping', especially with higher degrees of fuzziness.

Acknowledgment. We thank the modlab-team for help and stimulating discussions, especially Jürgen Paetz for carefully reading the manuscript and Norbert Dichter for maintenance of the computer cluster. This work was supported by the Beilstein-Institut zur Förderung der Chemischen Wissenschaften.

References

(1) Böhm, H.-J., Schneider, G., Eds. *Virtual Screening for Bioactive Molecules*; Wiley-VCH: Weinheim, 2000.
(2) Xue, L.; Bajorath, J. Molecular descriptors in chemoinformatics, computational combinatorial chemistry, and virtual screening, *Comb. Chem. High Throughput Screen.* **2000**, *3*, 363−372.
(3) Guner, O., Ed. *Pharmacophore Perception, Development and Use in Drug Design*; International University Line: La Jolla CA, 2000.
(4) Pickett, S. The biophore concept. In *Protein−Ligand Interactions*; Böhm, H.-J., Schneider, G., Eds.; Wiley-VCH: Weinheim, New York, 2003; pp 73−105.
(5) Greene, J.; Kahn, S.; Savoj, H.; Sprague, P.; Teig, S. Chemical function queries for 3D database search. *J. Chem. Inf. Comput. Sci.* **1994**, *34*, 1297−1308.
(6) Martin, Y. C.; Bures, M. G.; Danaher, E. A.; DeLazzer, J.; Lico, I.; Pavlik, P. A. A fast approach to pharmacophore mapping and its application to dopaminergic and benzodiazepine agonists. *J. Comput:-Aided Mol. Des.* **1993**, *7*, 83−102.
(7) Jones, G.; Willett, P.; Glen, R. C. A genetic algorithm for flexible molecular overlay and pharmacophore elucidation. *J. Comput:-Aided Mol. Des.* **1995**, *9*, 532−549.
(8) MOE, Molecular Operating Environment. Distributor: Chemical Computing Group, 1010 Sherbrooke St. West, #910, Montreal, Canada H3A.
(9) Vedani, A.; Dobler, M. Multidimensional QSAR in drug research. Predicting binding affinities, toxicity and pharmacokinetic parameters. *Prog. Drug Res.* **2000**, *55*, 105−135.

(10) Mason, J. S.; Cheney, D. L. Library design and virtual screening using multiple 4-point pharmacophore fingerprints. *Pac. Symp. Biocomput.* **2000**, 576−587.
(11) 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
(12) Xue, L.; Stahura, F. L.; Godden, J. W.; Bajorath, J. Fingerprint scaling increases the probability of identifying molecules with similar activity in virtual screening calculations. *J. Chem. Inf. Comput. Sci.* **2001**, *41*, 746−753.
(13) Horvath, D.; Mao, B. Neighbourhood behaviour. Fuzzy molecular descriptors and their influence on the relationship between structural similarity and property similarity. *QSAR Comb. Sci.* **2003**, *22*, 498−509.
(14) Fechner, U.; Franke, L.; Renner, S.; Schneider, P.; Schneider, P.; Schneider, G. Comparison of correlation vector methods for ligand-based similarity searching. *J. Comput.-Aided Mol. Des.* **2003**, *17*, 687−698.
(15) Good, A. C.; Kuntz, I. D. Investigating the extension of pairwise distance pharmacophore measures to triplet-based descriptors. *J. Comput.-Aided Mol. Des.* **1995**, *9*, 373−379.
(16) Mason, J. S.; Morize, I.; Menard, P. R.; Cheney, D. L.; Hulme, C.; Labaudiniere, R. F. New 4-point pharmacophore method for molecular similarity and diversity applications: Overview of the method and applications, including a novel approach to the design of combinatorial libraries containing privileged substructures. *J. Med. Chem.* **1999**, *42*, 3251−3264.
(17) Schneider, P.; Schneider, G. Collection of bioactive reference compounds for focused library design. *QSAR Comb. Sci.* **2003**, *22*, 713−718.
(18) CORINA. Distributor: Molecular Networks GmbH, Computerchemie, Nägelsbachstrasse 25, D-91052 Erlangen, Germany, http://www.mol-net.de/.
(19) Palomer, A.; Cabre, F.; Pascual, J.; Campos, J.; Trujillo, M.; Entrena, A.; Gallo, M.; Garcia, L.; Mauleon, D.; Espinosa, A. Identification of novel cyclooxygenase-2 selective inhibitors using pharmacophore models. *J. Med. Chem.* **2002**, *45*, 1402−1411.
(20) Patel, Y.; Gillet, V. J.; Bravi, G.; Leach, A. R. A Comparison of the pharmacophore identification programs: Catalyst, DISCO and GASP. *J. Comput.-Aided Mol. Des.* **2002**, *16*, 653−681.
(21) Woo, L. W.; Howarth, N. M.; Purohit, A.; Hejaz, H. A.; Reed, M. J.; Potter, B. V. Steroidal and nonsteroidal sulfamates as potent inhibitors of steroid sulfatase. *J. Med. Chem.* **1998**, *41*, 1068−1083.
(22) Supuran, C. T.; Scozzafava, A.; Casini, A. Carbonic anhydrase inhibitors. *Med. Res. Rev.* **2003**, *23*, 146−189.
(23) Marlowe, C. K.; Sinha, U.; Gunn, A. C.; Scarborough, R. M. Design, synthesis and structure−activity relationship of a series of arginine aldehyde factor Xa inhibitors. Part 1: structures based on the (D)-Arg-Gly-Arg tripeptide sequence. *Bioorg. Med. Chem. Lett.* **2000**, *10*, 13−16.
(24) Rudolf, K.; Eberlein, W.; Engel, W.; Wieland, H. A.; Willim, K. D.; Entzeroth, M.; Wienen, W.; Beck-Sickinger, A. G.; Doods, H. N. The first highly potent and selective non-peptide neuropeptide Y Y1 receptor antagonist: BIBP3226. *Eur. J. Pharmacol.* **1994**, *271*, R11−R13.
(25) Weber, A.; Casini, A.; Heine, A.; Kuhn, D.; Supuran, C. T.; Scozzafava, A.; Klebe, G. Unexpected nanomolar inhibition of carbonic anhydrase by COX-2-selective celecoxib: new pharmacological opportunities due to related binding site recognition. *J. Med. Chem.* **2004**, *47*, 550−557.
(26) Ho, Y. T.; Purohit, A.; Vicker, N.; Newman, S. P.; Robinson, J. J.; Leese, M. P.; Ganeshapillai, D.; Woo, L. W.; Potter, B. V.; Reed, M. J. Inhibition of carbonic anhydrase II by steroidal and nonsteroidal sulphamates. *Biochem. Biophys. Res. Commun.* **2003**, *305*, 909−914.

JM031139Y