Use of 3D QSAR Models for Database Screening: A Feasibility Study

Alexander Hillebrecht and Gerhard Klebe*

Institut für Pharmazeutische Chemie, Philipps-Universität Marburg, Marbacher Weg 6, 35032 Marburg, Germany

Received August 8, 2007

The applicability and scope of 3D QSAR methods (CoMFA, CoMSIA) to screen databases are examined. A protocol requiring minimal user intervention has been established to align training and test set molecules using FlexS. As model system isozymes of human carbonic anhydrase (hCA) are used, all results are exemplified studying affinity toward hCA II and selectivity between hCA I and II. The predictive power of the obtained models is assessed through prediction of 663 compounds not included in the training set and compared to 2D QSAR models derived from fragment (MACCS) or property (VSA) based descriptors. The predictive power is evaluated with respect to the following criteria: the numerical, concerning the absolute accuracy of prediction, and the categorical, characterizing the ability to assign a compound to the correct activity class.

INTRODUCTION

Quantitative Structure Activity Relationships (QSAR) describe the dependence of the biological activity or any other property under consideration (then sometimes termed QSPR) of a set of molecules with respect to their chemical structures. The various approaches differ in the type of descriptors employed to encode the structures (e.g., an overall property, topological indices, molecular field descriptors) and in the statistical learning method used to extract a correlation (e.g., linear regression, PLS, decision trees, artificial neural networks, support vector machines). Since the introduction of Hansch[1] and Free-Wilson[2] analysis in 1964, QSAR techniques have become widely used and established tools in the lead optimization process.[3]

QSAR methods require a training set of molecules for which the target property is known in advance to generate a model (except, e.g., kNN-QSAR[4] which is purely instance-based thus requiring no model deduction). Such a model can subsequently be used to predict the numeric (i.e., pK_i, IC_{50}) or categorical (i.e., active/inactive, high/low pK_i) activity of a novel molecule not contained in the training set. Furthermore, several of the statistical methods available provide information about the relevance of the applied descriptors for the activity/property being studied. For example, regression techniques return weights for each independent variable thus indicating through sign and magnitude how strongly and in which direction the respective descriptor modulates the dependent variable. This knowledge provides insights into the mechanism of action of the studied molecules and supports the synthesis planning of prospective modifications in order to improve the desired properties (direct design). However, all QSAR approaches—even those which do not produce easily interpretable output (like ANNs)—can be used to predict the activity of molecules (indirect design). Thus, QSAR models can be used for lead optimization[5] or database screening.[6−8]

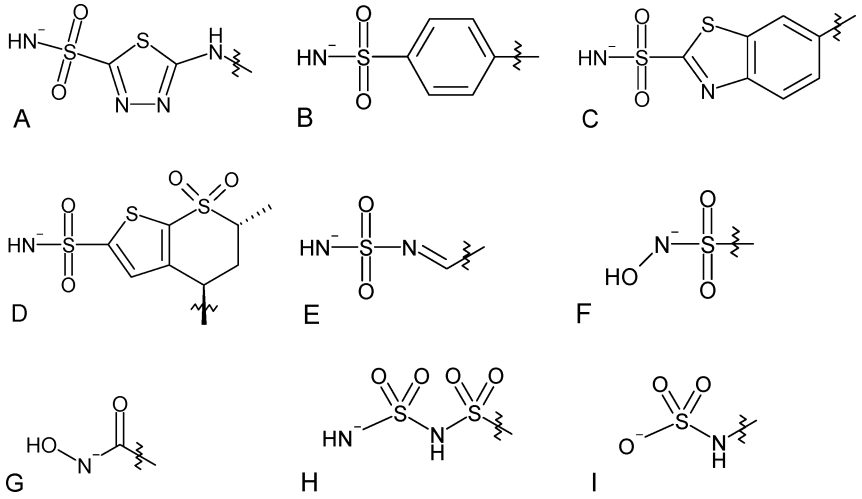
Despite such potential only a limited number of studies has been communicated that actually use 3D QSAR models for database screening. Moro et al. employed autocorrelation vectors of the molecular electrostatic potential (MEP) in conjunction with PLS analysis to screen a small-sized focused virtual combinatorial library containing 841 entries for potent antagonists of the human A_3 receptor.[9] The autocorrelation approach produces descriptors which encode the three-dimensional distribution of molecular properties, but their comparative evaluation does not depend on an alignment. Similarly, Pastor et al. use a modified auto- and cross-correlation transform to derive alignment-independent 3D descriptors from the well-known GRID molecular interaction fields. This GRIND[10] (GRid-INdependent Descriptors) approach has been used for virtual screening[11] and 3D QSAR analyses.[12] Murcia and Ortiz[13] described the development of a fully automatic virtual screening workflow consisting of docking and a subsequent COMBINE analysis. They established the models and simulated virtual screening experiments using factor Xa inhibitors. COMparative BINding Energy (COMBINE) analysis[14] is a receptor-based 3D QSAR approach using protein−ligand interaction energy terms as independent variables thus requiring knowledge or reasonable assumptions about the binding geometries. Zhang et al. performed a virtual screen for HMG-CoA reductase inhibitors using a combination of pharmacophore filtering, docking, and CoMFA predictions. Therein, the docking procedure provided the molecular alignment rule.[15]

In the present study we assess systematically the suitability of 3D QSAR models (CoMFA,[16,17] CoMSIA[18,19]) for large scale applications. 3D QSAR techniques require an appropriate spatial superimposition of the molecules under consideration prior to calculation of field descriptors. Various procedures have been pursued for this crucial step[20,21] many of them being entirely manual (rmsd fit, minimization in the binding pocket of the target protein) or at least requiring user intervention (selection of docking poses). Besides the subjective nature of a user-supervised alignment, the enor-

* Corresponding author e-mail: Klebe@staff.uni-marburg.de.

10.1021/ci7002945 CCC: $40.75
Published on Web 01/23/2008

Chart 1. Formulas of the Nine Scaffolds Contained in Training Set and Test Set Ligands.



mous time effort makes such an approach inappropriate for large screening scenarios. Thus, we decided to develop a protocol which employs the property- and interaction-based FlexS[22] program as alignment engine. FlexS is fast enough for screening purposes, easily extensible to combinatorial libraries, and uses a physicochemical description of the molecules in terms of Gaussian functions which is consistent with the methodology applied in CoMSIA. The statistical performance of the models obtained by this fully automated alignment procedure is compared to the results of a study previously conducted in our laboratory.[23] There, the same data set was used, but the molecules were manually aligned and minimized in the binding pocket.

In addition to 3D QSAR we also derived 2D models and applied both to an external test set of 663 molecules. Two kinds of 2D descriptors are used: The public version of the well-known MDL MACCS keys,[24,25] which encode the presence or absence of distinct structural fragments in a molecule, and the 32 partitioned van der Waals surface area (VSA) descriptors developed by Labute,[26] which are property based. The performance in terms of external predictivity is compared on a numerical and categorical basis: predictive r^2 and Spearman's r^2 assess the correlation between experimental and calculated activity values, Spearman's rank correlation coefficient characterizing the ability to correctly predict the rank *order* of the compounds according to their activity. While these parameters characterize the overall numerical accuracy of the models, the ability to enrich compounds with a desired activity is even more important for database screening purposes. In contrast to other virtual screening approaches such as docking/scoring or pharmacophore matching, however, one limitation of QSAR is that it works only on a rather limited piece of chemical space covered by the training set compounds. Hence, only focused libraries can be screened reliably, and in order to obtain enrichment or classification ratings a certain activity threshold has to be predefined for class assignment (e.g., if $pK_i > 7$, compound is classified "active"). In this study, we report sensitivity and specificity, classical enrichment plots, receiver operating characteristic (ROC) curves, the area under the ROC curve (AUC), and hit rates to compare different models from an objective point of view.[27,28]

All studies reported herein are based on pK_i values of sulfonamide-type inhibitors toward human carbonic anhydrase (hCA) isozymes. CAs are zinc containing hydrolases (EC 4.2.1.1), which catalyze the reversible hydration of carbon dioxide to bicarbonate under the release of one proton.[29,30] They are involved in a variety of important physiological processes, such as pH and CO_2 regulation, bone resorption, calcification, metabolic reactions, tumorigenicity, and electrolyte secretion.[31,32] Therefore, inhibitors of carbonic anhydrases offer the opportunity to treat several physiological disorders, e.g., as drugs against glaucoma, mountain sickness, and epilepsy.[33] Due to the high structural similarity and the broad tissue distribution of the different isoforms of CAs selectivity is an issue of major concern in the design of CA inhibitors. In the present study we used the pK_i value measured against hCA II (in the following sections referred to as pK_i(II)) as an example for affinity prediction and the difference of the pK_i values toward hCA I and hCA II (in the following sections referred to as ΔpK_i(I−II)) as an example for selectivity prediction.

METHODS

All 3D QSAR studies and all PLS analyses were performed using SYBYL7.1.[34] MACCS keys[24] and VSA descriptors[26] were calculated using MOE 2005.06.[35] VSA descriptor selection was accomplished by the SVL script (Scientific Vector Language) AutoQSAR.[36] Automation of ligand alignment was realized via the Python interface "pyflexs" running FlexS 1.20.2.[22] Compound prediction was implemented in SPL (Sybyl Programming Language).

Data Set and Preparation. In order to obtain a suitable training set 144 ligands were taken from our previous 3D QSAR study and subjected to the FlexS alignment procedure. For details about selection criteria and composition of this data set we refer to ref 16. The following procedure was pursued to automate ligand alignment: For each of the nine scaffolds (Chart 1, throughout this paper we will refer to these scaffolds by the corresponding capital letters **A** to **I**) comprised in the training and the test sets two items had to be defined: (1) a so-called MAPREF which is an anchoring fragment used to initiate the incremental construction algorithm of FlexS and (2) a reference ligand for spatial alignment. The assignment of atoms in each ligand to the corresponding MAPREF substructure is achieved by graph matching and does not require any user intervention. The

second item to be predefined is one complete reference ligand onto which the respective candidate molecules are superimposed. This reference should be at best a compound with maximum spatial extensions adopting a "representative" conformation for the individual chemical class, if possible extracted from a crystal structure. The utilization of a MAPREF substructure as alignment template ensures internal consistence of the produced alignment. Without such constraints placements lacking proper superimposition of the anchoring fragments with that of the reference ligand might achieve higher similarity scores because the overall similarity comprising various properties of the entire molecule of a distinct pose can be higher than the score obtained for the desired pose where common substructures are mutually overlaid. Using this procedure all except six molecules could be aligned ending up with a final training set size of 138 molecules.

The training set protocol was applied to 663 compounds serving as a real life test example. It has to be noted that this test set represents a rather unbalanced sample since the only selection criteria were availability of experimental data to be predicted and affiliation to one of the crude chemotypes **A** to **I** considered in the training set. Most of the compounds could be aligned applying default settings in FlexS, only for some rather rigid and large molecules the threshold for the minimum van der Waals overlap volume had to be reduced from 0.6 to 0.4 in order to obtain a superimposition solution.

3D QSAR Analyses. For CoMFA[16] analysis the interaction energies between a probe atom and the ligand atoms were calculated using a grid box of 26 × 34 × 25 points with 1 Å spacing, embedding all ligands with a margin of at least 4 Å in each direction. The same box dimensions were also used for CoMSIA[18] studies. A positively charged sp^3-carbon atom was used as a probe atom for calculating steric and electrostatic CoMFA fields applying SYBYL standard parameters (TRIPOS standard field, dielectric constant 1/r, cutoff 30 kcal/mol). CoMSIA fields were computed for steric, electrostatic, hydrophobic, and hydrogen-bonding properties, using a probe of charge +1, radius of 1, hydrophobicity and hydrogen-bonding properties of +1, and an attenuation factor α of 0.3 for the Gaussian distance-dependent function. All fields were scaled with the CoMFA_STD scaling procedure, assigning equal weights to each field. The response variables (pK_i(II), ΔpK_i(I−II)) were correlated with the field descriptors using SAMPLS[37] in a leave-one-out cross-validation analysis. The optimal number of PLS components was determined by subsequently extracting one more latent variable until the corresponding q^2 value is not further increased by more than five percent.[38] Afterward, a PLS analysis[39] was performed without cross-validation using the optimal number of components, applying no column filtering.

1D and 2D QSAR Analyses. The MDL MACCS keys have already proved useful for screening purposes in conjunction with PLS Discriminant Analysis (PLS-DA) models.[40] We decided to use them for building numeric 2D QSAR models. The public version was used that evaluates the presence of 166 distinct molecular fragments. Instead of computing binary fingerprints which store the information about the presence or the absence of one particular substructure the "counted" version was used resulting in 166 integers each capturing the exact number of occurrences of the fragment in a molecule. This protocol allows better differentiating molecules possessing not only qualitatively the same substructures but differing in counts. This is often the case in focused libraries.

In addition to such fragment oriented models also the property based VSA descriptors were used to derive 1D QSAR models. Each of the 32 VSA descriptors is calculated as the sum of all atomic contributions to the approximated van der Waals surface areas with the respective property falling into a certain range. In MOE this type of descriptor can be calculated for three properties: SlogP (calculated octanol/water partition coefficient), SMR (calculated molar refractivity), and PEOE (Gasteiger−Marsili partial charges). These descriptors show a rather low degree of correlation among each other, capture different aspects of protein−ligand interactions and transport phenomena, and their broad applicability to many QSAR related problems has been shown. In order to select a suitable subset of descriptors the AutoQSAR[36] procedure was applied. It identifies the "best" model based on leave-one-out cross-validation. This procedure should only generate rather crude 1D QSAR models. It will not evaluate the capacity of VSA descriptors exhaustively, and the obtained models should rather serve as a first overview for comparison to get an idea as to what can be achieved by a crude "quick and dirty" 1D model.

Evaluation of Predictivity. In order to assess and compare the predictive power of the different models, several statistical parameters as well as plots are reported: The predictive r^2 value is usually used to characterize the performance in terms of external predictivity of QSAR models. Additionally, Spearman's rank correlation coefficient was calculated which quantifies the ability of a model to correctly predict the relative order instead of the absolute numeric value of the modeled variable.

Since our study is intended to assess the predictive power of QSAR models for database screening purposes it is reasonable to apply also figures-of-merit commonly used in virtual screening, where the correct prediction of class memberships is even more important. Therefore, a somewhat arbitrary threshold had to be defined which determines class affiliation: For pK_i(II) the 2% or 5% of molecules with the highest (or lowest, respectively) activity were considered as the "high activity" (or "low activity") class, whereas the remaining part of the molecules is assigned to the complementary class. This setting simulates screening experiments where one wants to enrich compounds which possess a remarkably high affinity toward the target of interest, compared to the bulk of the training set. Defining the 2% or 5% of lowest activity as the class of interest, the situation represents antitarget modeling, where a distinct receptor must not be inhibited (this holds particularly for, e.g., cytochrome P450 or hERG channel blockers). The same aspects are examined for the activity difference ΔpK_i(I −II). This corresponds to a screening scenario with the aim of enriching compounds selective toward hCA I or hCA II, respectively. As direct measures for correct class prediction the sensitivity Se (also called "true positives rate" or "recall"), the specificity Sp, and the hit rate H (or "precision") are reported[27,28]

$$\mathrm{Se} = \frac{\mathrm{TP}}{\mathrm{TP} + \mathrm{FN}} \times 100\%$$

$$\mathrm{Sp} = \frac{\mathrm{TN}}{\mathrm{TN} + \mathrm{FP}} \times 100\%$$

$$\mathrm{H} = \frac{\mathrm{TP}}{\mathrm{TP} + \mathrm{FP}} \times 100\%$$

where FP and FN are the number of false positives and false negatives, and TP and TN are the number of true positives and negatives, respectively.

To give an even more comprehensive illustration of the results several kinds of plots are reported: (1) A plot displaying the predicted activity/selectivity value of a molecule on the ordinate versus its experimental one on the abscissa. (2) A classical enrichment plot. On the *x*-axis this graph displays the amount of database entries screened, and on the *y*-axis it shows the amount of actives retrieved from the database. (3) A receiver operating characteristic (ROC) curve. This type of plot is quite popular in many scientific areas like psychology, medicine, acoustics, or criminology to assess the ability of a diagnostic system to distinguish signal from noise. On the *x*-axis a ROC plot displays the term 1 − specificity which corresponds to the "noise" in the data set, and on the *y*-axis it shows the sensitivity which can be thought of as the "signal" that is to be identified by the ranking procedure. Nevertheless, its application in drug design and especially in virtual screening is still not standard although it exhibits some advantages compared to the usually applied enrichment curves.[28,41] Most important the area under the curve (AUC) of a ROC plot can be used to directly compare the achieved accuracy of a computer test. Furthermore, the shape of the "ideal curve" of an enrichment plot depends on the ratio of actives to inactives in the database, which is not the case for the ROC curves. Finally, enrichment plots only capture one aspect of a screening experiment, namely the power to retrieve actives, i.e., the sensitivity, whereas ROC curves also illustrate the second important aspect, the ability to discard inactives, i.e., the specificity.

RESULTS AND DISCUSSION

Comparison of 3D and 2D QSAR Models in Terms of Internal Predictivity. To get an impression about the internal consistency, Figure 1 displays the leave-one-out q^2 value as a crude measure of model performance. For all methods q^2 is smaller for selectivity prediction ($\Delta pK_i(I-II)$) than for affinity prediction ($pK_i(II)$) which is a consequence of the fact that $\Delta pK_i(I-II)$ is a difference of two single variables and thus contains the sum of both errors. In the case of CoMFA and CoMSIA the automated alignment yields slightly smaller q^2 values (the largest difference between manual and automated alignment: 0.055 for CoMFA, $pK_i(II)$). The observation that the differences with respect to the alignment method are smaller for CoMSIA can be attributed to the smoother Gaussian functional form used to derive the descriptors which makes this method less sensitive toward slight shifts of the molecules. Since the overall performance of the 3D QSAR methods is virtually the same, it can be stated that the automated alignment procedure is suitable for large scale applications.

The 1D and 2D QSAR approaches produce models of similar quality for prediction of $pK_i(II)$; however, in the case of selectivity prediction the decrease in q^2 is significantly higher compared to the 3D methods. This decrease is more pronounced for the VSA descriptors compared to the MACCS keys. Nevertheless, still highly significant models ($q^2 > 0.5$) can be obtained.

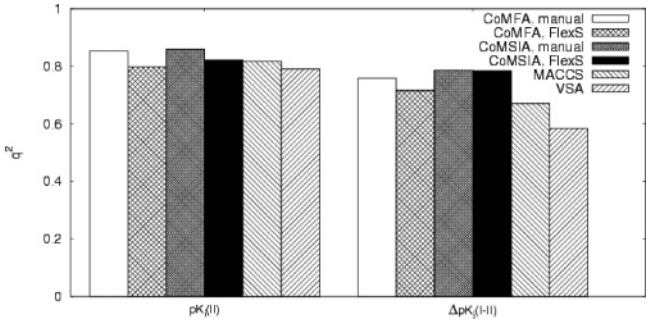


Figure 1. Comparison of performance of different models evaluated via cross-validation. The leave-one-out-q^2 is shown for prediction of $pK_i(II)$ (left) and $\Delta pK_i(I-II)$ (right), respectively.

Table 1. Statistical Results of the Different QSAR Analyses

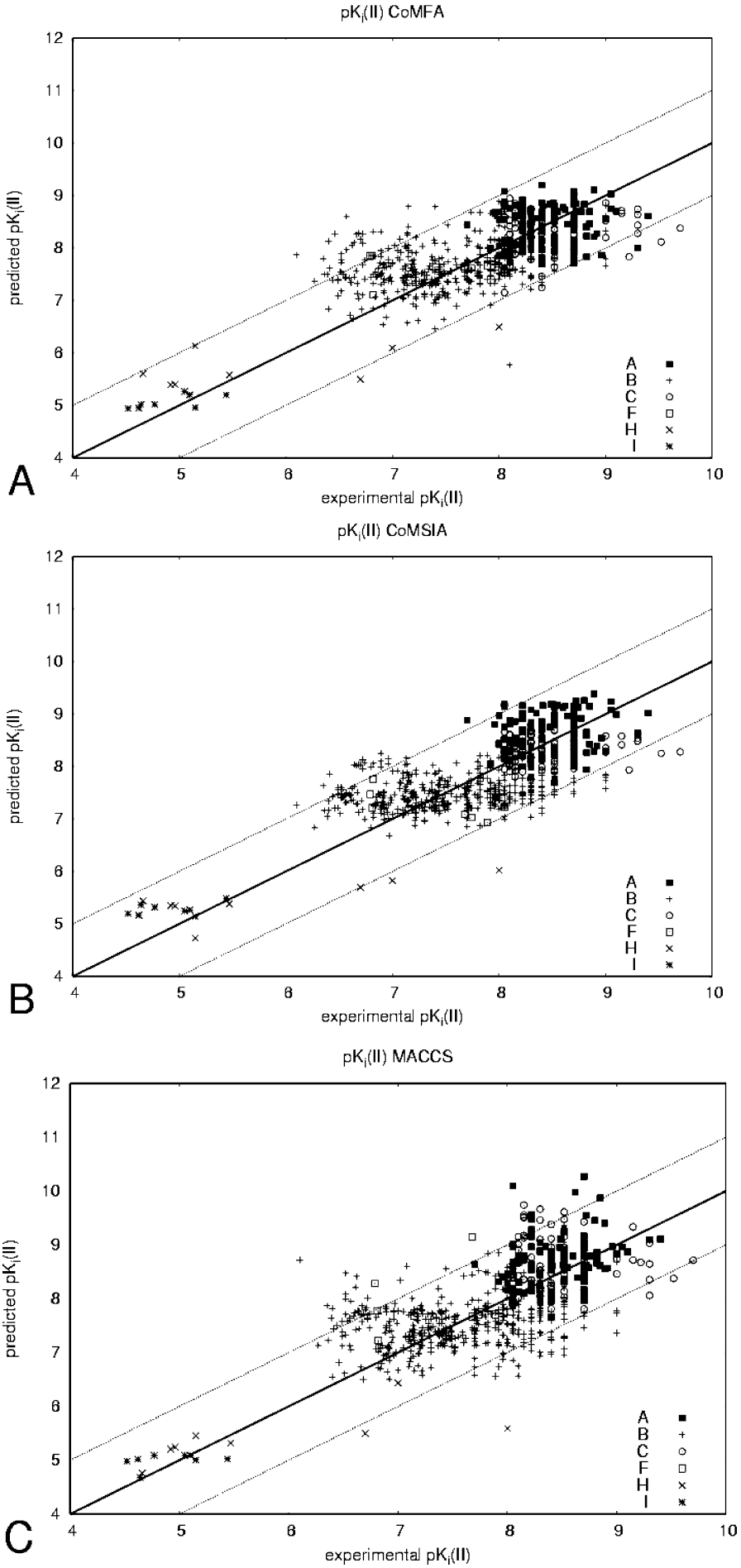
method	CoMFA		CoMSIA		MACCS	VSA
alignment	manual	FlexS	manual	FlexS		
dep. var.			$pK_i(II)$			
q^2	0.853	0.798	0.860	0.822	0.818	0.790
s_{PRESS}	0.504	0.593	0.489	0.552	0.556	0.629
r^2	0.949	0.880	0.943	0.867	0.884	0.837
S	0.297	0.457	0.313	0.476	0.444	0.553
F	423.840	243.854	453.840	441.585	264.532	27.33
no. comp.	6	4	5	2	4	17
dep. var.			$\Delta pK_i(I-II)$			
q^2	0.758	0.715	0.786	0.784	0.670	0.584
s_{PRESS}	0.598	0.629	0.552	0.548	0.684	0.799
r^2	0.977	0.851	0.950	0.905	0.802	0.660
S	0.184	0.455	0.368	0.364	0.529	0.720
F	633.455	190.350	330.962	316.402	188.962	14.028
no. comp.	9	4	4	4	3	13

Table 2. Numerical Measures of Predictivity for Different QSAR Methods

		CoMFA	CoMSIA	MACCS	VSA
pred. r^2	$pK_i(II)$	0.454	0.482	0.302	−0.710
	$\Delta pK_i(I-II)$	−0.079	0.001	−0.476	−0.893
Spearman's r^2	$pK_i(II)$	0.407	0.443	0.393	0.288
	$\Delta pK_i(I-II)$	0.115	0.118	0.069	0.000

The full set of statistical parameters is shown in Table 1.

External Numerical Predictivity: Correlation Coefficients. A test set of 663 sulfonamide type inhibitors was aligned with the above-mentioned FlexS protocol, subsequently $pK_i(II)$ and $\Delta pK_i(I-II)$ were predicted based on the 3D QSAR models previously obtained by FlexS alignment and the 1D/2D models (independent of conformation and orientation). Table 2 shows the predictive r^2 and Spearman's rank correlation coefficient for the different models. CoMSIA yields the best results in all cases, with CoMFA exhibiting slightly worse values. The MACCS approach performs significantly better than VSA; however, its predictive power is clearly worse compared to the 3D techniques. Regarding the size and structural imbalance of the test set with respect to the training set the results for prediction of $pK_i(II)$ are remarkably good for the 3D methods and still acceptable for the MACCS approach, whereas the VSA approach fails to make useful predictions. These findings confirm the well-known "beware of q^2 !" phenomenon, stating that a "good q^2" is by no means indicative that a model also possesses sufficient predictivity with respect to novel external compounds not regarded for training, though the relative ranking


pK$_i$(II) CoMFA
predicted pK$_i$(II)
experimental pK$_i$(II)
A
A
B
C
F
H
I
A
pK$_i$(II) CoMSIA
predicted pK$_i$(II)
experimental pK$_i$(II)
B
pK$_i$(II) MACCS
predicted pK$_i$(II)
experimental pK$_i$(II)
C

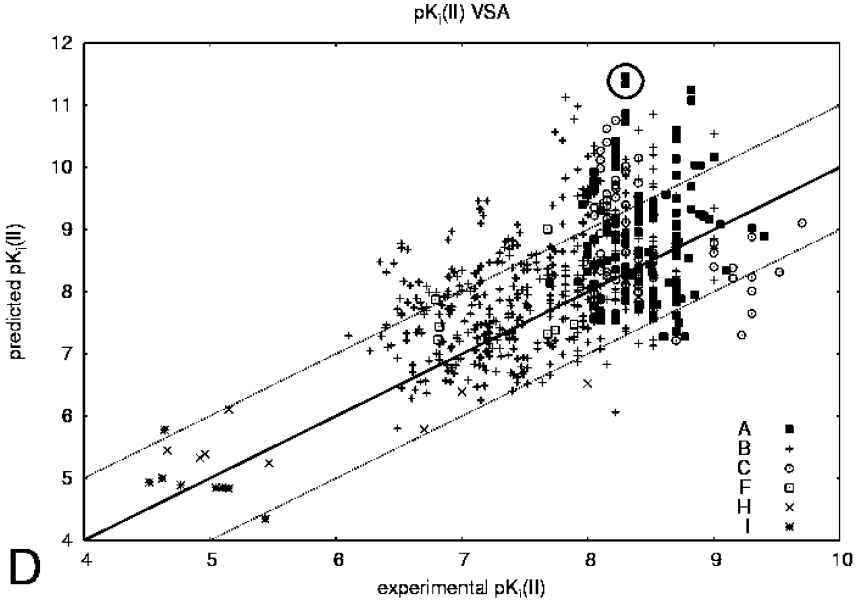


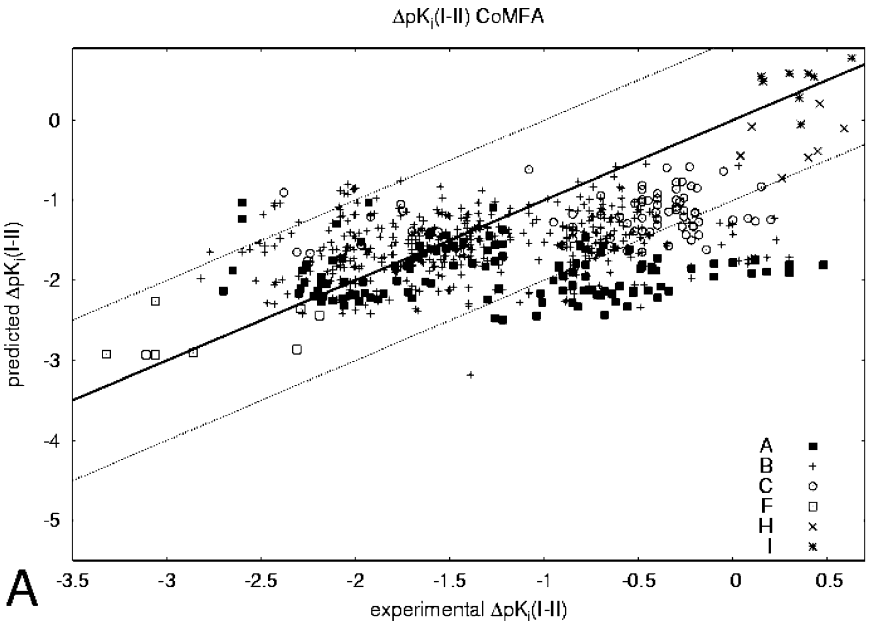
Figure 2. Plots of experimental versus predicted pK_i(II) values for the four QSAR approaches considered in this study. A: CoMFA; B: CoMSIA; C: MACCS; D: VSA. Dashed lines mark a range of ±1 logarithmic unit deviation; the solid line indicates perfect correlation. The shape of the points indicates the individual scaffold class of the respective ligands, the capital letters corresponding to the structural assignment given in Chart 1. The bold circle in part D marks two extreme outliers discussed in the text.

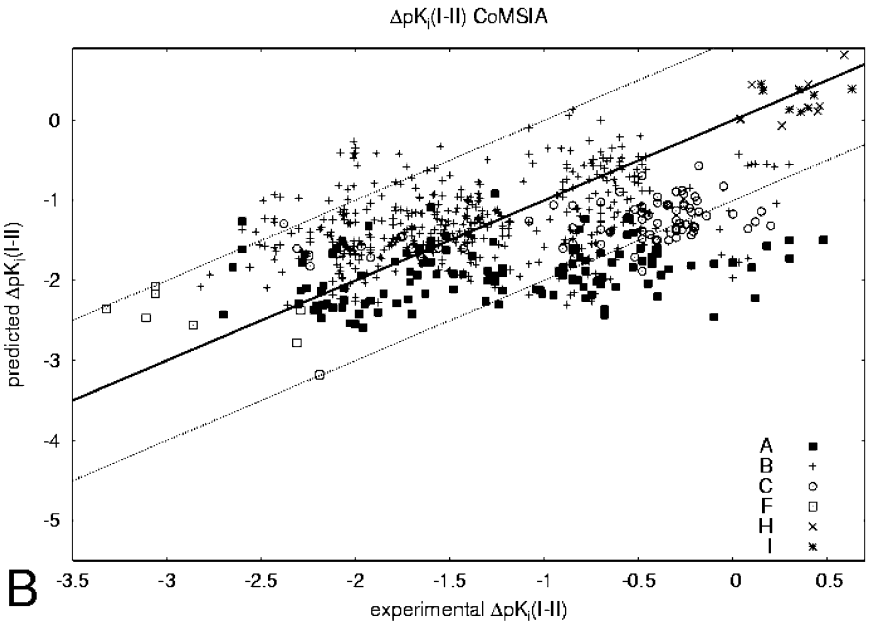
of the model quality can still be estimated based on the achieved q^2 in our study. Obviously all approaches are not able to give accurate numerical predictions for the selectivity variable $\Delta pK_i(I-II)$. Nevertheless, we will show in the following sections that these models are not completely useless with respect to screening purposes for which accurate numerical predictions are less important than correct assignment to affinity classes.

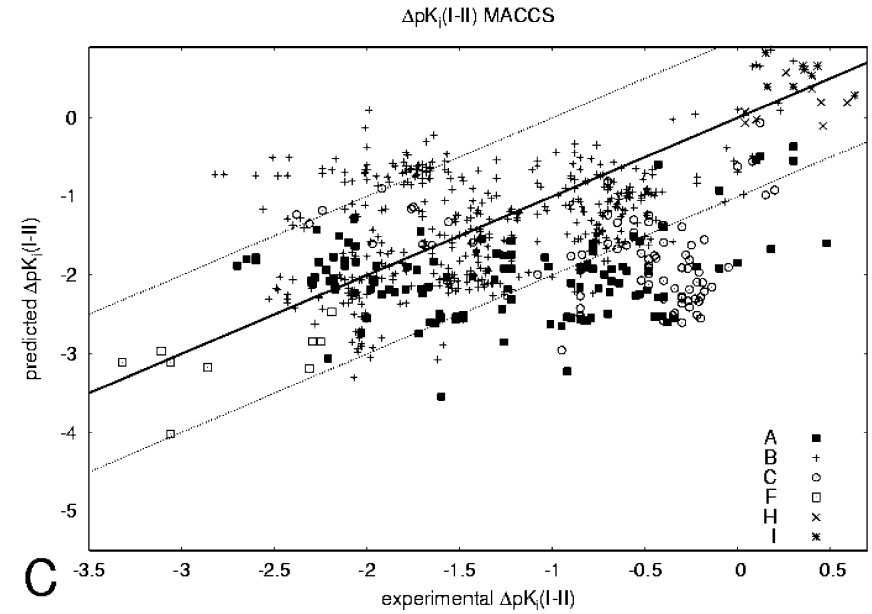
External Numerical Predictivity: Experimental versus Predicted Plots. Figure 2 shows the plots of experimental versus predicted pK_i(II) values for all four approaches. Each point corresponds to one molecule and indicates its membership to a chemical class by point shape. The distribution of points along the abscissa (i.e., the experimental pK_i(II)) reflects again the imbalance of the test set. Compounds with pK_i(II) values below 6.5 are rarely found, whereas those possessing pK_i(II) between 6.5 and 9.0 are clearly overrepresented. The plots demonstrate the value of CoMFA (Figure 2A) and CoMSIA (Figure 2B) for this large scale application since most of the test compounds are predicted correctly within ±1 logarithmic unit. The same holds for the MACCS key approach (Figure 2C) albeit with some more molecules falling outside this tolerance. Closer inspection of the plots reveals a tendency of the MACCS keys and even more of the VSA descriptors (Figure 2D) to overpredict some of the thiadiazolsulfonamides (scaffold **A**, Chart 1) and the benzothiazolsulfonamides (**C**). In order to shed some light on the reason for this finding, we pick two molecules marked by a bold circle in Figure 2D whose affinity is predicted more than 3 logarithmic units too high by the VSA model, whereas the other approaches provide a reasonable estimate. A projection of these molecules into the PCA space of the training set calculated using VSA descriptors reveals them as extreme structural outliers. They contain very bulky phenylpyridinium groups which are not present in any of the training set compounds. The most relevant descriptor responsible for the overprediction is SMR_VSA5 (strongly correlated to molecule volume and polarizability) which for most of the training set molecules does not exceed an upper limit of 160 but takes values of 212 and 231 for both ill-predicted compounds, respectively. Obviously, the field- and fragment-based descriptors are more robust with respect to structural extrapolation, at least in the case of our data set. It has to be noted, however, that in CoMFA/CoMSIA a descriptor used to evaluate a compound contains several thousand values, however, for MACCS only 166 and for VSA even only 32, respectively. Thus, it is not too surprising that a set of VSA values is only a crude approximative description of a molecule. Moreover, no information about the spatial distribution of the properties encoded by the VSA descriptors about the molecules is contained in the latter descriptors.

As expected deviations from correct predictions are generally worse for selectivity prediction of $\Delta pK_i(I-II)$ as shown in Figure 3. However, despite the poor predictive r^2 for all models the overall appearance of the plots suggests that the 3D QSAR models can give rough estimates in terms of selectivity. The group of thiadiazolesulfonamides (scaffold **A**, Chart 1) with experimental $\Delta pK_i(I-II) > -1$ is remarkably underpredicted by all four models. This can be easily explained by the fact that in the training set this chemical class has a mean $\Delta pK_i(I-II)$ of −1.91 with a maximum of −0.71. The poorly predicted subset, however, exhibits a mean $\Delta pK_i(I-II)$ of −0.52. Correct prediction would require clear extrapolation. In the case of benzothiazolsulfonamides (scaffold **C**) the MACCS and VSA models perform significantly worse compared to the 3D models.

External Categorical Predictivity: Sensitivity, Specificity, and Hit Rates. In order to assess the categorical external predictivity of the established models we will first report the resulting sensitivities (Se), specificities (Sp), and hit rates (H) after a distinct threshold for the modeled variable has been defined. Each compound is labeled "positive" or "negative" according to the above-mentioned arbitrarily chosen threshold. The thresholds are selected such that compounds with the highest/lowest 2% or 5% of pK_i(II)/


ΔpK$_i$(I-II) CoMFA
predicted ΔpK$_i$(I-II)
experimental ΔpK$_i$(I-II)
A
B
C
F
H
I
A


ΔpK$_i$(I-II) CoMSIA
predicted ΔpK$_i$(I-II)
experimental ΔpK$_i$(I-II)
A
B
C
F
H
I
B


ΔpK$_i$(I-II) MACCS
predicted ΔpK$_i$(I-II)
experimental ΔpK$_i$(I-II)
A
B
C
F
H
I
C

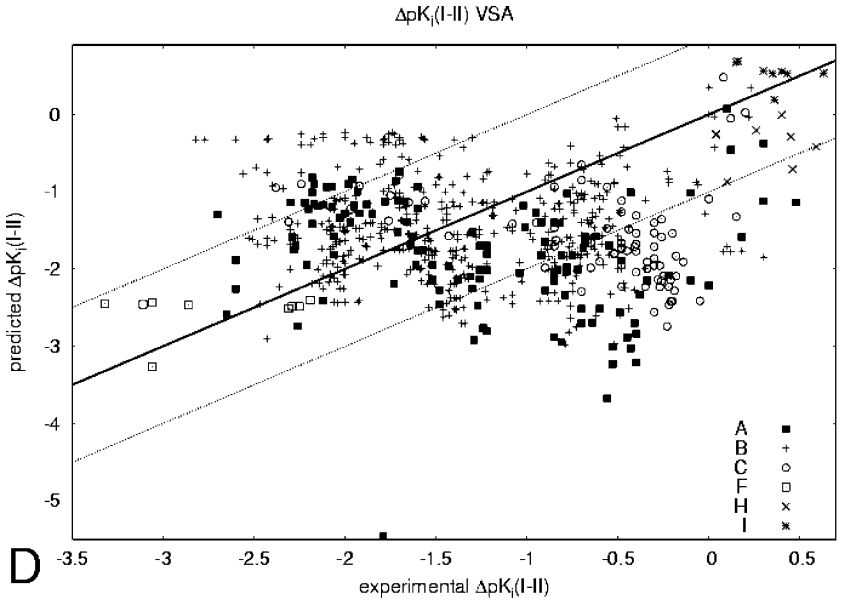


Figure 3. Plots of experimental versus predicted $\Delta pK_i(I-II)$ for the four QSAR approaches considered in this study. A: CoMFA; B: CoMSIA; C: MACCS; D: VSA. Dashed lines mark a range of ± 1 logarithmic unit deviation; the solid line indicates perfect correlation. The shape of the points indicates the individual scaffold class of the respective ligands, the capital letters corresponding to the structural assignment given in Chart 1.

$\Delta pK_i(I-II)$ are retrieved. In a screening scenario one is usually only interested in identifying compounds with "extreme" activities either to find those with high affinity for a particular target, with low affinity for an antitarget, or with extraordinary selectivity profiles. In a real life situation an even lower amount (≤1%) would be of interest due to the large size of databases screened, but for our present study this would result in a very small absolute number of molecules. Most likely rather unstable statistical results would be suggested. The same threshold will then be applied to the calculated values, and the comparison with the experimentally determined classification yields the assignment to "true positives/negatives" and "false positives/negatives" (TP, TN, FP, FN). The main difference between this kind of evaluation and the ROC and enrichment plots described below is that the latter methods monitor the evolution of Se or Sp in dependency on a variable threshold, whereas the approach applied here analyzes the classification performance taking the predicted values "as is".

The resulting sensitivities, specificities, and hit rates are shown in Figure 4 for the various approaches and thresholds. The plots reveal a trend toward the 1D/2D methods exhibiting a higher sensitivity, i.e., they tend to omit less actives than the 3D models. This is achieved at the cost of a reduced specificity as they also label many inactives as actives. This results in most cases in higher hit rates for the 3D QSAR approaches. The MACCS keys-based models perform significantly better compared to the VSA models; they yield high sensitivities in conjunction with reasonable specificities.

The plots also show that the results are generally better to identify compounds with minimal $pK_i(II)$ or maximal $\Delta pK_i(I-II)$, respectively, compared to the case of maximal $pK_i(II)$ or minimal $\Delta pK_i(I-II)$. This finding corresponds to the observation described above that compounds with very low pK_i (i.e., the classes of scaffolds **H** and **I**) are usually well predicted, whereas those with high $pK_i(II)$ (particularly the scaffolds **A** and **C**) are often overpredicted. In consequence corresponding results are obtained for the difference $\Delta pK_i(I-II)$.

External Categorical Predictivity: ROC Plots. Figure 5 shows two examples of ROC plots monitoring the screening progress. The main diagonal corresponds to a random classifier unable to discriminate signal from noise. Thus, for any possible threshold the same percentage for sensitivity (signal) and 1 − specificity (noise) is achieved. Its AUC is 50.0%, and any classifier better than random therefore has to produce an AUC above this lower limit. The curve of an ideal classifier coincides with the left and the top edge of the coordinate system and encloses an AUC of 100.0%.

Figure 5A shows the ROC plots for the retrieval of the 5% compounds with lowest $pK_i(II)$. For this example all four methods perform similarly well. For a real life scenario the left section of the plot is most interesting since it indicates how much signal (actives) can be identified by the model still discarding most of the noise (inactives). In our case about 35% of actives can be retrieved without extracting false positives. For the remaining part of the screening, the MACCS descriptors perform slightly worse compared to the other models. At higher noise levels ([1 − specificty]=0.4) the 3D methods slightly outperform the 1D and 2D approaches.

In Figure 5B, the ROC curves to identify the 5% compounds with lowest $\Delta pK_i(I-II)$ (i.e., the most selective ones for hCA II) are shown. Here, the differences between the four approaches are more pronounced. The worst performance is indicated for the VSA model intermediately even dropping below the random line. All approaches exhibit a poorer performance in the "interesting" left part of the plot compared to the previous example. In this area MACCS and 3D models perform similarly, whereas at specificities below 80% ([1 − specificity] > 0.2) 3D QSAR models outperform the 1D/2D models.

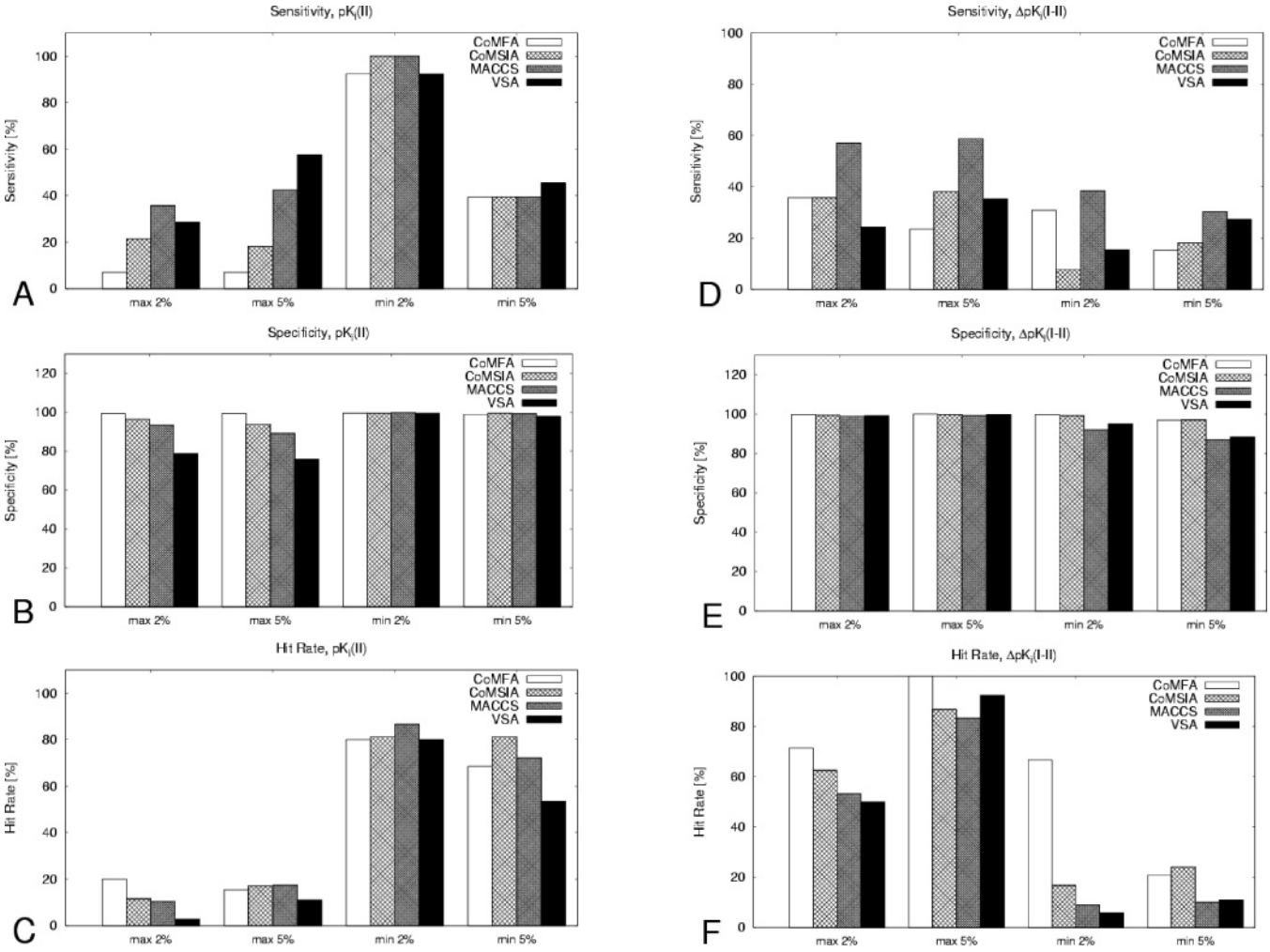


Figure 4. Comparison of sensitivity (A, D), specificity (B, E), and hit rate (C, F) for the four QSAR methods. A, B, C: pK_i(II); D, E, F: ΔpK_i(I−II).

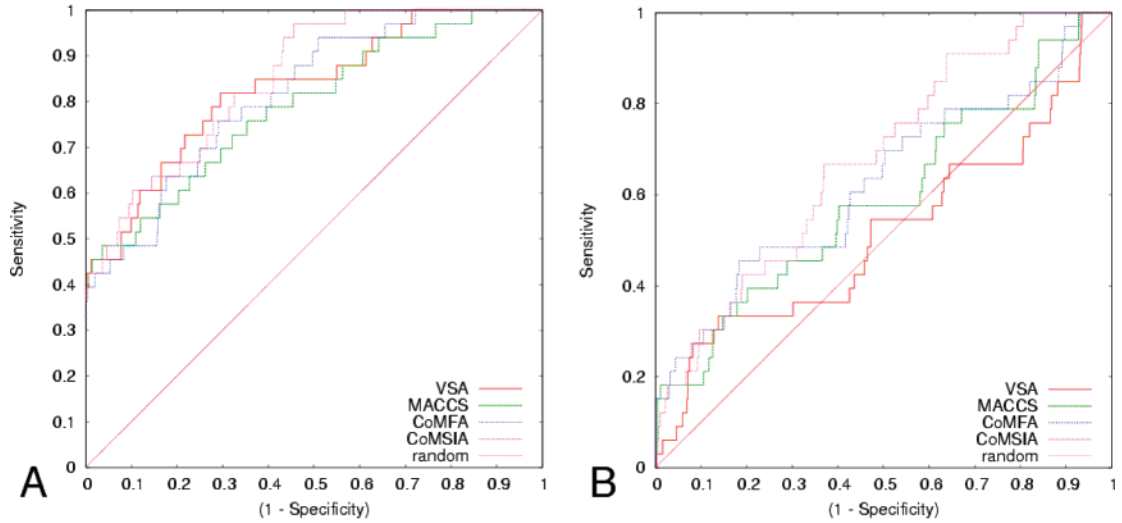


Figure 5. Receiver operating characteristic (ROC) curves for the retrieval of the 5% compounds with lowest pK_i(II) (A) and ΔpK_i(I−II) (B), respectively, applying the four QSAR approaches considered in this study. The main diagonal corresponds to a random classifier, whereas the left and top edge of the plot would show the ROC line of an ideal classifier.

In order to present a concise comparison for the other retrieval experiments, Figure 6 shows bar plots denoting the corresponding AUCs. For classification with respect to pK_i-(II) (Figure 6A), all approaches perform comparably well; however, the VSA models tend to be worse in identifying compounds with maximum pK_i(II). In the case of selectivity prediction (ΔpK_i(I−II), Figure 6B) the 1D/2D models outperform 3D QSAR in successfully retrieving the most hCA I selective compounds, whereas the opposite is true for identification of hCA II selective compounds. As demonstrated above, 3D QSAR methods obviously suffer less from the overprediction problem than the 1D/2D

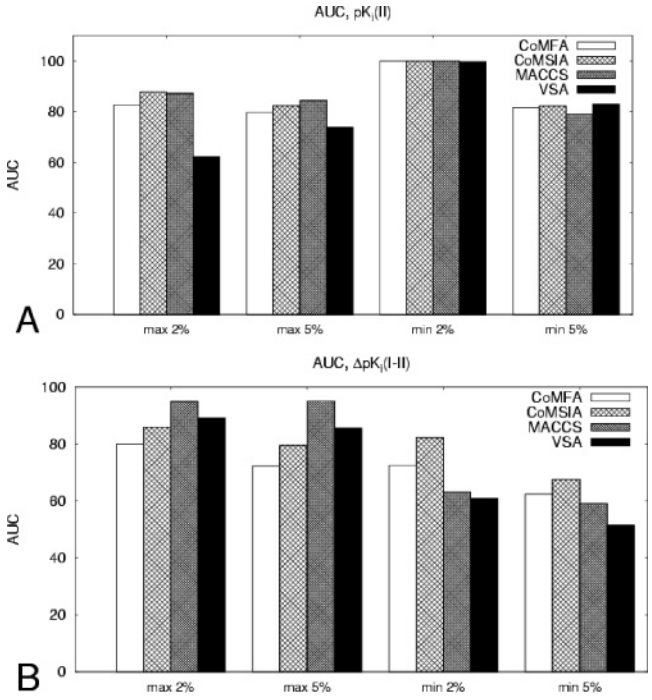


Figure 6. Bar plots displaying the area under the curve (AUC) values of the ROC curves for the four QSAR approaches considered in this study and activity thresholds obtained when pK_i(II) (A) and ΔpK_i(I−II) (B) are predicted.

techniques. The decreasing performance of the 3D models to classify the most hCA I selective compounds results from the fact that already the individual pK_i(I) values for this isoform are modeled with reduced accuracy (data not shown).

It has to be noted that the AUCs of the ROC curves allow for an overall comparison of the different approaches. However, the detailed characteristics of a classifier's performance can only be deduced considering the overall shape of a ROC curve.

External Categorical Prediction: Classical Enrichment Plots. The important information which cannot be derived from the ROC plots is the amount of actives identified when a certain subset of the database is screened. This type of information is provided by the classical enrichment plots usually used to visualize the screening performance in drug design. Similarly to the ROC curves the left part of the plots is the most interesting for practical applications. We will illustrate such an evaluation using the same examples as for the ROC plots.

Figure 7A shows the plot for the enrichment of the 5% compounds with lowest pK_i(II). Within the first few percent of the screened database all four methods successfully separate actives from inactives. For the remainder of the screening process all models show a similarly satisfying albeit not perfect performance.

Figure 7B displays the enrichment curves for the retrieval of the 5% compounds with highest selectivity toward hCA II. On the very left-hand side of the curves CoMFA, CoMSIA, and MACCS models discriminate satisfactorily for the first 1−2% of the database, whereas the VSA model does not depart from a random selection. For the remainder of the screening progress all models perform rather disappointingly; however, the 3D models still show some advantages. The VSA descriptors remain close to the performance of a random classifier.

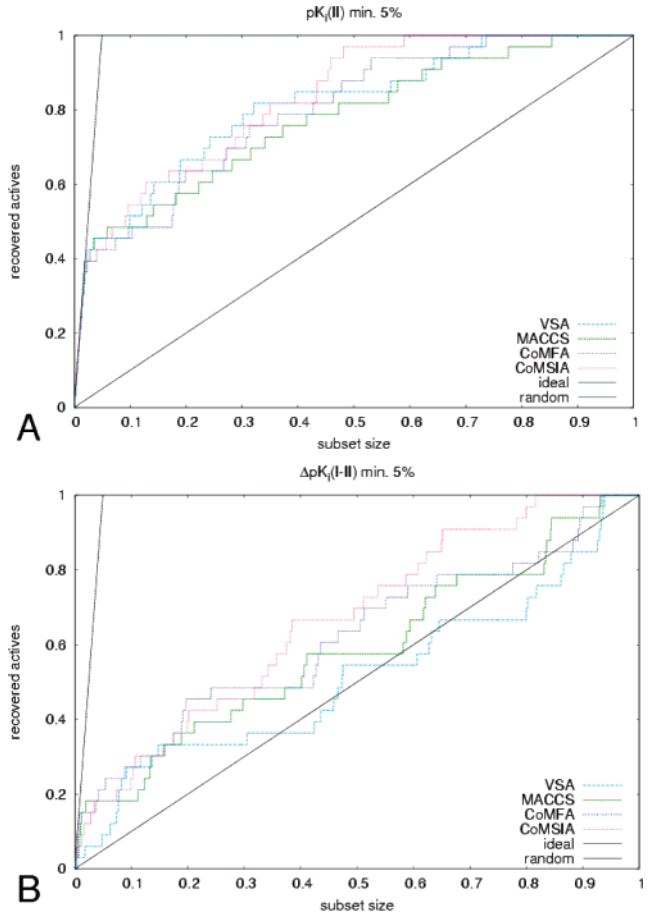


Figure 7. Enrichment plots for the retrieval of the 5% compounds with lowest pK_i(II) (A) and ΔpK_i(I−II) (B), respectively, applying the four QSAR approaches considered in this study. The main diagonal corresponds to a random selection, the steep line to the left to an ideal retrieval.

Interpretation of the MACCS Models. Besides the use of QSAR equations to predict novel compounds the coefficients of the equation indicate the relative importance of individual descriptors (usually multiplied by the standard deviation of the descriptor, denoted as stdev*coeff). With respect to the evaluation of the 3D QSAR models we refer to our recent paper[23] giving a detailed interpretation in terms of contour maps. Depending on the method applied, the resulting contours indicate either purely ligand-based, purely protein-based, or mixed protein−ligand-based information. In contrast to this former study where an alignment has been produced exploiting the protein binding pocket as a reference, here we have used a similarity-based alignment strategy. Therefore, conclusions based on contour maps have to be interpreted with some caution particularly referring to properties given by the protein environment. Nevertheless, since several crystal structures of complexes were used as reference some correlation between the properties of the binding pocket and the contour maps derived from the aligned ligands can be expected.

Each of the 166 MACCS descriptors captures frequency of occurrence of a distinct molecular fragment in a molecule. Thus, if the product stdev*coeff adopts large absolute values the corresponding fragment takes strong influence on the biological activity. If the sign is positive its occurrence enhances biological activity (or selectivity, respectively). If

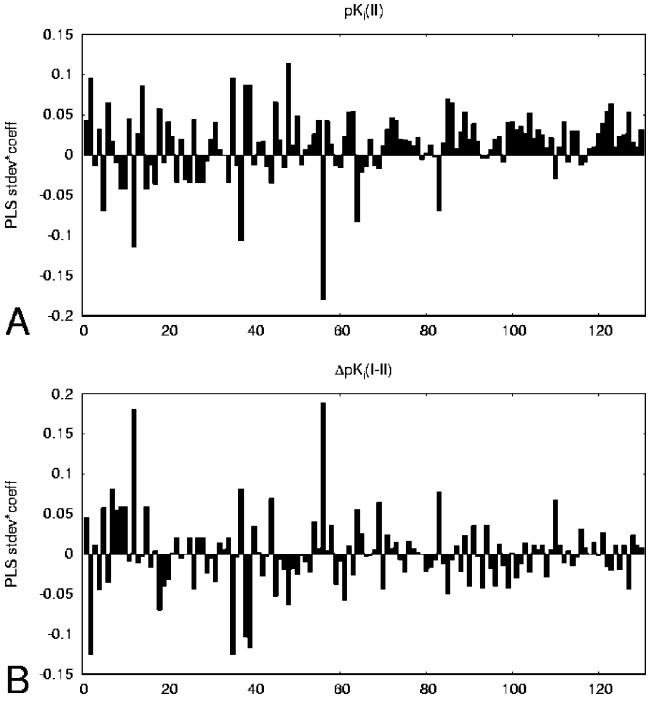


Figure 8. Bar plots of stdev*coeff derived from MACCS QSAR equations for prediction of pK_i(II) (A) and ΔpK_i(I−II) (B), respectively.

it is negative, the presence of that fragment in a molecule is detrimental for its activity. Figure 8A displays a plot of 130 stdev*coeff values derived from the QSAR equation for pK_i(II), one for each MACCS key (36 of the 166 were discarded due to zero variance). The highest peak is experienced by MACCS_81 (MACCS numbering corresponds to the original MDL numbering and not to the numbering scheme used in Figure 8), which corresponds to the fragment SA(A)A (where A is any atom except hydrogen). This pattern is highly abundant in the data set, but molecules exhibiting this fragment particularly frequently (two or three times, found in 25 molecules) achieve an extraordinarily high pK_i(II) (mean(subset) = 8.63, mean(all) = 7.17). The corresponding molecules possess in general more than one sulfonamide group mostly along with a thiophene ring. For example, the high-affinity thienothiopyranes (scaffold **A**) have one sulfonamide anchor and a second sulfonyl group in the thiopyrane ring, and additionally a third time the pattern is matched by the thiophene ring. An example indicating affinity decrease is fragment MACCS_89, which encodes the number of OAAAO substructures. This pattern occurs particularly in sulfamates (scaffold **I**) which have significantly lower pK_i(II) values (mean(subset) = 5.25) compared to the entire training set. This is due to the replacement of the sulfonamide anchor by a sulfamate group, which is known to be a poorer zinc binder.

Figure 8B shows a plot of stdev*coeff derived from the QSAR equation for ΔpK_i(I−II). The high value of the product stdev*coeff for MACCS_43 means that the substructure QHAQH (where Q is any atom except hydrogen and carbon) increases selectivity toward hCA I. The average for ΔpK_i(I−II) is −1.29 (i.e., most compounds inhibit hCA II stronger than hCA I). Molecules with higher occurrence of this fragment compared to the remainder of the training set (46 molecules with more than one occurrence) hence possess a higher ΔpK_i(I−II) (mean(subset) = 0.03). Compounds of this subset comprise structures with scaffold **H**, (thio)urea, and guanidinium moieties, respectively. An example for substructures increasing selectivity toward hCA II is MACCS_72 encoding OAAO. The average ΔpK_i(I−II) for molecules containing this fragment more than once is −2.64. Hydroxysulfonamides (scaffold **F**) as well as orthodimethoxy substituted phenyl rings are captured by this pattern. A comparison of the stdev*coeff plots shows that many of the values are anticorrelated, i.e., the same MACCS key is associated with a positive sign for pK_i(II) and a negative sign for ΔpK_i(I−II) and vice versa. This is a consequence of the fact that a factor increasing affinity toward hCA II often diminishes the difference ΔpK_i(I−II).

These examples show not only that the MACCS-based models are useful for prediction of novel compounds but also that the coefficients of the derived equation can be used to identify substructures of importance to influence a molecular property such as activity. Of course, the conclusions drawn from this type of analysis are only of statistical nature. There is no basis to assume a causal relationship between occurrence of a distinct fragment and its effect on biological activity. It might well be, that, as a hypothetical example, a compound class such as thiophenes most frequently contains several sulfonyl groups and exhibits at the same time higher affinities than the average of the training set. However, it remains unresolved whether the thiophene ring, sulfonyl groups, or the combination of both are responsible for this observation. Since the MACCS keys do not take the functional characteristics of the encoded groups into account with respect to protein−ligand interactions, they can only indirectly evidence a possible correlation. Another limitation of this approach is that many of the encoded fragments are rather small and do not represent a meaningful "chemical unit". The approach also does not consider any information about connectivity between single fragments. Thus, the method will fail if molecules are attempted to be predicted which possess a similar count of the same fragments as a subset of ligands from the training set but connected in a different way. Clearly, this shortcoming will not occur using a 3D QSAR method as an evaluation technique. Nevertheless, if narrow focused libraries with close similarity to the training set compounds are evaluated, the MACCS key based method will probably yield reasonable accuracy. An advantage of the fragmental description of the molecules is—despite ignorance of actual functional groups—the simple and straightforward translation into chemical structures. This is in contrast to many other property- or graph-based descriptors commonly applied in QSAR studies, e.g., the VSA descriptors used in our fourth approach to evaluate the data set. We therefore will refrain from a detailed interpretation. In principle the relevance of properties such as logP, MR etc. falling into distinct intervals could be assessed. However, the obtained information will remain rather indirect and general due to the global character of the descriptors. Particularly, it faces the problem of a nontrivial translation into chemical structures in accordance with the highlighted properties.

CONCLUSION

Within the present study we assessed the predictive power of QSAR approaches with respect to their applicability to

screen databases. Since we were especially interested in 3D methods, a protocol had to be established which is reliable and robust enough to produce consistent spatial alignments of the molecules under consideration. Due to the large number of compounds encountered in real life screening scenarios the protocol—once set up—has to be applicable without further manual intervention. FlexS in combination with automated recognition of a chemical compound class of ligands to be superimposed via the MAPREF methodology was chosen to successfully accomplish this task. We could demonstrate that CoMFA and CoMSIA models based on this alignment perform comparably well with similar models based on manually derived alignments using the protein's binding pocket as a reference point along with subsequent force field relaxation. Since the superimposition comprises a rather elaborate and time-consuming step we also tested the performance of alignment-free 1D and 2D QSAR models particularly with respect to database screening. Therefore, fragment-based MACCS descriptors and property-based VSA descriptors were computed based on the 2D molecular information. The external predictivity was assessed based on a test set of 663 compounds with known activities. Of course, this number of test molecules does not touch the size of a real library, but it should be sufficiently large for the intended benchmark test. In summary, the 3D QSAR models and the MACCS keys performed quite well with respect to affinity prediction (pK_i(II)), whereas the VSA descriptors did not achieve to establish models with comparable predictive power. In terms of numerical affinity prediction the 3D QSAR models significantly outperformed the 1D and 2D approaches. They tend to be more specific than the MACCS keys but at the cost of a lower sensitivity. The models are difficult to mutually rank against each other since relevance, predictive value, and applicability depend on the specific goal of the project, e.g., whether retrieval of only a few compounds with an enhanced activity is intended or whether as many actives as possible should be detected. Thus, even if the MACCS approach is easier to perform and computationally less demanding it does not make the 3D methods superfluous for screening purposes. Also with respect to the aspect of generality of the descriptors the molecular fields of a ligand approximate better the concept of molecular recognition, and second they should not suffer from the fact of missing connectivity information as the MACCS keys do. Most likely, the 3D QSAR methods are much more robust in handling data sets composed of compounds with structurally rather diverse molecular skeletons. Nevertheless, we could show that the PLS coefficients of the derived MACCS model can be interpreted meaningfully and used to extract knowledge about the influence of individual fragments on the dependent variables.

With respect to a quantitative selectivity prediction, only the internal consistency was convincing. None of the techniques was able to make satisfying numerical forecasts on this large data set. However, this can be attributed to the rather imbalanced composition of the test set and the generally higher error contained in a "composed variable". The results on categorical predictivity suggest that the 3D models can still give crude estimates about selectivity.

The main purpose of this study was to systematically assess the performance of 3D QSAR models with respect to database screening. Although the results are convincing, one has to keep in mind that they will depend on the training and test set composition and cannot be transferred generally. Furthermore, additional investigations considering other QSAR techniques (other statistical methods, 4D,[42] 5D,[43] 6D[44] approaches) need to be done to collect more experience on the scope and limitations of QSAR methods for database screening.

ACKNOWLEDGMENT

The authors acknowledge the kind support of BioSolveIT with respect to special FlexS parameters, in particular Markus Lilienthal. The Chemical Computing Group (CCG) is acknowledged for provision of one research license of MOE. The authors are grateful to Prof. Dr. Claudiu T. Supuran (University of Florence) for making the data set of CA I and CA II inhibitors available to us.

Supporting Information Available: Atomic coordinates of all molecules of the data set with assigned pK_i(I), pK_i(II), and ΔpK_i(I−II) values as SD file. This material is available free of charge via the Internet at http://pubs.acs.org.

REFERENCES AND NOTES

(1) Hansch, C.; Fujita, T. r-s-p analysis - A method for the correlation of biological activity and chemical structure. *J. Am. Chem. Soc.* **1964**, *86*, 1616−1626.
(2) Free, S. M., Jr.; Wilson, J. W. A Mathematical Contribution to Structure-Activity Studies. *J. Med. Chem.* **1964**, *53*, 395−9.
(3) Kubinyi, H. *QSAR: Hansch Analysis and Related Approaches*; VCH: Weinheim, 1993; Vol. 1.
(4) Zheng, W.; Tropsha, A. Novel variable selection quantitative structure−property relationship approach based on the k-nearest-neighbor principle. *J. Chem. Inf. Comput. Sci.* **2000**, *40*, 185−94.
(5) Brown, N.; Lewis, R. A. Exploiting QSAR methods in lead optimization. *Curr. Opin. Drug Discovery Dev.* **2006**, *9*, 419−24.
(6) Shen, M.; Beguin, C.; Golbraikh, A.; Stables, J. P.; Kohn, H.; Tropsha, A. Application of predictive QSAR models to database mining: identification and experimental validation of novel anticonvulsant compounds. *J. Med. Chem.* **2004**, *47*, 2356−64.
(7) Oloff, S.; Mailman, R. B.; Tropsha, A. Application of validated QSAR models of D1 dopaminergic antagonists for database mining. *J. Med. Chem.* **2005**, *48*, 7322−32.
(8) Tropsha, A. Application of Predicitve QSAR Models to Database Mining. In *Chemoinformatics in Drug Discovery*; Oprea, T. I., Ed.; Wiley-VCH: Weinheim, 2004; Vol. 23, pp 437−455.
(9) Moro, S.; Bacilieri, M.; Cacciari, B.; Bolcato, C.; Cusan, C.; Pastorin, G.; Klotz, K. N.; Spalluto, G. The application of a 3D-QSAR (autoMEP/PLS) approach as an efficient pharmacodynamic-driven filtering method for small-sized virtual library: application to a lead optimization of a human A3 adenosine receptor antagonist. *Bioorg. Med. Chem.* **2006**, *14*, 4923−*32*.
(10) Pastor, M.; Cruciani, G.; McLay, I.; Pickett, S.; Clementi, S. GRid-INdependent descriptors (GRIND): a novel class of alignment-independent three-dimensional molecular descriptors. *J. Med. Chem.* **2000**, *43*, 3233−43.
(11) Carosati, E.; Mannhold, R.; Wahl, P.; Hansen, J. B.; Fremming, T.; Zamora, I.; Cianchetta, G.; Baroni, M. Virtual screening for novel openers of pancreatic K(ATP) channels. *J. Med. Chem.* **2007**, *50*, 2117−26.
(12) Benedetti, P.; Mannhold, R.; Cruciani, G.; Ottaviani, G. GRIND/ALMOND investigations on CysLT1 receptor antagonists of the quinolinyl(bridged)aryl type. *Bioorg. Med. Chem.* **2004**, *12*, 3607−17.
(13) Murcia, M.; Ortiz, A. R. Virtual screening with flexible docking and COMBINE-based models. Application to a series of factor Xa inhibitors. *J. Med. Chem.* **2004**, *47*, 805−20.
(14) Ortiz, A. R.; Pisabarro, M. T.; Gago, F.; Wade, R. C. Prediction of drug binding affinities by comparative binding energy analysis. *J. Med. Chem.* **1995**, *38*, 2681−2691.
(15) Zhang, Q. Y.; Wan, J.; Xu, X.; Yang, G. F.; Ren, Y. L.; Liu, J. J.; Wang, H.; Guo, Y., Structure-based rational quest for potential novel inhibitors of human HMG-CoA reductase by combining CoMFA 3D QSAR modeling and virtual screening. *J. Comb. Chem.* **2007**, *9*, 131−8.

(16) Cramer, R. D. Comparative Molecular Field Analysis, (CoMFA). 1. Effect of Shape on Binding of Steroids to Carrier Proteins. *J. Am. Chem. Soc.* **1988**, *110*, 5959−5967.
(17) Cramer, R. D.; DePriest, S. A.; Patterson, D. E.; Hecht, P. The Developing Practice of Comparative Molecular Field Analysis. In *3D QSAR in Drug Design: Theory Methods and Applications*; Kubinyi, H., Ed.; ESCOM: Leiden, 1993; pp 443−485.
(18) Klebe, G.; Abraham, U.; Mietzner, T. Molecular similarity indices in a comparative analysis (CoMSIA) of drug molecules to correlate and predict their biological activity. *J. Med. Chem.* **1994**, *37*, 4130−46.
(19) Klebe, G. Comparative Molecular Similarity Indices Analysis: CoMSIA. *Perspect. Drug Discovery Des.* **1998**, *12/13/14*, 87−104.
(20) Klebe, G., Structural Alignment of Molecules. In *3D QSAR in Drug Design: Theory Methods and Applications*; Kubinyi, H., Ed.; ESCOM: Leiden, 1993; Vol. 1, pp 173−199.
(21) Lemmen, C.; Lengauer, T. Computational methods for the structural alignment of molecules. *J. Comput.-Aided Mol. Des.* **2000**, *14*, 215−32.
(22) Lemmen, C.; Lengauer, T.; Klebe, G. FLEXS: a method for fast flexible ligand superposition. *J. Med. Chem.* **1998**, *41*, 4502−20.
(23) Hillebrecht, A.; Supuran, C. T.; Klebe, G. Integrated approach using protein and ligand information to analyze selectivity- and affinity-determining features of carbonic anhydrase isozymes. *ChemMedChem* **2006**, *1*, 839−53.
(24) MDL Information Systems, Inc. 14600 Catalina Street, San Leandro, CA 94577.
(25) Durant, J. L.; Leland, B. A.; Henry, D. R.; Nourse, J. G. Reoptimization of MDL keys for use in drug discovery. *J. Chem. Inf. Comput. Sci.* **2002**, *42*, 1273−80.
(26) Labute, P. A widely applicable set of descriptors. *J. Mol. Graphics Modell.* **2000**, *18*, 464−77.
(27) Witten, I. H.; Frank, E. *Data Mining: Practical Machine Learning Tools and Techniques*, 2nd ed.; Morgan Kaufmann Publishers: San Fransisco, 2005.
(28) Triballeau, N.; Bertrand, H. O.; Acher, F. Are You Sure You Have a Good Model? In *Pharmacophores and Pharmacophore Searches*; Langer, T., Hoffmann, R. D., Eds.; Wiley-VCH: Weinheim, 2006; Vol. 32, pp 325−364.
(29) Maren, T. H. Carbonic anhydrase: chemistry, physiology, and inhibition. *Physiol. Rev.* **1967**, *47*, 595−781.
(30) Lindskog, S. Structure and mechanism of carbonic anhydrase. *Pharmacol. Ther.* **1997**, *74*, 1−20.
(31) Geers, C.; Gros, G. Carbon dioxide transport and carbonic anhydrase in blood and muscle. *Physiol. Rev.* **2000**, *80*, 681−715.
(32) Supuran, C. T.; Scozzafava, A. Carbonic Anhydrase Inhibitors. *Curr. Med. Chem.: Imm., Endoc., Metab. Agents* **2001**, *1*, 61−97.
(33) Supuran, C. T.; Scozzafava, A. Applications of carbonic anhydrase inhibitors and activators in therapy. *Expert Opin. Ther. Pat.* **2002**, *12*, 217−242.
(34) *SYBYL molecular modeling package*, *version 7.1*; Tripos Inc.: 1699 South Hanley Road, Suite 303, St. Louis, MO 63144, 2005.
(35) *MOE*; Chemical Computing Group: Montreal, Canada.
(36) AutoQSAR. In The SVL script AutoQSAR is freely available to MOE licensees and can be downloaded at http://svl.chemcomp.com.
(37) Bush, B. L.; Nachbar, R. B., Jr. Sample-distance partial least squares: PLS optimized for many variables, with application to CoMFA. *J. Comput.-Aided Mol. Des.* **1993**, *7*, 587−619.
(38) Thibaut, U.; Folkers, G.; Klebe, G.; Kubinyi, H.; Merz, A.; Rognan, D. Recommendations for CoMFA Studies and 3D QSAR Publications. In *3D QSAR in Drug Design: Theory Methods and Applications*; Kubinyi, H., Ed.; ESCOM: Leiden, 1993; Vol. 1, pp 711−716.
(39) Wold, S.; Johansson, E.; Cocchi, M. PLS - Partial Least-Squares Projections to Latent Structures. In *3D QSAR in Drug Design: Theory Methods and Applications*; Kubinyi, H., Ed.; ESCOM: Leiden, 1993; pp 523−550.
(40) Evers, A.; Hessler, G.; Matter, H.; Klabunde, T. Virtual screening of biogenic amine-binding G-protein coupled receptors: comparative evaluation of protein- and ligand-based virtual screening protocols. *J. Med. Chem.* **2005**, *48*, 5448−65.
(41) Triballeau, N.; Acher, F.; Brabet, I.; Pin, J. P.; Bertrand, H. O. Virtual screening workflow development guided by the "receiver operating characteristic" curve approach. Application to high-throughput docking on metabotropic glutamate receptor subtype 4. *J. Med. Chem.* **2005**, *48*, 2534−47.
(42) Hopfinger, A. J.; Wang, S.; Tokarski, J. S.; Jin, B.; Albuquerque, M.; Madhav, P. J.; Duraiswami, C. Construction of 3D-QSAR Models Using the 4D-QSAR Analysis Formalism. *J. Am. Chem. Soc.* **1997**, *119*, 10509−10524.
(43) Vedani, A.; Dobler, M. 5D-QSAR: the key for simulating induced fit? *J. Med. Chem.* **2002**, *45*, 2139−49.
(44) Vedani, A.; Dobler, M.; Lill, M. A. Combining protein modeling and 6D-QSAR. Simulating the binding of structurally diverse ligands to the estrogen receptor. *J. Med. Chem.* **2005**, *48*, 3700−3.

CI7002945