

Structure-Based Validation of the 3D-QSAR Technique MaP

Nikolaus Stiefl and Knut Baumann*

Department of Pharmacy and Food Chemistry, University of Wuerzburg, Am Hubland,
D 97074 Wuerzburg, Germany

Received October 19, 2004

For three target proteins with different binding pocket characteristics (size and shape, hydrophobicity, hydrogen-bonding) a structure-based validation of the translationally and rotationally invariant 3D-QSAR technique MaP is performed (MaP: Mapping Property distributions of molecular surfaces). The structure-based validation procedure comprises two steps: first, QSAR models are derived without using the information of the target protein. Second, the models are back-projected into the crystal structure of the binding pockets and interpreted. It is demonstrated that MaP is able to identify characteristics important for ligand binding in the cases studied here. Moreover, it is demonstrated that MaP is a versatile 3D-QSAR technique since good, predictive models could be obtained for all three data sets showing distinct characteristics.

INTRODUCTION

Recently developed QSAR techniques not only tend to be mere numerical descriptors but also rather efficiently encode the relevant molecular information and simultaneously optimize the interpretability of the descriptor.^{1–4} For most of these methods, interpretability is achieved by some sort of visualization procedure which enables the user to graphically analyze the obtained models. That way, properties that were identified as relevant by the QSAR technique can be exploited to propose novel structures. Unfortunately, the structure descriptors used do not necessarily encode the relevant properties for a specific ligand–receptor interaction but rather describe the differences in the data under scrutiny. Therefore, the models obtained do not give a blueprint of the receptor binding pocket. For example, if a specific structural backbone is important for biological activity but constant for all compounds in the data set, it cannot be identified as a structural requirement.

However, if the variation in biological activity is a consequence of the structural variation in the data, the user needs to be confident that the QSAR technique applied is able to identify the properties relevant for ligand–receptor interactions. In cases where structural information about the biological target is available, receptor-based QSAR techniques can be applied^{5–7} instead of computationally more expensive techniques.^{8–11} Since this prerequisite is often not available, a thorough validation of a novel 3D-QSAR technique with respect to its ability of identifying the relevant interactions is desirable. For instance, with methods that allow an easy visualization of the results, such a validation could comprise the back-projection of the target-independent QSAR results into the target's binding pocket.

In this contribution, such a structure-based validation of the translationally and rotationally invariant 3D-QSAR technique MaP (Mapping Property Distributions of Molecular

Surfaces)² is presented. MaP uses simple surface properties and a straightforward mathematical transformation to intuitively describe the interactions between ligand and receptor at their molecular interface. To thoroughly validate MaP's ability to identify relevant ligand–receptor interactions three different data sets were investigated. The data sets were chosen based on the differences of their relevant ligand–receptor interaction properties (electrostatic, hydrophobic, or hydrogen-bonding interactions) and on the availability of information about the target protein's binding pocket. Comparisons to the originally published QSAR models are given; however, the main focus of this contribution is the interpretation of the obtained results and with it the validation of the methodology.

MATERIALS AND METHODS

Data Sets. In total three data sets were studied. The complete set of structures and the corresponding activity information can be found in the Supporting Information. For the sake of brevity, in Tables 1–3 only the different backbones as well as the specific structures mentioned in the results section are listed.

The first data set is comprised of 49 inhibitors of the enzyme acetylcholinesterase (AChE) for which a structure-based 3D-QSAR model was developed using a docking – GRID/GOLPE approach.¹² In the original publication, 7 of the 49 compounds were synthesized based on the predictions made of the original data set of 42 compounds and served therefore as an independent test data set. Centrally active AChE inhibitors are promising candidates in the treatment of Alzheimer's disease.¹³ Today, a comprehensive knowledge about the relevant structure–activity relationships (SARs) is readily available. These were derived from various crystal structures with different ligands^{14–19} as well as different docking^{20,21} and molecular dynamics studies.²² Apart from general features such as a very long and narrow hydrophobic gorge leading to the enzymatic center, more specific interactions such as π – π as well as π –cationic interactions with the amino acids Trp84, Trp279, Phe330, and Phe331 are well

* Corresponding author phone: ++49 931 888-5473; fax: ++ 49 931 888-5494; e-mail: knut.baumann@mail.uni-wuerzburg.de. Corresponding author address: Department of Pharmacy, University of Wuerzburg, Am Hubland, D 97074 Wuerzburg, Germany.

described. These amino acids are also identified as relevant by MaP, which will be discussed in more detail later.

Recently, Muegge noted that well-defined and deeply buried binding pockets are particularly suited for docking and receptor-based QSAR studies.²³ Since these characteristics apply to the binding pocket of AChE, this enzyme is a popular target for receptor-based QSAR studies.^{12,24–26} However, it should be recalled that the model building step here does not include information about the enzyme structure. The target structure is only used for interpreting the resulting models. Biological activity data as well as information about the inhibitor structures and the data set composition were taken as compiled in ref 12.

The second data set comprised 44 nonnucleosidic inhibitors (NNI) of human immunodeficiency virus 1 reverse transcriptase (HIV-1 RT) for which a 3D-QSAR model was developed using a CoMFA approach that was further validated with docking experiments.²⁷ The multifunctional HIV-1 RT enzyme plays an important role in the replication of HIV and is therefore a major therapeutic target in the treatment of HIV.²⁸ RT is a dimeric protein with a polymerase site and an allosteric binding site. The investigated tetrahydroimidazo(4,5,1-JK)(1,4)-benzodiazepine-2(1*H*)-ones and -thiones (TIBO) bind to the allosteric site mainly with hydrophobic interactions and few hydrogen-bonding interactions.²⁷ These SARs are well supported by different high-resolution crystallographic coordinates for HIV-1 RT and TIBO derivatives.^{29,30}

Biological activity data as well as information about the inhibitor structures and the data set composition were taken as compiled in ref 27.

The third data set comprised 46 inhibitors of the poly-(ADP-ribose)polymerase-1 (PARP-1) with a benzamide moiety as the most prominent structural feature.³¹ PARP-1 is a chromatin-bound nuclear enzyme related to genomic repairs³² and the protection from neuronal lesions after cerebral ischemia.³³ For this data set another structure-based 3D-QSAR approach (docking – superpositioning – RECEPTOR^{34,35}) was applied to elucidate the topography of the active site of PARP-1. An analysis of the SARs available for the benzamide-type inhibitors led to the proposition of a pharmacophore that includes an *anti*-aryl-amide moiety and at least one aromatic ring. If the latter is substituted in position 3 biological activity is increased. The opposite is true for substituents in positions 2, 4, and 5. This pharmacophore is supported by hydrogen-bond interactions between the amide moiety with backbone atoms of Gly863 and π – π interactions between the aromatic moiety and Tyr907 found in different ligand-PARP-1 crystal structures.^{36–39} Biological activity data as well as information about the inhibitor structures and the data set composition were taken as compiled in ref 31.

Geometry Optimization. Even though MaP's sensitivity to minor conformational changes is rather low,⁴⁰ sensible conformations should be used for every compound under study. To obtain such conformations for all compounds a single reference molecule was built for each data set (see the respective tables in the Supporting Information). The criteria for choosing the reference compounds were a high biological activity, a good representation of the common structural skeleton, and a comparatively large size. The reference structures were 2D-3D converted within Alchemy

2000,⁴¹ and the AChE inhibitors were protonated at the basic piperidine nitrogen. Next the structures were geometry optimized within SYBYL⁴² using the MMFF94 force field (500 iterations, steepest descent, $\epsilon = 4$). Comparison of the resulting conformations with cocrystallized ligands of the respective target structures supported this procedure, since no significant differences could be identified. Using the template structures, all other compounds were built employing two different protocols. For each data set the first conformation ensemble was generated either by adding or removing the respective functional groups from the template structure followed by a geometry optimization (settings as above) where the common skeleton was kept fixed. The second set of conformations was generated for all compounds using FlexS⁴³ with its default settings. The same template molecules as above were used.

Theory of the MaP Descriptor. A comprehensive description of the technique can be found in previous publications.^{2,44} Briefly, MaP is a translationally and rotationally invariant 3D-QSAR method, i.e., the descriptor is alignment-independent. The basic idea of MaP is that ligands and receptors interact noncovalently at the interface of their respective molecular surfaces. In particular, steric, electrostatic, and hydrophobic properties play an important role for the affinity between the two molecules. Consequently, the MaP descriptor characterizes the molecular surface of a potentially active compound in terms of the distribution of its surface properties.

A three-step procedure is used to compute the descriptor: First, an approximation to the molecular surface is computed, where the surface is represented by equally distributed surface points. The default grid resolution to determine this discretized surface is 0.8 Å leading to approximately 1.6 surface points per Å². Next, each surface point is assigned one out of five property categories. The physicochemical properties currently used to characterize the molecular surface are H-bond donor (D) and H-bond acceptor (A) as well as hydrophobicity (H: hydrophilic, Lw: weakly hydrophobic, Ls: strongly hydrophobic). These properties are assigned to the surface points according to localized atomic properties of the molecule. The algorithms to determine these property categories can be found in ref 2. In the third step the property distribution of the surface points is mathematically encoded by a count statistics of potential 2-point pharmacophores (P2PP). A P2PP is characterized by the two properties of the involved surface points and their respective distance. Distance bins with a resolution of 1 Å are created for all property–property combinations starting at 0.5 Å (i.e. the borders are 0.5 Å, 1.5 Å, 2.5 Å, ...). Finally, a loop over all surface point pairs is initiated, and the occurrence of each specific P2PP (i.e. a particular property–property–distance combination) is counted and stored in a vector **v**. This count statistic implicitly encodes size and shape of the molecule by the distribution of distances.² This three-step procedure is depicted schematically in Figure 1. It is important to note, that each P2PP can be back-projected into the original molecular space. That way, a QSAR model consisting of several P2PPs can easily be visualized for interpretation.

Every single bin of the MaP-vector **v** is unambiguously characterized by the property–property–distance combination it encodes. Therefore, an abbreviation, that uses a character for each surface point property and a number for

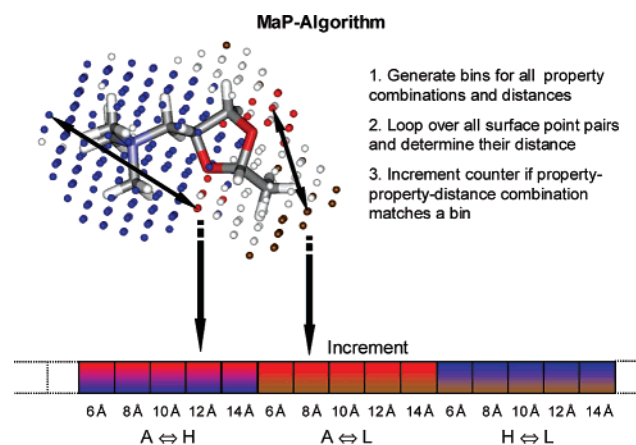


Figure 1. Flowchart for the calculation of the MaP-vector \mathbf{v} (A: red, H: blue, L: brown).

the mean intersurface-point distance will be used throughout the manuscript to simplify interpretation. An example would be a variable encoding a hydrogen-bond acceptor surface point (A) in a distance of 10 Å to a hydrophilic surface point (H), which will be written as AH₁₀.

Chemometric Methods. The data analysis was performed according to a standard protocol described in previous publications.^{2,44,45} For a data set of m molecules the MaP-vectors of dimension n are stacked to a matrix of dimensions $m \times n$. All variables (i.e. P2PPs) which are constant (variance equal to zero) are excluded. Principal component regression (PCR) is used by default as the regression technique. Since not all variables of the descriptor are related to the property under study, the most informative variables (MIV) are selected from the pool of all variables. For variable selection, MaP employs a stepwise regression,⁴⁶ i.e., a method that either adds or removes a variable from the current model in order to maximally improve the objective function (here: the cross-validated prediction error). The search is stopped in the first local optimum. To avoid that only weakly significant variables are introduced in the model, an early stopping rule (change in cross-validated root-mean-squared error of prediction < 3%) was introduced into the selection procedure. The advantages of using this early stopping rule are a reduced risk of chance correlation in variable selection and a better interpretability.

To avoid overfitting the stringent leave-multiple-out cross-validation (LMO-CV) procedure is employed as an objective function for variable selection.⁴⁷ The number of cross-validation runs (B) was set to three times the number of objects in the data set ($B = 3 \cdot m$) for all data sets. A reasonable default value for the percentage level of objects left out is 50.^{48,49} However, if this high percentage led to underfitting during the selection procedure, this value is lowered to 40 (see HIV-1 RT inhibitors).

Generally, models obtained by stepwise variable selection are prone to chance correlation. To check for the risk of chance correlations,^{50,51} validation of the variable selection results was performed in two ways. First, a permutation test based on the repetitive randomisation of the response vector was employed (y-scrambling) and second, the biological activity of an independent test set was predicted. The partition of the data sets into training and test sets can be found in the Supporting Information. Two different partitions were

used: the originally published test set and a larger representative test set (approximately 33%) generated according to an experimental design strategy (CADEX-split).⁵² This second test set is used because the random fluctuation of the prediction error (calculated as its relative standard deviation) is very high for small test data sets.⁵³ Therefore, the comparatively small original test data sets were augmented by larger ones. All models are summarized by measures characterizing the data fit (squared multiple correlation coefficient: R^2 , root-mean-squared error of calibration: RMSEC), the cross-validated performance (R^2_{CV} , cross-validated root-mean-squared error of prediction: RMSEP_{CV}), and the test set predictivity (R^2_{Test} , RMSEP_{Test}). The exact equations for computing these figures of merit comply with chemometric standards and can be found in previous publications.^{2,44,48}

Parameter Settings. Except for the PARP-1 inhibitors, MaP was applied using its default parameter settings (see Table 4). For the PARP-1 data set, it was known that a specific H-bond donor interaction is highly relevant for biological activity. Therefore, an additional differentiation between weak and strong H-bond donor groups was added. Therefore, the AM1 partial atomic charge of the hydrogen atom connected to the respective heteroatom of the donor group was incorporated as an additional cutoff value into the default parameter set.^{44,54} This cutoff was set to 0.19 (≤ 0.19 weak donor, Dw; > 0.19 strong donor, Ds).

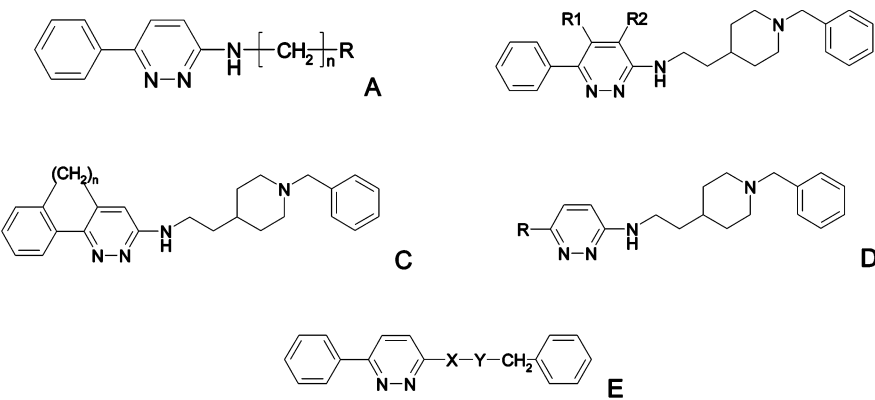
RESULTS AND DISCUSSION

The goal of this study was not to compare the MaP method to the established techniques already applied to the data sets under scrutiny but to study whether MaP is able to deduce the relevant and meaningful interactions between ligand and receptor. Consequently, the major part of this discussion will focus on the interpretation of the derived models with only numerical comparisons to the reference results.

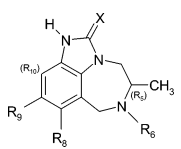
ACHe Inhibitors. Using the default settings (properties: {A, D, H, Lw, Ls}) the following characteristics were obtained for the data. The largest distances between two surface point pairs were 23.1 Å and 24.1 Å (\Rightarrow number of distance bins: $n_{distbin} = 23$ and $n_{distbin} = 24$) for the conformations obtained manually and with the help of FlexS, respectively. Accordingly, the size of the MaP vector \mathbf{v} was 345 and 360 for each compound of the data set. For all compounds, 46 and 47 out of these variables showed a constant variance and were deleted resulting in a final data matrix of 49×299 and 49×313 for the entire data set.

In a preliminary analysis of the data set (data set split¹² training/test data: 42/7) compound **7** was identified as outlier in \mathbf{Y} -space independent of the conformations used (leave-one-out cross-validated residual with and without variable selection $> 3 \cdot \text{RMSEP}_{CV-1}$). Therefore, this compound was removed from the data set for the rest of the analysis. Structural reasons for the different biological activity will be discussed below.

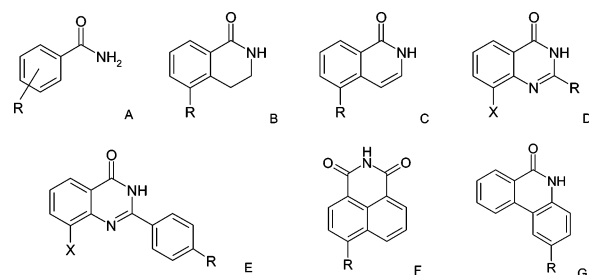
For both conformation ensembles the derived models showed equally good results with only numerical differences in model quality (see Table 5). Additionally, both models allowed a similar interpretation of the data set under scrutiny. These two facts emphasize the robustness of the MaP

Table 1. Structures and Biological Activity Data of Selected AChE Inhibitors


No	<i>n</i>	X	Y	R	Skeleton	pIC ₅₀
7	5	-	-		A	5.00
8	2	-	-		A	3.77
21	3	-	-	-	C	7.66
39	-	NHCH ₂ CH ₂		-	E	4.62
41	-	NHCH ₂ CO		-	E	3.92

Table 2. Structures and Biological Activity Data of Selected HIV-1 RT Inhibitors


no.	X	R ₆	R ₈	R ₉	R ₅ /R ₁₀	pIC ₅₀
27	S	CH ₂ CH=C(CH ₃) ₂	CN	H		7.25
28	S	CH ₂ CH=C(CH ₃) ₂	CH ₃	H		7.85
29	S	CH ₂ CH=C(CH ₃) ₂	Cl	H		8.34

Table 3. Structures and Biological Activity Data of Selected PARP-1 Inhibitors


no.	R	X	skeleton	pIC ₅₀
3	4-F		A	3.70
17	CH ₃		B	6.36

procedure against minor conformational changes. However, for the sake of comprehensibility only the results obtained for the FlexS conformations are discussed in the following.

PCR with leave-one-out cross-validation (LOO-CV) for selecting the number of latent variables (*q*) was employed

Table 4. Default Settings of the MaP-Procedure

parameter	value
properties	A, D, H, Lw, Ls
grid spacing for equally distributed surface points	0.8 Å
proximity distance $d_{\text{cut-off}}$	2 Å
Δd	2 Å
cutoff values for hydrophilic/hydrophobic surface points	
hydrophilic	property < 0
weakly hydrophobic	$0 \leq \text{property} < 0.12$
strongly hydrophobic	property ≥ 0.12
resolution (res) of the radial distribution function (distance bins)	1 Å

to build a model for the remaining 41 compounds. For the full model (i.e. no variable selection), an optimal dimensionality of $q = 7$ and a leave-one-out cross-validated squared multiple correlation coefficient $R^2_{\text{CV-1}}$ of 0.74 was obtained. For the best model obtained with leave-multiple-out cross-validation (LMO-CV) the optimal dimensionality was $q = 7$ with a $R^2_{\text{CV-50\%}}$ -value of 0.66. This model was obtained with a leave-50%-out cross-validation where the data set was randomly split 123 times ($3 \cdot m$) into construction and validation data. Using PLS instead of PCR did not change the statistical figures of merit significantly ($R^2_{\text{CV-1}} = 0.76$, $q = 4$; $R^2_{\text{CV-50\%}} = 0.68$, $q = 4$; data not shown in Table 5). The same was also true for the following two data sets. Therefore, PLS models are omitted here and will no longer be discussed in this manuscript.

Even though model quality for the full model was already quite good, a variable selection was employed to identify the most significant variables (MIV) for interpretation purposes. Objective function for simultaneously selecting variables and the optimal number of latent variables was LMO-CV with 50% of the data left out. The QSAR equation

Table 5. Results of the MaP-Procedure and the Field-Based Techniques for the AChE Data Set

RT/CE ^a	RMSEP _{CV-1} ^b	R ² _{CV-1} ^c	RMSEP _{CV-50%} ^d	R ² _{CV-50%} ^d	RMSEC ^e	R ^{2f}	RMSEP _{Test} ^g	R ² _{Test} ^g	m/m _{Test} ^h	n ⁱ	q ^j
MaP/M ^k	0.71	0.73	0.83	0.64	0.64	0.83	0.80	0.77	41/7	299	7
MaP/FS ^l	0.71	0.74	0.80	0.66	0.57	0.86	0.83	0.76	41/7	313	7
MaP-VS ^m /M	0.56	0.84	0.60	0.81	0.55	0.85	0.93	0.70	41/7	4	3
MaP-VS/FS	0.54	0.85	0.56	0.84	0.52	0.87	0.79	0.78	41/7	4	2
MaP-VS-Cdx ⁿ /M	0.53	0.85	0.57	0.82	0.49	0.90	0.84	0.67	32/16	4	4
MaP-VS-Cdx/FS	0.60	0.81	0.64	0.79	0.58	0.84	0.61	0.83	32/16	4	2
COMFA/M	0.59	0.81			0.35	0.94	0.75	0.80	42/7		3
COMFA/FS	0.59	0.81			0.25	0.97	1.06	0.60	42/7		3
COMSIA/M	0.56	0.83			0.37	0.93	0.79	0.78	42/7		3
COMSIA/FS	0.56	0.83			0.37	0.93	0.94	0.69	42/7		2

^a RT/CE: regression technique/conformation ensemble. ^b RMSEP_{CV-1}: leave-one-out cross-validated root-mean-squared error of prediction. ^c R²_{CV-1}: leave-one-out cross-validated coefficient of determination. ^d Same as *b* and *c* for leave-50%-out cross-validation. ^e RMSEC: root-mean-squared error of calibration. ^f R²: coefficient of determination. ^g Same as *b* and *c* for test set prediction. ^h *m*: number of objects. ⁱ *n*: number of variables. ^j *q*: number of latent variables. ^k M: manually obtained conformation ensemble. ^l FS: conformation ensemble obtained with FlexS. ^m VS: Variable selection with tabu search. ⁿ Cdx: CADEX split into training and test set.

obtained for the selected MIVs is

$$\hat{y} = 5.77 + 0.0019 \cdot \text{HLw}_3 + 0.0006 \cdot \text{LwLs}_{10} + 0.0013 \cdot \text{AH}_{10} - 0.0011 \cdot \text{LsLs}_{11}$$

$$R^2_{\text{CV-50\%}} = 0.84, \text{RMSEP}_{\text{CV-50\%}} = 0.56,$$

$$R^2 = 0.87, \text{RMSEC} = 0.52, m = 41, q = 2$$

Here, \hat{y} is the estimated biological activity, *m* is the number of objects of the training data, *q* is the number of principal components, and the data were mean-centered. Again, model quality of the derived model is very good, this time with a lower model dimensionality of *q* = 2. The variables of the QSAR equation are given as property–property–distance triplets, describing the underlying properties in an intuitive way. Variable HLw₃, encodes a hydrophilic (H) and a weak hydrophobic surface patch (Lw) in a mean distance of 3 Å, LwLs₁₀ a weak hydrophobic (Lw) and a strong hydrophobic surface patch (Ls) in a mean distance of 10 Å, AH₁₀ an acceptor (A) and a hydrophilic surface patch (H) in a mean distance of 10 Å, and LsLs₁₁ two strong hydrophobic surface patches (Ls) in a mean distance of 11 Å.

The results of the original study¹² are given here for the sake of completeness. After variable selection with a SRD/FFD (smart region definition/fractional factorial design) procedure, the following figures of merit were obtained: R²_{CV-50%} = 0.91, RMSEP_{CV-50%} = 0.41, R² = 0.99, RMSEC = 0.14, *m* = 42, *q* = 4, RMSEP_{Test} = 0.44. It should be noted that in this case *m* equals 42, i.e., compound **7** did not show up as an outlier. Given the latter, a direct comparison between the two models is not possible. However, it can be seen that both models give very good results with the original one being slightly superior. However, many more model selection steps were employed to arrive at this model.

For the interpretation it should be kept in mind that owing to mean-centering of the data, variables exhibiting a positive regression coefficient do not necessarily encode a positive contribution to biological activity and vice versa. For example if a compound is lacking a specific feature which is negatively correlated with biological activity (negative sign of the regression coefficient), its entries for this variable will be negative resulting in an enhanced predicted biological activity.

For this data set interpretation of the selected MIVs is straightforward and gives a good insight into the structural

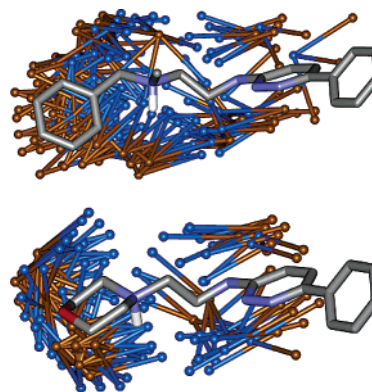


Figure 2. Back-projection of variable HLw₃ onto compounds **8** and **42**. Each line connecting two points shows a P2PP (H: blue, Lw: brown). This variable describes the presence of a protonated (quaternary) nitrogen atom and its surrounding environment. If either the distance to the aromatic amine is very short (**8**, top) or the lateral ring system is missing (**42**, bottom), the number of neighboring weakly hydrophobic surface points is small. This results in a low predicted biological activity.

necessities for a high biological activity. The QSAR equation reveals that the hydrophilic as well as the strongly hydrophobic surface properties are very important for biological activity. The largest hydrophilic surface area originates from the protonated basic aliphatic or piperidine nitrogen atom found in most structures (see Figure 1 in the Supporting Information). Structures that are not positively ionizable do not show a large hydrophilic surface area. As a result, they show small entries for the variables comprising the hydrophilic attribute (H). Since both regression coefficients containing the H-attribute are positively correlated with biological activity, these compounds are predicted to be weakly active (e.g. **41**). Moreover, variable HLw₃ encodes the environment around the protonated nitrogen atom. If other hydrophilic substructures are in close proximity to this strongly hydrophilic group, the surrounding weakly hydrophobic environment (Lw) is drastically reduced in size which reduces the magnitude of variable HLw₃. Such a spatial arrangement is found in compounds with too short a C-chain between the protonated nitrogen and the secondary amine of the aminopyridazine substructure (e.g. **8**, see Figure 2). Here, variable HLw₃ takes a value of 720, while the mean value is 1150. Another possibility for a reduced hydrophobic environment close to the hydrophilic surface area of the quaternary nitrogen is the lack of a lateral ring system as in

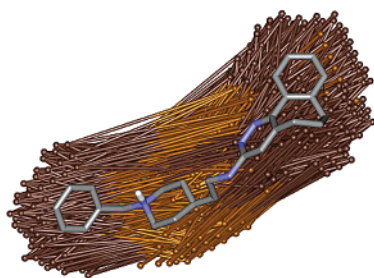


Figure 3. Back-projection of variable $LwLs_{10}$ onto compound **21** (Lw: light brown, Ls: dark brown) This variable describes extended lateral aromatic moieties and their distance to the weakly hydrophobic surface area surrounding the protonated nitrogen group as essential for biological activity. Moreover, this variable encodes global size and shape of the molecules.

minaprine (**42**) (see Figure 2). This reduces HLW_3 even further to about 390.

The main strongly hydrophobic surface areas originate at the lateral aromatic moieties of the compounds. To give a better understanding of the relevant QSARs identified by the model, two variables are described in more detail, whereas the other two are discussed in the Supporting Information. In contrast to variable HLW_3 variable $LwLs_{10}$ describes the position and the size of the lateral aromatic systems. The latter are encoded by the strong hydrophobic surface patches (Ls), whereas the weak hydrophobic surface patches stem from the same areas as in variable HLW_3 (i.e. aliphatic carbons close to the protonated nitrogen). Accordingly, a short C-chain between the central nitrogen atoms (Lw) as well as missing or small lateral aromatic moieties (Ls) result in reduced entries for this variable (i.e. this variable partly encodes the information of variable HLW_3). Put differently, compounds that comprise extended lateral aromatic moieties separated by a specific minimum distance show an increased predicted biological activity. One such compound fulfilling these requirements for a high biological activity is given in Figure 3.

The third variable found in the QSAR equation (AH_{10}) further emphasizes the importance of the location of the protonated nitrogen (H) with respect to the aminopyridazine (amine: A-attribute), while the fourth variable ($LsLs_{11}$) penalizes small-sized molecules.

To complete the picture a CoMFA as well as a CoMSIA study was carried out for both conformation ensembles (see Table 5). It can be seen that both field-based methods perform slightly worse on the FlexS conformations corroborating their sensitivity toward different alignments.

Validation of the derived model was completed with a permutation test on the training data ($m = 41$). The results are summarized as quantiles of the distribution of $R^2_{CV-50\%}$ -values resulting from 500 variable selections on the scrambled data (scrambling from scratch) which are referred to as $R^2_{CV-50\%,PT}$. The median $R^2_{CV-50\%,PT}$ was 0.08, the 95%-quantile 0.34, and the maximum value 0.53, respectively. Since all $R^2_{CV-50\%,PT}$ -values are smaller than the real $R^2_{CV-50\%}$, it can be concluded that the probability of chance correlation is quite low ($p < 0.002$).

Additionally, the training and test data set were merged and then split into a representative training set (32 objects) and a test set (16 objects, CADEX). Test set prediction for this as well as for the originally published test data set (7 objects, ORIG) was good. (CADEX: $R^2_{Test} = 0.83$; ORIG:

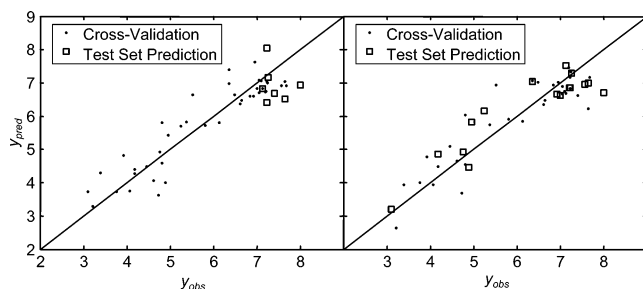


Figure 4. Observed activities plotted against the leave-one-out cross-validated and the predicted biological activities. Left: original test set. Right: CADEX split into training and test set.

$R^2_{Test} = 0.78$; see Table 5.) Interestingly, the biological activity of most original test set compounds was underestimated (see Figure 4). However, when visualizing the MIVs for these compounds this becomes clear, since none of the compounds of the training data set shows similar extended aromatic ring systems (extrapolation).

To illustrate the relevance of the identified MIVs, they were back-projected into the binding pocket of the donepezil–AChE complex (PDB entry: 1EVE). This projection gives a very good impression of the relevance of the derived model (see Figure 5). Variables $LwLs_{10}$, $LsLs_{11}$, and HLW_3 describe the elongated shape of the ligands and the π – π interactions between the aromatic moieties and Trp84 as well as Trp279. Additionally, variables HLW_3 and AH_{10} describe the position of the quaternary nitrogen atom of the ligands and with it the potential for cationic– π interactions with Phe330 and Phe331. If this charge is displaced, (e.g. in **39**) these interactions cannot form. Owing to the property coding of MaP, the cationic part of the nitrogen atom is encoded as hydrophilic surface property. This could easily be changed by changing the property assignment. However, since the interpretation of the model was straightforward, this was not necessary here.

Based on the interpretation of the model the behavior of the outlier in Y-space (**7**) is easy to grasp. Compared to other highly active compounds of the data sets, compound **7** shows restricted flexibility between the lateral phenyl and the piperidine ring system. Investigation of the donepezil–AChE complex substantiates the detrimental influence of this rigidity. The lateral phenyl ring system of donepezil is oriented parallel to Trp84 which is not possible for the conformationally restricted compound **7**.

HIV-1 RT Inhibitors. With default settings (properties: {A, D, H, Lw, Ls}) the following characteristics were obtained for the data. The largest distances between two surface point pairs were 15.3 Å and 14.8 Å ($\Rightarrow n_{distbin} = 15$ each) for the conformations obtained manually and with the help of FlexS, respectively. Accordingly, the size of the MaP vector \mathbf{v} was 225 for each compound of the data set independent of the chosen conformation. For all compounds, 14 and 15 out of these variables showed a constant variance and were deleted resulting in a final data matrix of 44×211 and 44×210 for the entire data set, respectively.

As the derived models for both conformation ensembles showed very similar model quality (see Table 6), only the results for the manually obtained conformations are discussed.

PCR with LOO-CV for selecting the number of latent variables was employed to build a model for the full data

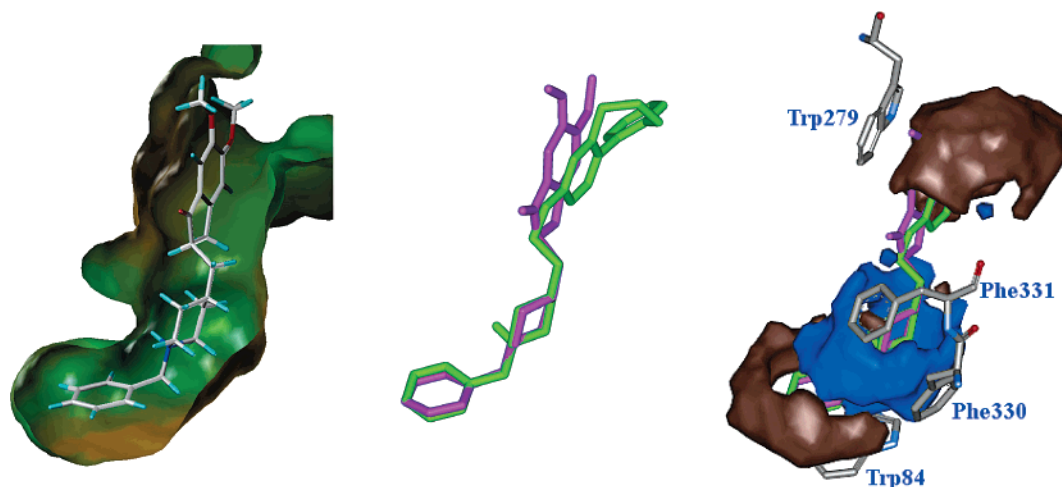


Figure 5. Left: crystal structure of the donepezil–AChE complex. The inhibitor fits nicely into the binding pocket. The surface of the enzyme is color-coded with continuous values according to its hydrophilic/hydrophobic properties (blue: hydrophilic, brown: hydrophobic). Middle: overlay of donepezil (pink) as found in its ligand–enzyme complex and **21** (green). Both conformations correspond very well. Right: strongly hydrophobic (dark brown) and hydrophilic (blue) surface parts of variables LwLs₁₀ and HLw₃, donepezil (pink), and **21** (green) back-projected into the binding pocket of AChE (represented by the amino acids Trp84, Phe330, Phe331, and Trp279). The identified MIVs characterize the known SARs very well.

Table 6. Results of the MaP-Procedure for the HIV-1 Data Set^a

RT/CE	RMSEP _{CV-1}	R^2_{CV-1}	RMSEP _{CV-50%}	$R^2_{CV-50\%}$	RMSEC	R^2	RMSEP _{Test}	R^2_{Test}	m/m_{Test}	n	q
MaP/M	0.79	0.63	0.81	0.62	0.77	0.67	-	-	44/-	211	2
MaP/FS	0.81	0.62	0.82	0.60	0.79	0.66	-	-	44/-	210	2
MaP-VS/M	0.82	0.64	0.84	0.63	0.81	0.67	0.37	0.79	37/7	1	1
MaP-VS/FS	0.85	0.62	0.88	0.59	0.84	0.65	0.46	0.68	37/7	1	1
MaP-VS-Cdx/M	0.66	0.72	0.68	0.69	0.63	0.75	0.94	0.59	29/15	3	1
MaP-VS-Cdx/FS	0.70	0.68	0.72	0.62	0.68	0.71	1.00	0.52	29/15	2	1

^a For comments and abbreviations see Table 5.

set. The model obtained showed an optimal dimensionality of $q = 2$ and an R^2_{CV-1} -value of 0.63 without variable selection. The model quality was virtually identical for LMO-CV ($q = 2$, $R^2_{CV-50\%} = 0.62$).

Interpretation of the model was done on the variable selected model. A single variable model was obtained with LMO-CV:

$$\hat{y} = 6.43 + 0.00085 \cdot \text{LwLs}_7$$

$$R^2_{CV-50\%} = 0.63, \text{RMSEP}_{CV-50\%} = 0.84,$$

$$R^2 = 0.67, \text{RMSEC} = 0.81, m = 37, q = 1$$

The identified MIV encodes a weakly hydrophobic (Lw) and a strongly hydrophobic surface patch in a mean distance of 7 Å. Comparing the originally published model²⁷ ($R^2_{CV-1} = 0.70$, $R^2 = 0.90$, $R^2_{Test} = 0.96$, $\text{RMSEP}_{Test} = 0.15$, $q = 4$) to the MaP model shows that model quality of both models is good, while the test set prediction for the original model is better (MaP: $R^2_{Test} = 0.79$). It should be noted, that in the original publication, no variable selection was applied and that compound **27** (enumeration as in ref 27) shows up twice (**27** and **31**, i.e., $m = 45$) with two different activity values.

Obviously, the different hydrophobic surface properties are the key factor to enhanced biological activity. With respect to this property three structural factors play an important role: the heteroatom in position X, the size of the substituent in position R₆, and the substituents R₈–R₁₀ of the aromatic ring system (see Table 2). LwLs₇ describes the distribution of these hydrophobic surface areas where for

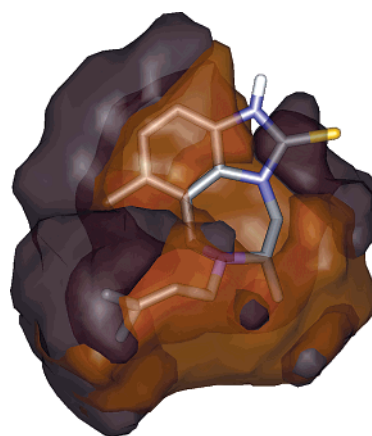


Figure 6. Back-projection of weak (Lw, light brown) and strong hydrophobic surface patches (Ls, dark brown) onto the highly active compound **28**.

the more active compounds the weak hydrophobic surface property is found along the aliphatic ring system as well as the methyl substituent in position R₅, whereas the strong hydrophobic surface extends mainly along the aromatic moiety, between the sulfur atom and the neighboring methyl group in position R₅, and the substituent in position R₆ (see Figure 6).

If either of these surface patches is reduced or has a changed property, the predicted biological activity will decrease. Within this data set three different types of surface modifications are found. Briefly, if either position R₆ is substituted with a too short and too weakly hydrophobic side

chain, heteroatom X is an oxygen atom instead of a sulfur atom or if positions R₈ and R₉ are substituted with an hydrophilic or too large a group, the entries in LwLs₇ are reduced and with it the predicted biological activity. In contrast to that, small hydrophobic substituents in position R₈ and R₉ result in higher entries in LwLs₇ and therefore in a higher predicted biological activity.

Validation was supplemented by permutation testing on the training data ($m = 37$). The results were as follows: the median $R^2_{CV-50\%,PT}$ was 0.03, the 95%-quantile 0.28, and the maximum value 0.50, respectively. Hence, the probability of chance correlation is low ($p < 0.002$). In addition to y-scrambling, the training and test data set were merged and then split into a representative training set (29 objects) and a test set (15 objects, CADEX). Test set prediction for this as well as for the originally published test data set (7 objects, ORIG) was satisfactory (CADEX: $R^2_{Test} = 0.59$; ORIG: $R^2_{Test} = 0.79$; see Table 6). For both data set partitions the MIV selected was LwLs₇ which corroborates the relevance and robustness of the models found.

As for AChE, the identified MIV was back-projected into the binding pocket of an HIV-1 RT complexed with a nonnucleoside inhibitor. For this purpose, the complex of RT/TIBO R86183 (PDB⁵⁵ code: 1HNV) was used since the cocrystallized ligand is also found in the studied data set (29). That way, the original binding mode can directly be utilized for the structure-based validation of the derived model. However, the conformation of the side chain in position R₆ differs slightly between the cocrystallized ligand and the one used in the QSAR study (see Figure 2 in the Supporting Information). This is due to the fact that the information about the conformation of the cocrystallized ligand was deliberately not used in the derivation of the conformations for the QSAR analysis. To easily interpret the back-projected surface patches the relevant MIV was computed and then back-projected into the binding pocket for the cocrystallized ligand.

Since the strongly hydrophobic surface patches are the more important part of LwLs₇, only those were visualized for the sake of interpretability. Figure 7 nicely highlights the strongly hydrophobic interactions of the 3,3-dimethylallyl side chain in R₆ with the pocket defined by Pro95, Tyr181, Tyr188, and Trp229, respectively, the hydrophobic interactions of the substituents in positions R₈ and R₉ with the amino acids Tyr188, Phe227, and Leu234, respectively, and the hydrophobic interactions of the surface located between the sulfur atom and the methyl group in position R₅ with Val179.

The structure–activity relationships (SAR) derived in ref 27 comprise, among others, the potential of the compounds to adopt a butterfly-like conformation and a specific location of lipophilic and electron rich-groups. These SARs and those identified by MaP correspond well.

PARP-1 Inhibitors. For the aforementioned reasons an additional property was used for this data set. Using the H-bond donor differentiation⁴⁴ of MaP (properties: {A, Dw, Ds, H, Lw, Ls}) the following characteristics were obtained for the data. The largest distances between two surface point pairs were 18.7 Å and 16.9 Å ($\Rightarrow n_{distbin} = 19$ and 17, respectively) for the conformations obtained manually and with the help of FlexS, respectively. Excluding constant variables yielded matrices of sizes 46 × 327 (manually) and 46 × 297 (FlexS). As model quality and interpretability was

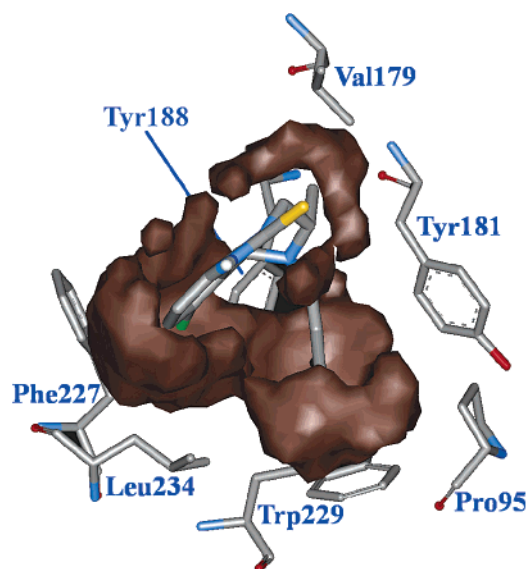


Figure 7. Back-projection of the strong hydrophobic surface points of LwLs₇. It can easily be seen that the major interactions necessary for a high biological activity are well described by this MIV (see text).

similar for both data sets (see Table 7), only the results for the manually obtained conformations are discussed.

Full model quality was unsatisfactory here (see Table 7). Variable selection on the data set using the extended property set yielded the following QSAR equation:

$$\hat{y} = 5.38 - 0.0100 \cdot \text{DsDs}_4 + 0.0052 \cdot \text{DsH}_6 + 0.0026 \cdot \text{LwLs}_3 - 0.0021 \cdot \text{LsLs}_{11}$$

$$R^2_{CV-50\%} = 0.70, \text{RMSEP}_{CV-50\%} = 0.63, \\ R^2 = 0.79, \text{RMSEC} = 0.56, m = 46, q = 4$$

Model quality is competitive with the originally published RECEPTOR approach.³¹ The latter uses five selected variables all of which are related to the enthalpy and entropy of ligand binding. In the RECEPTOR approach an R^2_{CV-1} of 0.56 and an R^2 of 0.66 was obtained for the full data set ($m = 46$) which was improved to $R^2_{CV-1} = 0.74$ and $R^2 = 0.80$ by eliminating four outliers ($m = 42$).

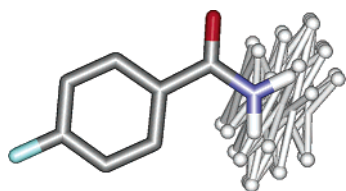
The interpretation of the selected MIVs for MaP is as follows: DsDs₄ is an indicator variable, penalizing compounds with amide functionality as opposed to lactam substructures (see Figure 8). Variable DsH₆ encodes the positive effect of a substitution with a strong H-bond donor group in position 8 of the ring system on biological activity. Finally, variables LwLs₃ and LsLs₁₁ describe mainly quinazoline-4-ones with large hydrophobic substituents in position 2. For these compounds, a methoxy instead of a methyl substitution in position X results in a decreased predicted biological activity, i.e., the latter two variables encode mainly the specific size and shape properties of these compounds.

Validation of the model was complemented by computing a permutation test on the entire data set ($m = 46$). The median $R^2_{CV-50\%,PT}$ was 0.01, the 95%-quantile 0.23, and the maximum value 0.42, respectively. Hence, the probability of chance correlation is low ($p < 0.002$).

Since the originally published test data set is very small (4 compounds³¹) and its derivation was not discussed in the original publication, this test data set was not used here.

Table 7. Results of the MaP-Procedure for the PARP-1 Data Set^a

RT/CE	RMSEP _{CV-1}	R ² _{CV-1}	RMSEP _{CV-50%}	R ² _{CV-50%}	RMSEC	R ²	RMSEP _{Test}	R ² _{Test}	m/m _{Test}	n	q
MaP/M	0.96	0.30	1.02	0.20	0.82	0.57	-	-	46/-	327	6
MaP/FS	0.87	0.42	1.00	0.23	0.75	0.65	-	-	46/-	297	6
MaP-VS/M	0.57	0.75	0.63	0.70	0.56	0.79	-	-	46/-	4	4
MaP-VS/FS	0.53	0.78	0.60	0.72	0.51	0.83	-	-	46/-	7	6
MaP-VS-Cdx/M	0.65	0.65	0.74	0.55	0.59	0.76	0.60	0.80	30/16	4	4
MaP-VS-Cdx/FS	0.62	0.69	0.67	0.63	0.57	0.78	0.59	0.81	30/16	4	4

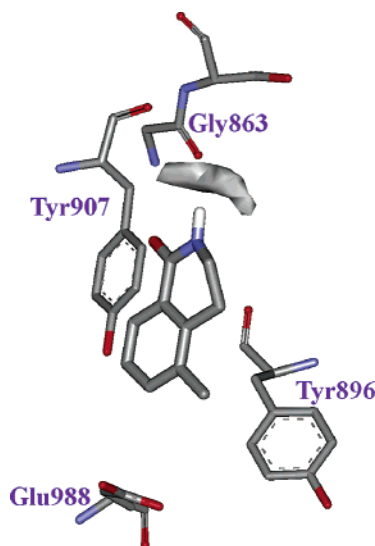
^a For comments and abbreviations see Table 5.**Figure 8.** Back-projection of variable DsDs₄ onto compound **3** (Ds: white). Almost only compounds with free amide substituents (low biological activity) show entries for this indicator-type variable.

Instead, the data set was split into a representative training set (30 objects) and a test set (16 objects) with the help of the CADEX algorithm (see Table 7). Test set prediction for this data set was good ($R^2_{\text{Test}} = 0.80$). Apart from minor distance deviations, the MIVs that were selected for this data set split (DsDs₄, LsLs₁₂, LwLs₃, DsH₅) were equivalent to those of the original model. This further substantiates the relevance of the models found.

As for the previous two data sets, the MIVs of the QSAR equation were back-projected into the binding pocket of PARP-1. After superimposition, no relevant variation in the binding pocket was observed among all available crystal structures (1PAX, 2PAX, 3PAX, 4PAX, 1EFY, 2PAW). Therefore, the complex with 3,4-dihydro-5-methyl-isoquinoline (compound **17**, PDB code: 1PAX) was used for interpretation.

Inspection of the strong H-bond donor surface patches of DsDs₄ and DsH₆ highlights that only one of the hydrogen atoms of the amide substructure can be part of the H-bond interactions between the ligand and the receptor (see Figure 9). Additionally, it can be seen that close to the aromatic ring system, a free carboxyl function is found (Glu988). A ring substitution in position 8 with a strong H-bond donor functionality (DsH₆) could then have two hypothetical outcomes. The first would be that the positively charged H-bond donor hydrogen atom could interact with the negative charge of the carboxyl function either directly (distance of the donor-hydrogen atom to either of the oxygen atoms is approximately 3 Å) or by a water bridge which is not visible in the X-ray structure.³¹ The second possibility is an increase in the π - π interactions between the aromatic ring system of the ligand and Tyr907 due to a lower LUMO.³¹

Inspection of the ligand binding pocket with back-projected variables LwLs₃ and LsLs₁₁ does not reveal any specific information about the negative effect of the methoxy-substitution. However, in one of the cocrystallized ligand-enzyme complexes (PDB code: 3PAX) the methoxy group is located in a sterically restrained environment between the aromatic ring systems of Tyr896 and Tyr907. This may indicate that the constraint imposed on the flexibility of this group upon binding is important for its reduced biological activity. If the same position is substituted with a methyl

**Figure 9.** Back-projection of the strong H-bond donor surface patches (grey) of Ds-type variables for compound **17** into the binding pocket of PARP-1 (PDB code: 1PAX). This visualization allows a direct validation of the importance of the identified MIVs. It can be seen that the hydrogen bond between the amide nitrogen of the ligand and the backbone amide oxygen of Gly863 is well described. In case of an amide substructure the H-bond donor surface patch is extended; however, only a small part of the patch is directed toward Gly863 which explains their lower activity.

group the contact surface as well as the π - π interactions between the aromatic ring systems will be increased without the constraints described above.

CONCLUSION

The validation of a novel methodology is an important step in its development process. For many QSAR techniques this validation is very difficult to perform since no direct measurement of meaning and relevance is available apart from model quality obtained for specific data sets. However, 3D-QSAR methods that allow a visualization of the derived models can easily be tested with regard to relevance by projecting the models into the binding pockets of receptors for which the binding mode of ligands is known. In this contribution, this visualization was performed for the novel translationally and rotationally invariant 3D-QSAR technique MaP. For this purpose, three diverse QSAR data sets for which a crystal structure of the biological target was available were used. Even though model quality for all three data sets was good, for validation purposes the study focused on the interpretation with and without the crystal structure of the relevant binding pocket. For all three data sets, interpretation of the respective QSAR equation was intuitive and straightforward. It allowed the identification of important surface features for the explanation of the variability in the data with

respect to the biological activity. To validate the models derived by MaP the relevant surface areas were back-projected into the crystal structures of the target protein. This visualization highlighted that the features identified without the crystal structure were actually relevant for the respective ligand–receptor interactions.

Therefore, it can be concluded that for the studied data sets the MaP technique is not only able to produce easily interpretable, high quality QSAR models without the need for a superimposition step, but that the derived models are actually able to describe relevant ligand–receptor interactions. However, no claim whatsoever is made that MaP can deduce an active site blueprint since this is in general not possible for QSAR methods. As for all QSAR techniques, only if the structural variation in the data set is addressing the relevant interaction sites a description of the structural features necessary for biological activity can be obtained.

ACKNOWLEDGMENT

This work was financially supported by the “Deutsche Forschungsgemeinschaft” (German Research Foundation); Grant: Sonderforschungsbereich 630, TP C5.

Supporting Information Available: Tables showing all compounds of the three data sets, a figure showing the hydrophilic surface for an AChE inhibitor, and a figure showing the superimposition of the cocrystallized TIBO R86183 with the conformation of that compound as used in this study. This material is available free of charge via the Internet at <http://pubs.acs.org>.

REFERENCES AND NOTES

- Cruciani, G.; Crivori, P.; Carrupt, P.-A.; Testa, B. Molecular Fields in Quantitative Structure-Permeation Relationships: the VolSurf Approach. *THEOCHEM* **2000**, *503*, 17–30.
- Stiefl, N.; Baumann, K. Mapping Property Distributions of Molecular Surfaces: Algorithm and Evaluation of a Novel 3D Quantitative Structure–Activity Relationship Technique. *J. Med. Chem.* **2003**, *46*, 1390–1407.
- 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*, 3233–3243.
- 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.
- Sippl, W.; Höltje, H.-D. Structure-Based 3D-QSAR – Merging the Accuracy of Structure-Based Alignments with the Computational Efficiency of Ligand-Based Methods. *THEOCHEM* **2000**, *503*, 31–50.
- Pan, D.; Tseng, Y.; Hopfinger, A. J. Quantitative Structure-Based Design: Formalism and Application of Receptor-Dependent RD-4D-QSAR Analysis to a Set of Glucose Analogue Inhibitors of Glycogen Phosphorylase. *J. Chem. Inf. Comput. Sci.* **2003**, *43*, 1591–1607.
- Zamora, I.; Afzelius, L.; Cruciani, G. Predicting Drug Metabolism: A Site of Metabolism Prediction Tool Applied to the Cytochrome P450 2C9. *J. Med. Chem.* **2003**, *46*, 2313–2324.
- Helms, V.; Wade, R. C. Computational Alchemy To Calculate Absolute Protein–Ligand Binding Free Energy. *J. Am. Chem. Soc.* **1998**, *120*, 2710–2713.
- Paulsen, M. D.; Ornstein, R. L. Binding Free Energy Calculations for P450cam-Substrate Complexes. *Protein Eng.* **1996**, *9*, 567–571.
- Nikitina, E.; Sulimov, V.; Zayets, V.; Zaitseva, N. Semiempirical Calculations of Binding Enthalpy for Protein–Ligand Complexes. *Int. J. Quantum Chem.* **2004**, *97*, 747–763.
- Hansson, T.; Marelus, J.; Aqvist, J. Ligand Binding Affinity Prediction by Linear Interaction Energy Methods. *J. Comput.-Aided Mol. Des.* **1998**, *12*, 27–35.
- Sippl, W.; Contreras, J. M.; Parrot, I.; Rival, Y. M.; Wermuth, C. G. Structure-Based 3D QSAR and Design of Novel Acetylcholinesterase Inhibitors. *J. Comput.-Aided Mol. Des.* **2001**, *15*, 395–410.
- Tariot, P. N. Maintaining Cognitive Function in Alzheimer Disease: How Effective are Current Treatments? *Alzh. Dis. Assoc. Dis.* **2001**, *15*, S26–S33.
- Harel, M.; Schalk, I.; Ehret-Sabatier, L.; Bouet, F.; Goeldner, M.; Hirth, C.; Axelsen, P. H.; Silman, I.; Sussman, J. L. Quaternary Ligand Binding to Aromatic Residues in the Active-Site Gorge of Acetylcholinesterase. *Proc. Natl. Acad. Sci. U.S.A.* **1993**, *90*, 9031–9035.
- Silman, I.; Harel, M.; Axelsen, P.; Raves, M.; Sussman, J. L. Three-Dimensional Structures of Acetylcholinesterase and of its Complexes with Anticholinesterase Agents. *Biochem. Soc. Trans.* **1994**, *22*, 745–749.
- Harel, M.; Quinn, D. M.; Nair, H. K.; Silman, I.; Sussman, J. L. The X-ray Structure of a Transition State Analogue Complex Reveals the Molecular Origins of the Catalytic Power and Substrate Specificity of Acetylcholinesterase. *J. Am. Chem. Soc.* **1996**, *118*, 2340–2346.
- Kryger, G.; Silman, I.; Sussman, J. L. Three-Dimensional Structure of a Complex of E2020 with Acetylcholinesterase from Torpedo Californica. *J. Physiol. (Paris)* **1998**, *92*, 191–194.
- Greenblatt, H. M.; Kryger, G.; Lewis, T.; Silman, I.; Sussman, J. L. Structure of Acetylcholinesterase Complexed with (–)-Galanthamine at 2.3 Å Resolution. *FEBS Lett.* **1999**, *463*, 321–326.
- Dvir, H.; Jiang, H. L.; Wong, D. M.; Harel, M.; Chetrit, M.; He, X. C.; Jin, G. Y.; Yu, G. L.; Tang, X. C.; Silman, I.; Bai, D. L.; Sussman, J. L. X-ray Structures of Torpedo Californica Acetylcholinesterase Complexed with (+)-Huperzine A and (–)-Huperzine B: Structural Evidence for an Active Site Rearrangement. *Biochemistry* **2002**, *41*, 10810–10818.
- Kapkova, P.; Stiefl, N.; Suerig, U.; Engels, B.; Baumann, K.; Holzgrabe, U. Synthesis, Biological Activity, and Docking Studies of New Acetylcholinesterase Inhibitors of the Bispiridinium Type. *Arch. Pharm. Med. Chem.* **2003**, *336*, 523–540.
- Pilger, C.; Bartoluci, C.; Lamba, D.; Tropsha, A.; Fels, G. Accurate Prediction of the Bound Conformation of Galanthamine in the Active Site of Torpedo Californica Acetylcholinesterase Using Molecular Docking. *J. Mol. Graph. Mol. Model.* **2001**, *19*, 288–296.
- Shen, T.; Tai, K.; Henchman, R. H.; McCammon, J. A. Molecular Dynamics of Acetylcholinesterase. *Acc. Chem. Res.* **2002**, *35*, 332–340.
- Muegge, I.; Podlogar, B. L. 3D-Quantitative Structure Activity Relationships of Biphenyl Carboxylic Acid MMP-3 Inhibitors: Exploring Automated Docking as Alignment Method. *Quant. Struct.-Act. Relat.* **2001**, *20*, 215–222.
- Bernard, P.; Kireev, D. B.; Chretien, J. R.; Fortier, P. L.; Coppet, L. Automated Docking of 82 N–Benzylpiperidine Derivatives to Mouse Acetylcholinesterase and Comparative Molecular Field Analysis with ‘Natural’ Alignment. *J. Comput.-Aided Mol. Des.* **1999**, *13*, 355–371.
- Recanatini, M.; Cavalli, A.; Belluti, F.; Piazzzi, L.; Rampa, A.; Bisi, A.; Gobbi, S.; Valenti, P.; Andrisano, V.; Bartolini, M.; Cavrini, V. SAR of 9-Amino-1,2,3,4-Tetrahydroacridine Based Acetylcholinesterase Inhibitors: Synthesis, Enzyme Inhibitory Activity, QSAR, and Structure-Based CoMFA of Tacrine Analogues. *J. Med. Chem.* **2000**, *43*, 2007–2018.
- Golbraikh, A.; Bernard, P.; Chretien, J. R. Validation of Protein-Based Alignment in 3D Quantitative Structure–Activity Relationships with CoMFA Models. *Eur. J. Med. Chem.* **2000**, *35*, 123–136.
- Barreca, M. L.; Carotti, A.; Carrieri, A.; Chimirri, A.; Monforte, A. M.; Pellegrini Calace, M.; Rao, A. Comparative Molecular Field Analysis (CoMFA) and Docking Studies of Non-Nucleoside HIV-1 RT Inhibitors (NNIs). *Bioorg. Med. Chem.* **1999**, *7*, 2283–2292.
- De Clercq, E. Non-Nucleoside Reverse Transcriptase Inhibitors (NNRTIs) for the Treatment of Human Immunodeficiency Virus Type 1 (HIV-1) Infections: Strategies to Overcome Drug Resistance Development. *Med. Res. Rev.* **1996**, *16*, 125–157.
- Das, K.; Ding, J.; Hsiou, Y.; Clark Jr., A. D.; Moereels, H.; Koymans, L.; Andries, K.; Pauwels, R.; Janssen, P. A.; Boyer, P. L.; Clark, P.; Smith Jr. R. H.; Kroeger Smith, M. B.; Micejda, C. J.; Hughes, S. H.; Arnold, E. Crystal Structures of 8-Cl and 9-Cl TIBO Complexed with Wild-Type HIV-1 RT and 8-Cl TIBO Complexed with the Tyr181Cys HIV-1 RT Drug-Resistant Mutant. *J. Mol. Biol.* **1996**, *264*, 1085–1100.
- Ding, J.; Das, K.; Moereels, H.; Koymans, L.; Andries, K.; Janssen, P. A.; Hughes, S. H.; Arnold, E. Structure of HIV-1 RT/TIBO R 86183 Complex Reveals Similarity in the Binding of Diverse Nonnucleoside Inhibitors. *Nat. Struct. Biol.* **1995**, *2*, 407–415.
- Costantino, G.; Macchiarulo, A.; Camaioni, E.; Pellicciari, R. Modeling of Poly(ADP-Ribose)Polymerase (PARP) Inhibitors. Docking of Ligands and Quantitative Structure–Activity Relationship Analysis. *J. Med. Chem.* **2001**, *44*, 3786–3794.
- Ha, H. C.; Snyder, S. H. Poly(ADP–Ribose)Polymerase-1 in the Nervous System. *Neurobiol. Dis.* **2000**, *7*, 225–39.
- Plaschke, K.; Kopitz, J.; Weigand, M. A.; Martin, E.; Bardenheuer, H. J. The Neuroprotective Effect of Cerebral Poly(ADP–Ribose)–

- Polymerase Inhibition in a Rat Model of Global Ischemia. *Neurosci. Lett.* **2000**, 284, 109–112.
- (34) Hahn, M. Receptor Surface Models. 1. Definition and Construction. *J. Med. Chem.* **1995**, 38, 2080–2090.
- (35) Hahn, M.; Rogers, D. Receptor Surface Models. 2. Application to Quantitative Structure–Activity Relationships Studies. *J. Med. Chem.* **1995**, 38, 2091–2102.
- (36) Ruf, A.; Rolli, V.; de Murcia, G.; Schulz, G. E. The Mechanism of the Elongation and Branching Reaction of Poly(ADP–Ribose)–Polymerase as Derived from Crystal Structures and Mutagenesis. *J. Mol. Biol.* **1998**, 278, 57–65.
- (37) White, A. W.; Almassy, R.; Calvert, A. H.; Curtin, N. J.; Griffin, R. J.; Hostomsky, Z.; Maegley, K.; Newell, D. R.; Srinivasan, S.; Golding, B. T. Resistance-Modifying Agents. 9.(1) Synthesis and Biological Properties of Benzimidazole Inhibitors of the DNA Repair Enzyme Poly(ADP–Ribose)Polymerase. *J. Med. Chem.* **2000**, 43, 4084–4097.
- (38) Ruf, A.; Mennissier de Murcia, J.; de Murcia, G.; Schulz, G. E. Structure of the Catalytic Fragment of Poly(ADP–Ribose)Polymerase from Chicken. *Proc. Natl. Acad. Sci. U.S.A.* **1996**, 93, 7481–7485.
- (39) Ruf, A.; de Murcia, G.; Schulz, G. E. Inhibitor and NAD⁺ Binding to Poly(ADP–Ribose)Polymerase as Derived from Crystal Structures and Homology Modelling. *Biochemistry* **1998**, 37, 3893–3900.
- (40) Stiefl, N.; Holzgrabe, U.; Baumann, K. A Critical Comparison of MaP (Mapping Property Distributions of Molecular Surfaces) to Established 3D-QSAR Techniques. In *Designing Drugs and Crop Protectants: Processes, Problems and Solutions*; Ford, M., Livingstone, D., Dearden, J., van de Waterbeemd, H., Eds.; Blackwell Publishing: Oxford, U.K., 2003; pp 195–197.
- (41) Alchemy 2000, Version 2.05, Tripos Inc., 1699 South Hanley Rd., St. Louis, Missouri, 63144, U.S.A., 1998.
- (42) SYBYL 6.8, Tripos Inc., 1699 South Hanley Rd., St. Louis, Missouri, 63144, U.S.A., 2001.
- (43) FlexS 1.10 BioSolveIT, An der Ziegelei 75, 53757 St. Augustin, Germany, 2003.
- (44) Stiefl, N.; Bringmann, G.; Rummey, C.; Baumann, K. Evaluation of Extended Parameter Sets for the 3D-QSAR Technique MaP: Implications for Interpretability and Model Quality Exemplified by Antimari-ally Active Naphthylisoquinoline Alkaloids. *J. Comput.-Aided Mol. Des.* **2003**, 17, 347–365.
- (45) Baumann, K.; Albert, H.; von Korff, M. A Systematic Evaluation of the Benefits and Hazards of Variable Selection in Latent Variable Regression. Part I: Search Algorithm, Theory and Simulations. *J. Chemom.* **2002**, 16, 339–350.
- (46) Draper, N. R.; Smith, H. *Applied Regression Analysis*; John Wiley & Sons: New York, U.S.A., 1981.
- (47) Baumann, K.; Stiefl, N. Validation Tools for Variable Subset Selection. *J. Comput.-Aided Mol. Des.* **2004**, 18, 549–592.
- (48) Baumann, K.; von Korff, M.; Albert, H. A Systematic Evaluation of the Benefits and Hazards of Variable Selection in Latent Variable Regression. Part II: Practical Applications. *J. Chemom.* **2002**, 16, 351–360.
- (49) Baumann, K. Distance Profiles (DiP): A translationally and rotationally invariant 3D structure descriptor capturing steric properties of molecules. *Quant. Struct.-Act. Relat.* **2002**, 21, 507–519.
- (50) Topliss, J. G.; Edwards, R. P. Chance Factors in Studies of Quantitative Structure–Activity Relationships. *J. Med. Chem.* **1979**, 22, 1238–1244.
- (51) Klopman, G.; Kalos, A. N. Causality in Structure–Activity Studies. *J. Comput. Chem.* **1985**, 6, 492–506.
- (52) Kennard, R. W.; Stone, L. A. Computer Aided Design of Experiments. *Technometrics* **1969**, 11, 137–148.
- (53) Faber, N. M. Estimating the Uncertainty in Estimates of Root Mean Square Error of Prediction: Application to Determining the Size of an Adequate Test Set in Multivariate Calibration. *Chemom. Intell. Lab. Sys.* **1999**, 49, 79–89.
- (54) Ghafourian, T.; Dearden, J. C. The Use of Atomic Charges and Orbital Energies as Hydrogen-Bonding-Donor Parameters for QSAR Studies: Comparison of MNDO, AM1 and PM3 Methods. *J. Pharm. Pharmacol.* **2000**, 52, 603–610.
- (55) Berman, H. M.; Westbrook, J.; Feng, Z.; Gilliland, G.; Bhat, T. N.; Weissig, H.; Shindyalov, I. N.; Bourne, P. E. The Protein Data Bank. *Nucleic Acids Res.* **2000**, 28, 235–42.

CI049683I