ELSEVIER

Journal of Molecular Structure (Theochem) 503 (2000) 31–50

THEO CHEM

www.elsevier.nl/locate/theochem

Structure-based 3D-QSAR—merging the accuracy of structure-based alignments with the computational efficiency of ligand-based methods ☆

W. Sippl*, H.-D. Höltje

Institute for Pharmaceutical Chemistry, Heinrich-Heine-University at Düsseldorf, Universitätsstr. 1, D-40225 Düsseldorf, Germany

Abstract

One of the major challenges in computational approaches to drug design is the accurate prediction of binding affinity of biomolecules. The strategies that can be applied for this purpose fall into two major categories—the indirect ligand-based and the direct receptor-based approach. In this contribution, we used a combination of both approaches in order to improve the prediction accuracy for drug molecules. The combined approach was tested on two sets of ligands for which the three-dimensional structure of the target receptor was known—estrogen receptor ligands and acetylcholinesterase inhibitors. The binding modes of the ligands under study were determined using an automated docking program (AUTODOCK) and were compared with available X-ray structures of corresponding protein–ligand complexes. The ligand alignments obtained from the docking simulations were subsequently taken as the basis for a comparative field analysis applying the GRID/GOLPE program. Using the interaction field derived with a water probe and applying the smart region definition variable selection, highly predictive models were obtained. The comparison of our models with interaction energy-based models and with traditional CoMFA models obtained using a ligand-based alignment indicates that the combination of structure-based and 3D-QSAR methods is able to improve the prediction ability of the underlying model.

Keywords: 3D-QSAR; Drug design

1. Introduction

One of the main objectives in today's drug design is the prediction of new biologically active compounds on the basis of previously synthesized molecules. The strategies that can be applied for this purpose fall into two major categories—the indirect ligand-based and the direct receptor-based approach. The common aim of both strategies is to understand structure–activity relationships and to employ this knowledge for proposing new compounds with enhanced activity and selectivity profile for a specific therapeutic target. The ligand-based methods, including traditional quantitative structure–activity relationships (QSAR) [1,2] and modern 3D-QSAR techniques such as the comparative molecular field analysis (CoMFA) [3,4], are based entirely on experimental structure–activity relationships for enzyme inhibitor or receptor ligands. For the direct receptor-based methods, which include molecular docking and advanced molecular dynamics simulations, the 3D-structure of a target enzyme or even a receptor–effector complex is required with atomistic resolution, generally determined by either X-ray crystallography, NMR spectroscopy or protein homology model building [5].

3D-QSAR methods, especially CoMFA, are

☆ Dedicated to Prof. Dr. B.C. Lippold on the occasion of his 60th birthday.

* Corresponding author.

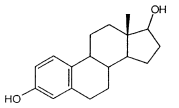
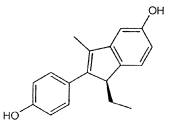
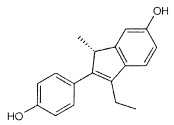
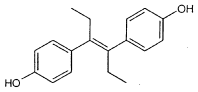
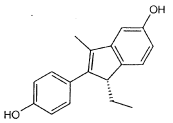
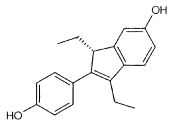
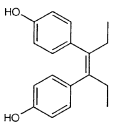
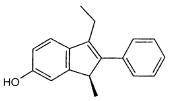
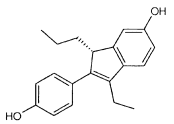
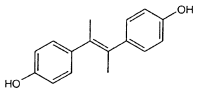
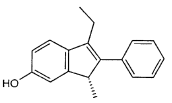
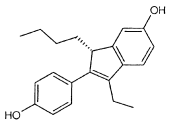
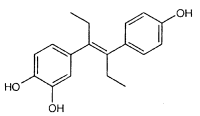
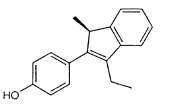
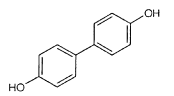
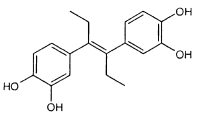
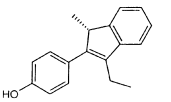
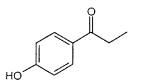
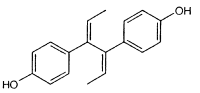
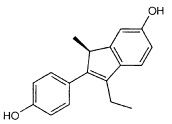
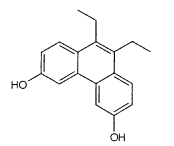
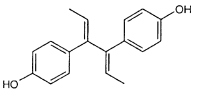
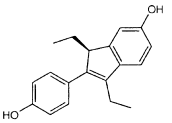
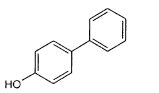
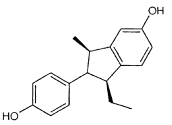
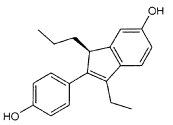
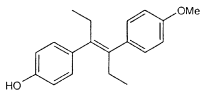
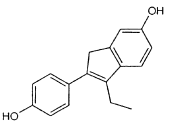
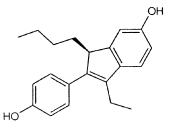
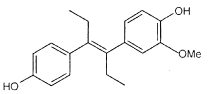
0166-1280/00/$ - see front matter
PII: S0166-1280(99)00361-9

nowadays used widely in drug design, since they are computationally not demanding and afford fast generation of QSARs from which biological activity of newly synthesized molecules can be predicted. The basic assumption in CoMFA is that a suitable sampling of the steric and electrostatic fields around a set of aligned molecules might provide all the information necessary for understanding their biological activities [3]. The suitable sampling is achieved by calculating interaction energies between each molecule and an appropriate probe at regularly spaced grid points surrounding the molecules. The resulting energies derived from simple potential functions (normally Coulomb and Lennard-Jones potential) can then be contoured to give a quantitative spatial description of molecular properties. If correlated with biological activity, 3D-fields can be generated, which describe the contribution of a region of interest surrounding the ligands to the target properties. However there is a main difficulty in the application of 3D-QSAR methods such as CoMFA. For a correct model, a spatial orientation of the ligands towards one another has to be found, which is representative for the relative differences in the binding geometry at the protein binding site. The success of a molecular field analysis is therefore completely determined by the quality of the choice of the superimposition of the studied molecules [6–8]. Therefore, the first step in a 3D-QSAR study is the generation of a reliable pharmacophore model. Many strategies have been reported for this purpose in the literature [4,9,10]. Depending on the molecular flexibility and the structural diversity of the investigated compounds, this task of unique pharmacophore generation becomes less feasible. Despite the difficulties concerning the molecular alignment, many successful 3D-QSAR studies applying the CoMFA approach have been reported in the last few years.

Structure-based methods are nowadays able to calculate fairly accurately the position and orientation of a potential ligand in a receptor binding site. This has been demonstrated by various docking studies described in the literature [11–14]. The direct methods yield important information concerning the spatial orientation of the ligands in the binding site and also towards other ligands binding to the same target. The major problem of today's docking programs is the inability to evaluate binding free energies correctly in order to rank different ligand–receptor complexes. Since docking programs generate huge amount of possible ligand–receptor complexes, it is impossible to determine, a priori, which ligand conformation represents the bioactive one. The problem predicting affinity has generated considerable interest in developing methods to calculate ligand affinity reliably for a widely diverse series of molecules binding to the same target protein of known structure [15–18]. For the calculation of ligand–receptor interaction energies, most approaches rely on molecular mechanics force fields that represent van der Waals and Coulombic interactions on the basis of empirical potentials. Other approaches use more simple scoring functions rather than calculating the affinity by molecular equations [17]. These methods commonly use available experimental data to obtain parameters for some relatively simple functionals that allow for fast estimation of the binding energy. The estimated binding energies or scores are widely used to discriminate between active and inactive ligands, for example in virtual database screening, but are mostly not accurate enough for 3D-QSAR analysis [18]. The main problem in affinity prediction is that the underlying molecular interactions are highly complex and various terms should be taken into account to quantify the free energy of the interaction process. Only rigorous methods, such as the free energy perturbation methods, are at the moment able to correctly predict binding affinity. Since these methods are computationally very intensive, such methods can not be applied to large ligand series, commonly studied in QSAR analysis.

One possibility to overcome these difficulties seems to be the combination of ligand and receptor-based approaches. It is quite appealing to combine the accuracy of the structure-based alignments with the computational efficiency of the ligand-based methods. According to this strategy the alignment is generated on the basis of the experimental or predicted position of molecules in the binding pocket and 3D-QSAR programs are then used to calculate the binding affinity. With respect to structure-based methods, the 3D-QSAR approach has the advantage of dealing only with the differences in affinity of a special series of compounds. In this case the interaction energy of each ligand is not essential, because some of the terms describing this energy (desolvation processes, entropic terms) take approximately the same value for

Table 1
Structure of estrogen receptor ligands included in the study

No	Structure	No	Structure	No	Structure
1		**11**		**21**	
2		**12**		**22**	
3		**13**		**23**	
4		**14**		**24**	
5		**15**		**25**	
6		**16**		**26**	
7		**17**		**27**	
8		**18**		**28**	
9		**19**		**29**	
10		**20**		**30**	

every ligand and can be neglected. The combination of direct and indirect methods was successfully employed recently by several groups [19–23].

In this contribution, we report on the application of structure-based 3D-QSAR methods, exemplified by recent molecular modeling studies on different classes of protein ligands from our own laboratories. Special emphasis will be placed on a detailed description of the combined receptor/ligand-based approach. In addition, we would like to compare the results obtained by such a combined strategy with results obtained from classical 3D-QSAR and structure-based approaches.

2. Case studies

2.1. Estrogen receptor ligands

The effect of estrogens is mediated by an intracellular estrogen receptor (ER), which belongs to the steroid/thyroid nuclear hormone superfamily. These receptors act as transcriptional activators via a direct interaction with DNA sequences, termed as response elements [24]. The biological action of estradiol (**1**), the most active endogenous estrogen, and other estrogen receptor ligands and their primary interactions with the receptor protein have been topics of much interest over the years [25]. Particularly, the elucidation of the structural requirements, which enable binding of estrogens to the receptor, has actively been pursued as a means to develop novel therapeutic agents. Natural and synthetic estrogen receptor ligands are used in a number of clinical applications such as, for example, breast cancer therapy, treatment and prevention of oestoporosis, and in estrogen replacement therapy of postmenopausal women [26].

Over the last years a large amount of structure-activity information for modified estrogens and nonsteroidal estrogen receptor ligands has been reported in the literature [27]. Several models of the estrogen receptor ligand pharmacophore have been published in the last few years [27–33]. For our investigation, we selected structure–activity data for a series of 30 structurally diverse estrogen receptor ligands originally reported by Sadler et al. [33]. The molecular structures of these 30 nonsteroidal estrogen receptor ligands are represented in Table 1. The objective of our selection was twofold. Firstly, the biological activity of the estrogen receptor ligands reported in this study, has been measured in the same laboratory under the same condition. This is a prerequisite for a successful 3D-QSAR analysis (a detailed discussion on the use of biological data in QSAR can be found in Ref. [7]). Secondly, the authors have performed a CoMFA analysis using a traditional ligand-based alignment rule. Therefore, it was quite interesting to see whether applying a structure-based 3D-QSAR approach could provide a better explanation of the biological data.

In 1997 the 3D-structure of the estrogen receptor has been resolved by Brzozowski et al. [34]. Until now four X-ray structures of the receptor liganded with different molecules—estradiol (**1**), diethylstilbestrol (**2**), a nonsteroidal compound, and two antagonists raloxifen and 4-hydroxytamoxifen—have been published [34,35]. The availability of several structurally diverse structures bound in the active site of the estrogen receptor provided important experimental information detailing the molecular alignment of the studied molecules.

We took the crystal structures of the human estrogen receptor-α-ligand binding domain liganded with estradiol and diethylstilbestrol (pdb code: 1ERE and 2ERD) from the Brookhaven Protein Databank. The two crystal structures were overlaid using the backbone atoms (Fig. 1). The estrogen receptor shows a nearly identical three-dimensional structure in these X-ray structures. The only major conformational differences are the orientation of two sidechains in the binding pocket—His524 and Met421. As the analysis of the crystal structures show, estradiol and diethylstilbestrol bind in similar way to the receptor. Both ligands show hydrogen bonds with Glu353 and His524 and interact in addition with hydrophobic residues at the binding pocket. Due to the size of diethylstilbestrol, His524 is turned somewhat further from the binding pocket in the corresponding complex.

In order to find out the correct binding mode for all the studied ligands, we performed a computational docking experiment. For our docking analysis we selected the program AUTODOCK (version 2.4) which has been shown to successfully reproduce experimentally observed binding modes [12,36,37]. The program is described in detail elsewhere [12].

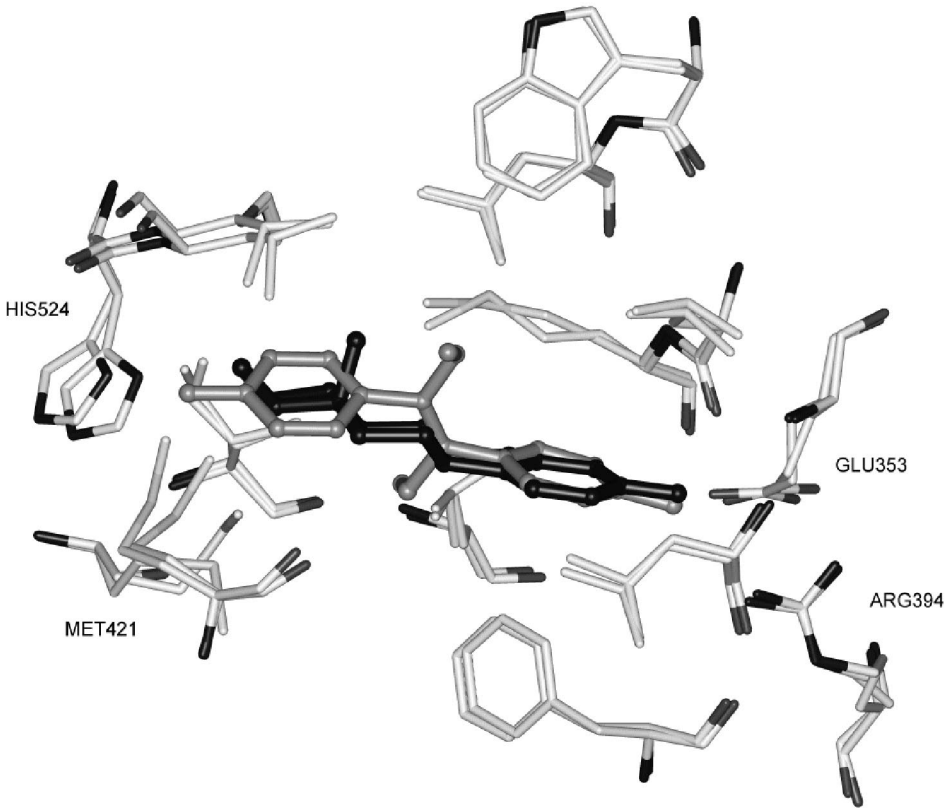


Fig. 1. Comparison of the X-ray structures of the estrogen receptor liganded with estradiol (dark-gray) and diethylstilbestrol (gray). The two crystal structures were overlaid using the backbone atoms. Only the amino acid residues in proximity to the binding pocket are shown for clarity.

AUTODOCK uses a simulated annealing procedure to explore the binding possibilities of a ligand in a binding pocket. The interaction energy of ligand and protein is evaluated using atom affinity potentials calculated on a grid similar to that described by Goodford [38]. All ligand atoms but no protein atoms were allowed to move during the docking simulation. For each ligand the simulation was composed of 100 docking runs, and the docked complexes were clustered with an RMSD tolerance of 0.7 Å. The derived complexes were then refined and the interaction energies were calculated using the YETI force field [39,40]. This force field uses more sophisticated energy potentials to calculate the binding energy compared to the more simple functions implemented in docking programs, such as AUTODOCK.

We started our docking studied using the two molecules, estradiol and diethylstilbestrol, for which the binding mode has been determined experimentally. These two ligands were taken as positive controls to test the program used for docking. Since the geometry of the binding pocket exhibits small differences in the two complexes, both protein structures were taken as target receptor for the docking procedure. The complexes, which were obtained using the corresponding crystal structure (IERE for estradiol and 2ERD for diethylstilbestrol), showed the lower interaction energies. In Fig. 2 the complex showing the lowest energy within the AUTODOCK/YETI procedure for diethylstilbestrol is overlaid with its corresponding crystal structure. Similar results were obtained for estradiol (not shown). As one can recognize, AUTODOCK was successful in reproducing the experimentally found binding position for estradiol and diethylstilbestrol (one may speak of reproduction if the root-mean square deviation (RMSD) is below 1–2 Å [11]). The RMSD values between the observed and calculated position are 0.21 Å for estradiol and 0.37 Å for diethylstilbestrol.

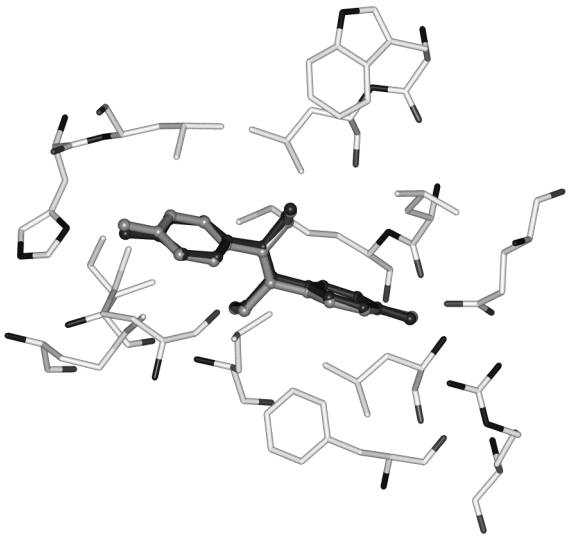


Fig. 2. Comparison of the position of diethylstilbestrol calculated by the AUTODOCK/YETI procedure (dark-gray) and the one observed in the crystal structure (RMSD = 0.37 Å).

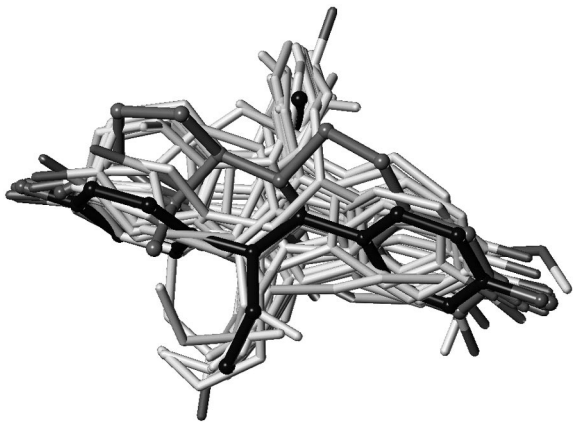


Fig. 3. Alignment of all investigated estrogen receptor ligands obtained by the docking simulation (estradiol is colored dark-gray, diethylstilbestrol is colored gray).

In the next step the ligand–receptor complexes for the resulting 28 ligands were computed in the same way. For all ligands the complex which possessed the lowest interaction energy was selected for the further investigation. Fig. 3 shows the superimposition of all 30 ligands (the protein structure is not shown for clarity). We focussed then on a possible correlation between the calculated interaction energies and the determined biological activities. The biological activities were taken from the original publication [33]. They were expressed as relative binding affinities (RBA) relative to estradiol which is set to 100. These numbers were then transformed to the decadic log values. The interaction energy was calculated as the sum of the van der Waals and the electrostatic energies. The YETI force field, AM1 partial charges and a distance dependent dielectric function were selected for the calculation. A correlation coefficient of $r^2 = 0.54$ and a cross-validated coefficient of $q^2 = 0.49$ were obtained for the 30 ligands (the calculated values are listed in Table 2, model C). Trends between experimental and calculated can be distinguished, but the standard deviation of $s = 0.83$ indicates the inaccuracy of the calculation. Also, compared to the results of the original CoMFA study by Sadler, who reported a cross-validated correlation coefficient of $q^2 = 0.796$, one can recognize the only moderate prediction. Since we have not considered solvation or entropic effects in our calculation of the binding energy, it is not surprising that the correlation is not high. This has also been reported by other authors who have applied this strategy [41–43].

With respect to the receptor-based methods, the 3D-QSAR approach has the advantage of dealing only with the differences in affinity of a special series of compounds. In this case the interaction energy of each ligand is not important, because some of the terms describing this energy (desolvation processes, entropic terms) take approximately the same value for every ligand. Since only differences in the binding affinity are regarded, these terms are not considered. Therefore we used the comparative field analysis to develop a quantitative structure-activity relationship for the investigated ligands.

The superimposition of the ligands derived from the molecular docking was taken as starting point for a comparative molecular field analysis. We used the GRID/GOLPE method for our analysis [38,44]. The interaction field between the ligands and a water probe were calculated using the GRID program employing a grid spacing of 1 Å. The GRID calculations gave 8740 variables for each compound. A major part of these variables is not important for describing the interaction between the ligand and the receptor and is introducing only noise in the statistical PLS analysis [45]. These variables were selected and eliminated using the SRD/FFD (Smart Region Definition/Factorial

Table 2
Actual and calculated activities

Compound	Actual	Model A[a]	Model B[b]	Model C[c]
1	2.00	1.77	1.99	1.98
2	2.46	2.25	2.44	1.97
3	−0.10	−0.26	0.01	0.31
4	1.52	1.93	1.44	1.35
5	2.00	1.62	1.98	2.31
6	1.40	1.47	1.39	*2.33*
7	−0.52	−0.68	−0.60	*1.51*
8	1.30	0.88	1.40	1.73
9	0.30	0.38	0.36	0.05
10	1.14	*1.90*	1.07	1.01
11	2.00	1.94	2.01	*0.72*
12	2.15	1.90	2.36	1.91
13	0.26	−0.04	0.17	0.04
14	−0.70	−0.24	−0.54	*0.53*
15	0.75	0.49	0.68	0.67
16	−0.05	−0.11	0.01	*0.92*
17	2.46	2.18	2.36	*1.61*
18	2.47	2.36	2.51	*0.78*
19	2.36	2.39	2.33	*1.72*
20	2.25	2.40	2.32	*1.61*
21	1.11	1.14	0.93	1.08
22	1.04	1.19	1.03	1.35
23	1.26	1.19	1.26	1.13
24	0.90	1.22	0.95	*−0.16*
25	−1.70	−1.50	−1.76	*−0.37*
26	−1.00	−1.18	−0.96	−0.70
27	−0.80	−0.73	−0.80	−1.09
28	−1.70	−1.51	−1.67	*−0.59*
29	1.30	1.49	1.29	*0.11*
30	1.00	0.77	0.94	1.02

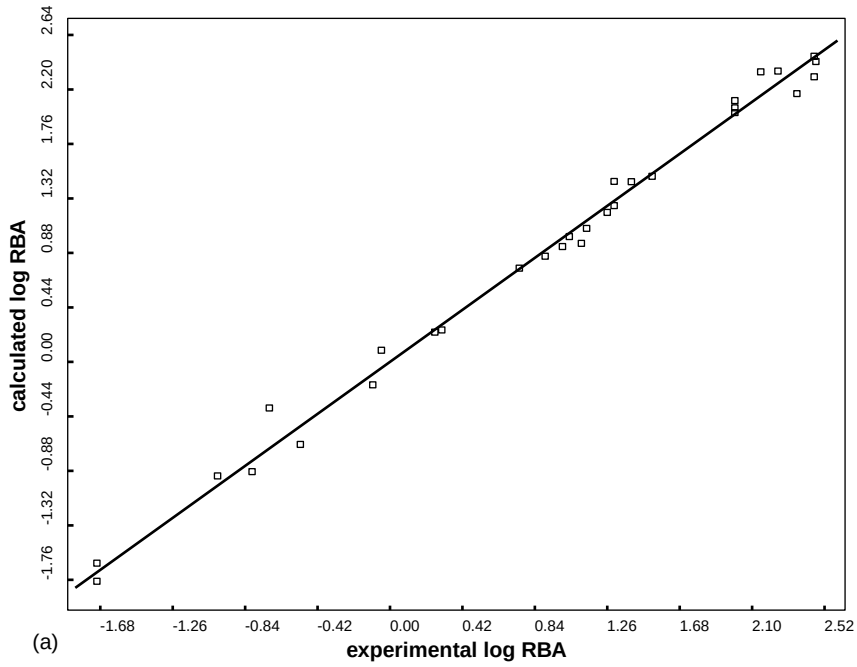
[a] Original CoMFA model of Sadler.
[b] GRID/GOLPE model.
[c] Interaction energy model. Italics mark deviations more than 0.5 log units from the actual activity.

Fractional Design) method within the GOLPE program (for a detailed description of this approach see [45]). The resulting models are of higher quality than models calculated without variable selection. Applying the SRD preselection procedure, the number of variables was reduced from 8340 to 1105 variables without changing the quality of the model. Region selection based on the FFD procedure further reduced the number of variables to 493 with major improvement of the quality of the model. To form the basis for a predictive statistical model, the method of partial least squares (PLS) regression was used to analyze the 30 compounds by correlating variations in their biological activities with variations in their interaction fields. The optimum number of principal components corresponding to the smallest standard error of prediction was determined by the leave-one-out (LOO) cross-validation procedure. Using the optimal number of principal components, the final PLS analysis was carried out without cross-validation to generate a predictive model with a conventional correlation coefficient. The LOO cross-validation method might lead to high q^2 values which do not necessarily reflect a general predictiveness of a model. Therefore we have performed a second cross-validation, using five groups of approximately the same size in which the objects were assigned randomly. In this method 80% of the compounds were randomly selected and a model is generated, which is then used to predict the remaining compounds (leave-20%-out). This cross-validation technique has been shown to yield better indices for the robustness and predictivity of a model than the normal LOO procedure [46].

The results of the statistical analysis are represented in Fig. 4. The analysis based on the structure-based alignment yielded a correlation coefficient with a cross-validated q^2 of 0.921 using four principle components. The conventional r^2 of this analysis is 0.992. This means, the model explains approximately 99% of the variance in ligand binding of the investigated compounds. The model expresses also good predictive ability, indicated by the high correlation coefficient of $q^2 = 0.900$ obtained by using the leave-20%-out cross-validation procedure.

The comparison of our results with the CoMFA results of Sadler, who reported a cross-validated q^2 value of 0.796, indicated that our model constructed on the basis of the receptor structure supplies a better explanation of the biological activities. This is also indicated by the small deviation of the calculated from the experimental values in our calculation (Table 2). The original CoMFA model is for example not able to explain the different activities of the two enantiomers **13** and **14**. The structure-based approach supplies here a unique explanation. (A detailed description of the interpretation of the results it out of the scope of this contribution and will be presented elsewhere [47].)

One of the important features of a GRID/GOLPE analysis is the possibility of translating back the PLS coefficients assigned to each variable to the



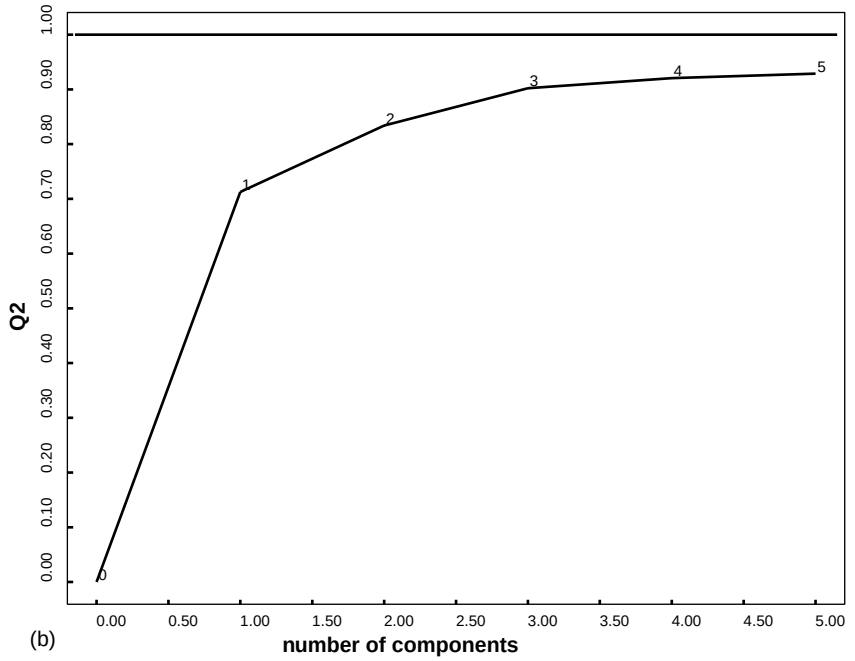


Fig. 4. GRID/GOLPE results. Fig. 4a shows the calculated vs experimental activity. Fig. 4b shows the cross-validated (LOO) squared correlation coefficients (q^2) for different model dimensionalities.

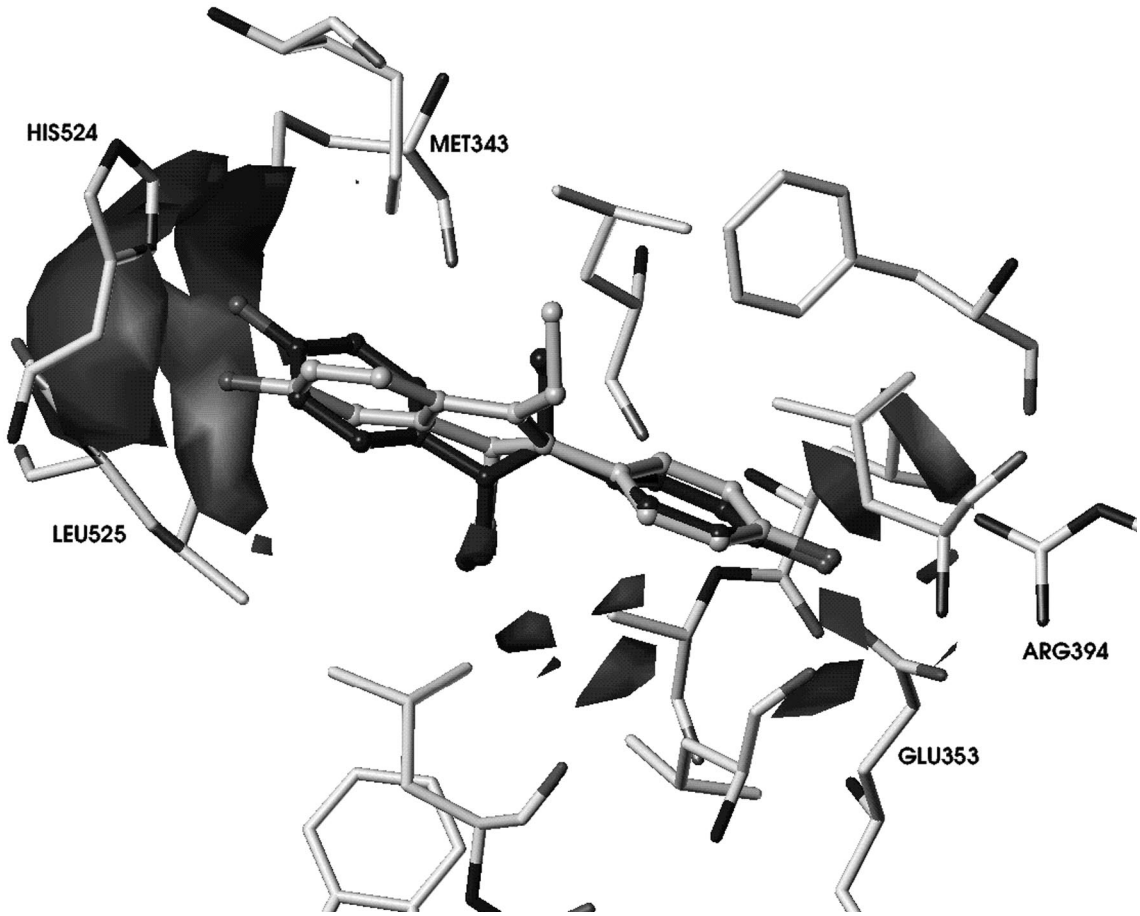


Fig. 5. Comparison between the PLS coefficient maps, contoured at −0.004, and the amino acid residues located close to the binding pocket. The region around His524 and the backbone of Leu525 and Gly521 contains only negative coefficients indicating that polar interactions here would increase the activity. (Diethylstilbestrol is colored dark-gray, and the most potent ligand—compound 17—is colored gray.)

3D-positions they occupy in real space. These values can be contoured at a particular significant level and can be displayed as grid plot of PLS coefficients. The contour coefficient maps indicate those areas in which the PLS model has found a high correlation between the ligand-probe interaction energy and the biological activity. In our analysis we have used the water probe, which is a polar group with the ability to participate in hydrogen bonds and electrostatic interactions. Consequently, areas containing negative coefficients correspond to areas in the binding pocket where energetically favorable interactions produce an increase in the activity. These regions tend to coincide with the presence of polar groups in the binding pocket. In contrast, as area with negative coefficients corresponds to a region where the presence of energetically favorable interactions decrease the biological activity. Such regions can appear for several reasons. For example, they may indicate a hydrophobic region where the presence of a polar group is unfavorable. Alternatively, these regions may represent a ligand–receptor interaction that may only be present through a different binding mode, which is less favorable for the activity [22]. Since the structure of the estrogen receptor was known, it was quite interesting to compare the results of our 3D-QSAR analysis, given by the PLS coefficient maps, with the chemical and geometrical properties of the binding site. It is necessary to note, that, in general such comparison should be attempted carefully. The PLS coefficient contour

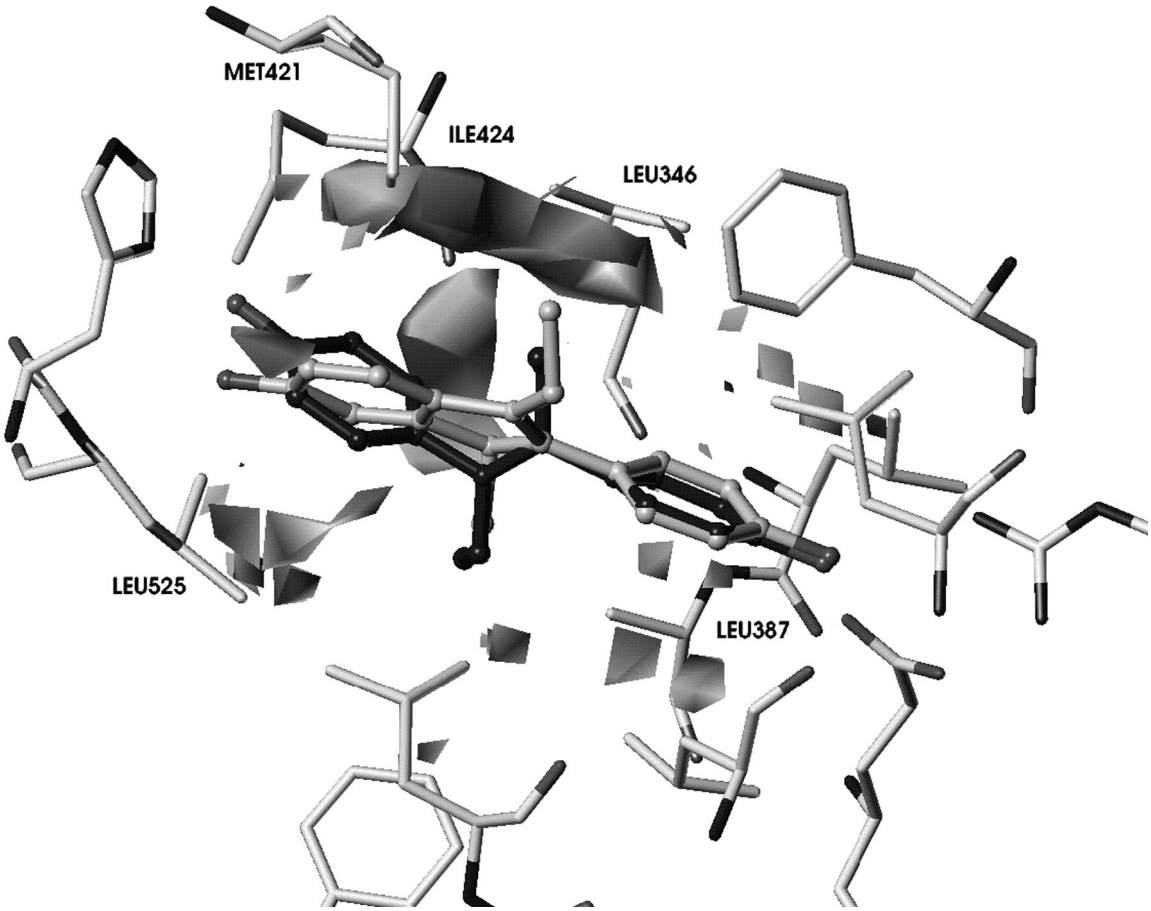
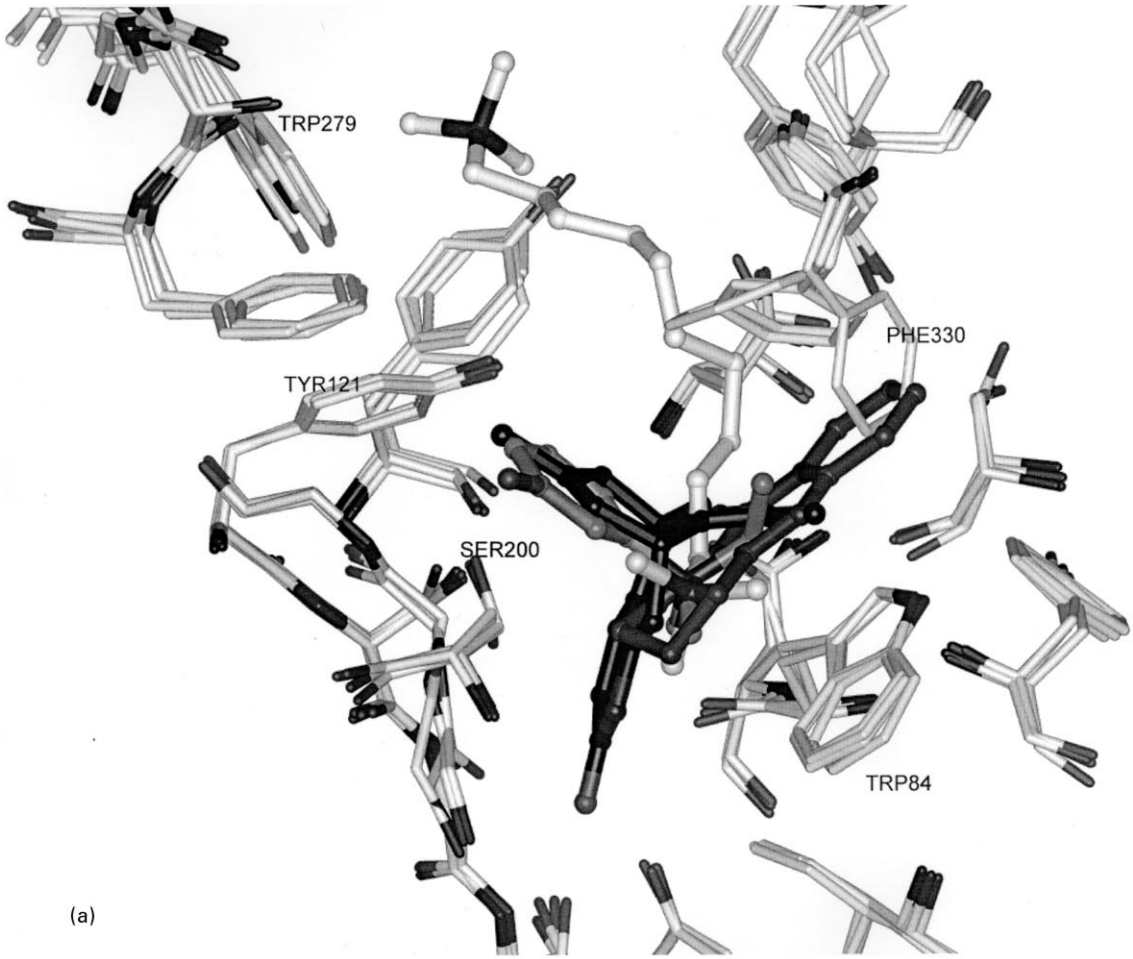


Fig. 6. Comparison between the PLS coefficient maps, contoured at 0.004, and the amino acid residues located close to the binding pocket. (Diethylstilbestrol is colored dark-gray, and the most potent ligand—compound 17—is colored gray).

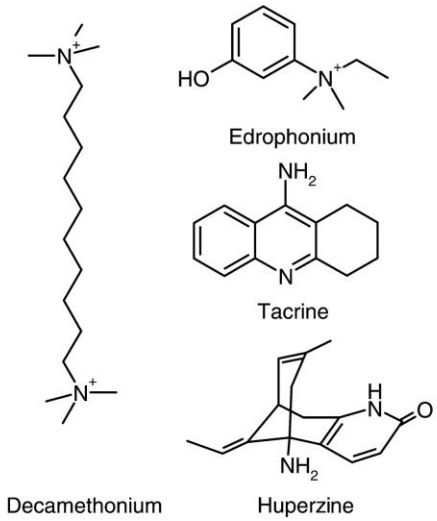
maps can by no means be regarded as a set of low resolution picture of the binding site, since the contour maps reflect only those regions in space, where the ligand–probe interaction energy is correlated with the biological activity [22,45]. Nevertheless when the alignment is based on the receptor structure one might expect a certain correlation.

We superimposed the coefficient contour maps and the active site of the estrogen receptor and compared them with each other. Figs. 5 and 6 show the grid plot of the PLS coefficients for the water probe. The contours in Fig. 5 represent negative coefficients under −0.004, while, the contours in Fig. 6 represent positive coefficients over 0.004. Since we used the water probe, the positive contour maps indicate the areas where polar interaction decrease activity and the negative contour maps show the regions where polar interaction increase activity. In general, we observed a nice agreement between the maps and the positions of particular amino acid residues in the active site. More specially, we find that a big positive field occupies a hydrophobic pocket close to Leu346,

Fig. 7. (a) Superimposition of the investigated crystal structures of AChE liganded with huperzine (black), tacrine (dark-gray), edrophonium (gray) and decamethonium (light-gray). Only the amino acid residues close to the binding site are displayed. The only major conformational difference between the four complexes is the orientation of the phenyl ring of Phe330, a residue located in the middle of the gorge. (b) Molecular structures of edrophonium, decamethonium, huperzine and tacrine.


TRP279
TYR121
SER200
PHE330
TRP84

(a)


HO
Edrophonium
NH2
N
Tacrine
H
N
O
NH2
Decamethonium
Huperzine

(b)

Met421 and Ile424. The most active ligands possess alkyl substituents, which are able to occupy this pocket. A second positive field coincides with Leu525, indicating that interactions between the aromatic ring system of the ligands and Leu525 influence ligand binding. The region around His524 and the backbone of Leu525 and Gly521 contains only negative coefficients (contours in Fig. 6) indicating that polar interactions here would increase the activity. Less pronounced fields are located close to Glu353 and Arg394. Since all studied ligands possess a hydroxyl group interacting with these two residues, it is not surprising that this region do not contribute to an explanation of the affinity variation. In conclusion, we obtained a nice complementary relationship between the coefficient contour maps and the binding site of the estrogen receptor which indicate the powerful ability of the GRID/GOLPE procedure for reproducing a meaningful 3D-QSAR model.

2.2. Acetylcholinesterase inhibitors

In the second example, we report on the application of the described combined approach to a series of aminopyridazine acetylcholinesterase (AChE) inhibitors [48]. According to the cholinergic hypothesis memory impairments in patients with Alzheimer's disease result from a deficit of cholinergic functions in the brain [49,50]. One promising strategy to overcome this deficit is the inhibition of the acetylcholinesterase, the enzyme responsible for the hydrolysis of acetylcholine. Many compounds were synthesized in the past and tested for acetylcholinesterase inhibition [50,51]. The chemical structures of these compounds are diverse, ranging from quarternary compounds—such as decamethonium or edrophonium—to formally neutral molecules such as tacrine or huperzine (see Fig. 7a and b). A novel family of therapeutically promising inhibitors are the benzylpiperidine derivatives including donepezil. Donepezil was recently introduced into the Alzheimer therapy [51]. The starting point for the development of aminopyridazines as AChE inhibitors was the finding, that the antidepressant minaprine (Fig. 8) shows weak inhibition of AChE [52,53]. Since minaprine has a unique structure among the known AChE inhibitors, it was taken as promising lead compound. The synthesized inhibitors, which are used within this modeling

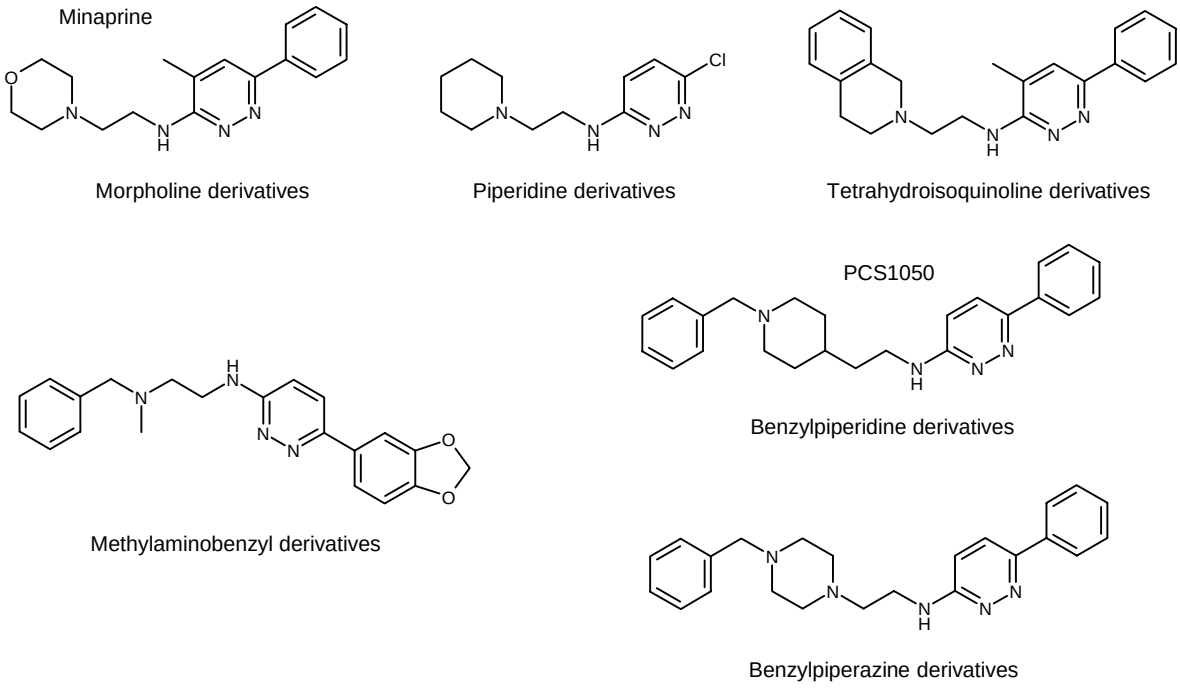


Fig. 8. Examples of the investigated aminopyridazine AChE inhibitors.

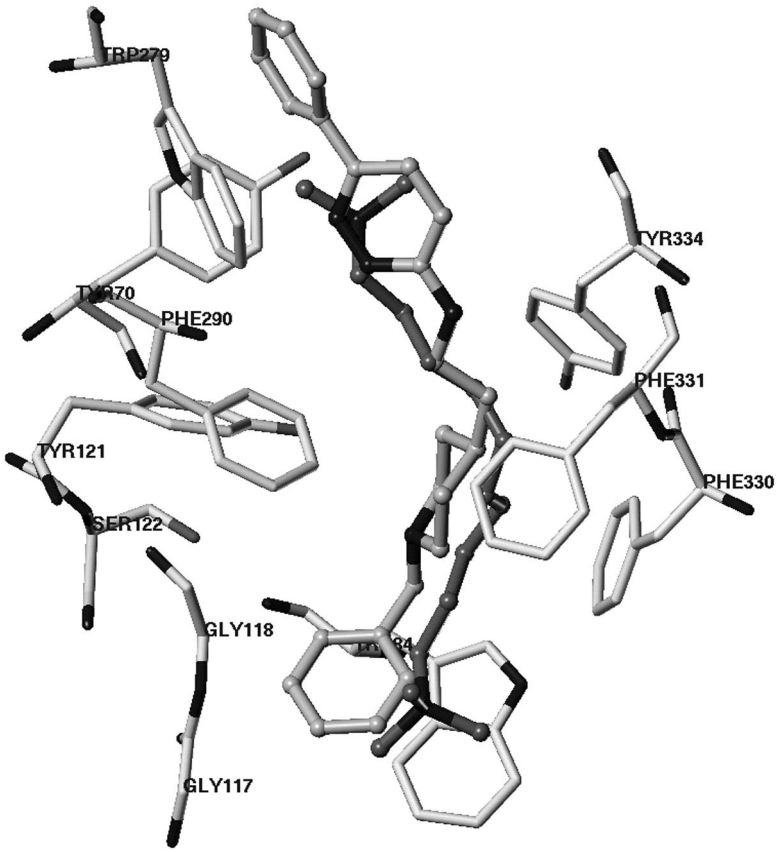


Fig. 9. Comparison between the predicted position of compound PCS1050 (gray) and the X-ray structure of decamethonium (dark-gray) are shown.

study, can be classified into six different families. The inhibitors are described in detail elsewhere [54,55]. Examples from each family are shown in Fig. 8.

The positioning of the molecules in a fixed lattice is the most important input variable in comparative molecular field analysis. In order to obtain a realistic alignment of the investigated inhibitors, we included also in this example the known crystal structures of AChE in our 3D-QSAR study. During the last few years four structures of AChE complexed with reversible inhibitors have been published [56–58]. Unfortunately up to now no X-ray structure is available for AChE complexed with the potent benzylpiperidine inhibitors. As in the investigation of the estrogen receptor ligands, we decided to use the AUTODOCK program in order to determine the exact position of the inhibitors in the binding pocket.

The detailed inspection of the four AChE-inhibitor X-ray structures, obtained from the Brokhaven Protein Databank (1ACL liganded with decamethonium, 2ACK liganded with edrophonium, 1ACJ liganded with tacrine and 1VOT liganded with hyperzine) yielded crucial information concerning the orientation of the inhibitors in the binding pocket. AChE shows a nearly identical three-dimensional structure in all known X-ray structures. The active site is located 20 Å from the protein surface at the bottom of a deep and narrow gorge. The only major conformational difference between the four complexes is the orientation of the phenyl ring of Phe330, a residue located in the middle of the gorge (Fig. 7a). Depending on the co-crystallized inhibitor this aromatic residue adopts a different conformation. However the positions of the four inhibitors in the

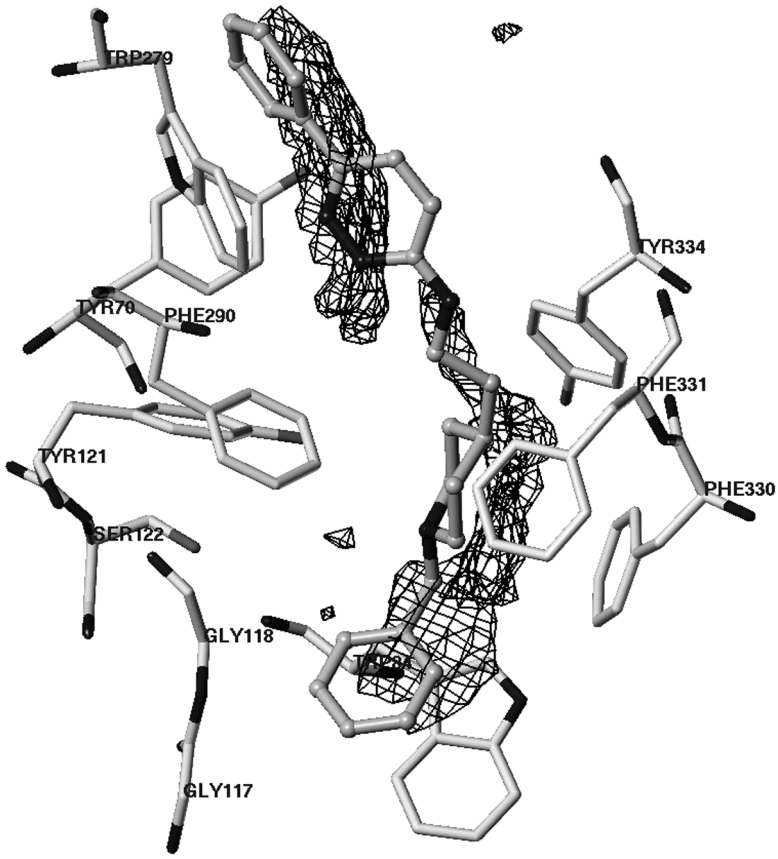


Fig. 10. Favorable regions of interaction between the hydrophobic DRY probe and the active site (contour level −0.6 kcal/mol). PCS1050 is displayed for comparison.

binding pocket are quite different. It seems improbable that a ligand-based method would be able to predict this alignment correctly.

As in the previous example, the known crystal structures of AChE were taken as positive controls for the performance of the AUTODOCK approach. The same parameters were taken as for the docking of the estrogen receptor ligands. First results we derived from the AUTODOCK/YETI calculations showed that the program was able to predict the experimentally determined positions of the inhibitors correctly. However, the experimentally observed complexes possessed not the lowest interaction energy. An exact analysis of the crystal structures of AChE showed that various water molecules are present in the binding pocket and are involved in the binding process. The analysis showed further that some of the water molecules are found in all four complexes in very similar places, independently of the size of the inhibitors. Therefore, we included these six conserved water molecules in our docking studies.

An excellent agreement between the calculated complexes and the crystal structures was observed when we considered the six structurally conserved water molecules during our docking studies. Not only are the RMSD between theoretically predicted and experimentally determined positions quite low (tacrine: 0.28 Å; huperzine: 0.51 Å; edrophonium: 0.71 Å; decamethonium: 1.15 Å), but also the positions found in the X-ray structure are in all cases those with the lowest interaction energy.

Encouraged by these results we applied the developed procedure to our data set of 48 aminopyridazine

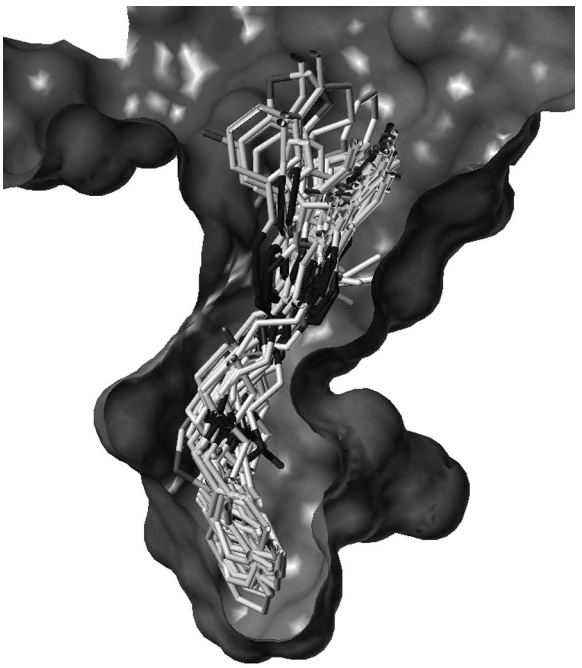


Fig. 11. This shows the superposition of all investigated inhibitors in the active site as obtained by the docking experiment. The protein surface is displayed for comparison. The positions of the protonated nitrogen atoms and the aminopyridazine groups are comparable for all the potent inhibitors.

inhibitors. Since the aminopyridazine derivatives possess a comparable size as decamethonium, and it is likely that they interact in a similar way with the binding site, we took the protein structure from the AChE-decamethonium complex for the docking experiment. Fig. 9 shows the predicted position of PCS 1050, a potent inhibitor, in comparison to the position of decamethonium observed in the corresponding crystal structure. The hydrophobic parts of the aminopyridazine inhibitors interact with an aromatic residue at the bottom of the gorge (Trp84), with aromatic residues in the middle of the gorge (Phe330, Phe331 and Tyr334) and with two aromatic residues at the entrance of the gorge (Trp279 and Tyr70). The benzyl ring of the inhibitor displays classic $\pi-\pi$ stacking with the aromatic ring of Trp84. It thus occupies the binding site for the quarternary ligands. The charged nitrogen of the piperidine makes a cation$-\pi$ interaction with Phe330 and electrostatic interactions with Tyr121.

Additional interactions of the piperidine ring occur with Phe331 and Tyr334. No direct hydrogen bonds were observed between the polar groups of the inhibitor and the binding site. Similar binding modes were observed for the other inhibitors [48]. To further validate the docking results we analyzed the binding pocket using the well-known program GRID [33]. GRID generates a contour map of the interaction energy versus the three-dimensional position of the probe with respect to the crystal structure of the protein. This information can lead to the prediction of how various functional groups of the inhibitors will interact in a specific region within the active site. Several probes were used to analyze the active site of AChE. The results were compared with the positions of inhibitors. We observed good agreement between the positions of the cationic head of the inhibitors and the contour maps obtained using the cationic trimethylammonium probe, as well as, between the location of the hydrophobic parts and the contour maps obtained using the hydrophobic DRY probe (as an example, the results for the DRY probe are shown in Fig. 10). This agreement further supported our docking results.

The alignment of all inhibitors, obtained by the docking procedure, is displayed together with the molecular surface of the binding pocket in Fig. 11. This alignment was further used as input for a comparative molecular field analysis using the water probe for the generation of the molecular field. Here 48 aminopyridazine derivatives [48] were included in a GRID/GOLPE analysis aimed to obtain information about the regions around the ligands which correlate with a variation in biological activity. The same computational steps as applied for the estrogen receptor ligands were applied to the AChE inhibitors. Using the SRD/FFD procedure within GOLPE we were able to reduce the number of variables from 16589 to 1238. Fig. 12 shows a plot of the experimental against calculated values and the values of the squared correlation coefficients (r^2) and of the squared coefficients (q^2) for different model dimensionalities (the experimental activities have been determined using AChE from Torpedo californica and are expressed as $-\log IC_{50}$ values [55]). Three components were found to be significant ($q^2 = 0.851$ and SDEP $= 0.33$). The conventional r^2 of this analysis is 0.972. This means, the model explains approximately 97% of the variance in ligand binding of the investigated

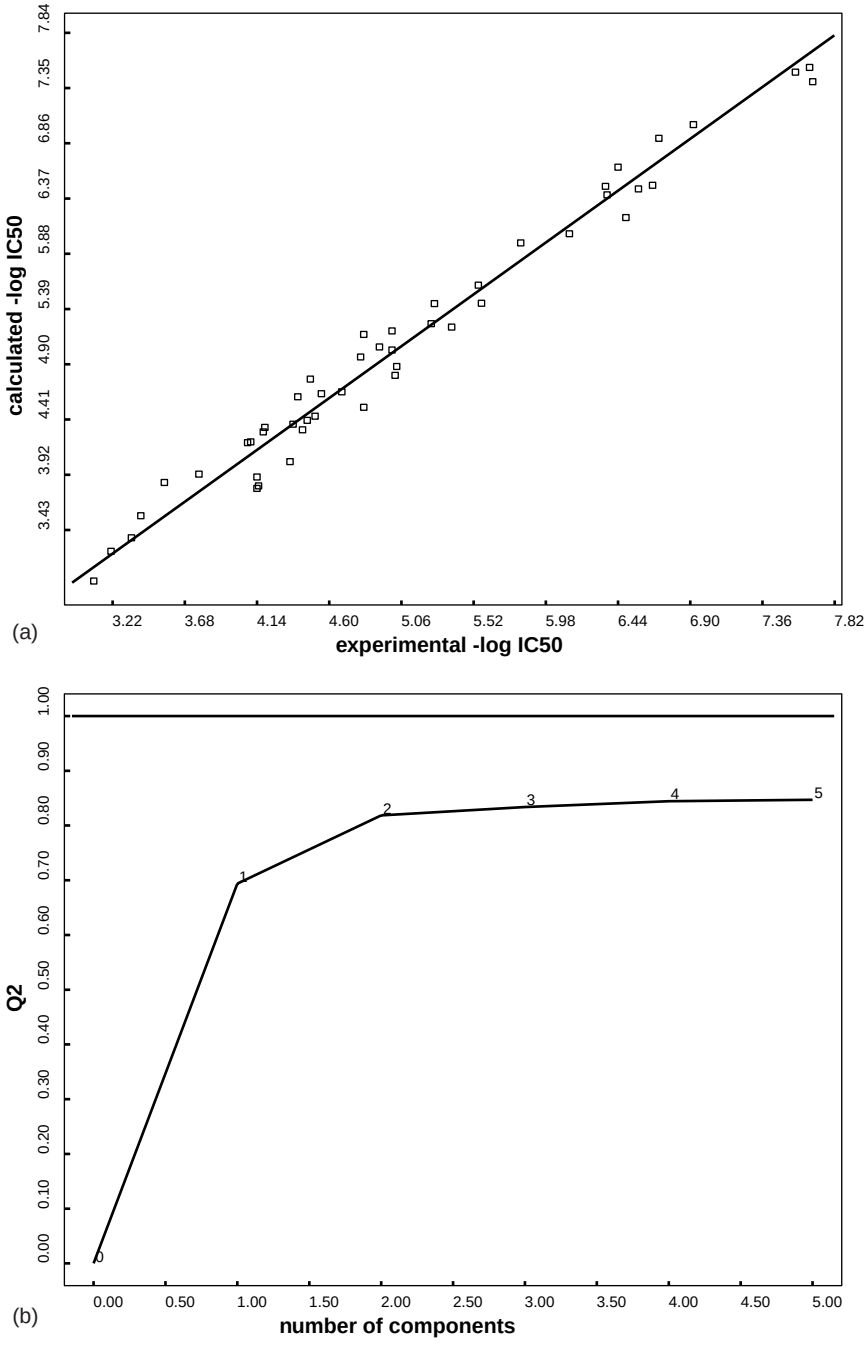


Fig. 12. GRID/GOLPE results for the AChE inhibitors: (a) calculated vs experimental activity; and (b) cross-validated squared correlation coefficients (q^2) for different model dimensionalities.

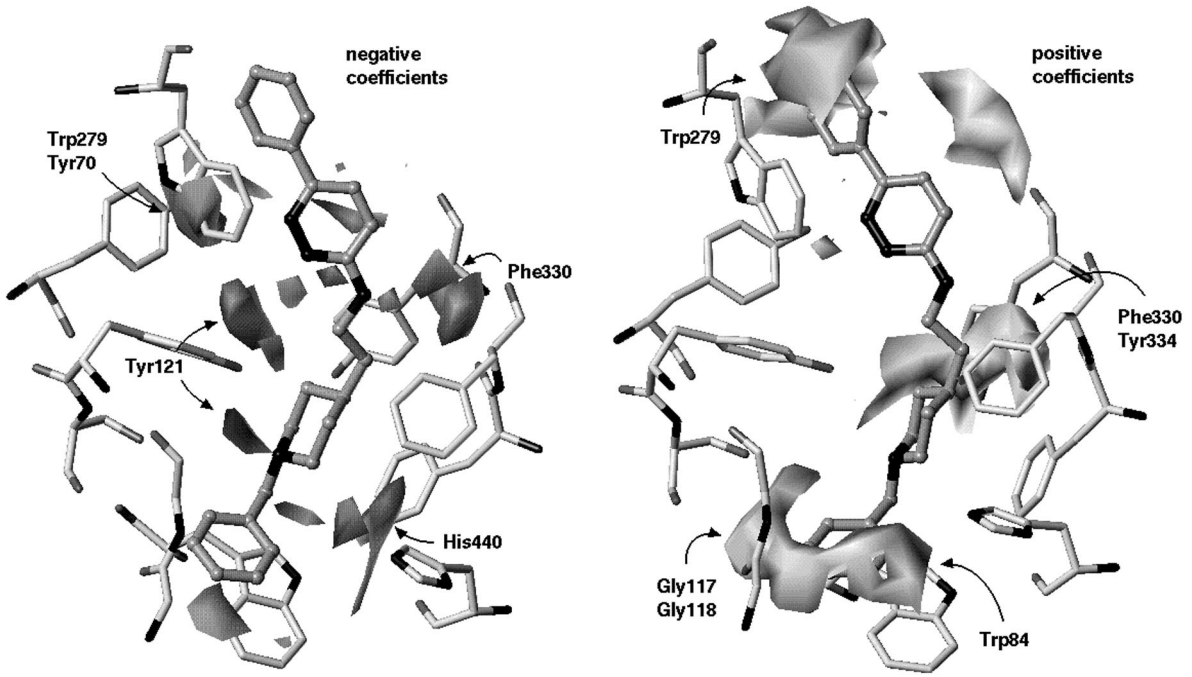


Fig. 13. Comparison between the PLS coefficient maps and the location of important residues in the binding pocket (indicated by the arrows).

compounds. The model expresses also excellent predictive ability, indicated by the high correlation coefficient of $q^2 = 0.830$ obtained by using the leave-20%-out cross-validation procedure.

Since the three-dimensional structure of our target is known, we were able to analyze the quality of the developed model by comparing the PLS coefficient maps of the inhibitors with the architecture of the active site. The regions, which the model indicates as important for the activity, should be close to the residues present in the binding pocket. Fig. 13 shows on the left side the negative PLS coefficient maps and on the right side the positive PLS coefficient maps. Since we used the water probe the positive contour maps indicate the areas where polar interaction decrease activity and the negative contour maps show the regions where polar interaction increase activity. In general we observed good agreement between the maps and the positions of important amino acid residues in the active site. The three main positive fields are close to the important aromatic residues in the gorge. The negative maps are more widely distributed, but also for these maps a clear correlation was found between the location of the maps and the position of polar amino acid residues.

Since the position of each inhibitor in the active site was calculated automatically the virtual testing of new compounds—not synthesized so far—seems to be a promising method for the design of new acetylcholinesterase inhibitors.

3. Conclusion

In this contribution we showed that the combination of ligand-based and receptor-based methods could lead to highly predictive and meaningful QSAR models. Besides the good predictiveness, the received models are also able to point out which interaction sites in the binding pocket might be responsible for the variance in biological activities.

In this context, it must be considered that a PLS analysis indicates only where a variation in the interaction fields is correlated with a variation in the biological activities. If all molecules of a data set would show a certain important interaction with the receptor, indicated by similar interaction energy at a particular

grid point for all compounds, this would not be reflected by the resulting PLS model. Thus the degree of correspondence depends strongly on the structural diversity of the studied ligands. If one considers these circumstances, important information can be obtained from a comparison of the contour maps and the binding site, which can then be integrated in the drug design process.

In the last decade, structure-based methods have become major tools in drug design, including lead finding and optimization [5]. It has also been shown, that structure-based methods are able to predict fairly accurate the position of ligands in receptor binding sites. Apart from the accurate prediction of experimental data, modern docking methods become ever more efficient. Meanwhile docking programs are developed, which can perform the docking of highly flexible ligands in a few minutes on a modern workstation [11,32]. The major problem is still the prediction of binding affinities, probably limited by the approximation used in today's force field methods. The application of 3D-QSAR methods—such as the GRID/GOLPE procedure—may facilitate the prediction of binding affinities if one has a series of compounds which bind in a similar way to a target protein.

Since a multivariate QSAR analysis considers only the information, which applies to the considered data set, advantages are offered in comparison to the more rigorous methods. The rigorous methods have to consider all influences on ligand binding, and must calculate the corresponding amounts correctly. We guess that a multivariate QSAR analysis is able to provide a kind of scoring function valid for a particular data set. Since the reported combined strategy is able to rapidly predict biological affinity, the method can be applied to large ligand series. As long as no methods are developed, which are able to solve the affinity prediction problem, structure-based 3D-QSAR is an exciting strategy for future drug design studies.

Acknowledgements

The authors thank Birgit Rau for technical assistance. We especially want to mention the contribution of Jean-Marie Contreras, Yveline Rival and Camille G. Wermuth (Universite Louis-Pasteur, Strasbourg) whose work has been the basis for the SAR data of the acetylcholinesterase inhibitors used in this study.

References

[1] C. Hansch, A. Leo, Exploring QSAR. Fundamentals and Applications in Chemistry and Biology, American Chemical Society, Washington, DC, 1995.
[2] H. Kubinyi, QSAR and 3D-QSAR in drug design, Drug Discovery Today 2 (1997) 457–467.
[3] R.D. Cramer III, D.E. Patterson, J.D. Bunce, Comparative molecular field analysis (CoMFA) 1. Effect of shape on binding of steroids to carrier proteins, J. Am. Chem. Soc. 110 (1988) 5959–5967.
[4] H. Kubinyi (Ed.), 3D-QSAR in Drug Design. Theory, Methods and Applications ESCOM Science Publishers BV, Leiden, 1993.
[5] J.D. Kuntz, Structure-based strategies for drug design and discovery, Science 257 (1992) 1078–1082.
[6] K. Kim, Non-linear dependencies in CoMFA, J. Comput. Aided Mol. Des. 7 (1993) 71–82.
[7] G. Folkers, A. Merz, D. Rognan, CoMFA: scope and limitations, in: H. Kubinyi (Ed.), 3D-QSAR in Drug Design. Theory, Methods and Applications, ESCOM Science Publishers BV, Leiden, 1993, pp. 583–618.
[8] G. Klebe, U. Abraham, On the prediction of binding properties of drug molecules by comparative molecular field analysis, J. Med. Chem. 36 (1993) 70–80.
[9] H. Van de Waterbeemd, B. Testa, G. Folkers, Computer-Assisted Lead Finding and Optimization. Current Tools for Medicinal Chemistry, Verlag Helvetica Chimica Acta, Basel, Switzerland, 1997.
[10] H.-D. Höltje, G. Folkers, in: R. Mannhold, H. Kubinyi, H. Timmerman (Eds.), Methods and Principles in Medicinal Chemistry: Vol. 5. Molecular Modeling—Basic Principles and Applications, VCH Verlags-gesellschaft, Weinheim, Germany, 1997.
[11] B. Kramer, M. Rarey, T. Lengauer, CASP2 Experiences with docking flexible ligands using FlexX, Proteins, Suppl. 1 (1997) 221–225.
[12] G.M. Morris, D.S. Goodsell, R. Huey, A.J. Olson, Distributed automatic docking of flexible ligands to proteins, J. Comput. Aided Mol. Des. 8 (1994) 243–256.
[13] T.P. Lybrand, Ligand–protein docking and rational drug design, Curr. Opin. Struct. Biol. 5 (1995) 224–228.
[14] E. Meng, B.K. Shoichet, I.D. Kuntz, Automated docking with grid-based energy evaluation, J. Comput. Chem. 13 (1992) 505–524.
[15] J.R.H. Tame, Scoring functions: a view from the bench, J. Comput. Aided Mol. Des. 13 (1999) 99–108.
[16] H.J. Böhm, Prediction of binding constants of protein ligands: a fast method for the prioritization of hits obtained from de-novo design or 3D database search programs, J. Comput. Aided Mol. Des. 12 (1998) 309–323.
[17] H.J. Böhm, The development of a simple empirical scoring function to estimate the binding constant for a protein–ligand

complex of known three-dimensional structure, J. Comput. Aided Mol. Des. 8 (1994) 243–256.
[18] C. Lemmen, T. Lengauer, G. Klebe, FLEXS: a method for fast flexible ligand superposition, J. Med. Chem. 41 (1998) 4502–4520.
[19] C.L. Waller, T.I. Oprea, A. Giolitti, G.R. Marshall, Three-dimensional QSAR of human immunodeficiency virus (i) protease inhibitors. 1. A CoMFA study employing experimentally-determined alignment rules, J. Med. Chem. 36 (1993) 4152–4160.
[20] S.A. De Priest, D. Mayer, C.B. Naylor, G.R. Marshall, 3D-QSAR of angiotensin-converting enzyme and thermolysin inhibitors: a comparison of COMFA models based on deduced and experimentally determined active-site geometries, J. Am. Chem. Soc. 115 (1993) 5372–5384.
[21] S.J. Cho, M.L. Garsia, J. Bier, A. Tropsha, Structure-based alignment and comparative molecular field analysis of acetylcholinesterase inhibitors, J. Med. Chem. 39 (1996) 5064–5071.
[22] M. Pastor, G. Cruciani, K. Watson, A strategy for the incorporation of water molecules present in a ligand binding site into a three-dimensional quantiative structure–activity relationship analysis, J. Med. Chem. 40 (1997) 4089–4102.
[23] R.J. Vaz, L.R. McLean, J.T. Pelton, Evaluation of proposed modes of binding of (2S)-2-[4-[[(3S)-1-acetimidoyl-3-pyrrolidinyl]oxyl]phenyl]-3-(7-amidino-2-naphtyl)-propanoic acid hydrochloride and some analogs to factor Xa using a comparative molecular field analysis, J. Comput. Aided Mol. Des. 12 (1998) 99–110.
[24] R.M. Evans, The Steroid and thyroid hormone receptor superfamily, Science 240 (1988) 889–895.
[25] E. Von Angerer, The Estrogen Receptor as a Target for Rational Drug Design, Landes, Austin, USA, 1995.
[26] F.J. Zeelen, Medicinal Chemistry of Steroids, Elsevier, Amsterdam, 1990.
[27] G.M. Anstead, K.E. Carlson, J.A. Katzenellenbogen, The estradiol pharmacophore: ligand structure–estrogen receptor binding affinity relationships and a model for the receptor binding site, Steroids 62 (1996) 268–303.
[28] H.-D. Höltje, N. Dall, A molecular modelling study on the hormone binding site of the estrogen receptor, Pharmazie 48 (1993) 243–246.
[29] D.F.V. Lewis, M.G. Parker, R.J.B. King, Molecular modeling of the human estrogen receptor and ligand interaction based on site-directed mutagenesis and amino acid sequence homology, J. Steroid. Biochem. Mol. Biol. 52 (1995) 55–65.
[30] R.A. Goldstein, J.A. Katzenellenbogen, Z.A. Luthey-Schulten, D.A. Seielstad, P.G. Wolynes, Three-dimensional model for the hormone binding domains of steroid receptors, Proc. Natl. Acad. Sci. USA 90 (1993) 9949–9953.
[31] L. Lemesle-Varloot, T. Ojasoo, J.P. Mornon, J.P. Raynaud, A Model for the determination of the 3d-spatial distribution of the functions of the hormone-binding domain of receptor that bind 3-keto-4-ene steroids, J. Steroid Bicohem. Mol. Biol. 41 (1992) 369–388.
[32] J.M. Wurtz, U. Egner, N. Heinrich, D. Moras, A. Mueller-Fahrnow, Three-dimensional models of estrogen receptor ligand binding domain complexes, based on related crystal structures and mutational and structure–activity relationships data, J. Med. Chem. 41 (1998) 1803–1814.
[33] B.R. Sadler, S.J. Cho, K.S. Ishaq, K. Chae, K.S. Korach, Three-dimensional structure-activity relationship study of nonsteroidal estogen receptor ligands using the comparative molecular field analysis/cross-validated r^2-guided region selection approach, J. Med. Chem. 41 (1998) 2261–2267.
[34] A.M. Brzozowski, A.C. Pike, Z. Dauter, R.E. Hubbard, T. Bonn, O. Engstrom, L. Ohman, G.L. Greene, J.A. Gustafsson, M. Carlquist, Molecular basis of agonism and antagonism in the estrogen receptor, Nature 389 (1997) 753–758.
[35] A.K. Shiau, D. Barstad, P.M. Loria, L. Cheng, P.J. Kushner, D.A. Agard, G.L. Greene, The structural basis of estrogen receptor/coactivator recognition and the antagonism of this interaction by tamoxifen, Cell 95 (1998) 927–937.
[36] D.S. Goodsell, G.M. Morris, A.J. Olson, Automated docking of flexible ligands: applications of AutoDock, J. Mol. Recognit. 9 (1996) 1–5.
[37] M.J. Rao, A.J. Olson, Modelling of factor Xa-inhibitor complexes: a computational flexible docking approach, Prot. Struct. Funct. Gen. 34 (1999) 173–183.
[38] P.J. Goodford, A computational procedure for determining energetically favourable binding sites on biologically important macromolecules, J. Med. Chem. 28 (1985) 849–857.
[39] A. Vedani, D.W. Huhta, A new force field for modelling metalloproteins, J. Am. Chem. Soc. 112 (1990) 269–280.
[40] A. Vedani, J.D. Dunitz, Lone-pair directionality of H-bond potential functions for molecular mechanics calculations: the inhibition of human carbonic anhydrase II by sulfonamides, J. Am. Chem. Soc. 107 (1985) 7653–7658.
[41] N.R. Taylor, M. von Itzstein, A structural and energetics analysis of the binding of a series of N-acetylneuraminic-acid-based inhibitors to influenza virus sialidase, J. Comput. Aided Mol. Des. 10 (1996) 233–246.
[42] H. Jiang, K. Chen, Y. Tang, J. Chen, Q. Li, Q. Wang, R. Ji, Molecular modelling and 3D-QSAR studies on the interaction mechanism of tripeptidyl thrombin inhibitors with human α-thrombin, J. Med. Chem. 40 (1997) 3085–3090.
[43] V.N. Viswanadhan, M.R. Reddy, A. Wlodawer, M.D. Varney, J. Weinstein, An approach to rapid estimation of relative binding affinities of enzyme inhibitors: application to peptidomimetic inhibitors of the human immunodeficiency virus type I protease, J. Med. Chem. 39 (1996) 705–712.
[44] M. Baroni, G. Constantino, G. Cruciani, D. Riganelli, R. Valigli, S. Clementi, Generating optimal linear PLS estimations (GOLPE): an advanced chemometric tool for handling 3D-QSAR problems, Quant. Struct.–Act. Relat. 12 (1993) 9–20.
[45] G. Cruciani, K. Watson, Comparative molecular field analysis using GRID force field and GOLPE variable selection methods in a study of inhibitors of glycogen phosphorylase b, J. Med. Chem. 37 (1994) 2589–2601.
[46] T.I. Oprea, A.E. Garcia, Three-dimensional quantitative structure activity relationships of steroid aromatase inhibitors, J. Comput. Aided Mol. Des. 10 (1996) 186–200.
[47] W. Sippl, Receptor-based three-dimensional quantitative structure-activity relationships of estrogen receptor ligands,

J. Comput. Aided. Mol. Des., submitted for publication, 1999.

[48] W. Sippl, J.M. Contreras, Y. Rival, C.G. Wermuth, Comparative molecular field analysis of aminopyridazine acetylcholinesterase inhibitors, Proceedings of the 12th European Congress on Quantitative Structure–Activity Relationships Copenhagen 23-28 (1998) 1998.

[49] L. Parnetti, U. Senin, P. Mecocci, Cognitive enhancement therapy for Alzheimer's disease. The way forward, Drugs Future 53 (1997) 752–768.

[50] E. Giacobini, From molecular structure to alzheimer therapy, Jpn. J. Pharmacol. 74 (1997) 225–241.

[51] M. Brufani, L. Filocamo, S. Lappa, A. Maggi, New acetylcholinesterase inhibitors, Drugs Future 22 (1997) 397–410.

[52] C.G. Wermuth, Aminopyridazines—an alternative route to potent muscarinic agonists with noccholinergic syndrome, Farmaco, Ed. Sci. 48 (1993) 253–274.

[53] C.G. Wermuth, G. Schlewer, J.J. Bourguignon, G. Maghioros, M.J. Bouchet, C. Moire, J.P. Kann, P. Worms, K. Biziere, 3-Aminopyridazine derivatives with atypical antidepressant serotonergic, and dopaminergic activities, J. Med. Chem. 32 (1989) 528–537.

[54] C.G. Wermuth, J.M. Contreras, S. Chayer, J. Pinto, P. Guilbaud, Y. Rival, J.J. Bourguignon, 3-Aminopyridazines as a source of acetylcholinesterase inhibitors, Acta Pharm. Hung. (1996) S3–S8.

[55] J.M. Contreras, Y. Rival, S. Chayer, J.J. Bourguignon, C.G. Wermuth, Aminopyridazines as acetylcholinesterase inhibitors, J. Med. Chem. 42 (1999) 730–741.

[56] J.L. Sussman, I. Silman, Atomic structure of acetylcholinesterase from torpedo californica: a prototypic acetylcholine-binding protein, Science 253 (1991) 872–879.

[57] M. Harel, J.L. Sussman, Quaternary ligand binding to aromatic residues in the active site gorge of acetylcholinesterase, Proc. Natl. Acad. Sci. USA 90 (1993) 9031–9035.

[58] M.L. Raves, M. Harel, Y.P. Pang, I. Silman, A.P. Kozikowski, J.L. Sussman, Structure of acetylcholinesterase complexed with the nootropic alkaloid huperzine A, Nat. Struct. Biol. 4 (1997) 57–63.