Journal of Computer-Aided Molecular Design, **14:** 559–572, 2000.
KLUWER/ESCOM

Receptor-based 3D QSAR analysis of estrogen receptor ligands – merging the accuracy of receptor-based alignments with the computational efficiency of ligand-based methods*

Wolfgang Sippl
Institute for Pharmaceutical Chemistry, Heinrich-Heine-University Düsseldorf, D-40225 Düsseldorf, Germany (Phone: ++49 211 8113848, Fax: ++49 211 8113847; E-mail: sippl@pharm.uni-duesseldorf.de)

Received 30 June 1999; Accepted 18 February 1999

Key words: AutoDock, CoMFA, docking, estrogen receptor, GOLPE, GRID, prediction of binding affinity, 3D QSAR

Summary

One of the major challenges in computational approaches to drug design is the accurate prediction of binding affinity of biomolecules. In the present study several prediction methods for a published set of estrogen receptor ligands are investigated and compared. The binding modes of 30 ligands were determined using the docking program AutoDock and were compared with available X-ray structures of estrogen receptor-ligand complexes. On the basis of the docking results an interaction energy-based model, which uses the information of the whole ligand-receptor complex, was generated. Several parameters were modified in order to analyze their influence onto the correlation between binding affinities and calculated ligand-receptor interaction energies. The highest correlation coefficient ($r^2 = 0.617$, $q^2_{\mathrm{LOO}} = 0.570$) was obtained considering protein flexibility during the interaction energy evaluation. The second prediction method uses a combination of receptor-based and 3D quantitative structure-activity relationships (3D QSAR) methods. The ligand alignment obtained from the docking simulations was taken as 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 (SRD) variable selection, a significant and robust model was obtained ($r^2 = 0.991$, $q^2_{\mathrm{LOO}} = 0.921$). The predictive ability of the established model was further evaluated by using a test set of six additional compounds. The comparison with the generated interaction energy-based model and with a traditional CoMFA model obtained using a ligand-based alignment ($r^2 = 0.951$, $q^2_{\mathrm{LOO}} = 0.796$) indicates that the combination of receptor-based and 3D QSAR methods is able to improve the quality of the underlying model.

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 ones. 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 inhibitors or receptor ligands. For the direct receptor-based approach, including molecular docking and advanced molecular dynamics simulations, the 3D structure of a target enzyme or even a receptor-effector complex is required with

*Dedicated to Prof. Dr. H.-D. Höltje on the occasion of his 60th birthday.

atomistic resolution, generally determined by either X-ray crystallography, NMR spectroscopy or protein homology model building [5].

3D QSAR methods, especially CoMFA, are 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 yields 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. 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 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 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 structures 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.

Receptor-based methods nowadays are 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 recently described in the literature [11–14]. The direct methods yield important information concerning the spatial orientation of the ligands at 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 a huge amount of possible ligand-receptor complexes, it is impossible to determine a priori which ligand conformation represents the bioactive one. The problem of predicting affinity has generated considerable interest in developing new methods to calculate ligand affinity reliably for a widely diverse series of molecules [15–19]. For the calculation of ligand-receptor interaction energies, most approaches rely on molecular mechanics force fields that represent van der Waals and electrostatic interactions on the basis of empirical potentials. Other approaches use more simple scoring functions rather than calculating the affinity by molecular mechanics equations [16–18]. 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 [19]. The main problem in affinity prediction is that the underlying molecular interactions are complex and various terms have to be taken into account to quantify the free energy of the interaction process. Only rigorous methods, such as the free energy pertubation methods, are at the moment able to correctly predict binding affinity. Since these methods are computationally very intensive, such methods cannot 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 [20–25]. It is quite appealing to combine the accuracy of a receptor-based alignment with the computational efficiency of a ligand-based method. According to this strategy the alignment is generated on the basis of the experimental or predicted position of molecules in a binding pocket and 3D QSAR programs are then used to calculate the binding affinity [20].

In the present study the combination of receptor-based and 3D QSAR methods will be analyzed and evaluated in order to accurately predict binding affinity. For this reason a set of 30 structurally diverse estrogen receptor ligands reported by Sadler et al. [26] was used as a test example. The objective of the selection was twofold. First, the biological activity of the estrogen receptor ligands reported in this study has been measured in the same laboratory under the same conditions, and second, the authors have performed a CoMFA analysis using a traditional ligand-based alignment rule. Therefore it was quite interesting to compare the results obtained by the receptor-based 3D QSAR approach with the original CoMFA results. In

addition, the 3D QSAR model will be compared with results obtained by correlating the biological activities with calculated ligand-protein interaction energies. Since the structure of the estrogen receptor has been solved recently the results of the present 3D QSAR analysis, given by the PLS coefficient maps, will be compared with chemical and geometrical properties of the ligand binding site.

Methods

Data set

For the investigation, structure-activity data for a series of 30 structurally diverse estrogen receptor ligands originally reported by Sadler et al. were selected [26]. The molecular structures of these 30 estrogen receptor ligands are represented in Table 1. The biological activities 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 of the RBA which were used in the present 3D QSAR analysis (see Table 3).

Crystal structures

In 1997 the 3D-structure of the estrogen receptor has been resolved by Brzozowski et al. [27]. Until now four X-ray structures of the receptor liganded with different molecules – two agonists, estradiol (**1**) and diethylstilbestrol (**2**) and two antagonists, raloxifen and 4-hydroxytamoxifen, have been published [27, 28]. To model estrogen receptor-ligand complexes, the coordinates of the human estrogen receptor-α-ligand binding domain liganded with estradiol and diethylstilbestrol (1ERE and 2ERD) were taken from the Protein Databank [29]. Polar hydrogen atoms were added and charges from AMBER [30] were loaded. The default charge states of titratable sites on the protein were chosen and all histidines were treated uncharged. AM1-ESP [31] charges were calculated for the ligands. All water molecules observed in the receptor crystal structure were deleted and the whole molecule was subjected to a minimization using the YETI force field [32] keeping all protein backbone atoms at fixed positions.

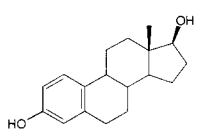
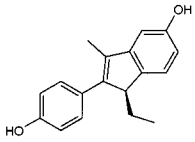
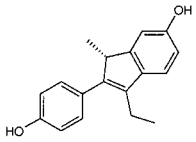
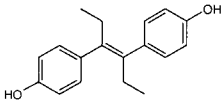
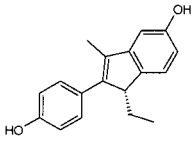
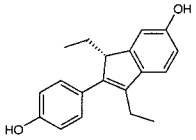
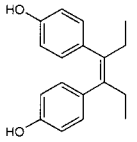
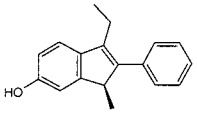
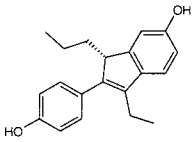
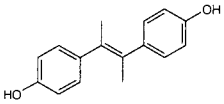
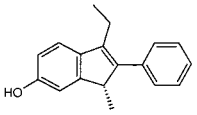
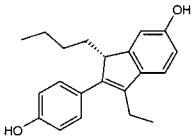
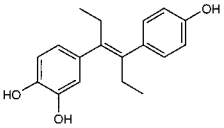
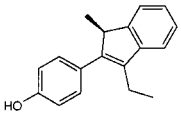
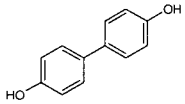
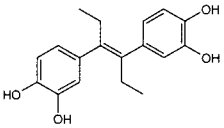
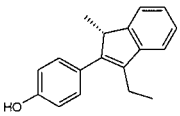
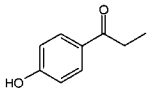
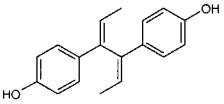
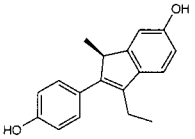
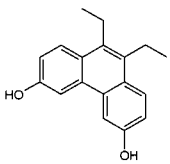
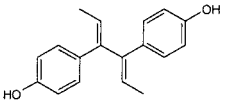
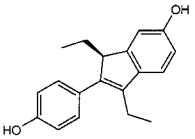
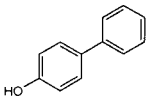
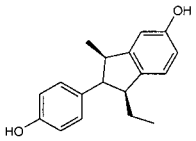
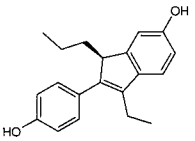
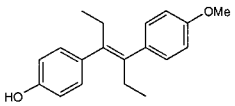
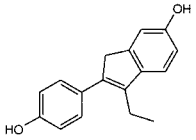
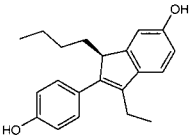
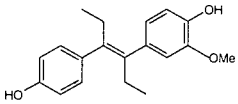
Docking simulations

The docking analysis was performed using the program AutoDock (version 2.4), which has been shown to successfully reproduce experimentally observed binding modes [12, 33, 34]. The program is described in detail elsewhere [12]. 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 [35]. The minimized ligand-free protein structures, as described above, were used as input structure for the docking simulations. 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, each of 50 cycles containing a maximum of 20 000 accepted and rejected steps. The simulated annealing procedure was started at high temperature (RT = 616 kcal/mol) and was decreased by a factor of 0.95 on each cycle. The 100 docked complexes were clustered with an RMSD tolerance of 0.7 Å. In a second step the complexes were further refined and the interaction energies were calculated using the YETI force field and a distance dependent dielectric constant ($D_{(r)} = 2r$) [32, 36, 37]. For this second step the first 20 complexes of the AutoDock energy ranking were selected. The protein structure was held fixed during the minimization, whereas the ligand was allowed to change its conformation and position in the binding pocket. Applying this minimization, the ligand conformation is relaxed into a neighbouring local energy minimum.

Molecular interaction fields

To validate the results obtained by the automated docking procedure, a series of GRID calculations [38] were performed. Program GRID is an approach to predict noncovalent interactions between a molecule of known three-dimensional structure (i.e., estrogen receptor) and a small group as a probe (representing chemical features of a ligand). The calculations were performed using version 16 of the GRID program and the crystal structures mentioned above. The calculations were performed on a cube (13 × 15 × 12 Å) including the binding pocket in order to search for binding sites complementary to the functional groups of the ligands. Several probes were used to study the active site of the estrogen receptor (hydroxyl, methyl and hydrophobic DRY probe). The calculated GRID contour maps were then viewed superimposed on the crystal structure of the estrogen receptor using the SYBYL 6.5 software [39].

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	

Interaction energy calculation

The correlation between biological activities and protein-ligand interaction energies was analyzed using the YETI force field and by variation of several parameters during the geometry optimization. The ligand-receptor complexes derived from the docking simulations were chosen for this investigation. The analysis was first performed with fixed protein geometry during the optimization process. Several sets of charges – Gasteiger-Hückel [40], AM1-ESP [31] and ab initio HF 6-31G [41] – were tested for their performance. A distance dependent dielectric constant was selected for all the calculations ($D_{(r)} = 2r$) [32]. In a second step the effect of protein flexibility during the optimization was analyzed by allowing atoms within 8 Å of the ligand to move during the optimization, while the rest of the protein was kept fixed (e.g. all amino acid residues close to the binding pocket and all ligand atoms were considered flexible in this protocol).

3D QSAR analysis

The GRID/GOLPE method was used within this study to perform a 3D QSAR analysis [35, 42]. The interaction field between the ligands and a water probe was calculated using the GRID program employing a grid spacing of 1 Å. The size of the grid box used for the calculation was defined in such a way that it extended approximately 4 Å beyond each of the molecules in each dimension (18 × 22 × 19 Å). A cutoff of +5 kcal/mol was applied in order to obtain a more symmetrical distribution of energy values. The GRID calculation 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 [43]. These variables were selected and eliminated applying the advanced pretreatment procedure and the SRD/FFD (Smart Region Definition/Fractional Factorial Design) variable selection within the GOLPE program (for a detailed description of this approach see [43–45]). Models obtained applying variable selection are in general of higher quality than models calculated without variable selection. The GOLPE advanced pretreatment eliminated those variables with absolute values smaller than 0.1 kcal/mol and variables with a standard deviation below 0.05. In addition, variables which take only two or three values were also removed. After this pretreatment the data set still contains 2200 variables. Subsequently, the SRD procedure was used to carry out the variable selection on groups of variables chosen according to their positions in 3D space [45]. Within the SRD procedure 400 seeds, a critical distance cutoff of 1.0 Å and a collapsing distance cutoff of 2.0 Å were used. The regions calculated were then used in an FFD variable selection procedure. The number of variables was reduced to 493 with major improvement of the quality of the model.

Model validation

To form the basis for a statistically significant 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 PLS 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 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 a second cross-validation, using five groups of approximately the same size in which the objects were assigned randomly, was performed. In this method 80% of the compounds are 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 of a model than the normal LOO procedure [46].

To further evaluate the predictive ability of the resulting model, a test set of six additional molecules was utilised (Figure 1). The test set includes the two phytoestrogens Coumestrol and Genistein, the phenolic compounds Bisphenol A and Hexestrol and the two steroids Estron and 17α-Estradiol [47]. The alignment of the six compounds was generated using the described AutoDock/Yeti procedure. Their activities were predicted using the generated GRID/GOLPE model.

Results and discussion

A detailed inspection of the crystal structures revealed that the estrogen receptor complexed with an agonist (estradiol or diethylstilbestrol) shows a nearly

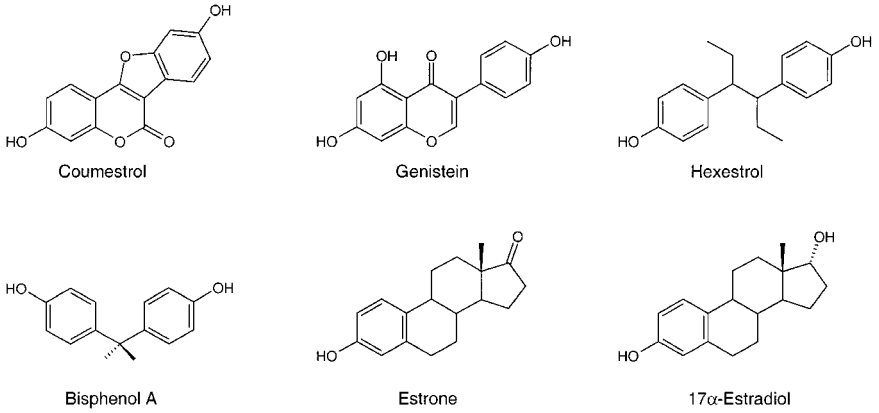


Figure 1. Compounds included in the test set.

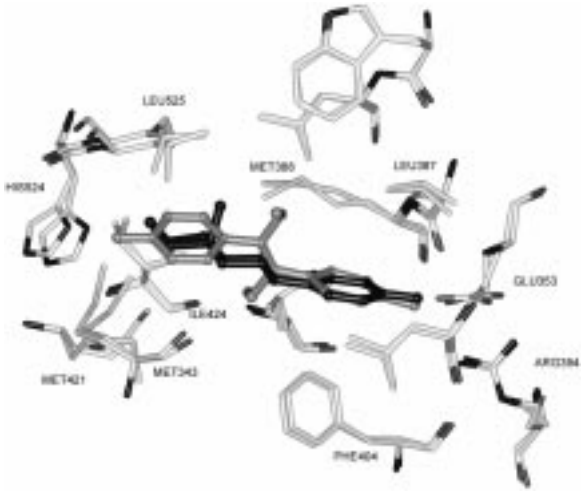


Figure 2. Comparison of the X-ray structures of the estrogen receptor liganded with estradiol (black) and diethylstilbestrol (grey). The two crystal structures were overlaid using the backbone atoms. Only the amino acid residues in proximity to the binding pocket are shown for clarity.

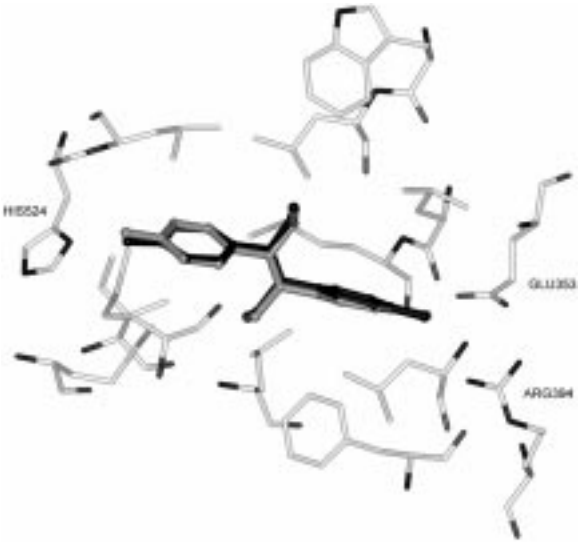


Figure 3. Comparison of the position of diethylstilbestrol calculated by the AutoDock/YETI procedure (black) and the one observed in the crystal structure (grey, RMSD = 0.37 Å).

identical three-dimensional structure. In contrast, the structure of the estrogen receptor liganded with an antagonist (raloxifen or 4-hydroxytamoxifen) shows major structural changes. Since all compounds included in the present study are agonists, only the crystal structures of the receptor-agonist complexes were selected for the investigation. The crystal structures were overlaid using their backbone atoms and are shown in Figure 2. Estradiol and diethylstilbestrol bind in a similar way to the receptor. One of the phenolic rings of diethylstilbestrol lies in the same position as the aromatic ring of estradiol. Both ligands form hydrogen bonds to the γ-carboxylate of Glu353, to the guanidinium of Arg394 and to the imidazole ring of His524. In addition, the ligands show van der Waals interactions with several residues at the binding pocket (Leu387, Met343, Leu346, Phe404, Met421, Ile424 and Leu525). The second hydroxyl group of diethylstilbestrol is located 1.7 Å from the position of the corresponding group of estradiol, but is still able to form a hydrogen bond to the imidazole of His524. The ethyl group of diethylstilbestrol, which projects perpendicularly from the plane of the aromatic systems, fits into an additional cavity formed by several hydrophobic amino acid residues. The only major conformational difference between both receptor-agonist complexes is the orientation of the side chains of His524 and Met421. Due to the different size of diethylstilbestrol

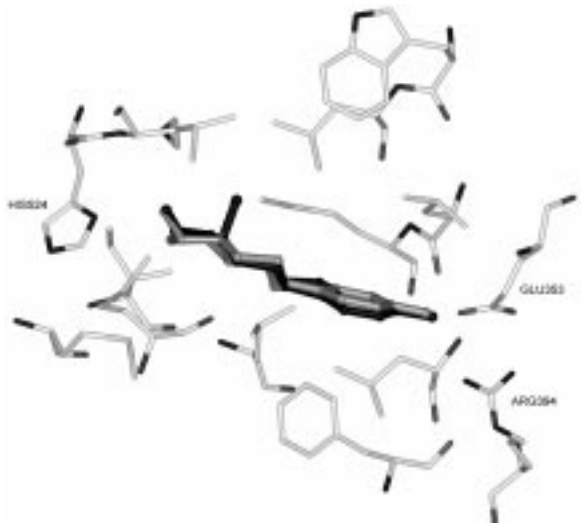


Figure 4. Comparison of the position of estradiol calculated by the AutoDock/YETI procedure (black) and the one observed in the crystal structure (grey, RMSD = 0.21 Å).

and estradiol these side chains adopt different packing orientations in response to each ligand.

The docking simulations were started using the two molecules, estradiol and diethylstilbestrol, for which the binding mode has been determined experimentally. These two ligands were taken as positive control to test the program used for docking. Since the geometry of the binding pocket exhibits small differences in the two crystal structures, both protein structures were taken as target for the docking procedure. AutoDock considers only the polar hydrogen atoms during the docking process, and uses simple potential functions for the calculation of the binding energy. Therefore, the obtained protein-ligand complexes were further refined applying a more sophisticated optimization procedure. All hydrogen atoms were added to the protein structure and the resulting complexes were minimized using the YETI force field. This force field has been validated on protein-ligand complexes and was therefore well suited for the present task [32]. During this optimization process only the ligand structure was allowed to relax. The protein-ligand interaction energies were then calculated using the refined complexes.

In Figures 3 and 4 the complexes showing the lowest energy within the AutoDock/YETI procedure for diethylstilbestrol and estradiol are overlaid with the corresponding crystal structures. The complex which was obtained using the corresponding crystal structure (1ERE for estradiol and 2ERD for diethylstilbestrol)

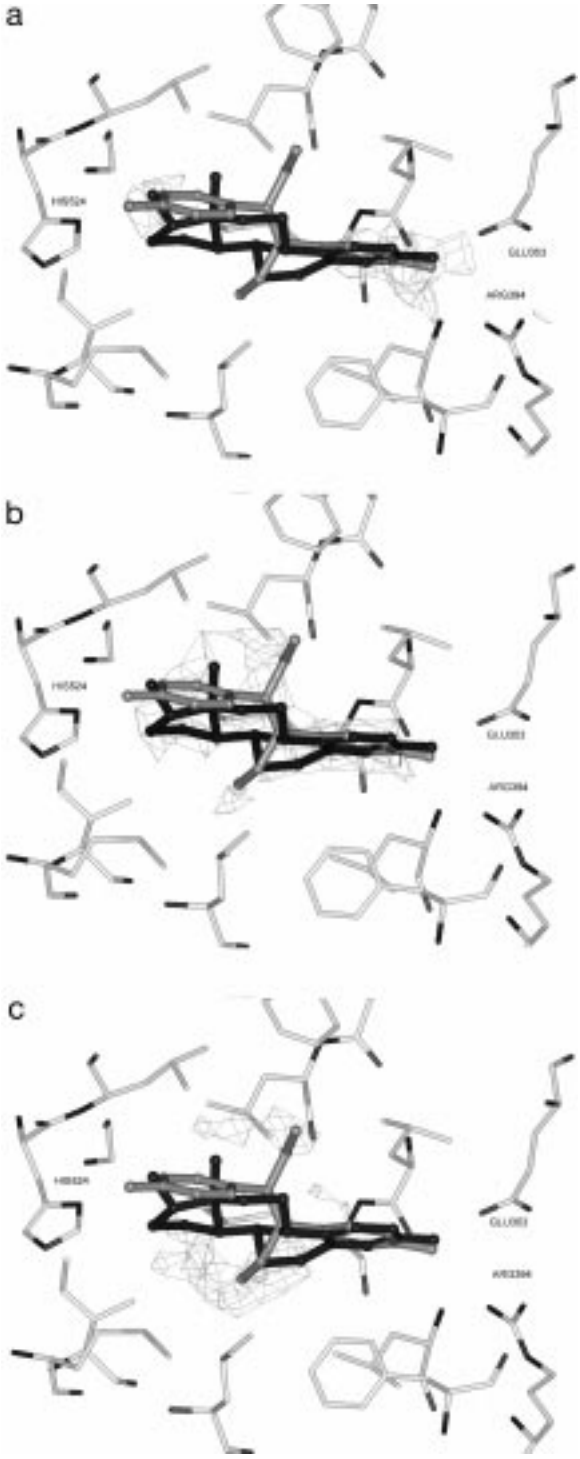


Figure 5. Comparison of the location of the ligands (diethylstilbestrol: grey, estradiol: black) in the binding pocket and the GRID interaction fields calculated (a) with the hydroxyl probe (contoured at an energy level of −5 kcal/mol), (b) with the methyl probe (contoured at an energy level of −2.5 kcal/mol) and (c) with the hydrophobic DRY probe (contoured at an energy level of −0.6 kcal/mol).

showed in both cases the lower interaction energy. Thus, AutoDock was successful in reproducing the experimentally found binding position and the YETI force field was able to correctly rank the obtained complexes (one may speak of reproduction if the root-mean square deviation (RMSD) is below 1–2 Å [12]). The RMSD values (all heavy atoms) between the observed and calculated position are 0.21 Å for estradiol and 0.37 Å for diethylstilbestrol. It turns out that the experimentally determined complexes were not correctly ranked using the AutoDock method solely. However, in both cases the experimentally determined complex was among the first five solutions found by AutoDock and the YETI optimization was successful in assigning the lowest interaction energy to the complex close to the crystal structure.

To further validate the results obtained by the automated docking procedure, the receptor binding pocket was analyzed using the program GRID. The hydroxyl, methyl and the hydrophobic DRY probes were used in order to identify and visualize the main interactions between the protein and the ligands. The two main regions of interaction for the hydroxyl probe correspond very well with the position of the aliphatic and aromatic hydroxyl groups of the potent ligands (Figure 5a). In an effort to study the areas of hydrophobic interaction the methyl and the hydrophobic DRY probe were tested on the binding pocket. Figure 5b shows the agreement between the molecular shape of diethylstilbestrol and estradiol and the interaction field of the methyl probe, contoured at -2.5 kcal/mol. The interaction field obtained with a methyl probe indicates mainly the van der Waals interactions and corresponds to the location and size of the non-polar parts of the ligands. The most negative areas obtained with the hydrophobic DRY probe (contoured at -0.6 kcal/mol) lie below and above the planar ring systems of the ligands and correspond to the positions of the alkyl groups of diethylstilbestrol and the other ligands (Figure 5c). The analysis showed that the GRID interaction fields are in close agreement with the experimentally determined positions of the ligands. They can therefore be used to check the results obtained by the AutoDock/YETI procedure.

In the next step the ligand-receptor complexes for the resulting 28 ligands were generated in the same way. For all ligands the complex which possessed the lowest interaction energy after the YETI optimization was selected. In most cases there was a clear discrimination between the interaction energies of the most favourable and the other complexes obtained for a particular ligand. In cases where two complexes possessed a similar interaction energy (within the range of 0.5 kcal/mol), the ligand conformation showing a better agreement with the calculated GRID interaction fields was selected. Figure 6 shows as an example the result of the docking simulation for the most potent ligand, the indenesterol derivative **18**. In comparison, diethylstilbestrol is shown in the experimentally derived position. The phenolic ring of compound **18** lies in the same position as the aromatic ring of diethylstilbestrol, forming a hydrogen bond to Glu353/Arg394. A second hydrogen bond is formed to the imidazole of His524, although the position of the hydroxyl group is different compared to the position of the corresponding group of diethylstilbestrol. The two ethyl substituents of ligand **18** occupy the same region in space as the alkyl chains of diethylstilbestrol. The position of the other ligands in the binding pocket are comparable. All ligands form a hydrogen bond to Glu353/Arg394 indicated by a similar position of the phenolic ring systems in the alignment. A second hydrogen bond to the imidazole of His524 is formed by the potent ligands. The derived alignment indicated further that the alkyl substituents of the ligands occupy similar regions in space corresponding to two hydrophobic cavities located on both sides of the planar ring system.

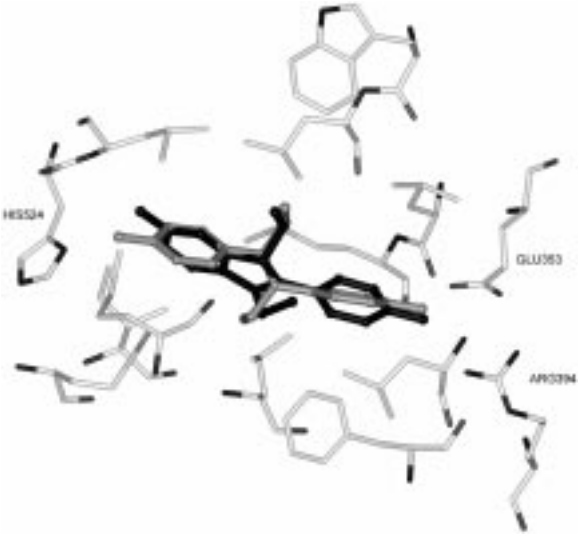


Figure 6. Superimposition of diethylstilbestrol (grey) and compound **18** (black) as obtained from the docking simulation.

In the next step the correlation between binding affinity and ligand-receptor interaction energy was investigated. We started with fixed receptor geometry

Table 2. Comparison of different optimization protocols (charge set, protein flexibility)

Model	r^{2}[a]	SDEC[b]	q^2_{LOO}[c]
Gasteiger-Hückel, receptor rigid	0.546	0.825	0.487
Gasteiger-Hückel, receptor flexible	0.568	0.836	0.502
AM1-ESP, receptor rigid	0.540	0.831	0.490
AM1-ESP, receptor flexible	0.617	0.758	0.570
HF-6-31G, receptor rigid	0.523	0.846	0.450
HF-6-31G, receptor flexible	0.561	0.812	0.499

[a] Correlation coefficient in fitting.
[b] Standard deviation of error of calculation.
[c] LOO cross-validated correlation coefficient.

during the geometry optimization and interaction energy calculation process. The statistical results are given in Table 2. The cross-validated correlation coefficients q^2 lie in the range between 0.45 and 0.49, depending on the chosen point charge model. No significant effects were observed when moving from Gasteiger to ab initio charges. Since the investigated compounds contain only carbon, hydrogen and oxygen atoms, even simple charge calculation methods work. Next, the protein flexibility during the geometry optimization process was studied by allowing atoms within 8 Å of the ligand to move during the optimization, while the rest of the receptor was kept fixed. Introducing protein flexibility improved the results slightly. The protocol with the highest correlation coefficient ($r^2 = 0.617$, SDEC $= 0.758$, $q^2_{LOO} = 0.570$) was obtained with AM1-ESP charges and considering protein flexibility.

The derived model is able to indicate trends between actual and predicted biological activities. However, the model cannot explain the differences in biological activity between the enantiomers of compounds **13/14**, **15/16** or **17/21** (see model C in Table 3). Also compared with the original CoMFA model of Sadler ($r^2 = 0.951$, SDEC $= 0.290$, $q^2_{LOO} = 0.796$, model A in Table 3), one recognises that the interaction energy model shows a less satisfactory prediction accuracy. Since solvation or entropic effects were not considered in the calculation of the binding energy, it is not surprising that the prediction of the activities is less accurate. Corresponding results applying the interaction energy approach have been reported recently by several groups [48–51].

Table 3. Actual and calculated activities (log RBA)

Comp.	Actual	Model A[a]	Model B[b]	Model C[c]
1	2.00	1.77	1.99	1.02
2	2.46	2.25	2.44	1.36
3	−0.10	−0.26	0.01	0.59
4	1.52	1.93	1.44	0.92
5	2.00	1.62	1.98	2.58
6	1.40	1.47	1.39	2.19
7	−0.52	−0.68	−0.60	1.31
8	1.30	0.88	1.40	1.41
9	0.30	0.38	0.36	0.65
10	1.14	1.90	1.07	0.67
11	2.00	1.94	2.01	0.47
12	2.15	1.90	2.36	1.97
13	0.26	−0.04	0.17	−0.21
14	−0.70	−0.24	−0.54	0.07
15	0.75	0.49	0.68	0.58
16	−0.05	−0.11	0.01	0.68
17	2.46	2.18	2.36	1.92
18	2.47	2.36	2.51	1.81
19	2.36	2.39	2.33	2.14
20	2.25	2.40	2.32	2.16
21	1.11	1.14	0.93	1.76
22	1.04	1.19	1.03	1.03
23	1.26	1.19	1.26	1.40
24	0.90	1.22	0.95	0.55
25	−1.70	−1.50	−1.76	−0.16
26	−1.00	−1.18	−0.96	−1.24
27	−0.80	−0.73	−0.80	−0.32
28	−1.70	−1.51	−1.67	−1.32
29	1.30	1.49	1.29	0.02
30	1.00	0.77	0.94	0.81

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

Another way to calculate binding affinity for a given set of molecules is the application of 3D QSAR techniques. The difference between receptor-based interaction and 3D QSAR models is that, while correlating binding affinity with interaction energy (in a receptor-based model), the contribution of each individual additive interaction is equally weighted before analysis, whereas 3D QSAR weights the contribution of the ligand-probe interactions at each grid point. The major problem of receptor-based interaction methods

is still the prediction of binding affinity, probably limited by the approximations used in today's force field methods. Several authors have pointed to the present limits of predicting binding affinity even from crystal structures of ligand-site complexes [52, 53]. Therefore, the 3D QSAR approach was applied to calculate the binding affinities of the estrogen receptor ligands.

The superimposition of the ligands derived from the molecular docking was taken as starting point for a comparative molecular field analysis using the strategy described in the Methods section. The LOO cross-validated q^2_{LOO} value for the initial model was 0.830 using five PLS components. The application of the SRD/FFD variable selection resulted in an improvement of the significance of the model. The analysis yielded a correlation coefficient with a cross-validated q^2_{LOO} of 0.921 using four PLS components with a standard error of prediction (SDEP) of 0.345 (Figure 7). The conventional r^2 of this analysis is 0.992 with a standard error of calculation (SDEC) of 0.112. This means that the model explains approximately 99% of the variance in ligand binding of the investigated compounds. The model is also robust, indicated by a high correlation coefficient of $q^2 = 0.900$ obtained by using the leave-20%-out cross-validation procedure.

The comparison of the derived model with the CoMFA results of Sadler, who reported a cross-validated q^2_{LOO} value of 0.796, indicates that the model constructed on the basis of the receptor structure supplies a better explanation of the biological activities ($q^2_{\mathrm{LOO}} = 0.921$). This is also indicated by the small deviation of the calculated from the experimental values (model B in Table 3). The original CoMFA model is for example not able to explain the activity of compound **10** or the difference in the activities of the two enantiomers **13** and **14**. The receptor-based approach supplies here a unique explanation.

Beside the ligands investigated by Sadler et al., other natural and synthetic compounds have been described in the literature which also bind at the estrogen receptor [47, 54, 55]. In order to test the predictive ability of our model some of these structurally diverse molecules were selected (Figure 1). Using the described AutoDock/Yeti procedure, the molecules were docked into the binding pocket and the resulting conformations were used for the prediction of the binding affinities. A standard error of prediction of 0.43 was obtained using the four component model (Table 4). The reasonable agreement between the predicted and observed RBAs substantiates the fitness of the GRID/GOLPE mode beyond the training set compounds. However, the presented method (which uses a rigid receptor structure during the docking simulation) is limited to molecules which bind to the same receptor conformation. Therefore, the derived model is not able to predict binding affinities for estrogen receptor antagonists such as raloxifen or tamoxifen.

Table 4. Observed vs. predicted activities for the test set

Compound	Observed	Predicted
Bisphenol A	−1.30	−0.57
Coumestrol	1.97	1.48
Estron	1.78	1.46
Genistein	0.70	0.66
Hexestrol	2.25	1.89
17α-Estradiol	1.76	1.55

One interesting feature of a GRID/GOLPE analysis is the possibility of translating back the PLS coefficients assigned to each variable to the 3D positions they occupy in real space [44]. These values can be contoured at a particular significant level and can be displayed as a grid plot of PLS coefficients. The contour coefficient maps indicate those areas in which the model has found a high correlation between the ligand-probe interaction energy and the biological activity. The water probe used for the calculation of the molecular field is a polar group with the ability to participate in hydrogen bonds and electrostatic interactions. Thus, areas containing negative coefficients correspond to areas in the binding pocket where energetically favourable interactions produce an increase in the activity. These regions correspond to the position of polar groups in the binding pocket. In contrast, an area with positive coefficients corresponds to a region where the presence of energetically favourable interactions decrease the biological activity [45]. Such regions can appear for several reasons. For example, they may indicate a hydrophobic region where the presence of a polar group is unfavourable. Furthermore, these regions may represent a ligand-receptor interaction that may only be present through a different binding mode, which is less favourable for the activity [24].

Since the structure of the estrogen receptor was known, it was quite interesting to compare the results of the present 3D QSAR analysis, given by the PLS coefficient maps, with the chemical and geometrical properties of the binding site. It is necessary to

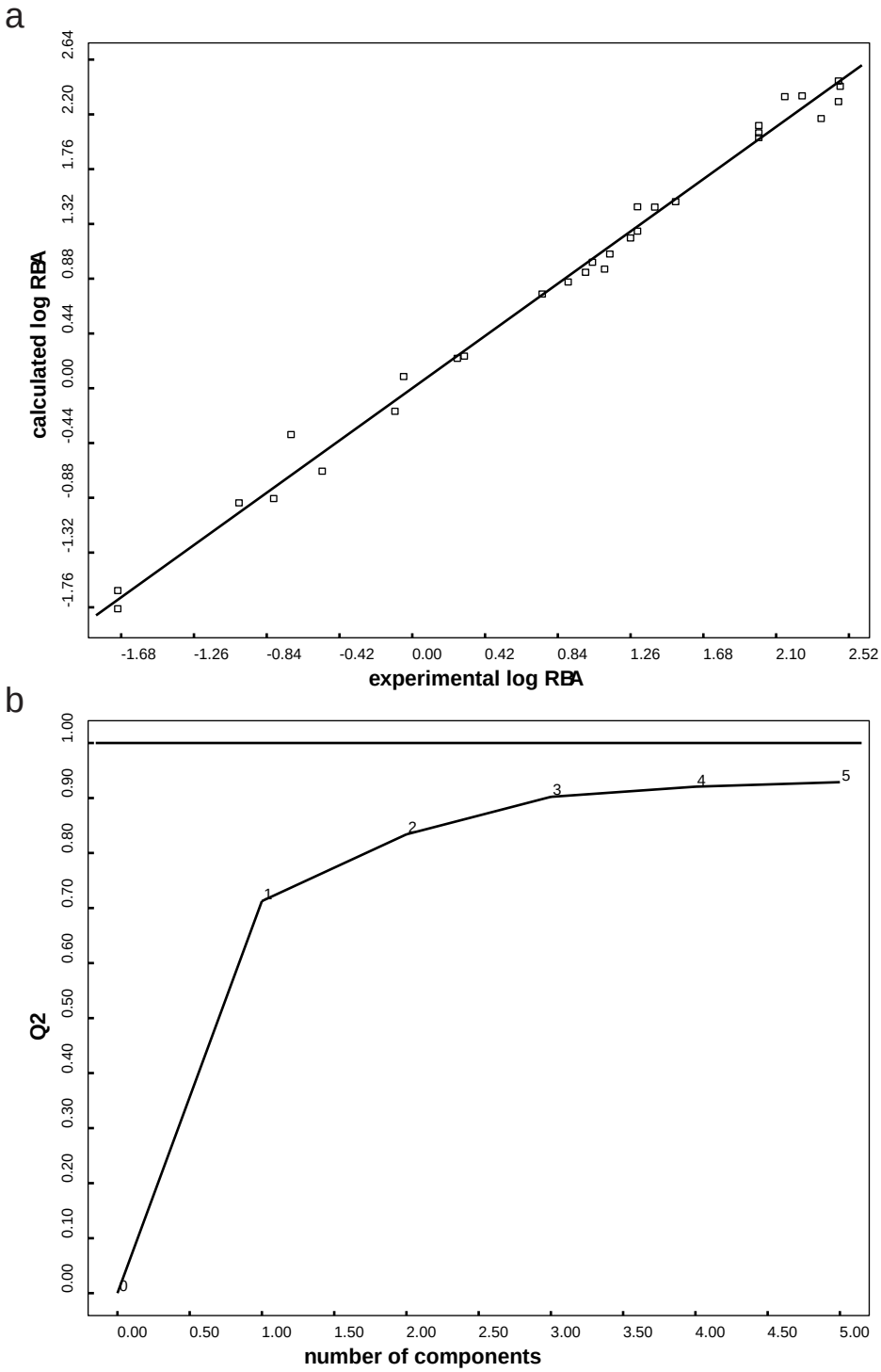


Figure 7. Statistical results for the resulting GRID/GOLPE model, (a) the calculated vs. experimental activity and (b) the cross-validated (LOO) squared correlation coefficients (q^2_{LOO}) for different model dimensionalities.

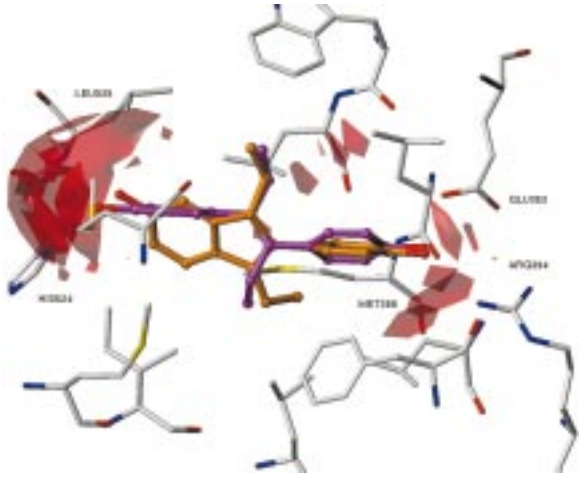


Figure 8. Comparison between the PLS coefficient maps, contoured at −0.004, and the amino acid residues located close to the binding pocket. The region around His524 and the backbone of Leu525 contains only negative coefficients indicating that polar interactions here would increase the activity (diethylstilbestrol is displayed in magenta, and the potent ligand – compound **17** – is coloured orange).

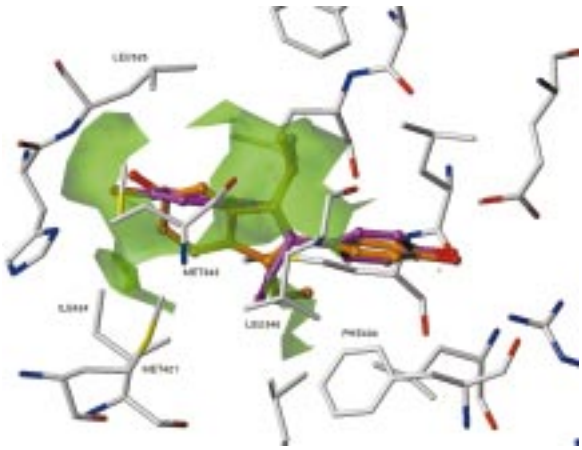


Figure 9. Comparison between the PLS coefficient maps, contoured at +0.004, and the amino acid residues located close to the binding pocket (same colour scheme as in Figure 8).

note that, in general, such a comparison should be attempted carefully. The PLS coefficient contour maps can by no means be regarded as a set of low resolution pictures of the binding site, since the contour maps reflect only those regions in space where the ligand-probe interaction energy is correlated with a variance of the biological activity. This has been demonstrated and extensively discussed by several authors [43–45, 56–58]. However, it provides an opportunity to interpret features indicated in the contour maps with respect to the protein environment.

The coefficient contour maps and the active site of the estrogen receptor were superimposed and compared with each other. Figures 8 and 9 show the grid plot of the PLS coefficients for the used water probe. The red contours in Figure 8 represent negative coefficients under −0.004 while the green contours in Figure 9 represent positive coefficients over 0.004. Since the water probe was used for the calculation, the positive contour maps indicate the areas where polar interactions decrease activity and the negative contour maps show the regions where polar interactions increase activity. In general an agreement between the maps and the positions of particular amino acid residues in the active site was observed. In particular, one recognises that a big positive field occupies a hydrophobic pocket close to Leu346, Met343 and Leu525. The most active ligands possess alkyl substituents which are able to occupy this pocket. Several smaller fields lie close to Phe404, Ile424 and Met421, indicating that the interactions between the hydrophobic parts of the ligands and these residues influence also ligand binding. The region around His524 and the backbone of Leu525 contains only negative coefficients, indicating that polar interactions would here increase the activity. Less pronounced fields are located close to Glu353, Arg394 and the polar backbone atoms of Met388. Since all studied ligands possess a hydroxyl group interacting with these residues, it is not surprising that this region does not contribute to an explanation of the variation in the biological activities.

Conclusions

In this study several prediction methods were examined on a data set of 30 estrogen receptor ligands. It was shown that the model derived by combining ligand-based and receptor-based methods yields a better explanation of the binding affinities than an interaction energy-based or a ligand-based CoMFA model. Beside the statistical significance and robustness, the receptor-based 3D QSAR model is 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 a similar interaction energy at a particular grid point for all compounds, this would not be re-

flected by the resulting PLS model. Thus the degree of correspondence depends strongly on the structural diversity of the training set. If one considers these circumstances, useful 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 [56, 57].

In the last decade, receptor-based methods have become major tools in drug design, including lead finding and optimization [5]. It has also been shown that receptor-based methods are able to predict fairly accurately the position of ligands in receptor binding sites. Apart from the accurately prediction of experimental data, modern docking methods become even 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, 33]. The major problem is still the prediction of binding affinities, probably limited by the approximations 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. Therefore, 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, combined receptor- and ligand-based 3D QSAR is an interesting strategy for future drug design studies.

Acknowledgements

The author wishes to thank Prof. S. Clementi and Dr G. Cruciani, University of Perugia, for providing the GOLPE software and Prof. H.-D. Höltje, Heinrich-Heine University Düsseldorf, for supporting the project.

References

1. Hansch, C. and Leo, A., Exploring QSAR. Fundamentals and Applications in Chemistry and Biology, American Chemical Society, Washington, DC, 1995.
2. Kubinyi, H., Drug Discovery Today, 2 (1997) 457.
3. Cramer III, R.D., Patterson, D.E. and Bunce, J.D., J. Am. Chem. Soc., 110 (1988) 5959.
4. Kubinyi, H. (Ed.), 3D QSAR in Drug Design. Theory, Methods and Applications, ESCOM, Leiden, 1993.
5. Kuntz, J.D., Science, 257 (1992) 1078.
6. Kim, K., J. Comput.-Aided Mol. Design, 7 (1993) 71.
7. Folkers, G., Merz, A. and Rognan, D., In Kubinyi, H. (Ed.), 3D QSAR in Drug Design. Theory, Methods and Applications, ESCOM, Leiden, 1993, p. 583.
8. Klebe, G. and Abraham, U., J. Med. Chem., 36 (1993) 70.
9. Van de Waterbeemd, H., Testa, B. and Folkers, G. (Eds) Computer-Assisted Lead Finding and Optimization. Current Tools for Medicinal Chemistry, Verlag Helvetica Chimica Acta, Basel, Switzerland, 1997.
10. Höltje, H.-D. and Folkers, G., Molecular Modeling: Basic Principles and Applications, VCH Verlagsgesellschaft,Weinheim, Germany, 1997.
11. Kramer, B., Rarey, M. and Lengauer, T., Proteins, Suppl. 1, (1997) 221.
12. Morris, G.M., Goodsell, D.S., Huey, R. and Olson, A.J., J. Comput.-Aided Mol. Design, 8 (1994) 243.
13. Lybrand, T.P., Curr. Opin. Struct. Biol., 5 (1995) 224.
14. Meng, E., Shoichet, B.K. and Kuntz. I.D., J. Comput. Chem., 13 (1992) 505.
15. Tame, J.R.H., J. Comput.-Aided Mol. Design, 13 (1999) 99.
16. Böhm, H.J., J. Comput.-Aided Mol. Design, 12 (1998) 309.
17. Böhm, H.J., J. Comput.-Aided Mol. Design, 8 (1994) 243.
18. Wang, R., Liu, L., Lai, L. and Tang, Y., J. Mol. Model., 4 (1998) 379.
19. Lemmen, C., Lengauer, T. and Klebe, G., J. Med. Chem., 41 (1998) 4502.
20. Waller, C.L., Oprea, T.I., Giolitti, A. and Marshall, G.R., J. Med. Chem., 36 (1993) 4152.
21. De Priest, S.A., Mayer, D., Naylor, C.B. and Marshall, G.R., J. Am. Chem. Soc., 115 (1993) 5372.
22. Cho, S.J., Garsia, M.L., Bier, J. and Tropsha, A., J. Med. Chem., 39 (1996) 5064.
23. Vaz, R.J., McLean, L.R. and Pelton, J.T., J. Comput.-Aided Mol. Design, 12 (1998) 99.
24. Sippl, W., Contreras, J.M., Rival, Y. and Wermuth, C.G., In Gundertofte, K. (Ed.), Molecular Modelling and Prediction of Bioactivity, Plenum Press, New York, NY, in press.
25. Pastor, M., Cruciani, G. and Watson, K., J. Med. Chem., 40 (1997) 4089.
26. Sadler, B.R., Cho, S.J., Ishaq, K.S., Chae, K. and Korach, K.S., J. Med. Chem., 41 (1998) 2261.
27. Brzozowski, A.M., Pike, A.C., Dauter, Z., Hubbard, R.E., Bonn, T., Engstrom, O., Ohman, L., Greene, G.L., Gustafsson, J.A. and Carlquist, M., Nature, 389 (1997) 753.
28. Shiau, A.K., Barstad, D., Loria, P.M., Cheng, L., Kushner, P.J., Agard, D.A. and Greene, G.L., Cell, 95 (1998) 927.
29. Bernstein, F.C., Koetzle, T.F., Williams, G.J.B., Meyer, E.F., Brice, M.D., Rodgers, J.R., Kennard, O., Shimanouchi, T. and Tasumi, M., J. Mol. Biol., 112 (1977) 535.
30. Weiner, S.J., Kollman, P.A., Case, D.A., Singh, U.C., Ghio, C., Alagona, G., Profeta, S. and Weiner, P.J., J. Am. Chem. Soc., 106 (1984) 765.

31. Singh, U.C. and Kollman, P.A., J. Comput. Chem., 5 (1984) 129.
32. Vedani, A. and Huhta, D.W., J. Am. Chem. Soc., 112 (1990) 112 269.
33. Goodsell, D.S., Morris, G.M. and Olson, A.J., J. Mol. Recogn., 9 (1996) 1.
34. Rao, M.J. and Olson, A.J., Proteins Struct. Funct. Genet., 34 (1999) 173.
35. Goodford, P.J., J. Med. Chem., 28 (1985) 849.
36. Vedani, A. and Dunitz, J.D., J. Am. Chem. Soc., 107 (1985) 7653.
37. PrGen 1.5.6, Biographics Laboratory, Basel, Switzerland.
38. GRID, v. 16, Molecular Discovery Ltd., Oxford, U.K.
39. SYBYL 6.5, Tripos Associates, Inc., St. Louis, MO, U.S.A.
40. Gasteiger, J. and Marsilli, M., Tetrahedron, 36 (1980) 3219.
41. Spartan 4.1.1, Wavefunction, Irvine, CA, U.S.A.
42. GOLPE version 4.0, Multivariate Infometric Analysis, Perugia, Italy.
43. Baroni, M., Constantino, G., Cruciani, G., Riganelli, D., Valigli, R. and Clementi, S., Quant. Struct.-Act. Relat., 12 (1993) 9.
44. Cruciani, G. and Watson, K., J. Med. Chem., 37 (1994) 2589.
45. Pastor, M., Cruciani, G. and Clementi, S., J. Med. Chem., 40 (1997) 1455.
46. Oprea, T.I. and Garcia, A.E., J. Comput.-Aided Mol. Design, 10 (1996) 186.
47. Tong, W., Perkins, R., Xing, L., Welsh, W.J. and Sheehan, D.M., Endocrinology, 138 (1997) 4022.
48. Taylor, N.R. and von Itzstein, M., J. Comput.-Aided Mol. Design, 10 (1996) 233.
49. Jiang, H., Chen, K., Tang, Y., Chen, J., Li, Q., Wang, Q. and Ji, R., J. Med. Chem., 40 (1997) 3085.
50. Viswanadhan, V.N., Reddy, M.R., Wlodawer, A., Varney, M.D. and Weinstein, J., J. Med. Chem., 39 (1996) 705.
51. Bursi, R. and Grootenhuis, P.D.J., J. Comput.-Aided Mol. Design, 13 (1999) 221.
52. Klebe, G., J. Mol. Biol., 237 (1994) 212.
53. Klebe, G. and Böhm, H., J. Recept. Signal Transduct. Res., 17 (1997) 459.
54. Waller, C.L., Oprea, T.I., Chae, K., Park, H.K., Korach, K.S., Laws, S.C., Wiese, T.E., Kelce, W.R. and Gray, L.E., Chem. Res. Toxicol., 9 (1996) 1240.
55. Waller, C.L., Minor, D.L. and McKinney, J.D., Environ. Health Perspect., 103 (1995) 702.
56. Böhm, M., Stürzebecher, J. and Klebe, G., J. Med. Chem., 42 (1999) 458.
57. Klebe, G. and Abraham, U., J. Comput.-Aided Mol. Design, 13 (1999) 1.
58. Kim, K.H., Greco, G. and Novellino, E., Perspect. Drug. Discov. Design, 12 (1998) 257.