CHEMMEDCHEM

DOI: 10.1002/cmdc.200800274

Mixed-Model QSAR at the Glucocorticoid Receptor: Predicting the Binding Mode and Affinity of Psychotropic Drugs

Morena Spreafico,[a] Beat Ernst,[a] Markus A. Lill,[b] Martin Smiesko,[a] and Angelo Vedani*[a]

The glucocorticoid receptor (GR) is a member of the nuclear receptor superfamily that affects immune response, development, and metabolism in target tissues. Glucocorticoids are widely used to treat diverse pathophysiological conditions, but their clinical applicability is limited by side effects. A prediction of the binding affinity toward the GR would be beneficial for identifying glucocorticoid-mediated adverse effects triggered by drugs or chemicals. By identifying the binding mode to the GR using flexible docking (software Yeti) and quantifying the binding affinity through multidimensional QSAR (software Quasar), we validated a model family based on 110 compounds, representing four different chemical classes. The correlation with the experimental data (cross-validated $r^2 = 0.702$; predictive $r^2 = 0.719$) suggests that our approach is suited for predicting the binding affinity of related compounds toward the GR. After challenging the model by a series of scramble tests, a consensus approach (software Raptor), and a prediction set, it was incorporated into our VirtualToxLab and used to simulate and quantify the interaction of 24 psychotropic drugs with the GR.

Introduction

Nuclear receptors regulate biological functions such as cell growth and differentiation, metabolic processes, reproduction and development, intracellular signaling, and can be involved in carcinogenesis through the control of gene expression.[1] Chemicals that disrupt the endocrine system interfere with the function of nuclear receptors, alter their functions, and consequently cause adverse health effects.[2] Besides natural and synthetic hormones, a broad variety of chemicals show endocrine-disrupting activity, for example, industrial chemicals such as cleaning agents, byproducts of industrial processes such as dioxins, as well as pesticides and plastic additives. They can either bind directly to, and thus block, a hormone receptor or affect the synthesis, transport, metabolism, or elimination of hormones. To date various nuclear-receptor-mediated hormonal responses to toxic compounds have been reported, including xenobiotic effects on the thyroid hormone receptor, the epidermal growth factor receptor, the aryl hydrocarbon receptor, and the androgen and estrogen receptors.[3]

Endocrine disruptors can alter the physiological function of the glucocorticoid receptor (GR), leading to the dysfunction of metabolic, immunological, and central nervous system processes. For example, there is evidence that low concentrations of arsenic[4] and polychlorinated biphenyls (PCBs)[5] can interfere with glucocorticoid activity. However, many aspects of endocrine disruption have yet to be elucidated, such as the long-term effects of low-level exposure. The presence of compounds with hormonal activity in the biosphere has become a worldwide environmental concern, which is addressed by various regulations in both the US and Europe. In 1996, the US Environmental Protection Agency's Office of Research and Development identified endocrine disruption as one of its top six research priorities and developed a risk-based research approach.[6] In December 2006, the European Union approved the Registration, Evaluation, and Authorization of CHemicals (REACH), a regulation that covers the production and use of chemical substances. In particular, additional authorization for substances of "very high concern", such as endocrine disruptors, is required. In Switzerland, the necessity for a coordinated interdisciplinary approach has also been recognized, and the National Research Programme on Endocrine Disruption (NRP50) was conducted 2001–2007.[7] A reliable method for the in silico identification of the endocrine-disrupting (or, more generally, toxic) potential of drugs and chemicals is therefore regarded as highly desirable by regulatory bodies as well as the pharmaceutical, chemical, and food industries.

The GR is a ligand-activated transcription factor that controls a wide variety of biological processes including an organism's development, metabolism, and immune response.[8] It can both activate and repress the transcription of target genes by binding to glucocorticoid-responsive elements or through cross-talk with other transcription factors, such as activator protein-1 or nuclear factor-κB.[9]

[a] *M. Spreafico, Prof. B. Ernst, Dr. M. Smiesko, Prof. A. Vedani*
Department of Pharmaceutical Sciences, University of Basel
Klingelbergstrasse 50, 4056 Basel (Switzerland)
Fax: (+41) 61-267-1552
E-mail: angelo.vedani@unibas.ch

[b] *Prof. M. A. Lill*
Department of Medicinal Chemistry and Molecular Pharmacology
Purdue University, West Lafayette, IN 47907 (USA)

Supporting information for this article is available on the WWW under http://dx.doi.org/10.1002/cmdc.200800274.

Ligands of the GR, the glucocorticoids, are in wide therapeutic use for their anti-inflammatory and immunosuppressive activities, for example, to treat asthma, allergic rhinitis, rheumatoid arthritis, and acute transplant rejection.[10] However, a range of side effects including osteoporosis, metabolic syndrome, impaired development, and blunted growth limits their clinical use.[11] Therefore, the identification of new glucocorticoids is still an endeavor in pharmaceutical research and development, and in silico tools that aid the rational design of glucocorticoids, particularly by quantifying their binding affinity, are consequently much needed.

Herein we discuss the development and validation of a 3D model for the GR that allows the screening of drugs and chemicals for potential GR activity. The underlying methodology was developed at the Biographics Laboratory 3R[12] and compiled into a virtual laboratory (*VirtualToxLab*) for the in silico identification of the toxic (endocrine-disrupting) potential of drugs and environmental chemicals.[13]

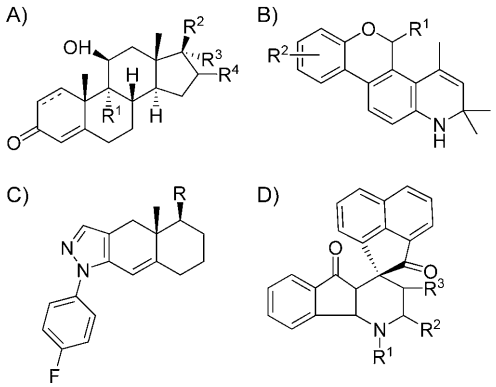

Figure 1. Compound classes A–D used in the QSAR study: A) steroids (**A01**–**A17**), B) quinoline derivatives (**B01**–**B30**), C) fluorophenylindazole derivatives (**C01**–**C55**), and D) spirocyclic dihydropyridine derivatives (**D01**–**D08**); all 110 structures (along with eight of the prediction set and the 24 psychotropic drugs) are given in the Supporting Information.

Methods

The 3D structure of the GR with bound dexamethasone (PDB code: 1M2Z, 2.5 Å resolution, $R_{free} = 0.267$)[14] was obtained from the Protein Data Bank.[15] For our study, we chose chain A, which includes 255 amino acid residues, dexamethasone, and 118 water molecules. First, all hydrogen atoms were added to the structure, and the correct protonation state for the histidine residues was assigned: His645 and His654 were protonated at their N^{δ} atoms, while His588, His726, and His775 were protonated at their respective N^{ε} atoms. Next, the hydrogen-bond network was constructed and optimized, and the structure was relaxed using the directional force field implemented in Yeti.[16]

The experimental affinity data for the 110 GR-binding compounds were obtained from multiple sources.[17] Most of the affinity data (73%) are available as IC_{50} values, the remaining as K_i values. For their use in our study, we converted the IC_{50} into K_i values using the Cheng–Prusoff relation.[18] The binding affinities range from 3.2×10^{-5} to 5×10^{-11} M, with the majority clustering within two orders of magnitude (10^{-7}–10^{-9} M). For prednisolone (compound code within this study: **A02**), a comparison of experimental binding affinities obtained from three different sources revealed a difference of more than one order of magnitude in its K_i value (from 2.4 nM[17c] to 32 nM[17a]), suggesting that the predictive power of the model might be somewhat jeopardized by this fact, as no model can be better than the underlying experimental data.

The ligand set comprises four different chemical classes: steroids, quinoline derivatives, fluorophenylindazole derivatives, and spirocyclic dihydropyridine derivatives (Figure 1). The structures of the compounds and their K_i values[17] are given in the Supporting Information.

Quinoline derivatives bear a stereocenter. For four compounds of the data set—(*R*)- and (*S*)-2,5-dihydro-10-methoxy-2,2,4-trimethyl-5-phenyl-1*H*-[1]benzopyrano[3,4-*f*]quinoline (**B25**, **B26**) and (*R*)- and (*S*)-2,5-dihydro-10-methoxy-2,2,4-trimethyl-5-(3,5-dichlorophenyl)-1*H*-[1]benzopyrano[3,4-*f*]quinoline (**B29**, **B30**)—the K_i values for individual stereoisomers are available. In both cases, the *S* isomer shows a higher affinity than the *R* counterpart (a factor of 33 for **B26**/**B25** and 114 for **B30**/**B29**).[17b] Docking studies with these stereoisomers to the GR are in agreement with this observation. Because most of the quinoline derivatives used were only tested as racemates, the affinity for the *S* isomers in our study were corrected by dividing their K_i values by a factor of two, in order to account for the content of the nearly inactive isomer in the sample.

The 3D structures of all ligands were generated and optimized in aqueous solution based on the AMBER* force field[19] as implemented in MacroModel 6.5.[20] Atomic partial charges (CM1) were generated with the AMSOL package.[21] To identify the lowest-energy conformation for all quinolines, fluorophenylindazoles, and spirocyclic dihydropyridines, a conformational search with MacroModel 6.5 was performed (10000 Monte Carlo iterations). For the steroids, only the compound with the highest experimental affinity was investigated, assuming that all the steroids share a common puckering of the B and C rings.[22] Each of the 110 ligands was individually docked into the receptor structure with the software Yeti. To account for induced fit, we used flexible docking based on the Monte Carlo search protocol, allowing adaptation of the amino acid side chains that contain at least one atom within 12 Å of any atom in the given ligand. An appropriate consideration of receptor flexibility would seem to be a prerequisite for the identification of feasible binding modes, particularly for ligands with bulky substituents (see below). In our study, we selected the lowest-energy pose of the most active compound for each chemical class and subsequently used it as a binding-mode template for the remaining compounds that belong to the same chemical class. Details of the binding of a representative compound from classes A–D (see Figure 1) at the GR are shown in Figure 2.

The data set was split into training and test compounds[23] in such a way to maximize the diversity of the training set with respect to binding affinity and chemical properties. This was done by grouping the compounds according to chemical class

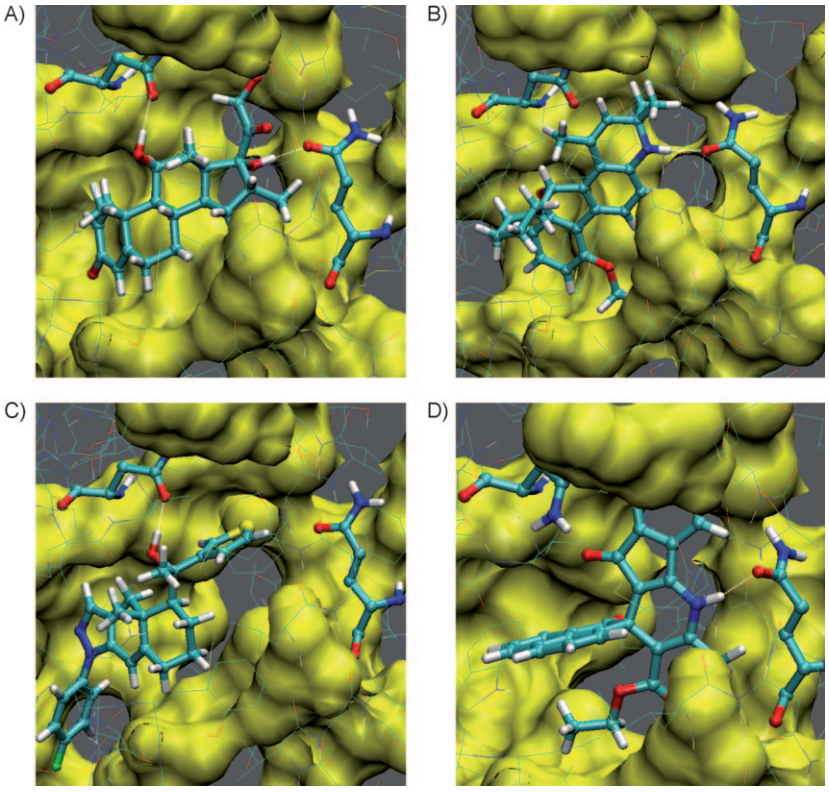

Figure 2. Details of the binding of a representative compound of classes A–D of the GR used for this QSAR study: the ligands A) **A01**, B) **B18**, C) **C12**, and D) **D02**, as well as residues Asn564 and Gln642 are colored by atom type; the residues lining the hydrophobic pocket are depicted in yellow.

(i.e., those that share the same scaffold) and ranking by affinity. For each group, the most and least active compounds were assigned to the training set. From the compounds remaining in the pool, those with different scaffolds and functionalities were selected to be part of the training set in order to maximize chemical diversity. For the QSAR simulations we used a 4:1 ratio, yielding 88 training and 22 test compounds.

Our aim was to establish a quantitative structure–activity relationship (QSAR) for the 110 compounds binding to the GR using the Quasar and Raptor technologies in consensus-scoring mode.[24] Both Quasar and Raptor are receptor-modeling concepts that allow multidimensional QSAR. Because the details of these pieces of software are published elsewhere,[12a,b,25] we summarize them here only briefly. We used the so-called mixed-modeling approach, in which the ligands are docked to the X-ray crystal structure, and their binding affinity is quantified using a quasi-atomistic binding-site model thereof. This approach yields more reliable binding energies than directly scoring the ligand–protein interactions at the experimental structure.[26]

In Quasar, the binding site of the protein is represented by a surrogate, which consists of a 3D surface surrounding the ligands superimposed in their bioactive conformation (as obtained, for example, from docking studies at the true biological receptor) at van der Waals distance. The topology of this surface mimics the shape of the binding site. This surface is then populated with quasi-atomistic properties corresponding to those of the amino acids: positively and negatively charged salt bridges; hydrogen-bond donors and acceptors; neutral, positively, and negatively charged hydrophobic properties; hydrogen-bond flip-flop, as well as solvent.[12b] Apart from accepting 4D compound input (conformations, poses, protonation states, and tautomers), Quasar allows specifically for the simulation of induced fit (corresponding to side-chain flexibility and moderate backbone motion at the true biological receptor), whereby six different protocols are evaluated simultaneously (5D-QSAR).[12b] The model family, typically consisting of 200–500 models, is evolved by using a genetic algorithm and provides an averaged prediction for each compound along with the variation over the model family. Quasar employs the following scoring function derived from the directional Yeti force field:[12b,16a]

$$E_{bind} = E_{ligand-receptor} - E_{ligand\ desolvation} - E_{ligand\ internal\ strain} - T\Delta S - E_{induced\ fit}$$
$$\text{for which } E_{ligand-receptor} = E_{electrostatic} + E_{vdW} + E_{H\ bonding} + E_{polarization} \quad (1)$$

$$\Delta G_{exp} = |a| \cdot E_{bind} + b \quad (2)$$

Using the ligands of the training set, a linear regression of the experimental (ΔG_{exp}) and calculated (E_{bind}) binding affinity is then obtained [Eq. (2)]. The coefficients a and b are derived from the correlation of the training set in cross-validation mode and are later applied to molecules of the test set or new compounds for which binding affinity should be predicted. In Quasar, the solvent contribution can be calculated explicitly, allowing the presence of solvent properties on the surrogate surface, or implicitly, where the solvation terms (ligand desolvation and solvent stripping) are independently scaled for each model within the surrogate family. Each different scaling reflects a different solvent accessibility of the binding site. Solvation terms are associated with weights that evolve throughout the simulation (6D-QSAR).[25]

In contrast to Quasar, the binding site in Raptor is represented by two 3D surfaces, again populated with quasi-atomistic properties. The two shells allow one to account for various induced-fit mechanisms, observed, for example, with agonists and antagonists or ligands that differ substantially in size.[12c] Induced fit may, of course, exert different shapes and magnitudes for the two shells. The model development employs a

multistep optimization protocol including domain assignment, tabu search, and local search.[12c] In Raptor, the scoring function includes directional terms for hydrogen bonding and hydrophobicity, as well as terms for the cost of topological adaptation and the changes in entropy upon ligand binding:

$$E_{bind} = E_{ligand-receptor} - T\,\Delta S - E_{induced\ fit} \quad (3)$$

for which $E_{ligand-receptor} = w\,[E_{H\ bonding\ (shell\text{-}1)} + E_{hydrophobic\ (shell\text{-}1)}] + (1.0-w)\cdot[E_{H\ bonding\ (shell\text{-}2)} + E_{hydrophobic\ (shell\text{-}2)}]$, and w is the interpolation weight between the two shells.[12c]

Results and Discussion

Protein flexibility remains a major challenge for quantifying the binding of small molecules to a macromolecular target. Although induced fit is understood to be a key phenomenon for ligand binding, docking software typically treats the ligand as a flexible entity but leaves the receptor structure rigid (or nearly rigid).[27] In the Yeti concept, local induced fit is taken into account by allowing amino acid side chains to move and thereby adopt an optimal conformation with respect to the bound ligand. In particular, the docking of glucocorticoids with bulky substituents at the 17β-position (see Figure 3) into the crystal structure of the glucocorticoid receptor may not yield the correct binding pose, because the binding site would not seem to be wide enough to accommodate large compounds such as desoxymethasone 21-cinnamate (**A08**). For other steroid receptors such as the progesterone receptor,[28] there is experimental evidence for a local induced fit to accommodate bulky substituents at position 17. A similar behavior has been postulated for the androgen receptor.[29] Likewise, bulky ligands may indicate steric interference with the GR binding pocket if the protein structure is kept rigid during docking (cyan ribbon in Figure 3). In contrast, energetically favorable binding poses are found when the side chains of the protein are allowed to adjust to ligand topology (green residues in Figure 3).

The ensemble of binding modes of the 110 compounds, identified using flexible docking as implemented in the software Yeti,[16b] was used as input for the QSAR technology Quasar.[12b] This mixed-modeling approach allows both the correct identification of binding mode and the reliable estimation of binding affinities. The Quasar simulations were based on a family of 200 receptor models that differ with respect to the properties mapped on their surface. The family of receptor models was evolved for 60 000 crossover cycles, corresponding to 300 generations. For cross-validation we selected a leave-*n*-out ($n = 10$) protocol. Protein flexibility was mimicked using a total of six induced-fit scenarios.[12b] Throughout the entire simulation, a static mutation rate of 0.02 was applied during transcription of the quasi-atomistic properties.

The model family converged at a cross-validated r^2 value of 0.702 for the 88 training compounds and a predictive r^2 value of 0.719 for the 22 test ligands (see Figure 4 and Table 1). On average (rms), the calculated binding affinity of the training and test ligands deviate from the experimental K_i value factors of 1.5 and 1.6, respectively. The maximal observed deviation of an individual compound corresponds to a factor of 9.9 in K_i for the training set and 4.9 for the test set. A representation of the receptor surrogate with bound dexamethasone (**A01**) along with key amino acid residues is depicted in Figure 5. In comparison with the crystal structure of the GR complexed with dexamethasone, the receptor surrogate generated by Quasar properly reproduces the corresponding properties at the position of the key amino acid residues Asn564, Arg611, and Gln642, as identified by X-ray at the quasi-atomistic level:[12b] an H-bond-donating domain (green) corresponding to the guanidinium moiety of Arg611, and H-bond-accepting domains (yellow) mimicking the carbonyl groups of Asn564 and Gln642 at the true biological receptor.

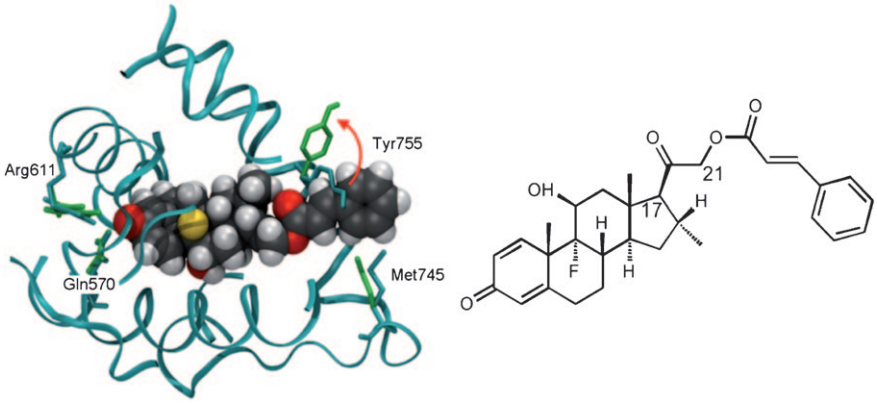

Figure 3. With a rigid protein structure, docking of desoxymethasone 21-cinnamate (**A08**, shown in space-filling mode; its chemical structure shown at right) to the crystal structure of the GR, obtained from the dexamethasone–GR complex (cyan ribbons and sticks), leads to unfavorable interactions. The following amino acid residues lining the binding pocket are shown: Tyr735, Met745, Arg611, and Gln570. Cyan stick representations refer to the positions of amino acids as observed in the crystal structure, whereas green sticks refer to the amino acid position after flexible docking. While the positions of Arg611 and Gln570 remain almost unchanged, flexible docking allows the side chains of Met745 and Tyr735 to move and accommodate the 21-cinnamate moiety of **A08** (the most evident movement of Tyr735 is indicated by the red arrow). At left, **A08** is rotated 180° vertically relative to the standard steroid depiction at right.

In multidimensional QSAR (mQSAR), it is of utmost importance to challenge a model, for example, by means of a scramble test, an external prediction set, or by consensus scoring. The scramble test, which is frequently used to assess the sensitivity of a model,[30] consists of a random shuffling of the binding affinities of the training set ligands with respect to the true affinity values. If, under these circumstances, the ligands of the test set are still predicted correctly, the model is worthless, as it is not sensitive to the biologi-

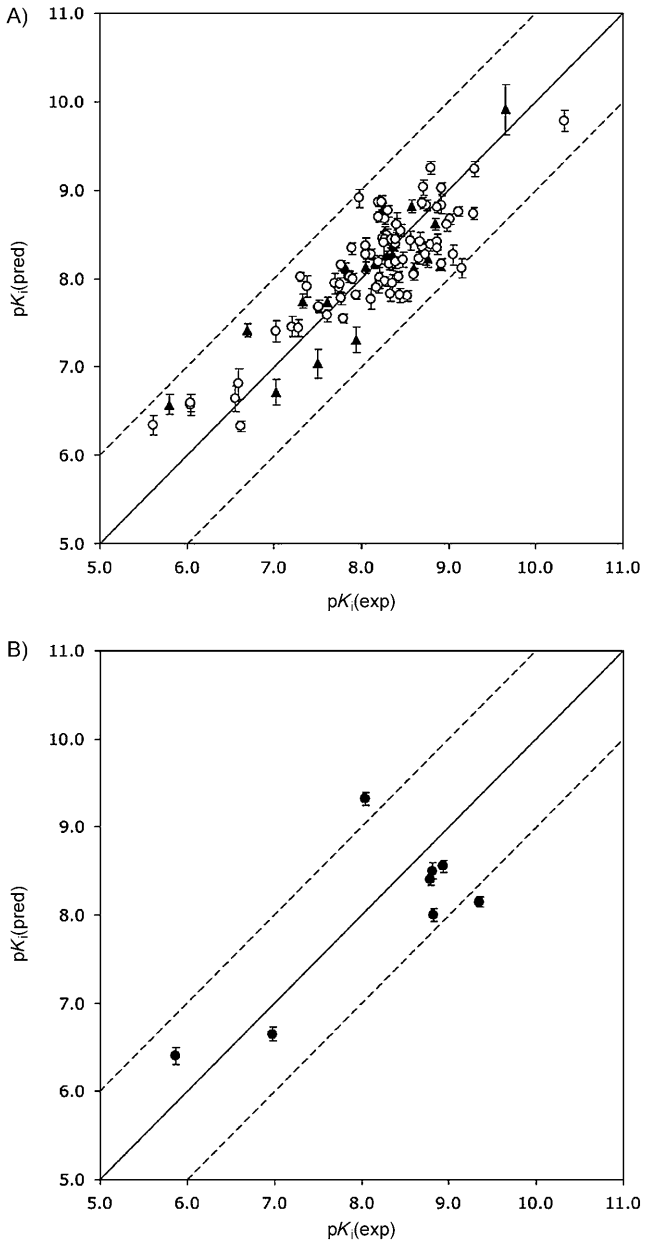

Figure 4. Comparison of experimental and predicted binding affinities of A) the training and test set and B) the external test set toward the GR, as obtained with Quasar. Ligands of the training set are depicted as open circles, those of the test set as filled triangles, and those of the prediction set as filled circles. Dashed lines are drawn at a factor of 10 from the experimental value.

Table 1. Summary of the Quasar and Raptor simulations with 88 training and 22 test compounds.[a]

Simulation	r^2	q^2	rmsd Training	Max. Training	p^2	rmsd Test	Max. Test
Quasar	0.710	0.702	1.5	9.9	0.719	1.6	4.9
Raptor	0.680	n/a	2.1	5.3	0.519	6.1	20.4

[a] r^2: correlation coefficient, q^2: cross-validated r^2, p^2: predictive r^2; the rmsd and maximal deviation from the experimental binding affinity are given as a factor (off) in K_i.

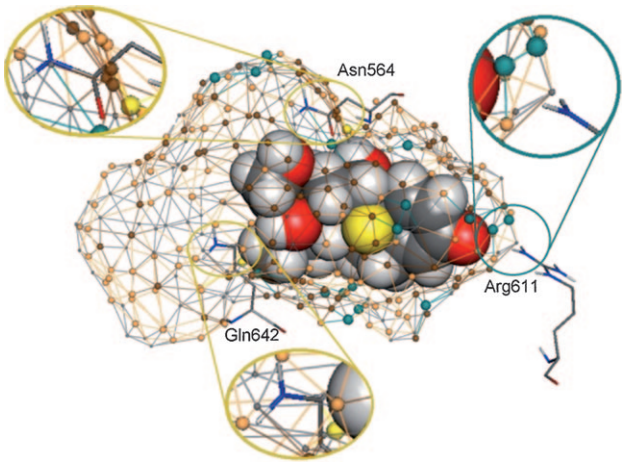

Figure 5. Representation of the GR surrogate (Quasar) with bound dexamethasone (**A01**, space-filling representation). The mapped quasi-atomistic properties are colored as follows: red (salt bridge, positively charged), blue (salt bridge, negatively charged), green (H-bond donor), yellow (H-bond acceptor), saddle brown (hydrophobic, positively charged), chocolate brown (hydrophobic, negatively charged), gray (hydrophobic, neutral). Key amino acid residues from the X-ray structure[24] (Asn564, Arg611, and Gln642) are shown as sticks. The analogies between model and receptor are circled and shown in larger size: green (H-bond donor) corresponding to Arg 611; yellow (H-bond acceptor) corresponding to the carbonyl group of Gln642, hidden in the back, and to the carbonyl of Asn564.

cal data. In our study, 20 scramble tests were performed with a different shuffling of the biological data for each scramble test. They yielded an average predictive r^2 value of -0.241 compared with $+0.719$ for the simulation using unscrambled values. Only a single simulation gave a positive predictive r^2 value (0.375); all others found no correlation (predictive $r^2 <$ 0.0), demonstrating that the model for the GR is indeed sensitive to the biological data.

Another sensitive issue is the possible overfitting of the model. We paid particular attention to this problem and monitored the evolution of the predictive r^2 value relative to the cross-validated r^2, and stopped the simulation at the point (300 generations) where the former starts to drop while the latter continues to rise. As our Quasar model family consists of 200 individual models, the scattering of the individual predictions around the mean value (see Figure 4A) is another indicator for possible overfitting. Although the ligands of the test set show a broader distribution, those of the training set scatter, on average, by a factor of 2–3 about their mean value. An overfitted model would result in rather small scatterings for the ligands of the training set.

To achieve consensus, we applied a second methodology (software Raptor). The Raptor simulation—using the same ligand alignment and selection—yielded an r^2 value of 0.680 and a predictive r^2 of 0.519. The comparison of predicted and experimental binding affinities is shown in Figure 6, and the performance coefficients are given in Table 1. When compared with Quasar, the Raptor simulation would seem to yield only modest predictive power, but considering the limited range of experimental activity (85% of compounds cluster within two orders of magnitude), the chemical diversity of the com-

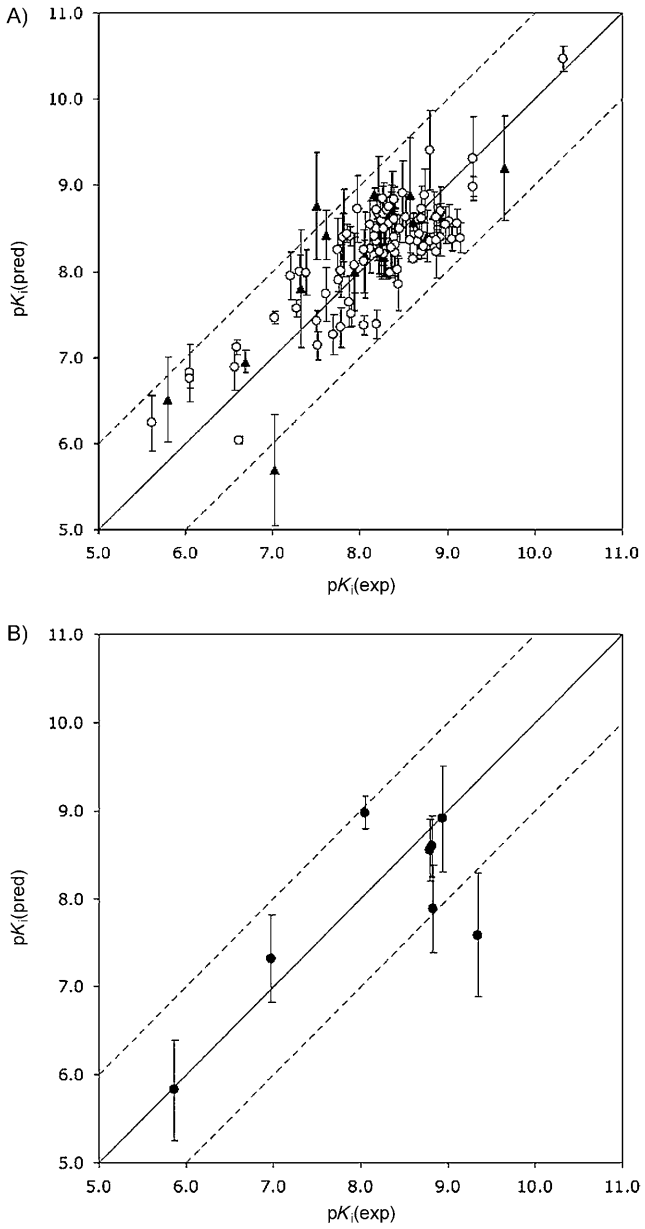

Figure 6. Comparison of experimental and predicted binding affinities of A) the training and test set and B) the external test set toward the GR as obtained with Raptor. Ligands of the training set are depicted as open circles, those of the test set as filled triangles, and those of the prediction set as filled circles. Dashed lines are drawn at a factor of 10 from the experimental value.

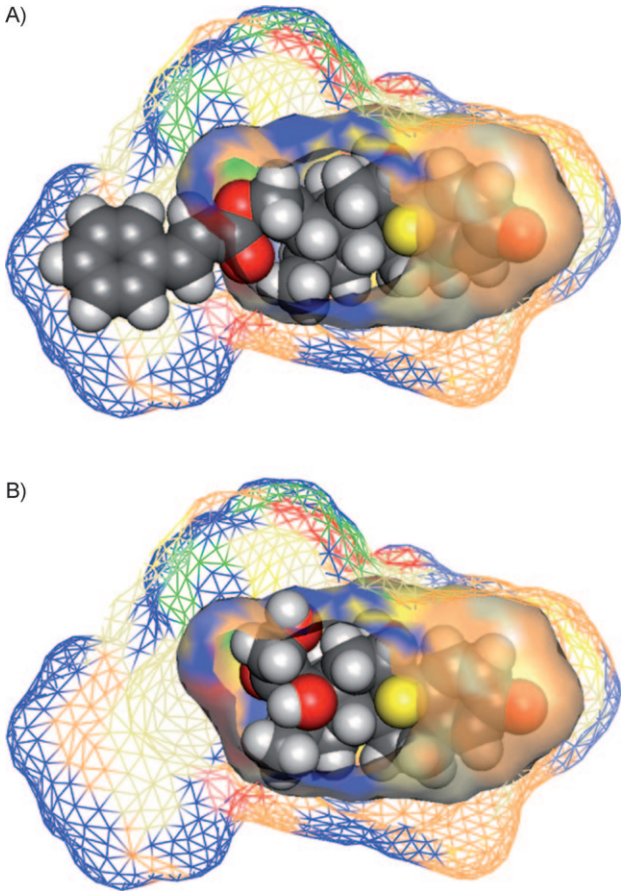

Figure 7. Raptor model of the GR binding site model with A) compound **A08** and B) dexamethasone (**A01**) bound (hydrophobic fields: beige, H-bond-donating propensity: blue, H-bond-accepting propensity: red, hydrogen-bond flip-flop: green). The inner shell (transparent surface) and outer shell (wire frame) are shown in different style to highlight the two shells of the surrogate (the front part of the receptor model has been clipped for clarity).

pounds, and the different literature sources for the affinities, the Raptor model can be considered acceptable in terms of quality. Moreover, its dual-shell representation can simulate induced fit more realistically, particularly with compounds that differ substantially in bulkiness, or in the presence of two induced-fit mechanisms. For example, the cinnamate substituent at position 21 of **A08** (see Figure 3) is snugly accommodated by the outer shell (Figure 7 A), indicating the necessity of side chain rearrangement in the binding pocket of the protein in order to allocate additional space for the large substituent. On the other hand, the inner shell hosts compounds such as dexamethasone (**A01**) that are characterized by a smaller volume (Figure 7 B). The comparison of predicted binding affinities obtained by the two approaches is shown in Figure 8. The average deviation is 0.32 logarithmic units (a factor of 2.1 in K_i). For only two compounds out of 110 (1.8%), both of which belong to the test set, the ratio is greater than a factor of 10 (in K_i). Thus, consensus between Quasar and Raptor predictions has been achieved. Whereas for prednisolone (**A02**), the threshold for acceptance is only slightly exceeded: 1.02 log units (a factor of 10.5 in K_i), the disagreement for 2,5′-dioxo-2′-phenyl-3′-ethoxycarbonylspiro(1,4′-acenapthene-1′,4′-dihydroindeno-[3,2-*b*]pyridine) (**D06**) is 1.73 log units (a factor of 53.7 in K_i).

Although the test set was not used for generating and optimizing the model, its performance was, of course, considered as a criterion for selecting the final model among all those generated. To truly challenge the model, a new independent set of compounds was identified in the literature,[31] and employed only for this validation step (the compound structures and their K_i values are given in the Supporting Information).

The chemical property domain of the prediction set (**P01**–**P08**) is included in the model's property space because the

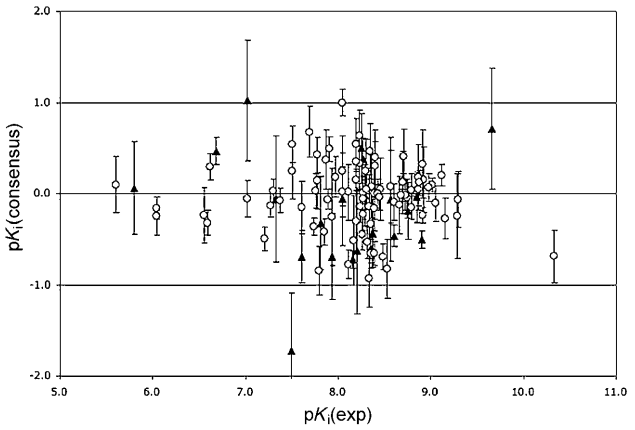

Figure 8. Consensus scoring using Quasar and Raptor. The quantity is expressed as $pK_{i\ consensus} = -\log(K_{i\ Quasar}/K_{i\ Raptor})$. Error bars indicate the cumulative standard deviation, $esd_{cumulative} = \sqrt{(esd^2_{Quasar} + esd^2_{Raptor})}$.

scaffold of the new compounds is already represented in the training set, and because the activity range is within the broader range of the activities of the training set, therefore allowing a reliable prediction instead of an extrapolation of binding affinities. The K_i values of the external compounds (as predicted by Quasar) are shown in Figure 4B. The predictive r^2 value is 0.538, the rmsd for this external set corresponds to a factor of 4.7 in K_i, and the maximal deviation is 17.9: an appreciable result considering that these ligands were not used for model construction and selection. The K_i values predicted by Raptor are shown in Figure 6B. The predictive r^2 value is 0.488, the rmsd for this external set corresponds to a factor of 20.3 in K_i, and the maximal deviation is 56.4.

Binding of psychotropic drugs to the GR

To assess the validity of the approach, we simulated the binding of 24 psychotropic drugs to the GR. For this task we used the protocol as implemented in the *VirtualToxLab*,[13,32] which includes a full conformational search in aqueous solution and the identification of the most probable protonation and tautomeric state at physiological pH, followed by automated, flexible docking and calculation of the binding affinity using 6D-QSAR.[25] The resulting binding affinities are given in Table 2. Because our GR model was trained using almost exclusively neutral species (87:1), we calculated the binding affinities of the psychotropic drugs for both the neutral and charged state (where applicable) and observed that the charged species—corresponding to the protonation state in aqueous solution at pH 7.4—typically yield higher affinities. This might, however, represent an artifact caused by the electrostatic contribution to the protein–ligand interaction. It is well known, of course, that the dielectric properties in the interior of a receptor may differ significantly from those in aqueous solution. Six compounds—bupropion, fluoxetine, lorazepam, methylphenidate, trimipramine, and venlafaxine—are marketed as racemic mixtures. Here, we simulated both stereoisomers (see Table 2). As an example, lorazepam is discussed in detail (Figure 9). Although the *R* isomer engages in hydrogen bonds with the GR, it features a ninefold weaker activity (Table 2). This is a consequence of the hydrophobic ligand–protein interactions, which are more pronounced for the *S* isomer. Table 3 lists the scaled interaction energies as obtained from the mQSAR simulation (Quasar).

Figure 10A reveals details of the binding of clomipramine to the GR. In the pose that contributes most (a total of eight were considered in the mQSAR), the proton of the dimethylammonium group points toward its own phenyl moiety, thus forming a charge–π interaction. Another pose found in the conformational search (not shown) shows the same H atom engaging in a weak hydrogen bond with the nearby Asn564 residue (highlighted in Figure 10). Likewise, the ammonium H atom of mirtazapine engages in a hydrogen bond with Asn564 (Figure 10B). However, for other tested compounds such as buspirone, fluoxetin, trimipramin, and trazodone (Figure 10C), no such possibility exists, and binding in the neutral state would make more sense. As our GR model was implemented in the *VirtualToxLab*, we calculated the binding affinity toward all 11 models (androgen, aryl hydrocarbon, estrogen α/β, glucocorticoid, mineralocorticoid, peroxisome proliferator-activated γ, and thyroid α/β receptors, as well as for the enzymes cytochrome P450 2A13 and 3A4). The results, along with those of >600 other compounds, are accessible through the internet.[33]

A wealth of information on the side effects of psychotropic drugs is published; summaries can be found, for example, in Wikipedia.[34] Some of the compounds analyzed in this study trigger adverse effects via the GR. These include alprazolam (increased glucocorticoid levels),[35] amitryptyline (induction of GR mRNA),[36] clomipramine (regulates GR expression),[37] desipramine (GR translocation, increased GR mRNA),[38] escitalopram (reversion of GR immunoreactivity),[38] fluoxetine (enhances GR function),[37,39] imipramine (partial agonist-like effects),[40] moclobemide (increased GR mRNA),[39] reboxetine (increased cortisol levels),[41] and sertraline (high GR responsiveness).[42] For other compounds, no or less clear effects were observed.[43] All this suggests that our computational protocol as implemented in the *VirtualToxLab* might be well suited for identifying such effects in silico.

Conclusions

Herein we present the generation and validation of a QSAR model for a series of 118 ligands of the GR (88 compounds in the training set, 22 in the test set, and eight in the prediction set). The model was obtained by combining flexible docking (software Yeti) and multidimensional QSAR (software Quasar and Raptor). The receptor surrogate is characterized by quasi-atomistic properties that reflect those of the true biological receptor. This suggests that the model is, at least in part, interpretable and correlated with structural properties. In contrast to other modeling studies on the GR,[17a,44] induced fit, a key mechanism for ligand binding, was explicitly simulated both in the docking phase and during the QSAR simulations. The Quasar simulation yielded a cross-validated r^2 value of 0.702 for the training set and a predictive r^2 of 0.719 for the internal

Table 2. Predicted binding affinities for 24 psychotropic drugs toward the GR.

Compound	Formal charge	Calculated binding affinity [M]	Relative binding affinity[a]
Alprazolam	neutral	1.8×10^{-8}	0.16
Amitriptyline	neutral	2.9×10^{-7}	0.010
	+1	6.5×10^{-9}	0.46
Bupropion (*R*/*S*)	neutral	8.6×10^{-7}/1.7×10^{-6}	0.0034/0.0018
	+1	3.6×10^{-8}/2.0×10^{-8}	0.082/0.15
Buspirone	neutral	1.3×10^{-6}	0.0023
	+1	1.3×10^{-10}	23
Clomipramine	neutral	5.0×10^{-9}	0.59
	+1	2.3×10^{-9}	1.3
Desipramine	neutral	2.7×10^{-8}	0.11
	+1	8.2×10^{-9}	0.36
Doxepin	neutral	1.1×10^{-8}	0.27
	+1	2.3×10^{-8}	0.13
Duloxetine	neutral	1.7×10^{-6}	0.0017
	+1	6.1×10^{-9}	0.49
Escitalopram	neutral	5.1×10^{-8}	0.058
	+1	4.8×10^{-9}	0.62
Flunitrazepam	neutral	6.9×10^{-8}	0.043
Fluoxetine (*R*/*S*)	neutral	1.8×10^{-6}/4.7×10^{-7}	0.0016/0.0063
	+1	2.7×10^{-10}/1.7×10^{-8}	11/0.17
Imipramine	neutral	6.6×10^{-7}	0.0045
	+1	1.5×10^{-9}	2.0
Lorazepam (*R*/*S*)	neutral	8.3×10^{-8}/9.1×10^{-9}	0.036/0.33
Methylphenidate (*R*/*S*)	neutral	3.9×10^{-6}/1.0×10^{-3}	0.00076/0.0030
	+1	1.1×10^{-7}/1.2×10^{-8}	0.027/2.5
Mirtazapine	neutral	5.5×10^{-7}	0.0054
	+1	2.0×10^{-10}	15
Moclobemide	neutral	8.4×10^{-7}	0.0035
Modafinil	neutral	1.2×10^{-8}	0.25
Nortriptyline	neutral	3.5×10^{-7}	0.0035
	+1	8.3×10^{-9}	0.36
Paroxetine	neutral	1.4×10^{-8}	0.21
	+1	6.8×10^{-10}	4.4
Reboxetine	neutral	5.1×10^{-8}	0.058
	+1	4.8×10^{-9}	0.62
Sertraline	neutral	9.0×10^{-8}	0.033
	+1	1.6×10^{-9}	1.9
Trazodone	neutral	2.8×10^{-7}	0.011
	+1	6.4×10^{-10}	4.6
Trimipramine (*R*/*S*)	neutral	4.6×10^{-7}/1.1×10^{-8}	0.0064/0.27
	+1	4.4×10^{-10}/1.1×10^{-8}	6.7/0.27
Venlafaxine (*R*/*S*)	neutral	6.6×10^{-7}/1.1×10^{-6}	0.0045/0.0027
	+1	7.3×10^{-9}/1.3×10^{-8}	0.41/0.23

[a] Relative binding affinity (RBA) with dexamethasone (**A01**) as reference compound: $K_{i\,exp}=2.963\times10^{-9}$ M $\rightarrow$ RBA = 1.0. RBA > 1.0 indicates a more active compound than dexamethasone; RBA < 1.0 indicates entities less active than dexamethasone.

test set. The predictive r^2 value for the external set of compounds is 0.538, and the rmsd is a factor of 4.7 in K_i. The Raptor simulation yielded a cross-validated r^2 value of 0.680 for the training set and a predictive r^2 of 0.519 for the internal test set. The predictive r^2 value for the external set of compounds is 0.488, and the rmsd is a factor of 20.3 in K_i. Model sensitivity was assessed by 20 scramble tests, and model robustness by consensus scoring. The model was then used for simulating and quantifying the binding of 24 psychotropic drugs to the GR. Limitations of the current model for the GR are sensitivity to the formal charge and the molecular weight of the compound to be tested. In the former instance, charged species would currently seem to be overestimated, as the model was trained using predominantly neutral species. The size of a compound matters, as for very small ligands the automated docking protocol might not sample enough poses (default = 25), whereas for large molecules leading to a significant induced fit (rmsd > 5 Å), the underlying protocol is unable to simulate and quantify such large conformational changes at the protein. The GR model has been added to the *VirtualToxLab* developed by the Biographics Laboratory 3R.[13,45] It currently includes 11 validated models for the androgen, aryl hydrocarbon, estrogen α/β, glucocorticoid, mineralocorticoid, peroxisome proliferator-activated γ, and thyroid α/β receptors, as well as for the enzymes cytochrome P450 2A13 and 3A4.

Supporting Information available: The chemical structures of the 142 compounds used in this study along with a comparison of experimental and predicted K_i values (Quasar and Raptor) as well as

Table 3. Quantitative aspects of (*R*/*S*)-lorazepam binding to the GR.[a]

Isomer	$E_{ligand-receptor}$	($E_{electrostatic}$	E_{vdW}	$E_{H\ bonding}$	$E_{polarization}$)	$E_{ligand\ desolvation}$	$T\Delta S$	$E_{ligand\ internal\ strain}$	$E_{induced\ fit}$	E_{bind}	K_i [nM]
R	−14.3	(−6.3	−2.1	−2.5	−3.4)	+3.6	+0.5	+0.4	+0.3	−9.5	83
S	−15.2	(−6.7	−4.4	−0.0	−4.1)	+3.6	+0.5	+0.1	+0.2	−10.8	9.1

[a] All energies are given in kcal mol^{-1}; E_{vdW}: energy contribution from van der Waals interactions; E_{bind}: calculated binding energy; K_i: calculated binding affinity. $E_{ligand-receptor} = E_{electrostatic} + E_{vdW} + E_{H\ bonding} + E_{polarization}$ [cf. Equation (1)].

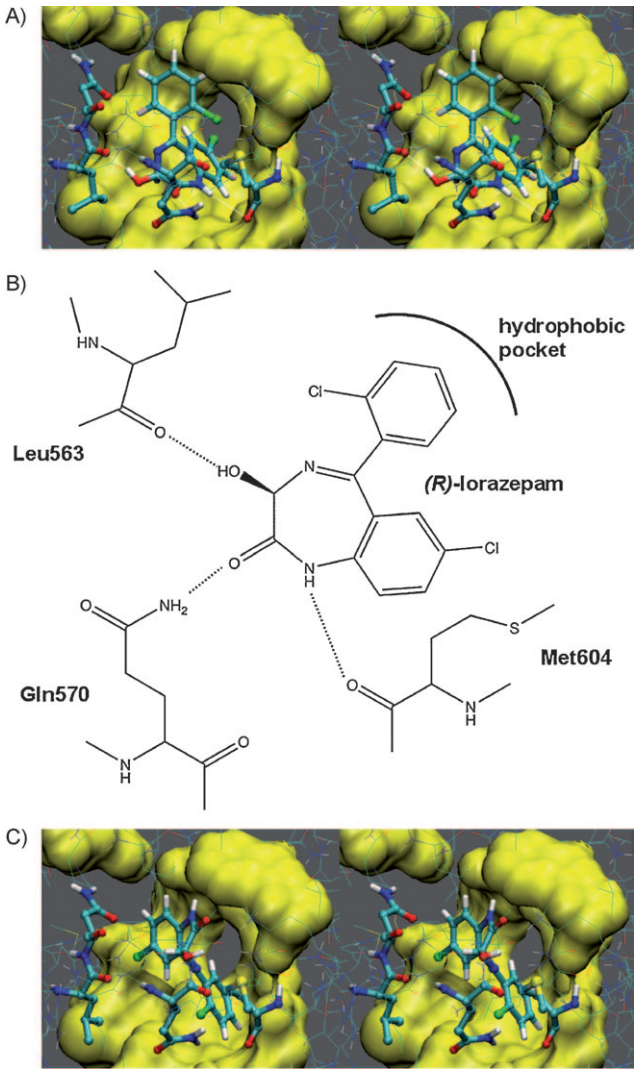

Figure 9. Details of the binding mode of lorazepam: Lorazepam, Leu563, and Asn564 are colored by atom type, and the residues lining the hydrophobic pocket are depicted in yellow. A) Stereo view: *R*-lorazepam engages in a stable hydrogen bond with Leu563 and two weak interactions with Gln570 and Met604 (dashed lines). B) Schematic representation of the hydrogen-bonding pattern associated with the binding of *R*-lorazepam to the GR. C) Stereo view: *S*-lorazepam does not engage in hydrogen bonds, but features stronger hydrophobic interactions.

details on the scramble tests are available as Supporting Information for this article.

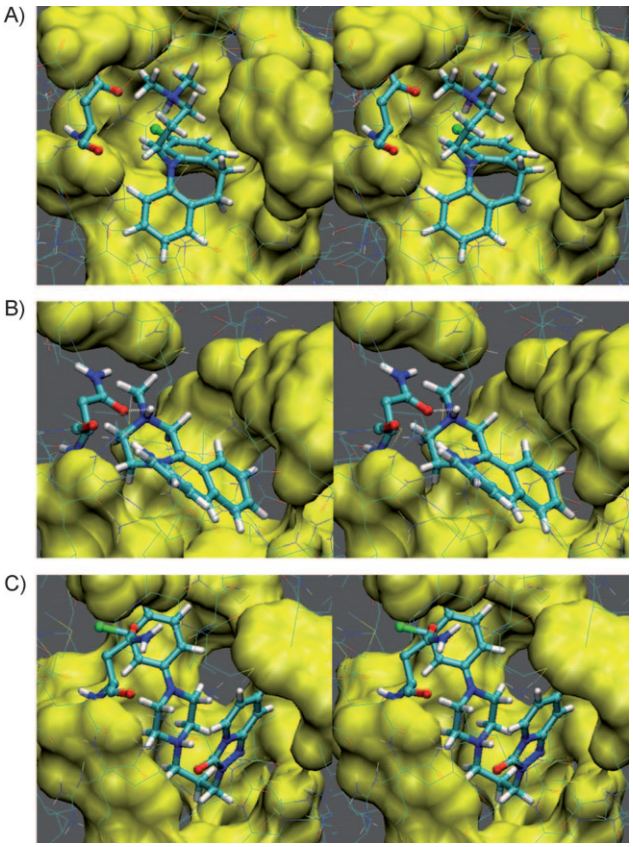

Figure 10. Details of the binding mode of three psychotropic drugs to the GR (stereo view). The ligands and Asn564 are colored by atom type, and the residues lining the hydrophobic pocket are depicted in yellow. A) Clomipramine: the ammonium H atom does not engage in a hydrogen bond with the GR. B) Mirtazapine: the hydrogen bond between the ammonium H atom and Asn564 (white dashed line) is too long (2.37 Å), not linear (127°), and therefore very weak. C) Trazodone: again, the ammonium H atom cannot engage in a hydrogen bond with the receptor.

Acknowledgement

Financial support by the Margaret and Francis Fleitmann Foundation, Luzern (Switzerland) and the Jacques and Dolly Gazan Foundation, Zug (Switzerland) is gratefully acknowledged.

Keywords: consensus scoring · flexible docking · glucocorticoid receptor · induced-fit simulation · multidimensional QSAR

[1] R. W. Weatherman, R. J. Fletterick, T. S. Scanlan, *Annu. Rev. Biochem.* **1999**, *68*, 559–581.
[2] T. Damstra, S. Barlow, A. Bergman, R. Kavlock, G. Van der Kraak, *Global Assessment of the State-of-the-Science of Endocrine Disruptors*, IPCS-WHO, **2002**, p. 1.
[3] J. A. McLachlan, S. F. Arnold, *Am. Sci.* **1996**, *84*, 452–461.
[4] R. C. Kaltreider, A. M. Davis, J. P. Lariviere, J. W. Hamilton, *Environ. Health Perspect.* **2001**, *109*, 245–251.
[5] M. Johansson, S. Nilsson, B. O. Lund, *Environ. Health Perspect.* **1998**, *106*, 769–772.
[6] URL: http://www.epa.gov/oscpmont/oscpendo/pubs/edspoverview/chronology.htm (access date: October 16, 2008).
[7] URL: http://www.nrp50.ch/ (access date: October 16, 2008).
[8] N. Z. Lu, S. E. Wardell, K. L. Burnstein, D. Defranco, P. J. Fuller, V. Giguere, R. B. Hochberg, L. McKay, J. M. Renoir, N. L. Weigel, E. M. Wilson, D. P. McDonnell, J. A. Cidlowski, *Pharmacol. Rev.* **2006**, *58*, 782–797.
[9] a) M. Beato, P. Herrlich, G. Schütz, *Cell* **1995**, *83*, 851–857; b) M. Karin, *Cell* **1998**, *93*, 487–490.
[10] P. J. Barnes, S. Pedersen, W. W. Busse, *Am. J. Respir. Crit. Care Med.* **1998**, *157*, S1–S53.
[11] T. Rhen, J. A. Cidlowski, *N. Engl. J. Med.* **2005**, *353*, 1711–1723.
[12] a) A. Vedani, H. Briem, M. Dobler, H. Dollinger, D. R. McMasters, *J. Med. Chem.* **2000**, *43*, 4416–4427; b) A. Vedani, M. Dobler, *J. Med. Chem.* **2002**, *45*, 2139–2149; c) M. A. Lill, A. Vedani, M. Dobler, *J. Med. Chem.* **2004**, *47*, 6174–6186; d) A. Vedani, H. Dollinger, K. M. Hasselbach, M. Dobler, M. A. Lill, *J. Med. Chem.* **2005**, *48*, 1515–1527.
[13] a) A. Vedani, M. Dobler, M. Spreafico, O. Peristera, M. Smiesko, *ALTEX* **2007**, *24*, 153–161; b) A. Vedani, M. Spreafico, O. Peristera, M. Dobler, M. Smiesko, *Chimia* **2008**, *5*, 322–328.
[14] R. B. Bledsoe, V. G. Montana, T. B. Stanley, C. J. Delves, C. J. Apolito, D. D. Mckee, T. G. Consler, D. J. Parks, E. L. Stewart, T. M. Willson, M. H. Lambert, J. T. Moore, K. H. Pearce, H. E. Xu, *Cell* **2002**, *110*, 93–105.
[15] URL: http://www.rcsb.org/pdb (access date: October 16, 2008).
[16] a) A. Vedani, D. W. Huhta, *J. Am. Chem. Soc.* **1990**, *112*, 4759–4767; b) URL: http://www.biograf.ch/downloads/yeti.pdf (access date: October 16, 2008).
[17] a) S. Hammer, I. Spika, W. Sippl, G. Jessen, B. Kleuser, H. D. Höltje, M. Schäfer, *Steroids* **2003**, *68*, 329–339; b) M. J. Coghlan, P. R. Kym, S. W. Elmore, A. X. Wang, J. R. Luly, D. Wilcox, M. Stashko, C. W. Lin, J. Miner, C. Tyree, M. Nakane, P. Jacobson, B. C. Lane, *J. Med. Chem.* **2001**, *44*, 2879–2885; c) S. W. Elmore, M. J. Coghlan, D. D. Anderson, J. K. Pratt, B. E. Green, A. X. Wang, M. A. Stashko, C. W. Lin, C. M. Tyree, J. N. Miner, P. B. Jacobson, D. M. Wilcox, B. C. Lane, *J. Med. Chem.* **2001**, *44*, 4481–4491; d) A. Ali, C. F. Thompson, J. M. Balkovec, D. W. Graham, M. L. Hammond, N. Quraishi, J. R. Tata, M. Einstein, L. Ge, G. Harris, T. M. Kelly, P. Mazur, S. Pandit, J. Santoro, A. Sitlani, C. Wang, J. Williamson, D. K. Miller, C. M. Thompson, D. M. Zaller, M. J. Forrest, E. Carballo-Jane, S. Luell, *J. Med. Chem.* **2004**, *47*, 2441–2452; e) M. Einstein, M. Greenlee, G. Rouen, A. Sitlani, J. Santoro, C. Wang, S. Pandit, P. Mazur, I. Smalera, A. P. M. Weaver, Y. Y. Zeng, L. Ge, T. Kelly, T. Paiva, W. Geissler, R. T. Mosley, J. Williamson, A. Ali, J. Balkovec, G. Harris, *J. Steroid Biochem. Mol. Biol.* **2004**, *92*, 345–356; f) C. J. Smith, A. Ali, J. M. Balkovec, D. W. Graham, M. L. Hammond, G. F. Patel, G. P. Rouen, S. K. Smith, J. R. Tata, M. Einstein, L. Ge, G. S. Harris, T. M. Kelly, P. Mazur, C. M. Thompson, C. F. Wang, J. M. Williamson, D. K. Miller, S. Pandit, J. C. Santoro, A. Sitlani, T. D. Yamin, E. A. O'Neill, M. D. Zaller, E. Carballo-Jane, M. J. Forrest, S. Luell, *Bioorg. Med. Chem. Lett.* **2005**, *15*, 2926–2931.
[18] H. C. Cheng, *J. Pharmacol. Toxicol. Methods* **2001**, *46*, 61–71.
[19] S. J. Weiner, P. A. Kollmann, D. A. Case, U. C. Singh, C. Ghio, G. Alagona, Jr., S. Profeta, P. Weiner, *J. Am. Chem. Soc.* **1984**, *106*, 765–784.
[20] F. Mohamadi, N. G. J. Richards, W. Guida, R. Liskamp, M. Lipton, C. Caufield, G. Chang, T. Hendrickson, W. C. Still, *J. Comput. Chem.* **1990**, *11*, 440–467.
[21] J. W. Storer, D. J. Giesen, C. J. Cramer, D. G. Truhlar, *J. Comput.-Aided Mol. Des.* **1995**, *9*, 87–110.
[22] J. P. Schmit, G. G. Rousseau, *Monogr. Endocrinol.* **1979**, *12*, 79–95.
[23] Varied notations are observed in the literature. Herein, training set refers to the compounds used to generate the model, and test set to those predicted by the model. The test set was subsequently used to select the best model among different boundary conditions for the simulations. In contrast, our prediction set refers to compounds solely tested by the final model.
[24] A. Vedani, M. Zumstein, M. A. Lill, B. Ernst, *ChemMedChem* **2007**, *2*, 78–87.
[25] A. Vedani, M. Dobler, M. A. Lill, *J. Med. Chem.* **2005**, *48*, 3700–3703.
[26] A. Vedani, M. Smiesko, in *Computational Structural Biology* (Eds.: M. Peitsch, T. Schwede) Imperial College Press, London, **2008**, pp. 549–572.
[27] A. R. Leach, B. K. Shoichet, C. E. Peishoff, *J. Med. Chem.* **2006**, *49*, 5851–5855.
[28] K. P. Madauss, S. J. Deng, R. J. H. Austin, M. H. Lambert, I. McLay, J. Pritchard, S. A. Short, E. L. Stewart, I. J. Uings, S. P. Williams, *J. Med. Chem.* **2004**, *47*, 3381–3387.
[29] M. A. Lill, F. Winiger, A. Vedani, B. Ernst, *J. Med. Chem.* **2005**, *48*, 5666–5674.
[30] S. Wold, L. Eriksson, in *Chemometrics Methods in Molecular Design* (Ed.: H. van de Waterbeemd) VCH, Weinheim, **1995**, pp. 309–318.
[31] N. Shah, T. S. Scanlan, *Bioorg. Med. Chem. Lett.* **2004**, *14*, 5199–5203.
[32] URL: http://www.biograf.ch/downloads/VirtualToxLab.pdf (access date: October 16, 2008).
[33] URL: http://www.biograf.ch/data/projects/virtualtoxlab_results.php (access date: October 16, 2008).
[34] URL: http://www.en.wikipedia.org/wiki (access date: October 16, 2008).
[35] J. E. Ottenweller, W. N. Tapp, B. H. Natelson, *Psychopharmacology* **1989**, *98*, 369–371.
[36] H. Vedder, U. Bening-Abu-Shach, S. Lanquillon, J. C. Krieg, *J. Psychiatr. Res.* **1999**, 33, 303–308.
[37] C. M. Pariante, A. Hye, R. Williamson, A. Makoff, S. Lovestone, R. W. Kerwin, *Neuropsychopharmacology* **2003**, *28*, 1553–1561.
[38] J. D. K. Uys, C. J. F. Muller, L. Marais, B. H. Harvey, D. J. Stein, W. M. U. Daniels, *Neuroscience* **2006**, *137*, 619–625.
[39] J. L. Yaul, J. Noble, K. E. Chapman, J. R. Seckl, *Mol Brain Res.* **2004**, *129*, 189–192.
[40] K. Mukherjee, A. Knisely, L. Jacobson, *Endocrinology* **2004**, *145*, 4185–4191.
[41] S. A. Hill, M. J. Taylor, C. J. Harmer, P. J. Cowen, *J. Psychopharmacol.* **2003**, *17*, 273–275.
[42] R. Yehuda, R. K. Yang, J. A. Golier, R. A. Grossman, L. M. Bierer, L. Tischler, *Neuropsychopharmacology* **2006**, *31*, 189–196.
[43] a) M. S. Lee, J. W. Yang, Y. H. Ko, H. Changsu, S. H. Kim, M. S. Lee, S. H. Joe, I. K. Jung, *Child Psychiatry Hum. Dev.* **2008**, *39*, 201–209; b) C. R. Sharpe, J.-P. Collet, E. Belzile, J. A. Hanley, J.-F. Boivin, *Br. J. Cancer* **2002**, *86*, 92–97.
[44] a) P. Buchwald, *J. Pharm. Sci.* **2005**, *94*, 2355–2379; b) M. Barker, M. Clackers, D. A. Demaine, D. Humphreys, M. J. Johnston, H. T. Jones, F. Pacquet, J. M. Pritchard, M. Salter, S. E. Shanahan, P. A. Skone, V. M. Vinader, I. Uings, I. M. McLay, S. J. F. Macdonald, *J. Med. Chem.* **2005**, *48*, 4507–4510.
[45] A. Vedani, A.-V. Descloux, M. Spreafico, B. Ernst, *Tox. Lett.* **2007**, *173*, 17–23.

Received: August 21, 2008
Revised: September 10, 2008
Published online on November 13, 2008

