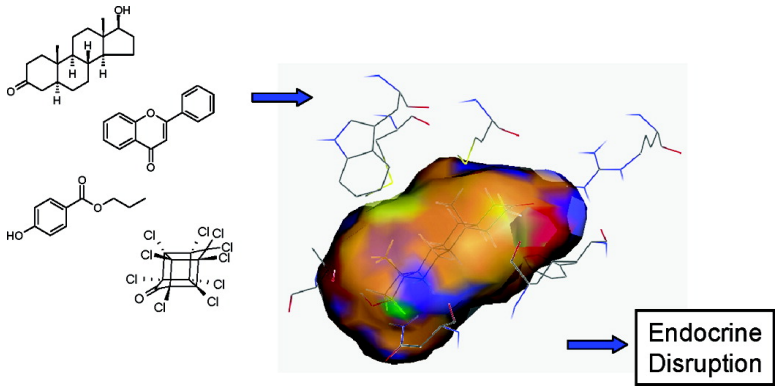
Journal of Medicinal Chemistry

Article

Impact of Induced Fit on Ligand Binding to the Androgen Receptor: A Multidimensional QSAR Study To Predict Endocrine-Disrupting Effects of Environmental Chemicals

Markus A. Lill, Fabienne Winiger, Angelo Vedani, and Beat Ernst

J. Med. Chem., **2005**, 48 (18), 5666-5674• DOI: 10.1021/jm050403f • Publication Date (Web): 13 August 2005



More About This Article

Additional resources and features associated with this article are available within the HTML version:

- Supporting Information
- Links to the 3 articles that cite this article, as of the time of this article download
- Access to high resolution figures
- Links to articles and content related to this article
- Copyright permission to reproduce figures and/or text from this article

View the Full Text HTML

Impact of Induced Fit on Ligand Binding to the Androgen Receptor: A Multidimensional QSAR Study To Predict Endocrine-Disrupting Effects of Environmental Chemicals

Markus A. Lill,[†,‡,*] Fabienne Winiger,[†] Angelo Vedani,[†,‡] and Beat Ernst[†]

Institute for Molecular Pharmacy, University of Basel, Klingelbergstrasse 50, CH-4056 Basel, Switzerland, and Biographics Laboratory 3R, Friedensgasse 35, CH-4056 Basel, Switzerland

Received April 28, 2005

We investigated the influence of induced fit of the androgen receptor binding pocket on free energies of ligand binding. On the basis of a novel alignment procedure using flexible docking, molecular dynamics simulations, and linear-interaction energy analysis, we simulated the binding of 119 molecules representing six compound classes. The superposition of the ligand molecules emerging from the combined protocol served as input for *Raptor*, a receptor-modeling tool based on multidimensional QSAR allowing for ligand-dependent induced fit. Throughout our study, protein flexibility was explicitly accounted for. The model converged at a cross-validated $r^2 = 0.858$ (88 training compounds) and yielded a predictive $r^2 = 0.792$ (26 test compounds), thereby predicting the binding affinity of all compounds close to their experimental value. We then challenged the model by testing five molecules not belonging to compound classes used to train the model: the IC_{50} values were predicted within a factor of 4.5 compared to the experimental data. The demonstrated predictivity of the model suggests that our approach may well be beneficial for both drug discovery and the screening of environmental chemicals for endocrine-disrupting effects, a problem that has recently become a cause for concern among scientists, environmental advocates, and politicians alike.

Introduction

Nuclear receptors represent the largest family of ligand-dependent eukaryotic transcription factors transforming extra- and intracellular signals into cellular responses by triggering the transcription of target genes. In particular, they mediate the effects of hormones and other endogenous ligands to regulate the expression of specific genes, thereby regulating development and metabolism. Among other members, this family includes receptors for the various steroid hormones, e.g. the androgen, estrogen, glucocorticoid, and progesterone receptor. Unbalanced production or cell insensitivity to specific hormones may result in diseases associated with human endocrine dysfunction.[1] Androgens and the androgen receptor (AR) play an essential role in the growth of normal prostate. They are, however, also involved in the development of prostate cancer,[2] representing the most common male malignancy in the United States. Both steroidal and nonsteroidal derivatives have shown clinical benefits as chemotherapeutic agents for prostate cancer. Still, several of these antiandrogens, for example cyprosterone, show overlapping effects with other hormonal systems.[3] As the binding sites of steroid hormone receptors share a common topology: a hydrophobic cavity accommodating the steroid-ligand scaffold and a hydrogen-bonding pattern, typically including the two terminal polar groups of the ligand molecule (cf. Figure 1, upper panel); these overlapping effects may well be seen as a common phenomenon.

In the complex of 5α-dihydrotestosterone with the androgen receptor,[4] the ligand forms a hydrogen bond via the oxygen of its carbonyl group in the A ring with Arg752 and two through its hydroxyl group with Asn705 and Tyr877, respectively. The aliphatic scaffold of the ligand molecule is accommodated by a hydrophobic pocket consisting of Leu707, Gly708, Trp741, Met742, Met745, Phe764, Met787, Leu873, and Met895.

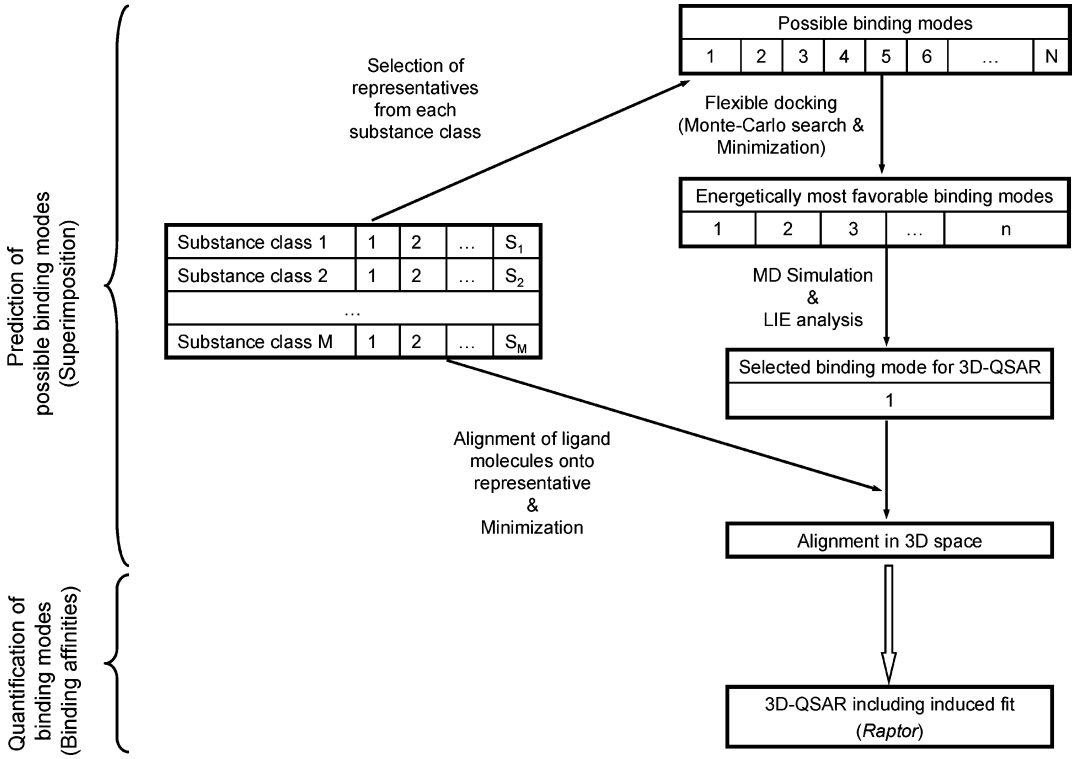
Many environmental chemicals, e.g. flavones or kepone, display similar structural properties: a hydrophobic core and one or two terminal polar groups. Consequently, they may bind to a nuclear receptor influencing the balance of the endocrine system.[1] The presence of these so-called endocrine disruptors in the biosphere has become a worldwide environmental concern. It has been concluded that such compounds elicit a variety of adverse effects in both humans and wildlife including promotion of hormone-dependent cancers, reproductive tract disorders, and a reduction in reproductive fitness. A variety of compounds in the environment have been shown to display agonistic or antagonistic activity toward the androgen receptor, including both natural products and synthetic compounds.

The concern over xenobiotics binding to the androgen receptor has created a need to both screen and monitor compounds expected to modulate endocrine effects. We therefore developed an in silico model to quantitatively predict the potential of structurally diverse ligands for binding to the androgen receptor using a multidimensional QSAR technique, specifically allowing for induced-fit. To identify the binding mode at the true biological receptor, a novel stepwise protocol consisting of flexible docking, molecular dynamics (MD) simulations, and

* Corresponding author. E-mail: markus@biograf.ch, Phone: ++41 61 2614259, Fax: ++41 61 2614258.
† University of Basel.
‡ Biographics Laboratory 3R.

10.1021/jm050403f CCC: $30.25
Published on Web 08/13/2005

Scheme 1. Flowchart of the Computational Procedure Including Alignment and Prediction of Binding Affinities



linear interaction-energy analysis (LIE) was developed (Scheme 1).

The Importance of Induced Fit for Ligand Binding to the Androgen Receptor. As the compounds binding to the androgen receptor (Figure 2) vary substantially in their structure, identifying the correct binding mode is all but unambiguous. The suitability of a QSAR model for predictive purposes depends critically on the alignment of the ligand molecules. Two recent QSAR studies[5] on the androgen receptor also used docking techniques to derive a protein-based alignment. However, in those studies, the receptor was

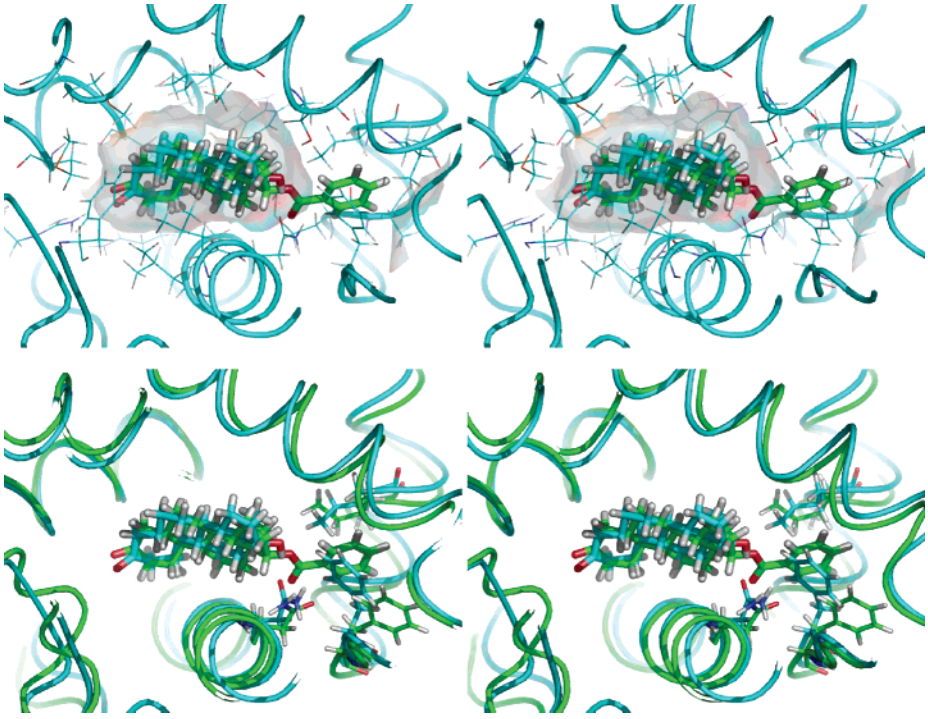


Figure 1. Stereo representation of DHT (color = cyan) and DHT benzoate binding (color = green) to the androgen receptor. Upper panel: Docking of DHT benzoate to the X-ray structure of the androgen receptor−DHT complex (color = cyan) would not provide enough space to accommodate additional steric bulk at the 17β-position. To display the ligand accessible volume of the binding pocket of the androgen receptor−DHT complex, the solvent-accessible surface is shown. Lower panel: Local induced fit as simulated with MD accommodates the additional volume of DHT benzoate.

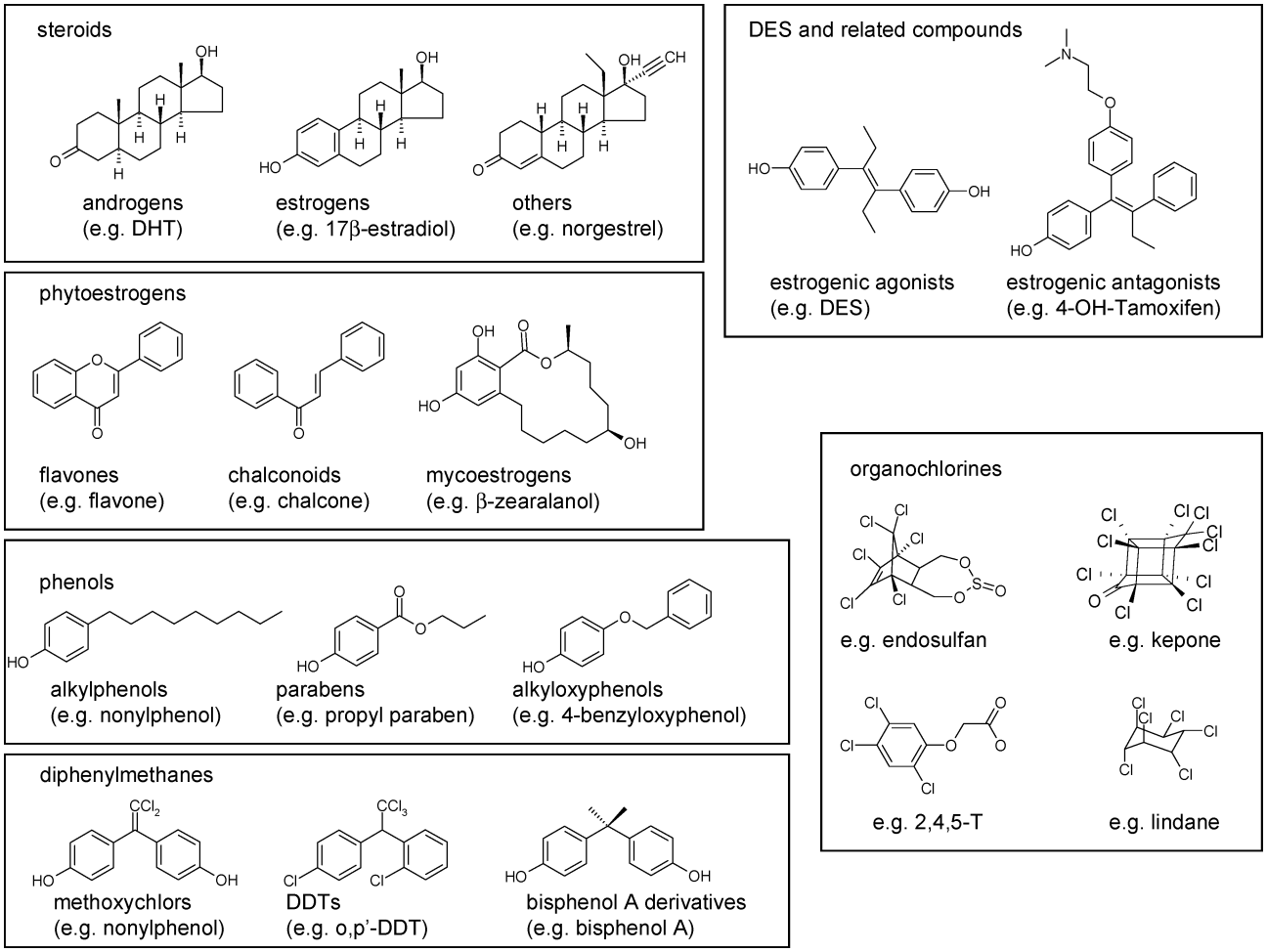


Figure 2. Compound classes and representative molecules used in our QSAR study.

kept rigid. Recently, it was demonstrated[6] that the progesterone receptor, another member of the steroid-hormone class of nuclear receptors, tolerates larger substituents at the 17α-position of the steroid scaffold as compared to the endogeneous ligand progesterone. These results postulate a local induced-fit and suggest that similar phenomena might be observed for the androgen receptor as well. Thus, DHT containing an additional benzoate moiety at the 17*β*-position binds with 260 nM to the androgen receptor. The experimental 3D structure of the androgen receptor−DHT complex as rigid template cannot provide enough space to accommodate the benzoate portion (Figure 1, upper panel). When performing MD simulations of DHT benzoate binding to the androgen receptor (Figure 1, lower panel), the side chain of Phe891 is relocated by about 3 Å, making room to accommodate the bulky benzoate fragment at the 17*β*-position; in addition, Ile899 and Asn705 reorient their side chains.

Materials and Methods

Ligand Data and Molecular Structure Building. The structural and pharmacological data for 119 androgen-receptor binding compounds were obtained from Fang et al.[7] The affinity measurements have been performed using a competitive binding assay with recombinant rat protein expressed in *Escherichia coli*. The amino acid sequence of the ligand binding domain is identical to that of the human androgen receptor ligand binding domain. ^{3}H-Methyltrienolone (R 1881) was used as the high-affinity ligand in the measurements.

The three-dimensional structure of all ligand molecules was generated using *MacroModel* 6.5[8] and optimized in aqueous solution based on the *AMBER** force field.[9] The partial charges for the flexible docking study were computed using the ESP methodology as implemented in *MOPAC*.[10]

Alignment Procedure. The 3D structure of the androgen receptor complexed by dihydrotestosterone (DHT) is available at a resolution of 2.0 Å[4] and was used as a template for docking (receptor-mediated alignment). To account for induced fit, we used flexible docking based on a Monte Carlo search protocol (software *Yeti*)[11] and allowing for an adaptation of the protein residue side-chains to each individual ligand molecule. Multiple docking solutions were identified with a similar score. Even a directional force field[11] is not truly suited for scoring ligand−protein interactions. As demonstrated by several authors, sampling different substates of the ligand−receptor space (e.g. free-energy perturbation) improves the quality of the prediction of binding affinities.[12] In our approach, we chose the orientations with the most favorable scores during flexible docking and, subsequently, performed MD simulations to generate an ensemble of configurations for each binding mode. To quantify the relative free energy of ligand binding of the different binding patterns, we applied linear-interaction energy (LIE) analysis.[13] LIE quantifies the free energy of a compound in a given binding mode, subtracting electrostatic and van der Waals interaction energies with solvent averaged over the entire simulation from the corresponding energies when bound to the protein:

$$\Delta G = \alpha(\langle E_{\mathrm{lig-prot}}{}^{\mathrm{elec}}\rangle - \langle E_{\mathrm{lig-solv}}{}^{\mathrm{elec}}\rangle) + \beta(\langle E_{\mathrm{lig-prot}}{}^{\mathrm{VdW}}\rangle - \langle E_{\mathrm{lig-solv}}{}^{\mathrm{VdW}}\rangle) \quad (1)$$

Table 1. Linear-Interaction Energy Analysis for Three Different Orientations of Nonylphenol and Dihydroxymethoxychlorolefin As Identified by the Flexible-Docking Search Protocol (all values are given in kcal/mol)'

Compound	Binding mode	$\langle E^{\mathrm{VdW}}{}_{\mathrm{lig-prot}}\rangle$	$\langle E^{\mathrm{elec}}{}_{\mathrm{lig-prot}}\rangle$	$\Delta\Delta G_{\mathrm{i-lowest}}{}^{a}$
Nonylphenol	A	−19.2	−5.8	1.7
	B	−20.4	−5.6	1.2
	C	−18.8	−9.5	0.0
Dihydroxymethoxychlorolefin	A	−18.5	−6.8	2.9
	B	−19.5	−9.2	1.2
	C	−18.5	−12.5	0.0

[a] The $\Delta\Delta G_{\mathrm{i-lowest}}$ values were calculated according to eq 2 using $\alpha = \beta = 0.5$.

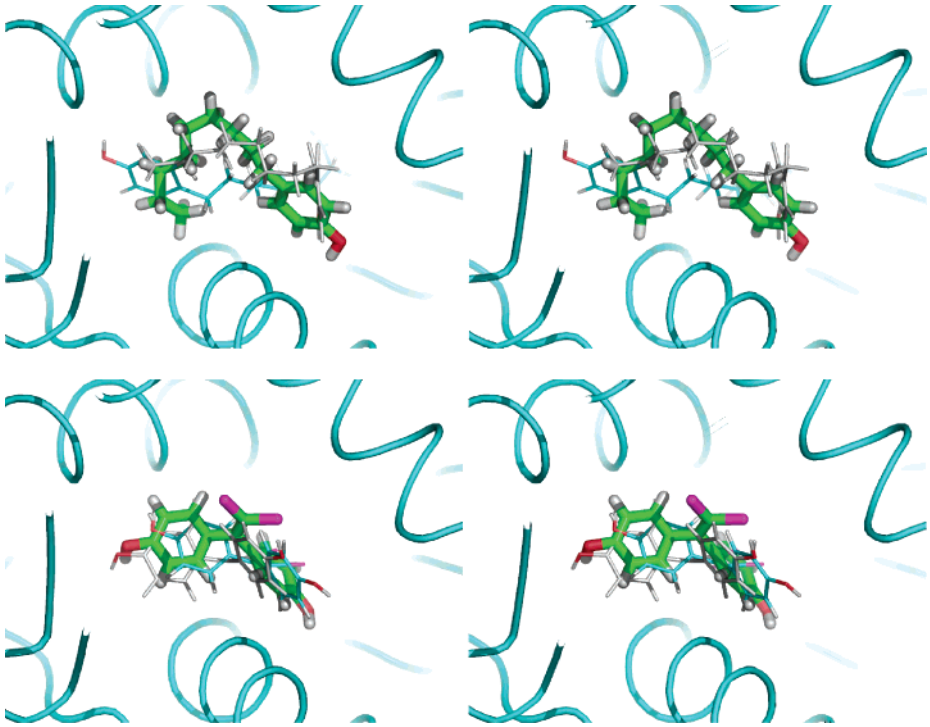


Figure 3. Stereo representation of the superposition of the three different orientations of nonylphenol (upper panel) and dihydroxymethoxychlorolefin (lower panel). The most favorable binding mode is highlighted in green (cf. Table 1).

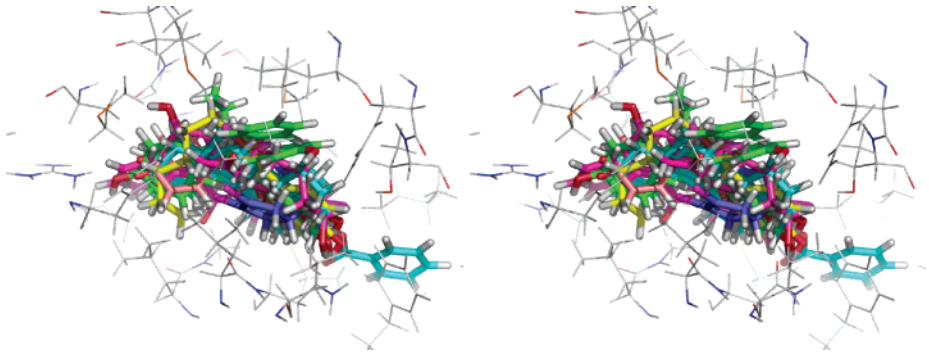


Figure 4. Superposition of the molecules representing the different compound classes (one each) used in our QSAR study. The binding pocket of the experimental androgen receptor−DHT complex displays steric clashes when used as a rigid template for docking.

For quantifying the relative energies of the two binding modes A and B, this equation reduces to

$$\Delta\Delta G_{\mathrm{AB}} = \Delta G_{\mathrm{A}} - \Delta G_{\mathrm{B}} = \alpha(\langle E_{\mathrm{lig-prot}}{}^{\mathrm{elec}}\rangle_A - \langle E_{\mathrm{lig-prot}}{}^{\mathrm{elec}}\rangle_B) + \beta(\langle E_{\mathrm{lig-prot}}{}^{\mathrm{VdW}}\rangle_A - \langle E_{\mathrm{lig-prot}}{}^{\mathrm{VdW}}\rangle_B) \quad (2)$$

as the interaction energies with the solvent are identical. While the parameter $\alpha = 0.5$ is theoretically derived,[13] values for the parameter β vary between 0.1 for predominantly hydrophilic binding sites and 0.9 for predominantly hydrophobic pockets.[14] As the androgen receptor features a large hydrophobic moiety and two polar regions, we chose $\beta = 0.5$.

As this procedure is computationally demanding,[15] we simplified the procedure assuming that the overall binding mode remains unaltered within each compound class. Consequently, we did not apply the analysis to each individual compound, but solely to one representative ligand per compound class (cf. Figure 2). The remaining molecules of the very compound class were then aligned to the representative ligand molecule, where the common scaffold served as template. Finally, each individual ligand−receptor complex was energetically refined allowing for subtleties in the ligand−receptor adaptation.

MD Simulations. The parameters for the ligand molecules were assigned using the *antechamber* module of *AMBER 7.0*.[16]

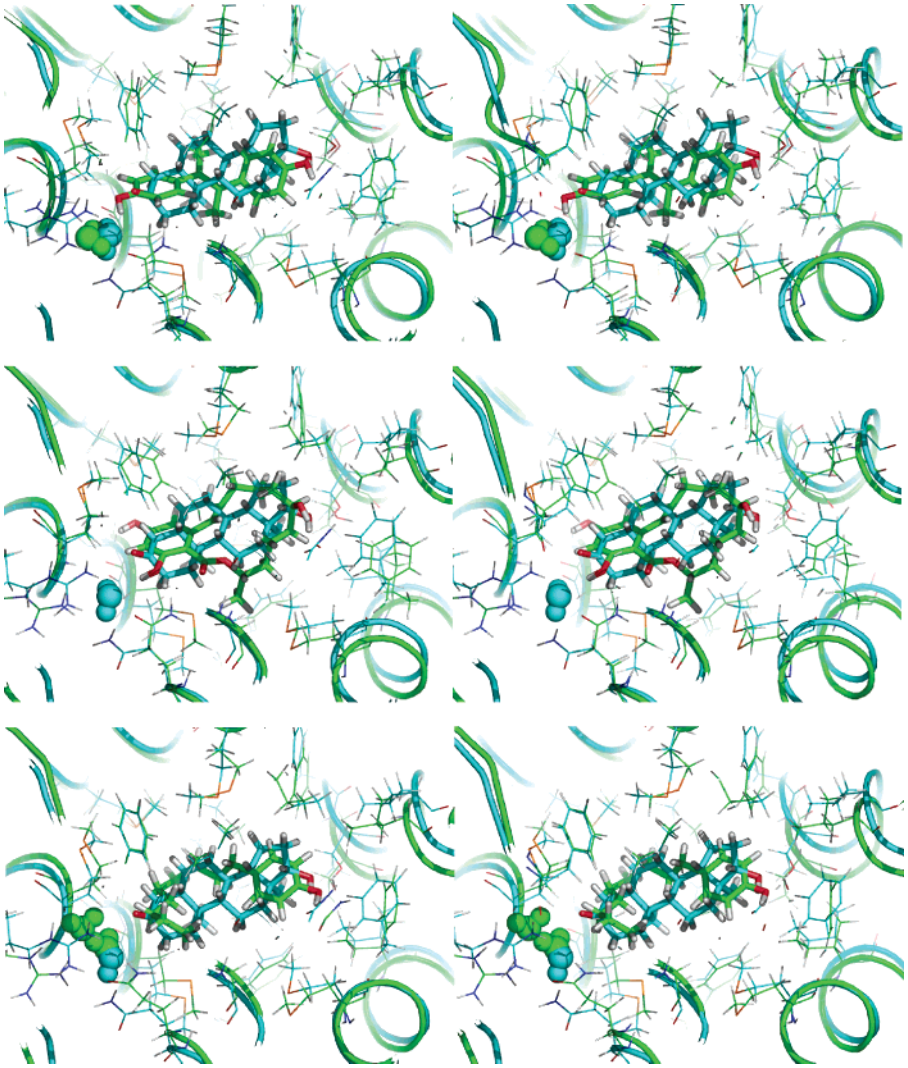


Figure 5. Structures of ligand−receptor complexes for dimethylstilbestrol (upper panel), α-zearalenol (middle panel), and nonylphenol (lower panel) binding to the AR, as obtained from MD simulations. The ligand molecules and protein are shown as sticks, water molecules as spheres. The carbon atoms are colored in green. The structures are overlaid onto the DHT−AR complex (colored in cyan) to demonstrate the observed induced protein fit.

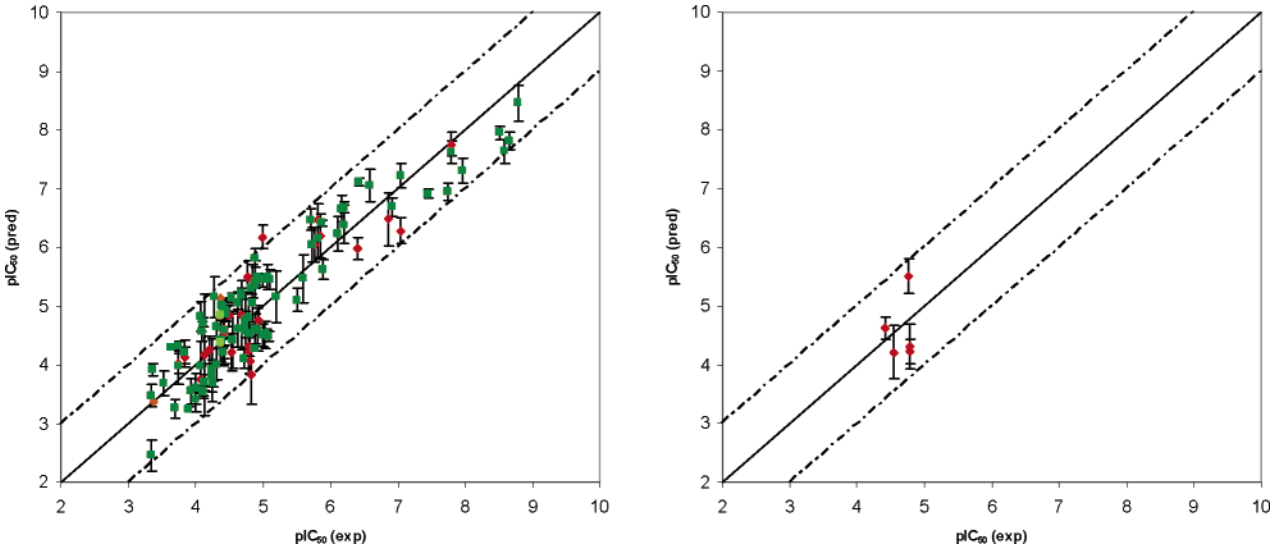


Figure 6. Comparison of predicted and experimental binding affinities for the 119 molecules used in our study. Left: training set (dark green/light green; model building) and test set (red/orange; model validation). Threshold compounds are shown in light green and orange, respectively. Right: prediction set.

Table 2. Experimental and Calculated IC_{50} for Compounds of the Training and Test Set Binding to the Androgen Receptor

Compound	pIC_{50} (exp)	pIC_{50} (pred)	error in pIC_{50}	Compound	pIC_{50} (exp)	pIC_{50} (pred)	error in pIC_{50}
			Test Set				
β-3β-Androstanediol	6.87	6.48	0.39	Nonylphenol	4.94	4.76	0.18
Methyltestosterone	7.80	7.76	0.04	4-*tert*-Butylphenol	3.84	4.12	0.29
Androstenediol	5.86	6.19	0.33	4-*sec*-Butylphenol	4.07	3.75	0.31
5α-Androstan-3β-ol	5.77	6.04	0.27	4-Heptyloxy-phenol	4.83	3.84	0.99
β-17β-Estradiol	6.39	5.98	0.41	Methylparaben	<3.37	3.17	
3-Deoxyestradiol	7.05	6.28	0.77	Vanillin	<3.37	2.12	
Progesterone	5.81	6.46	0.65	Monohydroxy-methoxychlorolefin	4.69	4.84	0.16
Prednisolone	<4.37	3.80		*p,p′*-DDD	4.81	4.07	0.74
Cholesterol	<4.37	4.84		Bisphenol A	4.13	4.17	0.04
Dimethylstilbestrol	4.85	5.43	0.58	4-Hydroxybenzophenone	3.74	4.02	0.28
7-Hydroxyflavone	<4.37	5.13		Chlordane	5.00	6.17	1.17
Chalcone	4.19	4.25	0.06	Aldrin	4.49	4.83	0.34
β-Zearalenol	4.42	4.54	0.12	Lindane	4.39	4.58	0.19
			Training Set				
Methyltrienolone (R1881)	8.51	7.95	0.56	Nafoxidine	4.89	4.62	0.27
Mibolerone	8.78	8.45	0.33	Clomiphene	4.87	5.34	0.46
Trenbolone	8.58	7.63	0.95	Flavone	4.11	3.53	0.58
α-3α-Androstanediol	5.70	6.47	0.77	6-Hydroxyflavone	3.74	4.32	0.58
5α-Androstan-17β-ol	7.96	7.31	0.65	Flavanone	4.26	3.68	0.58
β-Testosterone	5.51	5.10	0.40	4′-Hydroxyflavanone	4.24	3.81	0.43
α-Epitestosterone	7.80	7.62	0.18	6-Hydroxyflavanone	4.73	4.75	0.02
α-5α-Dihydrotestosterone	8.65	7.82	0.84	Genistein	4.07	3.97	0.09
β-5β-Dihydrotestosterone	6.42	7.11	0.69	Equol	4.13	3.70	0.43
11-Keto-ketosterone	7.05	7.23	0.17	Coumestrol	<4.37	3.57	
4-Androstenediol	6.20	6.39	0.18	4′-Hydroxychalcone	4.24	3.91	0.33
Testosterone proprionate	5.72	6.05	0.33	4-Hydroxychalcone	4.32	4.01	0.31
DHT benzoate	6.59	7.05	0.46	β-Zearalanol	4.80	4.79	0.01
4-Androstenedion	5.89	5.61	0.28	Zearalanone	4.37	4.41	0.04
Androsterone	4.39	4.98	0.59	α-Zearalenol	5.00	5.45	0.45
5α-Androstan-3,11,17-trione	4.88	5.82	0.95	4-*n*-Octylphenol	4.72	4.11	0.61
5,6-Didehydroisoandrosterone	4.53	5.13	0.60	4-Dodecylphenol	4.69	5.18	0.49
α-17α-Estradiol	4.12	4.72	0.61	4-*tert*-Amylphenol	4.11	3.56	0.55
4-Hydroxyestradiol	5.59	5.47	0.13	2-*sec*-Butylphenol	3.99	3.59	0.40
2-Hydroxyestradiol	5.07	5.49	0.42	4-Chloro-2-methylphenol	3.92	3.55	0.37
17-Deoxyestradiol	4.38	5.01	0.63	3-Chlorophenol	3.34	3.47	0.14
Estriol	3.36	3.91	0.55	Isoeugenol	3.69	3.25	0.43
Ethinylestradiol	5.09	5.45	0.36	Igepal CO-210	4.74	4.56	0.18
3-Methylestriol	4.27	5.15	0.88	Propylparaben	3.51	3.68	0.17
16β-OH-16α-Me-3Me-estradiol	4.43	4.94	0.50	4-Benzyloxy-phenol	3.63	4.29	0.67
Norethynodrel	5.81	6.15	0.33	Dihydroxymethoxychlorolefin	5.20	5.16	0.04
Norethindrone	6.92	6.69	0.24	*p,p′*-Methoxychlor	4.55	4.42	0.13
Norgestrel	7.73	6.94	0.79	HPTE	5.04	4.53	0.51
Promegestone	5.88	6.42	0.54	*p,p′*-Methoxy-chlorolefin	4.31	4.65	0.34
6α-Methyl-17α-hydroxyprogesterone	6.10	6.23	0.13	*o,p′*-DDT	4.83	5.31	0.48
6α-Methyl-17α-hydroxyprogesterone acetate	7.45	6.91	0.54	*o,p′*-DDD	4.98	4.51	0.48
Cyproterone acetate	6.20	6.68	0.49	*o,p′*-DDE	4.69	5.23	0.55
Corticosterone	4.63	4.62	0.01	*p,p′*-DDT	4.74	4.68	0.06
Cortisol	3.74	3.97	0.23	*p,p′*-DDE	4.81	4.52	0.29
Aldosterone	<4.37	4.85		Bisphenol B	4.42	4.59	0.16
Dexamethasone	4.09	4.56	0.48	p-Cumylphenol	4.40	4.21	0.19
Spironolactone	6.17	6.66	0.49	2,4-Dihydroxy-benzophenone	3.99	3.42	0.57
Diethylstilbestrol	4.85	5.06	0.21	Benzophenone	3.89	3.24	0.65
4,4′-Dihydroxystilbene	4.08	4.83	0.75	4,4′-Dihydroxybenzophenone	3.84	4.21	0.38
3,3′-Dihydroxyhexestrol	4.43	4.87	0.44	4-Hydroxybiphenyl	5.08	4.48	0.60
Hexestrol monomethyl ether	4.89	4.28	0.61	Endosulfan	4.63	5.05	0.43
trans-4-Hydroxystilbene	4.38	4.43	0.05	Heptachlor	4.87	5.39	0.52
Tamoxifen	4.92	4.59	0.33	Kepone	4.92	5.49	0.56
4-Hydroxytamoxifen	5.03	4.51	0.52	2,4,5-T	3.34	2.46	0.88

To be consistent with the underlying force field, the charges were reassigned using AM1-BCC.[17] A solvation cap of 30 Å around the geometric center of the binding pocket was created, and the whole system was minimized using the *sander* module. During the MD simulation (200 ps equilibration, slowly heating from 100 to 300 K, followed by 300 ps of data collection) all atoms more than 25 Å away from the center were restrained to their original position. The *mm_pbsa* module was chosen to compute the average interaction energies between ligand and receptor during the MD simulation, as needed for the LIE analysis.

Multidimensional QSAR. *Raptor*,[18] a receptor-modeling concept developed at our laboratory, explicitly and anisotropically allows for induced fit by a dual-shell representation of the three-dimensional binding-site model, mapped with physicochemical properties (hydrophobic character and hydrogen-bonding propensity) onto it. The inner shell is tailored using the most potent ligand of the training set, and the outer shell accommodates the topology of all molecules from the training set. The adaptation of both field and topology of the receptor surrogate to each ligand is achieved by combining a steric adjustment to the topology of the very ligand and a component due to the attraction or repulsion between ligand and receptor surrogate. The latter is obtained by correlating their physicochemical properties (hydrophobicity and hydrogen-bond propensity) in 3D space.

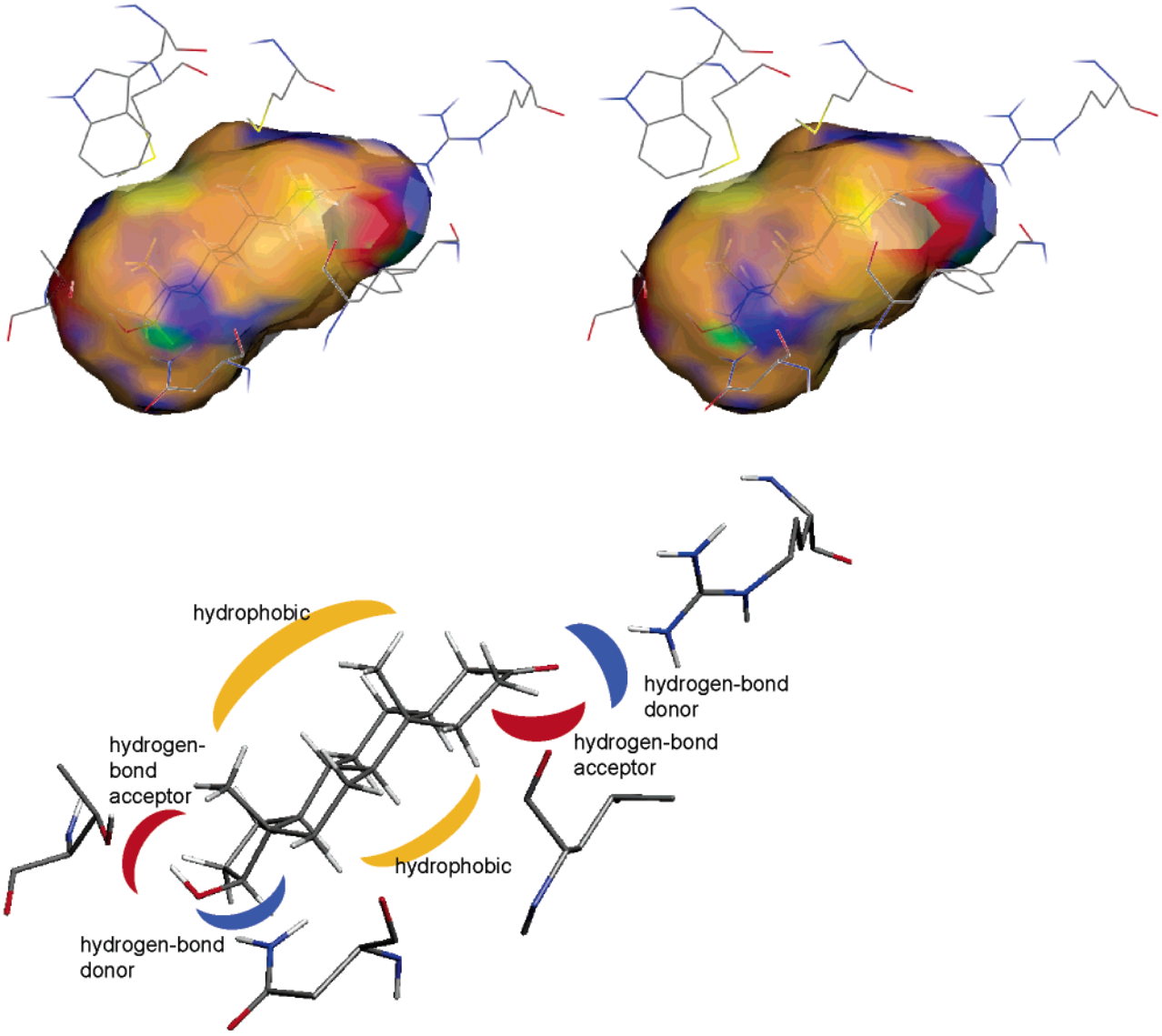


Figure 7. Upper panel: Stereo representation of the binding-site model of the androgen receptor as generated by the *Raptor* technology (beige = hydrophobic fields; blue = hydrogen-bond donating propensity; red = hydrogen-bond accepting propensity; green = hydrogen-bond flip-flop character). Only the inner shell of the *Raptor* model is shown with bound DHT; lower panel: Schematic representation of the physicochemical properties of the receptor model and their consistency with the experimental structure.

The underlying scoring function for evaluating ligand−protein interactions includes directional terms for hydrogen bonding ($\Delta G_{\mathrm{Hbond}}$) and hydrophobicity ($\Delta G_{\mathrm{Hphob}}$) as well as terms for the cost of the topological adaptation (ΔG_{IF}) and the changes in entropy ($T\Delta S$) upon ligand binding:[9]

$$\Delta G = \Delta G_{\mathrm{constant}} + \Delta G_{\mathrm{Hbond}} + \Delta G_{\mathrm{Hphob}} + T\Delta S + \Delta G_{\mathrm{IF}} \quad (3)$$

Experimental determination of binding affinity for weak inhibitors is often prevented due to limited solubility or limited sensitivity; thus, only an upper limit ('threshold') for IC_{50} values is accessible. These substances are typically neglected in QSAR studies, as no experimental binding affinity can be specified. However, these molecules contain valuable information about the binding site of the receptor, i.e., the information that they do not bind strongly to the receptor. Particularly, for lead-finding purposes and for estimating the toxic potential of chemicals, a computational model should also be able to separate strong- and moderate-binding inhibitors from weak and nonbinding compounds. Consequently, we extended the *Raptor* concept with a threshold option: the optimization algorithm forces the model to reproduce the binding affinity of the weak- and nonbinding ligand molecules to be lower than the experimental limit. Obviously, compounds which are experimentally measured to bind weaker than a threshold $IC_{50}^{(t)}$ and are correctly classified during the model optimization do not contribute to the lack-of-fit function. If the corresponding binding affinity of the ligand is predicted higher than the threshold, the lack-of-fit function applies a penalty proportional to $\Delta G^{(t)} - \Delta G$.

Results

Alignment Procedure. Table 1 and Figure 3 summarize the results of the novel alignment protocol for nonylphenol and dihydroxymethoxychlorolefin: the flexible-docking protocol identified three favorable binding modes for both ligands. For both compounds, LIE analysis favorized the energetically lowest entity by 89% ($\Delta\Delta G$ = 1.2 kcal/mol).

Figure 4 shows the relative orientation of a representative molecules of each compound class as identified by flexible docking, MD simulations, and LIE analysis.

The figure demonstrates that the androgen receptor−DHT complex as rigid entity would not provide enough space to accommodate the various compound classes resulting in steric clashes. This suggests the importance of induced-fit during ligand binding to the androgen receptor.

Figure 5 shows the resulting structures of the ligand−receptor complexes. If dimethylstilbestrol is binding to the AR (upper panel) the hydroxyl group, binding to the same moiety, as the carbonyl group at the 3-position of DHT does, induces a conformational flip of the side chain of Gln711. The hydroxyl group of dimethylstilbestrol subsequently engages in a hydrogen bond with the amide oxygen atom of the side chain of Gln711. The same induced fit is observed if α-zearalenol is binding to the AR (middle panel) building a hydrogen bond between its hydroxyl group and the amide nitrogen atom of the side chain of Gln711. In addition subtle changes (1−2 Å rms deviation) in the position of the hydrophobic side chains accommodate the topology of the very ligand molecule.

Multidimensional QSAR. Our QSAR study was based on 119 ligands; 88 molecules thereof were assigned to the training set, and the remaining 31 used as test and prediction compounds. Induced fit was accounted for by the dual-shell concept of *Raptor*. To allow for topological and physicochemical variation at the true biological receptor with different ligands bound, the *Raptor* results were averaged over 10 individual models defining a surrogate family.

The model converged at a cross-validated $r^2 = 0.858$ for the training compounds and yielded a predictive $r^2 = 0.792$ for the test compounds. Experimental and calculated IC_{50} values are compared in Figure 6 and Table 2. On the average, the predicted binding affinity of the training ligands deviates by a factor 1.7 from the experiment; those of the test set deviate by a factor 1.6 in IC_{50}. The maximal observed deviation of an individual molecule are 7.8 and 13.9, respectively. All seven weak- and nonbinding molecules were predicted to bind with affinities close to or weaker than the experimental threshold value, indicating that our model is able to distinguish these molecules from strong binders. A *scramble test*[19] (predictive $r^2 = -1.444$) demonstrated the sensitivity of the model family toward the biological data.

Comparison of the *Raptor* model with the binding site at the true biological receptor (Figure 7) shows that the hydrophobic pocket (represented by Leu707, Gly708, Trp741, Met742, Met745, Phe764, Met787, Leu873, and Met895) and hydrogen-bonding moieties (H-bond acceptors mimicking Tyr877; H-bond donors mimicking Arg752 and Asn705) are well identified by the model.

Prediction of External Compound Classes. To predict the androgen-receptor-mediated harmful potential of environmental chemicals, the virtual model has to be able to predict the binding affinity of structurally diverse compounds. Thus, using the optimized receptor model, we have addressed the following question: Can the receptor model predict the binding affinity of compounds within a foreign data set, i.e., substances from different structure classes than those used to train the system? We have selected four polychlorinated biphenyls and 3,4-diphenyltetrahydrofuran (Figure 8),

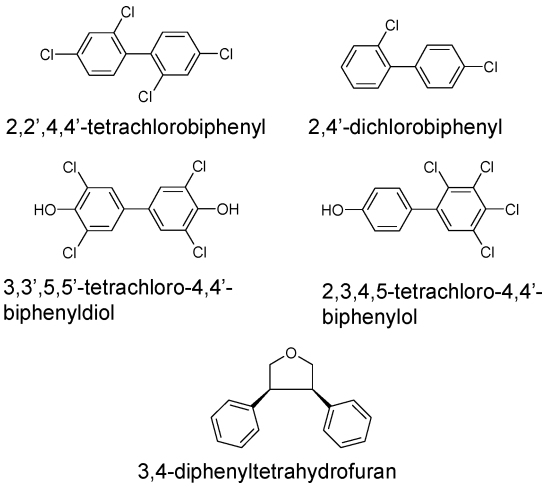


Figure 8. The four polychlorinated biphenyls and 3,4-diphenyltetrahydrofuran used to challenge the *Raptor* model for predicting affinities of structural compound classes different from the training and test set.

Table 3. Experimental and Calculated IC_{50} for the External Compound Classes Binding to the Androgen Receptor

Compound	pIC_{50} (exp)	pIC_{50} (pred)	error in pIC_{50}
2,3,4,5-Tetrachloro-4′-biphenylol	4.79	4.23	0.56
2,4′-Dichlorobiphenyl	4.79	4.31	0.48
2,2′,4,4′-Tetrachlorobiphenyl	4.77	5.50	0.74
3,3′,5,5′-Tetrachloro-4,4′-biphenyldiol	4.42	4.62	0.20
3,4-Diphenyltetrahydrofuran	4.54	4.21	0.33

compounds which are *not* represented in the training set (Figure 2). These two structure classes were randomly selected and separated from the 119 compounds prior to the QSAR simulation. The compounds were superimposed onto the ligand molecules of the training and test set using the same alignment protocol as described above. The binding affinities of these compounds were predicted using the *Raptor* model presented in the previous paragraph. The affinities lie well within 1 order of magnitude to the experimentally measured value (cf. Figure 6, right panel and Table 3).

Conclusions

To quantitatively predict relative free energies of ligand binding to the androgen receptor, it is of utmost importance to simulate induced fit. We have devised a new ligand-alignment protocol combining flexible docking, MD simulations, and LIE analysis. The resulting superposition of the ligand molecules of six diverse compound classes served as input for *Raptor*, a receptor-modeling concept based on multidimensional QSAR which allows for ligand-dependent induced fit. Protein flexibility was explicitly accounted for throughout our study. The resulting model allows the prediction of binding affinity of all compounds close to the experimental value. Most importantly, we could show that our concept, flexible docking combined with multidimensional QSAR, is able to predict molecules belonging to compound classes well outside those used for the training process.

The predicitivity of the model, demonstrated for the diverse test set as well as for two external compound classes, suggests that our approach is suited for the

prediction of endocrine-disrupting effects of environmental chemicals, both existing and hypothetical, a problem that has recently become a cause for concern among scientists, environmental advocates, and politicians alike. This has been underscored by the US legislation in 1995/6, by mandating that chemicals and formulations must be screened for potential endocrine-disrupting activity before they are manufactured or used in certain processes.[20]

Acknowledgment. This research was made possible through grants from the Swiss National Science Foundation (Grant No. 405040-104411) and the Jacques en Dolly Gazan Foundation, Zug, Switzerland.

References

(1) (a) Kavlock, R. J.; Daston, G. P.; DeRosa, C.; Fenner-Crisp, P.; Gray, L. E.; Kaattari, S.; Lucier, G.; Luster, M.; Mac, M. J.; Maczka, Z.; Miller, R.; Moore, J.; Rolland, R.; Scott, G.; Sheehan, D. M.; Sinks, T.; Tilson, H. A. Research needs for the risk assessment of health and environmental effects of endocrine disruptors: a report of the U. S. EPA-sponsored workshop. *Environ. Health Perspect.* **1996**, *104* (Suppl. 4), 715−740. (b) Colborn, T.; vom Saal, F. S.; Soto, A. M. Developmental effects of endocrine-disrupting chemicals in wildlife and humans (see comments). *Environ. Health Perspect.* **1993**, *101*, 378−384. (c) Carlsen, E.; Giwercman, A.; Keiding, N.; Skakkebaek, N. E. Evidence for decreasing quality of semen during past 50 years (see comments). *Br. Med. J.* **1992**, *305*, 609−613.

(2) Heinlein, C. A.; Chang, C. S. Androgen receptor in prostate cancer. *Endocr. Rev.* **2004**, *25*, 276−308.

(3) Neumann, F.; Jacobi, G. H. Antiandrogens in tumor therapy. *J. Clin. Oncol.* **1982**, *1*, 41−65.

(4) Sack, J. S.; Kish, K. F.; Wang, C.; Attar, R. M.; Kiefer, S. E.; An, Y.; Wu, G. Y.; Scheffler, J. E.; Salvati, M. E.; Krystek Jr., S. R.; Weinmann, R.; Einspahr, H. M. Crystallographic structure of the ligand-binding domains of the androgen receptor and its T877A mutant complexed with the natural agonist dihydrotestosterone. *Proc. Natl. Acad. Sci. U.S.A.* **2001**, *98*, 4904−4909.

(5) (a) Soderholm, A. A.; Lehtovuori, P. T.; Nyronen, T. H. Three-dimensional structure−activity relationships of nonsteroidal ligands in complex with androgen receptor ligand-binding domain. *J. Med. Chem.* **2005**, *48*, 917−925. (b) Bohl, C. E.; Chang, C.; Mohler, M. L.; Chen, J.; Miller, D. D.; Swaan, P. W.; Dalton, J. T. A ligand-based approach to identify quantitative structure−activity relationships for the androgen receptor. *J. Med. Chem.* **2004**, *47*, 3765−3776.

(6) Madauss, K. P.; Deng, S. J.; Austin, R. J. H.; Lambert, M. H.; McLay, I.; Pritchard, J.; Short, S. A.; Stewart, E. L.; Uings, I. J.; Williams, S. P. Progesterone receptor ligand binding pocket flexibility: Crystal structures of the norethindrone and Mometasone furoate complexes. *J. Med. Chem.* **2004**, *47*, 3381−3387.

(7) Fang, H.; Tong, W.; Branham, W. S.; Moland, C. L.; Dial, S. L.; Hong, H.; Xie, Q.; Perkins, R.; Owens, W.; Sheehan, D. M. Study of 202 natural, synthetic, and environmental chemicals for binding to the androgen receptor. *Chem. Res. Toxicol.* **2003**, *16*, 1338−1358.

(8) Mohamadi, F.; Richards, N. G. J.; Guida, W. C.; Liskamp, R.; Lipton, M.; Caufield, C.; Chang, G.; Hendrickson, T.; Still, W. C. MacroModel −An integrated software system for modeling organic and bioorganic molecules using molecular mechanics. *J. Comput. Chem.* **1990**, *11*, 440−467.

(9) Weiner, S. J.; Kollmann, P. A.; Case, D. A.; Singh, U. C.; Ghio, C.; Alagona, G.; Profeta Jr.; S.; Weiner, P. A new force field for molecular-mechanical simulation of nucleic acids and proteins. *J. Am. Chem. Soc.* **1984**, *106*, 765−784.

(10) Stewart, J. J. P. MOPAC − A semi-empirical molecular orbital program. *J. Comput.-Aid. Mol. Des.* **1990**, *4*, 1−105.

(11) (a) Vedani, A.; Huhta, D. W. A new force field for modeling metalloproteins. *J. Am. Chem. Soc.* **1990**, *112*, 4759−4767; b) http://www.biograf.ch/PDFS/Yeti.pdf.

(12) Reddy, M. R.; Erion, M. D., Eds. *Free energy calculations in rational drug design*; Kluwer Academic Publishers: Boston, 2001.

(13) Hansson, T.; Marelius, J.; Aqvist, J. Ligand binding affinity prediction by linear interaction energy methods. *J. Comput.-Aid. Mol. Des.* **1998**, *12*, 27−35.

(14) Wang, W.; Wang, J.; Kollman, P. A. What determines the van der Waals coefficient beta in the LIE (linear interaction energy) method to estimate binding free energies using molecular dynamics simulations? *Prot. Struct. Funct. Gen.* **1999**, *34*, 395−402.

(15) One simulation for each ligand molecule lasts 25 h on a Macintosh G5 1.8 GHz workstation.

(16) Case, D. A.; Pearlman, D. A.; Caldwell, J. W.; Cheatham III, T. E.; Wang, J.; Ross, W. S.; Simmerling, C. L.; Darden, T. A.; Merz, K. M.; Stanton, R. V.; Cheng, A. L.; Vincent, J. J.; Crowley, M.; Tsui, V.; Gohlke, H.; Radmer, R. J.; Duan, Y.; Pitera, J.; Massova, I.; Seibel, G. L.; Singh, U. C.; Weiner, P. K.; Kollman, P. A. 2002, AMBER 7, University of California, San Francisco.

(17) Araz Jakalian, Bruce L. Bush, David B. Jack, Christopher I. Bayly Fast, efficient generation of high-quality atomic charges. AM1-BCC model: I. Method. *J. Comput. Chem.* **2000**, *21*, 132−146.

(18) (a) Lill, M. A.; Vedani, A.; Dobler, M. Raptor −combining dual-shell representation, induced-fit simulation and hydrophobicity scoring in receptor modeling: Application towards the simulation of structurally diverse ligand sets. *J. Med. Chem.* **2004** *47*, 6174−6186. (b) http://www.biograf.ch/PDFS/Raptor.pdf.

(19) (a) Vedani, A.; Dobler, M. 5D-QSAR: The key for simulating induced fit? *J. Med. Chem.* **2002**, *45*, 2139−2149. (b) http://www.biograf.ch/PDFS/Quasar.pdf.

(20) (a) US Government: *Safe Drinking Water Act Amendment*. Public Law 104−182 (Section 136), 1996, http://www.epa.gov.safewater/sdwa/index.html. (b) US Government: *Food Quality Protection Act.* Public Law 104−170 (Section 408), 1996, http://www.fda.gov/opacom/laws/foodqual/fqpatoc.htm.

JM050403F