CHIMIA **2006**, *60*, No. 1/2

Chimia 60 (2006) 33–36

ISSN 0009–4293

Computational Pharmaceutical Chemistry – Novel Technologies for Lead Optimization and the Prediction of ADMET Properties

Markus A. Lill*

Abstract: The prediction of affinities of ligands binding to a target protein represents a major challenge in modern computer-aided drug design. To contribute towards this goal, we have developed a new technology to identify feasible binding modes of protein-bound, biomedically interesting molecules and to compute their binding affinity using multidimensional quantitative structure-activity relationships (QSAR). In our approach, the flexibility of the protein is explicitly simulated. Applications of the underlying technology to G protein-coupled receptors, nuclear receptors and cytochrome P450 show the ability of this approach to predict the binding affinity of diverse sets of ligands to a common protein, and suggest its potential to predict adverse or toxic effects of drugs and chemicals *in silico.*

Keywords: Computer-aided drug discovery · Flexible docking · Multidimensional QSAR · Prediction of binding affinities · Simulation of induced fit

Introduction

In the last two decades, computer-aided drug discovery (CADD) concepts have matured into powerful tools for identifying and optimizing lead structures. Based on the three-dimensional structure of target proteins, structure-based design has become a widespread approach to identify potential drug candidates *in silico*. While CADD techniques have been widely used to attain a qualitative understanding of ligand binding to proteins, the current challenge is to quantify their interaction. To compute the binding affinity of a given compound, an accurate prediction of relative free energies of binding, *e.g.* by using free energy perturbation calculations, would seem to be the method of choice. Unfortunately, the associated procedures are computationally demanding and limited to the comparison of affinities of structurally rather similar compounds. On the other hand, automated docking, as used for virtual screening, is up to now limited in its potential to reliably predict binding affinities. One major drawback is the underlying rigid-receptor assumption: the protein conformation remains unaltered during the docking process and cannot adapt its shape and properties to the ligand binding to it.

X-ray crystallography has clearly demonstrated the importance of induced fit upon ligand binding. Fig. 1, for example, shows the structures of 17α-estradiol [1] and Raloxifene [2] bound to the estrogen receptor [3]. If 17α-estradiol is bound, the ligand forms hydrogen bonds *via* both its hydroxyl groups with Glu 353/Arg 394 and His 524. The residual aromatic/aliphatic portion is accommodated by an extended hydrophobic pocket. When the estrogen receptor is complexed with Raloxifene, the estrogen receptor opens a small channel near the center of the binding pocket [2] by translocating Leu 540 approximately 10 Å. Within this channel, the alkylaminoethoxy side chain of Raloxifene is snugly accommodated. Additionally, Asp 351 is rotated towards the protonated piperidyl N-atom of Raloxifene, forming a salt bridge. As a result, the hydrophobic field and the hydrogen bond propensity created by the binding site are altered.

Recently, our group has developed a two-step procedure and the underlying technology to identify bioactive binding modes of ligands and to quantify their interaction with the target protein using multidimensional QSAR – explicitly taking protein flexibility into account. Application to G protein-coupled receptors, nuclear receptors and cytochrome P450 enzymes demonstrates the capability of this approach for predicting the binding affinities of structurally diverse sets of compounds, and its potential for quantifying receptor-mediated toxicity and drug–drug interactions.

Materials and Methods

Identification of Binding Modes

If the experimental structure of the target protein is available, automated docking is used to identify energetically favorable binding conformations for each ligand (software *Yeti*) [4]. During the docking process protein flexibility is accounted for by allowing the side chains of the amino acids to adapt to the compound binding to it. A Monte-Carlo protocol combines a global search for possible binding modes with

*Correspondence: Dr. M.A. Lill
Institute of Molecular Pharmacy
Department of Pharmaceutical Sciences
Pharmacenter
University of Basel
Klingelbergstrasse 50
CH-4056 Basel
and
Biographics Laboratory 3R
Friedensgasse 35
CH-4056 Basel
Tel.: +41 61 261 4259
Fax: + 41 61 261 4258
E-mail: markus.lill@unibas.ch and markus@biograf.ch

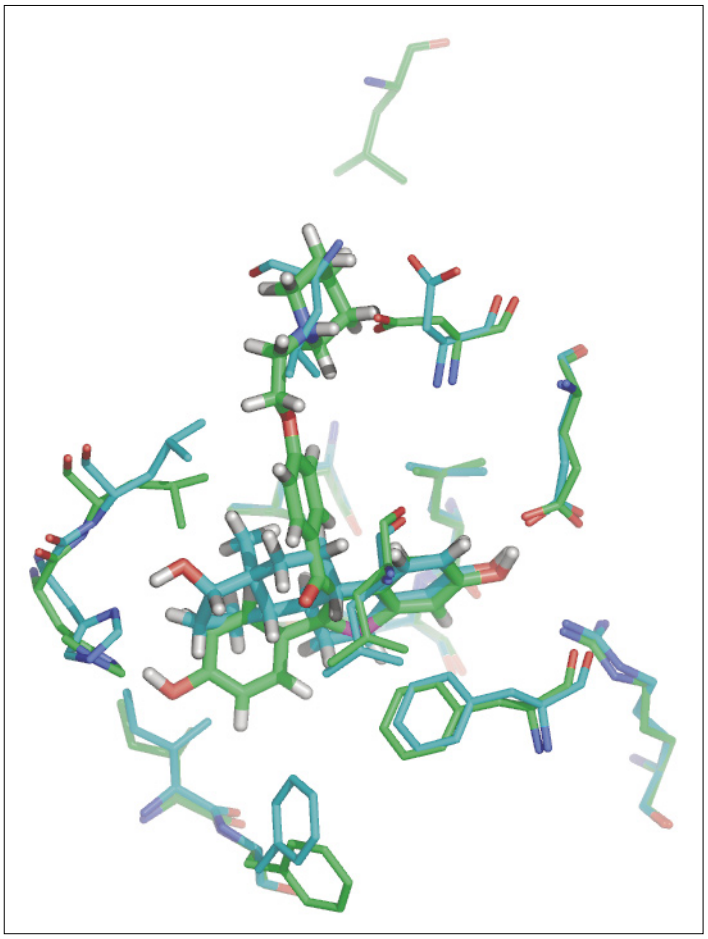

Fig. 1. Small molecule-protein structures with different ligands solved by means of X-ray crystallography. Superposition of 17α-estradiol (carbon atoms = cyan) and Raloxifene (carbon atoms = green) bound to the human estrogen receptor.

energy minimization to refine the docked protein–ligand structure, thus including ligand and protein flexibility. Additionally, the procedure allows for solvation of the ligand–protein complex throughout the Monte-Carlo search.

If no experimental data on the 3D structure of the protein are available, the identification of the conformation of the ligands and their relative orientation is predicted by maximizing the similarity of the physicochemical properties of the compounds in 3D space using the *Symposar* concept [5]. Both approaches identify an ensemble of possible ligand configurations. This ensemble of binding conformations results in a ligand alignment, represented as a 4D data set, and is used as input for the multidimensional QSAR technologies, *Quasar* [6] and *Raptor* [7], developed in our group, to quantify the binding energy.

Quantifying Ligand–Protein Interactions

QSAR is an area of computational research that builds mathematical, atomistic or virtual models to predict quantities such as the binding affinity, the toxic potential, and pharmacokinetic parameters of ligands. The idea behind QSAR is that structural features can be correlated with biological activity. Of particular interest for the biomedical research are QSAR based on three-dimensional models (3D-QSAR). The latter generate a rational model of the target protein, and allow for the quantification of the mutual interactions, including electrostatic forces, hydrogen bonds and hydrophobic contacts – the forces known to play a key role for both substrate recognition and specificity. In contrast to the real biological receptor, where the binding site is characterized by a 3D arrangement of amino acids, 3D-QSAR models typically represent this binding site by mapping physico-chemical properties onto a surface or a grid surrounding the ligand molecules, which are themselves superimposed in 3D space according to a pharmacophore hypothesis.

The receptor-modeling concepts, *Quasar* [6] and *Raptor* [7], use shells, mapped with physico-chemical properties onto them, to represent the surface of the binding site (Fig. 2). In *Quasar* these properties represent electrostatic, van der Waals and hydrogen-bonding particles. The binding affinity is calculated using a directional force field [8] to quantify the interaction between the ligand and receptor model, as well as including corrections for the ligand's desolvation, internal strain, entropy and the cost of induced protein fit:

$$\Delta G_{binding} = \Delta E_{force\ field} + \Delta E_{polarization} - \Delta G_{ligand\ desolvation} - T\Delta S - \Delta E_{internal\ strain} - \Delta G_{induced\ fit} \quad (1)$$

In *Raptor* an alternative scoring function was developed including specific terms for hydrophobic and hydrogen-bonding interactions between ligand and protein, thus, accounting for ligand desolvation implicitly:

$$\Delta G_{binding} = \Delta G_{constant} + \Delta G_{H\text{-}bonding} + \Delta G_{hydrophobic} - T\Delta S - \Delta G_{induced\ fit} \quad (2)$$

Both approaches allow the possibility of representing each ligand molecule as an ensemble of conformations, orientations, stereoisomers and protonation states (4D-QSAR), thereby reducing the bias in identifying the bioactive conformer. In addition, they explicitly simulate induced protein fit. While *Quasar* allows a topological accommodation of the binding site surface onto each individual ligand, *Raptor* anisotropically simulates induced fit by a dual-shell representation of the three-dimensional binding-site model. 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 ligand and a component due to the attraction or repulsion between ligand and receptor surrogate. The latter component is obtained by correlating their physico-chemical properties (hydrophobicity and hydrogen-bond propensity) in 3D space. *Raptor* also allows 'threshold' compounds for the modeling procedure to be defined, *i.e.* compounds which bind weaker than the resolution limit of the assay. Both technologies may be used independently, however, they exert their full predictive power when combined, aiming to reach consensus for the same ligand data.

Results

In the past, these technologies were applied to several systems of biomedical interest, ranging from lead optimization for protein kinases and G protein-coupled receptors to predicting adverse effects of drugs (Cytochrome P450 3A4) and environmental chemicals (estrogen, androgen, thyroid and aryl hydrocarbon receptor).

Lead Optimization for GPCRs

The methods have been used for lead-optimization purposes for ligands binding

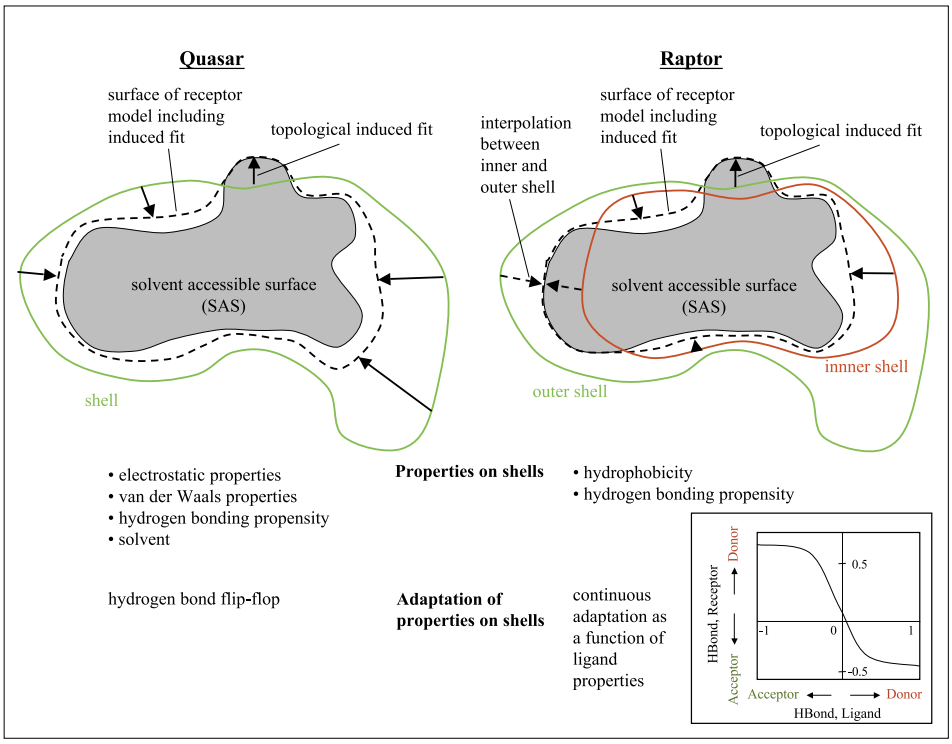

Fig. 2. Differences between the receptor modeling concepts *Quasar* (left) and *Raptor* (right). Top panel: sketch of a ligand (represented as its SAS; gray surface) in the receptor model represented as a single shell (green) in *Quasar* and as two shells (green and red) in *Raptor*. During the steric adaptation process the receptor model adapts (solid arrows) its topology to the shape of the ligand (dotted line). In *Raptor*, the fields generated by the protein binding site onto the ligand's SAS are computed by linear interpolation between the inner and the outer shell, if the ligand's SAS lies between those two shells (dashed arrows). Middle panel: The different properties mapped onto the shells representing the binding site surface of the receptor model. Bottom panel: In *Quasar* the character of the hydrogen-bond properties on the shell can flip depending on the ligand binding. In *Raptor* a continuous adaptation of hydrophobicity and hydrogen bond propensity to the ligand properties is allowed.

to the Neurokinin-1 [9] and CCR-3 [10] receptors. More recently, we studied the Bradykinin B_2 receptor using *Symposar* for automated alignment, and both *Quasar* and *Raptor* for quantifying their binding affinity in a consensus scoring manner. The resulting models are also used in the department's laboratory section [11].

Adverse Drug–Drug Interactions

Inhibition of Cytochrome P450 3A4 (CYP3A4) by small molecules represents a major mechanism associated with undesired drug–drug interactions responsible for a substantial number of late-stage failures in the pharmaceutical drug-development process. For a quantitative prediction of associated pharmacokinetic parameters, we developed a computational model, allowing us to predict the inhibitory potential of 48 structurally diverse molecules [12]. Based on the experimental structure of CYP3A4, we first sampled possible binding modes using automated docking (software *Yeti*) which includes protein flexibility and dynamic solvation throughout the docking process. The results are consistent with both X-ray and metabolism data [12]. Next, an ensemble of energetically favorable orientations was composed into a 4D dataset, to be used as input for a multidimensional QSAR technique (software *Raptor*). The QSAR modeling reached a cross-validated r^2 of 0.825 and a predictive r^2 of 0.659 (Fig. 3). On average, the predicted binding affinity of the training ligands deviates by a factor of 2.7 from the experimental value of K_i. Those of the test set deviate by a factor of 3.8.

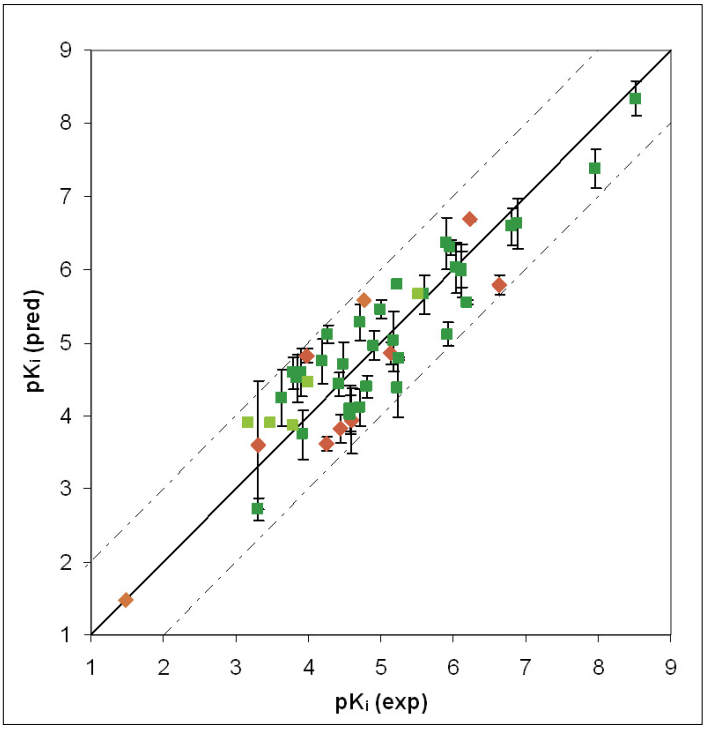

Fig. 3: Predicted *versus* experimental binding affinities for ligand molecules binding to CYP3A4. Data for the training set are shown in dark green, for the test set in red. The corresponding data for the threshold compounds are shown in light green for the training set and in orange for the test set.

Endocrine Disruption

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. Unbalanced production or cell insensitivity to specific hormones may result in diseases associated with human endocrine dysfunction [13].

Many environmental chemicals may bind to a nuclear receptor influencing the balance of the endocrine system. 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, such as promotion of hormone-dependent cancers, reproductive

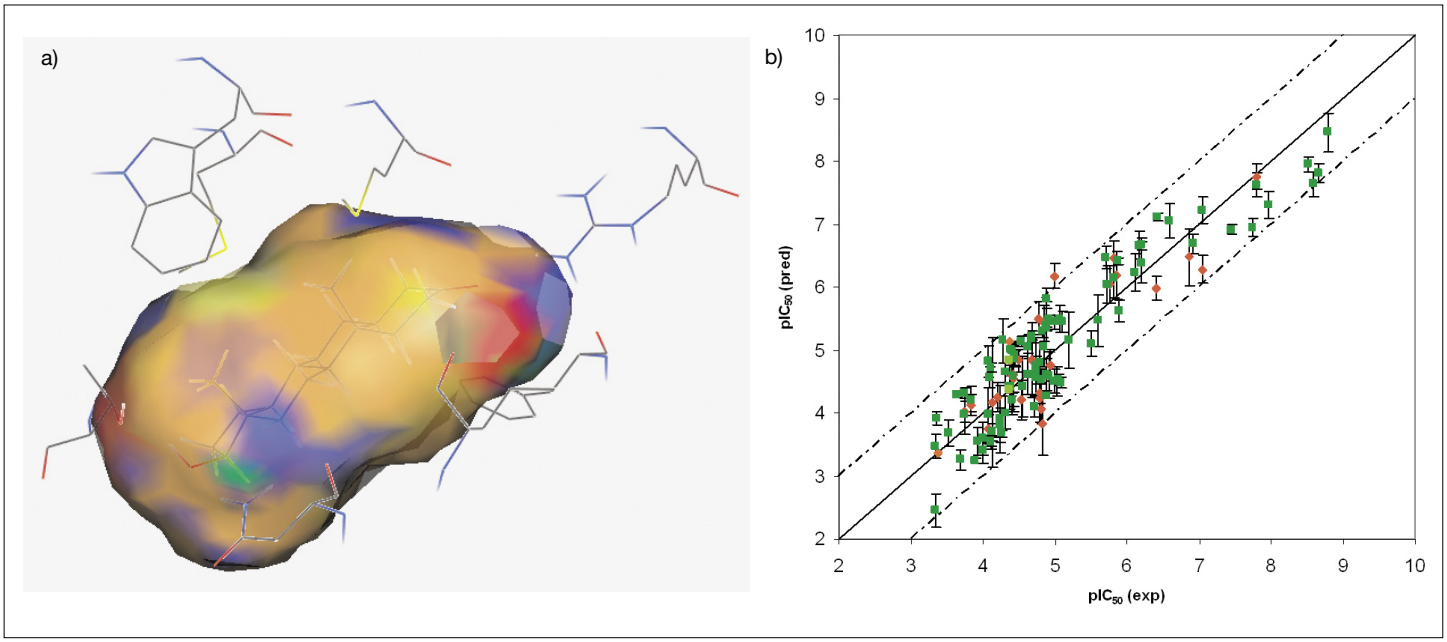

Fig. 4. a) Binding site model of the androgen receptor as generated by the *Raptor* technology (beige = hydrophobic fields; blue = hydrogen-bond donating propensity; red = hydrogen-bond accepting propensity; green = hydrogen-bond flip-flop character). Only the inner shell of the *Raptor* model is shown with bound DHT. b) Comparison of predicted and experimental binding affinities for the 119 molecules used in our study (Color scheme: see Fig. 2).

tract disorders, and a reduction in reproductive fitness.

A variety of compounds in the environment have been shown to display agonistic or antagonistic activity towards the androgen receptor, including both natural products and synthetic compounds. The concern over xenobiotics binding to the androgen receptor has created the need to both screen and monitor compounds expected to modulate endocrine effects. We therefore developed an *in silico* model, based on 119 molecules representing six compound classes, in order to quantitatively predict the potential of structurally diverse ligands for binding to the androgen receptor [14]. To identify the binding mode to the real biological receptor, a stepwise protocol consisting of flexible docking, molecular dynamics (MD) simulations and linear interaction-energy analysis (LIE) was developed [14]. The superposition of the ligand molecules emerging from the combined protocol served as input for *Raptor*. The model converged at a cross-validated r^2 of 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 values (Fig. 4). We then challenged the model by testing five molecules outside the 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.

Conclusions

To quantify the affinity of ligand–protein interactions, we developed a suite of novel technologies to identify probable binding modes and quantify their interaction with the target protein. A focus of our research is to study and model ligand-induced protein fit. These technologies would seem to be of huge interest for lead-optimization purposes, but are also relevant to quantify receptor-mediated toxicity or adverse drug-drug interactions in the context of predictive ADMET. The applications on CYP3A4 and on the androgen receptor show that these *in silico* methods represent promising approaches to increase the ability to predict and model relevant pharmacokinetic and toxicity endpoints, thereby accelerating the drug discovery process.

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.

Received: December 12, 2005

[1] pdb-code: 1ERE
[2] pdb-code: 1ERR
[3] A.M. Brzozowski, A.C.W. Pike, Z. Dauter, R.E. Hubbard, T. Bonn, O. Engstrom, L. Ohman, G.L. Greene, J.-A. Gustaffson, M. Carlquist, *Nature* **1997**, *389*, 753.
[4] *http://www.biograf.ch/PDFS/Yeti.pdf*.
[5] M.A. Lill, M. Dobler, A. Vedani, in preparation.
[6] a) A. Vedani, M. Dobler, M. Lill, *J. Med. Chem.* **2005**, *48*, 3700; b) A. Vedani, M. Dobler, *J. Med. Chem.* **2002**, *45*, 2139; c) *http://www.biograf.ch/PDFS/Quasar.pdf*.
[7] a) M.A. Lill, A. Vedani, M. Dobler, *J. Med. Chem.* **2004**, *47*, 6174; b) *http://www.biograf.ch/PDFS/Raptor.pdf*.
[8] A. Vedani, D. W. Huhta, *J. Am. Chem. Soc.* **1990**, *112*, 4759.
[9] A. Vedani, H. Briem, M. Dobler, H. Dollinger, D.R. McMasters, *J. Med. Chem.* **2000**, *43*, 4416.
[10] A. Vedani, H. Dollinger, K.-M. Hasselbach, M. Dobler, M.A. Lill, *J. Med. Chem.* **2005**, *48*, 1515.
[11] A. Vedani, O. Schwardt, S. Rabbani, *Chimia* **2006**, *60*, 2006, *1*, 73–81.
[12] M.A. Lill, M. Dobler, A. Vedani, *ChemMedChem*, **2006**, *1*, 73–81.
[13] a) R.J. Kavlock, G.P. Daston, C. DeRosa, P. Fenner-Crisp, L.E. Gray, S. Kaattari, G. Lucier, M. Luster, M.J. Mac, Z. Maczka, R. Miller, J. Moore, R. Rolland, G. Scott, D.M. Sheehan, T. Sinks, H.A. Tilson, *Environ. Health Perspect.* **1996**, *104*(Suppl. 4), 715; b) T. Colborn, F.S. vom Saal, A.M. Soto, *Environ. Health Perspect.* **1993**, *101*, 378; c) E. Carlsen, A. Giwercman, N. Keiding, N.E. Skakkebaek, *Br. Med. J.* **1992**, *305*, 609.
[14] M.A. Lill, F. Winiger, A. Vedani, B. Ernst, *J. Med. Chem.* **2005**, *48*, 5666.