



QSAR models in receptor-mediated effects: the nuclear receptor superfamily

Hong Fang^a, Weida Tong^{a,*}, William J. Welsh^b, Daniel M. Sheehan^{c,1}

^a*Logicon ROW Sciences, 3900 NCTR Road, MC 910, Jefferson, AR 72079, USA*

^b*Department of Pharmacology, Robert Wood Johnson Medical School, University of Medicine & Dentistry of New Jersey, 675 Hoes Lane, Piscataway, NJ 08854, USA*

^c*Division of Genetic and Reproductive Toxicology, National Center for Toxicological Research (NCTR), Jefferson, AR 72079, USA*

Abstract

The nuclear receptor (NR) superfamily is ligand-dependent transcriptional factors that mediate gene expression in humans and wildlife. These receptor-mediated effects are stimulated and/or inhibited by endogenous cognate ligands for each NR but also by exogenous substances including natural products and synthetic chemicals. The NRs and their ligands have thus attracted broad scientific interest, particularly in the pharmaceutical industry for drug discovery and in toxicology and environmental science for risk assessment as, for example, pertaining to endocrine disrupting chemicals. Besides advancing our fundamental knowledge of NR biology, these scientific efforts are generating relevant biological data on NR ligands particularly with respect to their binding affinities, receptor specificities, and agonist versus antagonist activities. These data from diverse sources serve as input for construction of quantitative structure–activity relationship (QSAR) models and related approaches that employ statistical regression techniques to correlate variations between the biological activities of NR ligands and their calculated structural and physicochemical properties. In this review, we attempt to summarize the substantial body of work in the published literature related to QSAR models for NR ligands, with special emphasis on different computational approaches and specific applications. Special attention is placed on the estrogen receptor, for which the greatest amount of relevant information is known at present. We also describe efforts to create ‘benchmark’ sets of high-quality biological data on NR ligands that may serve as resources for building statistically robust and predictive QSAR models.

Published by Elsevier Science B.V.

Keywords: Receptor-mediated effects; Nuclear receptors; Quantitative structure–activity relationships; Estrogen receptor; Endocrine disrupting chemicals

1. Introduction

Nuclear receptors (NRs) are a superfamily of ligand-dependent transcription factors that mediate the effects of hormones and other endogenous ligands to regulate the expression of specific genes. Members of the NR superfamily, which may number

* Corresponding author. Tel.: +1-870-543-7142; fax: +1-870-543-7382.

E-mail address: wtong@nctr.fda.gov (W. Tong).

¹ Present address: Daniel M. Sheehan and Associates, 1422 Scott Street, Little Rock, AR 72202, USA.

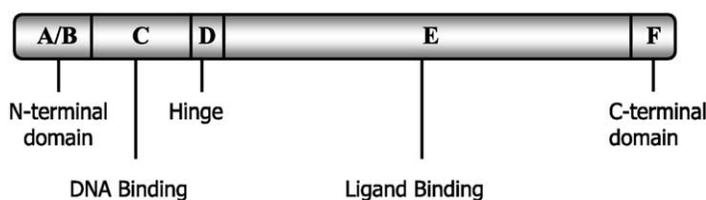


Fig. 1. Functional domains of NRs.

in the hundreds, include receptors for various steroid hormones (estrogen, androgen, progesterone, and several corticosteroids), retinoic acid (the retinoic acid receptor α , β , and γ isoforms, and the retinoid X receptor α , β , and γ isoforms), thyroid hormones, vitamin D, and dietary lipids (the peroxisome proliferator activated receptor (PPAR) α , β , and γ isoforms). A large number of 'orphan' NRs have also been identified whose cognate ligands are still unknown [1]. Diminished or excessive production of a particular hormone or target-cell insensitivity to a hormone is among the major problems related to human endocrine dysfunction diseases [2].

The NRs contain five functional domains from the N to C termini, designated A/B, C, D, E, and F (Fig. 1). While the C domain containing the DNA-binding domain is the region of highest sequence conservation in the superfamily, there is considerable variability across receptors in the A/B and D domain [3]. The E domain, known as the ligand-binding domain (LBD), is the sequence specifying ligand binding. Despite the low sequence homology (as low as 20%) between the LBDs of different NRs, analysis of crystal structure data [4–14] has revealed a remarkable similarity in their three-dimensional (3D) structures. By virtue of its significance for ligand specificity and receptor-mediated effects among the NRs, the LBD has attracted the greatest amount of scientific curiosity and attention.

Adopting the generally accepted depiction of the ligand–receptor interaction in NRs, the LBD can be envisaged as a molecular switch. Depending on whether the bound ligand is an agonist or antagonist, the C-terminal helix 12 (H12) is found in either one of two orientations. In its agonist-bound conformation, H12 serves as a 'lid' to close the ligand-binding pocket [15]. Thus, the activation function-2 (AF-2, located at the C-terminal of the LBD) [16] is brought into position to recruit co-activators and/or

co-repressors which bind to facilitate transcriptional initiation [17]. In its antagonist-bound conformation, however, H12 is positioned in a different orientation such that the transcriptional machinery remains deactivated. The preservation of the molecular switch, despite the low sequence homology in the LBD across NR members, argues that agonist/antagonist discrimination is critically important for receptor functioning.

There are a large number of ligands, diverse in both structure and source, which act through the NRs to produce receptor-mediated effects. Steroid hormones play a vital role in a wide variety of essential physiological processes including cell growth, sexual development, maintenance of salt balance, and sugar metabolism [18]. Estrogens elicit many cellular responses in target tissues and can exert both positive and negative effects on health and reproductive function. For example, estrogens are used beneficially for fertility control (oral contraception) and for relief of menopausal symptoms (estrogen replacement therapy). The adverse developmental effects of diethylstilbestrol (DES) demonstrate human fetal sensitivity to estrogenic chemicals. Progesterone is a hormone that functions to help regulate the menstrual cycle and plays a significant role in pregnancy. Progesterone has also been reported to both stimulate and inhibit the growth of experimental mammary tumors, dependent upon the dose and experimental model [19]. Retinoids have received considerable attention from their use in the therapeutic treatment of different diseases (e.g. cancer) and from their critical role in embryonic and fetal development [20,21].

It has also been recognized that numerous synthetic chemicals are capable of interfering with the normal signaling pathway by interacting with NRs [22]. Environmental chemicals, such as pesticides, herbicides, and plasticizers are capable of activating the PPAR leading to liver dysfunction and hepatoma. Numerous findings also show that certain

environmental chemicals have the ability to disrupt the endocrine system by mimicking the functions of natural hormones. These endocrine disrupting chemicals (EDCs) may exert adverse effects on humans and wildlife [23].

Quantitative structure–activity relationship (QSAR) models have proven their utility, from both the pharmaceutical and toxicological perspectives, for identification of chemicals that might interact with NRs. While their primary function in the pharmaceutical enterprise is lead discovery and optimization, QSAR models have played an essential role in toxicology as a priority setting tool for risk assessment. For example, public health concern about EDCs resulted in Federal legislation mandating the environmental protection agency (EPA) to regulate potential EDCs in drinking water and food additives [24]. Under this requirement, up to 87,000 existing chemicals will be experimentally evaluated for their potential to disrupt activities in the estrogen, androgen, and thyroid hormone systems [25]. In an attempt to reduce the time and expense in this prodigious task of screening and testing such a large number of chemicals, QSAR models are being developed to prioritize chemicals as to their endocrine disrupting potential for further experimental evaluation [26].

QSAR models offer numerous additional benefits beyond prediction [27], such as: (1) leveraging existing structure–activity data; (2) providing insights into mechanisms of action (e.g. agonist versus antagonist) or identifying alternative mechanisms (e.g. metabolism); (3) identifying key structural features associated with high/low activity; (4) suggesting new design strategies and synthetic targets; (5) narrowing the dose range for a planned assay; (6) assisting in generation of new hypotheses to guide further research; (7) revealing chemicals that deviate from the QSAR model and, therefore, from the presumed biological model.

Basically, QSAR models employ quantitative regression methods to correlate, and rationalize, variations in the biological activity of a structurally related series of chemicals with variations in their molecular structures as encoded in pre-selected quantities commonly known as molecular descriptors. The fundamental assumption inherent in every QSAR model is that a chemical's physical and chemical

properties and its biological activities are predicated by its structure [28].

The number of QSAR models derived from NR–ligand biological data is already fairly sizable. The list is surely to expand in coming years as our knowledge and appreciation of the critical role of NRs in biological function continue to grow. Reflecting the availability of relevant biological data, the largest number of QSAR models in the published literature is associated with estrogen receptor (ER) binding. A few papers have reported QSAR models for androgens, progesterone, retinoic acids and thyroid hormone. In this review, we will first provide an overview of various QSAR approaches applied for the NR superfamily. We will then proceed to survey the QSAR models available for ER with emphasis on two separate but related applications: structural characterization and risk assessment. Section 4 will summarize QSAR models for other members of the NRs, such as the androgen receptor (AR) and the progesterone receptor (PR). Section 5 will discuss ongoing efforts to assemble high-quality ‘benchmark’ sets of biological data associated with ligands for ER and other NRs that can be used to develop robust and predictive QSAR models for applications in risk assessment. These efforts may also provide a rich source of data for use as standards in the design, testing, and comparison of different QSAR models and approaches.

2. QSAR models associated with nuclear receptors

The NR superfamily has been the subject of various QSAR models and modeling approaches, of which comparative molecular field analysis (CoMFA) is predominant [29–34]. To construct a CoMFA model [35], a collection of chemicals with known activities (i.e. the training set) are first aligned together usually employing structural similarity as the basis for alignment. The aligned molecules are then embedded in a 3D grid, after which the steric and electrostatic fields are computed for each chemical at every grid point surrounding the molecules. The variations in these steric/electrostatic fields are then correlated with biological activity using partial least-squares (PLS) regression.

A major objective in applying CoMFA is to identify key structural and/or steric–electrostatic

characteristics shared by all or most of the ligands under study. An attractive feature of CoMFA is its capability to render a color-coded 3D contour map depicting regions in space around molecules where differences in steric and electrostatic fields are most strongly correlated with differences in activity. These 3D contour maps provide visual clues for modifying ligand structure so to either enhance or diminish activity. (Whereas enhanced activity of a lead chemical is sought in drug discovery, diminished activity of a toxicant is desired in toxicology.) One can also use these maps to infer critical features of the receptor-binding pocket based on arguments of ligand–receptor complementarity. This information is essential when the crystal structure of the receptor is unknown or not available. Even when the receptor crystal structure is available, CoMFA provides an additional source of information about the receptor from the perspective of the ligands. It should be noted that the crystal structure of a ligand–receptor complex provides but a single, albeit low-energy, snapshot of the actual dynamic biological system. Structural information derived from this crystal structure pertains only to the bound ligand in the strictest sense and certainly not to the wide structural diversity of ligands that bind to certain NRs, such as ER [36] and the orphan pregnane xenobiotic receptor (PXR) [37,38].

Besides CoMFA and related 3D-QSAR approaches, other QSAR approaches have been applied to the NRs and their ligands. So-called classical QSAR models attempt to correlate the biological activity of a series of ligands with their associated physicochemical properties or features that are usually calculated but sometimes derived from experimental measurements. A series of Hansch-type QSAR models have been developed separately for several chemical classes that bind to the ER [39,40]. These models considered only a few descriptors and found that indicators for the presence or absence of oxygen atoms, molar refraction (MR) and the Hammett σ parameter correlated significantly with activity. In fact, a comprehensive list of descriptors available to build classical QSAR models would number in the hundreds and perhaps thousands. Commercially available molecular modeling programs often include statistical tools to help in choosing which descriptors best encode for

structure–activity variation. For example, a genetic function approximation (GFA) approach developed by Rogers and Hopfinger [41] and implemented in Cerius² (<http://www.msi.com>) is a popular genetic algorithm-based statistical approach that is now widely used in QSAR model development [42].

Hologram QSAR (HQSAR), a novel fragment-based QSAR approach, was recently introduced by Tripos, Inc. (<http://www.tripos.com>). In HQSAR, each molecule in the dataset is divided into structural fragments that are then counted in bins of a fixed length array to form a molecular hologram. HQSAR has several attributes, including speed, reproducibility, and ease of use, that suggest its potential utility for prioritizing large numbers of chemicals for subsequent testing [34]. Tong et al. compared the performance of HQSAR with CoMFA for several ER datasets [34,43]. Although HQSAR and CoMFA achieved comparable results for two smaller datasets, CoMFA produced better results for a diverse dataset as judged by both internal and external validation.

Zheng and Tropsha [44] reported an automated variable selection QSAR method that is based on the k-Nearest Neighbor (kNN) principle. In this kNN-QSAR method, a chemical's activity is estimated as the mean activity value of its k nearest neighbors based on Euclidean distance in a multidimensional descriptor coordinate system. The method was tested on 58 ER ligands and demonstrated its effectiveness and generality.

Pharmacophore-based screening has become a common tool in the field of computer-aided drug design. A new method of rapid pharmacophore fingerprinting has been developed [45]; these fingerprints are used as descriptors to construct a QSAR model using PLS regression. Examples are given using the datasets reported by Kuiper et al. for both ER- α and ER- β [46]. The results are compared with previously published QSAR models for the same data to demonstrate the superiority of a full 3D, conformationally flexible approach. The QSAR model can be readily interpreted in structural/chemical terms.

Information extracted from molecular spectra has recently been applied to develop QSAR models associated with NR ligands [47]. The binding affinities of 45 progestagens (progesterone mimics) have been quantitatively modeled using the comparative spectra analysis (CoSA) approach, in which

experimental $^1\text{H-NMR}$, mass, simulated $^{13}\text{C NMR}$ and IR spectra were used separately or in combination to predict PR binding. The results are comparable with CoMFA, strongly supporting the use of spectroscopic fields in QSAR studies [48,49].

QSAR methods that incorporate information on ligand–receptor interactions have been investigated by a number of groups. The receptor coordinates are required either from crystal structure data or from homology modeling analysis, as is energy minimization of the bound ligand–receptor complex. Methods, such as VALIDATE [50], COMBINE [51] and free energy perturbation (FEP) [52] are the examples to use ligand–receptor interaction for QSAR models. As demonstrated in a recent study by Jayatilleke et al. [53], ligand-based and receptor-based approaches are highly compatible and yield more highly predictive QSAR models when employed in tandem. The utmost goal, whether in drug discovery or in computational toxicology, is to develop and apply QSAR models that provide the highest statistical quality and predictive ability. This can be achieved by using the full extent of information provided for the biological system under study. For some NRs, this information may include knowledge of the crystal structure of the receptor and/or receptor–ligand binary complex. Recently, Oostenbrink et al. [54] reported a single-step perturbation method allowing the calculation in a single simulation of relative free energy for a large number of polyaromatic hydrocarbons (PAHs) binding to the ER- α subtype. Agreement between the calculated and experimental results had a maximum deviation of only 3.3 kJ/mol. Moreover, this method is between four and six times less computer-time intensive as the thermodynamic integration method.

Predicting the receptor-binding affinity of biomolecules is one of the major challenges in computational approaches to drug design. Basically, two strategies are used: indirect ligand-based approaches (e.g. CoMFA, CoMSIA, CoMMA) and the direct receptor-based approach. Sippl [55,56] combined these two approaches and tested them on a set of ER ligands. The binding conformation was determined using an automated docking program [57], which was further verified through comparison with the crystal structures. The ligand alignments obtained from the docking simulations

were subsequently taken as the basis for a comparative field analysis using the GRID/GOLPE program [58]. The model constructed on the basis of the receptor structure supplies a better explanation of the binding activity.

3. QSAR models for estrogen receptor

Estrogens are widely prescribed in menopausal women for hormone replacement therapy to maintain bone mineral density and preserve cardiovascular health. Anti-estrogens, such as raloxifene [59,60] and tamoxifen, are being studied as agents to prevent breast cancer in woman at high risk.

Estrogens regulate the expression of specific genes and the secretion of certain hormones, and coordinate diverse processes, such as cell proliferation, cell differentiation and tissue organization through pleiotropic actions. Once estrogens reach the bloodstream, they may remain free or bind to serum estrogen-binding proteins like α -fetoprotein (AFP) in rodents [61,62] or sex hormone-binding globulin (SHBG) in humans [62]. Only the free (unbound) hormone is able to diffuse into the target cells, where it binds to the ER to form a hormone–receptor complex. The prevailing model suggests that this complex then interacts with an estrogen response element (ERE) of target genes and activate the transcriptional machinery [63,64].

Fang et al. [65] found a strong linear correlation for ER-binding affinities among a diverse group of chemicals assayed with ER from rat uterine cytosol and human ER- α . Furthermore, the ER-binding data also correlated strongly with the results from assays measuring estrogenicity using a downstream event, i.e. a yeast-based reporter gene assay and MCF-7 cell proliferation assay. These findings demonstrate that ER binding is the major determinant for ER-mediated effects. Therefore, modeling ER binding is essential for understanding the structural requirements for potential drug candidates and for identifying potential EDCs. Thus far, most QSAR models for ER binding have focused on (1) structural characterization of ligands to identify those features required for binding; and (2) identification of estrogenic EDCs for the purpose of priority setting.

3.1. Structural characterization

In the past few years, a number of QSAR models have been developed for ligand binding to the ER [29–34,43,44,66–68]. Most of these ER models were constructed using CoMFA [35].

The ER subfamily is comprised of ER- α and ER- β , which share $\sim 55\%$ sequence identity in their LBDs and $\sim 72\%$ identity (26 of 36 residues) within the ligand-binding pocket. The expression patterns of these two ER isoforms across tissues as well as their ligand-binding patterns were found to exhibit notable differences [46]. Tong et al. [33] employed CoMFA to identify and differentiate the structural features of estrogens responsible for ligand binding to ER- α and ER- β . In the CoMFA contour maps (Fig. 2), the green/yellow polyhedra denote regions around the ligands where an increase in steric bulk is favorable/unfavorable for binding, and the red/blue polyhedra denote regions in which negative/positive electrostatic potential is preferred for enhanced binding.

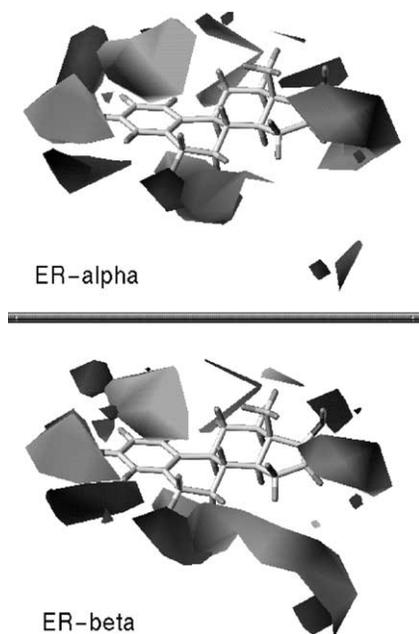


Fig. 2. CoMFA steric and electrostatic contour plots for estrogen binding to ER- α and ER- β . The 17β -estradiol molecule has been inserted inside the fields as a 3D geometrical reference. Enhanced RBA (i.e. higher-binding affinity) is associated with adding/subtracting steric bulk from green/yellow regions and with adding/subtracting positive electrostatic charge in the blue/red regions.

Comparison of the respective CoMFA contour plots between ER- α and ER- β revealed only subtle differences at the 7α position (Fig. 3) that nevertheless may be biologically relevant.

Several contour maps are reported in the literature from CoMFA studies on a number of datasets, including 58 structurally diverse ER ligands by Waller et al. [30], 42 steroidal congeners by Wiese et al. [31], 30 DES derivatives by Sadler et al. [68], 71 halogenated estradiol derivatives by Gantchev [69], and 130 structurally diverse ligands by Shi et al. [43]. These maps consistently indicated that negative charge around the 3-OH and 17β -OH of estradiol favored binding. These findings are consistent with historical observations as well as studies on the ER crystal structure [12] showing that the 3 and 17β -OH positions of estradiol are critical for high-affinity ER binding. The CoMFA contour provided by Gantchev [69] also showed that some small sterically bulky groups (e.g. CH_3) at the 17α , 7α and 11β positions are tolerated, in agreement with SAR results [70]. In several CoMFA models [30,31,33], steric intolerance in the vicinity of the steroid A-ring was interpreted to indicate that this region of the receptor's binding pocket exhibits a preference for planar aromatic rings. This explanation is consistent with both biological activity data [70,71] and structural information [11, 12] on ER.

QSAR models developed using CoMFA are sensitive to molecular alignment, a process which can be somewhat arbitrary and subjective. Consequently, other QSAR approaches have been evaluated for ER binding. Tong et al. [32] investigated the utility of structural descriptors to construct classical QSAR models for a set of ER ligands. They found that obtaining a statistically robust model is highly dependent on the ability of the selected descriptors

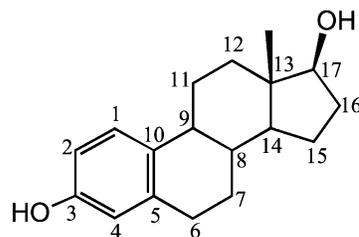


Fig. 3. Structure of 17β -estradiol. The same atom numbering scheme applies to other steroids mentioned in this paper.

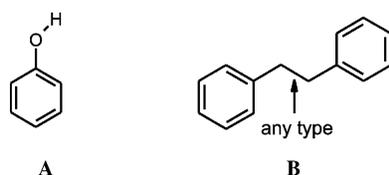


Fig. 4. Depiction of structural indicators deemed important for ER binding: (A) phenolic ring; (B) DES skeleton.

to encode the variation in activity with chemical structure. It is clear that more one knows at the molecular level about the estrogenic activity of the chemicals, the easier the task in selecting among the wide variety and types of specific molecular descriptors that correlate with binding [72].

For example, the crystal structure of the estradiol–ER complex reveals that the hydroxyl group at the three position of the phenolic A ring forms hydrogen bonds with Glu 353, Arg 394 and a conserved water molecule in the receptor-binding site [12], thus stabilizing the binding conformation. Moreover, the crystal structures of ER complexed with DES, 4-OH-tamoxifen and raloxifene reveal this same binding pattern [11]. These observations suggest that a phenolic ring is likely a common structural feature associated with tight binding to the receptor. By including an indicator for the phenolic ring (Fig. 4(A)), Gao et al. [73] was able to develop a statistically valid classification model using a binary QSAR approach. Shi et al. [43] recently combined the same phenolic ring indicator with the standard steric–electrostatic CoMFA descriptors for model development. The inclusion of this indicator enhanced the predictive ability of the model as measured by the increase in the cross-validated r^2 (q^2) from 0.65 to 0.71. Although the steric/electrostatic field contributed predominately to the ER binding, the 6.8% contribution from the structural indicator confirms the appropriateness in separately including the contribution of the phenolic A ring.

Gao et al. [73] also introduced a hexestrol indicator by recognizing that this structural common feature was associated with the high-estrogenic activity of DES analogues. The hexestrol indicator was used for a binary QSAR model. In an integrated system developed by Tong and co-workers [43,74–76] for use as a priority setting tool for EDCs, a similar structural indicator, DES skeleton (Fig. 4(B)) was used in one of

the beginning Phases of the ‘four-phase’ model (vide infra) as a structural alert to identify potential ER binders [74–76].

Commercial molecular modeling programs often feature a wide array of chemical structure descriptors available for developing QSAR models. However, the aforementioned examples demonstrate that inclusion of biologically relevant descriptors (which are not necessarily included in the commercial software) through benefit of expert knowledge can improve substantially the performance of the QSAR model, at least for ER binding.

3.2. Hazard identification

This particular application has focused most recently on development of QSAR models to predict ER-binding affinity for prioritizing potential estrogenic EDCs. Two QSAR-based systems, the common reactivity pattern (COREPA) approach [77,78] and the four-phase approach [74–76], are being evaluated by the EPA to determine their appropriateness for priority setting of potential EDCs.

The endocrine disruptor knowledge base (EDKB) project team [79] at the FDA’s National Center for Toxicological Research (NCTR) has developed an integrated systems for priority setting of EDCs [74–76]. The system is focused on minimizing possible false negatives since chemicals labeled as ‘inactive’ in the process are dropped into a lower priority category for experimental testing. For this purpose, different computational models have been rationally integrated into a four-phase scheme according to the nature of each model. A progressive phase paradigm is used as a screen to reduce the number of chemicals to be considered in the subsequent phase. Therefore, these four phases work in a hierarchical way to incrementally reduce the size of a dataset with increasing precision of prediction. Within each phase, different models have been selected to work complementarily in representing key activity-determining structure features to minimize the rate of false negatives. For predicting ER-binding affinity, the models comprised of the four phases following:

- *Phase I: Filtering*—Two rejection filters, molecular weight <94 or >1000 and no-ring structure, were used to significantly and with high confidence

eliminate those chemicals extremely unlikely to bind ER [76]. These two filters were validated on ~2000 chemicals whose ER activities were available from the literature.

- *Phase II: Active/inactive assignment*—The chemicals passing through Phase I were assigned as YES/NO for ER binding using three different methods, i.e. structural alerts, pharmacophore searching, and classification models. While structural alerts identify key 2D structural features associated with ER binding, pharmacophore search identifies 3D substructure important for ER binding. Classification models use pattern recognition to qualitatively categorize chemicals into active and inactive subsets on the basis of their similarity in physicochemical properties. In its current form, this Phase employs in parallel 11 models, three structural alerts, seven pharmacophores and one classification model to discriminate active from inactive chemicals. A chemical predicted to be active by any of these models is subsequently evaluated in the Phase III, while others are eliminated. Since each method incorporates and weighs differently the various structural features that endow a chemical with the ability to bind the ER, the combined outputs derived from the three approaches are complementary in minimizing false negatives. Moreover, combining the outputs of these 11 models provides a rational means to rank order the chemicals in decreasing order of potential activity [76].
- *Phase III: Quantitative predictions*—In this phase, a CoMFA model is used to make a more accurate quantitative activity prediction for chemicals from Phase II. Chemicals with higher predicted binding affinity are given higher priority for further evaluation in Phase IV. The CoMFA model demonstrated good statistical reliability using both internal and external validation [43].
- *Phase IV: Rule-based decision-making system*—In this final stage of the integrated priority setting approach, a rule-based (or knowledge-based) decision-making system is employed to foster definitive decision making. The system would be useful only after incorporating accumulated human knowledge and expertise (i.e. rules). Nonetheless, combining information from Phases II and III with other sources, such as production volume,

environmental fate and so forth should provide sufficient information to make a final decision on priority setting.

This approach has been validated by a number of existing datasets, including the NCTR ER-binding dataset [80], the E-SCREEN assay data [81], the yeast two-hybrid reporter gene assay data [82], and other datasets [30,83–87]. Thus far, the system has produced no false negatives as would be critical in priority setting for regulatory purpose. When the Phase I and II protocols were applied to 58,000 chemicals recognized by EPA as a representative subset of the 80,000 chemicals, some 9100 chemicals were identified as potential estrogens of which some 3600 have activity no less than 10^5 -fold below 17β -estradiol. Therefore, the method dramatically reduced the number of potential estrogens by some 83% and with a low rate of false negatives as required. The same integrated scheme is being extended to include endpoints of other endocrine disrupting mechanisms (e.g. AR binding).

COREPA is a unique SAR approach designed to analyze 'reactivity patterns' of structurally diverse but biologically similar chemicals [88]. COREPAs represent a set of the specific ranges of structural descriptors determined with the biological activity of concern. Constructing a SAR model using COREPA consisted of the following steps: (1) multiple conformers within 20 kcal/mol of the lowest energy structure are generated for each chemicals in a dataset; (2) a set of descriptors are calculated for each conformer of each chemical, thus the discrete distribution of the descriptor across the conformers is obtained for each chemical; (3) specify a cutoff to separate a dataset into two groups, active and inactive; (4) a set of parameters are determined by evaluating the degree of overlap between the distribution associated with active and inactive groups, which provided the maximum measure of similarity within groups and least overlap between groups; (5) the cutoff-dependent common reactive patterns are obtained as products of the probabilistic distributions for specific parameters associated with active or inactive chemicals; (6) a decision tree based on these parameters is established to construct an SAR model; (7) a cutoff is re-specified to separate the dataset, and steps 4–7 are repeated.

The COREPA models based on 26 steroids and 19 non-steroids [89,90] were recently reported for ER relative-binding affinity (RBA) by selecting cutoff of >150, 100–10, 10–1, and 1–0.1% [78,91]. However, no internal or external validation was performed in the study.

4. Models for other nuclear receptor superfamily

Second to ER in the number of published QSAR studies are the PR [42,92–94] and the AR [29,93]. Not many QSAR models have been developed for corticosteroids [95–97], retinoid acid [98] and thyroids [99]. This is mainly due to the limited amount of biological data for these receptors.

An AR CoMFA model reported by Waller et al. [29] was based on 28 structurally diverse natural, synthetic, and environmental chemicals, of which 21 were used as a training set and seven as a test set. The ability of the model to accurately predict-binding affinity of the testset molecules was demonstrated in the study.

A comparison of PR and AR binding using the CoMFA technique was reported by Loughney [93]. The contour map indicates that sensitivity to steric bulk in the region of the steroid A ring is greater for PR than for AR. This is consistent with the knowledge that a large class of the anti-androgens [100] have an electron withdrawing substituent in the A ring. Furthermore, the steric contour for the PR model indicates tolerance for steric bulk in the region of the 17 α position whereas, in contrast, binding affinity is decreased in the AR model. This is consistent with the rational design of orally active analogues of progesterone by adding substituents to the 17 α position. Chemicals, such as norethindrone (norethisterone) and (levo)norgetrel have lower AR-binding affinity but higher PR-binding activity.

In an early study, Ojasoo et al. [101] compared 3D structures of AR with PR where both receptor structures were derived from homology modeling. They found that empty space is present in both receptors around the 3-keto, 11 β , 7 α positions, and C21 positions for PR. The crystal structures for both AR [15] and PR [10] are now available. Comparison of the LBDs between AR and PR should allow us to

better understand structural requirements for both receptors.

So et al. reported [42] QSAR studies on PR binding by using several variable selection approaches, including forwarding stepping regression (FSR), GFA, generalized simulated annealing (GSA), and genetic neural network (GNN). A comparison of the predictive qualities for both training and test chemicals demonstrated that the GNN protocol achieved the best results.

An excellent review paper on PR QSAR and SAR models is reported by Bursi [92]. Thus, this topic will not be discussed further.

5. Compilation of standard biological datasets on NRs for developing statistically valid QSAR models

Many QSAR approaches have been developed in the field of drug discovery. The methods of choice are dependent on a number of factors. Comparison of these approaches by use of the same dataset permits evaluation of the quality of a method relative to other methods. The developers of the CoMFA technique, first introduced in 1988 [35], selected a dataset composed of 31 steroids with binding affinity for corticosteroid-binding globulins to introduce and validate the method. This dataset was divided into a training and test sets of 21 and 10 steroids, respectively. A variety of new QSAR methods were evaluated using this dataset to compare their performance with CoMFA, including SOMFA [102], CoMMA [103,104], CoMSIA [105,106], COMPASS [107], molecular similarity matrices [102], MS-WHIM [108], MEDV-13 [109], etc. [110–119]. Consequently, and inadvertently, this steroid dataset has become the de facto benchmark for assessing a QSARs acceptability compared with CoMFA as reviewed by Coats [120].

Evaluation of new methods based on this particular dataset has certain limitations, specifically for application in the NRs. One of the major concerns is that the dataset contains only 21 steroidal congeners with an activity range of less than three orders of magnitude. Such QSAR models, which explore only a small portion of chemical-structure and biological-activity space, fail to predict activity even for

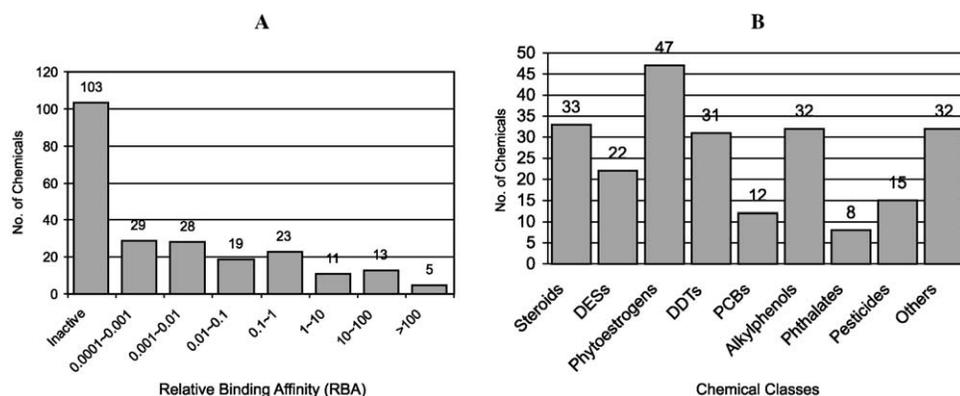


Fig. 5. The distribution of (A) binding activity and (B) chemical classes for the NCTR dataset. The activity is represented as RBA. The RBA for the endogenous ligand, 17 β -estradiol, was set to 100.

chemicals that differ only slightly in structure from the training set. Additionally, the structures of eight chemicals were drawn incorrectly in the original publication [115,120], and the activities for seven inactive chemicals ($K_i < 1 \times 10^5 \text{ M}^{-1}$) were arbitrarily assigned a specific value ($\text{p}K_i = -5.0$). These assumptions, uncertainties, and errors would argue against using this steroid dataset as a standard for developing, validating, or comparing various QSAR models.

Recently, the EDKB team at the US FDA's NCTR has reported a more adequate dataset to validate QSAR methods, particularly for use in the NRs [70,80]. They validated a rat ER-binding assay and measured the binding affinity for over 230 chemicals for use in QSAR model development. This 'NCTR dataset' contains chemicals that were selected to cover the structural diversity of chemicals that bind to ER with an activity distribution ranging over six orders of magnitude, which is an essential requirement for a robust predictive model applicable to structurally diverse estrogens (Fig. 5). The chemical selection process was highly interdisciplinary, involving computational chemists, biologists and experimental toxicologists, and has resulted in steady improvement in performance of the QSAR models [79]. In terms of experimental quality, number of chemicals, activity range and structural diversity, the NCTR dataset represents the most reliable and self-consistent dataset currently available on estrogens to build QSAR models. The NCTR dataset has been used recently to compare the performance of HQSAR to

CoMFA [43,70]. Along with the NCTR dataset, the Waller [121] and Kuiper [89] datasets were used as the testsets for external validation. This study demonstrated that these three datasets are superior to the earlier small datasets [34] for evaluating the relative strengths and weaknesses of CoMFA versus HQSAR.

6. Concluding remarks

It is clear that significant progress has been made in developing and validating QSAR models for the prediction of binding affinity for the NRs. The NRs have become the subject of numerous studies, not only for medicinal chemists and toxicologists but also for computational chemists, crystallographers, and statisticians. These models help us to understand the mechanisms in receptor binding, to predict binding affinity for environmental and other chemicals as tools for government regulation control, and to identify leads in drug discovery. The recent growth of lab-on-chip (i.e. microarray and protein array) technology [122–124] and advanced recombinant DNA technologies (cDNA cloning, Southern blotting, PCR, etc.) have in the last decade enabled rapid identification of macromolecules as well as their expression at sufficient purity and quantities adequate for structure determination. This dramatic increase in the availability of 3D structures for many receptors has greatly expanded the list of potential drug targets, and also provides a rich source of data for QSAR and related

computational techniques [125,126]. By combining ligand- and structure-based approaches in QSAR, accurate and biologically meaningful models should be possible for the receptor-mediated effects.

Acknowledgments

The authors wish to thank Roger Perkins (Logicon ROW Sciences, Jefferson, AR), Gary Timm (EPA, Washington, DC) and Jim Kariya (EPA, Washington, DC) for providing guidance and support to the EDKB project. The authors also thank Dr Willie Owens (The Procter & Gamble Company, Cincinnati, OH) for providing valuable input on the EDKB project. WJW wishes to acknowledge the financial support provided by a grant from the US Environmental Protection Agency's Science to Achieve Success (STAR) program.

References

- [1] V. Giguere, *Endocr. Rev.* 20 (1999) 689.
- [2] G.L. Zubay, W.W. Parson, D.E. Vance, *Principles of Biochemistry*, Wm. C. Brown Communications, Inc., Dubuque, IA, 1995.
- [3] D.J. Mangelsdorf, C. Thummel, M. Beato, P. Herrlich, G. Schutz, K. Umesono, B. Blumberg, P. Kastner, M. Mark, P. Chambon, R.M. Evans, *Cell* 83 (1995) 835.
- [4] N. Poujol, J.M. Wurtz, B. Tahiri, S. Lumbroso, J.C. Nicolas, D. Moras, C. Sultan, *J. Biol. Chem.* 275 (2000) 24022.
- [5] P.F. Egea, A. Mitschler, N. Rochel, M. Ruff, P. Chambon, D. Moras, *EMBO J.* 19 (2000) 2592.
- [6] A. Pike, A. Brzozowski, R. Hubbard, T. Bonn, A. Thorsell, O. Engstrom, J. Ljunggren, J. Gustafsson, M. Carlquist, *EMBO J.* 18 (1999) 4608.
- [7] D.M. Tanenbaum, Y. Wang, S.P. Williams, P.B. Sigler, *Proc. Natl Acad. Sci. USA* 95 (1998) 5998.
- [8] J. Uppenberg, C. Svensson, M. Jaki, G. Bertilsson, L. Jendeberg, A. Berkenstam, *J. Biol. Chem.* 273 (1998) 31108.
- [9] J.M. Wurtz, U. Egner, N. Heinrich, D. Moras, A. Mueller-Fahnow, *J. Med. Chem.* 41 (1998) 1803.
- [10] S.P. Williams, P.B. Sigler, *Nature* 393 (1998) 392.
- [11] A.K. Shiau, D. Barstad, P.M. Loria, L. Cheng, P.J. Kushner, D.A. Agard, G.L. Greene, *Cell* 95 (1998) 927.
- [12] A.M. Brzozowski, A.C. Pike, Z. Dauter, R.E. Hubbard, T. Bonn, O. Engstrom, L. Ohman, G.L. Greene, J.A. Gustafsson, M. Carlquist, *Nature* 389 (1997) 753.
- [13] J.P. Renaud, N. Rochel, M. Ruff, V. Vivat, P. Chambon, H. Gronemeyer, D. Moras, *Nature* 378 (1995) 681.
- [14] W. Bourguet, M. Ruff, P. Chambon, H. Gronemeyer, D. Moras, *Nature* 375 (1995) 377.
- [15] P.M. Matias, P. Donner, R. Coelho, M. Thomaz, C. Peixoto, S. Macedo, N. Otto, S. Joschko, P. Scholz, A. Wegg, S. Basler, M. Schafer, U. Egner, M.A. Carrondo, *J. Biol. Chem.* 275 (2000) 26164.
- [16] C.K. Glass, D.W. Rose, M.G. Rosenfeld, *Curr. Opin. Cell Biol.* 9 (1997) 222.
- [17] A. Mueller-Fahnow, U. Egner, *Curr. Opin. Biotechnol.* 10 (1999) 550.
- [18] W.L. Duax, D. Ghosh, V. Pletnev, J.F. Griffin, *Pure Appl. Chem.* 68 (1996) 1297.
- [19] C. Campagnoli, N. Biglia, M. Giai, P. Sismondi, *Zentralbl. Gynakol.* 119 (1997) 38.
- [20] A.S. Boyd, *Am. J. Med.* 86 (1989) 568.
- [21] W.K. Hong, L.M. Itri, *Retinoids and human cancer*, in: M.B. Sporn, A.B. Roberts, D.S. Goodman (Eds.), *The Retinoids: Biology, Chemistry and Medicine*, Raven, New York, 1994, p. 597.
- [22] L.E. Gray Jr., W.R. Kelce, T. Wiese, R. Tyl, K. Gaido, J. Cook, G. Klinefelter, D. Desaulniers, E. Wilson, T. Zacharewski, C. Waller, P. Foster, J. Laskey, J. Reel, J. Giesy, S. Laws, J. McLachlan, W. Breslin, R. Cooper, R. Di Giulio, R. Johnson, R. Purdy, E. Mihaich, S. Safe, C. Sonnenschein, W. Welshons, R. Miller, S. McMaster, T. Colborn, *Reprod. Toxicol.* 11 (1997) 719.
- [23] R.J. Kavlock, G.P. Daston, C. DeRosa, P. Fenner-Crisp, L.E. Gray, S. Kaattari, G. Lucier, M. Luster, M.J. Mac, C. Maczka, R. Miller, J. Moore, R. Rolland, G. Scott, D.M. Sheehan, T. Sinks, H.A. Tilson, *Environ. Health Perspect.* 104 (Suppl. 4) (1996) 715.
- [24] US Congress, Public Law, (1996) 104–182.
- [25] M. Patlak, *Environ. Sci. Technol.* 30 (1996) 540A.
- [26] EDSTAC, <http://www.epa.gov/Opptintr/Opptendo/Finalrpt.htm>, in: Final Report.
- [27] J.D. McKinney, A. Richard, C. Waller, M.C. Newman, F. Gerberick, *Toxicol. Sci.* 56 (2000) 8.
- [28] M. Johnson, G.M. Maggiora, *Concepts and Applications of Molecular Similarity*, Wiley, New York, 1990.
- [29] C.L. Waller, B.W. Juma, L.E. Gray Jr., W.R. Kelce, *Toxicol. Appl. Pharmacol.* 137 (1996) 219.
- [30] C.L. Waller, T.I. Oprea, K. Chae, H.K. Park, K.S. Korach, S.C. Laws, T.E. Wiese, W.R. Kelce, L.E. Gray Jr., *Chem. Res. Toxicol.* 9 (1996) 1240.
- [31] T.E. Wiese, L.A. Polin, E. Palomino, S.C. Brooks, *J. Med. Chem.* 40 (1997) 3659.
- [32] W. Tong, R. Perkins, R. Strelitz, E.R. Collantes, S. Keenan, W.J. Welsh, W.S. Branham, D.M. Sheehan, *Environ. Health Perspect.* 105 (1997) 1116.
- [33] W. Tong, R. Perkins, L. Xing, W.J. Welsh, D.M. Sheehan, *Endocrinology* 138 (1997) 4022.
- [34] W. Tong, D.R. Lewis, R. Perkins, Y. Chen, W.J. Welsh, D.W. Goddette, T.W. Heritage, D.M. Sheehan, *J. Chem. Inf. Comput. Sci.* 38 (1998) 669.
- [35] R.D. Cramer, D.E. Patterson, J.D. Bunce, *J. Am. Chem. Soc.* 110 (1988) 5959.

- [36] J.A. Katzenellenbogen, *Environ. Health Perspect.* 103 (Suppl. 7) (1995) 99.
- [37] S.A. Jones, L.B. Moore, J.L. Shenk, G.B. Wisely, G.A. Hamilton, D.D. McKee, N.C. Tomkinson, E.L. LeCluyse, M.H. Lambert, T.M. Willson, S.A. Kliewer, J.T. Moore, *Mol. Endocrinol.* 14 (2000) 27.
- [38] R.E. Watkins, G.B. Wisely, L.B. Moore, J.L. Collins, M.H. Lambert, S.P. Williams, T.M. Willson, S.A. Kliewer, M.R. Redinbo, *Science* 292 (2001) 2329.
- [39] H. Gao, J.A. Katzenellenbogen, R. Garg, C. Hansch, *Chem. Rev.* 99 (1999) 723.
- [40] R. Garg, A. Kurup, C. Hansch, *Crit. Rev. Toxicol.* 31 (2001) 223.
- [41] D. Rogers, A.J. Hopfinger, *J. Chem. Inf. Comput. Sci.* 34 (1994) 854.
- [42] S.S. So, S.P. van Helden, V.J. van Geerestein, M. Karplus, *J. Chem. Inf. Comput. Sci.* 40 (2000) 762.
- [43] L.M. Shi, H. Fang, W. Tong, J. Wu, R. Perkins, R. Blair, W. Branham, D. Sheehan, *J. Chem. Inf. Comput. Sci.* 41 (2001) 186.
- [44] W. Zheng, A. Tropsha, *J. Chem. Inf. Comput. Sci.* 40 (2000) 185.
- [45] M.J. McGregor, S.M. Muskal, *J. Chem. Inf. Comput. Sci.* 39 (1999) 569.
- [46] G.G. Kuiper, J.G. Lemmen, B. Carlsson, J.C. Corton, S.H. Safe, P.T. van der Saag, B. van der Burg, J.A. Gustafsson, *Endocrinology* 139 (1998) 4252.
- [47] R. Bursi, T. Dao, T. van Wijk, M. de Gooyer, E. Kellenbach, P. Verwer, *J. Chem. Inf. Comput. Sci.* 39 (1999) 861.
- [48] R.D. Beger, J.P. Freeman, J.O. Lay, J.G. Wilkes, D.W. Miller, *Toxicol. Appl. Pharmacol.* 169 (2000) 17.
- [49] R.D. Beger, J.P. Freeman, J.O. Lay Jr., J.G. Wilkes, D.W. Miller, *J. Chem. Inf. Comput. Sci.* 41 (2001) 219.
- [50] R.D. Head, M.L. Smythe, T.I. Oprea, C.L. Waller, S.M. Green, G.R. Marshall, *J. Am. Chem. Soc.* 118 (1996) 3959.
- [51] A.R. Ortiz, M.T. Pisabarro, F. Gago, R.C. Wade, *J. Med. Chem.* 38 (1995) 2681.
- [52] P. Kollman, *Chem. Rev.* 93 (1993) 2395.
- [53] P.R. Jayatilke, A.C. Nair, R. Zauhar, W.J. Welsh, *J. Med. Chem.* 43 (2000) 4446.
- [54] B.C. Oostenbrink, J.W. Pitera, M.M. van Lipzig, J.H. Meerman, W.F. van Gunsteren, *J. Med. Chem.* 43 (2000) 4594.
- [55] W. Sippl, *J. Comput.-Aided Mol. Des.* 14 (2000) 559.
- [56] W. Sippl, H.D. Holtje, *Theochem—J. Mol. Struct.* 503 (2000) 31.
- [57] G.M. Morris, D.S. Goodsell, R. Huey, A.J. Olson, *J. Comput.-Aided Mol. Des.* 8 (1994) 243.
- [58] G. Cruciani, K.A. Watson, *J. Med. Chem.* 37 (1994) 2589.
- [59] T.A. Grese, L.D. Pennington, J.P. Sluka, M.D. Adrian, H.W. Cole, T.R. Fuson, D.E. Magee, D.L. Phillips, E.R. Rowley, P.K. Shetler, L.L. Short, M. Venugopalan, N.N. Yang, M. Sato, A.L. Glasebrook, H.U. Bryant, *J. Med. Chem.* 41 (1998) 1272.
- [60] T.A. Grese, S. Cho, D.R. Finley, A.G. Godfrey, C.D. Jones, C.W. Lugar III, M.J. Martin, K. Matsumoto, L.D. Pennington, M.A. Winter, M.D. Adrian, H.W. Cole, D.E. Magee, D.L. Phillips, E.R. Rowley, L.L. Short, A.L. Glasebrook, H.U. Bryant, *J. Med. Chem.* 40 (1997) 146.
- [61] M.E. Baker, K.L. Medlock, D.M. Sheehan, *Proc. Soc. Exp. Biol. Med.* 217 (1998) 317.
- [62] D.M. Sheehan, M. Young, *Endocrinology* 104 (1979) 1442.
- [63] B. Gillesby, T. Zacharewski, *Environ. Toxicol. Chem.* 17 (1998) 3.
- [64] J.D. Norris, D. Fan, S.A. Kerner, D.P. McDonnell, *Mol. Endocrinol.* 11 (1997) 747.
- [65] H. Fang, W. Tong, R. Perkins, A. Soto, N. Prechtel, D.M. Sheehan, *Environ. Health Perspect.* 108 (2000) 723.
- [66] S. Bradbury, O. Mekenyan, A. GT, *Environ. Toxicol. Chem.* 15 (1996) 1945.
- [67] L. Xing, W.J. Welsh, W. Tong, R. Perkins, D.M. Sheehan, *SAR QSAR Environ. Res.* 10 (1999) 215.
- [68] B.R. Sadler, S.J. Cho, K.S. Ishaq, K. Chae, K.S. Korach, *J. Med. Chem.* 41 (1998) 2261.
- [69] T.G. Gantchev, H. Ali, J.E. van Lier, *J. Med. Chem.* 37 (1994) 4164.
- [70] H. Fang, W. Tong, L. Shi, R. Blair, R. Perkins, W.S. Branham, S.L. Dial, C.L. Moland, D.M. Sheehan, *Chem. Res. Toxicol.* 14 (2001) 280.
- [71] G.M. Anstead, K.E. Carlson, J.A. Katzenellenbogen, *Steroids* 62 (1997) 268.
- [72] W. Tong, W.J. Welsh, L. Shi, H. Fang, R. Perkins, *Environ. Toxicol. Chem.* (2002) in press.
- [73] H. Gao, C. Williams, P. Labute, J. Bajorath, in progress, 1998.
- [74] W. Tong, R. Perkins, J. Wu, L. Shi, M. Tu, H. Fang, R. Blair, W. Branham, D.M. Sheehan, An integrated computational approach for prioritizing potential estrogens, International Symposium on Environmental Endocrine Disruptors, Kobe, Japan, December 9–11, 1999.
- [75] L.M. Shi, W. Tong, H. Fang, R. Perkins, J. Wu, M. Tu, R. Blair, W. Branham, J. Walker, C. Waller, D. Sheehan, *SAR QSAR Environ. Res.* 13 (2002) 69.
- [76] H. Hong, W. Tong, H. Fang, L.M. Shi, Q. Xie, J. Wu, R. Perkins, J. Walker, W. Branham, D. Sheehan, *Environ. Health Perspect.* 110 (2002) 29.
- [77] G.T. Ankley, O.G. Mekenyan, V.B. Kamenska, P.K. Schmieder, S.P. Bradbury, *SAR QSAR Environ. Res.* 13 (2002) 365.
- [78] O.G. Mekenyan, V. Kamenska, P.K. Schmieder, G.T. Ankley, S.P. Bradbury, *Toxicol. Sci.* 58 (2002) 270.
- [79] R. Perkins, J. Anson, R. Blair, W.S. Branham, S. Dial, H. Fang, B.S. Hass, C. Moland, L. Shi, W. Tong, W. Welsh, D. Sheehan, The endocrine disruptor knowledge base (EDKB), a prototype toxicological knowledge base for endocrine disrupting chemicals, in: J.D. Walker (Ed.), *Handbook on Quantitative Structure Activity Relationships (QSARs) for Predicting Chemical Endocrine Disruption Potentials*, SETAC Press, Pensacola, FL, 2001.
- [80] R. Blair, H. Fang, W.S. Branham, B. Hass, S.L. Dial, C.L. Moland, W. Tong, L. Shi, R. Perkins, D.M. Sheehan, *Toxicol. Sci.* 54 (2000) 138.
- [81] A.M. Soto, C. Sonnenschein, K.L. Chung, M.F. Fernandez, N. Olea, F.O. Serrano, *Environ. Health Perspect.* 103 (Suppl. 7) (1995) 113.

- [82] T. Nishirara, J. Nishikawa, T. Kanayama, F. Dakeyama, K. Saito, S. Imagawa, S. Takatori, Y. Kitagawa, S. Hori, H. Utsumi, *J. Health Sci.* 46 (2000) 282.
- [83] K.W. Gaido, L.S. Leonard, S. Lovell, J.C. Gould, D. Babai, C.J. Portier, D.P. McDonnell, *Toxicol. Appl. Pharmacol.* 143 (1997) 205.
- [84] N.G. Coldham, M. Dave, S. Sivapathasundaram, D.P. McDonnell, C. Connor, M.J. Sauer, *Environ. Health Perspect.* 105 (1997) 734.
- [85] E.J. Routledge, J.P. Sumpter, *J. Biol. Chem.* 272 (1997) 3280.
- [86] E.J. Routledge, J. Parker, J. Odum, J. Ashby, J.P. Sumpter, *Toxicol. Appl. Pharmacol.* 153 (1998) 12.
- [87] C.A. Harris, P. Henttu, M.G. Parker, J.P. Sumpter, *Environ. Health Perspect.* 105 (1997) 802.
- [88] O. Mekenyan, N. Nikolova, S. Karabunarliev, S.P. Bradbury, G. Ankley, B. Hansen, *Quant. Struct.-Act. Relat.* 18 (1999) 139.
- [89] G.G. Kuiper, B. Carlsson, K. Grandien, E. Enmark, J. Haggblad, S. Nilsson, J.A. Gustafsson, *Endocrinology* 138 (1997) 863.
- [90] R. Bolger, T.E. Wiese, K. Ervin, S. Nestich, W. Checovich, *Environ. Health Perspect.* 106 (1998) 551.
- [91] S.P. Bradbury, V. Kamenska, P.K. Schmieder, G.T. Ankley, O.G. Mekenyan, *Toxicol. Sci.* 58 (2000) 253.
- [92] R. Bursi, M.B. Groen, *Eur. J. Med. Chem.* 35 (2000) 787.
- [93] D.A. Loughney, C.F. Schwender, *J. Comput.-Aided Mol. Des.* 6 (1992) 569.
- [94] S.P. van Helden, H. Hamersma, 3D-QSAR of the receptor binding of steroids: a comparison of multiple regression, neural networks and comparative molecular field analysis, in: F. Sanz, J. Giraldo, F. Manaut (Eds.), *QSAR and Molecular Modelling: Concepts, Computational Tools and Biological Applications*, Prous Science, Barcelona, 1995, p. 481.
- [95] W.C. Herndon, G. Rum, Three-dimensional topological descriptors and similarity of molecular structures: binding affinities of corticosteroids, in: F. Sanz, J. Giraldo, F. Manaut (Eds.), *QSAR and Molecular Modelling: Concepts, Computational Tools and Biological Applications*, Prous Science, 1995, Barcelona, p. 380.
- [96] G. Grassy, P. Trape, J. Bompard, B. Calas, G. Auzou, *J. Mol. Graph.* 13 (1995) 356.
- [97] J.C. Dore, J. Gilbert, T. Ojasoo, J.P. Raynaud, *J. Med. Chem.* 29 (1986) 54.
- [98] D. Douguet, E. Thoreau, G. Grassy, *Quant. Struct.-Act. Relat.* 18 (1999) 107.
- [99] T.C. Bruice, N. Kharasch, R.J. Winzler, *Arch. Biochem. Biophys.* 62 (1956) 305.
- [100] S.M. Singh, S. Gauthier, F. Labrie, *Curr. Med. Chem.* 7 (2000) 211.
- [101] T. Ojasoo, J. Delettre, J.P. Mornon, C. Turpin-VanDycke, J.P. Raynaud, *J. Steroid Biochem. Biol.* 27 (1987) 255.
- [102] D.D. Robinson, P.J. Winn, P.D. Lyne, W.G. Richards, *J. Med. Chem.* 42 (1999) 573.
- [103] B.D. Silverman, D.E. Platt, *J. Med. Chem.* 39 (1996) 2129.
- [104] B.D. Silverman, *Quant. Struct.-Act. Relat.* 19 (2000) 237.
- [105] G. Klebe, U. Abraham, T. Mietzner, *J. Med. Chem.* 37 (1994) 4130.
- [106] G. Klebe, Comparative molecular similarity indices analysis-comsia, in: H. Kubinyi, G. Folkers, Y.C. Martin (Eds.), *3D QSAR in Drug Design*, vol. 3, Kluwer Academic Publishers, Dordrecht, The Netherlands, 1998, p. 87.
- [107] N. Jain, K. Koile, D. Chapman, *J. Med. Chem.* 37 (1994) 2315.
- [108] G. Bravi, E. Gancia, P. Mascagni, M. Pegna, R. Todeschini, A. Zaliani, *J. Comput.-Aided Mol. Des.* 11 (1997) 79.
- [109] S.-S. Liu, C.-S. Yin, Z.-L. Li, S.-X. Cai, *J. Chem. Inf. Comput. Sci.* 41 (2001) 321.
- [110] C. de Gregorio, L.B. Kier, L.H. Hall, *J. Comput.-Aided Mol. Des.* 12 (1998) 557.
- [111] S.J. Cho, A. Tropsha, *J. Med. Chem.* 38 (1995) 1060.
- [112] U. Norinder, *J. Chemom.* 10 (1996) 95.
- [113] S. Anzali, G. Barnickel, M. Krug, J. Sadowski, M. Wagener, J. Gasteiger, J. Polanski, *J. Comput.-Aided Mol. Des.* 10 (1996) 521.
- [114] G.E. Kellogg, L.B. Kier, P. Gaillard, L.H. Hall, *J. Comput.-Aided Mol. Des.* 10 (1996) 513.
- [115] M. Wagener, J. Sadowski, J. Gasteiger, *J. Am. Chem. Soc.* 117 (1995) 7769.
- [116] M. Hahn, D. Rogers, *J. Med. Chem.* 38 (1995) 2091.
- [117] A.C. Good, S.S. So, W.G. Richards, *J. Med. Chem.* 36 (1993) 433.
- [118] J. Schnitker, R. Gopaldaswamy, G.M. Crippen, *J. Comput.-Aided Mol. Des.* 11 (1997) 93.
- [119] T.I. Oprea, D. Ciubotario, T.I. Sulea, Z. Simon, *Quant. Struct.-Act. Relat.* 12 (1993) 21.
- [120] E.A. Coats, *Perspect. Drug Disc. Des.* 12 (1998) 199.
- [121] C.L. Waller, D.L. Minor, J.D. McKinney, *Environ. Health Perspect.* 103 (1995) 702.
- [122] R.G. Sosnowski, E. Tu, W.F. Butler, J.P. O'Connell, M.J. Heller, *Proc. Natl Acad. Sci. USA* 94 (1997) 1119.
- [123] E.F. Nuwaysir, M. Bittner, J. Trent, J.C. Barrett, C.A. Afshari, *Mol. Carcinog.* 24 (1999) 153.
- [124] U. Scherf, D.T. Ross, M. Waltham, L.H. Smith, J.K. Lee, L. Tanabe, K.W. Kohn, W.C. Reinhold, T.G. Myers, D.T. Andrews, D.A. Scudiero, M.B. Eisen, E.A. Sausville, Y. Pommier, D. Botstein, P.O. Brown, J.N. Weinstein, *Nat. Genet.* 24 (2000) 236.
- [125] I.D. Kuntz, *Science* 257 (1992) 1078.
- [126] T.L. Blundell, *Nature* 384 (1996) 23.