

COMPARATIVE MOLECULAR FIELD ANALYSIS (CoMFA) MODEL USING A LARGE DIVERSE SET OF NATURAL, SYNTHETIC AND ENVIRONMENTAL CHEMICALS FOR BINDING TO THE ANDROGEN RECEPTOR*

H. HONG^a, H. FANG^a, Q. XIE^a, R. PERKINS^a, D. M. SHEEHAN^b and W. TONG^{c,†}

^aNorthrop Grumman Information Technology, Jefferson, AR 72079, USA; ^bDaniel M. Sheehan and Associates, 1422 Scott Street, Little Rock, AR 72202, USA; ^cCenter for Toxicoinformatics, Division of Biometry and Risk Assessment, National Center for Toxicological Research (NCTR), Jefferson, AR 72079, USA

(Received 14 July 2003; In final form 31 August 2003)

A large number of natural, synthetic and environmental chemicals are capable of disrupting the endocrine systems of experimental animals, wildlife and humans. These so-called endocrine disrupting chemicals (EDCs), some mimic the functions of the endogenous androgens, have become a concern to the public health. Androgens play an important role in many physiological processes, including the development and maintenance of male sexual characteristics. A common mechanism for androgen to produce both normal and adverse effects is binding to the androgen receptor (AR). In this study, we used Comparative Molecular Field Analysis (CoMFA), a three-dimensional quantitative structure–activity relationship (3D-QSAR) technique, to examine AR–ligand binding affinities. A CoMFA model with $r^2 = 0.902$ and $q^2 = 0.571$ was developed using a large training data set containing 146 structurally diverse natural, synthetic, and environmental chemicals with a 10^6 -fold range of relative binding affinity (RBA). By comparing the binding characteristics derived from the CoMFA contour map with these observed in a human AR crystal structure, we found that the steric and electrostatic properties encoded in this training data set are necessary and sufficient to describe the RBA of AR ligands. Finally, the CoMFA model was challenged with an external test data set; the predicted results were close to the actual values with average difference of 0.637 logRBA. This study demonstrates the utility of this CoMFA model for real-world use in predicting the AR binding affinities of structurally diverse chemicals over a wide RBA range.

Keywords: QSAR; CoMFA; Androgen receptor; Endocrine disrupting chemicals; Androgen; Environmental chemicals

INTRODUCTION

A number of environmental chemicals, by mimicking natural hormones, can disrupt crucial endocrine functions in experimental animals, wildlife and humans [1,2]. These chemicals, termed endocrine disrupting chemicals (EDCs), may exert adverse effects through a variety

*Presented at CMTPI 2003: Computational Methods in Toxicology and Pharmacology Integrating Internet Resources (Thessaloniki, Greece, September 17–19, 2003).

[†]Corresponding author. E-mail: wtong@nctr.fda.gov

of mechanisms, including binding to nuclear receptors such as the androgen receptor (AR). The scientific debate surrounding EDCs has grown contentious, in part owing to the fact that some suspected EDCs are high production volume and economically important chemicals, and that they may act at low and environmentally relevant doses. Both public and regulatory concerns led to government regulatory interest [3,4] which promoted increased research activity across Europe, Asia, and North America. In response to Congressional action, the U.S. Environmental Protection Agency (EPA) developed and is implementing a plan to screen and test for androgenic, estrogenic, and thyroid endpoints for a large number of chemicals, many of which occur in drinking water and food supplies [5].

Androgens play an important role in many physiologic processes, including the development and maintenance of male sexual characteristics, such as muscle and bone mass, prostate growth, spermatogenesis, and male hair pattern. While there may be several mechanisms of action for androgenic effects, the binding to AR is considered to be a necessary step for receptor-mediated androgenic action, including toxicity, for both agonists and antagonists. The AR is a member of the nuclear receptor superfamily that includes the steroid receptors, as well as the vitamin D, thyroid, retinoic acid, and orphan receptors. Several laboratories have assayed chemicals to determine their AR binding affinities [6–8]. Studies of the structure–activity relationships (SAR) of chemicals binding to AR can provide information on the structural features required for androgenic actions. These can also provide guidance for regulators in their evaluation of potential androgenic EDCs and for studies of AR-related diseases in drug discovery.

Computer-based SAR analysis is now routinely used to predict the pharmacological activity of chemicals [9,10]. In the EDC research, both qualitative and quantitative SAR models have been developed for ligand binding to various hormone receptors [11]. The largest number of QSAR models in the published literature are for estrogen receptor (ER) binding [12–20]. A few papers have reported QSAR models for other nuclear receptors [21–23]. Previous AR QSAR models have limited predictive capabilities for a number of reasons. They normally use small training sets that, in consequence, have limited structural diversity. Furthermore, these models have not been validated by an independent external test set to demonstrate their capability for predicting chemicals not included in the training set.

In this paper, we report a Comparative Molecular Field Analysis (CoMFA) model based on a large number (146) of chemicals selected to span a broad range of both chemical structural categories and affinities, and to include natural, synthetic, and environmental chemicals.

MATERIALS AND METHODS

AR Binding Assay and Training Data Set

The AR binding affinity for 146 chemicals was determined with an AR competitive binding assay using a recombinant rat AR ligand binding domain protein commercially available from PanVera [24]. The detailed assay protocol is described elsewhere [6]. Briefly, a chemical's binding activity was determined by competing with radiolabeled [³H]-R1881 for AR. The IC₅₀ (50% inhibition of [³H]-R1881 binding) for each competitor was determined. The relative binding affinity (RBA) for each competitor was calculated by dividing the IC₅₀ of R1881 by the IC₅₀ of the competitor and multiplying by 100 (RBA = 100 for R1881). The validated assay incubation conditions were 18–20 h at 4°C using 1 nM [³H]-R1881 and 1.84 AR concentrations. The competing chemical concentrations ranged from 4.28 × 10⁻⁹ M to 4.28 × 10⁻⁴ M. All assays were replicated a minimum of two times; the IC₅₀ values are the means of the replicate values. The training data set used in this study was designed to reflect

both the structural diversity of AR ligands and the wide range distribution of AR binding affinities; this is necessary for building a robust and valid QSAR model. The training chemicals have an AR logRBA range of about 6 units (Table I). The diverse classes of chemicals in the training set are shown in Fig. 1, while their logRBA distribution is shown in Fig. 2.

Molecular Modeling

The CoMFA descriptors and the development of a CoMFA model are dependent on the 3-dimensional (3D) structures of chemicals, i.e. the conformation of the molecules. Therefore, before constructing a CoMFA model, the conformations of each molecule in the training data set need to be generated. In this study, the relevant low energy conformation for each molecule was chosen for alignment onto the template structure, R1881. Molecular structures were constructed using the Sybyl 6.7 fragment database [25]. Structures were fully geometrically optimized using the standard Tripos force field [26] and the conjugate-gradient minimizer was used to minimize energy differences with a convergence criterion of 0.001 kcal/mol. The partial charge for every atom in a molecule was calculated using the Gasteiger–Marsili method [27].

CoMFA Alignment

CoMFA has been used widely to relate chemical structures to their chemical and biological properties [28–31]. The hypothesis of CoMFA is that the differences among chemicals in a target property, such as binding affinity, are often correlated with the differences in the non-covalent fields surrounding their structure. These fields, i.e. the steric (Lennard–Jones) and electrostatic (Coulombic) fields, are calculated at regular intervals throughout a defined region. While there are many adjustable parameters in CoMFA, the most important factor is the relative alignment of the individual molecules when their fields are calculated. Most alignment rules employ a least-squares fitting of pharmacophoric elements between a template molecule and other molecules in the training data set. In this study, we investigated the SAR of 146 diverse AR ligands from different structural classes. The selection of an appropriate template structure and the alignment of these diverse ligands to the template are important for developing a reliable CoMFA model. R1881 was used as the template structure, given that (i) it is one of the highest affinity AR ligands; (ii) its structure is rigid; and (iii) its binding conformation in the AR binding site has been determined in 3D space by single crystal X-ray crystallography [32]. The rationale and results of the alignment of all 145 chemicals onto the template structure are summarized in Table II.

Generating CoMFA Fields

The CoMFA fields are the interaction energies between a probe atom (or a molecule) and a set of aligned molecules, which are used to establish the three-dimensional quantitative structure–activity relationship (3D-QSAR) equations. To generate the CoMFA fields, a probe atom is systematically moved from one point to another for each aligned molecule within a defined 3D grid. At each grid point, the interaction energy is calculated between the probe and the target molecule. In this study, the 146 aligned molecules were placed in a 3D cubic lattice with 2 Å spacing and 2704 grid points (16 × 13 × 13). The steric (van der Waals) and electrostatic (Coulombic) interaction energies were calculated for each molecule at each grid point using a sp³ carbon probe with a +1.0 charge. Energies greater than +40 kcal/mol

TABLE I Experimental and CoMFA-calculated logRBA for 146 compounds

| Name | Measured | CoMFA | |
|---|----------|---------|--------|
| | | Fitting | LOO |
| 1,3-Diphenyltetramethyldisiloxane | - 3.13 | - 2.67 | - 1.07 |
| 11-keto-Testosterone | 0.54 | 1.01 | 1.28 |
| 16 β -OH-16 α -Me-3-Me-Estradiol | - 2.08 | - 1.61 | - 1.03 |
| 17 α -Estradiol | - 2.40 | - 1.49 | - 0.68 |
| 17-Deoxyestradiol | - 2.13 | - 1.90 | - 1.79 |
| 1-Methoxy-4-[1-propenyl]benzene | - 3.19 | - 3.10 | - 2.93 |
| 2-(4-Nitro-benzyl)-isoindole-1,3-dione | - 2.46 | - 2.42 | - 3.01 |
| 2-(4-OH-Benzyl)-isoindole-1,3-dione | - 2.76 | - 3.21 | - 3.37 |
| 2,2',4,4'-Tetrachlorobiphenyl | - 1.74 | - 1.84 | - 2.23 |
| 2,3,4,5-Tetrachloro-4'-biphenylol | - 1.73 | - 1.80 | - 2.12 |
| 2,4,5-T | - 3.18 | - 2.79 | - 2.39 |
| 2,4'-Dichlorobiphenyl | - 1.72 | - 2.18 | - 2.58 |
| 2,4-Dihydroxybenzophenone | - 2.53 | - 2.26 | - 2.31 |
| 2-Benzyl-isoindole-1,3-dione | - 3.12 | - 3.15 | - 2.62 |
| 2-OH-Estradiol | - 1.44 | - 0.82 | - 0.59 |
| 2- <i>sec</i> -Butylphenol | - 2.52 | - 2.14 | - 1.95 |
| 3,3',5,5'-Tetrachloro-4,4'-biphenyldiol | - 2.10 | - 1.71 | - 1.64 |
| 3,3'-Dihydroxyhexestrol | - 2.08 | - 1.88 | - 1.53 |
| 3,4-Diphenyltetrahydrofuran | - 1.98 | - 1.97 | - 2.50 |
| 3 α -Androstenediol | - 0.81 | - 0.18 | 0.10 |
| 3 β -Androstenediol | 0.36 | 0.48 | 0.53 |
| 3-Chlorophenol | - 3.17 | - 2.79 | - 2.52 |
| 3-Deoxyestradiol | 0.54 | - 0.65 | - 1.10 |
| 3-Methylestriol | - 2.25 | - 2.57 | - 2.70 |
| 4-(3,5-Diphenylcyclohexyl)phenol | - 2.27 | - 2.28 | - 1.86 |
| 4,4'-Dihydroxybenzophenone | - 2.67 | - 2.88 | - 2.98 |
| 4,4'-Dihydroxystilbene | - 2.44 | - 1.81 | - 1.63 |
| 4,4'-Sulfonyldiphenol | - 3.09 | - 2.91 | - 2.13 |
| 4-Amino butylbenzoate | - 2.85 | - 3.46 | - 3.64 |
| 4-Androstenediol | - 0.31 | - 0.14 | - 0.03 |
| 4-Androstenedione | - 0.62 | - 1.48 | - 2.03 |
| 4-Benzyloxyphenol | - 2.89 | - 2.79 | - 2.29 |
| 4-Chloro-2-methyl phenol | - 2.59 | - 2.56 | - 2.71 |
| 4'-Chloroacetoacetanilide | - 3.46 | - 3.35 | - 2.82 |
| 4-Dodecylphenol | - 1.81 | - 1.67 | - 1.57 |
| 4-Heptyloxybenzoic acid | - 2.74 | - 2.56 | - 1.96 |
| 4-Heptyloxyphenol | - 1.69 | - 1.82 | - 1.97 |
| 4-Hydroxybenzophenone | - 2.78 | - 2.67 | - 2.66 |
| 4-Hydroxybiphenyl | - 1.43 | - 1.91 | - 2.42 |
| 4'-Hydroxychalcone | - 2.27 | - 1.91 | - 1.89 |
| 4-Hydroxychalcone | - 2.19 | - 1.99 | - 2.18 |
| 4-Hydroxy-tamoxifen | - 1.49 | - 1.62 | - 1.66 |
| 4- <i>n</i> -Octylphenol | - 1.80 | - 1.76 | - 1.81 |
| 4-OH-Estradiol | - 0.91 | - 0.82 | - 0.79 |
| 4- <i>sec</i> -Butylphenol | - 2.44 | - 2.59 | - 2.68 |
| 4- <i>tert</i> -Amylphenol | - 2.39 | - 2.09 | - 2.00 |
| 4- <i>tert</i> -Butylphenol | - 2.67 | - 2.59 | - 2.51 |
| 5,6-Didehydroisoandrosterone | - 1.98 | - 1.89 | - 1.54 |
| 5 α -Androstane | - 3.32 | - 2.25 | - 1.51 |
| 5 α -Androstane-17 β -ol | 1.45 | 0.13 | - 0.20 |
| 5 α -Androstane-3,11,17-trione | - 1.64 | - 1.22 | - 0.75 |
| 5 α -Androstane-3 β -ol | - 0.74 | - 1.93 | - 2.36 |
| 6 α -Me-17 α -OH-Progesterone | - 0.41 | - 0.27 | - 0.12 |
| 6 α -Me-17 α -OH-Progesterone acetate | 0.94 | 0.77 | - 2.17 |
| 6-Hydroxyflavanone | - 1.78 | - 2.02 | - 2.16 |
| 6-Hydroxyflavone | - 2.77 | - 2.45 | - 2.32 |
| Aldrin | - 2.02 | - 1.86 | - 1.27 |
| Androstenediol | - 0.66 | 0.37 | 0.81 |
| Androsterone | - 2.12 | - 2.39 | - 2.13 |
| Aurin | - 1.70 | - 1.36 | - 0.86 |

TABLE I – *continued*

| <i>Name</i> | <i>Measured</i> | <i>CoMFA</i> | |
|--------------------------------------|-----------------|----------------|------------|
| | | <i>Fitting</i> | <i>LOO</i> |
| Bis(<i>n</i> -octyl) Phthalate | – 3.28 | – 3.40 | – 3.16 |
| Bisphenol A | – 2.39 | – 2.18 | – 1.68 |
| Bisphenol B | – 2.09 | – 1.93 | – 1.81 |
| Butylbenzylphthalate | – 2.07 | – 1.84 | – 1.84 |
| β -Zearalanol | – 1.72 | – 1.71 | – 2.38 |
| β -Zearalenol | – 2.09 | – 2.26 | – 2.63 |
| Carbaryl | – 3.12 | – 3.02 | – 2.81 |
| Chalcone | – 2.32 | – 2.00 | – 2.06 |
| Chlordane | – 1.51 | – 1.90 | – 2.40 |
| Clomiphene | – 1.64 | – 1.77 | – 2.13 |
| Corticosterone | – 1.87 | – 1.49 | – 0.09 |
| Cortisol | – 2.77 | – 2.52 | – 0.89 |
| Cyproterone acetate | – 0.32 | – 0.95 | – 1.38 |
| Dexamethasone | – 2.42 | – 2.26 | – 1.57 |
| Dibutyl adipate | – 2.73 | – 2.64 | – 2.92 |
| Diethyl phthalate | – 3.44 | – 3.23 | – 2.48 |
| Diethylstilbestrol (DES) | – 1.66 | – 1.79 | – 1.86 |
| Dihydrotestosterone (DHF) | 2.14 | 1.18 | 0.60 |
| Dihydrotestosterone benzoate | 0.07 | 0.07 | – 0.90 |
| Dihydroxymethoxychlor olefin | – 1.31 | – 1.22 | – 1.21 |
| Diisobutyl phthalate (DIBP) | – 2.22 | – 2.19 | – 2.13 |
| Diisobutyl adipate | – 2.84 | – 2.73 | – 2.81 |
| Diisononylphthalate | – 3.56 | – 3.56 | – 3.11 |
| Dimethylstilbestrol (DMS) | – 1.66 | – 1.71 | – 1.91 |
| Di- <i>n</i> -Butyl phthalate (DBuP) | – 1.95 | – 2.10 | – 2.31 |
| Endosulfan (technical grade) | – 1.87 | – 2.44 | – 2.94 |
| Enzophenone | – 2.63 | – 2.68 | – 2.80 |
| Epitestosterone | – 1.00 | – 1.09 | – 0.91 |
| Equol | – 2.39 | – 2.26 | – 2.63 |
| Estradiol (E2) | – 0.12 | – 0.70 | – 0.86 |
| Estriol (E3) | – 3.15 | – 2.93 | – 2.35 |
| Ethylparathion | – 2.05 | – 2.17 | – 1.97 |
| Ethinylestradiol (EE) | – 1.42 | – 1.05 | – 0.81 |
| Etiocolan-17 β -ol-3-one | – 0.10 | – 0.13 | – 0.04 |
| Fenpicionil | – 1.61 | – 1.41 | – 1.66 |
| Flavanone | – 2.25 | – 2.56 | – 2.58 |
| 4'-Hydroxyflavanone | – 2.48 | – 2.39 | – 2.06 |
| Flavone | – 2.40 | – 2.80 | – 2.97 |
| Flutamide | – 2.42 | – 2.49 | – 1.06 |
| Genistein | – 2.44 | – 2.68 | – 2.04 |
| Heptachlor | – 1.64 | – 1.23 | – 1.11 |
| Hexestrol, monomethyl ether | – 1.63 | – 1.67 | – 1.97 |
| HPTE | – 1.47 | – 1.76 | – 1.84 |
| Igepal CO-210 | – 1.78 | – 1.83 | – 2.60 |
| Isoeugenol | – 2.81 | – 2.55 | – 2.35 |
| Kepone | – 1.58 | – 1.53 | – 1.62 |
| Lindane (Gama-HCH) | – 2.12 | – 2.17 | – 2.11 |
| Linuron | – 2.25 | – 2.49 | – 2.76 |
| Methylparathion | – 2.26 | – 2.55 | – 2.54 |
| Methyltestosterone | 1.28 | 1.26 | 1.15 |
| Methyltrienolone (R1881) | 2.00 | 1.45 | 0.77 |
| Metolachlor | – 2.61 | – 2.72 | – 2.46 |
| Mibolerone | 2.27 | 2.36 | 1.56 |
| Monohydroxymethoxychlor olefin | – 1.84 | – 1.81 | – 1.79 |
| Nafoxidine | – 1.63 | – 1.55 | – 2.38 |
| Nonylphenol | – 1.57 | – 1.72 | – 1.88 |
| Nordihydroguaiaretic acid | – 2.28 | – 2.51 | – 3.84 |
| Norethindrone | 0.41 | 0.91 | 1.11 |
| Norethynodrel | – 0.70 | 0.41 | 1.03 |
| Norgestrel | 1.22 | 1.15 | 0.52 |

TABLE I – *continued*

| Name | Measured | CoMFA | |
|----------------------------------|----------|---------|--------|
| | | Fitting | LOO |
| <i>o,p'</i> -DDD | - 1.52 | - 1.57 | - 1.64 |
| <i>o,p'</i> -DDE | - 1.81 | - 1.91 | - 2.04 |
| <i>o,p'</i> -DDT | - 1.69 | - 1.72 | - 1.75 |
| <i>p,p'</i> -DDD | - 1.70 | - 1.68 | - 1.63 |
| <i>p,p'</i> -DDE | - 1.70 | - 2.06 | - 2.22 |
| <i>p,p'</i> -DDT | - 1.76 | - 1.82 | - 1.77 |
| <i>p,p'</i> -Methoxychlor | - 1.94 | - 2.42 | - 2.51 |
| <i>p,p'</i> -Methoxychlor olefin | - 2.20 | - 2.50 | - 2.58 |
| <i>p</i> -Cumyl phenol | - 2.11 | - 1.94 | - 1.87 |
| Procymidone | - 2.61 | - 2.97 | - 2.82 |
| Progesterone | - 0.70 | - 0.72 | - 0.48 |
| Promegestone | - 0.64 | - 0.47 | 0.56 |
| Propanil (DCPA) | - 2.22 | - 2.30 | - 2.44 |
| Propyl parabene | - 3.00 | - 2.81 | - 2.35 |
| Spirolactone | - 0.35 | - 0.13 | 0.79 |
| Tamoxifen | - 1.59 | - 1.47 | - 1.29 |
| Testosterone | 1.28 | 0.69 | 0.48 |
| Testosterone propionate | - 0.79 | - 0.86 | - 1.03 |
| <i>Trans</i> -4-Hydroxystilbene | - 2.13 | - 2.27 | - 2.37 |
| Trenbolone | 2.05 | 1.53 | 0.73 |
| Triphenyl phosphate | - 1.69 | - 1.66 | - 2.97 |
| Triphenylethylene | - 1.98 | - 2.07 | - 2.07 |
| Triphenylsilanol | - 2.05 | - 2.33 | - 2.80 |
| Vinclozolin | - 2.50 | - 2.74 | - 2.64 |
| Zearalanone | - 2.14 | - 1.73 | - 1.20 |
| Zearalenol | - 1.64 | - 1.79 | - 2.11 |

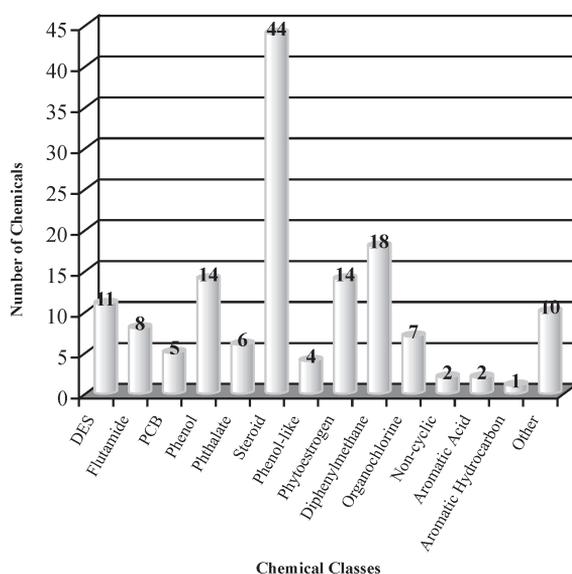


FIGURE 1 Chemical class distribution of the training set.

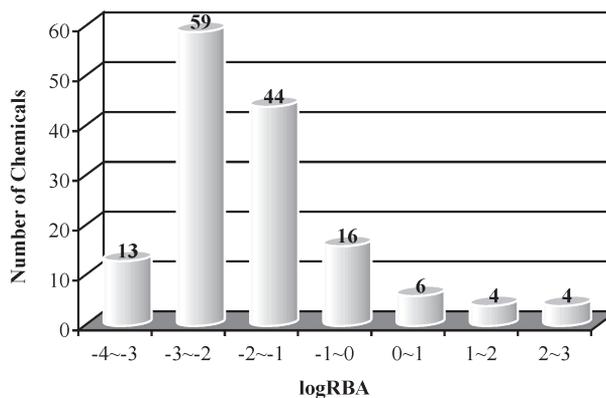


FIGURE 2 AR binding affinity distribution (expressed as logRBA) of the training set.

or less than -40 kcal/mol, which are considered to be inside the van der Waals surface, were truncated to the values of $+40$ kcal/mol and -40 kcal/mol, respectively.

PLS-QSAR

To form the basis for a predictive statistical model, the method of partial least squares (PLS) regression [33] was used to analyze the training data set of 146 chemicals by correlating variations in the AR logRBA with variations in their respective steric and electrostatic energies. In the PLS regression, the logRBA was the dependent variable while the energies were independent variables. The optimum number of principal components (PCs), corresponding to the smallest standard error of prediction, was determined by the leave-one-out (LOO) cross-validation procedure [34]. In this investigation, each chemical was systematically excluded once from the training set, after which its logRBA was predicted using the model derived from the remaining 145 chemicals. Combining the 146 predictions allows the calculation of a cross-validated r^2 (called q^2 hereafter to differentiate it from the correlation coefficient r^2). Using the optimal number of PCs, the final PLS analysis was carried out using all 146 chemicals without cross-validation to build a predictive QSAR model. This model was then evaluated using r^2 . The r^2 and q^2 parameters are two key measures of robustness of a QSAR model; the r^2 value is a measure of a model's goodness to fit the training set, while q^2 value is a measure of a model's predictive power. A model with $r^2 > 0.9$ and $q^2 > 0.5$ is generally considered to be both internally self-consistent and predictive [35].

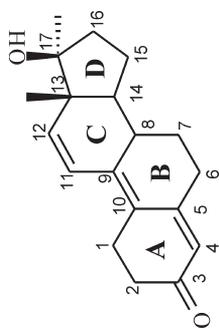
RESULTS AND DISCUSSION

CoMFA Results

The logRBA values calculated by CoMFA for the 146 chemicals are listed in Table I under "Fitting". The CoMFA model has a $r^2 = 0.902$ with a SE of 0.389. The relative contributions of the steric and electrostatic fields are 0.522 and 0.478, respectively, similar to those we reported for ER CoMFA models [12,14,19,20].

To evaluate the predictive power of this CoMFA model, a conventional LOO cross-validation was conducted. The predicted logRBA using LOO for the 146 chemicals are also summarized in Table I under "LOO". The parameter for measuring predictive power of

TABLE II CoMFA alignment rules. All molecules are aligned to the template structure of R 1881



| <i>Chemical classes*</i> | <i>Chemicals aligned</i> | <i>Alignment</i> | <i>Result</i> |
|--------------------------|--------------------------|--|---------------|
| Steroid | 43 | All steroids have the similar steroidal framework that contains the A, B, C, D rings. It is reasonably to hypothesize that steroids adopt a similar orientation in the binding site of AR. Thus, this class of molecules was aligned to the template based on the steroidal skeleton | |
| Diphenyl methane | 18 | The common structure of this class of chemicals has two benzene rings separated by the methane group. One of the benzene rings was aligned onto the ring A and another benzene ring was aligned to the ring C | |
| DES | 7 | Seven DES-like structures were aligned in such way that the two rings of DES were superimposed with the rings A and D of R1881 | |

TABLE II - continued

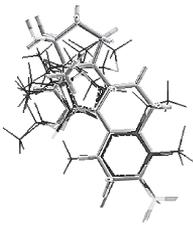
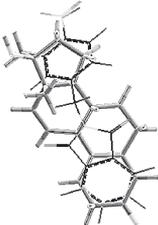
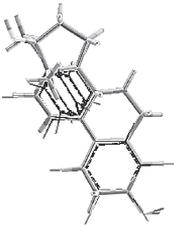
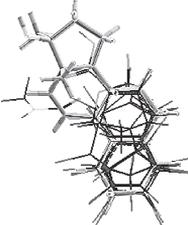
| <i>Chemical classes*</i> | <i>Chemicals aligned</i> | <i>Alignment</i> | <i>Result</i> |
|-------------------------------|--------------------------|---|--|
| Flutamide | 8 | Eight flutamide derivatives share structural commonalities with methylparathion and ethylparathion. These ten molecules were aligned to the template using the benzene ring that was superimposed to the ring A |  |
| Other | 2 | | |
| Other | 3 | Three isoindoles in this class were aligned to the template by superimposing their two benzene rings with the A and D rings |  |
| PCB | 5 | Five PCBs that have two benzene rings connected by a single bond were aligned by overlapping one of the benzene rings onto the ring A and the center of another benzene ring with the center of C ring |  |
| Organochlorine Phenol-like | 7 1 | For carbaryl, its naphthalene ring was superimposed with the rings A and B . Both lindane and 2,4,5-T were aligned by superimposing their six member rings with the A ring. Kepone, endosulfan, heptachlor, chlordane, and aldrin have fused rings; their centers were aligned to the centers of the rings A , B or C |  |

TABLE II – continued

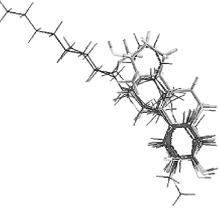
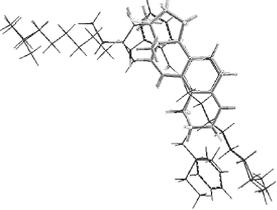
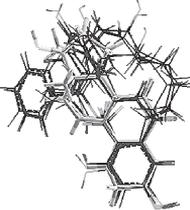
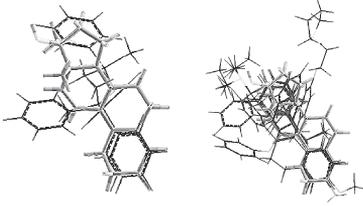
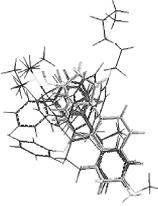
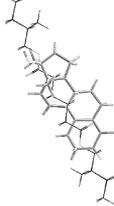
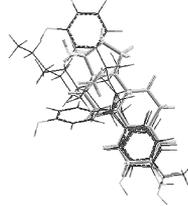
| Chemical classes* | Chemicals aligned | Alignment | Result |
|-------------------|-------------------|---|--|
| Phenol | 14 | It is reasonably assume that the phenolic hydroxyl of this class of chemicals mimic the function of 3-keto of R1881 in AR binding. Thus, Phenols was aligned to the template by superimposing their phenolic rings as well as the hydroxyl group to the ring A and 3-keto group, respectively |  |
| Phthalate | 6 | The unique structural characteristics of this class of chemicals are the two long and flexible branches on a benzene ring. Normally, these structures are capable of forming any shapes that are fitted in the shape of the AR binding domain. Thus, these molecules were aligned along the line between atom 3 and 17 of R1881 to maximize the occupancy of the binding site |  |
| Phytoestrogen | 14 | Flavanones and flavones contain three 6-member rings; these rings were superimposed to the rings A , B , D , correspondingly. Zearelanone derivatives have a benzene ring fused with a big ring; the benzene ring was superimposed to the ring A while the big ring was aligned with the rings B , C and D with maximizing steric superposition. Chalcone derivatives have two benzene rings connected by a three carbon atom chain; one benzene ring was aligned with the ring A |  |

TABLE II – continued

| <i>Chemical classes*</i> | <i>Chemicals aligned</i> | <i>Alignment</i> | <i>Result</i> |
|--------------------------|--------------------------|---|--|
| Other | 2 | Two siloxanes (Triphenylsilanol and 1,3-Diphenyltetramethyldisiloxane) were aligned to the template by superimposing one of their benzene rings to the ring A with maximizing the steric superposition of the rest of molecule to the rings B , C and D |  |
| DES | 4 | 5 antiestrogens along with aurin, 4-(3,5-Diphenylcyclohexyl)phenol, and triphenyl phosphate are triple rings-based structure. The alignment was mainly one of the rings superimposed to the ring A with maximizing superposition of the rest of molecules |  |
| Other | 2 | | |
| Phenol-like | 1 | | |
| Aromatic Hydrocarbon | 1 | | |
| Non-cyclic | 2 | Adipate derivatives have a non-cyclic and flexible structure. They were aligned along the line of the position 3 and 17 to maximize the steric overlapping to the template |  |
| Other | 1 | | |
| Phenol-like | 2 | | |
| Aromatic acid | 2 | 4,4'-sulfonyldiphenol, 4-amino butylbenzoate, nordihydroguaiaretic acid, 1-methoxy-4-[1-propenyl]benzene, and p-heptyloxybenzoic acid contain at least one benzene ring that was superimposed to the ring A with maximizing superposition of the rest of structure |  |

*The definition of chemical classes was described in our previous paper [6].

TABLE III Summary of the CoMFA statistical results

| <i>Statistics</i> | <i>CoMFA</i> |
|-------------------|--------------|
| r^2 | 0.902 |
| q^2 | 0.571 |
| SEE | 0.389 |
| <i>F</i> -value | 156.953 |
| PCs | 8 |
| Contributions (%) | |
| Steric | 0.522 |
| Electrostatic | 0.478 |

a QSAR model, q^2 , can be calculated from the residual between the predicted and experimental data. The q^2 for this CoMFA model is 0.571, indicating a good predictive power.

The statistical results of the CoMFA model are summarized in Table III.

CoMFA Coefficient Contour Map

The results of QSAR analysis by CoMFA, with its thousands of terms, is generally represented in the form of a 3D coefficient contour map. It is a valuable output from a CoMFA model in terms of studying regional specificity of molecular feature related to biological activity, which plays a large role in drug discovery, especially for lead optimization. It illustrates which areas in space around the molecules are associated with variation of biological activity. The contour map of the CoMFA model is plotted in Fig. 3. The structure of R1881 was placed in the map for reference. In Fig. 3, the green polyhedra are the regions where more bulky groups are expected to increase AR binding affinity while the yellow polyhedra are the regions where less bulky groups enhance AR binding affinity. A more positively charged group in the blue regions or a more negatively charged group in the red regions shows increased AR binding affinity.

To further verify the reliability and biological relevance of the CoMFA model, the CoMFA contour map was superimposed on the AR binding site of the X-ray crystal structure of the human AR binding domain [31] based on their common reference structure of R1881. Comparing the CoMFA contour map with the X-ray crystal structure reveals that the CoMFA contour map is consistent with the 3D shape of the AR binding site (Fig. 3). For example, the green regions (A, B, and C) where bulky groups favor binding correspond to the regions of the active site where unfilled spaces exist. This indicates that a bulky group in these regions increases the van de Waals interaction between a ligand and the AR, thus increasing the AR binding affinity.

The binding-favored negative charge regions (the red polyhedra) around the 3-keto and 17β -OH groups indicate a positive contribution to the affinity by either forming H-bonds between the ligand and AR or through electrostatic interaction between the negative charges of the ligand and the positive charges of amino acid residues of AR. In a SAR study in conjunction with the close examination of the ligand-AR crystal structures, we found that the 17β -OH forms two H-bonds with Asn705 and Thr877 while the 3-keto forms a H-bond with Arg752 [6]. This H-bond network is consistent with the CoMFA contour map. In addition, Arg 752 and Gln 711 are located near the red polyhedra of the contour map (Fig. 3). Arg is an amino acid residue with positive charge, indicating that a negative charge group placed around Arg will increase binding. This is consistent with the CoMFA results that the red polyhedra are the region where negative charge favors binding.

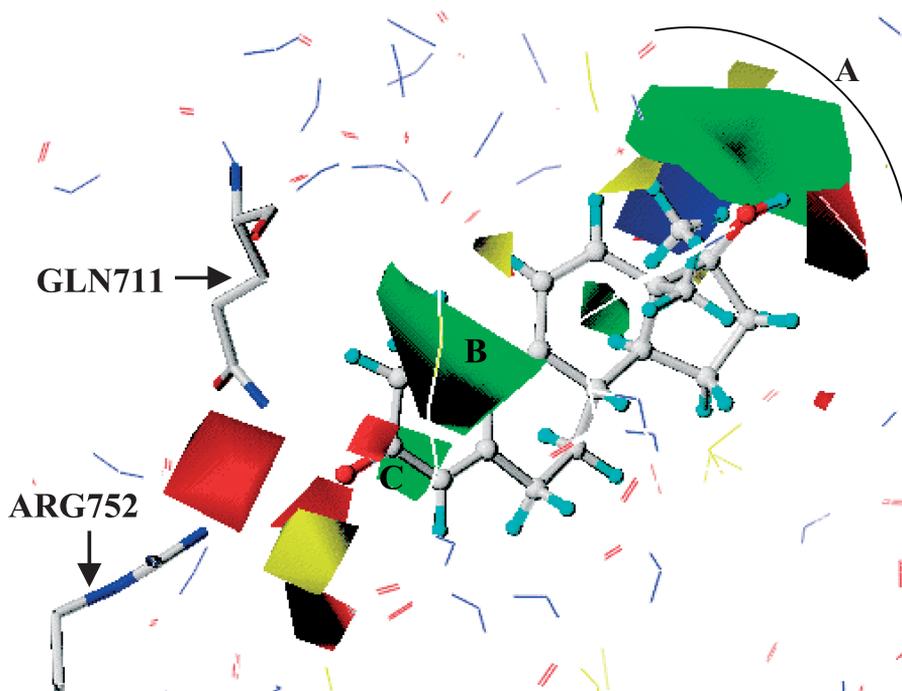


FIGURE 3 Superposition of the CoMFA contour map over the ligand binding site of the AR crystal structure by superimposing their common R1881 structures that presented as a ball and stick model. The AR protein is shown in a stick model with highlighting Gln711 and Arg752. For the CoMFA contour map, greater values of RBA are correlated with more bulky group near the green regions, less bulky groups near the yellow regions, more positive charge groups near the blue regions, and more negative charges near the red regions.

Prediction of Test Chemicals

AR, like ER [13], binds a wide range of chemical structures with varying affinities. This characteristic helps explain why diverse environmental chemicals can act, via AR, as androgenic or anti-androgenic endocrine disruptors [6]. The data presented here demonstrate this range of structural and affinity diversity. Consequently, the challenge in developing AR QSAR models is not only in constructing a statistically sound model (robust in both training and cross-validation steps) for such structurally diverse androgens, but more importantly in developing a model with the capability to accurately predict the activity of chemicals not included in the training set. Thus, we validated our CoMFA model by predicting the AR binding affinity of an external validation data set. This test set contained 8 chemicals with experimental binding data from another laboratory. They were neither tested in our assay nor included in the training set.

Waller *et al.* [22] reported the AR binding affinities for 28 chemicals using a competitive AR binding assay [7]. Of these 28 chemicals, 20 were also tested in our laboratory. Comparing the assay results from both labs for these 20 shared chemicals, we found that both assays are comparable [6]. Thus, the remaining 8 chemicals were used to challenge the model. It is important to note that there are some differences in the assays used by the two laboratories, which result a systematic shift found by comparing data from the 20 common chemicals [6]. Therefore, the CoMFA-predicted logRBAs in this study have to be converted to the scale of Waller's AR binding data in order to compare with the measured values. Converting our logRBA values to Waller's values was based on the relationship between the logRBA of the active chemicals in the assays from both laboratories. The equation was $\log\text{RBA}(\text{Waller}) = 1.09 \log\text{RBA}(\text{NCTR}) - 0.23$ with the $r^2 = 0.92$ [6]. The prediction results for the 8 chemicals are listed in Table IV.

TABLE IV CoMFA prediction result for the external testing data set

| Name | <i>logRBA</i> (Waller) | Predicted <i>logRBA</i> (NCTR) | Predicted <i>logRBA</i> (Waller) | Δ <i>logRBA</i> | Fold Difference* |
|---|---------------------------|-----------------------------------|-------------------------------------|------------------------|---------------------|
| 2,2',4,4',5,5'-Hexachlorobiphenyl | - 3.70 | - 2.33 | - 2.77 | - 0.93 | 8.52 (-) |
| Hydroxy-flutamide | - 0.95 | - 1.22 | - 1.56 | 0.61 | 4.07 (+) |
| Vinclozolin metabolites (M1) [†] | - 2.63 | - 1.61 | - 1.98 | - 0.65 | 4.42 (-) |
| Vinclozolin metabolites (M2) [‡] | - 1.20 | - 1.38 | - 1.73 | 0.53 | 3.42 (+) |
| Anadarone | - 2.70 | - 2.32 | - 2.76 | 0.06 | 1.14 (+) |
| 17 α -Hydroxyprogesterone | - 3.40 | - 1.55 | - 1.92 | - 1.48 | 30.23 (-) |
| Hydroxylinuron | - 2.91 | - 3.08 | - 3.59 | 0.68 | 4.76 (+) |
| 2,2',4',5,5'-Pentachloro-4-biphenylol | - 2.79 | - 2.27 | - 2.70 | - 0.09 | 1.22 (-) |

* +, increase; -, decrease.

[†] M1: 2-[[3,5-dichlorophenyl]-carbamoyloxy]-2-methyl-3-butenic acid.

[‡] M2: 3',5'-dichloro-2-hydroxy-2-methylbut-3-enamide.

Seven of the 8 chemicals had a prediction error (the difference between the measured and predicted *logRBA*) less than 1 *logRBA* unit, i.e. less than a 10-fold error. 17 α -hydroxyprogesterone had an error of 1.48 *logRBA* unit. Four predictions were higher than measured values and four were lower, suggesting the systematic error in prediction is low. The average prediction error for the external test data set is 0.63 *logRBA* unit. This demonstrates that the 3D-QSAR CoMFA model reported in this study has good predictive power. However, it would be useful to have a larger external data set for validation of our model.

CONCLUSION

QSAR is useful for evaluation of toxicity and/or discovery of lead chemicals in drug design. In this paper, a 3D-QSAR CoMFA model was developed using, to the best of our knowledge, the largest and most diverse data set that has been reported for chemical binding to the AR. This study confirms and also enhances our understanding of androgen ligand chemicals in the context of SAR across a wide range of RBAs and diverse chemical classes.

1. The reported CoMFA model is statistically significant with $r^2 = 0.902$ and $q^2 = 0.571$, demonstrating a sound SAR for AR binding affinity. This illustrates that the diverse AR ligands share structural commonalities important for AR binding. The model should be useful for identification of potential AR ligands and guiding synthesis of chemicals with increased or decreased affinity for AR.
2. The AR binding-important regions derived from the CoMFA model are consistent with the physical structure of the AR ligand binding site and the charge distribution of amino acid residue that obtained from the X-ray crystal structure of the AR binding domain bound with R1881. The CoMFA contour map correctly identifies the regions where positive/negative charge groups and more/less bulky groups increase or decrease the AR binding affinity for a chemical.
3. We demonstrated that the CoMFA model generated in this study has a reasonable predictive capability to estimate the AR binding affinity of an external testing data set.

Acknowledgements

The authors gratefully acknowledge the American Chemistry Council, the U.S. Environmental Protection Agency (EPA) and the FDA's Office of Women's Health for financial support.

References

- [1] Kavlock, R.J., Daston, G.P., DeRosa, C., Fenner-Crisp, P., Gray, L.E., Kaattari, S., Lucier, G., Luster, M., Mac, M.J., Maczka, C., Miller, R., Moore, J., Rolland, R., Scott, G., Sheehan, D.M., Sinks, T. and Tilson, H.A. (1996) "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.* **104**(Suppl. 4), 715–740.
- [2] Cooper, R.L. and Kavlock, R.J. (1997) "Endocrine disruptors and reproductive development: a weight-of-evidence overview", *J. Endocrinol.* **152**, 159–166.
- [3] US-Congress (1996) In: U.S.C. Editor, Vol. 21, pp. 346a(p).
- [4] US-Congress (1996) In: U.S.C. Editor, Vol. 42, pp. 300j–317.
- [5] EDSTAC: <http://www.epa.gov/opptintr/opptendo/finalrpt.htm>.
- [6] Fang, H., Tong, W., Branham, S.W., Dial, S.L., Moland, C.L., Hong, H., Xie, Q., Perkins, R., Owens, W. and Sheehan, D.M. (2003) "Structure–activity relationships of 202 natural, synthetic, and environmental chemicals for binding to the androgen receptor", *Chem. Res. Toxicol.* **202**, In press.
- [7] Kelce, W.R., Monosson, E., Gamcsik, M.P., Laws, S.C. and Gray, L.E. Jr. (1994) "Environmental hormone disruptors: evidence that vinclozolin developmental toxicity is mediated by antiandrogenic metabolites", *Toxicol. Appl. Pharmacol.* **126**, 276–285.
- [8] Dalton, J.T., Mukherjee, A., Zhu, Z., Kirkovsky, L. and Miller, D.D. (1998) "Discovery of nonsteroidal androgens", *Biochem. Biophys. Res. Commun.* **244**, 1–4.
- [9] Tong, W., Welsh, W.J., Shi, L., Fang, H. and Perkins, R. (2003) "SAR approaches and applications", *Environ. Toxicol. Chem.* **22**, 1680–1695.
- [10] Perkins, R., Fang, H., Tong, W. and Welsh, W.J. (2003) "Quantitative structure–activity relationship (QSAR) methods: Perspective on drug discovery and toxicology", *Environ. Toxicol. Chem.*, **22**, 1666–1679.
- [11] Fang, H., Tong, W. and Sheehan, D. (2003) "QSAR's in receptor-mediated Effects: The nuclear receptor superfamily", *J. Mol. Struct. (THEOCHEM)* **622**, 113–125.
- [12] Tong, W., Perkins, R., Xing, L., Welsh, W.J. and Sheehan, D.M. (1997) "QSAR models for binding of estrogenic compounds to estrogen receptor alpha and beta subtypes", *Endocrine* **138**, 4022–4025.
- [13] Fang, H., Tong, W., Shi, L., Blair, R., Perkins, R., Branham, W.S., Dial, S.L., Moland, C.L. and Sheehan, D.M. (2001) "Structure activity relationship for a large diverse set of natural, synthetic and environmental chemicals", *Chem. Res. Toxicol.* **14**, 280–294.
- [14] Shi, L.M., Fang, H., Tong, W., Wu, J., Perkins, R., Blair, R., Branham, W. and Sheehan, D. (2001) "QSAR models using a large diverse set of estrogens", *J. Chem. Inf. Comput. Sci.* **41**, 186–195.
- [15] Hong, H., Tong, W., Fang, H., Shi, L.M., Xie, Q., Wu, J., Perkins, R., Walker, J., Branham, W. and Sheehan, D. (2002) "Prediction of estrogen receptor binding for 58,000 chemicals using an integrated system of a tree-based model with structural alerts", *Environ. Health Perspect.* **110**, 29–36.
- [16] Gantchev, T.G., Ali, H. and van Lier, J.E. (1994) "Quantitative structure–activity relationships/comparative molecular field analysis (QSAR/CoMFA) for receptor-binding properties of halogenated estradiol derivatives", *J. Med. Chem.* **37**, 4164–4176.
- [17] Waller, C.L., Oprea, T.I., Chae, K., Park, H.K., Korach, K.S., Laws, S.C., Wiese, T.E., Kelce, W.R. and Gray, L.E. Jr. (1996) "Ligand-based identification of environmental estrogens", *Chem. Res. Toxicol.* **9**, 1240–1248.
- [18] Wiese, T.E., Polin, L.A., Palomino, E. and Brooks, S.C. (1997) "Induction of the estrogen specific mitogenic response of MCF-7 cells by selected analogues of estradiol-17 beta: a 3D QSAR study", *J. Med. Chem.* **40**, 3659–3669.
- [19] Tong, W., Perkins, R., Strelitz, R., Collantes, E.R., Keenan, S., Welsh, W.J., Branham, W.S. and Sheehan, D.M. (1997) "Quantitative structure–activity relationships (QSARs) for estrogen binding to the estrogen receptor: predictions across species", *Environ. Health Perspect.* **105**, 1116–1124.
- [20] Tong, W., Lewis, D.R., Perkins, R., Chen, Y., Welsh, W.J., Goddette, D.W., Heritage, T.W. and Sheehan, D.M. (1998) "Evaluation of quantitative structure–activity relationship methods for large-scale prediction of chemicals binding to the estrogen receptor", *J. Chem. Inf. Comput. Sci.* **38**, 669–677.
- [21] Singh, S.M., Gauthier, S. and Labrie, F. (2000) "Androgen receptor antagonists (antiandrogens): structure–activity relationships", *Curr. Med. Chem.* **7**, 211–247.
- [22] Waller, C.L., Juma, B.W., Gray, Jr, L.E. and Kelce, W.R. (1996) "Three-dimensional quantitative structure–activity relationships for androgen receptor ligands", *Toxicol. Appl. Pharmacol.* **137**, 219–227.
- [23] Loughney, D.A. and Schwender, C.F. (1992) "A comparison of progestin and androgen receptor binding using the CoMFA technique", *J. Comput. Aided Mol. Des.* **6**, 569–581.
- [24] Panvera: <http://www.panvera.com>.
- [25] SYBYL SYBYL molecular modeling system, version 6.7 (Tripose, Inc., St. Louis, MO 63144).
- [26] Clark, M.D.C.R. and Van Opendenbosch, N. (1989) "Validation of the general purpose Tripose 5.2 force field", *J. Comp. Chem.* **14**, 237–245.
- [27] Gasteiger, J. and Marsili, M. (1980) "Iterative partial equalization of orbital electronegativity—a rapid access to atomic charges", *Tetrahedron* **36**, 3219–3228.
- [28] Wilcox, R.E., Huang, W.H., Brusniak, M.Y., Wilcox, D.M., Pearlman, R.S., Teeter, M.M., DuRand, C.J., Wiens, B.L. and Neve, K.A. (2000) "CoMFA-based prediction of agonist affinities at recombinant wild type versus serine to alanine point mutated D2 dopamine receptors", *J. Med. Chem.* **43**, 3005–3019.
- [29] Desiraju, G.R., Gopalakrishnan, B., Jetti, R.K.R., Raveendra, D., Sarma, J.A.R.P. and Subramanya, H.S. (2000) "Three-dimensional quantitative structural activity relationship (3D-QSAR) studies of some 1,-diarylpyrzoles: analogue based design of selective cyclooxygenase-2 inhibitors", *Molecules* **5**, 945–955.

- [30] Puri, S., Chickos, J.S. and Welsh, W.J. (2002) "Three-dimensional quantitative structure–property relationship (3D-QSPR) models for prediction of thermodynamic properties of polychlorinated biphenyls (PCBs): enthalpy of vaporization", *J. Chem. Inf. Comput. Sci.* **42**, 299–304.
- [31] Lee, K.W. and Briggs, J.M. (2001) "Comparative molecular field analysis (coMFA) study of epothilones-tubulin depolymerization inhibitors: pharmacophore development using 3D QSAR methods", *J. Comput. Aided Mol. Des.* **15**, 41–55.
- [32] Matias, P.M., Donner, P., Coelho, R., Thomaz, M., Peixoto, C., Macedo, S., Otto, N., Joschko, S., Scholz, P., Wegg, A., Basler, S., Schafer, M., Egner, U. and Carrondo, M.A. (2000) "Structural evidence for ligand specificity in the binding domain of the human androgen receptor. Implications for pathogenic gene mutations", *J. Biol. Chem.* **275**, 26164–26171.
- [33] Wold, S., Albano, C., Dunn, W.I., Edlund, U., Esbensen, K., Geladi, P., Hellberg, S., Johansson, E., Lindberg, W. and Sjostrom, M. (1984) *Multivariate Data Analysis in Chemistry*. In: Kowalski, B. (Dordrecht, The Netherlands: Reidel, The Netherlands).
- [34] Cramer, R.D.I., Bunce, J.D. and Patterson, D.E. (1988) "Crossvalidation, bootstrapping, and partial least squares compared with multiple regression in conventional QSAR studies", *QSAR* **7**, 18–25.
- [35] Cramer, R.D., Patterson, D.E. and Bunce, J.D. (1988) "Comparative molecular field analysis (CoMFA). 1. Effect of shape on binding of steroids to carrier proteins", *J. Am. Chem. Soc.* **110**, 5959–5967.