Prediction of Estrogen Receptor Binding for 58,000 Chemicals Using an Integrated System of a Tree-Based Model with Structural Alerts

Huixiao Hong,[1] Weida Tong,[1] Hong Fang,[1] Leming Shi,[2] Qian Xie,[1] Jie Wu,[1] Roger Perkins,[1] John D. Walker,[3] William Branham,[4] and Daniel M. Sheehan[4]

[1]R.O.W. Sciences, Inc., Jefferson, Arkansas, USA; [2]BASF Corporation, Princeton, New Jersey, USA; [3]TSCA Interagency Testing Committee, U.S. Environmental Protection Agency, Washington, DC, USA; [4]Division of Genetic and Reproductive Toxicology, National Center for Toxicological Research, Jefferson, Arkansas, USA

A number of environmental chemicals, by mimicking natural hormones, can disrupt endocrine function in experimental animals, wildlife, and humans. These chemicals, called "endocrine-disrupting chemicals" (EDCs), are such a scientific and public concern that screening and testing 58,000 chemicals for EDC activities is now statutorily mandated. Computational chemistry tools are important to biologists because they identify chemicals most important for *in vitro* and *in vivo* studies. Here we used a computational approach with integration of two rejection filters, a tree-based model, and three structural alerts to predict and prioritize estrogen receptor (ER) ligands. The models were developed using data for 232 structurally diverse chemicals (training set) with a 10^6 range of relative binding affinities (RBAs); we then validated the models by predicting ER RBAs for 463 chemicals that had ER activity data (testing set). The integrated model gave a lower false negative rate than any single component for both training and testing sets. When the integrated model was applied to approximately 58,000 potential EDCs, 80% (~46,000 chemicals) were predicted to have negligible potential (log RBA < –4.5, with log RBA = 2.0 for estradiol) to bind ER. The ability to process large numbers of chemicals to predict inactivity for ER binding and to categorically prioritize the remainder provides one biologic measure to prioritize chemicals for entry into more expensive assays (most chemicals have no biologic data of any kind). The general approach for predicting ER binding reported here may be applied to other receptors and/or reversible binding mechanisms involved in endocrine disruption. *Key words:* endocrine-disrupting chemicals, estrogen receptor binding, relative binding affinties, risk assessment, structural alerts, tree-based models. *Environ Health Perspect* 110:29–36 (2002). [Online 10 December 2001] *http://ehpnet1.niehs.nih.gov/docs/2002/110p29-36hong/abstract.html*

Concern is growing among the scientific community, government regulators, and the public that endocrine-disrupting chemicals (EDCs) in the environment are adversely affecting human and wildlife health by disrupting endocrine function (*1,2*). Adverse outcomes have been observed in experimental animals and wildlife; potential effects in humans include reproductive and developmental toxicity, carcinogenesis, immunotoxicity, and neurotoxicity (*3*). EDCs may exert adverse effects through a variety of mechanisms, such as estrogen receptor (ER)-mediated mechanisms of toxicity.

The scientific debate surrounding EDCs has grown contentious, partly because some suspected EDCs are economically important chemicals, high in production volume. The public and regulatory concerns led to government regulatory actions and expanded research across Europe, Japan, and North America (*4,5*). In response to congressional action, the U.S. Environmental Protection Agency (EPA) established the Endocrine Disruptor Screening and Testing Advisory Committee (EDSTAC); EDSTAC recommended a plan to screen and test for estrogenic, androgenic, and thyroid end points for a large number of chemicals. To accomplish this, chemicals will be screened (tier 1) using a multiple end point strategy that includes more than 20 different *in vitro* and *in vivo* assays recommended by EDSTAC (*6*). Although more than 87,000 chemicals were initially selected for evaluation, many were polymers or otherwise unlikely to bind to steroid receptors, leaving about 58,000 chemicals for evaluation in tier 1. The number that will progress to the testing step (tier 2) (*7*) is not known. Processing chemicals through both tiers will require many years and extensive resources. Hence, the U.S. EPA has adopted an approach requiring priorities to be set before tier 1. Priority setting will use currently available information, such as production volume, human exposure, environmental fate and persistence, and biologic data. Priority setting rank-orders the most important chemicals for more resource-intensive and costly tier 1 evaluations.

Several types of hormonal activities, including estrogenic, androgenic, and thyroidal, are believed to contribute to endocrine disruption (*3*). Endocrine disruption can result from a variety of biologic mechanisms that interfere with these activities. Receptor binding is a major mechanism of toxicity for estrogens. Rapid methods for characterizing ER binding activity are important. Such methods should generate a small fraction of false negatives (chemicals predicted not to bind to their receptor, but which actually bind). False negatives constitute a crucial error because they will receive a relatively lower priority for evaluation in tier 1 and may remain in use for many years. Furthermore, the methods should provide reasonable quantitative accuracy for true positives, because those with higher affinities will generally be of higher priority. Computational methods, including structure–activity relationships (SARs), can predict receptor binding and therefore can be used to evaluate untested chemicals to provide biologic data for use in priority setting (*8–12*).

The first step in developing SAR models is acquisition of a training set of chemicals that have known activities. Information derived from structure of individual chemicals in the training set, such as hydrophobicity, structural fragments associated with activity (structural alerts), charged surface area, and so on, is called descriptors. Descriptors are then evaluated for their ability to predict the activity of the training set (model construction) and of other chemical data sets not used in the training set, but which have known activities (the testing set). This latter step is called external validation. With adequately validated performance, such models can be used to predict activities of untested potential EDCs. Numerous computational methods can be used to develop SAR models. The choice of methods depends on the nature of the application and the available data. For example, pharmacophores (three-dimensional substructures of active chemicals) are of great importance in drug discovery to generate potential candidates by rapidly searching large structure databases for chemicals with similar structures (*13*). Quantitative structure–activity relationships (QSARs) are used widely to correlate changes in biologic activity among chemicals with

Address correspondence to W. Tong, R.O.W. Sciences, Inc., 3900 NCTR Road, MC 910, Jefferson, AR 72079 USA. Telephone: (870) 543-7142. Fax: (870) 543-7382. E-mail: Wtong@nctr.fda.gov

This research was funded under an interagency agreement between the U.S. Environmental Protection Agency and the U.S. Food and Drug Administration's National Center for Toxicological Research. We gratefully acknowledge the American Chemistry Council and the FDA's Office of Women's Health for partial financial support.

Received 28 March 2001; accepted 12 June 2001.

changes in their descriptors (*14–16*). We have developed several pharmacophore and QSAR models (*17–20*) to predict the binding affinity of chemicals to the ER. Many of these models, together with those reported here, have been integrated into a four-phase system (*21*) that is an efficient tool in setting priorities for potential estrogenic EDCs.

In this study, we used a tree-based model and three structural alerts to evaluate 58,000 chemicals and predict those that have negligible potential to bind to the ER. The tree-based approach used a set of IF-THEN rules based on descriptors to determine a chemical's potential to bind to the ER. The depiction of the results provides a tree with a binary branching. The IF-THEN rule can look like this: If a chemical is steroidlike, then it goes to branch A. If not, it goes to branch B. Structural alerts are two-dimensional (2D) structural features that exist in most active chemicals. Before we applied the tree-based model and structural alerts to large data sets, we used several rejection filters to reduce the number of chemicals in the data set. Extensive validation has demonstrated that the rejection filters eliminated no active chemicals, even very weak binders. We found that the integrated combination of rejection filters, structural alerts, and tree-based models can be used to prioritize, with low false negative rates, the 58,000 chemicals under consideration.

Material and Methods

Data sets. In this study, we first constructed models using a training data set, and then validated them using a testing data set. Once the models were validated, they were applied to a real-world data set—a target data set for priority setting purposes.

Training data set. The reliability and predictability of a computational model not only depends on the computational method but varies significantly with the quality of the training data set. The training data set is used to generate valid rules and to guide decision making. To build a robust and predictive computational model for EDCs, it is important to have a training set of chemicals with broad structural diversity and an accurate and reproducible measure of biologic activity over a wide range. From a literature survey, we found that data may vary according to interlaboratory protocol and technique differences. Therefore, we established an in-house rat ER binding assay to provide data for model development, which were reported by Blair et al. (*22*) and Branham et al. (*23*).

The training data set was designed to reflect the structural diversity of the ER ligands and a wide distribution of ER binding affinities necessary for a robust model (*19*). Our training data set (National Center for Toxicological Research [NCTR] data set) of 232 chemicals has ER relative binding affinities (RBAs) that range over 10^6-fold; the RBA value for the endogenous ER ligand, 17β-estradiol (E_2), was set to 100. This NCTR data set has been used extensively to build and validate a series of computational models (*20,21*) proposed for priority setting. The cut-off log RBA value to distinguish ER binders from non-ER binders is set to –4.5, which is our lowest experimental resolution in the ER binding assay. The 131 ER binders have log RBAs ≥ –4.5, whereas 101 non-ER binders have log RBAs < –4.5.

Test data set. A computational model, once built, should be validated using external data sets to assess the potential rate of false positives and false negatives before it is applied to the target data set (a data set with unknown activity value). An ideal test data set should be directly related to the real problem in question. For this particular application, we developed the model to predict ER RBAs of 58,000 environmental chemicals, mostly pesticides and industrial chemicals. On the basis of this consideration, we selected a data set reported by Nishihara et al. (*24*) as a test data set. This data set contains 517 chemicals tested with the yeast two-hybrid assay, of which > 86% are pesticides and industrial chemicals. We used only 463 chemicals for this study after eliminating the ones that lacked unique structures, such as mixtures. Only 62 chemicals were categorized as active based on activity > 10% of 10^{-7}M E_2, as defined by Nishihara et al. (*24*). The majority of the chemicals were inactive, which is similar to the real-world situation where inactive chemicals should be a large portion in the target data set.

Target data set. Walker et al. (*25*) developed a database that contains a large and diverse collection of known pesticides and industrial chemicals as well as some food additives and drugs. The database contains 92,964 Chemical Abstract Service (CAS) registry numbers of chemicals that will probably have to be evaluated for their potential endocrine disruption. A final data set of 58,230 chemicals was used for this study after eliminating 34,573 chemicals for which structures were not available (*25*) and 161 chemicals for which three-dimensional structures could not be generated. The molecular structures of these chemicals were processed according to the following criteria (*26*): The records are valid (i.e., they contain completed structural information and there are no obvious errors in the structural description); counterions and solvent molecules were removed to obtain single structure records; and charges on acidic and basic groups were neutralized by adding or removing protons. This prevented structural differences caused by different protonation states, which might lead to differences in the calculation of the molecular descriptors.

Integrated system. The overview of the integrated system, based on rejection filters, a tree-based model, and three structural alerts, is shown in Figure 1. First, we used two rejection filters to eliminate nonbinders. Then we used three structural alerts and a tree-based model separately to predict ER binding activities of chemicals passing the rejection filters. Chemicals predicted to be active by the tree-based model or any one of the structural alerts were identified as potential ER binders. We desired that a large number of non-ER binders would be eliminated in this process, so that only a small fraction of the chemicals in the original data set would be ER binders that would need evaluation experimentally or be predicted using more precise models, such as QSARs (*20*).

Rejection filters. We used two rejection filters and excluded chemicals that matched any of these two filters (Figure 1). The first rejection filter is a molecular weight range, which was set to < 94 or > 1,000. The molecular weight of phenol, 94, was considered the lowest limit for a chemical to bind to ER, whereas a molecular weight of 1,000 was considered the upper limit as suggested by EDSTAC. The second rejection filter requires that an ER binder contain at least one ring structure of any size. This structural rejection filter is developed based on the fact that, from a large survey (*27*), there are no known estrogens lacking a ring.

Structural alerts. Structural alerts are key 2D structural fragments associated with ER

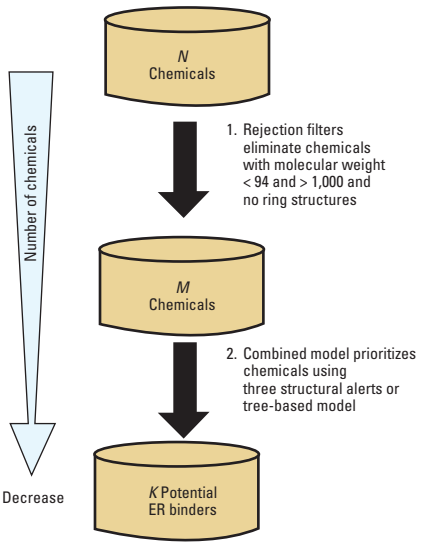

Figure 1. Overview of the integrated system that consists of two rejection filters, three structural alerts, and a tree-based model.

binding. Figure 2 depicts the three structural alerts [i.e., the steroid diethylstilbestrol (DES) and phenolic skeletons]. Chemicals containing any of these structural alerts were considered to be potential ER binders. We selected these structural alerts through careful SAR examination of a large number of chemicals with known binding affinities to the ER (*27*) in conjunction with knowledge of the recently reported ligand–ER crystal structures (*28,29*).

Tree-based model. Classifying or partitioning chemicals using the tree-based approach is based on the similar property principle, which states that structurally similar chemicals exhibit similar biologic properties (*30*). The measure of similarity between chemicals depends on the use of molecular descriptors calculated from their structures. Effective classification of chemicals depends critically on the nature of the molecular descriptors used. Thus, the tree-based model development consisted of two steps: selecting several descriptors using the genetic function approximation (GFA) method, and using these descriptors to construct a tree-based model.

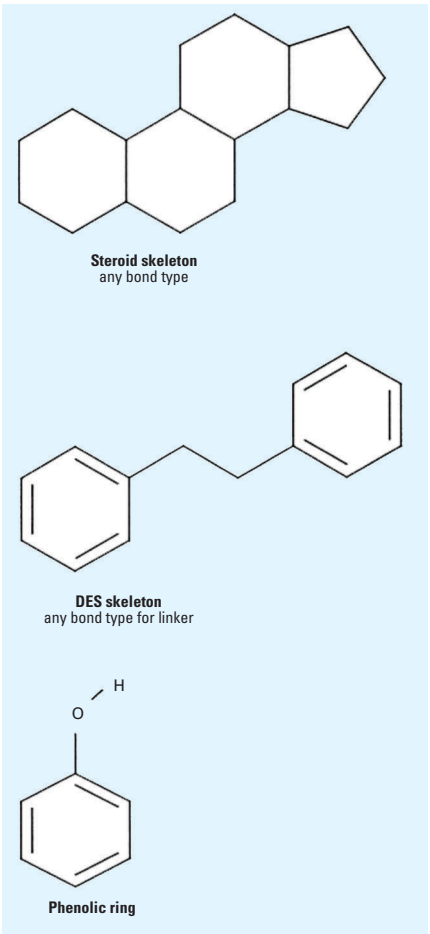

Figure 2. Three structural alerts. Chemicals containing any of these alerts are predicted to be active.

For the descriptor selection based on GFA, we investigated about 150 molecular descriptors. These descriptors cover a variety of structural information, including the conformational, electronic, information content, quantum mechanical, spatial, thermodynamic, and topologic nature of a structure. We selected an optimal, biologically relevant subset of descriptors for the tree-based model using GFA. GFA is a genetic algorithm-based statistical approach (*31,32*), which has been widely used for QSAR model development. Basically, GFA starts with a randomly selected set of descriptors from the original descriptor pool to generate a population of QSAR equations (100 equations in this study) using multivariate regression techniques such as the least-square regression method used in this study. Then the quality of each individual equation is estimated using a lack-of-fit (LOF) score function (*31,32*). These equations can be rank-ordered based on the model's quality. From evolution, parents (parent QSAR models) with good genes (descriptors) often produce better children (offspring QSAR models). Therefore, the two best QSAR models are chosen to "mate" and propagate their "genes" to offspring through the crossover operation, in which portions of the descriptors are taken from each parent QSAR equation and recombined to create the child. A good offspring QSAR model replaces the worst model in the equation pool. Mating needs to be repeated many times (20,000 times in this study) until there is no significant improvement in the model; at this time, good combinations of descriptors are discovered and spread through the population of QSAR models. It is reasonable to assume that the descriptors used more by QSAR models in the population should be more biologically relevant to the end point investigated. Therefore, we selected the top 10 descriptors that appeared most frequently in the QSAR model population for the tree-based model development.

A classification tree is the collection of many production rules, expressed as premise and conclusion (in the form IF ... THEN ...) and displayed in the form of a tree containing only binary branching. For example, a simple rule could be "if molecular weight > 300, then the chemicals are active." A tree-based model provides an alternative to linear and additive models for regression problems and to linear and additive logistic models for classification problems. Because tree-based model constructions are recursive in nature, they are also referred to as the recursive partitioning method (*33*) for pattern recognition in drug discovery. Depending on the nature of the activity data (end point), the tree can be constructed for either regression or classification. Each end node ("leaf of the tree") of a regression tree gives a quantitative prediction, whereas the classification tree gives categoric predictions. The classification tree is used most commonly in data analysis, where the end point is usually binomial (yes/no). In the present application, the tree-based model to classify chemicals into active and inactive ER RBA categories is implemented in S-plus software (*34*).

The development of a tree-based model consists of two steps, tree construction and pruning. In tree construction, a parent population is split into two offspring nodes that become parent populations for further splits. The splits are selected to maximally distinguish the response variable in the left and right nodes. Splitting continues until chemicals in each node are either in one activity category or cannot be split further to improve the model. To avoid overfitting of the training data, the tree needs to be cut down to a desired size using tree cost-complexity pruning. In this study, we tested all possible combinations in groups of three to six of these 10 descriptors selected by GFA in a combinatorial way to construct the tree-based models. We used the model giving the highest correct prediction rate for the NCTR data set for final application.

Validation of models. In this study, we assessed the quality of a model by several statistical measures, including false negative, false negative rate, false positive, false positive rate, and concordance. False positive is the number of chemicals predicted to be active by the model but inactive in the assay; thus the false positive rate is the false positives divided by the total number of inactive chemicals in the data set. In contrast, false negative is the number of chemicals predicted to be inactive by the model but active in the assay, and the false negative rate is the false negatives divided by the total number of active chemicals in the data set. Concordance is the overall agreement between the predicted and experimental results, positive and negative. The same criteria were also used to measure the predictivity of the model for the test data set.

General computations. We created and maintained chemical structures for the NCTR data set, the Nishihara et al. (*24*) data set, and the Walker et al. (*25*) data set using Molecular Design Limited (MDL) Information Systems' Integrated Scientific Information System (ISIS/Base 2.2.1) software (MDL Information System, San Leandro, CA) running on a personal computer. We identified three structural alerts for chemicals using the ISIS/Base software. The tree-based models were constructed using S-Plus (MathSoft, Inc., Cambridge, MA) software. Descriptor generation and

GFA analysis were performed using Cerius[2] (Molecular Simulation Inc., San Diego, CA).

Preprocessing and rejection filtering for the Walker data set were conducted on an SGI workstation using an in-house program developed in programming language C. The program checks and corrects the connection table of a structure in the sdf file format exported from ISIS Base, removes solvent molecules, and changes charge states by adding or removing protons if necessary. Then the program calculates molecular weight and counts the number of rings in a structure using a valid algorithm (*35*).

Results

The approach for the prediction of ER binding using several computational models consisted of two steps (Figure 1). First, we applied two rejection filters to eliminate inactive chemicals, and then used a tree-based model combined with three structural alerts to identify potential ER binders. The approach was developed based on the NCTR data set and validated by the Nishihara et al. (*24*) data set. We then used this system to identify potential ER binders in 58,000 chemicals (the Walker data set).

Rejection Filters

We investigated various parameters for potential rejection filters. The criteria used to select the filters are that a valid and efficient rejection filter should not generate any false negatives and should be able to significantly reduce the number of chemicals for experimental evaluation.

Solubility and permeability are important determinants for a chemical to bind to its target protein. Poor absorption or permeation of a chemical can greatly limit drug activity, even if it has good binding activity. Lipinski et al. (*36*) investigated these characteristics experimentally and computationally in drug design and development. They found that chemicals with poor absorption or permeation are more likely to have one of the following criteria in the structure: *a*) more than 5 hydrogen bond donors, *b*) 10 hydrogen bond acceptors, *c*) molecular weight > 500, and *d*) log*P* is > 5 (*36*). This is Lipinski's "rule of five," which has proven very useful in eliminating nondruglike chemicals in the early stage of drug discovery. The name does not imply that there are five rules; rather it is derived from the fact that the criteria in each rule are a numeric multiple of five. However, in applying these rules as rejection filters for the NCTR data set, we rejected 33 active chemicals by these rules (false negative rate = 25%), of which 15 chemicals actually were strong ER ligands (log RBA > 0). These rules might be useful for drug discovery purposes to identify potential but not all possible leads. However, for screening purposes in a regulatory context, all possible active chemicals, even potentially very weak ones, must be identified. This explains why false negatives are of great concern: It may be many years before the lowest priority chemicals go through screening and testing steps. Therefore, Lipinski's rule of five was not useful for predicting ER binding. However, because the rule of five includes criteria useful for *in vivo* activity, it may be used to improve predictions, in conjunction with binding data, for *in vivo* activity.

Two rejection filters, the molecular weight range and ring structure, met the criteria of not generating false negatives and being able to reduce the size of the data set. Chemicals matching any one of these two filters were excluded from subsequent models. As shown in Table 1, these two rejection filters correctly eliminated six inactive chemicals from the NCTR data set and 98 from the Nishihara data set, respectively. No false negatives were introduced using these two rejection filters. The sizes of the Nishihara and Walker data sets were reduced about 21% and 29%, respectively. This demonstrated that, for real-world applications, these two rejection filters might significantly reduce the number of chemicals for further evaluation with a minimum risk of introducing false negatives.

Structural Alerts

We used three structural alerts to identify potential ER binders. Each alert independently characterized the unique structural features important for ER binding. We found that the length and breadth of both the steroid skeleton and DES skeletons were filled well into the receptor-binding pocket, as illustrated in Figure 3. In addition, although most endogenous hormones contain the steroid skeleton, most strong estrogens have two benzene rings separated by two carbon atoms (*37*). It has been long understood that the phenolic ring is often associated with estrogenic activity (*38*). The contribution of the phenolic ring in binding is much more significant than any other structural feature (*27*). By overlaying the crystal structures of four ligand–ER complexes (E_2–ER, 4-hydroxytamoxifen–ER, raloxifene–ER, and DES–ER complexes) based on their common protein residues at the binding site, Shi et al. (*20*) found that the phenolic rings of all four ligands are closely positioned at the same location to allow hydrogen bond interactions with Glu 353, Arg 394 of the receptor, and a water molecule.

When we applied these structural alerts to the training and testing data sets, most active chemicals contained one of the structural alerts. The results of structural alerts searching on these data sets are summarized in Table 2. Of 131 active chemicals in the NCTR data set, 110 (84%) of the chemicals contained the phenolic ring, 30 (23%) contained the DES skeleton, and 22 (17%) contained the steroid skeleton. A total of 95% (124/131) of the active chemicals matched one or more of these structural alerts. For the Nishihara data set, about 90% (56/62) of the active chemicals were identified by these three structural alerts.

Tree-Based Model

The tree-based model classifies chemicals into active and inactive classes using a series

Table 1. Results of two rejection filters for the NCTR, Nishihara et al. (*24*), and Walker et al. (*25*) data sets.

Data sets	Data size	Eliminated by MW		Eliminated by ring		The number (%) of eliminated chemicals
		Active	Inactive	Active	Inactive	
NCTR	232	0	0	0	6	6 (2.6%)
Nishihara et al.	463	0	28	0	89	98 (21.2%)
Walker et al.	58,230	16,048		1,495		16,689 (28.7%)

This table lists the number of chemicals eliminated by either molecular weight (MW) range or lack of ring criteria as well as their combination. No active chemicals were rejected by these two filters.

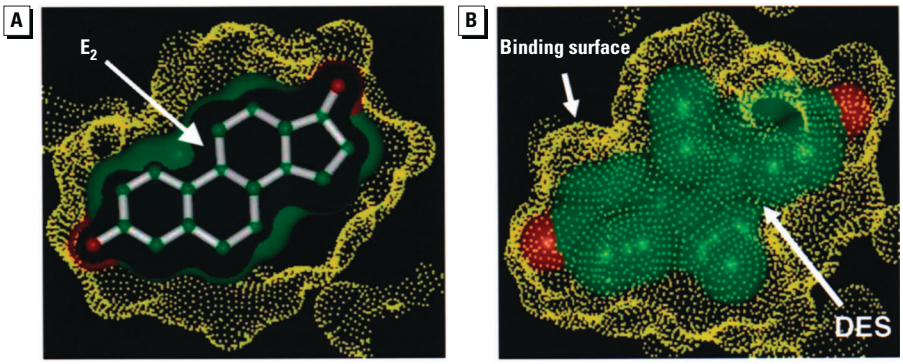

Figure 3. The surface of the ER binding site (yellow dots) bound with (*A*) E_2 and (*B*) DES.

of rules on the basis of descriptors. The descriptors characterize the structural similarities among chemicals with respect to the biologic activity modeled. Therefore, the critical task for model development is to evaluate a large number of possible molecular descriptors and identify the ones most related to the known active chemicals. We first calculated 153 molecular descriptors, of which we selected the top 10 using the GFA approach. Then we developed several tree-based models for the NCTR data set using combined groups of three to six of the top 10 descriptors. The model giving the best concordance was the final model. The final tree-based model consisted of five descriptors: phenolic ring index, log*P*, Jurs-PNSA-2, Jurs-RPCS, and shadow-XY fraction. The phenolic ring index indicates the presence or absence of the phenolic group in a chemical. The log*P* measures hydrophobicity of a chemical (*39*). The Jurs-PNSA-2 and Jurs-RPCS characterize the positive charged surface area of a molecule by combining molecular shape and electronic information (*40*). The shadow-XY fraction is a geometric descriptor related to the breadth of a molecule (*41*). Each descriptor encodes important structural characteristics for ER binding. For example, the phenolic group is considered the most important structural feature for ER binding (*38*). Our recent SAR study on a large number of xenoestrogens demonstrated that substitution at 7α and 11β position of E_2 enhanced ER binding by increasing the breadth of a chemical (*27*). The hydrophobicity and charged surface area are, in principle, critical for all receptor-binding systems (*36*).

The tree-based model using five optimal descriptors is summarized in Figure 4. The model identified the phenolic ring index as the most important descriptor for ER binding. If chemicals contained a phenolic moiety but also had log*P* values >1.49, they were more likely to be ER binders. In contrast, chemicals without a phenolic moiety were less likely to be ER binders unless they had relatively larger hydrophobicity (log*P*), charged surface area (Jurs-PNSA-2 and Jurs-RPCS), and breadth of the structure (shadow-XY).

Table 3 summarizes the results of the tree-based model on the training and testing data sets. The model had a concordance of about 88% for the NCTR data set. Of the 131 ER binders, 123 were correctly predicted to be active. Of the 101 non-ER binders, 81 were correctly predicted to be inactive. The false positive and false negative rates (Table 4) were 19.8% (20/101) and 6.1% (8/131), respectively. With the model applied to the Nishihara data set, the concordance was 82.5%, which is slightly lower than that for the training data set.

Table 2. Results of structural alerts for the training and testing data sets

Data sets	Data size	Active	Phenol ring		DES skeleton		Steriod skeleton		Any alerts	
			Hit[a]	Active[b]	Hit	Active	Hit	Active	Hit	Active
NCTR	232	131	131	110	34	30	31	22	157	124
Nishihara et al. (*24*)	463	62	120	54	17	8	11	7	131	56

[a]Hit, the number of chemicals containing the structural alert. [b]Active, the number of hit chemicals active in the assay.

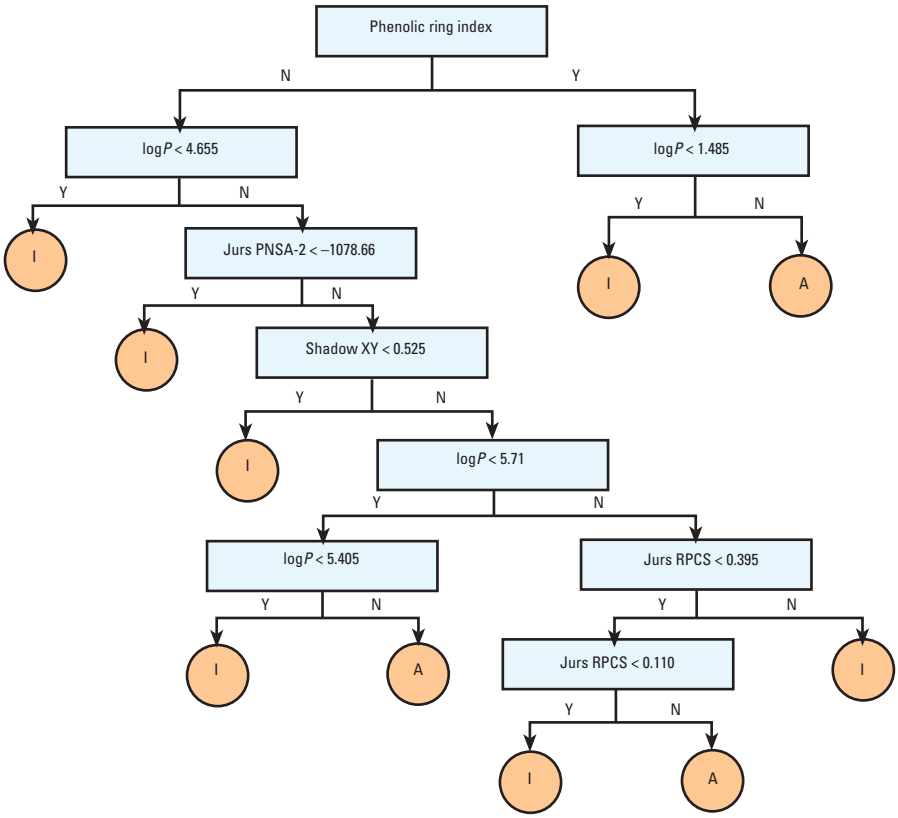

Figure 4. Tree-based model. The model displays a series of yes/no (Y/N) rules to classify chemicals into active (A) and inactive (I) categories based on five descriptors: phenolic ring index, log*P*, Jurs PNSA-2, shadow-XY, and Jurs RPCS. The squares represent the rules; the circle represents the categoric results.

Combination of the Tree-Based Model with the Structural Alerts

The tree-based model and structural alerts can independently identify most active chemicals with 6.1% and 4.6% false negative rates for the NCTR data set, and 12.9% and 9.7% for the Nishihara data set. Even though both models could be used to identify independently the potential ER binders for a variety of applications, it was desirable to reduce further the false negative rate for regulatory application. Thus, we studied a combination of the tree-based model with the three structural alerts for potential priority-setting applications. Chemicals predicted to be active by any of these models were considered to be ER binders. The combined model results for the training and testing data sets are summarized in Table 4. The combined models produced only 2.3% and 6.5% false negative rates for the NCTR and Nishihara data sets, respectively, which were nearly half of the false negative rates observed using individual models. It might be expected that the combined models would produce an increase in the false positive rate. Even though a low false negative rate is of great importance, a lower false positive rate is always desirable from an economic perspective. For the NCTR data set, the tree-based model and structural alerts yielded 19.8% and 34.7% false positive rates, and we observed only a slightly higher false positive rate (37.6%) for the combined model. We also observed similar results for the Nishihara data set, in which the false positive rate for the combined model, the tree model, and structural alerts was 21.2%, 18.2%, and 17.7%, respectively. This demonstrated that the combined model could significantly reduce the false negatives without significantly increasing screening costs resulting from an increased false positive rate.

It is worthwhile to point out that both the tree-based model and three structural alerts provided only a yes/no prediction. However, chemical potencies for *in vivo* uterotrophic response are significantly higher for high-affinity chemicals than for low-affinity chemicals. It is desirable that the strength of ER binding activity is ranked for chemicals predicted by the models because this would allow selection of the highest affinity-binding chemicals for testing. The tree-based model and three structural alerts not only identified different structural attributes associated with the ER binding activity, but also used different approaches to establish the chemical structure–ER binding activity relationship. It would be expected that the number of models predicting a chemical to be active should increase in direct proportion to its actual activity. In other words, chemicals can be ranked based on the number of positive predictions given by the three structural alerts and the tree-based model in an additive manner. As shown in Table 5, chemicals predicted to be active by none or only one of four models (three structural alerts and the tree-based model) were more likely to be inactive. In contrast, chemicals predicted to be active by more than two models were more likely to be ER binders. Specifically, if chemicals were assigned to be active by more than three models, they were likely to be in the highest activity range (more than 100,000-fold below E_2).

Priority Setting of 58,000 Chemicals

We applied the integrated system to prioritize the Walker data set, and the results are summarized in Tables 6 and 7. Of 58,230 chemicals in the data set, 16,689 chemicals were eliminated as inactive by the two rejection filters. The resulting 41,541 chemicals were predicted by the combined model, of which 34,638 chemicals were predicted to be inactive, and 6,903 chemicals were predicted to be active by at least 1 of 4 models in the combined approach. Through this process, it might be anticipated that less than 12% of the original chemicals might need to be tested for their potential ER activity. Table 7 shows that of 6,903 chemicals, only 104 chemicals were predicted to be active by more than 3 models—that is, in the most active category.

Discussion

The objective of priority setting is to rank a large number of chemicals for experimental evaluation from most important to least important. A number of criteria can be used for this purpose, such as production volume, persistence and fate in the environment, human exposure levels, and so on. Most of the 58,000 chemicals required for assay have no biologic data. Computer-based ER-binding prediction models can supply one element of such missing data to be used in priority setting.

We integrated a tree-based model with structural alerts and rejection filters for priority setting a large number of chemicals based on estimation of their logRBA range. The system reduced the number of environmental chemicals for assay by about 88%, with a minimum false negative rate. Tests on three data sets indicated that this integrated model is acceptable for priority setting of estrogenic EDCs. An important advantage of this system for priority setting is its efficiency of scale when applied to a large number of chemicals. When several end points are analyzed simultaneously, the efficiency of scale of computation is even more pronounced. The results from this integrated model, together with information on exposure level, production volume, and environmental persistence of chemicals, may be sufficient for prioritizing potential estrogens. A similar procedure appears appropriate for androgens and thyroids.

Computer-based priority setting is widely applied in drug discovery to identify potential drugs. The computational approaches used in the drug discovery process include models from the simple Lipinski's rule of five to classification/clustering and QSARs. The purpose of priority setting in drug discovery is to identify a few lead chemicals; sometimes even one good lead chemical is sufficient if it can be developed into a drug. It is not necessary to discover or design all possible lead chemicals. Thus, relatively high false negatives are tolerable, but false positives need to be low. Either the structural alerts or the tree model reported here might be good enough for such applications. However, prediction of ER binding affinity for chemicals as an element in priority setting requires a minimum false negative rate, because these chemicals will receive the lowest priority for entry into screening and testing steps. Results here show that the process we have designed can significantly reduce the number of chemicals for experimental evaluation with minimum false negative rate. Moreover, the system can rank

Table 3. Results of the tree-based model for the training and testing data sets.

Data set	Category	Chemicals	Predicted active	Predicted inactive	Percent concordance
NCTR	Active	131	123	8	87.9
	Inactive	101	20	81	
Nishihara et al.	Active	62	54	8	82.5
(*24*)	Inactive	401	73	328	

Table 4. False positive and false negative rates of the tree-based model, structural alerts, and combined model for the training and testing data sets.

Data sets	Error rates (%)	Tree-based model	Structural alerts	Combined model
NCTR	False positive	19.8	34.7	37.6
	False negative	6.1	4.6	2.3
Nishihara et al.	False positive	18.2	17.7	21.2
(*24*)	False negative	12.9	9.7	6.5

Table 5. Ranking the training and testing data sets by adding the prediction results from four models (three structural alerts and the tree model).

		The number of chemicals predicted to be active by			
Data sets	Activity range[a]	No model	One model	Two models	Three models[b]
NCTR	≥ –3.0	2	5	57	36
	< –3.0	1	2	28	0
	Inactive	63	20	16	2
Nishihara et al.	≥ –3.0	1	0	24	10
(*24*)	< –3.0	3	5	17	2
	Inactive	313	27	58	3

[a]The three experimental activity categories were defined as follows: ≥ –3.0, chemicals with activities larger than 100,000-fold below E_2; < –3.0, chemicals with activities less than 100,000-fold below E_2, but active in the assay. Inactive, chemicals are below the assay detection limit. For the NCTR data set the log RBA values were used, whereas log RP (relative potency) values were used for the Nishihara et al. data set. In both cases, the activity value for E_2 was equal to 2. The –3.0 value selected as a cutoff was suggested by the U.S. EPA, which might be used to distinguish active from inactive. [b]Because it is unlikely for chemicals to contain both steroid and DES skeletons, the maximum number of models predicting chemicals to be active was three for the data sets studied.

Table 6. Size reduction for the Walker et. al (*25*) data set using the integrated system shown in Figure 1.

Process	No. of chemicals	Resulting data size (%)
Original data size	58,230	Not applicable
Eliminated by two rejection filters	16,689	41,541 (71.3)
Predicted to be inactive by the combined model	34,638	6,903 (11.9)

chemicals based on their predicted ER binding activity. This ranking method can provide useful ER RBA data for use in screening.

Of the variety of xenoestrogens whose structures have little apparent resemblance to E_2, *o,p′*-DDT and kepone are of particular interest for their environmental persistence and wide industrial application (*42,43*). Of six DDT congeners (*o,p′*- and *p,p′*-DDT, *o,p′*- and *p,p′*-DDE, and *o,p′*- and *p,p′*-DDD) assayed, only *o,p′*-DDT reasonably binds ER about 100,000-fold below E_2 (*22*). Kepone was about 10,000 times weaker than E_2 in its affinity for ER (*22*). In contrast, its analogue, mirex, which has the carbonyl group replaced by two chlorine atoms, is inactive in binding. The system reported here correctly classified all these chemicals, demonstrating its strength in identification of apparently weakly ER ligands.

To assess the reliability and applicability of the models, the Nishihara et al. (*24*) data set was used to validate the system. However, the yeast two-hybrid assay used for this data set differs from the NCTR RBA assay. The ER competitive binding assay measures the binding affinity of a chemical for ER, whereas the yeast two-hybrid assay measures ER binding-dependent transcriptional and translational activity. These two assays differ in their sensitivity in distinguishing active from inactive chemicals, particularly for weak estrogens and antiestrogens (*37*). We compared the assay results for 80 common chemicals from both the Nishihara et al. and NCTR data sets, of which inconsistent assay results were observed for 12 chemicals. Specifically, of 30 active chemicals in the Nishihara data set, one chemical was found inactive in the NCTR data set; of 50 inactive chemicals in the Nishihara data set, 11 chemicals were found active in the NCTR data set. These observations show that even using the experimental data from the ER binding assay (the NCTR data set) to predict the experimental results from the yeast two-hybrid assay (the Nishihara et al. data set), there may be about a 15.0% (12/80) discrepancy, or 3.3% (1/30) false negative rates and 22% (11/50) false positive rates. In comparison, the combined model produced only 19.2% wrong predictions, or 6.5% false negative rates and 21.2% false positive rates. This demonstrated that the discrepancy produced by the combined model developed with the ER binding data was comparable to that by just using ER binding data alone for prediction of activity associated with a more complicated biologic mechanism. The observed false positives and false negatives for the Nishihara et al. data set are partially a result of the discrepancy between the two assays.

Table 7. Ranking 6,903 chemicals predicted to be active by the system for the Walker et. al (*25*) data set.

No. of models predicted as active	No. of chemicals
1	4,763
2	2,036
3	104

Practically any predictive system, whether using experimental or computational approaches, will produce some degree of error. Decreasing false negatives by modifying the criteria normally increases cost due to increasing false positives. The combined model minimized the false negative rate but increased the false positive rate. Therefore, we applied two rejection filters to eliminate chemicals that were most unlikely to be estrogens before applying the combined model to minimize the false positive. For the Nishihara et al. data set, 96 chemicals were eliminated by the rejection filters, of which 93 were also predicted to be inactive but three chemicals were predicted to be active by the combined model. By applying the rejection filters at the front end, we eliminated these three chemicals before using the combined model. Thus, the false positive rate was further reduced. Applying the rejection filters on a data set is an easy and rapid process, which has significant advantages for very large chemical data sets, such as the Walker et al. (*25*) data set.

The integrated system does not contain a three-dimensional pharmacophore model. A pharmacophore is a set of structural features (e.g., hydrogen bond donor, hydrogen bond acceptor, hydrophobic center) with associated geometry needed for a chemical to exhibit a certain type of biologic activity. It is normally important for modeling receptor binding systems. The lack of a pharmacophore model might be why three active chemicals—2,4′-dichlorobiphenyl, chalcone, and doisynoestrol—were predicted to be inactive. Recently, we have developed several pharmacophore models that complement the tree-based model and three structural alerts to form a more robust prediction system (*44*).

In summary, the results presented in this study demonstrate that the integrated model (two rejection filters, a tree-based model, and three structural alerts) shows great promise to screen a larger number of environmental chemicals for further experimental evaluation of estrogenic endocrine disruption. Numerous mechanisms are involved in endocrine disruption, which can be modeled using a similar approach to the one proposed in this article. Currently, a large volume of androgen receptor binding data is being generated by our group and by the U.S. EPA. A similar practice is being conducted in our lab to develop an integrated approach for predicting androgenic activity. These androgen models, together with the ones reported in this article, should provide rich potency-based information for incorporation into the U.S. EPA's Endocrine Disruption Priority Setting Database version 2.

References and Notes

1. Hileman B. Environmental estrogens linked to reproductive abnormalities, cancer. Chem Eng News 72:19–23 (1994).
2. Hileman B. Hormone disrupter research expands. Chem Eng News 75:24–25 (1997).
3. Kavlock RJ, Daston GP, DeRosa C, Fenner-Crisp P, Gray LE, Kaattari S, Lucier G, Luster M, Mac MJ, Maczka C, et al. 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 (1996).
4. The Food Quality Protection Act of 1996. Public Law 104-170, 1996.
5. The Safe Drinking Water Act. Public Law 104–182, 1996.
6. Gray LE Jr. Tiered screening and testing strategy for xenoestrogens and antiandrogens. Toxicol Lett 102–103:677–680 (1998).
7. Patlak M. A testing deadline for endocrine disrupters. Environ Sci Technol 30:540A–544A (1996).
8. Walker JD, Brink RH. New cost-effective, computerized approaches to selecting chemicals for priority testing consideration. In: Aquatic Toxicology and Environmental Fate, vol 11 (Suter GW, Lewis MA, eds). Philadelphia:ASTM, 1989;507–536.
9. Walker JD. Chemical selection by the TSCA interagency testing committee: use of computerized substructure searching to identify chemical groups for health effects, chemical fate and ecological effects testing. Sci Total Environ 109-110:691–700 (1991).
10. Walker JD. Estimation methods used by the TSCA interagency testing committee to prioritize chemicals for testing: exposure and biological effects scoring and structure activity relationships. Toxicol Model 1:123–141 (1995).
11. Walker JD, Gray DA. Past and future strategies for sorting and ranking chemicals: applications to the 1998 drinking water contaminants list chemicals. In: Identifying Future Drinking Water Contaminants. Washington, DC:National Academy Press, 1999;51–102.
12. Walker JD, Gray DA. The substructure-based computerized chemical selection expert system (SuCCESES): providing chemical right-to-know (CRTK) information on potential actions of structurally-related chemical classes on the environment and human health. In: Handbook on Quantitative Structure Activity Relationships (QSARs) for Predicting Effects of Chemicals on Environmental-Human Health Interactions (Walker JD, ed). Pensacola, FL:SETAC Press, 2001.
13. Hong H, Neamati N, Wang S, Nicklaus MC, Mazumder A, Pommier Y, Burke TRJ, Zhao H, Milne GWA. Discovery of HIV-1 integrase inhibitors by pharmacophore searching. J Med Chem 40:930–936 (1997).
14. Tong W, Perkins R, Strelitz R, Collantes ER, Keenan S, Welsh WJ, Branham WS, Sheehan DM. Quantitative structure–activity relationships (QSARs) for estrogen binding to the estrogen receptor: predictions across species. Environ Health Perspect 105:1116–1124 (1997).
15. Tong W, Perkins R, Xing L, Welsh WJ, Sheehan DM. QSAR models for binding of estrogenic compounds to estrogen receptor alpha and beta subtypes. Endocrinology 138:4022–4025 (1997).
16. Tong W, Perkins R, Sheehan DM. Perspectives on three-dimensional quantitative structure-activity relationship (3D-QSAR)/comparative molecular field analysis (CoMFA) in determining estrogenic effects. Jpn Chem Today 2:50–57 (1999).
17. Tong W, Lowis DR, Perkins R, Chen Y, Welsh WJ, Goddette DW, Heritage TW, Sheehan DM. 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 (1998).
18. Xing L, Welsh WJ, Tong W, Perkins R, Sheehan DM. Comparison of estrogen receptor alpha and beta subtypes based on comparative molecular field analysis (CoMFA). SAR QSAR Environ Res 10:215–237 (1999).
19. Perkins R, Anson J, Blair R, Branham WS, Chen Y, Dial S,

Fang H, Hass BS, Jackson M, Lu M, et al. The endocrine disruptor knowledge base (EDKB), a prototype toxicological knowledge base for endocrine disrupting chemicals. In: Handbook on Quantitative Structure Activity Relationships (QSARs) for Predicting Chemical Endocrine Disruption Potentials (Walker JD, ed). Pensacola, FL:SETAC Press, 2001.

20. Shi LM, Fang H, Tong W, Wu J, Perkins R, Blair R, Branham W, Sheehan D. QSAR models using a large diverse set of estrogens. J Chem Inf Comput Sci 41:186–195 (2001).
21. Shi LM, Tong W, Fang H, Perkins R, Wu J, Tu M, Blair R, Branham W, Walker JD, Waller C, Sheehan D. An integrated "4-Phase" approach for setting endocrine disruption priorities - Phase I and II prediction of estrogen receptor binding affinity. SAR QSAR Environ Res (in press).
22. Blair R, Fang H, Branham WS, Hass B, Dial SL, Moland CL, Tong W, Shi L, Perkins R, Sheehan DM. Estrogen receptor relative binding affinities of 188 natural and xenochemicals: structural diversity of ligands. Toxicol Sci 54:138–153 (2000).
23. Branham WS, Dial SL, Moland CL, Hass B, Blair R, Fang H, Shi L, Tong W, Perkins R, Sheehan DM. Unpublished data.
24. Nishihara T, Nishikawa J, Kanayama T, Dakeyama F, Saito K, Imagawa M, Takatori S, Kitagawa Y, Hori S, Utsumi H. Estrogenic activities of 517 chemicals by yeast two-hybrid assay. J Health Sci 46:282–298 (2000).
25. Walker JD, Waller CW, Kane S. The endocrine disruption priority setting database (EDPSD): a tool to rapidly sort and prioritize chemicals for endocrine disruption screening and testing. In: Handbook on Quantitative Structure Activity Relationships (QSARs) for Predicting Chemical Endocrine Disruption Potentials (Walker JD, ed). Pensecola, FL:SETAC Press, 2001.
26. Sadowski J, Kubinyi H. A scoring scheme for discriminating between drugs and nondrugs. J Med Chem 41:3325–3329 (1998).
27. Fang H, Tong W, Shi L, Blair R, Perkins R, Branham WS, Dial SL, Moland CL, Sheehan DM. Structure activity relationship for a large diverse set of natural, synthetic and environmental chemicals. Chem Res Toxicol 14(3):280–294 (2001).
28. Brzozowski AM, Pike AC, Dauter Z, Hubbard RE, Bonn T, Engstrom O, Ohman L, Greene GL, Gustafsson JA, Carlquist M. Molecular basis of agonism and antagonism in the oestrogen receptor. Nature 389:753–758 (1997).
29. Shiau AK, Barstad D, Loria PM, Cheng L, Kushner PJ, Agard DA, Greene GL. The structural basis of estrogen receptor/coactivator recognition and the antagonism of this interaction by tamoxifen. Cell 95:927–937 (1998).
30. Johnson M, Maggiora GM. Concepts and Applications of Molecular Similarity. New York:Wiley, 1990.
31. Clark DE, Westhead DR. Evolutionary algorithms in computer-aided molecular design. J Comput-Aided Mol Des 10:337–358 (1996).
32. Forrest S. Genetic algorithms-principles of natural selection applied to computation. Science 261:872–878 (1993).
33. Hawkins DM, Young SS, Rusinko A. Analysis of large structure-activity data set using recursive partitioning. Quant Struct-Act Rel 16:296–302 (1997).
34. Clark LA, Pregibon D. Tree-based methods. In: Modern Applied Statistics with S-PLUS, Chapter 9 (Venables WN, Ripley BD, eds). New York:Springer, 1997;413–430.
35. Hong H, Xin X. ESSESA, An expert system for structure elucidation from spectral analysis. 2. A novel algorithm of perception of the linear independent smallest set of smallest rings. Anal Chim Acta 262:179–191 (1992).
36. Lipinski CA, Lombardo F, Dominy BW, Feeney PJ. Experimental and computational approaches to estimate solubility and permeability in drug discovery and development settings. Adv Drug Delivery Rev 23:3–25 (1997).
37. Fang H, Tong W, Perkins R, Soto AM, Prechtl NV, Sheehan DM. Quantitative comparison of *in vitro* assays for estrogenic activity. Environ Health Perspect 108:723–729 (2000).
38. Anstead GM, Carlson KE, Katzenellenbogen JA. The estradiol pharmacophore: ligand structure-estrogen receptor binding affinity relationships and a model for the receptor binding site. Steroids 62:268–303 (1997).
39. Leffler JE, Grunwald E. Rates and Equilibrium Constants of Organic Reaction. New York:John Wiley and Sons, 1963.
40. Stanton DT, Jurs PC. Development and use of charge partial surface area structural descriptors in computer-aided quantitative structure-property relationship studies. Anal Chem 62:2323–2329 (1990).
41. Rohrbaugh RH, Jurs PC. Descriptions of molecular shape applied in studies of structure/activity and structure/property relationships. Anal Chim Acta 199:99–109 (1987).
42. McKim JM Jr, Choudhuri S, Wilga PC, Madan A, Burns-Naas LA, Gallavan RH, Mast RW, Naas DJ, Parkinson A, Meeks RG. Induction of hepatic xenobiotic metabolizing enzymes in female Fischer-344 rats following repeated inhalation exposure to decamethylcyclopentasiloxane (d5). Toxicol Sci 50:10–19 (1999).
43. Guzelian PS. Comparative toxicology of chlordecone (kepone) in humans and experimental animals. Annu Rev Pharmacol Toxicol 22:89–113 (1982).
44. Tong W, Perkins R, Wu J, Shi L, Tu M, Fang H, Blair R, Branham W, Sheehan DM. An integrated computational approach for prioritizing potential estrogenic endocrine disruptors. Proceedings of the International Symposium on Environmental Endocrine Disruptors, 8–9 December 1999, Kobe, Japan.