Estrogenic Activity of Anthraquinone Derivatives: In Vitro and In Silico Studies

Fei Li,^{†,§} Xuehua Li,[†] Jianping Shao,[†] Ping Chi,[†] Jingwen Chen,^{*,†} and Zijian Wang^{*,‡}

Key Laboratory of Industrial Ecology and Environmental Engineering (MOE), School of Environmental Science and Technology, Dalian University of Technology, Linggong Road 2, Dalian 116024, P. R. China, and State Key Laboratory of Environmental Aquatic Chemistry, Research Center for Eco-Environmental Sciences, Chinese Academy of Sciences, P.O. Box 2871, Beijing 100085, P. R. China

Received March 29, 2010

Comprehension of the ligand—receptor interactions is a prerequisite for constructing mechanism based quantitative structure—activity relationship (QSAR) models on xenoestrogenic activity. Molecular docking was performed to simulate the interactions between anthraquinone derivative (AQs) molecules and the estrogen receptor α (ER α). Hydrogen bonding, hydrophobic, and π - π interactions were found to be the dominant interactions between AQs and the receptor, which implied the estrogenic activities of the compounds. The recombinant yeast-based assay was employed to determine the estrogenic activities of 20 AQs. On the basis of the observed interactions between the AQs and ER α , appropriate molecular structural parameters were computed to develop a QSAR model. The polarizability term, the binding energy, the average molecular polarizability, the most negative formal charge in the molecule, and the average of the negative potentials on the molecular surface were significant parameters explaining the estrogenicity. The developed QSAR model had good robustness, predictive ability, and mechanism interpretability. The interactions between the AQs and ER α and the partition ability of the AQs into the biophase are main factors governing the estrogenic activities. Moreover, the applicability domain of the model was described.

1. Introduction

Unregulated estrogen receptor (ER) activation by xenoestrogens has been shown to disrupt the endocrine homeostasis of humans by mimicking the effects of the natural estrogen 17β -estradiol (E₂) (I). Xenoestrogens exert most of their biological effects via interactions with the estrogen receptor α (ER α ¹) (2), which is the predominant ER expressed in the uterus, kidney, and ovarian theca cells (3).

In recent years, quite a few test methods have been developed to screen xenoestrogens, including the recombinant yeast-based assay (4), the proliferation of the human breast cancer cell (MCF-7) (5), and the rat uterotrophic assay (6). For the recombinant yeast-based assay, the estrogenic responsiveness is primarily due to the ligand binding properties of the ER α (7). The assay is quick, convenient, cost-effective, and can provide valuable insights into the mechanisms of action but is restricted in its capacity to mimic whole animal metabolism and distribution (8). Many previous studies (9, 10) indicate that the yeast-based assay serves as a useful tool for screening hormone-like chemicals. The results determined by the yeast-

based assay agree with those from other estrogen-responsive assays, such as receptor-binding and cell proliferation assays (9, 11).

As the number of exhaustive xenoestrogens is much higher than that we could test, it is necessary to develop computational methods (in silico) to screen and predict the estrogenicity. On the basis of the relative binding affinities of compounds to the ER, several quantitative structure—activity relationships (QSARs) models have been established (12-14). According to the Organisation for Economic Co-operation and Development (OECD) and the U.S. Environmental Protection Agency (US EPA) (15), QSARs are promising tools for modeling and predicting the estrogenic activities of xenoestrogens. To facilitate the consideration of OSARs for regulatory purposes, the OECD issued guidelines for QSARs development and validation. Generally, QSAR models should be associated with the following information: a defined end point; an unambiguous algorithm; a defined domain of applicability; appropriate measures of goodnessof-fit, robustness and predictive power; and a mechanistic interpretation, if possible (16). The mechanistic interpretation is of paramount importance as QSAR models with clear mechanisms have high credibility, succinctness, and a definite boundary of the applicability domain. Thus, the molecular structural descriptors selected for constructing QSAR models should be based on mechanism analysis and facilitate the mechanism of interpretation.

To develop QSARs on xenoestrogens, it is necessary to understand the mechanisms of interaction between the xenobiotics and ER. Molecular docking was employed in some previous studies to observe the interactions between ligands and receptors (17, 18). For example, Celik et al. found that some polychlorinated biphenyls (PCBs), plasticizers, and pesticides

^{*} To whom correspondence should be addressed. (J.C.) Tel/Fax: +86-411-8470 6269. E-mail: jwchen@dlut.edu.cn. (Z.W.) Tel: +86-10-6284 9140. E-mail: wangzj@rcees.ac.cn.

[†] Dalian University of Technology.

^{*} Chinese Academy of Sciences.

[§] Current address: Yantai Institute of Coastal Zone Research, Chinese Academy of Sciences, Yantai, 264003, P. R. China.

¹ Abbreviations: AQs, anthraquinone derivatives; ERα, estrogen receptor α; E_{binding} , binding energy; QSARs, quantitative structure—activity relationships; PLS, partial least-squares; AD, applicability domain; R^2 , squared correlation coefficient; Q^2_{CUM} , the fraction of the total variation of the dependent variables that can be predicted by all of the extracted components; SE, standard error.

could bind in the steroid binding cavity, interacting with at least one of the two hydrophilic ends of the steroid binding site (18).

Anthraquinone derivatives (AQs) were selected in this study. The AQ dyes are the second largest class of dyes (only next to azo-dyes), and it has been estimated that 20% of the manufacturing and textile dyeing wastewater that may contain AQs is released to the environment (19). Previous studies suggest AQs may have estrogenic activities. For example, Kurihara et al. found that 2-hydroxyanthraquinone exhibited estrogenic activity (20). Matsuda et al. (21) found that alizalin and 2,6-dihydroxyanthraquinone as well as emodin with 2- and/or 6-hydroxyl groups showed potent estrogenic activity. However, the underlying mechanisms on the estrogenic activity and whether other AQs have the activity are still unclear.

In this study, an integrated *in vitro* and *in silico* approach was employed to investigate the estrogenic activity of AQs. Molecular docking was performed to define a model for the comprehension of the binding interactions between the ligand and ERα. The recombinant yeast-based assay was employed to determine the estrogenic activity of 20 AQ model compounds. By observing the ligand—receptor interactions, appropriate molecular structural parameters computed by the density function theory (DFT) were adopted to construct QSAR models. From the developed QSAR model, critical molecular structural features related to their estrogenic activity were identified. Furthermore, the developed model was externally validated, and the applicability domain was depicted.

2. Materials and Methods

2.1. Chemicals. The 20 AQs selected in this study were those that have been detected in environmental/biological samples (20, 21) and those with available standard congener samples. They were obtained from the State Key Laboratory of Fine Chemicals (Dalian University of Technology) and Sinochem Liaoning Imp. & Exp. Corp (>97% purity). Dimethyl sulfoxide (DMSO, GC, 99.5%), o-nitrophenyl β -D-galactopyranoside (o-NPG, \geq 98%), sodium dodecyl sulfate (SDS, 99%), histidine (99%), lysine (99%), yeast based nitrogen (99%), β -mercaptoethanol (99%), and 17β -estradiol (E₂, 98%) were purchased from Sigma (St. Louis, MO). The AQs were dissolved in DMSO, and the stock solutions were kept in the dark at 4 °C.

2.2. Recombinant Yeast-Based Assay. The recombinant yeast (*Saccharomyces cerevisiae*) cells were kindly provided by J. P. Sumpter from Brunel University. In this system, the DNA sequence of the human ER α , estrogen-responsive sequences, and the reporter gene lac-Z (encoding the enzyme β -galactosidase) were integrated into the yeast genome. In the presence of estrogens, β -galactosidase is synthesized and secreted into the medium, which is used to indicate the estrogenic activity (22).

The assay procedure described by Ma et al. was followed (23). Briefly, yeast cells were preincubated for 24 h at 30 °C. The cell density was adjusted to an absorbance of 0.75 at 600 nm. All of the samples were determined in a minimum of triplicates, and each triplicate included a positive control (E_2) and a negative control (DMSO).

The β -galactosidase activity (U) was calculated according to eq 1:

$$U = \frac{OD_{420} - OD'_{420}}{t \times V \times OD_{600}} \times D \tag{1}$$

where U is the activity of β -galactosidase, t is the incubation duration of the enzyme reaction, V is the volume of the test culture, D is diluting factor (D=6.6), OD_{600} is the cell density measured at 600 nm, OD_{420} and OD'_{420} stand for the cell density of the enzymic reaction supernatant and the blank measured at 420 nm, respectively.

The iterative four parameter curve fit method was used to obtain the dose—response relationships. The half maximal effective concentration (EC_{50}) was calculated from the dose—response curves. The estrogenic activities of the AQs were expressed by the logarithm of relative potency (RP) (24). The RP of each tested compound was calculated by the following equation:

$$RP = \frac{EC_{50}(E_2)}{EC_{50}(\text{tested compound})} \times 100$$
 (2)

where $EC_{50}(E_2)$ is the concentration of E_2 giving 50% induction in β -galactosidase activity.

2.3. Molecular Docking. The binding mode for the AQs to ERα was studied by CDOCKER, which has been incorporated into Discovery Studio 2.5 (Accelrys Software Inc.) through the Dock Ligands protocol. CDOCKER is an implementation of the docking tool based on the CHARMm forcefield that has been proved to be

Table 1. List of 20 Anthraquinone Derivatives and Their Observed and Predicted Estrogenic Activities and Binding Energy (E_{binding}) with ER α

				$\log RP$		
no.	compounds ^a	$EC_{50} (\mu M)$	E_{binding} (kJ/mol)	observed	predicted	residuals
1	1-aminoanthraquinone	9.14	-35.38	-1.66	-1.59	-0.07
2	2-aminoanthraquinone	9.80	-37.56	-1.69	-1.48	-0.22
3	1,4-diaminoanthraquinone*	1.91	-38.77	-0.98	-0.99	0.01
4	1,5-diaminoanthraquinone	1.74	-38.41	-0.94	-1.11	0.16
5	1-amino-2-bromoanthraquinone	3.99	-33.76	-1.30	-1.49	0.19
6	1-amino-4-bromoanthraquinone*	4.58	-36.96	-1.36	-1.45	0.09
7	1-amino-2,4-dibromoanthraquinone	5.38	-37.77	-1.43	-1.29	-0.14
8	1-amino-4-hydroxyanthraquinone	3.72	-38.62	-1.27	-1.32	0.05
9	1-amino-3-bromo-4-hydroxyanthraquinone	4.28	-39.08	-1.33	-1.19	-0.14
10	1,2-dihydroxyanthraquinone	5.14	-37.88	-1.41	-1.60	0.18
11	1,4-dihydroxyanthraquinone*	6.04	-38.02	-1.48	-1.61	0.13
12	1,5-dihydroxyanthraquinone	7.60	-36.60	-1.58	-1.55	-0.03
13	1-chloroanthraquinone*	10.74	-38.23	-1.73	-1.84	0.12
14	1,4,5,8-tetrachloroanthraquinone	8.34	-32.47	-1.62	-1.63	0.01
15	1,8-dichloroanthraquinone	14.49	-32.68	-1.86	-1.84	-0.02
16	anthraguinone	19.10	-37.31	-1.98	-1.96	-0.02
17	1-nitroanthraquinone*	5.38	-38.30	-1.43	-1.53	0.10
18	1,8-dihydroxy-4,5-dinitroanthraquinone	1.52	-50.00	-0.88	-0.86	-0.02
19	1,8-dinitroanthraquinone	4.08	-45.60	-1.31	-1.28	-0.03
20	1,5-dinitroanthraquinone	2.76	-49.72	-1.14	-1.23	0.09

^a Compounds marked with an asterisk (*) were selected to form the external validation set.

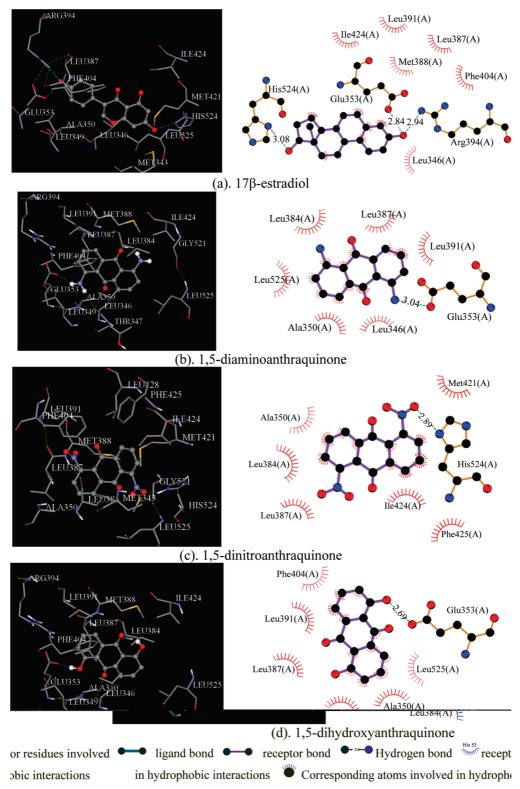


Figure 1. Hydrogen bonding (left) and hydrophobic interactions (right) between the AQs and ERα in the binding site.

viable (25). The crystal structure of ER α (PDB entry code: 1X7R) was extracted from the Brookhaven Protein Database (PDB http:// www.rcsb.org/pdb). In CDOCKER, random ligand conformations are generated through molecular dynamics, and a variable number of rigid-body rotations/translations are applied to each conformation to generate the initial ligand poses. The conformations are further refined by grid-based simulated annealing in the receptor active site, which makes the results accurate. The CDOCKER interaction energy between the AQs and ER α ($E_{binding}$) was finally computed. In addition, the electrostatic potentials of the ligand binding site for ERa were calculated by the electrostatic protocol that had been incorporated into Discovery Studio (version 2.5). From the docking analysis, insights into the interactions between the ligands and the receptor were gained, which facilitated the selection of appropriate molecular parameters to characterize the interactions in the subsequent QSAR studies.

2.4. Molecular Structural Parameter Generation and QSAR Development. It was hypothesized that the estrogenic activities of AQs depended on the following two processes: (a) the penetration of AQs through the biomembrane and reaching the target site of action, and (b) the interactions between the AQs and

 $ER\alpha$. A total of 15 theoretical parameters were computed and selected to characterize the processes.

The logarithm of octanol/water partition coefficient (log $K_{\rm OW}$) was purposely selected to describe the partition process. The parameters molecular volume (V), the average molecular polarizability (α) , and the polarizability term (π_1) (26) were also selected to partly describe the partition since they correlated with log $K_{\rm OW}$ (27). The following 11 parameters were selected to characterize intermolecular electrostatic interactions: the energy of the highest occupied molecular orbital $(E_{\rm HOMO})$, the energy of the lowest unoccupied molecular orbital $(E_{\rm LUMO})$, the most positive hydrogen atom in the molecule $(q^{\rm H}^+)$, the most negative formal charge in the molecule $(q^{\rm T})$, the electrophilicity index (ω) , the most positive and most negative values of the molecular surface potential $(V_{\rm S,max}, V_{\rm S,min})$, the averages of the positive and negative potentials on the molecular surface $(\bar{V}_{\rm s}^+, \bar{V}_{\rm s}^-)$, the average deviation of surface potential (Π) , and the balance parameter of the surface potential (τ) .

 $E_{\rm HOMO}$ and $E_{\rm LUMO}$ describe the electron donating/accepting character of a molecule (28, 29). $E_{\rm HOMO}$, $E_{\rm LUMO}$, $q{\rm H}^+$, and q^- were successfully employed in QSARs for characterizing intermolecular electrostatic interactions (30). ω measures the ability of a compound to soak up electrons. The relative binding affinity of some estrogen derivatives correlates strongly with ω (31). The molecular surface potential indicates the charge distribution in a molecule (32), which gauges the basicity and nucleophilicity of a molecule (33). These 6 potential derived parameters were used to rationalize the toxicity of 209 chlorinated diphenyls (34). $E_{\rm binding}$ was also considered, as binding to the receptor is the initiating step in endocrine disruption.

log $K_{\rm OW}$ values were computed by the EPI Suite 4.0 (http://www.epa.gov/oppt/exposure). V is defined as the volume inside a contour of 0.001 electrons/Bohr³ density. $\pi_{\rm I}$ was calculated as α/V . The equations for calculating ω and the 6 potential derived parameters (35) are described in Supporting Information.

All of the quantum chemical parameters were computed by the Gaussian 09 program (36). Initial geometries were optimized by semiempirical method PM3, then optimized at the hybrid density functional theory (DFT) using the B3LYP/6-31G(d,p) level (37). Solvent (water) effects were considered implicitly by the integral equation formulation of the polarized continuum model (IEFPCM) (38). The frequency analysis was performed on the optimized geometries to ensure that the systems had no imaginary vibration frequencies. The potentials on the molecular surface were calculated on the 0.001 electrons/Bohr³ surface.

The 20 AQs were randomly divided into a training set (75%) and a validation set (25%), as listed in Table 1. Partial least-squares (PLS) regression was performed for the model development as PLS can analyze data with strongly collinear, noisy, and numerous predictor variables (39). Simca-S (version 6.0, Umetri AB & Erisoft AB) was employed for the PLS analysis. Simca-S adopts a leavemany-out cross-validation to determine the number of PLS components (A). Cross-validation simulates how well a model predicts new data and gives a statistical Q^2_{CUM} (the fraction of the total variation of the dependent variables that can be predicted by all of the extracted components) for the model. The PLS analysis was performed repeatedly so as to eliminate redundant molecular structural parameters, as done in our previous studies (40). The model predictability was evaluated by external validation. The performance of external validation (41) was characterized by the external explained variance (Q^2_{EXT}) and the standard errors (SE), which are described in Supporting Information.

The applicability domain of the developed QSAR model was assessed by the Euclidean distance-based method, which was incorporated into the AMBIT Discovery (verion 0.04) (http://ambit.acad.bg/downloads/AmbitDiscovery/). The Euclidean distance method identifies the interpolation regions by assuming that the data are normally distributed (42). The normal distribution was confirmed by the Kolmogorov—Smirnov test (at the 95% confidence level).

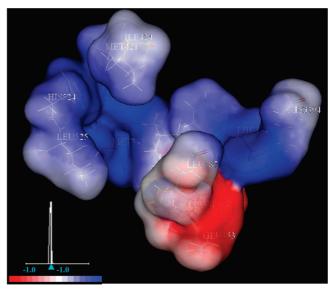


Figure 2. Electrostatic potential of the ligand binding site for ER α .

3. Results and Discussion

3.1. Estrogenic Activity Determined by the Recombined Yeast. The concentration range used in the assay was $10^{-11}-10^{-6}$ M for E_2 and $10^{-7}-10^{-4}$ M for the AQs. The determined EC_{50} values are listed in Table 1. EC_{50} for E_2 (2 nM) was similar to that reported by Chen et al. (43) and Maier et al. (44). Of the 20 AQs, 1,8-dihydroxy-4,5-dinitroanthraquinone was the most potent ($EC_{50} = 1.52 \mu M$).

3.2. Structural Analysis of Docking. The docking view of three representative AQs (1,5-diaminoanthraquinone, 1,5-dinitroanthraquinone and 1,5-dihydroxyanthraquinone) in the binding site of ER α was shown in Figure 1. At the deep end of the pocket, Glu353 (hER α numbering) serves as the anchoring points for the ligands, which explains the predominant contribution of the hydroxyl group to the binding affinity of the AQs. The ligands also interact with another polar region (His524) within the binding pocket.

As shown in Figure 1, hydrogen bonding, hydrophobic, and π - π interactions are characteristic interactions between the AQs and ERα. There are mainly three types of H-bonds: (a) H-bonds formed between the amino hydrogen of the AQs and the side chain oxygen of Leu346, (b) H-bonds formed between the nitro oxygen of the AQs and the hydrogen of the imidazole of His524, and (c) H-bonds formed between the hydroxyl hydrogen of the phenol ring with the carbonyl oxygen of Glu353. Acting as an anchor, the hydrogen-bonding intensely determines the 3D space position of the quinone ring and the benzene ring in the binding pocket, and facilitates the hydrophobic interaction of the AQs with the side chain of Leu384, Leu387, Ala350, Phe425, Met421, and Ile424. There are also $\pi - \pi$ interactions between the benzene rings of the AQs and Phe404 and Phe425. Figure 2 shows the electrostatic potential of the ligand-binding site for $ER\alpha$. The binding site has positive potentials. It can thus be concluded that the negative potentials of the AQ molecules facilitate their binding with $ER\alpha$.

The binding free energies ($E_{\rm binding}$) of the 20 AQs are listed in Table 2. A simple linear free energy relationship (LFER) between log RP and $E_{\rm binding}$ was obtained (Figure 3), which further proved that the binding to ER α was a key step for the AQs to exert their estrogenic activity. However, $E_{\rm binding}$ itself was not a good predictor for log RP, as indicated by the large prediction residuals for some AQs (Figure 3). It is necessary to develop poly parameter QSAR models on log RP.

Table 2	Values of	the Molecular	Structural	Parameters	Included	in the	OSAR Model ^a
Table 2.	values of	the Molecular	Structural	Parameters	mciuaea .	m me	OSAK Model

no.	compounds	$\pi_{ m I}$	E_{binding} (kJ/mol)	α (a.u.)	q- (a.c.u.)	\bar{V}_s^- (eV)
1	1-aminoanthraquinone	0.814	-35.378	133.342	-0.338	-0.018
2	2-aminoanthraquinone	0.824	-37.563	134.730	-0.305	-0.023
3	1,4-diaminoanthraquinone	0.878	-38.769	148.350	-0.337	-0.025
4	1,5-diaminoanthraquinone	0.864	-38.414	146.160	-0.338	-0.022
5	1-amino-2-bromoanthraquinone	0.809	-33.760	147.474	-0.335	-0.013
6	1-amino-4-bromoanthraquinone	0.799	-36.956	145.246	-0.336	-0.021
7	1-amino-2,4-dibromoanthraquinone	0.794	-37.765	159.767	-0.334	-0.017
8	1-amino-4-hydroxyanthraquinone	0.829	-38.619	139.076	-0.330	-0.028
9	1-amino-3-bromo-4-hydroxyanthraquinone	0.821	-39.079	153.859	-0.326	-0.023
10	1,2-dihydroxyanthraquinone	0.798	-37.880	132.252	-0.303	-0.026
11	1,4-dihydroxyanthraquinone	0.795	-38.016	131.885	-0.278	-0.029
12	1,5-dihydroxyanthraquinone	0.810	-36.604	133.529	-0.351	-0.022
13	1-chloroanthraquinone	0.762	-38.226	130.034	-0.298	-0.018
14	1,4,5,8-tetrachloroanthraquinone	0.745	-32.468	159.470	-0.268	-0.020
15	1,8-dichloroanthraquinone	0.750	-32.676	138.941	-0.294	-0.024
16	anthraquinone	0.771	-37.313	120.828	-0.303	-0.016
17	1-nitroanthraquinone	0.750	-38.303	134.102	-0.587	-0.036
18	1,8-dihydroxy-4,5-dinitroanthraquinone	0.765	-49.998	160.532	-0.585	-0.041
19	1,8-dinitroanthraquinone	0.730	-45.602	147.598	-0.577	-0.040
20	1,5-dinitroanthraquinone	0.731	-49.718	147.231	-0.589	-0.035

a a.u., atom unit; a.c.u., atom charge unit.

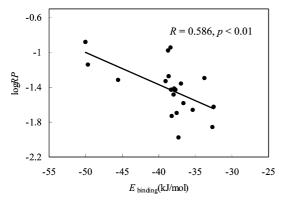


Figure 3. Plot of the log RP values vs the binding energy (E_{binding}) . RP, relative potency; E_{binding} , receptor binding energy (kJ/mol).

3.3. Development, Validation, and Applicability Domain of the QSAR Model. The following optimal QSAR model was developed, for which the values of the included molecular structural parameters are listed in Table 2.

$$\log RP = -8.08 + 4.51\pi_{\rm I} - 1.84 \times 10^{-2} E_{\rm binding} + 1.36 \times 10^{-2} \alpha - 6.70 \times 10^{-1} q^{-1} - 6.82 \bar{V}_s^{-1}$$

$$n = 15$$
, $A = 2$, $R^2 = 0.85$, $Q_{\text{CUM}}^2 = 0.73$,
 $SE = 0.12$ (training set), $Q_{\text{EXT}}^2 = 0.83$,
 $SE = 0.11$ (validation set), $p < 0.0001$

where p is the significance level.

The predicted log RP values and residuals are listed in Table 1. The R^2 value of the QSAR model was 0.85, indicating a good goodness-of-fit. Q^2_{CUM} of the QSAR is as high as 0.73, implying good robustness of the model. The difference between R^2 and Q^2_{CUM} does not exceed 0.3, indicating no overfitting in the model (45). As shown in Figure 4, the predicted log RP values were close to the observed values for both the validation and training sets. The Kolmogorov-Smirnov test for normality (at the 95% confidence level) confirmed a normal distribution of residuals, with the mean = 0.02, and standard deviation = 0.03. The model has acceptable predictability, as indicated by $Q^2_{\rm EXT} = 0.83$ and SE = 0.11. In summary, the developed QSAR model shows satisfactory performance.

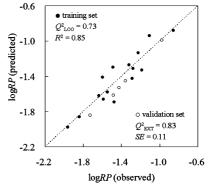


Figure 4. Plot of the observed vs predicted log RP values.

The applicability domain of the developed OSAR model is shown in Figure 5. All of the AQs in both the training and validation sets are in the domain, and none of them was particularly influential in the model space. Thus, the training set has great representativeness. All of the absolute residuals between the observed and predicted values are less than 3SE (Figure 6), and there are no outliers for the developed QSAR model. Thus, the developed QSAR model can be used to predict the estrogenic activity of other AQs to ERα.

Matsuda et al. (24) assayed the estrogenic activity of emodin, 2,6-dihydroxyanthraquinone, daidzein, and genistein using the competitive binding assay. The log RP values predicted by the established QSAR for emodin, 2,6-dihydroxyanthraquinone, and genistein are -1.50, -0.59, and -0.97, respectively, indicating that 2,6-dihydroxyanthraquinone is the most potent, followed by genistein and emodin. Thus, the predictions are consistent with the observations of Matsuda et al. (24).

3.4. Mechanistic Implications. The developed PLS model extracted 2 PLS components that were loaded on 5 predictor variables. Values of the variable importance in the projection (VIP) and PLS weights (W*) are listed in Table 3. From the W* values, one can see how the predictor variables and the response variable combine in the projections (PLS components) and how they relate to each other.

The first PLS component was loaded primarily on the 4 descriptors, E_{binding} , α , q^- , and \bar{V}_s^- (Table 3). These four descriptors remarkably govern log RP, as indicated by their large VIP values. The coefficients in the QSAR model indicate that $\log RP$ increases with $\pi_{\rm I}$ and α , but decreases with increasing

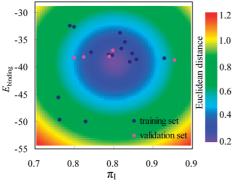


Figure 5. Applicability domain for the developed QSAR model characterized by the Euclidean distance-based approach.

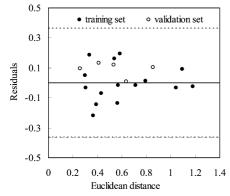


Figure 6. Plot of the Euclidean distance vs the residuals

Table 3. VIP Values and PLS Weights for the Optimal PLS Model

	VIP	W*c[1]	W*c[2]
$\pi_{ m I}$	1.167	0.137	0.950
$E_{\rm binding}$	0.996	-0.538	-0.121
α	0.986	0.495	0.513
q^-	0.947	-0.496	-0.007
$ar{ar{V}}_s^-$	0.881	-0.449	0.055

 $E_{\rm binding}$ and \bar{V}_s^- values. The electrostatic basicity term (q^-) is another significant contribution to $\log RP$. q^- itself has a negative value, and may characterize the hydrogen bond basicity of a molecule (30, 31). The AQs with low q^- values may have strong hydrogen bond basicity (accept protons more easily) and accordingly form H-bonds easily, exhibiting strong estrogenic activity. \bar{V}_s^- itself also has a negative value. The electrostatic potential analysis shows that the binding site prefers to bind with the negative potentials on the molecules. Hence, it is reasonable that in the developed PLS model $\log RP$ increases with decreasing \bar{V}_s^- values.

The second PLS component is loaded primarily on π_I and α . π_I describes the ability of electrons moving throughout a molecule (30) and may describe dispersive interactions between the AQs and ER α . log RP values increase with π_I and α values. As both π_I and α correlate with log $K_{\rm OW}$ and compounds with higher α values in an appropriate range (neither too small nor too large) tend to have large log $K_{\rm OW}$ values, the information of π_I and α condensed in the second PLS component may describe the partition of the AQs into the biophase.

4. Conclusions

Hydrogen bonding, hydrophobic, and $\pi-\pi$ interactions between AQs and ER α govern the estrogenic activities of the AQs. The developed QSAR model has good robustness, predictive ability, and mechanism interpretability, which can

be applied to predict the estrogenic activity of other AQs. AQs with lower q^- and \bar{V}_s^- values tend to have higher estrogenic activity. The partition ability of the AQs into the biophase also relates to their estrogenicity. Comprehension of the binding interactions between the ligands and the receptor through docking analysis is necessary for the development of mechanism-based QSAR models.

Acknowledgment. This study was supported by the High Technology Research and Development Program of China, the National Natural Science Foundation of China (20890113), and the Program for Changjiang Scholars and Innovative Research Team in Universities (IRT0813).

Supporting Information Available: The electrophilicity index, potential derived parameters, and external validation. This material is available free of charge via the Internet at http://pubs.acs.org.

References

- Coe, T. S., Hamilton, P. B., Hodgson, D., Paull, G. C., Stevens, J. R., Sumner, K., and Tyler, C. R. (2008) An environmental estrogen alters reproductive hierarchies, disrupting sexual selection in group-spawning fish. *Environ. Sci. Technol.* 42, 5020–5025.
- (2) Harris, H. A., Katzenellenbogen, J. A., and Katzenellenbogen, B. S. (2002) Characterization of the biological roles of the estrogen receptors, ER alpha and ER beta, in estrogen target tissues in vivo through the use of an ER alpha-selective ligand. *Endocrinology* 143, 4172–4177.
- (3) Kuiper, G. G. J. M., Carlsson, B., Grandien, K., Enmark, E., Haggblad, J., Nilsson, S., and Gustafsson, J. A. (1997) Comparison of the ligand binding specificity and transcript tissue distribution of estrogen receptors alpha and beta. *Endocrinology* 138, 863–870.
- (4) Routledge, E. J., and Sumpter, J. P. (1997) Structural features of alkylphenolic chemicals associated with estrogenic activity. *J. Biol. Chem.* 272, 3280–3288.
- (5) Gierthy, J. F., Arcaro, K. F., Li, A., Silber, P., Lloyd, S., and Yang, Y. (2003) Optimization and validation of the MCF-7 focus assay for estrogen modulators. *Toxicol. Sci* 72, 154–155.
- (6) Kanno, J., Onyon, L., Haseman, J., Fenner-Crisp, P., Ashby, J., and Owens, W. (2001) The OECD program to validate the rat uterotrophic bioassay to screen compounds for in vivo estrogenic responses: Phase 1. Environ. Health Perspect. 109, 785–794.
- (7) Lascombe, I., Beffa, D., Ruegg, U., Tarradellas, J., and Wahli, W. (2000) Estrogenic activity assessment of environmental chemicals using in vitro assays: Identification of two new estrogenic compounds. *Environ. Health Perspect.* 108, 621–629.
- (8) Shelby, M. D., Newbold, R. R., Tully, D. B., Chae, K., and Davis, V. L. (1996) Assessing environmental chemicals for estrogenicity using a combination of in vitro and in vivo assays. *Environ. Health Perspect.* 104, 1296–1300.
- (9) Nishikawa, J., Saito, K., Goto, J., Dakeyama, F., Matsuo, M., and Nishihara, T. (1999) New screening methods for chemicals with hormonal activities using interaction of nuclear hormone receptor with coactivator. *Toxicol. Appl. Pharmacol.* 154, 76–83.
- (10) Leusch, F. D. L., Jager, C. d., Levi, Y., Lim, R., Puijker, L., Sacher, F., Tremblay, L. A., Wilson, V. S., and Chapman, H. F. (2010) Comparison of five in vitro bioassays to measure estrogenic activity in environmental waters. *Environ. Sci. Technol.* 44, 3853–3860.
- (11) Soto, A. M., Sonnenschein, C., Chung, K. L., Fernandez, M. F., Olea, N., and Serrano, F. O. (1995) The E-screen assay as a tool to identify estrogens: an update on estrogenic environmental-pollutants. *Environ. Health Perspect.* 103, 113–122.
- (12) 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. (1996) Ligand-based identification of environmental estrogens. *Chem. Res. Toxicol.* 9, 1240–1248.
- (13) Shi, L. M., Fang, H., Tong, W. D., Wu, J., Perkins, R., Blair, R. M., Branham, W. S., Dial, S. L., Moland, C. I., and Sheehan, D. M. (2001) QSAR models using a large diverse set of estrogens. *J. Chem. Inf. Comput. Sci.* 41, 186–195.
- (14) Asikainen, A. H., Ruuskanen, J., and Tuppurainen, K. A. (2004) Consensus kNN QSAR: A versatile method for predicting the estrogenic activity of organic compounds in silico. A comparative study with five estrogen receptors and a large, diverse set of ligands. Environ. Sci. Technol. 38, 6724–6729.

- (15) US EPA (2009) High Throughput Pre-Screening (HTPS) and Quantitative Structure Activity Relationships (QSAR), http://epa.gov/endo/pubs/ prioritysetting/models.htm.
- (16) Organisation for Economic Co-Operation and Development (OECD) (2007) Guidance document on the validation of (Quantitative) Structure-Activity Relationships [(Q)SARs] models, http://www.oecd. org/dataoecd/55/22/38131728.pdf.
- (17) Amadasi, A., Mozzarelli, A., Meda, C., Maggi, A., and Cozzini, P. (2009) Identification of xenoestrogens in food additives by an integrated in silico and in vitro approach. Chem. Res. Toxicol. 22, 52–63.
- (18) Celik, L., Lund, J. D. D., and Schiott, B. (2008) Exploring interactions of endocrine-disrupting compounds with different conformations of the human estrogen receptor alpha ligand binding domain: a molecular docking study. *Chem. Res. Toxicol.* 21, 2195–2206.
- (19) Weber, E. J., and Adams, R. L. (1995) Chemical-mediated and sediment-mediated reduction of the azo-dye disperse-blue-79. *Environ.* Sci. Technol. 29, 1163–1170.
- (20) Kurihara, R., Shiraishi, F., Tanaka, N., and Hashimoto, S. (2005) Presence and estrogenicity of anthracene derivatives in coastal Japanese waters. *Environ. Toxicol. Chem.* 24, 1984–1993.
- (21) Matsuda, H., Shimoda, H., Morikawa, T., and Yoshikawa, M. (2001) Phytoestrogens from the roots of *Polygonum cuspidatum* (Polygonaceae): Structure-requirement of hydroxyanthraquinones for estrogenic activity. *Bioorg. Med. Chem. Lett.* 11, 1839–1842.
- (22) Routledge, E. J., and Sumpter, J. P. (1996) Estrogenic activity of surfactants and some of their degradation products assessed using a recombinant yeast screen. *Environ. Toxicol. Chem.* 15, 241–248.
- (23) Ma, M., Li, J., and Wang, Z. J. (2005) Assessing the detoxication efficiencies of wastewater treatment processes using a battery of bioassays/biomarkers. Arch. Environ. Contam. Toxicol. 49, 480–487.
- (24) Fang, H., Tong, W. D., Perkins, R., Soto, A. M., Prechtl, N. V., and Sheehan, D. M. (2000) Quantitative comparisons of in vitro assays for estrogenic activities. *Environ. Health Perspect.* 108, 723–729.
- (25) Wu, G. S., Robertson, D. H., Brooks, C. L., and Vieth, M. (2003) Detailed analysis of grid-based molecular docking: A case study of CDOCKER - A CHARMm-based MD docking algorithm. J. Comput. Chem. 24, 1549–1562.
- (26) Famini, G. R., and Wilson, L. Y. (1999) Using theoretical descriptors in linear free energy relationships: characterizing several polarity, acid and basicity scales. J. Phys. Org. Chem. 12, 645–653.
- (27) Nguyen, T. H., Goss, K. U., and Ball, W. P. (2005) Polyparameter linear free energy relationships for estimating the equilibrium partition of organic compounds between water and the natural organic matter in soils and sediments. *Environ. Sci. Technol.* 39, 913–924.
- (28) Honorio, K. M., and Da Silva, A. B. F. (2003) An AM1 study on the electron-donating and electron-accepting character of biomolecules. *Int. J. Quantum Chem.* 95, 126–132.
- (29) Riahi, S., Eynollahi, S., and Ganjali, M. R. (2009) Computational studies on anthraquinone-2-sulfonic acid, electrochemistry and electronic investigation. *Int. J. Electrochem. Sci. 4*, 1309–1318.
- (30) Colosi, L. M., Huang, Q. G., and Weber, W. J. (2006) Quantitative structure-activity relationship based quantification of the impacts of enzyme-substrate binding on rates of peroxidase-mediated reactions of estrogenic phenolic chemicals. J. Am. Chem. Soc. 128, 4041–4047.
- (31) Chattaraj, P. K., Sarkar, U., and Roy, D. R. (2006) Electrophilicity index. *Chem. Rev.* 106, 2065–2091.
- (32) Politzer, P., Abrahmsen, L., and Sjoberg, P. (1984) Effects of amino and nitro substituents upon the electrostatic potential of an aromatic ring. J. Am. Chem. Soc. 106, 855–860.

- (33) Colominas, C., Orozco, M., Luque, F. J., Borrell, J. I., and Teixido, J. (1998) A priori prediction of substituent and solvent effects in the basicity of nitriles. J. Org. Chem. 63, 4947–4953.
- (34) Chana, A., Concejero, M. A., de Frutos, M., Gonzalez, M. J., and Herradon, B. (2002) Computational studies on biphenyl derivatives. Analysis of the conformational mobility, molecular electrostatic potential, and dipole moment of chlorinated biphenyl: Searching for the rationalization of the selective toxicity of polychlorinated biphenyls (PCBs). Chem. Res. Toxicol. 15, 1514–1526.
- (35) Padmanabhan, J., Parthasarathi, R., Subramanian, V., and Chattaraj, P. K. (2006) Group philicity and electrophilicity as possible descriptors for modeling ecotoxicity applied to chlorophenols. *Chem. Res. Toxicol.* 19, 356–364.
- (36) Frisch, M. J., Trucks, G. W., Schlegel, H. B., Scuseria, G. E., Robb, M. A., Cheeseman, J. R., Scalmani, G., Barone, V., Mennucci, B., Petersson, G. A., Nakatsuji, H., Caricato, M., Li, X., Hratchian, H. P., Izmaylov, A. F., Bloino, J., Zheng, G., Sonnenberg, J. L., Hada, M., Ehara, M., Toyota, K., Fukuda, R., Hasegawa, J., Ishida, M., Nakajima, T., Honda, Y., Kitao, O., Nakai, H., Vreven, T., Montgomery, Jr., J. A., Peralta, J. E., Ogliaro, F., Bearpark, M., Heyd, J. J., Brothers, E., Kudin, K. N., Staroverov, V. N., Kobayashi, R., Normand, J., Raghavachari, K., Rendell, A., Burant, J. C. İyengar, S. S. Tomasi, J. Cossi, M. Rega, Millam, N. J., Klene, M. Knox, J. E., Cross, J. B., Bakken, V., Adamo, C., Jaramillo, J., Gomperts, R. E. Stratmann, O. Yazyev, A. J. Austin, R. Cammi, C. Pomelli, J. W. Ochterski, R. Martin, R. L., Morokuma, K., Zakrzewski, V. G., Voth, G. A., Salvador, P., Dannenberg, J. J., Dapprich, S., Daniels, A. D., Farkas, O., Foresman, J. B., Ortiz, J. V., Cioslowski, J., Fox, D. J. (2009) Gaussian 09, revision A.1, Gaussian, Inc., Wallingford, CT.
- (37) Arulmozhiraja, S., Shiraishi, F., Okumura, T., Iida, M., Takigami, H., Edmonds, J. S., and Morita, M. (2005) Structural requirements for the interaction of 91 hydroxylated polychlorinated biphenyls with estrogen and thyroid hormone receptors. *Toxicol. Sci.* 84, 49–62.
- (38) Huetz, P., Kamarulzaman, E. E., Wahab, H. A., and Mavri, J. (2004) Chemical reactivity as a tool to study carcinogenicity: Reaction between estradiol and estrone 3,4-quinones ultimate carcinogens and guanine. *J. Chem. Inf. Comput. Sci.* 44, 310–314.
- (39) Wold, S., Sjostrom, M., and Eriksson, L. (2001) PLS-regression: a basic tool of chemometrics. *Chemom. Intell. Lab. Syst.* 58, 109–130.
- (40) Chen, J. W., Harner, T., Ding, G. H., Quan, X., Schramm, K. W., and Kettrup, A. (2004) Universal predictive models on octanol-air partition coefficients at different temperatures for persistent organic pollutants. *Environ. Toxicol. Chem.* 23, 2309–2317.
- (41) Schüürmann, G., Ebert, R. U., Chen, J. W., Wang, B., and Kuhne, R. (2008) External validation and prediction employing the predictive squared correlation coefficient: test set activity mean vs training set activity mean. J. Chem. Inf. Model. 48, 2140–2145.
- (42) Jaworska, J., Nikolova-Jeliazkova, N., and Aldenberg, T. (2005) QSAR applicability domain estimation by projection of the training set in descriptor space: A review. Atla-Altern. Lab. Anim. 33, 445–459.
- (43) Chen, C. W., Hurd, C., Vorojeikina, D. P., Arnold, S. F., and Notides, A. C. (1997) Transcriptional activation of the human estrogen receptor by DDT isomers and metabolites in yeast and MCF-7 cells. *Biochem. Pharmacol.* 53, 1161–1172.
- (44) Maier, C. G. A., Chapman, K. D., and Smith, D. W. (1995) Differential estrogenic activities of male and female plant-extracts from 2 dioecious species. *Plant Sci. 109*, 31–43.
- (45) Golbraikh, A., and Tropsha, A. (2002) Beware of q²! J. Mol. Graphics Modell. 20, 269–276.

TX100118G