

Quantitative Structure-Activity Relationships (QSARs) for Estrogen Binding to the Estrogen Receptor: Predictions across Species

Weida Tong; Roger Perkins; Richard Strelitz; Elizabeth R. Collantes; Susan Keenan; William J. Welsh; William S. Branham; Daniel M. Sheehan

Environmental Health Perspectives, Vol. 105, No. 10. (Oct., 1997), pp. 1116-1124.

Stable URL:

http://links.jstor.org/sici?sici=0091-6765%28199710%29105%3A10%3C1116%3AQSR%28FE%3E2.0.CO%3B2-Q

Environmental Health Perspectives is currently published by The National Institute of Environmental Health Sciences (NIEHS).

Your use of the JSTOR archive indicates your acceptance of JSTOR's Terms and Conditions of Use, available at http://www.jstor.org/about/terms.html. JSTOR's Terms and Conditions of Use provides, in part, that unless you have obtained prior permission, you may not download an entire issue of a journal or multiple copies of articles, and you may use content in the JSTOR archive only for your personal, non-commercial use.

Please contact the publisher regarding any further use of this work. Publisher contact information may be obtained at <u>http://www.jstor.org/journals/niehs.html</u>.

Each copy of any part of a JSTOR transmission must contain the same copyright notice that appears on the screen or printed page of such transmission.

JSTOR is an independent not-for-profit organization dedicated to and preserving a digital archive of scholarly journals. For more information regarding JSTOR, please contact support@jstor.org.

Quantitative Structure–Activity Relationships (QSARs) for Estrogen Binding to the Estrogen Receptor: Predictions across Species

Weida Tong,¹ Roger Perkins,¹ Richard Strelitz,¹ Elizabeth R. Collantes,² Susan Keenan,² William J. Welsh,² William S. Branham,³ and Daniel M. Sheehan³

¹R.O.W Sciences, Jefferson, AR 72079 USA; ²Department of Chemistry, University of Missouri-St. Louis, St. Louis, MO 63121 USA; ³Division of Reproductive and Developmental Toxicology, National Center for Toxicological Research, Jefferson, AR 72079 USA

The recognition of adverse effects due to environmental endocrine disruptors in humans and wildlife has focused attention on the need for predictive tools to select the most likely estrogenic chemicals from a very large number of chemicals for subsequent screening and/or testing for potential environmental toxicity. A three-dimensional quantitative structure-activity relationship (QSAR) model using comparative molecular field analysis (CoMFA) was constructed based on relative binding affinity (RBA) data from an estrogen receptor (ER) binding assay using calf uterine cytosol. The model demonstrated significant correlation of the calculated steric and electrostatic fields with RBA and yielded predictions that agreed well with experimental values over the entire range of RBA values. Analysis of the CoMFA three-dimensional contour plots revealed a consistent picture of the structural features that are largely responsible for the observed variations in RBA. Importantly, we established a correlation between the predicted RBA values for calf ER and their actual RBA values for human ER. These findings suggest a means to begin to construct a more comprehensive estrogen knowledge base by combining RBA assay data from multiple species in 3D-QSAR based predictive models, which could then be used to screen untested chemicals for their potential to bind to the ER. Another QSAR model was developed based on classical physicochemical descriptors generated using the CODESSA (Comprehensive Descriptors for Structural and Statistical Analysis) program. The predictive ability of the CoMFA model was superior to the corresponding CODESSA model. Key words: CODESSA, CoMFA, endocrine disruptors, estrogen receptor, estrogens, quantitative structure-activity relationships, QSAR, relative binding affinity, species-to-species extrapolation, xenoestrogens. Environ Health Perspect 105:1116-1124 (1997). http://ehis.niehs.nih.gov

A significant number of compounds with a broad diversity of chemical structures, produced both in nature and by man, have estrogenic activity (1). Estrogens elicit many cellular responses in target tissues and can exert both positive and negative effects on health and reproductive function. For example, estrogens are used beneficially for fertility control (oral contraception) (2) and for relief of menopausal symptoms (estrogen replacement therapy) (3). The adverse developmental effects of diethylstilbestrol (DES) demonstrate human fetal sensitivity to estrogenic compounds (4), while other xenoestrogens appear to disrupt endocrine function in wildlife and in laboratory animals (5,6). Recently, concern about the adverse effects of chemical compounds with estrogenic activity on humans and other species has grown rapidly (7). Adverse effects on the development of reproductive capacity is the highest priority (8). Another manifest concern is xenoestrogen involvement in some common cancers in women (9). The emergence of the endocrine disruptor issue has resulted in new laws requiring evaluation of some chemicals found in foods or water for several types of hormonal activity (10,11).

Seventy thousand or more chemicals may ultimately need to be evaluated for estrogenic activity, and methods need to be developed to distinguish which of these should have highest priority for entry into the more expensive screening and testing procedures.

Because the chemical structures of xenoestrogens are highly diverse (12), estrogenicity is not readily deduced from simple inspection of chemical structure. Therefore, risk assessments for estrogenic chemicals are basically dependent on in vivo assays (13), such as uterine weight gain (14, 15), which measures responses in estrogen sensitive tissues, and multigeneration studies, which assess reproductive performance. Supporting data may come from reporter gene assays (16), cell proliferation assays (17), or from in vitro studies of competition of estrogen binding to the estrogen receptor (ER) (18). However, many in vivo assays are labor-intensive, time-consuming, and costly, which makes them impractical for routine screening and testing of a large number of chemicals.

Many estrogen responses are thought to be mediated via estrogen binding to the classical estrogen receptor (ER- α); this mechanism is the basis for the correlation between the ability of a chemical to compete for estradiol binding to ER and to induce estrogenic effects *in vivo*. Therefore, the determination of the pharmacophoric elements of ligands that bind to the ER is crucial to understanding the biological effects of estrogens (19).

Over the past 30 years, a large volume of estrogen-related data has accumulated in the literature. These data cover molecular biology to medicine, developmental roles to adverse effects, evolutionary conservation to reproduction, and many other diverse and important disciplines. This huge estrogen database could be transformed into a knowledge base for regulatory and research purposes in part by understanding the basis of the relationship between chemical structure and estrogenic activity. Quantitative structure-activity relationship (QSAR) analysis correlates chemical structure and a specific biological activity, and the derived models are used to predict activities of untested chemicals. Ultimately, it is reasonable to expect that a number of such models may be necessary to cover a number of possible mechanisms of effects of chemicals on estrogenicity (e.g., ER-B, aromatase inhibition, etc.).

Since the pioneering work of Hansch and colleagues (20) in the 1960s, QSAR models have been applied extensively in various areas in chemistry and biology (21). For environmental toxicology (22), QSAR is seen as a scientifically credible tool for predicting and classifying biological activities of chemicals when little or no actual data are available (23). QSAR studies generally involve two steps: first, descriptors (physicochemical parameters) are generated which encode for chemical structural information; and second, a statistical regression method correlates changes in structure with changes in biological activity. The compounds in the training set (i.e., the data set selected to construct the QSAR model) should be diverse both in chemical structure and biological activity to ensure a statistically robust model. The QSAR method typically assumes that chemicals function by a common mechanism. The model is then validated by predicting the biological activity for a test set (i.e., a group of chemicals not

Address correspondence to W. Tong, R.O.W. Sciences, Inc., 3900 NCTR Road, Mail Code-910, Jefferson, AR 72079 USA.

The authors gratefully acknowledge the generous support of the Food and Drug Administration's Office of Women's Health.

Received 1 April 1997; accepted 1 July 1997.

included among the training-set compounds). Once validated, these QSAR models can be used to predict activities of untested chemicals.

Recent advances in computing technology have extended classical QSAR to threedimensional QSAR (3D-QSAR) models that correlate the role of molecular shape with some endpoints which are usually biological in nature. A widely used approach for generating descriptors based on the three-dimensional structural information of molecules is comparative molecular field analysis (CoMFA) (24), which is recognized as a versatile and powerful tool in rational drug design (25) and related applications (26). CoMFA implicitly assumes that the ligand-receptor interaction is primarily noncovalent in nature, shape dependent, and invariant across the set of compounds under examination. To construct a CoMFA model, a collection of compounds with known activities (i.e., the training set) are first aligned together, usually employing structure similarity as the basis for alignment. The aligned molecules are then embedded in a three-dimensional grid, after which the steric and electrostatic fields are computed for each compound at every grid point surrounding the molecules. The variations in these computed steric-electrostatic fields are then correlated with variations in the observed biological activity [e.g., ER relative binding affinity (RBA)] using the multivariate (linear) regression technique of partial least-square (PLS), thus forming the basis of a statistical model potentially capable of predicting the biological activity of compounds outside of the training set.

Several CoMFA models have been developed for both natural and synthetic estrogens (27-31) and for related steroids and their receptors (32-34). However, these studies either included limited structural diversity in the training set or were not validated with a test set. There has been little effort to examine the relationships between estrogen binding across receptors from different species although this is an important issue inasmuch as interspecies extrapolation is one of the major sources of uncertainty in risk assessment.

The present study is the first in a series from our Estrogen Knowledge Base (EKB) program that seeks to rationalize the relationship between a structurally diverse set of estrogenic compounds and their RBAs to the ER using QSAR approaches. The ultimate goal of this EKB program is to derive computational models for estrogenicity that will serve as predictive tools to screen a large number of natural and synthetic chemicals from numerous structural classes. These tools will help identify chemicals that exhibit a high potential for binding to the estrogen receptor. The actual RBA will need to be established by binding studies and/or other short-term measures such as reporter gene or MCF-7 cell proliferation assays. Although the literature contains many different endpoints for biological activity, few data sets exist for large numbers of estrogens assayed under identical conditions. For this reason, we developed a QSAR model based on a specific endpoint from a single animal model, i.e., RBA data from calf ER. Subsequently, this model was used to predict RBAs in human ER. Success here may allow large data sets to be built from smaller sets, and by analogy with meta-analysis in epidemiology, we might thereby increase the resolving power of the model.

Because models developed with CoMFA are sensitive to alignment that can be somewhat arbitrary and discretionary (35), CoMFA results are difficult to duplicate. For this reason, we concurrently evaluated classical QSAR that uses physicochemical parameters in the formulation of regression equations and requires no alignment step. In the recent past, the CODESSA program (Comprehensive Descriptors for Structural and Statistical Analysis; Semichem, Shawnee, KS) has been successfully used in quantitative structure-property relationship (QSPR) studies to predict normal boiling points (36), gas chromatographic retention times (37), and glass transition temperatures (38), among other physical properties of compounds. We investigated the utility of the CODESSA-generated descriptors for QSAR studies and as a complementary tool to understand ER binding of estrogens.

Materials and Methods

Data sets for analysis. von Angerer and coworkers (39-41) have sought to develop potent antiestrogens based on nonsteroidal structures. Among their prototype structures are the 2-phenylindoles, for which extensive structure-activity relationship (SAR) studies have been done (39). We selected 53 estrogenic compounds from von Angerer's database as the training set to determine the steric and electrostatic requirements for recognition at the ER binding site. In most previous CoMFA studies of ER binding (27-31), the data sets either consisted of congeneric compounds or were not of good quality, which limited the general applicability of these models to serve as screens for potential estrogens. In the present study, several naturally occurring and synthetic estrogens (Fig. 1) were added to the 2-phenylindoles to increase the structural diversity and to span a larger range of RBA values within the training set. The chemical structures of the 53 compounds, organized by chemical

class, are shown in Figure 1 and Tables 1 and 2. The RBAs were calculated from a calf uterine estrogen receptor (calf ER) competitive binding assay with $[{}^{3}H]17\beta$ -estradiol (E₂). The RBA is the ratio of the molar concentrations of E₂ and the competing chemical required to decrease the receptor-bound radioactivity by 50%, multiplied by 100; thus, E₂ has an RBA of 100.

The test set comprised the 16 estrogenic compounds shown in Figure 2. Among these, the RBA values for 14 steroidal compounds were measured in a competitive ER binding assay using human MCF-7 cell cytosol (42). The binding affinities of 4 of the 16 compounds, estrone, estriol, moxestrol, and zindoxifen, were measured in the same calf ER binding assay used in the training set. Within the test set, only two estrogens, estrone and estriol, were measured in both binding assays.

Molecular modeling. The structures of the 2-phenylindole analogs were constructed from the Sybyl 6.1 fragment database (Tripos Inc., St. Louis, MO). The remaining compounds were built based on the crystal structures of estradiol and DES,

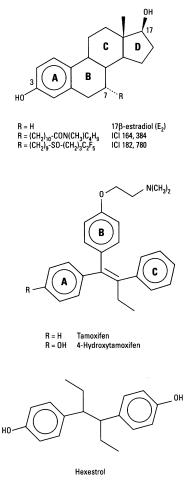
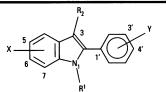


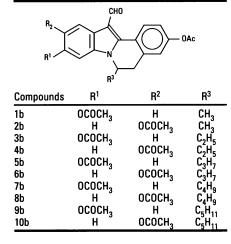
Figure 1. Chemical structures of steroidal and nonsteroidal estrogens in the training set.

 Table 1. Hydroxy-2-(hydroxyphenyl)indole derivatives in the training set

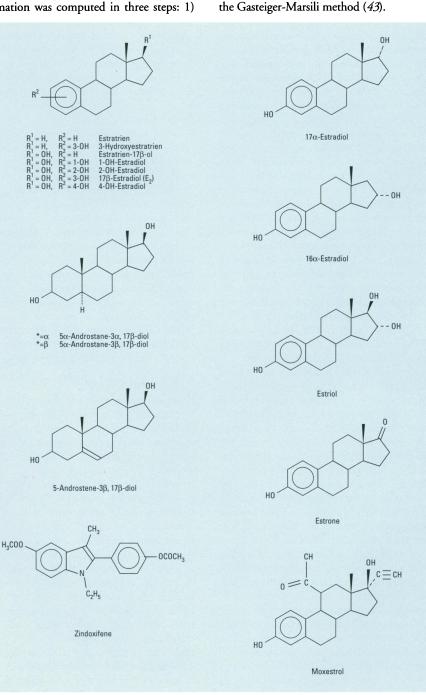


	X,Y=OH	Х,Ү=ОН		Position of	
Compound	s R ¹	R ²	X	Y	
1a	Н	Н	6	4´	
2a	Н	CH3	6	4´	
3a	н	C2H2	6	4′	
4a	Н	Τ̈́Η	5	4′	
5a	Н	CH₃	5	4′	
6a	CH ₃	Нĭ	6	4′	
7a	C₂H₅	н	6	4´	
8a	C ₃ H ₇	н	6	4′	
9a	C₄H₀	н	6	4´	
10a	CH ₃	CH3	6	4′	
11a	C2H2	CH,	6	4´	
12a	U ₂ H ₇	CH.	6	4´	
13a	i-C2H2	CH ₃	6	4′	
14a	СЙ3	C_2H_5	6	4′	
15a	C2H	C_2H_5	6	4′	
16a	C2H2	C_2H_5	6	4′	
17a	CH3	Τ̈́	5	4′	
18a	C,H,	н	5	4′	
19a	C3H2	н	5	4′	
20a	CH3	CH_3	5	4′	
21a	C ₂ H ₅	CH₃	5	4′	
22a	C3H7	CH,	5	4′	
23a	i-C ₃ H ₇	CH.	5	4´	
24a	C₄H ₉	CH ₃	5	4´	
25a	C5H11	CH3	5	4´	
26a	C2H2	C_2H_5	5	4′	
27a	C ₃ H ₇	C ₃ H ₇	5	4′	
28a	C ₂ H ₅	CH3	7	4´	
29a	C₂H₅	НĽ	6	3´	
30a	С́Н ₃	CH_3	6	3′	
31a	C ₂ H ₅	$CH_3^{"}$	6	3′	
32a	C ₃ H ₇	CH₃	6	3′	
33a	C₂H₅	ΗŤ	5	3′	
34a	CH3	CH₃	5	3′	
35a	C ₂ H ₅	CH3	5	3′	
36a	C ₃ H ₇	CH2	5	3′	
ZK119,010	(CH ₂) ₆ N(CH ₂ CH ₂) ₂	CH ₃	5	4´	

Table 2. Diacetoxy-12-formyl-5,6-dihydroindo- $lo[2,1-\alpha]$ isoquinolines in the training set



which were obtained from the Cambridge Structural Database (CSD; University Chemical Laboratory, Cambridge, U.K.). Because a single conformation is assumed in CoMFA studies for a ligand binding to a receptor, a binding conformation of the molecules being studied must be postulated (29). In the absence of experimental evidence on the binding conformation of estrogens, we used the structure based on the global minimum-energy conformation, which is standard practice in CoMFA studies (24). The global minimum-energy conformation was computed in three steps: 1)



the geometry of each molecule was opti-

mized to its nearest local minimum-energy

conformation to an energy gradient of 0.001

kcal/mol Å using the standard Tripos mole-

cular mechanics force field with a distance-

dependent (1/r) dielectric function; 2) these

energy-minimized structures were then sub-

jected to conformational analysis using a sys-

tematic search over all rotatable bonds at

10° increments; and finally, 3) the molecules

were reminimized by setting to their identi-

fied minimum-energy torsion angles. All

atomic partial charges were computed using

Figure 2. Chemical structures of the compounds in the test set.

CoMFA alignment and descriptors. To perform a CoMFA study, the molecules of interest must first be aligned to maximize superposition of their steric and electrostatic fields. The alignment procedure varies from molecule to molecule based on structural similarity or diversity. Most alignment rules employ a least-squares fitting of pharmacophoric elements between a designated template molecule (here E2, Fig. 1) and each of the other molecules in the training set. The pharmacophoric points of E_2 used for alignment were the centroids of the A- and D-rings and the C7 atom of the B-ring. This superposition provides maximal overlap of the rings and their substituents for the 2-phenylindoles (Tables 1, 2), the steroidal antiestrogens, ICI 182,780, ICI 164,384, and the triphenylethylene antiestrogens tamoxifen (TAM) and 4-hydroxytamoxifen (OH-TAM) (Fig. 1). For the 2-phenylindoles, the corresponding alignment points are the centroid of the phenyl ring of the indole moiety, the centroid of the 2-phenyl ring, and the indole nitrogen. The steroidal estrogens in Figure 1 were aligned in a similar manner. For the triphenylethylenes, the equivalent points are the centroids of the A- and C-rings, and the C1 atom of the Bring (or of the ethyl group in hexestrol).

This fitting procedure was followed by a field fit optimization to the template molecule. The field fit adjusts the geometry of the molecules to maximize the similarity of the steric and electrostatic fields between the template and training molecules. Because this procedure sometimes causes structural distortions, the molecules were subsequently reoptimized to relax the fitted molecules to the nearest local minimumenergy structure.

After alignment, the molecules were placed in a three-dimensional cubic lattice with 2 Å spacing. The steric and electrostatic fields were calculated at each mesh point using an sp³ carbon probe with +1.0 charge based on the van der Waals (LJ 12-6 potential) interactions and Coulombic interactions, respectively. The steric and electrostatic energy values were truncated to 30 kcal/mol.

PLS-QSAR. To form the basis for a predictive statistical model, the method of partial least squares (PLS) regression (44) was used to analyze the training set of 53 compounds by correlating their biological activities to the steric and electrostatic fields. In CoMFA studies, the number of steric–electrostatic field descriptors (independent variables, vector X) derived from the CoMFA field calculations is much larger than the number of training set compounds with associated activity data (dependent variables, vector Y). In this situation, PLS is particularly well suited to correlate these field variables with biological activity. PLS reduces the descriptors to a few principal components (PCs), which are linear combinations of the original descriptor variables and, hence, establishes a linear relationship between Y and X through these PCs. In order to determine the optimum number of PCs that yielded the smallest standard error of prediction, the leave-one-out (LOO) cross-validation procedure (45) was used. In this method, each compound is systematically excluded once from the data set, after which its activity is predicted by a model derived from the remaining compounds. Therefore, the cross-validated r^2 (termed q^2 hereafter) can be derived from:

$$q^2 = (SD - PRESS)/SD$$

where PRESS is the sum of squared differences between the actual and predicted activity data for each molecule during LOO cross-validation, and SD is the sum of squared deviations between the measured and mean activities of each molecule in the training set. Based on the optimal number of principal components derived from LOO cross-validation, the final PLS analysis was carried out without cross-validation to generate a predictive QSAR model with a conventional (correlation coefficient) r^2 . The correlation, in turn, was used to plot CoMFA color contours of the steric and electrostatic field characteristics, which offer potential insights into the relevant determinants of biological activity. The derived 3D-QSAR model was then employed to predict the binding affinity values of the 16 compounds in the test set (Fig. 2).

The r^2 and q^2 are two key measures of CoMFA model performance. The r^2 measures the model's goodness of fit to the training set activity data, with a value of r^2 greater than 0.9 normally indicating statistical significance. The q^2 is derived from the LOO cross-validation procedure, in which the stability of the model is tested by perturbing the regression coefficients by consecutively omitting each compound. Consequently, the q^2 is a measure of the robustness of the model, that is, its ability to predict. The q^2 is generally lower than the r^2 , and a model with $q^2 >$ 0.5 is normally considered to have a significant predictive ability (24).

Calculation of CODESSA descriptors. In the present application of the CODESSA methodology, as many as 365 descriptors were calculated for each compound in the training set. These descriptors were in one of the following categories: constitutional, topological, geometrical, electrostatic, and quantum-chemical. Most applications for

these descriptors have so far been for QSPR rather than QSAR studies. The simplest type are the constitutional descriptors, e.g., atom counts, molecular weight, etc., which reflect the molecular composition of the compound without regard to its geometry or electronic structure. For each molecule, topological descriptors include the Kier and Hall, Randic, and Wiener indices, which are most sensitive to molecular connectivity. Geometrical descriptors, such as the moment of inertia, molecular surface area, etc., require the 3D-coordinates of the atoms in the molecule. The electrostatic descriptors reflect the characteristics of charge distribution and can be calculated using one or more nonempirical procedures within the CODESSA program or by any quantumchemical program. Quantum chemical descriptors add important information to the conventional descriptors in terms of the internal electronic properties of molecules, which is otherwise not obtainable. To generate quantum chemical descriptors in the CODESSA program, all Sybyl energy minimized structures were submitted to semiempirical quantum chemical calculation with optimization using the AM1 model Hamiltonian in AMPAC (Semichem, Shawnee, KS). The set of CODESSA molecular descriptors were autoscaled, then LOO cross-validation (45) was carried out to determine the optimum number of PCs. The final QSAR model was determined by means of PLS without cross-validation.

Results

CoMFA-PLS. Table 3 contains a comparison of the RBAs (expressed as log RBA) predicted for the training set compounds using CoMFA-PLS with the corresponding experimental values. The largest log RBA residual is 0.41 (26a and 36a), corresponding to a maximum RBA variation of 2.6fold between calculated and experimental values. A plot of the calculated versus experimental RBAs is given by the open symbols in Figure 3. The correlation has a high r^2 (0.97) and q^2 (0.61), which indicates significant self-consistency and predictive capability. The CoMFA-PLS model required nine principal components to explain the variance in biological activity. The other statistical parameters associated with this model are standard error (SE) = 0.16, F = 130.3, and p<0.001.

The steric and electrostatic field contributions to the CoMFA model were 44% and 56%, respectively. The substantial contribution from the electrostatic fields is consistent with previous CoMFA models of ER binding (29-31). Figure 4 shows a color-coded three-dimensional contour map depicting regions in space around the

molecules where the variations in steric and electrostatic fields are most strongly correlated with variations in log RBA; an E_2 template molecule provides geometrical reference. The green and yellow polyhedra describe regions in space where increased steric bulk increases or decreases the RBA, respectively. Clearly, ligand–ER binding is sensitive to the length of the substituents at the indole N-1 position and at the steroid B-ring. This position is surrounded by an inner green region indicating that the presence of steric bulk in this region, such as the ethyl substituents in compounds 7a or 11a, will enhance RBA. Beyond the green region is a yellow region where additional steric bulk, such as in compounds 9a or 24a, will diminish RBA.

The red and blue polyhedra describe regions of space where an increase in negative charge is associated with increased and decreased RBA, respectively. The blue con-

 Table 3. Observed and calculated relative binding affinities (RBAs) for the calf estrogen receptor of the training set for both CoMFA and CODESSA models

		CoMFA-PLS		CODESSA-PLS	
Compounds	Experimental log RBA	Calculated log RBA	Residual	Calculated log RBA	Residual
1a	-2.00	-2.21	0.21	-2.15	0.15
2a	-1.22	-1.12	-0.10	-1.22	0.0
3a	-0.89	-1.07	0.18	-0.59	-0.30
4a	-2.00	-1.80	-0.20	-2.17	0.17
5a	-1.22	-1.01	-0.21	-1.19	-0.03
6a	0.58	0.53	0.05	-0.05	0.63
7a	1.20	1.21	-0.01	0.55	0.65
8a	0.93	0.95	-0.02	0.70	0.23
9a	0.63	0.70	-0.06	0.82	-0.19
10a	1.00	0.80	0.20	0.40	0.60
11a	1.52	1.39	0.13	0.67	0.85
12a	1.11	1.11	0.0	0.84	0.28
13a	1.11	0.86	0.25	0.75	0.36
14a	0.77	0.81	-0.04	0.68	0.09
15a	1.32	1.30	0.02	0.83	0.50
16a	1.28	1.34	-0.06	0.96	0.32
17a	-0.10	0.13	-0.23	0.00	-0.10
18a	0.76	0.73	0.04	0.45	0.31
19a	1.25	0.86	0.40	0.67	0.58
20a	0.66	0.60	0.06	0.35	0.32
21a	0.98	0.93	0.05	0.65	0.33
22a	1.20	1.24	-0.03	0.79	0.42
23a	0.54	0.77	-0.23	0.80	-0.25
24a	0.66	0.74	-0.08	0.89	-0.23
25a	0.36	0.58	-0.22	0.84	-0.48
26a	1.36	0.96	0.41	0.81	0.56
27a	0.23	0.55	-0.32	0.91	-0.68
28a	-1.70	-1.74	0.04	0.58	-2.28
29a	0.23	0.27	-0.04	0.55	-0.32
30a	-0.26	-0.36	0.10	0.36	-0.62
31a	0.48	0.81	-0.33	0.67	-0.19
32a	0.54	0.74	-0.20	0.88	-0.34
33a	0.23	0.25	-0.02	0.46	-0.23
34a	-0.22	-0.13	-0.09	0.33	-0.55
35a	0.34	0.30	0.05	0.69	-0.34
36a	0.87	0.46	0.03	0.83	0.04
ZK 119,010	1.32	1.51	-0.19	0.85	0.47
1b	0.23	0.28	-0.05	0.02	0.21
2b	0.59	0.62	-0.03	0.26	0.33
3b	0.33	0.14	-0.03	0.20	-0.09
4b	0.52	0.56	-0.04	0.47	0.05
5b	-0.30	-0.25	-0.05	0.28	-0.58
6b	0.42	0.25	0.05	0.36	0.06
7b	-0.15	-0.04	-0.11	0.30	-0.46
8b		0.26	0.06	0.38	-0.40
9b	0.32 -0.22	-0.26	0.00	0.30	-0.00
10b	0.22	0.17	0.04	0.43	-0.55
	2.00	1.95	0.05	1.73	-0.15
E ₂ ICI 164,384	0.75	0.67	0.05	0.82	-0.07
ICI 164,384 ICI 182,780	0.79		0.08	0.54	
TAM	-0.44	0.72			0.25
		-0.51	0.07	-0.08	-0.36
OH-TAM	0.83	1.00	-0.17	0.27	0.56
Hexestrol	1.34	1.41	-0.06	1.48	-0.13

tour surrounding the 3'-position of the phenyl moiety of the 2-phenylindoles suggests that the presence of the OH substituent at this position decreases RBA and gives a rational explanation for the lower RBA of compounds 29a to 36a compared with the corresponding analogs where substitution is at the 4'-position (compounds 7a, 10a–12a, 18a, and 20a–22a).

The CoMFA model so derived was used to predict the activity of the test compounds shown in Figure 2. Since four compounds (estrone, estriol, moxestrol, and zindoxifen) were assayed identically to those in the training set, they provide an additional means to assess the predictive significance of the model. There is very good agreement between the actual and predicted RBA values for these four test compounds with small residuals similar to that obtained for the training set results (Fig. 3 and Table 4). The model predicted estrone, estriol, and moxestrol to have moderate binding affinities, while zindoxifen is predicted to have relatively weak receptor binding.

As seen in Table 4, two of the four test compounds, estrone and estriol, were assayed with both calf and human ER. The RBA values obtained from both receptor assays are comparable. The availability of these data prompted us to take our model a step further by exploring the interspecies relationship for estrogen receptor binding. Hence our calf ER-derived CoMFA model was used to predict calf ER RBAs for the 14 compounds in Table 4 that had available RBAs for human ER. A comparison of the predicted calf ER RBA versus experimental human ER shows a positive correlation with a coefficient of 0.80. Compounds that bind poorly or show no binding affinity to human ER (e.g., estratrine, androstane) were predicted to bind similarly with the calf ER, as were compounds that had moderate binding affinity with human ER.

CODESSA-PLS. The CODESSA-PLS model for the same set of 53 compounds required three principal components to explain the variance in biological activity. As in CoMFA-PLS, the optimum number of components was determined using the LOO cross-validation procedure. The key statistical parameters for this model are $q^2 = 0.54$, $r^2 = 0.68$, SE = 0.55, F = 30.3, and p<0.001. The CODESSA results for the calculated log RBAs are likewise given in Table 3. A plot of the experimental versus calculated RBA values for the training set compounds is shown in Figure 5. It can be seen that there is an obvious outlier in compound 28a. Removing this compound from the training set improved the conventional r^2 (0.80). Table 1 shows that compound 28a is the only one

Abbreviations: PLS, partial least squares; E₂, 17β-estradiol; TAM, tamoxifen; OH-TAM, 4-hydroxytamoxifen.

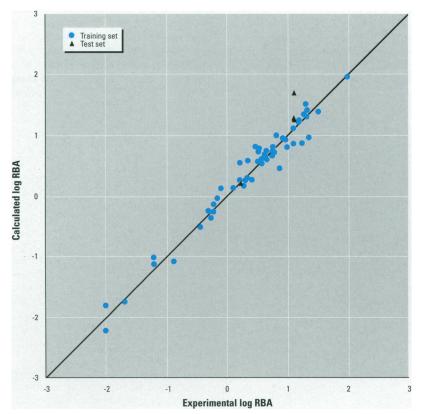


Figure 3. Plot of observed log relative binding affinity (log RBA) versus CoMFA-calculated log RBA for calf estrogen receptor. $r^2 = 0.97$; $q^2 = 0.61$. Percentage contribution from steric, 44%; from electrostatic, 56%.

in which the OH-substituent is at the 7position. It appears that the difference in the position of this functional group, while possibly encoded in the descriptors, cannot be easily distinguished by the CODESSA methodology. Indeed, for compounds 11a, 21a, and 28a, for example, where only the position of substituent X is changing, i.e., positions 6, 5, and 7, respectively, the calculated log RBA values are close to each other (calculated log RBA = 0.67, 0.65, and 0.58, respectively). On the contrary, the experimental RBA values for these compounds significantly decrease as the OH-substituent is moved from the 6-, 5-, and 7-positions (experimental log RBA = 1.52, 0.98, and -1.70, respectively).

The CODESSA-PLS methodology gives the PLS X-loadings, which were examined in order to delineate the relative contribution of each molecular descriptor to the regression model. The molecular descriptors with highly positive or highly negative PLS loadings (regarded as the most important) were associated with the quantum chemical and electrostatic descriptors that encode the features responsible for polar interactions between molecules. This is consistent with the CoMFA results in which the electrostatic field had a greater contribution to estrogenic activity than did steric fields.
 Table 4. Comparison of experimental log relative binding affinity (log RBA) for both the calf estrogen receptor (ER) and the human ER and the predicted log RBA obtained from CoMFA and CODESSA models

Number		Experimental log RBA		Predicted log RBA	
	Compounds	Human ER	Calf ER	CoMFA-PLS	CODESSA-PLS
1	17β-Estradiol	2.00	2.00	1.95	1.75
2	17α-Estradiol	1.34		1.60	1.75
3	16α-Estradiol	1.90		0.89	1.78
4	1-OH-Estradiol	-0.30		0.31	1.62
5	2-OH-Estradiol	1.85		0.95	2.18
6	4-OH-Estradiol	0.84		1.50	2.05
7	3-Hydroxyestratrien	1.90		1.27	1.16
8	Estratrien-17B-ol	1.04		0.51	1.21
9	Estratrien	<-1.30		-0.17	-0.10
10	Estrone	1.34	1.11	1.68	0.94
11	Estriol	1.23	1.11	1.24	1.70
12	5-Androstene-3β,17β-diol	-0.16		-0.67	1.99
13	5α-Androstane-3β,17β-diol	-0.30		-0.15	1.94
14	5α-Androstane-3α,17β-diol	<-1.30		-0.35	1.82
15	Moxestrol		1.11	1.27	1.37
16	Zindoxifen		0.23	0.20	0.30

PLS, partial least squares.

While the CODESSA descriptors are able only to explain 68% of the variation in the activity, the model is potentially capable of predicting the activity of compounds outside the training set, as demonstrated by the significant q^2 of 0.54. Figure 5 shows that there is good agreement of the predicted values with the experimental RBA in calf uterine cytosol for estrone, estriol, moxestrol, and zindoxifen.

As we already observed in the training set, the CODESSA-PLS model appears relatively insensitive to the position of substituents. This inherent limitation manifests itself in the predictions for compounds 1-6(Table 4), which differ from one another only by the position of the hydroxyl group. The RBA values predicted by CODESSA-PLS for all these compounds are not significantly different from each other. However,

Figure 4. CoMFA steric and electrostatic contour plots for binding affinity with the structure of 17β -estradiol shown inside the field. Enhanced rela-

tive binding affinity (i.e., higher binding affinity) is associated with adding/subtracting steric bulk from green/yellow regions and with adding/subtracting positive charge in the blue/red regions. The absence of color represents regions that

were unexamined by the current data set.

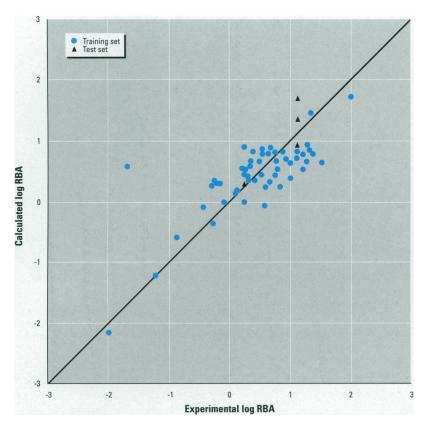


Figure 5. Plot of experimental versus CODESSA calculated log relative binding afffinity (log RBA) values for calf estrogen receptor (ER). $r^2 = 0.68$; $q^2 = 0.54$.

the experimental RBA values for compounds 1-OH-estradiol and 4-OH-estradiol (compounds 4 and 6) are much lower than the RBA for E_2 . Another notable disparity pertains to the androgenic compounds (compounds 12, 13, and 14 in Table 4), whose structures are very similar to E_2 except that the A-ring is not aromatic. Although these compounds have poor binding affinities, their predicted RBAs are close to that of E_2 . This prediction of abnormally high RBAs may be likely for these analogs because a structure with a nonaromatic A-ring was not represented in the training set.

Discussion

For this series of estrogenic chemicals, CoMFA demonstrated a significant correlation of its calculated steric and electrostatic fields with RBA and provided predictions in good agreement with experimental values, for calf ER RBAs. Calculated RBAs differed by less than 2.6-fold from experimental values, which traversed a 10,000fold range. The CoMFA coefficient contour plots revealed a consistent pattern of the chemical features largely responsible for the variations in RBA with E_2 as a template molecule (Figure 4). These features generally fall within three regions around the A, B, and D rings. The upper middle position of the molecule was relatively insensitive to structural changes. For the 2-phenylindoles, this result is consistent with previous observations that short alkyl groups in the indole N-1 position increase binding while there is a decrease in affinity with compounds containing more than three carbon atoms in the indole N-1 side chain (39).

The present results demonstrate the potential capability of QSAR models to rapidly assess the potential for compounds to bind to the ER. This capability clearly distinguishes the QSAR model from the actual measurement of RBAs, in vitro reporter gene and cell proliferation activities, or in vivo endpoints. Screening chemicals for their RBAs by use of QSAR models shows promise in terms of reducing the enormous time and expense of testing each and every compound exhaustively. As mentioned above, we recognize that those compounds identified by QSAR as binding to the ER will still require further experimental confirmation by laboratory studies. For this reason, QSAR models must be considered as only one component of a comprehensive screening and testing system that will determine the estrogenic impact of a compound on a particular species. It is hoped that QSAR models can serve as fast and reliable screens to identify those chemical structures most likely to exhibit estrogenic activity and to prioritize compounds for subsequent biological testing on the basis of their predicted RBAs. For these reasons, the level of precision required from a QSAR model will vary depending on the intended purpose.

We established a correlation between the predicted RBA values for calf ER and their actual RBA values for human ER. The largest difference observed between the two estrogen receptor sources is that for compound 16*a*-estradiol, whose predicted binding affinity is 10-fold below that found with human ER. This deviation is not strikingly larger than the 2.6-fold error in the calf ER RBA calculation. This may be due to species difference or in part to the lack of information in the training set regarding substituents at the 16-position of the steroid structure. This finding reinforces the necessity of having a diverse set of training molecules and suggests a rational way to evaluate which additional data would strengthen the model.

While further validation (by expanding the number of compounds) is needed, the satisfactory correlation across species indicates only a limited range of variability of RBAs for these two receptor sources. This is almost certainly a consequence of the well-defined evolutionary conservation of the ER primary structure (amino acid sequence) across a variety of species and the broader conservation of primary structure across a number of receptors that constitute the steroid hormone receptor superfamily (46). For example, there is only a sevenamino acid difference in the primary structures of ligand-binding domains between the human and murine ER (47, 48).

Prediction of chemical binding to the ER is important for regulatory purposes. Use of literature data sets obtained with comparable assays across species increases the number of chemical structures contributing to the model. Importantly, this may generate a set of metrics that improve interspecies extrapolation, a significant problem in risk assessment. This approach is therefore relevant to the development of tools to identify potential endocrine disruptors for subsequent evaluation, which, by virtue of wide environmental exposure, may act in a large number of animal species.

The QSAR models described here were developed to predict ER RBAs for numerous chemicals that have not been examined for estrogenic activity. In particular, legislation passed in 1996 requires the EPA to define a testing scheme for estrogenicity within 2 years (10, 11). Hence, models such as these offer a means to prioritize among a large number of chemicals, based on their potential as ER ligands, for testing in more expensive assays arranged in a tiered battery, with the most definitive *in vivo* tests at the apex. For this initial screening purpose, criteria must be developed for decision points for RBA values. This is not as straightforward as it may seem because other events, such as serum binding or metabolism, can alter potency. For example, in developing animals, some xenoestrogens show greatly increased potency relative to E_2 due to the presence of high affinity serum binding proteins for E_2 , which bind xenoestrogens with much lower affinity (49).

Regardless of the criteria applied to RBA values to define a chemical as estrogenic, it should be appreciated that perfect predictivity of RBAs is not required at the initial screening step. Rather, the incidence of false negatives (actual ER binding chemicals that appear negative in the screen) is of greatest concern, as these may not be further tested in a timely way. False positives (classifying a chemical as an ER ligand if the chemical actually does not bind) are of much less concern; they would be eliminated at higher tiers in the test battery. We are designing additional approaches that can be used in conjunction with the QSAR models such as described here to help in determining criteria for selecting chemicals for more extensive testing.

Our analysis indicates that the COMFA model is superior in precision to the CODESSA model for the present application. It appears that the CODESSA descriptors as implemented in the PLS procedure were unable to capture those factors that influence the variations in RBA and, therefore, limit the correlation between the RBAs and certain structural differences in the training set. It may also be possible that a nonlinear relationship exists between the descriptors and RBAs, which can be addressed by other nonlinear regression methods such as artificial neural networks.

Compared with CODESSA-PLS, CoMFA appears better able to explain estrogen receptor binding variation in terms of the steric and electrostatic requirements. However, when used in appropriate cases, CODESSA-PLS is certainly less prone to arbitrariness (e.g., alignment) and is procedurally less difficult to implement. Importantly, the level of precision with CoMFA seems more than adequate for the intended purpose, while the range of predictability of RBAs at 10,000-fold is quite large. Further development of this model is currently underway.

In summary, we have shown the feasibility of using computational methods to predict the RBAs of chemicals untested for estrogenic activity. Such predictions should allow selection of the highest priority chemicals for testing, resulting in information on estrogenic activity to be obtained more quickly than if no priority criteria existed.

REFERENCES

- 1. McLachlan JA, ed. Estrogens in the Environment. II: Influences on Development. New York:Elsevier, 1985.
- Mishell JR. Noncontraceptive benefits of oral contraceptives. J Reprod Med 38(suppl 12):1021–1029 (1993).
- Barrett-Connor E. Risk and benefits of replacement estrogen. Annu Rev Med 43:239–251 (1992).
- 4. Herbst AL, Bern HA. Developmental Effects of Diethylstilbestrol (DES) in Pregnancy. New York:Thieme-Stratton Inc., 1981.
- Colburn T, Clement C. Chemically-induced Alterations in Sexual and Functional Development: The Wildlife/Human Connection. Princeton, NJ: Princeton Scientific Publishing, 1992.
- Colburn T, vom Saal FS, Soto AM. Developmental effects of endocrine-disrupting chemicals in wildlife and humans. Environ Health Perspect 101(5):378–384 (1993).
- Hileman B. Environmental estrogens linked to reproductive abnormalities, cancer. Chem Eng News 72:19–23 (1994).
- Kavlock RJ, Daston GP, DeRosa C, Fenner-Crisp P, Gray LE, Kaattari, Lucier G, Luster M, Mac MJ, Maczka C. 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).
- Davis DL, Bradlow HL, Wolff M, Woodruff T, Hoel DG, Anton-Culver H. Medical hypothesis: xenoestrogens as preventable causes of breast cancer. Environ Health Perspect 101:372–377 (1993).
- Compilation of Laws Enforced by the U.S. Food and Drug Administration and Related Statutes. Vol 2. Washington, DC:U.S. Government Printing Office, 1996.
- 11. Safe Drinking Water Act Amendments of 1996. Public Law 104-182. 104th Congress, 1996.
- 12. Katzenellenbogen JA. The structural pervasiveness of estrogenic activity. Environ Health Perspect 103(suppl 7):99–101 (1995).
- Korach KS, McLachlan JA. Techniques for detection of estrogenicity. Environ Health Perspect 103(suppl 7):5-8 (1995).
- 14. Branham WS, Zehr DR, Sheehan DM. Differential sensitivity of rat uterine growth and epithelium hypertrophy to estrogens and antiestrogens. Proc Soc Exp Biol Med 203:297–303 (1993).
- Medlock KL, Branham WS, Sheehan DM. Effects of coumestrol and equol on the developing reproductive tract of the rat. Proc Soc Exp Biol Med 208:67–71 (1995).
- Miksicek RJ. Estrogenic flavonoids: structural requirements for biological activity. Proc Soc Exp Biol Med 208:44–50 (1995).
- 17. Soto AM, Sonnenschein C, Chung KL, Fernandez MF, Olea N, Serrano FO. The E-SCREEN assay as a tool to identify estrogens: an update on estrogenic enviromental pollutants. Environ Health Perspect 103(suppl 7):113–122 (1995).

- Liskowski L, Rose DP. Experience with a simple method for estrogen receptor assay in breast cancer. Clin Chim Acta 67:175–182 (1976).
- 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).
- Fujita T, Iwasa J, Hansch C. A new substituent constant, π, derived from partition coefficient. J Am Chem Soc 86:5175–5180 (1964).
- 21. Hansch C, Leo A. Exploring QSAR— Fundamentals and Applications in Chemistry and Biology. Washington, DC:The American Chemical Society, 1995.
- 22. Bradbury SP. Quantitative structure-activity relationship and ecological risk assessment: an overview of predictive aquatic toxicology research. Toxicol Lett 79:229–237 (1995).
- Hansch C, Telzer BR, Zhang L. Comparative QSAR in toxicology: examples from teratology and cancer chemotherapy of aniline mustards. Critical Rev Toxicol 25(1):67–89 (1995).
- Cramer RD, Patterson DE, Bunce JD. Comparative molecular field analysis (CoMFA).
 Effect of shape on binding of steroids to carrier proteins. J Am Chem Soc 110:5959–5967 (1988).
- 25. Tong W, Collantes E, Chen Y, Welsh WJ. A comparative molecular field analysis study of *N*benzylpiperidines as acetylcholinesterase inhibitors. J Med Chem 39(2):380–387.
- Collantes E, Tong W, Welsh WJ. Use of moment of inertia in comparative molecular field analysis to model chromatographic retention of nonpolar solutes. Anal Chem 68:2038–2043 (1996).
- Waller CL, Minor DL, McKinney JD. Using three-dimensional quantitative structure-activity relationships to examine estrogen receptor binding affinities of polychlorinated hydroxybiphenyls. Environ Health Perspect 103: 702-707 (1995).
- Waller CL, Oprea TI, Chae K, Park HK, Korach KS, Laws SC, Wiese TE, Kelce WR, Gray LE. Ligand-based identification of environmental estrogens. Chem Res Toxicol 9:1240–1248 (1996).
- Bradbury SP, Mekenyan OG, Ankley GT. Quantitative structure-activity relationships for polychlorinated hydroxybiphenyl estrogen receptor binding affinity—an assessment of conformer flexibility. Environ Toxicol Chem 15:1945–1954 (1996).
- Gantchev TG, Ali H, van Lier JE. Quantitative structure-activity relationships/comparative molecular field analysis (QSARs/CoMFA) for receptor-binding properties of halogenated estradiol derivatives. J Med Chem 37:4164–4176 (1994).
- Tong W, Perkins R, Xing L, Welsh WJ, Sheehan DM. QSAR models for binding of estrogenic compounds to estrogen receptor α and β subtypes. Endocrinology 138:4022–4025 (1997).
- Loughney DA, Schwender CF. A comparison of progestin and androgen receptor binding using the CoMFA technique. J Comput-Aided Mol Des 6:569–581 (1992).
- Oprea TI, Garcia AE. Three-dimensional quantitative structure-activity relationship of steroid aromatase inhibitors. J Comput-Aided Mol Des 10:186–200 (1996).
- 34. Waller CL, Juma BW, Gray LE, Kelce WR. Three-dimensional quantitative structure–activity relationships for androgen receptor ligands. Toxicol Appl Pharmacol 137:219–227 (1996).

- 35. Welsh WJ, Tong W, Collantes ER, Chickos JS, Gagarin SG. Enthalpies of sublimition and formation of polycyclic aromatic hydrocarbons (PAHs) derived from comparative molecular field analysis (CoMFA): application of moment of inertia for molecular alignment. Thermochem Acta 290:55–64 (1996).
- 36. Katritzky AR, Mu L, Lobanov VS, Karelson M. Correlation of boiling points with molecular structure. 1. A training set of 298 diverse organics and a test set of 9 simple inorganics. J Phys Chem 100:10400–10407 (1996).
- 37. Katritzky AR, Ignatchenko ES, Barcock RA, Lobanov VS. Prediction of gas chromatographic retention times and response factors using a general quantitative structure–property relationship treatment. Anal Chem 66:1799–1807 (1994).
- 38. Katritzky AR, Rachwal P, Law KW, Karelson M, Lobanov VS. Prediction of polymer glass transition temperatures using a general quantitative structure–property relationship treatment. J Chem Inf Comput Sci 36:879–884 (1996).
- 39. von Angerer E, Prekajac J, Strohmeier J. 2-Phenylindoles. Relationship between structure,

estrogen receptor affinity, and mammary tumor inhibiting activity in the rat. J Med Chem 27:1439–1447 (1984).

- 40. Polossek T, Ambros R, von Angerer S, Brandl G, Mannschreck A, von Angerer E. 6-Alkyl-12formylindolo[2,1-*a*]isoquinolines. Syntheses, estrogen receptor binding affinities, and stereospecific cytostatic activity. J Med Chem 35:3537–3547 (1992).
- 41. von Angerer E, Biberger C, Holler E, Koop R, Leichtl S. 1-Carbamoylalkyl-2-phenylindoles: relationship between side chain structure and estrogen antagonism. J Steroid Biochem Mol Biol 49:51–62 (1994).
- VanderKuur JA, Wiese T, Brooks SC. Influence of estrogen structure on nuclear binding and progesterone receptor induction by the receptor complex. Biochem 32:7002–7008 (1993).
- Gasteiger J, Marsili M. Iterative partial equalization of orbital electronegativity—a rapid access to atomic charges. Tetrahedron 36: 3219–3228 (1980).
- 44. Wold S, Albano C, Dunn WJ III, Edlund U, Esbensen K, Geladi P, Hellberg S, Johansson E,

Lindberg W, Sjostrom M. Multivariate data analysis in chemistry. In: Chemometrics: Mathematics and Statistics in Chemistry (Kowlaski B, ed). Dordrecht, The Netherlands: Reidel, 1984.

- 45. Cramer RD III, Bunce JD, Patterson DE. Crossvalidation, bootstrapping, and partial least squares compared with multiple regression in conventional QSAR studies. Quant Struct-Act Relat 7:18–25 (1988).
- Beato M. Gene regulation by steroid hormones. Cell 56:335–344 (1989).
- Green S, Walter P, Kumar V, Krust A, Bornert JM, Argos P, Chambon P. Human oestrogen receptor cDNA: sequence, expression and homology to v-erb-A. Nature 320:134–139 (1986).
- Greene GL, Gilna P, Waterfield M, Baker A, Hort Y, Shine J. Sequence and expression of human estrogen receptor complementary DNA. Science 231:1150–1154 (1986).
- Sheehan DM, Branham WS. Dissociation of estrogen-induced uterine growth and ornithine decarboxylase activity in the postnatal rat. Teratog Carcinog Mutagen 7:411–422 (1987).

IOAVAILABILIT

IBC's International Congress on Human Health **BIOAVAILABILITY**

Quantifying the Real Toxicity of Common Soil Contaminants

DECEMBER 11 & 12, 1997

RADISSON RESORT

SCOTTSDALE, ARIZONA

- Quantitative Use of Bioavailability Data in Risk Assessments
- Developing and Negotiating the Use of Bioavailability Adjustments
- Framework for a Proposed Good Bioavailability Practice
- Measuring Dermal Bioavailability of Dioxin, PCBs, and PAHs
- In Vitro Methods for Assessing the Bioavailability of Inorganics
- Validation of In Vitro Testing
- Methods to Evaluate the Systemic Bioavailability of PAHs
- Replacing Animal Studies with Synthetic Biofluids
- Phosphate Treatability Studies for Reducing Bioavailability

To register or request exhibiting information, contact IBC today.

IBC USA Conferences Inc. 225 Turnpike Road Southborough, MA 01772-1749 USA Telephone: (508) 481-6400 Fax: (508) 481-7911 E-mail: inq@ibcusa.com On-line: http://www.ibcusa.com/conf/bioavailability