Ligand-Based Identification of Environmental Estrogens

ARTICLE *in* CHEMICAL RESEARCH IN TOXICOLOGY · JANUARY 1997

Impact Factor: 3.53 · DOI: 10.1021/tx960054f · Source: PubMed

Chris L. Waller
Merck
45 PUBLICATIONS **2,505** CITATIONS
SEE PROFILE

Kenneth S Korach
National Institute of Environmental Health S…
456 PUBLICATIONS **35,225** CITATIONS
SEE PROFILE

Susan Laws
United States Environmental Protection Age…
53 PUBLICATIONS **4,186** CITATIONS
SEE PROFILE

Elena Wiese
Johannes Gutenberg-Universität Mainz
46 PUBLICATIONS **1,188** CITATIONS
SEE PROFILE

Ligand-Based Identification of Environmental Estrogens

Chris L. Waller,*,[†] Tudor I. Oprea,[‡] Kun Chae,[§] Hee-Kyoung Park,[§]
Kenneth S. Korach,[§] Susan C. Laws,[||] Thomas E. Wiese,[⊥] William R. Kelce,[||] and
L. Earl Gray, Jr.[||]

Experimental Toxicology and Reproductive Toxicology Divisions, National Health and Environmental Effects Research Laboratory, United States Environmental Protection Agency, Research Triangle Park, North Carolina 27711, Theoretical Biology and Biophysics, Los Alamos National Laboratory, Los Alamos, New Mexico 87545, Laboratory of Reproductive and Developmental Toxicology, National Institute of Environmental Health Sciences, Research Triangle Park, North Carolina 27709, and Curriculum in Toxicology, University of North Carolina, Chapel Hill, North Carolina 27599

Received March 25, 1996[⊗]

Comparative molecular field analysis (CoMFA), a three-dimensional quantitative structure−activity relationship (3D-QSAR) paradigm, was used to examine the estrogen receptor (ER) binding affinities of a series of structurally diverse natural, synthetic, and environmental chemicals of interest. The CoMFA/3D-QSAR model is statistically robust and internally consistent, and successfully illustrates that the overall steric and electrostatic properties of structurally diverse ligands for the estrogen receptor are both necessary and sufficient to describe the binding affinity. The ability of the model to accurately predict the ER binding affinity of an external test set of molecules suggests that structure-based 3D-QSAR models may be used to supplement the process of endocrine disruptor identification through prioritization of novel compounds for bioassay. The general application of this 3D-QSAR model within a toxicological framework is, at present, limited only by the quantity and quality of biological data for relevant biomarkers of toxicity and hormonal responsiveness.

Introduction

As scientific and public concern heightens over the potential of common environmental contaminants to disrupt the endocrine system, the need for a rapid and sensitive screening technique becomes apparent. Endocrine disruption may encompass a variety of mechanisms including but not limited to alterations in circulating hormone levels via disruption of steroid synthesis and plasma binding, competition with endogenous ligands for binding to steroid hormone receptors (*1*), inhibition of the receptor protein dimerization (*1*−*3*), and modulation of the dimer complexes ability to bind to the proper response element of DNA (*1, 3, 4*). Physical manifestations of these events range from inhibition of sex differentiation during development to potential involvement in testicular, prostate, uterine, and/or breast cancer. While many compounds spanning a variety of chemical classes have been examined for their endocrine disruption potential in *in vivo* and *in vitro* bioassays, to date quantitative structure−activity relationship (QSAR)[1] models for the structure-based identification of these compounds remain sparse. Although several QSAR models have been developed for binding to steroid hormone receptors (i.e., estrogen (*5*), progesterone (*6*), and androgen(*6*)), to plasma transport proteins (*7*), and to enzymes which modulate biosynthesis (*8*), the utility of many of these models as general toxicity prediction tools has been limited by the lack of structural diversity in the training set in that the training sets comprise congeneric molecules.

This study utilizes a series of natural and synthetic estrogen receptor (ER) ligands from several chemical classes (Figures 1−8) to construct a three-dimensional QSAR (3D-QSAR) model in order to overcome this deficiency. The extensively validated 3D-QSAR technique of comparative molecular field analysis (CoMFA) (*7*) was used to develop this model which correlates structural differences in molecules with their ability to compete for binding to the ER. While the CoMFA paradigm has its origins in the computer-aided drug design arena (*9*−*12*), we are now beginning to realize the utility of this technique in the process of hazard identification especially with toxicological mechanisms mediated by ligand−receptor protein interaction. We have previously published 3D-QSAR models for dioxin (*13*), estrogen (*5*), and androgen receptor ligands (*14*) which successfully demonstrate the capabilities of 3D-QSAR models of this type to predict receptor binding affinities of environmental contaminants.

The underlying hypothesis for 3D-QSAR models of receptor binding is that all molecules interact with the receptor at the same binding site in the same (or similar) mode. Therefore, it follows that a common pattern of atoms or functional groups in receptor ligands is required to facilitate the molecular recognition and binding processes (i.e., pharmacophore hypothesis). By superimposing the pharmacophoric elements of a series of known

* Author to whom correspondence should be addressed. (919) 541-7976 (phone), (919) 541-5394 (fax), waller@thor.herl.epa.gov (Internet e-mail).

[†] Experimental Toxicology Division, US EPA.

[‡] Los Alamos National Laboratory.

[§] National Institute of Environmental Health Sciences.

[||] Reproductive Toxicology Division, US EPA.

[⊥] University of North Carolina.

[⊗] Abstract published in *Advance ACS Abstracts,* September 15, 1996.

[1] Abbreviations: QSAR = quantitative structure−activity relationship, ER = estrogen receptor, 3D-QSAR = three-dimensional quantitative structure−activity relationship, CoMFA = comparative molecular field analysis, SEAL = steric and electrostatic alignment, DES = diethylstilbestrol, MSS = molecular spreadsheet, HINT = hydropathic interaction, DDT = 2,2-bis(*p*-chlorophenyl)-1,1,1-trichloroethane, PCB = polychlorinated biphenyl, HPTE = 2,2-bis(*p*-hydroxyphenyl)-1,1,1-trichloroethane, PLS = partial least squares, LOO = leave-one-out, PC = principal component, SEP = standard error of crossvalidated predictions, STDDEV*COEFF = standard deviation times coefficient.

S0893-228x(96)00054-9 CCC: $12.00

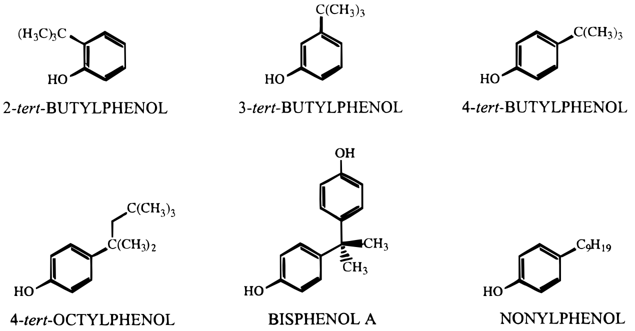

Figure 1. Structures of compounds in phenol subset.

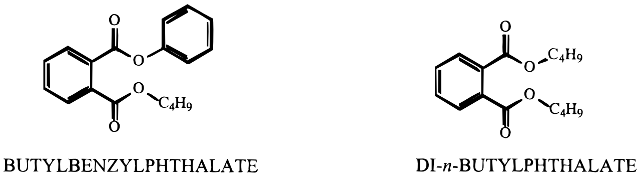

Figure 2. Structures of compounds in phthalate subset.

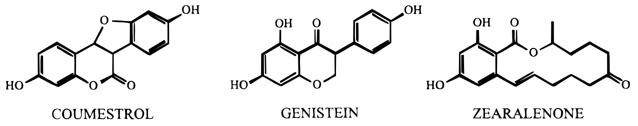

Figure 3. Structures of compounds in phytoestrogen subset.

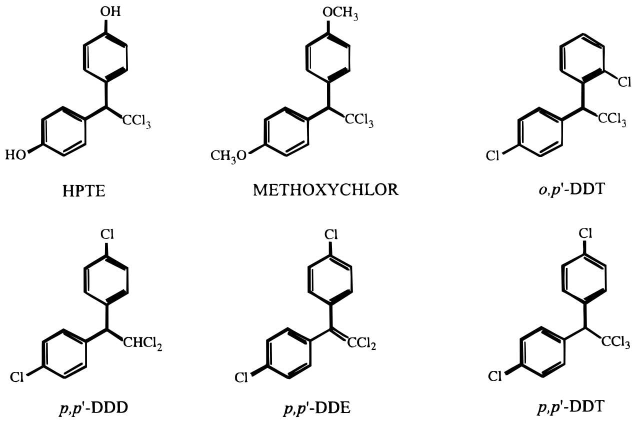

Figure 4. Structures of compounds in DDT subset.

ligands, it is possible to identify structural features which contribute to or detract from receptor-binding affinity. An automated molecular steric and electrostatic field alignment technique, SEAL (*15, 16*), was utilized in this study to identify the optimal alignment for the structurally diverse molecules in this study. The SEAL technique has provided a means by which a variety of putative steroid receptor ligands can be quickly aligned without detailed notions of molecular recognition on the part of the investigator.

The CoMFA model of this study may have significant utility as a three-dimensional data base searching tool (*17*), to the extent that large data bases of candidate molecules can now be rapidly screened and prioritized for compounds that should undergo subsequent biological and/or toxicological evaluation as estrogen acting endocrine discruptors.

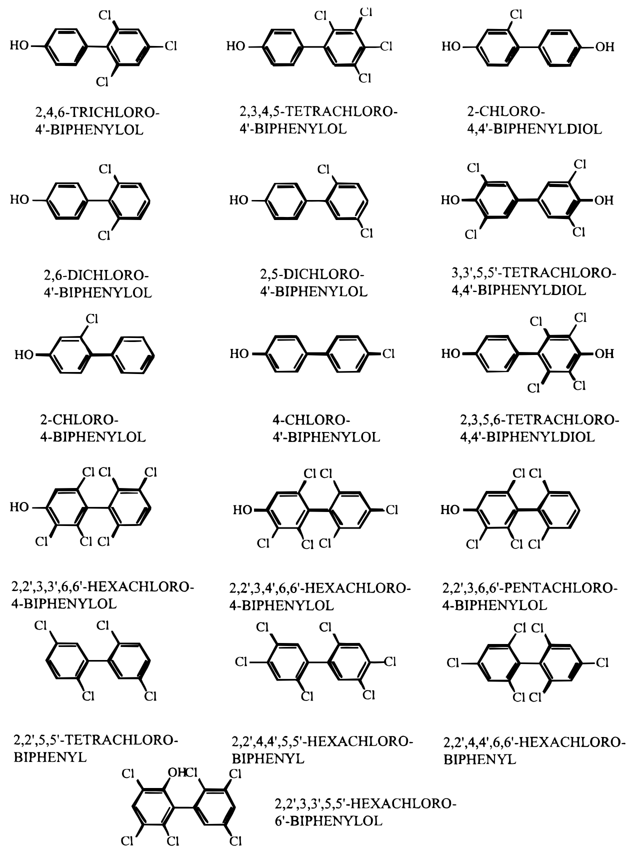
2,4,6-TRICHLORO-4'-BIPHENYLOL
2,3,4,5-TETRACHLORO-4'-BIPHENYLOL
2-CHLORO-4,4'-BIPHENYLDIOL
2,6-DICHLORO-4'-BIPHENYLOL
2,5-DICHLORO-4'-BIPHENYLOL
3,3',5,5'-TETRACHLORO-4,4'-BIPHENYLDIOL
2-CHLORO-4-BIPHENYLOL
4-CHLORO-4'-BIPHENYLOL
2,3,5,6-TETRACHLORO-4,4'-BIPHENYLDIOL
2,2',3,3',6,6'-HEXACHLORO-4-BIPHENYLOL
2,2',3,4',6,6'-HEXACHLORO-4-BIPHENYLOL
2,2',3,6,6'-PENTACHLORO-4-BIPHENYLOL
2,2',5,5'-TETRACHLORO-BIPHENYL
2,2',4,4',5,5'-HEXACHLORO-BIPHENYL
2,2',4,4',6,6'-HEXACHLORO-BIPHENYL
2,2',3,3',5,5'-HEXACHLORO-6'-BIPHENYLOL

Figure 5. Structures of compounds in PCB subset.

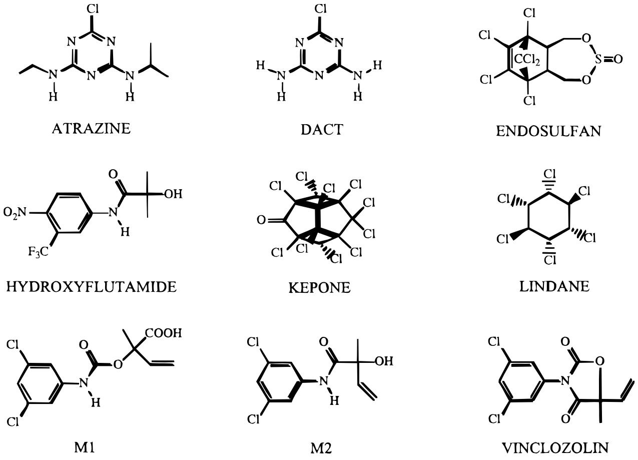

Figure 6. Structures of compounds in pesticide subset.

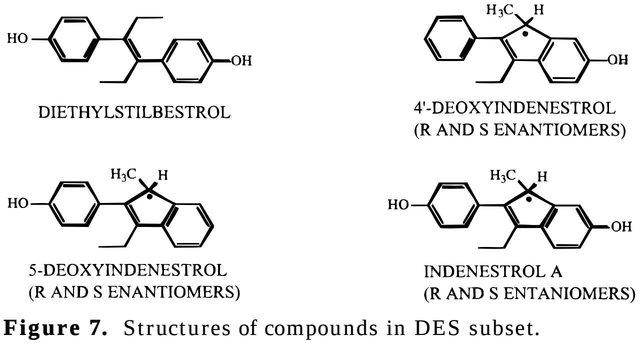
DIETHYLSTILBESTROL
4'-DEOXYINDENESTROL (R AND S ENANTIOMERS)
5-DEOXYINDENESTROL (R AND S ENANTIOMERS)
INDENESTROL A (R AND S ENTANIOMERS)

Figure 7. Structures of compounds in DES subset.

Methods

Biological Activity (Dependent Variable) Measurements. Competitive estrogen receptor binding assays were performed using mouse uterine cytosol preparation as described previously (*18−20*). In short, the binding affinities are determined as the equlibrium dissociation constant, K_i, defined as [R][I]/[RI], where [R], [I], and [RI] represent concentrations of receptor, inhibitor, and receptor−inhibitor complex. Under the specific conditions of the experiments, the K_i is equal to the concentration of the inhibitor which doubles the slope of the

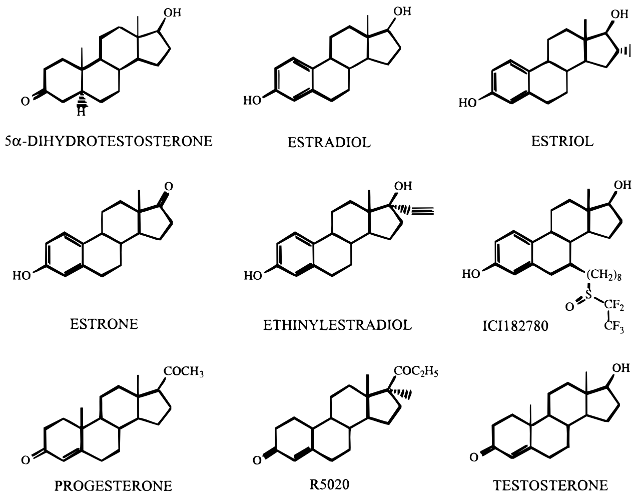

Figure 8. Structures of compounds in steroid subset.

double-reciprocal plot. The K_i values are utilized in the CoMFA/QSAR models as the negative log of K_i in μmol (pK_i).

Molecular Modeling. All molecules, with the exceptions of estradiol and diethylstilbestrol (DES), were constructed using the building tools of the SYBYL molecular modeling software package (*21*). The starting coordinates of estradiol and DES were retrieved from the Cambridge Structural Database. All molecules were geometry optimized using the standard Tripos force field(*22*) with the conjugate-gradient minimizer to an energy change convergence criterion of 0.001 kcal/mol/step. Flexible molecules were subjected to systematic conformational search routines using the SYBYL software. All rotatable bonds were scanned at 5° increments. The lowest energy conformer for a given molecule was then reminimized according to the criteria described above. The SYBYL minimized structures were then fully geometry optimized using the AM1 (*23*) model Hamiltonian in MOPAC (*24*). The semiempirically derived low energy structures with the corresponding atomic charges were used in all CoMFA/3D-QSAR studies presented in this paper.

Alignment Hypothesis. The rigid body alignment program SEAL (*16*) was used to align all low energy molecules to the template molecule estradiol. A series of preliminary studies suggested the following SEAL parameters. An attenuation factor of 0.25 was employed in the alignment process to provide for a balance of atom by atom pairings with field overlap. Steric, electrostatic, and hydrophobic field weighting factors of 1, 2, and 2, respectively, were utilized. These weightings, in combination with the attenuation factor, allowed the core ring structures of all steroids to be superimposed with estradiol while providing the less structurally similar molecules to be aligned predominantly according to their field similarities with the estradiol. Two hundred orientations were generated for each molecule with only the best retained for inclusion in the CoMFA/QSAR analyses.

CoMFA Interaction Energy Calculations. Steric (van der Waals) and electrostatic (Coulombic) interaction energies were computed between each molecule aligned relative to all others according to the techniques described above and a "probe atom" placed at regularly spaced (2.0 Å) intersections on a common grid. The probe atom was, by default, an sp^3 hybridized carbon with an effective radius of 1.53 Å and a charge of +1.0. The grid was constructed to extend at least 4.0 Å in all directions beyond the dimensions of the common volume of all superimposed molecules. All interaction energies (steric and electrostatic independently) measured for a given molecule were recorded in a QSAR table, with individual columns representing the interactions measured at a specific grid intersection. The interaction energies measured for subsequent molecules were placed in successive rows of the QSAR molecular spreadsheet (MSS).

The steric (van der Waals) interaction energies were computed using the 6−12 Lennard-Jones potential with truncation values of +30.0 and −30.0 kcal/mol. The electrostatic (Coulombic) interaction energies were computed as charge−charge interactions also with truncation values of +30.0 and −30.0 kcal/mol. Electrostatic interactions at "sterically forbidden" grid points (those with interaction energy values greater than 30.0 kcal/mol) were effectively ignored by setting their value to the mean value for that column. A distance dependent dielectric (1/*r*) was utilized.

Hydropathic Interaction (HINT (*25*)) Energy Calculations. Hydropathic interaction energies were computed in a similar manner as the steric and electrostatic field energies using the following empirically-derived potential: $A_t = \sum a_i s_i R_{it}$. The terms a_i and s_i represent the hydrophobic atom constant and the solvent-accessible surface area for atom *i*. R_{it} is a function of the distance between each atom *i* in the molecular system and the "probe atom" *t* and is computed as e^{-r}. As above, the interaction measured at each individual grid intersection is recorded in a separate (electrostatic-type only) column of the QSAR MSS. The grid definition used for the steric and electrostatic interaction energy calculations was implemented for the hydropathic calculations. The hydrophobic only HINT column type was utilized with no cutoff values imposed.

Statistical Analyses. All statistical analyses were performed using the partial least squares (PLS (*26*)) methodology as employed in the QSAR module of SYBYL 6.2 running on Silicon Graphics Onyx and IndigoII workstations. Initial analyses were performed using the leave-one-out (LOO) cross-validation (*27*) technique and ten principal components (PCs). The crossvalidated r^2 (termed q^2 herein) was computed as follows:

$$q^2 = (\mathrm{SD} - \text{“PRESS”})/\mathrm{SD} \qquad (1)$$

where SD is the sum of the squared deviations between the measured and mean binding affinities of the training set molecules, and "PRESS" is the sum of the squared deviations between the predicted and measured binding affinities for every molecule.

In an effort to increase the signal-to-noise ratio in the CoMFA analyses, individual steric and electrostatic columns corresponding to particular grid intersections exhibiting a standard deviation of less than 2.0 kcal/mol (termed minimum sigma) were not included in the crossvalidated analysis. A minimum sigma value of 0.5 kcal/mol was implemented in the analyses including the HINT columns. Therefore, only those columns displaying significant variability in the interaction values measured for different molecules were considered. This "column filtering" technique typically reduces the number of columns in the QSAR table by a factor of 10. The optimal number of components to be used in the subsequent non-crossvalidated analysis was determined as that which yielded the highest q^2 value and the lowest standard error of crossvalidated (LOO) predictions (SEP). For the non-crossvalidated analyses, the minimum sigma was reduced to 0.0 kcal/mol (all columns were included in the analyses).

Predictive Power of the CoMFA/QSAR Models. In order to more fully assess the general utility of the model based on the combination of CoMFA and HINT fields, the training set was divided into eight subclasses of structurally related molecules: phenols, phthalates, phytoestrogens, DDTs, PCBs, pesticides, DESs, and steroids. Eight CoMFA submodels were generated by *excluding all of the members of each class from the training set*. The binding affinity of each member of a particular class was then predicted using the model from which the entire class was excluded. In this manner, a more realistic (termed "unbiased" herein) estimate of the external predictive ability of the CoMFA/QSAR models is attained.

Results

SEAL Alignment. In Figure 9, representative molecules from the various subsets are displayed in their optimal orientation with the template molecule estradiol.

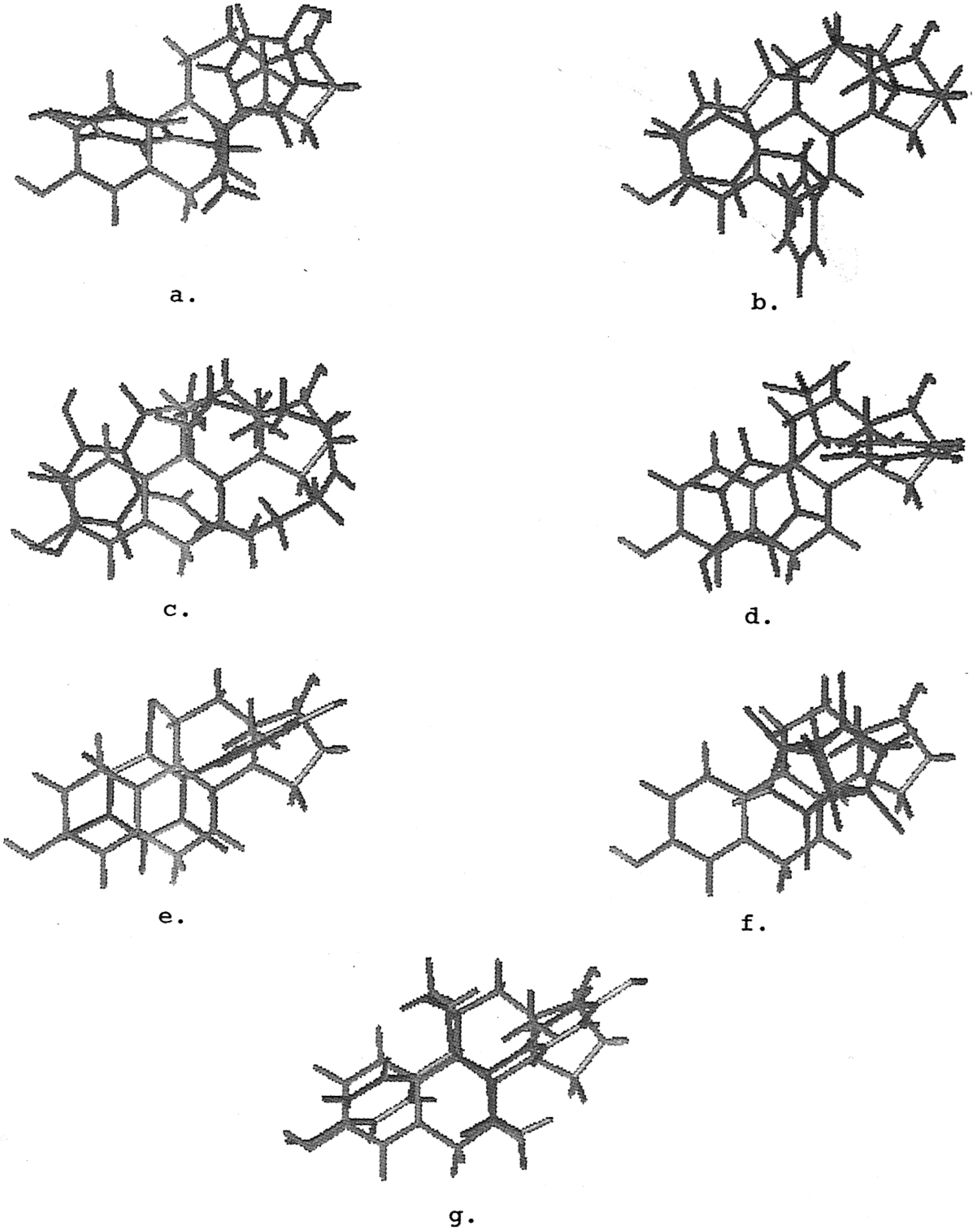

Figure 9. Alignment of selected chemicals with estradiol: (a) bisphenol A, (b) butylbenzylphthalate, (c) zearalenone, (d) HPTE, (e) 2,4,6-trichloro-4′-biphenylol, (f) kepone, and (g) DES.

The compounds bisphenol A (panel a), zearalenone (panel c), 2,2-bis(*p*-hydroxyphenyl)-1,1,1-trichloroethane, HPTE (panel d), and DES (panel g) all possess a phenolic ring moiety. In general, this ring was superimposed with the phenolic A-ring of estradiol. However, the phenolic ring of 2,4,6-trichloro-4′-biphenylol (panel e) was oriented over the estradiol D-ring with the 4′-hydroxyl group of the biphenyl in close proximity to the 17-hydroxyl of estradiol. Butyl benzyl phthalate (panel b) does not possess a phenolic ring as part of the structure. In this case, the optimal orientation found superimposed the aromatic ring of the phthalate moiety with the aromatic A-ring of estradiol. The optimal alignment for the highly chlorinated compound kepone (panel f) was found to be one in which the bulk of the structure is positioned over the B-, C-, and D-rings of estradiol with the carbonyl functionality of the kepone molecule pointing toward the 3-hydroxyl of the template molecule.

Training Set Models. The statistical results of the CoMFA/QSAR models using the entire training set of 58 structurally diverse molecules are summarized in Table 1. The analysis based on standard CoMFA steric and electrostatic potential energy fields yielded a leave-one-out crossvalidated correlation coefficient (q^2_{LOO}) of 0.467 using two principal components (PCs) with a standard error of crossvalidated predictions (SEP) of 1.456. The

Table 1. Statistical Results of CoMFA/QSAR Models

regression metric	CoMFA fields model	HINT field model	CoMFA with HINT field model
q^2_{LOO}	0.467 (2)[a]	0.481 (3)	0.590 (3)
SEP	1.456	1.450	1.289
r^2	0.808	0.691	0.881
s	0.873	1.119	0.694
F-test	116.019	40.194	133.522
p value	<0.0005	<0.0005	<0.0005
steric field contribution	38	N/A	27
electrostatic field contribution	62	N/A	45
hydrophobic field contribution	N/A	100	28

[a] The numbers in parentheses indicate the number of principle components (PC) used to construct the model.

Table 2. Actual versus Predicted Binding Affinities: Phenols

		predictions		
compd	actual pK_i	unbiased	cross-validated	fitted
2-*tert*-butylphenol	−2.367	−1.311 (0.2)[a]	−1.995	−1.985
3-*tert*-butylphenol	−2.597	−1.164 (0.6)	−1.806	−1.967
4-*tert*-butylphenol	−2.207	−1.476 (0.2)	−2.192	−2.176
4-*tert*-octylphenol	−0.121	0.335 (0.2)	0.940	0.363
bisphenol A	−0.164	−1.376 (0.3)	−2.012	−1.144
nonylphenol	0.080	−0.898 (2.6)	−0.982	−0.491
av absolute errors:		0.978	0.858	0.513

[a] Numbers in parentheses represent extrapolation (as %) required to make predictions.

conventional correlation coefficient (r^2) for this model was 0.808. The electrostatic field dominated this model, contributing approximately 62%.

The analysis based solely on HINT-derived hydrophobic fields expressed a slightly greater q^2_{LOO} of 0.481 using 3 PCs with a correspondingly lower SEP of 1.450. This model, however, was not as statistically robust, expressing an r^2 of 0.691.

The analysis based on the combination of CoMFA and HINT fields resulted in the most internally predictive model as indicated by the q^2_{LOO} value of 0.590 using 3 PCs. This model was the most statistically robust as well, expressing an r^2 of 0.881. The electrostatic field also made the dominant contribution (45%) in this model as well, with the hydrophobic and steric fields contributing approximately equally (28% and 27%, respectively).

Predictive Power of the CoMFA/QSAR Models. The training set data span approximately 8 log units of binding affinity. For a data set of this size expressing this degree of structural diversity, an entirely arbitrary acceptance factor of ±1.0 log unit has been adopted. The external unbiased, crossvalidated, and fitted predictions of binding affinity from the combined CoMFA/HINT field model are listed in Tables 2−9with corresponding average absolute errors of predictions listed for each compound subclass.

The unbiased average error of predictions for all compound subclasses was typically on the order of 1.0−2.0 log units. The binding affinities for the molecules of the PCB subclass were most closely predicted with an error of only 0.685 log unit, with the pesticide subclass being most poorly predicted with an error of 2.112 log units. The compounds comprising the phenol and PCB subclasses were predicted within an average absolute error of less than 1.0 log unit.

Table 3. Actual versus Predicted Binding Affinities: Phthalates

		predictions		
compd	actual pK_i	unbiased	cross-validated	fitted
butyl benzyl phthalate	−1.883	0.299 (10.2)[a]	0.599	−1.348
di-*n*-butyl phthalate	−2.002	−1.489 (6.7)	−1.359	−2.193
av absolute errors:		1.348	1.563	0.363

[a] Numbers in parentheses represent extrapolation (as %) required to make predictions.

Table 4. Actual versus Predicted Binding Affinities: Phytoestrogens

		predictions		
compd	actual pK_i	unbiased	cross-validated	fitted
coumestrol	1.032	−0.525 (8.8)[a]	−0.803	0.761
genistein	0.409	−0.035 (2.8)	0.278	1.016
zearalenone	2.222	−1.161 (6.2)	−0.417	1.567
av absolute errors:		1.795	1.535	0.511

[a] Numbers in parentheses represent extrapolation (as %) required to make predictions.

Table 5. Actual versus Predicted Binding Affinities: DDTs

		predictions		
compd	actual pK_i	unbiased	cross-validated	fitted
HPTE	1.301	−0.245 (1.7)[a]	−1.961	−0.624
methoxychlor	−1.839	0.064 (8.3)	0.105	−1.134
o,p′-DDT	−0.462	0.302 (2.8)	−0.862	−0.522
p,p′-DDD	−3.000	−0.033 (3.2)	−1.593	−2.230
p,p′-DDE	−3.000	−1.108 (2.4)	−3.374	−2.977
p,p′-DDT	−3.000	−0.305 (3.2)	−1.665	−2.198
av absolute errors:		1.961	1.454	0.714

[a] Numbers in parentheses represent extrapolation (as %) required to make predictions.

The LOO, or crossvalidated, predictions for most compound subclasses were marginally better in that the average absolute errors were reduced on the order of 0.1−0.5 log unit. The error for the phthalate subclass surprisingly was increased by approximately 0.1 log unit. Once again the chemicals of the PCB subclass were most closely predicted with an average absolute error of only 0.640 log unit. In addition to the phenol and DDT subclasses, the molecules comprising the steroid subclass were predicted within an average absolute error of 1.0 log unit.

The fitted, or non-crossvalidated, average errors of predictions for all chemical subclasses were reduced on the order of 0.3−1.0 log unit. The average absolute error for all subclasses was 1.0 log unit or less in this case. The binding affinities for the compounds comprising the phthalate and PCB subclasses were predicted within an average of 0.5 log unit or less.

CoMFA Contour Plots. In Figures 10−12, contour plots as the product of the standard deviation and the regression coefficient (STDDEV*COEFF) from the CoMFA/HINT QSAR model are presented. In Figure 10, green polyhedra are used to denote areas in space around the estradiol molecule in which increases in steric bulk brought about by larger training set molecules were demonstrated to enhance the estrogen receptor-binding ability. Yellow polyhedra are used to denote the converse situation. Steric bulk tolerance is noted in the vicinity of the steroid A-ring near the 2 and 3 positions and the

Table 6. Actual versus Predicted Binding Affinities: PCBs

compd	actual pK_i	predictions		
		unbiased	crossvalidated	fitted
2,4,6-trichloro-4′-biphenylol	1.316	1.329 (0.7)[a]	0.716	1.032
2,3,4,5-tetrachloro-4′-biphenylol	1.345	0.056 (1.8)	0.009	0.901
2-chloro-4,4′-biphenyldiol	0.939	0.889 (0.2)	0.353	0.920
2,6-dichloro-4′-biphenylol	0.519	1.376 (0.3)	0.615	0.815
2,5-dichloro-4′-biphenylol	0.446	0.836 (0.2)	0.933	0.981
3,3′,5,5′-tetrachloro-4,4′-biphenyldiol	−0.290	−1.064 (1.9)	0.372	−0.154
2-chloro-4-biphenylol	−0.509	−0.035 (0.2)	−0.666	−0.415
4-chloro-4′-biphenylol	−0.746	−1.238 (0.6)	−1.387	−1.051
2,3,5,6-tetrachloro-4,4′-biphenyldiol	−0.380	0.975 (1.4)	0.712	0.547
2,2′,3,3′,6,6′-hexachloro-4-biphenylol	−0.847	−0.330 (3.2)	−1.337	−1.179
2,2′3,4′,6,6′-hexachloro-4-biphenylol	−0.732	0.305 (3.6)	−0.296	−0.507
2,2′,3,6,6′-pentachloro-4′-biphenylol	−0.203	1.098 (3.6)	−0.853	−0.199
2,2′,5,5′-tetrachlorobiphenyl	−0.792	−0.993 (1.9)	−1.074	−0.922
2,2′,4,4′,5,5′-hexachlorobiphenyl	−0.934	−1.652 (4.7)	−1.594	−1.461
2,2′,4,4′,6,6′-hexachlorobiphenyl	−0.117	0.017 (1.3)	0.040	−0.043
2,2′,3,3′,5,5′-hexachloro-6-biphenylol	−0.812	−2.169 (4.3)	−2.317	−1.887
av absolute errors:		0.685	0.640	0.338

[a] Numbers in parentheses represent extrapolation (as %) required to make predictions.

Table 7. Actual versus Predicted Binding Affinities: Pesticides

compd	actual pK_i	predictions		
		unbiased	cross-validated	fitted
atrazine	-3.000	−0.286 (6.5)[a]	−1.531	−2.236
DACT	−3.000	−1.767 (33.3)	−2.719	−2.792
endosulfan	−2.778	0.672 (9.7)	0.809	−3.191
hydroxyflutamide	−3.000	−0.974 (19.2)	−1.622	−3.438
kepone	−0.146	−0.759 (2.1)	−2.436	−1.654
lindane	−3.000	−0.698 (7.6)	−0.533	−1.954
M1	−3.000	−1.227 (3.6)	−1.893	−2.935
M2	−3.000	−0.535 (6.7)	−2.363	−2.823
vinclozolin	-3.000	−0.570 (5.3)	−2.101	−2.833
av absolute errors:		2.112	1.568	0.532

[a] Numbers in parentheses represent extrapolation (as %) required to make predictions.

Table 8. Actual versus Predicted Binding Affinities: DESs

compd	actual pK_i	predictions		
		unbiased	cross-validated	fitted
diethylstilbestrol (DES)	3.155	0.680 (0.7)[a]	1.258	2.685
4′-deoxyindenestrol (*R*)	0.301	0.042 (0.5)	1.732	1.443
4′-deoxyindenestrol (*S*)	1.256	0.969 (0.2)	2.403	2.423
5-deoxyindenestrol (*R*)	0.955	−0.102 (1.0)	1.463	1.456
4′-deoxyindenestrol (*S*)	1.750	1.647 (0.2)	2.521	2.983
indenestrol A (*R*)	2.363	−0.400 (1.1)	0.084	1.212
indenestrol A (*S*)	3.456	1.550 (0.2)	1.827	3.060
av absolute errors:		1.264	1.380	1.001

[a] Numbers in parentheses represent extrapolation (as %) required to make predictions.

Table 9. Actual versus Predicted Binding Affinities: Steroids

compd	actual pK_i	predictions		
		unbiased	cross-validated	fitted
dihydrotestosterone	−1.000	0.402 (3.8)[a]	−1.761	−1.366
estradiol	2.585	0.989 (0.2)	2.180	2.096
estriol	1.854	1.380 (1.9)	2.999	2.550
estrone	2.357	0.560 (1.8)	1.796	1.869
ethinylestradiol	3.523	1.176 (3.4)	2.454	2.741
ICI182780	3.222	1.363 (30.6)	2.534	2.638
progesterone	−3.000	0.455 (8.6)	−1.315	−2.286
R5020	−0.072	0.839 (9.8)	−1.402	−0.926
testosterone	−1.462	0.559 (2.4)	−1.377	−1.290
av absolute errors:		1.762	0.859	0.572

[a] Numbers in parentheses represent extrapolation (as %) required to make predictions.

steroid D-ring near the 17 substituent. Steric bulk is not accommodated out of the plane of the A-ring of the steroid or near the 15 and 16 positions of the D-ring.

The corresponding electrostatic contour plots are presented in Figure 11. In this case, blue polyhedra denote areas in which positive potential is desired. Red polyhedra indicate that negative potential is preferred. The large blue polyhedron near the 3 position of the steroid A-ring probably resulted due to the common orientation of the hydrogen of the hydroxyl group of many of the training set molecules. The blue polyhedron near the 17 position of the D-ring indicates a similar situation in that hydroxyl appears the be the preferred substituent. Most descriptive are the areas of preference for negative potential which exist near the 3 and 17 positions of the steroid nucleus.

The hydrophobic contours shown in Figure 12 denote those areas which are most and least tolerant to hydrophobic moieties displayed as blue and red polyhedra, respectively. Hydrophobic bulk appears to be most desired in the vicinity of the steroid A-ring. Closer to the surface of the estradiol template are areas which do not accommodate hydrophobic bulk as well. These polyhedra are noted in the vicinity of the 1, 4, and 16 positions of the steroid core.

Discussion

SEAL Alignment. In our original publications in this series (*5, 28*), we attempted to utilize a consistent, repeatable, and mechanistically-relevant alignment rule which was designed to optimize the degree of steric and electrostatic potential energy field similarity between structurally-diverse ligands of the estrogen receptor and the natural ligand, estradiol. The high degree of subjectivity required to produce this alignment has been noted as a potential deficiency of the earlier models. The pharmacophore as originally proposed consisted of a six-membered ring with an electronegative substituent (usually hydroxyl) with another electronegative group held in a position more distal from the A-ring structure by a lipophilic molecular framework. With fewer biases (i.e., operator input), the SEAL algorithm essentially repli-

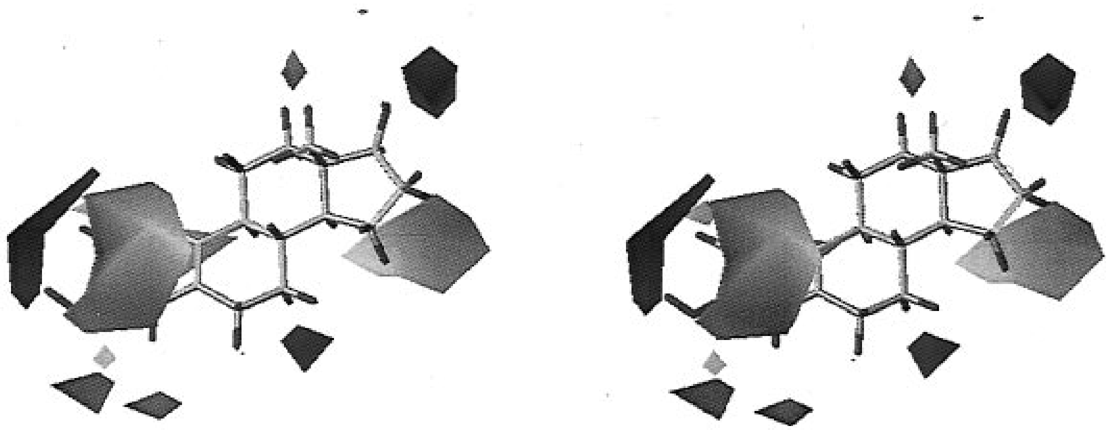

Figure 10. Crossed stereoview of steric field contribution contour plots. Molecule is estradiol. Green contours designate areas of steric bulk tolerance. Yellow contours designate areas of steric bulk intolerance.

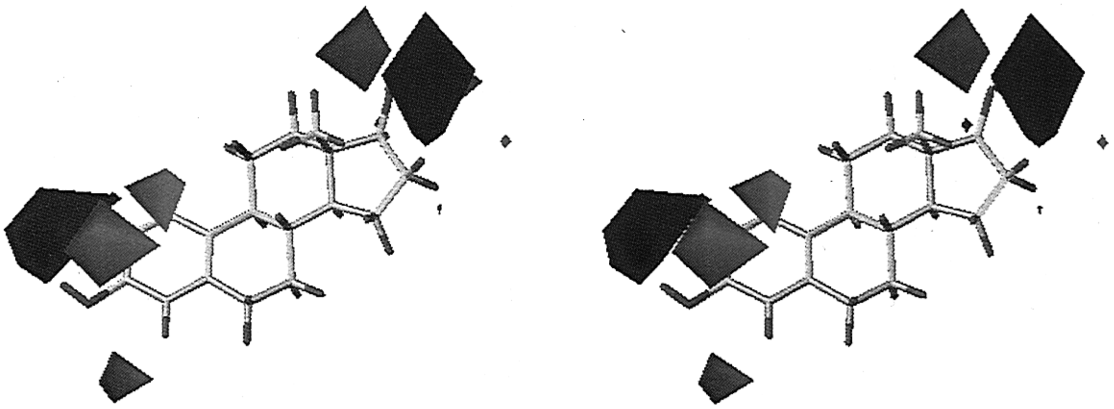

Figure 11. Crossed stereoview of electrostatic field contribution contour plots. Molecule is estradiol. Blue contours designate areas of positive potential tolerance. Red contours designate areas of negative potential tolerance.

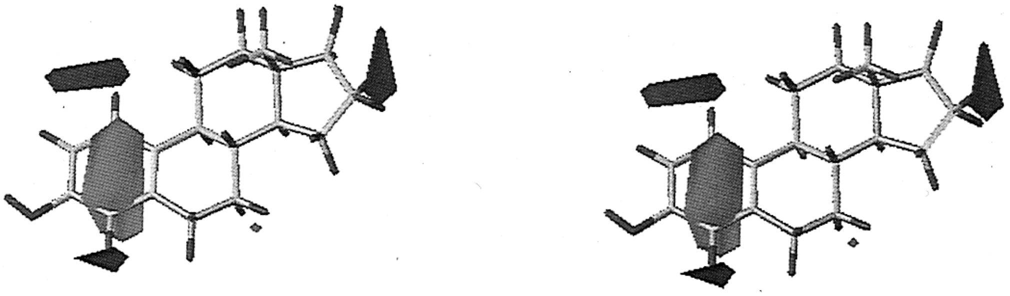

Figure 12. Crossed stereoview of hydrophobic field contribution contour plots. Molecule is estradiol. Orange contours designate areas of hydrophobic tolerance. Purple contours designate areas of hydrophobic intolerance.

cated this alignment in most cases, thus validating our original pharmacophore hypothesis. This model, while more refined than the original (*5*), can be viewed as being more general in that it comprises a more structurally diverse training data set. It can be argued that, in the case of certain hydroxylated PCBs and pesticides, the presumed to be critical hydrogen bonding functionality attached to the A-ring was moved in such a manner as to no longer allow a hydrogen bond to be formed with residues of the binding domain of the receptor. For certain members of the PCB subset of molecules, the phenolic ring was moved to overlap with the D-ring hydroxyl substituent while maximizing the degree of hydrophobic overlap in the central B- and C-ring regions (see Figure 9e). Common in all of the alignments was the overlap of aromatic A-rings when possible, thus suggesting the presence of a stacking interaction with the receptor. For pesticides such as kepone, the hydrogen bonding functionality was moved to such a degree that the carbonyl was in the vicinity the B-ring of the steroid nucleus while once again maximizing the degree of hydrophobic overlap in the B-, C-, and D-ring regions.

Training Set Models. The CoMFA/QSAR training set encompassed at least seven classes of chemical compounds of natural and synthetic origins. The standard CoMFA steric and electrostatic potential energy field model proved to be quite internally predictive and statistically robust, with the major contribution coming from the electrostatic field. The model based solely on the hydrophobic potential fields of the molecules was slightly, albeit statistically insignificantly, more internally predictive than the standard CoMFA fields model. The statistical robustness of the hydrophobic field model was somewhat decreased relative to the standard CoMFA fields model. This is indicative of the less descriptive nature of the singular field. The combination of CoMFA steric and electrostatic with hydrophobic fields resulted in a model with increased internal predictive ability and statistical robustness. The increase in the crossvalidated correlation coefficient (q^2) value possibly results from a synergistic interaction between the three fields. The increased standard correlation coefficient (r^2) is a result of the increase in the number of descriptors, as principal components, relative to the standard CoMFA fields model. The decrease in the contributions of the steric and electrostatic fields in this three field model relative to the two field CoMFA model is suggestive of a greater significance of the hydrophobic field to the model, at least with respect to the steric field.

Predictive Power of the CoMFA/QSAR Models. A novel technique of model validation was employed in this study. By separating the training set into distinct structural classes and using each as *external* test sets, an *unbiased* estimation of the true predictive ability for compounds outside the design space of the model is achieved. This predictive ability of the model for these compounds is ascertained by examining the percentage of points in the region which must be extrapolated in order to make the prediction. With the exception of nonylphenol, the predictions for the compounds of the phenol subset required extrapolation of less than 1%, which explains the small average absolute error of predictions for all phenols of less than 1.0 log unit. The greater degree of extrapolation required to make the predictions for the phthalate, phytoestrogens, DDT, DES, and steroid subsets of molecules indicates that these molecules possess unique structural information, relative to the training set molecules, and results in a higher average absolute error of predictions for these molecules. More precisely, the average absolute error for the unbiased predictions for the PCB subset was on the order of 0.7 log unit. The structural features of these molecules were well described by the training set molecules as indicated by the low degree of extrapolation (less than 5% in all cases) required to make these predictions. The unbiased predictions for the pesticide subset were much less precise than other sets with an average absolute error of predictions of greater than 2 log units. An examination of the degree of extrapolation necessary to make these predictions reveals a range between 2% and 33%, on average much greater than that required for the other compound subsets.

The leave-one-out crossvalidated predictions for all molecules were not significantly better than the unbiased predictions. This is somewhat surprising in that the inclusion of other members for structurally distinct chemical classes should add pertinent descriptive information to the training set and thus improve the prediction. The insignificant change is most probably indicative that the molecules in the training sets from the unbiased models already encompass the structural diversity of the leave-one-out training set models. The fitted predictions were much better in all cases, typically on the order of 1.0 log unit. This was expected since the final model should contain structural information extracted from all molecules.

CoMFA Contour Plots. In general, CoMFA contour plots provide the user with graphical evidence that the model (i.e., PLS regression) extracted the relevant information from the physicochemical data matrix. In certain instances, these contours can be used as qualitative biological activity prediction aids; this is probably of most utility in the computer-aided molecular design arena. On a larger scale, contour plots provide what may be viewed as negative images of the ligand binding domain of the receptor. Viewing CoMFA contours in this manner must be performed with caution due to the limited amount of information contained in the training set of ligands and the hypothetical nature of the alignment rule. With those caveats noted, an examination of the contours is warranted. Our original pharmacophore hypothesis suggested that the ligand binding domain of the estrogen receptor preferred compounds with two electronegative centers separated by a rigid hydrophobic framework. Figures 10−12 support this hypothesis. One could argue that the resulting contours represent the results of a self-fulfilling prophecy in that they are totally dependent on (1) the alignment rule and (2) the composition of the training set—two variables prone to subjective manipulation. This claim is somewhat invalidated in the present study by the less subjective molecular alignment technique (i.e., SEAL) employed and by the diverse nature of the training set molecules.

Conclusion

A QSAR model based on the three-dimensional structural characteristics of a diverse collection of estrogen receptor ligands has been presented. The predictive ability of the model was demonstrated using external test sets of structurally diverse molecules. The molecules contained in the test sets were sequentially incorporated into the training set to yield a CoMFA/QSAR model for ligands of the estrogen receptor, which effectively illustrates that potential ligands for the estrogen receptor can be identified solely on the basis of their physicochemical characteristics. The three-dimensional nature of the techniques utilized (i.e., SEAL, CoMFA, HINT) provided a means by which the validity of our original pharmacophore hypothesis could be tested. The results of the model in terms of predictive ability, statistical robustness, and graphical display of relevant physicochemical properties in the form of contour plots all suggest that the pharmacophore contains salient structural information.

Preliminary results from three-dimensional data base searches suggest that our model of the estrogen receptor pharmacophore is of sufficient detail as to provide for ligand-based identification of environmental estrogens. To date, this strategy has not been fully implemented; however, steps are being taken to integrate these emerging computational technologies with more traditional toxicological procedures, which would result in a more expedient hazard identification process. It is believed that three-dimensional data base searching techniques could be used to rapidly identify potential endocrine disruptors from data bases of thousands, perhaps millions, of compounds. QSAR models developed for the various endocrine systems (i.e., estrogen (*5*), androgen (*28*), etc.) could then be used to provide quick estimates of the biological potencies of these compounds, thus providing a scheme for prioritization for detailed *in vitro* and *in vivo* toxicological investigations.

Acknowledgment. The authors wish to acknowledge the analytical chemistry expertise of Ms. Nancy J. Andersen and Mr. John J. Falchi (PKB/ETD/NHEERL/USEPA). Drs. James D. McKinney and Linda S. Birnbaum (ETD/NHEERL/USEPA) are acknowledged for kindly reviewing the manuscript. Dr. Mary Bradley (Wyeth-Ayerst Research) is gratefully acknowledged for assistance with SEAL. One author (T.E.W.) wishes to acknowledge funding by USEPA training agreement T901915. Part of this work was supported by the U.S. Department of Energy through a Los Alamos National Laboratory LDRD-CD grant (T.I.O.). Disclaimer: This manuscript has been reviewed in accordance with the policies of the National Health and Environmental Effects Research Laboratory of the United States Environmental Protection Agency. Mention of trade names does not constitute endorsement or recommendation for use.

References

(1) Agarwal, M. K. (1994) Analysis of Steroid Receptor Domains with the Aid of Antihormones. *Int. J. Biochem.* **26**, 341−350.

(2) Parker, M. G., Arbuckle, N., Dauvois, S., Danielian, P., and White, R. (1993) Structure and Function of the Estrogen Receptor. *Ann. N.Y. Acad. Sci.* **June 11**, 119−125.

(3) Wong, C., Kelce, W. R., Sar, M., and Wilson, E. M. (1995) Androgen Receptor Antagonist versus Agonist Activities of the Fungicide Vinclozolin Relative to Hydroxyflutamide. *J. Biol. Chem.* **270**, 19998−20003.

(4) Hyder, S. M., Shipley, G. L., and Stancel, G. M. (1995) Estrogen Action in Target Cells: Selective Requirements for Activation of Different Hormone Response Elements. *Mol. Cell. Endocrinol.* **112**, 35−43.

(5) Waller, C. L., Minor, D. L., and McKinney, J. D. (1995) Examination of the Estrogen-Receptor Binding Affinities of Polychlorinated Hydroxybiphenyls Using Three-Dimensional Quantitative Structure−Activity Relationships. *Environ. Health Perspect.* **103**, 702−707.

(6) Loughney, D. A., and Schwender, C. F. (1992) A comparison of progestin and androgen receptor binding using the CoMFA technique. *J. Comput.-Aided Mol. Des.* **6**, 569−581.

(7) Cramer, R. D., Patterson, D. E., and Bunce, J. D. (1988) Comparative Molecular Field Analysis (CoMFA). 1. Effect of shape on Binding of Steroids to Carrier Proteins. *J. Am. Chem. Soc.* **110**, 5959−5967.

(8) Oprea, T. I., and Garcia, A. E. (1996) Three-Dimensional Quantitative Structure−Activity Relationship of Steroid Aromatase Inhibitors. *J. Comput.-Aided Mol. Des.* (in press).

(9) Thomas, B. F., Compton, D. R., Martin, B. R., and Semus, S. F. (1991) Modeling the Cannabinoid Receptor: A Three-Dimensional Quantitative Structure-Activity Analysis. *Mol. Pharmacol.* **40**, 656−665.

(10) Agarwal, A., and Taylor, E. W. (1993) 3-D QSAR for Intrinsic Activity of 5-HT$_{1A}$ Receptor Ligands by the Method of Comparative Molecular Field Analysis. *J. Comput. Chem.* **14**, 237−245.

(11) Waller, C. L., Oprea, T. I., Giolitti, A., and Marshall, G. R. (1994) 3D-QSAR of Human Immunodeficiency Virus (I) Protease Inhibitors. I. A CoMFA Study Employing Experimentally-Determined Alignment Rules. *J. Med. Chem.* **36**, 4152−4160.

(12) DePriest, S. A., Mayer, D., Naylor, C. B., and Marshall, G. R. (1993) 3D-QSAR of Angiotensin-Converting Enzyme and Thermolysin Inhibitors: A Comparison of CoMFA Models Based on Deduced and Experimentally Determined Active Site Geometries. *J. Am. Chem. Soc.* **115**, 5372−5384.

(13) Waller, C. L., and McKinney, J. D. (1992) Comparative Molecular Field Analysis of Polyhalogenated Dibenzo-*p*-dioxins, Dibenzofurans, and Biphenyls. *J. Med. Chem.* **35**, 3660−3666.

(14) Waller, C. L., and McKinney, J. D. (1995) Three-Dimensional Quantitative Structure−Activity Relationships of Dioxins and Dioxin-like Compounds: Model Validation and Ah Receptor Characterization. *Chem. Res. Toxicol.* **8**, 847−858.

(15) Quantum Chemistry Program Exchange #634.

(16) Kearsley, S. K., and Smith, G. M. (1990) An Alternative Method for the Alignment of Molecular Structures: Maximizing Electrostatic and Steric Overlap. *Tetrahedron Comput. Methodol.* **3**, 615−633.

(17) Martin, Y. C., Bures, M. G., and Willett, P. (1990) Searching Databases of Three-Dimensional Structures. In *Rev. in Computational Chemistry* (Lipkowitz, K., and Boyd, D., Eds.) pp 213−263, VCH Publishers, New York.

(18) Korach, K. S., Sarver, P., Chae, K., McLachlan, J. A., and McKinney, J. D. (1988) Estrogen Receptor-Binding Activity of Polychlorinated Hydroxybiphenyls: Conformationally Restricted Structural Probes. *Mol. Pharmacol.* **33**, 120−126.

(19) Chae, K., Gibson, M. K., and Korach, K. S. (1991) Estrogen Receptor Stereochemistry: Ligand Binding Orientation and Influence on Biological Activity. *Mol. Pharmacol.* **40**, 806−811.

(20) Laws, S. C., Carey, S. A., Kelce, W. R., Cooper, R. L., and Gray, L. E. (1996) An In Vitro Comparison of the Affinity of Hormones, Drugs, and Environmental Chemicals for Androgen, Estrogen, and Progesterone Receptors. *Toxicol. Appl. Pharmacol.* Submitted.

(21) SYBYL, Version 6.2, Tripos, Inc., St. Louis, MO.

(22) Clark, M., D., C. R., and Van Opdenbosch, N. (1989) Validation of the General Purpose Tripos 5.2 Force Field. *J. Comput. Chem.* **10**, 982−1012.

(23) Dewar, M. J. S., Zoebisch, E. G., Healy, E. F., and Stewart, J. J. P. (1985) AM1: A New General Purpose Quantum Mechanical Molecular Model. *J. Am. Chem. Soc.* **107**, 3902−3909.

(24) Quantum Chemistry Program Exchange #455.

(25) Kellogg, G. E., Semus, S. F., and Abraham, D. J. (1991) HINT: A new method of empirical hydrophobic field calculation for CoMFA. *J. Comput.-Aided Mol. Des.* **5**, 545−552.

(26) Wold, S., Ruhe, A., Wold, H., and Dunn, W. J. (1984) The Covariance Problem in Linear Regression. The Partial Least Squares (PLS) Approach to Generalized Inverses. *SIAM J. Sci. Stat. Comp.* **5**, 735−743.

(27) Cramer, R. D., Bunce, J. D., Patterson, D. E., and Frank, I. E. (1988) Crossvalidation, Bootstrapping, and Partial Least Squares Compared with Multiple Regression in Conventional QSAR Studies. *Quant. Struct.−Act. Relat.* **7**, 18−25.

(28) Waller, C. L., Juma, B. W., Gray, L. E., and Kelce, W. R. (1996) Three-Dimensional Quantitative Structure−Activity Relationships for Androgen Receptor Ligands. *Toxicol. Appl. Pharmacol.* **137,** 219−227.

TX960054F