2000

A Computationally Based Identification Algorithm for Estrogen Receptor Ligands: Part 2. Evaluation of a hERα Binding Affinity Model

Steven P. Bradbury
O. G. Mekenyan
V. Kamenska
P. K. Schmieder
G. T. Ankley

TOXICOLOGICAL SCIENCES **58,** 270–281 (2000)

A Computationally Based Identification Algorithm for Estrogen Receptor Ligands: Part 2. Evaluation of a hERα Binding Affinity Model

O. G. Mekenyan,* V. Kamenska,* P. K. Schmieder,† G. T. Ankley,† and S. P. Bradbury†,[1]

*Bourgas University "Prof. As. Zlatarov," Laboratory of Mathematical Chemistry, Department of Physical Chemistry, 118010 Bourgas, Bulgaria; and †U.S. Environmental Protection Agency, National Health and Environmental Effects Research Laboratory, Mid-Continent Ecology Division, 6201 Congdon Boulevard, Duluth, Minnesota 55804

Received April 17, 2000; accepted July 31, 2000

The objective of this study was to evaluate the capability of an expert system described in the previous paper (S. Bradbury *et al.*, *Toxicol. Sci.* 58, 253–269) to identify the potential for chemicals to act as ligands of mammalian estrogen receptors (ERs). The basis of the expert system was a structure activity relationship (SAR) model, based on relative binding affinity (RBA) values for steroidal and nonsteroidal chemicals derived from human ERα (hERα) competitive binding assays. The expert system enables categorization of chemicals into (RBA ranges of < 0.1, 0.1 to 1, 1 to 10, 10 to 100, and >150% relative to 17β-estradiol. In the current analysis, the algorithm was evaluated with respect to predicting RBAs of chemicals assayed with ERs from MCF7 cells, and mouse and rat uterine preparations. The best correspondence between predicted and observed RBA ranges was obtained with MCF7 cells. The agreement between predictions from the expert system and data from binding assays with mouse and rat ER(s) were less reliable, especially for chemicals with RBAs less than 10%. Prediction errors often were false positives, i.e., predictions of greater than observed RBA values. While discrepancies were likely due, in part, to species-specific variations in ER structure and ligand binding affinity, a systematic bias in structural characteristics of chemicals in the hERα training set, compared to the rodent evaluation data sets, also contributed to prediction errors. False-positive predictions were typically associated with ligands that had shielded electronegative sites. Ligands with these structural characteristics were not well represented in the training set used to derive the expert system. Inclusion of a shielding criterion into the original expert system significantly increased the accuracy of RBA predictions. With this additional structural requirement, 38 of 46 compounds with measured RBA values greater than 10% in hERα, MCF7, and rodent uterine preparations were correctly categorized. Of the remaining 129 compounds in the combined data sets, RBA values for 65 compounds were correctly predicted, with 47 of the incorrect predictions being false positives. Based upon this exploratory analysis, the modeling approach, combined with a high-quality training set of RBA values derived from a diverse set of chemical structures, could provide a credible tool for prioritizing chemicals with moderate to high ER binding affinity for subsequent *in vitro* or *in vivo* assessments.

Key Words: **structure activity relationships; expert systems; mammalian estrogen receptors; binding affinity; estrogen receptor ligands.**

Structure activity relationships (SARs) for predicting ligand-hormone receptor binding affinity have been proposed as screening tools to help prioritize untested compounds for more intensive investigations to assess potential effects on steroid signaling pathways (Ankley *et al.*, 1997; Bradbury *et al.*, 1998). Mekenyan *et al.* (1997, 1999) recently described the COmmon REactivity PAttern (COREPA) algorithm, which was developed specifically for this purpose. The algorithm is a 3-dimensional (3-D) SAR technique that assesses conformational flexibility of ligands. It permits identification and quantification of specific global and local stereoelectronic characteristics associated with the biological activity of a chemical, without the need to specify a predetermined toxicophore or the alignment of conformers to a lead compound.

In the companion paper, Bradbury *et al.* (2000) described a prototypical expert system for predicting human estrogen receptor alpha (hERα) binding affinity based on the COREPA algorithm. In that study, they defined stereoelectronic requirements associated with binding affinity of 45 steroidal and nonsteroidal ligands to the receptor. Reactivity patterns for hERα relative binding affinity (RBA; 17β-estradiol = 100%) were established, based on global nucleophilicity, interatomic distances between electronegative atoms, charge on heteroatoms, and electron donor capability of heteroatoms. These reactivity patterns were used to establish descriptor profiles, within the context of an expert system, to identify ligands with RBAs of >150%, 100 to 10%, 10 to 1%, and 1 to 0.1%. Using a "leave-one-out" evaluation, the reactivity patterns were determined to be stable and the resulting expert system correctly classified 30 of 45 compounds in this training set.

This article has been reviewed according to EPA guidelines. Mention of modeling or modeling approaches does not indicate endorsement by the EPA.

[1] To whom correspondence should be addressed. Fax: (218) 529-5015. E-mail: bradbury.steven@epa.gov.

TABLE 1
Ligands, Observed Relative Binding Affinities (RBA) to hER from MCF7 Cells, Source of Data (Ref), Number of Conformers Generated (N), and Associated Ranges of Heat of Formation, and Root Mean Square (RMS) Differences

No.	Ligand	RBA (%)	Ref.	N	ddH[kcal/mol]	RMS
1	Estradiol	100		4	–107.6159 to –96.8975	0.3343 to 0.8533
2	Estradiol–16α	80	V	3	–107.0931 to –96.6020	0.1832 to 0.4394
3	Estratrien–3–ol	40	Br	3	–62.8125 to –52.2316	0.2786 to 0.5524
4	Estrone	22	V	4	–87.0704 to –77.0100	0.1520 to 0.5643
5	Estradiol–17α	22	V	3	–107.8579 to –97.6046	0.2646 to 0.5818
6	9–11 Ene–estradiol	19.6	P	2	–84.9278 to –83.0444	0.4436
7	2–Hydroxyestratrien–17β–ol	18	Br	3	–107.6466 to –96.9750	0.3194 to 0.4266
8	6–Keto–estradiol–17β–ol	18	Br	1	–131.1277	0
9	Estriol	17	V	3	–152.4492 to –141.5218	0.3175 to 0.5773
10	4–Aminoestratriene–3,17β–diol	16	Br	2	–107.3821 to –105.3871	0.3041
11	4–Nitroestratriene–3,17β–diol	13	Br	5	–99.1052 to –88.3964	0.0972 to 0.3601
12	2–Aminoestratriene–3,17β–diol	12	Br	2	–108.0189 to –106.8755	0.2912
13	4–Fluoroestratrien–17βol	8	Br	3	–108.0670 to –97.4255	0.2997 to 0.3872
14	Estratrien–17β–ol	8	Br	3	–63.4706 to –52.8675	0.2669 to 0.3452
15	4–Nitroestratrien–3–ol,17–one	6	Br	7	–78.5453 to –68.2609	0.0575 to 0.5598
16	2–Aminoestratrien–17β–ol	4	Br	2	–64.7243 to –63.1693	0.2183
17	2–Fluoroestratrien–17β–ol	2	Br	3	–108.8062 to –98.2225	0.2952 to 0.3875
18	11β–Hydroxy–estradiol	1.68	P	3	–147.4003 to –137.0706	0.1703 to 0.9358
19	3–Methoxy estradiol–17β	1.4	Br	10	–101.0557 to –89.4757	0.1174 to 0.7713
20	2–Nitroestratriene–3,17β–diol	1	Br	4	–106.0532 to –95.4390	0.2188 to 0.5853
21	5–Androstene–3β,17β–diol	0.7	V	4	–136.9920 to –132.5624	0.1938 to 0.4582
22	5α–Androstane–3β,17β–diol	0.5	V	3	–162.4597 to –151.5775	0.3646 to 0.5040
23	11α–Hydroxy–estradiol	0.31	P	3	–151.4568 to –148.7675	0.6266 to 0.8112
24	4–Aminoestratrien–17β–ol	0.17	Br	2	–63.7826 to –61.5664	0.1552 to 0.2546
25	2–Nitroestratrien–3–ol,17–one	0.1	Br	4	–85.2338 to –78.9658	0.3211
26	11–Keto–estradiol	0.09	P	3	–130.3285 to –127.9250	0.2946 to 0.5596
27	4–Hydroxyestratrien–17β–ol	0.08	Br	3	–106.9196 to –96.1917	0.4330 to 0.4785
28	9β–Estradiol	0.07	P	2	–101.0173 to –100.9720	0.077
29	4–Nitroestratrien–17β–ol	<0.05	Br	5	–58.2424 to –47.6229	0.1915 to 0.5190
30	2–Nitroestratrien–17β–ol	<0.05	Br	2	–60.7037 to –59.2212	0.22429
31	Estratrien–17–one	<0.05	Br	3	–42.9864 to –32.9700	0.3066 to 0.3720
32	Estratriene	<0.05	Br	2	–18.5928 to –17.0508	0.2234
33	11–Keto–9β–estradiol	<0.05	P	2	–124.8020 to –124.5144	0.1461
34	5α–Androstane–3α17β–diol	<0.05	V	4	–160.7224 to –150.9430	0.1372 to 0.5407
35	4–Nonylphenol	0.021	B	196	–83.0323 to –78.3822	0.8358 to 1.7658
36	o,p′–DDT	0.0003	B	14	21.0963 to 32.7428	1.9169 to 7.3511

Note. V, VanderKuur *et al.* (1993); Br, Brooks *et al.* (1987); P, Palomino *et al.* (1994); B, Bolger *et al. (*1998).

To more completely evaluate the prototypical expert system, hERα ligand binding affinity for compounds not used in the original training set is required. Unfortunately, such data are not available in the open literature. However, a variety of ER binding affinity data sets for other experimental human and rodent models are available. While the use of RBA values from different experimental systems and species adds uncertainty to the evaluation of endpoint- and species-specific SARs, if the variability between experimental systems is within the desired precision of the predictions (i.e., variability across species and experimental systems is less than the variability across chemicals), such data can provide insights on the reliability of a model. In the current study, the 3-D SAR-based expert system derived from the hERα data set was assessed against RBA values for ER binding affinity obtained from MCF7 cells, and mouse and rat uterine preparations.

MATERIALS AND METHODS

ER Ligands and Receptor Binding Affinity

Relative binding affinity values of steroidal and nonsteroidal compounds to hER from MCF7 cells and rodent uterine ERs were used to evaluate the model described by Bradbury *et al.* (2000), which was based on RBA values derived from the hERα. Specifically, we assessed 3 data sets of 36, 35, and 58 chemicals, respectively, which were evaluated as to their affinity to MCF7 cells (Bolger *et al.,* 1998; Brooks *et al.,* 1987; Palomino *et al.,* 1994; VanderKuur *et al.,* 1993; Table 1), mouse uterine ER (Bolger *et al.,* 1998; Connor *et al.,* 1997; mER; Korach *et al.,* 1988; Table 2), and rat uterine ER (rER; Anstead *et al.,* 1989; Bolger *et al.,* 1998; Connor *et al.,* 1997; Gabbard

TABLE 2

Ligands, Observed Relative Binding Affinities (RBA) to Mouse Uterine ER (mER), Source of Data (Ref), Number of Conformers Generated (N), and Associated Ranges of Heat of Formation and Root Mean Square (RMS) Differences

No.	Ligand	RBA (%)	Ref.	N	ddH[kcal/mol]	RMS
1	Diethylstilbestrol (DES)	250	Ko	21	–47.5190 to –45.0619	0.4577 to 7.9498
2	Estradiol	100		4	–107.6159 to –96.8975	0.3343 to 0.8533
3	Estrone	60	B	4	–87.0704 to –77.0100	0.1520 to 0.5643
4	Tamoxifen	6	B	149	48.6071 to 56.0690	0.9294 to 14.9980
5	HPTE	3.25	BA	2	–56.7081 to –56.7009	3.0428
6	4 Hydroxy,2′,4′,6′–trichloro biphenyl	2.4	Ko	1	–12.8518	0
7	4,4′–Dihydroxy 2′–chloro biphenyl	1.1	Ko	2	–45.8849 to –45.6890	0.172
8	4 Hydroxy,2′,3′,4′,5′–tetrachloro biphenyl	1	Ko	3	–16.9179 to –16.8129	0.2006 to 2.2829
9	4–Nonylphenol	0.313	B*	196	–83.0323 to –78.3822	0.8358 to 11.7658
10	4–Hydroxyl 2′,6′dichloro biphenyl	0.26	Ko	2	–6.7187 to –6.6659	3.3605
11	4–t–Octylphenol	0.2	B	3	–54.1565 to –53.4612	3.3483 to 7.3376
12	4–Hydroxyl 2′,5′–dichloro biphenyl	0.2	Ko	2	–8.5800 to –8.3619	0.187
13	o,p’ DDT	0.17	B	14	21.0963 to 32.7428	1.9169 to 7.3511
14	4 Hydroxy,3,5,4′–trichloro biphenyl	0.1	Ko	3	–13.6828 to –11.4750	0.3064 to 0.4256
15	4,4′–Dihydroxy 3,5,3′,5′–tetrachloro biphenyl	0.074	Ko	3	–60.4322 to –58.0996	0.3419 to 0.4411
16	2′,3,3′,4′,5′–PentaCB–4–ol	0.072	C	4	–20.9217 to –20.8391	0.6420 to 4.8808
17	2,2′,3′,4′,6′–PentaCB–4–ol	0.044	C	2	–21.7134 to –21.7118	4.8634
18	4–Hydroxy 2–chloro biphenyl	0.04	Ko	3	–1.5043 to –1.3164	0.2039 to 2.1040
19	2,2′,3′,4′,5′–PentaCB–4–ol	0.033	C	2	–21.9928 to –21.9907	4.8319
20	2′,3,3′,5′,6′–PentaCB–4–ol	0.031	C	1	–20.4444	0
21	5α–Dihydrotestosterone	0.026	B	5	–143.5292 to –134.0742	0.2358 to 0.4780
22	4–Hydroxy 4′–chloro biphenyl	0.026	Ko	3	–4.0743 to –1.9763	0.2592 to 0.4097
23	4–Hydroxy biphenyl	0.02	Ko	4	3.1336 to 5.2432	1.4926 to 2.8044
24	4,4′–Dihydroxy biphenyl	$<$0.02	Ko	3	–41.1626 to –40.1363	0.3785 to 1.0316
25	4,4′–Dihydroxy 2′,3′,5′,6′–tetrachloro biphenyl	$<$0.02	Ko	1	–56.4487	0
26	Bisphenol A	0.018	B	2	–48.3145 to –48.2939	3.9247
27	2,2′,4′,6′–TetraCB–4–ol	0.018	C	1	–18.1477	0
28	Kepone	0.0145	BA	1	18.3543	0
29	2,2′,3′,5′,6′–PentaCB–4–ol	0.013	C	1	–21.7653	0
30	4–Nonylphenol	0.01	B*	196	–83.0323 to –78.3822	0.8358 to 11.8658
31	Methoxychlor	0.007	BA	20	–43.4469 to –39.5937	2.2637 to 6.1257
32	BBP	0.0034	B	64	–127.9410 to –117.1117	2.0314 to 9.7750
33	p,p′ DDT	0.0003	B	5	18.8319 to 20.4237	1.3120 to 6.7854
34	2′,3,4′,6′–TetraCB–4–ol	0	C	1	–16.8369	0
35	2′,3,3′,4′,6′–PentaCB–4–ol	0	C	2	–20.4800 to –20.4791	4.8799

Note. Ko, Korach *et al.* (1988); B, Bolger *et al.* (1998); C, Connor *et al.* (1997); A, average of multiple values from indicated reference.
*Chemical in more than one RBA range.

and Segaloff, 1983; Qian and Abul-Haji, 1990; Table 3). The RBA values of the ligands were calculated by dividing the concentration of test compound required to reduce the specific binding of radiolabeled 17β-estradiol (E_2) by 50% by the concentration of unlabeled E_2 required to achieve the same reduction. Species-specific RBA values obtained across studies were typically within an order of magnitude. With the exception of mouse uterine RBA values for 4-nonylphenol, average values are listed in Tables 1–3. In the case of 4-nonylphenol, two RBA values (0.313 and 0.01%) were employed in the analyses (note “compounds” 9 and 30 in Table 2).

The overlap of compounds between these data sets and those in the hERα knowledge base (Bradbury *et al.*, 2000) is summarized in Table 4. The hERα and mouse uterine data sets were most similar, with 14 compounds in common, whereas the hERα and rat uterine data sets had 4 compounds in common. For RBA values $>$1%, agreement between hERα values and those derived from the other biological models were within an order of magnitude; however, at lower RBA values, differences sometimes exceeded an order of magnitude.

ER Ligand Conformations and Molecular Descriptors

The 3-D structures of ligand conformers were generated based on the method of Ivanov *et al.* (1994), using torsion resolution, distance between nonbonded atoms and ring closure, and related parameters, as described in our companion study (see Bradbury *et al.*, 2000 and abbreviations given therein). Conformer geometry optimization was obtained with MOPAC 93 (Stewart, 1990, 1993), using the AM1 Hamiltonian with the key words $>$PRECISE$=$ and $>$NOMM$=$. For a given ligand, only conformers with a $\Delta H_f°$ within 20 kcal/mol of the $\Delta H_f°$ associated with the conformer with the absolute energy minimum were used (Tables 1–3). The conformers within this range of $\Delta H_f°$ were assumed to be energetically reasonable from a thermodynamic and kinetic perspective (Bradbury *et al.*, 1998, 2000; Ivanov *et al.*, 1998; Mekenyan *et al.*, 1997, 1999). As in our previous study (Bradbury *et al.*, 2000), it was assumed that conformers of each chemical could be considered as a statistical ensemble, based on the Boltzman’s statistics. Also included in Tables 1–3 are ligand-specific ranges of root mean square (RMS) differences between atoms

TABLE 3

Ligands, Observed Relative Binding Affinities (RBA) to Rat Uterine ER (rER), Source of Data (Ref), Number of Conformers Generated (N), and Associated Ranges of Heat of Formation, and Ranges of Root Mean Square (RMS) Differences

No.	Ligand	RBA (%)	Ref.	N	ddH[kcal/mol]	RMS
1	Diethylstilbestrol	470	B	21	–47.5190 to –45.0619	0.4577 to 7.9498
2	11β–Methyl estradiol–17β	124	G	1	–96.6593	0
3	D14 Estradiol–17β	107	G	7	–83.2854 to –70.4633	0.3367 to 1.2266
4	7α–Methyl estradiol–17β	104	G	3	–110.1453 to –99.8883	0.6168 to 0.7700
5	Estradiol	100		4	–107.6159 to –96.8975	0.3343 to 0.8533
6	1–Methyl; 3–ethyl; 6,4′–OH;2–phenylindene	81	An	12	–35.1787 to –31.5235	0.5139 to 3.1796
7	1,3–Diethyl; 6,4′–OH; 2–phenylindene	79	An	33	–40.9147 to –31.6182	0.3425 to 4.6881
8	7α–Methyl–D14 estradiol–17β	74	G	11	–85.3135 to –70.5533	0.5325 to 2.2718
9	7α–Methyl–estrone	68	G	4	–89.6667 to –80.0253	0.4258 to 0.8801
10	6–OH; 2,3–diphenylindenone–1	59	An	7	35.2282 to 37.0052	0.2739 to 2.8485
11	7α–Methyl–D14–estrone	52	G	6	–65.0343 to –48.5543	0.6721 to 2.3263
12	11β–Methyl–estrone	47	G	2	–76.6212 to –73.4133	0.6978
13	9α–Methyl–D14 estradiol–17β	41	G	2	–82.1867 to –74.4626	0.5254
14	9α–Methyl estradiol–17β	35	G	1	–105.4017	0
15	3–Ethyl; 6,4′–OH; 2–phenylindene	16	An	11	–31.8121 to –29.3166	0.2953 to 3.7704
16	1–Methyl; 6–OH; 2,3–diphenylindene	12	An	8	49.5270 to 50.5972	0.2555 to 3.6497
17	1,3–Diethyl; 4–OH; 2–phenylindene	9.3	An	38	3.3648 to 12.3888	0.3067 to 5.1054
18	D14–Estrone	9	G	6	–62.8227 to –47.0804	0.3437 to 1.4021
19	3–Phenyl; 6–OH; 2–phenylindene	8.9	An	8	52.9254 to 54.0814	0.5724 to 4.8400
20	Tamoxifen	6	Q	149	48.6071 to 56.0690	0.9294 to 14.9980
21	9α–Methyl–D14–estrone	6	G	2	–61.7544 to –53.7873	0.6163
22	7α–Methyl E2–17β 3–methyl ether	5.3	G	15	–103.7087 to –93.3945	0.0678 to 1.1556
23	11β–Methyl E2–17β 3methy ether	5.1	G	6	–97.2634 to –86.6012	0.5764 to 1.9878
24	9α–Methyl–estrone	5	G	1	–85.1291	0
25	3–Ethyl 4′–OH 2–phenylindenone–1	4.6	An	12	–6.0252 to –2.1746	1.4304 to 4.0225
26	7α–Methyl–D14 E2–17β3–methyl ether	3.1	G	33	–79.0601 to –60.1298	0.0980 to 2.9973
27	11β–[3–N,N–Dimethylamino–propoxy] estra–1,3,5(10)triene–3,17–diol	2.6	Q	60	–143.5618 to –123.7548	0.6246 to 9.5053
28	3–Ethyl 4′–OH 2–phenylindene	2.3	An	11	12.4237 to 15.7764	1.1698 to 3.2483
29	1,3–Diethyl 6–OH 2–phenylindene	2.2	An	43	3.4399 to 12.5355	0.4904 to 5.5741
30	11β–[2–N,N–Dimethylamino–ethoxy]estra–1,3,5(10)triene–3,17–diol	1.6	Q	20	–136.8826 to –118.1979	0.2646 to 6.3353
31	11β–Methyl D14 E2–17β 3–methyl ether	1.2	G	16	–77.3290 to –67.9761	0.1115 to 2.5966
32	3–Ethyl 6–OH 2–phenylindenone–1	1.2	An	12	–5.4638 to –1.8779	0.7238 to 4.3294
33	D14 E2–17β3–Methyl ether	0.8	G	20	–76.9186 to –66.3275	0.1354 to 1.3596
34	3–Ethyl 6–OH 2–phenylindene	0.58	An	11	12.5264 to 14.3330	0.3423 to 4.1639
35	4′–OH 2,3–Diphenylindenone–1	0.45	An	7	34.6060 to 36.6043	0.1824 to 2.7936
36	3–Phenyl 4′–OH 2–phenylindene	0.36	An	8	52.5707 to 54.0216	0.2870 to 3.9166
37	2,2′,3′,5′,6′–PentaCB–4–ol	0.14	C	1	–21.7653	0
38	2,2′,3′,4′,6′–PentaCB–4–ol	0.12	C	2	–21.7134 to –21.7118	4.8634
39	9α–Methyl–D14 E2–17β 3–methyl ether	0.1	G	11	–75.7626 to –66.8124	0.0830 to 1.1853
40	2′,3,3′,4′,5′–PentaCB–4–ol	0.082	C	4	–20.9217 to –20.8391	0.6420 to 4.8808
41	2′,3,3′,5′,6′–PentaCB–4–ol	0.068	C	1	–20.4444	0
42	2′,3,3′,4′,6′–PentaCB–4–ol	0.041	C	2	–20.4800 to –20.4791	4.8799
43	2,2′,3′,4′,5′–PentaCB–4–ol	0.036	C	2	–21.9928 to –21.9907	4.8319
44	3–(C6H4) 4“–OH 2–phenylindene	0.017	An	7	52.8334 to 53.9462	0.3024 to 4.8085
45	o,p′–DDT	0.01	B	14	21.0963 to 32.7428	1.9169 to 7.3511
46	2,3–Diphenylindenone–1	0.0095	An	7	79.1399 to 80.6384	1.1089 to 5.2588
47	2,2′,4′,6′–TetraCB–4–ol	0.0005	C	1	–18.1477	0
48	Dieldrin	0.0005	B	2	224.2979 to 224.5710	6.6002
49	2′,3,4′,6′–TetraCB–4–ol	0	C	1	–16.8369	0
50	11β–Methyl–D14–estrone 3–methyl ether	0	G	16	–56.9465 to –38.5028	0.1548 to 2.6509
51	9α–Methyl–D14–estrone 3–methyl ether	0	G	10	–55.3152 to –46.3932	0.1663 to 1.1154
52	7α–Methyl–D14–estrone 3–methyl ether	0	G	18	–58.6270 to –39.5063	0.0876 to 4.0087
53	D14–Estrone 3–methyl ether	0	G	20	–56.3380 to –43.4467	0.0846 to 1.6183
54	11β–Methyl–estrone–3–methyl ether	0	G	4	–70.3216 to –66.7593	0.5265 to 1.3343
55	9α–Methyl–estrone–3–methyl ether	0	G	6	–78.7993 to –77.2899	0.1260 to 0.5736
56	7α–Methyl–estrone–3–methyl ether	0	G	9	–83.1343 to –73.5141	0.5680 to 1.7791
57	Estrone–3–methyl ether	0	G	11	–80.6124 to –70.5378	0.0990 to 1.1738
58	9α–Methyl E2–17β 3–methyl ether	0	G	4	–99.0673 to –97.5325	0.1926 to 0.5440

Note. An, Anstead *et al.* (1989); B, Bolger *et al.* (1998); C, Connor *et al.* (1997); G, Gabbard and Segaloff (1983); Q, Qian and Abul–Hajj (1990).

TABLE 4
Summary of RBA Values (from Tables 1–3) of Chemicals Tested for ER Binding Affinity in More than One Biological Model

Ligand	hERα[a]	MCF7 (hER)[b]	rER[c]	mER[d]
Diethylstilbestrol	294	–	470	250
17β–Estradiol	100	100	100	100
Estrone	60	22	–	60
17α–Estradiol	58	22	–	–
Estriol	14	17	–	–
5–Androstene–3β,17β–diol	6	0.7	–	–
Tamoxifen	5.1	–	6	6
HPTE	1.7	–	–	3.25
o,p'–DDT	0.4	0.0003	0.01	0.13
4–Nonylphenol	0.3	0.021	–	0.313, 0.01,
4–t–Octylphenol	0.2	–	–	0.2
Kepone	0.2	–	–	0.0145
p,p′–DDT	0.06	–	–	0.0003
Bisphenol A	0.045	–	–	0.018
5α–Dihydrotestosterone	0.03	–	–	0.026
BBP	0.015	–	–	0.0034
Methoxychlor	0.012	–	–	0.0038, 0.01
Dieldrin	0.003	–	0.0005	–
2,2′,3′,5′,6′–PentaCB–4–ol	–	–	0.14	0.013
2,2′,3′,4′,6′–PentaCB–4–ol	–	–	0.12	0.044
2′,3,3′,4′,5′–PentaCB–4–ol		–	0.082	0.072
2′,3,3′,5′,6′–PentaCB–4–ol	–	–	0.068	0.031
2′,3,3′,4′,6′–PentaCB–4–ol	–	–	0.041	0.0
2,2′,3′,4′,5′–PentaCB–4–ol			0.036	0.033
2,2′,4′,6′–TetraCB–4–ol	–	–	0.0005	0.018
2′,3,4′,6′–TetraCB–4–ol	–	–	0.0	0.0

[a]Data from Kuiper *et al*. (1997) and Bolger *et al*. (1998).
[b]Data from Brooks *et al*. (1987), VanderKuur *et al*. (1993), Palomino *et al*. (1994), and Bolger *et al*. (1998).
[c]Data from Gabbard and Segaloff (1983), Anstead *et al*. (1989), Qian and Abul–Hajj (1990), Conner *et al*. (1997), and Bolger *et al*. (1998).
[d]Data from Korach *et al*. (1988), Conner *et al*. (1997), and Bolger *et al*. (1998).

of each conformer with the corresponding atoms in the lowest-energy conformer. As in our preceding study (Bradbury *et al.*, 2000), conformers of a given chemical within the specified 20 kcal/mol range of $\Delta H_f°$ often exhibited significant variation in potentially relevant electronic descriptors (data not shown). This observation is consistent with previous studies highlighting the necessity of including all energetically reasonable conformers when defining common reactivity patterns (Bradbury *et al.*, 1998, 2000; Mekenyan *et al.*, 1997, 1999).

To generate common reactivity patterns, the same set of global and local molecular descriptors used in our previous study (Bradbury *et al.*, 2000) were employed. These descriptors were associated with global nucleophilicity, heteroatom electronegativity and charge, and interatomic distances between heteroatoms.

Evaluation of the COREPA-Based hERα Ligand Reactivity Patterns

A summary of the COREPA method to assess hERα binding affinity was provided by Bradbury *et al.* (2000), while the conceptual basis and mathematical derivations for the method are reported elsewhere (Mekenyan *et al.*, 1997, 1999). Using this technique, a decision tree was developed to predict RBA ranges for chemical binding to hERα. The decision tree, based on the energy of the highest occupied molecular orbital (E_{HOMO}), interatomic distances between heteroatoms (d(R_R)), charge of a heteroatom (Q(R)), and donor delocalizabilities of heteroatoms ($S^E(R)$) was optimized to first minimize the probability of false negative identifications (i.e., underpredicted RBA values), while secondarily minimizing the number of false positive identifications. In the current study the decision tree was modified, as summarized below, using additional screens described by Bradbury *et al.* (2000), to further minimize the probability of false negative identifications. These modifications were applied with the realization that an increase in false positive predictions likely would result.

A "prescreen" reactivity pattern was used to eliminate those compounds whose RBA values were likely not to exceed 0.1%. Thus, conformers which had E_{HOMO} values of less than –9.95 eV, electronegative sites not meeting a $S^E(R)$ range of 0.239 to 0.277 $(a.u.)^2/eV$, or steroids not conforming to stereochemical requirements of the natural enantiomer were assigned a 0% probability of having an RBA value >0.1%. A reactivity pattern, with $E_{HOMO} > -8.99$ eV combined with a d(R_R) range of 11.77 to 12.22 Å between heteroatoms and a Q(R) range of −0.272 to −0.233 a.u. (imposed on both electronegative sites forming the d(R_R)), was employed to identify chemicals with an RBA value >150%. The reactivity pattern for the binding activity range 100 > RBA > 10% was based on an E_{HOMO} pattern > –9.44 eV, combined with d(R_R) ranges of 10.62 to 10.95, 10.38 to 10.51, or 11.50 to 11.80 Å, and the requirement that at least one of the heteroatoms in the distance range meet the Q(R) screen of −0.273 to −0.236 a.u. For the activity range of 10 > RBA > 1% a pattern was derived based on a $E_{HOMO} > -9.87$ eV, combined with distance screens of 9.38 to 9.93, 9.75 to 10.44, or 10.56 to 11.28 Å and a $S^E(R)$ pattern of 0.237 to 0.273 $(a.u.)^2/eV$, imposed on at least one electronegative site. Finally, the

TABLE 5
The Predicted RBA Ranges Using the Decision Tree Based on the hERα Reactivity Pattern, for Ligands with Measured Binding Affinity to hER from MCF7 Cells

			Predicted RBA ranges (%)			
No.	Ligand	Measured RBA (%)	00 > RBA > 10	10 > RBA > 1	1 > RBA > 0.1	RBA < 0.1
1	Estradiol	100 > RBA > 10	x			
2	Estradiol –16α	100 > RBA > 10	x			
3	Estratrien–3–ol	100 > RBA > 10			x	
4	Estrone	100 > RBA > 10	x			
5	Estradiol –17α	100 > RBA > 10	x			
6	9–11 Ene–estradiol	100 > RBA > 10	x			
7	2–Hydroxyestratrien–17β–ol	100 > RBA > 10		x		
8	6–Keto–estradiol–17β–ol	100 > RBA > 10	x			
9	Estriol	100 > RBA > 10	x			
10	4–Aminoestratriene–3,17β–diol	100 > RBA > 10	x			
11	4–Nitroestratriene–3,17β–diol	100 > RBA > 10		x		
12	2–Aminoestratriene–3,17β–diol	100 > RBA > 10	x			
13	4–Fluoroestratrien–17β–ol	10 > RBA > 1		x		
14	Estratrien–17β–ol	10 > RBA > 1			x	
15	4–Nitroestratrien–3–ol,17–one	10 > RBA > 1		x		
16	2–Aminoestratrien–17β–ol	10 > RBA > 1		x		
17	2–Fluoroestratrien–17β–ol	10 > RBA > 1		x		
18	11β–Hydroxy–estradiol	10 > RBA > 1	x			
19	3–Methoxy estradiol–17β	10 > RBA > 1		x		
20	2–Nitroestratriene–3,17β–diol	1 > RBA > 0.1		x		
21	5–Androstene–3β, 17β–diol	1 > RBA > 0.1		x		
22	5 α–Androstane–3β, 17β–diol	1 > RBA > 0.1				x
23	11α–Hydroxy–estradiol	1 > RBA > 0.1	x			
24	4–Aminoestratrien–17β–ol	1 > RBA > 0.1		x		
25	2–Nitroestratrien–3–ol,17–one	0.1 > RBA > 0.01		x		
26	11–Keto–estradiol	0.1 > RBA > 0.01	x			
27	4–Hydroxyestratrien–17β–ol	0.1 > RBA > 0.01		x		
28	9β–Estradiol	0.1 > RBA > 0.01				x
29	4–Nitroestratrien–17β–ol	0.1 > RBA > 0.01			x	
30	2–Nitroestratrien–17β–ol	0.1 > RBA > 0.01				x
31	Estratrien–17–one	0.1 > RBA > 0.01			x	
32	Estratriene	0.1 > RBA > 0.01				x
33	11–Keto–9β–estradiol	0.1 > RBA > 0.01				x
34	5α–Androstane–3α,17β–diol	0.1 > RBA > 0.01				x
35	4–Nonylphenol	0.1 > RBA > 0.01			x	
36	o,p′–DDT	0.01 > RBA			x	

Note. Chemicals are Assigned to the Highest Activity Range Predicted

reactivity pattern based on $E_{HOMO} > -9.93$ eV and $S^E(R)$ of 0.239 to 0.269, 0.248 to 0.279, or 0.300 to 0.330 $(a.u.)^2/eV$ was associated with the low binding activity range of 1 >RBA > 0.1%. These reactivity patterns were organized in a hierarchical decision tree that sequentially assessed the energetically reasonable conformers of a ligand in decreasing order of RBA ranges. Once identified as meeting a reactivity pattern for a particular binding activity range, a compound was assigned to that RBA range and not evaluated using patterns associated with lower RBA ranges.

The ER ligands in Tables 1–3 were used to evaluate the ability of this decision tree, and associated reactivity patterns, to predict RBA ranges for chemicals not included in the original hERα training set (see Bradbury *et al.*, 2000). Each energetically reasonable conformer of a chemical was processed through the decision tree by making use of an interpreter based on the SMILES algorithm, which permits the use of stereoelectronic structure-based rules. The decision tree provided a binary discrimination (i.e., a "yes" or "no" determination) of chemicals being within a specified RBA range. Thus, a chemical would be predicted to have an hERα affinity within a specified RBA range if at least one of its conformers met the associated reactivity pattern.

RESULTS AND DISCUSSION

MCF7 Cell Data Set (Table 5)

Nine of 12 chemicals with measured ER binding affinity within 10 > RBA > 100% were correctly categorized. Five out of these 9 compounds were not in the original hERα data set. Estratrien-3-ol (3; RBA = 107%) was incorrectly predicted to have 1 > RBA > 0.1%, while 2-hydroxyestratien-17β-ol (7; RBA = 79%) and 4-nitroestratriene-3,17β-diol (11;

RBA = 52%) were incorrectly predicted to have 10 > RBA > 1%. Explanations for these false-negative identifications include lack of a second electronegative site (3) or slightly lower global electron donor ability ($-9.57 < E_{HOMO} < -9.51$ eV) and shorter than required distances between electronegative atoms (9.1–9.9 Å) for compounds 7 and 11. 11β-Hydroxy-estradiol (18; RBA = 1.7%), 11α-hydroxy-estradiol (23; RBA = 0.31%) and 11-keto-estradiol (26; RBA = 0.1%) were incorrectly predicted as having RBAs greater than 10% (i.e., false positive identifications).

For compounds with observed RBA values between 1 and 10%, estratrien-17β-ol (14; RBA = 8%) was incorrectly predicted to have an RBA between 0.1 and 1%, due to the lack of a second electronegative site. False positive identifications using the reactivity pattern for the RBA range between 1 and 10% included 2-nitroestratriene-3,17β-diol (20), 5-androstene-3β, 17β-diol (21), 4-aminoestratrien-17β-ol (24), 2-nitroestratrien-3-ol,17-one (25), and 4-hydroxyestratien-17β-ol (27), with measured RBA values of 1, 0.7, 0.17, 0.1, and 0.08%, respectively. One of the false-positives, 5-androstene-3β, 17β-diol (21), was also in the hERα training set; however, in that biological model, the observed RBA was 6%. Thus, the discrepancy for this compound may reflect variability across receptor systems and/or laboratories.

For compounds with observed RBA values between 0.1 and 1%, 5α-androstane-3β, 17β-diol (22; RBA = 0.5%) was the only false negative identification, due to a very low global nucleophilicity ($-10.35 < E_{HOMO} < -10.31$ eV). Four compounds were false positive identifications for binding affinity in this range, including 29 (measured RBA < 0.05%), 31 (RBA < 0.05%), 35 (RBA = 0.021%), and 36 (RBA = 0.0003%). It should be noted that compounds 35 and 36 were evaluated using the hERα, with measured RBA values of 0.3 and 0.4%, respectively (Bradbury *et al.*, 2000), hence discrepancies between observed and predicted values for these compounds are likely due to variability across receptor systems.

In summary, of the 36 compounds in the MCF7 cell data set, 19 RBA ranges were correctly predicted, with RBA ranges for 5 compounds underpredicted (false negative identifications) and 12 overpredicted (false positives). Based on several compounds in common between the MCF7 cell data set and the hERα training set, it appears that the false positive identifications may be due, in part, to inherently higher binding affinity in the hERα system. For the false negative identifications with compounds having observed RBA values between 10 and 100%, the most notable discrepancy was observed for compound 3 and suggests that in some cases one, rather than two, electronegative sites may be required for ER binding in MCF7 cells.

Mouse Uterine Data Set (Table 6)

The common reactivity patterns for RBA ranges greater than 150% and for 10 to 100%, did not result in any false positive identifications. Compounds with measured RBA values greater than 10%, which were also in the hERα training set, were correctly discriminated.

For compounds with measured RBA values between 1 and 10%, 3 of 4 compounds were correctly categorized. Tamoxifen (4; RBA = 6%) was incorrectly predicted to have an RBA between 0.1 and 1%. This is consistent with the incorrect identification also observed for tamoxifen with the hERα data set (Bradbury *et al.*, 2000); this is due to the fact that the maximum interatomic distance between electronegative atoms for tamoxifen is much less than that specified in the common reactivity pattern. Of the 3 compounds whose RBA values were correctly predicted, two (6, 4-hydroxy-2′,4′,6′-trichlorobiphenyl, RBA = 2.4%; and 7, 4,4′-dihydroxy-2′-chlorobiphenyl, RBA = 1.1%) were not in the original hERα training set.

In terms of false positive classifications, the reactivity pattern for 10 > RBA > 1% incorrectly identified chemical 8 (4-hydroxy-2′,3′,4′,5′-tetrachlorobiphenyl; RBA = 1.0%) and several compounds with measured RBA values in the range of 0.01 to 0.1% (compounds 14–17, 19, 22, 24, 25, and 27) and measured RBA values less than 0.01% (compounds 31, 34, and 35). An analysis of this set of structures suggests that RBA values can be significantly reduced if the electronegative site (i.e., a charge greater than −0.3 a.u.) is shielded. With an additional rule added to the expert system in which ligands with atoms or fragments, either directly bonded to an electronegative heteroatom or in an ortho position to the heteroatom, were considered incapable of binding with an RBA >0.1%, the number of false positive identifications decreased from 13 to 5. The remaining false positive identifications were compounds 17, 22, 24, 27, and 34. Compound 8, with a measured RBA of 1.0%, was a false negative with a predicted RBA of <0.1%. This steric requirement was also observed in a recent application of the COREPA approach to the data set analyzed by Waller *et al.* (1996), which consisted of 9 steroidal and 49 nonsteroidal ligands (unpublished data).

The reactivity pattern for the range of 0.1 to 1% correctly predicted 5 of the 6 compounds with observed RBA values within this range, with 2 of the chemicals (10, 12) not in the original hERα training set. The pattern also identified 8 false positive ligands whose measured RBA values were 1 > RBA > 0.1% (compounds 18, 20, 23, 26, 29, 30, 32, and 33). In this case, use of the shielding rule did not reduce the number of false positives.

In summary, for the 35 compounds in the mER data set, 13 compounds were predicted to bind within the correct RBA ranges. Of the 22 incorrect classifications, one was a false negative and 21 were false positive predictions. Inclusion of the shielding rule in the expert system resulted in 20 correct classifications, with the number of false positive predictions reduced from 21 to 13 compounds, and with one additional false negative prediction. For the 7 compounds with measured RBA values between 150 and 1%, the one false negative prediction was associated with tamoxifen.

TABLE 6
The Predicted RBA Ranges Using the Decision Tree Based on the hERα Reactivity Pattern for Ligands with Measured Binding Affinity to Mouse Uterine Estrogen Receptors (mER)

			Predicted RBA ranges				
No.	Ligand	Measured RBA (%)	> 150	100 > RBA > 10	10 > RBA > 1	1 > RBA > 0.1	RBA < 0.1
1	Diethylstilbestrol	> 150	x				
2	Estradiol	100 > RBA > 10		x			
3	Estrone	100 > RBA > 10		x			
4	Tamoxifen	10 > RBA > 1				x	
5	HPTE	10 > RBA > 1			x		
6	4 Hydroxy,2′,4′,6′–trichloro biphenyl	10 > RBA > 1			x		
7	4,4′–Dihydroxy 2′–chloro biphenyl	10 > RBA > 1			x		
8	4 Hydroxy,2′,3′,4′,5′–tetrachlorobiphenyl	1 > RBA > 0.1			x		
9	4–Nonylphenol	1 > RBA > 0.1				x	
10	4–Hydroxyl 2′,6′dichloro biphenyl	1 > RBA > 0.1				x	
11	4–t–Octylphenol	1 > RBA > 0.1				x	
12	4–Hydroxyl 2′,5′–dichloro biphenyl	1 > RBA > 0.1				x	
13	o,p' DDT	1 > RBA > 0.1				x	
14	4 Hydroxy,3,5,4′–trichlorobiphenyl	0.1 > RBA > 0.01			x		
15	4,4′–Dihydroxy 3,5,3′,5′–tetrachlorobiphenyl	0.1 > RBA > 0.01			x		
16	2′,3,3′,4′,5′–PentaCB–4–ol	0.1 > RBA > 0.01			x		
17	2,2′,3′,4′,6′–PentaCB–4–ol	0.1 > RBA > 0.01			x		
18	4–Hydroxy 2–chloro biphenyl	0.1 > RBA > 0.01				x	
19	2,2′,3′,4′,5′–PentaCB–4–ol	0.1 > RBA > 0.01			x		
20	2′,3,3′,5′,6′–PentaCB–4–ol	0.1 > RBA > 0.01				x	
21	5a–Dihydrotestosterone	0.1 > RBA > 0.01					x
22	4–Hydroxy 4′–chloro biphenyl	0.1 > RBA > 0.01			x		
23	4–Hydroxy biphenyl	0.1 > RBA > 0.01				x	
24	4,4′–Dihydroxy biphenyl	0.1 > RBA > 0.01			x		
25	4,4′–Dihydroxy 2′,3′,5′,6′–tetrachlorobiphenyl	0.1 > RBA > 0.01			x		
26	Bisphenol A	0.1 > RBA > 0.01				x	
27	2,2′,4′,6′–TetraCB–4–ol	0.1 > RBA > 0.01			x		
28	Kepone	0.1 > RBA > 0.01					x
29	2,2′,3′,5′,6′–PentaCB–4–ol	0.1 > RBA > 0.01				x	
30	4–Nonylphenol	0.01 > RBA > 0.00				x	
31	Methoxychlor	0.01 > RBA > 0.00			x		
32	BBP	0.01 > RBA > 0.00				x	
33	p,p' DDT	0.01 > RBA > 0.00				x	
34	2′,3,4′,6′–TetraCB–4–ol	0.01 > RBA > 0.00			x		
35	2′,3,3′,4′,6′–PentaCB–4–ol	0.01 > RBA > 0.00			x		

With the inclusion of the shielding rule, the 13 false positive predictions included 9 biphenyl compounds, with measured RBA values typically between 0.1 and 0.01%. These low-affinity biphenyl compounds were not represented in the hERα training set, which may have led to a bias in the reactivity patterns for RBA values between 10 and 1% and 1 and 0.1%. Other remaining false positives were 30 (4-nonylphenol; RBA = 0.01), which was measured to have a mER binding affinity of 0.313 (9), similar to that measured for hERα (0.3), and 26 (bisphenol A), 32 (BBP), and 33 (p,p′-DDT), all with lower observed affinity to mouse receptors than previously measured for hERα (Table 4).

Rat Uterine Data Set (Table 7)

An evaluation of the RBA > 150% and 100 > RBA > 10% screening patterns against the rat RBA data set resulted in 5 false negative and 5 false positive predictions. The reactivity pattern derived for RBA > 150% correctly identified 1 (diethylstilbestrol; RBA = 470%), with no false positive identifications. The rER data set contained 3 compounds with observed RBA values between 100 and 150%, a range not available in the hERα training set. Although technically classified as false negatives, the hERα-based pattern for 100 > RBA > 10% identified 3 (D14 estradiol-17β; RBA = 107%) and 4 (7α-

TABLE 7

The RBA Ranges Predicted Using the Decision Tree Based on the hERα Reactivity Pattern for Ligands with Measured Affinity to Rat Uterine Estrogen Receptors (rER)

No.	Ligand	Measured RBA (%)	Predicted RBA ranges (%)					
			>150	150 > RBA > 100	100 > RBA > 10	10 > RBA > 1	1 > RBA > 0.1	RBA < 0.1
1	Diethylstilbestrol	>150	x					
2	11β-Methyl estradiol-17β	150 > RBA > 100				x		
3	D14 Estradiol-17β	150 > RBA > 100			x			
4	7α-Methyl estradiol-17β	150 > RBA > 100			x			
5	Estradiol	100 > RBA > 10			x			
6	1-Methyl; 3-ethyl; 6,4′-OH; 2-phenylindene	100 > RBA > 10			x			
7	1,3-Diethyl; 6,4′-OH; 2-phenylindene	100 > RBA > 10			x			
8	7α-Methyl-D14 estradiol-17β	100 > RBA > 10			x			
9	7α-Methyl-estrone	100 > RBA > 10			x			
10	6-OH; 2,3-Diphenylindenone-1	100 > RBA > 10					x	
11	7α-Methyl-D14-estrone	100 > RBA > 10			x			
12	11β-Methyl-estrone	100 > RBA > 10			x			
13	9α-Methyl-D14 estradiol-17β	100 > RBA > 10			x			
14	9α-Methyl estradiol-17β	100 > RBA > 10			x			
15	3-Ethyl; 6,4′-OH; 2-phenylindene	100 > RBA > 10			x			
16	1-Methyl; 6-OH; 2,3-diphenylindene	100 > RBA > 10					x	
17	1,3-Diethyl; 4-OH; 2-phenylindene	10 > RBA > 1					x	
18	D14-Estrone	10 > RBA > 1			x			
19	3-Phenyl; 6-OH; 2-phenylindene	10 > RBA > 1					x	
20	Tamoxifen	10 > RBA > 1					x	
21	9α-Methyl-D14-estrone	10 > RBA > 1			x			
22	7α-Methyl E2-17β 3-methyl ether	10 > RBA > 1				x		
23	11β-Methyl E2-17β 3-methyl ether	10 > RBA > 1				x		
24	9α-Methyl-estrone	10 > RBA > 1			x			
25	3-Ethyl 4′-OH 2-phenylindenone-1	10 > RBA > 1					x	
26	7α-Methyl-D14 E2-17β3-methyl ether	10 > RBA > 1				x		
27	11β-[3-N,N-Dimethyl-aminopropoxy] estra-1,3,5(10)triene-3,17-diol	10 > RBA > 1			x			
28	3-Ethyl 4′-OH 2-phenylindene	10 > RBA > 1					x	
29	1,3-Diethyl 6-OH 2-phenylindene	10 > RBA > 1					x	
30	11β-[2-N,N-Dimethyl-aminoethoxy] estra-1,3,5(10)triene-3,17-diol	10 > RBA > 1			x			
31	11β-Methyl D14 E2-17β 3-methyl ether	10 > RBA > 1				x		
32	3-Ethyl 6-OH 2-phenylindenone-1	10 > RBA > 1					x	
33	D14 E2-17β3-Methyl ether	1 > RBA > 0.1				x		
34	3-Ethyl 6-OH 2-phenylindene	1 > RBA > 0.1					x	
35	4′-OH 2,3-Diphenylindenone-1	1 > RBA > 0.1					x	
36	3-Phenyl 4′-OH 2-phenylindene	1 > RBA > 0.1					x	
37	2,2′,3′,5′,6′-PentaCB-4-ol	1 > RBA > 0.1					x	
38	2,2′,3′,4′,6′-PentaCB-4-ol	1 > RBA > 0.1				x		
39	9α-Methyl-D14 E2-17β 3-methyl ether	0.1 > RBA > 0.01				x		
40	2′,3,3′,4′,5′-PentaCB-4-ol	0.1 > RBA > 0.01				x		
41	2′,3,3′,5′,6′-PentaCB-4-ol	0.1 > RBA > 0.01					x	
42	2′,3,3′,4′,6′-PentaCB-4-ol	0.1 > RBA > 0.01				x		
43	2,2′,3′,4′,5′-PentaCB-4-ol	0.1 > RBA > 0.01				x		
44	3-(C6H4) 4″-OH 2-Phenylindene	0.1 > RBA > 0.01						x
45	o,p′-DDT	0.01 > RBA					x	
46	2,3-Diphenylindenone-1	0.01 > RBA						x
47	2,2′,4′,6′-TetraCB-4-ol	0.01 > RBA				x		
48	Dieldrin	0.01 > RBA					x	
49	2′,3,4′,6′-TetraCB-4-ol	0.01 > RBA				x		
50	11β-Methyl-D14-estrone 3-methyl ether	0.01 > RBA				x		
51	9α-Methyl-D14-estrone 3-methyl ether	0.01 > RBA				x		
52	7α-Methyl-D14-estrone 3-methyl ether	0.01 > RBA				x		
53	D14-Estrone 3-methyl ether	0.01 > RBA				x		
54	11β-Methyl-estrone-3-methyl ether	0.01 > RBA				x		
55	9α-Methyl-estrone-3-methyl ether	0.01 > RBA				x		
56	7α-Methyl-estrone-3-methyl ether	0.01 > RBA				x		
57	Estrone-3-methyl ether	0.01 > RBA				x		
58	9α-Methyl E2-17β 3-methyl ether	0.01 > RBA				x		

methyl estradiol-17β; RBA = 104%). The third chemical in this range (2; 11β-methyl estradiol-17β; RBA = 124%) was identified as having an RBA between 10 and 1%. For the 12 chemicals with observed RBA values between 10 and 100%, the reactivity pattern for 100 > RBA > 10% resulted in false negative predictions for 10 (6-hydroxy-2,3-diphenylindenone-1; RBA = 59%) and 16 (1-methyl-6-hydroxy; 2,3-diphenylindene; RBA = 12%), which were predicted to have RBAs between 0.1 and 1%. These incorrect predictions were due to short distances between electronegative sites and the lack of a second electronegative site, respectively. False positive identifications obtained with the 100 > RBA > 10% pattern included compounds 18 (RBA = 9%), 21 (RBA = 6%), 24 (RBA = 5%), 27 (RBA = 2.6%), and 30 (RBA = 1.6%). Using the additional filter for nonshielded electronegative sites (see *Mouse Uterine Data Set*), compounds 27 and 30 would not be predicted to have RBA values between 10 and 100%, but would be predicted to have RBA < 0.1, and therefore be identified as false negatives.

Application of the reactivity pattern for 10 > RBA > 1% to compounds with observed RBA values between 1 and 10% resulted in false negative predictions for 17 (1,3-diethyl-4-hydroxy-2-phenylindene; RBA = 9.3%), 19 (3-phenyl-6-hydroxy-2-phenylindene; RBA = 8.9%), 20 (tamoxifen; RBA = 6%), 25 (3-ethyl-4′-hydroxy-2-phenylindenone-1; RBA= 4.6%), 28 (3-ethyl-4′-hydroxy-2-phenylindene; RBA = 2.3%), 29 (1,3-diethyl-6-hydroxy-2-phenylindene; RBA = 2.2%), and 32 (3-ethyl-6-hydroxy-2-phenylindenone-1; RBA = 1.2%). For all of these compounds, the predicted RBA ranges were between 0.1 and 1%. Reasons for these incorrect classifications included the lack of a second electronegative site (17, 19, 25, 28, 29, and 32) or a small distance between electronegative sites (20; 4 Å). As noted in Table 7, the 10 > RBA > 1% screen resulted in 17 false-positive identifications, including 2 chemicals in the RBA range of 0.1 to 1% (33 and 38), 4 chemicals in the RBA range of 0.01 to 0.1% (39, 40, 42, and 43), and 11 chemicals with RBA values of less than 0.01% (47, 49–58). When the shielding screen for electronegative heteroatoms was employed, the number of false positive identifications decreased to 5 chemicals (33, 39, 47, 49, and 58), with one additional false negative, 38 (2,2′,3′,4′,6′-penta CB-4-ol; RBA = 0.12%), with a predicted RBA < 0.1.

Using the screening rule for RBA values between 1 and 0.1%, false positive identifications were noted for 41 (RBA = 0.068%), 45 (RBA = 0.01%), and 48 (0.0005%). Of these, 45 and 48 had greater hERα measured affinities (Table 4), again suggesting interspecies or interlaboratory differences as a basis for the discrepancy.

In summary, of the 58 compounds in the rat uterine data set, RBA ranges were correctly predicted for 21 ligands, with 12 false negative and 25 false positive classifications. Thirty-four compounds were, however, correctly classified with inclusion of the shielding rule, which decreased the number of false-positive identifications to 11 ligands, while increasing the number of false negatives to 15. For the 16 compounds with measured RBA values greater than 10%, there were 5 false negative identifications. For the 16 compounds with measured RBA values between 10 and 1%, 3 false positive identifications and two additional false negatives were noted (after inclusion of the shielding rule). Of the 7 ligands in this range incorrectly predicted to have an RBA value between 1 and 0.1%, the 4 most notable false negative predictions were associated with compounds whose measured RBA values ranged from 9.3 to 4.6%. The remaining 3 ligands had measured RBA values that ranged from 2.3 to 1.2%. Finally, for the 26 ligands with measured RBA values less than 1%, there was one false negative prediction generated upon inclusion of the shielding rule, and 8 false positive predictions.

Summary and Conclusions

Development and evaluation of similarity relationships to predict a specific type of biological activity based upon chemical structure requires the establishment of a knowledge base that contains training and evaluation sets of chemicals whose modes of action and potency are well defined. Minimizing the introduction of biological variability in the model development and evaluation process is critical to assessing the performance of SARs. It is also important that the training set of chemicals represents a range of structures, and associated properties, representative of the "chemical universe" of interest. Structure similarity relationships based on well-defined endpoints, and developed across a diverse set of chemicals, provide transparent models whose uncertainties can be better defined. It is also essential to define the required precision of a model to determine the data quality in the training and evaluation knowledge bases. In the present study, the model predicts RBA values within a factor of 10 ranges across 6 orders of magnitude. Consequently, variability of less than 10-fold in the training and evaluation data sets is not problematic. Of course, other applications of these training and evaluation data sets may require greater levels of precision and accuracy.

To develop a model to predict potential ER binding affinity from chemical structure, Bradbury *et al.* (2000) used a data set of 45 compounds that had been assayed with the hERα. Although a "leave-one-out" statistical approach was used to evaluate common reactivity patterns for 4 RBA ranges between 150 to 0.1%, it was not possible to independently assess the model with compounds not used in the training set. To more completely evaluate the model, an optimum approach would be to compare hERα RBA predictions to measured hERα values for chemicals not used in the training set. Unfortunately, such a data set did not exist in the open literature. While risking the addition of interspecies/test system variability in the evaluation process, the current study employed a data set of 99 structures that were assessed in MCF7 cells (hER) and rodent models (mER and rER) as a means to determine ability of the model to assess the activity of compounds not used in the original

training set. For those 17 compounds that overlapped with the original hERα data set, measured RBA values were within an order of magnitude.

The largest percentage of correct predictions was obtained for the MCF7 data set, which was most likely due to similarity in chemical structures between the two data sets, as well as the fact that the two systems are based on a similar, perhaps the same, receptor. The increase in incorrect predictions with the mouse and rat data sets was largely due to false positive identifications. The rate of false positive identifications was markedly reduced when an additional rule was included that required at least one electronegative atom to be unshielded. This occurred at the expense of a slight increase in the number of false negatives. The observation that the electronegative atom must be unshielded could reflect differences in the hERα and rodent receptors, or reflect a bias in the original training set where the occurrence of shielded electronegative atoms was rare. An analysis of the data set reported by Waller *et al.* (1996), where binding affinity was expressed in terms of pKi rather than RBA values, also indicated the need for an unshielded electronegative atom, which suggests the original hERα training set may not have had sufficient diversity in chemical structure. Assuming the false positive error rates for the rodent data sets are not primarily due to interspecies differences, but instead due to a bias in the original training set, a modification to the expert system rules that requires at least one electronegative atom be unshielded appears warranted.

Figure 1A summarizes the results of using the hERα-based reactivity patterns to predict RBA ranges from the combined hERα, MCF7 cell and rodent data sets using the expert system. Figure 1B represents predictions obtained with inclusion of the shielding rule. Eight false negatives were identified among the 46 ligands (31 compounds in the MCF7 cell, mER and rER data sets) whose measured RBA values were greater than 10%. No false positive identifications were observed for ligands with measured RBA values >10%. Thus, the interspecies comparison, undertaken assuming similarity between ER binding domains in the human and rodent assays, provides a reasonably robust screening result for ER ligands whose binding affinities are at least 10% of E_2, independent of the mammalian receptor system. Thirteen of the remaining 129 compounds with RBA values < 10% were incorrectly classified as false negatives, without the shielding rule (Fig. 1A), and 17 with the shielding rule (Fig. 1B). The number of false positives obtained for chemicals with RBA < 10% was 69 with the original expert system and 47 upon incorporation of the shielding rule. The increased occurrence of false positive identifications for chemicals with lower binding affinities is consistent with a lower level of biological similarity to E_2 and therefore a lower level of chemical similarity and specificity of reactivity patterns (Bradbury *et al.*, 2000).

As reported in a recent summary from a workshop sponsored by the U.S. Environmental Protection Agency (EPA), ranking and prioritization schemes for screening industrial chemicals

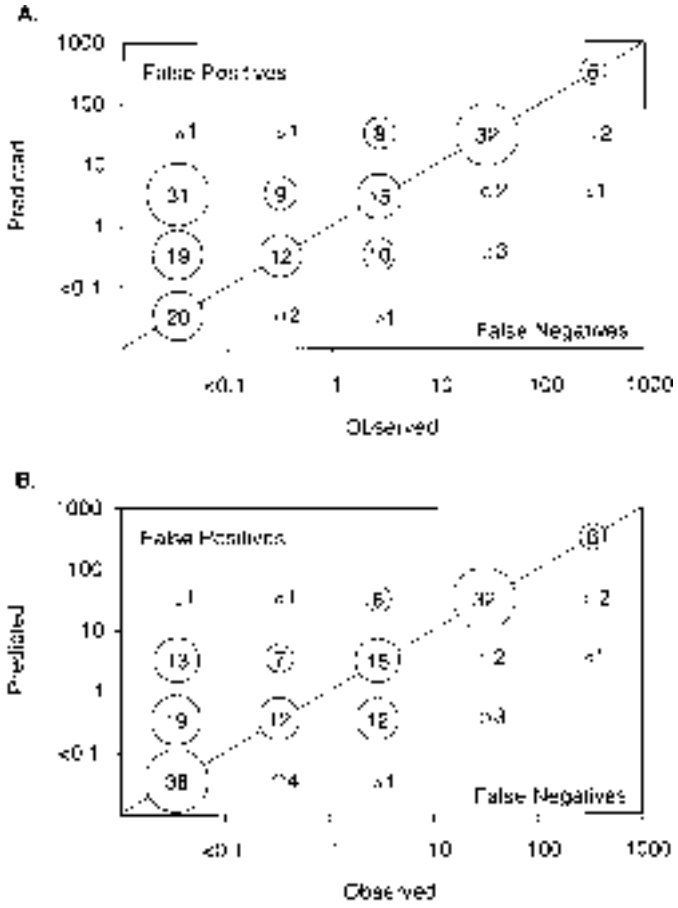

FIG. 1. Relationship of observed hERα, MCF7 cell, mouse uterine, and rat uterine RBAs to RBAs predicted from an expert system reported by Bradbury *et al.* (2000); (A) original model; (B) model with the additional requirement that at least one electronegative atoms, with an atomic charge of -0.3 or greater, must be unshielded.

for "endocrine-disrupting potential" relies exclusively on existing chemical-specific human health and wildlife exposure and effects data. Consequently, when prioritizing chemicals for testing there is a tendency to focus upon those compounds for which data exist. One approach to help obviate this bias is to utilize 3-D SARs to expand the knowledge base for prioritization (Meridian Institute, 1999). In a related workshop cosponsored by the Society of Toxicology and Environmental Chemistry—Europe, the Organization for Economic Cooperation, and Development, and the European Commission, SARs for receptor binding and gene expression were also endorsed as promising approaches to enhance prioritization efforts (Ankley *et al.*, 1997). The exploratory COREPA 3-D SAR technique described here was developed, in part, to facilitate this need for rapid and mechanistically credible evaluations of large chemical data sets, including the Toxic Substances Control Act (TSCA) chemical inventory, which contains more than 75,000 chemicals. Specifically, the development of COREPA-based expert systems and complementary training data sets to predict

ligand-receptor binding affinity are intended to support ranking paradigms for the EPA Endocrine Disruptor Priority Setting Database (Meridian Institute, 1999), and similar international programs, where prioritization decisions for testing thousands of chemicals in commerce for endocrine disruption are needed.

ACKNOWLEDGMENTS

This research was supported, in part, by a cooperative agreement between Bourgas University "Prof. As. Zlatarov" and U.S. EPA (CR822306-01-0); and by an agreement between the European Union and Bourgas University "Prof. As. Zlatarov" (EU Project IC20-CT98-0114, EDAEP). The authors thank Dr. Stoyan Karabunarliev for fruitful discussions. Comments on an earlier draft of the manuscript were provided by Chris Russom and Mike Hornung. Diane Spehar and Roger LePage assisted in manuscript preparation.

REFERENCES

Ankley, G., Bradbury, S. P., Hermens, J., Mekenyan, O. G., Tollefsen, K.-E. (1997). Current approaches to the use of structure activity relationships (SARs) in identifying the hazards of endocrine-disrupting chemicals to wildlife. *Proceedings of the EMWAT Workshop, April 1997*, pp. 19–40. SETAC-Europe, The Netherlands.

Anstead, G. M., Wilson, S. R., and Katzenellenbogen, J. A. (1989). 2-Arylindenes and 2-arilindenones: Molecular structures and considerations in the binding orientation of unsymmetrical non-steroidal ligands to the estrogen receptor. *J. Med. Chem.* **32,** 2163–2171.

Bolger, R., Wiese, T. E., Ervin, K., Nestich, S., and Checovich, W. (1998). Rapid screening of environmental chemicals for estrogen receptor. *Environ. Health Perspect.* **106,** 551–557.

Bradbury, S. P., Mekenyan, O. G., Ankley, G. T. (1998). The role of ligand flexibility in predicting biological activity: Structure-activity relationships for aryl hydrocarbon, estrogen, and androgen receptor-binding affinity. *Environ. Toxicol. Chem*. **17,** 15–25.

Bradbury, S. P., Kamenska, V., Schmieder, P. K., Ankley, G. T., and Mekenyan, O. G. (2000). A computationally based identification algorithm for estrogen receptor-ligands. Part I. Predicting hERα Binding Affinity. *Toxicol. Sci.* **58,** 253–269.

Brooks, S. C., Wappler, N. L., Corombos, J. D., Doherty, L. M. (1987). Estrogen structure-receptor function relationships. In *Recent Advances in Steroid Hormone Action* (V. K. Moudgil, Ed.), pp. 443–466. Walter de Gruyter, Berlin.

Connor, K., Ramamoorthy, K., Moore, M., Mustain, M., Chen, I., Safe, S., Zacharewski, T., Gillesby, B., Joyeux, A., and Balaguer, P. (1997). Hydroxylated polychlorinated biphenyls as estrogens and antiestrogens: Structure-activity relationships. *Toxicol. Appl. Pharmacol.* **145,** 111–123.

Gabbard, R. B., and Segaloff, A. (1983). Structure-activity relationships of estrogens. Effects of 14-dehydrogenation and axial methyl groups at C-7, C-9, C-11. *Steroids* **41,** 791–805.

Ivanov, J. M., Karabunarliev, S. H., and Mekenyan, O. G. (1994). 3DGEN: A system for an exhaustive 3D molecular design. *J. Chem. Inf. Comput. Sci.* **34,** 234-243.

Ivanov, J. M., Mekenyan, O. G., Bradbury, S. P, and Schuurmann, G. (1998). A kinetic analysis of the conformational flexibility of steroids. *Quant. Struc. Act. Relat.* **17,** 437–449.

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.

Kuiper, G. G. J. M., Carlsson, B., Grandien, K., Enmark, E., Haggblad, J., Nilsson, S., and Gustafsson, J.-K. (1997). Comparison of the ligand binding specificity and transcript tissue distribution of estrogen receptor a and b. *Endocrinology* **138,** 853–870.

Mekenyan, O. G., Ivanov, J. M., Karabunarliev, S. H., Bradbury, S. P., Ankley, G. T., and Karcher, W. (1997). A computationally-based hazard identification algorithm that incorporates ligand flexibility. 1. Identification of potential androgen receptor ligands. *Environ. Sci. Technol.* **31,** 3702–3711.

Mekenyan, O. G., Nikolova, N., Karabunarliev, S. H., Bradbury, S. P., Ankley, G. T., Hansen, B. (1999). New developments in a hazard identification algorithm for hormone receptor ligands. *Quant. Struct. Act. Relat.* **18,** 139–153.

Meridian Institute. (1999). EPA priority-setting workshop for the endocrine disruptor screening program. January 20–21, 1999. A meeting summary. Dillon, CO.

Palomino, E., Heeg, M. J., Horwitz, J. P., Polin, L., Brooks, S. C. (1994). Skeletal conformations and receptor binding in some 9,11-modified estradiols. *J. Steroid Biochem. Mol. Biol.* **50,** 75–85.

Qian, X., and Abul-Hajj, Y. J. (1990). Synthesis and biological activities of 11β-substituted estradiol as potential antiestrogens. *Steroids* **55,** 238–241

Stewart, J. J. (1990). MOPAC: A semiempirical molecular orbital program. *J. Comput.Aid. Mol. Des.* **4,** 1–105.

Stewart, J. J. (1993). *MOPAC 93*. Fujitsu Limited, Chiba-city, Chiba 261, Japan, and Stewart Computational Chemistry. Colorado Springs, CO.

VanderKuur, J. A., Wiese, T., and Brooks, S. C. (1993). Influence of estrogen structure on nuclear binding and progesterone receptor induction by the receptor complex. *Biochemistry* **32,** 7002–7008.

Waller, C. L., Oprea, T. I., Chae, K., Park, H.-K., Korach, K. S., Laws, S. C., Wiese, T. E., Kelce, W. R., and Gray, L. E., Jr. (1996). Ligand-based identification of environmental estrogens. *Chem. Res. Toxicol.* **9,** 1240–1248.