*Pure Appl. Chem.*, Vol. 75, Nos. 11–12, pp. 2363–2373, 2003.
© 2003 IUPAC

# Workshop 1.1

# Mechanism-based QSAR approach to the study of the toxicity of endocrine active substances*

C. Dias Selassie‡, R. Garg, and S. Mekapati

*Chemistry Department, Pomona College, Claremont, CA 91711, USA*

*Abstract*: Mechanism-based QSAR models for interactions between various ligands and the estrogenic receptor are developed by using well-developed physicochemical parameters. Common features of these QSARs are identified, and deficiencies in some datasets are highlighted. The relative binding affinities of various substituted hexestrols to estrogen receptors are examined in terms of their steric, electronic, and hydrophobic attributes. Different QSARs for hexestrols and tamoxifens reveal that steric effects are of overriding importance in variations of binding affinity. In the few cases where a large number of diverse substituents are located on aromatic rings, electronic effects emerge and suggest that electron-donating groups enhance binding to the receptor while hydrophobicity plays a marginal role in decreasing binding affinity. With substituted phenols bearing alkyl-type substituents and substituted hydroxy-biphenyls, the binding is strictly dependent on hydrophobicity and size. These QSAR models are described in detail and examined together to illustrate the utility of lateral validation in mechanistic interpretation.

## INTRODUCTION

The realization that agents from the environment of both synthetic and natural origin can produce reversible or irreversible biological effects in individuals or populations by interfering with the synthesis, transport, distribution, and binding of endogenous hormones has led to extensive investigations of these endocrine active substances (EASs) [1]. Many EASs are estrogenic in nature (xenoestrogens) and exert agonist and antagonist effects via the estrogen receptor (ER), which functions as a ligand-activated transcriptional regulator [2]. Those that are synthetic include pesticides, food antioxidants, and metabolites of nonionic surfactants; and naturally occurring ones are plant secondary metabolites and mold metabolites [3]. Mechanistic toxicological studies of environmental pollutants have sparked a great deal of scientific interest in the chemico-biological interactions between the ER and various EASs [4].

The general mechanisms for receptor-mediated responses are numerous and include binding to the receptor at the cell surface, cytoplasm, or nucleus. The interaction with a specific DNA sequence (estrogen response elements) and other coregulatory proteins results in alterations in gene transcription and ultimately changes in regulation of cell proliferation and differentiation. Thus, the mechanism of action of EASs is complex and multipronged. It could involve inhibition of hormone synthesis, metabolism, and gene activation.

The ER, which is the target of a large number of EASs, is a member of the nuclear hormone receptor gene superfamily. It is characterized by modular structural organization with distinct domains

*Report from a SCOPE/IUPAC project: Implication of Endocrine Active Substances for Human and Wildlife (J. Miyamoto and J. Burger, editors). Other reports are published in this issue, *Pure Appl. Chem.* **75**, 1617–2615 (2003).

‡Corresponding author

that are associated with DNA binding, hormone binding, and *trans*-activation [2]. There are two ER isoforms; ERα and ERβ, which exhibit some similar, but mostly distinct, tissue distribution patterns and vary in their ligand-binding ability [5]. Compounds with distinct and diverse structures bind to ER with varying degrees of affinity and potency [5,6]. 17β-Estradiol acts as a pure agonist, while other compounds such as Faslodex (ICI182780) act as antagonists [7], and still others, termed selective ER modulators, demonstrate both agonist and antagonist activity. Raloxifene and tamoxifen fall into the latter category. The diversity in structure and physicochemical properties of these various ER-binding ligands warrants a thorough examination of their binding affinities in terms of their electronic, steric, and hydrophobic attributes. Utilization of the quantitative structure–activity relationship (QSAR) paradigm allows for delineation of the structural attributes that determines affinity in binding [8].

Our mechanism-based approach to the study of the interactions of various ligands in cellular systems and with isolated estrogen receptors utilizes well-developed physicochemical parameters to develop appropriate models, which take on added significance when subjected to the process of lateral validation. Thus, an individual QSAR is considered to be weak when on its own and significant when bolstered by other similar QSAR. The availability of a sophisticated database, which allows for lateral validation of a biological QSAR from a physical organic viewpoint, as well as added biological perspective, makes this approach useful and amenable to mechanistic interpretation [9]. Specific examples of binding interactions between different types of estrogen receptors (mouse, rat, lamb, calf, human) and their nonsteroidal ligands will now be described. Common features of these QSARs will be pinpointed, and inadequacies in some data sets will be identified.

## METHODOLOGY

All physicochemical descriptors are auto-loaded, and multiple regression analysis is used to derive the QSARs that are executed with the C-QSAR program [9]. While comparing different QSARs, however, it must be borne in mind that variations in quality of testing in different laboratories will have an effect that cannot be estimated.

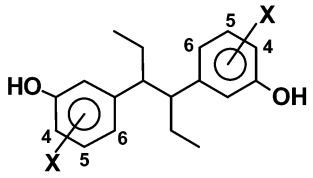
The parameters used in this article have been discussed in detail along with their applications [8]. In brief, Hammett σ, $\sigma^+$, and $\sigma^-$ constants are electronic parameters that apply to substituent effects on aromatic system. Normal σ values are utilized for substituents on aromatic systems where resonance between substituent and reaction center is minimal. $\sigma^+$ and $\sigma^-$ are employed where there is a strong resonance interaction between substituent and reaction center.

Clog *P* represents the calculated partition coefficient in octanol/water and is a measure of hydrophobicity of the whole molecules. π is the hydrophobic parameter for substituents attached to benzene. CMR is the calculated molar refractivity for the whole molecule. MR is calculated using the Lorentz–Lorenz equation and is described as follows: $[(n^2 - 1)/(n^2 + 2)](\mathrm{MW}/d)$, where $n$ is the refractive index, MW is the molecular weight, and $d$ is the density of a substance. Since there is a little variation in $n$, MR is largely a measure of volume with a small correction for polarizability. MR values are generally scaled by 0.1. MR can be used for a substituent or for the whole molecule. Clog *P* and CMR are calculated values for the neutral form of partially ionized molecules. B1, B5, and L are Verloop's STERIMOL parameters for substituents. B1 is a measure of the width of the first atom of a substituent, B5 is an attempt to define the overall volume, and L is the substituent length. All the biological data have been collected from the literature (see individual QSAR for respective references).

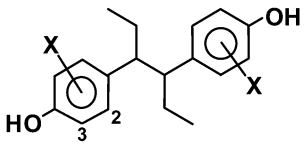
## RESULTS

### Metahexestrol and hexestrol derivatives

Metahexestrol and hexestrol derivatives were synthesized by Hartmann et al. [10,11], and their relative binding affinities (RBA) to estrogen receptors from calf uterine cytosol at 4 °C were determined. QSARs 1 and 2 were derived for di-X-metahexestrol **1** and di-X-hexestrol **2** derivatives, respectively.




**1**




**2**

*RBA of di-X-metahexestrols **1** to calf uterine cytosol ER at 4 °C/16 h*

$$\text{Log RBA} = -0.90(\pm0.32)\ \sigma_X^{+} - 1.26(\pm0.12)\ L_{4X} - 2.73(\pm0.33)\ B5_{5X} + 6.13(\pm0.61) \quad (1)$$

$n = 13$; $r^2 = 0.989$; $s = 0.141$; $q^2 = 0.969$; *outliers*: X = 4-$CH_2$OMe and 4-$NH_2$.

*RBA of di-X-hexestrols **2** to calf uterine cytosol ER at 4 °C/16 h*

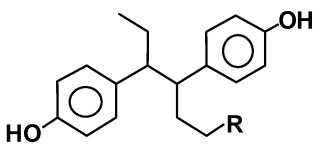
$$\text{Log RBA} = -0.65(\pm0.28)\ \sigma_X - 1.40(\pm0.27)\ MR_{3X} - 1.01(\pm0.54)\ MR_{2X} + 1.41(\pm0.26) \quad (2)$$

$n = 20$; $r^2 = 0.909$; $s = 0.281$; $q^2 = 0.864$; *outlier:* none

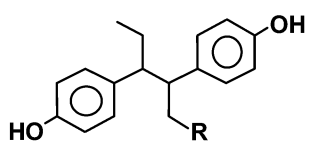
In these and following QSARs, $n$ is the number of compounds included in the analyses, $r$ is the correlation coefficient, $s$ is the standard deviation, and $q$ is the predictive (leave-one-out) correlation coefficient. The figures in parentheses are the 95 % confidence interval of regression coefficients and the intercept. The subscripts attached to parameter symbols represent substituents and/or their position. Substituents examined in the above two series are on the phenolic rings. In QSAR 1, X varies from H to such substituents as 4-F, -Cl, -Br, -Me, -Et, -$NO_2$, -$CH_2$OH, 5-OH, -Cl, 6-F, -Cl, and -Me; log RBA varies in a range of –2.00 (X = 5-OH) to 1.18 (6-Me). In QSAR 2, X is one of substituents such as H, 3-OH, -F, -Cl, -Br, -I, -Me, -Et, -$CH_2NMe_2$, -$CH_2OR$ (R = H, Me, Et), -$NO_2$, -$NH_2$, 2-OH, -F, -Cl, -Br, -Me, and Et; the log RBA variations are from –1.40 (X = 3-$CH_2NMe_2$) to 1.51 (X = 2-OH).

These QSARs contain only electronic and steric terms. The presence of Hammett constants indicates that electron-donating substituents enhance binding by either increasing electron density on the ring or decreasing the acidity of the phenolic group. Increased electron density on the phenolic ring would enhance dipolar interactions between the ring and electron-deficient amino acid residues in the binding domain. The presence of steric descriptors such as MR and Verloop's L and B5 terms and the negative dependence of RBA on them emphasize the constrained locale of the binding site and the deleterious effects of substituent bulk, particularly in the meta positions of both metahexestrols and hexestrols. Hydrophobicity is not significant in variations of these ligand–ER interactions.

Further modifications of the side chains of hexestrol to compound series **3** and **4** were undertaken by the Katzenellenbogen group [12,13].




**3**




**4**

*RBA of halogenated side chain hexestrols **3** to lamb uterine cytosol ER at 0 °C/16 h*

$$\text{Log RBA} = -0.56(\pm0.30)\ B1_R + 2.98(\pm0.52) \quad (3)$$

$n = 5$; $r^2 = 0.920$; $s = 0.090$; $q^2 = 0.754$

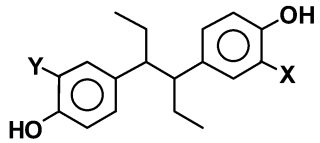
*RBA of functionalized side chain hexestrols **4** to lamb uterine cytosol ER at 0 °C/16 h*

$$\text{Log RBA} = 0.67(\pm0.25)\ \pi_R - 1.04(\pm0.24)\ MR_R + 2.75(\pm0.42) \quad (4)$$

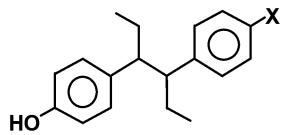
$n = 11$; $r^2 = 0.928$; $s = 0.274$; $q^2 = 0.845$; *outlier:* R = $CONH_2$

In QSAR 3, R varies among H, F, Cl, Br, and I; log RBA from 1.78 (R = I) to 2.48 (R = H). QSAR 4 includes R = $CH_2R'$ (R′ = H, F, Br, I, or OH) and COR″ (R″ = Me, *n*-Bu, OMe, O-*n*-Pent, $CHN_2$, or NH-*n*-Pent); log RBA varies from –0.82 (R = CONH-*n*-Pent) to 2.48 (R = Me). In both QSARs 3 and 4, a significant negative dependence on steric parameters indicates that the binding site has limited bulk tolerance. QSAR 4 pinpoints the importance of hydrophobic side chains in enhancing RBA.

Heiman et al. synthesized and evaluated the binding affinity of dihalogenated hexestrols **5** [14]. Utilization of this data resulted in the formulation of QSAR 5, while a similar study of 4-X-deoxyhexestrol derivatives **6** [15] led to the formulation of QSAR 6.




**5**




**6**

*RBA of aromatic ring-halogen hexestrols **5** to lamb uterine cytosol ER at 0 °C/16 h*

$$\text{Log RBA} = -1.25(\pm0.61)\ B1_Y - 0.71(\pm0.56)\ B1_X + 4.49(\pm0.99) \quad (5)$$

$n = 7$; $r^2 = 0.953$; $s = 0.218$; $q^2 = 0.842$

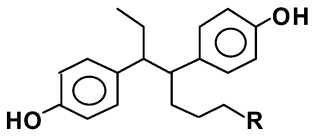
*RBA of 4-X-deoxyhexestrols **6** to immature rat uterine cytosol ER at 0°C/20~24 h*

$$\text{Log RBA} = -0.64(\pm0.17)\ MR_X - 1.24(\pm0.31)\ \sigma_X + 1.67(\pm0.\ 26) \quad (6)$$

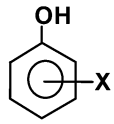
$n = 27$; $r^2 = 0.822$; $s = 0.286$; $q^2 = 0.760$; *outliers:* X = $SO_2N_3$, $NO_2$, $OCH_2COOH$, $OCH_2COCH_2Cl$, and $OCH_2COCH_2Br$

QSAR 5 includes compounds **5** with combinations of X (H, F, Br, or I) and Y (H, F, Br, or I). X and Y substituents are defined so that $B1_Y \geq B1_X$. The log RBA value varies between 0.42 (X = Y = Br) and 2.48 (X = Y = H). QSAR 6 is for compounds **6** in which X = H, OH, $OCH_2R$ [R = H, $CH_2OH$, $CH_2O(CH_2)_2Cl$, -$CHCH_2O$- (oxiranyl), or $CH{=}CH_2$], OCH2CH(OH)R′ (R′ = Me, CH2OH, CH2Cl, or CH2Br), OCH2COR″ (R″ = Me, CH=$N_2$, or OEt), $NH_2$, $N_3$, $SO_2F$, COOH, $CH_2R'''$ (R‴ = OH, Cl, Br, or OEt), COCH2R⁗ (R⁗ = H, Cl, or Br), CN, or C(=NH)OEt. The log RBA is in a range from –0.48 (X = $SO_2F$) to 2.48 (X = OH). The loss of a phenolic OH group in 4-X-deoxyhexestrol derivatives **6** clearly leads to a substantial decrease in binding affinity.

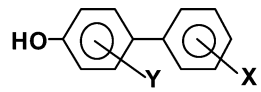
A similar study of binding of DES derivatives **7** to rat cytosolic ER [16] led to the formulation of QSAR 7, in which R = Me, COOPh, COOPh(4′-Cl), $CONEt_2$, or $CH_2X$ [X = $OCONH(CH_2)_2Cl$, OEt, OH, I, $OSO_2Me$, or Cl] and the log RBA varies from –0.52 ($COOC_6Cl_5$; $COOC_6H_4$-4′-Cl) to 1.0 (Me).




**7**




**8**




**9**

*RBA of DES derivatives **7** to rat uterine cytosol at 18°C/30 min*

$$\text{Log RBA} = -0.24(\pm 0.12)\ MR_R + 0.69(\pm 0.34) \quad (7)$$

$n = 10$; $r^2 = 0.725$; $s = 0.248$; $q^2 = 0.457$; *outliers:* R = COOH and $(CH_2)_4Me$

## Alkyl and phenyl phenols

Alkyl-phenolic compounds have also been implicated as weak ER ligands [14,17]. Using a recombinant yeast cell assay expressing human ER, the estrogenic potential of alkyl phenols **8** was determined in terms of the concentration (M) required for inducing half the maximal gene activation response produced by 17β-estradiol [17]. This data is now used to formulate QSAR 8. Substituents included are 2-*t*-Bu, -*s*-Bu, $-CHMe(CH_2)_7Me$, and $-CMe_2CH_2CMe_3$, 3-*t*-Bu, 4-Me, -Et, -*n*-Pr, -*t*-Bu, -*n*-Pent, -*n*-Hept, -*n*-Non, -CH(Me)R (R = Et, Hex, Oct, or Decyl), and $-C(Me)_2R'$ (R′ = Et, Pr, $CH_2$-*i*-Pr, or $CH_2$-*t*-Bu), and 2-Me-4-Non, the activity varies from 1.00 (4-Me) to 6.98 (4-$CMe_2CH_2$-*t*-Bu).

*Induction of half maximal gene activation by alkyl phenols **8***

$$\text{Log } 1/EC_{50} = 1.39(+/-\ 0.48)\text{Clog } P - 3.69(\pm 0.90)(\beta \cdot 10^{\text{Clog } P} + 1) + 1.05(\pm 0.24)\ MR_{4X} - 3.00(\pm 1.65) \quad (8)$$

$n = 21$; $r^2 = 0.927$; $s = 0.633$; $q^2 = 0.874$; Clog $P_o = 4.8$; *outliers*: X = 2,4-*t*-$Bu_2$, 2-*n*-Non-3-Me, and 2-*n*-Non-4-Me.

QSAR 8 shows a strong dependence of estrogenicity on hydrophobicity and size. The gene activation response increases as hydrophobicity increases up to a Clog *P* of 4.8 and then sharply decreases as hydrophobicity continues to increase. The extensive hydrophobic pocket may be constricted by some bulky amino acid residues at the binding site. The bulk of the substituent in the para position may help to anchor/lodge the substituent firmly within the confines of the hydrophobic pocket. The large number of diverse alkyl phenols in the set allows for a thorough evaluation of the role played by hydrophobicity.

In a similar study of the inhibition of ER binding of a smaller series of alkyl phenols **8** [18], the following QSAR 9 was developed. Despite the adequate range in hydrophobcity, a nonlinear model in Clog *P* could not be assessed because of the limited number of data points and lack of diversity in substituent type including 2-, 3-, 4-*t*-Bu, 4-*n*-Non, and 4-$C(Me_2)$R [R = *n*-Pent, and -Ph(4′-OH)]. The activity ranges from –2.60 (3-*t*-Bu) to 0.08 (4-*n*-Non). The positive dependence on bulk of substituents persists in this QSAR 9.

*Inhibition of mouse uterine ER binding by X-phenols **8***

$$\text{Log } 1/K_i = 0.63(\pm 0.26)\ MR_X - 2.67(\pm 0.75) \quad (9)$$

$n = 6$; $r^2 = 0.920$; $s = 0.405$; $q^2 = 0.858$; *outlier:* none

For the binding of a more diverse set of 4-hydroxybiphenyls **9** to mouse uterine cytosol ER [19], QSAR 10 indicates the bulk of the substituents in the phenyl ring (X) helps enhance binding to the ER, but those on the phenolic ring (Y) fail to do so.

*Half-maximal specific binding of 4-hydroxybiphenyls **9** relative to estradiol to mouse uterine ER*

$$\text{Log } 1/IC_{50} = 2.56(\pm0.82)\ MR_{2X} + 1.13(\pm0.95)\ MR_{4X} + 4.84(\pm0.46) \quad (10)$$

$n = 11$; $r^2 = 0.883$; $s = 0.294$; $q^2 = 0.778$; *outlier:* [X = 2,3,5,6-$Cl_4$-4-OH, Y = H].

In the hydroxybiphenyls **9**, X is either one of H, 4-OH, 4-Cl, 2,5-$Cl_2$, 2,6-$Cl_2$ 3,4,6-$Cl_3$, 2,3,4,5-$Cl_4$, 2-Cl-4-OH, and 3,5-$Cl_2$-4-OH, while Y is mostly H except for 3-Cl and 2,6-$Cl_2$. The activity range is from 5.00 (X = 4-OH, Y = H) to 7.38 (X = 2,4,6-$Cl_3$, Y = H).

### Tamoxifen derivatives

Tamoxifen derivatives exert their antiestrogenic action via an ER mediated process [20]. From a study by Gilbert et al. on the RBA of a novel series of basic diphenylethylenes **10**, the following QSAR 11 was developed. The combinations of substituents ($R_1$, $R_2$, and $R_3$) are from $R_1$ = *i*-Pr, *n*-Bu, *n*-Pent, *n*-Hex, *n*-Oct, and $CH_2Ph$, $R_2$ = OH, $O(CH_2)_2NMe_2$, $O(CH_2)_2NEt_2$, and $O(CH_2)_2N(CH_2)_4$, and $R_3$ = OH, $O(CH_2)_2NMe_2$, and $O(CH_2)_2NEt_2$; the activity range is between –1.70 [*n*-Hex, $O(CH_2)_2NEt_2$, and OH] and 0.00 [*i*-Pr, OH, and $O(CH_2)_2NEt_2$].

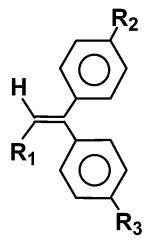
*RBA of diphenyl ethylenes **10** to lamb uterine cytosol ER at 25 °C/5 h*

$$\text{Log RBA} = 0.63(\pm0.26)I_Z - 0.64(\pm0.21)\ MR_{R1} + 0.36(\pm0.52) \quad (11)$$

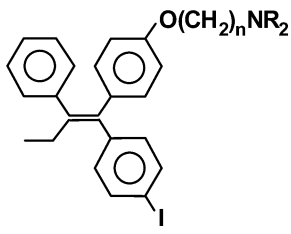
$n = 13$; $r^2 = 0.855$; $s = 0.204$; $q^2 = 0.744$; *outliers*: [$R_1 = CH_2Ph$, $R_2 = O(CH_2)_2N(Et)_2$, $R_3$ = OH]; and [$R_1$ = *n*-Oct, $R_2 = O(CH_2)_2N(Et)_2$, $R_3$ = OH].

Overall, the binding of these series of compounds to lamb uterine cytosol ER was very low as the intercept indicates. $I_Z$ is an indicator variable, which is unity when $R_3 = O(CH_2)_2NMe_2$ or $O(CH_2)_2NEt_2$, and $R_2$ = OH but zero when $R_3$ = OH and $R_2 = O(CH_2)_2NMe_2$, $O(CH_2)_2NEt_2$, or $O(CH_2)_2N(CH_2)_4$. An increase in the bulk of the $R_1$ group and "*E*-isomery" in the geometry are detrimental to binding to the ER.

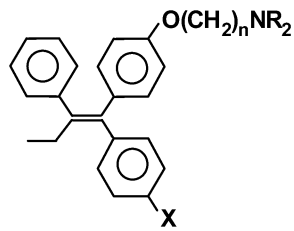
In an extensive study of tamoxifen analogs, Hardcastle et al. [21,22] focused on the following series **11** and **12** in which $NR_2 = NMe_2$ or $N(CH_2)_4$ (a pyrrolidine ring). With the iodine derivatives **11** ($n$ = 2~10), QSAR 12 was formulated. The log RBA varies from –0.40 ($n = 10$, $NR_2 = NMe_2$) to 1.40 ($n = 4$, $NR_2 = NMe_2$). Note that the indicator variable $I$ takes a value of 1 when $NR_2$ is $N(CH_2)_4$ and 0 when $NR_2$ is $NMe_2$.




10




11




12

*RBA of 4-iodo-tamoxifen analogs **11** to immature rat uterine cytosol at 4 °C/16 h*

$$\text{Log RBA} = -0.40(\pm0.10)\ \pi_4 + 0.36(\pm0.25)\ I + 1.18(\pm0.21) \quad (12)$$

$n = 16$, $r^2 = 0.860$, $s = 0.212$, $q^2 = 0.778$, *outlier:* none

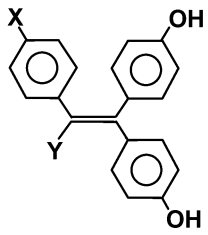
*RBA of tamoxifen analogs **12** to immature rat uterine cytosol ER at 4 °C /16h*

$$\text{Log RBA} = -0.48(\pm 0.14)\, L_{4X} + 3.18(\pm 0.63) \quad (13)$$

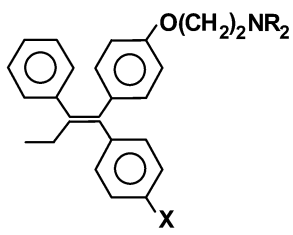
$n = 11$; $r^2 = 0.874$; $s = 0.183$; $q^2 = 0.827$; *outlier:* [X = H, $n = 2$, $NR_2 = NMe_2$]

In QSAR 12, the strong dependence on the hydrophilicity of the basic side chains is to be expected since there is a lack of variation in electronic character. The pyrrolidine ring on the side chain enhances RBA. For the 4-X-tamoxifen derivatives **12**, in which $n = 2 \sim 4$ and X = I, OH, *n*-Bu, or C≡CH with the activity varying from 0.30 ($n = 2$, $NR_2 = NMe_2$, X = *n*-Bu; $n = 2$, $NR_2 = N(CH_2)_4$, X = *n*-Bu) to 2.00 ($n = 2$, $NR_2 = NMe_2$, X = OH), a decrease in the length of the substituent leads to enhanced RBA. QSARs 12 and 13 suggest that substituents on the phenyl rings bind to restricted, polar space on the ER.

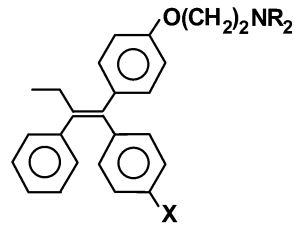
Ruenitz et al. examined the RBA of a small series of triarylethylenes **13** [23], in which X = H, F, or Br and Y = H or Cl, and log RBA ranges from 1.74 (X = Br, Y = H) to 2.25 (X = Y = H).




**13**




**14-*trans***




**14-*cis***

*RBA of triarylethylenes **13** to rat uterine ER at 4 °C/4 h*

$$\text{Log RBA} = -2.61(\pm 2.24)\sigma_X^{+} + 2.15(\pm 0.19) \quad (14)$$

$n = 4$; $r^2 = 0.926$; $s = 0.084$; $q^2 = 0.660$; *outlier*: [X = I ,Y = H]

The iodine derivative (X = I, Y = H) was not well predicted, and a size effect might be implicated. Unfortunately, the small number of analogs in the data set precludes further exploration with other parameters.

The analysis of a series of 4-substituted tamoxifen derivatives **14** (trans and cis) synthesized and evaluated by McCague et al. [24] led to the formulation of QSAR 15. The substituent X varies from H to Cl, Br, I, SH, $S(O)_nMe$ ($n = 0$~2), CHO, and $CH(O)CH_2$ (oxiranyl). The activity changes from –1.00 (*cis*-SMe) to 1.30 (*trans*-CHO). The data set includes both trans and cis geometric isomers. Trans and cis designate the relative positions of the ethyl group and the phenyl substituted with the basic side chain. Thus, an indicator variable $I_{cis}$ is assigned a value of 1 for *cis*-isomers.

*RBA of 4-X-tamoxifens **14** to calf uterine cytosol at 18 °C/30 min*

$$\text{Log RBA} = -1.20(\pm 0.33)\, I_{cis} + 0.93(\pm 0.68)\, \sigma_X + 0.25(\pm 0.26) \quad (15)$$

$n = 16$; $r^2 = 0.835$; $s = 0.303$; $q^2 = 0.779$; *outlier: trans*-$CH_2OH$

The positive Hammett ρ value (+0.93) indicates that electron-withdrawing groups on the aromatic ring enhances RBA. They would decrease the basicity of the basic side chain. The detrimental effects of cis configuration in binding are noted; thus the *trans*-isomers bind more effectively to the ER.

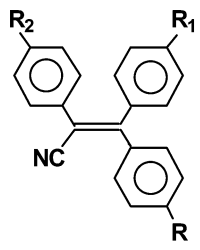
## DISCUSSION

Careful comparison of the various models developed for the binding of hexestrols to various mammalian ERs reveals a strong dependence on steric effects. Steric hindrance, whether represented by MR, which describes mostly volume, or Verloop's B1 (or B5 and L), which delineates width (length) of substituents, is of overriding importance in QSARs 1~7. Surprisingly variations in hydrophobicity do not play a significant role in these ligand–ER interactions except in QSAR 4 where modification is on the hexestrol side chain. A negative Hammett ρ value shows up when a diverse set of substituents is located on the aromatic rings. Merely extending an alkyl chain on an aromatic ring does not bring electronic diversity to a set of compounds. These features of hexestrol binding do suggest that although the ER may be overwhelmingly hydrophobic, the surface of the binding site is "craggy" and has lots of nooks and crannies, which make it "sterically challenged".

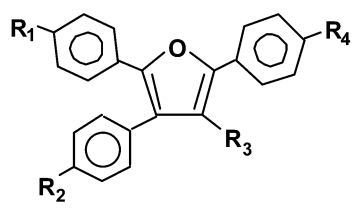
The negative correlation with bulk and size (as delineated by an indicator variable and Verloop's parameters) persists in QSARs 11, 13, and 15 of the binding of tamoxifen derivatives to the estrogen receptor. Electronic effects are only seen when the substituents on the phenyl ring vary in electronic character, otherwise hydrophilicity accounts for the variance in the data of a limited number of data sets. These results suggest that there must be patches of polar residues that are accessible by hydrophilic substituents on the aromatic rings of tamoxifen. QSAR 15 does not fit the normal pattern and reasons for its anomalous behavior are not apparent, although the incubation time is unusually short and thus equilibrium may have not been attained.

Phenolic compounds are rather interesting and their behavior is contradictory to what is seen with the substituted hexestrols and tamoxifens. Various phenols bearing alkyl type substituents have been assessed for their ability to interact with the receptor via a direct binding or yeast cell assay. In the two cases that were examined, the lack of diversity in substituent type precluded the inclusion of electronic parameters in QSARs 8 and 9. Thus, ER binding is strictly dependent on substantial hydrophobicity and minimal steric hindrance. A set of hydroxybiphenyls also behaves in a similar fashion as shown in QSAR 10. Phenols are smaller in size as compared to the hexestrols and tamoxifens and thus they have more flexibility. It has been assumed that the phenolic OH group generally superimposes itself on the OH group of the "A" ring of estradiol. But the strong dependence on hydrophobicity suggests that the binding mode of the X-phenols could be altered and the phenol could maneuver into the "D" ring region where the OH group could bind in the region of the 17-OH group of estradiol. The presence of an anchoring, polar amino acid residue could ground the phenols. The positive dependence on MR terms in QSARs 8, 9, and 10 suggests that a type of interaction toward an optimum could be occurring. Modeling studies would be invaluable in shedding light on this unusual behavior of substituted phenols binding to the ER.

Two comparative studies are worth mentioning—one, which examines different assay conditions and one which evaluates differential binding to α and β isoforms of the ER. Bignon et al. [25] examined the reactivity of a series of cyano-tamoxifen derivatives **15** under two experimental conditions from which QSARs 16 and 17 were formulated. Each of the substituents R, $R_1$, and $R_2$ in compounds **15** is either one of H, Me, OH, OMe, $OCHMe_2$, $O(CH_2)_2NEt_2$, and $NMe_2$.




**15**




**16**

*RBA of cyano-tamoxifen derivatives **15** to rat uterine cytosol ER at 25 °C/5 h*

$$\text{Log RBA} = 2.28(\pm 0.58)\, I_{OH} - 1.11(\pm 0.49)\, \pi_{4R} - 1.45(\pm 0.49) \quad (16)$$

$n = 23$; $r^2 = 0.861$; $s = 0.522$; $q^2 = 0.815$; *outlier:* [R = $R_1$ = $O(CH_2)_2NEt_2$, $R_2$ = H]

*RBA of cyano-tamoxifen derivatives **15** to rat uterine cytosol ER at 0 °C/2 h*

$$\text{Log RBA} = 2.08(\pm 0.47)\, I_{OH} - 0.46(\pm 0.39)\pi_{4R} - 0.94(\pm 0.38) \quad (17)$$

$n = 24$; $r^2 = 0.861$; $s = 0.437$; $q^2 = 0.805$; *outlier:* none

In QSAR 16, log RBA covers a range of –2.00 [R = $R_1$ = $OCHMe_2$, $R_2$= H and R = $R_1$ = H, $R_2$ = 4-$OCH_2CH_2N(Et)_2$] to 2.22 (R = $R_1$ = $R_2$ = OH). In QSAR 17, the lowest activity was also –2.00 (R = $R_1$ = $OCHMe_2$, $R_2$ = H), while the highest was 2.10 [R = OH, $R_1$ = $O(CH_2)_2NEt_2$, $R_2$ = H]. $I_{OH}$ is an indicator variable that denotes the positive impact of the presence of an OH moiety at any of R, $R_1$, and $R_2$ positions. A comparison of QSARs 16 and 17 notes strong similarities in the presence of the key descriptors: $I_{OH}$ and $\pi_4$. However, the magnitude of the sensitivity to hydrophilic effects varies considerably. It practically doubles at higher temperatures and longer reaction times. In both cases, RBA is enhanced by the presence of hydrophilic OH substituents, in particular, at the R position on the ring cis to the cyano group. The more negative $\pi_{4R}$ term in QSAR 16 than that in QSAR 17 would indicate that the extent of desolvation of the substituent at 25 °C exceeds that at 0 °C. This excellent study points out the danger of combining data from different assays which may have been run under different conditions.

Mortensen et al. compared the RBA of a series of triphenyl-furans **16** to isoforms of lamb uterine cytosol ER [26]. From their data, the following QSAR models 18 and 19 were formulated for the RBA to ERα and ERβ receptors, respectively.

*RBA of triphenylfurans **16** to ERα receptor of lamb uterine cytoxol at 0 °C/1~24 h*

$$\text{Log RBA} = -0.80(\pm 0.43)\, MR_{R3} + 1.15(\pm 0.46)\, I_{R2} + 1.75(\pm 0.51) \quad (18)$$

$n = 9$; $r^2 = 0.868$, $s = 0.227$; $q^2 = 0.729$; *outliers*: [$R_3$ = Me, $R_1$ = $R_2$ = $R_4$ = OH], [$R_3$ = Et, $R_1$ = $R_2$ = OH], [$R_3$ = Et, $R_1$ = $R_4$ = H, $R_2$ = OH]

*RBA of triphenylfurans **16** to ERβ receptor of lamb uterine cytoxol at 0 °C/18~24 h*

$$\text{Log RBA} = 2.46(\pm 0.99)\, MR_{R3} - 6.36(\pm 0.23) \log (\beta \cdot 10^{MR_{R3}} + 1) - 1.23(\pm 0.77) \quad (19)$$

$n = 10$; $r^2 = 0.899$; $s = 0.197$; $q^2 = 0.819$; MRO = 1.45, *outliers:* [$R_3$ = Pr, $R_1$ = $R_2$ = $R_4$ = OH], [$R_3$ = Et, $R_1$ = $R_4$ = H, $R_2$ = OH]

In the triphenyl-furans **16**, $R_1$, $R_2$, and $R_4$ = H or OH, and $R_3$ = Me, Et, Pr, Bu, or Ph(4-OH). The ERα binding activity ranges from 0.83 ($R_1$ = OH, $R_2$ = $R_4$ = H, $R_3$ = Et) to 2.15 ($R_1$ = $R_2$ = $R_4$ = H, $R_3$ = Et), while the ERβ binding activity ranges from –0.60 [$R_1$ = $R_2$ = $R_4$ = OH, $R_3$ = Ph(4-OH)] to 1.18 ($R_1$ = H, $R_2$ = $R_4$ = OH, $R_3$ = Pr).

QSAR 18 and 19 are significantly different, which suggests that the binding sites in the α and β receptors have marked differences in their topography. In QSAR 18, $I_{R2}$ is an indicator variable that acquires a value of 1 when $R_2$ = OH. Its presence suggests that a hydrogen bond formation could occur between the OH moiety and perhaps a polar residue at the binding site. The negative dependence on the bulk/polarizability parameter $MR_{R3}$ suggests that the binding site of $R_3$ substituents may be highly constrained and lined with residues of intermediate to marked polarity. The β receptor appears to be more extensive with less constraints and a moderate size cavity as implied by the fact that the optimum MR value of $R_3$ is 1.45. The lack of the $I_{R2}$ indicator variable in QSAR 19 suggests that the $R_2$ substituent on the phenyl ring at the 2-position may not be in close proximity to a H-bond acceptor or donor in the ERβ binding site.

## CONCLUDING REMARKS

The utility of the present type of "2D QSAR" in delineating receptor–ligand interactions in mostly congeneric series has been well documented while the process of lateral validation via comparative QSAR has helped bringing consistency to the various models [8]. Nevertheless, despite the proliferation of in vitro and in vivo assays to describe binding to steroid hormone receptors, the development of extensive 2D QSAR models has been limited.

The CoMFA (comparative molecular field analysis) is often viewed as an extension of QSAR and addressed as three-dimensional (3D). CoMFA has been utilized to develop several QSAR models for binding to various steroid hormone receptors [27]. However, despite the development of regression equations, the results are best mapped out and visualized as 3D color-coded contour plots. Semi-quantitative generalizations are drawn from these graphics but clear-cut quantitative aspects of the models are minimized or lost. Thus, rigorous and direct comparisons or extrapolations cannot be made between CoMFA models for different receptors, unlike "2D QSAR".

One of the attributes of CoMFA is its ability to combine diverse data sets and examine them as a whole. This operation can also be a weakness since it implicitly suggests that all the members of that set bind to a receptor at the "same" binding site and in the "same" mode. In the presence of definitive knowledge of the alignment and conformation of the individual molecules, it becomes necessary to examine numerous alignments and conformations and select the best combination that generates a CoMFA model with the highest predictive power (high $q^2$) [28]. With the 2D QSAR, the lateral validation of the individual QSARs allows for direct comparison of the descriptor terms, their contributions, and other statistical terms.

The action of endocrine-active substances is not limited to receptor binding and other non-receptor binding events should be considered. Our experimental results with simple and complex phenols (including eight estrogenic phenols) in a fast growing murine leukemia cell line (L1210), which acts as a model for fast growing cells (e.g., in embryogenesis), provide an underlying basis for our concern. The strong dependence of cytotoxicity of phenols on the OH bond-dissociation energy suggested that free-radical forming ability (coupled with weak hydrophobic contributions) was critical to inducing the cellular toxicity. The estrogenic phenols were included and were well predicted by this model [29]. The low coefficient with the log $P$ term suggests that interaction with a hydrophilic nucleotide sequence may be operative [29]. Thus, the ability of estrogenic phenols to interact with DNA at higher doses suggests that the effects of these endocrine-active substances could be diverse and of major concern.

Comparisons between our results and those previously published [4] on the binding of hexestrol and tamoxifen derivatives to the ER$\alpha$, establish a consistency in models, which emphasize the steric intolerance and moderate electron density requirements of the receptor. Hydrophobicity as represented either by Clog $P$ or $\pi$ is practically nonexistent and appears to not play a significant role in receptor binding. These results are also reflective of what emerged from an extensive analysis of estradiol derivatives [4]. The lack of hydrophobic terms in QSARs delineating ER–ligand interactions is perplexing. It may well be that the basic estradiol footprint (ABCD rings) present in estradiol, hexestrol, and tamoxifen derivatives occupy the sterically challenged, hydrophobic space only allowing limited mobility and subsequently variability to the substituents. Thus only small molecules like X-phenols will have options in binding site choices (either the A ring area or D ring environs).

Classical QSAR is invaluable in identifying structural features or properties that can affect binding to a receptor at the molecular level and thus impact biological activity. It can also pinpoint molecules that behave in an anomalous fashion. QSAR has great utility in mechanistic interpretation as well as predictability within a congeneric series, particularly with regard to binding affinities. QSAR models can be utilized effectively to predict ER-binding affinity within spanned space and thus have potential as priority setting tools for EASs.

## REFERENCES


1. J. A. Katzenellenbogen. *Environ. Health Perspect*. **103**, 99–101 (1995).
2. M. J. Tsai and B. W. O'Malley. *Annu. Rev. Biochem*. **63**, 451–486 (1994).
3. B. Gutendorf and J. Westendorf. *Toxicology* **166**, 79–89 (2001).
4. H. Gao, J. A. Katzenellenbogen, R. Garg, C. Hansch. *Chem. Rev.* **99**, 723–744 (1999).
5. G. G. J. M. Kuiper, B. Carlsson, J. Grandien, E. Enmark, J. Haggblad, S. Nilsson, J. A. Gustafsson. *Endocrinology* 863–870 (1997).
6. T. Barkhem, B. Carlsson, Y. Nilsson, E. Enmark, J. A. Gustafsson, S. Nilsson. *Mol. Pharmacol*. **54**, 105–112 (1998).
7. A. Howell, S. Downey, E. Anderson. *Eur. J. Cancer* **32A**, 516–588 (1996).
8. C. Hansch and A. Leo. *Exploring QSAR: Fundamentals and Applications in Chemistry and Biology*, American Chemical Society, Washington DC (1995).
9. C. Hansch, D. Hoekman, A. Leo, D. Weininger, C. D. Selassie. *Chem. Rev*. **102**, 783–812 (2002).
10. R. W. Hartmann, A. Heindl, H. Schonenberg. *J. Med. Chem*. **27**, 577–585 (1984).
11. R.W. Hartmann, W. Schwarz, H. Schonenberger. *J. Med. Chem.* **26**, 1137–1144 (1983).
12. R. Goswami, S. G. Harsy, D. F. Heiman, J. A. Katzenellenbogen. *J. Med. Chem*. **23**, 1002–1008 (1980).
13. S. W. Landvatter and J. A. Katzenellenbogen. *J. Med. Chem*. **25**, 1300–1307 (1982).
14. D. F. Heiman, S. G. Senderoff, J. A. Katzenellenbogen, R. J. Neeley. *J. Med. Chem.* **23**, 994–1002 (1980).
15. J. A. Katzenellenbogen, R. J. McGorrin, T. Tatee, R. J. Kempton, K. E. Carlson, D. H. Kinder. *J. Med. Chem*. **24**, 435–450 (1981).
16. K. Krohn, K. Kulikowski, G. Leclercq. *J. Med. Chem*. **32**, 1532–1538 (1989).
17. E. J. Routledge and J. P. Sumpter. *Environ. Toxicol. Chem*. **15**, 241–248 (1996).
18. C. L. Waller, T. I. Oprea, K. Chae, H. K. Park, K. S. Korach, S. C. Laws, T. E. Wiese, W. R. Kelce, L. E. Gray, Jr. *Chem. Res. Toxicol*. **9**, 1240–1248 (1996).
19. K. S. Korach, P. Sarver, K. Chae, J. A. McLachlan, J. D. McKinney. *Mol. Pharmacol*. **33**, 120–126 (1988).
20. J. Gilbert, M. Fuentes, T. Ojasoo, J.-C. Dore, M. Pons. *J. Med. Chem*. **40**, 1104–1111 (1997).
21. I. R. Hardcastle, M. G. Rowlands, R. M. Grimshaw, J. Houghton, M. Jarman, A. Sharff, S. Neidle. *J. Med. Chem*. **39**, 999–1004 (1996).
22. I. R. Hardcastle, M. G. Rowlands, J. Houghton, I. B. Parr, G. A. Potter, M. Jarman, K. J. Edwards, C. A. Laughton, J. O. Trent, S. Neidle. *J. Med. Chem*. **38**, 241–248 (1995).
23. P. C. Ruenitz, J. R. Bagley, N. T. Nanavati. *J. Med. Chem.* **31**, 1471–1475 (1988).
24. R. McCague, G. Leclercq, N. Legros, J. Goodman, G. M. Blackburn, M. Jarman, A. B. Foster. *J. Med. Chem*. **32**, 2527–2533 (1989).
25. E. Bignon, M. Pons, A. C. Crastes de Paulet, J. C. Dore, J. Gilbert, J. Abecassis, J. G. Miquel, T. Ojasoo, J. P. Raynaud. *J. Med. Chem*. **32**, 2902–2103 (1989).
26. D. S. Mortensen, A. L. Rodriguez, K. E. Carlson, J. Sun, B. S. Katzenellenbogen, J. A. Katzenellenbogen. *J. Med. Chem*. **44**, 3838–3848 (2001).
27. R. D. Cramer III, D. E. Patterson, J. D. Bunce. *J. Am. Chem. Soc*. **110**, 5959–5967 (1988).
28. R. A. Hughes, T. Hams, E. Altmann, D. McAllister, R. Vlahos, A. Robertson, M. Cushman, Z. Wang, A. G. Stewart. *Mol. Pharmacol*. **61**, 1053–1069 (2002).
29. C. D. Selassie, A. J. Shusterman, S. Kapur, R. P. Verma, L. Zhang, C. Hansch. *Perkin Trans. 2* 2729–2733 (1999).