0013-7227/97/$03.00/0
Endocrinology
Copyright © 1997 by The Endocrine Society
Vol. 138, No. 9
Printed in U.S.A.

QSAR Models for Binding of Estrogenic Compounds to Estrogen Receptor α and β Subtypes

WEIDA TONG, ROGER PERKINS, *R.O.W. Sciences, Jefferson, AR 72079*, LI XING, WILLIAM J. WELSH, *Department of Chemistry, University of Missouri-St. Louis, St. Louis, MO 63121*, and DANIEL M. SHEEHAN, *Division of Reproductive and Developmental Toxicology, National Center for Toxicological Research (NCTR), Jefferson, AR 72079.*

Abstract

We have developed Quantitative Structure-Activity Relationship (QSAR) models based on Comparative Molecular Field Analysis (CoMFA) for 31 estrogenic chemicals whose relative binding affinity (RBA) is available for both ER-α and ER-β. The models demonstrated a significant correlation (r^2>0.95) between the CoMFA-calculated steric/electrostatic fields and corresponding RBA data and a good predictive capability (q^2>0.6) based on cross-validation. The CoMFA models and contour plots obtained for ER-α and ER-β suggest a close similarity between the receptors in terms of mode of binding and provide a rational basis for ligand selectivity.

Introduction

Recently published information on the nucleotide and corresponding amino acid sequences of the estrogen receptor (ER-β) cDNA indicates a close homology with the previously described ER-α (1-3). Nevertheless, the expression patterns of ER-α and ER-β across tissues as well as their ligand binding patterns were found subsequently to exhibit notable differences (4). The potential significance of these findings, including the ligand selectivity, has also been explored (5). This latter issue is important in considering the activity of numerous endogenous estrogens as well as the xenoestrogens including phytoestrogens, environmental chemicals and estrogenic drugs.

For the last decade, drug discovery efforts have employed computer-based QSAR molecular models to describe ligand-receptor interactions and to predict those chemical structures that possess the most desired pharmacological characteristics (6). More recently, such models are finding use to predict the RBAs of xenoestrogens (7,8), some of which disrupt development and reproduction in wildlife (9). Here we report a three-dimensional QSAR based on CoMFA methodology (10) in order to identify and differentiate structural features of estrogens responsible for ligand binding to ER-α and ER- β.

Materials and Methods

Data sets for analysis

The RBA data used as the training set to construct the CoMFA models included both human ER-α and rat ER-β subtypes (4). The 31 chemicals (Table 1) comprised 19 steroids, several synthetic estrogens, several phytoestrogens, and two environmental estrogens. By encompassing both structural diversity and a wide range of RBAs, this training set was designed to yield robust models that can correlate the variances in RBA with chemical structure (as encoded in the steric/electrostatic CoMFA fields) and, moreover, identify differences between the ER-α and β subtypes better than less diverse data sets.

Received: 05/13/97

Molecular modeling

Chemical structures were generated from the Sybyl 6.3 fragment database (11) or obtained from the Cambridge Structural Database (CSD) (12). The global minimum-energy conformation of each molecule was computed in three steps: (i) the geometry of each molecule was optimized to its nearest local minimum-energy conformation to an energy gradient of 0.001 kcal/molÅ; (ii) these energy-minimized structures were then subjected to conformational analysis using a systematic search over all rotatable bonds at 10° increments; and finally (iii) the molecules were reminimized after the rotated torsion angles were set to the minimum-energy values identified in step (ii). All atomic partial charges were computed using the Gasteiger-Marsili method (13).

CoMFA alignment

CoMFA requires each molecule to be aligned to ensure maximal superposition of the steric and electrostatic fields of the training-set compounds. Most alignment rules employ a least-squares fitting of pharmacophoric elements between a designated template molecule (E_2 for this study) and the other molecules in the training set. For the steroidal compounds, these pharmacophoric elements were the centroids of the A-, B-, C- and D-rings (Figure 1). For the triphenylethylene anti-estrogens, the pharmacophoric elements were the centroids of the A- and C-rings and the C-1 of the B-ring which correspond to the centroids of the A- and D-rings and the C7 of the B-ring, respectively, of E_2 (8). Alignment of the triphenylene anti-estrogens in this way provided maximal

overlap of their rings and their substituents with the steroidal ICI 164,384 anti-estrogen. Similar considerations were applied for the alignment of the phytoestrogens and the xenoestrogens. For example, the pharmacophoric element for bisphenol A and methoxychlor were the centroids of the A- and B-rings and the vertex carbon atom of CPh_2.

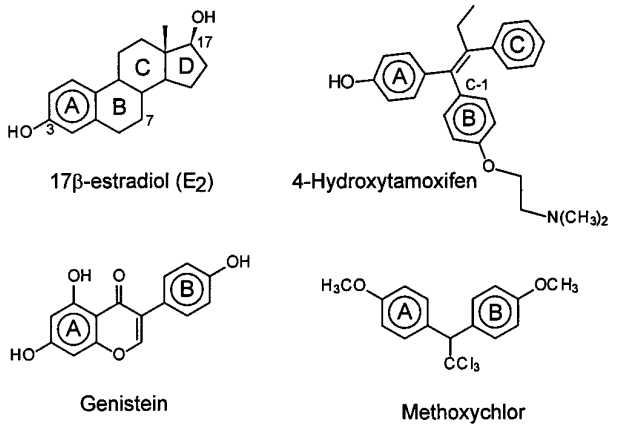

Figure 1: Structure of several chemicals in the training set

After alignment, the molecules were placed in a three-dimensional cubic lattice with 2 Å spacing. The steric (van der Waals) and electrostatic (Coulombic) fields were calculated for each molecule at each mesh point using an sp^3 carbon probe with +1.0 charge. Any calculated steric and electrostatic energies that were greater than 30 kcal/mol were truncated to this value.

PLS-QSAR

To form the basis for a predictive statistical model, the method of partial least squares (PLS) regression (14) was used to analyze the training set of 31 compounds by correlating variations in their biological activities with variations in their respective steric and electrostatic fields. The optimum number of principal components (PCs), corresponding to the smallest standard error of prediction, was determined by the leave-one-out (LOO) cross-validation procedure (15). In this method, each compound is systematically excluded once from the data set, after which its activity is predicted by a model derived from the remaining 30 compounds. Combining the 31 predictions allows the calculation of a cross-validated r^2 (termed q^2 hereafter). Using the optimal number of PCs, the final PLS analysis was carried out without cross-validation to generate a predictive QSAR model with a conventional correlation coefficient r^2. The r^2 and q^2 parameters are two key measures of significance of the CoMFA model. The r^2 value is a measure of the model's goodness of fit to the training-set activity data while the q^2 value is a measure of the model's predictive ability. A model for which $r^2 > 0.9$ and $q^2 > 0.5$ is generally considered to be both internally self-consistent and predictive (10).

Results

3D-QSAR/CoMFA models constructed separately for the ER-α and ER-β RBAs (Figure 2) yielded similar statistical results. Statistically significant values of q^2 (>0.6) and r^2 (>0.95) were obtained for both models, indicating good self-consistency and predictive ability. The contributions from the steric fields (40%) and electrostatic fields (60%) were identical in both models. Both models required only 4 PCs to optimally explain the variance in biological activity.

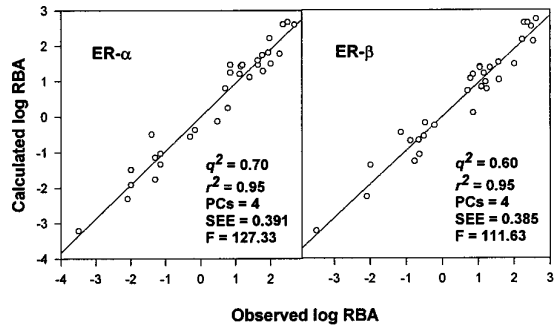

Figure 2: Plot of observed log RBA versus CoMFA calculated log RBA for ER-α and β

In order to increase the precision and predictive ability of the QSAR models, several factors were considered while developing these CoMFA models, such as reducing the lattice spacing and/or the value of the filtering energy. Reducing these two parameters did improve the q^2 value slightly but not enough to justify the extra computing time.

The effects of hydrophobic/hydrophilic character was also evaluated by adding values of the octanol/water partition coefficient (log*P*) to the CoMFA field descriptors for each compound in the training set. These log*P* values were calculated using the Ghose-Crippen method (16) available in Sybyl 6.3. The contribution of the log*P* descriptor to this CoMFA model was insignificant (3%) compared that of the steric and electrostatic fields (39% and 58%, respectively), indicating that the observed variations in the ER binding affinity among the training-set compounds are not explained by differences in their hydrophobic/hydrophilic character.

The original CoMFA models (i.e., 2 Å lattice spacing, 1.0 Kcal/mol filtering energy, no log*P* descriptors) were employed to derive color contour plots for both ER subtypes. To demonstrate the applicability of these CoMFA models to

compounds not included in the training set, the RBAs of test-set compounds were predicted for both receptors. The reasonable agreement between the predicted and observed RBAs (recorded as "less than" values) substantiates the fitness of the CoMFA model beyond the training set (Table 2).

Table 1: Observed and CoMFA-calculated log RBA values for binding to ER-α and ER-β.

Compounds	log RBA (α)		log RBA (β)	
	Obs.	Cal.	Obs.	Cal.
17β-Estradiol (E_2)	2.00	1.50	2.00	1.47
Estriol	1.15	1.43	1.32	1.37
Estrone	1.78	1.29	1.57	1.03
2-Hydroxy-E_2	0.85	1.47	1.04	1.40
4-Hydroxy- E_2	1.11	1.21	0.85	1.18
Moxestrol	1.63	1.59	0.70	0.72
ICI 164,384	1.93	1.82	2.22	2.16
17α-Estradiol	1.76	1.74	1.04	1.37
3α-Androstanediol	-1.15	-1.03	-0.52	-0.55
3β-Androstanediol	0.48	-0.12	0.85	0.10
4-Androstenediol	-0.30	-0.55	-0.22	-0.25
5-Androstenediol	0.78	0.25	1.23	0.77
Dehydroepiandrosterone	-1.40	-0.49	-1.15	-0.45
5α-Dihydrotestosterone	-1.30	-1.76	-0.77	-1.26
Nandrolone	-2.00	-1.49	-0.64	-1.07
Norethindrone	-1.15	-1.33	-2.00	-1.37
Norethynodrel	-0.16	-0.36	-0.66	-0.66
Testosterone	-2.10	-2.31	-2.10	-2.26
Progesterone	-3.50	-3.21	-3.50	-3.22
Diethylstilbestrol	2.67	2.60	2.47	2.53
Hexestrol	2.48	2.68	2.37	2.64
Dienestrol	2.35	2.61	2.61	2.75
Tamoxifen	0.85	1.25	0.78	1.07
4-Hydroxy-tamoxifen	2.25	1.78	2.53	2.12
Clomifene	1.40	1.12	1.08	0.83
Coumestrol	1.97	2.22	2.27	2.64
Genistein	0.70	0.81	1.56	1.52
β-Zearanol	1.20	1.47	1.15	1.21
Nafoxidine	1.64	1.47	1.20	0.97
Bisphenol A	-1.30	-1.14	-0.48	-0.18
Methoxychlor	-2.00	-1.91	-0.89	-0.68

Discussion

The CoMFA training set of 31 compounds covered at least seven structural classes, including mammalian, plant, synthetic and environmental estrogens and a complete estrogen antagonist. For both ER subtypes, CoMFA predicted RBA values are in good agreement with the corresponding experimental values. A distribution of residuals (i.e., differences between observed and calculated RBAs) was obtained ranging from 0.00 for norethynodrel (ER-β) to 0.75 for 3β-androstanediol (ER-β). The large residual for 3β-androstanediol compared with 3α-androstanediol is curious and will require further examination by incorporating more 3α- and 3β-substituted steroids. Consistent with previous QSAR and 3D-QSAR models for ER binding derived by us (8) and by other workers (7), the electrostatic field component made a slightly larger contribution than the steric field component to both CoMFA models. This finding might be related to observations that polar amino residues are well-represented in the hormone binding domain of ERs (1,17). An initial inspection of the CoMFA-calculated steric/electrostatic field contributions and corresponding regression models obtained for ER-α and ER-β would suggest a close similarity in the structural requirements for ligand binding. The ER-α and ER-β contour plots (Figure 2) are nearly identical on the left-hand side indicating that, for example, substitutions at the 3-position of the A-ring would affect binding to ER-α and ER-β about equally. Consistent with this deduction, the RBAs for 2-hydroxy-E_2 and 4-hydroxy-E_2 (Table 1) are proportionately the same for ER-α and ER-β compared to their respective RBAs for E_2. Also, both ERs are sensitive to adding steric bulk in the vicinity of the 17α-position on the steroid ring. However, closer examination of the respective CoMFA contour plots for ER-α and ER-β reveals subtle differences in their structural characteristics for binding that nevertheless may be biologically relevant. Of particular note, the green domain near the 17α position of E_2 is larger for ER-α than ER-β. This distinction suggests that increasing steric bulk in this region will enhance the binding affinity more for ER-α than for ER-β, consistent with the experimental findings for moxestrol and norethynodrel (Table 1).

Table 2: CoMFA-predicted log RBA values for the test set

Compounds	log RBA (α)		log RBA (β)	
	Obs.	Cal.	Obs.	Cal.
5α-androstanedione	<-2.0	-1.95	<-2.0	-2.11
5β-androstanedione	<-2.0	-2.10	<-2.0	-2.32
4-androstenedione	<-2.0	-2.22	<-2.0	-2.50
corticosterone	<-3.0	-2.41	<-3.0	-2.05

The present 3D-QSAR models will enable us to better quantify and visualize the unique specificity of different ER subtypes for various ligands. They will also further mechanistic studies of estrogen action and enable mass screenings of a large number of xenoestrogens for RBA values in order to prioritize them for further studies of possible estrogenicity.

Acknowledgment

The authors gratefully acknowledge the generous support of the FDA's Office of Women's Health.

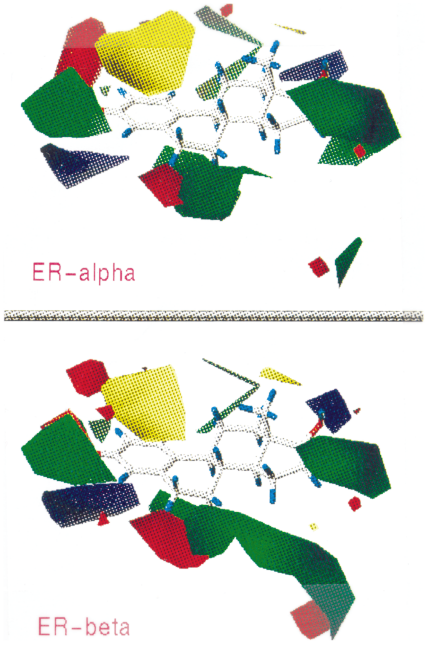

Figure 3 : CoMFA steric and electrostatic contour plots for estrogen binding to ER-α and ER-β (an E_2 molecule is placed inside the field for visualization purposes). Enhanced RBA (i.e., higher binding affinity) is associated with adding/subtracting steric bulk from green/yellow regions and with adding/subtracting positive charge in the blue/red regions. The absence of color represents regions that were unexamined by the current dataset.

References

1. Kuiper GGJM, Enmark E, Pelto-Huikko M, Nilsson S, Gustafsson J-Å 1996 Cloning of a novel estrogen receptor expressed in rat prostate and ovary. Proc Natl Acad Sci USA 93:5925-5930
2. Tremblay GB, Tremblay A, Copeland NG, Gilbert DJ, Jenkins NA, Labrie F, Giguere V 1997 Cloning, chromosomal localization, and functional analysis of the murine estrogen receptor β. Mol Endocrinology 11:353-365
3. Mosselman S, Polman J, Dijkema R 1996 ER-β Identification and characterization of a novel human estrogen receptor. FEBS Letters 392:49-53
4. Kuiper GGJM, Carlsson B, Grandien K, Enmark E, Haggblad J, Nilsson S, Gustafsson J-Å 1997 Comparison of the ligand binding specificity and transcript tissue distribution of estrogen receptors α and β. Endocrinology 138:863-870
5. Katzenellenbogen BS, Korach KS 1997 Editorial : a new actor in the estrogen receptor drama-enter ER-β. Endocrinology 138:861-862
6. Tong W, Collantes E, Chen Y, Welsh WJ 1995 A comparative molecular field analysis study of *N*-benzylpiperidines as acetylcholinesterase inhibitors. J Med Chem 39:380-387
7. Waller CL, Oprea TI, Chae K, Park HK, Korach KS, Laws SC, Wiese TE, Kelce WR, Gray LE 1996 Ligand-based identification of environmental estrogens. Chem Res Toxicol 9:1240-1248
8. Tong W, Perkins R, Strelitz R, Collantes ER, Keenan S, Welsh WJ, Branham WS, Sheehan DM 1997 Quantitative structure-activity relationships (QSARs) for estrogen binding to the estrogen receptor: prediction across species. Environ Health Perspect (submitted)
9. Kavlock RJ, Daston GP, DeRosa C, Fenner-Crisp P, Gray LE, Kaattari, Lucier G, Luster M, Mac MJ, Maczka C, Miller R, Moore J, Rolland R, Scott G, Sheehan DM, Sinks T, Tilson HA 1996 Research needs for the risk assessment of health and environmental effects of endocrine disruptors: A report of the USEPA-sponsored workshop. Environ Health Perspect 104:715-740
10. Cramer RD, Patterson DE, Bunce JD 1988 Comparative molecular field analysis (CoMFA). 1. Effect of shape on binding of steroids to carrier proteins. J Am Chem Soc 110:5959-5967
11. SYBYL Molecular Modeling System, Version 6.3; Tripos, Inc., St. Louis, MO 63144
12. Cambridge Structural Database, University Chemical Laboratory, Lensfield Rd., Cambridge CB2 1EW, England
13. Gasteiger J, Marsili M 1980 Iterative partial equalization of orbital electronegativity - a rapid access to atomic charges. Tetrahedron 36:3219-3228
14. Wold S, Albano C, Dunn WJ III, Edlund U, Esbensen K, Geladi P, Hellberg S, Johansson E, Lindberg W, Sjostrom M 1984 Multivariate data analysis in chemistry. In: Kowlaski B (Ed). In chemometrics: Mathematics and statistics in chemistry. Reidel, Dordrecht, The Netherlands
15. Cramer RD III, Bunce JD, Patterson DE 1988 Crossvalidation, bootstrapping, and partial least squares compared with multiple regression in conventional QSAR studies. Quant Struct-Act Relat 7:18-25
16. Ghose AK, Grippen GM 1986 Atomic physicochemical parameters for 3-dimensional structure-directed quantitative structure-activity-relationship 1. Partition-coefficients as a measure of hydrophobicity. J Comput Chem 7:565-577
17. Ekena K, Weis KE, Katzenellenbogen JA, Katzenellenbogen BS 1996 Identification of amino acids in the hormone binding domain of the human estrogen receptor important in estrogen binding. J Biol Chem 271:20053-20059