J Mol Model (2008) 14:1071–1082
DOI 10.1007/s00894-008-0338-4

ORIGINAL PAPER

Pharmacophore search for anti-fertility and estrogenic potencies of estrogen analogs

Sk. Mahasin Alam · Ria Pal · Shuchi Nagar · Md. Ataul Islam · Achintya Saha

Received: 21 January 2008 / Accepted: 11 June 2008 / Published online: 29 July 2008

Abstract Estrogen mediates its action following binding to the estrogen receptor to form an estrogen–receptor complex. The complex initiates gene transcription and produces estrogen-induced cell and/or tissue responses, i.e., estrogenic actions. High doses of estrogen can be used effectively as a contraceptive but are associated with side effects. Considering the long-term benefit-to-risk ratio of estrogen analogs as oral contraceptives, the present study was performed to deduce the active pharmacophore features required to differentiate the anti-fertility potency from the estrogenic activity of the steroidal motif. Implementing classical quantitative structure-activity relationship (QSAR) studies, substitution by an electron-donating group at the C_{17} position and the presence of a hydrogen bond acceptor at C_{11}, along with the orientation and conformational rigidity of the molecule, were found to be critically important features for estrogenic potency, including anti-fertility activity. However, low electron density at C_2 and high electronegativity at C_{16}, which may be due to substitution on those and/or neighboring atoms, favor contraceptive potency, whereas high electron density at C_5 and substitution by an electron-withdrawing group at C_7, which may confer hydrophobicity on the steroidal scaffold and an overall increment of electron affinity of the molecule, are favorable for estrogenicity. Further CATALYST-based 3D space modeling demonstrates that the presence of the aromatic ring (ring A), hydrophobic zone (ring B), and hydrogen bond acceptor at C_{17} in ring D, along with steric influence due to conformational rigidity of the compound, impart estrogenic contraceptive activity, but the presence of a second acceptor in ring A, and the critical distances between these features, selectively differentiate the anti-fertility potency from the estrogenic activity.

Keywords Estrogen · Anti-fertility and estrogenic potency · Quantitative structure-activity relationship · Pharmacophore space modeling

Introduction

Estrogen is the key regulator of the cellular processes involved in the development and maintenance of reproductive function [1]. It plays an important role in the reduction of bone resorption and increase of bone formation (and is used for the treatment and prevention of osteoporosis) [1, 2], retention of salt (sodium) and water [3, 4], and also has beneficial activities in the cardiovascular system that decreases the incidence of coronary heart disease [1, 3]. It also increases the level of HDL [4] and triglycerides [5], and decreases the level of LDL [6]. The estrogen receptor (ER)—a nuclear-activated transcriptional activator—mediates the effects of estrogen [7]. The estrogenic action of the ligand is thought to act predominantly by regulating gene expression after binding to the ER [7]. Ligand binding to the ER induces a conformational change in the receptor, which is important for the association of the receptor-DNA complex with transcriptional co-activators and the transcriptional components of the cell [8]. This association then culminates in the synthesis of estrogen-responsive genes and produces an estrogen-induced cell and/or tissue response [9]. The three

Electronic supplementary material The online version of this article (doi:10.1007/s00894-008-0338-4) contains supplementary material, which is available to authorized users.

S. M. Alam · R. Pal · S. Nagar · M. A. Islam · A. Saha (✉)
Department of Chemical Technology, University of Calcutta,
92, A.P.C. Road,
Kolkata 700 009, India
e-mail: achintya_saha@yahoo.com

major naturally occurring estrogens—estradiol (E2), estriol (E3) and estrone—are orally inactive. Synthetic estrogens are effective orally and used in replacement therapy, treatment of breast and prostate cancer, menstrual disorders, endometrial carcinoma, thromboembolic diseases, etc [3]. They are also effective as contraceptives at high doses but are associated with side effects including thrombic disorder, hypertension and certain types of malignancies. Other side effects include nausea, breast tenderness and fluid retention [3]. The mechanism of action of hormonal contraceptives is not clear. They probably act by preventing secretion of gonadotropins and thereby inhibiting ovulation, making cervical mucus thicker and thus making sperm penetration difficult. They may also make the endometrium hostile to the zygote [3].

Considering the long-term benefit-to-risk ratio of steroidal estrogen analogs as oral contraceptives, Peters et al [10–12] attempted to develop potential contraceptive agents with high anti-fertility activity accompanied by a reduction in other estrogenic activities. Structure-activity relationship (SAR) studies show that an alkyl or allyl substituent on the silicon side chain produces potent oral anti-fertility activity with reduced estrogenic activity [10]. Van der Waals volumes of estrogen analogs exhibit a useful correlation with biological activity [13]. Knowledge of the ER has allowed partial modeling of estrogenic activity. Different quantitative structure-activity relationships (QSAR) studies have been applied to predict the estrogenic activity of chemicals. Attempts range from the use of simple two-dimensional (2D) graph theoretical parameters to highly sophisticated three-dimensional (3D) approaches [13–18], e.g., structurally similar [19] and diverse [9, 20, 21] environmental estrogens have been classified through different molecular modeling techniques for exploring estrogenic activities. Although some work [22] has been done on the molecular modeling of contraceptive activity, no 3D pharmacophore hypothesis for anti-fertility potency has yet been explored. Consequently, the present work was initiated to study the steroidal scaffold as a small ligand [10–12], with a view to deducing the selectivity requirements in the active pharmacophore signal to differentiate anti-fertility potency from estrogenic activity, based on classical QSAR studies and the receptor-independent hypothesis. Due to the close structural similarity of this group of compounds to many prospective anti-fertility agents, it is possible that some pharmacodynamic similarity exists between these groups of compounds.

Materials and methods

The present work considered 53 silicon-substituted analogs of ethynyl estradiol [10], 31 {(triethyl) ethynyl}estradiol analogs [11] and 43 17-deoxy estrogen analogs [12] (Table 1), segregated into training (Tr, $n = 65$ for anti-fertility potency, $n = 97$ for estrogenic potency) and test (Ts, $n = 62$ for anti-fertility potency, $n = 30$ for estrogenic potency) sets for modeling aspects. The anti-fertility (A) and estrogenic (E) potencies of these compounds are considered as biological activity parameters (Table 1), and implemented as a logarithmic function, pA [$\log_{10}$ (1/A)] and pE [$\log_{10}$ (1/A)], respectively, for QSAR studies. The primary objective of this study was to generate relationships between the structure and corresponding activity using a multiple linear regression (MLR) approach, and to deduce the pharmacophore map using a receptor-independent space modeling technique. The2D structure of all compounds was first drawn in CS Chem Draw Ultra [23], then converted into a 3D structure using CS Chem 3D Pro [23]. Energy minimization of the compounds was performed in *MOPAC* module using the Austin Model 1 (AM1) method to locate global minima conformers. The energy-minimized structures were used to calculate different molecular properties, including physicochemical, electronic and spatial properties. The common atoms of this group of compounds have been numbered (Fig. 1) for computation of charge and electrotopological functions. Partial charge [24] was calculated using the Extended Huckel approach [23]. *E*-state indices [25] of the atoms were calculated using a JAVA-based program [26]. Some additional electronic (atomic charge functions, orbital energies) and physicochemical (ALogP, MR, etc) indices were generated using Tsar 3.3 [27]. The indicator variables used for modeling bioactivities were the presence of a methoxy group at position C_3 ($I_{3\text{-OMe}}$), a hydrogen bond acceptor at C_{11} ($I_{\_11\text{Hba}}$), and a substitutional requirement at the C_2 (I_{2S}), C_{15} (I_{15S}) and C_{16} (I_{16S}) positions. QSAR models were generated by MLR using standard and forward stepwise regression methods. The following statistical parameters [28, 29] were used to judge the statistical significance of the regression equation: correlation coefficient (R), explained variance (EV), variance ratio (F), standard error of estimate (s), and average of absolute values of calculated residuals (AVRES). The predictive power of the QSAR model was further judged by the leave-one-out (LOO) cross-validation method [30], which generates additional parameters such as predictive residual sum of squares (PRESS), standard deviation error of predictions (SDEP), average of absolute value of predicted residuals (Pres_{av}) and cross-validated variance (Q^2).

In addition, a group of compounds [10–12] (Tr, $n = 35$) (Table 2) containing the steroidal scaffold was selected for a receptor-independent pharmacophore space modeling study with regards to anti-fertility and estrogenic potencies. A 3D molecular space modeling study can deduce a pharmacophore hypothesis that can visualize the potential interaction between the ligand and receptor. A pharmacophore is a set

Table 1 Structural features and observed activities of steroidal analogs

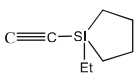
Comp	R	R_1	R_2	R_3	Observed Activity	
					Estrogenic	Anti-fertility
1	3-OH	H	-OH	-C≡C-Si(Me_2)-n-Bu	20.000	300.000
2[#]	3-OCH_3	H	-OH	-C≡C-$SiMe_3$	8.000	33.000
3	3-OH	H	-OH	-C≡C-$SiEt_3$	37.000	600.000
4[*]	3-OH	H	-OH	H	100.000	100.000
5	3-OH	H	-OH	-C≡C-Si(n-Pr)$_3$	25.000	200.000
6[#]	3-OH	H	-OH	-C≡C-Si(n-Bu)$_3$	2.500	-
7[#]	3-$OCOCH_3$	H	-OH	-C≡C-Si(n-Bu)$_3$	4.000	-
8[*,#]	3-OH	H	-OH	-C≡C-$SiCMe_3$	0.070	1.000
9	3-OH	H	-OH	-C≡C-Si (Me_2)$CH_2CH=CH_2$	24.000	250.000
10	3-OH	H	-OH	-C≡C-Si(Me_2)CH_2Br	27.000	200.00
11	3-OH	H	-OH	-C≡C-Si(Me_2)CH_2Ph	12.000	300.000
12[*,#]	3-OH	H	-OH	-C≡C-$SiPh_3$	0.500	-
13	3-OH	H	-OH	-C≡C-Si(Et_2)-n-C_6H_{13}	5.200	15.000
14	3-OH	H	-OH	-C≡C-Si(Et_2)-t-Bu	31.000	300.000
15	3-OH	H	-OH	-C≡C-Si(Et_2)-n-Pr	39.000	300.000
16[#]	3-OH	H	-OH		120.000	500.000
17	3-OH	H	-OH	-C≡C-Si(Me_2)-n-Pr	22.000	300.000
18	3-OH	H	-OH	-C≡C-Si(Et_2)-i-Pr	30.000	300.000
19	3-OH	H	-OH	-C≡C-Si(Me_2)-i-Pr	15.000	300.00
20[*]	3-OH	H	-OH	($SiMe_3$)C=CH_2	0.400	2.000
21[#]	3-OH	H	-OH	-C≡C-$SiMe_3$	7.000	33.000
22	3-OH	H	-OH	-C≡C-Si(Me)(Et)$CH=CH_2$	28.000	300.000
23	3-OH	H	-OH	-C≡C-Si(Et_2)OH	11.000	300.000
24[*,#]	3-OH	H	-OH	CH=CH$SiMe_3$(*E*)	0.200	-
25	3-OH	H	-OH	-C≡C-Si(Me)($CH_2Cl)_2$	39.000	150.000
26	3-OH	H	-OH	-C≡C-Si(Me_2)$CHCl_2$	18.000	300.000
27	3-OH	H	-OH	-C≡C-Si(Me_2)$CHClCH_3$	45.000	300.000
28	3-OH	H	-OH	-C≡C-Si(Me_2)$(CH_2)_2CH_2Cl$	25.000	200.000
29[*]	3-OH	H	-OH	-C≡C-Si Me_3(*Z*)	0.500	5.000
30	3-OH	H	-OH	-C≡C-Si(Me_2)$(CH_2)_2CF_3$	19.000	150.000
31	3-OH	H	-OH	-C≡C-Si(Me_2)CH_2Cl	17.000	150.00
32[#]	3-OH	H	-OH	-C≡C-Si(Me_2)Ph	7.000	33.000
33[*]	3-OH	H	-OH	($SiEt_3$)C=CH_2	0.500	4.000

Table 1 (continued)

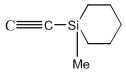
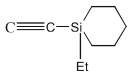
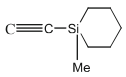
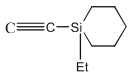
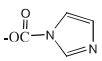
Comp	R	R_1	R_2	R_3	Observed Activity	
					Estrogenic	Anti-fertility
34	3-OH	H	-OH		25.000	250.000
35	3-OH	H	-OH		85.000	200.000
34	3-OH	H	-OH		25.000	250.000
35	3-OH	H	-OH		85.000	200.000
36	3-OH	H	-OH	-C≡C-Si(Me_2)-*t*-Bu	20.000	600.000
37[*,#]	3-OH	H	-OH	-CH=CHSiEt_3 (*E*)	0.040	1.300
38	3-OH	H	-OH	-C≡C-Si(Et_3)Me	40.000	600.000
39	3-OH	H	-OH	-C≡C-Si(Et_2)Ph	17.000	150.000
40	3-OH	H	-OH	-C≡C-Si(Et_2)H	29.000	600.000
41[*]	3-OH	H	-OH	CH=CHSiEt_3 (*Z*)	0.500	10.000
42	3-OH	H	-OH	-C≡C-Si(Et_2)CH_2CH=CH_2	44.000	300.000
43	3-OH	H	-OH	-C≡C-Si(Et_2)-*sec*-Bu	57.000	300.000
44	3-OH	H	-OH	-C≡C-Si(Me_2)-*sec*-Bu	20.000	300.000
45	3-OH	H	-OH	-C≡C-Si(Me_2)$(CH_2)_2$C≡N	37.000	300.000
46[*]	3-OH	H	-OH	[Si(n-Pr)$_3$]C=CH_2	0.500	2.000
47	3-OH	H	-OH	-C≡C-Si(Et_2)C≡CH	37.000	600.000
48	3-$OCOCH_3$	H	-OH	-C≡C-SiEt_3	34.000	600.000
49	3-$OCO(CH_2)_2CH_3$	H	-OH	-C≡C-SiEt_3	35.000	300.000
50[*,#]	3-OH	H	-OH	CH=CHSi(n-Pr)$_3$ (*E*)	0.050	-
51[#]	3-$OCO(CH_2)_2CH_3$	H	-$OCO(CH_2)_2CH_3$	-C≡C-SiEt_3	4.000	40.000
52[#]	3-OCH_3	H		-C≡C-SiEt_3	3.000	-
53		H	-OH	-C≡C-SiEt_3	24.000	300.000
54[*,#]	2-CH_2N-(n-Pr)$_2$	H	-OH	-C≡C-Si(Et)$_3$	0.300	-
55[#]	2-MeO	H	-OH	-C≡CH	0.300	-
56[#]	2-MeO	H	-OH	-C≡C-Si(Et)$_3$	0.060	-
57	2-Me	H	-OH	-C≡CH	5.500	40.000
58[*]	3-desoxy	H	-OH	-C≡CH	31.000	33.000
59	2-Me	H	-OH	-C≡C-Si(Et)$_3$	2.200	40.000
60[#]	2-CH_2N-(n-Pr)$_2$	H	-OH	-C≡CH	1.300	-
61[#]	4-CH_2CH=CH_2	H	-OH	-C≡CH	0.100	-
62[*,#]	4-CH_2CH=CH_2	H	-OH	-C≡CSi(Et)$_3$	-	-
63[*,#]	3-OH	6α-OH	-OH	-C≡CH	16.000	-

Table 1 (continued)

Comp	R	R_1	R_2	R_3	Observed Activity	
					Estrogenic	Anti-fertility
64	3-desoxy	H	-OH	-C≡CSi(Et)$_3$	13.000	100.000
65	3-OH	6β-OH	-OH	-C≡CSi(Et)$_3$	16.000	40.000
66[#]	3-OH	6β-Me	-OH	-C≡CH	1.800	-
67[*,#]	3-OH	6β-Me	-OH	-C≡CSi(Et)$_3$	0.200	7.500
68	3-OH	7α-OH	-OH	-C≡CH	0.100	66.000
69[#]	3-OH	7α-OH	-OH	-C≡CSi(Et)$_3$	0.300	20.000
70[#]	3-OH	6α, 7α-(OH)$_2$	-OH	-C≡CH	0.300	-
71[*]	3-OH	6α-OH	-OH	-C≡CSi(Et)$_3$	38.000	300.000
72[#]	3-OH	6α, 7α-(OH)$_2$	-OH	-C≡CSi(Et)$_3$	0.600	-
73[#]	3-OH	Δ^6	-OH	-C≡CH	2.000	20.000
74	3-OH	Δ^6	-OH	-C≡CSi(Et)3	2.000	200.000
75[*,#]	3-OCH$_3$	14β-H	-OH	-C≡CH	0.300	10.000
76[#]	3-OCH$_3$	15-OCH$_3$	-OH	-C≡CSi(Et)$_3$	3.000	-
77	3-OH	$\Delta^{6,8}$	-OH	-C≡CSi(Et)$_3$	1.000	66.000
78	3-OH	11β-OH	-OH	-C≡CH	0.800	200.000
79[*,#]	3-OH	16β-Me	-OH	-C≡CH	0.500	-
80	3-OH	11β-OH	-OH	-C≡CSi(Et)$_3$	14.400	64.000
81[#]	3-OCH$_3$	14β-OH	-OH	-C≡CSi(Et)$_3$	1.800	-
82	3-OCH$_3$	15-OCH$_3$	-OH	-C≡CH	0.800	10.000
83[*,#]	3-OH	16β-Me	-OH	-C≡CSi(Et)$_3$	0.200	-
84[#]	3-OH	$\Delta^{6,8}$	-OH	-C≡CH	0.600	10.000
85	3-OH	H	Me	H	0.300	20.000
86	3-OH	H	Et	H	1.000	66.000
87[*]	3-OH	11β-OCH$_3$	-OH	-C≡CH	-	1000.000
88	3-OH	H	Pr	H	0.800	20.000
89[#]	3-OH	H	Bu	H	0.400	-
90[#]	3-OH	H	CH_2OH	H	0.030	-
91[#]	3-OH	18-CH_3	Et	H	0.160	-
92[*,#]	3-OH	Δ^{15}	-C≡N	H	-	-
93[#]	3-OH	H	-C≡N	H	0.050	-
94[#]	3-OH	H	-OCOCH$_3$	H	0.010	-
95[#]	3-OH	H	-C≡CH	H	0.070	-
96[#]	3-OH	H	-CH=CH$_2$	H	0.260	10.000
97[*]	3-OH	H	-OH	H	27.000	10.000
98[#]	3-OH	H	-C≡CSi(Et)$_3$	H	0.040	-
99	3-OH	H	-CH_2CH_2Si(Et)$_3$	H	0.27	10.000
100[#]	3-OH	H	CF_2CH_3	H	0.090	-

Table 1 (continued)

Comp	R	R_1	R_2	R_3	Observed Activity	
					Estrogenic	Anti-fertility
101[#]	3-OH	H	-C(CH_3)=CH_2	H	0.08	-
102	3-OCH_3	H	-OH	Ph	0.690	10.000
103[*,#]	3-OCH_3	H	Ph	H	-	-
104[#]	3-OH	7α-OH	Et	H	0.040	-
105	3-OH	11β-OH	Et	H	0.030	20.000
106[#]	3-OH	11β-OCH_3	Et	H	37.000	-
107[*,#]	3-OH	11β-OH	17- =CH_2		0.020	10.000
108[#]	3-OH	11β-OH, 11α-CH_3	Et	H	0.050	-
109	3-OH	11- =CH_2	Et	H	64.000	400.000
110	3-OH	H	17- =O		12.000	10.000
111[*,#]	3-OCH_3	Δ^{15}	-OH	Ph	0.140	-
112	3-OH	H	17- =CH_2		0.500	40.000
113	3-OH	H	17- =$CHCH_3$		0.300	10.000
114[#]	3-OH	H	17- =CHC_2H_5		0.400	-
115[*,#]	3-OH	7α-Me	Et	H	15.000	-
116[#]	3-OH	H	17- =CHC3H7		0.200	-
117	3-OH	H	17- =CD2		0.650	67.000
118[#]	3-OH	H	17- =CHCHO		0.050	10.000
119[*,#]	3-OH	H	17- =NNH2		1.600	-
120[#]	3-OH	Δ^6	17- =CH2		0.040	-
121[#]	3-OH	H	17- =NOME		0.040	4.000
122	3-OH	H	17- =NOH		0.280	4.000
123[*,#]	3-OH	11-OXO	Et	H	-	-
124[#]	3-OH	H	17- =NNH(2-pyr)		0.100	-
125[#]	3-OH	H	17- =S		2.400	-
126[#]	3-OH	7α-OH	17- =CH2		0.100	-
127[#]	3-OH	11β-OH	17- =$CHCH_3$		0.040	-

*Estrogenic test compound; [#]Antifertility test compound

of functional groups/fragment types in a spatial arrangement that represents the interaction made in common by a set of small molecular ligands with a protein receptor [31]. The pharmacophore concept is based on the kinds of interaction observed in molecular reorganization, i.e., hydrogen bonding, charge, and hydrophobic interaction [32]. The receptor-independent pharmacophore hypothesis, generated by CATALYST [33], consists of analysis of features necessary for the bioactivity of the ligand arranged in 3D space that can explain the variance in activity of the molecules according to the geometric localization of their chemical features. In present work, the chemical features optimized for exploring the spatial pharmacophore map of this group of compounds are hydrogen bond (HB), acceptor-lipid (a) and donor (d), hydrophobic (p), and ring aromatic (r). To be retrieved as a hit, a candidate ligand

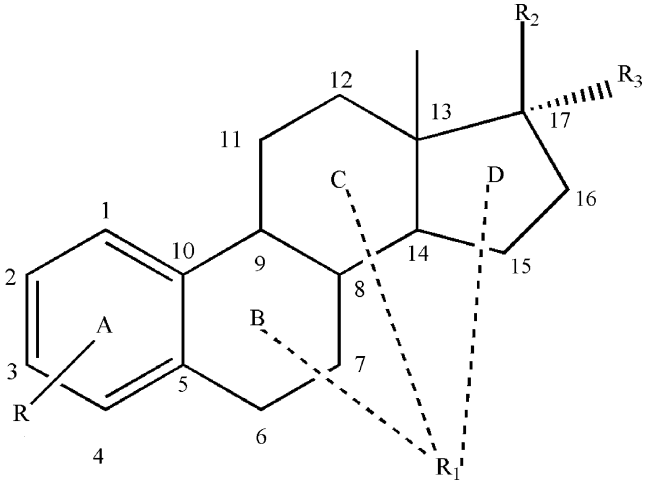


Fig. 1 General structure of estrogen analog

must possess appropriate functional groups that can simultaneously reside within the respective tolerance sphere of the pharmacophoric features. Each feature is associated with a weight (a measure of its suggested importance to the pharmacophore as a whole), and the better the overall superimposition of functional groups of the molecule to the appropriate features of the pharmacophore, the higher the score of the fit [34].

The different control parameters employed for hypothesis generation (called a Hypogen process) are spacing, uncertainty, and weight variation. Spacing is a parameter representing the minimum interfeature distance that may be allowed in the resulting hypotheses. In the present work, this was varied from 300 to 50. In the generated hypothesis, each feature signifies some degree of magnitude of the compound's activity. The level to which this magnitude is explored by the hypothesis generator is controlled by the weight variation parameter. This is varied in some cases from 1 to 2. In other cases, the default value of 0.3 is generally considered. The uncertainty parameter reflects the error of prediction, and denotes the standard deviation of a prediction error factor called the error cost. In the present work, values of 2 and 3 are considered as the uncertainty parameter. While generating the hypothesis, a total cost function comprising of three terms, viz., weight cost, error cost, and configuration cost, is minimized. Weight cost is a value that increases as the weight variation in the model deviates from the input weight variation value. The deviation between the estimated activity of the molecules in the training set and their experimentally determined value is the error cost. A fixed cost (ideal hypothesis cost) depends on the complexity of the hypothesis space being optimized and is also denoted as the configuration cost. The configuration cost is equal to the entropy of the hypothesis space. The CATALYST program [33] also calculates the cost of a null hypothesis that assumes no relationship in the data, and assumes that the experimental activities are normally distributed about their mean. Accordingly, the greater the difference (Δcost) between the total and the null costs, the more likely that the hypothesis does not reflect a chance correlation. The minimum difference between the total and null costs is taken as 60 bits for a hypothesis optimization. The generated hypothesis is further validated to nullify over prediction of bioactivities for inactive compounds through a process known as Hyporefine [33]. In this process, the steric interactions of the compounds in the hypothesis generation are considered, and if steric properties are crucial for bioactivity, then these are portrayed in the validated (Hyporefine) hypothesis. The quality of the generated hypothesis is judged through a cross-validation technique using CatScramble [33]. This validation procedure is based on Fischer's randomization test [29], where the biological activity data are randomized within a fixed chemical data set, and the Hypogen process is initiated to explore possibilities of other hypotheses of good predictive value. Logically, the

Table 2 Training and test subsets categorized for exploring pharmacophore features of steroid analogs. *Tr* Training set, *Ts* Test set

Activity	Set	Least-active subset (compound)	Intermediate subset (compound)	Most active subset (compound)	Number of compounds
Estrogenic	Tr	29, 46, 51, 57, 69, 70, 76, 85, 88, 96, 99, 113, 122	5, 9, 10, 22, 23, 30, 31, 34, 36, 39, 40, 64, 80	4, 15, 16, 25, 27, 35, 43, 45, 48	35
	Ts	2, 6, 7, 8, 12, 13, 20, 21, 24, 32, 33, 37, 41, 50, 52, 54, 55, 56, 59, 60, 61, 66, 67, 68, 72, 73, 74, 75, 77, 78, 79, 81, 82, 83, 84, 86, 89, 90, 91, 93, 94, 95, 98, 100, 101, 102, 104, 105, 107, 108, 111, 112, 114, 116, 117, 118, 119, 120, 121, 124, 125, 126, 127	1, 11, 17, 19, 26, 28, 42, 44, 53, 63, 65, 97, 110, 115	3, 14, 18, 38, 42, 47, 49, 58, 71, 106, 109	87
Anti-fertility	Tr	29, 46, 96, 99, 113, 122	4, 5, 9, 10, 25, 28, 30, 31, 34, 35, 39, 51, 57, 64, 69, 80, 85, 88	15, 16, 22, 23, 27, 36, 40, 43, 45, 48, 87	35
	Ts	8, 33, 37, 41, 67, 75, 82, 84, 97, 102, 107, 110, 118, 121	2, 13, 21, 32, 58, 59, 65, 68, 73, 74, 77, 86, 105, 112, 117	1, 3, 11, 14, 17, 18, 19, 26, 38, 42, 44, 47, 49, 53, 71, 78, 109	46

Table 3 Statistical quality of best quantitative structure-activity relationship (QSAR) models developed for anti-fertility and estrogenic potencies. *R* Correlation coefficient, *EV* explained variance, *F* variance ratio, *s* standard error of estimate, *AVRES* average of absolute values of calculated residuals, *PRESS* predictive residual sum of squares, *SDEP*standard deviation error of predictions, $Pres_{av}$ average of absolute value of predicted residuals, Q^2 cross-validated variance

Equation	*n*	Correlation statistics						Prediction statistics				
		R	R^2	EV (%)	*F*	*df*	*s*	AVRES	PRESS	SDEP	$Pres_{av}$	Q^2
1	97	0.890	0.792	77.85	57.234	6,90	0.508	0.409	28.536	0.542	0.448	0.745
2	65	0.946	0.894	88.53	99.769	5,59	0.235	0.179	4.175	0.253	0.202	0.865

hypothesis generated prior to scrambling should better attest to a good pharmacophore model. The explored pharmacophore models are further judged by estimating the fit score and activity of the test set.

In the present work, the number of conformers, generated by a simulated annealing technique [35], of each compound was limited to a maximum of 250 using the 'best conformer generation' method, with an energy cut-off of 20 kcal/mol. For hypothesis generation, biological activities are expressed as BA (1/A) for anti-fertility potency, and BE (1/E) for estrogenic potency. The Hypogen algorithm is forced to find pharmacophore models that contain at least one and at most two of all the input features.

Results

QSAR modeling

Structure–activity relationships were drawn to differentiate anti-fertility (Tr, n = 65) activity from the estrogenic (Tr, n = 97) potency of steroidal compounds [10–12], investigating physico-chemical (partition coefficient, hydrophobicity, steric and moments of inertia), electronic (atomic and partial charge functions, orbital energies) and electrotopological (E-state indices) features of molecular architecture in order to characterize unique pharmacophore features required for selective contraceptive activity. In all the regressional models, the 99% confidence intervals are shown in parentheses, and the F-values are also significant at the 99% confidence level.

From regressional analysis of Tr, the best univariate relationship for estrogenic potency was developed with the Extended Huckel partial charge function of atom C_{17} (C_{17}), which explains 39.53% variance in activity. The statistical quality of the relation was estimated to be

$$n = 97, R = 0.634, R^2 = 0.402, s = 0.839$$

and, in the case of bivariate relationships, the best significant relationship was explored with the same charge function of atom C_{17} (C_{17}) with an indicator (I_{11Hba}) that signifies the presence of a hydrogen bond acceptor at C_{11}, and explains 51.29% variance in activity. The quality of the model is estimated to be

$$n = 97, R = 0.723, R^2 = 0.523, s = 0.753$$

But, the best significant relationship for estrogenic potency was deduced to be

$$\begin{aligned} \mathrm{pE} = {} & 2.794(\pm 0.279) - 30.553(\pm 4.722)C_5 \\ & + 4.2678(\pm 0.809)C_7 - 6.546(\pm 0.560)C_{17} \\ & - 0.052(\pm 0.018)PMI_{\mathrm{Y}} - 0.303(\pm 0.076)E_{\mathrm{LUMO}} \\ & - 2.584(\pm 0.375)I\_{11\mathrm{Hba}} \end{aligned} \quad (1)$$

Where C_5 and C_7 are the same partial charge functions at atoms C_5 and C_7, respectively. PMI_Y is the principle moment of inertia at the y-axis and E_{LUMO} is the energy of the lowest unoccupied molecular orbital.

On the contrary, the best univariate model for anti-fertility potency was obtained with the charge function of atom C_{16} (C_{16}), which explains 65.68% variance in the activity; the statistical quality of the model is estimated to be

$$n = 65, R = 0.814, R^2 = 0.662, s = 0.407$$

and the best bivariate relationship was developed with the same charge functions of atoms C_{16} and C_{17}, and explains

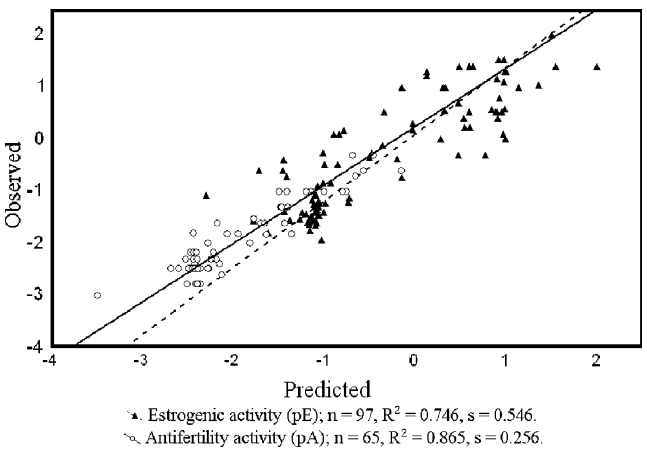


Fig. 2 Observed and predicted activities from quantitative structure-activity relationship (QSAR) models (Eqs. 1, 2) of estrogenic and antifertility potencies, respectively

Table 4 Hypotheses parameters observed in successive runs for estrogenic and antifertility potencies of steroidal analogs. *rmsd* Root mean square deviation, *WV* weight variation

Activity	Run No.	Method	Run parameters				Hypothesis no.	Output features[b]	Cost			R^d	rmsd
			Spacing (pm)	Uncertainty	WV	Input features[b]			Configuration	Null	Δ^c		
Estrogenic (n=35)	1	Hypogen	300	3	0.3	a, d, p, r	1	a x 2, p	14.1693	178.588	35.687	0.916	0.747
	2	Hypogen	250	3	0.3	a, d, p, r	1	a, p, r	15.6341	178.588	39.670	0.965	0.488
	3	Hypogen	200	3	0.3	a, d, p, r	1	a, p, r	15.8488	178.588	39.128	0.961	0.513
	4	Hypogen	150	3	0.3	a, d, p, r	1	a, p, r	15.8538	178.588	39.300	0.963	0.504
	5	Hypogen	100	3	0.3	a, d, p, r	1	a, p, r	15.8538	178.588	39.300	0.923	0.504
	6	Hypogen	50	3	0.3	a, d, p, r	1	a, p, r	15.8988	178.588	38.967	0.961	0.518
	7	Hypogen	300	2	0.3	a, d, p, r	1	a, p, r	14.1693	254.497	125.805	0.965	0.778
	8	Hypogen	250	2	0.3	a, d, p, r	1	a, p, r	15.3641	254.497	125.197	0.966	0.766
	9	Hypogen	200	2	0.3	a, d, p, r	1	a, p, r	15.8488	254.497	125.683	0.968	0.746
	10	Hypogen	150	2	0.3	a, d, p, r	1	a, p, r	15.8538	254.497	125.478	0.967	0.748
	11	Hypogen	100	2	0.3	a, d, p, r	1	a, p, r	15.8538	254.497	125.478	0.967	0.748
	12	Hypogen	50	2	0.3	a, d, p, r	1	a, p, r	15.8988	254.497	125.024	0.966	0.768
	13	Hypogen	200	2	1.0	a, d, p, r	1	a, p, r	15.8488	254.497	124.513	0.965	0.779
	14	Hypogen	200	2	2.0	a, d, p, r	1	a, p, r	15.8488	254.497	125.285	0.968	0.742
	15	Hyporefine of run 9	200	3	0.3	a, d, p, r	1	a, p, r, e	15.8488	254.497	126.312	0.969	0.731
Anti-fertility (n=35)	16	Hypogen	300	3	0.3	a, d, p, r	1	a x 2, p x 2	13.8929	154.637	15.151	0.907	0.613
	17	Hypogen	250	3	0.3	a, d, p, r	1	a x 2, p, r	14.8149	154.637	16.168	0.933	0.523
	18	Hypogen	200	3	0.3	a, d, p, r	1	a, d, p, r	15.0917	154.637	14.024	0.912	0.599
	19	Hypogen	150	3	0.3	a, d, p, r	1	a, p, r	15.0937	154.637	13.941	0.904	0.621
	20	Hypogen	100	3	0.3	a, d, p, r	1	a, p, r	15.0937	154.637	13.941	0.904	0.621
	21	Hypogen	50	3	0.3	a, d, p, r	1	a x 2, p, r	15.1372	154.637	14.247	0.915	0.587
	22	Hypogen	300	2	0.3	a, d, p, r	1	a, d, p x 2	13.8929	194.33	63.407	0.920	0.902
	23	Hypogen	250	2	0.3	a, d, p, r	1	a x 2, p, r	14.8149	194.33	65.170	0.935	0.814
	24	Hypogen	200	2	0.3	a, d, p, r	1	a x 2, p, r	15.0917	194.33	63.791	0.929	0.852
	25	Hypogen	150	2	0.3	a, d, p, r	1	a, p, r	15.0937	194.33	61.858	0.918	0.914
	26	Hypogen	100	2	0.3	a, d, p, r	1	a, p, r	15.0937	194.33	61.858	0.918	0.914
	27	Hypogen	50	2	0.3	a, d, p, r	1	a x 2, p, r	15.1372	194.33	64.882	0.936	0.812
	28	Hypogen	250	2	1.0	a, d, p, r	1	a x 2, p, r	14.8149	194.33	64.844	0.937	0.804
	29	Hypogen	250	2	2.0	a, d, p, r	1	a, p, r	14.8149	194.33	60.404	0.913	0.939
	30	Hyporefine of Run no.27	250	2	1	a, d, p, r	1	a x 2, p, r, e	14.8149	194.33	65.599	0.941	0.778

[b] *a* HB acceptor-lipid, *d* HB donor, *p* hydrophobic, *r* ring aromatic
[c] Δcost = null cost − total cost
[d] Correlation coefficient

74.45% variance in activity. The statistical quality of the model is estimated to be

$$n = 65, R = 0.867, R^2 = 0.752, s = 0.351$$

But the best model for anti-fertility potency can explain 88.53% variance in activity with good predictive property, with the relation:

$$\begin{aligned} \mathrm{pA} = & -3.003(\pm 0.230) + 6.367(\pm 1.152)C_2 \\ & -38.789(\pm 2.315)C_{16} - 2.299(\pm 0.365)C_{17} \\ & -0.043(\pm 0.012)PMI_{\mathrm{Y}} - 0.971(\pm 0.177)I_{\_11\mathrm{Hba}} \end{aligned} \quad (2)$$

Where C_2 is the extended Huckel partial charge on atom C_2.

The statistical parameters of the best relations (Eqs. 1, 2) developed are listed in Table 3, and independent variables used in the models are not intercorrelated ($R < 0.50$). The predicted (LOO cross-validation method) [30] activity from the models (Eqs. 1, 2) of Tr compounds are presented in Fig. 2.

Pharmacophore space modeling

The results of the 3D space-modeling study are presented in Table 4. Hypotheses 1 of runs 15 and 30 were considered to be the best hypotheses for estrogenic and anti-fertility

potencies, respectively. Selection of the models was characterized based on the highest cost difference (Δcost), the lowest root mean square deviation (rmsd) and highest correlation coefficient (R), and are shown in Table 4. The mapped pharmacophore features for both activities are illustrated in Fig. 3. The quality of hypotheses generated for estrogenic activity (hypothesis 1, run 15) and anti-fertility potency (hypothesis 1, run 30) were judged by a cross-validation technique using Fischer's randomization test [29] at the 99% confidence level, but no hypothesis generated better parameters than the original hypothesis in either case. Thus, the cross-validation analyses clearly indicated the superiority of the hypotheses (hypothesis 1, runs 15 and 30) considered for estrogenic and anti-fertility potencies. The observed vs estimated activities of the compounds for estrogenic and antifertility potencies are presented in Fig. 4, while estimated activities vs fit scores are depicted in Fig. 5. The results show that the presence of an HB-acceptor-lipid (a) at C_{17} in ring D, a hydrophobic (p) in ring B, and aromatic ring A (r) features, along with steric influence (e) are important for both estrogenic and anti-fertility potencies, exhibiting correlations of 96.8% and 93.7%, respectively. However, the presence of a second HB acceptor-lipid (a) in ring A is essential for selective anti-fertility activity. The distances (Fig. 3) between a1 and r are 8.405 and 7.987; between a1 and p are 6.993 and 6.337; and between r and p are 2.854 and 2.150 for estrogenic and anti-fertility potencies, respectively. The presence of a second acceptor (a2), with distances from a1, r and p of 10.611, 2.941 and 5.033, respectively, selectively distinguish anti-fertility activity from estrogenic potency.

Discussion

Reasonably well-predicting models for estrogenic, including anti-fertility, potencies of steroid compounds were obtained with cross-validated variance (CVV) [30] exceeding 75%. The models generated to differentiate anti-fertility and estrogenic activities also account for more than 75% of the variance in observed activities with low estimation errors. The best uni- and bi-variate relations for estrogenic potency demonstrate the importance of atom C_{17}, and the presence of a substituent that behaves as a hydrogen bond acceptor at C_{11}. The corresponding relations for anti-fertility activity indicate the importance of substitutions at atoms C_{16} and C_{17}. However, the best models (Eqs. 1, 2) reveal the importance of atom C_{17}, and the presence of a hydrogen bond acceptor at C_{11}, for estrogenic, including anti-fertility, activities. Furthermore, molecular orientation and conformational rigidity in the y-axis has a significant impact on both activities. In addition, the contribution of charge functions at atoms C_5 and C_7, along with increments of overall molecular electron affinity, influence estrogenic

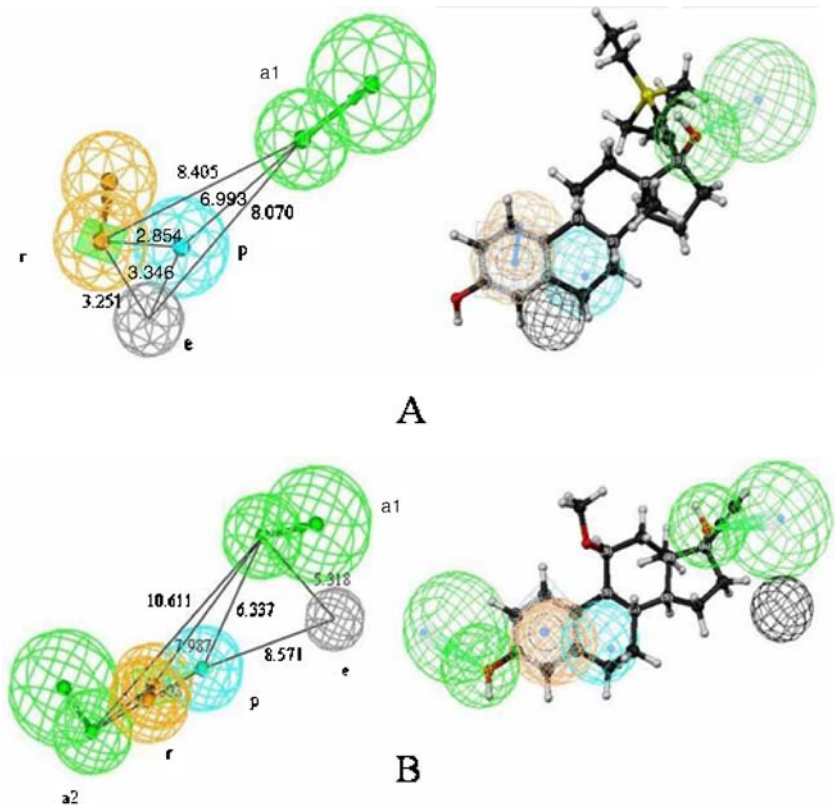


Fig. 3 Mapped pharmacophore features of estrogenic (**a**) and anti-fertility (**b**) potencies of compounds **16** and **87** (most active compounds in Tr). Features are HB acceptor-lipid (*a*), hydrophobic (*p*), ring aromatic (*r*) and excluded volume (*e*)

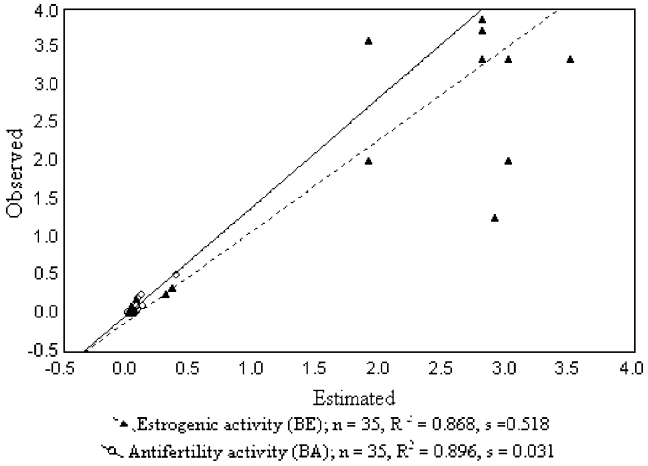


Fig. 4 Observed vs estimated activities obtained from best hypotheses (hypothesis 1, runs 15 and 30) of estrogenic and antifertility potencies

potency; however, the presence of suitable substituents at atoms C_2 and C_{16} selectively increase anti-fertility activity.

In both models, the negative coefficient of the partial charge function at atom C_{17} signifies that amplification of the partial charge or more electronegativity at C_{17} will result in an increase in both potencies. Similarly, increased electronegativity at atoms C_5 (Eq. 1) and C_{16} (Eq. 2) favors estrogenic and anti-fertility potencies, respectively, which can be achieved by substitution by an electron-donating group on target atoms, or a strong electron-withdrawing group on neighboring atoms. The negative contribution of the binary indicator, I_{11Hba} shows that the presence of a substituent at C_{11} that could behave as a hydrogen bond acceptor is essential for both estrogenic and contraceptive activities. Again, the negative contribution of the spatial parameter, PMI_Y in both cases (Eqs. 1, 2) indicates that an increase in the value of this parameter, i.e., greater conformational rigidity of the molecule, can increase both activities. Furthermore, the negative contribution of E_{LUMO} in Eq. 1 reveals that decreased electron affinity of the molecule is detrimental to estrogenic properties. Again, the positive coefficients of charge function at atoms C_7 (Eq. 1) and C_2 (Eq. 2) suggest that more positive change contribution on that atoms (atoms C_7 and C_2) will result in decreased estrogenic and anti-fertility potencies, respectively. Thus, substitutions by electron-withdrawing group (s) on target atoms that decrease charge functionality at atoms C_7 and C_2 will increase both activities.

In summary, the presence of an electron-donating substituent at C_{17}, and a hydrogen bond acceptor at C_{11}, along with high conformational rigidity of the steroidal estrogen, impart the estrogenic contraceptive property. Furthermore, high electron density at C_5, substitution by an electron-withdrawing group at C_7, and high electron affinity of the molecule aid estrogenic potency, whereas low electron density at C_2 and more electronegativity at C_{16} positions favor anti-fertility activity.

The quality of the best hypotheses generated in either case for pharmacophore space modeling of estrogenic and anti-fertility potencies are significant with regards to the cost differences, correlation coefficients and rmsd recorded. The best hypothesis (hypothesis 1 of run 9) and hyporefine (hypothesis 1 of run 15) on the same for estrogenic potency demonstrate around 97% correlation with activity, while that for anti-fertility activity (hypothesis 1 of runs 27 and 30) is in the range of 94%. In both cases, CatScrambling-based cross-validation demonstrates that none of the spreadsheets generated better parameters compared to the original hypotheses. The cross-validation results clearly indicate the superiority of the hypotheses selected, and also provide strong confidence (99% level) in the initial pharmacophore, i.e., hypothesis 1 in both cases.

From this study (Table 4), it can be deduced that one HB acceptor, hydrophobic and aromatic ring features, along with steric hindrance of the molecule, might function as prime biophores for both the potencies examined here. However, the presence of a second acceptor, and the critical interfeature distances (Fig. 3) differentiate the anti-fertility activity from the estrogenic potency. These studies can be corroborated further with the facts that one aromatic ring [36] is essential for any estrogenic activity that also includes contraceptive potency; and ring B of steroidal estrogen indicates a planar hydrophobic region [21], where the presence of the polar group is unfavorable for estrogenic potency. The crystal structure (1ERE) [37] of E_2 bound with ER, demonstrates that atoms C_{17} and C_{11} of the steroidal skeleton bind with the imidazole of His524, and the guanidinium of Arg394 by hydrogen bonds, whereas atom C_3 binds via another hydrogen bond to the γ-carboxylate of Glu353. The latter hydrogen bond is

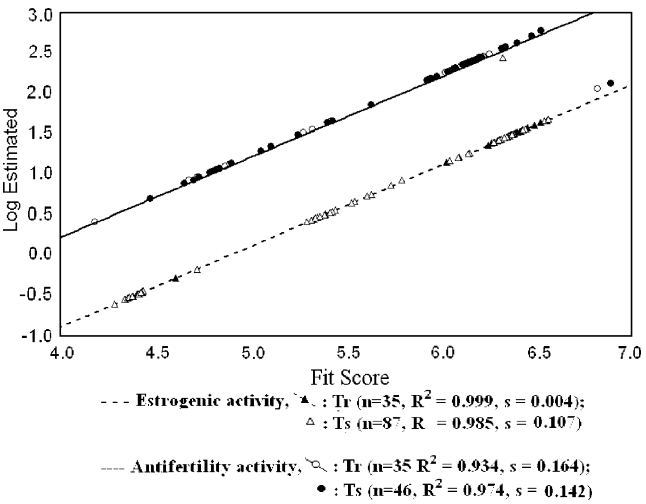


Fig. 5 Estimated vs fit score obtained from best hypotheses (hypothesis 1, runs 15 and 30) of estrogenic and antifertility potencies

essential for contraceptive potency in the absence of other estrogenic activities. The presence of a bulky substituent imparts a steric influence on the molecule, as greater orientation and conformational rigidity [38] was calculated in the projected hypotheses for both cases.

Thus, the mapped pharmacophore features (Fig. 3) for estrogenic and contraceptive potencies indicate that the presence of an aromatic ring (ring A), a hydrophobic zone (ring B) and a hydrogen bond acceptor at C_{17} (ring D), along with steric influence in the molecule, are critical for imparting both potencies, but that the presence of second acceptor in ring A selectively differentiates anti-fertility from estrogenic potency.

Conclusions

The present study characterizes some important pharmacophores for differentiating anti-fertility potency from estrogenic activity of steroidal derivatives. The work supports the facts that high electronegativity at atom C_{17}, the presence of a substituent that behaves as a hydrogen bond acceptor at atom C_{11}, and more bulky derivative(s) that impart greater orientation and conformational rigidity of the molecule, result in enhanced estrogenic, including contraceptive, potency. Additionally, high electronegativity at atom C_5, the presence of an electron-withdrawing substituent at atom C_7, which makes the steroidal skeleton more hydrophobic, and overall increased electron affinity of the molecule favor estrogenic activities. Substitutions in rings A and D that decrease electron density at atom C_2, and increase electronegativity at atom C_{16}, impart selective anti-fertility potency, distinct from other estrogenic activities. Pharmacophore space modeling studies also support the presence of aromatic ring A, a hydrophobic zone in ring B, and a hydrogen bond acceptor at atom C_{17} in ring D. These features, along with molecular steric hindrance, primarily govern both estrogenic and contraceptive potencies. However, the presence of a second acceptor in ring A and the critical distances between the features of the estrogen steroidal scaffold selectively differentiate anti-fertility potency from other estrogenic activities.

Acknowledgments Financial assistance from the University Grants Commission, New Delhi (India) under Major Research Project and University Potential of Excellence scheme under Modern Biology of University of Calcutta, Kolkata (India) is gratefully acknowledged.

References

1. McDonnell DP (1999) Trends Endocrino Metab 10:301–309
2. Park WC, Jordan VC (2002) Trends Mol Med 8:82–88
3. Mahapatra ABS (2001) Essentials of medical physiology, 1st edn. Current Books International, Kolkata, India, pp 374–376
4. Rang HP, Dale MM, Ritter JM, Moore PK (2003) In: Rang HP et al (eds) Pharmacology, 5th edn. Churchill Livingstone, New York, pp 431–433
5. Walsh BW, Schiff I, Rosner B, Greenberg L, Ravnikar V, Sacks FM (1991) N Engl J Med 325:1196–1204
6. Kuiper GG, Enmark E, Pelto-Huikko M, Nilsson S, Gustafsson JA (1996) Proc Natl Acad Sci USA 93:5925–5930
7. Chan L, O'Malley BW (1976) N Engl J Med 294:1322–1328, 1372–1381
8. Beekman JM, Allan GF, Tsai SY, Tsai MJ, O'Malley BW (1993) Mol Endocrinol 267:1266–1274
9. Sadler BR, Cho SJ, Ishaq KS, Chae K, Korach KS (1998) J Med Chem 41:2261–2267
10. Peters RH, Crowe DF, Tanabe M, Avery MA, Chong WKM (1987) J Med Chem 30:646–652
11. Peters RH, Crowe DF, Avery MA, Chong WKM, Tanabe M (1988) J Med Chem 31:572–576
12. Peters RH, Crowe DF, Avery MA, Chong WKM, Tanabe M (1989) J Med Chem 32:1642–1652
13. Schultz TW, Sinks GD, Cronin MTD (2000) Environmental Toxicol Chem 19:2637–2642
14. Saliner AG, Amat L, Carbo-Dorca R, Schultz TW, Cronin MTD (2003) J Chem Inf Comput Sci 43:1166–1176
15. Schmieder PK, Aptula AO, Routledge EJ, Sumpter JP, Mekenyan OG (2000) Environ Toxicol Chem 19:1727–1740
16. Bradbury SP, Kamenska V, Schmieder PK, Ankley GT, Mekenyan OGA (2000) Toxicol Sci 58:253–269
17. Waller CL, Oprea TI, Chae K, Park HK, Korach KS, Laws SC, Wiese TE, Kelce WR, Gray LE Jr (1996) Chem Res Toxicol 9:1240–1248
18. Shi LM, Fang H, Tong W, Wu J, Perkins R, Blair R, Branham W, Sheehan D (2001) J Chem Inf Comp Sci 41:186–195
19. Pasha F, Srivastava H, Singh P (2005) Mol Divers 9:215–220
20. Suzuki T, Ide K, Ishida M, Shapiro S (2001) J Chem Inf Comput Sci 41:718–726
21. Sippl W (2000) J Comp-Aided Mol Des 14:559–572
22. Mukherjee S, Mukherjee A, Saha A (2004) Bioorg Med Chem Lett 14:897–900
23. CS Chem3D Pro and CS MOPAC Pro 5.0 (1999) CambridgeSoft Corporation, Cambridge, MA
24. Dewar MJS (1969) The molecular orbital theory of organic chemistry. McGraw-Hill, New York
25. Kier LB, Hall LH (1990) Pharm Res 7:801–807
26. ETSA-CS (2007) Chem Tech, CU, Kolkata, West Bengal, India
27. TSAR 3.3 (2000) Oxford Molecular, Accelrys, San Diego, CA
28. Statistica 5.0 (1995) StatSoft, Tulsa, OK
29. Snedecor GW, Cochran WG (1967) Statistical methods, 6th edn. Iowa State University Press, Iowa
30. Wold S, Eriksson L (1995) Chemometric methods in molecular design. In: de Waterbeemd HV (ed) Methods and principles in medicinal chemistry, vol 1. VCH, Weinheim
31. Kristam R, Gillet VJ, Lewis RA, Thorner D (2005) J Chem Inf Model 45:461–476
32. Li H, Sutter J, Hoffman R (1999) In: Guner OF (ed) Pharmacophore perception, development, and use in drug design. International University Line, La Jolla, CA
33. Catalyst 4.11 (2005) Accelrys, San Diego, CA
34. Barnum D, Greene J, Smellie A, Sprague P (1996) J Chem Inf Comput Sci 36:563–571
35. Kirkpatrick S, Gelatt CD, Vecchi MP Jr (1983) Science 220:671–680
36. Jordan VC, Mittal S, Gosden B, Koch R, Lieberman ME (1985) Environ Health Perspect 61:97–110
37. Brzozowski AM, Pike ACW, Dauter Z, Hubbard RE, Bonn T, Engstrom O, Ohman L, Greene GL, Gustafsson JA, Carlquist M (1997) Nature 389:753–758
38. Mukherjee S, Mukherjee A, Saha A (2005) J Mol Struct Theochem 715:85–90