Available online at www.sciencedirect.com
ScienceDirect
Journal of Molecular Graphics and Modelling 26 (2008) 884–892
Journal of Molecular Graphics and Modelling
www.elsevier.com/locate/JMGM

Pharmacophore mapping of arylbenzothiophene derivatives for MCF cell inhibition using classical and 3D space modeling approaches☆

Subhendu Mukherjee [1], Shuchi Nagar [1], Sanchita Mullick, Arup Mukherjee, Achintya Saha *

Department of Chemical Technology, University of Calcutta, 92, A.P.C. Road, Kolkata 700009, India

Received 27 February 2007; received in revised form 11 June 2007; accepted 11 June 2007
Available online 16 June 2007

Abstract

Considering the worth of developing non-steroidal estrogen analogs, the present study explores the pharmacophore features of arylbenzothiophene derivatives for inhibitory activity to MCF-7 cells using classical QSAR and 3D space modeling approaches. The analysis shows that presence of phenolic hydroxyl group and ketonic linkage in the basic side chain of 2-arylbenzothiophene core of raloxifene derivatives are crucial. Additionally piperidine ring connected through ether linkage is favorable for inhibition of breast cancer cell line. These features for inhibitory activity are also highlighted through 3D space modeling approach that explored importance of critical inter features distance among HB-acceptor lipid, hydrophobic and HB-donor features in the arylbenzothiophene scaffold for activity.

Keywords: Pharmacophore mapping; Arylbenzothiophene; MCF cell-line inhibition; QSAR study; Space modeling

1. Introduction

Estrogen is known to play an important role in reproductive endocrinology, involved in growth and function of other tissues, such as the skeleton, cardiovascular and central nervous systems in both male and female [1,2]. It is also important for supporting homeostasis in a women's body, as evidenced by the progressive changes that occur at menopause, when ovarian estrogen synthesis stops around the age of 45. The decreased production of estrogen leads to certain postmenopausal pathologies, such as osteoporosis and coronary artery diseases (CAD) [3,4]. Hormone replacement therapy (HRT, specifically estrogen replacement therapy) is primarily used for the treatment of postmenopausal diseases and is found to be very beneficial for the treatment of osteoporosis. Despite the beneficial effects of HRT, it is observed that this therapy also encourages the development and growth of cancer in the breast and uterus. Thus search began for the treatment of the women reproductive cancer.

Selective estrogen receptor modulators (SERMs) [5,6] that show tissue-dependent agonistic or antagonistic behavior [7], are used as first line treatment for estrogen responsive breast cancer and for therapy against osteoporosis. The pure estrogen antagonists (anti-estrogens) are currently in clinical development for breast cancer treatment. SERMs are used as alternative and research is still on to obtain a variety of non-steroidal compounds, which interact with the estrogen receptor. These molecules are found to fully antagonize the effects of estrogen on uterine and mammary tissues, while mimicking the effects on bone and cardiovascular system [8]. Several group of compounds have been developed as SERM's, one such compound being raloxifene [9]. It is known that raloxifene is the first SERM to be available for the treatment of osteoporosis and is also being tested as a preventive measure for breast cancer and coronary heart disease (CHD) [9–11]. The present work has been designed to explore the pharmacophore features for MCF cells inhibition of 2-arylbenzothiophene core (Fig. 1) of raloxifene by implementing both classical as well as space modeling studies.

The pharmacophore concept is based on the kinds of interaction observed in molecular recognition, i.e., hydrogen bonding, charge and hydrophobic interaction and alternatively

☆ Presented in the 11th Annual International Conference on Research in Computational Molecular Biology (RECOMB 2007), San Francisco Bay Area, California, USA, April 21–25, 2007.

* Corresponding author. Tel.: +91 33 23508386; fax: +91 33 23519755.
E-mail address: achintya_saha@yahoo.com (A. Saha).

[1] Both the authors equally contributed in the work.

1093-3263/$ – see front matter © 2007 Elsevier Inc. All rights reserved.
doi:10.1016/j.jmgm.2007.06.003

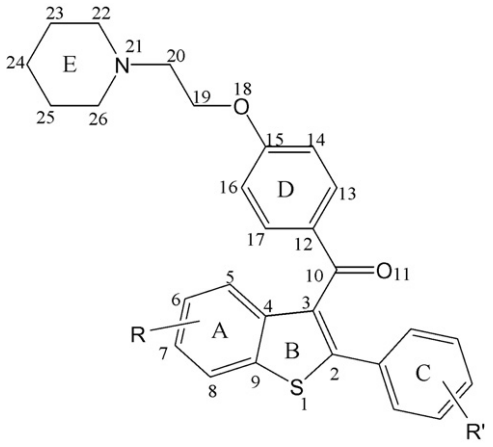


Fig. 1. General structure of arylbenzothiophene derivative.

can be used as a query in a 3D database search to identify new structural classes of potential lead compounds; and it can serve as a template for generating alignment for 3D QSAR analysis [12]. Two types of pharmacophore hypothesis are well established: receptor-based and receptor-independent pharmacophore. Receptor based pharmacophore mapping of SERMs [13–17] are commonly employed, but now a days receptor-independent pharmacophore mapping is growing interestingly for deriving bioactivities of diverse group of compounds, such as aromatase inhibitors [18], serotonin inhibitors [19] and antithrombin [20] agents and are now increasingly being handled by automated computational methods, such as *CATALYST* [21], *GASP*, *DISCO* which are commercially available programs [22–24]. Consequently, the present work is taken up to study the arylbenzothiophene scaffold as a small ligand [25] with a view to deduce the active pharmacophore signals based on receptor-independent hypothesis, using the CATALYST program [21], that can eventually aid in apprehending the tissue-specific effects of different compounds containing this unit.

2. Materials and methods

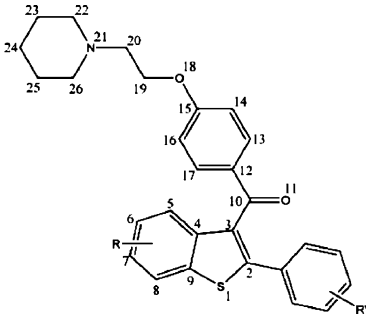
In the present study, 69 compounds of a series of raloxifene analogs which contain modifications of the 2-arylbenzothiophene core [25] are considered (Table 1) and are randomly segregated into training (Tr) and test (Ts) sets. For classical QSAR studies, the biological activity is expressed as logarithm to the base 10 of IC_{50} (pIC_{50}) to MCF-7 cells. Molecular (partition coefficient [26], hydrophobicity [26], steric [27,28] and thermodynamic factors [29], bulk, moments and orbital energies), electronic (Wang-Ford atomic charge [30] and extended Huckel partial charge [31–33] functions) and electrotopological (*E*-state indices) [34] properties have been explored for deriving classical QSAR models of MCF-7 cell-line inhibition of arylbenzothiophene derivatives using JAVA based program *ETSA–CS* [35], *TSAR* 3.3 [36], *CAChe* 6.1 [37] and *Chem 3D Pro* [38]. The common atoms of this group of compounds have been numbered for computation of charge and electro-topological functions. The indicator variables used for modeling the bioactivities are $I_{7\text{-OH}}$ and $I_{\text{Subs\_6}}$ that signify the presence of hydroxy substitution at atom C_7 and substitutional requirement at atom C_6, respectively.

Statistical analysis is performed by *Statistica* 5.0 [39] using standard and forward stepwise multiple regression methods. The different statistical parameters of the regression equation considered are: r or R (correlation coefficient), EV (explained variance), F (variance ratio) with d.f. (degree of freedom), s (standard error of estimate) and AVRES (average of absolute values of residuals). Leave-One-Out (LOO) cross-validation [40] is performed that generated PRESS (predictive residual sum of squares), SDEP (standard deviation of error of predictions), Pres_{av} (average of absolute value of predicted residuals) and Q^2 (cross-validated variance).

Pharmacophore space modeling study is also performed with this dataset using *CATALYST* 4.11 [21]. The biological activity is expressed as inhibitory concentration (IC_{50}) to MCF-7 cell-line. The pharmacophore models (hypotheses), generated by CATALYST [21], consist of an array of features necessary for bioactivity of the ligands arranged in 3D space that can explain the variance in activity of the molecules w.r.t. geometric localization of the chemical features present in them. To be retrieved as a hit, a candidate ligand must possess appropriate functional groups which can simultaneously reside within the respective tolerance spheres of the pharmacophoric features. Each feature is associated with a weight (a measure of its proposed 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 [22].

Different control parameters employed for hypothesis generation (called a HypoGen process) are spacing, uncertainty and weight variation. Spacing is a parameter representing the minimum inter-features distance that may be allowed in the resulting hypothesis. In the present work, the value of spacing used is 600 pm to accommodate maximum number of chemical features in the hypothesis. 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 2 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. A default value of 3 is considered as the uncertainty parameter in the present work. While generating hypothesis, a total cost function is minimized comprising of three terms, viz. weight cost, error cost and configuration cost. Weight cost is a value that increases as the weight variation in the model deviates from an ideal value of 2. The deviation between the estimated activities of the training set and their experimentally determined values is the error cost. A fixed cost that depends on the complexity of the hypothesis space being optimized is denoted as the configuration cost. The configuration cost is equal to the entropy of hypothesis space. The CATALYST program also calculates the cost of null hypothesis (null cost) that assumes no relation in the data, and the

Table 1
Structural features and inhibitory activity of arylbenzothiophene derivatives



Compound no.	Substituents		IC_{50} (nM)	Compound no.	Substituents		IC_{50} (nM)
	R	R′			R	R′	
1	H	H	300.0	[a]**36**	7-OH	Methyl	35.0
2	7-OMe	4′-OMe	300.0	[a]**37**	7-OH	Ethyl	20.0
3	7-OH	4′-OMe	1000.0	[a]**38**	7-OH	Cyclopentyl	5.0
4	7-OMe	4′-OH	250.0	[a]**39**	7-OH	Cyclohexyl	2.5
5	H	4′-OMe	100	[a]**40**	7-OH	trans-4′-OH-Cyclohexyl	2.0
6	H	4′-OH	35.0	[a]**41**	7-OH	4′-Pyridyl	100.0
7	7-OH	4′-Cl	1.0	**42**	7-OH	4′-OH	0.2
8	7-Cl	4′-OH	1000.0	**43**	7-OH	4′-F	2.3
9	7-OH	4′-Me	50.0	**44**	6,7-di (OH)	4′-OH	400.0
10	7-Me	4′-OH	300.0	**45**	6,7,8-tri (OMe)	4′-OMe	350.0
11	8-OH	4′-OH	300.0	**46**	6-F, 7-OH	4′-OH	3.0
12	5-OH	4′-OH	190.0	**47**	6,8-di (Me), 7-OH	4′-OH	500.0
13	6-OH	4′-OH	100.0	**48**	7-OH	2′-OH	10.0
14	5,7-di (OH)	4′-OH	350.0	**49**	7-OH	3′-OH	3.2
15	5,8-di (Me), 7-OH	4′-OH	100.0	**50**	7-OH	2′-Me	0.7
16	5,6-benzo, 7-OH	4′-OH	500.0	**51**	7-OH	3′-F	2.5
17	7-OH	H	2.5	**52**	7-OH	4′-Et	5.0
18	7-OH	4′-Ph	100.0	**53**	7-OH	4′-iPr	30.0
19	7-OMe	4′-CH_2OH	600.0	**54**	7-OH	4′-*n*Bu	10.0
20	7-OH	4′-CH_2SEt	100.0	**55**	7-OH	2′-Me, 4′-OH	2.0
21	7-OH	4′-CF_3	1000.0	**56**	7-OH	3′-Me, 4′-OH	1.0
22	7-OH	2′-OMe, 4′-OH	2.0	**57**	7-OH	3′-Cl, 4′-OH	2.3
23	7-OH	3′,5′-di (Me), 4′-OH	100.0	**58**	7-OH	3′-F, 4′-OH	0.3
24	7-OMe	3′,4′-OCH_2O	500.0	**59**	7-OH	4′-CO_2Me	50.0
25	7-OH	4′-NO_2	500.0	**60**	7-OH	4′-CO_2Et	50.0
26	7-CO_2Me	4′-OH	30.0	**61**	7-OH	4′-CO_2H	325.0
27	7-$CONH_2$	4′-OH	1000.0	**62**	7-OH	4′-C≡CH	0.8
28	7-COMe	4′-OH	60.0	[a]**63**	7-OH	1′-Napthyl	0.8
29	7-C≡CH	4′-OH	20.0	[a]**64**	7-OH	4′-OH-1′-napthyl	2.0
30	7-OH	4′-$CONH_2$	200.0	[a]**65**	7-OH	2′-Thienyl	20.0
31	7-OH	4′-CON(H)Me	40.0	[a]**66**	7-OH	3′-Thienyl	10.0
32	7-OH	4′-$CONMe_2$	20.0	[a]**67**	7-OH	Isopropyl	3.0
33	7-OH	4′-COMe	32.0	[a]**68**	7-OH	4′-Hydroxybenzyl	5.0
34	7-OH	4′-CH=CH_2	7.0	[a]**69**	7-OH	4′-Pyridyl *N*-oxide	100.0
[a]**35**	7-OH	2′-Napthyl	80.0				

[a] 2-phenyl replaced by R′ substituent directly at C_2; the common atoms have been numbered **1–26**.

experimental activities are distributed about their mean. Accordingly, the greater the difference (Δcost) between the null costs and the total cost, it is more likely that the hypothesis does not reflect a chance correlation.

In the present study, the number of conformers generated for each compound has been limited to maximum of 250 using 'best conformer generation' method with 20 kcal/mol energy *cutoff*. The input pharmacophore features used are HB acceptor-lipid (al), positive ionizable (pi) group, HB donor (d), hydrophobic (p) and ring aromatic (r). The HypoGen algorithm is forced to find pharmacophore model that contains maximum two same feature and others from the input chemical features. The quality of the generated hypothesis is adjudged through a cross-validation technique using Cat-Scramble [21], based on Fischer's randomization test [41], where the biological activity data are randomized within a fixed chemical data set and HypoGen process initiated to explore possibilities of other hypotheses of good predictive values. By logic, the hypothesis generated prior to scrambling should be better to attest for a good pharmacophore model.

Table 2
Calculated and predicted inhibitory activity to MCF-7 cell-line (IC_{50}) of training and test sets of arylbenzothiophene derivatives

Compound no.	Set	MCF-7 cell-line inhibition (pIC_{50})			Compound no.	Set	MCF-7 cell-line inhibition (pIC_{50})		
		Obs.	Calc.[a]	Pred.[a]			Obs.	Calc.[a]	Pred.[a]
1	Tr	2.477	2.316	2.300	**36**	Ts	1.544	–	0.572
2	Tr	2.477	1.968	1.889	**37**	Ts	1.301	–	0.414
3	Ts	3.000	–	1.636	**38**	Tr	0.699	0.512	0.500
4	Tr	2.398	1.997	1.930	**39**	Tr	0.398	0.336	0.331
5	Tr	2.000	2.347	2.402	**40**	Tr	0.301	0.369	0.375
6	Tr	1.544	1.260	1.128	**41**	Tr	2.000	1.486	1.457
7	Ts	0.000	–	1.033	**42**	Tr	−0.699	−0.100	0.040
8	Tr	3.000	2.641	2.532	**43**	Ts	0.362	–	1.525
9	Tr	1.699	1.134	1.105	**44**	Tr	2.602	2.526	2.500
10	Tr	2.477	2.608	2.631	**45**	Ts	2.544	–	3.445
11	Ts	2.477	–	1.599	**46**	Ts	0.477	–	2.724
12	Tr	2.279	2.788	2.947	**47**	Tr	2.699	2.440	2.345
13	Tr	2.000	1.980	1.972	**48**	Tr	1.000	0.573	0.549
14	Ts	2.544	–	1.997	**49**	Ts	0.505	–	1.720
15	Tr	2.000	2.373	2.547	**50**	Tr	−0.155	0.316	0.372
16	Tr	2.699	3.054	3.265	**51**	Ts	0.398	–	1.322
17	Ts.	0.398	–	1.280	**52**	Tr	0.699	1.230	1.255
18	Tr	2.000	1.948	1.939	**53**	Ts	1.477	–	0.113
19	Tr	2.778	2.587	2.553	**54**	Tr	1.000	0.852	0.846
20	Tr	2.000	1.365	1.334	**55**	Tr	0.301	0.385	0.394
21	Ts	3.000	–	1.630	**56**	Ts	0.000	–	0.847
22	Tr	0.301	0.090	0.054	**57**	Tr	0.362	0.961	0.990
23	Ts	2.000	–	0.962	**58**	Ts	−0.523	–	0.983
24	Ts	2.699	–	1.999	**59**	Tr	1.699	1.736	1.739
25	Ts	2.699	–	2.013	**60**	Tr	1.699	1.732	1.735
26	Tr	1.477	1.957	2.012	**61**	Ts	2.512	–	1.585
27	Ts	3.000	–	2.005	**62**	Ts	−0.097	–	1.452
28	Tr	1.778	1.860	1.871	**63**	Tr	−0.097	0.294	0.329
29	Tr	1.301	1.658	1.794	**64**	Tr	0.301	0.788	0.823
30	Tr	2.301	1.644	1.581	**65**	Tr	1.301	0.929	0.914
31	Tr	1.602	1.621	1.623	**66**	Tr	1.000	1.020	1.022
32	Tr	1.301	1.546	1.564	**67**	Tr	0.477	0.844	0.859
33	Tr	1.505	1.768	1.814	**68**	Tr	0.699	0.444	0.422
34	Tr	0.845	0.341	0.149	**69**	Ts	2.000	–	1.926
35	Ts	1.903	–	0.704					

Obs., observed values [25]; Calc., calculated values; Pred., predicted values. Tr, training and Ts, test set; $pIC_{50} = \log_{10}IC_{50}$.
[a] As per Eq. (1).

The coverage value of the hypothesis is again generated with the help of ''Best Flexible Search''. This pharmacophoric coverage is defined as the total number of compounds captured by the model [21]. The explored pharmacophore model is further adjudged by correlating binding interactions (Fig. 5a and b) exhibited by raloxifene in the X-ray crystal structures of ligand-bound human (1ERR) [42] and rat (1QKN) [43] ERs for confirmation.

3. Results and discussion

In the present study, molecular, electronic and electro-topological properties of 69 compounds of 2-arylbenzothiophene core [25], segregated into training (Tr) and test (Ts) sets (Table 1), are considered for modeling. The Tr comprised of 45 compounds (Compounds **1**, **2**, **4–6**, **8–10**, **12**, **13**, **15**, **16**, **18**, **19**, **20**, **22**, **26**, **28–34**, **38–42**, **44**, **47**, **48**, **50**, **52**, **54**, **55**, **57**, **59**, **60**, **63–68**) and rest 24 compounds are used as Ts. In the models, 95% confidence intervals are shown in parentheses and the *F*-values are significant at 99% confidence level. The regression constants for all relations are significant at 95%.

3.1. QSAR studies

From regression analysis of the Tr, it has been observed that the best significant univariate relation could be developed with Wang-Ford charge function of atom O_{11} that explained 46.448% variance and statistical quality of the relation is estimated to be

$$r = 0.690; \quad r^2 = 0.477; \quad s = 0.672; \quad n = 45$$

and in case of bivariate relationship, the best significant relationship has been explored with the same charge function of atom O_{11} (Ch_{11}) with an indicator ($I_{7\text{-OH}}$) that signifies the presence of hydroxyl group at C_7 and explained 58.984% variance in activity. The quality of this relationship has been estimated to be

$$R = 0.780; \quad R^2 = 0.608; \quad s = 0.588; \quad n = 45$$

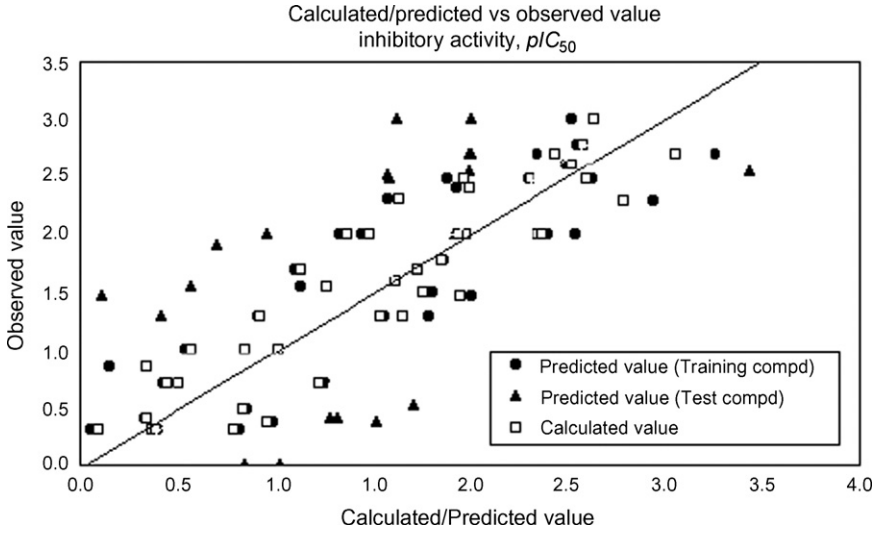


Fig. 2. Observed and calculated values as per Eq. (1).

But the best significant relationship for inhibition of MCF-7 cell-line has been deduced to be

$$pIC_{50} = -8.667(\pm 0.993)Ch_{10} - 29.136(\pm 10.139)Ch_{18} + 109.293(\pm 22.629)PCh_{26} - 0.916(\pm 0.141)I_{7-OH} + 0.975(\pm 0.215)I_{Subs\_6} - 11.479(\pm 5.254) \quad (1)$$

$n = 45$; $R = 0.918$; $R^2 = 0.842$; EV(%) = 82.196; $F = 41.628$; $df = 5, 39$; $s = 0.387$; AVRES = 0.304; PRESS = 7.931; SDEP = 0.420; Presav = 0.355; $Q^2 = 0.786$

where, Ch_{10}, Ch_{18} and PCh_{26} signify Wang-Ford atomic charge function of C_{10} and O_{18} and extended Huckel atomic partial charge on C_{26} respectively and $I_{Subs\text{-}6}$ indicates the substitutional requirement at atom C_6.

The calculated and predicted inhibitory activities of Tr and Ts compounds as per Eq. (1) have been delineated in Table 2 and Fig. 2. The independent variables used in Eq. (1) are not statistically intercorrelated ($r < 0.3$).

The best uni- and bivariate relations demonstrated the importance of atom O_{11} and presence of hydroxyl group at atom C_7 of arylbenzothiophene derivative for cell line inhibition. But, the overall best relation (Eq. (1)) can explain 92% correlation and exhibited importance of atoms C_{10}, O_{18} and C_{26}, presence of hydroxyl group at atom C_7 and substitutional requirement at atom C_6. In the model, the negative contribution of charge functions at atoms C_{10} and O_{18} reveal that more negative charge at these atomic positions will increase the bioactivity. Thus, the presence of ketonic functional group as well as ether linkage in the basic side-chain enhances the inhibitory activity of raloxifene to the breast cancer cells. The negative contribution of the binary indicator, $I_{7\text{-}OH}$ indicates that presence of phenolic hydroxyl substitution at C_7 is essential for inhibition of the cell-line. Again, the positive contribution of PCh_{26} suggests that more positive charge contribution on atom C_{26} in the piperidine ring will result in decreased potency. Thus substitution in the piperidine ring that decreases charge functionality at atom C_{26} will increase inhibitory activity. These results are further adjudged with the SAR studies [44,45], where pharmacophoric groups of raloxifene are described as importance of benzothiophene core, 7-phenolic hydroxyl group, ketonic 'hinge' and the basic amine chain for binding activity.

Again in the model (Eq. (1)) the positive contribution of I_{Subs_6} explains that presence of substitution at atom C_6 adjacent to atom C_7, has a negative effect on the potency. Comps. **44** and **47** have OH-substitution at C_7 and also other substitution at C_6 and are less potent compared to Comp. **42** having only OH substitution at C_7. Thus further substitution at atom C_6 can decrease the inhibitory activity. Local Intersection Volume (LIV) [46,47], a 3D local shape descriptor, is used for QSAR study of arylbenzothiophene–raloxifene analog for inhibitory activity as a case study. The model [13] ($n = 41$; $R^2 = 0.76$; $Q^2 = 0.68$) also supports the importance of 7-hydroxyl group, ether linkage and piperidine ring for the activity. But larger substituent at C_6 reduces the potency.

3.2. Space modeling

Pharmacophore space modeling studies for inhibitory activity of the compounds are randomly categorized into Tr

Table 3
Training and test subsets categorized in exploring pharmacophoric space of arylbenzothiophene derivatives

Set	Least-active subset (Compound)	Intermediate subset (Compound)	Most-active subset (Compound)	Number of compounds
Tr	**1**, **4**, **5**, **10**, **12**, **13**, **25**, **27**, **41**, **61**	**9**, **28**, **32**, **42**, **53**, **54**	**7**, **17**, **34**, **42**, **43**, **46**, **50**, **52**	24
Ts	**2**, **3**, **8**, **11**, **14**, **15**, **16**, **18**, **19**, **20**, **21**, **23**, **24**, **30**, **44**, **45**, **47**, **69**	**6**, **26**, **29**, **31**, **33**, **35**, **37**, **48**, **59**, **60**, **65**, **66**	**22**, **38**, **39**, **40**, **49**, **51**, **55**, **56**, **57**, **58**, **62**, **63**, **64**, **67**, **68**	45

Table 4
Hypotheses parameters observed in successive runs of training set

Run no.	Input features	Hypothesis no.	Pharmacophore features in generated hypothesis	Cost			*R*	rmsd
				Configuration	Total	Δ		
1	al, d, p, r, pi	1	al, d, 2 × p	15.491	105.530	26.941	0.928	0.774
		5	al, d, p, pi	15.491	109.240	23.231	0.888	0.954
2	al, d, r, pi	1	al, d, r	14.376	105.56	26.911	0.920	0.813
3	d, p, r, pi	1	d, 3 × p	13.363	104.721	27.750	0.915	0.833
4	al, p, r, pi	1	al, 3 × p	14.542	111.33	21.141	0.856	1.073
5	al, d, p, r	1	d, 3 × p	15.135	106.44	26.031	0.922	0.803
6	al, d, p, pi	1	al, d, 2 × p	14.530	104.94	27.531	0.924	0.793
7[a]	al, d, p, pi	–	–	27.101	–	–	–	–
8[b]	al, d, p, pi	–	–	27.101	–	–	–	–
9[a,b]	al, d, p, pi	–	–	38.711	–	–	–	–

al: HB acceptor-lipid; pi: positive ionizable group; d: HB donor; p: hydrophobic factor; r: ring aromatic; null cost = 132.471. Δcost = null cost − total cost; rmsd = rms deviation; spacing = 600 pm; $n = 24$; last 3 runs generated invalid hypothesis with configuration cost > 25.
[a] Weight variation = 1.
[b] Variable tolerance = 1.

and Ts sets as shown in Table 3. The results of this study have been stated in Tables 4 and 5. The mapped pharmacophore features of training set is described in Fig. 3 and estimated fit scores of the same Tr and Ts sets are delineated in Fig. 4. The hypothesis 1 of run number 1 of Tr set is adjudged to be the best hypothesis. This has been characterized in terms of mapped features, cost difference, root-mean-square divergence (rmsd) and the correlation coefficient of the generated hypothesis.

The hypothesis 1 of run number 1 of Tr has been subjected to ''Best Flexible Database Search'' against the entire set to obtain the coverage value of the hypothesis. The predictive capacities of the model is also analyzed using the SCORE command from the ''Generate Hypothesis'' workbench of CATALYST against the entire set to obtain the Scoring Correlation Coefficient (SCC > 0.5). On account of low Δcost generated in the hypotheses, hyporefine has not been performed.

Hydrogen bond (HB) acceptor-lipid (al), HB donor (d) and hydrophobic (p) features are found important in the Tr set (Table 4 and Fig. 3) exhibiting a correlation of 92.8%. The results demonstrated good predictive ability of the model as well. In CatScrambling based cross validation (Table 5), none of the hypothesis generated better parameters in comparison to original hypothesis. So, the cross-validation analyses clearly indicate the superiority of the hypothesis considered. The coverage ($n = 66$) and SCC (0.515) values recorded for the hypothesis of the training set suggest [21] that the hypothesis generated in the present work explained the variance in activity for most compounds. The projected hypothesis is further validated with some reference compounds (standard and non-

Table 5
Results of CatScramble cross-validation[a] of hypothesis 1 of run 1

Trial/Spreadsheet no.	Cost				*R*	rmsd
	Configuration	Fixed	Total	Δ		
1	14.711	98.287	117.163	15.308	0.791	1.269
2	15.491	98.287	115.961	16.510	0.814	1.206
3	15.536	98.333	115.764	16.707	0.821	1.185
4	13.693	96.489	117.486	14.985	0.773	1.318
5	13.774	96.571	119.978	12.493	0.754	1.365
6	15.481	98.278	115.930	16.541	0.820	1.187
7	13.160	95.957	119.942	12.529	0.738	1.401
8	15.018	97.815	113.028	19.443	0.847	1.104
9	15.410	98.197	120.312	12.159	0.809	1.230
10	14.313	97.110	107.898	24.573	0.891	0.941
11	15.251	98.047	110.934	21.537	0.867	1.033
12	14.016	96.813	124.230	8.241	0.688	1.506
13	15.013	97.810	112.530	19.941	0.848	1.098
14	14.565	97.362	120.333	12.138	0.746	1.381
15	15.047	97.844	115.126	17.345	0.818	1.195
16	15.250	98.046	115.233	17.238	0.826	1.170
17	14.881	97.677	126.323	6.148	0.672	1.537
18	14.712	97.510	110.723	21.748	0.869	1.030
19	15.543	98.340	121.661	10.810	0.745	1.384

Null cost = 132.471; Δcost = null cost – total cost; rmsd = rms deviation.
[a] 95% confidence level.

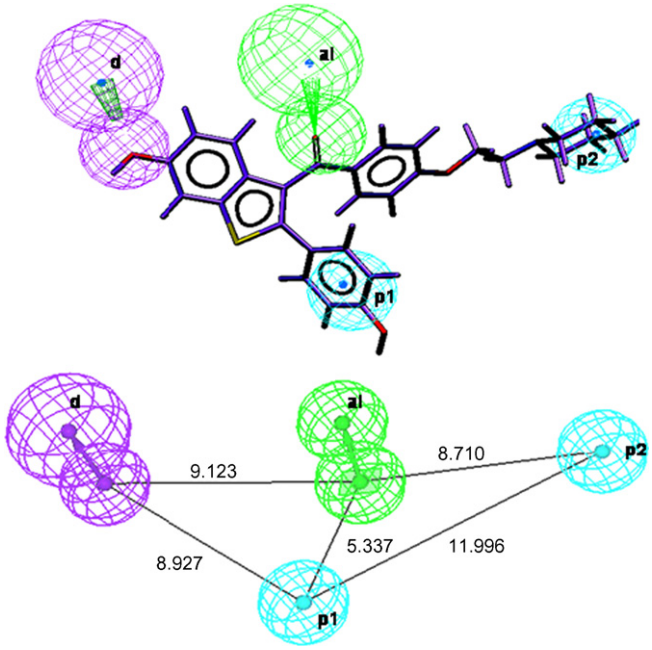


Fig. 3. Mapped pharmacophore features (h*ypothesis* 1 of *run* 1) of Compound **42** (highest active compound in training set 1); al: HB acceptor-lipid, d: HB donor and p: Hydrophobic factor; SCC = 0.51.

standard SERMs) by measuring fit scores and estimated values of such compounds are listed in Table 6.

From Fig. 3, it is evident that the benzothiophene hydroxyl group serves to be a HB-donor (d), the phenyl group attached as side-chain to this moiety and the piperidine moiety in the other side-chain serve as hydrophobic groups (p). Fig. 5 shows the ligand-binding domains (LBDs) in the X-ray crystal structures of raloxifene in human (Fig. 5a) [42] and rat (Fig. 5b) [43] ER varieties. From the figures, it can be observed that the benzothiophene hydroxyl group is within good H-bonding distance with E353 and R394 (in 1ERR) and with E260 and R301 (in 1QKN). This clearly explains significance of the hydrogen bond donor in the deduced pharmacophore model. Again, the phenyl group attached as side-chain to the benzothiophene moiety is found to be occupying a hydrophobic

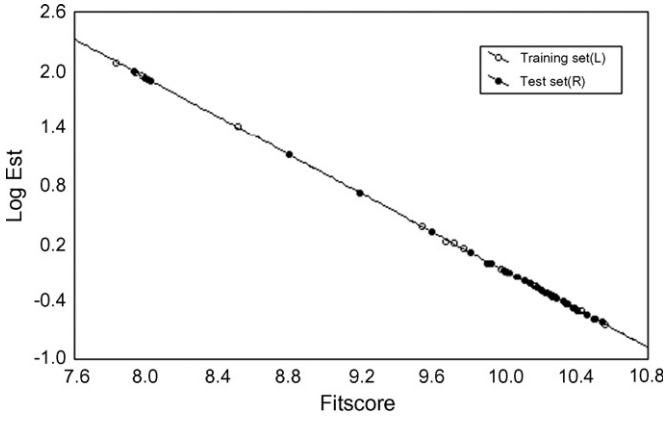


Fig. 4. Estimated vs. fit scores from best hypothesis of training and test sets. Training set ($n = 24$): Hypothesis 1 run no. 1: $R^2 = 1.000$, $s = 0.050$; Test set ($n = 45$): Hypothesis 1 run no. 1: $R^2 = 1.000$, $s = 0.050$.

Table 6
Estimated activities (IC_{50}) and fit scores of reference compounds

Compound	Est. activity (nM)	Fit score
Asigenin	3.9×10^2	7.33
Celecoxib	4.5×10^2	7.27
Clidium	5.7×10^4	5.17
Clomiphene	5.7×10^2	7.16
Diethylstilbesterol	3.6×10^4	5.36
Diphenylnaphthyl propylene	1.0×10^2	7.92
E2	8.0×10^4	5.02
EM-800 [48]	8.2×10	8.01
EM-652 [49]	7.7×10	8.03
Ephedrine	1.7×10^5	4.68
Etoposide	7.8×10	8.03
Gefitinib	1.0×10^2	7.90
Genistein	1.9×10^3	6.65
4-Hydroxy-tamoxifene	1.5×10^3	6.74
4-Hydroxy toremifene	3.8×10^2	7.34
4-Hydroxy trioxifene	7.5	9.05
Iodoxifen	1.6×10^2	7.71
Lasofoxifene	2.0×10^2	7.61
Letrozole	3.8×10^4	5.34
Luteolin	3.4×10^2	7.38
7-Methoxy flavone	3.6×10^4	5.36
Morphine	4.6×10^4	5.26
Nafoxidine	1.6×10^2	7.71
Norgestrel	3.6×10^4	5.36
Orphenadine	3.7×10^4	5.35
Phenytoin	3.7×10^4	5.35
Propranolol	3.6×10^4	5.36
Quercetin	3.7×10^2	7.35
Stanozolol	3.7×10^4	5.35
Tangeretin	4.2×10^2	7.30
Testosterone	3.7×10^4	5.35
Tetrahydroisoquinoline	8.0×10	8.02
TSE-424 [50]	9.9×10	7.92

Est.: Estimated activity.

region composed of the residues M388, F404 and L428 (in 1ERR) and M295 and I328 (in 1QKN). Thus, the importance of feature p1 (hydrophobic) in the model becomes further justified. The piperidine moiety can be found to be oriented towards a hydrophobic cleft made of L354, W383, L536 and L539 in the α-helical ambience (in case of 1ERR), which is not found to be so conserved in case of 1QKN. In spite of this, orientation of the piperidine chain in a hydrophobic cavity in case of 1ERR inevidently explains the importance of a hydrophobic group in this region. Significance of the HB acceptor-lipid (al) feature (Fig. 3) however cannot be explained through the crystal structures. It might be possible that this attribute is important as some kind of cell-permeation feature, as the activity profile considered in the present work is cell-line inhibition data rather than isolated protein based study findings.

These results can be further corroborated with the classical regressional analysis. It has been deduced from the QSAR model (Eq. (1)) that presence of phenolic hydroxyl substitution at atom C_7 and keto functional group in basic side chain are important for activity. These features behave as promising HB donor and HB acceptor-lipid features, respectively. Piperidine moiety connected through ether linkage in the aromatic ring D along with ring C can also behave as hydrophobic features of

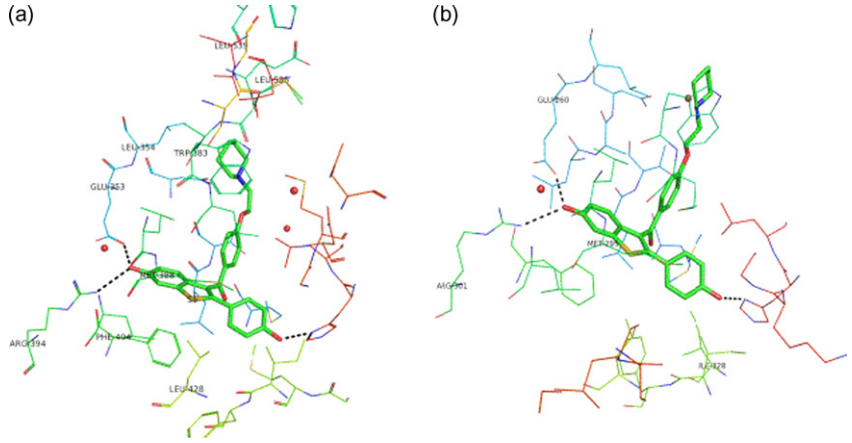


Fig. 5. Crystal structure of bound raloxifene in human [42] (a) and rat [43] (b) ER-LBD.

the molecule. The critical distances between the mapped pharmacophore features are also crucial for inhibitory activity.

4. Conclusion

In view of these observations, the present study could account for some of important pharmacophores of arylbenzothiophene derivatives for breast cancer cell inhibition. The study supports that the presence of ketonic group in the basic side-chain along with less positive charge on atom C_{26} of piperidine ring connected through ether linkage will be favorable for greater inhibition of the breast cancer cell-line. Additionally, presence of hydroxyl substitution at C_7 and no further substitution on atom C_6 might enhance the inhibitory activity. Pharmacophore space modeling studies also support the presence of HB acceptor-lipid, hydrophobic and HB donor along with critical distance among features in the raloxifene scaffold for inhibitory activity to the cell-line.

Acknowledgements

One of the author, S. Mukherjee wishes to thank CSIR, New Delhi (India) for the award of senior research fellowship. Financing assistance from AICTE, New Delhi (India) under RPS program and University TEQIP scheme are also thankfully acknowledged.

References

[1] E.P. Smith, J. Boyd, G.R. Frank, H. Takahashi, R.M. Cohen, B. Specker, T.C. Williams, D.B. Lubahn, K.S. Korach, Estrogen resistance caused by a mutation in the estrogen-receptor gene in a man, N. Engl. J. Med. 331 (1994) 1056–1061.

[2] D.R. Ciocca, R.L.M. Vargas, Estrogen receptors in human non-target tissues: biological and clinical implications, Endocrinol. Rev. 16 (1995) 35–62.

[3] R. Lindsay, Select steroids in the pathogenesis and prevention of osteoporosis, in: B.L. Riggs, L.J. Melton (Eds.), Osteoporosis: Etiology, Diagnosis, and Management, Raven Press, New York, 1988, pp. 333–358.

[4] R.K. Ross, M.C. Pike, T.M. Mack, B.E. Henderson, Oestrogen replacement therapy and cardiovascular disease, in: J.O. Drife, J.W.W. Studd (Eds.), HRT and Osteoporosis, Springer-Verlag, London, 1990, pp. 209–222.

[5] M. Dutertre, C.L. Smith, Molecular mechanisms of selective estrogen receptor modulator (SERM) action, J. Pharmacol. Exp. Ther. 295 (2000) 431–437.

[6] R.L. Rosati, P.D.S. Jardine, K.O. Cameron, D.D. Thompson, H.Z. Ke, Discovery and preclinical pharmacology of a novel, potent, nonsteroidal estrogen receptor agonist/antagonist, CP-336156, a diaryltetrahydronaphthalene, J. Med. Chem. 41 (1998) 2928–2931.

[7] W. Shao, M. Brown, Advances in estrogen receptor biology: Prospects for improvements in targeted breast cancer therapy, Breast Cancer Res. 6 (2004) 39–52.

[8] (a) H.G. Burger, Selective oestrogen receptor modulators, Hormone Res. 53 (2000) 25–29;
(b) S.R. Goldstein, S. Siddhanti, A.V. Ciaccia, L. Plouffe, A pharmacological review of selective oestrogen receptor modulators, Human Reprod. Update 6 (2000) 212–224;
(c) T.A. Grese, J.A. Dodge, Estrogen receptor modulators: effects in non-traditional target tissues, Annu. Rep. Med. Chem. 31 (1998) 181–190;
(d) R.F. Kauffman, H.U. Bryant, Selective estrogen receptor modulators, Drug News Perspect. 8 (1995) 531–539;
(e) G.L. Evans, R.T. Turner, Tissue-selective actions of estrogen analogs, Bone 17 (1995) 181S–190S.

[9] V.C. Jordan, S. Gapstur, M. Morrow, Selective estrogen receptor modulation and reduction in risk of breast cancer, osteoporosis, and coronary heart disease, J. Natl. Cancer Inst. 93 (2001) 1449–1457.

[10] V.C. Jordan, M. Morrow, Raloxifene as a multifunctional medicine? Current trials will show whether it is effective in both osteoporosis and breast cancer, Br. Med. J. 319 (1999) 331–332.

[11] V.C. Jordan, Estrogen, selective estrogen receptor modulation, and coronary heart disease: something or nothing, J. Natl. Cancer Inst. 93 (2001) 2–4.

[12] O.F. Guner, Pharmacophore Perception, Development, and Use in Drug Design, International University Line, California, 2000.

[13] E.F.F. da Cunha, R.C.A. Martins, M.G. Albuquerque, R.B. de Alencastro, LIV-3D-QSAR model for estrogen receptor ligands, J. Mol. Modell. 10 (2004) 297–304.

[14] J.M. Schmidt, J. Mercure, G.B. Tremblay, M. Page, M. Feher, R. Dunn-Dufault, M.G. Peter, P.R. Redden, De novo design, synthesis and evaluation of a non-steroidal diphenylnapthyl propylene ligand for the estrogen receptor, Bioorg. Med. Chem. 11 (2003) 1389–1396.

[15] D.S. Mortensen, A.L. Rodriguez, K.E. Carlson, J. Sun, B.S. Katzenellenbogen, J.A. Katzenellenbogen, Synthesis and biological evaluation of a novel series of furans: Ligands selective for estrogen receptor α, J. Med. Chem. 44 (2001) 3838–3848.

[16] B.R. Henke, T.G. Consler, N. Go, R.L. Hale, D.R. Hohman, S.A. Jones, A.T. Lu, L.B. Moore, J.T. Moore, L.A. Orband-Miller, R.G. Robinett, J. Shearin, P.K. Spearing, E.L. Stewart, P.S. Turnbull, S.L. Weaver, S.P. Williams, G.B. Wisely, M.H. Lambert, A new series of estrogen receptor modulators that display selectivity for estrogen receptor β, J. Med. Chem. 45 (2002) 5492–5505.

[17] G. Zong-Ru, Y. Xiang, X. Zhi-Bin, Molecular simulation of interactions between estrogen receptor and selective estrogen receptor modulators, Acta Pharmacol. Sin. 23 (2002) 208–212.
[18] D. Schuster, C. Laggner, T.M. Steindl, A. Palusczak, R.W. Hartmann, T. Langer, Pharmacophore modeling and in silico screening for new p450 19 (aromatase) inhibitors, J. Chem. Inf. Model. 46 (2006) 1301–1311.
[19] M.L. Lopez-Rodrýguez, B. Benhamu, T. de la Fuente, A. Sanz, L. Pardo, M. Campillo, A three-dimensional pharmacophore model for 5-hydroxytryptamine6 (5- HT6) receptor antagonists, J. Med. Chem. 48 (2005) 4216–4219.
[20] M.O. Taha, A.M. Qandil, D.D. Zaki, M.A. Aidamen, Ligand-based assessment of factor Xa binding site flexibility via elaborate pharmacophore exploration and genetic algorithm based QSAR modeling, Eur. J. Med. Chem. 40 (2005) 701–727.
[21] Catalyst, version 4.11, Accelrys, Inc., San Diego, CA, 2005.
[22] D. Barnum, J. Greene, A. Smellie, P. Sprague, Identification of common functional configurations among molecules, J. Chem. Inf. Comput. Sci. 36 (1996) 563–571.
[23] G. Jones, P. Willett, R.C. Glen, A genetic algorithm for flexible molecular overlay and pharmacophore elucidation, J. Comput. Aided Mol. Des. 9 (1995) 532–549.
[24] Y.C. Martin, M.G. Bures, E.A. Danaher, J. DeLazzer, I. Lico, P.A. Pavlik, A fast new approach to pharmacophore mapping and its application to dopaminergic and benzodiazepine agonists, J. Comput. Aided Mol. Des. 7 (1993) 83–102.
[25] A.G. Timothy, C. Stephen, R.F. Don, G.G. Alexander, D.J. Charles, W.L. Charles, J.M. Michael, M. Ken, D.P. Lewis, A.W. Mark, M.D. Adrian, W.C. Harlan, E.M. David, D.L. Phillips, R.R. Ellen, L.S. Lorri, L.G. Andrew, U.B. Henry, Structure–activity relationships of selective estrogen receptor modulators: modifications to the 2-arylbenzothiophene core of raloxifene, J. Med. Chem. 40 (1997) 146–167.
[26] H.V. de Waterbeemd, B. Testa, in: B. Testa (Ed.), Advances in Drug Research, vol. 16, Academic Press, New York, 1987, pp. 85–225.
[27] P.W. Atkins, Physical Chemistry, Freeman, New York, 1982.
[28] S.A. Wildman, G.M. Crippen, Prediction of physicochemical parameters by atomic contributions, J. Chem. Inf. Comput. Sci. 39 (1999) 868–873.
[29] C.J. Williford, E.P. Stevens, Strain energies as a steric descriptor in QSAR calculations, QSAR Combinatorial Sci. 23 (2004) 495–505.
[30] B. Wang, G. Ford, Atomic charges derived from a fast and accurate method for electrostatic potentials based on modified AM1 calculations, J. Comp. Chem. 15 (1994) 200–207.
[31] R. McWeeny, Coulson's Valence, Oxford University Press, London, 1979
[32] M.J.S. Dewar, The Molecular Orbital Theory of Organic Chemistry, McGraw-Hill, New York, 1969.
[33] J.E. House, Fundamentals of Quantum Chemistry, Elsevier Academic Press, California, 2005.
[34] L.B. Kier, Methods and principles in medicinal chemistry, in: H.V. de Waterbeemd (Ed.), Chemometric Methods in Molecular Design, vol. 2, VCH Publishers, Weinheim, 1995, pp. 39–44.
[35] ETSA-CS, Cyber-Mate, Srirampur, W. Bengal, India, 2003.
[36] TSAR 3D, version 3.3, Oxford Molecular Limited, Accelrys, Inc., San Diego, CA.
[37] CAChe, version 6.1, Life Science and Material Science Division, Fujitsu Ltd., Chiba, Japan.
[38] CS Chem3D Pro and CS MOPAC Pro, version 5.0, Cambridge Soft, Corporation, Cambridge, USA.
[39] Statistica, version 5.0, StatSoft, Inc., Tulsa, USA.
[40] S. Wold, L. Eriksson, Methods and principles in medicinal chemistry, in: H.V. de Waterbeemd (Ed.), Chemometric Methods in Molecular Design, vol. 2, VCH Publishers, Weinheim, 1995, p. 312.
[41] G.W. Snedecor, W.G. Cochran, Statistical Methods, sixth ed., Iowa State University Press, Iowa, 1967.
[42] A.M. Brzozowski, A.C. Pike, Z. Dauter, R.E. Hubbard, T. Bonn, O. Engstrom, L. Ohman, G.L. Greene, J.A. Gustafsson, M. Carlquist, Molecular basis of agonism and antagonism in the oestrogen receptor, Nature 389 (1997) 753–758.
[43] A.C. Pike, A.M. Brzozowski, R.E. Hubbard, T. Bonn, A.G. Thorsell, O. Engstrom, J. Ljunggren, J.A. Gustafsson, M. Carlquist, Structure of the ligand-binding domain of oestrogen receptor beta in the presence of a partial agonist and a full antagonist, EMBO J. 18 (1999) 4608–4618.
[44] T.A. Grese, J.P. Sluka, H.U. Bryant, G.J. Cullian, A.L. Glasebrook, C.D. Jones, K. Matsumoto, A.D. Palkowitz, M. Sato, J.D. Termine, M.A. Winter, M.A. Yang, J.A. Dodge, Molecular determinants of tissue selectivity in estrogen receptor modulators, Proc. Natl. Acad. Sci. 94 (1997) 14105–14110.
[45] T.A. Grese, J.P. Sluka, H.U. Bryant, H.W. Cole, J.R. Kim, D.E. Magee, E.R. Rowley, M. Sato, Benzopyran selective estrogen receptor modulators (SERMs): Pharmacological effects and structural correlation with raloxifene, Bioorg. Med. Chem. Let. 6 (1996) 903–908.
[46] H. Verli, M.G. Albuquerque, R.B. Alencastro, E.J. Barreiro, Local intersection volume: A new 3D descriptor applied to develop a 3D-QSAR pharmacophore model for benzodiazepine receptor ligands, Eur. J. Med. Chem. 37 (2002) 219–229.
[47] R.C.A. Martins, M.G. Albuquerque, R.B. Alencastro, Local intersection volume (LIV) descriptors: 3D-QSAR models for PGI_2 receptor ligands, J. Braz. Chem. Soc. 13 (2002) 816–821.
[48] S. Gauthier, B. Caron, J. Cloutier, Y.L. Dory, A. Favre, D. Larouche, J. Mailhot, C. Ouellet, A. Schwerdtfeger, G. Leblanc, C. Martel, J. Simard, Y. Merand, A. Belanger, C. Labrie, F. Labrie, (*S*)-(+)-4-[7-(2,2-Dimethyl-1-oxopropoxy)-4-methyl-2-[4-[2-(1-piperidinyl)-ethoxy]phenyl]-2H-1-benzopyran-3-yl]-phenyl-2,2-dimethylpropanoate (EM-800): a highly potent, specific, and orally active nonsteroidal antiestrogen, J. Med. Chem. 40 (1997) 2117–2122.
[49] J. Simard, R. Sanchez, D. Poirier, S. Gauthier, S.M. Singh, Y. Merand, A. Belanger, C. Labrie, F. Labile, Blockade of the stimulatory effect of estrogens, OH-tamoxifen, OH-toremifene, droloxifene, and raloxifene on alkaline phosphatase activity by the antiestrogen EM-800 in human endometrial adenocarcinoma Ishikawa cells, Cancer Res. 57 (1997) 3494–3497.
[50] C.P. Miller, M.D. Collini, B.D. Tran, H.A. Harris, Y.P. Kharode, J.T. Marzolf, R.A. Moran, R.A. Henderson, R.H.W. Bender, R.J. Unwalla, L.M. Greenberger, J.P. Yardley, M.A. Abou-Gharbia, C.R. Lyttle, B.S. Komm, Design, synthesis, and preclinical characterization of novel, highly selective indole estrogens, J. Med. Chem. 44 (2001) 1654–1657.