Identification of Structural Requirement of Estrogen Receptor Modulators using Pharmacoinformatics Techniques for Application to Estrogen Therapy

Md Ataul Islam[1#], Darshakkumar Ashokbhai Patel[2#], Savansinh Ghanshyamsinh Rathod[2#], Preeti Chunarkar[3] and Tahir S. Pillay*[1]

[1]*Department of Chemical Pathology, Faculty of Health Sciences, University of Pretoria and National Health Laboratory Service Tshwane Academic Division, Pretoria, South Africa.*
[2]*G.H.Patel Post Graduate Department of Computer Science and Technology (GDCST), Sardar Patel University, Anand, Gujarat, India.*
[3]*Department of Bioinformatics, Rajiv Gandhi Institute of Information Technology and Biotechnology, Bharati Vidyapeeth Deemed University, Pune-Satara Road, Pune, India.*

Correspondence should be addressed to T.S. Pillay, Department of Chemical Pathology, Faculty of Health Sciences, University of Pretoria, Private Bag X323, Arcadia, Pretoria, 0007
Email: tspillay@gmail.com
Phone: +27-123192114
Fax: +27-123283600

[#]Authors contributed equally

Abstract

An attempt was made in the present study to explore the structural requirements of known estrogen receptor (ER) modulators for biological activity using pharmacoinformatics approaches to elucidate critical functionalities for new, potent and less toxic chemical agents for successful application in estrogen therapy. For this purpose a group of non-steroidal ligands, 7-thiabicyclo[2.2.1]hept-2-ene-7-oxide derivatives were collected from the literature to perform quantitative structure-activity relationship (QSAR), pharmacophore and molecular docking studies. The 2D QSAR models (R^2_α = 0.857, se_α = 0.370, Q^2_α = 0.848, $R^2_{pred\text{-}\alpha}$ = 0.675, sp_α = 0.537; R^2_β = 0.874, se_β = 0.261, Q^2_β = 0.859, $R^2_{pred\text{-}\beta}$ = 0.659, sp_β = 0.408) explained that hydrophobicity and molar refractivity were crucial for binding affinity in both α- and β-subtypes. The space modeling study (R^2_α = 0.955, se_α = 1.311, Q^2_α = 0.932, $R^2_{pred\text{-}\alpha}$ = 0.737, sp_α = 0.497; R^2_β = 0.885, se_β = 1.328, Q^2_β = 0.878, $R^2_{pred\text{-}\beta}$ = 0.769, sp_β = 0.336) revealed the importance of HB donor and hydrophobic features for both subtypes, whereas, HB acceptor and aromatic ring were critical for α- and β-subtypes respectively. The functionalities developed in the QSAR and pharmacophore studies were substantiated by molecular docking which provided the preferred orientation of ligands for effective interaction at the active site cavity.

Key words: Estrogen receptor, SERMs, QSAR, Pharmacophore, Molecular Docking

Introduction

Estrogens are sex steroid hormones, secreted by the ovaries and testis, and control a number of physiological actions including the development of female reproductive system and secondary

characteristics, neuro-endocrine actions involved in the control of ovulation, the cyclic preparation of the reproductive tract for fertilization and implantation, and major actions on mineral, carbohydrate, protein and lipid metabolism(Lewis and Jordan 2005). These hormones also demonstrate remarkable effectiveness in deterrence and management of pre- and post-menopausal diseases(Rossouw et al. 2002; Yaffe et al. 1998). Hormone replacement therapy (HRT) is well known and the most common uses of estrogen agonists in which synthetic estrogens are administered to reduce osteoporotic fractures and improve menopausal symptoms(Rossouw et al. 2002). However HRT increases the risk of breast and uterus cancers(Chlebowski et al. 2003; Gehrig et al. 2004). The most common uses of estrogen agonists are in HRT and contraception, while estrogen antagonists are used in the treatment of hormone-responsive breast cancer and female infertility(Clarke et al. 2003).

The biochemical basis of estrogen actions is thought be primarily via the regulation of gene expression after binding to estrogen receptor (ER), which belongs to the nuclear receptor superfamily. It is often overexpressed in the tissue of breast cancer patients, which promotes the estrogen-dependent proliferation of cancer cells(Doisneau-Sixou et al. 2003; Foster et al. 2001a; Foster et al. 2001b; Holst et al. 2007). The nuclear receptor family members share a conserved structural architecture consisting of six structural domains, including ligand binding domain (LBD). The orientation of helix 12, located at the carboxy-terminus of the LBD is fundamental in distinguishing between agonists and antagonists(Pike et al. 2000). There are two subtypes of ER, namely ERα and ERβ that quite similar in overall structure. ERα is expressed predominantly in breast and uterine tissues, whereas ERβ is found chiefly in the brain, bone and vascular epithelium(Gustafsson 1999).

Selective estrogen receptor modulators (SERMs) are structurally diverse molecules which exhibit a unique pharmacological profile. Depending upon the target tissues SERMs are characterized by estrogen agonist action in some tissues but as estrogen antagonists in others(Maximov et al. 2013; Riggs and Hartmann 2003). The agonist and antagonist properties of SERMs arise from differentially expressed ERs due to ligand-dependent receptor conformational deviations, interactions with various co-activators and co-repressors and following changes in gene transcription(Pickar et al. 2010). Tamoxifen, raloxifene, toremifene etc. are potent SERMs that have been used to treat breast cancer and osteoporosis(Chmel et al. 2002; Fisher et al. 2005; Fisher et al. 1998). Successful SERMs are classified in generation, suggesting a progressive development in a process intended to improve the beneficial effects while reducing the harmful side effects associated with the earlier SERMs(Dowers et al. 2006).

Use of computational approaches in the pharmaceutical industry is a very popular and useful pharmacoinformatics application to search for potent and safe lead molecules. Traditional methods of drug discovery is a complex process and takes about 15 years of time and estimated to cost $0.8 to $1 billion to launch a drug into the market(Dalkas et al. 2012). It is reported that out of 5000 lead compounds, only 5 may enter into the preclinical phase and then to approval for human testing ultimately(Dalkas et al. 2012). This clearly indicates that the traditional approaches are cost-intensive and time-consuming, necessitating alternative approaches to save money, time and effort. The use of pharmacoinformatics techniques in drug discovery and development is rapidly gaining in popularity, implementation and appreciation(Kapetanovic 2008). The proper application of this technique can develop molecules with optimistic efficacy. Keeping in mind the objective of immense utility of SERMs

for the treatment of post-menopausal diseases, researchers in academia as well as in industry are dedicated to the development of synthetic therapeutic SERMs for estrogen therapy(Brogia et al. 2013; Chang et al. 2013; Lewis et al. 1995; Mukherjee et al. 2005; Smith et al. 2007; Zhang et al. 2013). Pharmacoinformatics techniques, ligand-based approaches such as 2D/3D quantitative structure activity relationship (QSAR) and pharmacophore mapping, and structure-based molecular docking method have become vital tools(Verma et al. 2010) for this purpose.

In the present study, a series of 7-thiabicyclo[2.2.1]hept-2-ene-7-oxide derivatives(Wang et al. 2012), reported as promising SERMs(Wang et al. 2012) for estrogen therapy were evaluated with their relative binding affinity for pharmacoinformatics studies to explore the physicochemical properties and 3D structural conformation of chemical moieties for selective estrogenic actions, through both 2D QSAR and pharmacophore studies. The QSAR modelling was carried out to obtained statistical validated models useful in exploring structure activity relationship (SAR) of the compounds, while pharmacophore concept was based on the type of interactions observed in the molecular recognition and alternatively can be used as a query in a 3D database search to identify new structural classes of potential lead compounds and also can serve as a template for generating alignment for 3D QSAR analysis.

Materials and methods

Pharmacoinformatics models were derived from a congeneric dataset of 7-thiabicyclo[2.2.1]hept-2-ene-7-oxide derivatives(Wang et al. 2012) (Table 1) to explore QSAR and pharmacophore modeling studies to determine pharmacophoric features of the small molecule required for binding affinity of ER subtypes. The most active compounds of both α- and β-subtypes are depicted in the Fig. 1. The relative binding affinity (*RBA*) of the of the dataset was converted into *kRBA* = log(1000x*RBA*) for QSAR study; while for pharmacophore modeling study it was expressed as *pRBA* = 1/*RBA*. The primary goal of the study was to generate a statistical relationship between the structure and corresponding activity, and deduce a pharmacophore map through receptor-independent space modeling techniques. Furthermore the functionalities derived in the receptor-independent studies were correlated with binding interactions observed between most active compounds of both α- and β-subtypes in the active site cavity of ER. The dataset (Table 1) was randomly divided into training set (n_{tr}=22) and test set (n_{ts}=10) for model generation and validation of generated models respectively.

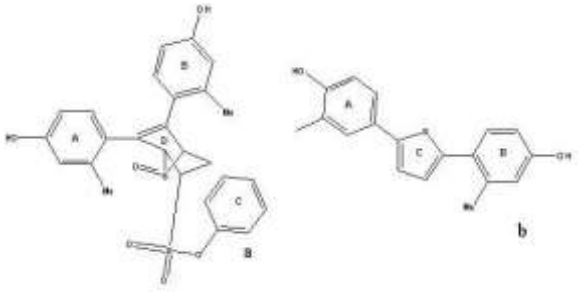
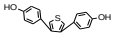


Fig. 1 Most active compounds of the dataset a) α-subtype and b) β-subtype

The different statistical parameters considered to judge the models were correlation coefficient (R^2), standard error of estimation (*se*), Leave-One Out (LOO) cross-validated correlation coefficient (Q^2) and modified correlation coefficient (r^2_m). Additionally, the variance ratio (*F*) with degree of freedom (*df*) and

explained variance (*EV*) for 2D-QSAR study were also considered, while different cost factors(Islam et al. 2008) were used for space modeling study. In order to evaluate the predictive power of the model, R^2_{pred} (correlation coefficient), s_p (error of prediction) and modified correlation coefficient ($r^2_{m(test)}$) of the test set were also determined in both studies.

Table 1 Chemical structures and relative binding affinity (*RBA*) of 7-thiabicyclo[2.2.1]hept-2-ene-7-oxide derivatives

SN	Struct.	*RBA*		SN	Struct.	*RBA*	
		ER_α	**ER_β**			**ER_α**	**ER_β**
C1		0.956	0.110	**C17**		0.016	0.014
C2		0.074	0.080	**C18**		3.49	0.014
C3		0.077	0.071	**C19**		0.310	0.021
C4		0.017	0.013	**C20**		2.48	0.010
C5		0.022	0.048	**C21**		1.18	0.022
C6		8.110	0.348	**C22**		0.516	0.014
C7		0.741	0.091	**C23**		0.236	0.019
C8		3.530	0.138	**C24**		0.109	0.009
C9		2.490	0.227	**C25**		0.009	0.007
C10		2.210	0.070	**C26**		0.026	0.019
C11		2.180	0.080	**C27**		0.005	0.008
C12		2.130	0.083	**C28**		0.005	0.011
C13		3.370	0.297	**C29**		0.035	0.055
C14		1.300	0.088	**C30**		0.572	1.690
C15		0.998	0.180	**C31**		2.16	4.860
C16		2.210	0.812	**C32**		0.793	0.306

QSAR study

QSAR is the mathematical relationship and statistically validated model which attempts to find a statistically significant correlation between chemical structure and biological activity using chemometric techniques. In drug discovery and design research, structure denotes the properties or descriptors of the molecules, their substituents or interaction energy fields, while biological activity refers to an experimental biological/biochemical endpoint such as binding affinity, toxicity or rate constants, while chemometric methods include MLR, PLS, PCA, PCR, ANN, GA *etc*. Various QSAR approaches have been developed gradually over a time span of more than a hundred years and served as a valuable

predictive tool, particularly in the design of pharmaceuticals and agrochemicals. The methods have evolved from Hansch and Free-Wilson's(Kubinyi 2004) one or two-dimensional linear free-energy relationships, via Crammer's three-dimensional QSAR(Cramer et al. 1988) to Hopfinger's fourth(Hopfinger and Tokarski 1997) and Vedani's fifth and sixth-dimensions(Vedani and Dobler 2002). One and two dimensional and related methods are commonly referred to as 'classical' QSAR methodologies. Irrespective of the type, all QSAR formalisms presume that every molecule included in the study binds to the same site of the same target receptor. However, the main difference between all these formalisms reside in the manner in which each one of them treats and represents structural properties of the molecules and extracts the quantitative relationships between the properties and activities.

The molecular descriptors were generated using molecular orbital environment (MOE) tool developed by Chemical Computing Group(MOE 2007)(MOE, 2007)(MOE, 2007)[35, 36][35, 36]. The compounds of the dataset were minimized using MMFF94 force field. The minimization algorithm used was the steepest descent followed by conjugate gradient method until it reached RMS gradient of 0.001 kcal/mol/Å. The QuaSAR module of MOE was considered for descriptors generation. More than 210 descriptors were obtained in this series of compound for QSAR study. All descriptors calculated in the MOE could be categorized into three classes. The 2D descriptors only based on atoms and connection information of the molecule for the calculation, i3D descriptors use 3D coordinate information about each molecule; however, they are invariant to rotations and translations of the conformation, whereas x3D descriptors known as external 3D descriptors, based on 3D coordinate information but also require an absolute frame of reference.

In order to obtain robust and good predictive models by predominant descriptors affecting binding affinity of the training set compounds, standard and forward stepwise methods of Statistica (http://www.statsoft.com) was used, considering descriptors as independent variable and *kRBA* as dependent variable. The descriptors with constant or near constant values were deleted and also in order to minimize redundant information auto-correlated (>0.5) descriptors were not considered in the model generation.

Pharmacophore space modeling study

The *HypoGen* algorithm in Discovery Studio(Accelrys 2013) was used to generate pharmacophore model for both α- and β-subtypes of ER ligands. Pharmacophore hypothesis can visualize potential interactions between ligand and active site of the receptor molecule. A pharmacophore is a set of functional group or fragment type in a spatial arrangement that represent the interaction made in common by a set of small molecular ligands with a protein receptor. The conformational models were generated for each ligand to ensure good coverage of conformational space within the minimum number of conformers.

The pharmacophore concept is based on the kinds of interaction observed in molecular recognition, *i.e.*, hydrogen bonding, charge, and hydrophobic interaction. In order to obtain the pharmacophore hypothesis, conformational models were generated for each ligand using algorithm, out of two, *i.e.* fast and best(Kristam et al. 2005). This was followed by the conformer generation, an algorithm which also

considers chemical features and conformers, and operates in two modes: *HipHop* and *HypoGen*. The *HipHop* generates pharmacophore models using active compounds only, whereas the *HypoGen* takes activity data into account and uses both active and inactive compounds in an attempt to identify a hypothesis that is common among the active compounds but not in the inactive compounds(Kristam et al. 2005). It builds top ten scoring hypothesis models with considering the training set, conformational models and chemical features by three steps: a constructive step, a subtractive step and an optimization step(Sadler et al. 1998). The constructive step generates hypotheses that are common among the most active compounds. In subtractive step, the hypotheses that fit to the inactive compounds are removed. Finally, the optimization step attempts to improve the score of the remaining hypotheses by applying small perturbation(Kristam et al. 2005; Li et al. 2000).

Validation of QSAR and pharmacophore models

Validation is an essential step of any pharmacoinformatics model to judge the predictivity and applicability as well as robustness. In the current study, both QSAR and pharmacophore models were validated by internal validation and test set prediction. Further the pharmacophore models were validated by cost function analysis and Fischer's randomization test.

Internal validation

The best models of both studies were validated internally using leave-one out (LOO) cross-validation method. In this procedure, randomly one compound was deleted from training set in each cycle and model regenerated using rest of the compounds with same parameters used in original model. The new generated model was used to predict the activity of deleted compound. The procedure was continued until all molecules of the training set deleted and activity predicted. The Q^2 and *se* were calculated based on predicted activity of training compounds. High Q^2 (>0.5) and low *se* (<0.5) explained better predictive ability of the model(Kubinyi et al. 1998). Further to confirm the good predictive ability of the training set compounds, r^2_m developed by Roy *et al.*(Ojha et al. 2011; Roy et al. 2012) was calculated. The r^2_m is a measure of the degree of deviation of the predicted activity from the observed ones. It is reported that model may be considered with r^2_m>0.5. Another parameter, ($[(R^2-R_0^2)/R^2]$ suggested by Golbraikh and Tropsha(Golbraikh and Tropsha 2002) were also calculated. It is reported that the value of the parameter less than 0.1 explained good predictive ability and robustness of the model. The R^2 and R^2_0 are the correlation coefficient between predicted versus observed activities for regressions with intercept and without intercept (through the origin) respectively.

Test set prediction

The prediction of test set compounds is a crucial step in order to verify whether selected model was able to accurately predict the activities of compound beyond the training set molecule. External validation provides the ultimate proof of the true predictivity of the model, and the predictive capacity of the model was judged best on statistical parameters, R^2_{pred} and s_p. The threshold value of R^2_{pred} is ≥0.5, whereas for s_p it is ≤0.5(Golbraikh and Tropsha 2002; Mitra et al. 2010). The value of R^2_{pred} depends on the mean observed activity of the training set compounds. Consequently high values of this parameter may be obtained for compounds with a wide range of activity data, but this may not indicate that the predicted activity values are very close to those observed. Though a good overall correlation is maintained, there may be a significant numerical difference between the two values. To better indicate the predictive ability of the model, modified r^2 [$r^2_{m(test)}$](Roy et al. 2009; Roy and Roy 2008) values were calculated and threshold value is reported as 0.5.

Cost function analysis

The statistical parameters employed for hypothesis generation were spacing, uncertainty, and weight variation. Spacing is a parameter representing the minimum inter-features distance that may be allowed in the resulting hypothesis. The weight variation is the level of magnitude explored by the hypothesis where each feature signifies some degree of magnitude of the compound's activity. 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 1.5 to 3.0 were considered as the uncertainty parameter. The total cost function is minimized encompassing three terms, *viz.*, weight cost, error cost, and configuration cost. Weight cost is value that increases as the weight variation in the model deviates from input weight variation value. The deviation between the estimated activity of the molecule in the training set and their

experimentally determined value is the error cost. A fixed 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 hypothesis space and should have a value <17 for a good pharmacophore model. In the selected models of both α- and β-subtypes the configuration cost was found less than 15. The HypoGen algorithm also calculates the cost of a null hypothesis that assumes no relationship in the data, and the experimental activities are normally distributed about their mean. Accordingly, the greater the difference (Δ*cost*) between the total and the null costs, it is more likely that the hypothesis does not reflect a chance correlation.

Fischer's randomization test

The Fischer's randomization test was used to ensure strong correlation between the chemical structures and the biological activity of the training set molecule. In this method, the biological activity was scrambled and assigning them new values. After that the pharmacophore hypotheses were generated using the same features and parameters as those used to develop the original pharmacophore hypotheses. If the randomization run generates better correlation coefficient and/or better statistical parameters than the original hypothesis may be considered to be developed by chance. Depending upon the statistical significance randomization run produces different number of spreadsheet. The statistical significance is given by following equation.

$$Significance = [1-(1+a)/b] \quad (1)$$

Where, *a* defined as total number of hypotheses having a total cost lower than best significant hypothesis, whereas *b* denoted by total number of *HypoGen* runs and random runs. In case of 95% confidence level 19 random spreadsheet are generated (b = 20) and every generated spreadsheet is submitted to *HypoGen* using the same experimental conditions as the initial run. In the present study, the developed pharmacophore model was checked at 95% confidence level that produced 19 spreadsheets.

Molecular docking

The molecular docking study was performed by using *LigandFit* of receptor-ligand interactions protocol in Discovery Studio(Accelrys 2013) after preparing ligand and receptor. For ligand preparation, all the duplicate structures were removed and ionization change, tautomer generation, isomer generation, and Lipinski filter were set to false. In order to prepare receptor, the hydrogen atoms were added to the protein molecules (PDB ID: 3ERT(Shiau et al. 1998) and 2QTU(Richardson et al. 2007) for α- and β-subtypes respectively), and water molecules were removed. The pH of the protein was set in the range of 6.5–8.5. The active site was selected based on the ligand binding domain of the bound ligand. *LigandFit* was chosen for receptor-ligand interaction and PLP1 was selected as the energy grid. The conformational search of the ligand poses was performed by Monte Carlo method with the torsional step size for polar hydrogen set to ten. The docking was performed with consideration of electrostatic energy with maximum internal energy 10,000 Cal. But no attempt was made to minimize the ligand-receptor complex (rigid docking). After completion of docking, the docked enzyme (protein–ligand complex) was analysed to investigate the type of interactions. Docking conformers of each compound were ranked according to their dock score function.

Results and Discussion

QSAR

The QSAR study was performed with the training set molecules of both α- and β-subtypes to obtain statistical relationships between molecular descriptors and binding affinity. The best QSAR models was developed with the contribution of *logP(o/w)*, *AM1_LUMO*, *balabanJ* (Table 2) and *SMR_VSA1* for α-subtype, whereas, *E_stain*, *AM1_dipol*, *GCUT-SLOGP_0* and *GCUT_SMR_1* (Table 2) for β-subtype. The predicted activities based on the both models are shown in the Table 3 and Fig. 4. In order to improve the models auto-correlated (|r|>0.5) descriptors were not considered in the model generation. Stepwise multiple linear regression method of Statistica was adopted to reach in the final model.

α-subtype

The best statistical model was developed with combination of *logP(o/w)*, *AM1_LUMO*, *balabanJ* and *SMR_VSA1* descriptors as below.

$$kRBA_{\alpha} = 4.687(\pm 0.992)balabanJ + 0.869(\pm 0.117)\log P(o/w) - 0.591(\pm 0.100)AM1\_LUMO + 0.039(\pm 0.007)SMR\_VSA1 - 10.843(\pm 1.880)$$

n_{tr} = 22, R^2 = 0.857, *se* = 0.370, Q^2 = 0.848, *EV* = 81.32%, *F(df)* = 23.860(4,17), n_{ts} = 10, R^2_{pred} = 0.675, s_p = 0.537, p<0.0001

The most important descriptor found in the best model of α-subtype was *balabanJ*, which is the connectivity topological index and a good descriptor for the shape of the molecule. Its large and positive coefficient in the model indicated that a bigger size and high branching of molecules might be favourable for binding affinity of ER ligands. Despite the fact that examples of QSAR studies based on Balaban index are rare in literature, a few recent studies (Balaban et al. 2005; Kim et al. 2002; Rastija and Medic-Saric 2009) and the present study showed that this index can be successfully used for this purpose. The coefficient of this parameter was found as 4.687 that indicates the superiority of the descriptor in the model. The second most important descriptor was found in the model, logP(o/w) with positive coefficient (0.869). The logP(o/w) is the logarithm value of ratio of the concentration of a chemical in n-octanol to that in water in a two-phase system at equilibrium. This coefficient has been shown to be one of the key parameters in QSAR/QSPR studies(Gupta and Prabhakar 2008; Platts et al. 2006; Saxena et al. 2003) and measure the hydrophobicity and hydrophilicity of a substance. Its positive regression coefficient in the model suggested that high hydrophobicity of the molecule might be favourable for their ability for membrane penetration and distribution into the organisms along with enhanced hydrophobic interactions at the active site cavity of ER. The third most important parameter found in the model was *AM1_LUMO*, the energy of the lowest unoccupied molecular orbital (*LUMO*) calculated using AM1 Hamiltonian. This descriptor reflects electrophilic reactivity and plays a important role in reductive processes in which an electron from a molecular donor anion or metallic surface is assumed to transfer into this orbital on an acceptor molecule(Burrow and Modelli 2013). The *orbital energy* has been successfully used in the development of QSAR models(Kupcewicz et al. 2013; Levet et al. 2013; Nandy et al. 2013; Nantasenamat et al. 2013). The negative coefficient of this descriptor in the model suggested that highly nucleophile compounds resulted in high binding affinity and might influence binding at the ERα subtype. The least important parameter among four descriptors in the model was *SMR_VSA1* i.e. molecular

refractivity (including implicit hydrogen). This property is basically sums the approximate exposed surface based on molar refractivity. The positive contribution of this parameter suggested that high molar refractivity value of the molecule was crucial for binding affinity.

Table 2 Descriptors used in the QSAR study

Mol. No.	***Descriptors***							
	[1]AMd	[2]AML	[3]BJ	[4]ES	[5]GS1	[6]GS2	[7]LP	[8]SV1
C1	6.680	-0.994	1.376	295.096	-0.216	-0.896	3.484	59.316
C2	7.501	-0.935	1.340	309.663	-0.216	-0.896	3.440	70.316
C3	7.171	-1.070	1.360	292.860	-0.216	-0.896	3.176	84.701
C4	5.489	-0.801	1.323	239.501	-0.216	-1.015	2.332	50.770
C5	7.173	-0.842	1.687	268.772	-0.215	-0.870	2.075	50.770
C6	3.215	-3.707	1.426	441.240	-0.188	-0.896	4.150	59.316
C7	7.535	-0.958	1.388	449.356	-0.189	-0.896	4.106	70.316
C8	6.381	-1.030	1.430	437.240	-0.188	-0.896	4.301	74.276
C9	4.151	-1.068	1.430	438.260	-0.183	-0.896	4.740	59.316
C10	5.917	-1.044	1.430	460.863	-0.183	-0.896	4.946	59.316
C11	6.164	-1.082	1.409	441.713	-0.188	-0.896	4.303	74.276
C12	5.929	-1.101	1.409	441.992	-0.188	-0.896	4.742	59.316
C13	6.109	-1.139	1.409	439.850	-0.188	-0.896	4.948	59.316
C14	5.743	-1.081	1.419	440.412	-0.188	-0.896	4.779	59.316
C15	5.249	-1.134	1.268	486.052	-0.188	-0.896	5.370	59.316
C16	5.263	-0.976	1.409	446.759	-0.207	-0.896	3.842	84.701
C17	5.073	-0.789	1.361	406.587	-0.216	-1.015	2.998	50.770
C18	5.351	-0.925	1.403	311.546	-0.177	-0.896	4.150	59.316
C19	6.395	-1.049	1.368	327.378	-0.183	-0.896	4.106	70.316
C20	6.394	-1.035	1.408	319.570	-0.177	-0.896	4.301	74.276
C21	3.734	-1.408	1.408	430.118	-0.176	-0.896	4.740	59.316
C22	7.381	-1.029	1.408	439.616	-0.177	-0.896	4.946	59.316
C23	5.034	-1.340	1.251	347.809	-0.183	-0.896	5.370	59.316
C24	6.414	-1.006	1.397	314.299	-0.187	-0.896	3.879	84.701
C25	5.836	-0.832	1.340	293.915	-0.216	-1.015	2.998	50.770
C26	5.570	-0.046	1.719	403.824	-0.160	-0.870	2.741	50.770
C27	6.363	-0.927	1.362	279.400	-0.216	-0.896	3.792	33.931
C28	7.162	-0.901	1.360	302.521	-0.216	-0.896	4.388	33.931
C29	6.138	-0.959	1.368	297.132	-0.216	-0.896	3.817	59.316
C30	0.628	-0.695	1.480	73.202	-0.216	-0.762	4.619	50.770
C31	0.655	-0.563	1.541	91.075	-0.168	-0.762	5.285	50.770
C32	0.646	-0.713	1.521	89.572	-0.147	-0.762	5.285	50.770

[1]AM1_dipole, [2]AM_LUMO, [3]BalabanJ, [4]E_Strain, [5] *GCUT_SMR_1,* [6] *GCUT_SLOGP_0,* [7] *logP(o/w),* [8]*SMR_VSA_1*

β-subtype

The best model of β-subtype was derived with importance of four parameters, *GCUT_SMA_1*, *GCUT_SLOGP_0*, *AM1_dipol* and *E-strain* (Table 2).

$$kRBA_{\beta} = -25.600(\pm 5.804)GCUT\_SMR\_1 + 13.544(\pm 2.029)GCUT\_SLOGP\_0 - 0.341(\pm 0.059)AM1\_dipol + 0.005(\pm 0.001)E\_strain + 9.226(\pm 1.420)$$

n_{tr} = 22, R^2 = 0.874, *se* = 0.261, Q^2 = 0.859, *EV* = 82.87%, *F(df)* = 20.573 (4, 17), n_{ts} = 10, R^2_{pred} = 0.659, s_p = 0.408, p<0.0001

Table 3 Observed and predicted binding affinity, and fit score of 7-thiabicyclo[2.2.1]hept-2-ene-7-oxide derivatives

S.N.	Relative Binding Affinity (*kRBA*)						Fit Score	
	Observed		Predicted					
			QSAR		Pharmacophore			
	α	***β***	***α***	***β***	***α***	***β***	***α***	***β***
C1	2.980*	2.041	2.525	1.788	2.036	1.556	12.512	4.080
C2	1.869*	1.903	1.711	1.579	1.041	1.690	12.517	4.124
C3	1.886	1.851*	2.013	1.610	1.799	1.625	10.275	5.148
C4	1.231*	1.114*	1.170	1.314	0.927	1.039	7.650	2.485
C5	1.342	1.681	1.338	1.801	0.942	1.549	9.418	4.072
C6	3.909	2.542*	3.941	2.966	3.935	2.867	12.412	6.391
C7	2.870*	1.959	2.528	1.546	2.840	2.073	15.316	4.596
C8	3.548	2.140	3.090	1.868	3.665	2.282	12.141	4.805
C9	3.396*	2.356*	2.914	2.513	3.897	2.288	15.373	4.812
C10	3.344*	1.845	3.079	2.009	3.750	2.083	15.226	6.307
C11	3.339*	1.903	3.025	1.964	3.082	1.963	15.458	4.487
C12	3.328*	1.919	2.837	2.037	3.784	1.911	15.260	4.435
C13	3.528	2.473	3.071	1.965	3.469	2.409	11.945	4.932
C14	3.114	1.944*	2.945	2.092	3.285	1.898	11.861	6.421
C15	2.999	2.255	2.741	2.400	3.201	2.546	11.677	5.070
C16	3.344	2.910	2.968	2.769	3.439	2.491	11.915	4.915
C17	1.204	1.146	0.581	1.272	1.158	1.256	9.634	3.779
C18	3.543*	1.146	2.190	1.322	3.281	0.982	12.758	3.505
C19	2.491	1.322	2.490	1.186	2.777	1.439	11.254	3.963
C20	3.394	1.000	2.990	1.007	2.987	1.319	11.463	3.842
C21	3.072	1.342*	3.011	1.435	2.995	1.614	10.971	5.138
C22	2.713	1.146	2.966	1.242	2.417	1.297	10.894	3.820
C23	2.373	1.279	2.783	1.751	2.584	0.988	11.060	3.512
C24	2.037	0.954	2.962	1.219	2.436	1.446	9.536	3.970
C25	0.954*	0.845	0.707	0.962	0.758	0.763	12.234	3.286
C26	1.415	1.279*	1.594	1.727	1.442	1.583	9.918	5.107
C27	0.699	0.903*	0.702	1.819	0.856	0.572	9.332	5.096
C28	0.699	1.041	1.195	1.660	0.892	1.137	9.368	3.661
C29	1.544	1.740	1.757	1.982	1.446	1.582	10.422	4.106
C30	2.757	3.228*	2.490	3.583	2.797	2.435	11.274	4.956
C31	3.334	3.687	3.279	3.435	3.247	3.773	11.523	6.297
C32	2.899	2.486*	3.272	2.891	2.928	2.461	11.404	4.985

*Test compounds

The most significant parameter was found in the model, *GCUT_SMR_1*, which is atomic charge contribution to the molar refractivity instead of partial charges. The negative contribution of the descriptor in the model indicated that less atomic refractivity might be influential for the binding affinity. The second most significant descriptor found was *GCUT_SLOGP_0*, which used atomic contribution to *logP* (using the Wildman and Crippen *SlogP* method(Wildman and Crippen 1999)) instead of partial charge. Positive coefficient of this descriptor suggested that high hydrophobicity value enhance the binding affinity as well as strengthen the interaction at the active site cavity. Dipole moment calculated using AM1 Hamiltonian was found crucial for the binding affinity towards β-subtype. This is the magnitude of dipole moments, indicating the strength and orientation behaviour of a molecule in an electrostatic field and it also behaves as a marker for chemical reactivity. The negative contribution of this descriptor indicated that a decrease in the charge density of the molecule might be crucial for binding affinity. The least significant parameter for the model of β-subtype was found to be *E_strain*. This descriptor is the current energy minus the value of the energy at a near local minimum. The current energy is calculated as for the E-descriptor. The local minima energy is the value of the E-descriptor after first performing an energy minimization. The model suggested that high *E_train* value of the molecule might be favourable for binding interaction towards ERβ.

Pharmacophore

The pharmacophore models was derived from training set (n_{tr}=22) molecules, whereas test set (n_{ts}=10) molecules was used to check the predictive ability of the model. *Hypogen* algorithm generates ten alternative pharmacophore hypotheses on each run describing the ER binding affinity of training set molecules. The statistical results of the pharmacophore study of both α- and β-subtypes are given in Table 4. The best model in both subtypes was adjudged in terms of mapped features, cost difference (Δ_{cost}), R^2 and *se* of the generated hypothesis. The predicted activities based on both models are shown in Table 3 and Fig. 4.

α-subtype

The best model of α-subtype (*Hypo 1*, Run Number 14 in Table 4) was mapped with most active compound of the dataset and shown in the Fig. 2. The selected model consists of the spatial arrangement of HB acceptor and aromatic ring along with hydrophobicity pharmacophoric features for binding affinity of the training set molecule. The null cost of all ten hypotheses of run number 14 (Table 4) was found to be 449.103 whereas the cost difference was 365.110. The configuration cost of all ten hypotheses of selected run was found to be 15.231 that explained the suitability of the model as it was reported that configuration cost must be less than 17 for attest a good pharmacophore model(Li et al. 2000). *Hypo 1* in run number 14 (Table 4) showed low total cost of 83.993, high cost difference of 365.110, less *se* value (1.311) and high correlation coefficient of 0.955 between observed and predicted binding affinity of the training set molecules.

The best model of the α-subtype explained that oxygen atom of sulfonic group attached to ring 'C' (Fig. 1a) was crucial for HB acceptor. Rings 'A' and 'C' (Fig. 1a) along with alkyl group attached to ring 'B' (Fig. 1a) were found to be important for imparting hydrophobicity of the molecule. The hydroxyl group attached to ring 'B' was revealed as HB donor and crucial for hydrogen bond interactions with key amino residues at the active site of ERα

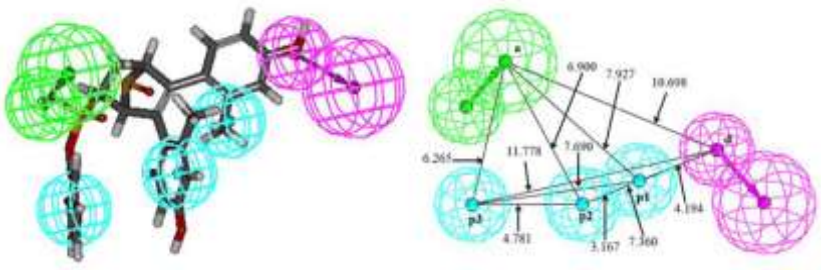


Fig. 2 Mapped with most active compound in training set of α-subtype; mapped features: HB acceptor (*a*), HB donor (*d*) and Hydrophobic (*p*).

β-subtype

The best model (*Hypo 1* of Run number 23 in Table 4) in β-subtype was derived with importance of HB donor, hydrophobicity and aromatic ring features. The final model was selected after successive variation of input parameters viz. spacing, uncertainty and weight variation keeping rest parameters as default. It was observed that at spacing 250, uncertainty 1.5 and weight variation 0.5 the hypothesis given maximum R^2 (0.885), less *se* (1.328) and high cost difference (Δ_{cost} = 223.170) values. The mapped pharmacophoric features and inter-feature distances are depicted in Fig. 3.

Table 4 Hypothesis parameters observed in successive runs for binding affinity to estrogen receptor

Run No.	Input parameters			Output parameters				
	[1]Sp	[2]U	[3]W	Cost		R^2	*se*	[5]OF
				Total	[4]Δ			
α-subtype								
1	300	3.0	0.3	94.630	113.598	0.924	0.593	a, d, 3xp
2	250	3.0	0.3	101.950	113.598	0.792	0.930	a, 2xr
3	200	3.0	0.3	99.176	113.598	0.828	0.841	2xp, r
4	150	3.0	0.3	98.603	113.598	0.826	0.846	2xp, r
5	100	3.0	0.3	101.732	113.598	0.814	0.919	2xa, r
6	300	2.5	0.3	108.052	128.585	0.684	1.385	2xd, 2xp
7	300	2.0	0.3	108.153	167.477	0.755	1.580	a, d, p, r
8	300	1.5	0.3	117.992	357.817	0.865	1.993	d, p, r
9	300	1.5	0.5	157.363	365.110	0.342	2.844	a, d, r
10	300	1.5	0.7	140.621	365.110	0.778	2.580	d, p, r
11	300	1.5	1.0	130.422	365.110	0.814	2.371	d, p, r
12	300	1.5	1.3	177.389	365.110	0.661	3.192	2xd, r
13	300	1.5	1.5	129.725	365.110	0.841	2.190	a, d, p
14	**300**	**1.5**	**2.0**	**83.993**	**365.110**	0.955	**1.311**	**a, d, 3xp**
β-subtype								
15	300	3.0	0.3	99.406	97.305	0.771	0.698	d, 2xp, r
16	250	3.0	0.3	99.411	97.306	0.797	0.656	a, 3xp, r
17	200	3.0	0.3	100.247	97.305	0.780	0.684	d, 2xp, r
18	150	3.0	0.3	100.509	97.3059	0.771	0.698	a, d, p, d
19	100	3.0	0.3	100.399	97.306	0.780	0.856	a, 2xp, r
20	250	2.5	0.3	98.261	103.511	0.776	0.826	2xa, 2xp
21	250	2.0	0.3	89.659	122.416	0.830	0.949	d, 2xp, r
22	250	1.5	0.3	86.371	223.174	0.889	1.304	d, p, 2xr
23	**250**	**1.5**	**0.5**	**87.394**	**223.174**	0.885	**1.328**	**d, p, 2xr**
24	250	1.5	1.0	89.494	223.174	0.876	1.386	d, 2xp, r
25	250	1.5	1.5	94.665	223.174	0.846	1.544	2xa, 2xp
26	250	1.5	2.0	90.036	223.174	0.874	1.390	a, 2xp, r

[1]Spacing, [2]Uncertainty, [3]Weight variation, [4](Null cost – Total cost), [5]Output features

The selected model suggested that hydroxyl group attached to ring ‘A’’ (Fig. 1(b)) was revealed as HB donor. The rings ‘B’’ and ‘C’’ were behaved as aromatic ring whereas, methyl group attached to ring ‘A’’ was crucial for imparting hydrophobicity of the molecule. The best hypotheses of both α- and β-subtypes were validated to nullify over prediction of the bioactivity for inactive compounds through hyporefine. This process considers the steric interaction of the compounds in hypothesis generation, but in the present work steric hindrance was not portrayed in the selected models of both subtypes and this means that the presence of steric hindrance of the molecules has no direct influence on binding affinity to ER.

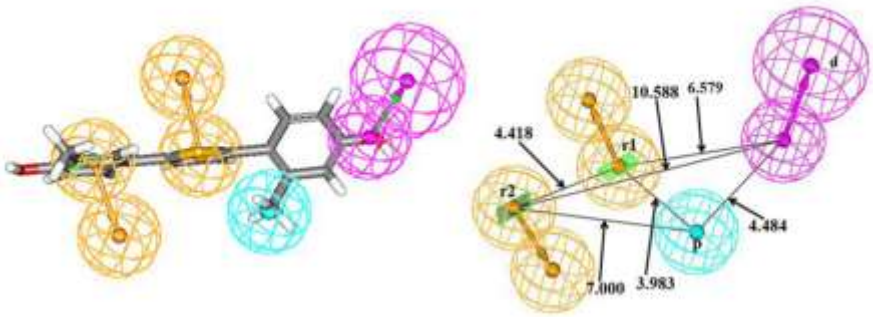


Fig. 3 Mapped with most active compound in training set of β-subtype; mapped features: HB donor (*d*), Hydrophobic (*p*) and Ring aromatics (*r*).

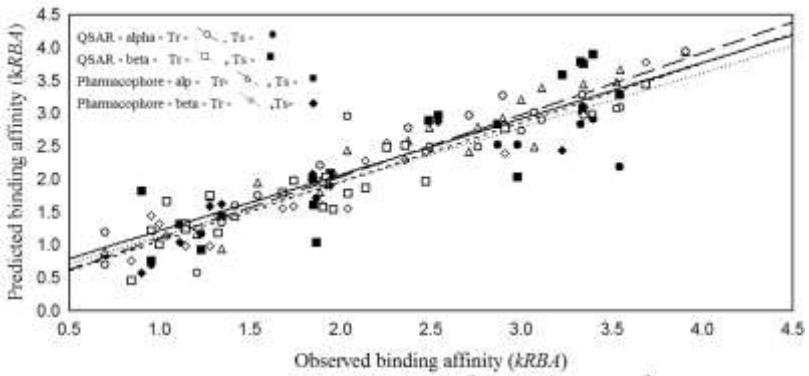
Validation of QSAR and pharmacophore models

Internal validation

The activity of the training compounds were predicted using LOO cross-validation method in case of both QSAR and pharmacophore studies of α- and β-subtypes. In QSAR study the Q^2 was found to be 0.848 and 0.859 for α- and β-subtypes respectively whereas *se* found 0.370 and 0.261 respectively. The r^2_m value was found to be 0.795 and 0.733 for α- and β-subtypes respectively. In addition to the above the Golbraikh-Tropsha metric ($[(r^2-r_0^2)/r^2]$) (Golbraikh and Tropsha 2002) was calculated for both subtypes. For α-subtype it was found to be 0.023 whereas for β-subtype it was 0.007. Values in both subtypes satisfied the criteria (<0.1) explained by Golbraikh et al.(Golbraikh and Tropsha 2002). In case of pharmacophore model Q^2 was found to be 0.932 and 0.878 for α- and β-subtypes respectively along with value of *se* 1.311 and 1.328 for α- and β-subtypes respectively. The r^2_m value for pharmacophore models obtained as 0.891 and 0.815 for α- and β-subtypes respectively. The statistical results (Q^2 and r^2_m >0.5) of the both studies explained that selected models were robust in nature.

Test set

The activity of the test compounds were predicted in both QSAR and pharmacophore studies. The correlation (*R*) between observed and estimated activity of test compounds was 0.933 and 0.920 for α- and β-subtypes respectively in QSAR study, whereas in pharmacophore it was 0.930 and 0.893 for α- and β-subtypes respectively. The R^2_{pred} also was calculated. In QSAR study, R^2_{pred} was found 0.675 with s_p = 0.537 and 0.659 with s_p = 0.408 for α- and β-subtypes respectively. The pharmacophore space modeling models were given R^2_{pred} of 0.737 and 0.769 for α- and β-subtypes respectively along with s_p of 0.497 and 0.336 for α- and β-subtypes respectively. For a better determination of the predictive abilities of the models, the values of $r^2_{m(test)}$ were also calculated. The value of this parameter determines whether the predicted activity values are close to the corresponding observed ones, since a high value of R^2_{pred} may not always indicate a low residual between the observed and predicted activity data. In QSAR study, the $r^2_{m(test)}$ was found to be 0.815 and 0.748 for α- and β-subtypes respectively, whereas in pharmacophore model it was 0.613 and 0.717 respectively for α- and β-subtypes respectively. It is observed that all models in the present study revealed with high R^2_{pred} (>0.5) and $r^2_{m(test)}$ (>0.5) values that explained the superiority of the models.



α-subtype: QSAR study: Tr: n=22, R^2=0.857, se=0.370, Q^2=0.848
Ts: n=10, R^2_{pred}=0.675, s_p=0.537

Pharmacophore: Tr: n=22, R^2=0.955, se=1.311, Q^2=0.932
Ts: n=10, R^2_{pred}=0.737, s_p=0.497

β-subtype: QSAR study: Tr: n=22, R^2=0.874, se=0.261, Q^2=0.859
Ts: n=10, R^2_{pred}=0.659, s_p=0.408

Pharmacophore: Tr: n=22, R^2=0.885, se=1.328, Q^2=0.878
Ts: n=10, R^2_{pred}=0.769, s_p=0.336

Fig. 4 Observed and predicted activity as per QSAR and pharmacophore models

Cost function analysis

The statistical parameters employed for hypothesis generation were spacing, uncertainty, and weight variation. Best pharmacophore models of both subtypes were developed with value of weight variation of 2.0 and 0.5 for α- and β-subtypes respectively while, values of 1.5 to 3.0 were found optimal in case of uncertainty parameter for α- and β-subtypes respectively. As it explained in the materials and methods the configuration cost is equal to the entropy of hypothesis space and should have a value <17 for a good pharmacophore model and in the best pharmacophore models of both α- and β-subtypes the configuration cost were found less tha 15. It is reported that greater the difference (Δ*cost*) between the total and the null costs, it is more likely that the hypothesis does not reflect a chance correlation. In the curent work, the Δ*cost* for α- and β-subtypes was found to be 365.110 and 223.174 respectively that clearly indicated that both models does not reflect a chance correlation. Above statitical parameters of both models undoubtedly indicated that both models were not by chance.

Fischer's randomization test

The quality of the hypotheses was further adjudged through a cross-validation technique, Fischer's randomization technique(Snedecor and Cochran 1967) at 99% confidence level, but none of the hypotheses generated better parameters in comparison to original hypotheses. So the cross-validated validation technique clearly indicates the superiority of the hypotheses consider for binding affinity to ER subtypes.

Binding interactions

The most active compound of both α- (**C6** in Table 1) and β (**C31** in Table 1)-subtypes were considered for molecular docking study to observe the binding interactions between catalytic amino residues and the ligand. The protein receptor molecules (PDB ID: 3ERT and 2QTU for α- and β-subtypes respectively)

were obtained from RSCB Protein Data Bank(Berman et al. 2000). The *LigandFit* module of Discovery Studio was used for the study. In both cases the molecular docking was given ten best poses with binding interactions at the active site cavity. Receptor-ligand complex for both subtypes were selected based on the dock score and number of binding interactions at the active site cavity of ER. The best dock pose of both α- and β-subtypes is depicted in the Fig. 5.

α-subtype

The molecular docking study of most active compound of α-subtype (Fig. 5(a)) revealed that one hydrogen bond formed between Thr347 and oxygen atom of sulphite group present in between rings 'C' and 'D' (Fig. 1(a)), while one bump interaction observed with Ala350 with same group. The rings 'A' and 'C' of the ligand (**C6**) were connected with amino residue Cys530 through a potential hydrogen bond, in addition ring 'B' was formed one bump interaction. Thr347 and Leu525 were found important to connect ring 'C' through bump and hydrogen bond interaction respectively.

The functionalities developed in the QSAR and pharmacophore studies were successfully correlated with binding interactions observed at the active site cavity of ERα. The importance of hydrophobicity, shape of the molecule, orbital energy and atomic refractivity in QSAR study adjudged through hydrogen bonding and bump interaction perceived between the ligand and catalytic residues in the active site. The pharmacophore model was suggested that sulphite group of the molecule might be crucial for HB acceptor that is validated by forming interactions of sulphite group with Thr347 and Ala350 amino acids. The hydroxyl group attached to ring 'A' was revealed as HB donor in the ligand-based pharmacophore model justified by the hydrogen bond interaction between Cys530 and same. The alkyl group attached to ring 'A' and phenyl rings 'B' and 'C' were important for imparting hydrophobicity of the molecules was successfully validated by interactions formed with Cys530, Thr347 and Leu525.

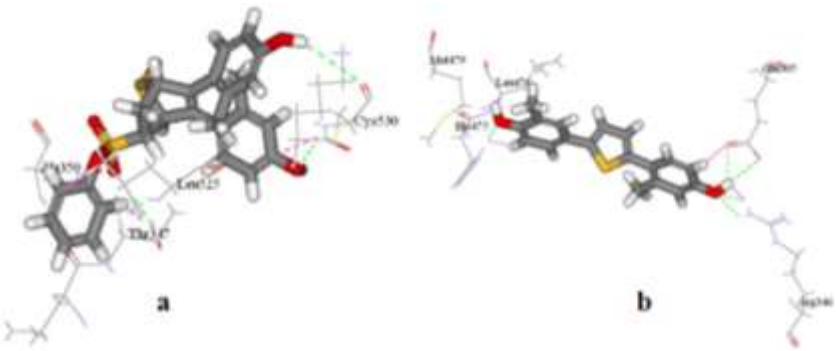


Fig. 5 Best docked pose a) α-subtype and b) β-subtype

β-subtype

The hydroxyl group attached to the ring 'B'' (Fig. 1(b)) formed two hydrogen bonds with each of the Glu305 and Arg346 amino residues. The same functional group of ring 'B'' formed one bump interaction with Arg346 residue. The Glu305 residue directly interacted with ring 'B'' via two bump interactions. The hydroxyl group attached to ring 'A'' was connected to His475 via one potential hydrogen bond and one bump interactions. The functional group of ring 'A'' was able to connect Leu476 and Met479 via bump interaction.

The important functionalities derived in the ligand-based QSAR and pharmacophore studies of β-subtype were successfully correlated with binding interactions found between most active compound β-subtype and catalytic residues of ERβ. The hydrophobicity, dipole moment, molar refractivity contributed by atoms of the molecule and the energy parameters were found to be critical in the QSAR model judged by the binding interaction observed in molecular docking study. In the pharmacophore study hydroxyl group attached to ring 'A'' (Fig. 1(b)) was important for HB donor which was validated by forming hydrogen bond interaction by His475 with the same. Ring 'B'' showed as an aromatic pharmacophoric feature that was validated by bump interactions with Glu305 and Met340 amino residues at the active site cavity of ERβ.

Conclusion

The present study examined the structure–activity relationship for non-steroidal analogs, SERMs that were being evaluated for a number of estrogen-related diseases, and highlighted some of the molecular properties and structural requirement for selective binding affinity to the ER subtypes. The 2D QSAR models were generated to identify descriptors that contributed significant correlation with biological activity, and search for pharmacophoric elements responsible for selective binding affinity to the ER subtypes. The models generated from the data set were validated by assessing their statistical significance and predictivity. Hydrophobicity in both QSAR and pharmacophore studies was observed to be prime feature for imparting binding affinity. Orbital energy, molar refractivity and dipole moments were revealed critical parameters for binding affinity towards ER. The space modelling study suggested that accumulation of more hydrophobic and nucleophile substituents to the molecules may provide better therapeutic effects for estrogen therapy. The receptor-based molecular docking study successfully correlated the functionality developed in the ligand-based QSAR and pharmacophore modelling. The polar Glu305, Arg346, Thr379, Trp383 and His475; and, non-polar Met340, Leu476, Met479, Leu525 and Leu536 were found to be the catalytic amino residues at the active site cavity of the ER. Crucial functionalities developed in the receptor-independent pharmacoinformatics study can be used to design and synthesis new chemical entities for effective application in estrogen therapy.

Acknowledgment

MA Islam and TS Pillay were funded by the University of Pretoria Vice Chancellor's post-doctoral fellowship and National Research Foundation (NRF), South Africa Innovation Post-doctoral fellowship schemes.

Conflict interests

The authors declare that they have no conflict interests.

References

Accelrys (2013) Discovery Studio Modeling Environment, Release 4.0. Accelrys Software Inc., San Diego

Balaban AT, Khadikar PV, Supuran CT, Thakur A, Thakur M (2005) Study on supramolecular complexing ability vis-a-vis estimation of pKa of substituted sulfonamides: dominating role of

Balaban index (J) Bioorganic & medicinal chemistry letters 15:3966-3973 doi:10.1016/j.bmcl.2005.05.136
Berman HM et al. (2000) The Protein Data Bank Nucleic acids research 28:235-242
Brogia S, Papazafiric P, Roussisd V, Tafi A (2013) 3D-QSAR using pharmacophore-based alignment and virtual screening for discovery of novel MCF-7 cell line inhibitors European Journal of Medicinal Chemistry 67:344–351
Burrow PD, Modelli A (2013) On the treatment of LUMO energies for their use as descriptors SAR and QSAR in environmental research 24:647-659 doi:10.1080/1062936X.2013.792873
Chang YH, Chen JY, Hor CY, Chuang YC, Yang CB, Yang CN (2013) Computational Study of Estrogen Receptor-Alpha Antagonist with Three-Dimensional Quantitative Structure-Activity Relationship, Support Vector Regression, and Linear Regression Methods International Journal of Medicinal Chemistry 2013:1-3
Chlebowski RT et al. (2003) Influence of estrogen plus progestin on breast cancer and mammography in healthy postmenopausal women: the Women's Health Initiative Randomized Trial JAMA 289:3243-3253 doi:10.1001/jama.289.24.3243
Chmel R, Rob L, Strnad P (2002) [What can we expect of raloxifene in the treatment of postmenopausal osteoporosis--views of a gynecologist] Ceska gynekologie / Ceska lekarska spolecnost J Ev Purkyne 67:187-191
Clarke R et al. (2003) Antiestrogen resistance in breast cancer and the role of estrogen receptor signaling Oncogene 22:7316-7339 doi:10.1038/sj.onc.1206937
Cramer RD, Patterson DE, Bunce JD (1988) Comparative molecular field analysis (CoMFA). 1. Effect of shape on binding of steroids to carrier proteins Journal of the American Chemical Society 110:5959-5967 doi:10.1021/ja00226a005
Dalkas GA, Vlachakis D, Tsagkrasoulis D, Kastania A, Kossida S (2012) State-of-the-art technology in modern computer-aided drug design Brifiefings in Bioinformatics 14:745-752
Doisneau-Sixou SF, Sergio CM, Carroll JS, Hui R, Musgrove EA, Sutherland RL (2003) Estrogen and antiestrogen regulation of cell cycle progression in breast cancer cells Endocrine-related cancer 10:179-186
Dowers TS, Qin ZH, Thatcher GR, Bolton JL (2006) Bioactivation of Selective Estrogen Receptor Modulators (SERMs) Chemical research in toxicology 19:1125-1137 doi:10.1021/tx060126v
Fisher B et al. (2005) Tamoxifen for the prevention of breast cancer: current status of the National Surgical Adjuvant Breast and Bowel Project P-1 study Journal of the National Cancer Institute 97:1652-1662 doi:10.1093/jnci/dji372
Fisher B et al. (1998) Tamoxifen for prevention of breast cancer: report of the National Surgical Adjuvant Breast and Bowel Project P-1 Study Journal of the National Cancer Institute 90:1371-1388
Foster JS, Henley DC, Ahamed S, Wimalasena J (2001a) Estrogens and cell-cycle regulation in breast cancer Trends in endocrinology and metabolism: TEM 12:320-327
Foster JS, Henley DC, Bukovsky A, Seth P, Wimalasena J (2001b) Multifaceted regulation of cell cycle progression by estrogen: regulation of Cdk inhibitors and Cdc25A independent of cyclin D1-Cdk4 function Molecular and cellular biology 21:794-810 doi:10.1128/MCB.21.3.794-810.2001
Gehrig PA, Bae-Jump VL, Boggess JF, Groben PA, Fowler WC, Jr., Van Le L (2004) Association between uterine serous carcinoma and breast cancer Gynecologic oncology 94:208-211 doi:10.1016/j.ygyno.2004.04.009
Golbraikh A, Tropsha A (2002) Beware of q2! Journal of molecular graphics & modelling 20:269-276
Gupta MK, Prabhakar YS (2008) QSAR study on tetrahydroquinoline analogues as plasmodium protein farnesyltransferase inhibitors: a comparison of rationales of malarial and mammalian enzyme inhibitory activities for selectivity Eur J Med Chem 43:2751-2767 doi:10.1016/j.ejmech.2008.01.025
Gustafsson JA (1999) Estrogen receptor beta--a new dimension in estrogen mechanism of action The Journal of endocrinology 163:379-383
Holst F et al. (2007) Estrogen receptor alpha (ESR1) gene amplification is frequent in breast cancer Nature genetics 39:655-660 doi:10.1038/ng2006
Hopfinger AJ, Tokarski JS (1997) Three-Dimensional QuantitativeStructure-Activity Relationship Analysis. In: Charifson PS (ed) Practical Applicationof Computer-Aided Drug Design. MarcelDekker, Inc., New York, USA, pp 105-164
Islam MA, Nagar S, Das S, Mukherjee A, Saha A (2008) Molecular design based on receptor-independent pharmacophore: application to estrogen receptor ligands Biological & pharmaceutical bulletin 31:1453-1460
Kapetanovic IM (2008) Computer-aided drug discovery and development (CADDD): in silico-chemico-biological approach Chemico-biological interactions 171:165-176 doi:10.1016/j.cbi.2006.12.006

Kim Y, Nam NH, You YJ, Ahn BZ (2002) Synthesis and cytotoxicity of 3,4-diaryl-2(5H)-furanones Bioorganic & medicinal chemistry letters 12:719-722
Kristam R, Gillet VJ, Lewis RA, Thorner D (2005) Comparison of conformational analysis techniques to generate pharmacophore hypotheses using catalyst Journal of chemical information and modeling 45:461-476 doi:10.1021/ci049731z
Kubinyi H (2004) 2D QSAR Models: Hansch and Free-Wilson Analyses. In: Bultinck P, Winter HD, Langenaeker W, Tollenaere JP (eds) Comput. Med. Chem. Drug Discov. Marcel Dekker, Inc, NewYork, USA, pp 539-570
Kubinyi H, Hamprecht FA, Mietzner T (1998) Three-dimensional quantitative similarity-activity relationships (3D QSiAR) from SEAL similarity matrices Journal of medicinal chemistry 41:2553-2564 doi:10.1021/jm970732a
Kupcewicz B, Balcerowska-Czerniak G, Malecka M, Paneth P, Krajewska U, Rozalski M (2013) Structure-cytotoxic activity relationship of 3-arylideneflavanone and chromanone (E,Z isomers) and 3-arylflavones Bioorganic & medicinal chemistry letters 23:4102-4106 doi:10.1016/j.bmcl.2013.05.044
Levet A et al. (2013) Quantitative structure-activity relationship to predict acute fish toxicity of organic solvents Chemosphere 93:1094-1103 doi:10.1016/j.chemosphere.2013.06.002
Lewis DF, Parker MG, King RJ (1995) Molecular modelling of the human estrogen receptor and ligand interactions based on site-directed mutagenesis and amino acid sequence homology J Steroid Biochem Mol Biol 52:55-65
Lewis JS, Jordan VC (2005) Selective estrogen receptor modulators (SERMs): mechanisms of anticarcinogenesis and drug resistance Mutation research 591:247-263 doi:10.1016/j.mrfmmm.2005.02.028
Li H, Sutter J, Hoffman R (2000) Pharmacophore Perception, Development, and Use in Drug Design. International University Line, California
Maximov PY, Lee TM, Jordan VC (2013) The discovery and development of selective estrogen receptor modulators (SERMs) for clinical practice Current clinical pharmacology 8:135-155
Mitra I, Saha A, Roy K (2010) Pharmacophore mapping of arylamino-substituted benzo[b]thiophenes as free radical scavengers Journal of molecular modeling 16:1585-1596 doi:10.1007/s00894-010-0661-4
MOE (2007) Molecular Operating Environment (MOE). Chemical Computing Group Inc., 1010 Sherbooke St. West, Suite #910, Montreal, QC, Canada
Mukherjee S, Saha A, Roy K (2005) QSAR of estrogen receptor modulators: exploring selectivity requirements for ER(alpha) versus ER(beta) binding of tetrahydroisoquinoline derivatives using E-state and physicochemical parameters Bioorg Med Chem Lett 15:957-961 doi:10.1016/j.bmcl.2004.12.048
Nandy A, Kar S, Roy K (2013) Development and validation of regression-based QSAR models for quantification of contributions of molecular fragments to skin sensitization potency of diverse organic chemicals SAR and QSAR in environmental research 24:1009-1023 doi:10.1080/1062936X.2013.821422
Nantasenamat C, Worachartcheewan A, Prachayasittikul S, Isarankura-Na-Ayudhya C, Prachayasittikul V (2013) QSAR modeling of aromatase inhibitory activity of 1-substituted 1,2,3-triazole analogs of letrozole Eur J Med Chem 69:99-114 doi:10.1016/j.ejmech.2013.08.015
Ojha PK, Mitra I, Das RN, Roy K (2011) Further exploring rm2 metrics for validation of QSPR models Chemometrics and Intelligent Laboratory Systems 107:194–205
Pickar JH, MacNeil T, Ohleth K (2010) SERMs: progress and future perspectives Maturitas 67:129-138 doi:10.1016/j.maturitas.2010.05.009
Pike VW, Law MP, Osman S, Davenport RJ, Rimoldi O, Giardina D, Camici PG (2000) Selection, design and evaluation of new radioligands for PET studies of cardiac adrenoceptors Pharmaceutica acta Helvetiae 74:191-200
Platts JA, Oldfield SP, Reif MM, Palmucci A, Gabano E, Osella D (2006) The RP-HPLC measurement and QSPR analysis of logP(o/w) values of several Pt(II) complexes Journal of inorganic biochemistry 100:1199-1207 doi:10.1016/j.jinorgbio.2006.01.035
Rastija V, Medic-Saric M (2009) QSAR study of antioxidant activity of wine polyphenols Eur J Med Chem 44:400-408 doi:10.1016/j.ejmech.2008.03.001
Richardson TI, Dodge JA, Wang Y, Durbin JD, Krishnan V, Norman BH (2007) Benzopyrans as selective estrogen receptor beta agonists (SERBAs). Part 5: Combined A- and C-ring structure-activity relationship studies Bioorganic & medicinal chemistry letters 17:5563-5566 doi:10.1016/j.bmcl.2007.08.009

Riggs BL, Hartmann LC (2003) Selective estrogen-receptor modulators -- mechanisms of action and application to clinical practice The New England journal of medicine 348:618-629 doi:10.1056/NEJMra022219

Rossouw JE et al. (2002) Risks and benefits of estrogen plus progestin in healthy postmenopausal women: principal results From the Women's Health Initiative randomized controlled trial Jama 288:321-333

Roy K, Mitra I, Kar S, Ojha PK, Das RN, Kabir H (2012) Comparative studies on some metrics for external validation of QSPR models Journal of chemical information and modeling 52:396-408 doi:10.1021/ci200520g

Roy PP, Paul S, Mitra I, Roy K (2009) On two novel parameters for validation of predictive QSAR models Molecules 14:1660-1701

Roy PP, Roy K (2008) On Some Aspects of Variable Selection for Partial Least Squares Regression Models QSAR & Combinatorial Science 27:302–313

Sadler BR, Cho SJ, Ishaq KS, Chae K, Korach KS (1998) Three-dimensional quantitative structure-activity relationship study of nonsteroidal estrogen receptor ligands using the comparative molecular field analysis/cross-validated r2-guided region selection approach Journal of medicinal chemistry 41:2261-2267 doi:10.1021/jm9705521

Saxena AK et al. (2003) QSAR studies in substituted 1,2,3,4,6,7,12,12a-octa-hydropyrazino[2',1':6,1]pyrido[3,4-b]indoles--a potent class of neuroleptics Bioorganic & medicinal chemistry 11:2085-2090

Shiau AK, Barstad D, Loria PM, Cheng L, Kushner PJ, Agard DA, Greene GL (1998) The structural basis of estrogen receptor/coactivator recognition and the antagonism of this interaction by tamoxifen Cell 95:927-937

Smith HM, Knox AJ, Zisterer DM, Lloyd DG, Meegan MJ (2007) Flexible estrogen receptor modulators: synthesis, biochemistry and molecular modeling studies for 3-benzyl-4,6-diarylhex-3-ene and 3,4,6-triarylhex-3-ene derivatives Med Chem 3:135-155

Snedecor GW, Cochran WG (1967) Statistical methods. Ames, 6 edn. Iowa

State University Press,

Vedani A, Dobler M (2002) Multidimensional QSAR: Moving from three- to five-dimensional concepts Quantitative Structure-Activity Relationships 21:382–390

Verma J, Khedkar VM, Coutinho EC (2010) 3D-QSAR in drug design--a review Current topics in medicinal chemistry 10:95-115

Wang P et al. (2012) Identification and structure-activity relationships of a novel series of estrogen receptor ligands based on 7-thiabicyclo[2.2.1]hept-2-ene-7-oxide Journal of medicinal chemistry 55:2324-2341 doi:10.1021/jm201556r

Wildman SA, Crippen GM (1999) Prediction of Physicochemical Parameters by Atomic Contributions Journal of chemical information and modeling 39:868-873

Yaffe K, Sawaya G, Lieberburg I, Grady D (1998) Estrogen therapy in postmenopausal women: effects on cognitive function and dementia Jama 279:688-695

Zhang L et al. (2013) Identification of putative estrogen receptor-mediated endocrine disrupting chemicals using QSAR- and structure-based virtual screening approaches Toxicology and applied pharmacology 272:67-76 doi:10.1016/j.taap.2013.04.032