MEDICINAL CHEMISTRY RESEARCH

ORIGINAL RESEARCH

Check for updates

In silico prediction of estrogen receptor subtype binding affinity and selectivity using 3D-QSAR and molecular docking

Wenliang Jiang[1] · Qinghua Chen[1] · Bo Zhou[2] · Fangfang Wang[1]

Abstract

Estrogen receptors (ERs) are significant protein targets that can alter normal homeostatic transcription and signaling pathways. And the development of selective ER ligands is involved in the pharmacology of anticancer drugs, leading to the synthesis of numerous ER selective ligands. In the present work, three dimensional quantitative structure–activity relationship (3D-QSAR) models, including comparative molecular field analysis (CoMFA) and comparative molecular similarity indices analysis (CoMSIA) in combination with molecular docking were performed on a series of ER ligands to investigate the potential relationship between the structural features of the compounds and its activity and selectivity. Good statistical significance was achieved for the CoMFA (ERα: $R^2_{cv} = 0.676$, $R^2_{pred} = 0.631$. ERβ: $R^2_{cv} = 0.578$, $R^2_{pred} = 0.5341$) and CoMISA (ERα: $R^2_{cv} = 0.693$, $R^2_{pred} = 0.5758$. ERβ: $R^2_{cv} = 0.653$, $R^2_{pred} = 0.615$) models. In addition, molecular docking results confirmed that the substituents at ring A, ring D and ring E were crucial for interacting with receptors ERα and ERβ, respectively, which were in agreement with the QSAR results. The substituents in ring A and ring F for ERα and residues Thr347/Ala302, Met388/Leu343, Leu391/Arg346, Met421/Ile376, Gly521/Leu476 for ERα and ERβ, respectively, contribute the most to the subtype selectivity. These results obtained from this work provide valuable information for the design of novel selective ER ligands with enhanced activity.

Keywords Estrogen receptor β; · Estrogen receptor α · CoMFA · CoMSIA · Molecular docking

Introduction

Estrogens are mainly involved in menstrual and estrous reproductive cycles. The physiological function is modulated largely by the estrogen receptors (ERs), which regulate various endocrine mechanisms, including development, homeostasis, and metabolism (Aranda and Pascual 2001). ERs are expressed in a cell-type and tissue specific manner, and the presence varies based on different disease state,

✉ Fangfang Wang
yu100288@163.com

[1] School of Life Science, Linyi University, Linyi 276000, China

[2] State Key Laboratory of Functions and Applications of Medicinal Plants, College of Basic Medical, Guizhou Medical University, Guizhou 550004, China

such as stroke and infertility (Weihua et al. 2003; Nilsson and Gustafsson 2010; Powel et al. 2012).

There are two subtypes for ERs, ERα, and ERβ, which share 53% sequence identity in the ligand binding domain, and they are highly similar in the DNA binding domain (Journé et al. 2008). However, the two subtypes exhibit distinct cellular and tissue distribution patterns. ERα is mainly expressed in the mammary gland, uterus, ovary, bone, male reproductive organs, prostate, liver, and adipose tissue. In contrast, ERβ is predominant in the prostate, bladder, ovary, colon, adipose tissue, and immune system (Kuiper et al. 1997; Fitzpatrick et al. 1999). Evidence has certified that estradiol and hormone replacement therapies would target ERα and ERβ, but the risk of breast and endometrial cancers, and thromboembolism will be triggered (Maximov et al. 2013). Many studies have proven that the ligands binding selectively to ERβ may be employed to cure certain cancers, endometriosis, inflammatory diseases, and cardiovascular (Minutolo et al. 2011). In additional, these compounds have an effect on regulating brain development and estrogen-induced promotion of neurogenesis (Zhao and Brinton 2005). Effective efforts

have also been made to develop subtype-selective compounds which selectively antagonize undesirable side effects for therapeutic purposes.

ERs possess conserved structurally and functionally distinct domains. The DNA-binding domain, as the most conserved domain, is involved in DNA recognition and binding, whereas ligand binding occurs in the carboxyl terminal ligand-binding domain, consisting of twelve α-helices (H1-H12) and a β-hairpin. The H12 helix plays important role in molecular switching (Dilis et al. 2007). When ligands binds, the H12 rests across H3 and H11, forming a groove to accommodate co-regulator and facilitate downstream activation process (Bruning et al. 2010). In addition, the amino terminal domain is not conserved and represents the most variable domain both in sequence and length (Nilsson et al. 2001).

Based on the structural characteristics of ERs, many studies have synthesized some potent compounds for ERs (Farzaneh and Zarghi 2016). Naturally occurring phytoestrogens and several modified analogs have been reported to possess about 40-fold selectivity for ERβ receptor (Miller et al. 2003), for example, diarylpropylnitrile (Meyers et al. 2001; Jun et al. 2003) and bipenyl (Korach et al. 1988; Kuiper et al. 1998) exhibited 70-fold selectivity for ERβ, additionally, tetrahydrochrysenes (Meyers et al. 1999), aryl benzothiophenes (Ulrich et al. 2002), and benzoxazenes (Yang et al. 2004) have been reported possessing 10- to 40-fold ERβ receptor selective binding. However, non-selective or low binding affinity compounds will be synthesized inevitably. Therefore, novel ER compounds with improved binding activity and selectivity should be explored.

It has been reported that several experimental methods making use of ER receptors or biological materials of human, rat, mouse, and calf can be used for screening the estrogenic chemicals (Pons et al. 1990; Soto et al. 1995; Reel et al. 1996; Shelby et al. 1996; Fang et al. 2000). However, it is time-consuming and material redundant for screening of novel compounds in vitro. Therefore, computational models are preferred now for some reasons, such as speed and low cost. For example, environmental estrogens have been classified by physicochemical properties using principal component analysis and hierarchical cluster analysis (Suzuki et al. 2001). Chafourian et al. have developed partial least squares (PLS) model to identify significant descriptors associated with ER binding affinity (Ghafourian and Cronin 2005). Marini et al. (2005) and Ji et al. (2008) have constructed the QSAR model to study the ER ligands using artificial neural networks. In addition, consensus kNN QSAR as a feasible method has been used for rapid screening of organic compounds having estrogenic activity (Asikainen et al. 2004). However, these classical approaches ignore the three dimensional conformation, often fail to elucidate molecular mechanism of ER ligands effectively.

Therefore, Tong et al. (1997) and Wolohan and Reichert (2003) performed comparative molecular field analysis (CoMFA) on a series of compounds binding to ERs, resulting in predictive QSAR models.

In the present paper, a multistep work combing 3D-QSAR and molecular docking was done to study the structure–activity relationship of the employed inhibitors, to elucidate the structural features required for binding affinity of ERs and selectivity for ERα over ERβ receptor. The results can be employed to guide the rational synthesis of novel selective ligands of ERα possessing increased binding affinity and higher selectivity.

Methods and materials

Datasets and biological activity

In this study, a total of 81 compounds were collected from the literatures (Blizzard et al. 2004a, 2004b, 2005; Chen et al. 2004; Kim et al. 2004) and employed for molecular modeling. To facilitate the QSAR analysis, the IC_{50} values were converted to the corresponding pIC_{50} ($-logIC_{50}$) values, which were further used as dependent variable for model construction. The structure and inhibitory activity are listed in Table 1. The whole dataset was divided into a training set for model generation and a test set (compounds marked with [a] in Table 1) for model validation. In splitting the original dataset, the structural diversity and the range of biological activity in the two sets were carefully considered to ensure the representative ability of the training set molecules.
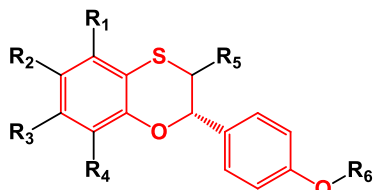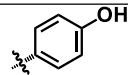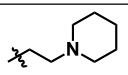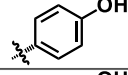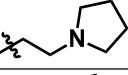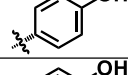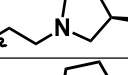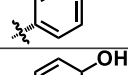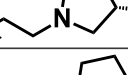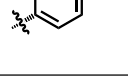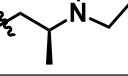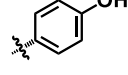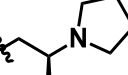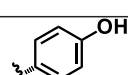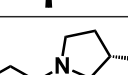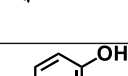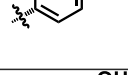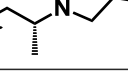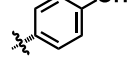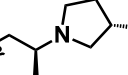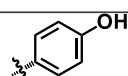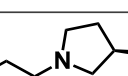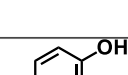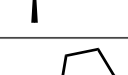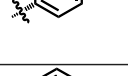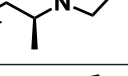
Molecular modeling
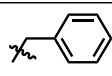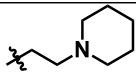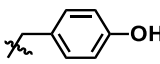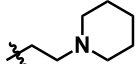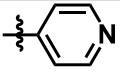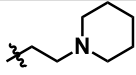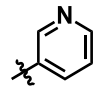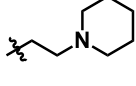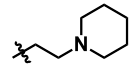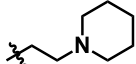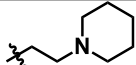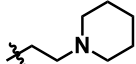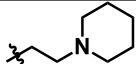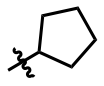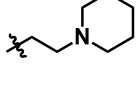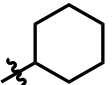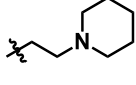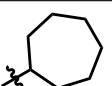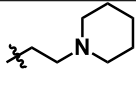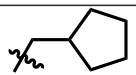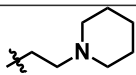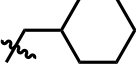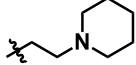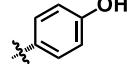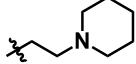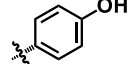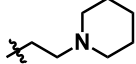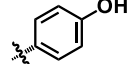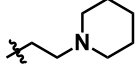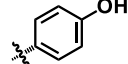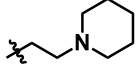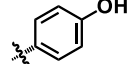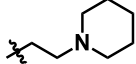
3D-QSAR studies and all molecular modeling were performed using the Sybyl-X 1.1 (Tripos Associates, St. Louis, MO). Initially, the structures of all compounds used in the model generation were sketched. Then the geometrical conformations were optimized using the Tripos force field with a distance-based dielectric and the Powell conjugate gradient algorithm (Clark et al. 1989). Different types of charges (Gasteiger and Marsili 1980) were calculated for each compound. Repeated minimizations were performed with the maximum iterations of 1000 to reach an energy convergence gradient value of 0.05 kcal/mol Å. Furthermore, the minimized structures were employed as the initial conformation for CoMFA and CoMSIA analyses.

Conformational sampling and alignment

In order to obtain reliable QSAR models, molecular alignment is considered as one of the most sensitive parameters
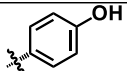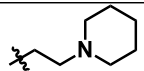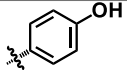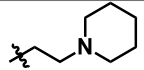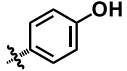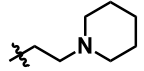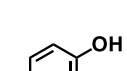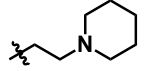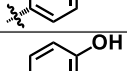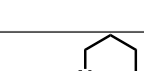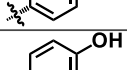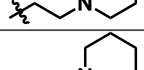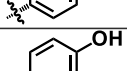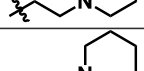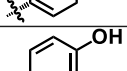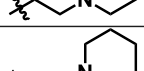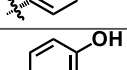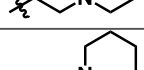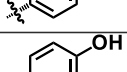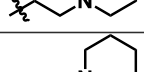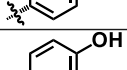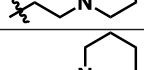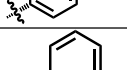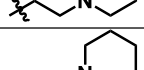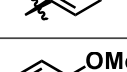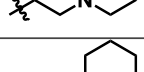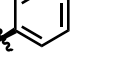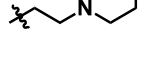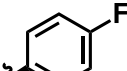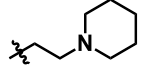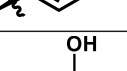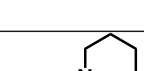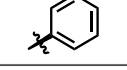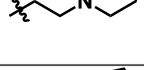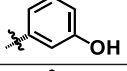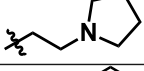
Table 1 Molecular structures of ER ligands and the binding affinity pIC_{50} values

Compound	R_1	R_2	R_3	R_4	R_5	R_6	$pIC_{50}/ER\beta$	$pIC_{50}/ER\alpha$
1[a]	H	OH	H	H			7.37	9.05
2	H	OH	H	H			7.19	8.59
3	H	OH	H	H			7.06	8.80
4[a]	H	OH	H	H			7.30	9.00
5	H	OH	H	H			7.35	9.00
6	H	OH	H	H			7.15	8.77
7	H	OH	H	H			7.06	9.10
8[a]	H	OH	H	H			7.05	9.22
9	H	OH	H	H			7.35	8.60
10	H	OH	H	H			7.28	8.89
11	F	OH	H	H			8.28	9.10
12	H	OH	H	H			7.16	8.92
13	F	OH	H	H			8.04	9.15

No.					Structure	Amine	Value 1	Value 2
14	H	OH	H	H			7.22	8.13
15	H	OH	H	H	(–OH)		6.51	8.57
16[a]	H	OH	H	H	(pyridine)		5.09	6.31
17	H	OH	H	H	(pyridine)		5.08	6.27
18	H	OH	H	H	(thiophene, S)		6.79	8.06
19	H	OH	H	H	(furan, O)		6.44	7.46
20	H	OH	H	H	$-CH(CH_3)_2$		6.98	8.15
21[a]	H	OH	H	H	$-C(CH_3)_3$		7.37	7.96
22	H	OH	H	H	$-CH_2CH(CH_3)_2$		7.42	8.52
23[a]	H	OH	H	H	(cyclopentyl)		7.51	8.60
24	H	OH	H	H	(cyclohexyl)		7.80	8.59
25	H	*OH*	H	H	(cycloheptyl)		7.55	8.12
26	H	OH	H	H	(cyclopentylmethyl)		7.47	8.52
27	H	OH	H	H	(cyclohexylmethyl)		7.40	8.41
28	H	H	OH	H	(–OH)		7.35	8.74
29	Me	H	OH	H	(–OH)		7.34	8.70
30	Et	H	OH	H	(–OH)		6.73	7.27
31[a]	OH	H	OH	H	(–OH)		6.10	7.60
32	F	H	OH	H	(–OH)		7.89	9.30
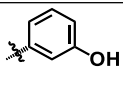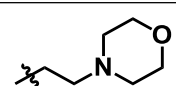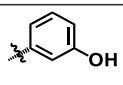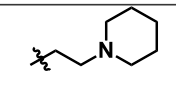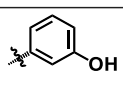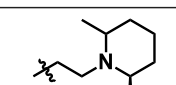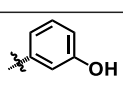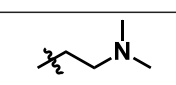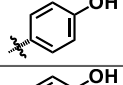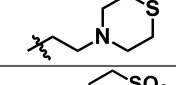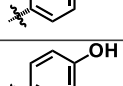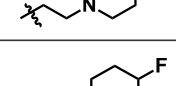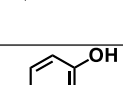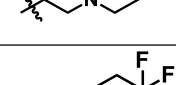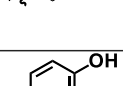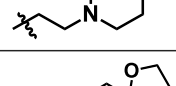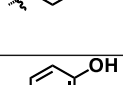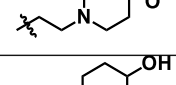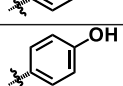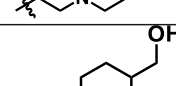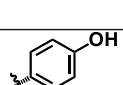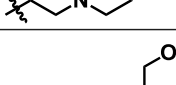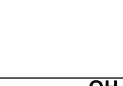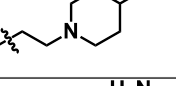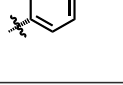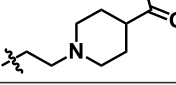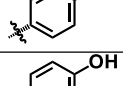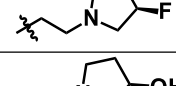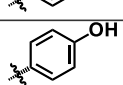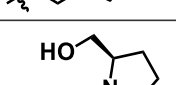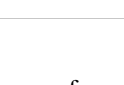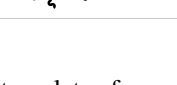
(Cho and Tropsha 1995; Thaimattam et al. 2005). In this study, the most potent compound was selected as a template, and the remaining compounds were aligned on the template by common substructure alignment (as shown in Figs. 1a and 2a). And three different alignment rules were employed to build the most reliable QSAR models: template ligand-based alignment, docking-based alignment, and scaffold-based alignment.

33[a]	Cl	H	OH	H			8.10	9.29
34	H	Cl	OH	H			6.28	7.64
35	H	Me	OH	H			6.23	7.46
36	H	H	OH	Me			7.27	8.92
37	H	H	OH	Et			7.60	8.33
38	F	OH	H	H			7.88	9.10
39	Cl	OH	H	H			7.85	8.62
40[a]	H	OH	Cl	H			7.14	8.27
41	H	OH	F	H			6.38	7.87
42	H	OH	H	Cl			6.98	8.28
43	H	OH	H	Me			6.46	8.28
44[a]	H	OH	H	H			6.52	7.92
45	H	OH	H	H			6.10	7.39
46[a]	H	OH	H	H			7.18	8.42
47[a]	H	OH	H	H			6.60	8.52
48[a]	H	OH	H	H			7.43	9.05
49	H	OH	H	H			6.94	8.39
50	H	OH	H	H			6.99	8.92

Template ligand-based alignment (*Alignment 1*): in this method, the most potent compound (compound 11 for ERβ and compound 78 for ERα) was selected as template to align the remaining compounds. The aligned results are shown in Figs. 1b and 2b.

Docking-based alignment (*Alignment 2*): all compounds were docked in the binding site of the receptors, and the conformation of each compound (possessing the highest total score) was used for model generation (shown in Figs. 1c and 2c).

51	H	OH	H	H	(3-OH phenyl)	(morpholine ethyl)	6.50	8.42
52	H	OH	H	H	(3-OH phenyl)	(piperidine ethyl)	6.96	9.10
53	H	OH	H	H	(3-OH phenyl)	(2,6-dimethylpiperidine ethyl)	6.46	8.47
54	H	OH	H	H	(3-OH phenyl)	(dimethylamino ethyl)	6.89	8.85
55[a]	H	OH	H	H	(4-OH phenyl)	(thiomorpholine ethyl)	7.14	8.82
56	H	OH	H	H	(4-OH phenyl)	(thiomorpholine-SO$_2$ ethyl)	6.61	8.70
57	H	OH	H	H	(4-OH phenyl)	(4-F piperidine ethyl)	6.94	8.60
58	H	OH	H	H	(4-OH phenyl)	(4,4-diF piperidine ethyl)	6.56	7.80
59	H	OH	H	H	(4-OH phenyl)	(dioxolane spiro piperidine ethyl)	6.57	8.70
60	H	OH	H	H	(4-OH phenyl)	(4-OH piperidine ethyl)	6.48	8.74
61	H	OH	H	H	(4-OH phenyl)	(4-CH$_2$OH piperidine ethyl)	6.36	8.66
62[a]	H	OH	H	H	(4-OH phenyl)	(4-CH$_2$CH$_2$OH piperidine ethyl)	5.95	8.14
63	H	OH	H	H	(4-OH phenyl)	(4-carboxamide piperidine ethyl)	6.44	8.66
64	H	OH	H	H	(4-OH phenyl)	(3-F pyrrolidine ethyl)	7.23	9.30
65	H	OH	H	H	(4-OH phenyl)	(3-OH pyrrolidine ethyl)	7.09	8.77
66[a]	H	OH	H	H	(4-OH phenyl)	(2-CH$_2$OH pyrrolidine ethyl)	6.87	8.70

Scaffold-based alignment (*Alignment 3*): the conformations of all compounds were generated from molecular docking, then, the most active compound was used as template for superimposition, which the procedure is similar to *Alignment 1*. And the result is shown in Figs. 1d and 2d.
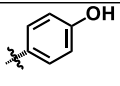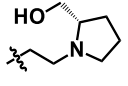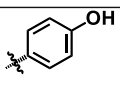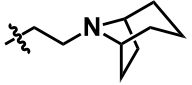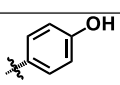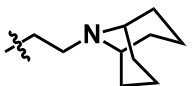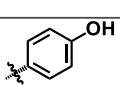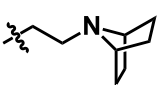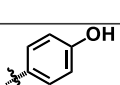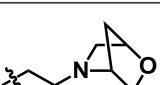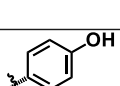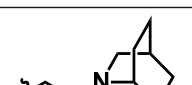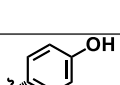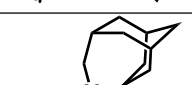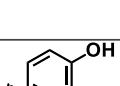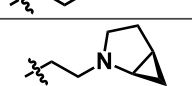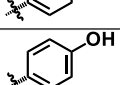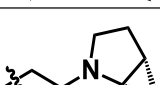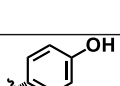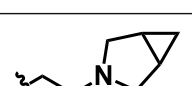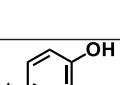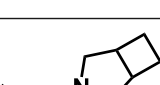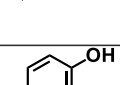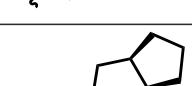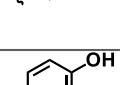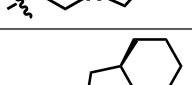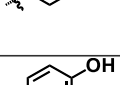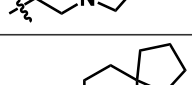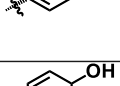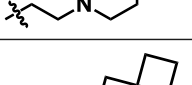
67	H	OH	H	H			7.09	8.41
68	H	OH	H	H			7.29	9.22
69	H	OH	H	H			7.33	9.22
70	H	OH	H	H			6.93	9.05
71[a]	H	OH	H	H			7.14	9.15
72	H	OH	H	H			7.28	8.89
73[a]	H	OH	H	H			7.40	9.22
74[a]	H	OH	H	H			7.48	9.22
75	H	OH	H	H			7.25	9.30
76	H	OH	H	H			6.87	9.15
77[a]	H	OH	H	H			7.07	9.10
78	H	OH	H	H			7.74	9.52
79[a]	H	OH	H	H			7.48	9.00
80	H	OH	H	H			7.41	8.92
81	H	OH	H	H			7.59	9.40

Fig. 1 a Compound 11 used as templates for template ligand-based alignment. The common substructure is shown in blue. **b–d** Present the alignments for ERβ from the *Alignment 1*, *2* and *3*, respectively
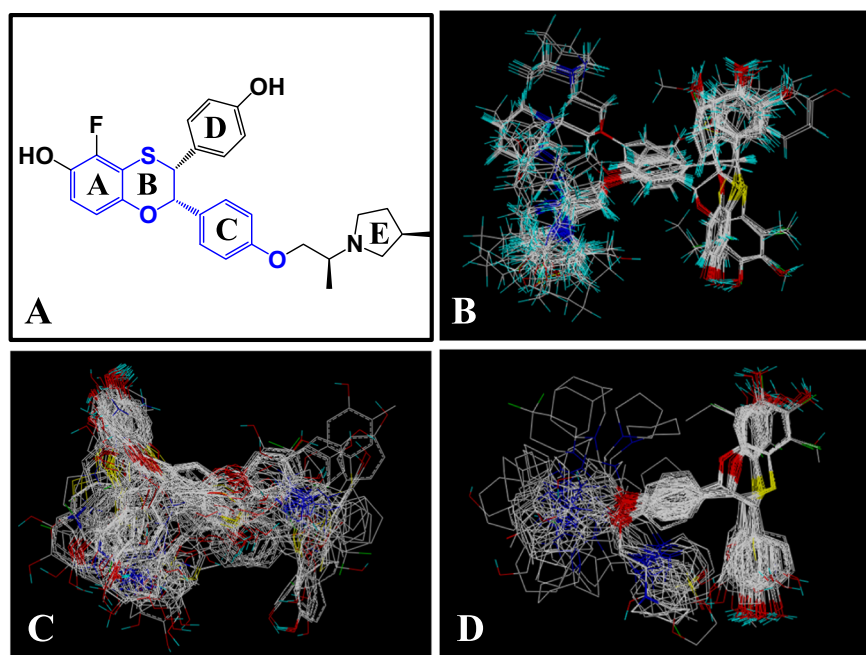
Fig. 2 a Compound 78 used as templates for template ligand-based alignment. The common substructure is shown in blue. **b–d** Present the alignments for ERα from the *Alignment 1*, *2* and *3*, respectively
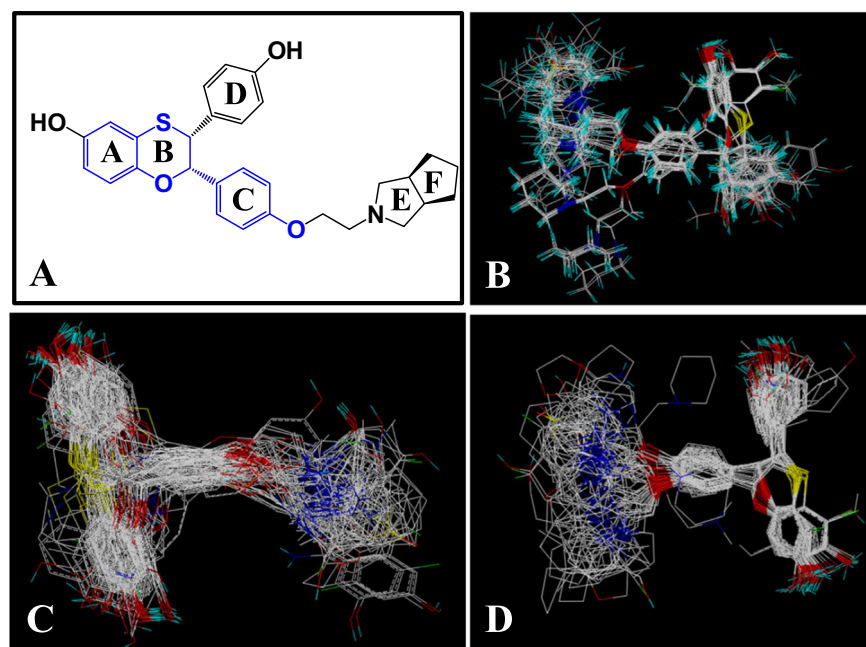
3D-QSAR studies

In the present work, CoMFA and CoMSIA methods were performed upon the alignments to correlate the structures with the inhibitory activity. The algorithms for CoMFA and CoMSIA techniques have been demonstrated in many literatures (Murumkar et al. 2010; Srivastava et al. 2010), therefore, only the different parameters used in the present work were introduced.

The steric field energies in the form of a Lennard–Jones 6-12 function and the electrostatic fields in form of Coulomb function were calculated at each lattice intersection point of a regularly spaced grid of 2.0 Å. An sp^3 carbon atom probe with a van der Waals radius of 1.52 Å and a charge of $+1.0$ was served as the probe atom to compute the CoMFA steric and electrostatic fields. And the energy cutoff values were set to 30 kcal/mol (Ståhle and Wold 1988). The generated steric and electrostatic fields were

further scaled by CoMFA standard requirements given in Sybyl.

CoMSIA exhibits greater robustness than CoMFA because of its feature of avoiding singularities at the atomic positions arising from the Lennard–Jones and dramatic changes of Coulomb potentials (Pirhadi and Ghasemi 2010). For CoMSIA method, the five different fields: steric (S), electrostatic (E), hydrophobic (H), hydrogen bond donor (D) and hydrogen bond acceptor (A) were calculated using the same probe atom as that used in CoMFA analysis. The attenuation factor and the column filtering were set to 0.3 and 2.0 kJ/mol, respectively, to speed up the analysis and reduce the noise.

PLS regression was adopted to derive linear relationship between CoMFA/CoMSIA descriptors and inhibitory activity values (Cramer et al. 1988; Goyal et al. 2014). Initially, the leave-one-out method was used in the cross-validation analysis to produce the cross-validated correlation coefficient R^0_{cv}, the lowest standard error of prediction and the optimal number of principal components (Nc). Then, non-cross-validation analysis was performed to calculate conventional R^2_{ncv} using the Nc obtained from the cross-validation analysis.

However, the cross-validated correlation coefficient R^2_{cv} alone is insufficient to estimate the capability of the derived QSAR models. Therefore, external validation should be conducted. The predictive ability of the QSAR models was evaluated with the test set. The predictive correlation (R^2_{pred}) was computed as follows:

$$R^2_{pred} = 1 - \frac{PRESS}{SD}, \tag{1}$$

where PRESS is the sum of squared deviation between the actual activity and predicted activity values for the test compounds; SD is the sum of squared deviations between the inhibitory activities of the test set and the mean activities of the training set compounds.

Finally, the CoMFA/CoMSIA results were also graphically interpreted by contour maps which were generated varying the spacing in the box and the weight for the standard deviation (StDev × Coeff) in relation to the parameters of the previous model. In addition, the most active inhibitors were set as template to validate the contour maps. The default value of contour by contribution, 80% for favored region and 10% for disfavored region, was set during contour map analysis.

Molecular docking

Molecular docking simulations were performed using AutoDock (version 4) to explore the interaction mechanism between inhibitors and receptors ERβ and ERα (Morris et al. 1998), the crystal structure 1NDE (Henke et al. 2002) and

1R5K (Wu et al. 2005), were extracted from the RSCB Protein Data Bank (http://www.rscb.org/pdb). A grid-based docking program was employed to analyze the binding modes. The AutoDock scoring function is described by Morris et al. (1998). The interaction energy was evaluated using atom affinity potentials calculated on a grid similar to that described by Goodford (1985). Prior to molecular docking, the original compounds and water molecules were removed from the crystal structure, then, polar hydrogen and united atom Kollman charge were assigned for the receptor.

For molecular docking, the LGA search method was used by randomizing initial position, orientation, and relative dihedrals. The grid box defining the binding search space was input as x center 107.542, −7.885, y center 6.849, 48.829, and z center −107.145, 332.15, respectively for ERβ and ERα. In addition, the maximum number of 10,000 retries and 27,000 generations were set for docking. Finally, the conformations with the lowest binding energy were extracted and aligned together for further QSAR analysis.

Calculation and selection of Dragon descriptors

In the present work, Dragon Professional, version 5.0 software (Todeschini et al. 2005) (http://www.talete.mi.it/index.htm) was employed to enhance the robustness and generalization of the derived QSAR models. For each block of descriptors, the constant variables were firstly removed. For the remaining descriptors, pairwise correlation analysis was conducted in R software (www.r-project.org) to exclude those ones with the standard deviation >0.5 (Wang et al. 2008). The HOMA and MATS3m descriptors for ERβ, and X1A and nPyridines descriptors for ERα showed important influences with $R^2 = 0.62$ and $R^2 = 0.68$.

Results and discussion

In the present work, different charges (Gasteiger–Huckel, Gasteiger–Marsili, MMFF944, and Huckel) were tried to obtain optimal models. It is clear that the models based on Gasteiger–Huckel charge exhibited better statistical results (Tables S1 and S12).

In addition, the QSAR models based on *Alignment 1* are superior to those based on *Alignment 2* and *Alignment 3*. Obviously, some models based on *Alignment 2* also have high R^2_{cv}, but the external predictive power is low. For *Alignment 3*, the results indicate that the derived models are not robust, and the points for the test set distribute unevenly around the regression line. The reason for this phenomenon is that inhibitors possessing different binding activities would exhibit different orientations in the binding site during molecular docking, especially the different substituents in the structure, leading to the results of alignment

Table 2 Statistical data of optimal QSAR models based on different superimposition methods for ERβ

Parameters	I		II		III	
	CoMFA	CoMSIA	CoMFA	CoMSIA	CoMFA	CoMSIA
Superimposition methods						
R^2_{cv}	0.578	0.653	0.571	0.528	0.361	0.468
R^2_{ncv}	0.878	0.866	0.918	0.773	0.496	0.669
SEE	0.204	0.214	0.163	0.273	0.399	0.324
F	63.417	56.983	154.948	46.835	28.050	57.531
R^2_{pred}	0.5341	0.615	0.4629	0.5596	0.4545	0.5017
SEP	0.379	0.343	0.375	0.393	0.450	0.410
N_c	6	6	4	4	2	2
Field contribution						
S	0.366	0.254	0.183	0.209	0.102	–
E	0.395	–	0.598	–	0.140	0.150
H	–	0.478	–	0.436	–	0.082
D	–	–	–	–	–	0.151
A	–	–	–	–	–	0.135
MATS3m	0.127	0.126	0.094	0.168	0.416	0.283
HOMA	0.112	0.142	0.125	0.187	0.342	0.199

R^2_{cv} = Cross-validated correlation coefficient using the leave-one-out methods, R^2_{ncv} = Non-cross-validated correlation coefficient, F = Ratio of R^2_{ncv} explained to unexplained = $R^2_{ncv}/(1-R^2_{ncv})$, R^2_{pred} = Predicted correlation coefficient for the test set of compounds, N_C = Optimal number of principal components

SEE standard error of estimate, *SEP* standard error of prediction, *S* steric, *E* electrostatic, *H* hydrophobic, *D* H-bond donor, *A* H-bond acceptor

are worse than the template ligand-based alignment. Therefore, the results obtained from *Alignment 1* are better than *Alignment 2* and *Alignment 3*.

3D-QSAR statistical results for ERβ

The CoMFA results for ERβ are summarized in Table 2, showing that the CoMFA model has high R^2_{cv} (0.578), F (63.417), and a small SEE (0.204), as well as non-cross-validated correlation coefficient R^2_{ncv} (0.878) with six optimum number of components, proving a good correlation between the experimental and the predicted activities (as shown in Fig. 3a). The statistical indexes indicate that the derived CoMFA model has strong predictive ability. The corresponding steric and electrostatic fields explain 36.6% and 39.5% of the total variance, respectively. In addition, molecular descriptors MATS3m (Caballero et al. 2008) and HOMA (Rodríguez-Fortea and Poblet 2014) also make 12.7 and 11.2% contribution to the model, which effectively enhance the fitting degree of the model. MATS3m is the Moran autocorrelation of lag 3 weighted by atomic masses belonging to 2D autocorrelation descriptors, derived directly from the molecular structure and the computation involves the summations of the autocorrelation functions corresponding to different fragment lengths. The geometrical descriptor HOMA is sensitive to the C–C bond distance. The definition of the descriptors is listed in Table 3.

The predictive capability of the CoMFA model is checked using the test set. Examination of the residuals between the actual and the predicted values suggests that compound 62 might be outlier. The outlier status is due to the large residual value which is nearly up to 2.0 log unit. Moreover, the predicted correlation coefficient R^2_{pred} (0.5341) represents that the prediction ability of the model is good.

Different models were generated by different combination of the five fields, as shown in Table S1. Among them, CoMSIA-SH model is superior to other models. The optimum model is built on six components and the statistical parameters are listed as follows: $R^2_{cv} = 0.653$, $R^2_{ncv} = 0.866$, SEE = 0.214, $F = 56.983$. It is also noted that the hydrophobic field makes a higher contribution (47.8%) to the inhibitory activity than that of the steric field (25.4%). Furthermore, the dragon descriptor MATS3m and HOMA also explain 12.6% and 14.2% of the total variance, which play significant roles in predicting the inhibitory activity.

In addition, the test set is used to evaluate the predictive accuracy of the CoMSIA model. The R^2_{pred} (0.615) confirms that the CoMSIA model is highly predictive. Similar to the CoMFA model, compound 62 is also considered to be outlier due to its low inhibitory ability. Figure 3b depicts the correlation between the observed and predicted pIC$_{50}$ values for the training set and the test set, and most of the points are uniformly distributed around the line, indicating that the built model has excellent predictive ability.

3D-QSAR statistical results for ERα

The same training and test set are used to derive CoMFA and CoMSIA models. The statistical details are summarized in Table 4. The results show that the optimal CoMFA model provides a leave-one-out R^2_{cv} of 0.676 (>0.5) with an optimal number of principal components of six. A correlation coefficient R^2_{ncv} of 0.902 with a low SEE of 0.221, and an F value of 81.692 are also obtained. In contribution, the steric and electrostatic field contributes 34.0% and 44.2%, respectively. The best CoMSIA model provides R^2_{cv} of 0.693 with an N_C of six, an R^2_{ncv} of 0.913 with a low SEE of 0.209 and an F value of 92.264. In CoMSIA model, the contribution of the electrostatic, hydrophobic, X1A, and nPyridines are 43.3%, 30.5%, 7.6% and 14.2%, respectively. Based on the field contribution, the electrostatic field is the most significant field in the CoMSIA model. The X1A (Kier and Hall 1986) property is calculated from the hydrogen-depleted molecular graph, which describes the molecular structure in topological terms. Functional group counts nPyridines is the number of Pyridines (Durgapal et al. 2018).

External validation is also carried out to further assess the reliability and the predictive ability of the developed models. The R^2_{pred} of 0.631, 0.5758 are achieved for CoMFA and CoMSIA model, respectively. The graph of predicted pIC_{50} versus experimental pIC_{50} is shown in Fig. 3c, d, further verifying the excellent external predictive ability of the models.

3D-QSAR contour maps for ERβ

The visualization of the QSAR results as 3D coefficient contour maps is one of the most absorbing feature for CoMFA and CoMSIA models. The contour maps for ERβ are shown in Figs. 3 and 4. And the most potent inhibitor 11 was selected as reference structure.

CoMFA contour maps

The CoMFA steric contour map is shown in Fig. 4a, where the sterically favorable regions are represented in green and the unfavorable regions in yellow. A small green contour

Fig. 3 The correlation plots of the actual versus the predicted pIC_{50} values using the training set based on the CoMFA and CoMSIA models obtained from the activity for ERβ/α. Graphs of the predicted versus the experimental pIC_{50} values of the optimal models. **a** CoMFA model for ERβ. **b** CoMSIA model for ERβ. **c** CoMFA model for ERα. **d** CoMSIA model for ERα
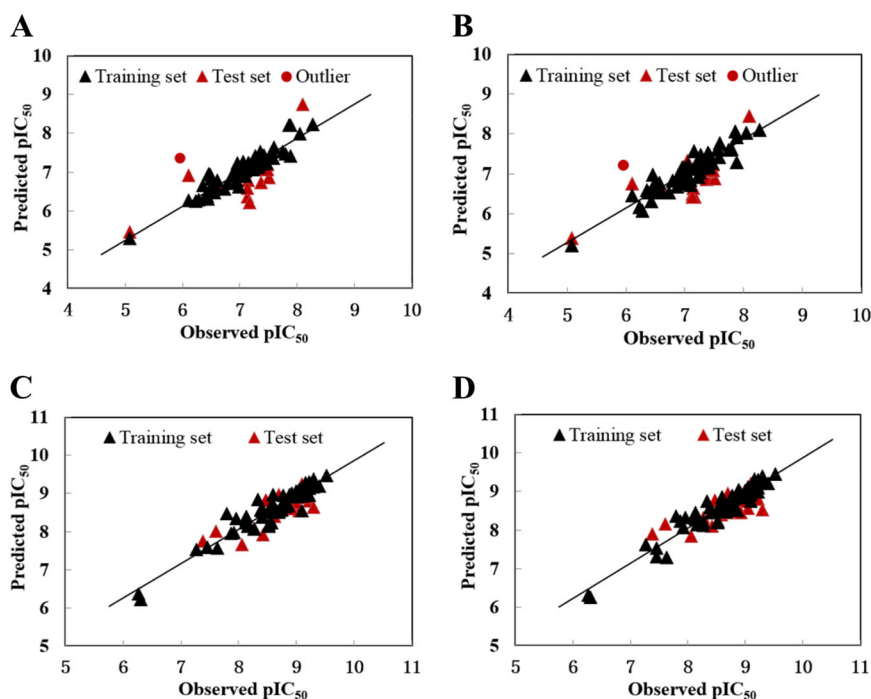
Table 3 Descriptors used in model construction

Symbol	Class	Meaning
MATS3m	2D autocorrelations	Moran autocorrelation-lag3/weighted by atomic masses
HOMA	Geometrical descriptors	Harmonic Oscillator Model of Aromaticity index
X1A	Connectivity indices	Average connectivity index chi-1
nPyridines	Functional group counts	Number of Pyridines

Table 4 Statistical data of optimal QSAR models based on different superimposition methods for ERα

Parameters	I		II		III	
	CoMFA	CoMSIA	CoMFA	CoMSIA	CoMFA	CoMSIA
Superimposition methods						
R^2_{cv}	0.676	0.693	0.561	0.557	0.557	0.559
R^2_{ncv}	0.902	0.913	0.631	0.632	0.630	0.654
SEE	0.221	0.209	0.411	0.410	0.411	0.401
F	81.692	92.264	99.059	99.585	98.961	53.887
R^2_{pred}	0.631	0.5758	0.2403	0.2614	0.2397	0.3246
SEP	0.403	0.392	0.448	0.450	0.450	0.453
Nc	6	6	1	1	1	2
Field contribution						
S	0.340	–	0.030	–	0.044	0.069
E	0.442	0.433	0.083	0.089	0.062	–
H	–	0.305	–	0.048	–	–
D	–	–	–	–	–	–
A	–	–	–	–	–	0.161
X1A	0.076	0.097	0.415	0.404	0.418	0.260
nPyridines	0.142	0.165	0.472	0.459	0.476	0.510

Fig. 4 CoMFA StDev × Coeff contour plots for ERβ ligands in combination of compound **11**. (**a** The steric contour map, where the green and yellow contours represent 80% and 20% level contributions, respectively. **b** The electrostatic contour map, where the blue and red contours represent 80% and 20% level contributions, respectively
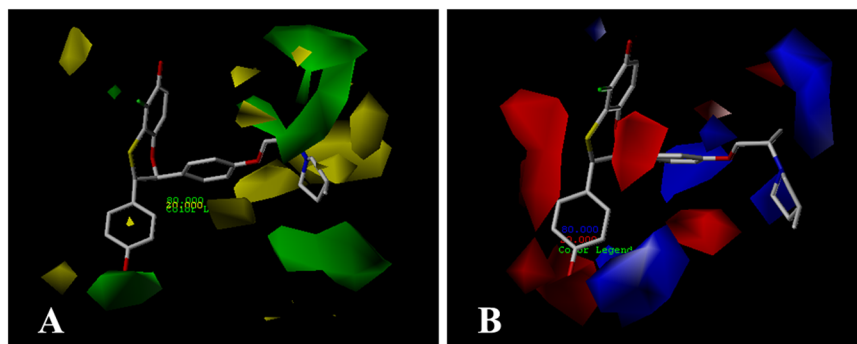
map located at the fluorine atom of ring A (Fig. 1a) indicates that bulky substituent is favored at this position, as illustrated by the fact that the inhibitory activity of compound 11 (F) is stronger than that of compound 10 (H), the same as compound 13 (F) and compound 12 (H). Another green contour map is covered the para position of ring D, indicating that this position is suitable for larger substituent. The higher activity of compound 22 (–CH$_2$CH(CH$_3$)$_2$) than compound 20 (i-Pr) is an example. A large green contour is mapped near ring E, illustrating that bulkier group at this position may increase the activity, as illustrated by the fact that the pIC$_{50}$ value of compound 59 (–N⟨⟩) is higher than that of compound 58 (–N⟨⟩F). For compound 48 and 49, the conformation of ring E is altered and extended to a series of yellow contour maps, meaning that the activity would be decreased when introducing large substituent. This can be explained by the fact that compound 48 having pyrrolidine

group, favored the inhibitory activity whereas in compound 49 harboring piperidine.

For electrostatic contour map (Fig. 4b), the blue regions indicate the favorable effect of positively charged groups, whereas the red contours indicate the positions in which electronegative groups may enhance the activity. A red contour near the substituent of ring D indicates that groups with negative charges may increase the activity. This can explain why the activity of compound 46 (F) is higher than compound 45 (OMe). In addition, compound 47 and 1 possessing electronegative substituents at this position exhibit higher activity than compound 46 and 14. A small blue contour map is present around ring A, suggesting that electropositive potential would enhance the inhibitory activity. Therefore, compound 31 holding –OH group decrease the activity compared with compound 30 (Et) and compound 29 (Me). In addition, the most potent compound

Fig. 5 CoMSIA StDev × Coeff
contour plots for ERβ ligands in
combination of compound **11**. **a**
The steric contour map, where
the green and yellow contours
represent 80% and 20% level
contributions, respectively. **b**
The hydrophobic contour map,
where the yellow and white
contours represent 80% and 20%
level contributions, respectively
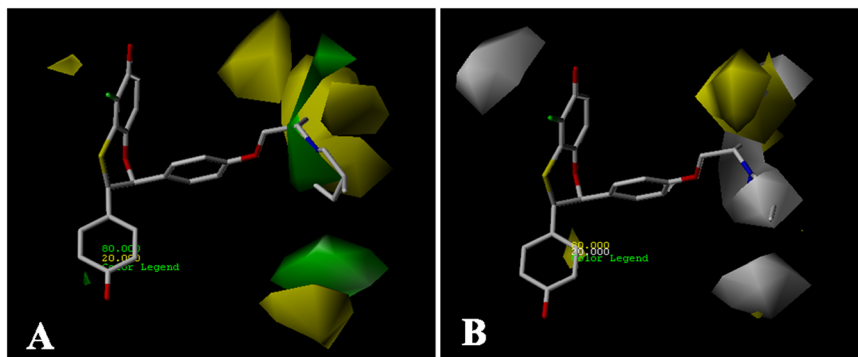
Fig. 6 CoMFA StDev × Coeff
contour plots for ERα ligands in
combination of compound **78**. **a**
The steric contour map, where
the green and yellow contours
represent 80% and 20% level
contributions, respectively. **b**
The electrostatic contour map,
where the blue and red contours
represent 80% and 20% level
contributions, respectively
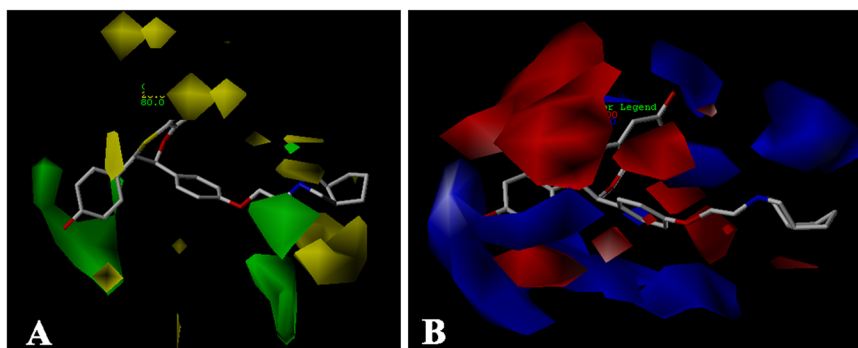
11 possesses electronegative substituent at this position, therefore, the activity can be further enhanced by modifying the structure. A series of blue contour maps found near ring E indicates that groups with positive charges may increase the activity. This can explain why compound 1 shows higher activity than compounds 55–67.

CoMSIA contour maps

The steric contour maps of CoMSIA (shown in Fig. 5a) are similar to those of CoMFA steric fields (Fig. 4a), thus not discussed here.

Hydrophobic contour maps are shown in Fig. 5b, where the yellow region indicates the favorable region for hydrophobic groups and the white region indicates an unfavorable region for hydrophobic groups. A yellow contour is present near ring D, illustrating a better ERβ activity if with properly hydrophobic substituent, which could be verified by compound 14 and 15, the former activity (⟋–◯) is larger than the latter (⟋–◯–OH). Around ring A, a white contour map indicates that hydrophilic groups are beneficial to the inhibitory activity. That is why hydrophilic –OH is applied to the most potent compound. Additionally, a white contour map is situated near ring E, therefore, compound 50 with cycloheptylamine tends to be higher than compound 49 with piperidine group. In addition, for compound 56 and 57, the orientation of the

substituent at ring E is changed, then extends into a yellow contour map, exactly explicate the higher potency of compound 55 with –N◯S than compound 56 with –N◯SO₂.

3D-QSAR contour maps for ERα

Similar to ERβ models, the most potent compound 78, which represents the general structure of the inhibitors was selected to analyze the contour maps.

CoMFA contour maps

The steric contour plots are depicted in Fig. 6a. Green maps refer to areas where bulky substituent would enhance the activity while yellow areas represent the opposite. In the electrostatic field maps (Fig. 6b), blue and red contour maps indicate regions where electron-donating and electron-withdrawing groups are favorable for the inhibitory activity, respectively.

As interpreted in Fig. 6a, there are some yellow contours falling around ring A, indicating a greater influence on ERα inhibitory activity. Accordingly, properly reducing the group volume would favor the activity. Therefore, compound 28 holding hydrogen atom in ring A possesses higher inhibitory activity than compound 34 (–Cl) and compound 35 (–Me). A large green polyhedron mapping around ring D is hypothesized that increased bulk of the substituent would

Fig. 7 CoMSIA StDev × Coeff contour plots for ERα ligands in combination of compound 17. **a** The electrostatic contour map, where the blue and red contours represent 80% and 20% level contributions, respectively. **b** The hydrophobic contour map, where the yellow and white contours represent 80% and 20% level contributions, respectively
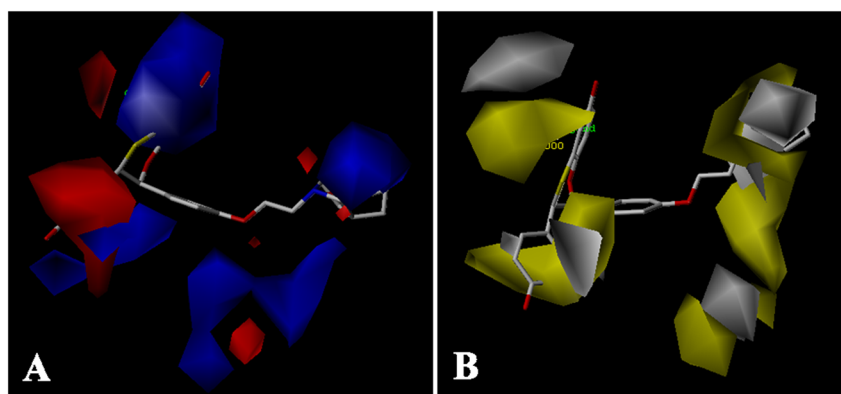
be beneficial. This can be explained by the fact that compound 22 having –CH$_2$CH(CH$_3$)$_2$, favored the activity, while in compound 20 having *i*-Pr at the substituent of ring D. The same pattern is followed in the case of compounds 14 and 15 (compound 15 = ⌇⟨⟩-OH > compound 14 = ⌇⟨⟩). A big yellow region surrounding ring E and ring F means that the activity would be decreased when introducing large substituents. This can be explained by the fact that compound 48 having pyrrolidine group, favor the biological activity, whereas in inhibitor 49 harboring piperidine group. Furthermore, the activity of compound 60 (–N⟨⟩-OH) is higher than compound 61 (–N⟨⟩-OH). At the linker between ring C and ring E, a green contour map indicates that bulky group would increase the inhibitory activity. This situation could be justified by comparing the activity of compound 7 (⌇-○-⟨) and compound 4 (⌇-○-⌇).

For the electrostatic filed (Fig. 6b), some blue contour maps (electropositive favorable) are observed around ring A. Comparison between compound 28 and compound 31 suggests that the hydroxyl group is less favorable. However, most of compounds applied in the present work have electronegative substituents, thus modification can be made at this position to improve the binding activity. One region of blue contour near ring E and ring F indicates that the presence of electropositive substituent is important for the binding activity. For example, compound 1 exhibits higher activity than compounds 55–67. In addition, a red electronegative region flanking the para position of ring D indicates that substitution of an electronegative group at this position would enhance the activity. This is the reason why compound 44 (H) and 46 (F) display difference in the activity. In case of compound 14 and 15, the activity is reduced from the ⌇⟨⟩-OH to ⌇⟨⟩. Furthermore, electropositive blue contour suggest that the polar group is favorable for the inhibitory potency. It can be seen that a blue contour is located around the above red contours, therefore, electropositive substituents with bulky volume

extending to this blue contour map are favorable for the activity.

CoMSIA contour maps

As shown in Fig. 7a, the electropositive favorable regions are represented in blue and electronegative regions in red. It can be observed that the electrostatic contour map of CoMSIA is similar to that of CoMFA (Fig. 6b), thus, only hyrophobic field will be discussed below.

The hydrophobic field is shown in Fig. 7b, yellow and white contours highlight areas where hydrophobic and hydrophilic properties are favored. The white contour at –OH of ring A indicates that this position is suitable for hydrophilic group, which explains the higher activity of compound 1 with hydrophilic –OH group than compound 34 (Cl) and compound 35 (Me). However, we also find a yellow region located above ring A, meaning that hydrophobic group can lead to high potency. At the same time, another yellow part is found surrounding ring D, indicating that hydrophobic groups are beneficial to the inhibitory activity. Things can get much clearer when the phenomenon is observed that compound 46 with hydrogen group is more potent than compound 47 (hydroxyl group). Besides, a white region is located around ring E and ring F, meaning that hydrophilic substituent is favorable here, followed by the illustration that compound 50 possesses higher potency than compound 49 for the reason that compound 50 has cycloheptylamine group, and compound 49 possesses piperidine substituent. Moreover, the result can also be applied to the comparison of compounds 57 (–N⟨⟩-F) and 60 (–N⟨⟩-OH) (compound 60 > compound 57).

In conclusion, according to the results of the 3D-QSAR analyses undertaken in the present work, the structural features of the inhibitors are quite compatible with the contour maps.

Fig. 8 **a** The ERβ active site amino acid residues around compound **11**. **b** The enlargement for the ligand in the binding site after molecular docking, which is displayed in stick, H-bonds are shown as dotted red lines, and the nonpolar hydrogens were removed for clarity
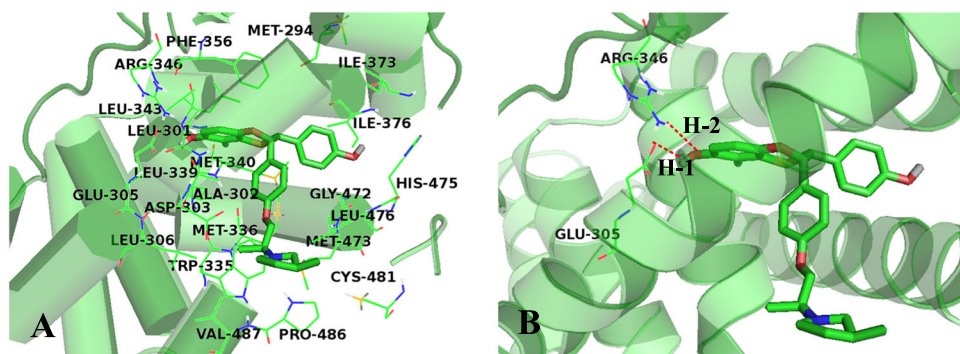
Fig. 9 **a** The ERβ active site amino acid residues around compound **17**. **b** The enlargement for the ligand in the binding site after molecular docking, which is displayed in stick, H-bonds are shown as dotted red lines, and the nonpolar hydrogens were removed for clarity
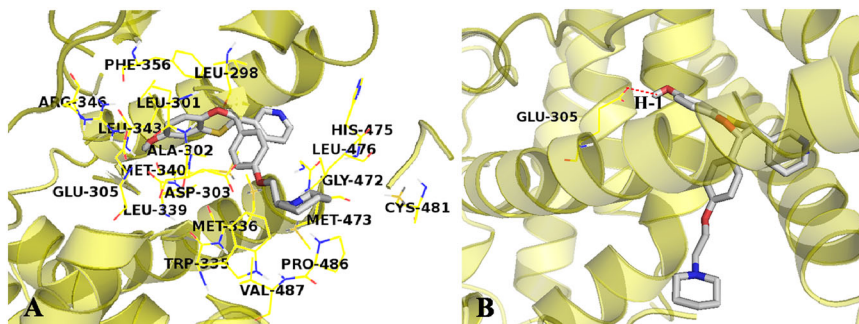
Molecular docking studies

To explore the probable ligand binding mode at the binding site, molecular docking study was carried out. In the present work, the highest active compounds and the lowest active compounds were selected for detailed analysis.

For ERβ

Compound **11** (the most potent) and compound **17** (the least active) were docked in the binding pocket of ERβ. The docked conformation of the inhibitors are shown in Figs. 8 and 9.

Docking results demonstrate that compound **11** is placed deep into the ERβ cavity enclosed by hydrophobic residues (Leu298, Leu301, Ala302, Leu306, Trp335, Met336, Leu339, Met340, Leu343, Phe356, Ile373, Ile376, Met473, Leu476, Pro486, and Val487) and hydrophilic residues (Asp303, Glu305, Arg346, Gly472, His475, Cys481) (shown in Fig. 8a). In addition, the hydroxyl group of ring A resides against the active pocket, and forms two hydrogen bonds with Glu305 and Arg346. The –OH of ring A interacts with the O atom of Glu305 (–O···HO, 2.20 Å, 163.9°) (H-1). The –OH of ring A forms hydrogen bond with Arg346 (–O···HN, 2.26 Å, 155.6°) (H-2) (Fig. 8b).

The substituent at ring A binds to a wide pocket (Fig. 8a), consisting of residues Leu339, Leu343, and Met340, which is in consistence with the CoMFA steric contour maps (Figs. 4a and 5a). Obviously, the group at ring D

extends into a large pocket bordered by residues Met336, Ile373, Ile376, Met473, His475, and Leu476, whose interaction resembles with the steric contour maps. Molecular docking studies also suggest that the substituents at ring E are almost oriented outside the active pocket, illustrating that bulky groups are accommodated, which is validated by the two green contour maps at this position in the CoMFA and CoMSIA models.

The hydroxyl group of ring D is faced with electropositive residue His475, indicating that electronegative groups are beneficial to the inhibitory activity, this is evident from the presence of a red contour map at this position. The hydroxyl group of ring A is involved in hydrogen bond interactions with the electronegative residue Gul305. This observation implies that electropositive substituents can improve the inhibitor potency, which is consistent with the blue contour map (Fig. 4b).

It is clear that the group at ring D is surrounded by hydrophobic residues (Met336, Ile373, Ile376, Met473, and Leu476), as indicated in Fig. 8a. The docked model reveals that the hydroxyl group of ring A forms hydrogen bond with the hydrophilic amino acid Glu305, where hydrophilc groups would favor the interaction between inhibitors and ERβ. These observations can be compared with the hydrophobic contour maps listed in Fig. 5b.

As can be seen in Fig. 9, the lowest compound **17** is accommodated in the active site composed of hydrophobic residues Leu298, Leu301, Ala302, Trp335, Met336, Leu339, Met340, Leu343, Phe356, Met473, Leu476,

Fig. 10 a The superposition of ERβ-11 (green) and ERβ-17 (cyan). **b** The superposition of ERα-87 (blue) and ERα-17 (purple)
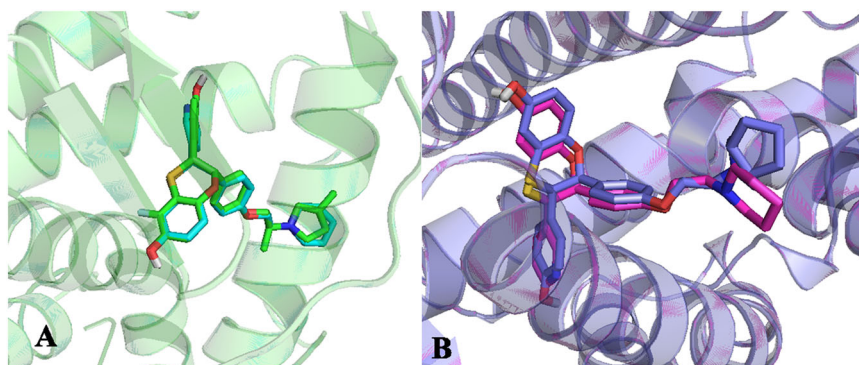
Fig. 11 a The ERα active site amino acid residues around compound 78. **b** The enlargement for the ligand in the binding site after molecular docking, which is displayed in stick, H-bonds are shown as dotted red lines, and the nonpolar hydrogens were removed for clarity
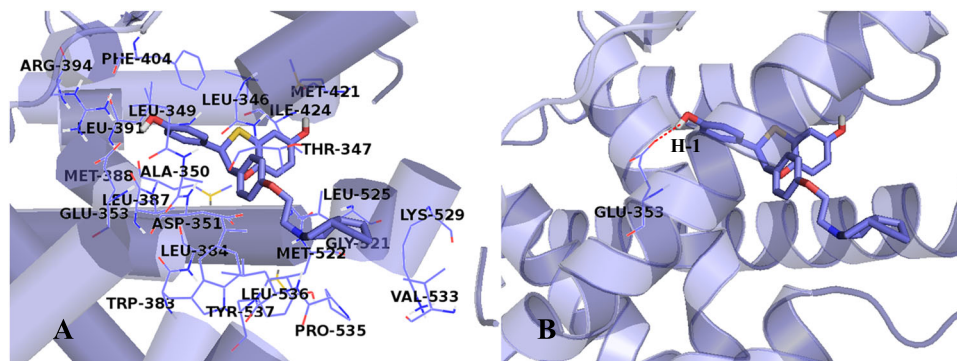
Pro486, Val487, and hydrophilic amino acids Asp303, Glu305, Arg346, Gly472, His475, and Cys481. This compound forms one hydrogen bond with the backbone of Glu305 (–O···HO, 2.16 Å, 164.3°) (H-1).

To gain insight on interactional distinction of the two complexes, the comparison was provided between compound 11 and compound 17 (as shown in Fig. 10a). Firstly, inhibitors 11 and 17 have the similar scaffold, and the superimposition of the two compounds in the binding pocket indicates that both compounds are well positioned in the binding pocket in the same way and display a similar binding mode. However, three structural differences are also existed for inhibitor 11 and inhibitor 17, which is described as follows: the fluorine atom (at ring A) of inhibitor 11 is absent in inhibitor 17; the phenol group of inhibitor 11 is substituted by the pyridine moiety of inhibitor 17; the 4-methylpiperidine group at ring C of inhibitor 11 is substituted by the piperidine moiety of inhibitor 17. These diversities make the hydroxyl group (at ring D) of compound 11 forming more stable electrostatic interaction with residue His475. As can be seen from Fig. 9, the altered substituent makes it far away from His475, and the electrostatic interactions are absent for compound 17. In addition, the orientation of ring E for compound 11 and compound 17 is changed, therefore, the conformation of the entire compound is shifted outward toward the binding cavity as compared with compound 11, which is detrimental

to ligand–receptor interaction, thus the binding activity of compound 17 is lower than compound 11.

For ERα

To study the detailed binding mode between inhibitor and ERα, molecular docking was performed with the lead inhibitor 78. As shown in Fig. 11a, the compound is positioned towards the hydrophobic amino acids (Leu346, Leu349, Ala350, Trp383, Leu384, Leu387, Met388, Leu391, Phe404, Met421, Ile424, Met522, Leu525, Val533, Pro535, and Leu536) and hydrophilic residues (Thr347, Asp351, Glu353, Arg394, Gly521, Lys529, and Tyr537). The ligand is anchored in the binding site perhaps via one hydrogen bond. The hydroxyl group of ring A acts as a donor to form hydrogen bond with the oxygen atom of Glu353 (–O···HO, 1.89 Å, 155.6°) (H-1).

Docking results show that substituent at ring A is buried into a large cavity composed of residues Met421, Ile424, Met522, and Leu525, indicating that the steric interaction would be favorable for the activity, which is similar to the above CoMFA contour map analysis (Fig. 6a). Meanwhile, the group at ring D is also fitted into a large pocket, suggests that bulky groups are favorable for the ligand–receptor interactions, evidenced by the green contour map. The substituents on the ring E and ring F are enclosed by residues Val533, Pro535, and Leu536, which is a small pocket,

suggests that large groups would cause steric hindrance with surrounding amino acids.

The docking investigations also reveal that the neutral environment (Met421, Ile424, Met522, and Leu525) would contribute favorable electrostatic interaction with the electropositive groups at ring A, thereby helping them to position into a stable conformation, proved by the blue contour map (Figs. 6b and 7a). The groups in ring E and ring F are oriented toward neutral amino acids Leu536, Pro535, Val533, and electronegative residue Asp351, which coincides with the blue contour map explained in CoMFA and CoMSIA.

One yellow contour map is situated above the hydroxyl group of ring D (Fig. 7b), which is close to the hydrophobic residues Met421, Ile424, Met522, and Leu525, indicates that hydrophobic groups would increase the inhibitory activity.

Moreover, compound 17 exhibits lower potency when binding to the receptor ERα, which is in agreement with the docking score. Docking results reveal that this compound binds to the same position of the receptor when compared with compound 87 (shown in Fig. 10b), in which residues Leu346, Thr347, Leu349, Ala350, Asp351, Glu353, Trp383, Leu384, Leu387, Met388, Leu391, Arg394, Phe404, Met421, Ile424, Phe425, Gly521, Leu525, Pro535, Leu536, and Tyr537 construct the binding pocket and interact with the inhibitor (Fig. 12).

The reasons for the activity discrepancy for compounds 78 and 17 are shown as follows: (1) for compound 17, ring D moiety possesses no hydrogen bond donor group, therefore, cannot form hydrogen bond interactions with the receptor; (2) at the terminal position of ring C, the two compounds undergo large conformation change, leading to
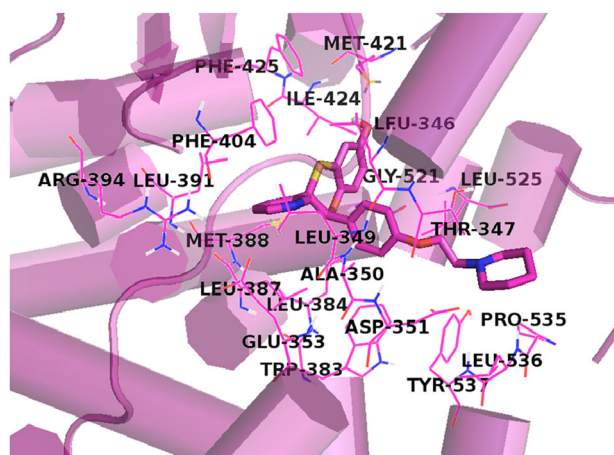
the unfavorable electrostatic interaction between compound 17 and the surrounding amino acids.

Selectivity for ERβ and ERα

Comparison of receptor binding activity

The correlation of the two subtype binding activity was investigated to analyze the selectivity. A mathematical model was constructed on the ERα and ERβ binding activity, as shown in Fig. 13. The correlation coefficient R^2 is 0.5312, suggesting that the activity for ERβ is in positive correlation with the activity for ERα, illustrating that the ERα binding compounds may also bind to the receptor ERβ. However, the potency of the binding activity is different. For compounds 11, 14, 21, 24, 25, 30, 37, and 39, the discrepancy of binding activity is lower than 1 when binding to ERα and ERβ. However, the other compounds possess stronger binding affinity to ERα. The situation of selectivity might originate from the receptor tissue distribution, which has been described in the section "Introduction". Therefore, a preferential interaction of related ligands selective for ERβ/ERα may mediate the divergent properties and reduce the occurrence of side effects.

Comparison of the proteins

In the present work, the selectivity was analyzed by comparing the sequence and structure. The docking complexes 1NDE-11 and 1R5K-78 were superimposed based on main-chain atoms. The sequence identity for ERα and ERβ is about 44.1%, the aligned RMSd is 0.747, with the Z-score of 8.5 (Fig. 14).

The inhibitor binding sites are shown in Fig. 14a, the superposition of the two complexes indicates that the inhibitors are located in the same binding site with similar pose. In addition, residue analysis of the two structures elucidates that the key residues Leu346 and Leu384 in the ERα

Fig. 12 a The ERα active site amino acid residues around compound 17. b The enlargement for the ligand in the binding site after molecular docking, which is displayed in stick, H-bonds are shown as dotted red lines, and the nonpolar hydrogens were removed for clarity
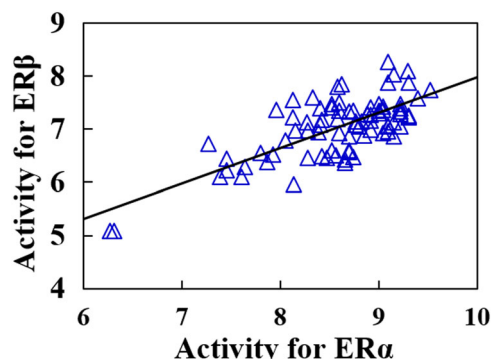

Fig. 13 A correlation plot of binding activities for ERβ and ERα of all compounds

Fig. 14 a The superposition of 1R5K-ERα (blue) and 1NDE-ERβ (green). **b** Alignments of the sequences of 1R5K and 1NDE for ERα and ERβ
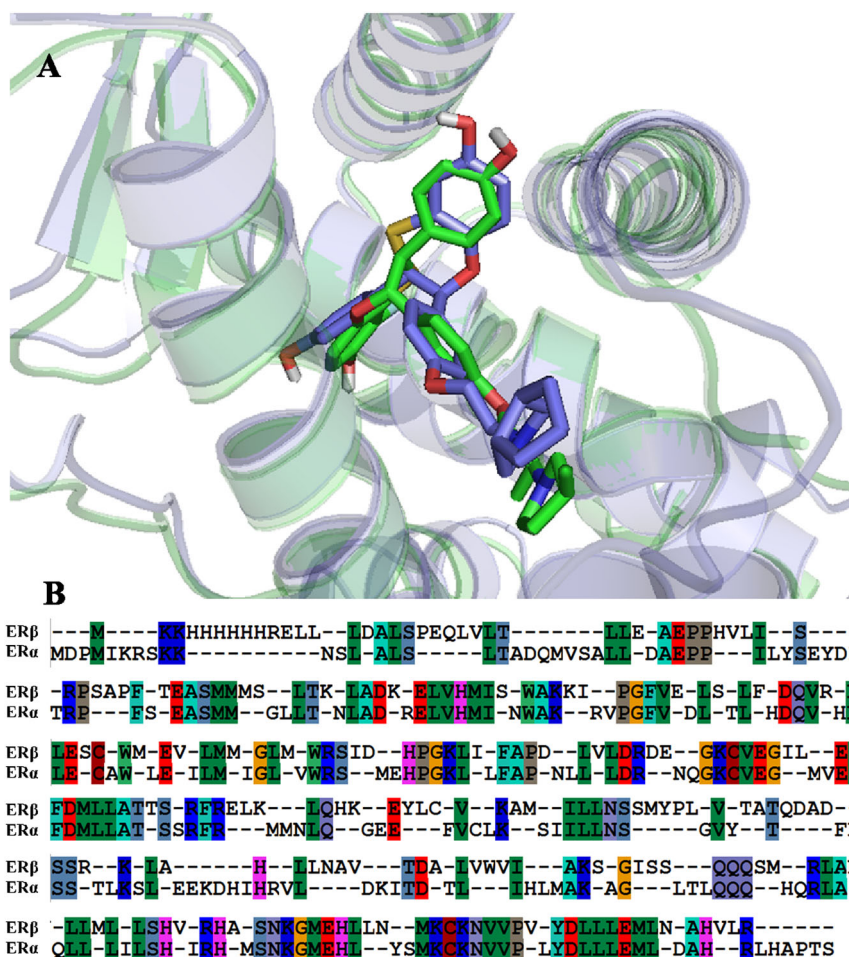
binding pocket is identical to ERβ. However, sequence analysis and molecular docking also reveal that obvious distinction is also existed for ERα and ERβ (Thr347/Ala302, Met388/Leu343, Leu391/Arg346, Met421/Ile376, Gly521/Leu476 for ERα, and ERβ, respectively), which may lead to the inhibitor selectivity.

Comparison of the QSAR results

The derived QSAR models for ERα and ERβ were also compared to understand the structural origin of the selectivity deeply. The statistical results are listed in Tables 2 and 4. The ERβ model shows similar contribution in the steric (36.6%) and electrostatic field (39.5%), while electrostatic contribution (44.2%) is higher than steric field (34.0%) for ERα, indicating that the electrostatic field is important for ERα. For CoMSIA models, the optimal QSAR models have been developed using different fields: steric and hydrophobic for ERβ, electrostatic and hydrophobic for ERα. In addition, the contribution of hydrophobic (47.8%) is higher than steric (25.4%) field, while the electrostatic (43.3%) is above than hydrophobic (30.5%) field for ERβ and ERα,

respectively. This illustrates that the interaction fields impacting the ligand binding might be different to ERβ and ERα. And the electrostatic field and the hydrophobic field are significant for ERα and ERβ, respectively.

Comparison of the contour maps

According to the QSAR and molecular docking results, the key structural features influencing the selectivity can be obtained. The structure–activity relationship is summarized (Fig. 15) as follows: (1) electropositive and hydrophilic groups at ring A, bulky, electronegative, and hydrophobic groups at ring D, electropositive and hydrophilic groups at ring E would be beneficial to the ERβ and ERα binding activity. (2) In addition, the differences are also appeared in the contour maps. ERβ has a green contour map around ring A while ERα is yellow here, which means that the binding activity of ERα can be improved by minifying the volume of the substituent at ring A. Another difference existed at the terminal position of ring F for ERα, the color is green for ERβ but yellow for ERα, further indicating that steric interaction is favorable for ERα, and unfavorable for ERβ.
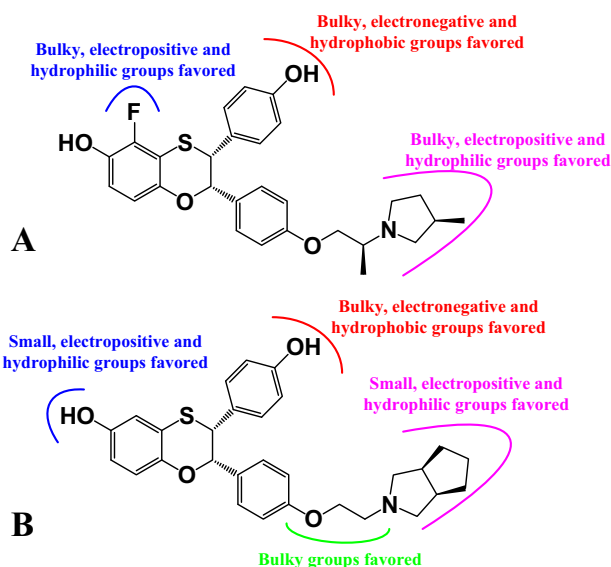
Fig. 15 a Structure–activity relationship revealed by QSAR studies for ERβ. **b** Structure–activity relationship revealed by QSAR studies for ERα

Comparison with the references

For this class of ER inhibitors, a series of computational models have been developed. In 2007, Salum et al. (2007) used both CoMFA and HQSAR models to describe the structure–activity relationship of ERα modulators. The results indicate the derived models are significant, which can be used for the design of novel ligands possessing high affinity and potency. The estrogen activity of 127 ERα modulators were collected by Wang et al. (2008) to establish QSAR models including the Bayesian-regularized neural network and multiple linear regression, which all had a better predictive power. In recent researches for virtual screening of ER ligands, pharmacophore and QSAR modeling as powerful tool have been built for ERβ ligands. And the developed models were used to screen the national center institute list of compounds to find new ERβ ligands (Taha et al. 2010). More recently, Huang et al. (2015) applied Discovery Studio sofware to build three-dimensional pharmacophore model for the ERα lignds and the ERβ ligands. Then the models were employed to virtual screening, and two leads (AH and AG) with novel scaffolds were discovered. However, these models are developed only for ERα, but not for ERβ when compared with the models constructed in our paper, and the selectivity problem is not analyzed.

Conclusion

QSAR models were conducted with 3D molecular descriptors for a series of ER inhibitors. Two types of

QSAR methods, CoMFA and CoMSIA, were used to investigate the relationship between the structure and inhibitory activity. In addition, molecular docking analysis was employed to elucidate the interaction mechanism and identify appropriate binding conformations interacting with ERβ and ERα. The main research conclusions are described as follows:

(1) CoMFA and CoMSIA models were developed based on three different alignment methods. Template ligand-based alignment derived QSAR models show satisfactory fitting ability and acceptable predictive ability. In addition, the effects of different charges were also evaluated, and the Gasteiger–Huckel charge was considered as the best in developing reliable QSAR models. For ERβ, hydrophobic descriptor plays a more significant role than steric and electrostatic descriptors. However, electrostatic interactions are found to be important field on ERα inhibitory activity. In addition, hydrogen bond interactions can also enhance the ERβ and ERα binding activity.

(2) Furthermore, for better understanding of the binding mode of inhibitors at the active pocket of ERβ and ERα, we conducted molecular docking simulations. As a result, some key residues Glu305, Met336, Leu339, Met340, Leu343, Arg346, Ile373, Met473, His475, and Leu476 are found to play critical role in maintaining the stability of ERβ-inhibitor 11, while amino acids Asp351, Glu353, Trp383, Met421, Ile424, Met522, Leu525, Val533, Pro535, and Leu536 are major factors for enhanced ERα binding activity.

(3) The ligand selectivity has been validated by different methods. Results indicate that substituents at ring A and ring F for ERα (ring E for ERβ), Thr347/Ala302, Met388/Leu343, Leu391/Arg346, Met421/Ile376, Gly521/Leu476 for ERα and ERβ, respectively, introduce selectivity. Therefore, the results of this study would provide useful guidelines for developing novel potent and selective inhibitors.

Compliance with ethical standards

References

Aranda A, Pascual A (2001) Nuclear hormone receptors and gene expression. Physiol Rev 81:1269–1304

Asikainen AH, Ruuskanen J, Tuppurainen KA (2004) Consensus kNN QSAR: a versatile method for predicting the estrogenic activity of organic compounds in silico. A comparative study with five estrogen receptors and a large, diverse set of ligands. Environ Sci Technol 38:6724–6729

Blizzard TA, Chen HY, Wu JY, Gude C, Kim S, Chan W, Birzin ET, Tien YY, Pai LY (2004a) Estrogen receptor ligands. Part 7: Dihydrobenzoxathiin SERAMs with bicyclic amine side chains. Bioorg Med Chem Lett 14:3861–3864

Blizzard TA, Dininno F, Chen HY, Kim S, Wu JY, Chan W, Birzin ET, Yang YT, Pai LY, Hayes EC (2005) Estrogen receptor ligands. Part 13: Dihydrobenzoxathiin SERAMs with an optimized antagonist side chain. Bioorg Med Chem Lett 15:3912–3916

Blizzard TA, Dininno F, Morgan JD, Wu JY, Chen HY, Kim S, Chan W, Birzin ET, Yi TY (2004b) Estrogen receptor ligands. Part 8: Dihydrobenzoxathiin SERAMs with heteroatom-substituted side chains. Bioorg Med Chem Lett 14:3865–3868

Bruning JB, Parent AA, Gil G, Zhao M, Nowak J, Pace MC, Smith CL, Afonine PV, Adams PD, Katzenellenbogen JA (2010) Coupling of receptor conformation and ligand orientation determine graded activity. Nat Chem Biol 6:837–843

Caballero J, Fernández M, González-Nilo FD (2008) Structural requirements of pyrido[2,3-]pyrimidin-7-one as CDK4/D inhibitors: 2D autocorrelation, CoMFA and CoMSIA analyses. Bioorg Med Chem 16:6103–6115

Chen HY, Kim S, Wu JY, Birzin ET, Chan W, Yi TY, Dahllund J, Dininno F, Rohrer SP, Schaeffer JM (2004) Estrogen receptor ligands. Part 3: The SAR of dihydrobenzoxathiin SERMs. Bioorg Med Chem Lett 14:2551–2554

Cho SJ, Tropsha A (1995) Cross-validated R2-guided region selection for comparative molecular field analysis: a simple method to achieve consistent results. J Med Chem 38:1060–1066

Clark M, Cramer RD, Van Opdenbosch N (1989) Validation of the general purpose Tripos 5.2 force field. J Comput Chem 10:982–1012

Cramer III RD, Bunce JD, Patterson DE, Frank IE (1988) Cross-validation, bootstrapping, and partial least squares compared with multiple regression in conventional QSAR studies. Quant Struct-Act Relatsh 7:18–25

Dilis R, Hanson RN, Nettles KW, Bruning JB, Gil G, O"Neill EE, Nowak J, Hughs A, Kim Y, Desombre ER (2007) Structural plasticity in the oestrogen receptor ligand-binding domain. EMBO Rep 8:563–568

Durgapal J, Bisht N, Alam M, Sharma D, Salman M, Nandi S (2018) QSAR and structure-based docking studies of aryl pyrido [2, 3-d] pyrimidin-7 (8H)-ones: an attempt to anticancer drug design. Int J Quant Struct-Prop Relatsh (IJQSPR) 3:43–73

Fang H, Tong W, Perkins R, Soto AM, Prechtl NV, Sheehan DM (2000) Quantitative comparisons of in vitro assays for estrogenic activities. Environ Health Perspect 108:723–729

Farzaneh S, Zarghi A (2016) Estrogen receptor ligands: a review (2013–2015). Sci Pharm 84:409–427

Fitzpatrick SL, Funkhouser JM, Sindoni DM, Stevis PE, Deecher DC, Bapat AR, Merchenthaler I, Frail DE (1999) Expression of estrogen receptor-β protein in rodent ovary. Endocrinology 140:2581–2591

Gasteiger J, Marsili M (1980) Iterative partial equalization of orbital electronegativity—a rapid access to atomic charges. Tetrahedron 36:3219–3228

Ghafourian T, Cronin MT (2005) The impact of variable selection on the modelling of oestrogenicity. SAR QSAR Environ Res 16:171–190

Goodford PJ (1985) A computational procedure for determining energetically favorable binding sites on biologically important macromolecules. J Med Chem 16:849–857

Goyal M, Dhanjal JK, Goyal S, Tyagi C, Hamid R, Grover A (2014) Development of dual inhibitors against Alzheimer's disease using fragment-based QSAR and molecular docking. Biomed Res Int 2014:1–12

Henke BR, Consler G, Go N, Hale RL, Hohman DR, Jones SA, Lu AT, Moore LB, Moore JT, Orband-Miller LA, Graham Robinett R, Shearin J, Spearing PK, Stewart EL, Turnbull PS, Weaver SL, Williams SP, Bruce Wisely G, Lambert MH (2002) A new series of estrogen receptor modulators that display selectivity for estrogen receptor β. J Med Chem 45:5492–5505

Huang W, Wei W, Yang Y, Zhang T, Shen ZR (2015) Discovery of novel selective ERα/ERβ ligands by multi-pharmacophore modeling and virtual screening. Chem Pharm Bull 63:780–791

Ji L, Wang X, Yang X, Liu S, Wang L (2008) Back-propagation network improved by conjugate gradient based on genetic algorithm in QSAR study on endocrine disrupting chemicals. Chin Sci Bull 53:33–39

Journé F, Body JJ, Leclercq G, Laurent G (2008) Hormone therapy for breast cancer, with an emphasis on the pure antiestrogen fulvestrant: mode of action, antitumor efficacy and effects on bone health. Exp Opin Drug Saf 7:241–258

Jun S, Jerome B, Katzenellenbogen JA, Katzenellenbogen BS (2003) Molecular basis for the subtype discrimination of the estrogen receptor-beta-selective ligand, diarylpropionitrile. Mol Endocrinol 17:247–258

Kier LB, Hall LH (1986) Molecular Connectivity in Structure Activity Analysis. Research Studies

Kim S, Wu j, Chen HY, Birzin ET, Chan W, Yang YT, Colwell L, Li S, Dahllund J, DiNinno F, Rohrer SP, Schaeffer JM, Hammond ML (2004) Estrogen receptor ligands. Part 4: The SAR of the syn-dihydrobenzoxathiin SERAMs. Bioorg Med Chem Lett 14:2741–2745

Korach KS, Sarver P, Chae K, Mclachlan JA, Mckinney JD (1988) Estrogen receptor-binding activity of polychlorinated hydroxybiphenyls: conformationally restricted structural probes. Mol Pharmacol 33:120–126

Kuiper GG, Carlsson B, Grandien K, Enmark E, Häggblad J, Nilsson S, Gustafsson JA (1997) Comparison of the ligand binding specificity and transcript tissue distribution of estrogen receptors α and β. Endocrinology 138:863–870

Kuiper GGJM, Lemmen JG, Carlsson B, Corton JC, Safe SH, Van Der Saag PT, Van Der Burg B, Gustafsson JA (1998) Interaction of estrogenic chemicals and phytoestrogens with estrogen receptor beta. Endocrinology 139:4252–4263

Marini F, Roncaglioni A, Novic M (2005) Variable selection and interpretation in structure-affinity correlation modeling of estrogen receptor binders. J Chem Inform Model 45:1507–1519

Meyers MJ, Sun J, Carlson KE, Katzenellenbogen BS, Katzenellenbogen JA (1999) Estrogen receptor subtype-selective ligands: asymmetric synthesis and biological evaluation of cis- and trans-5,11-dialkyl- 5,6,11, 12-tetrahydrochrysenes. J Med Chem 42:2456–2468

Meyers MJ, Sun J, Carlson KE, Marriner GA, Katzenellenbogen BS, Katzenellenbogen JA (2001) Estrogen receptor-beta potency-selective ligands: structure-activity relationship studies of diarylpropionitriles and their acetylene and polar analogues. J Med Chem 44:4230–4251

Miller CP, Collini MD, Harris HA (2003) Constrained phytoestrogens and analogues as ERβ selective ligands. Bioorg Med Chem Lett 13:2399–2403

Minutolo F, Macchia M, Katzenellenbogen BS, Katzenellenbogen JA (2011) Estrogen receptor β ligands: recent advances and biomedical applications. Med Res Rev 31:364–442

Morris GM, Goodsell DS, Halliday RS, Huey R, Hart WE, Belew RK, Olson AJ (1998) Automated docking using a Lamarckian genetic

algorithm and an empirical binding free energy function. J Comput Chem 19:1639–1662

Murumkar PR, Zambre VP, Yadav MR (2010) Development of predictive pharmacophore model for in silico screening, and 3D QSAR CoMFA and CoMSIA studies for lead optimization, for designing of potent tumor necrosis factor alpha converting enzyme inhibitors. J Comput-Aided Mol Des 24:143–156

Nilsson S, Gustafsson JÅ (2010) Estrogen receptors: their actions and functional roles in health and disease. Nucl Recept 8:91–141

Nilsson S, Mäkelä S, Treuter E, Tujague M, Thomsen J, Andersson G, Enmark E, Pettersson K, Warner M, Gustafsson J (2001) Mechanisms of estrogen action. Physiol Rev 81:1535–1565

Pirhadi S, Ghasemi JB (2010) 3D-QSAR analysis of human immunodeficiency virus entry-1 inhibitors by CoMFA and CoMSIA. Eur J Med Chem 45:4897–4903

Pons M, Gagne D, Nicolas JC, Mehtali M (1990) A new cellular model of response to estrogens: a bioluminescent test to characterize (anti) estrogen molecules. Biotechniques 9:450–459

Powel E, Shanle E, Brinkman A, Li J, Keles S, Wisinski KB, Huang W, Xu W (2012) Identification of estrogen receptor dimer selective ligands reveals growth-inhibitory effects on cells that co-express ERα and ERβ. PloS One 7:1–13

Reel JR, Iv JCL, Neal BH (1996) Survey and assessment of mammalian estrogen biological assays for hazard characterization. Fundam Appl Toxicol Off J Soc Toxicol 34:288–305

Rodríguez-Fortea A, Poblet JM (2014) Maximum aromaticity or maximum pentagon separation; which is the origin behind the stability of endohedral metallofullerenes? Faraday Discuss 173:201–213

Shelby MD, Newbold RR, Tully DB, Chae K, Davis VL (1996) Assessing environmental chemicals for estrogenicity using a combination of in vitro and in vivo assays. Environ Health Perspect 104:1296–1300

Salum LB, Polikarpov I, Andricopulo AD (2007) Structural and chemical basis for enhanced affinity and potency for a large series of estrogen receptor ligands: 2D and 3D QSAR studies. J Mol Graph Model 26:434–442

Soto AM, Sonnenschein C, Chung KL, Fernandez MF, Olea N, Serrano FO (1995) The E-SCREEN assay as a tool to identify estrogens: an update on estrogenic environmental pollutants. Environ Health Perspect 103:113–122

Srivastava V, Gupta S, Siddiqi M, Mishra B (2010) 3D-QSAR studies on quinazoline antifolate thymidylate synthase inhibitors by CoMFA and CoMSIA models. Eur J Med Chem 45:1560–1571

Ståhle L, Wold S (1988) Multivariate data analysis and experimental design in biomedical research. Prog Med Chem 6:291–338

Suzuki T, Ide K, Ishida M, Shapiro S (2001) Classification of environmental estrogens by physicochemical properties using principal component analysis and hierarchical cluster analysis. J Chem Inform Comput Sci 41:718–726

Taha MO, Tarairah M, Zalloum H, Abu-Sheikha G (2010) Pharmacophore and QSAR modeling of estrogen receptor β ligands and subsequent validation and in silico search for new hits. J Mol Graph Model 28:383–400

Thaimattam R, Daga PR, Banerjee R, Iqbal J (2005) 3D-QSAR studies on c-Src kinase inhibitors and docking analyses of a potent dual kinase inhibitor of c-Src and c-Abl kinases. Bioorg Med Chem 13:4704–4712

Todeschini R, Consonni V, Mauri A, Pavan M (2005) DRAGON, 5.3; Talete srl, Milan, Italy

Tong W, Perkins R, Xing L, Welsh WJ, Sheehan DM (1997) QSAR models for binding of estrogenic compounds to estrogen receptor α and β subtypes. Endocrinology 138:4022–4025

Ulrich S, Philippe S, Bischoff SF, Joachim N, Dominik F, Philipp F (2002) Toward selective ERbeta agonists for central nervous system disorders: synthesis and characterization of aryl benzthiophenes. J Med Chem 45:1399–1401

Wang Y, Li Y, Ding J, Jiang Z, Chang Y (2008) Estimation of bioconcentration factors using molecular electro-topological state and flexibility. SAR QSAR Environ Res 19:375–395

Wang Y, Li Y, Ding J, Wang Y, Chang YQ (2008) Prediction of binding affinity for estrogen receptor α modulators using statistical learning approaches. Mol Divers 12:93–102

Weihua Z, Andersson S, Cheng G, Simpson ER, Warner M, Gustafsson G (2003) Update on estrogen signaling FEBS Lett 546:17–24

Wolohan P, Reichert DE (2003) CoMFA and docking study of novel estrogen receptor subtype selective ligands. J Comput-Aided Mol Des 17:313–328

Wu YL, Yang XJ, Ren Z, McDonnell DP, Norris JD, Willson TM, Greene GL (2005) Structural basis for an unexpected mode of SERM-mediated ER antagonism. Mol Cell 18:413–424

Maximov PY, Lee TM, Craig Jordan V (2013) The discovery and development of selective estrogen receptor modulators (SERMs) for clinical practice. Curr Clin Pharmacol 8:135–155

Yang W, Wang Y, Ma Z, Golla R, Stouch T, Seethala R, Johnson S, Zhou R, Güngör T, Feyen JHM (2004) Synthesis and structure-activity relationship of 3-arylbenzoxazines as selective estrogen receptor beta agonists. Bioorg Med Chem Lett 14:2327–2330

Zhao L, Brinton RD (2005) Estrogen receptor β as a therapeutic target for promoting neurogenesis and preventing neurodegeneration. Drug Dev Res 66:103–117