Journal of Molecular Graphics and Modelling 28 (2010) 383–400

Contents lists available at ScienceDirect

Journal of Molecular Graphics and Modelling

journal homepage: www.elsevier.com/locate/JMGM

Pharmacophore and QSAR modeling of estrogen receptor β ligands and subsequent validation and *in silico* search for new hits

Mutasem O. Taha [a,*], Mai Tarairah [a], Hiba Zalloum [a], Ghassan Abu-Sheikha [b]

[a] Department of Pharmaceutical Sciences, Faculty of Pharmacy, University of Jordan, Queen Rania Street, Amman 11942, Jordan
[b] Department of Pharmaceutical Sciences, Faculty of Pharmacy, Al-Zaytoonah Private University of Jordan, Amman, Jordan

ARTICLE INFO

Article history:
Received 15 June 2009
Received in revised form 19 August 2009
Accepted 22 September 2009
Available online 30 September 2009

Keywords:
Estrogen receptor β
Pharmacophore modeling
QSAR
Docking
ROC
In silico screening

ABSTRACT

The pharmacophoric space of estrogen receptor beta (ERβ) was explored using a set of 119 known ligands. Subsequently, genetic algorithm and multiple linear regression analysis were employed to select optimal combinations of pharmacophoric models and physicochemical descriptors in self-consistent and predictive quantitative structure–activity relationships (QSARs) ($r^2_{96} = 0.79 - 0.83$, *F*-statistic = 40.96–36.20, $r^2_{LOO} = 0.74 - 0.76$ and r^2_{PRESS} against 23 external compounds = 0.54–0.56, respectively). Four binding hypotheses emerged in two optimal QSAR equations suggesting the existence of distinct binding modes accessible to ligands within ERβ binding pocket. The close similarity among the resulting pharmacophores prompted us to merge them in two hybrid models. The hybrid pharmacophores illustrated superior receiver operator characteristic curves (ROCs) and closely resembled binding interactions suggested by docking experiments. The resulting models and associated QSAR equations were employed to screen the national cancer institute (NCI) list of compounds and an *in house* built database of known drugs and agrochemicals to search for new ERβ ligands.

1. Introduction

Estrogen receptor (ER) is a ligand-activated transcription factor that belongs to the nuclear hormone receptor superfamily. The primary endogenous ligand for these receptors is 17β-estradiol. ER mediates the activity of estrogens in a variety of organs, including those in the reproductive, cardiovascular, skeletal, immune, and central nervous systems [1–3]. There are two subtypes of estrogen receptors encoded by different genes: ERα and ERβ [4]. Both ERs share distinct domains: DNA binding domains and hormone binding domains.

ERα and ERβ have modest overall sequence identity, differing greatly at the N-terminal domains, but having sequence conservation among DNA and ligand binding domains. The distribution of the two isoforms vary greatly and it is clear that they perform different biological functions [3,5,6,9]. Crystallographic structures of ligand complexes with ERα and ERβ show close binding pockets that differ only in two amino acids: Leu384 in ERα is replaced by Met336 in ERβ, and Met421 in ERα is replaced by Ile373 in ERβ [6–8].

ERβ agonists may be useful in a variety of clinical applications without triggering classic estrogenic side effects, e.g., uterine stimulation [3,9]. Potential applications of ERβ ligands include inflammatory bowel disease [10], rheumatoid arthritis [11], endometeriosis [12], and sepsis [13–15].

Despite extensive modeling efforts to understand ERα/ligand affinity and selectivity [87–92], the area of ERβ medicinal chemistry is generally devoid of significant modeling efforts except for some studies [1,2,8,16,17,91,92]. The presence of reasonable crystal structures for ERβ directed most modeling studies towards docking experiments [18–25]. However, although crystallographic data are considered the most reliable structural information that can be used for drug design, they are associated with some serious problems: (i) decisions whether to leave buried water molecules in the binding site or not have to be taken into consideration prior to ligand design and/or docking studies [26]. (ii) Inadequate resolution of crystallographic structures. The most reliable resolution for structure-based drug design has been recently recommended to be below 1.5 Å [27]. Incidentally, reported ERβ structures exhibited mediocre resolutions (1.8–3.0 Å) [1,8,11,28–35]. (iii) Crystallization-related artifacts, e.g., structural distortions inflected upon the three-dimensional structure of the ligand–protein complex during crystallization [36]. (vi) Crystallographic structures lack information on hydrogen atoms, so it should be appropriately assumed whether ionizable moieties exposed in the active site exist in their ionized form or not prior to *in silico* design [26,37]. Moreover, crystallographic structures generally ignore structural heterogeneity related to protein anisotropic motion and discrete conformational substrates [38].

* Corresponding author. Tel.: +962 777424750; fax: +962 65339649.
E-mail address: mutasem@ju.edu.jo (M.O. Taha).

1093-3263/$ – see front matter © 2009 Elsevier Inc. All rights reserved.
doi:10.1016/j.jmgm.2009.09.005

Table 1
The training subsets employed in exploring the pharmacophoric space of ERβ ligand, numbers correspond to compounds in n.

Training sets	Most-active[a]	Moderate active[a]	Least active[a]
A	**75**, **77**, **79**, **80**, **115**, **117**	**24**, **25**, **27**, **28**, **29**, **30**, **32**, **33**, **40**, **69**, **72**, **74**, **81**, **111**, **116**	**31**, **42**, **45**, **82**
B	**86**, **91**, **100**, **101**, **102**, **104**, **105**	**19**, **34**, **35**, **37**, **48**, **64**, **89**, **90**, **94**, **103**, **113**	**8**, **42**, **45**, **87**, **88**
C	**1**, **2**, **3**, **10**, **11**, **63**, **118**	**13**, **17**, **18**, **49**, **50**, **52**, **53**, **55**, **58**, **61**, **62**, **83**, **85**	**9**, **46**, **106**, **109**, **119**
D	**4**, **5**, **6**, **57**, **60**	**12**, **14**, **15**, **56**, **58**, **59**, **65**, **68**, **84**, **92**, **107**, **110**	**8**, **17**, **21**, **26**, **41**, **106**, **119**
E	**73**, **76**, **78**, **86**, **100**, **105**, **112**	**7**, **22**, **36**, **38**, **39**, **42**, **44**, **47**, **68**, **69**, **70**, **71**, **72**, **84**	**21**, **23**, **26**, **31**, **32**, **43**, **46**
F	**1**, **2**, **3**, **11**, **95**, **98**	**13**, **16**, **20**, **51**, **54**, **59**, **66**, **67**,**93**, **96**, **97**, **108**, **118**	**9**, **17**, **32**, **46**, **83**, **99**, **114**

[a] Potency categories as defined by Eqs. (1) and (2) in Section 2.1.4.

The continued interest in the development of new selective ERβ ligands combined with the lack of adequate ligand-based computer-aided modeling and drug discovery efforts in this area prompted us to explore the possibility of developing ligand-based three-dimensional (3D) pharmacophore(s) integrated within self-consistent QSAR model(s). The pharmacophore model(s) can be used as 3D search query(ies) to mine 3D libraries for new ERβ ligands, while the QSAR model helps to predict the biological activities of the captured compounds.

Pharmacophore modeling was performed employing CATALYST software. CATALYST enables automatic pharmacophore construction by using a collection of molecules with activities ranging over a number of orders of magnitude. CATALYST pharmacophores (hypotheses) explain the variability of activity of the molecules with respect to the geometric localization of the chemical features present in the molecules used to build it. The pharmacophore model consists of a collection of features necessary for the biological activity of the ligands arranged in 3D space, e.g., hydrogen-bond acceptors (HBA) and hydrogen-bond donors (HBD), hydrophobic regions (Hbic), aromatic rings (RingArom), positive ionizable (PosIon), etc. [56–62].

A total of 119 ERβ ligands were used in this study (Figure A and Table A in Supporting Information). These compounds were carefully collected from the literature such that they exhibit significantly dissimilar affinities to ERα and ERβ to allow access to selective models (pharmacophores and QSARs). Six training subsets were selected from this collection (Table 1). Each subset consisted of ligands of wide structural diversity. The affinities of the training subsets spanned over 4.0 orders of magnitude. Genetic algorithm and multiple linear regression statistical analysis were subsequently employed to select an optimal combination of complementary pharmacophores capable of explaining bioactivity variations among all ligands. Closely related QSAR-selected pharmacophores were merged in hybrid models that illustrated excellent ability to selectively capture active hits. The resulting models were employed as three-dimensional search queries to mine for new ERβ ligands.

2. Experimental

2.1. Molecular modeling

2.1.1. Software and hardware

The following software packages were utilized in the present research.

- CATALYST (Version 4.11), Accelrys Inc. (www.accelrys.com), USA.
- CERIUS2 (Version 4.10), Accelrys Inc. (www.accelrys.com), USA.
- CS ChemDraw Ultra 6.0, Cambridge Soft Corp. (http://www.cambridgesoft.Com), USA.

Pharmacophore and QSAR modeling studies were performed using CATALYST (HYPOGEN module) and CERIUS2 software suites from Accelrys Inc. (San Diego, CA, www.accelrys.com) installed on a Silicon Graphics Octane2 desktop workstation equipped with a dual 600 MHz MIPS R14000 processor (1.0 GB RAM) running the Irix 6.5 operating system. Structure drawing was performed employing ChemDraw Ultra 6.0 which was installed on a Pentium 4 PC.

2.1.2. Data set

The structures of 119 ERβ inhibitors and activators (**1**–**119**, Table A and Figure A in Supporting Information) were collected from recently published literature [2,3,6,8,9,11,16,17,39–55,86]. Although the collected ERβ ligands were gathered from different articles and were bioassayed employing two *in vitro* methodologies, it was possible to normalize their ERβ affinities based on the fact that compound **48** and **estradiol** were bioassayed by the two methods. The affinities were expressed as the concentrations of the test compounds that displaced 50% of the bound [^{3}H] 17β-estradiol (IC_{50}, nM). The logarithm of measured IC_{50} values were used in pharmacophore modeling and QSAR analysis thus correlating the data linear to the free energy change. The collected compounds were carefully selected such that they express significant affinity differences against ERα versus ERβ.

The two-dimensional (2D) chemical structures of the inhibitors were sketched using ChemDraw Ultra and saved in MDL-molfile format. Subsequently, they were imported into CATALYST, converted into corresponding standard 3D structures and energy minimized to the closest local minimum using the molecular mechanics CHARMm force field implemented in CATALYST. The resulting 3D structures were utilized as starting conformers for CATALYST conformational analysis.

2.1.3. Conformational analysis

The molecular flexibilities of the collected compounds were taken into account by considering each compound as a collection of conformers representing different areas of the conformational space accessible to the molecule within a given energy range. Accordingly, the conformational space of each ligand (**1**–**119**, Table A and Figure A in Supporting Information) was explored adopting the "best conformer generation" option within CATALYST, which is based on the generalized CHARMm force field implemented in the program. Default parameters were employed in the conformation generation procedure, i.e., for each molecule, a conformational ensemble was generated with an energy threshold of 20 kcal/mol from the local minimized structure that has the lowest energy level. A maximum of 250 conformers per molecule were generated [56].

2.1.4. Generation of pharmacophoric hypotheses

All 119 molecules with their associated conformational models were regrouped into a spreadsheet. The biological data of the inhibitors were reported with an "Uncertainty" value of 3, which means that the actual bioactivity of a particular inhibitor is assumed to be situated somewhere in an interval ranging from one-third to three-times the reported bioactivity value of that inhibitor [57–59]. Subsequently, six structurally diverse training subsets were carefully selected from the collection for pharmacophore modeling: **A**, **B**, **C**, **D**, **E**, and **F** in Table 1. Typically, CATALYST requires informative training sets that include at least

Table 2
Training sets and CATALYST run parameters employed for exploring ERβ pharmacophoric space.

Training subset[a]	Trial	Number of training compounds	Selected input features: types and ranges[b]	Number of output features	Spacing parameter[c]
A	**1**	25	HBA-F (0–3), HBD (0–3), Hbic (0–3), RingArom (0–3)	4–5	100
	2	25	HBA-F (0–3), HBD (0–3), Hbic (0–3), RingArom (0–3)	4–5	300
	3	25	HBA-F (0–3), HBD (0–3), Hbic (0–3), RingArom (0–3)	5–5	100
	4	25	HBA-F (0–3), HBD (0–3), Hbic (0–3), RingArom (0–3)	5–5	300
B	**1**	23	HBA-F (0–3), HBD (0–3), Hbic (0–3), RingArom (0–3)	4–5	100
	2	23	HBA-F (0–3), HBD (0–3), Hbic (0–3), RingArom (0–3)	4–5	300
	3	23	HBA-F (0–3), HBD (0–3), Hbic (0–3), RingArom (0–3)	5–5	100
	4	23	HBA-F (0–3), HBD (0–3), Hbic (0–3), RingArom (0–3)	5–5	300
C	**1**	25	HBA-F (0–3), HBD (0–3), Hbic (0–3), RingArom (0–3), PosIon (0–1)	4–5	100
	2	25	HBA-F (0–3), HBD (0–3), Hbic (0–3), RingArom (0–3), PosIon (0–1)	4–5	300
	3	25	HBA-F (0–3), HBD (0–3), Hbic (0–3), RingArom (0–3), PosIon (0–1)	5–5	100
	4	25	HBA-F (0–3), HBD (0–3), Hbic (0–3), RingArom (0–3), PosIon (0–1)	5–5	300
D	**1**	24	HBA (0–3), HBD (0–3), Hbic (0–3), RingArom (0–3), PosIon (0–1)	4–5	100
	2	24	HBA (0–3), HBD (0–3), Hbic (0–3), RingArom (0–3), PosIon (0–1)	4–5	300
	3	24	HBA (0–3), HBD (0–3), Hbic (0–3), RingArom (0–3), PosIon (0–1)	5–5	100
	4	24	HBA (0–3), HBD (0–3), Hbic (0–3), RingArom (0–3), PosIon (0–1)	5–5	300
E	**1**	28	HBA (0–3), HBD (0–3), Hbic (0–3), RingArom (0–3)	4–5	100
	2	28	HBA (0–3), HBD (0–3), Hbic (0–3), RingArom (0–3)	4–5	300
	3	28	HBA (0–3), HBD (0–3), Hbic (0–3), RingArom (0–3)	5–5	100
	4	28	HBA (0–3), HBD (0–3), Hbic (0–3), RingArom (0–3)	5–5	300
F	**1**	26	HBA (0–3), HBD (0–3), Hbic (0–3), RingArom (0–3), PosIon (0–1)	4–5	100
	2	26	HBA (0–3), HBD (0–3), Hbic (0–3), RingArom (0–3), PosIon (0–1)	4–5	300
	3	26	HBA (0–3), HBD (0–3), Hbic (0–3), RingArom (0–3), PosIon (0–1)	5–5	100
	4	26	HBA (0–3), HBD (0–3), Hbic (0–3), RingArom (0–3), PosIon (0–1)	5–5	300

[a] Correspond to training sets in Table 1.
[b] HBA: hydrogen-bond acceptor, HBA-F: hydrogen-bond acceptor including fluorine atoms, HBD: hydrogen-bond donor, RingArom: ring aromatic, Hbic: hydrophobic, PosIon: positive ionizable the allowed ranges of input features are in brackets.
[c] Other parameters were set to their default values.

16 compounds of evenly spread bioactivities over at least three and a h6alf logarithmic cycles. Lesser training lists could lead to chance correlation and thus faulty models [57–59].

The selected training sets were utilized to conduct 24 modeling runs to explore the pharmacophoric space of ERβ ligands (Table 2). The exploration process included altering interfeature spacing parameter (100 and 300 pm) and the maximum number of allowed features in the resulting pharmacophore hypotheses, as shown in Table 2.

Pharmacophore modeling employing CATALYST proceeds through three successive phases: the constructive phase, subtractive phase and optimization phase. During the constructive phase, CATALYST generates common conformational alignments among the most-active training compounds. Only molecular alignments based on a maximum of five chemical features are considered. The program identifies a particular compound as being within the most-active category if it satisfies Eq. (1) [57–59]:

$$(\mathrm{MAct} \times \mathrm{UncMAct}) - \frac{\mathrm{Act}}{\mathrm{UncAct}} > 0.0 \quad (1)$$

where "MAct" is the activity of the most-active compound in the training set, "Unc" is the uncertainty of the compounds and "Act" is the activity of the training compounds under question. However, if there are more than eight most-active inhibitors, only the top eight are used.

In the subsequent subtractive phase, CATALYST eliminates some hypotheses that fit inactive training compounds. A particular training compound is defined as being inactive if it satisfies Eq. (2) [57–59]:

$$\log(\mathrm{Act}) - \log(\mathrm{MAct}) > 3.5 \quad (2)$$

However, in the optimization phase, CATALYST applies fine perturbations in the form of vectored feature rotation, adding new feature and/or removing a feature, to selected hypotheses that survived the subtractive phase, in an attempt to find new models of enhanced bioactivity/mapping correlation, i.e., improved 3D-QSAR properties. Eventually, CATALYST selects the highest ranking models (10 by default) and presents them as the optimal pharmacophore hypotheses resulting from the particular automatic modeling run.

Our pharmacophore exploration efforts (24 automatic runs, Tables 1 and 2) culminated in 210 pharmacophore models of variable qualities.

2.1.5. Assessment of the generated hypotheses

When generating hypotheses, CATALYST attempts to minimize a cost function consisting of three terms: weight cost, error cost and configuration cost [56–61]. Weight cost is a value that increases as the feature weight in a model deviates from an ideal value of 2, which originates from the assumption that each pharmacophoric feature contributes 2 log cycles to affinity. The deviation between the estimated activities of the training set and their experimentally determined values adds to the error cost. Error cost provides the highest contribution to total cost and it is directly related to the capacity of the particular pharmacophore as 3D-QSAR model, i.e., in correlating the molecular structures to the corresponding biological responses. The activity of any compound can be estimated from a particular hypothesis through Eq. (3) [56].

$$\log(\text{estimated activity}) = I + \mathrm{Fit} \quad (3)$$

where I = the intercept of the regression line obtained by plotting the log of the biological activity of the training set compounds against the Fit values of the training compounds. The Fit value for any compound is obtained automatically employing Eq. (4) [56].

$$\mathrm{Fit} = \sum \text{mapped hypothesis features} \times W\left[1 - \sum\left(\frac{\mathrm{disp}}{\mathrm{tol}}\right)^2\right] \quad (4)$$

where Σmapped hypothesis features represents the number of pharmacophore features that successfully superimpose (i.e., map or overlap with) corresponding chemical moieties within the fitted compound, *W* is the weight of the corresponding hypothesis feature spheres. This value is fixed to 1.0 in CATALYST-generated models. disp is the distance between the center of a particular pharmacophoric sphere (feature centroid) and the center of the corresponding superimposed chemical moiety of the fitted compound; tol is the radius of the pharmacophoric feature sphere (known as tolerance, equals to 1.6 Å by default). $\Sigma(\text{disp/tol})^2$ is the summation of $(\text{disp/tol})^2$ values for all pharmacophoric features that successfully superimpose corresponding chemical functionalities in the fitted compound [56].

The third cost term, i.e., the configuration cost, penalizes the complexity of the hypothesis. This is a fixed cost, which is equal to the entropy of the hypothesis space. The more the numbers of features (a maximum of five) in a generated hypothesis, the higher is the entropy with subsequent increase in this cost.

The overall cost (total cost) of a hypothesis is calculated by summing over the three cost factors. However, error cost is the main contributor to total cost.

CATALYST calculates a fixed cost, which estimates the cost of the ideal hypothesis. CATALYST also calculates the cost of the null hypothesis, which presumes that there is no relationship in the data and that experimental activities are normally distributed about their mean. Accordingly, the greater the difference from the null hypothesis cost and the closer the difference from the fixed cost, the more likely that the hypothesis does not reflect a chance correlation [56,61,62]. In a successful automatic modeling run, CATALYST® ranks the generated models according to their total costs [56,60,62].

An additional approach to assess the quality of CATALYST-HYPOGEN pharmacophores is to cross-validate them using the Cat-Scramble algorithm implemented in CATALYST. This validation procedure is based on Fisher's randomization test [63]. We selected a 95% confidence level in this validation test, which instruct CATALYST to generate 19 random spreadsheets by the Cat-Scramble command. Subsequently, CATALYST-HYPOGEN is challenged to use these random spreadsheets to generate hypotheses using exactly the same features and parameters used in generating the initial unscrambled hypotheses [64]. Success in generating pharmacophores of comparable cost criteria to those produced by the original unscrambled data reduces the confidence in the training compounds and the unscrambled original pharmacophore models. Based on Fisher randomization criteria, all resulting 210 pharmacophores exceeded the 90% significance threshold for subsequent processing (clustering and QSAR analyses). Table 3 shows the success criteria of generated pharmacophores.

2.1.6. Clustering the generated pharmacophore hypotheses

The successful models (210) were clustered into 42 groups utilizing the hierarchical average linkage method available in CATALYST. Clustering was implemented in such a way that models generated from a particular training set were clustered into 5-membered subgroups. Subsequently, the highest ranking members, as judged based on their significance *F*-values, were selected to represent their clusters in subsequent QSAR modeling. Table 4 shows information about representative pharmacophores including their pharmacophoric features, success criteria and differences from corresponding null hypotheses. The table also shows the corresponding Cat.Scramble confidence levels for each representative pharmacophore.

Table 3
The general performance of pharmacophoric models generated for ERβ[a].

Training subset[b]	Trial[c]	Cost criteria[d]			Fisher confidence range (%)[e]
		Configuration	Total cost range	Null	
A	**1**	13.73	111.65–118.84	192.74	95.0
	2	9.30	123.06–139.68	192.74	95.0
	3[f]	–	–	–	–
	4[f]	–	–	–	–
B	**1**	13.92	100.35–111.57	149.58	95.0
	2	12.26	99.79–115.76	149.58	95.0
	3	11.73	105.08–114.80	149.58	95.0
	4	8.50	107.42–129.54	149.58	90.0 – 95.0
C	**1**	19.56	121.31–128.56	181.08	95.0
	2	18.60	118.26–134.47	181.08	95.0
	3	18.28	116.37–128.37	181.08	95.0
	4	16.20	127.96–133.01	181.08	95.0
D	**1**	18.40	117.92–121.44	175.17	95.0
	2	16.34	114.40–123.14	175.17	95.0
	3	17.22	115.66–121.24	175.17	95.0
	4	14.23	123.21–128.92	175.17	95.0
E	**1**	11.44	123.68–137.11	213.45	95.0
	2	7.82	124.18–159.22	213.45	90.0–95.0
	3	7.51	125.71–157.70	213.45	90.0–95.0
	4[f]	–	–	–	–
F	**1**	19.56	108.14–118.64	172.13	95.0
	2	18.60	113.19–117.49	172.13	95.0
	3	18.28	114.69–119.52	172.13	95.0
	4	16.20	114.80–118.57	172.13	95.0

[a] The detailed success criteria for selected representative pharmacophore models are shown in Table 4.
[b] The training subsets are as in Table 1.
[c] Selected input features and run parameters are as in Table 2.
[d] See Section 2.1.5 for detailed description and discussion regarding the different cost criteria.
[e] Calculated employing the Cat.Scramble method (see Section 2.1.5).
[f] These runs failed to yield any pharmacophore models.

Table 4
The performance of best representatives of clustered pharmacophore hypotheses generated for ERβ.

Training set[a]	Hypotheses[b,c]	Pharmacophoric features in generated hypotheses	Total cost	Cost of null hypothesis	Residual cost[d]	R[e]	F-statistic[f]	Cat.Scramble (%)
A	**HypoA/1/2**[g,h]	**2×HBD, Hbic, RingArom**	**112.00**	**192.74**	**80.74**	**0.94**	**29.48**	**95**
	HypoA/1/6	HBA-F, HBD, Hbic, RingArom	117.64	192.74	75.1	0.90	28.99	95
	HypoA/2/4	3×HBA-F, HBic	127.40	192.74	65.34	0.85	13.53	95
	HypoA/2/5	HBA-F, 2×HBD, Hbic	133.22	192.74	59.52	0.82	9.36	95
B	HypoB/1/4	HBA-F, HBD, 2×Hbic	106.48	149.59	43.11	0.90	26.35	95
	HypoB/1/6	HBD, 2×Hbic, RingArom	108.46	149.59	41.13	0.88	11.88	95
	HypoB/1/10	HBA-F, HBD, Hbic, RingArom	111.57	149.59	38.02	0.86	34.89	95
	HypoB/2/6	2×HBA-F, Hbic, RingArom	107.22	149.59	42.37	0.88	21.72	95
	HypoB/2/7	HBA-F, 2×Hbic, RingArom	111.41	149.59	38.18	0.85	19.36	95
	HypoB/3/1	HBA-F, HBD, 3×Hbic	105.08	149.59	44.51	0.90	30.71	95
	HypoB/3/10	HBD, 3×Hbic, RingArom	114.79	149.59	34.8	0.81	22.81	95
	HypoB/4/6	2×HBA-F, 3×Hbic	121.80	149.59	27.79	0.76	17.52	95
C	HypoC/1/4	HBD, 2×Hbic, PosIon, RingArom	125.18	181.08	55.9	0.89	4.46	95
	HypoC1/5	2×HBA-F, Hbic, RingArom	126.72	181.08	54.36	0.88	20.99	95
	HypoC1/8	HBA-F, HBD, Hbic, RingArom	127.63	181.08	53.45	0.88	22.45	95
	HypoC1/9	2×HBD, 3×Hbic	128.07	181.08	53.01	0.88	16.06	95
	HypoC/2/7	2×HBD, Hbic, RingArom	133.44	181.08	47.64	0.83	12.05	95
	HypoC/3/4	HBA-F, HBD, 3×Hbic	125.98	181.08	55.1	0.88	24.09	95
	HypoC/4/1	HBD, 2×Hbic, PosIon, RingArom	127.96	181.08	53.12	0.85	7.16	95
	HypoC/4/4	HBA-F, 2×Hbic, PosIon, RingArom	131.22	181.08	49.86	0.84	6.57	95
D	HypoD/1/4	HBA, HBD, 3×Hbic	118.72	175.17	56.45	0.90	19.03	95
	HypoD/2/1	HBA, Hbic, 2×RingArom	114.40	175.17	60.77	0.91	9.14	95
	HypoD/2/5	HBD, Hbic, 2×RingArom	121.81	175.17	53.36	0.87	16.09	95
	HypoD/2/9	HBA, HBD, Hbic, RingArom	123.11	175.17	52.06	0.86	10.27	95
	HypoD/3/8[h]	**HBD, 3×Hbic, RingArom**	**120.89**	**175.17**	**54.28**	**0.88**	**45.75**	**95**
	HypoD/4/2	3×Hbic, 2×RingArom	123.67	175.17	51.5	0.85	8.56	95
	HypoD/4/4[h]	**2×HBA, 2×Hbic, posIon**	**125.97**	**175.17**	**49.2**	**0.84**	**6.53**	**95**
	HypoD/4/10[h]	**HBA, 2×Hbic, PosIon, RingArom**	**128.92**	**175.17**	**46.25**	**0.83**	**6.73**	**95**
E	HypoE/1/5	HBD, 2×Hbic, RingArom	130.53	213.45	82.92	0.90	37.48	95
	HypoE/1/7	HBA, 2×Hbic, RingArom	131.96	213.45	81.49	0.89	35.21	95
	HypoE/1/8	3×Hbic, RingArom	132.49	213.45	80.96	0.89	15.73	95
	HypoE/2/3	HBA, 3×Hbic	131.83	213.45	81.62	0.88	19.05	95
	HypoE/3/2	HBA, HBD, 3×Hbic	141.71	213.45	71.74	0.831	0.06	95
	HypoE/3/7	2×HBA, 3×Hbic	153.33	213.45	60.12	0.763	27.88	90
F	HypoF/1/5	HBA, HBD, Hbic, RingArom	118.08	172.13	54.05	0.92	4.03	95
	HypoF/1/7	HBD, Hbic, posIon, RingArom	118.21	172.13	53.92	0.94	3.34	95
	HypoF/1/10	HBA, Hbic, 2×RingArom	118.64	172.13	53.49	0.94	6.23	95
	HypoF/2/1	2×HBA, Hbic, PosIon	113.19	172.13	58.94	0.96	6.67	95
	HypoF/2/9	HBA, Hbic, PosIon, RingArom	117.45	172.13	54.68	0.94	9.96	95
	HypoF/3/4	HBD, 3×Hbic, PosIon	118.45	172.13	53.68	0.93	4.96	95
	HypoF/3/10	HBA, HBD, 3×Hbic	119.52	172.13	52.61	0.93	9.81	95
	HypoF/4/8	HBD, 2×Hbic, PosIon, RingArom	118.32	172.13	53.81	0.93	7.61	95

[a] Correspond to training sets in Table 1.
[b] Correspond to runs in Table 2.
[c] Best models from their respective clusters, as judged based on F-statistic.
[d] The difference between the total cost and the cost of the corresponding null hypotheses.
[e] The correlation coefficients between bioactivity estimates and bioactivities of corresponding training set compounds.
[f] Fisher statistic calculated based on the linear regression between the fit values of collected ligands (**1**–**119**, Table A and figure A in Supporting Information) against pharmacophore hypothesis (employing the "best fit" option and Eq. (4)) and their respective ERβ affinities.
[g] HypoA/1/2 corresponds to the second binding hypothesis generated in the first automatic run trial for set A (see Table 2).
[h] Bolded pharmacophores emerged in the best QSAR equations (bolded).

2.1.7. QSAR modeling

A subset of 96 compounds taken from the total list of ligands (**1**–**119**, Table A and Figure A in Supporting Information) was utilized as a training set for QSAR modeling. However, since it is essential to access the predictive power of the resulting QSAR models on external set of inhibitors, the remaining 23 molecules (*ca.* 20% of the dataset) were employed as test subset for validating the QSAR models. The test molecules were selected as follows: the 119 compounds were ranked according to their IC_{50} values, and then every fifth compound was selected for the test set starting from the high-potency end. This selection considers the fact that the test molecules must represent a range of biological activities similar to that of the training set. The selected test ligands are: **5**, **11**, **13**, **14**, **18**, **28**, **31**, **32**, **35**, **36**, **40**, **44**, **48**, **59**, **60**, **63**, **71**, **76**, **85**, **90**, **100**, **107**, **114** (Table A and Figure A in Supporting Information).

The logarithm of measured $1/IC_{50}$ (nM) values were used in QSAR, thus correlating the data linear to the free energy change. The chemical structures of the inhibitors were imported into CERIUS2 as standard 3D single conformer representations in SD format. Subsequently, different descriptor groups were calculated for each compound employing the C2.DESCRIPTOR module of CERIUS2. The calculated descriptors included various simple and valence connectivity indices, electrotopological state indices and other molecular descriptors (e.g., logarithm of partition coefficient, polarizability, dipole moment, molecular volume, molecular weight, molecular surface area, etc.) [65]. Furthermore, the training compounds were fitted (using the Best fit option in

Table 5
Statistical criteria of high-ranking QSAR models.

Model	Terms[a]	$r^2_{(96)}$ [b]	*F*-value	$r^2_{(LOO)}$ [c]	$r^2_{(BS)}$ [d]	$r^2_{(PRESS)}$ [e]	PRESS[f]	Pharmacophoric terms in QSAR model[g]
A	5	0.634	39.502	0.591	0.635	0.553	14.690	HypoA/1/2, HypoD/3/8
B	6	0.723	47.139	0.678	0.724	0.509	16.141	HypoA/1/2, HypoD/3/8
C[g]	9	0.790	40.963	0.735	0.790	0.540	15.139	HypoA/1/2, HypoD/3/8, HypoD/4/4
D	5	0.644	41.217	0.600	0.645	0.605	12.987	HypoA/1/2, HypoD/3/8
E	6	0.702	42.450	0.657	0.702	0.625	12.350	HypoA/1/2, HypoD/3/8
F[g]	12	0.826	36.199	0757	0.818	0.561	14.434	HypoA/1/2, HypoD/3/8, HypoD/4/10

[a] Number of explanatory terms including the intercept.
[b] Non-cross-validated correlation coefficient for 96 training compounds.
[c] Cross-validation correlation coefficients determined by the leave-one-out technique.
[d] Bootstrapping correlation coefficient.
[e] Predictive r^2 determined for the 23 test compounds.
[f] The sum of squared deviations between predicted and actual activity values for every molecule in the test set of 23 compounds.
[g] These QSAR equations were selected to predict the ERβ inhibitory activities of captured hits.

CATALYST) against the representative pharmacophores (42 models, Table 4), and their fit values were added as additional descriptors. The fit value for any compound is obtained automatically via Eq. (4) [56].

Genetic function approximation (GFA) was employed to search for the best possible QSAR regression equation capable of correlating the variations in biological activities of the training compounds with variations in the generated descriptors, i.e., multiple linear regression modeling (MLR). We manually removed sparse variables prior to QSAR analysis to avoid overwhelming GFA-MLR with large number of poor descriptors.

Our preliminary diagnostic trials suggested the following optimal GFA parameters: explore linear, quadratic and spline equations at mating and mutation probabilities of 50%; population size = 500; number of genetic iterations = 30,000 and lack-of-fit (LOF) smoothness parameter = 1.0 [65]. However, to determine the optimal number of explanatory terms (QSAR descriptors), it was decided to scan and evaluate all possible QSAR models resulting from 3 to 17 explanatory terms.

All QSAR models were validated employing leave-one-out cross-validation (r^2_{LOO}), bootstrapping (r^2_{BS}) and predictive r^2 (r^2_{PRESS}) calculated from the test subsets. The predictive r^2_{PRESS} is defined as:

$$r^2_{PRESS} = SD - \frac{PRESS}{SD} \quad (5)$$

where SD is the sum of the squared deviations between the biological activities of the test set and the mean activity of the training set molecules, PRESS is the squared deviations between predicted and actual activity values for every molecule in the test set.

Descriptor-scanning identified several high-quality QSAR models, of which models **C** and **F** (see Table 5 and Eqs. (10) and (11)) exhibited the best overall statistical criteria, and therefore, were selected to predict the inhibitory actions of our *in silico* hits. Fig. 1 shows the plots of experimental versus fitted (training set) and predicted (testing set) compound affinities to ERβ as calculated by QSAR models **C** and **F**. Four pharmacophore hypotheses emerged in these optimal QSAR models, namely, HypoA/1/2, HypoD/3/8, HypoD/4/4 and HypoD/4/10. Table 6 shows the three-dimensional coordinates of the four pharmacophores, while Figs. 2 and 3 show the pharmacophoric features of the four models, and how they map training compounds **117** (IC_{50} = 0.136 nM) and **5** (IC_{50} = 0.6 nM).

2.1.8. Receiver operator characteristic curve analysis (ROC analysis)

Successful pharmacophore models (i.e., HypoA/1/2, HypoD/3/8, HypoD/4/4, HypoD/4/10 and Merged models, see below) were further validated by assessing their abilities to selectively capture diverse ERβ active compounds from a large list of decoys employing ROC analysis.

Therefore, it was necessary to prepare valid evaluation structural database (testing set) that contain an appropriate list of decoy compounds in combination with diverse list of known active compounds. The decoy list was prepared as described by Verdonk et al. [66,67]. Briefly, the decoy compounds were selected based on three basic one-dimensional (1D) properties that allow the assessment of distance (*D*) between two molecules (e.g., *i* and *j*): (i) the number of hydrogen-bond donors (NumHBD); (ii) number of hydrogen-bond acceptors (NumHBA), and (iii) count of nonpolar atoms (NP, defined as the summation of Cl, F, Br, I, S and C

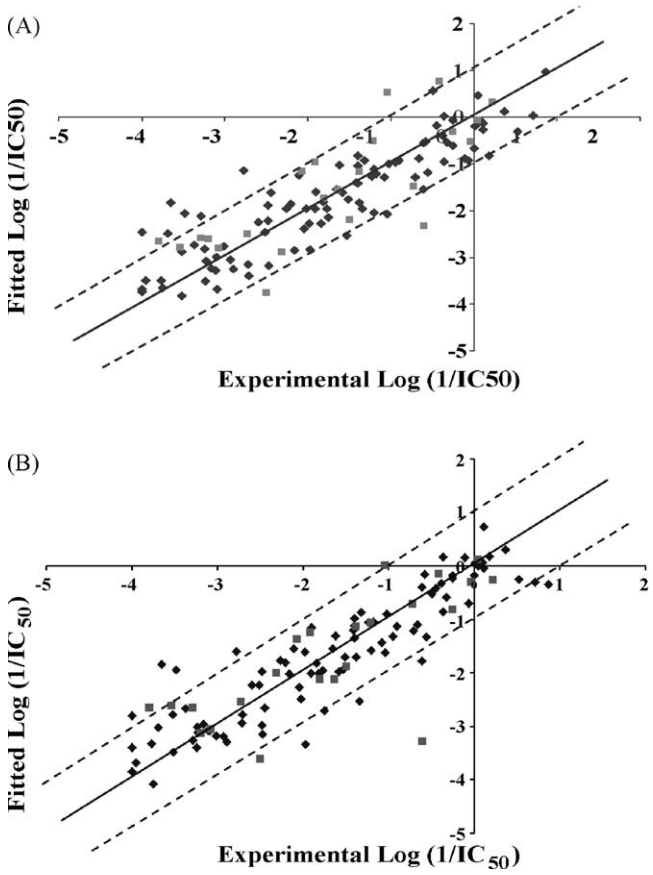


Fig. 1. (A) Experimental versus fitted (◆, 96 compounds, $r^2_{LOO} = 0.735$) and predicted (■, 23 compounds, $r^2_{PRESS} = 0.540$) bioactivities calculated from QSAR model **C** (in Table 5 and Eq. (11)) for ERβ ligands. (B) Experimental versus fitted (◆, 96 compounds, $r^2_{LOO} = 0.757$) and predicted (■, 23 compounds, $r^2_{PRESS} = 0.561$) bioactivities calculated from QSAR model **F** (in Table 5 and Eq. (11)) for ERβ ligands. The solid lines are the regression lines for the fitted and predicted bioactivities of training and test compounds, respectively, whereas the dotted lines indicate the 1 log point error margins.

Table 6
Pharmacophoric features and corresponding weights, tolerances and 3D coordinates of HypoA/1/2[a], HypoD/3/8[b], HypoD/4/4[c] and HypoD/4/10[d].

Model	Definitions	Chemical features						
		HBD		HBD		Hbic	RingArom	
HypoA/1/2[a]	Weights	2.17686		2.17686		2.17686	2.17686	
	Tolerances	1.60	2.20	1.60	2.20	1.60	1.60	1.60
	Coordinates X	2.58	4.75	−2.20	−2.16	1.42	−1.14	0.81
	Y	5.22	6.81	−5.22	−7.06	2.96	−2.58	−4.14
	Z	−1.14	0.19	−0.06	−2.43	−0.66	−0.09	1.57
Model	Definitions	Chemical features						
		HBD		Hbic	Hbic	Hbic	RingArom	
HypoD/3/8[b]	Weights	1.51398		1.51398	1.51398	1.51398	1.51398	
	Tolerances	1.60	2.20	1.60	1.60	1.60	1.60	1.60
	Coordinates X	−3.66	−6.63	−2.45	1.20	0.64	−2.62	−2.28
	Y	3.91	3.98	4.13	1.97	5.65	3.85	0.77
	Z	5.38	5.81	0.65	−0.44	−3.02	3.22	2.48
Model	Definitions	Chemical features						
		HBA		HBA		Hbic	Hbic	PosIoniz
HypoD/4/4[c]	Weights	1.36973		1.36973		1.36973	1.36973	1.36973
	Tolerances	1.60	2.20	1.60	2.20	1.60	1.60	1.60
	Coordinates X	2.58	3.50	−3.88	−2.49	−3.56	−3.96	4.86
	Y	0.04	0.81	3.11	5.42	−2.07	−0.40	−0.42
	Z	−1.34	−4.10	4.53	5.86	−3.36	0.68	1.57
Model	Definitions	Chemical features						
		HBA		Hbic	Hbic	PosIons	RingArom	
HypoD/4/10[d]	Weights	1.211		1.211	1.211	1.211	1.211	
	Tolerances	1.60	2.20	1.60	1.60	1.60	1.60	1.60
	Coordinates X	−2.48	−4.47	3.44	2.22	−3.13	1.86	0.41
	Y	0.73	2.91	0.86	5.16	−2.82	−0.78	−2.20
	Z	1.88	2.57	−1.92	0.32	2.91	−3.86	−1.65

[a] HypoA/1/2 is hypothesis number 2 generated in run number 1 based on set A.
[b] HypoD/3/8 is hypothesis number 8 generated in run number 3 based on set D.
[c] HypoD/4/4 is hypothesis number 4 generated in run number 4 based on set D.
[d] HypoD/4/10 is hypothesis number 10 generated in run number 4 based on set D.

atoms in a particular molecule). For each active compound in the testing set, the distance to the nearest other active compound is assessed using their Euclidean Distance (Eq. (6)):

$$D(i,j) = \sqrt{(\mathrm{NumHBD}_i - \mathrm{NumHBD}_j)^2 + (\mathrm{NumHBA}_i - \mathrm{NumHBA}_j)^2 + (\mathrm{NP}_i - \mathrm{NP}_j)^2} \quad (6)$$

The minimum distances are then averaged over all active compounds (Dmin). Subsequently, for each active compound in the testing set, 36 decoys were randomly chosen from the ZINC database [68]. The decoys were selected in such a way that they did not exceed Dmin distance from their corresponding active compound.

Moreover, to further diversify the actives members, we excluded any active compound having zero distance ($D(i,j)$) from other active compound(s) in the testing set. Active testing compounds were defined as those possessing ERβ affinities ranging from 0.136 to 1000 nM. The testing set included 39 active compounds and 1383 ZINC compounds.

The testing set (1422 compounds) was screened by each particular pharmacophore employing the "Best flexible search" option implemented in CATALYST, while the conformational spaces of the compounds were generated employing the "Fast conformation generation option" implemented in CATALYST. Compounds missing one or more features were discarded from the hit list. The *in silico* hits were scored employing their fit values as calculated by Eq. (4).

ROC curve analysis describes the sensitivity (*Se* or true positive rate, Eq. (7)) for any possible change in the number of selected compounds (**n**) as a function of $(1 - Sp)$. *Sp* is defined as specificity or true negative rate (Eq. (8)) [67–69].

$$Se = \frac{\text{number of selected actives}}{\text{total number of actives}} = \frac{\mathrm{TP}}{\mathrm{TP} + \mathrm{FN}} \quad (7)$$

$$Sp = \frac{\text{number of discarded inactives}}{\text{total number of inactives}} = \frac{\mathrm{TN}}{\mathrm{TN} + \mathrm{FP}} \quad (8)$$

where TP is the number of active compounds captured by the virtual screening method (true positives), FN is the number of active compounds discarded by the virtual screening method, TN is the number of discarded decoys (presumably inactives), while FP is the number of captured decoys.

If all molecules scored by a virtual screening (VS) protocol with sufficient discriminatory power are ranked according to their score (i.e., fit values), starting with the best-scored molecule and ending with the molecule that got the lowest score, most of the actives will have a higher score than the decoys. Since some of the actives will be scored lower than decoys, an overlap between the distribution of active molecules and decoys will occur, which will lead to the prediction of false positives and false negatives [67,69]. The selection of one score value as a threshold strongly influences the ratio of actives to decoys and therefore the validation of a VS method. However, ROC curve method avoids the selection of a threshold by considering all *Se* and *Sp* pairs for each score threshold [69]. A ROC curve is plotted by setting the score of the active molecule as the first threshold. Afterwards, the number of decoys within this cutoff is counted and the corresponding *Se* and *Sp* pair is calculated. This calculation is repeated for the active molecule with the second highest score and so forth, until the scores of all actives are considered as selection thresholds.

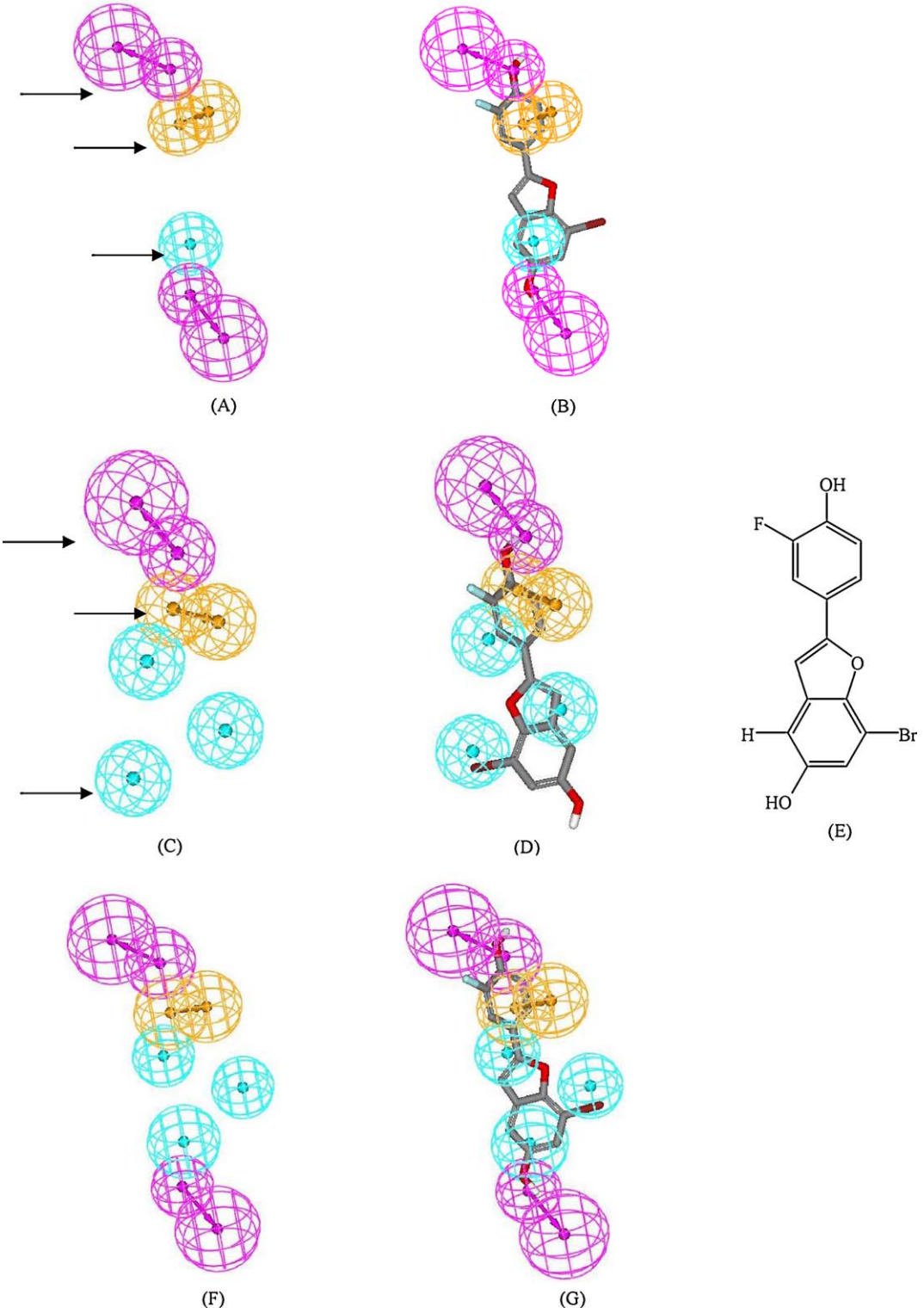


Fig. 2. Features of HypoA/1/2, HypoD/3/8 and HypoA/1/2-D/3/8: HBD as violet vectored spheres, Hbic as blue spheres, and RingArom as orange vectored spheres: (A) HypoA/1/2, (B) HypoA/1/2 fitted against **117** (IC_{50} = 0.136 nM), (C) HypoD/3/8 (D) HypoD/3/8 fitted against **117**, (E) structure of **117**, (F) merged HypoA/1/2-D/3/8, and (G) Hybrid HypoA/1/2-D/3/8 fitted against **117**. Arrows point to closely positioned common features in HypoA/1/2 and HypoD/3/8 allowing pharmacophore merging.

An ideal ROC curve continues as a horizontal straight line to the upper-right corner where all actives and all decoys are retrieved, which corresponds to *Se* = 1 and *Sp* = 0. In contrast to that, the ROC curve for a set of actives and decoys with randomly distributed scores tends towards the $Se = 1 - Sp$ line asymptotically with increasing number of actives and decoys [67]. The success of a particular virtual screening workflow can be judged from the following criteria (shown in Table 7):

(1) *Area under the ROC curve* (*AUC*) [69]. In an optimal ROC curve an AUC value of 1 is obtained; however, random distributions cause an AUC value of 0.5. Virtual screening that performs

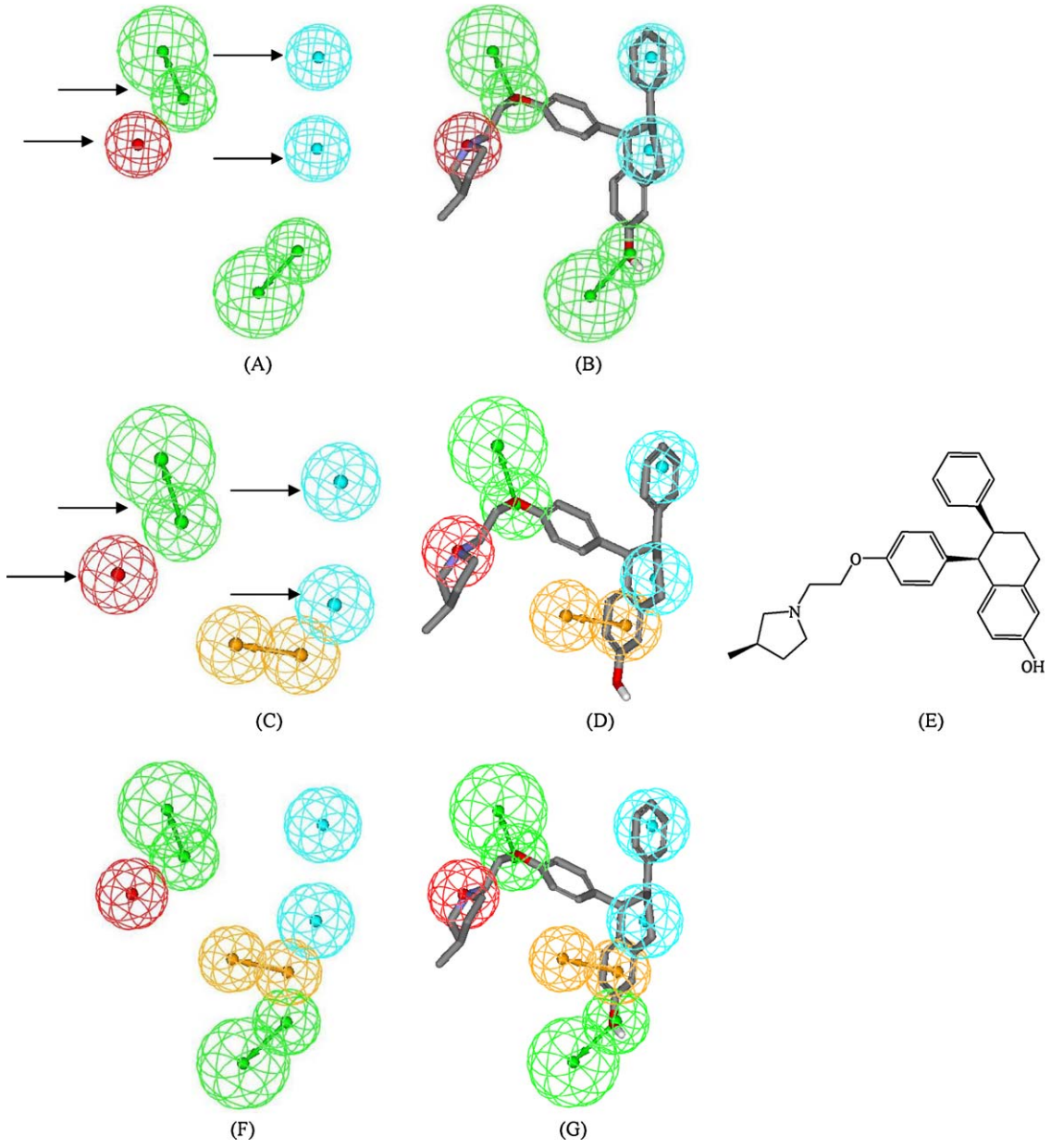


Fig. 3. Pharmacophoric features of HypoD/4/4, HypoD/4/10 and HypoD/4/4-D/4/10: HBA as green vectored spheres, Hbic as blue spheres, PosIons as a red sphere, and RingArom as orange vectored spheres, (A) HypoD/4/4 (B) HypoD/4/4 fitted against **5** (IC_{50} = 0.6 nM), (C) HypoD/4/10 (D) HypoD/4/10 fitted against **5**, (E) structure of **5**, (F) Hybrid HypoD/4/4-D/4/10, and (G) Merged HypoD/4/4-D/4/10 fitted against **5**. Arrows point to closely positioned common features in HypoD/4/4 and HypoD/4/10 allowing pharmacophore merging.

better than a random discrimination of actives and decoys retrieve an AUC value between 0.5 and 1, whereas an AUC value lower than 0.5 represents the unfavorable case of a virtual screening method that has a higher probability to assign the best scores to decoys than to actives [67,69].

(2) *Overall accuracy* (*ACC*) describes the percentage of correctly classified molecules by the screening protocol (Eq. (9)). Testing compounds are assigned a binary score value of zero (compound not captured) or one (compound captured) [67,70,71].

$$\mathrm{ACC} = \frac{\mathrm{TP} + \mathrm{TN}}{N} = \frac{A}{N} Se + \left(1 - \frac{A}{N}\right) Sp \tag{9}$$

where N is the total number of compounds in the testing database, A is the number of true actives in the testing database.

(3) *Overall specificity* (*SPC*) describes the percentage of discarded inactives by the particular virtual screening workflow. Inactive test compounds are assigned a binary score value of zero (compound not captured) or one (compound captured) regardless to their individual fit values [67,70,71].

(4) *Overall true positive rate* (*TPR or overall sensitivity*) describes the fraction percentage of captured actives from the total number of actives. Active test compounds are assigned a binary score value of zero (compound not captured) or one (compound captured) regardless to their individual fit values.

(5) *Overall false negative rate* (*FNR or overall percentage of discarded actives*) describes the fraction percentage of active compounds discarded by the virtual screening method. Discarded active test compounds are assigned a binary score value of zero (compound not captured) or one (compound captured) regardless to their individual fit values.

Table 7
The success criteria of different QSAR-selected pharmacophore models and their merged offspring.

Pharmacophore model	ROC criteria[a]				
	AUC[b]	ACC[c]	SPC[d]	TPR[e]	FNR[f]
HypoA/1/2	0.953	0.9599	0.9857	0.342	0.0142
HypoD/3/8	0.862	0.9599	0.9681	0.763	0.0318
HypoD/4/4	0.981	0.9599	0.989	0.263	0.011
HypoD/4/10	0.958	0.9599	0.990	0.237	0.01
HypoA/1/2-D/3/8[g]	0.964	0.9599	0.988	0.289	0.012
HypoD/4/4-D/4/10[h]	0.987	0.9599	0.989	0.263	0.011

[a] ROC: receiver operator characteristic curve.
[b] AUC: area under the curve.
[c] ACC: overall accuracy.
[d] SPC: overall specificity.
[e] TPR: overall true positive rate.
[f] FNR: overall false negative rate.
[g] Hybrid hypothesis generated from HypoA/1/2 and HypoD/3/8.
[h] Hybrid hypothesis generated from HypoD/4/4 and HypoD/4/10.

Table 7 and Figs. 4 and 5 show the success criteria of the four QSAR-selected models and hybrid pharmacophores (see section below).

2.1.9. Merging QSAR-selected pharmacophores

The fact that HypoA/1/2 and HypoD/3/8 share closely positioned HBD, RingArom and Hbic features (see Fig. 2), while HypoD/4/4 and HypoD4/10 share closely positioned PosIons, HBA, and two Hbic features (Fig. 3), prompted us to merge the two pairs into two hybrid models: HypoA/1/2-D/3/8 and HypoD/4/4-D/4/10. Common features were merged in such a way that the tolerance constrains of the parent models are averaged in the offspring hybrid pharmacophores.

HypoA/1/2 and HypoD/3/8 were merged at a distance tolerance threshold of 0.7 Å, i.e., any two closely positioned features, from the originating models, were merged in a single feature in the offspring model provided their centers occur within a maximum distance of 0.7 Å. Similarly, HypoD/4/4 and HypoD4/10 were merged at a distance tolerance threshold of 0.5 Å.

2.1.10. Docking experiments

The three-dimensional structures of **5** and **117** (Table A and figure A in Supporting Information) were sketched in Chemdraw Ultra and imported in CERIUS2 in SD file format. The ligands were assumed to exist in their unionized states and were assigned partial atomic charges using the default Gasteiger method implemented in CERIUS2 [72,73].

The 3D coordinates of ERβ were retrieved from the Protein Data Bank (PDB code: 2FSZ, resolution 2.20 Å). Hydrogen atoms were added to the protein utilizing CERIUS2 templates for protein residues. Gasteiger charges were assigned to the protein atoms as implemented in LigandFit® [73,74]. Explicit water molecules were removed from the protein and the protein structure was used in docking without energy minimization. LigandFit® considers the flexibility of the ligand and treats the receptor as rigid. The two structures were docked employing the following configurations: (i) Monte Carlo search parameters: number of trials = 15,000; search step for torsions with polar hydrogens = 30.0 degree. (ii) The RMS threshold for ligand-to-binding site shape match was set to 2.0 employing a maximum of 2.0 binding site partitions. (iii) Interaction energy parameters: the interaction energies were

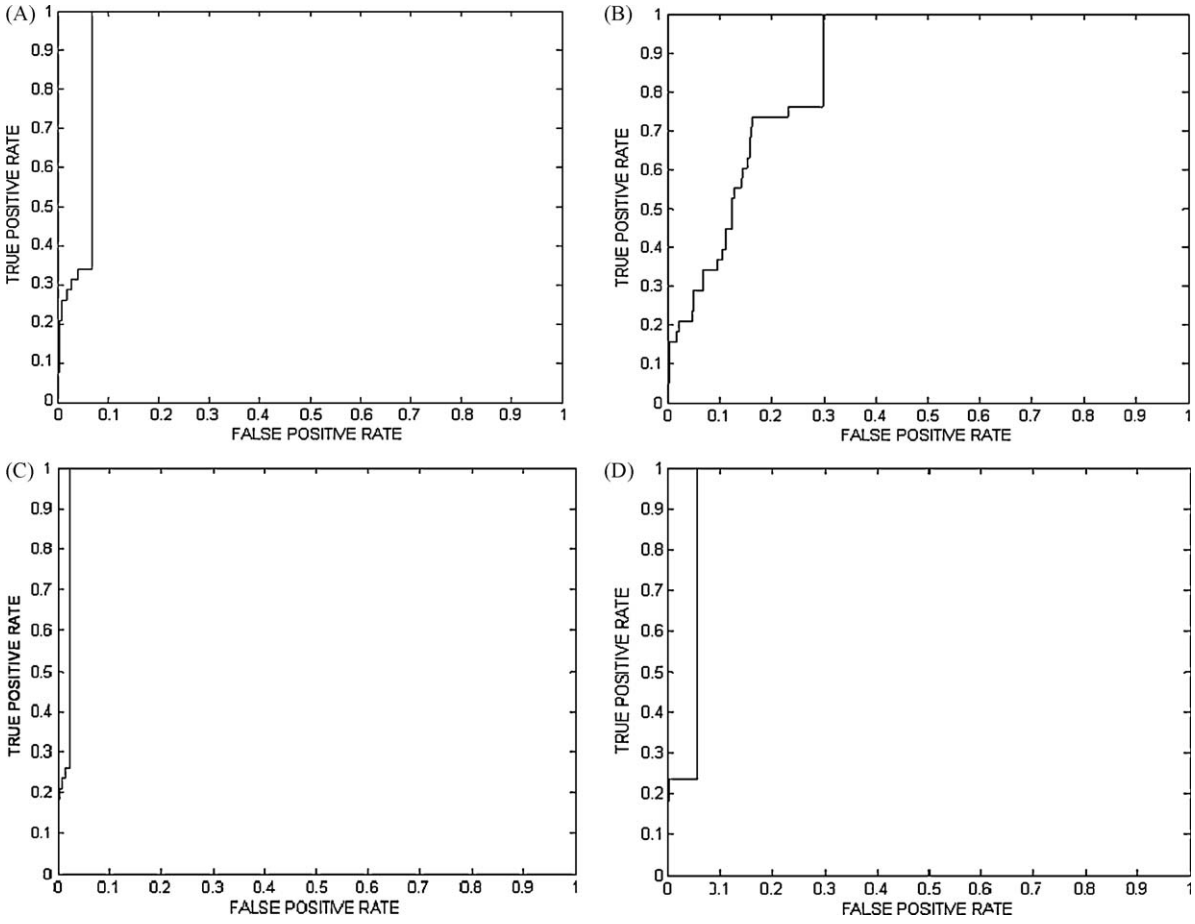


Fig. 4. Received operator curves (ROCs) conducted for QSAR-selected models: (A) HypoA/1/2, (B) HypoD/3/8, (C) HypoD/4/4, and (D) HypoD/4/10.

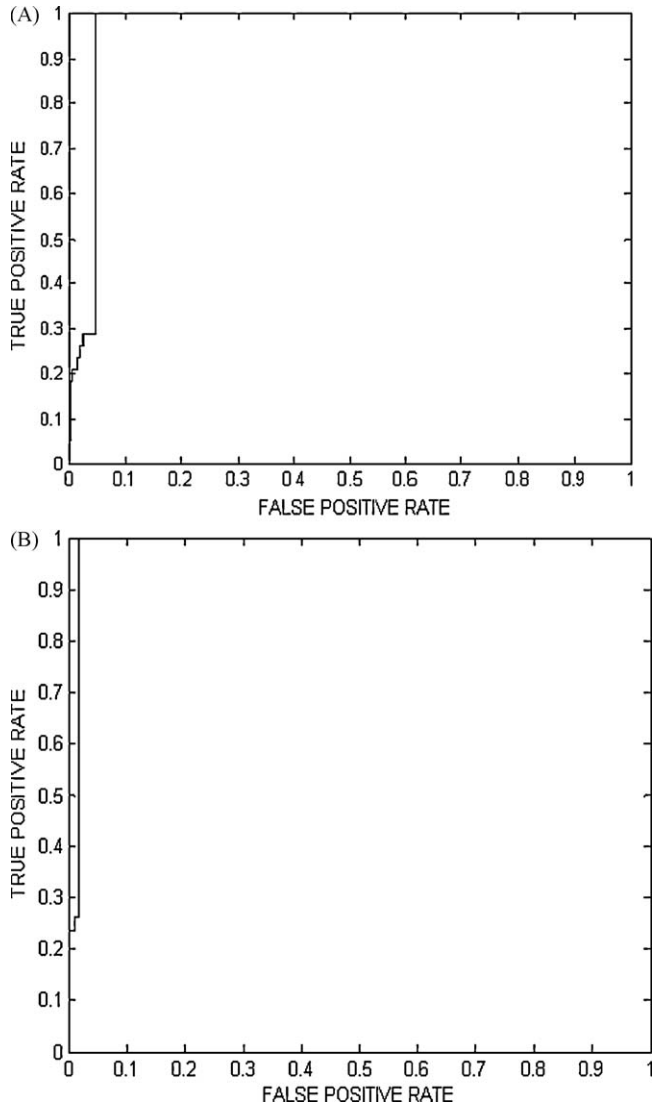


Fig. 5. Received operator curves (ROCs) for merged pharmacophore models: (A) HypoA/1/2-D/3/8 and (B) HypoD/4/4-D/4/10.

assessed employing CFF force field (version 1.02) with a nonbonded cutoff distance of 10.0 Å and distance dependent dielectric. An energy grid extending 3.0 Å from the binding site was implemented. The interaction energy was estimated by tri-linear interpolation value using soft potential energy approximations [74]. (iv) Rigid body ligand minimization parameters: 100 iterations of rigid body minimization were applied to every orientation of the docked ligand. Eventually, a maximum of 10 optimal conformers/poses were saved for each molecule for subsequent scoring. (v) The docked conformers/poses were scored using PLP1 scoring function [75].

2.1.11. *In silico screening for new ERβ inhibitors*

HypoA/1/2-D/3/8 and HypoD/4/4-D/4/10 were employed as 3D search queries to screen two 3D flexible structural databases, namely, the national cancer institute (NCI) list of compounds and our *in house* built database of known drugs and agrochemicals (DAG). Screening was performed employing the "Best Flexible Database Search" option implemented within CATALYST. NCI hits were subsequently filtered based on Lipinski's and Veber's rules [76,77]. However, DAG hits were processed without subsequent post-filtering. Table 8 details the number of captured and refined compounds by each model.

Table 8
Numbers of captured hit compounds by hybrid pharmacophores HypoA/1/2-D/3/8 and HypoD/4/4-D/4/10.

3D database[a]	Post-screening filtering[b]	Pharmacophore models	
		HypoA/1/2-D/3/8	HypoD/4/4-D/4/10
NCI	Before	3635	2808
	After	1176	409
DAG[c]		73	57

[a] NCI: national cancer institute list of available compounds (238,819 structures), DAG: the list of established drugs and agrochemicals (2602 structures).
[b] Post-screening filtering employing Lipinski's and Veber's rules. Two Lipinsk's violations were tolerated.
[c] This list of compounds was virtually screened without post-screening filtering.

Surviving hits were fitted on the original QSAR-selected pharmacophores, i.e., HypoA/1/2, HypoD/3/8, HypoD/4/4 and HypoD/4/10 using the "best fit" option within CATALYST. The resulting fit values together with the relevant molecular descriptors were substituted in the optimal QSAR models **C** and **F** (Table 5 and Eqs. (10) and (11)). Fig. 7 and Table 9 show the highest ranking hits as predicted by successful QSAR models.

3. Results and discussion

3.1. Data mining and conformational coverage

The literature was extensively surveyed to collect as many structurally diverse ERβ ligands as possible (**1**–**119**, see Table A and Figure A in Supporting Information) [2,3,6,8,9,11,16,17,39–55]. The collected compounds were carefully selected from around 800 reported inhibitors in such a way to guarantee dissimilar affinities to ERα and ERβ, and therefore, to allow access to selective ERβ pharmacophore models. The selected ligands have a 10-fold average preference to either estrogen receptor, i.e., compared to the other. We emphasized this trend particularly in the "most potent" and "least potent" categories (as defined by Eqs. (1) and (2)) which illustrate 18 and 5.5 fold average affinity preferences to ERβ and ERα, respectively. For example, **102** has ERα IC50 = 176 nM and ERβ IC50 = 0.8 nM, **105** (ERα IC50 = 23 nM and ERβ IC50 = 0.30 nM), **112** (ERα IC50 = 97 nM and ERβ IC50 = 1 nM). On the other hand, **17** has ERα IC50 = 11 nM and ERβ IC50 = 4900 nM, **20** (ERα IC50 = 8 nM and ERβ IC50 = 1231 nM), **66** (ERα IC50 = 1.4 nM and ERβ IC50 = 129 nM), etc.

The 2D structures of the inhibitors were imported into CATALYST and converted automatically into reasonable 3D structures. The resulting 3D conformers were used as starting points for conformational analysis and in the determination of various molecular descriptors for QSAR modeling. The conformational space of each ligand was extensively sampled utilizing the poling algorithm implemented within the CONFIRM module of CATALYST [60]. Efficient conformational coverage should minimize conformation-related noise during pharmacophore generation and validation stages. Pharmacophore generation and pharmacophore-based search procedures are known for their sensitivity to inadequate conformational sampling of training compounds [78].

3.2. Exploration of ERβ pharmacophoric space

CATALYST enables automatic pharmacophore construction by using a collection of at least 16 molecules with bioactivities spanning over 3.5 orders of magnitude [56–61]. It implements an optimization algorithm that evaluates large number of potential models for a particular target through fine perturbations to

Table 9
High-ranking hit molecules captured by hybrid pharmacophores HypoA/1/2-D/3/8 and HypoD/4/4-D/4/10 and their corresponding QSAR estimates from Eqs. (10) and (11).

Hit[a]	Name or NCI code	Fit values against[b]				QSAR estimates			
						Eq. (10)		Eq. (11)	
		HypoA/1/2	HypoD/3/8	HypoD/4/4	HypoD/4/10	$\log(1/IC_{50})$	IC_{50} (nM)	$\log(1/IC_{50})$	IC_{50} (nM)
120	33338	8.47	5.30	0	0	−0.14	1.38	−0.83	6.76
121	34319	8.30	5.93	0	0	−0.03	1.06	−0.81	6.43
122	57537	6.15	5.26	0	0	−0.41	2.57	−1.13	13.44
123	66810	5.80	6.19	0	0	0.39	0.41	0.34	0.46
124	68053	6.82	5.14	0	0	−0.08	1.21	−0.79	6.11
125	86423	7.19	6.54	0	0	−0.63	4.26	−1.71	51.22
126	328436	5.75	5.05	0	0	0.43	0.37	0.41	0.39
127	366211	7.12	5.16	0	0	−0.40	2.50	−1.06	11.43
128	370875	8.27	6.02	0	0	1.15	0.07	0.40	0.40
129	402886	7.89	5.87	0	0	0.15	0.71	−0.79	6.20
130	605857	5.11	5.89	0.89	0.15	−0.89	7.78	−2.09	121.83
131	652097	8.16	6.86	0	0	0.14	0.73	−1.084	12.13
132	691341	7.47	4.30	0	0	0.24	0.58	−0.51	3.24
133	692899	6.81	5.24	3.74	1.42	−0.16	1.46	−0.43	2.71
134	692902	6.75	5.63	3.70	1.40	0.34	0.46	−0.013	1.03
135	693113	7.47	5.92	0	0	0.37	0.43	0.074	0.84
136	371300	6.75	6.14	0	0	0.20	0.63	0.05	0.90
137	Fexofenadine	3.83	5.25	1.90	0.45	−0.34	2.18	−1.71	51.55
138	Flupenthixol	0	0	0.99	1.002	−2.01	102.06	−2.16	144.95
139	Phenolsulfonphthalein	6.21	3.98	0	0	−2.24	174.70	−0.97	9.28
140	Hydroxystilbamidine	2.50	3.51	0	0	−2.47	291.59	−2.67	468.2
141	Masoprocol	8.38	5.83	0	0	−0.47	2.93	−1.17	14.87
142	Methylergonovine	5.29	0	0	0	−3.07	1169.8	−3.17	1478.84
143	o-Dianisidine	5.39	4.46	0	0	−3.49	3111.9	−3.76	5687.32
144	Terfenadine	5.07	4.99	1.70	0.51	0.33	0.47	−0.54	3.49
145	Topotecan	5.66	3.44	0	0	−2.38	236.97	−1.58	37.86
146	Liothyronine	7.79	5.64	1.88	2.76	−0.65	4.43	−1.57	36.74

[a] Hits shown in Fig. 7.
[b] Best fit values against each binding hypothesis calculated by Eq. (4).

hypotheses that survived the subtractive and constructive phases (see Section 2.1.4) [59]. The number of the evaluated pharmacophores is indicated by the configuration (Config.) cost calculated for each modeling run (see Section 2.1.5). It is generally recommended that the Config. cost of any CATALYST run not to exceed 17, which correspond to 2^{17} hypotheses to be assessed by CATALYST, in order to guarantee thorough analysis of all models [60].

The number of investigated pharmacophores is a function of training compounds, selected input chemical features and other CATALYST control parameters [60]. Restricting the extent of explored pharmacophoric space should improve the efficiency of optimization via allowing elaborate evaluation of limited number of pharmacophoric models. However, extensive restrictions imposed on the pharmacophoric space might reduce the possibility of discovering optimal pharmacophoric hypotheses, as they might occur outside the "boundaries" of the pharmacophoric space.

Therefore, we decided to explore the pharmacophoric space of ERβ ligands under reasonably imposed "boundaries" through sixteen CATALYST automatic runs and employing six carefully selected training subsets from the collected compounds: **A**, **B**, **C**, **D**, **E** and **F** in Table 1. The training compounds in these subsets were selected in such away to guarantee maximal 3D diversity and continuous ERβ affinities spread over more than 3.5 logarithmic cycles. Furthermore, the training ligands were selected in such a way that differences in their ERβ affinities are primarily attributable to the presence or absence of pharmacophoric features (e.g., HBA or HBD or Hbic or RingArom or PosIoniz, Table 2) rather than steric shielding and/or bioactivity-enhancing or -reducing auxiliary groups (e.g., electron donating or withdrawing groups). We gave special emphasis to the 3D diversity of the most-active categories in the training subsets as they have significant influence on the extent of the evaluated pharmacophore space during the Constructive Phase of CATALYST algorithm (see Section 2.1.4).

Guided by our reasonably restricted pharmacophoric exploration concept, we restricted the software to explore pharmacophoric models incorporating from zero to three HBA, HBD, Hydrophobic, and RingArom features and from zero to one PosIon feature, as shown in Table 2. Furthermore, we instructed CATALYST to explore only 4- and 5-featured pharmacophores, i.e., ignore models of lesser number of features as CATALYST can generate pharmacophore models of a maximum of 5 features only [56–61]. This restriction has the advantage of narrowing the investigated pharmacophoric space and better representing the feature-rich nature of known ERβ ligands.

In each run, the resulting binding hypotheses were automatically ranked according to their corresponding "total cost" values (see Section 2.1.5). CATALYST also calculates the cost of the null hypothesis, which presumes that there is no relationship in the data and that experimental activities are normally distributed about their mean. Accordingly, the greater the difference from the null hypothesis cost (residual cost, Table 4) the more likely that the hypothesis does not reflect a chance correlation [56–61].

CATALYST includes additional validation technique known as Cat.Scramble [56]. This procedure is based on Fisher's randomization test [63]. In this test the biological data and the corresponding structures are scrambled several times, and the software is challenged to generate pharmacophoric models from the randomized data. The confidence in the parent hypotheses (i.e., generated from unscrambled data) is lowered proportional to the number of times the software succeeds in generating binding hypotheses from scrambled data of apparently better cost criteria than the parent hypotheses (see Section 2.1.5). Tables 3 and 4 show Fisher criteria for the different runs.

Eventually, 210 pharmacophore models emerged from 24 automatic CATALYST modeling rounds. All models illustrated good overall performance and Fisher confidence levels $\geq$90%, as shown in Table 3. The resulting pharmacophores were clustered and the

best representatives (42 models) were used in subsequent QSAR analysis. Table 4 shows the performance criteria of representative binding models.

Clearly from Table 4, all models shared comparable features and acceptable statistical criteria. Emergence of several comparable pharmacophore models suggests the existence of multiple binding modes assumed by different ERβ ligands within the binding pocket. Therefore, it is quite challenging to select any particular pharmacophore hypothesis as a sole representative of the binding process.

Table 5 shows the statistical criteria of the resulting top-ranking QSAR models. Eqs. (10) and (11) show the details of the best QSAR models, i.e., models **C** and **F** in Table 5, respectively. Fig. 1 shows the corresponding scatter plots of experimental versus estimated bioactivities for the training and testing ligands based on both equations.

$$\begin{aligned}\log(1/\mathrm{IC}_{50}) = {} & -8.53 + 0.25\,\mathrm{HypoD/4/4} + 0.12\,\mathrm{HypoA/1/2} + 0.27\,\mathrm{HypoD/3/8} + 0.37\,\mathrm{SaaO} \\ & + 0.63\,{}^{0}\chi - 5.7 \times 10^{-3}\,\mathrm{JursDPSA2} - 18.27\,\mathrm{JursFNSA3} - 0.63\,\mathrm{AtypeO60} \\ r^2_{96} = 0.79,\ & F\text{-statistic} = 40.96, \quad r^2_{\mathrm{BS}} = 0.79, \quad r^2_{\mathrm{LOO}} = 0.74, \quad r^2_{\mathrm{PRESS(23)}} = 0.54\end{aligned} \tag{10}$$

$$\begin{aligned}\log(1/\mathrm{IC}_{50}) = {} & -4.87 + 0.25\,\mathrm{HypoD/3/8} + 0.10\,\mathrm{HypoA/1/2} + 0.25\,\mathrm{HypoD/4/10} \\ & + 0.42\,{}^{3}\chi^{v}_{P} + 1.86\,\mathrm{AtypeC43} - 1.08 \times 10^{-3}\,\mathrm{PMImag} - 1.79\,\mathrm{LUMO} \\ & - 0.42\,\mathrm{AtypeO60} + 3.27\,\mathrm{AtypeC35} - 0.57\,\mathrm{AtypeO58} + 0.25\log P \\ r^2_{96} = 0.83, \quad & F\text{-statistic} = 36.20, \quad r^2_{\mathrm{BS}} = 0.82, \quad r^2_{\mathrm{LOO}} = 0.76, \quad r^2_{\mathrm{PRESS(23)}} = 0.56\end{aligned} \tag{11}$$

3.3. QSAR modeling

Pharmacophoric hypotheses provide excellent insights into ligand-macromolecule recognition and can be used as 3D search queries to mine for new biologically interesting scaffolds. However, their predictive value as 3D-QSAR models is usually limited by steric shielding and bioactivity-enhancing or -reducing auxiliary groups [61]. This point combined with the fact that pharmacophore modeling of ERβ ligands furnished several binding hypotheses of comparable statistical criteria prompted us to employ classical QSAR analysis to search for the best combination of pharmacophore(s) and other 2D descriptors capable of explaining bioactivity variation across the whole list of collected inhibitors (**1–119** in Table A and figure A in supporting Information). The fact that the collected inhibitors exhibit distinct (generally contradictory) affinity profiles against ERα and ERβ confers additional selectivity on the resulting QSAR models, i.e., extra to pharmacophoric selectivity discussed earlier.

We employed genetic function approximation and multiple linear regression analysis (GFA-MLR-QSAR) to search for an optimal QSAR equation(s). GFA-MLR-QSAR selects optimal descriptor combinations based on the Darwinian concept of genetic evolution whereby the statistical criteria of regression models from different descriptor combinations (chromosomes) are employed as fitness criteria [65]. GFA-MLR-QSAR analysis was employed to explore various combinations of pharmacophores and other structural descriptors and to evaluate their statistical properties as predictive QSAR models. The fit values obtained by mapping the 42 representative hypotheses against all collected ERβ ligands (**1–119**, Table A and Figure A in Supporting Information) were enrolled as independent variables (genes) in a cycle of GFA-MLR-QSAR analysis over 30,000 iterations employing Friedman's LOF fitness criterion (see Section 2.1.7) [65,82]. However, since it is essential to access the predictive power of the resulting QSAR models on an external set of compounds, we randomly selected 23 molecules and employed them as external test molecules for validating the QSAR models (r^2_{PRESS}). Moreover, all QSAR models were cross-validated automatically using the leave-one-out cross-validation in CERIUS2 [65,79].

where r^2_{96} is the correlation coefficient, r^2_{LOO} is the leave-one-out correlation coefficient, r^2_{BS} is the bootstrapping regression coefficient and r^2_{PRESS} is the predictive r^2 determined for the 23 test compounds [65,79]. HypoA/1/2, HypoD/3/8, HypoD/4/4 and HypoD/4/10 represent the fit values of the training compounds against the four pharmacophore models as calculated from Eq. (4). SaaO is the electrotopological sum descriptor for aromatic oxygen atoms. ${}^{0}\chi$ and ${}^{3}\chi^{v}_{P}$ are the zero-order simple and third-order path valence connectivity indices, respectively. AtypeO60, AtypeO58, AtypeO60, AtypeC35 and AtypeC43 are atom-type-based AlogP descriptors related to oxygen and carbon atoms. JursFNSA3 is a Jurs charged partial surface area descriptor encoding for the fractional charged partial surface area (=partial positive solvent-accessible surface area × (total positive charge/total molecular solvent-accessible surface area)), while JursDPSA2 is the difference in total charge weighted surface areas obtained by subtracting the total charge weighted positive solvent-accessible surface area (=partial positive solvent-accessible surface area × total positive charge) minus total charge weighted negative solvent-accessible surface area (=partial negative solvent-accessible surface area × total negative charge) [65]. LUMO is the lowest unoccupied molecular orbital; log *P* is the calculated logarithm of the n-octanol/water partition coefficient, PMImag is the principal moment of inertia.

The emergence of generally orthogonal pharmacophoric models (of average cross-correlation r^2 = 0.36) in Eqs. (10) and (11) suggests the existence several complementary binding modes accessible to ligands within the binding pocket of ERβ: i.e., a particular pharmacophore can optimally explain the affinities of some training inhibitors, while the other explains the remaining inhibitors. Similar conclusions were reached about the binding pockets of other targets, e.g., factor Xa, GSK-3beta, bacterial Mur F and HSL [80–83].

Fig. 2 shows the pharmacophoric features of models HypoA/1/2 and HypoD/3/8 and how they map most-active training compound **117** (IC_{50} = 0.136 nM), while Fig. 3 shows the pharmacophoric features of HypoD/4/4 and HypoD/4/10 and how they map most-active training compound **5** (IC_{50} = 0.6 nM). Table 6 shows the *X*, *Y*, and *Z* coordinates of the pharmacophores in Eqs. (10) and (11).

Emergence of PMImag, SaaO and connectivity indices is suggestive of certain role played by the ligands' topology in the binding process. However, despite their predictive significance, these descriptors have generally poor information content. On the other hand, the combined emergence of hydrophilicity indicators, i.e., JursDPSA2, JursFNSA3, AtypeO60, AtypeO58, and AtypeO60, in association with negative regression terms and hydrophobic

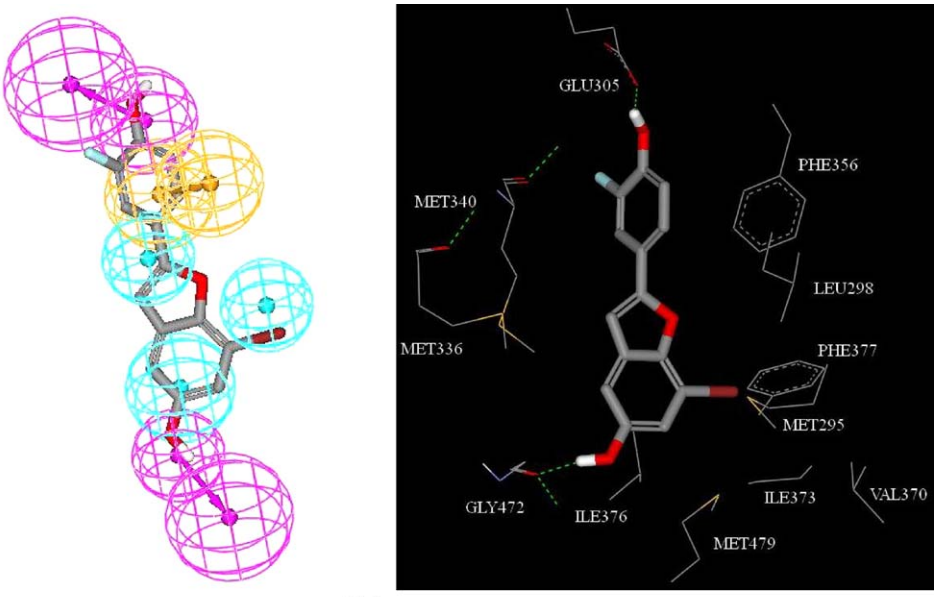
descriptors, i.e., log *P*, AtypeC35, and AtypeC43, in association with positive regression coefficients strongly suggest an inverse relationship tying ligand/ERβ affinity and ligand hydrophilicity. This is unsurprising, as ligands need to dehydrate prior to binding to the hydrophobic pocket of ERβ.

Finally, emergence of LUMO in Eq. (11) in association with a negative slope suggests a direct relationship between ligand/ERβ affinity and ligands' electrophilicities. We believe this trend is explainable by the fact that ERβ/ligand binding probably involves aromatic ring stacking with PHE356, which is expected to be stronger with electron-deficient rings, i.e., those of more negative LUMO energies.

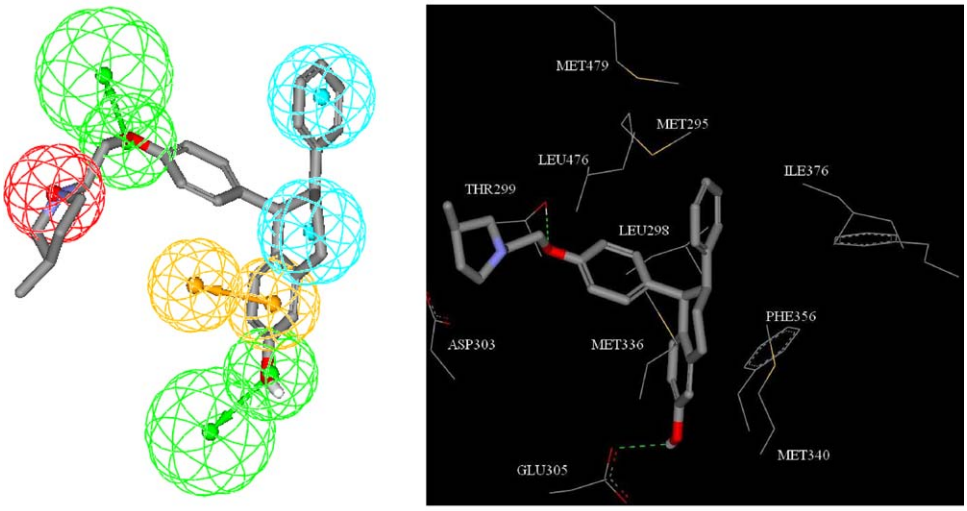
3.4. ROC analyses of successful pharmacophores

To further validate the resulting models (both QSARs and pharmacophores), we subjected our QSAR-selected pharmacophores to receiver operator characteristic (ROC) curve analyses. In ROC analysis, the ability of a particular pharmacophore model to correctly classify a list of compounds as actives or inactives is indicated by the area under the curve (AUC) of the corresponding ROC curve as well as other parameters, i.e., overall accuracy, overall specificity, overall true positive rate, and overall false negative rate (see Section 2.1.8).

Table 7 and Fig. 4 show the ROC curves of QSAR-selected pharmacophores. Generally, all four models illustrated excellent overall performance, as they scored average AUCs above 85%. However, HypoD/4/4 performed best with AUC value of 98%. This is not unexpected, as the presence of PosIons and two HBA features in this model points to the hydrophilic nature and therefore the selectivity of this pharmacophore. While on the other hand, the hydrophobic nature of HypoD/3/8 seems to explain its relatively inferior selectivity. Well-positioned hydrophilic groups promote selective ligand–receptor interactions and accordingly promote pharmacophoric selectivity and ROC-AUC. The midway performance of HypoA/1/2 and HypoD/4/10, i.e., between those of HypoD/4/4 and HypoD/3/8 is probably related to the intermediacy of their hydrophobic-hydrophilic nature. Nevertheless, despite the mediocre AUC of HypoD/3/8 it captured 76.3% of the true actives albeit with relatively high-negative rate (3.18%).



(A)



(B)

Fig. 6. Mapping hybrid pharmacophore models against potent training compounds compared to their docked poses into ERβ (protein databank code: 2FSZ, resolution 2.20 Å). (A) The docked pose of **117** compared to the way it fits HypoA/1/2-D/3/8 and (B) the docked pose of **5** compared to the way it maps HypoD/4/4-D/4/10.

3.5. Merging similar pharmacophore models

The close resemblance of QSAR-selected pharmacophore models combined with the urge to further optimize their ROC profiles prompted us to merge similar models into hybrid pharmacophores: HypoA/1/2-D/3/8 and HypoD/4/4-D/4/10, as in Figs. 2 and 3.

ROC analysis of the new hybrid models illustrated significant improvements in AUC compared to parent models, as shown in Table 7 and Figs. 4 and 5. The improvement is particularly apparent in HypoA/1/2-D/3/8 case, as the average AUC improved from 0.908 for models HypoA/1/2 and HypoD/3/8 to become 0.964. Furthermore the pharmacophoric features of the new hybrid models agree nicely with binding interactions proposed by docking experiments, as shown in the next section.

3.6. Comparison of pharmacophore model with the active site of ERβ

Despite the problems associated with crystallographic structures (see Section 3.1), binding pharmacophoric features obtained by pharmacophore-QSAR analysis can be compared with crystallographic ERβ active site to identify probable residues important for activity [80–82,84].

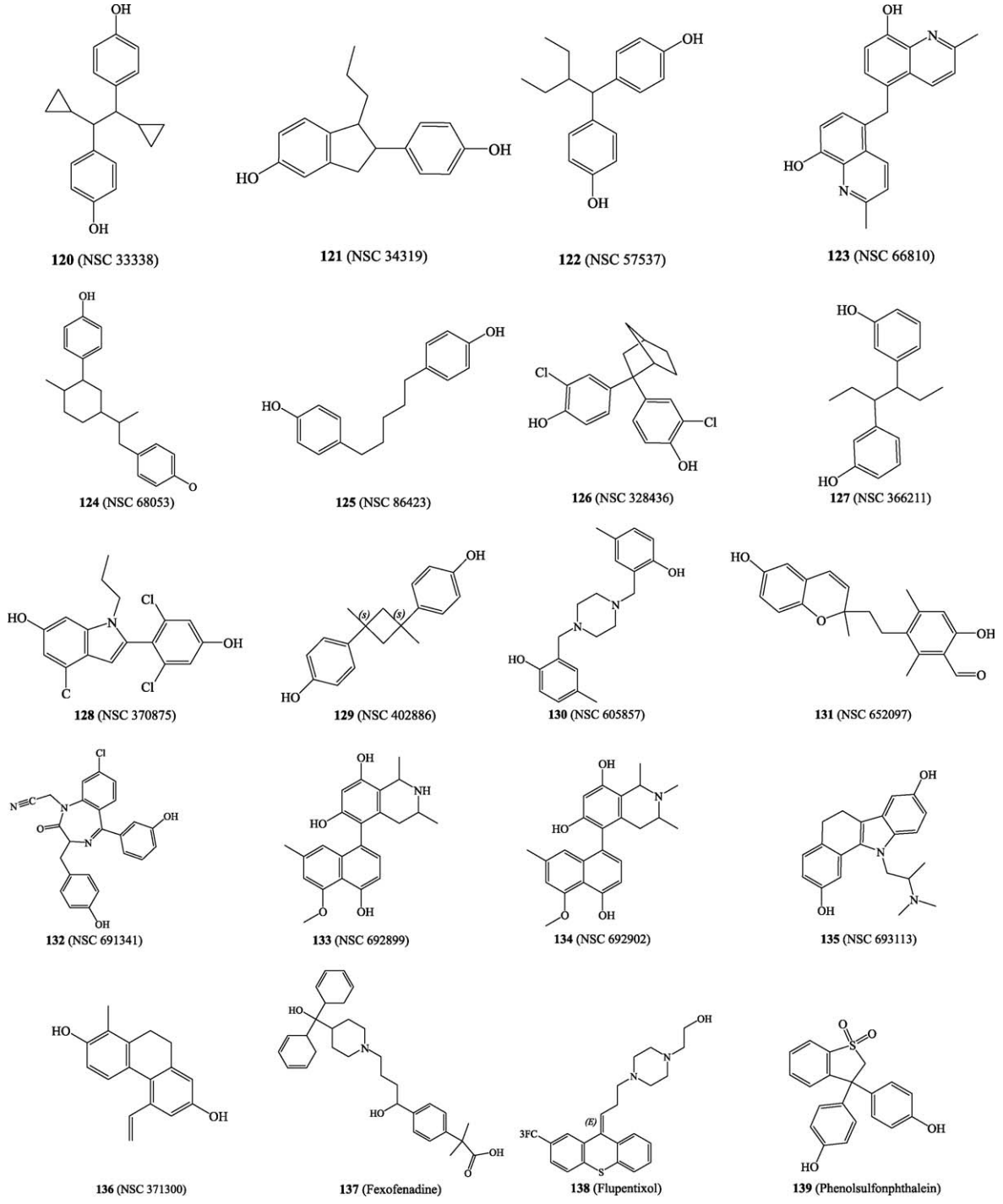


Fig. 7. Chemical structures of highest ranking *in silico* hits (as suggested by the best QSAR model and the associated pharmacophores).

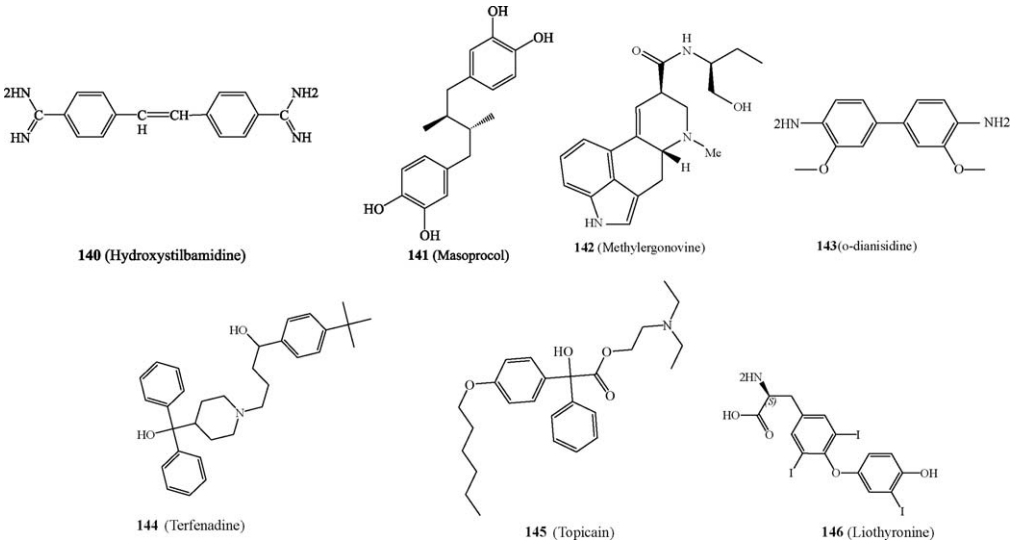


Fig. 7. (*Continued*).

Pharmacophoric features of HypoA/1/2-D/3/8 and HypoD/4/4-D/4/10, as well as the alignment of compounds **117** and **5** (Table A and Figure A in Supporting Information) as proposed by these binding hypotheses were compared with the structures of **117** and **5** docked into the binding site of ERβ (Fig. 6). The docking experiment was performed employing LIGANDFIT docking engine [73,74]. A marked similarity was observed between features proposed by the merged pharmacophore models and ligand binding features in the docked structures.

In the docked pose of **117** (Fig. 6a), the hydroxyl group of the *o*-fluorophenol fragment is hydrogen-bonded to the carboxylate of GLU305. This interaction corresponds to mapping the phenolic hydroxyl against a hydrogen-bond donor in HypoA/1/2-D/3/8 (Fig. 6a). Similarly, in the docked pose of **117**, the hydroxy group of the *m*-iodophenol fragment is hydrogen-bonded to the peptidic carbonyl of GLY472 (Fig. 6a). This interaction corresponds nicely to mapping this phenolic hydroxyl against a hydrogen-bond donor in HypoA/1/2-D/3/8 (Fig. 6a).

Furthermore, the hydrophobic amino acids LEU298, PHE377, MET295, VAL370, ILE373, MET479, ILE376, MET336 and MET340 seem to furnish suitable hydrophobic environment for effective interactions with the iodobenzofuran fragment of **117** (Fig. 6a). This conduct agrees nicely with fitting this central fragment against three hydrophobic features in HypoA/1/2-D/3/8 (Fig. 6a). Finally, the docking experiment suggests that the *o*-flurophenol fragment of **117** stacks against the aromatic ring of PHE356, which correlates with mapping this moiety against a RingArom feature in HypoA/1/2-D/3/8 (Fig. 6a).

Likewise, the docked pose of **5** within ERβ binding site closely correlates with way it maps HypoD/4/4-D/4/10 (Fig. 6b): the docking experiment placed the aromatic ether oxygen of **5** adjacent to the hydroxyl side chain of THR299 suggesting the existence of mutual hydrogen-bonding (Fig. 6b), which seems to correspond nicely with mapping this group against a hydrogen-bond acceptor feature in the hybrid pharmacophore HypoD/4/4-D/4/10 (Fig. 6b). Furthermore, the docking experiment suggests the involvement of the phenolic hydroxyl group of **5** in hydrogen-bonding with the carboxylate side chain of GLU305 (Fig. 6b). This interaction is probably represented by mapping this phenolic hydroxyl against a hydrogen-bond acceptor in HypoD/4/4-D/4/10 (Fig. 6b). On the other hand, the docking engine placed the positively charged pyrrolidine fragment of **5** at 4.4 Å from the carboxylate moiety of ASP303 suggesting the existence of effective mutual electrostatic attraction (Fig. 6b). This corresponds to mapping the pyrrolidine group of **5** against positive ionizable feature in HypoD/4/4-D/4/10 (Fig. 6b). Moreover, the docking software stacks the phenolic ring of **5** against the benzene of PHE356, which clearly correlates with mapping this ring against a RingArom feature in HypoD/4/4-D/4/10 (Fig. 6b). Finally, the amino acids MET336, LEU298, LEU476, MET295, MET479, and ILE376 seem to furnish hydrophobic pocket suitable for interaction with the central phenylcyclohexenyl fragment of **5**. This conduct corresponds with fitting this fragment against two hydrophobic features in HypoD/4/4-D/4/10 (Fig. 6b).

3.7. In silico screening

Models HypoA/1/2-D/3/8 and HypoD/4/4-D/4/10 were employed as 3D search queries against two available 3D flexible structural databases, namely, the NCI list of compounds (238,819 compounds) and our *in house* built database of known drugs and agrochemicals (DAG, 2602 compounds). HypoA/1/2-D/3/8 and HypoD/4/4-D/4/10 captured 3708 and 2865 hits, respectively. Hits are defined as those compounds that have their chemical groups spatially overlap (map) with corresponding features within the particular pharmacophoric model.

NCI hits were subsequently filtered via Lipinski's and Veber's drug-likeness rules [76,77] to further optimize our search results. Table 8 details the number of captured hits and surviving hits after refinement by Lipinski's and Veber's rules. The remaining hits were fitted against HypoA/1/2, HypoD/3/8, HypoD/4/4 and HypoD/4/10 and their fit values were substituted in QSAR Eqs. (10) and (11) to determine their predicted bioactivities. The highest ranking hits, together with their estimated bioactivities, are shown in Fig. 7 and Table 9.

Interestingly, one of our drug hits, fexophenadine (**137**), was reported to possess potent anti-rheumatic and anti-inflammatory properties [85] suggesting their ERβ affinities, which further validates our molecular modeling approach and conclusions.

4. Conclusion

ERβ selective agonists have utility as novel anti-inflammatory agents without triggering classic estrogenic side effects such as uterine stimulation. The pharmacophoric space of ERβ ligands was explored via six diverse sets of compounds, of significantly dissimilar ERα/ERβ affinities, and using CATALYST to identify high-quality binding model(s). Subsequently, genetic algorithm and

multiple linear regression analysis were employed to access optimal QSAR models capable of explaining ERβ affinity variation across 119 collected ERβ ligands. Four pharmacophores emerged in two QSAR equations, suggesting the existence of more than one binding mode accessible to ligands within ERβ binding pocket. The close similarity between the pharmacophores prompted us to merge them into two hybrid models that illustrated better ROC profiles and resembled binding features suggested by docking experiments.

The pharmacophoric models were employed to propose several potent ERβ ligands retrieved via *in silico* mining of two structural databases. Our results suggest usefulness of the combination of pharmacophoric exploration and QSAR analyses as a tool for finding new ERβ ligands.

Acknowledgements

This project was partially sponsored by the Faculty of Graduate Studies (M.Sc. Thesis of Mai Tarairah). The authors thank the Deanship of Scientific Research and Hamdi-Mango Center for Scientific Research at the University of Jordan for their generous funds. The authors also thank Dr. M.K. Mohammad for his useful discussions.

Appendix A. Supplementary data

Supplementary data associated with this article can be found, in the online version, at doi:10.1016/j.jmgm.2009.09.005.

References

[1] E.S. Manas, Z.B. Xu, R.J. Unwalla, W.S. Somers, Understanding the selectivity of genistein for human estrogen receptor-β using X-ray crystallography and computational methods, Structure 12 (2004) 2197–2207.
[2] J.W. Ullrich, R.J. Unwalla, R.R. Singhaus, H.A. Harris, R.E. Mewshaw, Estrogen receptor β ligands: design and synthesis of new 2-phenyl-isoindole-1,3-diones, Bioorg. Med. Chem. Lett. 17 (2007) 118–122.
[3] A.T. Vu, A.N. Campbell, H.A. Harris, R.J. Unwalla, E.S. Manas, R.E. Mewshaw, ERβ ligands. Part 6: 6H-chromeno[4,3-b]quinolines as a new series of estrogen receptor b-selective ligands, Bioorg. Med. Chem. Lett. 17 (2007) 4053–4056.
[4] M.D. Angelis, F. Stossi, K.A. Carlson, B.S. Katzenellenbogen, J.A. Katzenellenbogen, Indazole estrogens: highly selective ligands for the estrogen receptor β, J. Med. Chem. 48 (2005) 1132–1144.
[5] E.H. Kong, A.C.W. Pike, R.E. Hubbard, Structure and mechanism of the oestrogen receptor, Biochem. Soc. Trans. 31 (2003) 56–59.
[6] R.J. Edsall, H.A. Harris, E.S. Manas, R.E. Mewshaw, ERβ ligands. Part 1: The discovery of ERβ selective ligands which embrace the 4-hydroxy-biphenyl template, Bioorg. Med. Chem. 11 (2003) 3457–3474.
[7] A.C.W. Pike, A.M. Brzozowski, R.E. Hubbard, T. Bonn, A.G. Thorsell, O. Engstrom, J. Ljunggren, J.-A. Gustafsson, M. Carlquist, Structure of the ligand-binding domain of oestrogen receptor beta in the presence of a partial agonist and a full antagonist, EMBO J. 18 (1999) 4608–4618.
[8] M.S. Malamas, E.S. Manas, R.E. McDevitt, I. Gunawan, Z.B. Xu, M.D. Collini, C.P. Miller, T. Dinh, R.A. Henderson, J.C. Keith, H.A. Harris, Design and synthesis of aryl diphenolic azoles as potent and selective estrogen receptor-β ligands, J. Med. Chem. 47 (2004) 5021–5040.
[9] C. Yang, R. Edsall, H.A. Harris, Z. Xiaochun, E.S. Manas, R.E. Mewshawa, ERb ligands. Part 2: Synthesis and structure–activity relationships of a series of 4-hydroxy-biphenyl-carbaldehyde oxime derivatives, Bioorg. Med. Chem. 12 (2004) 2553–2570.
[10] H.A. Harris, L.M. Albert, Y. Leathurby, M.S. Malamas, R.E. Mewshaw, C.P. Miller, Y.P. Kharode, J. Marzolf, B.S. Komm, R.C. Winneker, D.E. Frail, R.A. Henderson, Y. Zhu, J.C. Keith Jr., Evaluation of an estrogen receptor-β agonist in animal models of human disease, Endocrinology 144 (2003) 4241–4249.
[11] B.O. Nilsson, Modulation of the inflammatory response by estrogens with focus on the endothelium and its interactions with leukocyte, Inflamm. Res. 56 (2007) 269–273.
[12] H.A. Harris, K.L. Bruner-Tran, X. Zhang, K.G. Osteen, C.R. Lyttle, A selective estrogen receptor-β agonist causes lesion regression in an experimentally induced model of endometriosis, Hum. Reprod. 20 (2005) 936–941.
[13] P.A. Cristofaro, S.M. Opal, J.E. Palardy, N.A. Parejo, J. Jhung, J.C. Keith, H.A. Harris, WAY-202196, a selective estrogen receptor-beta agonist, protects against death in experimental septic shock, Crit. Care Med. 34 (2006) 2188–2193.
[14] R.H. Straub, The complex role of estrogens in inflammation, Endocr. Rev. 28 (2007) 521–574.
[15] H.A. Harris, Estrogen receptor-beta: recent lessons from in vivo studies, Mol. Endocrinol. 21 (2007) 1–13.
[16] T.A. Blizzard, C. Gude, J.D. Morgan, W. Chan, E.T. Birzin, M. Mojena, C. Tudela, F. Chen, K. Knecht, Q. Su, B. Kraker, R.T. Mosley, M.A. Holmes, N. Sharma, P.M.D. Fitzgerald, S.P. Rohrer, M.L. Hammond, Androstenediol analogs as ER-β-selective SERMs, Bioorg. Med. Chem. Lett. 16 (2006) 834–838.
[17] W. Sun, L.D. Cama, e.T. Birzin, S. Warrier, L. Locco, R. Mosley, M.L. Hammond, S.P. Rohrer, 6H-benzo[c]chromen-6-one derivatives as selective ERβ agonists, Bioorg. Med. Chem. Lett. 16 (2006) 1468–1472.
[18] G. Kiss, N.W. Allen, Automated docking of estrogens and SERMs into an estrogen receptor alpha and beta isoform using the PMF forcefield and the Lamarckian genetic algorithm, Theor. Chem. Acc. 117 (2007) 305–314.
[19] T. Dottorini, P. Cozzini, Probing the binding of ligands to estrogen receptor using an empirical system, Int. J. Quant. Chem. 106 (2006) 641–646.
[20] Z. Papoutsi, E. Kassi, N. Fokialakis, S. Mitakou, G. Lambrinidis, E. Mikros, P. Moutsatsou, Deoxybenzoins are novel potent selective estrogen receptor modulators, Steroids 72 (2007) 693–704.
[21] I. Barrett, M.J. Meegan, R.B. Hughes, M. Carr, A.J.S. Knox, N. Artemenko, G. Golfis, D.M. Zisterer, D.G. Lloyd, Synthesis, biological evaluation, structural–activity relationship, and docking study for a series of benzoxepin-derived estrogen receptor modulators, Bioorg. Med. Chem. 16 (2008) 9554–9573.
[22] M. Spreafico, E. Boriani, E. Benfenati, M. Novic, Structural features of diverse ligands influencing binding affinities to estrogen alpha and estrogen beta receptors. Part II. Molecular descriptors calculated from conformation of the ligands in the complex resulting from previous docking study, Mol. Divers. 11 (2007) 171–181.
[23] A.J.S. Knox, M.J. Meegan, V. Sobolev, D. Frost, D.M. Zisterer, D.C. Williams, D.G. Lloyd, Target specific virtual screening: optimization of an estrogen receptor screening platform, J. Med. Chem. 50 (2007) 5301–5310.
[24] P. Wolohan, D.E. Reichert, CoMSIA and docking study of rhenium based estrogen receptor ligand analogs, Steroids 72 (2007) 247–260.
[25] P. Wolohan, D.E. Reichert, CoMFA and docking study of novel estrogen receptor subtype selective ligands, J. Comput. Aid Mol. Des. 17 (2003) 313–328.
[26] K.F. Koehler, S.N. Rao, J.P. Snyder, in: N.C. Cohen (Ed.), Modeling Drug-Receptor Interactions. Guidebook on Molecular Modeling in Drug Design, Academic Press, CA, 1996, pp. 253–255.
[27] N.R.A. Beeley, C. Sage, GPCRs: an update on structural approaches to drug discovery, Targets 2 (2003) 19–25.
[28] B.R. Henke, T.G. Consler, N. Go, R.L. Hale, D.R. Hohman, S.A. Jones, A.T. Lu, L.B. Moore, J.T. Moore, L.A. Orband-Miller, R.G. Robinett, J. Shearin, P.K. Spearing, E.L. Stewart, P.S. Turnbull, S.L. Weaver, S.P. Williams, G.B. Wisely, M.H. Lambert, A new series of estrogen receptor modulators that display selectivity for estrogen receptor β, J. Med. Chem. 45 (2002) 5492–5505.
[29] A.K. Shiau, D. Barstad, J.T. Radek, M.J. Meyers, K.W. Nettles, B.S. Katzenellenbogen, J.A. Katzenellenbogen, D.A. Agard, G.L. Greene, Structural characterization of a subtype-selective ligand reveals a novel mode of estrogen receptor antagonism, Nat. Struct. Biol. 9 (2002) 359–364.
[30] B.H. Norman, T.I. Richardson, J.A. Dodge, L.A. Pfeifer, G.L. Durst, Y. Wang, J.D. Durbin, V. Krishnan, S.R. Dinn, S. Liu, J.E. Reilly, K.T. Ryter, Benzopyrans as selective estrogen receptor β agonists (SERBAs) functionalization of the benzopyran A-ring, Bioorg. Med. Chem. Lett. 17 (2007) 5082–5085.
[31] R.E. McDevitt, M.S. Malamas, E.S. Manas, R.J. Unwalla, Z.B. Xu, C.P. Miller, H.A. Harris, Estrogen receptor ligands: design and synthesis of new 2-arylindene-1-ones, Bioorg. Med. Chem. Lett. 15 (2005) 3137–3142.
[32] R.E. Mewshaw, S.M. Bowen, H.A. Harris, Z.B. Xu, E.S. Manas, S.T. Cohn, ERβ ligands. Part 5: Synthesis and structure–activity relationships of a series of 4′-hydroxyphenyl-aryl-carbaldehyde oxime derivatives, Bioorg. Med. Chem. Lett. 17 (2007) 902–906.
[33] A.M. Brzozowski, A.C.W. Pike, Z. Dauter, R.E. Hubbard, T. Bonn, O. Engstrom, L. Ohman, G.L. Greene, J.-A. Gustafsson, M. Carlquist, Molecular basis of agonism and antagonism in the oestrogen receptor, Nature 389 (1997) 753–758.
[34] Y. Wang, N.Y. Chirgadze, S.L. Briggs, S. Khan, E.V. Jensen, T.P. Burris, A second binding site for hydroxytamoxifen within the coactivator-binding groove of estrogen receptor β, Proc. Natl. Acad. Sci. U.S.A. 103 (2006) 9908–9911.
[35] B.H. Norman, J.A. Dodge, T.I. Richardson, P.S. Borromeo, C.W. Lugar, S.A. Jones, K. Chen, Y. Wang, G.L. Durst, R.J. Barr, C. Montrose-Rafizadeh, H.E. Osborne, R.M. Amos, S. Guo, A. Boodhoo, V. Krishnan, Benzopyrans are selective estrogen receptor β agonists with novel activity in models of benign prostatic hyperplasia, J. Med. Chem. 49 (2006) 6155–6157.
[36] B. Waszkowycz, in: A.L. Harvey (Ed.), New Methods for Structure-based De Novo Drug Design. Advances in Drug Discovery Techniques, John Wiley & Sons, Chichester, 1998, pp. 150–153.
[37] R.A. Silverman, The Organic Chemistry of Drug Design and Drug Action, Academic Press, San Diego, 1991.
[38] M.A. DePristo, P.I.W. de Bakker, T.L. Blundell, Heterogeneity and inaccuracy in protein structures solved by X-ray crystallography, Structure 12 (2004) 831–838.
[39] T.A. Blizzard, F. DiNinno, J.D. Morgan, W. Chan, E.T. Birzin, L.Y. Pai, E.C. Hayes, C.A. DaSilva, R.T. Mosley, Y.T. Yang, S.P. Rohrer, F. DiNinno, M.L. Hammond, Estrogen receptor ligands. Part 14: Application of novel antagonist side chains to existing platforms, Bioorg. Med. Chem. Lett. 15 (2005) 5124–5128.
[40] T.A. Blizzard, F. DiNinno, H.Y. Chen, S. Kim, J.Y. Wu, W. Chan, E.T. Birzin, Y.T. Yang, L.Y. Pai, E.C. Hayes, C.A. DaSilva, S.P. Rohrer, J.M. Schaeffer, M.L. Hammond, Estrogen receptor ligands. Part 13: Dihydrobenzoxathiin SERMs with an optimized antagonist side chain, Bioorg. Med. Chem. Lett. 15 (2005) 3912–3916.
[41] T.A. Blizzard, F. DiNinno, J.D. Morgan, J.Y. Wu, H.Y. Chen, S. Kim, W. Chan, E.T. Birzin, Y.T. Yang, L.Y. Pai, Z. Zhang, E.C. Hayes, C.A. DaSilva, W. Tang, S.P. Rohrer, J.M. Schaeffer, M.L. Hammond, Estrogen receptor ligands. Part 8: Dihydrobenzoxathiin SERMs with heteroatom-substituted side chains, Bioorg. Med. Chem. Lett. 14 (2004) 3865–3868.

[42] T.A. Blizzard, F. DiNinno, J.D. Morgan, H.Y. Chen, J.Y. Wu, C. Gude, S. Kim, W. Chan, E.T. Birzin, Y.T. Yang, L.Y. Pai, Z. Zhang, E.C. Hayes, C.A. DaSilva, W. Tang, S.P. Rohrer, J.M. Schaeffer, M.L. Hammond, Estrogen receptor ligands. Part 7: Dihydrobenzoxathiin SERMs with bicyclic amine side chains, Bioorg. Med. Chem. Lett. 14 (2004) 3861–3864.
[43] H.Y. Chen, S. Kim, J.Y. Wu, E.T. Birzin, W. Chan, Y.T. Yang, J. Dahllund, F. DiNinno, S.P. Rohrer, J.M. Schaeffer, M.L. Hammond, Estrogen receptor ligands. Part 3: The SAR of dihydrobenzoxathiin SERMs, Bioorg. Med. Chem. Lett. 14 (2004) 2551–2554.
[44] H.Y. Chen, K.D. Dykstra, E.T. Birzin, K. Frisch, W. Chan, Y.T. Yang, R.T. Mosley, F. DiNinno, S.P. Rohrer, J.M. Schaeffer, M.L. Hammond, Estrogen receptor ligands. Part 1: The discovery of flavanoids with subtype selectivity, Bioorg. Med. Chem. Lett. 14 (2004) 1417–1421.
[45] M.D. Collini, D.H. Kaufman, E.S. Manas, H.A. Harris, R.A. Henderson, Z.B. Xu, R.J. Unwalla, C.P. Miller, 7-Substituted 2-phenyl-benzofurans as ERβ selective ligands, Bioorg. Med. Chem. Lett. 14 (2004) 4925–4929.
[46] K.D. Dykstra, L. Guo, E.T. Birzin, W. Chan, Y.T. Yang, E.C. Hayes, C.A. DaSilva, L.Y. Pai, R.T. Mosley, B. Kraker, P.M.D. Fitzgerald, F. DiNinno, S.P. Rohrer, J.M. Schaeffera, M.L. Hammonda, Estrogen receptor ligands. Part 16: 2-Aryl indoles as highly subtype selective ligands for ERα, Bioorg. Med. Chem. Lett. 17 (2007) 2322–2328.
[47] S. Kim, J. Wu, H.Y. Chen, E.T. Birzin, W. Chan, Y.T. Yang, Colwell, L. a Susan Li, J. Dahllund, F. DiNinno, S.P. Rohrer, J.M. Schaeffer, M.L. Hammond, Estrogen receptor ligands. Part 4: The SAR of the syn-dihydrobenzoxathiin SERAMs, Bioorg. Med. Chem. Lett. 14 (2004) 2741–2745.
[48] S. Kim, J.Y. Wu, E.T. Birzin, K. Frisch, W. Chan, L.Y. Pai, Y.T. Yang, R.T. Mosley, P.M.D. Fitzgerald, N. Sharma, J. Dahllund, A.-G. Thorsell, F. DiNinno, S.P. Rohrer, J.M. Schaeffer, M.L. Hammond, Estrogen receptor ligands. II. Discovery of benzoxathiins as potent selective estrogen receptor α modulators, J. Med. Chem. Lett. 47 (2004) 2171–2175.
[49] J. Liu, E.T. Birzin, W. Chan, Y.T. Yang, L.Y. Pai, C. DaSilva, E.C. Hayes, R.T. Mosley, F. DiNinno, S.P. Rohrer, J.M. Schaeffer, M.L. Hammond, Bioorg. Med. Chem. Lett. 15 (2005) 715–718.
[50] C.P. Miller, M.D. Collini, H.A. Harris, Constrained phytoestrogens and analogues as ERβ selective ligands, Bioorg. Med. Chem. Lett. 13 (2003) 2399–2403.
[51] D.L. Parker, D. Meng, R.W. Ratcliffe, R.R. Wilkening, D.M. Sperbeck, M.L. Greenlee, L.F. Colwell, S. Lambert, E.T. Birzin, K. Frisch, S.P. Rohrer, S. Nilsson, A.-G. Thorsell, M.L. Hammond, Triazolo-tetrahydrofluorenones as selective estrogen receptor beta agonists, Bioorg. Med. Chem. Lett. 16 (2006) 4652–4656.
[52] Q. Tan, T.A. Blizzard, J.D. Morgan, E.T. Birzin, W. Chan, Y.T. Yang, L.-Y. Pai, E.C. Hayes, C.A. DaSilva, S. Warrier, J. Yudkovitz, H.A. Wilkinson, N. Sharma, P.M.D. Fitzgerald, S. Li, L. Colwell, J.E. Fisher, S. Adamski, A.A. Reszka, D. Kimmel, F. DiNinno, S.P. Rohrer, L.P. Freedman, J.M. Schaeffer, M.L. Hammond, Estrogen receptor ligands. Part 10: Chromanes: old scaffolds for new SERMs, Bioorg. Med. Chem. Lett. 15 (2005) 1675–1681.
[53] Q. Tan, E.T. Birzin, W. Chan, Y.T. Yang, L.Y. Pai, E.C. Hayes, C.A. DaSilva, F. DiNinno, S.P. Rohrer, J.M. Schaeffer, M.L. Hammond, Estrogen receptor ligands. Part 5: The SAR of dihydrobenzoxathiins containing modified basic side chains, Bioorg. Med. Chem. Lett. 14 (2004) 3747–3751.
[54] K.J. Wildonger, R.W. Ratcliffe, R.T. Mosley, M.L. Hammond, E.T. Birzin, S.P. Rohrer, Tetrahydrofluorenones with conformationally restricted side chains as selective estrogen receptor beta ligands, Bioorg. Med. Chem. Lett. 16 (2006) 4462–4466.
[55] R.R. Wilkening, R.W. Ratcliffe, A.K. Fried, D. Meng, W. Sun, L. Colwell, S. Lam- bert, M. Greenlee, S. Nilsson, A. Thorsell, M. Mojena, C. Tudela, K. Frisch, W. Chan, E.T. Birzin, S.P. Rohrer, M.L. Hammond, Estrogen receptor b-subtype selective tetrahydrofluorenones: use of a fused pyrazole as a phenol bioisostere, Bioorg. Med. Chem. Lett. 16 (2006) 3896–3901.
[56] CATALYST version 4.11 Users' Manual, Accelrys Software Inc., San Diego, 2005.
[57] K. Poptodorov, T. Luu, R. Hoffmann, in: T. Langer, R.D. Hoffmann (Eds.), Pharmacophore Model Generation Software Tools. Methods and principles in Medicinal Chemistry, Pharmacophores and Pharmacophores Searches, Wiley–VCH, Weinheim, 2006, pp. 17–47.
[58] Y. Kurogi, O. Güner, Pharmacophore modeling and three dimensional database searching for drug design using catalyst, Curr. Med. Chem. 8 (2001) 1035–1055.
[59] H. Li, J. Sutter, R. Hoffmann, in: O.F. Güner (Ed.), HypoGen: An Automated System for Generating 3D Predictive Pharmacophore Models. Pharmacophore Perception, Development and Use in Drug Design, International University Line, CA, 2000 pp. 173–189.
[60] J. Sutter, O. Güner, R. Hoffman, H. Li, M. Waldman, in: O.F. Güner (Ed.), Effect of Variable Weights and Tolerances on Predictive model Generation. Pharmacophore Perception, Development and Use in Drug Design, International University Line, CA, 2000, pp. 501–511.
[61] I.B. Bersuker, S. Bahçeci, J.E. Boggs, in: O.F. Güner (Ed.), The Electron-Conformational Method of Identification of Pharmacophore and Anti-Pharmacophore Shielding. Pharmacophore Perception, Development and Use in Drug Design, International University Line, CA, 2000, pp. 457–473.
[62] P.W. Sprague, R. Hoffmann, in: H. Waterbeemd, B. Testa, G. Folkers (Eds.), CATALYST Pharmacophore models and their Utility as Queries for Searching 3D Databases. Computer Assisted Lead Finding and Optimization-Current Tools for Medicinal Chemistry, VHCA, Basel, 1997, pp. 230–240.
[63] R. Fisher, The Principle of Experimentation Illustrated by a Psycho-Physical Experiments, Hafner Publishing Co., New York, 1966.
[64] E.M. Krovat, T. Langer, Nonpeptide angiotensin II receptor antagonists: chemical feature based pharmacophore identification, J. Med. Chem. 46 (2003) 716–726.
[65] CERIUS2 version 4.10 QSAR Users' Manual, Accelrys Inc., San Diego, CA, 2005.
[66] M.L. Verdonk, L. Marcel, V. Berdini, M.J. Hartshorn, W.T.M. Mooij, C.W. Murray, R.D. Taylor, P. Watson, Virtual screening using protein–ligand docking: avoiding artificial enrichment, J. Chem. Inf. Comput. Sci. 44 (2004) 793–806.
[67] J. Kirchmair, P. Markt, S. Distinto, G. Wolber, T. Langer, Evaluation of the performance of 3D virtual screening protocols: RMSD comparisons, enrichment assessments, and decoy selection—what can we learn from earlier mistakes? J. Comput. Aided. Mol. Des. 22 (2008) 213–228.
[68] J.J. Irwin, B.K. Shoichet, ZINC—a free database of commercially available compounds for virtual screening, J. Chem. Inf. Comput. Sci. 45 (2005) 177–182.
[69] N. Triballeau, F. Acher, I. Brabet, J.-P. Pin, H.-O. Bertrand, Virtual screening workflow development guided by the "Receiver Operating Characteristic" curve approach application to high-throughput docking on metabotropic glutamate receptor subtype 4, J. Med. Chem. 48 (2005) 2534–2547.
[70] M. Jacobsson, P. Liden, E. Stjernschantz, H. Bostroem, U. Norinder, Improving structure-based virtual screening by multivariate analysis of scoring data, J. Med. Chem. 46 (2003) 5781–5789.
[71] H. Gao, C. Williams, P. Labute, J. Bajorath, Binary quantitative structure–activity relationship (QSAR) analysis of estrogen receptor ligands, J. Chem. Inf. Comput. Sci. 39 (1999) 164–168.
[72] J. Gasteiger, M.A. Marsili, new model for calculating atomic charges in molecules, Tetrahedron Lett. 34 (1978) 3181–3184.
[73] CERIUS2 LigandFit User Manual, Accelrys Inc., San Diego, CA, 2000, pp. 3–48.
[74] C.M. Venkatachalam, X. Jiang, T. Oldfield, M. Waldman, LigandFit: a novel method for the shape-directed rapid docking of ligands to protein active sites, J. Mol. Graph. Modell. 21 (2003) 289–307.
[75] D.K. Gehlhaar, G.M. Verkhivker, P.A. Rejto, C.J. Sherman, D.B. Fogel, L.J. Fogel, S.T. Freer, Molecular recognition of the inhibitor AG-1343 by HIV-1 protease: conformationally flexible docking by evolutionary programming, Chem. Biol. 2 (1995) 317–324.
[76] C.A. Lipinski, F. Lombardo, B.W. Dominy, P.J. Feeney, Experimental and computational approaches to estimate solubility and permeability in drug discovery and development settings, Adv. Drug Del. Rev. 46 (2001) 3–26.
[77] D.F. Veber, S.R. Johnson, H.Y. Cheng, B.R. Smith, K.W. Ward, K.D. Kopple, Molecular properties that influence the oral bioavailability of drug candidates, J. Med. Chem. 45 (2002) 2615–2623.
[78] R.P. Sheridan, S.K. Kearsley, Why do we need so many chemical similarity search methods, Drug Discov. Today 7 (2002) 903–910.
[79] L.F. Ramsey, W.D. Schafer, The Statistical Sleuth, Wadesworth Publishing Co., USA, 1997.
[80] M.O. Taha, Y. Bustanji, M.A. Al-Ghussein, M. Mohammad, H. Zalloum, I. Al-Masri, N. Atallah, Pharmacophore modeling, quantitative structure–activity relationship analysis, and in silico screening reveal potent glycogen synthase kinase-3β inhibitory activities for cimetidine, hydroxychloroquine, and gemifloxacin, J. Med. Chem. 51 (2008) 2062–2077.
[81] M.O. Taha, N. Atallah, A.G. Al-Bakri, C. Paradis-Bleau, H. Zalloum, Kh.S. Younis, R.C. Levesque, Discovery of new MurF inhibitors via pharmacophore modeling and QSAR analysis followed by in-silico screening, Bioorg. Med. Chem. 16 (2008) 1218–1235.
[82] M.O. Taha, A.M. Qandil, D.D. Zaki, M.A. AlDamen, Ligand-based assessment of factor Xa binding site flexibility via elaborate pharmacophore exploration and genetic algorithm-based QSAR modeling, Eur. J. Med. Chem. 40 (2005) 701–727.
[83] M.O. Taha, L.A. Dahabiyeh, Y. Bustanji, H. Zalloum, S. Saleh, Combining ligand-based pharmacophore modeling, quantitative structure–activity relationship analysis and in silico screening for the discovery of new potent hormone sensitive lipase inhibitors, J. Med. Chem. 51 (2008) 6478–6494.
[84] M.O. Taha, Y. Bustanji, A.G. Al-Bakri, M. Yousef, W.A. Zalloum, I.M. Al-Masri, N. Atallah, Discovery of new potent human protein tyrosine phosphatase inhibitors via pharmacophore and QSAR analysis followed by in silico screening, J. Mol. Graph. Modell. 25 (2007) 870–884.
[85] T. Tamura, S. Masaki, K. Ohmori, A. Karasawa, Effect of olopatadine and other histamine H1 receptor antagonists on the skin inflammation induced by repeated topical application of oxazolone in mice, Pharmacology 75 (2005) 45–52.
[86] T.A. Blizzard, F. DiNinno, J. Morgan, D. Chen, H.Y. Wu, J.Y. Kim, S. Chan, W. Birzin, E.T. Yang, Y.T. Pai, L.-Y. Fitzgerald, P.M.D. Sharma, N. Li, Y. Zhang, Z. Hayes, E.C.A. DaSilva, W. Tang, S.P. Rohrer, J.M. Schaeffer, M.L. Hammond, Estrogen receptor ligands. Part 9: Dihydrobenzoxathiin SERAMs with alkyl substituted pyrrolidine side chains and linkers, Bioorg. Med. Chem Lett. 15 (2005) 107–113.
[87] D. Muthas, Y.A. Sabnis, M. Lundborg, A. Karlen, Is it possible to increase hit rates in structure-based virtual screening by pharmacophore filtering? An investigation of the advantages and pitfalls of post-filtering, J. Mol. Graph. Modell. 26 (2008) 1237–1251.
[88] K.W. Nettles, J.B. Bruning, G. Gil, E.E. O'Neill, J. Nowak, A. Hughs, Y. Kim, E.R. DeSombre, R. Dilis, R.N. Hanson, A. Joachimiak, G.L. Greene, Structural plasticity in the oestrogen receptor ligand-binding domain, EMBO Reports 8 (2007) 563–568.
[89] L.A. Abou-Zeid, A.M. El-Mowafy, Differential recognition of resveratrol isomers by the human estrogen receptor-α: molecular dynamics evidence for stereoselective ligand binding, Chirality 16 (2004) 190–195.
[90] S. Bradbury, V. Kamenska, P. Schmieder, G. Ankley, O. Mekenyan, A computationally based identification algorithm for estrogen receptor ligands: Part 1. Predicting hERα binding affinity, Toxicol. Sci. 58 (2000) 253–269.
[91] L. Xing, W.J. Welsh, W. Tong, R. Perkins, D.M. Sheehan, Comparison of estrogen receptor α and β subtypes based on comparative molecular field analysis (CoMFA), SAR QSAR Environ. Res. 10 (1999) 215–237.
[92] S. Agatonovic-Kustrin, J.V. Turner, Molecular structural characteristics of estrogen receptor modulators as determinants of estrogen receptor selectivity, Mini Rev. Med. Chem. 8 (2008) 943–951.