Structural Chemistry (2023) 34:1731–1742
https://doi.org/10.1007/s11224-022-02086-9

ORIGINAL RESEARCH

Optimization, spectral characterization, QSAR, and molecular docking analyses of newly designed boron compounds

Esra Çetiner[1] · Koray Sayın[1] · Yener Ünal[2]

Received: 27 April 2022 / Accepted: 18 October 2022 / Published online: 23 November 2022
© The Author(s), under exclusive licence to Springer Science+Business Media, LLC, part of Springer Nature 2022

Abstract
In silico analyses of new designed boron compounds were done in detail. In this study, a total of 110 compounds were investigated and optimized at B3LYP-D3/6-31G(d) level in the water. There are two compound groups in this study which are the SCUD and D groups. While the SCUD group contains newly designed boron compounds, the D group contains synthesized compounds by the third person. Spectral characterizations of the whole compounds were performed using IR and NMR spectrum. A total of eight QSAR models were derived using D-group compounds. The biological activity of boron compounds in the SCUD group was calculated, and inhibitor candidates from boron compounds were determined. Molecular docking of selected nineteen compounds was performed against the target protein. Finally, three compounds which are SCUD 28, SCUD 52, and SCUD 65 can be inhibitor candidates. They exhibit better results than that of tamoxifen which is using clinical treatment.

Keywords Boron compounds · Molecular docking · QSAR · Selective estrogen receptor modulator · DFT

Introduction

Carbon chemistry has been studied intensively for the last 200 years. Although it is the neighbor of carbon in the periodic table, the chemistry of boron has attracted attention in recent years and its properties in the field of medicine have begun to be studied. Boron should not be more than 18 mg in the human body [1], but considering its biological properties and potential, it has been used in pharmacological drug design [2]. Especially, it has also been reported that boron compounds show at least two times better biological activity than their carbonaceous derivatives [3]. Boron-containing bioactive molecules can be studied in two parts. These are molecules containing only one boron atom and boron clusters composed of boron atoms which are represented in Fig. 1.

Boron compounds are used in adhesives, paints, soaps, detergents, fiber optics, flame retardants, fuel additives, glass, and many other fields [4]. Its application in medicine started in the 1960s with boron neutron capture therapy in cancer treatment and gained momentum. In recent years, many properties of boron compounds in the field of health have been examined, and it has been emphasized that these compounds have anti-cancer, anti-HIV, and anti-rheumatoid arthritis activities, as well as drug carrier properties, and are effective in diagnosing cancer. In this study, the activity of new 1-(diphenylboranyl)piperidine and triphenylboranamine derivatives against the estrogen receptor will be examined. Quantitative structure–activity relationship (QSAR) models are regression technique which is used in many research field. Like other regression models, QSAR relates a set of "predictor" variables (X) to the potency of the response variable (Y), while QSAR models relate the predictor variables to a categorical value of the response variable. QSAR models are mainly used in the predicting biological activity of the newly designed compound and chemicals. These types of analysis can be used not only for the prediction of biological activity but also for toxicity and chemical properties.

In 2004, the QECD member countries adopted five principles for the validation of QSAR models for regulatory aims. With respect to these principles, QSAR models should be associated with the following items: (1) a defined endpoint, (2) an unambiguous algorithm, (3) a defined domain

✉ Koray Sayın
krysayin@gmail.com

1 Department of Chemistry, Faculty of Science, Sivas Cumhuriyet University, Sivas, TÜRKİYE

2 Department of Statistics and Computer Science, Faculty of Science, Sivas Cumhuriyet University, Sivas, TÜRKİYE

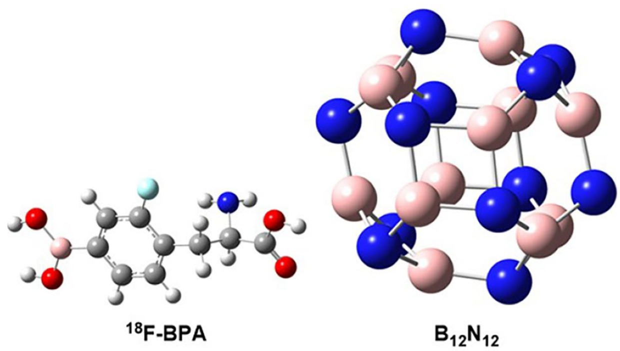

Fig. 1 Molecular structure of some boron compounds

of applicability (AD), (4) appropriate measures of goodness-of-fit, robrobustnessd predictivity, and (5) a mechanistic interpretation. Especially, the third item expresses the need to define and AD for QSAR models. Because QSAR models are reductionist, and they could generate reliable predictions using different quantum chemical descriptors which are inevitably associated with limitations. Actually, AD is organized to help to understand to express the scope and limitations of models. But, the AD concept can be implicit in the published articles. For instance, the model has been developed from a training set of chemicals that belong to a single chemical class or that are considered to share a common mechanism of action. In other cases, the AD concept has been explicitly defined [5]. There are a lot of approaches to define AD of QSAR models such as according to structural rules, range of descriptor variable, continuous descriptor variable, the application of multiple linear regression, tolerance volume, and decision tree analysis [5–7].

In this article, a total of eighty boron compounds (SCUD Group) were designed and given in Supplemental Material. Additionally, thirty similar boron compounds (D group) are taken into consideration for QSAR analyses and are given in Supplemental Material. These thirty boron compounds have been synthesized by Das and coworkers in 2015 [8]. All these compounds are fully optimized at B3LYP-D3/6-31G(d) level in the water. A conductor-like-polarized continuous pattern (C-PCM) solvation model is used to consider solute–solvent interactions. The structure and spectral analyses (IR and NMR) of designed boron compounds are done in detail. Then, the electronic properties of these compounds are examined by contour plots of frontier molecular orbitals and molecular electrostatic potential (MEP) maps. A quantitative structure–activity relationship (QSAR) analysis is performed using thirty compounds. These compounds are divided into two parts: test (5 compounds) and analysis groups (25 compounds). A total of 246 quantum chemical descriptors are calculated using Maestro software. The list of quantum chemical descriptors is given in the Supplemental Material. Studied compounds are eliminated using the determined eight QSAR models. Docking analyses of inhibitor candidates with better results than tamoxifen are performed. In this analysis, a total of 20 proteins, ERα and ERβ, are used. These proteins were selected from the protein data bank and GeneCards as 1ERE [9], 1PCG [10], 3ERT [11], 4Q50 [12], 5U2B [13], 5UFW [14], 6C42 [14], 6DF6 [15], 6VJD, 6VIG, 1HJ1 [16], 1U3Q [17], 1X7J [18], 2I0G [19], 2NV7 [20], 2YLY [21], 2Z4B [22], 2QTU [23], 3OLL [24], and 5TOA [25]. As a result, a molecule that could show a better effect than tamoxifen was determined.

Materials and methods

Optimization

Fully optimization calculations were performed using Gaussian software [26, 27]. Initially, the whole compounds in this study were pre-optimized in the universal force field (UFF) molecular mechanic method in order not to waste time and not to encounter errors in future optimization calculations. In subsequent optimization calculations, the Becke-3-parameter-Lee–Yang–Parr (B3LYP) hybrid functional was used as a calculation method with the D3 version of Grimme's dispersion. 6-31G(d) was selected as the basis set and the C-PCM method was used to consider solute–solvent interaction. All calculations were done in the water phase. Furthermore, ChemDraw software was used as utilities throughout the study [28].

Spectral analysis

Infrared (IR) and nuclear magnetic resonance (NMR) spectrum are calculated at the same level of theory. In the analysis of the IR spectrum, the VEDA 4XX program was used [29]. In the NMR spectrum, chemical shift values of carbon and hydrogen atoms are calculated using Eq. (1). In this stage, tetramethylsilane (TMS) was calculated at the same level of theory.

$$\delta = \delta_{TMS} - \delta_{Compound} \tag{1}$$

QSAR analysis

The used quantum chemical descriptors of D and SCUD group compounds were calculated using the Maestro program. Initially, the relationship between experimental IC_{50} and quantum chemical descriptors was investigated in detail. In derivating of QSAR model, the regression method was used. The five parameters with the highest correlation with the IC_{50} variable were included in the model. The IC_{50} variable was

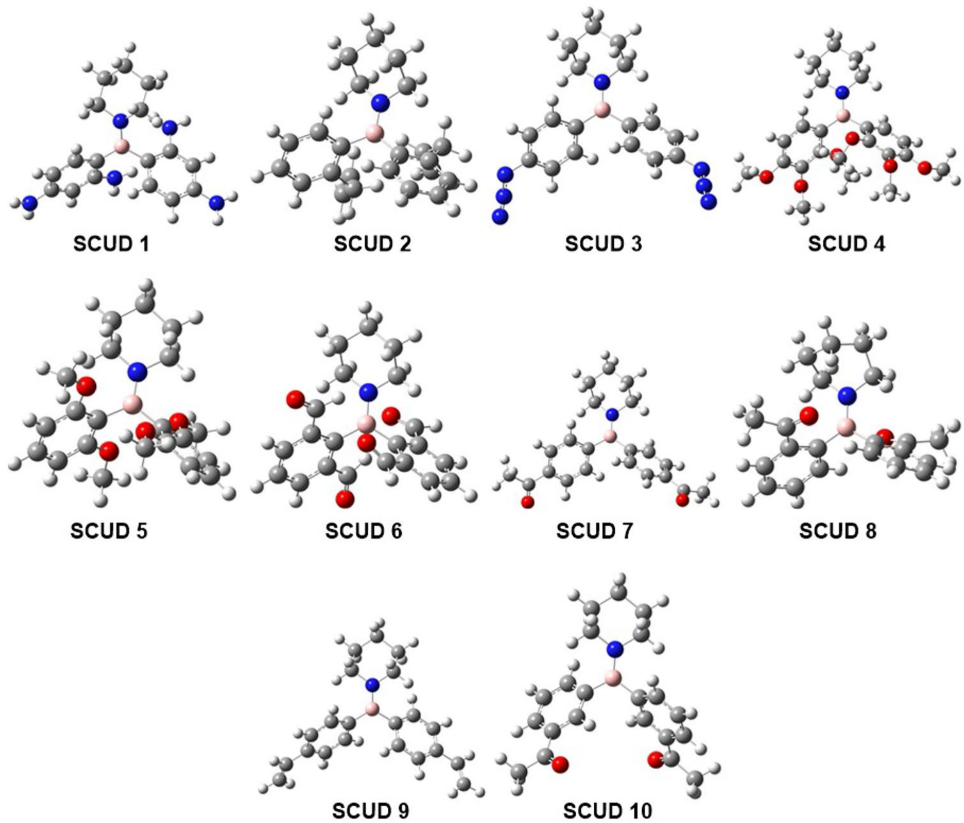

Fig. 2 The optimized structure of SCUD 1–10

taken as the dependent variable, and a multiple regression model was created with other variables. The significance of the obtained models was examined. In addition to this, the R square value of the independent variables' explanation ratio of the dependent variable was calculated. Estimated IC_{50} values were calculated from the obtained regression model. The correlation coefficient between the estimates obtained with the IC_{50} values of the control group was calculated. It was determined that there is a very high correlation between the actual IC_{50} values and the predicted values.

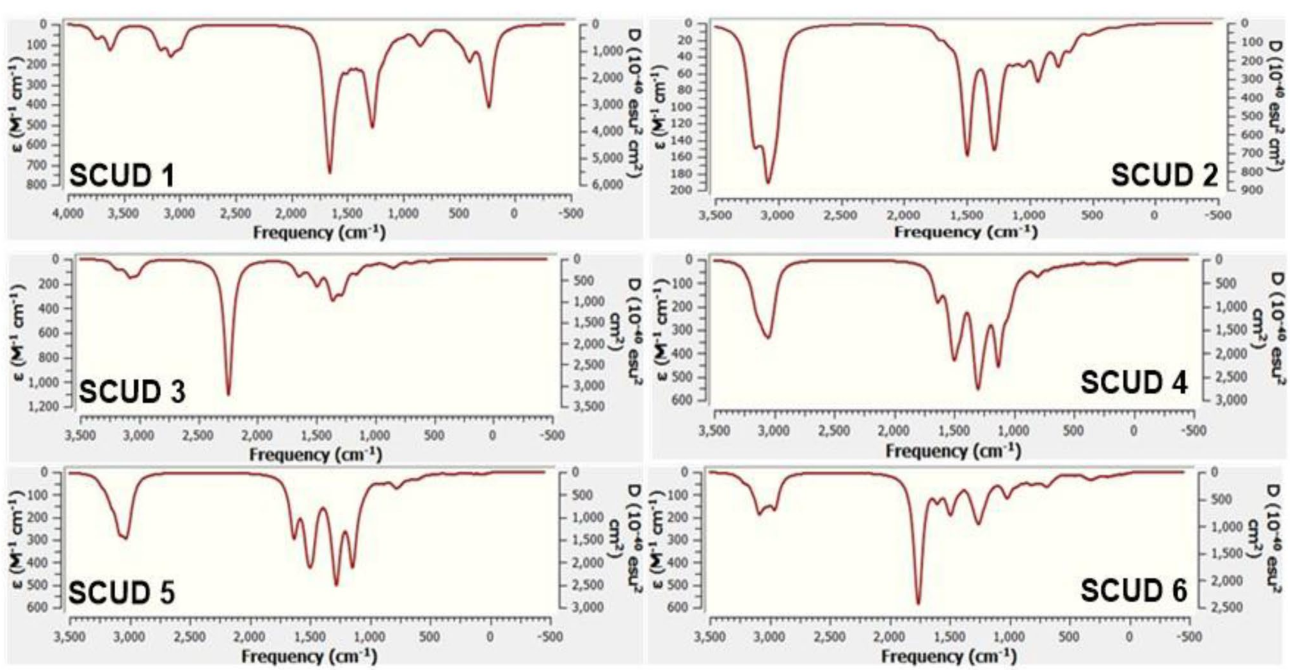

Fig. 3 IR spectrum of SCUD 1–6

Table 1 Calculated vibrational frequencies (cm^{-1}) of SCUD 1–3

SCUD 1		SCUD 2		SCUD 3	
Frequency[a]	Mode[b]	Frequency[a]	Mode[b]	Frequency[a]	Mode[b]
3756	STRE(NH)	3178	STRE(CH)	3183	STRE(CH)
3624	STRE(NH)	3078	STRE(CH)	3078	STRE(CH)
3170	STRE(CH)	1492	STRE(NB), BEND(HCN)	2244	STRE(NN)
3080	STRE(CH)	1285	STRE(NB), BEND(HCN), BEND(HCC)	1651	STRE(CC)
1657	STRE(CC)	1052	STRE(NC), STRE(CC)	1493	STRE(NB), BEND(HCH)
1274	STRE(NB), BEND(HCC), BEND(HCN)	933	TORS(HCCC)	1365	STRE(NN), STRE(NC)
849	TORS(HCCN), OUT(NCCC)	771	TORS(HCCC)	1287	STRE(NB), BEND(HCC), BEND(HCN)
		677	STRE(CC)	1160	STRE(NN), STRE(NC)

[a]in cm^{-1}

[b]*STRE* stretching, *BEND* bending, *TORS* torsion, *OUT* out of planar

Molecular docking

Selected compounds are prepared for docking calculation using the LigPrep module in Maestro software. The acidity of calculations is selected as 7 ± 2. Then target proteins which are 1ERE, 1PCG, 3ERT, 4Q50, 5U2B, 5UFW, 6C42, 6DF6, 6VJD, 6VIG, 1HJ1, 1U3Q, 1X7J, 2I0G, 2NV7, 2YLY, 2Z4B, 2QTU, 3OLL, and 5TOA were prepared using protein preperation module. The receptor-binding domain of them is defined using Grid Generation. Then molecular docking calculations were performed [30–33]. In these calculations, four parameters which are docking score (DS), van der Walls energy (E_{vdW}), Coloumb interaction energy (E_{Coul}), and total interaction energy (E_{Total}) were examined and evaluated analyses. The ground state structures of phenyl urea derivatives were obtained from computational calculations.

Results and discussion

Optimized structures

The designed compounds (SCUD Group) and previously synthesized by third parties (D Group) compounds are optimized at B3LYP-D3/6-31G(d) level in the water. Optimized structures of SCUD 1–10 are represented in Fig. 2. Additionally, the optimized structures of SCUD 11–80 and D1–D30 are given in Supplemental Material.

According to optimized structures, SCUD compounds are boron nitrite derivatives. The environment of boron compounds is found as trigonal planar. The geometric parameters on the structure are found as good agreement with the results of published articles [34–36]. As for the D groups, the big difference of these compounds is the carbon–carbon double bond on the structure. It is also known that the B-N structure

Table 2 Calculated vibrational frequencies (cm^{-1}) of SCUD 4–6

SCUD 4		SCUD 5		SCUD 6	
Frequency[a]	Mode[b]	Frequency[a]	Mode[b]	Frequency[a]	Mode[b]
3052	STRE(CH)	3072	STRE(CH)	3087	STRE(CH)
1638	STRE(CC)	3033	STRE(CH)	2960	STRE(CH)
1497	STRE(NB)	1630	STRE(CC)	1762	STRE(CO)
1299	STRE(CB)	1502	BEND(HCH)	1600	STRE(CC)
1129	STRE(OC)	1284	BEND(HCC)	1490	STRE(CC)
801	TORS(HCCC)	1145	STRE(OC)	1260	STRE(CC)
		782	TORS(CCCC), OUT(OCCC)	1019	STRE(CC)
				685	BEND(OCC), BEND(CCC)

[a]in cm^{-1}

[b]*STRE* stretching, *BEND* bending, *TORS* torsion, *OUT* out of planar

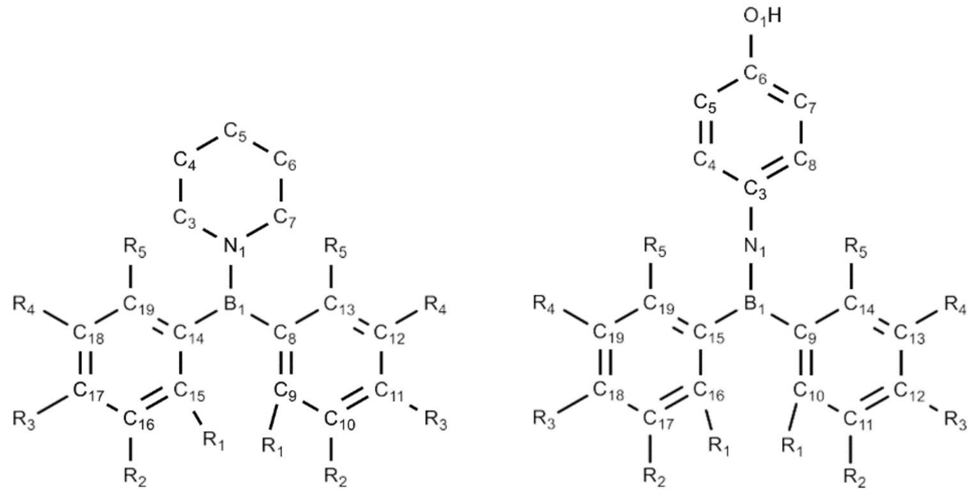

Fig. 4 Atomic labeling of atoms in of SCUD compounds

is an elemental isomer of the C=C structure. At the same time, it has been reported that the biological activities of compounds containing B-N bonds are at least two times more effective than elemental isomers containing C=C bonds [3]. As structurally, it is determined that the geometric parameters of D group compounds were quite compatible with similar structures [37–39].

The IR spectrum of SCUD groups

The IR spectrum is one of the mainly used spectral techniques for the characterization of chemicals. The IR spectrum can be obtained as computationally. In our study, the IR spectrum of eighty SCUD compounds is calculated and analyzed using VEDA 4XX software. IR spectrum of SCUD 1–6 are represented in Fig. 3. Additionally, VEDA analyses of these compounds are given in Tables 1 and 2. IR spectrum and VEDA analyses of other compounds in the SCUD group are given in Supplemental Material.

According to Tables 1 and 2, the vibration modes of labeled frequencies are given. However, calculated frequencies are harmonic while experimental frequency is anharmonic. Therefore, some differences are encountered.

Table 3 Chemical shift values (ppm) of carbon atoms in SCUD 1–5 compounds

Assignment	SCUD 1	SCUD 2	SCUD 3	SCUD 4	SCUD 5
C3	45.7	51.2	50.5	50.0	50.3
C4	23.3	29.1	29.8	29.9	29.3
C5	22.0	25.9	26.3	26.6	26.4
C6	25.4	29.3	29.4	29.3	28.1
C7	47.1	50.8	51.3	50.2	50.3
C8	105.5	138.5	134.1	125.5	112.9
C9	141.5	126.8	128.6	122.6	153.4
C10	85.9	118.1	112.5	100.9	90.9
C11	135.7	121.9	131.9	147.5	122.8
C12	90.5	123.6	112.2	136.4	96.3
C13	127.6	137.1	131.6	149.5	153.4
C14	103.5	134.6	133.9	125.2	113.1
C15	138.7	129.9	131.6	122.9	153.1
C16	84.3	118.3	128.9	101.1	96.0
C17	135.6	122.2	112.4	147.3	122.8
C18	91.5	123.0	112.5	136.6	96.2
C19	127.2	138.2	131.6	149.3	153.5

Table 4 Chemical shift values (ppm) of hydrogen atoms in SCUD 1–5 compounds

Assignment	SCUD 1	SCUD 2	SCUD 3	SCUD 4	SCUD 5
C3*H*	3.0	3.1	3.8	3.8	3.6
C3*H′*	2.4	3.6	3.0	1.6	3.0
C4*H*	0.9	1.5	1.7	1.9	1.5
C4*H′*	1.4	1.9	1.9	1.9	1.5
C5*H*	0.9	1.9	1.9	1.7	1.6
C5*H′*	1.2	1.7	1.7	1.5	1.8
C6*H*	0.9	1.6	1.6	1.5	1.4
C6*H′*	0.9	1.6	1.5	3.8	2.1
C7*H*	3.2	3.5	3.2	3.1	2.9
C7*H′*	2.6	3.3	3.8	-	3.5
C9*H*	-	-	7.0	-	3.5
C10*H*	4.6	7.1	7.0	-	-
C11*H*	-	7.2	-	6.5	7.1
C12*H*	5.0	7.0	6.7	7.2	6.3
C13*H*	6.2	6.9	7.6	7.3	-
C15*H*	-	7.8	7.7	6.5	-
C16*H*	4.8	7.3	6.7	-	6.2
C17*H*	-	7.3	-	-	7.1
C18*H*	4.9	8.1	6.9	-	6.1
C19*H*	5.9	-	7.0		-

Nevertheless, there is good agreement between calculated and published data [40–43].

Simulated NMR spectrum.

An NMR spectrum of designed boron compounds is calculated. Chemical shift values of hydrogen and carbon atoms are calculated using Eq. (1). Additionally, TMS is used as reference material in the calculation of chemical shift values of related atoms. Atomic labeling of studied compounds is represented in Fig. 4. Additionally, chemical shift values of carbon and hydrogen atoms of SCUD 1–5 are given in Tables 3 and 4, respectively. NMR results for other compounds in the SCUD group are given in Supplemental Material.

According to NMR data of studied boron compounds, chemical shift values of aliphatic and aromatic carbon atoms are calculated in the range between 22–50 and 122–140 ppm, respectively. As for the hydrogen atoms, chemical shift values of hydrogen atoms coordinated to oxygen and nitrogen atoms are calculated nearly 4 and 6 ppm, respectively. Additionally, chemical shift values of hydrogen atoms on the benzene ring are calculated in the range of 6.1–8.1 ppm. All calculated chemical shift values are in agreement with published data and article [40–43]. It can be said that spectral characterization of the designed compounds is done in detail.

Electronic properties

The electronic properties of chemicals play an important role on the determination of interaction mechanism, the

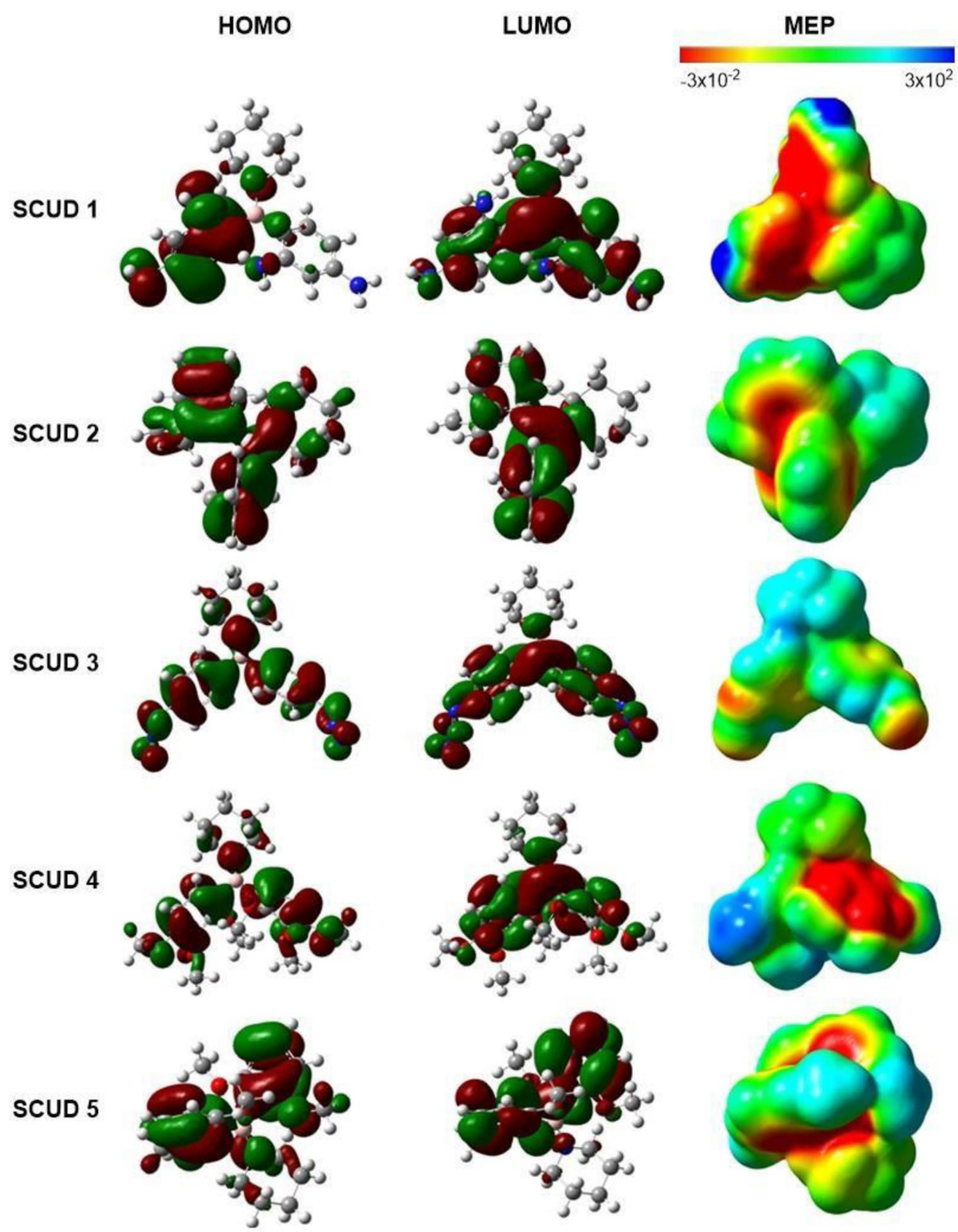

Fig. 5 Contour diagram of frontier molecular orbitals and MEP maps of SCUD 1–5

Table 5 Derived QSAR Models

$IC_{50}{}^{c} = -54{,}418 - 5112\ X_1 – 0.956\ X_2 – 0.044\ X_3 + 19{,}528X_4 – 7384\ X_5$						
Model 1	X1: Balaban distance connectivity index	X2: atoms in ring system	X3: PEOE5	X4: PEOE12	X5: addition electronic charge	
	R^2	MSS	MSS/p	RSS	F	Sig.
	1.000	9527.237	1905.447	0.489	66,184.518	0.0001
	Adj. R^2	Std. error of estimates	PRESS	Q^2		
	0.99	0.16968	0.000184	0.999		
$IC_{50}{}^{c} = 1768{,}911 + 7402\ X_1 – 49{,}888X_2 – 1731{,}403X_3 + 1072X_4 – 7027\ X_5$						
Model 2	X1: polarity	X2: Narumi simple topological	X3: path/walk 2—Randic shape index	X4: PEOE3	X5: HumanOralAbsorption	
	R^2	MSS	MSS/p	RSS	F	Sig.
	0.958	9127.256	1825.451	400.470	77.491	0.0001
	Adj. R^2	Std. Error of estimates	PRESS	Q^2		
	0.946	4.85356	0.50051	0.999		
$IC_{50}{}^{c} = 640{,}191 - 24{,}609X_1 – 41{,}292X_2 – 33{,}152X_3 + 52{,}583X_4 + 252{,}414X_5$						
Model 3	X1: quadratic	X2: mean square distance Balaban	X3: topological charge index of order 4	X4: topological charge index of order 5	X5: topological charge index of order 7	
	R^2	MSS	MSS/p	RSS	F	Sig.
	0.995	9481.682	1896.336	46.045	700.140	0.0001
	Adj. R^2	Std. Error of estimates	PRESS	Q^2		
	0.994	1.64576	48.83518	0.9948		
$IC_{50}{}^{c} = 295{,}831 + 2261X_1 – 33{,}670X_2 + 267{,}442X_3 – 274{,}026X_4 + 0.0040X_5$						
Model 4	X1: molecular electrotopological variation	X2: maximal electrotopological negative variation	X3: second Mohar	X4: reciprocal distance Randic-type index	X5:Bzzz	
	R^2	MSS	MSS/p	RSS	F	Sig.
	0.975	9288.643	1857.729	239.084	132.093	0.0001
	Adj. R^2	Std. error of estimates	PRESS	Q^2		
	0.968	3750	292,974	0.969		
$IC_{50}{}^{c} = 38{,}113 + 210{,}139\ X_1 – 85{,}861X_2 + 2275{,}609X_3 + 1860{,}835X_4 – 0.404X_5$						
Model 5	X1:topological charge index of order 6	X2:connectivity index chi-4	X3:path/walk 4—Randic shape index	X4:path/walk 5—Randic shape index	X5:PercentHumanOralAbsorption	
	R^2	MSS	MSS/p	RSS	F	Sig.
	0.966	9205.123	1841.025	322.664	97.015	0.0001
	Adj. R^2	Std. error of estimates	PRESS	Q^2		
	0.956	4.356	1488.844	0.843		
$IC_{50}{}^{c} = 1357{,}636 + 323{,}093\ X_1 – 116{,}978X_2 + 33{,}345\ X_3 – 47{,}691X_4 – 0.584\ X_5$						
Model 6	X1:topological charge index of order 9	X2:3-path Kier alpha-modified shape index	X3: Kier flexibility	X4: ring perimeter	X5: S	
	R^2	MSS	MSS/p	RSS	F	Sig.
	0.968	9221.937	1844.387	305.790	102.536	0.0001
	Adj. R^2	Std. error of estimates	PRESS	Q^2		
	0.958	4.24118	343.913	0.963		
$IC_{50}{}^{c} = -347{,}528 - 191{,}135X_1 + 9239{,}329\ X_2 – 0.544X_3 + 2542\ X_4 – 0.005X_5$						
Model 7	X1: eccentric	X2: radial centric	X3: PEOE4	X4: dipole Y	X5: Byyy	
	R^2	MSS	MSS/p	RSS	F	Sig.
	0.967	9214.064	1842.813	313.662	98.878	0.0001
	Adj. R^2	Std. error of estimates	PRESS	Q^2		
	0.957			−0.569		

Table 5 (continued)

$IC_{50}{}^{c} = 12{,}303 - 84{,}392X_1 + 203{,}507\ X_2 + 1629{,}611X_3 – 69{,}589X_4 + 1931{,}620X_5$						
Model 8	X1: Topological charge index of order 2	X2: Topological charge index of order 5	X3: Mean topological charge index of order 4	X4: Connectivity index chi-4	X5: path/walk 3—Randic shape index	
	$_R2$	MSS	MSS/p	RSS	F	Sig.
	0.969	9233.398	1846.680	294.329	106.661	0.0001
	Adj. R^2	Std. error of estimates	PRESS	Q^2		
	0.960	4.16095	1963.793	0.794		

active site of compounds, and the molecular effectiveness of compound surface, etc. For these aims, different plots of maps can be used and contour diagram of frontier molecular orbitals and molecular electrostatic potential (MEP) maps are calculated for each boron compounds. While the contour diagram of frontier molecular orbitals and MEP maps of SCUD 1–5 are represented in Fig. 5, the results for other compounds are represented in Supplemental Materials.

According to Fig. 5, HOMO electrons are delocalized on the benzene rings of the studied compound. In the contour plot of LUMO, electrons are mainly delocalized on the benzene rings of the compounds, too. Especially, it can be easily seen that π electrons play an essential role in having this feature. While the contour plot of frontier molecular orbitals shows special zones that can be active, MEP maps show the reactive zones on the molecular surface. The reactivity of π electrons is seen easily from MEP maps of related compounds, too.

Quantitative structure–activity relationship (QSAR) analyses

In the event of a change in the structure of any series of molecules, biological activity also creates positive or negative changes. Accordingly, a systematic cause-effect relationship is called a structure–activity relationship (SAR). The main purpose of SAR is to determine the consequences of changes in the structure, and then, considering these results, to determine which changes in the chemical structure and properties will provide better biological activity. Using this definition, the biological activities (or properties, reactivity) of new or untested chemicals can be determined by QSAR models, based on the chemical structures of similar compounds with known biological activities in the studied molecule series. In this study, group D compounds were used only for QSAR analysis, and D1–D25 compounds were determined as the analysis group, while D26–D30 compounds were determined as the test group. Molecular descriptors are used in QSAR analysis, and these descriptors vary as structural, topological, electrostatic, geometric, and quantum chemical descriptors. Structural descriptors are known as parameters that give simple definitions about the molecule, and these descriptors are the number of heteroatoms in the molecule. Topological descriptors are parameters that provide information about the binding order in a molecule. Examples of these descriptors are the Weiner index and the Randic index. Another type of descriptor that can be used in QSAR is electrostatic descriptors and gives us information about the molecular charge distribution. Geometric descriptors are one of the descriptor groups that can be used in QSAR analysis and provide us with information about the size and shape of the molecule. The last set of descriptors that can be used is quantum chemical descriptors, which are parameters related to the electronic structure of the molecule. Regression analyses are done between experimental IC_{50} values and calculated parameters. A total of eight QSAR models are derived and given in Table 5.

For both simple and multiple linear regression analyses, a number of measures of statistical fit are commonly applied. Some of the statistical comparison criterion is R^2 and R^2 adj. The fact that these values are close to 1.00 is considered as a measure of the relationship between the mathematical model and the independent input variables. The difference between R^2 and R^2 adj values is obtained by re-calculating the possible meaningless factors in the model.

$$R^2 = \frac{MSS}{TSS} = 1 - \frac{RSS}{TSS}$$

MSS: model sum of squares $MSS = \sum_i (\hat{y}_i - \overline{y})^2$

RSS is the sum of the squares (residual) $RSS = \sum_i (y_i - \hat{y}_i)^2$

TSS: the total sum of the squares $TSS = \sum_i (y_i - \overline{y})^2$

The standard error of estimate measures the dispersion of the observed values from the regression line. The smaller the value of s the higher the reliability of the prediction. However, it is not recommended to have a standard error of estimate smaller than the experimental error of the biological data, as this indicates an overfitted model. The cross-validated explained variance or cross-validated correlation coefficient Q2 is used as a measure of the goodness of internal power to predict. It is calculated by the formula:

$$Q_2 = 1 - \frac{PRESS}{TSS}$$

where PRESS is the predictive error sum of squares, that is, the sum of the squares of the differences (residuals) between the experimental and predicted responses when predictions are made for objects left out of the training sets [44].

The calculated biological activity ($IC_{50}{}^{c}$) of D26–D30 is given in Table 6. According to this table, the regression coefficient between experimental and calculated biological activity is found to be higher than 0.95. It shows that derived QSAR models are so good for our studied compounds. Despite the well results obtained, the most consistent and good results were obtained with models 1, 2, and 3. When models 1, 2, and 3 were examined in detail, Model 2 was selected for further studies because the results obtained in model 2 were both consistent with the general trend and had a good regression coefficient. In model 2, selected descriptors are polarity is member of the electronic properties while other parameters are related with the topological parameters.

Calculated IC_{50} values of studied boron compounds in group SCUD are calculated and given in the Supplemental Material. Additionally, the biological activity of tamoxifen is calculated using model 2. It is desirable that the IC50 value be less than one hundred. However, with the derived model, IC_{50} values greater than 100 can be calculated. This does not mean that the engineered compounds are ineffective. Because it can play a role in determining the biological activity orientations of these compounds. In addition, the theoretical IC_{50} value of tamoxifen, which is used clinically, was calculated as 355. In this study, a total of nineteen compounds which their IC_{50} values of less than 355 were selected to perform molecular docking analyses. As a result, SCUD 1, SCUD 11, SCUD 12, SCUD 24, SCUD 28, SCUD 29, SCUD 31, SCUD 40, SCUD 51, SCUD 52, SCUD 57, SCUD 62, SCUD 64, SCUD 65, SCUD 68, SCUD 69, SCUD 71, SCUD 75, SCUD 78, and SCUD 80 are selected for further analyses.

In terms of the AD of the derived QSAR models, it can be easily said that the compounds used in the derivation of the QSAR model are not completely similar to each other, but are close to each other. Furthermore, multilinear regression (MLR) analysis is used in the derivation. On the other hand, a lot of quantum chemical descriptors such as structural, topological, electrostatic, and geometric are scanned to find more harmonic ones. The AD of QSAR models is implicited in the derivation of models [5]. Additionally, the applicability of QSAR models is investigated above. So, it can be said that derived QSAR models give logical results (Table 6), especially model 2 is find as the best.

Molecular docking analysis

There are analysis methods such as molecular structure descriptors, charge densities, QSAR, and molecular

Table 6 Calculated IC_{50} values for D25–D30 compounds using derived QSAR models and regression coefficient (R^2) between experimental and calculated ones

	Compound	IC_{50}	$IC_{50}{}^{c}$	R^2
Model 1	D26	100.00	100.02	
	D27	2.31	2.33	
	D28	1.18	0.81	0.999
	D29	0.02	0.06	
	D30	0.01	0.15	
Model 2	D26	100.00	102.37	
	D27	2.31	8.82	
	D28	1.18	8.67	0.999
	D29	0.02	3.65	
	D30	0.01	2.73	
Model 3	D26	100.00	99.46	
	D27	2.31	3.91	
	D28	1.18	0.52	0.999
	D29	0.02	3.63	
	D30	0.01	1.48	
Model 4	D26	100.00	98.58	
	D27	2.31	9.43	
	D28	1.18	0.37	0.997
	D29	0.02	5.93	
	D30	0.01	3.08	
Model 5	D26	100.00	85.67	
	D27	2.31	4.83	
	D28	1.18	1.97	0.996
	D29	0.02	5.99	
	D30	0.01	9.94	
Model 6	D26	100.00	97.04	
	D27	2.31	5.81	
	D28	1.18	0.31	0.998
	D29	0.02	0.18	
	D30	0.01	5.23	
Model 7	D26	100.00	67.51	
	D27	2.31	15.88	
	D28	1.18	28.37	0.944
	D29	0.02	23.84	
	D30	0.01	8.50	
Model 8	D26	100.00	96.98	
	D27	2.31	9.98	
	D28	1.18	11.20	0.990
	D29	0.02	22.01	
	D30	0.01	11.34	

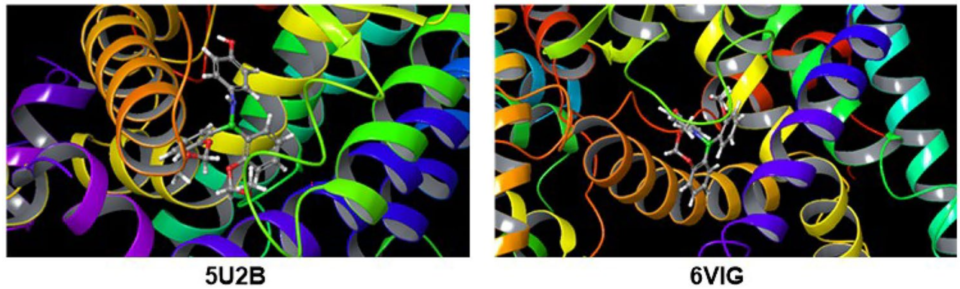

Fig. 6 Molecular docking structure of SCUD 65 with 5U2B and 6VIG

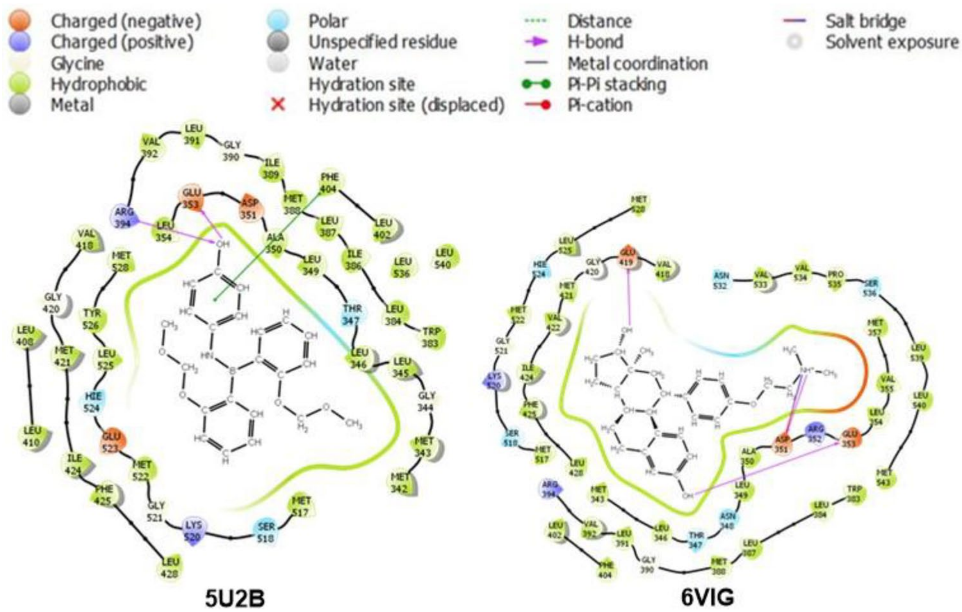

Fig. 7 Interaction maps of SCUD 65 with 5U2B and 6VIG

docking that can be used to predict the biological activities of molecules. With these methods, the many features of molecules can be antibacterial, antifungal, antimalarial, anticancer, etc. can be seen. In this study, molecular docking analyses of boron compounds, which are predicted to be effective in QSAR analysis, against target proteins were performed. A total of 20 proteins are used. Selected proteins are 1ERE, 1PCG, 3ERT, 4Q50, 5U2B, 5UFW, 6C42, 6DF6, 6VJD, and 6VIG belong to ERα, while 1HJ1, 1U3Q, 1X7J, 2I0G, 2NV7, 2YLY, 2Z4B, 2QTU, 3OLL, and 5TOA belong to ERβ protein. The structures of these proteins are shown in the Supplemental Material. Additionally, the x–y-z coordinates of the receptor binding region of the proteins are given in Supplemental Material. According to docking results, some inhibitor candidates are interacted with the target protein and some of them are not interacted. Molecules interacting with ERα and ERβ proteins are given in the table given in Supplemental Material, with a "+" sign and a "−" sign for those that do not.

According to obtained results, studied boron compounds are effective against ERα while they are inactive against ERβ. Additionally, the anticancer properties of selected boron compounds are compared with tamoxifen's results. It is seen that only SCUD 65 exhibits a better effect than that of tamoxifen. Furthermore, SCUD 28 and SCUD 52 exhibit similar effect with tamoxifen. Calculated docking score, van der Walls interaction energy, Coulomb interaction energy, and total interaction energy for selected boron compounds are given in Supplemental Material. The docking structure and interaction map of SCUD 65 with 5U2B and 6VIG are represented in Figs. 6 and 7, respectively.

Conclusion

In this study, compounds in SCUD and D groups are optimized in the water phase at the B3LYP-D3/6-31G(d) level. IR spectra of boron compounds were calculated and PED analyzes were performed with the VEDA program. NMR spectra of boron compounds were calculated and chemical shift values of carbon and hydrogen atoms in the compounds were calculated. QSAR analysis was performed using compounds in group D. Eight models were derived using D1-D25 compounds and the reliability of the derived models was investigated using the test group. Model 2 was judged to be the best. Theoretical IC_{50} values of the compounds in the SCUD group and tamoxifen were calculated using model 2. Compounds that were better than tamoxifen were decided. Molecular docking analyzes were performed between the molecules and target proteins. It is seen that only SCUD 65 exhibits a better effect than that of tamoxifen. Furthermore, SCUD 28 and SCUD 52 exhibit similar effect with tamoxifen. SCUD 65 can be a good inhibitor candidate for estrogen receptors.

Supplementary Information The online version contains supplementary material available at https://doi.org/10.1007/s11224-022-02086-9.

Acknowledgements This research was made possible by TUBITAK ULAKBIM, High Performance, and the Grid Computing Center (TR-Grid e-Infrastructure).

Author contribution EÇ performed the drug design and computational analyses. KS designed the analyses and consistent guidance; analyzed the data, manuscript preparation, and review; edited the final version; and submitted it for publication. YU performed statictical analyses.

Funding This work is supported by the Scientifc Research Project Fund of Sivas Cumhuriyet University under the project numbers RGD-020.

Availability of data and material All experimental data were included in the article.

Declarations

Conflict of interest/Competing interests The authors declare no competing interests.

References

1. Murray FJ (1995) A human health risk assessment of boron (boric acid and borax) in drinking water. Regul Toxicol Pharmacol 22(3):221–230
2. Fayaz A, Hosmane NS, Zhu Y (2020) Boron chemistry for medical applications. Molecules 25(4):828
3. Scorei IR (2011) Calcium fructoborate: plant-based dietary boron as potential medicine for cancer therapy. Front Biosci S 3:205
4. Kar Y, Şen N, Demirbaş A (2006) Boron minerals in Turkey, boron minerals in Turkey, their application areas and importance for the country's economy. Minerals Energ 3(4):2–10
5. Netzeva TI, Worth AP, Aldenberg T, Benigni R, Cronin MTD, Gramatica P, Jaworska JS, Kahn S, Klopman G, Marchant CA, Myatt G, Nikolova-Jeliazkova N, Patlewicz GY, Perkins R, Roberts DW, Schultz TW, Stanton DT, van de Sandt JJM, Tong W, Veith G, Yang C (2005) Current status of methods for defining the applicability domain of (quantitative) structure–activity relationships. ATLA 33:155–173
6. Roy K, Kar S, Ambure P (2015) On a simple approach for determining applicability domain of QSAR models. Chemom Intell Lab Syst 145:22–29
7. Gajewicz A (2018) How to judge whether QSAR/read-across predictions can be trusted: a novel approach for establishing a model's applicability domain. Environ Sci Nano 5:408–421
8. Das BC, Thapa P, Karki R, Schinke C, Das S, Kambhampati S, Lai A, Kahraman M, Govek S, Nagasawa J, Bonnefous C, Julien J, Douglas K, Sensintaffar J, Lu N, Lee K-J, Aparicio A, Kaufman J, Qian J, Shao G, Prudente R, Moon MJ, Joseph JD, Darimont B, Smith ND (2015) Identification of GDC-0810 (ARN-810), an orally bioavailable selective estrogen receptor degrader (SERD) that demonstrates robust activity in tamoxifen-resistant breast cancer xenografts. J Med Chem 58:4888
9. Brzozowski AM, Pike AC, Dauter Z, Hubbard RE, Bonn T, Engstrom O, Ohman L, Greene GL, Gustafsson JA, Carlquist M (1997) Molecular basis of agonism and antagonism in the oestrogen receptor. Nature 389:753–758
10. Leduc AM, Trent JO, Wittliff JL, Bramlett KS, Briggs SL, Chirgadze NY, Wang Y, Burris TP, Spatola AF (2003) Helix-stabilized cyclic peptides as selective inhibitors of steroid receptor-coactivator interactions. Proc Natl Acad Sci USA 100:11273–11278
11. Shiau AK, Barstad D, Loria PM, Cheng L, Kushner PJ, Agard DA, Greene GL (1998) Human estrogen receptor alpha ligand-binding domain in complex with 4-hydroxytamoxifen. Cell 95:927–937
12. Fanning SW, Mayne CG, Dharmarajan V, Carlson KE, Martin TA, Novick SJ, Toy W, Green B, Panchamukhi S, Katzenellenbogen BS, Tajkhorshid E, Griffin PR, Shen Y, Chandarlapaty S, Katzenellenbogen JA, Greene GL (2016) Estrogen receptor alpha somatic mutations Y537S and D538G confer breast cancer endocrine resistance by stabilizing the activating function-2 binding conformation. Elife 5:e12792
13. Stender JD, Nwachukwu JC, Kastrati I, Kim Y, Strid T, Yakir M, Srinivasan S, Nowak J, Izard T, Rangarajan ES, Carlson KE, Katzenellenbogen JA, Yao XQ, Grant BJ, Leong HS, Lin CY, Frasor J, Nettles KW, Glass CK (2017) Structural and molecular mechanisms of cytokine-mediated endocrine resistance in human breast cancer cells. Mol Cell 65:1122–1135
14. Fanning SW, Hodges-Gallagher L, Myles DC, Sun R, Fowler CE, Green BD, Harmon CL, Greene GL, Kushner PJ (2018) Estrogen receptor alpha ligand binding domain in complex with OP1154. Nat Commun 9:2368
15. Zhang B, Kiefer JR, Blake RA, Chang JH, Hartman S, Ingalla ER, Kleinheinz T, Mody V, Nannini M, Ortwine DF, Ran Y, Sambrone A, Sampath D, Vinogradova M, Zhong Y, Nwachukwu JC, Nettles KW, Lai T, Liao J, Zheng X, Chen H, Wang X, Liang J (2019) Unexpected equivalent potency of a constrained chromene enantiomeric pair rationalized by co-crystal structures in complex with estrogen receptor alpha. Bioorg Med Chem Lett 29:905–911
16. Pike ACW, Brzozowski AM, Carlquist M (2001) Rat estrogen receptor beta ligand-binding domain in complex with pure antiestrogen Ici164 384. Structure 9:145
17. Malamas MS, Manas ES, McDevitt RE, Gunawan I, Xu ZB, Collini MD, Miller CP, Dinh T, Henderson RA, Keith JC Jr, Harris HA (2004) Crystal structure of estrogen receptor beta complexed with CL-272. J Med Chem 47:5021–5040
18. Manas ES, Xu ZB, Unwalla RJ, Somers WS (2004) Understanding the selectivity of genistein for human estrogen receptor-beta using X-ray crystallography and computational methods. Structure 12:2197–2207
19. Norman BH, Dodge JA, Richardson TI, Borromeo PS, Lugar CW, Jones SA, Chen K, Wang Y, Durst GL, Barr RJ, Montrose-Rafizadeh C, Osborne HE, Amos RM, Guo S, Boodhoo A, Krishnan V (2006) Benzopyrans are selective estrogen receptor Beta agonists with novel activity in models of benign prostatic hyperplasia. J Med Chem 49:6155–6157
20. Mewshaw RE, Bowen SM, Harris HA, Xu ZB, Manas ES, Cohn ST (2007) ER beta ligands. Part 5: synthesis and structure-activity relationships of a series of 4'-hydroxyphenyl-aryl-carbaldehyde oxime derivatives. Bioorg Med Chem Lett 17:902–906
21. Roberts LR, Armour D, Barker C, Bazin R, Bess K, Brown A, Favor D, Ellis D, MacKenny M, Pullen N, Stennett A, Strand L, Styles M, Phillips C (2011) Sulfonamides as selective estrogen receptor beta agonists. Bioorg Med Chem Lett 21:5680

22. Richardson TI, Dodge JA, Durst GL, Pfeifer LA, Shah J, Wang Y, Durbin JD, Krishnan V, Norman BH (2007) Benzopyrans as selective estrogen receptor beta agonists (SERBAs). Part 3: synthesis of cyclopentanone and cyclohexanone intermediates for C-ring modification. Bioorg Med Chem Lett 17:4824–4828
23. Richardson TI, Dodge JA, Wang Y, Durbin JD, Krishnan V, Norman BH (2007) Estrogen receptor beta ligand-binding domain complexed to a benzopyran ligand. Bioorg Med Chem Lett 17:5563–5566
24. Moecklinghoff S, Rose R, Ottmann C, Brunsveld L (2010) Crystal structure of phosphorylated estrogen receptor beta ligand binding domain. ChemBioChem 11:2251–2254
25. Textor L, Nascimento AS, Polikarpov I (2017) Crystal Structure of ER beta bound to Estradiol. Sci Rep 7:3509
26. GaussView, Version 6.1, Roy Dennington, Todd A. Keith, and John M. Millam, Semichem Inc., Shawnee Mission, KS, 2016
27. Gaussian 16, Revision B.01, Frisch MJ, Trucks GW, Schlegel HB, Scuseria GE, Robb MA, Cheeseman JR, Scalmani G, Barone V, Petersson GA, Nakatsuji H, Li X, Caricato M, Marenich AV, Bloino J, Janesko BG, Gomperts R, Mennucci B, Hratchian HP, Ortiz JV, Izmaylov AF, Sonnenberg JL, Williams-Young D, Ding F, Lipparini F, Egidi F, Goings J, Peng B, Petrone A, Henderson T, Ranasinghe D, Zakrzewski VG, Gao J, Rega N, Zheng G, Liang W, Hada M, Ehara M, Toyota K, Fukuda R, Hasegawa J, Ishida M, Nakajima T, Honda Y, Kitao O, Nakai H, Vreven T, Throssell K, Montgomery JA, Peralta JE JR, Ogliaro F, Bearpark NJ, Heyd JJ, Brothers EN, Kudin KN, Staroverov VN, Keith TA, Kobayashi R, Normand J, Raghavachari K, Rendell AP, Burant JC, Iyengar SS, Tomasi J, Cossi M, Millam JM, Klene M, Adamo C, Cammi R, Ochterski JW, Martin RL, Morokuma K, Farkas O, Foresman JB, Fox DJ (2016) Gaussian, Inc., Wallingford CT
28. PerkinElmer (2012) ChemBioDraw Ultra Version (13.0.0.3015)
29. Jamroz MH (2010) Vibrational energy distribution analysis VEDA 4, Warsaw, 2004–2010
30. Friesner RA, Murphy RB, Repasky MP, Frye LL, Greenwood JR, Halgren TA, Sanschagrin PC, Mainz DT (2006) Extra precision glide: docking and scoring incorporating a model of hydrophobic enclosure for protein-ligand complexes. J Med Chem 49:6177–6196
31. Schrödinger Release 2021–2: Epik, Schrödinger, LLC, New York, NY, 2021
32. Schrödinger Release 2021–2: Protein Preparation Wizard; Epik, Schrödinger, LLC, New York, NY, 2021; Impact, Schrödinger, LLC, New York, NY; Prime, Schrödinger, LLC, New York, NY, 2021
33. Schrödinger Release 2021–2: SiteMap, Schrödinger, LLC, New York, NY, 2021
34. Kaya S, Sayin K, Erkan S, Karakaş D (2022) Investigations of structural, spectral (IR and NMR) and in silico analyses of boron compounds as SERM inhibitor. Chemical Data Collections 37:100816
35. Üngördü A, Sayin K (2019) Quantum chemical calculations on sparfloxacin and boron complexes. Chem Phys Lett 733:136677
36. Sayin K, Üngördü A (2019) Investigations of structural, spectral and electronic properties of enrofloxacin and boron complexes via quantum chemical calculation and molecular docking. Spectrochim Acta Part A Mol Biomol Spectrosc 220:117102
37. Gorbunova Y, Zakusilo DN, Boyarskaya IA, Vasilyev AV (2020) Reactions of 3-arylpropenenitriles with arenes under superelectrophilic activation conditions: Hydroarylation of the carbon-carbon double bond followed by cyclization into 3-arylindanones. Tetrahedron 76:131264
38. Zhang X, Wang X, Wang B, Ding Z-J, Li C (2020) Carbon-carbon double bond in pillar[5]arene cavity: selective binding of cis/trans-olefin isomers. Chin Chem Lett 31:3230–3232
39. Zhou H-B, Nettles KW, Bruning JB, Kim Y, Joachimiak A, Sharma S, Carlson KE, Stossi F, Katzenellenbogen BS, Greene GL, Katzenellenbogen JA (2007) Elemental isomerism: a boron-nitrogen surrogate for a carbon-carbon double bond increases the chemical diversity of estrogen receptor ligands. Chem Biol 14:659–669
40. Tüzün B, Sayin K (2019) Investigations over optical properties of boron complexes of benzothiazolines. Spectrochim Acta Part A Mol Biomol Spectrosc 208:48–56
41. Sayin K, Karakaş D (2018) Quantum chemical investigation of levofloxacin-boron complexes: a computational approach. J Mol Struct 1158:57–65
42. Çöpçü B, Sayin K, Karakaş D (2021) Investigations substituent effect on structural, spectral and optical properties of phenylboronic acids. J Mol Struct 1227:129550
43. Sayin K (2019) Research of the substituent effect on non-linear optical properties of bis (difluoroboron)-1,2-bis((1H-pyrrol-2-yl) methylene)hydrazine (BOPHY) derivatives: molecular simulation analyses. Spectrochim Acta Part A Mol Biomol Spectrosc 212:380–387
44. Gramatica P (2007) Principles of QSAR models validation: internal and external. QSAR Comb Sci 26(5):694–701

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Springer Nature or its licensor (e.g. a society or other partner) holds exclusive rights to this article under a publishing agreement with the author(s) or other rightsholder(s); author self-archiving of the accepted manuscript version of this article is solely governed by the terms of such publishing agreement and applicable law.