QSAR, Ligand Based Design and Pharmacokinetic Studies of Parviflorons Derivatives as Anti-Breast Cancer Drug Compounds Against MCF-7 Cell Line

Hadiza Lawal Abdulrahman¹ · Adamu Uzairu¹ · Sani Uba¹

Received: 28 July 2020 / Accepted: 12 November 2020 / Published online: 29 November 2020
© The Author(s) 2020

Abstract

The anti-proliferative activities of Novel series of Parviflorons against MCF-7 breast cancer cell line was explored via in-silico studies like Quantitative Structure–Activity Relationship QSAR, designing new compounds and analyzing the pharmacokinetics properties of the designed compounds. From QSAR, model one emerged the best from the statistical assessments of (R^2) = 0.9444, (R^2_{adj}) = 0.9273, (Q^2) = 0.8945 and (R^2_{pred}) of 0.6214. The model was used in designing new derivative compounds, with higher effectiveness against estrogen positive breast cancer (MCF-7). The pharmacokinetics analysis carried out on the newly designed compounds showed that all the compounds passed the drug-likeness test and also the Lipinski rule of five, and they could further proceed to pre-clinical tests. The results indicates that the derivative compounds would serve as potent cure to estrogen positive breast cancer (MCF-7 cell line).

Keywords QSAR analysis · Ligand based design · Pharmacokinetics · Breast cancer

1 Introduction

After heart disease, cancer is the seconding cause of mortality amongst mankind, in 2018 about 18.1 million cases were diagnosed. Of all the cancer disease, mammary tumor is frequently amongst women and is the second death related disease amongst women folk. After the collective effort towards concurring this disease, the tumor remains a big challenge globally [7]. In the survival of mammary tumor there is a bigger difference globally, with an approximation of 5 years of 80% in the advanced countries to below 40% for emerging countries. Advancing countries battle with resources and infrastructural limitations that challenges the course of successful mammary tumor outcome by early detection, diagnosis and management [4].

Computer-aided drug discovery (CADD) and designing ensures the best possible lead compound, it reduces the cost related to discovering a drug and it also reduces the time taken for the drug to pass through other stages before its ready for usage. It is a fundamental way in the drugs discovery arena. CADD techniques ascertains principal molecule by assessing, predicting the potency, the probable side effect and also assist in correcting drug-likeness of the compounds [11]. Drug compounds that have low drug-likeness and ADMET properties wont progress to pre-clinical research, irrespective of the high biological activity. ADMET is one of the main properties used in analyzing a drug compound, though a significant progress was made when a lot of consideration was given to such properties recently [17].

There has been an increase of breast cancer occurrences, which is still the most substantial cause of mortality among the female being. Despite the headway made in managing breast cancer, the search for a curative treatment is still ongoing as most times, the tumor becomes resistance to this treatment with a short time. Although a number of crucial studies and clinical trials have significantly contributed to the enhancement of mammary tumor care, many cancer cases and pathway often remain yet unknown to the majority of clinicians [19]. Recently [14] reported the anti-proliferative

✉ Hadiza Lawal Abdulrahman
azezalawal@gmail.com

Adamu Uzairu
adamuuzairu@yahoo.com

Sani Uba
saniuba10@yahoo.com

¹ Department of Chemistry, Ahmadu Bello University, P.M.B. 1044, Zaria, Nigeria

activities of some novel compounds of Parviflorons derivatives against MCF-7 cell line. This study is aimed at building a mathematical QSAR, design new Parviflorons compounds based on a derived QSAR model and to furthermore ascertain the pharmacokinetic properties of the newly designed drug compounds.

Luminal type breast cancer (MCF-7) are Estrogen receptor (ER)/progesterone receptor (PR)—positive type which are caused by the over expression of estrogen receptor α (ER α). It accounts for about 70% of the mammary tumor patients tagged as ER positive (ER+). The constant activation of ER α by estrogens induces the proliferation of cancer cell [12].

2 Methodology

2.1 Computational Information

2.1.1 Hardware and Software

The computer details used in this research is; 7th generation HP pavilion Intel R, core i7-7500u RAM 12.00 GB running on a windows 10 operating system. The software's used to carry out this research includes Spartan' 14 (version 1.1.2), Material studio (V8), Auto dock visualizer version 4.2, Pyrex software version, Chemdraw software version 12.0.2, PADEL-Descriptor Software V2.20 and DTC data lab software and Microsoft word Office Excel 2013 version.

2.2 QSAR Assessment

2.2.1 Data Gathering

Twenty-six (26) novel derivative compounds of Parviflorons derivatives against MCF-7 cell line with their anti-proliferative activities reported in inhibitory concentration (IC_{50}), against breast cancer (MCF-7) cell line were reported from [14] article.

2.2.2 Anti-proliferative Activities and Geometry Optimization

The IC_{50} values were normalized to pIC_{50} using scale of logarithm $\{pIC_{50} = -\log_{10}(IC_{50} \times 10^{-6})\}$. The tabulated anti-proliferative activities (IC_{50}) and pIC_{50} of the derivatives are shown in Table 1, measured in concentration of micro molar (μM). QSAR analysis requires intensive attention for the whole job to be executed. At the start, drawing the structure is a crucial step for the calculation of molecular descriptors as the independent variables. In this research, Chemdraw V (12.0.2) was used in drawing Parviflorons derivatives and converted to 3D format

Table 1 Parviflorons derivatives and its activities

No	Structure	IC_{50}	pIC_{50}
1		6.1	5.2147
2		6.4	5.1938
3		5.8	5.2366
4		6.0	5.2219

Table 1 (continued)

No	Structure	IC ₅₀	pIC ₅₀
5		5.9	5.2291
6		39.0	4.4089
7		20.4	4.6904

Table 1 (continued)

No	Structure	IC ₅₀	pIC ₅₀
8		2.8	5.5528
9		6.4	5.1938
10		7.1	5.1487

Table 1 (continued)

No	Structure	IC ₅₀	pIC ₅₀
11		17.6	4.7545
12		4.4	5.3565
13		8.1	5.0915
14		5.2	5.2830

Table 1 (continued)

No	Structure	IC ₅₀	pIC ₅₀
15		17.1	4.7670
16		17.9	4.7471
17		22.9	4.6401

Table 1 (continued)

No	Structure	IC ₅₀	pIC ₅₀
18		27.2	4.6402
19		20.9	4.6799
20		17.0	4.7696

Table 1 (continued)

No	Structure	IC ₅₀	pIC ₅₀
21		19.1	4.7180
22		26.3	4.5800
23		21.9	4.6596

Table 1 (continued)

No	Structure	IC ₅₀	pIC ₅₀
24		24.1	4.6170
25		34.2	4.4650
26		29.4	4.5317

for geometric optimization using Spartan 14 V (1.1.4) software, using Density Functional Theory (DFT) with B3LYP, 6-311G basis set, for the geometric optimization of the compounds [3]. The aim of optimization is to acquire a more appropriate 3-dimensional structure that is very close with the original 3-dimensional molecular structure. Therefore the molecular parameters may well represent the main physicochemical properties of the observed molecule [15].

2.2.3 Molecular Descriptors Calculations and Pretreatment

26 derivative compounds of Parviflorons were converted to SDF format after optimization. Pharmaceutical Data Exploration Laboratory Software V (2.20) was used in calculating physicochemical descriptors [18]. The descriptors were pretreated using Data Pre-treatment software GUI 1.2 [1] to remove irrelevant values.

2.2.4 Division of Data Set and Model Building

Kennard-Stone algorithm [13] method was utilized to distribute the derivatives into training or calibration set and test or validation set to build the model [2]. The calibration set is used to develop a calibrated model that would be used in predicting the bio-activities of the validation set of molecules. Version 8 of Material studio software was utilized in constructing a mathematical model with Genetic Function Approximation (GFA) technique. The dependent variable is the anti-proliferative activities (pIC₅₀) and the independent variable are model parameters (descriptors) which were obtained using Pharmaceutical Data Exploration Laboratory Software V (2.20).

2.2.5 Model Validation (Internal)

Internal validation employs the derivative compounds used in generating the model and checks for core effectiveness. Cross-Validation (CV) procedure is commonly utilized as an internal validation technique for the derived model, mostly one compound from the train set is removed, The n-1 (n = the total molecules) molecules are utilized in building the model using the calibration or train set. The anti-proliferative activity of the compound removed is calculated once, the method is repeated n times for every molecule, thus every molecule having a calculated activity [6]. Such procedure is known as leave-one-out (LOO) technique. It's given as:

$$Q_{cv}^2 = 1 - \left[\frac{\sum (Y_{pred} - Y_{exp})^2}{\sum (Y_{exp} - Y_{mintraining})^2} \right]$$

$Y_{training}$, Y_{exp} , and Y_{pred} are the average activities (pIC₅₀) of training set, bio-activities (IC₅₀) and prediction inhibition concentration of the train set [5]. The coefficient of correlation for the cross-validated technique R^2 is given as:

$$R^2 = 1 - \left[\frac{\sum (Y_{exp} - Y_{pred})^2}{\sum (Y_{exp} - Y_{training})^2} \right]$$

Where Y_{exp} and Y_{pred} are averages of the actual and predicted activity of the training sets [16]. It is a research tool used in estimating the prediction power of the statistical model that was acquired from a regression technique.

2.2.6 Model Validation (External)

A built model with excellent good fit and an approved prediction can still be faulty in an actual relationship between (model descriptors) predictor variables and (bio-activity) response variables. The degree of potency of the built model (equation) is analyzed by external validation, it calculates the degree of fitness of the model. The criteria proposed by Golbraikh and Tropsha for an effective built model with good predictive power are stated as follows;

- a. $R^2_{pred} > 0.6$
- b. $< \frac{r^2 - r_o^2}{r^2} 0.1$
- c. $0.85 < k < 1.15$ or $0.85 < k' < 1.15$.

where r^2 is the squared correlation coefficient between the actual and calculated activity, r_o^2 is the correlation coefficient squared between the actual and calculated activity, and k and k' are the regression slopes passing through the origin [2].

2.2.7 QSAR Applicability Domain of Model

The goal of an applicability domain methods is for estimating individually, the reliability of each generated model [8]. A model validation should be within the training domain and its essential for the compounds to be assessed as fitting within the domain to ascertain the model. An applicability domain is evaluated by the leverage value for every molecule. The leverage (L) defines the applicability domain of the generated equation [20]. It is formulated as;

$$L_i = x_i (X^T X)^{-k} x_i^T (i = K, \dots, P).$$

Where X^T is the matrix transpose of X used in constructing the model, X_i is matrix of train compounds of I and X is the $n \times k$ matrix of train set descriptors. (H^*) is the warning leverage, it is a prediction tool that checks for outliers. It's written as;

$$H^* = \frac{3(p+1)}{m}$$

p equals to the total structural descriptors and m is the total compounds of train sets. The William's plot (A plot of standardized values versus the leverage values) of both the training (calibration) and test (validation) set. Molecules that fall within the warning leverages on the plot are the predicted compounds.

2.2.8 Computational Pharmacokinetics (Drug-Likeness)

SwissADME was used in analyzing the drug-likeness of the newly designed compounds. Furthermore, the designed compounds was checked for their compliance with Lipinski's rule of five [10], a well-used criteria to comprehend if a compound can be orally absorbed or not, such as: molecular weight (MW) ≤ 500 , octanol/water partition coefficient ($AlogP$) ≤ 5 , number of hydrogen bond donors (HBDs) ≤ 5 and number of hydrogen bond acceptors (HBAs) ≤ 10.6 . According to the Rule of Five, a drug compound would not be orally active if it violates two or more of the four rules [9].

3 Results and Discussion

3.1 Insilico QSAR Investigation

Insilico QSAR investigation was used in finding a simple mathematical equation that was used in calculating an enhanced anti-proliferative activities from structures of Parviflorons derivatives. The QSAR investigation also correlated the molecular descriptors (model parameters) with the physico-chemical properties of the 26 derivative compounds (bio-activities) using statistical techniques. Based on the Genetic Function Approximation (GFA) technique employed, four QSAR models were generated to predict the anti-proliferative activities of Parviflorons derivatives. Model 1 (one) passed both internal and external validation with correlation coefficient squared (R^2) of 0.9444, correlation coefficient adjusted squared (R^2_{adj}) of 0.9273, cross validation coefficient (Q^2) of 0.8945. The external validation of (R^2_{pred}) of 0.6214 for model 1 was calculated using the model descriptors from the test set as shown in Tables 2 and 3. The robustness of the QSAR models were assessed using the reliability of the train set and predicted pIC_{50} of the test set, which agrees with the criteria proposed by Golbraikh and Tropsha ($R^2_{pred} > 0.6$) for an effective QSAR model as shown in Table 5.

Model 1

$$Y = 0.157509087 \times nX + 110477157080 \times MATS3e - 1.703000586 \times GATS5e + 0.574341593 \times MLFER_BO + 5.589384.$$

The robustness of the QSAR models were assessed by the reliability of the calibration set and calculated pIC_{50} of the validation set. The Experimental, predicted and the residual values of Parvifloron derivatives are shown in Table 4. The low residual value is obtained from the difference between the anti-proliferative and calculated activity, indicating the high predictive power of the model. Both internal and external validation conforms model 1 to be very stable and highly effective.

Table 2 External validation of model 1

Name	pIC ₅₀	nX	MATS3e	GATS5e	MLFER_BO	Y _{pred}	Y _{pred} -Y _{obs}
1	5.2147	0	0.0768	1.1787	1.459	5.3011	0.0864
19	4.5654	1	0.0393	1.2185	0.930	4.6568	0.0913
25	4.6170	0	0.0159	1.0686	1.1560	4.6166	- 0.0014
3	5.2366	0	0.1157	1.3261	1.5340	5.5404	0.3038
14	5.0915	0	0.0601	1.0769	1.8800	5.5249	0.4334
8	5.5528	0	0.1100	1.3110	0.8590	5.1210	- 0.4309
17	4.7471	1	0.0076	0.9358	0.9300	4.7747	0.0275
13	5.3565	0	0.0707	1.1310	1.7210	5.4626	0.1060

Table 3 Continuation of descriptor calculation used in external validation of model 1

(Y _{pred} -Y _{obs}) ²	Y _{mintrn}	(Y _{mintrn} -Y _{obs})	(Y _{mintrn} -Y _{obs}) ²
0.0074	4.8299	- 0.3848	0.1481
0.0083	4.8299	0.2644	0.0699
1.94E-06	4.8299	0.2119	0.0449
0.0923	4.8299	- 0.4065	0.1654
0.1879	4.8299	- 0.2616	0.0685
0.1857	4.8299	- 0.7220	0.5227
0.0007	4.8299	0.0827	0.0068
0.0112	4.8299	- 0.5267	0.2774

$$\sum(Y_{ob} - Y_{pred})^2 = 0.4936, \sum(Y_{obs} - Y_{train})^2 = 1.3036, \therefore R^2_{test} = 1 - (0.7130/1.5357) = 0.6214$$

Table 5 defines the model parameters (descriptors) in the calculated model, the descriptors were used in verifying the model both internally and externally. They were calculated using PADEL-Descriptor Software V2.20 from (Abdullahi et al. [2]). The effectiveness and predicting power of the generated model was assessed using internal and external validation analysis, the model conformed with the least approved QSAR model values, indicating that model can be used in designing new Parvifloronderivatives compounds with better anti-breast cancer activity as seen in table 6.

Statistical analysis was used in evaluating the individual contribution of each molecular descriptor in the QSAR model, i.e. the Mean effect and VIF (Variance Inflation Factor). The coefficient of the mean effect values are used to either increase or decrease the effect of the descriptors. Therefore, increasing nX, GATS5e and MLFER_BO would increase the bio-activities of the derivative compounds (positive coefficient) while decreasing MATS3e would also increase the bio-activities of the derivative compounds (negative coefficient) as proven in Table 7. VIF (Variance Inflation Factor) gives a degree of the inter-relationship amongst the model parameters. The VIF scores were within the approved value of 1–5, indicating that there is no collinearity between the bio-activity and model parameters (descriptors) of the constructed model, as shown in Table 7.

Table 4 The bio-activities (pIC₅₀), calculated activities and residual values of model 1

No	pIC ₅₀	Predicted IC ₅₀	Residual
1	5.2147	5.3011	0.0864
2*	5.1938	5.1603	0.0335
3	5.2366	5.5404	0.3038
4*	5.2219	5.2735	- 0.0516
5	5.2291	5.1210	0.1072
6	4.4089	4.4058	0.0031
7	4.6904	4.6790	0.0104
8	5.5528	5.1210	- 0.4309
9*	5.1938	5.1420	0.0518
10	5.1487	5.2251	- 0.0764
11	4.7545	4.8763	- 0.1218
12	5.3565	5.4626	0.1060
13	5.0915	5.5249	0.4334
14*	5.2830	5.2454	0.0386
15*	4.7670	4.7365	0.0305
16	4.7471	4.7747	0.0275
17	4.6402	4.6516	- 0.0115
18*	4.5654	4.6568	0.0913
19	4.6799	4.6175	0.0624
20*	4.7696	4.7780	- 0.0085
21	4.7180	4.5996	0.1194
22	4.5800	4.5292	0.0509
23	4.6596	4.7609	- 0.1014
24	4.6170	4.6166	- 0.0014
25	4.4650	4.5503	- 0.0843
26*	4.5317	4.5830	- 0.0523

The compounds with (*) are the test set while the compounds without (*) are the train set

Figure 1 shows a graph of observed activities against the calculated activities of both the test set and the train set of compounds. The plot showed that the predicted activity was in good agreement with its experimental values as shown in Table 2, conforming to the effectiveness and stability of the model generated.

Table 5 Definition of descriptors and their classes for the model

Name	Definition	Class
nX	Number of halogen atoms	2D
MATS3e	Moran autocorrelation-lag 3/weighted by Sanderson electronegativities	2D
GATS5e	Geary autocorrelation-lag 5/weighted by Sanderson electronegativities	2D
MLFER_BO	Overall or summation solute hydrogen bond basicity	2D

Table 6 Golbraikh and Tropsha approved QSAR model standards

Parameter	Threshold score	Model score
Q^2_{cv}	$Q^2_{cv} > 0.5$	0.5958
R^2_{train}	$R^2_{train} > 0.6$	0.729
R^2_{test}	$R^2_{test} > 0.6$	0.672
$ r^2_o - r^2_\sigma $	$ r^2_o - r^2_\sigma < 0.3$	0.019
K	$0.85 < k < 1.15$	1.032
K'	$0.85 < k < 1.15$	0.967
$(r^2 - r^2_o)/r^2$	$(r^2 - r^2_o)/r^2 < 0.1$	0.018
$(r^2 - r^2_\sigma)/r^2$	$(r^2 - r^2_\sigma)/r^2 < 0.1$	0.098

Figure 2 shows the values of both test and train set spread on both sides of zero point on the plot, showing no systematic errors between the standardized residual versus the anti-proliferative activity (Experimental activity).

Figure 3 shows the standardized residuals against the leverage values also called William's plot. Most of the compounds fell within the applicability domain from the calculated leverage of ($L=0.833$), only 3 compounds we

found outside the applicability domain which might be due to a slight changes in their molecular structure as compared with other molecules in the data set.

3.2 Ligand-Based Drug Design

Eight (8) new Parviflorons derivative compounds were designed using the ligand based approach. The lead compounds (4 and 16) were chosen due to their low residual values and high pIC_{50} values as shown in Table 4. This approach uses the molecular descriptors obtained from the mathematical QSAR model and adjustments were made on the lead compounds (4 and 16) based on the definition of the molecular descriptors nX, having a positive coefficient (this mean adding either of the halogen atoms, which includes F, Cl, Br, I etc. at different structural positions) and GATS5e also having a positive coefficient (this also means adding electronegative compounds such as OH, OCH_3 etc.) as shown in Table 5. The newly designed compounds and their new calculated activities are seen visually in Table 8.

Table 7 Statistical analysis of model 1 parameters

	nX	MATS3e	GATS5e	MLFER_BO	VIF	Mean Effect
nX	1	− 0.6337	− 0.6707	− 0.1514	0.0169	0.0337
MATS3e	− 0.6337	− 0.6334	0.8728	0.1169	− 0.9919	− 0.0185
GATS5e	− 0.6707	0.8728	1	0.36288	0.7478	0.9562
MLFER_BO	− 0.1514	0.1169	0.36288	1	1.1292	0.0286

Fig. 1 plot of predicted activities versus inhibition concentration **Fig. 2** A graph of standardized residual against bioactivities (Experimental activities)

Fig. 3 The William's plot

3.3 Physicochemical and ADME Properties (Pharmacokinetics) of Designed Parvifloron Compounds

There are lot of designed compounds that fail to become drugs. Efficiency and safety of the drug to the system are the main cause of drug failure, these indicates the absorption, distribution, metabolism, excretion, and toxicity (ADMET) properties of compounds plays a major role in every step of the drug discovery pipeline. Therefore, it is compulsory to discover potent compounds with effective ADMET properties (Guan et al. [9]). All the newly designed compounds were assessed for their drug-likeness (ADME and physicochemical properties). None of the designed compounds violated two rules out of the Lipinski rule of five; a prominent principle used in certifying the drug-likeness of a compound, this shows that all the designed compounds passed the drug-likeness test as shown in Table 9, making the compounds a breakthrough in finding the cure to triple-negative breast cancer. Figure 4 shows the bioavailability radar for molecules 1 and 6. The Bioavailability Radar gives an initial scan at the drug-likeness of the compound.

4 Conclusion

Parvifloron derivatives showed a more promising anti-breast cancer drug candidate against MCF-7 cell line via QSAR studies and pharmacokinetics analysis. Based on the statistical analysis from the mathematical model obtained from QSAR studies showed that increasing nX, GATS5e and MLFER_BO descriptors will increase the anti-proliferative activities of Parvifloron derivatives while decreasing MATS3e would also increase the anti-proliferative of Parvifloron derivatives as a standard anti-breast cancer drug

Table 8 Newly designed imidazole derivative compounds with their new predicted activities (pIC_{50})

S/No	Compounds	New predicted activity (pIC_{50})
1		6.6970
2		5.4845
3		5.4540

Table 8 (continued)

S/No	Compounds	New predicted activity (pIC ₅₀)
4		5.5084
5		5.5707
6		5.4600

Table 8 (continued)

S/No	Compounds	New predicted activity (pIC ₅₀)
7		5.4586
8		5.2738

agent. The effectiveness and predicting power of the generated model was assessed using internal and external validation analysis, the model conformed with the least approved values, indicating that model can be used in designing new Parvifloron derivatives compounds with better anti-breast cancer activity. The molecular descriptors (nX and GATS5e) had more significant and based on their mean effect, adjustment were made on the fragments of the lead compounds (4 and 16) to design eight new Parvifloron derivative compounds with a higher predicted activity against MCF-7 cell line.

Furthermore, the pharmacokinetics analysis (drug-likeness test) carried out on the newly designed Parvifloron

Table 9 Physicochemical and ADME properties (Pharmacokinetics) of designed imidazole derivative compounds against MCF-7 cell line

S/No	MW (mg/mol)	nAH	nRB	HBA	HBD	MR	TPSA (Å ²)	iLOGP	BBB	PAIN	Brenk
1	454.51	6	5	6	1	121.98	63.60	4.55	no	0	0
2	487.41	6	5	4	1	132.09	63.60	4.58	No	0	0
3	487.41	6	5	4	1	132.09	63.60	4.43	No	0	0
40	454.51	6	5	6	1	121.98	63.60	4.19	No	0	0
5	446.36	0	2	3	3	107.30	60.69	3.43	Yes	0	1
6	478.26	0	2	3	3	110.08	60.69	2.75	Yes	0	1
7	399.36	0	2	3	3	102.21	60.69	3.32	Yes	0	0
8	389.36	0	2	3	3	103.93	60.69	3.37	Yes	0	1

MW Molecular weight (<500 mg/mol), nAH number of aromatic heavy atoms, nRB rotatable bonds, HBA Hydrogen bond acceptors, HBD Hydrogen bond donors, MR molecular refractivity, TPSA Topological polar surface area, BBB blood–brain barrier

Fig. 4 the bioavailability radar for molecules 1 and 6

compounds revealed that all the compounds passed drug-likeness test (ADME and other physicochemical properties) and they also adhered to the Lipinski rule of five: a criteria used in evaluating the drug-likeness of compounds. This concludes that the compounds can move on to the next step of pre-clinical trial, showing a great discovery for medicine in finding permanent solutions to breast cancer (MCF-7 cell line).

Funding This research did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors.

Compliance with Ethical Standards

Conflict of interest No conflict of interest.

Ethics Approval and Consent to Participate Not applicable.

Consent of Publication Not applicable.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

References

1. Abdulrahman HL, Uzairu A, Uba S (2020) Computer modeling of some anti-breast cancer compounds. Struct Chem. <https://doi.org/10.1007/s11224-020-01608-7>
2. Abdullahi M, Shallangwa GA, Uzairu A (2020) In silico QSAR and molecular docking simulation of some novel aryl sulfonamide

- derivatives as inhibitors of H5N1 influenza A virus subtype. Beni-Suef Univ J Basic Appl Sci 9(1):2
3. Arthur DE, Uzairu A, Mamza P, Abechi SE, Shallangwa G (2018) Activity and toxicity modelling of some NCI selected compounds against leukemia P388ADR cell line using genetic algorithm-multiple linear regressions. J King Saud Univ Sci 32:324
 4. Akram M, Iqbal M, Daniyal M, Khan AU (2017) Awareness and current knowledge of breast cancer. Biol Res 1(1):1–44
 5. Brandon. V., and Orr, A. (2015). Comprehensive R archive network (CRAN): Rproject.org.
 6. Deeb O (2012) Recent applications of quantitative structure-Activity relationships in drug design. Medicinal Chemistry and Drug Design. InTech, London, pp 55–82
 7. El-Kashef H, Badr G, El-Maali NA, Sayed D, Melnyk P, Lebegue N, El-Khalek RA (2020) Synthesis of a novel series of (Z)-3, 5-disubstituted thiazolidine-2, 4-diones as promising anti-breast cancer agents. Bioorg Chem 5(3):61–97
 8. Eriksson L, Jaworska J, Worth AP, Cronin MT, McDowell RM, Gramatica P (2003) 4Methods for reliability and uncertainty assessment and for applicability evaluations of classification-and regression-based QSARs. Environ Health Perspect 111(10):1361–1375
 9. Guan L, Yang H, Cai Y, Sun L, Di P, Li W, Tang Y (2019) ADMET-score—a comprehensive scoring function for evaluation of chemical drug-likeness. MedChemComm 10(1):148–157
 10. Hou Y, Zhu L, Li Z, Shen Q, Xu Q, Li W, Gong P (2019) Design, synthesis and biological evaluation of novel 7-amino-[1, 2, 4] triazolo [4, 3-f] pteridinone, and 7-aminotetrazolo [1, 5-f] pteridinone derivative as potent antitumor agents. Eur J Med Chem 163:690–709
 11. Leelananda SP, Lindert S (2016) Computational methods in drug discovery. Beilstein J Org Chem 12(1):2694–2718
 12. Jordan VC, Brodie AM (2007) Development and evolution of therapies targeted to the estrogen receptor for the treatment and prevention of breast cancer. Steroids 72(1):7–25
 13. Kennard RW, Stone LA (1969) Computer aided design of experiments. Technometrics 11:137–148
 14. Miyajima Y, Saito Y, Takeya M, Goto M, Nakagawa-Goto K (2019) Synthesis of 4-epi-Parviflorons A, C, and E: structure-activity relationship study of antiproliferative abietane derivatives. J Org Chem 84(6):3239–3248
 15. Putri DE, Pranowo HD, Haryadi WINARTO (2019) Study on anti-tumor activity of novel 3-substituted 4 anilino-coumarin derivatives using quantitative structure-activity relationship (QSAR). Mater Sci Forum 948:101–108
 16. Tropsha A, Gramatica P, Gombar VK (2003) The importance of being earnest: validation is the absolute essential for successful application and interpretation of QSPR models. Mol Inform 22:69–77
 17. Yang H, Sun L, Wang Z, Li W, Liu G, Tang Y (2018) ADMETopt: a web server for ADMET optimization in drug design via scaffold hopping. J ChemInf Model 58(10):2051–2056
 18. Yap CW (2011) PaDEL-descriptor: an open source software to calculate molecular descriptors and fingerprints. J Comput Chem 32:1466–1474
 19. Uysal E (2017) Top100 cited classic articles in breast cancer research. Eur J breast health 13(3):129
 20. Veerasamy R, Rajak H, Jain A, Sivadasan S, Varghese CP, Agrawal RK (2011) Validation of QSAR models-strategies and importance. Int J Drug Des Discov 3:511–519