ORIGINAL RESEARCH

Computer modeling of some anti-breast cancer compounds

Hadiza Lawal Abdulrahman [1] · Adamu Uzairu [1] · Sani Uba [1]

Abstract
The research was aimed at exploring the biological activities of novel series of β-lactam derivatives against MCF-7 breast cancer cell lines via computer modeling such as quantitative structure-activity relationship (QSAR), designing new compounds and analyzing the drug likeliness of designed compounds. The QSAR model was highly robust as it also conforms to the least minimum requirement for QSAR model from the statistical assessments with a correlation coefficient squared (R^2) of 0.8706, correlation coefficient adjusted squared (R^2_{adj}) of 0.8411, and cross-validation coefficient (Q^2) of 0.7844. The external validation of R^2_{pred} was calculated as 0.6083 for model 4. The model parameters (MATS5i and MATS1s) were used in designing new derivative compounds with higher potency against estrogen-positive breast cancer. The pharmacokinetics test on the restructured compounds revealed that all the compounds passed the drug likeness test and they could further proceed to clinical trials. These reveal a breakthrough in medicine, in the research for breast cancer drug with higher effectiveness against the MCF-7 cell line.

Keywords QSAR · Ligand-based design · Pharmacokinetics analysis · Breast cancer · MCF-7 cell line

Introduction

Cancer, a dreadful malady, is also referred to as malignant tumors, and it is a heterogeneous tumor that affects almost all parts of the body. Mammary tumor is among the common diseases that brings about morbidity and mortality among the female being [14]. The cases that were encountered with most anti-cancer drugs during the past three decades include drug ineffectiveness, no drug selectivity, growing side effects, and the drug becoming resistant to the tumor [14]. Regardless of the up-to-date diagnostic and therapeutic advancement, mammary tumor is still the common cause of mortality and the primary utmost cancer among women worldwide [2]. Therefore, more effective and safe therapeutic agents are urgently needed with many pathways to increase the positive outcome of the patients clinically [14].

Computational approaches are regularly employed in almost all modern drug discovery effort, and robust success has been attained for computer-aided lead generation and optimization. These techniques are mostly accurate, faster, and cost-efficient. CADD represents more recent applications of software tools in the designing of lead candidate [13]. CADD technique is basically divided into two sections which are the structure-based (SB) and ligand-based (LB) drug discovery. The CADD technique employed depends on the crystal structure (receptor) available. One of the computer-aided tools for drug discovery and design includes quantitative structure-activity relationship (QSAR).

Quantitative structure-activity relationship (QSAR) is a current technique used in optimizing template molecules and re-designing new drug compounds. QSAR calculates the activities, toxicities, and carcinogenicities of molecules obtained from the definition of the molecular parameters from a derived mathematical equation [5]. It is also a known arithmetic relationship connecting molecular compounds and biological activities for a library of molecules quantitatively [3]. An arithmetic equation is generated from the structural info of a well-derived calibration compound and corresponding biological activities, while the model is validated using some validation compounds for which the biological activities are accessible [12].

Malebari et al. [10] reported 43 novel β-lactam derivative compounds as potent inhibitors against estrogen-positive MCF-7 cell line. The purpose of this research is to utilize

✉ Hadiza Lawal Abdulrahman
azeezalawal@gmail.com

Adamu Uzairu
adamuuzairu@yahoo.com

Sani Uba
saniuba10@yahoo.com

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

ligand-based drug design to design new β-lactam derivative compounds based on an established QSAR mathematical model as inhibitors against estrogen-positive breast cancer (MCF-7 cell line) and, furthermore, to test for the pharmaco-kinetic properties of the designed compounds.

Methodology

QSAR process

Data collection

Forty-four (44) new derivative compounds of β-lactams with thier individual inhibitory concentration (IC_{50}) against breast cancer (MCF-7 cell line) were attained from 14 publications.

Bio-activities

The bio-activities of β-lactam derivative compounds were measured in inhibitory concentration (IC_{50}). The scale of logarithm ($pIC_{50} = -\log10 (IC_{50} \times 10^{-6})$) was applied to equalize the IC_{50} values. The IC_{50} and pIC_{50} values of the derivatives are seen in Table 1. It is measured in the concentration of micromolar (μM).

Geometry optimization

This technique was used to get a desirable structure that would be the closest to the initial structural condition [11]. 2D sketching of Chemdraw V (12.0.2) was employed in drawing β-lactam derivatives, and then, it was uploaded on Spartan 14 V (1.1.4) for geometrical optimization. [1]. The template molecule is seen in Fig. 1.
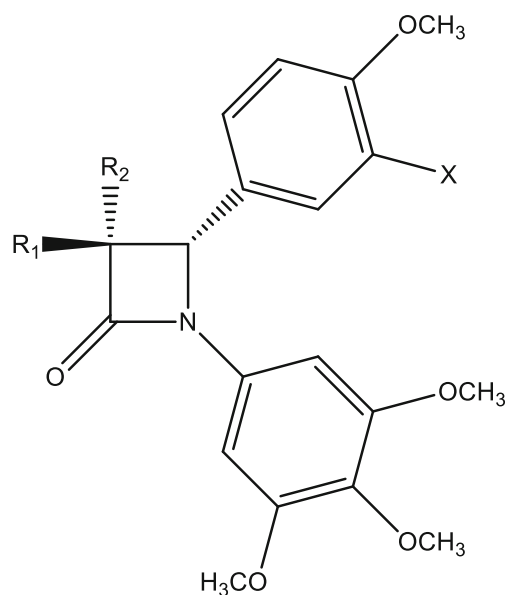
Table 1 β-Lactam derivatives and its bio-activities

No/S.	X	R1	R2	IC_{50}	$logIC_{50}$
1	F	C_6H_5	H	0.017	7.7696
2e	F	OC_6H_5	H	0.073	7.1367
3	F	Cl	H	0.040	7.3979
4	F	Cl	Cl	0.173	6.7610
5	F	C_2H_3	H	0.038	7.4202
6	F	C_3H_5	H	0.323	6.4908
7	Cl	C_6H_5	H	0.230	6.6383
8	Cl	OC_6H_5	H	0.430	6.3665
9	Cl	Cl	H	0.091	7.0400
10	Cl	Cl	Cl	0.140	6.8539
11	Cl	C_2H_3	H	0.120	6.9208
12	Cl	C_3H_5	H	0.270	6.5686
13	Br	C_6H_5	H	0.651	6.1864
14	Br	OC_6H_5	H	0.818	6.0872
15	Br	Cl	H	0.603	6.2197
16	Br	Cl	Cl	1.120	5.9508
17	Br	C_2H_3	H	0.088	7.0555
18	Br	C_3H_5	H	0.284	6.5467
19	I	C_6H_5	H	0.648	6.1884
20	I	OC_6H_5	H	0.208	6.6819
21	I	Cl	H	0.421	6.3757
22	I	Cl	Cl	0.885	6.0531
23	I	C_2H_3	H	0.215	6.6676
24	I	C_3H_5	H	1.354	5.8684
25	CH_V	C_6H_5	H	0.462	6.3354
26	CH_3	OC_6H_5	H	2.235	5.6507
27	CH_3	Cl	H	0.433	6.3635
28	CH_3	Cl	Cl	0.747	6.1267
29	CH_3	C_2H_3	H	0.328	6.4841
31	F	H	H	0.022	7.6576
32	Cl	H	H	0.140	6.8539
33	Br	H	H	0.060	7.2218
34	I	H	H	0.821	6.0857
35	CH_3	H	H	0.317	6.4989
36	F	OH	H	0.022	7.6577
37	Cl	OH	H	0.012	7.9208
38	Br	OH	H	0.044	7.3565
39	I	OH	H	0.100	7.0000
40	CH_3	OH	H	0.005	8.3010
41	F	NH_2	H	0.483	6.3161
42	Cl	NH_2	H	0.141	6.8508
43	OH	C_6H_5	H	0.004	8.3979
44	H	C_6H_5	H	0.015	7.8239

Fig. 1 Template molecule of β-lactam derivatives

Table 2 Externally validating Eq. 4

No.	logIC$_{50}$	ALogP	ATSC0i	MATS5i	MATS1s	ETA_Beta_ns_d	Y_{pred}
22	6.0531	0.7305	65.7792	−0.1148	−0.0233	2.5	5.8359
3	7.3979	0.0003	84.6455	0.0638	−0.0331	2.5	7.5686
28	6.1267	1.0375	65.9475	−0.1293	−0.0298	2.0	5.8965
18	6.5467	0.0355	74.8107	−0.0640	−0.0059	2.5	6.2287
36	7.6576	1.4952	86.0986	0.0348	−0.0867	2.5	7.2948
31	7.6576	0.3256	85.3414	0.0519	−0.0351	2.5	7.2794
42	6.8508	1.1099	67.3467	−0.1138	−0.0981	2.5	6.9724
6	6.4908	96.3243	96.3243	0.0592	−0.0205	2.5	6.7437
10	6.8539	61.3644	61.3644	−0.0350	−0.0251	2.5	6.4113
23	6.6676	75.1465	75.1465	−0.0517	−0.0162	2.5	6.2759
11	6.9208	70.4746	70.4746	0.0171	−0.0179	2.5	6.9492
12	6.5686	74.1892	74.1892	−0.0344	−0.0071	2.5	6.4584
26	5.6507	83.5391	83.5395	−0.1170	−0.0499	2.5	6.3432
27	6.3635	66.7514	66.7514	−0.1385	−0.0277	2.0	6.3223
33	7.2218	63.5676	63.5676	−0.0955	−0.0238	2.5	6.5371

Model parameters

The model parameters were obtained for the whole derived compounds of β-lactams using Pharmaceutical Data Exploration Laboratory Software V (2.20) [16].

Pretreatment and division of data set

The values from the PADEL V (2.20) were prepped using graphical user interface (GUI) 1.2 (Data Pretreatment software), to remove relentless and undesired descriptor values [1]. Kennard-Stone algorithm [8] was employed in splitting the derivatives into calibration and validation fragments in other to construct the equation (model).

Model building and validation

A mathematical equation was built using the train set as predictor variable, while the pIC$_{50}$ was used as the predicted variable by using genetic function approximation (GFA) technique of Material Studio Software V 8. The obtained equations were evaluated using Friedman formula [4].

$$LOF = \frac{SEE}{M\left[1 - \beta\left(\frac{c + d \times p}{M}\right)\right]^2}$$

where SEE is the standard estimated error; it is given as

$$SEE = \sqrt{\frac{(Y_{exp} - Y_{pred})^2}{N - P - 1}}$$

C is the summation of the model definitions, p is the total number of model descriptors, M is the sum of prediction set,

and d is a user-defined smoothing parameter [9]. The model is verified externally using the correlation coefficient (R^2). The nearer R^2 value is to 0.1, the better the regression fitness. R^2 is calculated 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 means of the actual and calculated activities of the calibration sets [1]. $Y_{minraining}$ indicates the average pIC$_{50}$ of the train set molecules [7].

Table 3 Externally validating Eq. 4 (continuation)

$Y_{pred} - Y_{obs}$	$(Y_{pred} - Y_{obs})^2$	Y_{mintrn}	$(Y_{mintrn} - Y_{obs})$	$(Y_{mintrn} - Y_{obs})^2$
−0.2171	0.0471	6.8258	−0.7728	0.5971
0.1706	0.0291	6.8258	0.5721	0.3273
−0.2302	0.0529	6.8258	−0.6991	0.4888
−0.3170	0.1011	6.8258	−0.2791	0.0770
−0.3627	0.1316	6.8258	0.8318	0.6918
−0.3782	0.1420	6.8258	0.8318	0.6918
0.1217	0.0148	6.8258	0.0240	0.0006
0.2529	0.0630	6.8258	−0.3350	0.1122
−0.4426	0.1959	6.8258	0.0281	0.00079
−0.3917	0.1534	6.8258	−0.1582	0.0250
0.0284	0.0008	6.8258	0.0950	0.0090
−0.1102	0.0121	6.8258	−0.2572	0.0661
0.6925	0.4795	6.8258	−0.0751	1.3808
−0.0412	0.0017	6.8258	−0.4623	0.2137
−0.6848	0.4687	6.8258	0.39960	0.1569

$\sum(Y_{ob} - Y_{pred})^2 = 1.8961$ $\sum(Y_{obs} - Y_{train})^2 = 4.8401$ $\therefore R^2 \text{ test} = 1 - (1.8961/4.8401) = 0.60826$

R^2-based metrics could be a bit misleading, and mean absolute error (MAE)–based metrics are known to be straightforward determinant for predicting errors. MAE values are kept constant, and the effectives of the calculated activity is known if 5% of the residual values are absent [12]. It is expressed as:

$$MAE = \frac{1}{n} \, x\sum\{Y_{obs} - Y_{pred}\}$$

Modeling assessment

The mathematical equation (model) produced is made to undertake statistical test like cross-validated test, R^2 Fisher's test, and R^2 predicted.

Applicability domain

A model validation should be within the training domain and it is 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 [15]. It is formulated as:

$$L_i = x_i \left(X^T X\right)^{-k} x_i{}^T \ (i = K\cdots, P)$$

where X^T is the matrix transpose of X used in constructing the equation, X_i is the matrix of prediction sets of I, and X is the n x k matrix of train set descriptors. (E^*) is the warning leverage; it is a predictive tool that tests for outliers. It is written as:

$$E^* = \frac{3(h+1)}{m}$$

h stands for the total structural parameters and m is the total compounds of train sets. William's plot is a plot of standardized values vs. leverage values of both the training (calibration) and test (validation) sets. Molecules that stay within the calculated H^* on the graph are the calculated compounds.

Drug likeness analysis

The ADME properties of a molecule are an important determinant of its therapeutic potency. ADME and bio-availability test play an important role in the drug likeness of new drug molecules [17]. In this research, SwissADME was employed in evaluating the physicochemical properties, pharmacokinetics, and drug similarity of the designed compounds. Furthermore, the designed compounds were checked to ensure compliance with five rules of Lipinski [6].

Table 4 The bio-activities (pIC$_{50}$), calculated activities, and residual values of model 4

No/S.	IC$_{50}$	logIC$_{50}$	Residual
1	0.017	7.7696	0.2090
2	0.073	7.1367	−0.1669
3*	0.040	7.3979	0.1707
4	0.173	6.7610	−0.7543
5	0.038	7.4202	0.4821
6*	0.323	6.4908	0.2527
7	0.230	6.6383	−0.0807
8	0.430	6.3665	−0.3466
9	0.091	7.0400	0.2780
10*	0.140	6.8539	−0.4426
11*	0.120	6.9208	0.0284
12*	0.270	6.5686	−0.1102
13	0.651	6.1864	−0.1520
14	0.818	6.0872	−0.2899
15	0.603	6.2197	−0.1724
16	1.120	5.9508	−0.0207
17	0.088	7.0555	0.0709
18*	0.284	6.5467	−0.3170
19	0.648	6.1884	−0.0468
20	0.208	6.6819	0.4945
21	0.421	6.3757	0.2257
22*	0.885	6.0531	−0.2171
23*	0.215	6.6676	−0.3917
24	1.354	5.8684	0.0072
25	0.462	6.3354	0.0373
26*	2.235	5.6507	0.6925
27*	0.433	6.3635	−0.0412
28*	0.747	6.1267	−0.2302
29	0.328	6.4841	−0.2059
31*	0.022	7.6576	−0.3782
32	0.140	6.8539	0.0197
33*	0.060	7.2218	−0.6848
34	0.821	6.0857	−0.0899
35	0.317	6.4989	0.03045
36*	0.022	7.6577	−0.3627
37	0.012	7.9208	0.2397
38	0.044	7.3565	−0.1076
39	0.100	7.0000	0.1606
40	0.005	8.3010	−0.0981
41	0.483	6.3161	−0.0707
42*	0.141	6.8508	0.1217
43	0.004	8.3979	0.1478
44	0.015	7.8239	0.3198

The compounds with an asterisk (*) are called the validation sets

Table 5 Description of model parameters and its classes for Eq. 4

Name	Definition	Class	Mean effect
ALogP2	Square of AlogP.	2D	0.1436
ATSC0i	Centered Broto-Moreau autocorrelation—lag 0/weighted by first ionization potential	2D	0.5085
MATS5i	Moran autocorrelation—lag 5/ weighted by first ionization potential	2D	0.1443
MATS1s	Moran autocorrelation—lag 1/weighted by I-state	2D	−0.3687
ETA_Beta_ns_d	A measure of ion electrons entering in resonance	2D	0.5724

Result and discussion

QSAR investigation

QSAR analysis was used in building a simple mathematical equation in calculating an enhanced biological activity from β-lactam derivatives. The QSAR analysis also correlated the molecular descriptors (model parameters) with the physicochemical properties of the 43 derivative compounds (bioactivities) using statistical methods. Based on the genetic function approximation (GFA) technique employed, 4 mathematical equations were produced to predict the biological activities of β-lactam derivatives. Model 4 (four) passed both internal and external validation with a correlation coefficient squared (R^2) of 0.8706, correlation coefficient adjusted squared (R^2_{adj}) of 0.8411, and cross-validation coefficient (Q^2) of 0.7844. The external validation of R^2_{pred} of 0.6083 for model 4 was calculated using the model descriptors from the test set as shown in Tables 2 and 3; the MAE was found to be close to zero which reconfirms the strength of the equation (model) [12]. The effectiveness of the equations was measured by the reliability of the calibration set and calculated pIC_{50} of the validation set, which agrees with the criteria proposed by Golbraikh and Tropsha ($R^2_{pred} > 0.6$) for a robust equation as seen in Table 6.

$$Y = -0.619650812*ALogp2 + 11.748891707*MATS5i - 6.583353004*GATS5e + 8.783141246*GATS1s + 0.245270016*nHBAcc + 9.277548 \tag{1}$$

$$Y = -0.683762136*ALogp2 + 8.522074982*MATS5i + 8.041720811*GATS1s - 0.257469582*CrippenLogP - 0.418234209*ETA\_Beta\_ns\_d + 4.354169 \tag{2}$$

$$Y = -0.570977806*ALogp2 + 10.984779979*MATS5i + 9.660965568*GATS1s - 1.970549610*GATS5s - 0.532187476*ETA\_Beta\_ns\_d + 5.00093 \tag{3}$$

$$Y = -0.708666994*ALogp2 - 0.014820037*ATSC0i + 7.327733722*MATS5i - 19.009577528*MATS1s - 0.526316186*ETA\_Beta\_ns\_d + 9.042494 \tag{4}$$

Table 6 Statistical analysis of model 1 parameters

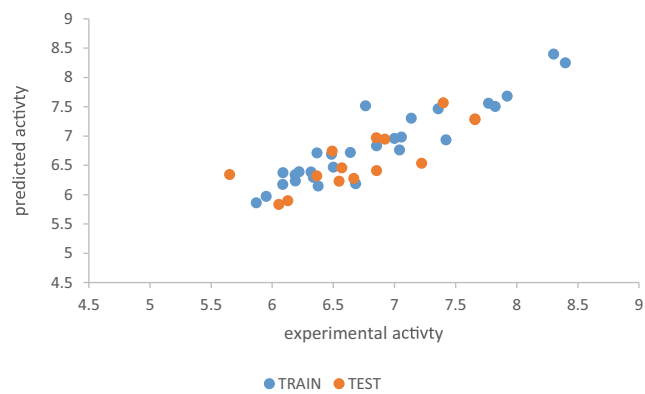|	ALogP2	ATSC0i	MATS5i	MATS1s	ETA_Beta_ns_d	VIF	P value
ALogP2	1					1.1421	4.61E−06
ATSC0i	0.1769	1				2.1360	0.0504
MATS5i	0.1416	0.7186	1			2.2741	7.3E−07
MATS1s	0.2759	0.0548	0.2166	1		1.2296	9.37E−09
ETA_Beta_ns_d	−0.0574	0.2359	0.2923	0.2571	1	1.1511	6.94E−05

Fig. 2 Plot of predicted activities versus inhibition concentration
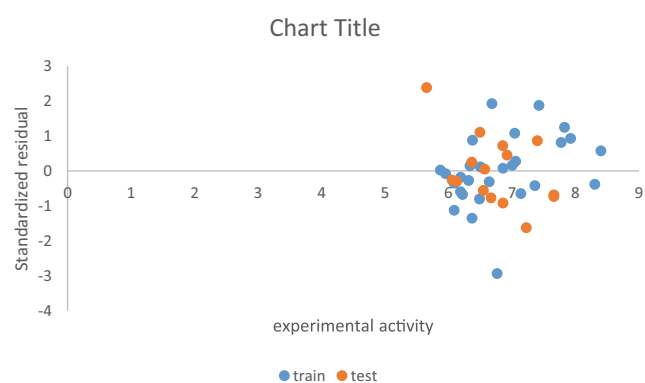

Fig. 3 A graph of standardized residual against bio-activities (experimental activities)

The effectiveness of the equation was measured by the value of the calibration set and calculated pIC_{50} of the validation set. The observed, calculated, and residual values of β-lactam compounds are seen in Table 4. The low residual value
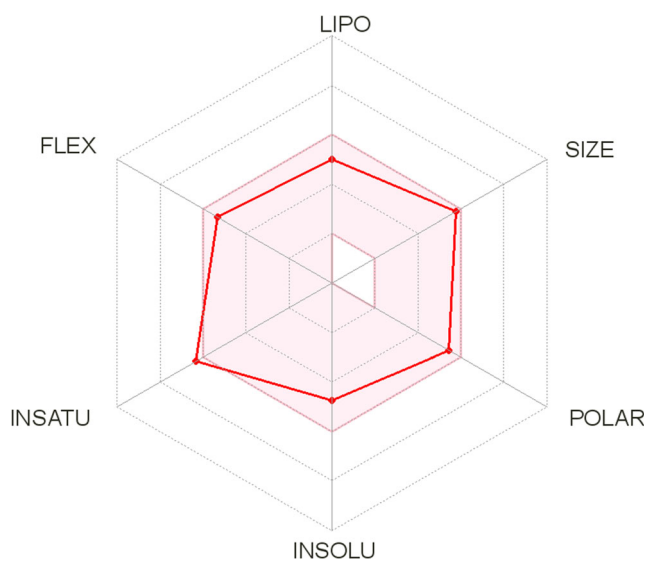

Fig. 4 A plot of leverages versus standard residual (William's plot)

Table 7 Newly designed β-lactam derivative compounds with their new inhibition concentration (pIC_{50})

S/No.	Structures	New predicted activity (pIC_{50})
1		9.6702
2		9.5546
3		10.6637
4		9.8178
5		9.0060
6		10.3570

from the difference between the biological activities and calculated activities displays the effectiveness of the equation.

Table 8 Pharmacokinetic test of the designed β-lactam compounds

S/No.	MW (g/mol)	nAH	nRB	HBA	HBD	MR	TPSA (Å²)	iLOGP	BBB	PAINS	Brenk
1	437.83	12	7	8	2	110.06	114.76	2.69	No	0	0
2	436.84	12	7	7	2	111.20	120.55	2.41	No	0	0
3	437.83	12	7	8	2	110.06	114.76	2.51	No	0	0
4	479.83	18	8	8	2	130.40	114.76	3.54	No	0	0
5	479.48	18	8	8	2	130.40	114.76	2.88	No	0	0
6	478.49	18	8	7	2	131.54	120.55	3.08	No	0	0

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

Both internal and external validations confirm model 4 to be very potent and extremely effective.

Table 5 shows the definition of the descriptors that were used in building the mathematical model; the descriptors were used in validating the model both internally and externally.

The mean effect of the mathematical model was executed statistically to evaluate the contribution of each model parameter individually. From the coefficient of the mean effect values, ALogP2, ATSC0i, MATS5i, and MATS1 had a positive coefficient, meaning that increasing the model parameters would increase the biological activities of the derivatives. Furthermore, ETA_Beta_ns_d having a negative coefficient means that decreasing the model parameter would also increase the biological activities of the derivative compounds as proven in Table 6. Variance inflation factor (VIF) gives a degree of the inter-relationship among the model parameters. The VIF scores were within the approved value of 1–5, indicating that there is no co-linearity between the bio-activities and model parameters of the derived model, as shown in Table 6. The null hypothesis shows no significant connection amid the bio-activity and model parameters of the derived equation at $p > 0.05$. At a 95% confidence level, the P values of the model parameters were below 0.05. Therefore, the null hypothesis is rejected and the alternative hypothesis is accepted. This indicates that there is no co-linearity between the bio-activity and model parameters of the constructed model, as shown in Table 6.

Figure 2 shows a plot of observed activities against the calculated activities of both the test set and the train set of β-lactam derivatives. The graph 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 built model.

Figure 3 is a graph of standardized against experimental activity; from the plot, it is shown that 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 biological activity (experimental activity).

Figure 4 shows a diagram of standardized residual against leverage values also called William's plot. All the compounds fell within the applicability domain from the calculated leverage of $L = 0.6429$, thou 1 compound was outside the applicability domain which might be due to a slight change in the molecular structure as compared with the remaining molecules in the data set.

Ligand-based drug design

Six (6) new β-lactam derivative compounds were designed using the ligand-based approach. This approach uses the molecular descriptors obtained from the mathematical QSAR model; adjustments were made on the lead compounds (37 and 43) based on the definition of the molecular descriptors (ATSC0i and MATS55i) as shown in Table 5. The newly

Fig. 5 The bio-availability radar is for compounds 37 and 43
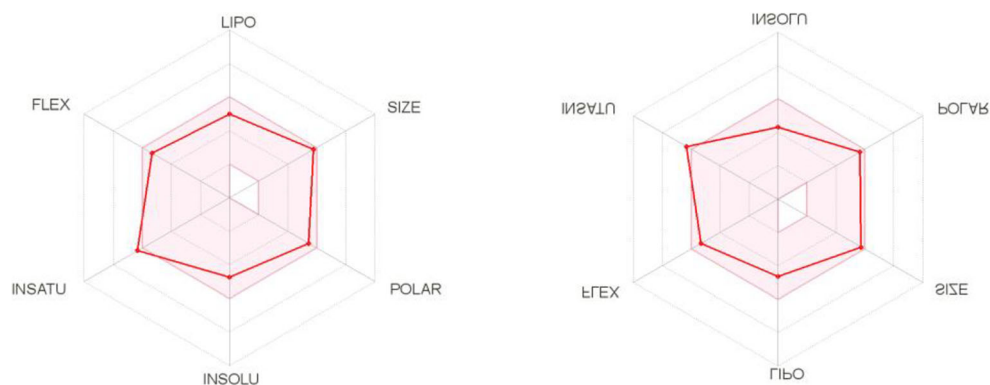
designed compounds with their new calculated activities are seen visually in Table 7.

Computational pharmacokinetics of the designed compounds

The physicochemical properties of the designed derivatives were explored for its drug-like properties. Compounds 37 and 43 revealed the characteristics of the effective drug-like template compounds as seen in Table 8. The compounds showed no violation of Lipinski's rule of five, high GI absorption, 0.55 oral bio-availability score, and zero PAINS alerts + (pain-assay interference structural alerts) indorsing its dependability for further clinical trials. The bio-availability radar is for compounds 37 and 43 and is shown in Fig. 5; it gives a quick glance at the pharmacokinetic properties of the structures.

Conclusion

QSAR and pharmacokinetics analysis carried out on the β-lactam derivatives proved the derivative compounds to be standard anti-breast cancer agents against MCF-7 cell line. The effectiveness of the generated QSAR model was assessed using internal and external validation test; the model conformed to the minimum approved values, indicating the equation could be used in designing new β-lactam derivative compounds with enhanced anti-cancer activities. The statistical analysis carried out on the QSAR model showed that ALogP2, ATSC0i, MATS5i, and MATS1s had a positive coefficient, meaning that increasing the model parameters would increase the biological activities of the β-lactam derivative compounds, while ETA_Beta_ns_d having a negative coefficient means decreasing the model parameter would also increase the biological activities of the derivative compounds. Compounds 37 and 43 were chosen as template compounds in designing 6 new derivative compounds because they had higher predicted activity and low residual values. The molecular descriptors (MATS5i and MATS1s) had more significance, and based on their mean effect, adjustments were made on the fragments of the template compounds.

Furthermore, the pharmacokinetic analysis (drug likeliness test) carried out on the newly designed compounds revealed that all the compounds passed the drug likeness test (ADME and other physicochemical properties) and they also had zero violation to Lipinski rule of five: a standard measure used in assessing the drug likeness of molecules. This concludes that the compounds can move on to the next step of pre-clinical trial, proving a tremendous discovery for medicine in finding permanent solutions to estrogen-positive breast cancer (MCF-7 cell line).

Compliance with ethical standards

References

1. Abdulrahman HL, Uzairu A, Uba S (2020) Computational pharmacokinetic analysis on some newly designed 2-anilinopyrimidine derivative compounds as anti-triple-negative breast cancer drug compounds. Bull Natl Res Centre 44:1–8

2. Ahmed EY, Latif NAA, El-Mansy MF, Elserwy WS, Abdelhafez OM (2020) VEGFR-2 inhibiting effect and molecular modeling of newly synthesized coumarin derivatives as anti-breast cancer agents. Bioorg Med Chem 28(5):115328

3. Dearden JC (2017) The history and development of quantitative structure-activity relationships (QSARs). Oncology: breakthroughs in research and practice. IGI Global, Pennsylvania, pp 67–117

4. Friedman JH (1991) Multivariate adaptive regression splines. Ann Stat:1–67

5. Gramatica P (2020) Principles of QSAR modeling: comments and suggestions from personal experience. Int J Quant Structure-Property Relationsh (IJQSPR) 5(3):1–37

6. Hou Y, Zhu L, Li Z, Shen Q, Xu Q, Li W et al (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

7. Ibrahim MT, Uzairu A, Uba S, Shallangwa GA (2020) Computational modeling of novel quinazoline derivatives as potent epidermal growth factor receptor inhibitors. Heliyon 6(2):e03289

8. Kennard RW, Stone LA (1969) Computer aided design of experiments. Technometrics 11:137–148

9. Khaled KF, Abdel-shafi NS (2011) Quantitative structure and activity relationship modeling study of corrosion inhibitors: genetic function approximation and molecular dynamics simulation methods. Int J Electrochem Sci 6:4077–4094

10. Malebari AM, Fayne D, Nathwani SM, O'Connell F, Noorani S, Twamley B et al (2020) β-Lactams with antiproliferative and antiapoptotic activity in breast and chemoresistant colon cancer cells. Eur J Med Chem 189:112050

11. 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). In:

Materials science forum (Vol. 948, pp. 101-108). Trans Tech Publications

12. Roy K, Das RN, Ambure P, Aher RB (2016) Be aware of error measures. Further studies on validation of predictive QSAR models. Chemom Intell Lab Syst 152:18–33

13. Sehgal SA, Mirza AH, Tahir RA, Mir A (2018) Quick guideline for computational drug design. Bentham Science Publishers, Sharjah

14. Tantawy ES, Amer AM, Mohamed EK, Abd Alla MM, Nafie MS (2020) Synthesis, characterization of some pyrazine derivatives as anti-cancer agents: in vitro and in Silico approaches. J Mol Struct 1210:128013

15. 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

16. Yap CW (2011) PaDEL-descriptor: an open source software to calculate molecular descriptors and fingerprints. J Comput Chem 32:1466–1474

17. Zhang C, Li Q, Meng L, Ren Y (2020) Design of novel dopamine D2 and serotonin 5-HT2A receptors dual antagonists toward schizophrenia: an integrated study with QSAR, molecular docking, virtual screening and molecular dynamics simulations. J Biomol Struct Dyn 38(3):860–885