

RESEARCH

Open Access



QSAR, molecular docking studies, ligand-based design and pharmacokinetic analysis on Maternal Embryonic Leucine Zipper Kinase (MELK) inhibitors as potential anti-triple-negative breast cancer (MDA-MB-231 cell line) drug compounds

Hadiza Abdulrahman Lawal^{*} , Adamu Uzairu and Sani Uba

Abstract

Background: Cancer of the breast is known to be among the top spreading diseases on the globe. Triple-negative breast cancer is painstaking the most destructive type of mammary tumor because it spreads faster to other parts of the body, with high chances of early relapse and mortality. This research would aim at utilizing computational methods like quantitative structure–activity relationship (QSAR), performing molecular docking studies and again to further design new effective molecules using the QSAR model parameters and to analyze the pharmacokinetics “drug-likeness” properties of the new compounds before they could proceed to pre-clinical trials.

Results: The QSAR model of the derivatives was highly robust as it also conforms to the least minimum requirement for QSAR model from the statistical assessments of (R^2) = 0.6715, (R^2_{adj}) = 0.61920, (Q^2) = 0.5460 and (R^2_{pred}) of 0.5304, and the model parameters (AATS6i and VR1_Dze) were used in designing new derivative compounds with higher potency. The molecular docking studies between the derivative compounds and Maternal Embryonic Leucine Zipper Kinase (MELK) protein target revealed that ligand 2, 9 and 17 had the highest binding affinities of – 9.3, – 9.3 and – 8.9 kcal/mol which was found to be higher than the standard drug adriamycin with – 7.8 kcal/mol. The pharmacokinetics analysis carried out on the newly designed compounds revealed that all the compounds passed the drug-likeness test and also the Lipinski rule of five.

Conclusions: The results obtained from the QSAR mathematical model of parthenolide derivatives were used in designing new derivatives compounds that were more effective and potent. The molecular docking result of parthenolide derivatives showed that compounds 2, 9 and 17 had higher docking scores than the standard drug adriamycin. The compounds would serve as the most promising inhibitors (MELK). Furthermore, the pharmacokinetics analysis 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 adhered to the Lipinski rule of five. This gives a great breakthrough in medicine in finding the cure to triple-negative breast cancer (MBA-MD-231 cell line).

*Correspondence: azeezalawal@gmail.com
Department of Chemistry Ahmadu, Bello University, P.M.B. 1044, Zaria,
Nigeria

Keywords: QSAR, Molecular docking, Ligand-based design, Pharmacokinetic analysis

Background

Cancer has long been a challenging malady to the human race with great occurrence and death rates. Nearly 8 million humans pass away from tumors yearly, and 14.2 million different cancer patients are treated worldwide yearly (Xu et al. 2019). Cancer of the breast is the most reoccurring type of disease and the second leading cause of death among women globally. Yearly, approximately 1–1.3 million cases of breast cancer are detected globally (Ge et al. 2019). The high occurrence and death rate of cancer occur from the fact that there are over 200 kinds of cancer and it is indeed hard to diagnose at a premature stage (Lolak et al. 2019).

Triple-negative breast cancer (TNBC) is painstaking the most destructive type of mammary tumor because

it spreads faster to other parts of the body, with high chances of early relapse and mortality (Hu et al. 2012). Yearly, an estimation of over a million female beings is detected with mammary tumor and TNBC is responsible for almost 15–20% of the overall breast cancer detected (Jo et al. 2019). TNBC does not express estrogen receptor (ER), progesterone receptor and human epidermal growth factor 2 (HER2) (Hu et al. 2012).

Early detection, understanding of the cause and pathway of this disease and advancing in treatment have played a key role in decreasing breast cancer death rates during the past few years. Chemotherapy remains the main key to complete therapy since it would extend and terminate tumor cells faster within the human system (Kaplan 2013). Computational methods of drug discovery

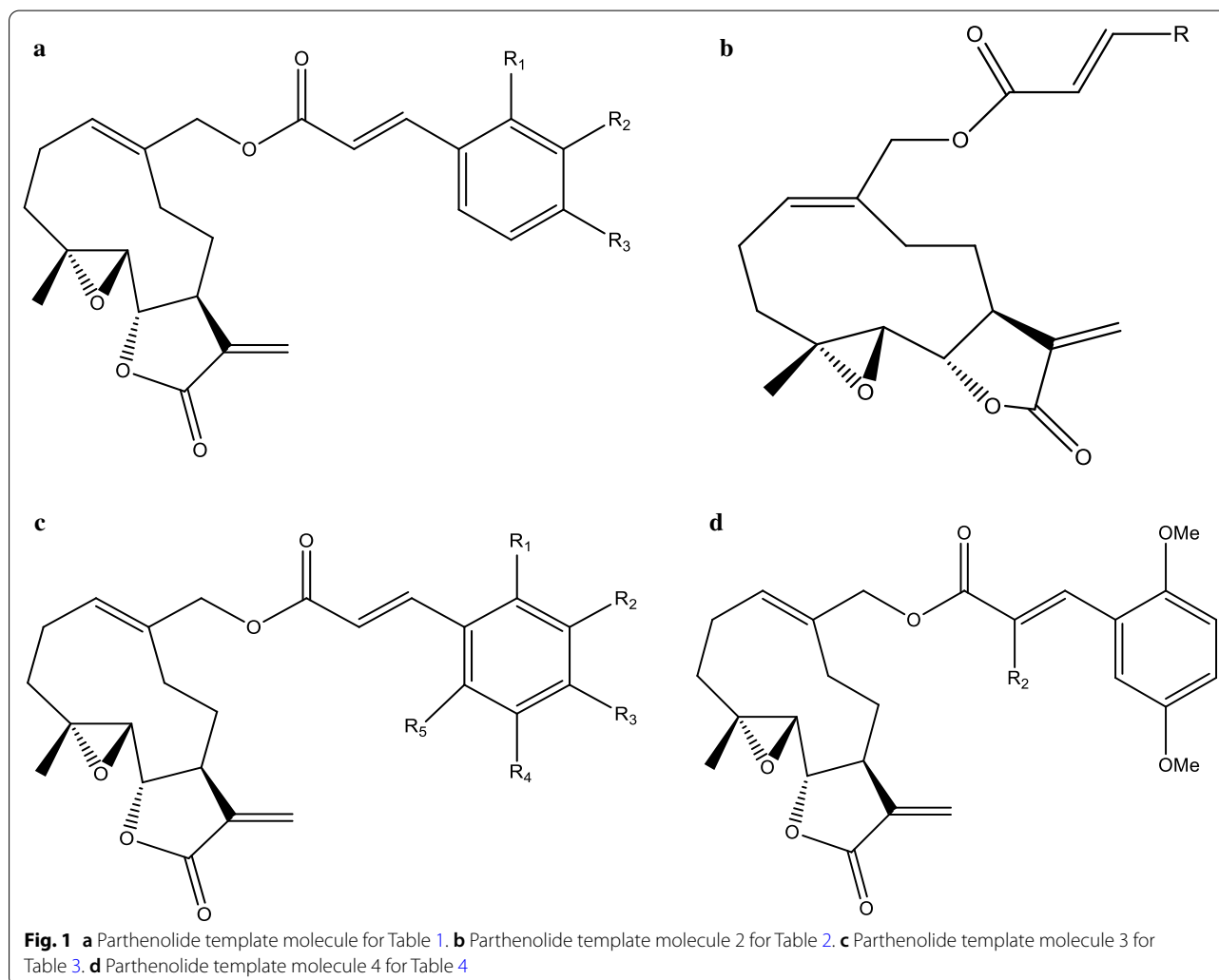


Table 1 The inhibitory concentration of parthenolide derivatives for template molecule 1

No.	R ₁	R ₂	R ₃	Observed activities (μM)	Calculated activities pIC ₅₀
1	H	H	H	1.35	5.8697
2	CF ₃	H	H	0.74	6.1308
3	F	H	H	1.80	5.7447
4	Cl	H	H	1.42	5.8477
5	Br	H	H	1.33	5.8761
6	Me	H	H	1.79	5.7471
7	OMe	H	H	0.46	6.3372
8	OEt	H	H	0.34	6.4685
9	H	CF ₃	H	0.83	6.0809
10	H	F	H	1.73	5.7610
11	H	Cl	H	1.78	5.7496
12	H	Br	H	1.64	5.7852
13	H	Me	H	1.08	5.9666
14	H	OMe	H	1.84	5.7352
15	H	H	CF ₃	2.52	5.5986
16	H	H	F	1.32	5.8794
17	H	H	Cl	1.22	5.9136
18	H	H	Br	1.02	5.9914
19	H	H	Me	2.27	5.6430
20	H	H	i-Pr	1.33	5.8761
21	H	H	OMe	1.73	5.7610
22	H	H	OEt	1.58	5.8013
23	H	H	OAc	1.15	5.9393
24	H	H	NO ₂	1.93	5.7144
25	H	H	CN	1.80	5.7447
26	H	H	OH	1.89	5.7235
27	H	OMe	OH	2.01	5.6968

have proven to be faster than traditional methods of drug discovery, and it saves more time, resources and also tends to have higher efficacy with less toxicity. Efficiency and safety of the drug to the system are the two major causes leading to drug failure. Therefore, it is compulsory to find potent molecules with better ADMET properties “drug-likeness” (Guan et al. 2019).

Novel 61 series of parthenolide derivatives were obtained from the literature of (Ge et al. 2019) as inhibitors against the MDA-MB-231 cell line. Edupuganti et al. 2017 suggest that MELK with Protein Data Bank (PDB: 4BKY) has therapeutic value for targeting various cancers, especially the critical survival functions like TNBCs.

The purpose of this research is to utilize computational methods like quantitative structure–activity relationship (QSAR), in calculating the activities of parthenolide derivative compounds as inhibitors against breast cancer cell line MDA-MB-231 based on an established QSAR mathematical model, molecular docking studies

in understanding how the ligands (parthenolide compounds) interrelate with a protein receptor Maternal Embryonic Leucine Zipper Kinase (MELK), again, to further design new effective molecules using the QSAR model parameters and to analyze the pharmacokinetics “drug-likeness” properties of the new compounds.

Methods

QSAR studies

Data set

In total, 61 derivative compounds of parthenolide were obtained from (Ge et al. 2019) article.

Anti-proliferative activities

Their bioactivities were changed to (pIC₅₀) using the formula below. It was measured in inhibitory concentration (IC₅₀) values in the micromolar concentration of (μM) (Abdulrahman et al. 2020a, b, c). The parent compounds as shown in Fig. 1a–d are the structural derivatives of parthenolide. Tables 1, 2, 3, 4 and 5 show the various substituents measured in (IC₅₀) and (pIC₅₀) that were attached to the parent compound.

$$pIC_{50} = -\log_{10} (IC_{50} \times 10^{-6}).$$

Structural optimization

The data set was drawn on software (V12.0.2) of Chem-Draw and was transferred to Spartan 14 software (V1.1.4) for structural optimization, setting up the parameters density functional theory (DFT) at B3LYP, 6-31G/ basis set (Abdulrahman et al. 2020a, b, c).

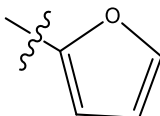
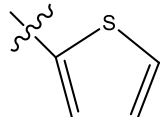
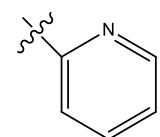
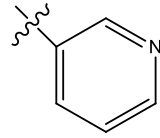
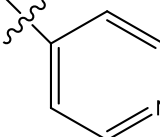
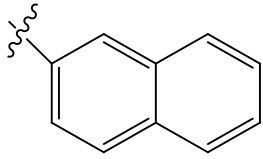
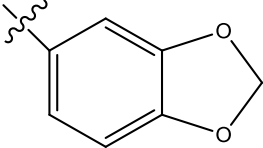
Molecular descriptor calculations

In total, 61 derivative compounds of parthenolide were converted to SDF format after optimization. Pharmaceutical Data Exploration Laboratory Software V (2.20) was used in calculating physicochemical descriptors (Yap et al. 2011). The descriptors were pretreated using Data Pre-treatment software GUI 1.2. (Arthur et al. 2020; Abdulrahman et al. 2020a, b, c) to remove irrelevant values.

Training set and test set division

For a valid QSAR model to be obtained, the pretreated set must be divided into train and test set. The division should satisfy the following conditions: (i) it is essential for all the train set descriptive points to be evenly shared within the entire area used by the data set (Noolvi and Patel 2013). (ii) All descriptor values should be near the train set descriptor values. (iii) All characteristic molecular elements in the test set should spread within the

Table 2 The inhibitory activities of parthenolide derivatives for template molecule 2

No.	R	Observed Activities IC ₅₀ (μM)	Calculated Activities pIC ₅₀
28		2.25	5.6478
29		1.80	5.7447
30		2.29	5.6402
31		3.27	5.4855
32		2.79	5.5544
33		1.48	5.8297
34		1.80	5.7447

descriptor space and should also be close to the train set (Noolvi and Patel 2013). The algorithm of Kennard–Stone was employed in dividing the set into 30% test and 70% train set (Kennard and Stone 1969).

Model building

Multiple linear regression technique was used in constructing a model in Version 8 of Material studio software. It was used to show the correlation between the dependent variables (pIC₅₀) and the independent variables as the 2D model descriptors (Abdulrahman et al. 2020a, b, c). The model is suited so that the total squared difference between the actual and predicted value of the set of anti-proliferative activity is diminished. In regression analysis, the reliant mean of the dependent variable,

(pIC₅₀), relies on the independent descriptors (Abdullahi et al. 2020).

Validating the model (internal)

The prediction of the model built must be verified on a data set that was never used in building the model at first (Tropsha et al. 2003). The models obtained from the internal validation of the train test (sixty-one compounds) were evaluated using Friedman formula (Friedman 1991).

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

Table 3 The inhibitory activity of parthenolide derivatives for template molecule 3

No.	R ₁	R ₂	R ₃	R ₄	R ₅	IC ₅₀ (μM)	pIC ₅₀
35	OMe	OMe	H	H	H	1.74	5.7595
36	OMe	H	OMe	H	H	0.55	6.2596
37	OMe	H	H	OMe	H	0.26	6.5850
38	OMe	H	H	H	OMe	0.25	6.6021
39	H	OMe	OMe	H	H	1.95	5.7090
40	H	OMe	H	OMe	H	2.02	5.6946
41	OMe	OMe	OMe	H	H	1.77	5.7520
42	OMe	H	OMe	OMe	H	1.02	5.9914
43	OMe	H	OMe	H	OMe	1.09	5.9626
44	H	OMe	OMe	OMe	H	0.95	6.0223
45	F	H	H	H	F	1.59	5.7986
46	Cl	H	H	H	Cl	1.59	5.7986
47	Br	H	H	H	Br	1.50	5.8239

d equals user-defined smoothing parameter, C equals the sum of model definitions, M equals the sum of derivatives on the train set, while p is the summation of model definition (Abdulrahman et al. 2020a, b, c). SEE is the standard estimated error. The smaller SEE is, the more enhanced the model would be.

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

The model from the regression structure becomes.

$$Y = k_1x_1 + k_2x_2 + k_3x_3 + n.$$

Y is the response variable, 'k' is the coefficient of regression corresponding to 'x's that are the model parameters which are the predictor variables, and n is the regression constant.

(R²) is the regression coefficient, and it is the degree of fitness of the equation of regression. It signifies the variation section in the data that is described by regression. The nearer R² value is 0. 1, the better the regression fitness. R² is given as:

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

where Y_{exp} and Y_{pred} are the biological and calculated activities of the train set. Y_{mintraining} indicates the average pIC₅₀ of the train set molecules. Table 2 shows the accepted least required values in assessing a QSAR model (Ibrahim et al. 2018). (R²_{pred}) is the predicted validation parameter, and it is assessed by calculating the prediction power of the model. An (R²_{pred}) above 0.5 shows a high prediction power and robustness of the model.

Applicability domain of QSAR

Model validation should be within the training domain, and the performance of the compounds needs to be assessed within the domain to determine the fitness of the model. The applicability domain is evaluated using William's plot. The warning leverage is used in sorting the compounds within a particular space on the graph known as the applicability domain, compounds that fall within the space on the plot are the approved predicted compounds (Veerasingam et al. 2011). It is formulated as;

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

where j equals the total model parameters and m is the total molecules of train sets.

Molecular docking studies

Some compounds with high pIC₅₀ were chosen for molecular docking studies with Maternal Embryonic Leucine Zipper Kinase (MELK) as protein target. The crystal structure was obtained from RCSB PDB (<https://www.rcsb.org>) with the ID, 4BKY. The crystal structure was prepared with Discovery Studio Software and converted to PDB format. The binding affinity of the ligand-protein complex was calculated with AutoDock Vina of PyRX software employed in calculating (Abdulrahman et al. 2020a, b, c). Visualizer of Discovery Studio was used to understand the ligand-protein target interactions.

Computational pharmacokinetics (drug-likeness)

The SwissADME, a free web tool used in evaluating the pharmacokinetics, drug-likeness (physicochemical and ADME properties) and medicinal chemistry friendliness of small molecules (Daina et al. 2017), would be used in testing the drug-likeness of the newly designed compounds.

Table 4 The inhibitory activity of parthenolide derivatives for template molecule 4

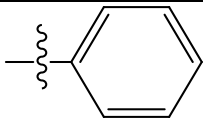
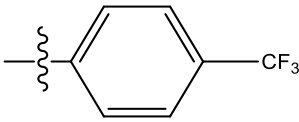
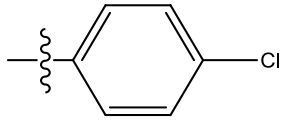
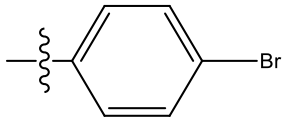
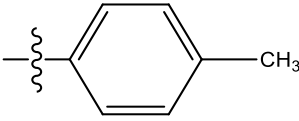
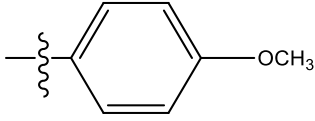
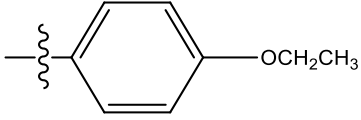
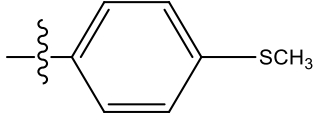
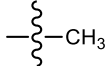
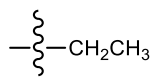
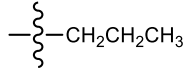
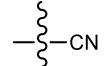
No.	R ₂	Actual Activities (IC ₅₀)	Calculated Activities (pIC ₅₀)
48		2.01	5.6968
49		1.98	5.7033
50		2.02	5.6946
51		2.01	5.6968
52		2.95	5.5302
53		2.51	5.6003
54		1.87	5.7282
55		1.87	5.7282
56		1.76	5.7545
57		2.25	5.6478
58		7.53	5.1232
59		2.35	5.6289

Table 5 The inhibitory activities of parthenolide derivatives

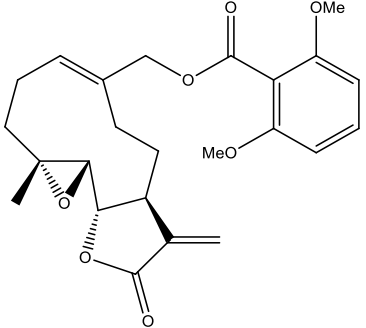
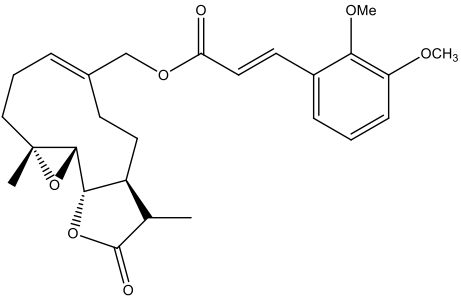
No.	Structures	Actual Activities (IC ₅₀)	Calculated Activities (pIC ₅₀)
60		3.55	5.4498
61		7.60	5.1192

Table 6 Suggested values used in the assessment of QSAR model

Characters	Names	Values
R ²	Coefficient of determination	≥ 0.6
P _(95%)	Confidence interval at 95% confidence level	< 0.05
Q ²	Squared cross-validation coefficient	≥ 0.5
R ² -Q ²	Difference between R ² and Q ²	< 0.3
N _{test set}	Least number of the test set	≥ 5
R ² _{ext}	Coefficient of determination of external validation	≥ 0.5

Furthermore, some physicochemical properties and positive controls of the designed compounds were checked using the on-line tool for their adaptability with Lipinski's rule of five (Hou et al. 2019). Lipinski and co-workers proposed the "Rule of Five" in 1997, which was the original and most known rule-based filter for drug-likeness of a molecule, distinguishing whether a molecule can be orally absorbed well or not, following the criteria: molecular weight (MW) ≤ 500, octanol/water partition coefficient (AlogP) ≤ 5, number of hydrogen bond donors (HBD's) ≤ 5 and number of hydrogen bond acceptors (HBAs) ≤ 10.6. According to the Rule of Five, a molecule would not be

orally active if it violates two or more of the four rules (Guan et al. 2019).

Results

QSAR of parthenolide derivatives

The goal of QSAR is to establish a model from the obtained descriptor that has a higher performance than the experimental values. In this research, parthenolide derivatives went through a quantitative structure–activity relationship with its actual activities.

The Model

$$\begin{aligned}
 pIC_{50} = & 0.112976697 * (AATS6i) \\
 & + 0.010871443 * (ATSC8i) \\
 & + 1.482833475 * (GATS5s) \\
 & + 0.000807496 * (VR1_Dze) \\
 & - 0.303714684 * (C2SP3) - 12.639784038.
 \end{aligned}$$

The model best predicts the biological activities of parthenolide derivatives against the MDA-MB-231 cell line. The model also conforms to the least proposed requirement in QSAR modeling as indicated in Table 2.

Tables 7 and 8 reveal how the model was validated externally by carrying out calculations using the values of the descriptors obtained from the test set. The

Table 7 Model parameters used in validating the model externally

Name	pIC ₅₀	AATS6i	ATSC8i	GATS5s	VR1_Dze	C2SP3	Y _{Pred}
1	5.8697	161.5169	48.7326	0.8808	292.3360	6	5.8575
20	5.8761	162.0662	40.5484	0.8599	336.7858	6	5.8355
25	5.7447	161.1486	43.7089	0.8297	443.9988	6	5.8080
26	5.7235	161.4648	47.3323	0.7449	293.8548	6	5.6361
4	5.8477	161.4031	48.9774	0.8945	288.8093	6	5.8647
41	5.7520	161.6404	55.2788	0.9363	360.7402	6	6.0801
42	5.9914	161.9008	47.3935	0.8602	358.8289	6	5.9095
43	5.9626	162.5176	82.5586	0.8348	356.6440	6	6.3221
48	5.6968	158.9117	24.3199	1.0608	405.1766	6	5.6558
5	5.8761	161.1950	49.3864	0.8916	288.7227	6	5.8413
52	5.5302	158.5990	38.7462	1.0551	405.2722	6	5.7680
53	5.6003	159.3748	13.5155	1.0230	410.7551	6	5.5391
57	5.6478	160.7295	22.3043	1.0283	421.4272	7	5.5005
58	5.1232	161.2564	29.4073	1.0188	398.3031	8	5.3008
6	5.7471	161.839	42.3780	0.8672	288.5304	6	5.8016
60	5.4498	163.1298	37.0344	0.8237	307.2417	7	5.5363
16	5.8794	162.1464	52.3900	0.6452	289.6975	6	5.6169
8	6.4685	161.2731	51.2615	0.8828	536.4764	6	6.0575

Table 8 Continuation of the external validation of the model

Y _{pred} - Y _{obs}	(Y _{pred} - Y _{obs}) ²	Y _{mintrin}	(Y _{mintrin} - Y _{obs})	(Y _{mintrin} - Y _{obs}) ²
- 0.0122	0.0001	5.8285	- 0.0412	0.0017
- 0.0406	0.0017	5.8285	- 0.0477	0.0023
0.0633	0.0040	5.8285	0.0838	0.0070
- 0.0875	0.0076	5.8285	0.1040	0.0110
0.0160	0.0003	5.8285	- 0.0192	0.0004
0.3281	0.1077	5.8285	0.0765	0.0058
- 0.0819	0.0067	5.8285	- 0.1629	0.0265
0.3595	0.1292	5.8285	- 0.1341	0.0170
- 0.0410	0.0017	5.8285	0.1317	0.0173
- 0.0348	0.0012	5.8285	- 0.0477	0.0023
0.2388	0.0570	5.8285	0.2983	0.0880
- 0.0612	0.0037	5.8285	0.2282	0.0521
- 0.1473	0.0217	5.8285	0.1807	0.0326
0.1776	0.0315	5.8285	0.7053	0.4974
0.0544	0.0020	5.8285	0.0814	0.0066
0.0865	0.0075	5.8285	0.3787	0.14344
- 0.2625	0.0689	5.8285	- 0.0509	0.0026
- 0.4110	0.1689	5.8285	- 0.6400	0.4096

$$\sum(Y_{obs} - Y_{pred})^2 = 0.53041 \quad \sum(Y_{obs} - \bar{Y}_{train})^2 = 1.32576 \therefore R^2_{test} = 1 - (0.53041 / 1.32576) = 0.5304$$

calculated external validation (R^2_{pred}) was 0.5304, which agrees with the least requirements as shown in Table 6.

Table 9 shows the experimental, calculated and residual values of the derivative compounds. The low residual values were calculated from the predicted activities. The low residual values show the potency of the built model.

Table 10 defines the model parameters (descriptors) used in building the model, and the descriptors were used in verifying the model both internally and externally. The mean effect was calculated statistically, and it shows the collinearity of the model parameters and the activities of the derivative compounds in the built model.

Table 11 shows correlation between the descriptors generated from the model (VIF) and the P values of the descriptors. The VIF was calculated using the equation;

$$VIF = \frac{1}{(1 - X^2)}$$

X^2 is the coefficient of correlation (Myers 1990).

The P values estimate the statistical difference between the model parameters at 95% confidence level.

A straight line graph was obtained from plotting the activities (predicted activity against experimental activity) of both the test set and train set of the derivative compounds as shown in Fig. 2. The anti-proliferative activities showed a good connection as shown on the plot.

Table 9 The bioactivities (pIC₅₀), predicted (pIC₅₀) and residual of the model

Structure	Bioactivities(pIC ₅₀)	Predicted(pIC ₅₀)	Residual
1*	5.8697	5.8575	0.0122
2	6.1308	6.0889	0.0419
3	5.7447	5.7767	-0.0310
4*	5.8477	5.8647	-0.0170
5*	5.8761	5.8413	0.0348
6	5.7471	5.8016	-0.0544
7	6.3372	5.9952	0.3421
8*	6.4685	6.0575	0.4110
9	6.0809	6.1495	-0.0686
10	5.7610	5.6714	0.0906
11	5.7496	5.8640	-0.1144
12	5.7852	5.8690	-0.0848
13	5.9666	5.9944	-0.0278
14	5.7352	5.7595	-0.0243
15	5.5986	5.4421	0.1565
16*	5.8794	5.6169	0.2625
17	5.9136	5.8413	0.0723
18	5.9914	5.8049	0.1865
19	5.6430	5.7113	-0.0674
20*	5.8761	5.8355	0.0406
21	5.7610	5.8648	-0.1020
22	5.8013	5.8464	-0.0450
23	5.9393	5.7953	0.1430
24	5.7144	5.7631	-0.0487
25*	5.7447	5.8080	-0.0633
26*	5.7235	5.6361	0.0875
27	5.6968	5.8680	-0.1722
28	5.6478	5.6913	-0.0435
29	5.7447	5.9183	-0.1736
30	5.6402	5.4665	0.1737
31	5.4855	5.5797	-0.0943
32	5.5544	5.5653	-0.0109
33	5.8297	5.8392	-0.0095
34	5.7447	5.9110	-0.1672
35*	5.7595	6.1619	-0.4025
36	6.2596	6.1104	0.1492
37	6.5850	6.4285	0.1566
38	6.6021	6.2714	0.3307
39	5.7090	5.8368	-0.1268
40	5.6946	5.7957	-0.1011
41*	5.7520	6.0801	-0.3281
42*	5.9914	5.9095	0.0819
43*	5.9626	6.3221	-0.3595
44	6.0223	5.8656	0.1568
45	5.7986	5.7778	0.0208
46	5.7986	5.8829	-0.0843
47	5.8239	5.8352	-0.0113
48*	5.6968	5.6558	0.0410
49	5.7033	5.6116	0.0917

Table 9 (continued)

Structure	Bioactivities(pIC ₅₀)	Predicted(pIC ₅₀)	Residual
50	5.6946	5.6849	0.0098
51	5.6968	5.7200	-0.0242
52*	5.5302	5.7680	-0.2388
53*	5.6003	5.5391	0.0612
54	5.7282	5.6032	0.1249
55	5.7282	5.6806	0.0476
56	5.7545	5.8740	-0.1196
57*	5.6478	5.5005	0.1473
58*	5.1232	5.3008	-0.1776
59	5.6289	5.7336	-0.1046
60*	5.4498	5.5363	-0.0865
61	5.6383	5.6725	-0.0342

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

A graph of standardized residual versus biological activity of the entire data set was plotted as shown in Fig. 3. The compounds of both the test and train set spread on both sides of the y-axis, defining the potency of the model.

The applicability domain was evaluated using the William's plot. The William's plot is a graph of standardized residual against leverages, as shown in Fig. 4. The warning leverage is used in sorting the compounds within a particular space, and the compounds that fell within the space on the plot were the actual predicted compounds. The warning leverage was found to be ($h^* = 0.47$).

Molecular docking results

Molecular docking interaction on compounds of parthenolide derivatives with the protein target, Maternal Embryonic Leucine Zipper Kinase (MELK), was performed. The molecular docking results are summarized in Table 12. The binding affinity was calculated using PyRX software, while Discovery Studio Software was used in understanding and visualizing the various interaction formed between the ligand (derivative compounds) and the binding pose of the receptor (MELK).

Figure 5 shows the prepared receptor and ligand that was used in the docking studies, Figs. 6 and 7 show the 2D interaction of the complexes 9 and 2, while Fig. 8 shows the 3D interaction of the prepared ligand 9 and 2 to form complexes.

Ligand base drug design

Fifteen (15) new compounds new parthenolide derivative compounds were designed using the ligand-based approach. The ligand-based design used the derived mathematical model obtained from the QSAR studies,

Table 10 Definition of descriptors and their classes for the model

Name	Definition	Class	Mean effect
ATSC6i	Centered Broto-Moreau autocorrelation—lag 6/weighted by first ionization potential	2D	0.9884
ATSC8i	Centered Broto-Moreau autocorrelation-lag 8/weighted by first ionization potential	2D	0.0262
GATS5s	Geary autocorrelation-lag 5/weighted by L I-state	2D	0.0695
VR1_Dze	Randic-like eigenvector-based index from Barysz matrix/weighted by Sanderson electronegativities	2D	0.0166
C2SP3	Singly bound carbon bound to two other carbons	2D	−0.1006

Table 11 Statistical analysis of the descriptor used in the QSAR model

	AATS6i	ATSC8i	GATS5s	VR1_Dze	C2SP3	VIF	P value
AATS6i	1					2.077006	0.006999
ATSC8i	0.374295	1				1.272532	0.000162
GATS5s	−0.65093	−0.3855	1			1.984035	3.01E−05
VR1_Dze	0.000492	−0.16932	0.0273	1		1.059332	0.000661
C2SP3	0.280075	0.185489	0.013492	−0.15324	1	1.226411	−4.6729

and adjustment was made on the lead compounds (1 and 13) based on the definition of the molecular descriptors obtained from the model. Table 13 shows the newly designed compounds with their new calculated activities.

Physicochemical and ADME properties (pharmacokinetics) of designed parthenolide compounds

A compound with poor drug-likeness and poor ADMET properties will not be allowed to progress into pre-clinical research, regardless of the high anti-proliferative activities. ADME properties are one of the main pipelines in drug discovery, for drug-likeness of a molecules to be assessed, and it must pass through the pipeline of drug discovery. All the newly designed compounds were assessed for their drug-likeness (pharmacokinetics analysis) as shown in Table 14.

Discussion

QSAR of parthenolide derivatives

From the statistical parameters, the squared coefficient of correlation (R^2) was 0.6715, correlation coefficient adjusted squared (R^2_{adj}) was 0.61920, cross-validation coefficient (Q^2) was 0.5460, and external validation (R^2_{pred}) was 0.5304. How the external validation of the model was verified is shown in Tables 7 and 8, which also conforms to the least requirement in QSAR modeling as shown in Table 6. The actual, calculated and residual values of parthenolide derivative compounds are shown in Table 9. The low residual value is the outcome between the difference between actual and predicted

activities showing the high performance of the model. The model certified both internal and external parameters, thereby confirming the model to have high performance, very stable and robust.

The model's parameters are defined in Table 10 (names, definition and class). The mean effect (Table 10) shows the contribution of each descriptor in the constructed model, and the positive coefficient and values of the descriptor from the mean effect showed AATS6i has more impact followed by GATS5s, ATSC8i and VR1_Dze, therefore increasing the positive effect of the descriptors would increase the biological activities of parthenolide derivatives, while C2SP3 carries a negative effect and gives the least contribution in the model, reducing its negative effect would also contribute in increasing biological activities of parthenolide derivatives as proven in Tables 7 and 8. There was no much inter-correlation between the model parameters from the statistical analysis of Variance Inflation Factor (VIF), making the model highly stable. The null hypothesis shows no significant connection amid the bioactivity and model parameters of the constructed model 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 a co-linearity between the bioactivity and model parameters of the constructed model, as shown in Table 11.

Figure 2 shows a plot of actual activities against the predicted activities of both the test set and the train set of compounds. The plot showed that the predicted

Table 12 Binding affinity, interaction type, bond type and distances in some complexes

Complex	Binding affinity (kcal/mol)	Amino acid	Bond type	Interaction	Bond Distance(A ⁰)
9	−9.3	THR19	Hydrogen Bond	Conventional Hydrogen Bond	2.74925
		GLY21	Hydrogen Bond; Halogen	Conventional Hydrogen Bond; Halogen (Fluorine)	2.4619
		ARG53	Hydrogen Bond; Halogen	Conventional Hydrogen Bond; Halogen (Fluorine)	2.98056
		CYS89	Hydrogen Bond	Conventional Hydrogen Bond	3.34925
		GLY18	Hydrogen Bond	Carbon Hydrogen Bond	3.35524
		ILE17	Hydrogen Bond	Carbon Hydrogen Bond	3.59598
		ASP150	Halogen	Halogen (Fluorine)	3.33001
		ASP150	Electrostatic	Pi-Anion	4.20988
		PHE22	Hydrophobic	Pi-Pi T-shaped	5.96743
		ILE17	Hydrophobic	Alkyl	5.00307
		VAL25	Hydrophobic	Alkyl	5.43404
		LEU139	Hydrophobic	Alkyl	4.42983
		ILE17	Hydrophobic	Alkyl	3.89836
		PHE22	Hydrophobic	Pi-Alkyl	5.15308
		TYR88	Hydrophobic	Pi-Alkyl	4.67658
		ALA23	Hydrophobic	Pi-Alkyl	4.44028
		2	−9.3	LYS134	Hydrogen Bond
GLY18	Hydrogen Bond			Carbon Hydrogen Bond	3.50504
ASP150	Hydrogen Bond			Carbon Hydrogen Bond	3.40604
THR19	Hydrogen Bond			Carbon Hydrogen Bond	3.35007
PHE22	Hydrophobic			Pi-Sigma	3.72576
ALA23	Hydrophobic			Alkyl	3.10836
ALA38	Hydrophobic			Alkyl	4.3497
ILE17	Hydrophobic			Alkyl	5.0501
VAL25	Hydrophobic			Alkyl	4.61294
LEU139	Hydrophobic			Alkyl	4.86094
VAL25	Hydrophobic			Pi-Alkyl	4.17368
ALA38	Hydrophobic			Pi-Alkyl	4.84776
LEU86	Hydrophobic			Pi-Alkyl	5.45664
ILE149	Hydrophobic			Pi-Alkyl	4.54951
17	−8.9	GLY21	Hydrogen Bond	Conventional Hydrogen Bond	2.62202
		ILE17	Hydrophobic	Pi-Sigma	3.47363
		ILE17	Hydrophobic	Alkyl	4.41825
		LEU139	Hydrophobic	Pi-Alkyl	5.25912

activity was in good agreement with its experimental values as shown in Table 6, conforming to the effectiveness and strength of the built model. Figure 3 shows how the entire molecules spread on both negative and positive sides of zero point on the y-axis of the plot, indicating no systematic errors between the standardized residual versus inhibition concentration (experimental activity). Figure 4 shows the standardized residuals against the leverage values also called William's plot. The better part of the derivatives fell within the calculated leverage of ($h=0.47$); except 8 compounds we found to be outside the warning leverage which might be due to a slight

difference in structure compared to other compounds in the data set.

Molecular docking studies

Molecular docking interaction on compounds of parthenolide derivatives with the protein target, Maternal Embryonic Leucine Zipper Kinase (MELK), was performed. Compounds 2, 9 and 17 were chosen because of their high binding score. Compounds 2, 9 and 17 had the highest docking score of − 9.3, − 9.3 and − 8.9 kcal/mol as shown in Table 12. The visual examination of docked complexes was done by assessing the hydrogen bond

Table 13 Newly designed parthenolide derivative compounds with their new predicted activities (pIC_{50})

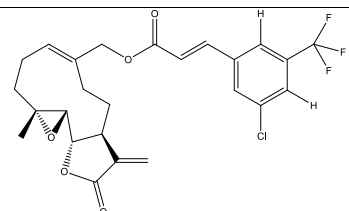
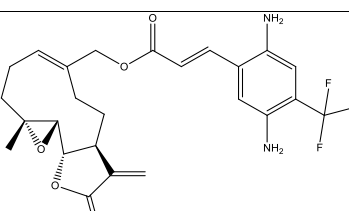
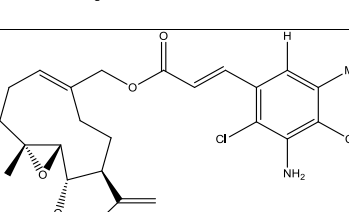
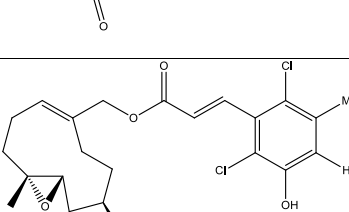
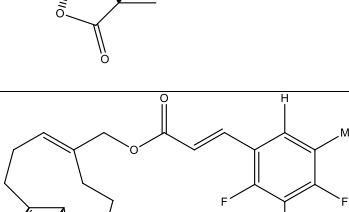
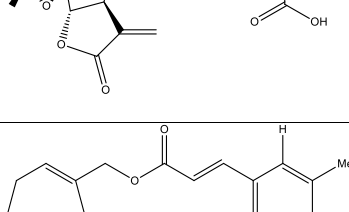
S/No.	Structures	New predicted activity (pIC_{50})
1		6.1312
2		6.0335
3		6.3248
4		6.2884
5		6.2044
6		6.2357

Table 13 (continued)

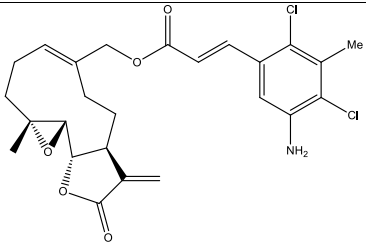
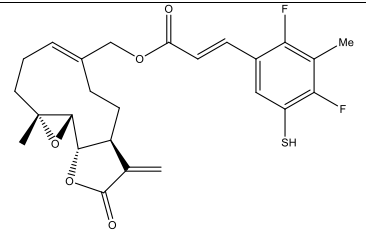
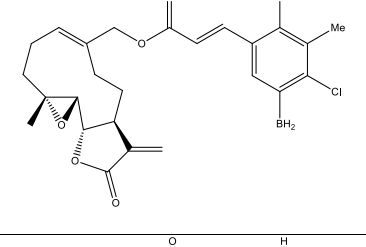
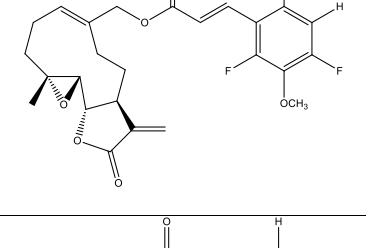
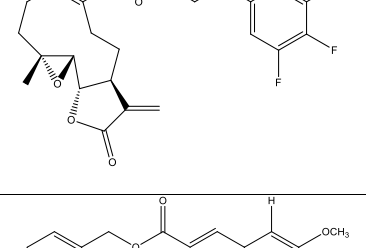
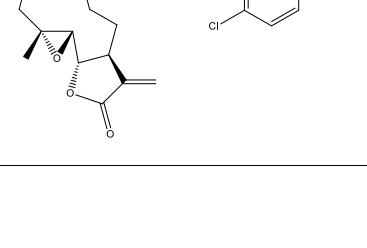
7		6.3281
8		6.0766
9		6.3470
10		6.1056
11		5.9117
12		5.9920

Table 13 (continued)

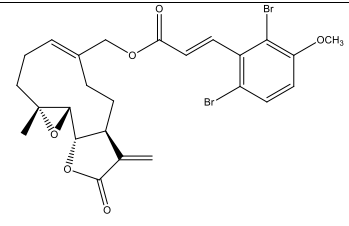
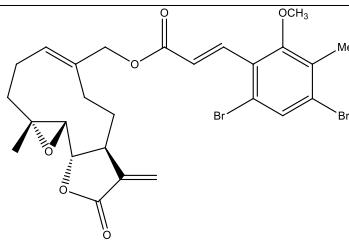
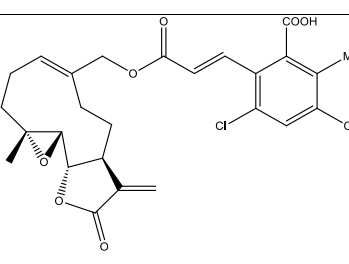
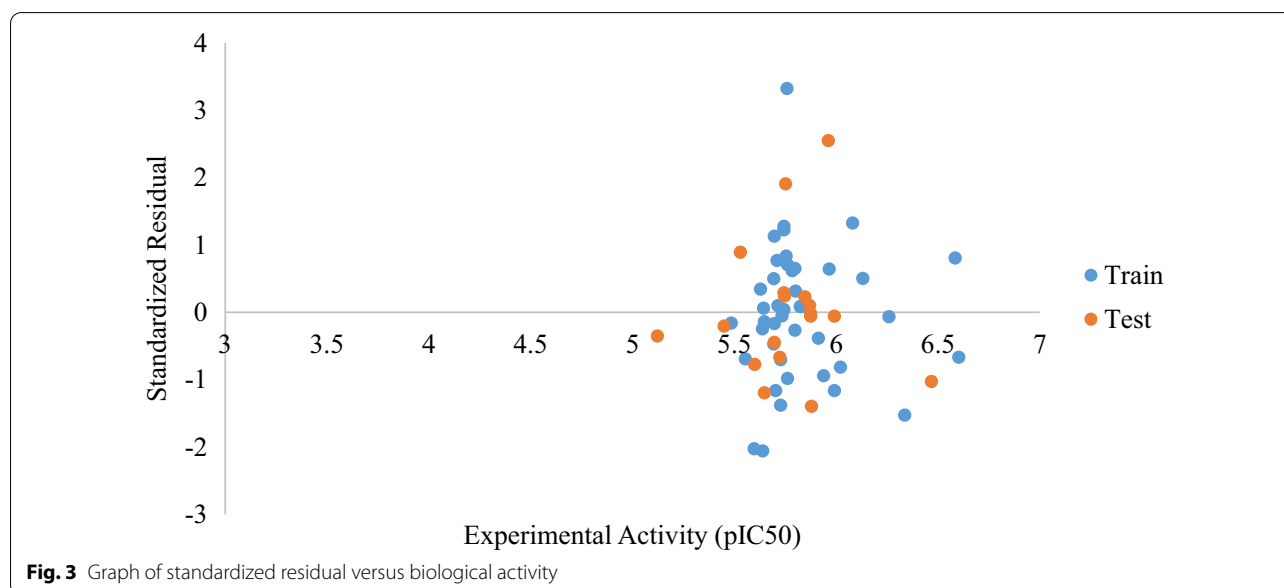
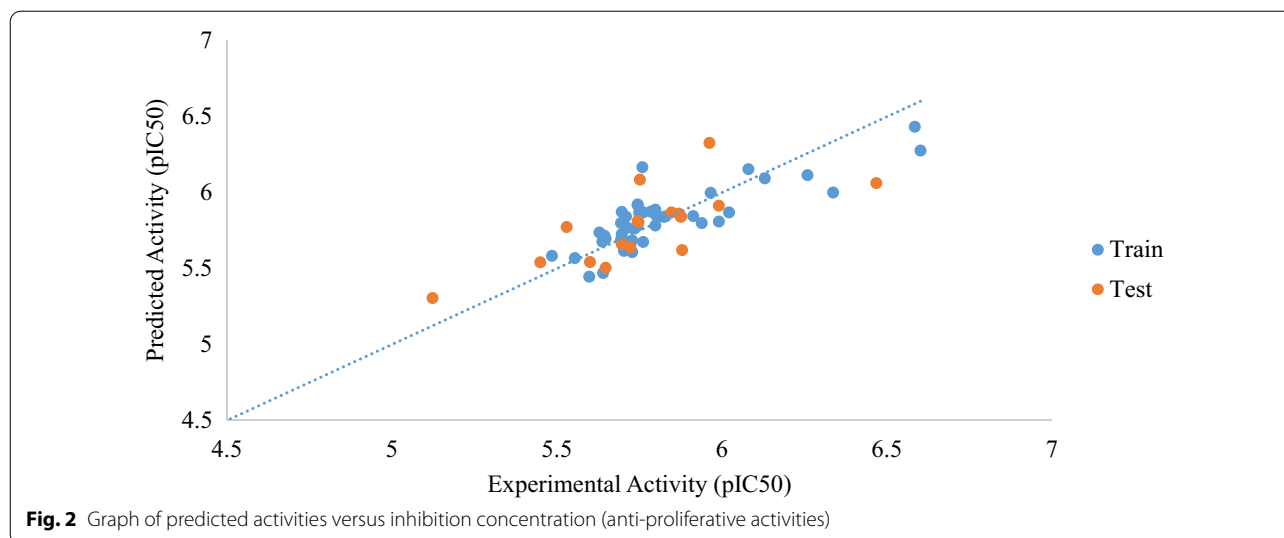
13		5.9804
14		6.1148
15		6.3606

Table 14 Physicochemical and ADME properties (pharmacokinetics) of designed parthenolide compounds against the MDA-MB-231 cell line

S/No	MW (g/mol)	nAH	nRB	HBA	HBD	MR	TPSA (\AA^2)	iLOGP	BBB	PAINS	Brenk
1	496.9	6	6	8	0	119.86	65.13	4.02	NO	0	4
2	492.49	6	6	8	2	123.66	117.17	3.41	NO	0	5
3	492.39	6	5	5	1	129.24	91.15	4.33	NO	0	5
4	493.38	6	5	6	1	126.86	85.36	3.9	NO	0	4
5	488.48	6	6	9	1	121.69	102.43	3.23	NO	0	4
6	581.29	6	5	5	1	134.62	91.15	4.15	NO	0	5
7	492.39	6	5	5	1	129.24	91.15	3.91	NO	0	5
8	476.53	6	5	7	0	121.98	103.93	4.08	NO	0	5
9	489.2	6	5	5	0	133.01	65.13	0	NO	0	5
10	460.47	6	6	8	0	116.26	74.36	3.45	NO	0	4
11	442.26	6	5	7	0	117.94	65.13	0.00	NO	0	5
12	458.93	6	6	6	0	121.35	74.36	4.02	NO	0	4
13	582.28	6	6	6	0	131.74	74.36	4.24	NO	0	4
14	596.30	6	6	6	0	136.71	74.36	4.65	NO	0	4
15	521.39	6	6	7	1	131.79	102.43	3.67	NO	0	4

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

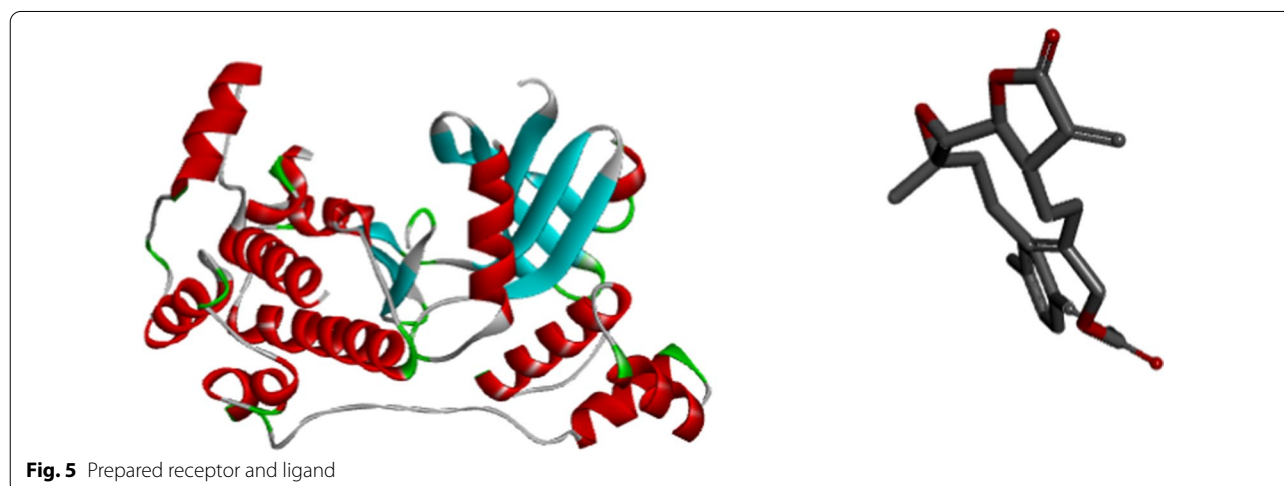
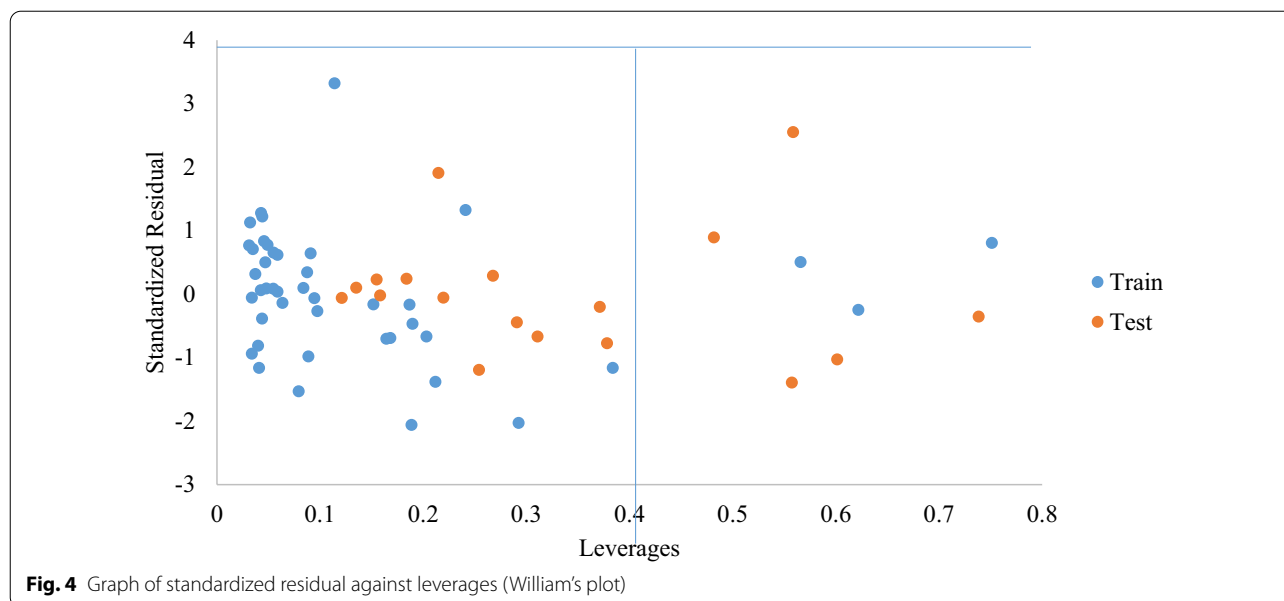


interaction, hydrogen bond length and the hydrophobic interaction as shown in Table 12.

Compound 9 showed backbone conventional hydrogen bonds (2.74925, 2.4619, 2.98056, 3.34925, 3.35524 and 3.59598 Å⁰) with THR19, GLY21, ARG53, CYS89, GLY18 and ILE17 of the protein target site. Furthermore, it formed hydrophobic bonds (5.96743, 5.00307, 5.43404, 4.42983, 3.89836, 5.15308, 4.67658 and 4.44028 Å⁰) with PHE22, ILE17, VAL25, LEU139, ILE17, PHE22, TYR88 and ALA23 of the target site. It formed a halogen and

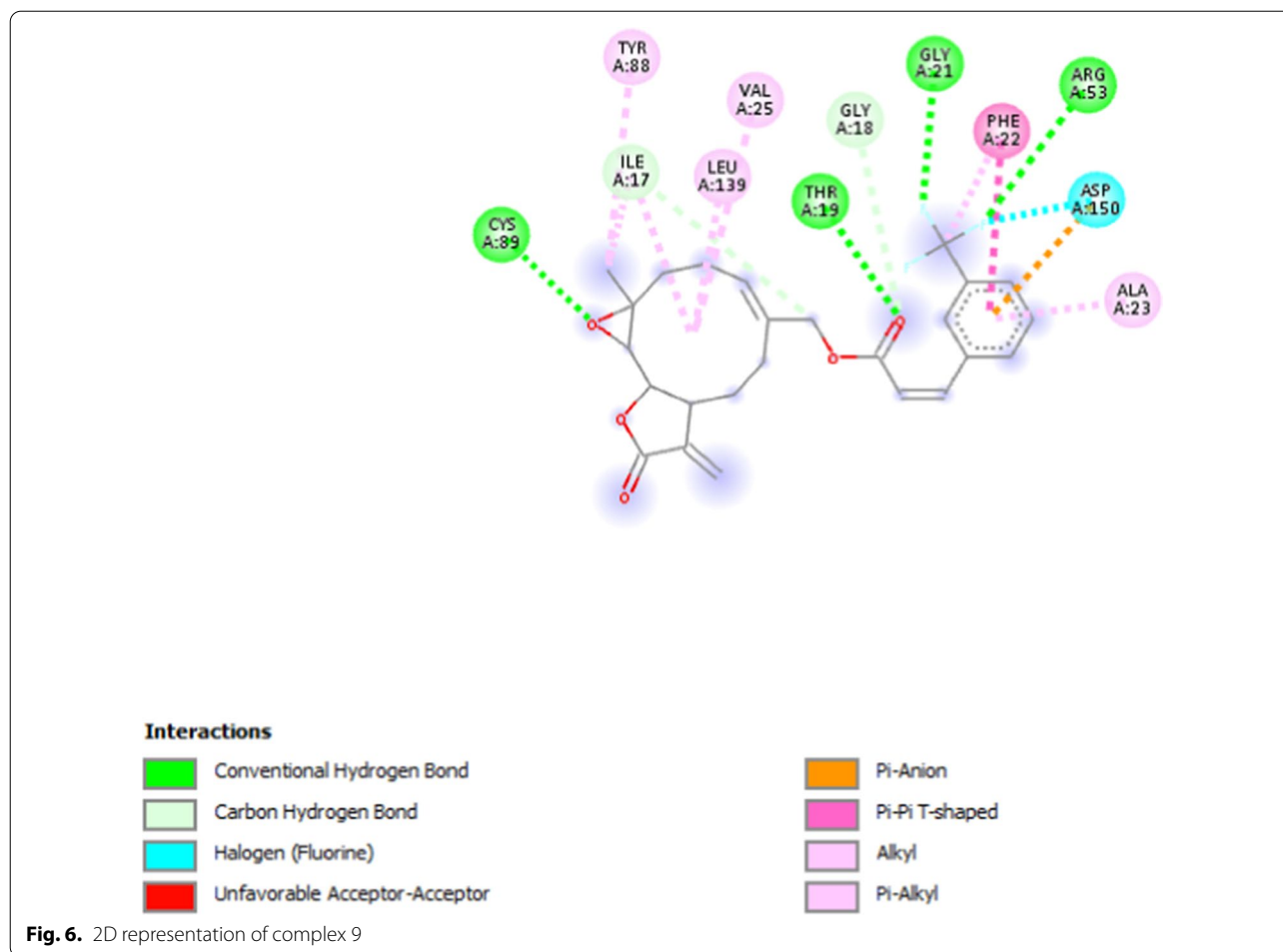
electrostatic bond (3.33001 and 4.20988 Å⁰) with amino acids of ASP150. The carbonyl group of the compound acts as a hydrogen acceptor and formed one hydrogen bond with THR19, while two of the fluorine (Halogens) attached to the phenyl group of the compound, act as a hydrogen donor and formed two hydrogen bonds with GLY21 and ARG53 of the protein crystal.

Compound 2 also showed four hydrogen bonding (1.91438, 3.50504, 3.40604 and 3.35007 Å⁰) with LYS134, GLY18, ASP150 and THR19 of the active site of the



receptor. The derivative also formed hydrophobic bonds (3.72576, 3.10836, 4.3497, 5.0501, 4.61294, 4.86094, 4.17368, 4.84776, 5.45664 and 4.54951 Å) with amino acids of PHE22, ALA23, ALA38, ILE17, VAL25, LEU139, VAL25, ALA38, LEU86 and ILE149 of the protein target. The carbonyl of the compound attached to 3-methylenedihydrofuran-2(3H)-one acts as a hydrogen acceptor to form one hydrogen bond with LSY134.

Compound 17 also formed hydrogen and hydrophobic bonds with the protein receptor at different bond distances as shown in Table 12. The hydrogen and hydrophobic interactions show that the ligands have high potency against the MELK receptor. The prepared receptor and ligand are shown in Fig. 5. The 3D and 2D representations of the binding pose are shown in Figs. 6 and 7, and how the ligands bind firmly to the active sites



of the receptor and also a detailed interaction to form complexes.

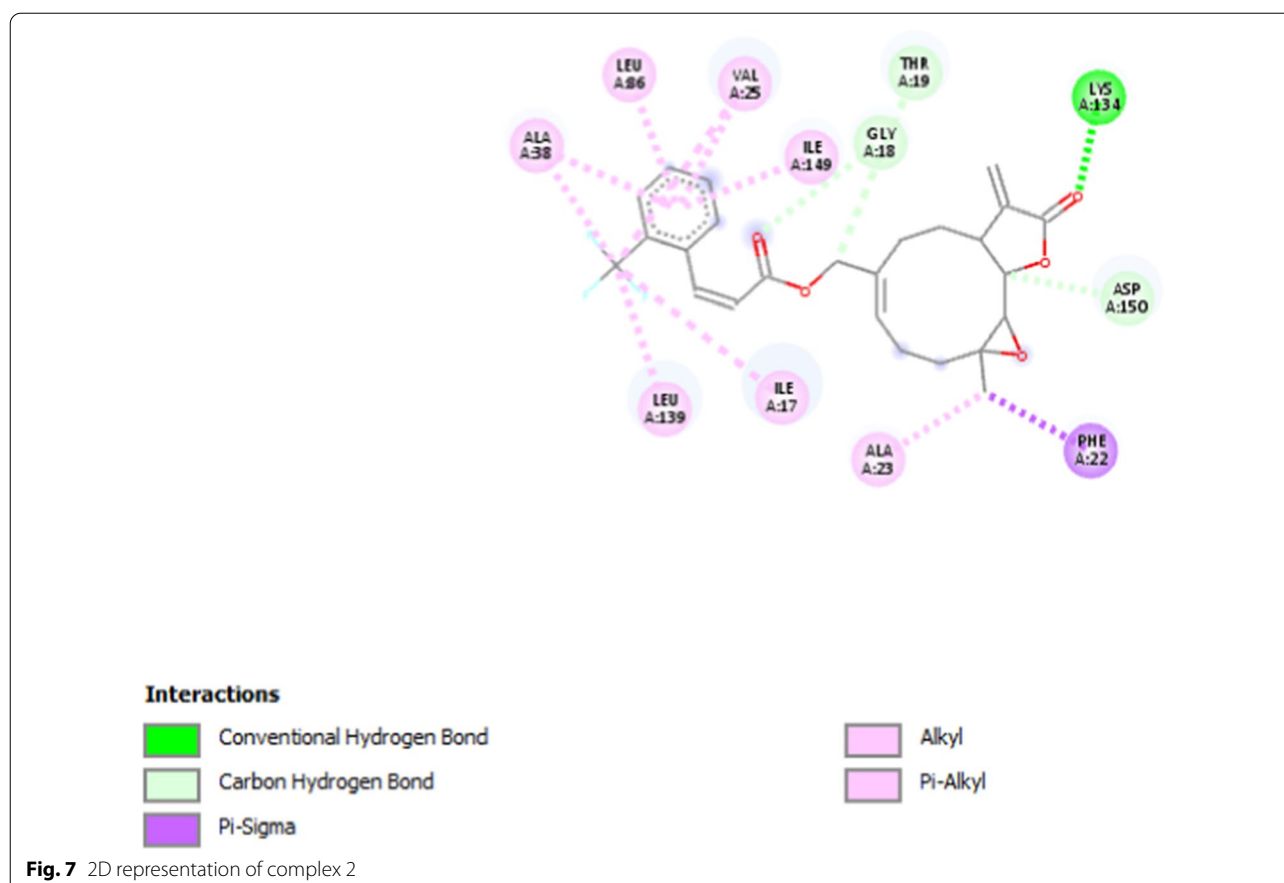
Ligand-based drug design

Compounds 13 and 1 were chosen as lead compounds in the design of fifteen (15) new parthenolide derivative compounds due to their high predicted activity and low residual values as shown in Table 7. From the mean effect of the descriptors, AATS6i had a greater positive impact, while VR1_Dze had the least positive impact on the model. According to AATS6i (Centered Broto-Moreau autocorrelation—lag 6/weighted by first ionization potential) and VR1_Dze (Randic-like eigenvector-based index from Barysz matrix/weighted by Sanderson electronegativities) descriptors, adding more ionization potential and electronegative elements

would increase the potency of the chosen templates. The structural modification occurred by adding more ionization potential and electronegative elements to template molecules (compounds 1 and 13). The predicted activities of the newly designed compounds were higher than that of the template molecules as shown in Table 13.

Physicochemical and ADME properties (pharmacokinetics) of designed parthenolide compounds

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, and this shows that all the designed compounds passed the drug-likeness test as shown in Table 14, making the compounds a breakthrough in finding the cure to triple-negative breast cancer. Figure 9 shows the bioavailability RADAR of molecules 2, 5 and 8. The bioavailability RADAR enables a first glance at the drug-likeness of a molecule. The pink area signifies the ideal range for each properties (lipophilicity: XLOGP3 between -0.7 and $+5.0$, size: MW between 150 and 500 g/mol, polarity: TPSA between 20 and 130 \AA^2 , solubility: log S not higher than 6, saturation: fraction of carbons in the sp^3 hybridization not less than 0.25 and flexibility: no more than 9 rotatable bonds) (Daina et al. 2017). The GI—gastrointestinal absorption was also high for all the designed molecules.

Conclusion

The results obtained from the QSAR mathematical model of parthenolide derivatives were used in designing new derivatives compounds that were more effective and potent. The molecular docking result of parthenolide derivatives showed that compounds 2, 9 and 17 had higher docking scores than the standard drug adriamycin. The compounds would serve as the most promising inhibitors (MELK). Furthermore, the pharmacokinetics analysis carried out on the newly designed compounds revealed that all the compounds can move on to the next step of pre-clinical trial because they passed drug-likeness test (ADME and other physicochemical properties) and they also adhered to the Lipinski rule of five: a major principle used in analyzing the drug-likeness of small compounds. This gives a great breakthrough in medicine in finding the cure to triple-negative breast cancer (MBA-MD-231 cell line).

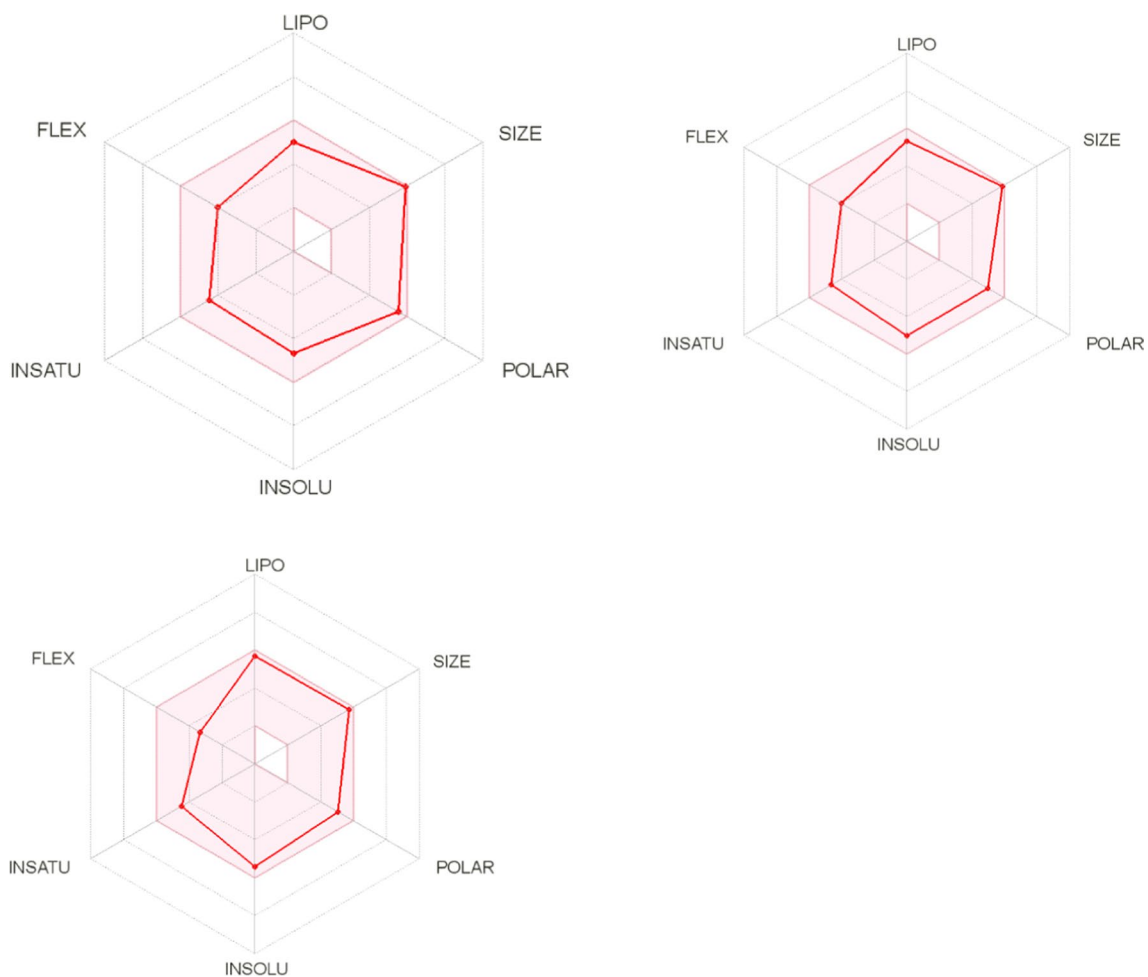


Fig. 8 The bioavailability RADAR of molecules 12, 5 & 8

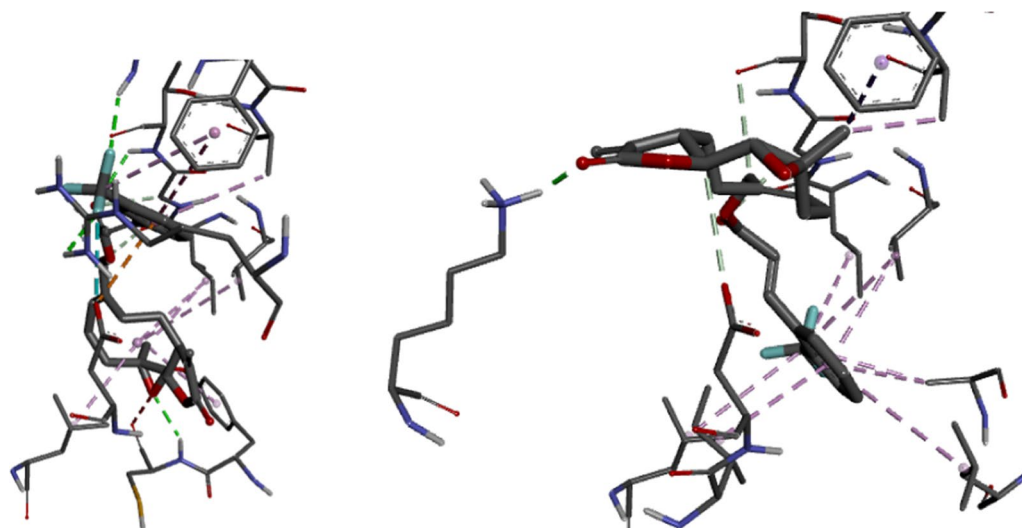


Fig. 9 3D representation of complex 9 and 2

Abbreviations

ADME properties: Absorption, distribution, metabolism and elimination properties; MELK: Maternal Embryonic Leucine Zipper Kinase; QSAR: Quantitative structure–activity relationship; TNBCs: Triple-negative breast cancers.

Acknowledgements

Authors acknowledge the technical effort of the physical chemistry department, Ahmadu Bello University Zaria for their immense advice and various contributions toward the successful completion of the research.

Authors' contributions

All author's participated accordingly, HAL carried out the computational analysis, AU arranged the manuscript according to the journal format, and SU re-confirmed the results and edited the manuscript to ensure that error was minimized. All authors read and approved the final manuscript.

Funding

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

Availability of data and materials

Not applicable.

Declarations

Ethics approval and consent to participate

Not applicable.

Consent for publication

Not applicable.

Competing interests

The authors declare no competing interest.

Received: 27 May 2020 Accepted: 14 April 2021

Published online: 17 May 2021

References

- Abdullahi M, Uzairu A, Shallangwa GA, Mamza P, Arthur DE, Ibrahim MT (2020) In-silico modelling studies on some C14-urea-tetrandrine derivatives as potent anti-cancer agents against prostate (PC3) cell line. *J King Saud Univ Sci* 32(1):770–779
- Abdulrahman HL, Uzairu A, Uba S (2020a) QSAR, ligand based design and pharmacokinetic studies of parviflorons derivatives as anti-breast cancer drug compounds against MCF-7 cell line. *Chem Africa* 29:1–3
- Abdulrahman HL, Uzairu A, Uba S (2020b) Computer modeling of some anti-breast cancer compounds. *Struct Chem* 15:1–9
- Abdulrahman HL, Uzairu A, Uba S (2020c) In silico studies of some 2-anilino-pyrimidine derivatives as anti-triple-negative breast cancer agents. *Beni-Suef Univ J Basic Appl Sci* 9:1–2
- Arthur DE, Uzairu A, Mamza P, Abechi SE, Shallangwa G (2020) Activity and toxicity modelling of some NCI selected compounds against leukemia P388ADR cell line using genetic algorithm-multiple linear regressions. *Journal of King Saud University-Science* 32(1):324–331
- Daina A, Michielin O, Zoete V (2017) SwissADME: a free web tool to evaluate pharmacokinetics, drug-likeness and medicinal chemistry friendliness of small molecules. *Sci Rep* 7:42717
- Edupuganti R, Taliaferro JM, Wang Q, Xie X, Cho EJ, Vidhu F, Ren P, Anslyn EV, Bartholomeusz C, Dalby KN (2017) Discovery of a potent inhibitor of MELK that inhibits expression of the anti-apoptotic protein Mcl-1 and TNBC cell growth. *Bioorg Med Chem* 25(9):2609–2616
- Friedman JH (1991) Multivariate adaptive regression splines. *The annals of statistics*. 19:1–67
- Ge W, Hao X, Han F, Liu Z, Wang T, Wang M, Chen N, Ding Y, Chen Y, Zhang Q (2019) Synthesis and structure-activity relationship studies of parthenolide derivatives as potential anti-triple negative breast cancer agents. *Eur J Med Chem* 166:445–69
- Guan L, Yang H, Cai Y, Sun L, Di P, Li W, Liu G, Tang Y (2019) ADMET-score—a comprehensive scoring function for evaluation of chemical drug-likeness. *MedChemComm* 10(1):148–157
- Hou Y, Zhu L, Li Z, Shen Q, Xu Q, Li W, Liu Y, 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
- Hu K, Law JH, Fotovati A, Dunn SE (2012) Small interfering RNA library screen identified polo-like kinase-1 (PLK1) as a potential therapeutic target for breast cancer that uniquely eliminates tumor-initiating cells. *Breast Cancer Res* 14(1):1–5
- Iqbal J, Ejaz SA, Khan I, Ausekle E, Miliutina M, Langer P (2019) Exploration of quinolone and quinoline derivatives as potential anticancer agents. *DARU J Pharm Sci* 27(2):613–626
- Jalali-Heravi M, Kyani A (2004) Use of computer-assisted methods for the modeling of the retention time of a variety of volatile organic compounds: a PCA-MLR-ANN approach. *J Chem Inf Comput Sci* 44(4):1328–1335
- Jo J, Kim SH, Kim H, Jeong M, Kwak JH, Han YT, Jeong JY, Jung YS, Yun H (2019) Discovery and SAR studies of novel 2-anilino-pyrimidine-based selective inhibitors against triple-negative breast cancer cell line MDA-MB-468. *Bioorg Med Chem Lett* 29(1):62–65
- Kaplan W. Background Paper 6.5 Cancer and Cancer Therapeutics. World Health Organization (ed) Priority medicines for Europe and the world: update. 2013:6–5.
- Kennard RW, Stone LA (1969) Computer aided design of experiments. *Technometrics* 11(1):137–148
- Lolak N, Akocak S, Bua S, Supuran CT (2019) Design, synthesis and biological evaluation of novel ureido benzenesulfonamides incorporating 1, 3, 5-triazole moieties as potent carbonic anhydrase IX inhibitors. *Bioorg Chem* 82:117–122
- Myers RH (1990) Classical and modern regression application. Duxbury Press, CA
- Noolvi MN, Patel HM (2013) A comparative QSAR analysis and molecular docking studies of quinazoline derivatives as tyrosine kinase (EGFR) inhibitors: A rational approach to anticancer drug design. *J Saudi Chem Soc* 17(4):361–379
- Qu FZ, Xiao SN, Wang XD, Zhang Y, Su GY, Zhao YQ (2019) Semi-synthesis and anti-tumor activity of novel 25-OCH3-PPD derivatives incorporating aromatic moiety. *Bioorg Med Chem Lett* 29(2):189–193
- 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. *QSAR Comb Sci* 22(1):69–77
- 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
- Xu P, Chu J, Li Y, Wang Y, He Y, Qi C, Chang J (2019) Novel promising 4-anilino-quinazoline-based derivatives as multi-target RTKs inhibitors: design, molecular docking, synthesis, and antitumor activities in vitro and vivo. *Bioorg Med Chem* 27(20):114938
- Yap CW (2011) PaDEL-descriptor: an open source software to calculate molecular descriptors and fingerprints. *J Comput Chem* 32(7):1466–1474

Publisher's Note

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