


2D- and 3D-QSAR modelling, molecular docking and *in vitro* evaluation studies on 18 β -glycyrrhetic acid derivatives against triple-negative breast cancer cell line

Aparna Shukla, Rekha Tyagi, Sanjeev Meena, Dipak Datta, Santosh Kumar Srivastava & Feroz Khan

To cite this article: Aparna Shukla, Rekha Tyagi, Sanjeev Meena, Dipak Datta, Santosh Kumar Srivastava & Feroz Khan (2019): 2D- and 3D-QSAR modelling, molecular docking and *in vitro* evaluation studies on 18 β -glycyrrhetic acid derivatives against triple-negative breast cancer cell line, Journal of Biomolecular Structure and Dynamics, DOI: [10.1080/07391102.2019.1570868](https://doi.org/10.1080/07391102.2019.1570868)

To link to this article: <https://doi.org/10.1080/07391102.2019.1570868>

 View supplementary material 

 Accepted author version posted online: 28 Jan 2019.
Published online: 22 Feb 2019.

 Submit your article to this journal 

 Article views: 31

 View Crossmark data 



2D- and 3D-QSAR modelling, molecular docking and *in vitro* evaluation studies on 18 β -glycyrrhetic acid derivatives against triple-negative breast cancer cell line

Aparna Shukla^a, Rekha Tyagi^b, Sanjeev Meena^c, Dipak Datta^c, Santosh Kumar Srivastava^b and Feroz Khan^a

^aMetabolic and Structural Biology Department, CSIR-Central Institute of Medicinal and Aromatic Plants, Lucknow, Uttar Pradesh, India;

^bMedicinal Chemistry Division, Central Institute of Medicinal and Aromatic Plants, Lucknow, Uttar Pradesh, India; ^cBiochemistry Division, CSIR-Central Drug Research Institute (CDRI), Lucknow, Uttar Pradesh, India

Communicated by Ramaswamy H. Sarma

ABSTRACT

Triple-negative breast cancers (TNBCs) are one of the most aggressive and complex forms of cancers in women. TNBCs are commonly known for their complex heterogeneity and poor prognosis. The present work aimed to develop a predictive 2D and 3D quantitative structure–activity relationship (QSAR) models against metastatic TNBC cell line. The 2D-QSAR was based on multiple linear regression analysis and validated by Leave-One-Out (LOO) and external test set prediction approach. QSAR model presented regression coefficient values for training set (r^2), LOO-based internal regression (q^2) and external test set regression (pred_r^2) which are 0.84, 0.82 and 0.75, respectively. Five properties, Epsilon4 (electronegativity), ChiV3cluster (valence molecular connectivity index), chi3chain (retention index for three-membered ring), TNN5 (nitrogen atoms separated through 5 bond distance) and nitrogen counts, were identified as important structural features responsible for anticancer activity of MDA-MB-231 inhibitors. Five novel derivatives of glycyrrhetic acid (GA) named GA-1, GA-2, GA-3, GA-4 and GA-5 were semi-synthesised and screened through the QSAR model. Further, *in vitro* activities of the derivatives were analysed against human TNBC cell line, MDA-MB-231. The result showed that GA-1 exhibits improved cytotoxic activity to that of parent compound (GA). Further, atomic property field (APF)-based 3D-QSAR and scoring recognise C-30 carboxylic group of GA-1 as major influential factor for its anticancer activity. The significance of C-30 carboxylic group in GA derivatives was also confirmed by molecular docking study against cancer target glyoxalase-I. Finally, the oral bioavailability and toxicity of GA-1 were assessed by computational ADMET studies.

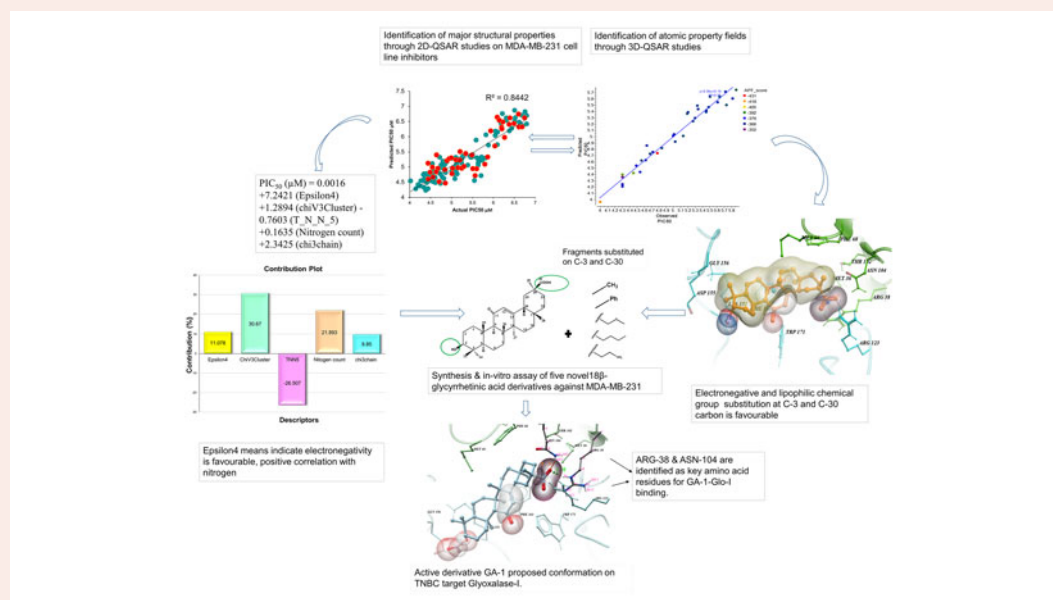
ARTICLE HISTORY

Received 30 October 2018

Accepted 8 January 2019

KEYWORDS

QSAR; breast cancer; triple-negative breast cancer; glycyrrhetic acid; MDA-MB-231; glyoxalase-I



Introduction

Breast cancer is the most frequently diagnosed cancer and the second leading cause of female deaths worldwide. In majority of cases, mortality is due to its metastatic dissemination to distant sites (Polyak, 2011). Despite the enormous medical importance of metastasis, its molecular underpinning remains insufficiently understood because of its intertumour and intratumour heterogeneity (Bianchini et al., 2016; Ovcaricek, Frkovic, Matos, Mozina, & Borstnar, 2011). Breast carcinomas, demarcated as triple-negative breast cancers (TNBC), are highly aggressive and do not express progesterone receptor (PR), estrogen receptor (ER) and human epidermal growth factor receptor2 (HER2) (Thike et al., 2010). Consequently, it is resistant to hormone-targeted therapies, and only 20% of TNBC respond well to standard chemotherapy using anthracycline-based (doxorubicin plus cyclophosphamide) or paclitaxel chemotherapy, etc. (Zardavas, Baselga, & Piccart, 2013). Thus, in current breast cancer research, developing improved treatment for metastatic TNBC is one of the highest priorities. Several researches have been carried out to understand metastatic TNBC cells sensitivity towards plant-based different chemical scaffolds (Iqbal et al., 2018). Recent prognosis works on TNBC focusing on targets PARP1, mTOR, TGF- β from Notch signalling, Wnt/ β -catenin and Hedgehog pathways (Badve et al., 2011; Jamdade et al., 2015; Wein et al., 2018).

Glycyrrhiza glabra, an Indian medicinal herb, also known as licorice, contains biologically active triterpenoid glycyrrhizic acid (GL). GL is a diglucopyranosiduronic acid of the glycyrrhetic acid (GA) (Tewari et al., 2017). Substantial research has been carried out on human liver metabolism of GL. Krähenbühl, Hasler, and Krapf (1994), revealed GL transformed into its aglycone form GA through intestinal bacteria when orally administered. In cancer research, GA is mostly explored for its activity against human hepatocellular carcinoma (HCC) cells since an earlier work revealed the existence of GA receptor on rat and human hepatocytes surface (Negishi et al., 1991). It has been reported that anti-HCC response of GA is mediated through inhibition of immune response by regulating T cells, cell cycle arrest, induction of cell apoptosis and autophagy (Cai et al., 2017). Evidently, GA has been identified to exhibit remarkable anticancer activities. Therefore, over the past decade, GA has been serving as a good structural template for more potent anticancer agents. Many groups have studied the effects of structural modification in GA on the cytotoxicity of various human cancer cell lines (Xu et al., 2017). Despite these capabilities, the mechanism of action of GA in metastatic TNBC has not been investigated so far.

Notably, our earlier work on GA and its novel derivatives against breast cancer MCF-7 displayed good anticancer potency (Yadav, Kalani, Khan, & Srivastava, 2013; Yadav et al., 2014). Therefore, the present work was designed to combat metastatic TNBC cell lines using biological effects of GA and its novel derivatives. The work includes chemical feature identification of metastatic TNBC cell inhibitors through regression-based quantitative structure–activity relationship (QSAR) model (Yadav & Khan, 2013). Further, five novel GA

derivatives were semi-synthesised and screened through the prepared QSAR model. The derivatives were further investigated for *in vitro* activity in metastatic breast cancer cell line MDA-MB-231. Subsequently, atomic property field (APF)-based 3D-QSAR model was generated to explore APF and structure–activity relationship of synthesised derivatives. The anticancer mechanism of action of GA derivatives on TNBC drug targets was explored through molecular docking studies. In TNBC cells, enzyme Glyoxalase-I (GLO-I) inhibition leads to increased level of alpha-oxoaldehydes that cause increase in apoptosis and suppress migration and invasion of metastatic breast cancer. Therefore, GLO-I is considered as one of the promising TNBC targets. Here, molecular docking and 3D-QSAR modelling were performed considering Glyoxalase-I as a promising TNBC drug target. The oral bio-availability and possible toxicity were also assessed through computational ADMET (absorption, distribution, metabolism, excretion and toxicity) analysis.

Materials and method

Computational 2D-QSAR modelling for GA derivative designing

Dataset collection and structure preparation

The modelling set comprises 144 compounds, metastatic TNBC cell line, MDA-MB-231 inhibitors (Tables S1 and S2, training and test set compounds, respectively, Supplementary materials) collected from the ChEMBL database and reported literatures (Gao et al., 2014; Goldbrunner et al., 1997; He et al., 2015; Motiwala et al., 2013; Li, Feng, Song, Li, & Huai, 2016). The modelling set exhibits plant product-inspired scaffolds and comprises two to five fused ring skeletons (Table S3, Supplementary material).

The structural drawing and geometry cleaning of the modelling set compounds were performed through ChemBioOffice suite Ultra v12.0 (2015) software (CambridgeSoft Corp., UK). Further, each compound was subjected to energy minimisation to get optimised bond distance, bond angles and set dihedrals by applying Merck molecular force field (MMFF). Moreover, the method adds additional properties to the compounds including initial potential energy, root-mean-square (RMS) gradient, MMFF energy and minimisation criteria.

Chemical descriptors calculation

For QSAR model generation, the compounds were denoted by structural descriptors or physicochemical properties. Computation for descriptors was done by using VLife MDS v4.4 (2014) software (VLife Technologies, NovaLead Pharma Pvt. Ltd., India). Software VLife calculates structural descriptors, belonging to major classes *viz.*, (a) physicochemical descriptors, (b) extended topochemical descriptors and (c) alignment-independent descriptors.

QSAR model generation

Primarily, the dataset of 144 compounds was divided into 70% as training set and 30% as external test set applying random selection technique using Vlife. The PIC_{50} (μM) value was assigned as dependent variable. PIC_{50} is a negative logarithm of IC_{50} value, expressed in molar concentration. The physico-chemical properties or structural descriptors were considered as independent variables. Invariable descriptors with zero or equal values were deleted. The regression coefficient for training set (100 compounds) was calculated by using Equation (1), where y_i and \hat{y}_i signify the actual and predicted PIC_{50} of i^{th} compound, respectively, whereas y_{mean} is the average/mean value of actual PIC_{50} of training set compounds

$$r^2 = \frac{\sum (y_i - \hat{y}_i)^2}{\sum (y_i - \hat{y}_{\text{mean}})^2} \quad (1)$$

QSAR model generation criteria and parameter used for feature selection

The 2D-QSAR model was developed by applying multiple linear regression (MLR) approach. A stepwise forward-backward selection criterion was applied for feature/descriptor extraction. Continuous multiple variable-based MLR model was generated, step by step depending on Fischer coefficient values (F values). The F test gives the statistical significance of the descriptor. The $F_{\text{test in}}$ and $F_{\text{test out}}$ values were set at 4 and 3, respectively. The predicted descriptors were identified by these stepping criteria. The search is terminated when the addition of additional variables is no longer needed. Before model development, inter-correlated descriptors (correlation >0.70) were discarded.

2 D-QSAR model validation

To scrutinise the predictability of developed model, leave-one-out (q^2 , LOO), external set predictions (r^2_{pred}) and r^2_m matrix were calculated (Cramer III, Bunce, Patterson, & Frank, 1988; Golbraikh & Tropsha, 2002; Ojha et al., 2011). An LOO-based cross-validated regression coefficient (q^2) was calculated based on Equation (2) (Shen et al., 2002), where y_i and \hat{y}_i signify the actual and predicted PIC_{50} values for i^{th} compound, respectively, whereas y_{mean} is the average/mean value of actual PIC_{50} of training set compounds

$$q^2 = 1 - \frac{\sum (y_i - \hat{y}_i)^2}{\sum (y_i - y_{\text{mean}})^2} \quad (2)$$

The regression for external test set (r^2_{pred}) was calculated by using Equation (3) (Kier & Hall, 1977; Golbraikh & Tropsha, 2002). r^2_{pred} validate the model predictability for external compounds and verify the model predicted result's reliability

$$r^2_{\text{pred}} = 1 - \frac{\sum (y_{i(\text{test})} - \hat{y}_{i(\text{test})})^2}{\sum (y_{i(\text{test})} - y_{\text{mean}(\text{test})})^2} \quad (3)$$

Randomisation test

The robustness of generated model was also assessed by calculating Z score values using Equation (4), where h , μ and σ

signify r^2 of original dataset, average values of r^2 s for random datasets and standard deviation for random dataset, respectively. The calculated Z score should be higher than the tabulated value Z_c as reported by Zheng and Tropsha (2000). The higher Z score indicates that the null hypothesis is rejected and the model generated from the actual dataset is statistically significant

$$Z = (h - \mu) / \sigma \quad (4)$$

r^2_m matrix for QSAR model validation

The external predictability of developed model was also checked by using r^2_m matrix method calculating r^2_0 , r^2_0 , r^2_m , r^2_m , \bar{r}^2_m and r^2_m (Ojha et al., 2011). The r^2_m matrix is measured by using r^2 and r^2_0 , where r^2 and r^2_0 signify correlation between observed and predicted values with and without intercept for the regression line, respectively. Statistical parameters r^2_0 , r^2_0 , r^2_m , r^2_m , and \bar{r}^2_m were computed as mentioned by Ojha, Mitra, Das, and Roy, (2018). The equation used for r^2_m parameter calculation is given in Equation (5)

$$r^2_m = r^2 \left(1 - \left| \sqrt{r^2_0 - r^2} \right| \right) \quad (5)$$

Applicability domain (AD) assessment of 2D-QSAR model

A statistically validated model predicted results are considered to be reliable only when the query set compound falls within its AD. Here, three most important criteria were adopted to check the AD of developed model viz., (i) the biological space cover of whole dataset, (ii) the chemical space cover by training and test sets, and (iii) distance-based distribution of training and test sets in structure and activity space (Jaworska, Nikolova-Jeliazkova, & Aldenberg, 2005). A 3D principal component analysis (PCA) was applied to compute the projection of chemical space of test set within training set (Adhikari et al., 2017; Amin et al., 2018). A structure similarity-based hierarchical cluster analysis was done to assess structure relatedness of training, test and query set compounds (Figure S1, Supplementary material).

APF-based 3D-QSAR modelling study

An APF-based 3D-QSAR was also performed on congeneric series of 42 GA derivatives. An APF-based 3D-QSAR thoroughly describes the spatial arrangement of structural features that bestow specific activity to the molecule, since a 3D-QSAR model reliability is highly dependent upon the structural filed alignment, Therefore, for 3D-QSAR studies, a congeneric series of 42 GA derivatives were taken instead of using a total of 144 compounds so as to get more homogeneous structure space.

A series of 42 GA derivatives with known inhibition activity against MDA-MB-231 were collected from 2D-QSAR dataset and reported literatures (Yang et al., 2016, ; Gao, et al., 2014). Their structures were drawn and converted to 3D ICM object using ICM-Chemist v3.8-6a 2018 (Molsoft L.L.C., San Diego, CA) software (Abagyan, 2018, <http://www.molsoft.com/icm-chemist-pro.html>; Totrov, 2008). The set of 42 GA

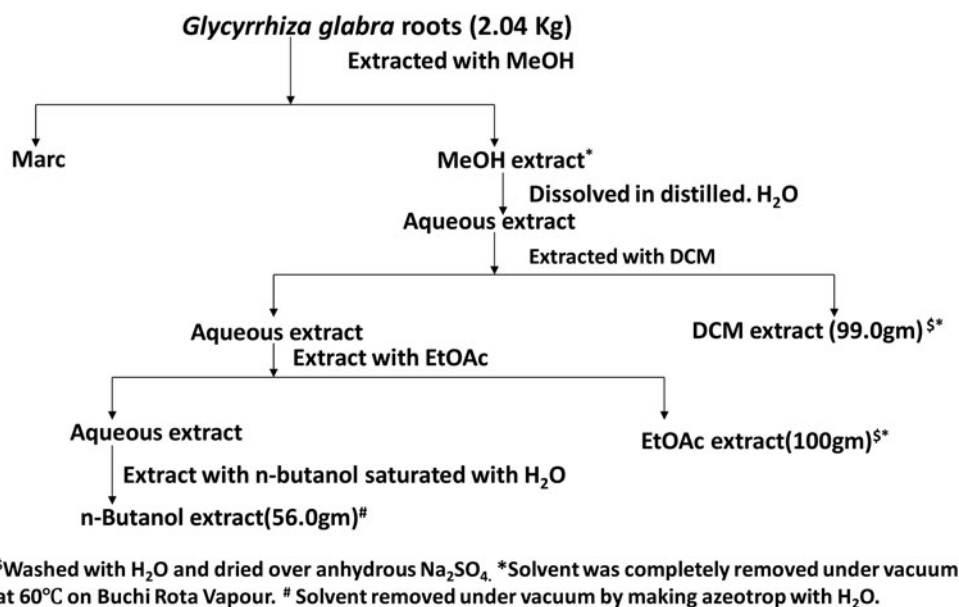


Figure 1. A schematic procedure for extraction and fractionation of *Glycyrrhiza glabra* roots.

derivatives was randomly split into 37 training and 5 test sets using DS v3.5. The crystal structure of enzyme Glyoxalase-I (GLO-I) bound with GA (2.3 resolution) was retrieved from protein database (PDB: 4PV5). The GLO-I-bound conformation of GA was taken as the rigid template structure to which training set structures were aligned based on their APF energy fields. The APF fields of co-crystallised GA are depicted in Figure S5 under [supplementary materials](#). Afterward, for each molecule, 3D-based continuous atomic potentials were generated and approximated based on regular space grid. These continuous potentials represent seven physicochemical properties *viz.*, hydrogen bond donor (blue blob) and acceptors (red blob), sp²-hybridised carbon atoms, molecules lipophilicity (yellow blobs), charge, molecule size and electronegativity or positivity.

Therefore, the training set of 37 compounds can be represented by 259 descriptors. The training and test set molecules were aligned on generated APF fields of co-crystallised GA. For quantitative prediction of novel compound, a partial least square (PLS)-based optimal weight distribution was assigned to each molecule based on its APF components. The optimal number of latent vectors for PLS was established by LOO cross-validation on the training set. Then, the weighted contributions of each APF components were added together. For external validation, randomly selected five compounds were assigned predicted binding values by calculating their fit within the combined QSAR-APF score. The model further utilised to design and screen novel GA derivatives GA-1, GA-2, GA-3, GA-4 and GA-5 based on their APF alignment.

Chemistry

Extraction and chemical synthesis

Five novel 18β-GA derivatives were designed and synthesised modifying C-3 and C-30 positions. [Figure 4\(a\)](#) represents compound preparation scheme-1, i.e. synthesis of 3-O-acyl

derivatives of GA and 5b. [Figure 4\(b\)](#) compound preparation scheme-2, i.e. synthesis of amide derivatives of 3-O-acyl GA.

Isolation of 3β-hydroxy-11-oxoolean-12-en-29-oic acid (GA) from *G. glabra*

Extraction and fractionation of *G. glabra* roots

The roots of *G. glabra* were air-dried under shade and then powdered. This powdered material (2.04 kg) was extracted with methanol (4 × 5 L) at room temperature. The combined methanol extract was subjected for complete solvent removal at 40 °C under vacuum. This dried methanolic extract was dissolved in distilled water (2 L) and successively extracted with dichloromethane, ethyl acetate and *n*-butanol (4 × 400 ml). The combined dichloromethane, ethyl acetate and *n*-butanol extracts were separately subjected under vacuum distillation at 40 °C to yield dichloromethane (99.0 g), ethyl acetate (100.0 g) extracts and *n*-butanol (56.0 g) as given in [Figure 1](#).

Isolation of GL from *n*-BuOH extract of *G. glabra* by flash chromatography

A glass flash column with internal diameter 3 cm and length 23 cm was used. The Flash column was packed with silica gel-H of TLC grade (without binder). The column was tightly packed using vacuum followed by elution of the column with a non-polar solvent (hexane) to ensure good packing of the column. Before loading the extract, glass column was completely dried, and then 1.00 g of BuOH extract of *G. glabra* was dissolved in small amount of methanol, and with the help of a pipette, it was spread onto the glass column without using vacuum to form a uniform band. The above step was followed by complete drying of the glass column under vacuum. Gradient elution of flash was carried out with

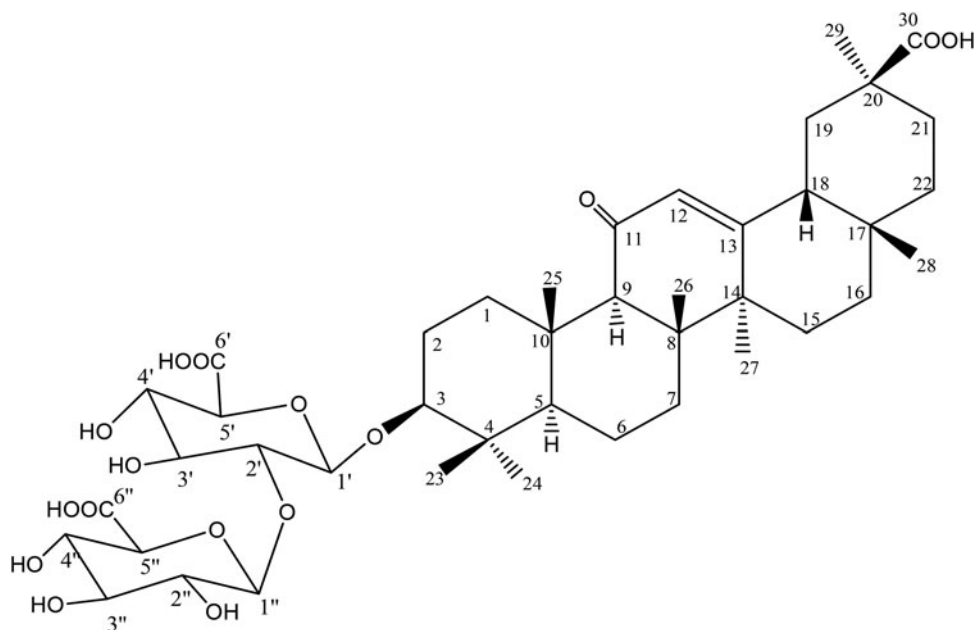


Figure 2. Structure of glycyrrhizic acid.

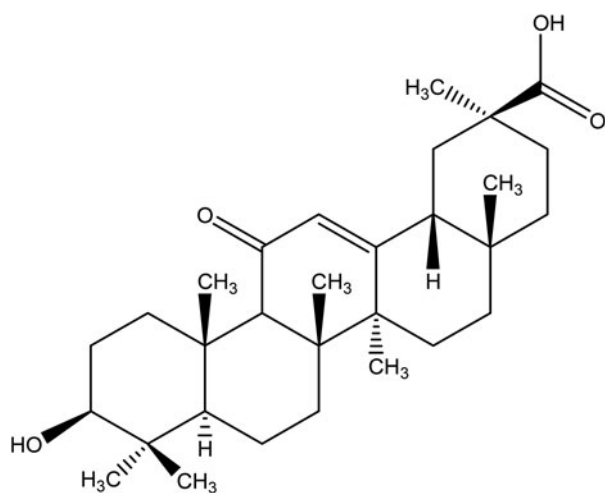


Figure 3. Structure of glycyrrhetic acid (GA).

a mixture of CHCl_3 :MeOH in increasing order (up to 50% MeOH). Fractions of 50 ml each were collected. A total of 149 fractions were collected and pooled based on their TLC profile. Excellent separation was achieved due to fine particle size (average size $10\ \mu\text{m}$) of silica gel-H. The pooled fractions 42–58 (650 mg) eluted with CHCl_3 :MeOH (85:15) were homogeneous and characterised as GL based on their ^1H and ^{13}C NMR spectroscopic data (Figure 2).

Acidic hydrolysis of GL to GA

GL (650.0 mg obtained from Flash chromatographic fractions 42–58) was dissolved in 25 ml of 10% H_2SO_4 solution in MeOH, and the reaction mixture was refluxed for 3–4 h, which was further diluted with water, neutralised with 10% NaOH solution and then extracted thrice with CHCl_3 . The combined CHCl_3 extract was dried under vacuum, which afforded aglycone (450 mg). This aglycone was purified over flash using silica gel-H. A total of 148 fractions were collected

and pooled based on their TLC profile. The fractions 29–46 eluted with CHCl_3 :MeOH (99:1) afforded homogeneous product (GA, 250 mg) characterised as GA on the basis of their ^1H and ^{13}C NMR spectroscopic data, 18 β -GA (Figure 3).

Semi-synthesis of GA derivatives

The chemical reactions for the synthesis of 3-O-acyl derivatives and 3-O-acetyl amide derivatives are depicted in Schemes 1 and 2, respectively. All the acyl derivatives were synthesised by taking GA and corresponding acyl chloride (2 equivalent) and a catalytic amount of 4-(*N,N*-dimethyl) aminopyridine (DMAP) into dry pyridine as solvent and refluxing the reaction mixture for 8 h up to 80°C (Figure 4(a)). Reaction mixture was then neutralised with 5% HCl solution and extracted thrice with ethyl acetate. The combined ethyl acetate fraction was washed with water, dried over anhydrous Na_2SO_4 and solvent removed under vacuum to yield the crude product. Further, the crude product was purified by column chromatography which afforded the desired products.

All the 3-O-acetyl amide derivatives were semi-synthesised by treating 3-O-acetyl GA with oxalyl chloride (2 equivalent) in dry dichloromethane (DCM) for 3 h followed by adding corresponding amines (1.5 equivalent) and triethylamine under nitrogen atmosphere (Figure 4(b)). The reaction mixture was stirred for 4 h at room temperature. The reaction was quenched with H_2O (10 ml), and the organic phase was separated. The aqueous phase was extracted with CH_2Cl_2 ($3 \times 30\ \text{ml}$). The combined organic phase was dried over Na_2SO_4 , filtered and evaporated under vacuum to give the crude product. The products were purified by column chromatography, which afforded the desired derivatives. All the GA derivatives were characterised on the basis of their ^1H and ^{13}C NMR spectroscopic data.

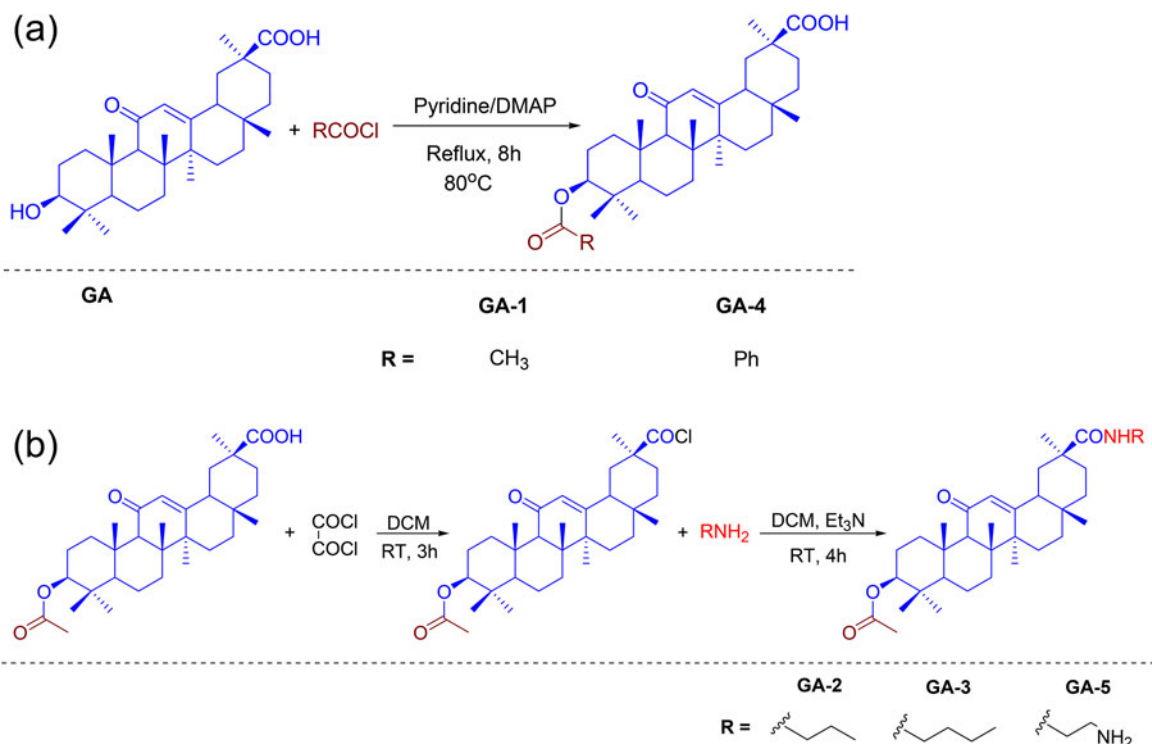


Figure 4. a. Scheme 1. Synthesis of 3-O-acyl derivatives of GA. b. Scheme 2. Synthesis of amide derivatives of 3-O-acetyl GA.

Characterisation of GL, GA and GA derivatives (GA1–GA5)

All the GA derivatives were characterised on the basis of their ¹H and ¹³C NMR spectroscopic data as given below:

GL:

¹H NMR (300 MHz, C₅H₅N) : δ 0.76–1.32 (3H each all s, 7 × tert.CH₃), 2.32 (s, 1H; 9H), 1.97 (3H, s, C-32), 4.2 (1H, dd, J = 6.8 and 8.7 Hz, 3 α-H), 5.56 (1H, s, H-12), 5.20 (1H, d, H-1'), 3.23 (1H, m, H-2'), 3.62 (1H, m, H-3' and H-4'), 4.45 (1H, d, H-5'), 12.22 (1H, s, H-6'), 4.86 (1H, d, H-1''), 3.53 (1H, m, H-2''), H-3'' and H-4''), 4.5 (1H, d, H-5''), 12.3 (1H, s, H-6'')

¹³C NMR (C₅H₅N, 75 MHz): 39.9 (C-1), 27.0 (C-2), 88.7 (C-3), 37.8 (C-4), 55.5 (C-5), 18.1 (C-6), 33.2 (C-7), 43.7 (C-8), 62.4 (C-9), 37.6 (C-10), 199.9 (C-11), 128.9 (C-12), 169.9 (C-13), 45.8 (C-14), 28.3 (C-15), 26.8 (C-16), 32.4 (C-17), 48.9 (C-18), 41.9 (C-19), 44.3 (C-20), 31.8 (C-21), 38.6 (C-22), 28.0 (C-23), 16.8 (C-24), 17.0 (C-25), 19.0 (C-26), 23.7 (C-27), 28.7 (C-28), 28.9 (C-29), 179.4 (C-30), 104.1 (C-1'), 83.1 (C-2'), 75.8 (C-3'), 71.7 (C-4'), 76.5 (C-5'), 170.5 (C-6'), 105.2 (C-1''), 75.4 (C-2''), 76.9 (C-3''), 71.8 (C-4''), 75.9 (C-5''), 170.6 (C-6'').

GA:

¹H NMR (300 MHz, CDCl₃): δ 0.76–1.32 (3H each all s, 7 × tert.CH₃) 2.32 (s, 1H, 9H), 3.36 (1H, dd, J = 6.8 and 8.5 Hz, 3 α-H) 5.62 (1H, m, H-12).

¹³C NMR: 39.9 (C-1), 27.0 (C-2), 78.1 (C-3), 37.8 (C-4), 55.5 (C-5), 18.1 (C-6), 33.2 (C-7), 43.7 (C-8), 62.4 (C-9), 37.5 (C-10), 199.1 (C-11), 128.9 (C-12), 169.9 (C-13), 45.8 (C-14), 28.3 (C-15), 26.8 (C-16), 32.4 (C-17), 48.9 (C-18), 41.9 (C-19), 44.3 (C-20), 31.8 (C-21), 37.5 (C-22), 27.8 (C-23), 16.8 (C-24), 17.0 (C-25), 19.0 (C-26), 23.7 (C-27), 28.8 (C-28), 28.9 (C-29), 179.4 (C-30).

GA-1:

¹H NMR (CDCl₃, 300 MHz): δ 0.86–1.35 (3H each, all s, 7 × tert CH₃), 4.32 (1H, m, H-3), 5.61 (1H, s, H-12), 2.04 (3H, s, H-2').

¹³C NMR (CDCl₃, 75 MHz): δ_C 39.2 (C-1), 26.8 (C-2), 81.6 (C-3), 38.5 (C-4), 55.4 (C-5), 17.8 (C-6), 33.1 (C-7), 43.6 (C-8), 62.1 (C-9), 37.3 (C-10), 200.8 (C-11), 128.8 (C-12), 169.9 (C-13), 45.9 (C-14), 28.4 (C-15), 26.8 (C-16), 32.3 (C-17), 48.6 (C-18), 41.2 (C-19), 44.2 (C-20), 31.6 (C-21), 38.1 (C-22), 28.9 (C-23), 16.8 (C-24), 17.1 (C-25), 19.1 (C-26), 23.7 (C-27), 28.9 (C-28), 29.8 (C-29), 182.2 (C-30), 171.5 (C-1'), 21.7 (C-2').

GA-2:

¹H NMR (CDCl₃, 300 MHz): δ 0.83–1.34 (3H each, all s, 7 × tert CH₃), 4.48 (1H, m, H-3), 5.60 (1H, s, H-12), 2.00 (3H, s, H-2'), 3.24 (2H, m, H-1''), 0.86 (3H, t, J = 7.5 Hz, H-3'').

¹³C NMR (CDCl₃, 75 MHz): δ 39.2 (C-1), 26.9 (C-2), 81.0 (C-3), 37.9 (C-4), 55.4 (C-5), 17.8 (C-6), 33.1 (C-7), 43.6 (C-8), 62.1 (C-9), 37.3 (C-10), 200.3 (C-11), 128.8 (C-12), 169.7 (C-13), 45.8 (C-14), 28.4 (C-15), 26.9 (C-16), 32.3 (C-17), 48.6 (C-18), 41.6 (C-19), 43.9 (C-20), 31.9 (C-21), 38.4 (C-22), 28.9 (C-23), 16.7 (C-24), 17.0 (C-25), 19.1 (C-26), 23.7 (C-27), 28.9 (C-28), 30.0 (C-29), 176.0 (C-30), 171.3 (C-1'), 21.6 (C-2'), 42.3 (C-1''), 23.9 (C-2''), 11.8 (C-3'').

GA-3:

¹H NMR (CDCl₃, 300 MHz): δ_C 0.85–1.37 (3H each, all s, 7 × tert CH₃), 4.46 (1H, m, H-3), 5.62 (1H, s, H-12), 2.02 (3H, s, H-2'), 3.29 (2H, m, H-1''), 0.85 (3H, t, J = 7.5 Hz, H-4'').

¹³C NMR (CDCl₃, 75 MHz): δ_C 39.2 (C-1), 26.8 (C-2), 81.0 (C-3), 37.9 (C-4), 55.4 (C-5), 17.8 (C-6), 33.1 (C-7), 43.6 (C-8), 62.1 (C-9), 37.3 (C-10), 200.4 (C-11), 128.8 (C-12), 169.8 (C-13), 45.8 (C-14), 28.4 (C-15), 26.8 (C-16), 32.3 (C-17), 48.6 (C-18), 42.3 (C-19), 43.9 (C-20), 31.9 (C-21), 38.4 (C-22), 28.9 (C-23), 16.8 (C-24), 17.0 (C-25), 19.1 (C-26), 23.7 (C-27), 28.9 (C-28), 30.0 (C-29), 176.0 (C-30), 171.4 (C-1'), 21.7 (C-2'), 39.8 (C-1''), 33.1 (C-2''), 20.4 (C-3''), 14.0 (C-4'').

GA-4:

^1H NMR (CDCl_3 , 300 MHz): δ 0.84–1.40 (3H each, all s, 7 \times tert CH_3), 4.79 (1H, dd, $J=6.3$ & 8.9 Hz, H-3), 5.74 (1H, s, H-12), 7.47–8.13 (5H, m, Ar-H).

^{13}C NMR (CDCl_3 , 75 MHz): δ_{C} 39.2 (C-1), 26.8 (C-2), 81.7 (C-3), 38.2 (C-4), 55.5 (C-5), 17.8 (C-6), 32.3 (C-7), 43.6 (C-8), 62.2 (C-9), 37.4 (C-10), 200.6 (C-11), 128.9 (C-12), 172.5 (C-13), 45.9 (C-14), 28.6 (C-15), 26.8 (C-16), 31.3 (C-17), 48.6 (C-18), 38.9 (C-19), 45.9 (C-20), 30.1 (C-21), 37.4 (C-22), 28.9 (C-23), 16.8 (C-24), 17.4 (C-25), 19.1 (C-26), 23.8 (C-27), 28.9 (C-28), 29.8 (C-29), 176.0 (C-30), 172.5 (C-1'), 130.0 (C-2'), 130.6 (C-3' and C-7'), 128.9 (C-4' and C-6'), 133.1 (C-5').

GA-5:

^1H NMR (CDCl_3 , 300 MHz): δ 0.99–1.35 (3H each, all s, 7 \times tert CH_3), 4.48 (1H, m, H-3), 5.67 (1H, s, H-12), 2.04 (3H, s, H-2'), 3.82 (2H, t, $J=6.6$ Hz, H-1''), 2.95 (2H, t, $J=6.6$ Hz, H-2'').

^{13}C NMR (CDCl_3 , 75 MHz): δ_{C} 39.2 (C-1), 26.8 (C-2), 80.6 (C-3), 38.0 (C-4), 55.4 (C-5), 17.8 (C-6), 33.1 (C-7), 43.7 (C-8), 62.2 (C-9), 37.4 (C-10), 201.0 (C-11), 128.7 (C-12), 171.4 (C-13), 45.9 (C-14), 28.4 (C-15), 26.8 (C-16), 32.2 (C-17), 48.6 (C-18), 41.9 (C-19), 44.1 (C-20), 30.1 (C-21), 38.4 (C-22), 29.0 (C-23), 16.8 (C-24), 17.1 (C-25), 19.1 (C-26), 23.7 (C-27), 28.4 (C-28), 29.9 (C-29), 177.0 (C-30), 171.4 (C-1'), 21.7 (C-2'), 45.9 (C-1''), 41.9 (C-2'').

In vitro cytotoxicity evaluation

Preparation of test sample solutions

The test samples 18 β -GA and its derivatives GA-1, GA-2, GA-3, GA-4 and GA-5 were weighed in micro-centrifuge tubes, and stock solutions of 20 mM were made by dissolving the samples in DMSO. Stocks are stored at -20°C . A working solution of 12.5, 25, 50, 100 and 200 μM concentrations was made by diluting the stock solution in culture medium.

Cell culture

The MDA-MB-231 (Organism: Homo sapiens, Tissue/site: breast metastatic, Cell type: epithelial, TNBC) were procured from American Type Culture Collection (ATCC) and cultured as per manual instructions. The cells were cultured and maintained in RPMI-1640 medium at 37°C and 5% $\text{CO}_2/95\%$ air in a humidified incubator and were regularly examined microscopically for stable phenotype.

SRB assay

Addition of cells: the cells were dispensed in a flat-bottom 96-well plate. To each well, 100 μl of the cell suspension containing 10,000–15,000 cells was added. Further, the cells were incubated at 37°C in 5% $\text{CO}_2/95\%$ air concentration for 24 h, prior to the addition of test samples.

Test samples addition: a working solution of 100 μl of test sample was added to the cell monolayer to give a final concentration of 200 μM . A series of four dilutions 12.5, 25, 50 and 100 μM for each derivative in three replicates were included.

Negative (Vehicle) controls: in every assay plate, DMSO was added in 0.1% concentration as vehicle control. The final concentration of DMSO was 0.1% in all assay wells. Finally,

the plates were incubated at 37°C in 5% CO_2 concentration for 48 h.

Addition of sulphorhodamine B assay and colorimetric reading: once the treatment period was done, after 48 h incubation, cold 50% trichloroacetic acid (TCA Sigma-Aldrich (St. Louis, MO), 50 μl /well) was added on top of the medium to fix the cells attached to substratum and incubated for 1 h at 4°C . After that, a five-time gentle wash was given to the plate on a slowly running tap water to remove dead cells, culture medium and TCA. After washing, the plates were air-dried. Further, 50 μl /well of SRB solution was added to the dried plate and left at room temperature for 30 min. After incubation, unbound SRB dye was removed by four to five times washing with 1% (v/v) glacial acetic acid. Plates were allowed to air dry at room temperature. Further, 150 μl of 10 mM Tris base solution was added to each well to solubilise the protein-bound dye, and plate is shaken for 15 min on a gyratory shaker. Finally, the absorbance was taken at 510 nm using a plate reader.

Data analysis

Percentage of cell growth inhibition in the presence of the test sample is calculated as follows:

$$\text{Percentage of cells killed} = 100 - \left[\frac{\text{Mean OD}_{\text{test}}}{\text{Mean OD}_{\text{control}}} \right] + 100$$

Identification of therapeutic targets for GA in MDA-MB-231

Based on recently published report, a 48 h treatment of MDA-MB-231 cells with 20 μM /l GA attenuate cellular glutathione (GSH) level causes apoptosis in TNBC cancers (Cai et al., 2017). The cellular GSH level is controlled by GLO-I and Topoisomerase-II as reported by Silva et al. (2013) and Cameron et al. (1999). However, a web-based target identification tools, viz., Stitch-DB (<http://stitch.embl.de>) and Swiss target prediction, identified hydroxysteroid 11-beta-dehydrogenase-1(1HSD1) as possible binding targets for GA (Figure S8, Supplementary information). Therefore, an approach of molecular docking-based screening was applied for GA and derivatives to rank possible MDA-MB-231 targets viz., GLO-I, TOPO-II and 11HSD1 based on their degree of binding energies.

Molecular docking and APF-based scoring

The crystal structure of GLO-I (PDB: 4PV5) bound with GA (2.3 \AA resolution) was retrieved from protein database (Zhang et al., 2015). The GA derivative structures were converted to ICM object using ICM Molsoft-chemist v3.8-6 (2018) (Molsoft LLC, San Diego, CA). It uses Monte Carlo minimisation in the APF's potentials in conjunction with standard MMFF94 force field energies. The method was developed by Totrov (2008), and Grigoryan et al. (2010). Protein-ligand docking tool FlexX provided by LeadIT software v2.1.6, 2017 (BioSolveIT GmbH, Sankt FeAugustin,

Table 1. The statistical parameters and their calculated values for training set of QSAR model.

S. No.	Statistical qualities (training set)	Parameter explanation	Value	Reported acceptable range
1.	N	Training set, 70% whole dataset	100	
2.	r^2	Regression coefficient for training set	0.8442	>0.6 (Golbraikh & Tropsha,2002)
3.	q^2	Regression coefficient for leave-one-out (LOO) validation	0.8282	>0.5 (Golbraikh & Tropsha,2002)
4.	F -test	Fisher test	101.656	High value is good
5.	Z score for r^2	Randomisation test for r^2	14.767	>1.28 at SD 0.10, (Zheng & Tropsha,2000)
6.	Z score for q^2	Randomisation test for q^2	14.617	>1.28 at SD 0.10, (Zheng & Tropsha,2000)
7.	r^2_0	Correlation regression without intercept	0.8439	
8.	r^2_0	Reciprocal of r^2_0 , i.e. taking predicted value in x-axis while calculation	0.8219	
9.	r^2_m	Correlation between actual and predicted values with intercept and without intercept while calculation	0.8301	>0.5 (Ojha et al.,2011)
10.	r^2_m	Reciprocal of r^2_m , i.e. taking predicted value in x-axis	0.7181	>0.5 (Ojha et al.,2011)
11.	r^2_m	Average of r^2_m and r^2_m	0.7741	>0.5 (Ojha et al.,2011)
12.	Δr^2_m	Absolute difference between r^2_m and r^2_m	0.1119	<0.2 (Ojha et al.,2011)

Germany, www.biosolveit.de/LeadIT) was used to perform molecular interaction study as proposed by Kramer, Rarey, and Lengauer (1999). The amino acids within 10 Å region from GA active site of enzyme GLO-I were selected to get more flexibility in interaction study. The number of pose generation was set at 10, and computed pose with minimum energy (RMSD) was selected for comparative study.

Computational assessment for oral bioavailability and toxicity

All the five GA derivatives were also studied for their oral bioavailability by calculating various pharmacokinetic parameters such as plasma protein binding, blood–brain barrier penetration capacity, intestinal absorption, hepatotoxicity and oral bioavailability. Furthermore, the derivatives were evaluated for toxicity risk screening using Discovery Studio v3.5 TOPKAT (Toxicity Prediction by Komputer Assisted Technology) tool. TOPKAT is a Quantitative Structure Toxicity Relationship (QSTR)-based tool developed by Accelrys Inc. USA (<http://accelrys.com>) licensed to CSIR-CIMAP, Lucknow (www.cimap.res.in). The module utilises highly robust and cross-validated QSTR models to predict toxicity. The module applies patented Optimal Predictive Space (OPS) which is a unique multivariate descriptor space for result interpretation (Ensein, 1988; Ensein et al., 1987). The module computes

the toxic and environmental effects of compounds exclusively from their chemical structures. DS-TOPKAT module searches fragments within query molecule based on molecular fingerprint similarity with the training set compounds. The TOPKAT toxicity prediction results for unknown compound are calculated based on probability score, Bayesian scores and Mahalanobis distance (structure similarity) from the centre of the training set compounds. DS-TOPKAT gives predictions for a range of toxicological end points, including mutagenicity, developmental toxicity, rodent carcinogenicity, rat chronic Lowest Observed Adverse Effect Level (LOAEL), rat Maximum Tolerated Dose (MTD) and rat oral LD₅₀ (Table S6, Supplementary material).

Results and discussion

2D-QSAR model development and validation results

The developed QSAR identifies activity-inducing features of 144 MDA-MB-231 inhibitors selected in the model development (Tables S1 and S2, Supplementary material). The developed QSAR model was validated through various statistical approaches viz., LOO, external test set prediction (r^2_{pred}), Z scores and r^2_m matrix calculation. The results of statistical parameters are summarised in Table 1.

QSAR MLR equation

$$\begin{aligned} \text{PIC}_{50(\mu\text{M})} = & 0.0016 + 7.2421(\text{Epsilon4}) \\ & + 1.2894(\text{chiV3Cluster}) - 0.7603(\text{T}_N\text{N}_5) \\ & + 0.1635(\text{Nitrogen count}) + 2.3425(\text{chi3chain}) \end{aligned} \quad (6)$$

where N (training set, 70% of 144 MDA-MB-231 inhibitors) = 100, n (test set, 30% of 144 MDA-MB-231 inhibitors) = 44, r^2 (regression coefficient for training set) = 0.8442, $R2se$ = 0.3063, q^2 (regression coefficient for LOO validation) = 0.8282, $q2se$ = 0.3214, Fisher test = 101.6555, predicted r^2 (regression coefficient for external test set) = 0.7532, pred $r2se$ = 0.3659, Z score R^2 = 14.76689, Z score q^2 = 14.61679 and Z score pred r^2 = 4.11170.

The QSAR model attains good correlation coefficient of 0.84 for training set of 100 inhibitors. The fitness plot between observed and predicted PIC₅₀ values is presented in

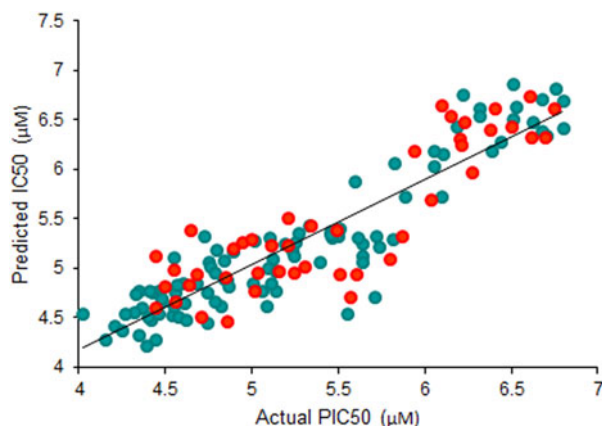


Figure 5. Regression curve (MLR model) for actual and predicted PIC₅₀ of 144 natural scaffold-based inhibitors of metastatic TNBC cell line MDA-MB-231. Training and test set compounds are highlighted in blue and red dots, respectively.

Table 2. The statistical parameters and their calculated values for test set of QSAR model.

S. No.	Statistical qualities (test set)	Parameter explanation	Value	Reported acceptable range
1.	n	Test set, 30% of whole dataset	44	
2.	r_{pred}^2	Regression coefficient for test set	0.7532	>0.5 (Golbraikh & Tropsha,2002)
3.	Z score for r_{pred}^2	Randomisation test for r_{pred}^2	4.11170	>1.28 at SD 0.10, (Zheng & Tropsha,2000)
4.	r_0^2	Correlation regression without intercept	0.7410	
5.	r_0^2	Reciprocal of r_0^2 , i.e. taking predicted value in x-axis while calculation	0.7299	
6.	r_m^2	Correlation between actual and predicted values with intercept and without intercept while calculation	0.6702	>0.5 (Ojha et al.,2011)
7.	r_m^2	Reciprocal of r_m^2 , i.e. taking predicted value in x-axis	0.6384	>0.5 (Ojha et al.,2011)
8.	r_m^2	Average of r_m^2 and r_m^2	0.6543	>0.5 (Ojha et al.,2011)
9.	Δr_m^2	Absolute difference between r_m^2 and r_m^2	0.0317	<0.2 (Ojha et al.,2011)

Figure 5. The high value of cross-validation (LOO) regression (q^2) of 0.82 indicates that training set compounds (blue dots) exhibit less statistical noise. Regression coefficient for randomly selected 44 external test set (Pred R^2) was found to be 0.75. The test set regression infers the good predictability of model for unknown compounds (red dots), and a small measure of error (0.0016) indicates data comprehensiveness. Additionally, an even distribution of residual values around the axis line indicates good model quality (Figure S3, Supplementary material). Furthermore, high value for Fishers test, $F = 101.65$, again verified robustness of the model. Also, high Z scores of 14.76689, 14.61679 and 4.11170 for r^2 , q^2 and pred r^2 , respectively, supported the good model quality. The computed statistical qualities for training and test sets are summarised in Tables 1 and 2 with their reported cut-off values.

Based on Ojha et al. (2011) the r_{pred}^2 is not a true evidence for model prediction ability, since r_{pred}^2 depends on training set mean and therefore greatly influenced by training set and test set selections. However, r_m^2 matrix shows the predictability of the model for whole dataset. For test set, the acceptable range for parameters, r_{pred}^2 , r_m^2 and r_m^2 , is 0.5, Δr_m^2 should be less than 0.2 and r_{mbar}^2 should be more than 0.5 (Ojha et al., 2011). In the present case, the r^2 , r^2 (LOO), r_m^2 , r_m^2 , r_{mbar}^2 and Δr_m^2 values for training set are 0.84, 0.82, 0.83, 0.71, 0.77 and 0.11, respectively (Table 1). All statistical parameters for training set were found within their cut-off limits (Table 1). For test set, r_{pred}^2 , r_m^2 and r_m^2 were found to be 0.75, 0.67 and 0.63, respectively. The calculated values of r_{mbar}^2 and Δr_m^2 for test set are 0.65 and 0.03, respectively, that are within their cut-off limits (Table 2). The computed r_m^2 matrix validates the reliability and robustness of developed QSAR model for anticancer activity prediction of unknown compounds. Also, the model identified features help to explore the structure-based inhibition mechanism of MDA-MB-231inhibitors.

QSAR model identified 2D structural properties and description

Generated Equation (6) explains the model extracted out of five important descriptors that determine the cytotoxic potential of MDA-MB-231 inhibitors which are (i) Epsilon4 that signifies the measure of electronegativity count, (ii) ChiV3cluster indicates valence molecular connectivity index, (iii) chi3chain represents retention index for three-membered

ring, (iv) TNN5 stands for nitrogen atoms separated through 5-bond distances and (v) nitrogen counts, i.e. number of nitrogen atoms in the molecule.

The topochemical descriptor Epsilon4 signifies the measure of electronegativity count and contributes 11% to the biological activity (PIC₅₀). The descriptor chiV3Cluster belongs to valence molecular connectivity index of third-order cluster (Shen et al., 2002). It is known that molecular connectivity indices are mostly successful among other topological properties in compound property estimation, since these indices are based on contingent chemical, structural and mathematical ground. The important advantage of the molecular connectivity model comprises its flexibility, to quantify general as well as local structural properties. The percentage contribution for identified descriptors is presented in Figure S4 under supplementary material. Figure S4 describes that descriptor chiV3Cluster contributes 30% to biological activity of training set compounds, whereas TNN5 descriptor that defines two nitrogen atoms separated through 5-bond distances showed inverse relationship to the biological activity. However, descriptor nitrogen count shows positive effect and contributes 21% to the activity (PIC₅₀). Additionally, Equation (6) indicates that nitrogen-containing functional groups may increase the biological activity, though chemical fragments containing nitrogen atoms departed by long chain (TNN5) might not be very favourable. Overall, the model suggests that maximum contribution hails from the descriptors chiV3Cluster, Epsilon4 and nitrogen count (Figure S4, Supplementary material).

2 D-QSAR model AD assessment results

A PCA analysis indicates that 44 test set compounds fall within the structure space of training set compounds. Figure 6 shows generated PCA graph for training set (blue sphere) and test set (yellow sphere). The figure illustrates a uniform distribution of test set within the vector space of training set compounds. The figure defines the test set as a true representative of training set. Also, a broad biological activity space of 10^{-1} – 10^1 μM for training and test sets indicates that the data were comprehensive. A correlation matrix for PIC₅₀ and extracted descriptor (Epsilon4, chiV3Cluster, T_N_N_5, Nitrogen count and chi3chain) was also generated (Table S4, Supplementary material). It showed that the developed model based on the selected descriptors is well established.

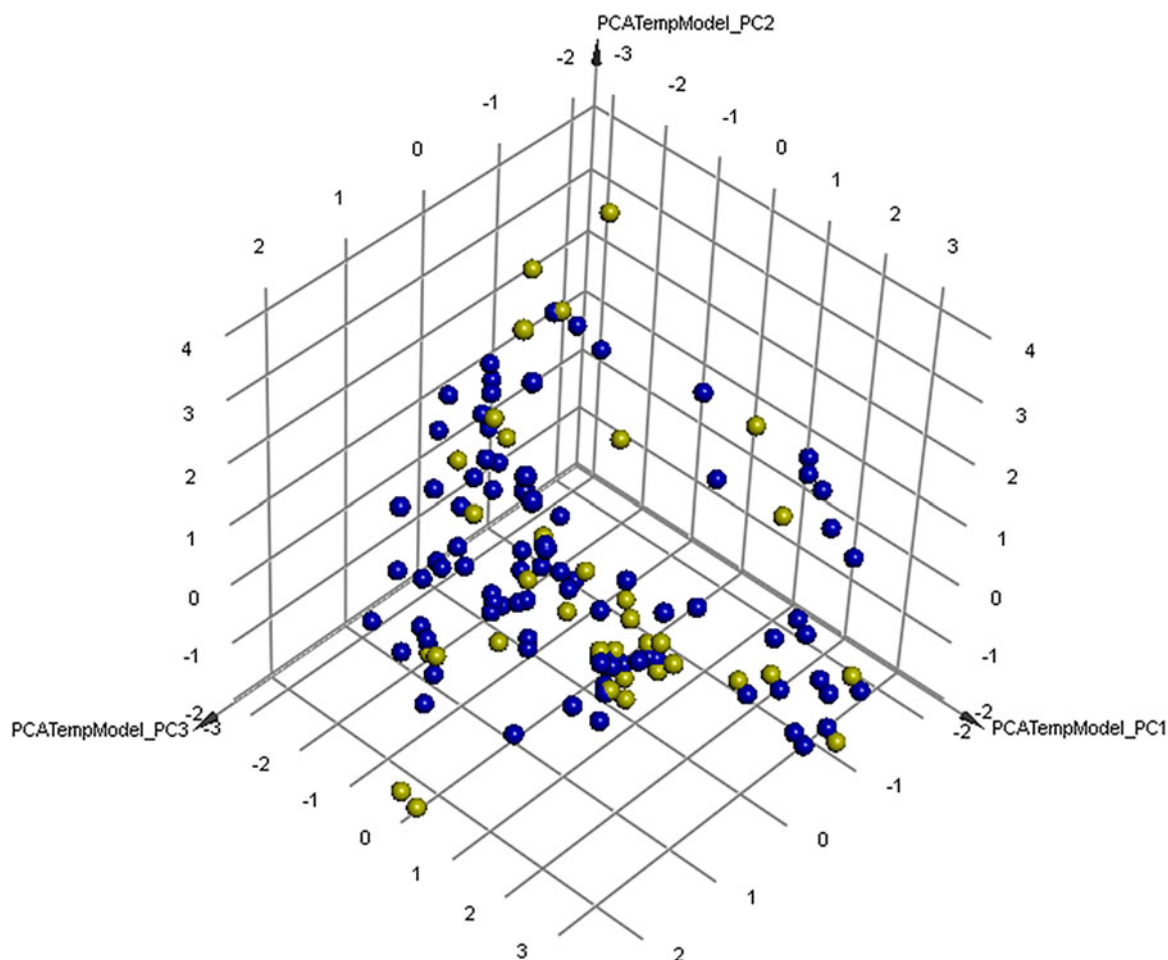


Figure 6. Generated 3D Principal Component Analysis (PCA) to ascertain uniform distribution of test set (yellow sphere) within property vector space of training set (blue sphere).

A UPGMA-based hierarchical cluster analysis (Tanimoto structure similarity distance 0–0.7) of 144 dataset compounds indicate that training set, test and five GA derivatives come within the AD of QSAR model. Also, the heat map generation for chemical properties *viz.*, molecular weight, logP, polar surface area, maximum ring size, minimum ring size, maximum fused rings and number of rotatable bonds also indicates the optimal chemical property range (Figure S1, Supplementary material).

Identified 3D structural property fields through 3D-QSAR studies

At this point, an attempt was made to generate APF-based 3D-QSAR analysis to systematically describe the structural atomic field level of novel GA derivatives. The 3D-QSAR analysis performed by Atom Property Fields (APF) methods was developed by Totrov (2008). For this purpose, a set of congeneric series of 42 GA derivatives with *in vitro* inhibition activity against MDA-MB-231 cell line were selected. The dataset structures were flexibly aligned to the generated property fields of co-crystallised GA on GLO-I. The co-crystallised GA on GLO-I binding site is depicted in Figure 7. The generated model presented a good regression coefficient of 0.96 for training set compounds. Also, the external test set-

based regression reverted a good predictive regression coefficient of 0.82 (Tables 3 and 4). The regression plots between observed and predicted PIC_{50} for training and test sets are shown in Figures 8 and 9. The calculated results of statistical parameters training and test sets are compiled in Tables 3 and 4, respectively. All statistical properties were found within their cut-off limit. The results indicate that generated model is robust enough to give consistent prediction for novel GA derivatives. Henceforth, novel GA compounds namely GA-1, GA-2, GA-3, GA-4 and GA-5 were aligned on training set and their PIC_{50} values were predicted through developed 3D-QSAR model. The model presented IC_{50} for GA derivatives ranges from 44.26 to 103.75 μ M. Based on 3D-QSAR model predicted results, it has been expected that designed derivatives may show moderate activity against TNBC cell line.

Semi-synthesis and SRB-based *in vitro* cytotoxicity assay results for GA derivatives GA-1, GA-2, GA-3, GA-4 and GA-5 against metastatic TNBC cell line MDA-MB-231

Our design concept for GA derivatives, GA-1, GA-2, GA-3, GA-4 and GA-5, was to introduce structural variations at C-3 and C-30 positions to improve the anticancer efficiency. The 2D-QSAR model extracted descriptor Epsilon4 indicates that

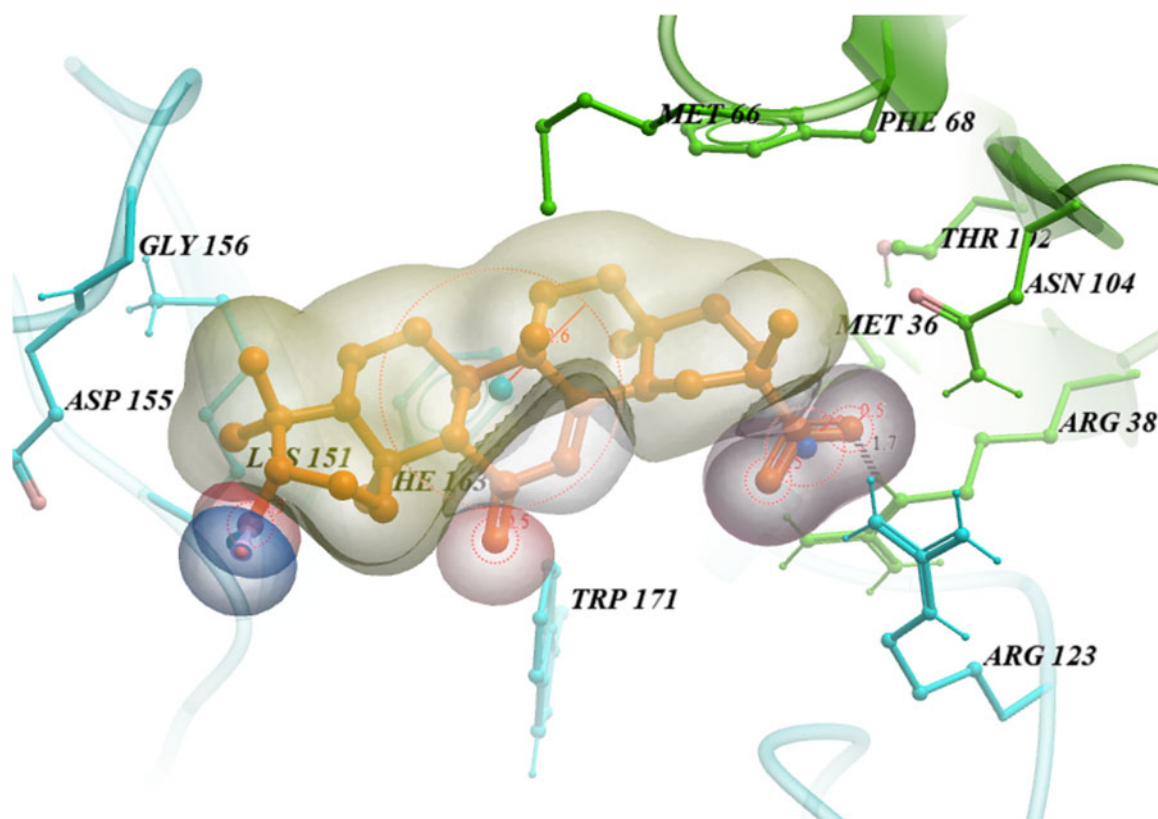


Figure 7. The zoom-in view of glyoxalase-I bound glycyrrhetic acid atomic property fields represented with CPK form. The white blob represented equipotential contour of lipophilic property. Red and blue blobs on carbon-3 (C-3) and carbon-30 (C-30) represented equipotential contour of hydrogen bond acceptor and donor, respectively.

Table 3. The statistical parameters and their calculated values for training set of APF 3D-QSAR model.

S. No.	Statistical qualities (training set)	Parameter explanation	Value	Reported acceptable range
1.	N	Training set	37	
2.	SelfMAE	Mean absolute error	0.0771089	
3.	Test r^2	Regression coefficient for training set	0.963138	>0.6 (Golbraikh & Tropsha,2002)
4.	selfRMSE	Root-mean-square error	0.0999803	
5.				
6.	Self-spearman	Spearman regression coefficient	0.98056	
7.	r_0^2	Correlation regression without intercept	0.9632	
8.	r_0^2	Reciprocal of r_0^2 , i.e. taking predicted value in x -axis while calculation	0.9617	
9.	r_m^2	Correlation between actual and predicted values with intercept and without intercept while calculation	0.90555	>0.5 (Ojha et al.,2011)
10.	r_m^2	Reciprocal of r_m^2 , i.e. taking predicted value in x -axis	0.9203	>0.5 (Ojha et al.,2011)
11.	r_m^2	Average of r_m^2 and r_m^2	0.9129	>0.5 (Ojha et al.,2011)
12.	Δr_m^2	Absolute difference between r_m^2 and r_m^2	-0.0147	<0.2 (Ojha et al.,2011)

Table 4. The statistical parameters and their calculated values for test set of APF 3D-QSAR model.

S. No.	Statistical qualities (test set)	Parameter explanation	Value	Reported acceptable range
1.	n	Test set	5	
2.	SelfMAE	Mean absolute error	0.2772	
3.	r^2	Regression coefficient for test set	0.82	>0.5 (Golbraikh & Tropsha,2002)
4.	test r^2 (LOO)	Regression coefficient for test set leave-one-out (LOO) validation	0.6502	>0.6 (Golbraikh & Tropsha,2002)
5.	testRMSE	Root-mean-square error	0.3279	
6.	Test Spearman	Spearman regression coefficient	0.8207	
7.	r_0^2	Correlation regression without intercept	0.8196	
8.	r_0^2	Reciprocal of r_0^2 , i.e. taking predicted value in x -axis while calculation	0.7659	
9.	r_m^2	Correlation between actual and predicted values with intercept and without intercept while calculation	0.8043	>0.5 (Ojha et al.,2011)
10.	r_m^2	Reciprocal of r_m^2 , i.e. taking predicted value in x -axis	0.6269	>0.5 (Ojha et al.,2011)
11.	r_m^2	Average of r_m^2 and r_m^2	0.7157	>0.5 (Ojha et al.,2011)
12.	Δr_m^2	Absolute difference between r_m^2 and r_m^2	0.17737	<0.2 (Ojha et al.,2011)

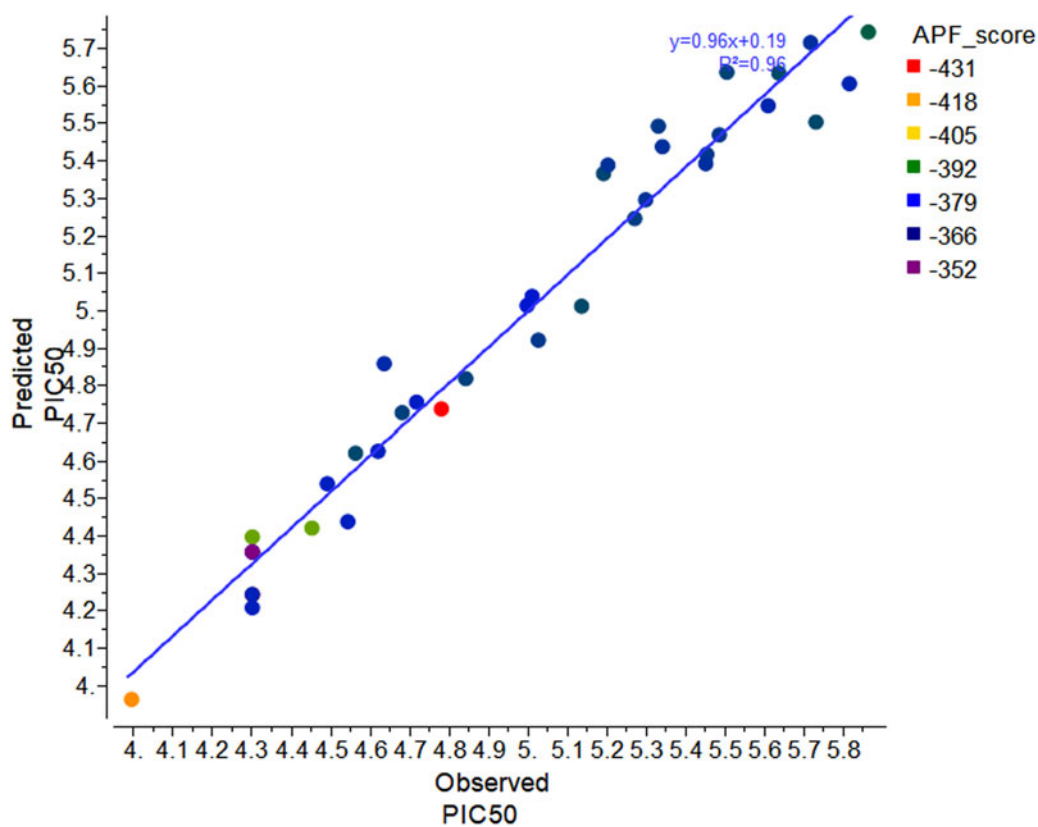


Figure 8. Regression plot for training set compounds for 3D-QSAR model. Different compounds were highlighted with different colour codes based on APF score of training set compounds.

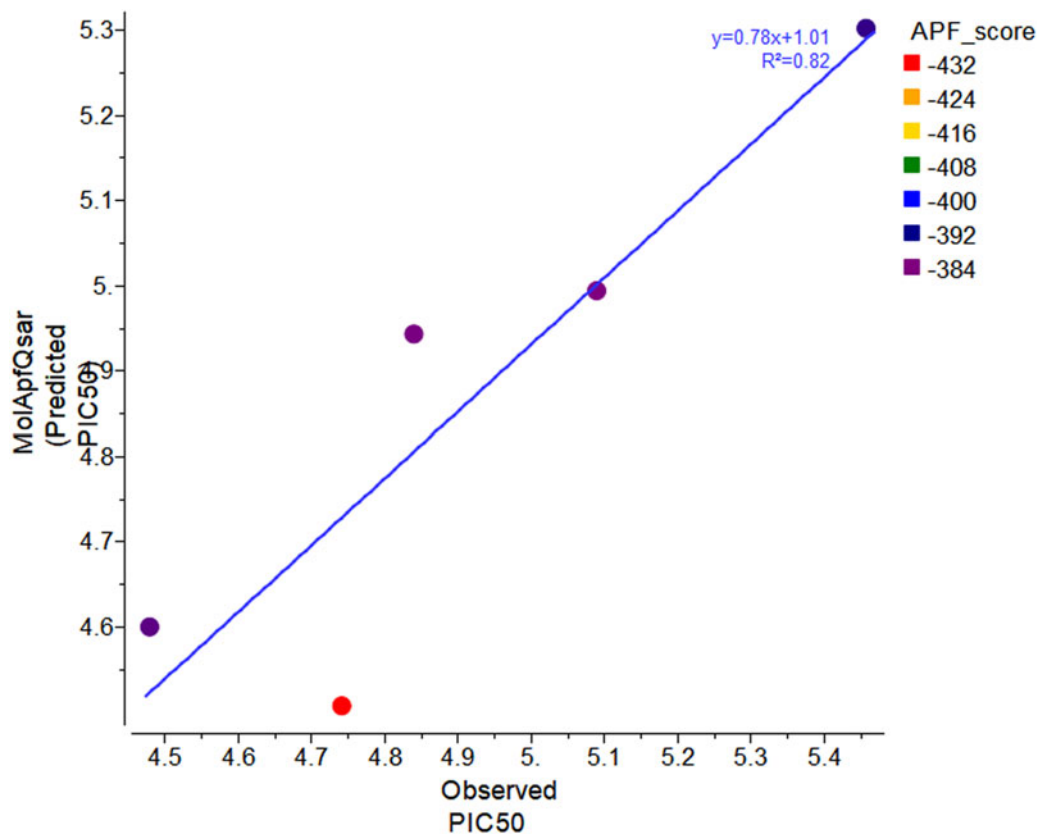


Figure 9. Regression plot for test set compounds for 3D-QSAR model. Different compounds were highlighted with different colour codes based on their APF score.

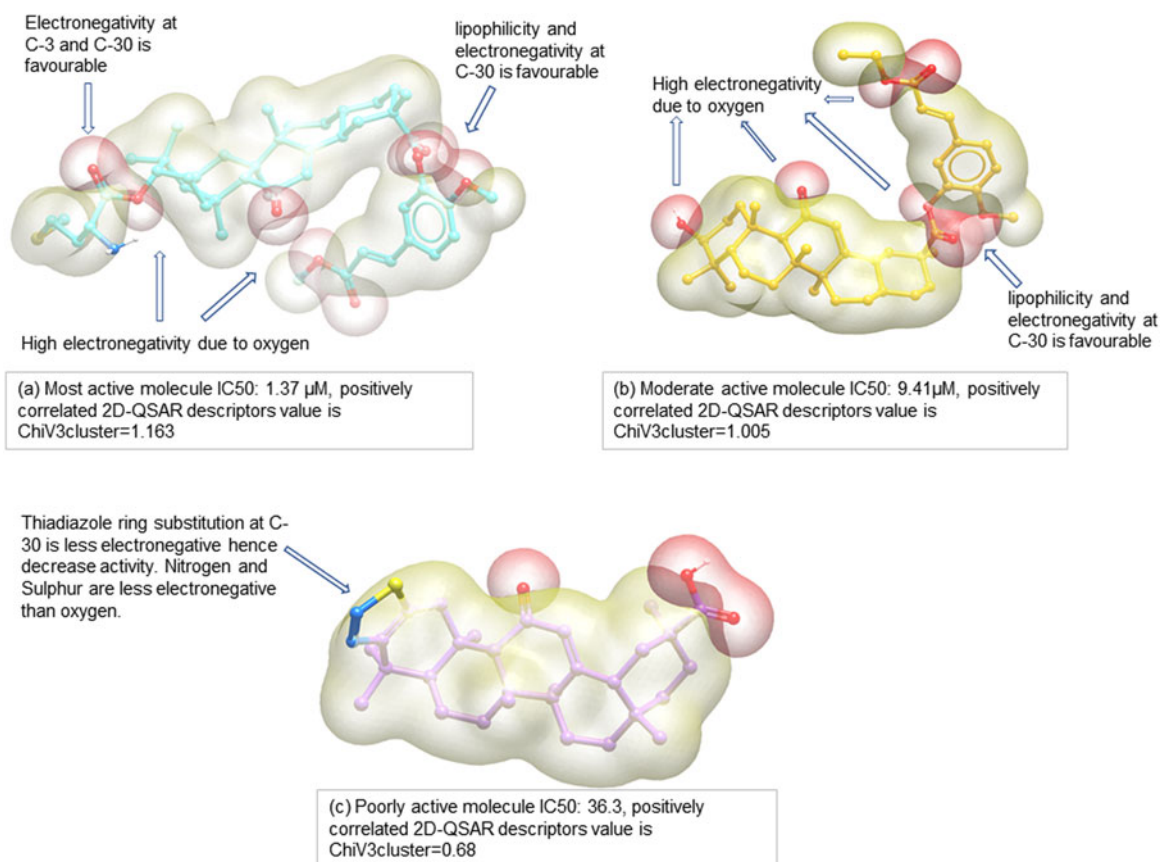


Figure 10. The APFs map of equipotent contour of most active, moderately active and least active GA derivatives used in 2D- and 3D-QSAR modelling. (a) Most active GA derivative with IC_{50} : 1.37 μM , showing modification at C-3 and C-30 carbons with lipophilic branches. The most active GA derivative exhibits high 2D-QSAR descriptor $ChiV3cluster$ value; 1.163 (b) moderately active GA derivative with IC_{50} : 9.41 μM , showing modification at C-30 carbon with lipophilic fragment. The molecule exhibits moderate $ChiV3cluster$ value; 1.005 (c) least active GA derivative with IC_{50} : 50 μM , showing modification of C-3 carbon with 1,2,3 thiaziazol group decreases the overall activity of molecule. Least active derivative showed low 2D-QSAR descriptor $ChiV3cluster$ value 0.68. The white blob represented equipotential contour of lipophilic property. Red blob represented equipotential contour of hydrogen bond acceptor.

electronegativity on GA causes favourable effects on biological activities. A positive correlation with electronegativity was also found in 3D-QSAR studies (Figure 10). 2D-QSAR descriptor $ChiV3cluster$ suggested that less branching at GA scaffold is favourable. Hence, small fragments *viz.*, propyl amide, butyl amide and amino ethyl amide were substituted at C-30 carbon (GA-1, GA-2, GA-3 and GA-5). The structures of prepared derivative are given in Figure 4(a,b). However, 3D-QSAR-based property fields suggest that lipophilicity and electronegativity play governing role in anticancer properties of GA derivatives. Therefore, a derivative with benzoate group substitution at C-3 position was also prepared (GA-4). The detailed analysis of 2D- and 3D-QSAR-based structure–activity relationships is illustrated in Figure 10. Based on QSAR model studies, five novel derivatives of GA named GA-1, GA-2, GA-3, GA-4 and GA-5 were semi-synthesised with modifications at C-3 and C-30 positions and screened through the developed model (Figure 4(a,b)).

Further, a dose-dependent *in vitro* cytotoxicity of GA and derivatives was investigated against metastatic TNBC cell line MDA-MB-231. The 48 h exposure of derivatives GA-1, GA-3 and GA-4 inhibits MDA-MB-231 cells with IC_{50} 76.5, 91.79 and 116.07 μM , respectively (Table 5). Derivative GA-1 was found the most active as it showed the most cytotoxic

activity against MDA-MB-231 cells. However, GA-2 and GA-5 were found least effective as they indicated 35.56 and 25.63% cancer cells inhibition at maximum concentration of 200 μM , respectively. GA-3 and GA-4 showed moderate inhibition potentials of 91.79 and 116.07 μM , respectively.

2 D and 3D-QSAR model results and their correlation with *in vitro* activity of GA-1, GA-2, GA-3, GA-4 and GA-5

In order to prospectively validate the generated 2D- and 3D-QSAR models, the anticancer activities of the novel GA derivatives were calculated and compared with the *in vitro* activity. Relevance for 2D-QSAR was based on the data homogeneity constructed using natural scaffold-based training set with fused ring structures (two to five rings) (Figure S3, Supplementary material). The GA derivatives are pentacyclic triterpene. The 2D-QSAR model predicted that IC_{50} was in the range of 49–18 μM . Not much difference in activities between GA-1, GA-2 and GA-3 was predicted because of their high topological similarity. Hence, an atomic potential field-based 3D-QSAR was applied to recognise biologically significant structural features of GA derivatives. The dataset for 3D-QSAR

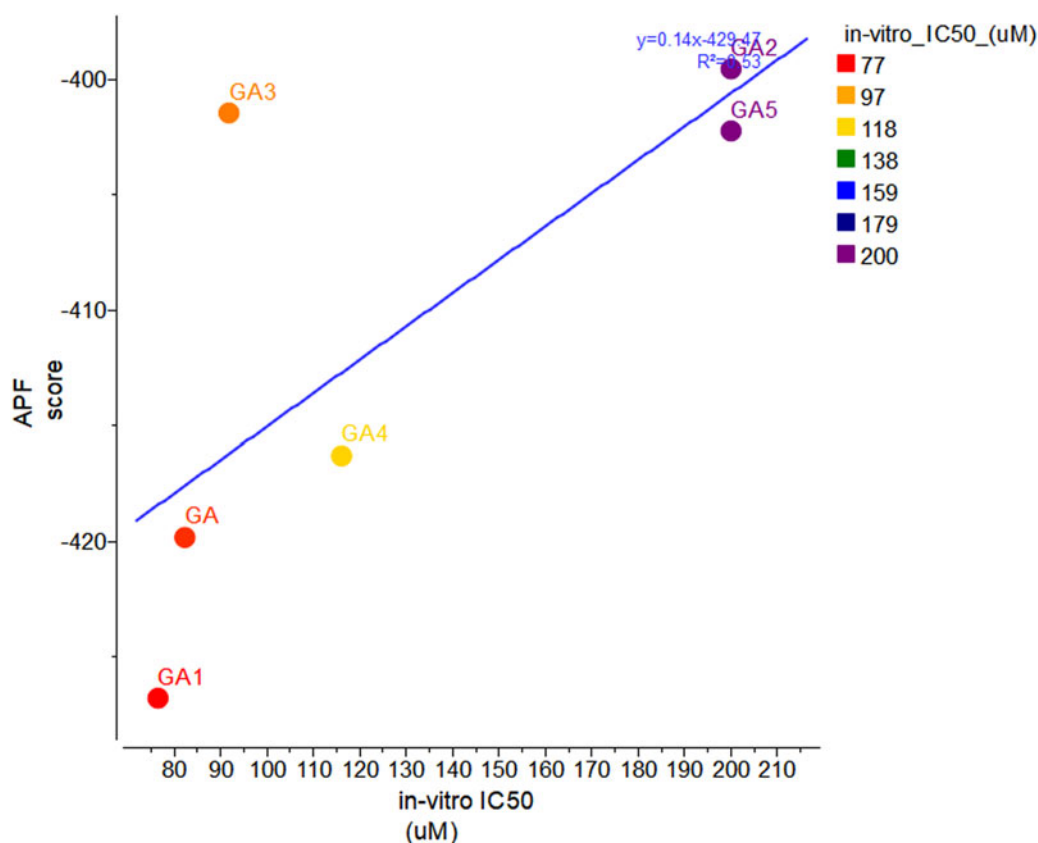


Figure 11. A plot between *in vitro* activity and APF scores of GA derivatives, GA-1, GA-2, GA-3, GA-4 and GA-5. The figure illustrates that there is a positive correlation between atomic property fields and breast cancer inhibition of GA derivatives.

was based on congeneric series of GA derivatives. Consequently, it has been expected that APF-based 3D-QSAR model might present more specific results for derivatives. The model presented IC₅₀ for GA derivatives ranges from 44.26 to 103.75 μ M. The 3D-QSAR prediction provided more variations in PIC₅₀ of GA derivative. Also, a positive correlation between APF scores and *in vitro* activity indicates correlation between APFs score in negative and cytotoxic activity (Figure 11).

The results of 2D- and 3D-QSAR models and *in vitro* activities on GA derivatives indicate that 3-O-acyl derivative named GA-1 was more significant in terms of biological activity. It has been found that modification at C-30 carboxylic group with amide group in GA-2, GA-3, GA-4 and GA-5 resulted in a decrease in cytotoxic potential against MDA_MB-231. Moreover, APF 3D-QSAR-based predicted activities for derivatives were found comparable to their *in vitro* IC₅₀ values. Therefore, the results indicate that APF-based 3D-QSAR performed well in predicting the biological activities of studied compounds.

Mode of action study, binding energy and APF scores of GA-1, GA-2, GA-3, GA-4 and GA-5 with anticancer target GLO-I

Breast cancer majorly depends on glycolysis as energy source based on Warburg effect (Fonseca-Sánchez et al., 2012; Sullivan, Gui, & Vander Heiden, 2016). During glycolysis, a highly reactive compound known as methyl glyoxalase is formed. GLO-I metabolises and inactivates methylglyoxalase

produced through glycolysis, making GLO-I inhibitors as potential anti-tumour agents (Cai et al., 2016; Silva et al., 2013). Inhibition of GLO-I resulted in the accumulation of α -oxoaldehydes at cytotoxic levels and reverse multi-drug resistance (MDR). RT-PCR and Western blot analysis of metastatic breast cancer MDA-MB-231 often show high expression of GLO-I. Additionally, knockdown study on GLO-I enzyme suppresses migration and invasion and promotes apoptosis in metastatic breast cancer cells (Guo et al., 2016). Conventional and most considered GLO-I inhibitors are coenzyme GSH analogs, which exhibits efficient inhibition *in vitro* (Silva et al., 2013) and reverse MDR; thus, GLO-I inhibitors have been proposed as efficient anti-tumour agents. However, these GSH-based inhibitors suffer from poor pharmacokinetic properties and are difficult to use as lead structure for the further design of small molecule. Alternatively, non-GSH analog natural inhibitors include flavonoids, methylgerfelin (MGI), indomethacine, zopolrestat and curcumin and its derivatives showed good pharmacokinetic properties (Zhang et al., 2015). Notably, the carboxylic group of these molecules mimics glycyl and γ -glutamyl residue moieties of GSH to form hydrogen bonds with the glycyl and glutamyl sites, respectively, in the GSH binding site (Figure S5, Supplementary material). Recently, Cai et al., (2017) have reported MDA-MB-231 treatment with 20 μ M/I of GA for 48 h which causes apoptosis by inducing GSH inhibition.

Here, a docking-based screening of GA derivatives presented good binding energy with predicted target GLO-I in comparison to other two targets 11HSD1 and TOPO-II (Table

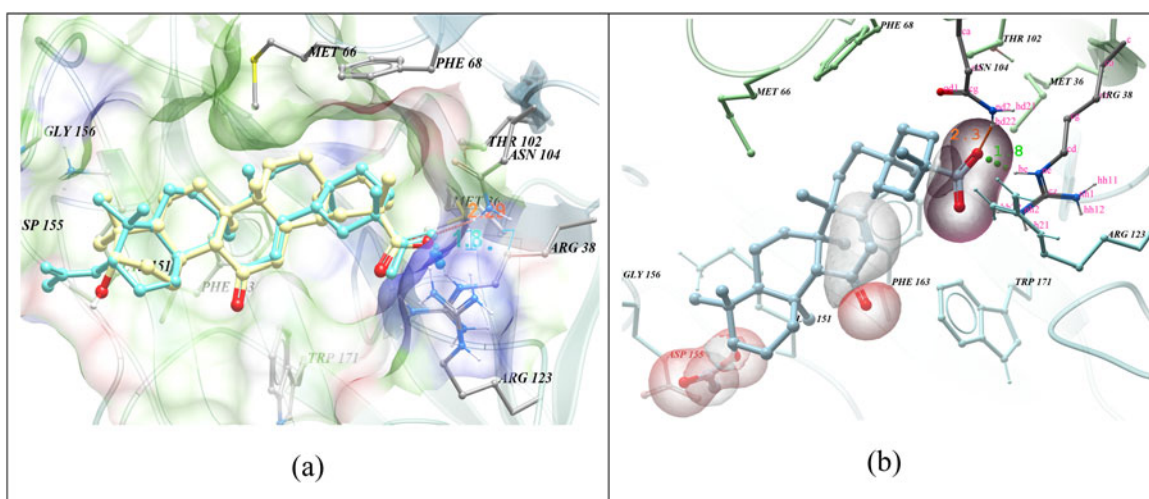


Figure 12. Atomic property field-based docking model of derivative GA-1 (cyan colour ball and stick form, $IC_{50} = 76.5 \mu\text{M}$) bound to Glyoxalase-I GSH binding site. (a) The superimposition between glycyrrhetic acid (white ball and stick form, oxygen atoms highlighted with red) and GA-1 illustrates that the GA-1 exhibits similar binding conformation as that of co-crystallised glycyrrhetic acid. (b) A close view of GA-1 and key amino acid residue binding. Orange and green ball lines represent hydrogen bonds with GA-1 C-30 negatively charged oxygen atom ($-\text{RCOO}^-$) with polar hydrogens of amino acid residues ARG-38 (1.8 Å) and ASN-104 (2.3 Å), respectively.

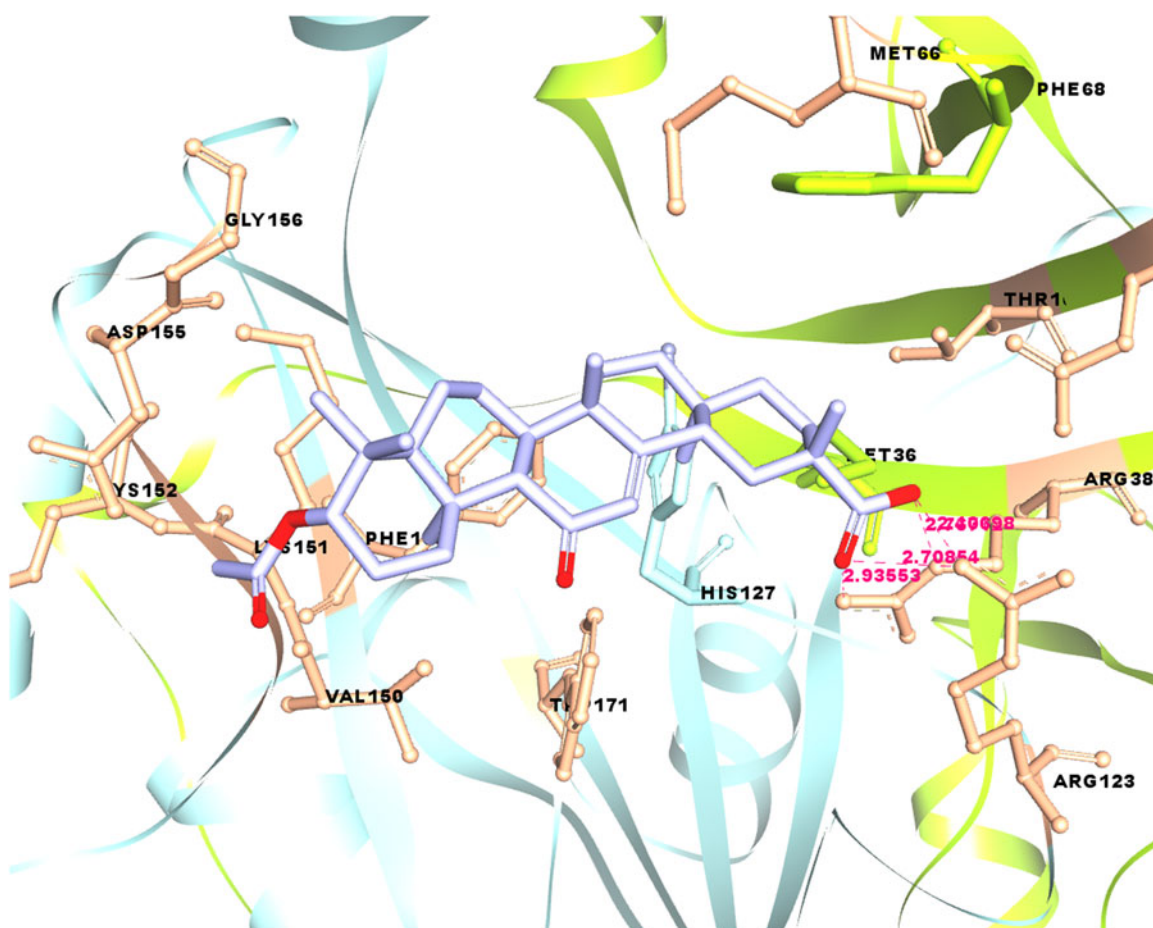


Figure 13. Proposed binding pose of derivatives GA-1 (with binding energy -16.997 kJ/mol) on GSH binding site of Glyoxalase-I. C-30 carboxylic group of GA-1 making two hydrogen bonds with key amino acids ARG-38 and ARG-123.

S5, Supplementary material). Therefore, a non-GSH-based ligand namely GA-1, GA-2, GA-3, GA-4 and GA-5 was analysed for their binding affinity towards GLO-I enzyme.

The mammalian GLO-I exhibits two binding sites with zinc as a cofactor at its catalytic binding site. One of the binding

sites is glycyI site specific for GSH binding as presented in Figure 7. The key amino acid residues for glycyI site are LYS150A, GLY155A, LYS156A, and LEU160A AND PHE162A. However, glutamyl site exhibited key amino acids, ARG37B, ASN103B and ARG122A. While the Zn^{2+} catalytic site

Table 5. SRB-based *in vitro* cytotoxic activity of GA, GA-1, GA-2, GA-3, GA-4 and GA-5 against metastatic triple-negative breast cancer cell line MDA-MB-231.

Compound	<i>In vitro</i> IC ₅₀ (μM)
Glycyrrhetic acid	82.29
GA-1	76.5
GA-2	>200
GA-3	91.79
GA-4	116.07
GA-5	>200

Table 6. Compliance of atomic property field score and FlexX binding energy of Glycyrrhetic acid, GA-1, GA-2, GA-3, GA-4 and GA-5 on Glyoxalase-I GSH binding site.

Compound	APF ^a score	FlexX binding energy (kJ/mol)
Glycyrrhetic acid (positive control)	-419.617182	-25.561
GA-1	-425.82269	-16.997
GA-2	-402.04735	-9.7
GA-3	-401.184344	-7.293
GA-4	-415.679134	-21.232
GA-5	-402.136577	-10.419

^aAtomic property field (APF)-based score calculated through ICM-Chemist v3.-6a (Molsoft L.L.C., USA, licensed to CSIR-CIMAP, Lucknow, India).

coordinating residues include GLN32, HIS126 and GLU172. As reported by Zhang et al. (2015), for a non-GSH analogs, GA does not require metal Zn²⁺ coordination. Therefore, in the present work, zinc ion was excluded while docking process as GA is a pentacyclic triterpenoid. From the APF-based alignment and scoring, it is evident that carboxylic group at C-30 position plays a critical role in GA-GLO-I binding. The C-30 carboxylic group hydrogen atom is highly polar and causes negative charge over oxygen after ionisation. This leads to hydrogen bonding between C-30 -RCOO⁻ with polarised hydrogen atoms present in amino acid residues ARG-38 (1.8 Å) and ASN-104 (2.3 Å) (electrophilic centres) of GLO-I binding pocket (Figure 12).

In accordance with the results, the most active GA derivative GA-1 (IC₅₀ = 76.5 μM) showed highest APF score of -425.82 (Table 6). For this purpose, the parent compound GA with APF -219.67 (IC₅₀ = 82.29) was taken as the positive control to identify structure-activity relationship based on their APF score. The least active derivatives GA-2 (IC₅₀ >200 μM) and GA-5 (IC₅₀ >200 μM) showed minimum APF score of -402.047 and -402.14, respectively. Likewise, derivatives GA-3 and GA-4 with moderate IC₅₀ of 91.79 and 116.07 μM accordingly presented a moderate APF score of -401.18 and -415.68, respectively. Consequently, the APF alignment score and IC₅₀ values indicated that the carboxylic group at C-30 in GA-1 and GA-4 in some way plays a major role in enzyme receptor binding.

Correspondingly, the FlexX-based binding energy calculation of GA derivatives on GLO-I was also found in close agreement with APF guided structure-based screening results (Table 6). The results of molecular docking of GA and its derivatives against GLO-I approved that GA-1 exhibited the highest binding affinity of -16.997 kJ/mol in comparison to derivatives GA-2 (-9.7 kJ/mol), GA-3 (-7.293 kJ/mol) and GA-5 (-10.419 kJ/mol). However, GA-4 also showed good

binding affinity (-21.232 kJ/mol). The binding pose analysis of active GA-1 showed that amino acid residues ARG-38 and ARG-123 form hydrogen bonds with C-30 carboxylic group of GA-1. The proposed binding pose of GA-1 on GSH binding site of GLO-I is represented in Figure 13.

Oral bioavailability and toxicity risk assessment results

Rodent (mouse/rat) carcinogenic probability based on data from National Toxicological programme (NTP) showed GA and its derivatives as non-carcinogen. Conversely, the US Food and Drug Administration (FDA)-based data predicted GA and its derivatives as carcinogenic compounds. This contraindication was resolved by considering Weight of Evidence (WOE) prediction that indicates GA and its derivatives may possess carcinogenic character. Many anticancer compounds often possess carcinogenic characters since they target proliferating cells and cause developmental toxicity in developing embryos. However, Ames mutagenic prediction showed derivatives as non-mutagenic. Here, GA-1, GA-2 and GA-3 were also found non-toxic for developmental mutagenicity. Additionally, GA-1 showed moderate to severe skin effect and ocular irritancy. Detailed compliance for computational toxicity analysis is provided in supplementary Table S6. Lastly, the computational results showed that GA and its derivatives have good aerobic biodegradability, and hence they are non-persistent and safe to the environment.

Conclusions

In this study, the QSAR model predicted that IC₅₀ and SRB assay-based biological activities of GA derivatives, GA-2, GA-3, and GA-4 were found comparable against triple-negative breast cancer cell line MDA-MB-231. This indicated that the model extracted structural features *viz.*, Epsilon4 (measure of electronegativity count), ChiV3cluster (valence molecular connectivity index), chi3chain (retention index for three-membered ring), TNN5 (nitrogen atoms separated through 5-bond distances) and nitrogen counts had significant contribution to the biological activity. The results also signify that OH group substitution with acyl group at C-3 position increases the compound lipophilicity, thereby increasing the cytotoxicity potential against TNBC breast cancer cell line MDA-MB-231. Conversely, substitution at C-30 position with propyl amide, butyl amide and amino ethyl amide resulted in decreased cytotoxicity. However, C-30 substitution with butyl amide did not cause any significant difference, whereas the cytotoxicity was decreased due to addition of benzoate group at C-3 position. Overall, the results suggested that C-30 carboxylic group is crucial for GA-based cytotoxic activity. Therefore, an addition of 3-O-acetyl group at C-3 increases GA lipophilicity, thereby improving the cytotoxicity.

Additionally, APF-based scoring and FlexX-based binding affinity with GLO-I, a highly expressing enzyme in metastatic TNBC breast cancers, confirmed that GA and GA-1 exhibited maximum binding affinity. APF-based flexible alignment of GA-1 with co-crystallised GA again established the significance of C-30 carboxyl group as it serves to make three

hydrogen bonds with GLO-I key amino acids ARG-38, ARG-123 and ASN-104. Thus, it is a novel work reporting active natural leads screened virtually using different molecular modelling approaches and *in vitro* validation of predicted results tested on triple-negative breast cancer cell lines. This study will be helpful in early lead discovery against meta-static breast cancers.

Acknowledgments

We are thankful to the Director, CSIR-CIMAP, Lucknow, India for rendering essential research facilities and support. SKS acknowledges the Indian Council of Medical Research (ICMR), New Delhi, for the Emeritus Medical Scientist at CSIR-CIMAP, Lucknow. We acknowledge Jawaharlal Nehru University (JNU), New Delhi, for PhD registration and administrative support. The CIMAP communication number is CIMAP/PUB/2018/40.

Author's contribution

AS, FK and SKS designed the detailed experiments, performed the study, and collected and analyzed the data. RK and SKS synthesised and characterised the derivatives. SM and DD performed *in-vitro* SRB assay. All authors analysed the results and approved the final manuscript.

Disclosure statement

All authors declare no conflict of interest.

Funding

AS acknowledges the Department of Science & Technology (DST), Government of India, New Delhi, for financial assistance through WOS-B fellowship (GAP-339; S. No. DST/KIRAN/SoRF-PM/007/2015/CG) at CSIR-CIMAP.

ORCID

Santosh Kumar Srivastava  <http://orcid.org/0000-0003-2558-6549>

References

- Abagyan, R. (2018). Retrieved from <http://www.molsoft.com/icm-chemist-pro.html>
- Adhikari, N., Amin, S. A., Jha, T., & Gayen, S. (2017). Integrating regression and classification-based QSARs with molecular docking analyses to explore the structure-antiaromatase activity relationships of letrozole-based analogs. *Canadian Journal of Chemistry*, 95(12), 1285–1295. doi:10.1139/cjc-2017-0419. doi:10.1139/cjc-2017-0419
- Amin, S. A., Bhargava, S., Adhikari, N., Gayen, S., & Jha, T. (2018). Exploring pyrazolo [3, 4-d] pyrimidine phosphodiesterase 1 (PDE1) inhibitors: A predictive approach combining comparative validated multiple molecular modelling techniques. *Journal of Biomolecular Structure and Dynamics*, 36(3), 590–608. doi:10.1080/07391102.2017.1288659. doi:10.1080/07391102.2017.1288659
- Badve, S., Dabbs, D. J., Schnitt, S. J., Baehner, F. L., Decker, T., Eusebi, V., ... Palacios, J. (2011). Basal-like and triple-negative breast cancers: A critical review with an emphasis on the implications for pathologists and oncologists. *Modern Pathology*, 24(2), 157. doi:10.1038/modpathol.2010.200
- Bianchini, G., Balko, J. M., Mayer, I. A., Sanders, M. E., & Gianni, L. (2016). Triple-negative breast cancer: Challenges and opportunities of a heterogeneous disease. *Nature Reviews Clinical Oncology*, 13(11), 674–690. doi:10.1038/nrclinonc.2016.66.
- Cai, Y., Xu, Y., Chan, H. F., Fang, X., He, C., & Chen, M. (2016). Glycyrrhetic acid mediated drug delivery carriers for hepatocellular carcinoma therapy. *Molecular Pharmaceutics*, 13(3), 699–709. doi:10.1021/acs.molpharmaceut.5b00677
- Cai, Y., Zhao, B., Liang, Q., Zhang, Y., Cai, J., & Li, G. (2017). The selective effect of glycyrrhizin and glycyrrhetic acid on topoisomerase II α and apoptosis in combination with etoposide on triple negative breast cancer MDA-MB-231 cells. *European Journal of Pharmacology*, 809, 87–97. doi:10.1016/j.ejphar.2017.05.026. doi:10.1016/j.ejphar.2017.05.026
- Cameron, A. D., Ridderström, M., Olin, B., Kavarana, M. J., Creighton, D. J., & Mannervik, B. (1999). Reaction mechanism of glyoxalase I explored by an X-ray crystallographic analysis of the human enzyme in complex with a transition state analogue. *Biochemistry*, 38(41), 13480–13490. doi:10.1021/bi990696c
- Cramer, I. I. I., R. D., Bunce, J. D., Patterson, D. E., & Frank, I. E. (1988). Crossvalidation, bootstrapping, and partial least squares compared with multiple regression in conventional QSAR studies. *Quantitative Structure-Activity Relationships*, 7(1), 18–25. doi:10.1002/qsar.19880070105. doi:10.1002/qsar.19880070105
- Enslein, K. (1988). An overview of structure-activity-relationships as an alternative to testing in animals for carcinogenicity, mutagenicity, dermal and eye irritation, and acute oral toxicity. *Toxicology and Industrial Health*, 4(4), 479–498. doi:10.1177/074823378800400407. doi:10.1177/074823378800400407
- Enslein, K., Borgstedt, H. H., Blake, B. W., & Hart, J. B. (1987). Prediction of rabbit skin irritation severity by structure activity relationships. *In Vitro Toxicology*, 1, 129–147.
- Fonseca-Sánchez, M. A., Rodríguez Cuevas, S., Mendoza-Hernández, G., Bautista-Piña, V., Arechaga Ocampo, E., Hidalgo Miranda, A., ... López-Camarillo, C. (2012). Breast cancer proteomics reveals a positive correlation between glyoxalase 1 expression and high tumor grade. *International Journal of Oncology*, 41(2), 670–680. doi:10.3892/ijo.2012.1478. doi:10.3892/ijo.2012.1478
- Gao, C., Dai, F. J., Cui, H. W., Peng, S. H., He, Y., Wang, X., & Qiu, W. W. (2014). Synthesis of novel heterocyclic ring-fused 18 β -glycyrrhetic acid derivatives with antitumor and antimetastatic activity. *Chemical Biology & Drug Design*, 84(2), 223–233. doi:10.1111/cbdd.12308
- Golbraikh, A., & Tropsha, A. (2002). Beware of q²!. *Journal of Molecular Graphics & Modelling*, 20(4), 269–276. doi:10.1016/S1093-3263(01)00123-1. doi:10.1016/S1093-3263(01)00123-1
- Goldbrunner, M., Loidl, G., Polossek, T., Mannschreck, A., & von Angerer, E. (1997). Inhibition of tubulin polymerization by 5, 6-dihydroindolo [2, 1-a] isoquinoline derivatives. *Journal of Medicinal Chemistry*, 40(22), 3524–3533. doi:10.1021/jm970177c
- Grigoryan, A. V., Kufareva, I., Totrov, M., & Abagyan, R. A. (2010). Spatial chemical distance based on atomic property fields. *Journal of Computer-Aided Molecular Design*, 24(3), 173–182. doi:10.1007/s10822-009-9316-x
- Guo, Y., Zhang, Y., Yang, X., Lu, P., Yan, X., Xiao, F., ... Meng, Q. H. (2016). Effects of methylglyoxal and glyoxalase I inhibition on breast cancer cells proliferation, invasion, and apoptosis through modulation of MAPKs, MMP9, and Bcl-2. *Cancer Biology & Therapy*, 17(2), 169–180. doi:10.1080/15384047.2015.1121346
- He, S., Dong, G., Wang, Z., Chen, W., Huang, Y., Li, Z., ... Sheng, C. (2015). Discovery of novel multiacting topoisomerase I/II and histone deacetylase inhibitors. *ACS Medicinal Chemistry Letters*, 6(3), 239–243. doi:10.1021/ml500327q
- Iqbal, J., Abbasi, B. A., Batool, R., Mahmood, T., Ali, B., Khalil, A. T., ... Ahmad, R. (2018). Potential phytochemicals for developing breast cancer therapeutics: Nature's healing touch. *European Journal of Pharmacology*, 827, 125–148. doi:10.1016/j.ejphar.2018.03.007
- Jamdade, V. S., Sethi, N., Mundhe, N. A., Kumar, P., Lahkar, M., & Sinha, N. (2015). Therapeutic targets of triple-negative breast cancer: A review. *British Journal of Pharmacology*, 172(17), 4228–4237. doi:10.1111/bph.13211
- Jaworska, J., Nikolova-Jeliazkova, N., & Aldenberg, T. (2005). QSAR applicability domain estimation by projection of the training set descriptor space: A review. *ATLA-Nottingham*, 33(5), 445.

- Kier, L. B., & Hall, L. H. (1977). Nature of structure–activity-relationships and their relation to molecular connectivity. *European Journal of Medicinal Chemistry*, 12 (4), 307–312.
- Krähnenbühl, S., Hasler, F., & Krapf, R. (1994). Analysis and pharmacokinetics of glycyrrhizic acid and glycyrrhetic acid in humans and experimental animals. *Steroids*, 59(2), 121–126. doi:10.1016/0039-128X(94)90088-4. doi:10.1016/0039-128X(94)90088-4
- Kramer, B., Rarey, M., & Lengauer, T. (1999). Evaluation of the FLEXX incremental construction algorithm for protein–ligand docking. *Proteins: Structure, Function, and Genetics*, 37(2), 228–241. doi:10.1002/(SICI)1097-0134(19991101)37:2<228::AID-PROT8>3.0.CO;2-8
- Li, Y., Feng, L., Song, Z. F., Li, H. B., & Huai, Q. Y. (2016). Synthesis and anticancer activities of glycyrrhetic acid derivatives. *Molecules*, 21(2), 199. doi:10.3390/molecules21020199
- Motiwala, H. F., Bazzill, J., Samadi, A., Zhang, H., Timmermann, B. N., Cohen, M. S., & Aubé, J. (2013). Synthesis and cytotoxicity of semisynthetic withalongoide A analogues. *ACS Medicinal Chemistry Letters*, 4(11), 1069–1073. doi:10.1021/ml400267q
- Negishi, M., Irie, A., Nagata, N., & Ichikawa, A. (1991). Specific binding of glycyrrhetic acid to the rat liver membrane. *Biochimica et Biophysica Acta (BBA)-Biomembranes*, 1066(1), 77–82. doi:10.1016/0005-2736(91)90253-5. doi:10.1016/0005-2736(91)90253-5
- Ojha, P. K., Mitra, I., Das, R. N., & Roy, K. (2011). Further exploring rm2 metrics for validation of QSPR models. *Chemometrics and Intelligent Laboratory Systems*, 107(1), 194–205. doi:10.1016/j.chemolab.2011.03.011. doi:10.1016/j.chemolab.2011.03.011
- Ovaricek, T., Frkovic, S., Matos, E., Mozina, B., Borstnar, S. T. N. B. (2011). Triple negative breast cancer – prognostic factors and survival. *Radiation and Oncology* 45(1), 46–52. doi:10.2478/v10019-010-0054-4
- Polyak, K. (2011). Heterogeneity in breast cancer. *Journal of Clinical Investigation*, 121(10), 3786–3788. doi:10.1172/JCI60534
- Roy, K., Ambure, P., Kar, S., & Ojha, P. K. (2018). Is it possible to improve the quality of predictions from an “intelligent” use of multiple QSAR/QSPR/QSTR models? *Journal of Chemometrics*, 32(4), e2992. doi:10.1002/cem.2992. doi:10.1002/cem.2992
- Shen, M., LeTiran, A., Xiao, Y., Golbraikh, A., Kohn, H., & Tropsha, A. (2002). Quantitative structure – activity relationship analysis of functionalized amino acid anticonvulsant agents using k nearest neighbor and simulated annealing PLS methods. *Journal of Medicinal Chemistry*, 45(13), 2811–2823. doi:10.1021/jm010488u
- Silva, M. S., Gomes, R. A., Ferreira, A. E., Freire, A. P., & Cordeiro, C. (2013). The glyoxalase pathway: The first hundred years and beyond. *Biochemical Journal*, 453(1), 1–15. doi:10.1042/BJ20121743
- Sullivan, L. B., Gui, D. Y., & Vander Heiden, M. G. (2016). Altered metabolite levels in cancer: Implications for tumour biology and cancer therapy. *Nature Reviews Cancer*, 16(11), 680. doi:10.1038/nrc.2016.85. doi:10.1038/nrc.2016.85
- Tewari, D., Mocan, A., Parvanov, E. D., Sah, A. N., Nabavi, S. M., Huminiecki, L., ... Atanasov, A. G. (2017). Ethnopharmacological approaches for therapy of jaundice: Part II. Highly used plant species from acanthaceae, Euphorbiaceae, Asteraceae, Combretaceae, and Fabaceae Families. *Frontiers in Pharmacology*, 8, 519. doi:10.3389/fphar.2017.00519
- Thike, A. A., Cheok, P. Y., Jara-Lazaro, A. R., Tan, B., Tan, P., & Tan, P. H. (2010). Triple-negative breast cancer: Clinicopathological characteristics and relationship with basal-like breast cancer. *Modern Pathology*, 23(1), 123. doi:10.1038/modpathol.2009
- Totrov, M. (2008). Atomic property fields: Generalized 3D pharmacophoric potential for automated ligand superposition, pharmacophore elucidation and 3D QSAR. *Chemical Biology & Drug Design*, 71(1), 15–27. doi:10.1111/j.1747-0285.2007.00605.x
- Totrov, M. (2011). Ligand binding site superposition and comparison based on Atomic Property Fields: identification of distant homologues, convergent evolution and PDB-wide clustering of binding sites. *BMC Bioinformatics*, 12(Suppl 1), S35. doi:10.1186/1471-2105-12-S1-S35
- Wein, L., Luen, S. J., Savas, P., Salgado, R., & Loi, S. (2018). Checkpoint blockade in the treatment of breast cancer: current status and future directions. *British Journal of Cancer*, 119(1), 4–11. doi:10.1038/s41416-018-0126-6
- Xu, B., Wu, G. R., Zhang, X. Y., Yan, M. M., Zhao, R., Xue, N. N., ... Wang, P. L. (2017). An overview of structurally modified glycyrrhetic acid derivatives as antitumor agents. *Molecules*, 22(6), 924. doi:10.3390/molecules22060924.
- Yadav, D., Kalani, K., Khan, F., & Srivastava, S. (2013). QSAR and docking based semi-synthesis and *in vitro* evaluation of 18 β -glycyrrhetic acid derivatives against human lung cancer cell line A-549. *Medicinal Chemistry*, 9(8), 1073–1084. doi:10.2174/1573406411309080009
- Yadav, D. K., Kalani, K., Singh, A. K., Khan, F., Srivastava, S. K., & Pant, A. B. (2014). Design, synthesis and *in vitro* evaluation of 18 β -glycyrrhetic acid derivatives for anticancer activity against human breast cancer cell line MCF-7. *Current Medicinal Chemistry*, 21(9), 1160–1170. doi:10.2174/1573406411309080009
- Yadav, D. K., & Khan, F. (2013). QSAR, docking and ADMET studies of camptothecin derivatives as inhibitors of DNA topoisomerase-I. *Journal of Chemometrics*, 27(1-2), 21–33. doi:10.1002/cem.2488. doi:10.1002/cem.2488
- Zardavas, D., Baselga, J., & Piccart, M. (2013). Emerging targeted agents in metastatic breast cancer. *Nature Reviews Clinical Oncology*, 2013, 10(4), 191. doi:10.1038/nrclinonc.2013.29
- Zhang, H., Huang, Q., Zhai, J., Zhao, Y.-N., Zhang, L.-P., Chen, Y.-Y., ... Hu, X.-P. (2015). Structural basis for 18- β -glycyrrhetic acid as a novel non-GSH analog glyoxalase I inhibitor. *Acta Pharmacologica Sinica*, 36(9), 1145. doi:10.1038/aps.2015.59
- Zheng, W., & Tropsha, A. (2000). Novel variable selection quantitative structure – property relationship approach based on the k-nearest-neighbor principle. *Journal of Chemical Information and Computer Sciences*, 40(1), 185–194. doi:10.1021/ci980033m