



ISSN: 0976-3031

Available Online at <http://www.recentscientific.com>

CODEN: IJRSFP (USA)

*International Journal of Recent Scientific Research*  
Vol. 10, Issue, 01(E), pp. 30529-30539, January, 2019

**International Journal of  
Recent Scientific  
Research**

DOI: 10.24327/IJRSR

## Research Article

### PPI OF HSP90 AND 3D-QSAR EVALUATION STUDY OF PURINE DERIVATIVES AS ANTI-CARCINOGENIC SUBSTRATES OF HSP90-OVEREXPRESSED MALIGNANT FEMALE BREAST CANCER

Abiodun Julius Arannilewa<sup>1,2\*</sup>, Oluwaseun Suleiman Alakanse<sup>1</sup>, Ifedayo Michael Obaidu<sup>2,3</sup>, Oluwagbenga Ajibade Oyeyemi<sup>2</sup>, Ayomikun Emmanuel Olugbode<sup>1</sup>, Oluwaseyi Israel Malachi<sup>4</sup>, Ebunolorun Ibukunoluwa Ayo<sup>2</sup>, Funmilola Favour Anjorin<sup>6</sup>, Oluwaferanmi Victor Adefolaju<sup>5</sup>, Samuel Adebowale Oyerinde<sup>2</sup>, Samuel Olufemi Araoyinbo<sup>7</sup>, Adepeju Oluwatosin Ogunsola<sup>1</sup>, Elijah Olorunleke Ige<sup>1</sup> and George Oche Ambrose<sup>1</sup>

<sup>1</sup>Department of Biochemistry, University of Ilorin, Ilorin, Kwara State, Nigeria

<sup>2</sup>Department of Biochemistry, Ekiti State University, Ado Ekiti, Ekiti State, Nigeria

<sup>3</sup>Department of Biochemistry, University of Ibadan, Ibadan, Oyo State, Nigeria

<sup>4</sup>Food Production Department, Goodfoods Inc., Montreal, QC, Canada

<sup>5</sup>Department of Biochemistry, Federal University of Technology, Minna, Niger State, Nigeria

<sup>6</sup>Department of Chemical Pathology, University of Ibadan, Ibadan, Oyo State, Nigeria

<sup>7</sup>Department of Chemistry, University of Ibadan, Ibadan, Oyo State, Nigeria

DOI: <http://dx.doi.org/10.24327/ijrsr.2019.1001.3077>

#### ARTICLE INFO

##### Article History:

Received 6<sup>th</sup> October, 2018

Received in revised form 15<sup>th</sup> November, 2018

Accepted 12<sup>th</sup> December, 2018

Published online 28<sup>th</sup> January, 2019

##### Key Words:

HSP90, Breast Cancer, 3D-QSAR, Purine Derivatives, Protein-protein Interaction (PPI)

#### ABSTRACT

Breast cancer accounts for 9 % of global cases of cancer and represents the second most common of all types of cancer. It is only found in women and represents about 22.9 % of global cases of cancer in women. Pharmacologically inhibiting HSP 90 and a resultant inhibition of its client protein have been shown to be a great therapeutic target in the treatment of breast carcinoma because these inhibitors have the potential of suppressing multiple oncogenic signaling pathways.

Sixty (60) compounds from literature work; 42 train and 18 test dataset compounds were selected for this study respectively. Best selected 3D physiochemical descriptors with significant positive contributions towards the development of the model includes; RDFM5, MoRSEN28, MoRSEC20, RDFM9, MoRSEV11, RDFC16, RDFE15, ASPAN, E3u, RDFC25, RDFU8, and MoRSEV13 which all contribute significantly to the bioactivity. MLR statistical regression analysis expressed potential predictive power compared to PLS method. The 3D QSAR model developed by the trained dataset with  $R^2 = 0.97463$ ,  $Q^2 = 0.900$ ,  $R_m^2 = 0.8654$ , and  $r^2 = 0.6415$  for external validation depicts the predictive potential of the 3D QSAR model. The 3D-QSAR model may lead to a better understanding of structural requirements of 8-arylsulfanyl, 8-arylsulfoxyl, and 8-arylsulfonyl adenines derivatives which could also aid in designing novel anti-carcinogenic molecules.

Copyright © Abiodun Julius Arannilewa *et al*, 2019, this is an open-access article distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution and reproduction in any medium, provided the original work is properly cited.

#### INTRODUCTION

Molecular chaperones play vital roles in proteostasis and cellular homeostasis, most importantly during heat shock. They are a class of protein that helps other functionally inactive protein (otherwise referred to as clients or substrates) achieve functionality by interacting with them. The wholistic effect of these interaction-activation-dissociation processes is a response

to stress that brings about homeostasis<sup>1</sup>. HSP90, a molecular chaperone, is particularly overexpressed during heat shock response and has been shown to be essential in other cellular processes including steroid hormone receptor signaling in association with its cofactors<sup>2</sup>.

Breast cancer accounts for 9 % of global cases of cancer and represents the second most common of all types of cancer<sup>3</sup>. It is

\*Corresponding author: Abiodun Julius Arannilewa

Department of Biochemistry, University of Ilorin, Ilorin, Kwara State, Nigeria

only found in women and represents about 22.9 % of global cases of cancer in women. In fact, statistical data has shown that of all women given birth to between 2007 and 2009, it is likely that about 12.38 % of them show symptoms of breast cancer at a point in their lifetime<sup>3</sup>. A number of different malignancies are involved in breast cancer. Each of which is having its diverse behavior and distinct genetic profile with varying response to therapy<sup>4</sup>. Current treatment depends on molecular tumor subtypes which have been defined clinically by the expression of estrogen receptor (ER), progesterone receptor (PR), and human epidermal growth factor receptor type 2 (HER2)<sup>5</sup>. These molecules form part of the HSP90 chaperone complex which has been shown to represent a promising target for pharmacological modulation in the treatment of breast cancer<sup>5</sup>. In about 20-25 % of ER $\alpha$  positive breast cancer, HER2 gene is overexpressed<sup>6</sup> and several evidences have been given to support its role in breast cancer pathogenesis<sup>7</sup>

HSP 90 is a dimeric protein with an N-terminal ATP-binding domain (N), a C-terminal dimerization domain (C) and a middle domain (M) that is essential for client binding as its three functional units. The N- and C-terminals have been shown to be drugable and crucial in providing sites of interaction for proteins, co-chaperone and in assisting the function of chaperone<sup>8,9,10</sup>. HSP90 action is a mechanism that involves the ATP-binding of the N-terminal domain, hydrolysis of the ATP, causing a change in the conformation of chaperone and a subsequent interaction with co-chaperone and client proteins<sup>5</sup>. Pharmacologically inhibiting HSP 90 and a resultant inhibition of its client protein have been shown to be a great therapeutic target in the treatment of breast carcinoma because these inhibitors have the potential of suppressing multiple oncogenic signaling pathways at the same time thereby decreasing the chance of molecular feedback loops and mutations causing resistance to tumor<sup>11</sup>.

Most HSP 90 inhibitors that act on the N-terminal domain via targeting ATP have made their way to clinical trials. ATP binding by these inhibitors prevents hydrolysis and conformational changes of chaperone, thus stopping the orchestrated chaperone cycle. This obstruction causes ubiquitination, proteosomal degradation of client proteins and induction of HSP 70 co-chaperone<sup>12</sup>. This degradation of clients such as the human epidermal growth factor receptor type 2 (HER2), epidermal growth factor receptor (EGFR), antiapoptotic kinase (Akt), MET receptor tyrosine kinase, vascular epithelial growth factor receptor (VEGFR), androgen receptor (AR), estrogen receptor (ER), and mutant p53 may block several key carcinoma promoting cellular pathways including the phosphatidylinositol-3-kinase (PI3-kinase)-Akt and RAS-RAF-MEK-ERK pathways<sup>5</sup>

Of particular interest among the client proteins is HER2 which has shown great sensitivity along with perturbation of other interesting components of the HER2 pathway including Akt<sup>13, 14, 15, 16</sup>. In HER2-positive breast cancer, antitumor activity of HSP90 inhibitors has been confirmed by preclinical data making way for clinical development<sup>17, 18</sup>. These inhibitors have been used independently in several studies or as a combination therapy with anti-HER2 antibody including trastuzumab, lepatinib<sup>5</sup> and pertuzumab<sup>19</sup>

Several HSP90 inhibitors have been identified including geldanamycin (GM) and its derivative tanespimycin (17-AAG), carbonmonooxide via heme oxygenase-1, alvespimycin (17-DMAG), retaspimycin (IPI-504), novobicin and DCZ3112 which targets HSP90-CDC37 interaction as against C-terminal ATP target by traditional inhibitors<sup>19</sup>

While most of these inhibitors have shown great promises in the treatment of breast cancer, majority of them including GM has been limited by their low selectivity, high toxicity and the induction of heat shock response<sup>19</sup>. This has necessitated the need to look for a novel inhibitor that would overcome these limitations without itself having other limitations which would be more of a disadvantage.

The critical role of HSP90 in malignant breast cancer can only be ascertained when developed potent and potential inhibitors are evaluated by clinicians. On this ground, QSAR studies of 8-arylsulfanyl, 8-arylsulfoxyl, and 8-arylsulfonyl adenines derivatives as antagonists of HSP90 were carried out.

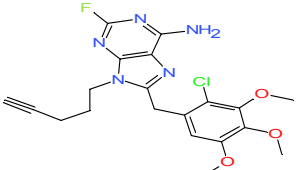
The present study aims at rationalizing inhibitory derivatives of 8-arylsulfanyl, 8-arylsulfoxyl, and 8-arylsulfonyl adenine scaffold to provide insight for future studies. Structural Activity Relationship (SAR) in QSAR is elucidated via measurement of physicochemical (descriptors) properties which are correlation of biological activity; this leads to the generation of mathematical models. In this study, an attempt has been made to develop QSAR models adopting the multiple linear regression (MLR) and partial least squares (PLS) methodologies. The concept of the training and test sets has been introduced for the prediction of HSP90 inhibitory activity of structurally varied sets of compounds.

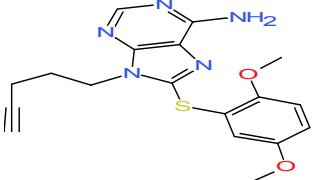
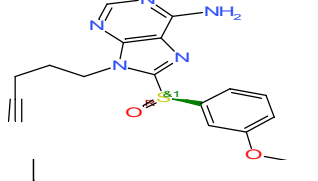
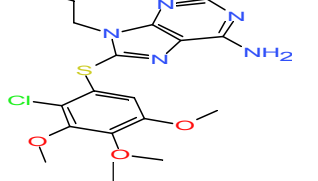
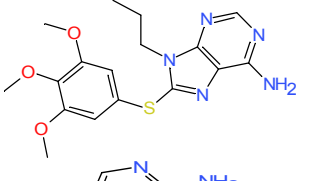
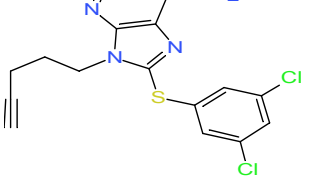
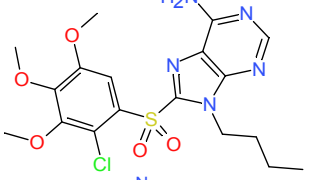
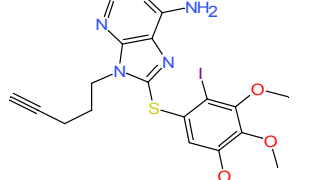
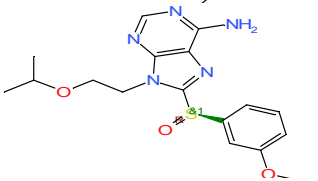
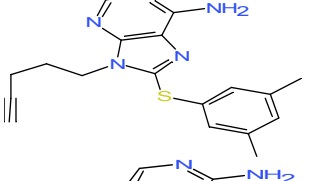
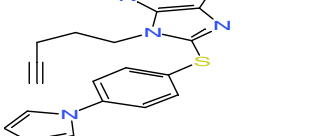
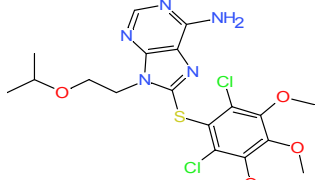
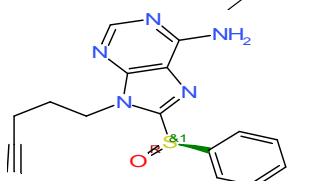
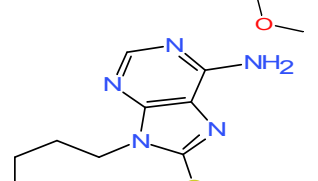
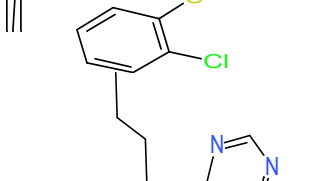
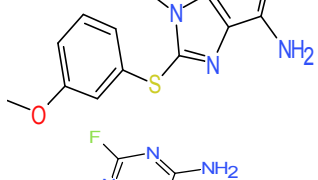
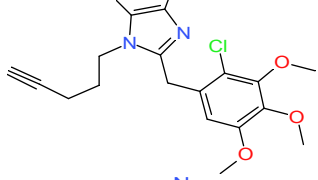
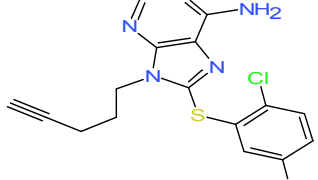
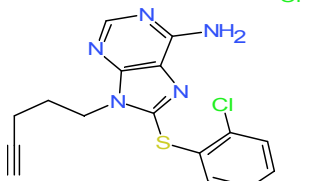
## MATERIALS AND METHODS

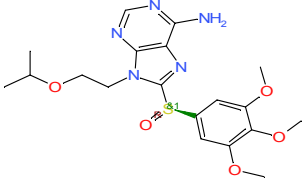
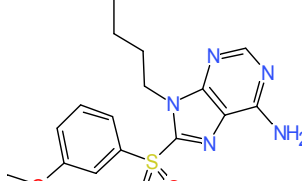
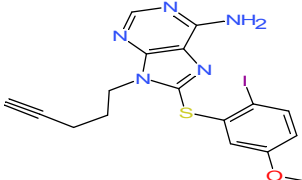
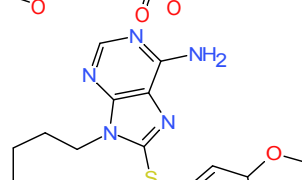
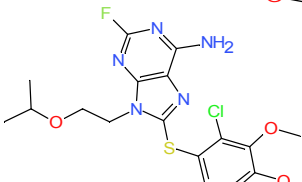
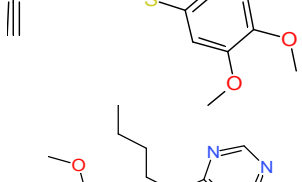
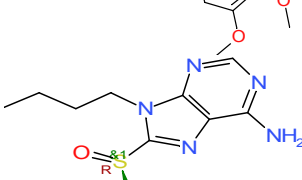
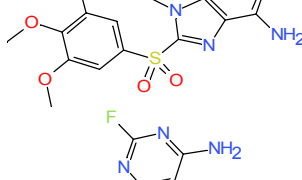
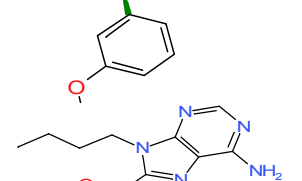
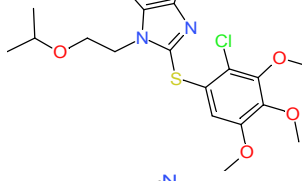
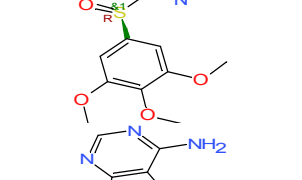
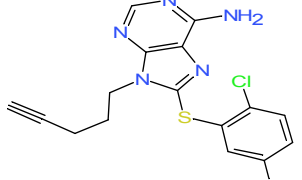
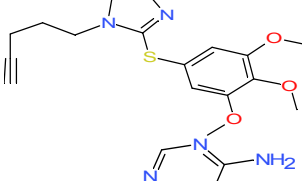
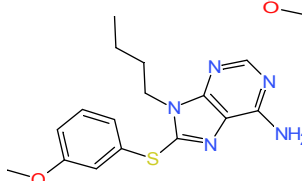
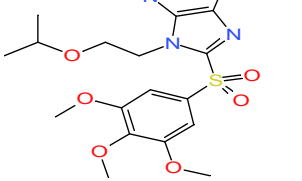
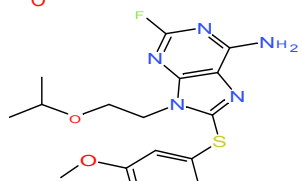
### Experimental data

In this present molecular modelling study, a set of 60 compounds 8-arylsulfanyl, 8-arylsulfoxyl, and 8-arylsulfonyl adenines were retrieved from the ChEMBL database ([www.ebl.ac.uk/chembl](http://www.ebl.ac.uk/chembl)) with the accession ID of ChEMBL1143602. This dataset depicts an *in vitro* antagonistic activity of the derivate in term of IC<sub>50</sub> ( $\mu$ M) against HSP90. The IC<sub>50</sub> values were converted into pIC<sub>50</sub> = (-Log (IC<sub>50</sub>X)). The molecular structure of 8-arylsulfanyl, 8-arylsulfoxyl, and 8-arylsulfonyl adenines are indicated in the (Table 1) below with observed activities.

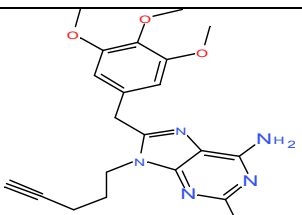
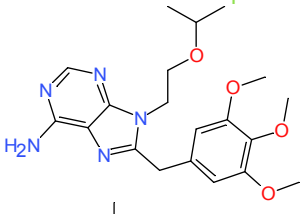
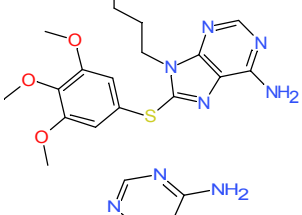
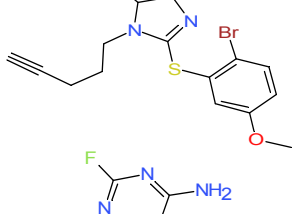
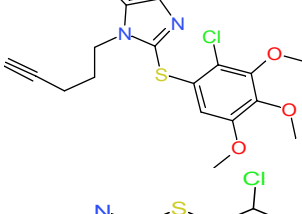
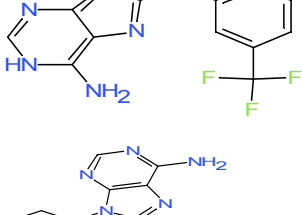
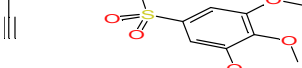
**Table 1** Derivatives of 8-arylsulfanyl, 8-arylsulfoxyl, and 8-arylsulfonyl adenines and pIC<sub>50</sub> values

S/N	Compound ID	Structure	pIC <sub>50</sub>
1	ChEMBL112953		3.98

2	CHEMBL273986		5.82
3	CHEMBL189537		5.49
4	CHEMBL191268		4.50
5	CHEMBL191641		4.64
6	CHEMBL192013		4.05
7	CHEMBL192937		3.84
8	CHEMBL193056		4.44
9	CHEMBL193672		4
10	CHEMBL195096		4.15
11	CHEMBL364730		4.46
12	CHEMBL370247		3.92
13	CHEMBL189537		5.0
14	CHEMBL190439		4.60
15	CHEMBL195869		3.71
16	CHEMBL112953		4.42
17	CHEMBL364450		5.74
18	CHEMBL364456		4.58
19	CHEMBL365118		4.67

20	CHEMBL370092		3.45	28	CHEMBL372013		4.20
21	CHEMBL191635		4.18	29	CHEMBL190518		4.24
22	CHEMBL191786		6.24	30	CHEMBL190654		4.57
23	CHEMBL192649		5.48	31	CHEMBL191786		4.07
24	CHEMBL192761		4.24	32	CHEMBL192062		5.66
25	CHEMBL190518		4.36	33	CHEMBL195869		5.46
26	CHEMBL193313		4.65	34	CHEMBL439852		4.27
27	CHEMBL193672		4.16	35	CHEMBL427169		4.39

36	CHEMBL364272		3.80	45	CHEMBL191241		4.38
37	CHEMBL370092		3.91	46	CHEMBL192013		4.55
38	CHEMBL370247		4.36	47	CHEMBL192062		4.22
39	CHEMBL191635		4.86	48	CHEMBL193259		4.54
40	CHEMBL191668		6.09	49	CHEMBL424824		4.68
41	CHEMBL191920		4.18	50	CHEMBL362972		4.50
42	CHEMBL194716		4.58	51	CHEMBL364239		3.77
43	CHEMBL363196		4.32	52	CHEMBL364239		3.83
44	CHEMBL370286		4.23	53	CHEMBL370286		3.33

54	CHEMBL349231		4.12
55	CHEMBL352683		4.77
56	CHEMBL191641		4.54
57	CHEMBL192145		4.18
58	CHEMBL193282		5.09
59	CHEMBL415775		5.74
60	CHEMBL193310 (1)		3.89

### Accession of chemical structures

The canonical smiles retrieved from accession ID of CHEMBL1143602 were converted into SDF using data Warrior software version 4.7.2. The generated 3D-QSAR model was derived from the training dataset of 42 molecules while the predictive potential of the model was validated by the test set of 18 molecules with uniformly distributed biological activities. (Table 2) shows the observed and predicted biological activities of the training and test datasets.

### Geometry optimization and Descriptors generation

The geometries of the derived SDF from data Warrior software version 4.7.2 were optimized using the online web server

(<http://www.scbdd.com/mopac-optimization>) so make conformation have the least potential energy. MOPAC software packages were used. To develop a QSAR model, activity compounds are to be quantitatively represented by molecular descriptors. The ChemDec web server ([http://www.scbdd.com/chemopy\\_desc](http://www.scbdd.com/chemopy_desc)) was used to generate 3D descriptors under the following categories: Geometric descriptors, CPSA descriptors, RDF descriptors, WHIM descriptors, and MoRSE descriptors. The pretreatment of the independent variables (i.e descriptors) was carried out using DataPreTreatmentGUI 1.2 software. This works on the principles of eliminating invariable (constant columns) and other descriptors based correlation coefficient cut-off of 0.99 and variance cut-off of 0.0001.

### Data normalization

Variability in distribution and range of each variable, calculated value of the descriptors and corresponding biological activity justifies subjection to statistical min-max normalization techniques using Normalize The Data software version 1.0. The software works based on the principles of adjusting the minimum and maximum value of each variable to a uniform range between 0 and 1 according to the equation below.

$$X_{\text{normalized}} = \frac{xi - X_{\text{min}}}{X_{\text{max}} - X_{\text{min}}}$$

Where represents the min-max normalized value, xi represents the value of interest,  $X_{\text{min}}$  represents the minimum value, and  $X_{\text{max}}$  represents the maximum value.

### Selection of training and test set

The dataset of 60 molecules was divided into 18 and 42 test and train dataset respectively using Dataset Division GUI v1.2 based on Kennard-Stone method (Paul and Mukhopadhyay, 2004). The method was used to for both MLR and PLS model with pIC50 values as dependent variable, while the calculated 3D descriptor represent the independent variables.

### Model Validation

Outstanding achievements recorded in the field of drug discovery and design has been showed to enormously rest on the fulcrum of physicochemical and structural features in cohort with bioactivity. Constrained conclusions and justification in the use of a developed QSAR model rest on its ability to predict unknown chemicals with some degree of certainty. Want of validation of a model as lead to the development of models that falsely predict the biological activity which good more harm than good, hence validation of a QSAR model is most critical part in QSAR model studies and development<sup>20, 21</sup>

### Internal validation

Methodologies used in internal validation in this study include; cross validation ( $Q^2$ ), least squares fit ( $R^2$ ), and adjusted  $R^2$  ( $R^2_{\text{adj}}$ ). Least square fit ( $R^2$ ), is the most common methodology used in internal validation of a QSAR model. This method is similar to linear regression in that it is used in the determination of correlation between the experiment and predicted activities. The discrepancy between  $R^2$  and  $R^2_{\text{adj}}$  is

(<3), this validate the fact that the number of descriptors used in the development of a QSAR model is acceptable.

$$R^2 = \left[ \frac{N \sum XY - (\sum X)(\sum Y)}{\sqrt{[(N \sum X^2 - (\sum X)^2)(N \sum Y^2 - (\sum Y)^2)]}} \right]^2$$

Another method used in internal validation of a QSAR model is the cross-validation (CV,  $Q^2$ ,  $q^2$ , or Jack-knifing) methodology. This repeats the regression process many times on subsets of data. During this process, each molecule is left out once, the predicted values of the missing molecule is used to compute the R value. The overall  $R^2$  for giving QSAR model is usually larger in comparison to cross validation  $R^2$ . This used in the evaluation of the predictive potency of an equation. In addition to this, over-fitting is brought into consideration; this refers to the phenomenon in which the QSAR predictive model may well delineate relationship between predictors and response but fail subsequently in the provision of valid predictions for new compounds. This becomes fishy when  $R^2$  is 25% significantly larger than  $Q^2$  or when the difference is (>0.3). A good predictability potential of a QSAR model is obtained when  $R^2 - Q^2 \leq 0.3$ <sup>22, 23</sup>

$$Q^2 = 1 - \frac{\text{PRESS}}{\sum_{i=1}^N (Y_i - Y_m)^2}$$

$$\text{PRESS} = \sum_{i=1}^N (Y_{\text{pred}, i} - Y_i)^2$$

The data value not used to construct the CV model is represented by  $Y_i$

#### External Validation

Coefficients of determination ( $R^2$ ) (predicted against observed  $r_0^2$ , and observed against predicted activities  $r_0'$ ), correlation coefficient  $R$  between predicted and observed activities, and slope  $k$  and  $k'$  regression lines via the origin are statistical techniques used in the determination of the predictive power of a QSAR model.

$$R^2_{\text{pred}} = 1 - \frac{\sum_{i=1}^{\text{test}} (Y_{\text{exp}} - Y_{\text{pred}})^2}{\sum_{i=1}^{\text{test}} (Y_{\text{exp}} - Y_u)^2}$$

Where ( $Y_u$ ) is the average value of dependent viable for the train set.

In 2008 Roy and Paul proposed the use of  $r_m^2$  for determination of external predictive potential of a QSAR model.

$$R_m^2 = r_2 (\sqrt{r^2 - r'^2})$$

Where  $r^2$  is the correlation coefficient between observed and predicted values and  $r'^2$  is correlation coefficient between observed and predicted set at the intercept of zero. A model is considered acceptable and good when the overall  $r_m^2 > 0.5$ <sup>21, 24</sup>

#### 3D-QSAR model development

In this study, 3D-QSAR models were developed using PLS and MLR methods to screen potential leads against HSP90 within the training set (42 compounds). The total number of 3D descriptors (497) was calculated for each of the 60 compounds using Chemopy algorithm imbedded within ChemDec web server ([http://www.scbdd.com/chemopy\\_desc](http://www.scbdd.com/chemopy_desc)). Subsequently, a robust QSAR model equation was derived by MLR; descriptors without relevance were rid of via forward stepwise method

leading to selection of (20) 3D descriptors in the final QSAR regression equation (Figure 5a-6b). The regression equation portrays a relationship in the form of a linear equation that best depicts individual data points.

$$Y = C + b_1x_1 + b_2x_2 + b_3x_3 + \dots$$

Where Y is the dependent variable, "b's" are regression coefficients for corresponding 'x's' (independent variable), 'c' is a regression constant.

Latent variables which are combinations of the original variables are easily find by PLS model. In circumstances where dataset are imbedded with high inter-correlated descriptors and descriptors exceeding number of observations, PLS have been proved to be useful. The optimum and maximum latent variables were determined based on leave one out cross validation methodology. Same descriptors (497) were calculated using ChemDec web server ([http://www.scbdd.com/chemopy\\_desc](http://www.scbdd.com/chemopy_desc)) and subsequently selected as in MLR analysis. Irrelevant descriptors without relevance were rid of via forward stepwise forward algorithm using S-MLR v1.2 software at a variance cut-off (0.001) and inter-correlation cut-off of 0.99.

#### Protein-Protein-Interaction of HSP90

The FASTA amino acid sequence of HSP90 (UniProtKB-P07900 (HSP90A\_HUMAN)) protein was retrieved from UniProt Database (<https://www.uniprot.org/>). The FASTA sequence was then subsequently uploaded on STRING database ([string-db.org](http://string-db.org)). Interaction search was carried out at the highest confidence of 0.90 interactions, maximum of 50 nodes interactions at the first shell and the second shell respectively.

## RESULTS AND DISCUSSION

### 3D-QSAR study

Stepwise MLR methodology was used to remove descriptors of high correlation based on inter-correlation coefficients of the descriptors at a correlation regression cut-off of 0.99. The below equation (Table 2) depicts that the model obtained via Multiple Linear Regression Model (MLR) showed a good  $R^2$ ,  $R^2_{\text{adjusted}}$ , and cross-validation (CV,  $Q^2$ ,  $q^2$ , or Jack-knifing) values within the required thresholds which are parameters needed for internal validation (Table 3) of the developed 3D-QSAR model. The discrepancy between  $R^2$  and  $R^2_{\text{adj}}$  is (<0.3), has been used to validate the fact that the number of descriptors used in the development of a QSAR model is acceptable. And based on this studies and required threshold (<3), the difference between  $R^2$  and  $R^2_{\text{adjusted}}$  was acceptable (0.0242). In addition to  $Q^2$  use in ascertaining predictive potential, it also plays critical roles in determination of over-fitness of a QSAR model equation. Based on the values obtained for  $R^2$  and  $Q^2$ , 0.0746 was the generated difference ( $R^2 - Q^2$ ) which is still within the required threshold for the determination of over-fitness ( $R^2 - Q^2 \leq 0.3$ ) of a QSAR model. The scatter plot which was plotted for observed and predicted pIC50 value for train and test set was indicated in the (Figure 1a and 1b) respectively.

**Table 2** 3D-QSAR model using Multiple Linear Regression Model (MLR)

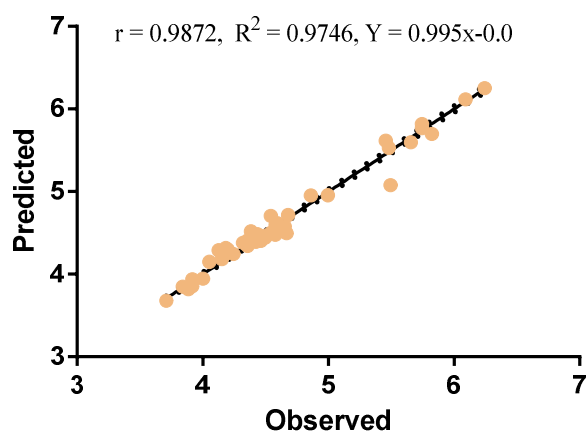
MLR model
$\text{PIC50} = -3.45494(+/-0.86422) + 0.05688(+/-0.00835)\text{RDFM5} + 1.37228(+/-0.65614)\text{MoRSEN28} + 4.04899 (+/-0.81756)\text{MoRSEC20} + 0.08909 (+/-$



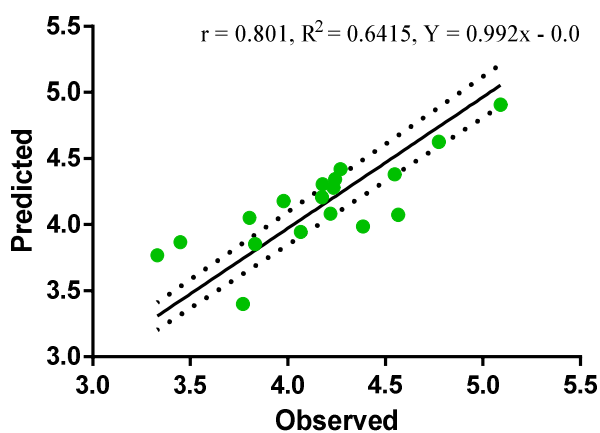
0.0089) RDFM9 + 0.89385(+/-0.23368) MoRSEV11 + 4.06859(+/-0.48409) RDFC16 - 0.28706(+/-0.13583) MoRSEM12 -0.49782(+/-0.1141) MoRSEU14 + 0.05637(+/-0.00743) RDFE15 + 13.62806(+/-1.69935) ASPAN + 3.76225(+/-0.56353) E3u + 1.5645(+/-0.26527) MoRSEM30 - 0.19928(+/-0.0937) MoRSEM9 + 8.12644(+/-2.36935) RDFC25 +0.03013(+/-0.00839) RDFU8 - 0.14481(+/-0.02683) RDFP5 - 2.6622(+/-0.53341) P2m + 0.56675(+/-0.2028) MoRSEV13 - 0.75569(+/-0.33411) Petitj3D - 0.32532(+/-0.50184) RDFC20

**Table 3** Multiple Linear Regression Model (MLR) internal and external validation parameters

Internal Validation Parameters	Threshold value
$R^2 = 0.97463$ , $R^2_{\text{adjusted}} = 0.95047$ , $\text{PRESS} = 0.44449$ , $F = 40.33859$	$Q^2 = 0.90004$ Passed (Threshold value $Q^2 > 0.5$ )
	$r^2 = 0.64154$ Passed (Threshold value $r^2 > 0.6$ )
<b>Leave-One-Out (LOO) Result</b>	$ r_0^2 - r^2  = 0.22638$ Passed (Threshold value $ r_0^2 - r^2  < 0.3$ )
$Q^2 = 0.900$ , Average $R_m^2 = 0.8654$	$k_1 = 0.00342$ [ $(r^2 - r_0^2)/r^2$ ] = 0.00059
<b>External Validation Parameters</b>	OR*
$r^2 = 0.6415$ , $r_0^2 = 0.6412$ , reverse $r_0^2 = 0.41478$ , $R^2_{\text{Pred}} = 0.8436$	$k' = 0.99289$ [ $(r^2 - r_0^2)/r^2$ ] = 0.35346 Passed (Threshold value: $[0.85 < k < 1.15$ and $((r^2 - r_0^2)/r^2) < 0.1$ ]



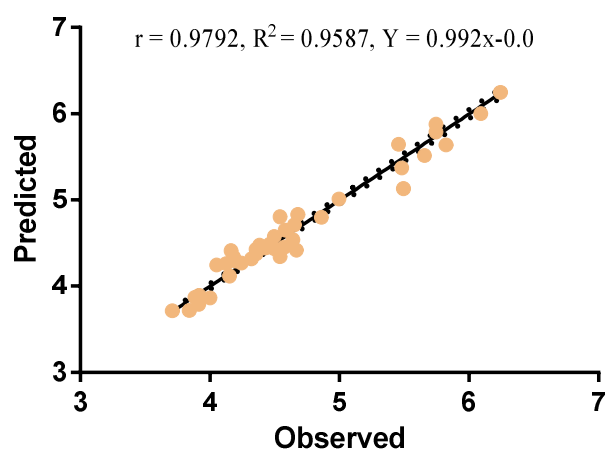
**Figure 1** MLR analysis showing the correlation between observed and predicted pIC50 values for the Training set



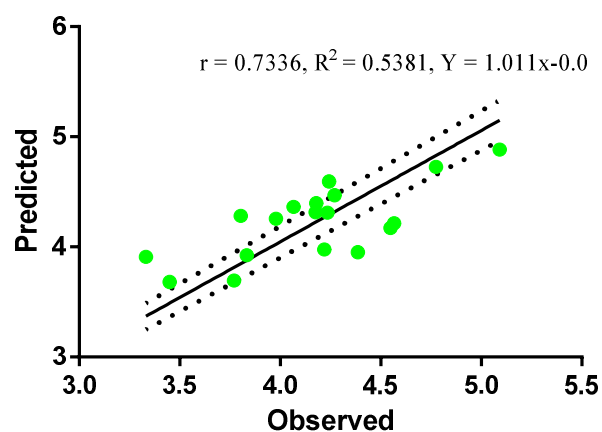
**Figure 2** MLR analysis showing the correlation between observed and predicted pIC50 values for the Test set

The PLS regression was commenced with 497 descriptors. Descriptors with insignificant regression coefficients were removed until the optimum  $Q^2$  was reached. The number of descriptors used for PLS model was found to be the same as that of the MLR regression analysis. The PLS model equation was obtained by PLS regression analysis. Figure 1-4 depicts the scatter plot between observed and predicted values for training and test set for MLR and PLS regression (Table 4 and 5) analysis respectively.

The 3D-QSAR equation fitted with MLR present a significant relationship between dependent variable (pIC50) and independent variables (selected 3D descriptors). The  $R^2$  regression coefficient (0.9746) depicts an excellent and strong ~ 97% correlation between the bioactivity and selected calculated descriptors in the training dataset, while the value of Jack-knifing ( $Q^2$ ) depicts an approximate 90% (0.900) prediction power and accuracy of this 3D QSAR model in contrast to PLS regression analysis whose  $R^2$  regression coefficient is ~ 96% and Jack-knifing ( $Q^2$ ) value is ~82%.



**Figure 3** PLS analysis showing the correlation between observed and predicted pIC50 values for the Train set



**Figure 4** PLS analysis showing the correlation between observed and predicted pIC50 values for the Test set

**Table 4** 3D-QSAR model using Partial Linear Regression Model (PLS)

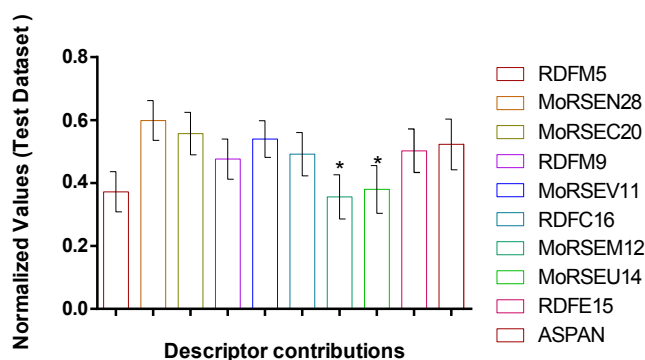
PLS model
$\text{PIC50} = 0.00 + 0.52435 (\text{RDFM5}) + 0.16919 (\text{MoRSEN28}) + 0.17851$



$$(MoRSEC20) + 0.8699 (RDFM9) + 0.25968 (MoRSEV11) + 0.4563 (RDFC16) - 0.40636 (MoRSEM12) - 0.32786 (MoRSEU14) + 0.4919 (RDFF15) + 0.74763 (ASPAN) + 0.39579 (E3u) + 0.44726 (MoRSEM30) - 0.11591 (MoRSEM9) + 0.20099 (RDFC25) + 0.16366 (RDFU8) - 0.45952 (RDFF5) - 0.30715 (P2m) + 0.18549 (MoRSEV13) - 0.14935 (Petitj3D) - 0.01257 (RDFFC20)$$

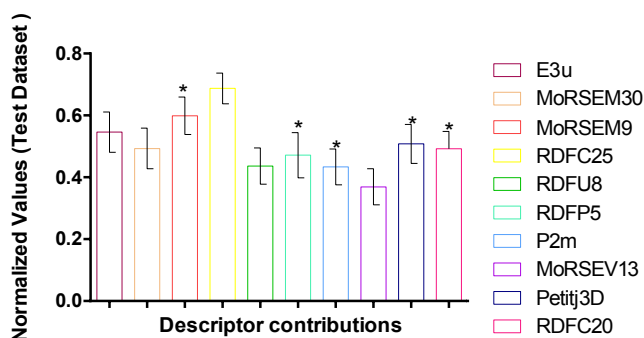
**Table 5** Partial Linear Regression Model (PLS) internal and external validation parameters

Internal Validation Parameters	Threshold value
$R^2 = 0.95874$	$Q^2 = 0.90004$ Passed (Threshold value $Q^2 > 0.5$ )
<b>Leave-One-Out (LOO) Result</b>	$r^2 = 0.64154$ Passed (Threshold value $r^2 > 0.6$ )
$Q^2 = 0.82479$ , Average $R_m^2$ (Train; LOO) = 0.76737, $MAE_{Train} = 0.20813$ , $SD_{Train} = 0.17464$	$ r^2 - r_0^2  = 0.22638$ Passed (Threshold value $ r^2 - r_0^2  < 0.3$ )
<b>External Validation Parameters</b>	$k = 1 = 0.0342$ [ $(r^2 - r_0^2)/r^2$ ] = 0.00059
	OR*
$Q^2F1 = 0.78842$ , $Q^2F2 = 0.51307$ , $R_m^2$ (Test) = 0.36207, $CCC_{Test} = 0.69454$ , $MAE_{Test} = 0.25954$ , $SD_{Test} = 0.14848$	$k' = 0.99289$ [ $(r^2 - r_0^2)/r^2$ ] = 0.35346 Passed (Threshold value: [ $0.85 < k < 1.15$ and $((r^2 - r_0^2)/r^2) < 0.1$ ])



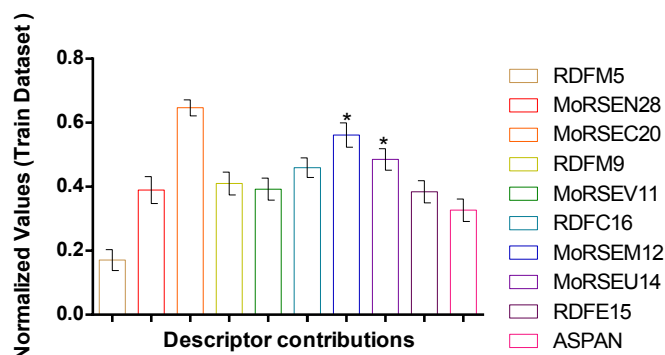
**Figure 5a** 3D Descriptor contributions for MLR and PLS analysis for test dataset

\* Descriptors with negative contribution



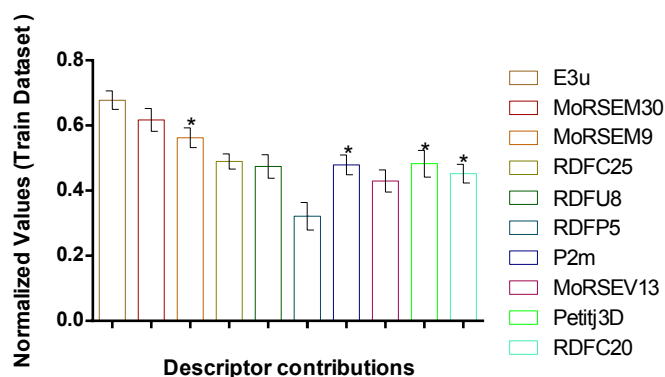
**Figure 5b** 3D Descriptor contributions for MLR and PLS analysis for test dataset

\* Descriptors with negative contribution



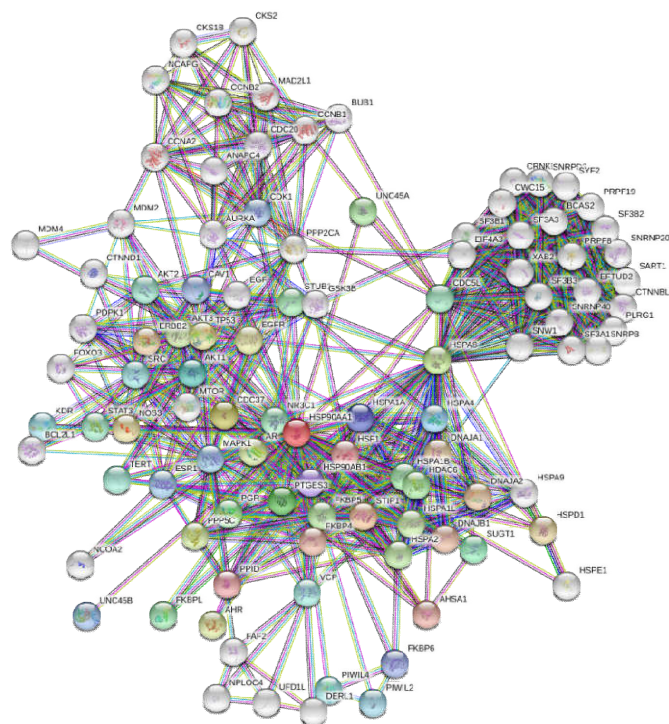
**Figure 6a** 3D Descriptor contributions for MLR and PLS analysis for train dataset

\* Descriptors with negative contribution

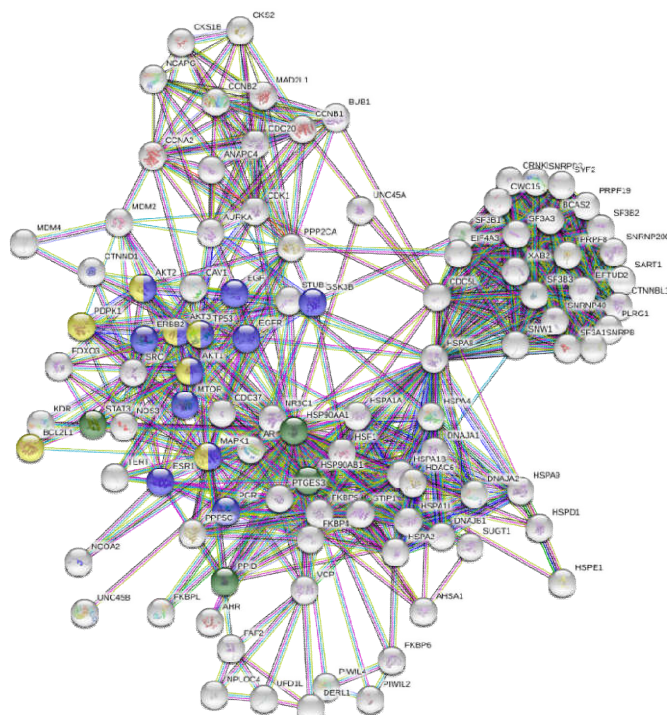


**Figure 6b** 3D Descriptor contributions for MLR and PLS analysis for train dataset

\* Descriptors with negative contribution



**Figure 7a** Protein-Protein Interaction of HSP90 indicating 101 nodes, 750 edges, with 0.745 as the average local clustering coefficient. Colored nodes; query protein and first shell of interactors, while white nodes are second shell of interactors.



**Figure 7b** Protein-Protein Interaction of HSP90. Blue nodes indicate proteins (MTOR, ESR1, PGR, AKT 1-3, EGFR, ERBB2, EGF, GSK3B) that play critical roles along breast cancer pathophysiology; Yellow nodes indicate proteins (AKT 1-2, PDPK1, BCL2L1, MAPK1, and TP53) with apoptotic role, deep green nodes indicate necroptotic roles of HSP90AA1, PTGES3, PPIID, and STAT3

Varied proteins associated with cell growth and survival has been linked with the maturation and stabilization roles of Heat shock protein (HSP90). In addition to this, a recent study affirmed that the growth and aversion of apoptosis by carcinogenic cells has also been linked with Hsp90 and its inhibition has been strongly associated with induction of apoptosis. Programmed cell death can also be triggered by varied extrinsic and intrinsic signaling pathways. Specifically, necroptosis, ferroptosis, and pyroptosis are examples of regulated cell death associated with inflammation which could be either beneficial or harmful<sup>25</sup>. Most Hsp90 inhibitors bind either the N-terminal or C-terminal to inhibit ATPase activity<sup>26</sup> and some also bind the middle domain region where client protein binds to elicit wide range of physiological actions. 8-arylsulfanyl, 8-arylsulfoxyl, and 8-arylsulfonyl adenines been an antagonist of Hsp90 may disrupt the complex association of Hsp90 and CDC37, hence triggers apoptosis and disruption of necroptotic signaling pathways<sup>26</sup> in carcinogenic cells. Further Clinical studies are highly recommended.

## CONCLUSION

In this study, we evaluated the structure-activity relationship using 3D descriptors of a series of 8-arylsulfanyl, 8-arylsulfoxyl, and 8-arylsulfonyl adenines in order to evaluate HSP90 inhibitors. Both MLR and PLS regression analysis was used to develop statistically significant models which was further validated via cross validation using LOO methodology. The models show good predictive potentials for HSP90 inhibitors which can be used to predict new HSP90 inhibitors. These 3D QSAR models and protein-protein-interaction stipulated in these studies could provide reliable tools for the

design of HSP90 inhibitors with varied pharmacodynamics and pharmacokinetic properties.

## References

- 1 Florian HS. Maximilian MB. Johannes B. The HSP90 chaperone machinery. *Nature reviews | Molecular cell biology* 2017; volume18: 359
- 2 Pratt WB. Toft DO. Steroid receptor interactions with heat shock protein and immunophilin chaperones. *Endocr. Rev.* 1997; 18: 306–360
- 3 Ferlay J. Shin HR. Bray *Fet al.* Estimates of worldwide burden of cancer in 2008: GLOBOCAN 2008. *Int J Cancer.*2010; 127: 2893–917
- 4 Perou CM. Sorlie T. Eisen MB *et al.* Molecular portraits of human breast tumours. *Nature.*2000; 406: 747-52
- 5 De Mattos-Arruda L. Javier C. Breast cancer and HSP90 inhibitors: Is there a role beyond the HER2-positive subtype? *The Breast.*2012; 21: 604-607
- 6 Slamon DJ. Godolphin W. Jones LA *et al.* Studies of the HER-2/neu proto-oncogene in human breast and ovarian cancer. *Science* 1989; 244 (4905): 707-712
- 7 Allred DC. Clark GM. Molina R *et al.* Overexpression of HER-2/neu and its relationship with other prognostic factors change during the progression of in situ to invasive breast cancer. *Hum. Pathol.*1992; 23 (9): 974-979
- 8 Pearl LH. Prodromou C. Structure and mechanism of the Hsp90 molecular chaperone machinery. *Annu Rev Biochem.*2006; 75: 271-94
- 9 Workman P. Burrows F. Neckers L. Rosen N. Drugging the cancer chaperone HSP90: combinatorial therapeutic exploitation of oncogene addiction and tumor stress. *Ann N Y Acad Sci.*2007; 1113: 202-16
- 10 Trepel J, Mollapour M, Giaccone G, Neckers L. (2010). Targeting the dynamic HSP90 complex in cancer. *Nat Rev Cancer*; 10:537-49
- 11 Maloney A. Workman P. HSP90 as a new therapeutic target for cancer therapy: the story unfolds. *Expert OpinBioIther.*2002; 2: 3-24
- 12 Prodromou C. Pearl LH. Structure and functional relationships of Hsp90. *Curr Cancer Drug Targets.*2003; 3: 301-23
- 13 Mimnaugh EG. Chavany C. Neckers L. Polyubiquitination and proteosomal degradation of the p185c-erbB-2 receptor protein-tyrosine kinase induced by geldanamycin. *J Biol Chem.* 1996; 271: 22796-801
- 14 Basso AD. Solit DB. Chiosis G *et al.* Akt forms an intracellular complex with heat shock protein 90 (Hsp90) and Cdc37 and is destabilized by inhibitors of Hsp90 function. *J Biol Chem.*2002; 277: 39858-66
- 15 Xing H. Weng D. Chen G *et al.* Activation of fibronectin/ PI-3K/Akt2 leads to chemoresistance to docetaxel by regulating surviving protein expression in ovarian and breast cancer cells. *Cancer Lett.*2008; 261:108-19
- 16 Xu W. Soga S. Beebe K *et al.* Sensitivity of epidermal growth factor receptor and ErbB2 exon 20 insertion mutants to Hsp90 inhibition. *Br J Cancer.*2007; 97: 741-4

- 17 Workman P. Burrows F. Neckers L. Rosen N. Drugging the cancer chaperone HSP90: combinatorial therapeutic exploitation of oncogene addiction and tumor stress. *Ann N Y AcadSci*; 2007; 1113: 202-16.
- 18 Basso AD. Solit DB. Munster PN. Rosen N. Ansamycin antibiotics inhibit Akt activation and cyclin D expression in breast cancer cells that overexpress HER2. *Oncogene*.2002; 21: 1159-66
- 19 Chen X. Liu. P. Wang Q *et al.* DCZ3112, a novel Hsp90 inhibitor, exerts potent antitumor activity against HER2-positive breast cancer through disruption of Hsp90-Cdc37 interaction, *Cancer Letters* 2018, doi: 10.1016/j.canlet.2018.07.012
- 20 Kolossov E. Stanforth R. The quality of QSAR models: problems and solutions. *SAR and QSAR in Environmental Research*. 2007; 18(1-2): 89-100
- 21 Roy PP. Roy K. On some aspects of variable selection for partial least squares regression models. *QSAR & Combinatorial Science*. 2008; 27(3): 302-313
- 22 Leach AR. Leach AR. *Molecular modelling: principles and applications*. Pearson education. 2001
- 23 Shao. J. Linear model selection by cross-validation. *Journal of the American statistical Association*. 1993; 88(422): 486-494
- 24 Golbraikh A. Tropsha A. Beware of q<sup>2</sup>!. *Journal of molecular graphics and modelling*, 2002; 20(4): 269-276
- 25 Zhou W. Yuan J. Necroptosis in health and diseases. In *Seminars in cell & developmental biology*.2014;Vol. 35: pp. 14-23
- 26 Patki JM. Pawar SS. *Pathol. Oncol. Res.* 2013; 19: 631–640

**How to cite this article:**

Abiodun Julius Arannilewa et al. 2019, PPI of Hsp90 and 3d-qsar Evaluation Study of Purine Derivatives As Anti-Carcinogenic substrates of hsp90-Overexpressed Malignant Female Breast Cancer. *Int J Recent Sci Res.* 10(01), pp. 30529-30539. DOI: <http://dx.doi.org/10.24327/ijrsr.2019.1001.3077>

\*\*\*\*\*