Special Conference Edition, June, 2023 http://dx.doi.org/10.4314/bajopas.v14i1.44S

Bayero Journal of Pure and Applied Sciences, 14(1): 270 - 277
ISSN 2006 – 6996

QUANTITATIVE STRUCTURE-ACTIVITY RELATIONSHIP (QSAR) MODELLING OF SOME 2, 4-DIAMINOPYRIMIDINES DERIVATIVES AS EPIDERMAL GROWTH FACTOR RECEPTOR (EGFR) MUTANT INHIBITORS

Muhammad Tukur Ibrahim[1]* Adamu Uzairu[1], Umma Muhammad[2] and Aisha Muhammad Danmallam[3].

[1]Department of Chemistry, Faculty of Physical Science, Ahmadu Bello University, P.M.B 1045, Zaria, Kaduna State Nigeria.

[2]Department of Science Laboratory Technology, School of Technology, Kano State Polytechnic, Kano State Nigeria.

[3]Department of Chemistry, Federal University Birnin Kebbi, Kebbi State Nigeria.

*Corresponding author E-mail: muhdtk1988@gmail.com (M. T. Ibrahim)

ABSTRACT

Non-small cell lung cancer (NSCLC) accounted for about 80% to 85% of the lung cancers while the Oat cell lung cancer accounted for about 15 to 20% remaining lung cancer cases. The key approaches used in the management of NSCLC include operation, radiotherapy, targeted therapy and chemotherapy. In spite of development in the management modalities, the diagnosis in NSCLC patients has not significantly improved. The aim of this research was to carry out QSAR modelling on some EGFR mutant inhibitors to develop a highly predicted mathematical model with good internal and external validations. Density Functional Theory method at B3LYP/6-311G* level of theory was used to identify the optimum conformation of the investigated compounds. A Multi-linear regression Genetic Function Algorithm method (MLR-GFA) was used to build four different mathematical models using QSAR modeling technique. Out of the four models built, the first model was chosen and reported due to its statistical significance with $R^2 = 0.952$, $R^2_{adj} = 0.936$, $Q^2_{CV} = 0.898$, $R^2_{test} = 0.852$ and $cRp^2 = 0.840$. The R^2 value of the chosen model displayed that the model can be able to describe about 95.2% of the variation in the EGFR inhibitory activities of the 2,4-diaminopyrimidines derivatives. The chosen model was further subjected to applicability domain to measure the quality of estimation by the chosen model. All the studied compounds were found to be within the chemical space of the model which ascertain the quality of estimation of the chosen model and foil the waste of the results acquired by the model.

Key words: QSAR Modelling, 2,4-diaminopyrimidines, EGFR Mutant, Inhibitors.

INTRODUCTION

Cancer of lung is known to be the principal source of tumor-linked mortality worldwide. Non-small cell lung cancer (NSCLC) and Small cell lung cancer are the two main classifications of lung cancer, respectively (Bistrović *et al.*, 2018). The common and lethal type of lung cancer with close to 2.0 million reported cases and survival rate of not less than 20% in each 5-years after diagnosis at all stages and survival rate of not more than 2.0% in each 5-years at the fourth (IV) stage was NSCLC that comprises bulky cell carcinoma, glandular carcinoma, and squamous cell carcinoma (Khuri, 2016; Tièche *et al.*, 2016). NSCLC accounted for about 80% to 85% of the lung cancers while the Oat cell lung cancer accounted for about 15 to 20% remaining lung cancer cases(Ibrahim *et al.*, 2021a). The key approaches used in the management of NSCLC include operation, radiotherapy targeted therapy and chemotherapy. In spite of development in the management modalities, the diagnosis in NSCLC patients has not significantly improved (Wang *et al.*, 2022).

In 1956, epidermal growth factor receptor (EGFR) was first uncovered by a scientist called Stanley Cohen while working in the USA at University of Vanderbilt (Ibrahim *et al.*, 2022; Sato *et al.*, 2018). As distributed in the mammalian cell surfaces as well as cells epithelial, fibroblasts, cells glial, keratinocytes, EGFR which belongs to tyrosine kinase receptor family plays a significant role in key cellular processes which include propagation, variation, movement, apoptosis, and angiogenesis and in many cancers. EGFR overexpression has been described in several cancers. EGFR was also identified as the main target for the combating of tumor related illness like NSCLC (Hanan *et al.*, 2016; Ibrahim *et al.*, 2021b). Consequently, developing potent EGFR inhibitors is one of the research hotspots for the management of human tumor most especially NSCLC (Othman *et al.*, 2021; Tu *et al.*, 2017).

EGFR Tyrosine kinase inhibitors (TKIs) targeting the NSCLC, such as Erlotinib and Gefitinib, were initially designed to inhibit the growth of tumors at the EGFR ATP-binding site. Moreover, these small molecules were used as the initial choice in the management of NSCLC for the activated, classical and the gatekeeper classes of EGFR mutations (Hsu *et al.*, 2018; Passaro *et al.*, 2021). Erlotinib as well as Gefitinib provide substantial medical assistance in the management of NSCLC by protecting the EGFR mutations in exon 19 (activating (L858R)) and exon 21 (classical (del. E746-A750)), respectively(Song *et al.*, 2017). Besides,

patients with activating (L858R) EGFR mutations, usually recovered when treated with gefitinib and erlotinib the so-called first-generation EGFR-TKIs. Though, the period for their efficiency was actually inadequate as a result of the establishment of resistance by the EGFR-T790M mutation (gatekeeper mutation) that was observed in about half (50%) of patients with such EGFR-T790M mutation (Ibrahim *et al.*, 2020). Dacomitinib and Afatinib were subsequently designed and developed to control the established resistance by the EGFR-T790M mutation to the EGFR-TKIs of the first-generation (Pawara *et al.*, 2021; Song *et al.*, 2017). Regrettably, the efficacy of Dacomitinib and Afatinib was inadequate as a result of severe side effects which include rashes of the skin and gastrointestinal toxicity (Patel *et al.*, 2017; Pawara *et al.*, 2021). In order to also fix and resolve the resistance by EGFR-T790M mutation while being more selective to the $EGFR^{WT}$ and observed toxicity to the second-generation irreversible EGFR-TKIs, Osimertinib (AZD9291) and Rociletinib (CO-1686) were further designed and established. But Osimertinib and Rociletinib could not accomplish the specified role owing the immergence of resistance by the EGFR-C797S mutation (Chen *et al.*, 2017; Karnik *et al.*, 2021; Pawara *et al.*, 2021).

Pyrimidine in general belongs to an electron rich nitrogen containing heterocycle compounds with different biological and pharmacological activities, and is used for therapeutic applications in a broad spectrum of human diseases (Kim *et al.*, 2020; Mahapatra *et al.*, 2021). The aim of this research is to carry out QSAR modelling on some EGFR mutant inhibitors to develop a highly predicted mathematical model with good internal and external validation parameters.

METHODS

Sourcing of dataset and the transformation of EGFR inhibitory activities

Twenty-seven (27) set of 2,4-diaminopyrimidines derivatives as EGFR mutant inhibitors were synthesized and reported by Cheng and co-workers (Chen *et al.*, 2017). For the purpose of this study, EGFR inhibitory activities against EGFR kinase (the defined biological activities for this research) were carefully chosen and utilized as stipulated by the OECD (organization for economic co-operation and development) principle 1, which clearly stated that for any QSAR study there should be a well-defined inhibitory activity. The EGFR kinases inhibitory activities (EGFR IC_{50}) were reported in their micro-molar (μM) unit and were subsequently transformed into their logarithm form (pIC_{50}) by means of equation 1(Ibrahim *et al.*, 2021b). Table 1 presents the SMILES and the transformed EGFR pIC_{50} of the 27 set of 2,4-diaminopyrimidines derivatives.

$$pIC_{50} = -\log (IC_{50} * 10^{-6}) \quad \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots (1)$$

Table 1 SMILES and the EGFR pIC_{50} of the 2,4-diaminopyrimidines derivatives.

S/N	SMILES	pIC_{50}
1	CN1CCN(CC1)c1ccc(cc1)Nc1nccc(n1)Nc1ccc2c(c1)cc[nH]2	6.658
2	CN1CCN(CC1)c1ccc(cc1)Nc1ncc(c(n1)Nc1ccc2c(c1)cc[nH]2)C(F)(F)F	6.917
3	CN1CCN(CC1)c1ccc(cc1)Nc1ncc(c(n1)Nc1ccc2c(c1)cc[nH]2)[N+](=O)[O-]	6.043
4	COc1cnc(nc1Nc1ccc2c(c1)cc[nH]2)Nc1ccc(cc1)N1CCN(CC1)C	6.810
5	CN1CCN(CC1)c1ccc(cc1)Nc1ncc(c(n1)Nc1ccc2c(c1)cc[nH]2)C	6.767
6	CN1CCN(CC1)c1ccc(cc1)Nc1ncc(c(n1)Nc1ccc2c(c1)cc[nH]2)F	6.742
7	CN1CCN(CC1)c1ccc(cc1)Nc1ncc(c(n1)Nc1ccc2c(c1)cc[nH]2)Br	7.456
8	CN1CCN(CC1)c1ccc(cc1)Nc1ncc(c(n1)Nc1ccc2c(c1)cc[nH]2)c1ncc2c(n1)[nH]nc2	6.854
9	CN1CCN(CC1)c1ccc(cc1)Nc1ncc(c(n1)Nc1ccc2c(c1)[nH]cc2)Br	7.721
10	CN1CCN(CC1)c1ccc(cc1)Nc1ncc(c(n1)Nc1ccc2c(c1)cn[nH]2)Br	7.409
11	CN1CCN(CC1)c1ccc(cc1)Nc1ncc(c(n1)Nc1ccc2c(c1)[nH]nc2)Br	7.367
12	CN1CCN(CC1)c1ccc(cc1)Nc1ncc(c(n1)Nc1ccc2c(c1)ccn2Cc1cccc(c1)F)Br	7.310
13	CN1CCN(CC1)c1ccc(cc1)Nc1ncc(c(n1)Nc1ccc2c(c1)n(C)cc2)Br	7.678
14	CN1CCN(CC1)c1ccc(cc1)Nc1ncc(c(n1)Nc1ccc2c(c1)CCC2)Br	7.456
15	CN1CCN(CC1)c1ccc(cc1)Nc1ncc(c(n1)Nc1ccc2c(c1)CCN2C(=O)C)Br	5.222
16	CN1CCN(CC1)c1ccc(cc1)Nc1ncc(c(n1)Nc1ccc2c(c1)N(CC2)C(=O)C)Br	5.066
17	CN1CCN(CC1)c1ccc(cc1)Nc1ncc(c(n1)Nc1ccc2c(c1)cccn2)Br	5.108
18	CN1CCN(CC1)c1ccc(cc1)Nc1ncc(c(n1)Nc1ccc2c(c1)nccn2)Br	6.627
19	C=CC(=O)Nc1cc(ccc1N(CCN(C)C)C)Nc1ncc(c(n1)Nc1ccc2c(c1)[nH]cc2)Br	6.699
20	C=CC(=O)Nc1cc(ccc1N1CCN(CC1)C)Nc1ncc(c(n1)Nc1ccc2c(c1)[nH]cc2)Br	4.785
21	C=CC(=O)Nc1cc(Nc2ncc(c(n2)Nc2ccc3c(c2)[nH]cc3)Br)c(cc1N1CCN(CC1)C)OC	5.569
22	C=CC(=O)Nc1cc(ccc1N1CCN(CC1)C)Nc1ncc(c(n1)Nc1ccc2c(c1)n(C)cc2)Br	4.556
23	C=CC(=O)Nc1cc(Nc2ncc(c(n2)Nc2ccc3c(c2)n(C)cc3)Br)c(cc1N1CCN(CC1)C)OC	5.387
24	C=CC(=O)Nc1cc(ccc1N(CCN(C)C)C)Nc1ncc(c(n1)Nc1ccc2c(c1)CCC2)Br	6.222
25	C=CC(=O)Nc1cc(ccc1N1CCN(CC1)C)Nc1ncc(c(n1)Nc1ccc2c(c1)CCC2)Br	4.558
26	C=CC(=O)Nc1cc(Nc2ncc(c(n2)Nc2ccc3c(c2)CCC3)Br)c(cc1N1CCN(CC1)C)OC	5.495
27	C=CC(=O)Nc1cc(Nc2nccc(n2)Nc2ccc3c(c2)n(C)cc3)c(cc1N1CCN(CC1)C)OC	5.076

2D-structures sketching of the obtained dataset and optimum conformational search

The 2D-conformations/structures of the obtained dataset were sketched with the help of Chemdraw software (Anebi *et al.*, 2021; Mills, 2006). The optimum conformational search for all the 27 set of 2,4-diaminopyrimidines derivatives in this study were achieved at the MMFF with DFT at B3LYP/6-311G* theory level, respectively (Kohn *et al.*, 1996; Pawara *et al.*, 2021).

Determination, pre-treatment and separation of chemical descriptors

The determination of chemical descriptors in this research work was achieved by the use of PaDEL descriptor tool (also known as pharmaceutical data exploration laboratory (PaDEL) tool kit) (Grisoni *et al.*, 2018; Yap, 2011). PaDEL descriptor tool is able to compute up to about 1875 chemical descriptors such as one-dimensional (1D), two-dimensional (2D) and three-dimensional (3D) chemical descriptors. Version 1.2 of the data pre-treatment software which remove constant, unproductive and redundant chemical descriptors was adopted to checkmate and pre-treat the determined chemical descriptors (Ambure *et al.*, 2015). Furthermore, version 1.2 of data division software was adopted for the separation of the pre-treated chemical descriptors into about 21 compounds as model training set (80%) and about 6 compounds as test set (20%), respectively (Kennard and Stone, 1969).

Variable selection and evaluation

According to OECD principle 2, in any QSAR modelling study an ambiguous algorithm must be dodged/avoided. As such, multi-linear regression genetic function algorithm (MLR-GFA) technique was utilized as the procedure for the variable selection (model development**)** as a result of its high relevance in molecular modelling (specifically QSAR modelling). Herein, the determined inhibitory activities (EGFR kinases inhibitory activities) were taken as the dependent variables and the determined, treated and training set descriptors were taken as independent variables, respectively (Hansch and Fujita, 1964; Ibrahim *et al.*, 2018).

OECD principle 3 states that, the applicability domain (AD) of a model needs to be revealed so as to ensure whether the model can effectively make good prediction of inhibitory activities or not of new compounds that were not used initially in generating the model. For this reason in this study, the leverage technique was employed to display the AD of the studied model (Tropsha *et al.*, 2003).

OECD principle 4 states that, the productivity of a model must be assessed by subjecting the model into rigorous validation and evaluation before that model can be referred as a valid one. Besides that, the model must respectively pass those validation and evaluation processes. Among these validation processes are internal validation, external validation and MLR-Y-randomization test, respectively (Bajorath, 2015).

RESULTS AND DISCUSSION

On Twenty-seven (27) set of 2,4-diaminopyrimidines derivatives as EGFR mutant inhibitors, QSAR modelling was employed to build a mathematical model with good predictive performance. The whole Twenty-seven (27) set of 2,4-diaminopyrimidines derivatives was divided into 21 compounds as model training set (80%) and about 6 compounds as test set (20%), respectively. The 21 compounds (training set) were used to develop three models. Among the three models established, the first model as represented by equation 2 was chosen and reported due to its high values of R^2 = 0.952, R^2_{adj} = 0.936, Q^2_{CV} = 0.898, R^2_{test} = 0.852 and cRp^2 = 0.840 as presented in Table 2 (Tropsha and Golbraikh, 2010; Veerasamy *et al.*, 2011). The R^2 value of the chosen model displayed that the model can be able to describe about 95.2% of the variation in the EGFR inhibitory activities of the 2,4-diaminopyrimidines derivatives. The theoretical R^2 value of the model was better and higher than its equivalent R^2_{adj} which depicts the importance of the model and ascertained the model's free from over parameterization. The explanation and categories of the descriptors contained in the reported model is presented in Table 2.

Chosen model

pIC₅₀ = 175.454347482***AATSC0c**+ 9.116285871***MATS5v**– 18.126176241***SpMin1_Bhm**– 11.365539786***SpMin6_Bhv**+9.910911477***CIC0**+8.124386 ……………………………….2

Table 2 Statistical parameters of model 1 and accepted values for assessment of QSAR models

Symbol	Name	Accepted Value	Chosen Model
R^2	Squared correlation co-efficient	≥0.6	0.952
R^2_{adj}	Adjusted squared correlation co-efficient	≥0.6	0.936
Q^2_{cv}	Cross-Validation Co-efficient	≥0.5	0.898
R^2- Q^2	Difference between R^2 and Q^2	≤0.3	0.054
$N_{(ext., \& test set)}$	Minimum number of external and test set	≥ 5	27
R^2_{test}	Squared correlation co-efficient of test set	≥0.5	0.852
cRp^2	Y-scrambling parameter	≥0.5	0.840

The values of the variation inflation factor (VIF) of the molecular descriptors contained in the training model building set ware calculated in order to find out whether there was an occurrence of problem associated to multicollinearity in the model or not. These computed values are expected to be less than 10 signifying the competence of the model. If the computed values are less than 10 then the molecular descriptors contained in the model are said to be statistically significant and orthogonal (Table 3). The computed VIF values of the molecular descriptors contained in the model were all less than 10 which

displayed the goodness of the model. And also, the model was free from problems associated to multicollinearity, as such the model can be accepted and used (Beheshti *et al.*, 2016).

The mean effect on the molecular descriptors contained in the training model building set were as well calculated in order to determine the impact and significance of a molecular descriptor in comparison to others in the model on the EGFR inhibitory activities of the 2,4-diaminopyrimidines derivatives against EGFR kinases. The positive or negative coefficients of each value of a molecular descriptor indicate the magnitude, impact and significance of such property on the EGFR inhibitory activities of the investigated compounds. Based on the computed mean effect of the molecular descriptors contained in the reported model (Table 3), SpMin1_Bhm molecular descriptor has the highest value of +20.63691, followed by SpMin6_Bhv with a value of +9.554734, followed by AATSC0c with a value of -1.32315, followed by MATS5v with a value of -2.94932, and then lastly CIC0 with a value of -24.9192, respectively. The co-efficient and sign of SpMin1_Bhm molecular descriptor shows that if such a property is increased in the main structures of the investigated compounds their potency might rise and reverse is the case. This applies to all molecular descriptors with positive signs. For molecular descriptors with a negative sign like AATSC0c, when such property is decreased from the main structures of the investigated compounds, their potency as well might rise and vice versa. Also, this applies to all molecular descriptors with negative signs. The trend in mean effect of these molecular descriptors is shown below (Adedirin *et al.*, 2018).

SpMin1_Bhm > **SpMin6_Bhv** > **AATSC0c** > **MATS5v** > **CIC0**

Table 3 Representation, explanation, category, VIF, and ME of the molecular descriptors contained in the chosen model.

S/No	Representation	Explanation	Category	VIF	ME
1	**AATSC0c**	Average centered Broto-Moreau autocorrelation – lag 0 / weighted by charges	2D	3.294636	-1.32315
2	**MATS5v**	Moran autocorrelation – lag 5 / weighted by van der Waals volumes	2D	2.557591	-2.94932
3	**SpMin1_Bhm**	Smallest absolute eigenvalue of Burden modified matrix – n 1 / weighted by relative mass	2D	1.263577	20.63691
4	**SpMin6_Bhv**	Smallest absolute eigenvalue of Burden modified matrix – n 6 / weighted by relative van der Waals volumes	2D	9.271082	9.554734
5	**CIC0**	Complementary information content index (neighborhood symmetry of 0-order)	2D	7.676824	-24.9192

The scatter plot of biological EGFR inhibitory activities against the modelled EGFR inhibitory activities of both the training and test sets for the chosen model which is presented in Figure 1 was used to establish the consistency of the chosen model. The internal squared correlation co-efficient value (R^2) of 0.952 of the chosen model was detected to have agreed and also be exact with that inferred from Figure 2 (R^2 value of 0.952). This further established the reliability of the chosen model. The scatter plot of biological EGFR inhibitory activities against computed residuals of both the training and test sets for the chosen model as portrayed by Figure 2 is utilized to identify whether there are methodological and systematic errors in the chosen model. The unanticipated spread of the computed residuals on Figure 3 further affirmed that chosen model was free from the presence of the methodological and systematic errors.

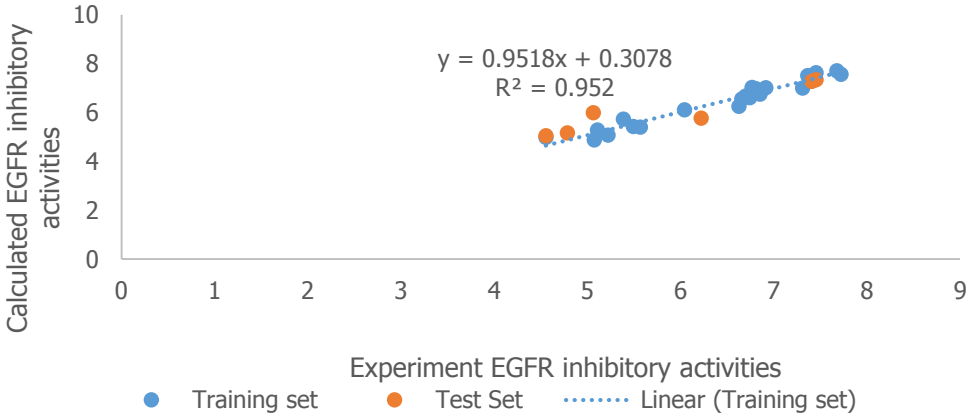

Figure 1 Experimental EGFR inhibitory activities against calculated activities

Figure 2 Experimental EGFR inhibitory activities against residual values

The experimental EGFR inhibitory activities, calculated activities and the calculated residual values of both the training and test sets for the equation 2 is revealed in Table 4. An equation (model) can be assumed to be high predictive if at all the computed residual values (the difference between the actual experiment and the calculated activities) are relatively low. For the chosen model, it was recognized to be high predictive by the low computed residual as essentially seen on Table 4.

Table 4 Experimental, calculated EGFR inhibitory activities and residual values.

Name	Experimental activities	Calculated activities	Residuals
1	6.658	6.554	0.103
2	6.917	7.003	-0.085
3	6.043	6.100	-0.057
4	6.810	6.973	-0.163
5	6.767	7.023	-0.256
6	6.742	6.597	0.145
7 [V]	7.456	7.332	-0.124
8	6.854	6.748	0.106
9	7.721	7.559	0.162
10 [V]	7.409	7.264	-0.144
11	7.367	7.503	-0.137
12	7.310	6.999	0.310
13	7.678	7.708	-0.030
14	7.456	7.627	-0.171
15	5.222	5.067	0.154
16 [V]	5.066	5.981	0.915
17	5.108	5.279	-0.171
18	6.627	6.248	0.379
19	6.699	6.661	0.038
20 [V]	4.785	5.163	0.378
21	5.569	5.403	0.166
22	4.556	4.978	-0.423
23	5.387	5.724	-0.337
24 [V]	6.222	5.766	-0.456
25 [V]	4.558	5.037	0.480
26	5.495	5.427	0.068
27	5.076	4.876	0.199

[V] = Validation test set

The Williams diagram of the reported equation 2 is presented in Figure 3. It evidently depicted the applicability domain (also known as the chemical space) of the chosen model. From the Figure (williams diagram), no compound with a standardized residuals > $\pm 3\delta$ (outlier compound) and compound with leverage values > h*= 0.857 (influential compounds) were recognized. Meaning that all the compounds were found to be within the chemical space of the chosen model which ascertain the quality of estimation made by the chosen model.

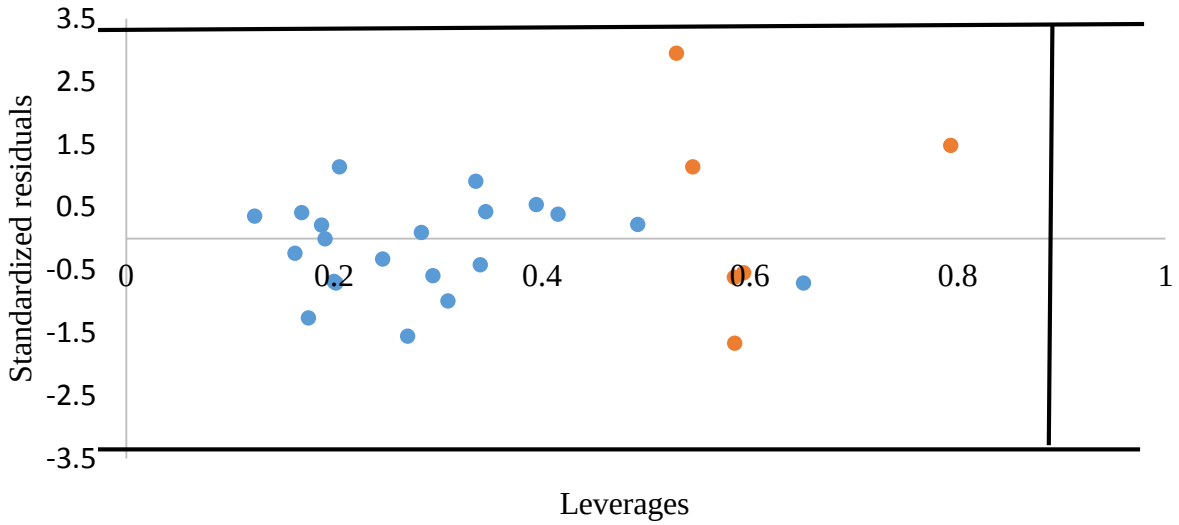

Figure 3 Williams Diagram of the chosen model

The MLR-Y scrambling correlation by chance test on the 21 compounds training test for the chosen model was shown on Table 5. Ten (10) diverse equations or models are developed by varying the EGFR inhibitory activities and keeping the chemical descriptors unchanged in the MLR-Y scrambling correlation by chance test. The 10 diverse equations or models developed ware expected to have low R^2 and Q^2 values. As a result of this, the chosen model was found to be robust with low R^2 and Q^2 values for numerous counts. Furthermore, the chosen model was not generated by chance correlation as the term ($^{C}R^2_P$) to assess the test has value of 0.840 which is greater than 0.5 the threshold limit set, respectively.

Table 5 Randomization chance correlation test.

Model	R	R^2	Q^2
Original	0.950918	0.904245	0.798187
Random 1	0.639738	0.409265	0.003664
Random 2	0.397006	0.157614	-0.41182
Random 3	0.254866	0.064957	-0.707
Random 4	0.144868	0.020987	-0.62367
Random 5	0.349818	0.122373	-0.6628
Random 6	0.222061	0.049311	-0.52743
Random 7	0.505503	0.255533	-0.32732
Random 8	0.375678	0.141134	-0.40384
Random 9	0.206491	0.042639	-0.725
Random 10	0.431415	0.186119	-0.2768
	Random Models Parameters		
Average r:	0.352744		
Average r^2:	0.144993		
Average Q^2:	-0.4662		
cRp^2:	0.840		

CONCLUSION

A very high predictive QSAR model was established and reported on twenty-seven (27) set of 2,4-diaminopyrimidines derivatives as EGFR mutant inhibitors with high R^2 value of 0.952, R^2_{adj} value of 0.936, Q^2_{CV} value of 0.898, R^2_{test} value of 0.852 and cRp^2 value of 0.840 as the validation parameters, respectively. The Williams diagram showed that all the compounds were within the chemical space of the reported model which determined the quality of estimation by the reported model and halt the waste of the results established by the reported model. As such, the model is observed to make prediction of the activities of the compounds that were not used in building the model.

ACKNOWLEDGMENTS

The authors sincerely acknowledged Ahmadu Bello University, Zaria for its technical support in the course of this research.

REFERENCES

Adedirin, O., Uzairu, A., Shallangwa, G. A., and Abechi, S. E. (2018). QSAR and molecular docking based design of some n-benzylacetamide as?-aminobutyrate-aminotransferase inhibitors. *The Journal of Engineering and Exact Sciences, 4*(1), 0065-0084.

Ambure, P., Aher, R. B., Gajewicz, A., Puzyn, T., and Roy, K. (2015). "NanoBRIDGES" software: open access tools to perform QSAR and nano-QSAR modeling. *Chemometrics and Intelligent Laboratory Systems, 147*, 1-13.

Anebi, E., Ibrahim, M., Shallangwa, G., Isyaku, S., Abdulsalam, S., and Danmallam, A. (2021). Molecular Docking Study, Drug-Likeness and Pharmacokinetic Properties (ADMET) Prediction of Some Novel Thiophene Derivatives as Salmonella typhi Inhibitors. *Bayero Journal of Pure and Applied Sciences, 14*(2), 235-244.

Bajorath, A. T. J. (2015). Computational Methods for Drug Discovery and Design: ACS Publications.

Beheshti, A., Pourbasheer, E., Nekoei, M., and Vahdani, S. (2016). QSAR modeling of antimalarial activity of urea derivatives using genetic algorithm–multiple linear regressions. *Journal of Saudi Chemical Society, 20*(3), 282-290.

Bistrović, A., Krstulović, L., Harej, A., Grbčić, P., Sedić, M., Koštrun, S., . . . Raić-Malić, S. (2018). Design, synthesis and biological evaluation of novel benzimidazole amidines as potent multi-target inhibitors for the treatment of non-small cell lung cancer. *European Journal of Medicinal Chemistry, 143*, 1616-1634.

Chen, L., Fu, W., Feng, C., Qu, R., Tong, L., Zheng, L., . . . Cai, Y. (2017). Structure-based design and synthesis of 2, 4-diaminopyrimidines as EGFR L858R/T790M selective inhibitors for NSCLC. *European Journal of Medicinal Chemistry, 140*, 510-527.

Grisoni, F., Ballabio, D., Todeschini, R., and Consonni, V. (2018). Molecular descriptors for structure–activity applications: a hands-on approach *Computational Toxicology* (pp. 3-53): Springer.

Hanan, E. J., Baumgardner, M., Bryan, M. C., Chen, Y., Eigenbrot, C., Fan, P., . . . Purkey, H. E. (2016). 4-Aminoindazolyl-dihydrofuro [3, 4-d] pyrimidines as non-covalent inhibitors of mutant epidermal growth factor receptor tyrosine kinase. *Bioorganic & Medicinal Chemistry Letters, 26*(2), 534-539.

Hansch, C., and Fujita, T. (1964). p-σ-π Analysis. A Method for the Correlation of Biological Activity and Chemical Structure. *Journal of the American Chemical Society, 86*(8), 1616-1626.

Hsu, W.-H., Yang, J.-H., Mok, T., and Loong, H. (2018). Overview of current systemic management of EGFR-mutant NSCLC. *Annals of Oncology, 29*, i3-i9.

Ibrahim, M. T., Uzairu, A., and Muhammad, U. (2022). In silico design, anti-proliferative activity modeling, molecular docking and pharmacokinetic properties prediction of some non-small cell lung cancer (NSCLC) therapeutic agents. *Bayero Journal of Pure and Applied Sciences, 13*(1), 72-79.

Ibrahim, M. T., Uzairu, A., Shallangwa, G. A., and Ibrahim, A. (2018). Computational studies of some biscoumarin and biscoumarin thiourea derivatives as alfa-glucosidase inhibitors. *The Journal of Engineering and Exact Sciences, 4*(2), 0276-0285.

Ibrahim, M. T., Uzairu, A., Shallangwa, G. A., and Uba, S. (2020). Lead Identification of Some Anti-Cancer Agents with Prominent Activity Against Non-small Cell Lung Cancer (NSCLC) and Structure-Based Design. *Chemistry Africa, 3*(4), 1023-1044.

Ibrahim, M. T., Uzairu, A., Shallangwa, G. A., and Uba, S. (2021a). Molecular docking investigation and pharmacokinetic properties prediction of some anilinopyrimidines analogues as egfr t790m tyrosine kinase inhibitors. *Egyptian Journal of Basic and Applied Sciences, 8*(1), 203-213.

Ibrahim, M. T., Uzairu, A., Uba, S., and Shallangwa, G. A. (2021b). Design of more potent quinazoline derivatives as EGFRWT inhibitors for the treatment of NSCLC: a computational approach. *Future Journal of Pharmaceutical Sciences, 7*(1), 1-11.

Karnik, K. S., Sarkate, A. P., Tiwari, S. V., Azad, R., Burra, P. V., and Wakte, P. S. (2021). Computational and Synthetic approach with Biological Evaluation of Substituted Quinoline derivatives as small molecule L858R/T790M/C797S triple mutant EGFR inhibitors targeting resistance in Non-Small Cell Lung Cancer (NSCLC). *Bioorganic Chemistry, 107*, 104612.

Kennard, R. W., and Stone, L. A. (1969). Computer aided design of experiments. *Technometrics, 11*(1), 137-148.

Khuri, F. (2016). Lung cancer and other pulmonary neoplasms. *Goldman-Cecil Medicine, 2*, 1303-1313.

Kim, T., Cho, S., Oh, H., Hur, J., Kim, H., Choi, Y.-H., . . . Kim, S.-H. (2020). Design of anticancer 2, 4-diaminopyrimidines as novel anoctamin 1 (ANO1) ion channel blockers. *Molecules, 25*(21), 5180.

Kohn, W., Becke, A. D., and Parr, R. G. (1996). Density functional theory of electronic structure. *The Journal of Physical Chemistry, 100*(31), 12974-12980.

Mahapatra, A., Prasad, T., and Sharma, T. (2021). Pyrimidine: A review on anticancer activity with key emphasis on SAR. *Future Journal of Pharmaceutical Sciences, 7*(1), 1-38.

Mills, N. (2006). ChemDraw Ultra 10.0 CambridgeSoft, 100 CambridgePark Drive, Cambridge, MA 02140. www. cambridgesoft. com. Commercial Price: 1910fordownload, 2150 for CD-ROM; Academic Price: 710fordownload, 800 for CD-ROM: ACS Publications.

Othman, I. M., Alamshany, Z. M., Tashkandi, N. Y., Gad-Elkareem, M. A., Anwar, M. M., and Nossier, E. S. (2021). New pyrimidine and pyrazole-based compounds as potential EGFR inhibitors: Synthesis, anticancer, antimicrobial

evaluation and computational studies. *Bioorganic Chemistry, 114*, 105078.

Passaro, A., Mok, T., Peters, S., Popat, S., Ahn, M.-J., and De Marinis, F. (2021). Recent advances on the role of EGFR tyrosine kinase inhibitors in the management of NSCLC with uncommon, non exon 20 insertions, EGFR mutations. *Journal of Thoracic Oncology, 16*(5), 764-773.

Patel, H., Pawara, R., Ansari, A., and Surana, S. (2017). Recent updates on third generation EGFR inhibitors and emergence of fourth generation EGFR inhibitors to combat C797S resistance. *European Journal of Medicinal Chemistry, 142*, 32-47.

Pawara, R., Ahmad, I., Surana, S., and Patel, H. (2021). Computational identification of 2, 4-disubstituted amino-pyrimidines as L858R/T790M-EGFR double mutant inhibitors using pharmacophore mapping, molecular docking, binding free energy calculation, DFT study and molecular dynamic simulation. *In Silico Pharmacology, 9*(1), 1-22.

Sato, J. D., Okamoto, T., Barnes, D., Hayashi, J., Serrero, G., and McKeehan, W. L. (2018). A tribute to Dr. Gordon Hisashi Sato (December 24, 1927–March 31, 2017). *In Vitro Cellular & Developmental Biology-Animal, 54*(3), 177-193.

Song, Z., Huang, S., Yu, H., Jiang, Y., Wang, C., Meng, Q., . . . Li, Y. (2017). Synthesis and biological evaluation of morpholine-substituted diphenylpyrimidine derivatives (Mor-DPPYs) as potent EGFR T790M inhibitors with improved activity toward the gefitinib-resistant non-small cell lung cancers (NSCLC). *European Journal of Medicinal Chemistry, 133*, 329-339.

Tièche, C. C., Peng, R.-W., Dorn, P., Froment, L., Schmid, R. A., and Marti, T. M. (2016). Prolonged pemetrexed pretreatment augments persistence of cisplatin-induced DNA damage and eliminates resistant lung cancer stem-like cells associated with EMT. *BMC cancer, 16*(1), 1-15.

Tropsha, A., and Golbraikh, A. (2010). Predictive quantitative structure-activity relationships modeling. *Handbook of chemoinformatics algorithms, 33*, 211.

Tropsha, A., Gramatica, P., and Gombar, V. K. (2003). The importance of being earnest: validation is the absolute essential for successful application and interpretation of QSPR models. *QSAR & Combinatorial Science, 22*(1), 69-77.

Tu, Y., Wang, C., Xu, S., Lan, Z., Li, W., Han, J., . . . Zhu, W. (2017). Design, synthesis, and docking studies of quinazoline analogues bearing aryl semicarbazone scaffolds as potent EGFR inhibitors. *Bioorganic & Medicinal Chemistry, 25*(12), 3148-3157.

Veerasamy, R., Rajak, H., Jain, A., Sivadasan, S., Varghese, C. P., and Agrawal, R. K. (2011). Validation of QSAR models-strategies and importance. *Int. J. Drug Des. Discov, 3*, 511-519.

Wang, D., Zhou, J., Fang, W., Huang, C., Chen, Z., Fan, M., . . . Luo, L. (2022). A multifunctional nanotheranostic agent potentiates erlotinib to EGFR wild-type non-small cell lung cancer. *Bioactive materials, 13*, 312-323.

Yap, C. W. (2011). PaDEL-descriptor: An open source software to calculate molecular descriptors and fingerprints. *Journal of computational chemistry, 32*(7), 1466-1474.