

2D QSAR approach to develop newer generation molecules active against ERBB2 receptor kinase as potential Anticancer Agent

Supriyo Saha^{1*}, Amartya Banerjee² and Ayan Rudra²

¹*Sardar Bhagwan Singh Institute of Biomedical Sciences and Research, Balawala, Dehradun.*

²*Guru Ram Das Post Graduate Institute of Management and Technology, Dehradun.*

Abstract

Quantitative Structure Activity Relationship analysis was performed using 159 molecules active against ERBB2 kinase inhibitory activity, where IC₅₀ was converted into pIC₅₀. The statistically suitable QSAR model was $PIC_{50} = -0.27351(\pm 1.85955) - 0.02792(\pm 0.08389) ALogP - 2.14626(\pm 1.14052) SpMin3_Bhv - 1.32984(\pm 0.13918) ntN + 0.21209(\pm 0.02129) ETA_Beta_ns$ (SEE :0.63346, r^2 :0.56219, r^2 adjusted :0.54627, F :35.31334 (DF :4, 110) which suggest that electron-richness of the molecule (ETA_Beta_ns) will create positive response and Ghose-Crippen LogKow value (ALogP), Smallest absolute eigenvalue of Burden modified matrix - n³ / weighted by relative van der Waals volumes (SpMin3_Bhv) and number of atom type E state (ntN) cause negative response in biological activity. Then the model was validated by Golbraikh and Tropsha acceptable criteria as Q^2 : 0.51805 Passed (Threshold value $Q^2 > 0.5$), r^2 :0.60327 Passed (Threshold value $r^2 > 0.6$), $|r^2 - r^2_{0}| < 0.18983$ Passed (Threshold value $|r^2 - r^2_{0}| < 0.3$). The External Validation Parameters (Without Scaling): r^2 :0.60327 and external Validation parameters (After Scaling): rm^2 :0.5955. The model was cross validated using Leave-One-Out (LOO) process with Q^2 :0.51805, PRESS: 48.59098, SDEP :0.65002. So the QSAR model was suitable for future development.

Keywords: ERBB2 receptor, PADEL, Stepwise regression, FA-MLR, Golbraikh and Tropsha acceptable model, LOO Method

1. Introduction

The ErbB/HER protein-tyrosine kinases, which include the epidermal growth factor receptor (EGFR), consist of a growth-factor-binding ectodomain, a single transmembrane segment, an intracellular protein-tyrosine kinase catalytic domain, and a tyrosine-containing cytoplasmic tail[1]. In humans, the family includes HER1 (EGFR, ErbB1), HER2 (Neu, ErbB2), HER3 (ErbB3), and HER4 (ErbB4). ErbB2/HER2, the second member of the group, has several unique properties[2]. First, ErbB2 lacks a known direct ligand; in order to function, it must work as a co-receptor, or heterodimerization partner, for other ErbB receptors that possess stimulatory ligands[3][4][5]. Second, unlike other ErbB receptors, ErbB2 overexpression can cause malignant transformation without the expression of a growth factor. This observation suggests that ErbB2 has a high level of constitutive (ligand-independent) activity, and ErbB2 expression above a specific threshold can drive tumor growth[6]. ErbB2 is a cell surface receptor and when bind with the growth factors in activated (homo-/hetero- dimerized) form it promotes gene regulation leads to cellular division and growth. ErbB2 probably also plays a role in cell adhesion, cell specialization, and cell movement. Insufficient ErbB signaling in humans is associated with the development of neurodegenerative diseases, such as multiple sclerosis and Alzheimer's Disease[7], while excessive ErbB signaling is associated with the development of a wide variety of types of solid tumor.

Over expression of ErbB2 gene causes production of excessive amount of ErbB2 protein (ErbB2 receptor) that triggers continuous cell proliferation and may lead to tumor formation. Over expression of this ErbB2 gene is found in 25% of breast cancer patient.

*** Correspondence Info**

Supriyo Saha

Department of Pharmaceutical Sciences,

Sardar Bhagwan Singh Post Graduate Institute of Biomedical Sciences and Research, Balawala, Dehradun – 248161, Uttarakhand, India.

E-mail: supriyo9@gmail.com

2. Materials and Methods

To construct a QSAR model, a set of theoretical and constructive descriptors were calculated by use of PaDEL-descriptor: open source software, ToMoCoMD and QSAR Model was constructed by use of MLR Plus Validation Tool. By using PADEL and ToMoCoMD we were calculated 1875 descriptors include Ghose-Crippen Log Ko/w, Ghose-Crippen molar refractivity, Sum of the atomic polarizabilities (including implicit hydrogens), Wildman- Crippen LogP and MR, Wildman-Crippen MR, Eccentric Connectivity Index: topological descriptor combining distance and adjacency information, H Bond Acceptor Count Number of hydrogen bond acceptors, McGowan characteristic volume, Wiener Polarity Number. All the explanations of relevant descriptors were enlisted in Table 1. A descriptor represents a quantitative property depends on the molecular structure. Theoretical descriptors are advantageous due to its free from uncertainty of experimental measurement and can be calculated for compounds before synthesis. Theoretical descriptors employed in this QSAR study to model as an inhibitor of ERBB2 receptor as potential anticancer agent.

2.1. Dataset and Descriptor Calculation

Dataset of 159 ERBB2 receptor antagonist was downloaded from <http://crdd.osdd.net>. All the Molecules SMILES are transferred into .mol format by ACDLABS and structures were optimized. Calculate the 2D and 3D descriptor using PADEL descriptor[8] and ToMoCoMD[9] software. Table 1 showed the detail dataset along with chemical structure, IC₅₀ value and pIC₅₀ value and Table 2 results some useful descriptor with explanation.

2.2. Descriptor Pretreatment

Cutoff the inter correlated descriptor selected using V-WSP as variance cut off 0.0001 and correlation coefficient value 0.99.

2.3. Dataset Division

Total dataset was divided into Train and Test using Kennard Stone method as 115 molecules in Training set and 44 molecules in Test set.

2.4. Suitable Descriptor Selection

Suitable descriptor selection using Stepwise MLR as F values 3.9 to 4.0. Then best subset was selected using 4 descriptor combination and r² cut off value 0.6.

2.5. The chemometric tool

For the development of QSAR equation two methods were implemented; (1) Stepwise regression (2) multiple linear regressions with factor analysis as pre processing factor analysis for variable selection (FA-MLR).

2.6. Stepwise regression

Multi step linear equation, a multistep equation was built by step by step. The basic procedure involved: (i) identifying an initial model (ii) repeating the previous step by altering descriptor or variable combination to achieve better f and r² value. (iii) calibrate the equation by justify the values in between observed and predicted values (Roy *et al.*, 2014). The stepwise MLR was performed using statistical software SPSS and it was judged by parameters as explained variance (r²a), correlation coefficient (r), standard error of estimate (s) and variance ratio (F) at a specified degree of freedom (DF). All accepted MLR equation had regression level significant at 95 and 99% levels. the generated qsar equation was validated by leave one out or LOO method using Minitab software and different parameters like cross validation r² (q²), standard deviation based on press (S_{PRESS}) and standard deviation of error of prediction (SDEP)[10].

2.7. FA-MLR

In this case a final statistical tool was used to develop a QSAR relation, factor analysis as a data pre processing step to identify the important factor to identify the important variables contributing the response variable by avoiding co linear value. The data matrix is first standardized and correlation matrix and subsequently reduced correlation matrix. An eigen value problem is then solved and the factor pattern can be obtained from the corresponding eigen vectors. The main objectives are to display multidimensional data in space of lower dimensionality with minimum loss of information (explaining > 95% of variance of data matrix) and to extract the basic features behind the data with ultimate goal of interpretation[11].

2.8. QSAR Equation Development

MLR Plus valid software was used to developed QSAR equation, where IC₅₀ was converted pIC₅₀ value.

2.9. QSAR Equation Validation

Golbraikh and Tropsha acceptable model criteria's to validate a QSAR Equation

1. Q^2 value is Passed (Threshold value $Q^2 > 0.5$).
2. r^2 value is Passed (Threshold value $r^2 > 0.6$).
3. $|r^2 - r'^2|$ value is Passed (Threshold value $|r^2 - r'^2| < 0.3$).

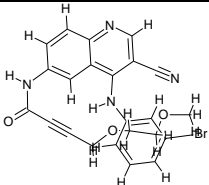
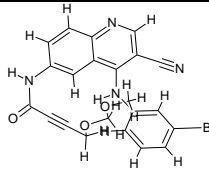
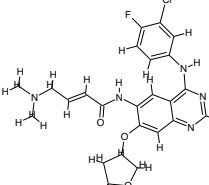
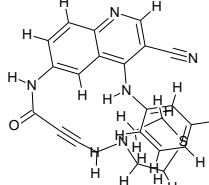
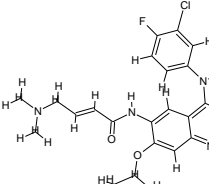
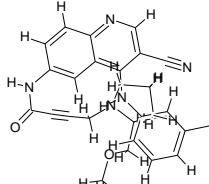
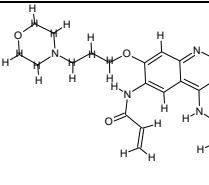
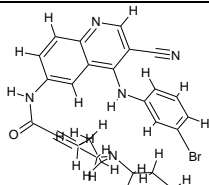
2.10. QSAR Equation Validation

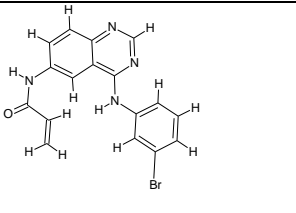
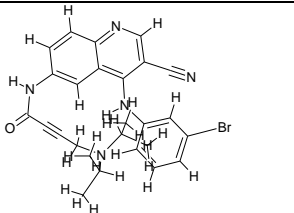
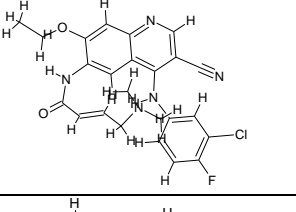
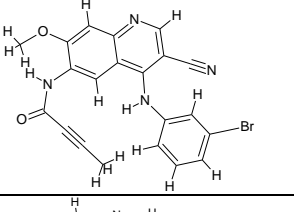
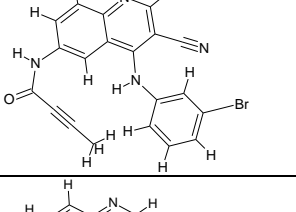
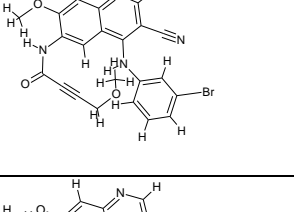
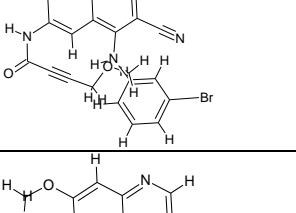
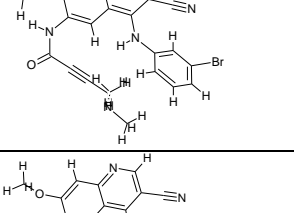
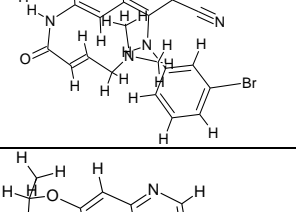
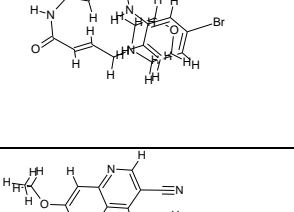
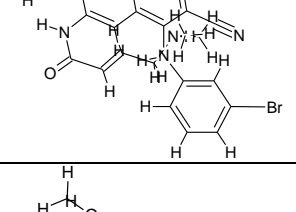
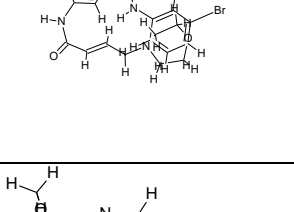
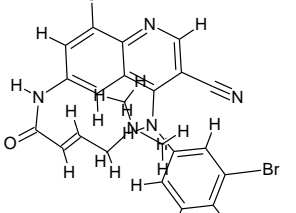
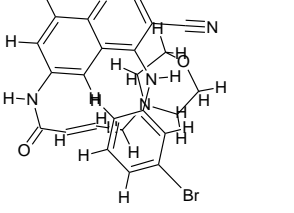
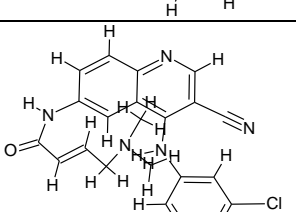
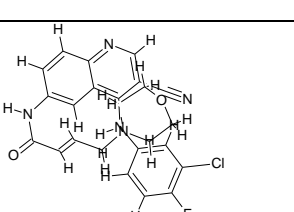
The model was cross validated using Leave-One-Out (LOO) process.

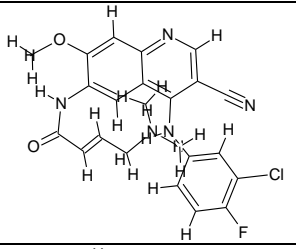
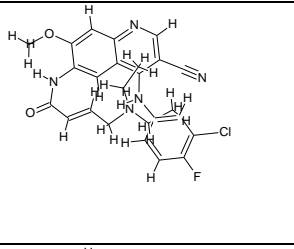
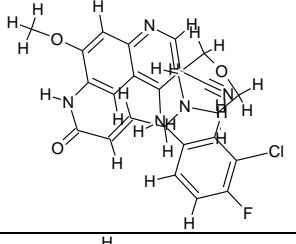
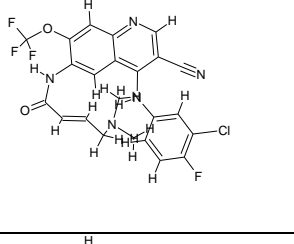
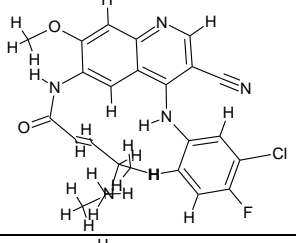
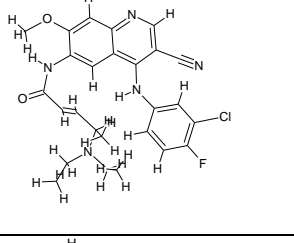
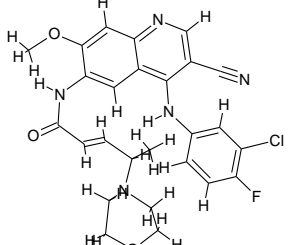
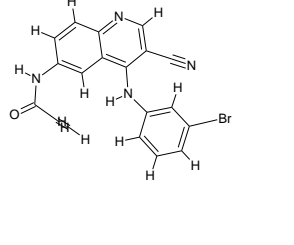
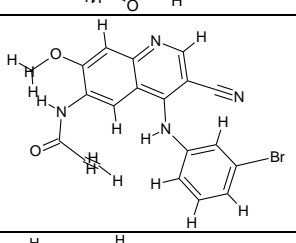
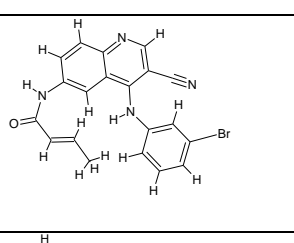
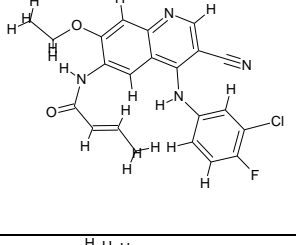
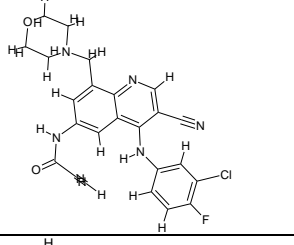
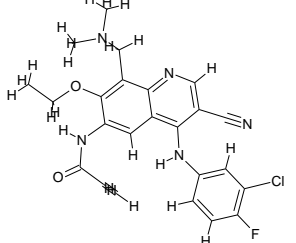
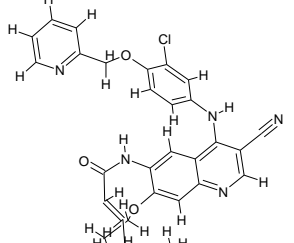
3. Results and Discussion

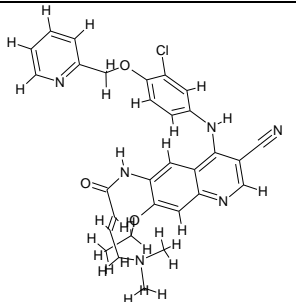
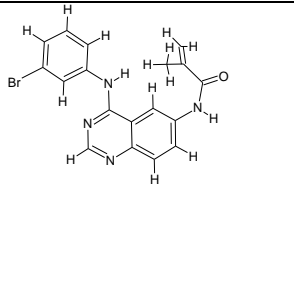
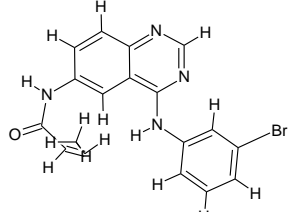
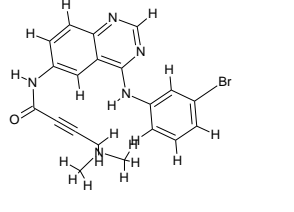
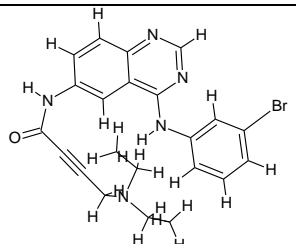
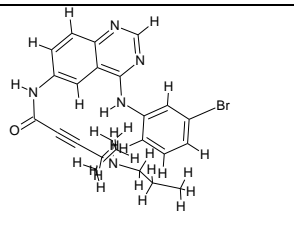
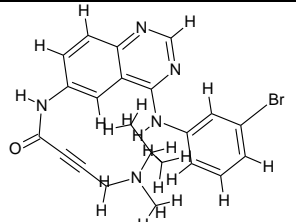
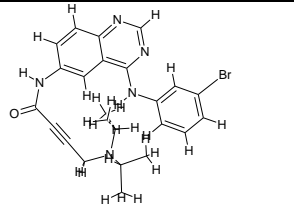
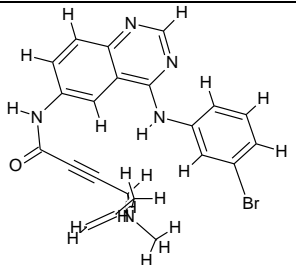
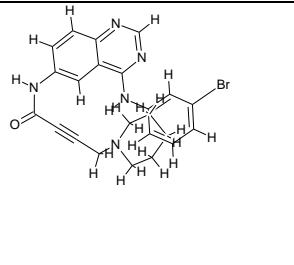
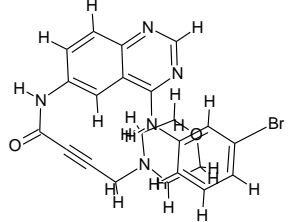
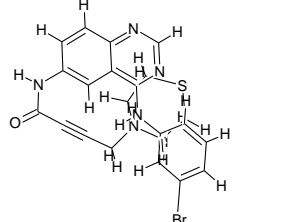
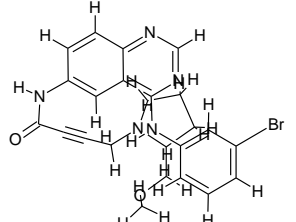
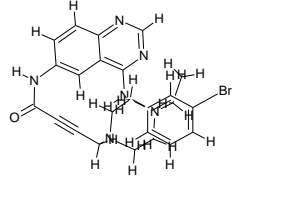
The statistically suitable QSAR model was $pIC_{50} = -0.27351(+/-1.85955) - 0.02792(+/-0.08389) ALogP - 2.14626(+/-1.14052) SpMin3_Bhv - 1.32984(+/-0.13918) ntN + 0.21209(+/-0.02129) ETA_Beta_ns$. Statistical information: (SEE: 0.63346, $r^2:0.56219$, r^2 adjusted: 0.54627, F: 35.31334 (DF: 4, 110)). The model suggest that electron-richness of the molecule (ETA_Beta_ns) will create positive response and Ghose-Crippen LogKow value (ALogP), Smallest absolute eigen value of Burden modified matrix - n 3 / weighted by relative van der Waals volumes (SpMin3_Bhv) and number of atom type E state (ntN) cause negative response in biological activity. Model was validated by Golbraikh and Tropsha acceptable criteria as $Q^2: 0.51805$ Passed (Threshold value $Q^2 > 0.5$), $r^2:0.60327$ Passed (Threshold value $r^2 > 0.6$), $|r^2 - r'^2|:0.18983$ Passed (Threshold value $|r^2 - r'^2| < 0.3$). External Validation Parameters (Without Scaling): $r^2:0.60327$ and external Validation parameters (After Scaling): $rm^2:0.5955$. The model was cross validated using Leave-One-Out (LOO) process with $Q^2:0.51805$, PRESS: 48.59098, SDEP :0.65002. The graphical representation in between $Y_{observed}$ and $Y_{predicted}$ showed a good correlation in Figure 1 which represent that most of the data were overlapped.

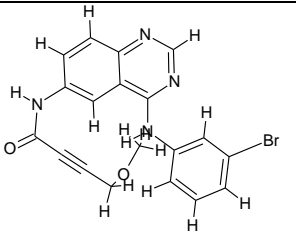
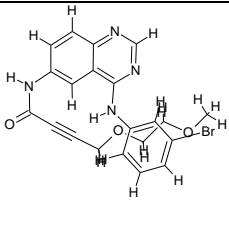
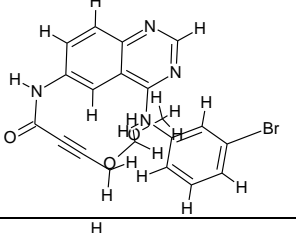
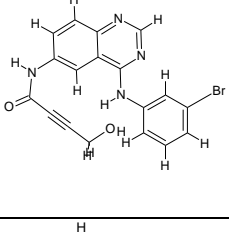
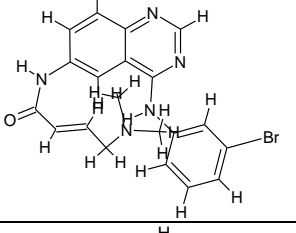
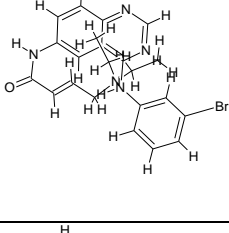
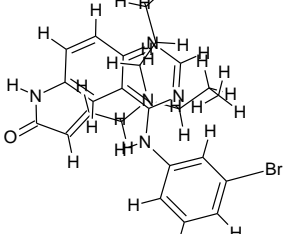
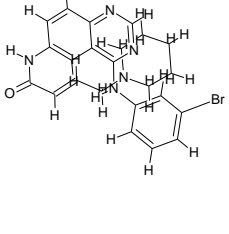
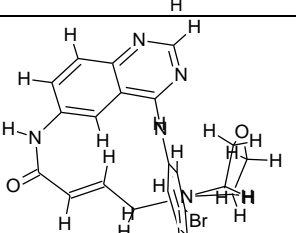
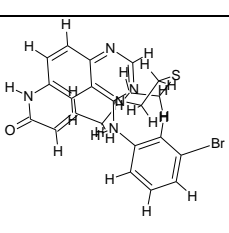
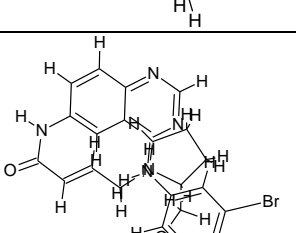
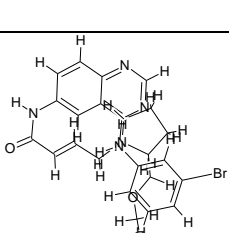
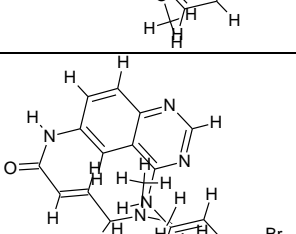
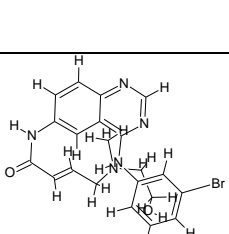
Table 1: Detail Dataset of 159 molecules with IC_{50} and pIC_{50} values

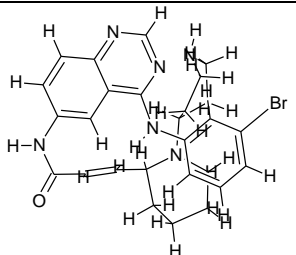
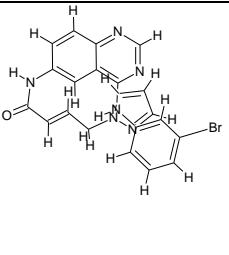
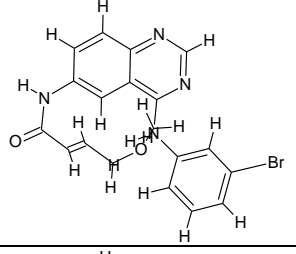
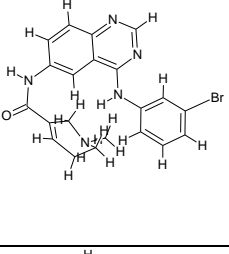
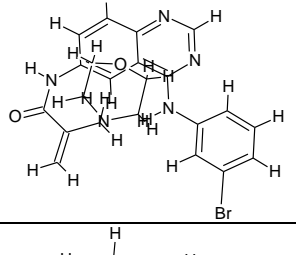
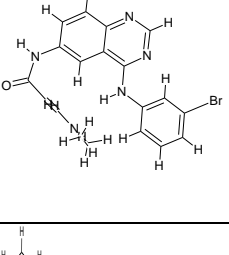
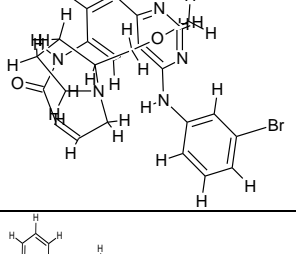
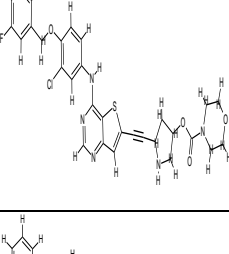
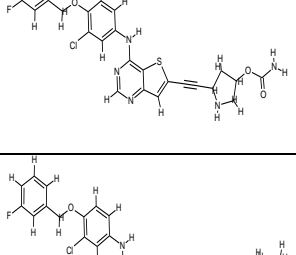
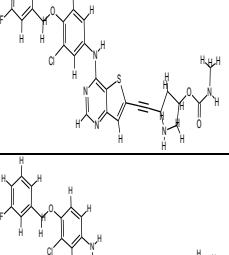
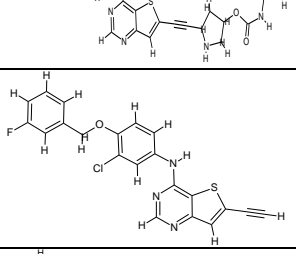
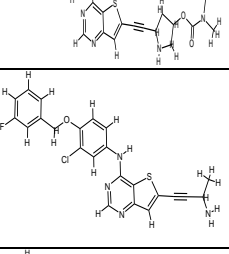
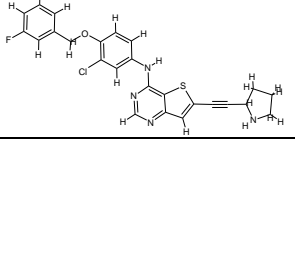
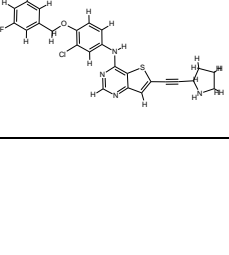
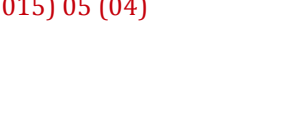

SN	Structure	IC_{50}	pIC_{50}	SN	Structure	IC_{50}	pIC_{50}
1		24.9nM	1.603	9		0.38mM	0.420
2		14nM	1.853	10		1.04mM	-0.017
03		127nM	0.896	11		1.22mM	-0.0863
4		0.114mM	0.943	12		0.37mM	0.432

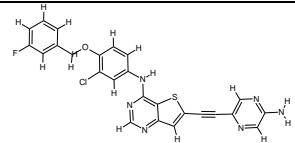
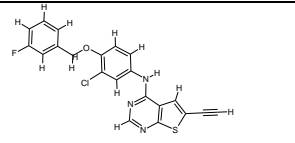
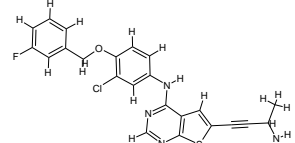
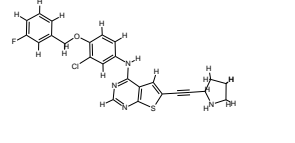
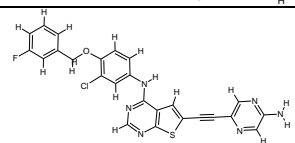
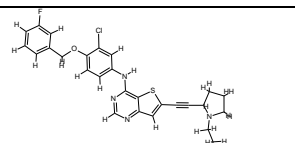
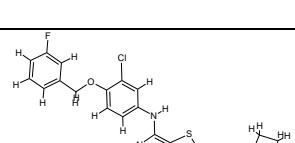
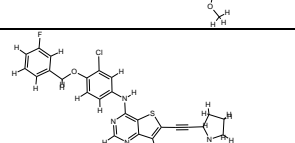
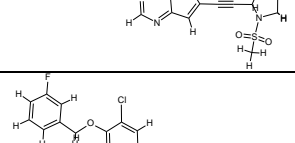
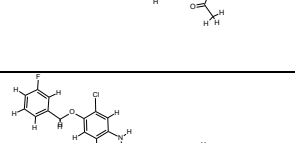
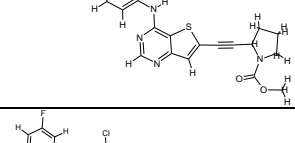
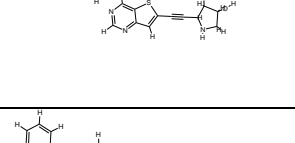
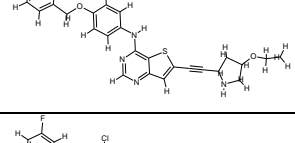
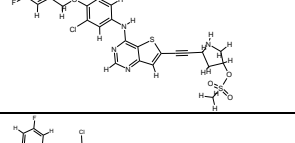
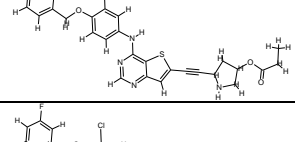
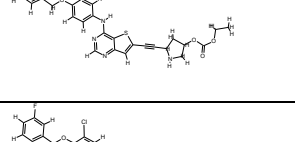
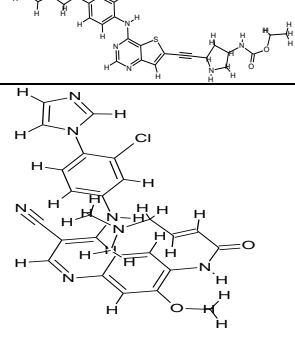
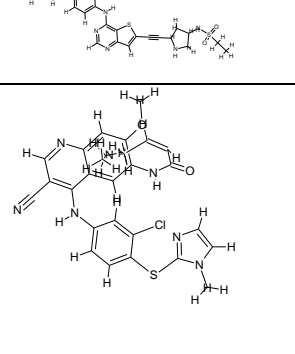
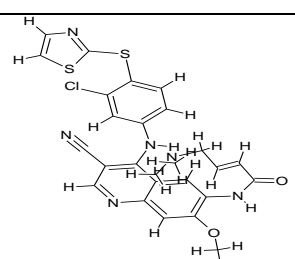
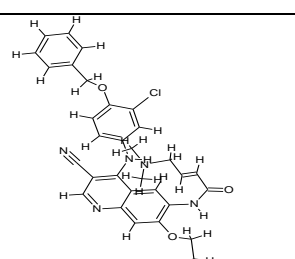
5		1.23mM	-0.089	13		0.32mM	0.495
6		3.87mM	-0.587	14		2.12mM	-0.326
7		0.30mM	0.522	15		0.38mM	0.420
8		0.45m	0.346	16		0.18mM	0.744
17		4.16mM	-0.619	18		5.66mM	-0.753
19		34.13mM	-1.533	20		2.97mM	-0.472
21		33.95mM	-1.530	22		9.35mM	-0.971
23		23.59mM	-1.372	24		1.95mM	-0.290

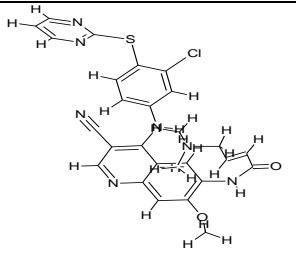
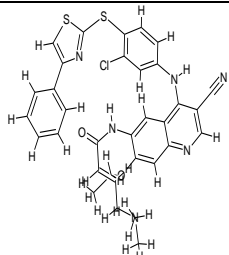
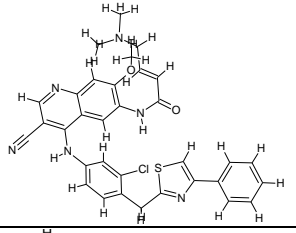
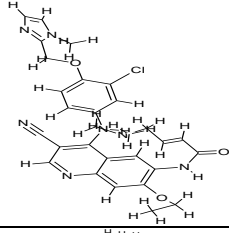
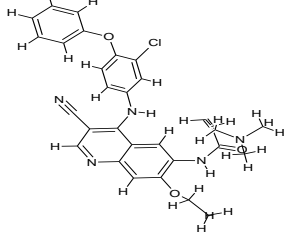
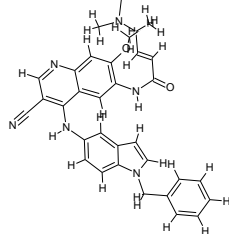
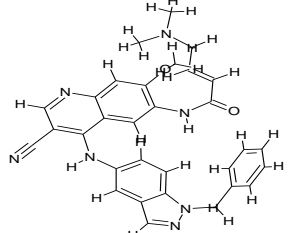
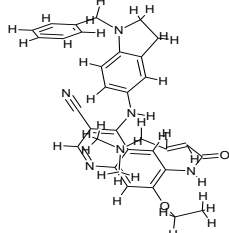
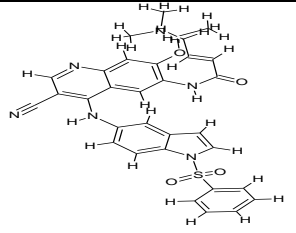
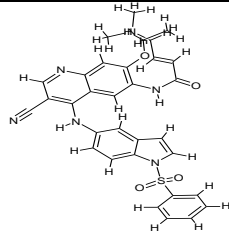
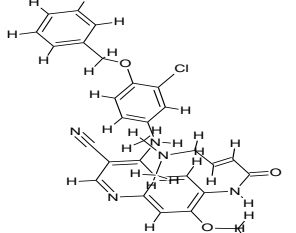
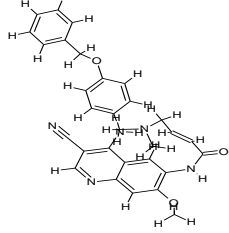
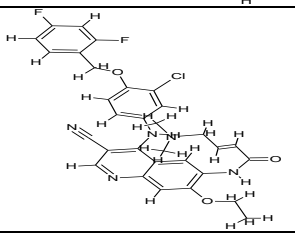
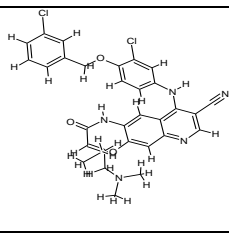
25		8.74mM	-0.942	26		4.49mM	-0.652
27		7.80mM	-0.892	28		4.13mM	-0.615
29		11.82mM	-1.072	30		9.58mM	-0.981
31		33.95mM	-1.531	32		4.68mM	-0.670
33		4.50mM	-0.653	34		21.51mM	-1.332
35		4.39mM	-0.642	36		1.21mM	-0.0827
37		0.059mM	1.229	38		15.24mM	-1.183

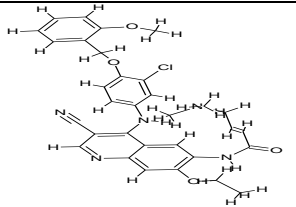
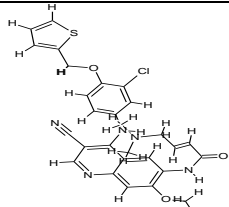
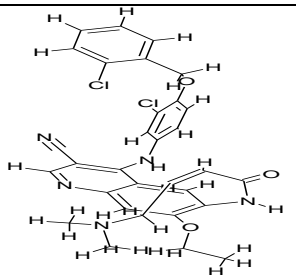
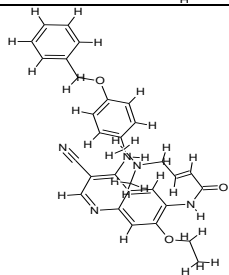
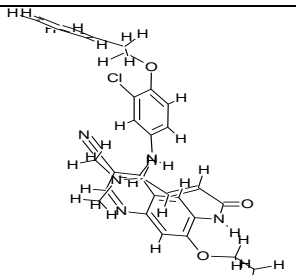
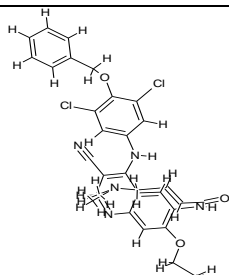
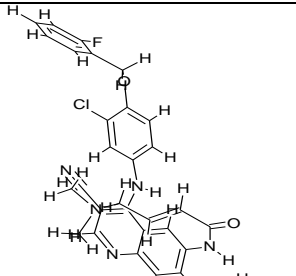
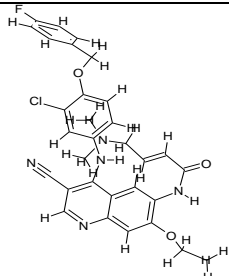
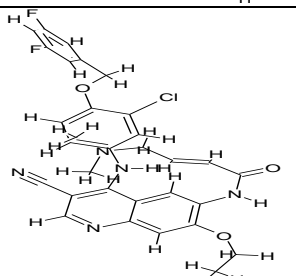
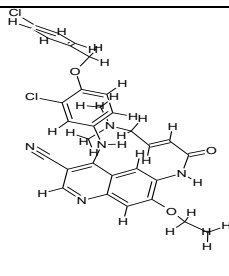
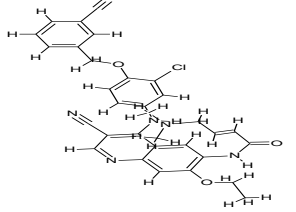
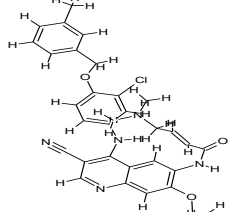
39		21.65mM	-1.335	40		0.014mM	1.853
41		0.038mM	1.420	42		0.177mM	0.752
43		0.113mM	0.946	44		0.162mM	0.790
45		0.085mM	1.071	46		0.073mM	1.136
47		0.053mM	1.276	48		0.124mM	0.906
49		0.117mM	0.932	50		0.183mM	0.737
51		0.033mM	1.481	52		0.116mM	0.935

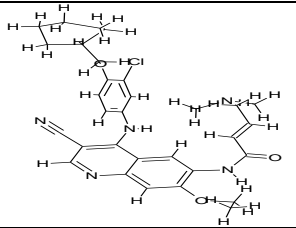
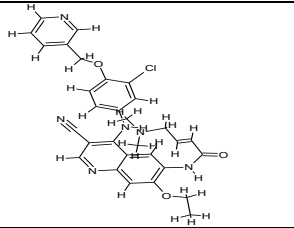
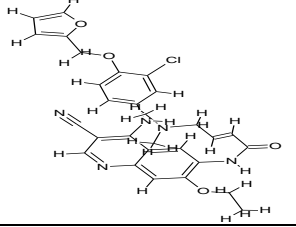
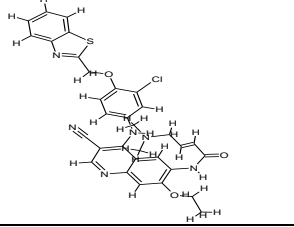
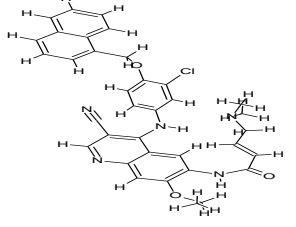
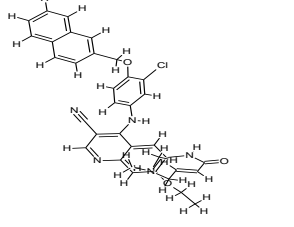
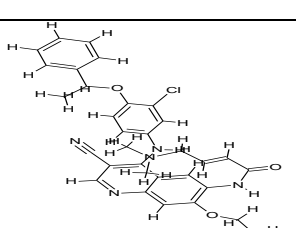
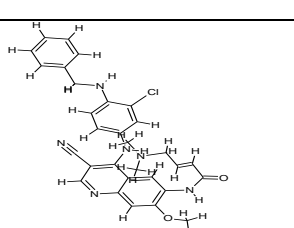
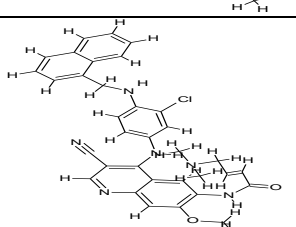
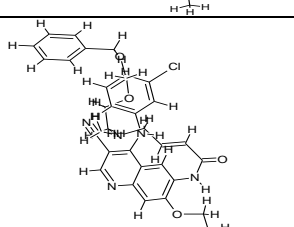
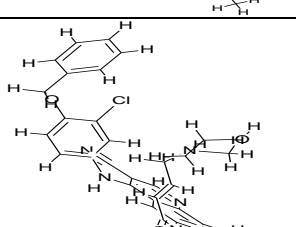
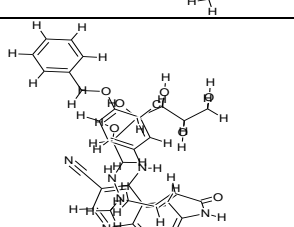
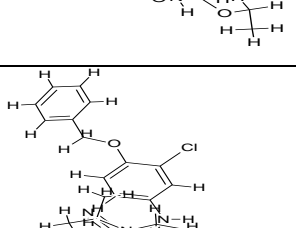
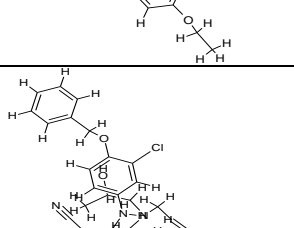
53		0.122mM	0.913	54		0.016mM	1.796
55		0.301mM	0.521	56		1.164mM	-0.066
57		1.428mM	-0.154	58		1.559mM	-0.193
59		4.182mM	-0.621	60		2.808mM	-0.448
61		1.912mM	-0.281	62		0.733mM	0.134
63		0.634mM	0.198	64		0.414mM	0.383
65		1.592mM	-0.201	66		8.832mM	-0.946

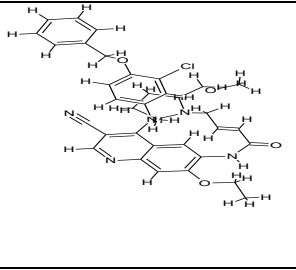
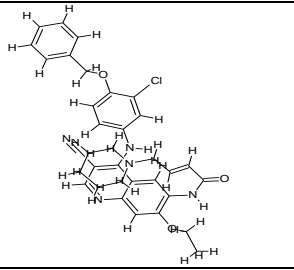
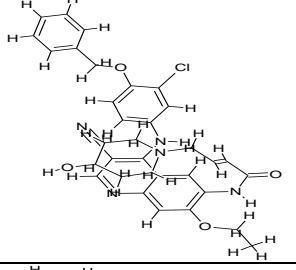
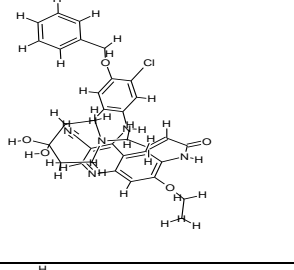
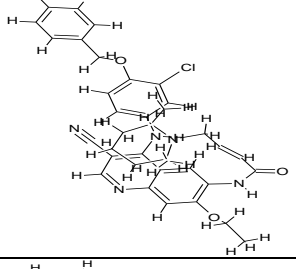
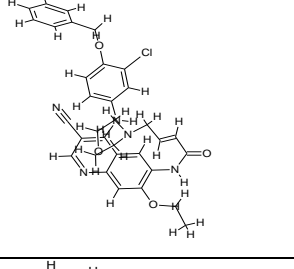
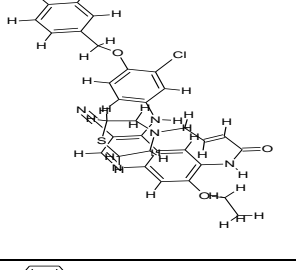
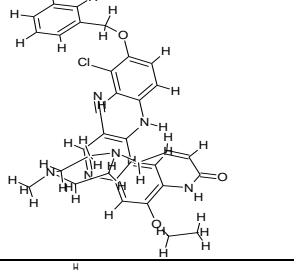
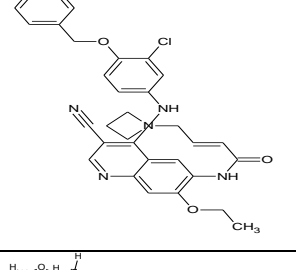
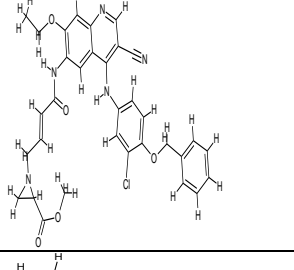
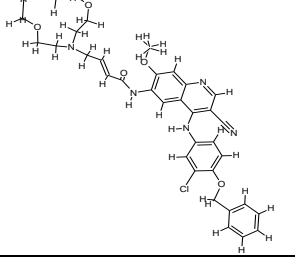
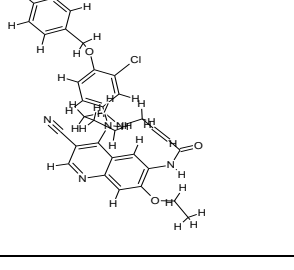
67		5.042mM	-0.702	68		0.458mM	0.339
69		2.921mM	-0.465	70		0.084mM	1.075
71		0.016mM	1.796	72		0.020mM	1.698
73		0.043mM	1.366	74		0.068mM	1.167
75		0.012mM	1.921	76		0.027mM	1.568
77		0.013mM	1.886	78		0.054mM	1.267
79		0.010mM	2.0	80		0.019mM	1.721
81		0.044mM	1.356	82		0.054mM	1.267

83		0.052mM	1.283	84		0.11mM	0.958
85		0.088mM	1.055	86		0.038mM	1.420
87		0.120mM	0.921	88		0.011mM	1.958
89		0.098mM	1.008	90		0.040mM	1.398
91		0.090mM	1.045	92		0.135mM	0.869
93		0.078mM	1.107	94		0.087mM	1.060
95		0.80mM	0.0969	96		0.155mM	0.809
97		0.083mM	1.081	98		0.076mM	1.119
99		0.348mM	0.458	100		0.013mM	1.886
101		0.074mM	1.131	102		0.625mM	0.204

103		0.184mM	0.735	104		0.015mM	1.824
105		0.075mM	1.124	106		0.230mM	0.638
107		0.025mM	1.602	108		0.033mM	1.481
109		0.039mM	1.408	110		0.363mM	0.440
111		0.081mM	1.092	112		0.008mM	2.096
113		0.044mM	1.356	114		0.131mM	0.883
115		0.010mM	2.0	116		0.498mM	0.303

117		0.281mM	0.551	118		0.406mM	0.391
119		0.732mM	0.135	120		0.078mM	1.108
121		0.388mM	0.411	122		0.024mM	1.619
123		0.161mM	0.793	124		0.132mM	0.879
125		0.028mM	1.552	126		0.269mM	0.570
127		0.062mM	1.207	128		0.064mM	1.194

129		0.02mM	1.698	130		0.018mM	1.744
131		0.076mM	1.119	132		0.103mM	0.987
133		0.083mM	1.081	134		0.183mM	0.737
135		0.027mM	1.568	136		0.966mM	0.015
137		0.028mM	1.553	138		0.104mM	0.983
139		0.171mM	0.767	140		0.025mM	1.602
141		0.169mM	0.772	142		0.058mM	1.236

143		0.412mM	0.385	144		0.059mM	1.229
145		0.029mM	1.537	146		0.54mM	0.267
147		0.04mM	1.397	148		0.196mM	0.708
149		0.603mM	0.219	150		0.059mM	1.229
151		0.069mM	1.161	152		0.227mM	0.643
153		0.061mM	1.214	154		0.246mM	0.609

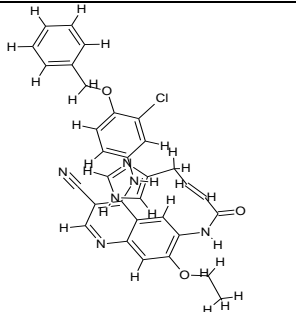
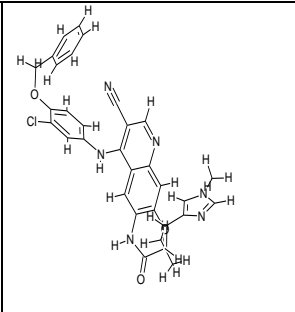
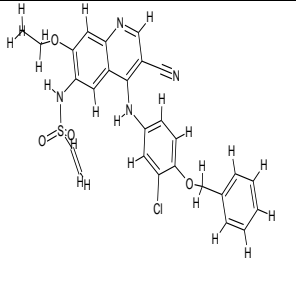
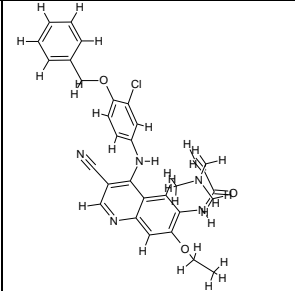
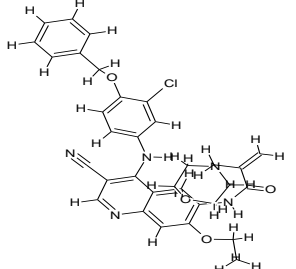
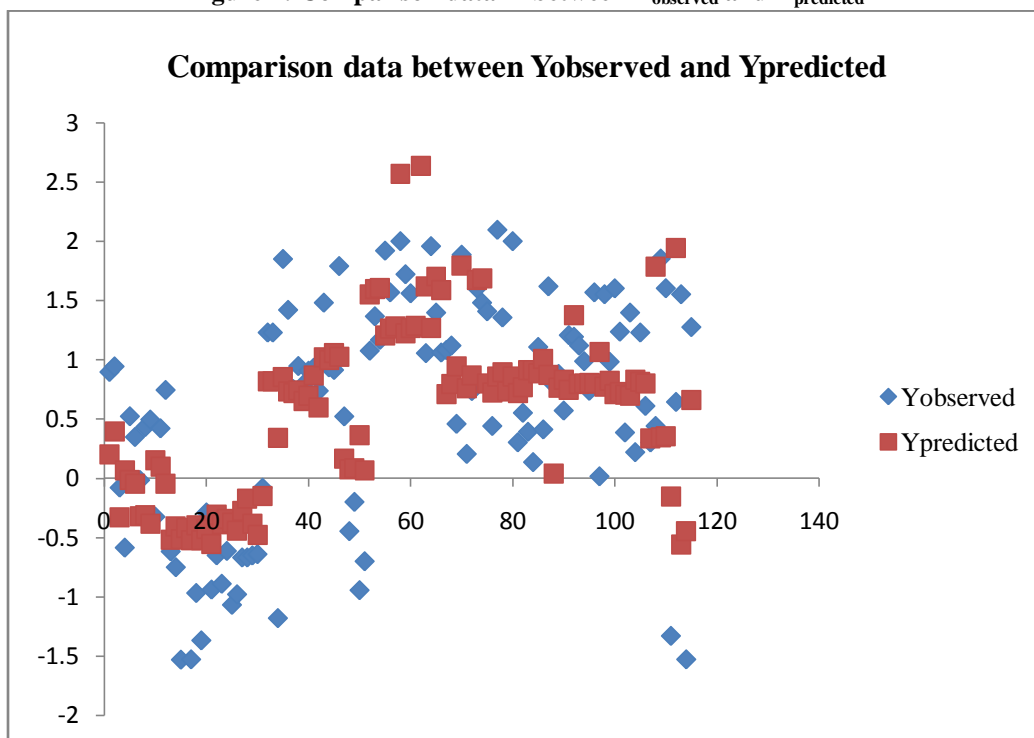
155		0.089mM	1.051	156		0.5mM	0.301
157		0.0363mM	1.440	158		1.1	-0.0415
159		3.9	-0.591				

Table 2: List of relevant descriptor with explanation

SN	Abbreviation descriptors	Explanation of descriptors
1.	AlogP	Ghose-Crippen LogKow
2.	SpMin3_Bhv	Burden modified eigenvalues
3.	ntN	Total number of Nitrogen Atoms
4.	ETA_Beta_ns	A measure of electron-richness of the molecule
5.	Crippen MR	Crippen's molar refractivity
6.	McGowan Volume	McGowan characteristic volume
7.	VABC	Van der Waals volume calculated
8.	nRing	No of Ring
9.	nRotb	No of Rotatable Bonds
10.	Phia Kappa	flexibility Index
11.	Bac	Balaban Centric Index
12.	AlogP	Ghose-Crippen LogKow
13.	Crippen LogP	Crippen's LogP
14.	XLogP	XLogP
15.	AMR	Molar refractivity
16.	TopoPSA	Topological Polar Surface Area
17.	Wpol	Weiner polarity number
18.	MW	Molecular Weight
19.	ETA	Electro Topochemical Descriptor

Figure 1: Comparison data in between Y_{observed} and $Y_{\text{predicted}}$ 

3. Conclusion

This total work is an attempt to deliver a platform for the new research to develop a better new generation chemical entity active against ErbB2 receptor kinase as potential Anticancer Agent. ErbB2 is a remarkably fantastic marker discovered in the field of breast cancer and our future aim is to give birth of some novel molecule to treat breast cancer.

References

- [1] Robert R. The ErbB/HER receptor protein-tyrosine kinases and cancer, *Biochemical and Biophysical Research Communications*. 2004; 319: 1–11.
- [2] Burgess. A.W., Cho. H.S., An open-and-shut case? Recent insights into the activation of EGF/ErbB receptors. *Mol. Cell*. 2003; 12: 541–552.
- [3] Yarden. Y., Slivkowski. M.X., Untangling the erbB signaling network. *Nat. Rev. (Mol. Cell. Biol.)*. 2001; 2: 127–137.
- [4] Porta. D. G., Beerli. R.R., Daly. J.M., Hynes. N.E., ErbB-2, the preferred heterodimerization partner of all ErbB receptors, is a mediator of lateral signaling. *EMBO J*. 1997; 16: 1647–1655.
- [5] Citri. A., Skaria. K.B., Yarden. Y., The deaf and the dumb: the biology of ErbB-2 and ErbB-3, *Exp. Cell Res*. 2003; 284: 54–65.
- [6] Bubli. E.M., Yarden. Y., The EGF receptor family: spearheading a merger of signaling and therapeutics. *Curr. Opin. Cell Biol*. 2007; 19 (2): 124–134.
- [7] Cho. H.S., Leahy. D.J., Structure of the extracellular region of HER3 reveals an interdomain tether. *Science*. 2002; 297 (5585): 1330–1333.
- [8] Yap. C.W., PaDEL-descriptor: open source software to calculate molecular descriptors and fingerprints. *J. Comput Chem*. 2011; 32(7): 1466–74.
- [9] Garcia.J. Marrero-Ponce. C. R., Acevedo-Martínez. Y., Barigye. L., Valdes-Martini. S. J., Contreras-Torres. J. R., QuBiLS-MIDAS: A Parallel Free-Software for Molecular Descriptors Computation based on Multi-Linear Algebraic Maps. *J. Comput. Chem*. 2014; 35: 1395–1409.
- [10] Roy. K., Kar. S. The rm2 metrics and regression through origin approach: reliable and useful validation tools for predictive QSAR models (Commentary on 'Is regression through origin useful in external validation of QSAR models?'). *Eur. J Pharm Sci*. 2014; 62: 111–114.
- [11] Roy K, Mitra I. On various metrics used for validation of predictive QSAR models with applications in virtual screening and focused library design. *Comb Chem High Throughput Screen*. 2010; 14(6): 450–74.