AMERICAN JOURNAL OF PHARMTECH RESEARCH

Journal home page: http://www.ajptr.com/

QSAR STUDY OF EGFR (EPIDERMAL GROWTH FACTOR RECEPTOR) INHIBITORS-A RATIONAL APPROACH IN DEVELOPMENT OF ANTICANCER DRUGS.

Sharin Varasani[1*], Aravind Badiger[2], Gopkumar P[1].

1. Drug Discovery Lab, Shree Dhanvantary Pharmacy College, Kim, Surat, Gujarat
2. Shree Dhanvantary Pharmaceutical Analysis and Research Center, Kim, Surat, Gujarat

ABSTRACT

Epidermal Growth Factor Receptor (EGFR) is known to play vital role in many cellular signalling pathways and hence is considered as a potent target for cancer. Inhibition of this enzyme has been reported to be beneficial by various workers. QSAR study of Quinazolines as EGFR was performed using accelrys discovery studio client (DSV-Version 3.0) as the modelling tool. A total of 67 selected molecules were considered for the development of QSAR model. Partial least square model of the data generated exhibited a very good linear relation between the training set of compounds with that of the reported activity as well as the test set of compounds with the predicted activity. The 4 statistical analysis performed revealed following observations; Training data set r^2= 0.701, q^2 (Cross validated r^2) = 0.616 validated by internal validation with correlation of coefficient (r2) of 0.848 and cross validated r2 (q2) of 0.573 and external set of compounds with a predictive correlation of coefficient of 0.900. A total of 9 descripters pruned on the study explained the structural correlation of quinazolines with EGFR. The model developed can be used to predict bioefficacy of unknown molecules 4-[1,3-benzothiazol-2-yl]-N-[(1E)-(4-nitrophenyl)methylene]aniline as EGFR inhibitors. Further a hypothetical model to describe the interaction between the predicted molecules with EGFR is proposed and this hypothetical model explains the possibility of Met769 and Gln767 as the possible binding sites. The activity is observed in the preliminary cytotoxic activity (MTT assay). The study calls for the development of the molecules predicted as bio efficacious in this model and a quantitative inhibitory analysis of EGFR.

Key word: EGFR, QSAR, r^2, q^2

*Corresponding Author Email: sharinvarasani@yahoo.com

Received 25 July 2011, Accepted 29 July 2011

Please cite this article in press as: Varasani S. *et al*., QSAR Study of EGFR(Epidermal Growth Factor Receptor)Inhibitors-A Rational Approach in Development of Anticancer Drugs**.** American Journal of PharmTech Research 2011.

INTRODUCTION

Many of the tyrosine kinase enzymes are involved in cellular signaling pathways and regulate key cell functions such as proliferation, differentiation and anti-apoptotic signaling. Unregulated activation of these enzymes, through mechanisms such as point mutations or over expression, can lead to a large percentage of clinical cancers[1,2]. The importance of tyrosine kinase enzymes in health and disease is further underscored by the existence of aberrations in tyrosine kinase enzymes signaling occurring in inflammatory diseases and diabetes. Inhibitors of tyrosine kinase as a new kind of effective anticancer drug are important mediators of cellular signal transduction that affects growth factors and oncogenes on cell proliferation[3,4]. The development of tyrosine kinase inhibitors has therefore become an active area of research in pharmaceutical science. Epidermal growth factor receptor (EGFR) which plays a vital role as a regulator of cell growth is one of the intensely studied tyrosine kinase targets of inhibitors. EGFR is over expressed in numerous tumors, including those derived from brain, lung, bladder, colon, breast, head and neck. EGFR hyper activation has also been implicated in other diseases including polycystic kidney disease, psoriasis and asthma[5-7]. Since the hyper activation of EGFR has been associated with these diseases, inhibitor of EGFR has potential therapeutic value and it has been extensively studied in the pharmaceutical industry.

QSAR is a technique that quantifies the relationship between structure of a drug molecule and its biological activity as mathematical equation. Thus a QSAR model correlates the biological activity with molecular descriptors. A QSAR attempts to find consistent relationships between the variations in the values of molecular properties[8] and the biological activity for a series of compounds The numerical values resulting from the operation of a given descriptor on 2D molecular structure are normally used in QSAR as independent variables to correlate and predict various experimental physical properties (QSPR) or biological activities (QSAR)[9]. For different properties different descriptors will be used in QSAR equation.

Derivatives of quinazoline have been studied for their structure activity relations and EGFR inhibitory activity. Quantitative Structure-Activity Relationship (QSAR) study is basically concerned with correlation of structure and property/activity. Several molecular descriptors are used to quantify the structural features of lead molecule. The purpose of using QSAR-Descriptors is to calculate the properties of molecules that serve as numerical descriptions or characterizations of molecules in other calculations such as diversity analysis or combinatorial library design. In view of the above facts and to quantify the structural features required for the

series of quinazoline based inhibitors to inhibit EGFR, we report QSAR study of these inhibitors with the objective to develop a significant model to predict inhibitory activity of EGFR inhibitors.

MATERIALS AND METHODS

All computational work was performed on Pentium IV workstation-using Discovery Studio Client (DSV v3.0), developed by Acclerys software Inc., Japan. A total of 67 compounds were selected for the present study. All the compounds was drawn using Molecule Window and Sketching tool and then subjected to conformational analysis and energy minimization using prepare ligand protocol. Since each molecules possess more than one conformation, From this list the conformer with lowest absolute energy was selected, the lowest energy conformer in the conformation database of the particular molecule is considered the most stable conformation and the best conformation for the QSAR study for that molecule, hence a database is created from the lowest energy conformer of each molecule. Lowest energy conformer was transferred to molecule window and molecular descriptors were calculated. DSV calculates 560 subdescriptors from three major classes of descriptors; – 1D descriptor 2D descriptors, 3D descriptors.

67 Quinazolines derivatives were divided into two sets. 60 compounds were taken in for the training set (Table 1) and 7 compounds were taken in for the test set (Table 2). IC_{50} values for EGFR inhibition were transformed into –log (IC_{50}*10-6) *i.e.* pIC_{50}. The QSAR model was developed using pIC_{50} as activity field and partial least square (PLS) method for performing regression analysis. PLS regression method was used to analyze the internal interaction between the data matrices including the independent variable matrix of molecular descriptors of DSV and the dependent variable matrix of pIC_{50}. The PLS algorithm transforms the original variables into PLS components from dataset, such that each PLS component is correlated with the dependent data to a maximal possible extent. Statistical parameters like r^2 (Coefficient of correlation) and q^2 (Cross-Validated coefficient of correlation) were obtained after fitting the data by PLS. The model was cross validated to get prediction values, residual values. Cross validated cross validated r2 (q2) were also obtained. Using the prediction fit performed in training set model, predicted pIC_{50} values of test set were evaluated. Out of 560 descriptors, 9 subdescriptors, pruned from 4 major descriptor sets were selected and are listed in Table 3. The model developed was optimized using LOO (Leave One Out method).Use of different set of descriptors and subdescriptors are particularly requires the model to be validated by external

predictive power i.e. external prediction is address the latent variables in case of PLS analysis. Hence, a set of 7 molecules covering different quinazolines derivatives was employed as test set to evaluate the predictivity of training set. After validate the model predict the activity of the unknown molecule as EGFR inhibitor. Here basic moiety selected for the derivatization was 1, 3-benzothiazol-2-amine. First create the database of the new molecule and then the new derivatives thus generated were incorporated into the test data set and their activity was predicted on the basis of the PLS model generated for the training set. pIC_{50} value is a measure of the potency of the compound. When the pIC_{50} value is higher, the compound is more potent. With the highest pIC_{50} value of the compound, was synthesized and perform the cytotoxic effect of that synthesized molecule.

Different sets of subdescriptors employed in this study, met the orthogonality of the model, which ensures that each descriptor is encoding different properties from others. Traditional descriptors such as, logP *etc.,* are more relevant to drug transport or pharmacokinetics than the receptor affinity. Further, traditional descriptors consider the whole molecules property rather than the distinguishing details of sub-structural differences.

RESULTS AND DISSCUSION

Training set of 67 derivatives of Quinazoline, which represents quinazoline with structural modification on aromatic amine, and attached to core nucleus were employed. A test set of 7 molecules used for evaluation, encompassed a wide range of structures to validate the model. Correlation Plot of pIC50 between experimental values and the predicted values in training set is shown in Figure **1**. Also, Correlation Plot of pIC50 between experimental values and the predicted values in test set by considering the model developed by training set (model fit) is shown in Figure **2**.

Following statistical measures were used to correlate biological activity and molecular descriptors:

N = number of samples, r2 = coefficient of correlation, q2 = cross validated coefficient of correlation, r^2 pred. = coefficient of correlation of test set.

Model developed has a correlation of coefficient (r2) of 0.701 and cross validated r2 (q2) of 0.616. The model developed predicts 87.87 % of variance and is validated by internal validation with correlation of coefficient (r2) of 0.848 and cross validated r2 (q2) of 0.573 and external set of compounds with a predictive correlation of coefficient of 0.900. Results

comparing observed and predicted pIC50 along with residual values are shown in Table 1. The predictivity of model evaluated by test set of compounds is mentioned in Table 2.

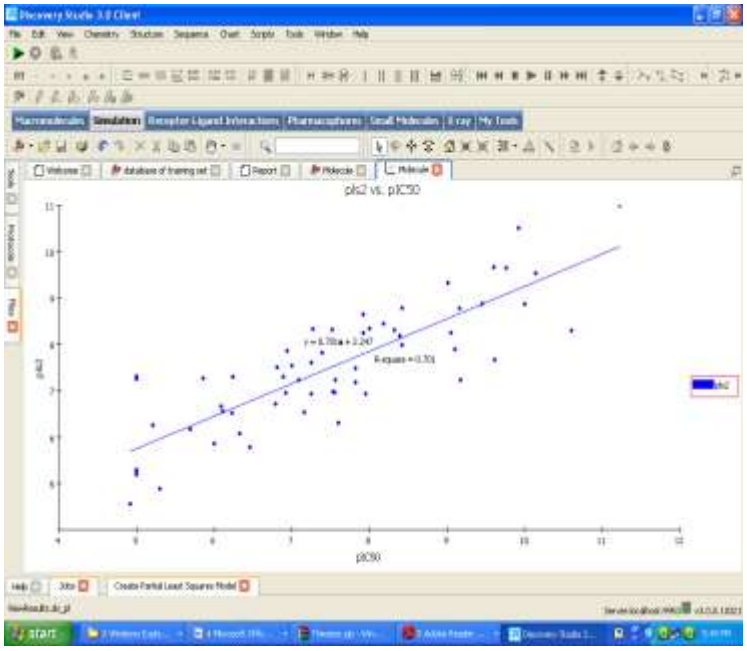


Fig 1: PLS Molecular Point Plot for training set

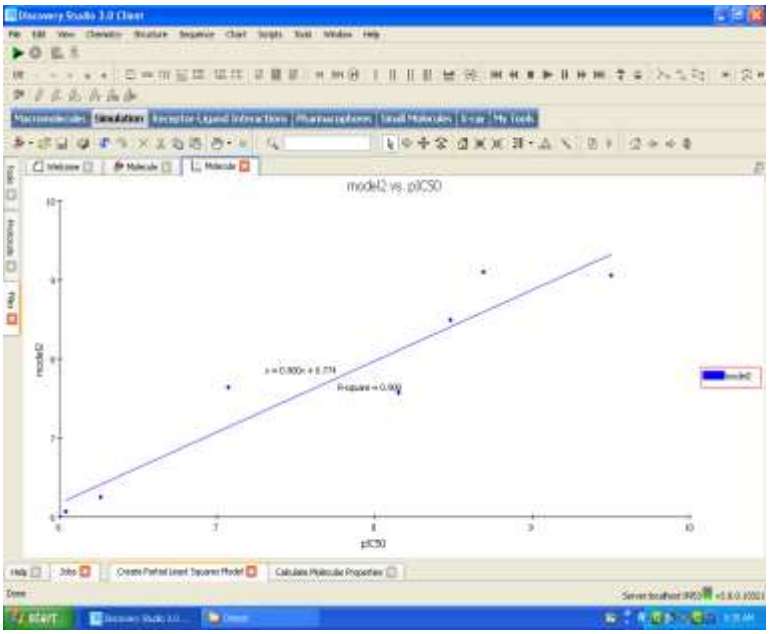
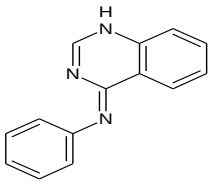
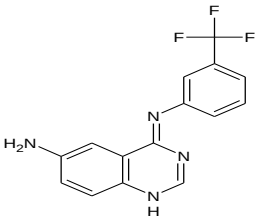
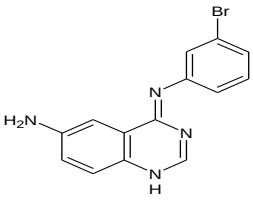
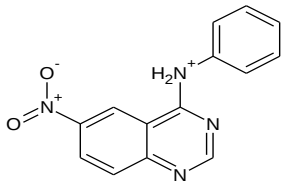
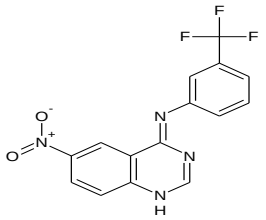
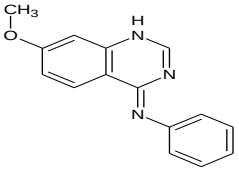


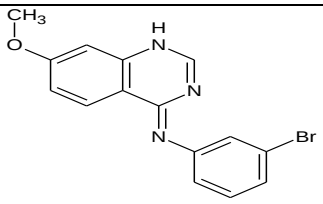
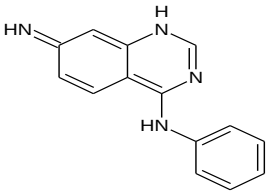
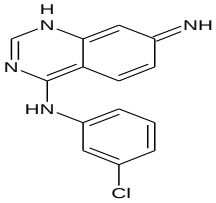
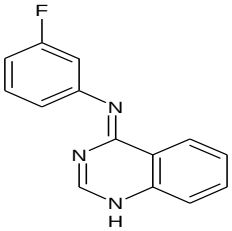
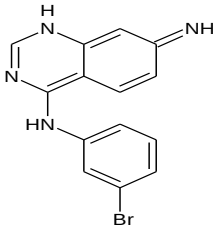
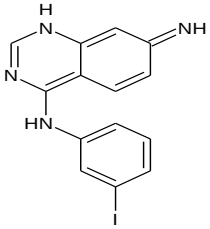
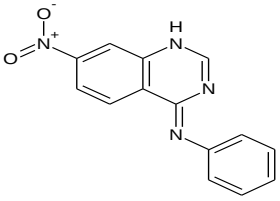
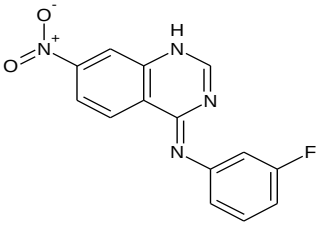
Fig 2: PLS Molecular Point Plot (Test set)

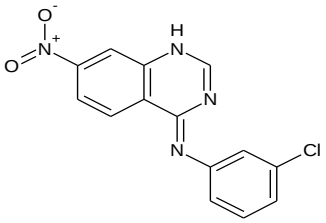
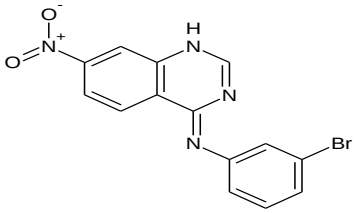
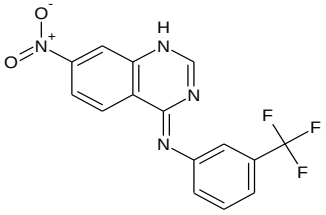
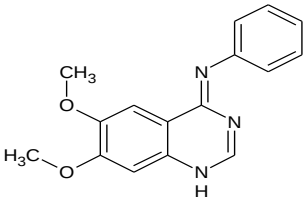
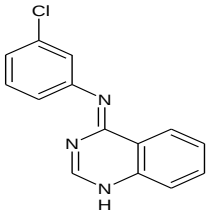
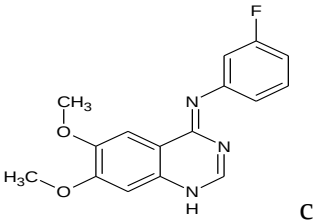
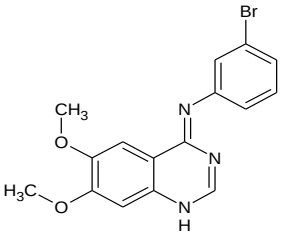
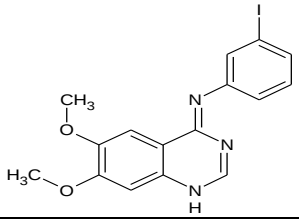
9 sub-descriptors pruned from 4 major groups of descriptors were adequate to explain different properties of quinazoline moiety. The Major group of descriptors involved sub groups like AlogP, Molecular property counts, Surface area and volume and topological descriptors. AlogP predicts logP of molecules parameterizing the contributions of various atoms to the overall molecular partition coefficient using constrained least squares fitting to a training set of

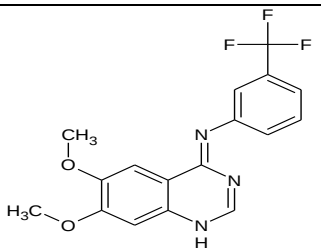
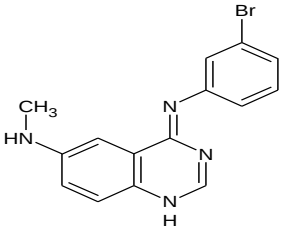
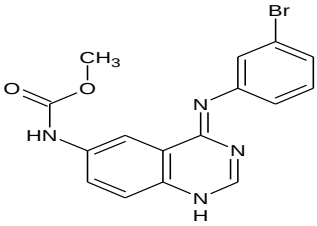
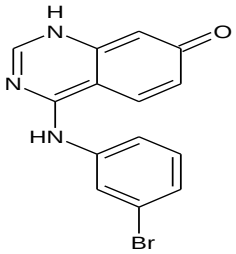
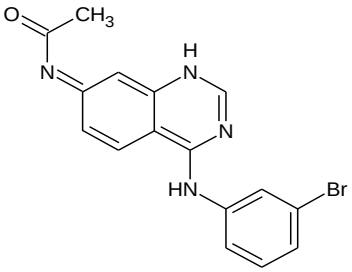
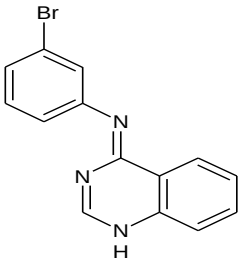
compounds with experimentally measured partition coefficients. In order to get reasonable correlations, the most common elements contained in drugs (hydrogen, carbon, oxygen, sulphur, nitrogen and halogens) are divided into several different atom types depending on the environment of the atom within the molecule. Alogp provides a rough estimate for a wide variety of molecules.

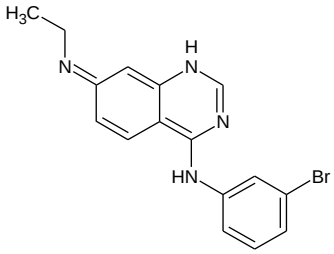
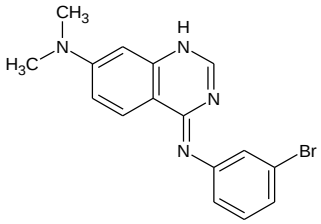
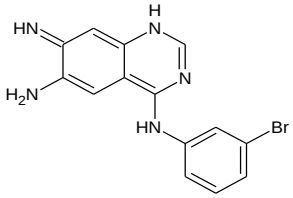
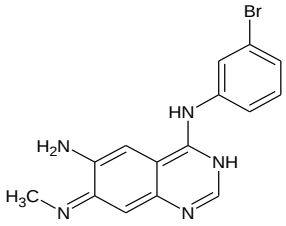
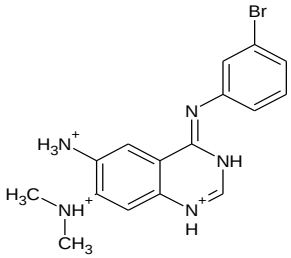
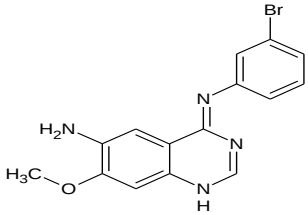
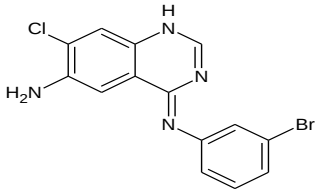
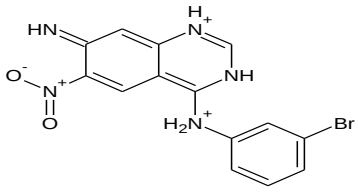
Table 1. Structures, Experimental and Predicted Activity of quinazolines Used in Training Set for EGFR Inhibition

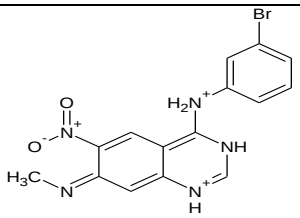
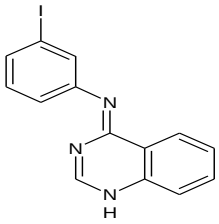
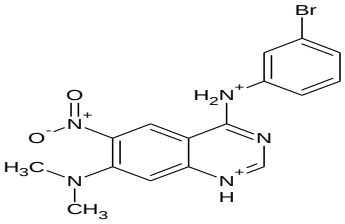
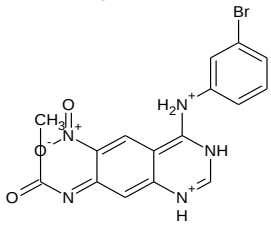
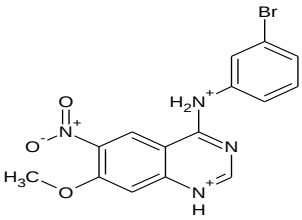
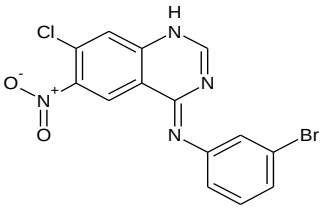
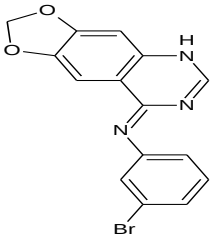
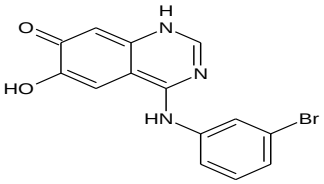
Name	Structure	Experimental pIC_{50}	Predicted pIC_{50}	Residuals
egfr 1		6.46	5.78378	0.676216
egfr 10		6.24	7.30152	-1.06152
egfr 11		9.1	7.89381	1.20619
egfr 12		5.3	4.88614	0.413862
egfr 14		5	5.26211	-0.262111
egfr 15		6.92	6.95909	-0.0390907

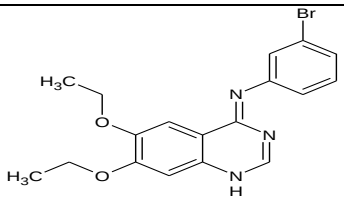
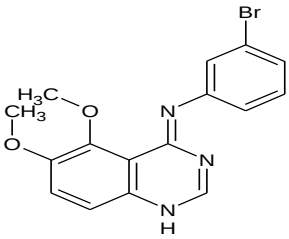
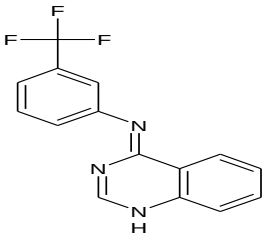
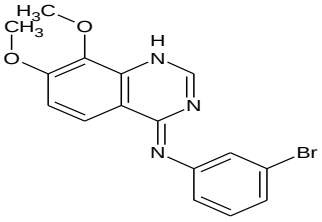
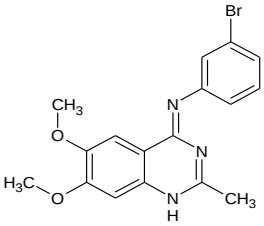
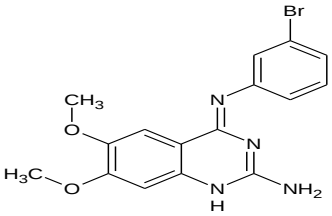
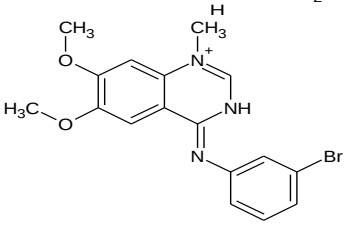
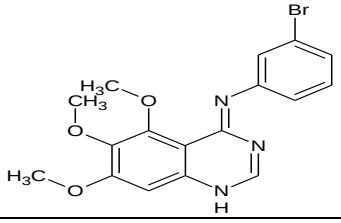
egfr 16		8	8.34383	-0.343835
egfr 17		7	7.54124	-0.541239
egfr 19		9.6	9.66531	-0.0653093
egfr 2		7.25	7.6103	-0.360305
egfr 20		10	8.86876	1.13124
egfr 21		9.45	8.87632	0.573681
egfr 23		4.92	4.55843	0.361566
egfr 24		5.21	6.25475	-1.04475

egfr 25		6.09	6.66825	-0.578252
egfr 26		6	5.85988	0.140125
egfr 28		5	5.29219	-0.292193
egfr 29		7.53	6.98123	0.548771
egfr 3		7.63	8.03812	-0.408115
egfr 30		8.42	8.78416	-0.364158
egfr 32		10.6	8.30092	2.29908
egfr 33		9.05	8.25243	0.797566

egfr 34		9.61	7.66858	1.94142
egfr 35		8.39	8.18555	0.20445
egfr 37		7.92	8.24738	-0.327376
egfr 38		8.32	8.31046	0.00953651
egfr 39		7.39	7.82096	-0.43096
egfr 4		7.56	7.23305	0.326954

egfr 41		7.92	8.64877	-0.728771
egfr 42		7.95	6.93546	1.01454
egfr 43		9.92	10.5114	-0.591363
egfr 44		9.16	8.77948	0.380517
egfr 45		6.79	6.7153	0.0746987
egfr 46		8.42	7.98698	0.433018
egfr 47		8.18	8.4462	-0.266203
egfr 48		7.27	8.32959	-1.05959

egfr 49		7.16	6.53586	0.624142
egfr 5		7.09	7.23205	-0.142045
egfr 50		5.69	6.17055	-0.480552
egfr 51		7.55	6.9614	0.588596
egfr 52		7.82	7.48591	0.334087
egfr 53		7.6	6.30892	1.29108
egfr 54		7.82	7.18393	0.636074
egfr 55		9.76	9.64899	0.111007

egfr 56		11.22	10.9984	0.221578
egfr 59		5.86	7.26909	-1.40909
egfr 6		6.23	6.52042	-0.290419
egfr 60		5	7.31249	-2.31249
egfr 61		5	5.19163	-0.191626
egfr 62		6.33	6.07745	0.252553
egfr 63		6.81	7.50557	-0.69557
egfr 64		9.17	7.23541	1.93459

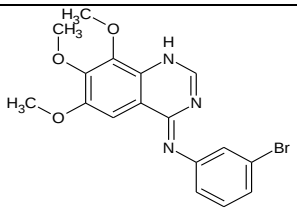
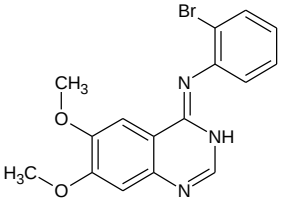
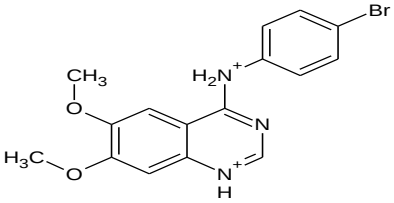
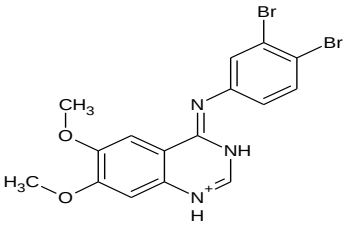
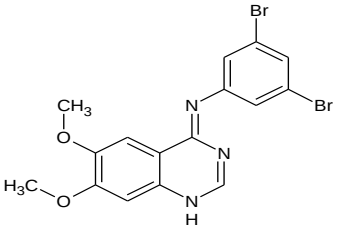
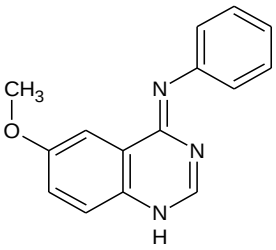
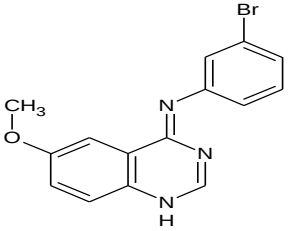
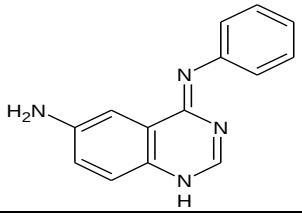
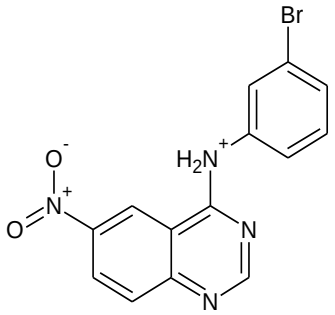
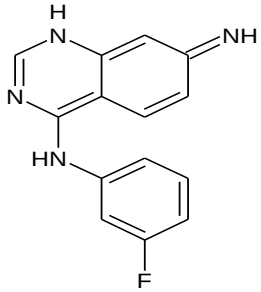
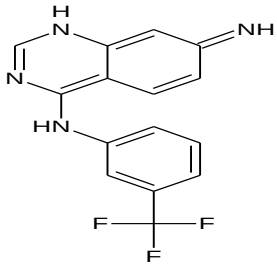
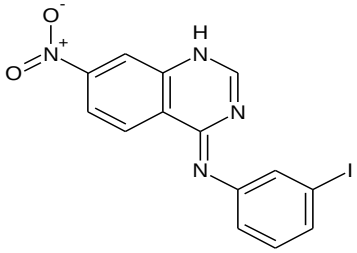
egfr 65		5	7.25451	-2.25451
egfr 66		6.89	7.30428	-0.41428
egfr 67		9.01	9.32766	-0.317661
egfr 68		10.14	9.53603	0.603967
egfr 69		6.94	7.86263	-0.922627
egfr 7		7.25	6.9339	0.316103
egfr 8		7.52	8.319	-0.799002
egfr 9		6.11	6.56219	-0.452186

Table 2. Structures, Experimental and Predicted Activity of quinazolines Used in Test Set for EGFR Inhibition

Name	Structure	Experimental pIC_{50}	Predicted pIC_{50}	Residuals
egfr 13		6.04	6.06846	-0.0284566
egfr 18		8.69	9.10297	-0.412973
egfr 22		8.48	8.49568	-0.0156771
egfr 27		6.26	6.25053	0.00946516

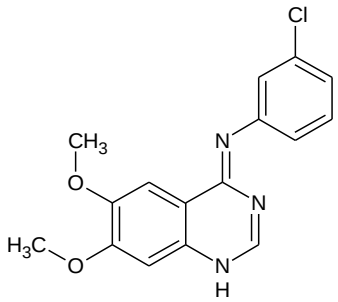
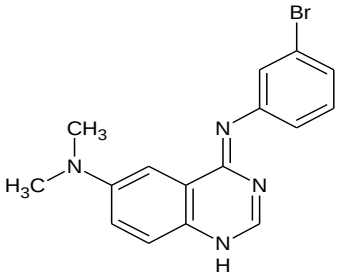
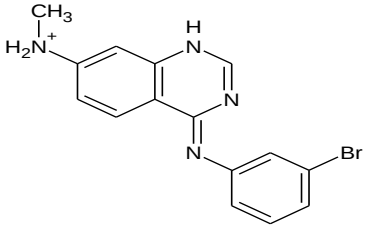
egfr 31		9.5	9.05971	0.440287
egfr 36		7.07	7.64015	-0.570149
egfr 40		8.15	7.5725	0.577503

Table 3 : Molecular Descriptors Used in Present Study:

Sr.no.	Descriptors	Description
2D Descriptors : Use the atoms and connection information of the molecule for the calculation		
ALogP Descriptors		
1.	ALogP	Log of the octanol-water partition coefficient using Ghose and Crippen's method.
2.	ALogP_Count	An array of 120 numbers that correspond to the 120 Ghose and Crippen atom types. The content of each array element is the number of times that particular atom type was seen in the molecule.
Molecular property count descriptors		
3.	Num_AromaticRings	Base rings that are aromatic.
4.	Num_H_Acceptors	Hydrogen Bond Acceptors are defined as heteroatoms (Oxygen, Nitrogen, Sulfur, or Phosphorus) with one or more lone pairs, excluding atoms with positive formal charges,

		amide and pyrrole-type Nitrogens, and aromatic Oxygen and Sulfur atoms in heterocyclic rings.
5.	Num_H_Donors	Hydrogen Bond Donors are defined as heteroatoms (Oxygen, Nitrogen, Sulfur, or Phosphorus) with one or more attached Hydrogen atoms.
Surface area and Volume		
6.	Molecular_Fractional PolarSurfaceArea	The ratio of the polar surface area divided by the total surface area.
Topological Descriptors		
7.	JX	Balaban Indices
8.	PHI	Kappa Shape Indices
9.	Kappa_3	Kappa Shape Indices

Molecular property count calculates of the number atoms, bonds, rings, chains, hydrogen acceptor and donors, etc. in the molecules that are necessary for a molecule to perform biological activities under consideration.

Molecular surface area and volume considers surface area, polar surface area, solvent accessible surface area, solvent accessible volume using 2D approximation and ratios using these properties in different combination which have direct or indirect effects on biological activities of molecules.

In tyrosine kinase family, EGFR is a growth factor receptor that induces cell differentiation and proliferation upon activation through the binding of one of its ligands. The receptor is located at the cell surface, where the binding of a ligand activates a tyrosine kinase in the intracellular region of the receptor. EGFR is thought to be involved in the development of cancer. EGFR inhibitors are candidate therapeutic agents for the treatment of various cancers.

The ATP binding site of EGFR based on published crystal structure reviles useful information by the virtue of which different binding interaction of quinazoline derivatives in comparison to establish EGFR inhibitor like Erlotinib can be understood.

According to the pharmacophore model of the ATP binding pocket of EGFR, five regions conserved throughout the protein kinases are classifiable. These include adenine region, hydrophobic region I and II, phosphate binding region, and sugar pocket. The aniline moiety of erlotinib is inserted into the hydrophobic pocket of Val702, Met742, and Leu764, denoted as hydrophobic region I. Two significant hydrogen bonds include Met769 - NH N1 quinazoline and Gln767 – CO HC_2 quinazoline in erlotinib can be assimilated by the interaction of

aromatic ring of benzothiazole moiety of predicted compound, that is 4-[1,3-benzothiazol-2-yl]-N-[(1E)-(4-nitrophenyl)methylene]aniline[10].(Figure 3)

This possible interaction of the drug with the ATP binding site of the enzyme provides possible docking interaction of the drug. Enhance observed pIC_{50} value (7.88) is quite justifiable.

The similar type of activity is observed in the preliminary cytotoxic activity. Where in % inhibition of the tumour cell(concentration of the cells @ 5000cells/well) was found to be in line with that of comparative drug (gefitinib) thus indicating the potential of this drug.

Further investigation is requiring understanding the bio efficacy of these drugs on different cell lines and effect of the drug in vivo. Further efforts are also require understanding the identification of impurities of synthesized drug.

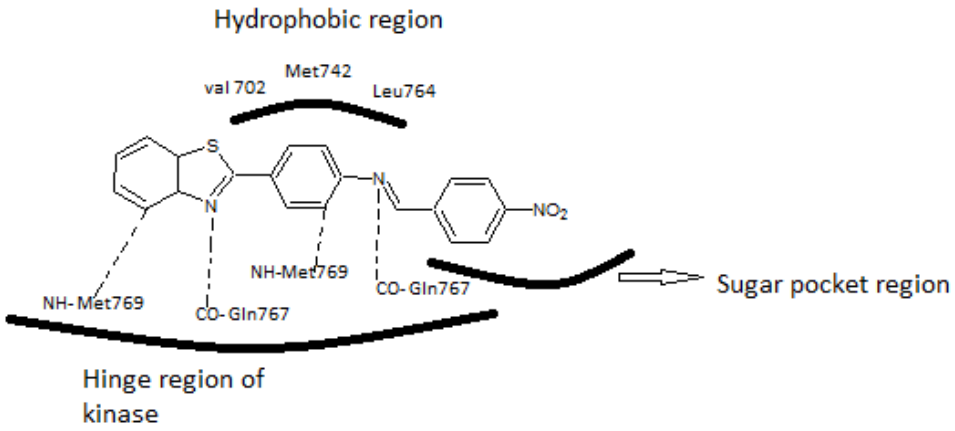


Figure 3. Hypothetical Interaction

CONCLUSION.

In conclusion, the model developed to predict the structural features of quinazolines to inhibit EGFR reveals useful information about the structural features required for the molecules. In the models developed AlogP, molecular properties, molecular property Count, and topological descriptors were the major contributing descriptors. Descriptor values obtained helps us to understand the structural features required by ATP binding site of EGFR. Since a wide range of structures were incorporated in the training set, molecular Property Count, Topological descriptors also influenced the predictivity of model and that derived model is to predict the activity of unknown molecule as a EGFR inhibitor. The results obtained from QSAR study consider not only wide range of structures, but also various physicochemical interactions

involved in enzyme inhibitor complex. Hence the model proposed in this work is useful in describing QSAR of quinazoline derivatives as EGFR inhibitors and can be employed to design new molecule with specific inhibitory activity as EGFR inhibitor.

In order to evaluate the efficacy of model generated we synthesized the molecules with good pIC_{50} value prediction and characterize the assay structure followed by bio efficacy study.

On completion of preliminary cytotoxicity study to understand the inhibitory effect of test compound in comparison with standard drug against the drug sample of leukemic cell (concentration of cell@5000 cells/well) was carried out.

Results of preliminary cytotoxic study (% inhibition=34.14%) indicate that the experimental values obtained and predicted value are in line with each other. Results are as follows:

The model used to predict the molecules' bio efficacy with the better accuracy.

The values predicted and model developed on the basis of EGFR inhibition categorically which in turn is correlated to the anticancer activity.

In our experimentation, though the synthesized molecule had a predicted pIC_{50} value=7.88 denote elicited the cytotoxic effect as per expectation.

Here, the standard drug gefitinib is a categorical EGFR inhibitor but also has shown good experimental preliminary cytotoxic effect (% inhibition= 84.22%).

REFERENCES

1. Kurup A, Garg R, Hansch C. Comparative QSAR study of tyrosine kinase inhibitors. Chem Rev 2001;101: 2573-2600.
2. Palmer BD, Kraker AJ, Hartl BG, Panopoulos AD, Panek RL, Batley BL *et al*,. Structure activity relationships for 5-substituted 1-phenylbenzimidazolesas selective inhibitors of the platelet-derived growth factor receptor. J Med Chem 1999; 42: 2373-2382.
3. Oblak M, Randic M, Solmajer T. Quantitative structure-activity relationship of flavonoid analogues. 3. Inhibition of p56lck protein tyrosine kinase. J Chem Inf Comput Sci 2000 Jul-Aug; 40(4): 994-1001.
4. Naumann T, Matter H. Structural classification of protein kinases using 3D molecular interaction field analysis of their ligand binding sites: target family landscapes. J Med Chem 2002; 45: 2366-2378.
5. Bridges AJ, Cody DR. Synthesis and Src Kinase Inhibitory Activity of a Series of 4-Phenylamino-3-quinolinecarbonitriles J Med Chem. 1996; 39: 267-276.

6. Cuiling Ma, Kimberly A, Hong Lin, Gang Chen, Chanshu Huang, Jia Luo.Tje role of epidermal growth factorreceptor in ethanol-mediated inhibition of activator protein-1 transactivation. Biochem Pharmacol 2005; 69: 1785-1794.
7. Albuschat R, Löwe W, Weber M, Luger P, Jendrossek V. 4- Anilinoquinazolines with Lavendustin A subunit as inhibitors of epidermal growth factor receptor tyrosine kinase: syntheses, chemical and pharmacological properties. Eur J Med Chem 2004; 39: 1001-1011.
8. Allen B. Richon ,Stanley S. Young. An introduction to QSAR methodology-Network Science. Network Science Corporation. 1997
9. Donald P. Visco RSP. The signature molecular descriptor. 1. Using extended valence sequences in QSAR and QSAR studies. J Chem Inf Comput Sci 2003; 43: 707-720.
10. Alexander J. Bridges HZ, Donna R Cody .Tyrosine Kinase Inhibitors. 8. An Unusually Steep Structure-Activity Relationship for Analogues of 4-(3-Bromoanilino)-6,7-dimethoxyquinazoline (PD 153035), a Potent Inhibitor of the Epidermal Growth Factor Receptor. J Med Chem 1996; 39: 267-276.