Journal of Saudi Chemical Society (2013) **17**, 361–379

King Saud University

Journal of Saudi Chemical Society

www.ksu.edu.sa
www.sciencedirect.com

ORIGINAL ARTICLE

A comparative QSAR analysis and molecular docking studies of quinazoline derivatives as tyrosine kinase (EGFR) inhibitors: A rational approach to anticancer drug design

Malleshappa N. Noolvi *, Harun M. Patel

Computer Aided Drug Discovery Department, ASBASJSM College of Pharmacy, Bela (Ropar) 14011, Punjab, India

Received 10 April 2011; accepted 29 April 2011
Available online 6 May 2011

KEYWORDS

Protein tyrosine kinase (EGFR);
Multiple linear regression (MLR);
Principal component regression (PCR);
Partial least squares (PLS);
kNN-MFA;
Molecular docking

Abstract In this paper, an attempt was made to develop a quantitative structure–activity relationship (2D and 3D QSAR) and molecular docking studies on a series of quinazoline derivatives acting as protein tyrosine kinases (EGFR) inhibitors. 2D QSAR was performed using multiple linear regression (MLR), principal component regression (PCR) and partial least squares regression (PLS) methods. Among these three methods, multiple linear regression (MLR) method has come out with a very promising result as compared to other two methods. According to Model-1 by MLR anticancer activity of quinazoline derivatives were influenced by individual (H-donor count, and XlogP) and alignment independent descriptor (T_C_Br_1, T_2_O_1 and T_2_N_7) help in understanding the effect of substituent at different position of quinazolines. The contribution plot of steric and electrostatic field interactions generated by 3D-QSAR shows interesting results in terms of internal and external predictability. Molecular field analysis was applied for the generation of steric and electrostatic descriptors based on aligned structures. Steric and electrostatic field effects are discussed in the light of contribution plot generated. Finally, molecular docking analysis was carried out to better understand the interactions between EGFR target and inhibitors in this series. Hydrophobic and hydrogen bond interactions lead to identification of active binding sites

* Corresponding author. Address: Department of Pharmaceutical Chemistry, ASBASJSM College of Pharmacy, Bela (Ropar) 14011, Punjab, India. Tel.: +91 9417563874; fax: +91 1881263655.
E-mail addresses: mnoolvi@yahoo.co.uk (M.N. Noolvi), hpatel_38@yahoo.com (H.M. Patel).

Peer review under responsibility of King Saud University.
doi:10.1016/j.jscs.2011.04.017

Production and hosting by Elsevier

of EGFR protein in the docked complex. The present study is more versatile than the earlier reported methods. Hence the model proposed in this work can be employed to design new derivatives of quinazoline with specific tyrosine kinase (EGFR) inhibitory activity.

1. Introduction

A considerable amount of experimental studies have been carried out with 4-anilinoquinazoline derivatives which are potent and highly selective inhibitors of epidermal growth factor (EGFR) phosphorylation at the ATP binding site. These compounds cause inhibition of EGFR produced by abnormal signal transduction via hyperactivation of tyrosine protein kinases due to overexpression or mutation, thus leading to anticancer activities against human lung cancer, breast cancer, squamous head, and neck carcinomas (Herbst et al., 2004). A number of 5-substituted 4-anilinoquinazoline derivatives were synthesized by Ballard et al. (2005) and these compounds were evaluated in erbB2 and EGFR kinase assays measuring inhibition of phosphorylation at the ATP binding site. Rewcastle et al. (1995) prepared two series of 4-(phenylmethyl) amino and 4-(3-bromophenyl) amino quinazoline compounds and evaluated their inhibitory activities against EGFR tyrosine kinases that ultimately led to structure–activity relationships of these compounds. Structure–activity relationships of a series of quinazoline derivatives studied by Gibson et al. (1997) identified 4-(4-isoquinolylamino) quinazoline and 4-(trans-2-phenyl cyclopropyl amino) quinazoline as potent EGFR inhibitors against a tumour xenograft model (A431 vulval carcinoma in nude mice). In order to study the structure–activity relationships, Hennequin et al. (2006) synthesized a number of 4-anilinoquinazoline compounds, and it was shown that anilinoquinazolines possessing C-6 aminomethyl side-chains act as potent and selective inhibitors of EGFR kinase. Structure–activity relationships for 4-anilinoquinazolines and modelling of the binding of these compounds to EGFR have also been studied by Denny (2001). Bridges et al. (1996) synthesized numerous 4-anilinoquinazoline derivatives acting as EGFR-mediated potential tyrosine kinase inhibitors, and the anticancer activities of these compounds against human A431 carcinoma cell vesicles have been reported. The development of tyrosine kinase inhibitors has therefore become an active area of research in pharmaceutical science. One could not, however, confirm that the compounds synthesized would always possess good inhibitory activity to Tyrosine kinase, while experimental assessments of inhibitory activity of these compounds are time-consuming and expensive. Consequently, it is of interest to develop a prediction method for biological activities before the synthesis. Quantitative structure–activity relationship (QSAR) searches information relating chemical structure to biological and other activities by developing a QSAR model. Using such an approach one could predict the activities of newly designed compounds before a decision is being made whether these compounds should be really synthesized and tested.

With the above facts and in continuation of our research for newer anticancer agent (Noolvi and Patel, 2011a,b,c; Noolvi et al., 2011a,b; Noolvi et al., 2010; Manjula et al., 2009; Badiger et al., 2006) in the present study, we reported 2D, 3D-QSAR and molecular docking studies on a series of EGFR inhibitors to provide further insight into the key structural features required to design potential drug candidates of this class.

2. Materials and methods

2.1. 2D-QSAR methodology

2.1.1. Data set

Forty-five molecules belonging to 4-anilino quinazoline derivatives as EGFR tyrosine kinase inhibitors were taken from the literature and used for QSAR analysis (Bridges et al., 1996). The above reported series of quinazoline derivatives showed wide variations in their structures and potency profiles. Various 2D-QSAR models were generated for this series using multiple linear regression (MLR), principal component regression (PCR) and partial least squares (PLS) regression methods and those which come out with promising results are discussed here. QSAR models were generated by a training set of 34 (MLR), 33 (PCR) and 30 (PLS) molecules. Predictive power of the resulting models was evaluated by a test set of 11 (MLR), 12 (PCR) and 15 (PLS) molecules with uniformly distributed biological activities. The test set was selected based on the criteria given by Oprea et al. (1994). The structures of all the compounds along with their actual and predicted biological activities are presented in Table 1.

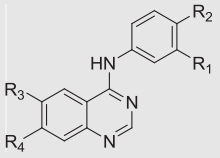
2.1.2. Biological activities

The biological activities were converted into the corresponding pIC_{50} values, where IC_{50} value represents the drug in molar concentration that causes 50% inhibition of phospholipase Cγ1 phosphorylation by EGFR. All the IC_{50} values had been obtained using the same assay method (Fry et al., 1994). The IC_{50} values of reference compounds were checked to ensure that no difference occurred between different groups. The pIC_{50} values of the molecules under study spanned a wide range from 5 to 11. Since some compound exhibited insignificant/no inhibition, such compounds were excluded from the present study.

2.1.3. Computational data

The data set used for the QSAR analyses contains 45 molecules belonging to quinazoline derivatives as tyrosine kinase (EGFR) inhibitors. All the structures of the compounds were drawn in 2D-APPL mode of software and exported to 3D model. The modelling analyses, calculations, and visualizations for 2D QSAR were performed using the V-Life Molecular Design Suite 3.0 (Vlife MDS). The compounds were then subjected to conformational analysis and energy minimization using Montocarlo conformational search with RMS gradient of 0.001 kcal/mol using a MMFF force field. Montocarlo conformational search method is similar to the RIPS method that generates a new molecular conformation by randomly perturb-

Table 1 Structure, experimental and predicted activity of quinazolines used in training and test set using MLR analysis.



S. No.	R_1	R_2	R_3	R_4	pIC_{50}[a]		Residual
					Exp.	Pred.	
1	H	H	H	H	6.46	6.68	−0.22
2	CH_3	H	H	H	6.04	5.89	0.15
3	Cl	H	H	H	7.63	7.85	−0.22
4	Br	H	H	H	7.56	7.34	0.22
5	I	H	H	H	7.09	6.81	0.28
6^T	CF_3	H	H	H	6.23	6.45	−0.22
7^T	Br	H	NO_2	H	6.04	5.87	0.17
8	Br	H	OCH_3	H	6.45	6.23	0.22
9^T	Br	H	H	NO_2	6.00	6.45	−0.45
10	Br	H	H	OCH_3	8.00	7.65	0.35
11^T	Br	H	OH	OH	9.76	9.12	0.64
12^T	Br	H	NH_2	NH_2	9.92	9.68	0.24
13	F	H	H	H	7.25	7.75	−0.5
14	H	H	OCH_3	H	7.24	7.48	−0.24
15	H	H	NH_2	H	6.11	5.76	0.35
16^T	CF_3	H	NH_2	H	6.24	6.04	0.2
17	H	H	OCH_3	H	6.92	6.86	0.06
18	H	H	H	NH_2	7.00	7.56	−0.56
19	CF_3	H	H	NH_2	8.48	8.38	0.1
20	F	H	H	NO_2	5.21	5.36	−0.15
21	Cl	H	H	NO_2	6.09	6.45	−0.36
22	I	H	H	NO_2	6.26	6.12	0.14
23	H	H	OCH_3	OCH_3	7.53	7.65	−0.12
24	F	H	OCH_3	OCH_3	8.42	8.23	0.19
25	Cl	H	OCH_3	OCH_3	9.50	9.56	−0.06
26	I	H	OCH_3	OCH_3	9.05	8.76	0.29
27^T	CF_3	H	OCH_3	OCH_3	9.61	9.89	−0.28
28^T	Br	H	$NHCH_3$	H	8.39	8.12	0.27
29	Br	H	$N(CH_3)_2$	H	7.07	7.56	−0.49
30	Br	H	$NHCOOCH_3$	H	7.92	7.86	0.06
31	Br	H	H	OH	8.32	8.12	0.2
32^T	Br	H	H	OH	7.39	7.86	−0.47
33	Br	H	H	$NHCOCH_3$	8.15	8.54	−0.39
34	Br	H	H	$NHCH_3$	7.92	7.68	0.24
35^T	Br	H	H	NHC_2H_5	7.95	7.46	0.49
36^T	Br	H	H	$N(CH_3)_2$	9.16	9.45	−0.29
37	Br	H	NH_2	$NHCH_3$	6.79	6.54	0.25
38	Br	H	NH_2	$N(CH_3)_2$	7.95	8.68	−0.73
39	Br	H	NH_2	OCH_3	8.18	7.56	0.62
40	Br	H	NH_2	Cl	7.16	6.87	0.29
41	Br	H	NO_2	$NHCH_3$	7.82	7.57	0.25
42	Br	H	NO_2	OCH_3	7.60	6.87	0.73
43	Br	H	OC_2H_5	OC_2H_5	11.22	11.65	−0.43
44	Br	$O(CH_2CH_2CH_3)$	$O(CH_2CH_2CH_3)$		9.76	9.56	0.20
45	H	Br	OCH_3	OCH_3	9.01	8.56	0.45

Expt. = experimental activity, Pred. = Predicted activity, T = test set.
[a] $-\text{Log}\,(IC_{50} * 10^{-6})$.

ing the position of each coordinate of each atom in molecule. Most stable structure for each compound was generated after energy minimization and used for calculating various physicochemical descriptors.

2.1.4. Molecular descriptors

The various descriptors selected for 2D QSAR were vdW Surface Area (van der Waals surface area of the molecule), −ve Potential Surface Area (total van der Waals surface area

with negative electrostatic potential of the molecule), +ve Potential Surface Area (total van der Waals surface area with positive electrostatic potential of the molecule) dipole moment, Y compDipole (y component of the dipole moment), element count, slogP, path count, cluster, distance based topological indices, connectivity index, hydrophobic and hydrophilic areas like SA most hydrophilic (most hydrophilic value on the vdW surface by Audry Method using Slogp), SA most hydrophobic–hydrophilic distance (distance between most hydrophobic and hydrophilic point on the vdW surface by Audry method using Slogp), SA hydrophilic area (vdW surface descriptor showing hydrophilic surface area by Audry Method using SlogP) and SK most hydrophilic (most hydrophilic value on the vdW surface by Kellog method using Slogp), radius of gyration, Wiener's index, moment of inertia, semi-empirical descriptors, HOMO (Highest occupied molecular orbital), LUMO (lowest unoccupied molecular orbital), heat of formation and ionization potential. The list of descriptors along with their description is given in Table 2. Besides these all, alignment independent descriptors were also calculated.

2.1.5. Selection of training and test set

In order to obtain a validated QSAR model for the purpose of meaningful prediction, an available dataset should be divided into the training and test sets. For the prediction statistics to be reliable, the test set must include at least five compounds (Golbraikh and Tropsha, 2002). Ideally, the division into the training and test set must satisfy the following three conditions: (i) all representative compound-points of the test set in the multidimensional descriptor space must be close to those of the training set. (ii) All representative points of the training set must be close to those of the test set. (iii) The representative points of the training set must be distributed within the whole area occupied by the entire dataset (Golbraikh and Tropsha, 2002).

The dataset of 45 molecules was divided into training and test set by sphere exclusion (SE) method (Shen et al., 2002) for MLR, PCR and PLS model with dissimilarity value of 2.4, 2.8 and 3.0, respectively, using pIC_{50} activity field as dependent variable and various 2D descriptors as independent variables. In classical sphere exclusion algorithm the molecules are selected whose similarities with each of the other selected molecules are not higher than a defined threshold (Shen et al., 2002). Each selected molecule generates a hyper-sphere around itself, so that any molecule inside the sphere is excluded from the selection in the train set and driven towards the test set. The number of compounds selected and the diversity among them can be determined by adjusting the radius of the sphere (R). The different statistical models were developed using multiple linear regression (MLR), principal component regression (PCR) and partial least squares (PLS) regression methods. The equations were found to derive 2D-QSAR equation from different model building method (multiple regression, principle component regression and partial least squares regression) coupled with stepwise forward–backward variable selection method for assuming the biological activity with the help of physico-chemical descriptor values. Only those equations are discussed in Section 3 which come out with promising results from three methods (MLR, PCR and PLS).

2.1.6. Statistical parameters

Statistical measures used for the evaluation of models were the number of compounds in regression n, the regression coefficient r^2, the F-test (Fischer's value) for statistical significance F, the cross-validated correlation coefficient q^2 and the standard error of estimation r^2 and q^2. The regression coefficient r^2 is a relative measure of fit by the regression equation. It represents the part of the variation in the observed data that is explained by the regression. The correlation coefficient values closer to 1.0 represent the better fit of the regression. The F-test reflects the ratio of the variance explained by the model and the variance due to the error in the regression. High values of the F-test indicate that the model is statistically significant. Validation parameter, predictive r^2 (r^2_pred) was calculated for evaluating the predictive capacity of the model. A value of r^2_pred greater than 0.5 indicates the good predictive capacity of the QSAR model.

2.1.7. Model validation

This is done to test the internal stability and predictive ability of the QSAR models. Developed QSAR models were validated by the following procedure.

2.1.7.1. Internal validation. Internal validation was carried out using leave-one-out (q^2, LOO) method. For calculating q^2, each molecule in the training set was eliminated once and the activity of the eliminated molecule was predicted by using the model developed by the remaining molecules. The q^2 was calculated using the equation which describes the internal stability of a model.

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

where y_i and $\hat{y}_i$ are the actual and predicted activity of the ith molecule in the training set, respectively, and y_{mean} is the average activity of all molecules in the training set.

2.1.7.2. External validation. The predictive ability of the selected model was also confirmed by external validation of test set compounds which is also denoted with pred_r^2. The pred_r^2 value is calculated as follows:

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

where y_i and $\hat{y}_i$ are the actual and predicted activity of the ith molecule in the training set, respectively, and y_{mean} is the average activity of all molecules in the training set.

2.1.7.3. Randomization test. To evaluate the statistical significance of the QSAR model for an actual dataset, one-tail hypothesis testing was used (Golbraikh and Tropsha, 2003; Gilbert, 1976).The robustness of the models for training sets was examined by comparing these models to those derived for random datasets. Random sets were generated by rearranging the activities of the molecules in the training set. The statistical model was derived using various randomly rearranged activities (random sets) with the selected descriptors and the corresponding q^2 were calculated. The significance of the models hence obtained was derived based on a calculated Z score (Gilbert, 1976; Cramer et al., 1988). A Z score value is calculated by the following formula:

Table 2 List of 2D descriptors with their corresponding classification.

Descriptor classes	Descriptors name/Id
Constitutional descriptors	Mol. Wt. H-Acceptor Count, Rotatable bond count, No. halogen atoms, No. ester groups, No. rigid bonds, No. Aromatic rings, No. double bonds, No. total atoms, No. single bonds
Physico-chemical descriptors	Volume, slogp, smr, polarizability AHC, polarizability AHP, SKlogP value, water solubility, buffer solubility, SK_MP, AMR value, SKlogS value, SKlogPvp, SKlog S_buffer, SK_BP, AlogP98 value, solvation Free Energy, AlogP98 002 C, AlogP98 024 C, AlogP98 026 C, AlogP98 040 C, AlogP98 47 H, AlogP98 051 H, AlogP98 001 C, AlogP98 005 C, AlogP98 025 C, AlogP98 029 C, AlogP98 046 H, AlogP98 060 O, AlogP98 089 Cl
Electrostatic descriptors	Max. positive charge, max. positive hydrogen charge, local dipole index, relative negative charge, PPSA2, PNSA1, PNSA3, DPSA2, FPSA1, FPSA3, FNSA2, WPSA1, WPSA3, WNSA2, hydrophobic SA – MPEOE, negative charged polar SA – MPEOE, SAAA1, SAAA3, CHAA2, SCAA1, SCAA3, HRNCS, HRNCG, max negative charge, total positive charge, charge polarization, polarity parameter, relative positive charge, PPSA1, PPSA3, PNSA2, DPSA1, DPSA3, FPSA2, FNSA1, FNSA3, WPSA2, WNSA1, WNSA3, RNCS, positive charged polar SA – MPEOE, SAAA2, CHAA1, CHAA3, SCAA2, +ve potential surface area, −ve potential surface area, most +ve potential, most −ve potential, average potential, average +ve potential, average −ve potential, most +ve & −ve potential distance
Topological descriptors	chi1, chi4, chi5, chiV0, chiV1, chiV2, chiV3, chiV4, chiV5, 0PathCount, 3PathCount, 4PathCount, 5PathCount, chi6chain, chiV6chain, chi3Cluster, chiV3Cluster, 3ClusterCount, chi4pathCluster, chiV4pathCluster, 4pathClusterCount, kappa3, k1alpha, k2alpha, k3alpha, total structure connectivity index, Chi 3 path, Chi 4 path/cluster, VChi 0, VChi 2, VChi 3 cluster , VChi 4 cluster, VChi 5 path, Kier shape 2, Kier alpha 1, Kier alpha 3, Kier symmetry index, Chi 0, Chi 2, Chi 3 cluster, Chi 5 path, VChi 1, VChi 3 path, VChi 4 path, VChi 4 path/cluster, Kier shape 1, Kier shape 3, Kier alpha 2, Kier flexibility, Kier steric descriptor, Delta Chi 0, Delta Chi 2, Delta Chi 3 cluster, Delta Chi 5 path, difference chi 1, difference chi 3, difference chi 5, Delta Chi 1, Delta Chi 3 path, Delta Chi 4 path, Delta Chi 4 path/cluster, difference chi 0, difference chi 2, difference chi 4, IC, CIC, IAC total, I_adj_mag, I_dist_mag, I_edge_adj_mag, I_edge_adj_deg_mag, I_edge_dist_mag, BIC, SIC, I_adj_equ, I_adj_deg_equ, I_dist_equr, I_edge_adj_equ, I_edge_adj_deg_equ, I_edge_dist_equ, charge index 1, charge index 3, charge index 5, charge index 7, charge index 9, valence charge index 2, valence charge index 4, valence charge index 6, valence charge index 8, valence charge index 10, bound charge index 2, bound charge index 4, bound charge index 6, bound charge index 8, bound charge index 10, valence bound charge index 1, valence bound charge index 3, valence bound charge index 5, valence bound charge index 7, valence bound charge index 9, charge index 0, charge index 2, charge index 4, charge index 6, charge index 8, charge index 10, valence charge index 1, valence charge index 3, valence charge index 5, valence charge index 7, valence charge index 9, global topological charge index, bound charge index 1, bound charge index 3, bound charge index 5, bound charge index 7, bound charge index 9, valence bound charge index 0, valence bound charge index 2, valence bound charge index 4, valence bound charge index 6, valence bound charge index 8, valence bound charge index 10, Wiener index, Harary index, 2-MTI prime, Graph diameter, Graph Petitjean, Eccentric adjacency index, Odd–even index, ring degree-distance index, Balaban index JY, superpendentic index, Centralization_distance_matrix, Dispersion_distance_matrix, SC-3 cluster, SC-4 cluster, SC-5 path, SC-7 path, SC-9 path, solvation chi 0, solvation chi 2, solvation chi 3 cluster, solvation chi 4 cluster, solvation chi 5 path, VS-3, VS-5, molecular walk count 3, molecular walk count 5, Path/walk 3, Narumi HTI, Pogliani index, degree complexity, graph distance complexity, mean square distance index, edge Wiener index, edge MTI, edge connectivity index, hyper Wiener index, 1st Zagreb, quadratic index, 2-MTI, Gutman MTI, graph radius, eccentric connectivity index, Platt number, vertex degree-distance index, Balaban index JX, Xu, Unipolarity_distance_matrix, SC-3 path, SC-4 path SC-4 path/cluster, SC-6 path, SC-8 path, SC-10 path, solvation chi 1, solvation chi 3 path, solvation chi 4 path, solvation chi 4 path/cluster, VS-2, VS-4, molecular walk count 2, molecular walk count 4, Path/walk 2, Narumi ATI, Narumi GTI, ramification index, graph vertex complexity, graph distance index, mean distance deviation, Edge Hyper Wiener index, Edge Gutman MTI, DistTopo, connectivity index, radius of gyration, MomInertia *X*, MomInertia *Y*, MomInertia *Z*, Balaban index J, Hosoya index, Id, Id Average, Idw, SsCH3count, SssCH2count, SaaCHcount, SaasCcount, SssssCcount, SdOcount, SsCH3E-index, SaaCHE-index, SdssCE-index, SaasCE-index, SaasN(Noxide)E-index, SdOE-index, SssOE-index, SssSE-index, SddssS(sulfate)E-index, SsClEindex, SsFE-index, E-state SsCH3, E-state SaasC, E-state S_hydrophobic, E-state S_hydrophobic_unsat, E-state S_polar, E-state S_negative_charged_group, E-state SHCsatu, E-state SH_hydrophobic, E-state SaaCH, E-state SdssC, E-state SaaaC, E-state SaasN, E-state SssO, E-state SsCl, E-state S_hydrophobic_sat, E-state S_none, E-state S_hbond_acceptor, E-state SHaaCH, E-state SHCsats
Semi-empirical descriptors	Heat of formation, HOMO energy, LUMO energy, sum of absolute charges, X compDipole, Y compDipole, Z compDipole, dipole moment, quadrupole 1, quadrupole 2, quadrupole 3, QM dipole X, QM dipole Y, QM dipole Z, QM dipole magnitude, XX polarizability, YY polarizability, ZZ polarizability, XY polarizability, XZ polarizability, YZ polarizability, average polarizability, XA most hydrophobic, XA average hydrophobicity, XA most hydrophobic hydrophilic distance, XK hydrophobic area, XK most hydrophobic, XK most hydrophilic, XK average, XK most hydrophobic hydrophilic distance, SA hydrophobic area, SA hydrophilic area, SA most hydrophobic, SA most hydrophilic, SA average, SA average hydrophobicity, SA average hydrophilicity, SA most hydrophobic hydrophilic distance, SK hydrophobic area, SK hydrophilic area, SK most hydrophobic, SK most hydrophilic, SK average, SK average hydrophobicity, SK average hydrophilicity, SK most hydrophobic hydrophilic distance, polar surface area excluding P and S, polar surface area including P and S

$$Z \text{ score} = \frac{(h-\mu)}{\sigma} \quad (3)$$

where h is the q^2 value calculated for the actual dataset, μ the average q^2, and σ is its standard deviation calculated for various iterations using models built by different random datasets.

2.2. 3D QSAR methodology

Like many 3D QSAR methods (Cramer et al., 1988; Sharaf et al., 1986) k-nearest neighbour molecular field analysis (kNN-MFA) requires suitable alignment of given set of molecules. This is followed by generation of a common rectangular grid around the molecules. The steric and electrostatic interaction energies are computed at the lattice points of the grid using a methyl probe of charge +1. These interaction energy values are considered for relationship generation and utilized as descriptors to decide nearness between molecules. The term descriptor is utilized in the following discussion to indicate field values at the lattice points. The optimal training and test sets were generated using the sphere exclusion algorithm. This algorithm allows the construction of training sets covering descriptor space occupied by representative points. Once the training and test sets were generated, kNN methodology was applied to the descriptors generated over the grid.

2.2.1. k-Nearest neighbour (kNN) method

The kNN methodology relies on a simple distance learning approach whereby an unknown member is classified according to the majority of its k-nearest neighbours in the training set. The nearness is measured by an appropriate distance metric (e.g., a molecular similarity measure calculated using field interactions of molecular structures). The standard kNN method is implemented simply as follows (Sharaf et al., 1986): (1) calculate distances between an unknown object (u) and all the objects in the training set. (2) Select k objects from the training set most similar to object u, according to the calculated distances. (3) Classify object u with the group to which the majority of the k objects belongs. An optimal k value is selected by optimization through the classification of a test set of samples or by leave-one out cross-validation. The variables and optimal k values were chosen using different variable selection methods as described below.

2.2.1.1. kNN-MFA with simulated annealing. Simulated annealing (SA) is the simulation of a physical process, 'annealing', which involves heating the system to a high temperature and then gradually cooling it down to a preset temperature (e.g., room temperature). During this process, the system samples possible configurations distributed according to the Boltzmann distribution so that at equilibrium, low energy states are the most populated. The SA kNN-MFA method employs the kNN classification principle combined with the SA variable selection procedure. For each predefined number of variables (Vn) it seeks to optimize the following using stochastic sampling and simulated annealing as an optimization tool; (i) the number of nearest neighbours (k) used to estimate the activity of each molecule and (ii) the selection of variables from the original pool of all molecular descriptors that are used to calculate similarities between molecules (i.e., distances in Vn-dimensional descriptor space). The SA kNN-MFA reported here is similar to that described by Zheng and Tropsham (2000) and can be summarized as follows:

(1) Generate a trial solution to the underlying optimization problem; i.e., a kNN-MFA model is built based on a random selection of descriptors. (2) Calculate the value of the fitness function, which characterizes the quality of the trial solution to the underlying problem, i.e., the q^2 value for a kNN-MFA model. (3) Perturb the trial solution to obtain a new solution; i.e., change a fraction of the current trial solution descriptors to other randomly selected descriptors and build a new kNN-MFA model for the new trial solution. (4) Calculate the value of the fitness function (q^2_{new}) for the new trial solution. (5) Apply the optimization criteria: if $q^2_{curr} \leqslant q^2_{new}$ the new solution is accepted and used to replace the current trial solution; if $q^2_{curr} > q^2_{new}$, the new solution is accepted only if the Metropolis criterion is satisfied; i.e.

$$\text{rnd} < e^{-(q^2_{curr}-q^2_{new})/T} \quad (4)$$

where rnd is a random number uniformly distributed between 0 and 1 and T is a parameter analogous to the temperature in the Boltzmann distribution. (6) Steps 3–5 are repeated until the termination condition is satisfied. The temperature-lowering scheme and the termination condition used in this work have been adapted from Sun et al. (1994).

Thus, when a new solution is accepted or when a preset number of successive steps of generating trial solutions (20 steps) do not lead to a better result, the temperature is lowered by 10% (the default initial temperature is 1000 K). The calculations are terminated, when either the current temperature of simulations reaches 10^{-6} K or the ratio between the current temperature and the temperature corresponding to the best solution found equals 10^{-6}.

2.2.1.2. kNN-MFA with stepwise (SW) variable selection. This method employs a stepwise variable selection procedure combined with kNN to optimize (i) the number of nearest neighbours (k) and (ii) the selection of variables from the original pool as described in simulated annealing. The step by-step search procedure begins by developing a trial model with a single independent variable and adds independent variables, one step at a time, examining the fit of the model at each step. The method continues until there are no more significant variables remaining outside the model.

2.2.1.3. kNN-MFA with genetic algorithm. Genetic algorithms (GA) first described by Holland (1975) mimic natural evolution and selection. In biological systems, genetic information that determines the individuality of an organism is stored in chromosomes. Chromosomes are replicated and passed onto the next generation with selection criteria depending on fitness. Genetic information can however be altered through genetic operations such as mutation and crossover. In GAs, each "chromosome" is a set of genes, which constitutes a candidate solution to the discrimination problem. A population of "chromosomes" is used. The passage of each "chromosome" to the next generation is determined by its relative fitness, i.e., the closeness of its properties to those desired. Random combinations and/or changes of the transmitted "chromosomes" produce variations in the next generation of "offspring". Better the fitness (correspondence with desired properties), greater is the chance of that chromosome being selected for transmission. Optimal or near optimal solutions are obtained through evolu-

tion over many generations. There are four major components of GA: chromosome generation, fitness assessment, selection, and mutation.

This method employs a stochastic variable selection procedure, combined with kNN, to optimize (i) the number of nearest neighbours (k) and (ii) the selection of variables from the original pool as described in simulated annealing. The implementation of GA based kNN-MFA involved the following steps:

(1) Generate the initial population of chromosomes (candidate solutions) by randomly selecting genes (descriptors) from the pool of available genes. (2) Calculate pairwise Euclidean distances for all pair of molecules with respect to each chromosome. (3) Calculate the fitness of each chromosome using a weighted kNN cross-validation procedure. (4) Select chromosomes for mating pool by roulette wheel selection. (5) Apply uniform crossover and mutation operations on the mating pool chromosomes to create a new population of offspring. (6) Calculate fitness of each offspring using a weighted kNN cross-validation procedure. (7) Replace the least fit chromosomes in an initial population with the best offspring. (8) Repeat steps 2–7 until the convergence criteria or the maximum number of generations is reached.

2.3. Biological activity data

Forty-five 4-anilino quinazoline derivatives having anticancer activities by EGFR kinase inhibition were considered in the present study and used for kNN-MFA analysis (Bridges et al., 1996). kNN-MFA (3DQSAR) models were generated for these derivatives using stepwise variable selection method. The dataset of 45 molecules was divided into the training (32 molecules) and test set (13 molecules) by sphere exclusion method. The experimental biological activities, in the form of IC_{50} were converted into pIC_{50} ($-\log IC_{50}$), where IC_{50} represents the concentration of these compounds that produce 50% kinase inhibition. Our aim is to utilize these activity data for the development of a valid 3D-QSAR model based on steric and electrostatic fields that gives a deep insight into structure property–activity correlations.

2.3.1. Geometry optimization

Three-dimensional quantitative structure–activity relationship studies of 4-anilinoquinazoline derivatives were carried out by using Molecular Design Suite software version 3.5 (Molecular Design Suit 3.5, Vlife Science). Three-dimensional structures of all compounds have been constructed using MDS 3.5 and their geometries were subsequently optimized to make the conformations having least potential energy. Energy minimizations were performed using Merck molecular force field (MMFF) and MMFF charge (Halgren, 1996) followed by considering distance-dependent dielectric constant of 1.0 and convergence criterion of 0.01 kcal/mol. The total energy of a conformation can be calculated using MMFF by the relation.

$$E_{\text{total}} = E_{\text{B}} + E_{\text{A}} + E_{\text{BA}} + E_{\text{OOP}} + E_{\text{T}} + E_{\text{vdw}} + E_{\text{elec}} \quad (5)$$

where,

E_{B} = energy of bond stretching;
E_{A} = energy of angle bending;
E_{BA} = energy of bond stretching and angle bending;
E_{OOP} = out-of-plane bending energy;
E_{T} = torsion energy term;
E_{vdw} = van der Waals energy;
E_{elec} = electrostatic energy.

2.3.2. Alignment of molecules

Molecular alignment is a crucial step in 3D-QSAR study to obtain meaningful results. This method is based on moving of molecules in 3D space, which is related to the conformational flexibility of molecules. The goal is to obtain optimal alignment between the molecular structures necessary for ligand–receptor interactions (Cramer et al., 1988). All molecules in the data set were aligned by template-based method using 4-anilinoquinazoline as template, where a template is built by considering common substructures in the series. A highly bioactive energetically stable conformation in this class of compounds is chosen as a reference molecule on which other molecules in the data set are aligned, considering template as a basis for the alignment.

2.3.3. Computation of steric and electrostatic fields

The aligned biologically active conformations of 4-anilinoquinazolines are used for the calculation of molecular fields. Molecular fields are the steric and electrostatic interaction energies which are used to formulate a relationship between steric and electrostatic properties together with the biological activities of compounds. Each conformation is taken in turn, and the molecular fields around it are calculated. This is done by generating three-dimensional rectangular grids around the molecule and calculating the interaction energy between the molecule and probe group placed at each grid point. Steric and electrostatic fields are computed at each grid point considering MMFF charges (Halgren, 1996). Methyl probe of charge +1 with 10.0 kcal/mole electrostatic and 30.0 kcal/mole steric cutoff were used for fields generation. A value of 1.0 is assigned to the distance-dependent dielectric constant. Steric and electrostatic field descriptors were calculated using Lennard–Jones and Coulomb potentials (Cramer et al., 1988; Hoskuldsson, 1995). In the present study, molecular field analysis coupled with stepwise forward–backward variable was applied to obtain a 3D-QSAR model based on steric and electrostatic descriptors. The calculated steric and electrostatic field descriptors were used as independent variables and pIC_{50} values were used as dependent variables in the present study (Hoskuldsson, 1995) to derive the 3D-QSAR models using MDS software.

2.3.4. Cross-validation using weighted k-nearest neighbour

The standard leave-one-out procedure was implemented as described by Zheng and Tropsham (2000) and can be summarized as follows. (1) A molecule in the training set was eliminated, and its biological activity was predicted as the weighted average activity of the k most similar molecules (Eq. (6)). The similarities were evaluated as the inverse of Euclidean distances between molecules (Eq. (7)) using only the subset of descriptors corresponding to the current trial solution.

$$W_i = \frac{\exp -d_j}{\sum_{\text{k-nearest neighbours}} \exp(-d_j)}$$

$$\hat{y}_i = \sum w_i y_i \quad (6)$$

$$d_{ij} = \left[\sum_{k=1}^{V_n} (X_{i,k} - X_{j,k})^2 \right]^{1/2} \quad (7)$$

(2) Step 1 was repeated until every molecule in the training set has been eliminated and its activity predicted once. (3) The cross-validated r^2 (q^2) value was calculated using Eq. (8), where y_i and $\hat{y}_i$ are the actual and predicted activities of the ith molecule, respectively, and y mean is the average activity of all molecules in the training set. Both summations are over all molecules in the training set. Since the calculation of the pair wise molecular similarities, and hence the predictions, were based upon the current trial solution, the q^2 obtained is indicative of the predictive power of the current kNN-MFA model

$$q^2 = 1 - \frac{\sum(y_i - \hat{y}_i)^2}{\sum(y_i - y_{\text{mean}})^2} \tag{8}$$

(4) Steps 1–3 were repeated for k 2, 3, 4, etc. Formally, the upper limit of k is the total number of molecules in the data set. However, the best value has been empirically found to lie between 1 and 5. The k value that led to the highest q^2 value was chosen for the current kNN-MFA model.

Since the final equations are not very useful to represent efficiently the kNN-MFA models, 3D master grid maps of the best models are displayed. They represent area in space where steric and electrostatic field interactions are responsible for the observed variation of the biological activity.

2.3.4.1. External validation. The following procedure was applied for external validation. (1) Predict the biological activity of a molecule in the test set as the weighted average activity of the k most similar molecules in the training set (Eq. (6)). The similarities were evaluated as the inverse of Euclidean distances between molecules (Eq. (7)) as calculated using the descriptors determined by the current model. (2) Step 1 was repeated for every molecule in the test set. (3) The predicted $r^2_{(\text{pred\_}r^2)}$ value was calculated using Eq. (9), where y_i and $\hat{y}_i$ are the actual and predicted activities of the ith molecule in test set, respectively, and y_{mean} is the average activity of all molecules in the training set. Both summations are over all molecules in the test set. The pred_r^2 value is indicative of the predictive power of the current kNN-MFA model for external test set.

$$\text{pred\_}r^2 = 1 - \frac{\sum(y_i - \hat{y}_i)^2}{\sum(y_i - y_{mean})^2} \tag{9}$$

2.3.4.2. Randomization test. To evaluate the statistical significance of the QSAR model for an actual data set, we have employed a one-tail hypothesis testing (Golbraikh and Tropsha, 2003; Gilbert, 1976). The robustness of the QSAR models for experimental training sets was examined by comparing these models to those derived for random data sets. Random sets were generated by rearranging biological activities of the training set molecules. The significance of the models hence obtained was derived based on calculated Z score (Golbraikh and Tropsha, 2003; Gilbert, 1976).

2.3.5. Statistical parameters for 3D QSAR models

The QSAR models were evaluated using following statistical measures: n, number of observations (molecules); Vn, number of descriptors; k, number of nearest neighbours; q^2, cross-validated r^2 (by the leave-one-out method); pred_r^2, predicted r^2 for the external test set; Z score, the Z score calculated by q^2 in the randomization test; best_ran_q^2, the highest q^2 value in the randomization test.

2.4. Molecular docking studies

Piecewise linear pairwise potential (PLP)-based molecular docking of 4-anilino quinazoline has been performed using the docking module of Molecular Design Software, (Gehlhar et al., 1995; Verkhivker et al., 2000) which involves the use of the PLP function summed over energy interactions between all pairs of protein and ligand atoms.

Molecular docking energy evaluations are usually carried out with the help of scoring function. There are several scoring functions such as dock score, PLP score, potential of mean

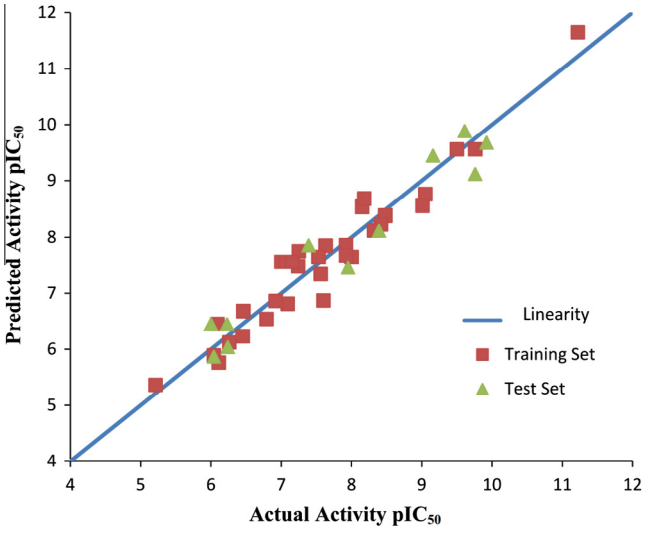


Figure 1 Graph of actual vs. predicted activities for training and test set molecules by multiple linear regression model (MLR).

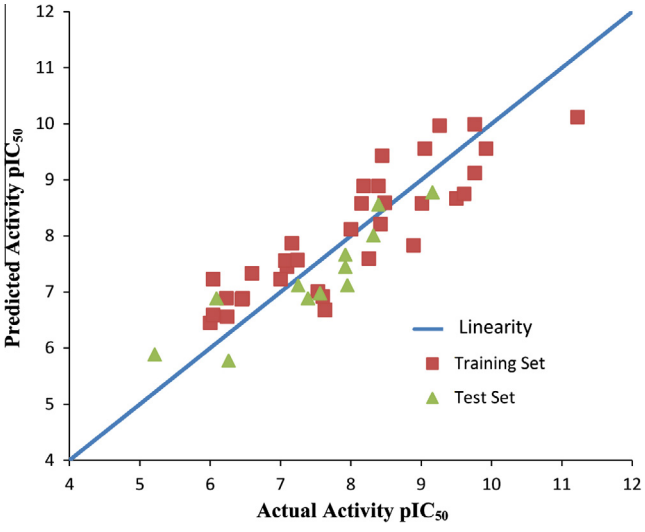


Figure 2 Graph of actual vs. predicted activities for training and test set molecules by principal component regression model (PCR).

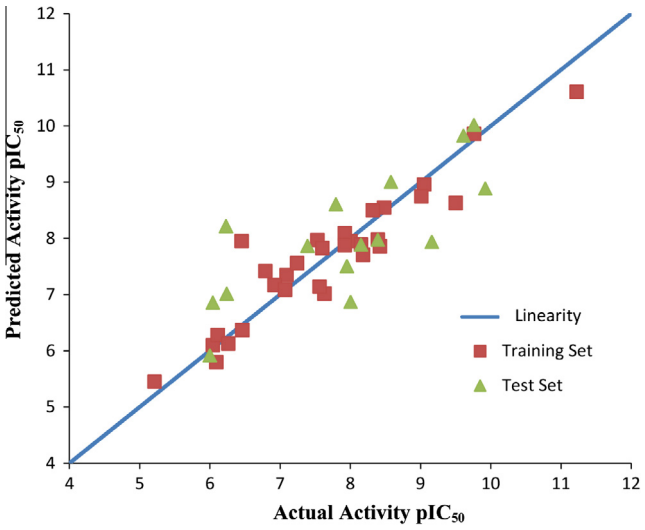


Figure 3 Graph of actual vs. predicted activities for training and test set molecules by partial least square regression model (PLS).

Table 4 Statistical parameters of MLR, PCR and PLS by 2D QSAR.

Parameters	MLR	PCR	PLS
N	34	33	30
Df	28	27	25
r^2	0.9212	0.7837	0.7622
q^2	0.8290	0.6905	0.6833
F test	52.1234	30.4505	40.0549
r^2 se	0.6879	0.6726	0.6091
q^2 se	0.7198	0.8045	0.7029
pred_r^2	0.6170	0.5623	0.5736
best_ran_r^2	0.62099	0.47789	0.4523
best_ran_q^2	0.32287	0.22874	0.2576
Z score_ran_r^2	6.66361	8.40938	8.18328
Z score_ran_q^2	6.32934	7.66924	6.90778
α_ran_r^2	0.00000	0.00000	0.00000
α_ran_q^2	0.00000	0.00000	0.00000

MLR = multiple linear regression, PCR = principal component regression, PLS = partial least squares, N = number of molecules of training set, Df = degree of freedom, r^2 = coefficient of determination, q^2 = cross-validated r^2, pred_r^2 = r^2 for external test set, Z score = the Z score calculated by q^2 in the randomization test, best_ran_q^2 = the highest q^2 value in the randomization test and α_ran_q^2 = the statistical significance parameter obtained by the randomization test.

force (PMF) score, steric and electrostatic score, etc. For our interest, the energy of interactions between ligand and protein was calculated in terms of PLP score, which depends upon the following different atom type parameter: hydrogen-bond donors, and hydrogen bond acceptors, both donors and acceptors, and non polar atoms such as carbon.

The PLP function is incorporated by the MDS Vlife Science software in the GRIP docking method which calculates the ligand-receptor binding affinity in terms of the PLP score. The PLP score is designed to enable flexible docking of ligands to perform a full conformational and positional search within a rigid binding site. All the optimized conformers were docked into active binding site of EGFR target protein that can be obtained in co-crystallized state with erlotinib (protein data bank, PDB entry 1M17), which was considered as the reference to define the active binding site in the present investigation (Stamos et al., 2002). Water molecules and HET ATOM-like bound ligand data were removed from the PDB file of EGFR protein during docking study. The crystal structure was refined using Vlife Science's MDS 3.0 software (Vlife MDS 3.0, 2007a). The refinement of the crude PDB structure of receptor was done by completing the incomplete residues. The co-crystallized ligand lying within the receptor was modified by assigning missing bond order and hybridization states. The side chain hydrogens were then added to the crystal structure and their positions were optimized up to the rms gradient 1 by aggregating the other part of the receptor. The optimized receptor was then saved as mol file and used for docking simulation. The 2D structure of the compounds were built and then converted into the 3D with the help of Vlife MDS 3.0 software (Vlife MDS 3.0, 2007b). The 3D structures were then

Table 3 Molecular descriptors contributing in the present study.

Descriptor	Description
Individual	
H-donor count	This descriptor signifies number of hydrogen bond donor atoms
XlogP	This descriptor signifies ratio of solute concentration in octanol & water and generally termed as octanol water partition coefficient
Rotatable bond count	This descriptor signifies number of rotatable bonds
Molecular weight	This descriptor signifies the molecular weight of the compound
Alignment independent descriptors	
T_C_Br_1	This is the count of number of carbon atoms separated from bromine atoms by one bond distance
T_2_N_7	This is the count of number of double bonded atoms separated from nitrogen atoms by seven bond distance
T_2_O_1	This is the count of number of double bonded atoms separated from oxygen atoms by one bond distance
T_C_N_1	This is the count of number of carbon atoms separated from nitrogen atoms by one bond distance
T_C_O_1	This is the count of number of carbon atoms separated from oxygen atoms by one bond distance

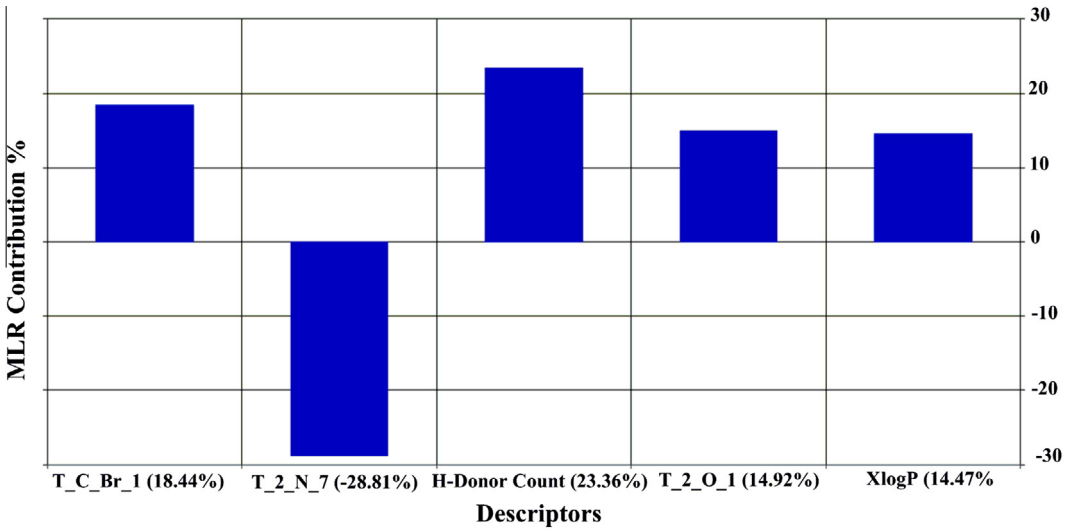


Figure 4 Plot of percentage contribution of each descriptor in developed MLR model explaining variation in the activity.

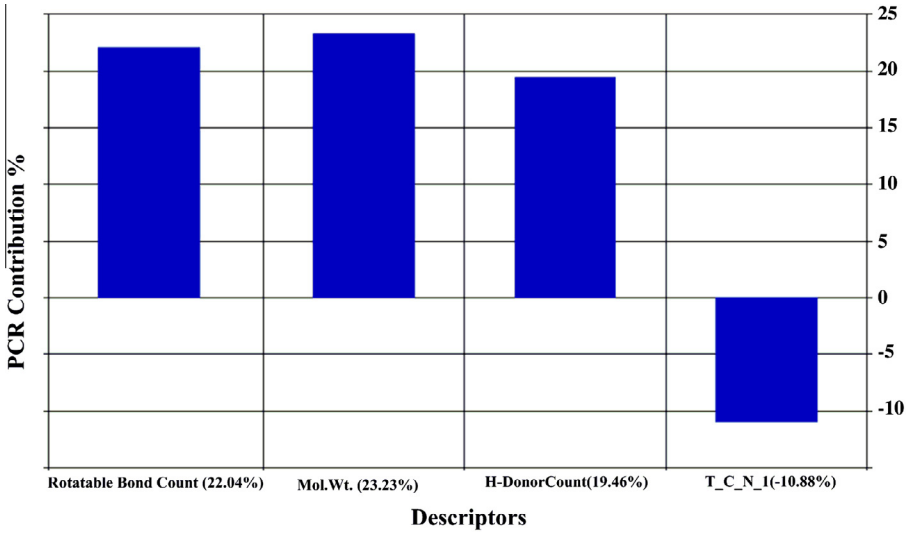


Figure 5 Plot of percentage contribution of each descriptor in developed PCR model explaining variation in the activity.

energetically minimized up to the rms gradient of 0.01 using Merck Molecular Force Field (MMFF) (Vlife MDS 3.0, 2007c). Conformers of compounds were then generated by Monte Carlo method. In doing so, all rotatable bonds of the ligand were selected and number of seeds used for searching the conformational space was set 5. All the conformers were then energetically minimized up to the rms gradient of 0.01 and then saved in separate folder (Vlife MDS 3.0, 2007d). The active site selection was done by choosing the cavity having maximum hydrophobic surface area. The docking simulation was done using GRIP batch docking. In this, all generated conformers of one ligand were put as one batch in GRIP docking wizard. Likewise, the batches for all other ligands were put. All the confirmers were virtually docked at the defined cavity of the receptor. The parameters fixed for docking simulation was like this-number of placement: 50, rotation angle of: 10°, exhaustive method, scoring function: dock score. By rotation angle, ligand would be rotated inside the receptor cavity to generate different ligand poses inside the receptor cavity. By placements, the method will check all the 50 possible placements into the active site pocket and will result out few best placements out of 50. For each ligand, all the conformers with their best placements and their dock score will be saved in output folder. The method also highlights the best placements of best conformer of one particular ligand which is having best (minimum) dock score. The ligand forming most stable drug-receptor complex is the one which is having minimum dock score. After docking simulation, the best docked conformer of each ligand and receptor were merged and their complex was then energetically optimized by defining radius of 10 Å measured from the docked ligand. Stepwise energy optimization was done by first hydrogens; second side chains and finally the backbone of receptor (Vlife MDS 3.0, 2007e). The optimized complexes were then checked for various interaction of ligand with receptor like hydrogen bonding, hydrophobic bonding and van der Waal's interaction. The binding affinity was evaluated by the binding free energy (ΔGb, kcal/mol), hydrogen bonding interaction, hydrophobic interaction and

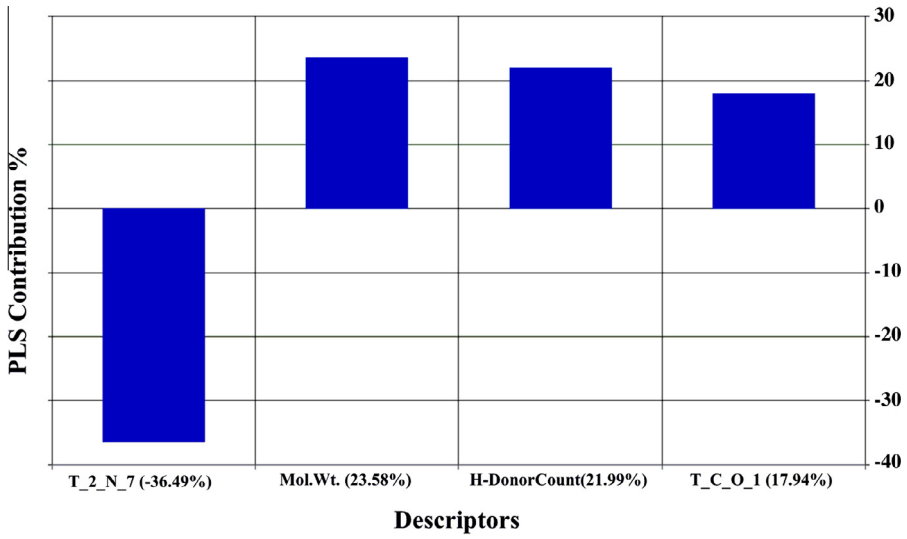


Figure 6 Plot of percentage contribution of each descriptor in developed PLS model explaining variation in the activity.

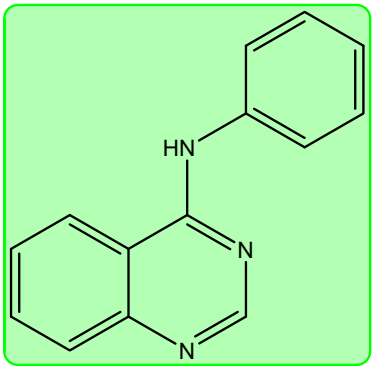


Figure 7 4-Anilinoquinazoline (template).

RMSD values. Docking interaction of synthesized compounds is discussed in Section 3.

3. Result and discussion

3.1. 2D QSAR modelling and its validation

With regard to QSAR modelling, our first goal was to establish a predictive model with a reasonable number of input features to ensure good generalization performance. While correlating various descriptors with biological activity is the most important means to study structure–activity relationships, the interest lies in deciding when to stop adding a new descriptor to the model. Thus, the optimal model should use the minimum number of descriptors to obtain the best fit. To achieve this, a well accepted method is to find out the saturation point, a point beyond which there is no considerable improvement in the regression coefficient (r^2 and q^2) values even if a new descriptor is added. MLR, PCR and PLS techniques were used in the present study for selecting a significant set of descriptors in order to build the significant models. In this section, the prediction performances of the method proposed by three different models (MLR, PCR and PLS) were evaluated.

The MLR, PCR and PLS models predicted the training data with a r^2 of 0.9212, 0.7837 and 0.7622 together with q^2 estimating to 0.8290, 0.6905 and 0.6833, respectively. The plots of calculated versus observed values of are shown in Figs. 1–3 in the case of all three models. Eq. (10) appears to be the best QSAR model obtained by the multiple linear regression (MLR) analysis. In this study we have limited the number of presented equations to this of the best regression model of the whole set. The model is given as follows together with the statistical and validation parameters.

$$\begin{aligned} \mathrm{pIC}_{50} = {} & 0.2354(\mathrm{T\_C\_Br\_1}) + 1.0187(\text{H-donor count}) \\ & + 0.1785(\mathrm{T\_2\_O\_1}) - 0.4166(\mathrm{T\_2\_N\_7}) \\ & + 0.6829(\mathrm{XlogP}) - 3.5517. \end{aligned} \tag{10}$$

$N = 34$, Df = 28, $r^2 = 0.9212$, $q^2 = 0.8290$, $F = 52.1234$, pred_$r^2 = 0.6170$, α_ran_$r^2 = 0.00000$, α_ran_$q^2 = 0.00000$, best_ran_$r^2 = 0.62099$, best_ran_$q^2 = 0.32287$, Z score_ran_$r^2 = 6.66361$ and Z score_ran_$q^2 = 6.32934$.

Eq. (10) shows the positive contribution of T_C_Br_1 (18.44%) (This is the count of number of carbon atom separated from bromine atom by one bond distance) and XlogP (14.47%) (This descriptor signifies ratio of solute concentration in octanol and water and generally termed as Octanol–Water partition coefficient) indicating the importance of electron withdrawing group at 4-anilino portion of quinazoline ring (compounds **7**–**12** and similar analogues).

The other alignment independent descriptor, i.e., T_2_N_7 (−28.81%) is inversely proportional, which indicates that the increase in the number of bonds between NH_2, NH-, NHCO-OCH_3, $NHCH_3$, NHC_2H_5 groups and quinazoline ring at C-6 (R3) and C-7 (R4) diminish the activity. This diminishing influence is augmented by minimizing chain length between NHCO-OCH_3, $NHCH_3$, NHC_2H_5 groups and quinazoline ring at C-6 (R3) and C-7 (R4) (compounds **28**–**30** and similar analogues).

The direct relationship of hydrogen donor count descriptor (23.36%) reveals the importance of hydrogen bonding for selective potent molecule design.

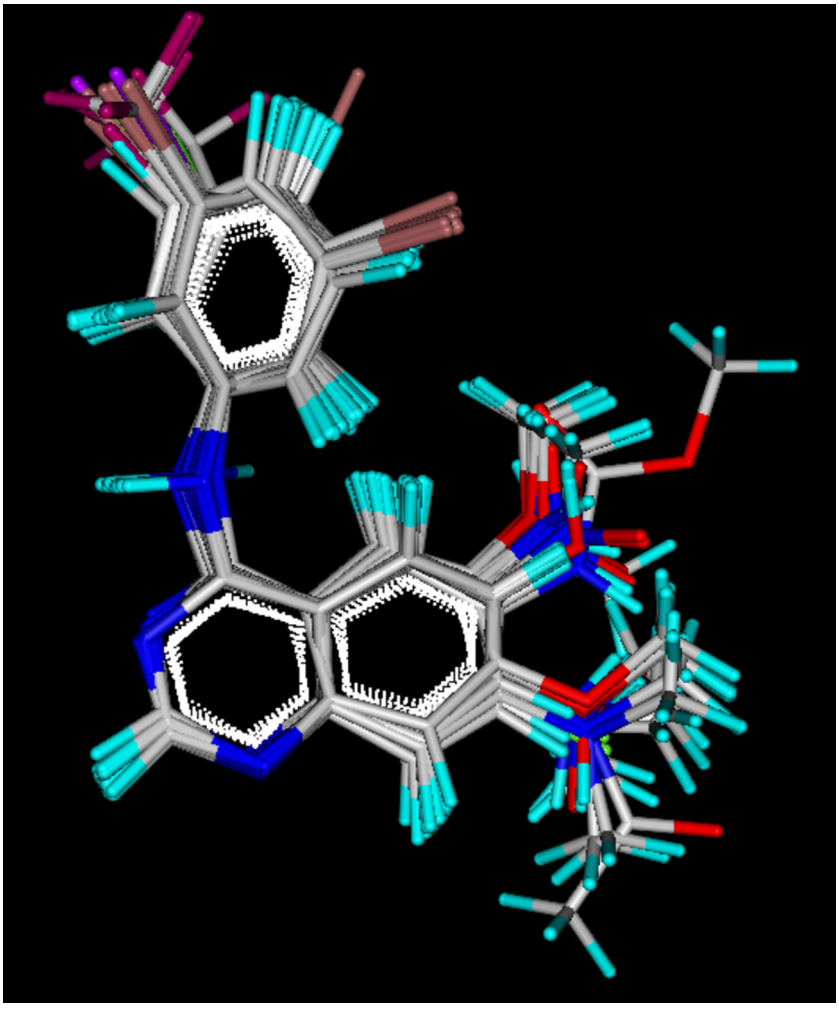


Figure 8 3D view of aligned molecules.

The positive contribution of next important alignment independent descriptor T_2_O_1 (14.92%), which represents number of double bonded atoms separated from oxygen atom by one bond distance is directly proportional to the activity. It reveals that presence of alkoxy group such as OCH_3, OC_2H_5 at C-6 (R_3) and C-7 (R_4) position of quinazoline is favourable for the activity (compounds **23**–**27** and similar analogues).

The models obtained by PCR methodology also give a good prediction in terms of the r^2 value and is denoted by Eq. (11).

$$\begin{aligned} pIC_{50} = {} & 0.3544(\text{Rotatable bond count}) \\ & + 0.7857(\text{H-donor count}) \\ & + 0.0120(\text{molecular weight}) \\ & - 0.3263(\text{T\_C\_N\_1}) + 7.7591. \end{aligned} \tag{11}$$

$N = 33$, Df = 27, $r^2 = 0.7837$, $q^2 = 0.6905$, $F = 30.4505$, pred_$r^2 = 0.5623$, α_ran_$r^2 = 0.00000$, α_ran_$q^2 = 0.00000$, best_ran_$r^2 = 0.47789$, best_ran_$q^2 = 0.22874$, Z score_ran_$r^2 = 8.40938$ and Z score_ran_$q^2 = 7.66924$.

In the above equation, Rotatable bond count (22.04%) and molecular weight (23.23%) descriptors have a positive influence on the EGFR inhibitory activity values of the compounds.

The next important alignment independent descriptor is T_C_N_1 (−10.88%), which can be defined as the count of number of carbon atom separated from other nitrogen atom by one bond distance in a molecule. It reveals that NH_2, NH–, $NHCOOCH_3$, $NHCH_3$, NHC_2H_5 groups should not be directly attached with quinazoline ring at C-6 (R3) and C-7 (R4).

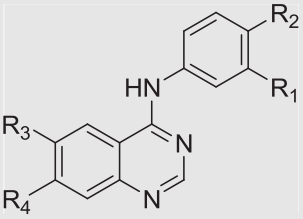
The models obtained on the same 4-anilino quinazolines by PLS regression analysis is given below as Eq. (12) along with the corresponding statistical parameters.

$$\begin{aligned} pIC_{50} = {} & 0.8304(\text{H-donor count}) \\ & + 0.0114(\text{molecular weight}) \\ & + 0.2403(\text{T\_C\_O\_1}) - 0.4861(\text{T\_2\_N\_7}) \\ & + 4.0271. \end{aligned} \tag{12}$$

$N = 30$, Df = 25, $r^2 = 0.7622$, $q^2 = 0.6833$, $F = 40.0549$, pred_$r^2 = 0.5736$, α_ran_$r^2 = 0.00000$, α_ran_$q^2 = 0.00000$, best_ran_$r^2 = 0.4523$, best_ran_$q^2 = 0.2576$, Z score_ran_$r^2 = 8.18328$ and Z score_ran_$q^2 = 6.90778$.

Alignment independent descriptor T_2_N_7 is common between MLR and PLS while as molecular weight descriptor is common between PCR and PLS; only differs from each other in their percentage of contribution. The definition of the

Table 5 Structure, experimental and predicted activity of quinazolines used in training and test set using kNN-MFA 3D QSAR model.



S. No.	R_1	r^2	R_3	R_4	pIC_{50}[a]		Residual
					Exp.	Pred.	
1	H	H	H	H	6.46	6.78	−0.32
2	CH_3	H	H	H	6.04	6.23	−0.19
3	Cl	H	H	H	7.63	6.88	0.75
4	Br	H	H	H	7.56	7.34	0.22
5[T]	I	H	H	H	7.09	7.24	0.66
6	CF_3	H	H	H	6.23	6.56	−0.33
7	Br	H	NO_2	H	6.04	6.21	−0.17
8	Br	H	OCH_3	H	6.45	6.31	0.14
9	Br	H	H	NO_2	6.00	5.78	0.22
10[T]	Br	H	H	OCH_3	8.00	7.76	0.24
11	Br	H	OH	OH	9.76	9.64	0.12
12[T]	Br	H	NH_2	NH_2	9.92	9.45	0.47
13	F	H	H	H	7.25	7.45	−0.2
14	H	H	OCH_3	H	7.24	7.01	0.23
15	H	H	NH_2	H	6.11	6.45	−0.34
16[T]	CF_3	H	NH_2	H	6.24	6.58	−0.34
17	H	H	OCH_3	H	6.92	6.71	0.21
18	H	H	H	NH_2	7.00	6.62	0.38
19[T]	CF_3	H	H	NH_2	8.48	8.89	−0.41
20	F	H	H	NO_2	5.21	5.68	−0.47
21	Cl	H	H	NO_2	6.09	6.14	−0.05
22	I	H	H	NO_2	6.26	6.43	−0.17
23	H	H	OCH_3	OCH_3	7.53	7.68	−0.15
24	F	H	OCH_3	OCH_3	8.42	8.78	−0.36
25[T]	Cl	H	OCH_3	OCH_3	9.50	9.21	−9.71
26[T]	I	H	OCH_3	OCH_3	9.05	9.25	−0.2
27	CF_3	H	OCH_3	OCH_3	9.61	9.21	0.4
28	Br	H	$NHCH_3$	H	8.39	8.21	0.18
29	Br	H	$N(CH_3)_2$	H	7.07	7.12	−0.05
30[T]	Br	H	$NHCOOCH_3$	H	7.92	7.38	0.54
31	Br	H	H	OH	8.32	7.89	0.43
32	Br	H	H	OH	7.39	7.89	−0.5
33[T]	Br	H	H	$NHCOCH_3$	8.15	7.67	0.48
34	Br	H	H	$NHCH_3$	7.92	7.12	0.8
35	Br	H	H	NHC_2H_5	7.95	7.45	0.5
36[T]	Br	H	H	$N(CH_3)_2$	9.16	9.89	−0.73
37[T]	Br	H	NH_2	$NHCH_3$	6.79	5.34	1.45
38	Br	H	NH_2	$N(CH_3)_2$	7.95	7.34	0.61
39[T]	Br	H	NH_2	OCH_3	8.18	8.67	−0.49
40	Br	H	NH_2	Cl	7.16	6.85	0.31
41	Br	H	NO_2	$NHCH_3$	7.82	7.36	0.46
42	Br	H	NO_2	OCH_3	7.60	6.86	0.74
43	Br	H	OC_2H_5	OC_2H_5	11.22	11.56	−0.34
44[T]	Br	$O(CH_2CH_2CH_3)$	$O(CH_2CH_2CH_3)$		9.76	8.45	1.31
45	H	Br	OCH_3	OCH_3	9.01	8.34	0.67

Expt. = experimental activity, Pred. = predicted activity, T = test set.
[a] $-\text{Log}\,(IC_{50} * 10^{-6})$.

remaining descriptor that was found to be dominating in the developed QSAR models is given below.

The positive contribution of next alignment independent descriptor T_C_O_1 (17.94%), which represents number of

carbon atom separated from oxygen atom by one bond distance is directly proportional to the activity. It reveals that presence of alkoxy group such as OCH_3, OC_2H_5 at C-6 (R_3) and C-7 (R_4) position of quinazoline is favourable for the activity (compounds **23**–**27** and similar analogues).

Also it can be observed that all the developed QSAR models share one descriptor in common, i.e. H-donor count, showing the importance of this descriptor in governing the activity.

Among these three models, multiple linear regression (MLR) has come out with a very promising result as compared to other two methods. The observed and predicted pIC_{50} along with residual values are shown in Table 1, list of 2D descriptor in Table 2, list of descriptor contributing in the present study Table 3 and statistical data is shown in Table 4. The descriptors which contribute to the activity by MLR, PLS and PCR are shown in Figs. 4–6 respectively.

3.2. 3D QSAR modelling and its validation

In the present study, kNN-MFA model is developed coupled with stepwise variable selection method to develop 3D-QSAR models of 4-anilinoquinazoline derivatives based on steric and electrostatic fields. The structure of 4-anilinoquinazoline template is shown in Fig. 7. A highly bioactive energetically stable conformation in this class of compounds is chosen as a reference molecule on which other molecules in the data set are aligned, considering template as a basis for the alignment. The aligned view of 4-anilinoquinazolines is presented in Fig. 8. The total data set was divided into training and test sets using the sphere exclusion algorithm for diversity of the sampling procedure. Compounds marked with (T) in Table 5 were selected as test set molecules. The quality of the model was assessed by cross-validated q^2 in the training set and external validation was performed by calculating predictive r^2 (Pred_r^2) from the test-set compounds.

During the kNN-MFA investigation, dissimilarity value for the selection of training and test by spherical exclusion method of 3.0000–5.3000 were investigated. For each dissimilarity value, kNN-MFA model was built. The dissimilarity value of 3.600 produced a significant result as compared to the other. In this study we have limited the number of presented equations to this of the best regression model of the whole set. The kNN-MFA model showed significant correlation coefficient q^2 (r^2) of 0.8123, , coefficient of correlation of predicted data set (pred_r^2se) of 0.7912, r^2 for external test set (pred_r^2) 0.6243, degree of freedom 25 and k-nearest neighbour of 2 proved a good conventional statistical correlation which have been obtained.

S_192, E_188, E_151, E_288 and E_159 are the steric and electrostatic field energy of interactions between probe (CH_3) and compounds at their corresponding spatial grid points of 192, 188, 151, 288 and 159. The above model is validated by predicting the biological activities of the test molecules, as indicated in Table 5.

The plot of observed versus predicted activities for the test compounds is represented in Fig. 9. From Table 5 it is evident that the predicted activities of all the compounds in the test set are in good agreement with their corresponding experimental activities and optimal fit is obtained. The external predictability of the above 3D-QSAR model using the test set was determined by Pred_r^2, which is 0.7243. So the above results indicate that 3D-QSARmodel for EGFR generates 72.43% external model prediction (Table 6).

3.2.1. Steric and electrostatic contribution plot

The plot of contributions of steric and electrostatic field interactions (Fig. 10) indicates relative regions of the local fields (steric and electrostatic) around the aligned molecules (Abraham et al., 1993). Green and blue balls represent steric and electrostatic field effects, respectively. In the QSAR model, steric descriptors with positive coefficients represent regions of high steric tolerance; bulky substituent is favourable in this region. Steric descriptors with negative coefficients indicate

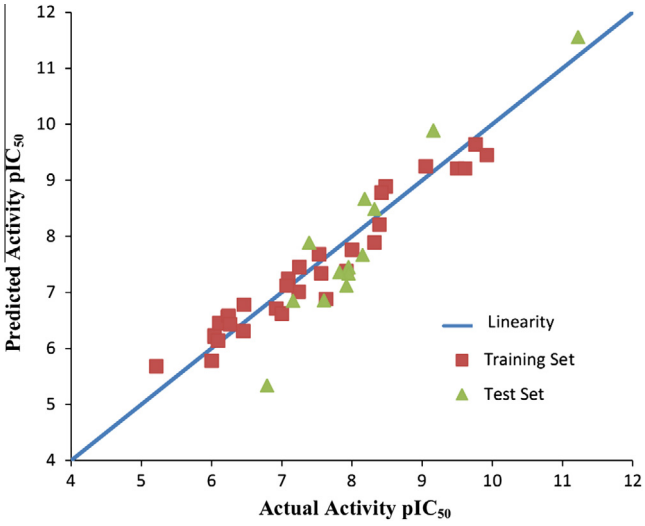


Figure 9 Graph of actual vs. predicted activities for training and test set molecules by kNN-MFA model.

Table 6 Stastical results of kNN-MFA method.

Parameters	kNN-MFA model
N	33
k	2
q^2	0.8123
pred_r^2	0.7243
pred_r^2se	0.7912
Z score	8.3245
best_ran_q^2	0.51099
α_ran_q^2	0.00000
Descriptors	E_151 (0.6423, 2.1154)
	E_159 (3.7094, 4.0670)
	E_188 (−1.1273, −1.1411)
	S_192 (30.0000, 30.0000)
	E_288 (10.0000, 10.0000)
Vn	5

N, number of molecules in training set; Vn, number of descriptors; k, number of nearest neighbours; q^2, cross-validated r^2 (by the leave-one out method); pred_r^2, predicted r^2 for the external test set; Z score, the Z score calculated by q^2 in the randomization test; best_ran_q^2, the highest q^2 value in the randomization test and α_ran_q^2, the statistical significance parameter obtained by the randomization test.

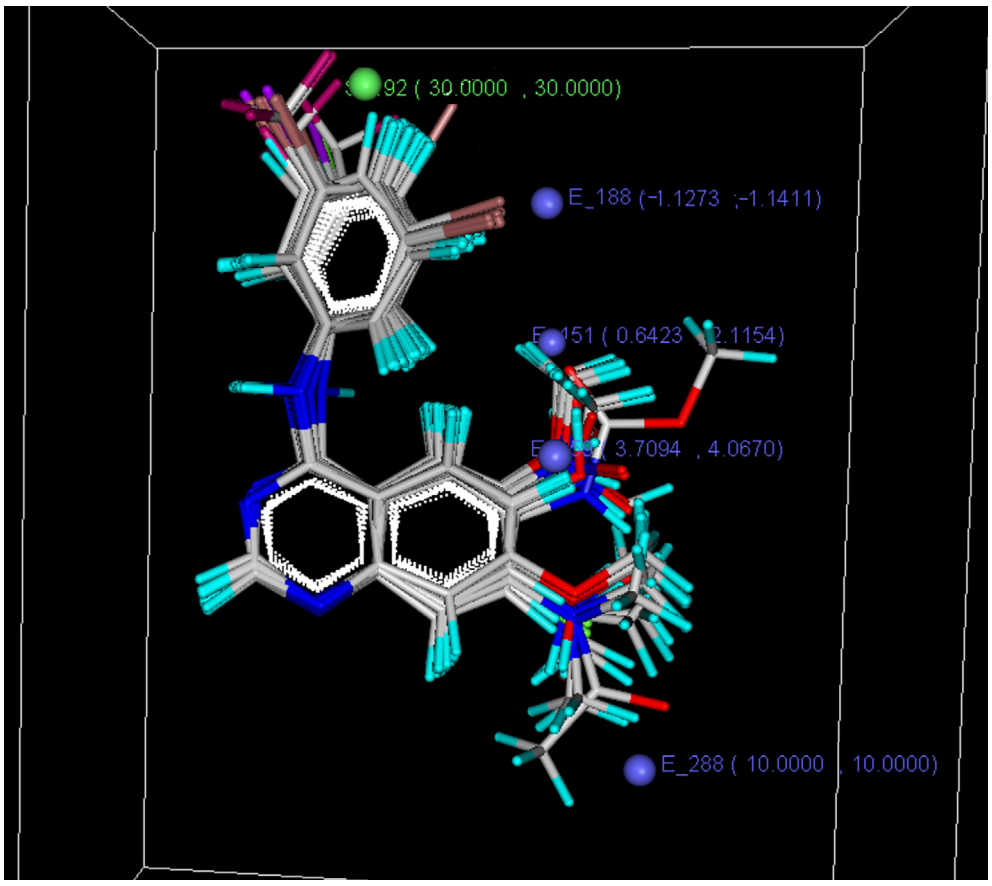


Figure 10 Contribution plot steric and electrostatic field of interactions by kNN-MFA models.

regions where bulky substituent is disfavoured. Electrostatic field descriptors with positive coefficients represent regions where electropositive (electron-donating) groups are favorable, whereas negative coefficient indicates that electronegative (electron-withdrawing) groups are favourable in this region (Nandi and Bagchi, 2009).

From 3D-QSAR kNN-MFA model (Table 6) and Fig. 10 it is observed that electrostatic field with negative coefficient (E_188) on the anilino moiety, indicating that electronegative groups are favourable on this site and presence of electronegative groups increase the activity of 4-anilinoquinazoline compounds (compounds **7**–**12** and similar analogue). Electrostatic descriptor with positive coefficient (E_151, E_159 and E_288) around 6 and 7-position of the quinazoline ring corroborates that electropositive (electron-donating) group is preferred at 7-position of quinazolines. These results are in close agreement with the experimental observations that compounds **23**–**27** contain alkoxy group at 6- and 7-positions. These compounds produce greater activity due to electropositive substituents on the 6- and 7-positions of the quinazoline ring (Bridges et al., 1996). A bulky aromatic anilino substituent is essential at 4-position of the quinazoline ring for producing kinase inhibition, as indicated by the presence of steric field with positive coefficient (S_192) in this region. It is inferred from the docking results that the 4-anilino moiety is located in a deep hydrophobic pocket formed by ASP 831, LYS 721, THR 830, VAL 702 and LEU 820. Thus, the contribution plot arising out of 3D-QSAR studies provide some useful insights for better understanding of the structural features of these compounds responsible for producing significant EGFR kinase inhibitory activity, which conforms with the docking results.

3.3. Docking study

From the docking study, it is clear that the quinazoline ring is surrounded by hydrophobic residues, as indicated in Table 7. The anilino group substituted at the 4 position of quinazoline ring and itself quinazoline ring of natural ligand Erlotinib (AQ499A) is bounded by hydrophobic pockets consisting of residue such as ASP 831, LYS 721,THR 830,VAL 702 and MET 769, LEU 768, ALA 719, THR 766, LEU 820,GLY 772, PHE 771, respectively (Fig. 11A and B).

The docked models reveal that N-1 of the quinazoline forms hydrogen bond with hydrogen atom of amino backbone of MET-769. The hydrogen bonding distances for natural ligand (Erlotinib) and highly active compound **43** are 1.677 Å and 2.339 Å respectively as shown in (Fig. 11C and D). The hydrogen bonding distance for remaining compounds are shown in Table 7. The quinazoline ring plays a crucial role for producing biological activity by interacting with MET 769, an important active residue for binding affinity of the inhibitor, which correlates with the results obtained from crystallographic study of erlotinib–EGFR (Stamos et al., 2002). These interactions underscore the importance of nitrogen

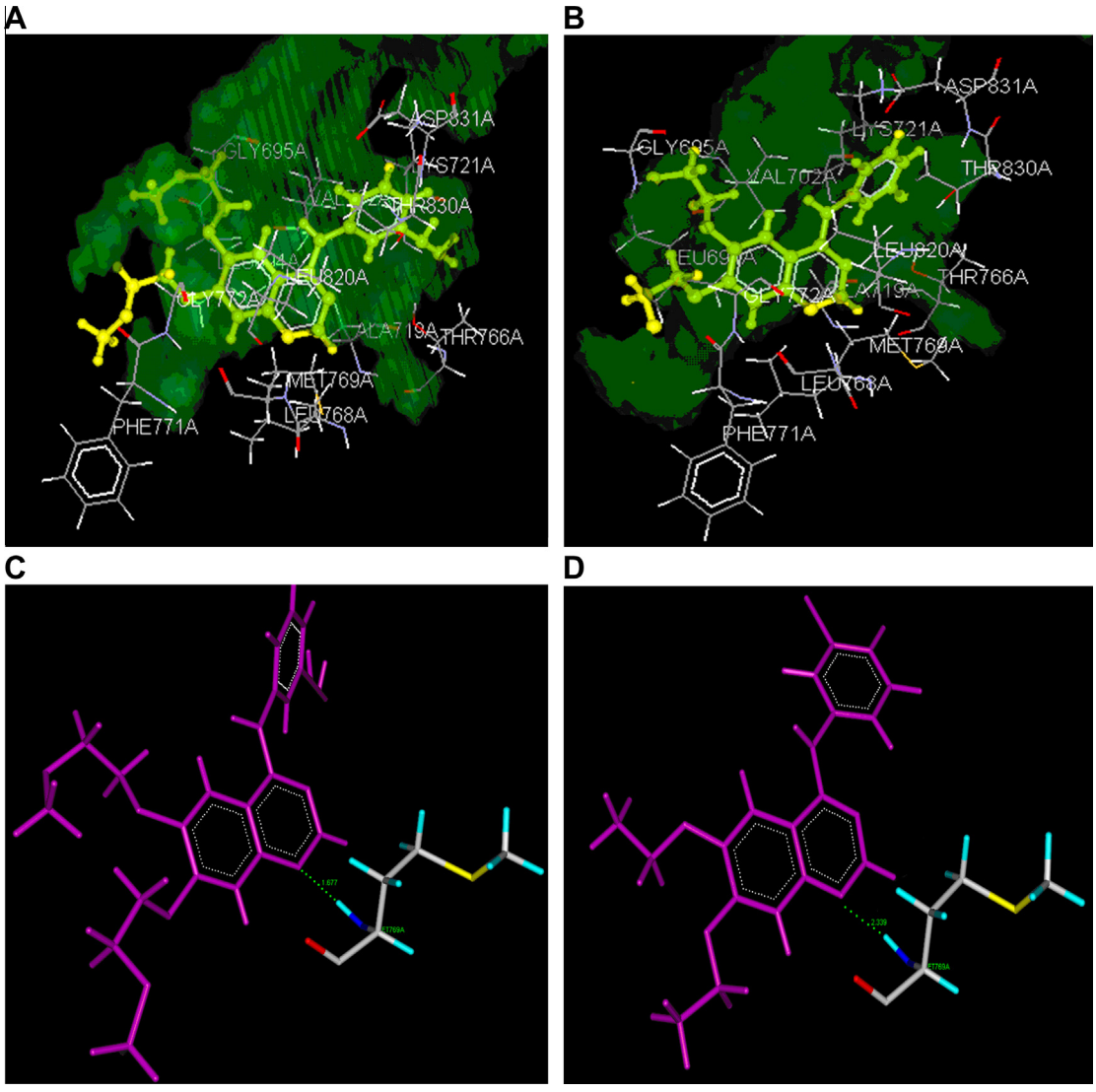


Figure 11 Docking interaction of natural ligand (Erlotinib) and highly active compound **43**; where images [A (natural ligand (Erlotinib)] and [B (highly active compound **43**)] shows hydrophobic interaction of natural ligand (Erlotinib) and highly active compound **43** with the active binding sites of EGFR, represented by molecular surface; the bound ligands (Erlotinib and highly active compound **43**) are represented as stick models (green colour). The residues within 5 Å of the inhibitor are displayed. In images [C (natural ligand (Erlotinib)] and [D (highly active compound **43**)], dotted line represents H-bonding between N-1 of quinazoline ring of Erlotinib (1.677 Å) and highly active compound **43** (2.339 Å) with hydrogen atom of amino backbone of MET-769 (stick model).

atoms for binding and subsequent inhibitory capacity. The minimum PLP score of −73.45 kcal/mol with root mean square standard deviation (RMSD) and binding free energy (ΔG_b) of 3.21 Å and −15.05 kcal/mol, respectively, for compound **43** indicates high binding affinity of the ligand towards EGFR. For compound **43**, the methylene carbons of ethoxy group at 6 and 7 position of the quinazoline produce strong hydrophobic interaction with GLY 695 and LEU 694. The moderately active compound **24** produces good PLP score of −68.26 kcal/mol with root mean square standard deviation (RMSD) 5.73 Å, but it is less active than compound **43** since the methoxy group at 6 and 7 position of quinazoline is associated with hydrophobic interaction with GLY 695 having lower hydropathy indice, which may decrease the activity of the compound. Compound **20** shows poor affinity towards EGFR, as denoted by PLP score of −51.23 kcal/mol and there are no hydrophobic interactions at 6 and 7 positions due to presence of $-NO_2$ as deactivating group.

4. Concluding remarks

In the present 2D QSAR investigation, all proposed QSAR models were statistically significant. However Model-1 by multiple linear regression analysis could be considered as best one in terms of excellent internal and external predictive abilities. According to Model-1 (MLR) anticancer activity of quinazoline derivatives was influenced by individual (H-donor count, and XlogP) and alignment independent descriptor (T_C_Br_1, T_2_O_1 and T_2_N_7) help in understanding the effect of sub-

Table 7 Docking results based on hydrogen bonding, hydrophobic interaction, PLP dock score, binding free energy and root mean square deviation (RMSD).

Compounds	Hydrogen bonding distance between N-1 of quinazoline and H atom of amino acid backbone of MET 769 (Å)	Hydrophobic Interaction (within 5 Å)			PLP dock score (kcal/mol)	ΔG_b (binding free energy calculated by docking study (kcal/mol)	RMSD (root mean square standard deviation) (Å)
		Anilino moiety	Quinazoline ring	Substituent attached to 6 and 7-position of quinazoline			
Erlotinib (lead)	1.677	ASP 831 LYS 721 THR 830 VAL 702	MET 769 LEU 694 GLY 772 LEU 768 ALA 719 LEU 820 THR 766 VAL 702	GLY 695 LEU 694 PHE 771 GLY 772	−88.32	−18.43	0.25
Highly active comp. **43**	2.339	ASP 831 LYS 721 THR 830	MET 769 LEU 694 GLY 772 LEU 768 ALA 719 LEU 820 THR 766 VAL 702 PHE 771	GLY 695 LEU 694	−73.45	−15.05	3.21
Moderately active comp. **24**	2.562	ASP 831 LYS 721 THR 830	MET 769 LEU 694 GLY 772 LEU 768 ALA 719 LEU 820 THR 766 VAL 702	GLY 695	−68.26	−12.58	5.73
Low active comp. **20**	3.21	ASP 831 LYS 721 THR 830	LEU 694 GLY 772 LEU 768 ALA 719 LEU 820	No hydrophobic interaction due to presence of $-NO_2$ as deactivating group	−51.23	−8.34	9.38

stituent at different position of quinazolines. The result obtained from 2D QSAR study suggests that electron withdrawing group on aniline portion of quinazoline ring enhances the lipophilicity of compounds and favors the EGFR inhibition. It also suggests that long chain alkoxy group at 6 and 7 position of quinazoline ring favours the activity. The contribution plot of steric and electrostatic field interactions generated by 3D-QSAR shows that electron withdrawing groups at aniline moiety are favorable. This finding is in close agreement with the structures of these compounds, where presence of electron withdrawing groups is found in the anilino moiety. It also suggests that bulky electron-donating groups are favourable at 6- and 7-position of the template. This finding supports the experimental observations, where presence of bulky electronegative groups at 6 and 7-position signifies increase in activities of compounds. From the molecular docking studies, it is evident that hydrophobic groups substituted at 6- and 7-positions of the quinazoline ring possessing strong hydrophobic interactions with nonpolar active residues are likely to enhance EGFR kinase inhibition. The quinazoline ring plays a crucial role for producing biological activity by interacting with MET 769, an important active residue for binding affinity of the inhibitor, which correlates with the results obtained from crystallographic study of erlotinib–EGFR. These interactions underscore the importance of nitrogen atoms for binding and subsequent inhibitory capacity. The present study is more versatile than the earlier reported methods. Hence the model proposed in this work can be employed to design new derivatives of quinazoline with specific tyrosine kinase (EGFR) inhibitory activity.

Acknowledgements

The authors would like to thank the Director General, Department of Science and Technology, New Delhi for funding the project (Grant No. SR/FT/LS-0083/2008), Chairman, Captain

M.P. Singh and Sardar Sangat Singh Longia, Secretary AS-BASJSM College of Pharmacy for providing necessary facilities.

References

Abraham, D.J., Kellogg, G.E., Kubinyi, H., 1993. 3D QSAR in Drug Design: Theory, Methods and Applications. ESCOM, Leiden.

Badiger, A.M., Noolvi, M.N., Nayak, P.V., 2006. QSAR study of benzthiazole derivatives as p56 lck inhibitors. Lett. Drug Des. Discov. 3, 550–560.

Ballard, P., Bradbury, R.H., Hennequin, L.F.A., Hickinson, D.M., Johnson, P.D., Kettle, J.G., Klinowska, T., Morgentin, R., Ogilvie, D.J., Olivier, A., 2005. 5-Substituted 4-anilinoquinazolines as potent, selective and orally active inhibitors of erbB2 receptor tyrosine kinase. Bioorg. Med. Chem. Lett. 15, 4226–4229.

Bridges, A.J., Zhou, H., Cody, D.R., Rewcastle, G.W., McMichael, A., Showalter, H.D.H., Fry, D.W., Kraker, A.J., Denny, W.A., 1996. Tyrosine kinase inhibitors. 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. 39, 267–276.

Cramer, R.D., Patterson, D.E., Bunce, J.D., 1988. Comparative molecular field analysis (CoMFA). 1. Effect of shape on binding of steroids to carrier proteins. J. Am. Chem. Soc. 110, 5959–5967.

Denny, W.A., 2001. The 4-anilinoquinazoline class of inhibitors of the erbB family of receptor tyrosine kinases. IL Farmaco 56, 51–56.

Fry, D.W., Kraker, A.J., McMichael, A., Ambroso, L.A., Nelson, J.M., Leopold, W.R., Connors, R.W., Bridges, A.J., 1994. A specific inhibitor of the epidermal growth factor receptor tyrosine kinase. Science 265, 1093.

Gehlhar, D.K., Verkhivker, G.M., Rejto, P.A., Sherman, C.J., Fogel, D.B., Fogel, L.J., Freer, S.T., 1995. Molecular recognition of the inhibitor AG-1343 by HIV-1 protease: conformationally flexible docking by evolutionary programming. Chem. Biol. 2, 317–324.

Gibson, K.H., Grundy, W., Godfrey, A.A., Woodburn, J.R., Ashton, S.E., Curry, B.J., Scarlett, L., Barker, A.J., Brown, D.S., 1997. Epidermal growth factor receptor tyrosine kinase: structure–activity relationships and antitumour activity of novel quinazolines. Bioorg. Med. Chem. Lett. 7, 2723–2728.

Gilbert, N., 1976. Statistics. W.B. Saunders Co., Philadelphia PA.

Golbraikh, A., Tropsha, A., 2002. Predictive QSAR modeling based on diversity sampling of experimental datasets for the training and test set selection. J. Comput. Aided Mol. Des. 16, 357–369.

Golbraikh, A., Tropsha, A., 2003. QSAR modeling using chirality descriptors derived from molecular topology. J. Chem. Inf. Comput. Sci. 43, 144–154.

Halgren, T.A., 1996. Merck molecular force field. III. Molecular geometries and vibrational frequencies. J. Comput. Chem. 17, 553–586.

Hennequin, L.F.A., Ballard, P., Boyle, F.T., Delouvrie, B., Ellston, R.P.A., Halsall, C.T., Harris, C.S., Hudson, K., Kendrew, J., Pease, J.E., Ross, H.S., Smith, P., 2006. Novel 4-anilinoquinazolines with C-6 carbon-linked side chains: synthesis and structure–activity relationship of a series of potent, orally active, EGF receptor tyrosine kinase inhibitors. Bioorg. Med. Chem. Lett. 16, 2672–2676.

Herbst, R.S., Fukuoka, M., Baselga, J., 2004. Gefitinib – a novel targeted approach to treating cancer. Nat. Rev. Cancer 4, 956–965.

Holland, J., 1975. Adaptation in Natural and Artificial Systems. University of Michigan Press.

Hoskuldsson, A., 1995. A combined theory for PCA and PLS. J. Chemometr. 9, 91–123.

Manjula, S.N., Noolvi, M.N., Parihar, K.V., Manohara Reddy, S.A., Ramani, V., Gadad, A.K., Singh, G., Kutty, N.G., Rao, M., 2009. Synthesis and anti-tumour activity of optically active thiourea and their 2-aminobenzothiazole derivatives: a novel class of anticancer agents. Eur. J. Med. Chem. 44, 2923–2929.

Molecular Design Suite 3.5, VLife Technologies, Pune, India. < www.vlifesciences.com > .

Nandi, S., Bagchi, M.C., 2009. QSAR of amino pyrido[2,3-*d*]pyrimidin-7-yl derivatives: anticancer drug design by computed descriptors. J. Enzyme Inhib. Med. Chem. 24, 937–948.

Noolvi, M.N., Patel, H.M., 2011a. Synthesis and in vitro antitumor activity of substituted quinazoline and quinoxaline derivatives: search for anticancer agent. Eur. J. Med. Chem. 46, 2327–2346.

Noolvi, M.N., Patel, H.M., 2011b. 2D QSAR studies on a series of quinazoline derivatives as tyrosine kinase (EGFR) inhibitor: an approach to design anticancer agents. Lett. Drug Des. Discov. 7, 556–586.

Noolvi, M.N., Patel, H.M. 2011c. Synthesis, method optimization, anticancer activity of 2,3,7-trisubstituted quinazoline derivatives and targeting EGFR-tyrosine kinase by rational approach. Arab. J. Chem., in press.

Noolvi, M.N., Patel, H.M., Bhardwaj, B., 2010. 2D QSAR studies on a series of 4-anilino quinazoline derivatives as tyrosine kinase (EGFR) inhibitor: an approach to design anti cancer agents. Digest J. Nanomater. Biostruct. 5, 387–401.

Noolvi, M.N., Patel, H.M., Agrawal, S., Zambre, A., Badiger, A. 2011a. Synthesis, antimicrobial and cytotoxic activity of novel 4 azetidine-2-one derivatives of 1H-benzimidazole. Arab. J. Chem., in press.

Noolvi, M.N., Patel, H.M., Bhardwaj, B. 2011b. An approach to design anticancer agents by 2D QSAR studies on a series of quinazoline analogues as tyrosine kinase (erbB-2) inhibitors. Med. Chem. 7, in press.

Oprea, T.I., Waller, C.L., Marshall, G.R., 1994. Three-dimensional quantitative structure–activity relationship of human immunodeficiency virus (I) protease inhibitors. 2. Predictive power using limited exploration of alternate binding modes. J. Med. Chem. 37, 2206.

Rewcastle, G.W., Denny, W.A., Bridges, A.J., Zhou, H., Cody, D.R., McMichael, A., Fry, D.W., 1995. Tyrosine kinase inhibitors. 5. Synthesis and structure–activity relationships for 4-[(phenylmethyl) amino]- and 4-(phenylamino) quinazolines as potent adeonosine-triphosphate binding site inhibitors of the tyrosine kinase domain of the epidermal growth factor receptor. J. Med. Chem. 38, 3482–3487.

Sharaf, M.A., Illman, D.L., Kowalski, B.R., 1986. Chemometrics. Wiley, New York.

Shen, M., LeTiran, A., Xiao, Y., Golbraikh, A., Kohn, H., Tropsha, A., 2002. Quantitative structure–activity relationship analysis of functionalized amino acid anticonvulsant agents using k nearest neighbor and simulated annealing PLS methods. J. Med. Chem. 45, 2811–2823.

Stamos, J., Sliwkowski, M.X., Eigenbrot, C., 2002. Structure of the epidermal growth factor receptor kinase domain alone and in complex with a 4-anilinoquinazoline inhibitor. J. Biol. Chem. 277, 46265–46272.

Sun, L., Xie, Y., Song, X., Wang, J., Yu, R., 1994. Cluster analysis by simulated annealing. Comput. Chem. 18, 103–108.

Verkhivker, G.M., Bouzida, D., Gehlhaar, D.K., Rejto, P.A., Arthurs, S., Colson, A.B., Freer, S.T., Larson, V., Luty, B.A., Marrone, T., Rose, P.W., 2000. Deciphering common failures in molecular docking of ligand-protein complexes. J. Comput. Aided Mol. Des. 14, 731–751.

VLife MDS 3.0 Documentation. 2007a. Tutorial: Engine Build Molecule, pp. 1–12.

VLife MDS 3.0 Documentation. 2007b. Tutorial: Protein Complex Optimization, pp. 1–8.

VLife MDS 3.0 Documentation. 2007c. Tutorial: Engine Build Molecule, pp. 17–20.
VLife MDS 3.0 Documentation. 2007d. Tutorial: Engine Build Molecule, pp. 9–14.
VLife MDS 3.0 Documentation. 2007e. Tutorial: Biopredicta Protein Complex Optimization, pp. 2–11.
Zheng, W., Tropsham, A., 2000. Novel variable selection quantitative structure–property relationship approach based on the k-nearest-neighbor principle. J. Chem. Inf. Comput. Sci. 40, 185–194.