

## 3D-QSAR studies on quinazoline derivatives as EGFR-T790M inhibitors by Comparative Molecular Field Analysis (CoMFA)

Cuizhu Ge<sup>1</sup>, Fucheng Song<sup>2</sup>, Hua Gao<sup>1</sup>, Honglin Zhai<sup>3</sup>, Hongzong Si<sup>4\*</sup>

<sup>1</sup>Department of Pharmaceutical Chemistry, Qingdao University School of Pharmacy, Qingdao, Shandong 266021, China

<sup>2</sup>School of Public Health, Qingdao University, Qingdao, Shandong 266071, China

<sup>3</sup>Department of Chemistry, Lanzhou University, Lanzhou, 730000, China

<sup>4</sup>Institute for Computational Science and Engineering, Laboratory of New Fibrous Materials and Modern Textile, The Growing Base for State Key Laboratory, Qingdao University, Qingdao, Shandong 266071, China

**Abstract:** To explore the structure activity correlation of quinazoline derivatives as EGFR-T790M inhibitors, three-dimensional quantitative structure-activity relationship (3D-QSAR) modeling were carried out. A satisfactory model with good predictive abilities have been developed by CoMFA, the cross validated  $Q^2=0.651$ , non-cross-validated  $R^2=0.988$ . Based on these results, we design novel compounds as potential inhibitors of EGFR-T790M with good predicted activities. The results of 3D-QSAR will guide the research and development of novel EGFR-T790M kinase inhibitors.

**Keywords:** EGFR; EGFR-T790M kinase inhibitors; Quinazoline derivatives; 3D-QSAR; CoMFA

Received 10 November 2015, Revised 22 November 2015, Accepted 25 November 2015

\*Corresponding Author: Hongzong Si, sihz@qdu.edu.cn

### 1. Introduction

Epidermal growth factor receptor (EGFR) is a transmembrane glycoprotein, consisting of a single polypeptide chain of 1186 amino acids that belongs to the ErbB family of receptor tyrosine kinase (RTK) [1,2]. This family includes EGFR/ErbB1/HER1, ErbB2/HER2/Neu, ErbB3/HER3, and ErbB4/HER4 [3]. EGFR is a cell-surface receptor, that binding of ligand (for example EGF and TGF- $\alpha$ ) to the extracellular domain of EGFR leads to receptor dimerization and activation of signaling pathways [4,5]. EGFR signaling pathways play a key role in the regulation of cell proliferation, survival and differentiation [6]. Several studies have strongly confirmed that EGFR overexpression and mutations is associated with many malignant diseases, highlighting EGFR as an interesting target in cancer therapy [7].

Lung cancer is the leading cause of death from cancer worldwide, non-small-cell lung cancer (NSCLC) constitutes more than 80% of all the cases of lung cancer [8]. The majority of patients with NSCLC are diagnosed at an advanced stage. Systemic platinum-based doublet chemotherapy combined with taxanes, pemetrexed and gemcitabine is the standard first-line therapy for advanced NSCLC [9]. However, it provides only a modest benefit in survival. The discovery of EGFR tyrosine kinase inhibitors (TKIs) has led to a new therapy for lung cancer treatment [10]. Most importantly, many NSCLC patients are with EGFR mutations, particularly in Asian countries [11]. The major EGFR mutations are exon 19 deletion mutations (Del-19s) and the L858R point mutation in exon 21 [12]. Erlotinib and gefitinib, reversible EGFR tyrosine-kinase inhibitors (TKIs), show stronger

binding affinity with mutants EGFR than wild-type EGFR. However, the efficacy of gefitinib and erlotinib become limited if a second-site T790M mutation in EGFR kinase domain [13]. So designing new irreversible inhibitors that overcome EGFR-T790M resistance has been proved to be quite challenging reports. In this study, we have built CoMFA models based on the structures of some quinazoline derivatives that show inhibition of EGFR-T790M kinase to provide helping for design novel EGFR-T790M potent kinase inhibitors.

### 2. Materials and methods

#### 2.1. Dataset

The structures of the quinazoline derivatives and their biological activities of 55 compounds were selected from the literature [14]. We converted their activity values IC<sub>50</sub> (nM) into pIC<sub>50</sub> in order to use it as dependent variables for subsequent 3D-QSAR analysis. These 55 compounds were divided into a training set of 44 compounds and a test set of 11 compounds. The structures of the compounds were constructed using ChemDraw Ultra 8.0, and then were optimized and calculated in SYBYL 7.3. Partial atomic charges were calculated by the Gasteiger-Hückel method [15]. After that energy minimization using the Tripos force field with a distance dependent-dielectric, Powell gradient algorithm [16] with a convergent threshold of a maximum deviate of 0.05 kcal/(mol\*Å) and a maximum of 1,000 iterations. The structures, biological activity and predicted activity values were given in Table 1.

Table 1 Structures and activity of the quinazoline derivatives as EGFR-T790M tyrosine kinase inhibitors.

No.		R <sub>3</sub> O-	R <sub>4</sub>	IC <sub>50</sub> (nM)	pIC <sub>50</sub>		
					Obs.	Pred.	Res.
1			H	102.2	6.991	7.059	-0.068
2			H	1.8	8.745	8.690	0.055
3			H	115.1	6.939	6.939	0.000
4			H	195.8	6.708	6.828	-0.120
5			H	560.4	6.252	6.097	0.155
6			H	72.2	7.141	7.134	0.007
7			H	134.9	6.870	6.979	-0.109
8*			H	100.1	7.000	7.021	-0.021
9*			H	32.2	7.492	7.421	0.071
10			H	104.7	6.980	7.072	-0.092
11			H	167.3	6.777	6.638	0.139
12			H	5.8	8.237	8.318	-0.081

13*			H	121.1	6.917	7.540	-0.623
14				1.2	8.921	8.958	-0.037
15*				573.6	6.241	7.457	-1.216
16				102.3	6.990	6.994	-0.004
17				91	7.041	7.195	-0.154
18				5.4	8.268	8.048	0.220
19				10	8.000	7.928	0.072
20				1.2	8.921	8.844	0.077
21*				3	8.523	8.484	0.039
22				510.2	6.292	6.334	-0.042
23				133.9	6.873	6.858	0.015
24*				8.1	8.092	7.665	0.427
25				9	8.046	8.024	0.022
26				1.2	8.921	8.912	0.009
27*				147.4	6.832	7.417	-0.585
28				6	8.222	8.205	0.017

29				8.1	8.092	8.114	-0.022
30*				5.3	8.276	7.746	0.530
31		$-\text{OCH}_2\text{CH}_2\text{OCH}_3$		2.1	8.678	8.782	-0.104
32*				520.2	6.284	6.981	-0.697
33		$-\text{OCH}_3$		26.9	7.570	7.701	-0.131
34		$-\text{OCH}_2\text{CH}_2\text{OCH}_3$		8.9	8.051	7.954	0.097
35		$-\text{OCH}_2\text{CH}_3$		5.5	8.260	8.309	-0.049
36*				109.4	6.961	6.614	0.347
37		$-\text{OCH}_2\text{CH}_2\text{OCH}_3$		356.6	6.448	6.387	0.061
38		$-\text{OCH}_2\text{CH}_3$		0.9	9.046	9.016	0.030
39		$-\text{OCH}_3$		2.1	8.678	8.682	-0.004
40				6.5	8.187	8.286	-0.099
41		$-\text{OCHF}_2$		9.6	8.018	8.061	-0.043
42		$-\text{OCH}_2\text{CF}_3$		6.9	8.161	8.206	-0.045
43		$-\text{OCH}_3$		141	6.851	6.857	-0.006
44				7	8.155	8.088	0.067
45				5.3	8.276	8.297	-0.021
46		$-\text{OCH}_3$		2.7	8.569	8.450	0.119
47		$-\text{OCH}_2\text{CH}_3$		6.1	8.215	8.303	-0.088

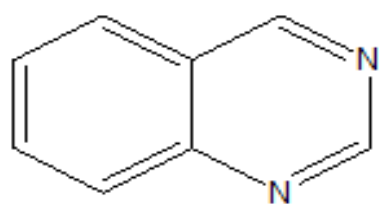
48		-OCH <sub>3</sub>		3.3	8.481	8.626	-0.145
49		-OCH <sub>2</sub> CH <sub>3</sub>		2.4	8.620	8.593	0.027
50				5.6	8.252	8.335	-0.083
51		-OCH <sub>3</sub>		7.9	8.102	8.027	0.075
52		-OCH <sub>2</sub> CH <sub>3</sub>		4.2	8.377	8.375	0.002
53				6.4	8.194	8.048	0.146
54*				1	9.000	8.483	0.517
55		-OCH <sub>3</sub>		1.8	8.745	8.609	0.136

\* test set

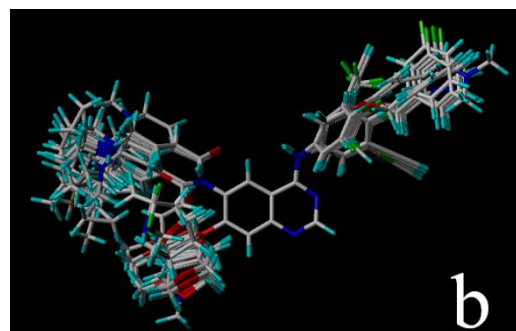
## 2.2. Molecular alignment

Molecular alignment is considered as one of the most important factors to construct 3D-QSAR models. The accuracy and reliability of the models directly rely on the quality of the structural alignment. In this study, we selected the highest active compound 38 as

template molecule, picked out the common fragment, and then aligned the rest of molecules on the template molecule by the Align Database command in SYBYL. The common fragment and the alignment of all the compounds in training set were shown in Figure 1.



a



b

Figure 1. a is the common fragment and b is the alignment of all the compounds in training set.

## 2.3. CoMFA model generation

To construct 3D-QSAR model, the CoMFA descriptors were used as independent variables and pIC<sub>50</sub> values as the dependent variable. In CoMFA study, steric (S) and electrostatic (E) fields were calculated using the Tripos force field with distance dependent-dielectric constant at each lattice intersection in a regularly spaced (2Å) grid taking sp<sup>3</sup> carbon atom as steric probe and a +1.0 unit charge as electrostatic probe. The SYBYL default energy cutoff of 30 kcal/mol was set for both steric and electrostatic fields.

The partial least squares (PLS) regression analysis

was used to build a linear correlation between these CoMFA descriptors to the biological activity values. PLS was performed in two stages. First, cross-validation analysis was done using the leave-one-out (LOO) method to determine the optimum number of components (ONC) and cross-validated correlation coefficient (Q<sup>2</sup>). Second, the obtained optimal component was applied to generate the final PLS regression models with non-cross-validation method. The non-cross-validated models were assessed by the non-cross-validated correlation coefficient (R<sup>2</sup>), the standard error of estimation (SEE), and F probability value (F value).

The robustness and the statistical liability of the models were further validated using bootstrap analysis for 100 runs [17] and Y-randomization test for 20 times [18]. In the Y-randomization test, new QSAR models were developed by scrambling the dependent variable vector randomly and keeping the original independent variable unchanged [19]. The Y-randomization results were evaluated by the parameter  $R_p^2$  which penalized the model for the difference between squared mean correlation coefficient ( $Q^2$ ) of randomized models and squared correlation coefficient ( $R^2$ ) of the non-randomized model [20]. The parameter was calculated by the following equation:

$$R_p^2 = R^2 * \sqrt{R^2 - R_r^2} \quad (1)$$

The parameter  $R_p^2$  should be greater than 0.5 for an acceptable model and ensures that the models are not obtained by chance.

The external predictive capability of CoMFA model were evaluated by predicting the activity of the test set, which were excluded during the model generation. The external  $R_{ext}^2$  was calculated using the following equation [17]:

$$R_{ext}^2 = 1 - \frac{\sum_{i=1}^{ntest} (y_i - \tilde{y}_i)^2}{\sum_{i=1}^{ntest} (y_i - \tilde{y}_{ir})^2} \quad (2)$$

where  $ntest$  is the number of compounds that constitute the validation data set (test set),  $\tilde{y}_{ir}$  is the averaged value for the dependent variable for the training set, and  $y_i$ ,  $\tilde{y}_i$ , ( $i = 1, \dots, ntest$ ) are the experimental values and the 3D-QSAR model predictions of the dependent variable over the available validation set.

To further validate the predictive ability of the derived models, the following equation was used [21]:

$$R_{m(overall)}^2 = r^2 * (1 - \sqrt{r^2 - r_0^2}) \quad (3)$$

where  $r^2$  is squared correlation coefficient between observed and predicted  $pIC_{50}$  values of 55 quinazoline analogues, and  $r_0^2$  is squared correlation coefficient with intercept set to zero. The parameter determines whether the range of predicted activity values for the whole dataset is really close to the observed activity or not. The value of  $R_{m(overall)}^2$  should be greater than 0.5 for a reasonable model.

### 3. Results and discussion

#### 3.1. CoMFA statistical results

The statistical parameters obtained from the CoMFA analyses of quinazoline analogues were summarized in Table 2. As can be seen, the CoMFA

model based on steric and electrostatic fields yielded a cross-validated  $Q^2$  of 0.651 with 8 optimum number of components (ONC=8), a non-cross-validated  $R^2$  of 0.988, a low standard error of estimation (SEE) of 0.098, and an F value of 356.844. The contributions of the steric and electrostatic fields were 50.3% and 49.7%, respectively.

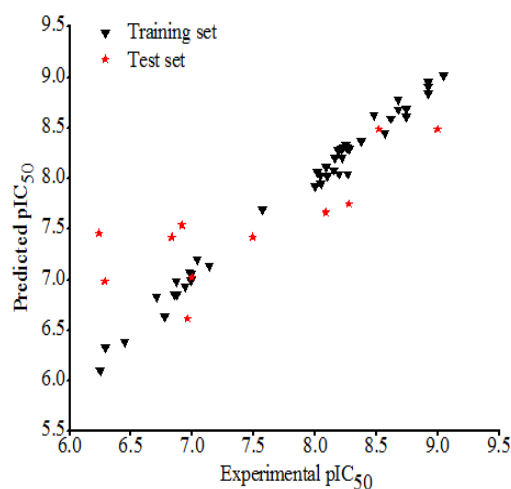
**Table 2 Statistical results of CoMFA model.**

PLS statistics	CoMFA
$Q^2$	0.651
ONC	8
$R^2$	0.988
SEE	0.098
F value	356.844
	Field contribution %
Steric	50.3
Electrostatic	49.7

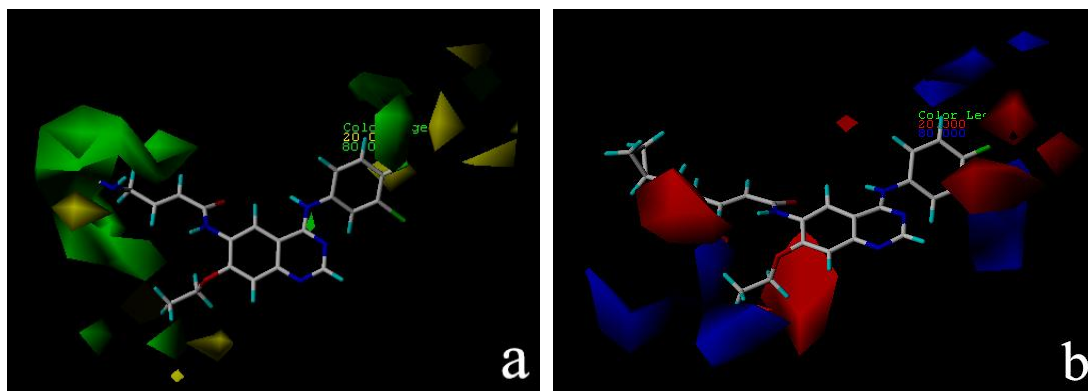
#### 3.2. Validation of the CoMFA model

To measure the bias of the original calculations, a bootstrapping analysis for 100 runs was performed,  $R_{boot}^2=0.995$  and  $SEE_{boot}=0.002$ . The higher value of  $R_{boot}^2$  and lower  $SEE_{boot}$  confirmed the robustness of the models. To further evaluate the robustness of the developed CoMFA model, Y-randomization test was performed. The  $Q^2$  (-0.650~0.112) and  $R_p^2$  (0.760) demonstrated that the constructed CoMFA model was stable.

In addition, the external validation method was also performed to assess the reliability and applicability of the built model.  $R_{ext}^2=0.679$ ,  $R_{m(overall)}^2=0.767$  confirmed the reliable predictive ability of the developed CoMFA model. Figure 2 showed a good relationship between predicted  $pIC_{50}$  values and experimental ones.



**Figure 2. Plot of predicted  $pIC_{50}$  values versus experimental ones.**



**Figure 3.** Stdev\*Coeff CoMFA contour maps superimposed with compound 38. (a) Steric field, (b) Electrostatic field.

To visualize the different fields effecting on the target compounds in 3D space, the non-cross validated CoMFA model was used to generate the 3D color-coded contour maps. The result of CoMFA was mapped using the “Stdev\*Coeff” option in SYBYL7.3. The contributions of favorable and unfavorable levels were maintained as the default value of 80% and 20%, respectively. The steric contour map was represented in green and yellow contours, in which green contours indicated regions where bulky group would be favorable, while the yellow contours represented regions where bulky group would be unfavorable. The electrostatic contour map was indicated by blue and red contours, where electropositive group near blue and electronegative group near red were favorable for increasing biological activity. The compound 38 was used as a reference structure. The contour maps of CoMFA model were shown in Figure 3.


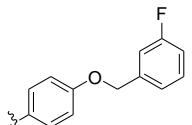
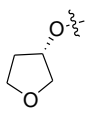
As shown in Figure 3a, the green contours near the R<sub>4</sub> position indicated bulky groups at this position would increase the potency. It may explained why

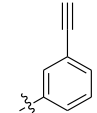
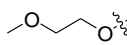
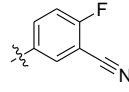
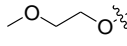
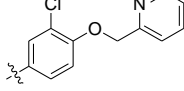
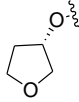
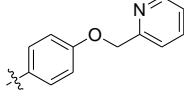
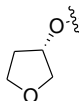
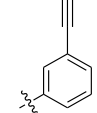
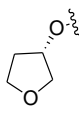
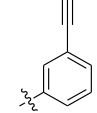
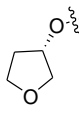
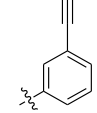
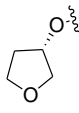
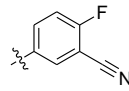
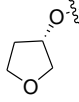
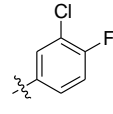
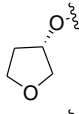
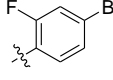

compounds 24, 25, 26 which possessed a -CH<sub>2</sub>N(CH<sub>3</sub>)<sub>2</sub> group at R<sub>4</sub> position showed significantly increased activities than compounds 7, 8, 9. In Figure 3b, a blue contour near the R<sub>3</sub> position indicated electropositive groups substituent would increase the biological activity. It may be explained why compound 39 showed increased activities than compounds 41, 42. Near the R<sub>4</sub> position, there is a red contour which indicated electronegative groups substituent would increase the biological activity.

### 3.4. Design novel compounds

As a validation of the conclusion from CoMFA model, the R<sub>4</sub> position substituted with electronegative group will increase the biological activity. So the R<sub>4</sub> position is substituted with a series of electronegative groups and the modified structures and their predicted pIC<sub>50</sub> values are shown in Table 3.

**Table 3** Design novel compounds and predicted pIC<sub>50</sub>.

No.		R <sub>3</sub> O-	R <sub>4</sub>	pIC <sub>50</sub> Predicted by CoMFA
56			OH	7.309

57			OH	9.076
58			OH	7.292
59			OH	7.547
60			OH	8.155
61			OH	8.979
62			F	8.277
63			Br	8.389
64			OH	7.708
65			OH	8.737
66			OH	8.808

#### 4. Conclusion

The development of new irreversible inhibitors that overcome EGFR-T790M resistance has been proved to be quite challenging reports. In this work, we use 3D-QSAR method to explore the structure-activity relationship of EGFR-T790M inhibitors to EGFR-T790M kinase. A good predictive capabilities CoMFA model was generated and used for design novel potential EGFR-T790M inhibitors. In addition, a series of validation methods, including bootstrapping, Y-randomization as well as external validation was carried out to validate the model. The validation results illustrate that the established CoMFA model is steady and reliable. Consequently, our study could be applied to further optimize and design new EGFR-T790M potent kinase inhibitors.

#### References

- [1] Swati A, Deepa S, Anjani Kumar T, Pooja P, Ankur K, Gurmeet S, Anil KM. Studies for development of novel quinazolinones: New biomarker for EGFR. *Spectrochim Acta A*, 143 2015 309-318.
- [2] Chetan Y, Dipesh B, Imran V, Sushilkumar P, Ambikanandan M. Epidermal growth factor receptor targeting in cancer: A review of trends and strategies. *Biomaterials*, 34 2013 8690-8707.
- [3] Solange P, Stefan Z, Alex AA. Oral epidermal growth factor receptor tyrosine kinase inhibitors for the treatment of non-small cell lung cancer: Comparative pharmacokinetics and drug-drug interactions. *Cancer Treat Rev*, 40 2014 917-926.
- [4] Hassan E, Hanane I, Anouar B, Fouad K, Lamia G, Amina M, Lamia J, Youssef B, Hind M, Ibrahim E. Frequency and type of epidermal growth factor receptor mutations in moroccan patients with lung adenocarcinoma. *J Thorac Oncol*, 8(9) 2013 1212-1214.
- [5] Goffin JR, Zbuk K. Epidermal growth factor receptor: pathway, therapies, and pipeline. *Clin Ther*, 35(9) 2013 1282-1303.



- [6] Mudjupa C, Abdelhamed S, Refaat A, Yokoyama S, Saiki I, Vajragupta O. Lead compound bearing caffeic scaffold induces EGFR suppression in solid tumor cancer cells. *J Appl Biomed*, 13 2015 305-317.
- [7] Wang XB, Zhou HY. Molecularly targeted gemcitabine-loaded nanoparticulate system towards the treatment of EGFR overexpressing lung cancer. *Biomed Pharmacother*, 70 2015 123-128.
- [8] Baek JH, Sun J, Mina YJ, Cho EK, Cho BC, Kim J, Ahn M, Park K. Efficacy of EGFR tyrosine kinase inhibitors in patients with EGFR-mutated non-small cell lung cancer except both exon 19 deletion and exon 21 L858R: A retrospective analysis in Korea. *Lung Cancer*, 87 2015 148-154.
- [9] Narita Y, Matsushima Y, Shirowa T, Chiba K, Nakanishi Y, Kurokaw T, Urushihara H. Cost-effectiveness analysis of EGFR mutation testing and gefitinib as first-line therapy for non-small cell lung cancer. *Lung Cancer*, 90 2015 71-77.
- [10] Roengvoraphoj M, Tsongalis GJ, Dragnev KH, Rigas JR. Epidermal growth factor receptor tyrosine kinase inhibitors as initial therapy for non-small cell lung cancer: Focus on epidermal growth factor receptor mutation testing and mutation-positive patients. *Cancer Treat Rev*, 39 2013 839-850.
- [11] Wolfram B, Michael T. EGFR-TKI resistant non-small cell lung cancer (NSCLC): New developments and implications for future treatment. *Lung Cancer*, 77 2012 2-8.
- [12] Kaneda T, Hata A, Tomioka H, Tanaka K, Kaji R, Fujita S, Tomii K, Katakami N. Possible differential EGFR-TKI efficacy among exon 19 deletional locations in EGFR-mutant non-small cell lung cancer. *Lung Cancer*, 86 2014 213-218.
- [13] Ma L, Wang DD, Huang YQ, Wong MP, Lee VHF, Yan H. Decoding the EGFR mutation-induced drug resistance in lung cancer treatment by local surface geometric properties. *Comput Biol Med*, 63 2015 293-300.
- [14] Zhang L, Yang YY, Zhou HJ, Zheng QM, Li YH, Zheng SS, Zhao SY, Chen D, Fan CW. Structure-activity study of quinazoline derivatives leading to the discovery of potent EGFR-T790M inhibitors. *Eur J Med Chem*, 102 2015 445-463.
- [15] Johann G, Mario M. Iterative partial equalization of orbital electronegativity-a rapid access to atomic charges. *Tetrahedron*, 22 1980 3219-3228.
- [16] Clark M, Cramer RD, Vanopdenbosch N. Validation of the general-purpose Tripos 5.2 force-field. *J Comput Chem*, 10 1989 982-1012.
- [17] Mouchlis VD, Melagraki G, Mavromoustakos T, Kollias G, Afantitis A, Melagraki G. Molecular Modeling on Pyrimidine-Urea Inhibitors of TNF- $\alpha$  Production: An Integrated Approach Using a Combination of Molecular Docking, Classification Techniques, and 3D-QSAR CoMSIA. *J Chem Inf Model*, 52 2012 711-723.
- [18] Bhadoriya KS, Sharma MC, Jain SV. Pharmacophore modeling and atom-based 3D-QSAR studies on amino derivatives of indole as potent isoprenylcysteine carboxyl methyltransferase (Icmt) inhibitors. *J Mol Struct*, 1081 2015 466-476.
- [19] Damre MV, Gangwal RP, Dhoke GV, Lalit M, Sharma D, Khandelwal K, Sangamwar AT. 3D-QSAR and molecular docking studies of amino-pyrimidine derivatives as PknB inhibitors. *J Taiwan Inst Chem E*, 45 2014 354-364.
- [20] Partha PR, Somnath P, Indrani M, Kunal R. On two novel parameters for validation of predictive QSAR models. *Molecules*, 14 2009 1660-1701.
- [21] Yu Z, Li XC, Ge CZ, Si HZ, Cui LH, Gao H, Duan YB, Zhai HL. 3D-QSAR modeling and molecular docking study on Mer kinase inhibitors of pyridine-substituted pyrimidines. *Molecular Diversity*, 19 2015 135-147.