



If you did not document it, you did not do it..



3D QSAR STUDIES ON A SERIES OF QUINAZOLINE DERRIVATIVES AS TYROSINE KINASE (EGFR) INHIBITOR: THE k-NEAREST NEIGHBOR MOLECULAR FIELD ANALYSIS APPROACH.

Malleshappa N. Noolvi* and Harun M. Patel

Department of Pharmaceutical Chemistry, ASBASJSM College of Pharmacy, Bela (Ropar)-14011, Punjab, India

ABSTRACT: Epidermal growth factor receptor (EGFR) protein tyrosine kinases (PTKs) are known for its role in cancer. Quinazoline have been reported to be the molecules of interest, with potent anticancer activity and they act by binding to ATP site of protein kinases. ATP binding site of protein kinases provides an extensive opportunity to design newer analogs. With this background, we report an attempt to discern the structural and physicochemical requirements for inhibition of EGFR tyrosine kinase. The k-Nearest Neighbor Molecular Field Analysis (kNN-MFA), a three dimensional quantitative structure activity relationship (3D- QSAR) method has been used in the present case to study the correlation between the molecular properties and the tyrosine kinase (EGFR) inhibitory activities on a series of quinazoline derivatives. kNN-MFA calculations for both electrostatic and steric field were carried out. The master grid maps derived from the best model has been used to display the contribution of electrostatic potential and steric field. The statistical results showed significant correlation coefficient r^2 (q^2) of 0.846, r^2 for external test set (pred_r^2) 0.8029, coefficient of correlation of predicted data set (pred_r^2se) of 0.6658, degree of freedom 89 and k nearest neighbor of 2. Therefore, this study not only casts light on binding mechanism between EGFR and its inhibitors, but also provides hints for the design of new EGFR inhibitors with observable structural diversity.

KEYWORDS: Quinazoline, Tyrosine kinase (E GFR), k-Nearest Neighbor Molecular Field Analysis (kNN-MFA).

received on 12-04-2010
modified on 03-05-2010
accepted on 10-06-2010
available online 15-08-2010
www.jbclinpharm.com

INTRODUCTION

Many of the tyrosine kinase enzymes are involved in cellular signaling pathways and regulate key cell functions such as proliferation, differentiation, anti-apoptotic signaling and neurite outgrowth. Unregulated activation of these enzymes, through mechanisms such as point mutations or over expression, can lead to a large percentage of clinical cancers [1, 2]. The importance of tyrosine kinase enzymes in health and disease is further underscored by the

existence of aberrations in tyrosine kinase enzymes signaling occurring in inflammatory diseases and diabetes. Inhibitors of tyrosine kinase as a new kind of effective anticancer drug are important mediators of cellular signal transduction that affects growth factors and oncogenes on cell proliferation [3, 4]. The development of tyrosine kinase inhibitors has therefore become an active area of research in pharmaceutical science. Epidermal growth factor receptor (EGFR) which plays a vital role as a regulator of cell growth is one of the intensely studied tyrosine kinase targets of inhibitors. EGFR is overexpressed in numerous tumors, including those

*Corresponding Author:
E-mail: mnoolvi@yahoo.co.uk

derived from brain, lung, bladder, colon, breast, head and neck. EGFR hyper activation has also been implicated in other diseases including polycystic kidney disease, psoriasis and asthma [5-7]. Since the hyper activation of EGFR has been associated with these diseases, inhibitor of EGFR has potential therapeutic value and it has been extensively studied in the pharmaceutical industry.

One could not, however, confirm that the compounds designed would always possess good inhibitory activity to EGFR, 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.

Many different approaches to QSAR have been developed over the years. The rapid increase in three-dimensional structural information (3D) of bio-organic molecules, coupled with the development of fast methods for 3D structure alignment (e.g. active analogue approach), has led to the development of 3D structural descriptors and associated 3D QSAR methods. The most popular 3D QSAR methods are comparative molecular field analysis (CoMFA) and comparative molecular similarity analysis (CoMSIA) [8, 9]. The CoMFA method involves generation of a common three dimensional lattice around a set of molecules and calculation of the steric and electrostatic interaction energies at the lattice points. The interaction energies are numerically very high when a lattice point is very close to an atom and special care needs to be taken in order to avoid problems arising because of this. The CoMSIA method avoids these problems by using similarity function represented as Gaussian. This information around the molecule is converted into numerical data using the partial least squares (PLS) method that reduces the dimensionality of data by generating components. However, a major disadvantage is that PLS attempts to fit a linear curve among all the points in the data set. Further, the PLS method does not offer scope for improvement in results. It has been observed from several reports that the predictive

ability of PLS method is rather poor due to fitting of a linear curve between the available points. In the case of the CoMSIA method, molecular similarity is evaluated and used instead of molecular field, followed by PLS analysis.

Variable selection methods have also been adopted for optimal region selection in 3D QSAR methods and shown to provide improved QSAR models as compared to the original CoMFA technique. For example, GOLPE was developed using chemometric principles, and q2-GRS was developed on the basis of independent analyses of small areas (or regions) of near molecular space to address the issue of optimal region selection in CoMFA [10, 11]. These considerations provide an impetus for the development of fast, generally nonlinear, variable selection methods for performing molecular field analysis. With the above facts and in continuation of our research for newer anti-cancer agent [12, 13] in the present study, we report here the development of a new method (kNN-MFA) that adopts a k-nearest neighbor principle for generating relationships of molecular fields with the experimentally reported activity to provide further insight into the key structural features required to design potential drug candidates of this class. This method utilizes the active analogue principle that lies at the foundation of medicinal chemistry.

COMPUTATIONAL METHODS

A. Methodology

We hereby report the models, as generated by kNN-MFA in conjunction with stepwise (SW) forward-backward variable selection methods. In the kNN-MFA method, several models were generated for the selected members of training and test sets, and the corresponding best models are reported herein. VLife Molecular Design Suite (VLifeMDS), allows user to choose probe, grid size, and grid interval for the generation of descriptors. The variable selection methods along with the corresponding parameters are allowed to be chosen, and optimum models are generated by maximizing q2. k-nearest neighbor 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 [14]. 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.

1. Nearest Neighbor (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 neighbors 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: Calculate distances between an unknown object (u) and all the objects in the training set; select k objects from the training set most similar to object u, according to the calculated distances; and 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 [15].

2. 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.

3. kNN-MFA with Stepwise (SW) Variable Selection

This method employs a stepwise variable selection procedure combined with kNN to optimize the number of nearest neighbors (k) and the selection of variables from the original pool as described in simulated annealing.

4. kNN-MFA with Genetic Algorithm

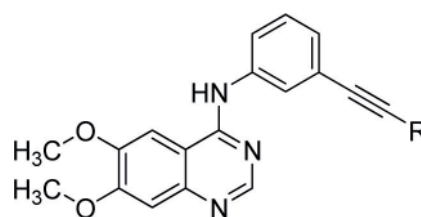
Genetic algorithms (GA) first described by Holland [16] mimic natural evolution and selection. In biological systems, genetic information that determines the individuality of an organism is stored in chromosomes. Chromosomes are replicated and passed onto the next generation with selection criteria depending on fitness.

B. Chemical Data

One hundred twenty six quinazoline derivatives as tyrosine kinase (EGFR) inhibitors were taken from the literature and used for kNN-MFA analysis [17-28]. The above reported quinazoline derivatives showed wide variation in their structure and potency profiles. kNN-MFA (3DQSAR) models were generated for these derivatives using a training set of 98 molecules. Predictive power of the resulting models was evaluated by a test set of 28 molecules with uniformly distributed biological activities. Selection of test set molecules was made by considering the fact that test set molecules represent structural features similar to compounds in the training set [29]. The structures of all compounds along with their actual and predicted biological activities are shown in Table 1(A-Y).

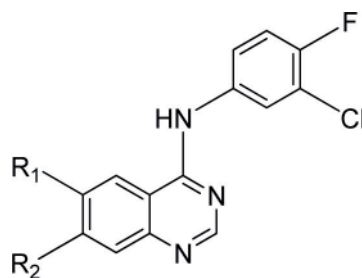
Table 1 (A-W): Structure, Experimental and Predicted Activity of Quinazolines Used in Training and Test Set using model 1

Table 1A:



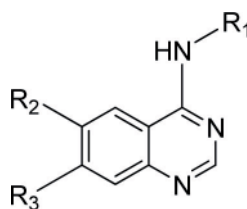
Sr.No.	Index	R	IC ₅₀ ^a (μM)	pIC ₅₀ ^b		Residual
				Exp.	Pred.	
1	4	C ₆ H ₅	0.921	6.0357	5.7791	0.2566
2	5	(CH ₂) ₂ OH	1.643	5.7843	6.2187	-0.4344
3	6	(CH ₂) ₃ OH	0.402	6.3957	5.8729	0.5228
4	7	C(CH ₃) ₂ OH	1.362	5.8658	6.009	-0.1432

Table 1B:



Sr.No.	Index	R	R ₂	IC ₅₀ ^a (μM)	pIC ₅₀ ^b		Residual
					Exp.	Pred.	
5	2		H	0.050	7.3010	7.0557	0.2453
6	3		H	0.088	7.0555	7.0183	0.0372
7	9		H	1.640	5.7851	6.8871	-1.102
8	1		OCH ₃	0.039	7.4089	7.4771	-0.0682
9	8 ^T		H	0.126	6.8996	6.3894	0.5102

Table 1C:



Sr.No.	Index	R ₁	R ₂	R ₃	IC ₅₀ ^a (μM)	pIC ₅₀ ^b		Residual
						Exp.	Pred.	
10	10 ^T		H	H	4.100	5.3872	5.1073	0.2799
11	11		OCH ₃	OCH ₃	0.148	6.8297	6.0700	0.7597
12	12		OCH ₃	OCH ₃	0.234	6.6307	6.5306	0.1001

Table 1C: Continued

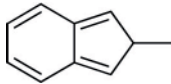
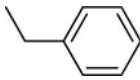
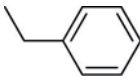
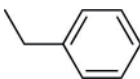
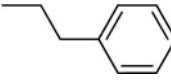
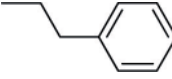
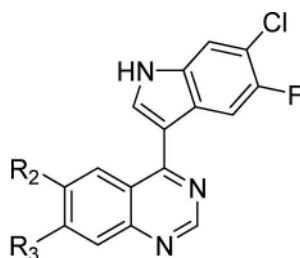
Sr.No.	Index	R ₁	R ₂	R ₃	IC ₅₀ ^a (μM)	pIC ₅₀ ^b		Residual
						Exp.	Pred.	
13	13		OCH ₃	OCH ₃	1.00	6.00	6.0671	-0.061
14	14t		H	H	0.320	6.4948	7.0614	-0.5666
15	15 ^T		H	OCH ₃	0.058	7.2365	7.0613	0.1752
16	16		OCH ₃	OCH ₃	0.010	8.000	8.4797	-0.4797
17	17		H	H	0.086	7.0655	7.3682	-0.3027
18	18		H	OCH ₃	0.029	7.5376	7.0593	0.4783

Table 1D:



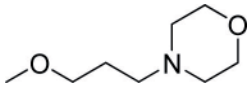
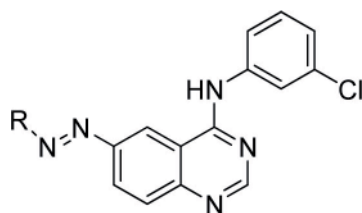
Sr.No.	Index	R ₂	R ₃	IC ₅₀ ^a (μM)	pIC ₅₀ ^b		Residual
					Exp.	Pred.	
19	19 ^T	OCH ₃	OCH ₃	0.209	6.6798	7.9607	-1.2809
20	20	OCH ₂ CH ₂ OCH ₃	OCH ₂ CH ₂ OCH ₃	0.472	6.3260	6.3022	0.0238
21	21		OCH ₃	0.333	6.4775	5.9600	0.5175

Table 1E:



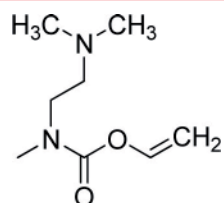
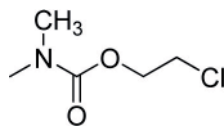
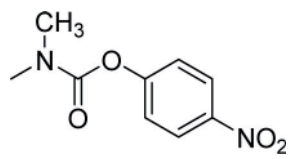
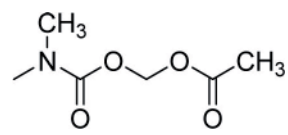
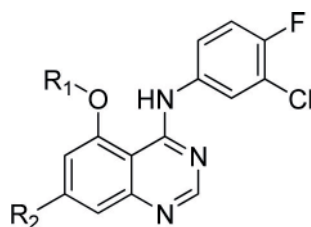
Sr.No.	Index	R	IC ₅₀ ^a (μM)	pIC ₅₀ ^b		Residual
				Exp.	Pred.	
22	22		0.18	6.7447	6.8724	-0.1277
23	23 ^T		0.50	6.3010	7.5469	-1.2459
24	24		0.77	6.1135	6.0487	0.0648
25	25		0.63	6.2006	6.4705	-0.2699

Table 1F:



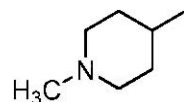
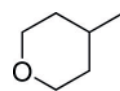
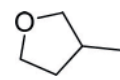
Sr.No.	Index	R ₁	R ₂	IC ₅₀ ^a (μM)	pIC ₅₀ ^b		Residual
					Exp.	Pred.	
26	120		OCH ₃	0.021	7.6777	6.6367	1.041
27	125		OCH ₃	1.384	5.8588	6.0627	-0.2039
28	126		OCH ₃	0.187	6.7281	6.0047	0.7234

Table 1F: Continued

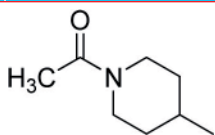
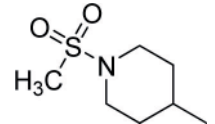
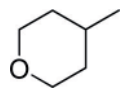
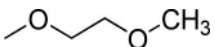
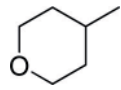
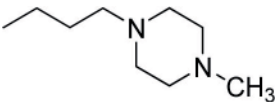
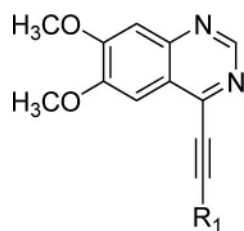
Sr.No.	Index	R ₁	R ₂	IC ₅₀ ^a (μM)	pIC ₅₀ ^b		Residual
					Exp.	Pred.	
29	121		OCH ₃	0.873	6.0589	5.4980	0.5689
30	122		H	9.601	5.0176	4.8863	0.1313
31	123			0.671	6.1732	6.6764	-0.504
32	124			0.066	7.1804	6.9526	0.2278

Table 1G:



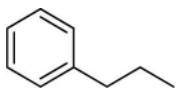
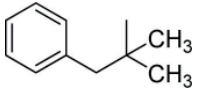
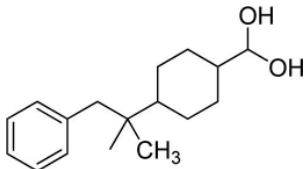
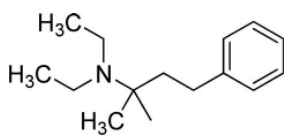
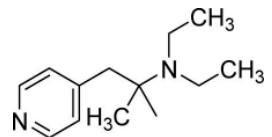
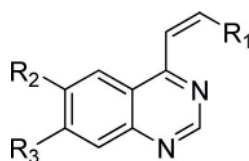
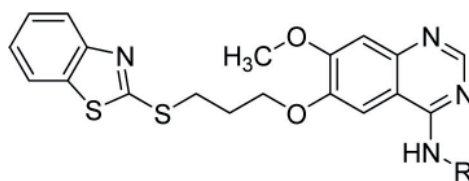
Sr.No.	Index	R	IC ₅₀ ^a (μM)	pIC ₅₀ ^b		Residual
				Exp.	Pred.	
33	33		0.014	7.8538	7.5036	0.8178
34	36 ^T		0.013	7.8860	7.783	0.103
35	37		0.033	7.4814	7.3368	0.1446
36	26		1.200	5.9208	6.1254	-0.2046
37	27		0.200	6.6989	7.5001	-0.8012

Table 1H:



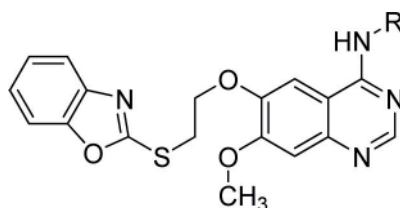
Sr.No.	Index	R ₁	R ₂	R ₃	IC ₅₀ ^a (μM)	pIC ₅₀ ^b		Residual
						Exp.	Pred.	
38	28		OCH ₃	OCH ₃	0.004	8.3979	7.2130	1.1849
39	30		OCH ₂ OCH ₃	OCH ₂ OCH ₃	0.015	7.8239	7.1639	0.66
40	31		OCH ₂ OCH ₃	OCH ₂ OCH ₃	0.005	8.3010	7.5943	0.7067
41	32		OCH ₂ OCH ₃	OCH ₂ OCH ₃	0.140	6.8538	6.1981	0.6557

Table 1I:



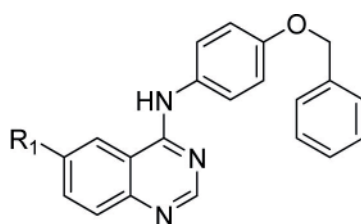
Sr.No.	Index	R	IC ₅₀ ^a (μM)	pIC ₅₀ ^b		Residual
				Exp.	Pred.	
42	39		5.35	5.2700	5.1301	0.1399
43	38		4.05	5.3925	5.0316	0.3609
44	56		5.87	5.2313	5.4640	-0.2327
45	57		14.17	4.8486	4.9251	-0.0765
46	58		14.25	4.8461	5.0317	-0.1856

Table 1J:



Sr.No.	Index	R	IC ₅₀ ^a (μM)	pIC ₅₀ ^b		Residual
				Exp.	Pred.	
47	59		4.05	5.3925	5.0124	0.3801
48	60		4.03	5.3946	5.3931	0.0015
49	61		9.08	5.0416	5.2016	-0.16
50	62		7.25	4.7632	5.0482	-0.285
51	40		13.06	4.8840	5.2166	-0.3326
52	41		12.09	4.9175	5.0044	-0.0869

Table 1K:

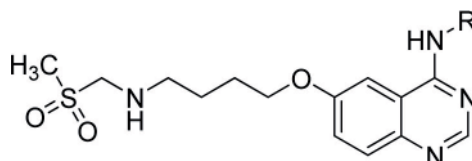


Sr.No.	Index	R	IC ₅₀ ^a (μM)	pIC ₅₀ ^b		Residual
				Exp.	Pred.	
53	91 ^T	OH	0.098	7.0087	7.2968	-0.2881
54	97		0.263	6.5800	6.8503	-0.2703
55	98	OCH ₃	0.074	7.1307	7.3250	-0.1943

Table 1k: Continued

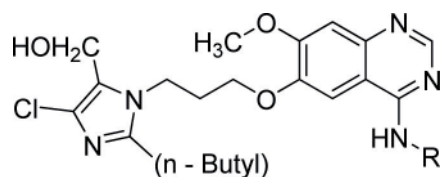
Sr.No.	Index	R	IC ₅₀ ^a (μM)	pIC ₅₀ ^b		Residual
				Exp.	Pred.	
56	101		0.068	7.1674	6.9495	0.2179
57	99		0.104	6.9829	7.1621	-0.1792
58	100		0.830	6.0809	7.0556	-0.9747
59	102		0.074	7.1307	6.9241	0.2066
60	103		0.074	7.1307	7.1457	-0.015

Table 1L:



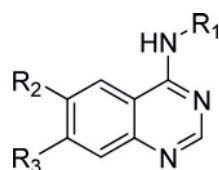
Sr.No.	Index	R	IC ₅₀ ^a (μM)	pIC ₅₀ ^b		Residual
				Exp.	Pred.	
61	95		0.011	7.9586	7.3971	0.5615
62	93		0.020	7.6989	8.0204	-0.3215
63	94		0.093	7.0315	7.1333	-0.1018
64	92		0.027	7.5686	7.1199	0.4487
65	96		0.024	7.6197	7.9548	-0.3351

Table 1M:



Sr.No.	Index	R	IC ₅₀ ^a (μM)	pIC ₅₀ ^b		Residual
				Exp.	Pred.	
66	49		8.28	5.0819	5.7111	-0.6292
67	50		10.57	4.9759	5.2131	-0.2372
68	51		20.02	4.6985	5.3927	-0.6942
69	52		9.96	5.0017	4.9603	0.0414
70	53		8.20	5.0861	5.2265	-0.1404
71	54		17.69	4.7522	4.8375	-0.0853
72	55		10.03	4.9986	4.8323	0.1663

Table 1N:



Sr.No.	Index	R ₁	R ₂	R ₃	IC ₅₀ ^a (μM)	pIC ₅₀ ^b		Residual
						Exp.	Pred.	
73	104			H	0.008	8.096	7.791	0.305
74	105			OCH ₃	0.515	6.288	6.452	-0.164

Table 1N: Continued

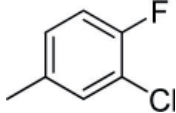
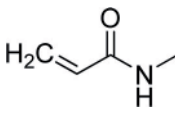
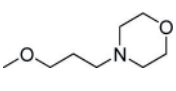
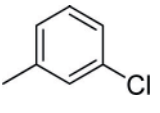
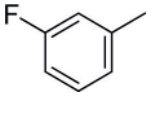
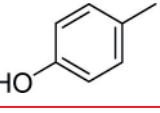
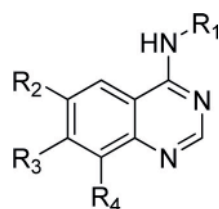
Sr.No.	Index	R ₁	R ₂	R ₃	IC ₅₀ ^a (μM)	pIC ₅₀ ^b		Residual
						Exp.	Pred.	
75	106				0.074	7.130	7.463	-0.333
76	113		OCH ₃	OCH ₃	0.03	7.522	7.329	0.193
77	107T		OCH ₃	OCH ₃	0.025	7.602	8.00	-0.398
78	108		OCH ₃	OCH ₃	0.10	7.00	7.617	-0.617

Table 1O:



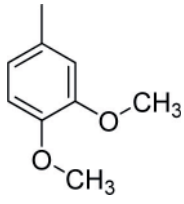
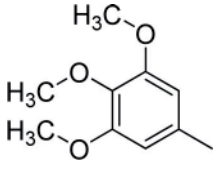
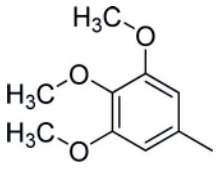
Sr.No.	Index	R ₁	R ₂	R ₃	R ₄	IC ₅₀ ^a (μM)	pIC ₅₀ ^b		Residual
							Exp.	Pred.	
79	109T		OCH ₃	OCH ₃	H	0.35	6.4559	6.5467	-0.0908
80	110		OCH ₃	OCH ₃	OCH ₃	18	4.6989	5.0431	-0.3442
81	111		OCH ₃	OCH ₃	OCH ₃	17	4.6089	7.2100	-2.6011

Table 1O: Continued

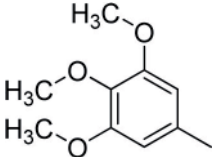
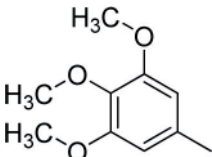
Sr.No.	Index	R ₁	R ₂	R ₃	R ₄	IC ₅₀ ^a (μM)	pIC ₅₀ ^b		Residual
							Exp.	Pred.	
82	112		OCH ₃	OCH ₃	C1 at C-2	19	4.6980	5.2305	-0.5325
83	114		OCH ₃	H	H	0.10	7.000	7.1685	-0.1685

Table 1P:

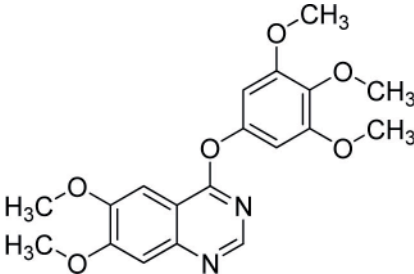
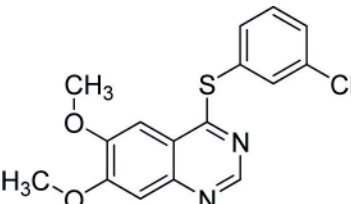
Sr.No.	Index	Molecules	IC ₅₀ ^a (μM)	pIC ₅₀ ^b		Residual
				Exp.	Pred.	
84	115		0.30	6.3010	6.9534	-0.6524
85	116		0.01	8.000	8.2981	-0.2981

Table 1Q:

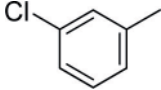
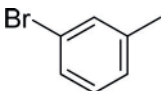
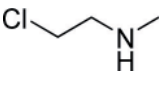
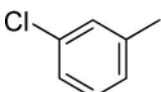
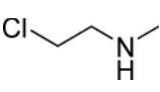
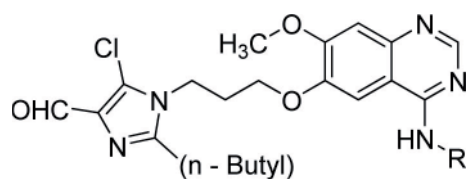
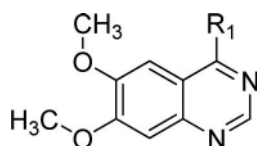
Sr.No.	Index	R ₁	R ₂	IC ₅₀ ^a (μM)	pIC ₅₀ ^b		Residual
					Exp.	Pred.	
86	117 ^T		NH ₂	0.2	6.698	7.061	-0.363
87	118 ^T			0.01	8.00	6.385	1.615
88	119 ^T			0.026	7.585	7.061	0.524

Table 1R:



Sr.No.	Index	R	IC ₅₀ ^a (μM)	pIC ₅₀ ^b		Residual
				Exp.	Pred.	
89	42		5.02	5.2992	5.0294	0.2698
90	43		3.51	5.4546	5.0438	0.4108
91	44		13.99	4.8539	5.0600	-0.2061
92	45		3.000	5.5228	5.3152	0.3708
93	46		6.44	5.1911	4.8536	0.3375
94	47		12.00	4.9208	5.0481	-0.1273
95	48		10.05	4.9978	5.2263	-0.2285

Table 1S:



Sr.No.	Index	R	IC ₅₀ ^a (μM)	pIC ₅₀ ^b		Residual
				Exp.	Pred.	
96	88t		0.009	8.0457	7.3779	0.6678

Table 1S: Continued

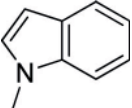
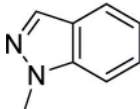
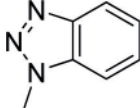
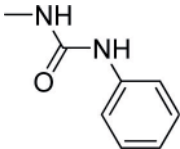
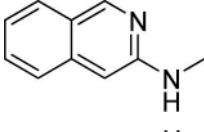
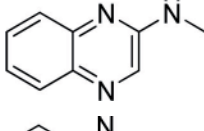
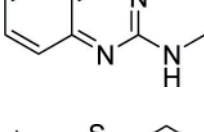
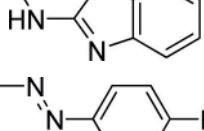
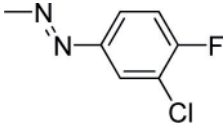
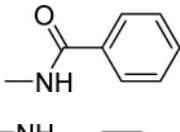
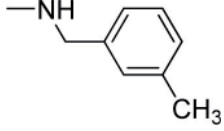
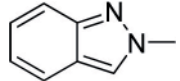
Sr.No.	Index	R	IC ₅₀ ^a (μM)	pIC ₅₀ ^b		Residual
				Exp.	Pred.	
97	89		0.40	6.0506	6.0381	0.0125
98	90 ^t		0.64	6.2009	5.9567	0.2442
99	70 ^t		0.73	6.3009	6.5277	-0.2268
100	71 ^t		0.65	6.4049	6.3490	0.0559
101	73 ^t		0.66	6.1804	7.0103	-0.8299
102	72		0.01	8.000	8.3896	-0.3896
103	74		0.82	6.0861	6.3725	-0.2864
104	75 ^t		0.70	6.2054	6.1541	-0.3356
105	76		0.73	6.1366	6.0857	0.0509
106	77		0.67	6.3040	6.7578	-0.4538
107	79 ^t		0.17	6.7695	6.8587	-0.0937
108	80 ^t		0.72	6.5421	6.3405	0.2016

Table 1S: Continued

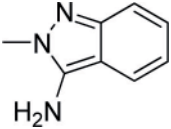
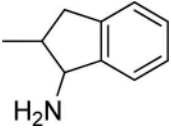
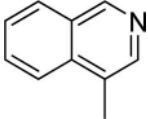
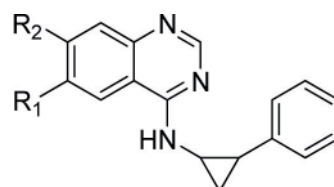
Sr.No.	Index	R	IC ₅₀ ^a (μM)	pIC ₅₀ ^b		Residual
				Exp.	Pred.	
110	82 ^T		0.35	6.4422	6.5112	-0.069
111	83 ^T		0.91	6.0409	6.3538	-0.3129
112	84 ^T		0.0064	8.1938	6.6823	1.5115

Table 1T:



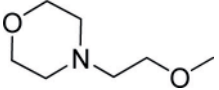
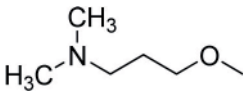
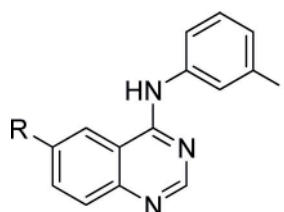
Sr.No.	Index	R ₁	R ₂	IC ₅₀ ^a (μM)	pIC ₅₀ ^b		Residual
					Exp.	Pred.	
113	69 ^T	OCH ₃	OCH ₃	0.01	8.000	7.9253	0.0747
114	78	OCOCH ₃	OCH ₃	0.084	7.0757	6.8913	0.1844
115	85 ^T	OH	OCH ₃	0.39	6.4089	6.3825	0.0264
116	86		OCH ₃	0.3	6.5228	6.4423	0.0805
117	87		OCH ₃	0.027	7.5686	7.9773	-0.4087

Table 1U:



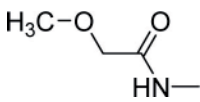
Sr.No.	Index	R	IC ₅₀ ^a (μM)	pIC ₅₀ ^b		Residual
				Exp.	Pred.	
118	67		0.072	7.1426	7.5408	-0.3982

Table 1U: Continued

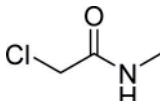
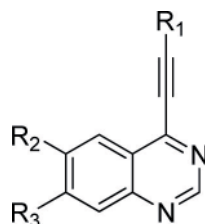
Sr.No.	Index	R	IC ₅₀ ^a (μM)	pIC ₅₀ ^b		Residual
				Exp.	Pred.	
119	68		0.02	7.6989	7.5209	0.178

Table 1V:



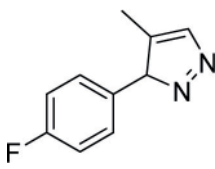
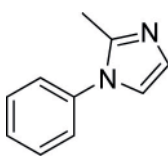
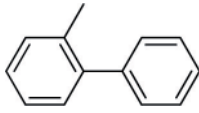
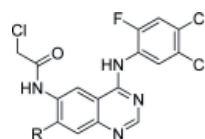
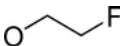
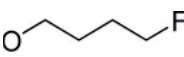
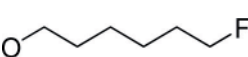
Sr.No.	Index	R	R ₂	R ₃	IC ₅₀ ^a (μM)	pIC ₅₀ ^b		Residual
						Exp.	Pred.	
120	35 ^T		OC ₂ OCH ₃	OC ₂ OCH ₃	0.0018	8.7447	6.6367	2.108
121	34		OC ₂ OCH ₃	OC ₂ OCH ₃	0.012	7.9208	7.930	-0.0092
122	29		OCH ₃	OCH ₃	0.026	7.5850	7.4378	0.1472

Table 1W:



Sr.No.	Index	R	IC ₅₀ ^a (μM)	pIC ₅₀ ^b		Residual
				Exp.	Pred.	
123	63		0.008	8.0969	7.0047	1.0992
124	64		0.023	6.6382	7.4201	-0.7819
125	65		0.038	7.4202	6.6383	0.7819
126	66	H	0.010	8.000	8.2915	-0.2915

Expt. = Experimental activity, Pred. = Predicted activity

a = Compound concentration required to inhibit tumor cell proliferation by 50%

b = -Log (IC₅₀ * 10⁻⁶): Training data set developed using model 1

T Test Set

C] Biological activities

126 quinazoline derivatives having different substitution were divided into two sets, 98 (75%) molecules were taken for the training set and 28 (25%) compounds were taken in for the test set. IC₅₀ (μM) values for EGFR inhibition were transformed into $-\log(\text{IC}_{50} \times 10^{-6})$ i.e. pIC₅₀ [30]. Since some compounds exhibited insignificant/no inhibition, such compounds were excluded from the present study. All the IC₅₀ values had been obtained using the in vitro MTT assay method [31, 32]. The IC₅₀ values of reference compounds were checked to ensure that no difference occurred between different groups. The pIC₅₀ values of the molecules under study spanned a wide range from 5 to 9.

D] Data Set

All computational work was performed on Apple workstation (8-core processor) using Vlife MDS QSAR plus software developed by Vlife Sciences Technologies Pvt Ltd, Pune, India, on windows XP operating system. All the compounds were drawn in Chem DBS using fragment database and then subjected to energy minimization using batch energy minimization method.

E] Molecular Modeling and Alignment

Conformational search were carried out by systemic conformational search method and lowest energy conformers were selected. All the compounds were aligned by template based method. The selection of template molecule for alignment was done by considering the following facts: a) the most active compound; b) the lead or commercial compound; c) the compound containing the greatest number of functional group [33, 34]. Generally, the low energy conformer of the most active compound is selected as a reference [35]. In the present study, all the compounds were aligned against minimum energy conformation of most active compound no.28 (Figure1) by using quinazoline nucleus as template shown in Figure 2.

F] Selection of Training and Test Set

The dataset of 126 molecules was divided into training and test set by Sphere Exclusion (SE) method for model 1, model 2 and model 3 having dissimilarities values of 8.2, 8.3 and 8.1 respectively with pIC₅₀ activity field as dependent variable and various 3D descriptors calculated for the compounds as independent variables.

Figure1: Reference molecule (28) used for alignment by template based alignment

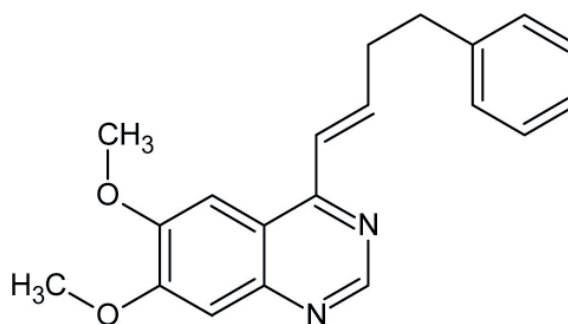
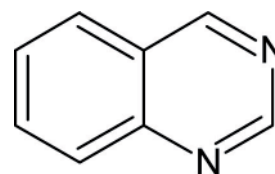


Figure 2: Quinazoline moiety as a template for alignment



G] Cross-Validation Using Weighted K-Nearest Neighbor.

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

a) Internal Validation.

(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. 1). The similarities were evaluated as the inverse of Euclidean distances between molecules (eq. 2) using only the subset of descriptors corresponding to the current trial solution

$$w_i = \frac{\text{Exp}(-d_j)}{\sum \text{Exp}(-d_j)}$$

k –Nearest neighbor

$$\hat{y}_i = \sum w_i y_i \quad (1)$$

$$d_{ij} = \sum_{k=1}^m (X_{ik} - X_{jk})^2)^{1/2} \quad (2)$$

(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. 3, where y_i and \hat{y}_i are the actual and pre-

dicted activities of the i th molecule, respectively, and y_{mean} is the average k -Nearest neighbor activity of all molecules in the training set. Both summations are over all molecules in the training set. Since the calculation of the pairwise 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.

$$q^2 = 1 - \frac{\sum (y_i - \hat{y}_i)^2}{\sum (y_i - y_{mean})^2} \quad (3)$$

b) External Validation.

The predicted r^2 ($pred_r^2$) value was calculated using eq 4, where y_i and \hat{y}_i are the actual and predicted activities of the i th 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.

$$pred_r^2 = 1 - \frac{\sum (y_i - \hat{y}_i)^2}{\sum (y_i - y_{mean})^2} \quad (4)$$

Both summations are over all molecules in the test set. Thus, the $pred_r^2$ value is indicative of the predictive power of the current model for external test set.

c) Randomization Test.

To evaluate the statistical significance of the QSAR model for an actual data set, we have employed a one-tail hypothesis testing [36–37]. 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 [36–37].

$$Zscore = \frac{(h - \mu)}{\sigma} \quad (5)$$

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 build by different random datasets. The probability (α) of significance of randomization test is derived by comparing Z score value with Z score

Figure 3: Hierarchical Graph Showing Uniform Distribution of Training and Test Set

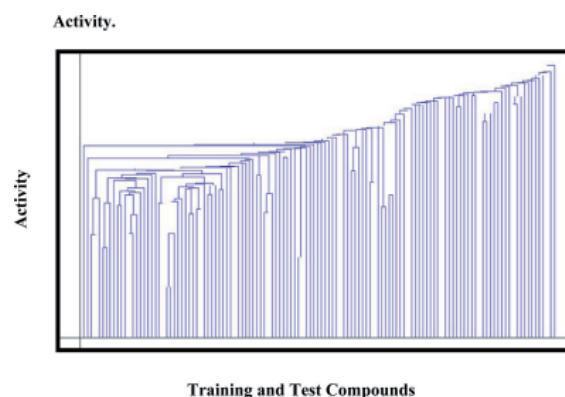


Figure 4: Graph of Actual vs. Predicted activities for training and test set molecules from the kNN-MFA model 1, A) Training set (Blue dots) B) Test Set (Yellow dots)

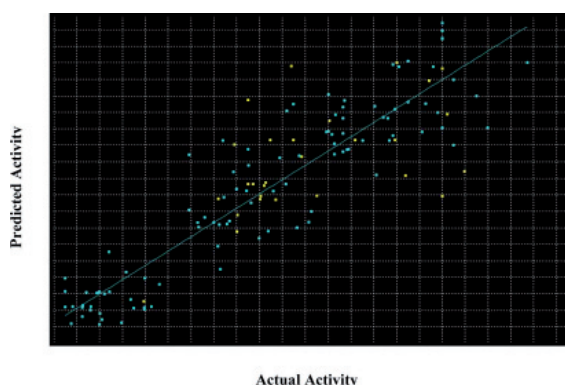
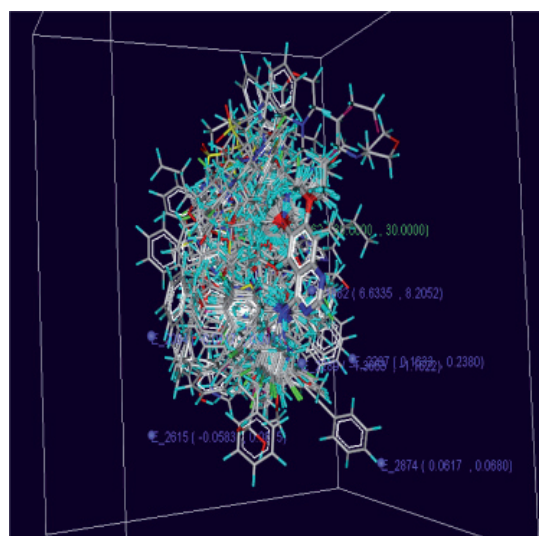


Figure 5: 3D- alignment of molecule with the important steric and electrostatic point contributing to the model with range of values shown in parenthesis



critical value, if Z score value is less than 4.0; otherwise it is calculated by the formula as given in the literature. For example, a Z score value greater than 3.10 indicates that there is a probability (α) of less than 0.001 that the QSAR model constructed for the real dataset is random. The randomization test suggests that all the developed models have a probability of less than 1% that the model is generated by chance.

EXPERIMENTAL

All the one hundred twenty six compounds were built on workstation of molecular modeling software VlifeMDS, which is a product Vlife Sciences Pvt Ltd., India [38]. We hereby report the models, as generated by kNN-MFA in conjunction with stepwise (SW) forward-backward variable selection methods shown in Table 2.

In the present kNN-MFA study, (-13.2343 to 19.1320) \times (-12.0268 to 15.04940) \times (-11.2513 to 15.4959) A⁰grid at the interval of 2.00 was generated around the aligned compounds. The steric and electrostatic interaction energies are computed at the lattice points of the grid using a methyl probe of charge +1 of Gasteiger-Marsili type. These interaction energy values are considered for relationship generation and utilized as descriptors to decide nearness between molecules. The QSAR models were developed using forward-backward variable selec-

tion method with pIC₅₀ activity field as dependent variable and physico-chemical descriptors as independent variable having cross-correlation limit of 1, 0.8 and 0.9 for mode 1, model 2 and model 3 respectively. Selection of test and training set was done by sphere exclusion method having dissimilarity value of 8.2, 8.3 and 8.1 for mode 1, model 2 and model 3 respectively. Variance cut off point was 0.0. Numbers of maximum and minimum neighbors were 5 and 2 respectively.

The method described above has been implemented in software, Vlife Molecular Design Suite (VlifeMDS), [38] which allows user to choose probe, grid size, and grid interval for the generation of descriptors. The variable selection methods along with the corresponding parameters are allowed to be chosen, and optimum models are generated by maximizing q^2 .

Steps Involved In kNN-MFA Method

1. Molecules are optimized before alignment optimization is done by MOPAC energy minimization and optimization is necessary process for proper alignment of molecules around template.
2. kNN-MFA method requires suitable alignment of given set of molecules, alignment are template based.
3. This is followed by generation of common rectangular grid around the molecules, the steric and electrostatic interaction energies are computed at the lattice

Table 2: Stastical Results of kNN-MFA method

Parameters	Model 1 (Dissimilarity value =8.2)	Model 2 (Dissimilarity value =8.3)	Model 3 (Dissimilarity value =8.1)
n	98	98	98
k	2	2	2
q ²	0.8463	0.7487	0.7802
pred_r ²	0.8029	0.7192	0.7412
pred_r ² se	0.6658	0.7126	0.6533
Z score	15.25775	11.5635	14.23195
best_ran_q ²	0.03060	0.1225	0.1092
α _ran_q ²	0.0001	0.001	0.001
Descriptors	E_1882 (6.6335, 8.2052)	E_1882 (6.6335, 8.2052)	S_1099 (11.5811, 12.2127)
	S_1462 (30.0000, 30.0000)	E_1515 (-0.1446, 0.2255)	E_2289 (-1.6082,-0.7847)
	E_2289 (-1.3653,-1.1622)	S_2892 (-0.0329, -0.0067)	S_1631 (-0.3820,-0.2460)
	E_2287 (0.1633, 0.2380)	S_1462 (30.0000, 30.0000)	E_1911 (-3.9843,-33125)
	E_2615 (-0.0583, 0.0815)	S_734 (-0.0194 -0.0098)	E_2272 (-0.3129,-0.1297)
	E_2874 (0.0617, 0.0680)		S_512 (-0.0214,-0.0165)
Vn	06	05	06

- points of the grid using a methyl probe of charge +1.
- The optimal training and test set were generated using sphere exclusion method.
 - Model was generated by various kNN methods, and models validated internally and externally by leave one out, external validation.
 - Predict the activity of test set of compounds.

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.

RESULT AND DISCUSSION

Training set of 98 and test set of 28 quinazoline derivatives having different substitution were employed. Following statistical measure was used to correlate biological activity and molecular descriptors: n = number of molecules, V_n = number of descriptors, k = number of nearest neighbor, df = degree of freedom, r^2 = coefficient of determination, q^2 = cross validated r^2 (by the leave-one out method), $pred_r^2 = r^2$ for external test set, $pred_r^2se$ = coefficient of correlation of predicted data 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 = the statistical significance parameter obtained by the randomization test.

Selecting training and test set by spherical exclusion method, Unicolumn statics shows that the max of the test is less than max of train set and the min of the test set is greater than of train set shown in Table 3, which is prerequisite analysis for further QSAR study. The above result shows that the test is interpolative i.e. derived within the min-max range of the train set. The mean and standard deviation of the train and test provides insight to the relative difference of mean and point density distribution of the two sets. In this case the mean in the test set higher than the train set shows the presence of relatively

more active molecules as compared to the inactive ones. Also the similar standard deviation in both set indicates that the spread in both the set with their respective mean is comparable.

The activity distribution graph shows the comparison between the activity of training and test set. It can be observed from Hierarchical Graph that the test set molecule activities lie within the range of training set, shown in Figure 3.

The observed and predicted pIC_{50} along with residual values for model 1 are shown in Table 1(A-W). The plot of observed vs. predicted activity is shown in Figure 4 From the plot it can be seen that kNN-MFA model is able to predict the activity of training set quite well (all points are close to regression line) as well as external.

During the kNN-MFA investigation, dissimilarity value for the selection of training and test by spherical exclusion method of range 8.0000 to 9.5000 were investigated. The dissimilarity value of 8.200 produced a significant result as compare to the 8.100 and 8.300 shown in Table 2. Further increase in resolution have produced decrease in model quality. From the Table 2 it was observed that the results were less sensitive to resolution of dissimilarity value.

It is known that the CoMFA method provides significant value in terms of a new molecule design, when contours of the PLS coefficients are visualized for the set of molecules. Similarly, the kNN-MFA models provide direction for the design of new molecules in a rather convenient way. The points which contribute to the kNN-MFA models 1 is displayed in Figure 5. The range of property values for the chosen points may aid in the design of new potent molecules (Figure 5). The range is based on the variation of the field values at the chosen points using the most active molecule and its nearest neighbor set.

The q^2 , $pred_r^2$, V_n and k value of kNN-MFA with model 1, 2 and 3 were (0.8463, 0.8029, 06/2) (0.7487, 0.7192, 05/2) and (0.7802, 0.7412, 06/2) respectively. Among these three methods, model 1 have better q^2 (0.8463) and $pred_r^2$ (0.8029) than other two models, model 1 correctly predicts activity 84.63% and 80.29% for the training and test set respectively. It uses 1 steric and 5 electronic descriptors with 2 k nearest neighbor to evaluate activity of new molecule.

The model is validated by $\alpha_ran_q^2 = 0.0001$, $best_ran_q^2 = 0.03060$, and $Z_score_ran_q^2 =$

Table 3: Unicolumn Statics of Training and Test Set

Unicolumn Statics	Average	Max	Min	Std. Deviation
For Training Set	6.4542	8.7447	4.6985	1.1079
For Test Set	6.8568	8.1938	5.3872	0.7417

15.25775. The randomization test suggests that the developed model have a probability of less than 1% that the model is generated by chance.

The kNN MFA models obtained by using all the three dissimilarity values showed that electrostatic and steric interactions plays major role in determining biological activity. S₁₄₆₂ in model 1, S₁₄₆₂, S₂₈₉₂, S₇₃₄ in model 2 and S₁₀₉₉, S₁₆₃₁, S₅₁₂ in model 3 are steric field descriptors similarly E₁₈₈₂, E₂₂₈₉, E₂₂₈₇, E₂₆₁₅, E₂₈₇₄ in model 1, E₁₅₁₅, E₁₈₈₂ in model 2, and E₂₂₈₉, E₁₉₁₁, E₂₂₇₂ model 3 are electrostatic field descriptors. It can also be noted that electrostatic descriptor E₁₈₈₂ and steric descriptor S₁₄₆₂ was common in the Model 1 and Model 2 using forward-backward variable selection method implying the significant role of these descriptors in electrostatic and steric field interaction for the structure activity relationship.

Negative value in electrostatic field descriptors indicates that negative electronic potential is required to increase activity and more electronegative substituents group is preferred in that position, positive range indicates that group that imparting positive electrostatic potential is favorable for activity so less electronegative group is preferred in that region. Similarly negative range in steric descriptors indicates that negative steric potential is favorable for activity and less bulky substituents group is preferred in that region, Positive value of steric descriptors reveals that positive steric potential is favorable for increase in activity and more bulky group is preferred in that region. n, number of observations (molecules); V_n, number of descriptors; k, number of nearest neighbors; q₂, cross-validated r² (by the leave-one out method); pred_r², predicted r² for the external test set; Zscore, the Zscore calculated by q₂ in the randomization test; best_ran_q₂, the highest q₂ value in the randomization test and α _ran_q₂, the statistical significance parameter obtained by the randomization test.

CONCLUSION

In conclusion, the model developed to predict the structural features of quinazoline to inhibit EGFR tyrosine kinase, reveals useful information about the structural features requirement for the molecule. In all three optimized models, Model 1 is giving very significant results. The master grid obtained for the various kNN-MFA models show that negative value

in electrostatic field descriptors indicates the negative electronic potential is required to increase activity and more electronegative substituents group is preferred in that position, positive range indicates that the group which imparts positive electrostatic potential is favorable for activity so less electronegative group is preferred in that region. Negative range in steric descriptors indicates that negative steric potential is favorable for activity and less bulky substituents group is preferred in that region, Positive value of steric descriptors reveals that positive steric potential is favorable for increase in activity and more bulky group is preferred in that region. On the basis of the spatial arrangement of the various shapes, electrostatic and steric potential contributions model proposed in this work is useful in describing QSAR of quinazoline derivatives as EGFR tyrosine kinase inhibitor and can be employed to design new derivatives of quinazoline with specific inhibitory activity.

ACKNOWLEDGEMENT

The authors would like to thank Director General, Department of Science and Technology, New Delhi for funding the project (Grant.No.SR/FT/LS-0083/2008) and Sardar Sangat Singh Longia, Secretary ASBASJSM College of Pharmacy for providing the necessary facilities.

REFERENCES

1. Kurup R, Garg C and Hansch. *Chem. Rev.* 2001; 101: 2573-2600..
2. Palmer BD, Krake AJ, Hartl BG, et al. *J. Med. Chem.* 1999; 42: 2373-2382.
3. Oblak M, Randic M and Solmajer T. *J. Chem. Inf. Comput. Sci.* 2000; 40: 994-1001.
4. Naumann T and Matter T. *J. Med. Chem.* 2002; 45: 2366-2378.
5. Bridges AJ, Zhou H, Cody DR, et al. *J. Med. Chem.* 1996; 39: 267-276.
6. Ma KA, Bower H, Lin G, et al. *Biochem. Pharmacol.* 2005;69:1785-1794.
7. Albuschat R, Lowe W, Weber M, et al. *Eur. J. Med. Chem.* 2004; 39: 1001-1011.
8. Cramer RD, Patterson DE and Bunce JD. *J. Am. Chem. Soc.* 1988;110: 5959.
9. Klebe G, Abraham U and Mietzner T. *J. Med. Chem.* 1994; 37: 24.
10. Baroni M, Costantino G, Cruciani G, et al. *An Advanced Chemometric Tool for Handling 3D-QSAR Problems. Quant. Struct.-Act. Relat.* 1993; 12: 9.
11. Cho SJ and Tropsha A. *J. Med. Chem.* 1995; 38: 1060.
12. Manjula SN, Noolvi MN, Parihar KV, et al. *Synthesis and anti-tumour activity of optically active thiourea and their 2-aminobenzothiazole derivatives: A novel class of anticancer agents. Eur. J. Med. Chem.* 2009: 1-7.

13. Badiger AM, Noolvi MN and Nayak PV. QSAR study of Benzthiazole derivatives as p56 lck inhibitors. *Lett. Drug .Des. Discov.* 2006; 3:550-560.
14. Golbraikh A and Tropsha A. *J. Chem. Inf. Comput. Sci.* 2003; 43: 144.
15. Sharaf MA, Illman DL and Kowalski BR. *Chemometrics*. Wiley, New York, 1986.
16. Holland, J. *Adaptation in Natural and Artificial Systems*, University of Michigan Press, 1975.
17. Kitano Y, Suzuki T, Kawahara E, et al. Synthesis and inhibitory activity of 4-alkenylquinazolines: Identification of new scaffolds for potent EGFR tyrosine kinase inhibitors. *Bioorg. Med. Chem. Lett.* 2007; 17: 5863-5867.
18. Chandregowda V, Kush AK and Reddy C. Synthesis and in vitro antitumour activities of novel 4 – Anilinoquinazolines derivatives. *Eur. J. Med. Chem.* 2009; 44: 3046-3055.
19. Mishani E, Abourbeh G, Rozen Y, et al. Novel carbon -11 labeled 4-dimethylamino but-2-enoic acid [4 –(phenylamino) –quinazoline-6-yl] –amides: potential PET bioprobes for molecular imaging of EGFR positive tumours. *Nucl. Med. Biol.* 2004; 31: 469-476.
20. Dissoki S, Aviv Y, Laky D, et al. The effect of the [18F] – PEG group on tracer qualification of [4-(phenylamino) – quinazoline-6-yl] –amide moiety – an EGFR putative irreversible inhibitor. *Int. J. Appl. Radiat. Isot.* 2007; 65: 1140-1151.
21. Abouzid K and Shouman S. Design, synthesis and in vitro anti-tumour activity of 4-aminoquinoline and 4-aminoquinazoline derivatives targeting EGFR tyrosine kinase. *Bioorg. Med. Chem.* 2008; 16: 7543-7551.
22. Shaul M, Abourbeh G, Jacobson O, et al. Novel iodine -124 labeled EGFR inhibitors as potential PET agents for molecular maging in cancer. *Bioorg. Med. Chem.* 2004; 12: 3421-3429.
23. Gibson KH, Grundy W, Godfrey AA, et al. Epidermal growth factor receptor tyrosine kinase: Structure activity relationship and anti-tumour activity of novel quinazoline. *Bioorg. Med. Chem. Lett.* 1997; 7: 2723-2728.
24. Zhang Y, Cockrill S, Guntrip S, et al. Synthesis and SAR of potent EGFR / erB2 dual inhibitors. *Bioorg. Med. Chem. Lett.*, 2004, 14: 111-114.
25. Liu L, Yuan T, Liu H, et al. Synthesis and biological evaluation of substituted 6-alkynyl – 4 – Anilinoquinazolines derivatives as potent EGFR inhibitors. *Bioorg. Med. Chem. Lett.* 2007, 17; 63:73-6377.
26. Bridges A, Cody D, Zhou H, et al. Enantioselective inhibition of the epidermal growth factor receptor tyrosine kinase by 4-(\square - phenethylamino) quinazolines. *Bioorg. Med. Chem.* 1995; 3: 1651-1656.
27. Lowe W and Luth AEJ. Synthesis of 4- (indole -3-yl) quinazolines – A new class off epidermal growth factor receptor tyrosine kinase inhibitors. *Eur. J. Med. Chem.* 2008; 43: 1478-1488.
28. Rachid Z, MacPhee M, Williams C, et al. Design and synthesis of new stabilized combi-triazenes for targeting solid tumours expressing the epidermal growth fator receptor (EGFR) or its closest homologue HER2. *Bioorg. Med. Chem. Lett.* 2009, article in press.
29. Murthy VS and Kulkarni VM. 3D-QSAR CoMFA and CoMSIA on Protein Tyrosine Phosphatase 1B Inhibitors. *Bioorg. Med. Chem.* 2002; 10: 2267-2282.
30. Sachan N, Kadam SS and Kulkarni VM. Human Protein Tyrosine Phosphatase 1B Inhibitors: QSAR by Genetic Function Approximation. *J. Enzyme Inhib. Med. Chem.* 2007; 22: 267-276.
31. Rusnak DW, Lackey K, Affleck K, et al. *Mol. Cancer. Ther.* 2001; 1:85.
32. Skehan P, Storeng R, Scudiero D, et al. *J. Natl. Cancer. Inst.* 1990; 82:1107.
33. Agarwal A and Taylor EW. 3-D QSAR for intrinsic activity of 5-HT1a receptor ligands by the method of comparative molecular field analysis. *J. Computat. Chem.* 1993; 14: 237–245.
34. Baurin N, Vangrevelinghe E and Allory LM. 3D-QSAR CoMFA study on imidazolinergic i2 ligands: a significant model through a combined exploration of structural diversity and methodology. *J. Med. Chem.* 2000; 43: 1109–1122.
35. Xu M, Zhang A, Han S and Wang L. Studies of 3D-Quantitative Structure-Activity Relationships on a set of nitroaromatic compounds: CoMFA advanced CoMFA and CoMSIA. *Chemosphere.* 2002; 48: 707-715.
36. Zheng W and Tropsha A. *J. Chem. Inf. Comput. Sci.* 2000; 40: 185.
37. Gilbert N, *Statistics*, Saunders WB Co, Philadelphia, PA. 1976.
38. VLifeMDS2.0, Molecular Design Suite, Vlife Sciences Technologies Pvt. Ltd., Pune, India, 2004, (www.vlifesciences.com).