Design of EGFR Kinase Inhibitors: A Ligand-Based Approach and Its Confirmation with Structure-Based Studies

Aparna Vema,[a] Sunil K. Panigrahi,[a] G. Rambabu,[c] B. Gopalakrishnan,[b,*] J. A. R. P. Sarma[c,*] and Gautam R. Desiraju[a,*]

[a]*School of Chemistry, University of Hyderabad, Hyderabad 500 046, India*
[b]*TATA Consultancy Services, 5-9-62 6th Floor, Khan Lateef Khan Estate, Fateh Maidan Road, Hyderabad 500 001, India*
[c]*gvk bioSciences Pvt. Ltd., #210 'My Home Tycoon', 6-3-1192 Begumpet, Hyderabad 500 016, India*

Abstract—Three-dimensional quantitative structure–activity relationship (3D-QSAR) models were developed for 100 anilinoquinazolines, inhibiting epidermal growth factor receptor (EGFR) kinase. The studies included molecular field analysis (MFA) and receptor surface analysis (RSA). The cross-validated r^2 (r^2_{cv}) values are 0.81 and 0.79 for MFA and RSA, respectively. The predictive ability of these models was validated by 28 test set molecules. The results of the best QSAR model were further compared with structure-based investigations using docking studies with the crystal structure of EGFR kinase domain. The results helped to understand the nature of substituents at the 6- and 7-positions, thereby providing new guidelines for the design of novel inhibitors.
© 2003 Elsevier Ltd. All rights reserved.

Introduction

Exposure of cells to growth factors, cytokines, hormones, neurotransmitters and other signaling molecules initiates reversible phosphorylation of tyrosine residues on regulatory proteins. This leads to a cascade of signal transduction events responsible for regulation of cell proliferation.[1] Of these signaling pathways, the epidermal growth factor receptor (EGFR) system has drawn attention because of findings that deregulation of this receptor system is a significant factor in the genesis or progression of several human cancers.[2] Drugs targeting EGFR fall into three main categories depending on the receptor region targeted: extracellular, intracellular and nuclear. Small molecule inhibitors that target the intracellular EGFR appear to be the most promising approach towards treating EGFR mediated cancers. These molecules act by binding either reversibly or irreversibly to the C-terminal tyrosine kinase domain of EGFR, thereby inhibiting autophosphorylation of the receptor and therefore activation.[3]

Anilinoquinazolines are the most developed class of drugs that inhibit EGFR kinase intracellularly.[4] These compounds are being studied actively by many research groups,[5–9] and as a result drug candidates of this class have already reached various phases of clinical trials (Fig. 1). Structure–activity relationship (SAR) studies reveal the nature of desirable substituents on the anilinoquinazoline moiety. Electron withdrawing, lipophilic substituents at the 3-position of aniline are favourable with Cl and Br being optimal. Similarly, electron-donating groups at the 6- and 7-positions of quinazoline are preferred.[10] Bulky substituents appear to be tolerated at the 6- and 7-positions.[11] 3D-QSAR studies by Hou et al. have described the region around the 7-position as more electronegative than that near the 6-position.[12] In the present study, 3D-QSAR studies were carried out on a series of EGFR kinase inhibitors in order to provide further insight into the key structural features required to design potential drug candidates of this class. The QSAR studies include Molecular Field Analysis (MFA) and Receptor Surface Analysis (RSA).[13,14] Here, we present our observations on the role of differential substitution at the 6- and 7-positions.

*Corresponding author. Tel.:+91-40-2301-0510x4828; fax:+91-40-2301-0567 e-mail: desiraju@uohyd.ernet.in (G.R. Desiraju); tel.: +91-40-5567-1012; fax: +91-40-5567-1111 e-mail: gopal@atc.tcs.co.in (B. Gopalakrishnan); tel.: +91-40-5551-9990x135; fax: +91-40-5562-6885 e-mail: sarma@gvkbio.com (J.A.R.P. Sarma).
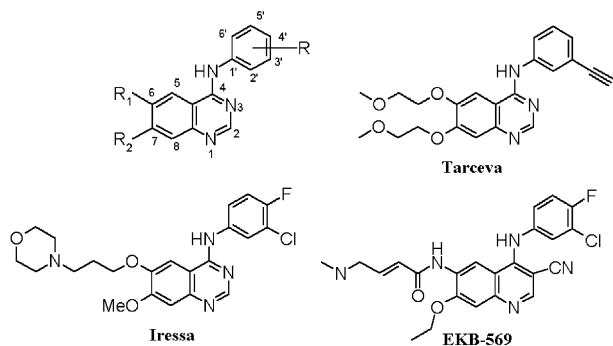
Figure 1. Atom numbering scheme for the anilinoquinazolines and some derivatives that are currently in clinical trials.
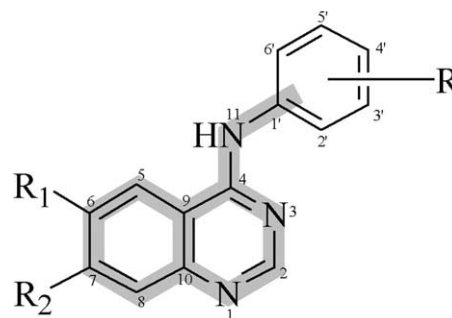
Figure 2. Anilinoquinazolines skeleton with atom numbering scheme. The shaded region includes the 12 atoms used for alignment.

Computational Methods

Selection of molecules

A set of 128 compounds reported as EGFR tyrosine kinase inhibitors was compiled (Table 1).[10,11,15–18] The compounds under study belong to six structurally different families. These groups include quinazolines (family A: 45 compounds), pyrido[3,2-*d*]pyrimidines (family B: 7 compounds], pyrido[4,3-*d*]pyrimidines (family C: 49 compounds), pyrido[3,4-*d*]pyrimidines (family D: 5 compounds), pyrido[2,3-*d*]pyrimidines (family E: 6 compounds), pyrimido[4,5-*d*]pyrimidines (family F: 16 compounds). The biological activities were converted into the corresponding pIC_{50} values ($-\log IC_{50}$), where IC_{50} value represents the drug in molar concentration that causes 50% inhibition of phopholipase Cγ1 phosphorylation by EGFR. All the IC_{50} values had been obtained using the same assay method.[19] The IC_{50} values of reference compounds were checked to ensure that no difference occurred between different groups. The pIC_{50} values of the molecules under study spanned a wide range from 5 to 11. About 75% of the 128 compounds that is, 100 compounds were selected as the training set and the remaining 25% that is, 28 compounds were included in the test set. The test set was selected based on the criteria given by Oprea et al.[20]
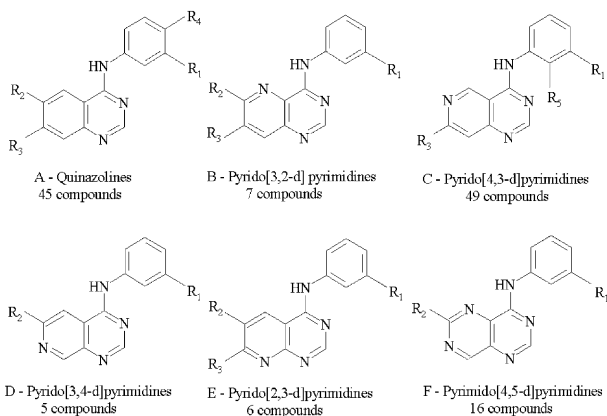
Molecular modeling

The X-ray crystal structure of the inhibitor (4-[3-hydroxyanilino]-6,7-dimethoxyquinazoline) bound to the active site of CDK2 (PDB entry 1DI8)[21] was used as a template to construct the three-dimensional models of all the compounds. All the structures were initially minimized by OFF methods using the steepest descent algorithm with a gradient convergence value of 0.001 kcal/mol. Partial atomic charges were calculated using the Gasteiger method.[22] Further geometry optimization was carried out for each compound with the MOPAC 6 package using the semi-empirical AM1 Hamiltonian.[23] The most active compound **43** was used as the template for the superimposition of the rest of the molecules. The alignment was carried out using the Align module in *Cerius*[2] [24] with 12 atoms selected for superimposition as shown in Figure 2.

MFA

MFA studies were performed with the QSAR module of *Cerius*[2]. The molecular field was created using proton and methyl groups as probes, which represent electrostatic and steric fields, respectively. These fields were sampled at each point of a regularly spaced grid of 1 Å. The co-ordinates of the MFA grid box were (−3.547, −6.216, −7.450) and (14.452, 17.783, 6.540) for the lower and upper corners, respectively. The total grid points generated were 7125. A number of spatial and structural descriptors such as polarizability, dipole moment, radius of gyration, number of rotatable bonds, molecular volume, principal moments of inertia, AlogP, number of hydrogen bond donors and acceptors and molar refractivity were considered along with the steric and electrostatic fields. Only 10% of the total descriptors whose variance was highest were considered for further analysis. Regression analysis was carried out using Genetic Partial Least Squares (G/PLS) method consisting of over 50,000 generations with a population size of 100. The optimal number of components was set to 5. An energy cutoff of ±30.0 kcal/mol was set for both steric and electrostatic contributions. The smoothing parameter, *d* was set to 1.0 to control the bias in the scoring factors between equations with different number of terms. The length of the final equation was fixed to nine terms. Cross-validation was performed with the leave-one-out procedure. PLS analysis was scaled, with all variables normalized to a variance of 1.0.

RSA
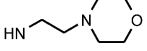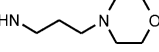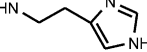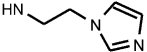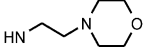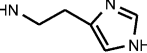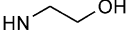
The same set of aligned molecules was considered for the generation of a receptor surface, using the *van der Waals field function*. The contribution of each molecule in the generation of receptor surface is proportional to its biological activity. During this process, various chemical properties, namely charge, electrostatic potential, hydrogen bonding propensity and hydrophobicity associated with each surface point were calculated automatically. Each molecule was docked into this generated receptor. The interaction energy between the molecule and the receptor was evaluated and added to the study table. Also, the interaction energies of each molecule with different probes such as methyl group (steric) and a

Table 1. Structures of 128 EGFR kinase inhibitors in this study

A - Quinazolines
45 compounds

B - Pyrido[3,2-d] pyrimidines
7 compounds

C - Pyrido[4,3-d]pyrimidines
49 compounds

D - Pyrido[3,4-d]pyrimidines
5 compounds

E - Pyrido[2,3-d]pyrimidines
6 compounds

F - Pyrimido[4,5-d]pyrimidines
16 compounds

Molecule no.	Class	Substitution					pIC_{50}
		R_1	R_2	R_3	R_4	R_5	
1	A	—	—	—	—	—	6.46
2	A	Me	—	—	—	—	6.04
3	A	Cl	—	—	—	—	7.63
4	A	Br	—	—	—	—	7.56
5	A	I	—	—	—	—	7.09
6	A	CF_3	—	—	—	—	6.23
7	A	Br	NO_2	—	—	—	6.04
8	A	Br	OMe	—	—	—	6.45
9	A	Br	—	NO_2	—	—	6.00
10	A	Br	—	OMe	—	—	8.00
11[a]	A	Br	OH	OH	—	—	9.76
12[a]	A	Br	NH_2	NH_2	—	—	9.92
13	A	F	—	—	—	—	7.25
14	A	—	OMe	—	—	—	7.25
15	A	—	NH_2	—	—	—	6.11
16	A	CF_3	NH_2	—	—	—	6.24
17	A	—	OMe	—	—	—	6.92
18	A	—	—	NH_2	—	—	7.00
19[a]	A	CF_3	—	NH_2	—	—	8.48
20	A	F	—	NO_2	—	—	5.21
21	A	Cl	—	NO_2	—	—	6.09
22	A	I	—	NO_2	—	—	6.26
23	A	—	OMe	OMe	—	—	7.53
24	A	F	OMe	OMe	—	—	8.42
25[a]	A	Cl	OMe	OMe	—	—	9.50
26[a]	A	I	OMe	OMe	—	—	9.05
27	A	CF_3	OMe	OMe	—	—	9.61
28[a]	A	Br	NHMe	—	—	—	8.39
29	A	Br	$N(Me)_2$	—	—	—	7.07
30	A	Br	NHCOOMe	—	—	—	7.92
31	A	Br	—	OH	—	—	8.32
32	A	Br	—	NHAc	—	—	7.39
33[a]	A	Br	—	NHMe	—	—	8.15
34	A	Br	—	NHEt	—	—	7.92
35[a]	A	Br	—	$N(Me)_2$	—	—	7.95
36	A	Br	NH_2	NHMe	—	—	9.16
37[a]	A	Br	NH_2	$N(Me)_2$	—	—	6.79
38	A	Br	NH_2	OMe	—	—	8.42
39	A	Br	NH_2	Cl	—	—	8.18
40	A	Br	NO_2	NHMe	—	—	7.16
41	A	Br	NO_2	OMe	—	—	7.82
42	A	Br	NO_2	Cl	—	—	7.60
43	A	Br	OEt	OEt	—	—	11.22
44	A	Br	OPr	OPr	—	—	9.76
45[a]	A	H	OMe	OMe	Br	—	9.01
46[a]	B	Br	—	—	—	—	7.46
47	B	Br	NH_2	—	—	—	8.11
48	B	Br	Cl	—	—	—	7.74
49[a]	B	Br	F	—	—	—	7.35
50	B	Br	NHMe	—	—	—	8.50
51	B	Br	$N(Me)_2$	—	—	—	8.01

(continued)

Table 1 (*continued*)

Molecule no.	Class	Substitution					pIC$_{50}$
		R$_1$	R$_2$	R$_3$	R$_4$	R$_5$	
52	B	Br	OMe	—	—	—	8.36
53[a]	C	Br	—	—	—	—	7.45
54	C	Br	—	NHAc	—	—	7.53
55	C	Br	—	F	—	—	7.88
56	C	Br	—	OMe	—	—	7.40
57[a]	C	—	—	NH$_2$	—	—	6.60
58[a]	C	NO$_2$	—	NH$_2$	—	—	7.39
59	C	—	—	NH$_2$	—	Br	6.61
60[a]	C	Br	—	NH$_2$	—	—	8.00
61	C	—	—	NH$_2$	Br	—	7.04
62	C	—	—	NH$_2$	CF$_3$	—	5.32
63	C	—	—	NH$_2$	—	OMe	5.43
64	C	OMe	—	NH$_2$	—	—	6.88
65[a]	C	—	—	NH$_2$	OMe	—	6.17
66[a]	C	—	—	NH$_2$	—	NH$_2$	5.27
67	C	N(Me)$_2$	—	NH$_2$	—	—	5.74
68	C	—	—	NH$_2$	N(Me)$_2$	—	5.31
69	C	F	—	NH$_2$	—	—	6.07
70	C	Cl	—	NH$_2$	—	—	6.92
71	C	OH	—	NH$_2$	—	—	7.15
72	C	Me	—	NH$_2$	—	—	7.39
73	D	Br	—	—	—	—	7.29
74	D	Br	Cl	—	—	—	7.39
75	D	Br	F	—	—	—	6.90
76	D	Br	OMe	—	—	—	8.58
77	D	Br	NH–CH$_2$–C$_6$H$_4$–OMe	—	—	—	8.63
78	E	Br	—	—	—	—	6.16
79[a]	E	Br	—	NH$_2$	—	—	6.02
80	E	Br	—	F	—	—	6.16
81	E	Br	—	NHMe	—	—	7.28
82	E	Br	—	N(Me)$_2$	—	—	6.48
83	E	Br	—	OMe	—	—	6.58
84	F	H	NHMe	—	—	—	7.88
85	F	Br	Cl	—	—	—	7.08
86[a]	F	Br	NH$_2$	—	—	—	8.82
87	F	Br	NHMe	—	—	—	9.11
88	F	Br	N(Me)$_2$	—	—	—	9.02
89	F	Br	OMe	—	—	—	8.42
90	F	Br	HN–CH$_2$CH$_2$–morpholine	—	—	—	9.09
91[a]	F	Br	HN–CH$_2$CH$_2$CH$_2$–morpholine	—	—	—	8.53
92	F	Br	HN–CH$_2$CH$_2$–(imidazol-4-yl)	—	—	—	9.60
93	F	Br	HN–CH$_2$CH$_2$–(imidazol-1-yl)	—	—	—	8.63
94[a]	F	Me	Cl	—	—	—	6.42
95	F	Me	NH$_2$	—	—	—	7.76
96	F	Me	NHMe	—	—	—	8.36
97[a]	F	Me	N(Me)$_2$	—	—	—	8.39
98	F	Me	HN–CH$_2$CH$_2$–morpholine	—	—	—	8.63
99	F	Me	HN–CH$_2$CH$_2$–(imidazol-4-yl)	—	—	—	8.52
100[a]	C	Br	HN–CH$_2$CH$_2$–OH	—	—	—	9.61

(*continued*)

Table 1 (*continued*)

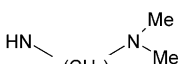| Molecule no.	Class	Substitution					pIC$_{50}$
		R$_1$	R$_2$	R$_3$	R$_4$	R$_5$	
101	C	Br	MeN⌒OH	—	—	—	8.58
102	C	Br	HN⌒OH, OH	—	—	—	9.03
103	C	Br	MeN⌒OH, OH	—	—	—	8.49
104	C	Br	N(⌒OH)(⌒OH)	—	—	—	7.85
105	C	Br	N⌒OH, ⌒OH	—	—	—	7.92
106[a]	C	Br	HN⌒N(Me)(Me)	—	—	—	7.34
107	C	Br	HN⌒⌒N(Me)(Me)	—	—	—	8.05
108[a]	C	Br	HN(CH$_2$)$_4$N(Me)(Me)	—	—	—	8.13
109	C	Br	HN(CH$_2$)$_5$N(Me)(Me)	—	—	—	8.07
110	C	Br	MeN(CH$_2$)$_2$N(Me)(Me)	—	—	—	7.39
111	C	Br	HN⌒morpholine	—	—	—	8.49
112[a]	C	Br	HN⌒⌒morpholine	—	—	—	8.72
113	C	Br	HN⌒⌒⌒morpholine	—	—	—	8.26
114	C	Br	HN⌒⌒N-piperazine-NMe	—	—	—	8.30
115	C	Br	HN⌒N(⌒OH)(⌒OH)	—	—	—	8.03
116	C	Br	HN⌒⌒N(⌒OH)(⌒OH)	—	—	—	8.92
117	C	Br	HN⌒NHNH$_2$	—	—	—	8.14
118[a]	C	Br	HN⌒imidazole	—	—	—	9.29
119	C	Br	HN⌒⌒imidazole	—	—	—	9.04
120	C	Br	HN⌒COOH	—	—	—	8.82

(*continued*)

Table 1 (continued)

Molecule no.	Class	Substitution					pIC$_{50}$
		R$_1$	R$_2$	R$_3$	R$_4$	R$_5$	
121	C	Br	HN–...–OH (C=O)	—	—	—	9.21
122	C	Br	HN–...–OH (C=O)	—	—	—	9.55
123	C	Br	MeN–...–OH (C=O)	—	—	—	7.79
124	C	Br	HN–...–S(=O)(=O)–OH	—	—	—	8.85
125	C	Me	HN–(CH$_2$)$_4$–N(Me)(Me)	—	—	—	8.26
126	C	Me	HN–...–N(morpholine)	—	—	—	8.03
127	C	Me	HN–...–N(piperazine)NMe	—	—	—	8.25
128	C	Me	HN–...–N(imidazole)	—	—	—	8.45

[a]Test set.

proton (electrostatic) positioned along the grid points throughout the receptor surface were added to the study table. Only 10% of the total descriptors whose variance was highest were considered as independent data to perform further analysis. Regression analysis was carried out using the G/PLS method as mentioned above.

Structure-based studies

Docking studies were carried out using *GOLD*.[25,26] The most active molecule **43** was docked into the active site of EGFR kinase (PDB entry 1M17).[27] Initially, hydro-

gen atoms were added to the protein, considering all the residues in their neutral form. The added hydrogens and water molecules were minimized while keeping all the heavy atoms fixed. The minimization followed steepest descent and conjugate gradient algorithms for at least 1000 iterations each using CHARMM[28] force field in *Insight II*.[29] The EGFR active site was defined using the inhibitor *erlotinib* {[6,7-bis(2-methoxy-ethoxy)quinazoline-4-yl]-(3-ethynylphenyl)amine} and included all amino acid residues within 10 Å radius from the centre of the quinazoline ring. Docking was performed using the default parameters of *GOLD*.[30] The protein active site conformations for the 10 best conformations of the ligand were retrieved for further analysis.

Table 2. Statistical details for MFA and RSA models

	MFA	RSA
r^2_{cv} [a]	0.81	0.79
r^{2} [b]	0.84	0.85
PRESS[c]	23.75	26.35
N[d]	5.0	5.0
r^2_{pred} [e]	0.761	0.841
r^2_{bs} [f]	0.845	0.824
LSE[g]	0.188	0.190

[a]Cross-validated r^2.
[b]Conventional r^2.
[c]Predicted sum of squared residuals.
[d]Number of components.
[e]Predictive r^2.
[f]Bootstrap r^2.
[g]Least square error.

Results and Discussion

The statistical details of the 3D QSAR models are given in Table 2. The actual and predicted activities, obtained from MFA and RSA 3D-QSAR models, of both the training and test sets are listed in Tables 3 and 4, respectively. Scatter-plots of actual versus predicted activities for both training and test set molecules obtained from MFA and RSA are shown in Figure 3.

MFA

An MFA model with a cross-validated r^2 (r^2_{cv}) value = 0.81 and a conventional r^2 = 0.84 was developed. The higher r^2_{cv} of MFA than that of RSA makes it pos-
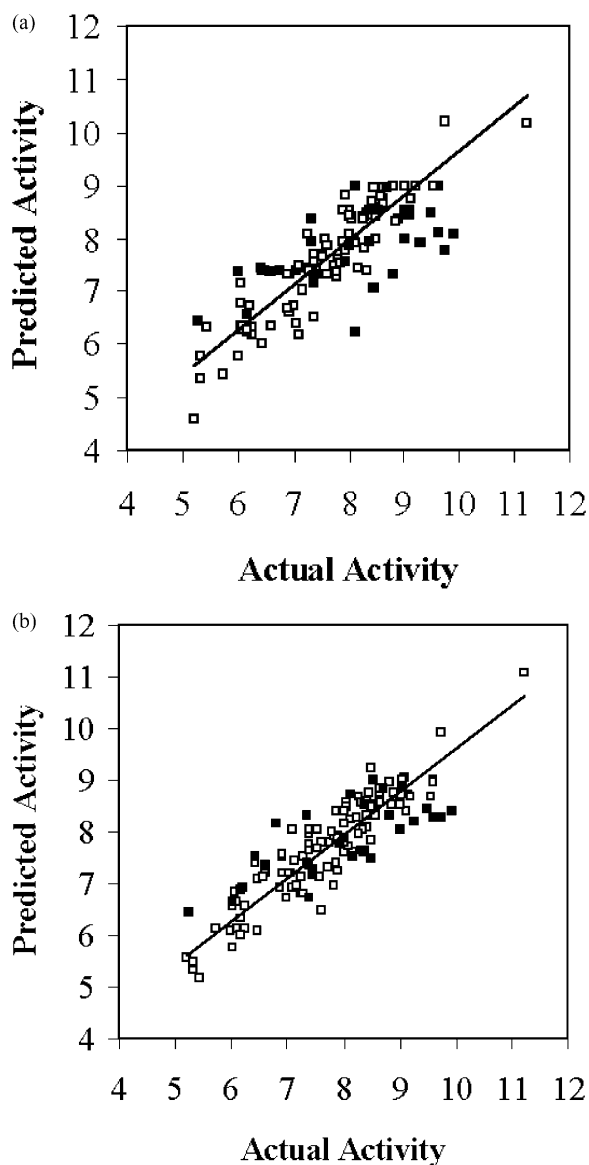
Table 3. Actual and predicted activities of the training set molecules

S. no.	Molecule ID	Actual activity	Predicted activity	
		pIC_{50}	MFA	RSA
1	1	6.46	6	6.09
2	2	6.04	7.15	6.57
3	3	7.63	7.32	6.5
4	4	7.56	7.96	7.11
5	5	7.09	6.19	6.93
6	6	6.23	6.73	6.93
7	7	6.04	6.35	5.76
8	8	6.45	7.4	7.4
9	9	6	5.78	6.07
10	10	8	8.38	7.67
11	13	7.25	8.08	6.83
12	14	7.25	7.45	7.12
13	15	6.11	6.28	6.63
14	16	6.24	6.19	6.52
15	17	6.92	6.61	7.53
16	18	7	6.73	6.74
17	20	5.21	4.58	5.57
18	21	6.09	6.33	6.11
19	22	6.26	6.32	6.13
20	23	7.53	7.67	8.03
21	24	8.42	8.43	8.07
22	27	9.61	8.99	8.99
23	29	7.07	7.41	7.22
24	30	7.92	8.51	7.23
25	31	8.32	7.81	7.63
26	32	7.39	7.67	7.94
27	34	7.92	7.76	7.9
28	36	9.16	8.75	8.74
29	38	8.42	8.68	8.32
30	39	8.18	7.45	7.6
31	40	7.16	7.03	6.96
32	41	7.82	7.59	6.96
33	42	7.6	7.84	7.79
34	43	11.22	10.16	11.09
35	44	9.76	10.2	9.91
36	47	8.11	7.98	7.72
37	48	7.74	7.48	7.34
38	50	8.5	8.46	7.85
39	51	8.01	8.51	7.82
40	52	8.36	8.47	8.03
41	53	7.45	7.64	7.68
42	55	7.88	7.93	7.4
43	56	7.4	7.45	8.05
44	59	6.61	6.35	7.19
45	61	7.04	6.39	7.21
46	62	5.32	5.33	5.33
47	63	5.43	6.31	5.17
48	64	6.88	6.7	6.9
49	67	5.74	5.41	6.12
50	68	5.31	5.74	5.05
51	69	6.07	6.75	6.83
52	70	6.92	7.31	7.2
53	71	7.15	7.45	7.46
54	72	7.39	7.29	7.36
55	73	7.29	7.33	7.52
56	74	7.39	7.54	7.64
57	75	6.9	7.3	7.58
58	76	8.58	8.56	8.35
59	77	8.63	8.72	8.58
60	78	6.16	6.2	6.02
61	80	6.16	6.26	6.33
62	81	7.28	7.41	6.82
63	82	6.48	7.38	7.09
64	83	6.58	7.34	7.13
65	84	7.88	8.52	8.38
66	85	7.08	7.47	8.05
67	87	9.11	8.52	8.38
68	88	9.02	8.51	8.58
69	89	8.42	8.46	8.62
70	90	9.09	8.45	9.04
71	92	9.6	9	8.95
72	93	8.63	8.65	8.75
73	95	7.76	7.25	7.79
74	96	8.36	7.41	7.59
75	98	8.63	8.94	8.85
76	99	8.52	8.42	8.51
77	101	8.58	8.77	8.32
78	102	9.03	8.43	8.54
79	103	8.49	7.99	9.23
80	104	7.85	7.51	7.9
81	105	7.92	8.81	7.87
82	107	8.05	8.38	8.55
83	109	8.07	8.37	8.47
84	110	7.39	6.52	7.78
85	111	8.49	8.38	8.48
86	113	8.26	8.38	8.68
87	114	8.3	8.38	8.57
88	115	8.03	8.07	8.38
89	116	8.92	8.37	8.77
90	117	8.14	7.9	8.22
91	119	9.04	8.99	8.99
92	120	8.82	8.98	8.97
93	121	9.21	8.97	8.67
94	122	9.55	8.97	8.67
95	123	7.79	7.35	8
96	124	8.85	8.33	8.54
97	125	8.26	8.37	7.97
98	126	8.03	8.45	8.18
99	127	8.25	8.37	8.27
100	128	8.45	8.95	8.76

Table 4. Actual and predicted activities of the test set molecules

| S. no.	Molecule ID	Actual activity	Predicted activity	
		pIC_{50}	MFA	RSA
1	11	9.76	7.76	8.29
2	12	9.92	8.08	8.39
3	19	8.48	7.06	7.49
4	25	9.5	8.47	8.43
5	26	9.05	8.5	8.89
6	28	8.39	8.54	7.65
7	33	8.15	6.22	7.51
8	35	7.95	7.54	7.75
9	37	6.79	7.4	8.15
10	45	9.01	7.96	8.04
11	46	7.46	7.31	7.18
12	49	7.35	7.95	7.42
13	53	7.45	7.3	7.28
14	57	6.6	7.33	7.37
15	58	7.39	7.14	6.71
16	60	8	7.86	7.89
17	65	6.17	6.55	6.87
18	66	5.27	6.43	6.46
19	79	6.02	7.35	6.66
20	86	8.82	7.32	8.31
21	91	8.53	8.51	9.02
22	94	6.42	7.45	7.53
23	97	8.39	7.95	8.52
24	100	9.61	8.1	8.27
25	106	7.34	8.36	8.33
26	108	8.13	9	8.74
27	112	8.72	8.93	8.86
28	118	9.29	7.9	8.21

(a)

(b)

Figure 3. Scatter-plots of actual verses predicted activity for both training (□) and test sets (■) using (a) MFA (b) RSA.

sible to interpolate the order of activities of the training set molecules. The steric (CH_3) and electrostatic (H^+) descriptors in the QSAR equation of MFA (**1**) specify the regions where variations in the structural features (steric or electrostatic), of different compounds in the training set, lead to increased or decreased activities. The number accompanying descriptors, in eq 1 represents its position in the three-dimensional MFA grid.

$$
\begin{aligned}
\text{Activity} = {} & 6.3273 + 0.018513\,(H^+/1223) \\
& + 0.03961(CH_3/2871) - 0.038226 \\
& \times (H^+/1194) - 0.028351(H^+/5124) \\
& + 0.03765(CH_3/\,2315) + 0.035301 \\
& \times (H^+/1239) + 0.023949(CH_3/5111) \\
& + 0.038927(H^+/5139) \qquad\qquad (1)
\end{aligned}
$$

The presence of ($H^+/1239$) and ($H^+/1223$) descriptors in the vicinity of the *meta* position of the aniline ring emphasizes the electrostatic environment required at this position (Fig. 4). Any moderately electron withdrawing group like a halogen at this position could enhance the EGFR kinase inhibitory activity. Our observation is in agreement with the SAR studies of Rewcastle et al.[10] The presence of steric descriptor ($CH_3/2315$) close to this position indicates the importance of steric interactions. A methyl group at this position would satisfy only the steric requirement but not the electrostatic requirement while Br, an isostere of methyl, would fulfill both steric and electrostatic stipulations. Hence, the activity of molecules **4**, **60**, **85**, **86**, **87**, **88** and **90** are higher than that respectively of **2**, **72**, **94**, **95**, **96**, **97** and **98**, wherein Br replaces a methyl group. The presence of two electrostatic descriptors near the 7-position of the quinazoline nucleus, namely ($H^+/5124$) with a negative coefficient and ($H^+/5139$) with a positive coefficient in the final QSAR eq 1, describes a subtle balance of electrostatic parameters
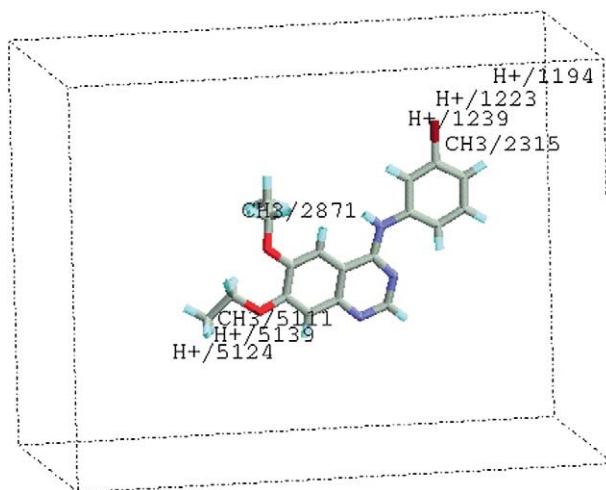
Figure 4. Molecule **43** enclosed inside the grid with 3D points of the QSAR equation.

required at this position. The need for moderate electron donating groups with appropriate steric parameters is evident with the trend shown by the molecules **79** $<$ **82** $<$ **83**.

A high value of r_{cv}^2 alone, however, is an insufficient criterion for a QSAR model to be robust and highly predictive.[31] The predictive power of the model was therefore validated with the test set molecules. The predictive power of the model generated, is calculated by

$$r_{pred}^2 = \frac{SD\text{-}PRESS}{SD} \qquad (2)$$

where SD is the sum of squared deviations between the biological activities of each molecule and the mean activity of the training set molecules and PRESS is the sum of squared deviations between the predicted and actual activity values for every molecule in the test set. The prediction of the model was reasonably good with a predictive r^2 (r_{pred}^2) value of 0.76.

RSA

A receptor surface model with $r_{cv}^2 = 0.79$ and conventional $r^2 = 0.85$ was developed. The QSAR equation generated by RSA is given in eq 3

$$
\begin{aligned}
\text{Activity} = {} & 6.2786 - 7.34427(\text{Ele}/1391) + 3.30427 \\
& \times (\text{VDW}/4591) + 10.2277(\text{VDW}/2861) \\
& - 2227.15(\text{Ele}/3761) + 6.31853 \\
& \times (\text{VDW}/3761) - 4.42666(\text{Ele}/3841) \\
& + 1.73629(\text{Ele}/4081) + 0.005889(\text{MW}) \qquad (3)
\end{aligned}
$$

The equation consists of eight molecular field descriptors. The positioning of these descriptors, indicated by a number along with the descriptors, on the model explains the nature of the substituents required. At the meta position of the anilino group, the presence of both steric (VDW/2861) and electrostatic (Ele/3761) descriptors explains the need of a bulky and electron withdrawing substituent. Br, satisfying both these criteria, is the right substituent at this position. The same was apparent from MFA discussed earlier. The steric descriptor (VDW/4591) present near the 6-position of the quinazoline ring indicates that bulky linear substituents are necessary at this position. Further, the electrostatic descriptors (Ele/1391) and (Ele/3841) present at the 6- and 7-positions of the quinazoline ring, respectively, indicate that these linear substituents should be electron withdrawing in nature. The presence of O or N atoms at these positions allows linear substituents to orient towards the descriptors. Inclusion of MW as one of the crucial descriptors in the final QSAR eq 3 is in accordance with the Lipinski's rule of five.[32]

When charge is mapped on to the receptor surface, it shows a red colour around the *meta* position of the aniline ring (Fig. 5). This indicates that an electron-withdrawing group at this position would be essential. Further, there are two red regions around the 1- and 3-positions of the quinazoline ring and a blue region at the 2-position. This indicates that electronegative atoms are vital at the 1- and 3-positions. The electron withdrawing nature of these atoms makes the adjacent region electron deficient. Thus, the region near the 2-position is blue (electropositive). Similarly, a blue region is in juxtaposition with the red region at the 4-position, indicating that the anilino nitrogen is crucial. The RSA model was further validated using the same test set molecules as in MFA. RSA predicted 93% of molecules correctly with an error less than 1.5.

MFA verses RSA

The r_{pred}^2 values, obtained for both the models are listed in Table 2. RSA exhibited higher predictivity with r_{pred}^2 values 0.84, while MFA had a lower r_{pred}^2 value 0.76. However, MFA had a higher value of r_{cv}^2. This r_{cv}^2 of MFA indicates good internal consistency but the lower r_{pred}^2 value implies that it cannot predict molecules outside the training set better than RSA. While MFA takes into consideration only steric and electrostatic parameters, RSA considers hydrophobicity also along with these parameters. Inclusion of this additional parameter to develop the model could be one of the reasons of its higher predictivity over MFA.

Structural requirements at the 6- and 7-positions

In the receptor surface model, the presence of steric descriptor (VDW/4591) only near the 6-position but not the 7-position shows that the substitutional requirements for these two positions are different. It is quite evident from this that bulky substituents are tolerable at the 6-position rather than at the 7-position. Also, when the receptor surface model was mapped with various properties like charge, hydrophobicity and hydrogen bonding capability, certain variations in the structural features of the substituents at 6- and 7-positions were observed.

1. Pale red colour near the 6-position as in Figure 5
2. Light brown colour (less hydrophobic) at the 6-position as in Figure 6
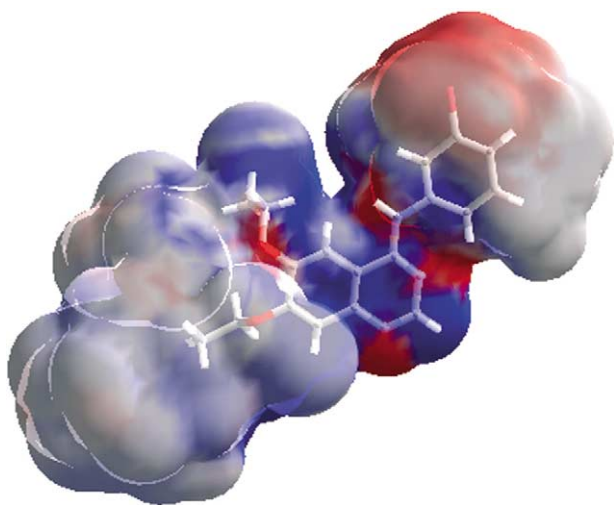3. Pale green colour (H-bond acceptor) at the 6-position as in Figure 7.

All the above features indicate that substituents at the 6-position should be more electronegative, hydrophilic and possess greater H-bond acceptor propensity than substituents at the 7-position. Though the differences at these positions are seemingly minimal, it appears to be significant in ligand binding. This is contrary to the reports of Rewcastle et al.,[10] generally stating that electron-donating groups at both the 6- and 7-positions are

acceptable. The above observations clearly differentiate the nature of substituents at these positions and are different from those reported by Hou et al.[12] Till date such differences have not been described, although some of the promising drug candidates, like Iressa and EKB-569, satisfy these criteria.
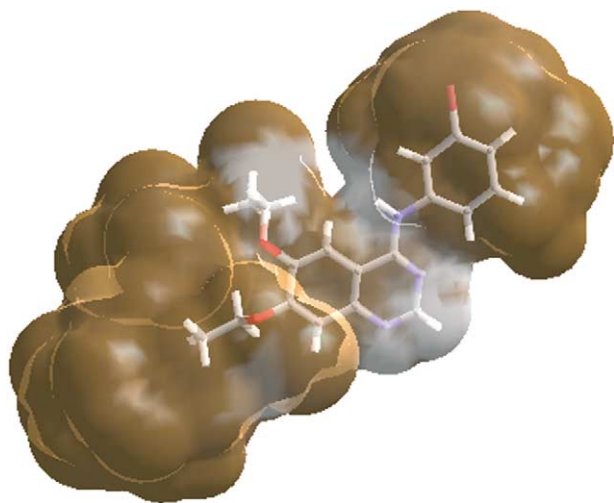
Correlation of RSA results with docking

The results of RSA were compared with docking studies using the crystal structure of EGFR kinase domain (1M17.pdb) that was released during the progress of this work. The most active molecule **43** was docked into the active site using *GOLD*. The active site of protein along with the docked molecule is shown in Figure 8. The molecule sits in the active site in such a way that its aniline ring protrudes inside one of the hydrophobic pockets, while the substituent at 7-position faced another hydrophobic region. Also, the substituent at the
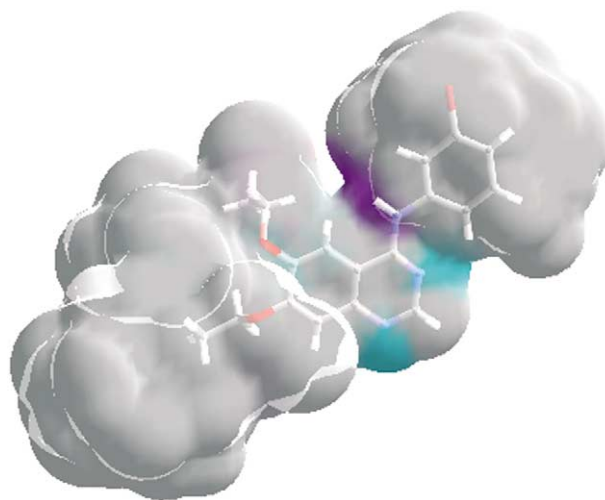
6-position is directed towards the hydrophilic region that is occupied by the phosphate chain of ATP (substrate). The Gauss–Connolly surface (computed in *MOE*[33]) of the active site (Fig. 9) also shows a slight variation in the regions around the 6- and 7-positions. A pale blue (electropositive) region is seen near the 7-position, while the region around 6-position is pale red (electronegative). This indicates that the receptor

Figure 7. Receptor surface model with hydrogen bonding property mapped onto it. Hydrogen bond donor property is shown in purple and hydrogen bond acceptor property is shown in green. Molecule **43** is placed within the receptor surface generated.
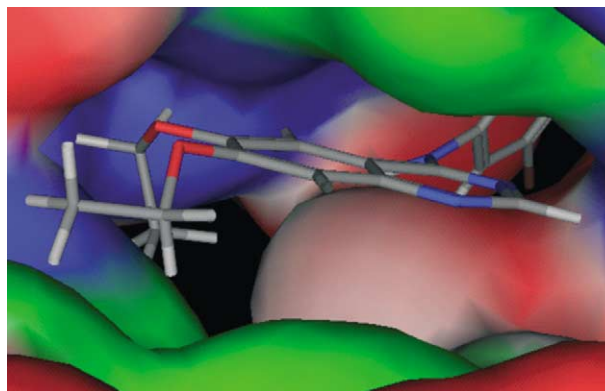

Figure 5. Receptor surface model with charge mapped onto it. Positive regions are shown in blue and negative regions shown in red. Molecule **43** is placed within the receptor surface generated.

Figure 8. Active site of EGFR with molecule **43** docked into it. Hydrophilic regions are shown in blue, hydrophobic regions in green and solvent accessible surface in red.
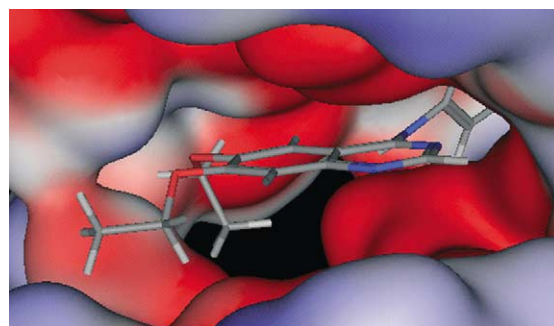

Figure 6. Receptor surface model with hydrophobic property mapped onto it. Hydrophobic regions are shown in brown and hydrophilic regions in white. Molecule **43** is placed within the receptor surface generated.

Figure 9. Active site of EGFR with molecule **43** docked into it. Blue regions are electropositive and red regions indicate electronegative character.

surface around the 6-position is relatively electronegative when compared to that around the 7-position. Thus, our observations concur well with that of the crystal structure.

Conclusions

The MFA and RSA 3D-QSAR models developed for anilinoquinazoline derivatives provide similar information about the structural requirements for their activity with RSA exhibiting a better predictive capability than MFA. The inclusion of hydrophobic field in RSA in addition to the normal steric and electrostatic parameters used in MFA could be one of the reasons for its higher predictive capability. The model differentiates the 6- and 7-positions and indicates that substituents at the 6-position should have higher electronegativity, hydrophilicity, hydrogen bonding propensity and bulky in size. These observations were compared and confirmed with docking studies carried out with the most active molecule **43** of the series. The results could help in the design of new molecules with promising activity in the future.

Acknowledgements

References and Notes

1. Traxler, P.; Lydon, N. *Drugs of the Future* **1995**, *20*, 1261.
2. Levitzki, A.; Gazit, A. *Science* **1995**, *267*, 1782.
3. Pedersen, M. W.; Poulsen, H. S. *Sci. Med.* **2002**, *8*, 206.
4. Bridges, A. J. *Chem. Rev.* **2001**, *101*, 2541.
5. Traxler, P.; Bold, G.; Buchdunger, E.; Caravatti, G.; Furet, P.; Manley, P.; Reilly, T. O.; Wood, J.; Zimmermann, J. *Med. Res. Rev.* **2001**, *21*, 499.
6. (a) Smaill, J. B.; Showalter, H. D. H.; Zhou, H.; Bridges, A. J.; McNamara, D.; Fry, D. W.; Nelson, J. M.; Sherwood, V.; Vincent, P. W.; Roberts, B. J.; Elliott, W. L.; Denny, W. A. *J. Med. Chem.* **2001**, *44*, 429. (b) Smaill, J. B.; Rewcastle, G. W.; Loo, J. A.; Greis, K. D.; Chan, O. H.; Reyner, E. L.; Lipka, E.; Showalter, H. D. H.; Vincent, P. W.; Elliott, W. L.; Denny, W. A. *J. Med. Chem.* **2000**, *43*, 1380. (c) Smaill, J. B.; Palmer, B. D.; Rewcastle, G. W.; Denny, W. A.; McNamara, D. J.; Dobrusin, E. M.; Bridges, A. J.; Zhou, H.; Showalter, H. D. H.; Winters, R. T.; Leopold, W. R.; Fry, D. W.; Nelson, J. M.; Slintak, V.; Elliot, W.; Roberts, B. J.; Vincent, P. W.; Patmore, S. J. *J. Med. Chem.* **1999**, *42*, 1803.
7. Kurup, A.; Garg, R.; Hansch, C. *Chem. Rev.* **2001**, *101*, 2573.
8. Gibson, K. H.; Grundy, W.; Godfrey, A. A.; Woodburn, J. R.; Ashton, S. E.; Curry, B. J.; Scarlett, L.; Barker, A. J.; Brown, D. S. *Bioorg. Med. Chem. Lett.* **1997**, *7*, 2723.
9. Wissner, A.; Floyd, M. B.; Rabindran, S. K.; Nilakantan, R.; Greenberger, L. M.; Shen, R.; Wang, Y. F.; Tsou, H. R. *Bioorg. Med. Chem. Lett.* **2002**, *12*, 2893.
10. Rewcastle, G. W.; Denny, W. A.; Bridges, A. J.; Zhou, H.; Cody, D. R. *J. Med. Chem.* **1995**, *38*, 3482.
11. Bridges, A. J.; Zhou, H.; Cody, D. R.; Rewcastle, G. W.; McMichael, A.; Showalter, H. D. H.; Fry, D. W.; Kraker, A. J.; Denny, W. A. *J. Med. Chem.* **1996**, *39*, 267.
12. Hou, T.; Zhu, L.; Chen, L.; Xu, X. *J. Chem. Inf. Comput. Sci.* **2003**, *43*, 273.
13. Hahn, M. *J. Med. Chem.* **1995**, *38*, 2080.
14. Hahn, M.; Rogers, D. *J. Med. Chem.* **1995**, *38*, 2091.
15. Thompson, A. M.; Bridges, A. J.; Fry, D. W.; Kraker, A. J.; Denny, W. A. *J. Med. Chem.* **1995**, *38*, 3780.
16. Rewcastle, G. W.; Palmer, B. D.; Thompson, A. M.; Bridges, A. J.; Cody, D. R.; Zhou, H; Fry, D. W.; McMichael, A.; Denny, W. A. *J. Med. Chem.* **1996**, *39*, 1823.
17. Rewcastle, G. W.; Bridges, A. J.; Fry, D. W.; Rubin, J. R.; Denny, W. A. *J. Med. Chem.* **1997**, *40*, 1820.
18. Thompson, A. M.; Murray, D. K.; Elliott, W. L.; Fry, D. W.; Nelson, J. A.; Showalter, H. D. H.; Roberts, B. J.; Vincent, P. W.; Denny, W. A. *J. Med. Chem.* **1997**, *40*, 3915.
19. Fry, D. W.; Kraker, A. J.; McMicheal, A.; Ambroso, L. A.; Nelson, J. M.; Leopold, W. R.; Connors, R. W.; Bridges, A. J. *Science* **1994**, *265*, 1093.
20. Oprea, T. I.; Waller, C. L.; Marshall, G. R. *J. Med. Chem.* **1994**, *37*, 2206.
21. Shewchuk, L.; Hassell, A.; Wisley, B.; Rocque, W.; Holmes, W.; Veal, J.; Kuyper, L. F. *J. Med. Chem.* **2000**, *43*, 133.
22. Gasteiger, J.; Marsili, M. *Tetrahedron* **1980**, *36*, 3291.
23. Dewar, M. J. S.; Zoebisch, E. G.; Healy, E. F.; Stewart, J. J. P. *J. Am. Chem. Soc.* **1985**, *107*, 3902.
24. Cerius² Molecular Modelling Program Package, Molecular Simulations (Accelrys) Inc.: San Diego, CA 92121-3752. USA
25. Jones, G.; Willett, P.; Glen, R. C.; Leach, A. R.; Taylor, R. *J. Mol. Biol.* **1997**, *267*, 727.
26. Jones, G.; Willett, P.; Glen, R. C. *J. Mol. Biol.* **1995**, *245*, 43.
27. Stamos, J.; Sliwkowski, M. X.; Eigenbrot, C. *J. Biol. Chem.* **2002**, *277*, 46265.
28. Brooks, B. R.; Bruccoleri, R. E.; Olafson, B. D.; States, D. J.; Swaminathan, S.; Karplus, M. *J. Comp. Chem.* **1983**, *4*, 187.
29. *INSIGHT II 97 Molecular Modelling Program Package*; Molecular Simulations (Accelrys) Inc.: San Diego, CA, 1997.
30. GOLD 2.0, Cambridge Crystallographic Data Centre, 12 Union Road, Cambridge CB2 1EZ, UK.
31. Golbraikh, A.; Tropsha, A. *J. Mol. Graph. Mod.* **2002**, *20*, 269.
32. Lipinski, C. A.; Lombardo, F.; Dominy, B. W.; Feeney, P. J. *Adv. Drug. Del. Rev.* **1997**, *23*, 3.
33. *MOE 2002.03;* Chemical Computing Group Inc.: Montreal H3A 2R7, Canada.