Available online at www.sciencedirect.com
ScienceDirect
European Journal of Medicinal Chemistry 43 (2008) 781–791
EUROPEAN JOURNAL OF MEDICINAL CHEMISTRY
http://www.elsevier.com/locate/ejmech

Original article

Towards predictive inhibitor design for the EGFR autophosphorylation activity

Amor A. San Juan [a,b,*]

[a] Life Science Division, Korea Institute of Science and Technology, P.O. Box 131, Cheongryang, Seoul 130-650, South Korea
[b] School of Science, University of Science and Technology, 52 Eoeun-dong, Yuseong-gu, Daejeon 305-333, South Korea

Received 25 January 2007; received in revised form 13 June 2007; accepted 15 June 2007
Available online 5 July 2007

Abstract

Inhibition of the epidermal growth factor receptor (EGFR) tyrosine kinase is one among the pivotal targets for the treatment of cancer. The structural investigation directly halting the EGFR autophosphorylation is expected to give insights into alternatively blocking the aberrant activity of EGFR. The three-dimensional quantitative structure–activity relationship (3D-QSAR) models were developed from the systematic search conformer-based alignment method. Models derived from the training set of 95 compounds showed superior CoMFA as compared with CoMSIA (CoMFA: $q^2 = 0.50$, $r^2 = 0.74$, $N = 5$, $F = 48.83$, $r^2_{\text{pred}} = 0.56$ while CoMSIA: $q^2 = 0.48$, $r^2 = 0.62$, $N = 2$, $F = 72.70$, $r^2_{\text{pred}} = 0.51$). Validation of the models by test set prediction of 26 compounds was in good agreement with the experimental results. Further validation by molecular docking superimposition into the 3D-QSAR contour maps was found in agreement with each other. We identified that the structural modification of compound **19** by attachment of a bulky group on pyrrole ring along with an electronegative group on quinazoline ring and a hydrogen-bond donor on methyl formate opens a new avenue towards the optimization of novel chemical entities to develop potent inhibitors for EGFR autophosphorylation.

Keywords: 3D-QSAR; Autophosphorylation; EGFR; Ligand design; Lung cancer

1. Introduction

The epidermal growth factor receptor (EGFR) was the first receptor tyrosine kinase (RTK) discovered [1]. EGFR family (EGFR/ErbB1/HER1, ErbB2/HER2, ErbB3/HER3 and ErbB4/HER4) is composed of an extracellular ligand-binding domain, followed by a single transmembrane domain and a cytoplasmic domain containing a conserved protein tyrosine kinase [2]. EGFR, ErbB2 and ErbB4 contain the catalytically competent kinase domains and can form heterodimers with each other [3]. ErbB3 contains an inactive kinase domain, but it can pair with and activate the other members of the family [3]. The critical roles of EGFR members include the regulation of cell proliferation, differentiation and migration [4]. The binding of EGFR to its cognate ligands leads to autophosphorylation of RTK followed by the activation of signal transduction pathways [5]. Apparently, the abnormal activation of EGFR is over-expressed in several cancers: non-small-cell lung, breast, colorectal, bladder, prostate and ovary [6,7]. EGFR was found abundantly expressed in most patients with non-small-cell lung cancer (NSCLC) [8]. Molecular targeted therapies approved for EGFR aberrant activity include the anti-cancer drugs erlotinib (Tarceva; OSI Pharmaceuticals) and gefitinib (Iressa; Astra-Zeneca Pharmaceuticals). Gefitinib and erlotinib are quinazoline-based EGFR inhibitors which reversibly prevent the ATP binding and autophosphorylation activity [9–11]. The clinical development of gefitinib has proved to be more problematic due to toxic side effects but erlotinib achieved breakthrough in trials [12].

* Corresponding address. Life Science Division, Korea Institute of Science and Technology, P.O. Box 131, Cheongryang, Seoul 130-650, South Korea. Tel.: +82 2 958 6850.
E-mail address: amor.san_juan@up.edu.ph

0223-5234/$ - see front matter © 2007 Elsevier Masson SAS. All rights reserved.
doi:10.1016/j.ejmech.2007.06.006

Recently, EGFR inhibitors undergo resistance caused by the tyrosine kinase mutation or autophosphorylation of downstream signaling molecules [13]. The active site mutants of EGFR undergo oncogenic transformation associated with the constitutive autophosphorylation of EGFR, Shc phosphorylation and STAT pathway activation [14]. There are three kinase domain mutations that are associated with drug resistance: an exon 19 point mutation (D761Y), an exon 20 point mutation (T790 M), and an exon 20 insertion (D770_N771insNPG) [15]. Few EGFR mutations were found to have clinical response to gefitinib and erlotinib which selectively activates the antiapoptotic pathways [14,15]. However, the transformation by an exon 20 insertion was reported to be resistant to these inhibitors [16]. Hence, the treatment of cancers harboring EGFR exon 20 insertions may require the development of alternative strategies for kinase inhibition. The fact that 80–90% of lung cancer patients do not have mutant EGFR encourages an additional research to develop novel small molecule inhibitors that can effectively inhibit both wild type and mutant EGFR proteins.

The design of new EGFR agents has been an intense research in developing anti-cancer treatments. Quinazolines are the most developed class of drugs that inhibit intracellularly the EGFR kinase. Drug candidates of this class such as iressa, tarceva and EKB-569 have already reached various stages of clinical trials. Although several applications of 3D-QSAR for the design of EGFR inhibitors were published [17–21], the uniqueness of the current work is the inclusion of new dataset that is not available in previous studies. The ligand design towards the uncontrolled cell division of autophosphorylation could be an alternate route for EGFR inhibition. In this case, targeting the autophosphorylation activity of EGFR is expected to develop potent and selective inhibitors for EGFR.

The present study constructs 3D-QSAR models for the EGFR inhibitors involved in cellular autophosphorylation. Ligand-based 3D-QSAR studies are quite useful to rationalize the structural requirements for enzyme inhibitory activity. The 3D-QSAR techniques of comparative molecular field analysis (CoMFA) [22] and comparative molecular similarity indices analysis (CoMSIA) [23] were employed to examine the structure–activity relationship of the EGFR inhibitors. The generated models provide structural insights for the steric, electrostatic, hydrophobic and hydrogen-bond interactions that influence the inhibition of EGFR autophosphorylation. Additionally, the key structural elements pivotal to design new EGFR agents are also proposed.

2. Methodology

2.1. Dataset

Data from four separately published articles using identical binding assay protocols [24–27] were combined for the CoMFA and CoMSIA studies. Briefly, the autophosphorylation activity assays were performed on the cytoplasmic domain of EGFR. A total of 122 compounds were pooled, yielding a dataset that spanned >4 logarithmic (log) units in terms of pIC_{50} and contained four distinct core structures (Table 1). The experimental biological activities of the dataset compounds are evenly distributed: 66 weakly active compounds ($pIC_{50} < 6.0$), 42 moderately active compounds ($6.0 < pIC_{50} < 7.0$), and 14 highly active compounds ($pIC_{50} > 7.0$). Manual selection of training (95 compounds) and test (26 compounds) sets was based on structural diversity and wide range of activity. A careful selection of the test set compounds represents a range of biological activity similar to the training set. The inhibition constant (IC_{50}) values, i.e., the concentration (nM) of inhibitor that produces 50% inhibition to EGFR autophosphorylation were converted into pIC_{50} ($-\log IC_{50}$) and subsequently used as dependent variables in the QSAR analyses.

Molecular modeling calculations were performed by SYBYL v7.2 [28] on a Silicon Graphics Octane (R1200) workstation with an IRIX 6.5 operating system. The chemical structures of the compounds were built by employing the Sketch option in SYBYL. Initially, the structures were minimized using the Tripos force field [29] and Gasteiger–Huckel charges [30]. Subsequently, the conjugate gradient algorithm was implemented for the optimization of structures. The algorithm removes the instability constraints by successive iteration steps and finally terminates when the convergence reached 0.005 kcal mol^{-1}.

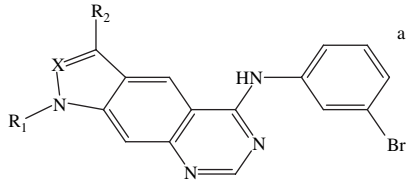
2.2. Alignment of compounds

In the present study, two conformer-based alignment methods were employed. First, the systematic search conformer-based alignment (SSCBA) aligns the compounds based on minimum energy conformation. Second, the docked conformer-based alignment (DCBA) employed molecular docking to align the compounds at the best binding mode of the receptor. The SSCBA process initially searched the global minimum energy conformation of the most active compound **19**. During the systematic conformational search, the torsional angles about pre-selected rotatable bonds were varied in a systematic manner and examined to detect and reject unacceptable conformations. After the best conformation of **19** was obtained, all compounds were superimposed using the reference atoms (Fig. 1) to attain maximum overlap of molecular features contributing to activity. Conversely, in DCBA the compounds were aligned by docking to position the side chains and pharmacophore atoms into the active site of the protein. During the docking process, the compounds find their most probable binding position to the receptor. The compound with the best binding mode identified was used to superimpose the rest of the compounds.

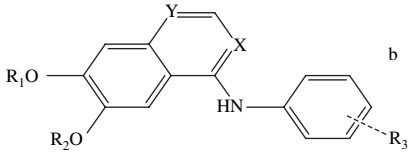
2.3. Setup for CoMFA and CoMSIA

Comparative molecular field analysis (CoMFA) models of steric and electrostatic fields were based on both Lennard–Jones and Coulomb potentials, respectively [22]. The CoMFA was performed to evaluate the steric and electrostatic properties associated with the activity of compounds. Initially,

Table 1
Dataset structure and experimental activity synthesized by Palmer et al. [24], Wissner et al. [25,26], and Smith et al. [27]



Compound	R_1	X	R_2	pIC_{50}
a				
1[e]	H	N	H	7.70
2	Me	N	H	7.89
3[e]	$CH_2CH(OH)CH_2OH$	N	H	6.50
4	$(CH_2)_2NMe_2$	N	H	6.58
5	$(CH_2)_2N(CH_2CH_2)_2O$	N	H	6.72
6	$(CH_2)COOH$	N	H	5.92
7[e]	H	C	H	7.66
8	Me	C	H	7.66
9	$CH_2CH(OH)CH_2OH$	C	H	7.31
10[e]	$(CH_2)_2NMe_2$	C	H	6.84
11	$(CH_2)_3NMe_2$	C	H	6.64
12[e]	$(CH_2)_2N(CH_2CH_2)_2O$	C	H	7.80
13	$(CH_2)_3N(CH_2CH_2)_2O$	C	H	7.11
14[e]	$(CH_2)COOH$	C	H	5.75
15[e]	H	C	$CH_2N(CH_2CH_2OH)_2$	6.87
16	H	C	CH_2NMe_2	8.01
17	H	C	$CH_2N(CH_2CH_2)_2O$	8.02
18	H	C	$CH_2N(Me)(CH_2)_2NMe_2$	7.09
19	H	C	$CH_2N(Me)CH_2COOMe$	8.02
20	H	C	$CH_2N(Me)CH_2COOH$	6.23



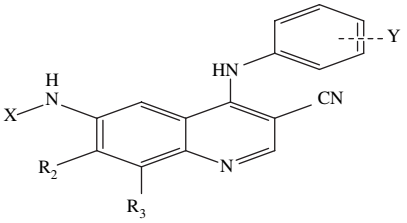
Compound	X	Y	R_1	R_2	R_3	pIC_{50}
b						
21	C–CN	N	C_2H_5	C_2H_5	3-Br	6.38
22	C–CN	N	C_2H_5	C_2H_5	3-Cl, 4-F	6.01
23	C–CN	N	CH_3	CH_3	3-Br	6.72
24	C–CN	N	CH_3	CH_3	3-Cl	6.74
25	C–CN	N	CH_3	CH_3	$3\text{-}CF_3$	5.84
26	C–CN	N	CH_3	CH_3	$3,4\text{-di-}OCH_3$	4.30
27[e]	C–CN	N	CH_3	CH_3	3-F	6.13
28	C–CN	N	CH_3	CH_3	3-CN	4.85
29	C–CN	N	CH_3	CH_3	4-F	4.32
30[e]	C–CN	N	CH_3	CH_3	$3\text{-}COCH_3$	5.13
31[e]	C–CN	N	CH_3	CH_3	$3\text{-}NH_2$	6.09
32	C–CN	N	CH_3	CH_3	$3\text{-}N(CH_3)_2$	6.07
33	C–CN	N	CH_3	CH_3	$3\text{-}NO_2$	6.06
34	C–CN	N	CH_3	CH_3	3-Cl, 4-F	6.27
35	C–CN	N	CH_3	CH_3	4-Cl, 2-F	4.86
36	C–CN	N	CH_3	CH_3	3-OH	5.20
37[e]	C–CN	N	CH_3	CH_3	$4\text{-}CH_3$	5.56
38	C–CN	N	CH_3	CH_3	$3\text{-}CONH_2$	5.11
39	C–CN	N	CH_3	CH_3	$3\text{-Br, }4\text{-}CH_3$	5.75
40	C–CN	N	CH_3	CH_3	3-Cl, 4-OH	5.04
41	C–CN	N	CH_3	CH_3	3,5-di-Cl, 4-OH	5.50
42[e]	C–CN	N	CH_3	CH_3	5-Cl, 2-OH	4.77
43	C–CN	N	CH_3	CH_3	$3\text{-}SCH_3$	6.12
44	C–CN	N	CH_3	CH_3	$3\text{-Cl, }4\text{-}CH_3$	5.83
45	C–CN	N	CH_3	CH_3	$4\text{-}CH(CH_3)_2$	6.64
46	C–CN	N	CH_3	CH_3	$2\text{-}CH(CH_3)_2$	6.44
47[e]	C–CN	N	CH_3	CH_3	$3\text{-}CH(CH_3)_2$	4.21

(*continued on next page*)

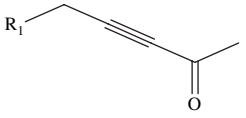
Table 1 (*continued*)

Compound	X	Y	R_1	R_2	R_3	pIC_{50}
48	C–CN	N	CH_3	CH_3	4-Br	5.18
49	C–CN	N	CH_3	CH_3	2-Br	5.38
50[e]	C–CN	N	CH_3	CH_3	3-CF_3, 4-F	6.80
51	C–CN	N	CH_3	CH_3	2-CH_3, 3-Br	6.08
52	C–CN	N	CH_3	CH_3	3-CH_3, 4-Br	6.21
53	C–CN	N	CH_3	CH_3	3-N_3	5.66
54	C–CN	N	CH_3	CH_3	3-$NHCOCH_3$	4.16
55	C–CN	N	C_2H_5	CH_3	3-Br	4.67
56	C–CN	N	CH_3	C_2H_5	3-Br	4.48
57	C–CHO	N	CH_3	CH_3	3-Br	5.47
58	C–CO_2H	N	CH_3	CH_3	3-Br	4.47
59	C–$CONH_2$	N	CH_3	CH_3	3-Br	4.82
60	C–CN	N	H	H	3-Br	6.46
61	C–CN	N	CH_3OCH_2	CH_3OCH_2	3-Br	5.17
62	C–CN	N	$CH_3O(CH_2)_2$	$CH_3O(CH_2)_2$	3-Br	5.96
63	C–CN	N	CH_2		3-Br	5.17
64	C–CN	N	$(CH_2)_3$		3-Br	5.29
65	N	C–CN	CH_3	CH_3	3-Br	4.15
66	N	C–CN	CH_3	CH_3	3-Cl	4.05
67	N	N	CH_3	CH_3	3-Br	7.15

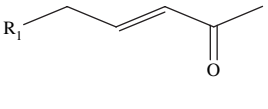
c



X =

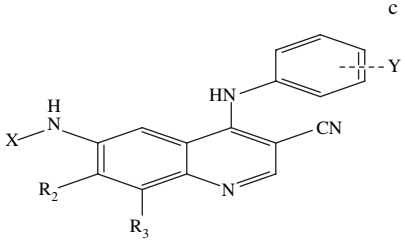


I

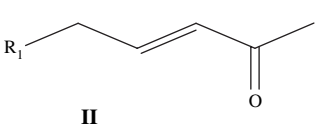


II

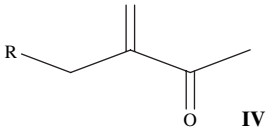
68	II	3-Cl, 4-F	$N(CH_3)_2$	OC_2H_5	H	7.10
69[e]	I	3-Br	H	H	H	6.03
70	I	3-Br	OCH_3	H	H	6.02
71	I	3-Br	$O(CH_2)_2OCH_3$	H	H	6.03
72[e]	I	3-Br	OCH_2OCH_3	H	H	6.48
73	I	3-Br	$N(CH_2CH_2)_2S$	H	H	6.26
74	I	3-Br	$N(CH_2)_3CHCH_2OCH_3$	H	H	6.82
75[e]	I	3-Br	$N(CH_2CH_2CH_3)_2$	H	H	6.89
76[e]	I	3-Br	$N[CH(CH_2)_2]_2$	H	H	6.60
77	I	3-Br	H	OCH_3	H	6.04
78[e]	I	3-Br	OCH_3	OCH_3	H	6.10
79	I	3-Br	$N(CH_3)_2$	OCH_3	H	7.05
80	II	3-Br	$N(CH_3)_2$	H	H	6.51
81	II	3-Br	$N(CH_3)_2$	OCH_3	H	6.10
82	II	3-Br	$N(CH_2CH_2)_2O$	OCH_3	H	5.42
83	II	3-Br	$N(CH_3)_2$	OC_2H_5	H	5.80
84	II	3-Br	$N(CH_2CH_2)_2O$	OC_2H_5	H	4.99



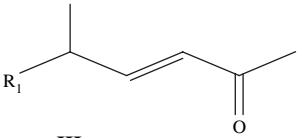
X =



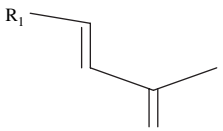
II



IV



III



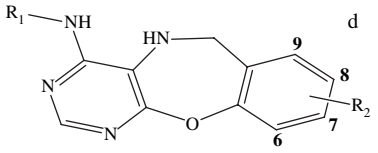
V

c

85	II	3-Br	$N(CH_3)_2$	H	OCH_3	5.54
86	II	3-Br	$N(CH_2CH_2)_2O$	H	OCH_3	5.27
87[e]	II	3-Cl, 4-F	$N(CH_3)_2$	H	H	6.19
88	II	3-Cl, 4-F	$N(CH_2CH_2)_2O$	H	H	5.64
89	II	3-Cl, 4-F	$N(CH_3)_2$	OCH_3	H	5.63

Table 1 (*continued*)

Compound	X	Y	R_1	R_2	R_3	pIC_{50}
90	II	3-Cl, 4-F	$N(C_2H_5)_2$	OCH_3	H	5.72
91	II	3-Cl, 4-F	$N(CH_2CH_2)_2O$	OCH_3	H	4.87
92[e]	II	3-Cl, 4-F	$N(CH_3)_2$	OCF_3	H	5.58
93	III	3-Cl, 4-F	$N(CH_3)_2$	OCH_3	H	5.28
94	III	3-Cl, 4-F	$N(C_2H_5)_2$	OCH_3	H	5.32
95	III	3-Cl, 4-F	$N(CH_2CH_2)_2O$	OCH_3	H	4.98
96	V	3-Br	H	H	H	6.23
97	V	3-Br	H	OCH_3	H	6.09
98	V	3-Br	CH_3	H	H	5.79
99	V	3-Cl, 4-F	CH_3	OC_2H_5	H	5.12
100[e]	IV	3-Cl, 4-F	$N(CH_2CH_2)_2O$	H	H	5.57
101	IV	3-Cl, 4-F	$N(CH_3)_2$	OC_2H_5	H	6.25



Compound	R_1	R_2	pIC_{50}
[d]			
102	CH_2Ph	8-Br	4.52
103	CH_2Ph	8-OCH_3	5.07
104[e]	*m*-Br Ph	8-Br	5.14
105	*m*-Br Ph	8-F	5.16
106	*m*-Br Ph	7-OCH_3	5.72
107[e]	*m*-Br Ph	8-OCH_3	6.00
108	*m*-Cl, *p*-F Ph	8-Br	5.08
109	*m*-Cl, *p*-F Ph	8-Cl	5.37
110	*m*-Cl, *p*-F Ph	8-F	5.14
111	*m*-Cl, *p*-F Ph	7-OCH_3	5.05
112	*m*-Cl, *p*-F Ph	8-OCH_3	5.62
113[e]	*m*-C_6H_5 Ph	8-F	5.00
114	*m*-C_6H_5 Ph	8-OCH_3	5.01
115	*m*-Cl, *p*-F Ph	8-NH_2	5.04
116	*m*-Br Ph	8-$NHSO_2CH_3$	5.15
117	*m*-Br Ph	8-NHCOCH=CH_2	5.27
118	*m*-Cl, *p*-F Ph	8-NHCOCH=CH_2	5.05
119[e]	*m*-Br Ph	H	5.09
120	*m*-Cl, *p*-F Ph	H	5.57
121	*m*-Br Ph	7,8-$(OCH_3)_2$	6.33
122	*m*-Cl, *p*-F Ph	7,8-(OCH3)2	6.16

[a] Quinazoline-based EGFR inhibitors active against autophosphorylation activity synthesized in 1997.
[b] Quinolinecarbonitrile-based EGFR inhibitors active against autophosphorylation activity synthesized in 2000.
[c] Quinolinecarbonitrile-based EGFR inhibitors active against autophosphorylation activity synthesized in 2003.
[d] Tricyclicazepine-based EGFR inhibitors active against autophosphorylation activity synthesized in 2006.
[e] Test set compound.

a three-dimensional cubic lattice with 2 Å grid spacing was generated automatically around the molecules to ensure that the grid extended the molecular dimensions by 4 Å in all directions. Next, the steric and electrostatic field energies were calculated using an sp^3 carbon probe with van der Waals radius of 1.52 Å and a +1 charge. Energies were truncated to ± 30 kcal mol^{-1} and the electrostatic contribution was ignored at lattice points with maximum steric interactions. Finally, the CoMFA fields automatically generated were block scaled by the CoMFA standard method in SYBYL.

Comparative molecular similarity index analysis (CoMSIA) was employed to determine the hydrophobic and hydrogen-bond properties related with the activity of compounds [23]. CoMSIA descriptors were derived by the same lattice box used for the CoMFA calculations. All five CoMSIA similarity index fields (steric, electrostatic, hydrophobic, hydrogen-bond donor, and hydrogen-bond acceptor) were evaluated using the sp^3 carbon probe. The CoMSIA models from hydrophobic and hydrogen bonds were calculated between the grid point and each atom of the molecule by a Gaussian function [23]. The implementation of a Gaussian distance function generates smoother sampling of the descriptor fields around the molecules. The attenuation factor's default value of 0.30 was used, which is the standard distance dependence

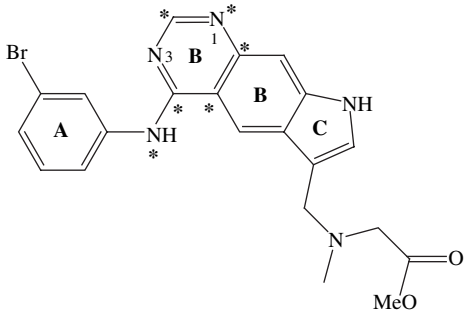


Fig. 1. The six reference atoms used for alignment with template compound **19** are starred.

of molecular similarity. The effect of using the standard attenuation factor is shown in contour maps with prominent molecular features. Conversely, a higher value of attenuation factor yields fewer molecular features.

2.4. Partial least square (PLS) analysis

PLS method was employed to linearly correlate the 3D-QSAR fields with the inhibitory activity values. Correlation was performed by using the fields as independent variables and the pIC_{50} values as dependent variables. The result of the linear correlation corresponds to a regression equation with thousands of coefficients. To select and verify the best model, leave-one-out (LOO) [31] cross-validation was employed. In this technique, compounds are excluded from the dataset, and the activity of each removed compound is predicted by a new model derived from the remaining compounds in the set. The optimum number of components (N) which corresponds to the highest cross-validated q^2 and the lowest standard error of prediction (SEP) was evaluated. To shorten the time of computational analysis and reduce the background noise, the minimum-sigma (column filtering) was set to 2 kcal mol^{-1} for CoMFA and 1.0 kcal mol^{-1} for CoMSIA. Further, region-focusing technique available in advanced CoMFA module in SYBYL with grid size value of 1 Å was implemented to refine the model. The technique suppresses the noise by increasing the weights for those lattice points which were most pertinent to the model [32].

2.5. Model validations

The standard method of evaluating the quality of a QSAR model is to assess the three standard squared correlation coefficients: r^2, q^2, and r^2_{pred}. Generally, a robust statistical model obtained from CoMFA and CoMSIA should have high values (>0.50) of r^2, q^2, and r^2_{pred}. Measurement by cross-validated q^2 determines the internal predictive ability of the model while the conventional correlation coefficient r^2 evaluates the internal consistency of the model. Predictive r^2 (r^2_{pred}) validates the model by evaluating the correlation between actual and predicted inhibition. The predictive r^2 was calculated overall of the test set while the r^2 and q^2 were calculated based on the overall compounds in the training set. The biological activities of the test set were predicted using the model derived from the training set. After the calculated activities were obtained, the overall predictive performance of the model was expressed by the predictive r^2 value [33].

Molecular docking is another way to validate a model obtained from CoMFA or CoMSIA. Docking was reported to be useful in validating the 3D-QSAR results [34–36]. The validation was performed by consistency check between the contour maps and the binding site. In this study, FlexX interfaced with the SYBYL was employed in the docking analysis [37]. FlexX is a fast and automated program that considers ligand conformational flexibility by an incremental fragment placing technique [37,38]. Initially, the ligand and protein coordinates were prepared. The charges of the ligands were removed and replaced by formal charges. The ligand and water molecules were removed from the 1XKK PDB code [39] protein structure and subsequently added with hydrogen atoms and the Kollman-all atom charges. Hydrogen atoms and the Kollman-all atom charges were added to the whole protein. Next, the active site for docking was selected for residues located within a distance of 6.5 Å from the co-crystallized ligand. Charged residues located at the binding site were protonated and ionized at the physiological pH of 7.0. Redocking the bound crystal structure of the ligand to its active site validated the reliability of FlexX. The low root-mean-square-deviation of 0.03 Å for the inhibitor–protein complex lapatinib–1XKK indicates reproducibility of the docked complex and hence the reliability of FlexX.

3. Results and discussion

3.1. Molecular alignment from two conformers

The best method for the alignment of compounds was judged based on PLS statistical analysis. The first conformer SSCBA yields a higher cross-validated q^2 as compared with the second conformer DCBA (Table 2). After applying the region-focusing and removal of outliers, the q^2 slightly increased from 0.47 to 0.50 for SSCBA and from 0.36 to 0.37 for DCBA. The analysis of the outliers (**20** and **54**) removed showed some unique features that are not present in the training set. Further, these two compounds gave >1 log unit of residuals, which is the difference between the corresponding values of the experimental and predicted binding affinities. The results showed significantly better predictivity for the systematic conformational search alignment rather than the docking-based alignment. The superposition of compounds by pairwise fitting of common structural elements to the reference compound yields reduced noise in calculation. The findings seem bizarre that the docked geometries result in inferior models compared with those models based on simplistic molecular superposition. This can be explained by the fact that docking a different ligand to the receptor may often lead to unreliable results due to inability of the current docking programs to model protein flexibility. Although a suitable alignment should be based on the proper orientation of the ligands towards their experimental position in a binding pocket,

Table 2
CoMFA and CoMFA region-focusing for training set

Parameters	Conformer: SSCBA		Conformer: DCBA	
	CoMFA	CoMFA (RF[k])	CoMFA	CoMFA (RF[k])
q^{2}[a]	0.47	0.50	0.36	0.37
r^{2}[b]	0.72	0.74	0.64	0.64
SEP[c]	0.67	0.65	0.73	0.73
SEE[d]	0.48	0.47	0.54	0.55
F[e]	46.04	48.83	40.72	31.04
Component[f]	5	5	4	5
Outliers[g]		20, 54		20, 54
Fractions[h]				
Steric	0.48	0.53	0.55	0.49
Electrostatic	0.52	0.47	0.45	0.51
r^{2}_{bs}[i]	0.82	0.80	0.73	0.76
r^{2}_{pred}[j]	0.49	0.56	0.43	0.52

[a] Cross-validated correlation coefficient.
[b] Non-cross-validated correlation coefficient.
[c] Standard error of prediction.
[d] Standard error of estimate.
[e] F-test value.
[f] Optimum number of components.
[g] Outliers removed.
[h] Relative field contributions.
[i] Bootstrap in 100 runs.
[j] Predictive r^2.
[k] Region focus with 1 Å grid.

the results in this study indicate that the overall orientation of the structurally aligned molecules primarily affects the predictability of the models [40]. As shown in Fig. 2a and b, the docking conformer-based alignment appeared to have maximum overlap of molecular superposition. Conversely, the systematic search conformer-based alignment appeared to have varied overlap at C-ring position (refer to Fig. 1). The results of the varying molecular superposition at C-ring in SSCBA gave better PLS statistics as compared with DCBA. This can be inferred by the diversity of scaffolds that is reflected by the alignment which yields better predictive ability of the models. Based on the above observations, the SSCBA alignment method was chosen for further analysis.

3.2. CoMFA and CoMSIA

The best statistical model for CoMFA shows the highest cross-validated q^2 value of 0.50 with five components, the conventional r^2 value of 0.74 and predictive r^2 value of 0.56. Similarly, a better statistical quality of the best CoMSIA was obtained from SSCBA (Table 3). Several statistical parameters such as r^2, q^2, r^2_{pred}, N and F were used to evaluate the robustness of a QSAR model. Firstly, consideration was based on the high q^2 value at optimum number of components (N). The CoMSIA models from the two conformer-based alignments were selected preliminarily. For the SSCBA conformer, the best models include the SEDA: steric, electrostatic, donor and acceptor and SDA: steric, donor and acceptor, while for the DCBA conformer include the SEDA and SEHD: steric, electrostatic, hydrophobic and donor. Subsequently, consideration of the best models is based on high r^2_{pred}. Tropsha [41] has emphasized that high q^2 value (>0.50) is necessary but not a sufficient condition for a predictive QSAR model. Once models with low q^2, low r^2 and high N values were ruled out, the best models should be chosen based on high r^2_{pred} [41]. Only the SEDA and SDA models from the SSCBA conformer have performed well ($r^2_{pred} > 0.50$) in the prediction for the activity of test compounds. Finally, the consideration is based on a lower value of standard error of estimate (SEE) which measures the accuracy of the predicted model. In this case, the SEDA model yielded the lowest SEE value of 0.51. Overall, based on q^2, r^2, r^2_{pred}, N, SEE as well as the highest value of Fisher F-test, the best models generated from SSCBA include the SE fields from CoMFA as well as the SDA fields from CoMSIA. Thus the models that were included in the graphical analysis were the steric, electrostatic, hydrogen-bond donor and acceptor fields. These graphical contour maps could be

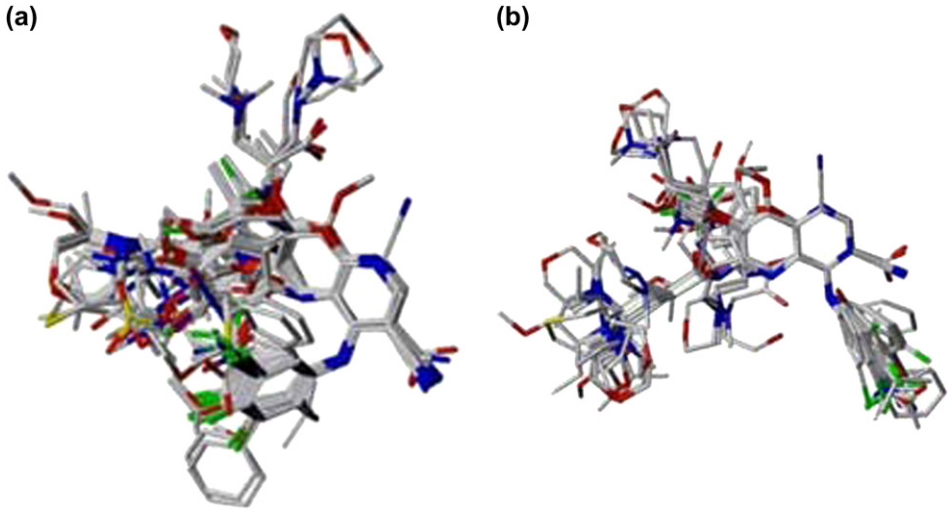


Fig. 2. Comparison of conformational alignments. (a) Systematic search conformer-based alignment (SSCBA). (b) Docked conformer-based alignment (DCBA).

Table 3
Best CoMSIA models from different conformers

Parameters	Conformer: SSCBA		Conformer: DCBA	
	SEDA[j]	SDA[k]	SEDA[j]	SEHD[l]
q^2[a]	0.42	0.48	0.46	0.44
r^2[b]	0.69	0.62	0.68	0.59
SEP[c]	0.70	0.65	0.67	0.68
SEE[d]	0.51	0.56	0.52	0.58
F[e]	49.17	72.70	63.50	65.29
Component[f]	4	2	3	2
Fractions[g]				
Steric	0.14	0.30	0.15	0.15
Electrostatic	0.38		0.30	0.29
Hydrophobic				0.29
Donor	0.25	0.30	0.30	0.28
Acceptor	0.24	0.40	0.24	
r^2_{bs}[h]	0.76	0.79	0.72	0.63
R^2_{pred}[i]	0.54	0.51	0.45	0.45

[a] Cross-validated correlation coefficient.
[b] Non-cross-validated correlation coefficient.
[c] Standard error of prediction.
[d] Standard error of estimate.
[e] F-test value.
[f] Optimum number of components.
[g] Relative field contributions.
[h] Bootstrap in 100 runs.
[i] Predictive r^2.
[j] Steric, electrostatic, donor, acceptor.
[k] Steric, donor, acceptor.
[l] Steric, electrostatic, hydrophobic, donor.

selected and used as the final screening tool for lead optimization in designing a drug for EGFR autophosphorylation.

3.3. Graphical interpretation of contour maps

Fig. 3a shows the steric contour map for CoMFA. Green polyhedra represent a steric group that confers an increased activity while yellow polyhedra represent a bulky group that results in a decreased activity. The steric fields indicate that bulky substituents on the C-ring down to trimethylamine (TMA) side chain will improve the activity. The interpretation matched with the structural features and biochemical activities of compounds **1**–**20**. In general, the substitutions at R_1 position (**1**–**14**) of C-ring yielded lower activities as compared with R_2 position (**15**–**20**). Further, the bulky group at R_2 position of C-ring showed higher activity. In fact, the attachment of TMA in **16** yielded a remarkable 183.5-fold increase in potency as compared with **14**. Conversely, a large yellow block found at the methyl formate indicated unfavorable steric group substitution. To illustrate this, the presence of a carboxylic group in **20** showed 60.8-fold decreased in activity, in contrast with **19** that showed favorable activity.

Fig. 3b depicts the electrostatic contour map for CoMFA. A red contour indicates that an electronegative group will favor the activity while a blue contour will reduce the activity. The red blocks encompassing the B-ring and methyl formate indicated favored activity by the presence of electron rich functionalities. The favorable electronegative group at N_3 position of the B-ring is shown by the high activities of carbonitrile analogues (**21**–**25**, **27**, **31**–**34**, **45**, **46**). As a result, when this position is substituted by aldehyde (**57**), carboxylic acid (**58**) or amide (**59**) groups it leads to significant loss in potency by 18-fold, 180-fold and 80-fold, respectively. Furthermore, the relocation of a nitrile group from N_3 to N_1 position of the B-ring (**65** and **66**) showed decrease activities by 373-fold and 490-fold, respectively. In fact, the unsubstituted nitrogen atoms of quinazoline in **67** showed highly favored activity. Hence, the two nitrogen atoms at the B-ring play an important role in binding affinity to the receptor. Conversely, blue contours were found farther away from the A-, B- and C-ring where low electron density is expected to increase the activities. Particularly the tricyclic ring, i.e., pyrrolo-quinazoline is surrounded by a big blue block that indicates an electropositive group enhances the activity. For example, the inhibitory activities of compounds **2**–**5**, **11**, and **13** are arranged as **2** < **13** < **5** < **11** < **4** < **3**, in the order of increasing activity.

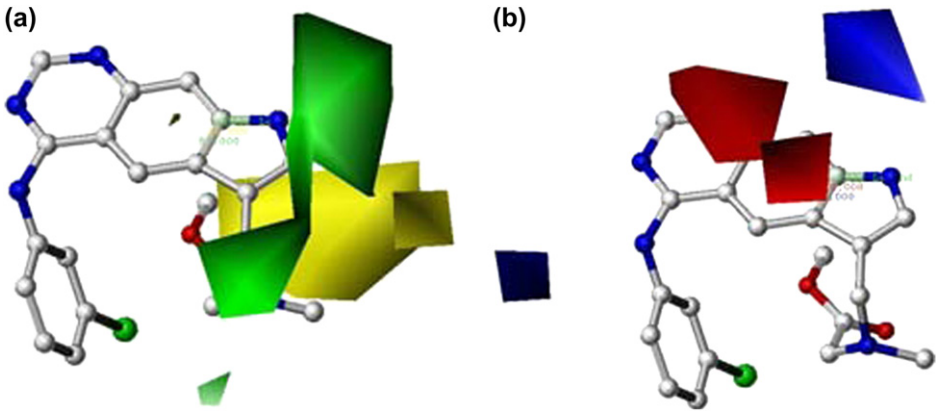


Fig. 3. CoMFA (a, steric; b, electrostatic) region-focused contour maps obtained from SSCBA. Green (yellow) regions indicate where bulky groups increase (decrease) the EGFR autophosphorylation activity. Red (blue) regions indicate where electronegative groups increase (decrease) the EGFR autophosphorylation activity. The most potent compound **19** is shown in ball and stick representation. (For interpretation of the references to color in this figure legend and text citation, the reader is referred to the web version of this article.)

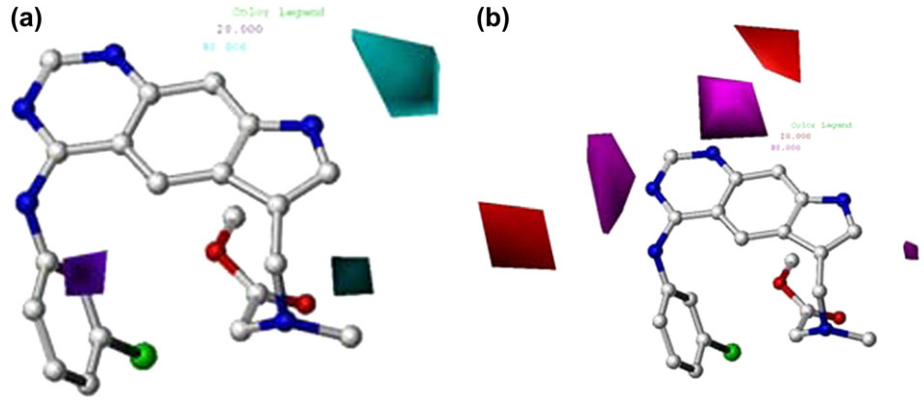


Fig. 4. CoMSIA contour map obtained from SSCBA method (a, hydrogen-bond donor; b, hydrogen-bond acceptor) around the potent compound **19** (shown in ball and stick representation). Cyan (purple) regions indicate where hydrogen-bond donors increase (decrease) the EGFR autophosphorylation activity. Magenta (red) regions indicate where hydrogen-bond acceptors increase (decrease) the EGFR autophosphorylation activity. (For interpretation of the references to color in this figure legend and text citation, the reader is referred to the web version of this article.)

The hydrogen-bond contours for CoMSIA provided an additional structural insight for the probable binding site of the ligand–receptor complex. Cyan represents the ligand that favors proton donor, which is associated with the proton acceptor at the receptor site. In Fig. 4a, a cyan contour near the C_3 position of the A-ring shows that hydrogen-bond donor favors activity. The strong base analogue **16**, the weak base **17**, and the neutral ester **19** all showed autophosphorylation activity superior to that of the parent compound **7**, with IC_{50}'s below 10 nM. Further, the substitution of TMA in **16** yielded a two-fold increase in potency, but after carboxylic acid was attached to TMA in **20** the potency reduced to 60-fold. In hydrogen-bond acceptor contour map (Fig. 4b), there is a magenta block close to the N_1 position of B-ring, and a red block just a little farther away. This means that the N_1 substitution should have electron rich atoms to be closely connected into an electron-donating atoms to improve the activity. This can be explained by a reduced activity when the nitrogen at N_1 position of **67** was replaced by carbon linked to nitrile (**65** and **66**), which yield the decrease in potency by 1012-fold and 1261-fold, respectively. The results indicate that the inactivity of carbonitrile analogue favors the direct interaction of N_1 in quinoline with the anionic binding site of the receptor.

3.4. Model validations

Model validation is the most important step in model building. This study validates the model based on quantitative (predictive r^2) and qualitative (molecular docking) evaluations. Twenty-seven compounds selected in the test set were used to verify the validity of the CoMFA and CoMSIA models. All compounds in the test datasets were predicted well, with almost all residuals <1 log unit (Table 4). Additionally, the good correlation between the observed and predicted values for the training set is evident from the tabulated results (Table S1, see Supplementary data) and corresponding model plots with $r^2 = 0.74$ for CoMFA (Fig. 5a) and $r^2 = 0.61$ for CoMSIA (Fig. 5b). Further, the relatively good predictive r^2 values of 0.56 (CoMFA) and 0.51 (best CoMSIA) suggest that the models are predictive. As shown by a good predictive ability of the models, an additional validation by docking was carried out. Molecular docking further verified the consistency of 3D-QSAR contour maps in the active site of EGFR. Two hydrogen bonds were formed between the most active compound **19** with the Lys745 and Met793 residues in the active site (Fig. 6a). Generally, the projection of the 3D-QSAR contour

Table 4
Residuals of test set prediction for CoMFA and CoMSIA models

Compound	pIC_{50}[a]	Residuals	
		CoMFA	CoMSIA
1	7.70	0.34	0.09
3	6.50	−0.34	−0.66
7	7.66	0.32	0.27
10	6.84	0.26	0.25
12	7.80	1.17	1.03
14	5.75	−0.57	−0.81
15	6.87	−0.86	−1.15
27	6.13	0.61	0.60
30	5.13	−0.59	−0.46
31	6.09	0.45	0.80
37	5.56	0.11	0.03
42	4.77	−0.76	−0.63
47	4.21	−1.72	−1.43
50	6.80	1.21	1.24
69	6.03	−0.41	0.09
72	6.48	0.25	0.55
75	6.89	0.22	0.46
76	6.60	0.05	0.17
78	6.10	−0.37	0.29
87	6.19	0.15	0.26
92	5.58	0.07	−0.26
100	5.57	−0.57	−0.83
104	5.14	−0.24	−0.44
107	6.00	0.42	0.23
113	5.00	−0.01	−0.45
119	5.09	−0.26	−0.44
Variance[b]		0.38	0.43

[a] Experimental pIC_{50} for EGFR autphosphorylation activity.
[b] Population variance: RSS/n for n compounds.

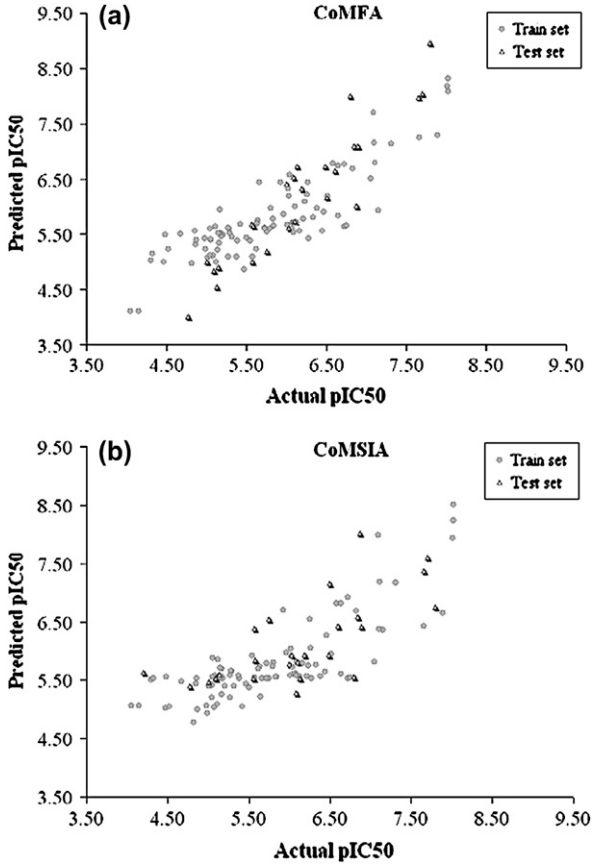


Fig. 5. (a) The correlation coefficient ($r^2 = 0.74$) between experimental pIC_{50} and CoMFA predicted pIC_{50} for training and test sets. (b) The correlation coefficient ($r^2 = 0.61$) between experimental pIC_{50} and CoMSIA predicted pIC_{50} for training and test sets.

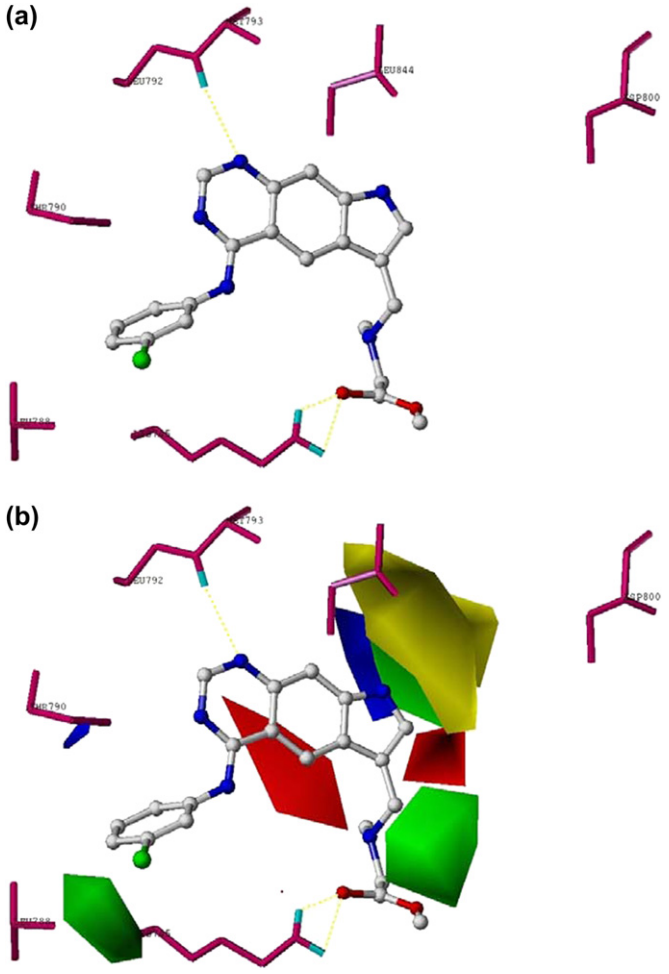


Fig. 6. (a) Ligand–protein contact of compound **19** bound to EGFR (PDB ID: 1XKK). Two hydrogen bonds (shown in yellow dotted lines) were mediated in Met793 and Lys745 residues in the active site. (b) CoMFA electrostatic contour (focused tripos standard field) superimposed on the receptor active site (compound **19** is shown in the background). (For interpretation of the references to color in this figure legend and text citation, the reader is referred to the web version of this article.)

maps into the active site showed complementarity that indicates a unified pharmacophore model. Fig. 6b shows a red contour near the quinazoline moiety of **19** that is found projected along the hydrophilic pocket Met793. This is consistent with the formation of a hydrogen bond between the electronegative N_1 atom with the hydrogen side chain of Met793. Further superposition of CoMSIA hydrogen-bond acceptor at the binding site water accessible surface was found complementary (Fig. S1a and b). The magenta blocks encompassing the B-ring of **19** represent regions where electron-donating residues were found in this position of the receptor. This was satisfied by the presence of charged residues Met793 and Thr790. In fact, the methyl hydrogen of residue Thr790 is in close contact with N_3 of quinoline. Moreover, the residue Met793 formed a hydrogen bond with N_1 of B-ring. Overall, the lipophilic surface of **19** was found in contact with hydrophilic part of the receptor.

3.5. Associated structural modeling study

Pednekar et al. [42] published a 3D-QSAR model for 34 of the Wissner et al. [26] quinolinecarbonitrile analogues which correspond to compounds **68–101** (Table 1) in the present study. The combined methodologies of CoMFA and genetic function approximation (GFA) were employed in the generation of their 3D-QSAR models using thermodynamic, electronic and spatial descriptors. Their best model had $q^2 = 0.46$, $r^2 = 0.99$, $F = 690.82$, and $r^2_{pred} = 0.52$. Note that their best model is for fewer ligands and is rather less reliable than the present study according to the statistical measures q^2, r^2 and F. Additionally their r^2_{pred} is relatively good for their small, six-compound test set. A comparison of their results with this study in terms of the r^2_{pred} values indicates validity for 27 compounds in the test set. According to their work the steric region is favored at the quinoline moiety while this study showed favorable steric substitution at the pyrrole moiety directly attached to the quinoline. There is a consensus between the present study and Pednekar's report that an electrostatic group is favored at the quinoline region. The applied GFA method from their study showed complementary analysis with their generated contour maps. Similarly, the docking method in this study showed complementary analysis of the

contour maps to the EGFR active site. Overall, the combination of these two works indicates the relative contributions of steric and electrostatic substitutions to increase the EGFR inhibition.

4. Conclusion

The present investigation designs predictive 3D-QSAR models for autophosphorylation activity of EGFR. The inhibition of autophosphorylation as an alternate route to control the activation of the EGFR will have implications for the development of new EGFR-blocking drugs. Structural replacements by larger substituents to the C-ring down to trimethylamine moiety, electronegative groups around the B-ring, and hydrogen-bond donor to the A-ring of the most potent compound **19** are necessary to increase the EGFR activity against autophosphorylation. The robustness of the constructed 3D-QSAR models was validated by a good predictive r^2 and further consistency between the contour maps and docking to the active site. The structural requirements of the ligand as well as the plausible binding mode of the inhibitor to the EGFR obtained from the present study can be utilized in the design of new and more potent agents for EGFR autophosphorylation inhibition.

Appendix. Supplementary data

Supplementary data associated with this article can be found in the online version, at doi:10.1016/j.ejmech.2007.06.006. Table S1 shows the comparison between the predicted and experimental activities for the training set. Fig. S1 shows the hydrogen-bond donor superimposed to the active site.

References

[1] G. Carpenter, L. King Jr., S. Cohen, Nature 276 (1978) 409–410.
[2] J. Schelessinger, Cell 110 (2002) 669–672.
[3] Y. Yarden, M.X. Sliwkowski, Nat. Rev. Mol. Cell Biol. 2 (2001) 127–137.
[4] Y. Yarden, Eur. J. Cancer 37 (2001) S3–S8.
[5] R.S. Herbst, Int. J. Radiat. Oncol., Biol., Phys. 59 (2004) 21–26.
[6] D.S. Salomon, R. Brandt, F. Ciardiello, N. Normanno, Crit. Rev. Oncol. Hematol. 19 (1995) 183–232.
[7] J. Mendelsohn, Endocr. Relat. Cancer 8 (2001) 3–9.
[8] V. Rusch, J. Baselga, C. Cordon-Cardo, J. Orazem, M. Zaman, S. Hoda, J. McIntosh, J. Kurie, E. Dmitrovsky, Cancer Res. 53 (1993) 2379–2385.
[9] M.H. Cohen, G.A. Williams, R. Sridhara, G. Chen, R. Pazdur, Oncologist 8 (4) (2003) 303–306.
[10] R.S. Herbst, M. Fukuoka, J. Baselga, Nat. Rev. Cancer 4 (2004) 956–965.
[11] A.F. Cohen, E.A. Dubois, H. van Bronswijk, Ned. Tijdschr. Geneeskd. 49 (2006) 2703–2704.
[12] D. Blackhall, M. Ranson, N. Thatcher, Lancet Oncol. 7 (2006) 499–507.
[13] R. Bianco, T. Troiani, G. Tortora, F. Ciardiello, Endocr. Relat. Cancer 12 (2005) S159–S171.
[14] R. Sordella, D.W. Bell, D.A. Haber, J. Settleman, Science 305 (2004) 1163–1167.
[15] G.J. Riely, K.A. Politi, V.A. Miller, W. Pao, Clin. Cancer Res. 12 (2006) 7232–7241.
[16] H. Greulich, T.H. Chen, W. Feng, P.A. Jänne, J.V. Alvarez, M. Zappaterra, S.E. Bulmer, D.A. Frank, W.C. Hahn, W.R. Sellers, M. Meyerson, PloS Med. 2 (2005) 1167–1176.
[17] T. Hou, L. Zhu, L. Chen, X. Xu, J. Chem. Inf. Comput. Sci. 43 (2003) 273–287.
[18] H. Assefa, S. Kamath, J.K. Buolamwini, J. Comput. Aided Mol. Des. 17 (2003) 475–493.
[19] A. Vema, S.K. Panigrahi, G. Rambabu, B. Gopalakrishnan, J.A. Sarma, G.R. Desiraju, Bioorg. Med. Chem. 11 (2003) 4643–4653.
[20] S. Kamath, J.K. Buolamwini, J. Med. Chem. 46 (2003) 4657–4668.
[21] G. Chen, X. Luo, W. Zhu, C. Luo, H. Liu, C.M. Puah, K. Chen, H. Jiang, Bioorg. Med. Chem. 12 (2004) 2409–2417.
[22] R.D. Cramer III, D.E. Patterson, J.D. Bunce, J. Am. Chem. Soc. 110 (1988) 5959–5967.
[23] G. Klebe, U. Abraham, T. Mietzner, J. Med. Chem. 37 (1994) 4130–4146.
[24] B.D. Palmer, S. Trumpp-Kallmeyer, D.W. Fry, J.M. Nelson, H. Showalter, W.A. Denny, J. Med. Chem. 40 (1997) 1519–1529.
[25] A. Wissner, D.M. Berger, D.H. Boschelli, M.B. Floyd Jr., L.M. Greenberger, B.C. Gruber, B.D. Johnson, N. Mamuya, R. Nilakantan, M.F. Reich, R. Shen, H.R. Tsou, E. Upeslacis, Y.F. Wang, B. Wu, F. Ye, N. Zhang, J. Med. Chem. 43 (2000) 3244–3256.
[26] A. Wissner, E. Overbeek, M.F. Reich, M.B. Floyd, B.D. Johnson, N. Mamuya, E.C. Rosfjord, C. Discafani, R. Davis, X. Shi, S.K. Rabindran, B.C. Gruber, F. Ye, W.A. Hallett, R. Nilakantan, R. Shen, Y.F. Wang, L.M. Greenberger, H.R. Tsou, J. Med. Chem. 46 (2003) 49–63.
[27] L. Smith II, E.L. Piatnitski, A.S. Kiselyov, X. Ouyang, X. Chen, S. Burdzovic-Wizemann, Y. Xu, Y. Wang, R.L. Rosler, S.N. Patel, H.H. Chiang, D.L. Milligan, J. Columbus, W.C. Wong, J.F. Doody, Y.R. Hadari, Bioorg. Med. Chem. Lett. 16 (2006) 1643–1646.
[28] SYBYL 7.2, Tripos Inc., 1699 Hanley Road, St. Louis, MO 63144, 2006.
[29] M. Clark, R. Cramer III, N. Van Opdenbosch, J. Comput. Chem. 10 (1989) 982–1012.
[30] A. Streitweiser, Molecular Orbital Theory for Organic Chemists, Wiley, New York, 1961.
[31] L. Stahle, S. Wold, Prog. Med. Chem. 25 (1988) 292–338.
[32] F. Lindgren, P. Geladi, S. Rannar, S. Wold, J. Chemom. 8 (1994) 349–363.
[33] A. Leach, V. Gillet (Eds.), An Introduction to Chemoinformatics, Kluwer Academic Publisher, Dordrecht, United Kingdom, 2003.
[34] W. Sippl, H. Höltje, J. Mol. Struct. (Theochem) 503 (2000) 31–50.
[35] J. Buolamwini, H. Assefa, J. Med. Chem. 45 (2002) 841–852.
[36] G. Desiraju, B. Gopalakrishnan, R. Jetti, A. Nagaraju, D. Raveendra, J. Sarma, M. Sobhia, R. Thilagavathi, J. Med. Chem. 45 (2002) 4847–4857.
[37] M. Rarey, B. Kramer, T. Lengauer, G. Klebe, J. Mol. Biol. 261 (1996) 470–489.
[38] B. Kramer, M. Rarey, T. Lengauer, Proteins: Struct., Funct., Genet. 37 (1999) 228–241.
[39] E.R. Wood, A.T. Truesdale, O.B. McDonald, D. Yuan, A. Hassell, S.H. Dickerson, B. Ellis, C. Pennisi, E. Horne, K. Lackey, K.J. Alligood, D.W. Rusnak, T.M. Gilmer, L. Shewchuk, Cancer Res. 64 (2004) 6652–6659.
[40] W. Sippl, J. Comput. Aided Mol. Des. 14 (2000) 559–572.
[41] A. Tropsha, in: T. Oprea (Ed.), Cheminformatics in Drug Discovery, Wiley, Weinheim, 2005, p. 437.
[42] D.V. Pednekar, M.A. Kelkar, S.R. Pimple, K.G. Akamanchi, Med. Chem. Res. 13 (2004) 605–618.