J Mol Model (2010) 16:263–277
DOI 10.1007/s00894-009-0534-x

ORIGINAL PAPER

# In silico QSAR studies of anilinoquinolines as EGFR inhibitors

**Farhan Ahmad Pasha · Muhammad Muddassar · Anil Kumar Srivastava · Seung Joo Cho**

Received: 15 February 2009 / Accepted: 9 April 2009 / Published online: 10 July 2009


**Abstract** Members of the epidermal growth factor receptor (EGFR) family of proteins are frequently overactive in solid tumors. A relatively new therapeutic approach to inhibit the kinase activity is the use of ATP-competitive small molecules. In silico techniques were employed to identify the key interactions between inhibitors and their protein receptors. A series of EGFR inhibitory anilinoquinolines was studied within the framework of hologram quantitative structure activity relationship (HQSAR), density functional theory (DFT)-based QSAR, and three-dimensional (3D) QSAR (CoMFA/CoMSIA). The HQSAR analysis implied that substitutions at certain sites on the inhibitors play an important role in EGFR inhibition. DFT-based QSAR results suggested that steric and electronic interactions contributed significantly to the activity. Ligand-based 3D-QSAR and receptor-guided 3D-QSAR analyses such as CoMFA and CoMSIA techniques were carried out, and the results corroborated the previous two approaches. The 3D QSAR models indicated that steric and hydrophobic interactions are dominant, and that substitution patterns are an important factor in determining activity. Molecular docking was helpful in identifying a bioactive conformer as well as a plausible binding mode. The docked geometry-based CoMFA model with steric and electrostatic fields effect gave $q^2$=0.66, $r^2$= =0.94 with $r^2_{predictive}$=0.72. Similarly, CoMSIA with hydrophobic field gave $q^2$=0.59, $r^2$=0.85 with $r^2_{predictive}$=0.63. Bulky groups around site 3 of ring "C", and hydrophilic and bulky groups at position 6 of ring "A" are desirable, with a hydrophobic and electron-donating group at site 7 of ring "A" being helpful. Accordingly, potential EGFR inhibitors may be designed by modification of known inhibitors.

**Keywords** 3D-QSAR · CoMFA · Epidermal growth factor receptor · Density functional theory · Anilinoquinazoline · Kinase inhibitors

**Dedication** This manuscript is dedicated to Dr. P. P. Singh, M. L. K. College Balrampur on the occasion of his 75th birthday.

F. Ahmad Pasha
Computational Science Center,
Korea Institute of Science and Technology,
PO Box 131, Cheongryang,
Seoul 130-650, Korea

S. Joo Cho (✉)
Research Center for Resistant Cells, Chosun University,
Gwangju 501-759, Republic of Korea
e-mail: chosj@chosun.ac.kr

S. Joo Cho
College of Medicine, Chosun University,
375 Seosuk-dong,
Dong-gu Gwangju 501-759, Rep. of Korea

M. Muddassar
School of Science, University of Science and Technology,
52 Eoeun-dong, Yuseong-gu,
Daejeon 305-333, South Korea

A. Kumar Srivastava
Department of Chemistry, M. L. K. P. G. College,
Balrampur, India

*Present Address:*
F. Ahmad Pasha
Institute de Biologie Structurale,
41, Rue Jules Horowitz,
Grenoble-Cedex 38027, France

## Introduction

Epidermal growth factor receptor (EGFR) is a growth factor receptor kinase that has been implicated in different kinds of cancer [1]. EGFR is overexpressed in numerous tumors derived from brain, lung, bladder, head, and neck [2, 3], implicating EGFR inhibitors as potential anticancer drugs. Recently, a large number of compounds have been

synthesized and evaluated as EGFR inhibitors, with special attention being paid to compounds having a phenyl amino pyrimidine moiety in their structures [4–8]. Computational techniques and quantitative structure activity relationship (QSAR) analysis have been applied successfully to explore the possibilities of potential inhibitors further. The combined application of seemingly disparate methods such as hologram quantitative structure activity relationship (HQSAR), and density functional theory (DFT)-based QSAR, as well as three-dimensional (3D)-QSAR techniques such as comparative molecular field analysis (CoMFA) and comparative molecular similarity analysis (CoMSIA) have proven quite successful [9–15].

Anilinoquinolines are well-known EGFR inhibitors, as demonstrated in 2D and 3D-QSAR studies [16–18]. The X-ray structure of EGFR in complex with an inhibitor has also been published [19], which will facilitate further structure-based design. In this study, we report on HQSAR, DFT-based QSAR, ligand-based 3D-QSAR, and receptor-guided 3D-QSAR analyses of 58 anilinoquinoline derivatives. HQSAR analysis was performed by varying parameters concerning atom (A), bond (B), connectivity (C), hydrogen atom (H) and H-bonding donor/acceptor (D/A). DFT-based descriptors were used to study electronic interactions. The 3D field analyses utilized CoMFA [20] and CoMSIA [21]. The EGFR structure was taken from the protein data bank (http://www.rcsb.org/pdb/; PDB 1M17) and used as a receptor site for anilinoquinolines.

## Theory

DFT-based descriptors have proven immensely useful in predicting the reactivity of atoms and molecules as well as site selectivity [22, 23]. In DFT-based frontier molecular orbital examinations—highest occupied molecular orbital ($\varepsilon_{HOMO}$) and lowest unoccupied molecular orbital ($\varepsilon_{LUMO}$)—several global chemical descriptors of molecules such as hardness (η) [23] and global softness (S) [23], chemical potential (μ) [22], electronegativity (χ) [24], and electrophilicity index (ω) [25] have been used widely in QSAR. Reasonably, the $\varepsilon_{HOMO}$ $\varepsilon_{LUMO}$ energies play important roles in different reactions and can be used as descriptors.

The softness of an atom in a molecule was first described by Klopman [26] and modified by Singh et al. [27]. The Klopman Equations 1 and 2 are as follows.

$$E_m^{\ddagger} = IP_m - a^2(IP_m - EA_m) - \left[\frac{\chi_s\left(C_s^m\right)^2}{R_r}\right](1 - 1/\varepsilon)\left[q_r + 2b^2\chi_r\left(C_r^m\right)^2\right] \quad (1)$$

$$E_n^{\ddagger} = IP_n - b^2(IP_n - EA_n) - \left[\frac{\chi_s\left(C_s^n\right)^2}{R_S}\right](1 - 1/\varepsilon)\left[q_S - 2b^2\chi_s\left(C_s^n\right)^2\right] \quad (2)$$

Where

| | |
|---|---|
| $E_n^{\ddagger}$ | Softness of Lewis acid |
| $E_m^{\ddagger}$ | Softness of a Lewis base |
| $IP$ | Ionization potential of an atom in a molecule |
| $EA$ | Electron affinity of an atom in a molecule |
| $\varepsilon$ | Dielectric constant of the medium in which reaction is carried out |
| $R$ and $EA$ | Radius and charge of atom s and r |
| $C$ | Electron density |
| $\chi_s$ | $q - (q-1)\sqrt{k}$ and $k$=0.75 |
| $a$&$b$ | Variational parameters defined as $a^2 + b^2$=1 |

The ionization potential (IP), electron affinity (EA), charge ($q$), and electron density ($C$) of an atom in a molecule are essential requirements for solution of the Klopman Equations. The IP calculation of an atom in a molecule has been described [28], as has EA [29]. The usefulness of the difference between softness of highest acidic and highest basic site (Eq. 3) has been recently reviewed [35].

$$\Delta E_{nm}^{\ddagger} = \left|E_n^{\ddagger} - E_m^{\ddagger}\right| \quad (3)$$

The charge and electron density of an atom in a molecule can be obtained by DFT calculations. Water medium is taken, hence the value of the dielectric constant was used in the calculation [30].

The molar refractivity is a constitutive-additive property that can be calculated using the Lorenz-Lorentz formula (Eq. 4)

$$MR = \frac{(n^2 - 1 * MW)}{(n^2 + 2 * d)} \quad (4)$$

Where MW is the molecular weight, $n$ is the refractive index, and $d$ the density. The value of MR depends only on the wave longitude of the light used to measure the refractive index.

## Materials and methods

The basic structure of the anilinoquinazolines is shown in Fig. 1.

A set of 58 anilinoquinolines with EGFR inhibitory activity were identified from the literature [31] (Table 1). The compounds were divided randomly into a training set ($n$=46) and test set ($n$=12). The $IC_{50}$ values (i.e., the concentration (μM) of inhibitor that produces 50% inhibition of EGFR) were converted into $pIC_{50}$=−log$IC_{50}$ (Table 1).

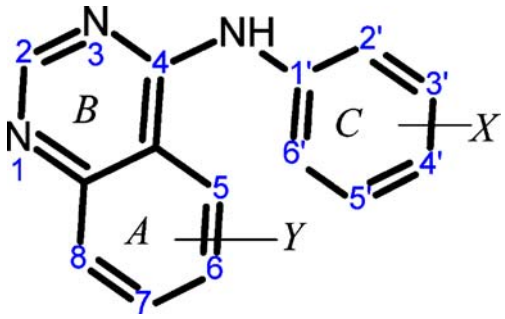


**Fig. 1** Basic structure of anilinoquinolines

**Table 1** $pIC_{50}$ values of 4-(X-phenylamino)-Y-quinazoline derivatives

| Compound no. | X | Y | pIC50 |
|---|---|---|---|
| 1 | H | H | 6.46 |
| 2 | 3-F | H | 7.25 |
| 3 | 3-Cl | H | 7.64 |
| 4 | 3-Br | H | 7.57 |
| 5 | 3-I | H | 7.1 |
| 6 | 3-CF3 | Ha | 6.24 |
| 7 | H | 6-OMe | 7.26 |
| 8 | 3-Br | 6-OMe | 7.52 |
| 9 | H | 6-NH2 | 6.11 |
| 10 | 3-CF3 | 6-NH2 | 6.24 |
| 11 | 3-Br | 6-NH2 | 9.11 |
| 12 | H | 6-NO2 | 5.3 |
| 13 | 3-Br | 6-NO2 | 6.05 |
| 14 | H | 7-OMe | 6.92 |
| 15 | 3-Br | 7-OMe | 8 |
| 16 | H | 7-NH2 | 7 |
| 17 | 3-F | 7-NH2 | 8.7 |
| 18 | 3-Cl | 7-NH2 | 9.6 |
| 19 | 3-Br | 7-NH2 | 10 |
| 20 | 3-I | 7-NH2 | 9.46 |
| 21 | 3-CF3 | 7-NH2 | 8.48 |
| 22 | H | 7-NO2 | 4.92 |
| 23 | 3-F | 7-NO2 | 5.22 |
| 24 | 3-Cl | 7-NO2 | 6.09 |
| 25 | 3-Br | 7-NO2 | 6 |
| 26 | 3-I | 7-NO2 | 6.27 |
| 27 | H | 6,7-di-OMe | 7.54 |
| 28 | 3-F | 6,7-di-OMe | 8.42 |
| 29 | 3-Cl | 6,7-di-OMe | 9.51 |
| 30 | 3-Br | 6,7-di-OMe | 10.6 |
| 31 | 3-I | 6,7-di-OMe | 9.05 |
| 32 | 3-CF3 | 6,7-di-OMe | 9.62 |
| 33 | 3-Br | 6-NHMe | 8.4 |
| 34 | 3-Br | 6-NMe2 | 7.08 |
| 35 | 3-Br | 7-OH | 8.33 |
| 36 | 3-Br | 7-NHMe | 8.16 |
| 37 | 3-Br | 7-NHC2H5 | 7.92 |
| 38 | 3-Br | 7-NMe2 | 7.96 |
| 39 | 3-Br | 6,7-di-NH2 | 9.92 |
| 40 | 3-Br | 6-NH2,7-NHMe | 9.16 |
| 41 | 3-Br | 6-NH2,7-NMe2 | 6.8 |
| 42 | 3-Br | 6-NH2,7-OMe | 8.42 |
| 43 | 3-Br | 6-NH2, 7-Cl | 8.19 |
| 44 | 3-Br | 6-NO2, 7-NH2 | 7.28 |
| 45 | 3-Br | 6-NO2, 7-NHMe | 7.17 |
| 46 | 3-Br | 6-NO2, 7-NMe2 | 5.7 |
| 47 | 3-Br | 6-NO2,7-OMe | 7.82 |
| 48 | 3-Br | 6-NO2,7-Cl | 7.6 |
| 49 | 3-Br | 6,7-di-OHa | 9.77 |
| 50 | 3-Br | 6,7-di-OC2H5 | 11.22 |
| 51 | 3-Br | 6,7-di-OC3H7 | 9.77 |
| 52 | 3-Br | 6,7-di-OC4H9 | 6.98 |
| 53 | 3-Br | 5,6-di-OMe | 5.86 |
| 54 | 3-Br | 5,6,7-tri-OMe | 9.17 |
| 55 | 2-Br | 6,7-di-OMe | 6.89 |
| 56 | 4-Br | 6,7-di-OMe | 9.02 |
| 57 | 3,4-di-Br | 6,7-di-OMe | 10.14 |
| 58 | 3,5-di-Br | 6,7-di-OMe | 6.95 |

## Computational details

### *Hologram quantitative structure activity relationship*

HQSAR is a promising tool with which to establish the relationship between the structure of a compound and its biological activities. Advantages of this approach include the ability to achieve molecular alignment and conformational specification, as well as selection and calculation or measurement of physicochemical descriptors. In this study, several HQSAR models were developed using Sybyl 7.3 software (Tripos, St. Louis, MO). HQSAR was performed using fragment parameters A, B, C, H, and DA. Several combinations of these parameters were considered using fragment size (3–8), with 12 default series of hologram lengths of 53, 59, 61, 71, 83, 97, 151, 199, 257, 307, 353, and 401 bins. An optimum number of components (LV) were selected in each case.

### *DFT analysis*

All geometries were optimized at the B3LYP/6-31G level using polarizable continuum model (PCM) solvation for water in the Gaussian03 package [31] (http://www.gaussian.com/

citation_g03.htm). The total energy, electrophilicity index, and other values were calculated using standard working equations given in theory. The Klopman atomic softness in terms of $E_n$ and $E_m$ values were calculated at each atom of every molecule using necessary values from Gaussian results. The highest $E_n$ value (corresponding to the highest acidic site) and highest $E_m$ values (corresponding to the highest basic site) of each molecule were identified and their difference ($\Delta E_{nm}$) was used as a descriptor with other important properties such as total energy and molar refractivity (MR) [32].

### *Three-dimensional-quantitative structure activity relationship analysis*

The analysis utilized CoMFA and CoMSIA, which themselves utilized Sybyl ver. 7.3 software (Tripos) running on a Linux cluster. Initially, molecular geometries were minimized using Tripos force field (TFF) [33] in conjunction with Gasteiger–Huckel charges, distance-dependent dielectric, and conjugate gradient methods. The convergence criterion was 0.01 kcal/mol. Two different geometrical schemes (AM1 and docked-based), in conjunction with three kinds of charges, namely Gasteiger Huckel (GH), Gasteiger Marsili (GM), and molecular mechanics (MM), were used to develop the QSAR models. In CoMFA, both the steric and electrostatic fields were used as descriptors. In CoMSIA, all five fields (steric, electrostatic, hydrophobic, donor, and acceptor) were used as descriptors.

### *Comparative molecular field analysis*

The steric and electrostatic potential fields for CoMFA was calculated at each lattice intersection of a regularly spaced grid of 2.0 Å. The lattice was defined automatically and extended up to four units past van der Waals volume of all molecules in the X, Y, and Z directions. The van der Waals potential and columbic terms, which represent steric and electrostatic fields, respectively, were calculated using TFF and distance-dependent dielectric constant. An $sp^3$ carbon atom with a van der Waals radius of 1.52 Å and charge of +1.0 served as the probe atom to calculate the steric and electrostatic fields. The steric and electrostatic contributions were truncated to ±30 kcal $mol^{-1}$ and electrostatic contributions were ignored at lattice intersections with maximum steric interactions. The CoMFA steric and electrostatic fields were generated and scaled by the CoMFA standard option given in the Sybyl software.

### *Comparative molecular similarity analysis*

The reported CoMSIA method is based on molecular similarity indices with the same lattice box used for CoMFA. Molecular similarity was expressed in terms of five different properties (steric, electrostatic, hydrophobic, H-bond donors, and acceptors) calculated using a $C^+$ probe atom with a radius of 1.0 Å placed at a regular grid spacing of 2.0 Å. CoMSIA similarity indices ($A_F$) for molecule $j$ with atoms $i$ at a grid point $q$ were calculated using Eq. 5

$$A_{F,K}^{q}(j) = -\sum \omega_{prob,k}\omega_{ik}e^{-\alpha r^2}iq \tag{5}$$

where $k$ represents the following physicochemical properties: steric electrostatic, hydrophobic, H-bond donor, and H-bond acceptor. A Gaussian-type distance dependence was used between grid point $q$ and each atom $i$ of the molecule. The default value (0.3) was used as the attenuation factor ($\alpha$). The steric indices are related to the third power of the atomic radii, electrostatic descriptors are derived from atomic partial charges, hydrophobic fields are derived from atom-based parameters [34] and H-bond donor and acceptor indices are obtained by a rule-based method based on experimental results [35].

## Statistics

### *Multiple linear regression analysis*

Multiple linear regression analysis (MLR) analysis was performed using DataFit software [36] (http://www.curvefitting.com/index.html) to derive QSAR models. Quantum chemical descriptors were used as independent variables, and the $pIC_{50}$ value was used as the dependent variable. A systematic search was performed to determine the significant descriptors. To minimize the effect of co-linearity and avoid redundancy, a correlation matrix was developed and the exact linear variables were not used in final model.

### *Partial least square analysis and validation of QSAR models*

To derive 3D-QSAR models, CoMFA and CoMSIA descriptors were used as independent variables, and $pIC_{50}$ values were used as the dependent variable. The partial least squares (PLS) method was used to linearly correlate these CoMFA and CoMSIA descriptors to activity. The CoMFA cutoff values were set at 30 kcal $mol^{-1}$ for both steric and electrostatic fields, and all fields were scaled by the default options in Sybyl. Cross-validation analysis was performed using the leave one out (LOO) method, in which one compound is removed from the data set and its activity predicted using the model derived from the rest of the dataset. The cross-validated correlation coefficient ($q^2$) resulted in the optimum number of components, and the

lowest standard error of prediction considered for further analysis was calculated using Eq. 6:

$$q^2 = 1 - \frac{\sum_y \left(\gamma_{pred} - \gamma_{observed}\right)^2}{\sum_y \left(\gamma_{observed} - \gamma_{mean}\right)^2} \quad (6)$$

$$PRESS = \sum_y \left(\gamma_{predicted} - \gamma_{observed}\right)^2 \quad (7)$$

Where, $\gamma_{pred}$, $\gamma_{actual}$ and $\gamma_{mean}$ are the predicted, actual, and mean values of the target property ($pIC_{50}$), respectively, and PRESS is the sum of squared deviation between predicted and observed activities of the training set molecules (calculated using Eq. 7). Non-cross-validated PLS analyses were performed with a column filter value of 2.0 to reduce analysis time with a small effect on the $q^2$ values. To achieve robustness and statistical confidence in the derived models, bootstrapping analysis was used for 10 runs. To assess the predictive power of the derived QSAR models, the activity of a test set of 12 molecules was predicted. The predictive abilities of the models were expressed by the predictive $r^2$ value, which is analogous to cross-validated $r^2$ ($q^2$) as calculated using Eq. 8:

$$r^2_{pred} = \frac{SD - PRESS}{SD} \quad (8)$$

where SD is the sum of the squared deviations between the biological activities of the test set and mean activities of the training molecules.

## Results

### Hologram quantitative structure activity relationship analysis

HQSAR was performed on 58 anilinoquinolines using a training set of 46 molecules and a test set of 12 molecules (Table 1). The statistical summary of the HQSAR analysis shown in Table 2 indicates that the best model obtained by the combination of fragment A/H type with length 199 provided good statistics ($q^2$=0.67, $r^2$=0.90, Ensemble= 0.52, SE=0.51, Component=6, length=199). The activities of the test and training sets predicted by this model are reported in Table 3 with other statistical values. The model is quite reliable as it is based on the relationship of substituents and hydrogen atoms with biological activity. Since the structures of the current series are very similar, and the variation of activity is a function of substituents, the presence of any substituents or hydrogen at certain sites might be useful to describe the activity. The fragment contribution to the activity of molecules 23 and 30 are displayed in Figs. 2 and 3, respectively. In Fig. 2, ring "A" contains a nitro ($NO_2$) group, which strongly disfavors activity, as is apparent by the brown color. This moiety affected the contribution to activity of both rings "A" and "B". Compound 30 has a similar skeleton, with two methoxy groups at ring "A"; both rings "A" and "B" contributed to activity. Ring C in compound 23 has no substitution and the hydrogen atom disfavored activity, while in compound 30 ring C contributed positively to activity due to one chloro substituent. These results imply that an electron-withdrawing group such as the $NO_2$ moiety at position 6 of ring "A" is unfavorable for activity, as shown by the red color in Fig. 2. The effect of such a group was spread up to ring "B", which contributed negatively to activity. Electron-donating groups such as the $OCH_3$ moiety at ring "A" were favorable for activity. This was supported by the present observation of the retention of an electron-donating group at ring "A" by the highly active molecule. Similarly, an electronegative group at ring "C" contributed to enhanced activity.

**Table 2** Regression summary of hologram quantitative structure activity relationship (HQSAR) models. $q^2$ Cross validated correlation coefficient, $r^2$ correlation coefficient, *SE* standard error, *n* number of PLS component, *L* length of hologram, *A* atoms, *B* bonds, *C* connections, *H* hydrogen atoms, *DA* donor and acceptor atoms

| | $q^2$ | $r^2$ | *E* | SE | *L* |
|---|---|---|---|---|---|
| A | 0.45 | 0.88 | 0.34 | 0.55 | 353 |
| B | 0.57 | 0.87 | 0.46 | 0.59 | 61 |
| C | 0.45 | 0.86 | 0.38 | 0.6 | 401 |
| H | 0.68 | 0.84 | 0.62 | 0.65 | 97 |
| DA | 0.45 | 0.66 | 0.4 | 0.92 | 257 |

### DFT-based QSAR

The HQSAR study demonstrated that the biological activity of the tested anilinoquinolines is significantly related to the molecular substituents. Accordingly, we utilized DFT-based QSAR to consider electronic interactions. All molecules were fully optimized at DFT B3LYP/6-31G level, and the electronic populations were calculated using the Mulliken population analysis (MPA) scheme. The most important global electronic properties, such as the electronegativity and electrophilicity indices, were calculated and correlation with biological activity was attempted in conjunction with molar refractivity (MR) and solvent accessible surface area (SASA). This approach was unsuccessful. Atomic properties such as atomic softness $E_n$ and $E_m$ (given by KLOPMAN) were also calculated at all sites of every molecule.

**Table 3** Observed and predicted activities of training and test sets by two different HQSAR models. *$pIC_{50}$* Observed activity, *$PA_{19}$* predicted activity by model 19, *NC* net hologram contribution, *$PA_{AVG}$* average predicted activity, *$SE_{PRED}$* standard error in prediction, *SEM* standard error of mean

| Compound no. | $pIC_{50}$ | $PA_{19}$ | NC | $PA_{AVG}$ | $SE_{PRED}$ | SEM |
|---|---|---|---|---|---|---|
| Training set | | | | | | |
| 1 | 6.46 | 5.84 | −1.19 | 5.83 | 0.26791 | 0.080778 |
| 2a | 7.25 | 6.105 | −0.44 | 6.431 | 0.224356 | 0.067646 |
| 5 | 7.1 | 6.911 | −0.923 | 7.049 | 0.195892 | 0.059064 |
| 6 | 6.24 | 6.091 | −0.761 | 6.276 | 0.531036 | 0.160113 |
| 7 | 7.26 | 6.46 | −0.852 | 6.491 | 0.190511 | 0.057441 |
| 8 | 7.52 | 7.58 | −0.362 | 7.693 | 0.16289 | 0.049113 |
| 9 | 6.11 | 6.411 | −1.033 | 6.683 | 0.418944 | 0.126316 |
| 13 | 6.05 | 5.957 | −1.047 | 6.061 | 0.31476 | 0.094904 |
| 14 | 6.92 | 7.251 | −0.497 | 7.194 | 0.259751 | 0.078318 |
| 15 | 8 | 8.371 | −0.244 | 8.396 | 0.243991 | 0.073566 |
| 19 | 10 | 9.598 | 1.037 | 9.422 | 0.399012 | 0.120307 |
| 20 | 9.46 | 9.549 | 0.976 | 9.438 | 0.383097 | 0.115508 |
| 21 | 8.48 | 8.728 | 0.951 | 8.665 | 0.407253 | 0.122791 |
| 22 | 4.92 | 5.254 | −1.197 | 5.012 | 0.182816 | 0.055121 |
| 23 | 5.22 | 5.519 | −1.427 | 5.612 | 0.256013 | 0.077191 |
| 24 | 6.09 | 5.933 | −1.195 | 6.041 | 0.244761 | 0.073798 |
| 25 | 6 | 6.375 | −0.942 | 6.214 | 0.15754 | 0.047500 |
| 26 | 6.27 | 6.326 | −1.094 | 6.23 | 0.124337 | 0.037489 |
| 27 | 7.54 | 8.183 | 0.094 | 7.866 | 0.164572 | 0.049620 |
| 28 | 8.42 | 8.448 | 0.22 | 8.466 | 0.240048 | 0.072377 |
| 29 | 9.51 | 8.862 | 0.538 | 8.895 | 0.283039 | 0.085340 |
| 30 | 10.6 | 9.304 | 0.624 | 9.068 | 0.253834 | 0.076534 |
| 31 | 9.05 | 9.255 | 0.728 | 9.084 | 0.206406 | 0.062234 |
| 33a | 8.4 | 7.821 | −0.173 | 7.976 | 0.285217 | 0.085996 |
| 35 | 8.33 | 8.316 | −0.182 | 8.296 | 0.265145 | 0.079944 |
| 36 | 8.16 | 7.821 | −0.173 | 7.976 | 0.285217 | 0.085996 |
| 37 | 7.92 | 7.443 | −0.115 | 7.706 | 0.311056 | 0.093787 |
| 38 | 7.96 | 7.983 | −0.056 | 7.703 | 0.575695 | 0.173579 |
| 39 | 9.92 | 9.892 | 1.246 | 9.774 | 0.262678 | 0.079201 |
| 40 | 9.16 | 9.033 | 0.713 | 8.781 | 0.181971 | 0.054866 |
| 41 | 6.8 | 6.994 | −0.811 | 7.08 | 0.358769 | 0.108173 |
| 42 | 8.42 | 8.123 | −0.031 | 8.462 | 0.288241 | 0.086908 |
| 43 | 8.19 | 8.351 | 0.147 | 8.618 | 0.223424 | 0.067365 |
| 44 | 7.28 | 8.02 | 0.243 | 7.995 | 0.31187 | 0.094032 |
| 45 | 7.17 | 7.263 | −0.353 | 7.263 | 0.258279 | 0.077874 |
| 46 | 5.7 | 5.582 | −2.044 | 5.688 | 0.295499 | 0.089096 |
| 47 | 7.82 | 7.291 | −0.246 | 7.26 | 0.2877 | 0.086745 |
| 48 | 7.6 | 7.101 | −0.343 | 7.263 | 0.138883 | 0.041875 |
| 49 | 9.77 | 10.234 | 0.822 | 9.798 | 0.616906 | 0.186004 |
| 50 | 11.22 | 11.33 | 3.063 | 11.345 | 0.301932 | 0.091036 |
| 51 | 9.77 | 10.056 | 1.741 | 9.672 | 0.310653 | 0.093666 |
| 52 | 6.98 | 6.845 | −1.041 | 7.055 | 0.185805 | 0.056022 |
| 53 | 5.86 | 7.079 | −0.954 | 7.272 | 0.348746 | 0.105151 |
| 54 | 9.17 | 8.475 | 0.281 | 8.533 | 0.158539 | 0.047801 |
| 55 | 6.89 | 7.871 | 0.094 | 8.085 | 0.152051 | 0.045845 |
| 58a | 6.95 | 10.224 | 1.15 | 10.069 | 0.426528 | 0.128603 |

**Table 3** (continued)

| Compound no. | $pIC_{50}$ | $PA_{19}$ | NC | $PA_{AVG}$ | $SE_{PRED}$ | SEM |
|---|---|---|---|---|---|---|
| Test set | | | | | | |
| 3 | 7.64 | 6.519 | −0.661 | 6.859 | 0.341503 | 0.102967 |
| 4 | 7.57 | 6.96 | −0.465 | 7.033 | 0.228562 | 0.068914 |
| 10 | 6.24 | 6.661 | −0.586 | 7.129 | 0.776584 | 0.234149 |
| 11 | 9.11 | 7.531 | −0.263 | 7.886 | 0.339778 | 0.102447 |
| 12 | 5.3 | 4.837 | −1.733 | 4.859 | 0.40408 | 0.121835 |
| 16 | 7 | 8.478 | 0.467 | 8.219 | 0.311324 | 0.093868 |
| 17 | 8.7 | 8.742 | 0.569 | 8.82 | 0.433235 | 0.130625 |
| 18 | 9.6 | 9.156 | 0.858 | 9.248 | 0.415339 | 0.12523 |
| 32 | 9.62 | 8.434 | 0.096 | 8.312 | 0.460672 | 0.138898 |
| 34 | 7.08 | 7.983 | −0.056 | 7.703 | 0.575695 | 0.173579 |
| 56 | 9.02 | 9.007 | 0.33 | 8.772 | 0.22075 | 0.066559 |
| 57 | 10.14 | 9.438 | 0.716 | 9.451 | 0.252834 | 0.076232 |

The highest $E_n$ value (highest electron accepting tendency) and highest $E_m$ value (highest electron donating tendency) were identified and used as descriptors. Besides these descriptor values, one indicator parameter ($I$) was also used. All molecules having two substitutions at ring "A" were allotted values of $I=1$ and the remainder had $I=0$. The DFT-based regression Eq. 9 was developed, and showed a high correlation as clear from values $r^2_{CV}=0.75$ and $r^2=0.80$.

$$PA_{DFT} = -2.0703\varepsilon_{HOMO} - 0.7483\varepsilon_{LUMO} + 0.043_W - 0.194MR + 2.0234I - 0.036\Delta E_{nm} + 17.82$$
$$r^2{}_{CV} = 0.75r^2 = 0.80 \quad (9)$$

This model involved steric parameters such as MR and molecular weight, and electronic parameters like eigenvalues of the frontier orbital and the difference of softness $\Delta E_{nm}$. The study revealed that steric bulk and electronic interaction contributed significantly to EGFR inhibitory activity. The predicted activities of the training and test set are reported in Table 4.

## Three-dimensional-quantitative structure activity relationship analysis

3D-QSAR analyses, particularly CoMFA and CoMSIA, require a well-defined geometry of the molecule because the molecular activities are very sensitive to conformation. This stable and well-defined geometry may be obtained from quantum chemical methods, which is quite reliable in the absence of crystal structure of ligand receptor complexes; the crystal structure of EGFR (1M17) with a ligand very similar to the current series was recently reported [18].

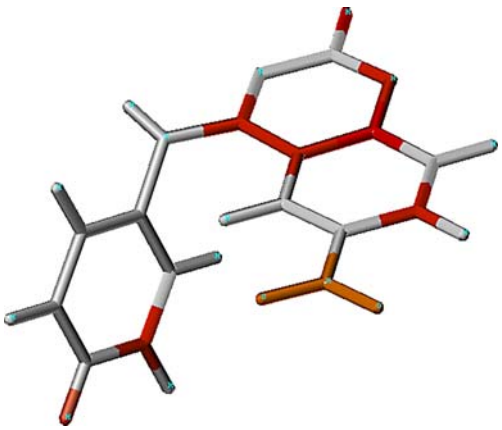

**Fig. 2** Contribution of parameter A (atoms)/H (hydrogen atoms) to least active molecule activity

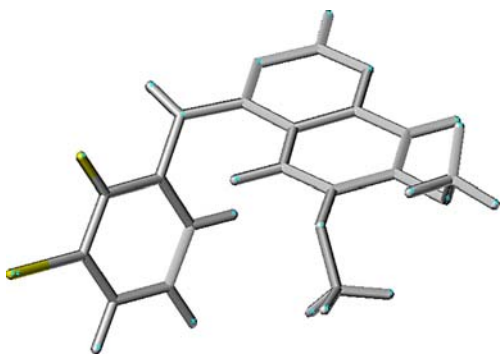

**Fig. 3** Contribution of parameter A/H to highest active molecule activity

**Table 4** Observed and predicted activities of training and test set of compounds by density functional theory (DFT) models. $\varepsilon_{HOMO}$ Energy of highest occupied molecular orbital, $\varepsilon_{LUMO}$ energy of lowest unoccupied molecular orbital, *MW* molecular weight, *MR* molar refractivity, *I* indicator parameter, $\Delta E_{nm}$ difference in acidic and basic softness, *pIC50* observed activity, *PA* predicted activity

| Compound no. | $\varepsilon_{HOMO}$ | $\varepsilon_{LUMO}$ | MW | MR | I | $\Delta E_{nm}$ | pIC50 | PA |
|---|---|---|---|---|---|---|---|---|
| Training set | | | | | | | | |
| 1 | 5.660096 | 1.469448 | 221.261 | 67.632 | 0 | −136.447 | 6.46 | 6.321 |
| 2 | 5.796156 | 1.551084 | 239.251 | 67.849 | 0 | −136.518 | 7.25 | 6.713 |
| 5 | 6.041065 | 1.795992 | 347.157 | 80.04 | 0 | −100.639 | 7.1 | 7 |
| 6 | 5.877793 | 1.578296 | 289.259 | 73.606 | 0 | −131.467 | 6.24 | 7.374 |
| 7 | 5.660096 | 1.415024 | 251.287 | 74.095 | 0 | −136.834 | 7.26 | 6.412 |
| 8 | 5.823368 | 1.523872 | 330.183 | 81.718 | 0 | −136.631 | 7.52 | 7.9 |
| 9 | 5.469612 | 1.197328 | 236.276 | 72.333 | 0 | −137.975 | 6.11 | 6.707 |
| 13 | 6.013852 | 3.102168 | 345.155 | 82.58 | 0 | −135.125 | 6.05 | 6.747 |
| 14 | 5.551248 | 1.49666 | 251.287 | 74.095 | 0 | −136.555 | 6.92 | 6.566 |
| 15 | 5.660096 | 1.605508 | 330.183 | 81.718 | 0 | −136.584 | 8 | 8.175 |
| 19 | 5.17028 | 1.387812 | 315.172 | 79.955 | 0 | −138.493 | 10 | 9.117 |
| 20 | 5.4424 | 1.605508 | 362.172 | 84.741 | 0 | −143.641 | 9.46 | 9.67 |
| 21 | 5.197492 | 1.415024 | 304.274 | 78.306 | 0 | −132.88 | 8.48 | 8.689 |
| 22 | 5.905004 | 2.966108 | 266.259 | 74.957 | 0 | −134.675 | 4.92 | 5.143 |
| 23 | 6.041065 | 2.99332 | 284.249 | 75.173 | 0 | −134.93 | 5.22 | 5.583 |
| 24 | 6.095488 | 3.020532 | 300.704 | 79.762 | 0 | −134.759 | 6.09 | 5.26 |
| 25 | 6.068276 | 2.99332 | 345.155 | 82.58 | 0 | −134.874 | 6 | 6.706 |
| 26 | 6.285972 | 3.619196 | 392.155 | 87.365 | 0 | −99.264 | 6.27 | 5.592 |
| 27 | 5.387976 | 1.659932 | 281.313 | 80.559 | 1 | −106.537 | 7.54 | 7.756 |
| 28 | 5.687309 | 1.469448 | 299.304 | 80.775 | 1 | −137.171 | 8.42 | 9.119 |
| 29 | 5.741732 | 1.49666 | 315.759 | 85.363 | 1 | −136.911 | 9.51 | 8.793 |
| 30 | 5.71452 | 1.523872 | 360.21 | 88.181 | 1 | −136.817 | 10.6 | 10.192 |
| 31 | 5.959428 | 1.795992 | 407.21 | 92.967 | 1 | −100.498 | 9.05 | 9.26 |
| 33 | 5.4424 | 1.333388 | 329.199 | 85.449 | 0 | −137.35 | 8.4 | 8.089 |
| 35 | 5.632884 | 1.523872 | 316.157 | 76.949 | 0 | −136.876 | 8.33 | 8.626 |
| 36 | 5.387976 | 1.659932 | 329.199 | 85.449 | 0 | −131.683 | 8.16 | 7.753 |
| 37 | 5.143068 | 1.415024 | 343.225 | 90.197 | 0 | −137.258 | 7.92 | 8.326 |
| 38 | 5.143068 | 1.469448 | 343.225 | 89.684 | 0 | −137.035 | 7.96 | 8.377 |
| 39 | 5.143068 | 1.170116 | 330.186 | 84.656 | 0 | −140.31 | 9.92 | 9.135 |
| 40 | 5.115856 | 1.170116 | 344.213 | 90.149 | 0 | −139.863 | 9.16 | 8.711 |
| 41 | 5.415188 | 1.251752 | 358.24 | 94.384 | 0 | −137.718 | 6.8 | 7.734 |
| 42 | 5.4424 | 1.22454 | 345.198 | 86.419 | 0 | −137.667 | 8.42 | 8.682 |
| 43 | 5.768944 | 1.523872 | 349.617 | 84.76 | 0 | −137.344 | 8.19 | 8.283 |
| 44 | 5.605672 | 3.020532 | 360.169 | 87.28 | 0 | −135.67 | 7.28 | 7.405 |
| 45 | 5.605672 | 3.074956 | 374.196 | 92.774 | 0 | −135.271 | 7.17 | 6.886 |
| 46 | 5.551248 | 2.884472 | 388.223 | 97.008 | 0 | −135.369 | 5.7 | 6.926 |
| 47 | 5.98664 | 3.047744 | 375.181 | 89.043 | 0 | −135.254 | 7.82 | 6.885 |
| 48 | 6.068276 | 3.12938 | 379.6 | 87.385 | 0 | −135.025 | 7.6 | 7.159 |
| 49 | 5.632884 | 1.387812 | 332.156 | 78.643 | 0 | −137.273 | 9.77 | 9.101 |
| 50a | 5.687309 | 1.49666 | 388.263 | 97.677 | 0 | −137.041 | 11.22 | 7.615 |
| 51a | 5.687309 | 1.49666 | 416.317 | 106.726 | 0 | −137.041 | 9.77 | 7.064 |
| 52 | 5.660096 | 1.469448 | 444.37 | 115.928 | 0 | −137.091 | 6.98 | 6.562 |
| 53 | 6.013852 | 1.741568 | 360.21 | 88.181 | 0 | −100.477 | 5.86 | 6.071 |
| 54 | 5.660096 | 1.49666 | 390.236 | 94.645 | 0 | −136.879 | 9.17 | 8.34 |
| 55 | 5.741732 | 1.551084 | 360.21 | 88.181 | 0 | −137.747 | 6.89 | 8.125 |
| 58a | 5.823368 | 1.605508 | 439.106 | 95.804 | 0 | −136.879 | 6.95 | 9.798 |

**Table 4** (continued)

| Compound no. | $\varepsilon_{HOMO}$ | $\varepsilon_{LUMO}$ | MW | MR | I | $\Delta E_{nm}$ | pIC50 | PA |
|---|---|---|---|---|---|---|---|---|
| Test set | | | | | | | | |
| 3 | 5.85058 | 1.578296 | 255.706 | 72.437 | 0 | 136.346 | 7.64 | 6.39 |
| 4 | 5.82336 | 1.551084 | 300.157 | 75.255 | 0 | 136.462 | 7.57 | 7.836 |
| 10 | 5.57846 | 1.306176 | 304.274 | 78.306 | 0 | 132.213 | 6.24 | 7.957 |
| 11 | 5.55124 | 1.306176 | 315.172 | 79.955 | 0 | 137.773 | 9.11 | 8.363 |
| 12 | 5.85058 | 3.074956 | 266.259 | 74.957 | 0 | 135.103 | 5.3 | 5.189 |
| 16 | 5.11586 | 1.306176 | 236.276 | 72.333 | 0 | 138.779 | 7 | 7.386 |
| 17 | 5.17028 | 1.3606 | 254.266 | 72.549 | 0 | −138.55 | 8.7 | 7.957 |
| 18 | 5.19742 | 1.387812 | 270.721 | 77.137 | 0 | 138.391 | 9.6 | 7.691 |
| 32 | 5.74173 | 1.523872 | 349.312 | 86.532 | 1 | 131.828 | 9.62 | 9.806 |
| 34 | 5.4424 | 1.3606 | 343.225 | 89.684 | 0 | 137.337 | 7.08 | 7.849 |
| 56 | 5.63288 | 1.523872 | 360.21 | 88.181 | 1 | 136.879 | 9.02 | 10.36 |
| 57 | 5.74173 | 1.578296 | 439.106 | 95.804 | 1 | 136.977 | 10.14 | 12.014 |

## Molecular alignment

3D alignment of the structures of all the molecules was performed by two different geometrical schemes (1 and 2).

### *Geometrical scheme 1*

In this scheme, a ligand-based technique was used. Random search based, minimum energy conformers were fully optimized at a semi-empirical AM1 level with gnorm= 0.01 and T=3,600. The fully optimized structures were aligned over the template (molecule 30) and used for CoMFA and CoMSIA with three kinds of charges.

*Docking* A co-crystal structure of a very similar ligand was reported recently as PDB 1M17 [18]. This protein was obtained from PDB and modified using the FlexX program (BioSolvIT, Sankt Augustin, Germany). The ligand-based active site was defined at a distance of 6.5 Å. Compound 30 was successfully docked to the receptor site, and 100 possible conformers were generated. The best-fit mode was identified on the basis of total score and similarity to the co-crystallized ligand (Fig. 4).

Residue numbers 694, 702, 719,721, 738, 742, 754, 766, 768, 769, 773, 820, 830, and 831 surrounding the active site, and residue 769 are directly involved in hydrogen bonding with the pyridine ring of the inhibitor [18]. N1 of the quinazoline accepts a hydrogen bond from the Met769 amide nitrogen. The other quinazoline nitrogen atom (N3) is not within hydrogen bonding distance of the Thr766 side chain (4.1 Å), but a water molecule bridges this gap. Such a water molecule has been described [37]. Compound 30 displayed the same surrounding interactions and a direct hydrogen bond with residue 769. This prominent binding mode was used as a template to design other ligands of the series for geometrical scheme 2.

### *Geometrical scheme 2*

This scheme was based on the docked structure of compound 30 to the receptor protein (PDB file 1M17). All ligands were designed by modification of the aforementioned template and were minimized at TFF level within the receptor site. During minimization all protein residues and common moieties of the ligands were constrained. These minimized structures were aligned on the template displayed in Fig. 5, which was used for CoMFA and CoMSIA with three kinds of charges.

## Comparative molecular field analysis

GH, GM, and MM charges were applied to both geometrical schemes. Six models ($CoMFA_1$–$CoMFA_6$) were developed. $CoMFA_1$–$CoMFA_3$ were based on AM1 geometries with GH, GM, and MM charges, respectively. $CoMFA_4$–$CoMFA_6$ were based on docked geometries with GH, GM, and MM charges, respectively. The result of both geometries was good, with the docked geometries being statistically significant as reported in Table 5. The best-fit model ($CoMFA_6$) was based on the docked geometry and an MM charge having high LOO values of $q^2$=0.66 and $r^2$=0.94. This model involved both steric and bulk interactions, but the steric contribution was dominant in the interaction. The model was validated successfully with the test set of 12 compounds and a $r^2_{predictive}$ of 0.72 was obtained. The model is also statistically reliable as it is

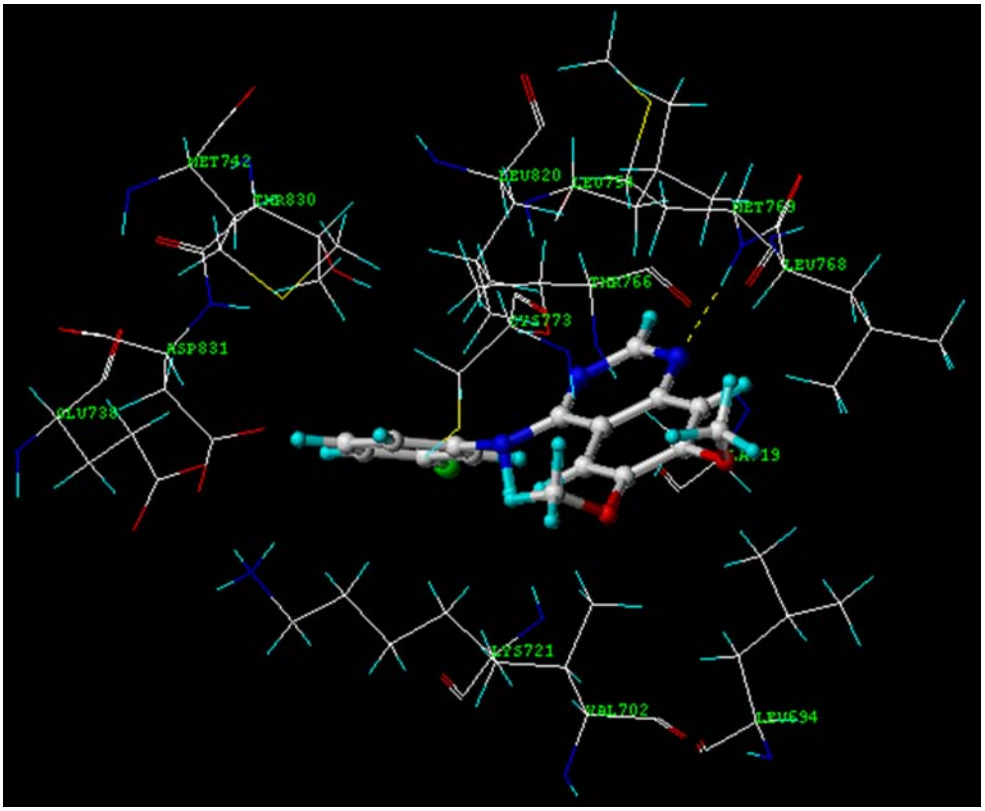


**Fig. 4** Best docked mode of compound 30

based on the best binding mode, all molecules have nearly the same conformation as the co-crystal structure, and all have a key contact hydrogen bonding between residue 769 and the nitrogen atom of ring "B". All molecules acquired the same binding site and an almost equal interaction with the hinge region. The activities predicted by CoMFA are summarized in Table 6.

*CoMFA map*

3D-CoMFA contour maps of the best-fit model $CoMFA_6$ are displayed in Fig. 6 with both kinds of field effects on compound 30. Green contours indicate the area in which steric bulk substations might affect activity beneficially and the yellow region is favorable for small groups.

The blue contour indicates the region where a positive group was required for high activity, while the red zone indicates the region favorable for negative groups. The green contour was evident around position 3 of ring "C" and position 7 of ring A, indicating that a bulkier group would favor higher activity. A small red contour was evident close to position 6 of ring A, indicating that a negative group would favor higher activity. A blue contour also appeared after the red contour at the same position, which was a clear indication that any elongation of the negative group chain should be positive in nature. A methoxy and ethoxy group would be desirable at this position, in support of previous findings[15].

Comparative molecular similarity analysis

CoMSIA was conducted similarly to CoMFA. CoMFA established that docked geometries and MM charges were the more reliable of the parameters tested. CoMSIA involved only docked geometries with MM charges and the five field descriptors. $CoMSIA_1$–$CoMSIA_6$ were based on steric, electrostatic, hydrophobic, hydrogen bond donor, and hydrogen bond acceptor field effects, respectively. The hydropho-

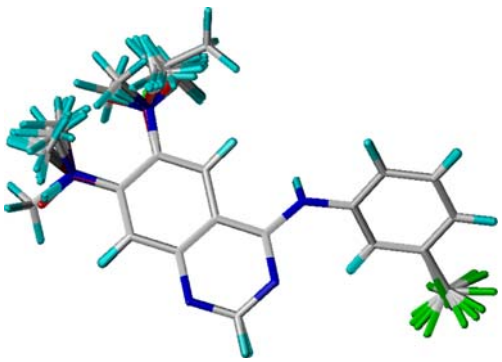

**Fig. 5** Geometrical scheme-2-based aligned structures of all ligands

**Table 5** Regression summary of comparative molecular field analysis (CoMFA) results. $q^2$ Cross validated correlation coefficient, $n$ number of component, $r^2$ correlation coefficient, $F$ Fisher value, $SE$ standard error

| Compound no. | Geometry | Charge[a] | Field | $r^2$ | $n$ | $q^2$ | $F$ | SE |
|---|---|---|---|---|---|---|---|---|
| 1 | AM1 | GH | 0.40S/0.60E | 0.39 | 8 | 0.95 | 90.6 | 0.35 |
| 2 | AM1 | GM | 0.43S/0.57E | 0.30 | 10 | 0.97 | 94.73 | 0.313 |
| 3 | AM1 | MM | 0.45S/0.55E | 0.49 | 9 | 0.97 | 134.6 | 0.27 |
| 4 | Docked | GH | 0.53S/0.47E | 0.64 | 7 | 0.93 | 65 | 0.44 |
| 5 | Docked | GM | 0.51S/0.49E | 0.65 | 7 | 0.93 | 63.34 | 0.45 |
| 6 | Docked | MM | 0.57S/0.42E | 0.66 | 7 | 0.94 | 72.6 | 0.42 |

[a] *GH* Gasteiger Huckel, *GM* Gasteiger Marsili, *MM* molecular mechanics

**Table 6** Observed and predicted activities of training and test set of compounds by best-fit docking based 3D-QSAR models. $pIC_{50}$ Observed activity, $PA_{CoMFA}$ predicted activity by CoMFA, $PA_{CoMSIA}$ predicted activity by CoMSIA

| Compound no. | $pIC_{50}$ | $PA_{CoMFA}$ | Residual | $PA_{CoMSIA}$ | Residual |
|---|---|---|---|---|---|
| Training set | | | | | |
| 1 | 6.46 | 5.647 | 0.813 | 6.365 | 0.095 |
| 2a | 7.25 | 6.126 | 1.124 | 6.873 | 0.377 |
| 5 | 7.1 | 7.239 | −0.139 | 7.772 | −0.672 |
| 6 | 6.24 | 6.489 | −0.249 | 6.55 | −0.31 |
| 7 | 7.26 | 6.985 | 0.275 | 6.787 | 0.473 |
| 8 | 7.52 | 7.727 | −0.207 | 8.01 | −0.49 |
| 9 | 6.11 | 6.182 | −0.072 | 6.406 | −0.296 |
| 13 | 6.05 | 6.235 | −0.185 | 6.563 | −0.513 |
| 14 | 6.92 | 7.081 | −0.161 | 7.459 | −0.539 |
| 15 | 8 | 8.581 | −0.581 | 8.599 | −0.599 |
| 19 | 10 | 9.381 | 0.619 | 9.074 | 0.926 |
| 20 | 9.46 | 9.233 | 0.227 | 9.326 | 0.134 |
| 21 | 8.48 | 8.75 | −0.27 | 8.041 | 0.439 |
| 22 | 4.92 | 4.772 | 0.148 | 4.735 | 0.185 |
| 23 | 5.22 | 5.017 | 0.203 | 5.47 | −0.25 |
| 24 | 6.09 | 6.037 | 0.053 | 5.974 | 0.116 |
| 25 | 6 | 6.18 | −0.18 | 6.285 | −0.285 |
| 26 | 6.27 | 6.136 | 0.134 | 6.41 | −0.14 |
| 27 | 7.54 | 8.055 | −0.515 | 8.078 | −0.538 |
| 28 | 8.42 | 8.547 | −0.127 | 8.585 | −0.165 |
| 29 | 9.51 | 9.249 | 0.261 | 9.021 | 0.489 |
| 30 | 10.6 | 9.379 | 1.221 | 9.258 | 1.342 |
| 31 | 9.05 | 9.202 | −0.152 | 9.389 | −0.339 |
| 33a | 8.4 | 7.333 | 1.067 | 7.316 | 1.084 |
| 35 | 8.33 | 8.256 | 0.074 | 9.125 | −0.795 |
| 36 | 8.16 | 8.145 | 0.015 | 7.579 | 0.581 |
| 37 | 7.92 | 8.088 | −0.168 | 8.281 | −0.361 |
| 38 | 7.96 | 8.088 | −0.128 | 6.755 | 1.205 |
| 39 | 9.92 | 9.785 | 0.135 | 8.923 | 0.997 |
| 40 | 9.16 | 9.058 | 0.102 | 8.491 | 0.669 |
| 41 | 6.8 | 6.706 | 0.094 | 7.54 | −0.74 |
| 42 | 8.42 | 8.416 | 0.004 | 8.436 | −0.016 |

**Table 6** (continued)

| Compound no. | $pIC_{50}$ | $PA_{CoMFA}$ | Residual | $PA_{CoMSIA}$ | Residual |
|---|---|---|---|---|---|
| 43 | 8.19 | 8.584 | −0.394 | 8.132 | 0.058 |
| 44 | 7.28 | 7.988 | −0.708 | 7.761 | −0.48 |
| 45 | 7.17 | 7.434 | −0.264 | 7.451 | −0.281 |
| 46 | 5.7 | 5.58 | 0.12 | 5.412 | 0.288 |
| 47 | 7.82 | 7.526 | 0.294 | 7.472 | 0.348 |
| 48 | 7.6 | 7.253 | 0.347 | 7.03 | 0.57 |
| 49 | 9.77 | 9.742 | 0.028 | 9.665 | 0.105 |
| 50 | 11.22 | 11.634 | −0.414 | 11.777 | −0.557 |
| 51 | 9.77 | 9.526 | 0.244 | 10.026 | −0.256 |
| 52 | 6.98 | 6.771 | 0.209 | 6.612 | 0.368 |
| 53 | 5.86 | 6.672 | −0.812 | 7.43 | −1.57 |
| 54 | 9.17 | 8.636 | 0.534 | 8.517 | 0.653 |
| 55 | 6.89 | 7.318 | −0.428 | 6.736 | 0.154 |
| 58a | 6.95 | 9.443 | −2.493 | 9.333 | −2.383 |
| Test set | | | | | |
| 3 | 7.64 | 7.154 | 0.486 | 7.365 | 0.275 |
| 4 | 7.57 | 7.221 | 0.349 | 7.685 | −0.115 |
| 10 | 6.24 | 7.436 | −1.196 | 6.761 | −0.521 |
| 11 | 9.11 | 7.885 | 1.225 | 7.567 | 1.543 |
| 12 | 5.3 | 4.642 | 0.658 | 5.391 | −0.091 |
| 16 | 7 | 7.646 | −0.646 | 7.5 | −0.5 |
| 17 | 8.7 | 8.153 | 0.547 | 8.071 | 0.629 |
| 18 | 9.6 | 9.227 | 0.373 | 8.749 | 0.851 |
| 32 | 9.62 | 8.753 | 0.867 | 8.341 | 1.279 |
| 34 | 7.08 | 8.162 | −1.082 | 8.04 | −0.96 |
| 56 | 9.02 | 8.454 | 0.566 | 7.762 | 1.258 |
| 57 | 10.14 | 9.502 | 0.638 | 9.003 | 1.137 |

bic field effect-based model ($CoMSIA_3$) provided the best result, with LOO values of $q^2$=0.59, $r^2$=0.85, and $r^2_{predictive}$= 0.63, which was statistically superior. The activities predicted by $CoMSIA_3$ are summarized in Table 6. Table 7 presents the regression summary of CoMSIA models.

*CoMSIA map*

Figure 7 displays the CoMSIA map with the hydrophobic field effect. A purple contour favorable for hydrophobic groups was evident around position 7 of ring A. A small cyan contour favorable for hydrophilic groups appeared at a specific distance from position 6. The results indicate that a hydrophobic group near position 7 of ring A would have a beneficial effect on activity, while a hydrophilic group is desirable at position 6 for better activity.

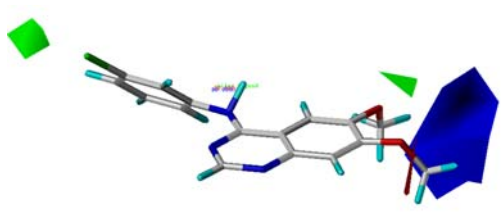

**Fig. 6** Comparative molecular field analysis (CoMFA) contour map of anilinoquinolines

## Discussion

The first SAR analysis of this series of compounds [38] was based on elementary QSAR obtained from physico-chemical aspects. In that study, sites 6 and 7 of ring "A" were identified as being suitable for bulkier groups but there was no quantitative support. Similar studies using HQSAR are still lacking, despite the high predictability. The present HQSAR-based model clearly indicates the contribution of substitutions at sites 6 and 7 of ring "A" to activity. QSAR of same series, and models derived by electronic constant and log P values have been reported

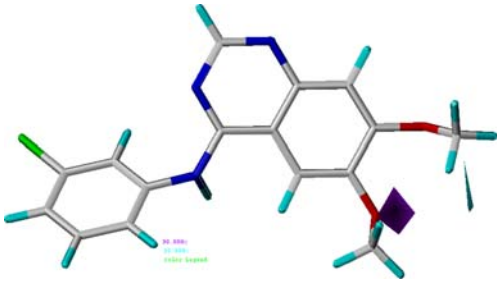

**Fig. 7** CoMSIA contour map of anilinoquinolines

[37], as has the hydrophobic related electronic interaction. The current DFT-based QSAR study reveals that steric and electronic interactions contribute significantly to activity. The current DFT model ($r^2_{CV}$=0.75 $r^2$=0.80) is quite compatible with a previous study ($q^2$=0.81 and $r^2$=0.85, components=5) [37], and the current DFT model was validated additionally by other methods.

A similar series of compounds was studied by another group using another co-crystal structure (1D18; an analogous protein) as a starting geometry, with the remaining analyses utilizing a ligand-based technique [18]. All molecules were optimized using PM3 Hamiltonian and MM charges. In this latter study, the best-fit models (CoMFA $q^2$=0.643, CoMSIA $q^2$=0.651) imply that the steric and electrostatic fields are the most important factors for receptor interaction. However, the study omitted to include other aspects including hydrophobic interactions and hydrogen bonding. Overall, the authors reported that substitutions at position 3 of aniline need an optimal volume and electron density to achieve a maximal inhibitory effect, while electron-donating groups at positions 6 and 7 contribute to improved activity. The predictive $r^2$ for CoMFA was the same as that found in the present study, but for CoMSIA, $r^2_{predictive}$ was 0.80 with nine components and two fields. An earlier 3D QSAR study involved a similar series of compounds [19]. In this case, the authors used an AM1 charge along with docked conformers; CoMFA $q^2$=0.64 and $r^2$=0.97 with a contribution of all available five fields and six components was reported. In contrast to this previous study, here we used three different charges with docked geometries; MM charges were superior than AM1 ($q^2$=0.66 and $r^2_{predictive}$ =0.72 for the test set). Similarly, CoMSIA using docked geometries with MM charge produced $q^2$=0.58 and $r^2_{predictive}$ =0.63 for the test set, with only a hydrophobic field effect. An earlier study using CoMFA reported a major contribution of steric field, while in CoMSIA the major contribution was from the electrostatic field [17]. The present CoMSIA study indicates that the hydrophobic field provides the major contribution. The basis of this dichotomy requires further elucidation.

The current study included ligand-based models with AM1 geometries, which showed significant QSAR, while receptor-guided geometry was used in geometrical scheme-2. The X-ray structure (PDB 1M17) demonstrates that the hinge contact involves the nitrogen moiety of ring B with residue Met769. In this X-ray structure, the compound adopts a similar orientation as for 4-anilino-quinazoline molecules complexed with cyclin-dependent kinase 2 (CDK2) (PDB codes 1di8 and 1di9, respectively) [37]. The EGFR kinase domain (EGFRK) adopts a bilobate-fold involving the N and C lobes. These two lobes are separated by a cleft similar to those in which ATP, ATP analogues, and ATP-competitive inhibitors have been found to bind [19]. In both the apo-EGFRK and inhibitor-bound forms of EGFRK, there is a salt bridge between Lys721 and Glu738 [19], which indicates that EGFR does not require large rearrangements within the N-lobe for catalytic competence. The cancer drug Erlotinib lies with the N1- and C8-containing edge of the quinazoline directed toward the segment connecting the N- and C-lobes, with the ether linkages projecting past the connecting segment into the solvent, and the anilino substituent on the opposite end sequestered in a hydrophobic pocket [19]. In the case of docked-based models, we used the same receptor site as was obtained from the crystal structure. Optimization of all ligands was carried out within the receptor site by freezing the whole protein as well as the central common moiety of ligands, and substituents were allowed to move only under the influence of the receptor site; thus the surrounding

**Table 7** Regression summary of comparative molecular similarity analysis (CoMSIA) results. $q^2$ Cross validated correlation coefficient, *n* number of component, $r^2$ correlation coefficient, *F* Fisher value, *SE* standard error, $r^2_{bs}$ boot strapping correlation coefficient, *SD* standard deviation, $r^2_{predictive}$ predictive correlation coefficient

| No. | Geometry | Charge | Field | $q^2$ | n | $r^2$ | F | SE | $r_{bs}^2$ | SD | $r_{predictive}^2$ |
|---|---|---|---|---|---|---|---|---|---|---|---|
| 1 | Docked | MM | Steric | 0.098 | – | – | – | – | – | – | – |
| 2 | Docked | MM | Electrostatic | 0.352 | 2 | – | – | – | – | – | – |
| 3 | Docked | MM | Hydrophobic | 0.58 | 8 | 0.85 | 23.73 | 0.65 | 0.92 | 0.02 | 0.63 |
| 4 | Docked | MM | Donor | 0.047 | | – | – | – | – | – | – |
| 5 | Docked | MM | Acceptor | 0.21 | 2 | – | – | – | – | – | – |

ligands and the hinge contact was common for all other ligands as per the crystal structure. This resulted in a superior model, as evidenced by the high values of $q^2$=0.66 and $r^2$=0.94 with $r^2_{predictive}$=0.72.

## Conclusions

HQSAR analysis reveals that an electron-donating group at site 6 of ring "A" and electronegative group at site 3 of ring "C" favor higher activity of EGFR inhibitory anilinoquinolines, and that the effect of such groups is global in nature. The least active (compound-23) and most active (compound-30) molecules (Figs. 2, 3) are similar in structure, and the same fragment shows diverse effects due to electron-donating and electron-withdrawing substitutions. The importance of steric bulk with electronic interaction is evident. The ligand-based 3D QSAR model has proven significant, but a more definitive conclusion requires consideration of the receptor site. The receptor-guided model has a high value of $q^2$=0.79 and $r^2$=0.93 ,suggesting that ligands of high activity can be obtained by substituting site 6 ring "A" with electron-donating and hydrophobic groups, while site 7 favors bulky and hydrophilic groups.

**Acknowledgments** This work was supported by the Korea Science and Engineering Foundation (KOSEF) grant funded by the Korea government (MEST) through the Research Center for Resistant Cells (R13-2003-009).

## References

1. Kamath S, Buolamwini JK (2006) Targeting EGFR and HER-2 receptor tyrosine kinases for cancer drug discovery and development. Med Res Rev 26:569–594
2. Salomon DS, Brandt R, Ciardiello F, Normanno N (1995) Epidermal growth factor-related peptides and their receptors in human malignancies. Crit Rev Oncol Hematol 19:183–232
3. Gullick WJ (1991) prevalence of aberrant expression of the epidermal growth-factor receptor in human cancers. Br Med Bull 47:87–98
4. Rewcastle GW, Denny WA, Bridges AJ, Zhou HR, Cody DR, Mcmichael A, Fry DW (1995) Tyrosine kinase inhibitors. 5. Synthesis and structure-activity-relationships for 4-[(phenylmethyl) amino]-quinazolines and 4-(phenylamino)quinazolines as potent adenosine 5′-triphosphate binding-site inhibitors of the tyrosine kinase domain of the epidermal growth-factor receptor. J Med Chem 38:3482–3487
5. Thompson AM, Bridges AJ, Fry DW, Kraker AJ, Denny WA (1995) Tyrosine kinase inhibitors. 7. 7-amino-4-(phenylamino) pyrido[4, 3-d]pyrimidines and 7-amino-4-[(phenylmethyl)amino] pyrido[4, 3-d]pyrimidines—a new class of inhibitors of the tyrosine kinase-activity of the epidermal growth-factor receptor. J Med Chem 38:3780–3788
6. Smaill JB, Palmer BD, Rewcastle GW, Denny WA, McNamara DJ, Dobrusin EM, Bridges AJ, Zhou HR, Showalter HDH, Winters RT, Leopold WR, Fry DW, Nelson JM, Slintak V, Elliot WL, Roberts BJ, Vincent PW, Patmore SJ (1999) Tyrosine kinase inhibitors. 15. 4-(phenylamino)quinazoline and 4-(phenylamino) pyrido[d] pyrimidine acrylamides as irreversible inhibitors of the atp binding site of the epidermal growth factor receptor. J Med Chem 42:1803–1815
7. Rewcastle GW, Palmer BD, Bridges AJ, Showalter HDH, Li S, Nelson J, McMichael A, Kraker AJ, Fry DW, Denny WA (1996) Tyrosine kinase inhibitors.9. Synthesis and evaluation of fused tricyclic quinazoline analogues as ATP site inhibitors of the tyrosine kinase activity of the epidermal growth factor receptor. J Med Chem 39:918–928
8. Smaill JB, Showalter HDH, Zhou HR, Bridges AJ, McNamara DJ, Fry DW, Nelson JM, Sherwood V, Vincent PW, Roberts BJ, Elliott WL, Denny WA (2001) Tyrosine kinase inhibitors. 18. 6-substituted 4-anilinoquinazolines and 4-anilinopyrido [3, 4-d] pyrimidines as soluble, irreversible inhibitors of the epidermal growth factor receptor. J Med Chem 44:429–440
9. Singh PP, Srivastava HK, Pasha FA (2004) DFT-based QSAR study of testosterone and its derivatives. Bioorg Med Chem 12:171–177
10. Pasha FA, Srivastava HK, Singh PP (2005) QSAR study of estrogens with the help of PM3-based descriptors. Int J Quant Chem 104:87–100
11. Srivastava HK, Pasha FA, Singh PP (2005) Atomic softness-based QSAR study of testosterone. Int J Quant Chem 103:237–245
12. Muddassar M, Pasha FA, Chung HW, Yoo KH, Oh CH, Cho SJ (2008) Receptor guided 3D-QSAR: a useful approach for designing of IGF-1R inhibitors. J Biomed Biotechnol 2008: article ID 837653, doi:10.1155/2008/837653
13. Pasha FA, Dal Nam K, Cho SJ (2007) CoMFA based quantitative structure toxicity relationship of azo dyes. Mol Cell Toxicol 3:145–149
14. Pasha FA, Muddassar M, Chung HW, Cho SJ, Cho H (2008) Hologram and 3D-quantitative structure toxicity relationship studies of azo dyes. J Mol Model 14:293–302
15. Pinto-Bazurco M, Tsakovska I, Pajeva I (2006) QSAR and 3D QSAR of inhibitors of the epidermal growth factor receptor. Int J Quant Chem 106:1432–1444
16. Assefa H, Kamath S, Buolamwini JK (2003) 3D-QSAR and docking studies on 4-anilinoquinazoline and 4-anilinoquinoline epidermal growth factor receptor (EGFR) tyrosine kinase inhibitors. J Comp Aided Mol Des 17:475–493
17. Pednekar DV, Kelkar MA, Pimple SR, Akamanchi KG (2004) 3D QSAR studies of inhibitors of epidermal growth factor receptor [EGFR] using CoMFA and GFA methodologies. Med Chem Res 13:605–618
18. Stamos J, Sliwkowski MX, Eigenbrot C (2002) Structure of the epidermal growth factor receptor kinase domain alone and in complex with a 4-anilinoquinazoline inhibitor. J Biol Chem 277:46265–46272
19. Cramer RD, Patterson DE, Bunce JD (1988) Comparative molecular-field analysis (CoMFA). 1. effect of shape on binding of steroids to carrier proteins. J Am Chem Soc 110:5959–5967
20. Klebe G, Abraham U, Mietzner T (1994) Molecular similarity indexes in a comparative analysis (CoMSIA) of drug molecules to correlate and predict their biological activity. J Med Chem 37:4130–4146
21. Chattaraj PK, Nath S, Sannigrahi AB (1994) Hardness, chemical-potential, and valency profiles of molecules under internal rotations. J Phys Chem 98:9143–9145
22. Parr RG, Chattaraj PK (1991) Principle of maximum hardness. J Am Chem Soc 113:1854–1855
23. Parr RG, Donnelly RA, Levy M, Palke WE (1978) Electronegativity—density functional viewpoint. J Chem Phys 68:3801–3807
24. Parr RG, Von Szentpaly L, Liu SB (1999) Electrophilicity index. J Am Chem Soc 121:1922–1924

25. Klopman G (1968) Chemical reactivity and concept of charge- and frontier-controlled reactions. J Am Chem Soc 90:223
26. Singh PP, Srivastava SK, Srivastava AK (1980) Matching between Lewis-acids and Lewis-bases on the basis of quantitative softness values and their relation with the stability of the complexes. J Inorg Nucl Chem 42:521–532
27. Dewar MJS, Morita T (1969) Ground states of conjugated molecules. XII. Improved calculations for compounds containing nitrogen or oxygen. J Am Chem Soc 91:796
28. Singh PP, Srivastava HK, Pasha FA (2004) DFT-based QSAR study of testosterone and its derivatives. Bioorg Med Chem 12:171–177
29. Daniels DJ (1996) Surface-penetrating radar. Institute of Electrical Engineers, London
30. Kurup A, Garg R, Hansch C (2001) Comparative QSAR study of tyrosine kinase inhibitors. Chem Rev 101:2573–2600
31. Frisch MJ (2004) Gaussian, Inc. Wallingford CT
32. Glasstone S (1954) Textbook of physical chemistry D. Van Nostrand, New York
33. Clark M, Cramer RD, Vanopdenbosch N (1989) Validation of the general-purpose Tripos 5.2 force-field. J Comp Chem 10:982–1012
34. Viswanadhan VN, Ghose AK, Revankar GR, Robins RK (1989) Atomic physicochemical parameters for 3 dimensional structures directed quantitative structure–activity relationships.4. additional parameters for hydrophobic and dispersive interactions and their application for an automated superposition of certain naturally occurring nucleoside antibiotics. J Chem Inf Comp Sci 29:163–172
35. Klebe G (1994) The use of composite crystal-field environments in molecular recognition and the de novo design of protein ligands. J Mol Biol 237:212–235
36. DataFit Ev Oakdale Engineering, Oakdale, PA
37. Shewchuk L, Hassell A, Wisely B, Rocque W, Holmes W, Veal J, Kuyper LF (2000) Binding mode of the 4-anilinoquinazoline class of protein kinase inhibitor: X-ray crystallographic studies of 4-anilinoquinazolines bound to cyclin-dependent kinase 2 and p38 kinase. J Med Chem 43:133–138
38. Bridges AJ, Zhou H, Cody DR et al (1996) Tyrosine kinase inhibitors. 8. An unusually steep structure-activity relationship for analogues of 4-(3-bromoanilino)-6, 7-dimethoxyquinazoline (PD 153035), a potent inhibitor of the epidermal growth factor receptor. J Med Chem 39:267–276