Mapping the Binding Site of a Large Set of Quinazoline Type EGF-R Inhibitors Using Molecular Field Analyses and Molecular Docking Studies

Tingjun Hou, Lili Zhu, Lirong Chen, and Xiaojie Xu*

College of Chemistry and Molecular Engineering, Peking University, Beijing 100871, P. R. China

In the current work, three-dimensional QSAR studies for one large set of quinazoline type epidermal growth factor receptor (EGF-R) inhibitors were conducted using two types of molecular field analysis techniques: comparative molecular field analysis (CoMFA) and comparative molecular similarity indices analysis (CoMSIA). These compounds belonging to six different structural classes were randomly divided into a training set of 122 compounds and a test set of 13 compounds. The statistical results showed that the 3D-QSAR models derived from CoMFA were superior to those generated from CoMSIA. The most optimal CoMFA model after region focusing bears significant cross-validated r^2_{cv} of 0.60 and conventional r^2 of 0.92. The predictive power of the best CoMFA model was further validated by the accurate estimation to these compounds in the external test set, and the mean agreement of experimental and predicted $\log(IC_{50})$ values of the inhibitors is 0.6 log unit. Separate CoMFA models were conducted to evaluate the influence of different partial charges (Gasteiger−Marsili, Gasteiger−Hückel, MMFF94, ESP-AM1, and MPA-AM1) on the statistical quality of the models. The resulting CoMFA field map provides information on the geometry of the binding site cavity and the relative weights of various properties in different site pockets for each of the substrates considered. Moreover, in the current work, we applied MD simulations combined with MM/PBSA (Molecular mechanics/Possion-Boltzmann Surface Area) to determine the correct binding mode of the best inhibitor for which no ligand−protein crystal structure was present. To proceed, we define the following procedure: three hundred picosecond molecular dynamics simulations were first performed for the four binding modes suggested by DOCK 4.0 and manual docking, and then MM/PBSA was carried out for the collected snapshots. The most favorable binding mode identified by MM/PBSA has a binding free energy about 10 kcal/mol more favorable than the second best one. The most favorable binding mode identified by MM/PBSA can give satisfactory explanation of the SAR data of the studied molecules and is in good agreement with the contour maps of CoMFA. The most favorable binding mode suggests that with the quinazoline-based inhibitor, the N3 atom is hydrogen-bonded to a water molecule which, in turn, interacts with Thr 766, not Thr 830 as proposed by Wissner et al. (*J. Med. Chem.* **2000**, *43*, 3244). The predicted complex structure of quinazoline type inhibitor with EGF-R as well as the pharmacophore mapping from CoMFA can interpret the structure activities of the inhibitors well and afford us important information for structure-based drug design.

INTRODUCTION

The epidermal growth factor receptor (EGF-R) is a 170 000-dalton membrane glycoprotein composed of an extracellular EGF-binding domain (621 amino acids), a single transmembrane region (23 hydrophobic amino acids), and an intracellular domain (542 amino acids) that exhibits protein tyrosine kinase activity.[1] The binding of EGF to the receptor activates a cyclic nucleotide-independent, tyrosine-specific protein kinase, which phosphorylates various endogenous membrane proteins including the EGF receptor molecule itself and many other exogenous substrates.[2,3] The protein phosphorylation is a critical mechanism for regulating protein function in many cell regulatory processes, including T-cell and B-cell activation, mitogenesis, differentiation and development, angiogenesis, platelet activation, neurotransmitter signaling, cell cycle control, and growth control.[4] Therefore it is not surprising to see that functional perturba-

tions of protein-tyrosine phosphatases result in many diseases, including prostate cancer, head and neck cancer, ovarian cancer, nonsmall-cell lung cancer, and bladder cancer.[5−9] The overexpression or inappropriate expression of the EGF receptor or its ligands, EGF and transforming growth factor-α, can produce loss of growth control and the unregulated cell proliferation associated with malignancy.[10]

Small molecules which can selectively inhibit the EGF-R could therefore have great therapeutic potential in the treatment of malignant and nonmalignant epithelial diseases. Recently, a number of reports have shown that a broad class of 4-anilinoquinazolines are potent and highly selective inhibitors of EGF-R phosphorylation, resulting from competitive binding at the ATP site.[11−17] Several pharmaceutical firms and research groups have established programs based on these inhibitors, and now at least three such compounds have entered or will be entering clinical trials.[18−20] Several essential chemical groups for binding to EGF-R have been described.[11−17] The pyrimidine ring is mandatory, and a free NH linker is clearly optimal. Electron-withdrawing lipophilic

* Corresponding author phone: 8610-62757456; e-mail: xiaojxu@chem.pku.edu.cn.

substituents on the 3-position of the aniline are favorable, with chlorine and bromine optimal, and electron-donating groups at the 6- and 7-postion of the quinazoline are preferred.[13] To our knowledge, the detailed quantitative structure−activity relationship of these compounds with respect to kinase activity at the 3D-level has never been reported. So the first aim of the present study was to obtain a 3D-QSAR model based on six groups of compounds reported and to take insight into the main intermolecular interactions between the structure and inhibitory potency of the EGF-R inhibitors. To reach our research objectives two kinds of 3D-QSAR techniques have been used, including comparative molecular field analysis (CoMFA) and comparative molecular similarity indices analysis (CoMSIA).

The comparative molecular field analysis (CoMFA) is a well-documented and validated technique for the study of quantitative structure−activity relationship (QSAR) at the 3-D level.[21,22] The basic assumption for CoMFA is that the observed biological properties can be well-understood or -correlated with the suitable samplings of the steric and electrostatic fields surrounding a set of ligands. Recently, another 3D-QSAR procedure, comparative molecular similarity indices analysis (CoMSIA), has been reported.[23] In CoMSIA, a distance-dependent Gaussian-type functional form has been introduced, which can avoid singularities at the atomic positions and the dramatic changes of potential energy for those grids in the proximity of the surface. Meantime, no arbitrary definition of cutoff limits is required in CoMSIA. The unique differences between conventional CoMFA and CoMSIA are the field type and the potential function. In CoMSIA, similarity is expressed in terms of different physicochemical properties: steric occupancy, partial atomic charges, local hydrophobicity, and H-bond donor and acceptor properties. Moreover, in CoMSIA, a Gaussian-type distance-dependent function has been used to calculate different kinds of physicochemical properties.

However, CoMFA or CoMSIA is unable to appropriately describe all binding forces, being based only on standard steric and electrostatic molecular fields to model receptor−ligand interactions. The direct and complete representations of the binding forces between ligand and receptor can only be fully described by the complex structures. Because the 3D crystal structure of EGF-R is unavailable, it is relatively difficult to get a binding mode for the inhibitors in complex with EGF-R. In previous publications, the binding modes of the quinazoline type inhibitors in the ATP binding site of EGF-R have been reported by several groups.[12,16,17] It is interesting to find that these reported binding modes are quite different. In the Traxler's work, from SAR studies, a binding mode for 4-(phenylamino)-7*H*-pyrrolo[2,3-*d*]pyrimidines as well as for the structurally related 4-(phenylamino) quinazolines at the ATP-binding site of the EGF-R tyrosine kinase was proposed.[12] In the proposed binding mode, the pyrrole NH(7) and the N(1) of the pyrimidine ring form a similar bidentate hydrogen bond donor−acceptor system as ATP. In Palmer's work, the authors proposed another structure of quinazoline-based inhibitors complexed with EGF-R.[16] In this model, the pyrrole ring in pyrroloquinazolines occupies the entrance of the ATP binding pocket of the enzyme, with the pyrrole nitrogen located at the bottom of the cleft and the pyrrole C-3 position pointing toward a pocket corresponding to the ribose binding site of ATP. The N1 of the quinazoline

ring forms a hydrogen bond with the NH backbone of Met 769. The N3 of the quinazoline ring forms another hydrogen bond with the side chain of Thr 766. A recent paper by Wissner et al. reported a new binding mode of a quinazoline-type inhibitor complexed with EGF-R.[17] The new mode suggests that to quinazoline-based inhibitors, the N1 atom of the quinazoline forms a hydrogen bond with the backbone NH of Met 769 and the N3 atom forms a hydrogen bond to a water molecule. This water molecule, in turn, interacts with the hydroxyl group of Thr 830. However, until now, the correct binding mode of quinazolines in EGF-R has not been really clear. Because in the previous molecular modeling concerned with the interactions between EGF-R and quinazolines, the calculations were usually based on manual molecular docking and simple molecular minimizations, which could not give the dynamic illustration of the binding of quinazolines in EGF-R. So the other important aim of us is to recognize the correct binding mode of the quinazoline inhibitors in the ATP binding site of EGF-R. Here, we selected the most active quinazoline-type inhibitor as the studied case, and we proposed several inhibitor/EGF-R complexes as the possible binding models. Then, the binding free energies between EGF-R and the inhibitor were computed by using the MM/PBSA approach. Molecular dynamics (MD) simulations were used to generate the thermally averaged ensemble of conformations of the ligand−protein complexes. Based on the trajectories from MD simulations, the free energies of binding were calculated using molecular mechanics, continuum solvent model, surface area estimation, and normal model analysis. We expect that the best binding mode predicted by MM/PBSA can successfully suggest the structure−activity relationship, and give us important structural information for drug design.

Recently, MM/PBSA based on molecular dynamics (MD) sampling has been drawn much more attention. The MM/PBSA technique was proposed for evaluating solvation and binding free energies of macromolecules and their complexes.[24] In the method, the average total free energy of the system, G, is evaluated as

$$G = E_{MM} + G_{PB} + G_{NP} - TS \qquad (1)$$

where G_{PB} is the polar solvation energy in continuum solvent, usually computed using a finite-difference Poisson−Boltzmann (PB) model,[25,26] and G_{NP} is the nonpolar solvation energy, which is often obtained from the solvent-accessible surface area (SA). E_{MM} denotes the sum of molecular mechanical (MM) energies of the molecules from internal, electrostatic, and van der Waals energies. The last term in eq 1 is the solute entropy and can be estimated by a combination of classical statistical formulas and normal-mode analysis.[27] According to eq 2, the binding free enery of a noncovalent associating for a protein−ligand system can be computed as

$$\Delta G_{bind} = G_{complex} - G_{protein} - G_{ligand} = \\ \Delta E_{MM} + \Delta G_{PB} + \Delta G_{NP} - T\Delta S \quad (2)$$

The thermally averaged energy terms in eq 2 are obtained from the MD sampling. The ensemble of structures for the uncomplexed protein or ligand are directly generated by
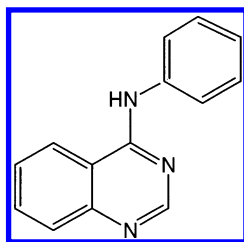
Quinazoline Type EGF-R Inhibitors

J. Chem. Inf. Comput. Sci., Vol. 43, No. 1, 2003 **275**

Table 1. EGFR Tyrosine Kinase Inhibition Data for 4-Anilinoquinazolines

A 1-58   A 59   A 60   A 61-69

no.	R_1	R_2	X	log(1/C) obsd	log(1/C) calc	residue	no.	R_1	R_2	X	log(1/C) obsd	log(1/C) calc	residue
A1	H	H	H	−2.54	−2.55	0.01	A35	NHMe	H	Br	0.60	−1.09	0.49
A2	H	H	F	−1.75	−2.08	0.33	A36	NMe$_2$	H	Br	−1.92	−1.01	−0.91
A3	H	H	Cl	−1.36	−1.81	0.45	A37	NHCO$_2$Me	H	Br	−1.08	−0.56	−0.52
A4*	H	H	Br	−1.43	−1.55	0.12	A38	H	OH	Br	−0.67	−0.46	−0.21
A5	H	H	I	−1.90	−1.78	−0.12	A39	H	NHAc	Br	−1.60	−1.29	−0.31
A6	H	H	CF$_3$	−2.76	−2.55	−0.21	A40	H	NHMe	Br	−0.85	−0.74	−0.11
A7	OMe	H	H	−1.74	−2.50	0.76	A41	H	NHEt	Br	−1.08	−1.26	0.18
A8	OMe	H	Br	−1.48	−1.04	−0.44	A42	H	NMe$_2$	Br	−1.04	−1.51	0.47
A9	NH$_2$	H	H	−2.89	−2.47	−0.42	A43	NH$_2$	NH$_2$	Br	0.92	1.14	−0.22
A10	NH$_2$	H	CF$_3$	−2.76	−1.94	−0.82	A44	NH$_2$	NHMe	Br	0.16	0.48	−0.32
A11	NH$_2$	H	Br	0.11	−0.93	1.04	A45	NH$_2$	NMe$_2$	Br	−2.20	−2.03	−0.17
A12*	NO$_2$	H	H	−3.70	−4.37	0.67	A46	NH$_2$	OMe	Br	−0.58	−0.20	−0.38
A13	NO$_2$	H	Br	−2.95	−2.89	−0.06	A47	NH$_2$	Cl	Br	−0.81	−1.11	0.30
A14	NO$_2$	H	CF$_3$	−4.00	−3.89	−0.11	A48	NO$_2$	NH$_2$	Br	−1.72	−1.76	0.04
A15	H	MeO	H	−2.08	−2.31	0.23	A49*	NO$_2$	NHMe	Br	−1.83	−1.10	−0.73
A16	H	MeO	Br	−1.00	−0.78	−0.22	A50	NO$_2$	NMe$_2$	Br	−3.30	−3.31	0.01
A17	H	NH$_2$	H	−2.00	−1.37	−0.63	A51	NO$_2$	NHAc	Br	−1.45	−1.59	0.14
A18	H	NH$_2$	F	−0.30	−0.38	0.08	A52	NO$_2$	OMe	Br	−1.18	−1.26	0.08
A19	H	NH$_2$	Cl	−0.60	−0.13	−0.47	A53	NO$_2$	Cl	Br	−1.40	−1.38	−0.02
A20	H	NH$_2$	Br	1.00	0.14	0.86	A54	OCH$_2$O		Br	−1.18	−1.18	0.00
A21*	H	NH$_2$	I	0.46	0.05	0.41	A55	OH	OH	Br	0.77	0.71	0.06
A22	H	NH$_2$	CF$_3$	−0.52	−0.89	0.37	A56	OEt	OEt	Br	2.22	1.94	0.26
A23	H	NO$_2$	H	−4.08	−4.30	0.22	A57	OPr	OPr	Br	0.77	0.82	−0.05
A24	H	NO$_2$	F	−3.79	−3.53	−0.26	A58*	OBu	OBu	Br	−2.02	−0.09	−1.93
A25	H	NO$_2$	Cl	−2.91	−3.19	0.28	A59		5,6-diOMe		−3.14	−3.23	0.09
A26	H	NO$_2$	Br	−3.00	−2.92	−0.08	A60		7,8-diOMe		−4.00	−4.01	0.01
A27	H	NO$_2$	I	−2.73	−3.13	0.40	A62	2-NH2		3′-Br	−2.67	−2.78	0.11
A28	H	NO$_2$	CF$_3$	−4.00	−3.94	−0.06	A64	5-OMe		3′-Br	0.17	0.05	0.12
A29	OMe	OMe	H	−1.46	−0.71	−0.75	A65	8-OMe		3′-Br	−4.00	−3.86	−0.14
A30*	OMe	OMe	F	−0.58	0.33	−0.91	A66	H		2′-Br	−2.11	−1.54	−0.57
A31	OMe	OMe	Cl	0.51	0.61	−0.10	A67	H		4′-Br	0.02	0.10	−0.08
A32	OMe	OMe	Br	1.60	0.87	0.73	A68	H		3′,4′-Br	1.14	0.50	0.64
A33	OMe	OMe	I	0.05	0.64	−0.59	A69*	H		3′,5′-diBr	−2.05	0.20	−2.25
A34	OMe	OMe	CF$_3$	0.62	−0.06	0.68							

using the trajectory of the complex and simply removing the atoms of ligand or protein, respectively. By now, MM/PBSA has been successfully applied to binding free energy calculations for several systems.[28−33] In a recent paper of Wang's et al., the author applied the MM/PBSA technique to determine the binding mode between HIV-1 RT and efavirenz (also known as SUSTIVATM, for which the crystal structure has not been published).[32] The authors calculated the binding free energies between HIV-1 RT and efavirenz for the five possible binding modes and successfully recognized the correct one, for which the MM-PBSA energy is about 5 kcal/mol favorable than the second best binding mode. The final structure of the MD simulations is very close to a measured resolution 2.3-angstrom crystal structure solved by the Dupont Pharmaceuticals recently, and the rmsd of the ligand and its surroundings (about 50 residues) is 1.0 angstrom. Encouraged by Wang's results, we decided to apply MM/PBSA to determine the binding mode of the quinazoline-type inhibitor with EGF-R. We expect that the correct binding mode can be recognized by the computed binding free energies.

MATERIAL AND METHODS

Data Set. Since a small molecule called PD 153035 (4-(3-bromoanilino)-6,7-dimethoxyquinazoline) was reported by David W. Fry et al. in 1994, various modifications on almost all the possible groups of this lead compound have been performed.[11−17] To develop reliable CoMFA and CoMSIA models, 135 quinazoline derivatives with diverse structures and distinct activities were selected from four literatures of the same research group, and therefore the data representing activities possess comparability.[13−16] The biological activity of each compound was expressed as the value of IC$_{50}$, which was the concentration of drugs to inhibit the phosphorylation of a 14-residue fragment of phospholipase Cγ1 by EGF-R, and −log(IC$_{50}$) was used for the 3D-QSAR analysis. In case of molecules where IC$_{50}$ values were not available as absolute values, the highest concentration tested was used as IC$_{50}$ values. 122 analogues listed in Tables 1, 2, 3 and 4 were randomly selected as a training set, and the rest acted as a test set (labeled with asterisks in Table 1). Selection of training set and test set molecules was made by considering

Figure 1. The core structure used to perform molecular alignment.

the fact that test set molecules represent a range of biological activity and chemical classes similar to that of a training set. Thus the test set is the true representation of the training set.

CoMFA and CoMSIA Studies. 1. Molecular Modeling. The molecular geometries of all compounds in Table 1 were modeled using the SYBYL 6.5.[34] The initial structures were minimized using molecular mechanism with MMFF94 force field.[35] For the molecules with flexible side chains, conformation analyses were preformed using the systematic search procedure.

2. Calculations of the Atomic Charges. Five different kinds of partial atomic charges were considered: (1) Gasteiger–Marsili charges,[36–38] (2) Gasteiger–Hückel charge,[39] (3) MMFF94 charges,[35] (4) the electrostatic potential fit charges at the AM1 level (ESP-AM1),[40] and (5) the Mulliken population analysis atomic charges at the AM1 level (MPA-AM1).[40] Both of the Gasteiger–Marsili and the Gasteiger–Hückel methods calculate atomic charges based on information about the atoms and their connectivity within the molecule. The MMFF94 atomic charges are simply computed based on the bond increment parameters in MMFF94 force field. The calculations of Gasteiger–Marsili, Gasteiger–Hückel, and MMFF94 charges in SYBYL were automated with SYBYL programming language (SPL) scripts. The calculations of MPA-AM1 and MPA-AM1 charges were calculated using MOPAC 6.0.

3. Molecular Alignment. Superimposition of the molecules was carried by a combination technique of atom-by-atom fit and field fit.

Alignment 1. Considering the molecules studied here bearing large conjugated systems, all compounds besides **B1** to **B13** and **B20** to **B34** were superimposed by an atom-by-atom least-squares fit as implemented in the SYBYL FIT option with the unsubstituented template shown in Figure 1 and compound A56 with the best biological activity as the reference molecule.

Alignment 2. For the compounds **B1** to **B13** and **B20** to **B34**, the linking groups between the quinazoline chromophore and the phenyl 4-side chain were replaced by the other groups rather than NH, so the molecular aggregates

for these compounds were produced by sterically and electrostatically aligning the molecules using the Align module in Cerius2,[41] and the compound **A56** was also treated as the reference molecule. The field alignment fixed the position of the reference molecule and moved the second molecule to a random starting position with respect to the first molecule. Optimal electrostatic and steric complementarities between the molecules were achieved by the implementation of a simple two-component force-field algorithm. In order for this procedure to be carried out, partial charges were required for each molecule. The partial charges from electrostatic fits by MOPAC calculations were assigned, and van der Waals parameters available in CVFF force field were used to estimate the steric field.[42]

Figure 2 shows the stereoview of the aligned molecules (including the test set) within the grid box used to generate the CoMFA or CoMSIA columns.

4. CoMFA Setup. The overlapped molecules were surrounded by a 3D grid of points in the three dimensions extending at least 4 Å beyond the union volume occupied by the superimposed molecules. The default sp^3 carbon atom with $+1|e|$ charge was selected as the probe atom for the calculations of the steric and electrostatic fields around the aligned molecules. Values of steric and electrostatic energies were truncated at 30 kcal/mol. In the CoMFA study, besides the usually used steric and electrostatic fields, the H-bonding fields were also included. The scheme proposed by Bohacek and McMartin was applied to generate the H-bonding fields.[43] In CoMFA, paired hydrogen bond acceptor and donor fields were created. Lattice points were assigned an energy of 0 if they were not near H-bond acceptor (or donor) atoms, or if H-bonding interactions were forbidden by steric congestion. Lattice points in sterically allowed regions close to acceptor (or donor) atoms were assigned a nominal energy equal to the designated steric cutoff values. Lattice points could be H-bond accepting only if they were not H-bond donating. For the "Advanced CoMFA" module in SYBYL 6.6, once an H-bond CoMFA column has been created, the acceptor component is nominally a "steric" field type and the donor component is nominally an "electrostatic" field type.

To choose the optimum number of components (ONC) and check the statistical significance of the models, leave-one-out (LOO) cross-validation technique was used by the enhanced version of PLS, the SAMPLES method.[44] Subsequently, the final 3D-QSAR models were obtained using ONC without cross-validation. Column filtering (minimum σ) was used at the default value of 2 kcal/mol in the cross-validation part.

5. CoMSIA Setup. The CoMSIA study was carried out in the "QSAR" module in SYBYL with five kinds of
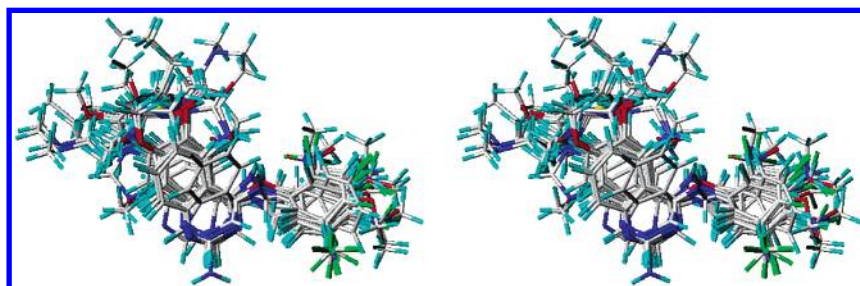

Figure 2. Stereoview of all aligned compounds in training set and test set.

physicochemical properties in CoMSIA, which include steric contributions by the third power of the atomic radii, electrostatics by partial charges, hydrophobicities by atom-based hydrophobic parameters, and H-bond properties by suitably placed pseudoatoms, using a common probe with 1 Å radius, +1 charge, +1 hydrophobicity, and H-bond property of +1. The extent and orientation of the grids surrounding the tested molecules were the same as those in the CoMFA study. The attenuation factor, α, which is the coefficient of the squared mutual distance in the Gaussian-type function in the calculation of similarity indices, was set to 0.3. The statistical evaluation for the CoMSIA analyses was performed in the same manner as described for CoMFA.

Homology Modeling of EGF-R. The sequence alignment between EGF-R and cAPK was performed according to the work reported by Singh et al.[45] The homology modeling was carried out by using the homology modeling technique. The coordinates of the sequence conserved regions came directly from the template protein. The coordinates for loops were assigned by searching the Brookhaven Protein Data Bank (PDB). In these calculations, a standard rotamer library was used, and a cutoff distance of 10 Å was applied for treatment of nonbonded interaction. The regions of insertion and deletion were minimized while holding the remainder of the structure fixed, thus perserving the overall structures. The entire structure was then subjected to conjugate gradient energy minimization for 500 cycles to convergence in order to remove any steric contacts. Energy minimization calculations were carried out using the DISCOVER program in INSIGHTII, the prediction of the EGF-R model was preformed using the Homology program in INSIGHTII.[46]

Then the ATP was inserted into the EGF-R ATP binding site. The construction of the complex structure of EGF-R with ATP was divided into two stages. First, the structural alignment was used to superimpose EGF-R and cAPK together. Then, ATP was extracted from cAPK and merged into EGF-R. During this model-building process, the structure of EGF-R was altered manually (especially torsional angles of side chains) to avoid unacceptable atom bumps. Then molecular minimization and molecular dynamics in AMBER 6.0 were performed to relax the obtained EGF-R/ATP complex.[47] We derived atom partial charges of ATP by semiempirical AM1 geometry optimization and subsequent single-point Hartree−Fock (HF)/6-31G* calculations of the electrostatic potential, to which the charges were fitted using the RESP fitting technique.[48] Then the EGF-R/ATP complex was solvated with a 20 Å sphere of TIP3P water molecules, and MD simulations were subsequently performed for a total of 300 ps at $T = 300$ K. The resulting structure was fully minimized.

Construction of the Initial Models for MD Simulations. Considering the crystal structure of EGF-R complexed with **A56** never been reported, it is a challenge for us to construct the initial structures for MD simulations. For **A56**, the binding orientation was first estimated with docking program DOCK 4.0.[49,50] The obtained EGF-R/ATP complex was used to generate the receptor site and the energetic grid for the following docking calculations. The docking procedure was applied as the following ansatz: the sphere centers (a set of overlapping spheres) of the ATP binding site were identified by *sphgen* program; then a box was created to enclose the spheres to be used for docking, and the energetic grid was

created by *grid* program to minimize the overall computational time; finally, flexible docking was performed to determine the possible binding orientation of **A56** in the active site.

Moreover, the other three kinds of possible binding orientations of **A56** were proposed by manual docking based on the prior publications.[12,16−17] We simply removed the ATP molecule from the minimized EGF-R/ATP complex and replaced it by the molecule **A56**. The initial orientations of **A56** were determined according to the structures proposed in the prior publications.[12,16−17]

MM/PBSA Calculations. All MD simulations were carried out at 300 K with the ligand bound to the protein with a cap of waters around the complex filled up to 20 Å from the center of mass of the ligand. The explicit solvent model TIP3P water was used. The SHAKE procedure was employed to constrain all bonds involving at least one hydrogen atom. The time step of the simulations was 2.0 fs with a cutoff 12 Å for the nonbonded interactions. The nonbonded pairs were updated every 30 steps. Prior to the MD simulations, the system was minimized with harmonic position constraints for all protein heavy atoms. The constrains are 5000, 1000, 100, and 10 kcal/mol/Å². Subsequently, a cycle of minimizaration was done to relax all the atoms without constraints. The maximum minimization steps were 10 000, and the convergence criterion for energy gradient was 0.5 kcal/mol/Å². All residues within 16 Å were allowed to move, while the other atoms were restrained by a 50 kcal/mol/Å² harmonic force. MD simulation procedures for the protein−ligand systems involved (1) 150 ps of MD simulations for equilibrium and (2) 150 ps of MD simulations for data collection. In the data collection stage, every 500 fs, the snapshot was recorded in the trajectory file.

The energy terms in eq 2 were calculated separately. The gas-phase molecular mechanical (MM) energy was averaged over all the snapshots. All MM calculations were preformed using the *anal* module in AMBER with a nonbonded cutoff of 99 Å and a dielectric constant of 1 in the absence of any solvent.

The electrostatic contribution to the solvation energy, ΔG_{PB}, was calculated using the DELPHI,[51,52] which solves the Poisson−Boltzmann equation numerically and calculates the electrostatic energy according to the electrostatic potential. The grid size was defined as 0.5 Å. The radius of the probe molecule was set to 1.4 Å. The partial charges used in PB calculations were taken from the AMBER parameter set (protein) and the ESP fittings (ligand). The radii of atoms were taken from the PARSE parameter set.[53] The solvent accessible surface (*SAS*) were calculated using the MSMS program.[54] The nonpolar contribution to the desolvation free energy was calculated as $0.00542 \times SAS + 0.92$ kcal/mol.

The conformational entropy contribution (translation, rotation, and vibration) to the binding free energy was evaluated using normal-mode analysis. The normal-mode analysis was carried out using the *nmode* module in AMBER 6.0.[55] We must point out that the *nmode* calculation was a N^3 problem, where N is the number of atoms in question, so it is very time-consuming. To simplify the calculations, the residues within an 8 Å sphere around the ligand were cut out from an MD snapshot for each ligand−protein complex. The open valences were saturated by adding hydrogen atoms in SYBYL. The corresponding uncomplexed reactants were

Table 2. EGFR Tyrosine Kinase Inhibition Data for 4-Anilino- and 4- (Benzylamino)quinazolines and Related Heterocycle

a

b

B-14

B-15

B-16

B-17

B-19
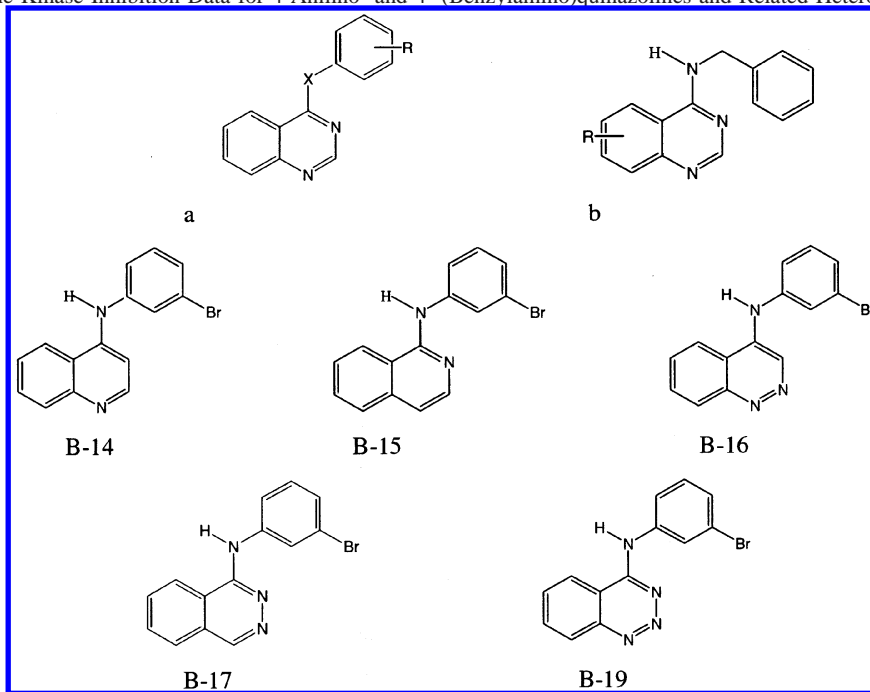
no.	type	R	X	log(1/C) obsd	log(1/C) calc	residue	no.	type	R	X	log(1/C) obsd	log(1/C) calc	residue
B1	a	H	$NH(CH_2)_2$	−3.61	−3.35	−0.26	B16				−4.00	−3.95	−0.05
B2	a	H	$NHCH_2$	−2.51	−3.00	0.49	B17				−5.00	−5.30	0.30
B3	a	4-Cl	$NHCH_2$	−3.85	−3.02	−0.83	B19				−3.00	−3.08	0.08
B4	A	4-OMe	$NHCH_2$	−4.00	−4.06	0.06	B20	b	5-NO_2		−3.90	−3.77	−0.13
B6	A	H	Nme	−5.00	−5.01	0.01	B21	b	5-NH_2		−4.00	−3.55	−0.45
B7	A	3-OMe	NH	−2.93	−2.54	−0.39	B22	b	5-OMe		−3.69	−3.28	−0.41
B8*	A	3-Me	NH	−2.96	−2.26	−0.70	B23	b	6-NO_2		−4.00	−4.22	0.22
B9	a	3-Cl	NH	−1.36	−1.55	0.19	B24	b	6-NH_2		−3.15	−3.46	0.31
B10	a	3-Br	NH	−1.43	−1.59	0.16	B25*	b	6-OME		−2.30	−2.73	0.43
B11	a	3-I	NH	−1.90	−1.68	−0.22	B26	b	7-NO_2		−3.77	−3.97	0.20
B12	a	3-CF_3	NH	−2.76	−2.50	−0.26	B28	b	7-OME		−1.76	−1.92	0.16
B13	a	3-Br	O	−2.88	−3.12	0.24	B32	b	6-OMe,7-OH		−2.77	−2.69	−0.08
B14				−3.74	−3.58	−0.16	B33	b	6-OH,7-OMe		−1.75	−1.82	0.07
B15*				−5.00	−4.67	−0.33	B34	b	6,7-$(OMe)_2$		−1.00	−1.71	0.71

generated by removing the atoms of the protein and atoms from the reduced complex structure. Then, each of the structures was fully minimized using conjugate gradient strategy for 50 000 steps in the presence of a distance-dependent dielectric ($\epsilon = 4r_{ij}$) using the *sander* module in AMBER. Consequently, the structures were further minimized using newton-Raphson technique for 500 steps using the *nmode* module in AMBER. Last, the entropy for each structure was calculated using classical statistical formulas and normal-mode analysis. To reduce the statistical error, for each structure, totally five snapshots at 30 ps, 60 ps, 90 ps, 120 ps, and 150 ps were extracted and analyzed. The final entropy estimate was averaged over the five snapshots.

RESULTS AND DISCUSSIONS

CoMFA Models. Often for QSARs developed with CoMFA a shift in the q^2 values is observed as the grid spacing is altered.[55−58] To examine this possibility with these data, the different grid boxes with 1.0, 1.5, 2.0, 2.5, and 3.0 Å grid spacing, respectively, were used for the CoMFA calculations. The influence of the different grid spacing to CoMFA model is obvious. Only from the q^2 after leave-one-out cross-validation, the model with the grid spacing of

2.0 Å was selected as the best model. The explanation for the variation of q^2 with the grid spacing is straightforward. A lower grid spacing (1.0 or 1.5 Å) may generate more noise in the PLS calculations and require a greater computational effort. While the grid spacing is defined as a larger value (2.5 or 3.0 Å), some important information about field properties in some regions may be lost. The following discussion concerned with CoMFA will only refer to the models generated from 2.0 Å grid spacing.

The CoMFA models using different set of partial charges are shown in Table 6. From the statistical quality, all of them exhibit good statistical quality between the predicted and experimentally determined values of $\log IC_{50}$: model 1 ($r^2 = 0.93$; $q^2 = 0.52$), model 2 ($r^2 = 0.92$; $q^2 = 0.53$), model 3 ($r^2 = 0.87$; $q^2 = 0.54$), model 4 ($r^2 = 0.88$; $q^2 = 0.55$), and model 5 ($r^2 = 0.90$; $q^2 = 0.53$). It is obvious that the way in which charges are calculated does not greatly affect the results. Considering the internal predictive ability of model 4 are slightly higher than those of the other models, and Gasteiger−Marsili partial charges are used in the further studies.

CoMSIA Models. At 2.0 Å grid spacing, the CoMSIA models using different field combinations are shown in Table

Table 3. EGFR Tyrosine Kinase Inhibition Data for Tricyclic Anilinoquinazoline Derivatives
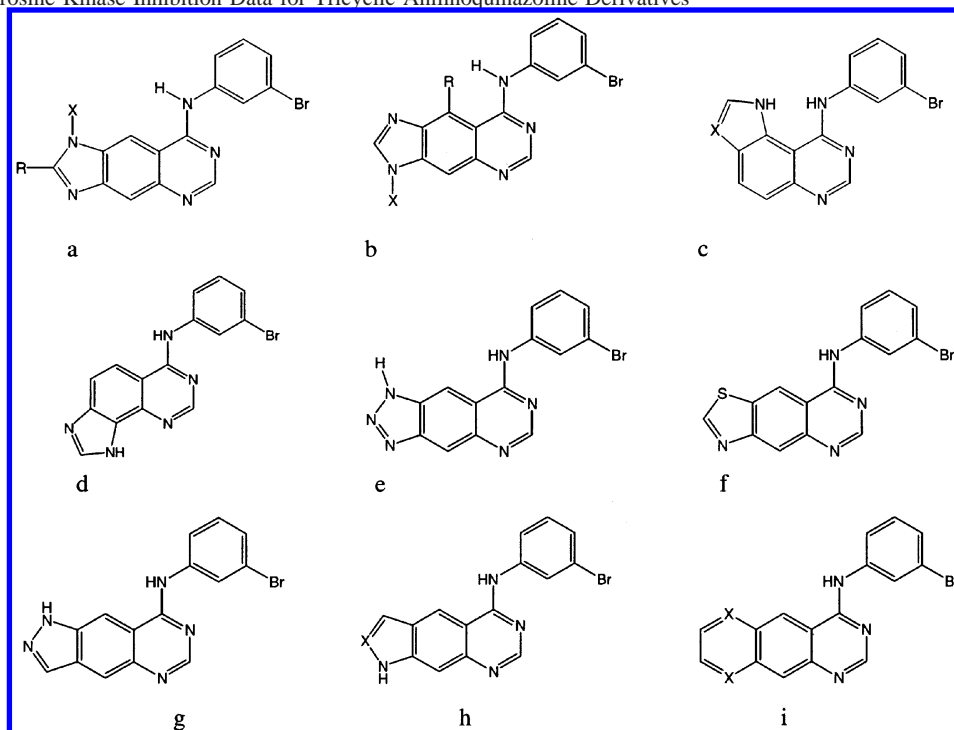
no.	type	X	R	log(1/C) obsd	log(1/C) calc	residue	no.	type	X	R	log(1/C) obsd	log(1/C) calc	residue
C1	a	H	H	2.10	2.04	0.06	C9	d			−2.43	−2.13	−0.30
C2*	a	H	Me	0.54	2.12	−1.58	C10	e			−0.61	−0.74	0.13
C3	a	Me	H	2.00	1.82	0.18	C11	f			−1.64	−1.43	−0.21
C4	a	$(CH_2)_2NMe_2$	H	−0.12	−0.32	0.20	C12	g			0.47	0.90	−0.43
C5	b	Me	H	1.60	1.09	0.51	C13	h	N		0.36	0.04	0.32
C6	b	$(CH_2)_2NMe_2$	H	−1.34	−0.77	−0.57	C14	h	CH		0.36	0.33	0.03
C7	b	$(CH_2)_2NMe_2$	Cl	−2.31	−2.64	0.33	C15	c	CH		−0.09	−0.09	0.00
C8	c	N		−1.46	−0.91	−0.55	C17	i	N		−0.23	−0.22	−0.01
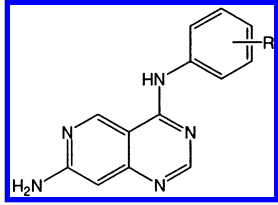
7. It can be seen that using steric and electrostatic fields only, the quality of the CoMSIA model ($q^2 = 0.40$) was poorer than the CoMFA model ($q^2 = 0.55$). After considering the H-bonding fields, the q^2 value of the CoMSIA model does not show obvious variation. After adding the hydrophobic field, the predictive power of the 3D-QSAR model ($q^2 = 0.423$) increases slightly, indicating that the biological activity may exist some relationships with the hydrophobic field. The best 3D-QSAR was derived from using all five fields afforded by CoMSIA, which possesses the best predictive power ($q^2 = 0.43$).

Using all five fields, the sensitivity of the CoMSIA models to different grid spacings was also investigated. The difference of the q^2 values between the best and the worst CoMSIA models is about 0.05 unit, which is obviously smaller than that between the best and the worst CoMFA models. The instabilities of CoMFA can be attributed to the shape and steepness of the Lennard-Jones potential and in consequence to the required arbitrary fixation of cutoff values. For example, in some cases, some atoms of the aligned molecules are near some grid points, if we use the potentials in CoMFA, the Lennard-Jones potentials will increase to large values near these grids, and generally the large potentials will be truncated to the cutoff value. The contributions of those grids using cutoff values will introduce certain unpredicted instability to the final models.

The Predictive Ability of the Best CoMFA Model. To investigate the influence of the H-bonding field to the quality of the CoMFA models, H-bonding fields were introduced to the CoMFA study by using the "Advanced CoMFA" module in SYBYL. Following the addition of the H-bonding fields, the predictive power of the CoMFA model ($q^2 = 0.54$) was decreased a little.

Furthermore, the region focusing technique, available in the 'Advanced CoMFA' module in SYBYL, was used to refine the model by increasing the weights for those lattice points which were most pertinent to the model. After region focusing, the CoMFA model with the steric and electrostatic field only demonstrates minor improvement ($q^2 = 0.56$). While to the CoMFA model with the steric, electrostatic and H-bonding fields, the q^2 after regions focusing was greatly promoted ($q^2 = 0.60$). The interpretation of the significant improvement of the CoMFA model 3 in Table 9 is straightforward. The H-bonding interactions between ligand and receptor are usually located and oriented, which may be quite different from the common electrostatic interactions. So, only a small set of H-bonding fields on the grid points are meaningful, and the other H-bonding fields on the grid points will bring significant noise to the subsequent PLS analysis. After considering regions focusing, the weights for those meaningful grid points are increased, and the noise will be greatly decreased. After considering regions focusing,
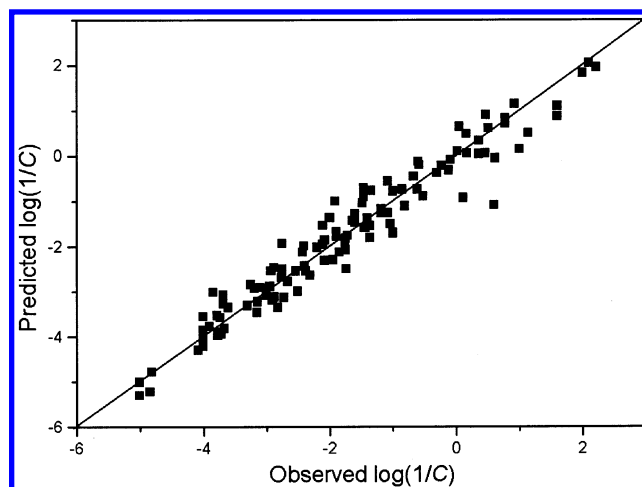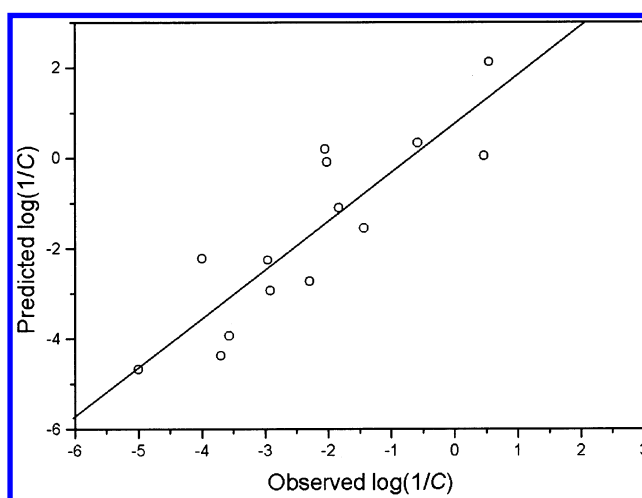
Table 4. EGFR Tyrosine Kinase Inhibition Data for 7-Aminopyrido[4,3-d]pyrimidine Analogues

no.	R	log(1/C) obsd	log(1/C) calc	residue
D1	H	−2.40	−2.44	0.04
D2	2′-NO$_2$	−3.10	−2.91	−0.19
D3	3′-NO$_2$	−1.60	−1.47	−0.13
D4	4′-NO$_2$	−4.81	−4.78	−0.03
D5	2′-Br	−2.38	−2.55	0.17
D6	3′-Br	−1.00	−1.69	0.69
D7	4′-Br	−1.95	−2.30	0.35
D8*	2′-CF$_3$	−4.00	−2.22	−1.78
D9	3′-CF$_3$	−1.18	−1.17	−0.01
D10	4′-CF$_3$	−3.67	−3.82	0.15
D11*	2′-OMe	−3.57	−3.93	0.36
D12	3′-OMe	−2.11	−1.96	−0.15
D13	4′-OMe	−2.83	−3.35	0.52
D14	2′-NH$_2$	−3.72	−3.94	0.22
D15	3′-NH$_2$	−3.19	−2.93	−0.26
D16	4′-NH$_2$	−3.77	−3.90	0.13
D17	2′-NMe$_2$	−4.84	−5.22	0.38
D18	3′-NMe$_2$	−3.25	−2.85	−0.40
D19	4′-NMe$_2$	−3.69	−3.08	−0.61
D20*	3′-F	−2.92	−2.94	0.02
D21	3′-Cl	−2.08	−1.86	−0.22
D22	3′-I	−2.41	−1.99	−0.42
D23	3′-OH	−1.85	−2.13	0.28
D24	3′-Me	−1.60	−1.42	−0.18

Figure 3. Comparison of experimental log(1/C) with calculated log(1/C) using the best CoMFA model.

Figure 4. Plot of the actual prediction of the best CoMFA model in Table 9.

the CoMFA model with the steric, electrostatic, and H-bonding fields was treated as the best CoMFA model. The leave-one-out cross-validated PLS analysis of the model resulted in the model with a q^2 of 0.60 using 12 principle components. The non-cross-validated PLS analysis yields the model with a higher r^2 of 0.93 with a very low standard error of estimate (*SD*) 0.46 (Table 9). Moreover, from the comparison of four 3D-QSAR models for the prediction on twelve molecules in the test sets, it can be found that the best CoMFA model also possesses good predictive ability ($r_{pred}^2 = 0.93$, $SSE = 0.93$) with the average absolute error of 0.60 log units across a range of 2.03 log units. The biological activities (log1/C), the calculated activities using the best CoMFA model, and the residue values from the observed values for the training set were shown in Tables 1, 2, 3 and 4 and Figure 3. The predicted biological activities of the test-set compounds were listed in Tables 1, 2, 3 and 4 and Figure 4. The derived model was satisfactory from the viewpoint of statistical significance and actual predictive ability.

CoMFA Contour Maps. The CoMFA steric, electrostatic, and H-bonding fields from the final non-cross-validated analysis were plotted as three-dimension colored contour maps in Figures 5−9. The field energies at each lattice point were calculated as the scalar results of the coefficient and the standard deviation associated with a particular column of the data table (stdev*coff), always plotted as the percentage of contribution to the CoMFA equation. These maps show regions where differences in molecular fields are associated with differences in biological activity. The maps do not show what is common to all molecules of a set, and hence one cannot expect to generate a complete image of the receptor.

Table 5. Influence of Different Grids on the CoMFA Models

	CoMFA(1)	CoMFA(2)	CoMFA(3)	CoMFA(4)	CoMFA(5)
q^2	0.51	0.51	0.52	0.40	0.38
r^2	0.97	0.93	0.92	0.58	0.63
standard error	0.29	0.45	0.45	1.10	1.02
F	260.18	173.02	170.22	80.56	67.66
n	15	9	9	2	3
grid spacing (Å)	1.0	1.5	2.0	2.5	3.0
fraction					
steric	0.39	0.41	0.42	0.37	0.45
electrostatic	0.61	0.59	0.58	0.64	0.55

QUINAZOLINE TYPE EGF-R INHIBITORS

J. Chem. Inf. Comput. Sci., Vol. 43, No. 1, 2003 **281**

Table 6. Influence of the Different Charges on the CoMFA Models[a]

	CoMFA(1)	CoMFA(2)	CoMFA(3)	CoMFA(4)	CoMFA(6)
charge model	MMFF94	ESP-AM1	Gast_Huck	Gasteiger	MPA-AM1
q^2	0.52	0.53	0.54	0.55	0.53
r^2	0.93	0.92	0.87	0.88	0.56
standard error	0.45	0.50	0.61	0.59	0.90
F	170.22	125.36	97.50	106.50	107.98
n	9	10	8	8	9
fraction					
steric	0.42	0.48	0.50	0.46	0.52
electrostatic	0.58	0.52	0.50	0.54	0.48

[a] Grid spacing: 2.0 Å.

Table 7. CoMSIA Models Using Different Field Combinations□Grid Spacing: 2.0 Å□

	CoMSIA(1)	CoMSIA(2)	CoMSIA(3)	CoMSIA(4)
	steric + electrostatic	steric + electrostatic + H-bonding	steric + electrostatic + hydrophobic	steric + electrostatic + H-bonding + hydrophobic
q^2	0.40	0.40	0.42	0.43
r^2	0.84	0.70	0.74	0.75
standard error	0.69	0.93	0.86	0.85
F	66.36	68.23	66.85	70.32
n	9	4	5	5
fraction				
steric	0.26	0.13	0.15	0.09
electrostatic	0.75	0.34	0.41	0.23
hydrophobic			0.44	0.26
H-bond acceptor		0.25		0.21
H-bond donor		0.28		0.21

Table 8. Influence of Different Grids on the CoMSIA Models

	CoMSIA(1)	CoMSIA(2)	CoMSIA(3)	CoMSIA(4)	CoMSIA(5)
q^2	0.41	0.41	0.43	0.45	0.39
r^2	0.75	0.75	0.75	0.75	0.71
standard error	0.86	0.86	0.85	0.86	0.92
F	68.37	68.49	70.32	67.85	70.47
n	5	5	5	5	4
grid spacing (Å)	1.0	1.5	2.0	2.5	3.0
fraction					
steric	0.09	0.09	0.09	0.09	0.10
electrostatic	0.24	0.24	0.23	0.25	0.29
hydrophobic	0.27	0.27	0.26	0.24	0.28
H-bond acceptor	0.20	0.20	0.21	0.21	0.19
H-bond donor	0.20	0.20	0.21	0.22	0.14

Table 9. CoMFA Models after Considering H-Bonding Field and Region Focusing

	CoMFA(1)	CoMFA(2)	CoMFA(3)
q^2	0.48	0.56	0.60
r^2	0.86	0.88	0.93
standard error	0.65	0.58	0.46
F	98.98	107.50	124.51
n	7	9	12
fraction			
steric	0.22	0.51	0.29
electrostatic	0.27	0.49	0.35
H-bond acceptor	0.30		0.23
H-bond donor	0.21		0.13

Figure 5. The contour plot of the CoMFA steric fields.

The steric contour map for the best CoMFA model depicts regions around the molecules where enhanced EGF-R binding affinity is associated with increasing (green) and decreasing (yellow) steric bulk (see Figures 5 and 6). The most active compound **A56** was treated as the reference molecule. The steric contours show that around the anilino group there exist four sterically unfavorable regions, in which three smaller ones are located at the 4′-, 5′-, and 6′-positions, and the other larger one is located near the 1′- and 2′-positions. These yellow regions indicate that additional steric interaction in these regions would lead to a decreased activity. On the 1′-, 2′-, 4′-, 5′-, and 6′-positions, only very small groups are preferred. Substitution of the aniline side chain with larger groups will greatly reduce activity, suggesting limited bulk tolerance. For example, if we constituent the Br or Cl atoms on the positions to the CF_3 group, the activities of the molecules are greatly reduced. Here, it should
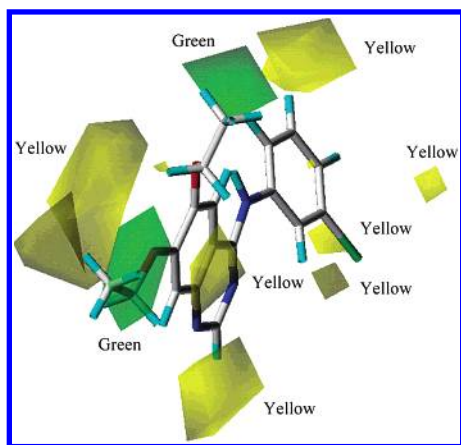
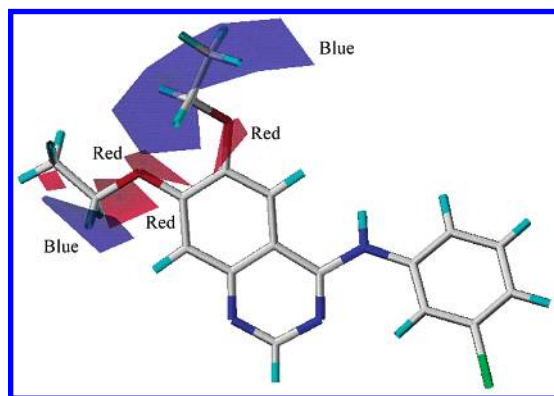**Figure 6.** The contour plot of the CoMFA steric fields.

Figure 7. The contour plot of the CoMFA electrostatic fields.
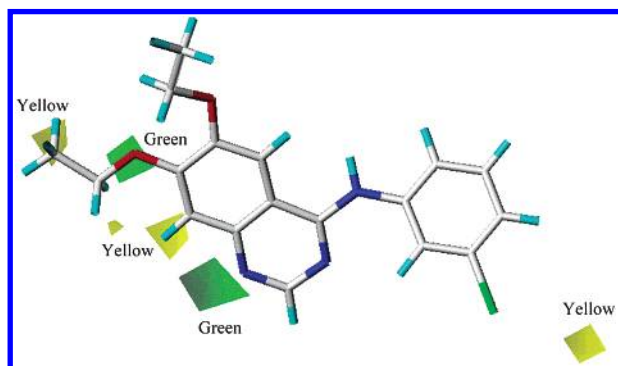

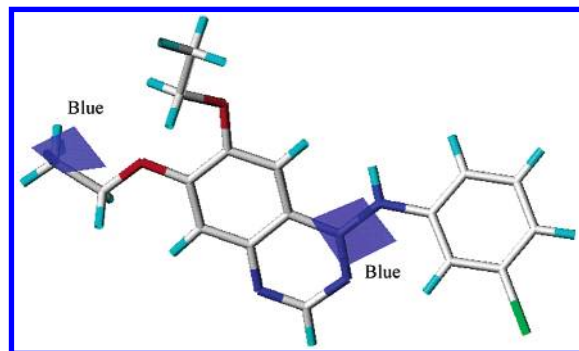**Figure 8.** The contour plot of the H-bonding acceptor fields.

Figure 9. The contour plot of the H-bonding donor fields.

be noted that the substitution of the hydrogen atom with the F, Cl, Br, or I atoms will greatly enhance the biological activities. The enhancement of the biological activities should be caused by the greater hydrophobic interactions induced by the substituented F, Cl, Br, or I atom. The above analyses

demonstrate that the analino group may be located in a hydrophobic pocket deep in the binding cleft, which may produce strong hydrophobic interaction with the binding pocket. Around the bicyclic ring system, there exist two large yellow regions and two large green regions. One yellow region is located near the 8-position in the quinazoline ring, and large groups linked to this position may produce bad atom bumps with EGF-R, for example, when substitution of two hydrogen atoms in compound **A4** with two OMe-groups to give compound **A60** inhibitory potency was decreased more than 100-fold. The other yellow region is situated close to the 6- and 7-positions in the quinazoline ring, while the two green regions are also situated close to the 6- and 7-positions. To have a deep insight of the contour maps around the quinazoline, **A56** in Figure 5 was rotated to another orientation (Figure 6). The two green regions and two yellow regions mean that near these positions steric bulk can be readily tolerated, but the orientation of the groups linked to the 6- and 7-positions in the quinazoline ring should be specific. For example, the large substituents at the 6- and 7-positions of the quinazoline and pyrido[d]pyrimidines and at the 3-position of the tricyclic pyrrolo- and pyrazolo-quinazolines can be tolerated without a major loss of affinity.

The electrostatic contour plot on the set of 122 compounds shows three red regions near the 6- and 7-positions on quinazoline and two blue regions nearby. The initial structure−activity relationships of Rewcastle et al. indicate that electron-donating groups at the 6- and 7-positions of quinazoline are preferred.[13] But from the contour map of electrostatic field, it seems that the requirement of the electrostatic features of the groups on these two sites is somewhat different. In many cases, the blue and red regions are usually located nearby. It is not very surprising, because a charge withdrawing or donating group will make the groups linked with it bear more positive or negative charges, which is usually expressed by the adjacent blue and red regions. The blue and red regions are usually neighboring, but the distributed range will tell us which kind of electrostatic feature is more important. There are two red regions near the groups at the 7-position of quinazoline. That is to say, the negative charges in these two regions are very important to ligand binding. These two red regions may explain why the introduction of N or O atoms with negative charges will greatly enhance the biological activity. For example, the substitution of the hydrogen atom at the 7-position in compound **A4** (IC_{50}: 27) with one amino group gives compound **A20** (IC_{50}: 0.1) increased inhibitory potency more than 270-fold. The substitution of the hydrogen atom at the 7-position in compound **A8** (IC_{50}: 30) with one methoxyl group gives compound **A32** (IC_{50}: 0.025) increased inhibitory potency more than 1200-fold. We think that the O or N atom with negative charges at the 7-position will produce strong electrostatic interaction or even stable hydrogen bonds with protein. Compared with the red regions near the 7-position, the red region near the 6-position was obviously smaller, but the blue region at the end of the substituted group at the 6-position was widely distributed. That is to say, a relatively large group with a positive charge in the blue region was favorable to enhance the affinity.

The graphical interpretation of the field contributions of the H-bonding is shown in Figures 8 and 9. In principle, they should highlight the areas near which H-bonding donor

QUINAZOLINE TYPE EGF-R INHIBITORS

J. Chem. Inf. Comput. Sci., Vol. 43, No. 1, 2003 **283**

Table 10. Binding Free Energies between EGFR and Inhibitor for Four Types of Binding Models (kcal/mol)[a]

	Model I	Model II	Model III	Mode IV
ΔE_{vdw}	−44.8 (2.3)	−42.5 (1.9)	−47.9 (2.1)	−45.4 (2.0)
ΔE_{ele}	−21.6 (3.2)	−4.7 (2.4)	−12.2 (4.9)	−22.0 (2.5)
ΔG_{PB}	48.7 (3.2)	29.7 (4.2)	42.3 (5.3)	39.5 (2.5)
ΔG_{SA}	−5.6 (0.1)	−5.2 (0.1)	−5.4 (0.2)	−5.1 (0.1)
$-T\Delta S$	17.9 (1.5)	16.8 (1.6)	17.4 (1.5)	16.2 (1.5)
$\Delta\Delta G_1$	−5.4	−5.9	−5.8	−16.8

[a] All the energies are in kcal/mol.

or acceptor on the ligand can form H-bonds with the receptor to influence binding affinity. From the fraction of field, the H-bonding acceptor field is a little more important than the other fields. The contour shows that more H-bonding acceptor groups in the green regions increases potency, whereas more H-bonding acceptor groups in the yellow regions decrease potency. In Figure 8, there are two distinct green regions: one near the O atom of the group at the 7-position and the other near the N1 atom in the quinazoline ring. The atoms near these two positions may act as H-bonding acceptor and produce H-bonding interactions with H-bonding donors in protein. Compared with H-bonding acceptor field, the H-bonding donor field seems less important. In Figure 9, the H-bonding donor favorable areas are represented by blue contours, while the H-bonding donor unfavorable areas are represented by red contours. In Figure 9, two H-bonding donor features can be observed, which are indicated by two blue contours. One blue region is located near the NH linker, that is to say, the NH linker may form the stable H-bonds with some residues in protein. The other blue region is located near the group at the 7-position of quinazoline, which is not direct to be interpreted.

Binding Mode Prediction of EGF-R/A56. Three-hundred picosecond MD simulations were performed for four possible binding modes. For all modes the systems were well equilibrated after 150 ps MD simulations, and the last 150 ps were selected for further processing analysis. After equilibration, the MD simulations are very stable, and the root-mean-square deviations of the main chain are smaller than 0.10 Å and those of the ligand are smaller than 0.35 Å.

The free energies of binding using MM/PBSA for four possible binding modes are shown in Table 10. The binding mode IV has the most favorable binding free energy, which is −16.8 kcal/mol, about 11 kcal/mol more negative than the second best binding mode (binding mode −5.9 kcal/mol). For the four binding modes, the van der Waals and the electrostatic energies are very different, which means that in different binding modes the energetic complementarity between ligand and protein is quite different. We conclude that in this case both the van der Waal and electrostatic terms are very important for determining the different binding orientations. Here, the electrostatic term includes not only ligand−protein interactions but also both ligand and protein desolvation upon binding. In the next several paragraphs, we will discuss the four binding modes in details.

1. Binding Mode I. For compound **A56**, both flexible docking and rigid docking give similar binding orientations. Figure 10 shows the averaged structure of MD simulations for the complex after molecular docking. In this figure, it can be seen that the inhibitor occupies the binding space of ATP. If we consider the structure−activity relationship in prior publications, this kind of binding mode can be easily eliminated. In previous work, it has been proven that substitution of the anilino side chain with larger groups greatly reduces activity, which suggests that the anilino group is located deep in the binding cleft and produces close contact with EGF-R.[12−16] In Figure 10, we find that the Br atom at the 5-position of the anilino moiety is oriented to the outside of the ATP binding pocket, which means that substitution of the Br atom with the other larger groups will not greatly affect the binding affinities of ligand. The predicted complexes based on molecular docking cannot give reasonable explanation to the reported relationship between the structures and the activities.

2. Binding Mode II. Figure 11 shows the initial structure proposed by Traxler et al.[12] and the conformer from averaging the MD trajectory. From Figure 11 it can be observed that the spatial position of the inhibitor greatly changes. After the least-squares fit for the C_α of protein, we found that the center of mass of the inhibitor moves about 5 Å. That is to say, the inhibitor cannot be stably located at
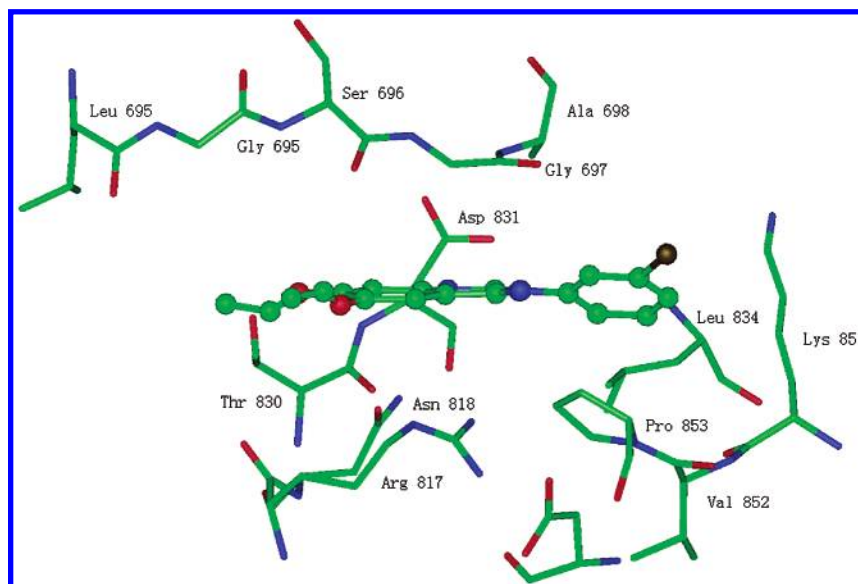
Figure 10. The averaged conformer of MD simulations from the complexed structure based on DOCK 4.0.
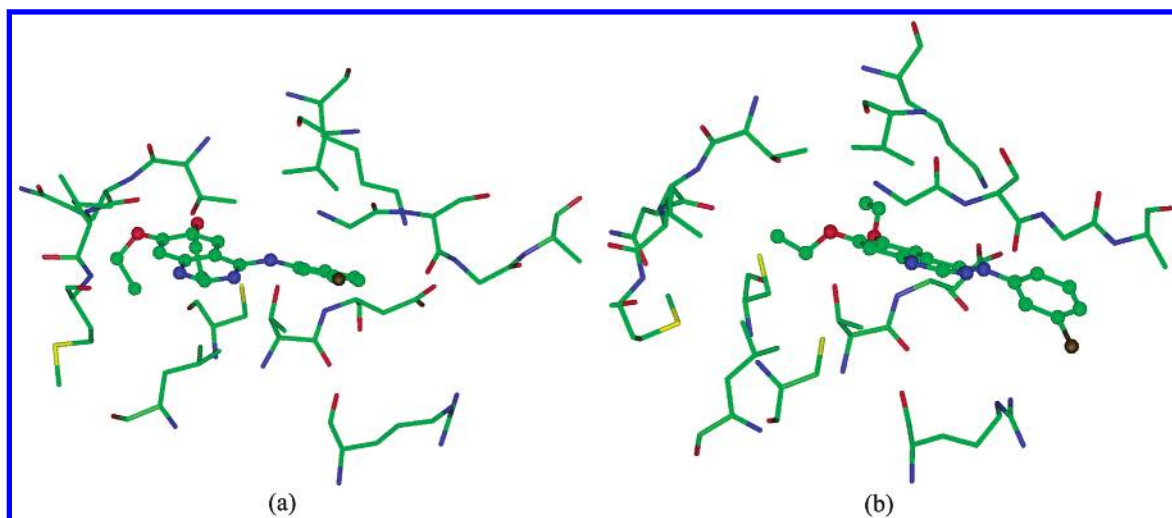
Figure 11. The complexed structure propsed by Traxler et al. (a) the initial structure and (b) the averaged structure of MD simulations.
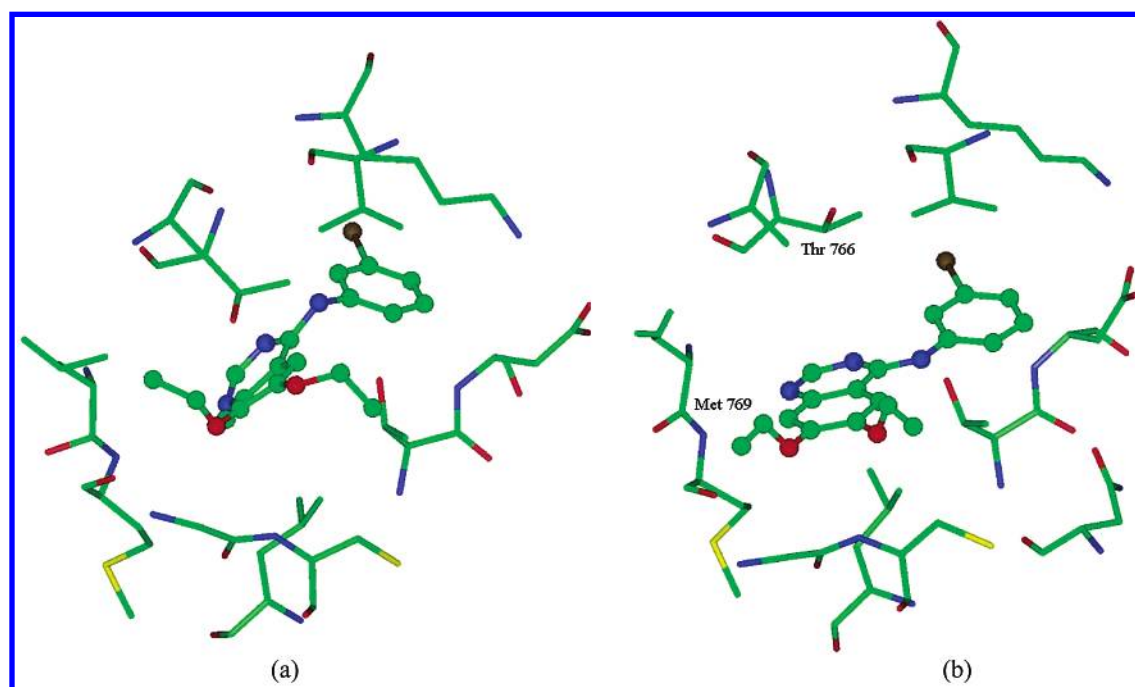

Figure 12. The complexed structure propsed by Palmer et al. (a) the initial structure and (b) the averaged strucuture of MD simulations.

the position proposed by Traxler et al. Moreover, the predicted complex shows that the Br atom linked to the anilino moiety is oriented outside the binding pocket, which is unreasonable.

   3. Binding Mode III. Figure 12 shows the structure from averaging the MD trajectory based on the binding mode proposed by Palmer et al.[16] In Palmer's model, the N-1 atom of the quinazoline ring forms a hydrogen bond with the NH backbone of Met-769. A second hydrogen bond can be formed with N-3 of the quinazoline ring and the side chain of Thr-766, which is located at the beginning of the extended coil stretch deep in the binding cleft. In the initial structure of binding mode III, the torsional angles of the Thr-766 and the position of the inhibitor were manually adjusted to ensure that these two hydrogen bonds were formed. After optimizations of molecular mechanics, it can be found that these two hydrogen bonds can be maintained. But after long-term MD simulations, it is interesting to find that the positions of the protein and inhibitor change obviously. Figure 12 shows that in the structure averaging the MD trajectory the N-1 of the

quinazoline ring can also form a stable hydrogen bond with the NH backbone of Met-769. But the distance between N-3 of the quinazoline ring and the OH atom of Thr-766 increases greatly, which is about 6 Å, indicating that the N-3 of the quinazoline ring cannot form a stable hydrogen bond with the OH atom of Thr-766. Although compared with the binding mode I and II, the binding mode III can partly give a satisfactory explanation of the SAR data of the published quinazoline-type inhibitors; our simulations also indicate that this binding mode is unstable.

   4. Binding Mode IV. The binding mode IV has the most favorable binding free energy. For binding mode IV, the predicted binding model proposed by Wissner et al.[17] was treated as the initial model to perform MD simulations. According to the Wissner's model, the N1 atom of the quinazoline forms a hydrogen bond with the backbone NH of Met 769 and the N3 atom forms a hydrogen bond to a water molecule. This water molecule, in turn, forms a hydrogen bond with the hydroxyl group of Thr 830. So in the initial structure, a water molecule was manually placed
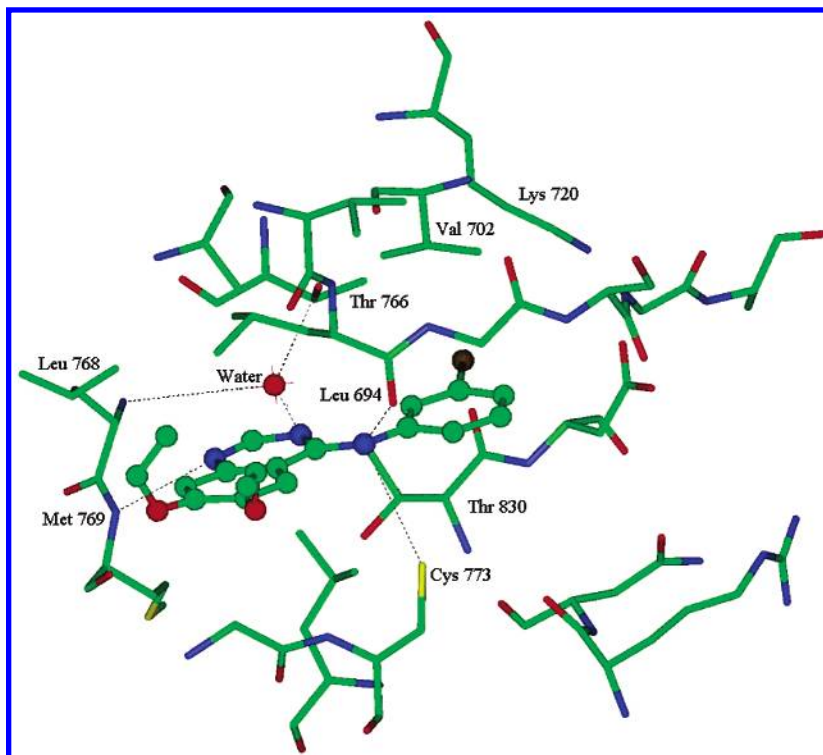
QUINAZOLINE TYPE EGF-R INHIBITORS

J. Chem. Inf. Comput. Sci., Vol. 43, No. 1, 2003 **285**

Figure 13. The averaged complexed structure of MD simulations proposed by Wissner et al.

between the N3 atom and the hydroxyl group of Thr 830. Figure 13 shows the complexed structure from averaging the MD trajectory. It is interesting to note that the position of the bridge water has changed significantly. The water molecule cannot be stably located between the N3 atom and the hydroxyl group of Thr 830, while it moves to the position between the N3 atom and the hydroxyl group of Thr-766. When we carefully examined the trajectories from the MD simulations, we could find that after the initial 50 MD simulation, the position of the bridge water molecules did not change significantly. The oxygen of the bridge water is about 3.2 Å away from the N3 atom and 2.9 Å from the oxygen atom the Thr 766, forming stable bifuricated hydrogen bonds. It is obvious that the water molecule can form two stable hydrogen bonds with the N3 atom and the side chain of Thr 766. Moreover, the oxygen atom of the water molecule is located within 4 Å of the N atom of Leu 768, which may produce relatively strong electrostatic interaction with the N atom of Leu 768. During MD simulations, the hydrogen bond between the N1 atom and the backbone NH of Met 769 was well reserved. In the CoMFA contour maps of H-bonding fields, the N1 atom in quinazoline ring was also determined as a potential H-bonding acceptor.

Beside these two important hydrogen bonds, several favorable interactions between the ligand and the enzyme are clearly demonstrated in Figure 13. The NH linker can form favorable electrostatic interactions with the NH atom of Leu 694 and the SH atom of Cys 773, which is also in good agreement with the CoMFA H-bonding contour maps. The 5-bromoanilino substituent is located in a deep, moderately sized pocket adjacent to the adenine binding pocket. The bottom of this inhibitor is made up of Val 702, Lys 721, Thr 766, and Thr 830. The side chains of these four residues are more hydrophobic, which can produce relative strong van der Waals contacts and hydrophobic interactions.

The above CoMFA steric contour maps also show that hydrophobic, sterically demanding substituents filling this pocket increase affinity, which agrees with the binding mode IV. The binding mode IV shows that two ethoxyl groups linked to the 6- and 7-positions in quinazoline are oriented outside without obvious unfavorable atom bumps, which is consistent with two green regions in Figures 5 and 6. Moreover, the ethoxyl group at the 7-position of the quinazoline can form relatively strong van de Waals and hydrophobic interactions with the nonpolar side chains of Leu 694 and Leu 768.

This binding mode was found to satisfy all the SAR data described previously. Previous work shows that the substituents linked to the anilino group should be small in size, which can be simply interpreted by the surface complementary between ligand and receptor. Because the anilino group is located deep in the binding pocket and produces close contact with the receptor, introduction of large groups on the anilino will be unfavorable to the surface and energetic complementarity between ligand and receptor and greatly decrease the ligand binding. In the binding mode IV, it seems that the influence of the N1 and N3 atoms of the quinazoline ring to the binding affinities should be quite different. We believe that the binding affinity of inhibitor is more sensitive to the changes of the N1 atom to the other nonpolar atoms than those of the N3 atom to the other nonpolar atoms, because the N1 atom can form a stable hydrogen bond with EGF-R, while the N3 atom only forms an indirect interaction with the receptor through a bridge water molecule. The above conclusion is in good agreement with the previous experiments.[17] If we change the N3 atom of a ligand to other atoms, link one strong H-bond donor or acceptor at this site, and make the ligand directly interact with EGF-R through H-bonding, the biological activities of the ligand should be improved. The binding mode shows that the introduction of large groups at the 5- and 8- positions is not desirable for
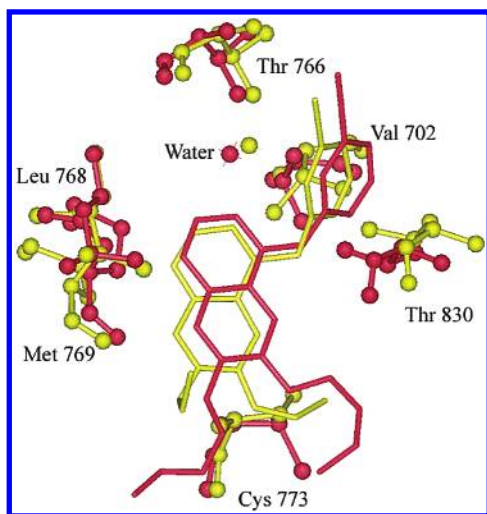
Figure 14. Alignment of the model structure (yellow, binding mode IV) and the crystal structure (red) for EGF-R/quinazoline complex. The least-squares fit was only performed for the heavy atoms of residue shown in the figure and the bridge water molecule and the common structure of the ligands. The residues and the bridge water molecules are shown in ball-and-stick; the ligands are shown in stick.

high potency, because larger groups at these two positions may produce unfavorable van der Waals contact with EGF-R. So if we expect to get compounds with increased potency by introducing new groups in the quinazoline ring, only modification of groups at positions 6 and 7 in the quinazoline ring should be suggested. Moreover, the N atom in anilino is very important. If the N atom is modified to the other atoms, the potency of inhibitor should be greatly reduced, which can be suggested by the strong electrostatic interactions between the atom and two residues in protein (Leu 694 and Cys 773).

When we began this project, the crystal structure of EGF-R had not been reported, so we applied homology modeling to construct the 3D structure of EGF-R. But after the manuscript had been examined by two experts, one reviewer indicated that recently the X-ray structure of EGFR-R in complex with a quinaoline-type inhibitor (very similar to A56) had been released from the Brookhaven Protein Data Bank (PDB code: 1M17). So according to the information afforded by the reviewer, we searched the Brookhaven Protein Data Bank and really found 1M17. The revised day of this structure is September 4, 2002, but the reference of 1M17 has not been reported until now.[59] Here, it is very challenging for us to compare the X-structure of 1M17 with the complexed structure predicted by us. When we gave a simple observation of 1M17 in InsightII, we found that the crystal structure was quite close to the binding mode suggested by us. To perform a careful and clear comparison between our model and the X-ray structure, we only extracted several important residues near the ligands from the predicted complex and 1M17, respectively. Figure 14 shows the alignment of the model structure and the crystal structure. During the alignment based on the least-squares fit, the heavy atoms of the residue, the heavy atoms of the common structures of these two ligands, and two water molecules were used as fit centers. From this figure, one can see that the position and orientation of the inhibitor was well predicted and overall the root-mean-square deviation is only 1.2 Å. It is really more encouraging that the bridge water molecule

predicted by us was also found in the crystal structure, and the predicted interacting form of this bridge water was even nearly the same as that in the crystal structure. In 1M17, a bridge water is also located at the position between between the N3 atom and the hydroxyl group of Thr 766. The oxygen of the bridge water is 2.8 Å away from the N3 atom and 3.1 Å from the oxygen atom of Thr 766, which is in good agreement with our prediction. In Wissner's work, the authors indicated the N3 atom may interact with protein through a water molecule, but they fail to predict the correct interacting form of this bridge water. In Wissner's model, the bridge water molecule forms a hydrogen bond with the hydroxyl group of Thr 830, not Thr 766 as shown in the crystal structure and binding model IV in Figure 14. Although the precise structure of the model proposed by Wissner et al. is unavailable, due to the failure in the prediction of the bridge water, the binding model proposed by Wissner et al. should show some crucial discrepancies with the crystal structure.

At last, it is necessary to point out that the DOCK calculations in this paper do not give the proper orientation of **A56** in EGF-R. The reason lies in the fact that the present DOCK program does not consider the flexibility of the protein. In this paper, the anilino substituent is located in a deep, moderately sized pocket adjacent to the adenine binding pocket, which is not occupied by ATP. In the initial structure of EGF-R for molecular docking, the binding pocket contacting with the anilino group is not fully opened, so the anailino group cannot be properly placed into the binding pocket by the DOCK program without unfavorable atom bumps.

REFERENCES AND NOTES

(1) Yarden, Y.; Harari, I.; Schlessinger, J. *J. Biol. Chem.* **1985**, *260*, 315−319.
(2) Yarden, Y.; Schlessinger, J. *Biochemistry* **1987**, *26*, 1434−1442.
(3) Cohen, S.; Carpenter, G.; King, L., Jr. *J. Biol. Chem.* **1980**, *255*, 4834−4842.
(4) Halaban, R. *Cancer Met. Rev.* **1991**, *10*, 129−140.
(5) Cohen, D. W.; Simak, R.; Rair, W. R.; Melamed, J.; Scher, H. I.; Cordon-Cardo, C. *J. Urol.* **1994**, *152*, 2120−2124.
(6) Grandis, J. R.; Melhem, M. F.; Gooding, W. E.; Day, R.; Holst, V. A.; Wagener, M. M.; Drenning, S. D.; Tweardy, D. J. *J. Natl. Cancer Inst.* **1998**, *90*, 824−832.
(7) Morishigie, K. I.; Kurachi, H.; Ameniya, K.; Fujita, Y.; Yamamoto, T.; Mikaye, A.; Tanizawa, O. *Cancer Res.* **1991**, *51*, 5322−5328.
(8) Rusch, V.; Klimstra, D.; Venkatraman, E.; Pisters, P. W. T.; Langenfeld, J.; Dmitrovsky, E. *Clin. Cancer Res.* **1997**, *3*, 5515−522.
(9) Thogersen, V. B.; Jorgensen, P. E.; Sorensen, B. S.; Bross, P.; Orntoft, T.; Wolf, H.; Nexo, E. *J. Clin. Lab Invest.* **1999**, *59*, 267−277.
(10) Plowman, G. D.; Ullrich, A.; Shawver, L. K. *Drug News Perspect.* **1994**, *7*, 334−337.
(11) Fry D. W.; Kraker, A.; J.; MaMichael, A.; Ambroso, L. A.; Nelson, J. M.; Leopold, W. R.; Connors, R. W.; Bridges, A. J. *Science* **1994**, *265*, 1093−1095.
(12) Traxler, P. M.; Furet, P.; Mett, H.; Buchdunger, E.; Meyer, T.; Lydon, N. J. *Med. Chem.* **1996**, *39*, 2285−2292.
(13) Rewcastle, G. W.; Denny, W. A.; Bridges, A. J. Zhou, H. R.; Cody, D. R.; McMichael, A.; Fry, D. W. *J. Med. Chem.* **1995**, *38*, 3482−3487.

QUINAZOLINE TYPE EGF-R INHIBITORS

J. Chem. Inf. Comput. Sci., Vol. 43, No. 1, 2003 **287**

(14) Bridges, A. J.; Zhou, H. R.; Cody, D. R.; Rewcastle, G. W.; McMichael, A.; Hollis Showalter, H. D.; Fry, D. W.; Kraker, A. J.; Denny, W. A. *J. Med. Chem.* **1996**, *39*, 267−276.
(15) Gordon W. R.; Palmer, B. D.; Bridges, A. J.; Hollis Showalter, H. D.; Sun, L.; Nelson, J.; McMichael, A.; Kraker, A. J.; Fry, D. W.; Denny, W. A. *J. Med. Chem.* **1996**, *39*, 918−928.
(16) Palmer, B. D.; Trumpp-Kallmeyer, S.; Fry, D. W.; Nelson, J. M.; Hollis Showalter, H. D.; Denny, W. A. *J. Med. Chem.* **1997**, *40*, 1519−1529.
(17) Wissner, A.; Berger, D. M.; Boschelli, D. H.; Floyd, M. B.; Greenberger, L. M.; Gruber, B. C.; Johnson, B. D.; Mamuya, N.; Nilakantan, R.; Reich, M. F.; Shen, R.; Tsou, H. R.; Upeslacis, E.; Wang, Y. F.; Wu, B. Q.; Ye, F.; Zhang, N. *J. Med. Chem.* **2000**, *43*, 3244−3256.
(18) Bridges, A. J. *Cur. Med. Chem.* **1999**, *6*, 825−843.
(19) Traxler, P. M. *Exp. Opin. Ther. Patents* **1997**, *7*, 571−588.
(20) Traxler, P. M. *Exp. Opin. Ther. Patents* **1998**, *8*, 1599−1625.
(21) Cramer, R. D., III; Patterson, D. E.; Bunce, J. D. *J. Am. Chem. Soc.* **1988**, *110*, 5959−5967.
(22) Hou, T. J.; Xu, X. J. *Prog. Chem.* **2001**, *13*, 436−440.
(23) Klebe, G.; Abraham, U.; Mietzner, T. *J. Med. Chem.* **1994**, *37*, 4130−4146.
(24) Srinvasan, J.; Cheatham, T. E.; Cieplak, P.; Kollman, P. A.; Case, D. A. *J. Am. Chem. Soc.* **1998**, *120*, 9401−9509.
(25) Honig, B.; Nicholls, A. *Science* **1995**, *268*, 1144−1149.
(26) Sitkoff, D.; Sharp, K. A.; Honig, B. *J. Phys. Chem.* **1994**, *98*, 1978−1988.
(27) Wilson, E. B.; Decius, J. C.; Cross, P. C. *Molecular Vibrations*; McGraw-Hill: New York, 1955.
(28) Chong, L. T.; Duan, Y.; Wang, L.; Massova, I.; Kollman, P. A. *Proc. Natl. Acad. Soc.* **1999**, *96*, 14330−14335.
(29) Massova, I.; Kollman, P. A. *Perspect. Drug Discov. Des.* **2000**, *18*, 113−135.
(30) Massova, I.; Kollman, P. A. *J. Am. Chem. Soc.* **1999**, *121*, 8133−8134.
(31) Kuhn, B.; Kollman, P. A. *J. Med. Chem.* **2000**, *43*, 3786−3791.
(32) Wang, J. M.; Morin, P.; Wang, W.; Kollman, P. A. *J. Am. Chem. Soc.* **2001**, *123*, 5221−5230.
(33) Hou, T. J.; Xu, X. J. *J. Phys. Chem. B* **2002**, *106*, 5527−5535.
(34) Tripos 6.5 User Guide, Tripos Inc. St. Louis, U.S.A. 1995. http://www.tripos.com.
(35) Halgren, T. A. *J. Comput. Chem.* **1996**, *17*, 490−519.
(36) Gasteiger, J.; Marsili, M. *Tetrahedron* **1980**, *36*, 3219−3228 1980.
(37) Marsili, Ml; Gasteiger, J. *Croat. Chem. Acta* **1980**, *53*, 601−614.
(38) Gasteiger, J.; Marsili, M. *Organ. Magn. Reson.* **1981**, *15*, 353−360.
(39) Streitwieser, A. *Molecular Orbital Theory for Organic Chemists*; Wiley: New York, 1961.
(40) Stewart, J. J. P. MOPAC 6.0, Quantum Chemical Program Exchange 455, 1990.
(41) Cerius2 User Guide, Molecular Simulation Inc., San Diego, U.S.A., 2001.
(42) Dauber-Osguthorpe, P.; Roberts, V. A.; Osguthorpe, D. J.; Wolff, J.; Genest, M.; Hagler, A. T. *Proteins* **1988**, *4*, 31.
(43) Bohacek, R. S.; McMartin, C. *J. Med. Chem.* **1992**, *35*, 1671−1684.
(44) Bush, B. L.; Nachbar, R. B. *J. Comput.-Aided Mol. Des.* **1993**, *7*, 587−619.
(45) Singh, J.; Dohbrusin, E. M.; Fry, D. W.; Haske, T.; Whitty, A.; McNamara, D. J. *J. Med. Chem.* **1997**, *40*, 1130−1135.
(46) Insight II 2000 User Guide, MSI Inc., San Diego, U.S.A., 2000.
(47) Case, D. A.; Pearlman, D. A.; Caldwell, J. W.; Cheatham, T. E., III; Ross, W. S.; Simmerling, C. L.; Darden, T. A.; Merz, K. M.; Stanton, R. V.; Cheng, A. L.; Vincent, J. J.; Crowley, M.; Tsui, V.; Radmer, R. J.; Duan, Y.; Pitera, J.; Massova, I.; Seibel, G. L.; Singh, U. C.; Weiner, P. K.; Kollman, P. A. AMBER 6, University of California, San Francisco, 1999.
(48) Cleplak, P.; Cornell, W. D.; Bayly, C.; Kollman, P. A. *J. Comput. Chem.* **1994**, *16*, 1357−1377.
(49) Kuntz, I. D.; Blaney, J. M.; Oatley, S. J.; Langridge, R.; Ferrin, T. E. *J. Mol. Biol.* **1982**, *161*, 269−288.
(50) Ewing, T. J. A.; Kuntz, I. D. *J. Comput. Chem.* **1997**, *18*, 1175−1189.
(51) Honig, B.; Nicholls, A. *Science* **1995**, *268*, 1144−1149.
(52) Gilson, M. K.; Honig, B. *Proteins* **1998**, *4*, 7−18.
(53) Sitkoff, D.; Sharp, K. A.; Honig, B. *J. Phys. Chem.* **1994**, *98*, 1978−1988.
(54) Sanner, M. F.; Olson, A. J.; Spehner, J. *Biopolymer* **1996**, *38*, 305−320.
(55) Hou, T. J.; Li, Y. Y.; Xu, X. J. *J. Mol. Model.* **2000**, *6*, 438−445.
(56) Hou, T. J.; Li, Z. M.; Li, Z.; Liu, J.; Xu, X. J. *J. Chem. Inf. Comput. Sci.* **2000**, *40*, 1002−1009.
(57) L. L. Zhu, T. J. Hou, X. J. Xu, *J. Mol. Model.* **2001**, *7*, 223−230.
(58) T. J. Hou, X. J. Xu, *Chemometr. Intell. Lab.* **2001**, *56*, 123−132.
(59) Stamos, J.; Sliwkowski, M. X.; Eigenbrot, C. (To be published).