Available online at www.sciencedirect.com
SCIENCE DIRECT®
Bioorganic & Medicinal Chemistry 12 (2004) 2409–2417
Bioorganic & Medicinal Chemistry

Elucidating inhibitory models of the inhibitors of epidermal growth factor receptor by docking and 3D-QSAR ☆

Gang Chen,[a,b] Xiaomin Luo,[a,*] Weiliang Zhu,[b] Cheng Luo,[a] Hong Liu,[a] Chum Mok Puah,[b] Kaixian Chen[a] and Hualiang Jiang[a,*]

[a]*Drug Discovery & Design Center and State Key Laboratory of Drug Research, Shanghai Institute of Materia Medica, Shanghai Institutes of Biological Sciences, Chinese Academy of Sciences, 555 Zuchongzhi Road, Shanghai 201203, PR China*
[b]*School of Chemical & Life Sciences, Technology Centre for Life Sciences, Singapore Polytechnic, 500 Dover Road, Singapore 139651, Singapore*

Received 16 December 2003; revised 30 January 2004; accepted 31 January 2004

Abstract—Epidermal growth factor receptor (EGFR) protein tyrosine kinases (PTKs) are attractive targets for anti-tumor drug design. Although thousands of their ligands have been studied as potential inhibitors against PTKs, there is no QSAR study that covers different kinds of inhibitors with observable structural diversity. However, by using this approach, we could mine far more useful information. Hence in order to better understand the binding model and the relationship between the physicochemical properties and the inhibitory activities of different kind of various inhibitors, molecular docking and 3D-QSAR, viz. CoMFA and CoMSIA, were combined to study 124 reported inhibitors with different scaffolds. Based on the docked binding conformations, highly reliable and predictive 3D-QSAR models were derived, which reveal how steric, electrostatic, and hydrophobic interactions contribute to inhibitors' bioactivities. This result also demonstrates that it is possible to include different kinds of inhibitors with observable structural diversity into one 3D-QSAR study. Therefore, this study not only casts light on binding mechanism between EGFR and its inhibitors, but also provides new hints for de novo design of new EGFR inhibitors with observable structural diversity.

1. Introduction

Protein tyrosine kinases (PTKs) play important roles in activating numerous signal transduction pathways within cells, leading to cellular proliferation, differentiation, and various regulatory mechanisms.[1,2] Epidermal growth factor receptor (EGFR), which was identified as a kind of PTK in 1980s that belongs to a large family of the transmembrane growth factor receptor PTKs, is activated by the binding of epidermal growth factor (EGF).[3] Mistakenly regulated activity or over-expression of the receptor have been demonstrated to be related to many human cancers such as breast and liver cancers,[3–6] leading many to believe that EGFR PTK is an attractive target for anti-tumor drug discovery.[7–11] Thus by reasoning, inhibitors of the EGFR PTK are expected to suppress the kinase's activity, with great therapeutic potential for the treatment of malignant and nonmalignant tumors.[11]

In last decade or so, several kinds of compounds were reported as tyrosine kinase inhibitors.[3,12–21] The total number of intensively studied chemicals were estimated to be in their thousands.[17] Although many of them possess potent inhibitory activities against the tyrosine kinase, they often lack selectivity or show only weak cellular and in vivo potency.[3,17] Therefore, it is still a big challenge to develop successful drugs targeting at EGFR. Quantitative structure–activity relationship at the three-dimensional (3D) level (3D-QSAR) is in general considered as a powerful approach for developing new drug leads based on small ligand structures. Although there are some reports on 3D-QSAR, such as comparative molecular field analysis (CoMFA)[22] and comparative molecular similarity index analysis (CoMSIA),[23] studies on the EGFR inhibitors are limited to one kind of inhibitors.[17,24–26] Moreover no 3D-QSAR

Keywords: Inhibitors; EGFR; Docking; QSAR; CoMFA; CoMSIA.
☆ Supplementary data associated with this article can be found, in the online version, at doi:10.1016/j.bmc.2004.02.001
* Corresponding authors. Tel.: +86-21-50806600x1203; fax: +86-21-62797092; e-mail: hljiang@mail.shcnc.ac.cn

0968-0896/$ - see front matter © 2004 Elsevier Ltd. All rights reserved.
doi:10.1016/j.bmc.2004.02.001

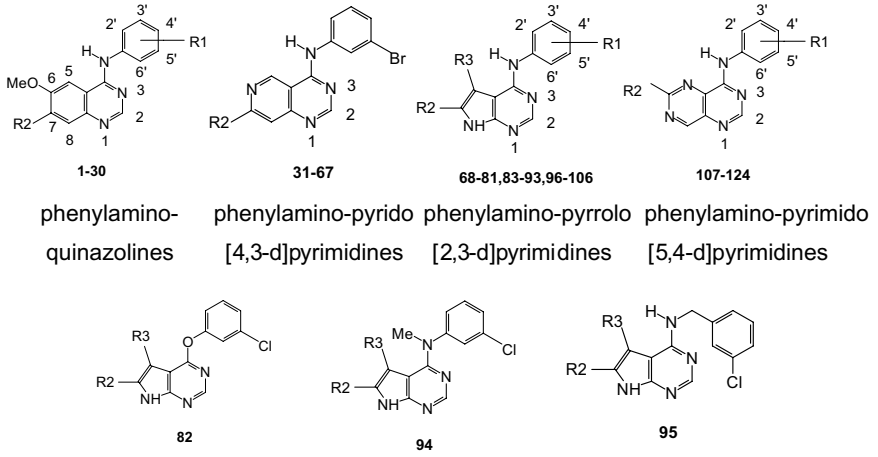
studies have been found on the structures that covered two or more different kinds of ligands. One reason may lie in the difficulty of aligning compounds that have distinct differences in structures, for the key step of using with CoMFA or CoMSIA lies in the alignment. Nevertheless, there should be abundant information hidden behind the observed results, and waiting to be mined. A 3D-QSAR study on ligands with observable structure diversity, if possible, will definitely lead to more universal and robust QSAR models for designing novel compounds in inhibiting EGFR. To address such powerful models covering different types of ligands, we carried out CoMFA and CoMSIA studies on 124 reversibly ATP-competitive inhibitors of EGFR PTK, belonging to four types of reported EGFR inhibitors in this research. They are phenylamino-quinazolines, phenylamino-pyrido[4,3-*d*]pyrimidines, phenylamino-pyrrolo[2,3-*d*]pyrimidines, and phenylamino-pyrimido[5,4-*d*]pyrimidines.

2. Materials and methods

2.1. Structures and activities of inhibitors

The structures of 124 EGFR TK inhibitors and their inhibitory activities ($-\log IC_{50}$s) are shown in Table 1.[18–21]

Table 1. Structure, experimental, and predicted activities of CoMFA and CoMSIA of the 124 inhibitors



No.	R_1				R_2	R_3	$\log 1/C$		
	2′	3′	4′	5′			Obsd	CoMFA	CoMSIA
1	F	Br			1-(1,2,3-Triazolyl)-$(CH_2)_2O$		6.40	6.160	6.516
2	F	Br			$MeN(CH_2CH_2)_2CH–CH_2O$		6.30	6.352	6.569
3	F	Cl			$MeO(CH_2)_2O$		7.00	6.818	6.658
4	F	Cl			4-Me-piperazinyl-$(CH_2)_3O$		7.12	7.203	6.836
5	F	Br			4-Me-piperazinyl-$(CH_2)_3O$		7.00	6.939	7.742
6*	F	Cl			4-Me-piperazinyl-$(CH_2)_2O$		6.70	6.752	6.734
7	F	Cl			4-Morpholinyl-$(CH_2)_3O$		7.00	6.967	6.574
8	F	Cl			4-Morpholinyl-$(CH_2)_2O$		6.40	5.988	6.578
9	F	Cl			1-Pyrrolidinyl-$(CH_2)_3O$		7.00	7.037	6.864
10	F	Cl			$(CH_2)_4N–CH_2CH=CH–CH_2O$		7.05	6.745	7.188
11	F	Br			$(CH_2)_4N–CH_2CH=CH–CH_2O$		7.00	6.995	6.825
12	F	CN			$(CH_2)_4N–CH_2CH=CH–CH_2O$		6.10	6.105	6.050
13	F	Cl			4-Pyridyl-N(Me)–$(CH_2)_2O$		6.52	5.794	6.239
14	F	Cl			$MeN(CH_2CH_2)_2CH–CH_2O$		6.52	6.399	7.058
15	F	Me			$MeN(CH_2CH_2)_2CH–CH_2O$		6.00	6.119	6.302
16*	F	Cl	OH		$MeN(CH_2CH_2)_2CH–CH_2O$		7.00	6.960	6.988
17	F	Me	OH		$MeN(CH_2CH_2)_2CH–CH_2O$		6.52	6.536	6.375
18	F	Cl		F	$MeN(CH_2CH_2)_2CH–CH_2O$		6.35	6.599	6.200
19	F	Br		F	$MeN(CH_2CH_2)_2CH–CH_2O$		6.52	6.719	6.581
20	F	Cl			$HN(CH_2CH_2)_2CH–CH_2O$		6.52	6.542	6.325
21	F	Br			$HN(CH_2CH_2)_2CH–CH_2O$		6.70	6.566	6.731
22	F	Me			$HN(CH_2CH_2)_2CH–CH_2O$		6.30	6.432	6.092
23	F	Cl		F	$HN(CH_2CH_2)_2CH–CH_2O$		6.30	6.214	5.935
24	F	Br		F	$HN(CH_2CH_2)_2CH–CH_2O$		6.35	6.242	6.578
25	F	Cl			$MeN(CH_2CH_2)_2CH–CH_2CH_2O$		6.52	6.671	6.953
26*	F	Br			$MeN(CH_2CH_2)_2CH–CH_2CH_2O$		6.70	6.992	7.087

Table 1 (*continued*)

No.	R_1				R_2	R_3	log 1/*C*		
	2′	3′	4′	5′			Obsd	CoMFA	CoMSIA
27	F	Br			$HN(CH_2CH_2)_2CH–CH_2CH_2O$		6.70	6.515	6.356
28	F	Cl			(*R*)-$MeN(CH_2)(CH_2)_3CH–CH_2O$		7.00	6.974	7.029
29	F	Br			(*R*)-$MeN(CH_2)(CH_2)_3CH–CH_2O$		6.70	6.427	7.314
30	F	Br			(*S*)-$MeN(CH_2)(CH_2)_3CH–CH_2O$		7.00	7.373	7.702
31		Br			NH_2		8.00	8.653	8.509
32		Br			NHMe		9.89	9.268	9.372
33		Br			NMe_2		10.05	9.133	9.536
34		Br			$NHCH_2CH_2OH$		9.62	9.294	9.367
35		Br			$N(Me)CH_2CH_2OH$		8.59	8.743	8.727
36*		Br			$NHCH_2CH(OH)CH_2OH$		9.04	8.887	8.802
37		Br			$N(Me)CH_2CH(OH)CH_2OH$		8.49	8.577	9.157
38		Br			$NHCH(CH_2OH)_2$		7.85	8.169	8.573
39		Br			$N(CH_2CH_2OH)_2$		7.92	7.728	8.001
40		Br			$NHCH_2(CHOH)_4CH_2OH$		8.32	8.019	8.320
41		Br			$N(Me)CH_2(CHOH)_4CH_2OH$		8.49	8.339	8.631
42		Br			$NH(CH_2)_2NMe_2$		7.35	7.048	7.253
43		Br			$NH(CH_2)_3NMe_2$		8.06	8.135	8.133
44		Br			$NH(CH_2)_4NMe_2$		8.13	8.814	8.601
45		Br			$NH(CH_2)_5NMe_2$		8.08	7.966	7.969
46*		Br			$N(Me)(CH_2)_2NMe_2$		7.40	7.753	8.075
47		Br			$NH(CH_2)_2$Nmorph		8.49	8.538	9.062
48		Br			$NH(CH_2)_3$Nmorph		8.72	8.801	8.997
49		Br			$NH(CH_2)_4$Nmorph		8.27	7.848	8.524
50		Br			$NH(CH_2)_3$N(O)morph		9.13	9.231	8.785
51		Br			$NH(CH_2)_3$NMepip		8.31	8.244	7.670
52		Br			$NH(CH_2)_2N(CH_2CH_2OH)_2$		8.04	7.966	7.665
53		Br			$NH(CH_2)_3N(CH_2CH_2OH)_2$		8.92	8.817	8.398
54		Br			$NHNH_2$		8.15	8.330	8.004
55		Br			$NH(CH_2)_3$(1-imidazolyl)		9.29	9.397	9.161
56*		Br			$NH(CH_2)_2$(4-imidazolyl)		9.04	9.481	9.372
57		Br			$NHCH_2COOH$		8.82	8.971	8.647
58		Br			$NHCH_2CH_2COOH$		9.21	8.806	9.349
59		Br			$NH(CH_2)_3COOH$		9.55	9.718	9.570
60		Br			$N(Me)CH_2COOH$		7.80	7.771	8.663
61		Br			$NHCH_2CH_2SO_3H$		8.85	9.492	9.441
62		Me			NHMe		8.85	8.707	8.442
63		Me			$NH(CH_2)_4NMe_2$		8.27	8.704	7.825
64		Me			$NH(CH_2)_3$Nmorph		8.03	7.895	8.202
65		Me			$NH(CH_2)_3$Nmepipazin		8.25	8.416	8.229
66*		Me			$NH(CH_2)_3$(1-imidazolyl)		8.46	8.119	7.876
67		Me			$NH(CH_2)_2$(4-imidazolyl)		8.14	8.094	7.952
68		Cl			Me	Me	7.57	6.814	6.988
69		Cl			C_6H5	Me	8.00	7.301	7.923
70		Cl			Me	C_6H5	6.64	6.723	6.856
71		Cl			C_6H5	C_6H5	7.02	7.683	7.595
72		Cl			CONHMe	H	8.52	8.576	7.924
73		Cl			Pyridin-2-yl	H	8.15	7.558	7.712
74		Cl			4-OMe–C_6H_4	H	7.82	7.700	8.022
75		Cl			4-OH–C_6H_4	H	8.52	8.376	8.568
76*		Cl			4-NH_2–C_6H_4	H	8.52	7.984	7.782
77		Cl			3-NH_2–C_6H_4	H	8.40	8.577	8.021
78		Cl			4-COOH–C_6H_4	H	9.00	9.022	8.903
79					Me	Me	5.72	5.742	6.122
80		Me			Me	Me	6.24	6.369	6.227
81		F			Me	Me	6.26	6.476	6.024
82		Cl			Me	Me	5.60	6.122	6.654
83		Br			Me	Me	7.60	6.880	7.230
84		OH			Me	Me	5.90	5.791	5.424
85		OMe			Me	Me	5.92	6.302	5.993
86*		CF_3			Me	Me	5.72	6.228	6.341
87		CO_2H			Me	Me	4.79	4.775	4.870
88		CN			Me	Me	6.70	6.224	5.856
89			Me		Me	Me	5.68	6.162	5.843
90			Et		Me	Me	4.11	4.782	5.649
91			F		Me	Me	6.25	6.280	6.166

(*continued on next page*)

Table 1 (*continued*)

No.	R_1				R_2	R_3	log 1/*C*		
	2′	3′	4′	5′			Obsd	CoMFA	CoMSIA
92	Cl				Me	Me	4.12	5.151	4.918
93		Cl		Cl	Me	Me	6.77	6.911	7.086
94		Cl			Me	Me	6.30	5.836	5.933
95					Me	Me	5.97	5.955	5.729
96*					$-CH_2CH_2CH_2CH_2-$		6.51	6.406	6.411
97		Me			$-CH_2CH_2CH_2CH_2-$		6.09	7.426	7.002
98		Cl			$-CH_2CH_2CH_2CH_2-$		7.54	7.561	7.199
99		Br			$-CH_2CH_2CH_2CH_2-$		7.34	7.141	7.709
100		OH			$-CH_2CH_2CH_2CH_2-$		6.38	6.237	5.999
101		OMe			$-CH_2CH_2CH_2CH_2-$		6.07	5.988	5.745
102		CF_3			$-CH_2CH_2CH_2CH_2-$		6.44	5.697	6.535
103		CO_2H			$-CH_2CH_2CH_2CH_2-$		4.59	4.774	5.314
104			Me		$-CH_2CH_2CH_2CH_2-$		6.96	6.804	6.336
105					$-CH_2CH_2CH_2CH_2-$		8.22	7.431	6.387
106*					$-CH{=}CHCH{=}CH-$		5.16	5.554	5.985
107					Cl		5.59	6.887	6.393
108					NHMe		7.89	7.479	7.685
109		Br			Cl		7.09	7.099	7.044
110		Br			NH_2		8.82	8.646	8.415
111		Br			NHMe		9.12	8.702	8.428
112		Br			NMe_2		9.02	9.232	9.099
113		Br			OMe		8.42	8.511	8.877
114		Br			$NH(CH_2)_2NMe_2$		7.46	7.426	8.226
115		Br			$NH(CH_2)_2$morpholide		9.09	8.830	8.711
116*		Br			$NH(CH_2)_3$morpholide		8.54	8.483	7.740
117		Br			$NH(CH_2)_2$(4-imidazolyl)		9.60	9.489	9.005
118		Br			$NH(CH_2)_3$(1-imidazolyl)		8.64	8.731	8.257
119		Me			Cl		6.42	7.260	7.084
120		Me			NH_2		7.77	8.385	7.882
121		Me			NHMe		8.37	8.278	8.150
122		Me			NMe_2		8.40	8.677	7.599
123		Me			$NH(CH_2)_2$morpholide		8.64	8.521	8.246
124		Me			$NH(CH_2)_2$(4-imidazolyl)		8.52	8.521	8.413

* Inhibitors are selected for test set.

The 3D structures of these compounds were constructed by using molecular modeling software package Sybyl6.8.[27] The geometries of these inhibitors were optimized using the Tripos force field[28] with Gasteiger–Hückel charges.[29–31] The Powell method available in the Maximin2 module encoded in Sybyl6.8 was used for energy minimization using an 8 Å nonbond cutoff and an energy convergence gradient value of 0.005 kcal/(mol Å).

2.2. EGFR structure

The X-ray crystal structure of the EGFR PTK in complex with inhibitor AQ4774[32,33] was recovered from the Protein Data Bank (PDB, http://www.rcsb.org/pdb/, entry code 1M17).[34] AQ4774 occupies the ATP binding pocket. Although residues of the enzyme were missing from 965 to 976 in the structure, they are too far away from the binding site. So we did not repair these residues for the protein structure. Subsequently, the residues from 977 to 996, which code the C-terminal of the protein and are far from the binding pocket as well, were also removed. After adding all the necessary hydrogen atoms, the AQ4774–protein complex was relaxed 200 steps by using the MMFF94 force field with MMFF94 charges.[35]

2.3. Molecular docking

DOCK4.0[36] was employed to dock all the inhibitors listed in Table 1 into the EGFR PTK binding pocket. Briefly, the molecular surface of the pocket was calculated using the AutoMS[37] program. Then, SPHGEN[38] was used to create the sphere clusters for the pocket. The steric and electrostatic environment of the pocket was evaluated with the program Grid.[36] During the docking simulations, the inhibitors were regarded as flexible and subjected to an energy minimization. For each inhibitor, the top 10 docked binding conformations were reported. We noticed that the skeletons of compounds **1**–**67** and **107**–**124** in Table 1 are somewhat similar to that of AQ4774. This skeleton is composed of a naphthalene-like nitrogen heterocondensed ring and a substituted benzene ring. All of the ring atoms of this skeleton were used to calculate the root mean square deviation (rmsd) between AQ4774 and each reported conformations. For compounds **68**–**106**, the atoms of the two six-membered rings are used to estimate rmsd. The binding confor-

mation of each inhibitor with least rmsd among the 10 reported conformations of each inhibitor was selected for further QSAR analyses.

2.4. QSAR analyses

To construct good predictive QSAR models and to evaluate the contributions of electrostatic, steric, and hydrophobic effects to the activities of EGFR PTK inhibitors, CoMFA[22] and CoMSIA[23] analyses were performed upon the binding conformations and their alignments derived from docking calculation.

2.5. CoMFA

For CoMFA analysis, steric and electrostatic fields were calculated using a sp^3 carbon as steric probe, and a +1 point charge as the electrostatic probe. Steric and electrostatic interactions were calculated using Tripos force field with a distance-dependent dielectric constant at all intersection in a regularly spaced (2 Å) grid. The minimum sigma (column filtering) was set to 2.0 kcal/mol to improve the signal-to-noise ratio by omitting those lattice points whose energy variation was below this threshold. A cutoff of 30 kcal/mol was adopted and the regression analysis was carried out using the full cross-validated partial least-squares (PLS) method (leave one out) with CoMFA standard options for scaling of variables.[39,40] The final model (non-cross-validated conventional analysis) was developed and yielded the highest cross-validated q^2 value with the optimum number of components equal to that yielding the highest q^2.

2.6. CoMSIA

For CoMSIA analysis, three physicochemical properties, steric, electrostatic, and hydrophobic fields, were evaluated. The steric contribution was reflected by the third power of the atomic radii of the atoms. Electrostatic properties were calculated based on the atomic charges. An atom-based hydrophobicity was assigned according to the parameterization developed by Viswanadhan and co-workers.[41] The lattice dimensions were selected with a sufficiently large margin (>4 Å) to enclose all molecules. Any singularities were avoided at atomic positions in CoMSIA fields because a Gaussian-type distance dependence of the physicochemical properties was adopted, thus no arbitrary cutoffs were required. In general, similarity indices, $A_{F,K}$ between the compounds of interest and a probe atom placed at the intersections of the lattice could be calculated with the equation below,[23]

$$A^q_{F,K}(j) = -\sum_{i=1}^{n} w_{probe,k} w_{ik} e^{-ar^2_{iq}}$$

where q represents a grid point, i is summation index over all atoms of molecule j under computation, w_{ik} is the actual value of the physicochemical property k of atom i, $w_{prob,k}$ is the value of the probe atom. In this study, similarity indices were computed using a probe atom ($w_{prob,k}$) with charge +1, radius 1 Å, hydrophobicity +1, and attenuation factor α 0.3 for the Gaussian-type distance. The statistical evaluation for the CoMSIA analyses was performed in the same way as described in CoMFA. Molecular modeling, docking, and QSAR-related calculations were carried out on a SGI O2 workstation with Sybyl6.8 and DOCK4.0.

3. Results and discussion

3.1. Interaction models

3.1.1. The structure of ATP binding site. The crystal structure of EGFR PTK revealed that the ATP binding pocket consists of Thr766, Gln767, Leu768, Met769, Gly772, Thr830, Asp831, Val702, Lys721, Ala719, Glu738, and Met742.[33] This binding pocket can be divided into five regions including two hydrophobic regions and an adenine region.[3] Hydrophobic region I shaped by Ala719-Lys721, Leu764-Thr766, Thr830 and Asp831 is located deeply inside the ATP binding pocket (refer to Supporting material, Fig. S1). Hydrophobic region II mainly comprises Leu694 and Gly776.[3] These two hydrophobic regions are almost opposite each other to the adenine region.

Docking simulation resulted in two different binding models for the inhibitors to EGFR PTK, designated as model 1 and model 2. Figure 1 illustrated these two binding models taking inhibitor **94** as an example. For model 1, the nitrogen atom N1 of the pyrrolo[2,3-*d*]pyrimidine (refer to Table 1 for atomic numbering) forms a hydrogen bond with the NH group of Met769, while the anilino group lies in a hydrophobic pocket I. The binding between AQ4774 and EGFR PTK in the crystal structure belongs to model 1. For model 2, although the nitrogen atom N1 of the pyrrolo[2,3-*d*]pyrimidine also forms a hydrogen bond with the NH group of Met769 (Fig. 1), the anilino group lies in the

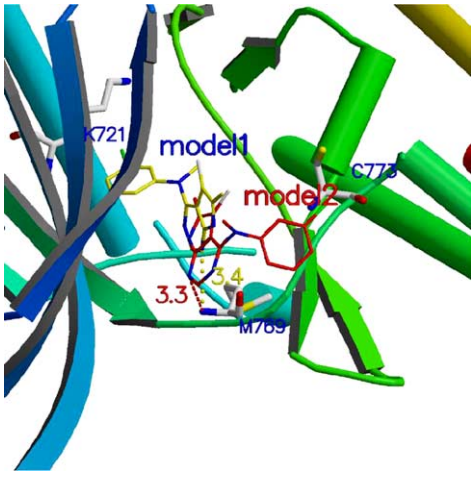


Figure 1. Binding models 1 and 2 of inhibitor **94**.

hydrophobic pocket II. Therefore, the orientation of anilino group is in opposite directions in these two binding models.

These two binding models are very similar to each other in terms of binding pattern.[42] The anilino moiety binds with a hydrophobic pocket in both models, while there is a hydrogen bond between the inhibitors and the NH backbone of Met769. The only difference is that the inhibitor binds to different hydrophobic pockets. Many researchers concluded that hydrophobic interaction dominates the binding between EGFR PTK and inhibitors.[3,19,42] Our docking result shows that the electrostatic interaction energy is no larger than 25% of the overall interaction energy between inhibitor and EGFR PTK, demonstrating that the hydrophobic interaction is essential to the binding. Meanwhile, we also noticed that all of the 124 inhibitors possess model 1 in binding to EGFR while only a few inhibitors adopt model 2. Therefore, model 1 was used in the QSAR analyses.

3.1.2. Binding conformations of the inhibitors. The DOCK4.0 predicted binding conformation of AQ4774 is shown in Figure 2 with the X-ray crystallographic obtained conformational superposition.[32,33] The rmsd of heavy atoms between these two conformations is ~1.46 Å. A large part of the rmsd is from $-OC_2H_4OCH_3$, which is very flexible and is almost located outside the binding pocket to access the solvent. The rmsd between the two conformations without $-OC_2H_4OCH_3$ moiety is only ~0.97 Å, indicating the parameters for the docking simulation are reasonable in reproducing the X-ray structure. DOCK4.0 and its parameter set could be extended to search the enzyme-binding conformations for other inhibitors accordingly.

Figure 3 illustrates 124 best conformations of the 124 inhibitors, chosen from the docked conformations according to their rmsds to AQ4774. It shows that the common ring structures were superimposed each other rather well. Based on this set of binding conformations and their alignment, CoMFA and CoMSIA analyses were performed.

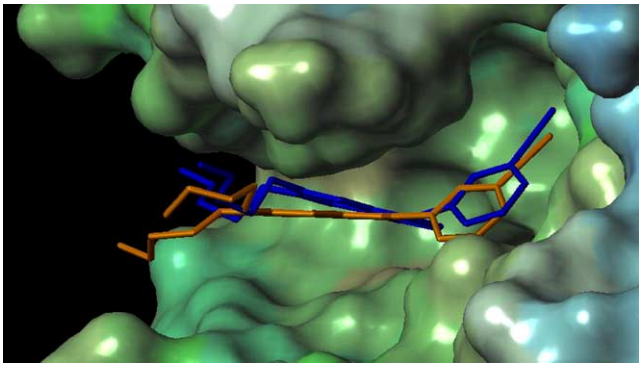


Figure 2. Docked (brown) and X-ray crystallographic (blue) conformations of AQ4774.

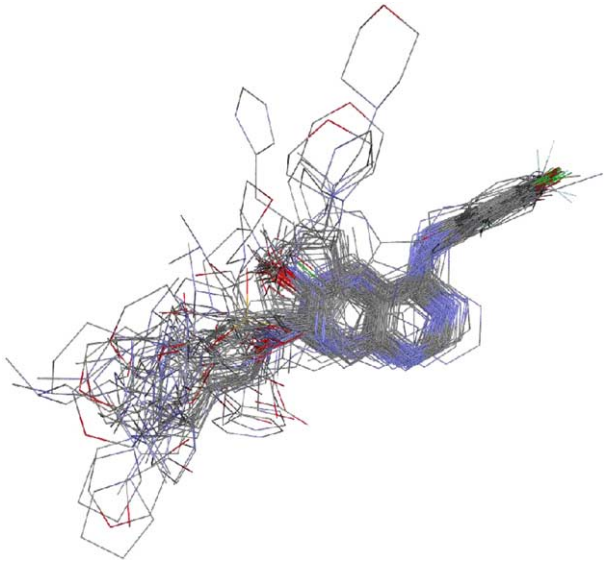


Figure 3. The docked 124 best binding conformations in the ATP binding site of EGFR.

3.1.3. Hydrogen-bonding interactions. Our docking results showed that all inhibitors of EGFR PTK were involved in a hydrogen bonding with the residue Met769 in the binding pocket (Fig. 1). This hydrogen bond is in fact crucial for the binding activities of EGFR inhibitors, in which more than a 3700-fold loss in inhibitory potency will be caused when substituted N1 with a carbon atom.[42] A second hydrogen bond can be formed between the N3 of the quinazoline ring and the side chain of Thr766, which is located at the beginning of the extended coil stretched deep into the binding cleft. Removal of the N3 of the quinazoline ring results in a moderate loss of affinity (a 200-fold loss in inhibitory potency), indicating that this hydrogen bond is probably less important than the first one.[42] Our docking result is in agreement with Palmer's result that most of the 124 inhibitors formed the first hydrogen bond, and a few of them formed the second hydrogen bond (refer to Supporting material, Fig. S2, for the details of the interactions between four classes of inhibitors and EGFR, which was derived by the program LIGPLOT[43]).

3.2. 3D-QSAR models

3.2.1. CoMFA. During 3D-QSAR analyses, we selected 112 (unasterisked molecules in Table 1) compounds for model construction and 12 compounds (asterisked molecules in Table 1) as testing set for model validation. The result of CoMFA analysis is summarized in Table 2. PLS analysis gave a correlation with a cross-validated q^2 of 0.591 and an optimum component number of 7. The non-cross-validated PLS analysis was repeated with the optimum number of components, determined by the cross-validated analysis, to give a r^2 of 0.904, $F_{7,103} = 139.9$, and an estimated standard error of 0.401. These values indicate a good conventional statistical correlation. Thus, the CoMFA model has a good predictive ability. The steric descriptors explain 50.6% of variance, while the electrostatic descriptors explain

Table 2. Statistical indexes of CoMFA and CoMSIA models based on 112 inhibitors binding conformers

	CoMFA	CoMSIA
PLS statistics		
q^2	0.591	0.588
r^2	0.904	0.856
s	0.401	0.487
F	139.900	125.904
Optimal components	7	5
Contribution		
Steric	0.506	0.212
Electrostatic	0.494	0.366
Hydrophobic		0.422

49.4%. The predicted inhibitory activities of the 112 compounds were listed in Table 1. The correlation between observed activities and the predicted values was depicted in Figure 4a. The good correlation suggests that a reliable CoMFA model was successfully constructed.

The CoMFA result was usually represented as 3D 'coefficient contour'. It shows regions where variations of steric and electrostatic nature in the structural features of the different molecules contained in the training set lead to increases or decreases in the activity. The CoMFA steric and electrostatic fields were presented as contour plots in Figure 5. To aid in visualization, compound **42** was displayed in the maps. Green colored regions indicated where increased steric bulk is associated with enhanced activity, and yellow colored regions suggested where increased steric bulk is unfavorable to activity. The biggest yellow steric region was found near the C5 and N6 of pyrido[4,3-*d*]pyrimidine moiety in the CoMFA contour map, suggesting that the modification of the inhibitors with bulky group to this region would lead to decrease in activity. In fact, no EGFR PTK inhibitors have a substituent at the site of C5. Two small yellow contours around the C2 of pyrido[4,3-*d*]pyrimidine moiety and 6′ position of anilino moiety suggest that the region near residue Val702 is quite sterically restricted. Another yellow region near R2 substituent moiety indicates that bulky R2 group is not beneficial to the inhibitory activity.

Regions where more positive charge is favorable to inhibitory activity are indicated in blue, while regions where increased negative charge is favorable to activity were indicated in red. Hence, red polyhedral near the R2 or R3 substituent shows that more negatively charged R2 or R3 groups should improve inhibitory activity. This is in agreement with fact that many inhibitors have atoms, such as N or O, with negative partial atomic charge near this region in their structures.

3.2.2. CoMSIA. CoMSIA analysis results are also summarized in Table 2. The CoMSIA model has a cross-validated q^2 of 0.588 for five optimum

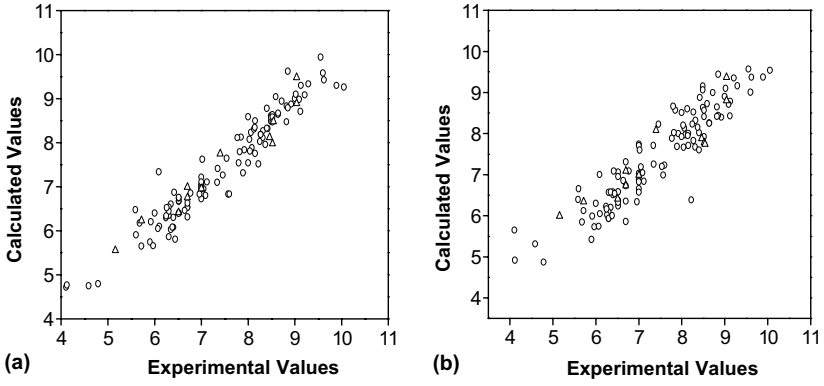


Figure 4. Correlation between predicted and the observed activities. ○: Compounds of the training set ($r^2 = 0.904$ and 0.856 for the two QSAR models, respectively); △: compounds of the testing set ($r = 0.966$ and 0.906 for the two QSAR models, respectively). (a) CoMFA result; (b) CoMSIA result.

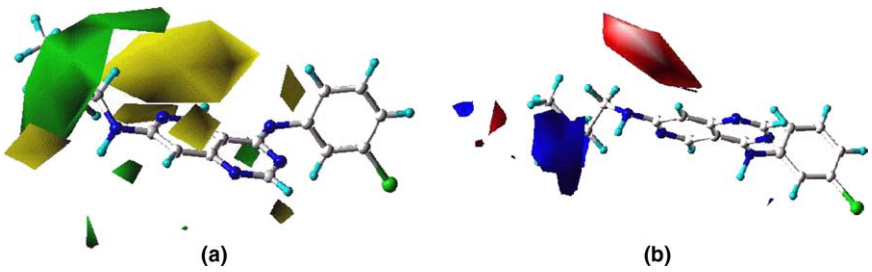


Figure 5. CoMFA contour maps in combination with inhibitor **42**: (a) the steric field distribution; (b) the electrostatic field distribution. The inhibitor is shown in ball-and-stick. Sterically favored areas in green; sterically unfavored areas in yellow. Positive potential favored areas in blue; positive potential unfavored areas in red.

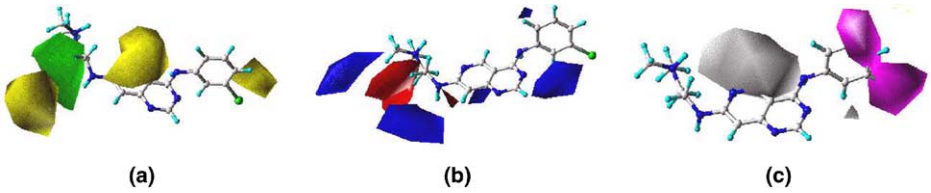


Figure 6. CoMSIA contour maps in combination with inhibitor **42**: (a) the steric field distribution; (b) the electrostatic field distribution; (c) the hydrophobic field distribution. Sterically favored areas in green; sterically unfavored areas in yellow. Positive potential favored areas in blue; positive potential unfavored areas in red. Hydrophobic favored areas in magenta; hydrophilic favored areas in white.

components. The non-cross-validated PLS analysis was repeated with the optimum number of components, giving a r^2 of 0.856, $F_{5,103} = 125.9$, and a estimated standard error of 0.487. These data demonstrate that the CoMSIA model is also fairly predictive. The predicted inhibitory potencies of the 112 compounds are listed in Table 1, and the correlation between predicted and observed activities was shown in Figure 4b. These data also suggest that a good CoMSIA model was achieved. The steric field descriptors explain 21.2% of the variance, while the electrostatic is 36.6%, and the hydrophobic field explains the rest, with 42.2%, suggesting that the hydrophobic interaction is more important than the electrostatic.

The steric and electrostatic fields of CoMSIA, as shown in Figure 6a and b, are generally in accordance with the field distribution of CoMFA maps. The hydrophobic analysis of CoMSIA was depicted in Figure 6c, where magenta and white colored regions represent hydrophobic and hydrophilic favorable to bioactivity. Precisely, the white colored region is the site where the hydrophilic side chains of Cys773 and Lys721 interact with the inhibitor, while the magenta color polyhedral is where the phenyl group of inhibitors interacts with the hydrophobic region I within the binding pocket (Supporting material, Fig. S1). The agreement of the CoMSIA contour with the structural topological properties of the EGFR PTK binding site demonstrates again the reasonability of the constructed CoMSIA model.

3.2.3. Validation of the 3D-QSAR models. In order to validate the QSAR models, we randomly selected inhibitor **6** as starting point, inhibitors with number of $10n + 6$ ($n = 1, 2, \ldots, 11$) were included in the testing set. Altogether, 12 inhibitors were selected as testing compounds, which were shown in Table 1 (labeled with asterisks) and Figure 4 (with triangle symbol). The predicted $-\log IC_{50}$ values by using our CoMFA and CoMSIA models are in good agreement with the observed data with a statistically tolerable error range (Table 1), $r = 0.966$ and 0.906 for CoMFA and CoMSIA models, respectively, demonstrating again that the CoMFA and CoMSIA models, which based on 112 inhibitors are reliable and have a good ability of prediction. From these results, the method that used in 3D-QSAR study are reliable and could be used for designing new inhibitors against EGFR.

4. Conclusions

In this study, molecular docking and 3D-QSAR methods were successfully combined to investigate the relationship between molecular bioactivities and their structures with observable geometrical diversity. The 124 best binding conformations of 124 EGFR PTK inhibitors, which belong to four different classes, were determined using docking. Based on the binding conformations and their alignments inside the binding pocket, highly reliable and predictive 3D-QSAR models were derived, which are consistent with the observed results and could be traced and mapped back to the original ligand–receptor binding model. This result also demonstrates that it is possible to construct a universal 3D-QSAR model using different types of inhibitors with observable structural diversity. Therefore, our new QSAR models not only lead to a better understanding of inhibition's mechanism, but also provide useful guidelines for designing new inhibitors of observable structural diversity.

Acknowledgements

This work was supported by grants from the State Key Program of Basic Research of China (2002CB512802), the 863 Hi-Tech Program (2001AA235051, 2001AA235071, 2001AA231111, 2002AA233011, and 2001AA235041), and the National Natural Science Foundation of China (29725203, 20072042, and 30070891).

References and notes

1. Aaronson, S. A. *Science* **1991**, *254*, 1146.
2. Alexander, J. B. *Chem. Rev.* **2001**, *101*, 2541.
3. Traxler, P.; Furet, P. *Pharmacol. Ther.* **1999**, *82*, 195.
4. Cance, W. G.; Liu, E. T. *Breast Cancer Res. Treat.* **1995**, *35*, 105.
5. Chrysogelos, S. A.; Dickson, R. B. *Breast Cancer Res. Treat.* **1994**, *29*, 29.
6. Yang, E. B.; Wang, D. F.; Cheng, L. Y.; Mack, P. *Cancer J.* **1997**, *10*, 319.
7. Yang, E. B.; Guo, Y. J.; Zhang, K.; Chen, Y. Z.; Mack, P. *Biochim. Biophys. Acta* **2001**, *1550*, 144.
8. Traxler, P.; Bold, G.; Buchdunger, E.; Caravatti, G.; Furet, P.; Manley, P.; O'Reilly, T.; Wood, J.; Zimmermann, J. *Med. Res. Rev.* **2001**, *21*, 499.

9. Traxler, P.; Bold, G.; Frei, J.; Lang, M.; Lydon, N.; Mett, H.; Buchdunger, E.; Meyer, E.; Mueller, M.; Furet, P. *J. Med. Chem.* **1997**, *40*, 3601–3616.
10. Khazaie, K.; Schirrmacher, V.; Lichtner, R. B. *Cancer Metast. Rev.* **1993**, *12*, 255.
11. Fry, D. W.; Kraker, A. J.; McMichael, A.; Ambroso, L. A.; Nelson, J. M.; Leopold, W. R.; Connors, R. W.; Bridges, A. J. *Science* **1994**, *265*, 1093.
12. Fry, D. W.; Bridges, A. J.; Denny, W. A.; Doherty, A.; Greis, K. D.; Hicks, J. L.; Hook, K. E.; Keller, P. R.; Leopold, W. R.; Loo, J. A.; McNamara, D. J.; Nelson, J. M.; Sherwood, V.; Smaill, J. B.; Trumpp-Kallmeyer, S.; Dobrusin, E. M. *Proc. Natl. Acad. Sci. U.S.A.* **1998**, *95*, 12022.
13. Rewcastle, G. W.; Murray, D. K.; Elliott, W. L.; Fry, D. W.; Howard, C. T.; Nelson, J. M.; Roberts, B. J.; Vincent, P. W.; Showalter, H. D.; Winters, R. T.; Denny, W. A. *J. Med. Chem.* **1998**, *41*, 742.
14. Smaill, J. B.; Palmer, B. D.; Rewcastle, G. W.; Denny, W. A.; McNamara, D. J.; Dobrusin, E. M.; Bridges, A. J.; Zhou, H.; Showalter, H. D.; Winters, R. T.; Leopold, W. R.; Fry, D. W.; Nelson, J.; Slintak, V.; Elliot, W. L.; Roberts, B. J.; Vincent, P. W.; Patmore, S. J. *J. Med. Chem.* **1999**, *42*, 1803.
15. Rewcastle, G. W.; Palmer, B. D.; Thompson, A. M.; Bridges, A. J.; Cody, D. R.; Zhou, H.; Fry, D. W.; McMichael, A.; Denny, W. A. *J. Med. Chem.* **1996**, *39*, 1823.
16. Wakeling, A. E.; Barker, A. J.; Davies, D. H.; Brown, D. S.; Green, L. R.; Cartlidge, S. A.; Woodburn, J. R. *Breast Cancer Res. Treat.* **1996**, *38*, 67.
17. Bridges, A. J. *Chem. Rev.* **2001**, *101*, 2573.
18. Hennequin, L. F.; Stokes, E. S. E.; Thomas, A. P.; Johnstone, C.; Ple, P. A.; Ogilvie, D. J.; Dukes, M.; Wedge, S. R.; Kendrew, J.; Curwen, J. O. *J. Med. Chem.* **2002**, *45*, 1300.
19. Traxler, P. M.; Furet, P.; Mett, H.; Buchdunger, E.; Meyer, T.; Lydon, N. *J. Med. Chem.* **1996**, *39*, 2285.
20. Thompson, A. M.; Murray, D. K.; Elliott, W. L.; Fry, D. W.; Nelson, J. A.; Showalter, H. D. H.; Roberts, B. J.; Vincent, P. W.; Denny, W. A. *J. Med. Chem.* **1997**, *40*, 3915.
21. Rewcastle, G. W.; Bridges, A. J.; Fry, D. W.; Rubin, J. R.; Denny, W. R. *J. Med. Chem.* **1997**, *40*, 1820.
22. Cramer, M.; Cramer, R. D., III; Jones, D. M. *J. Am. Chem. Soc.* **1988**, *110*, 5959.
23. Klebe, G.; Abraham, U.; Mietzner, T. *J. Med. Chem.* **1994**, *37*, 4130.
24. Peng, T.; Pei, J. F.; Zhou, J. J. *J. Chem. Inf. Comput. Sci.* **2003**, *43*, 298.
25. Zhu, L. L.; Hou, T. J.; Chen, L. R.; Xu, X. J. *J. Chem. Inf. Comput. Sci.* **2001**, *41*, 1032.
26. Kamath, S.; Abraham, U.; Buolamwini, J. K. *J. Med. Chem.* **2003**, *46*, 4657.
27. Sybyl Version 6.8, St Louis (MO), Tripos Associates, 2001.
28. Clark, M.; Cramer, R. D.; Opdenbosch, N. V. Validation of the General Purpose Tripose 5.2 Force Field. *J. Comput. Chem.* **1989**, *10*, 982.
29. Gasteiger, J.; Marsili, M. *Tetrahedron* **1980**, *36*, 3219.
30. Marsili, M.; Gasteriger, J. *Croat. Chem. Acta* **1980**, *53*, 601.
31. Purcell, W. P.; Singer, J. A. *J. Chem. Eng. Data* **1967**, *12*, 235.
32. Bernstein, F. C.; Koetzle, T. F.; Williams, G. J.; Meyer, E. E., Jr.; Brice, M. D.; Rodgers, J. R.; Kennard, O.; Shimanouchi, T.; Tasumi, M. *J. Mol. Biol.* **1977**, *112*, 535.
33. Stamos, J.; Sliwkowski, M. X.; Eigenbrot, C. *J. Biol. Chem.* **2002**, *277*, 46265.
34. Berman, H. M.; Westbrook, J.; Feng, Z.; Gilliland, G.; Bhat, T. N.; Weissig, H.; Shindyalov, I. N.; Bourne, P. E. The Protein Data Bank. *Nucl. Acids Res.* **2000**, *28*, 235.
35. Halgren, T. *J. Am. Chem. Soc.* **1990**, *112*, 4710.
36. Ewing, T. J. A.; Kuntz, I. D. *J. Comput. Chem.* **1997**, *18*, 1175.
37. Connolly, M. L. *Science* **1983**, *221*, 709.
38. Kuntz, I. D.; Blaney, J. M.; Oatley, S. J.; Langridge, R.; Ferrin, T. E. *J. Mol. Biol.* **1982**, *161*, 269.
39. Solis, F. J.; Wets, R. J. B. *Math. Oper. Res.* **1981**, *6*, 19.
40. Weiner, S. J.; Kollman, P. A.; Case, D. A.; Singh, C.; Ghio, G.; Alagona, S.; Profeta, P.; Weiner, P. *J. Am. Chem. Soc.* **1984**, *106*, 765.
41. Ghose, A. K.; Viswanadhan, V. N.; Wendoloski, J. J. *J. Comb. Chem.* **1999**, *1*, 55.
42. Palmer, B. D.; Trumpp-Kallmeyer, S.; Fry, D. W.; Nelson, J. M.; Showalter, H. D. H.; Denny, W. A. *J. Med. Chem.* **1997**, *40*, 1519.
43. Wallace, A. C.; Laskowski, R. A.; Thornton, J. M. *Protein Eng.* **1995**, *8*, 127.