

In silico identification of EGFR-T790M inhibitors with novel scaffolds: start with extraction of common features

Mingli Xiang^{1,*}

Kai Lei^{1,*}

Wenjie Fan¹

Yuchun Lin²

Gu He¹

Mingli Yang³

Lijuan Chen¹

Yirong Mo⁴

¹The State Key Laboratory of Biotherapy, West China Hospital, West China Medical School, Sichuan University, Chengdu, People's Republic of China; ²Department of Chemical and Biomolecular Engineering, University of California, Berkeley, CA, USA; ³Institute of Atomic and Molecular Physics, Sichuan University, Chengdu, People's Republic of China; ⁴Department of Chemistry, Western Michigan University, Kalamazoo, MI, USA

*These authors contributed equally to this work

Background: Epidermal growth factor receptor (EGFR) is an attractive therapeutic target for a number of human tumors including non-small cell lung cancer (NSCLC). Most patients with NSCLC and somatic mutations have shown a dramatic initial clinical response to reversible EGFR inhibitors. The clinical efficacy of reversible inhibitors is, however, ultimately limited due to the emergence of drug resistance, which is usually conferred by the EGFR T790M mutation. Importantly, irreversible, synthetic small molecule inhibitors are currently evaluated and some of them have been shown to overcome the acquired resistance that is oftentimes observed in these patients. Thus far, irreversible EGFR inhibitors as a drug class have not received regulatory approval due in part to their poor effectiveness at clinically achievable concentrations. Therefore, there is an urgent need to discover and develop novel, potent irreversible inhibitors against the EGFR T790M mutation.

Material and methods: In the following study, we report a novel "hybrid strategy" to identify irreversible EGFR inhibitors with active scaffolds starting with the identification and extraction of a common chemical reactive feature and a pharmacophore feature. The chemical reactive feature was elucidated by investigating 138 currently known irreversible inhibitors at B3LYP/6-31G(d) level using the density function theory method. The pharmacophore feature was extracted from the same inhibitors using pharmacophore modeling. Based on these unique features, two constraints were set while calibrating the protocols of in silico screening. Compounds bearing these specific features were obtained from the National Cancer Institute diversity database to form our subsequent library. Finally, a structure based virtual screening against the library was conducted using standard protocols validated in our lab.

Results: Twenty-eight candidate compounds that demonstrated antitumor activity and that had novel scaffolds different from commonly known quinazoline/quinoline analogs were obtained. The interaction modes between three representative candidates and our model system are similar to that between the model system and the reference compound T-001, which has previously been reported to be one of the most potent of the 138 irreversible inhibitors.

Conclusion: The hybrid strategy starting with the extraction of common features is an effective approach to design potential irreversible inhibitors with novel scaffolds and therefore to obtain lead molecules in the selection process. These candidates possessing unique scaffolds have a strong likelihood to act as further starting points in the preclinical development of potent irreversible T790M EGFR inhibitors.

Keywords: mutant EGFR, NCI database, virtual screening, drug resistant, quantum chemical calculation, pharmacophore modeling

Introduction

As key regulators of critical cellular processes, the ErbB protein family or epidermal growth factor receptor (EGFR) family has received much attention for several decades.¹⁻⁸

Correspondence: Mingli Xiang
The State Key Laboratory of Biotherapy, West China Hospital, West China Medical School, Sichuan University, China # 1 Keyuan Road 4, Gaopeng Street, High Technological Development Zone, Chengdu, 610041, People's Republic of China
Tel +86 28 8516 4063
Fax +86 28 8516 4060
Email xiang_mingli@scu.edu.cn

The human EGFR family consists of four members: EGFR [(Human Epidermal Growth Factor Receptor) HER1/ErbB1], HER2 (ErbB2), HER3 (ErbB3), and HER4 (ErbB4).^{2,3,9–11} They are structurally related receptor tyrosine kinases (RTKs) sharing a similar molecular architecture.^{3,10,12–14} Each of them comprises identical extracellular ligand-binding regions, a single hydrophobic transmembrane segment, and a cytoplasmic region. The extracellular region contains four sub-domains (I–IV)^{12–14} and the cytoplasmic region comprises a conserved protein tyrosine kinase (TK) catalytic domain as well as a carboxy terminal tail with tyrosine autophosphorylation sites.^{2,3}

It is well recognized that ErbB members share remarkable homology in their endocellular TK domains, but are distinct in their extracellular component and carboxy terminal tails.¹³ The ectodomain structure of ErbB2, for example, is radically different from the others.¹⁴ ErbB2 has a fixed conformation that resembles the ligand-activated state of EGFR and ErbB3. Within the extracellular region of ErbB2, a unique sub-domain I–III interaction buries the ligand binding site and makes the site not accessible for interaction.¹⁴ As such, ErbB2 lacks a ligand-binding domain to interact with a growth factor ligand. Although the intracellular TK domain of ErbB receptors is highly conserved, the kinase domain of ErbB3 has a substitution in critical amino acids, which results in no ErbB3 intrinsic kinase activity.^{3,13,15,16} ErbB2 and ErbB3 are non-autonomous TKs. They form heterodimeric complexes with other ErbBs that are capable of generating potent downstream signaling. In contrast, the other two members are autonomous. When bound to ligand growth factors, the receptor dimerization is induced and intracellular protein TK is activated with subsequent initiation of numerous downstream signaling events, which ultimately leads to cell proliferation, migration, and differentiation.^{3,13} Aberrant ErbB receptor activation and their intracellular signaling pathways and mutations in RTKs have been causally linked to cancers, diabetes, inflammation, severe bone disorders, arteriosclerosis, and angiogenesis.^{12,13,17}

Among all of the four members, EGFR was the first receptor protein-TK to be sequenced¹⁰ and linked directly to human tumors.^{3,14} Not only is it involved in development of numerous types of human cancers^{3,14,18} but it is even the host cofactor for hepatitis C virus entry as well.¹⁹ It has been intensely pursued as a therapeutic target.^{4,10,14,17,20} One of the most clinically advanced strategies toward aberrant ErbB receptor activation is small-molecule inhibition of the protein TK catalytic domain^{13,14,21–24} in which small-molecule TK inhibitors (TKIs) compete with adenosine triphosphate (ATP) in the TK domain. The first generation TKIs (erlotinib,

gefitinib,⁴ and lapatinib, for instance) are reversible small-molecule inhibitors. They prevent autophosphorylation of the EGFR TK by competing with ATP.²⁰ But cellular oncogenic transformation caused by different EGFR mutations results in differential sensitivity to the first generation TKIs.^{24,25} Although most patients with NSCLC and somatic mutations showed a dramatic initially clinical response to EGFR reversible inhibitors, they gradually developed acquired resistance to these drugs²⁶ due to the EGFR T790M mutation.^{4,6,7,16,21,24,27} This mutation leads to a change in the 3-dimensional structure of the TK domain and prevents reversible inhibitors from binding to EGFR.^{11,23,24} In this situation, various second generation EGFR TKIs are currently being evaluated. In contrast to the first generation TKIs that join to the catalytic site in the EGFR TK domain through classic competitive binding with ATP, second generation TKIs such as HKI-272 (neratinib),¹⁵ EKB-569 (pelitinib),²⁸ BIBW2992 (tovok),^{15,20,29} and PF00299804 (dacomitinib),^{15,20,30} which are shown in Figure 1, form covalent bonds to a conserved cysteine residue⁵ within the active site. The obvious advantage of these compounds over the first generation entities is the ability to prevent the development of acquired resistance or overcome acquired resistance in patients previously treated with the first generation of drugs.^{4,5,11,17,24,26,28,31} Although clinical trials with a number of such compounds have shown promising results, irreversible EGFR inhibitors as a drug class have not received regulatory approval.^{17,28} In fact, some of them did not demonstrate effectiveness to overcome this mutant at clinically achievable concentrations.^{16,25,31,32} It is clear that much work remains to be done on the road to develop second generation EGFR TKIs.

It was shown by our detailed review that the main group of currently known EGFR TKIs are almost all based on a quinazoline/quinoline scaffold.^{28,31,33–37} It is believed that the quinazoline/quinoline scaffold may not be the most potent or specific for inhibiting T790M EGFR.⁴ Therefore, to discover and develop novel, potent irreversible inhibitors against the EGFR T790M is quite important. Compounds with cores different from the quinazoline/quinoline scaffold and possessing inhibitory activity against T790M EGFR could represent promising starting points as lead compounds. Undoubtedly, 2nd generation inhibitors against T790M EGFR must have some common features and it is therefore important to find those compounds by revealing and utilizing these common features.

In this study, we report a novel hybrid strategy, which combines quantum chemical calculation, pharmacophore modeling, and structure-based drug design, to identify

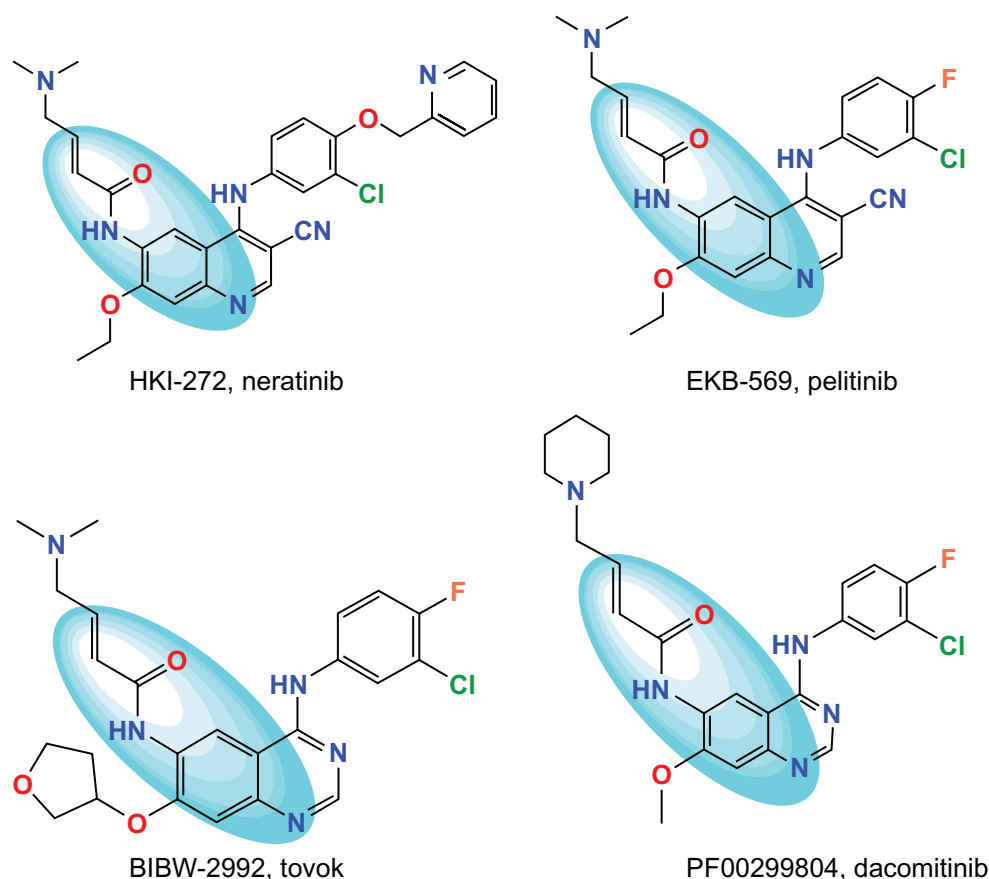


Figure 1 Structures of irreversible TK inhibitor.^{15,20,28–30} The colored ellipsoid marks the common features.

Abbreviation: TK, tyrosine kinase.

irreversible inhibitors with novel scaffolds. It started with the extraction of a common chemical reactive feature and a pharmacophore feature. Compounds bearing those specific common features were obtained from the National Cancer Institute (NCI) database (release 3 files) and were docked to the binding site of the model system with a calibrated protocol. Finally, 28 compounds that demonstrated antitumor activity and bear novel scaffolds different from commonly known quinazoline/quinoline analogs were identified. For three representative candidates, their binding modes to the model system were predicted.

Material and methods

As an attractive target, EGFR TK plays a crucial role in cancer therapy. Although several inhibitors of EGFR TK have been clinically validated for the treatment of patients with cancers during the past several years,²⁶ the search for novel and more potent irreversible inhibitors against this target is still meaningful and considerably challenging. Our aim is to identify irreversible inhibitors with novel scaffolds against T790M EGFR starting

with the identification and extraction of the common features from currently known irreversible inhibitors. The schematic representation chart of our workflow is shown in Figure 2.

Extraction of common features

In view of the properties of currently known irreversible inhibitors that can form covalent bonds with a conserved cysteine residue within the active site of EGFR, we presumed that there must exist some common features among them. Obviously, the features must at least include a chemical reactive feature and a pharmacophore feature. The former is related to the forming of covalent bonds and the latter is essential for a drug's biological activity.

Reactive site of irreversible inhibitors

At first, 138 TKIs (Table S1 in supplementary material) with biological activity values (IC_{50}) ranging from 0.002 to 13.570 μ M were selected from published documents.^{34–37} It was noted that almost all of them contain either an acrylamide group or a propargyl amide group.

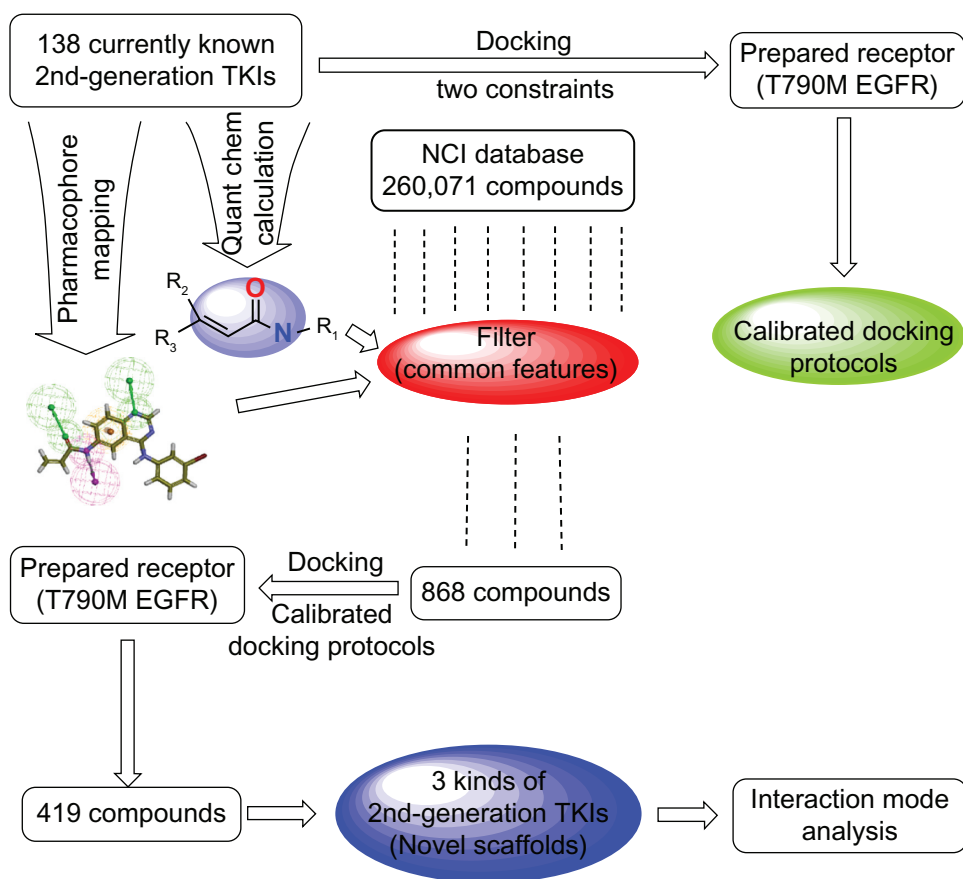


Figure 2 Schematic diagram of workflow for identifying potential EGFR-T790M inhibitors with novel scaffolds.

Abbreviations: EGFR, epidermal growth factor receptor; TKI, tyrosine kinase inhibitor; NCI, National Cancer Institute.

The exact location of Michael receptor in irreversible TKIs was determined by optimizing structures of 138 TKIs and calculating their reactive sites. The structures were drawn using MarvinSketch (ChemAxon Ltd., Budapest, Hungary).³⁸ Next, their geometries were obtained at the same operating environment by calculating the lowest energy conformer using a MMFF94 force field.³⁹ Later, these geometries were further optimized with density functional theory (DFT) method at B3LYP/6-31G(d) level by employing Gaussian 03.⁴⁰ Finally, the Natural Bond Orbital (NBO) charge distribution of molecules was calculated using the B3LYP/6-31G(d)//B3LYP/6-31G(d) method. For a molecule containing N electrons, independent calculations were made for the corresponding N -electron molecule and $(N + 1)$ -electron molecule with the same molecular geometry.⁴¹ The variation of NBO charges on an atom of an $(N + 1)$ -electron molecule relative to the N -electron molecule reflects the site reactivity of nucleophilic attack.⁴¹ It is obvious that the atom with the largest variation of NBO charge is most probably the reactive site of a molecule that involves a

Michael addition reaction with the sulfur atom of Cysteine (Cys)797 of EGFR.

Pharmacophore feature

For all of the 138 currently known 2nd generation TKIs, their common pharmacophore feature was extracted as follows.

At first, conformations of these 138 compounds were generated with Omega 2.2.5 (OpenEye Scientific Software Inc., Santa Fe, NM, USA)⁴² using the geometry optimized at B3LYP/6-31G(d) level with DFT method. Then, pharmacophore models were built by employing Discovery Studio (Accelrys Inc., San Diego, CA, USA)⁴³ with compound T-001 (Figure 3A),³⁷ a reference compound that is one of the most potent irreversible TKIs among these 138 compounds. By manually adjusting the number and types of the function groups and through mapping all conformations of these 138 TKIs onto the pharmacophore models, the common pharmacophore feature of these inhibitors was finally identified and extracted. As shown in Figure 3, the common pharmacophore feature of these 2nd generation inhibitors consists of one

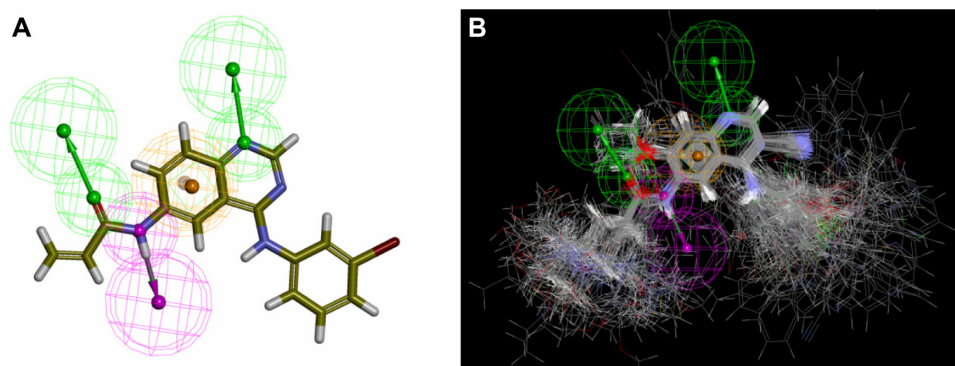


Figure 3 The common pharmacophore feature generated from compound T-001³⁷ (A) and 138 tyrosine kinase inhibitors (TKIs)^{34–37} superimposed on the feature (B). Aromatic hydrophobic center is represented by the orange wire mesh ball. Hydrogen bond acceptor is depicted with the green arrow and wire mesh balls while the hydrogen bond donor is described by the pink arrow and wire mesh balls.

hydrogen bond donor, one aromatic hydrophobic center, and two hydrogen bond acceptors.

After that, another 21 irreversible inhibitors^{4,34–37} coupled with two reversible inhibitors⁴ (Table S2 in supplementary materials) were used to test whether these extracted features are effective enough to identify potential irreversible inhibitors.

Docking with constraints

As a reliable, cost-effective, and time-saving technique, structure based virtual screening (SBVS) has shown a powerful ability to discover lead compounds.³² In the SBVS pipeline, preparation of receptor, determination of docking protocols and preparation of a compound library are the committed steps.

Preparation of receptor

TK cSrc was chosen as the model system of the mutant EGFR and its 3-dimensional structure was obtained from the Protein Data Bank (PDB ID: 2QQ7).⁴⁴ Water molecules in the crystallographic complex were stripped. Preparation of receptor for docking was done using FRED RECEPTOR 2.2.5 (OpenEye Scientific Software Inc., Santa Fe, NM, USA).⁴⁵ Four hundred different molecule sensors which were stored internally by the program were used to describe the binding site for docking. The size of the box is 6,140 cubic Angstroms. The inner contour and outer contour filters, which were detected with 400 sensor molecules, are 92 cubic and 1,442 cubic angstroms, respectively.

Identification of docking protocols

As many as 252 docking strategies were tested to find a satisfactory way upon which TKIs with potent activities could be enriched at the top of high-score ranking.

At first, conformers of the 138 TKIs were created via Omega 2.2.5⁴² and the number of conformers for each TKI

was limited up to 2,000. Then, in total, 252 schedules were tested for obtaining optimal docking strategy. At the conformer searching stage, four scoring functions including Shapegauss,⁴⁶ Plp,⁴⁷ Chemgauss2,⁴⁸ Chemgauss3⁴⁸ were chosen for exhaustive searching of conformers. At the optimization stage, seven scoring functions (Chemgauss2,⁴⁸ Chemgauss3,⁴⁸ Chemscore,⁴⁸ Shapegauss,⁴⁶ Plp,⁴⁷ OEChemscore, and Screenscore⁴⁹) were hired for rigid optimization of poses. For each one of the above 28 combinations, nine scoring functions (those seven used in rigid optimization plus Zapbind⁵⁰ and consensus score) were used for evaluating the interaction between the TKIs and the active site of EGFR. Scoring functions employed for identifying optimal docking protocols are listed in Table 1.

In the process of docking currently known irreversible TKIs, two constraints were added. The first one was that each potential irreversible TKI should provide a hydrogen acceptor to form a hydrogen bond with the hinge region M793. There is strong evidence that inhibitors make a hydrogen

Table 1 Scoring functions employed for identifying optimal docking protocols at different stages*

Exhaustive search	Rigid optimization	Interaction evaluation
Shapegauss ⁴⁶	Shapegauss ⁴⁶	Shapegauss⁴⁶
Plp ⁴⁷	Plp ⁴⁷	Plp ⁴⁷
Chemgauss2⁴⁸	Chemgauss2 ⁴⁸	Chemgauss2 ⁴⁸
Chemgauss3 ⁴⁸	Chemgauss3 ⁴⁸	Chemgauss3 ⁴⁸
	Chemscore ⁴⁸	Chemscore ⁴⁸
	OEChemscore	OEChemscore
	Screenscore⁴⁹	Screenscore ⁴⁹
		Zapbind ⁵⁰
		Consensus score

Notes: *In total, 252 docking strategies were gained using the combinations of scoring functions at different stages. Scoring functions employed in the optimal docking protocols are marked with bold in the table.

Abbreviation: Plp, linear pairwise potential.

bond to the hinge region atom N-H in M793 in the protein kinase family.^{4,5,8,17,28,51,52} The other constraint was to require an alkyne carbon or an alkene carbon being within 4 ± 1 angstrom range of the sulfur atom of the conserved Cys797 residue, which would allow for covalent bond formation.⁵³ This is supported by our quantum chemical calculations that demonstrated that the reactive site of the known irreversible TKIs locates at the β -carbon of the acrylamide or propargyl amide group.

Preparation of the molecule library

As described above, common features of currently known irreversible TKIs were extracted (Figures 1–3). Commonly, each of those TKIs has an acrylamide group or a propargyl amide group. At the same time, they all bear one hydrogen bond donor, one aromatic hydrophobic center, and two hydrogen bond acceptors. We expected that potential irreversible TKIs with novel scaffolds would have such features. These features were used as a filter (Figure 2). Compounds in the NCI database (260,071) were filtered using these features. Those that do not meet these features were eliminated. After filtering, only 868 compounds in the NCI database were remained. Then, reference compound T-001 and another three irreversible inhibitors⁴ with novel scaffolds mixed into these molecules. Finally, conformations of these compounds were generated using Omega 2.2.5⁴² and the maximum number of conformations was set to 2,000.

Constraint virtual screening and interaction mode analysis

With the optimal docking protocols mentioned above and two added constraints, 868 compounds, from the NCI database and bearing the common features, were docked to the binding site of a model system using FRED program version 2.2.5.⁴⁵ Potential irreversible TKIs were obtained from the top 200 of the SBVS result. For the reference compound T-001³⁷ and the screened potential irreversible TKIs with representative scaffolds, the interaction between them and the model system was analyzed and visualized using DS visualizer 3.1 (Accelrys Inc., San Diego, CA, USA).⁴³

Results and discussion

Common features

The irreversible TKIs can form covalent bonds with a conserved cysteine residue (Cys797) within the active site of T790M EGFR. Apparently, as shown in Figure 1 and

Table S1 in supplementary material, currently known 2nd generation TKIs^{34–37} contain either an acrylamide group or a propargyl amide group.

Quantum chemical calculations revealed that the reactive site of the known irreversible TKIs, which form a covalent bond to the conserved Cys797 within T790M EGFR, locates at the β -carbon of the acrylamide or propargyl amide group (Figure 4 and Table 2). This is the theoretical basis for us to set one of the constraints while conducting SBVS.

It is interesting to note that often the β -carbon of acrylamide group or propargyl amide group in the currently known irreversible TKIs is the most electrophilic, as shown in Figure 4A and D. While results of NBO analysis show that the β -carbon is no longer the most electrophilic atom when an electron-withdrawing atom directly links to the β -carbon (Figure 4B) or the nitrogen atom at position three of the quinazoline group is replaced by a carbon linked with a cyano group (Figure 4C). The β -carbon becomes one of the most electrophilic atoms in the molecule. Even though it still keeps the tendency to attack the nucleophilic thiol group of the conserved Cys797 in T790M EGFR, the bioactivity of the molecule is much less than its analog (Figure 4E, 4F and 4G).

Except for the chemical reactive feature mentioned above, the currently known irreversible TKIs have a common pharmacophore feature that was extracted previously. As shown in Figure 3, the common pharmacophore feature can be described to include one hydrogen bond donor, one aromatic hydrophobic center, and two hydrogen bond acceptors. It is known that pharmacophore refers to the 3-dimensional arrangement of functional groups essential to a drug's biological activity.^{54,55} A diverse set of lead compounds, which potentially possess desired biological activity but have totally different chemical scaffolds, often share a common pharmacophore.⁵⁶

In a sense, compounds sharing these common chemical reactive and pharmacophore features could be guaranteed to form covalent bonds to the conserved Cys797. At the same time, they possess the desired biological activity and have different chemical scaffolds as well.

Selection of model system

There are many 3-dimensional structures of T790M EGFR determined using X-ray crystallography. We choose TK cSrc (PDB ID: 2QQ7)⁴⁴ as the model system of the mutant EGFR based on the following considerations. At first, the resolution of 2QQ7 is 2.38 angstrom and the crystal structure is

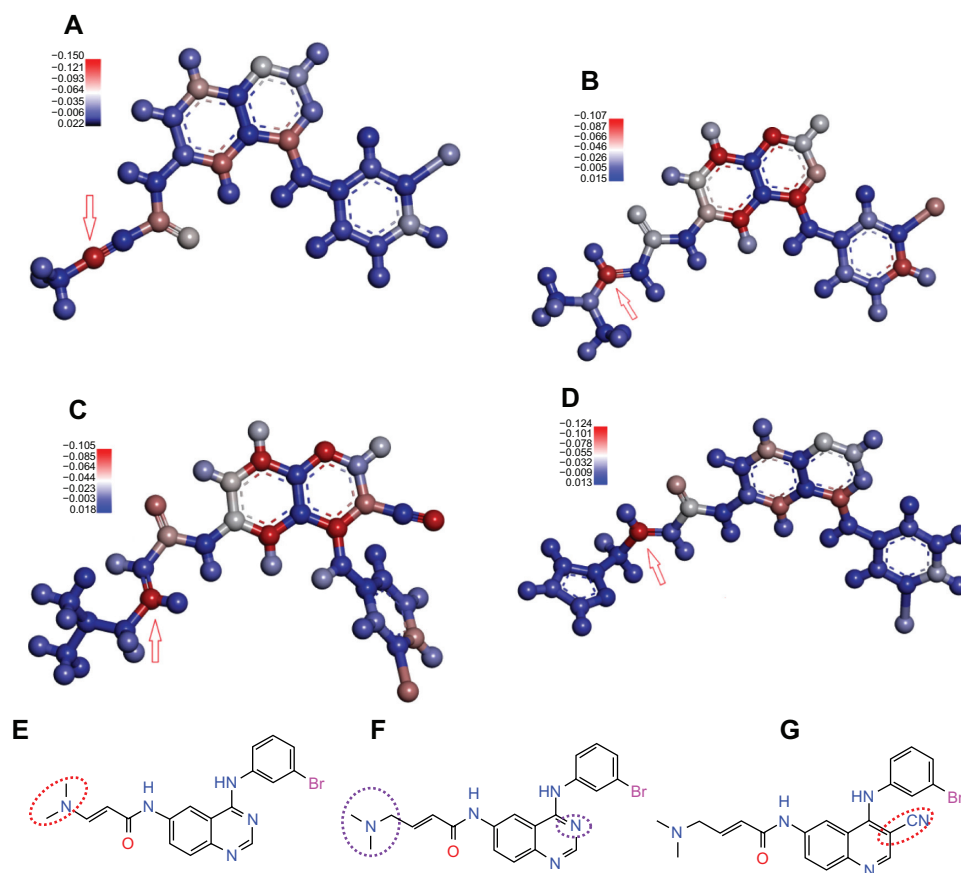


Figure 4 The electrophilic reactive site of some known irreversible TKIs^{34–37} calculated at B3LYP/6-31G(d) level with DFT method by employing Gaussian 03⁴⁰ and the effect of replacement on the bioactivity of irreversible TKI. Electrophilic center was highlighted by the red color and the nucleophilic center was painted with a blue color. Red arrow points to the β -carbon of the acrylamide group or propargyl amide group.

Notes: T-005, 0.008 μ M (A); T-084, 0.250 μ M (B); T-092, 0.310 μ M (C); T-125, 1.124 μ M (D); T-084, 0.250 μ M (E); T-008, 0.011 μ M (F); T-092, 0.310 μ M (G).

Abbreviations: DFT, density functional theory; TKI, tyrosine kinase inhibitor.

complexed with an irreversible inhibitor. It was believed that when the resolution value of crystallographic data is lower than 2.5 angstrom the modeling of details in the protein structures is much more subjective than desired.^{57,58} It was also confirmed that selecting structures with poor resolution may produce false predicted conformations.⁵⁹ Thus, as a rule of thumb, data worse than 2.5 angstrom resolution should not be considered reliable for virtual library screening applications.⁶⁰

In addition, cSrc is a tractable target for the structural investigation of irreversible EGFR. It has many advantages⁴⁴ for the study of homologous protein kinases over EGFR. The kinase domains between EGFR and cSrc share high identity⁴⁴ in structure and sequence. The sequence identity between their ATP binding sites is as high as 70%. Most importantly, cSrc has been validated as a model system for the structural investigation of irreversible EGFR inhibitors.³³

Constrained docking against the NCI database

Having a proper pose within the binding site is indispensable for ligand being effective. In addition, adding constraints is an effective way to gain proper poses of ligands. Therefore, two constraints, which are discussed above, were added while docking molecules to the binding site of the model system.

With those two constraints, 138 known irreversible TKIs were docked to the binding site of the model system. Finally, an optimal strategy that provides the highest enrichment rate was identified from among 252 combinations. Scoring functions Chemgauss2,⁴⁸ Screenscore,⁴⁹ and Shapegauss⁴⁶ were used for exhaustive searching of conformers, rigid optimization of poses, and interaction evaluation, respectively. According to this strategy, five potent TKIs were enriched within the top six of the 138 compounds. These results are presented in Table 3.

Table 2 NBO charge distribution and the relevant change of charge distribution in the N-electron and (N + 1)-electron molecule for N electrons systems T-005 and T-008

No	T-005				T-008			
	Element	q(N)	q(N + 1)	Δq	Element	q(N)	q(N + 1)	Δq
1	C	0.153	0.133	-0.021	C	0.166	0.141	-0.025
2	C	-0.239	-0.257	-0.017	C	-0.232	-0.246	-0.013
3	C	-0.203	-0.278	-0.074	C	-0.200	-0.268	-0.068
4	C	0.173	0.179	0.006	C	0.172	0.176	0.004
5	C	-0.130	-0.115	0.015	C	-0.133	-0.119	0.014
6	C	-0.240	-0.333	-0.093	C	-0.248	-0.329	-0.081
7	N	-0.497	-0.554	-0.057	N	-0.491	-0.547	-0.056
8	C	0.262	0.218	-0.044	C	0.260	0.215	-0.044
9	N	-0.560	-0.591	-0.030	N	-0.559	-0.587	-0.027
10	C	0.466	0.376	-0.090	C	0.462	0.373	-0.089
11	N	-0.568	-0.577	-0.009	N	-0.571	-0.578	-0.007
12	C	0.171	0.174	0.003	C	0.173	0.170	-0.003
13	C	-0.274	-0.294	-0.020	C	-0.273	-0.290	-0.016
14	C	-0.090	-0.088	0.002	C	-0.091	-0.090	0.001
15	C	-0.267	-0.315	-0.049	C	-0.267	-0.313	-0.046
16	C	-0.212	-0.226	-0.014	C	-0.212	-0.224	-0.012
17	C	-0.265	-0.279	-0.014	C	-0.268	-0.285	-0.017
18	Br	0.051	0.008	-0.043	Br	0.055	0.016	-0.039
19	N	-0.605	-0.597	0.008	N	-0.614	-0.606	0.008
20	C	0.635	0.552	-0.083	C	0.667	0.617	-0.050
21	C	-0.139	-0.117	0.022	C	-0.340	-0.353	-0.013
22	O	-0.586	-0.649	-0.063	O	-0.615	-0.675	-0.060
23	C	0.073	-0.045	-0.118	C	-0.119	-0.234	-0.116
24	C	-0.750	-0.730	0.020	C	-0.290	-0.279	0.011
25	H	0.235	0.208	-0.028	N	-0.490	-0.488	0.002
26	H	0.256	0.227	-0.029	C	-0.471	-0.467	0.004
27	H	0.256	0.241	-0.015	C	-0.477	-0.472	0.004
28	H	0.217	0.186	-0.031	H	0.280	0.265	-0.016
29	H	0.423	0.419	-0.004	H	0.257	0.231	-0.026
30	H	0.290	0.289	0.000	H	0.208	0.186	-0.022
31	H	0.253	0.235	-0.019	H	0.217	0.189	-0.028
32	H	0.244	0.227	-0.017	H	0.413	0.408	-0.004
33	H	0.235	0.229	-0.006	H	0.291	0.291	0.000
34	H	0.429	0.410	-0.018	H	0.254	0.237	-0.017
35	H	0.270	0.249	-0.021	H	0.244	0.228	-0.016
36	H	0.266	0.243	-0.023	H	0.227	0.220	-0.007
37	H	0.269	0.242	-0.027	H	0.414	0.400	-0.014
38					H	0.218	0.198	-0.020
39					H	0.246	0.230	-0.016
40					H	0.212	0.190	-0.023
41					H	0.233	0.222	-0.011
42					H	0.230	0.218	-0.012
43					H	0.190	0.179	-0.011
44					H	0.225	0.221	-0.004
45					H	0.231	0.217	-0.014
46					H	0.228	0.234	0.006
47					H	0.190	0.179	-0.011

Note: β -carbon atoms are in bold. Data obtained from Tsou HR, Mamuya N, Johnson BD, et al. 6-Substituted-4-(3-bromophenylamino)quinazolines as putative irreversible inhibitors of the epidermal growth factor receptor (EGFR) and human epidermal growth factor receptor (HER-2) tyrosine kinases with enhanced antitumor activity. *J Med Chem.* 2001;44(17):2719–2734.³⁷

Abbreviations: NBO, natural bond orbital; q, electric charge.

Together with reference compound T-001 and another three irreversible inhibitors,⁴ 868 molecules retrieved from the NCI database containing 260,071 compounds were docked to the binding site of the model system according to the optimal docking strategy. Of these, 419 compounds successfully docked to the site.

By manually analyzing compounds among the top 200 compounds, 28 of them are found to have novel scaffolds different from the currently known quinoline/quinazoline core and demonstrated bioactivity against various types of tumors. The average antitumor activities (GI_{50}) of these 28 compounds are in the order of magnitude of 10^{-5} μ g/mL (Table 4).

Table 3 The IC_{50} value and Shapegauss⁴⁶ score of five potent TKIs^{34–37} enriched within the top six of 138 second-generation TKIs

TKIs	IC_{50} ($\mu\text{m} \cdot \text{L}^{-1}$)	Shapegauss score
T-013	0.016	–629.309
T-016	0.020	–606.586
T-019	0.024	–602.346
T-106	0.360	–601.050
T-018	0.024	–598.813
T-009	0.011	–596.112

Abbreviations: IC_{50} , half maximal inhibitory concentration; TKI, tyrosine kinase inhibitor.

Validation of the hybrid strategy

The committed step in the pipeline of SBVS is to validate docking strategy.⁶¹ In order to test whether the extracted common features could be used to identify potential irreversible inhibitors, 23 molecules were collected to work as a testing set. This group comprised compound T-001, another 17 irreversible inhibitors^{34–37} with the quinoline scaffold, three irreversible inhibitors⁴ with a core different from quinoline, and two reversible inhibitors (gefitinib and erlotinib).⁴ As expected, the common features worked as a filter: reversible inhibitors were filtered out due to the fact that they lack the chemical

reactive feature and only irreversible inhibitors with the quinoline scaffold or other scaffolds remained. As shown in Figure 5, they could be mapped onto the pharmacophore feature. It was demonstrated that the extracted common features could work as an effective filter to identify irreversible inhibitors whether they contain the quinoline scaffold or not.

Before conducting a SBVS, four reference compounds (compound T-001,³⁷ WZ-3146,⁴ WZ-4002,⁴ and WZ-8040⁴) with a scaffold different from quinoline/quinazoline were mixed into a molecule library that includes 868 compounds retrieved from the NCI database with the aid of a common feature filter. According to the optimal docking protocols, these molecules were docked to the binding site of the model system. As demonstrated in Table 4, the scorings of reference compound T-001 and the other three irreversible inhibitors with a scaffold different from quinoline/quinazoline are among the top 150 high score ranking results. These four irreversible inhibitors and the 28 potential inhibitors could be mapped onto the extracted pharmacophore feature (Figures 5 and 6). The results show that the hybrid strategy is an effective way to identify potent irreversible inhibitors bearing novel scaffolds different from quinoline/quinazoline analogs.

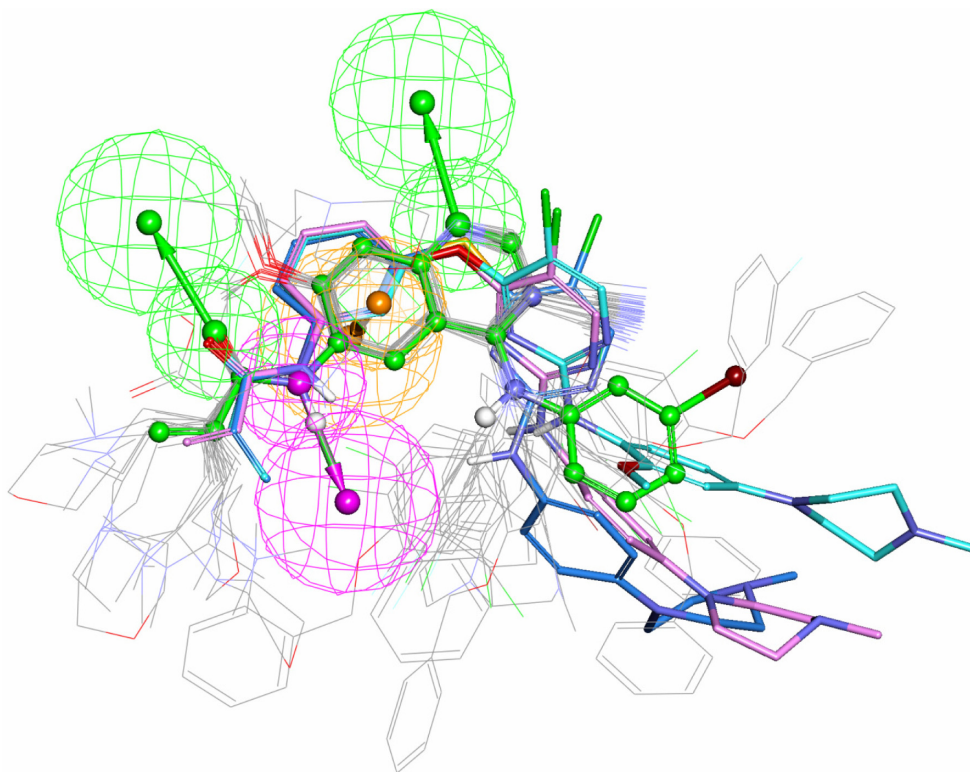


Figure 5 In total, 21 irreversible TKIs^{34–37} in the testing set (refer to Table S2 in supplementary materials) were mapped on the common pharmacophore feature generated from compound T-001.³⁷ T-001 is depicted by the stick and its carbon atoms are colored in green. WZ-3146⁴ is depicted by the stick and its carbon atoms are colored pink; WZ-4002⁴ is depicted by the stick and its carbon atoms are colored sky blue; WZ-8040⁴ is depicted by the stick and its carbon atoms are colored blue.

Abbreviation: TKI, tyrosine kinase inhibitor.

Table 4 The Shapegauss⁴⁶ score and antitumor activities of the 28 potential irreversible TKIs plus four reference irreversible TKIs^{4,37}

Serial number	NSC number	Shapegauss score	Position among the top 200	Average GI ₅₀ * (μg · mL ⁻¹)	Classification
1	621158	-500.858	42	9.85×10^{-5}	Class A
2	623897	-494.190	46	9.80×10^{-5}	Class A
3	630912	-472.511	65	2.14×10^{-5}	Class A
4	637681	-469.007	72	9.43×10^{-5}	Class A
5	633920	-462.811	81	7.63×10^{-5}	Class A
6	637635	-462.406	84	8.46×10^{-5}	Class A
7	637606	-459.648	89	7.61×10^{-5}	Class A
8	637638	-458.330	91	5.54×10^{-6}	Class A
9	630913	-456.322	93	6.49×10^{-5}	Class A
10	623881	-452.136	98	7.35×10^{-5}	Class A
11	630908	-452.072	99	6.60×10^{-5}	Class A
12	637630	-451.936	100	5.67×10^{-5}	Class A
13	637698	-450.357	104	4.25×10^{-5}	Class A
14	623867	-450.046	106	8.94×10^{-5}	Class A
15	642576	-443.251	111	2.17×10^{-5}	Class B
16	25603	-440.210	114	1.13×10^{-5}	Class B
17	637614	-436.258	118	1.13×10^{-5}	Class A
18	623882	-430.112	122	2.55×10^{-5}	Class A
19	637605	-429.136	124	1.18×10^{-5}	Class A
20	637626	-428.069	125	8.06×10^{-5}	Class A
21	111135	-426.822	126	2.01×10^{-5}	Class B
22	623877	-421.301	134	1.97×10^{-5}	Class A
23	623879	-415.805	142	1.09×10^{-5}	Class A
24	637706	-409.342	152	5.50×10^{-5}	Class A
25	637700	-408.819	154	9.56×10^{-5}	Class A
26	636097	-406.973	160	7.16×10^{-5}	Class A
27	623875	-400.459	170	1.54×10^{-5}	Class A
28	57443	-381.532	198	5.46×10^{-5}	Class C
29	T-001 ³⁷	-419.784	137	0.002**	
30	WZ3146 ⁴	-471.704	66	0.029**	
31	WZ4002 ⁴	-526.444	21	0.047**	
32	WZ8040 ⁴	-469.951	69	0.009**	

Notes: *The average GI₅₀ overall cell lines from National Cancer Institute (NCI) Cancer Screening Results, December 2010; **in μM/L.

Abbreviations: TKI, tyrosine kinase inhibitor; GI₅₀, the concentration required to achieve 50 % growth inhibition; NSC, Nomenclature Standards Committee.

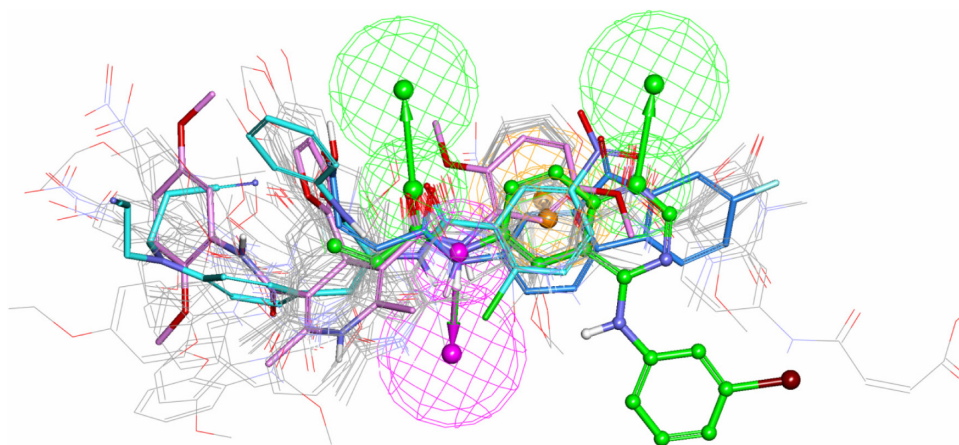


Figure 6 Potential irreversible TKIs that had been identified from the NCI database (refer to Table 4) were mapped onto the common pharmacophore feature generated from reference compound T-001.³⁷ T-001 is depicted by the stick and its carbon atoms are colored green. Representative compounds NSC-621158, NSC-642576, and NSC-57443 are depicted by the stick and their carbon atoms are colored pink, sky blue, and blue, respectively.

Abbreviations: NCI, National Cancer Institute; NSC, Nomenclature Standards Committee; TKI, tyrosine kinase inhibitor.

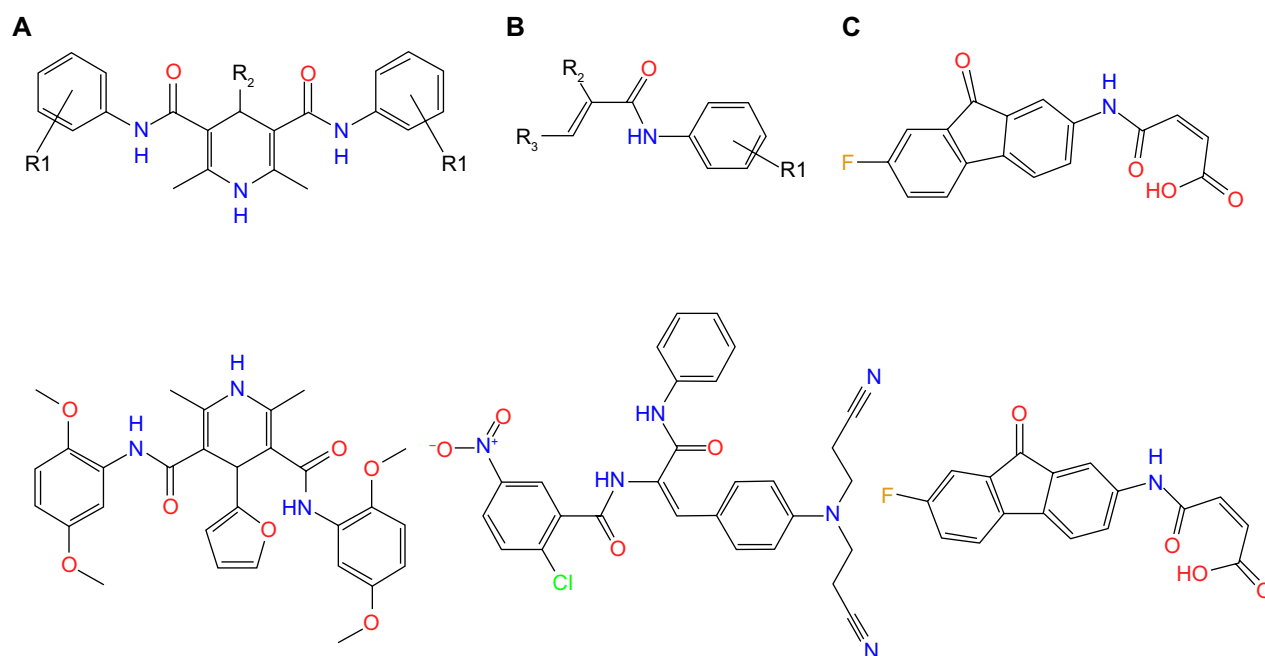


Figure 7 Three classes of potential irreversible TKIs with novel scaffolds and their corresponding representative molecules.

Notes: Class A: NSC number 621158 (**A**); Class B: NSC number 642576 (**B**); Class C: NSC number 57443 (**C**).

Abbreviations: NSC, Nomenclature Standards Committee; TKI, tyrosine kinase inhibitor.

Interaction mode analysis

It is easy to identify that these 28 compounds belong to three types of scaffolds (Figure 7 and Table S3 in supplementary materials). Class A is a group of symmetric compounds. There are two symmetric acrylamide functional groups in this class of compound and the alkene moieties are parts of the dimethyl-1,4-dihydropyridine group. Class B is asymmetric. An aromatic ring is directly linked to the N atom of the acrylamide functional group in both class A and class B, while in class C, it is a fluorenone group that directly binds to the N atom of the acrylamide.

From each class, we selected one representative molecule that had the highest score among its own class. Together with reference compound T-001,³⁷ the interaction modes between them and the mutant EGFR were analyzed with the aid of Discovery Studio.⁴³

According to a pharmacophore model^{27,51} of the ATP binding site in protein kinases, the conserved site is divided into five regions: (1) adenine region, which contains two amino acid residues, Glutamine (Gln)791 and Methionine (Met)793 on the hinge region corresponding to Glutamic acid (Glu)339 and Met341 in TK cSrc; (2) sugar pocket, which is the region extending from Cys797 (corresponding to Cys345 in TK cSrc) toward the solvent; (3) hydrophobic region I, which is composed of two residues Threonine (Thr)790 and Thr854, corresponding to Met338 and Alanine (Ala)403 in TK cSrc; (4) hydrophobic region II, which is formed by residues

Leucine (Leu)718 and Glycine (Gly)796, corresponding to Leu273 and Gly344 in tyrosine kinase cSrc; and (5) phosphate binding region, which is largely solvent exposed. The ATP binding site in TK cSrc is shown in Figure 8.

As shown in Figure 9, the four selected molecules similarly form hydrogen bonds with the N-H of residue Met341 and the β -carbon of acrylamide is near the S atom of Cys345. Surely, all of them have an interaction with the sugar region of the binding site. In hydrophobic region II, they interact with Leu273 and Gly344, while in the adenine region they only

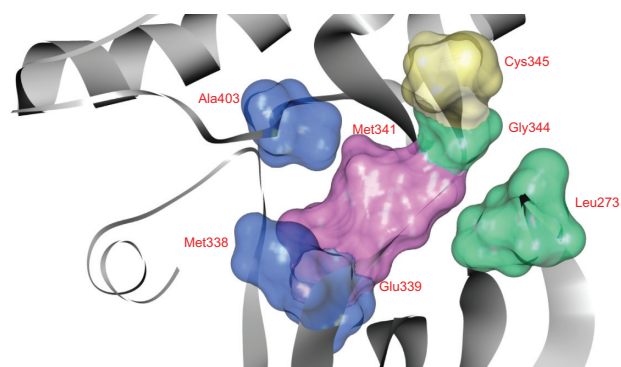


Figure 8 Pharmacophore model^{27,51} of the ATP binding site in tyrosine kinase cSrc (PDB ID: 2QQ7).⁴⁴ This site is represented by the solid surface. Adenine region, sugar region, hydrophobic region I, and hydrophobic region II are colored pink, yellow, blue, and green, respectively.

Abbreviations: ATP, adenosine triphosphate; PDB, protein data bank; Ala, alanine; Gln, glutamine; Glu, glutamic acid; Met, methionine; Thr, threonine; Leu, leucine; Gly, glycine; Cys, cysteine

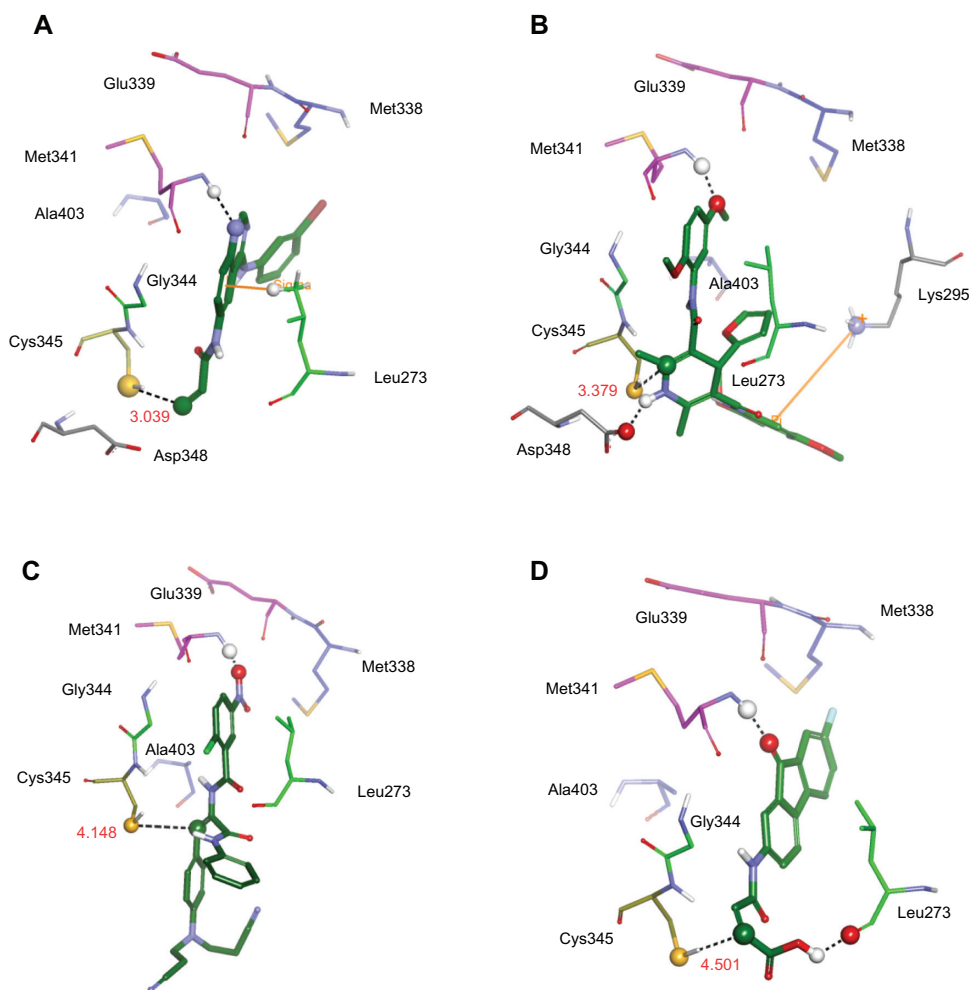


Figure 9 The illustrations showing the interaction between the model system and three representative compounds as well as between the reference compound T-001³⁷ and the model system. Residues are displayed in thin stick style. Carbon atoms of residues corresponding to adenine region, sugar region, hydrophobic region I, and hydrophobic region II are colored pink, yellow, blue, and light green, respectively. Compounds are represented by a thick stick and their carbon atoms are colored green. Interactions such as hydrogen bond, cation- π interaction, and π - σ interaction are highlighted.

Notes: Compounds in (A–D) correspond to T-001,³⁷ NSC-621158, NSC-642576, and NSC-57443, respectively.

Abbreviations: Ala, alanine; Gln, glutamine; Glu, glutamic acid; Met, methionine; Thr, threonine; Leu, leucine; Gly, glycine; Cys, cysteine.

interact with Met341; Glu339 is far from them. Conversely, reference compound T-001,³⁷ compound A (NSC-621158), and compound B (NSC-642576) do not interact with residues in the hydrophobic region I. In contrast, compound C (NSC-57443) interacts with Met338 in this region (Figure 9D).

Specifically, it was noted that there is a π - σ interaction between T-001³⁷ and the side chain of Leu273 (Figure 9A). For compound A, a hydrogen bond interaction with the side chain of Asp348 is observed. Meanwhile, a cation- π interaction is noted between this compound and the side chain of Lys295 (Figure 9B). The hydrogen bond formed between compound B and N-H of Met341 is quite strong. The distance between the heavy atoms with high electro-negativity values is only 2.254 angstrom. The 3,3'-(phenylazanediy) dipropanenitrile moiety of compound B extends to solvent (Figure 9C). Additional interaction between compound C

and the hydrophobic region II was also observed. That is, a hydrogen bond forms between compound C and the carbonyl oxygen of Leu273 (Figure 9D).

Conclusion

In this work, we were trying to find irreversible TKIs with novel scaffolds. At first, quantum chemical calculations and pharmacophore modeling were employed to extract common features of 138 currently known irreversible TKIs. The β -carbon of acrylamide group or propargyl amide group was identified by quantum chemical calculations as the electrophilic reactive site or one of the sites in irreversible TKIs. Commonly, each such TKI bears one aromatic hydrophobic center, one hydrogen donor, and two hydrogen acceptors.

Next, optimal docking protocols were identified and validated. By adding two constraints and docking 138

currently known irreversible TKIs to the binding site of a model system, optimal docking protocols were identified from selecting the best one among 252 types of docking strategies. According to this optimal strategy, five potent irreversible TKIs were enriched within the top six of the 138 TKIs.

From the NCI diversity database (260,071 compounds), 868 molecules were retrieved with the aforementioned common features. These molecules plus another four irreversible inhibitors were then docked to the binding site of a model system by employing optimal docking protocols. Finally, 28 potential irreversible TKIs, which belong to three classes of novel scaffolds and exhibited antitumor activity, were gained from the top 200 with high score rankings. The interaction modes between three representative candidates and the model system are similar to that between the model system and the reference compound T-001, one of the most potent inhibitors among the identified 138 irreversible inhibitors. Still, there are subtle differences in their interactions with the model system, which contribute to their different inhibition activities.

The interaction scores of the four irreversible inhibitors are among the top 150 with high score rankings. Average antitumor activities (GI_{50}) of the 28 compounds are in the order of magnitude of $10^{-5} \mu\text{g} \cdot \text{mL}^{-1}$. The hybrid strategy starting with the extraction of common features is shown to be an effective approach to identify irreversible inhibitors with novel scaffolds. Candidates possessing unique scaffolds have a strong likelihood to act as further starting points in the preclinical development of potent irreversible inhibitors against T790M EGFR.

Acknowledgments

This work was supported by the National Key Programs of China during the 12th Five-Year Plan Period (Grant 2012ZX09103-101-017). We would like to acknowledge the OpenEye Scientific Software and ChemAxon for their academic licenses. The authors would like to acknowledge Dr Carl K Edwards for his review of this manuscript.

Disclosure

The authors report no conflicts of interest in this work.

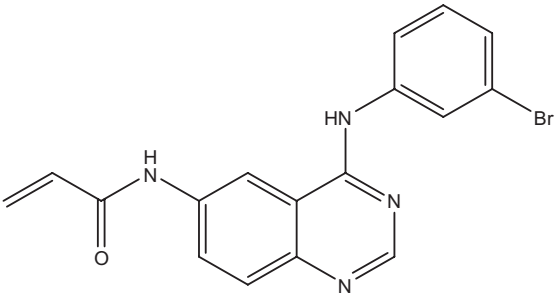
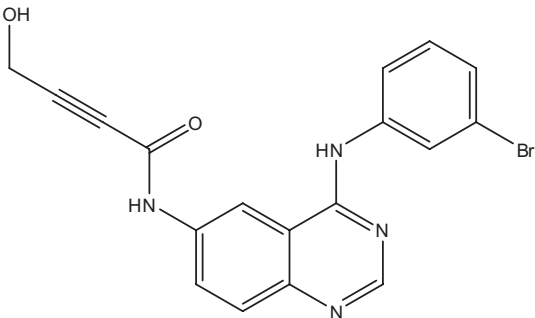
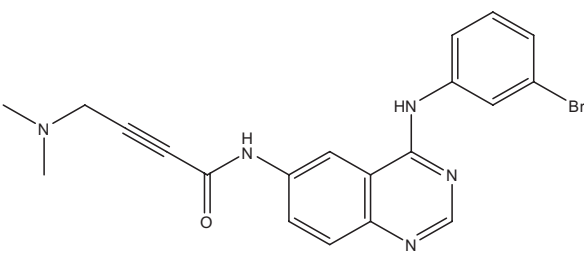
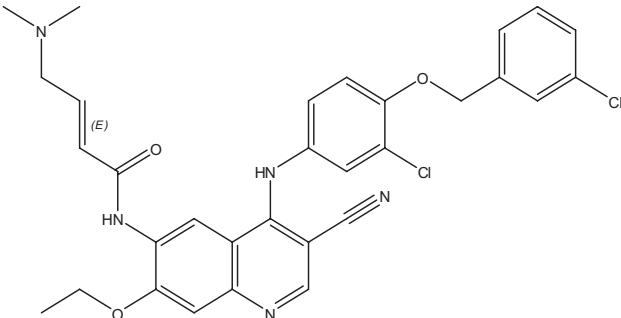
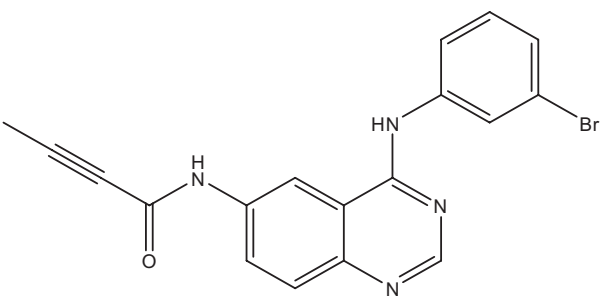
References

- Shimizu N, Behzadian MA, Shimizu Y. Genetics of cell surface receptors for bioactive polypeptides: binding of epidermal growth factor is associated with the presence of human chromosome 7 in human-mouse cell hybrids. *Proc Natl Acad Sci U S A*. 1980;77(6):3600–3604.
- Schlessinger J. Cell signaling by receptor tyrosine kinases. *Cell*. 2000;103(2):211–225.
- Yarden Y, Sliwkowski MX. Untangling the ErbB signalling network. *Nat Rev Mol Cell Biol*. 2001;2(2):127–137.
- Zhou W, Ercan D, Chen L, et al. Novel mutant-selective EGFR kinase inhibitors against EGFR T790M. *Nature*. 2009;462(7276):1070–1074.
- Carmi C, Cavazzoni A, Vezzosi S, et al. Novel irreversible epidermal growth factor receptor inhibitors by chemical modulation of the cysteine-trap portion. *J Med Chem*. 2010;53(5):2038–2050.
- Bai F, Liu H, Tong L, et al. Discovery of novel selective inhibitors for EGFR-T790M/L858R. *Bioorg Med Chem Lett*. 2012;22(3):1365–1370.
- Li S, Guo C, Sun X, et al. Synthesis and biological evaluation of quinazoline and quinoline bearing 2,2,6,6-tetramethylpiperidine-N-oxyl as potential epidermal growth factor receptor(EGFR) tyrosine kinase inhibitors and EPR bio-probe agents. *Eur J Med Chem*. 2012;49:271–278.
- Sato T, Watanabe H, Tsuganezawa K, et al. Identification of novel drug-resistant EGFR mutant inhibitors by in silico screening using comprehensive assessments of protein structures. *Bioorg Med Chem*. 2012;20(12):3756–3767.
- Salomon DS, Brandt R, Ciardiello F, Normanno N. Epidermal growth factor-related peptides and their receptors in human malignancies. *Crit Rev Oncol Hematol*. 1995;19(3):183–232.
- Roskoski R. The ErbB/HER receptor protein-tyrosine kinases and cancer. *Biochem Biophys Res Commun*. 2004;319(1):1–11.
- Gazdar AF. Activating and resistance mutations of EGFR in non-small-cell lung cancer: role in clinical response to EGFR tyrosine kinase inhibitors. *Oncogene*. 2009;28 Suppl 1:S24–S31.
- Lemmon MA, Schlessinger J. Cell signaling by receptor tyrosine kinases. *Cell*. 2010;141(7):1117–1134.
- Lurje G, Lenz HJ. EGFR signaling and drug discovery. *Oncology*. 2009;77(6):400–410.
- Hynes NE, Lane HA. ERBB receptors and cancer: the complexity of targeted inhibitors. *Nat Rev Cancer*. 2005;5(5):341–354.
- Doebele RC, Oton AB, Peled N, Camidge DR, Bunn PA. New strategies to overcome limitations of reversible EGFR tyrosine kinase inhibitor therapy in non-small cell lung cancer. *Lung Cancer*. 2010;69(1):1–12.
- Sequist LV. Second-generation epidermal growth factor receptor tyrosine kinase inhibitors in non-small cell lung cancer. *Oncologist*. 2007;12(3):325–330.
- Yun CH, Mengwasser KE, Toms AV, et al. The T790M mutation in EGFR kinase causes drug resistance by increasing the affinity for ATP. *Proc Natl Acad Sci U S A*. 2008;105(6):2070–2075.
- Schneider MR, Wolf E. The epidermal growth factor receptor ligands at a glance. *J Cell Physiol*. 2009;218(3):460–466.
- Lupberger J, Zeisel MB, Xiao F, et al. EGFR and EphA2 are host factors for hepatitis C virus entry and possible targets for antiviral therapy. *Nat Med*. 2011;17(5):589–595.
- Spicer JF, Rudman SM. EGFR inhibitors in non-small cell lung cancer (NSCLC): the emerging role of the dual irreversible EGFR/HER2 inhibitor BIBW 2992. *Target Oncol*. 2010;5(4):245–255.
- Rocha-Lima CM, Soares HP, Racz LE, Singal R. EGFR targeting of solid tumors. *Cancer Control*. 2007;14(3):295–304.
- Okamoto I. Epidermal growth factor receptor in relation to tumor development: EGFR-targeted anticancer therapy. *FEBS J*. 2010;277(2):309–315.
- Ciardiello F, Tortora G. EGFR antagonists in cancer treatment. *New Engl J Med*. 2008;358(11):1160–1174.
- Kobayashi S, Ji H, Yuza Y, et al. An alternative inhibitor overcomes resistance caused by a mutation of the epidermal growth factor receptor. *Cancer Res*. 2005;65(16):7096–7101.
- Greulich H, Chen TH, Feng W, et al. Oncogenic transformation by inhibitor-sensitive and -resistant EGFR mutants. *PLoS Med*. 2005;2(11):e313.
- Riely GJ. Second-generation epidermal growth factor receptor tyrosine kinase inhibitors in non-small cell lung cancer. *J Thorac Oncol*. 2008;3(6 Suppl 2):S146–S149.

27. Liu B, Bernard B, Wu JH. Impact of EGFR point mutations on the sensitivity to gefitinib: insights from comparative structural analyses and molecular dynamics simulations. *Proteins*. 2006;65(2):331–346.
28. Wissner A, Mansour TS. The development of HKI-272 and related compounds for the treatment of cancer. *Arch Pharm (Weinheim)*. 2008;341(8):465–477.
29. Li D, Ambrogio L, Shimamura T, et al. BIBW2992, an irreversible EGFR/HER2 inhibitor highly effective in preclinical lung cancer models. *Oncogene*. 2008;27(34):4702–4711.
30. Engelman JA, Zejnullahu K, Gale CM, et al. PF00299804, an irreversible pan-ERBB inhibitor, is effective in lung cancer models with EGFR and ERBB2 mutations that are resistant to gefitinib. *Cancer Res*. 2007;67(24):11924–11932.
31. Kwak EL, Sordella R, Bell DW, et al. Irreversible inhibitors of the EGF receptor may circumvent acquired resistance to gefitinib. *Proc Natl Acad Sci U S A*. 2005;102(21):7665–7670.
32. Godin-Heymann N, Ulkus L, Brannigan BW, et al. The T790M “gatekeeper” mutation in EGFR mediates resistance to low concentrations of an irreversible EGFR inhibitor. *Mol Cancer Ther*. 2008;7(4):874–879.
33. Blair JA, Rauh D, Kung C, et al. Structure-guided development of affinity probes for tyrosine kinases using chemical genetics. *Nat Chem Biol*. 2007;3(4):229–238.
34. Tsou HR, Overbeek-Klumpers EG, Hallett WA, et al. Optimization of 6,7-disubstituted-4-(arylamino)quinoline-3-carbonitriles as orally active, irreversible inhibitors of human epidermal growth factor receptor-2 kinase activity. *J Med Chem*. 2005;48(4):1107–1131.
35. Wissner A, Overbeek E, Reich MF, et al. Synthesis and structure-activity relationships of 6,7-disubstituted 4-anilinoquinoline-3-carbonitriles. The design of an orally active, irreversible inhibitor of the tyrosine kinase activity of the epidermal growth factor receptor (EGFR) and the human epidermal growth factor receptor-2 (HER-2). *J Med Chem*. 2003;46(1):49–63.
36. Wissner A, Brawner Floyd MB, Rabindran SK, et al. Syntheses and EGFR and HER-2 kinase inhibitory activities of 4-anilinoquinoline-3-carbonitriles: analogues of three important 4-anilinoquinazolines currently undergoing clinical evaluation as therapeutic antitumor agents. *Bioorg Med Chem Lett*. 2002;12(20):2893–2897.
37. Tsou HR, Mamuya N, Johnson BD, et al. 6-Substituted-4-(3-bromophenylamino)quinazolines as putative irreversible inhibitors of the epidermal growth factor receptor (EGFR) and human epidermal growth factor receptor (HER-2) tyrosine kinases with enhanced antitumor activity. *J Med Chem*. 2001;44(17):2719–2734.
38. Marvin, <http://www.chemaxon.com> [computer program]. Version 5.6.0.1: ChemAxon Ltd; 2011.
39. Halgren TA. Merck molecular force field. I. Basis, form, scope, parameterization, and performance of MMFF94. *J Comput Chem*. 1996;17(5–6):490–519.
40. Gaussian 03 [computer program]. Version Revision C.02 ed. Wallingford CT: Gaussian, Inc; 2004.
41. Yang W, Mortier WJ. The use of global and local molecular parameters for the analysis of the gas-phase basicity of amines. *J Am Chem Soc*. 1986;108(19):5708–5711.
42. Boström J. Reproducing the conformations of protein-bound ligands: a critical evaluation of several popular conformational searching tools. *J Comput Aided Mol Des*. 2001;15(12):1137–1152.
43. Discovery Studio Modeling Environment [computer program]. Version Release 3.1. San Diego: Accelrys Software Inc; 2012.
44. Michalczyk A, Klüter S, Rode HB, et al. Structural insights into how irreversible inhibitors can overcome drug resistance in EGFR. *Bioorg Med Chem*. 2008;16(7):3482–3488.
45. FRED program. <http://www.eyesopen.com> [computer program]. Version 2.2.5: OpenEye Scientific Software Inc., Santa Fe, NM, USA; 2012.
46. Nicholls A, Grant JA, Brown FK. Gaussian docking functions. *Bio-polymers*. 2003;68(1):76–90.
47. Verkhivker GM, Bouzida D, Gehlhaar DK, et al. Deciphering common failures in molecular docking of ligand-protein complexes. *J Comput Aided Mol Des*. 2000;14(8):731–751.
48. Eldridge MD, Murray CW, Auton TR, Paolini GV, Mee RP. Empirical scoring functions: I. The development of a fast empirical scoring function to estimate the binding affinity of ligands in receptor complexes. *J Comput Aided Mol Des*. 1997;11(5):425–445.
49. Stahl M, Rarey M. Detailed analysis of scoring functions for virtual screening. *J Med Chem*. 2001;44(7):1035–1042.
50. Grant JA, Pickup BT, Nicholls A. A smooth permittivity function for Poisson-Boltzmann solvation methods. *J Comput Chem*. 2001;22(6):608–640.
51. Traxler P, Furet P. Strategies toward the design of novel and selective protein tyrosine kinase inhibitors. *Pharmacol Ther*. 1999;82(2–3):195–206.
52. Cavasotto CN, Ortiz MA, Abagyan RA, Piedrafita FJ. In silico identification of novel EGFR inhibitors with antiproliferative activity against cancer cells. *Bioorg Med Chem Lett*. 2006;16(7):1969–1974.
53. Rode HB, Sos ML, Grütter C, Heynck S, Simard JR, Rauh D. Synthesis and biological evaluation of 7-substituted-1-(3-bromophenylamino)isoquinoline-4-carbonitriles as inhibitors of myosin light chain kinase and epidermal growth factor receptor. *Bioorg Med Chem*. 2011;19(1):429–439.
54. Dror O, Shulman-Peleg A, Nussinov R, Wolfson HJ. Predicting molecular interactions in silico: I. A guide to pharmacophore identification and its applications to drug design. *Curr Med Chem*. 2004;11(1):71–90.
55. Güner OF. History and evolution of the pharmacophore concept in computer-aided drug design. *Curr Top Med Chem*. 2002;2(12):1321–1332.
56. Xiang M, Cao Y, Fan W, Chen L, Mo Y. Computer-aided drug design: lead discovery and optimization. *Comb Chem High Throughput Screen*. 2012;15(4):328–337.
57. Davis AM, Teague SJ, Kleywegt GJ. Application and limitations of X-ray crystallographic data in structure-based ligand and drug design. *Angew Chem Int Ed Engl*. 2003;42(24):2718–2736.
58. Tuccinardi T. Docking-based virtual screening: recent developments. *Comb Chem High Throughput Screen*. 2009;12(3):303–314.
59. Plewczynski D, Łażniewski M, von Grotthuss M, Rychlewski L, Ginalski K. VoteDock: consensus docking method for prediction of protein-ligand interactions. *J Comput Chem*. 2011;32(4):568–581.
60. Schapira M, Abagyan R, Totrov M. Nuclear hormone receptor targeted virtual screening. *J Med Chem*. 2003;46(14):3045–3059.
61. Ravindranathan KP, Mandiyan V, Ekkati AR, Bae JH, Schlessinger J, Jorgensen WL. Discovery of novel fibroblast growth factor receptor 1 kinase inhibitors by structure-based virtual screening. *J Med Chem*. 2010;53(4):1662–1672.

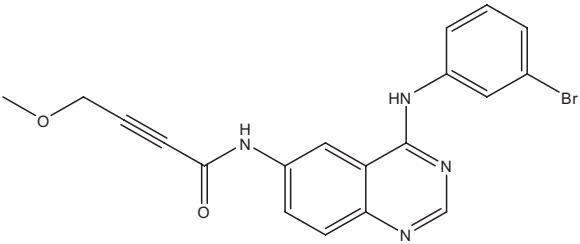
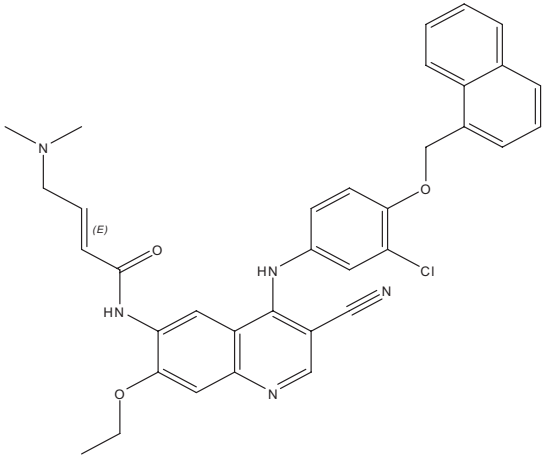
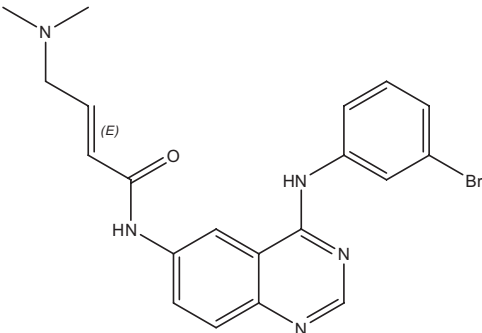
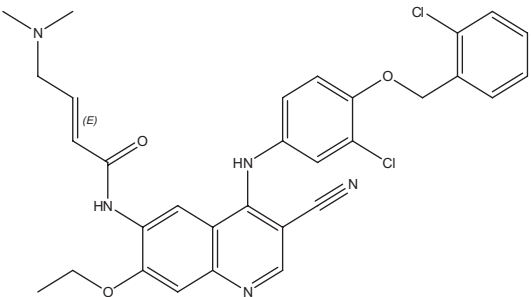
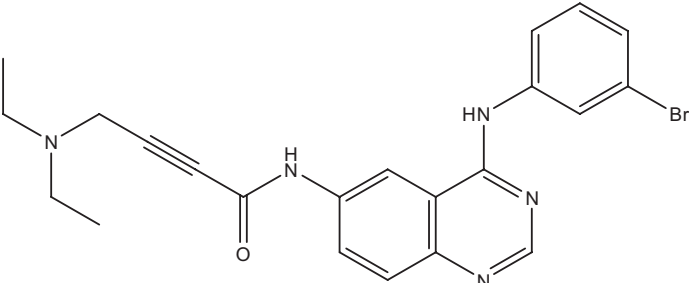
Supplementary tables

Table S1 138 currently known irreversible tyrosine kinase inhibitors¹⁻⁴ used for extracting common features and calibrating docking protocols

Serial number	Structure	EGFR IC ₅₀ (μM)
T-001		0.002
T-002		0.002
T-003		0.006
T-004		0.007
T-005		0.008

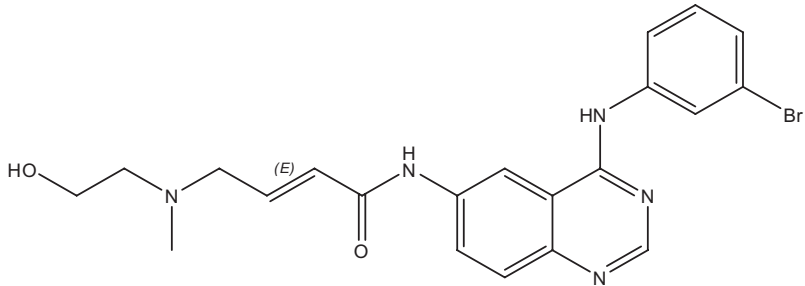
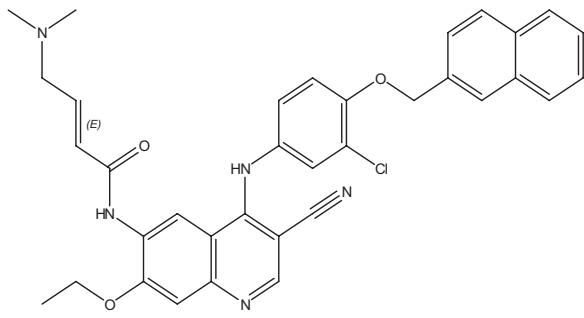
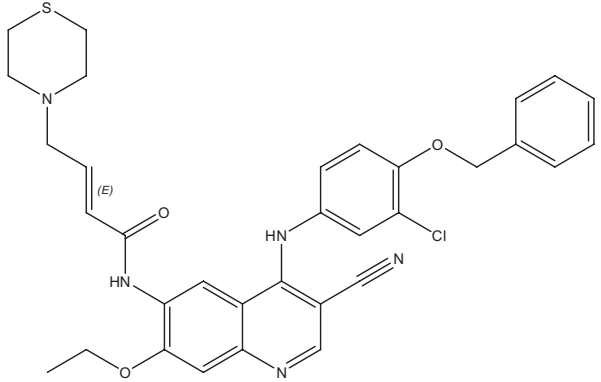
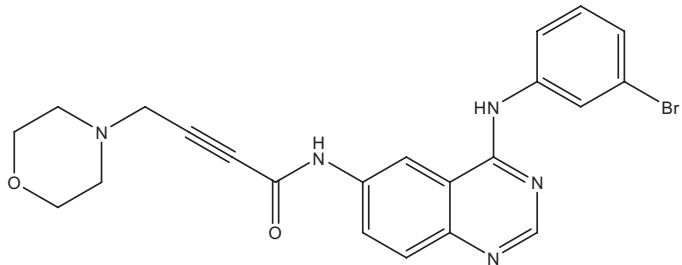
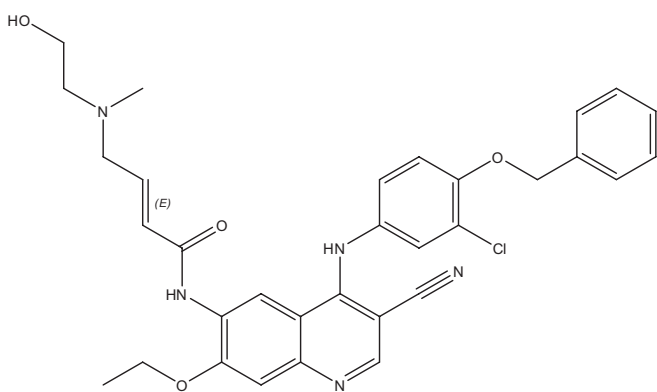
(Continued)

Table SI (Continued)

Serial number	Structure	EGFR IC ₅₀ (μM)
T-006		0.009
T-007		0.010
T-008		0.011
T-009		0.011
T-010		0.012

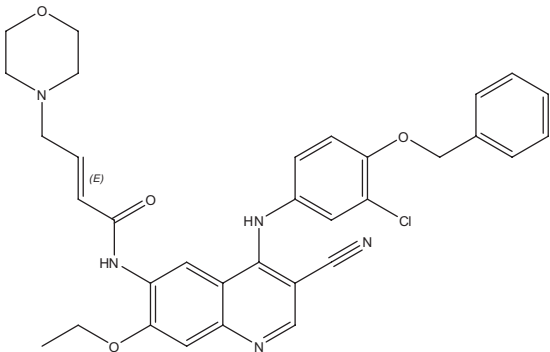
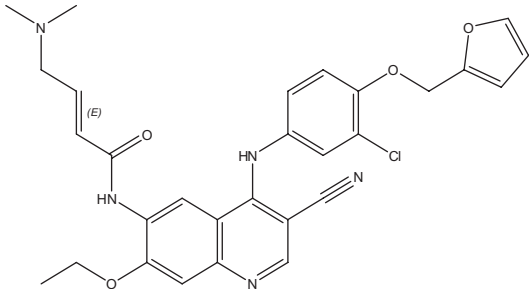
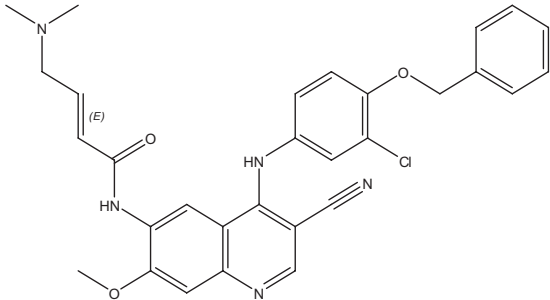
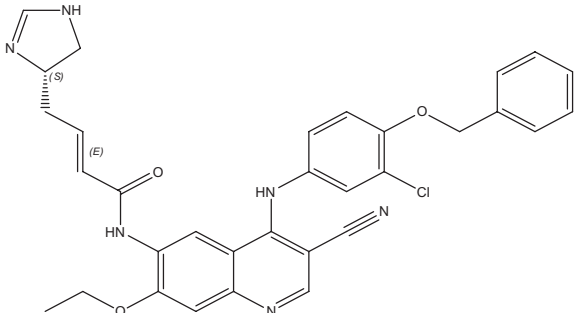
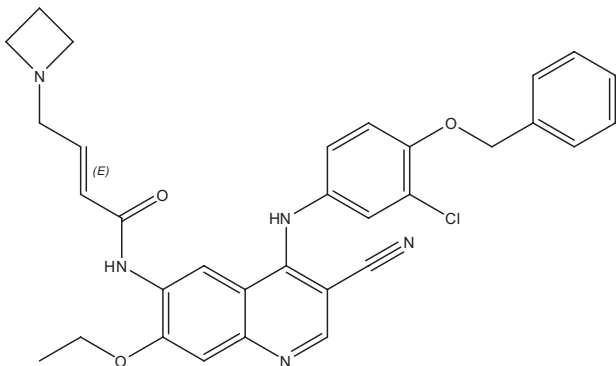
(Continued)

Table S1 (Continued)

Serial number	Structure	EGFR IC ₅₀ (μM)
T-011		0.014
T-012		0.015
T-013		0.016
T-014		0.020
T-015		0.020

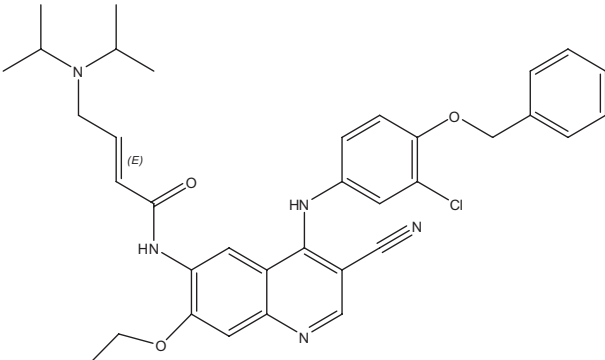
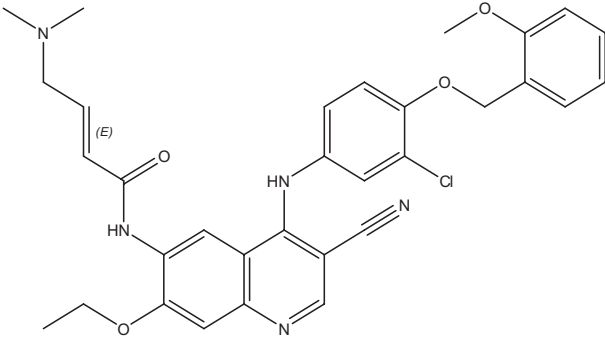
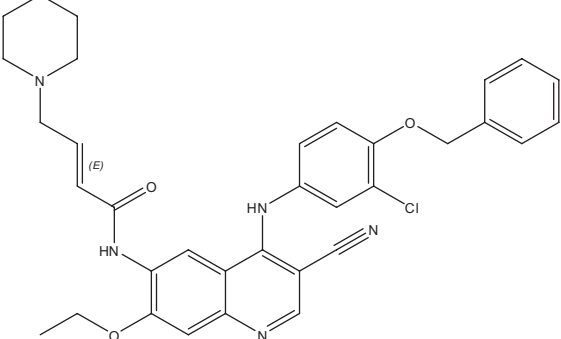
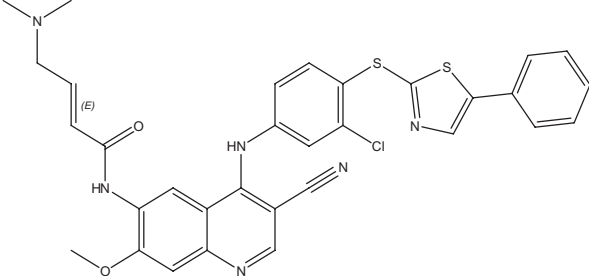
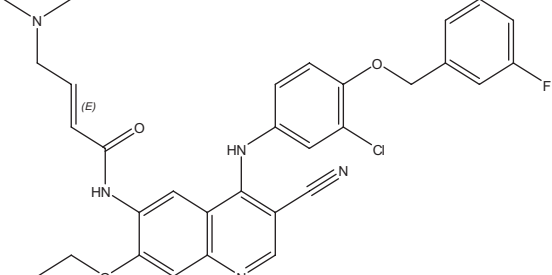
(Continued)

Table S1 (Continued)

Serial number	Structure	EGFR IC ₅₀ (μM)
T-016		0.020
T-017		0.023
T-018		0.024
T-019		0.024
T-020		0.026

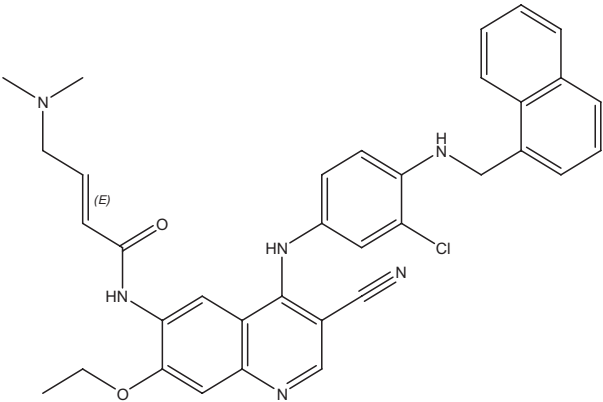
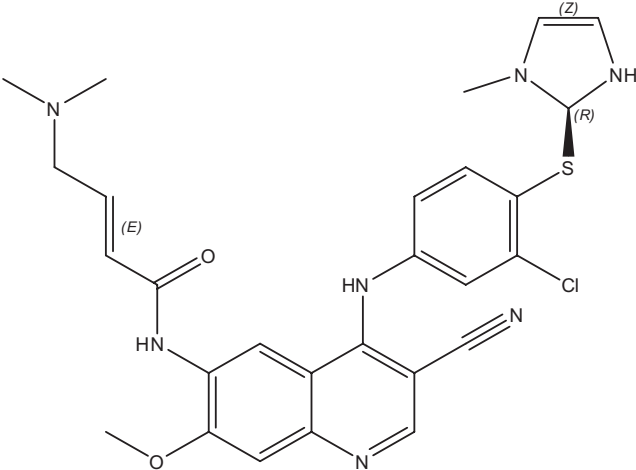
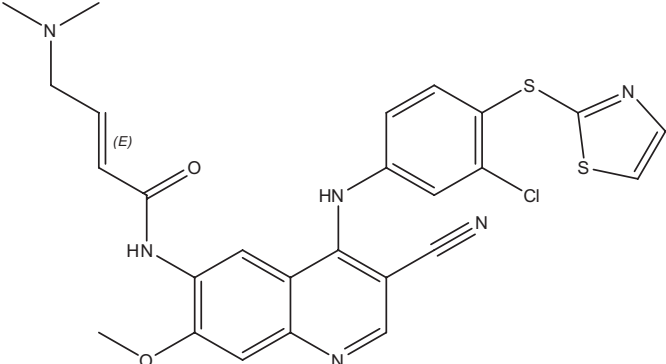
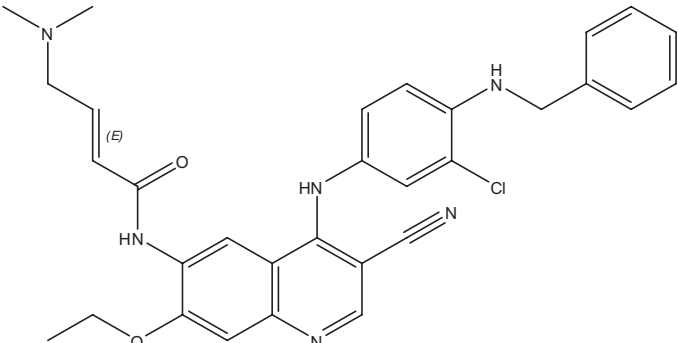
(Continued)

Table S1 (Continued)

Serial number	Structure	EGFR IC ₅₀ (μM)
T-021		0.030
T-022		0.031
T-023		0.031
T-024		0.032
T-025		0.034

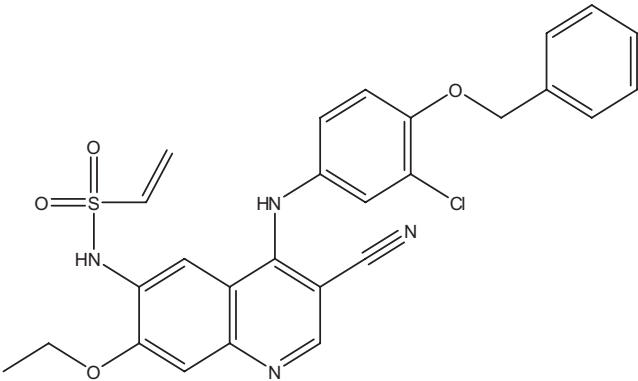
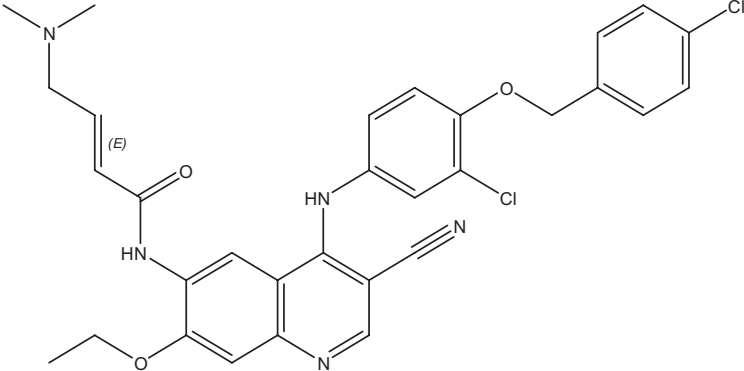
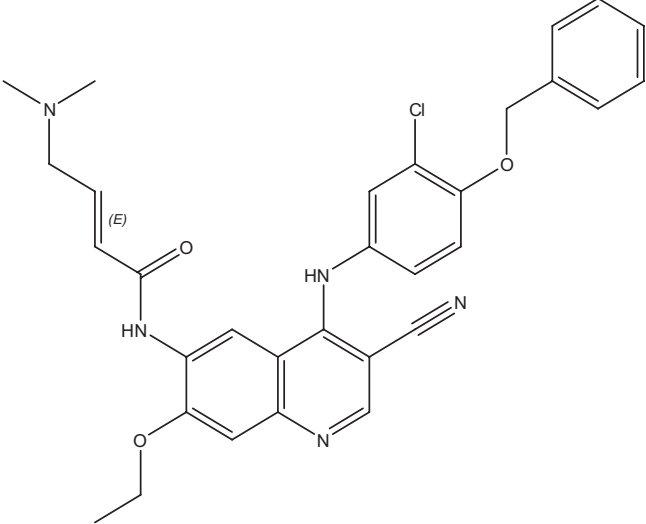
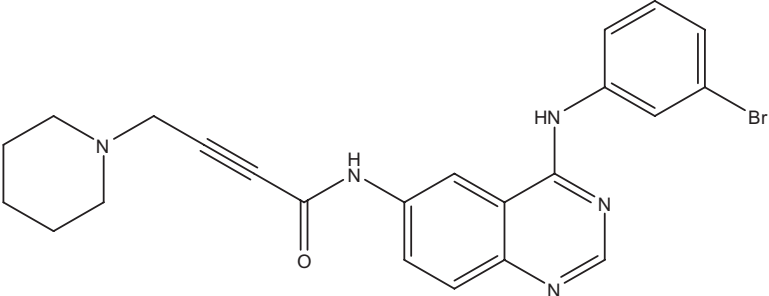
(Continued)

Table SI (Continued)

Serial number	Structure	EGFR IC ₅₀ (μM)
T-026		0.034
T-027		0.041
T-028		0.042
T-029		0.045

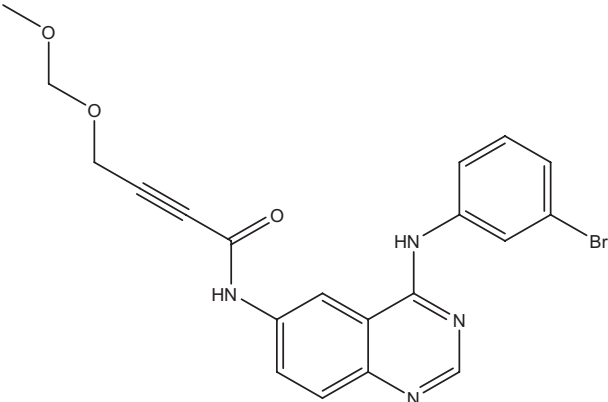
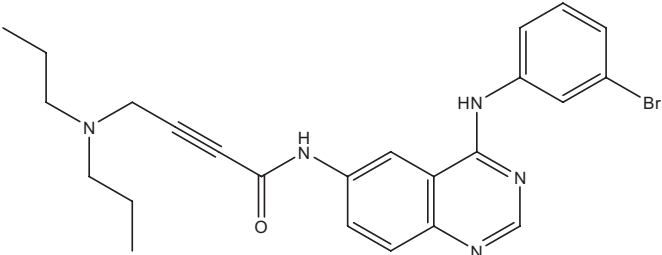
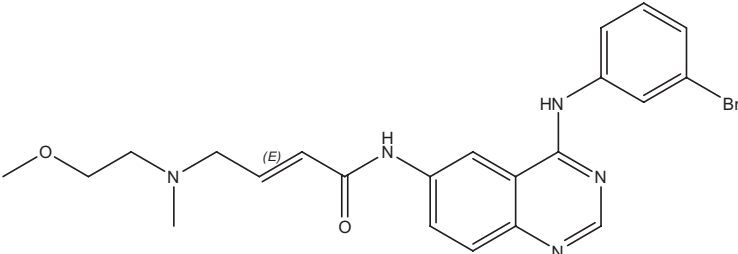
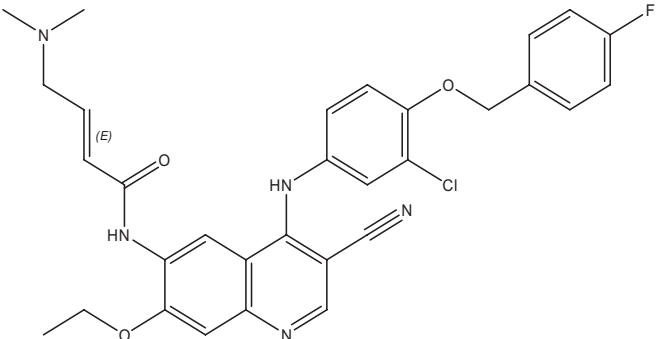
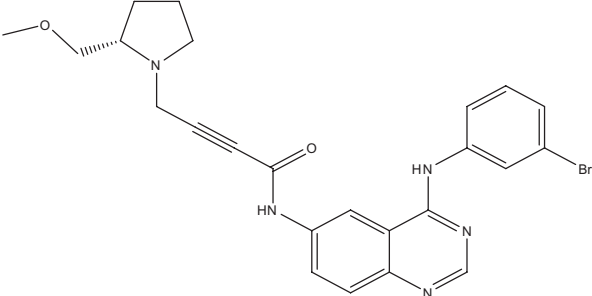
(Continued)

Table S1 (Continued)

Serial number	Structure	EGFR IC ₅₀ (μM)
T-030		0.045
T-031		0.051
T-032		0.053
T-033		0.057

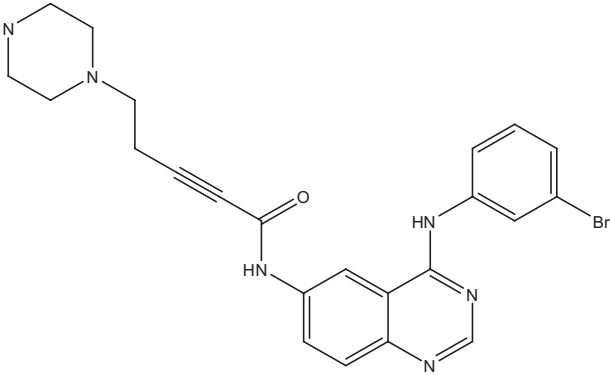
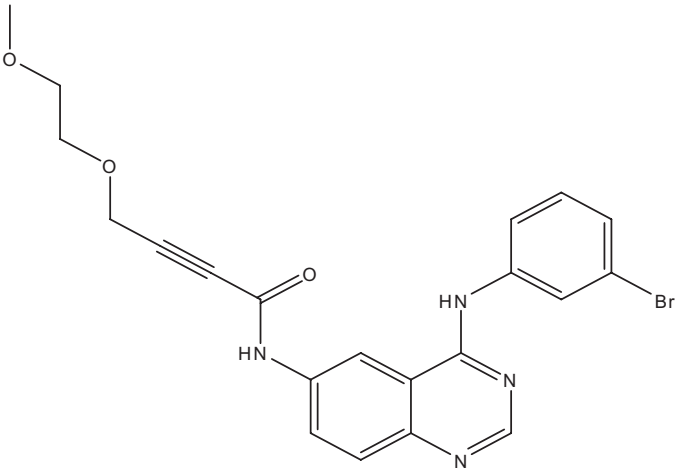
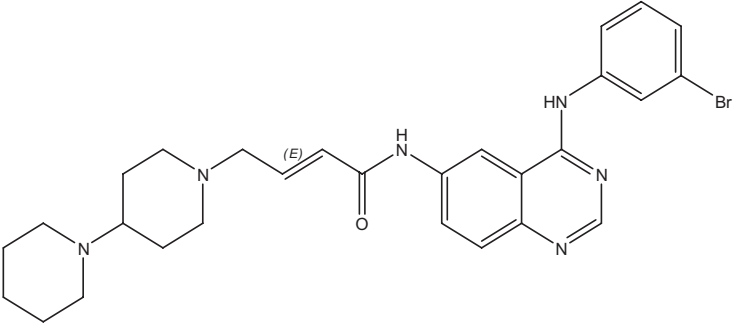
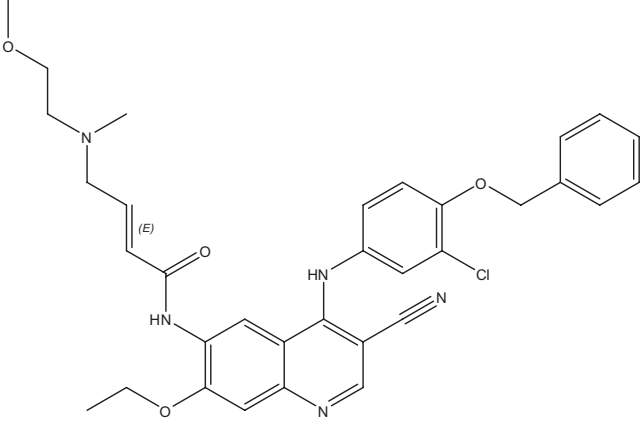
(Continued)

Table SI (Continued)

Serial number	Structure	EGFR IC ₅₀ (μM)
T-034		0.059
T-035		0.062
T-036		0.062
T-037		0.064
T-038		0.065

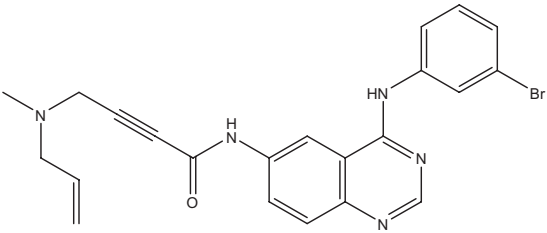
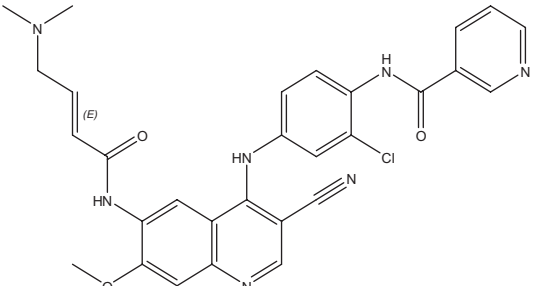
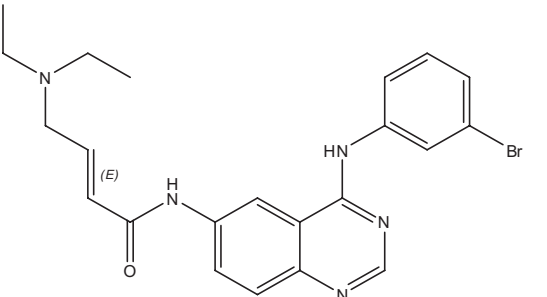
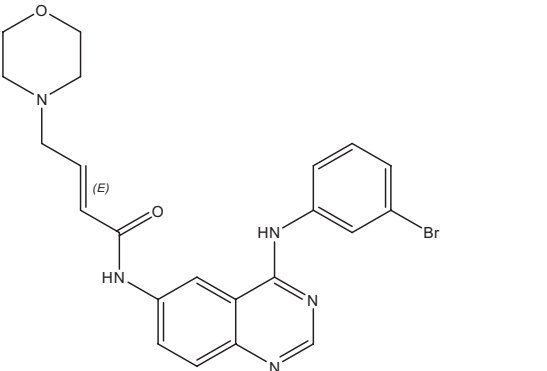
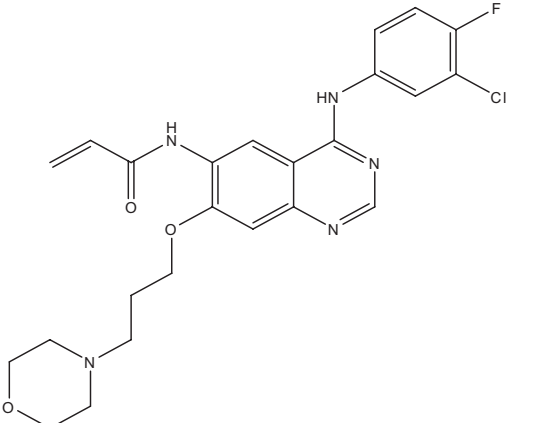
(Continued)

Table S1 (Continued)

Serial number	Structure	EGFR IC ₅₀ (μM)
T-039		0.065
T-040		0.065
T-041		0.068
T-042		0.068

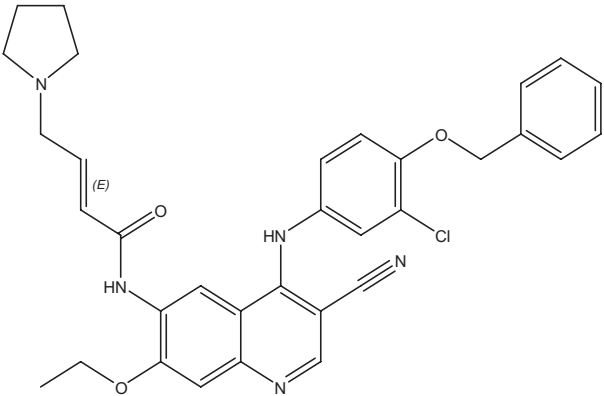
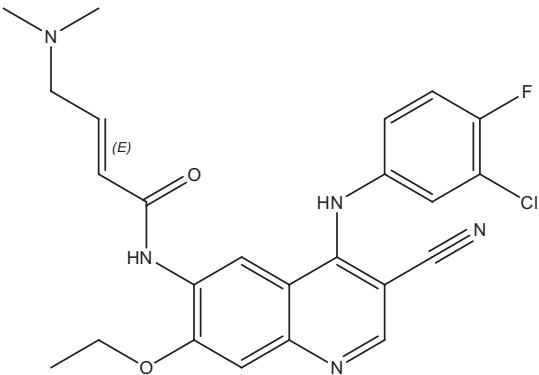
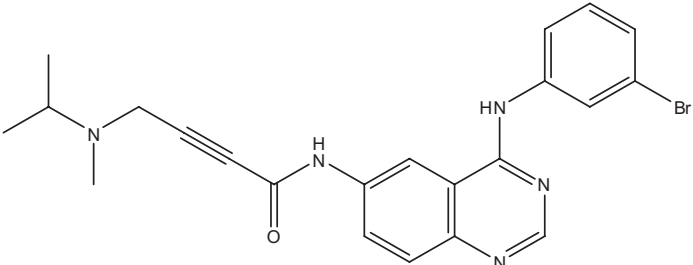
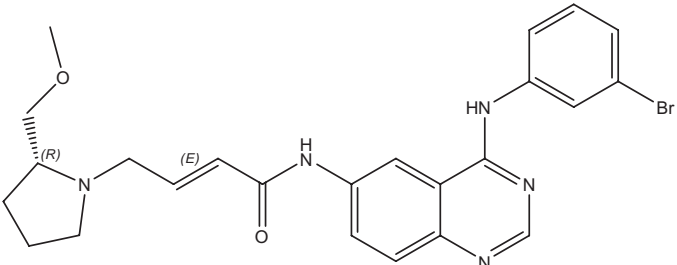
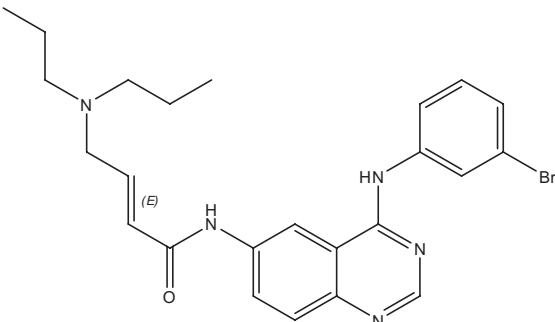
(Continued)

Table S1 (Continued)

Serial number	Structure	EGFR IC ₅₀ (μM)
T-043		0.071
T-044		0.072
T-045		0.073
T-046		0.074
T-047		0.074

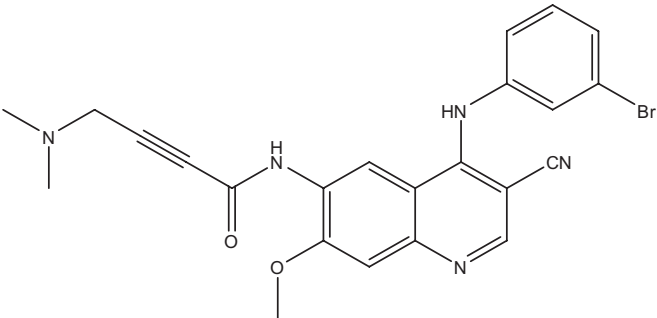
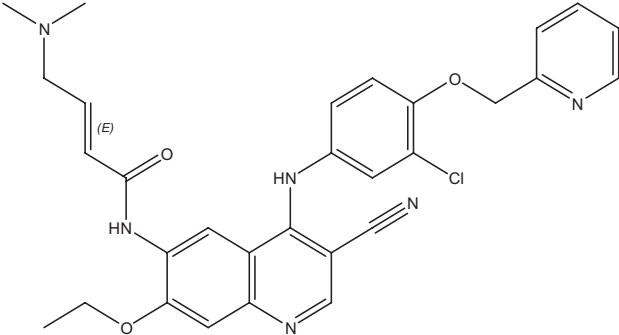
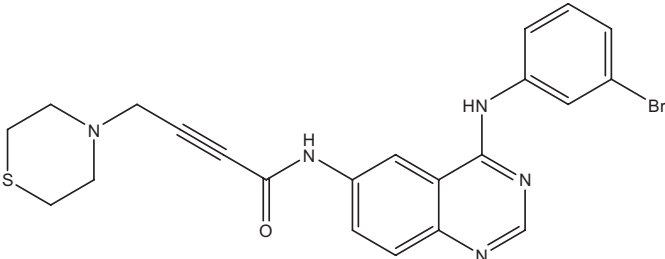
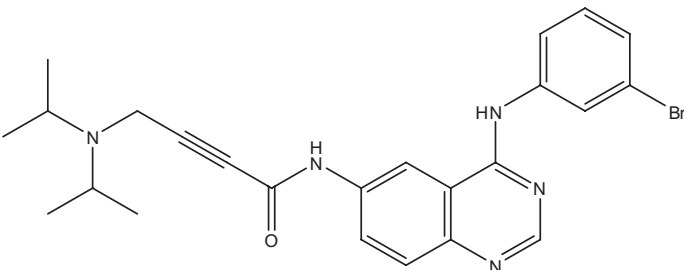
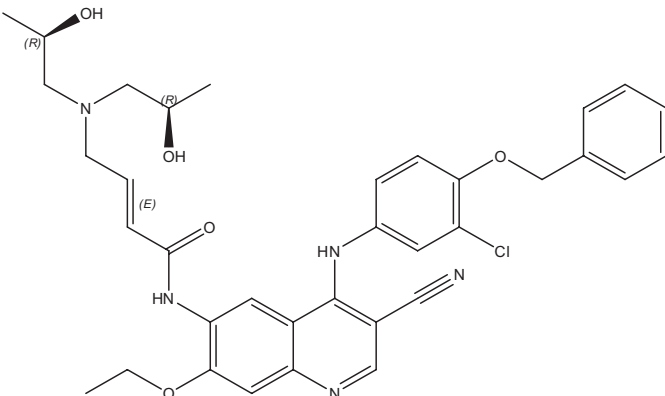
(Continued)

Table S1 (Continued)

Serial number	Structure	EGFR IC ₅₀ (μM)
T-048		0.078
T-049		0.080
T-050		0.085
T-051		0.085
T-052		0.090

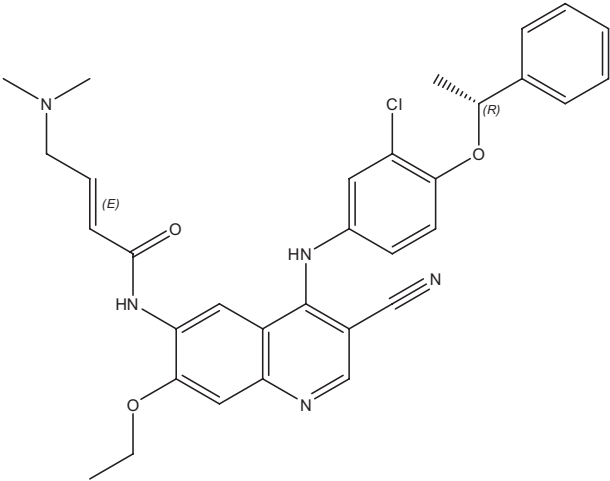
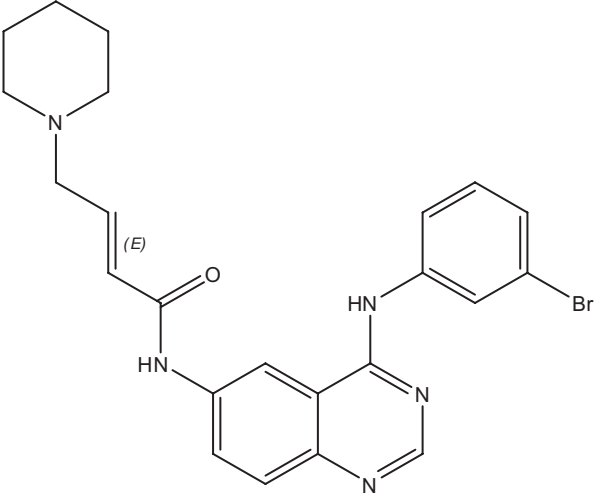
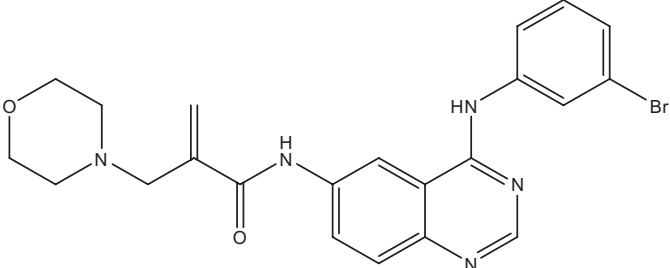
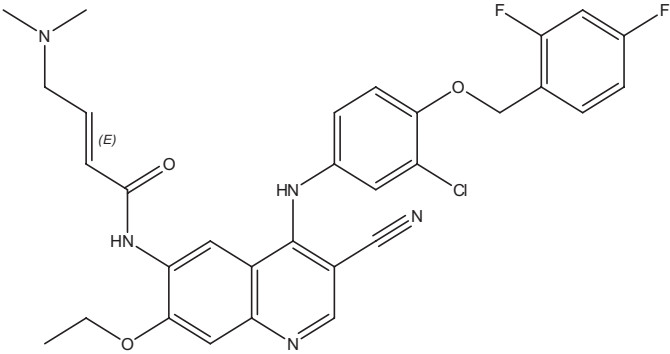
(Continued)

Table SI (Continued)

Serial number	Structure	EGFR IC ₅₀ (μM)
T-053		0.090
T-054		0.092
T-055		0.095
T-056		0.097
T-057		0.098

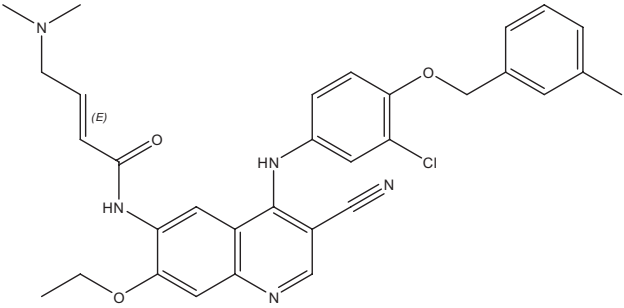
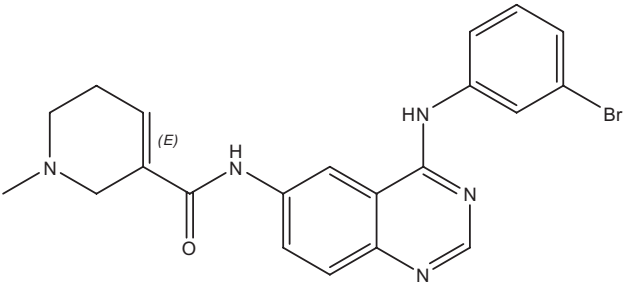
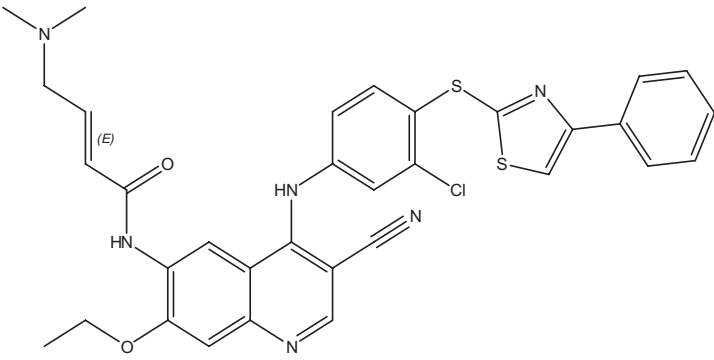
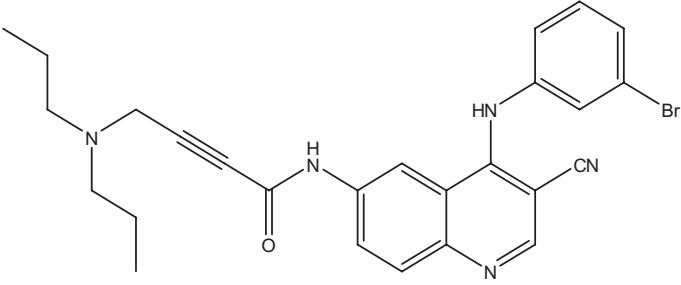
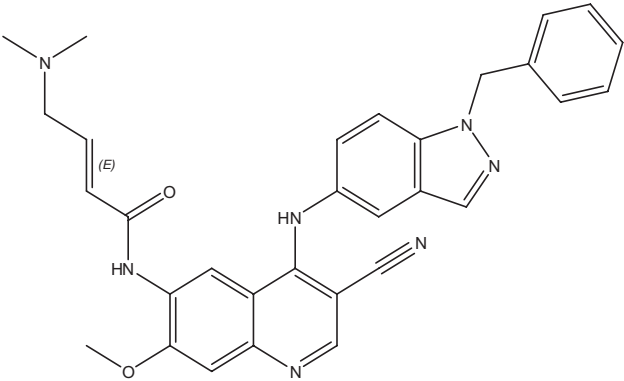
(Continued)

Table S1 (Continued)

Serial number	Structure	EGFR IC ₅₀ (μM)
T-058		0.100
T-059		0.106
T-060		0.107
T-061		0.110

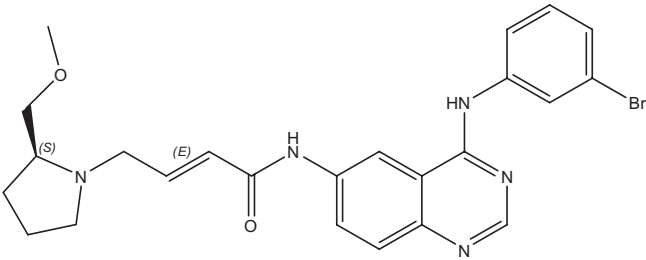
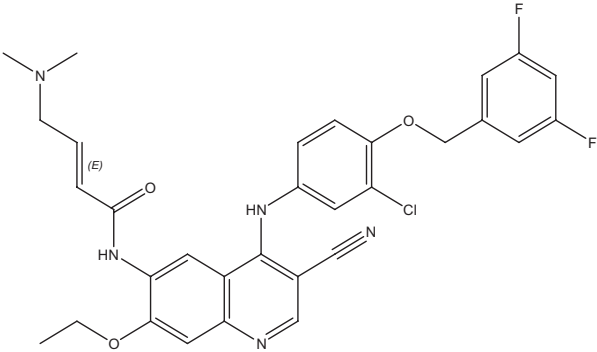
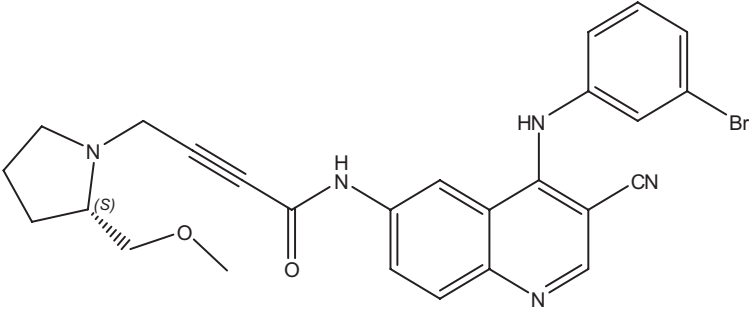
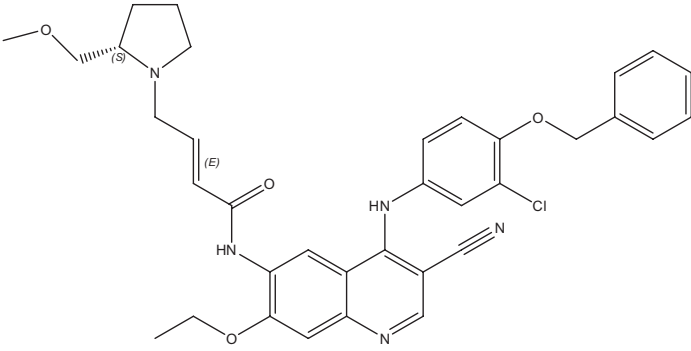
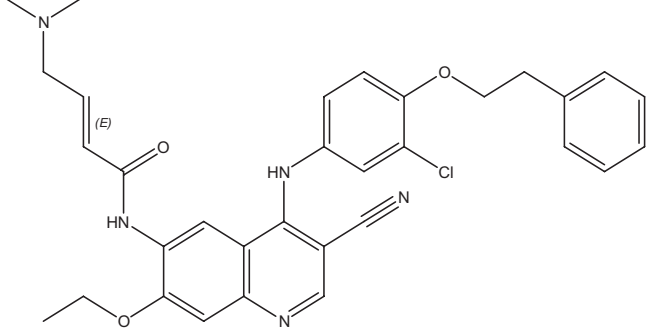
(Continued)

Table S1 (Continued)

Serial number	Structure	EGFR IC ₅₀ (μM)
T-062		0.111
T-063		0.115
T-064		0.125
T-065		0.130
T-066		0.130

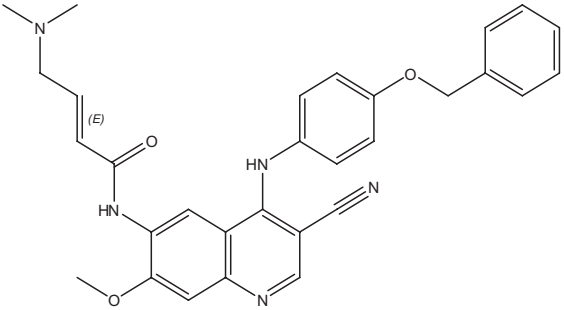
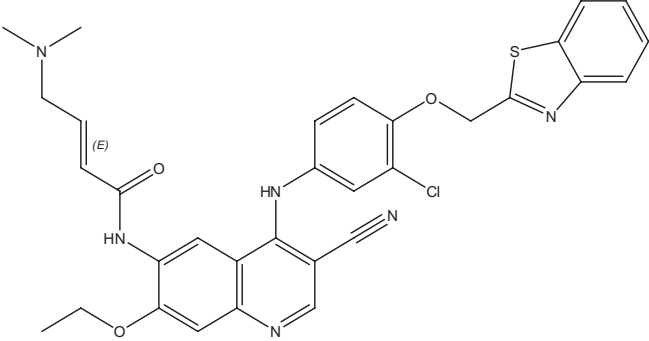
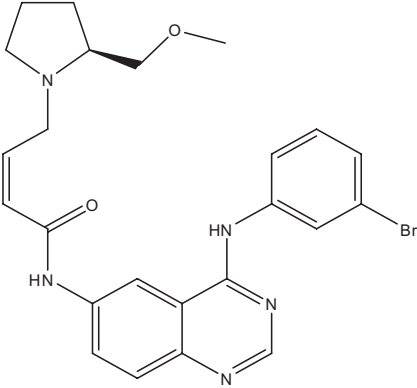
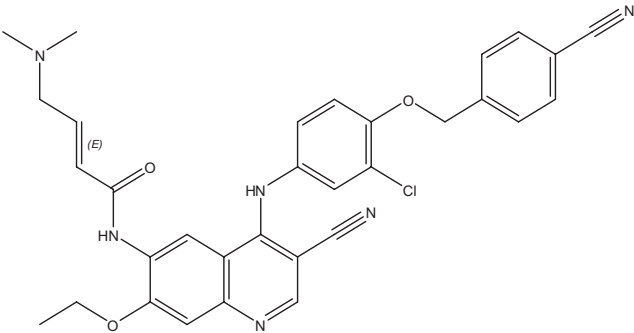
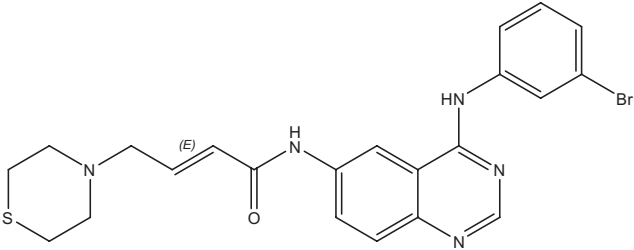
(Continued)

Table S1 (Continued)

Serial number	Structure	EGFR IC ₅₀ (μM)
T-067		0.131
T-068		0.140
T-069		0.150
T-070		0.151
T-071		0.165

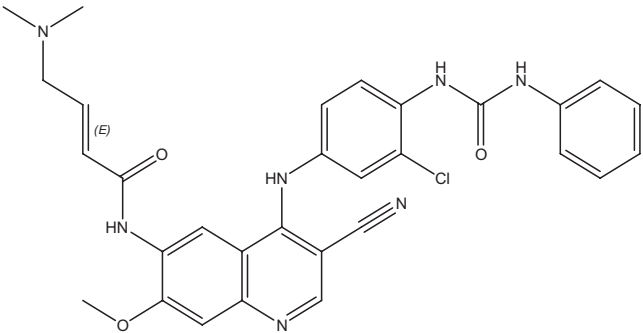
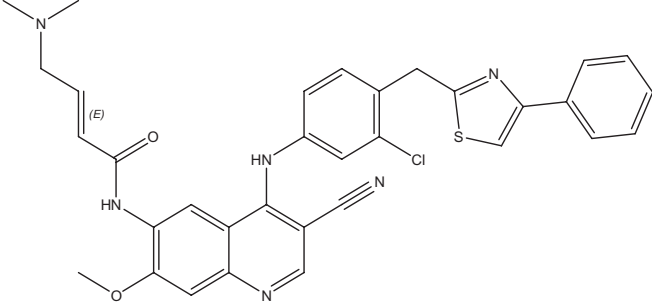
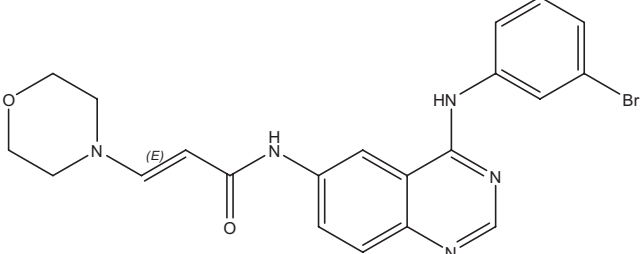
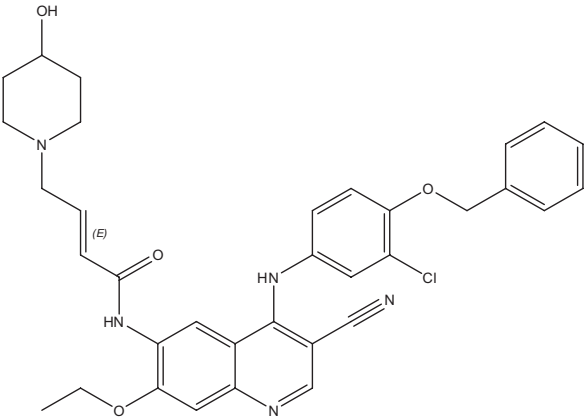
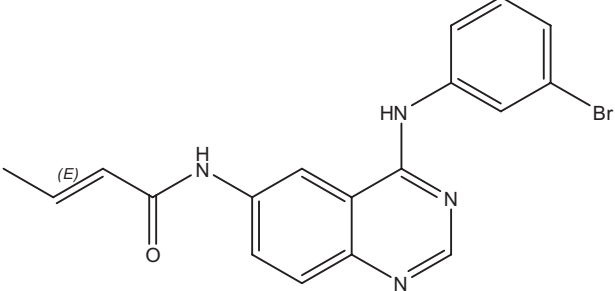
(Continued)

Table S1 (Continued)

Serial number	Structure	EGFR IC ₅₀ (μM)
T-072		0.185
T-073		0.188
T-074		0.194
T-075		0.195
T-076		0.203

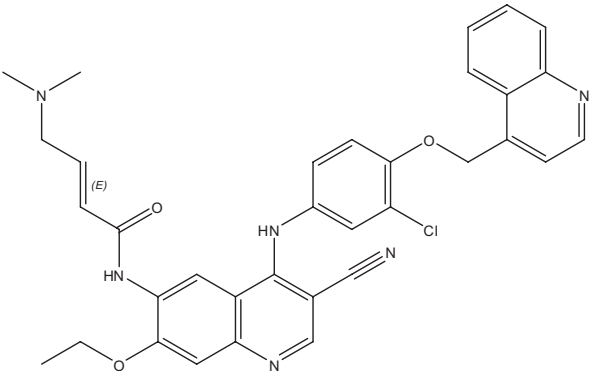
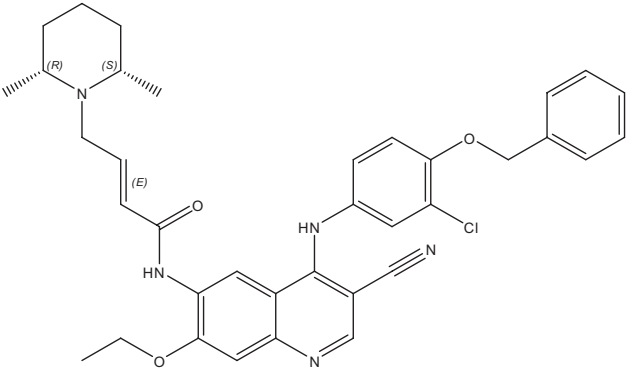
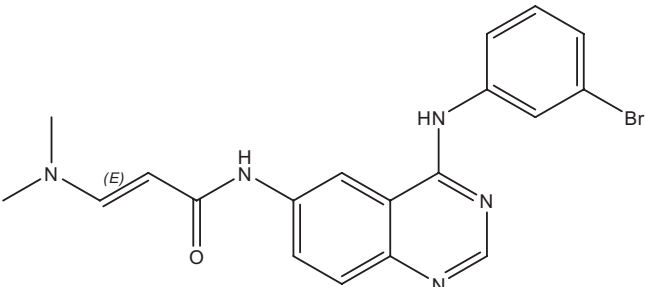
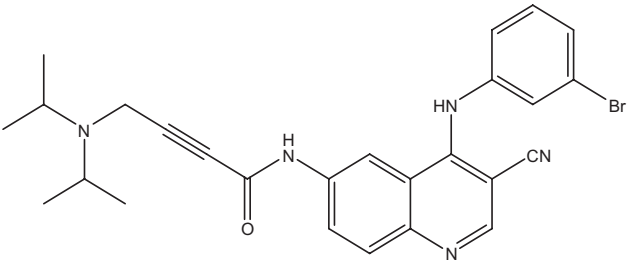
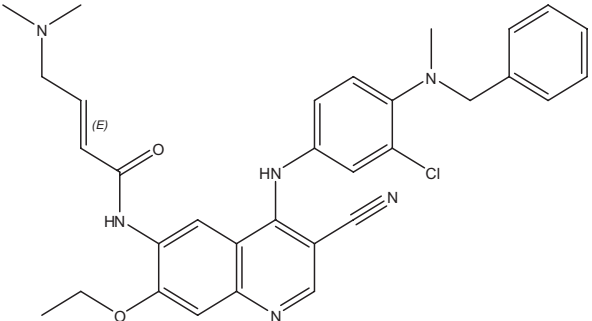
(Continued)

Table S1 (Continued)

Serial number	Structure	EGFR IC ₅₀ (μM)
T-077		0.210
T-078		0.214
T-079		0.215
T-080		0.221
T-081		0.236

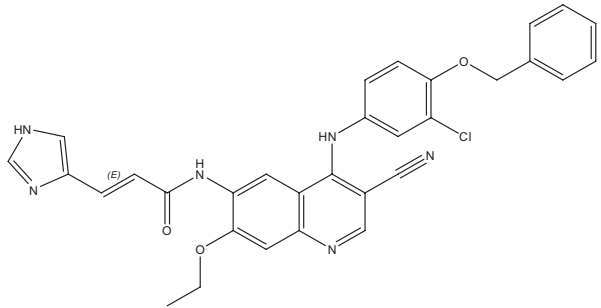
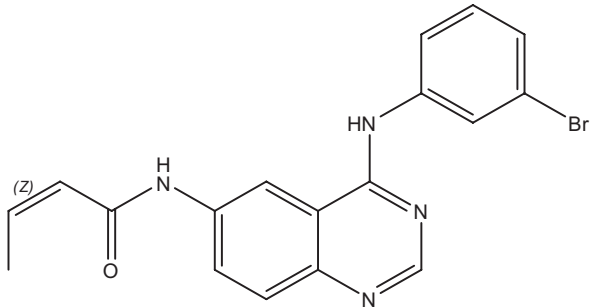
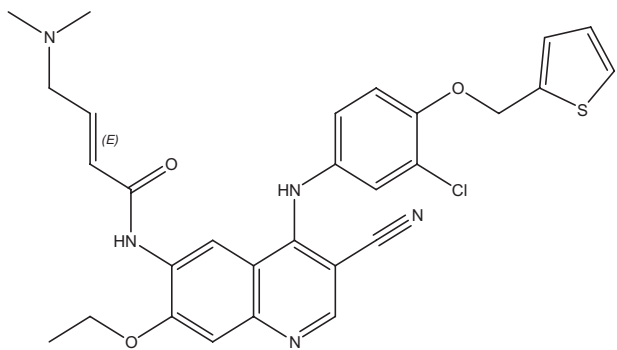
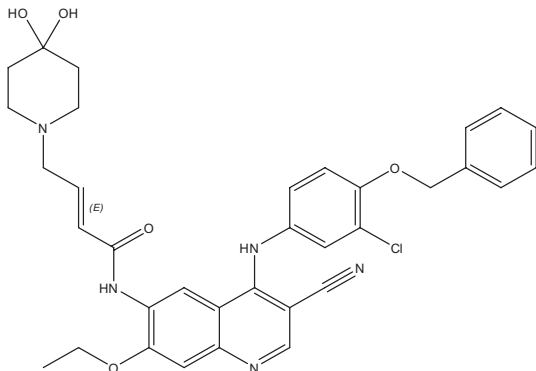
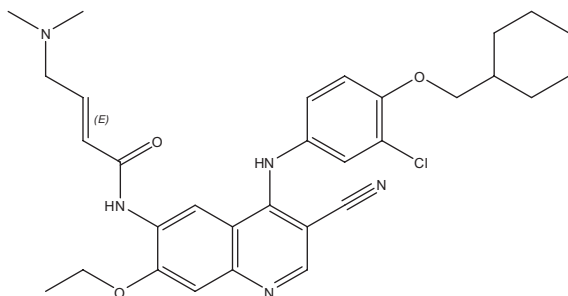
(Continued)

Table SI (Continued)

Serial number	Structure	EGFR IC ₅₀ (μM)
T-082		0.238
T-083		0.247
T-084		0.250
T-085		0.250
T-086		0.250

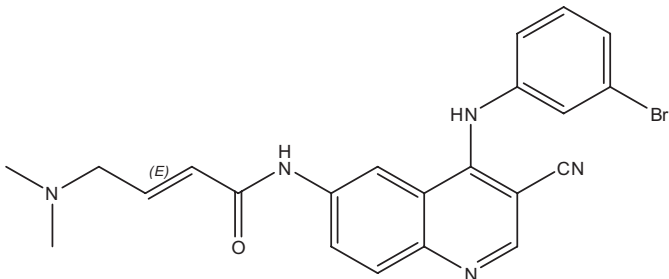
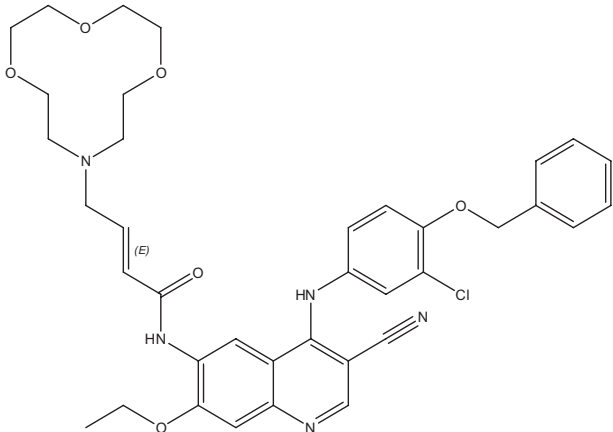
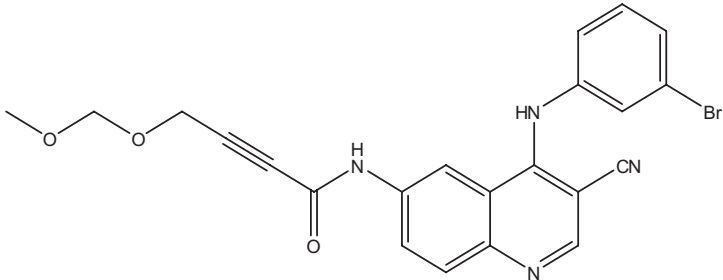
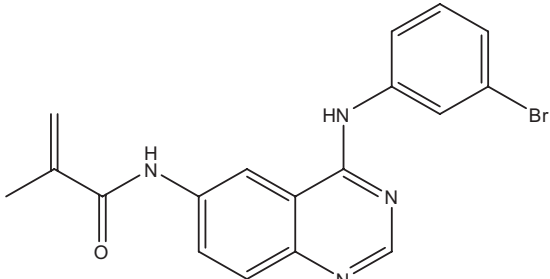
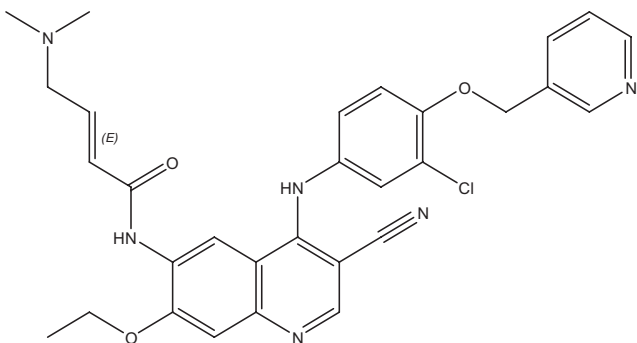
(Continued)

Table S1 (Continued)

Serial number	Structure	EGFR IC ₅₀ (μM)
T-087		0.255
T-088		0.269
T-089		0.277
T-090		0.282
T-091		0.285

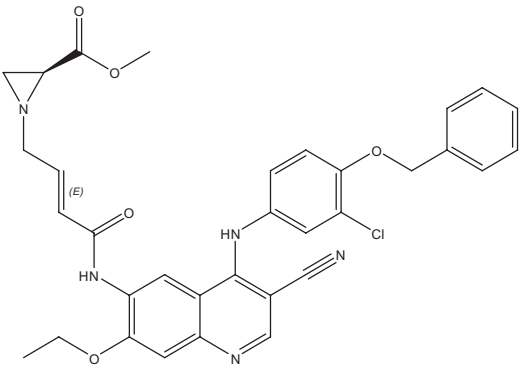
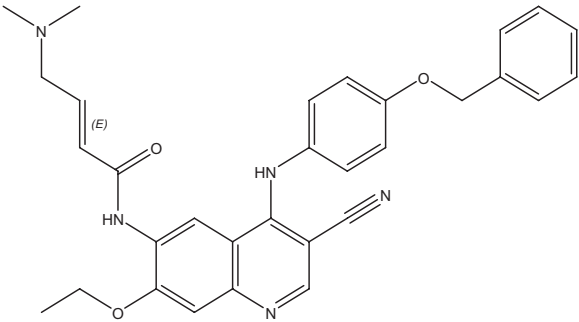
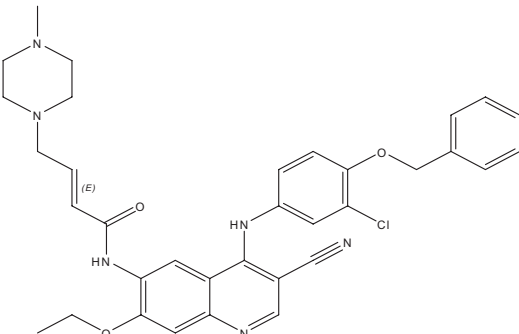
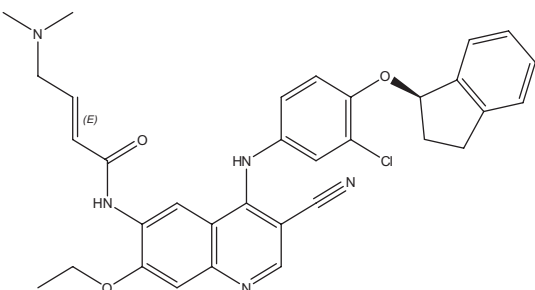
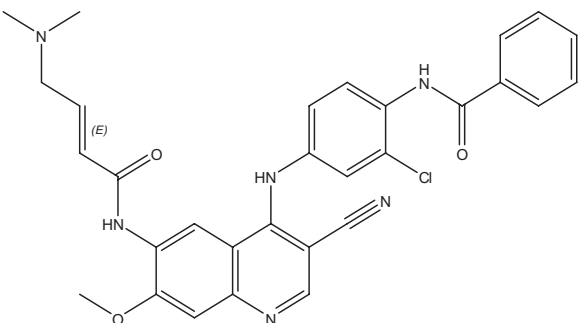
(Continued)

Table SI (Continued)

Serial number	Structure	EGFR IC ₅₀ (μM)
T-092		0.310
T-093		0.321
T-094		0.330
T-095		0.339
T-096		0.341

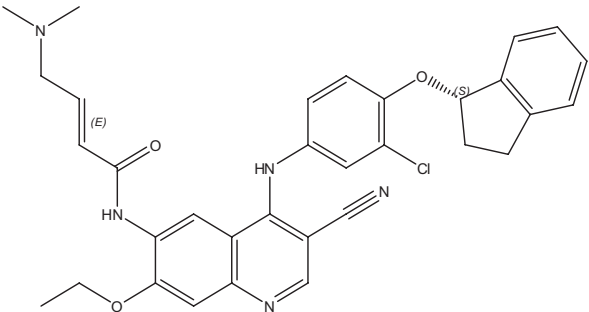
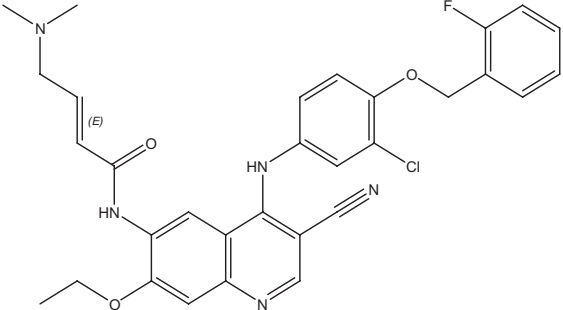
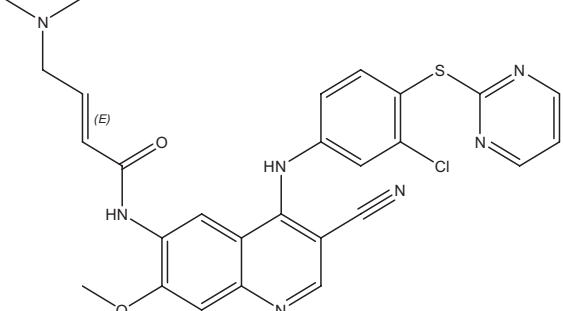
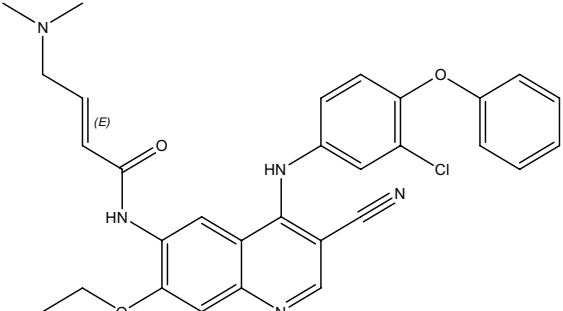
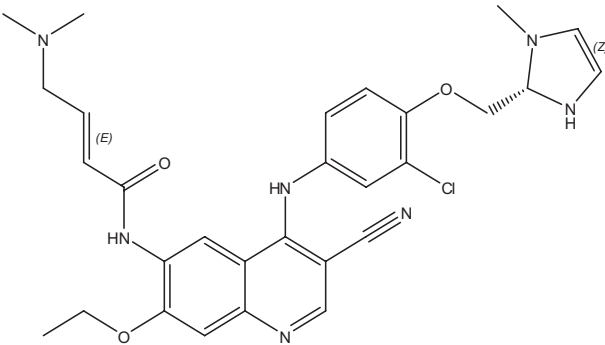
(Continued)

Table S1 (Continued)

Serial number	Structure	EGFR IC ₅₀ (μM)
T-097		0.343
T-098		0.345
T-099		0.360
T-100		0.380
T-101		0.400

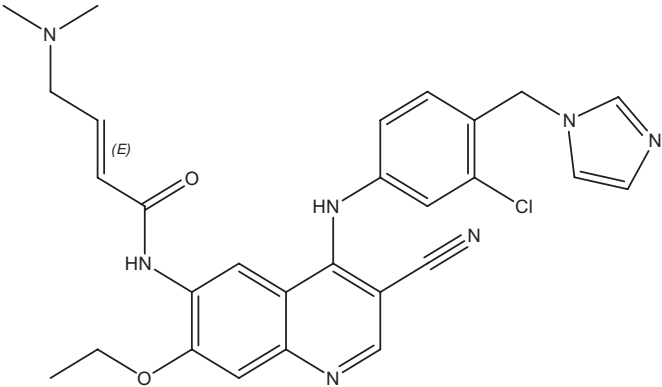
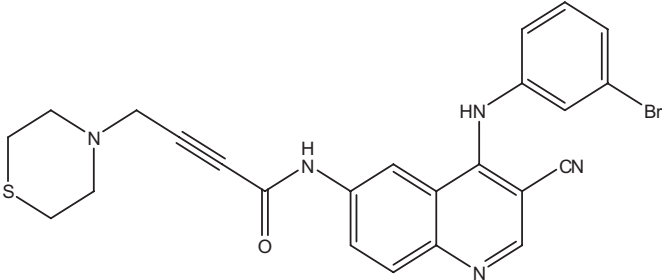
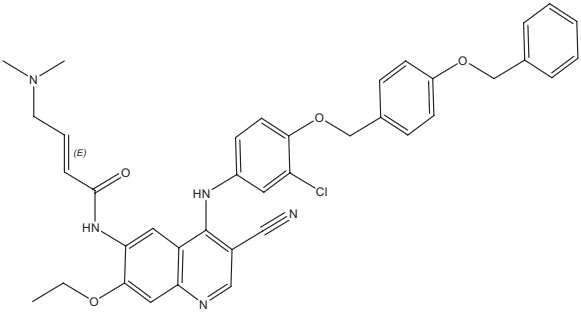
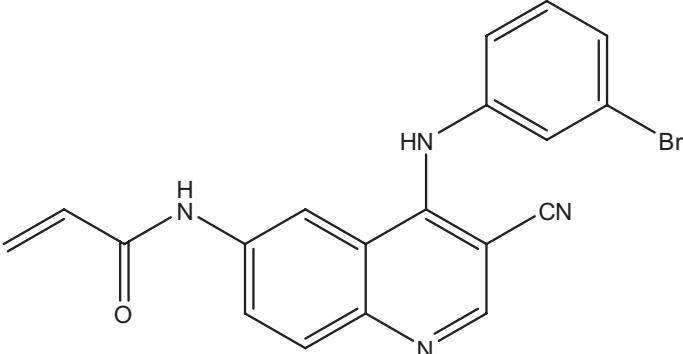
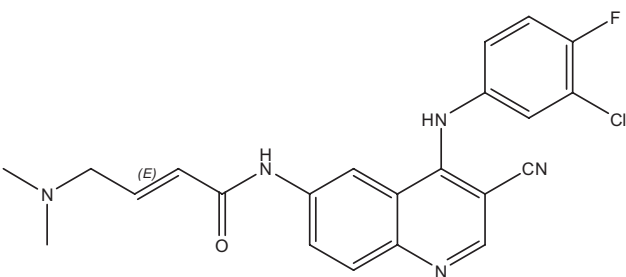
(Continued)

Table S1 (Continued)

Serial number	Structure	EGFR IC ₅₀ (μM)
T-102		0.400
T-103		0.401
T-104		0.403
T-105		0.445
T-106		0.500

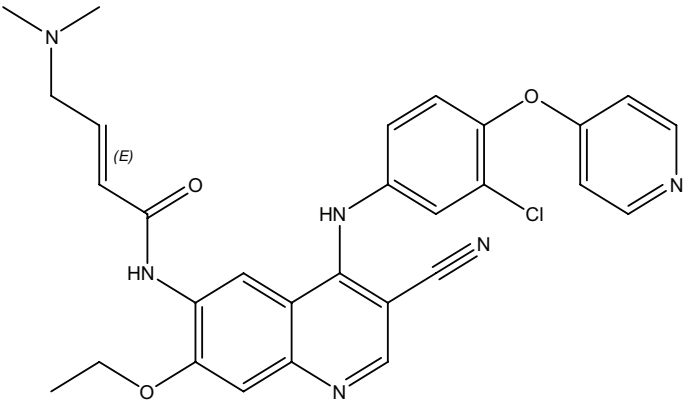
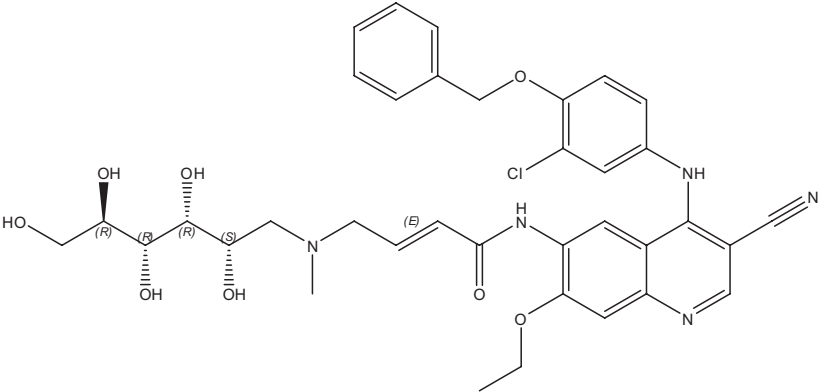
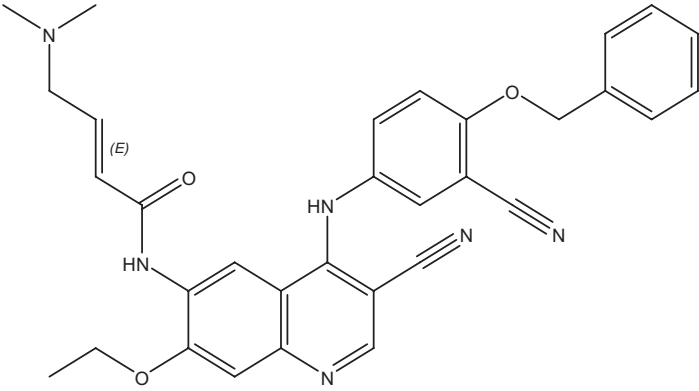
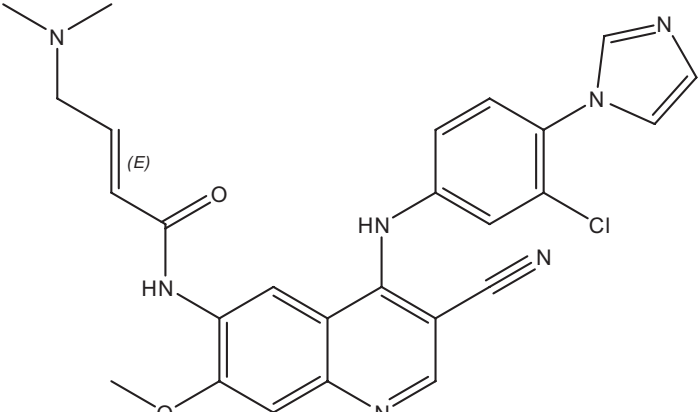
(Continued)

Table S1 (Continued)

Serial number	Structure	EGFR IC ₅₀ (μM)
T-107		0.519
T-108		0.550
T-109		0.574
T-110		0.590
T-111		0.650

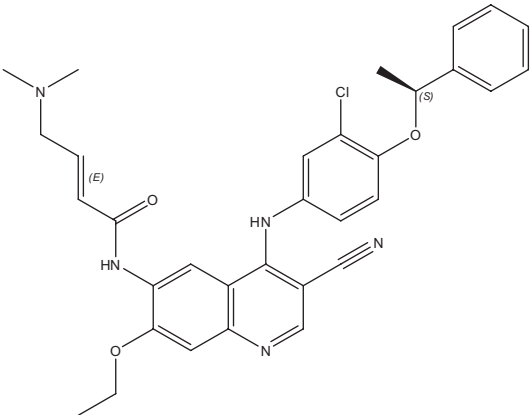
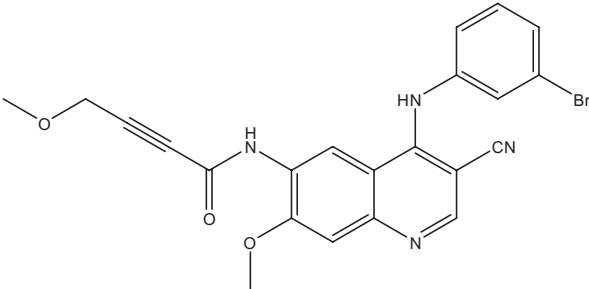
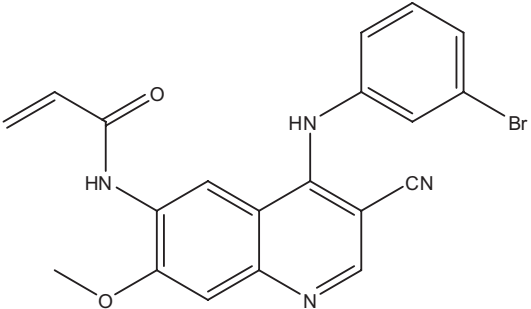
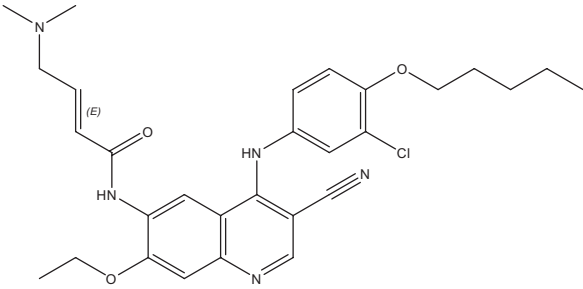
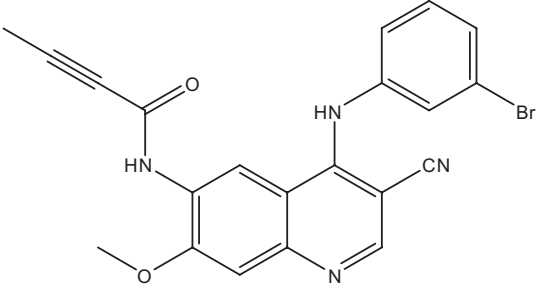
(Continued)

Table SI (Continued)

Serial number	Structure	EGFR IC ₅₀ (μM)
T-112		0.650
T-113		0.719
T-114		0.768
T-115		0.770

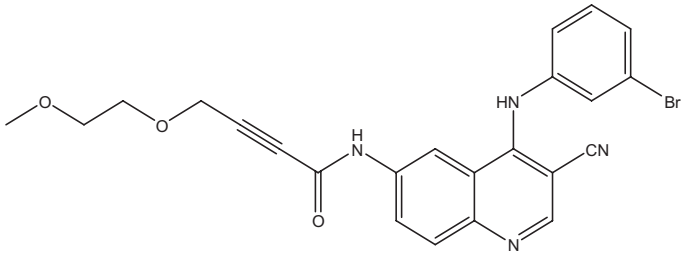
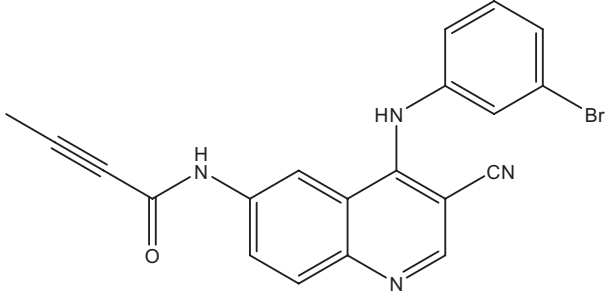
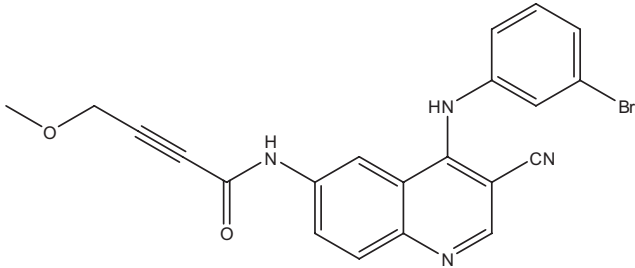
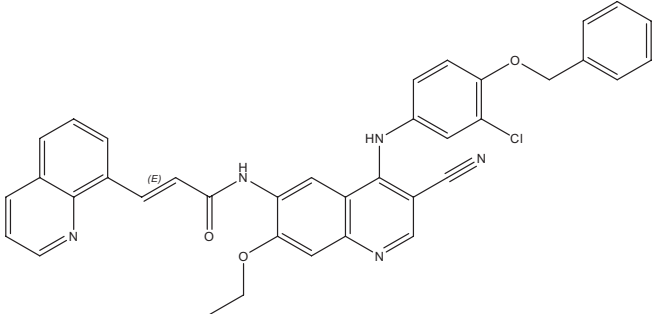
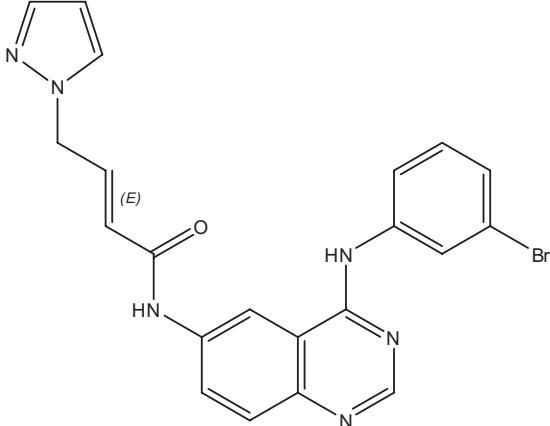
(Continued)

Table S1 (Continued)

Serial number	Structure	EGFR IC ₅₀ (μM)
T-116		0.770
T-117		0.790
T-118		0.810
T-119		0.894
T-120		0.910

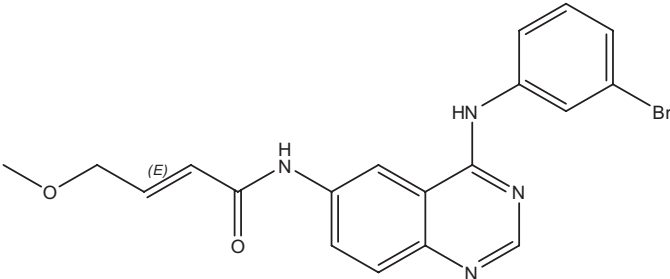
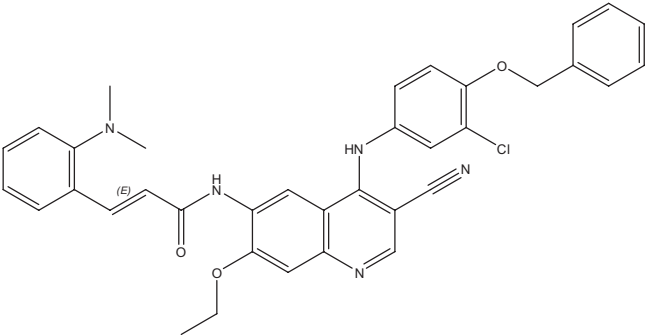
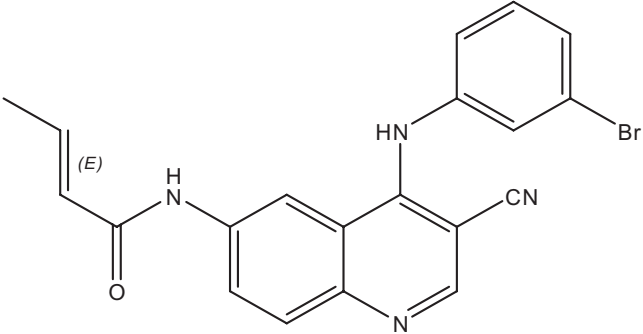
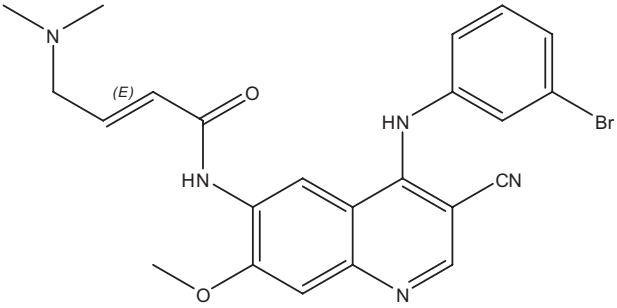
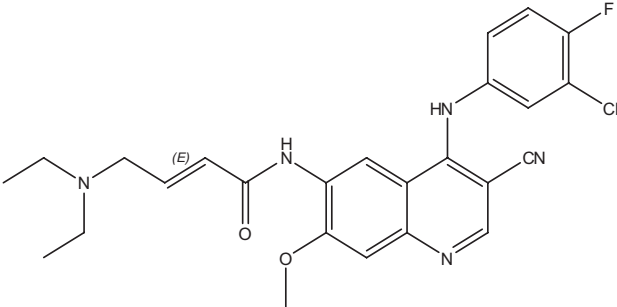
(Continued)

Table SI (Continued)

Serial number	Structure	EGFR IC ₅₀ (μM)
T-121		0.930
T-122		0.940
T-123		0.960
T-124		0.990
T-125		1.124

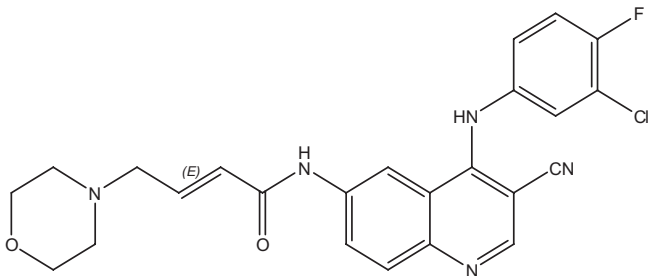
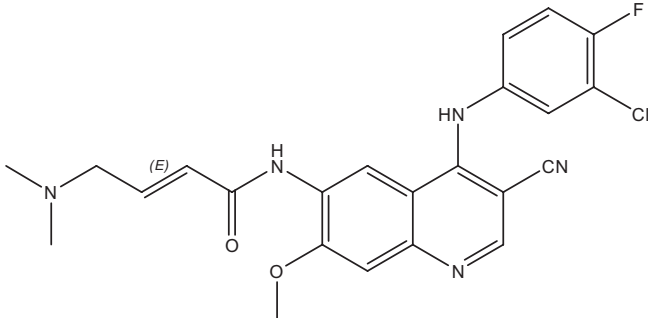
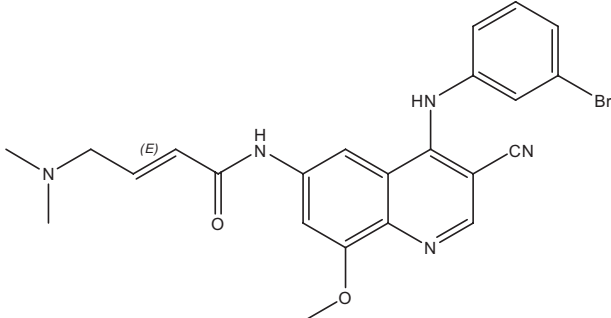
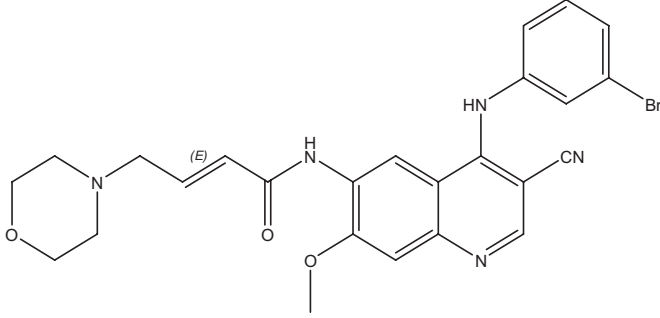
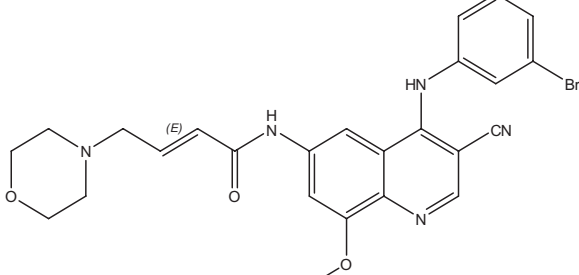
(Continued)

Table S1 (Continued)

Serial number	Structure	EGFR IC ₅₀ (μM)
T-126		1.132
T-127		1.569
T-128		1.620
T-129		1.790
T-130		1.910

(Continued)

Table SI (Continued)

Serial number	Structure	EGFR IC ₅₀ (μM)
T-131		2.290
T-132		2.320
T-133		2.880
T-134		3.830
T-135		5.390

(Continued)

Table S1 (Continued)

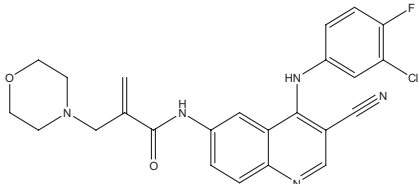
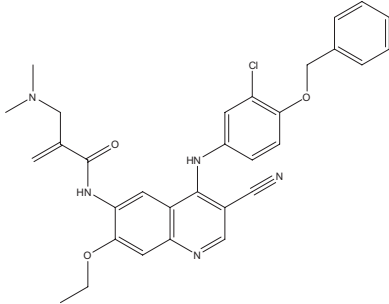
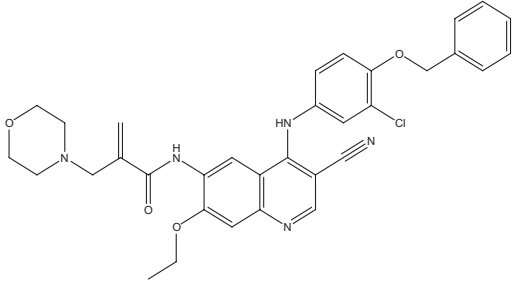
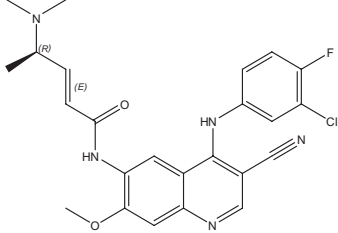
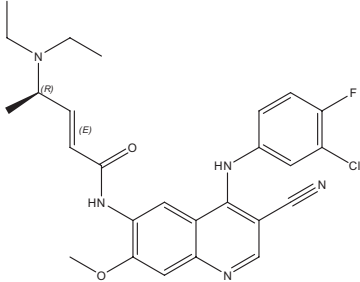
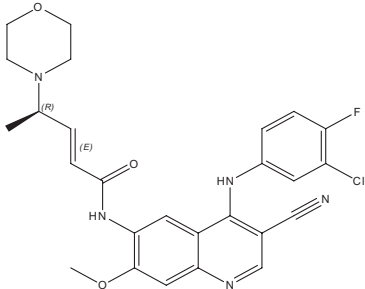
Serial number	Structure	EGFR IC ₅₀ (μM)
T-136		7.500
T-137		12.700
T-138		13.570

Abbreviations: EGFR, epidermal growth factor receptor; IC₅₀, half maximal inhibitory concentration; (S), sinister; (R), rectus; (Z), zusammen; (E), entgegen.

References

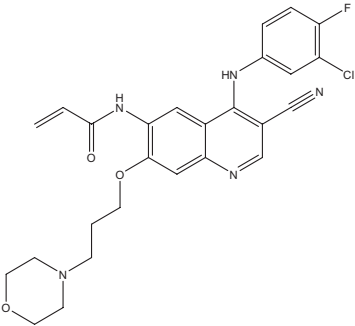
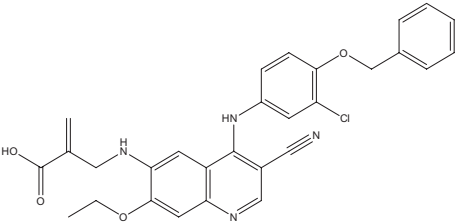
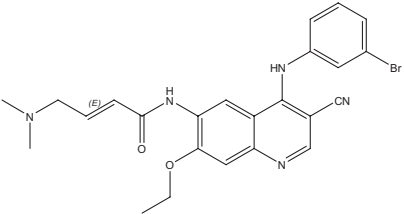
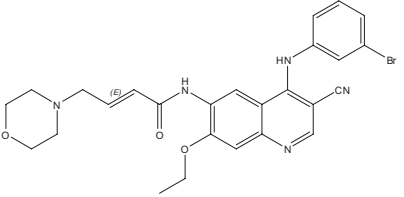
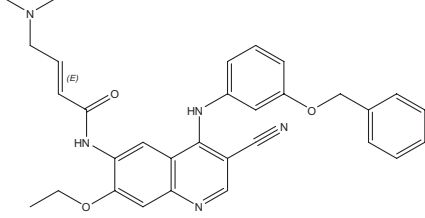
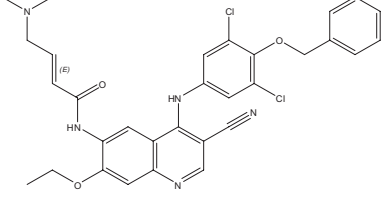
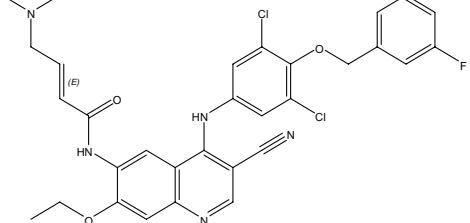
1. Wissner A, Brawner Floyd MB, Rabindran Sridhar K, et al. Syntheses and EGFR and HER-2 kinase inhibitory activities of 4-anilinoquinoline-3-carbonitriles: analogues of three important 4-anilinoquinazolines currently undergoing clinical evaluation as therapeutic antitumor agents. *Bioorg Med Chem Lett*. 2002;12(20):2893–2897.
2. Wissner A, Overbeek E, Reich MF, et al. Synthesis and Structure-Activity Relationships of 6,7-Disubstituted 4-Anilinoquinoline-3-carbonitriles. The Design of an Orally Active, Irreversible Inhibitor of the Tyrosine Kinase Activity of the Epidermal Growth Factor Receptor (EGFR) and the Human Epidermal Growth Factor Receptor-2 (HER-2). *J Med Chem*. 2003;46(1):49–63.
3. Tsou H-R, Overbeek-Klumpers EG, Hallett WA, et al. Optimization of 6,7-disubstituted-4-(arylamino)quinoline-3-carbonitriles as orally active, irreversible inhibitors of human epidermal growth factor receptor-2 kinase activity. *J Med Chem*. 2005;48(4):1107–1131.
4. Tsou H-R, Mamuya N, Johnson BD, et al. 6-Substituted-4-(3-bromophenylamino)quinazolines as Putative Irreversible Inhibitors of the Epidermal Growth Factor Receptor (EGFR) and Human Epidermal Growth Factor Receptor (HER-2) Tyrosine Kinases with Enhanced Antitumor Activity. *J Med Chem*. 2001;44(17):2719–2734.

Table S2 Structures of 21 irreversible¹⁻⁵ and two reversible⁵ TKIs used for testing the effect of the hybrid strategy to identify irreversible inhibitors

Serial number	Structure	T-790M EGFR IC ₅₀ (μM)
V-01		2.700
V-02		0.223
V-03		0.075
V-04		5.220
V-05		4.840
V-06		10.590

(Continued)

Table S2 (Continued)

Serial number	Structure	T-790M EGFR IC ₅₀ (μM)
V-07		0.080
V-08		1.191
V-09		1.600
V-10		10.320
V-11		13.000
V-12		0.163
V-13		0.362

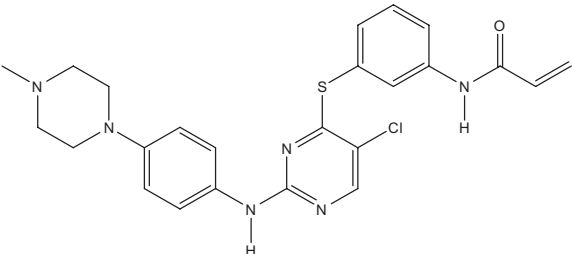
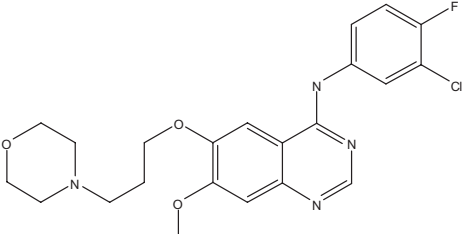
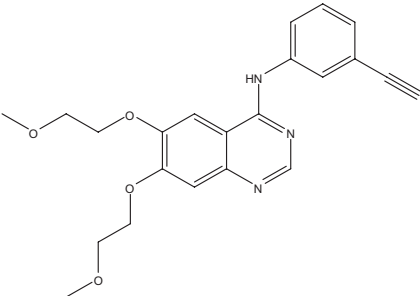
(Continued)

Table S2 (Continued)

Serial number	Structure	T-790M EGFR IC ₅₀ (μM)
V-14		0.340
V-15		0.100
V-16		2.640
V-17		0.560
T-001		0.002
wz-3146		0.029 ⁵
wz-4002		0.047 ⁵

(Continued)

Table S2 (Continued)

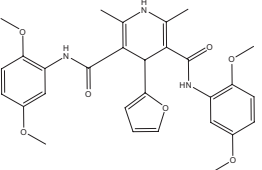
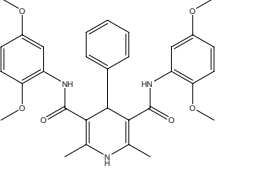
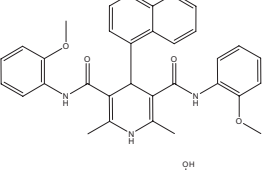
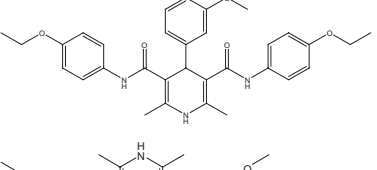
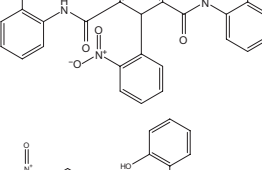
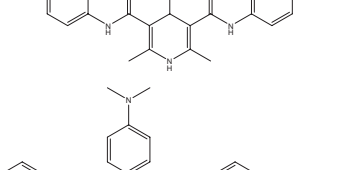
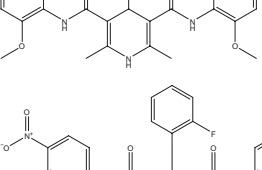
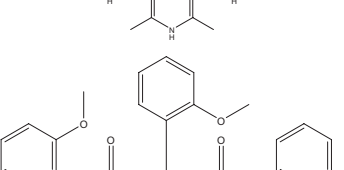
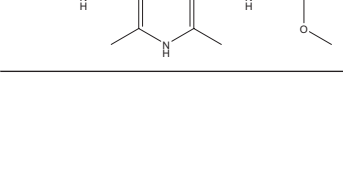
Serial number	Structure	T-790M EGFR IC ₅₀ (μM)
wz-8040		0.009 ⁵
Gefitinib		>3.3 ⁵
Erlotinib		

Abbreviations: TKI, tyrosine kinase inhibitor; EGFR, epidermal growth factor receptor; IC₅₀, half maximal inhibitory concentration; (R), rectus; (E), entgegen.

References

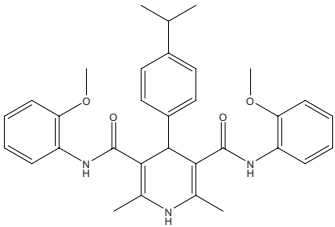
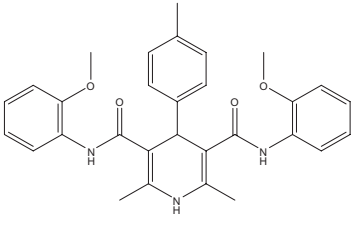
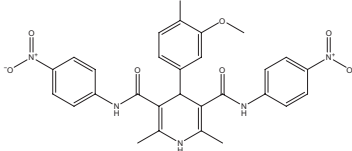
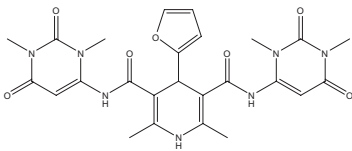
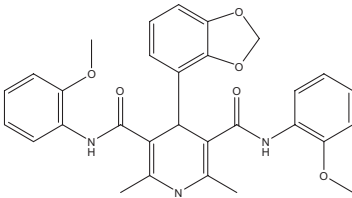
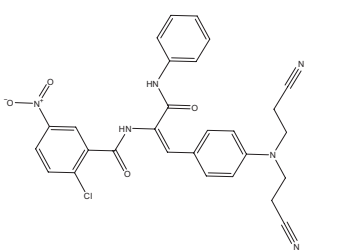
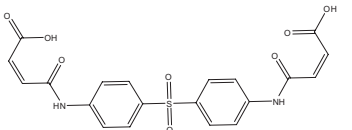
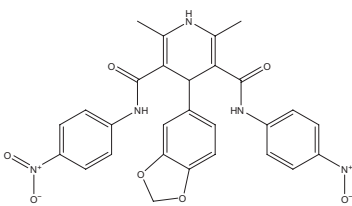
1. Wissner A, Brawner Floyd MB, Rabindran Sridhar K, et al. Syntheses and EGFR and HER-2 kinase inhibitory activities of 4-anilinoquinoline-3-carbonitriles: analogues of three important 4-anilinoquinazolines currently undergoing clinical evaluation as therapeutic antitumor agents. *Bioorg Med Chem Lett.* 2002;12(20):2893–2897.
2. Wissner A, Overbeek E, Reich MF, et al. Synthesis and Structure-Activity Relationships of 6,7-Disubstituted 4-Anilinoquinoline-3-carbonitriles. The Design of an Orally Active, Irreversible Inhibitor of the Tyrosine Kinase Activity of the Epidermal Growth Factor Receptor (EGFR) and the Human Epidermal Growth Factor Receptor-2 (HER-2). *J Med Chem.* 2003;46(1):49–63.
3. Tsou H-R, Overbeek-Klumpers EG, Hallett WA, et al. Optimization of 6,7-disubstituted-4-(arylamino)quinoline-3-carbonitriles as orally active, irreversible inhibitors of human epidermal growth factor receptor-2 kinase activity. *J Med Chem.* 2005;48(4):1107–1131.
4. Tsou H-R, Mamuya N, Johnson BD, et al. 6-Substituted-4-(3-bromophenylamino)quinazolines as Putative Irreversible Inhibitors of the Epidermal Growth Factor Receptor (EGFR) and Human Epidermal Growth Factor Receptor (HER-2) Tyrosine Kinases with Enhanced Antitumor Activity. *J Med Chem.* 2001;44(17):2719–2734.
5. Zhou W, Ercan D, Chen L, et al. Novel mutant-selective EGFR kinase inhibitors against EGFR T790M. *Nature.* 2009;462(7276):1070–1074.

Table S3 The structures of 28 potential irreversible TKIs with three classes of novel scaffolds identified by docking molecules from NCI database (release 3 files) to the binding site of T790M EGFR using FRED 2.2.5¹ (OpenEye Scientific Software Inc., Santa Fe, NM, USA).

Serial no	NSC no	Shapegauss score ²	Structure	Position among the top 200	Classification of scaffold
1	621158	-500.858		42	Class A
2	623897	-494.190		46	Class A
3	630912	-472.511		65	Class A
4	637681	-469.007		72	Class A
5	633920	-462.811		81	Class A
6	637635	-462.406		84	Class A
7	637606	-459.648		89	Class A
8	637638	-458.330		91	Class A
9	630913	-456.322		93	Class A

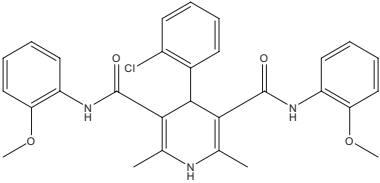
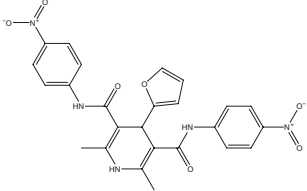
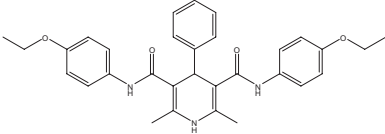
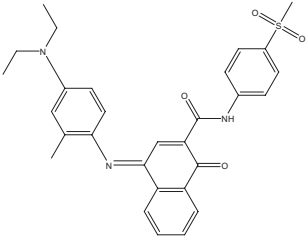
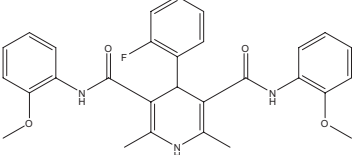
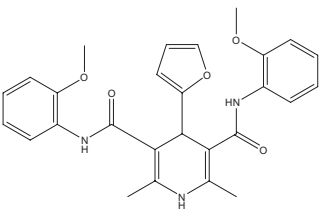
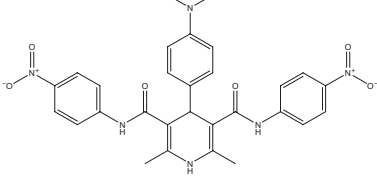
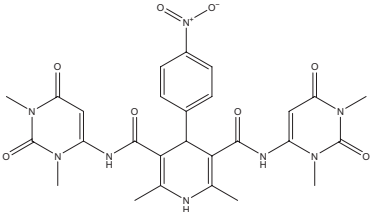
(Continued)

Table S3 (Continued)

Serial no	NSC no	Shapegauss score ²	Structure	Position among the top 200	Classification of scaffold
10	623881	-452.136		98	Class A
11	630908	-452.072		99	Class A
12	637630	-451.936		100	Class A
13	637698	-450.357		104	Class A
14	623867	-450.046		106	Class A
15	642576	-443.251		111	Class B
16	25603	-440.210		114	Class B
17	637614	-436.258		118	Class A

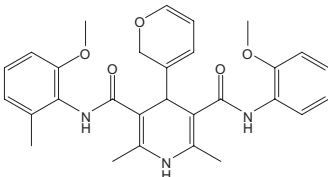
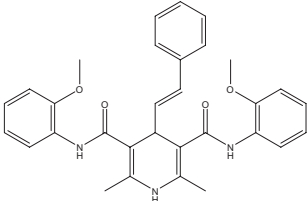
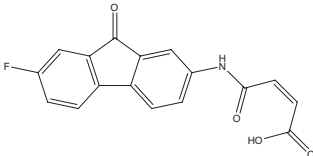
(Continued)

Table S3 (Continued)

Serial no	NSC no	Shapegauss score ²	Structure	Position among the top 200	Classification of scaffold
18	623882	-430.112		122	Class A
19	637605	-429.136		124	Class A
20	637626	-428.069		125	Class A
21	111135	-426.822		126	Class B
22	623877	-421.301		134	Class A
23	623879	-415.805		142	Class A
24	637706	-409.342		152	Class A
25	637700	-408.819		154	Class A

(Continued)

Table S3 (Continued)

Serial no	NSC no	Shapegauss score ²	Structure	Position among the top 200	Classification of scaffold
26	636097	-406.973		160	Class A
27	623875	-400.459		170	Class A
28	57443	-381.532		198	Class C

Abbreviations: TKIs, tyrosine kinase inhibitors; EGFR, epidermal growth factor receptor; NCI, National Cancer Institute; NSC, Nomenclature Standards Committee; IC₅₀, half maximal inhibitory concentration; (S), sinister; (R), rectus; (Z), zusammen; (E), entgegen.

References

1. *FRED program*. <http://www.eyesopen.com> [computer program]. Version 2.2.5: OpenEye Scientific Software Inc., Santa Fe, NM, USA; 2012.
2. Nicholls A, Grant JA, Brown FK. Gaussian Docking Functions. *Biopolymers*. 2003;68:76–90.

Drug Design, Development and Therapy

Publish your work in this journal

Drug Design, Development and Therapy is an international, peer-reviewed open-access journal that spans the spectrum of drug design and development through to clinical applications. Clinical outcomes, patient safety, and programs for the development and effective, safe, and sustained use of medicines are a feature of the journal, which

Submit your manuscript here: <http://www.dovepress.com/drug-design-development-and-therapy-journal>

has also been accepted for indexing on PubMed Central. The manuscript management system is completely online and includes a very quick and fair peer-review system, which is all easy to use. Visit <http://www.dovepress.com/testimonials.php> to read real quotes from published authors.

Dovepress