ACCELERATED PUBLICATION

Erlotinib binds both inactive and active conformations of the EGFR tyrosine kinase domain

Jin H. PARK*†[1], Yingting LIU‡[1], Mark A. LEMMON*†[2] and Ravi RADHAKRISHNAN†‡[2]

*Department of Biochemistry and Biophysics, Perelman School of Medicine, University of Pennsylvania, Philadelphia, PA 19104, U.S.A., †Graduate Group in Biochemistry and Molecular Biophysics, Perelman School of Medicine, University of Pennsylvania, Philadelphia, PA 19104, U.S.A., and ‡Department of Bioengineering, University of Pennsylvania, Philadelphia, PA 19104, U.S.A.

Erlotinib and gefitinib, tyrosine kinase inhibitors used to block EGFR (epidermal growth factor receptor) signalling in cancer, are thought to bind only the active conformation of the EGFR-TKD (tyrosine kinase domain). Through parallel computational and crystallographic studies, we show in the present study that erlotinib also binds the inactive EGFR-TKD conformation, which may have significant implications for its use in EGFR-mutated cancers.

Key words: cancer, epidermal growth factor receptor (EGFR), erlotinib, gefitinib, inhibitor, molecular dynamics, receptor tyrosine kinase (RTK), X-ray crystallography.

INTRODUCTION

Activating mutations in the EGFR (epidermal growth factor receptor) are now well accepted as oncogenic drivers in such cancers as NSCLC (non-small-cell lung cancer) [1] and glioblastoma [2,3]. Several drugs that inhibit EGFR signalling are now in clinical use, including antibody therapeutics and ATP-competitive small molecule TKIs (tyrosine kinase inhibitors) [4]. Approved EGFR-targeted TKIs include erlotinib and gefitinib (which are EGFR-specific), lapatinib (which inhibits both EGFR and HER2/ErbB2) and vandetanib [which inhibits EGFR, VEGFR (vascular endothelial growth factor receptor) and Ret]. Many more are currently being developed and/or tested.

Not all EGFR-driven cancers are sensitive to EGFR-targeted TKIs [4]. In EGFR-mutated NSCLC, for example, primary resistance is seen in approximately 25 % of cases. Where responses are seen, acquired resistance to first-generation TKIs invariably occurs [5], often through secondary EGFR mutations [6,7]. Intriguingly, the selectivity of response to different EGFR-targeted TKIs appears to depend on how the receptor is activated. For example, EGFR activated by TKD (tyrosine kinase domain) mutations in NSCLC can be inhibited by erlotinib and CI-1033 (canertinib) but not by lapatinib or HKI-272 (neratinib), whereas the same receptor activated by extracellular mutations in glioblastoma is conversely inhibited by lapatinib and HKI-272 (neratinib) but not by erlotinib or CI-1033 (canertinib) [3].

Previous structural studies of EGFR-TKD bound to different TKIs [8] suggest that individual drugs differ in their preference for the 'active' compared with 'inactive' conformation of the kinase, although they are all strictly type I TKIs [9] in that they bind EGFR-TKD in the 'DFG (Asp-Phe-Gly)-in' conformation.

Crystal structures of erlotinib-bound [10] or gefitinib-bound [11] EGFR-TKD show the kinase in the same active conformation, allosterically stabilized by a characteristic asymmetric dimer formed in the crystals [12]. By contrast, crystals of EGFR-TKD bound to lapatinib [13] or HKI-272 (neratinib) [7] show a distinct Src-like inactive conformation in which a short activation loop α-helix makes autoinhibitory interactions with the displaced αC helix [12]. On the basis of these studies, erlotinib and gefitinib are thought to bind only the active EGFR-TKD conformation, whereas lapatinib and neratinib are thought to bind only the inactive conformation.

Despite these structural views, biochemical and binding studies argue that EGFR-targeted TKIs may not in fact be so conformationally selective. Extensive direct binding studies suggested that erlotinib, gefitinib and lapatinib all bind similarly (within 3-fold) to both wild-type inactive and mutationally activated forms of EGFR-TKD [14]. Erlotinib and gefitinib affinities were also reported [15] not to be affected by a C-lobe mutation (V924R) that prevents formation of the activated asymmetric dimer [12], again suggesting that these inhibitors bind similarly to inactive and active EGFR-TKD. Moreover, studies of near full-length EGFR [16] indicated that the apparent K_i value for inhibition by lapatinib is increased only 1.8-fold upon receptor activation. It should be noted, however, that some other published data do suggest affinity differences between active and inactive EGFR-TKD for some of these inhibitors. For example, the apparent K_i for lapatinib inhibition of near full-length EGFR was increased >25-fold when activated by the oncogenic L834R mutation [16]. Blocking EGFR activation with cetuximab also increased the apparent K_i for erlotinib [16], suggesting that inactive EGFR has a reduced affinity for erlotinib. Moreover,

one set of direct gefitinib-binding studies suggested that this TKI binds more tightly to the activated (L834R) EGFR-TKD than to the inactive conformation [11]. On balance, it therefore remains unclear whether EGFR inhibitors are truly conformationally selective. It is clear, however, that a full understanding of which TKIs bind more tightly to which mutated EGFR variants found in patients could have significant potential impact on inhibitor choice in the clinic.

To investigate conformational selectivity of EGFR inhibitors further, we undertook computational and crystallographic studies of erlotinib binding to EGFR-TKD in its active and inactive states. Our computational studies predicted that erlotinib binds both inactive and active TKD conformations with similar affinities. We used X-ray crystallography to test this prediction, and describe a structure of inactive EGFR-TKD with bound erlotinib. The findings of the present study cast further doubt on the conformational selectivity of EGFR inhibitors, and also suggest that other characteristics, such as differences in inhibitor dissociation [13] or cycling [17] rates, may underlie the distinct effects of erlotinib and lapatinib on NSCLC and glioblastoma cell lines [3], and will need to be understood to counter TKI resistance.

EXPERIMENTAL

Plasmid construction

DNA encoding kinase domain residues 672–998 (mature protein numbering), equivalent to residues 696–1022 (precursor protein numbering) of human EGFR (NCBI sequence NP_005219.2) was amplified by PCR using primers that included an N-terminal His$_6$ tag and SpeI/XhoI restriction sites. The PCR product was subcloned into pFastBac1(Invitrogen) for generating recombinant baculovirus with the Bac-to-Bac system (Invitrogen) for protein expression in *Spodoptera frugiperda* Sf9 cells. To generate the V924R mutation, the codon for Val^{924} (mature protein numbering) was substituted with an arginine codon using the QuikChange® method (Agilent Technologies).

Protein production and purification

Sf9 cells at $(1.5–2) \times 10^6$/ml were infected with recombinant baculovirus, and harvested by centrifugation after 3 days. Cells expressing histidine-tagged EGFR$^{672-998}$/V924R ($\sim$3 litres of medium) were lysed by sonication in 100 ml of lysis buffer [20 mM Tris/HCl (pH 8.0), containing 500 mM NaCl, 5 mM 2-mercaptoethanol and protease inhibitor cocktail (Roche)]. After centrifugation at 40 000 *g* for 30 min to remove cell debris, the supernatant was passed through a 0.45 μm filter and incubated with Ni-NTA (Ni^{2+}-nitrilotriacetate) agarose beads (Qiagen) for 1 h at 4°C. Beads were washed in 50 column volumes of lysis buffer, and bound EGFR-TKD protein was eluted in lysis buffer containing increasing concentrations of imidazole. Eluted protein was then further purified using an UnoQ anion-exchange column (Bio-Rad Laboratories) equilibrated with 20 mM Tris/HCl (pH 8.0), containing 5% glycerol and 2 mM DTT (dithiothreitol), and eluting with a gradient from 75 mM to 1 M NaCl over 20 column volumes. EGFR-TKD protein was then subjected to a final step of size-exclusion chromatography using a Superdex 200 column (GE Healthcare) equilibrated in 20 mM Tris/HCl (pH 8.0), containing 150 mM NaCl and 2 mM DTT. In total 1–2 mg of purified EGFR$^{672-998}$/V924R protein was typically obtained per litre of Sf9 cell culture.

Table 1   Data collection and refinement statistics (molecular replacement)

Each dataset was collected from a single crystal. Values for highest resolution shell are shown in parentheses.

Parameter	EGFR/erlotinib
Data collection	
Space group	$C222_1$
Cell dimensions	
a, b, c (Å)	78.0, 114.3, 84.9
α, β, γ (°)	90, 90, 90
Resolution (Å)	50–2.75
R_{sym}	0.159 (0.494)
I/σ	12.8 (2.1)
Completeness (%)	96.6 (82.1)
Redundancy	4.8 (2.8)
Refinement	
Resolution (Å)	50–2.75
Number of reflections	9413
R_{work}/R_{free}	0.23/0.25
Number of atoms	
Protein	2201
Ligand	29
Water	34
B-factors	
Protein	47.2
Ligand	47.5
Water	43.3
RMSDs	
Bond lengths (Å)	0.009
Bond angles (°)	1.081

Crystallization and structure determination

Crystals were obtained using the hanging-drop vapour diffusion method, by mixing equal volumes of protein and reservoir solutions and equilibrating over the reservoir solution at 21°C. EGFR-TKD protein was concentrated to 6 mg/ml in 20 mM Tris/HCl (pH 8.0), containing 150 mM NaCl and 2 mM DTT. Crystals were obtained with a reservoir solution of 0.25 M sodium thiocyanate (pH 6.9) and 27% (w/v) PEG [poly(ethylene glycol)] 3350, and when 10 mM taurine had been included as 'additive' in the hanging drop. Crystals were soaked for 2 h at 21°C in mother liquor containing 1 mM erlotinib. Crystals were cryo-protected in reservoir solution with 20% (w/v) glycerol added and flash frozen in liquid nitrogen. Diffraction data were collected at beamline 23ID-D of GM/CA@APS (Advanced Photon Source), where crystals diffracted to 2.75 Å (1 Å = 0.1 nm), and were processed using HKL2000 [18] (see Table 1). The structure was solved by molecular replacement using Phaser [19] with the inactive EGFR (V924R)-TKD structure (PDB code 3GT8 [20]) as the search model. Repeated cycles of manual building/rebuilding using Coot [21] were alternated with rounds of refinement using REFMAC [19,22], plus composite omit maps calculated using CNS [23]. PHENIX [24] and TLS refinement [25] were used in the later stages. Co-ordinates, parameter files and molecular topology of erlotinib were generated by PRODRG [26]. Data collection and refinement statistics are shown in Table 1. One molecule of EGFR$^{672-998}$/V924R is present in the asymmetric unit, and the model of its structure complexed with erlotinib includes amino acids 679–709 and 714–960 (mature EGFR numbering). Structural figures were generated with PyMOL (http://www.pymol.org).

System preparation and molecular docking

Active EGFR-TKD was modelled on the basis of PDB entries 1M17 (which also provided the initial erlotinib conformation) [10] and 2ITX [11], and the L834R mutant was modelled on the basis of PDB entry 2ITV [11]. Inactive EGFR-TKD was modelled based on PDB entries 2GS7 [12] and 1XKK [13]. Protein and ligand conformations were prepared using the Protein Preparation Wizard and LigPrep protocols from Schrödinger Software. All docking simulations used the OPLS (Optimized Potentials for Liquid Simulations) force field [30], and used Schrödinger's IFD (Induced Fit Docking) package [31]. Ligand was first docked to rigid protein using Glide XP [32]. For the resulting top 20 complex conformations, the protein side chains within 5.0 Å of the ligand in that pose were subjected to conformational search and minimized using Prime [33] and the ligand was redocked to the 20 new receptor conformations.

Parameterization of erlotinib for MD (molecular dynamics)

For MD-based analysis of EGFR-TKD–inhibitor interactions, we first generated a CHARMM format force field for erlotinib by following the procedure detailed in the Supplementary Online Data (at http://www.biochemj.org/bj/448/bj4480417add.htm), adding nine new atom types to the CHARMM27 [34] topology file to represent new atom types in erlotinib (see Supplementary Figure S1 at http://www.biochemj.org/bj/448/bj4480417add.htm). Tests of erlotinib parameterization are shown in Supplementary Figure S2 and Supplementary Table S1 (at http://www.biochemj.org/bj/448/bj4480417add.htm).

MD simulations

Conformations generated from IFD were energy-minimized and subsequently equilibrated by performing MD using the CHARMM27 force field [34]. Each system was subjected to constant temperature and constant pressure MD runs at 300 K and 101.325 kPa, followed by constant temperature equilibrium at 300 K with periodic boundaries enforced and long-range electrostatics taken into consideration for 10 ns.

MM/PBSA (molecular mechanics/Poisson–Boltzmann surface area) calculation

We used MM/PBSA [35,36] to calculate the absolute binding free energy and compared it with the Glide [32] score based on the 10 ns MD simulation. MM/PBSA energies of each modelled complex conformation were calculated as the average of the single-point MM/PBSA energy of 500 snapshots taken from the 10 ns simulation. Water molecules and salt ions were removed from the trajectory before calculation. The molecular mechanics energy, U_{MM}, was evaluated by averaging energies over all structures using an infinite cut-off for non-bonded interactions. The electrostatic contribution to the solvation free energy, W_{PB}, was calculated using Poisson–Boltzmann Solver in CHARMM [37]. The reference system had a solvent dielectric constant of 1 and salt concentration of 0 M, and the solvated system had a dielectric constant of 80 and salt concentration of 100 mM. Non-polar contribution to the solvation free energy, W_{SA}, was approximated with the surface area model $\Delta W_{SA} = [0.00542$ kcal/mol/Å$^2] \times SASA + 0.92$ kcal/mol [38], where the SASA (surface-accessible surface area) was estimated using a 1.4 Å solvent probe radius. Terms for entropy changes in the MM/PBSA score were neglected.

Table 2   Erlotinib-binding energies calculated by Glide, MM/PBSA and FEP for wild-type and L834R EGFR-TKD in the active conformation

Molecule	Glide [32] score (kcal/mol)	MM/PBSA (kcal/mol)	FEP (kcal/mol)
Wild-type	− 9.34	− 28.2	
L834R	− 9.35	− 30.7	
$\Delta\Delta G$	− 0.01	− 2.5	0.83 ± 1.16

Erlotinib in active EGFR-TKD *model*
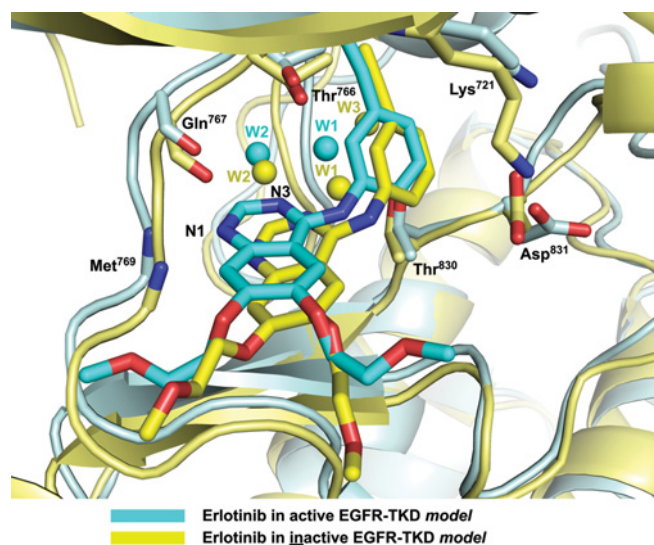Erlotinib in *inactive* EGFR-TKD *model*

Figure 1   Comparison of erlotinib binding to active and inactive EGFR-TKD models

On the basis of the computational models described in the present study, erlotinib is shown bound to active EGFR-TKD (cyan) and inactive EGFR-TKD (yellow). The protein structure in the background is similarly coloured (cyan for active, yellow for inactive). Functional groups in several EGFR-TKD residues that interact directly or indirectly with the bound erlotinib are shown. The backbone amide of Met[769] donates a hydrogen bond to N1 of the erlotinib quinazoline moiety. The backbone carbonyl of Gln[767] and side chains of Thr[766] and Thr[830] participate in a hydrogen-bonding network to which water molecules (W1 and W2) also contribute. The Lys[721] and Asp[831] side chains are shown for reference. Polypeptide in the foreground has been removed for clarity.

RESULTS AND DISCUSSION

Computational analysis of erlotinib binding to EGFR-TKD

To investigate the possible basis for preferential binding of erlotinib to the active conformation of EGFR-TKD, we used computational approaches. Docking erlotinib on to wild-type EGFR-TKD in its active conformation, as described in the Experimental section, closely reproduced the binding mode observed crystallographically in PDB entry 1M17 [10] (see Supplementary Online Data and Supplementary Figure S4 at http://www.biochemj.org/bj/448/bj4480417add.htm). Erlotinib was stable in this bound conformation (cyan in Figure 1) throughout a 10 ns MD run (Supplementary Figure S5 at http://www.biochemj.org/bj/448/bj4480417add.htm). Notably, two water molecules (W1 and W2 in Figure 1) were also found to remain in the binding pocket, contributing to a stable network of hydrogen bonds involving residues Thr[766], Gln[767] and Thr[830] of EGFR-TKD. Docking erlotinib into L834R-mutated EGFR-TKD gave very similar results (see Supplementary Online Data). Moreover, estimates of binding energy using Glide [32], MM/PBSA calculations (see the Experimental section) or FEP
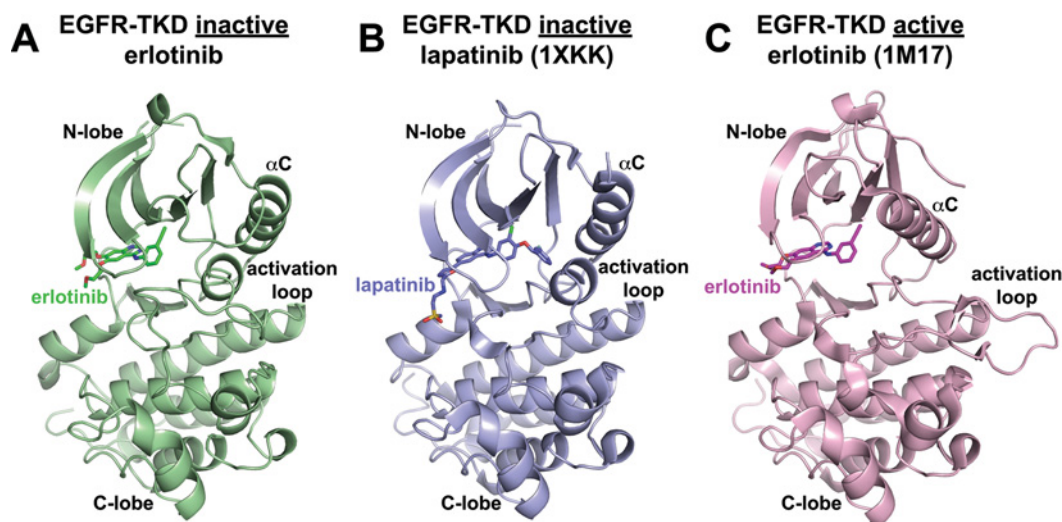
Figure 2   Crystal structures of TKI-bound EGFR-TKD

(**A**) The 2.75 Å resolution structure of V924R-mutated EGFR-TKD in its inactive form bound to erlotinib (green) as described in the text and Table 1. (**B**) Crystal structure from PDB entry 1XKK [13] showing lapatinib (blue) bound to inactive EGFR-TKD. (**C**) Crystal structure from PDB entry 1M17 [10] showing erlotinib (magenta) bound to active EGFR-TKD. Note that the αC helix is in the 'in' or 'active' position in (**C**), but in the 'out' or 'inactive' position in (**A**) and (**B**), and that the characteristic short α-helix seen in the inactive EGFR-TKD activation loop is present in (**A**) and (**B**), but not (**C**). As described in the text, the crystal structure shown in (**A**) confirms our computational findings that erlotinib can bind to the inactive EGFR-TKD conformation.

(free energy perturbation) calculations (see Supplementary Online Data) indicated no increase in erlotinib binding affinity to the L834R mutant (Table 2).

Computational studies of erlotinib binding to inactive EGFR-TKD

To challenge the assumption that erlotinib, gefitinib and other related inhibitors only bind to (and stabilize) the active EGFR-TKD conformation, we used the same computational approaches to ask whether erlotinib can bind to the inactive conformation of the kinase. Surprisingly, docking erlotinib into a crystal structure of the inactive EGFR-TKD (PDB entry 1XKK [13]), after removing lapatinib from the model, yielded a very similar Glide [32] score (-9.72 kcal/mol) to that seen for active EGFR-TKD (-9.34 kcal/mol). Moreover, when the two models were overlaid (Figure 1), the orientation of erlotinib and the conformation of its binding site were very similar in active (cyan) and inactive (yellow) EGFR-TKD models. N1 of the erlotinib quinazoline moiety receives a hydrogen bond from the amide nitrogen of Met[769] in both cases. The two water molecules (W1 and W2) mentioned above form essentially the same hydrogen-bonding network in each model, which also involves the Gln[767] backbone carbonyl, the Thr[766] side chain, nitrogen N3 of the erlotinib quinazoline moiety and the Thr[830] side chain (Figure 1). An additional third strongly bound water molecule (W3 in Figure 1) was seen in the model of erlotinib-bound inactive EGFR-TKD. Distances between erlotinib and protein (and between bound waters and erlotinib) were stable throughout a 10 ns MD simulation, as plotted in Supplementary Figure S7 (at http://www.biochemj.org/bj/448/bj4480417add.htm). Thus our computational studies failed to provide any explanation for why erlotinib might bind preferentially to the active conformation of EGFR-TKD. Yun et al. [11] suggested that favourable van der Waals interactions between gefitinib and the back of the ATP-binding cleft in active EGFR-TKD may be lost in the inactive state, but this was not seen in our modelling.

Crystal structure of erlotinib bound to inactive EGFR-TKD

Following the unexpected suggestion from our computational studies that erlotinib binds similarly to active and inactive EGFR-TKD, we sought to crystallize the kinase domain in its inactive conformation with erlotinib bound. Previous crystal structures of active EGFR-TKD with erlotinib [10] or gefitinib [11] bound were obtained by soaking the drug into pre-formed crystals. However, it is now clear that wild-type EGFR-TKD (or variants harbouring NSCLC mutations) always adopt the active conformation in crystals [10,12], because of EGFR-TKD's intrinsic tendency to form asymmetric dimers in which the C-lobe of one molecule associates with the N-lobe of its neighbour to stabilize the active conformation. Although EGFR-TKD co-crystallized with lapatinib [13], HKI-272 [7] or related inhibitors adopt the inactive conformation, wild-type EGFR-TKD only crystallizes in the inactive conformation when the asymmetric dimer is disrupted by a mutation in its interface (such as V924R in the C-lobe) [12,20] or by a Mig6-derived peptide that binds to the C-lobe dimerization site [39].

By soaking erlotinib into crystals formed by a V924R-mutated EGFR-TKD variant, we were able to determine the structure of erlotinib bound to inactive EGFR-TKD at a resolution of 2.75 Å (Table 1). As shown in Figure 2, our structure of inactive EGFR-TKD closely resembles that seen in PDB entry 1XKK [13] (in complex with lapatinib). The RMSD (root mean square deviation) between α carbon positions in these two structures was 1.30 Å, and the α carbon RMSD for overlay of our inactive structure with PDB entry 3GT8 [20] was 1.42 Å. As illustrated in Figure 3, the crystal structure shows erlotinib (green in Figure 3A) to be slightly shifted in the binding site compared with its predicted position in the model (yellow in Figure 3A); to the left in the aspect of the Figure by approximately 1 Å. The distance between the amide nitrogen of Met[769] in EGFR-TKD and N1 of the erlotinib quinazoline moiety is maintained at 2.9 Å in the crystal structure and 3.2 Å in the model. The slight shift of erlotinib is associated with a similar shift in the polypeptide
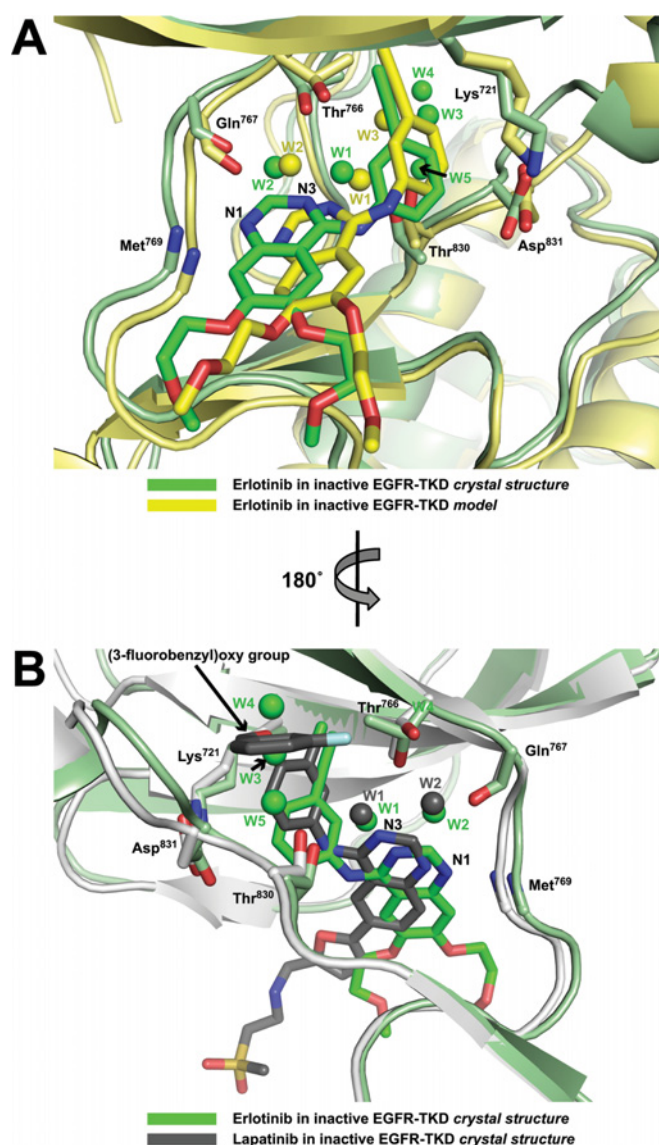
Figure 3   Erlotinib binding to inactive EGFR-TKD

(**A**) The crystallographically observed mode of erlotinib binding to inactive EGFR-TKD (green) is compared with the computational model (yellow). Protein structure in the background coloured green for the crystal structure and yellow for the model. (**B**) The mode of erlotinib binding to inactive EGFR-TKD observed crystallographically (green) is compared with the mode of lapatinib binding to inactive EGFR-TKD in PDB entry 1XKK [13] (grey), in a view rotated by 180° about a vertical axis compared with that seen in (**A**). Waters W1–W5 are labelled for the erlotinib/inactive EGFR-TKD structure. Waters W1 and W2 are also found in the lapatinib complex [13]. W3–W5 in the erlotinib complex structure lie in a pocket occupied by the (3-fluorobenzyl)oxy group of lapatinib in PDB entry 1XKK. Functional groups, side chains and bound waters are shown as described in Figure 1. Polypeptide in the foreground has been removed for clarity.

backbone in the region of residues 767–770. Other key side chains (Thr^{766}, Thr^{830}, Asp^{831} and Lys^{721}) shown in Figure 3(A) are located very similarly in the model and crystal structure. Importantly, the pattern of water molecules predicted by our model was also observed crystallographically (compare yellow and green waters in Figure 3A). W1 and W2 are in very similar positions in the crystal structure and model, providing crystallographic evidence that the predicted water-mediated hydrogen-bonding network involving the quinazoline moiety of erlotinib, and Thr^{766}/Gln^{767}/Thr^{830} of EGFR-TKD is maintained in the inactive

conformation. Interestingly, equivalent waters are also seen in the published crystal structure of lapatinib-bound inactive EGFR-TKD (Figure 3B) [13]. Moreover, a crystallographic water was seen in our erlotinib/inactive EGFR-TKD structure close to the position of W3 in the computational model, and near the side chains of Thr^{766} and Thr^{830}. Additional waters labelled W4 and W5 in Figure 3(A) are seen in our erlotinib/EGFR-TKD crystal structure and not the model. Interestingly, as shown in Figure 3(B), waters W3–W5 in our structure lie in the same pocket within inactive EGFR-TKD that is occupied by the (3-fluorobenzyl)oxy group of lapatinib in PDB entry 1XKK [13]; these waters also interact with the EGFR-TKD backbone in parts of the αC helix (Met^{742}) and Phe^{832} of the DFG motif.

Conclusions

It is generally assumed that the EGFR inhibitors gefitinib and erlotinib bind selectively to the active conformation of EGFR-TKD, whereas lapatinib selectively binds the inactive configuration. A preference of lapatinib for inactive EGFR-TKD can be rationalized, since it has a bulky [(3-fluorobenzyl)oxy] substituent on its aniline ring that projects into the space opened up upon displacement of the αC helix in the inactive kinase (as shown in Figure 2B). For erlotinib, however, no significant difference in the binding affinity for inactive and active EGFR-TKD conformations could be detected using a range of computational approaches. This led us to confirm crystallographically that erlotinib can indeed bind to EGFR-TKD in its inactive conformation (Figures 2A and 3). These findings have several important implications.

First, our findings complicate suggestions in the literature that erlotinib and gefitinib drive EGFR dimerization by stabilizing the active configuration. Since asymmetric dimerization of EGFR-TKD promotes its acquisition of the active conformation, the converse should also be true: that stabilization of the active TKD conformation drives dimerization. Thus if erlotinib and gefitinib selectively bind and stabilize the active EGFR-TKD conformation, their binding should drive asymmetric dimer formation. Indeed, it has been reported that gefitinib stabilizes formation of EGFR-containing heterodimers in cells [40]. Moreover, Springer and colleagues have reported in electron microscopy studies [15,41] that gefitinib promotes dimerization of near full-length EGFR by inducing structures that resemble asymmetric EGFR-TKD dimers. The finding in the present study that erlotinib also binds to inactive (monomeric) EGFR-TKD is difficult to reconcile with the simple interpretation of these reports, as are other published data indicating that neither stabilizing the active conformation of EGFR-TKD with oncogenic mutations nor disrupting the asymmetric dimer interface with a V924R mutation (and thus stabilizing the inactive conformation) alters the affinity of EGFR-TKD for gefitinib or erlotinib [14,15].

Secondly, the findings of the present study argue that the relative abilities of erlotinib/gefitinib-type EGFR TKIs and lapatinib/neratinib TKIs to inhibit different mutationally activated EGFR variants is more complicated than previously thought. Given the results of the present study, the fact that NSCLC mutants and glioblastoma mutants are selectively inhibited by erlotinib and lapatinib respectively [3] seems unlikely simply to reflect stabilization of different conformational states of EGFR-TKD in the two cancers (active for NSCLC, inactive for glioblastoma). Rather, the different sensitivities of the EGFR mutants are likely to reflect more complicated conformational, and dynamic, characteristics, as suggested by a previous study [17]. It is clear, for example, that lapatinib dissociates from EGFR-TKD much

more slowly than erlotinib or gefitinib [13] (although covalent inhibitors dissociate even more slowly, yet still show differential effects). More extensive analysis is required to identify the differences between these groups of inhibitors, which ultimately should aid in tailoring the type of inhibitor used clinically to the mode in which EGFR has been aberrantly activated.

AUTHOR CONTRIBUTION

Jin Park performed all of the experimental and crystallographic work, guided by Mark Lemmon. Yingting Liu performed all computational modelling work guided by Ravi Radhakrishnan. Jin Park, Yingting Liu, Mark Lemmon and Ravi Radhakrishnan all interpreted results and prepared the paper.

REFERENCES

1   Sharma, S. V., Bell, D. W., Settleman, J. and Haber, D. A. (2007) Epidermal growth factor receptor mutations in lung cancer. Nat. Rev. Cancer **7**, 169–181

2   Cancer Genome Atlas Research Network (2008) Comprehensive genomic characterization defines human glioblastoma genes and core pathways. Nature **455**, 1061–1068

3   Vivanco, I., Robins, H. I., Rohle, D., Campos, C., Grommes, C., Nghiemphu, P. L., Kubek, S., Oldrini, B., Chheda, M. G., Yannuzzi, N. et al. (2012) Differential sensitivity of glioma-versus lung cancer-specific EGFR mutations to EGFR kinase inhibitors. Cancer Discovery **2**, 458–471

4   Pao, W. and Chmielecki, J. (2010) Rational, biologically based treatment of EGFR-mutant non-small-cell lung cancer. Nat. Rev. Cancer **10**, 760–774

5   Jänne, P. A., Gray, N. and Settleman, J. (2009) Factors underlying sensitivity of cancers to small-molecule kinase inhibitors. Nat. Rev. Drug Discovery **8**, 709–723

6   Yoshikawa, S., Kukimoto-Niino, M., Parker, L., Handa, N., Terada, T., Fujimoto, T., Terazawa, Y., Wakiyama, M., Sato, M., Sano, S. et al. (2012) Structural basis for the altered drug sensitivities of non-small cell lung cancer-associated mutants of human epidermal growth factor receptor. Oncogene, doi:10.1038/onc.2012.21

7   Yun, C. H., Mengwasser, K. E., Toms, A. V., Woo, M. S., Greulich, H., Wong, K. K., Meyerson, M. and Eck, M. J. (2008) The T790M mutation in EGFR kinase causes drug resistance by increasing the affinity for ATP. Proc. Natl. Acad. Sci. U.S.A. **105**, 2070–2075

8   Eck, M. J. and Yun, C. H. (2010) Structural and mechanistic underpinnings of the differential drug sensitivity of EGFR mutations in non-small cell lung cancer. Biochim. Biophys. Acta **1804**, 559–566

9   Liu, Y. and Gray, N. S. (2006) Rational design of inhibitors that bind to inactive kinase conformations. Nat. Chem. Biol. **2**, 358–364

10 Stamos, J., Sliwkowski, M. X. and Eigenbrot, C. (2002) Structure of the epidermal growth factor receptor kinase domain alone and in complex with a 4-anilinoquinazoline inhibitor. J. Biol. Chem. **277**, 46265–46272

11 Yun, C. H., Boggon, T. J., Li, Y., Woo, M. S., Greulich, H., Meyerson, M. and Eck, M. J. (2007) Structures of lung cancer-derived EGFR mutants and inhibitor complexes: mechanism of activation and insights into differential inhibitor sensitivity. Cancer Cell **11**, 217–227

12 Zhang, X., Gureasko, J., Shen, K., Cole, P. A. and Kuriyan, J. (2006) An allosteric mechanism for activation of the kinase domain of epidermal growth factor receptor. Cell **125**, 1137–1149

13 Wood, E. R., Truesdale, A. T., McDonald, O. B., Yuan, D., Hassell, A., Dickerson, S. H., Ellis, B., Pennisi, C., Horne, E., Lackey, K. et al. (2004) A unique structure for epidermal growth factor receptor bound to GW572016 (Lapatinib): relationships among protein conformation, inhibitor off-rate, and receptor activity in tumor cells. Cancer Res. **64**, 6652–6659

14 Fabian, M. A., Biggs, 3rd, W. H., Treiber, D. K., Atteridge, C. E., Azimioara, M. D., Benedetti, M. G., Carter, T. A., Ciceri, P., Edeen, P. T., Floyd, M. et al. (2005) A small molecule-kinase interaction map for clinical kinase inhibitors. Nat. Biotechnol. **23**, 329–336

15 Lu, C., Mi, L. Z., Schürpf, T., Walz, T. and Springer, T. A. (2012) Mechanisms for kinase-mediated dimerization of the EGF receptor. J. Biol. Chem., doi:10.1074/jbc.M112.414391

16 Wang, Z., Longo, P. A., Tarrant, M. K., Kim, K., Head, S., Leahy, D. J. and Cole, P. A. (2011) Mechanistic insights into the activation of oncogenic forms of EGF receptor. Nat. Struct. Mol. Biol. **18**, 1388–1393

17 Barkovich, K. J., Hariono, S., Garske, A. L., Zhang, J., Blair, J. A., Fan, Q. W., Shokat, K. M., Nicolaides, T. and Weiss, W. A. (2012) Kinetics of inhibitor cycling underlie therapeutic disparities between EGFR-driven lung and brain cancers. Cancer Discovery **2**, 450–457

18 Otwinowski, Z. and Minor, W. (1997) Processing of X-ray diffraction data collected in oscillation mode. Methods Enzymol. **276**, 307–326

19 Collaborative Computational Project, Number 4 (1994) The CCP4 suite: programs for protein crystallography. Acta Crystallogr. Sect. D Biol. Crystallogr. **50**, 760–763

20 Jura, N., Endres, N. F., Engel, K., Deindl, S., Das, R., Lamers, M. H., Wemmer, D. E., Zhang, X. and Kuriyan, J. (2009) Mechanism for activation of the EGF receptor catalytic domain by the juxtamembrane segment. Cell **137**, 1293–1307

21 Emsley, P. and Cowtan, K. (2004) Coot: model-building tools for molecular graphics. Acta Crystallogr. Sect. D Biol. Crystallogr. **60**, 2126–2132

22 Murshudov, G. N., Skubák, P., Lebedev, A. A., Pannu, N. S., Steiner, R. A., Nicholls, R. A., Winn, M. D., Long, F. and Vagin, A. A. (2011) REFMAC5 for the refinement of macromolecular crystal structures. Acta Crystallogr. Sect. D Biol. Crystallogr. **67**, 355–367

23 Brunger, A. T., Adams, P. D., Clore, G. M., DeLano, W. L., Gros, P., Grosse-Kunstleve, R. W., Jiang, J. S., Kuszewski, J., Nilges, M., Pannu, N. S. et al. (1998) Crystallography & NMR system: a new software suite for macromolecular structure determination. Acta Crystallogr. Sect. D Biol. Crystallogr. **54**, 905–921

24 Adams, P. D., Afonine, P. V., Bunkóczi, G., Chen, V. B., Davis, I. W., Echols, N., Headd, J. J., Hung, L.-W., Kapral, G. J., Grosse-Kunstleve, R. W. et al. (2010) PHENIX: a comprehensive Python-based system for macromolecular structure solution. Acta Crystallogr. Sect. D Biol. Crystallogr. **66**, 213–221

25 Winn, M. D., Isupov, M. N. and Murshudov, G. N. (2001) Use of TLS anisotropic displacements in macromolecular refinement. Acta Crystallogr. Sect. D Biol. Crystallogr. **57**, 122–133

26 Schüttelkopf, A. W. and van Aalten, D. M. (2004) PRODRG: a tool for high-throughput crystallography of protein-ligand complexes. Acta Crystallogr. Sect. D Biol. Crystallogr. **60**, 1355–1363

27 Reference deleted

28 Reference deleted

29 Reference deleted

30 Jorgensen, W. L. and Tiradorives, J. (1988) The OPLS potential functions for proteins: energy minimizations for crystals of cyclic-peptides and crambin. J. Am. Chem. Soc. **110**, 1657–1666

31 Sherman, W., Day, T., Jacobson, M. P., Friesner, R. A. and Farid, R. (2006) Novel procedure for modeling ligand/receptor induced fit effects. J. Med. Chem. **49**, 534–553

32 Friesner, R. A., Murphy, R. B., Repasky, M. P., Frye, L. L., Greenwood, J. R., Halgren, T. A., Sanschagrin, P. C. and Mainz, D. T. (2006) Extra Precision Glide: docking and scoring incorporating a model of hydrophobic enclosure for protein-ligand complexes. J. Med. Chem. **49**, 6177–6196

33 Jacobson, M. P., Pincus, D. L., Rapp, C. S., Day, T. J. F., Honig, B., Shaw, D. E. and Friesner, R. A. (2004) A hierarchical approach to all-atom protein loop prediction. Proteins **55**, 351–367

34 Foloppe, N. and MacKerell, A. D. (2000) All-atom empirical force field for nucleic acids: I. Parameter optimization based on small molecule and condensed phase macromolecular target data. J. Comput. Chem. **21**, 86–104

35 Wang, J., Morin, P., Wang, W. and Kollman, P. A. (2001) Use of MM-PBSA in reproducing the binding free energies to HIV-1 RT of TIBO derivatives and predicting the binding mode to HIV-1 RT of efavirenz by docking and MM-PBSA. J. Am. Chem. Soc. **123**, 5221–5230

36 Swanson, J. M. J., Henchman, R. H. and McCammon, J. A. (2004) Revisiting free energy calculations: a theoretical connection between MM/PBSA and direct calculation of the association free energy. Biophys. J. **86**, 67–74

37   Im, W., Beglov, D. and Roux, B. (1998) Continuum Solvation Model: computation of electrostatic forces from numerical solutions to the Poisson-Boltzmann equation. Comput. Phys. Commun. **111**, 59–75

38   Sanner, M. F., Olson, A. J. and Spehner, J. C. (1996) Reduced surface: an efficient way to compute molecular surfaces. Biopolymers **38**, 305–320

39   Zhang, X., Pickin, K. A., Bose, R., Jura, N., Cole, P. A. and Kuriyan, J. (2007) Inhibition of the EGF receptor by binding of MIG6 to an activating kinase domain interface. Nature **450**, 741–744

40   Anido, J., Matar, P., Albanell, J., Guzmán, M., Rojo, F., Arribas, J., Averbuch, S. and Baselga, J. (2003) ZD1839, a specific epidermal growth factor receptor (EGFR) tyrosine kinase inhibitor, induces the formation of inactive EGFR/HER2 and EGFR/HER3 heterodimers and prevents heregulin signaling in HER2-overexpressing breast cancer cells. Clin. Cancer Res. **9**, 1274–1283

41   Mi, L. Z., Lu, C., Li, Z., Nishida, N., Walz, T. and Springer, T. A. (2011) Simultaneous visualization of the extracellular and cytoplasmic domains of the epidermal growth factor receptor. Nat. Struct. Mol. Biol. **18**, 984–989

SUPPLEMENTARY ONLINE DATA
Erlotinib binds both inactive and active conformations of the EGFR tyrosine kinase domain

Jin H. PARK*†[1], Yingting LIU‡[1], Mark A. LEMMON*†[2] and Ravi RADHAKRISHNAN†‡[2]

*Department of Biochemistry and Biophysics, Perelman School of Medicine, University of Pennsylvania, Philadelphia, PA 19104, U.S.A., †Graduate Group in Biochemistry and Molecular Biophysics, Perelman School of Medicine, University of Pennsylvania, Philadelphia, PA 19104, U.S.A., and ‡Department of Bioengineering, University of Pennsylvania, Philadelphia, PA 19104, U.S.A.

EXPERIMENTAL

Erlotinib parameterization

The CHARMM27 force field [1] consists of the following terms:

$$U(\vec{R}) = \sum_{\text{bonds}} K_b(b - b_0)^2 + \sum_{\text{angles}} K_\theta(\theta - \theta_0)^2$$
$$+ \sum_{\text{dihedrals}} K_\chi[1 + \cos(n\chi - \delta)]$$
$$+ \sum_{\text{nonbond}} \varepsilon_{ij}\left[\left(\frac{R_{ij}^{\min}}{r_{ij}}\right)^{12} - \left(\frac{R_{ij}^{\min}}{r_{ij}}\right)^{6}\right] + \sum_{\text{nonbond}} \frac{q_i q_j}{D r_{ij}}$$

where K_b, K_θ and K_χ are the bond, angle and dihedral force constants; b_0, θ_0 and χ_0 represent the equilibrium value of bond, angle and dihedral; $R_{ij}^{\min}$ and ε_{ij} are, respectively, the distance between atoms i and j at which the LJ (Lennard–Jones) potential is zero and the depth of the LJ potential well for the same pair of atoms; D is the effective dielectric constant; and q_i is the partial atomic charge on atom i. All of these parameters need to be defined for each atom type of erlotinib.

An erlotinib molecule is depicted in Figure S1, with atom types labelled. As indicated in Figure S1, parameterization of erlotinib required nine new atom types to be added to the CHARMM topology file. Initial partial atomic charges (q_i) for these atom types were calculated using a CHELPG (CHarges from ELectrostatic Potentials using a Grid)-based method [2] in the *ab initio* electronic structure package GAUSSIAN [3], by fitting the molecular mechanics-derived electrostatic potential to that obtained quantum mechanically. The van der Waals constants ($R_{ij}^{\min}$ and ε_{ij}) were transferred from existing CHARMM parameters and were not modified during refinement as their values depend mostly on atomic properties and are transferable to the molecular environment. The equilibrium constants (b_0 and θ_0) were obtained from optimized structures of erlotinib based on its conformation in an erlotinib/EGFR-TKD crystal structure [4] using *ab initio* calculations, and were not changed during optimization. Initial estimates of all missing intermolecular force-field constants K_b, K_θ and K_χ, and the equilibrium constants for the dihedral terms, namely n and δ, were made based on analogy with existing CHARMM parameters. For the carbon–carbon triple bond in erlotinib (between CC_3 atoms) we first performed an *ab initio* rigid potential energy surface scan along the bond length, and then used a parabolic potential function to fit the potential surface for an initial estimate of this bond constant.
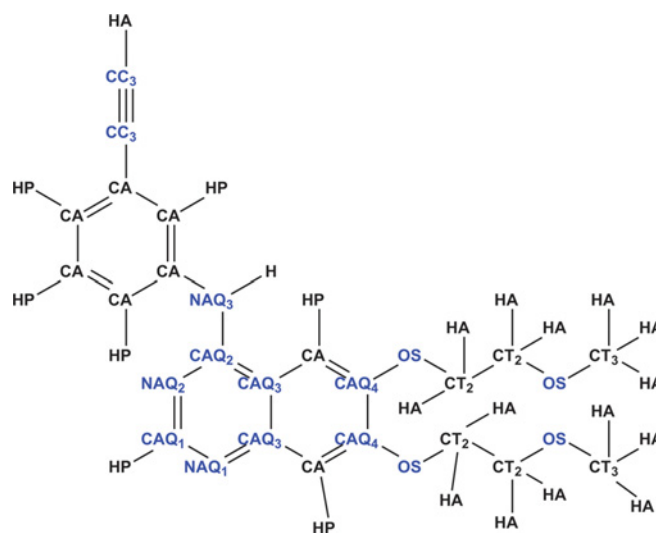

Figure S1   Atom types defined for erlotinib

The nine new atom types that required parameterization are shown in blue text (OS, CC_3, CAQ_1, CAQ_2, CAQ_3, CAQ_4, NAQ_1, NAQ_2 and NAQ_3).

Partial atomic charges (q_i) were adjusted to reproduce inter-action energies and geometry of hydrogen-bond donor/acceptor atoms in the inhibitor compound with water molecules: water molecules were placed proximal to the three nitrogen atoms of erlotinib. Hydrogen bonds between erlotinib and these three waters were individually optimized via *ab initio* calculations using the 6-31G* basis set with fixed monomer geometries. The partial atomic charges in CHARMM were manually adjusted to reproduce the *ab initio* geometric and energetic results. Following the strategy used in generating CHARMM force fields for other biomolecules [1,5,6], the *ab initio* interaction energies are scaled by 1.16 and the distances are offset by -0.2Å.

The force constants K_b, K_θ and K_x were refined by reproducing the vibrational eigenvalues and eigenvectors from *ab initio* calculations following the procedure used by Vaiana et al. [7], which ensured that both vibrational frequencies and vibrational modes (defined by eigenvectors) calculated from CHARMM and those from *ab initio* methods match closely. In this algorithm, the current parameter set is used for energy minimization and the calculation of normal modes v_i^C and $\overline{\chi}_i^c$ (eigenvalues and eigenvectors) with CHARMM. Each of the modes is projected to the eigenvector sets $\overline{\chi}_i^G$ (the corresponding quantity calculated

---

[1]   These authors contributed equally to this work.

[2]   Correspondence may be addressed to either of these authors (email mlemmon@mail.med.upenn.edu or rradhak@seas.upenn.edu).

Co-ordinates and structure factors of the erlotinib-bound EGFR[672–998]/V924R structure have been deposited with the PDB under the accession code 4HJO.

Table S1 Water–erlotinib interactions and erlotinib dipole moment calculated *ab initio* (GAUSSIAN) and using CHARMM with erlotinib parameters

The *ab initio* interaction energies have been scaled by a factor of 1.16 (see the Experimental section) and the distances are scaled by − 0.2 Å. The dipole moment (measured in Debyes) calculated by Gaussian was 4.87 and by CHARMM was 5.07.

Hydrogen bond	Interaction energies (kcal/mol)		Distances (Å)	
	GAUSSIAN	CHARMM	GAUSSIAN	CHARMM
N1...HOH	− 6.69	− 6.61	1.93	1.91
N3...HOH_2	− 5.33	− 5.30	2.12	2.01
NH...OHH_2	− 6.52	− 6.52	2.44	2.63

from GAUSSIAN) and a best projection (mode i in CHARMM projecting to mode j_{max} in GAUSSIAN), is obtained according to two criteria: (i) there is a one-to-one correspondence of the two-sets of eigenvectors, and (ii) $\prod_i \frac{1}{\max_j (\overline{x}_i^c \cdot \overline{x}_j^G)}$ is a minimum.

Then, the penalty function value is calculated by

$$\sigma = \sqrt{\frac{\sum\limits_{i=1}^{3N-6} \omega_i (v_i^C - v_{j\max}^G)^2}{3N-6}}, \; \omega_i = \frac{1}{\max_j(\overline{x}_i^C \cdot \overline{x}_j^G)}.$$

In the ideal case, $v_i^C = v_{j\max}^G \overline{x}_i^c \cdot \overline{x}_j^G = \delta_{ij}$, which implies that the CHARMM parameter set can perfectly reproduce the frequency spectrum from GAUSSIAN.

For the dihedral potential surface fitting, a relaxed potential surface scan for key dihedral angle (C4-C4-N1-C6) is performed by both CHARMM and GAUSSIAN and defined as $\overline{D}^C$, $\overline{D}^G$ and the penalty function is defined as

$$\sigma = \sqrt{\sum\limits_{i=1}^{NGRID} (D_i^C - D_i^G)^2}$$

An automated procedure using a GA (genetic algorithm) [8] was developed for the refinement of the intermolecular force-field constants and the dihedral potential energy surfaces. We iteratively repeated the procedure to refine (i) partial charges, (ii) frequencies and eigenvectors and (iii) the dihedral surface scan until reaching a force field with which the target data, the water interaction, the dihedral potential energy surface and the vibrational normal modes, calculated by CHARMM match well with the corresponding values calculated by GAUSSIAN. A demonstration of the successful application of this procedure for parameterizing erlotinib is shown in Figure S2 and Table S1. This parameter set is optimized to be consistent with the CHARMM27 force field, and is ready to be used further in MD simulations involving the inhibitor.

FEP

ΔG_B and ΔG_U (see Figure S3) for the effects of the L834R mutation were calculated by the FEP method [9], using the alchemical free energy method in NAMD 2.7b2 [10] with the dual-topology paradigm. For each state (bound or unbound), FEP calculations were performed in both forward and backward directions to ensure convergence and to obtain error bars. For each direction, the perturbation was divided into 72 windows
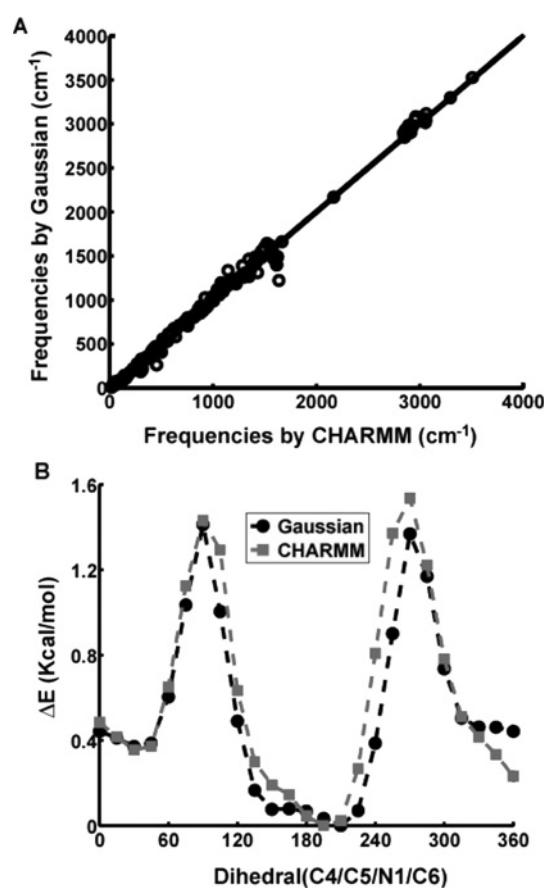

Figure S2 Target data matching between electronic structure (GAUSSIAN) and molecular mechanics calculations (CHARMM)

(**A**) One-to-one correspondence of frequencies. The frequencies are matched based on eigenvectors. (**B**) The Dihedral surface energy scan.

[$\lambda = 0$, 10^{-6}, 10^{-5}, 10^{-4}, 10^{-3}, 10^{-2}, 0.005, 0.01, 0.015, 0.02, 0.03-0.1 (with an interval of 0.01), 0.1–0.9 (with an interval of 0.02), 0.9–0.98 (with an interval of 0.01), 0.098, 0.0985, 0.099, 0.995, 0.999, 0.9999, 0.99999, 0.999999, 1]. In each window, the system was equilibrated for 20 ps and run for another 100 ps for data collection. Larger window sizes and longer simulations were also tested to ensure that this set-up provides reasonable convergence in the final binding affinity. To avoid 'end-point catastrophes' [11], the soft-core potential was used to gradually scale the unbonded interaction potential. For appearing particles, van der Waals interactions were linearly coupled to the simulation from $\lambda = 0$ (fully decoupled) to $\lambda = 1$ (fully coupled), and electrostatic interactions were coupled to the simulation over the range $\lambda = 0.5$ to $\lambda = 1$. For the vanishing particles, the van der Waals interactions were linearly decoupled from the simulation over the value range 0 to 1, and the electrostatic interactions were decreased gradually from $\lambda = 0$ to $\lambda = 0.5$.

RESULTS AND DISCUSSION

Erlotinib binding to active EGFR-TKD

To investigate the basis for possible preferential binding of erlotinib to the active conformation of EGFR-TKD, we used computational approaches. We first docked erlotinib on to two different wild-type active conformation EGFR-TKD crystal structures using the docking algorithm Glide [12] (see the
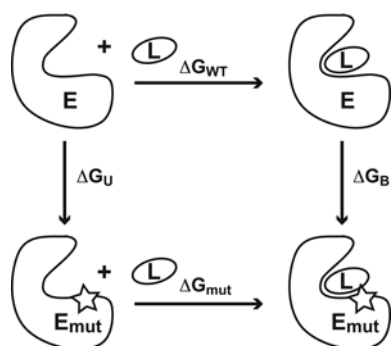
Figure S3   Thermodynamic cycle for calculating change in erlotinib binding energy caused by EGFR-TKD mutation

The change in ΔG for drug binding caused by an EGFR-TKD mutation ($\Delta\Delta G_{WT\rightarrow mut}$) can be calculated based on this thermodynamic cycle as the difference between the free energy changes caused by the particular mutation in the bound state (ΔG_B) and the unbound state (ΔG_U).

Experimental section in the main text). One structure was an EGFR-TKD–erlotinib complex (PDB entry 1M17 [4]) and the other was an active conformation of EGFR-TKD with bound p[NH]ppA (adenosine 5′-[β,γ-imido]triphosphate) (PDB entry 2ITX) [13]. In both cases, erlotinib was predicted to bind in a very similar orientation to that seen crystallographically [4], as shown in Figure S4, with N1 of the erlotinib quinazoline moiety accepting a predicted hydrogen bond from the amide nitrogen of Met[769]. The docked structures were subjected to MD simulations (see the Experimental section in the main text), which showed that erlotinib is stable in the location shown in Figure S4 throughout the entire course of a 10 ns run (Figure S5). The aniline moiety remains in a pocket defined by Leu[764] (not shown), Lys[721] and Thr[766], whereas the 'tails' of the erlotinib molecule are quite flexible. In our MD simulations, two water molecules (cyan W1 and W2 in Figure S4) entered the binding pocket and formed a stable network of hydrogen bonds involving residues Thr[766], Gln[767] and Thr[830] after the initial 2 ns of equilibration. W1 is also seen in the published crystal structure (magenta in Figure S4), and forms a predicted hydrogen bond with both the N3 nitrogen of the erlotinib quinazoline moiety and the side-chain hydroxy group of the gatekeeper residue Thr[766]. The second water molecule seen in our model (W2), but not observed in the 2.6 Å resolution crystal structure [4], appears to bridge W1 to the carbonyl oxygen of Gln[767]. Very similar observations were made when erlotinib was docked similarly on to the structure of EGFR-TKD harbouring a L834R activating mutation (PDB entry 2ITZ [13]), with the modelled inhibitor overlaying almost exactly with that seen in Figure S4, and with the same water involvement in binding seen in a subsequent MD run (Figure S5, green trace). A very similar hydrogen-bonding network involving two water molecules was reported in a previous computational study using a different force field [14], lending further confidence to these results.

Comparison of erlotinib binding to wild-type and L834R EGFR-TKD

To compare the erlotinib-binding affinities of wild-type and L834R-mutated EGFR-TKD, we calculated the absolute binding
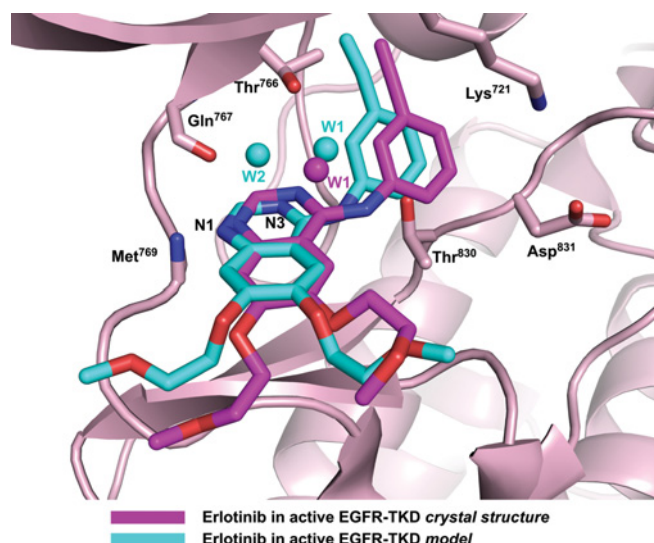

Erlotinib in active EGFR-TKD *crystal structure*
Erlotinib in active EGFR-TKD *model*

Figure S4   Erlotinib binding to active EGFR-TKD in the crystal structure and model

Erlotinib is shown bound to active EGFR-TKD in our computational model (cyan) and in the erlotinib/EGFR-TKD (active conformation) crystal structure reported in PDB entry 1M17 [4] (magenta). Functional groups in several EGFR-TKD residues that interact directly or indirectly with the bound erlotinib are shown, and the cartoon from only 1M17 is shown. The backbone amide of Met[769] donates a hydrogen bond to N1 of the erlotinib quinazoline moiety. The backbone carbonyl of Gln[767] and side chains of Thr[766] and Thr[830] participate in a hydrogen-bonding network to which water molecules (W1 and W2 in our model; W1 in 1M17) also contribute. The Lys[721] and Asp[831] side chains are shown for reference. Polypeptide in the foreground has been removed for clarity.

Table S2   Cumulative free energy in forward and backward directions for the bound and unbound states in FEP calculations

Direction	ΔG_B	ΔG_U	$\Delta\Delta G$
Wild-type→L834R	41.714	41.033	0.681
L834R→wild-type	−40.556	−39.575	−0.981

energy using both the Glide docking package [12] and MM/PBSA calculations (see the Experimental section), as listed in Table 2 of the main text. The Glide docking scores and MM/PBSA energies are very similar for both wild-type EGFR-TKD and the variant containing the L834R activating mutation. To further interrogate possible differences, we also calculated the relative binding affinity difference based on the thermodynamic cycle in Figure S3 and FEP calculations. In the FEP calculations, the Helmholtz free energy difference between wild-type EGFR-TKD and the L834R system for the bound and unbound states (ΔG_B and ΔG_U respectively) were calculated in both forward and backward directions to check for convergence. The cumulative free energy differences are shown in Figure S6 (see also Table S2). The calculated $\Delta\Delta G_{WT\rightarrow mut}$ values are 0.68 kcal/mol and 0.98 kcal/mol respectively in the forward and backward directions, again indicating that erlotinib binds with a very similar affinity to wild-type and L834R-mutated EGFR-TKD.
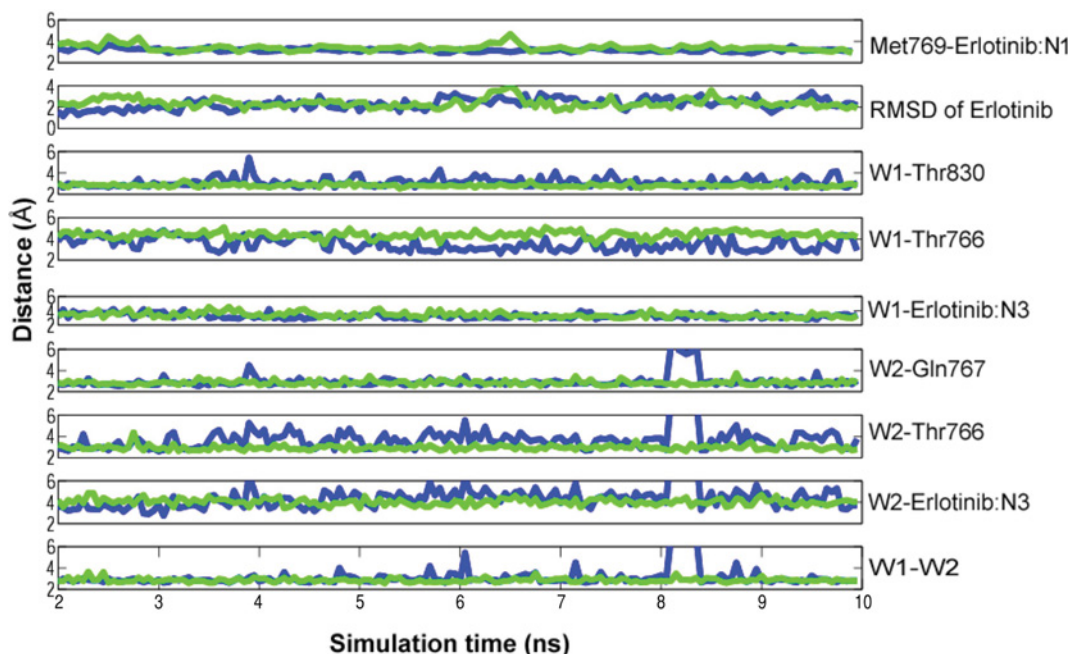
Figure S5   RMSD of erlotinib and key distances between EGFR-TKD and erlotinib, as well as the distance of waters from EGFR-TKD or erlotinib monitored during a 10 ns MD simulation

Data for the wild-type active-conformation simulation are blue, and data for the L834R (active conformation) simulation are green. Data from the first 2 ns (pre-equilibration) are not shown.
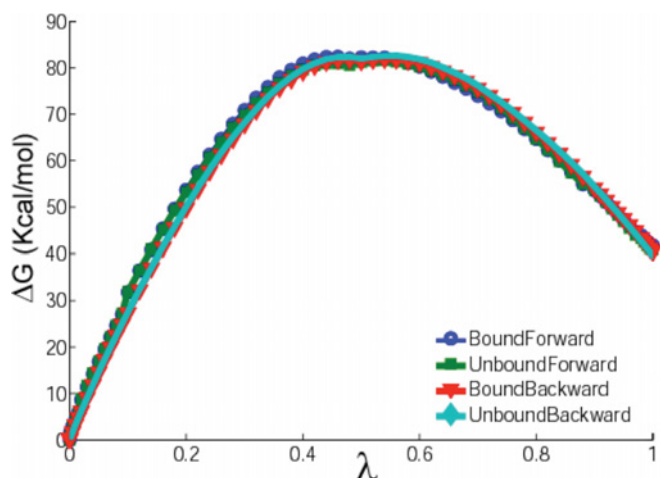

Figure S6   Cumulative free energy (kcal/mol) calculated in FEP studies for bound and unbound systems in both forward and backward directions

The backward direction energies were scaled to the same zero point as the forward direction. For the FEP calculations, the van der Waals and electrostatic interactions are separately scaled. For appearing particles, van der Waals interactions are linearly coupled to the simulation from $\lambda = 0$ (fully decoupled) to $\lambda = 1$ (fully coupled), and electrostatic interactions are coupled into the simulation over the range $\lambda = 0.5$ to $\lambda = 1$. For vanishing particles, van der Waals interactions are linearly decoupled from the simulation over the value range of 0–1, and the electrostatic interactions are decreased gradually from $\lambda = 0$ to $\lambda = 0.5$.
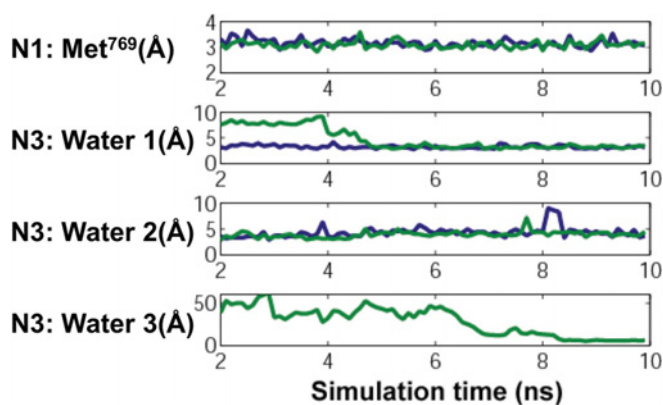

Figure S7   Comparison of RMSD values for critical hydrogen bonds in active and inactive wild-type simulations

Values for the distance between N1 of the erlotinib quinazoline moiety and Met[769], as well as the distance from erlotinib N3 of the three waters shown in the model in Figure 3(A) of the main text were monitored during the 10 ns simulation (omitting the first 2 ns), and are plotted in green. For comparison, data for the wild-type active EGFR-TKD conformation (as in Figure S5) are shown in blue.

REFERENCES

1 Foloppe, N. and MacKerell, A. D. (2000) All-atom empirical force field for nucleic acids: I. Parameter optimization based on small molecule and condensed phase macromolecular target data. J. Comput. Chem. **21**, 86–104

2 Breneman, C. M. and Wiberg, K. B. (1990) Determining atom-centered monopoles from molecular electrostatic potentials: the need for high sampling density in formamide conformational analysis. J. Comput. Chem. **11**, 361–373

3 Frisch, M. J., Trucks, G. W., Schlegel, H. B., Scuseria, G. E., Robb, M. A., Cheeseman, J. R., Scalmani, G., Barone, V., Mennucci, B., Petersson, G. A. et al. (2009), Gaussian 09, Gaussian, Inc, Wallingford CT

4 Stamos, J., Sliwkowski, M. X. and Eigenbrot, C. (2002) Structure of the epidermal growth factor receptor kinase domain alone and in complex with a 4-anilinoquinazoline inhibitor. J. Biol. Chem. **277**, 46265–46272

5 MacKerell, A. D., Bashford, D., Bellott, M., Dunbrack, R. L., Evanseck, J. D., Field, M. J., Fischer, S., Gao, J., Guo, H., Ha, S. et al. (1988) All-atom empirical potential for molecular modeling and dynamics studies of proteins. J. Phys. Chem. B. **102**, 3586–3616

6 Pavelites, J. J., Gao, J. L., Bash, P. A. and Mackerell, A. D. (1997) A molecular mechanics force field for NAD$^+$, NADH, and the pyrophosphate groups of nucleotides. J. Comput. Chem. **18**, 221–239

7 Vaiana, A. C., Cournia, Z., Costescu, I. B. and Smith, J. C. (2005) AFMM: a molecular mechanics force field vibrational parametrization program. Comput. Phys. Commun. **167**, 34–42

8 Morris, G. M., Goodsell, D. S., Halliday, R. S., Huey, R., Hart, W. E., Belew, R. K. and Olson, A. J. (1988) Automated docking using a Lamarckian genetic algorithm and an empirical binding free energy function. J. Comput. Chem. **19**, 1639–1662

9 Radmer, R. J. and Kollman, P. A. (1997) Free energy calculation methods: a theoretical and empirical comparison of numerical errors and a new method for qualitative estimates of free energy changes. J. Comput. Chem. **18**, 902–919

10 Phillips, J. C., Braun, R., Wang, W., Gumbart, J., Tajkhorshid, E., Villa, E., Chipot, C., Skeel, R. D., Kale, L. and Klaus Schulten, K. (2005) Scalable molecular dynamics with NAMD. J. Comput. Chem. **26**, 1781–1802

11 Beutler, C., Mark, A. E., van Schaik, R. C., Gerber, P. R. and van Gunsteren, W. F. (1994) Avoiding singularities and numerical instabilities in free energy calculations based on molecular simulations. Chem. Phys. Lett. **222**, 529–539

12 Friesner, R. A., Murphy, R. B., Repasky, M. P., Frye, L. L., Greenwood, J. R., Halgren, T. A., Sanschagrin, P. C. and Mainz, D. T. (2006) Extra Precision Glide: docking and scoring incorporating a model of hydrophobic enclosure for protein-ligand complexes. J. Med. Chem. **49**, 6177–6196

13 Yun, C. H., Boggon, T. J., Li, Y., Woo, M. S., Greulich, H., Meyerson, M. and Eck, M. J. (2007) Structures of lung cancer-derived EGFR mutants and inhibitor complexes: mechanism of activation and insights into differential inhibitor sensitivity. Cancer Cell **11**, 217–227

14 Balius, T. E. and Rizzo, R. C. (2009) Quantitative prediction of fold resistance for inhibitors of EGFR. Biochemistry **48**, 8435–8448