BJ Structure

Biochem. J. (2012) **448**, 417–423 (Printed in Great Britain) doi:10.1042/BJ20121513

online data

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:

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

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.

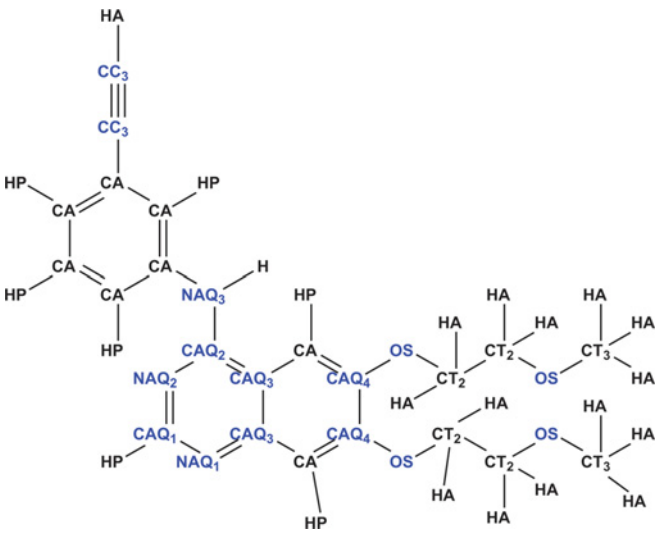

Figure S1 Atom types defined for erlotinib

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 interaction 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.

	Interaction energies (kcal/mol)		Distances (Å)	
Hydrogen bond	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

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

In the ideal case, $v_i^C = v_{j\max}^G \overline{\chi}_i^c \cdot \overline{\chi}_j^G = \delta_{ij}$, 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

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

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

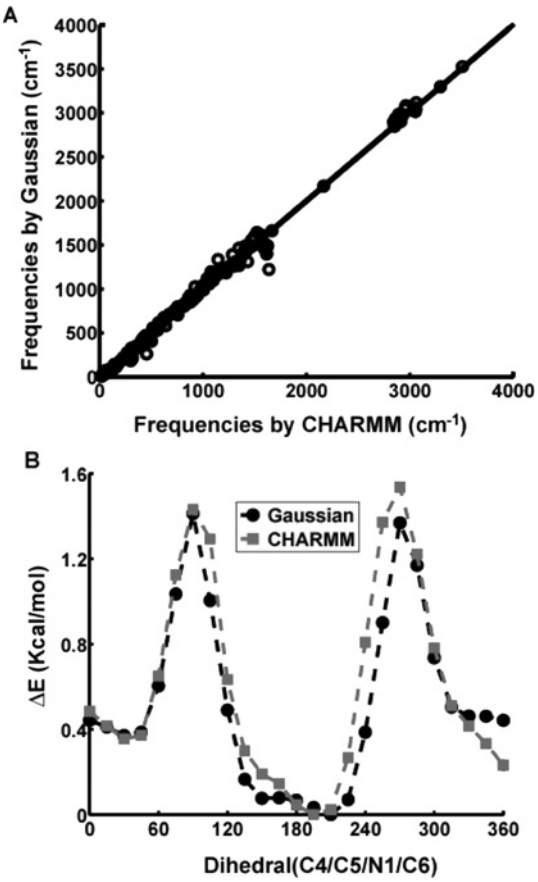

Figure S2 Target data matching between electronic structure (GAUSSIAN) and molecular mechanics calculations (CHARMM)

(**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

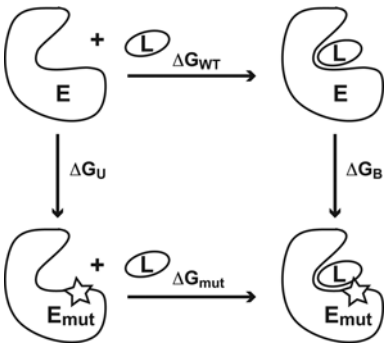

Figure S3 Thermodynamic cycle for calculating change in erlotinib binding energy caused by EGFR-TKD mutation

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

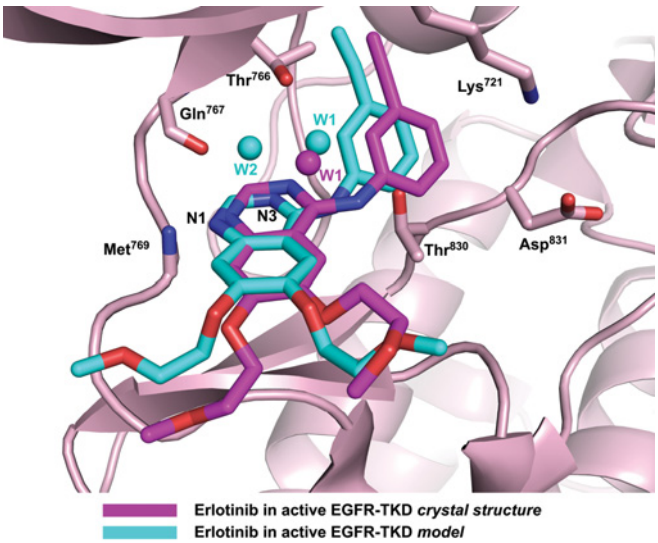

Figure S4 Erlotinib binding to active EGFR-TKD in the crystal structure and model

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.

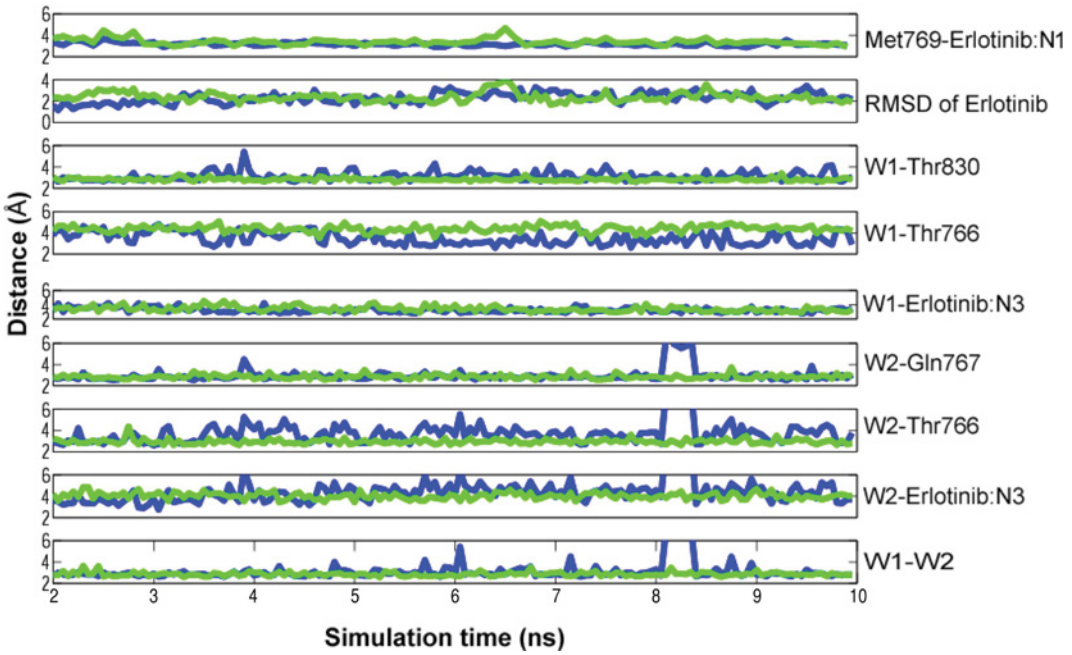

Figure S5 RMSD of erlotinib and key distances between EGFR-TKD and erlotinib, as well as the distance of waters from EGFR-TKD or erlotinib monitored during a 10 ns MD simulation

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.

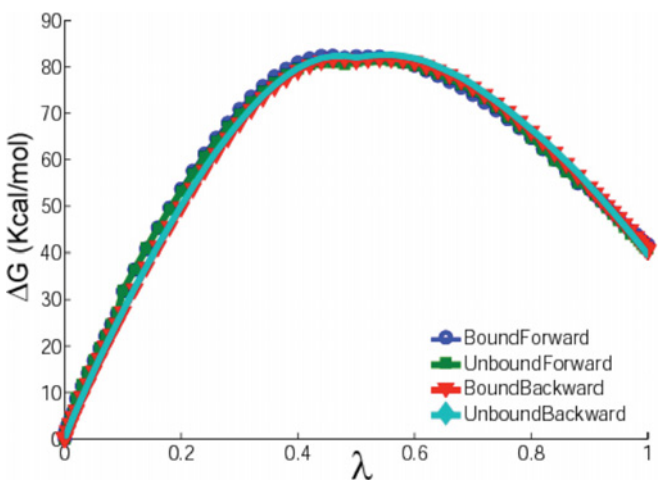

Figure S6 Cumulative free energy (kcal/mol) calculated in FEP studies for bound and unbound systems in both forward and backward directions

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 $\lambda = 0.5$.

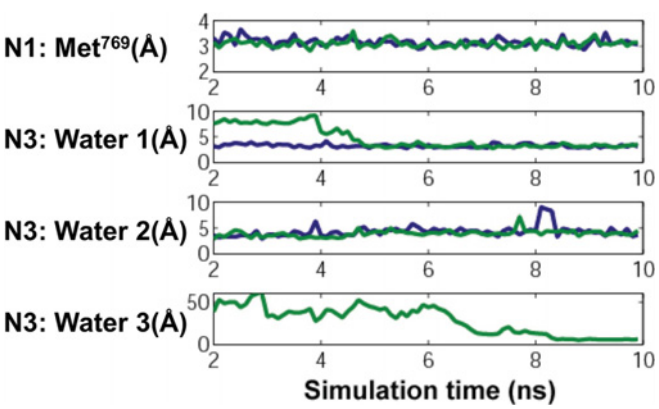

Figure S7 Comparison of RMSD values for critical hydrogen bonds in active and inactive wild-type simulations

Values for the distance between N1 of the erlotinib quinazoline moiety and Met769, 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

Received 1 October 2012/19 October 2012; accepted 29 October 2012
Published as BJ Immediate Publication 29 October 2012, doi:10.1042/BJ20121513