Effects of Single Nucleotide Polymorphisms on the Binding of Afatinib to EGFR: A Potential Patient Stratification Factor Revealed by Modelling Studies

Srinivasaraghavan Kannan[1*], Daniel Shao-Weng Tan[2,3,4] Chandra Verma[1,5,6*]

[1]Bioinformatics Institute (A*STAR), 30 Biopolis Street, 07-01 matrix, Singapore 138671

[2]Division of Medical Oncology, National Cancer Centre Singapore, 11 Hospital Drive, Singapore 169610

[3]Cancer Therapeutics Research Laboratory, National Cancer Centre Singapore, 11 Hospital Drive, Singapore 169610

[4]Cancer Stem Cell Biology, Genome Institute of Singapore, 60 Biopolis Street, 02-01, Singapore 138672

[5]School of Biological Sciences, Nanyang Technological University, 60 Nanyang Drive, Singapore 637551, Singapore

[6]Department of Biological Sciences, National University of Singapore, 14 Science Drive 4, Singapore 117543, Singapore

***Corresponding authors:** Srinivasaraghavan Kannan, Bioinformatics Institute (A*STAR), 30, Biopolis Street, 07-01 Matrix, Singapore 138671, Singapore, E-mail: raghavk@bii.a-star.edu.sg; Tel: +65 6478 8353; Fax: +65 6478 9048. Chandra S. Verma, Bioinformatics Institute (A*STAR), 30, Biopolis Street, 07-01 Matrix, Singapore 138671, Singapore, E-mail: chandra@bii.a-star.edu.sg; Tel: +65 6478 8273; Fax: +65 6478 9048.

Abstract

The anticancer drug afatinib has been found to be more effective at inhibiting the oncogenic EGFR mutant exon 19 deletion (19del) over the oncogenic EGFR mutant L858R. The underlying mechanism has been hypothesized to result from differences in structural constraints introduced by the mutations and stabilizing interactions afforded by a buried water molecule in 19del (*Kannan S. et al. Scientific Reports 2017, 7: 1540*). The COSMIC cancer database is mined for EGFR sequences to discover that several mutations in the form of Single Nucleotide Polymorphisms (SNPs) line this hydration cavity. In this work, the effects of these SNPs on the affinity of afatinib for $EGFR^{WT}$ and oncogenic mutants $EGFR^{L858R}$ and $EGFR^{19del}$ were studied using Free Energy Perturbation and Thermodynamic Integration calculations. The simulations reveal that several SNPs have significant effects on the affinity of afatinib for the mutant EGFRs carrying the SNPs and may thus have clinical implications relating to emergence of resistance to afatinib, thus potentially impacting the choice of EGFR inhibitors in the clinic.

Introduction

Mutations arise in proteins that are targeted by drugs as an adaptive response to evolutionary pressures; this is likely true at least for drugs that compete with the binding of the natural substrates to protein targets. For example, the break point cluster-Abelson tyrosine kinase (BCR-ABL) is a target for chronic leukemia therapy and is targeted by the wonder drug Gleevec (STI-571); the T315I "gatekeeper" mutation in the ABL domain of the target leads to the loss of an important hydrogen bond between the target and this drug, rendering the drug ineffective[1]. Similarly, the epidermal growth factor receptor (EGFR) is implicated in non-small cell lung carcinoma (NSCLC) and is targeted by drugs such as gefitinib. The T790M "gatekeeper" mutation in the EGFR is resistant to gefitinib because the competitiveness of the inhibitor relative to ATP is significantly weakened[2]. Another compelling case is that of the ALK (anaplastic lymphoma kinase) protein targeted by the small molecule crizotinib used for treating some forms of NSCLC; it is rendered ineffective by mutations such as L1196M, C1156Y, and G1202R[3].

The kinase domain of EGFR, made up of two lobes, is activated when a series of structural changes take place: a flexible helix (αC-helix), which is part of the dimer interface, undergoes a conformational change (to what is referred to as αC-In conformation) resulting in the formation of a salt bridge between a conserved glutamate (E762) from the αC-helix and a conserved Lysine (K745) from strand β3, which also coordinates ATP through its phosphate. A long flexible loop (termed the activation loop or A-loop), which, together with the P-loop, mediates interactions with the ATP and the substrate binding lobes, undergoes conformational changes to enable substrate binding and phosphorylation (Figure 1). Hence, it is no surprise that some of the frequently occurring and key oncogenic mutations in EGFR such as the common drivers of NSCLC, L858R and exon 19 deletion (19del) induce these conformational changes, resulting in abnormal activation and drug sensitivity.

While several studies have examined how resistance mutations in proteins alter the binding affinities of drugs, only a few studies have evaluated the effects of single-nucleotide polymorphisms (SNPs) on drug binding. Sun et al have systematically investigated the effects of 879 SNPs in 271 human drug targets[4]. Using simulations, they show that only half the SNPs are located within 12 Å of the drug binding sites and affected the binding of a few drugs[4]. Yang et al constructed a comprehensive database called "Variations and Drugs

(VnD)" which provides information on disease-related genes, proteins, drugs, structural changes caused by non-synonymous SNPs (nsSNPs)[5]. More recently they complemented this with a database of germline and somatic non-synonymous single-nucleotide variations (nsSNVs) that are associated with different cancer types and are part of the drug binding sites in the associated proteins[6]. By mapping the identified mutations onto the protein-drug complex structures, they identified 3133 nsSNVs that could affect the binding of the drug molecules[7]. We extend these studies by carrying out a detailed study quantifying the effects of SNPs in EGFR on the binding of the drug afatinib to the kinase domain of EGFR, combining long MD simulations and Thermodynamic Integration (TI) and Free energy perturbation (FEP) calculations. Free energy computations have been successful at reproducing, in close agreement with experimental data (RMS error of less than 1 kcal/mol), the free energies of binding of small-molecule ligands to proteins[8]. Their common use has been limited by the lack of computational power until recently. However, with the developments in computer power, recent studies have demonstrated the predictive power of these methods in quantifying the thermodynamics of interactions between proteins and small molecules, proteins and antibodies[8-13].

We recently developed structural and computational models to understand the higher efficacy of the irreversible EGFR inhibitor afatinib observed in a clinical trial in patients carrying the $EGFR^{19del}$ mutation compared to those carrying the $EGFR^{L858R}$ mutation[14]. Modelling and molecular simulations suggested that differential structural constraints imposed by the two mutants on afatinib could account, at least partly, for this difference. In particular, the $EGFR^{19del}$ mutant was found to engage the drug with strong hydrogen and halogen bond interactions and also had a single buried water molecule which was part of the network of these interactions (Figure 2A); in contrast, in the $EGFR^{L858R}$ mutant and the $EGFR^{WT}$ kinase, this region was occupied by few loosely ordered water molecules (Figure 2B, 2C) which did not bind the drug as strongly as in $EGFR^{19del}$. The observations resulted in the speculation that the interactions of this bound water molecule together with the rigidity of the αC-helix contribute to the preferential binding of afatinib to the $EGFR^{19del}$ mutation over $EGFR^{L858R}$ and $EGFR^{WT}$. We further speculated that known SNPs that surround the buried water molecule may result in differences in the shape of the hydration cavity and hence affect the affinity for afatinib (Figure 2A). We now report a TI and FEP based analysis of the effects of these SNPs on the binding of afatinib to $EGFR^{WT}$, $EGFR^{L858R}$ and $EGFR^{19del}$ mutants. This

we believe is the first report exploring the effects of SNPS on the free energies of the binding of a drug to a target protein.

Eleven different mutations corresponding to the reported SNPS were introduced into the structures of $EGFR^{WT}$, $EGFR^{L858R}$ and $EGFR^{19del}$ bound with afatinib, as reported earlier[14]. Different numbers of conserved water molecules identified[14] in the previous study (one, two and four for $EGFR^{19del}$, $EGFR^{L858R}$ and $EGFR^{WT}$ respectively) are included here. For each SNP, the mutant EGFR–afatinib–water molecules complex was constructed by mutating the sidechain and the system prepared for calculations as discussed in[14]. The modelled complexes were subjected to MD simulations using Amber16 with ff03 and gaff force fields for the protein and afatinib respectively. The covalent bond between afatinib and Cys797 was modelled/parameterized as mentioned in[14.] Each complex system was simulated for 250 ns in triplicates. Binding energies were estimated by two TI and FEP based calculations: one was performed on apo EGFR and the other on the EGFR-afatinib complex. The difference in binding energy upon mutation was computed using a thermodynamic cycle (more details in methods). Multistate Bennett Acceptance Ratio (MBAR) was used for the analysis of the FEP data as MBAR is shown to improve the accuracy of the predicted free energies with better sampling and convergence[15, 16]. The methodology used here was motivated by several recent studies where TI and FEP based calculations were shown to reproduce the experimentally determined mutation-dependent binding free energy differences with root mean square errors (RMSE) of ~ 1 kcal/mol[17 - 21].

Results and Discussions

We first benchmarked the ability of the simulations to reproduce the binding energetics of the buried water molecule since there is simulation and experimental data available in the literature. The double-decoupling technique with thermodynamics integration (TI) in MD simulations has been shown to successfully reproduce the free energy of solvating (hydrating) a water molecule[22] and of embedding a water molecule in the binding pockets of proteins[22, 23]. Our calculations yielded the free energy of a water molecule in bulk solvent to be +6.5 kcal/mol which is in good agreement with the experimental value of +6.3 kcal/mol[24]. We next computed the free energy of decoupling the buried water in the $EGFR^{19del}$-afatinib complex to be +10.3 kcal/mol. This suggests that moving a water molecule from bulk solvent into the binding site in the $EGFR^{19del}$-afatinib complex is favoured by ~3.8 kcal/mol

(following the guidelines of ref 22, 23, the water molecule was constrained in the binding site of the *EGFR*^{19del}-afatnib complex with a harmonic force constant of 1.1 kcal/mol). The FEP/MBAR calculations also suggested that the water molecule prefers to be solvated by the protein binding site by ~4.3 kcal/mol. We next enquire whether any of the SNPs affect (a) the free energy of insertion of this water into the binding cavity and (b) the binding free energy of afatinib to *EGFR*^{19del}. Of the 11 SNPs studied here, 3 reside in the A-loop (yellow carbon atoms in Figure 2), 4 in the αC-helix (green and blue carbon atoms in Figure 2) and the remaining in strands β2 (orange carbon atoms in Figure 2) and β3 (grey carbon atoms in Figure 2) located in the N-lobe of the kinase. Except for the DFG motif (where D is replaced by G and G is replaced by E), all the other SNPs involve substitutions of small hydrophobic sidechains with bulkier or branched hydrophobic sidechains; none of these are involved in direct contacts with afatinib.

We first look at the effects of the SNPs in the background of *EGFR*^{19del}. Ile759 contributes to packing in the hydrophobic pocket; its mutation to the shorter sidechain of Val results in decrease of the stability of W1 and a slight increase (~1 kcal/mol) in the affinity for afatinib (Figure 3). The long sidechain of Met (M766) shields the buried water molecule (we will refer to this as W1). Substitution of M766 with bulkier Phe results in occlusion of the binding site of water and its displacement (reduction of residence time of water to ~10 ns); nevertheless this results in increased packing interactions of the 3chloro-4flurophenyl group of afatinib, and an associated increase in the affinity of afatinib by ~2 kcal/mol. Substitution of M766 with the smaller Val or Thr results in the loss of packing, thus the water molecule at W1 becomes less ordered (residence time decreases from 90 ns in *EGFR*^{19del} to 70 ns and 45 ns in M766V and M766T respectively; Figure 3) and this is associated with a reduction in the affinity of afatinib by 3-4 kcal/mol. (Figure 3). The side chain of Leu777 shields the water; its mutation to Gln does not affect the residence time of W1 or of the affinity for afatinib as the linear and flexible polar side chain of Gln is easily accommodated and makes additional interactions with W1. Substitution of Leu777 with bulkier Phe results in destabilization of the packing, especially of spatially contiguous Leu788, which allows the water to exit, thus reducing its residence time to ~60% although the resulting hole is not compensated for by any altered packing in the pocket; these structural perturbations result in reducing the affinity for afatinib by ~3.3 kcal/mol (Figure 3). Mutations of Leu778 (which shields W1) to smaller Val or larger Phe result in altered structural packing (substitution to smaller Val results in larger flexibility of W1; substitution to larger Phe results in occlusion of the water site) and

enabling the water to exit, resulting in reductions in its residence time to ~10 - 20 ns; again there are no structural rearrangements to compensate for the loss of water; the resulting affinity for afatinib reduces by ~1.4-1.8 kcal/mol (Figure 3). The DFG motif residues Asp855, Phe856 and Gly857 are not in direct contact with W1, but are in close proximity. Substitutions Asp855Gly and Gly857Glu result in structural perturbations, the residence times of W1 reduce to 75-85 ns. The Phe856Tyr mutation results in a ne*w* hydrogen bond between the side chain hydroxyl of Tyr856 and W1 and the residence time of W1 is ~85 ns (in the $EGFR^{19del}$, the residence time is 92 ns). The associated affinity for afatinib for this SNP increases by up to ~4 kcal/mol while for the other two is less than ~1 kcal/mol (Figure 3). In general while the destabilization and loss of the buried water W1 can be deleterious for afatinib interactions, the location of the mutation is critical (as in M766F) as it can result in the sidechain compensating for the loss of the water and increase the affinity for afatinib. In summary, the free energy calculations suggest that the binding of afatinib can be affected in $EGFR^{19del}$ patients carrying different SNPs and can result in either destabilizing the affinities for afatinib by up to 4.5 kcal/mol or stabilizing by up to 2.7 kcal/mol (Figure 3B, C). Calculations of the energetics of the buried water molecule (in the case of the 7 SNPs where the water remains stably bound) suggest that it contributes favourably, with binding free energies (removal of water from bulk into the binding pocket) of ~3.0 kcal/mol (except for D855G and G857E; Figure 3D). The contribution is >6.0 kcal/mol in the case of D855G but only within 1 kcal/mol in the case of G857E. In both these cases, the conserved water molecule is not in direct contact with the residue and therefore the contribution is likely the result of non-specific interactions rather than hydration at site W1.

We next investigated the effect of these mutations in the background of the activating mutation $EGFR^{L858R}$ which accounts for 40% of NSCLC. Simulations of the $EGFR^{L858R}$ – afatinib complex were carried out with two buried water molecules that were identified earlier (Figure 2B). Most SNPs result in a decrease in hydration either resulting from substitutions with larger sidechains which occlude the hydration sites or from substitution with shorter side chains which result in loss of packing and an associated increase in flexibility/exit of water molecule W1 and W2 (W1 and W2 corresponds to the two hydration sites identified earlier[14], buried water at these sites are shown as cyan spheres in Figure 2). In some SNPs (M766T and L777Q), increased interactions result in retention of the water molecules. Again, the SNPs at the DFG motif (D855G, F856Y and G857E) are not involved in direct contacts with the water molecules or with the bound drug, but showed increased

residence times for the water molecules as a result of increased structural stability around the hydration site. While in general the effects on hydration associated with the SNPs do not seem to be correlated with the changes in affinity for afatinib (Figure 4 A, C), F856Y is associated with a significantly increased affinity (~10 kcal/mol), similar to this same SNP in $EGFR^{19del}$ (Figure 4 C).

In the case of $EGFR^{WT}$ the αC-helix and the hydrophobic pocket are characterized by increased flexibility, with the result that the water molecules in the pocket are very dynamic (residence times of only ~10 ns), except in the case of M766T, which shows increased residence times (~30 ns) of water molecules (Figure 4 B, D). Similar increase in hydration of M766T was recently reported from a combined simulation/experimental study of the $EGFR^{M766T}$ complexes with the small molecules erlotinib, gefitinib and lapatinib[25]; this activating mutation is known to be resistant to both erlotinib and gefitinib (both bind to the active form of the kinase), but is inhibited by lapatinib (stabilizes the inactive form of the kinase). This difference was attributed to the different water mediated interactions involving Thr766 and the bound drugs. Similarly, the M766T SNP, despite being hydrated in our study, appears to destabilize the binding of afatinib and we speculate that either the patients carrying this SNP will be resistant to afatinib or, that such a mutation may arise in response to afatinib therapy. Once again, there appears to be no correlation between the residence times of the water molecules and the changes in affinity of afatinib.

Conclusions

In summary, the affinity for afatinib appears to be modulated by hydration at W1 in $EGFR^{19del}$, while it is agnostic to hydration in $EGFR^{L858R}$ or $EGFR^{WT}$. In $EGFR^{19del}$, SNPs which displace the water W1 result in a loss in affinity for afatinib. It is also likely that this is linked to the increase in flexibility of αC helix in the order $EGFR^{19del} < EGFR^{L858R} < EGFR^{WT}$; most SNPs are located in or near αC. The agreement of our free energy calculations with experimental data on the energetics of water provide confidence in extending these calculations to the effects of SNPs on the affinity for afatinib. It is clear that the presence of SNPs can result in significant effects on the affinity for afatinib. This may in turn have clinical consequences in terms of dosing and/or the emergence of resistant mutations in response to afatinib administration and if validated, such an approach could provide additional means to stratify patients to specific EGFR inhibitors.

Materials and Methods

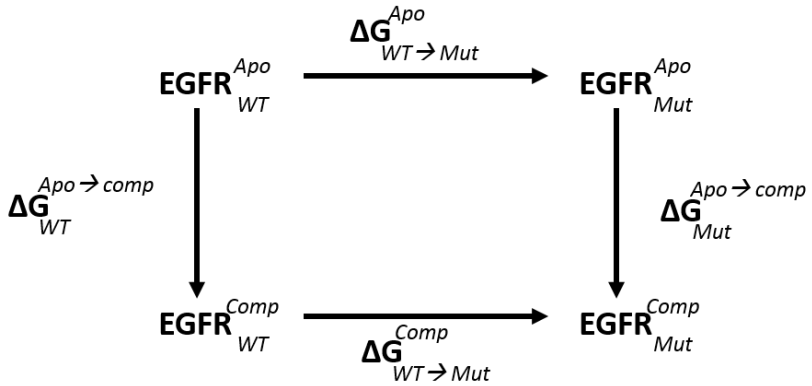
Molecular Dynamics Simulations

We used molecular dynamics (MD) simulations to investigate the structural dynamics of the mutant specific covalent binding of the anticancer drug afatinib to EGFR. The structures of the kinase domain (KD) of $EGFR^{WT}$, mutant EGFR and the corresponding models with SNPs, all in complex with afatinib, were taken from our previous study[14]. Since the mutations $EGFR^{L858R}$ and $EGFR^{19del}$ are known to be activating and the drug afatinib considered here is known to bind to the active forms of the KDs, only the active forms of $EGFR^{WT}$, $EGFR^{L858R}$, *and* $EGFR^{19del}$ are considered in this study. All the modelled structures were then subjected to atomistic MD simulations with the *pmemd.cuda* module of the program Amber16[26],using the Amber ff03 force field[27] and the general Amber force field (GAFF)[28] following the protocol set out in our previous work[14]. In the current study, simulations were carried out for 250 ns in triplicates for each system.

Residence times of water molecules were computed as a summation of times over the trajectory during which corresponding particular site was hydrated. The calculation is agnostic to the identity of the water and only accounts for the hydration of the site (ie to account for water exchange). Last 100 ns of the MD simulations trajectory was used for hydration analysis (residence time). Simulation trajectories were visualized using VMD[29] and figures were generated using PyMOL[30].

Binding Free energy calculations

Relative binding free energies were calculated using the thermodynamic cycle shown below:



From this cycle, we can calculate the difference in the free energy for the binding of the ligand to the wildtype versus and mutant.

where ΔG is the binding free energy of mutating WT protein, either bound to ligand or unbound.

$$\Delta\Delta G_{bind} = \Delta G^{Apo \rightarrow Comp}_{Mut} - \Delta G^{Apo \rightarrow Comp}_{WT}$$

$$\Delta\Delta G_{bind} = \Delta G^{comp}_{WT \rightarrow Mut} - \Delta G^{Apo}_{WT \rightarrow Mut}$$

The $\Delta G_{WT \rightarrow Mut}$ energy differences in the apo and complex forms are calculated using alchemical transformation methods (thermodynamic integration or TI and free energy petrubration or FEP). The free energy differences between the WT and mutant proteins were calculated by gradually perturbing from one to the other in a series of discrete steps, represented by λ values. The λ values vary from 0 to 1 corresponding respectively to the initial A and final B states. Two different transformations needed to be simulated: WT to mutant in the apo state and WT to mutant when both are complexed. We use a total of 11 λ windows (0.0, 0.10, 0.20, 0.30, 0.40, 0.50, 0.60, 0.70, 0.80, 0.90, 1.0) for the TI and FEP/MBAR simulations at 300 K. At each λ window we carried out 500 steps of steepest descent minimization, a 50 ps density equilibration run, and a 5ns NPT production run. The conformation of the mutant residue was taken from the most populated conformation sampled during standard MD simulations (previous paragraph). A time step of 1 fs is used together with the SHAKE algorithm[31]. Ewald sums[32] with a 9 Å cutoff in the real part, isotropic pressure scaling, and a Langevin type thermostat to maintain the temperature at 300 K were also used. To improve convergence, "soft-core" potentials were applied both to the Lennard-Jones and the Coulombic potentials as implement in AMBER 16[26]. This makes it possible to perform both charge and van der Waals mutations in the same step and therefore the alchemical transformation was carried out in a single step, referred to as the single-transformation approach (STA). Free energy derivatives ($\partial V/\partial\lambda$) were collected independently for each λ from the production run. In the TI method, the free energy difference is calculated from the integral of $\partial V(\lambda)/\partial\lambda$ from 0 to 1, where V is the potential energy. In the case of FEP, free energy differences were calculated using the multi Bennett acceptance ratio (MBAR) method as implemented in Amber 16[26]. We followed the running

average procedure as used in ref 18 for the analysis of statistical uncertainty. For each λ point, a running average of $\langle \partial V/\partial \lambda \rangle_\lambda$ on every 500 ps interval was calculated.

Competing financial interests

The authors declare no competing financial interest. SK and CV are founders of SiNOPSEE Therapeutics, Singapore; however there is no conflict of interest with this work.

Acknowledgement

The authors thank National Super Computing Centre (NSCC) for computing facilities.

Reference

1. Branford, S.; Rudzki, Z.; Walsh, S.; Grigg, A.; Arthur, C.; Taylor, K.; Herrmann, R.; Lynch, K.P.; Hughes, T.P. High frequency of Point Mutations Clustered within the Adenosine Triphosphate-Binding Region of BCR/ABL in Patients with Chronic Myeloid Leukemia or Ph-positive Acute Lymphoblastic Leukemia who Develop Imatinib (STI571) Resistance. *Blood*. **2002,** 99, 3472–3475.
2. Yun, C.H.; Mengwasser, K.E.; Toms, A.V.; Woo, M.S.; Greulich, H.; Wong, K.K.; Meyerson, M.; Eck, M.J. The T790M Mutation in EGFR Kinase Causes Drug Resistance by Increasing the Affinity for ATP. *Proc. Natl. Acad. Sci. USA.* **2008,** 105, 2070–2075.
3. Awad, M.M.; Shaw, A.T. ALK Inhibitors in Non–Small Cell Lung Cancer: Crizotinib and Beyond. *Clin. Adv. Hematol. Oncol.* **2014,** 7, 429-439.
4. Sun, H.Y.; Ji, F.Q.; Fu, L.Y.; Wang, Z.Y.; Zhang, H.Y. Structural and Energetic Analyses of SNPs in Drug Targets and Implications for Drug Therapy. *J. Chem. Inf. Model.* **2013,** 12, 3343-3351.
5. Yang, J.K; Oh, S.; Ko, G.; Park, S.J.; Kim, W.Y.; Lee, B.; Lee, S. VnD: a Structure-Centric Database of Disease-Related SNPs and Drugs. *Nucleic Acids Res.* **2011,** 39, D939–944.
6. Wu, T.J.; Shamsaddini, A.; Pan, Y.; Smith, K.; Crichton, D.J.; Simonyan, V.; Mazumder, R. A Framework for Organizing Cancer-Related Variations from Existing Databases, Publications and NGS Data using a High-performance Integrated Virtual Environment (HIVE). *Database*. **2014,** 2014: bau022.

7. Yan. C.; Pattabiraman, N.; Goecks, J.; Lam, P.; Nayak, A.; Pan, Y.; Torcivia-Rodriguez, J.; Voskanian, A.; Wan, Q.; Mazumder, R. Impact of Germline and Somatic Missense Variations on Drug Binding Sites. *The Pharmacogenomics Journal.* **2017,** 17, 128–136.
8. Wang, L.; Wu, Y.; Deng, Y.; Kim, B.; Pierce, L.; Krilov, G.; Lupyan, D.; Robinson, S.; Dahlgren, M.K.; Greenwood, J.; Romero, D.L.; Masse, C.; Knight, J.L.; Steinbrecher, T.; Beuming, T.; Damm, W.; Harder, E.; Sherman, W.; Brewer, M.; Wester, R.; Murcko, M.; Frye, L.; Farid, R.; Lin, T.; Mobley, D.L.; Jorgensen, W.L.; Berne, B.J.; Friesner, R.A.; Abel, R. Accurate and Reliable Prediction of Relative Ligand Binding Potency in Prospective Drug Discovery by Way of a Modern Free-Energy Calculation Protocol and Force Field. *J. Am. Chem. Soc.* **2015,** 137, 2695-2703.
9. Clark, A.J.; Gindin, T.;, Zhang, B.; Wang, L.; Abel, R.; Murret, C.S.; Xu, F.; Bao, A.; Lu, N.J.; Zhou, T.; Kwong, P.D.; Shapiro, L.; Honig, B.; Friesner, R.A. Free Energy Perturbation Calculation of Relative Binding Free Energy between Broadly Neutralizing Antibodies and the gp120 Glycoprotein of HIV-1. *J. Mol. Biol.* **2017,** 429, 930-947.
10. Steinbrecher, T.; Abel, R.; Clark, A.; Friesner, R.; Free Energy Perturbation Calculations of the Thermodynamics of Protein Side-Chain Mutations. *J. Mol. Biol.* **2017,** 429, 923-929.
11. Steinbrecher, T.; Zhu, C.; Wang, L.; Abel, R.; Negron, C.; Pearlman, D.; Feyfant, E.; Duan, J.; Sherman, W. Predicting the Effect of Amino Acid Single-Point Mutations on Protein Stability—Large-Scale Validation of MD-Based Relative Free Energy Calculations. *J. Mol. Biol.* **2017,** 429, 948-963.
12. Lai, P.K.; Kaznessis, Y.N.; Free Energy Calculations of Microcin J25 Variants Binding to the FhuA Receptor. *J. Chem. Theory Comput.* **2017,** 13, 3413–3423.
13. Lee, T.S.; Hu, Y.; Sherborne, B.; Guo, Z.; York, D.M. Toward Fast and Accurate Binding Affinity Prediction with pmemdGTI: An Efficient Implementation of GPU-Accelerated Thermodynamic Integration. *J. Chem. Theory Comput.* **2017,** 13, 3077–3084.
14. Kannan, S.; Pradhan, M.R.; Tiwari, G.; Tan, W.C.; Chowbay, B.; Tan, E.H.; Tan, D.S.; Verma, C. Hydration Effects on the Efficacy of the Epidermal Growth Factor Receptor Kinase Inhibitor Afatinib. *Scientific Reports*. **2017,** 7, 1540.
15. König, G.; Boresch, S. Non-Boltzmann Sampling and Bennett's Acceptance Ratio Method: how to Profit from Bending the rules. *J. Comput. Chem.* **2011,** 32, 1082– 1090.
16. Mikulskis, P.; Genheden, S.; Ryde, U. A Large-scale Test of Free-Energy Simulation Estimates of Protein-Ligand Binding Affinities. *J. Chem. Inf. Model.* **2014,** 54, 2794–2806.

17. Zhu, S.; Travis, S.M.; Elcock, A.H.. Accurate Calculation of Mutational Effects on the Thermodynamics of Inhibitor Binding to p38α MAP Kinase: A Combined Computational and Experimental Study. *J. Chem. Theory Comput.* **2013,** 9, 3151–3164.
18. Park, J.; McDonald, J.J.; Petter, R.C.; Houk, K.N. Molecular Dynamics Analysis of Binding of Kinase Inhibitors to WT EGFR and the T790M Mutant. *J. Chem. Theory Comput.* **2016,** 12, 2066–2078.
19. Cole, D.J.; Janecek, M.; Stokes, J.E.; Rossmann, M.; Faver, J.C.; McKenzie, G. J.; Venkitaraman, A.R.; Hyvonen, M.; Spring, D.R.; Huggins, D.J.; Jorgensen, W. L. Computationally-Guided Optimization of Small-Molecule Inhibitors of the Aurora A kinase–TPX2 Protein–Protein Interaction. *Chem. Commun. Camb. Engl.* **2017 ,** 53, 9372-9375.
20. Mondal, J.; Tiwary, P.; Berne, B.J. How a Kinase Inhibitor Withstands Gatekeeper Residue Mutations. *J. Am. Chem. Soc.* **2016,** 138, 4608-4615.
21. Mena-Ulecia, K.; Gonzalez-Norambuena, F.; Vergara-Jaque, A.; Poblete, H.; Tiznado, W.; Caballero. J. Study of the Affinity Between the Protein Kinase PKA and Homoarginine-containing Peptides Derived from Kemptide: Free Energy Perturbation (FEP) Calculations. *J. Comput. Chem.* **2018,** 39, 986–992.
22. Hamelberg, D.; McCammon, J.A. Standard Free Energy of Releasing a Localized Water Molecule from the Binding Pockets of Proteins: Double-Decoupling Method. *J. Am. Chem. Soc.* **2004,** 126, 7683-7689.
23. Barillari, C.; Taylor, J.; Viner, R.; Essex, J. W Classification of Water Molecules in Protein Binding Sites. *J. Am. Chem. Soc.* **2007,** 129, 2577-2587.
24. Bennaim, A.; Marcus, Y. J. Solvation Thermodynamics of Nonionic Solutes. *Chem. Phys.* **1984,** 81, 2016−2027.
25. Ruan, Z.; Katiyar, S.; Kannan, N. Computational and Experimental Characterization of Patient Derived Mutations Reveal an Unusual Mode of Regulatory Spine Assembly and Drug Sensitivity in EGFR Kinase. *Biochemistry.* **2017**, 56, 22-32.
26. Case, D.A.; Betz, R.M.; Cerutti, D.S.; Cheatham, III, T.E.; Darden, T.A.; Duke, R.E.; Giese, T.J.; Gohlke, H.; Goetz, A.W.; Homeyer, N.; Izadi, S.; Janowski, P.; Kaus, J.; Kovalenko, A.; Lee, T.S.; LeGrand, S.; Li, P.; Lin, C.; Luchko, T.; Luo, R.; Madej, B.; Mermelstein, D.; Merz, K.M.; Monard, G.; Nguyen, H.; Nguyen, H.T.; Omelyan, I.; Onufriev, A.; Roe, D.R.; Roitberg, A.; Sagui, C.; Simmerling, C.L.; Botello-Smith, W.M.; Swails, J.; Walker, R.C.; Wang, J.; Wolf, R.M.; Wu, X.; Xiao, L.; Kollman, P.A. AMBER 16, University of California, San Francisco.**2016.**

27. Ponder, J.W.; Case, D.A. Case. Force Fields for Protein Simulations. *Adv. Prot. Chem.* **2003,** 66, 27-85.
28. Wang, J.; Wolf, R.M.; Caldwell, J.W.; Kollman, P.A.; Case, D.A. Development and Testing of a General Amber Force Field. *J. Comput. Chem.* **2004,** 25, 1157–1174.
29. Humphrey, W.; Dalke, A.; Schulten, K. VMD—Visual Molecular Dynamics. *J. Mol. Graph.* **1996,** 14, 33–38.
30. De Lano, W. The PyMOL Molecular Graphics System. San Carlos CA, USA: De Lano Scientific. **2002.**
31. Miyamoto, S.; Kollman, P.A. Settle: An Analytical Version of the SHAKE and RATTLE Algorithm for Rigid Water Models. *J. Comput. Chem*. **1992,** 13, 952–962.
32. Darden, T.; York, D.; Pedersen, L. Particle mesh Ewald: An N_log(N) method for Ewald Sums in Large Systems. *J. Chem. Phys*. **1993,** 98, 10089–10092.

Figure Legends

Figure 1: Overview of the structure of $EGFR^{WT}$ kinase in its active and inactive forms, with the helix αC adopting the αC-In (active, magenta) and αC-Out (inactive, cyan) states and the activation loop adopting conformations associated with the active (green) and inactive (orange) conformations. The critical residues (K745 and E762) that make the salt bridge (in magenta dashes) in the active state are shown as sticks and the salt bridge interaction is shown as red dashes.

Figure 2: Structure of afatinib (part; magenta) bound to (A) $EGFR^{19del}$ *(B)* $EGFR^{L858R}$ (C) $EGFR^{WT}$ sampled during the MD simulations. Water molecules (cyan sphere) observed to mediate interactions between the halogen atom of afatinib and the αC helix of (A) $EGFR^{19del}$ (B) $EGFR^{L858R}$ (C) $EGFR^{WT}$. Hydrogen bond (magenta) and halogen bond (black) interactions are shown as dotted lines. The critical residues that make the salt bridge (K745 and E762; K745 is shown as green sticks and is not labelled for clarity; its salt bridging partner E762 is shown as spheres to highlight the cavity but is not a SNP) are shown. All SNPs within 5 Å of the water molecule are shown as spheres.

Figure 3: (A) Residence times of water molecules observed during the MD simulations of $EGFR^{19del}$(green) (SNPs) complexed with afatinib (B) Free energies of mutating the sidechains from wild type to the SNP in $EGFR^{19del}$ complexed to afatinib calculated using TI (blue) and FEP/MBAR(red) along with the associated errors (C) the calculated free energies (using the FEP/MBAR method) are mapped onto the structure of $EGFR^{19del}$ bound to afatinib (the surface is coloured from blue to red, corresponding to ΔΔG of -4 kcal/mol to +4 kcal/mol) (D) Free energy of solvation of the buried water molecule in the various SNP containing $EGFR^{19del}$ – afatinib complexes, calculated using the double-decoupling approach (the blue and red colours correspond to energies calculated using TI and FEP/MBAR respectively) along with the associated errors ; only those SNPs were considered where the water molecule had a residence time of 60 ns as in (A).

Figure 4: (A, B) Residence times of water molecules observed during the MD simulations of (A) $EGFR^{L858R}$ (B) $EGFR^{WT}$ complexed to afatinib and the corresponding energetic analysis (C: $EGFR^{L858R}$, D: $EGFR^{WT}$) using the TI (blue) and FEP/MBAR (red) approaches along with the associated errors.

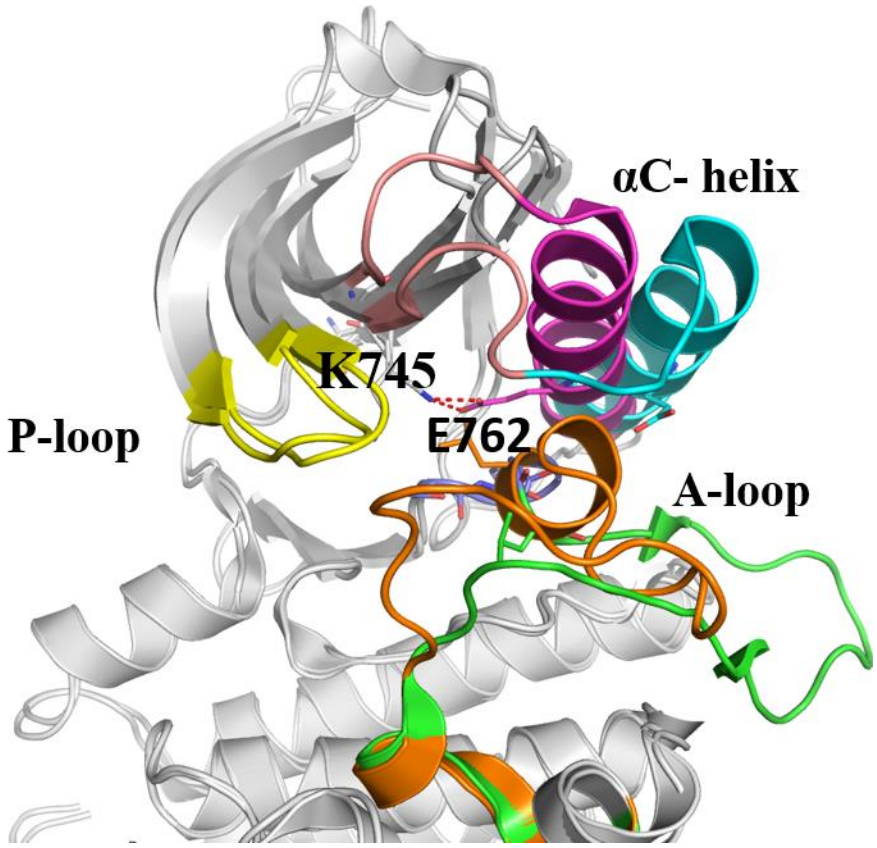


Figure 1

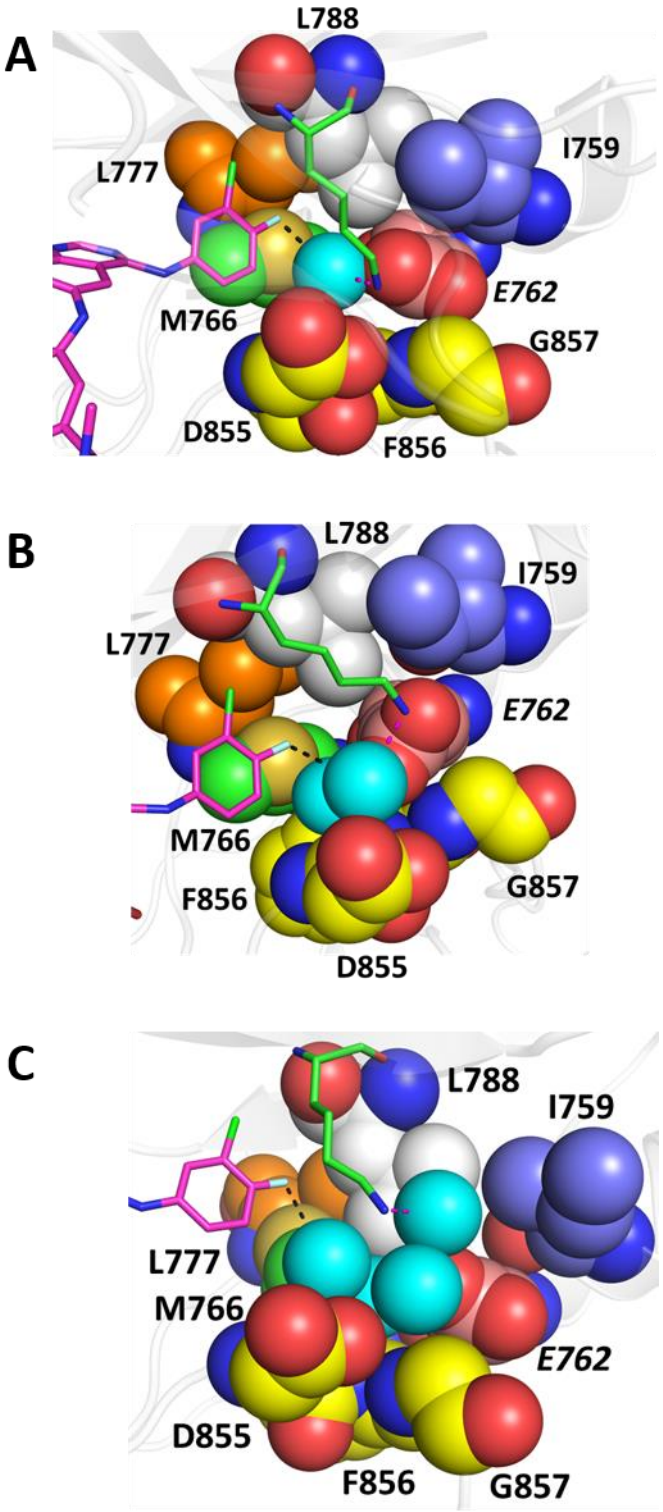

A
L788
I759
L777
E762
M766
G857
D855
F856
B
L788
I759
L777
E762
M766
F856
G857
D855
C
L788
I759
L777
M766
E762
D855
F856
G857

Figure 2

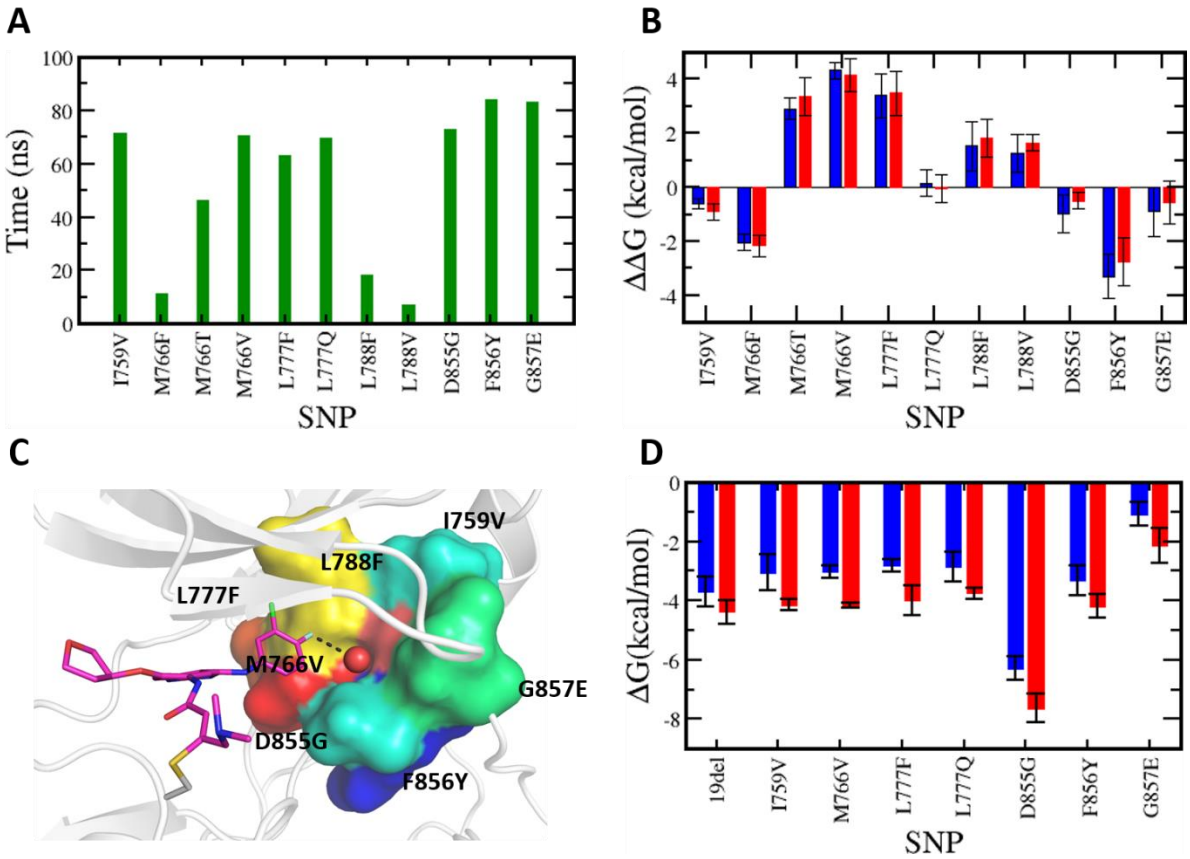

A
Time (ns)
100
80
60
40
20
0
I759V
M766F
M766T
M766V
L777F
L777Q
L788F
L788V
D855G
F856Y
G857E
SNP
B
ΔΔG (kcal/mol)
4
2
0
-2
-4
I759V
M766F
M766T
M766V
L777F
L777Q
L788F
L788V
D855G
F856Y
G857E
SNP
C
I759V
L788F
L777F
M766V
G857E
D855G
F856Y
D
ΔG(kcal/mol)
0
-2
-4
-6
-8
19del
I759V
M766V
L777F
L777Q
D855G
F856Y
G857E
SNP

Figure 3

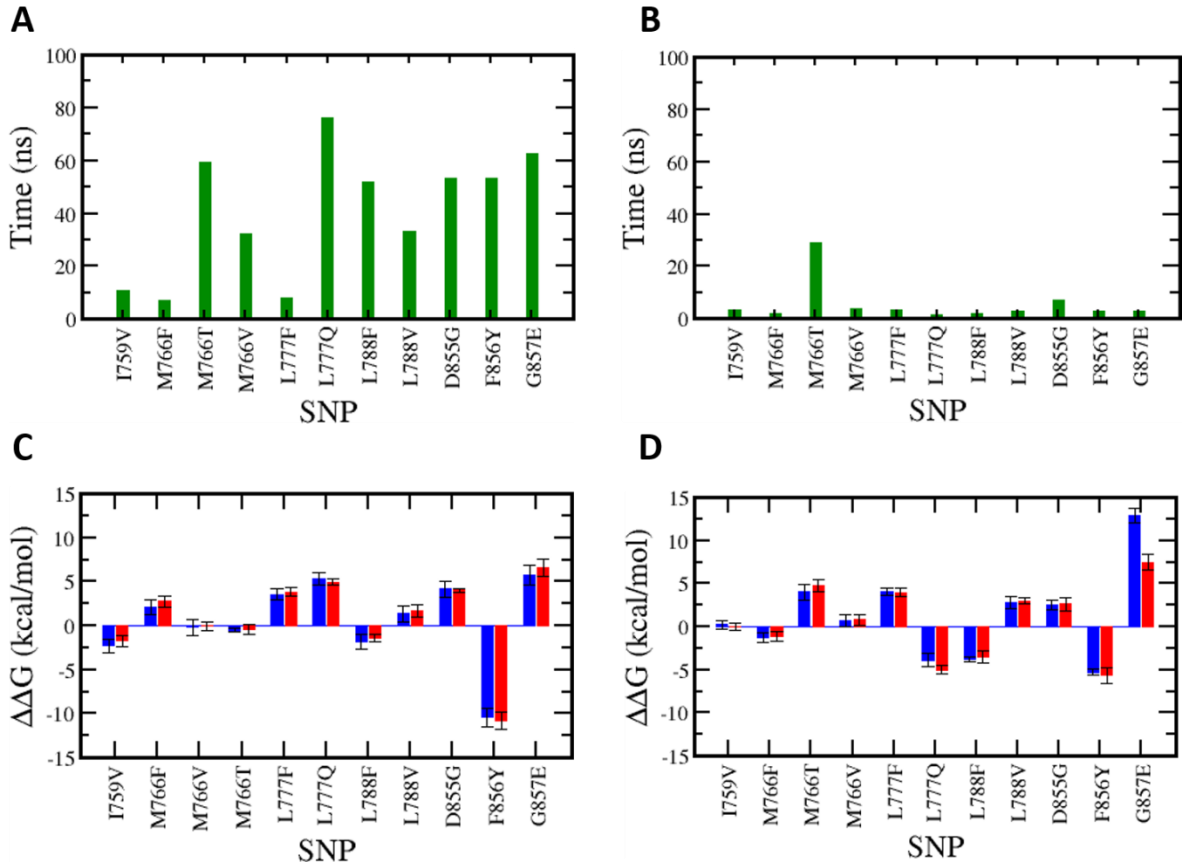

A
B
C
D
Time (ns)
ΔΔG (kcal/mol)
SNP
I759V
M766F
M766T
M766V
L777F
L777Q
L788F
L788V
D855G
F856Y
G857E

Figure 4

TOC Graphics