COMMUNICATIONS BIOLOGY

ARTICLE

DOI: 10.1038/s42003-018-0075-x OPEN

Predicting resistance of clinical Abl mutations to targeted kinase inhibitors using alchemical free-energy calculations

Kevin Hauser[1], Christopher Negron[1], Steven K. Albanese[2,3], Soumya Ray[1], Thomas Steinbrecher[4], Robert Abel[1], John D. Chodera[3] & Lingle Wang[1]

The therapeutic effect of targeted kinase inhibitors can be significantly reduced by intrinsic or acquired resistance mutations that modulate the affinity of the drug for the kinase. In cancer, the majority of missense mutations are rare, making it difficult to predict their impact on inhibitor affinity. We examine the potential for alchemical free-energy calculations to predict how kinase mutations modulate inhibitor affinities to Abl, a major target in chronic myelogenous leukemia (CML). These calculations have useful accuracy in predicting resistance for eight FDA-approved kinase inhibitors across 144 clinically identified point mutations, with a root mean square error in binding free-energy changes of $1.1^{1.3}_{0.9}$ kcal mol^{-1} (95% confidence interval) and correctly classifying mutations as resistant or susceptible with 88^{93}_{82}% accuracy. This benchmark establishes the potential for physical modeling to collaboratively support the assessment and anticipation of patient mutations to affect drug potency in clinical applications.

[1] Schrödinger, New York, NY 10036, USA. [2] Louis V. Gerstner, Jr. Graduate School of Biomedical Sciences, Memorial Sloan Kettering Cancer Center, New York, NY 10065, USA. [3] Computational and Systems Biology Program, Sloan Kettering Institute, Memorial Sloan Kettering Cancer Center, New York, NY 10065, USA. [4] Schrödinger-GmbH, Q7 23, 68161 Mannheim, Germany. Correspondence and requests for materials should be addressed to L.W. (email: lingle.wang@schrodinger.com)

Targeted kinase inhibitors are a major therapeutic class in the treatment of cancer. A total of 38 selective small-molecule kinase inhibitors have now been approved by the FDA[1], including 34 approved to treat cancer, and perhaps 50% of all current drugs in development target kinases[2]. Despite the success of selective inhibitors, the emergence of drug resistance remains a challenge in the treatment of cancer[3–10] and has motivated the development of second- and then third-generation inhibitors aimed at overcoming recurrent resistance mutations[11–15].

While a number of drug resistance mechanisms have been identified in cancer (e.g., induction of splice variants[16], or alleviation of feedback[17]), inherent or acquired missense mutations in the kinase domain of the target of therapy are a major form of resistance to tyrosine kinase inhibitors (TKI)[10,18,19]. Oncology is entering a new era with major cancer centers now deep sequencing tumors to reveal genetic alterations that may render subclonal populations susceptible or resistant to targeted inhibitors[20], but the use of this information in precision medicine has lagged behind. It would be of enormous value in clinical practice if an oncologist could reliably ascertain whether these mutations render the target of therapy resistant or susceptible to available inhibitors; such tools would facilitate the enrollment of patients in mechanism-based basket trials[21,22], help prioritize candidate compounds for clinical trials, and aid the development of next-generation inhibitors.

While some cancer missense mutations are highly recurrent and have been characterized clinically or biochemically, a long tail of rare mutations collectively accounts for the majority of clinically observed missense mutations (Fig. 1a), leaving clinicians and researchers without knowledge of whether these uncharacterized mutations might lead to resistance. While rules-based and machine learning schemes are still being assessed in oncology contexts, work in predicting drug response to microbial resistance has shown that rare mutations present a significant challenge to approaches that seek to predict resistance to therapy[23]. Clinical cancer mutations may impact drug response through a variety of mechanisms by altering kinase activity, ATP affinity, substrate specificities, and the ability to participate in regulatory interactions, compounding the difficulties associated with limited datasets that machine learning approaches face. In parallel with computational approaches, high-throughput experimental techniques such as MITE-Seq[24] have been developed to assess the impact of point mutations on drug response. However, the complexity of defining selection schemes that reliably correlate with in vivo drug effectiveness and long turn-around times might limit their ability to rapidly and reliably impact clinical decision-making.

Physics-based approaches could be complementary to machine-learning and experimental techniques in predicting changes in TKI affinity due to mutations with few or no prior clinical observations. Alchemical free-energy methods permit receptor-ligand binding energies to be computed rigorously, including all relevant entropic and enthalpic contributions[25]. Encouragingly, kinase:inhibitor binding affinities have been predicted using alchemical free-energy methods with mean unsigned errors of 1.0 kcal mol^{-1} for CDK2, JNK1, p38, and Tyk2[26–33]. Recently, one study has hinted at the potential utility of alchemical free-energy calculations in oncology by predicting the impact of a single clinical mutation on the binding free energies of the TKIs dasatinib and RL45[34].

Here, we ask whether physical modeling techniques may be useful in predicting whether clinically identified kinase mutations lead to drug resistance or drug sensitivity. We perform state-of-the-art relative alchemical free-energy calculations using FEP+[26], recently demonstrated to achieve sufficiently good accuracy to drive the design of small-molecule inhibitors for a broad range of

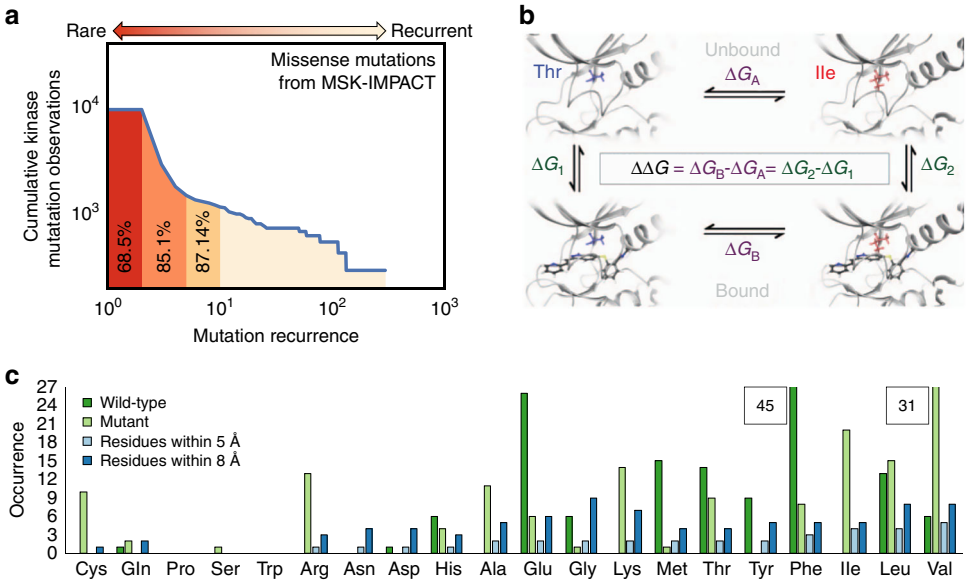


Fig. 1 Relative alchemical free-energy calculations can be used to predict affinity changes of FDA-approved selective kinase inhibitors arising from clinically identified mutations in their targets of therapy. **a** Missense mutation statistics derived from 10,336 patient samples subjected to Memorial Sloan Kettering-Integrated Mutation Profiling of Actionable Cancer Targets (MSK-IMPACT) deep sequencing panel[20] show that 68.5% of missense kinase mutations in cancer patients have never been observed previously, while 87.4% have been observed no more than ten times; the vast majority of clinically observed missense kinase mutations are unique to each patient. **b** To compute the impact of a clinical point mutation on inhibitor binding free energy, a thermodynamic cycle can be used to relate the free energy of the wild-type and mutant kinase in the absence (top) and presence (bottom) of the inhibitor. **c** Summary of mutations studied in this work. Frequency of the wild-type (dark green) and mutant (green) residues for the 144 clinically-identified Abl mutations used in this study (see Table 1 for data sources). Also shown is the frequency of residues within 5 Å (light blue) and 8 Å (blue) of the binding pocket. The ordering of residues along the *x*-axis corresponds to the increasing occurrence of residues within 5 Å of the binding pocket. The number of wild-type Phe residues ($n = 45$) and mutant Val residues ($n = 31$) exceeded the limits of the y-axis

targets during lead optimization[25–27,35], to calculate the effect of point mutation on the binding free energy between the inhibitor and the kinase receptor (Fig. 1b, c). We compare this approach against a fast but approximate physical modeling method implemented in Prime[36] (an MM-GBSA approach) in which an implicit solvent model is used to assess the change in minimized interaction energy of the ligand with the mutant and wild-type kinase. We consider whether these methods can predict a ten-fold reduction in inhibitor affinity (corresponding to a binding free-energy change of 1.36 kcal mol^{-1}) to assess baseline utility. As a benchmark, we compile a set of reliable inhibitor ΔpIC_{50} data for 144 clinically identified mutants of the human kinase Abl, an important oncology target dysregulated in cancers like chronic myelogenous leukemia (CML), for which six[1] FDA-approved TKIs are available. While ΔpIC_{50} can approximate a dissociation constant ΔK_D, other processes contributing to changes in cell viability might affect IC_{50} in ways that are not accounted for by a traditional binding experiment, motivating a quantitative comparison between ΔpIC_{50} and ΔK_D. The results of this benchmark demonstrate the potential for FEP+ to predict the impact that mutations in Abl kinase have on drug binding, and a classification accuracy of 88^{93}_{82}% (for all statistical metrics reported in this paper, the 95% confidence intervals (CI) is shown in the form of $\left(x^{\text{upper}}_{\text{lower}}\right)$), an RMSE of $1.07^{1.26}_{0.89}$ kcal mol^{-1}, and an MUE of $0.79^{0.92}_{0.67}$ kcal mol^{-1} was achieved.

Results

A benchmark of ΔpIC_{50}s for predicting mutational resistance. To construct a benchmark evaluation dataset, we compiled a total of 144 ΔpIC_{50} measurements of Abl:TKI affinities, summarized in Table 1 while ensuring all measurements for an individual TKI were reported in the same study from experiments run under identical conditions. 131 ΔpIC_{50} measurements were available across the six TKIs with available co-crystal structures with wild-type Abl—26 for axitinib and 21 for bosutinib, dasatinib, imatinib, nilotinib, and ponatinib. 13 ΔpIC_{50} measurements were available for the two TKIs for which docking was necessary to generate Abl:TKI structures—7 for erlotinib and 6 for gefitinib. For added diversity, this set includes TKIs for which Abl is not the primary target—axitinib, erlotinib, and gefitinib. All mutations in this benchmark dataset have been clinically observed (Supplementary Table 1). Due to the change in bond topology required by mutations involving proline, which is not currently supported by the FEP+ technology for protein residue mutations, the three mutations H396P (axitinib, gefitinib, erlotinib) were excluded from our assessment. As single-point mutations were highly represented in the Memorial Sloan Kettering-Integrated Mutation Profiling of Actionable Cancer Targets (MSK-IMPACT) study analyzed in Fig. 1a, we excluded double mutations from this work. However, the impact of mutations from multiple sites can potentially be modeled by sequentially mutating each site and this will be addressed in future work.

Experimental ΔpIC_{50} measurements for wild-type and mutant Abl were converted to $\Delta\Delta G$ in order to make direct comparisons between physics-based models and experiment. However, computation of experimental uncertainties were required to understand the degree to which differences between predictions and experimental data were significant. Since experimental error estimates for measured IC_{50}s were not available for the data in Table 1, we compared that data to other sources that have published IC_{50}s for the same mutations in the presence of the same TKIs (Fig. 2a–c). Cross-comparison of 97 experimentally measured $\Delta\Delta G$s derived from cell viability assay IC_{50} data led to an estimate of experimental variability of $0.32^{0.36}_{0.28}$ kcal mol^{-1} root mean square error (RMSE) that described the expected repeatability of the measurements. Because multiple factors influence the IC_{50} aside from direct effects on the binding affinity we also compared $\Delta\Delta G$s derived from ΔpIC_{50}s with those derived from binding affinity measurements (ΔK_d) for which data for a set of 27 mutations was available (Fig. 2d). The larger computed RMSE of $0.81^{1.04}_{0.59}$ kcal mol^{-1} represents an estimate of the lower bound of the RMSE to the IC_{50}-derived $\Delta\Delta G$s that we might hope to achieve with FEP+ or Prime, which were performed using non-phosphorylated models, when comparing sample statistics directly. Comparing 31 mutations for which phosphorylated and non-phosphorylated ΔK_ds were available, we found a strong correlation between the $\Delta\Delta G$s derived from those data ($r = 0.94$, Supplementary Figure 1).

Most mutations do not significantly reduce TKI potency. The majority of mutations do not lead to resistance by our 10-fold affinity loss threshold: 86.3% of the co-crystal set ($n = 113$) and 86.8% of the total set ($n = 125$). Resistance mutations, which are likely to result in a failure of therapy, constitute 13.7% of the co-crystal set ($n = 18$) and 13.2% of the total set of mutations ($n = 19$). The ΔpIC_{50}s for all 144 mutations are summarized in Supplementary Tables 2–7. Two mutations exceeded the dynamic range of the assays ($IC_{50} > 10{,}000$ nM); as these two mutations

Table 1 Public ΔpIC_{50} datasets for 144 Abl kinase mutations and eight TKIs with corresponding wild-type co-crystal structures used in this study

TKI	N_{mut}	R	S	PDB	(kcal mol^{-1}) $\lvert\Delta G_{max} - \Delta G_{min}\rvert$	Source	(kcal mol^{-1}) ΔG_{WT}
Axitinib	26	0	26	4wa9	2.05	52	−8.35
Bosutinib	21	4	17	3ue4	2.79	79	−9.81
Dasatinib	21	5	16	4xey	5.08	79	−11.94
Imatinib	21	5	16	1opj	2.16	79	−9.19
Nilotinib	21	4	17	3cs9	3.88	79	−10.74
Ponatinib	21	0	21	3oxz	1.00	79	−11.70
Subtotal	131	18	113				
Erlotinib	7	1	6	Dock to 3ue4	1.73	82	−9.77
Gefitinib	6	0	6	Dock to 3ue4	1.79	82	−8.84
Total	144	19	125				

N_{mut} Total number of mutants for which ΔpIC_{50} data was available
Number of **R**esistant, **S**usceptible mutants using 10-fold affinity change threshold
PDB Source PDB ID, or *Dock to 3ue4*, which used 3ue4 as the receptor for Glide-SP docking inhibitors without co-crystal structure
ΔG_{WT} Binding free energy of inhibitor to wild-type Abl, as estimated from IC_{50} data

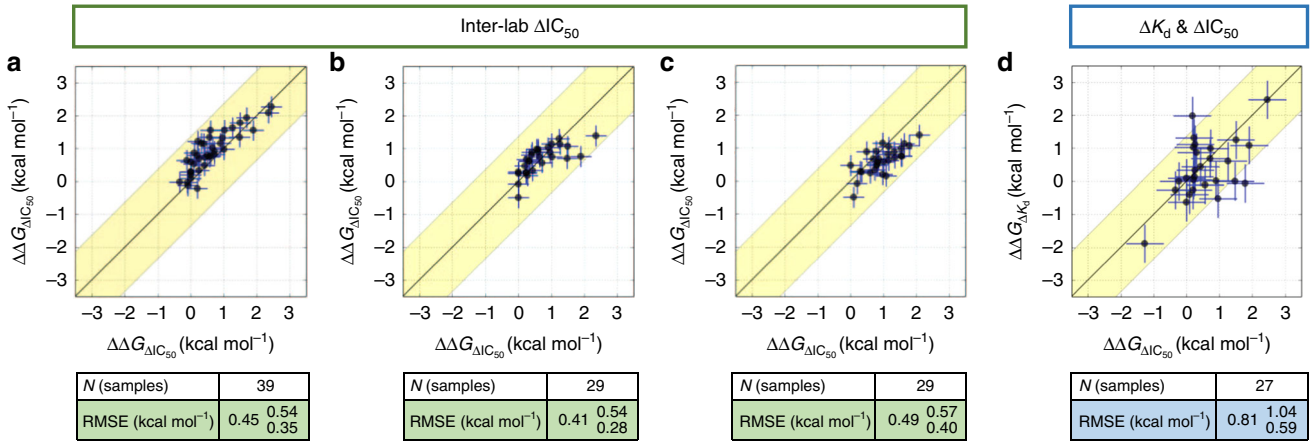


Fig. 2 Cross-comparison of the experimentally measured effects that mutations in Abl kinase have on ligand binding, performed by different labs. $\Delta\Delta G$ was computed from publicly available ΔpIC_{50} or ΔpK_d measurements and these values of $\Delta\Delta G$ were then plotted and the RMSE between them reported. **a** ΔpIC_{50} measurements (*X*-axis) from ref. [79] compared with ΔpIC_{50} measurements (*Y*-axis) from ref. [81]. **b** ΔpIC_{50} measurements (*X*-axis) from ref. [79] compared with ΔpIC_{50} measurements (*Y*-axis) from ref. [80]. **c** ΔpIC_{50} measurements (*X*-axis) from ref. [81] compared with ΔpIC_{50} measurements (*Y*-axis) from ref. [80]. **d** ΔpIC_{50} measurements (*X*-axis) from ref. [79] compared with ΔpK_d measurements (*Y*-axis) from ref. [82] using non-phosphorylated Abl kinase. Scatter plot error bars in (**a**–**c**) are ±standard error (SE) taken from the combined 97 inter-lab $\Delta\Delta G$s derived from the ΔpIC_{50} measurements, which was $0.32^{0.36}_{0.28}$; the RMSE was $0.45^{0.51}_{0.39}$ kcal mol^{-1}. Scatter plot error bars in (**d**) are the ±standard error (SE) of $\Delta\Delta G$s derived from ΔpIC_{50} and ΔpK_d from a set of 27 mutations, which is $0.58^{0.74}_{0.42}$ kcal mol^{-1}; the RMSE was $0.81^{1.04}_{0.59}$ kcal mol^{-1}

clearly raise resistance, we excluded them from quantitative analysis (RMSE and MUE) but included them in truth table analyses and classification metrics (accuracy, specificity, and sensitivity).

FEP+ predicts affinity changes for clinical Abl mutants. Figure 1b depicts the thermodynamic cycle that illustrates how we used relative free-energy calculations to compute the change in ligand binding free energy in response to the introduction of a point mutation in the kinase (Fig. 1c). From prior experience with relative alchemical free-energy calculations for ligand design, good initial receptor-ligand geometry was critical to obtaining accurate and reliable free-energy predictions[26], so we first focused on the 131 mutations in Abl kinase across six TKIs for which wild-type Abl:TKI co-crystal structures were available. Figure 3 summarizes the performance of predicted binding free-energy changes ($\Delta\Delta G$) for all 131 mutants in this set for both a fast MM-GBSA physics-based method that only captures interaction energies for a single structure (Prime) and rigorous alchemical free-energy calculations (FEP+). Scatter plots compare experimental and predicted free-energy changes ($\Delta\Delta G$) and characterize the ability of these two techniques to predict experimental measurements. Statistical uncertainty in the predictions and experiment-to-experiment variability in the experimental values are shown as ellipse height and widths, respectively. The value for experimental variability was 0.32 kcal mol^{-1}, which was the standard error computed from the cross-comparison in Fig. 2. For FEP+, the uncertainty was taken to be the standard error of the average from three independent runs for a particular mutation, while Prime results are deterministic and are not contaminated by statistical uncertainty.

To better assess whether discrepancies between experimental and computed $\Delta\Delta G$s simply arise for known forcefield limitations or might indicate more significant effects, we incorporated an additional error model in which the forcefield error was taken to be a random error $\sigma_{FF} \approx 0.9$ kcal mol^{-1}, a value established form previous benchmarks on small molecules absent conformational sampling or protonation state issues[37]. Thin error bars in Fig. 2 represent the overall estimated error due to both this forcefield error and experimental variability or statistical uncertainty[38,39].

To assess overall quantitative accuracy, we computed both RMSE—which is rather sensitive to outliers, and mean unsigned error (MUE). For Prime, the MUE was $1.16^{1.37}_{0.96}$ kcal mol^{-1} and the RMSE was $1.72^{2.00}_{1.41}$ kcal mol^{-1}. FEP+, the alchemical free-energy approach, achieved a significantly higher level of quantitative accuracy with an MUE of $0.82^{0.95}_{0.69}$ kcal mol^{-1} and an RMSE of $1.11^{1.30}_{0.91}$ kcal mol^{-1}. Notably, alchemical free-energy calculations come substantially closer than MMGBSA approach to the minimum achievable RMSE of $0.81^{1.04}_{0.59}$ kcal mol^{-1} (due to experimental error; Fig. 2) for this dataset.

FEP+ accurately classifies affinity changes for Abl mutants. While quantitative accuracy (MUE, RMSE) is a principle metric of model performance, an application of potential interest is the ability to classify mutations as raising resistance to a specific TKI. To characterize the accuracy with which Prime and FEP+ classified mutations in a manner that might be therapeutically relevant, we classified mutations by their experimental impact on the binding affinity as susceptible (affinity for mutant is diminished by no more than 10-fold, $\Delta\Delta G \le 1.36$ kcal mol^{-1}) or as resistant (affinity for mutant is diminished by least 10-fold, $\Delta\Delta G > 1.36$ kcal mol^{-1}). Summary statistics of experimental and computational predictions of these classes are shown in Fig. 2 (bottom) as truth tables (also known as confusion matrices).

The simple minimum-energy scoring method Prime correctly classified 9 of the 18 resistance mutations in the dataset while merely 85 of the 113 susceptible mutations were correctly classified (28 false positives). In comparison, the alchemical free-energy method FEP+, which includes entropic and enthalpic contributions as well as explicit representation of solvent, correctly classified 9 of the 18 resistance mutations while a vast majority, 105, of the susceptible mutations were correctly classified (merely 8 false positives). Prime achieved a classification accuracy of $0.72^{0.79}_{0.64}$, while FEP+ achieved an accuracy that is significantly higher (both in a statistical sense and in overall magnitude), achieving an accuracy of $0.87^{0.92}_{0.81}$. Sensitivity (also called true positive rate) and specificity (true negative rate) are also informative statistics in assessing the performance of a binary classification scheme. For Prime, the sensitivity was $0.50^{0.73}_{0.25}$, while the specificity was $0.75^{0.83}_{0.67}$. To put this in perspective, a CML

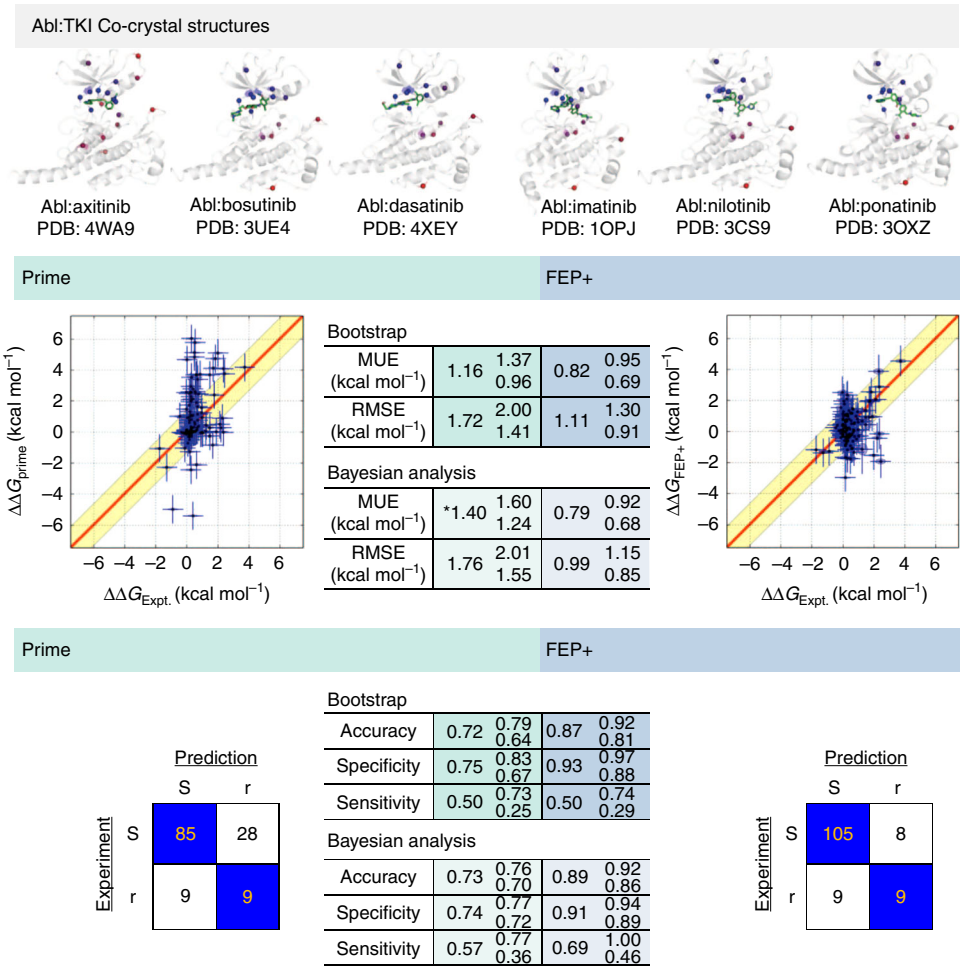


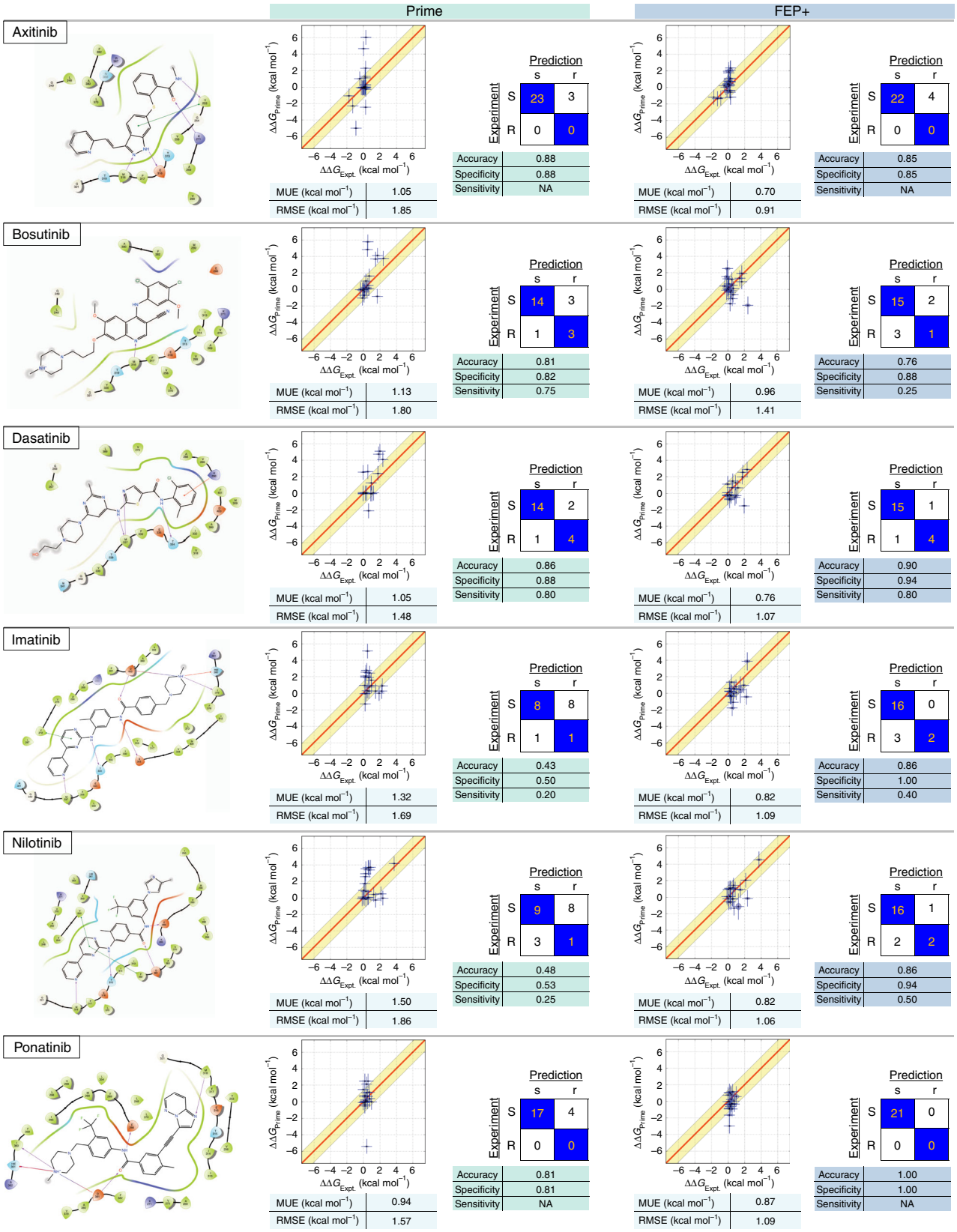
Fig. 3 Comparison of experimentally measured binding free-energy changes ($\Delta\Delta G$) for 131 clinically observed mutations and 6 targeted kinase inhibitors (TKI). Co-crystal structures are publicly available for wild-type Abl kinase (see Methods) bound to these inhibitors. Top panel: Abl:TKI co-crystal structures (protein is gray; TKI is green) with positions of point mutations shown as spheres colored from blue (near) to red (far) by relative distance from the inhibitor. Middle panel: Scatter plots show Prime and FEP+ computed $\Delta\Delta G$ compared to experiment. Variability (ellipses) in experimental $\Delta\Delta G$ (standard error between IC_{50}-derived $\Delta\Delta G$ measurements made by different labs, 0.32 kcal mol^{-1}) and computed $\Delta\Delta G$ ($\pm\sigma = 0$ kcal mol^{-1} for Prime while for FEP+ the standard error of the mean from 3 independent runs). Experimental error bars (σ_{exp}) are the standard error between ΔpIC_{50} and ΔK_d measurements, 0.58 kcal mol^{-1}. To better highlight true outliers unlikely to simply result from expected forcefield error, we presume forcefield error ($\sigma_{FF} \approx 0.9$ kcal mol^{-1}[137]) also behaves as a random error, and represent the total estimated statistical and forcefield error $\left(\sqrt{\sigma_{FF}^2 + \sigma_{exp/cal}^2}\right)$ as vertical error bars. The yellow region indicates area in which predicted $\Delta\Delta G$ is within 1.36 kcal mol^{-1} of experiment. Two mutations were beyond the concentration limit of the assay and were not plotted; $N = 129$. Bottom panel: Truth tables and classification results include T315I/dasatinib and L248R/imatinib; 131 points were used. Truth tables of classification accuracy, sensitivity and specificity using two-classes (resistant: $\Delta\Delta G > 1.36$ kcal/mol; $\Delta\Delta G \leq 1.36$ kcal/mol). For MUE, RMSE, and classification statistics, sub/superscripts denote 95 % CIs. For Prime, *MUE highlights that the Bayesian model yields a value for MUE that is noticeably larger than MUE for observed data due to the non-Gaussian error distribution of Prime

patient bearing a resistance mutation in the kinase domain of Abl has an equal chance of Prime correctly predicting this mutation would be resistant to one of the TKIs considered here, while if the mutation was susceptible, the chance of correct prediction would be ~75%. By contrast, the classification specificity of FEP+ was substantially better. For FEP+, the sensitivity was $0.50^{0.74}_{0.29}$ while the specificity was $0.93^{0.97}_{0.88}$. There is a very high probability that FEP+ will correctly predict that one of the eight TKIs studied here will remain effective for a patient bearing a susceptible mutation.

How reliant are classification results on choice of cutoff? Previous work by O'Hare et al. utilized TKI-specific thresholds for dasatinib, imatinib, and nilotinib[40], which were ~2 kcal mol^{-1}. Supplementary Figure 2 shows that when our classification threshold was increased to a 20-fold change in binding (1.77 kcal mol^{-1}), FEP+ correctly classified 8 of the 13 resistant mutations and with a threshold of 100-fold change in binding (2.72 kcal mol^{-1}), FEP+ correctly classified the only two resistant mutations (T315I/dasatinib and T315I/nilotinib). With the extant multilayered and multinodal decision-making algorithms used by experienced oncologists to manage their patients' treatment, or by medicinal chemists to propose candidate compounds for clinical trials, the resistant or susceptible cutoffs could be selected with more nuance than the simple 10-fold affinity threshold we consider here. With a larger affinity change cutoff, for example, the accuracy with which physical models predict resistance mutations increases beyond 90% (Supplementary Figure 2). For the alchemical

approach, classification accuracy was $0.92^{0.96}_{0.87}$ when an affinity change cutoff of 20-fold was used while using an affinity change cutoff of 100-fold further improved the accuracy to $0.98^{1.00}_{0.96}$.

Bayesian analysis can estimate the true error. The statistical metrics—MUE, RMSE, accuracy, specificity, and sensitivity—discussed above are based on analysis of the apparent performance of the observed modeling results compared with the



observed experimental data via sample statistics. However, this analysis considers a limited number of mutants, and both measurements and computed values are contaminated with experimental or statistical error. To obtain an estimate of the intrinsic performance of our physical modeling approaches, accounting for known properties of the experimental variability and statistical uncertainties, we used a hierarchical Bayesian model to infer posterior predictive distributions from which expectations and 95% predictive intervals could be obtained. The results of this analysis are presented in Fig. 3 (central tables).

FEP+ is significantly better than Prime at predicting the impact of mutations on TKI binding affinities, as the apparent performance as well as the intrinsic performance were well-separated outside their 95% CI in nearly all metrics. Applying the Bayesian model, the MUE and RMSE for FEP+ was $0.79^{0.92}_{0.68}$ and $0.99^{1.15}_{0.85}$ kcal mol^{-1}, respectively ($N = 129$). For the classification metrics accuracy, specificity, and sensitivity, the model yields $0.89^{0.92}_{0.86}$, $0.91^{0.94}_{0.89}$, and $0.69^{1.00}_{0.46}$, respectively ($N = 131$). The intrinsic RMSE and MUE of Prime was $1.76^{2.01}_{1.55}$ and $1.40^{1.60}_{1.24}$ kcal mol^{-1} ($N = 129$), respectively, and the classification accuracy, specificity, and sensitivity was $0.73^{0.76}_{0.70}$, $0.74^{0.77}_{0.72}$, and $0.57^{0.77}_{0.36}$, respectively ($N = 131$). The intrinsic MUE of Prime obtained by this analysis is larger than the observed MUE reflecting the non-Gaussian, fat-tailed error distributions of Prime results.

How transferable is FEP+ across the six TKIs? The impact of point mutations on drug binding are not equally well predicted for the six TKIs. Figure 4 expands the results in Fig. 3 on a TKI-by-TKI basis to dissect the particular mutations in the presence of a specific TKI. Prime and FEP+ correctly predicted that most mutations in this dataset ($N = 26$) do not raise resistance to axitinib, though FEP+ predicted 4 false positives compared with 3 false positives by Prime. The MUE and RMSE of FEP+ was excellent for this inhibitor, $0.70^{0.93}_{0.50}$ and $0.91^{1.14}_{0.64}$ kcal mol^{-1}, respectively. While the classification results for bosutinib ($N = 21$) were equally well predicted by Prime as by FEP+, FEP+ was still able to achieve superior, but not significant, predictive performance for the quantitative metrics MUE and RMSE, which were $0.96^{1.42}_{0.55}$ and $1.41^{1.97}_{0.77}$ kcal mol^{-1}, respectively (FEP+) and $1.13^{1.83}_{0.60}$ and $1.80^{2.62}_{0.92}$ kcal mol^{-1}, respectively (Prime). For dasatinib, FEP+ achieved an MUE and RMSE of $0.76^{1.13}_{0.49}$ and $1.07^{1.57}_{0.59}$ kcal mol^{-1}, respectively, whereas the results were, as expected, less quantitatively predictive for Prime ($N = 20$). The results for imatinib were similar to those of dasatinib above, where the MUE and RMSE for FEP+ were $0.82^{1.15}_{0.53}$ and $1.09^{1.43}_{0.69}$ kcal mol^{-1}, respectively (N = 20). Nilotinib, a derivative of imatinib, led to nearly identical quantitative performance results for FEP+ with an MUE and RMSE of $0.82^{1.12}_{0.57}$ and $1.06^{1.39}_{0.69}$ kcal mol^{-1}, respectively ($N = 21$). Similar to axitinib, ponatinib presented an interesting case because there were no mutations in this dataset that raised resistance to it. Despite the wide dynamic range in the computed values of $\Delta\Delta G$ for other inhibitors, FEP+ correctly predicted a narrow range of $\Delta\Delta G$s for this drug. This is reflected in the MUE and RMSE of $0.87^{1.16}_{0.62}$ and $1.09^{1.46}_{0.70}$ kcal mol^{-1}, respectively, which are in-line with the MUEs and RMSEs for the other TKIs.

Understanding the origin of mispredictions. Resistance mutations that are mispredicted as susceptible are particularly critical because they might mislead the clinician or drug designer into believing the inhibitor will remain effective against the target. Which resistance mutations did FEP+ mispredict as susceptible? Nine mutations were classified by FEP+ to be susceptible when experimentally measured ΔpIC_{50} data indicate the mutations should have increased resistance according to our 10-fold affinity cutoff for resistance. Notably, the 95% CI for five of these mutations included the 1.36 kcal mol^{-1} threshold, indicating these misclassifications are not statistical significant when the experimental error and statistical uncertainty in FEP+ are accounted for: bosutinib/L248R ($\Delta\Delta G_{FEP+} = 1.32^{1.94}_{0.70}$ kcal mol^{-1}), imatinib/E255K ($\Delta\Delta G_{FEP+} = 0.43^{3.05}_{-2.19}$ kcal mol^{-1}), imatinib/Y253F ($\Delta\Delta G_{FEP+} = 0.95^{1.64}_{0.26}$ kcal mol^{-1}), and nilotinib/Y253F ($\Delta\Delta G_{FEP+} = 0.89^{1.69}_{0.09}$ kcal mol^{-1}). The bosutinib/V299L mutation was also not significant because the experimental $\Delta\Delta G$, $1.70^{2.33}_{1.08}$ kcal mol^{-1}, included the 1.36 kcal mol^{-1} cutoff; the value of $\Delta\Delta G$ predicted by FEP+ for this mutation was $0.91^{1.02}_{0.79}$ kcal mol^{-1}, the upper bound of the predicted value was within 0.06 kcal mol^{-1} of the lower bound of the experimental value.

Four mutations, however, were misclassified to a degree that is statistically significant: dasatinib/T315A, bosutinib/T315I, imatinib/E255V, and nilotinib/E255V. For dasatinib/T315A, although the T315A mutations for bosutinib, imatinib, nilotinib, and ponatinib were correctly classified as susceptible, the predicted free-energy changes for these four TKIs were consistently more negative than the corresponding experimental measurements, like dasatinib/T315A, indicating there might be a generic driving force contributing to the errors in T315A mutations for these five TKIs. Abl is known to be able to adopt many different conformations (including DFG-in and DFG-out), and it is very likely that the T315A mutation induces conformational changes in the apo protein[41], the inadequate sampling of which may have led to the errors for the T315A mutation. By comparison, the T315I mutations for axitinib, bosutinib, imatinib, nilotinib, and ponatinib were all accurately predicted with the exception of bosutinib/T315I being the only misprediction, suggesting an issue specific to bosutinib. The interactions between the 2,4-dichloro-5-methoxyphenyl ring in bosutinib and the positively charged amine of the catalytic Lys271 may not be accurately captured by the fixed-charge OPLS3 force field, possibly leading to the misprediction for bosutinib/T315I mutation.

Insufficient sampling might also belie the imatinib/E255V and nilotinib/E255V mispredictions because they reside in the highly flexible P-loop. Since E255V was a charge change mutation, we utilized a workflow that included a transmutable explicit ion (see Methods). The distribution of these ions in the simulation box around the solute might not have converged to their equilibrium state on the relatively short timescale of our simulations (5 ns),

Fig. 4 Physical modeling accuracy in computing the impact of clinical Abl mutations on selective inhibitor binding. Ligand interaction diagrams for six selective FDA-approved TKIs for which co-crystal structures with Abl were available (left). Comparisons for clinically observed mutations are shown for FEP+ (right) and Prime (left). For each ligand, computed *vs.* experimental binding free energies ($\Delta\Delta G$) are plotted with MUE and RMSE (units of kcal mol^{-1}) depicted below. Truth tables are shown to the right. Rows denote true susceptible (S, $\Delta\Delta G \leq 1.36$ kcal mol^{-1}) or resistant (R, $\Delta\Delta G >$ kcal mol^{-1}) experimental classes using a 1.36 kcal mol^{-1} (10-fold change) threshold; columns denote predicted susceptible (s, $\Delta\Delta G \leq$ kcal mol^{-1}) or resistant (r, $\Delta\Delta G >$ kcal mol^{-1}). Correct predictions populate diagonal elements (orange text), incorrect predictions populate off-diagonals. Accuracy, specificity, and sensitivity for two-class classification are shown below the truth table. Elliptical point sizes and error bars in the scatter plots depict estimated uncertainty/variability and error, respectively, ($\pm\sigma$) of FEP+ values (vertical size) and experimental values (horizontal size). Note: The sensitivity for axitinib and ponatinib is NA, because there is no resistant mutation for these two drugs

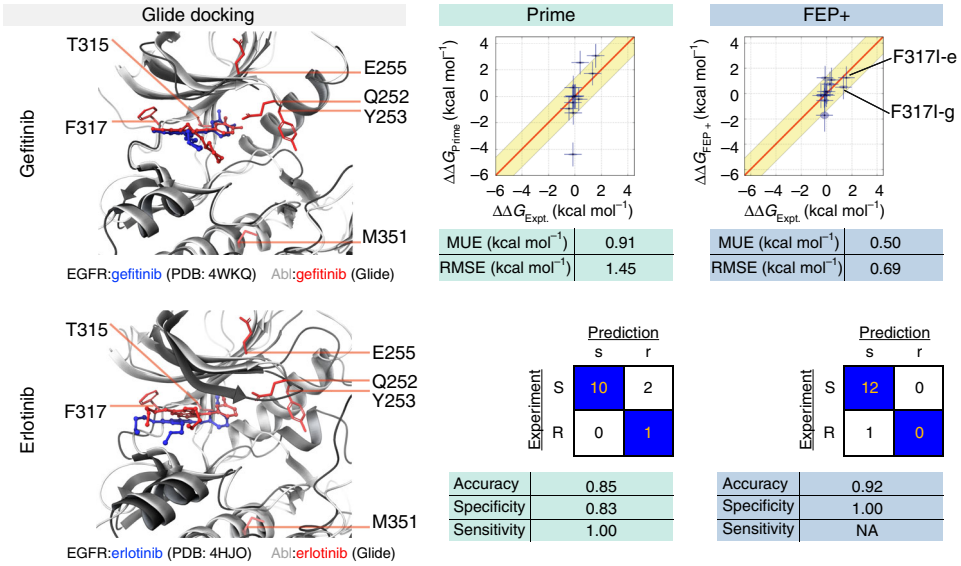


Fig. 5 Predicting resistance mutations using FEP+ for inhibitors for which co-crystal structures with wild-type kinase are not available. The docked pose of Abl:erlotinib is superimposed on the co-crystal structure of EGFR:erlotinib; erlotinib docked to Abl (light gray) is depicted in green and erlotinib bound to EGFR (dark gray) is depicted in blue. The docked pose of Abl:gefitinib is superimposed on the co-crystal structure of EGFR:gefitinib; gefitinib docked to Abl (light gray) is depicted in green and gefitinib bound to EGFR (dark gray) is depicted in blue. The locations of clinical mutants for each inhibitor are highlighted (red spheres). The overall RMSEs and MUEs for Prime (center) and FEP+ (right) and two-class accuracies are also shown in the figure. Computed free-energy changes due to the F317I mutation for erlotinib (−e) and gefitinib (−g) are highlighted in the scatter plot. FEP+ results are based on the docked models prepared with crystal waters added back while the Prime (an implicit solvent model) results are based on models without crystallographic water

and the insufficient sampling of ion distributions coupled with P-loop motions might lead to misprediction of these two mutations.

How strongly is accuracy affected for docked TKIs? To assess the potential for utilizing physics-based approaches in the absence of a high-resolution experimental structure, we generated models of Abl bound to two TKIs—erlotinib and gefinitib—for which co-crystal structures with wild-type kinase are not currently available. In Fig. 5, we show the Abl:erlotinib and Abl:gefitinib complexes that were generated using a docking approach (Glide-SP, see Methods). These two structures were aligned against the co-crystal structures of EGFR:erlotinib and EGFR:gefinitib to highlight the structural similarities between the binding pockets of Abl and EGFR and the TKI binding mode in Abl versus EGFR. As an additional test of the sensitivity of FEP+ to system preparation, a second set of Abl:erlotinib and Abl:gefitinib complexes was generated in which crystallographic water coordinates were transferred to the docked inhibitor structures (see Methods).

Alchemical free-energy simulations were performed on 13 mutations between the two complexes; 7 mutations for erlotinib and 6 mutations for gefitinib. The quantitative accuracy of FEP+ in predicting the value of $\Delta\Delta G$ was excellent—MUE and RMSE of $0.58^{0.86}_{0.33}$ and $0.80^{1.09}_{0.44}$ kcal mol^{-1}, respectively, if crystal waters are omitted, and $0.50^{0.78}_{0.26}$ kcal mol^{-1} and $0.69^{0.97}_{0.35}$ kcal mol^{-1} if crystal waters were restored after docking. Encouragingly, these results indicate that our initial models of Abl bound to erlotinib and gefitinib were reliable because the accuracy and dependability of our FEP+ calculations were not sensitive to crystallographic waters. Our secondary concern was the accuracy with which the approach classified mutations as resistant or susceptible.

While the results presented in (Fig. 5) indicate that FEP+ is capable of achieving good quantitative accuracy when a co-crystal structure is unavailable, it is important to understand why a mutation was predicted to be susceptible but was determined experimentally to be resistant. F317I was the one mutation that increased resistance to erlotinib (or gefitinib) because it destabilized binding by more than 1.36 kcal mol^{-1}—$1.35^{1.67}_{1.03}$ kcal mol^{-1} (gefitinib) and $1.58^{1.90}_{1.26}$ kcal mol^{-1} (erlotinib), but the magnitude of the experimental uncertainty means we are unable to confidently discern whether this mutation induces more than 10-fold resistance to either TKI. Therefore, the one misclassification by FEP+ in Fig. 5 is not statistically significant and the classification metrics presented there underestimate the nominal performance of this alchemical free-energy method.

Discussion

The results presented in this work are summarized in Table 2. The performance metrics summarized in Table 2 indicates that the set of 131 mutations for the six TKIs in which co-crystal structures were available is on par with the complete set (144 mutations), which included results based on Abl:TKI complexes generated from docking models. The performance results for the 13 mutations for the two TKIs (erlotinib and gefitinib) in which co-crystal structures were unavailable exhibited good quantitative accuracy (MUE and RMSE) and good classification power.

Overall ($N = 144$), the MM-GBSA approach Prime classified mutations with good accuracy $\left(0.73^{0.80}_{0.66}\right)$ and specificity $\left(0.76^{0.84}_{0.69}\right)$ while the alchemical approach FEP+ was a significant improvement in classification accuracy $\left(0.88^{0.93}_{0.82}\right)$ and specificity $\left(0.94^{0.98}_{0.89}\right)$. The quantitative accuracy with which Prime was able to predict the experimentally measured change in Abl:TKI binding ($N = 142$) characterized by RMSE and MUE was $1.70^{1.98}_{1.40}$ and $1.14^{1.35}_{0.93}$ kcal mol^{-1}, respectively. In stark contrast, the quantitative accuracy of FEP+ was statistically superior to Prime with an RMSE and an MUE of $1.07^{1.26}_{0.89}$ and $0.79^{0.92}_{0.67}$ kcal mol^{-1}, respectively.

Table 2 Summary of FEP+ and Prime statistics in predicting mutational resistance or sensitivity to FDA-approved TKIs

Dataset	Method	N_{quant}	MUE (kcal mol^{-1})	RMSE (kcal mol^{-1})	N_{class}	Accuracy	Specificity	Sensitivity
all	FEP+	142	$0.79^{0.92}_{0.67}$	$1.07^{1.26}_{0.89}$	144	$0.88^{0.93}_{0.82}$	$0.94^{0.98}_{0.89}$	$0.47^{0.69}_{0.25}$
all	Prime	142	$1.14^{1.35}_{0.93}$	$1.70^{1.98}_{1.40}$	144	$0.73^{0.80}_{0.66}$	$0.76^{0.84}_{0.69}$	$0.53^{0.76}_{0.30}$
xtals	FEP+	129	$0.82^{0.95}_{0.69}$	$1.11^{1.30}_{0.91}$	131	$0.87^{0.92}_{0.81}$	$0.93^{0.97}_{0.88}$	$0.50^{0.74}_{0.29}$
xtals	Prime	129	$1.16^{1.37}_{0.96}$	$1.72^{2.00}_{1.41}$	131	$0.72^{0.79}_{0.64}$	$0.75^{0.83}_{0.67}$	$0.50^{0.73}_{0.25}$
axitinib	FEP+	26	$0.70^{0.93}_{0.50}$	$0.91^{1.14}_{0.64}$	26	$0.85^{0.96}_{0.69}$	$0.85^{0.96}_{0.69}$	NA
axitinib	Prime	26	$1.05^{1.71}_{0.53}$	$1.85^{2.61}_{0.96}$	26	$0.88^{1.00}_{0.73}$	$0.88^{1.00}_{0.73}$	NA
bosutinib	FEP+	21	$0.96^{1.42}_{0.55}$	$1.41^{1.97}_{0.77}$	21	$0.76^{0.95}_{0.57}$	$0.88^{1.00}_{0.71}$	$0.25^{1.00}_{0.00}$
bosutinib	Prime	21	$1.13^{1.83}_{0.60}$	$1.80^{2.62}_{0.92}$	21	$0.81^{0.95}_{0.62}$	$0.82^{1.00}_{0.62}$	$0.75^{1.00}_{0.00}$
dasatinib	FEP+	20	$0.76^{1.13}_{0.49}$	$1.07^{1.57}_{0.59}$	21	$0.90^{1.00}_{0.76}$	$0.94^{1.00}_{0.79}$	$0.80^{1.00}_{0.33}$
dasatinib	Prime	20	$1.05^{1.54}_{0.61}$	$1.48^{1.92}_{0.95}$	21	$0.86^{1.00}_{0.71}$	$0.88^{1.00}_{0.69}$	$0.80^{1.00}_{0.33}$
imatinib	FEP+	20	$0.82^{1.15}_{0.53}$	$1.09^{1.43}_{0.69}$	21	$0.86^{1.00}_{0.71}$	$1.00^{1.00}_{1.00}$	$0.40^{0.83}_{0.00}$
imatinib	Prime	20	$1.32^{1.81}_{0.91}$	$1.69^{2.26}_{1.15}$	21	$0.43^{0.67}_{0.24}$	$0.50^{0.75}_{0.25}$	$0.20^{0.67}_{0.00}$
nilotinib	FEP+	21	$0.82^{1.12}_{0.57}$	$1.06^{1.39}_{0.69}$	21	$0.86^{1.00}_{0.67}$	$0.94^{1.00}_{0.80}$	$0.50^{1.00}_{0.00}$
nilotinib	Prime	21	$1.50^{1.97}_{1.06}$	$1.86^{2.25}_{1.43}$	21	$0.48^{0.67}_{0.24}$	$0.53^{0.75}_{0.29}$	$0.25^{1.00}_{0.00}$
ponatinib	FEP+	21	$0.87^{1.16}_{0.62}$	$1.09^{1.46}_{0.70}$	21	$1.00^{1.00}_{1.00}$	$1.00^{1.00}_{1.00}$	*NA*
ponatinib	Prime	21	$0.94^{1.54}_{0.50}$	$1.57^{2.44}_{0.69}$	21	$0.81^{0.95}_{0.62}$	$0.81^{0.95}_{0.62}$	*NA*
Glide	FEP+	13	$0.50^{0.78}_{0.26}$	$0.69^{0.97}_{0.35}$	13	$0.92^{1.00}_{0.77}$	$1.00^{1.00}_{1.00}$	$0.00^{0.00}_{0.00}$
Glide	Prime	13	$0.91^{1.56}_{0.39}$	$1.45^{2.22}_{0.54}$	13	$0.85^{1.00}_{0.62}$	$0.83^{1.00}_{0.58}$	$1.00^{1.00}_{0.00}$

Accuracy, specificity, and sensitivity were computed to assess two-class prediction performance:
resistant ($\Delta\Delta G > 1.36$ kcal mol^{-1}) or susceptible ($\Delta\Delta G \leq 1.36$ kcal mol^{-1})
95% CIs (sub-/superscripts) were estimated from 1000 bootstrap replicates. The sensitivity for axitinib and ponatinib is NA, because there is no resistant mutation for these two drugs
N_{quant} Number of mutations for which quantitative metrics were evaluated, N_{class} number mutations for which classification metrics were evaluated, *All* all mutations, *xtals* all mutations for which co-crystal structures were available, *Glide* erlotinib and gefitinib

From the perspective of a clinician, classification rate would be an important metric to measure the predictive power of technologies such as Prime and FEP+. To test the hypothesis that reducing the large spread in Prime predictions could improve its classification rate, we scaled the computed relative free energies and recalculated the performance metrics (Supplementary Table 8). As expected, the MUE and RMSE were improved but the specificity of Prime was drastically diminished. Scaling FEP+ eliminated its sensitivity and a naive model (all $\Delta\Delta G$s = 0.00 kcal mol^{-1}) had zero sensitivity. Lastly, we constructed a consensus model in which free energies were a weighted average of scaled Prime and FEP+. This model also had zero sensitivity.

To address the impact of picking a cutoff to classify predicted free energies as resistant or sensitizing, we computed ROC curves for the various predicted datasets: Prime, FEP+, naive model, and consensus model (Supplementary Figure 3). ROC-AUC for FEP+ was $0.75^{0.90}_{0.61}$ ($n = 144$); ROC-AUC for Prime was $0.66^{0.81}_{0.52}$ ($n = 144$); ROC-AUCs for the naive model and consensus model were $0.50^{0.50}_{0.50}$ ($n = 144$) and $0.78^{0.90}_{0.67}$ ($n = 144$), respectively. These results show that Prime has poor discriminatory power (ROC-AUC in [0.6,0.7]) while FEP+ has fair discriminatory power (ROC-AUC in [0.7,0.8]).

A hierarchical Bayesian approach was developed to estimate the intrinsic accuracy of the models when the noise in the experimental and predicted values of $\Delta\Delta G$ was accounted for. Utilizing this approach, the MUE and RMSE for Prime was found to be $1.39^{1.58}_{1.23}$ and $1.75^{1.98}_{1.55}$ kcal mol^{-1} ($N = 142$), respectively. The accuracy, specificity, and sensitivity of Prime was found using this method to be $0.74^{0.76}_{0.71}$, $0.75^{0.77}_{0.73}$, and $0.59^{0.78}_{0.40}$ (N = 144) respectively. The MUE and RMSE of FEP+ was found to be $0.76^{0.87}_{0.66}$ and $0.95^{1.09}_{0.82}$ kcal mol^{-1} ($N = 142$), respectively, which is significantly better than Prime. Likewise, a clearer picture of the true classification accuracy, specificity, and sensitivity of FEP+ was found—$0.90^{0.93}_{0.86}$, $0.92^{0.95}_{0.90}$, and $0.68^{1.00}_{0.46}$, respectively.

The high accuracy of FEP+ is very encouraging, and the accuracy can be further improved with more accurate modeling of a number of physical chemical effects not currently considered by the method. While highly optimized, the fixed-charged OPLS3[37] force field can be further improved by explicit consideration of polarizability effects[42], as hinted by some small-scale benchmarks[43]. These features could be especially important for bosutinib, whose 2,4-dichloro-5-methoxyphenyl ring is adjacent to the positively charged amine of the catalytic Lys271. Many simulation programs also utilize a long-range isotropic analytical dispersion correction intended to correct for the truncation of dispersion interactions at finite cutoff, which can induce an error in protein–ligand binding free energies that depends on the number of ligand heavy atoms being modified;[44] recently, efficient Lennard–Jones PME methods[45,46] and perturbation schemes[44] have been developed that can eliminate the errors associated with this truncation. While the currently employed methodology for alchemical transformations involving a change in system charge reduces artifacts that depend on the simulation box size and periodic boundary conditions, the explicit ions that were included in these simulations may not have sufficiently converged to their equilibrium distributions in these relatively short simulations. Kinases and their inhibitors are known to possess multiple titratable sites with either intrinsic or effective pK_as near physiological pH, while the simulations here treat protonation states and proton tautomers fixed throughout the bound and unbound states; the accuracy of the model can be further improved with the protonation states or tautomers shift upon binding or mutation considered[47,48]. Similarly, some systems display significant salt concentration dependence[49], while the simulations for some systems reported here did not rigorously mimic all aspects of the experimental conditions of the cell viability assays.

While we have shown that predicting the direct impact of mutations on the binding affinity of ATP-competitive TKIs for a single kinase conformation has useful predictive capacity, many additional physical effects that can contribute to cell viability are not currently captured by examining only the predicted change in inhibitor binding affinity. For example, kinase missense mutations can also shift the populations of kinase conformations (which may affect ATP and inhibitor affinities differentially), modulate ATP affinity, modulate affinity for protein substrate, or modulate the ability of the kinase to be regulated or bounded by scaffolding proteins. While many of these effects are in principle tractable by physical modeling in general it is valuable to examine

our mispredictions and outliers to identify whether any of these cases are likely to induce resistance (as observed by ΔpIC_{50} shifts) by one of these alternative mechanisms.

A simple threshold of 10-fold TKI affinity change is a crude metric for classifying resistance or susceptibility due to the myriad biological factors that contribute to the efficacy of a drug in a person. In addition to affecting the binding affinity of inhibitors, missense mutations can also cause drug resistance through other physical mechanisms including induction of splice variants or alleviation of feedback. While the current study only focused on the effect of mutation on drug binding affinity, resistance from these other physical mechanisms could be similarly computed using physical modeling. For example, some mutations are known to activate the kinase by increasing affinity to ATP, which could be computed using free-energy methods like FEP.

In this communication, we tested the hypothesis that FEP+, a fully automated relative-alchemical free-energy workflow, had reached the point where it can accurately and reliably predict how clinically observed mutations in Abl kinase alter the binding affinity of eight FDA-approved TKIs. To establish the potential predictive impact of current-generation alchemical free-energy calculations—which incorporate entropic and enthalpic effects and the discrete nature of aqueous solvation—compared to a simpler physics-based approach that also uses modern forcefields but scores a single minimized conformation, we employed a second physics-based approach (Prime). This simpler physics-based model was able to capture a useful amount of information to achieve substantial predictiveness with an MUE of $1.14^{1.35}_{0.93}$ kcal mol^{-1} ($N = 142$), RMSE of $1.70^{1.98}_{1.40}$ kcal mol^{-1}, respectively ($N = 142$), and classification accuracy of $0.73^{0.80}_{0.66}$ ($N = 144$). Surpassing these good results, we went on to demonstrate that FEP+ is able to achieve superior predictive performance—MUE of $0.79^{0.92}_{0.67}$ kcal mol^{-1} ($N = 142$), RMSE of $1.07^{1.26}_{0.89}$ kcal mol^{-1} ($N = 142$), and classification accuracy of $0.88^{0.93}_{0.82}$ ($N = 144$). While future enhancements to the workflows for Prime and FEP + to account for additional physical and chemical effects are likely to improve predictive performance further, the present results are of sufficient quality and achievable on a sufficiently rapid timescale (with turn-around times ~6 h/calculation) to impact research projects in drug discovery and the life sciences. This work illustrates how the domain of applicability for alchemical free-energy methods is much larger than previously appreciated, and might further be found to include new areas as research progresses: aiding clinical decision-making in the selection of first- or second-line therapeutics guided by knowledge of likely subclonal resistance; identifying other selective kinase inhibitors (or combination therapies) to which the mutant kinase is susceptible; supporting the selection of candidate molecules to advance to clinical trials based on anticipated activity against likely mutations; facilitating the enrollments of patients in mechanism-based basket trials; and generally augmenting the armamentarium of precision oncology.

Methods

System preparation. All system preparation utilized the Maestro Suite (Schrödinger) version 2016-4. Comparative modeling to add missing residues using a homologous template made use of the Splicer tool, while missing loops modeled without a template used Prime. All tools employed default settings unless otherwise noted. The Abl wild-type sequence used in building all Abl kinase domain models utilized the ABL1_HUMAN Isoform IA (P00519-1) UniProt gene sequence spanning S229–K512. Models were prepared in non-phosphorylated form. We used a residue indexing convention that places the Thr gatekeeper residue at position 315 to match common usage; an alternate indexing convention utilized in experimental X-ray structures for Abl:imatinib (PDB: 1OPJ)[50] and Abl:dasatinib (PDB: 4XEY)[51] was adjusted to match our convention.

Complexes with co-crystal structures. Chain B of the experimental structure of Abl:axitinib (PDB: 4WA9)[52] was used, and four missing residues at the N and C termini were added using homology modeling with PDB 3IK3[53] as the template following alignment of the respective termini of the kinase domain. Chain B was selected because chain A was missing an additional 3 and 4 residues at the N and C termini, respectively, in addition to 3- and 20-residue loops, both of which were resolved in chain B. All missing side chains were added with Prime. The co-crystal structure of Abl:bosutinib (PDB: 3UE4)[54] was missing 4 and 10 N- and C-terminal residues, respectively, in chain A that were built using homology modeling with 3IK3 as the template. All loops were resolved in chain A (chain B was missing two residues in the P-loop, Q252 and Y253). All missing side chains were added with Prime. The co-crystal structure of Abl:dasatinib (PDB: 4XEY)[51] was missing 2 and 9 N- and C-terminal residues, respectively, that were built via homology modeling using 3IK3 as the template. A 3 residue loop was absent in chain B but present in chain A; chain A was chosen. The co-crystal structure of Abl:imatinib (PDB: 1OPJ)[50] had no missing loops. Chain B was used because chain A was missing two C-terminal residues that were resolved in chain B. A serine was present at position 336 (index 355 in the PDB file) and was mutated to asparagine using Prime to match the human wild-type reference sequence (P00519-1). The co-crystal structure of Abl:nilotinib (PDB: 3CS9)[55] contained four chains in the asymmetric unit all of which were missing at least one loop. Chain A was selected because its one missing loop involved the fewest number of residues of the four chains; chain A was missing 4 and 12 N- and C-terminal residues, respectively, that were built using homology modeling with 3IK3 as the template. A 4-residue loop was missing in chain A (chain B and C were missing two loops, chain D was missing a five residue loop) that was built using Prime. The co-crystal structure of Abl:ponatinib (PDB: 3OXZ)[56] contained only one chain in the asymmetric unit. It had two missing loops, one 4 residues (built using Prime) and one 12 residues (built using homology modeling with 3OY3[56] as the template). Serine was present at position 336 and was mutated to Asn using Prime to match the human wild-type reference sequence (P00519-1). Once the residue composition of the six Abl:TKI complexes were normalized to have the same sequence, the models were prepared using Protein Preparation Wizard. Bond orders were assigned using the Chemical Components Dictionary and hydrogen atoms were added. Missing side chain atoms were built using Prime. Termini were capped with *N*-acetyl (N terminus) and *N*-methyl amide (C terminus). If present, crystallographic water molecules were retained. Residue protonation states (e.g., Asp381 and Asp421) were determined using PROPKA[57] with a pH range of 5–9. Ligand protonation state was assigned using PROPKA with pH equal to the experimental assay. Hydrogen bonds were assigned by sampling the orientation of crystallographic water, Asn and Gln flips, and His protonation state. The positions of hydrogen atoms were minimized while constraining heavy atoms coordinates. Finally, restrained minimization of all atoms was performed in which a harmonic positional restraint (25.0 kcal mol^{-1} Å^{-2}) was applied only to heavy atoms. Supplementary Table 9 summarizes the composition of the final models used for FEP.

Complexes without co-crystal structures. Co-crystal structures of Abl bound to erlotinib or gefitinib were not publicly available. To generate models of these complexes, Glide-SP[58] was utilized to dock these two compounds into an Abl receptor structure. Co-crystal structures of these two compounds bound to EGFR were publicly available and this information was used to obtain initial ligand geometries and to establish a reference binding mode against which our docking results could be structurally scored. The Abl receptor structure bound to bosutinib was used for docking because its structure was structurally similar to that of EGFR in the erlotinib- (PDB: 4HJO)[59] and gefitinib-bound (PDB: 4WKQ)[60] co-crystal structures. Abl was prepared for docking by using the Protein Preparation Wizard (PPW) with default parameters. Crystallographic waters were removed but their coordinates retained for a subsequent step in which they were optionally reintroduced. Erlotinib and gefitinib protonation states at pH 7 ± 2 were determined using Epik[61]. Docking was performed using the Glide-SP workflow. The receptor grid was centered on bosutinib. The backbone NH of Met318 was chosen to participate in a hydrogen bonding constraint with any hydrogen bond donor on the ligand. The hydroxyl of T315 was allowed to rotate in an otherwise rigid receptor. Ligand docking was performed with enhanced sampling; otherwise default settings were used. Epik state penalties were included in the scoring. The 16 highest ranked (Glide-SP score) poses were retained for subsequent scoring. To determine the docked pose that would be subsequently used for free-energy calculations, the ligand heavy-atom RMSD between the 16 poses and the EGFR co-crystal structures (PDB IDs 4HJO and 4WKQ) was determined. The pose in which erlotinib or gefitinib most structurally resembled the EGFR co-crystal structure (lowest heavy-atom RMSD) was chosen as the pose for subsequent FEP+. Two sets of complex structures were subjected to free-energy calculations to determine the effect of crystal waters: In the first set, without crystallographic waters, the complexes were prepared using Protein Prep Wizard as above. In the second set, the crystallographic waters removed prior to docking were added back, and waters in the binding pocket that clashed with the ligand were removed.

Force field parameter assignment. The OPLS3 forcefield[37] version that shipped with Schrödinger Suite release 2016-4 was used to parameterize the protein and ligand. Torsion parameter coverage was checked for all ligand fragments using

Force Field Builder. The two ligands that contained a fragment with a torsion parameter not covered by OPLS3 were axitinib and bosutinib; Force Field Builder was used to obtain these parameters. SPC parameters[62] were used for water. For mutations that change the net change of the system, counterions were included to neutralize the system with additional Na+ and Cl− ions added to achieve 0.15 M excess to mimic the solution conditions of the experimental assay.

Prime (MM-GBSA). Prime was used to predict the geometry of mutant side chains and to calculate relative changes in free energy using MM-GBSA single-point estimates[36]. VSGB[63] was used as the implicit solvent model to calculate the solvation free energies for the four states (complex/wild-type, complex/mutant, apo protein/wild-type, and apo protein/mutant) and $\Delta\Delta G$ calculated using the thermodynamic cycle depicted in Fig. 1b. Unlike FEP (see below), which simulates the horizontal legs of the thermodynamic cycle, MM-GBSA models the vertical legs by computing the interaction energy between the ligand and protein in both wild-type and mutant states, subtracting these to obtain the $\Delta\Delta G$ of mutation on the binding free energy.

Alchemical free-energy perturbation calculations using FEP+. Alchemical free-energy calculations were performed using the FEP+ tool in the Schrödinger Suite version 2016-4, which offers a fully automated workflow requiring only an input structure (wild-type complex) and specification of the desired mutation. The default protocol was used throughout: It assigns protein and ligand force field parameters (as above), generates a dual-topology[64] alchemical system for transforming wild-type into mutant protein (whose initial structure is modeled using Prime), generates the solvent-leg endpoints (wild-type and mutant apo protein), and constructs intermediate windows spanning wild-type and mutant states. Simulations of the apo protein were setup by removing the ligand from the prepared complex (see System Preparation) followed by an identical simulation protocol as that used for the complex. Charge-conserving mutations utilized 12 λ windows (24 systems) while charge-changing mutations utilized 24 λ windows (48 systems). Each system was solvated in an orthogonal box of explicit solvent (SPC water[62]) with box size determined to ensure that solute atoms were no less than 5 Å (complex leg) or 10 Å (solvent leg) from an edge of the box. For mutations that change the net charge of the system, counterions were included to neutralize the charge of the system, and additional Na+ and Cl− ions added to achieve 0.15 M excess NaCl to mimic the solution conditions of the experimental assay. The artifact in electrostatic interactions for charge change perturbations due to periodic boundary conditions in MD simulations are corrected based on the method proposed by Rocklin et al.[65], where the difference in solvation free energy of the solute under non-periodic boundary condition and that under periodic boundary condition is approximated by Poisson–Boltzmann method and serves as the correction term for each system.

System equilibration was automated. It followed the default 5-stage Desmond protocol: (i) 100 ps with 1 fs time steps of Brownian dynamics with positional restraints of solute heavy atoms to their initial geometry using a restraint force constant of 50 kcal mol^{-1} Å^{-2}; this Brownian dynamics integrator corresponds to a Langevin integrator in the limit when $\tau \to 0$, modified to stabilize equilibration of starting configurations with high potential energies; particle and piston velocities were clipped so that particle displacements were limited to 0.1 Å, in any direction. (ii) 12 ps MD simulations with 1 fs time step using Langevin thermostat at 10 K with constant volume, using the same restraints; (iii) 12 ps MD simulations with 1 fs time step using Langevin thermostat and barostat[66] at 10 K and constant pressure of 1 atmosphere, using the same restraints; (iv) 12 ps MD simulations with 1 fs time step using Langevin thermostat and barostat at 300 K and constant pressure of 1 atmosphere, using the same restraints; (v) a final unrestrained equilibration MD simulation of 240 ps with 2 fs time step using Langevin thermostat and barostat at 300 K and constant pressure of 1 atmosphere. Electrostatic interactions were computed with particle-mesh Ewald (PME)[45] and a 9 Å cutoff distance was used for van der Waals interactions. The production MD simulation was performed in the NPT ensemble using the MTK method[67] with integration time steps of 4, 4, and 8 fs, respectively, for the bonded, near, and far interactions following the RESPA method[68] through hydrogen mass repartitioning[69]. Production FEP+ calculations utilized Hamiltonian replica exchange with solute tempering (REST)[70], with automated definition of the REST region. Dynamics were performed with constant pressure of 1 atmosphere and constant temperature of 300 K for 5 ns in which exchanges between windows was attempted every 1.2 ps.

Because cycle closure could not be used to reduce statistical errors via path redundancy[70], we instead performed mutational free-energy calculations in triplicate by initializing dynamics with different random seeds. The relative free energies for each mutation in each independent run were calculated using BAR[71,72]. The reported $\Delta\Delta G$ was computed as the mean of the computed $\Delta\Delta G$ from three independent simulations. Triplicate simulations were performed in parallel using four NIVIDA Pascal Architecture GPUs per alchemical free-energy simulation (12 GPUs in total), requiring ~6 h in total to compute $\Delta\Delta G$.

Obtaining $\Delta\Delta G$ from ΔpIC_{50} benchmark set data. Reference relative free energies were obtained from three publicly available sources of ΔpIC_{50} data (Table 1). Under the assumption of Michaelis–Menten binding kinetics (pseudo first-order, but relative free energies are likely consistent), the inhibitor is competitive with ATP (eq:ic50). This assumption has been successfully used to estimate relative free energies[34,73–75] using the relationship between IC_{50} and competitive inhibitor affinity K_i,

$$IC_{50} = \frac{K_i}{1 + \frac{[S_0]}{K_M}}. \quad (1)$$

If the Michaelis constant for ATP (K_M) is much larger than the initial ATP concentration S_0, the relation in eq:ic50 will tend towards the equality $IC_{50} = K_i$. The relative change in binding free energy of Abl:TKI binding due to protein mutation is simply,

$$\Delta\Delta G = -RT \ln \frac{IC_{50,WT}}{IC_{50,mut}} \quad (2)$$

where $IC_{50,WT}$ is the IC_{50} value for the TKI binding to the wild-type protein and $IC_{50,mut}$ is the IC_{50} value for the mutant protein. R is the ideal gas constant and T is taken to be room temperature (300 K).

As alluded to above, relating ΔpIC_{50}s to $\Delta\Delta G$s assumes that the Michaelis constant for ATP is much larger than the initial concentration of ATP, and that the experimentally observed ΔpIC_{50} change is solely from changes in kinase:TKI binding affinity. In practice, not all of these assumptions may hold. For example, the experimentally observed ΔpIC_{50} might depend on the metabolism of drugs, and for drugs with different mechanisms of action than directly binding to the kinase binding pocket (e.g., binding to the transition structures of kinases, target gene amplification, up/downregulation of positive-/negative-feedback effectors, diminished synergism of pro-apoptotic machinery, decoupling of the target from cell survival circuits)[76,77], their inhibition ability might not correlate well with binding affinity. However, the comparison between ΔpIC_{50} and ΔK_D is presented in Fig. 2d, and this comparison indicates the assumptions we used to relate ΔpIC_{50} to $\Delta\Delta G$ are reasonable for the dataset we studied.

Quantitative accuracy metrics. MUE was calculated by taking the average absolute difference between predicted and experimental estimates of $\Delta\Delta G$. RMSE was calculated by taking the square root of the average squared difference between predicted and experimental estimates of $\Delta\Delta G$. MUE depends linearly on errors such that large and small errors contribute equally to the average value, while RMSE depends quadratically on errors, magnifying their effect on the average value.

Truth tables. Two-class truth tables were constructed to characterize the ability of Prime and FEP+ to correctly classify mutations as susceptible ($\Delta\Delta G \le 1.36$ kcal mol^{-1}) or resistant ($\Delta\Delta G > 1.36$ kcal mol^{-1}), where the 1.36 kcal mol^{-1} threshold represents a 10-fold change in affinity. Accuracy was calculated as the fraction of all predictions that were correctly classified as sensitizing, neutral, or resistant. Sensitivity and specificity were calculated using a binary classification of resistant ($\Delta\Delta G > 1.36$ kcal mol^{-1}) or susceptible ($\Delta\Delta G \le 1.36$ kcal mol^{-1}). Specificity was calculated as the fraction of correctly predicted non-resistant mutations out of all truly susceptible mutations **S**. Sensitivity was calculated as the fraction of correctly predicted resistant mutations out of all truly resistant mutations, **R**. The number of susceptible mutations was 113 for axitinib, bosutinib, dasatinib, imatinib, nilotinib and ponatinib, and 12 for erlotinib and gefitinib; the number of resistant mutations **R** was 18 for axitinib, bosutinib, dasatinib, imatinib, nilotinib, and ponatinib, and 1 for erlotinib and gefitinib.

Consensus model. First, Prime and FEP+ ($n = 142$) were scaled by minimizing their RMSE to experiment by optimizing slope using linear regression. The resulting (minimum) RMSE was used in a subsequent step to combine the scaled FEP+ and scaled Prime free energies with inverse-variance weighted averaging.

ROC. A ROC curve was generated by computing the true positive rate (sensitivity) and the true negative rate (specificity) when the classification cutoff differentiating resistant from sensitizing mutations is changed for (only) the predicted values of $\Delta\Delta G$. Cutoffs were chosen by taking the minimum and maximum value of $\Delta\Delta G$ for a dataset (Prime or FEP+), and iteratively computing specificity and sensitivity in steps of 0.001 kcal mol^{-1}, which by this definition will be in the range [0,1]. Experimental positives and negatives were classified with the 1.36 kcal mol^{-1} cutoff. ROC-AUC was computed using the trapezoidal rule.

Estimating uncertainties of physical-modeling results. 95% symmetric CI (95%) for all performance metrics were calculated using bootstrap by resampling all datasets with replacement, with 1000 resampling events. Confidence intervals were estimated for all performance metrics and reported as $x^{x_{high}}_{x_{low}}$ where x is the mean statistic calculated from the complete dataset (e.g., RMSE), and x_{low} and x_{high} are the values of the statistic at the 2.5th and 97.5th percentiles of the value-sorted list of the bootstrap samples. Uncertainty for $\Delta\Delta G$s was computed by the standard

deviation between three independent runs (using different random seeds to set initial velocities), where the 95% CI was $[\Delta\Delta G - 1.96 \times \sigma_{\mathrm{FEP+}}, \Delta\Delta G + 1.96 \times \sigma_{\mathrm{FEP+}}]$ kcal mol^{-1}. 1σ used in plots for FEP+ and experiment; 0σ for Prime.

Bayesian hierarchical model to estimate intrinsic error. We used Bayesian inference to estimate the true underlying prediction error of Prime and FEP+ by making use of known properties of the experimental variability (characterized in Fig. 2) and statistical uncertainty estimates generated by our calculations under weak assumptions about the character of the error.

We presume the true free-energy differences of mutation i, $\Delta\Delta G_i^{\mathrm{true}}$, comes from a normal background distribution of unknown mean and variance,

$$\Delta\Delta G_i^{\mathrm{true}} \sim \left(\mu_{\mathrm{mut}}, \sigma_{\mathrm{mut}}^2\right) i = 1, \ldots, M \tag{3}$$

where there are M mutations in our dataset. We assign weak priors to the mean and variance

$$\mu_{\mathrm{mut}} \sim U(-6, +6) \tag{4}$$

$$\sigma_{\mathrm{mut}} \propto 1 \tag{5}$$

where we limit $\sigma > 0$.

We presume the true computational predictions (absent statistical error) differ from the (unknown) true free-energy difference of mutation $\Delta\Delta G_i^{\mathrm{true}}$ by normally distributed errors with zero bias but standard deviation equal to the RMSE for either Prime or FEP+, the quantity we are focused on estimating:

$$\Delta\Delta G_{i,\mathrm{Prime}}^{\mathrm{true}} \sim \left(\Delta\Delta G_i^{\mathrm{true}}, \mathrm{RMSE}_{\mathrm{Prime}}^2\right) \tag{6}$$

$$\Delta\Delta G_{i,\mathrm{FEP+}}^{\mathrm{true}} \sim \left(\Delta\Delta G_i^{\mathrm{true}}, \mathrm{RMSE}_{\mathrm{FEP+}}^2\right) \tag{7}$$

In the case of Prime, since the computation is deterministic, we actually calculate $\Delta\Delta G_{\mathrm{Prime}}^{\mathrm{true}}$ for each mutant. For FEP+, however, the computed free-energy changes are corrupted by statistical error, which we also presume to be normally distributed with standard deviation $\sigma_{\mathrm{calc},i}$,

$$\Delta\Delta G_{i,\mathrm{FEP+}} \sim \left(\Delta\Delta G_{i,\mathrm{FEP+}}, \sigma_{i,\mathrm{FEP+}}^2\right) \tag{8}$$

where $\Delta\Delta G_{i,\mathrm{FEP+}}$ is the free energy computed for mutant i by FEP+, and $\sigma_{i,\mathrm{FEP+}}$ is the corresponding statistical error estimate.

The experimental data we observe is also corrupted by error, which we presume to be normally distributed with standard deviation σ_{exp}:

$$\Delta\Delta G_{i,\mathrm{exp}} \sim \left(\Delta\Delta G_i, \sigma_{\mathrm{exp}}^2\right) \tag{9}$$

Here, we used an estimate of $K_{\mathrm{d}}-$ and IC_{50}-derived $\Delta\Delta G$ variation derived from the empirical RMSE of 0.81 kcal mol^{-1}, where we took $\sigma_{\mathrm{exp}} \approx 0.81/\sqrt{2} = 0.57$ kcal mol^{-1} to ensure the difference between two random measurements of the same mutant would have an empirical RMSE of 0.81 kcal mol^{-1}.

Under the assumption that the true $\Delta\Delta G$ is normally distributed and the calculated value differs from the true value via a normal error model, it can easily be shown that the MUE is related to the RMSE via

$$\mathrm{MUE} = \int \mathrm{d}x_{\mathrm{true}}\, p(x_{\mathrm{true}}) \int \mathrm{d}x_{\mathrm{calc}}\, p(x_{\mathrm{calc}}|x_{\mathrm{true}})|x_{\mathrm{calc}} - x_{\mathrm{true}}| \tag{10}$$

$$= \int \mathrm{d}x_{\mathrm{true}} \frac{1}{\sqrt{2\pi\sigma_{\mathrm{true}}^2}} e^{-\frac{(x_{\mathrm{true}} - \mu_{\mathrm{true}})^2}{2\sigma_{\mathrm{true}}^2}} \int \mathrm{d}x_{\mathrm{calc}} \frac{1}{\sqrt{2\pi\sigma_{\mathrm{calc}}^2}} e^{-\frac{(x_{\mathrm{calc}} - \mu_{\mathrm{true}})^2}{2\sigma_{\mathrm{calc}}^2}} |x_{\mathrm{calc}} - x_{\mathrm{true}}| \tag{11}$$

$$= \sqrt{\frac{2}{\pi}}\mathrm{RMSE} \tag{12}$$

The model was implemented using PyMC3[78], observable quantities were set to their computed or experimental values, and 5000 samples drawn from the posterior (after discarding an initial 500 samples to burn-in) using the default NUTS sampler. Expectations and posterior predictive intervals were computed from the marginal distributions obtained from the resulting traces.

Code availability. Scripts used for statistics analysis (including the Bayesian inference model) can be found at the following URL: https://github.com/kehauser/Predicting-resistance-of-clinical-Abl-mutations-to-targeted-kinase-inhibitors-using-FEP.

Data availability. All relevant data are publicly available: compiled experimental datasets, input files for Prime and FEP+, and computational results that support our findings can be found at GitHub by following the URL: https://github.com/kehauser/Predicting-resistance-of-clinical-Abl-mutations-to-targeted-kinase-inhibitors-using-FEP.

Received: 8 January 2018 Accepted: 15 May 2018
Published online: 13 June 2018

References

1. Roskoski, R. Jr. USFDA approved protein kinase inhibitors. http://www.brimr.org/PKI/PKIs.htm (2017).
2. Santos, R. et al. A comprehensive map of molecular drug targets. *Nat. Rev. Drug Discov.* **16**, 19–34 (2016).
3. Shah, N. P. et al. Multiple BCR-ABL kinase domain mutations confer polyclonal resistance to the tyrosine kinase inhibitor imatinib (STI571) in chronic phase and blast crisis chronic myeloid leukemia. *Cancer Cell* **2**, 117–125 (2002).
4. Buczek, M., Escudier, B., Bartnik, E., Szczylik, C. & Czarnecka, A. Resistance to tyrosine kinase inhibitors in clear cell renal cell carcinoma: From the patient's bed to molecular mechanisms. *Biochim. Biophys. Acta Rev. Cancer* **1845**, 31–41 (2014).
5. Huang, L. & Fu, L. Mechanisms of resistance to EGFR tyrosine kinase inhibitors. *Acta Pharm. Sin. B* **5**, 390–401 (2015).
6. Meyer, S. C. & Levine, R. L. Molecular pathways: molecular basis for sensitivity and resistance to JAK kinase inhibitors. *Clin. Cancer Res.* **20**, 2051–2059 (2014).
7. Davare, M. A. et al. Structural insight into selectivity and resistance profiles of ROS1 tyrosine kinase inhibitors. *Proc. Natl Acad. Sci. USA* **112**, E5381–E5390 (2015).
8. Van Allen, E. M. et al. The genetic landscape of clinical resistance to RAF inhibition in metastatic melanoma. *Cancer Discov.* **4**, 94–109 (2014).
9. Rani, S. et al. Neuromedin U: a candidate biomarker and therapeutic target to predict and overcome resistance to HER-tyrosine kinase inhibitors. *Cancer Res.* **74**, 3821–3833 (2014).
10. Holohan, C., Van Schaeybroeck, S., Longley, D. B. & Johnston, P. G. Cancer drug resistance: an evolving paradigm. *Nat. Rev. Cancer* **13**, 714–726 (2013).
11. Weisberg, E., Manley, P. W., Cowan-Jacob, S. W., Hochhaus, A. & Griffin, J. D. Second generation inhibitors of BCR-ABL for the treatment of imatinib-resistant chronic myeloid leukaemia. *Nat. Rev. Cancer* **7**, 345–356 (2007).
12. Lu, Y. X., Cai, Q. & Ding, K. Recent developments in the third generation inhibitors of Bcr-Abl for overriding T315I mutation. *Curr. Med. Chem.* **18**, 2146–2157 (2011).
13. Juchum, M., Günther, M. & Laufer, S. A. Fighting cancer drug resistance: opportunities and challenges for mutation-specific EGFR inhibitors. *Drug Resist. Updat.* **20**, 12–28 (2015).
14. Song, Z., Wang, M. & Zhang, A. Alectinib: a novel second generation anaplastic lymphoma kinase (ALK) inhibitor for overcoming clinically-acquired resistance. *Acta Pharm. Sin. B* **5**, 34–37 (2015).
15. Neel, D. S. & Bivona, T. G. Resistance is futile: overcoming resistance to targeted therapies in lung adenocarcinoma. *Npj Precis. Oncol.* **1**, 1–6 (2017).
16. Gruber, F., Hjorth-Hansen, H., Mikkola, I., Stenke, L. & TA, J. A novel BCR-ABL splice isoform is associated with the L248V mutation in CML patients with acquired resistance to imatinib. *Leukemia* **20**, 2057–60 (2006).
17. Chandarlapaty, S. et al. AKT inhibition relieves feedback suppression of receptor tyrosine kinase expression and activity. *Cancer Cell* **19**, 58–71 (2011).
18. Knight, Z. A., Lin, H. & Shokat, K. M. Targeting the cancer kinome through polypharmacology. *Nat. Rev. Cancer* **10**, 130 (2010).
19. Housman, G. et al. Drug resistance in cancer: an overview. *Cancers* **6**, 1769–1792 (2014).
20. Zehir, A. et al. Mutational landscape of metastatic cancer revealed from prospective clinical sequencing of 10,000 patients. *Nat. Med.* **23**, 703–713 (2017).
21. Redig, A. J. & Jänne, P. A. Basket trials and the evolution of clinical trial design in an era of genomic medicine. *J. Clin. Oncol.* 33, 975–977 (2015).
22. Hyman, D. M., Taylor, B. S. & Baselga, J. Implementing genome-driven oncology. *Cell* **168**, 584–599 (2017).
23. Pesesky, M. W. et al. Evaluation of machine learning and rules-based approaches for predicting antimicrobial resistance profiles in gram-negative bacilli from whole genome sequence data. *Front. Microbiol.* **7**, 1887 (2016).
24. Melnikov, A., Rogov, P., Wang, L., Gnirke, A. & Mikkelsen, T. S. Comprehensive mutational scanning of a kinase in vivo reveals substrate-dependent fitness landscapes. *Nucleic Acids Res.* **42**, e112–e112 (2014).
25. Chodera, J. D. et al. Alchemical free energy methods for drug discovery: progress and challenges. *Curr. Opin. Struct. Biol.* **21**, 150–160 (2011).

26. Wang, L. et al. 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.* **137**, 2695–2703 (2015).
27. Abel, R. et al. Accelerating drug discovery through tight integration of expert molecular design and predictive scoring. *Curr. Opin. Struct. Biol.* **43**, 38–44 (2017).
28. Aldeghi, M., Heifetz, A., Bodkin, M. J., Knapp, S. & Biggin, P. C. Accurate calculation of the absolute free energy of binding for drug molecules. *Chem. Sci.* **7**, 207–218 (2016).
29. Cappel, D. et al. Relative binding free energy calculations applied to protein homology models. *J. Chem. Inf. Model.* **56**, 2388–2400 (2016).
30. Clark, A. J. et al. Free energy perturbation calculation of relative binding free energy between broadly neutralizing antibodies and the gp120 glycoprotein of hiv-1. *J. Mol. Biol.* **429**, 930–947 (2017).
31. Steinbrecher, T. et al. Predicting the effect of amino acid single-point mutations on protein stabilitylarge-scale validation of md-based relative free energy calculations. *J. Mol. Biol.* **429**, 948–963 (2017).
32. Ford, M. C. & Babaoglu, K. Examining the feasibility of using free energy perturbation (FEP+) in predicting protein stability. *J. Chem. Inf. Model.* **57**, 1276–1285 (2017).
33. Zou, J., Song, B., Simmerling, C. & Raleigh, D. Experimental and computational analysis of protein stabilization by Gly-to- d -Ala substitution: a convolution of native state and unfolded state effects. *J. Am. Chem. Soc.* **138**, 15682–15689 (2016).
34. Mondal, J., Tiwary, P. & Berne, B. J. How a kinase inhibitor withstands gatekeeper residue mutations. *J. Am. Chem. Soc.* **138**, 4608–4615 (2016).
35. Lovering, F. et al. Imidazotriazines: spleen tyrosine kinase (Syk) inhibitors identified by free-energy perturbation (FEP). *ChemMedChem* **11**, 217–233 (2016).
36. Rapp, C., Kalyanaraman, C., Schiffmiller, A., Schoenbrun, E. L. & Jacobson, M. P. A molecular mechanics approach to modeling protein–ligand interactions: relative binding affinities in congeneric series. *J. Chem. Inf. Model.* **51**, 2082–2089 (2011).
37. Harder, E. et al. OPLS3: a force field providing broad coverage of drug-like small molecules and proteins. *J. Chem. Theory Comput.* **12**, 281–296 (2016).
38. Huang, J. & MacKerell, A. D. CHARMM36 all-atom additive protein force field: validation based on comparison to NMR data. *J. Comput. Chem.* **34**, 2135–2145 (2013).
39. Maier, J. A. et al. ff14SB: improving the accuracy of protein side chain and backbone parameters from ff99SB. *J. Chem. Theory Comput.* **11**, 3696–3713 (2015).
40. O'Hare, T. Combined Abl inhibitor therapy for minimizing drug resistance in chronic myeloid leukemia: Src/Abl inhibitors are compatible with imatinib. *Clin. Cancer Res.* **11**, 6987–6993 (2005).
41. Shan, Y. et al. A conserved protonation-dependent switch controls drug binding in the Abl kinase. *Proc. Natl. Acad. Sci.* **106**, 139–144 (2009).
42. Demerdash, O., Yap, E.-H. & Head-Gordon, T. Advanced potential energy surfaces for condensed phase simulation. *Annu. Rev. Phys. Chem.* **65**, 149–174 (2014).
43. Jiao, D., Golubkov, P. A., Darden, T. A. & Ren, P. Calculation of protein–ligand binding free energy by using a polarizable potential. *Proc. Natl. Acad. Sci. USA* **105**, 6290–6295 (2008).
44. Shirts, M. R., Mobley, D. L., Chodera, J. D. & Pande, V. S. Accurate and efficient corrections for missing dispersion interactions in molecular simulations. *J. Phys. Chem. B* **111**, 13052–13063 (2007).
45. Essmann, U. et al. A smooth particle mesh Ewald method. *J. Chem. Phys.* **103**, 8577–8593 (1995).
46. Wennberg, C. L., Murtola, T., Hess, B. & Lindahl, E. Lennard–Jones lattice summation in bilayer simulations has critical effects on surface tension and lipid properties. *J. Chem. Theory Comput.* **9**, 3527–3537 (2013).
47. Onufriev, A. V. & Alexov, E. Protonation and pK changes in protein–ligand binding. *Q. Rev. Biophys.* **46**, 181–209 (2013).
48. Martin, Y. C. Let's not forget tautomers. *J. Comput. Aided Mol. Des.* **23**, 693–704 (2009).
49. Jensen, J. Calculating pH and salt dependence of protein-protein binding. *Curr. Pharm. Biotechnol.* **9**, 96–102 (2008).
50. Nagar, B. et al. Structural basis for the autoinhibition of c-Abl tyrosine kinase. *Cell* **112**, 859–871 (2003).
51. Lorenz, S., Deng, P., Hantschel, O., Superti-Furga, G. & Kuriyan, J. Crystal structure of an SH2–kinase construct of c-Abl and effect of the SH2 domain on kinase activity. *Biochem. J.* **468**, 283–291 (2015).
52. Pemovska, T. et al. Axitinib effectively inhibits BCR-ABL1(T315I) with a distinct binding conformation. *Nature* **519**, 102–105 (2015).
53. O'Hare, T. et al. AP24534, a Pan-BCR-ABL inhibitor for chronic myeloid leukemia, potently inhibits the T315I mutant and overcomes mutation-based resistance. *Cancer Cell* **16**, 401–412 (2009).
54. Levinson, N. M. & Boxer, S. G. Structural and spectroscopic analysis of the kinase inhibitor bosutinib and an isomer of bosutinib binding to the Abl tyrosine kinase domain. *PLoS ONE* **7**, e29828 (2012).
55. Weisberg, E. et al. Characterization of AMN107, a selective inhibitor of native and mutant Bcr-Abl. *Cancer Cell* **7**, 129–141 (2005).
56. Zhou, T. et al. Structural mechanism of the Pan-BCR-ABL inhibitor ponatinib (AP24534): lessons for overcoming kinase inhibitor resistance: structural mechanism of ponatinib. *Chem. Biol. Drug Des.* **77**, 1–11 (2011).
57. Li, H., Robertson, A. D. & Jensen, J. H. Very fast empirical prediction and rationalization of protein pKa values. *Proteins Struct. Funct. Bioinformatics* **61**, 704–721 (2005).
58. Friesner, R. A. et al. Glide: a new approach for rapid, accurate docking and scoring. 1. Method and assessment of docking accuracy. *J. Med. Chem.* **47**, 1739–1749 (2004).
59. Park, J. H., Liu, Y., Lemmon, M. A. & Radhakrishnan, R. Erlotinib binds both inactive and active conformations of the EGFR tyrosine kinase domain. *Biochem. J.* **448**, 417–423 (2012).
60. Yosaatmadja, Y. & Squire, C. 1.85 angstrom structure of EGFR kinase domain with gefitinib. https://doi.org/10.2210/pdb4WKQ/pdb (2014)
61. Shelley, J. C. et al. Epik: A software program for pK a prediction and protonation state generation for drug-like molecules. *J. Comput. Aided Mol. Des.* **21**, 681–691 (2007).
62. Berendsen, H. J. C., Postma, J. P. M., van Gunsteren, W. F. & Hermans, J. in *Intermolecular Forces* (ed. Pullman, B.) 14, 331–342. (Springer, Dordrecht, 1981).
63. Shivakumar, D. et al. Prediction of absolute solvation free energies using molecular dynamics free energy perturbation and the OPLS force field. *J. Chem. Theory Comput.* **6**, 1509–1519 (2010).
64. Pearlman, D. A. A comparison of alternative approaches to free energy calculations. *J. Phys. Chem.* **98**, 1487–1493 (1994).
65. Rocklin, G. J., Mobley, D. L., Dill, K. A. & Hünenberger, P. H. Calculating the binding free energies of charged species based on explicit-solvent simulations employing lattice-sum methods: an accurate correction scheme for electrostatic finite-size effects. *J. Chem. Phys.* **139**, 184103 (2013).
66. Feller, S. E., Zhang, Y., Pastor, R. W. & Brooks, B. R. Constant pressure molecular dynamics simulation: The langevin piston method. *J. Chem. Phys.* **103**, 4613–4621 (1995).
67. Martyna, G. J., Tobias, D. J. & Klein, M. L. Constant pressure molecular dynamics algorithms. *J. Chem. Phys.* **101**, 4177–4189 (1994).
68. Tuckerman, M., Berne, B. J. & Martyna, G. J. Reversible multiple time scale molecular dynamics. *J. Chem. Phys.* **97**, 1990–2001 (1992).
69. Hopkins, C. W., Le Grand, S., Walker, R. C. & Roitberg, A. E. Long-time-step molecular dynamics through hydrogen mass repartitioning. *J. Chem. Theory Comput.* **11**, 1864–1874 (2015).
70. Wang, L., Berne, B. J. & Friesner, R. A. On achieving high accuracy and reliability in the calculation of relative protein-ligand binding affinities. *Proc. Natl Acad. Sci. USA* **109**, 1937–1942 (2012).
71. Bennett, C. H. Efficient estimation of free energy differences from Monte Carlo data. *J. Comput. Phys.* **22**, 245–268 (1976).
72. Shirts, M. R., Bair, E., Hooker, G. & Pande, V. S. Equilibrium free energies from nonequilibrium measurements using maximum-likelihood methods. *Phys. Rev. Lett.* **91**, 140601 (2003).
73. Price, D. J. & Jorgensen, W. L. Computational binding studies of human pp60c-src SH2 domain with a series of nonpeptide, phosphophenyl-containing ligands. *Bioorg. Med. Chem. Lett.* **10**, 2067–2070 (2000).
74. Luccarelli, J., Michel, J., Tirado-Rives, J. & Jorgensen, W. L. Effects of water placement on predictions of binding affinities for p38α MAP kinase inhibitors. *J. Chem. Theory Comput.* **6**, 3850–3856 (2010).
75. Michel, J., Verdonk, M. L. & Essex, J. W. Protein-ligand binding affinity predictions by implicit solvent simulations: a tool for lead optimization? *J. Med. Chem.* **49**, 7427–7439 (2006).
76. Barouch-Bentov, R. & Sauer, K. Mechanisms of drug resistance in kinases. *Expert Opin. Invest. Drugs* **20**, 153–208 (2011).
77. McDermott, U. et al. Identification of genotype-correlated sensitivity to selective kinase inhibitors by using high-throughput tumor cell line profiling. *Proc. Natl. Acad. Sci. USA* **104**, 19936–19941 (2007).
78. Salvatier, J., Wiecki, T. V. & Fonnesbeck, C. Probabilistic programming in python using pymc3. *PeerJ Comput. Sci.* **2**, e55 (2016).
79. Schrock, A., Chen, T.-H., Clackson, T. & Rivera, V. M. Comprehensive analysis of the in vitro potency of ponatinib, and all other approved BCR-ABL tyrosine kinase inhibitors (TKIs), against a panel of single and compound BCR-ABL mutants. *Blood* **122**, 3992 (2013).
80. O'Hare, T., Eide, C. A. & Deininger, M. W. Bcr-abl kinase domain mutations, drug resistance, and the road to a cure for chronic myeloid leukemia. *Blood* **110**, 2242–2249 (2007).
81. Soverini, S. et al. Contribution of abl kinase domain mutations to imatinib resistance in different subsets of philadelphia-positive patients: By the gimema

working party on chronic myeloid leukemia. *Clin. Cancer Res.* **12**, 7374–7379 (2006).
82. Davis, M. I. et al. Comprehensive analysis of kinase inhibitor selectivity. *Nat. Biotechnol.* **29**, 1046–1051 (2011).

Acknowledgements

We thank Daniel Robinson (Schrödinger), Sonya M. Hanson (MSKCC), and Gregory A. Ross (MSKCC) for helpful discussions. J.D.C. acknowledges support from NIH National Cancer Institute Cancer Center Core Grant P30 CA008748; J.D.C. and S.K.A. acknowledge support from the Sloan Kettering Institute, Cycle for Survival, and NIH grant R01 GM121505. K.H. acknowledges help from Wei Chen (Schrödinger) and Anthony Clark (Schrödinger) for instructions on running mutations changing the net charge of the system, and Simon Gao (Schrödinger) for assistance in computational resources.

Author contributions

K.H., J.D.C., C.N., R.A., and L.W. designed the research; K.H., S.A., T.S., and L.W. identified experimental datasets; K.H. and L.W. performed the simulations; K.H., C.N., S.K.A., S.R., T.S., R.A., J.D.C., and L.W. analyzed the data; K.H., J.D.C., S.K.A., and L.W. wrote the paper.

Additional information

Supplementary information accompanies this paper at https://doi.org/10.1038/s42003-018-0075-x.

Competing interests: J.D.C. is a member of the Scientific Advisory Board for Schrödinger Inc. S.R. is a former employee of Schrödinger Inc.; and K.H., C.N., R.A., T.S., and L.W. are employees of Schrödinger Inc.