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*]

[1]Schrödinger, New York, NY 10036; [2]Gerstner Sloan Kettering Graduate School, Memorial Sloan Kettering Cancer Center, New York, NY 10065; [3]Computational and Systems Biology Program, Sloan Kettering Institute, Memorial Sloan Kettering Cancer Center, New York, NY 10065; [4]Schrödinger, GmbH, Q7 23, 68161 Mannheim, Germany

***For correspondence:** lingle.wang@schrodinger.com (LW)

Abstract 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. This complicates the practice of precision medicine, pairing of patients with clinical trials, and development of next-generation inhibitors. Here, 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). We find these calculations can achieve useful accuracy in predicting resistance for a set of eight FDA-approved kinase inhibitors across 144 clinically-identified point mutations, achieving a root mean square error in binding free energy changes of $1.1^{1.3}_{0.9}$ kcal/mol (95% confidence interval) and correctly classifying mutations as resistant or susceptible with 88^{93}_{82}% accuracy. Since these calculations are fast on modern GPUs, this benchmark establishes the potential for physical modeling to collaboratively support the rapid assessment and anticipation of the potential for patient mutations to affect drug potency in clinical applications.

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

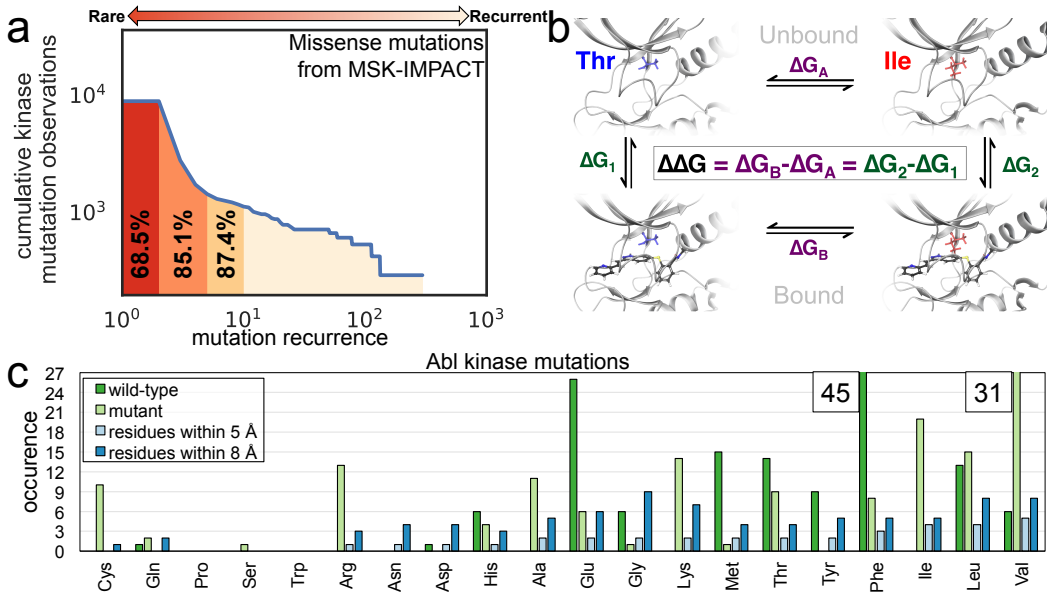


Figure 1. Relative alchemical free-energy calculations can be used to predict affinity changes of FDA-approved selective kinase inhibitors arising from clinically-identified mutations in their targets of therapy. *(a)* Missense mutation statistics derived from 10,336 patient samples subjected to MSK-IMPACT 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. *(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 number of wild-type Phe residues (n=45) and mutant Val residues (n=31) exceeded the limits of the y-axis.

basket trials [21, 22], help prioritize candidate compounds for clinical trials, and aid the development of next-generation inhibitors.

The long tail of rare kinase mutations frustrates prediction of drug resistance

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 (***Figure 1***a), 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.

Alchemical free-energy methods can predict inhibitor binding affinities

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. Modern atomistic molecular mechanics forcefields such as OPLS3 [25], CHARMM [26], and AMBER FF14SB [27] have reached a sufficient level of maturity to enable the accurate and reliable prediction of receptor-ligand binding free energy. Alchemical free-energy methods permit receptor-ligand binding energies to be computed

rigorously, including all relevant entropic and enthalpic contributions [28]. Encouragingly, kinase:inhibitor binding affinities have been predicted using alchemical free-energy methods with mean unsigned errors of 1.0 kcal/mol for CDK2, JNK1, p38, and Tyk2 [29, 30]. Beyond kinases, alchemical approaches have predicted the binding affinity of BRD4 inhibitors with mean absolute errors of 0.6 kcal/mol [31]. Alchemical methods have also been observed to have good accuracy (0.6 kcal/mol mean unsigned error for Tyk2 tyrosine kinase) in the prediction of relative free energies for ligand transformations within a complex whose receptor geometry was generated using a homology model [32].

Alchemical approaches can predict the impact of protein mutations on free energy

Alchemical free-energy calculations have also been used to predict the impact of mutations on protein-protein binding [33] and protein thermostabilities [34]. Recent work has found that protein mutations can be predicted to be stabilizing or destabilizing with a classification accuracy of 71% across ten proteins and 62 mutations [35]. The impact of Gly to D-Ala mutations on protein stability was predicted using an alchemical approach with a similar level of accuracy [36]. 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 [37].

Assessing the potential for physical modeling to predict resistance to FDA-approved TKIs

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+ [29], recently demonstrated to achieve sufficiently good accuracy to drive the design of small-molecule inhibitors for a broad range of targets during lead optimization [28–30, 38]. We compare this approach against a fast but approximate physical modeling method implemented in Prime [39] (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) 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 (x^{upper}_{lower})), an RMSE of $1.07^{1.26}_{0.89}$ kcal/mol, and an MUE of $0.79^{0.92}_{0.67}$ kcal/mol was achieved.

Results

Free energy calculations can recapitulate the impact of clinical mutations on TKI affinity

Alchemical free-energy calculations utilize a physics-based approach to estimate the free energy of transforming one chemical species into another, incorporating all enthalpic and entropic contributions in a physically consistent manner [28, 40–42]. While relative alchemical free-energy calculations have typically been employed in optimizing small molecules for increased potency or selectivity [29, 38, 42, 43], a complementary alchemical approach can be used to compute the impact of point mutations on ligand binding affinities. ***Figure 1***b 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. In the *bound* leg of the cycle, the wild-type protein:ligand complex is transformed into the mutant protein:ligand complex. In the *unbound* leg of the cycle, the *apo* protein is transformed from wild-type into mutant. To achieve reliable predictions with short relative free-energy calculations, a reliable receptor:ligand complex structure is required with the assumption that the binding mode of wild-type and mutant are similar. In this work, high-resolution co-crystal structures of wild-type Abl bound to an inhibitor were utilized when available. To assess the potential for using docked inhibitor poses, we also examined two systems for

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

					(kcal/mol)		(kcal/mol)
TKI	**N_{mut}**	**R**	**S**	**PDB**	$\lvert\Delta G_{max} - \Delta G_{min}\rvert$	**Source**	ΔG_{WT}
axitinib	26	0	26	4wa9	2.05	[44]	−8.35
bosutinib	21	4	17	3ue4	2.79	[45]	−9.81
dasatinib	21	5	16	4xey	5.08	[45]	−11.94
imatinib	21	5	16	1opj	2.16	[45]	−9.19
nilotinib	21	4	17	3cs9	3.88	[45]	−10.74
ponatinib	21	0	21	3oxz	1.00	[45]	−11.70
subtotal	**131**	**18**	**113**				
erlotinib	7	1	6	*Dock to 3ue4*	1.73	[46]	−9.77
gefitinib	6	0	6	*Dock to 3ue4*	1.79	[46]	−8.84
total	**144**	**19**	**125**				

N_{mut}: Total number of mutants for which ΔpIC_{50} data was available.
Number of **R**esistant, **S**usceptical 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.

which co-crystal structures were not available (Abl:erlotinib and Abl:gefitinib) and used docking to generate initial coordinates.

Compiled ΔpIC_{50} data provides a benchmark 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***, taking care to ensure 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 (***Table S1***). 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 IMPACT study analyzed in ***Figure 1***a, 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 ΔΔG 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 (***Figure 2***a,b,c). Cross-comparison of 97 experimentally measured ΔΔGs derived from cell viability assay IC_{50} data led to an estimate of experimental variability of $0.32^{0.36}_{0.28}$ kcal/mol 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—the focus of this study—we also compared ΔΔGs derived from ΔpIC_{50}s with those derived from binding affinity measurements (ΔK_d) for which data for a limited set of 27 mutations was available (***Figure 2***d); the larger computed RMSE of $0.81^{1.04}_{0.59}$ kcal/mol represents an estimate of the lower bound of the RMSE to the IC_{50}-derived ΔΔGs that we might hope to achieve with FEP+ or Prime, which were performed using

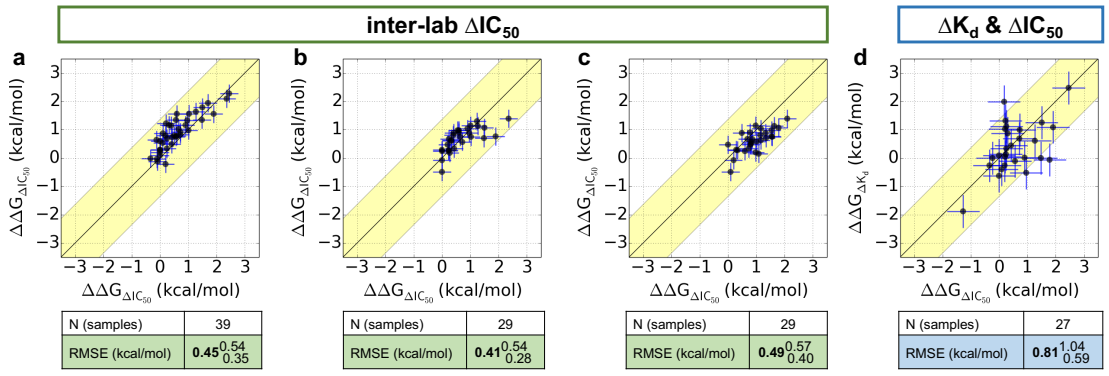


Figure 2. Cross-comparison of the experimentally measured effects that mutations in Abl kinase have on ligand binding, performed by different labs. ΔΔG was computed from publicly available ΔpIC_{50} or ΔpK_d measurements and these values of ΔΔG were then plotted and the RMSE between them reported. (**a**) ΔpIC_{50} measurements (X-axis) from [45] compared with ΔpIC_{50} measurements (Y-axis) from [47]. (**b**) ΔpIC_{50} measurements (X-axis) from [45] compared with ΔpIC_{50} measurements (Y-axis) from [48]. (**c**) ΔpIC_{50} measurements (X-axis) from [47] compared with ΔpIC_{50} measurements (Y-axis) from [48]. (**d**) ΔpIC_{50} measurements (X-axis) from [45] compared with ΔpK_d measurements (Y-axis) from [46] using non-phosphorylated Abl kinase. Scatter plot error bars in a,b,and c are ±standard error (SE) taken from the combined 97 inter-lab ΔΔGs 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. Scatter plot error bars in d are the ±standard error (SE) of ΔΔGs derived from ΔpIC_{50} and ΔpK_d from a set of 27 mutations, which is $0.58^{0.74}_{0.42}$; the RMSE was $0.81^{1.04}_{0.59}$ kcal/mol.

non-phosphorylated models, when comparing sample statistics directly. In comparing 31 mutations for which phosphorylated and non-phosphorylated ΔK_ds were available, we found a strong correlation between the ΔΔGs derived from those data (r=0.94, Supplementary ***Figure S1***); the statistics of that comparison are similar to those of the inter-lab variability comparison.

Most clinical 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 ***Table S2—Table S7*** in the Supplementary Information. Two mutations exceeded the dynamic range of the assays (IC_{50} >10,000 nM); as these two mutations clearly raise resistance, we excluded them from quantitative analysis (RMSE and MUE) but included them in truth table analyses and classification metrics (accuracy, specificity, sensitivity).

How accurately does physical modeling predict affinity changes for clinical Abl mutants?

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 [29], 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 (ΔΔG) 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 (ΔΔG) 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, which was the standard error computed from the cross-comparison in ***Figure 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 (see Methods).

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 $\sigma_{\mathrm{FF}} \approx 0.9$ kcal/mol, a value established form previous benchmarks on small molecules absent conformational sampling or protonation state issues [25]. Thin error bars in ***Figure 2*** represent the overall estimated error due to both this forcefield error and experimental variability or statistical uncertainty.

To assess overall quantitative accuracy, we computed both root-mean-squared error (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 and the RMSE was $1.72^{2.00}_{1.41}$ kcal/mol. 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 and an RMSE of $1.11^{1.30}_{0.91}$ kcal/mol. 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 (due to experimental error; ***Figure 2***) for this dataset.

How accurately can physical modeling classify mutations as susceptible or resistant?

While quantitative accuracy (MUE, RMSE) is a principle metric of model performance, an application of potential interest is the ability to classify mutations as causing 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, $\Delta\Delta G \leq 1.36$ kcal/mol) or as *resistant* (affinity for mutant is diminished by least 10-fold, $\Delta\Delta G > 1.36$ kcal/mol). Summary statistics of experimental and computational predictions of these classes are shown in ***Figure 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 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 sensitive are classification results to choice of cutoff?

Previous work by O'Hare et al. utilized TKI-specific thresholds for dasatinib, imatinib, and nilotinib [49], which were ~2 kcal/mol. Supplementary ***Figure S2*** shows that when our classification threshold was increased to a 20-fold change in binding (1.77 kcal/mol), FEP+ correctly classified 8 of the 13 resistant mutations and with a threshold of 100-fold change in binding (2.72 kcal/mol), 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 S2***). For the alchemical approach, the two-class 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 two-class 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

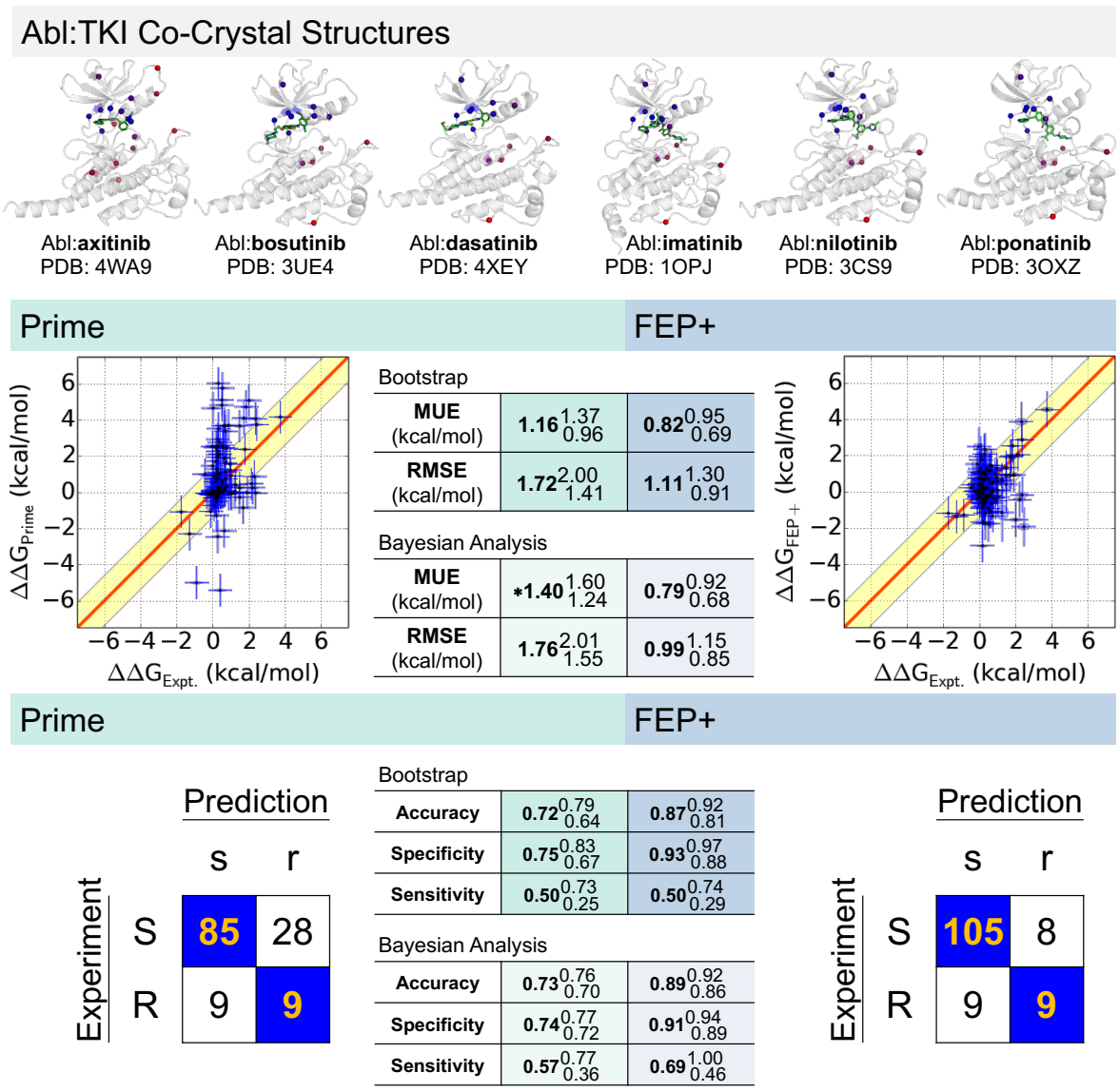


Figure 3. Comparison of experimentally-measured binding free-energy changes (ΔΔG) for 131 clinically observed mutations and 6 selective kinase inhibitors for which co-crystal structures of wild-type kinase with inhibitor are available. *Top panel:* Abl:TKI co-crystal structures used in this study with locations of clinical mutants for each inhibitor highlighted (colored from blue to red for residues nearest to farthest from ligand) in relation to TKI (green sticks) on the corresponding Abl:TKI wild-type crystal structure. *Middle panel:* Scatter plots show Prime and FEP+ computed ΔΔG compared to experiment, with ellipse widths and heights ($\pm\sigma$) for experiment and FEP+ respectively. The red diagonal line indicates when prediction equals experiment, while the yellow shaded region indicates area in which predicted ΔΔG is within 1.36 kcal/mol of experiment (corresponding to a ten-fold error in predicted affinity change). ΔΔG < 0 denotes the mutation increases the susceptibility of the kinase to the inhibitor, while ΔΔG > 0 denotes the mutation increases the resistance of the kinase to the inhibitor. The two mutations that were beyond the concentration limit of the assay (T315I/dasatinib, L248R/imatinib) were not plotted; 129 points were plotted. Truth tables of classification accuracy, sensitivity and specificity using two-classes. *Bottom panel:* Truth tables and classification results include T315I/dasatinib and L248R/imatinib; 131 points were used. For MUE, RMSE, and truth table performance statistics, sub/superscripts denote 95 % CIs. Variability in the experimental data is shown as ellipse widths and uncertainty in our calculations is shown as ellipse heights. Experimental variability was computed as the standard error between IC_{50}-derived $\Delta\Delta G$ measurements made by different labs, 0.32 kcal/mol. The statistical uncertainty in the Prime calculations was zero because the method is deterministic ($\sigma_{cal} = 0$), while the uncertainty in the FEP+ calculations was reported as the standard error, σ_{cal}, of the mean of the predicted $\Delta\Delta G$s from three independent runs. To better highlight true outliers unlikely to simply result from expected forcefield error, we presume forcefield error ($\sigma_{FF} \approx 0.9$ kcal/mol [25]) also behaves as a random error, and represent the total estimated statistical and forcefield error ($\sqrt{\sigma_{FF}^2 + \sigma_{exp/cal}^2}$) as vertical error bars. The horizontal error bars for the experiment (σ_{exp}) was computed as the standard error between ΔpIC_{50} and ΔK_d measurements, 0.58 kcal/mol. 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.

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 (detailed in the Methods) to infer posterior predictive distributions from which expectations and 95% predictive intervals could be obtained. The results of this analysis are presented in ***Figure 3*** (central tables).

FEP+ is significantly better than Prime at predicting the impact of mutations on TKI binding affinities, as the apparent performance (using the original observations) as well as the intrinsic performance (where Bayesian analysis was used to correct for statistical uncertainty or experimental variation) were well-separated outside their 95% confidence intervals in nearly all metrics. Applying the Bayesian model, the MUE and RMSE for FEP+ was $0.79^{0.92}_{0.68}$ kcal/mol and $0.99^{1.15}_{0.85}$ kcal/mol 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 $1.76^{2.01}_{1.55}$ kcal/mol and $1.40^{1.60}_{1.24}$ kcal/mol (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.

Is the impact of point mutations on drug binding equally well-predicted for 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 ***Figure 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, $0.70^{0.93}_{0.50}$ kcal/mol and $0.91^{1.14}_{0.64}$ kcal/mol 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 highly significant, predictive performance for the quantitative metrics MUE and RMSE, which were $0.96^{1.42}_{0.55}$ kcal/mol and $1.41^{1.97}_{0.77}$ kcal/mol respectively (FEP+) and $1.13^{1.83}_{0.60}$ kcal/mol and $1.80^{2.62}_{0.92}$ kcal/mol respectively (Prime). For dasatinib, FEP+ achieved an MUE and RMSE of $0.76^{1.13}_{0.49}$ kcal/mol and $1.07^{1.57}_{0.59}$ kcal/mol 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 $0.82^{1.15}_{0.53}$ kcal/mol and $1.09^{1.43}_{0.69}$ kcal/mol respectively (N=20). Nilotinib, a derivative of imatinib, led to nearly identical quantitative performance results for FEP+ with an MUE and RMSE of $0.82^{1.12}_{0.57}$ kcal/mol and $1.06^{1.39}_{0.69}$ kcal/mol 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 ΔΔG for other inhibitors, FEP+ correctly predicted a very narrow range of ΔΔGs for this drug. This is reflected in the MUE and RMSE of $0.87^{1.16}_{0.62}$ kcal/mol and $1.09^{1.46}_{0.70}$ kcal/mol 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 (false negatives) 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% confidence intervals for five of these mutations spanned the 1.36 kcal/mol 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), imatinib/E255K ($\Delta\Delta G_{FEP+}$=$0.43^{3.05}_{-2.19}$ kcal/mol), imatinib/Y253F ($\Delta\Delta G_{FEP+}$=$0.95^{1.64}_{0.26}$ kcal/mol), and nilotinib/Y253F ($\Delta\Delta G_{FEP+}$=$0.89^{1.69}_{0.09}$ kcal/mol). The bosutinib/V299L mutation was also not significant because the experimental ΔΔG, $1.70^{2.33}_{1.08}$ kcal/mol, included the 1.36 kcal/mol cutoff; the value of ΔΔG predicted by FEP+ for this mutation was $0.91^{1.02}_{0.79}$ kcal/mol, the upper bound of the predicted value was within 0.06 kcal/mol of the lower bound of the experimental value.

Four mutations, however, were misclassified to a degree that is statistically significant given their 95% con-

fidence intervals: 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 much more negative than the corresponding experimental measurements, just as for 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 will induce conformational changes in the apo protein [50], which was not adequately sampled in the relatively short simulations, leading to the errors for T315A mutations for these TKIs. 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 complex electrostatic interactions between the 2,4-dichloro-5-methoxyphenyl ring in bosutinib and the adjacent positively charged amine of the catalytic Lys271 may not be accurately captured by the fixed-charge OPLS3 force field, 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), and the insufficient sampling of ion distributions coupled with P-loop motions might lead to misprediction of these two mutations.

How accurately can the impact of mutations be predicted 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 ***Figure 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 ΔΔG was excellent—MUE and RMSE of $0.58^{0.86}_{0.33}$ kcal/mol and $0.80^{1.09}_{0.44}$ kcal/mol respectively if crystal waters are omitted, and $0.50^{0.78}_{0.26}$ kcal/mol and $0.69^{0.97}_{0.35}$ kcal/mol 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 (***Figure 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—$1.35^{1.67}_{1.03}$ kcal/mol (gefitinib) and $1.58^{1.90}_{1.26}$ kcal/mol (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 ***Figure 5*** is not statistically significant and the classification metrics presented there underestimate the nominal performance of this alchemical free-energy method.

Discussion

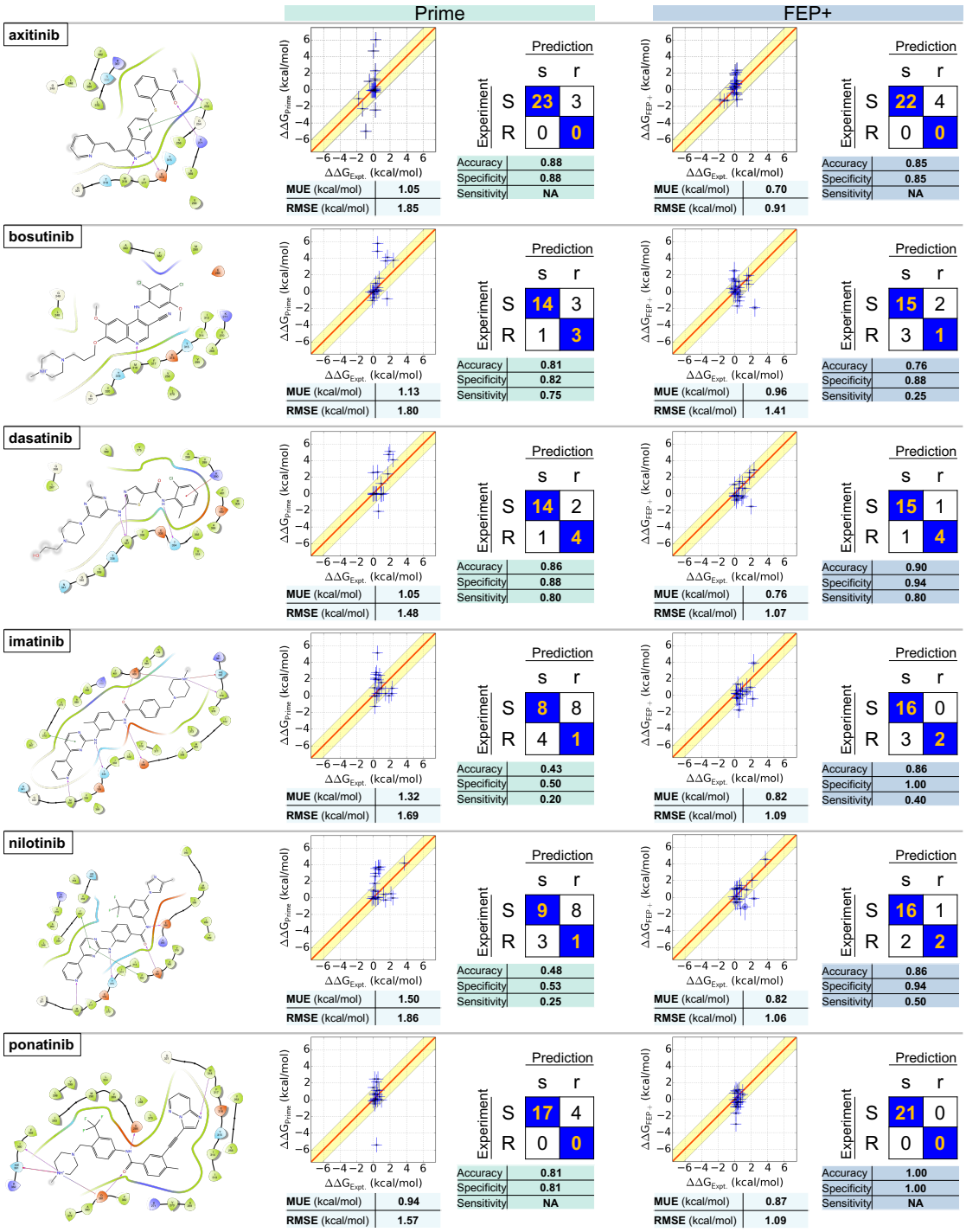


Figure 4. Physical modeling accuracy in computing the impact of clinical Abl mutations on selective inhibitor binding. Ligand interaction diagrams for six selective FDA-approved tyrosine kinase inhibitors (TKIs) for which co-crystal structures with Abl were available (left). Comparisons for clinically-observed mutations are shown for FEP+ (right) and Prime (left). For each ligand, computed *vs.* experimental binding free energies (ΔΔG) are plotted with MUE and RMSE (units of kcal/mol) depicted below. Truth tables are shown to the right. Rows denote *true* susceptible (S, ΔΔG ≤ 1.36 kcal/mol) or resistant (R, ΔΔG > 1.36 kcal/mol) experimental classes using a 1.36 kcal/mol (10-fold change) threshold; columns denote *predicted* susceptible (s, ΔΔG ≤ 1.36 kcal/mol) or resistant (r, ΔΔG > 1.36 kcal/mol). 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.

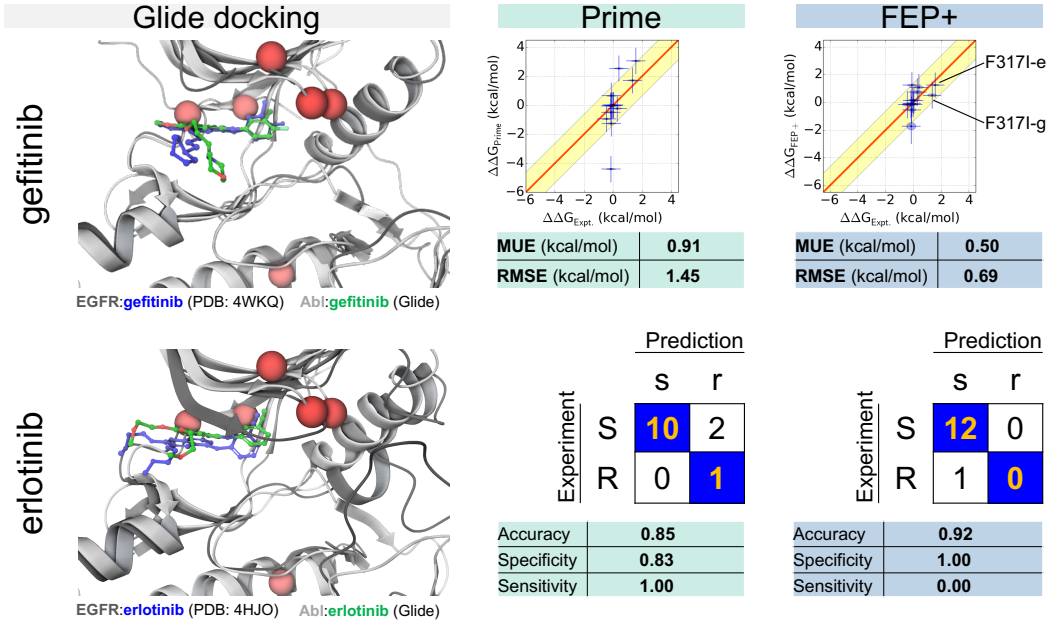


Figure 5. Predicting resistance mutations using FEP+ for inhibitors for which co-crystal structures with wild-type kinase are not available. The docked pose of Abl:erlotinib is superimposed on the co-crystal structure of EGFR:erlotinib; erlotinib docked to Abl (light gray) is depicted in green and erlotinib 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.

Physics-based modeling can reliably predict when a mutation elicits resistance to therapy

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 ($0.73^{0.80}_{0.66}$) and specificity ($0.76^{0.84}_{0.69}$) while the alchemical approach FEP+ was a significant improvement in classification accuracy ($0.88^{0.93}_{0.82}$) and specificity ($0.94^{0.98}_{0.89}$). 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 $1.70^{1.98}_{1.40}$ kcal/mol and $1.14^{1.35}_{0.93}$ kcal/mol respectively. In stark contrast, the quantitative accuracy of FEP+ was statistically superior to Prime with an RMSE and an MUE of $1.07^{1.26}_{0.89}$ kcal/mol and $0.79^{0.92}_{0.67}$ kcal/mol respectively.

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 (by 1/2, 1/3, and by 0.23, which was the optimal factor that gives lowest RMSE) and recalculated the classification metrics (***Table S8***). As expected, the MUE and RMSE were improved but the specificity of Prime was drastically diminished; as MUE and RMSE improved, it became increasingly unable to identify resistance mutations. Scaling FEP+ eliminated its sensitivity and a naïve model (where all free energies were set to 0.00 kcal/mol) had zero sensitivity. Lastly, we constructed a consensus model in which free energies were a weighted average of scaled Prime and FEP+. However, this model also had no sensitivity. It appears difficult to improve upon the predictive power of FEP+ by statistical operations.

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 (scaled and non-scaled), FEP+ (scaled and non-scaled), naïve model, and consensus model (constructed from scaled Prime and scaled FEP+, see above). ROC curves are independent of a linear transformation on the predicted dataset. Therefore, ROC curves and ROC-AUCs for scaled and non-scaled Prime were identical, as well as scaled and non-scaled FEP+. ROC curves for these six sets of predictions are presented in Supplementary ***Figure S3***. 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 naïve 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 apparently has poor discriminatory power (ROC-AUC in [0.6,0.7]) while FEP+ apparently has fair discriminatory power (ROC-AUC in [0.7,0.8]).

Hierarchical Bayesian model estimates global performance (N=144)

A hierarchical Bayesian approach was developed to estimate the intrinsic accuracy of the models when the noise in the experimental and predicted values of ΔΔG was accounted for. Utilizing this approach, the MUE and RMSE for Prime was found to be $1.39^{1.58}_{1.23}$ kcal/mol and $1.75^{1.98}_{1.55}$ kcal/mol (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 $0.76^{0.87}_{0.66}$ kcal/mol and $0.95^{1.09}_{0.82}$ kcal/mol (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.

Examining the physical and chemical features of outliers

Current alchemical approaches neglect effects that will continue to improve accuracy

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 [25] force field can be further improved by explicit consideration of polarizability effects [51], as hinted by some small-scale benchmarks [52]. 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

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)	RMSE (kcal/mol)	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	$0.85^{1.00}_{0.62}$	$0.83^{1.00}_{0.58}$	$1.00^{1.00}_{0.00}$

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

Accuracy, specificity, and sensitivity were computed to assess two-class prediction performance:
resistant ($\Delta\Delta G > 1.36$ kcal/mol) or *susceptible* ($\Delta\Delta G \leq 1.36$ kcal/mol).

95% CIs (sub-/superscripts) were estimated from 1000 bootstrap replicates. Note: The sensitivity for axitinib and ponatinib is NA, because there is no resistant mutation for these two drugs.

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 [53]; recently, efficient Lennard-Jones PME methods [54, 55] and perturbation schemes [53] 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 (see Methods) 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 [56, 57]. Similarly, some systems display significant salt concentration dependence [58], while the simulations for some systems reported here did not rigorously mimic all aspects of the experimental conditions of the cell viability assays.

Experimentally observed IC_{50} changes can be caused by other physical mechanisms

While we have shown that predicting the direct impact of mutations on the binding affinity of ATP-competitive tyrosine kinase inhibitors 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. These physical mechanisms might affect the IC_{50}s of cell viability assays but not necessarily the binding affinity of the inhibitors. While many of these effects are in principle tractable by physical modeling in general (and alchemical free energy methods in particular), it is valuable to examine our mispredictions and outliers to identify whether any of these cases is likely to induce resistance (as observed by ΔpIC_{50} shifts) by one of these alternative mechanisms.

Other physical mechanisms of resistance are likely similarly computable.

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. Except for 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 the same thermodynamic cycle utilized here for inhibitors.

Conclusion

Revolutionary changes in computing power—especially the arrival of inexpensive graphics processors (GPUs)—and software automation have enabled alchemical free-energy calculations to impact drug discovery and life sciences projects in previously unforeseen ways. 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, which uses an implicit model of solvation to score the energetic changes in interaction energy that arise from the mutation, 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 (N=142), RMSE of $1.70^{1.98}_{1.40}$ kcal/mol 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 (N=142), RMSE of $1.07^{1.26}_{0.89}$ kcal/mol (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 turnaround times ~6 hours/calculation) to impact research projects in drug discovery and the life sciences. With exponential improvements in computing power, we anticipate the domains of applicability for alchemical free-energy methods such as FEP+ will take on increasingly integrated roles to impact projects. 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) [59] and Abl:dasatinib (PDB: 4XEY) [60] was adjusted to match our convention.

Complexes with co-crystal structures. Chain B of the experimental structure of Abl:axitinib (PDB: 4WA9) [44] was used, and four missing residues at the N- and C-termini were added using homology modeling with PDB 3IK3 [61] 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) [62] 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) [60] 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) [59] 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) [63] 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) [64] 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 [64] 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 [65] with a pH range of 5.0–9.0. 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/Å^2) was applied only to heavy atoms. ***Table S9*** 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 [66] 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) [67] and gefitinib-bound (PDB: 4WKQ) [68] 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 7.0±2.0 were determined using Epik [69]. 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 [25] 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 [70] 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 [39]. VSGB [71] 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 ΔΔG calculated using the thermodynamic cycle depicted in ***Figure 1***b. 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 ΔΔG 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 [72] 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 [70]) 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.* [73].

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/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 [74] 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) [75] and a 9 Å cutoff distance was used for van de Waals interactions. The production MD simulation was performed in the NPT ensemble using the MTK method [76] with integration time steps of 4 fs, 4 fs, and 8 fs respectively for the bonded, near, and far interactions following the RESPA method [77] through hydrogen mass repartitioning [78]. Production FEP+ calculations utilized Hamiltonian replica exchange with solute tempering (REST) [79], 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 [79], 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 [80, 81] The reported ΔΔG was computed as the mean of the computed ΔΔG 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 hours in total to compute ΔΔG.

Obtaining ΔΔG 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 (***Equation 1***). This assumption has been successfully used to estimate relative free energies [37, 82–84] using the relationship between IC$_{50}$ and competitive inhibitor affinity K_i,

$$\mathrm{IC}_{50} = \frac{K_i}{1 + \frac{[S_0]}{K_M}}. \tag{1}$$

If the Michaelis constant for ATP (K_M) is much larger than the initial ATP concentration S_0, the relation in ***Equation 1*** 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,

$$\Delta\Delta G = -RT \ln \frac{\mathrm{IC}_{50,WT}}{\mathrm{IC}_{50,mut}} \tag{2}$$

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 ΔΔGs 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-/down-regulation of positive-/negative-feedback effectors, diminished synergism of pro-apoptotic machinery, decoupling of the target from cell survival circuits) [85, 86], their inhibition ability might not correlate well with binding affinity. However, the comparison between ΔpIC$_{50}$ and ΔK_D is presented in ***Figure 2***d, and this comparison indicates the assumptions we used to relate ΔpIC$_{50}$ to ΔΔG are reasonable for the dataset we studied.

Assessing prediction performance

Quantitative accuracy metrics

Mean unsigned error (MUE) was calculated by taking the average absolute difference between predicted and experimental estimates of ΔΔG. Root-mean square error (RMSE) was calculated by taking the square root of the average squared difference between predicted and experimental estimates of ΔΔG. 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 (ΔΔG ≤ 1.36 kcal/mol) or resistant (ΔΔG > 1.36 kcal/mol), where the 1.36 kcal/mol 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 (ΔΔG > 1.36 kcal/mol) or susceptible (ΔΔG ≤ 1.36 kcal/mol). 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 ΔΔG. Cutoffs were chosen by taking the minimum and maximum value of ΔΔG for a data set (Prime or FEP+), and iteratively computing specificity and sensitivity in steps of 0.001 kcal/mol, which by this definition will be in the range [0,1]. Experimental positives and negatives were classified with the 1.36 kcal/mol cutoff. ROC-AUC was computed using the trapezoidal rule.

Estimating uncertainties of physical-modeling results

95% symmetric confidence intervals (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 $x^{x_{\text{high}}}_{x_{\text{low}}}$ 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 2.5^{th} and 97.5^{th} percentiles of the value-sorted list of the bootstrap samples. Uncertainty for ΔΔGs was computed by the standard deviation between three independent runs (using different random seeds to set initial velocities), where the 95% CI was [ΔΔG−1.96×$\sigma_{\text{FEP+}}$, ΔΔG+1.96×$\sigma_{\text{FEP+}}$] kcal/mol. 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 ***Figure 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, $\Delta\Delta G_i^{\text{true}}$, comes from a normal background distribution of unknown mean and variance,

$$\Delta\Delta G_i^{\text{true}} \sim \mathcal{N}(\mu_{\text{mut}}, \sigma^2_{\text{mut}}) \; i = 1, \ldots, M \tag{3}$$

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

$$\mu_{\text{mut}} \sim U(-6, +6) \tag{4}$$
$$\sigma_{\text{mut}} \propto 1 \tag{5}$$

where we limit $\sigma > 0$.

We presume the true computational predictions (absent statistical error) differ from the (unknown) true free energy difference of mutation $\Delta\Delta G_i^{\text{true}}$ 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:

$$\Delta\Delta G_{i,\text{Prime}}^{\text{true}} \sim \mathcal{N}(\Delta\Delta G_i^{\text{true}}, \text{RMSE}_{\text{Prime}}^2) \tag{6}$$
$$\Delta\Delta G_{i,\text{FEP+}}^{\text{true}} \sim \mathcal{N}(\Delta\Delta G_i^{\text{true}}, \text{RMSE}_{\text{FEP+}}^2) \tag{7}$$

In the case of Prime, since the computation is deterministic, we actually calculate $\Delta\Delta G_{\text{Prime}}^{\text{true}}$ 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 $\sigma_{\text{calc},i}$,

$$\Delta\Delta G_{i,\text{FEP+}} \sim \mathcal{N}(\Delta\Delta G_{i,\text{FEP+}}, \sigma_{i,\text{FEP+}}^2) \tag{8}$$

where $\Delta\Delta G_{i,\text{FEP+}}$ is the free energy computed for mutant i by FEP+, and $\sigma_{i,\text{FEP+}}$ 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}:

$$\Delta\Delta G_{i,\text{exp}} \sim \mathcal{N}(\Delta\Delta G_i, \sigma_{\text{exp}}^2) \tag{9}$$

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

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

$$\text{MUE} = \int dx_{\text{true}}\, p(x_{\text{true}}) \int dx_{\text{calc}}\, p(x_{\text{calc}}|x_{\text{true}})\, |x_{\text{calc}} - x_{\text{true}}| \tag{10}$$
$$= \int dx_{\text{true}} \frac{1}{\sqrt{2\pi\sigma_{\text{true}}^2}} e^{-\frac{(x_{\text{true}}-\mu_{\text{true}})^2}{2\sigma_{\text{true}}^2}} \int dx_{\text{calc}} \frac{1}{\sqrt{2\pi\sigma_{\text{calc}}^2}} e^{-\frac{(x_{\text{calc}}-\mu_{\text{true}})^2}{2\sigma_{\text{calc}}^2}} |x_{\text{calc}} - x_{\text{true}}| \tag{11}$$
$$= \sqrt{\frac{2}{\pi}}\,\text{RMSE} \tag{12}$$

The model was implemented using PyMC3 [87], 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.

Data availability

Compiled experimental datasets, input files for Prime and FEP+ and computational results can be found at the following URL: https://goo.gl/6cC8Bu

Code availability

Scripts used for statistics analysis (including the Bayesian inference model) can be found at the following URL: https://goo.gl/6cC8Bu

Acknowledgments

We thank Daniel Robinson (Schrödinger), Sonya M. Hanson (MSKCC), and Gregory A. Ross (MSKCC) for helpful discussions. JDC acknowledges support from NIH National Cancer Institute Cancer Center Core Grant P30 CA008748; JDC and SKA acknowledge support from the Sloan Kettering Institute, Cycle for Survival, and NIH grant R01 GM121505. KH 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.

Disclosures

JDC is a member of the Scientific Advisory Board for Schrödinger Inc.

Author Contributions

KH, JDC, CN, RA, and LW designed the research; KH, SA, TS, and LW identified experimental datasets; KH and LW performed the simulations; KH, CN, SKA, SR, TS, RA, JDC, and LW analyzed the data; KH, JDC, SKA, and LW wrote the paper.

References

[1] **Robert Roskoski Jr**. USFDA Approved Protein Kinase Inhibitors. . 2017; http://www.brimr.org/PKI/PKIs.htm, updated 3 May 2017.

[2] **Santos R**, Ursu O, Gaulton A, Bento AP, Donadi RS, Bologa CG, Karlsson A, Al-Lazikani B, Hersey A, Oprea TI, Overington JP. A Comprehensive Map of Molecular Drug Targets. Nat Rev Drug Discov. 2016 Dec; 16(1):19–34. doi: 10.1038/nrd.2016.230.

[3] **Shah NP**, Nicoll JM, Nagar B, Gorre ME, Paquette RL, Kuriyan J, Sawyers CL. 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. 2002 Aug; 2(2):117–125.

[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. Biochimica et Biophysica Acta (BBA) - Reviews on Cancer. 2014; 1845(1):31 – 41. http://www.sciencedirect.com/science/article/pii/S0304419X13000437, doi: https://doi.org/10.1016/j.bbcan.2013.10.001.

[5] **Huang L**, Fu L. Mechanisms of Resistance to EGFR Tyrosine Kinase Inhibitors. Acta Pharm Sin B. 2015; 5(5):390–401.

[6] **Meyer SC**, Levine RL. Molecular Pathways: Molecular Basis for Sensitivity and Resistance to JAK Kinase Inhibitors. Clin Cancer Res. 2014; 20(8):2051–2059. doi: 10.1158/1078-0432.CCR-13-0279.

[7] **Davare MA**, Vellore NA, Wagner JP, Eide CA, Goodman JR, Drilon A, Deininger MW, O?Hare T, Druker BJ. Structural Insight into Selectivity and Resistance Profiles of ROS1 Tyrosine Kinase Inhibitors. Proc Natl Acad Sci. 2015; 112(39):E5381–E5390. doi: 10.1073/pnas.1515281112.

[8] **Van Allen EM**, Wagle N, Sucker A, Treacy DJ, Johannessen CM, Goetz EM, Place CS, Taylor-Weiner A, Whittaker S, Kryukov GV, Hodis E, Rosenberg M, McKenna A, Cibulskis K, Farlow D, Zimmer L, Hillen U, Gutzmer R, Goldinger SM, Ugurel S, et al. The Genetic Landscape of Clinical Resistance to RAF Inhibition in Metastatic Melanoma. Cancer Discov. 2014; 4(1):94–109. doi: 10.1158/2159-8290.CD-13-0617.

[9] **Rani S**, Corcoran C, Shiels L, Germano S, Breslin S, Madden S, McDermott MS, Browne BC, OtextquoterightDonovan N, Crown J, Gogarty M, Byrne AT, OtextquoterightDriscoll L. Neuromedin U: A Candidate Biomarker and Therapeutic Target to Predict and Overcome Resistance to HER-Tyrosine Kinase Inhibitors. Cancer Res. 2014; 74(14):3821–3833. doi: 10.1158/0008-5472.CAN-13-2053.

[10] **Holohan C**, Van Schaeybroeck S, Longley DB, Johnston PG. Cancer Drug Resistance: An Evolving Paradigm. Nat Rev Cancer. 2013 Sep; 13(10):714–726. doi: 10.1038/nrc3599.

[11] **Weisberg E**, Manley PW, Cowan-Jacob SW, Hochhaus A, Griffin JD. Second Generation Inhibitors of BCR-ABL for the Treatment of Imatinib-Resistant Chronic Myeloid Leukaemia. Nat Rev Cancer. 2007 May; 7(5):345–356. doi: 10.1038/nrc2126.

[12] **Y Lu X**, Cai Q, Ding K. Recent Developments in the Third Generation Inhibitors of Bcr-Abl for Overriding T315I Mutation. Curr Med Chem. 2011 May; 18(14):2146–2157. doi: 10.2174/092986711795656135.

[13] **Juchum M**, Günther M, Laufer SA. Fighting Cancer Drug Resistance: Opportunities and Challenges for Mutation-Specific EGFR Inhibitors. Drug Resist Updat. 2015 May; 20:12–28. doi: 10.1016/j.drup.2015.05.002.

[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. 2015 Jan; 5(1):34–37. doi: 10.1016/j.apsb.2014.12.007.

[15] **Neel DS**, Bivona TG. Resistance Is Futile: Overcoming Resistance to Targeted Therapies in Lung Adenocarcinoma. Npj Precis Oncol. 2017 Dec; 1(1). doi: 10.1038/s41698-017-0007-0.

[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. Leuk Off J Leuk Soc Am Leuk Res Fund UK. 2006 Dec; 20:2057–60.

[17] **Chandarlapaty S**, Sawai A, Scaltriti M, Rodrik-Outmezguine V, Grbovic-Huezo O, Serra V, Majumder PK, Baselga J, Rosen N. AKT Inhibition Relieves Feedback Suppression of Receptor Tyrosine Kinase Expression and Activity. Cancer Cell. 2011 Jan; 19(1):58–71. doi: 10.1016/j.ccr.2010.10.031.

[18] **Knight ZA**, Lin H, Shokat KM. Targeting the Cancer Kinome through Polypharmacology. Nat Rev Cancer. 2010; 10(2):130.

[19] **Housman G**, Byler S, Heerboth S, Lapinska K, Longacre M, Snyder N, Sarkar S. Drug Resistance in Cancer: An Overview. Cancers. 2014 Sep; 6(3):1769–1792. doi: 10.3390/cancers6031769.

[20] **Zehir A**, Benayed R, Shah RH, Syed A, Middha S, Kim HR, Srinivasan P, Gao J, Chakravarty D, Devlin SM, Hellmann MD, Barron DA, Schram AM, Hameed M, Dogan S, Ross DS, Hechtman JF, DeLair DF, Yao J, Mandelker DL, et al. Mutational Landscape of Metastatic Cancer Revealed from Prospective Clinical Sequencing of 10,000 Patients. Nat Med. 2017 May; 23(6):703–713. doi: 10.1038/nm.4333.

[21] **Redig AJ**, Jänne PA. Basket Trials and the Evolution of Clinical Trial Design in an Era of Genomic Medicine. American Society of Clinical Oncology; 2015.

[22] **Hyman DM**, Taylor BS, Baselga J. Implementing Genome-Driven Oncology. Cell. 2017 Feb; 168(4):584–599. doi: 10.1016/j.cell.2016.12.015.

[23] **Pesesky MW**, Hussain T, Wallace M, Patel S, Andleeb S, Burnham CAD, Dantas G. Evaluation of Machine Learning and Rules-Based Approaches for Predicting Antimicrobial Resistance Profiles in Gram-Negative Bacilli from Whole Genome Sequence Data. Front Microbiol. 2016 Nov; 7. doi: 10.3389/fmicb.2016.01887.

[24] **Melnikov A**, Rogov P, Wang L, Gnirke A, Mikkelsen TS. Comprehensive Mutational Scanning of a Kinase *in Vivo* Reveals Substrate-Dependent Fitness Landscapes. Nucleic Acids Res. 2014 Aug; 42(14):e112–e112. doi: 10.1093/nar/gku511.

[25] **Harder E**, Damm W, Maple J, Wu C, Reboul M, Xiang JY, Wang L, Lupyan D, Dahlgren MK, Knight JL, Kaus JW, Cerutti DS, Krilov G, Jorgensen WL, Abel R, Friesner RA. OPLS3: A Force Field Providing Broad Coverage of Drug-like Small Molecules and Proteins. J Chem Theory Comput. 2016 Jan; 12(1):281–296. doi: 10.1021/acs.jctc.5b00864.

[26] **Huang J**, MacKerell AD. CHARMM36 All-Atom Additive Protein Force Field: Validation Based on Comparison to NMR Data. J Comput Chem. 2013 Sep; 34(25):2135–2145. doi: 10.1002/jcc.23354.

[27] **Maier JA**, Martinez C, Kasavajhala K, Wickstrom L, Hauser KE, Simmerling C. ff14SB: Improving the Accuracy of Protein Side Chain and Backbone Parameters from ff99SB. J Chem Theory Comput. 2015 Aug; 11(8):3696–3713. doi: 10.1021/acs.jctc.5b00255.

[28] **Chodera JD**, Mobley DL, Shirts MR, Dixon RW, Branson K, Pande VS. Alchemical Free Energy Methods for Drug Discovery: Progress and Challenges. Curr Opin Struct Biol. 2011 Apr; 21(2):150–160. doi: 10.1016/j.sbi.2011.01.011.

[29] **Wang L**, Wu Y, Deng Y, Kim B, Pierce L, Krilov G, Lupyan D, Robinson S, Dahlgren MK, Greenwood J, Romero DL, Masse C, Knight JL, Steinbrecher T, Beuming T, Damm W, Harder E, Sherman W, Brewer M, Wester R, 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. 2015 Feb; 137(7):2695–2703. doi: 10.1021/ja512751q.

[30] **Abel R**, Mondal S, Masse C, Greenwood J, Harriman G, Ashwell MA, Bhat S, Wester R, Frye L, Kapeller R, et al. Accelerating drug discovery through tight integration of expert molecular design and predictive scoring. Current opinion in structural biology. 2017; 43:38–44.

[31] **Aldeghi M**, Heifetz A, Bodkin MJ, Knapp S, Biggin PC. Accurate Calculation of the Absolute Free Energy of Binding for Drug Molecules. Chem Sci. 2016; 7(1):207–218. doi: 10.1039/C5SC02678D.

[32] **Cappel D**, Hall ML, Lenselink EB, Beuming T, Qi J, Bradner J, Sherman W. Relative Binding Free Energy Calculations Applied to Protein Homology Models. J Chem Inf Model. 2016; 56(12):2388–2400. doi: 10.1021/acs.jcim.6b00362.

[33] **Clark AJ**, Gindin T, Zhang B, Wang L, Abel R, Murret CS, Xu F, Bao A, Lu NJ, Zhou T, et al. Free Energy Perturbation Calculation of Relative Binding Free Energy between Broadly Neutralizing Antibodies and the gp120 Glycoprotein of HIV-1. Journal of molecular biology. 2017; 429(7):930–947.

[34] **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. Journal of molecular biology. 2017; 429(7):948–963.

[35] **Ford MC**, Babaoglu K. Examining the Feasibility of Using Free Energy Perturbation (FEP+) in Predicting Protein Stability. J Chem Inf Model. 2017 Jun; 57(6):1276–1285. doi: 10.1021/acs.jcim.7b00002.

[36] **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. 2016 Dec; 138(48):15682–15689. doi: 10.1021/jacs.6b09511.

[37] **Mondal J**, Tiwary P, Berne BJ. How a Kinase Inhibitor Withstands Gatekeeper Residue Mutations. J Am Chem Soc. 2016; 138(13):4608–4615. doi: 10.1021/jacs.6b01232.

[38] **Lovering F**, Aevazelis C, Chang J, Dehnhardt C, Fitz L, Han S, Janz K, Lee J, Kaila N, McDonald J, Moore W, Moretto A, Papaioannou N, Richard D, Ryan MS, Wan ZK, Thorarensen A. Imidazotriazines: Spleen Tyrosine Kinase (Syk) Inhibitors Identified by Free-Energy Perturbation (FEP). ChemMedChem. 2016 Jan; 11(2):217–233. doi: 10.1002/cmdc.201500333.

[39] **Rapp C**, Kalyanaraman C, Schiffmiller A, Schoenbrun EL, Jacobson MP. A Molecular Mechanics Approach to Modeling Protein–Ligand Interactions: Relative Binding Affinities in Congeneric Series. J Chem Inf Model. 2011 Sep; 51(9):2082–2089. doi: 10.1021/ci200033n.

[40] **Shirts MR**, Mobley DL, Chodera JD. Chapter 4 Alchemical Free Energy Calculations: Ready for Prime Time? In: *Annual Reports in Computational Chemistry*, vol. 3 Elsevier; 2007.p. 41–59.

[41] **Mobley DL**, Klimovich PV. Perspective: Alchemical Free Energy Calculations for Drug Discovery. J Chem Phys. 2012 Dec; 137(23):230901. doi: 10.1063/1.4769292.

[42] **Abel R**, Mondal S, Masse C, Greenwood J, Harriman G, Ashwell MA, Bhat S, Wester R, Frye L, Kapeller R, Friesner RA. Accelerating Drug Discovery through Tight Integration of Expert Molecular Design and Predictive Scoring. Curr Opin Struct Biol. 2017 Apr; 43:38–44. doi: 10.1016/j.sbi.2016.10.007.

[43] **Kuhn B**, Tichý M, Wang L, Robinson S, Martin RE, Kuglstatter A, Benz J, Giroud M, Schirmeister T, Abel R, Diederich F, Hert J. Prospective Evaluation of Free Energy Calculations for the Prioritization of Cathepsin L Inhibitors. J Med Chem. 2017 Mar; 60(6):2485–2497. doi: 10.1021/acs.jmedchem.6b01881.

[44] **Pemovska T**, Johnson E, Kontro M, Repasky GA, Chen J, Wells P, Cronin CN, McTigue M, Kallioniemi O, Porkka K, Murray BW, Wennerberg K. Axitinib Effectively Inhibits BCR-ABL1(T315I) with a Distinct Binding Conformation. Nature. 2015 Feb; 519(7541):102–105. doi: 10.1038/nature14119.

[45] **Schrock A**, Chen TH, Clackson T, Rivera VM. 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. 2013; 122(21):3992–3992.

[46] **Davis MI**, Hunt JP, Herrgard S, Ciceri P, Wodicka LM, Pallares G, Hocker M, Treiber DK, Zarrinkar PP. Comprehensive Analysis of Kinase Inhibitor Selectivity. Nat Biotechnol. 2011 Oct; 29(11):1046–1051. doi: 10.1038/nbt.1990.

[47] **Soverini S**, Colarossi S, Gnani A, Rosti G, Castagnetti F, Poerio A, Iacobucci I, Amabile M, Abruzzese E, Orlandi E, Radaelli F, Ciccone F, Tiribelli M, di Lorenzo R, Caracciolo C, Izzo B, Pane F, Saglio G, Baccarani M, Martinelli G. 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. Clinical Cancer Research. 2006; 12(24):7374–7379. http://clincancerres.aacrjournals.org/content/12/24/7374, doi: 10.1158/1078-0432.CCR-06-1516.

[48] **O'Hare T**, Eide CA, Deininger MW. Bcr-Abl kinase domain mutations, drug resistance, and the road to a cure for chronic myeloid leukemia. Blood. 2007; 110(7):2242–2249.

[49] **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. 2005 Oct; 11(19):6987–6993. doi: 10.1158/1078-0432.CCR-05-0622.

[50] **Shan Y**, Seeliger MA, Eastwood MP, Frank F, Xu H, Jensen MØ, Dror RO, Kuriyan J, Shaw DE. A Conserved Protonation-Dependent Switch Controls Drug Binding in the Abl Kinase. Proc Natl Acad Sci. 2009; 106(1):139–144.

[51] **Demerdash O**, Yap EH, Head-Gordon T. Advanced Potential Energy Surfaces for Condensed Phase Simulation. Annu Rev Phys Chem. 2014 Apr; 65(1):149–174. doi: 10.1146/annurev-physchem-040412-110040.

[52] **Jiao D**, Golubkov PA, Darden TA, Ren P. Calculation of Protein–ligand Binding Free Energy by Using a Polarizable Potential. Proc Natl Acad Sci. 2008; 105(17):6290–6295.

[53] **Shirts MR**, Mobley DL, Chodera JD, Pande VS. Accurate and Efficient Corrections for Missing Dispersion Interactions in Molecular Simulations. J Phys Chem B. 2007 Nov; 111(45):13052–13063. doi: 10.1021/jp0735987.

[54] **Essmann U**, Perera L, Berkowitz ML, Darden T, Lee H, Pedersen LG. A Smooth Particle Mesh Ewald Method. J Chem Phys. 1995 Nov; 103(19):8577–8593. doi: 10.1063/1.470117.

[55] **Wennberg CL**, 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. 2013 Aug; 9(8):3527–3537. doi: 10.1021/ct400140n.

[56] **Onufriev AV**, Alexov E. Protonation and pK Changes in Protein–ligand Binding. Q Rev Biophys. 2013 May; 46(02):181–209. doi: 10.1017/S0033583513000024.

[57] **Martin YC**. Let's Not Forget Tautomers. J Comput Aided Mol Des. 2009 Oct; 23(10):693–704. doi: 10.1007/s10822-009-9303-2.

[58] **Jensen J**. Calculating pH and Salt Dependence of Protein-Protein Binding. Curr Pharm Biotechnol. 2008 Apr; 9(2):96–102. doi: 10.2174/138920108783955146.

[59] **Nagar B**, Hantschel O, Young MA, Scheffzek K, Veach D, Bornmann W, Clarkson B, Superti-Furga G, Kuriyan J. Structural Basis for the Autoinhibition of C-Abl Tyrosine Kinase. Cell. 2003; 112(6):859–871.

[60] **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. 2015 Jun; 468(2):283–291. doi: 10.1042/BJ20141492.

[61] **O'Hare T**, Shakespeare WC, Zhu X, Eide CA, Rivera VM, Wang F, Adrian LT, Zhou T, Huang WS, Xu Q, Metcalf CA, Tyner JW, Loriaux MM, Corbin AS, Wardwell S, Ning Y, Keats JA, Wang Y, Sundaramoorthi R, Thomas M, et al. AP24534, a Pan-BCR-ABL Inhibitor for Chronic Myeloid Leukemia, Potently Inhibits the T315I Mutant and Overcomes Mutation-Based Resistance. Cancer Cell. 2009 Nov; 16(5):401–412. doi: 10.1016/j.ccr.2009.09.028.

[62] **Levinson NM**, Boxer SG. Structural and Spectroscopic Analysis of the Kinase Inhibitor Bosutinib and an Isomer of Bosutinib Binding to the Abl Tyrosine Kinase Domain. PLoS ONE. 2012 Apr; 7(4):e29828. doi: 10.1371/journal.pone.0029828.

[63] **Weisberg E**, Manley PW, Breitenstein W, Brüggen J, Cowan-Jacob SW, Ray A, Huntly B, Fabbro D, Fendrich G, Hall-Meyers E, Kung AL, Mestan J, Daley GQ, Callahan L, Catley L, Cavazza C, Mohammed A, Neuberg D, Wright RD, Gilliland DG, et al. Characterization of AMN107, a Selective Inhibitor of Native and Mutant Bcr-Abl. Cancer Cell. 2005 Feb; 7(2):129–141. doi: 10.1016/j.ccr.2005.01.007.

[64] **Zhou T**, Commodore L, Huang WS, Wang Y, Thomas M, Keats J, Xu Q, Rivera VM, Shakespeare WC, Clackson T, Dalgarno DC, Zhu X. Structural Mechanism of the Pan-BCR-ABL Inhibitor Ponatinib (AP24534): Lessons for Overcoming Kinase Inhibitor Resistance: Structural Mechanism of Ponatinib. Chem Biol Drug Des. 2011 Jan; 77(1):1–11. doi: 10.1111/j.1747-0285.2010.01054.x.

[65] **Li H**, Robertson AD, Jensen JH. Very Fast Empirical Prediction and Rationalization of Protein pKa Values. Proteins Struct Funct Bioinforma. 2005 Oct; 61(4):704–721. doi: 10.1002/prot.20660.

[66] **Friesner RA**, Banks JL, Murphy RB, Halgren TA, Klicic JJ, Mainz DT, Repasky MP, Knoll EH, Shelley M, Perry JK, Shaw DE, Francis P, Shenkin PS. Glide: A New Approach for Rapid, Accurate Docking and Scoring. 1. Method and Assessment of Docking Accuracy. J Med Chem. 2004 Mar; 47(7):1739–1749. doi: 10.1021/jm0306430.

[67] **Park JH**, Liu Y, Lemmon MA, Radhakrishnan R. Erlotinib Binds Both Inactive and Active Conformations of the EGFR Tyrosine Kinase Domain. Biochem J. 2012 Dec; 448(3):417–423. doi: 10.1042/BJ20121513.

[68] **Yosaatmadja Y**, Squire CJ. 1.85 Angstrom Structure of EGFR Kinase Domain with Gefitinib. . 2014 Nov; doi: 10.2210/pdb4wkq/pdb.

[69] **Shelley JC**, Cholleti A, Frye LL, Greenwood JR, Timlin MR, Uchimaya M. Epik: A Software Program for pK a Prediction and Protonation State Generation for Drug-like Molecules. J Comput Aided Mol Des. 2007 Dec; 21(12):681–691. doi: 10.1007/s10822-007-9133-z.

[70] **Berendsen HJC**, Postma JPM, van Gunsteren WF, Hermans J. Interaction Models for Water in Relation to Protein Hydration. In: Pullman B, editor. *Intermolecular Forces*, vol. 14 Dordrecht: Springer Netherlands; 1981.p. 331–342. doi: 10.1007/978-94-015-7658-1_21.

[71] **Shivakumar D**, Williams J, Wu Y, Damm W, Shelley J, Sherman W. Prediction of Absolute Solvation Free Energies Using Molecular Dynamics Free Energy Perturbation and the OPLS Force Field. J Chem Theory Comput. 2010 May; 6(5):1509–1519. doi: 10.1021/ct900587b.

[72] **Pearlman DA**. A Comparison of Alternative Approaches to Free Energy Calculations. J Phys Chem. 1994 Feb; 98(5):1487–1493. doi: 10.1021/j100056a020.

[73] **Rocklin GJ**, Mobley DL, Dill KA, Hünenberger PH. 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. 2013 Nov; 139(18):184103. doi: 10.1063/1.4826261.

[74] **Feller SE**, Zhang Y, Pastor RW, Brooks BR. Constant pressure molecular dynamics simulation: The Langevin piston method. The Journal of Chemical Physics. 1995; 103(11):4613–4621. https://doi.org/10.1063/1.470648, doi: 10.1063/1.470648.

[75] **Essmann U**, Perera L, Berkowitz ML, Darden T, Lee H, Pedersen LG. A smooth particle mesh Ewald method. The Journal of Chemical Physics. 1995; 103(19):8577–8593. https://doi.org/10.1063/1.470117, doi: 10.1063/1.470117.

[76] **Martyna GJ**, Tobias DJ, Klein ML. Constant pressure molecular dynamics algorithms. The Journal of Chemical Physics. 1994; 101(5):4177–4189. https://doi.org/10.1063/1.467468, doi: 10.1063/1.467468.

[77] **Tuckerman M**, Berne BJ, Martyna GJ. Reversible multiple time scale molecular dynamics. The Journal of Chemical Physics. 1992; 97(3):1990–2001. https://doi.org/10.1063/1.463137, doi: 10.1063/1.463137.

[78] **Hopkins CW**, Le Grand S, Walker RC, Roitberg AE. Long-Time-Step Molecular Dynamics through Hydrogen Mass Repartitioning. Journal of Chemical Theory and Computation. 2015; 11(4):1864–1874. http://dx.doi.org/10.1021/ct5010406, doi: 10.1021/ct5010406, pMID: 26574392.

[79] **Wang L**, Berne BJ, Friesner RA. On Achieving High Accuracy and Reliability in the Calculation of Relative Protein-Ligand Binding Affinities. Proc Natl Acad Sci. 2012 Feb; 109(6):1937–1942. doi: 10.1073/pnas.1114017109.

[80] **Bennett CH**. Efficient Estimation of Free Energy Differences from Monte Carlo Data. J Comput Phys. 1976; 22:245–268.

[81] **Shirts MR**, Bair E, Hooker G, Pande VS. Equilibrium Free Energies from Nonequilibrium Measurements Using Maximum-Likelihood Methods. Phys Rev Lett. 2003 Oct; 91(14). doi: 10.1103/PhysRevLett.91.140601.

[82] **Price DJ**, Jorgensen WL. Computational Binding Studies of Human Pp60c-Src SH2 Domain with a Series of Nonpeptide, Phosphophenyl-Containing Ligands. Bioorg Med Chem Lett. 2000 Sep; 10(18):2067–2070. doi: 10.1016/S0960-894X(00)00401-7.

[83] **Luccarelli J**, Michel J, Tirado-Rives J, Jorgensen WL. Effects of Water Placement on Predictions of Binding Affinities for P38α MAP Kinase Inhibitors. J Chem Theory Comput. 2010 Dec; 6(12):3850–3856. doi: 10.1021/ct100504h.

[84] **Michel J**, Verdonk ML, Essex JW. Protein-Ligand Binding Affinity Predictions by Implicit Solvent Simulations: A Tool for Lead Optimization? J Med Chem. 2006 Dec; 49(25):7427–7439. doi: 10.1021/jm061021s.

[85] **Barouch-Bentov R**, Sauer K. Mechanisms of drug resistance in kinases. Expert opinion on investigational drugs. 2011 Feb; 20(2):153–208.

[86] **McDermott U**, Sharma SV, Dowell L, Greninger P, Montagut C, Lamb J, Archibald H, Raudales R, Tam A, Lee D, Rothenberg SM, Supko JG, Sordella R, Ulkus LE, Iafrate AJ, Maheswaran S, Njauw CN, Tsao H, Drew L, Hanke JH, et al. Identification of genotype-correlated sensitivity to selective kinase inhibitors by using high-throughput tumor cell line profiling. Proceedings of the National Academy of Sciences of the United States of America. 2007 Dec; 104(50):19936–19941.

[87] **Salvatier J**, Wiecki TV, Fonnesbeck C. Probabilistic programming in Python using PyMC3. PeerJ Computer Science. 2016 Apr; 2:e55. https://doi.org/10.7717/peerj-cs.55, doi: 10.7717/peerj-cs.55.

[88] **Gruber FX**, Lundán T, Goll R, Silye A, Mikkola I, Rekvig OP, Knuutila S, Remes K, Gedde-Dahl T, Porkka K, Hjorth-Hansen H. BCR-ABL Isoforms Associated with Intrinsic or Acquired Resistance to Imatinib: More Heterogeneous than Just ABL Kinase Domain Point Mutations? Med Oncol. 2012 Mar; 29(1):219–226. doi: 10.1007/s12032-010-9781-z.

[89] **Redaelli S**, Mologni L, Rostagno R, Piazza R, Magistroni V, Ceccon M, Viltadi M, Flynn D, Passerini CG. Three novel patient-derived BCR/ABL mutants show different sensitivity to second and third generation tyrosine kinase inhibitors. American Journal of Hematology. 2012; 87(11):E125–E128. https://onlinelibrary.wiley.com/doi/abs/10.1002/ajh.23338, doi: 10.1002/ajh.23338.

[90] **Cortes JE**, Kantarjian H, Shah NP, Bixby D, Mauro MJ, Flinn I, O'Hare T, Hu S, Narasimhan NI, Rivera VM, Clackson T, Turner CD, Haluska FG, Druker BJ, Deininger MWN, Talpaz M. Ponatinib in Refractory Philadelphia Chromosome–Positive Leukemias. N Engl J Med. 2012 Nov; 367(22):2075–2088. doi: 10.1056/NEJMoa1205127.

[91] **Branford S**. 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 May; 99(9):3472–3475. doi: 10.1182/blood.V99.9.3472.

[92] **Press RD**, Willis SG, Laudadio J, Mauro MJ, Deininger MWN. Determining the rise in BCR-ABL RNA that optimally predicts a kinase domain mutation in patients with chronic myeloid leukemia on imatinib. Blood. 2009; 114(13):2598–2605. http://www.bloodjournal.org/content/114/13/2598, doi: 10.1182/blood-2008-08-173674.

Supplementary Information

TITLE

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

AUTHORS

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

AFFILIATIONS

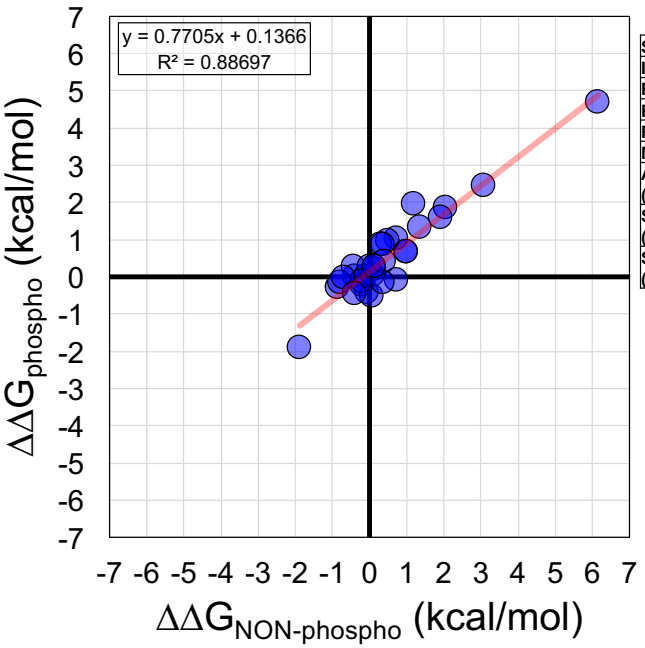
[1]Schrödinger, New York, NY 10036. [2]Gerstner Sloan Kettering Graduate School, Memorial Sloan Kettering Cancer Center, New York, NY 10065. [3]Computational and Systems Biology Program, Sloan Kettering Institute, Memorial Sloan Kettering Cancer Center, New York, NY 10065. [4]Schrödinger, GmbH, Q7 23, 68161 Mannheim, Germany.

CORRESPONDING AUTHOR

*Corresponding Author: *lingle.wang@schrodinger.com (LW)*

CONTENTS

- ***Figure S1***: Comparison of 31 mutations for which phosphorylated and non-phosphorylated ΔK_ds were available.
- ***Figure S2***: Truth tables with varying classification cutoffs for each TKI.
- ***Figure S3***: ROC curves for non-scaled and scaled FEP+, non-scaled and scaled Prime, a consensus model and a naïve model.
- ***Table S1***: IC_{50} experiment-derived reference ΔΔG data.
- ***Table S2***: Axitinib: experimental IC_{50} data and alchemical free-energy ΔΔGs.
- ***Table S3***: Bosutinib: experimental IC_{50} data and alchemical free-energy ΔΔGs.
- ***Table S4***: Dasatinib: experimental IC_{50} data and alchemical free-energy ΔΔGs.
- ***Table S5***: Imatinib: experimental IC_{50} data and alchemical free-energy ΔΔGs.
- ***Table S6***: Nilotinib: experimental IC_{50} data and alchemical free-energy ΔΔGs.
- ***Table S7***: Ponatinib: experimental IC_{50} data and alchemical free-energy ΔΔGs.
- ***Table S8***: Summary of statistics of scaled predictions, a naïve model, and a consensus model.
- ***Table S9***: Summary of the preparation of the 6 Abl:TKI co-crystal structure complexes.



Summary Statistics

Slope	0.77
Intercept (kcal/mol)	0.14
Pearson's R	0.94
R²	0.89
RMSE (kcal/mol)	0.50
MUE (kcal/mol)	0.39
Accuracy (cut=1.36 kcal/mol)	0.94
Specificity (cut=1.36 kcal/mol)	0.96
Sensitivity (cut=1.36 kcal/mol)	0.80

Raw Data (adapted de Davis et al.)

TKI	ddG_NON-P (kcal/mol)	ddG_phos (kcal/mol)
axitinib	2.05	1.85
	1.37	1.32
	-0.86	-0.27
	0.74	1.02
	-1.88	-1.89
bosutinib	0.99	0.69
	-0.05	-0.40
	-0.39	0.00
	-0.20	-0.23
	3.08	2.47
dasatinib	0.74	-0.07
	0.06	-0.53
	-0.09	0.00
	0.15	0.20
	6.16	4.69
imatinib	1.20	1.98
	0.49	0.99
	1.00	0.67
	0.29	0.88
	ND	**ND**
nilotinib	0.35	0.87
	0.11	0.04
	-0.43	0.29
	0.00	0.29
	2.50	**ND**
gefitinib	1.92	1.59
	0.39	0.41
	-0.78	-0.16
	-0.22	-0.13
	0.38	-0.15
erlotinib	**ND**	1.36
	0.12	0.29
	-0.70	-0.03
	-0.41	-0.44
	ND	**ND**

Figure S1. Comparison of 31 mutations for which phosphorylated and non-phosphorylated ΔK_ds were available. Scatter plot compares ΔΔGs (derived from the ΔK_ds) and contains the best-fit line with slope 0.77 and intercept 0.14. Summary statistics for this comparison are also shown. The raw ΔΔGs used for this comparison were adapted from [46]; kino-bead data for ponatinib was not available.

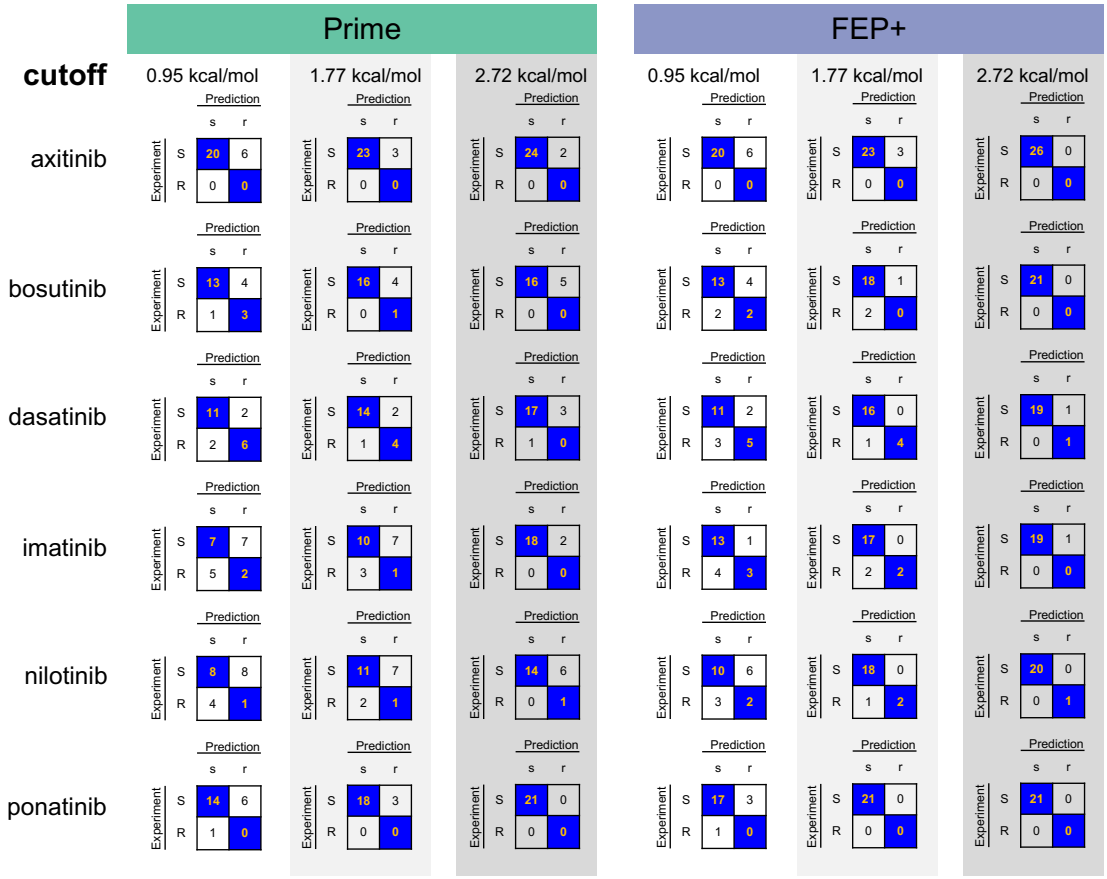


Figure S2. TKI-by-TKI truth tables with increasingly large classification cutoffs. Truth tables for the six TKIs (axitinib, bosutinib, dasatinib, imatinib, nilotinib, and ponatinib) using Prime (left, green) and FEP+ (right, blue) with classification cutoff values defining whether mutations are susceptible (S, experiment; s, prediction) or resistant (R, experiment; r, prediction). A mutation is susceptible if $\Delta\Delta G \leq$ **cutoff** or resistant if $\Delta\Delta G >$ **cutoff**.

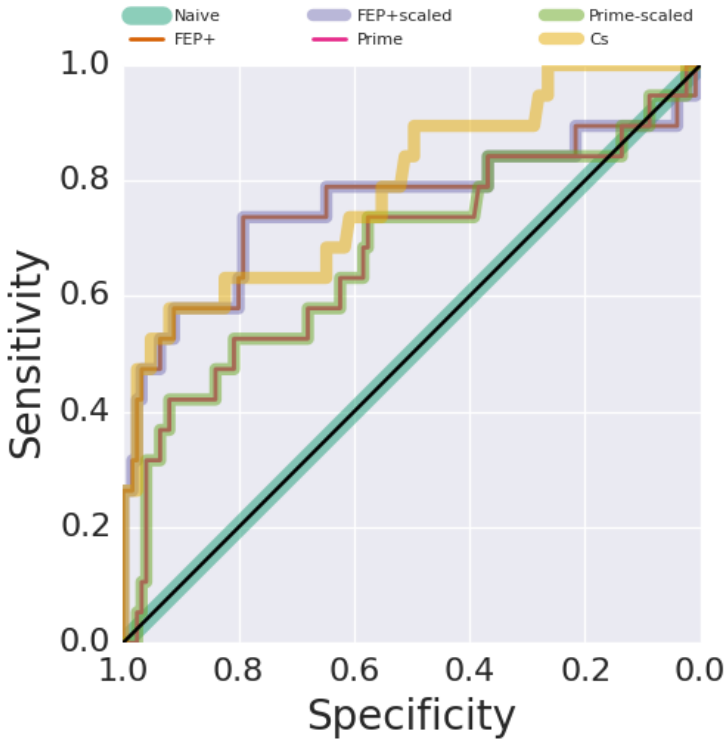


Figure S3. ROC curves for non-scaled and scaled FEP+, non-scaled and scaled Prime, a consensus model and a naïve model. ROC-AUC for scaled and non-scaled FEP+ was $0.75^{0.90}_{0.61}$ (n=144); ROC-AUC for scaled and non-scaled Prime was $0.66^{0.81}_{0.52}$ (n=144); ROC-AUCs for the naïve model and consensus model were $0.50^{0.50}_{0.50}$ (n=144) and $0.78^{0.90}_{0.67}$ (n=144) respectively. Optimal scaling factors (a=0.34 for FEP+; a=0.23 for Prime) obtained using linear regression (m=142) were applied to the full dataset (n=144), which was used in this ROC analysis. ROC-AUC interpretations: [0.50,0.60], failure; [0.60,0.70], poor; [0.70,0.80], fair; [0.80,0.90], good; [0.90,1.00], excellent.

Table S1. ΔΔG data derived from publicly available ΔpIC_{50} measurements and sources of mutation clinical-observation

Mutation	axitinib ΔΔG (kcal/mol)	bosutinib ΔΔG (kcal/mol)	dasatinib ΔΔG (kcal/mol)	imatinib ΔΔG (kcal/mol)	nilotinib ΔΔG (kcal/mol)	ponatinib ΔΔG (kcal/mol)	gefitinib ΔΔG (kcal/mol)	erlotinib ΔΔG (kcal/mol)	Source of Clinical-Observation
M244V	-0.11	0.43	0.00	0.21	-0.13	0.00	nd	nd	A
L248R	0.31	1.50	0.65	2.33	2.15	0.58	nd	nd	B
L248V	0.32	0.56	0.55	0.64	0.33	0.17	nd	nd	A,C
G250E	0.27	0.11	0.41	1.01	0.60	0.30	nd	nd	A,C,D
Q252H	0.20	nd	nd	nd	nd	nd	-0.44	-0.13	A
Y253F	0.26	-0.34	0.24	1.90	1.48	0.30	-0.17	0.00	C
Y253H	0.03	nd	nd	nd	nd	nd	nd	nd	A,C,D
E255K	0.26	0.56	0.90	1.50	1.27	0.41	-0.11	-0.11	A,C,D
E255V	0.30	0.66	1.02	2.22	2.36	1.00	nd	nd	A,C
D276G	0.18	nd	nd	nd	nd	nd	nd	nd	C
E279K	-0.03	nd	nd	nd	nd	nd	nd	nd	C
E292L	0.03	nd	nd	nd	nd	nd	nd	nd	E
V299L	-0.88	1.70	1.24	0.23	0.28	0.17	nd	nd	C
T315A	-0.45	0.32	2.02	0.51	0.72	0.17	nd	nd	C
T315I	-1.27	2.45	5.08	2.32	3.75	0.41	nd	-0.15	C,D
T315V	-1.73	nd	nd	nd	nd	nd	nd	nd	B
F317C	nd	0.50	1.86	0.28	0.04	0.00	nd	nd	A[g]
F317I	nd	0.71	1.79	0.17	0.30	0.51	1.35	1.58	C
F317L	0.23	0.09	0.96	0.72	0.20	0.17	0.29	0.40	C,D
F317R	0.27	nd	nd	nd	nd	nd	nd	nd	B
F317V	0.28	1.72	2.36	0.97	0.33	0.72	nd	nd	C
M343T	0.21	nd	nd	nd	nd	nd	nd	nd	F[h]
M351T	-0.24	0.19	0.00	0.42	0.00	0.17	0.05	-0.08	A,C,D
E355A	nd	0.02	0.24	0.47	0.11	0.51	nd	nd	C
F359C	nd	-0.01	0.00	0.77	0.68	0.41	nd	nd	C
F359I	0.10	0.04	0.24	0.28	0.86	0.77	nd	nd	A
F359V	0.07	-0.11	0.00	0.32	0.60	0.17	nd	nd	A,C
L384M	0.06	nd	nd	nd	nd	nd	nd	nd	F[i]
H396R	0.25	-0.10	0.00	0.40	0.25	0.17	nd	nd	A[j]
F486S	0.05	nd	nd	nd	nd	nd	nd	nd	A[k]
E459K	nd	0.35	0.41	0.66	0.55	0.30	nd	nd	C

A: Gruber et al. ([88])
B: Redaelli et al. ([89])
C: Cortes et al. ([90])
D: Branford et al. ([91])
E: Press et al. ([92])
F: Shah et al. ([3])
[g]: F317C observed with Δ27-183
[h]: M343T observed as compound mutation with H396R
[i]: L384M observed as compound mutation with M343T
[j]: H396R observed as compound mutation with F486S
[k]: F486S observed as compound mutation with H396R

Table S2. Axitinib: experimental IC_{50} values and alchemical free-energy ΔΔGs for each mutation.

	Expt. IC_{50} (nM)	Expt. ΔΔG (kcal/mol)	Prime ΔΔG (kcal/mol)	FEP+$_{Run1}$ ΔΔG (kcal/mol)	FEP+$_{Run1}$ BAR err (kcal/mol)	FEP+$_{Run2}$ ΔΔG (kcal/mol)	FEP+$_{Run2}$ BAR err (kcal/mol)	FEP+$_{Run3}$ ΔΔG (kcal/mol)	FEP+$_{Run3}$ BAR err (kcal/mol)	$\Delta\Delta G_{Av}$ (kcal/mol)	SE (kcal/mol)
wild-type	823										
M244V	690	-0.11	-0.10	-0.40	0.41	-0.35	0.41	-0.43	0.41	-0.39	0.02
L248R	1393	0.31	-0.06	2.13	0.43	2.42	0.45	2.46	0.43	2.34	0.10
L248V	1399	0.32	6.02	-1.32	0.41	-1.04	0.42	-1.22	0.42	-1.19	0.08
G250E	1295	0.27	0.31	-0.35	0.41	-0.71	0.41	-0.74	0.41	-0.60	0.13
Q252H	1155	0.20	-0.18	0.07	0.43	0.30	0.42	0.29	0.43	0.22	0.08
Y253F	1275	0.26	1.11	0.77	0.43	0.23	0.43	1.15	0.45	0.72	0.27
Y253H	867	0.03	4.65	1.14	0.47	0.38	0.49	-0.19	0.45	0.44	0.39
E255K	1282	0.26	0.12	1.30	0.44	0.63	0.43	1.10	0.44	1.01	0.20
E255V	1350	0.30	-0.29	0.98	0.42	1.04	0.42	1.26	0.43	1.09	0.09
D276G	1105	0.18	-0.01	0.03	0.42	0.64	0.42	0.44	0.43	0.37	0.18
E279K	778	-0.03	-0.15	0.06	0.42	-0.22	0.43	1.27	0.43	0.37	0.46
E292L	863	0.03	-0.00	0.53	0.43	0.35	0.42	0.31	0.42	0.40	0.07
V299L	188	-0.88	-5.00	-1.08	0.42	-1.39	0.42	-1.37	0.42	-1.28	0.10
T315A	389	-0.45	0.99	0.09	0.43	0.24	0.47	0.31	0.42	0.21	0.06
T315I	98	-1.27	-2.30	-1.26	0.42	-1.50	0.45	-1.39	0.43	-1.38	0.07
T315V	45	-1.73	-1.07	-1.10	0.41	-1.32	0.42	-1.15	0.48	-1.19	0.07
F317L	1220	0.23	1.29	-0.64	0.41	-0.10	0.41	-0.38	0.41	-0.37	0.16
F317R	1286	0.27	-2.46	2.64	0.46	2.27	0.51	1.38	0.47	2.10	0.37
F317V	1320	0.28	2.29	0.45	0.42	0.70	0.42	0.75	0.42	0.63	0.09
M343T	1175	0.21	-0.04	-0.26	0.54	-0.50	0.53	-0.58	0.50	-0.45	0.10
M351T	553	-0.24	-0.07	-0.25	0.41	-0.03	0.41	0.37	0.41	0.03	0.18
F359I	975	0.10	-0.04	1.89	0.41	1.60	0.42	1.78	0.41	1.76	0.08
F359V	933	0.07	-0.07	2.68	0.42	1.55	0.42	1.64	0.41	1.96	0.36
L384M	916	0.06	-0.01	-0.07	0.41	0.27	0.41	0.23	0.41	0.14	0.11
H396R	1247	0.25	-0.02	0.36	0.42	1.23	0.41	0.65	0.42	0.75	0.26
F486S	897	0.05	-0.09	0.65	0.47	1.14	0.46	0.44	0.48	0.74	0.21

BAR err: Bennett Acceptance Ratio error.

$\Delta\Delta\mathbf{G}_{Av}$: Average of three independent FEP+ runs.
SE: Standard Error between three independent FEP+ runs.

Table S3. Bosutinib: experimental IC_{50} values and alchemical free-energy ΔΔGs for each mutation.

	Expt. IC_{50} (nM)	Expt. ΔΔG (kcal/mol)	Prime ΔΔG (kcal/mol)	FEP+$_{Run1}$ ΔΔG (kcal/mol)	FEP+$_{Run1}$ BAR err (kcal/mol)	FEP+$_{Run2}$ ΔΔG (kcal/mol)	FEP+$_{Run2}$ BAR err (kcal/mol)	FEP+$_{Run3}$ ΔΔG (kcal/mol)	FEP+$_{Run3}$ BAR err (kcal/mol)	$\Delta\Delta G_{Av}$ (kcal/mol)	SE (kcal/mol)
wild-type	71										
M244V	147	0.43	0.02	-0.28	0.41	-0.11	0.41	-0.08	0.41	-0.16	0.06
L248R	874	1.50	3.67	1.00	0.43	1.63	0.43	1.33	0.43	1.32	0.18
L248V	182	0.56	5.77	0.37	0.41	0.72	0.42	0.38	0.42	0.49	0.12
G250E	85	0.11	-0.30	0.28	0.43	0.63	0.43	-1.07	0.43	-0.05	0.52
Y253F	40	-0.34	-0.03	0.21	0.45	0.02	0.43	0.95	0.43	0.39	0.28
E255K	181	0.56	0.49	-1.01	0.43	-1.30	0.43	-1.01	0.43	-1.11	0.10
E255V	214	0.66	0.11	-0.47	0.42	-0.51	0.43	-0.91	0.43	-0.63	0.14
V299L	1228	1.70	-0.85	0.97	0.43	0.90	0.42	0.85	0.42	0.91	0.03
T315A	122	0.32	1.00	-1.61	0.41	-1.61	0.41	-1.97	0.41	-1.73	0.12
T315I	4338	2.45	3.75	-2.32	0.43	-2.21	0.42	-1.26	0.42	-1.93	0.34
F317C	165	0.50	4.83	1.04	0.41	1.27	0.41	1.22	0.42	1.18	0.07
F317I	232	0.71	1.61	0.16	0.41	0.07	0.42	0.02	0.41	0.08	0.04
F317L	82	0.09	-0.71	0.05	0.41	0.47	0.41	0.24	0.41	0.25	0.12
F317V	1280	1.72	4.12	1.98	0.42	1.50	0.42	2.25	0.42	1.91	0.22
M351T	97	0.19	0.02	0.36	0.42	0.82	0.41	0.71	0.41	0.63	0.14
E355A	74	0.02	0.13	-0.20	0.44	0.13	0.43	0.27	0.43	0.07	0.14
F359C	70	-0.01	-0.09	3.02	0.42	2.51	0.42	1.97	0.43	2.50	0.30
F359I	76	0.04	-0.06	0.66	0.41	1.74	0.41	1.43	0.42	1.28	0.32
F359V	59	-0.11	-0.06	0.98	0.43	1.69	0.41	1.91	0.42	1.53	0.28
H396R	60	-0.10	-1.07	0.62	0.42	-0.07	0.42	-0.93	0.43	-0.13	0.45
E459K	127	0.35	0.26	-0.69	0.42	0.23	0.42	-0.54	0.42	-0.33	0.28

BAR err: Bennett Acceptance Ratio error.

$\Delta\Delta\mathbf{G}_{Av}$: Average of three independent FEP+ runs.
SE: Standard Error between three independent FEP+ runs.

Table S4. Dasatinib: experimental IC_{50} values and alchemical free-energy ΔΔGs for each mutation.

	Expt. IC_{50} (nM)	Expt. ΔΔG (kcal/mol)	Prime ΔΔG (kcal/mol)	FEP+$_{Run1}$ ΔΔG (kcal/mol)	FEP+$_{Run1}$ BAR err (kcal/mol)	FEP+$_{Run2}$ ΔΔG (kcal/mol)	FEP+$_{Run2}$ BAR err (kcal/mol)	FEP+$_{Run3}$ ΔΔG (kcal/mol)	FEP+$_{Run3}$ BAR err (kcal/mol)	$\Delta\Delta G_{Av}$ (kcal/mol)	SE (kcal/mol)
wild-type	2										
M244V	2	0.00	-0.10	0.05	0.41	-0.37	0.41	-0.43	0.41	-0.25	0.15
L248R	6	0.65	-2.13	1.40	0.42	1.50	0.43	1.51	0.42	1.47	0.04
L248V	5	0.55	2.60	0.58	0.42	0.70	0.41	0.79	0.41	0.69	0.06
G250E	4	0.41	-0.00	-0.54	0.43	-0.31	0.43	0.01	0.44	-0.28	0.16
Y253F	3	0.24	0.00	-0.21	0.43	-0.24	0.43	-0.03	0.44	-0.16	0.07
E255K	9	0.90	-0.08	-0.30	0.43	-0.17	0.44	-1.05	0.43	-0.51	0.27
E255V	11	1.02	-0.08	0.06	0.42	-0.80	0.42	-0.12	0.42	-0.29	0.26
V299L	16	1.24	0.01	0.83	0.41	0.36	0.42	0.77	0.42	0.65	0.15
T315A	59	2.02	5.09	-1.74	0.41	-1.65	0.41	-1.23	0.41	-1.54	0.16
T315I	10000	5.08	-2.69	5.63	0.43	4.69	0.44	5.50	0.43	5.27	0.29
F317C	45	1.86	4.72	2.63	0.42	2.32	0.42	2.62	0.41	2.52	0.10
F317I	40	1.79	2.38	1.94	0.41	2.04	0.41	1.94	0.41	1.97	0.03
F317L	10	0.96	1.22	1.26	0.41	1.42	0.41	1.08	0.41	1.25	0.10
F317V	104	2.36	4.08	3.12	0.42	2.84	0.42	2.68	0.42	2.88	0.13
M351T	2	0.00	0.04	0.04	0.41	0.14	0.41	0.00	0.42	0.06	0.04
E355A	3	0.24	0.00	-0.24	0.43	-0.87	0.45	-1.25	0.44	-0.79	0.29
F359C	2	0.00	-0.03	1.24	0.42	0.68	0.41	1.38	0.42	1.10	0.21
F359I	3	0.24	-0.02	-0.50	0.42	-0.33	0.42	-1.14	0.42	-0.66	0.25
F359V	2	0.00	-0.03	-0.87	0.41	0.57	0.42	-0.62	0.41	-0.31	0.44
H396R	2	0.00	2.53	-0.76	0.43	-0.09	0.43	-0.06	0.43	-0.30	0.23
E459K	4	0.41	0.00	-0.68	0.42	-0.17	0.42	-0.07	0.41	-0.31	0.19

T315I was beyond the concentration limit of the assay (10,000 nM).

BAR err: Bennett Acceptance Ratio error.
$\Delta\Delta\mathbf{G}_{Av}$: Average of three independent FEP+ runs.
SE: Standard Error between three independent FEP+ runs.

Table S5. Imatinib: experimental IC_{50} values and alchemical free-energy ΔΔGs for each mutation.

	Expt. IC_{50} (nM)	Expt. ΔΔG (kcal/mol)	Prime ΔΔG (kcal/mol)	FEP+$_{Run1}$ ΔΔG (kcal/mol)	FEP+$_{Run1}$ BAR err (kcal/mol)	FEP+$_{Run2}$ ΔΔG (kcal/mol)	FEP+$_{Run2}$ BAR err (kcal/mol)	FEP+$_{Run3}$ ΔΔG (kcal/mol)	FEP+$_{Run3}$ BAR err (kcal/mol)	$\Delta\Delta G_{Av}$ (kcal/mol)	SE (kcal/mol)
wild-type	201										
M244V	287	0.21	-0.08	0.15	0.41	0.43	0.41	0.17	0.41	0.25	0.09
L248R	10000	2.33	1.92	1.92	0.43	2.52	0.44	2.34	0.43	2.26	0.18
L248V	586	0.64	1.89	-1.04	0.41	-1.02	0.42	-1.20	0.41	-1.09	0.06
G250E	1087	1.01	0.92	0.16	0.41	0.02	0.41	0.12	0.41	0.10	0.04
Y253F	4908	1.90	-0.02	0.87	0.43	0.65	0.42	1.34	0.44	0.95	0.20
E255K	2487	1.50	0.25	-0.12	0.44	1.95	0.44	-0.55	0.44	0.43	0.77
E255V	8322	2.22	0.24	-0.72	0.42	-0.02	0.42	-0.53	0.43	-0.42	0.21
V299L	295	0.23	-1.29	0.66	0.41	0.26	0.42	-0.37	0.42	0.18	0.30
T315A	476	0.51	5.10	-1.39	0.41	-1.86	0.41	-2.09	0.44	-1.78	0.21
T315I	9773	2.32	0.88	4.23	0.43	4.23	0.42	3.14	0.44	3.87	0.36
F317C	324	0.28	2.10	0.27	0.42	-0.18	0.41	0.45	0.42	0.18	0.19
F317I	266	0.17	0.94	0.59	0.41	0.66	0.41	0.48	0.41	0.58	0.05
F317L	675	0.72	0.74	0.58	0.41	0.53	0.41	0.38	0.41	0.50	0.06
F317V	1023	0.97	1.57	0.71	0.42	0.79	0.42	0.80	0.41	0.77	0.03
M351T	404	0.42	-0.02	1.72	0.41	1.03	0.42	1.20	0.42	1.32	0.21
E355A	441	0.47	0.29	0.13	0.43	0.08	0.44	0.14	0.43	0.12	0.02
F359C	728	0.77	2.43	0.88	0.42	0.47	0.41	0.33	0.42	0.56	0.17
F359I	324	0.28	1.95	-0.13	0.41	-0.87	0.41	0.08	0.41	-0.31	0.29
F359V	346	0.32	2.53	-0.66	0.41	0.02	0.41	-0.27	0.42	-0.30	0.20
H396R	395	0.40	2.76	-0.39	0.41	-0.38	0.42	-0.39	0.42	-0.39	0.00
E459K	612	0.66	0.24	-0.09	0.43	-0.09	0.42	-0.08	0.42	-0.09	0.00

T315I was beyond the concentration limit of the assay (10,000 nM).

BAR err: Bennett Acceptance Ratio error.
$\Delta\Delta\mathbf{G}_{Av}$: Average of three independent FEP+ runs.
SE: Standard Error between three independent FEP+ runs.

Table S6. Nilotinib: experimental IC_{50} values and alchemical free-energy ΔΔGs for each mutation.

	Expt. IC_{50} (nM)	Expt. ΔΔG (kcal/mol)	Prime ΔΔG (kcal/mol)	FEP+$_{Run1}$ ΔΔG (kcal/mol)	FEP+$_{Run1}$ BAR err (kcal/mol)	FEP+$_{Run2}$ ΔΔG (kcal/mol)	FEP+$_{Run2}$ BAR err (kcal/mol)	FEP+$_{Run3}$ ΔΔG (kcal/mol)	FEP+$_{Run3}$ BAR err (kcal/mol)	$\Delta\Delta G_{Av}$ (kcal/mol)	SE (kcal/mol)
wild-type	15										
M244V	12	-0.13	-0.11	0.15	0.41	-0.21	0.41	0.21	0.41	0.05	0.13
L248R	549	2.15	0.48	2.05	0.43	2.12	0.47	1.93	0.43	2.03	0.06
L248V	26	0.33	3.53	-0.50	0.42	-0.39	0.41	-0.92	0.41	-0.60	0.16
G250E	41	0.60	0.05	0.06	0.41	-0.27	0.41	-0.38	0.41	-0.20	0.13
Y253F	179	1.48	-0.27	1.09	0.43	0.42	0.42	1.16	0.42	0.89	0.24
E255K	127	1.27	0.41	-2.24	0.48	-1.52	0.46	0.33	0.46	-1.14	0.77
E255V	784	2.36	-0.03	0.31	0.42	-0.25	0.43	-0.55	0.43	-0.16	0.25
V299L	24	0.28	2.94	-0.18	0.41	0.21	0.41	0.15	0.41	0.06	0.12
T315A	50	0.72	3.38	-1.33	0.41	-1.31	0.41	-1.39	0.41	-1.34	0.02
T315I	8091	3.75	4.16	4.29	0.43	5.00	0.42	4.34	0.43	4.54	0.23
F317C	16	0.04	0.90	1.34	0.41	0.88	0.41	0.60	0.41	0.94	0.22
F317I	25	0.30	-0.18	1.24	0.41	1.17	0.41	0.82	0.41	1.08	0.13
F317L	21	0.20	1.74	1.03	0.41	1.07	0.41	1.09	0.41	1.06	0.02
F317V	26	0.33	0.77	1.16	0.41	0.68	0.42	1.07	0.42	0.97	0.15
M351T	15	0.00	0.09	-0.06	0.41	-0.09	0.42	-0.46	0.42	-0.20	0.13
E355A	18	0.11	-0.06	-0.46	0.43	-1.01	0.43	-0.32	0.43	-0.60	0.21
F359C	47	0.68	3.68	1.32	0.41	1.44	0.41	1.52	0.41	1.43	0.06
F359I	64	0.86	3.70	1.05	0.41	1.13	0.41	0.74	0.41	0.97	0.12
F359V	41	0.60	3.67	1.00	0.41	1.08	0.41	1.38	0.42	1.15	0.12
H396R	23	0.25	2.58	-0.07	0.42	0.21	0.42	0.03	0.42	0.06	0.08
E459K	38	0.55	-0.00	-0.17	0.42	-0.46	0.42	-0.10	0.42	-0.24	0.11

BAR err: Bennett Acceptance Ratio error.

$\Delta\Delta\mathbf{G}_{Av}$: Average of three independent FEP+ runs.
SE: Standard Error between three independent FEP+ runs.

Table S7. Ponatinib: experimental IC_{50} values and alchemical free-energy ΔΔGs for each mutation.

	Expt. IC_{50} (nM)	Expt. ΔΔG (kcal/mol)	Prime ΔΔG (kcal/mol)	FEP+$_{Run1}$ ΔΔG (kcal/mol)	FEP+$_{Run1}$ BAR err (kcal/mol)	FEP+$_{Run2}$ ΔΔG (kcal/mol)	FEP+$_{Run2}$ BAR err (kcal/mol)	FEP+$_{Run3}$ ΔΔG (kcal/mol)	FEP+$_{Run3}$ BAR err (kcal/mol)	$\Delta\Delta G_{Av}$ (kcal/mol)	SE (kcal/mol)
wild-type	3										
M244V	3	0.00	-0.13	0.07	0.41	-0.28	0.41	0.12	0.41	-0.03	0.13
L248R	8	0.58	2.48	1.40	0.43	0.96	0.43	1.10	0.44	1.15	0.13
L248V	4	0.17	2.48	-1.82	0.42	-1.23	0.42	-1.96	0.42	-1.67	0.22
G250E	0.021	0.30	0.17	-0.32	0.43	-0.25	0.43	-0.71	0.46	-0.43	0.14
Y253F	5	0.30	0.05	0.85	0.43	1.32	0.44	0.77	0.43	0.98	0.17
E255K	6	0.41	1.05	-0.27	0.48	-0.66	0.48	0.03	0.47	-0.30	0.20
E255V	16	1.00	-0.04	1.19	0.43	0.94	0.43	-0.41	0.43	0.57	0.50
V299L	4	0.17	-0.29	-0.56	0.41	-0.55	0.41	-1.42	0.41	-0.84	0.29
T315A	4	0.17	-0.51	-2.90	0.41	-3.15	0.41	-2.92	0.41	-2.99	0.08
T315I	6	0.41	-5.42	0.51	0.42	0.90	0.42	0.91	0.42	0.77	0.13
F317C	3	0.00	1.45	0.44	0.41	0.98	0.42	0.80	0.41	0.74	0.16
F317I	7	0.51	0.62	-0.76	0.41	-1.03	0.41	-1.02	0.41	-0.94	0.09
F317L	4	0.17	0.57	-1.08	0.41	-0.83	0.41	-0.85	0.41	-0.92	0.08
F317V	10	0.72	1.14	0.05	0.41	-0.21	0.41	0.24	0.42	0.03	0.13
M351T	4	0.17	-0.12	0.89	0.41	1.66	0.41	0.65	0.41	1.07	0.30
E355A	7	0.51	0.01	0.12	0.44	-0.52	0.44	-0.55	0.43	-0.32	0.22
F359C	6	0.41	2.12	0.25	0.42	-0.35	0.43	0.73	0.42	0.21	0.31
F359I	11	0.77	0.34	-0.66	0.41	-0.38	0.41	0.06	0.41	-0.33	0.21
F359V	4	0.17	0.74	0.11	0.41	-0.28	0.41	0.08	0.42	-0.03	0.13
H396R	4	0.17	-0.04	0.19	0.49	0.10	0.45	-1.41	0.48	-0.37	0.52
E459K	5	0.30	-0.00	-0.51	0.42	-0.78	0.42	-0.63	0.42	-0.64	0.08

BAR err: Bennett Acceptance Ratio error.

$\Delta\Delta\mathbf{G}_{Av}$: Average of three independent FEP+ runs.
SE: Standard Error between three independent FEP+ runs.

Table S8. Summary of statistics of scaled predictions, a naïve model, and a consensus model.

Method	Scaling factor	MUE (kcal/mol) [N=142]	RMSE (kcal/mol) [N=142]	Accuracy [N=144]	Specificity [N=144]	Sensitivity [N=144]
Prime	1.00	$1.14^{1.35}_{0.94}$	$1.70^{1.97}_{1.40}$	$0.73^{0.80}_{0.65}$	$0.76^{0.83}_{0.68}$	$0.53^{0.78}_{0.29}$
Prime	0.50	$0.64^{0.76}_{0.53}$	$0.91^{1.06}_{0.77}$	$0.84^{0.90}_{0.78}$	$0.90^{0.95}_{0.84}$	$0.42^{0.65}_{0.20}$
Prime	0.33	$0.53^{0.62}_{0.44}$	$0.76^{0.87}_{0.63}$	$0.87^{0.92}_{0.81}$	$0.96^{0.99}_{0.92}$	$0.26^{0.47}_{0.08}$
Prime	0.23	$0.49^{0.59}_{0.40}$	$0.73^{0.86}_{0.60}$	$0.86^{0.92}_{0.81}$	$0.99^{1.00}_{0.97}$	$0.00^{0.00}_{0.00}$
FEP+	1.00	$0.79^{0.91}_{0.67}$	$1.07^{1.27}_{0.89}$	$0.88^{0.93}_{0.81}$	$0.94^{0.98}_{0.89}$	$0.47^{0.72}_{0.22}$
FEP+	0.34	$0.55^{0.64}_{0.47}$	$0.78^{0.91}_{0.65}$	$0.88^{0.93}_{0.83}$	$1.00^{1.00}_{1.00}$	$0.11^{0.27}_{0.00}$
Naive	—	$0.57^{0.69}_{0.46}$	$0.87^{1.04}_{0.70}$	$0.87^{0.92}_{0.81}$	$1.00^{1.00}_{1.00}$	$0.00^{0.00}_{0.00}$
Consensus	—	$0.47^{0.56}_{0.39}$	$0.71^{0.84}_{0.59}$	$0.87^{0.92}_{0.81}$	$1.00^{1.00}_{1.00}$	$0.00^{0.00}_{0.00}$

Table S9. Summary of the preparation of the 6 Abl:TKI co-crystal structure complexes.

				Experimental structure				Prepared model used for simulations							
PDB	**Receptor**	**Ligand**	**Chains**	**# Water**[a]	**# Rec. atoms,** (Chain) [b]	**# Aminos,** (Chain)	**Chain** used	**# Water**	**# Rec.** atoms	**# Rec.** aminos	**# Ash**	**# Glh**	**# Hip**	**# Lig.** atoms	**Het. atom**[d] w/ proton
4wa9	Abl	Axit	A, B	305	2219 (B)	276 (B)	B	131	4580	284	Ash421	0	0	46	neutral
3ue4	Abl	Bosut	A, B	152	2187 (A)	270 (A)	A	89	4581	284	Ash421 Ash381	0	0	66	NBI,4401
4xey	Abl	Dasat	A, B	0	2195 (A)	269 (A)	A	0	4581	284	Ash421[c] Ash381[c]	0	0	59	neutral
1opj	Abl	Imat	A, B	231	2336 (B)	288 (B)	B	104	4579	284	0	0	0	69	N51,4767
3cs9	Abl	Nilot	A, B, C, D	266	2142 (A)	264 (A)	A	99	4579	284	0	0	0	61	neutral
3oxz	Abl	Ponat	A	89	2152 (A)	268 (A)	A	89	4580	284	0	0	0	67	N3,2155

[a] Total number of water molecules.

[b] Count includes N-Acetyl/N-terminal (6 atoms) and N-methylamide/C-terminal (6 atoms) capping groups.
[c] Original index in experimental structure was Ash440, Ash400.
[d] (PDB atom name), (PDB serial).