ACS central science

# Predicting Kinase Inhibitor Resistance: Physics-Based and Data-Driven Approaches

Matteo Aldeghi,* Vytautas Gapsys, and Bert L. de Groot*

Computational Biomolecular Dynamics Group, Max Planck Institute for Biophysical Chemistry, Am Fassberg 11, 37077 Göttingen, Germany

S *Supporting Information*

ABSTRACT: Resistance to small molecule drugs often emerges in cancer cells, viruses, and bacteria as a result of the evolutionary pressure exerted by the therapy. Protein mutations that directly impair drug binding are frequently involved in resistance, and the ability to anticipate these mutations would be beneficial in drug development and clinical practice. Here, we evaluate the ability of three distinct computational methods to predict ligand binding affinity changes upon protein mutation for the cancer target Abl kinase. These structure-based approaches rely on first-principle statistical mechanics, mixed physics- and knowledge-based potentials, and machine learning, and were able to estimate binding affinity changes and identify resistant mutations with remarkable accuracy. We expect that these complementary approaches will enable the routine prediction of resistance-causing mutations in a variety of other target proteins.
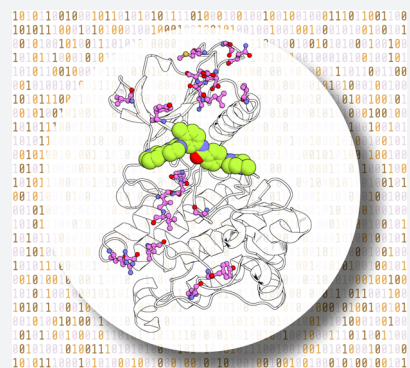
INTRODUCTION

Drug resistance is one of the chief challenges to be overcome in the development of robust anticancer and antimicrobial therapies. While resistance can emerge via multiple mechanisms, such as increased drug efflux or activation of alternative signaling pathways, it is often caused by protein mutations directly impacting drug binding.[1,2] Anticipating these resistance-causing mutations is of interest for personalized medicine, as it would inform treatment decisions based on the patient's genotype[3,4] and aid the development of combination therapies. It would also benefit drug development by allowing the parallel exploration of inhibitors with different resistance profiles. While large-scale experimentation is feasible,[5] it is neither cheap nor convenient, and accurate computer predictions mitigate the need for systematic experimentation.

Protein kinases are among the most exploited drug targets, with 48 inhibitors approved to date in the United States.[6] The majority of these inhibitors target tyrosine kinases,[6] which play a critical role in the modulation of growth factor signaling.[7,8] Hence, tyrosine kinase inhibitors (TKIs) are employed against a number of malignancies, like chronic myelogenous leukemia (CML) and nonsmall-cell lung cancer.[7,8] In particular, TKIs targeting the human kinase Abl are the first-line therapy for the treatment of CML.[9] However, susceptibility to resistance requires continued development of new-generation inhibitors.[8] For instance, in nonsmall-cell lung cancer, acquired resistance inevitably occurs within 1−2 years of starting the therapy.[10] In CML, more than 25% of patients switch TKI due to intolerance or resistance,[11] the latter being most often caused by mutations in Abl.[8] Because kinases display a long tail of rare and uncharacterized mutations,[12] the sensitivity of many clinically identified kinase mutants to TKI treatment is often

unknown. Thus, rapid and accurate computational approaches could impact clinical decision-making by providing oncologists with a first indication of whether the observed mutation is likely to cause resistance to certain inhibitors.

Here, we show how both physics-based and data-driven computational approaches can be used to accurately estimate the change in affinity of TKIs for the human kinase Abl caused by single-point mutations. To test the different methodologies, we used a data set of 144 Abl:TKI affinity changes ($\Delta\Delta G$) across 8 TKIs caused by 31 Abl mutations (Figure 1), which was compiled by Hauser et al.[12] using publicly available cell viability IC_{50} data. The computational approaches tested were based on (i) molecular dynamics (MD) simulations[13] with a nonequilibrium free energy calculation protocol;[14−16] (ii) Rosetta, a modeling program that uses mixed physics- and knowledge-based potentials;[17,18] and (iii) machine learning (ML).[19] For completeness, we discuss our results also in light of those obtained by Hauser et al.[12] with the OPLS3 force field[20] and a MD-based approach similar to ours (Table 1).

RESULTS AND DISCUSSION

The performance of all calculations was assessed by the root-mean-square error (RMSE) between calculated and experimentally measured $\Delta\Delta G$ values, the Pearson correlation coefficient (r), and the area under the precision recall curve (AUPRC).[21] The latter measures the ability to classify mutations as resistant or susceptible. In particular, precision measures which fraction of mutations classified as resistant are actually resistant, whereas recall measures which fraction of

a

31 Abl kinase mutations

b

8 inhibitors

Axitinib

Bosutinib

Dasatinib

Imatinib

Nilotinib

Ponatinib

Erlotinib

Gefitinib
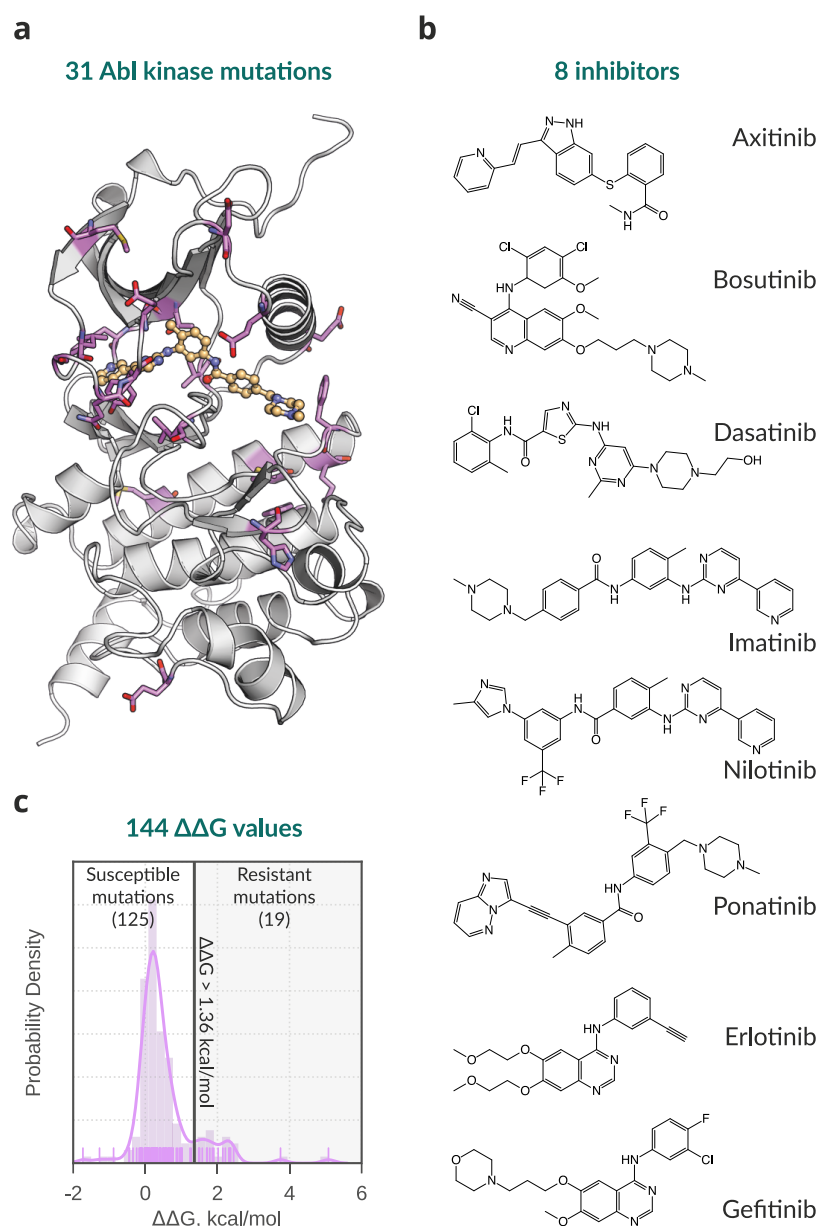
c

144 ΔΔG values

Figure 1. Data set of Abl kinase mutations and associated TKI affinity changes ($\Delta\Delta G$) studied. (a) Structure of human Abl kinase (PDB-ID 1OPJ) with imatinib (light orange) bound. Mutated wild-type residues are shown as violet sticks. (b) Name and chemical structure of the 8 TKIs studied. (c) Distribution of the 144 experimental $\Delta\Delta G$ values. The line at $\Delta\Delta G = 1.36$ kcal/mol separates mutations defined as resistant from susceptible.

resistant mutations are correctly classified as resistant. Consistent with previous work,[12] resistant mutations were defined as those causing more than a 10-fold drop in affinity ($\Delta\Delta G_{exp} > 1.36$ kcal/mol). On this data set, given the fraction of resistant mutations (Figure 1c), a random classifier would return an AUPRC of 0.13.

First, we estimated changes in TKI affinity with an approach based on the first-principles of statistical mechanics. MD simulations were used to sample the isothermal–isobaric ensemble while a specific residue was "alchemically" mutated[14] into another one. From the nonequilibrium work required to transform the residue, it is possible to recover the equilibrium free energy difference[22,23] and, using a suitable thermodynamic cycle (Figure S1), a rigorous estimate of $\Delta\Delta G$. We tested multiple force fields, and, on this data set, Amber force fields performed better than Charmm (Figure S2, Table S1); halogen

bond modeling did not improve the Charmm results (Figure S3). For simplicity, we discuss only the results obtained with the best-performing Amber (A99)[24−27] and Charmm (C22)[28−31] force fields (Table 1). The performance of A99 was comparable to that of OP3 (Figure 2),[12] which was obtained by Hauser et al.[12] with the FEP+ program by Schrödinger Inc. and the recently optimized proprietary OPLS3 force field.[20] C22 performed significantly worse than A99 and OP3 in terms of correlation and classification ability. Surprisingly, combining force fields in a consensus approach (considering half of the simulations for each parent force field) did not generally improve the results (Figure S4). It is conceivable that, on this data set, the benefits of the additional sampling obtained with one force field outweighs those expected by combining multiple force fields.[16] We also did not observe a dependence of the accuracy of the calculation on

Table 1. Summary of the Approaches Used, Their Performance, and Computational Cost[a]

			Approximate cost per $\Delta\Delta G$ estimate		Performance		
Abbreviation	Method	Force field or scoring function	Hardware	Compute hours	RMSE (kcal/mol)	Pearson	AUPRC
OP3	Molecular Dynamics	OPLS3[b]	1 GPU[c]	72[g]	$1.07_{0.89}^{1.25}$	$0.49_{0.19}^{0.69}$	$0.56_{0.32}^{0.76}$
C22	Molecular Dynamics	Charmm22* and CGenFF v 3.0.1	10 CPU[d] cores and 1 GPU[e]	59	$1.03_{0.85}^{1.21}$	$0.24_{0.01}^{0.44}$	$0.25_{0.12}^{0.48}$
A99	Molecular Dynamics	Amber99sb*-ILDN and GAFF v 2.1	10 CPU[d] cores and 1 GPU[e]	59	$0.91_{0.77}^{1.05}$	$0.44_{0.24}^{0.59}$	$0.56_{0.32}^{0.77}$
A99l	Molecular Dynamics	Amber99sb*-ILDN and GAFF v 2.1	10 CPU[d] cores and 1 GPU[e]	98[g]	$0.91_{0.74}^{1.09}$	$0.42_{0.20}^{0.59}$	$0.51_{0.27}^{0.75}$
R15	Rosetta	REF15	1 CPU[f] core	32	$0.72_{0.60}^{0.83}$	$0.67_{0.45}^{0.81}$	$0.53_{0.29}^{0.74}$
R16	Rosetta	βNOV16	1 CPU[f] core	32	$0.83_{0.70}^{0.96}$	$0.59_{0.35}^{0.74}$	$0.39_{0.18}^{0.60}$
ML1	Machine Learning	n/a	1 CPU[f] core	0.02	$0.87_{0.68}^{1.06}$	$0.12_{-0.04}^{0.29}$	$0.20_{0.10}^{0.39}$
ML2	Machine Learning	n/a	1 CPU[f] core	0.02	$0.68_{0.55}^{0.80}$	$0.57_{0.34}^{0.72}$	$0.47_{0.25}^{0.68}$

[a]For the performance measures, the point estimates from the original samples and their 95% bootstrapped confidence intervals are shown (x_{lower}^{upper}). RMSE: root-mean-square error; AUPRC: area under the precision-recall curve. [b]Data for the MD calculations with the OPLS3 force field refer to and were taken from Hauser et al.[12] [c]Nvidia (Pascal architecture). [d]Intel Xeon E5-2630 v 4. [e]Nvidia GeForce GTX 1080 Titan. [f]Intel Xeon (Broadwell architecture). [g]Time for charge-conserving mutations. For charge-changing mutations the simulation time is double.

the net charge change of the system upon mutation (Figure S5).

While our free energy protocol used a total of 216 ns of MD per $\Delta\Delta G$ estimate,[16] OP3 used 720 ns for charge-changing mutations, and 360 ns for charge-conserving ones.[12] We thus also tested a more expensive protocol (*A99l*), which matched exactly the simulation time invested by OP3. *A99l* improved upon the uncertainty (Figure S6), but not upon the accuracy (Figure 2), of the A99 estimates. Lower uncertainty is expected when more calculation repeats are used as in this case. On the other hand, accuracy is not necessarily proportional to the length of the simulations, as longer simulations allow for better sampling but also larger deviation from the experimentally determined structure in a potentially artificial fashion. Overall, this rigorous physics-based approach returned accurate $\Delta\Delta G$ estimates (e.g., for A99, RMSE = 0.91 kcal/mol, r = 0.44, and AUPRC = 0.56) and was able to discriminate between resistant and susceptible mutants well. However, a strong dependence of the results on the force field was observed. In addition, compared to the approaches discussed later, the MD based calculations required considerable computational resources (~2.5 days on 10 CPU cores and one GPU per $\Delta\Delta G$ estimate).

As an approach that is neither rigorously physics-based nor purely statistical, we tested a recently proposed Rosetta protocol[18] that was found to perform well also on protein–ligand binding.[16] This method samples multiple conformations of wild-type and mutant proteins with a Monte Carlo algorithm and estimates $\Delta\Delta G$ with the all-atom Rosetta energy function,[17] a mixed knowledge- and physics-based potential, by averaging over the generated wild-type and mutant ensembles. We found this approach (R15) to be highly accurate on this data set (Figure 2)—with low absolute errors (RMSE = 0.72 kcal/mol), strong correlation (r = 0.67), and good classification performance (AUPRC = 0.53)—while requiring only moderate computational resources (~32 h on a single CPU core per $\Delta\Delta G$ estimate). In addition to the standard REF15 scoring function, we tested the βNOV16 function (R16), which returned slightly worse results than R15 for this system (Figure 2). As had been noticed before,[16] here too, complementing Rosetta with the MD results via simple averaging generally resulted in enhanced performance (Figure

S7). For instance, combining A99 and R15, we obtained a particularly strong consensus estimate (Figure S8) with RMSE = 0.62 kcal/mol, r = 0.71, and AUPRC = 0.61. This can be attributed to the complementary performance of the two approaches: while the MD based calculations showed better precision performance and a lower number of false positives, Rosetta displayed better recall and a lower number of false negatives (Table S2, Figure 2).

As a purely data-driven approach, we employed extremely randomized regression trees,[19] a ML technique that uses ensembles of decision trees similarly to random forest. As input features, we used terms derived from the crystallographic protein–ligand structures (e.g., hydrogen bonds, nonpolar contacts, residue-ligand distance), ligand and residue physicochemical properties, and fast empirical scoring functions.[32−34] However, the way information from these sources is transformed to produce a $\Delta\Delta G$ estimate does not rely on any specific physical model. The technical details of the ML model construction are provided in the SI Methods. An advantage of this statistical approach is speed: once the necessary input features have been computed, a $\Delta\Delta G$ estimate can be obtained in seconds.

The ML model was trained and validated on a subset of the Platinum database,[35] and then tested on the TKI data set. The training/validation set contained 484 point-mutations with associated $\Delta\Delta G$ values across 84 proteins and 143 ligands. We computed a total of 128 features, and then used a greedy algorithm to select any number of features, up to 40, which minimized the mean-squared error of 10-fold cross-validation. The folds were built such that each of them would contain a unique set of proteins not present in the other folds. This procedure identified an optimal set of 29 features (Data S3). After training, the model was tested on the TKI data set (ML1 in Figure 2). Given no tyrosine kinase was present in the training set, we effectively assessed whether the model could extrapolate to this protein target. ML1 performed well in terms of RMSE (0.87 kcal/mol) but was not significantly better than random in terms of correlation and classification ability (r = $0.12_{-0.04}^{0.29}$, AUPRC = $0.20_{0.10}^{0.39}$), and was overall comparable to the MD results obtained with the Charmm force fields (Figure 2, Table S1). These results are similar to those obtained with
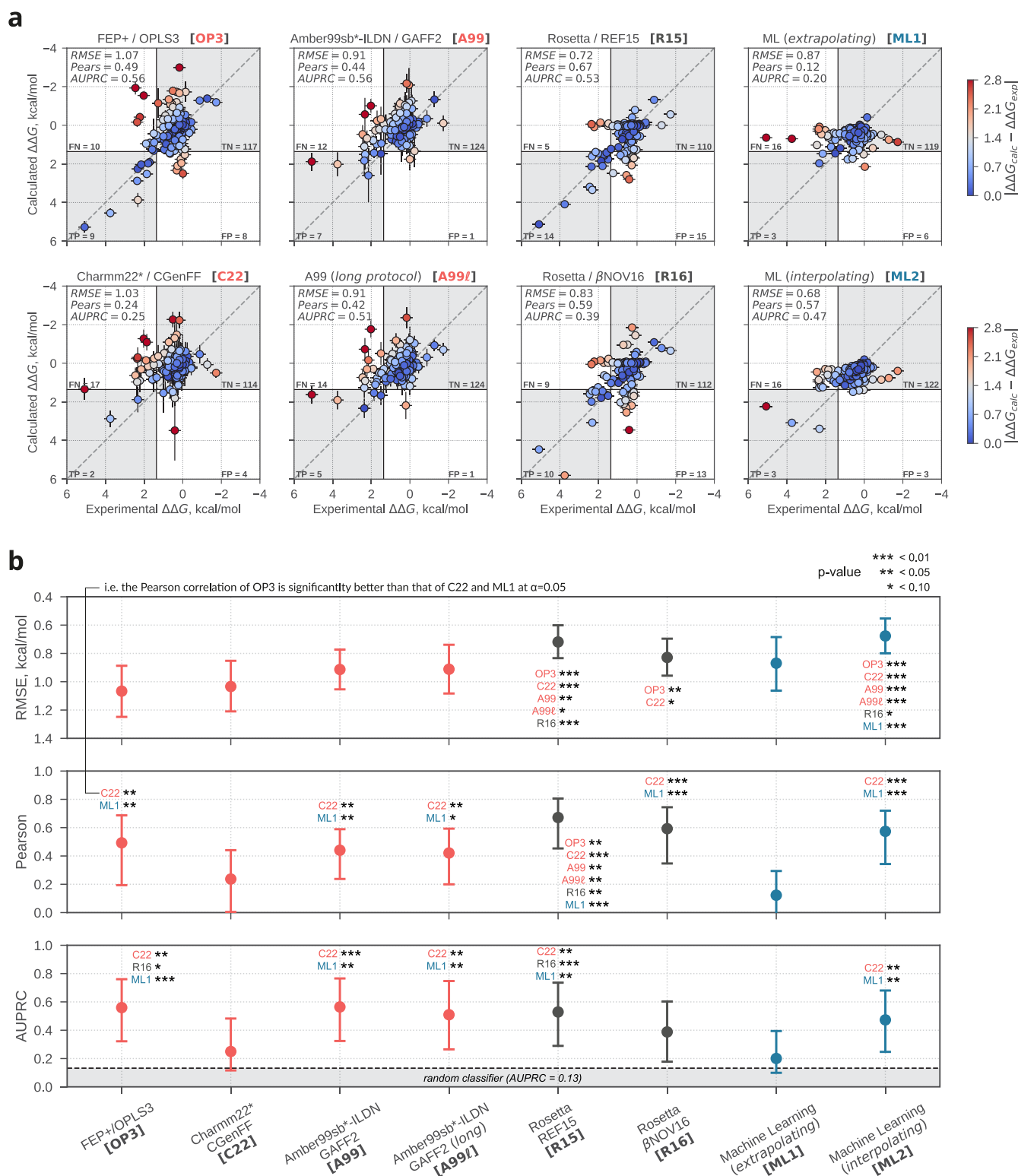
Figure 2. Accuracy of the ΔΔG estimates. (a) Scatter plots of experimental versus calculated ΔΔG values. The identity is shown as a dashed gray line. The four quadrants indicate the location of true positives (TP), true negatives (TN), false positives (FP), and false negatives (FN) according to the definition of resistant and susceptible mutations used (resistant if $\Delta\Delta G_{exp} > 1.36$ kcal/mol, susceptible otherwise)[12] and an equivalent cutoff (1.36 kcal/mol) for the calculated ΔΔG values. Each ΔΔG estimate is color-coded according to its absolute error with respect to the experimental ΔΔG value; at 300 K, an error of 1.4 kcal/mol corresponds to a ~10-fold error in K_d change, and an error of 2.8 kcal/mol to a ~100-fold error in K_d change. (b) Summary of the performance of the ΔΔG estimates across approaches in terms of RMSE, Pearson correlation, and AUPRC; point estimates from the original samples and 95% bootstrapped confidence intervals are shown (SI Methods). Differences at three levels of significance are reported using labels within the chart: e.g., a "C22**" label above the RMSE mark of OP3 indicates that the RMSE of OP3 is significantly lower (i.e., better agreement with experiment) than that of C22 at $\alpha = 0.05$.

another ML model[36] based on Gaussian process regression ([Figure S9]).

Given that ML is expected to perform best when trained on the most relevant data,[37] we asked whether the model performance would improve when trained on tyrosine kinase mutations only. In this case, we effectively assessed the interpolative ability of the model, as it was trained and tested on the TKI data set (ML2). To do this, we used 8-fold nested cross-validation. At each iteration, the model selected a subset of the 128 features via 7-fold cross validation with 7 inhibitors and was then tested on the eighth. This was done 8 times to obtain predictions for all TKIs and thus all entries in the data set. ML2 accurately estimated $\Delta\Delta G$ values and could discriminate well between resistant and susceptible mutations (RMSE = 0.68 kcal/mol, r = 0.57, AUPRC = 0.47).

If the objective is only to identify resistant mutations[38] rather than assessing their impact on ligand binding quantitatively, then AUPRC is the most relevant performance measure to examine. In this regard, MD based calculations with the OPLS3 or Amber99 force fields (OP3, A99, A99*l*), and Rosetta with the REF15 scoring function (R15) provided the best performance ([Figure 3], [Figure S10]). Furthermore, in
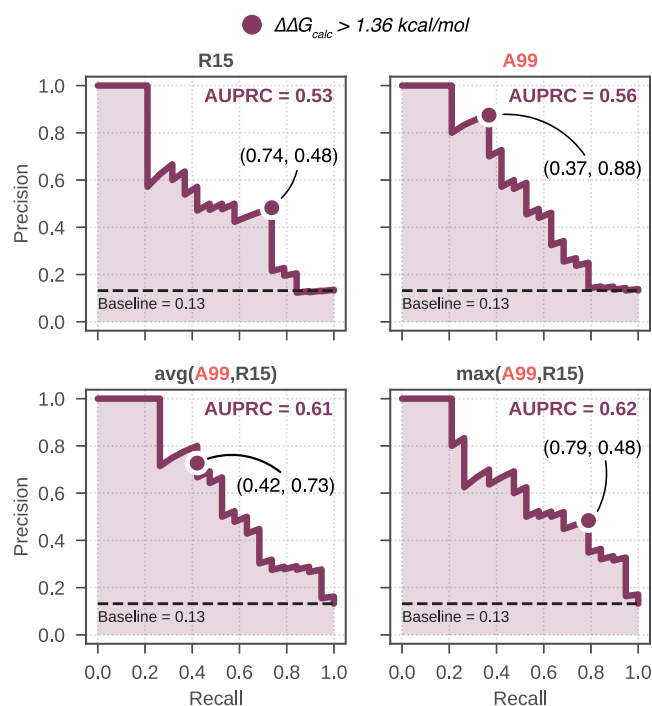

Figure 3. Precision recall curves for selected approaches. A99 and R15 have been combined to give two consensus results: in "avg(A99,R15)", the results of A99 and R15 have been averaged; in "max(A99,R15)", for each mutation, the most positive $\Delta\Delta G$ estimate among A99 and R15 was selected. The curve for a random estimator is shown as a dashed black line (baseline with AUPRC of 0.13). The precision and recall corresponding to the conventional threshold of $\Delta\Delta G_{calc}$ > 1.36 kcal/mol is reported and marked by a circle on the curves.

this scenario, one is likely willing to accept a larger fraction of false positives if this means recovering more resistant mutations. Thus, a suitable consensus strategy is to select the most positive $\Delta\Delta G$ estimate among those available to maximize the recall (at the expense of precision) at a given threshold ([Figure S11]). Using this approach with A99 and R15

returns an AUPRC of 0.62 and, at the conventional threshold of $\Delta\Delta G_{calc}$ > 1.36 kcal/mol, a recall of 0.79 and a precision of 0.48 ([Figure 3]; [Table S2]). This means that about half of the mutations classified as resistant are in fact resistant, and 79% of the resistant mutations in the data set have been identified.

From the results presented, it emerges that the latest computational approaches, from physics-based to data-driven ones, are able to predict TKI affinity changes upon Abl mutation and identify resistance-causing mutations. However, the different methods have different strengths and weaknesses. The MD-based calculations achieved a remarkable performance considering that force field parameters are based on fundamental physical properties of organic molecules (e.g., quantum mechanical energies and electrostatic potentials, liquid densities, and enthalpies of vaporization), rather than binding affinities directly. This, together with the rigorous statistical mechanical treatment, might lie at the basis of the demonstrated generalizability[16,39] of the approach across protein−ligand systems. The downside is the comparably higher computational cost required to obtain precise $\Delta\Delta G$ estimates. On the other hand, ML predictions proved accurate only when the binding affinities for Abl:TKI were used directly for training. With this information, however, a $\Delta\Delta G$ estimate could be generated in a matter of seconds to minutes on a single CPU core, with most of the time spent collecting the necessary structure-based features. Thus, the optimal approach choice in a prospective scenario will depend on the amount of prior information available for the system of interest, the nature of the study (a large-scale scan versus an in-depth study of a few mutations), as well as the computational resources at hand. That being said, we found that Rosetta struck a good balance between accuracy, generalizability, and computational cost. Similar to MD, this approach considers an ensemble of protein−ligand conformations,[18] and its scoring function is based on simple physical and structural properties of organic molecules and biomolecules,[17] so that it does not require binding affinities to the system of interest for training. In addition, approximations such as the use of implicit solvation and a more limited conformational sampling result in a moderate computational cost.

■ CONCLUSION

In summary, we have shown how fundamentally different structure-based approaches are able to predict TKI resistance to similar or better accuracy than reported thus far. Taking advantage of multiple techniques in a consensus fashion resulted in remarkable accuracy and tuning of the classification performance. These results suggest that a diverse set of computational approaches, each with its own strengths and weaknesses, is now available for the reliable estimate of resistance-causing protein mutations. It is our hope that the availability of multiple complementary approaches will enable the routine prediction of resistance-causing mutations across many other protein targets.

■ ASSOCIATED CONTENT

⑤ Supporting Information

The Supporting Information is available free of charge on the [ACS Publications website] at DOI: [10.1021/acscentsci.9b00590].

Methods, including supporting figures and data tables (PDF)

Data S1: detailed information on the data set and numerical results for all calculations (XLSX)

Data S2: input files pertaining to the molecular-dynamics-based free energy calculations and Rosetta calculations (ZIP)

Data S3: input files and *Jupyter* notebooks used for the training and testing of the machine learning model. Available at https://doi.org/10.5281/zenodo.3350897

■ AUTHOR INFORMATION

Corresponding Authors
*E-mail: maldegh@gwdg.de.
*E-mail: bgroot@gwdg.de.
ORCID
Matteo Aldeghi: 0000-0003-0019-8806
Vytautas Gapsys: 0000-0002-6761-7780
Bert L. de Groot: 0000-0003-3570-3534

Notes
The authors declare no competing financial interest.

■ ACKNOWLEDGMENTS

■ REFERENCES

(1) Lovly, C. M.; Shaw, A. T. Molecular Pathways: Resistance to Kinase Inhibitors and Implications for Therapeutic Strategies. *Clin. Cancer Res.* **2014**, *20* (9), 2249−2256.

(2) Housman, G.; Byler, S.; Heerboth, S.; Lapinska, K.; Longacre, M.; Snyder, N.; Sarkar, S. Drug Resistance in Cancer: An Overview. *Cancers* **2014**, *6* (3), 1769−1792.

(3) Zehir, A.; Benayed, R.; Shah, R. H.; Syed, A.; Middha, S.; Kim, H. R.; Srinivasan, P.; Gao, J.; Chakravarty, D.; Devlin, S. M.; Hellmann, M. D.; Barron, D. A.; Schram, A. M.; Hameed, M.; Dogan, S.; Ross, D. S.; Hechtman, J. F.; DeLair, D. F.; Yao, J. J.; Mandelker, D. L.; Cheng, D. T.; Chandramohan, R.; Mohanty, A. S.; Ptashkin, R. N.; Jayakumaran, G.; Prasad, M.; Syed, M. H.; Rema, A. B.; Liu, Z. Y.; Nafa, K.; Borsu, L.; Sadowska, J.; Casanova, J.; Bacares, R.; Kiecka, I. J.; Razumova, A.; Son, J. B.; Stewart, L.; Baldi, T.; Mullaney, K. A.; Al-Ahmadie, H.; Vakiani, E.; Abeshouse, A. A.; Penson, A. V.; Jonsson, P.; Camacho, N.; Chang, M. T.; Won, H. H.; Gross, B. E.; Kundra, R.; Heins, Z. J.; Chen, H. W.; Phillips, S.; Zhang, H.; Wang, J.; Ochoa, A.; Wills, J.; Eubank, M.; Thomas, S. B.; Gardos, S. M.; Reales, D. N.; Galle, J.; Durany, R.; Cambria, R.; Abida, W.; Cercek, A.; Feldman, D. R.; Gounder, M. M.; Hakimi, A. A.; Harding, J. J.; Iyer, G.; Janjigian, Y. Y.; Jordan, E. J.; Kelly, C. M.; Lowery, M. A.; Morris, L. G. T.; Omuro, A. M.; Raj, N.; Razavi, P.; Shoushtari, A. N.; Shukla, N.; Soumerai, T. E.; Varghese, A. M.; Yaeger, R.; Coleman, J.; Bochner, B.; Riely, G. J.; Saltz, L. B.; Scher, H. I.; Sabbatini, P. J.; Robson, M. E.; Klimstra, D. S.; Taylor, B. S.; Baselga, J.; Schultz, N.; Hyman, D. M.; Arcila, M. E.; Solit, D. B.; Ladanyi, M.; Berger, M. F. Mutational Landscape of Metastatic Cancer Revealed from Prospective Clinical Sequencing of 10,000 Patients. *Nat. Med.* **2017**, *23* (6), 703−713.

(4) Fowler, P. W.; Cole, K.; Gordon, N. C.; Kearns, A. M.; Llewelyn, M. J.; Peto, T. E. A.; Crook, D. W.; Walker, A. S. Robust Prediction of Resistance to Trimethoprim in Staphylococcus Aureus. *Cell Chem. Biol.* **2018**, *25* (3), 339−349.

(5) Melnikov, A.; Rogov, P.; Wang, L.; Gnirke, A.; Mikkelsen, T. S. Comprehensive Mutational Scanning of a Kinase in Vivo Reveals Substrate-Dependent Fitness Landscapes. *Nucleic Acids Res.* **2014**, *42* (14), No. e112.

(6) Roskoski, R. Properties of FDA-Approved Small Molecule Protein Kinase Inhibitors. *Pharmacol. Res.* **2019**, *144*, 19−50.

(7) Bhullar, K. S.; Lagarón, N. O.; McGowan, E. M.; Parmar, I.; Jha, A.; Hubbard, B. P.; Rupasinghe, H. P. V. Kinase-Targeted Cancer Therapies: Progress, Challenges and Future Directions. *Mol. Cancer* **2018**, *17* (1), 1−20.

(8) Arora, A.; Scholar, E. M. Role of Tyrosine Kinase Inhibitors in Cancer Therapy. *J. Pharmacol. Exp. Ther.* **2005**, *315* (3), 971−979.

(9) Jiao, Q.; Bi, L.; Ren, Y.; Song, S.; Wang, Q.; Wang, Y. S. Advances in Studies of Tyrosine Kinase Inhibitors and Their Acquired Resistance. *Mol. Cancer* **2018**, *17* (1), 1−12.

(10) Camidge, D. R.; Pao, W.; Sequist, L. V. Acquired Resistance to TKIs in Solid Tumours: Learning from Lung Cancer. *Nat. Rev. Clin. Oncol.* **2014**, *11*, 473.

(11) Patel, A. B.; O'Hare, T.; Deininger, M. W. Mechanisms of Resistance to ABL Kinase Inhibition in Chronic Myeloid Leukemia and the Development of Next Generation ABL Kinase Inhibitors. *Hematol. Oncol. Clin. North Am.* **2017**, *31* (4), 589−612.

(12) Hauser, K.; Negron, C.; Albanese, S. K.; Ray, S.; Steinbrecher, T.; Abel, R.; Chodera, J. D.; Wang, L. Predicting Resistance of Clinical Abl Mutations to Targeted Kinase Inhibitors Using Alchemical Free-Energy Calculations. *Commun. Biol.* **2018**, *1* (1), 70.

(13) Abraham, M. J.; Murtola, T.; Schulz, R.; Páll, S.; Smith, J. C.; Hess, B.; Lindahl, E. GROMACS: High Performance Molecular Simulations through Multi-Level Parallelism from Laptops to Supercomputers. *Software X* **2015**, *2*, 1−7.

(14) Gapsys, V.; Michielssens, S.; Seeliger, D.; de Groot, B. L. Pmx: Automated Protein Structure and Topology Generation for Alchemical Perturbations. *J. Comput. Chem.* **2015**, *36* (5), 348−354.

(15) Gapsys, V.; Michielssens, S.; Seeliger, D.; de Groot, B. L. Accurate and Rigorous Prediction of the Changes in Protein Free Energies in a Large-Scale Mutation Scan. *Angew. Chem., Int. Ed.* **2016**, *55* (26), 7364−7368.

(16) Aldeghi, M.; Gapsys, V.; de Groot, B. L. Accurate Estimation of Ligand Binding Affinity Changes upon Protein Mutation. *ACS Cent. Sci.* **2018**, *4* (12), 1708−1718.

(17) Alford, R. F.; Leaver-Fay, A.; Jeliazkov, J. R.; O'Meara, M. J.; DiMaio, F. P.; Park, H.; Shapovalov, M. V.; Renfrew, P. D.; Mulligan, V. K.; Kappel, K.; Labonte, J. W.; Pacella, M. S.; Bonneau, R.; Bradley, P.; Dunbrack, R. L.; Das, R.; Baker, D.; Kuhlman, B.; Kortemme, T.; Gray, J. J. The Rosetta All-Atom Energy Function for Macromolecular Modeling and Design. *J. Chem. Theory Comput.* **2017**, *13* (6), 3031−3048.

(18) Barlow, K. A.; Ó Conchúir, S.; Thompson, S.; Suresh, P.; Lucas, J. E.; Heinonen, M.; Kortemme, T. Flex DdG: Rosetta Ensemble-Based Estimation of Changes in Protein-Protein Binding Affinity upon Mutation. *J. Phys. Chem. B* **2018**, *122* (21), 5389−5399.

(19) Geurts, P.; Ernst, D.; Wehenkel, L. Extremely Randomized Trees. *Mach. Learn.* **2006**, *63* (1), 3−42.

(20) Harder, E.; Damm, W.; Maple, J.; Wu, C.; Reboul, M.; Xiang, J. Y.; Wang, L.; Lupyan, D.; Dahlgren, M. K.; Knight, J. L.; Kaus, J. W.; Cerutti, D. S.; Krilov, G.; Jorgensen, W. L.; Abel, R.; Friesner, R. A. OPLS3: A Force Field Providing Broad Coverage of Drug-like Small Molecules and Proteins. *J. Chem. Theory Comput.* **2016**, *12* (1), 281−296.

(21) Saito, T.; Rehmsmeier, M. The Precision-Recall Plot Is More Informative than the ROC Plot When Evaluating Binary Classifiers on Imbalanced Datasets. *PLoS One* **2015**, *10* (3), No. e0118432.

(22) Jarzynski, C. Equilibrium Free-Energy Differences from Nonequilibrium Measurements: A Master-Equation Approach. *Phys. Rev. E: Stat. Phys., Plasmas, Fluids, Relat. Interdiscip. Top.* **1997**, *56* (5), 5018−5035.

(23) Crooks, G. E. Entropy Production Fluctuation Theorem and the Nonequilibrium Work Relation for Free Energy Differences. *Phys. Rev. E: Stat. Phys., Plasmas, Fluids, Relat. Interdiscip. Top.* **1999**, *60* (3), 2721−2726.

(24) Hornak, V.; Abel, R.; Okur, A.; Strockbine, B.; Roitberg, A.; Simmerling, C. Comparison of Multiple Amber Force Fields and

Development of Improved Protein Backbone Parameters. *Proteins: Struct., Funct., Genet.* **2006**, *65* (3), 712−725.

(25) Best, R. B.; Hummer, G. Optimized Molecular Dynamics Force Fields Applied to the Helix-Coil Transition of Polypeptides. *J. Phys. Chem. B* **2009**, *113* (26), 9004−9015.

(26) Lindorff-Larsen, K.; Piana, S.; Palmo, K.; Maragakis, P.; Klepeis, J. L.; Dror, R. O.; Shaw, D. E. Improved Side-Chain Torsion Potentials for the Amber Ff99SB Protein Force Field. *Proteins: Struct., Funct., Genet.* **2010**, *78* (8), 1950−1958.

(27) Wang, J.; Wolf, R. M.; Caldwell, J. W.; Kollman, P. A.; Case, D. A. Development and Testing of a General Amber Force Field. *J. Comput. Chem.* **2004**, *25* (9), 1157−1174.

(28) MacKerell, a D.; Bashford, D.; Bellott, M.; Dunbrack, R. L.; Evanseck, J. D.; Field, M. J.; Fischer, S.; Gao, J.; Guo, H.; Ha, S.; Joseph-McCarthy, D.; Kuchnir, L.; Kuczera, K.; Lau, F. T.; Mattos, C.; Michnick, S.; Ngo, T.; Nguyen, D. T.; Prodhom, B.; Reiher, W. E.; Roux, B.; Schlenkrich, M.; Smith, J. C.; Stote, R.; Straub, J.; Watanabe, M.; Wiórkiewicz-Kuczera, J.; Yin, D.; Karplus, M. All-Atom Empirical Potential for Molecular Modeling and Dynamics Studies of Proteins. *J. Phys. Chem. B* **1998**, *102* (18), 3586−3616.

(29) Mackerell, A. D.; Feig, M.; Brooks, C. L. Extending the Treatment of Backbone Energetics in Protein Force Fields: Limitations of Gas-Phase Quantum Mechanics in Reproducing Protein Conformational Distributions in Molecular Dynamics Simulation. *J. Comput. Chem.* **2004**, *25* (11), 1400−1415.

(30) Piana, S.; Lindorff-Larsen, K.; Shaw, D. E. How Robust Are Protein Folding Simulations with Respect to Force Field Parameterization? *Biophys. J.* **2011**, *100* (9), L47−L49.

(31) Vanommeslaeghe, K.; Hatcher, E.; Acharya, C.; Kundu, S.; Zhong, S.; Shim, J.; Darian, E.; Guvench, O.; Lopes, P.; Vorobyov, I.; MacKerell, A. D. CHARMM General Force Field: A Force Field for Drug-like Molecules Compatible with the CHARMM All-Atom Additive Biological Force Fields. *J. Comput. Chem.* **2009**, *31* (4), 671−690.

(32) Salentin, S.; Schreiber, S.; Haupt, V. J.; Schroeder, M.; Adasme, M. F. PLIP: Fully Automated Protein-Ligand Interaction Profiler. *Nucleic Acids Res.* **2015**, *43* (W1), W443−W447.

(33) Trott, O.; Olson, A. J. AutoDock Vina: Improving the Speed and Accuracy of Docking with a New Scoring Function, Efficient Optimization, and Multithreading. *J. Comput. Chem.* **2009**, *31* (2), 455−461.

(34) Wang, C.; Zhang, Y. Improving Scoring-Docking-Screening Powers of Protein-Ligand Scoring Functions Using Random Forest. *J. Comput. Chem.* **2017**, *38* (3), 169−177.

(35) Pires, D. E. V.; Blundell, T. L.; Ascher, D. B. Platinum: A Database of Experimentally Measured Effects of Mutations on Structurally Defined Protein-Ligand Complexes. *Nucleic Acids Res.* **2015**, *43* (D1), D387−D391.

(36) Pires, D. E. V.; Blundell, T. L.; Ascher, D. B. MCSM-Lig: Quantifying the Effects of Mutations on Protein-Small Molecule Affinity in Genetic Disease and Emergence of Drug Resistance. *Sci. Rep.* **2016**, *6* (1), 29575.

(37) Ross, G. A.; Morris, G.; Biggin, P. One Size Does Not Fit All: The Limits of Structure-Based Models in Drug Discovery. *J. Chem. Theory Comput.* **2013**, *9*, 4266−4274.

(38) Carter, J.; Walker, T. M.; Walker, A. S.; Whitfield, M.; Peto, T. E. A.; Crook, D. W.; Fowler, P. W. Prediction of Pyrazinamide Resistance in Mycobacterium Tuberculosis Using Structure-Based Machine Learning Approaches. *Bio Rxiv* **2019**, 518142.

(39) Cournia, Z.; Allen, B.; Sherman, W. Relative Binding Free Energy Calculations in Drug Discovery: Recent Advances and Practical Considerations. *J. Chem. Inf. Model.* **2017**, *57* (12), 2911−2937.