Resistor: an algorithm for predicting resistance mutations using Pareto optimization over multistate protein design and mutational signatures

Nathan Guerin[1], **Teresa Kaserer**[2,*], **Bruce R. Donald**[1,3,4,5,*]

[1]*Department of Computer Science, Duke University, Durham, USA.* [2]*Institute of Pharmacy/Pharmaceutical Chemistry, University of Innsbruck, Austria.* [3]*Department of Biochemistry, Duke University Medical Center, USA.* [4]*Department of Chemistry, Duke University, USA.* [5]*Department of Mathematics, Duke University, USA.*

[*]*Corresponding authors,* `brd+recomb22@cs.duke.edu` *and* `teresa.kaserer@uibk.ac.at`

Abstract: Resistance to pharmacological treatments, such as to antibiotics or cancer therapeutics, ablates the efficacy of drugs and is a major public health challenge. Accurate, prospective prediction of resistance mutations would allow for design of drugs less susceptible to conferred resistance. Here we report RESISTOR—a novel structure- and sequence-based algorithm for predicting resistance mutations. RESISTOR optimizes over four objectives to find the Pareto frontier of these resistance-causing criteria. The first two axes of optimization are provable approximations to the relative change in binding affinity (ΔK_a) of a drug and endogenous ligand upon a protein's mutation. These ΔK_a predictions are made by the provable thermodynamic- and ensemble-based multistate computational protein design algorithm K^* after an initial sequence filter using the multistate design algorithm COMETS from the computational protein design software OSPREY. By virtue of pruning using COMETS, RESISTOR inherits the empirical sublinearlity characteristics of the COMETS sequence search, rendering RESISTOR, to our knowledge, the first provable structure-based resistance prediction algorithm that is sublinear in the size of the sequence space. This is important because the size of sequence space is exponential in the number of residue positions that can mutate to confer resistance. On the third axis RESISTOR uses empirically derived mutational signatures to determine the probability that a particular single- or double-point mutation will occur in a given context. The fourth axis of optimization is over "mutational hotspots", or those locations in the protein where multiple amino acids are predicted to confer resistance. As validation of the algorithm, we applied RESISTOR to a number of tyrosine kinase inhibitors used to treat lung adenocarcinoma and melanoma through inhibition of EGFR or B-Raf kinase activity. In so doing, we searched over a set of 1257 sequences in EGFR and 1214 sequences in B-Raf with an average conformation space size of $\sim 5.9 \times 10^{10}$, using experimental structures when available and docked complexes when not. This search generated a set of predicted resistance mutations that we compared to known resistance mutations in EGFR and report herein that RESISTOR correctly identified eight clinically significant resistance mutations, including the "gatekeeper" T790M mutation to erlotinib and gefitinib and five known resistance mutations to osimertinib. This demonstrates that by exploiting the wealth of structural and sequence data available in the form of molecular structures and mutational signatures, RESISTOR is a general method for predicting resistance mutations that can be applied to a wide variety of cancer, antimicrobial, antiviral and antifungal drug targets. RESISTOR is available as part of OSPREY on GitHub and is free and open source software.

Acknowledgements: We thank all members of the Donald lab for helpful discussions, the NIH (grants R01-GM078031, R01-GM118543, and R35-GM144042 to BRD), and the Austrian Science Fund (grant P34376-B to TK) for funding.

1. Introduction

Acquired resistance to therapeutics is a pressing public health challenge that affects maladies from bacterial and viral infections to cancer [1–6]. There are several different ways cancer cells acquire resistance to treatments, including drug inactivation, drug efflux, DNA damage repair, cell death inhibition, and escape mutations, among others [2]. While it is unlikely that medicinal chemists will be able to address all of the resistance-conferring mechanisms in cancer cells, progress can be made in each discrete area, leading to the development of more durable therapeutics. To that end, several structure-based computational techniques for therapeutic design and resistance prediction have been proposed.

One such technique is based on the substrate-envelope hypothesis. In short, the substrate-envelope hypothesis states that drugs designed to have the same interactions as the endogenous substrate in the active site will be unlikely to lose efficacy because any mutation that ablates binding to the drug would also ablate binding to the endogenous substrate [7]. C. Schiffer and B. Tidor's labs developed the substrate-envelope hypothesis for targeting drug-resistant HIV strains [7–10]. Their design technique has been successfully applied to develop compounds with reduced susceptibility to drug-resistant HIV proteases [10].

Another computational technique is to use ensemble-based positive and negative design [11]. There are two specific ways that point mutations can confer resistance to therapeutics: they can decrease binding affinity to the therapeutic or they can increase binding to the endogenous ligand [11, 12]. Protein design with the goal of decreasing binding is known as *negative design*, and increasing binding is known as *positive design*. As a concrete example, consider the case of a drug that inhibits the tyrosine kinase activity of the epidermal growth factor receptor (EGFR) to treat lung adenocarcinoma. Here, an active site mutation could sterically prevent the inhibitor from entering the active site [13]. On the other hand, a different mutation might have no effect on an enzyme's interactions with the drug, but instead accelerate the rate of EGFR's tyrosine kinase activity on its native substrates, resulting in the total phosphorylization of downstream tyrosines returning to its original, pre-inhibited, state [14, 15]. Because these two distinct pathways to therapeutic resistance exist, it is necessary to predict resistance mutations using both positive and negative design. In other words, predicting resistance can be reduced to predicting a ratio of the change in K_a upon mutation of the protein:endogenous ligand and protein:drug complexes.

K_a is an equilibrium constant measuring the binding and unbinding of a ligand to a receptor. It is defined as:

$$K_a = \frac{k_{\text{on}}}{k_{\text{off}}} = \frac{[RL]}{[R]\,[L]},$$

where k_{on} and k_{off} are the on- and off-rate constants, and $[RL]$, $[R]$, and $[L]$ the equilibrium concentrations of, respectively, the receptor-ligand complex, unbound receptor, and unbound ligand. K_a is the reciprocal of the disassociation constant K_d. K^* is an algorithm implemented in the OSPREY computational protein design software that provably approximates K_a [16, 17].

Our lab developed a provable, ensemble-based method using positive and negative K^* design to computationally predict and experimentally validate resistance mutations in protein targets [11]. We then applied this methodology to prospectively predict resistance mutations in dihydrofolate reductase when Staphylococcus aureus was treated with a novel antifolate [12] (also confirmed *in vivo* [12, 18]), demonstrating the utility of correctly predicting potential escape mutations during the drug discovery process. For cancer therapeutics, Kaserer and Blagg used OSPREY to combine positive and negative K^* design with mutational signatures and hotspot identification to both retrospectively and prospectively predict clinically significant resistance mutations [19]. Their technique combined sequence, in the form of trinucleotide mutational probabilities, and structures, in the form of positive and negative K^* design, to predict resistance mutations.

From these previous works, it is clear that multiple criteria must be combined to decide whether a mutation confers resistance. Often it is the human designers themselves who must choose arbitrary weights for different criteria. Yet multi-objective, or Pareto, optimization techniques would allow designers to com-

bine multiple criteria without choosing arbitrary decision thresholds. Pareto optimization for protein design has been employed by Chris Bailey-Kellogg, Karl Griswold, and co-workers [20–25]. One such example is PEPFR (Protein Engineering Pareto FRontier), which enumerates the entire Pareto frontier for a set of different criteria such as stability vs. diversity, affinity vs. specificity, and activity vs. immunogenicity [26]. Algorithmically, PEPFR combined divide-and-conquer with dynamic or integer programming to achieve an algorithm where the number of divide-and-conquer "divide" steps required for the search over design space is linear only in the number of Pareto optimal designs. To our knowledge, Pareto optimization has yet to be applied to predicting resistance mutations, making RESISTOR the first algorithm to employ Pareto optimization for predicting resistance mutations.

Instead of merely finding a single solution optimizing a linear combination of functions, Pareto optimization finds all consistent solutions optimizing multiple objectives such that no solution can be improved for one objective without making another objective worse. Specifically, let Λ be the set of possible solutions to the multi-objective optimization problem, and let $\lambda \in \Lambda$. Let $\mathscr{F}$ be a set of objective functions and $f \in \mathscr{F}$, where $f : \Lambda \to \mathbb{R}$ is one objective function. A particular solution λ is said to *dominate* another solution λ' when

$$f(\lambda) \leq f(\lambda') \text{ for all } f \in \mathscr{F}, \text{ and}$$
$$g(\lambda) < g(\lambda') \text{ for at least one } g \in \mathscr{F}.$$

A solution λ is *Pareto optimal* if it is not dominated. RESISTOR combines ensemble-based positive and negative design, cancer-specific mutational signature probabilities, and hotspots to identify not only the Pareto frontier, but also the Pareto ranks of all candidate sequences.

The inclusion of mutational signature probabilities in Pareto optimization is possible because distinct mutational processes cause different types of cancers [27, 28]. Specifically, these mutational processes drive the type and frequency of DNA base substitutions. Each different signature is postulated to be associated with a biological process (such as ABOPEC activity [28]) or a causative agent (such as tobacco use), although not all associations are definitively known. What is certain is that particular signatures tend to appear in particular types of cancer. For example, 12 single-base substitution signatures, 2 double-base substitution signatures, and 7 indel signatures were found in a large set of melanoma samples, with many of those signatures associated with ultraviolet light exposure [28]. Building on the work of Alexandrov et al. [27], Kaserer and Blagg combined the multiple signatures found in each cancer type to generate overall single-base substitution probabilities [19]. RESISTOR uses these probabilities to compute the overall probability that mutation events will occur in a gene independent of changes to protein fitness. This amino acid mutational probability is one of the axes we optimize over.

The most computationally complex part of provable, ensemble-based multistate design entails computing the K^* scores of the different design states. This is largely because for biological accuracy it is necessary to use K^* with continuous sidechain flexibility [29]. Though OSPREY has highly-optimized GPU routines for continuous flexibility [17], energy minimization over a combinatorial number of sequences in a continuous space is, in practice, computationally expensive. Having a method to reduce the number of sequences evaluated would greatly decrease the computational cost. COMETS is an empirically sublinear algorithm that provably returns the minimum of an arbitrary combination of multiple sequence states [30]. RESISTOR uses COMETS to prune sequences whose predicted binding with the drug improves and binding with the endogenous ligand deteriorates. While COMETS does not compute the full partition function, it provides a useful method to efficiently prune a combinatorial sequence space. This is particularly important when one considers resistant protein targets with more than one resistance mutation.

The epidermal growth factor receptor (EGFR) and serine/threonine-protein kinase B-Raf (B-Raf in short) are two oncogenes associated with, respectively, lung adenocarcinoma and melanoma. EGFR is a tyrosine kinase and B-Raf is a serine/threonine kinase. Both kinases are conformationally flexible, but two confor-

mations are particularly determinative to their kinase activity—the "active" and "inactive" conformations. Oncogenic mutations to EGFR include L858R and deletions in exon 19, both of which constitutively activate EGFR [31, 32]. Likewise, V600E is the most prevalent constitutively activating mutation in B-Raf [33]. Numerous drugs have been developed to treat the EGFR L858R and B-Raf V600E mutations. The first generation inhibitors erlotinib and gefitinib competitively inhibit ATP binding in EGFR's active site, whereas binding by the third generation osimertinib is irreversible [34–36]. For B-Raf, the therapeutics dabrafenib, vemurafenib, and encorafenib target the V600E mutation and are in clinical use, and PLX8394 is in clinical trials [37–40]. Use of RESISTOR to predict resistance mutations to these drugs would provide strong validation of the efficacy of this novel approach.

By presenting RESISTOR, this article makes the following contributions:

1. A novel multi-objective optimization algorithm that combines four axes of resistance-causing criteria to rank candidate mutations.
2. The use of COMETS as a provable and empirically sublinear pruning algorithm that removes a combinatorial number of candidate sequences before expensive ensemble evaluation.
3. A validation of RESISTOR that correctly predicted eight clinically significant resistance mutations in EGFR, providing explanatory ensemble-bound structural models for acquired resistance.
4. Prospective predictions and explanatory structural models of resistance mutations for four drugs targeting B-Raf mutations in melanoma.
5. An implementation of RESISTOR in our laboratory's free and open source computational protein design software OSPREY [17].

2. Methods

The Pareto optimization in RESISTOR optimizes over four axes: structure-based positive design, structure-based negative design, sequence-based mutational probabilities, and the count of resistance-causing mutations at a given amino acid location. The objective function maximizes the ΔK_a of the positive design (the protein bound to the endogenous ligand), minimizes the ΔK_a of the negative designs (the protein bound to the drug), maximizes the mutational probability, and maximizes the count of resistance-causing mutations per amino acid. Fig. 1 shows an overview how these axes are implemented in our algorithm.

2.1. Structure-based Positive and Negative Design

We use the K^* algorithm in OSPREY to predict an ε-accurate approximation to the binding affinity (K_a) in four states: 1) the wildtype structure bound to the endogenous ligand; 2) the wildtype structure bound to the therapeutic; 3) the mutated structure bound to the endogenous ligand; and 4) the mutated structure bound to the therapeutic. This ε-accurate approximation is called the K^* *score* [16, 17]. In order to calculate the K^* score of a protein:ligand complex, it is necessary to have a structural model of the atomic coordinates. Experimentally-determined complexes have been solved for EGFR bound to an analog of its endogenous ligand (PDB id 2itx), to erlotinib (1m17), gefitinib (4wkq), and to osimertinib (4zau) [41–44]. Similarly, we used the crystal structure for B-Raf bound to dabrafenib (4xv2) and vemurafenib (3og7) [45, 46]. Experimentally-determined complexes of B-Raf bound to encorafenib, PLX-8394, and an ATP analog do not exist, so we instead docked the inhibitors into B-Raf in its activated conformation (for additional details on model selection and preparation see appendix A). We used these predicted complex structures for our resistance predictions.

We added new functionality into OSPREY that simplifies the process of performing computational mutational scans. A *mutational scan* refers to the process of computing the K^* score of every possible amino acid mutation within a radius of a target. RESISTOR uses this functionality to calculate the four K^* scores for each amino acid within a 5 Å radius of the drug or the endogenous ligand. This generated a search space of 2471 sequences. We then set all residues with sidechains within 3 Å of the mutating residue to be

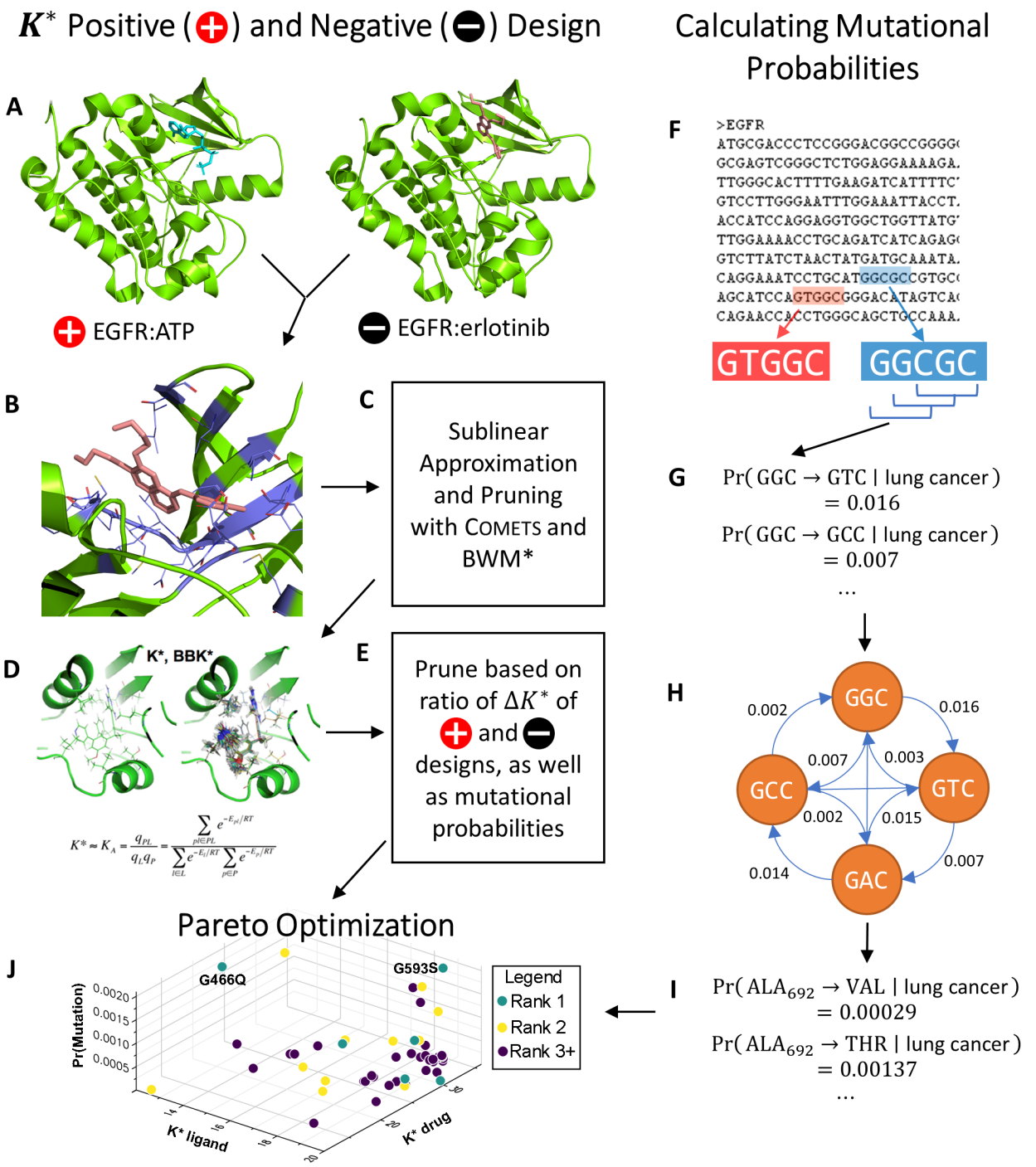


Fig. 1: **RESISTOR finds the Pareto frontier from OSPREY positive and negative designs, mutational probabilities, and resistance hotspots.** Schematic representation of how RESISTOR makes predictions. (A) Two structures are required as input to OSPREY to compute postive and negative design K^* scores. The structure for positive design is the protein (green) bound to the endogenous ligand (blue), for the negative design the protein is bound to the drug (pink). (B) All residues within 5 Å (purple) of the drug are allowed to mutate to any other amino acid. (C) COMETS is used as an efficient, sublinear algorithm to quickly prune infeasible mutations. BWM* is used with a fixed branch width to compute a polynomial-time approximation to the K^* score. (D) Candidate mutations that pass the COMETS pruning step have their positive and negative K^* scores computed in OSPREY. We recommend using the BBK* with MARK* algorithm as it is the fastest for computing K^* scores. (E) Candidate resistance mutations are pruned according to the criteria described in Sec. 2.4. (F) RESISTOR computes mutational probabilities using a protein's coding DNA along with cancer-specific trinucleotide mutational probabilities (G), sliding a window over 5′- and 3′-flanked codons. (H) RESISTOR employs a recursive graph algorithm to compute the probability that a particular amino acid will mutate to another amino acid (I). (J) Finally, RESISTOR uses Pareto optimization on the positive and negative K^* scores, the mutational probabilities, and hotspot counts to predict resistance mutants.

continuously flexible for the RESISTOR K^* designs. Each sequence has an associated conformation space size dependent on the total number of mutable and flexible residues, which one can use as a heuristic to estimate the difficulty of computing a complex's partition function. The average conformation space size of each sequence was $\sim 5.9 \times 10^{10}$ conformations, thus computing the partition functions is only possible using OSPREY's pruning and provable ε-approximation algorithms [17, 29, 47]. The change in the K^* score upon mutation for the endogenous ligand (positive design) and drug (negative design) become two of the four axes of optimization. These two axes also form the basis of a pruning step (described in Sec. 2.4).

2.2. *Computing the Probability of Amino Acid Mutations*

To convert the trinucleotide to trinucleotide probabilities into amino acid to amino acid mutational probabilities, RESISTOR constructs a directed graph with the trinucleotides as nodes and the probability that one trinucleotide mutates into another trinucleotide as directed edges. It then reads the cDNA of the protein in a sliding window of $5'$- and $3'$-flanked codons, since the two DNA bases flanking a codon are necessary to determine the probabilities of either the first or third base of a codon mutating. We designed a recursive algorithm to traverse the graph and find all codons that can be reached within n single-base mutations, where n is an input parameter. The algorithm then translates the target codons into amino acids and, as a final step, sums the different probabilities on each path to an amino acid into a single amino acid mutational probability (see Fig. 1F-I). One can either (a) precompute a cancer-specific codon-to-codon lookup table consisting of every $5'$- and $3'$-flanked codon to its corresponding amino acid mutational probabilities, or (b) read in a sequence's cDNA and compute the mutational probabilities on the fly. The benefit of (a) is it only needs to be done once per cancer type and can be used on an arbitrary number of sequences. On the other hand, when assigning mutational probabilities to proteins that have strictly fewer than 4^5 amino acids, it is faster to compute the amino-acid specific mutational signature on the fly. In both cases, the algorithm is strictly polynomial and bounded by $O(kn^9)$, where k is the number of codons with flanking base pairs (upper-bounded by 4^5) and n is the number of mutational steps allowed, which in the case of RESISTOR is 2. An implementation of this algorithm is included in the free and open source OSPREY GitHub repository [17].

2.3. *Identifying Mutational Hotspots*

After calculating the positive and negative change in affinity ΔK_a and determining the mutational probability of each amino acid, RESISTOR prunes the set of candidate mutations (see section 2.4). Post-pruning, it counts the number of mutations at each amino acid location. This count is important to determine whether a residue location is likely to become a "mutational hotspot", namely a residue location where many mutations are predicted to confer resistance. Correctly identifying mutational hotspots is vital because they indicate that a drug is dependent on the wildtype identity of the amino acid at that location, and it is likely that many mutations away from that amino acid will cause resistance. Consequently, the fourth axis used in RESISTOR's Pareto optimization is the count of predicted resistance-conferring mutations per residue location.

2.4. *Reducing the Positive Prediction Space*

Prior to carrying out the multi-objective optimization to identify predicted resistance mutations, we prune the set of candidates. First, we introduce a cut-off based on the ratio of K^* scores of positive and negative designs, adapted from Kaserer and Blagg's 2018 cut-off [19]. We determine the average of the K^* scores for the drug and endogenous ligand across all of the wildtype designs for the same protein. The cut-off c is:

$$c = \frac{c_0 K^*_{\overline{L}}}{K^*_{\overline{D}}},$$

where c_0 is a user-specified constant, $K^*_{\overline{L}}$ is the average of the K^* scores for the wildtype protein bound to

the endogenous ligand, and $K^*_{\overline{D}}$ is the average of the K^* score for the wildtype protein bound to the drug. We recommend in practice to set c_0 to be greater than the range $\left(K^*_{\max} - K^*_{\min}\right)$ of wildtype K^* scores—we set it to 100 for the tyrosine kinase inhibitor (TKI) predictions.[1] A mutation m is predicted to be *resistant* when:

$$\frac{K^*_L(m)}{K^*_D(m)} > c,$$

where $K^*_L(m)$ is the K^* score of the endogenous ligand bound to the mutant, and $K^*_D(m)$ is the K^* score of the drug bound to the mutant.

We further prune the predicted resistance mutation candidates by removing all mutations that cannot arise within two DNA base substitutions. Whether an amino acid can be reached within two DNA base substitutions is determined by the algorithm described in section 2.2, and if it cannot, then that particular mutation is assigned a mutational probability of 0 and pruned.

3. Results

3.1. RESISTOR *Identifies 8 Known Resistance Mutations in EGFR*

We evaluated a total of 1257 sequences across the three TKIs for EGFR. Among these sequences, the average conformation space size for computing a complex's partition function was $\sim 1.3 \times 10^7$. After we ran the RESISTOR algorithm on these sequences, 108 mutants were predicted as resistance-conferring candidates from a purely thermodynamic and probabilistic basis. We then computed the Pareto frontier over the four axes for each drug. Remarkably, RESISTOR correctly predicted eight clinically significant resistance mutants, with 7 of the 8 in the Pareto frontier and the remaining mutant in the 2nd Pareto rank (see Table 1).

3.1.1. EGFR Treated with Erlotinib and Gefitinib

Of the 462 sequences evaluated for the TKI erlotinib, RESISTOR identified 50 as candidate resistance mutations. Pareto ranking placed 19 sequences on the frontier, 13 sequences in the second rank, and 11, 6, and 1 sequences in the third, fourth, and fifth ranks, respectively. RESISTOR correctly identified two clinically significant mutations, T790M and G796D, as being on the Pareto frontier [48, 49]. This is encouraging because T790M is, by far, the most prevalent resistance mutation that occurs in lung adenocarcinoma treated with erlotinib [56]. Similarly, for gefitinib, RESISTOR evaluated 438 sequences and identified 22 as candidate resistance mutants. The most troublesome clinically significant mutant, T790M, is found on the Pareto frontier.

3.1.2. EGFR and Osimertinib

RESISTOR evaluated 357 OSPREY-predicted structures of EGFR bound with osimertinib and EGFR bound with its endogenous ligand. Of those, 36 were predicted as resistance candidates. Pareto optimization placed 16 sequences on the frontier, 2 sequences in rank 2, 8 sequences in rank 3, 1 sequence in rank 4, and 5 sequences in rank 5. RESISTOR correctly identified five clinically significant resistance mutations to osimertinib: L792H, G796R, G796S, G796D, and G796C [50–55], and while L792H was in the 2nd Pareto rank, all of the other correctly predicted resistance mutations are on the Pareto frontier.

Two osimertinib resistance mutations in particular stand out: L792H and G796D (see Fig. 2). Both of these mutants have appeared in the clinic [50–52, 55]. OSPREY generated an ensemble of the bound positive and negative complexes upon mutation, providing an explanatory model for how resistance occurs. In both cases, the mutant sidechains are much bulkier than the wildtype sidechain (Fig. 2A and D) and thus are predicted to clash with the original osimertinib binding pose (Fig. 2B and E). Consequently, in both cases the ligand

[1] In the future, c_0 could be learned from running RESISTOR on a resistance mutation dataset for homologous systems and examining the K^* scores.

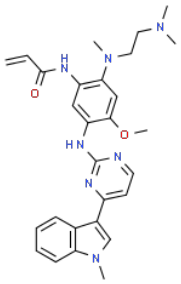
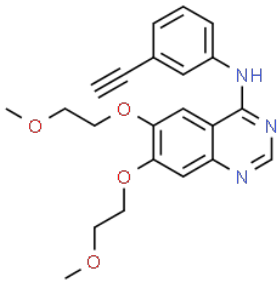
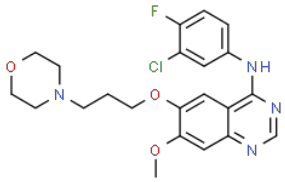
RESISTOR Identifies Clinically-Relevant Resistance Mutations in EGFR		
Osimertinib	Erlotinib	Gefitinib
		
L792H# G796R†# G796S# G796D# G796C#	T790M†# G796D#	T790M †#

Table 1: **RESISTOR correctly identified 8 resistance mutations in EGFR to erlotinib, gefitinib, and osimertinib.** For osimertinib, G796R, G796S, G796D, and G796C were on the RESISTOR-identified Pareto frontier. L792H was in the 2nd Pareto rank. For erlotinib, both T790M and G796D were on the Pareto frontier. For gefitinib, T790M was also on the Pareto frontier. Previous studies have documented all of these resistance mutations as occurring in the clinic [48–55]. † indicates that RESISTOR predicted the mechanism of resistance to be improved binding of the endogenous ligand to the mutant. # indicates that RESISTOR predicted the mechanism of resistance to be decreased binding of the drug to the mutant. Note that these predicted mechanisms are only attributed here if the predicted change in the $\log_{10}(\Delta K^*) \geq 0.5$.

is predicted to translate and rotate to create additional space for the mutant sidechains (Fig. 2C and F). We hypothesize that this movement weakens the other molecular interactions osimertinib makes in the EGFR active site.

In the case of G796D, there are additional factors that contribute to acquired resistance. First, the mutation to aspartate introduces a negative charge, which probably leads to electrostatic repulsion with the carbonyl oxygen of the osimertinib amide (Fig. 2F, highlighted with a dashed oval). In addition, the exit vector of the hydrogen bound to the amide nitrogen does not allow a hydrogen bond with the aspartate. Second, the allyl-group of osimertinib must be in close proximity to C797 for covalent bond formation. In fact, C797 is so important to osimertinib's efficacy that mutations at residue 797 confer resistance [57, 58]. Even if osimertinib still binds to G796D, the allyl group would have to move away from C797 (Fig. 2F, highlighted with a black arrow). This would prevent covalent bond formation and thus reduce the efficacy of osimertinib considerably. Lastly, it is likely that the mutation away from glycine reduces the conformational flexibility of the loop, incurring an entropic penalty while also plausibly making it more difficult to properly align osimertinib and C797.

3.2. RESISTOR *Predicts New Resistance Mutations in B-Raf and Provides Structural Models*

In addition to retrospective validation by comparison to existing clinical data for EGFR, we used RESISTOR to predict how mutations in the B-Raf active site could confer resistance. RESISTOR identified a number of prospective resistance mutations to vemurafenib, dabrafenib, encorafenib, and PLX8394. On the Pareto frontier for vemurafenib are 13 mutations, for dabrafenib 16 mutations, for encorafenib 15 mutations, and for PLX8394 15 mutations. The full set of predictions are included in appendix table C. Experimental data on resistance caused by escape mutations in B-Raf is sparse, but one existing source is from Wagenaar et al, who

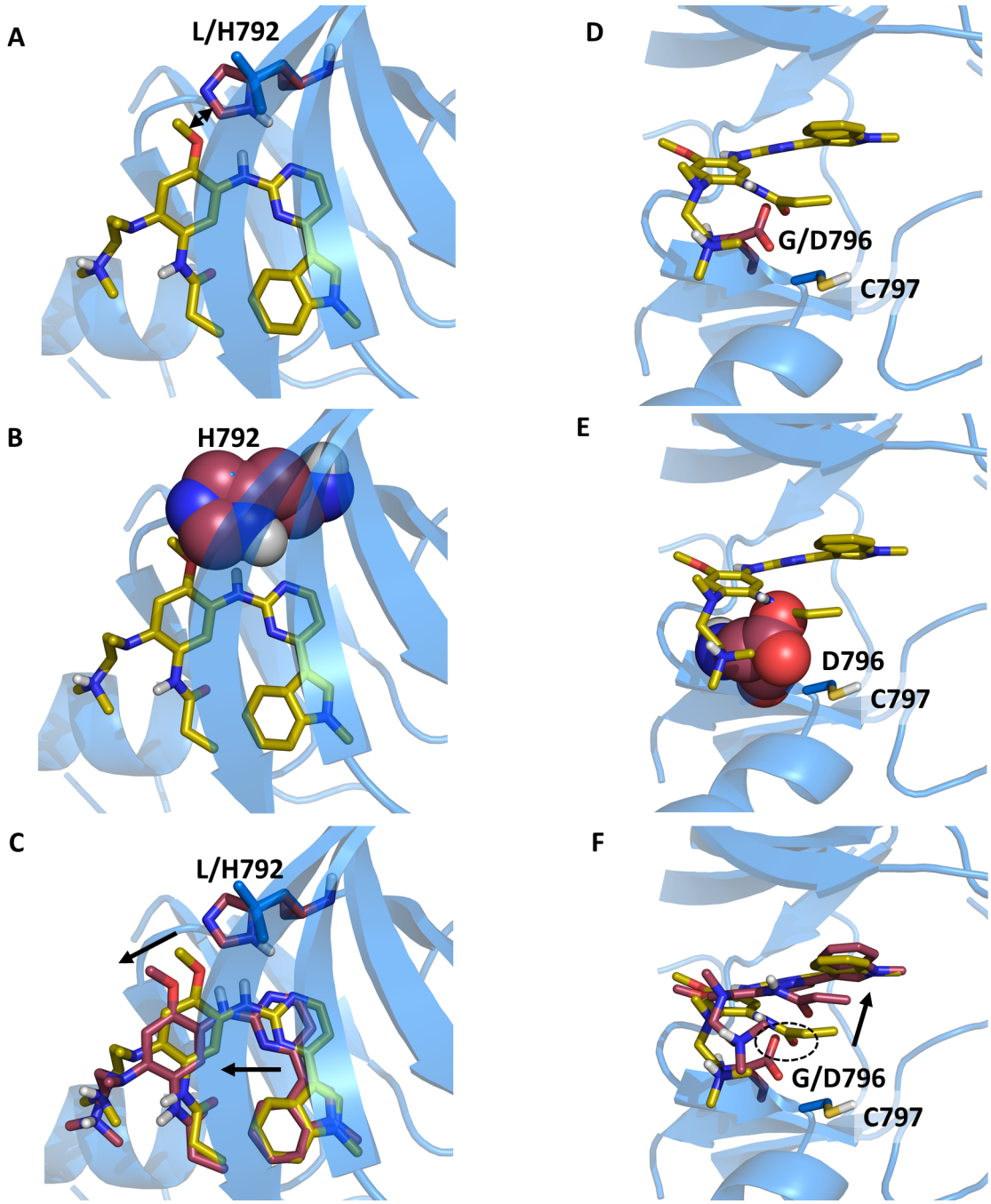


Fig. 2: **Structural models predicted by OSPREY agree with experimental data and explain mechanisms of resistance.** Structural models predicted by OSPREY of EGFR wildtype (blue) and resistance mutations (red) bound to osimertinib (yellow sticks). The histidine (A) and glutamate (D) side chains (red sticks) in the EGFR L792H (A) and G796D (D) mutations are bulkier than the wildtype leucine (A) and glycine (C) residues (blue sticks). They clash with osimertinib in its original binding pose as highlighted by the sphere representation in panels B and E. (C+F) To allow for accommodation of osimertinib in the modelled EGFR mutant structures (red sticks), the inhibitor's position within the binding pocket moves from the experimentally determined binding pose (yellow sticks). Movements are indicated by black arrows. (F) In case of the G796D mutation, the carboxylate moiety of D796 is predicted to be in close proximity to the osimertinib amide oxygen (highlighted with the dashed circle), thus leading to electrostatic repulsion. This mutation site is adjacent to C797, which reacts with the allyl-moiety of osimertinib to form a covalent bond in the wildtype. Due to the steric and electrostatic properties of the G796D mutant, the allyl group is located further away from C797 in the model, thus preventing covalent bond formation. The movement of the allyl group is indicated by the black arrow.

carried out saturation mutagenesis on the active site in A375 cells and measured IC_{50} values for vemurafenib with a number of different mutants [59]. Two of RESISTOR's predictions agree with the IC_{50} data: T529M and T529L. Furthermore, RESISTOR also provided structural ensembles demonstrating how these mutations are predicted to confer resistance. This is useful because there are no empirical structures of B-Raf with the V600E and the T529M/L mutations. Predictions about cancer systems for which we lack experimental data, such as B-Raf with dabrafenib, encorafenib, and PLX-8394, could be further tested either via binding assays such as biolayer interferometry and isothermal titration calorimetry or via saturation mutagenesis with corresponding IC_{50} data. Completing these wet lab experiments will provide further validation that would be interesting future work.

3.3. Complexity

There are a number of distinct steps in RESISTOR, each of which has its own complexity. While there are sublinear K^* algorithms, such as BBK* [60] with MARK* [47], these algorithms so far have only been applied to positive and negative design with optimization of specific multiple objectives, such as minimizing/maximizing the bound (respectively unbound) state partition functions and their ratios for computing binding affinity or stability. COMETS [30] provably does multistate design optimizing arbitrary constrained linear combinations of GMEC energies, but COMETS does not model the partition functions required for calculating binding affinity. A provable ensemble-based algorithm analogous to COMETS for arbitrary multistate design optimization is yet to be developed. Thus, general multistate K^* design remains, unfortunately, a problem linear in the number of sequences and thus exponential in the number of mutable residues.

Computing K^* itself, as a ratio of partition functions built from the thermodynamic ensembles of the bound to unbound states, can be expensive [61–63]. In order to reduce the number of K^* problems to solve, COMETS is employed as a pruning mechanism for all sequences in which there are more than one mutation. Without COMETS, RESISTOR would need to compute sN K^* scores, where s is the number of states and N is the number of sequences. With COMETS, RESISTOR is able to avoid computing many of these K^* scores, as COMETS has been shown in practice to reduce the number of required GMEC calculations by over 99% and to reduce N for continuous designs by 96%, yielding an overall speedup of over 5×10^5-fold [30]. Since in this study we considered only single residue mutations we omitted the COMETS pruning step, but in any use of RESISTOR that considers multiple simultaneously mutable residues we believe COMETS' empirical sublinearity will make the difference between feasible and infeasible searches.

Moreover, by using an approximation containing fixed partition function size and sparse residue interaction graphs, we can use the BWM* algorithm [64] to compute the K^* scores in time $O(nw^2q^{\frac{3}{2}w} + kn\log q)$, where w is the branch-width and q the number of rotamers per residue. When we have $w = O(1)$ this is polynomial time. In this study we found that the ε-approximation algorithms using adaptively-sized partition functions, such as BBK* with MARK*, were fast enough. However, for larger problems the sparse approximations allow us to approximate the necessary K^* scores for resistance prediction in time exponential only in the branch-width, and thus polynomial time for fixed branch-widths.

4. Conclusions

RESISTOR fills an important void in the science of predicting resistance mutations by providing an algorithm to enumerate the entire Pareto frontier of multiple resistance-causing criteria. It is, to our knowledge, the first algorithm to apply Pareto optimization to predicting resistance mutations. By categorizing predicted resistance mutations by their Pareto rank, it allows the drug discovery community to prioritize escape mutations on the Pareto frontier. RESISTOR also provides structural justification for the mechanism of each predicted escape mutation by generating an ensemble of predicted structural models upon mutation. In this study, we have applied RESISTOR to predict resistance mutations in EGFR and B-Raf for a number of different therapeutics. While RESISTOR as designed herein optimizes over 4 objectives, as a general method

any number of diverse objectives could be added. It can be applied not only to cancer therapeutics, but also to antimicrobial or antiviral drug design. Thus RESISTOR provides an important tool to the drug discovery community to design drugs less prone to resistance.

References

1. Centers for Disease Control and Prevention. *Antibiotic / Antimicrobial Resistance* July 2020. https://www.cdc.gov/drugresistance/index.html.
2. Housman, G., Byler, S., Heerboth, S., Lapinska, K., Longacre, M., Snyder, N. & Sarkar, S. Drug resistance in cancer: an overview. *Cancers* **6,** 1769–1792 (2014).
3. Zahreddine, H. & Borden, K. Mechanisms and insights into drug resistance in cancer. *Frontiers in pharmacology* **4,** 28 (2013).
4. Assaraf, Y. G., Brozovic, A., Goncalves, A. C., Jurkovicova, D., Linē, A., Machuqueiro, M., Saponara, S., Sarmento-Ribeiro, A. B., Xavier, C. P. & Vasconcelos, M. H. The multi-factorial nature of clinical multidrug resistance in cancer. *Drug Resistance Updates* **46,** 100645 (2019).
5. Gupta, R. K., Jordan, M. R., Sultan, B. J., Hill, A., Davis, D. H., Gregson, J., Sawyer, A. W., Hamers, R. L., Ndembi, N., Pillay, D., *et al.* Global trends in antiretroviral resistance in treatment-naive individuals with HIV after rollout of antiretroviral treatment in resource-limited settings: a global collaborative study and meta-regression analysis. *The Lancet* **380,** 1250–1258 (2012).
6. Vasan, N., Baselga, J. & Hyman, D. M. A view on drug resistance in cancer. *Nature* **575,** 299–309 (2019).
7. Altman, M. D., Ali, A., Kumar Reddy, G. K., Nalam, M. N., Anjum, S. G., Cao, H., Chellappan, S., Kairys, V., Fernandes, M. X., Gilson, M. K., Schiffer, C. A., Rana, T. M. & Tidor, B. HIV-1 protease inhibitors from inverse design in the substrate envelope exhibit subnanomolar binding to drug-resistant variants. *Journal of the American Chemical Society* **130,** 6099–6113 (2008).
8. Prabu-Jeyabalan, M., Nalivaika, E. & Schiffer, C. A. Substrate shape determines specificity of recognition for HIV-1 protease: analysis of crystal structures of six substrate complexes. *Structure* **10,** 369–381 (2002).
9. King, N. M., Prabu-Jeyabalan, M., Nalivaika, E. A., Wigerinck, P., De Béthune, M.-P. & Schiffer, C. A. Structural and thermodynamic basis for the binding of TMC114, a next-generation human immunodeficiency virus type 1 protease inhibitor. *Journal of virology* **78,** 12012–12021 (2004).
10. Shen, Y., Altman, M. D., Ali, A., Nalam, M. N., Cao, H., Rana, T. M., Schiffer, C. A. & Tidor, B. Testing the substrate-envelope hypothesis with designed pairs of compounds. *ACS chemical biology* **8,** 2433–2441 (2013).
11. Frey, K. M., Georgiev, I., Donald, B. R. & Anderson, A. C. Predicting resistance mutations using protein design algorithms. *Proceedings of the National Academy of Sciences* **107,** 13707–13712 (2010).
12. Reeve, S. M., Gainza, P., Frey, K. M., Georgiev, I., Donald, B. R. & Anderson, A. C. Protein design algorithms predict viable resistance to an experimental antifolate. *Proceedings of the National Academy of Sciences* **112,** 749–754 (2015).
13. Yan, X.-E., Zhu, S.-J., Liang, L., Zhao, P., Choi, H. G. & Yun, C.-H. Structural basis of mutant-selectivity and drug-resistance related to CO-1686. *Oncotarget* **8,** 53508 (2017).
14. Yun, C.-H., Mengwasser, K. E., Toms, A. V., Woo, M. S., Greulich, H., Wong, K.-K., Meyerson, M. & Eck, M. J. The T790M mutation in EGFR kinase causes drug resistance by increasing the affinity for ATP. *Proceedings of the National Academy of Sciences* **105,** 2070–2075 (2008).

15. Yoshikawa, S., Kukimoto-Niino, M., Parker, L., Handa, N., Terada, T., Fujimoto, T., Terazawa, Y., Wakiyama, M., Sato, M., Sano, S., *et al.* Structural basis for the altered drug sensitivities of non-small cell lung cancer-associated mutants of human epidermal growth factor receptor. *Oncogene* **32,** 27–38 (2013).

16. Georgiev, I., Lilien, R. H. & Donald, B. R. The minimized dead-end elimination criterion and its application to protein redesign in a hybrid scoring and search algorithm for computing partition functions over molecular ensembles. *Journal of computational chemistry* **29,** 1527–1542 (2008).

17. Hallen, M. A., Martin, J. W., Ojewole, A., Jou, J. D., Lowegard, A. U., Frenkel, M. S., Gainza, P., Nisonoff, H. M., Mukund, A., Wang, S., Holt, G. T., Zhou, D., Dowd, E. & Donald, B. R. OSPREY 3.0: Open-source protein redesign for you, with powerful new features. *Journal of computational chemistry* **39,** 2494–2507 (2018).

18. Reeve, S. M., Scocchera, E. W., Narendran, G., Keshipeddy, S., Krucinska, J., Hajian, B., Ferreira, J., Nailor, M., Aeschlimann, J., Wright, D. L. & Anderson, A. C. MRSA isolates from united states hospitals carry dfrg and dfrk resistance genes and succumb to propargyl-linked antifolates. *Cell chemical biology* **23,** 1458–1467 (2016).

19. Kaserer, T. & Blagg, J. Combining mutational signatures, clonal fitness, and drug affinity to define drug-specific resistance mutations in cancer. *Cell chemical biology* **25,** 1359–1371 (2018).

20. Parker, A. S., Choi, Y., Griswold, K. E. & Bailey-Kellogg, C. Structure-guided deimmunization of therapeutic proteins. *Journal of Computational Biology* **20,** 152–165 (2013).

21. Choi, Y., Griswold, K. E. & Bailey-Kellogg, C. Structure-based redesign of proteins for minimal T-cell epitope content. *Journal of computational chemistry* **34,** 879–891 (2013).

22. Salvat, R. S., Parker, A. S., Choi, Y., Bailey-Kellogg, C. & Griswold, K. E. Mapping the Pareto optimal design space for a functionally deimmunized biotherapeutic candidate. *PLoS Comput Biol* **11,** e1003988 (2015).

23. Griswold, K. E. & Bailey-Kellogg, C. Design and engineering of deimmunized biotherapeutics. *Current opinion in structural biology* **39,** 79–88 (2016).

24. Choi, Y., Ndong, C., Griswold, K. E. & Bailey-Kellogg, C. Computationally driven antibody engineering enables simultaneous humanization and thermostabilization. *Protein Engineering, Design and Selection* **29,** 419–426 (2016).

25. Salvat, R. S., Verma, D., Parker, A. S., Kirsch, J. R., Brooks, S. A., Bailey-Kellogg, C. & Griswold, K. E. Computationally optimized deimmunization libraries yield highly mutated enzymes with low immunogenicity and enhanced activity. *Proceedings of the National Academy of Sciences* **114,** E5085–E5093 (2017).

26. He, L., Friedman, A. M. & Bailey-Kellogg, C. A divide-and-conquer approach to determine the Pareto frontier for optimization of protein engineering experiments. *Proteins: Structure, Function, and Bioinformatics* **80,** 790–806 (2012).

27. Alexandrov, L. B., Nik-Zainal, S., Wedge, D. C., Aparicio, S. A., Behjati, S., Biankin, A. V., Bignell, G. R., Bolli, N., Borg, A., Børresen-Dale, A.-L., *et al.* Signatures of mutational processes in human cancer. *Nature* **500,** 415–421 (2013).

28. Alexandrov, L. B., Kim, J., Haradhvala, N. J., Huang, M. N., Ng, A. W. T., Wu, Y., Boot, A., Covington, K. R., Gordenin, D. A., Bergstrom, E. N., *et al.* The repertoire of mutational signatures in human cancer. *Nature* **578,** 94–101 (2020).

29. Gainza, P., Roberts, K. E. & Donald, B. R. Protein design using continuous rotamers. *PLoS computational biology* **8,** e1002335 (2012).

30. Hallen, M. A. & Donald, B. R. COMETS (Constrained Optimization of Multistate Energies by Tree Search): A provable and efficient protein design algorithm to optimize binding affinity and specificity with respect to sequence. *Journal of Computational Biology* **23,** 311–321 (2016).

31. Harrison, P. T., Vyse, S. & Huang, P. H. *Rare epidermal growth factor receptor (EGFR) mutations in non-small cell lung cancer* in *Seminars in cancer biology* **61** (2020), 167–179.

32. Lynch, T. J., Bell, D. W., Sordella, R., Gurubhagavatula, S., Okimoto, R. A., Brannigan, B. W., Harris, P. L., Haserlat, S. M., Supko, J. G., Haluska, F. G., *et al.* Activating mutations in the epidermal growth factor receptor underlying responsiveness of non–small-cell lung cancer to gefitinib. *New England Journal of Medicine* **350,** 2129–2139 (2004).

33. Davies, H., Bignell, G. R., Cox, C., Stephens, P., Edkins, S., Clegg, S., Teague, J., Woffendin, H., Garnett, M. J., Bottomley, W., *et al.* Mutations of the BRAF gene in human cancer. *Nature* **417,** 949–954 (2002).

34. Dowell, J., Minna, J. D. & Kirkpatrick, P. Erlotinib hydrochloride. *Nature Reviews Drug Discovery* **4** (2005).

35. Herbst, R. S., Fukuoka, M. & Baselga, J. Gefitinib—a novel targeted approach to treating cancer. *Nature Reviews Cancer* **4,** 956–965 (2004).

36. Soria, J.-C., Ohe, Y., Vansteenkiste, J., Reungwetwattana, T., Chewaskulyong, B., Lee, K. H., Dechaphunkul, A., Imamura, F., Nogami, N., Kurata, T., *et al.* Osimertinib in untreated EGFR-mutated advanced non–small-cell lung cancer. *New England journal of medicine* **378,** 113–125 (2018).

37. Ballantyne, A. D. & Garnock-Jones, K. P. Dabrafenib: first global approval. *Drugs* **73,** 1367–1376 (2013).

38. Bollag, G., Tsai, J., Zhang, J., Zhang, C., Ibrahim, P., Nolop, K. & Hirth, P. Vemurafenib: the first drug approved for BRAF-mutant cancer. *Nature reviews Drug discovery* **11,** 873–886 (2012).

39. Shirley, M. Encorafenib and binimetinib: first global approvals. *Drugs* **78,** 1277–1284 (2018).

40. Janku, F., Sherman, E., Parikh, A., Feun, L., Tsai, F., Allen, E., Zhang, C., Severson, P., Inokuchi, K., Walling, J., *et al.* Interim results from a phase 1/2 precision medicine study of PLX8394-a next generation BRAF inhibitor. *European Journal of Cancer* **138,** S2–S3 (2020).

41. Yun, C.-H., Boggon, T. J., Li, Y., Woo, M. S., Greulich, H., Meyerson, M. & Eck, M. J. Structures of lung cancer-derived EGFR mutants and inhibitor complexes: mechanism of activation and insights into differential inhibitor sensitivity. *Cancer cell* **11,** 217–227 (2007).

42. Stamos, J., Sliwkowski, M. X. & Eigenbrot, C. Structure of the epidermal growth factor receptor kinase domain alone and in complex with a 4-anilinoquinazoline inhibitor. *Journal of Biological Chemistry* **277,** 46265–46272 (2002).

43. Yosaatmadja, Y., Squire, C., McKeage, C. & Flanagan, M. 1.85 angstrom structure of EGFR kinase domain with gefitinib. *To Be Publ* (2014).

44. Yosaatmadja, Y., Silva, S., Dickson, J. M., Patterson, A. V., Smaill, J. B., Flanagan, J. U., McKeage, M. J. & Squire, C. J. Binding mode of the breakthrough inhibitor AZD9291 to epidermal growth factor receptor revealed. *Journal of structural biology* **192,** 539–544 (2015).

45. Zhang, C., Spevak, W., Zhang, Y., Burton, E. A., Ma, Y., Habets, G., Zhang, J., Lin, J., Ewing, T., Matusow, B., *et al.* RAF inhibitors that evade paradoxical MAPK pathway activation. *Nature* **526,** 583–586 (2015).

46. Hodis, E., Watson, I. R., Kryukov, G. V., Arold, S. T., Imielinski, M., Theurillat, J.-P., Nickerson, E., Auclair, D., Li, L., Place, C., *et al.* A landscape of driver mutations in melanoma. *Cell* **150,** 251–263 (2012).

47. Jou, J. D., Holt, G. T., Lowegard, A. U. & Donald, B. R. Minimization-Aware Recursive K*: A Novel, Provable Algorithm that Accelerates Ensemble-Based Protein Design and Provably Approximates the Energy Landscape. *Journal of Computational Biology* **27,** 550–564 (2020).

48. Helena, A. Y., Arcila, M. E., Rekhtman, N., Sima, C. S., Zakowski, M. F., Pao, W., Kris, M. G., Miller, V. A., Ladanyi, M. & Riely, G. J. Analysis of tumor specimens at the time of acquired resistance to EGFR-TKI therapy in 155 patients with EGFR-mutant lung cancers. *Clinical cancer research* **19,** 2240–2247 (2013).

49. Avizienyte, E., Ward, R. A. & Garner, A. P. Comparison of the EGFR resistance mutation profiles generated by EGFR-targeted tyrosine kinase inhibitors and the impact of drug combinations. *Biochemical Journal* **415,** 197–206 (2008).

50. Chen, K., Zhou, F., Shen, W., Jiang, T., Wu, X., Tong, X., Shao, Y. W., Qin, S. & Zhou, C. Novel mutations on EGFR Leu792 potentially correlate to acquired resistance to osimertinib in advanced NSCLC. *Journal of Thoracic Oncology* **12,** e65–e68 (2017).

51. Yang, Z., Yang, N., Ou, Q., Xiang, Y., Jiang, T., Wu, X., Bao, H., Tong, X., Wang, X., Shao, Y. W., *et al.* Investigating novel resistance mechanisms to third-generation EGFR tyrosine kinase inhibitor osimertinib in non–small cell lung cancer patients. *Clinical Cancer Research* **24,** 3097–3107 (2018).

52. Ou, S.-H. I., Cui, J., Schrock, A. B., Goldberg, M. E., Zhu, V. W., Albacker, L., Stephens, P. J., Miller, V. A. & Ali, S. M. Emergence of novel and dominant acquired EGFR solvent-front mutations at Gly796 (G796S/R) together with C797S/G and L792F/H mutations in one EGFR (L858R/T790M) NSCLC patient who progressed on osimertinib. *Lung Cancer* **108,** 228–231 (2017).

53. Fairclough, S. R., Kiedrowski, L. A., Lin, J. J., Zelichov, O., Tarcic, G., Stinchcombe, T. E., Odegaard, J. I., Lanman, R. B., Shaw, A. T. & Nagy, R. J. Identification of osimertinib-resistant EGFR L792 mutations by cfDNA sequencing: oncogenic activity assessment and prevalence in large cfDNA cohort. *Experimental hematology & oncology* **8,** 1–6 (2019).

54. Li, D., Yang, D., Cui, S., Pan, E., Yang, P. & Dai, Z. NGS-Based ctDNA Profiling After the Resistance of Second-Line Osimertinib for Patient with EGFR-Mutated Pulmonary Adenocarcinoma. *OncoTargets and therapy* **14,** 4261 (2021).

55. Zheng, D., Hu, M., Bai, Y., Zhu, X., Lu, X., Wu, C., Wang, J., Liu, L., Wang, Z., Ni, J., *et al.* EGFR G796D mutation mediates resistance to osimertinib. *Oncotarget* **8,** 49671 (2017).

56. Tate, J. G., Bamford, S., Jubb, H. C., Sondka, Z., Beare, D. M., Bindal, N., Boutselakis, H., Cole, C. G., Creatore, C., Dawson, E., *et al.* COSMIC: the catalogue of somatic mutations in cancer. *Nucleic acids research* **47,** D941–D947 (2019).

57. Thress, K. S., Paweletz, C. P., Felip, E., Cho, B. C., Stetson, D., Dougherty, B., Lai, Z., Markovets, A., Vivancos, A., Kuang, Y., *et al.* Acquired EGFR C797S mutation mediates resistance to AZD9291 in non–small cell lung cancer harboring EGFR T790M. *Nature medicine* **21,** 560–562 (2015).

58. Arulananda, S., Do, H., Musafer, A., Mitchell, P., Dobrovic, A. & John, T. Combination osimertinib and gefitinib in C797S and T790M EGFR-mutated non–small cell lung cancer. *Journal of Thoracic Oncology* **12,** 1728–1732 (2017).

59. Wagenaar, T. R., Ma, L., Roscoe, B., Park, S. M., Bolon, D. N. & Green, M. R. Resistance to vemurafenib resulting from a novel mutation in the BRAFV 600 E kinase domain. *Pigment cell & melanoma research* **27,** 124–133 (2014).

60. Ojewole, A. A., Jou, J. D., Fowler, V. G. & Donald, B. R. BBK*(Branch and Bound Over K*): A Provable and Efficient Ensemble-Based Protein Design Algorithm to Optimize Stability and Binding Affinity Over Large Sequence Spaces. *Journal of Computational Biology* **25,** 726–739 (2018).

61. Valiant, L. G. The complexity of computing the permanent. *Theoretical computer science* **8,** 189–201 (1979).

62. Nisonoff, H. *Efficient Partition Function Estimation in Computational Protein Design: Probabalistic Guarantees and Characterization of a Novel Algorithm* Bachelor's Thesis (Duke University, 2015).

63. Viricel, C., Simoncini, D., Barbe, S. & Schiex, T. *Guaranteed weighted counting for affinity computation: Beyond determinism and structure* in *International Conference on Principles and Practice of Constraint Programming* (2016), 733–750.

64. Jou, J. D., Jain, S., Georgiev, I. S. & Donald, B. R. BWM*: A novel, provable, ensemble-based dynamic programming algorithm for sparse approximations of computational protein design. *Journal of Computational Biology* **23,** 413–424 (2016).

65. Sastry, G. M., Adzhigirey, M., Day, T., Annabhimoju, R. & Sherman, W. Protein and ligand preparation: parameters, protocols, and influence on virtual screening enrichments. *Journal of computer-aided molecular design* **27,** 221–234 (2013).

66. Schrödinger, LLC. *Maestro* version Schrödinger Release 2020-3. Jan. 1, 2021. https://schrodinger.com.

67. Farid, R., Day, T., Friesner, R. A. & Pearlstein, R. A. New insights about HERG blockade obtained from protein modeling, potential energy mapping, and docking studies. *Bioorganic & medicinal chemistry* **14,** 3160–3173 (2006).

68. Sherman, W., Day, T., Jacobson, M. P., Friesner, R. A. & Farid, R. Novel procedure for modeling ligand/receptor induced fit effects. *Journal of medicinal chemistry* **49,** 534–553 (2006).

69. Sherman, W., Beard, H. S. & Farid, R. Use of an induced fit receptor structure in virtual screening. *Chemical biology & drug design* **67,** 83–84 (2006).

70. Jones, G., Willett, P., Glen, R. C., Leach, A. R. & Taylor, R. Development and validation of a genetic algorithm for flexible docking. *Journal of molecular biology* **267,** 727–748 (1997).

71. Haling, J. R., Sudhamsu, J., Yen, I., Sideris, S., Sandoval, W., Phung, W., Bravo, B. J., Giannetti, A. M., Peck, A., Masselot, A., *et al.* Structure of the BRAF-MEK complex reveals a kinase activity independent role for BRAF in MAPK signaling. *Cancer cell* **26,** 402–413 (2014).

72. Xu, W., Doshi, A., Lei, M., Eck, M. J. & Harrison, S. C. Crystal structures of c-Src reveal features of its autoinhibitory mechanism. *Molecular cell* **3,** 629–638 (1999).

73. Chemical Computing Group ULC. *Molecular Operating Environment (MOE)* 2019.

Competing interests: BRD is a founder of Ten63 Therapeutics, Inc.

Appendices

The following are optional appendices providing additional information to substantiate the claims made in the main paper. Appendix A describes the steps we took to a) prepare empirical protein structures for K^* scoring, and; b) to create the B-Raf structures with the various inhibitors for which no empirical structures exist. Appendix B shows the complete set of predictions RESISTOR made for EGFR with erlotinib, gefitinib, and osimertinib. Appendix C shows the complete set of predictions RESISTOR made for B-Raf with vemurafenib, encorafenib, dabrafenib, and PLX-8394.

A. Preparation of Empirical and Docked Structures for K^* Predictions

A.1. EGFR input structures

The crystal structures used for the EGFR predictions were adopted from Kaserer and Blagg's 2018 publication [19]. A full description of the PDB entries used can be found in that paper's section *Table S7*, and details on how the structures were prepared for OSPREY predictions is in that paper's section *Structure Selection and Preparation*.

A.2. B-Raf input structures

The crystal structures of vemurafenib (PDB id 3og7 [46]) and dabrafenib (PDB id 4xv2 [45]) in complex with B-Raf V600E were selected as input for RESISTOR. Both structures have been prepared using the default setting of the Protein Preparation Wizard [65] in Maestro [66]. In the case of encorafenib and PLX8394, crystal structures of structurally closely related, but not the identical, molecules were available. These experimental complexes were used to generate encorafenib and PLX8394 models. Encorafenib was docked into PDB id 4xv3 [45] using the default settings of the induced fit docking procedure in Maestro [66–69]. For validation, the co-crystallized ligand PLX7922 was re-docked. The highest scored docking pose of encorafenib was selected for further investigation. We found that the conserved substructures in encorafenib and PLX7922 aligned very well in this docking pose.

For PLX8394, re-docking of the co-crystallized ligand PLX7904 (PDB id 4xv1 [45]) failed with the induced fit docking procedure, but was successful using a rigid docking workflow in GOLD version 5.8.0 [70]. The binding site was defined as 6 Å around the ligand and the water molecule HOH905 was set to toggle and spin. The default settings of all other parameters were used.

An experimental structure of the endogenous ligand ADP was available, however, B-Raf adopted in inactive conformation in this complex. Apo B-Raf in its active conformation (PDB id 4mne [71]) was thus combined with ANP-bound protein kinase c-src (PDB id 2src [72]) to generate an active, endogenous ligand-bound B-Raf complex. This model was used as template to build a B-Raf:ADP homology model in the Molecular Operating Environment [73] using the default settings. This included refinement steps to resolve potential steric clashes in the rather crude ANP-B-Raf input template.

For all complexes, water molecules not involved in mediating interactions between the ligand and the target were deleted and only residues with a 12 Å radius around the ligand were kept in the final input structures.

B. EGFR Pareto Frontier

Table 2: All RESISTOR resistance mutation predictions for EGFR with erlotinib. "WT AA" is the wildtype identity of the amino acid. "Mut AA" is the resistance mutation. "Sig Prob" is the mutational signature probability for the mutation from "WT AA" to "Mut AA" in lung adenocarcinoma. "K* ATP" is the K^* score of the endogenous ligand. "K* Erlotinib" is the K^* score of erlotinib. "Count" is number of resistance mutations at the position. "Rank" is the Pareto rank of the mutation. Note: K^* scores are in $\log_{10}$ units where possible and 0 where there is predicted to be no binding.

Position	WT AA	Mut AA	Sig Prob	K* ATP	K* Erlotinib	Count	Rank
718	LEU	PHE	2.47E-04	17.16	0	8	1
718	LEU	HIP	4.20E-04	18.86	-62.44	8	1
718	LEU	HIE	4.20E-04	18.92	-61.9	8	1
723	PHE	VAL	3.16E-04	19.14	0	5	1
723	PHE	LEU	8.27E-03	19.05	0	5	1
726	VAL	PHE	5.09E-04	19.71	0	2	1
743	ALA	ASP	1.09E-02	13.51	0	4	1
790	THR	LYS	7.38E-03	19.54	22.01	4	1
790	THR	MET	6.02E-03	19.79	23.89	4	1
791	GLN	PRO	2.30E-03	19.22	0	3	1
791	GLN	LYS	1.63E-02	19.1	0	3	1
796	GLY	TRP	2.06E-05	19.13	0	12	1
796	GLY	LEU	4.41E-05	19.55	-25.46	12	1
796	GLY	GLU	1.54E-04	18.88	1.16	12	1
796	GLY	PHE	1.76E-04	19.48	4.68	12	1
796	GLY	ARG	2.86E-03	19.54	9.36	12	1
796	GLY	ASP	5.32E-03	19.15	18.29	12	1
796	GLY	CYS	3.84E-03	19.28	21.71	12	1
796	GLY	SER	6.43E-03	19.23	22.24	12	1
718	LEU	GLY	8.49E-06	18.15	0	8	2
718	LEU	TRP	1.75E-05	17.96	0	8	2
718	LEU	HID	4.20E-04	18.9	-60.92	8	2
718	LEU	ARG	2.38E-03	19.41	22.5	8	2
726	VAL	TRP	4.82E-05	19.51	0	2	2
745	LYS	ILE	2.43E-04	19.05	0	5	2
745	LYS	MET	5.16E-03	18.97	0	5	2
790	THR	ARG	1.39E-03	19.32	11.4	4	2
791	GLN	GLY	1.81E-05	19.06	0	3	2
796	GLY	TYR	4.34E-05	19.48	-13.32	12	2
796	GLY	ASN	5.25E-05	19.36	21.0	12	2
796	GLY	HIE	1.88E-05	19.55	23.68	12	2
800	ASP	GLY	1.53E-03	19.13	0	1	2
718	LEU	LYS	2.70E-04	19.22	22.98	8	3
723	PHE	ASP	8.41E-07	18.98	0	5	3
745	LYS	HIE	6.76E-05	18.85	0	5	3
745	LYS	THR	1.26E-03	18.71	0	5	3
790	THR	ASN	2.19E-04	19.16	21.87	4	3
793	TRP	ASN	1.04E-04	18.93	0	1	3

Position	WT AA	Mut AA	Sig Prob	K* ATP	K* Erlotinib	Count	Rank
796	GLY	THR	3.45E-05	19.28	16.88	12	3
844	LEU	TRP	1.90E-05	19.02	-17.34	4	3
844	LEU	HID	4.20E-04	18.8	22.65	4	3
844	LEU	HIE	4.20E-04	18.74	22.63	4	3
854	THR	ASN	2.62E-04	19.09	21.08	1	3
723	PHE	ALA	5.72E-07	18.98	0	5	4
723	PHE	GLY	5.92E-07	18.91	0	5	4
743	ALA	CYS	7.34E-05	14.53	0	4	4
743	ALA	GLU	7.74E-05	4.17	0	4	4
745	LYS	HID	6.76E-05	18.81	0	5	4
844	LEU	HIP	4.20E-04	18.57	22.42	4	4
743	ALA	ARG	1.73E-05	6.55	0	4	5

Table 3: All RESISTOR resistance mutation predictions for EGFR with gefitinib. "WT AA" is the wildtype identity of the amino acid. "Mut AA" is the resistance mutation. "Sig Prob" is the mutational signature probability for the mutation from "WT AA" to "Mut AA" in lung adenocarcinoma. "K* ATP" is the K^* score of the endogenous ligand. "K* Gefitinib" is the K^* score of gefitinib. "Count" is number of resistance mutations at the position. "Rank" is the Pareto rank of the mutation. Note: K^* scores are in $\log_{10}$ units where possible and 0 where there is predicted to be no binding.

Position	WT AA	Mut AA	Sig Prob	K* ATP	K* Gefitinib	Count	Rank
718	LEU	PHE	2.47E-04	17.16	-23.15	7	1
718	LEU	TRP	1.75E-05	17.96	1.76	7	1
718	LEU	HIP	4.20E-04	18.86	4.5	7	1
718	LEU	HID	4.20E-04	18.9	4.92	7	1
718	LEU	HIE	4.20E-04	18.92	4.99	7	1
718	LEU	ARG	2.38E-03	19.41	23.69	7	1
743	ALA	GLU	7.74E-05	4.17	-49.27	2	1
777	TYR	HID	2.81E-04	19.04	0	1	1
790	THR	ARG	1.39E-03	19.32	12.73	4	1
790	THR	LYS	7.38E-03	19.54	23.21	4	1
790	THR	MET	6.02E-03	19.79	25.06	4	1
796	GLY	LEU	4.41E-05	19.55	22.57	1	1
844	LEU	TRP	1.90E-05	19.02	-28.99	4	1
718	LEU	LYS	2.70E-04	19.22	24.45	7	2
726	VAL	PHE	5.09E-04	19.71	25.26	1	2
743	ALA	ARG	1.73E-05	6.55	-4.31	2	2
745	LYS	ILE	2.43E-04	19.05	16.99	1	2
790	THR	ASN	2.19E-04	19.16	23.52	4	2
844	LEU	HIP	4.20E-04	18.57	23.69	4	2
844	LEU	HID	4.20E-04	18.8	23.97	4	2
844	LEU	HIE	4.20E-04	18.74	23.94	4	2
854	THR	ASN	2.62E-04	19.09	20.61	1	2

Table 4: All RESISTOR resistance mutation predictions for EGFR with osimertinib. "WT AA" is the wildtype identity of the amino acid. "Mut AA" is the resistance mutation. "Sig Prob" is the mutational signature probability for the mutation from "WT AA" to "Mut AA" in lung adenocarcinoma. "K* ATP" is the K^* score of the endogenous ligand. "K* Osimertinib" is the K^* score of osimertinib. "Count" is number of resistance mutations at the position. "Rank" is the Pareto rank of the mutation. Note: K^* scores are in $\log_{10}$ units where possible and 0 where there is predicted to be no binding.

Position	WT AA	Mut AA	Sig Prob	K* ATP	K* Osimertinib	Count	Rank
718	LEU	TRP	1.75E-05	17.96	0	9	1
718	LEU	PHE	2.47E-04	17.16	0	9	1
718	LEU	HIP	4.20E-04	18.86	-38.77	9	1
718	LEU	HIE	4.20E-04	18.92	-38.57	9	1
718	LEU	MET	1.08E-02	19.44	25.52	9	1
719	GLY	VAL	1.70E-02	14.7	20.75	2	1
726	VAL	TRP	4.82E-05	19.51	0	3	1
743	ALA	ASP	1.09E-02	13.51	17.5	3	1
796	GLY	TRP	2.06E-05	19.13	-97.39	14	1
796	GLY	TYR	4.34E-05	19.48	-61.59	14	1
796	GLY	PHE	1.76E-04	19.48	-41.78	14	1
796	GLY	LEU	4.41E-05	19.55	-3.75	14	1
796	GLY	ARG	2.86E-03	19.54	10.61	14	1
796	GLY	ASP	5.32E-03	19.15	14.96	14	1
796	GLY	CYS	3.84E-03	19.28	21.85	14	1
796	GLY	SER	6.43E-03	19.23	24.76	14	1
718	LEU	HID	4.20E-04	18.9	-37.61	9	2
718	LEU	ARG	2.38E-03	19.41	20.16	9	2
723	PHE	ILE	2.09E-03	19.5	25.68	1	2
792	LEU	HIP	4.86E-03	18.88	24.98	3	2
792	LEU	HIE	4.86E-03	18.93	25.07	3	2
792	LEU	HID	4.86E-03	18.98	25.15	3	2
796	GLY	GLU	1.54E-04	18.88	2.49	14	2
796	GLY	HIE	1.88E-05	19.55	11.32	14	2
796	GLY	ASN	5.25E-05	19.36	18.38	14	2
718	LEU	LYS	2.70E-04	19.22	24.7	9	3
719	GLY	THR	4.02E-05	17.76	20.4	2	3
726	VAL	ARG	2.32E-05	17.68	21.62	3	3
726	VAL	LYS	5.39E-05	17.07	21.87	3	3
743	ALA	GLU	7.74E-05	4.17	0.47	3	3
796	GLY	HID	1.88E-05	19.49	11.69	14	3
796	GLY	THR	3.45E-05	19.28	22.43	14	3
844	LEU	TRP	1.90E-05	19.02	22.12	1	3
796	GLY	HIP	1.88E-05	19.46	12.09	14	4
718	LEU	GLY	8.49E-06	18.15	24.02	9	5
743	ALA	ARG	1.73E-05	6.55	12.58	3	5

C. B-Raf Pareto Frontier

Table 5: All RESISTOR resistance mutation predictions for B-Raf with dabrafenib. "WT AA" is the wildtype identity of the amino acid. "Mut AA" is the resistance mutation. "Sig Prob" is the mutational signature probability for the mutation from "WT AA" to "Mut AA" in melanoma. "K* ATP" is the K^* score of the endogenous ligand. "K* Dabrafenib" is the K^* score of dabrafenib. "Count" is number of resistance mutations at the position. "Rank" is the Pareto rank of the mutation. Note: K^* scores are in $\log_{10}$ units where possible and 0 where there is predicted to be no binding.

Position	WT AA	Mut AA	Sig Prob	K* ATP	K* Dabrafenib	Count	Rank
466	GLY	ARG	5.84E-02	10.53	-167.92	11	1
466	GLY	LYS	2.38E-02	11.79	-52.4	11	1
466	GLY	GLU	2.19E-01	12.89	21.32	11	1
471	VAL	LEU	4.43E-04	19.67	25.1	6	1
508	THR	ARG	2.95E-04	18.59	-118.81	4	1
535	SER	PRO	1.30E-03	18.65	0	1	1
593	GLY	PHE	1.99E-06	20.07	0	16	1
593	GLY	TYR	3.58E-05	19.86	0	16	1
593	GLY	ARG	7.80E-04	16.17	0	16	1
593	GLY	GLU	2.76E-04	18.73	-60.32	16	1
593	GLY	ASN	1.34E-03	19.16	-39.82	16	1
593	GLY	ASP	1.63E-02	18.89	-29.79	16	1
593	GLY	CYS	1.66E-03	19.06	17.46	16	1
593	GLY	VAL	9.35E-04	19.18	28.45	16	1
593	GLY	ILE	4.55E-05	19.8	30.24	16	1
593	GLY	SER	6.09E-02	18.83	34.27	16	1
466	GLY	GLN	7.24E-05	12.55	11.37	11	2
466	GLY	ASP	7.51E-04	17.06	18.64	11	2
466	GLY	VAL	2.51E-03	13.44	29.23	11	2
467	SER	PRO	7.37E-04	18.85	30.42	1	2
481	ALA	LYS	2.74E-04	17.32	-4.22	8	2
481	ALA	LEU	7.06E-05	18.68	9.97	8	2
481	ALA	GLU	1.21E-03	17.92	22.11	8	2
505	LEU	ARG	8.61E-04	18.58	16.53	5	2
508	THR	LYS	9.16E-04	18.59	27.22	4	2
514	LEU	ARG	5.22E-05	17.18	21.32	12	2
514	LEU	ILE	1.65E-03	18.4	32.55	12	2
529	THR	PHE	3.77E-05	15.91	-125.31	11	2
529	THR	MET	1.74E-05	18.65	-8.16	11	2
529	THR	ASN	9.96E-04	18.55	34.54	11	2
593	GLY	HIE	1.66E-05	19.58	0	16	2
593	GLY	THR	2.67E-05	19.06	27.33	16	2
464	GLY	GLN	7.24E-05	2.95	11.05	1	3
466	GLY	THR	5.47E-05	14.85	29.21	11	3
481	ALA	ILE	5.14E-05	14.2	21.38	8	3
481	ALA	VAL	1.46E-03	17.77	33.45	8	3
505	LEU	SER	2.78E-05	18.58	35.02	5	3
514	LEU	PRO	1.21E-03	18.12	33.34	12	3
527	ILE	LEU	6.37E-05	18.61	33.5	1	3

Position	WT AA	Mut AA	Sig Prob	K* ATP	K* Dabrafenib	Count	Rank
593	GLY	HIP	1.66E-05	19.56	0	16	3
514	LEU	SER	2.27E-04	18.1	34.19	12	4
593	GLY	HID	1.66E-05	19.55	0	16	4
471	VAL	MET	8.44E-06	18.96	27.83	6	5
481	ALA	ARG	1.53E-05	17.02	-15.13	8	5
481	ALA	ASP	4.39E-06	19.21	31.9	8	5
508	THR	GLU	1.05E-05	18.59	35.0	4	5
514	LEU	PHE	1.53E-05	18.19	0	12	5
578	LYS	TYR	5.24E-06	18.39	-143.76	1	5
593	GLY	TRP	4.78E-06	18.83	0	16	5
593	GLY	LEU	3.87E-07	19.35	-85.52	16	5
466	GLY	PRO	1.77E-07	18.27	0	11	6
466	GLY	TRP	2.10E-06	13.1	0	11	6
466	GLY	LEU	4.82E-06	9.74	-39.41	11	6
466	GLY	CYS	3.82E-06	17.62	27.45	11	6
469	GLY	PRO	1.77E-07	18.75	0	1	6
471	VAL	PRO	5.74E-07	17.88	0	6	6
471	VAL	ARG	3.03E-07	18.99	23.05	6	6
471	VAL	GLU	9.01E-07	18.82	31.61	6	6
481	ALA	GLN	1.80E-06	18.37	15.47	8	6
508	THR	GLN	2.51E-06	18.59	33.59	4	6
513	ILE	ARG	1.06E-06	18.57	-30.17	2	6
513	ILE	TYR	2.73E-06	18.57	27.86	2	6
514	LEU	HIP	3.37E-07	17.68	25.98	12	6
514	LEU	HID	3.37E-07	17.72	26.06	12	6
514	LEU	LYS	1.68E-06	17.37	32.25	12	6
514	LEU	MET	2.51E-06	18.42	33.68	12	6
528	VAL	ARG	5.57E-07	18.61	-72.34	1	6
529	THR	TYR	1.12E-06	18.58	-10.9	11	6
529	THR	ARG	4.98E-07	18.6	-4.95	11	6
529	THR	LYS	1.74E-06	18.57	4.7	11	6
529	THR	LEU	3.10E-06	18.56	26.71	11	6
532	CYS	HID	4.44E-06	14.39	26.54	7	6
532	CYS	HIP	4.44E-06	14.51	26.79	7	6
532	CYS	HIE	4.44E-06	12.48	25.2	7	6
532	CYS	ILE	1.16E-06	18.82	34.28	7	6
532	CYS	VAL	2.11E-06	18.7	34.77	7	6
471	VAL	HID	6.58E-07	17.61	33.79	6	7
505	LEU	GLY	6.90E-08	18.58	34.62	5	7
505	LEU	GLN	1.84E-06	18.6	34.82	5	7
514	LEU	HIE	3.37E-07	17.72	27.35	12	7
514	LEU	GLY	3.88E-08	18.03	33.68	12	7
514	LEU	ALA	9.24E-07	18.09	34.25	12	7
529	THR	HID	6.68E-08	18.58	-10.11	11	7
529	THR	HIE	6.68E-08	18.58	-4.75	11	7
529	THR	ASP	2.37E-06	18.51	34.7	11	7
531	TRP	PRO	5.44E-07	16.76	0	1	7

Position	WT AA	Mut AA	Sig Prob	K* ATP	K* Dabrafenib	Count	Rank
505	LEU	ALA	5.28E-07	18.58	35.0	5	8
529	THR	HIP	6.68E-08	18.58	-6.63	11	8

Table 6: All RESISTOR resistance mutation predictions for B-Raf with vemurafenib. "WT AA" is the wildtype identity of the amino acid. "Mut AA" is the resistance mutation. "Sig Prob" is the mutational signature probability for the mutation from "WT AA" to "Mut AA" in melanoma. "K* ATP" is the K^* score of the endogenous ligand. "K* Vemurafenib" is the K^* score of vemurafenib. "Count" is number of resistance mutations at the position. "Rank" is the Pareto rank of the mutation. Note: K^* scores are in $\log_{10}$ units where possible and 0 where there is predicted to be no binding.

Position	WT AA	Mut AA	Sig Prob	K* ATP	K* Vemurafenib	Count	Rank
471	VAL	LEU	4.43E-04	19.67	29.39	4	1
481	ALA	THR	1.77E-02	18.99	30.69	9	1
529	THR	ILE	2.02E-02	18.57	29.33	10	1
535	SER	PRO	1.30E-03	18.65	0	1	1
593	GLY	PHE	1.99E-06	20.07	0	16	1
593	GLY	TYR	3.58E-05	19.86	0	16	1
593	GLY	ARG	7.80E-04	16.17	-231.93	16	1
593	GLY	ASN	1.34E-03	19.16	-103.95	16	1
593	GLY	ASP	1.63E-02	18.89	-26.78	16	1
593	GLY	CYS	1.66E-03	19.06	19.89	16	1
593	GLY	VAL	9.35E-04	19.18	21.09	16	1
593	GLY	ILE	4.55E-05	19.8	23.08	16	1
593	GLY	SER	6.09E-02	18.83	31.33	16	1
463	ILE	TYR	2.55E-05	15.94	-124.43	4	2
481	ALA	GLU	1.21E-03	17.92	11.12	9	2
481	ALA	VAL	1.46E-03	17.77	28.52	9	2
505	LEU	PHE	5.44E-03	18.58	27.01	3	2
505	LEU	ARG	8.61E-04	18.58	27.41	3	2
508	THR	ARG	2.95E-04	18.59	12.91	2	2
508	THR	LYS	9.16E-04	18.59	21.18	2	2
514	LEU	ILE	1.65E-03	18.4	30.08	11	2
532	CYS	ARG	8.12E-04	13.04	-10.66	9	2
593	GLY	HIE	1.66E-05	19.58	0	16	2
593	GLY	GLU	2.76E-04	18.73	-12.82	16	2
593	GLY	THR	2.67E-05	19.06	20.72	16	2
481	ALA	LEU	7.06E-05	18.68	18.19	9	3
481	ALA	LYS	2.74E-04	17.32	18.46	9	3
514	LEU	ARG	5.22E-05	17.18	22.28	11	3
514	LEU	GLN	8.77E-05	17.02	27.9	11	3
514	LEU	SER	2.27E-04	18.1	30.81	11	3
529	THR	MET	1.74E-05	18.65	-8.52	10	3
529	THR	PHE	3.77E-05	15.91	5.09	10	3
593	GLY	HIP	1.66E-05	19.56	0	16	3
481	ALA	ILE	5.14E-05	14.2	24.57	9	4
593	GLY	HID	1.66E-05	19.55	0	16	4
471	VAL	MET	8.44E-06	18.96	31.15	4	5

Position	WT AA	Mut AA	Sig Prob	K* ATP	K* Vemurafenib	Count	Rank
481	ALA	ARG	1.53E-05	17.02	19.55	9	5
481	ALA	ASP	4.39E-06	19.21	24.15	9	5
514	LEU	PHE	1.53E-05	18.19	0	11	5
593	GLY	TRP	4.78E-06	18.83	0	16	5
593	GLY	LEU	3.87E-07	19.35	-237.44	16	5
463	ILE	HIE	5.14E-06	18.06	30.11	4	6
463	ILE	HID	5.14E-06	18.08	30.75	4	6
466	GLY	PRO	1.77E-07	18.27	0	1	6
471	VAL	PRO	5.74E-07	17.88	0	4	6
471	VAL	GLU	9.01E-07	18.82	31.34	4	6
481	ALA	GLN	1.80E-06	18.37	-15.45	9	6
514	LEU	HIP	3.37E-07	17.68	25.66	11	6
514	LEU	HID	3.37E-07	17.72	25.78	11	6
514	LEU	GLY	3.88E-08	18.03	30.29	11	6
514	LEU	ALA	9.24E-07	18.09	30.82	11	6
516	PHE	ARG	4.47E-06	18.58	29.7	1	6
529	THR	TYR	1.12E-06	18.58	-114.94	10	6
529	THR	ARG	4.98E-07	18.6	-34.39	10	6
529	THR	LYS	1.74E-06	18.57	-7.92	10	6
529	THR	LEU	3.10E-06	18.56	28.04	10	6
532	CYS	HIP	4.44E-06	14.51	-0.94	9	6
532	CYS	HIE	4.44E-06	12.48	-2.8	9	6
532	CYS	ILE	1.16E-06	18.82	27.39	9	6
532	CYS	VAL	2.11E-06	18.7	29.58	9	6
463	ILE	HIP	5.14E-06	17.51	30.23	4	7
505	LEU	MET	3.58E-07	18.6	25.88	3	7
514	LEU	HIE	3.37E-07	17.72	25.91	11	7
529	THR	HIE	6.68E-08	18.58	10.11	10	7
531	TRP	PRO	5.44E-07	16.76	0	1	7
532	CYS	HID	4.44E-06	14.39	-0.66	9	7
532	CYS	THR	2.67E-07	18.32	30.91	9	7
514	LEU	GLU	5.81E-08	17.36	28.0	11	8
529	THR	HIP	6.68E-08	18.58	10.86	10	8
529	THR	HID	6.68E-08	18.58	11.36	10	8

Table 7: All RESISTOR resistance mutation predictions for B-Raf with encorafenib. "WT AA" is the wildtype identity of the amino acid. "Mut AA" is the resistance mutation. "Sig Prob" is the mutational signature probability for the mutation from "WT AA" to "Mut AA" in melanoma. "K* ATP" is the K^* score of the endogenous ligand. "K* Encorafenib" is the K^* score of encorafenib. "Count" is number of resistance mutations at the position. "Rank" is the Pareto rank of the mutation. Note: K^* scores are in $\log_{10}$ units where possible and 0 where there is predicted to be no binding.

Position	WT AA	Mut AA	Sig Prob	K* ATP	K* Encorafenib	Count	Rank
471	VAL	LEU	4.43E-04	19.67	28.68	10	1
481	ALA	LEU	7.06E-05	18.68	-24.05	8	1
481	ALA	GLU	1.21E-03	17.92	10.6	8	1
529	THR	ILE	2.02E-02	18.57	31.23	12	1

Position	WT AA	Mut AA	Sig Prob	K* ATP	K* Encorafenib	Count	Rank
532	CYS	ARG	8.12E-04	13.04	-19.78	9	1
535	SER	PRO	1.30E-03	18.65	0	1	1
593	GLY	TYR	3.58E-05	19.86	0	17	1
593	GLY	ARG	7.80E-04	16.17	-2.33	17	1
593	GLY	ILE	4.55E-05	19.8	2.23	17	1
593	GLY	VAL	9.35E-04	19.18	7.05	17	1
593	GLY	ASN	1.34E-03	19.16	28.19	17	1
593	GLY	ASP	1.63E-02	18.89	28.87	17	1
593	GLY	PHE	1.99E-06	20.07	30.05	17	1
593	GLY	CYS	1.66E-03	19.06	34.56	17	1
593	GLY	SER	6.09E-02	18.83	35.04	17	1
471	VAL	PHE	7.85E-04	16.37	26.0	10	2
481	ALA	LYS	2.74E-04	17.32	7.23	8	2
514	LEU	ARG	5.22E-05	17.18	20.73	10	2
529	THR	MET	1.74E-05	18.65	-18.87	12	2
529	THR	PHE	3.77E-05	15.91	0.72	12	2
529	THR	ASN	9.96E-04	18.55	33.77	12	2
536	SER	ASN	1.37E-02	18.53	34.35	3	2
583	PHE	TYR	4.08E-03	18.68	29.81	9	2
593	GLY	HIE	1.66E-05	19.58	0	17	2
593	GLY	GLU	2.76E-04	18.73	22.97	17	2
593	GLY	THR	2.67E-05	19.06	24.19	17	2
593	GLY	ALA	2.54E-04	18.8	35.61	17	2
463	ILE	TYR	2.55E-05	15.94	9.07	3	3
481	ALA	ILE	5.14E-05	14.2	30.63	8	3
481	ALA	VAL	1.46E-03	17.77	34.23	8	3
505	LEU	HIP	1.46E-03	18.59	35.98	2	3
514	LEU	GLN	8.77E-05	17.02	31.64	10	3
536	SER	ASP	5.03E-05	18.21	33.54	3	3
583	PHE	VAL	2.14E-04	17.05	33.2	9	3
583	PHE	ILE	3.16E-03	17.45	34.49	9	3
583	PHE	SER	2.62E-03	16.86	34.31	9	3
593	GLY	HIP	1.66E-05	19.56	0	17	3
505	LEU	HID	1.46E-03	18.59	36.08	2	4
593	GLY	HID	1.66E-05	19.55	0	17	4
471	VAL	PRO	5.74E-07	17.88	0	10	5
471	VAL	ARG	3.03E-07	18.99	18.89	10	5
471	VAL	MET	8.44E-06	18.96	21.35	10	5
481	ALA	GLN	1.80E-06	18.37	-7.07	8	5
481	ALA	ARG	1.53E-05	17.02	1.87	8	5
481	ALA	ASP	4.39E-06	19.21	30.87	8	5
514	LEU	PHE	1.53E-05	18.19	-25.54	10	5
529	THR	ARG	4.98E-07	18.6	-6.65	12	5
532	CYS	HIP	4.44E-06	14.51	-40.5	9	5
532	CYS	HIE	4.44E-06	12.48	-40.74	9	5
533	GLU	PRO	3.22E-07	18.63	0	1	5
593	GLY	LEU	3.87E-07	19.35	19.61	17	5

Position	WT AA	Mut AA	Sig Prob	K* ATP	K* Encorafenib	Count	Rank
593	GLY	TRP	4.78E-06	18.83	19.1	17	5
463	ILE	HIE	5.14E-06	18.06	35.28	3	6
471	VAL	GLU	9.01E-07	18.82	35.77	10	6
529	THR	TYR	1.12E-06	18.58	19.14	12	6
529	THR	LYS	1.74E-06	18.57	20.27	12	6
529	THR	HIE	6.68E-08	18.58	21.51	12	6
529	THR	LEU	3.10E-06	18.56	22.42	12	6
531	TRP	PRO	5.44E-07	16.76	0	1	6
532	CYS	HID	4.44E-06	14.39	-40.35	9	6
532	CYS	ILE	1.16E-06	18.82	30.25	9	6
532	CYS	VAL	2.11E-06	18.7	30.4	9	6
583	PHE	ARG	5.91E-06	17.17	32.32	9	6
583	PHE	THR	1.13E-05	16.91	32.78	9	6
583	PHE	MET	5.94E-06	18.39	35.48	9	6
463	ILE	HID	5.14E-06	18.08	35.51	3	7
471	VAL	HIP	6.58E-07	17.42	30.52	10	7
471	VAL	HID	6.58E-07	17.61	31.08	10	7
471	VAL	TYR	1.33E-06	16.37	31.46	10	7
514	LEU	HIP	3.37E-07	17.68	31.73	10	7
514	LEU	HID	3.37E-07	17.72	31.87	10	7
514	LEU	LYS	1.68E-06	17.37	33.97	10	7
514	LEU	MET	2.51E-06	18.42	35.5	10	7
529	THR	HID	6.68E-08	18.58	20.75	12	7
529	THR	ASP	2.37E-06	18.51	34.9	12	7
536	SER	LEU	5.47E-07	17.44	30.59	3	7
583	PHE	TRP	9.45E-07	13.08	24.17	9	7
583	PHE	GLY	1.52E-06	16.63	33.81	9	7
471	VAL	HIE	6.58E-07	17.28	30.93	10	8
514	LEU	HIE	3.37E-07	17.72	32.04	10	8
529	THR	HIP	6.68E-08	18.58	20.97	12	8
532	CYS	THR	2.67E-07	18.32	35.12	9	8
514	LEU	GLU	5.81E-08	17.36	32.45	10	9
514	LEU	GLY	3.88E-08	18.03	35.47	10	9

Table 8: All RESISTOR resistance mutation predictions for B-Raf with PLX8394. "WT AA" is the wildtype identity of the amino acid. "Mut AA" is the resistance mutation. "Sig Prob" is the mutational signature probability for the mutation from "WT AA" to "Mut AA" in melanoma. "K* ATP" is the K^* score of the endogenous ligand. "K* PLX8394" is the K^* score of PLX8394. "Count" is number of resistance mutations at the position. "Rank" is the Pareto rank of the mutation. Note: K^* scores are in $\log_{10}$ units where possible and 0 where there is predicted to be no binding.

Position	WT AA	Mut AA	Sig Prob	K* ATP	K* PLX8394	Count	Rank
471	VAL	LEU	4.43E-04	19.67	29.38	3	1
513	ILE	PHE	1.46E-03	18.57	0	1	1
529	THR	ILE	2.02E-02	18.57	12.61	14	1
535	SER	PRO	1.30E-03	18.65	0	5	1
535	SER	LEU	1.56E-03	19.09	24.04	5	1

Position	WT AA	Mut AA	Sig Prob	K* ATP	K* PLX8394	Count	Rank
593	GLY	PHE	1.99E-06	20.07	0	16	1
593	GLY	TYR	3.58E-05	19.86	0	16	1
593	GLY	ARG	7.80E-04	16.17	-248.98	16	1
593	GLY	GLU	2.76E-04	18.73	-61.35	16	1
593	GLY	ASN	1.34E-03	19.16	-56.56	16	1
593	GLY	ASP	1.63E-02	18.89	-29.16	16	1
593	GLY	VAL	9.35E-04	19.18	6.66	16	1
593	GLY	CYS	1.66E-03	19.06	12.77	16	1
593	GLY	ILE	4.55E-05	19.8	22.76	16	1
593	GLY	SER	6.09E-02	18.83	28.13	16	1
463	ILE	TYR	2.55E-05	15.94	-260.1	5	2
481	ALA	LEU	7.06E-05	18.68	2.47	8	2
481	ALA	LYS	2.74E-04	17.32	3.05	8	2
481	ALA	GLU	1.21E-03	17.92	12.04	8	2
505	LEU	ARG	8.61E-04	18.58	24.96	4	2
508	THR	ARG	2.95E-04	18.59	-115.96	2	2
508	THR	LYS	9.16E-04	18.59	12.12	2	2
514	LEU	ILE	1.65E-03	18.4	21.52	10	2
514	LEU	ARG	5.22E-05	17.18	21.09	10	2
529	THR	PHE	3.77E-05	15.91	-106.19	14	2
529	THR	MET	1.74E-05	18.65	-28.68	14	2
529	THR	VAL	4.99E-05	18.72	28.05	14	2
529	THR	ASN	9.96E-04	18.55	27.98	14	2
532	CYS	ARG	8.12E-04	13.04	-0.71	9	2
535	SER	ILE	3.91E-04	19.02	28.17	5	2
535	SER	TYR	2.62E-03	18.68	29.19	5	2
593	GLY	HIE	1.66E-05	19.58	0	16	2
593	GLY	THR	2.67E-05	19.06	6.63	16	2
471	VAL	PHE	7.85E-04	16.37	27.32	3	3
481	ALA	ILE	5.14E-05	14.2	16.28	8	3
481	ALA	VAL	1.46E-03	17.77	27.92	8	3
514	LEU	VAL	5.22E-04	18.3	27.8	10	3
514	LEU	PRO	1.21E-03	18.12	29.31	10	3
593	GLY	HIP	1.66E-05	19.56	0	16	3
593	GLY	HID	1.66E-05	19.55	0	16	4
471	VAL	MET	8.44E-06	18.96	30.12	3	5
481	ALA	ARG	1.53E-05	17.02	-7.0	8	5
481	ALA	ASP	4.39E-06	19.21	26.65	8	5
505	LEU	TYR	8.68E-06	18.59	13.95	4	5
514	LEU	PHE	1.53E-05	18.19	0	10	5
535	SER	ARG	1.91E-06	18.84	26.28	5	5
593	GLY	TRP	4.78E-06	18.83	0	16	5
593	GLY	LEU	3.87E-07	19.35	-172.14	16	5
463	ILE	HIE	5.14E-06	18.06	14.78	5	6
463	ILE	HIP	5.14E-06	17.51	14.5	5	6
463	ILE	HID	5.14E-06	18.08	23.91	5	6
481	ALA	GLN	1.80E-06	18.37	-57.06	8	6

Position	WT AA	Mut AA	Sig Prob	K* ATP	K* PLX8394	Count	Rank
516	PHE	THR	3.77E-06	18.56	29.08	1	6
528	VAL	ARG	5.57E-07	18.61	21.75	1	6
529	THR	TYR	1.12E-06	18.58	0	14	6
529	THR	ARG	4.98E-07	18.6	-73.94	14	6
529	THR	LEU	3.10E-06	18.56	-40.25	14	6
529	THR	LYS	1.74E-06	18.57	-16.49	14	6
532	CYS	HIP	4.44E-06	14.51	10.43	9	6
532	CYS	HIE	4.44E-06	12.48	8.45	9	6
532	CYS	ILE	1.16E-06	18.82	28.7	9	6
532	CYS	VAL	2.11E-06	18.7	29.0	9	6
505	LEU	MET	3.58E-07	18.6	28.4	4	7
505	LEU	GLN	1.84E-06	18.6	29.15	4	7
514	LEU	HIP	3.37E-07	17.68	24.27	10	7
514	LEU	HIE	3.37E-07	17.72	24.93	10	7
514	LEU	HID	3.37E-07	17.72	25.66	10	7
529	THR	HIP	6.68E-08	18.58	-75.31	14	7
529	THR	HID	6.68E-08	18.58	-74.57	14	7
529	THR	HIE	6.68E-08	18.58	-73.8	14	7
529	THR	ASP	2.37E-06	18.51	27.31	14	7
529	THR	CYS	1.13E-07	18.53	29.46	14	7
531	TRP	PRO	5.44E-07	16.76	0	1	7
532	CYS	HID	4.44E-06	14.39	10.57	9	7
514	LEU	GLU	5.81E-08	17.36	26.07	10	8
514	LEU	LYS	1.68E-06	17.37	28.06	10	8