

Comprehensive mutational scanning of a kinase *in vivo* reveals substrate-dependent fitness landscapes

Alexandre Melnikov, Peter Rogov, Li Wang, Andreas Gnirke and Tarjei S. Mikkelsen*

Broad Technology Labs, Broad Institute, Cambridge, MA 02142, USA

Received April 22, 2014; Revised May 20, 2014; Accepted May 22, 2014

ABSTRACT

Deep mutational scanning has emerged as a promising tool for mapping sequence–activity relationships in proteins, ribonucleic acid and deoxyribonucleic acid. In this approach, diverse variants of a sequence of interest are first ranked according to their activities in a relevant assay, and this ranking is then used to infer the shape of the fitness landscape around the wild-type sequence. Little is currently known, however, about the degree to which such fitness landscapes are dependent on the specific assay conditions from which they are inferred. To explore this issue, we performed comprehensive single-substitution mutational scanning of APH(3')II, a Tn5 transposon-derived kinase that confers resistance to aminoglycoside antibiotics, in *Escherichia coli* under selection with each of six structurally diverse antibiotics at a range of inhibitory concentrations. We found that the resulting local fitness landscapes showed significant dependence on both antibiotic structure and concentration, and that this dependence can be exploited to guide protein engineering. Specifically, we found that differential analysis of fitness landscapes allowed us to generate synthetic APH(3')II variants with orthogonal substrate specificities.

INTRODUCTION

To enable cost-efficient generation of comprehensive and relatively unbiased amino acid substitution libraries for deep mutational scanning (1–8) and related applications, we developed a highly multiplexed approach to site-directed mutagenesis that we refer to as Mutagenesis by Integrated Tiles (MITE; Figure 1). Briefly, we design a set of long deoxyribonucleic acid (DNA) oligonucleotides that encode all desired mutations, but are otherwise homologous to the template sequence. The oligonucleotides are organized into ‘tiles’, where those within each tile differ in a central variable region but share identical 5'- and 3'-ends. The tiles

are staggered such that their variable regions collectively span the entire template. The oligonucleotides are synthesized on a programmable microarray (9) and individual tiles are then polymerase chain reaction (PCR) amplified using primers complementary to their shared ends. To avoid hybridization and extension of partially overlapping oligonucleotides, the tiles can be split into two non-overlapping pools that are synthesized and amplified separately. Finally, the PCR products from each tile are inserted into linearized plasmids that carry the remaining template sequence using multiplexed *in vitro* recombination. Because only a portion of each resulting mutant sequence is derived from the oligonucleotide pool, this approach limits the impact of synthesis errors but still benefits from the cost efficiency and flexibility of microarray-based DNA synthesis. Importantly, by allowing explicit synthesis of each desired mutation, MITE can provide both a more uniform mutational spectrum and better control over library composition than that which can be achieved by error-prone PCR or simple degenerate oligonucleotide synthesis.

In this report, we describe the application of MITE followed by sequencing to perform comprehensive single-substitution mutational scanning of the Tn5 transposon-derived aminoglycoside-3'-phosphotransferase-II (APH(3')II), a 264 amino acid residues kinase that confers resistance to a variety of aminoglycoside antibiotics (10), with the goal of elucidating which residues are essential for its activity and whether these residues are the same for different substrate structures and concentrations.

MATERIALS AND METHODS

Plasmid construction and cloning

All sequences used in this study, including gene synthesis products, oligonucleotide libraries and primers, are provided as Supplementary Data 1. Primers and individual oligonucleotides were synthesized by Integrated DNA Technologies (Coralville, IA, USA), the two 200mer oligonucleotide pools containing mutagenesis tiles were synthesized by Agilent (Santa Clara, CA, USA) as previously described (9), the wild-type (WT) APH(3')II coding region template for library construction was synthesized by GenScript USA (Piscataway, NJ, USA), and the single- and

*To whom correspondence should be addressed. Tel: +1-617-714-7000; Fax: +1-617-714-8102; Email: tarjei@broadinstitute.org

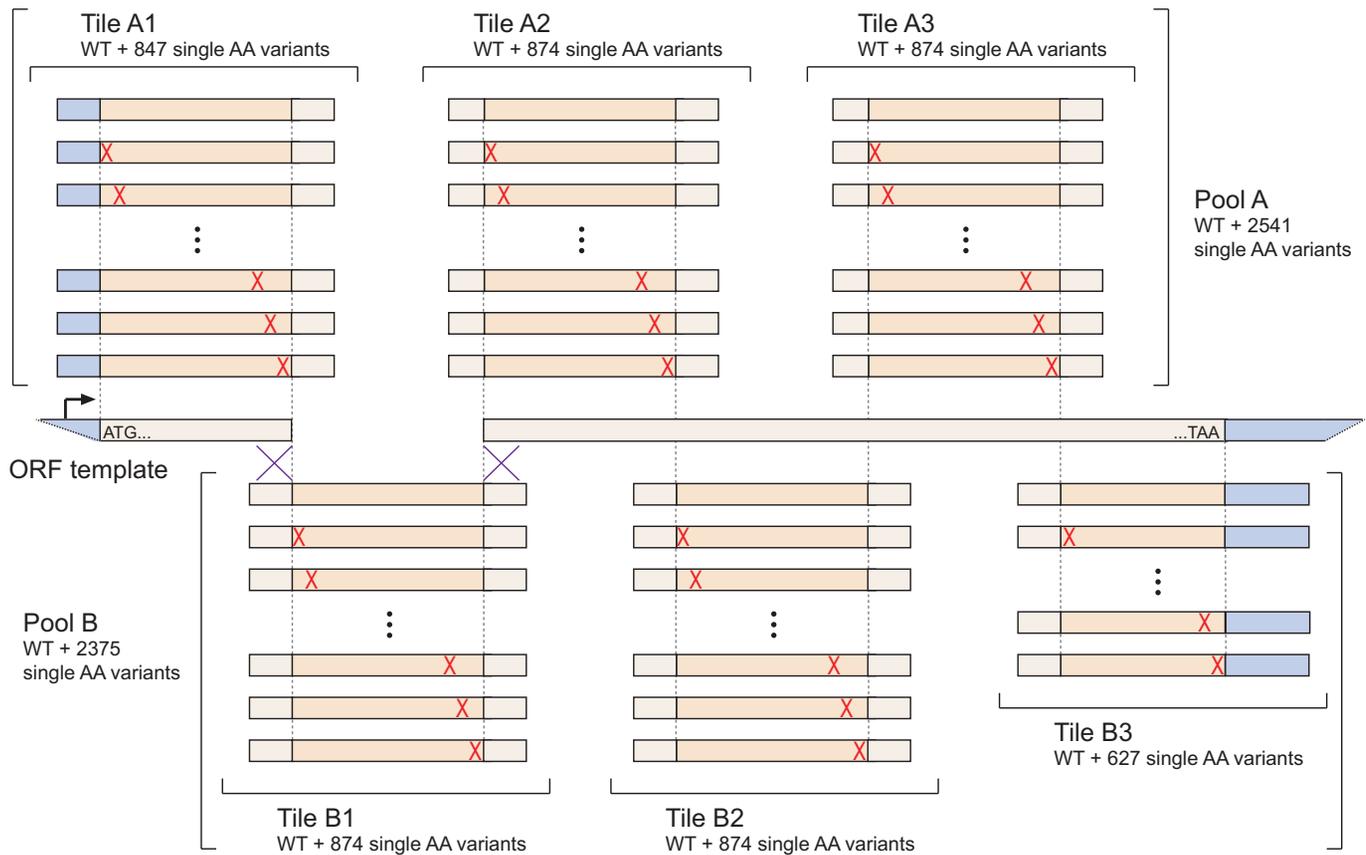


Figure 1. Overview of MITE. Oligonucleotides encoding all desired amino acid substitutions in a template open reading frame (ORF) are synthesized in two pools. The oligonucleotides in each pool are organized into non-overlapping tiles with shared 5' - and 3' -ends that facilitate selective PCR amplification. After amplification, each tile is inserted into a linearized expression vector that contains the remainder of the ORF using sequence- and ligation-independent cloning techniques.

multi-substitution synthetic ORF variants for validation of specificity-determining amino acid residues were synthesized by Gen9 (Cambridge, MA, USA).

To generate a plasmid vector carrying a constitutive EM7 promoter (Life Technologies, Carlsbad, CA, USA) and a synthetic gene encoding APH(3')II (*neo*; UniProtKB entry name: KKA2.KLEPN) flanked by two *SacI* restriction sites, we first combined a DNA fragment containing EM7 (template: EM7_pBR322; primers: EM7_Amp_F/R) with a PCR-linearized pBR322 backbone (template: pBR322; primers: pBR322_EM7_F/R) using the In-Fusion PCR Cloning System (Clontech Laboratories, Mountain View, CA, USA). The resulting plasmid (pBR322[EM7]) was then re-linearized by inverse PCR (primers: pBR322_tetRLin_F/R) and combined with a PCR amplified *neo* ORF fragment (template: KKA2.KLEPN_opt; primers: KKA2_pBR322_F/R) by In-Fusion to replace the *tetR* ORF of pBR322 with *neo*. The final product (pBR322[EM7-*neo*]) was isolated using QIAprep Spin Miniprep kits (Qiagen, Gaithersburg, MD, USA) and verified by Sanger sequencing (primers: pBR322_Seq_F/R).

To generate the APH(3')II single-substitution libraries, full-length oligonucleotides were first isolated from the synthesized pools using 10% TBE-Urea polyacrylamide

gels (Life Technologies, Carlsbad, CA, USA). Each pair of APH(3')II tiles and corresponding linearized pBR322[EM7-*neo*] plasmids were PCR amplified by Herculase II DNA Polymerase (Agilent, Santa Clara, CA, USA) with their respective primers (primer prefix: KKA2.TileAmp for tiles, KKA2.LinAmp for backbones), size selected on 1% E-Gel EX agarose gels (Life Technologies, Carlsbad, CA, USA), purified with MinElute Gel Extraction kits (Qiagen, Gaithersburg, MD, USA) and combined using In-Fusion reactions. Each reaction was separately transformed into Stellar chemically competent cells (Clontech Laboratories, Mountain View, CA, USA), the transformants were grown in LB media with carbenicillin (50 $\mu\text{g}/\text{ml}$) and plasmid DNA libraries were isolated using QIAprep Spin Miniprep kits (Qiagen, Gaithersburg, MD, USA). Finally, complete substitution libraries were generated by pooling equimolar amounts of the resulting six single-tile plasmid libraries.

To generate plasmids encoding selected single- and multiple-substitution APH(3')II variants, we amplified the Gen9 synthetic ORFs using Herculase II DNA Polymerase (primers: KKA2_pBR322_F/R) and inserted them into a PCR-linearized backbone (primers: pBR322_tetLin_F/R, template: pBR322[EM7]) using In-Fusion reactions. The resulting plasmids (pBR322[EM7-K1] through

pBR322[EM7-K44]) were verified by Sanger sequencing and preserved in *Escherichia coli* as glycerol stock.

Cell culture and selection

To determine the minimum inhibitory concentration (MIC) of each of the six aminoglycosides: kanamycin, ribostamycin, G418, amikacin, neomycin, paromomycin (Sigma-Aldrich, St Louis, MO, USA) in *E. coli* expressing WT APH(3')II, Stellar cells were transformed with pBR322[EM7-*neo*], recovered in SOC medium (New England Biolabs (NEB), Ipswich, MA, USA) at 37°C for 1 h, diluted 1:100 with LB and then divided into 96-well growth blocks containing LB with carbenicillin (50 µg/ml) and 2-fold serial dilutions of aminoglycosides. After growth at 37°C with shaking for 24 h, the culture densities were assessed by absorbance at 600 nm (A_{600}) using a NanoDrop 8000 (Thermo Scientific, Billerica, MA, USA). The MIC (minimum inhibitory concentrations) for each aminoglycoside was estimated as the lowest dilution at which A_{600} was <0.025.

To perform mutational scanning, Stellar cells were transformed with 10 ng (~3 fmol) of mutant library plasmids, recovered in SOC medium at 37°C for 1 h, diluted into 15 ml LB with carbenicillin (50 µg/ml) and one of the aminoglycosides at 1:1, 1:2, 1:4 or 1:8 dilutions of the following concentrations (estimated 1:1 WT MICs): 225 µg/ml for kanamycin, 2500 µg/ml for ribostamycin, 5 µg/ml for G418, 10 µg/ml for amikacin (see notes in Supplementary Data 2), 40 µg/ml for neomycin and 320 µg/ml for paromomycin. The cultures were incubated in 50 ml tubes at 37°C with shaking for 24 h, pelleted by centrifugation and frozen at -20°C. Plasmids were isolated from the pellets using QIAprep Spin Miniprep kits (Qiagen, Gaithersburg, MD, USA). Each transformation and selection was performed in duplicate, using each of the two independently generated mutant libraries.

To establish the substrate specificity of the synthetic APH(3')II variants relative to WT APH(3')II, glycerol stocks of the relevant clones (pBR322[EM7-K1] through pBR322[EM7-K44]) were streaked onto LB agar plates and cultured overnight at 37°C. Single colonies were then inoculated into 1 ml LB with 50 µg carbenicillin, incubated at 37°C with shaking for 6 h, diluted 1:100 into LB media, and then split into 96-well deep well plates containing LB with carbenicillin (50 µg/ml) and 2-fold serial dilutions of aminoglycosides dilution series. We note that the longer recovery in these follow-up experiments increased the absolute MICs compared to those we measured immediately following transformation).

MITE-Seq

To sequence and count mutations, the APH(3')II coding regions were first isolated from the plasmid pools by SacI digest (NEB, Ipswich, MA, USA) followed by agarose gel purification. The SacI fragments were ligated into high molecular weight concatemers using T4 DNA ligase (NEB, Ipswich, MA, USA). The concatemers were fragmented and converted to sequencing libraries using Nextera DNA Sample Prep kits (Illumina, San Diego, CA, USA). Library fragments from the 200–800 nt size range were selected using

agarose gels and then sequenced on Illumina MiSeq instruments using 2 × 150 nt reads. The reads were subsequently aligned to a reference sequence consisting of concatenated copies of the WT APH(3')II coding region using BWA version 0.5.9-r16 with default parameters (11). The number of occurrences of each amino acid at each position along the coding region was then counted and tabulated from the aligned reads. Only codons for which all three sequenced and aligned nucleotides had phred quality scores ≥30 were included in these counts.

Computational analysis

Data processing and statistical analysis, including computation of Pearson's and Spearman's correlation coefficients, χ^2 -statistics and associated *P*-values, were performed using the Enthought Python Distribution (<http://www.entthought.com>), with IPython (12), NumPy version 1.6.1, SciPy version 0.10.1 and matplotlib version 1.1.0 (13). Rendering of the APH(3')II crystal structure (PDB accession: 1ND4) was performed using PyMOL version 1.5 (Schrödinger, Portland, OR, USA). The rendering of its secondary structure was derived from PDBsum at EMBL-EBI (<http://www.ebi.ac.uk/pdbsum/>). The evolutionary conservation profile for APH(3')II was obtained from ConSurf-DB (14).

The magnitude of changes in the total abundance of mutant amino acids at each position after selection was estimated as

$$\Delta\text{Mut} = (\text{Mut}_{\text{selected}}/\text{WT}_{\text{selected}})/(\text{Mut}_{\text{input}}/\text{WT}_{\text{input}})$$

where $\text{Mut}_{\text{input}}$ and WT_{input} are, respectively, the observed counts of all mutant and WT amino acids at that position in the input library and $\text{Mut}_{\text{selected}}$ and $\text{WT}_{\text{selected}}$ are the corresponding observed counts after selection.

The magnitude of changes in the abundance of each specific amino acid at each position after selection was estimated as

$$\Delta\text{AA} = (\text{AA}_{\text{selected}}/\neg\text{AA}_{\text{selected}})/(\text{AA}_{\text{input}}/\neg\text{AA}_{\text{input}})$$

where AA_{input} and $\neg\text{AA}_{\text{input}}$ are, respectively, the observed counts of that amino acid and all other amino acids at that position in the input library and $\text{AA}_{\text{selected}}$ and $\neg\text{AA}_{\text{selected}}$ are the corresponding observed counts after selection.

The statistical significance of a deviation of any ΔMut or ΔAA from 1.0 was estimated using a χ^2 -test for independence on a 2 × 2 contingency table that contained the four corresponding counts with a pseudocount of 1 added to each. To correct for multiple hypothesis testing, the Benjamini–Hochberg procedure was applied to identify the 5% false discovery rate (FDR) threshold (15). We note that ΔMut or ΔAA values >1.0 do not necessarily indicate higher fitness relative to the WT APH(3')II (i.e. positive selection), because the majority of ORFs that contributed to the count of WT residues at one position still carried substitutions at other positions.

Substitutions that favored growth under selection with one aminoglycoside relative to another were identified by requiring $\Delta\text{AA} < 1.0$ at 5% FDR across two replicates under selection with the first and $\Delta\text{AA} \geq 1.0$ across two replicates for the same substitution under selection with the other,

as well as a minimum difference of 0.5 between the \log_{10} -transformed Δ AA values, at the indicated concentrations. These thresholds were established empirically to select a limited number of high confidence candidates.

RESULTS AND DISCUSSION

To perform mutational scanning of APH(3')II under multiple selective conditions, we first designed and synthesized six 200 nucleotide (nt) tiles that encoded all possible single amino acid substitutions across APH(3')II. Each tile contained a 140 nt variable region flanked by 30 nt constant ends. To test whether the cost-efficiency of our method could be further improved by synthesizing multiple mutant libraries in parallel, we also included corresponding tiles for nine homologous proteins, resulting in two non-overlapping pools of 26 250 and 23 666 distinct oligonucleotide sequences (Supplementary Data 1). We found that all six APH(3')II tiles could be selectively amplified from these pools with minimal optimization of PCR conditions. We then inserted these tiles into plasmids that carried the corresponding remainders of the APH(3')II coding sequence. Each amplification and multiplexed cloning reaction was performed in duplicate to generate two mutant libraries.

To characterize the resulting libraries, we first shotgun-sequenced their APH(3')II coding regions to a depth of $\sim 120\,000\times$ (Supplementary Figure S1). At this depth, we could detect 4993 of 4997 possible substitutions and found that the frequency of each mutant amino acid at each position was close to the expected value from an ideal single substitution library (observed median = 0.016%, interquartile range = 0.011–0.023%, expected = 0.020%). Amino acid mutations caused by single nucleotide substitutions were slightly overrepresented, which is likely due to the combined effects of synthesis- and sequencing-related errors. We next sub-cloned 90 plasmids from the libraries for full-length sequencing of the coding region and found that the majority (90%) of these clones carried a single amino acid substitution, while the rest carried either none (2.2%) or two substitutions (7.8%). To ensure full representation and minimize the impact of synthesis errors, each library was titrated to contain plasmids from $\sim 10^7$ transformants.

To perform mutational scanning *in vivo*, we cultured *E. coli* transformed with APH(3')II substitution libraries in liquid media supplemented with decreasing concentrations of one of six aminoglycoside antibiotics with diverse structures and a wide range of potencies: kanamycin, ribostamycin, G418, amikacin, neomycin or paromomycin (Supplementary Figure S2). To normalize the selective pressures induced by the different antibiotics, we first established their MICs for *E. coli* transformed with WT APH(3')II and then used 1:1, 1:2, 1:4 and 1:8 dilutions of WT MICs for the respective library selections. We note that, with the exception of G418 and amikacin, even the 1:8 dilutions were sufficient to inhibit growth of untransformed *E. coli*. After overnight selection, plasmids were isolated from the surviving cells and their coding regions were shotgun sequenced to compare the amino acid composition of the selected libraries to that of the original input libraries (median coverage = $51\,400\times$; Supplementary Data 2).

We began our analysis by examining changes in the relative abundance of mutant versus WT amino acids at each position after selection with kanamycin (Figure 2). At the highest concentration (1:1 WT MIC), we observed a significant depletion of mutant amino acids at ~ 210 of 263 positions (Δ Mut; χ^2 -test at the 5% FDR threshold, as determined by the Benjamini–Hochberg procedure (15)). At lower concentrations, we observed both a decrease in the number of positions with significant depletion of mutations (~ 155 , 130 and 113 at 1:2, 1:4 and 1:8 WT MIC, respectively; 5% FDR) and an increase in the diversity of amino acids that were tolerated at the remaining positions. Conservative substitutions to amino acids with biochemical properties that were similar to the WT residue were on average ~ 1.4 -fold more likely to be tolerated compared to non-conservative substitutions (10), regardless of the kanamycin concentration (Supplementary Figure S3a). The tolerance for specific substitutions appeared, however, to be both concentration- and position-dependent (Δ AA; Figure 2a). Importantly, the observed changes in amino acid frequencies were highly consistent between independent selections (Pearson's $r = 0.76$ – 0.91 across the four kanamycin dilutions; Supplementary Figure S4).

To better understand the patterns of selection, we projected them onto a previously reported co-crystal structure of APH(3')II in complex with kanamycin (16) (Δ Mut; Figure 2b). These projections show that while mutations are depleted from positions all over the protein at high concentrations, the positions that do not tolerate mutations even at low concentrations tend to cluster internally and near the active site of the enzyme. The projections also provide insights into selection patterns at specific positions. For example, residues 190, 157 and 154 are all aspartates in the vicinity of the active site in the WT enzyme (Supplementary Figure S3b). At 1:1 WT MIC, we observed depletion of all amino acids except the WT aspartates at 190 and 157, while both aspartate and glycine were tolerated at 154. At more moderate selection pressures (1:4 WT MIC), aspartate remained the only amino acid tolerated at 190, aspartate and asparagine were both tolerated at 157, while a wide variety of amino acids, including glycine, valine and arginine, were tolerated at 154. Consistent with its apparent immutability, structural and functional studies have previously identified Asp190 as the catalytic residue in APH(3')II and it is conserved across all known aminoglycoside phosphotransferases (16). The crystal structure also shows that Asp157 forms a stabilizing hydrogen bond with the second ring of kanamycin, which explains why substitutions to amino acids other than the very similar asparagine might be deleterious. In contrast, Asp154 does not appear to directly interact with the substrate, which is consistent with the higher tolerance we observed for non-conservative substitutions at this position. These results support the quantitative and qualitative accuracy of our mutational scanning data, but also suggest that the distribution of relative fitness among the mutants is highly sensitive to the conditions in which they are assayed. We note that if this is a general phenomenon, then establishing quantitative control over selection pressures will be critical to interpretation of deep mutational scanning experiments.

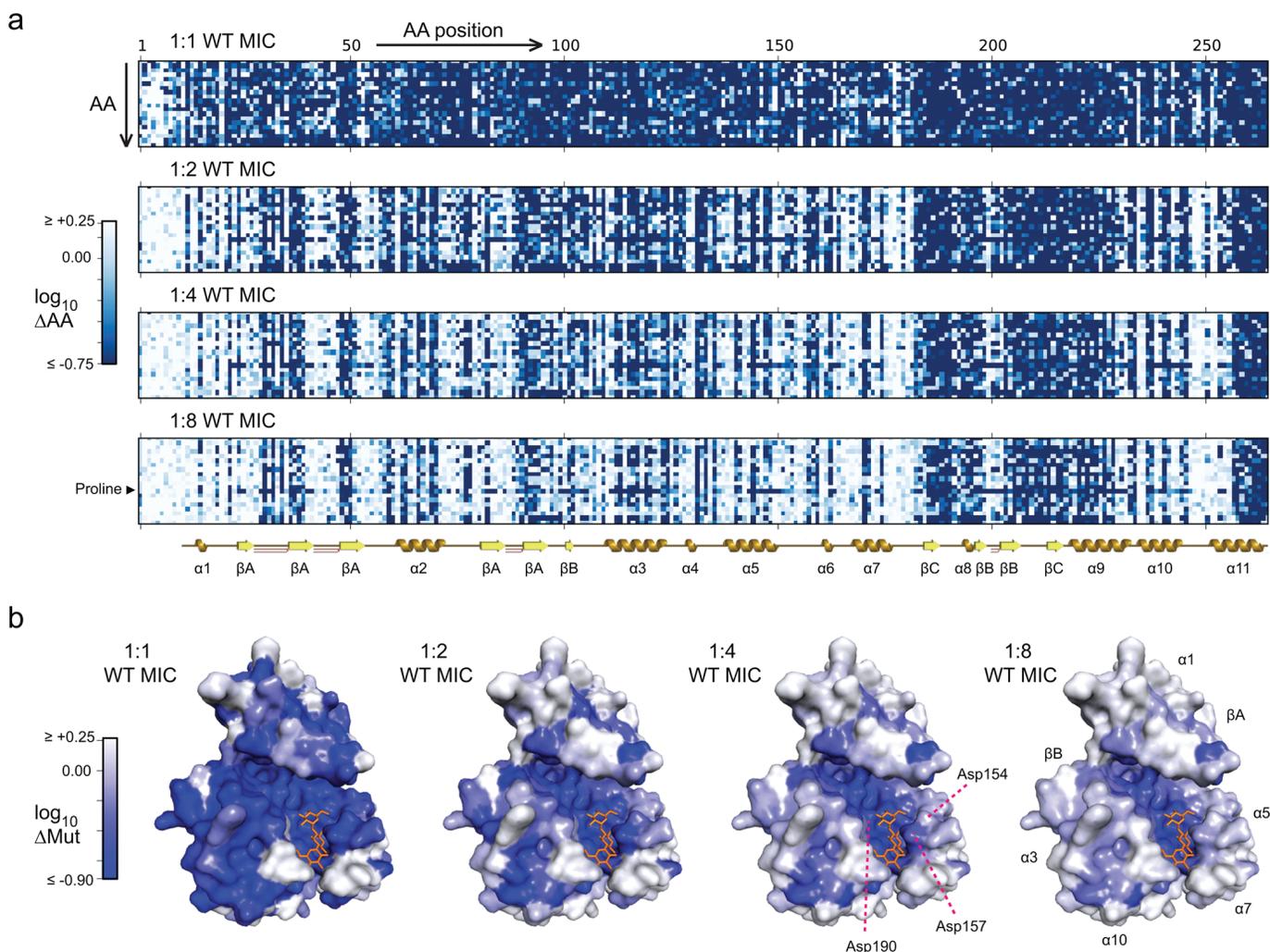


Figure 2. Mutational scanning of APH(3')II using kanamycin selection. **(a)** Visual representation of the changes in abundance of each specific amino acid (ΔAA) at each position after selection at four different concentrations. The color in each matrix entry corresponds to the change relative to the input library. The known secondary structure of APH(3')II is shown for reference, including alpha helices (gold), beta sheet strands (yellow) and beta hairpins (red; the first 10 residues are unstructured). The highly deleterious effect of proline substitutions, which impose unique structural constraints, stand out at low concentrations. **(b)** Projections of the observed changes in the abundance of mutant amino acids (ΔMut) onto a crystal structure of APH(3')II in complex with kanamycin (PDB accession: 1ND4). Numerical values are provided in Supplementary Data 2.

To compare the selection patterns induced by kanamycin in our experiments to those that have molded APH(3')II over evolutionary timescales, we examined a conservation profile derived from alignment of 133 homologs (14). We found positive rank correlations between evolutionary conservation and depletion of mutant amino acids at most positions across the protein (from Spearman's $\rho = 0.51$ at ~1:1 WT MIC to $\rho = 0.74$ at 1:8 WT MIC; Supplementary Figure S5a). The most common discrepancies were positions that were relatively conserved, yet highly tolerant of mutations under selection with kanamycin. For example, we identified 11 conserved residues that consistently tolerated non-WT amino acids even at 1:1 WT MIC (Supplementary Figure S5b). Interestingly, 7 of the 11 residues directly flank the active site, but none of them appear to interact with kanamycin in the WT enzyme (16). This may reflect that kanamycin is not the substrate, or at least not the only substrate, that has driven the natural evolution of APH(3')II,

and consequently that artificial selection with kanamycin will tolerate or even favor some substitutions that would be deleterious in other contexts.

We next examined the effects of selection using the five other aminoglycosides. These substrates generated concentration- and position-dependent selection patterns that were qualitatively similar to those generated by kanamycin (Supplementary Figures S6–S10). To quantitatively compare the patterns, we ranked the 20 amino acids by the relative changes in their abundances after selection and then computed the rank correlations at each position for each pair of conditions. We found that selection with any two aminoglycosides at matched concentrations showed positive correlations at the majority of positions (for example, median Spearman's $\rho = 0.79$ for kanamycin and paromomycin at 1:2 WT MIC and $\rho > 0$ at 235 of 264 positions at 5% FDR; Supplementary Figure S11), which implies that the relative fitness values of the mutants were largely similar.

We did, however, notice some substitutions that were consistently tolerated under selection with one substrate but depleted under selection with another. We hypothesized that further analysis of these exceptions might identify residues that influence the relative activity of APH(3')II on different substrates.

To identify such specificity-determining residues, we queried our data for individual substitutions that showed significant depletion after kanamycin selection at 1:2 WT MIC (at 5% FDR) but no trend toward depletion after selection with a second aminoglycoside at its highest concentration. These stringent criteria identified a handful of substitutions that appeared to be well-tolerated in the presence of ribostamycin, G418 or paromomycin but not kanamycin (11, 3 and 4 single substitutions, respectively; Supplementary Figure S12a; Supplementary Table S1). To confirm, we re-synthesized APH(3')II variants containing a selection of these substitutions and compared their specificities toward the second aminoglycoside to that of WT APH(3')II (here, we define specificity as the ratio of the MIC of the second aminoglycoside over the MIC of kanamycin). In six out of nine cases, the tested substitution led to a ≥ 2 -fold increase in specificity toward the second aminoglycoside (range: 2- to 8-fold; see Supplementary Table S2). Moreover, in nine of eleven additional test cases, we found that combining two or more substitutions that favored the same aminoglycoside increased the specificity to a greater extent than any one of the substitutions did alone (range: 4- to 32-fold; Supplementary Table S2). This indicates that our approach accurately identified specificity-determining residues in APH(3')II and that these residues can have additive effects. We note that the observed specificity increases were largely due to selective loss of activity on kanamycin, rather than an increase in the activity on the second substrate, which suggests that there may be limited overlap between specificity-determining and gain-of-function mutations.

To expand our search for specificity-determining residues, we next looked for substitutions that showed significant depletion after kanamycin selection at 1:2 WT MIC (at 5% FDR) but no trend toward depletion by the second aminoglycoside at its 1:2 WT MIC, or *vice versa*. These less stringent criteria identified additional candidates in every pairwise comparison (Figure 3a–e; Supplementary Table S1). Many of these candidate substitutions flanked the active site, particularly at or near residues that interact with the second and third ring of kanamycin in the WT crystal structure (in particular, the $\alpha 5$ – $\alpha 7$ and $\alpha 9$ – $\alpha 10$ helices). This has been reported as an expected hotspot for specificity-determining residues, because the first ring is constrained by being the phosphate acceptor, while the remaining rings show significant structural diversity among the aminoglycosides (16). Candidate substitutions were also frequently found in a relatively flexible part of the N-terminal domain that extends over the phosphate donor (ATP) and aminoglycoside binding sites (in particular, positions 27–57), and in the linker region between the N-terminal domain and the central core (in particular, position 96). These substitutions may influence the presentation of the two substrates near the catalytic residue, but the absence of ATP in the crystal structure limits

potential insights from the available model. To confirm the expected functional effects of the identified substitutions, we synthesized and determined the substrate-specificity of five pairs of designed APH(3')II variants, where each pair incorporated all substitutions that were preferentially tolerated by either kanamycin or the second aminoglycoside in the respective comparison (if multiple candidate substitutions were found at one position, we picked the one that showed the largest difference in abundance). Nine of the ten tested variants showed changes in specificity that favored the targeted substrate (range: 2- to 128-fold; Figure 3f and Supplementary Table S2). The 10th variant had minimal activity on both substrates, which was likely a consequence of the relatively large number of substitutions in this variant (24 compared to 12 or less in the others). Indeed, in six of the nine successful cases, the expected increase in specificity also came with a concomitant but lesser decrease in absolute activity (as measured by the relevant MIC; Supplementary Table S2), which suggests a precarious balance between the two properties. Strikingly, the two variants that were designed to favor or disfavor paromomycin relative to kanamycin showed a >2000 -fold combined change in their specificities. In this case, a single round of MITE-enabled mutational scanning was therefore sufficient to derive two synthetic enzyme variants with essentially orthogonal substrate specificities from one non-specific progenitor (Figure 3g).

In summary, MITE coupled with deep sequencing allowed us to perform comprehensive single-substitution mutational scanning of the complete APH(3')II protein coding sequence with sufficient accuracy to identify individual kinase activity- and substrate specificity-determining residues. We emphasize that structural information or molecular dynamics simulations was not required or directly utilized to identify these residues. Our approach is therefore applicable even when accurate structural models are unavailable or intractable. Assuming that indirect selection strategies can be devised, it can also be extended to proteins and activities that do not confer a direct growth advantage. MITE can readily be adapted to introduce additional types of mutations, such as small insertions or deletions, or defined neighboring substitutions.

Notably, we have recently used a similar approach to identify activity- and specificity-determining nucleotides in mammalian gene regulatory elements (8). In this previous report, we fit data from multi-substitution mutational scanning of enhancers to linear sequence–activity models, and then used these models to predict combinations of mutations that would improve enhancer activity. In both cases, we found that combining mutational scanning data from two separate conditions is an effective method for identifying mutations that change the ratio of the activities in two conditions in the desired direction. Engineering families of biomolecules with orthogonal activities is a central challenge in biotechnology and synthetic biology (2,17) due to (i) the difficulty of performing direct selection for the absence of activity in one or more conditions and (ii) the apparently limited overlap between mutations that increase activity in one condition versus those that increase specificity toward the same condition. We expect that performing parallel assays of designed mutant libraries under mul-

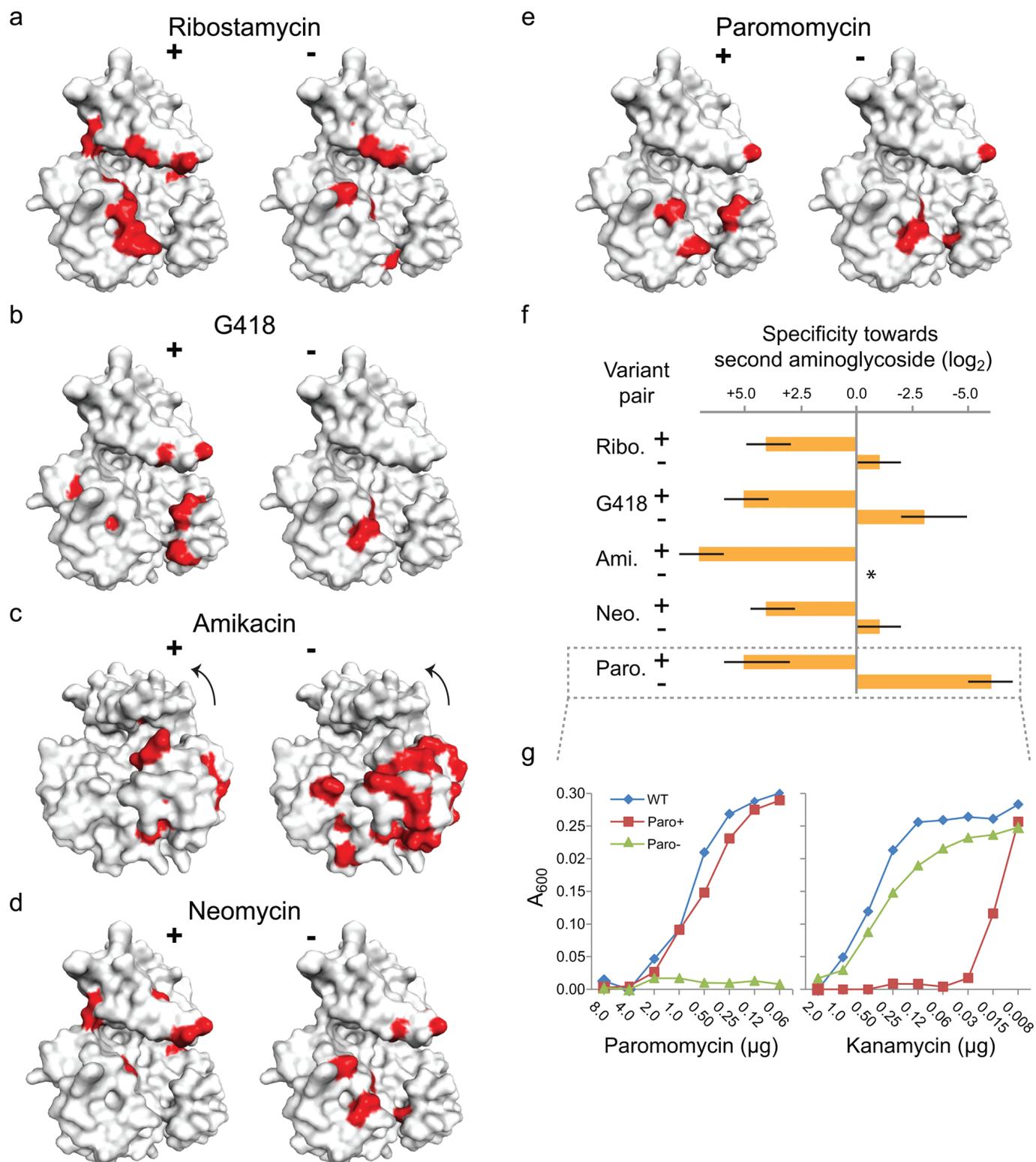


Figure 3. Identification of specificity-determining residues in APH(3')II. (a–e) The positions of substitutions that are specifically tolerated (+) or depleted (–) under selection with the indicated aminoglycoside relative to selection with kanamycin are shown in red. The identities of the substitutions are listed in Supplementary Tables S1 and S2. Arrows indicate a rotation of the structure. (f) Specificity toward the indicated aminoglycoside for pairs of synthetic enzyme variants designed to favor (+) or disfavor (–) this substrate over kanamycin, relative to that of WT APH(3')II. Bars show the medians and error bars show the ranges observed over 2–3 independent cultures, bounded by the resolution of the assay (2-fold dilutions; see also Supplementary Table S2). The variant predicted to favor kanamycin over amikacin (Ami.–) showed minimal activity on both substrates. (g) Optical density in *Escherichia coli* cultures transformed with WT APH(3')II or synthetic variants designed to specifically favor paromomycin (Paro+) or kanamycin (Paro–), after selection with each of these two aminoglycosides.

tiple conditions, followed by integrative analysis and iterative synthesis of new variants, will be a useful strategy for dissecting and engineering a wide variety of biomolecular activities and interactions.

ACCESSION NUMBERS

Illumina sequencing data has been submitted to the NCBI Sequence Read Archive (SRA) under SRA Study accession SRP042113. PDB ID: 1ND4.

SUPPLEMENTARY DATA

[Supplementary Data](#) are available at NAR Online.

ACKNOWLEDGMENT

The authors would like to thank E.M. LeProust and S. Chen for oligonucleotide library synthesis, and X. Zhang, A. Majithia, G. S. Ramachandran, P. Gatti, members of the Broad Technology Labs and members of the Lander laboratory for discussions.

FUNDING

Broad Institute. Funding for open access charge: Broad Institute.

Conflict of interest statement. None declared.

REFERENCES

- Fowler,D.M., Araya,C.L., Fleishman,S.J., Kellogg,E.H., Stephany,J.J., Baker,D. and Fields,S. (2010) High-resolution mapping of protein sequence-function relationships. *Nat. Methods*, **7**, 741–746.
- Araya,C.L. and Fowler,D.M. (2011) Deep mutational scanning: assessing protein function on a massive scale. *Trends Biotechnol.*, **29**, 435–442.
- Hietpas,R.T., Jensen,J.D. and Bolon,D.N. (2011) Experimental illumination of a fitness landscape. *Proc. Natl. Acad. Sci. U.S.A.*, **108**, 7896–7901.
- Hinkley,T., Martins,J., Chappey,C., Haddad,M., Stawiski,E., Whitcomb,J.M., Petropoulos,C.J. and Bonhoeffer,S. (2011) A systems analysis of mutational effects in HIV-1 protease and reverse transcriptase. *Nat. Genet.*, **43**, 487–489.
- Pitt,J.N. and Ferré-D'Amaré,A.R. (2010) Rapid construction of empirical RNA fitness landscapes. *Science*, **330**, 376–379.
- Patwardhan,R.P., Lee,C., Litvin,O., Young,D.L., Pe'er,D. and Shendure,J. (2009) High-resolution analysis of DNA regulatory elements by synthetic saturation mutagenesis. *Nat. Biotechnol.*, **27**, 1173–1175.
- Kinney,J., Murugan,A., Callan,C.G. and Cox,E.C. (2010) Using deep sequencing to characterize the biophysical mechanism of a transcriptional regulatory sequence. *Proc. Natl. Acad. Sci. U.S.A.*, **107**, 9158–9163.
- Melnikov,A., Murugan,A., Zhang,X., Tesileanu,T., Wang,L., Rogov,P., Feizi,S., Gnirke,A., Callan,C.G., Kinney,J.B. *et al.* (2012) Systematic dissection and optimization of inducible enhancers in human cells using a massively parallel reporter assay. *Nat. Biotechnol.*, **30**, 271–277.
- LeProust,E.M., Peck,B.J., Spirin,K., McCuen,H.B., Moore,B., Namsaraev,E. and Caruthers,M.H. (2010) Synthesis of high-quality libraries of long (150mer) oligonucleotides by a novel depurination controlled process. *Nucleic Acids Res.*, **38**, 2522–2540.
- Betts,M.J. and Russell,R.B. (2003) Amino acid properties and consequences of substitutions. In: Barnes,MR and Gray,IC (eds), *Bioinformatics for Geneticists*. Wiley, Hoboken, pp. 218–316.
- Li,H. and Durbin,R. (2009) Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics*, **25**, 1754–1760.
- Pérez,F. and Granger,B.E. (2007) IPython: a system for interactive scientific computing. *Comput. Sci. Eng.*, **9**, 21–29.
- Hunter,J.D. (2007) Matplotlib: a 2D graphics environment. *Comput. Sci. Eng.*, **9**, 90–95.
- Goldenberg,O., Erez,E., Nimrod,G. and Ben-Tal,N. (2008) The ConSurf-DB: pre-calculated evolutionary conservation profiles of protein structures. *Nucleic Acids Res.*, **37**, D323–D327.
- Benjamini,Y. and Hochberg,Y. (1995) Controlling the false discovery rate: a practical and powerful approach to multiple testing. *J. R. Stat. Soc. Ser. B.*, **57**, 289–300.
- Nurizzo,D., Shewry,S.C., Perlin,M.H., Brown,S.A., Dholakia,J.N., Fuchs,R.L., Deva,T., Baker,E.N. and Smith,C.A. (2003) The crystal structure of Aminoglycoside-3'-Phosphotransferase-IIa, an enzyme responsible for antibiotic resistance. *J. Mol. Biol.*, **327**, 491–506.
- Slusarczyk,A.L., Lin,A. and Weiss,R. (2012) Foundations for the design and implementation of synthetic genetic circuits. *Nat. Rev. Genet.*, **13**, 406–420.