

Video Article

# A Protocol for Functional Assessment of Whole-Protein Saturation Mutagenesis Libraries Utilizing High-Throughput Sequencing

Michael Allen Stiffler<sup>1</sup>, Subu K Subramanian<sup>1</sup>, Victor H Salinas<sup>1</sup>, Rama Ranganathan<sup>1</sup>

<sup>1</sup>Green Center for Systems Biology, University of Texas Southwestern Medical Center

Correspondence to: Michael Allen Stiffler at [Michael.Stiffler@utsouthwestern.edu](mailto:Michael.Stiffler@utsouthwestern.edu), Rama Ranganathan at [rama.ranganathan@utsouthwestern.edu](mailto:rama.ranganathan@utsouthwestern.edu)

URL: <https://www.jove.com/video/54119>

DOI: [doi:10.3791/54119](https://doi.org/10.3791/54119)

Keywords: Molecular Biology, Issue 113, mutagenesis, saturation mutagenesis, next generation sequencing, high throughput sequencing, TEM-1 beta-lactamase, antibiotic resistance, orthogonal primers

Date Published: 7/3/2016

Citation: Stiffler, M.A., Subramanian, S.K., Salinas, V.H., Ranganathan, R. A Protocol for Functional Assessment of Whole-Protein Saturation Mutagenesis Libraries Utilizing High-Throughput Sequencing. *J. Vis. Exp.* (113), e54119, doi:10.3791/54119 (2016).

## Abstract

Site-directed mutagenesis has long been used as a method to interrogate protein structure, function and evolution. Recent advances in massively-parallel sequencing technology have opened up the possibility of assessing the functional or fitness effects of large numbers of mutations simultaneously. Here, we present a protocol for experimentally determining the effects of all possible single amino acid mutations in a protein of interest utilizing high-throughput sequencing technology, using the 263 amino acid antibiotic resistance enzyme TEM-1  $\beta$ -lactamase as an example. In this approach, a whole-protein saturation mutagenesis library is constructed by site-directed mutagenic PCR, randomizing each position individually to all possible amino acids. The library is then transformed into bacteria, and selected for the ability to confer resistance to  $\beta$ -lactam antibiotics. The fitness effect of each mutation is then determined by deep sequencing of the library before and after selection. Importantly, this protocol introduces methods which maximize sequencing read depth and permit the simultaneous selection of the entire mutation library, by mixing adjacent positions into groups of length accommodated by high-throughput sequencing read length and utilizing orthogonal primers to barcode each group. Representative results using this protocol are provided by assessing the fitness effects of all single amino acid mutations in TEM-1 at a clinically relevant dosage of ampicillin. The method should be easily extendable to other proteins for which a high-throughput selection assay is in place.

## Video Link

The video component of this article can be found at <https://www.jove.com/video/54119/>

## Introduction

Mutagenesis has long been employed in the laboratory to study the properties of biological systems and their evolution, and to produce mutant proteins or organisms with enhanced or novel functions. While early approaches relied on methods which produce random mutations in organisms, the advent of recombinant DNA technology enabled researchers to introduce select changes to DNA in a site-specific manner, *i.e.*, site-directed mutagenesis<sup>1,2</sup>. With current techniques, typically using mutagenic oligonucleotides in a polymerase chain reaction (PCR), it is relatively facile to create and assess small numbers of mutations (*e.g.*, point mutations) in a given gene<sup>3,4</sup>. It is far more difficult however when the goal approaches, for example, the creation and assessment of all possible single-site (or higher-order) mutations.

While much has been learned from early studies attempting to assess large numbers of mutations in genes, the techniques used were often laborious, for example requiring the assessment of each mutation independently using nonsense suppressor strains<sup>5-7</sup>, or were limited in their quantitative ability due to the low sequencing depth of Sanger sequencing<sup>8</sup>. The techniques used in these studies have largely been supplanted by methods utilizing high-throughput sequencing technology<sup>9-12</sup>. These conceptually simple approaches entail creating a library comprising a large number of mutations, subjecting the library to a screen or selection for function, and then deep-sequencing (*i.e.*, on the order of  $>10^6$  sequencing reads) the library obtained before and after selection. In this way, the phenotypic or fitness effects of a large number of mutations, represented as the change in population frequency of each mutant, can be assessed simultaneously and more quantitatively.

We previously introduced a simple approach for assessing libraries of all possible single amino acid mutations in proteins (*i.e.*, whole-protein saturation mutagenesis libraries), applicable to genes with a length longer than the sequencing read length<sup>11,13</sup>. First, each amino acid position is randomized by site-directed mutagenic PCR. During this process, the gene is split into groups composed of contiguous positions with a total length accommodated by the sequencing platform. The mutagenic PCR products for each group are then combined, and each group independently subjected to selection and high-throughput sequencing. By maintaining a correspondence between the location of mutations in the sequence and the sequencing read length, this approach has the advantage of maximizing sequencing depth: while one could simply sequence such libraries in short windows without splitting into groups (*e.g.*, by a standard shotgun sequencing approach), most reads obtained would be wild-type and thus the majority of sequencing throughput wasted (*e.g.*, for a whole-protein saturation mutagenesis library of a 500 amino acid protein sequenced in 100 amino acid (300 bp) windows, at minimum 80% of reads will be the wild-type sequence).

Here, a protocol is presented which utilizes high-throughput sequencing for the functional assessment of whole-protein saturation mutagenesis libraries, using the above approach (outlined in **Figure 1**). Importantly, we introduce the usage of orthogonal primers in the library cloning process to barcode each sequence group, which allows them to be multiplexed into one library, subjected simultaneously to screening or selection, and then de-multiplexed for deep sequencing. Since the sequence groups are not subjected to selection independently, this reduces the workload and ensures that each mutation experiences the same level of selection. TEM-1  $\beta$ -lactamase, an enzyme which confers high-level resistance to  $\beta$ -lactam antibiotics (e.g., ampicillin) in bacteria is used as a model system<sup>14-16</sup>. A protocol is described for the assessment of a whole-protein saturation mutagenesis library of TEM-1 in *E. coli* under selection at an approximate serum level for a clinical dose of ampicillin (50  $\mu$ g/ml)<sup>17,18</sup>.

## Protocol

Note: See **Figure 1** for outline of protocol. Several steps and reagents in the protocol require safety measures (indicated with "CAUTION"). Consult material safety data sheets before use. All protocol steps are performed at RT unless other indicated.

## 1. Prepare Culture Media and Plates

1. Prepare and sterilize by autoclaving 1 L purified water, 100 ml Super Optimal Broth (SOB; **Table 1**), 1 L Luria-Bertani broth (LB; **Table 2**) and 1 L LB-agar (**Table 3**). Prepare separately and sterilize three culture flasks each containing 1 L LB.  
Note: Throughout the protocol "water" refers to autoclave-sterilized purified water; SOB, LB and LB-agar refer to the autoclave-sterilized solutions.
2. Prepare a 12 mg/ml stock of Tet by dissolving 0.12 g tetracycline hydrochloride in 10 ml of 70% ethanol. Sterilize using a 0.2  $\mu$ m filter and store at 4 °C protected from light.
3. Cool LB-agar to 50 °C and then add 1 ml of Tet stock (final concentration of 12  $\mu$ g/ml Tet). Pour into petri plates and cool at RT protected from light. Store at 4 °C protected from light.

## 2. Construction of the Whole-gene Saturation Mutagenesis Library

Note: Primers; completed PCRs, restriction digests and ligations; and purified DNA samples can be stored at -20 °C.

### 1. Designing mutagenesis primers

1. To mutagenize each amino acid position to all possible amino acids, design a pair of complementary mutagenesis primers (sense/forward and antisense/reverse) for each amino acid position with the following guidelines:
  1. Replace the codon corresponding to the amino acid to be mutagenized by NNS (where N is a mixture of all four nucleotide bases and S is a mixture of cytosine (C) and guanine (G)) and center in the primer, flanked by approximately 15 nucleotides on each side.
  2. Ensure that the 5' and 3' ends terminate in C or G and that the melting temperature ( $T_m$ ) is approximately 70 °C<sup>3</sup>. Use the computational script NNS\_PrimerDesign.m (See Supplementary Code File) to design NNS mutagenesis primers according to these guidelines.
2. Order primers from a commercial source. For ease of use, have them synthesized in 96-well plate format and pre-diluted in water to 50  $\mu$ M, with one set of plates containing the sense mutagenesis primers and another the antisense primers.
3. Fill a pipette basin with water and use a multichannel pipette to transfer 95  $\mu$ l to 263 wells over three 96-well plates. Dilute the primers 20-fold to 2.5  $\mu$ M by using a multichannel pipette to transfer 5  $\mu$ l from the 96-well plates containing the sense mutagenesis primers to the cognate wells of the plates containing water.
4. Repeat protocol step 2.1.3 to dilute the antisense mutagenesis primers.

### 2. Synthesis of NNS sub-libraries for each amino acid position by two-step PCR site-directed mutagenesis

1. Perform the first-round mutagenic PCRs. For each mutagenesis primer, prepare a 25  $\mu$ l PCR reaction using pBR322\_AvrII plasmid as template and primers AatII\_F or AvrII\_R (sense mutagenesis primers paired with primer AvrII\_R, and antisense mutagenesis primers paired with primer AatII\_F; total of 526 PCRs). See **Table of Materials** for AatII\_F and AvrII\_R sequences.
  1. Prepare a PCR "master mix" by adding the reagents from **Table 4** to a 15 ml conical tube. Transfer to a pipette basin. Use a multichannel pipette to transfer 15  $\mu$ l to 263 wells over three 96-well PCR plates. Use a multichannel pipette to transfer 10  $\mu$ l from the 96-well plates containing the diluted sense mutagenesis primers to the cognate wells in the PCR plates.
  2. Cover each PCR plate with a 96-well plate seal. Centrifuge at 200 x g for 2 min.
  3. Transfer the PCR plates to thermocycler and run the following program: 98 °C for 30 sec; 20 cycles: 98 °C for 10 sec, 55 °C for 20 sec, 72 °C for 1 min; 72 °C for 2 min; hold at 4 °C.
  4. Repeat protocol steps 2.2.1.1 - 2.2.1.3 for the 96-well plates containing the diluted antisense mutagenesis primers.
2. Perform the second-round mutagenic PCRs. For each amino acid position, prepare a 25  $\mu$ l PCR reaction using primers AatII\_F and AvrII\_R, and the mixed and diluted first-round mutagenic PCR products as a template (total of 263 PCRs).
  1. Fill a pipette basin with water and use a multichannel pipette to transfer 198  $\mu$ l to 263 wells over three 96-well plates.
  2. Combine and dilute 100-fold the mutagenic PCR products for each amino acid position by using a multichannel pipette to first transfer 1  $\mu$ l from the 96-well PCR plates containing the PCR products resulting from the sense mutagenesis primers to the cognate wells of the plates containing water. Then repeat the transfer for the PCR products resulting from the antisense mutagenesis primers.
  3. Prepare a PCR "master mix" by adding the reagents from **Table 5** to a 15 ml conical tube. Transfer to a pipette basin. Use a multichannel pipette to transfer 24  $\mu$ l to 263 wells over three 96-well PCR plates. Use a multichannel pipette to transfer 1  $\mu$ l from the 96-well plates containing the mixed and diluted first-round mutagenic PCR products to the cognate wells in the PCR plates.

4. Cover each PCR plate with a 96-well plate seal. Centrifuge at approximately 200 x g for 2 min. Transfer plates to thermocycler and run the same program as in protocol step 2.2.1.3.
3. Analyze results of the second-round mutagenic PCRs by gel electrophoresis. Ensure that all products are of the correct size and absent of contaminating products.
  1. Add 2 ml of 2x gel loading dye to a pipette basin and then use a multichannel pipette to transfer 6 µl to 263 wells over three 96-well plates. Use a multichannel pipette to transfer 6 µl from the 96-well PCR plates containing the second-round mutagenic PCR products to the cognate wells of the 96-well plates containing dye.
  2. Prepare a 1.5% agarose gel with 0.2 µg/ml ethidium bromide (CAUTION).
  3. Load DNA ladder in first and last lanes of each row. Then use a multichannel pipette to load 10 µl of samples from protocol step 2.2.3.1.
  4. Run the gel at 100 V for 40 min and image on a UV transilluminator.
  5. Repeat protocol steps 2.2.3.2 - 2.2.3.4 until all samples are analyzed.
4. Accurately measure the concentration of each NNS sub-library PCR product using a dsDNA quantitation reagent.
  1. Transfer approximately 15 ml EB buffer to a pipette basin. Use a multichannel pipette to transfer 49 µl to 263 wells over three 96-well black-walled, clear bottom assay plates.
  2. Use a multichannel pipette to transfer 1 µl of each second-step PCR product (protocol step 2.2.2) to the cognate wells of the 96-well assay plates.
  3. Prepare a DNA concentration standard curve by diluting lambda phage DNA to 2 ng/µl in 300 µl EB buffer and then make ten two-fold dilutions (for total of 11 concentrations). Transfer 50 µl to first eleven columns of a row of one of the 96-well assay plates from the previous step which contains no sample; to the twelfth column add 50 µl of EB buffer (reagent blank).
  4. Prepare dsDNA quantitation reagent by adding 75 µl of reagent (see **Table of Materials**) to a 15 ml conical tube, then add 15 ml of EB buffer. Mix by inverting tube and then transfer to a pipette basin. Protect reagent from light.
  5. Use a multichannel pipette to transfer 50 µl of prepared dsDNA quantitation reagent to each well of the assay plates. Mix by pipetting up-and-down. Incubate plates at RT for 5 min protected from light.
  6. Measure fluorescence of each sample using a microplate reader and standard fluorescein wavelengths (excitation 485 nm, emission 520 nm; 0.1 sec).
  7. Subtract the fluorescence value of the reagent blank from all the samples. Generate a standard curve from the fluorescence measurements of lambda phage samples. Calculate the concentration of each sample using their respective fluorescence measurements and the standard curve.

### 3. Cloning of NNS sub-libraries into selection vectors

1. Mix 100 ng of each NNS sub-library PCR product into five NNS sub-library groups. Following manufacturers' instructions, clean up samples using a DNA purification kit and then measure concentration using a dsDNA quantitation reagent.  
Note: Each group is composed of approximately 53 contiguous amino acids positions spaced along the TEM-1 sequence (NNS sub-library groups 1 - 5 are comprised of positions 26 - 78, 79 - 132, 133 - 183, 184 - 236, and 237 - 290, respectively; numbering according to Ambler *et al.*<sup>19</sup>).
2. Create cloning vectors for each NNS sub-library group.
  1. Prepare five 100 µl PCRs according to **Table 6**, using primers AvrII\_F and AatII\_OP1\_R - AatII\_OP5\_R, and plasmids pBR322\_OP1-5 as template (AatII\_OP1\_R paired with pBR322\_OP1, etc.).
  2. Transfer to thermocycler and run the following program: 98 °C for 30 sec; 25 cycles: 98 °C for 10 sec, 55 °C for 20 sec, 72 °C for 1.5 min; 72 °C for 2 min; hold at 4 °C. See **Table of Materials** for sequences of AatII\_R and AvrII\_F.
  3. Prepare a 1% agarose gel with 0.2 µg/ml ethidium bromide (CAUTION).
  4. Add 20 µl of 6x gel loading dye to each PCR sample. Load the first lane of the gel with DNA ladder; load entire volume of each sample, skipping at least one well between samples.
  5. Run gel at 100 V for 50 min.
  6. Visualize gel using a long-wavelength UV illuminator (CAUTION). Excise slices containing the PCR product at ~3,500 bp; transfer to separate microfuge tubes. Gel slices can be stored at -20 °C.
  7. Following manufacturers' instructions, purify samples using a gel extraction kit and measure concentration using a dsDNA quantitation reagent.
3. For both the NNS sub-library groups (protocol step 2.3.1) and cloning vectors (protocol step 2.3.2), set up restriction digests with AatII and AvrII enzymes according to **Table 7**. Incubate at 37 °C for 1 hr. Following manufacturers' instructions, clean up samples using a DNA purification kit and then measure concentration using a dsDNA quantitation reagent.
4. Set up ligation reactions following **Table 8** for each restriction-digested NNS sub-library group with cognate restriction-digested cloning vector (NNS sub-library group 1 with pBR322\_OP1, etc.). Incubate at RT for 1 hr. Clean up reactions using a DNA purification kit according to manufacturer's instructions; elute DNA with 20 µl of water.
5. Transform the entirety of the purified ligation reactions into library-efficient *E. coli* cells by electroporation.
  1. Thaw electrocompetent *E. coli* cells and then place cells and purified ligation reactions on ice.
  2. Transfer 10 µl thawed cells to each purified ligation reaction and then transfer to electroporation cuvette. Electroporate at 1.8 kV.
  3. Recover cells by resuspending in 1 ml SOB. Incubate for 1 hr at 37 °C.
  4. Resuspend 10 µl of each recovery culture in 990 µl LB; spread 100 µl on LB-agar plates containing 12 µg/ml Tet. Incubate plates O/N (~16 hr) at 37 °C.
  5. For each recovery culture, prepare a 250 ml culture flask with 50 ml LB and 50 µl Tet stock. Transfer to flask the remaining ~1 ml of recovery culture. Incubate O/N (~16 hr) at 37 °C with vigorous shaking (~200 rpm).
6. Count the number of colonies on each plate. Calculate the number of successful transformants as  $N * V^C / V^P$ , where  $N$  is the number of colonies,  $V^C$  is the recovery culture volume (1,000 µl) and  $V^P$  is volume of the recovery culture plated (1 µl).

Note: To ensure complete coverage of all mutations, as a rule of thumb the number of successful transformants should be  $\geq 100$ -fold over the number of expected mutations. Each NNS sub-library has  $\sim 53$  positions, so the expected number of mutations is 53 positions  $\times$  32 codons/position  $\approx 1.7 \times 10^3$ ; to give a library size  $\geq 100$ -fold ( $\geq 1.7 \times 10^5$ ) there should be  $\geq 170$  colonies on each plate.

7. According to manufacturer's instructions, isolate plasmid DNA from cultures using a plasmid purification kit and then measure concentrations using a dsDNA quantitation reagent. Mix together 100 ng of each plasmid. This creates the final whole-protein saturation mutagenesis library.

### 3. Selection of the TEM-1 Whole-protein Saturation Mutagenesis Library for Antibiotic Resistance

#### 1. Preparation of the pre-selection culture.

1. Dilute the plasmid from protocol step 2.3.7 to 0.5 ng/ $\mu$ l in water and transfer 20  $\mu$ l to a microcentrifuge tube. Perform transformation, recovery, plating and O/N growth as previously described in protocol step 2.3.5, except transfer 1  $\mu$ l of the SOB recovery culture to 999  $\mu$ l LB.
2. Count the number of colonies. To ensure complete coverage of all mutations there should be  $\geq 100$  colonies, indicating  $\geq 10^6$  successful transformants ( $100 \times 263$  positions  $\times$  32 codons/position  $\approx 10^6$ ).
3. Measure the concentration of the 50 ml O/N culture.
  1. Prepare an LB blank by adding 1 ml LB to a spectrophotometer cuvette. Measure OD600 on a spectrophotometer.
  2. Dilute the O/N culture 10-fold by resuspending 100  $\mu$ l in 900  $\mu$ l LB. Measure the OD600. Subtract the OD600 reading of the blank and multiply by 10 to give the OD600 of the O/N culture.
4. Pre-warm the three culture flasks from protocol step 1.1 for  $\sim 30$  min at 37 °C. Dilute the O/N culture to OD600 = 0.1 and add 1 ml to one flask (final OD600 = 0.001). This is the "pre-selection culture".
5. Incubate the "pre-selection culture" at 37 °C with vigorous shaking (200 rpm). Periodically monitor growth by measuring OD600 as in protocol step 3.1.3 (it is not necessary to dilute culture 10-fold) until OD600 = 0.1 ( $\sim 2.5$  hr).
6. Transfer 100 ml of the pre-selection culture to two 50 ml conical tubes. Centrifuge at 4,000  $\times$  g for 6 min at 4 °C. Remove most of supernatant and combine into a single 15 conical tube. Repeat centrifugation and remove all supernatant. Store at -20 °C.

#### 2. Selection for ampicillin resistance.

1. While the pre-selection culture is incubating, prepare a 50 mg/ml stock of Amp in water by dissolving 0.5 g sodium ampicillin in 10 ml water. Sterilize using a 0.2  $\mu$ m filter and store at 4 °C.
2. To the other two flasks, add Tet to a final concentration of 12  $\mu$ g/ml and a volume of the pre-selection culture such that the final OD600 = 0.001. To one flask, add 1 ml Amp, for a final concentration of 50  $\mu$ g/ml - this is the "selection culture".
3. Incubate the cultures at 37 °C with vigorous shaking (200 rpm). Monitor growth of the culture for which no ampicillin was added in the previous step, until OD600 = 0.1 ( $\sim 2.5$  hr). At this time, also measure the OD600 of the selection culture.
4. Divide the OD600 of the selection culture into 0.1 and multiply by 100 ml. Transfer this volume ( $\sim 400$  ml) to 50 ml conical tubes and centrifuge at 4,000  $\times$  g for 6 min at 4 °C. Remove most of supernatant and combine into a single 15 conical tube. Repeat centrifugation and remove all supernatant.
5. According to manufacturer's instructions, isolate plasmid DNA from the pre-selection (protocol step 3.1.6) and selection (protocol step 3.2.4) culture cell pellets and then measure concentrations using a dsDNA quantitation reagent.

### 4. High-throughput Sequencing to Determine the Fitness Effects of Mutations

#### 1. Preparation of samples for high-throughput sequencing

1. Prepare 25  $\mu$ l PCRs to de-multiplex the NNS sub-library groups with orthogonal primers
  1. Prepare a PCR master mix according to **Table 9**; transfer 23  $\mu$ l to ten PCR tubes.
  2. Add 1  $\mu$ l of 0.5 ng/ $\mu$ l purified plasmid DNA from the pre-selection culture to PCR tubes 1 - 5 and from the selection culture to tubes 6 - 10.
  3. Mix together 50  $\mu$ l of 50  $\mu$ M forward orthogonal primers OP1\_F - OP5\_F with the respective reverse orthogonal primers OP1\_R - OP5\_R. In the same order, transfer 1  $\mu$ l to PCR tubes 1 - 5 and 6 - 10. See **Table of Materials** for sequences of OP1\_F - OP5\_F and OP1\_R - OP5\_R.
  4. Transfer PCR tubes to thermocycler. Run the following program: 98 °C for 30 sec; 20 cycles: 98 °C for 10 sec, 55 °C for 20 sec, 72 °C for 1.5 min; 72 °C for 2 min; hold at 4 °C.
2. Prepare 25  $\mu$ l PCRs to isolate each of the NNS sub-library groups
  1. Dilute 100-fold the ten PCRs from protocol step 4.1.1.4 by transferring 1  $\mu$ l of each to separate PCR tubes and adding 99  $\mu$ l of water. Mix, then pipette out 99  $\mu$ l and discard.
  2. Mix together 50  $\mu$ l of 50  $\mu$ M forward primers Group1\_F - Group5\_F with the respective reverse primers Group1\_R - Group5\_R. In the same order, transfer 1  $\mu$ l to PCR tubes 1 - 5 and 6 - 10. See **Table of Materials** for sequences of Group1\_F - Group5\_F and Group1\_R - Group5\_R.
  3. Prepare a PCR master mix according to **Table 9**, transfer 23  $\mu$ l to each PCR tube. Transfer PCR tubes to thermocycler; run the same program as in protocol step 2.2.1.3.
3. Carry out the final 25  $\mu$ l PCRs to add indexing sequences
  1. Dilute 100-fold the ten PCRs from protocol step 4.1.2.3 by transferring 1  $\mu$ l of each to separate PCR tubes and adding 99  $\mu$ l of water. Mix, then pipette out 99  $\mu$ l and discard.
  2. Prepare a PCR master mix according to **Table 9**, transfer 23  $\mu$ l to each PCR tube.

3. For tubes with template originating from NNS sub-library groups 1-5, transfer 0.5 µl per tube forward primers 501\_F - 505\_F respectively. For tubes with template resulting from the pre-selection and selection cultures, transfer 0.5 µl per tube reverse primers 701\_R and 702\_R, respectively. See **Table of Materials** for sequences of 501\_F - 505\_F, and 701\_R and 702\_R.
4. Transfer PCR tubes to the thermocycler and run the program from step 2.2.1.3.
4. Mix and purify samples
  1. Measure concentrations using a dsDNA quantitation reagent according to manufacturer's instructions. Mix 100 ng of each PCR product into a single microcentrifuge tube.
  2. Prepare a 2% agarose gel with 0.2 µg/ml ethidium bromide (CAUTION).
  3. Add 6x gel loading dye to the mixed PCR products sample. Load the first lane of the gel with DNA ladder; load entire volume of the sample.
  4. Run gel at 100 V for 50 min. Visualize gel using a long-wavelength UV illuminator (CAUTION). Excise slice containing the PCR product at ~360 bp; transfer to microfuge tube. Gel slice can be stored at -20 °C.
  5. Following manufacturers' instructions, purify sample using a gel extraction kit and measure concentration using a dsDNA quantitation reagent. This is the final sample for high-throughput sequencing.
5. Sequence on a high-throughput sequencing platform (see **Table of Materials** for platform used in this protocol).
  1. Calculate sample concentration in nM as  $\left[\frac{(ng/10^3)}{(bp * 660)}\right] * 10^9$ , where  $ng$  is the concentration of the sample in ng/µl and  $bp$  is the sequence length of the sample DNA (~360 bp). Dilute sample to 4 nM in EB buffer.
  2. Follow manufacturer's instructions to denature sample and dilute to 9 pM in hybridization buffer.
  3. Load 600 µl of sample into reagent cartridge. Sequence following manufacturer's instructions and on-screen prompts.

## 2. Analysis of sequencing data

1. Download FLASH (Fast Length Adjustment of Short reads)<sup>20</sup>, place into folder along with fastq.gz files obtained from sequencing.
2. Use FLASH to join the fastq.gz files corresponding to the paired-end reads for each pair of indices (forward and reverse reads for each NNS sub-library group for the pre-selection and selection cultures).
  1. Open Command Prompt and change directory to folder in protocol step 4.2.1. Join each pair of reads using command: flash <mates1.fastq.gz> <mates2.fastq.gz>, where mates1.fastq.gz and mates2.fastq.gz are the files containing the forward and reverse reads, respectively.
3. After joining each pair of reads, place the out.extendedFragments.fastq output file into separate folders for results from the pre-selection or selection cultures. Rename the out.extendedFragments.fastq output file according to the NNS sub-library group to which it corresponds (*i.e.*, 1.fastq, 2.fastq, etc.).
4. Run the computational script NNS\_DataAnalyzer.m (See Supplementary Code File) from each folder to compute the counts for each single amino acid mutation, and the counts for the wild-type, for each NNS sub-library group.
5. Calculate the fitness effect  $F_i^a$  of each mutation  $a$  at each position  $i$  as the base ten logarithm of the ratio of counts obtained in the selection ( $N_i^{a,sel}$ ) versus the pre-selection ( $N_i^{a,pres}$ ) condition, relative to the wild-type:

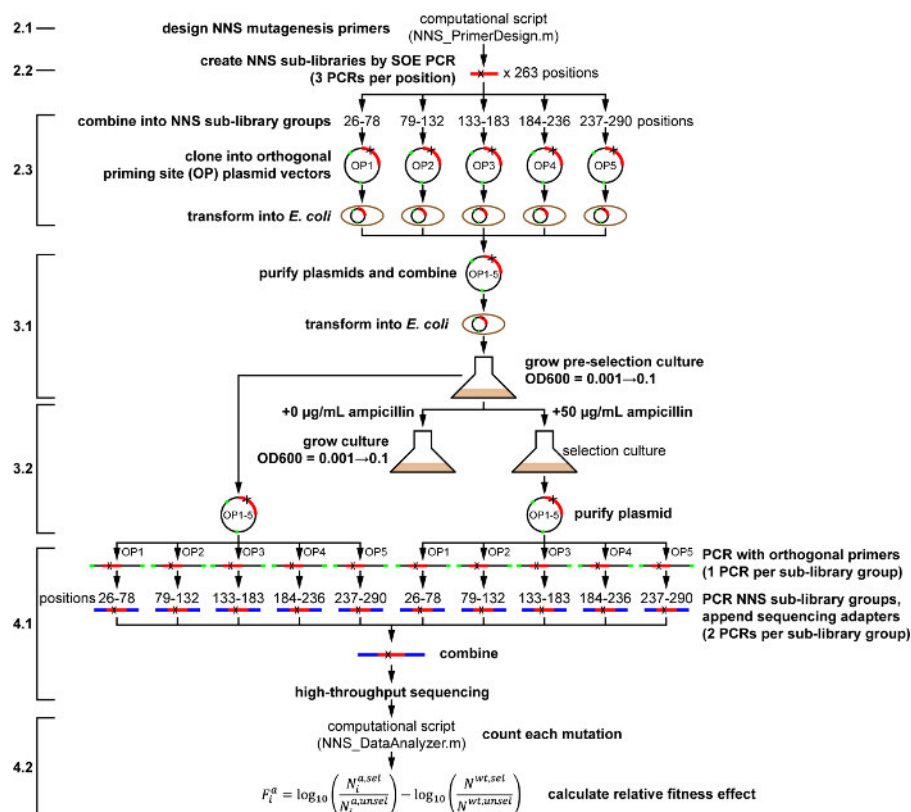
$$F_i^a = \log_{10} \frac{N_i^{a,sel}}{N_i^{a,pres}} - \log_{10} \frac{N_i^{wt,sel}}{N_i^{wt,pres}}.$$

## Representative Results

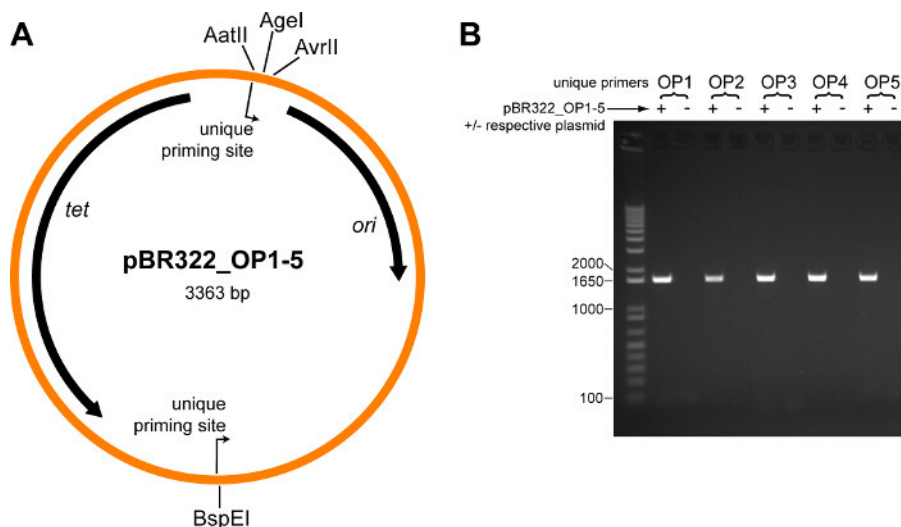
The plasmid map for the five modified pBR322 plasmids containing orthogonal priming sites (pBR322\_OP1 - pBR322\_OP5) is shown in **Figure 2A**. To test whether the orthogonal primers are specific, PCRs were performed using each pair of orthogonal primers individually, along with all five pBR322\_OP1-5 plasmids, or with all plasmids minus the plasmid matching the orthogonal primer pair. The correct product was only obtained when the matching plasmid was included, and no product of any size was obtained in its absence (**Figure 2B**).

A representative experiment was performed following the protocol described in this text (**Figure 1**). Following processing (protocol section 4.2),  $6.2 \times 10^6$  reads from the pre-selection condition and  $6.3 \times 10^6$  reads from the 50 µg/ml ampicillin selection condition were obtained. The counts obtained for each amino acid mutation from the pre-selection condition display a characteristic log-normal distribution (**Figure 3A**)<sup>13</sup>. At least one count from sequencing of the pre-selection culture for 98.9% of mutations (58 had no counts), and greater than 100 counts for 91.2% (465 had less than 100 counts) were obtained. **Figure 3B** depicts the relative fitness effect ( $F_i^a$ ) for each mutation at each position of TEM-1; the distribution of  $F_i^a$  is shown in **Figure 3C**. Under selection at 50 µg/ml ampicillin, most mutations have a neutral or nearly-neutral fitness effect ( $\approx 0$ , corresponding to white pixels in **Figure 3B** and the large peak in **Figure 3C**). A small fraction of mutations at this concentration have substantial effects on fitness ( $< 0$ , corresponding to blue pixels in **Figure 3B** and the left tail in **Figure 3C**); expectedly, these include mutations within the highly conserved active site residues (S70, K73, S130, D131, N132, K234 and G236)<sup>13,14</sup>. In contrast, few mutations considerably increase fitness over that of TEM-1 ( $> 0$ , corresponding to red pixels in **Figure 3B**), as might be expected since TEM-1 is highly efficient at ampicillin hydrolysis ( $k_{cat}/K_m \approx 10^7 \text{ M}^{-1}\text{s}^{-1}$ )<sup>14</sup>.

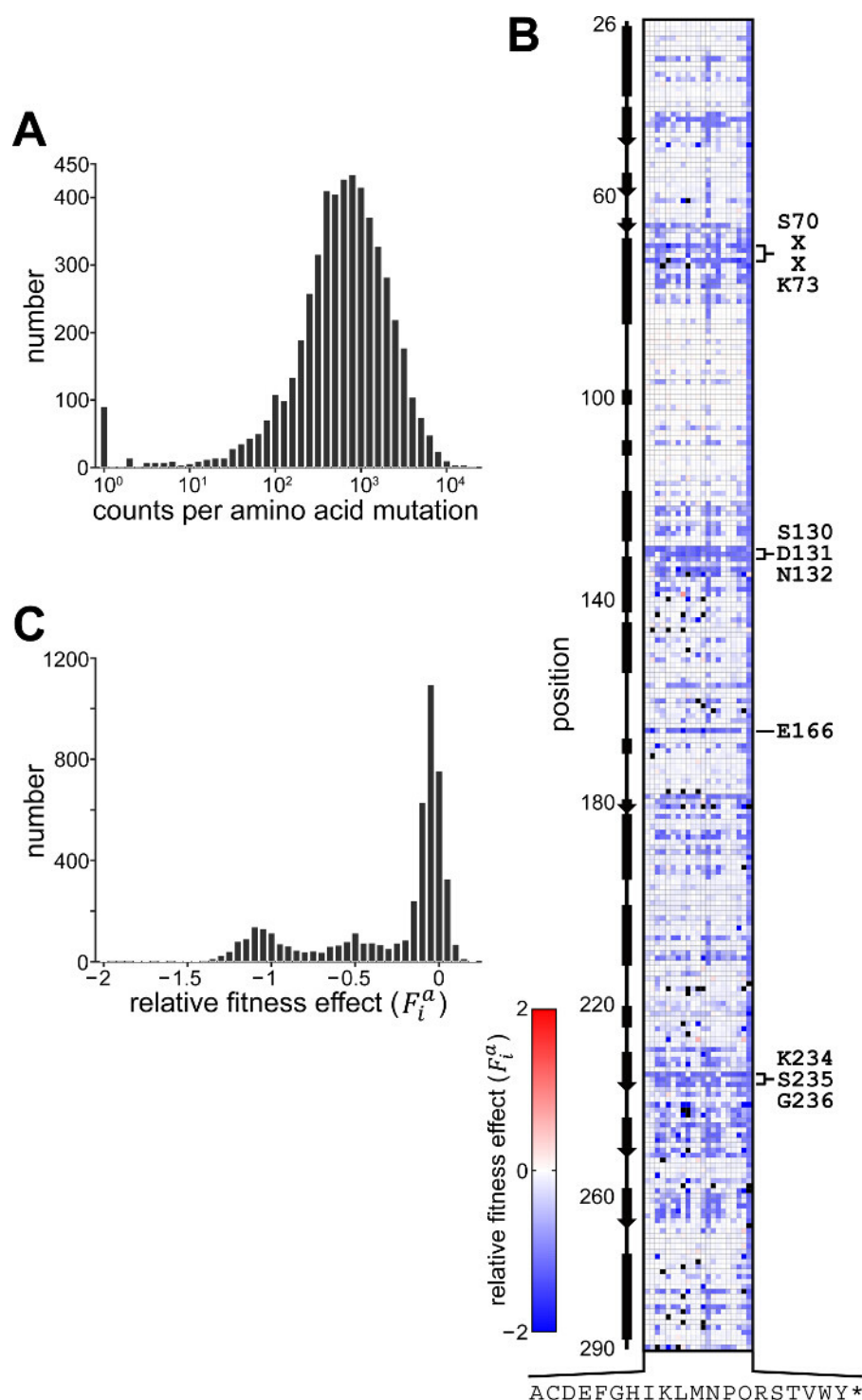




**Figure 1. Outline of Whole-protein Saturation Mutagenesis Protocol for TEM-1 β-lactamase under Ampicillin Selection.** Actions as described in the protocol are shown in bold. Numbered protocol steps are shown at left, for reference to the main text. [Please click here to view a larger version of this figure.](#)



**Figure 2. Validation of Orthogonal Priming Site Plasmid Vectors. (A)** Plasmid map for the five modified pBR322 plasmids containing orthogonal priming sites (pBR322\_OP1 - pBR322\_OP5). Location and direction of the orthogonal priming sites are indicated. Locations of several restriction sites are labeled; the TEM-1 whole-protein saturation mutagenesis library (which includes the entire TEM-1 gene and promoter) is cloned in-between the AatII and AvrII restriction sites. Abbreviations: *tet* tetracycline resistance gene, *ori* origin of replication. **(B)** Each pair of orthogonal primers (OP1-OP5) was tested in a PCR containing all five pBR322\_OP1-5 plasmids (+), or with all plasmids minus the respective plasmid (-). Shown is an ethidium bromide stained agarose gel (1% w/v) loaded with each PCR reaction, size separated by electrophoresis. The expected size product is 1,628 bp; the first lane is a DNA ladder, sizes of relevant standards are indicated. [Please click here to view a larger version of this figure.](#)



**Figure 3. Results of TEM-1  $\beta$ -lactamase Whole-protein Saturation Experiment.** (A) Histogram showing the distribution of counts for each amino acid mutation obtained from high-throughput sequencing of the library from the pre-selection condition. For simplicity, mutations with zero counts (53 mutations) are shown as having a count of one. (B) Fitness effects of all single amino acid mutations in TEM-1 under selection at 50  $\mu$ g/ml ampicillin. Shown is the data matrix containing the relative fitness effect ( $F_i^a$ ) depicted colorimetrically with blue representing deleterious effect, red a positive effect and white no fitness effect relative to wild-type. Mutations for which no counts were obtained from the pre-selection culture are colored black. Rows depict positions along the primary sequence and columns indicate mutation to one of twenty amino acids or stop codon (indicated by one-letter code, \* is stop codon); the secondary structure of TEM-1 is indicated at left and several highly conserved motifs within the active site are indicated at right. (C) Histogram showing the distribution of relative fitness effects. Shown are results for those mutations with >100 counts obtained from sequencing the pre-selection culture. [Please click here to view a larger version of this figure.](#)

Reagent	Mass or Volume	Comment
Tryptone	2 g	
Yeast extract	0.5 g	
Sodium chloride	0.06 g	
Potassium chloride	0.02 g	
Magnesium sulfate	0.24 g	
Purified water	100 ml	

**Table 1. Super Optimal Broth (SOB).** Reagent names and quantities used in preparing 100 ml SOB (Protocol step 1.1).

Reagent	Mass or Volume	Comment
Tryptone	10 g	
Yeast extract	5 g	
Sodium chloride	10 g	
Purified water	1 L	

**Table 2. Luria-Bertani Broth (LB).** Reagent names and quantities used in preparing 1 L LB (protocol step 1.1).

Reagent	Mass or Volume	Comment
Tryptone	10 g	
Yeast extract	5 g	
Sodium chloride	10 g	
Agar	15 g	
Purified water	1 L	

**Table 3. LB-agar.** Reagent names and quantities used in preparing LB-agar (protocol step 1.1).

Reagent	Volume	Comment
5x PCR buffer	1,450 $\mu$ l	
PCR additive	1,450 $\mu$ l	
2 mM dNTPs	725 $\mu$ l	2 mM each nucleotide
50 $\mu$ M AatII_F or AvrII_R primer	145 $\mu$ l	
1 ng/ $\mu$ l pBR322_AvrII plasmid	145 $\mu$ l	
2 units/ $\mu$ l DNA polymerase	72.5 $\mu$ l	
water	363 $\mu$ l	

**Table 4. First-round Mutagenic PCR Master Mix.** Reagent names and quantities for preparing the master mix for the first-round mutagenic PCR (protocol step 2.2.1). Total quantity is sufficient for 290 25  $\mu$ L reactions.

Reagent	Volume	Comment
5x PCR buffer	1,450 $\mu$ l	
PCR additive	1,450 $\mu$ l	
2 mM dNTPs	725 $\mu$ l	2 mM each nucleotide
50 $\mu$ M AatII_F primer	145 $\mu$ l	
50 $\mu$ M AvrII_R primer	145 $\mu$ l	
2 units/ $\mu$ l DNA polymerase	72.5 $\mu$ l	
water	2,973 $\mu$ l	

**Table 5. Second-round Mutagenic PCR Master Mix.** Reagent names and quantities for preparing the master mix for the second-round mutagenic PCR (protocol step 2.2.2). Total quantity is sufficient for 290 25  $\mu$ l reactions.



Reagent	Volume	Comment
5x PCR buffer	20 $\mu$ l	
PCR additive	20 $\mu$ l	
2 mM dNTPs	10 $\mu$ l	2 mM each nucleotide
50 $\mu$ M AvrII_F primer	2 $\mu$ l	
50 $\mu$ M AatII_OP1_R - AatII_OP5_R primer	2 $\mu$ l	one primer per reaction, paired with respective plasmid
1 ng/ $\mu$ l pBR322_OP1-5 plasmid	2 $\mu$ l	one plasmid per reaction
2 units/ $\mu$ l DNA polymerase	1 $\mu$ l	
water	43 $\mu$ l	

**Table 6. Cloning Vector PCR.** Reagent names and quantities for preparing the PCR to make the cloning vectors (protocol step 2.3.2).

Reagent	Volume	Comment
10x restriction enzyme buffer	5 $\mu$ l	
4 units/ $\mu$ l AvrII	2.5 $\mu$ l	
20 units/ $\mu$ l AatII	0.5 $\mu$ l	
DNA to restriction digest	volume for 500 ng	
water	to 50 $\mu$ l total volume	

**Table 7. Restriction Digests.** Reagent names and quantities for restriction digests of cloning vectors and NNS sub-library groups (protocol step 2.3.3).

Reagent	Volume	Comment
10x T4 DNA ligase buffer	5 $\mu$ l	
purified restriction-digested NNS sub-library group DNA	volume for 48 ng	
purified restriction-digested cloning vector DNA	volume for 52 ng	
400 units/ $\mu$ l T4 DNA ligase	1 $\mu$ l	
water	to 20 $\mu$ l total volume	

**Table 8. Ligations.** Reagent names and quantities for ligations of cloning vectors with restriction-digested NNS sub-library groups in a 1:3 vector:insert molar ratio (protocol step 2.3.4).

Reagent	Volume	Comment
5x PCR buffer	55 $\mu$ l	
PCR additive	55 $\mu$ l	
2 mM dNTPs	27.5 $\mu$ l	2 mM each nucleotide
2 units/ $\mu$ l DNA polymerase	2.75 $\mu$ l	
water	113 $\mu$ l	

**Table 9. PCR Reagents for Preparing Samples for High-throughput Sequencing.** Reagent names and quantities for preparing PCR master mixes used for de-multiplexing with orthogonal primers (4.1.1), isolating NNS sub-library groups (protocol step 4.1.2) and adding indexing sequences (protocol step 4.1.3). Total quantity is sufficient for 11 25  $\mu$ l reactions.

## Discussion

Here a protocol is described for performing the functional assessment of whole-protein saturation mutagenesis libraries, using high-throughput sequencing technology. An important aspect of the method is the use of orthogonal primers during the cloning process. Briefly, each amino acid position is randomized by mutagenic PCR, and mixed together into groups of positions whose combined sequence length is accommodated by high-throughput sequencing. These groups are cloned into plasmid vectors containing pairs of orthogonal priming sites, mixed together and subjected to selection, then de-multiplexed using the orthogonal primers, and subsequently deep sequenced. Since mutations are confined within the sequencing read length limit, this approach maximizes the number of useful reads containing mutations for genes of size longer than the sequencing read length. In addition, this technique allows for the simultaneous or "one-batch" selection of the entire mutational library, reducing the workload as well as the possibility that mutations experience different levels of selection. Practically, the critical steps in the protocol largely concern organization: during the cloning process (protocol step 2.3) one must ensure the correct mixing of mutagenic PCR products into groups and their subsequent cloning into the correct orthogonal priming site vector; during the preparation of the sample for sequencing (protocol

step 4.1), the correct orthogonal primers, as well as primers to isolate each of the NNS sub-library groups and add indexing sequences, must be used.

The three main steps of the protocol - library construction, selection, and sequencing - can be modified in several aspects. During library construction one could introduce mutations using a variety of techniques, for example, by error-prone PCR, or by constructing the gene using oligonucleotides synthesized by doping in a small fraction of alternative nucleotides<sup>21</sup>. One could construct the library to include double or higher-order mutations within segments of the protein (*i.e.*, NNS sub-library groups), or in alternate genotype backgrounds<sup>22</sup>. Importantly, all modifications to the library construction step however should satisfy the criterion that the correspondence between the location of mutations in the sequence and the sequencing read length is maintained. This criterion therefore excludes the application of the protocol towards comprehensive studies of multiple mutations across a protein. Modifications to the second part of the protocol include alternate selection conditions: different  $\beta$ -lactam types (or combinations) and concentrations, external stress conditions (*e.g.*, temperature, nutrient levels), host type (*e.g.*, different types of bacteria), or different sampling times (hours to days). For example, in previous work we examined the fitness effects of all single amino acid mutations in TEM-1 under different concentrations of ampicillin, and under the third-generation cephalosporin cefotaxime<sup>13</sup>. With regards to the third step of the protocol, we currently do not recommend deviating from the choice of the sequencing platform used here (see Table of Materials). While sequencing read lengths are indeed currently longer in other platforms, the number of reads obtainable is currently far lower; in general the accuracy to which the effect of a mutation can be determined is proportional to the number of reads obtained (see equation in protocol step 4.2.5).

Largely for simplicity, the protocol uses TEM-1  $\beta$ -lactamase as a model system, however the methodology described here can be extended to other systems for which a high-throughput selection or screening assay is in place. Constructing such assays is however often non-trivial: First, a strategy for compartmentalization of gene (mutation) and protein together must be established, for example within a cell, liquid droplet (as in a microfluidics platform), or by phage display. Second, and most importantly, a *quantitative* connection between protein function and a selectable phenotype, or fitness, must be established. For enzymes involved in metabolism or antibiotic resistance, the ability of cells to grow in nutrient drop-out or antibiotic media is often a direct function of enzymatic activity. A more synthetic approach could be used in other systems, for example by linking protein-protein binding affinity to reporter gene (*e.g.*, fluorescent protein) expression in bacteria or yeast<sup>11,23</sup>, or using a fluorogenic enzyme substrate in a microfluidics system<sup>24</sup>. Lastly, such an assay must be scalable, to address the size of a whole-protein mutagenesis library.

In summary, a high-throughput sequencing-based approach for the functional assessment of whole-protein saturation mutagenesis libraries is described here. Central to the approach is the construction of the mutagenesis library within segments along the gene, and the utilization of orthogonal primer barcodes to tag each segment for multiplexing and de-multiplexing the library. We anticipate that this protocol could be readily applied to other proteins for which an appropriate high-throughput selection or screen has been developed.

## Disclosures

The authors declare they have no competing financial interests

## Acknowledgements

R.R. acknowledges support from the National Institutes of Health (RO1EY018720-05), the Robert A. Welch Foundation (I-1366), and the Green Center for Systems Biology.

## References

- Hutchison, C. A., 3rd *et al.* Mutagenesis at a specific position in a DNA sequence. *J Biol Chem.* **253** (18), 6551-6560 (1978).
- Mullis, K. B., & Faloona, F. A. Specific synthesis of DNA *in vitro* via a polymerase-catalyzed chain reaction. *Methods Enzymol.* **155** 335-350 (1987).
- Papworth, C., Bauer, J. C., Braman, J. and Wright, D. A. Site-directed mutagenesis in one day with >80% efficiency. *Strategies.* **9** (3), 3-4 (1996).
- Higuchi, R., Krummel, B., & Saiki, R. K. A general method of *in vitro* preparation and specific mutagenesis of DNA fragments: study of protein and DNA interactions. *Nucleic Acids Res.* **16** (15), 7351-7367 (1988).
- Rennell, D., Bouvier, S. E., Hardy, L. W., & Poteete, A. R. Systematic mutation of bacteriophage T4 lysozyme. *J Mol Biol.* **222** (1), 67-88 (1991).
- Markiewicz, P., Kleina, L. G., Cruz, C., Ehret, S., & Miller, J. H. Genetic studies of the lac repressor. XIV. Analysis of 4000 altered *Escherichia coli* lac repressors reveals essential and non-essential residues, as well as 'spacers' which do not require a specific sequence. *J Mol Biol.* **240** (5), 421-433 (1994).
- Kleina, L. G., & Miller, J. H. Genetic studies of the lac repressor. XIII. Extensive amino acid replacements generated by the use of natural and synthetic nonsense suppressors. *J Mol Biol.* **212** (2), 295-318 (1990).
- Huang, W., Petrosino, J., Hirsch, M., Shenkin, P. S., & Palzkill, T. Amino acid sequence determinants of beta-lactamase structure and activity. *J Mol Biol.* **258** (4), 688-703 (1996).
- Fowler, D. M. *et al.* High-resolution mapping of protein sequence-function relationships. *Nat Methods.* **7** (9), 741-746 (2010).
- Hietpas, R. T., Jensen, J. D., & Bolon, D. N. Experimental illumination of a fitness landscape. *Proc Natl Acad Sci U S A.* **108** (19), 7896-7901 (2011).
- McLaughlin, R. N., Jr., Poelwijk, F. J., Raman, A., Gosal, W. S., & Ranganathan, R. The spatial architecture of protein function and adaptation. *Nature.* **491** (7422), 138-142 (2012).
- Deng, Z. *et al.* Deep sequencing of systematic combinatorial libraries reveals beta-lactamase sequence constraints at high resolution. *J Mol Biol.* **424** (3-4), 150-167 (2012).

13. Stiffler, M. A., Hekstra, D. R., & Ranganathan, R. Evolvability as a Function of Purifying Selection in TEM-1 beta-Lactamase. *Cell*. **160** (5), 882-892 (2015).
14. Matagne, A., Lamotte-Brasseur, J., & Frere, J. M. Catalytic properties of class A beta-lactamases: efficiency and diversity. *Biochem J*. **330** ( Pt 2) 581-598 (1998).
15. Salverda, M. L., De Visser, J. A., & Barlow, M. Natural evolution of TEM-1 beta-lactamase: experimental reconstruction and clinical relevance. *FEMS Microbiol Rev*. **34** (6), 1015-1036 (2010).
16. Weinreich, D. M., Delaney, N. F., Depristo, M. A., & Hartl, D. L. Darwinian evolution can follow only very few mutational paths to fitter proteins. *Science*. **312** (5770), 111-114 (2006).
17. Stewart, S. M., Fisher, M., Young, J. E., & Lutz, W. Ampicillin levels in sputum, serum, and saliva. *Thorax*. **25** (3), 304-311 (1970).
18. Giachetto, G. *et al.* Ampicillin and penicillin concentration in serum and pleural fluid of hospitalized children with community-acquired pneumonia. *Pediatr Infect Dis J*. **23** (7), 625-629 (2004).
19. Ambler, R. P. *et al.* A standard numbering scheme for the class A beta-lactamases. *Biochem J*. **276** ( Pt 1), 269-270 (1991).
20. Magoc, T., & Salzberg, S. L. FLASH: fast length adjustment of short reads to improve genome assemblies. *Bioinformatics*. **27** (21), 2957-2963 (2011).
21. Melamed, D., Young, D. L., Gamble, C. E., Miller, C. R., & Fields, S. Deep mutational scanning of an RRM domain of the *Saccharomyces cerevisiae* poly(A)-binding protein. *RNA*. **19** (11), 1537-1551 (2013).
22. Bank, C., Hietpas, R. T., Jensen, J. D., & Bolon, D. N. A systematic survey of an intragenic epistatic landscape. *Mol Biol Evol*. **32** (1), 229-238 (2015).
23. Dove, S. L., Joung, J. K., & Hochschild, A. Activation of prokaryotic transcription through arbitrary protein-protein contacts. *Nature*. **386** (6625), 627-630 (1997).
24. Romero, P. A., Tran, T. M., & Abate, A. R. Dissecting enzyme function with microfluidic-based deep mutational scanning. *Proc Natl Acad Sci U S A*. **112** (23), 7159-7164 (2015).