Journal of Visualized Experiments

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

--Manuscript Draft--

Manuscript Number:	JoVE54119R3
Full Title:	A Protocol for Functional Assessment of Whole-Protein Saturation Mutagenesis Libraries Utilizing High-Throughput Sequencing
Article Type:	Invited Methods Article - JoVE Produced Video
Keywords:	mutagenesis, saturation mutagenesis, next generation sequencing, high throughput sequencing, TEM-1 beta-lactamase, antibiotic resistance, orthogonal primers
Manuscript Classifications:	10.1.293.69.500: Synthetic Biology; 8.1.158.273.180.800: Systems Biology; 8.1.158.273.904: Synthetic Biology
Corresponding Author:	Michael Allen Stiffler, Ph.D University of Texas Southwestern Medical Center at Dallas Dallas, TX UNITED STATES
Corresponding Author Secondary Information:	
Corresponding Author E-Mail:	Michael.Stiffler@utsouthwestern.edu;mstiffler@post.harvard.edu
Corresponding Author's Institution:	University of Texas Southwestern Medical Center at Dallas
Corresponding Author's Secondary Institution:	
First Author:	Michael Allen Stiffler, Ph.D
First Author Secondary Information:	
Other Authors:	Subu K Subramanian
	Victor H Salinas
	Rama Ranganathan, MD, PhD
Order of Authors Secondary Information:	
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 β-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 β-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.
Author Comments:	
Additional Information:	
Question	**Response**

If this article needs to be "in-press" by a certain date to satisfy grant requirements, please indicate the date below and explain in your cover letter.	

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

AUTHORS:
Michael A. Stiffler
Green Center for Systems Biology
University of Texas Southwestern Medical Center
Dallas, TX, USA
mstiffler@post.harvard.edu

Subu K. Subramanian
Green Center for Systems Biology
University of Texas Southwestern Medical Center
Dallas, TX, USA
subramanian.subramanian@utsouthwestern.edu

Victor H. Salinas
Green Center for Systems Biology
University of Texas Southwestern Medical Center
Dallas, TX, USA
victor.salinas@utsouthwestern.edu

Rama Ranganathan
Green Center for Systems Biology
University of Texas Southwestern Medical Center
Dallas, TX, USA
rama.ranganathan@utsouthwestern.edu

CORRESPONDING AUTHORS:
Michael A. Stiffler
mstiffler@post.harvard.edu

Rama Ranganathan
rama.ranganathan@utsouthwestern.edu

KEYWORDS:
mutagenesis, saturation mutagenesis, next generation sequencing, high throughput sequencing, TEM-1 beta-lactamase, antibiotic resistance, orthogonal primers

SHORT ABSTRACT:
We present a protocol for the functional assessment of comprehensive single-site saturation mutagenesis libraries of proteins utilizing high-throughput sequencing. Importantly, this approach uses orthogonal primer pairs to multiplex library construction and sequencing. Representative results using TEM-1 β-lactamase selected at a clinically relevant dosage of ampicillin are provided.

LONG 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 β-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 β-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.

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[1,2]. 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[3,4]. 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[5-7], or were limited in their quantitative ability due to the low sequencing depth of Sanger sequencing[8]. The techniques used in these studies have largely been supplanted by methods utilizing high-throughput sequencing technology[9-12]. 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[11,13]: 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 β-lactamase, an enzyme which confers high-level resistance to β-lactam antibiotics (e.g. ampicillin) in bacteria is used as a model system[14-16]. 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 μg/mL)[17,18].

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 room temperature unless other indicated.

1. Prepare culture media and plates.

1.1. 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.

1.2. 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 μm filter and store at 4 °C protected from light.

1.3. Cool LB-agar to 50 °C and then add 1 mL of Tet stock (final concentration of 12 μg/mL Tet). Pour into petri plates and cool at room temperature 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.

2.1. Designing mutagenesis primers.
2.1.1. 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:

2.1.1.1. 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.

2.1.1.2. Ensure that the 5' and 3' ends terminate in C or G and that the melting temperature (T_m) is approximately 70 °C[3]. Use the computational script NNS_PrimerDesign.m (See Supplementary Code File) to design NNS mutagenesis primers according to these guidelines.

2.1.2. 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 µM, with one set of plates containing the sense mutagenesis primers and another the antisense primers.

2.1.3. Fill a pipette basin with water and use a multichannel pipette to transfer 95 µL to 263 wells over three 96-well plates. Dilute the primers 20-fold to 2.5 µM by using a multichannel pipette to transfer 5 µL from the 96-well plates containing the sense mutagenesis primers to the cognate wells of the plates containing water.

2.1.4. Repeat protocol step 2.1.3 to dilute the antisense mutagenesis primers.

2.2. Synthesis of NNS sub-libraries for each amino acid position by two-step PCR site-directed mutagenesis.
2.2.1. Perform the first-round mutagenic PCRs. For each mutagenesis primer, prepare a 25 µL 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.

2.2.1.1. 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 µL to 263 wells over three 96-well PCR plates. Use a multichannel pipette to transfer 10 µL from the 96-well plates containing the diluted sense mutagenesis primers to the cognate wells in the PCR plates.

2.2.1.2. Cover each PCR plate with a 96-well plate seal. Centrifuge at 200 x g for 2 min.

2.2.1.3. 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.

2.2.1.4. Repeat protocol steps 2.2.1.1 – 2.2.1.3 for the 96-well plates containing the diluted antisense mutagenesis primers.

2.2.2. Perform the second-round mutagenic PCRs. For each amino acid position, prepare a 25 µL 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).

2.2.2.1. Fill a pipette basin with water and use a multichannel pipette to transfer 198 µL to 263 wells over three 96-well plates.

2.2.2.2. Combine and dilute 100-fold the mutagenic PCR products for each amino acid position by using a multichannel pipette to first transfer 1 µL 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.

2.2.2.3. 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 µL to 263 wells over three 96-well PCR plates. Use a multichannel pipette to transfer 1 µL from the 96-well plates containing the mixed and diluted first-round mutagenic PCR products to the cognate wells in the PCR plates.

2.2.2.4. 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.

2.2.3. 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.

2.2.3.1. 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.

2.2.3.2. Prepare a 1.5% agarose gel with 0.2 µg/mL ethidium bromide (CAUTION).

2.2.3.3. 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.

2.2.3.4. Run the gel at 100 V for 40 min and image on a UV transilluminator.

2.2.3.5. Repeat protocol steps 2.2.3.2 – 2.2.3.4 until all samples are analyzed.

2.2.4. Accurately measure the concentration of each NNS sub-library PCR product using a dsDNA quantitation reagent.

2.2.4.1. 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.

2.2.4.2. 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.

2.2.4.3. 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).

2.2.4.4. 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.

2.2.4.5. 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 room temperature for 5 min protected from light.

2.2.4.6. Measure fluorescence of each sample using a microplate reader and standard fluorescein wavelengths (excitation 485 nm, emission 520 nm; 0.1 sec).

2.2.4.7. 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.

2.3. Cloning of NNS sub-libraries into selection vectors.

2.3.1. 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.*[19]).

2.3.2. Create cloning vectors for each NNS sub-library group.

2.3.2.1. 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.).

2.3.2.2. 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.

2.3.2.3. Prepare a 1% agarose gel with 0.2 µg/mL ethidium bromide (CAUTION).

2.3.2.4. 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.

2.3.2.5. Run gel at 100 V for 50 min.

2.3.2.6. Visualize gel using a long-wavelength UV illuminator (CAUTION). Excise slices containing the PCR product at ~3500 bp; transfer to separate microfuge tubes. Gel slices can be stored at -20 °C.

2.3.2.7. Following manufacturers' instructions, purify samples using a gel extraction kit and measure concentration using a dsDNA quantitation reagent.

2.3.3. 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.

2.3.4. 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 room temperature for 1 hr. Clean up reactions using a DNA purification kit according to manufacturer's instructions; elute DNA with 20 µL of water.

2.3.5. Transform the entirety of the purified ligation reactions into library-efficient *E. coli* cells by electroporation.

2.3.5.1. Thaw electrocompetent *E. coli* cells and then place cells and purified ligation reactions on ice.

2.3.5.2. Transfer 10 µL thawed cells to each purified ligation reaction and then transfer to electroporation cuvette. Electroporate at 1.8 kV.

2.3.5.3. Recover cells by resuspending in 1 mL SOB. Incubate for one hour at 37 °C.

2.3.5.4. 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 overnight (~16 hr) at 37 °C.

2.3.5.5. 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 overnight (~16 hr) at 37 °C with vigorous shaking (~200 rpm).

2.3.6. 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 (1000 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 ≥100-fold over the number of expected mutations. Each NNS sub-library has ~53 positions, so the expected number of mutations is 53 positions × 32 codons/position ≈ 1.7×10^3; to give a library size ≥100-fold (≥1.7×10^5) there should be ≥170 colonies on each plate.

2.3.7. 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.

3.1. Preparation of the pre-selection culture.

3.1.1. Dilute the plasmid from protocol step 2.3.7 to 0.5 ng/µL in water and transfer 20 µL to a microcentrifuge tube. Perform transformation, recovery, plating and overnight growth as previously described in protocol step 2.3.5, except transfer 1 µL of the SOB recovery culture to 999 µL LB.

3.1.2. Count the number of colonies. To ensure complete coverage of all mutations there should be ≥100 colonies, indicating ≥10^6 successful transformants (100 × 263 positions × 32 codons/position ≈ 10^6).

3.1.3. Measure the concentration of the 50 mL overnight culture.

3.1.3.1. Prepare an LB blank by adding 1 mL LB to a spectrophotometer cuvette. Measure OD600 on a spectrophotometer.

3.1.3.2. Dilute the overnight culture 10-fold by resuspending 100 µL in 900 µL LB. Measure the OD600. Subtract the OD600 reading of the blank and multiply by 10 to give the OD600 of the overnight culture.

3.1.4. Pre-warm the three culture flasks from protocol step 1.1 for ~30 min at 37 °C. Dilute the overnight culture to OD600 = 0.1 and add 1 mL to one flask (final OD600 = 0.001). This is the "pre-selection culture".

3.1.5. 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 (~2.5 hr).

3.1.6. Transfer 100 mL of the pre-selection culture to two 50 mL conical tubes. Centrifuge at 4000 x 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.

3.2. Selection for ampicillin resistance.

3.2.1. 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 µm filter and store at 4 °C.

3.2.2. To the other two flasks, add Tet to a final concentration of 12 µg/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 µg/mL – this is the "selection culture".

3.2.3. 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 (~2.5 hr). At this time, also measure the OD600 of the selection culture.

3.2.4. Divide the OD600 of the selection culture into 0.1 and multiply by 100 mL. Transfer this volume (~400 mL) to 50 mL conical tubes and centrifuge at 4000 x 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.

3.2.5. 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.

4.1. Preparation of samples for high-throughput sequencing.

4.1.1. Prepare 25 µL PCRs to de-multiplex the NNS sub-library groups with orthogonal primers.

4.1.1.1. Prepare a PCR master mix according to Table 9; transfer 23 µL to ten PCR tubes.

4.1.1.2. Add 1 µL of 0.5 ng/µL purified plasmid DNA from the pre-selection culture to PCR tubes 1 – 5 and from the selection culture to tubes 6 – 10.

4.1.1.3. Mix together 50 µL of 50 µM forward orthogonal primers OP1_F – OP5_F with the respective reverse orthogonal primers OP1_R – OP5_R. In the same order, transfer 1 µL to PCR tubes 1 – 5 and 6 – 10. See Table of Materials for sequences of OP1_F – OP5_F and OP1_R – OP5_R.

4.1.1.4. 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.

4.1.2. Prepare 25 µL PCRs to isolate each of the NNS sub-library groups.

4.1.2.1. Dilute 100-fold the ten PCRs from protocol step 4.1.1.4 by transferring 1 µL of each to separate PCR tubes and adding 99 µL of water. Mix, then pipette out 99 µL and discard.

4.1.2.2. Mix together 50 µL of 50 µM forward primers Group1_F – Group5_F with the respective reverse primers Group1_R – Group5_R. In the same order, transfer 1 µL to PCR tubes

1 – 5 and 6 – 10. See Table of Materials for sequences of Group1_F – Group5_F and Group1_R – Group5_R.

4.1.2.3. Prepare a PCR master mix according to Table 9, transfer 23 µL to each PCR tube. Transfer PCR tubes to thermocycler; run the same program as in protocol step 2.2.1.3.

4.1.3. Carry out the final 25 µL PCRs to add indexing sequences.

4.1.3.1. Dilute 100-fold the ten PCRs from protocol step 4.1.2.3 by transferring 1 µL of each to separate PCR tubes and adding 99 µL of water. Mix, then pipette out 99 µL and discard.

4.1.3.2. Prepare a PCR master mix according to Table 9, transfer 23 µL to each PCR tube.

4.1.3.3. 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.

4.1.3.4. Transfer PCR tubes to the thermocycler and run the program from step 2.2.1.3.

4.1.4. Mix and purify samples.

4.1.4.1. Measure concentrations using a dsDNA quantitation reagent according to manufacturer's instructions. Mix 100 ng of each PCR product into a single microcentrifuge tube.

4.1.4.2. Prepare a 2% agarose gel with 0.2 µg/mL ethidium bromide (CAUTION).

4.1.4.3. 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.

4.1.4.4. 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.

4.1.4.5. 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.

4.1.5. Sequence on a high-throughput sequencing platform (see Table of Materials for platform used in this protocol).

4.1.5.1. Calculate sample concentration in nM as $\left[\left(ng/10^3\right)/\left(bp*660\right)\right]*10^9$, 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.

4.1.5.2. Follow manufacturer's instructions to denature sample and dilute to 9 pM in hybridization buffer.

4.1.5.3. Load 600 µL of sample into reagent cartridge. Sequence following manufacturer's instructions and on-screen prompts.

4.2. Analysis of sequencing data.
4.2.1. Download FLASh (Fast Length Adjustment of Short reads)[20], place into folder along with fastq.gz files obtained from sequencing.

4.2.2. 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).

4.2.2.1. 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.

4.2.3. After joining each pair of reads, place the out.extendedFrags.fastq output file into separate folders for results from the pre-selection or selection cultures. Rename the out.extendedFrags.fastq output file according to the NNS sub-library group to which it corresponds (i.e. 1.fastq, 2.fastq, etc.).

4.2.4. 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.

4.2.5. 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 ($N_i^{a,pre}$) condition, relative to the wild-type: $F_i^a = \log_{10}\frac{N_i^{a,sel}}{N_i^{a,pre}} - \log_{10}\frac{N^{wt,sel}}{N^{wt,pre}}$.

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), 6.2×10^6 reads from the pre-selection condition and 6.3×10^6 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)[13]. 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 ($F_i^a \approx 0$, 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 ($F_i^a << 0$, 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)[13,14]. In contrast, few mutations considerably increase fitness over that of TEM-1 ($F_i^a >> 0$, corresponding to red pixels in Figure 3B), as might be expected since TEM-1 is highly efficient at ampicillin hydrolysis ($k_{cat}/K_m \approx 10^7$ $M^{-1}s^{-1}$)[14].

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.

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 1628 bp; the first lane is a DNA ladder, sizes of relevant standards are indicated.

Figure 3. Results of TEM-1 β-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 μg/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.

Table 1. Super Optimal Broth (SOB). Reagent names and quantities used in preparing 100 mL SOB (protocol step 1.1).

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

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

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 µL reactions.

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 µL reactions.

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

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

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).

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 µL 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[21]. 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[22]. *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 β-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[13]. 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 β-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[11,23], or using a fluorogenic enzyme substrate in a microfluidics system[24]. 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.

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.

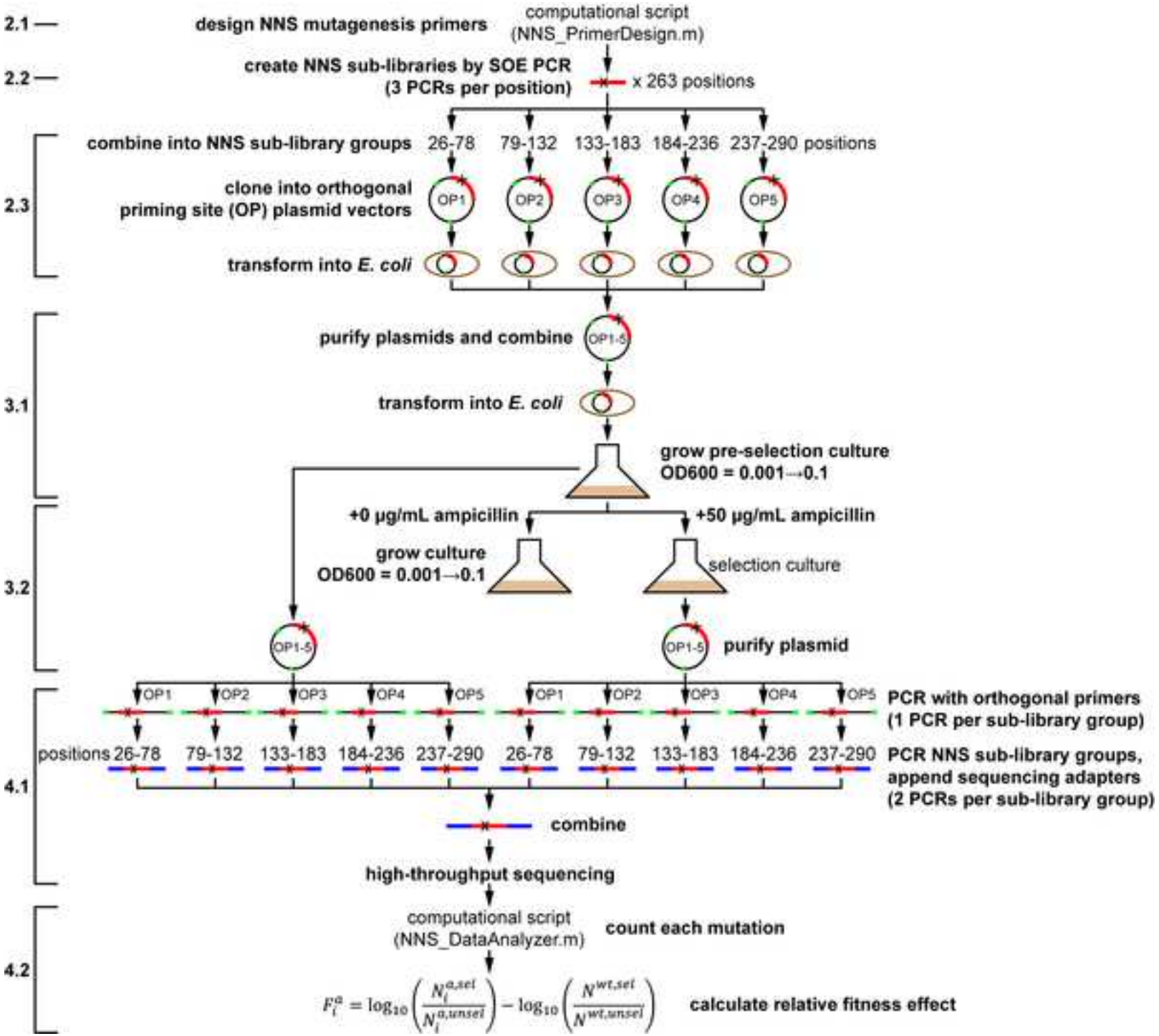
DISCLOSURES:
The authors declare they have no competing financial interests

REFERENCES:
1 Hutchison, C. A., 3rd *et al.* Mutagenesis at a specific position in a DNA sequence. *J Biol Chem.* **253** (18), 6551-6560 (1978).
2 Mullis, K. B. & Faloona, F. A. Specific synthesis of DNA in vitro via a polymerase-catalyzed chain reaction. *Methods Enzymol.* **155** 335-350 (1987).
3 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).
4 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).
5 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).
6 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, doi:10.1006/jmbi.1994.1458 (1994).
7 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, doi:10.1016/0022-2836(90)90126-7 (1990).
8 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, doi:10.1006/jmbi.1996.0279 (1996).
9 Fowler, D. M. *et al.* High-resolution mapping of protein sequence-function relationships. *Nat Methods.* **7** (9), 741-746, doi:10.1038/nmeth.1492 (2010).
10 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, doi:10.1073/pnas.1016024108 (2011).
11 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, doi:10.1038/nature11500 (2012).
12 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, doi:10.1016/j.jmb.2012.09.014 (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, doi:10.1016/j.cell.2015.01.035 (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).

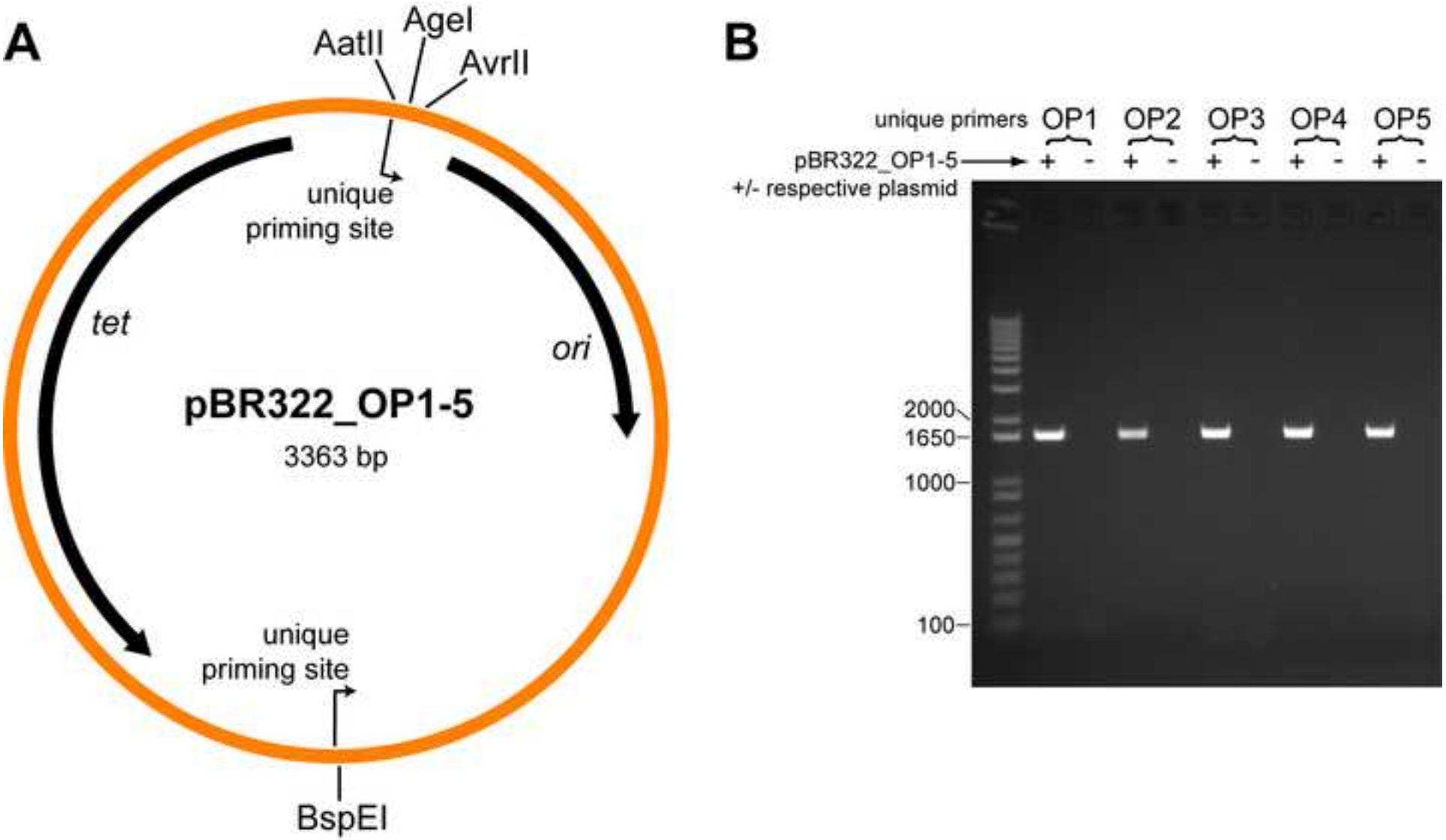
15 Salverda, M. L., De Visser, J. A. & Barlow, M. Natural evolution of TEM-1 beta-lactamase: experimental reconstruction and clinical relevance. *FEMS Microbiol Rev.* **34** (6), 1015-1036, doi:10.1111/j.1574-6976.2010.00222.x (2010).
16 Weinreich, D. M., Delaney, N. F., Depristo, M. A. & Hartl, D. L. Darwinian evolution can follow only very few mutational paths to fitter proteins. *Science.* **312** (5770), 111-114, doi:10.1126/science.1123539 (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, doi:10.1093/bioinformatics/btr507 (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, doi:10.1261/rna.040709.113 (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, doi:10.1093/molbev/msu301 (2015).
23 Dove, S. L., Joung, J. K. & Hochschild, A. Activation of prokaryotic transcription through arbitrary protein-protein contacts. *Nature.* **386** (6625), 627-630, doi:10.1038/386627a0 (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, doi:10.1073/pnas.1422285112 (2015).

Figure

Click here to download Figure Stiffler_Figure1_JoVE_rev.tif

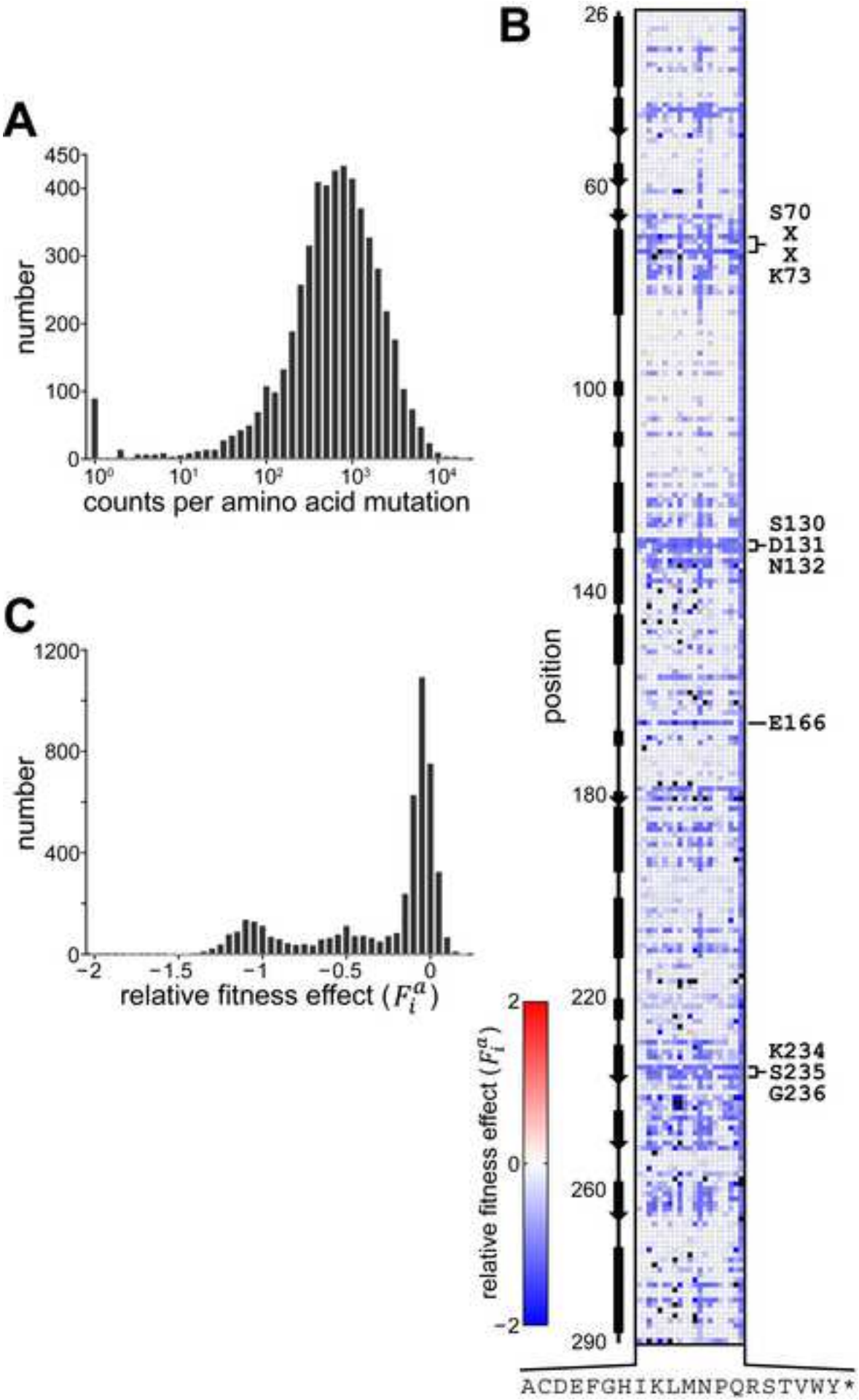


Figure



Figure

Click here to download Figure Stiffler_Figure3_JoVE.tif



Table

Reagent	Mass or Volume	Comment
Typtone	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

Click here to download Table Table 2.xlsx

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

Table

Click here to download Table Table 3.xlsx

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

Table

Reagent	Volume	Comment
5X PCR buffer	1450 µL	
PCR additive	1450 µL	
2 mM dNTPs	725 µL	2 mM each nucleotide
50 µM AatII_F or AvrII_R primer	145 µL	
1 ng/µL pBR322_AvrII plasmid	145 µL	
2 units/µL DNA polymerase	72.5 µL	
water	363 µL	

Table

Reagent	Volume	Comment
5X PCR buffer	1450 µL	
PCR additive	1450 µL	
2 mM dNTPs	725 µL	2 mM each nucleotide
50 µM AatII_F primer	145 µL	
50 µM AvrII_R primer	145 µL	
2 units/µL DNA polymerase	72.5 µL	
water	2973 µL	

Table

Reagent	Volume
5X PCR buffer	20 µL
PCR additive	20 µL
2 mM dNTPs	10 µL
50 µM AvrII_F primer	2 µL
50 µM AatII_OP1_R - AatII_OP5_R primer	2 µL
1 ng/µL pBR322_OP1-5 plasmid	2 µL
2 units/µL DNA polymerase	1 µL
water	43 µL

Comment

2 mM each nucleotide

one primer per reaction, paired with respective plasmid
one plasmid per reaction

Table

Reagent	Volume	Comment
10X restriction enzyme buffer	5 µL	
4 units/µL AvrII	2.5 µL	
20 units/µL AatII	0.5 µL	
DNA to restriction digest	volume for 500 ng	
water	to 50 µL total volume	

Table

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

Table

Reagent	Volume	Comment
5X PCR buffer	55 µL	
PCR additive	55 µL	
2 mM dNTPs	27.5 µL	2 mM each nucleotide
2 units/µL DNA polymerase	2.75 µL	
water	113 µL	

Name of Material/ Equipment	Company	Catalog Number
Typtone	Research Products Intl. Corp.	T60060-1000.0
Yeast extract	Research Products Intl. Corp.	Y20020-500.0
Sodium chloride	Fisher Scientific	BP358-212
Potassium chloride	Sigma-Aldrich	P9333-500G
Magnesium sulfate	Sigma-Aldrich	M7506-500G
Agar	Fisher Scientific	BP1423-500
Tetracycline hydrochloride	Sigma-Aldrich	T7660-5G
petri plates	Corning	351029
MATLAB	Mathworks	http://www.mathworks.com/products/matlab/
Oligonucleotide primers	Integrated DNA Technologies	https://www.idtdna.com/pages/products/dna-rna/custom-dna-oligos
pBR322_AvrII	available upon request	
pBR322_OP1 – pBR322_OP5	available upon request	
Q5 high-fidelity DNA polymerase	New England Biolabs	M0491L
15 mL conical tube	Corning	430025
Multichannel pipettes (Eppendorf ResearchPlus)	Eppendorf	
PCR plate, 96 well	Fisher Scientific	14230232
96 well plate seal	Excel Scientific	F-96-100
Veriti 96-well thermal cycler	Applied Biosystems	4375786
6X gel loading dye	New England Biolabs	B7024S
Agarose	Research Products Intl. Corp.	20090-500.0
Ethidium bromide	Bio-Rad	161-0433
UV transilluminator (FOTO/Analyst ImageTech)	Fotodyne Inc.	http://www.fotodyne.com/content/ImageTech_gel_documentation
EB buffer	Qiagen	19086

96-well black-walled, clear bottom assay plates	Corning	3651
Lambda phage DNA	New England Biolabs	N3011S
PicoGreen dsDNA reagent	Invitrogen	P7581
Victor 3V microplate reader	PerkinElmer	
DNA purification kit	Zymo Research	D4003
Microcentrifuge tubes	Corning	3621
Long-wavelength UV illuminator	Fisher Scientific	FBUVLS-80
Agarose gel DNA extraction buffer	Zymo Research	D4001-1-100
AatII	New England Biolabs	R0117S
AvrII	New England Biolabs	R0174L
T4 DNA ligase	New England Biolabs	M0202S
EVB100 electrocompetent *E. coli*	Avidity	EVB100
Electroporator (*E. coli* Pulser)	Bio-Rad	1652102
Electroporation cuvettes	Bio-Rad	165-2089
Spectrophotometer (Ultrospec 3100 pro)	Amersham Biosciences	80211237
50 mL conical tubes	Corning	430828
Plasmid purification kit	Macherey-Nagel	740588.25
8 well PCR strip tubes	Axygen	321-10-551
Qubit dsDNA HS assay kit	Invitrogen	Q32854
Qubit assay tubes	Invitrogen	Q32856
Qubit fluorometer	Invitrogen	Q32866
Ampicillin sodium salt	Akron Biotechnology	50824296
MiSeq reagent kit v2 (500 cycles)	Illumina	MS-102-2003
MiSeq desktop sequencer	Illumina	http://www.illumina.com/systems/miseq.html
FLASh software	John Hopkins University - open source	http://ccb.jhu.edu/software/FLASH/
AatII_F		
AvrII_R		

AvrII_F

AatII_OP1_R

AatII_OP2_R

AatII_OP3_R

AatII_OP4_R

AatII_OP5_R
OP1_F
OP1_R
OP2_F
OP2_R
OP3_F
OP3_R
OP4_F
OP4_R
OP5_F
OP5_R

Group1_F

Group1_R

Group2_F

Group2_R

Group3_F

Group3_R

Group4_F

Group4_R

Group5_F

Group5_R

501_F

502_F

503_F

504_F

505_F

701_R

702_R

Comments/Description

25 nmol scale, standard desalting

pBR322 plasmid modified to contain AvrII restriction site downstream of the TEM-1 gene

five modified pBR322 plasmids each containing a pair of orthogonal priming sites

includes 5X PCR buffer and PCR additive (GC enhancer)

dsDNA quantitation reagent, used in protocol step 2.2.4

dsDNA quantitation reagent

alternatively, one could sequence on Illumina HiSeq platform
software to merge paired-end reads from next-generation sequencing data

GATAATAATGGTTTCTTAGACGTCAGGTGGC
CTTCACCTAGGTCCTTTTAAATTAAAAATGAA
G

CTTCATTTTTAATTTAAAAGGACCTAGGTGAA
G
ACCTGACGTCCGTATTTCAACTGTCCGGTCTA
AGAAACCATTATTATCATGACATTAAC
ACCTGACGTCCGCTCACGGAGTGTACTAATTA
AGAAACCATTATTATCATGACATTAAC
ACCTGACGTCGTACGTCTGAACTTGGGACTTA
AGAAACCATTATTATCATGACATTAAC
ACCTGACGTCCCGTTCTCGATACCAAGTGATA
AGAAACCATTATTATCATGACATTAAC
ACCTGACGTCGTCCGTCGGAGTAACAATCTTA
AGAAACCATTATTATCATGACATTAAC
GACCGGACAGTTGAAATACG
CGACGTACAGGACAATTTCC
ATTAGTACACTCCGTGAGCG
AGTATTAGGCGTCAAGGTCC
AGTCCCAAGTTCAGACGTAC
GAAAAGTCCCAATGAGTGCC
TCACTTGGTATCGAGAACGG
TATCACGGAAGGACTCAACG
AGATTGTTACTCCGACGGAC
TATAACAGGCTGCTGAGACC

ACACTCTTTCCCTACACGACGCTCTTCCGATCT
NNNNNGCATTTTGCCTACCGGTTTTTGC
GTGACTGGAGTTCAGACGTGTGCTCTTCCGAT
CTNNNNNTCTTGCCCGGCGTCAAC

ACACTCTTTCCCTACACGACGCTCTTCCGATCT
NNNNNGAACGTTTTCCAATGATGAGCAC

GTGACTGGAGTTCAGACGTGTGCTCTTCCGAT
CTNNNNNGTCCTCCGATCGTTGTCAGAAG

ACACTCTTTCCCTACACGACGCTCTTCCGATCT
NNNNNAGTAAGAGAATTATGCAGTGCTGCC

GTGACTGGAGTTCAGACGTGTGCTCTTCCGAT
CTNNNNNTCGCCAGTTAATAGTTTGCGC

ACACTCTTTCCCTACACGACGCTCTTCCGATCT
NNNNNCCAAACGACGAGCGTGACAC

GTGACTGGAGTTCAGACGTGTGCTCTTCCGAT
CTNNNNNGCAATGATACCGCGAGACCC
ACACTCTTTCCCTACACGACGCTCTTCCGATCT
NNNNNCGGCTGGCTGGTTTATTGC
GTGACTGGAGTTCAGACGTGTGCTCTTCCGAT
CTNNNNNTATATGAGTAAACTTGGTCTGACA
G
AATGATACGGCGACCACCGAGATCTACACTA
TAGCCTACACTCTTTCCCTACACGAC

AATGATACGGCGACCACCGAGATCTACACAT
AGAGGCACACTCTTTCCCTACACGAC
AATGATACGGCGACCACCGAGATCTACACCC
TATCCTACACTCTTTCCCTACACGAC

AATGATACGGCGACCACCGAGATCTACACGG
CTCTGAACACTCTTTCCCTACACGAC

AATGATACGGCGACCACCGAGATCTACACAG
GCGAAGACACTCTTTCCCTACACGAC
CAAGCAGAAGACGGCATACGAGATCGAGTA
ATGTGACTGGAGTTCAGACGTG

CAAGCAGAAGACGGCATACGAGATTCTCCGG
AGTGACTGGAGTTCAGACGTG

jove
JOURNAL OF VISUALIZED EXPERIMENTS
1 Alewife Center #200
Cambridge, MA 02140
tel. 617.945.9051
www.jove.com

ARTICLE AND VIDEO LICENSE AGREEMENT

Title of Article: A Protocol for Whole-Protein Saturation Mutagenesis Utilizing High-Throughput Sequencing

Author(s): Michael A. Stiffler, Subu K. Subramanian, Victor H. Salinas and Rama Ranganathan

Item 1 (check one box): The Author elects to have the Materials be made available (as described at http://www.jove.com/publish) via: ☒ Standard Access ☐ Open Access

Item 2 (check one box):

☒ The Author is NOT a United States government employee.

☐ The Author is a United States government employee and the Materials were prepared in the course of his or her duties as a United States government employee.

☐ The Author is a United States government employee but the Materials were NOT prepared in the course of his or her duties as a United States government employee.

ARTICLE AND VIDEO LICENSE AGREEMENT

1. Defined Terms. As used in this Article and Video License Agreement, the following terms shall have the following meanings: **"Agreement"** means this Article and Video License Agreement; **"Article"** means the article specified on the last page of this Agreement, including any associated materials such as texts, figures, tables, artwork, abstracts, or summaries contained therein; **"Author"** means the author who is a signatory to this Agreement; **"Collective Work"** means a work, such as a periodical issue, anthology or encyclopedia, in which the Materials in their entirety in unmodified form, along with a number of other contributions, constituting separate and independent works in themselves, are assembled into a collective whole; **"CRC License"** means the Creative Commons Attribution-Non Commercial-No Derivs 3.0 Unported Agreement, the terms and conditions of which can be found at: http://creativecommons.org/licenses/by-nc-nd/3.0/legalcode; **"Derivative Work"** means a work based upon the Materials or upon the Materials and other pre-existing works, such as a translation, musical arrangement, dramatization, fictionalization, motion picture version, sound recording, art reproduction, abridgment, condensation, or any other form in which the Materials may be recast, transformed, or adapted; **"Institution"** means the institution, listed on the last page of this Agreement, by which the Author was employed at the time of the creation of the Materials; **"JoVE"** means MyJove Corporation, a Massachusetts corporation and the publisher of *The Journal of Visualized Experiments;* **"Materials"** means the Article and / or the Video; **"Parties"** means the Author and JoVE; **"Video"** means any video(s) made by the Author, alone or in conjunction with any other parties, or by JoVE or its affiliates or agents, individually or in collaboration with the Author or any other parties, incorporating all or any portion of the Article, and in which the Author may or may not appear.

2. Background. The Author, who is the author of the Article, in order to ensure the dissemination and protection of the Article, desires to have the JoVE publish the Article and create and transmit videos based on the Article. In furtherance of such goals, the Parties desire to memorialize in this Agreement the respective rights of each Party in and to the Article and the Video.

3. Grant of Rights in Article. In consideration of JoVE agreeing to publish the Article, the Author hereby grants to JoVE, subject to **Sections 4** and **7** below, the exclusive, royalty-free, perpetual (for the full term of copyright in the Article, including any extensions thereto) license (a) to publish, reproduce, distribute, display and store the Article in all forms, formats and media whether now known or hereafter developed (including without limitation in print, digital and electronic form) throughout the world, (b) to translate the Article into other languages, create adaptations, summaries or extracts of the Article or other Derivative Works (including, without limitation, the Video) or Collective Works based on all or any portion of the Article and exercise all of the rights set forth in (a) above in such translations, adaptations, summaries, extracts, Derivative Works or Collective Works and (c) to license others to do any or all of the above. The foregoing rights may be exercised in all media and formats, whether now known or hereafter devised, and include the right to make such modifications as are technically necessary to exercise the rights in other media and formats. If the "Open Access" box has been checked in **Item 1** above, JoVE and the Author hereby grant to the public all such rights in the Article as provided in, but subject to all limitations and requirements set forth in, the CRC License.

jove JOURNAL OF VISUALIZED EXPERIMENTS
1 Alewife Center #200
Cambridge, MA 02140
tel. 617.945.9051
www.jove.com

4. Retention of Rights in Article. Notwithstanding the exclusive license granted to JoVE in **Section 3** above, the Author shall, with respect to the Article, retain the non-exclusive right to use all or part of the Article for the non-commercial purpose of giving lectures, presentations or teaching classes, and to post a copy of the Article on the Institution's website or the Author's personal website, in each case provided that a link to the Article on the JoVE website is provided and notice of JoVE's copyright in the Article is included. All non-copyright intellectual property rights in and to the Article, such as patent rights, shall remain with the Author.

5. Grant of Rights in Video – Standard Access. This **Section 5** applies if the "Standard Access" box has been checked in **Item 1** above or if no box has been checked in **Item 1** above. In consideration of JoVE agreeing to produce, display or otherwise assist with the Video, the Author hereby acknowledges and agrees that, Subject to **Section 7** below, JoVE is and shall be the sole and exclusive owner of all rights of any nature, including, without limitation, all copyrights, in and to the Video. To the extent that, by law, the Author is deemed, now or at any time in the future, to have any rights of any nature in or to the Video, the Author hereby disclaims all such rights and transfers all such rights to JoVE.

6. Grant of Rights in Video – Open Access. This **Section 6** applies only if the "Open Access" box has been checked in **Item 1** above. In consideration of JoVE agreeing to produce, display or otherwise assist with the Video, the Author hereby grants to JoVE, subject to **Section 7** below, the exclusive, royalty-free, perpetual (for the full term of copyright in the Article, including any extensions thereto) license (a) to publish, reproduce, distribute, display and store the Video in all forms, formats and media whether now known or hereafter developed (including without limitation in print, digital and electronic form) throughout the world, (b) to translate the Video into other languages, create adaptations, summaries or extracts of the Video or other Derivative Works or Collective Works based on all or any portion of the Video and exercise all of the rights set forth in (a) above in such translations, adaptations, summaries, extracts, Derivative Works or Collective Works and (c) to license others to do any or all of the above. The foregoing rights may be exercised in all media and formats, whether now known or hereafter devised, and include the right to make such modifications as are technically necessary to exercise the rights in other media and formats. For any Video to which this Section 6 is applicable, JoVE and the Author hereby grant to the public all such rights in the Video as provided in, but subject to all limitations and requirements set forth in, the CRC License.

7. Government Employees. If the Author is a United States government employee and the Article was prepared in the course of his or her duties as a United States government employee, as indicated in **Item 2** above, and any of the licenses or grants granted by the Author hereunder exceed the scope of the 17 U.S.C. 403, then the rights granted hereunder shall be limited to the maximum rights permitted under such statute. In such case, all provisions contained herein that are not in conflict with such statute shall remain in full force and effect, and all provisions contained herein that do so conflict shall be deemed to be amended so as to provide to JoVE the maximum rights permissible within such statute.

8. Likeness, Privacy, Personality. The Author hereby grants JoVE the right to use the Author's name, voice, likeness, picture, photograph, image, biography and performance in any way, commercial or otherwise, in connection with the Materials and the sale, promotion and distribution thereof. The Author hereby waives any and all rights he or she may have, relating to his or her appearance in the Video or otherwise relating to the Materials, under all applicable privacy, likeness, personality or similar laws.

9. Author Warranties. The Author represents and warrants that the Article is original, that it has not been published, that the copyright interest is owned by the Author (or, if more than one author is listed at the beginning of this Agreement, by such authors collectively) and has not been assigned, licensed, or otherwise transferred to any other party. The Author represents and warrants that the author(s) listed at the top of this Agreement are the only authors of the Materials. If more than one author is listed at the top of this Agreement and if any such author has not entered into a separate Article and Video License Agreement with JoVE relating to the Materials, the Author represents and warrants that the Author has been authorized by each of the other such authors to execute this Agreement on his or her behalf and to bind him or her with respect to the terms of this Agreement as if each of them had been a party hereto as an Author. The Author warrants that the use, reproduction, distribution, public or private performance or display, and/or modification of all or any portion of the Materials does not and will not violate, infringe and/or misappropriate the patent, trademark, intellectual property or other rights of any third party. The Author represents and warrants that it has and will continue to comply with all government, institutional and other regulations, including, without limitation all institutional, laboratory, hospital, ethical, human and animal treatment, privacy, and all other rules, regulations, laws, procedures or guidelines, applicable to the Materials, and that all research involving human and animal subjects has been approved by the Author's relevant institutional review board.

10. JoVE Discretion. If the Author requests the assistance of JoVE in producing the Video in the Author's facility, the Author shall ensure that the presence of JoVE employees, agents or independent contractors is in accordance with the relevant regulations of the Author's institution. If more than one author is listed at the beginning of this Agreement, JoVE may, in its sole discretion, elect not take any action with respect to the Article until such time as it has received complete, executed Article and Video License Agreements from each such author. JoVE reserves the right, in its absolute and sole discretion and without giving any reason therefore, to accept or decline any work submitted to JoVE. JoVE and its employees, agents and independent contractors shall have

jove JOURNAL OF VISUALIZED EXPERIMENTS
1 Alewife Center #200
Cambridge, MA 02140
tel. 617.945.9051
www.jove.com

ARTICLE AND VIDEO LICENSE AGREEMENT

full, unfettered access to the facilities of the Author or of the Author's institution as necessary to make the Video, whether actually published or not. JoVE has sole discretion as to the method of making and publishing the Materials, including, without limitation, to all decisions regarding editing, lighting, filming, timing of publication, if any, length, quality, content and the like.

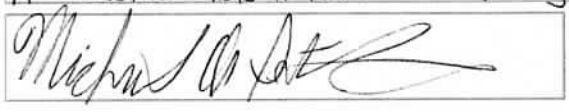
11. Indemnification. The Author agrees to indemnify JoVE and/or its successors and assigns from and against any and all claims, costs, and expenses, including attorney's fees, arising out of any breach of any warranty or other representations contained herein. The Author further agrees to indemnify and hold harmless JoVE from and against any and all claims, costs, and expenses, including attorney's fees, resulting from the breach by the Author of any representation or warranty contained herein or from allegations or instances of violation of intellectual property rights, damage to the Author's or the Author's institution's facilities, fraud, libel, defamation, research, equipment, experiments, property damage, personal injury, violations of institutional, laboratory, hospital, ethical, human and animal treatment, privacy or other rules, regulations, laws, procedures or guidelines, liabilities and other losses or damages related in any way to the submission of work to JoVE, making of videos by JoVE, or publication in *JoVE* or elsewhere by JoVE. The Author shall be responsible for, and shall hold JoVE harmless from, damages caused by lack of sterilization, lack of cleanliness or by contamination due to the making of a video by JoVE its employees, agents or independent contractors. All sterilization, cleanliness or decontamination procedures shall be solely the responsibility of the Author and shall be undertaken at the Author's expense. All indemnifications provided herein shall include JoVE's attorney's fees and costs related to said losses or damages. Such indemnification and holding harmless shall include such losses or damages incurred by, or in connection with, acts or omissions of JoVE, its employees, agents or independent contractors.

12. Fees. To cover the cost incurred for publication, JoVE must receive payment before production and publication the Materials. Payment is due in 21 days of invoice. Should the Materials not be published due to an editorial or production decision, these funds will be returned to the Author. Withdrawal by the Author of any submitted Materials after final peer review approval will result in a US$1,200 fee to cover pre-production expenses incurred by JoVE. If payment is not received by the completion of filming, production and publication of the Materials will be suspended until payment is received.

13. Transfer, Governing Law. This Agreement may be assigned by JoVE and shall inure to the benefits of any of JoVE's successors and assignees. This Agreement shall be governed and construed by the internal laws of the Commonwealth of Massachusetts without giving effect to any conflict of law provision thereunder. This Agreement may be executed in counterparts, each of which shall be deemed an original, but all of which together shall be deemed to me one and the same agreement. A signed copy of this Agreement delivered by facsimile, e-mail or other means of electronic transmission shall be deemed to have the same legal effect as delivery of an original signed copy of this Agreement.

A signed copy of this document must be sent with all new submissions. Only one Agreement required per submission.

CORRESPONDING AUTHOR:

Name:	Michael A. Stiffler or Rama Ranganathan
Department:	Green Center for Systems Biology
Institution:	University of Texas Southwestern Medical Center
Article Title:	A Protocol for Whole-Protein Saturation Mutagenesis Utilizing High-Throughput Sequencing
Signature:	 Date: 11 Sept 2015

Please submit a signed and dated copy of this license by one of the following three methods:

1) Upload a scanned copy of the document as a pfd on the JoVE submission site;
2) Fax the document to +1.866.381.2236;
3) Mail the document to JoVE / Attn: JoVE Editorial / 1 Alewife Center #200 / Cambridge, MA 02139

For questions, please email submissions@jove.com or call +1.617.945.9051

December 9, 2015

Dear Sephorah Zaman,

Please find attached our revised manuscript with changes made to incorporate your questions and the comments of the reviewers. We have also included revisions to Figure 1, and Tables 4, 5, 6, 9 and the Table of Materials. All changes to the revised protocol text have been tracked (in Microsoft Word); responses to your comments and those of the reviewers are provided below. Thank you again for considering our work.

Sincerely,
Michael Stiffler

1) All of your previous revisions have been incorporated into the most recent version of the manuscript. In addition, Editor may have made minor copy edits to your manuscript and formatting changes to comply with the JoVE format. Please maintain these changes. On the JoVE submission site, you can find the updated manuscript under "file inventory" and download the microsoft word document. ***Please use this updated version for any future revisions and track all changes using the track changes function in Microsoft Word****.*

Changes were made and followed in track changes on document 54119_R1_100815.docx.

2) With any additional highlighting (see Visualization notes below), please ensure that the length of the highlighted portion of the protocol doesn't exceed 2.75 pg.

We have checked that the length of highlighted text does not exceed 2.75 pgs.

3) Formatting: Email addresses should not be noted with superscripts on the title page.

Subscripts have been removed.

4) Visualization: The highlighting in the manuscript is at times discontinuous. Please check that all vital steps in the protocol are highlighted. For example:

We have attempted to make the highlighted portions more continuous, and more focused on the vital steps of the protocol.

a) 2.2.1.3-Is centrifugation always performed?

This is now highlighted.

b) 2.2.2.5-Is this step integral to the generation of the sublibraries mentioned in 2.4?

This step is now highlighted.

c) Is Step 2.3.2.2 integral to the production of cloning vectors?

This step is now highlighted.

d) Is 4.1.2.3 a necessary step?

This step is now highlighted.

e) When revising the highlighted portion of the manuscript, note that Section 4.2 should not be filmed, given that it deals with running a MATLAB script/complex equations.

This section is not highlighted.

5) Branding: Q5 is a registered trademark of NEB. Please remove the commercial term "Q5" from all Tables other than the Table of Materials. (use the general terms of "Reaction Buffer" or "DNA Polymerase"). In addition, replace "High QC Enhancer" with the general term "QC Enhancer."

Tables 4, 5, 6, and 9 have been changed accordingly. The table of materials has also been edited (to note that "PCR additive" called for in these tables is called "GC enhancer" as included with Q5 polymerase).

6) Discussion: Please elaborate on the limitations of this procedure, and emphasize the critical steps of the method.

We have stressed in the Discussion the important limitation of this procedure, being its application to comprehensive studies of the effects of multiple mutations across a protein (please see response to suggested revision 2 for Reviewer #1). We have also made additions to the first paragraph of the Discussion to highlight critical steps of the method.

7) Please take this opportunity to thoroughly proofread your manuscript to ensure that there are no spelling or grammatical errors. Your JoVE editor will not copy-edit your manuscript and any errors in your submitted revision may be present in the published version.

Done.

8) Please disregard the comment below if all of your figures are original.
If you are re-using figures from a previous publication, you must obtain explicit permission to re-use the figure from the previous publisher (this can be in the form of a letter from an editor or a link to the editorial policies that allows you to re-publish the figure). Please upload the text of the re-print permission (may be copied and pasted from an email/website) as a Word document to the Editorial Manager site in the "Supplemental files (as requested by JoVE)" section. Please also cite the figure appropriately in the figure legend, i.e. "This figure has been modified from [citation]."

All figures are original.

Response to Reviewers' comments

We sincerely thank the reviewers for interesting and valuable comments on our work. We hope this protocol will serve as a resource for others wishing to carry-out similar high-throughput mutagenesis studies. Below we provide a description and point-by-point response to each of the reviewer's comments. We have addressed the points through modifications of the text and figures. Please note that reviewers' comments are italicized and our responses in standard font.

Reviewers' comments:

Reviewer #1:

Manuscript Summary:

This manuscript presents a comprehensive experimental and computational protocol for performing saturation mutagenesis and selection to quantify fitness effects of all possible amino-acid mutations to a protein of interest. The protocol uses TEM-1 beta-lactamase in E. coli as a model protein under the selective pressure of ampicillin, a model system the authors recently used in a study of the relationship between selection strength and evolvability (Stiffler et al., 2015, Cell 160, 882-892).

This protocol represents a technological advance because it allows for sequencing throughput to be conserved by splitting the gene into several groups which are each independently mutagenized and cloned into a distinct plasmid backbones with unique primer binding sites for each subgroup. This allows for each mutation sub-library to be independently targeted during preparation of sequencing samples, thus avoiding the sequencing of gene regions that will not contain mutations in a given sub-library.

The protocol is detailed enough so that in can be replicated, but there are some areas detailed below that require revisions to more clearly depict the technological advance (sub-libraries linked to unique orthogonal primers) and its unique strengths.

1. Figure 1 currently depicts all plasmids and DNA fragments identically, without visual clarification of the orthogonal primer pairs and sub-library regions. Since the use of orthogonal primer pairs physically linked to each sub-library is a key element of this protocol, Figure 1 should be revised to more clearly convey this idea. The figure can be improved by more precisely highlighting the region of each sub-library group in the schematics of the plasmids and PCR amplicons. In other words, the red bar (mutagenized region) should be about 1/5th the length of the entire gene region and should be placed at the appropriate location within the depiction of each sub-library. Furthermore, the orthogonal primer sites are not represented in the diagram (all primer binding sites are shown using the same green marker). Each orthogonal primer pair could be depicted with a different color, and/or labeled OP-1 F/R, OP-2 F/R, etc, to show that there are unique sequences in the backbone for each sub-library. These changes together should visually clarify the idea of sub-libraries and how they are amplified with orthogonal primers to create targeted sequencing libraries that focus on mutated regions of the gene.

We have modified Figure 1 to more clearly indicate the usage of the orthogonal primers pairs. Admittedly, the figure is quite large so we wished to be as conservative as possible in terms of adding additional information, labels, etc. For instance, we attempted to label each orthogonal primer pair with a different color as suggested, however this became unwieldy when attempting to depict the combined plasmid library. We settled on labeling the individual plasmid and primers with OP1, OP2, etc as suggested. Likewise, we attempted to more clearly indicate the sub-library groups within the gene (the red region is meant to depict the entire gene, the X is meant to generically indicate mutation), but this was not amenable to labeling the combined plasmid library; in that case we hope that the indication of positions comprising each NNS sub-library group (26-78, 79-132, etc) will suffice to indicate the sub-library group regions.

2. Regarding the novelty of this protocol, which is the use of orthogonal primer pairs linked to each sub-library, I think the authors need to revise their claims about what limitations this overcomes vs. what this doesn't achieve. The statement that this protocol "accommodates genes of length significantly longer than sequencing read lengths" (line 40) requires further explanation and clarification. Indeed, as the authors state in the introduction, more "standard" shotgun sequencing approaches can and have been used to obtain mutation frequency data for similar experiments, with the disadvantage that many reads are wasted reading wild-type sequences. The conservation of sequencing depth by confining reads to regions containing mutations seems to be the only advantage in using the orthogonal primer / sub-library system over a shotgun approach to sequencing the entire gene. This ability to gain more information with less sequencing depth is certainly an improvement over standard protocols, but I suggest the authors revise statements such as "this technique overcomes the limitations of short sequencing reads" (line 578) since there are other limitations of short sequencing reads (such as the inability to observe mutation linkage across an entire gene) which this method does not address.

We thank the reviewer for making this point. We have attempted to be clearer about the advantages, and limitations, of the methods we proposed.

Regarding advantages: We have modified the text to focus on the important advantages: we state that splitting the gene into sub-library groups (more specifically, creating a correspondence between the location of the mutations within the gene and the sequencing read length) has the advantage of maximizing sequencing read depth/number of useful reads. The use of orthogonal primers to barcode each sub-library group gives the additional advantages of reducing workload and assuring that all mutations experience the same level of selection, since the sub-library groups can all be combined and subjected to simultaneous selection as opposed to having to do selections independently for each sub-library group. We have made changes to the short abstract, long abstract, Introduction (paragraphs three and four), and Discussion (paragraphs one and four). We have limited stating that the advantage of the protocol is “accommodating genes of length significantly longer than sequencing read lengths”. We now simply mention in the introduction (third paragraph) that our previous approach of using sub-library groups for evaluation of whole-protein saturation mutagenesis libraries is “applicable to genes of size longer than the sequencing read length limit”, and similarly in the Discussion (first paragraph).

Regarding limitations, we acknowledge that the method is limited in the ability to examine the effects of multiple mutations (mutation linkage) across a gene. We stress that one

could study multiple mutations if such mutations are confined within the NNS sub-library groups, or alternatively if the library is re-constructed in an alternate genotype background. We have modified the text by italicizing the sentence in paragraph two of the Discussion: "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". We followed this with the additional sentence: "This criterion therefore excludes the application of the protocol towards comprehensive studies of multiple mutations across a protein".

3. The example documented in the analysis code appears to use 10 sub-libraries (there are 10 "forward" and 10 "reverse" anchor sequences specified, but I might not be understanding the meaning of the anchor sequences completely). The protocol in this manuscript describes using 5 sub-libraries. For consistency I recommend the authors revise the analysis code to match the subgroups used in the protocol and/or clarify in the code exactly what the "anchor sequences" are.

There should be five "forward" and five "reverse" anchor sequences defined in the code, we have fixed the code accordingly and thank the reviewer for catching this. We have added a few additional lines to the code to describe the meaning of the anchor sequences (these are sequences which directly flank the sub-library groups and are introduced during the preparation of samples for high-throughput sequencing (protocol step 4.1). Since they do not contain mutations they can be used during the processing of the sequencing data to pull-out the sub-library groups).

4. On line 582, I believe "mutations" should be replaced with "sub-libraries". In conjunction with the revision #2 suggested above, I think the authors can further clarify here that without orthogonal primers, if one still wants to conserve sequencing depth by confining reads to regions with mutations, each sub-library would need to be subjected to selection separately (raising the possibility that sub-libraries experience different levels of selection).

We did mean mutations here, in that sense that since all the sub-library groups are mixed and selected together, all the mutations are likewise present during selection and should thus experience the same level of selection (as opposed to if the sub-libraries were selected independently, then all mutations might not experience the same level of selection, as the reviewer points out). We have clarified in the text that this is an advantage of using the orthogonal primers in combination with the sub-library strategy (see response to suggested revision 2 above).

5. In protocol step 2.3.5.2, it appears that 10 ul of cells are diluted into 20 ul of DNA obtained in step 2.3.4 and then electroporated. Is that indeed the case? Just want to make sure this isn't a typo and some smaller amount of DNA is meant to be added to the cells, since standard electroporation protocols recommend adding no more DNA than some small percentage (perhaps 5-10%?) of the total volume of cells.

We do use the entire 20 µL volume obtained after purifying the ligation reaction for electroporation, transforming into 10 µL of cells. While typically for transformations the volume

of DNA in water added is a small percentage of cell volume, in this case we wish to obtain a large library (thus the entire ligation is transformed) and, more importantly, are using highly-concentrated cells. Admittedly, we haven't carried out a comprehensive examination of the effect of DNA and cell volumes on transformation efficiencies, though anecdotally transforming only half the volume of the ligation into the same volume of cells produced lower transformation efficiency.

Major Concerns:
N/A

Minor Concerns:
N/A

Additional Comments to Authors:
N/A

Reviewer #2:
Manuscript Summary:
This article presents a protocol for the generation of saturation mutagenesis libraries over a length of sequence by dividing up library generation in multiple sublibraries, each covering a specific segment. This protocol also optimizes the high-throughput sequencing of the libraries even when the length is longer than the typical read length of NGS. Thus, this protocol can be very useful for the comprehensive exploration of a protein's fitness landscape.
The techniques presented here, including the segmental approach for construction of the mutagenesis library, have been previously described. There is one significant innovation, though, namely the utilization of orthogonal primer barcodes. The barcodes tag each segment for multiplexing purposes, allowing the simultaneous selection of all sublibraries and maximizing relevant sequencing coverage.

Major Concerns:
My main criticism is that this article is very hard to follow because Figure 1 does not provide enough detail about the different PCR reactions involved. I think the article needs a separate Figure 1a explaining these different PCR reactions, particularly the ones involved in the generation of NNS sublibraries (with current Fig 1 becoming Fig 1b).

We do admit that Figure 1 is quite large and risks creating confusion (as opposed to its intended goal of simplifying communication) with the addition of more information, labels, etc. We made a few modest changes in an attempt to address the reviewer's concerns: we indicated the number of PCRs required at the relevant steps, e.g. "create NNS sub-libraries by SOE PCR (3 PCRs per position)". Further, we have indicated to the left of the figure the actual protocol steps, for reference to the main text.

Minor Concerns:
1. In several places(including the title) authors describe the system for mutagenesis as "mutagenesis using high-thorughput sequencing"; this is misleading because mutagenesis is not

generated by sequencing. It is designed to facilitate sequencing. So I would suggest changing the title to: "A Protocol for whole-Protein Saturation Mutagenesis Efficiently Coupled to High Throughput Sequencing for Comprehensive Measurement of Fitness Landscape".
Other sentences that have this problem include:
Line 572: "mutagenesis, using high throughput sequencing technology"
Line 621: "high-throughput sequencing-based approach for whole-protein saturation.

We thank the reviewer for making this point. Our original intention was for "saturation mutagenesis using high-throughput sequencing" to be short for both the construction of a single-site randomized amino acid library, as well as the subsequent assessment of the effects of each mutation (via selection and high-throughput sequencing). This we acknowledge is confusing and requires clarification. We have edited the text (and the title) accordingly, e.g. changing "protocol for comprehensive single-site saturation mutagenesis utilizing high-throughput sequencing" to "protocol for the functional assessment of comprehensive single-site saturation mutagenesis libraries utilizing high-throughput sequencing". Changes were made to the title, short and long abstract, Introduction (paragraph four), and Discussion (paragraphs one and four).

2. The discussion does a good job of explaining the advantages and possible alternative uses of the system, but could elaborate more on the disadvantages. For example, restricting randomization to individual segments prevents exploration of epistatic interations involving mutations outside the specific segments. There is the uneven representation of random individual mutants (illustrated in Fig. 3a) also restricts access to sequence space. For example, if you try to evolve gain-of-function mutations some solutions may be much harder to find than others.

We acknowledge that the method is limited in the ability to examine the effects of multiple mutations across a gene. The first reviewer also pointed this out, and we have attempted to clarify this point with changes to the main text (please see our response to suggested revision 2 for the first reviewer above).

With regards to an uneven representation of random mutations: Since the accuracy of determining a mutational effect in our method is based on the number of read counts (equation in protocol step 4.2.5), and we wish to do this for all possible single amino acid mutations, it is imperative that the library is constructed in a manner to achieve as even a distribution of mutations as possible. Each mutation is created as a separate PCR product (protocol step 2.2.1-2.2.3) and accurately quantified (protocol step 2.2.4), and all positions mixed in equal amounts according to this quantification (protocol step 2.3.1). Given this careful mixing of the library, the distribution in Figure 3a should represent a best-case scenario (or fairly close to it) for the even mixing of a whole-protein saturation mutagenesis library; we thus feel our method provides a fairly even representation of mutations. We acknowledge that the NNS mutation scheme and SOE PCR method introduce biases, so some improvement could be had in future work.

Sequence space (at the single amino acid distance) is relatively unrestricted in our representative results, as 98.9% of all mutations had at least one count from sequencing of the pre-selection culture. As the effect of each mutation in the library is measured, as long as the mutation is in the library to begin with and the effect is sufficient (statistically significant increase in number of counts in selected versus pre-selected population), the method should not present limitations to evolving (or more specifically, finding) gain-of-function mutations. In fact, it should improve identification of such mutations since they do not have to increase to a high

proportion of the population to be identified, as in a typical directed evolution experiment employing selection.

Additional Comments to Authors:
N/A

Reviewer #3:
Manuscript Summary:
The current manuscript describes a high-throughput sequencing-based approach for whole-protein engineering via individual position saturation mutagenesis. 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. Therefore, the limitation of sequence reading length is mitigated.
The authors thoroughly detailed the three main steps, library construction, selection, and sequence amplification and analysis, involved in the protocol. The presented result clearly documents the usefulness of their approach in individual position saturation mutagenesis based protein engineering.
Sequence reading length has been a major limitation of using next-generation-sequencing (NGS) approach in protein engineering. The reported approach provides a solution to the problem. Future researches could take advantage of the reported concept to further simplify the process. For example the reviewer felt that the construction of orthogonal primer restricted sub-library is too laborious. Much simpler experiment can be utilized to reach the same goal.

We agree that the method could be simplified in some respects, depending on the goals of the researcher. We stress that our protocol takes great pains to ensure that the distribution of mutations is as even as possible, such that all mutations are accurately quantified (please see response to suggested revisions 2 for Reviewer #2 above). If one relaxes this criterion, along with the criterion that the library consists only of all possible single amino acid mutations, library construction could be simplified. As we mention to in the Discussion, one could make mutations by error-prone PCR or similar technique. As long as the mutations are confined into sub-library windows, our approach could be applied. Further, if one performed a standard shotgun sequencing approach, construction of the library into sub-library groups could be done away with, but with the loss of sequencing depth (please see response to suggested revision 2 for Reviewer #1).

Major Concerns:
N/A

Minor Concerns:
N/A

Additional Comments to Authors:
N/A

Supplemental code file (if applicable)

Click here to access/download

Supplemental code file (if applicable)

NNS_PrimerDesign.m

Supplemental code file (if applicable)