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*