Massively parallel single-nucleotide mutagenesis using reversibly terminated inosine

Gabe Haller[1], David Alvarado[1], Kevin McCall[1], Robi D Mitra[2], Matthew B Dobbs[1,3] & Christina A Gurnett[1,4,5]

Large-scale mutagenesis of target DNA sequences allows researchers to comprehensively assess the effects of single-nucleotide changes. Here we demonstrate the construction of a systematic allelic series (SAS) using massively parallel single-nucleotide mutagenesis with reversibly terminated deoxyinosine triphosphates (rtITP). We created a mutational library containing every possible single-nucleotide mutation surrounding the active site of the TEM-1 β-lactamase gene. When combined with high-throughput functional assays, SAS mutational libraries can expedite the functional assessment of genetic variation.

The difficulty of assessing the effects of genetic variants in the human population is a major obstacle confronting precision medicine. The power and utility of rapid, accurate mutation library production has recently been demonstrated[1–5]. Several large-scale mutagenesis methods have been used to screen both genes and noncoding loci. Mutagenesis libraries have been created by using a variety of methods, including error-prone or inosine-containing PCR, or by using synthetic oligonucleotides[1–10]. However, each of these approaches has its limitations. While error-prone PCR is inexpensive, products often have multiple mutations and high transition:transversion ratios; and the production of oligonucleotides requires substantial up-front financial investment.

Here we demonstrate the construction of a SAS by performing mutagenesis of segments of DNA using rtITPs. By combining cycle termination used in Sanger sequencing, reversible termination used in Illumina sequencing, and inosine's ability to base pair with each of the four bases, we have systematically incorporated single inosine molecules into DNA molecules, allowing the introduction of one (and only one) mutation per molecule during PCR amplification.

Briefly, we generated rtITPs through sequential enzymatic reaction of reversibly terminated deoxyATPs (**Supplementary Fig. 1** and **Supplementary Note**). This multistep synthesis was required because of adenosine deaminase's inability to deaminate ATP and the lack of commercially available rtITP. A linear amplification was then performed using a biotinylated forward primer, a special polymerase (Firebird475, Firebird Biomolecular Sciences, LLC) required for the incorporation of reversibly terminated nucleotides, and a dNTP pool containing >50% w/v rtITP (**Fig. 1a**). Incorporation of rtITP molecules and the termination of their extension was confirmed by amplification with a 5′-ROX-labeled primer (**Supplementary Fig. 2**). For short products, bands consisting of products >20 bp and less than full length were gel extracted to remove unwanted full-length product and unextended primers. Linear amplification products were subsequently isolated by hybridization to streptavidin-coated magnetic beads. The 3′-O-NH2 termination moiety on the rtITPs was then removed by exposure to sodium nitrite (0.7 M, pH 5.5), and products were extended using a high-fidelity polymerase. Beads were then washed to remove template DNA and amplified to produce PCR products containing each of the four alternative nucleotides at each site where inosines were incorporated.

We used SAS mutagenesis to create a library of the ampicillin-resistance gene (*AmpR*) encoding TEM-1 β-lactamase. First, 217 bp containing a portion of the active site of the ampicillin-resistance gene was amplified and cloned inframe into a wild-type β-lactamase plasmid also containing the kanamycin-resistance gene, allowing the plasmid library to replicate efficiently without ampicillin selection. Completely overlapping paired-end sequencing of this SAS library unselected for ampicillin resistance revealed that 33% of clones created with a 1:1 ratio of rtITP:dNTPs and >50% created with a 4:1 ratio of rtITP:dNTPs contained one (and only one) mutation. Importantly, only ~1% of clones contained >1 mutation for the 1:1 ratio and ~5% for the 4:1 ratio (**Fig. 1b**). This is a marked reduction in secondary mutations compared with previous methods, the most effective of which produced a similar rate of single mutations (33–47%) but a substantial proportion of molecules with secondary mutations (21–35%) (ref. 2). We initially limited our mutagenesis reactions to <250 bp so that they could be fully sequenced from both directions to distinguish true mutations from sequencing errors. We observed good correlation between true mutation counts and observed mutation counts, however, suggesting that shotgun sequencing can suffice to calculate mutation effect sizes (**Supplementary Fig. 3**). The nucleotide composition of generated mutations within the SAS library was similar at each nucleotide, resulting in an average transition:transversion ratio of 0.48 (**Fig. 1c** and **Supplementary Table 1**), while error-prone PCR showed the expected inflated transition:transversion ratio of 3.2.

[1]Department of Orthopaedic Surgery, Washington University, St. Louis, Missouri, USA. [2]Center for Genome Sciences and Systems Biology, Department of Genetics, Washington University School of Medicine, St. Louis, Missouri, USA. [3]Shriners Hospital for Children, St. Louis, Missouri, USA. [4]Department of Neurology, Washington University, St. Louis, Missouri, USA. [5]Department of Pediatrics, Washington University, St. Louis, Missouri, USA. Correspondence should be addressed to C.A.G. (gurnettc@neuro.wustl.edu).

RECEIVED 5 FEBRUARY; ACCEPTED 30 AUGUST; PUBLISHED ONLINE 3 OCTOBER 2016; DOI:10.1038/NMETH.4015

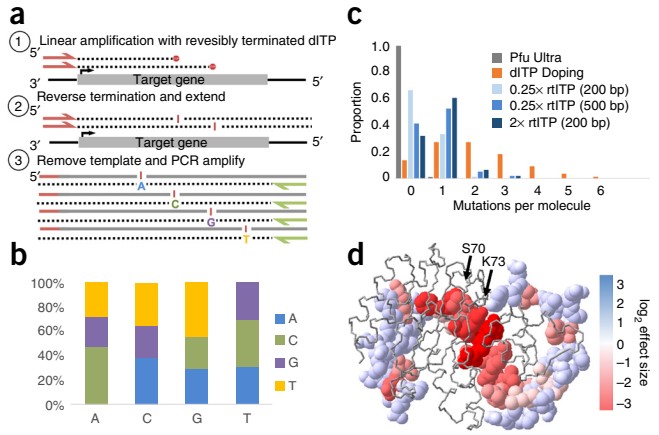

Figure 1 | Schematic and library characteristics of SAS mutagenesis. (**a**) Schematic of SAS mutagenesis. (**b**) Average proportions of nucleotide changes observed in the two SAS libraries (217 and 687 bp). The *Y*-axis is the original nucleotide, and colors represent the nucleotide to which the position was mutated. (**c**) Histogram of mutations per molecule for libraries under various reaction conditions. For Pfu Ultra, standard PCR was performed using Pfu as the polymerase. For dITP doping, dITP was included at a 1:1 ratio with dNTPs using Taq as the polymerase. (**d**) Crystal structure of TEM-1 β-lactamase with mutated sites pictured with space-filled molecular model and nontested sites with backbone only. Sites are colored according to their $\log_{10}$ effect size after ampicillin selection, with red indicating a greater negative effect size.

Using this SAS mutagenized library, we assessed the functional impact of mutations present within the first 217 bp of the *AmpR* gene. Upon submitting the SAS library to selection, we observed strong depletion of mutations resulting in nonconservative (i.e., hydrophobic to polar, etc.), nonsynonymous amino acid substitutions ($P < 2 \times 10^{-20}$, Mann–Whitney *U*) (**Supplementary Fig. 4**) with specific residues within the β-lactamase active site (S70 and K73) among the most strongly depleted with effect sizes dropping off further away from the active site (**Fig. 1d**). We also observed a strong correlation of our observed enrichment scores with those described previously[11] (Pearson's $r^2 = 0.31–0.68$; **Supplementary Fig. 5**).

Lastly, we sequenced the distal 687-bp segment of the *AmpR* gene to demonstrate the ability of SAS mutagenesis to create long mutational libraries. We appended unique molecular identifiable (UMI) sequences via primers with 20-mer random sequences followed by a universal 5′ overhang during the final PCR amplification in order to uniquely identify mutant DNA molecules during sequencing (**Supplementary Fig. 6**). Similar rates of singleton mutations (~50% total or ~8% singleton mutant molecules per 100 bp mutated) were observed within this long SAS library with slight drop-off along the length of the SAS library out to approximately 1 kb (**Supplementary Fig. 7**). The rate of incorporation of each nucleotide was also comparable to that observed for the first 200 bp of the SAS library (**Supplementary Tables 1** and **2**). These data suggest that long SAS libraries can be created without substantial change in mutagenic properties along the length of the product.

Mutagenesis using rtITP enables the rapid construction of comprehensive libraries containing all possible single-nucleotide changes within a region of DNA for a fraction of the cost of current methods. While saturation mutagenesis of splice sites and coding genes can now be performed at their native loci using CRISPR–Cas9 technologies[1], these experiments have thus far been limited because of size restrictions of oligonucleotide arrays. SAS mutagenesis will enable rapid, cost-effective production of homology-directed repair template pools to simultaneously assess the functional impact of a library of mutations within coding and noncoding loci. Overall, integration of SAS mutagenesis with high-throughput functional testing will enable the rapid assessment and fine-scale understanding of the millions of DNA variants present within the human population.

METHODS

Methods and any associated references are available in the online version of the paper.

Accession codes. Protein Data Bank, accession 1FQG for TEM-1β-lactamase.

Note: Any Supplementary Information and Source Data files are available in the online version of the paper.

ACKNOWLEDGMENTS

We thank C. Cruchaga and J. Budde and members of the Gurnett/Dobbs Lab for helpful discussion. This work was supported by a postdoctoral research fellowship (84291-STL) from the Shriners Hospital for Children (G.H.); US National Institutes of Health (NIH) grants (R01AR067715-01) (C.A.G. and M.B.D.) and (R01NS076993) (R.D.M.); and Shriners Hospital for Children research grant (85200-STL) (C.A.G. and M.B.D.)

AUTHOR CONTRIBUTIONS

G.H., D.A., R.D.M., M.B.D. and C.A.G. designed the study and wrote the manuscript. G.H. and K.M. performed experiments. All authors contributed to and approved the final manuscript.

COMPETING FINANCIAL INTERESTS

The authors declare competing financial interests: details are available in the online version of the paper.

1. Findlay, G.M., Boyle, E.A., Hause, R.J., Klein, J.C. & Shendure, J. *Nature* **513**, 120–123 (2014).
2. Kitzman, J.O., Starita, L.M., Lo, R.S., Fields, S. & Shendure, J. *Nat. Methods* **12**, 203–206 (2015).
3. Patwardhan, R.P. *et al. Nat. Biotechnol.* **30**, 265–270 (2012).
4. Patwardhan, R.P. *et al. Nat. Biotechnol.* **27**, 1173–1175 (2009).
5. Starita, L.M. *et al. Genetics* **200**, 413–422 (2015).
6. Smith, R.P. *et al. Nat. Genet.* **45**, 1021–1028 (2013).
7. Cirino, P.C., Mayer, K.M. & Umeno, D. *Methods Mol. Biol.* **231**, 3–9 (2003).
8. Copp, J.N., Hanson-Manful, P., Ackerley, D.F. & Patrick, W.M. *Methods Mol. Biol.* **1179**, 3–22 (2014).
9. McCullum, E.O., Williams, B.A., Zhang, J. & Chaput, J.C. *Methods Mol. Biol.* **634**, 103–109 (2010).
10. Gao, Y. *et al. Wei Sheng Wu Xue Bao* **54**, 97–103 (2014).
11. Stiffler, M.A., Hekstra, D.R. & Ranganathan, R. *Cell* **160**, 882–892 (2015).

ONLINE METHODS

Production of reversibly terminated deoxyinosine triphosphate. 3′-O-NH2-deoxyATP (Firebird Biomolecular Sciences, LLC) was enzymatically converted to 3′-O-NH2-deoxyinosine triphosphate by the sequential addition and heat inactivation of recombinant shrimp alkaline phosphatase (rSAP; New England BioLabs) to create 3′-O-NH2-deoxyadenosine; deamination with adenosine deaminase (Sigma) to create 3′-O-NH2-deoxyinosine; and a mixture of T4 PNK, pyruvate kinase and myokinase (adenylate kinase) (Sigma) to create 3′-O-NH2-deoxyinosine triphosphate. This product was then used directly in the SAS mutagenesis protocol. The ability of rtITP molecules to be incorporated into linear amplification products and to terminate polymerase extension was tested using a 5′-ROX-labeled primer identical in sequence to the biotinylated primer used to perform SAS mutagenesis on the *AmpR* gene. Fluorescently labeled linear amplification products were produced using varying concentrations of either rtATP or rtITP (**Supplementary Fig. 2**).

SAS mutagenesis. Linear amplification of target DNA was performed using a biotinylated primer, a 1:1 or 1:4 ratio of dNTPs and rtITP and Firebird Taq 475 (Firebird Biomolecular Sciences, LLC). This polymerase was specifically developed to incorporate 3′-O-NH2 linked nucleotides. The products were then bound to streptavidin beads and washed. Beads were then exposed to 70 mM sodium nitrite, pH 5.5, to reverse the termination, washed and cycle extended in the presence of template DNA. Upon extension, beads were washed, and DNA was eluted using 0.1 M NaOH followed by neutralization with 1 M Tris–HCl. Beads with bound DNA were then used as the DNA template for a PCR to produce final mutagenized products with randomly inserted nucleotides (A, C, G and T) in place of each inosine. A detailed protocol can be found in **Supplementary Note**.

Cloning and functional selection of TEM-1 β-lactamase. A dual selection plasmid (plasmid pGH1) was created by cloning the full TEM-1 β-lactamase gene amplified from plasmid pCMV6-XL6 into the plasmid pCR-Blunt-II-TOPO, which contains the kanamycin-resistance gene (*KanR*). A 217 bp segment containing the active site was then amplified from plasmid pGH1 containing both the *AmpR* and the *KanR* genes, and SAS mutagenesis was performed on the segment as described above. The inverse of the 217 bp segment within plasmid pGH1 was then amplified using primers that were the reverse complement of those used to amplify the 217 bp fragment. This PCR product was then Gibson assembled with the SAS library derived from the 217 bp fragment. This plasmid library was then transformed into XL10-Gold ultracompetent cells and grown in lysogeny broth (LB) medium to saturation in the presence of 50 μg/ml kanamycin. For ampicillin selection, kanamycin outgrowth was diluted 1:1,000 in 100 ml of selective media containing 200 μg/ml ampicillin. Cells were cultured for 2 h, centrifuged, washed by resuspending in 1 ml LB without antibiotic, centrifuged and washed again, then resuspended in 5 ml LB media without antibiotic and grown overnight to saturation. Plasmids were purified using Qiagen Mini Plasmid purification kit.

Sequencing and enrichment analysis. Mutated segments were sequenced in both directions with 250 bp paired-end reads using an Illumina MiSeq. The consensus sequence of each read pair was aligned using Novoalign. Mutation counts were obtained from pileups of aligned consensus sequencing reads. Enrichment scores were determined using the following formula for each observed mutation.

$$F_i^a = \log_{10}\left[\frac{N_i^{a,\mathrm{sel}}}{N_i^{a,\mathrm{unsel}}}\right] - \log_{10}\left[\frac{N_i^{\mathrm{wt,sel}}}{N_i^{\mathrm{wt,unsel}}}\right]$$

For validation, enrichment scores were compared to those observed previously for mutations in the TEM-1 β-lactamase gene using linear regression of enrichments scores compared with scores obtained previously. Projection of enrichment scores onto the crystal structure of TEM-1 β-lactamase (PDB:1FQG) was performed using Swiss-PBD viewer 4.1.0. Enrichment scores (F_i^a) were determined for allele a at position i comparing allele counts in the selected library ($N_i^{a,\mathrm{sel}}$) to the unselected library ($N_i^{a,\mathrm{unsel}}$) relative to wild-type allele counts.

Data availability. The data that support the findings of this study are available from the corresponding author upon request.