nature chemical biology

SUPPLEMENTARY INFORMATION

https://doi.org/10.1038/s41589-020-0599-5

In the format provided by the authors and unedited.

Multiplexed genomic encoding of non-canonical amino acids for labeling large complexes

Bijoy J. Desai and Ruben L. Gonzalez Jr. ✉

Department of Chemistry, Columbia University, New York, NY, USA. ✉e-mail: rlg2118@columbia.edu

SUPPLEMENTARY INFORMATION

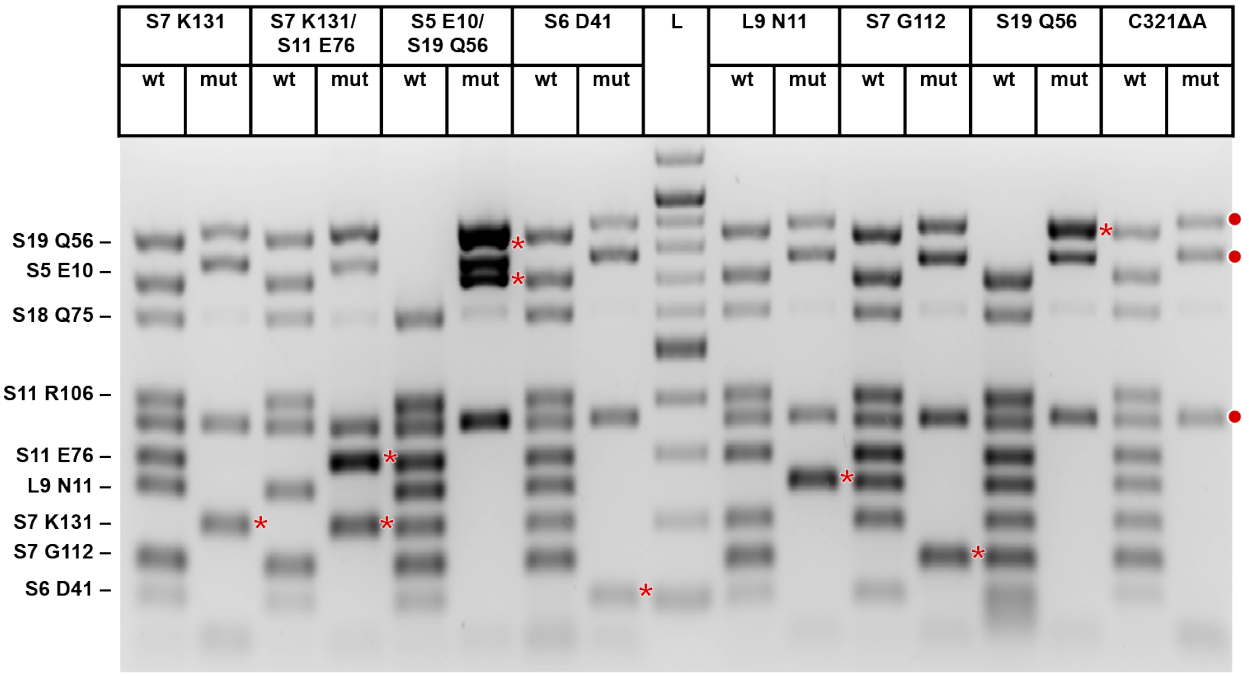
Multiplexed genomic encoding of non-canonical amino acids for labeling large complexes

Bijoy J. Desai and Ruben L. Gonzalez, Jr.*

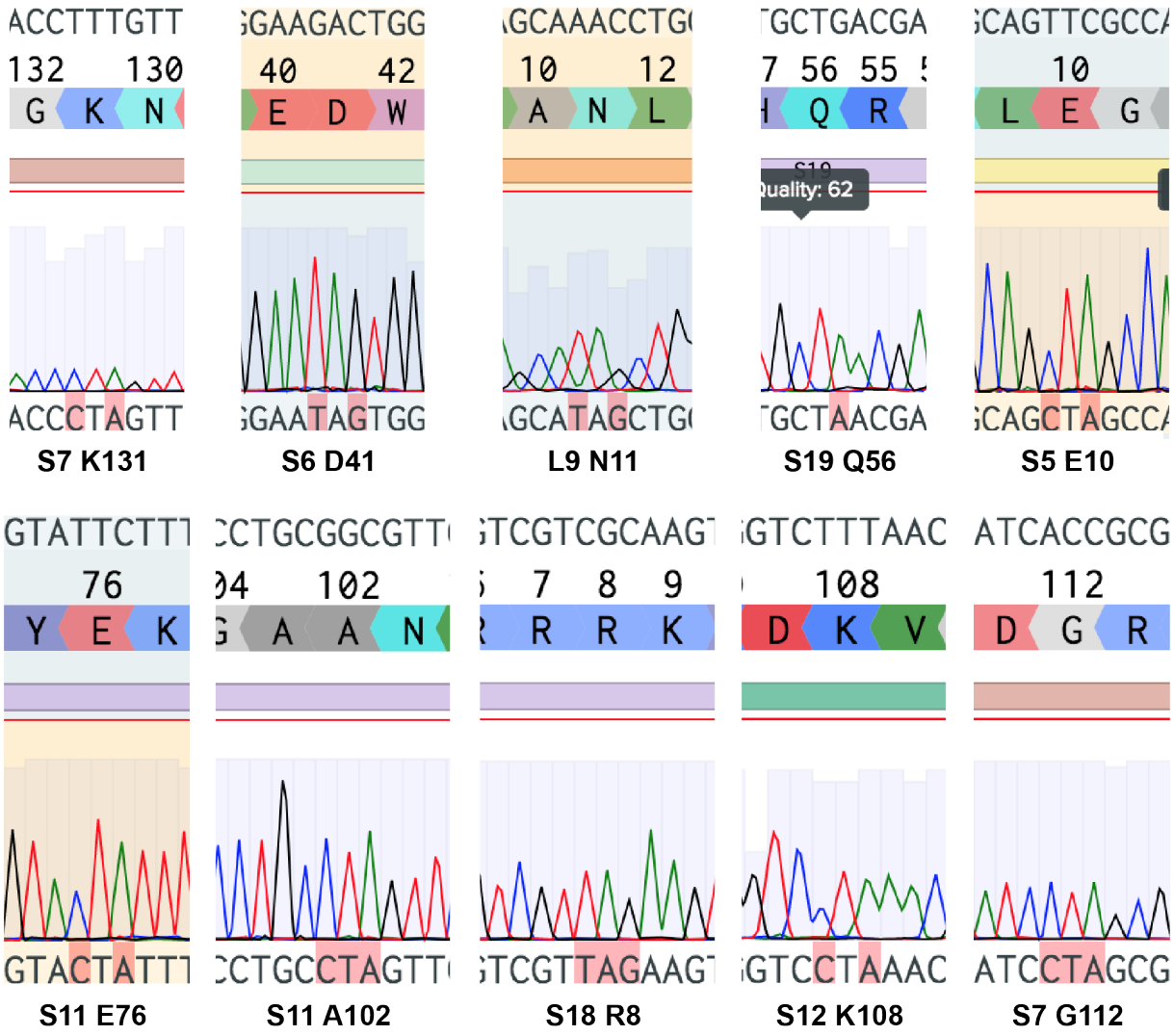
Department of Chemistry, Columbia University, 3000 Broadway, MC3126, New York, NY 10027, USA

*To whom correspondence should be addressed: Ruben L. Gonzalez Jr., Department of Chemistry, Columbia University, 3000 Broadway, MC3126, New York, NY 10027, USA, Tel.: (212) 854-1096; Fax: (212) 932-1289; Email: rlg2118@columbia.edu

Supplementary Figures

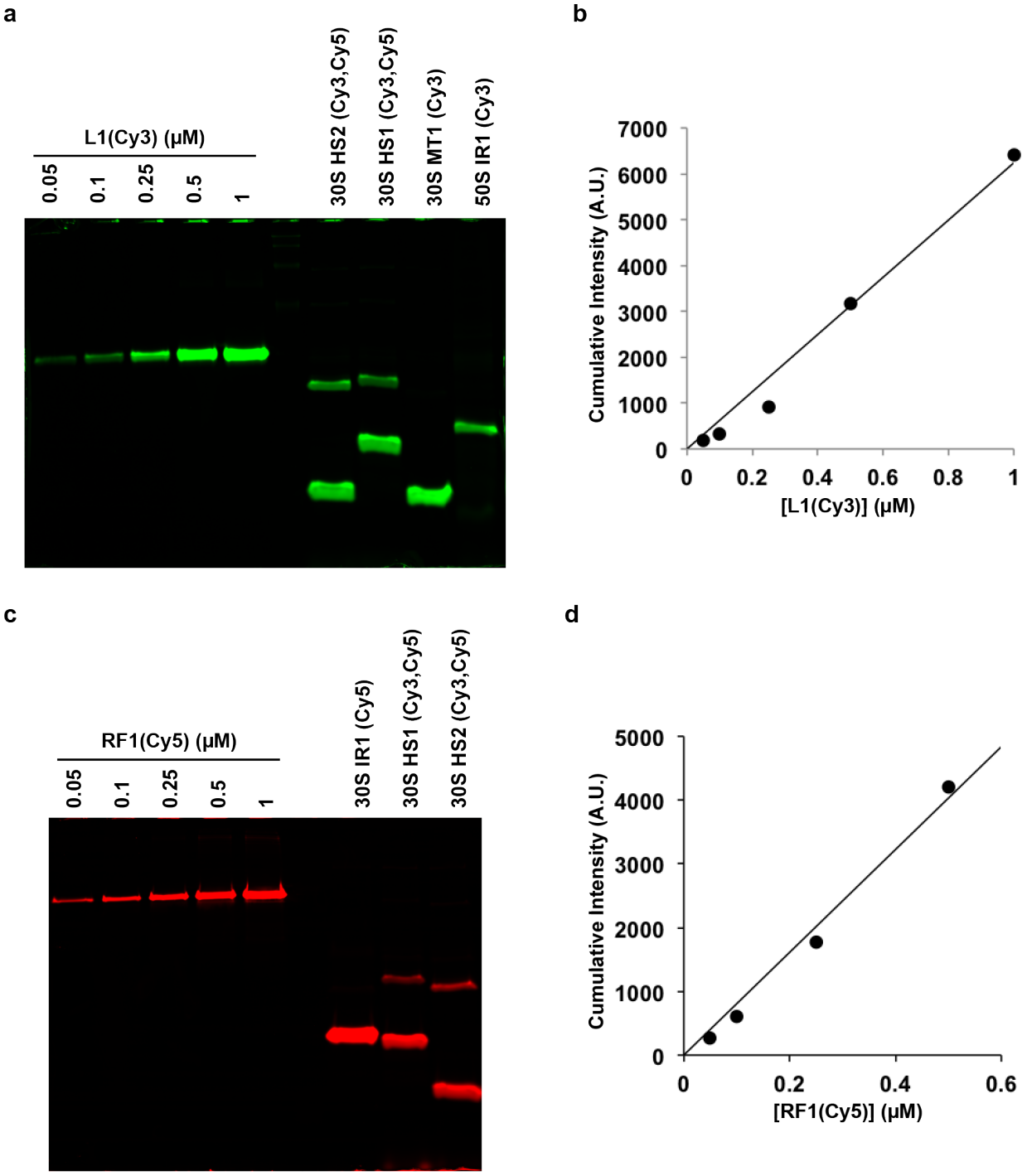


Supplementary Figure 1. MASC-PCR-based identification of colonies containing genomic TAG mutations in ribosomal protein genes isolated after six to eight rounds of MGE. 1.5 % agarose gel of MASC-PCR reactions performed using primers targeting either wild-type (wt) or mutant (mut) alleles and isolated single bacterial colonies as templates. Primers were designed such that each colony carrying a TAG mutation would give a uniquely sized PCR product (as indicated on the left-hand side of the gel). In these MASC-PCR reactions, the presence of a particular TAG mutation is indicated by the presence of a PCR product of a size corresponding to the TAG mutation in the lane labeled mut (indicated by the red ⋆s) and the concomitant absence of a PCR product of the same size in the lane labeled wt. Extraneous PCR products that were present in all wt or mut reactions were identified by performing MASC-PCR reactions on our tailored, parent C321ΔA strain (indicated by the red •s). The above experiment was repeated two more times, with similar results.



Supplementary Figure 2. Sanger sequencing-based confirmation of colonies containing genomic TAG mutations in ribosomal protein genes isolated after six to eight rounds of MGE. Each panel shows an alignment of the wildtype reference nucleotide- and amino acid sequences of the targeted ribosomal protein gene (Top Two Rows), Sanger sequencing trace (Middle Row), and nucleotide sequence corresponding to the Sanger sequencing trace (Bottom Row) for PCR products amplified from the genomic mutation region using isolated mutant colonies initially identified in the MASC-PCR screening. The amino acid sequences are shown as single letter codes. The Sanger sequencing traces are color-coded based on nucleotide: A is shown in green, T is shown in red, C is shown in blue, and G is shown in black. The

successfully mutated nucleotides are highlighted in pink in the nucleotide sequence corresponding to the Sanger sequencing trace.



Supplementary Figure 3. Quantification of labeling efficiency for 30S and 50S subunits purified from HS1, HS2, MT1, and IR1 mutant strains. Representative fluorescence emissions scans at excitation wavelengths of 532 nm for (a) Cy3 and 635 nm for (c) Cy5 of SDS-PAGE gels containing 12 μl of the indicated concentrations of Cy3- or Cy5-labeled protein

standards (L1 [Cy3] and RF1 [Cy5], respectively) and 12 µl of 1 µM of Cy3- and/or Cy5-labeled 30S or 50S subunits isolated from the indicated mutant strains as described in Methods. Linear standard curves plotted using the average cumulative fluorescence intensities of the protein bands from three independent experiments, corresponding to different concentrations of (b) L1 [Cy3] and (d) RF1 [Cy5]. The data points (black filled circles) in each standard curve were fit to a linear function of the form $y = mx + b$ (black lines), where y is the cumulative fluorescence intensity, x is the concentration of L1 [Cy3] or RF1 [Cy5], m is the slope, and b is y-intercept, which was required to go through zero. The equation for the standard curve describing L1 [Cy3] was $y = 6220x + 0$ and fit the data with an R^2 of 0.98 and the equation for the standard curve describing RF1 [Cy5] was $y = 8019.4x + 0$ and fit the data with an R^2 of 0.99.

Supplementary Tables

Supplementary Table 1. DNA oligonucleotides used for MGE. The nucleotides corresponding to the mutation are denoted in red and phosphorothioated nucleotides are followed by an asterisk (*).

Target for Mutation	DNA Oligonucleotide Sequence
S7 G112	G*C*A*T*CAGAAAGTTCGTTCGCCAGGCGCAGAGCCATGGATTTATC**CT** **A**GCGTTTACGAGCAGCTTCAACGATCCAACGCATTGCCAGAGCAT
S7 K131	A*C*G*T*TTCTTAACTGCAGTACC**C**T**A**GTTTTCTGCAGCATCAGAAAGTT CGTTCGCCAGGCGCAGAGCCATGGATTTATCACCGCGTTTACGA
S11 A102	A*A*T*G*TTAGTGATGCGGAAACCTGC**CTA**GTTCAGAGCACGAATAGTA GATTCGCGGCCTGGACCCGGACCTTTAACCATAACTTCCAGATT
S18 Q75	T*A*C*C*TTATCCTCTCAAAGTCGTATTAATGGACCGTGACCGATTACT**A** ATGGCGATCAGTGTACGGCAGCAGGGACAGGTAGCGAGCGCGTT
S18 R8	T*A*G*T*CGATCTCTTGAACGCCTTCCGCGGTGAAACGGCAGAACTT**CT** **A**ACGACGGAAATAACGTGCCATATGGTTAGTCTCCAGAATCTATC
S6 D41	T*T*G*T*GCAGTTTGTTGATCGGGTAAGCCAGCTGACGGCGGCCCCA**C**T **A**TTCCAGACGGTGGATCTTGCCTTCTGCACCAGTGATGGCAGCAG
L9 N11	G*C*A*T*AGCCCGCTTTAACGTTTACCTGATCACCCAGGCTACCCAG**C**T **A**TGCTACTTTATCAAGCAGAATAACTTGCATTACCTTATCCTCTC
S11 E76	G*G*A*C*CCGGACCTTTAACCATAACTTCCAGATTCTTGATGCCGTA**CTA** TTTCACGGCGTCAGCGCAACGCTCTGCTGCAACCTGAGCTGCAA
S5 E10	A*C*G*G*TTTTAGATACGCGGTTTACCGCGATCAGCTTTTCCTGCAG**C**T**A** GCCAGCTTGTTTTTCGATGTGAGCCATCTTACACCTCTACCTTA
S12 K108	G*A*A*C*GAGCCTGCTTACGGTC**C**T**A**AACGCCGGAGCAGTCAAGCGCA CCACGTACGGTGTGGTAACGAACACCCGGGAGGTCTTTAACACGA
S19 Q56	A*G*T*T*TGTGACCAACCATTTCGTCGGTTACAAATACCGGAACGTGCT**A** ACGACCATTATGGACAGCGATGGTCAAACCGATCATGTTAGGAA
S13 D11	TGCCGACGCCATAAATCGAAGTTAATGCGATTACGGCATGCTTATG**C**T**A** AGGAATGTTAATGCCTGCTATACGGGCCACTATGCACTCCT
S11 R106	A*A*C*C*GTTATGAGGGATCGGAGTCACATCAGTAATGTTAGTGAT**CTA**G AAACCTGCGGCGTTCAGAGCACGAATAGTAGATTCGCGGCCTG
RNase I A9	G*T*A*T*GAGTTCCACACCCATTATGAAAGCATTCTGGCGTAACGCC**TAA** TTGCTCGCGGTTTCTCTGCTTCCCTTCTCTTCTGCCAACGCCT

Supplementary Table 2. Percent enrichment of target mutations after performing the specified number of MGE cycles in the EcNR2 strain.

Target for mutation	Number of bp mutated	Number of MGE cycles	Percent enrichment[1]
L9 N11	2	6	7.4
S18 R8	3	6	3.1
S18 Q75	1	6	2.1

[1] Percent enrichment was measured by dividing the dividing the number of mutant colonies detected by MASC-PCR and confirmed via Sanger sequencing by the total number of colonies that were screened by MASC-PCR and multiplying the result by 100.

Supplementary Table 3. Labeling efficiencies of targeted positions quantified by interpolating the cumulative fluorescence intensity of each labeled protein band from an SDS-PAGE gel of labeled 30S and/or 50S subunits from a linear standard curve generated using Cy3- or Cy5-labeled protein standards of defined quantities and labeling efficiencies (**Methods**, **Supplementary Fig. 3**).

smFRET Signal	Labeling Position	Labeling Efficiency (%)[1]
IR1	L9 N11	29.2 ± 0.9
	S6 D41	96.3 ± 2.3
HS1	S7 G112	14.7 ± 0.6
	S11 A102	75.5 ± 3.1
HS2	S7 K131	22.3 ± 0.7
	S18 R8	70.8 ± 4.4
MT1	S18 R8	70.8 ± 4.4

[1]Averages and standard deviations were calculated from three independent experiments.

Supplementary Table 4. Types of labels and corresponding applications that can be enabled by using multiplexed genomic encoding of bioorthogonal ncAAs to site-specifically label MBCs.

Label Type	Applications
Optical labels	Fluorescence imaging (including smFRET, super-resolution imaging, *etc*.)
Affinity labels	Purification and isolation
Surface tethering labels	Fluorescence imaging (particularly smFRET), surface plasmon resonance, *etc*.
Microsphere labels	Atomic force microscopy, single-molecule force spectroscopy, *etc*.
Cross-linking reagents	Biochemical assays, proteomics, *etc.*
Spin labels	Nuclear magnetic resonance and electron spin resonance spectroscopies
Post-translation modifications	Biochemical assays

Supplementary Table 5. Sequences of mRNAs and DNA oligonucleotide used in smFRET experiments.

Name	Sequence
Bio-mRNA	5'-Biotin-CAACCUAAAACUUACACAAAUUAAAAAGGAAAUAGACAUGUUCAAAGUCGAAAAAUCUACUGCU-3'
NonBio-mRNA	5'-GGCAACCUAAAACUUACACAGGGCCCUAAGGAAAUAAAAAUGUUUAAACGUAAAUCUACUGCUGAACUCGCUGCACAAAUGGCUAAACUGAAUGGCAAUUAAGGAUC-3'
Cy5-DNA-Bio	5'-Cy5-TGTAAGTTTTAGGTTGCC-Biotin-3'