Supporting Information

Bacterial lyso-form lipoproteins are synthesized via an *intra*-molecular acyl chain migration

Krista M. Armbruster, Gloria Komazin, Timothy C. Meredith

Contents

TABLE S1. Primers used in this study.

Primer Name	Description	Primer Sequence
TM692	5'-*lnt*::Spt^r	GCCTCATTAATTGAACGCCAGCGCATTCGCCTGCTGCTGGCGATTCCC CTGCTCGCGCAG
TM693	*lnt*::Spt^r-3'	CACAGCAGCAAAACCAAACAATGCCGTCAGCACCCACAGCGGGCTTG AACGAATTGTTAG
KA1053	5'-pET22b(+) *Eflit*	CTTTAAGAAGGAGATATACATATGATGAGGTCGAAATTGCG
KA1054	pET22b(+) *Eflit*-3'	GTGGTGGTGGTGGTGGTGGCCGCCGCCTTTTTTTAAAGAACGTTTTCCG
TM1121	5'-*fadE*::Tet^r	GATTTTGAGTATTCTCGCTACGGTTGTCCTGCTCGGCGCGTTCAAATGT AGCACCTGAAG
TM1122	*fadE::*Tet^r-3'	GTCGCCAGCTCTTCCGGATCAAAGTCATCAACGTTAATACTGGAATCCG TTAGCGAGGTG
TM1123	5'-*fadE* upcheck	CCATATCATCACAAGTGGTCAGACC
TM1124	*fadE* downcheck-3'	CAACTTTCCGCACTTTCTCCGGCAAC
TM1130	5'-pKA522 PA3286	CTGTGAAGTGAAGCTGAGAGAAGAGGAACAACTATGCATAAAGCC
TM1131	pKA522 PA3286-3'	GGACTCTGGGGTTCGGCCACGATCAGTGTTTAC
KA1245	5'-*fadR::*Tmp^r	GTCATTAAGGCGCAAAGCCCGGCGGGTTTCGCGGAAGAGTACGGATA GACGGCATGCACG
KA1246	*fadR::*Tmp^r-3'	CCAAATCTCGCCACTCTCATGCCCATAGCGACGCACTGTGAATCCGTTG CTGCCAC
KA1247	5'-*fadR* upcheck	GGTATGATGAGTCCAAC
KA1248	*fadR* downcheck-3'	CATCGAGTTGCTGGAACG
TM1476	5'-*lolCDE*	GACCGACCAAAAGCTACAGCAACCAGACGG
TM1477	*lolCDE*-3'	TTACTGTCGGGAATTCTTGTTTTAATGTACTGCC
TM1614	5'-*fabH* upcheck	GGACACCTCTGACGAGTGGATTG
TM1615	*fabH* downcheck-3'	CACGATCGGTTGGATCGCAGGTG
TM1689	5'-*fabH::*Cm^{r}FRT	CGCCACATTGCCGCGCCAAACGAAACCGTTTCAACCATGGGCTATGA ATATCCTCCTTAG
TM1690	*fabH::*Cm^{r}FRT-3'	CCGCCCCAGATTTCACGTATTGATCGGCTACGCTTAATGCATGTAGG CTGGAGCTGCTTC
TM1722	5'-*fadD(*V451A)	GCTAATGCGCAGGAATCCTTCTTCATCC
TM1723	*fadD*(V451A)-3'	TTCCTGCGCATTAGCGATCGTAAAAAAGAC
TM1990	5'-pET22b(+) *Eflit*	AGAAGGAGATATACATATGATGAGGTCG
SS1314	pET22b(+) *Eflit*-3'	GGTGGTGGTGCTCGAGTTCACTTTTCTCTC
SS1562	5'-pET22b(+) *phoA*	GAAGGAGATATACATATGAAACAAAGCACTATTGC
SS1316	pET22b(+) *phoA*-3'	GGTGGTGGTGCTCGACGCGGTTTTATTTCA
SS1317	5'-pET22b(+) *lacZ*	CTTTAAGAAGGAGATATACAAACAGCTATGACCATG
SS1318	pET22b(+) *lacZ*-3'	GGTGGTGGTGCTCGAGTTCAAGCAAGCTTTT
SS1502	5'-(+9) *lacZ*	GCCGTCGTTTTACAACG
SS1503	5'-(+13) *phoA*	TTACTGTTTACCCCTGTG
SS1275	*Eflit*K212 fusion-3'	TTTTTTTAAAGAACGTTTTCCG
SS1276	5'-*Eflit*K212 *lacZ* fusion	CGTTCTTTAAAAAAAGCCGTCGTTTTACAAC
SS1504	5'-*Eflit*K212 *phoA*	CACAGGGGTAAACAGTAATTTTTTTAAAGAACG

	fusion	
SS1505	*Eflit*D52 *lacZ* fusion-3'	CGTTGTAAAACGACGGCATCAACCTCTACGTC
SS1506	*Eflit*D52 *phoA* fusion-3'	CACAGGGGTAAACAGTAAATCAACCTCTACGTC
SS1507	*Eflit*R129 *lacZ* fusion-3'	CGTTGTAAAACGACGGCCCGTATCAATGTCCA
SS1508	*Eflit*R129 *phoA* fusion-3'	CACAGGGGTAAACAGTAACCGTATCAATGTCCA
SS1509	*Eflit*D173 *lacZ* fusion-3'	CGTTGTAAAACGACGGCATCAGTCAGCGGATT
SS1510	*Eflit*D173 *phoA* fusion-3'	CACAGGGGTAAACAGTAAATCAGTCAGCGGATT
TM1987	5'-pET *Eflit* streptag	TGGAGTCATCCTCAATTTGAAAAATAATGAAGAGAGAAAAGTGAAC
TM1988	pET *Eflit* streptag-3'	TTGAGGATGACTCCACGCTGAACCGGATCCTTTTTTTAAAGAACG
TM1924	5'-*Eflit* Mutseq	CCGTGACCAACTTTAAGAAGGAGATATAC
TM1925	*Eflit* Mutseq-3'	CCAGTGCTCGAGTTCACTTTTCTCTC
TM2182	5'-*Eflit* H89	TTTTATGAAGTGAAAAAATTGTTTTTATTGTGCTATG
TM2183	*Eflit* H89-3'	TTTCACTTCATAAAAMNNAAAAGCCGCGCTATCAGAC
TM2184	5'-*Eflit* H157	GGATTATTTTTTAACAACGATGCTTGGTTATTTAATCC
TM2185	*Eflit* H157-3'	GTTAAAAAATAATCCMNNAAACGCAATAAAAAATTGATCAAAGCC
TM2149	5'-pET22 to PLI50 Ppen *Xba*I	TATTGTCGACTCTAGACTCTAGAAATAATTTTGTTTAAC
TM2150	pET22 to PLI50 Ppen *Asc*I-3'	TTTAGAATAGGCGCGCCCTCGAGTTCACTTTTCTCTC
TM1939	5'-*Eflit* L6C	TCGAAATGTCGTTTAAGGGAAACAC
TM1940	*Eflit* L6C-3'	TAAACGACATTTCGACCTCATCATA
TM1941	5'-*Eflit* V54C	GATCGTTGTACCTTGATTAAAAATT
TM1942	*Eflit* V54C-3'	CAAGGTACAACGATCAACCTCTACG
TM1958	5'-*Eflit* A165C	AACGATTGTTGGTTATTTAATCCGC
TM1959	*Eflit* A165C-3'	TAACCAACAATCGTTGTTAAAAAATAATCC
TM1945	5'-*Eflit* L171C	AATCCGTGTACTGATCCGATTATCA
TM1946	*Eflit* L171C-3'	ATCAGTACACGGATTAAATAACCAA
TM1947	5'-*Eflit* S209C	AAACGTTGTTTAAAAAAAGGCGGCG
TM1948	*Eflit* S209C-3'	TTTTAAACAACGTTTTCCGATTAAA

TABLE S2. Topology predictions for *E. faecalis* Lit.

	CCTOP (74)	DAS (75)	OCTOPUS (76)	Phobius (77)	Predict Protein (78)	SPLIT (79)	TMHMM (80)	TMpred (81)	TOP CONS (82)
Cytoplasmic	1-11	1-11	1-13	1-11	1-10	1-9	1-11	1-10	1-12
TM1	12-31	12-30	14-34	12-32	11-32	10-33	12-34	11-32	13-33
Periplasmic	32-95	31-95	35-96	33-95	33-95	34-94	35-95	33-95	34-94
TM2	96-115	96-112	97-117	96-115	96-117	95-116	96-118	96-113	95-115
Cytoplasmic	116-136	113-134	118-129	116-135	118-126	117-132	119-138	114-132	116-129
TM3	137-156	135-155	130-150	136-162	127-149	133-160	139-161	133-156	130-150
Periplasmic	157-186	156-183	151-182	163-181	150-182	161-179	162-182	157-183	151-181
TM4	187-206	184-205	183-203	182-206	183-206	180-204	183-205	185-206	182-202
Cytoplasmic	207-212	206-212	204-212	207-212	207-212	205-212	206-212	207-212	203-212

Output from several protein topology predictor programs reveals uncertainty in the boundaries of the periplasmic region between transmembrane (TM) regions 3 and 4.

74. Dobson, L., Remenyi, I., and Tusnady, G. E. (2015) CCTOP: a Consensus Constrained TOPology prediction web server. *Nucleic Acids Res.* **43**, W408-412
75. Cserzo, M., Wallin, E., Simon, I., von Heijne, G., and Elofsson, A. (1997) Prediction of transmembrane alpha-helices in prokaryotic membrane proteins: the dense alignment surface method. *Protein Eng.* **10**, 673-676
76. Viklund, H., and Elofsson, A. (2008) OCTOPUS: improving topology prediction by two-track ANN-based preference scores and an extended topological grammar. *Bioinformatics* **24**, 1662-1668
77. Kall, L., Krogh, A., and Sonnhammer, E. L. (2004) A combined transmembrane topology and signal peptide prediction method. *J. Mol. Biol.* **338**, 1027-1036
78. Yachdav, G., Kloppmann, E., Kajan, L., Hecht, M., Goldberg, T., Hamp, T., Honigschmid, P., Schafferhans, A., Roos, M., Bernhofer, M., Richter, L., Ashkenazy, H., Punta, M., Schlessinger, A., Bromberg, Y., Schneider, R., Vriend, G., Sander, C., Ben-Tal, N., and Rost, B. (2014) PredictProtein--an open resource for online prediction of protein structural and functional features. *Nucleic Acids Res.* **42**, W337-343
79. Juretic, D., Zoranic, L., and Zucic, D. (2002) Basic charge clusters and predictions of membrane protein topology. *J. Chem. Inf. Comput. Sci.* **42**, 620-632
80. Krogh, A., Larsson, B., von Heijne, G., and Sonnhammer, E. L. (2001) Predicting transmembrane protein topology with a hidden Markov model: application to complete genomes. *J. Mol. Biol.* **305**, 567-580
81. Hofmann, K., and Stoffel, W. (1993) TMbase-a database of membrane spanning proteins segments. *Biol. Chem. Hoppe-Seyler* **374**
82. Tsirigos, K. D., Peters, C., Shu, N., Kall, L., and Elofsson, A. (2015) The TOPCONS web server for consensus prediction of membrane protein topology and signal peptides. *Nucleic Acids Res.* **43**, W401-407

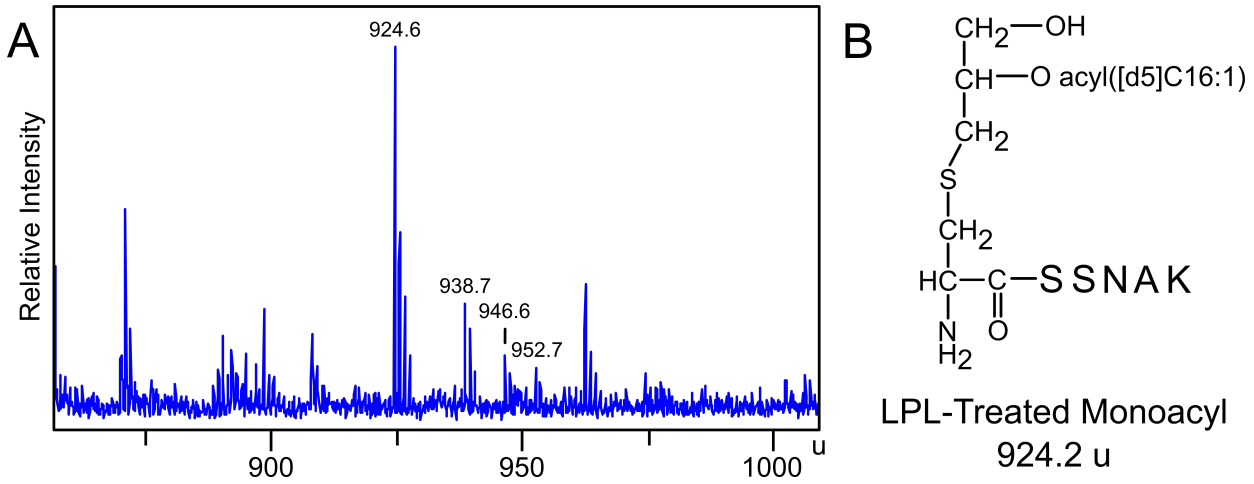

FIGURE S1. LPL-treatment of [d5]-DA-LppK58A strep removes the *sn*-1 acyl chain. *A*. [d5]-DA-LppK58A strep was treated with *Pseudomonas* spp. lipoprotein lipase (LPL) and analyzed by MALDI-TOF MS. The prominent peak at 924.6 u indicates a loss of the *sn*-1 linked saturated acyl chain ($C_{16:0}$). *B.* Regiospecific assignment of the *sn*-2 linked monounsaturated $C_{16:1}$ acyl chain.

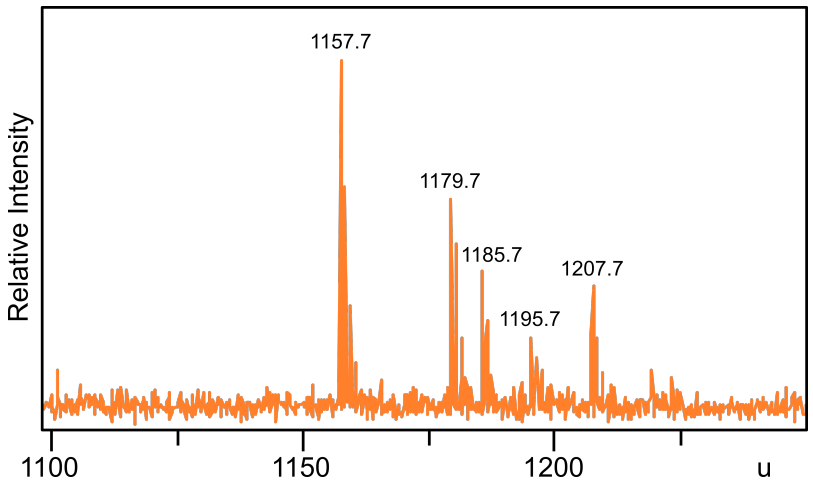

FIGURE S2. MALDI-TOF MS of lyso-LppK58A strep recovered from *in vitro* reaction with *E. faecalis* Lit. Parent spectra of the *N*-terminal lipopeptide of trypsinized LppK58A strep ($C_{32:1}$ total acyl chain length, 1157.7 u, protonated and 1179.7 u, sodiated) isolated from the *in vitro* reconstitution reactions with membranes expressing only *E. faecalis* Lit. The Δ28 u ions correspond to acyl chain length heterogeneity ($C_{34:1}$ total acyl chain length, 1185.7 u, protonated and 1207.7 u, sodiated).

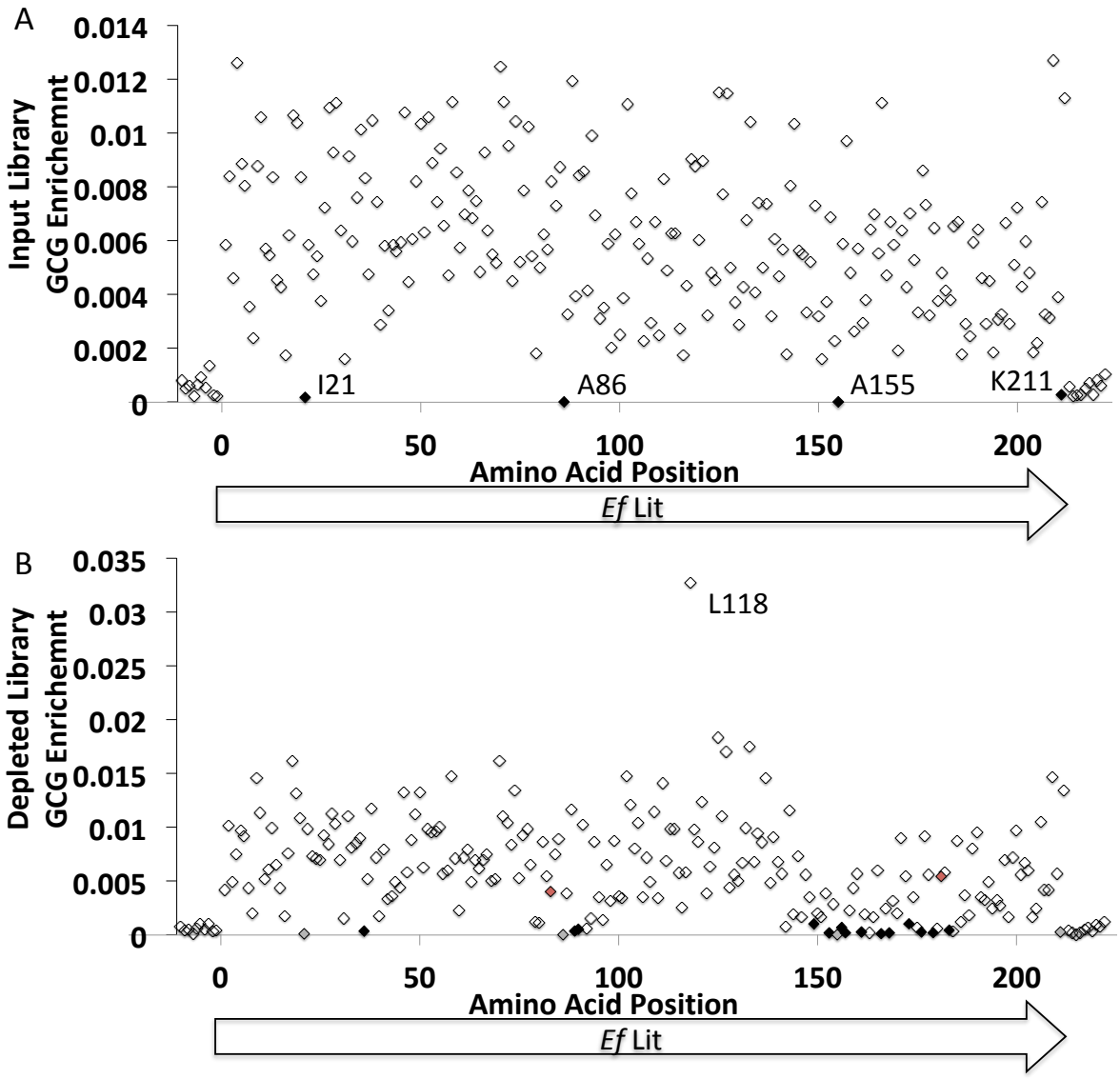

FIGURE S3. Mut-seq analysis of *E. faecalis* Lit by Ala scanning. *A.* The enrichment of each Ala encoding GCG triplet was calculated from the next generation read counts and plotted against the corresponding position in the *lit* gene. In-frame triplets upstream (10 amino acids) and downstream (8 amino acids) of the coding region which were not targeted for mutagenesis are shown for comparison. Residues I21 and K211 were not sufficiently represented in the input library to be considered in the analysis of important residues, while A86 and A155 are encoded by native GCG alanine triplets and mutagenesis results in synonymous codon replacement (*black* shaded). *B.* The enrichment of the GCG triplets from the same input library passaged in the absence of arabinose inducer. The respective plasmid encoded Lit allele is the sole means for lipoprotein *N*-acylation and growth rescue under these culture conditions. Residues for which importance could not be determined are *gray* shaded (listed in panel *A* above), while residues further evaluated for function are indicated (important for growth rescue are *black* shaded while negative controls are *red* shaded and correspond to Fig. 6 from text).

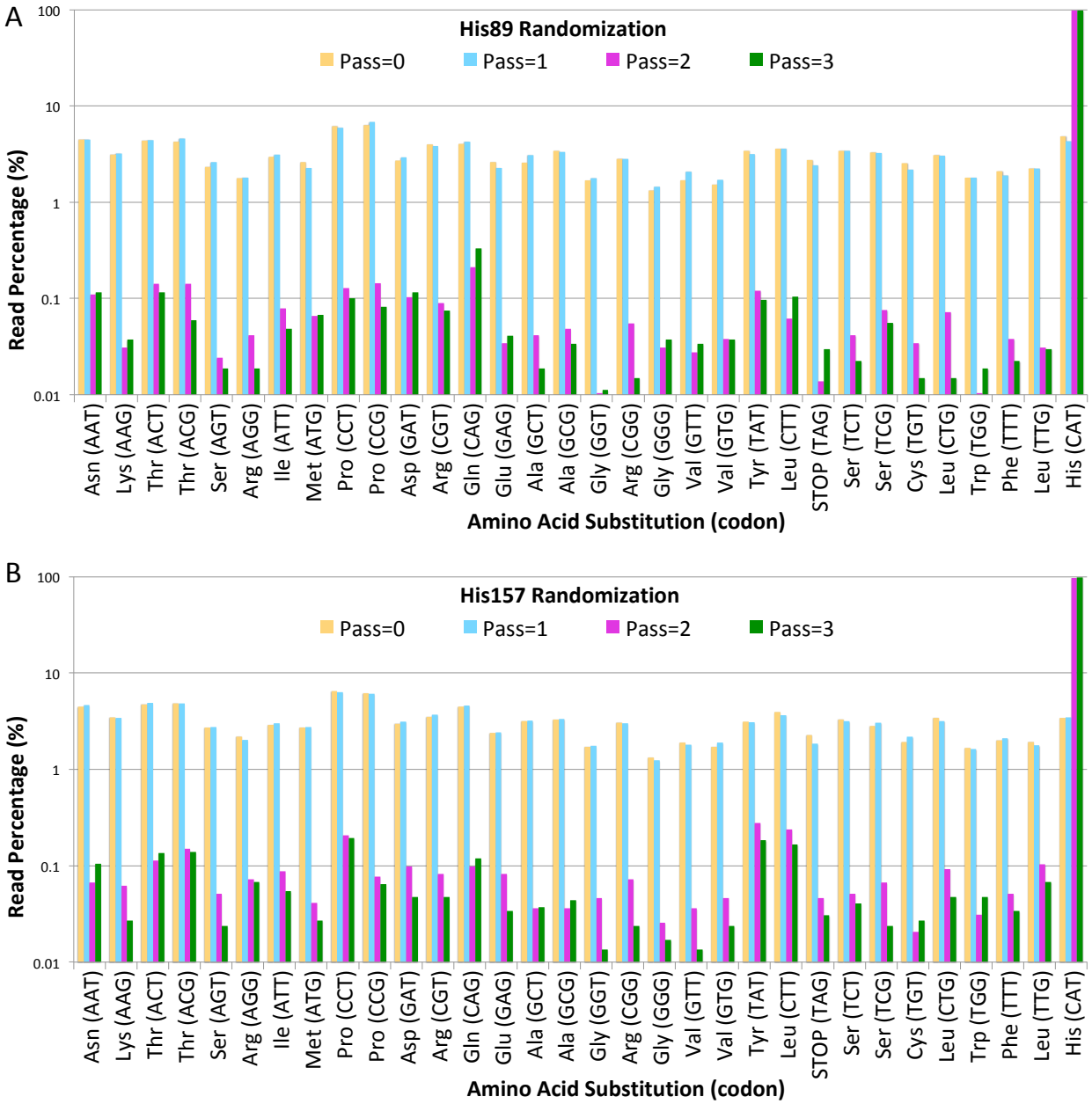

FIGURE S4. Randomization of His89 and His157 codon. A library of *E. faecalis lit* alleles was generated by introducing the NNK triplet into the His89 (*A*) or His 157 (*B*) position and transformed into *E. coli* Δ*lpp* P_{BAD}-*lnt* to evaluate rescue after three successive passages in the absence of arabinose inducer**.** Plasmid pools were isolated and sequenced to calculate read count percentages for each possible NNK triplet variant within the total population of the input (Pass=0) and ensuing sequential passages (Pass=1, 2, or 3).