Discovery and biological characterization of geranylated RNA in bacteria

Christoph E. Dumelin, Yiyun Chen, Aaron M. Leconte, Y. Grace Chen, David R. Liu

SUPPLEMENTARY INFORMATION

Supplementary Methods

General. Unless otherwise noted, all starting materials were obtained from commercial suppliers and were used without further purification.

Bacterial growth. Organisms were grown using the conditions shown in Supplementary Table 6. When required, sodium selenite was added to increase the total selenium concentration. For isotope labeling of RNA with ^{13}C and ^{15}N, *E. coli* were grown in 2 L of minimal media (6 g/L Na_2HPO_4, 3 g/L KH_2PO_4, 1 g/L NH_4Cl, 0.5 g/L NaCl, 1 g/L glucose, 2 mM $MgSO_4$, 0.1 mM $CaCl_2$), with either ^{13}C-labeled glucose or ^{15}N-labeled NH_4Cl. All organisms were grown to OD_{600} = 0.6 to 0.8 and centrifuged for 10 min at 10,000 x g. The resultant pellet was subjected to lysis.

Detergent cell lysis. The cell pellet as resuspended in 20 mL of 32 mM NaOAc, 2 mM EDTA, 1% SDS, pH 4.5, and mixed by vortexing vigorously. After incubation on ice for 15 min, the lysate was cleared by centrifugation (10 min at 17,000 x g) and the supernatant was carried forward.

Hot phenol cell lysis. The cell pellet was resuspended in 40 mL hot Buffer AE (50 mM NaOAc, 10 mM EDTA, pH 5.3, prewarmed to 65° C). 4 mL 10% SDS was immediately added and the solution was mixed by vortexing for 15 s. The resuspension was split into four 50 mL conical tubes; to each tube, 12 mL phenol (preheated to 65° C, pre-equilibrated with buffer AE) was added and the resulting mixture was mixed by vortexing for one minute. The mixture was incubated at 65° C for five minutes, with vortexing for 5 s once per minute, and immediately transferred to ice for 2 minutes. The cooled solution was then centrifuged for 10 min at 15,000 x g at 4° C. The aqueous layer was carried forward.

RNA purification. Cell lysates were extracted three times with acid-phenol chloroform (Ambion), and the aqueous layer was washed once with chloroform. An equal volume of isopropanol was added to the resulting aqueous extract followed by centrifugation for 20 min at 15,000 x g. The resulting pellet was dissolved in 1 mL of 50 mM NH_4OAc, pH 4.5, and subjected to size-exclusion chromatography using a NAP10 column (GE Healthcare). The macromolecular fraction was lyophilized and dissolved in 0.75 mL of 1x Turbo DNase buffer (Ambion), treated with 30 U TURBO DNase (Ambion) at 25 ºC for 30 min and then with 45 U proteinase K (New England Biolabs) at 25 °C for 30 min. The resulting solution was extracted twice with acid-phenol chloroform, washed with chloroform, and again subjected to size-exclusion chromatography using NAP10 columns equilibrated in 50 mM NH4OAc, pH 4.5. The macromolecular fraction was lyophilized and stored at -80 ºC.

RNA alkaline hydrolysis. To hydrolyze ges2U-containing RNA to mononucleosides, RNA was digested as above and lyophilized. Samples were redissolved in 200 µL 100 mM $(NH_4)_2CO_3$, pH 9 and incubated for 15 minutes at 94 °C and 3 hours at 37 °C with 100 U alkaline phosphatase, followed by a second purification by size exclusion chromatography and lyophilization. Alternatively, hydrolysis to mononucleosides was performed by digestion with 100 U alkaline phosphatase (Lee Biosolutions).

Nucleotide-specific nuclease digestion. To determine the nucleotide position of geranylation, isolated RNA or RNA from pull-down experiments (see below) was digested either with 100 U of RNase T1 (Roche) or 200 µg RNase A (Qiagen) for 3 hours at 37 °C in 500 µL of 50 mM NH_4OAc, pH 6. RNase T1 digestion reactions were lyophilized without further purification while RNase A digestion reactions were purified by size-exclusion chromatography and lyophilized as described above.

LC/MS analysis. For the comparison of abundance of different nucleosides in various *E. coli* strains, RNA was digested with nuclease P1 and alkaline phosphatase (Lee Biosciences) to mononucleosides. Samples were dissolved in 50 µL of 0.1% NH_3HCOOH and LC was performed using a linear gradient from 0.1% aqueous formic acid (A2) to acetonitrile (B2) on an Aquity UPLC BEH C18 column (1.7 µm, 2.1 mm x 100 mm, Waters) at a constant flow rate of 0.3 mL/min. The mobile phase composition was as follows: 100% A2 for 5 min; linear increase over 17 min to 100% B2; maintain at 100% B2 for 3 min; return to 100 % A2 over 5 min. Electrospray ionization conditions were as described above with the detector was operating in

positive mode. LC/MS/MS experiments for confirmation of the individual mononucleosides were performed at collision energies of 10, 20, and 30 eV.

For the comparison of chemically synthesized and biologically generated ges2U samples, we used a linear LC gradient from 0.1% aqueous formic acid (A2) to acetonitrile (B2) on an Aquity UPLC BEH C18 column (1.7 µm, 2.1 mm x 100 mm, Waters) at a constant flow rate of 0.3 mL/min. The mobile phase composition was as follows: 100% A2 for 1 min; linear increase over 21 min to 100% B2; maintain at 100% B2 for 8 min; return to 100 % A2 over 5 min. Electrospray ionization was used with a capillary voltage of 3 kV, a sampling cone voltage of 20 V, collision energies of 2, 10, 15, 20, and 25 eV, a LM resolution of 13, and the detector was operating in positive ion mode.

For analysis of oligonucleotides resulting from digestion of RNA with RNase T1, samples were dissolved in 50 µL of deinionized water and run using a linear gradient from 6 mM triethylammonium bicarbonate (A3) to methanol (B3) on an Aquity UPLC BEH C18 column (1.7 µm, 2.1 mm x 100 mm, Waters) at a constant flow rate of 0.3 mL/min. The mobile phase composition was a linear gradient from 100% A3 to 100% B3 over 60 min. Electrospray ionization was used with a capillary voltage of 3 kV, a sampling cone voltage of 40 V, collision energies of 2, 10, 20, and 30 eV and the detector was operating in negative ion mode.

Data analysis of small molecule-RNA conjugate screening results. Replicates of three nuclease P1-digested RNA samples and three control samples were used to identify candidate hits. Ions with average intensities below 50 counts were not considered for further analysis. The XCMS program[1] was used to quantify ion abundance as it stepped through each ion chromatogram; the step size was set to 0.050 Da. Integrated ion abundances were averaged among replicates, and the enrichment values reported represent the ratios of average ion abundances in the nuclease-treated samples to average ion abundances in the heat-inactivated nuclease (control) samples.

X-Ray crystallography of geranyl-2-thio-uracil. Geranyl-2-thiouridine (ges2U) was hydrolyzed with 1 equivalent of trifluoroacetic acid in water to obtain the geranyl-2-thiouracil. The product was crystallized in a mixture of methanol and ethyl acetate. A single crystal was mounted on a diffractometer and data was collected at 100 K. The intensities of the reflections were collected using a Bruker APEX II DUO CCD diffractometer ($Cu_{K\alpha}$ radiation, λ=1.54178 Å) equipped with an Oxford Cryosystems nitrogen flow apparatus. The collection method involved

1.0° scans in ω at 30°, 55°, 80° and 105° in 2θ. Data integration to 0.84 Å resolution was carried out using SAINT V7.46 A (Bruker diffractometer, 2009) with reflection spot size optimization. Absorption corrections were made with the SADABS software package (Bruker diffractometer, 2009). The structure was solved by the direct methods procedure and refined by least-squares methods again F^2 using SHELXS-97 and SHELXL-97 (Sheldrick, 2008). Non-hydrogen atoms were refined anisotropically, and hydrogen atoms were allowed to ride on the respective atoms. Crystal data as well as details of data collection and refinement are summarized in Supplementary Table 4, Geometric parameters and hydrogen-bond parameters are shown in Supplementary Table 5 . The ORTEP plot was produced with SHELXL-97, and the other drawings were produced with Accelrys DS Visualizer 2.0 (Accelrys, 2007). Complete list of software used: *APEX2* v2009.3.0 (Bruker-AXS, 2009), *SAINT* 7.46A (Bruker-AXS, 2009), *SHELXS97* (Sheldrick, 2008), *SHELXL97* (Sheldrick, 2008), Bruker *SHELXTL*. Crystallographic data have been deposited at the Cambridge Crystallographic Data Centre and copies can be obtained on request, free of charge, by quoting the publication citation and the deposition number CCDC 892714.

Quantitation of ges2U levels per cell. To establish a standard curve relating mass spectrometry ion counts with ges2U levels, 0.78, 1.56, 3.13, 6.25, 12.5, and 25 pmols of synthetic ges2U were analyzed by LC/MS. The ges2U signal from 50 μg of total *E. coli* Δ*mnmE* RNA was compared to the standard curves to determine the number of pmols of cellular ges2U per μg RNA. Based on a yield of 2.5 to 3 mg of total RNA per L and 4.6×10^{11} *E. coli* cells per litre of culture at OD_{600} = 0.6 (determined by plating serial dilutions of such a culture), the number of ges2U molecules per *E. coli* cell was estimated to be 398±81. The percentage of geranylated nucleotides was calculated by dividing this number by the sum of the number of $tRNA^{Glu}_{UUC}$, $tRNA^{Lys}_{UUU}$, and $tRNA^{Gln}_{UUG}$ per *E. coli* per cell previously determined using two different methods (13993 at 2.5 doublings per hour[2], 2.8%; 5910[3], 6.7%).

Size fractionation of total RNA. Size fractionation was performed using high-pressure liquid chromatography (Agilent 1100) with an anion-exchange Dionex DNAPac PA200 4x250mm analytical column. Mobile phase A was 300 mM sodium perchlorate in water and mobile phase B was 20% acetonitrile in water. The flow rate was a constant 1.2 mL/min, temperature was 80 °C and the mobile phase composition was as follows: linear increase from 30% A to 70% A over 5 min; linear increase over 35 min to 100% A; maintain at 100% A for 5 min; return to 70% A

over 2 min; return to 30% A over 3 min. Detection was carried out at 260 nm and 230 nm. 50 μg of total RNA from *E. coli* were injected for each size fractionation run. Fractions were collected every 5 min and the corresponding fractions from four to ten runs were combined. The first two fractions (0-5 min and 5-10 min elution) were concentrated using Millipore's Centriprep Centrifugal Filter Unit with Ultracel-3 membrane (YM-3) and the rest of the fractions were concentrated using the Centriprep Centrifugal Filter Unit with Ultracel-10 membrane (YM-10) by centrifugation (3,000 x g) until the total volume was less than 1.5 mL. The retentate was lyophilized and precipitated with ethanol prior to nuclease P1 digestion. Digestion by nuclease P1, purification, and mass spectrometric analysis was performed as described above.

Pull-down of individual tRNAs. Individual tRNAs were isolated using a slightly modified version of a previously described protocol.[4] In short, 200 μL of streptavidin sepharose slurry (GE Healthcare) was washed twice with phosphate buffered saline (PBS), individually incubated for 10 minutes at 25 ºC with 15 nmol of 3' biotinylated pull-down oligonucleotides specific for the tRNA sequences of interest, and subsequently washed twice with PBS. The sequences of the pull-down oligonucleotides were as follows: CCT GTT ACC GCC GTG AAA GGG CGG TGT CC-biotin (Glu), AGG GAA TGC CGG TAT CAA AAA CCG GTG CCT-biotin (Gln), AGT GAC ATA CGG ATT AAC AGT CCG CCG TTC-biotin (Asn), TGC GAC CAA TTG ATT AAA AGT CAA CTG CTC-biotin (Lys). DNA oligonucleotides were synthesized on a PerSeptive Biosystems Expedite 8909 DNA synthesizer using commercially available standard base monomers and 3'-BiotinTEG-CPG from Glen Research. 7.5 mg of *E. coli* tRNA (Roche) were dissolved in 1 mL of 10 mM Tris-HCl, 0.9 M tetramethylammonium chloride, 0.1 mM EDTA, pH 7.8. To this solution sepharose beads bound to the pull-down oligonucleotides were added and incubated for 10 minutes at 65 ºC. The beads were cooled to 25 ºC over 10 min and transferred to Ultrafree-MC spin columns PVDF 5 μM (Millipore) to remove unbound RNA. The resin was washed six times with 10 mM Tris-HCl, pH 7.8. Bound tRNAs were eluted twice by successive additions of 200 μL deionized water and incubation for 10 minutes at 65 ºC. The eluted RNA was digested either with nuclease P1 or RNase T1 as described above.

Expression and partial purification of SelU. *E.coli* selU containing a C-terminal hexahistidine tag was cloned into a pET vector. The protein was expressed in BL21(DE3) cells in 1 L of LB / 50 μg/mL carbenicillin by addition of 1 mM isopropyl-β-D-thiogalactopyranoside at OD_{600} = 0.6.

After 3 hours at 24 °C cells were harvested by centrifugation. Cell pellets were resuspended in 30 mL of buffer A (50 mM sodium phosphate, pH 8.0, 300 mM NaCl, 10 mM 2-mercaptoethanol, 5% glycerol) containing 20 units TURBO DNase and EDTA-free Protease Inhibitor Cocktail (Roche). Cells were disrupted by sonication. After centrifugation for 20 min at 17,000 *G*, the clarified extract was loaded onto 1 mL of Ni-NTA resin (Qiagen) pre-equilibrated in buffer A. After washing with 10 mL of buffer A and 10 mL of buffer A containing 10 mM imidazole, the bound protein was eluted with 6 mL buffer A containing 200 mM imidazole. The eluted protein was dialyzed twice against 1 L of 50 mM Tricine-NaOH, pH 8.0, 1 mM dithiothreitol, 10% glycerol. Purification of the dialyzed protein was then repeated as described above except the protein was elution in a stepwise manner using 2 mL buffer A containing, successively, 20, 40, 60, 80, 100, or 200 mM imidazole. All fractions were analyzed by SDS-PAGE. The elution fractions with 100 mM imidazole contained the most pure protein and were dialyzed twice against 1 L of 50 mM Tricine-NaOH, pH 8.0, 1 mM dithiothreitol, 10% glycerol. After dialysis the protein was either used immediately or stored at -20 °C. Characterization of the protein by UV spectroscopy showed similar absorption values at 260 and 280 nm, suggesting that a significant portion of the protein may be associated with oligonucleotides.

***In vitro* geranylation with SelU.** 140 nM partially purified selU-His_6 was incubated with 500 μg of *E. coli* tRNA (Roche) in the presence or absence of 135 μM ammonium geranyl pyrophosphate (Sigma) in 100 μL of 10 mM Tricine-NaOH, pH 8.0, 0.2 mM dithiothreitol, 2% glycerol at 25 °C. After 24 h the RNA was digested with Nuclease P1 / alkaline phosphatase and analyzed by LC/MS as described above.

Complementation with selU. The USER cloning method[5] was used to assemble a plasmid containing *selU*, *bla*, and *repA* and the origin of replication of pSC101. The *selU* genes including 200 nucleotides upstream of the translation initiation ATG as well 140 nucleotides downstream of the stop codon were PCR amplified from *E. coli BW25113* and *S. typhimurium LT2*. The ampicillin resistance gene *bla* was taken from pUC19, and *repA* and the origin of replication originate from pSC101. The point mutations G67E and G67R were introduced in a separate USER cloning step.

Codon usage reporter assay. USER cloning was used to assemble a plasmid containing *luxAB* under control of the tet promotor, *aadA*, *tetR*, *rop* and the origin of replication from ColE1. The luciferase fusion *luxAB* was constructed as previously described.[6] In addition, six codons were inserted between the starting ATG and the second codon of *luxAB*. The resulting first eight codon sequences were as follows: Glu(GAA): ATG GAA GAA GAA GAA GAA GAA AAA, Glu(GAG): ATG GAG GAG GAG GAG GAG GAG AAA.

For luciferase expression from the Glu(GAA) and Glu(GAG) reporter constructs in *E. coli* and *E. coli* Δ*selU*, 1 mL LB Lennox cultures containing 40 μg/mL spectinomycin, 200 ng/mL anhydrotetracyclin were inoculated with 1 μL saturated overnight starter cultures grown in the absence of anhydrotetracyline. For luciferase expression from the Glu(GAA) and Glu(GAG) reporter constructs in *E. coli* Δ*selU* + *selU* and *E. coli* Δ*selU* + *selU*(G67E), 1 mL LB Lennox cultures containing 40 μg/mL spectinomycin, 200 ng/mL anhydrotetracyclin, 50 μg/mL carbenicillin were inoculated with 1 μL saturated overnight starter cultures grown in the absence of anhydrotetracyline. Three replicates of each culture were incubated at 37 °C for 5 hours and an approximate optical density of OD_{600} = 0.5 was reached. 150 μL of each culture was transferred into clear-bottomed plates and assayed for luciferase activity by addition of 1 μL of 10% decanal and detection of luminescence. The individual counts were normalized by the respective optical densities. The three replicates were averaged and normalized to the signal obtained for Glu(GAA) in *E. coli*.

Transcript level measurement in the codon usage assay. RNA was isolated from 1 mL of *E. coli* culture grown to OD = 0.4 as described for the luciferase assay using RNeasy kit (Qiagen). Reverse transcription and quantitative PCR were carried out using the following primer pairs for the transcript of the bacterial luciferase gene fusion luxAB and the α-subunit of bacterial RNA polymerase rpoA as a reference.[7] luxAB forward: GGT AAC CCT TAT GTC GCT GC; luxAB backward: AAC TTA ATA TCA TCG GTA GGC C; rpoA forward: AAG CTG GTC ATC GAA ATG GAA; rpoA backward: GCC GCA CGA CGA ATC G. Reverse transcription was performed as follows: 0.5 μL of RNA extracts (1/100 dilution in deinionized water) and 50 pmol of primers were added to 9.5 μL deinionized water and heated to 80 °C for 3 min. Subsequently, 100 U M-MuLV Reverse Transcriptase (NEB), 2 U RNase inhibitor (NEB), 5 nmol dNTPs, 100 nmol DTT, and 1.2 μL of 10x M-MuLV Reverse Transcriptase Reaction Buffer (NEB) were added and incubated at 42 °C for 60 min, 50 °C for 10 min, and 94 °C for 15 min. Quantitative PCR was performed by adding 1 μL of the reverse transcription reaction mixture to 25 μL 1x iQ™ SYBR®

Green Supermix (Biorad) containing 16 pmol of each primer pair. QPCR reactions were performed using a CFX-96 Real-Time PCR System with a C1000 Thermal Cycler (Bio-Rad). The 40 reaction cycles were as follows: 94 °C for 30 sec; 50 °C for 30 sec; 72 °C for 1 min.

Frameshift reporter assay. USER cloning was used to assemble a plasmid containing the GST-MBP fusion protein under control of the tet promotor, *aadA*, *tetR*, *rop* and the origin of replication from ColE1. GST was subcloned from the pGEX vector, and *malE* was subcloned from the pMal-c2x vector. The following sequences were inserted for frameshift detection between GST and MBP: +1 frameshift: CCA AGC TTA **GCC AAGC** ACG ATC CCC TTT AAT; -1 frameshift: CAG GAT ACG AT**A AAA AA G**CC ATA GC TGG C. The frameshifting sequence is highlighted in bold. The plasmid backbone was amplified by PCR from the codon usage reporter construct.

GST-MBP fusion proteins were expressed for 12 hours at 37 °C in 100 mL LB Lennox cultures containing 40 μg/mL spectinomycin, 200 ng/mL anhydrotetracyclin when using wild-type *E. coli* and *E. coli* Δ*selU*, and in 100 mL LB Lennox cultures containing 40 μg/mL spectinomycin, 200 ng/mL anhydrotetracyclin, 50 μg/mL carbenicillin when using *E. coli* Δ*selU* + *selU* and *E. coli* Δ*selU* + *selU*(G67E). The cultures were centrifuged for 10 min at 10,000 x g and resuspended in 25 mL PBS, 50 U TURBO DNase, 250 μg RNase A, 1/2 tablet EDTA-free protease inhibitor cocktail (Roche). Cells were lysed by French press and cell debris was removed by centrifugation for 10 min at 17,000 x g. The cell lysate was incubated with 500 μL of glutathione sepharose (GE Healthcare) for 2 hours at 4 °C. The resin was transferred to Ultrafree-MC spin columns PVDF 5 μM and washed five times with 400 μL PBS. The purified protein was eluted twice using 500 μL 10 mM Tris-HCl, 10 mM glutathione, pH 7.8 and lyophilized. The samples were dissolved in 75 μL 1x protein gel loading buffer and were subsequently analyzed on a 12% SDS-PAGE gel. Proteins on the gel were transferred by electroblot onto a PVDF membrane (Millipore) pre-soaked in methanol. The membranes were blocked in PBS, 0.05% Tween 20, 3% BSA for 1 h and incubated with murine α-GST antibody (Abgent) in PBS, 3 % BSA for 1 h at room temperature. The membrane was washed three times with PBS, 0.05% Tween 20 and treated with the secondary antibody IRDye 800CW Goat Anti-Mouse IgG (1/10,000 dilution, Li-COR Biosciences) in blocking buffer (Li-COR Biosciences) for 1 h. The membrane was washed three times with PBS, 0.05% Tween 20 and visualized using an Odyssey infrared imaging system (Li-COR Biosciences). Images were analyzed using Odyssey imaging software version 2.0.

RTqPCR of tRNAs. Reverse transcription and quantitative PCR were carried out as described above using the following primers and 65 ng of RNA from the individual total RNA extractions. Glu forward: GTC CCC TTC GTC TAG AGG, Glu backward: CGA ACC CCT GTT ACC GC, Lys forward: GGG TCG TTA GCT CAG TTG G, Lys backward: CGA ACC TGC GAC CAA TTG, Gln forward: TGG GGT ATC GCC AAG CG, Gln backward: CGA ACC AGG GAA TGC CG, Asn forward: TCC TCT GTA GTT CAG TCG G, Asn backward: CGA ACC AGT GAC ATA CGG.

Aminoacylation level assays of individual tRNAs. Total RNA was isolated from *E. coli* as described above. De-aminoacylated RNA was prepared by incubation of RNA in 100 mM Tris-HCl, pH 9.6 for 30 minutes at 25 °C. Charged tRNA was separated from uncharged tRNA by gel electrophoresis of 10 μg RNA samples on a 8% acidic denaturing polyacrylamide gel (300 V, 10 V/cm)[8]. Electroblotting onto a BrightStar®-Plus Positively Charged Nylon Membrane (Life Technologies) was performed at 500 mA for 1.5 hours in 0.5x TBE using a semi-dry blotting device (TE70XP, Hoefer Inc). The RNA was crosslinked to the membrane in 175 mM EDC-HCl, 175 mM N-methyl-imidazole, pH 8 for 2 hours at 65 °C. Oligonucleotides complementary to nucleotides 25 to 50 in the individual tRNAs of interest were labeled using γ-^{32}P-ATP (100 nCi / pmol oligonucleotide) and polynucleotide kinase (NEB) followed by purification using Illustra MicroSpin G-25 spin columns (GE Healthcare). The membrane was incubated in 5 mL ULTRAhyb®-Oligo buffer (Life Technologies) at 42 °C for 30 min, followed by hybridization with 10 pmol of ^{32}P-labeled oligo at 42 °C for 12 h in 5 mL ULTRAhyb®-Oligo buffer, and three washes with 5 mL of NorthernMax® Low Stringency Wash Buffer (Life Technologies). The membrane was visualized by Phosphor-Imager analysis.

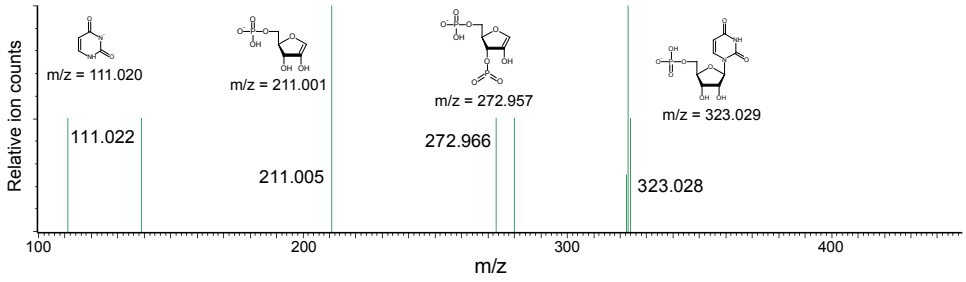
Supplementary References

1. Smith, C.A., Want, E.J., O'Maille, G., Abagyan, R. & Siuzdak, G. XCMS: processing mass spectrometry data for metabolite profiling using nonlinear peak alignment, matching, and identification. *Anal Chem* **78**, 779-87 (2006).
2. Dong, H., Nilsson, L. & Kurland, C.G. Co-variation of tRNA abundance and codon usage in Escherichia coli at different growth rates. *J Mol Biol* **260**, 649-63 (1996).
3. Jakubowski, H. & Goldman, E. Quantities of individual aminoacyl-tRNA families and their turnover in Escherichia coli. *J Bacteriol* **158**, 769-76 (1984).
4. Yokogawa, T., Kitamura, Y., Nakamura, D., Ohno, S. & Nishikawa, K. Optimization of the hybridization-based method for purification of thermostable tRNAs in the presence of tetraalkylammonium salts. *Nucleic Acids Res* **38**, e89 (2010).
5. Geu-Flores, F., Nour-Eldin, H.H., Nielsen, M.T. & Halkier, B.A. USER fusion: a rapid and efficient method for simultaneous fusion and cloning of multiple PCR products. *Nucleic Acids Res* **35**, e55 (2007).
6. Boylan, M.O. et al. Construction of a fused LuxAB gene by site-directed mutagenesis. *J Biolumin Chemilumin* **4**, 310-6 (1989).
7. Goh, S., Boberek, J.M., Nakashima, N., Stach, J. & Good, L. Concurrent growth rate and transcript analyses reveal essential gene stringency in Escherichia coli. *PLoS One* **4**, e6061 (2009).
8. Varshney, U., Lee, C.P. & RajBhandary, U.L. Direct analysis of aminoacylation levels of tRNAs in vivo. Application to studying recognition of Escherichia coli initiator tRNA mutants by glutaminyl-tRNA synthetase. *J Biol Chem* **266**, 24712-8 (1991).

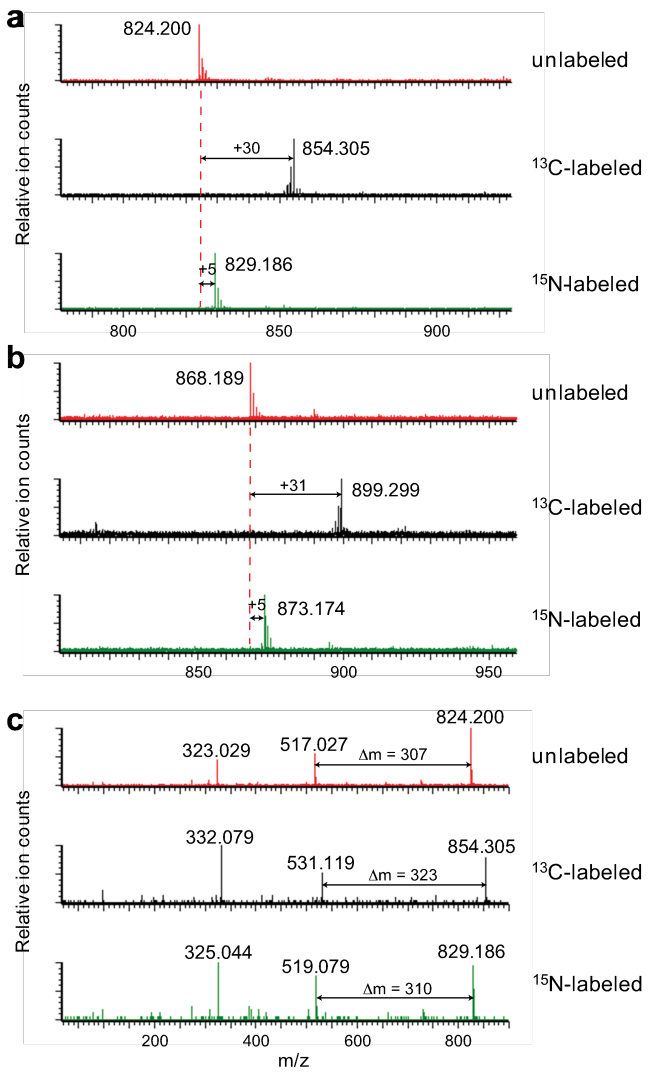
Supplementary Results

Retention time (min)	m/z	Nuclease / control enrichment	RNA modification
6.7	532.136	25	AMP-tryptophan
8.3	698.324	3	
9.6	796.069	4	
10.1	909.467	3	
10.8	868.189	12	
11.1	828.052	4	
11.7	824.200	32	
13.3	528.197	3	
13.3	501.186	3	

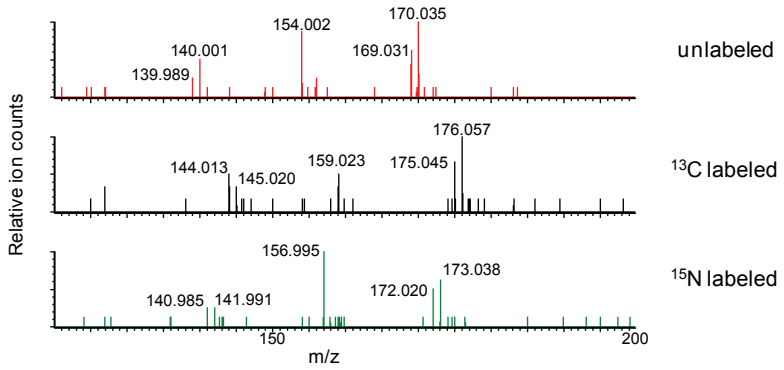
Supplementary Table 1. Complete list of unknown species discovered in E. coli ≥ 3-fold more abundant in the nuclease-treated sample than in the control sample. The most hydrophobic aminoacyl-adenosine monophosphate, AMP-tryptophan, is shown for comparison.



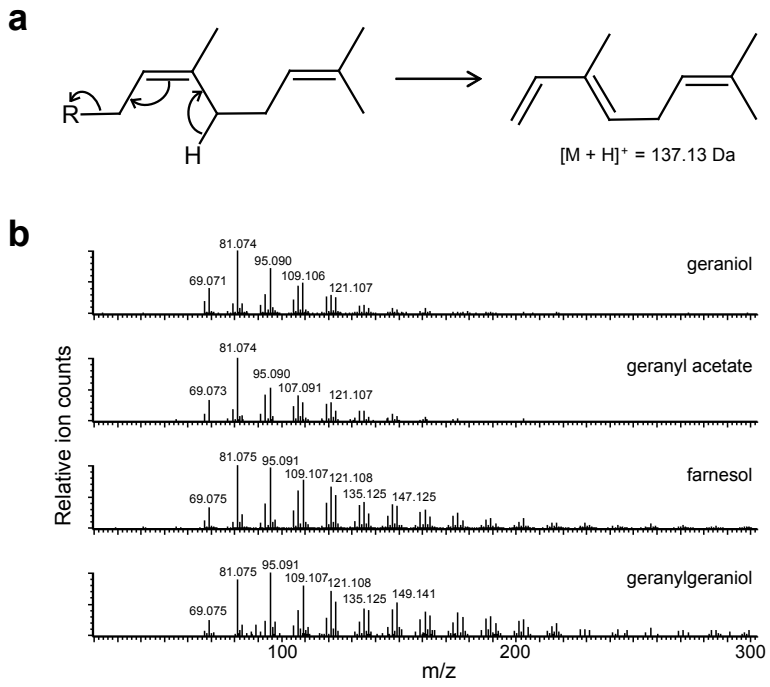
Supplementary Figure 1. Negative ion mode MS/MS of 517.029 reveals 5'-X-U-3' connectivity. The presence of a ribose diphosphate without a base and the absence of uridine diphosphate strongly suggest that the unknown modified nucleobase is on the 5' end of the dinucleotide.



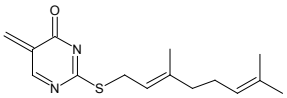
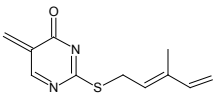
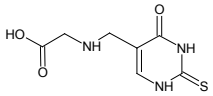
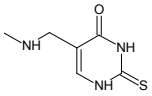
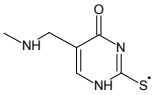
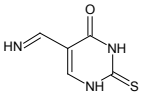
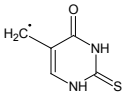
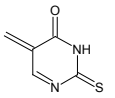
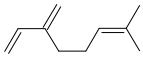
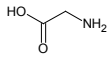
Supplementary Figure 2. MS and MS/MS of unknown dinucleotides from unlabeled as well as ^{13}C and ^{15}N labeled RNA. The spectra of the unlabeled RNA are red, the ^{13}C-labeled RNA spectra are black, and the ^{15}N-labeled RNA spectra are green. (a, b) Negative ion mode MS of 825 (a) and 869 (b) indicate the number of carbons and nitrogen atoms in the dinucleotides. (c) Negative ion mode MS/MS of 825 indicates the molecular weight of the unknown nucleobase in unlabeled, ^{13}C-, and ^{15}N-labeled RNA.



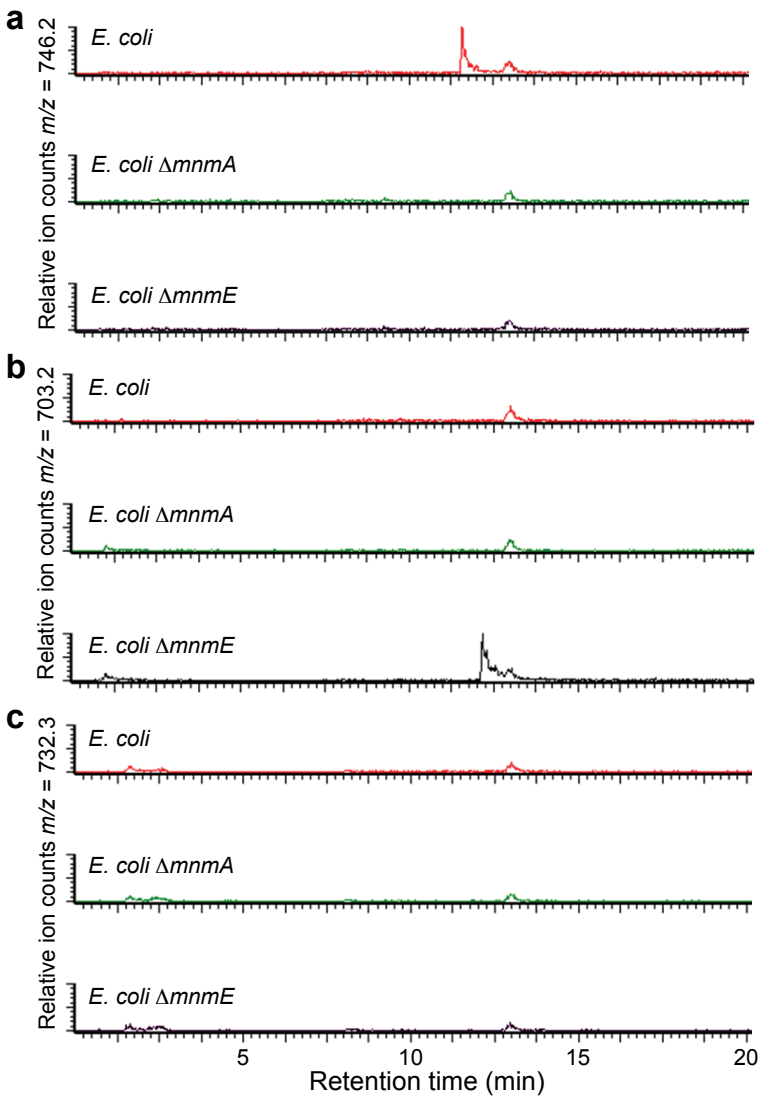
Supplementary Figure 3. Negative mode MS/MS of the unknown nucleobase 307.171 from unlabeled as well as ^{13}C-and ^{15}N-labeled RNA. The spectra are colored as in Supplementary Figure 2.



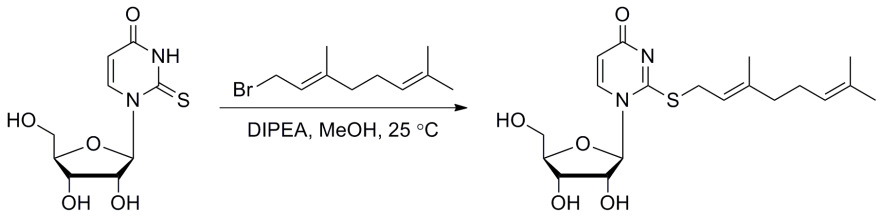
Supplementary Figure 4. (a) Putative mechanism of fragmentation of a geranyl containing molecule resulting in loss of a $C_{10}H_{16}$ moiety. (b) Positive ion mode MS/MS of geraniol, geranyl acetate, farnesol, and geranyl geraniol exhibits characteristic isoprene fragments with m/z = 81.07 as a dominant ion.

Parent	Observed $[M+H]^+$ *m/z*	Observed $[M-H]^-$ *m/z*	# of C atoms	# of N atoms	Proposed structure	MW
307/351	277.139		15	2		276.13
307/351	221.075		11	2		220.067
351	216.045		7	3		215.036
307	172.054	170.035	6	3		171.047
307		169.031	6	3		170.039
307		154.002	5	3		155.015
307		140.001	5	2		141.12
307/351	141.012	139.989	5	2		140.004
307/351	137.135		10	0		136.125
307/351	81.07		6	0		80.062
351	76.04		2	1		75.032

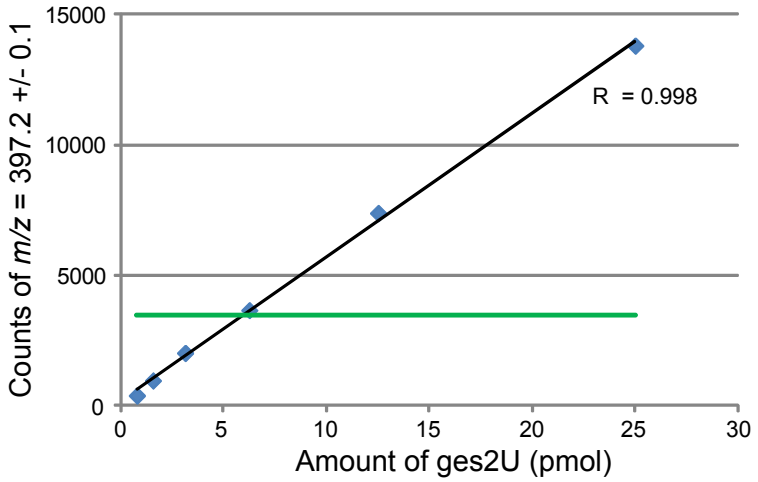
Supplementary Table 2. Summary of fragments observed in MS/MS of unlabeled and ^{13}C- or ^{15}N- labeled nucleobases including the proposed structures and their calculated molecular weights.



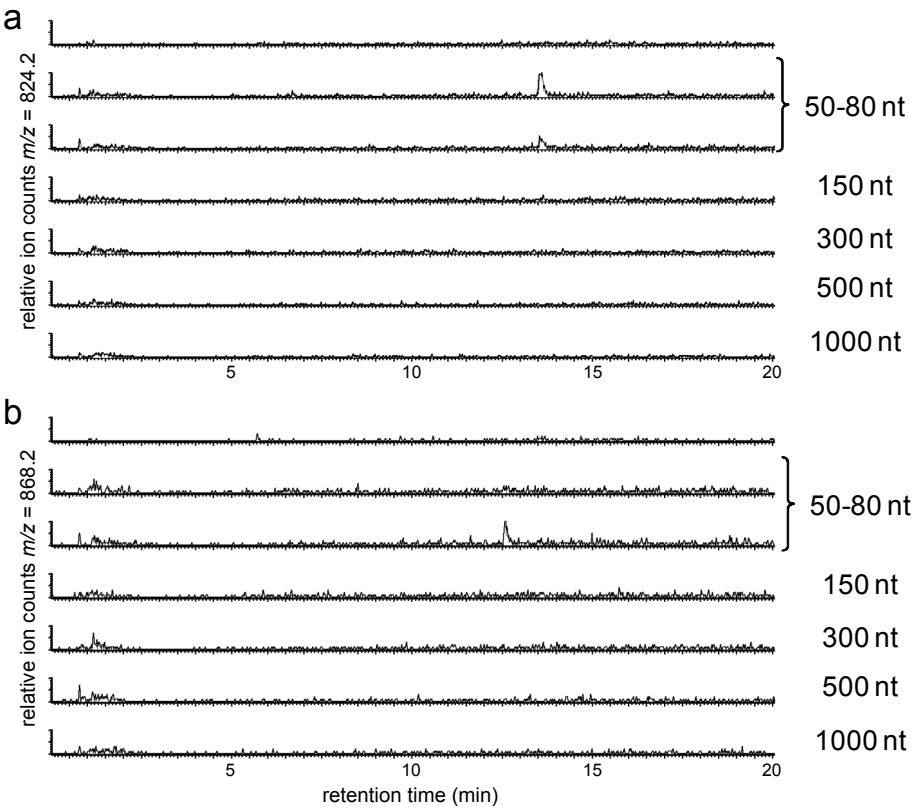
Supplementary Figure 5. Extracted ion chromatogram of positive mode MS analysis of RNA from *E. coli*, *E. coli* Δ*mnmA* and *E. coli* Δ*mnmE* strains. The individual panels depict the *m/z* for the dinucleotide containing ge-mnm5s2U (a), ge-s2U (b), and ge-mnm5U (c).



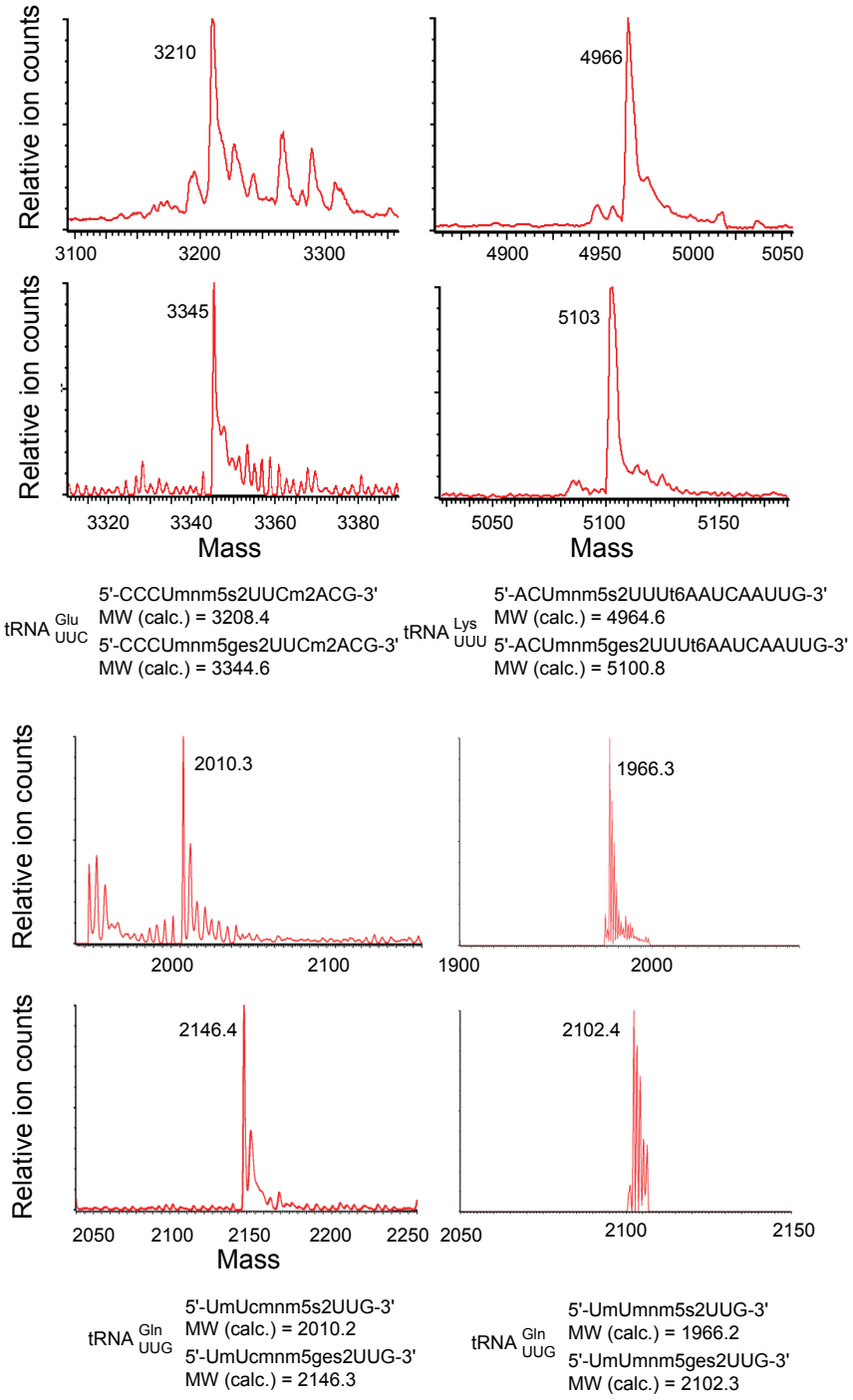
Supplementary Figure 6. Reaction scheme for the synthesis of ges2U.



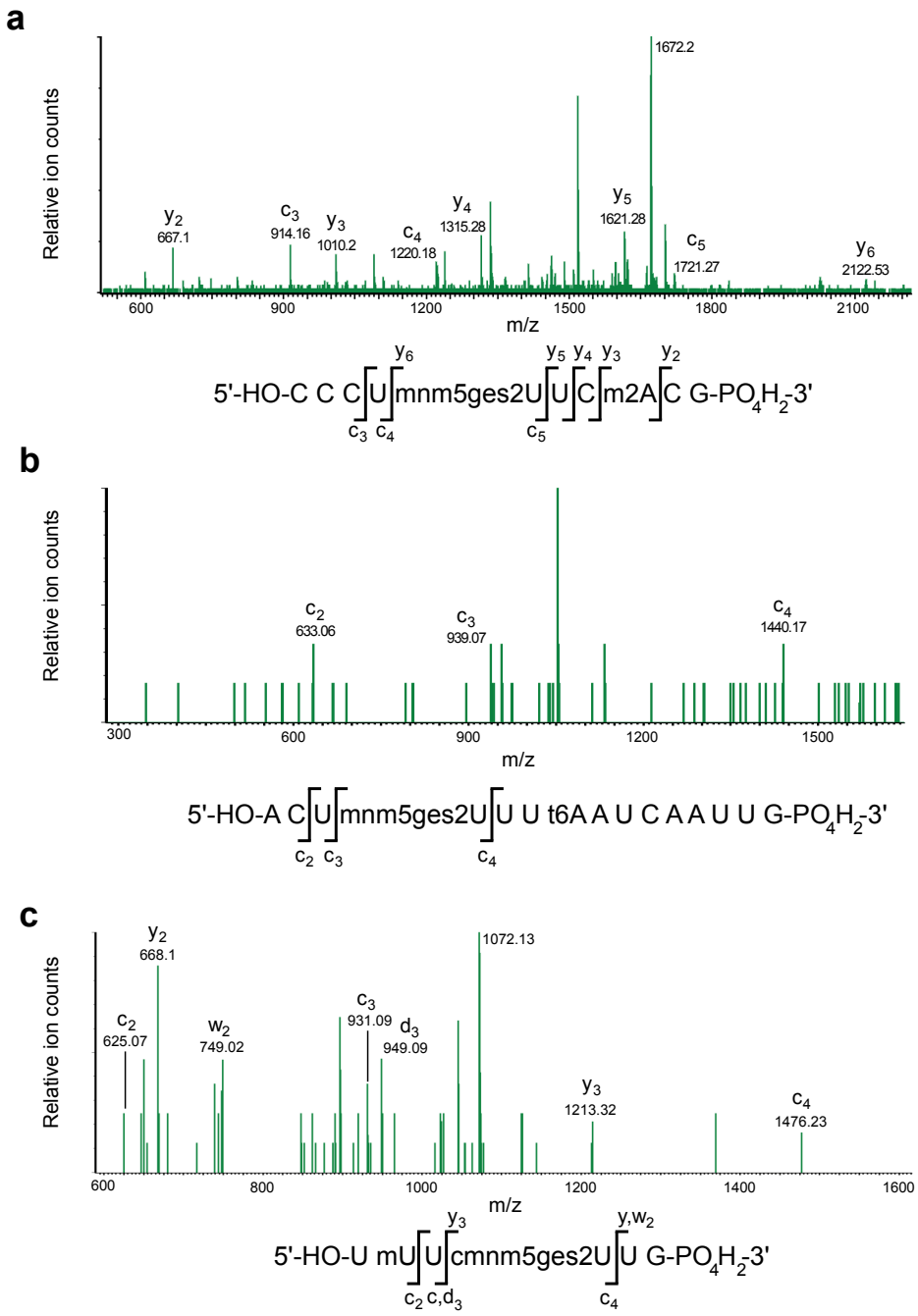
Supplementary Figure 7. Standard curve relating known ges2U quantities to ion counts. This curve was used to quantitate the amount of ges2U per *E. coli* Δ*mnmE* cell. The average ges2U signal arising from 50 μg of total RNA is indicated with a green line, which corresponds to 398±81 copies of ges2U per cell.



Supplementary Figure 8. Size fractionation of total RNA. The extracted ion chromatogram of the dinucleotide *m/z* = 824.2 (a) and *m/z* = 868.2 (b) in RNA fractionated by size prior to P1 digestion indicates that the dinucleotides are present in RNA 50-80 nucleotides in length.



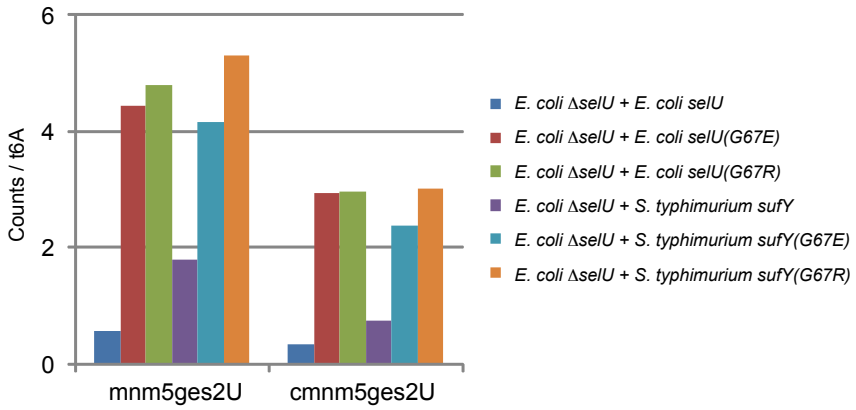
Supplementary Figure 9. RNase T1 digestion analysis of isolated tRNAs. Calculated and observed deconvoluted masses for the individual digestion fragments are in line with modification at U33 or U34. The predicted nucleotide sequences of the digested products with the site of geranylation at U34 are shown.



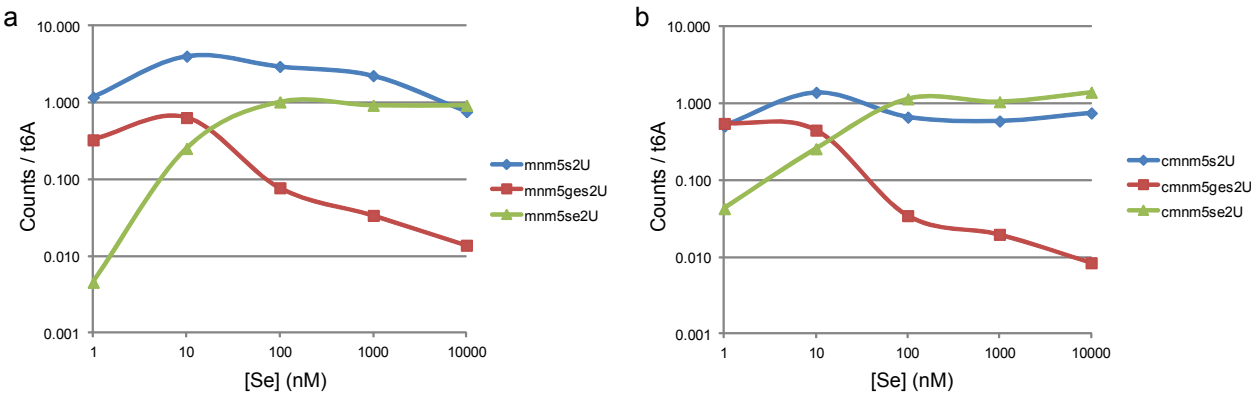
Supplementary Figure 10. MS/MS fragmentation of the T1 digestion products of (a) $tRNA^{Glu}_{UUC}$, (b) $tRNA^{Lys}_{UUU}$, and (c) $tRNA^{Gln}_{UUG}$ reveals U34 as the site of geranylation in all three tRNAs. (a) Sequential fragmentation of CCCUmnm5ges2UUCm2ACG (2^- m/z = 1672.2). (b) Sequential fragmentation of ACUmnm5ges2UUUt6AAUCAAUUG (3^- m/z = 1699.24). (c) Sequential fragmentation of UUmUUcmnm5s2UUG (2^- m/z = 1072.13).

Organism	(c)mnm5ges2U	sufY / selU
Escherichia coli	+	+
Enterobacter aerogenes	+	+
Pseudomonas aeruginosa	+	+
Salmonella typhimurium	+	+
Vibrio fischeri	-	-
Bacillus subtilis	-	-
Streptomyces venezuelae	-	-
Saccharomyces cerevisae	-	-
Schizosaccharomyces pombe	-	-
Bos taurus	-	-
Triticum (tRNA)	-	-

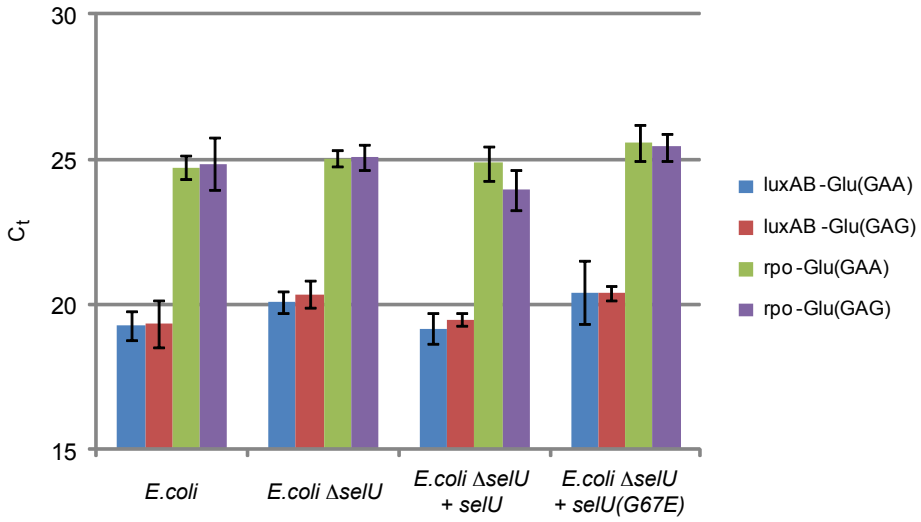
Supplementary Table 3. The organisms in which mnm5ges2U and cmnm5ges2U were observed when analyzing total RNA or tRNA (from *Triticum*) correlates with the presence of the tRNA 2-selenouridine synthase gene (sufY, selU). The samples of *S. cerevisae*, *S. pombe*, *B. taurus*, *Triticum* do not contain a geranylated form of the modified nucleotides present at U34 of $tRNA^{Lys}_{UUU}$, $tRNA^{Glu}_{UUC}$, and $tRNA^{Gln}_{UUG}$ in eukaryotes, which are 5-carboxymethylaminomethyl-2′-O-methyluridine, 5-carboxymethylaminomethyluridine, 5-methoxycarbonylmethyl-2-thiouridine, 2′-O-methyluridine, 5-taurinomethyl-2-thiouridine, 2-thiouridine, 5-methoxycarbonylmethyluridine, and 5-methoxycarbonylmethyl-2-thiouridine.



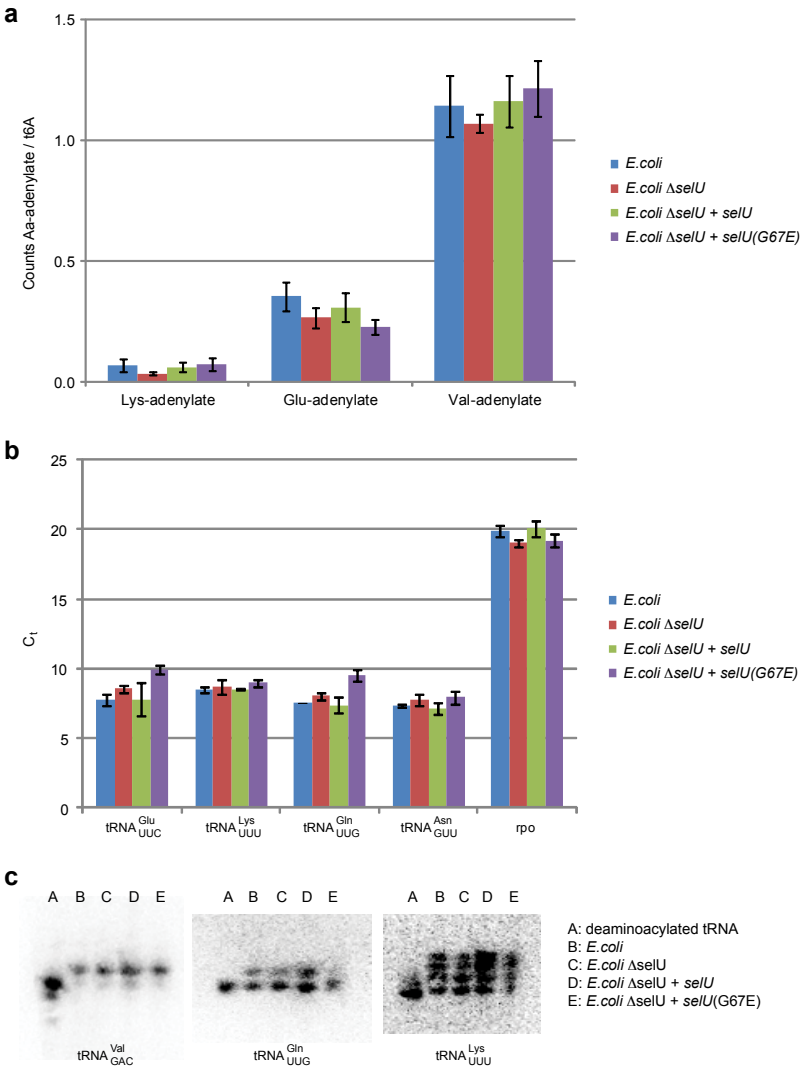
Supplementary Figure 11. Complementation of *E.coli* Δ*selU* with *selU*, *selU*(G67E), and *selU* (G67R) from *E.coli* and *S. typhimurium* shows that both mutants result in elevated levels of geranylated nucleotides. The Y-axis values reflect the relative ratios of ion counts of the individual nucleosides versus the ion counts of t6A.



Supplementary Figure 12. Levels of mnm5s2U, mnm5ges2U, and mnm5se2U (a) and cmnm5s2U, cmnm5ges2U, and cmnm5se2U (b) at varying concentrations of selenium in growth media shows a switch from geranylation to selenation between 10-100 nM selenium. The Y-axis values reflect the relative ratios of ion counts of the individual nucleosides versus the ion counts of t6A. Note that the levels of mnm5s2U and cmnm5s2U are underestimated due to saturation of the ion detector for these species under the conditions of this experiment.

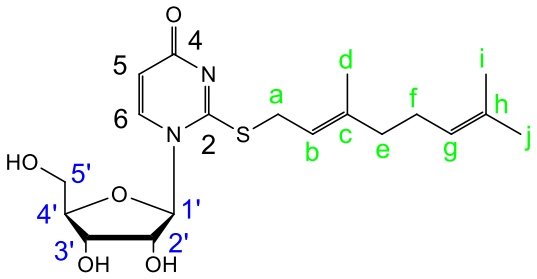


Supplementary Figure 13. Reverse transcription quantitative PCR analysis of luciferase (luxAB) and a housekeeping gene (rpo) shows no significant strain- or construct- dependent change in transcript levels given as cycle threshold values (C_t). The error bars represent the standard deviation of three independent reverse transcription quantitative PCR (RT-qPCR) experiments.

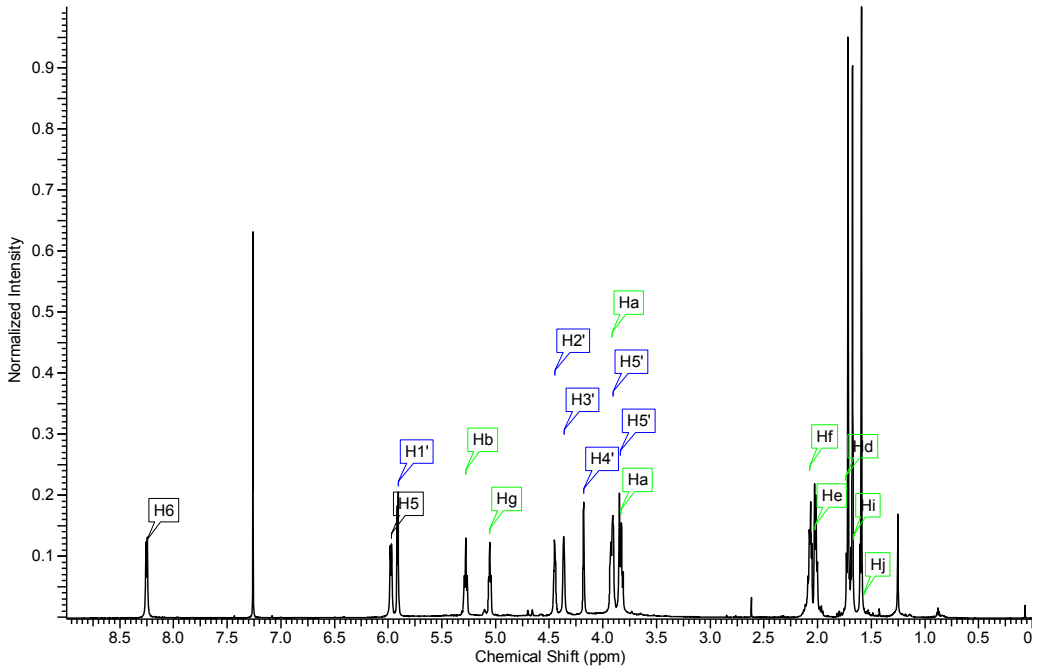


Supplementary Figure 14. (a) Determination of the relative amount of aminoacyl adenylate for lysine, glutamate, and valine in *E. coli* by LC/MS shows no substantial change due to geranylation levels. The error bars represent the standard deviation of three independent samples. (b) Measurement of RNA levels of $tRNA^{Glu}_{UUC}$, $tRNA^{Lys}_{UUU}$, $tRNA^{Gln}_{UUG}$, $tRNA^{Asn}_{GUU}$ and a control gene in *E.* coli by RT-qPCR reveal no dependence on geranylation levels. The error bars represent the standard deviation of three independent samples. (c) Analysis of tRNA aminoacylation levels by Northern blot. The ratio of charged:uncharged tRNA species is similar for Lys- and Val-tRNA under different geranylation levels (lanes B-E). In contrast, Gln-tRNA

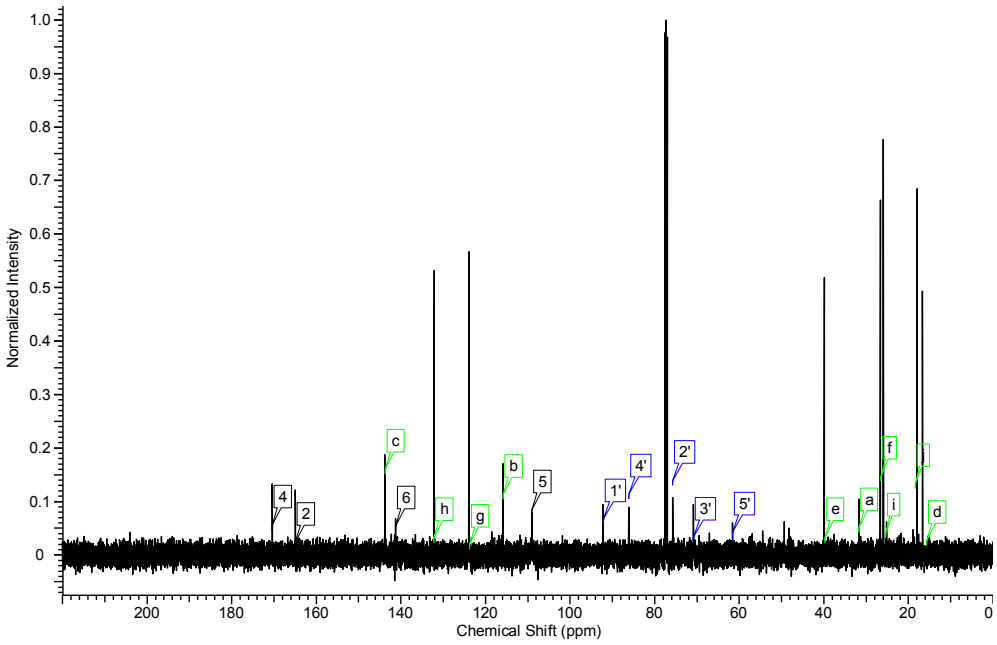
(center gel) shows substantially reduced aminoacylation under high geranylation conditions (lane E vs. lanes B-D).



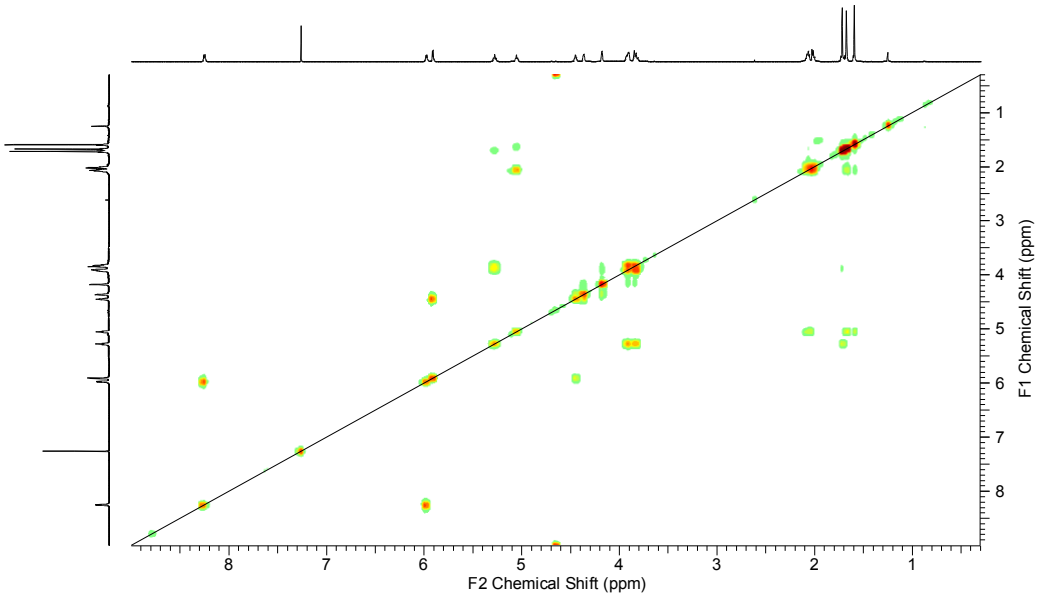
Supplementary Figure 15. Structure of ges2U with annotation of atoms referred to in the Supplementary Figures below.



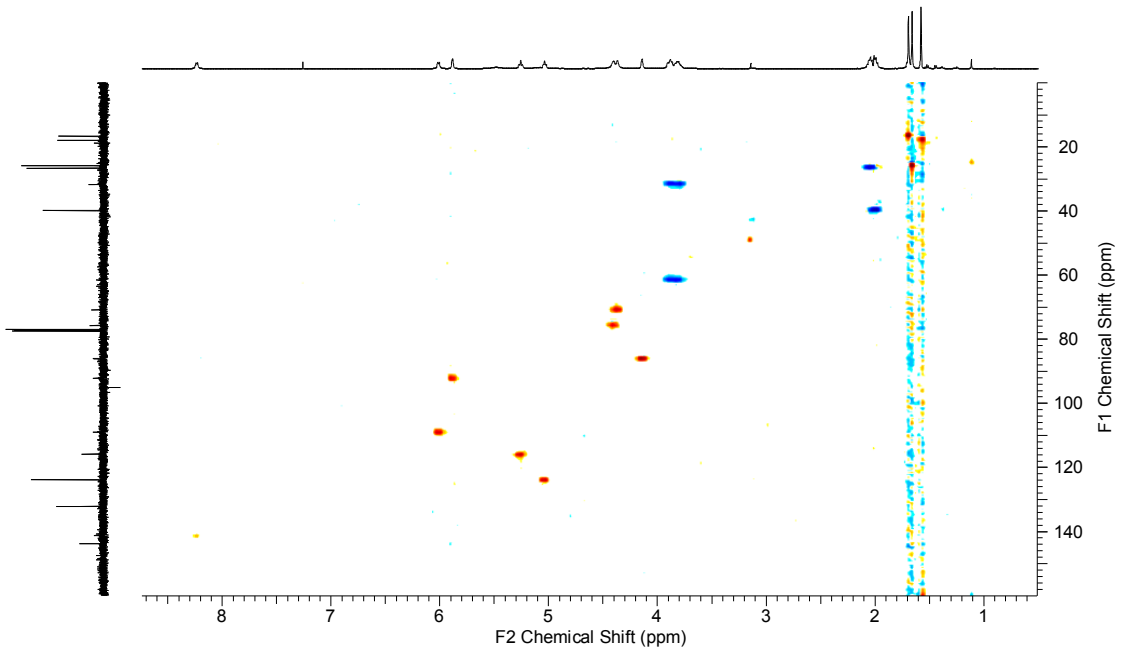
Supplementary Figure 16. ^{1}H-NMR characterization of synthetic ges2U. Peaks are annotated according to Supplementary Figure 14.



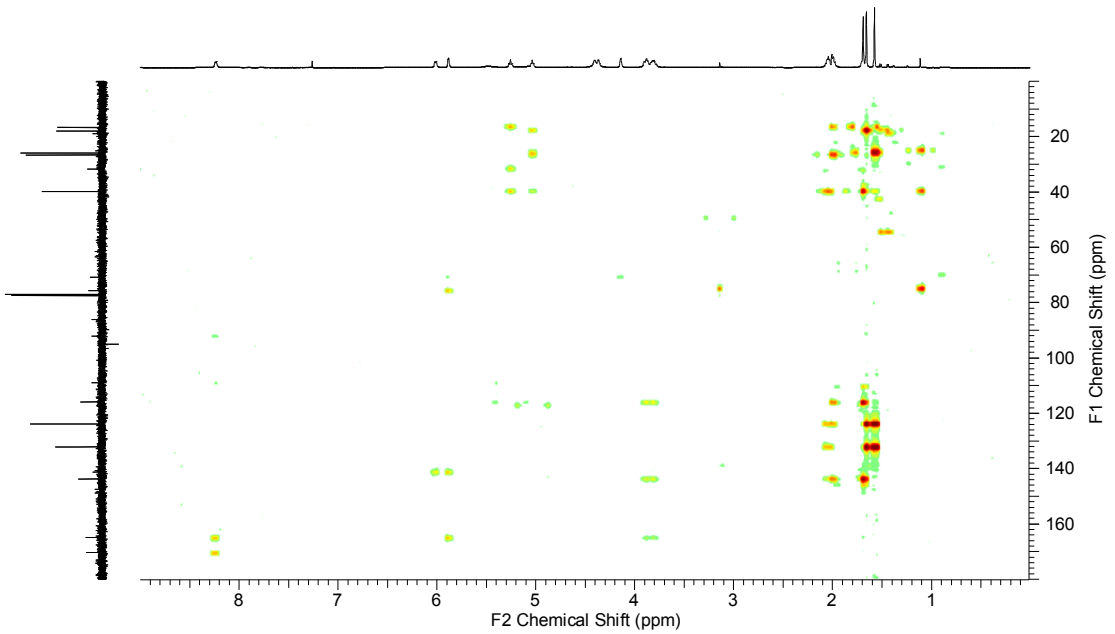
Supplementary Figure 17. ^{13}C-NMR characterization of synthetic ges2U. Peaks are annotated according to Supplementary Figure 14.



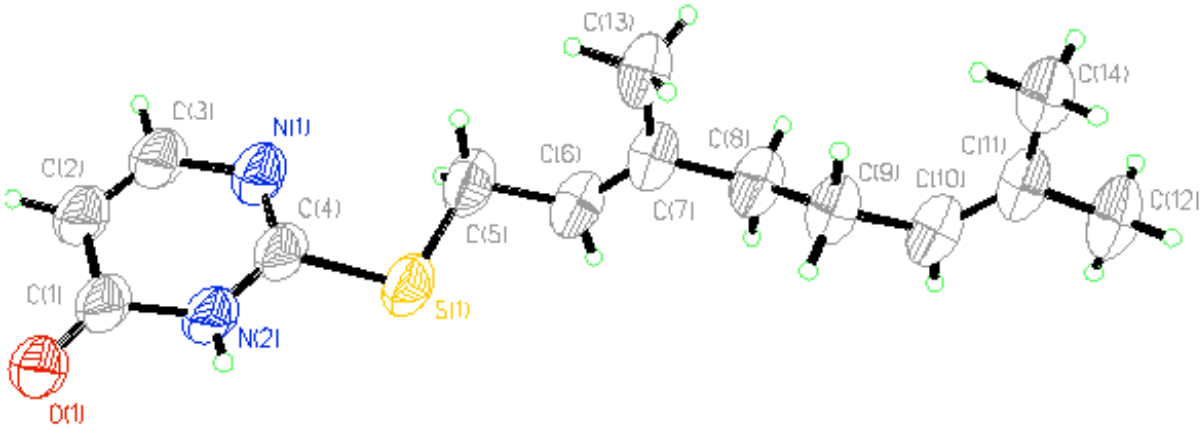
Supplementary Figure 18. gCOSY NMR spectrum of synthetic ges2U.



Supplementary Figure 19. gHSQC NMR spectrum of synthetic ges2U.



Supplementary Figure 20. gHMBC NMR spectrum of synthetic ges2U.



Supplementary Figure 21. Perspective views showing 50% probability displacement of synthetic geranyl-2-thiouracil based on X-ray diffraction.

Supplementary Table 4. Experimental details of geranyl-2-thiouracil crystallography.

	YCIV184001
Crystal data	
Chemical formula	$C_{14}H_{20}N_2OS$
M_r	264.38
Crystal system, space group	Triclinic, $P\bar{1}$
Temperature (K)	100
a, *b*, *c* (Å)	5.7216 (3), 7.6375 (3), 16.9440 (8)
α, β, γ (°)	87.624 (3), 82.534 (3), 75.252 (4)
V (Å^3)	709.94 (6)
Z	2
Radiation type	Cu *K*α
μ (mm^{-1})	1.94
Crystal size (mm)	0.36 × 0.12 × 0.02
Data collection	
Diffractometer	Bruker D8 goniometer with CCD area detector diffractometer
Absorption correction	Multi-scan *SADABS*
T_{min}, T_{max}	0.542, 0.962
No. of measured, independent and observed [$I > 2\sigma(I)$] reflections	11725, 2335, 1891
R_{int}	0.055
Refinement	
$R[F^2 > 2\sigma(F^2)]$, $wR(F^2)$, S	0.068, 0.180, 1.03
No. of reflections	2335
No. of parameters	170
No. of restraints	0
H-atom treatment	H atoms treated by a mixture of independent and constrained refinement
$\Delta\rho_{max}$, $\Delta\rho_{min}$ (e Å^{-3})	0.42, -0.17
Symmetry codes	(i) -*x*+3, -*y*+1, -*z*+1

Supplementary Table 5. Geometric parameters of the solved geranyl-2-thiouracil structure. Bond distances are in Å, and bond angles are in degrees.

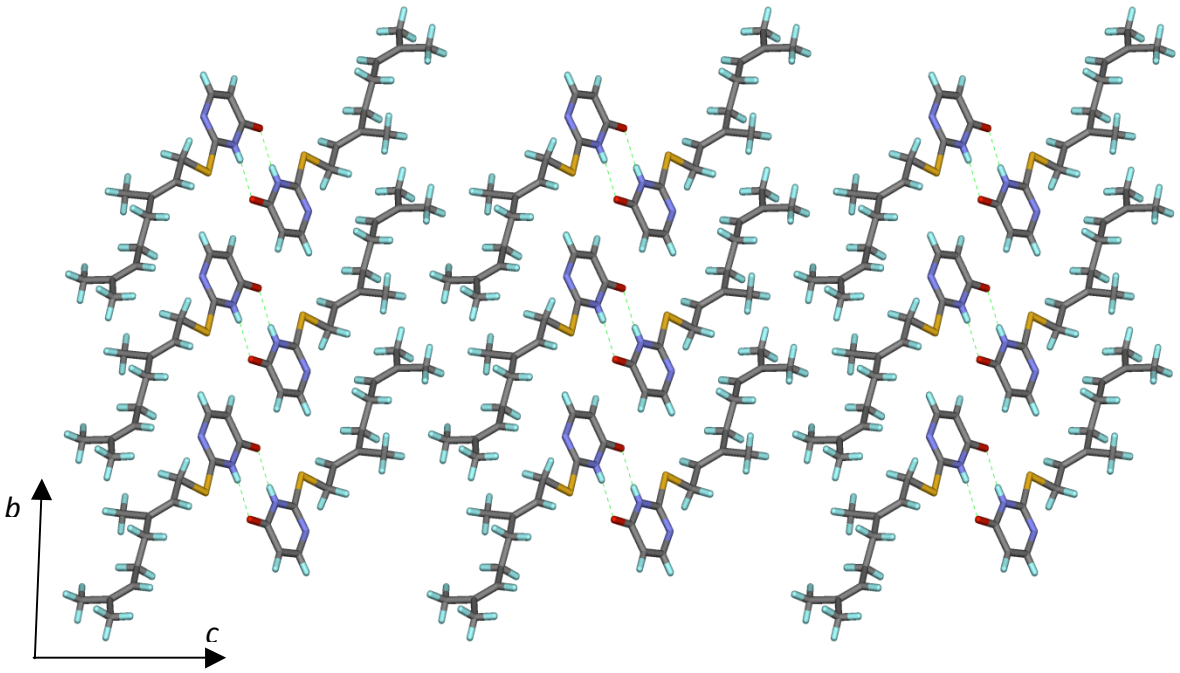
S1—C4	1.760 (4)	C8—C9	1.528 (6)
S1—C5	1.806 (5)	C8—H8A	0.9900
O1—C1	1.230 (5)	C8—H8B	0.9900
N1—C4	1.299 (5)	C9—C10	1.508 (5)
N1—C3	1.377 (5)	C9—H9A	0.9900
N2—C4	1.342 (6)	C9—H9B	0.9900
N2—C1	1.394 (5)	C10—C11	1.320 (6)
N2—H2	0.76 (4)	C10—H10	0.9500
C1—C2	1.433 (6)	C11—C14	1.489 (6)
C2—C3	1.340 (6)	C11—C12	1.516 (5)
C2—H2A	0.9500	C12—H12A	0.9800
C3—H3	0.9500	C12—H12B	0.9800
C5—C6	1.499 (5)	C12—H12C	0.9800
C5—H5A	0.9900	C13—H13A	0.9800
C5—H5B	0.9900	C13—H13B	0.9800
C6—C7	1.326 (6)	C13—H13C	0.9800
C6—H6	0.9500	C14—H14A	0.9800
C7—C13	1.488 (6)	C14—H14B	0.9800
C7—C8	1.508 (5)	C14—H14C	0.9800
C4—S1—C5	100.6 (2)	C9—C8—H8B	108.5
C4—N1—C3	114.3 (4)	H8A—C8—H8B	107.5
C4—N2—C1	123.8 (4)	C10—C9—C8	111.1 (3)
C4—N2—H2	124 (3)	C10—C9—H9A	109.4
C1—N2—H2	112 (3)	C8—C9—H9A	109.4

O1—C1—N2	120.5 (4)	C10—C9—H9B	109.4
O1—C1—C2	127.3 (4)	C8—C9—H9B	109.4
N2—C1—C2	112.2 (4)	H9A—C9—H9B	108.0
C3—C2—C1	119.8 (4)	C11—C10—C9	128.5 (4)
C3—C2—H2A	120.1	C11—C10—H10	115.8
C1—C2—H2A	120.1	C9—C10—H10	115.8
C2—C3—N1	125.4 (4)	C10—C11—C14	125.3 (4)
C2—C3—H3	117.3	C10—C11—C12	120.4 (4)
N1—C3—H3	117.3	C14—C11—C12	114.3 (4)
N1—C4—N2	124.5 (4)	C11—C12—H12A	109.5
N1—C4—S1	121.4 (3)	C11—C12—H12B	109.5
N2—C4—S1	114.1 (3)	H12A—C12—H12B	109.5
C6—C5—S1	108.8 (3)	C11—C12—H12C	109.5
C6—C5—H5A	109.9	H12A—C12—H12C	109.5
S1—C5—H5A	109.9	H12B—C12—H12C	109.5
C6—C5—H5B	109.9	C7—C13—H13A	109.5
S1—C5—H5B	109.9	C7—C13—H13B	109.5
H5A—C5—H5B	108.3	H13A—C13—H13B	109.5
C7—C6—C5	126.9 (4)	C7—C13—H13C	109.5
C7—C6—H6	116.6	H13A—C13—H13C	109.5
C5—C6—H6	116.6	H13B—C13—H13C	109.5
C6—C7—C13	124.7 (4)	C11—C14—H14A	109.5
C6—C7—C8	120.0 (4)	C11—C14—H14B	109.5
C13—C7—C8	115.2 (4)	H14A—C14—H14B	109.5
C7—C8—C9	115.0 (3)	C11—C14—H14C	109.5
C7—C8—H8A	108.5	H14A—C14—H14C	109.5
C9—C8—H8A	108.5	H14B—C14—H14C	109.5
C7—C8—H8B	108.5		

C4—N2—C1—O1	179.8 (4)	C5—S1—C4—N2	-176.6 (3)
C4—N2—C1—C2	-0.3 (6)	C4—S1—C5—C6	-171.9 (3)
O1—C1—C2—C3	179.8 (5)	S1—C5—C6—C7	-112.4 (5)
N2—C1—C2—C3	-0.1 (6)	C5—C6—C7—C13	4.1 (8)
C1—C2—C3—N1	0.2 (7)	C5—C6—C7—C8	-173.6 (4)
C4—N1—C3—C2	0.2 (7)	C6—C7—C8—C9	-125.3 (5)
C3—N1—C4—N2	-0.6 (6)	C13—C7—C8—C9	56.7 (6)
C3—N1—C4—S1	179.9 (3)	C7—C8—C9—C10	-178.4 (4)
C1—N2—C4—N1	0.7 (7)	C8—C9—C10—C11	119.5 (5)
C1—N2—C4—S1	-179.8 (3)	C9—C10—C11—C14	1.6 (7)
C5—S1—C4—N1	3.0 (4)	C9—C10—C11—C12	179.5 (4)

The parameters of one hydrogen bond in the structure are as follows:

D—H···*A*	*D*—H (Å)	H···*A* (Å)	*D*···*A* (Å)	*D*—H···*A* (°)
N2—H2···O1^{i}	0.76 (4)	2.01 (5)	2.771 (5)	173 (4)



Supplementary Figure 22. Synthetic geranyl-2-thiouracil crystal architecture viewed along the *a*-axis direction.

Supplementary Table 6. List of organisms used for RNA isolation.

Name	Source	Genotype	Growth media	Growth temperature	Lysis method
Escherichia coli BW25113	Keio collection at Yale	F-, Δ(araD-araB)567, ΔlacZ4787(::rrnB-3), λ-, rph-1, Δ(rhaD-rhaB)568, hsdR514	LB lennox broth	37 °C	detergent
Escherichia coli ΔmnmA	Keio collection at Yale (JW1119-1)	F-, Δ(araD-araB)567, ΔlacZ4787(::rrnB-3), λ-, ΔtrmU721::kan, rph-1, Δ(rhaD-rhaB)568, hsdR514	LB lennox broth	37 °C	detergent
Escherichia coli ΔmnmE	Keio collection at Yale (JW3684-1)	F-, Δ(araD-araB)567, ΔlacZ4787(::rrnB-3), λ-, rph-1, ΔtrmE737::kan, Δ(rhaD-rhaB)568, hsdR514	LB lennox broth	37 °C	detergent
Escherichia coli ΔselU	Keio collection at Yale (JW0491-1)	F-, Δ(araD-araB)567, ΔlacZ4787(::rrnB-3), ΔybbB786::kan, λ-, rph-1, Δ(rhaD-rhaB)568, hsdR514	LB lennox broth	37 °C	detergent
Enterobacter aerogenes	ATCC #15038		LB lennox broth	30 °C	detergent
Pseudomonas aeruginosa	Sagathalian lab, Harvard University		LB lennox broth	37 °C	detergent
Salmonella typhimurium LT2	Chen lab, Harvard University		LB lennox broth	37 °C	detergent
Vibrio fischeri	ATCC #700601		Zobell Marine Broth 2216 (VWR)	30 °C	detergent
Bacillus subtilis	Losick lab, Harvard University		LB lennox broth	37 °C	phenol
Streptomyces venezuelae	ATCC #10595		MYME	30 °C	phenol
Saccharomyces cerevisiae W303a	ATCC		YPD	30 °C	phenol
Schizosaccharomyces pombe h972-	Winston lab, Harvard Medical School		YES	30 °C	phenol

MYME: 100 g/L sucrose, 10 g/L maltose, 5 g/L peptone, 3 g/L yeast extract, 3 g/L malt extract