

Synthesis, base pairing and structure studies of geranylated RNA

Rui Wang^{1,2,†}, Sweta Vangaveti^{2,†}, Srivathsan V. Ranganathan², Maria Basanta-Sanchez², Phensinee Haruehanroengra^{1,2}, Alan Chen^{1,2} and Jia Sheng^{1,2,*}

¹Department of Chemistry, University at Albany, State University of New York, Albany, NY 12222, USA and ²The RNA Institute, University at Albany, State University of New York, Albany, NY 12222, USA

Received February 18, 2016; Revised May 30, 2016; Accepted June 2, 2016

ABSTRACT

Natural RNAs utilize extensive chemical modifications to diversify their structures and functions. 2-Thiouridine geranylation is a special hydrophobic tRNA modification that has been discovered very recently in several bacteria, such as *Escherichia coli*, *Enterobacter aerogenes*, *Pseudomonas aeruginosa* and *Salmonella Typhimurium*. The geranylated residues are located in the first anticodon position of tRNAs specific for lysine, glutamine and glutamic acid. This big hydrophobic terpene functional group affects the codon recognition patterns and reduces frameshifting errors during translation. We aimed to systematically study the structure, function and biosynthesis mechanism of this geranylation pathway, as well as answer the question of why nature uses such a hydrophobic modification in hydrophilic RNA systems. Recently, we have synthesized the deoxy-analog of S-geranyluridine and showed the geranylated T-G pair is much stronger than the geranylated T-A pair and other mismatched pairs in the B-form DNA duplex context, which is consistent with the observation that the geranylated tRNA^{Glu}_{UUC} recognizes GAG more efficiently than GAA. In this manuscript we report the synthesis and base pairing specificity studies of geranylated RNA oligos. We also report extensive molecular simulation studies to explore the structural features of the geranyl group in the context of A-form RNA and its effect on codon–anticodon interaction during ribosome binding.

INTRODUCTION

Natural RNA systems utilize a variety of chemical modifications to achieve structural and functional specificity and diversity. Currently, there are over 150 natural modifications that have been discovered in mRNA, rRNA, tRNA

and non-coding RNA of all three primary phylogenetic domains (Archaea, Bacteria and Eukarya) (1,2). These modifications have been increasingly demonstrated to play critical roles in many biological processes and be highly involved in many human diseases (3–6). Additionally, it is believed that these chemical modifications are few of the most evolutionarily conserved properties in RNAs, and some of the modified nucleobases are relics of the RNA World, where they may have enhanced the chemical diversity of RNA prior to the emergence of proteins (7). Therefore, studying the structures and functions, such as base pairing specificity and enzymatic recognition properties of these RNA chemical modifications, will be significant for further elucidation of RNA functions as both genetic information carriers and biological catalysts, the development of new RNA-targeted therapeutics, as well as origin of life studies.

tRNAs contain over 90 different chemical modifications that play central regulatory roles in codon–anticodon recognition, tRNA charging with specific amino acids and peptide synthesis during translation (8–16). Uracil is the most modified nucleobase and among the nearly 60 known uridine modifications, more than 15 contain C2 thiolation, which is known as 2-thiouridine and its C5 derivatives (1). The 2-thiouridine geranylation, in which a large hydrophobic geranyl group is covalently linked to the sulphur atom of the 2-thiouridine derivatives (as shown in Figure 1, compound 3, 4, 5), has been discovered very recently as a new natural RNA modification in several bacteria including *Escherichia coli*, *Enterobacter aerogenes*, *Pseudomonas aeruginosa* and *Salmonella Typhimurium*, at a frequency of up to 6.7% (~400 geranylated nucleotides per cell) (17). These modified residues are located in the first anticodon position (position 34) of tRNAs specific for lysine, glutamine and glutamic acid. The enzyme SelU, the selenouridine synthase that is known to transfer the 2-thiouridine to 2-selenouridine, another natural RNA analog, is also responsible for the biosynthesis of these geranylated nucleosides (Scheme 1) (18–20). This suggests that the RNA geranylation might also play regulatory roles in the selenonucleo-

*To whom correspondence should be addressed. Tel: +1 518 437 4419; Fax: +1 518 442 3462; Email: jsheng@albany.edu

†These authors contributed equally to the paper as first authors.

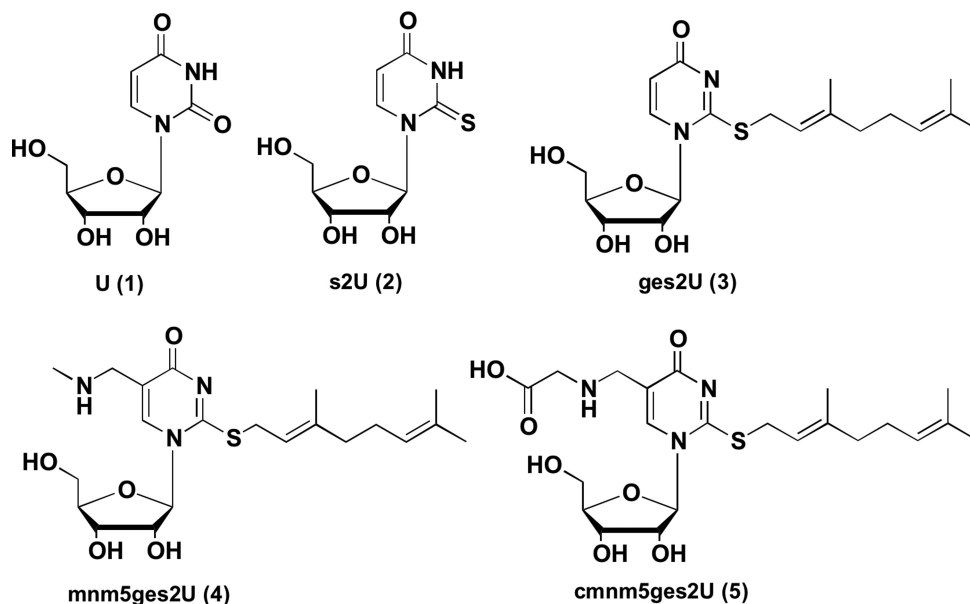
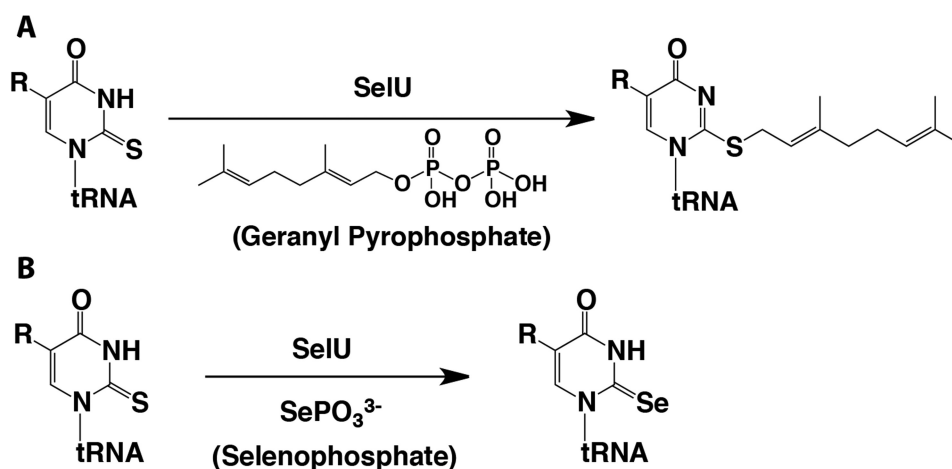


Figure 1. Chemical structures of uridine (U, 1), 2-thiouridine (s2U, 2), geranylated 2-thiouridine (ges2U, 3), geranylated 5-methylaminomethyl-2-thiouridine (mnm5ges2U, 4) and geranylated 5-carboxymethylaminomethyl-2-thiouridine (cmnm5ges2U, 5).



Scheme 1. The selenouridine synthase (SelU) has dual functions: (A) to install a geranyl group to the sulphur atom in the presence of geranyl pyrophosphate; and (B) to replace sulphur with selenium in the presence of selenophosphate. The R group represents the mnm- or cmnm- in position 5 of compound 4 and 5 in Figure 1.

side biosynthesis, an important biosynthetic pathway in live cells (21,22). Although the geranylated 2-thiouridine might merely be the intermediate product in the transformation of 2-thiouridine to 2-selenouridine (23,24), the question of why nature uses such a bulky hydrophobic group (the only terpene functionality discovered so far) in hydrophilic RNA systems remains unresolved. From an evolutionary point of view, it is quite reasonable to speculate that such hydrophobic terpene groups might be chemical relics from ancient RNA-mediated lipid synthesis (25).

This hydrophobic geranyl group has been demonstrated to affect the codon recognition and decrease the frameshifting errors during translation (17). In our studies, we have synthesized the deoxy-analog of S-geranyluridine and showed the geranylated T-G pair is much stronger than geranylated T-A and other mismatched pairs in the B-form

DNA duplex context. Our findings are consistent with the observation that the geranylated $\text{tRNA}^{\text{Glu}}_{\text{UUC}}$ recognizes GAG more efficiently than GAA in previous studies (26). In this manuscript, we present the synthesis and base pairing specificity studies of geranylated RNA oligos, as well as extensive molecular simulation studies to explore the structural features of the geranyl group in A-form RNA duplex in comparison with B-form DNA, and its effect on codon-anticodon interaction during ribosome binding.

MATERIALS AND METHODS

Synthesis of 2-thio-geranyluridine phosphoramidite

1-((2R,3R,4R,5R)-5-((bis(4-methoxyphenyl)(phenyl)methoxy)methyl)-3-((tert-butyl)dimethylsilyl)oxy)-4-hydroxytetrahydrofuran-2-yl)-

2-(((*E*)-3,7-dimethylocta-2,6-dien-1-yl)thio)pyrimidin-4(1*H*)-one (**8**). The synthesis of compound **7** in Scheme 2 are presented in the Supplementary Data. A solution of **7** (573 mg, 1 mmol), geranyl bromide (0.6 ml, 3 mmol) and *N,N*-diisopropylethylamine (0.53 ml, 3 mmol) in MeOH (10 ml) was stirred at 25°C for 12 h. The resulting reaction mixture was quenched with water (10 ml), washed with brine (8 × 30 ml) and dried over Na₂SO₄. The crude product was concentrated *in vacuo* and directly subjected to silica gel chromatography. A sticky liquid (651 mg, 92% yield) was obtained after column flash chromatography, TLC R_f = 0.4 (1% MeOH in CH₂Cl₂). ¹H nuclear magnetic resonance (NMR) (400 MHz, CD₃OD) δ 7.77 (d, *J* = 8.0 Hz, 1H, H-6), 7.46 (d, *J* = 6.8 Hz, 2H), 7.37 (d, *J* = 4.4 Hz, 2H), 7.30 (d, *J* = 8.4 Hz, 2H), 7.22-7.09 (m, 4H), 6.19 (d, *J* = 7.6 Hz, 1H), 5.83 (d, *J* = 8.0 Hz, 1H), 5.42 (t, *J* = 7.6 Hz, 1H), 5.02 (d, *J* = 5.6 Hz, 1H), 4.27-4.24 (m, 1H), 4.06 (br s, 1H), 3.99-3.92 (m, 2H), 3.82 (d, *J* = 8.4 Hz, 1H), 3.76 (s, 3H), 3.75 (s, 3H), 3.64 (d, *J* = 11.6 Hz, 1H), 3.42 (d, *J* = 4.4 Hz, 1H), 1.96 (br s, 4H), 1.72 (s, 3H), 1.63 (s, 3H), 1.49 (s, 3H), 0.81 (s, 9 H), 0.05 (s, 3H), 0.01 (s, 3H); ¹³C NMR (100 MHz, CD₃OD) δ 169.98, 164.37, 158.99, 158.83, 144.83, 142.05, 140.09, 135.25, 135.05, 130.19, 129.88, 128.93, 127.73, 127.63, 127.37, 127.02, 126.53, 126.28, 123.39, 116.84, 112.74, 112.33, 108.37, 90.72, 87.69, 77.43, 71.20, 63.83, 54.19, 54.18, 54.15, 40.64, 31.33, 27.38, 26.42, 25.89, 19.30, 17.78, 16.69, -5.32, -5.44; HRMS (ESI-TOF): molecular formula, C₄₆H₆₀N₂O₇SSi; [M+H]⁺: 814.3970 (calc. 814.1283).

3-(((2*R*,3*R*,4*R*,5*R*)-2-((bis(4-methoxyphenyl)(phenyl) methoxy)methyl)-4-((tert-butyl)dimethylsilyloxy)-5-(2-(((*E*)-3,7-dimethyl-octa-2,6-dien-1-yl)thio)-4-oxypyrimidin-1(4*H*)yl) tetrahydrofuran-3-yl)oxy)-(diisopropylamino)phosphino)propanenitrile (**9**). To a solution of compound **8** (200 mg, 0.31 mmol) in CH₂Cl₂ (5.0 ml), DIPEA (0.22 ml, 0.62 mmol) was added at 0°C under Ar atmosphere. 2-Cyanoethyl *N,N*-diisopropylchlorophosphoramidite (0.14 ml, 0.12 mmol) was carefully injected to this mixture over 15 min at 0°C. The reaction was warmed to ambient temperature for 5 h before being quenched with sodium bicarbonate solution. The solvent was extracted with CH₂Cl₂ and the organic layer was evaporated. Flash column chromatography was used for purification, TLC R_f = 0.4 (1% MeOH in CH₂Cl₂), to yield **9** as a white sticky solid. ¹H NMR (400 MHz, CD₃OD) δ 7.69 (d, *J* = 8.4 Hz, 1H, H-6), 7.43-7.21 (m, 10 H, H-Ar), 6.79 (d, *J* = 8.4 Hz, 2H, H-Ar), 6.72 (d, *J* = 8.4 Hz, 2H, H-Ar), 6.20 (d, *J* = 7.6 Hz, 1H), 5.80 (d, *J* = 7.6 Hz, 1H), 5.41 (d, *J* = 8.4 Hz, 1H), 5.02 (d, *J* = 8.0 Hz, 1H), 4.60 (s, 2H), 4.46-4.42 (m, 1H), 4.41 (s, 1H), 4.05-3.90 (m, 4H), 3.78 (s, 3H), 3.76 (s, 3H), 2.72 (dd, *J* = 12.0, 6.4, Hz, 2H), 2.15 (s, 6H), 1.95 (s, 4H), 1.72 (s, 3H), 1.64 (s, 3H), 1.50 (s, 3H), 1.29 (s, 4H), 1.21 (t, *J* = 6.4 Hz, 12H), 0.83 (s, 9H, H-^tBu), 0.06 (s, 3H, H-Me), 0.03 (s, 3H, H-Me); ¹³C NMR (100 MHz, CD₃OD) δ 164.35, 158.99, 158.86, 144.82, 142.06, 136.23, 136.12, 134.76, 130.55, 130.20, 128.04, 127.36, 126.61, 123.45, 116.77, 112.78, 108.42, 101.91, 90.45, 63.89, 58.57, 54.16, 43.01, 42.88, 29.86, 25.90, 24.99, 24.40, 23.58, 23.49, 19.59, 19.52, 17.85, 16.31, 15.22, -6.74, -6.94; ³¹P NMR (162 MHz, CD₃OD)

δ 147.52, 147.28; HRMS (ESI-TOF): molecular formula, C₅₅H₇₇N₄O₈PSSi; [M+H]⁺: 1013.3993 (calc. 1013.3461).

Synthesis of RNA oligonucleotides

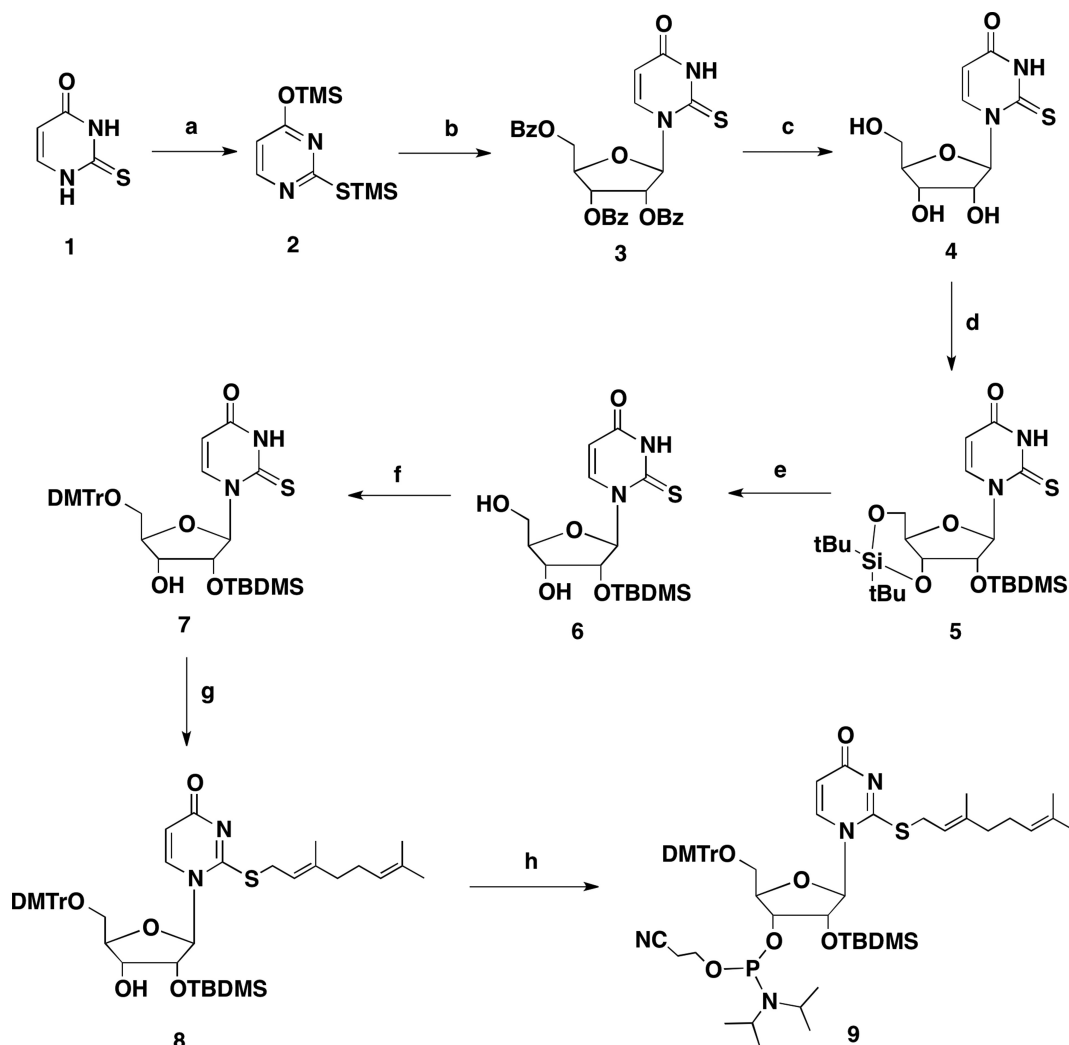
The oligonucleotides were chemically synthesized at 1.0 μmol scales by solid phase synthesis using a MerMade MM8 synthesizer. The geranyl uridine phosphoramidite was dissolved in dichloromethane to a concentration of 0.07 M. I₂ (0.02 M) in THF/Py/H₂O solution was used as an oxidizing reagent. Coupling was carried out using 5-ethylthio-1*H*-tetrazole solution (0.25 M) in acetonitrile for 6-min, for both native and modified phosphoramidites. About 3% trichloroacetic acid in methylene chloride was used for the 5'-detritylation. Synthesis was performed on control-pore glass (CPG-500) immobilized with the appropriate nucleoside through a succinate linker. All the reagents used are standard solutions obtained from ChemGenes Corporation. The oligonucleotide was prepared in DMTr-off form. After synthesis, the oligos were cleaved from the solid support and fully deprotected with concentrated ammonium solution at room temperature for 14 h. The solution was evaporated to dryness by Speed-Vac concentrator and was desilylated using a triethylamine trihydrogen fluoride (Et₃N•3HF) solution at 65°C for 2.5 h. The reaction was quenched with 1 ml of water and the RNA was precipitated by adding 0.2 ml of 3 M sodium acetate and 6 ml of *n*-butanol. The solution was cooled to -80°C for 1 h before the RNA was recovered by centrifugation and finally dried under vacuum.

HPLC analysis and purification

RNA oligonucleotides were purified by ion-exchange HPLC using a PA-100 column from Dionex at a flow rate of 1 ml/min. Buffer A was pure water, and buffer B contained 2M ammonium acetate (pH 7.1). The RNA oligonucleotides were eluted with a linear gradient of 0–50% buffer B over 20 min. The collected fractions were lyophilized, desalted with Waters Sep-Pac C18 columns and re-concentrated.

Enzymatic hydrolysis and UHPLC-MS/MS analysis of geranyl-RNA oligo

Major RNA nucleoside standards were purchased (Sigma-Aldrich Co). Isotopically labeled guanosine [¹³C][¹⁵N]G used as the internal standard (IS) was a gift from Cambridge Isotope Laboratories, Inc. Prior to UHPLC-MS/MS analysis, 1.0 pg/μl [¹³C][¹⁵N]-G was added, as IS, to 100 ng of total RNA to be hydrolyzed to the composite mononucleosides via a two-step enzymatic hydrolysis. The first step of phosphodiester bond cleavage was accomplished with nuclease P1, resulting in nucleoside-5'-monophosphates. Optimum nuclease P1 activity was achieved at pH 5.5 by the addition of 1/10 of the volume of 1.0 M ammonium acetate at pH 5.5. For each 0.5 absorbance unit of RNA, two units of nuclease P1 were added and incubated overnight at 37°C. The second step of nucleoside preparation uses bacterial alkaline phosphatase (BAP) to cleave the 5'-phosphate from the nucleosides resulting in individual nucleosides and



Scheme 2. Synthesis of *S*-2-thiouridine phosphoramidite **9**. Reagents and conditions: (a) TMSCl, HMDS; (b) 2,3,5-tri-*O*-benzoyl- β -D-ribofuranose, SnCl₄, 1,2-dichloroethane; (c) NaOMe, MeOH; (d) Di-*tert*-butylsilyl ditriflate; TBDMS-Cl, Imidazole, DMF; (e) HF•Py, Py; (f) DMTrCl, Py; (g) geranyl bromide, DIPEA, MeOH; (h) (*i*-Pr₂N)₂P(Cl)OCH₂CH₂CN, (*i*-Pr)₂NEt, CH₂Cl₂.

phosphoric acid. For optimum BAP activity, the pH was adjusted to pH 8.3 by adding 1/10 of the volume of 1.0 M ammonium bicarbonate pH 8.3. One unit of BAP was added for each 0.5 absorbance units of RNA and incubated at 37°C for 2 h. The nucleoside products were lyophilized and stored at -20°C. Samples were reconstituted in 0.01% formic acid in RNase-free water to a final concentration of 1 ng/ μ l prior to UHPLC-MS/MS analysis. A negative control sample was included in the dataset. The control comprised of the enzymes, IS and reagents used during the enzymatic hydrolysis.

Hydrolyzed oligos were subjected to chromatography using a Waters ACQUITY I-Class UPLCTM (Waters, USA) liquid chromatographic system equipped with a binary pump and autosampler that was maintained at 4°C. A Waters ACQUITY UPLCTM HSS T3 column (2.1 \times 50 mm 1.7 μ m) and a HSS T3 guard column (2.1 \times 5 mm, 1.8 μ m) were used for the separation. The assay was completed at a flow rate of 0.2 ml/min and column temperature of 25°C. Mobile phases included RNase-free water (18.0 M Ω cm⁻¹)

containing 0.01% formic acid (Buffer A) and 50% acetonitrile in aqueous 0.01% formic acid pH 3.5 (Buffer B). Tandem MS analysis of DNA nucleosides was performed on a Waters XEVO TQ-STM (Waters, USA) triple quadrupole mass spectrometer equipped with an electrospray ionization (ESI) source maintained at 150°C with the capillary voltage set at 1kV. The desolvation gas, nitrogen, was maintained at 500 l/h and the desolvation temperature was set at 500°C. The cone gas flow and nebulizer pressure were set to 150 l/h and seven bars, respectively. Composition analysis was performed in ESI positive-ion mode using multiple-reaction monitoring (MRM). All of the commercial and synthesized nucleoside standards were characterized individually by direct infusion MS. The ion transitions, cone voltage and collision energy used for UHPLC-MS/MS were determined using MassLynx V4.1Intellistart software. Retention times and the corresponding protonated molecular and product ion pairs [MH⁺/BH²⁺] were obtained for each individual nucleoside.

UV and circular dichroism (CD) spectroscopy

UV-Vis absorption spectra were measured on a Perkin-Elmer Lambda 900 UV-Vis spectrometer. Methanol was used as the solvent to dissolve the nucleoside. Circular dichroism (CD) spectra were recorded with 5 μ M RNA duplexes in 20 mM phosphate buffer (pH 7.0) on a JASCO-815 spectropolarimeter at 20°C over a wavelength range of 200–320 nm using a 1 cm path length quartz cuvette with a scanning speed of 100 nm/min, bandwidth of 1.0 nm and D. I. T of 1.0 s. Each spectrum was averaged from an accumulation of four scans and baseline-corrected against the buffer.

Thermodenaturation of the RNA duplexes

Solutions of the duplex RNAs (1 μ M) were prepared by dissolving the purified RNAs in sodium phosphate (10 mM, pH 6.5) buffer containing 100 mM NaCl. The solutions were heated to 85°C for 3 min, then cooled slowly to room temperature and stored at 4°C overnight before melting temperature (T_m) measurements. Prior to thermal denaturation, the geranyl RNA duplex was bubbled under argon for 5 min. Data points for each denaturing curves were acquired at 260 nm by heating and cooling from 5 to 85°C four times at a rate of 0.5°C/min, using Cary-300 UV-Visible spectrometer equipped with a temperature controller system. The block temperature was used as the standard. The thermodynamic parameters of each duplex strand were obtained by fitting the melting curves in the Meltwin software (27).

Simulation methods

DNA/RNA duplexes. To study the ges2U modification in the context of the duplex using molecular dynamics (MD) simulations, we developed AMBER (28) type force-field parameters for the atoms of the modified nucleoside. For obtaining the partial charges on the atoms, we used the on-line RESP charge-fitting server, REDS (29). The geometry of the modified nucleoside was energy minimized, and Hartree-Fock level theory and 6–31G* basis-sets were employed to arrive at a set of partial charges (30). AMBER-99 force-field parameters were used for bonded interactions (28), and AMBER-99 parameters with Chen-Garcia corrections were used for LJ interactions (31). The unmodified RNA and DNA duplexes were constructed using nucleic acid builder suite of AMBER 11 package. The sequence of the RNA duplex is 5'-GGACUXCUGCAG-3' and 3'-CCUGAYGACGUC-5', consistent with the experimental work presented here. X and Y correspond to U/ges2U and A/G, respectively. The DNA duplex sequence is 5'-CTTCTXGTCCG-3' and 3' GAAGAYCAGGC-5', consistent with our previous work (26). X and Y are the DNA counterparts of the above mentioned RNA nucleotides. Using the WebMO graphical editor, we performed mutations such as, U to ges2U, T to ges2T and A to G, to get the different RNA and DNA duplexes for MD simulation studies. The simulation system included the DNA or RNA duplex in a solution of 1M NaCl solution in a 3D periodic box. The box size was 6 \times 6 \times 6 nm³, containing the duplex, 152 Na⁺ ions, 130 Cl⁻ ions and roughly 6600 water molecules.

The system was subjected to energy minimization to prevent any overlap of atoms, followed by a 10 ns equilibration MD run. 10 parallel simulations with the same equilibrated starting conformations but with different starting velocities were performed for 100 ns each, totaling 1 μ s of production run for each DNA/RNA duplex. Coordinates were stored every picosecond for further analysis.

Ribosome. To study the effects of ges2U modification on codon–anticodon interactions, two crystal structures of 16S ribosomal subunit of *Thermus thermophilus* containing the anticodon stem loop (ASL) of tRNA^{Lys} and the mRNA codon fragment in the decoding center, were obtained from the protein data bank (PDB ID: 1XMO, 1XMQ). The two structures differ in their codon sequences (1XMO has AAG while 1XMQ has AAA as the codon). In order to investigate the effects of modification, simulations were set up for each codon–anticodon pair, with and without the ges2U modification at position 34 of the ASL of tRNA^{Lys}. To further include the effects of the naturally occurring modifications at the position 5 of the base, the methyl-amino-methyl (mmm) modified 2-thiouridine was used to mutate corresponding U34. The missing atoms in the ribosome-associated proteins were fixed using MOE (32). The simulation system included the ribosomal subunit with all its associated proteins, the ASL of the tRNA and the mRNA codon fragment in a solution of 1 M KCl in a 3D periodic box. The box size was 26 \times 26 \times 26 nm³, containing \sim 11 800 K⁺ ions, \sim 10 700 Cl⁻ ions and roughly 539 600 water molecules (1.6×10^6 atoms). The system was subjected to energy minimization to prevent any overlap of atoms, followed by a 1 ns equilibration and 5 ns production MD run. Coordinates of the RNA components (rRNA, tRNA, mRNA) of the system were stored every 1 ps for further analysis. A schematic of the simulation set up is illustrated in Supplementary Figure S23.

All MD simulations were performed using Gromacs-4.6.3 package (33). The MD simulations incorporated leap-frog algorithm with a 2 fs time-step to integrate the equations of motion. The system was maintained at 300 K, using the velocity rescaling thermostat (34). The pressure was maintained at 1 atm using the Berendsen (35) and Parrinello-Rahman barostat (36) for equilibration and production runs, respectively. The long-ranged electrostatic interactions were calculated using particle mesh Ewald (PME) (37) algorithm with a real space cut-off of 1.2 nm. LJ interactions were also truncated at 1.2 nm. TIP3P model was used represent the water molecules, and LINCS (38) algorithm was used to constrain the motion of hydrogen atoms bonded to heavy atoms.

RESULTS AND DISCUSSION

Synthesis of geranyl-2-thiouridine phosphoramidite and geranyl-RNA oligonucleotide

We started the synthesis of 2-thiouridine (4, Scheme 2) through regular Vorbrüggen glycosylation of the protected ribofuranose with silylated 2-thiouracil (2) in the presence of Tin (IV) chloride, followed by the deprotection of the benzoyl groups using base treatment. The simultaneous silylation of 3' and 5'-hydroxyl groups with Di-*tert*-

butylsilyl (DTBS) ditriflate followed by the 2'-protection with *tert*-butyldimethylsilyl (TBDMS) group gave the silylated 2-thiouridine compound **5**. Compound **5** was then selectively desilylated with hydrogen fluoride in pyridine and tritylated with trityl chloride at the 5' position to generate the key intermediate **7** with satisfying yield. The geranyl group was installed to the 2-sulfur center before the final phosphoramidite building block **9** was synthesized. It is noteworthy that the geranyl-2-thiouridine nucleoside, which is the 2-thio-geranylated product of compound **4**, shows a UV absorption peak at 240 nm, compared to the 260 nm for native uridine (Supplementary Figure S24). We have demonstrated previously that this *S*-2-thio-geranyl group is well compatible with the solid phase synthesis conditions by using dichloromethane as the solvent. In order to test its stability under fluoride treatment, which is necessary for regular RNA deprotection and purification process, we treated the *S*-2-geranyl-thiouridine nucleoside with Et₃N•3HF/Et₃N at 65°C for 4 h. No decomposed product was detected. After we applied the phosphoramidite to the solid phase synthesizer and made the geranylated RNA oligonucleotide, the ESI-MS of the purified RNA showed the correct mass with the presence of the geranyl group (Supplementary Figure S25).

UHPLC-MS/MS analysis of geranyl nucleoside in RNA oligo

In order to further confirm the incorporation and positions of geranyl modification in RNA oligonucleotide, we hydrolyzed the geranyl strand and its native counterpart to mononucleosides via nuclease P1 and BAP treatment. The resultant nucleoside mixture was separated by ultra-high performance liquid chromatography (UHPLC) in less than 20 min and detected by a tandem mass spectrometry (MS/MS) method that was recently published (39). The geranyl modified nucleoside was confirmed using MS/MS where the molecular ions [MH⁺] are fragmented at the glycosidic bond providing the product ion [BH₂⁺] and the neutral sugar residue. The resultant [MH⁺]/[BH₂⁺] ion-pair (Supplementary Table S1) were then monitored using MRM. Individual MRMs selected to verify the oligonucleotide composition and successful incorporation of geranyl-nucleoside included ges2U 397.3→265.1; G 284.1→152.1; A 268.2→136.0; U 245.1→133.0; C 243.2→112.1 (Supplementary Figure S26). *S*-2-thiouridine (ges2U) was detected only on the target oligo (Supplementary Figure S26A) in comparison to the native RNA oligo, which was included as a negative control (Supplementary Figure S26B). The synthetic geranyl-nucleoside, which is the 2-thio-geranylated product of compound **4**, was used to develop the MRM method and also as our positive control (Supplementary Figure S26C).

Thermal denaturation and base pairing studies of geranyl-RNA duplex

With the geranylated RNA strand in hand, we studied the base pairing stability and specificity of geranyl-uridine (ges2U) against other bases through thermal denaturation experiments. The UV-*T*_m curves of native and ger-

anylated RNA duplexes, 5'-GGACUXCUGCAG-3' and 3'-CCUGAYGACGUC-5', with Watson-Crick and other non-canonical base pairs (X pairs with Y) are shown in Supplementary Figure S27A and B. Their detailed temperature data are summarized in Table 1. Compared to our previous DNA duplex results, this hydrophobic geranyl group in the Watson-Crick face of RNA duplex has stronger effects on decreasing the overall stability of duplexes containing both native and non-canonical base pairs. As indicated in Table 1, one single geranylation decreases the *T*_m by 26.2°C, corresponding to a Δ*G*⁰ reduction of 11.0 kcal/mol, for the native duplex (entry 1 versus 5). The *T*_m drops 13.3°C for the U-G mismatch containing duplex after geranylation (entry 2 versus 6), 13.7°C for the U-C mismatch one (entry 3 versus 7) and 14.3°C for the U-U mismatch one (entry 4 versus 8), corresponding to a Δ*G*⁰ reduction of 6.7, 4.4 and 5.5 kcal/mol, respectively. When directly comparing the Watson-Crick base pairs (U-A and ges2U-A) with their own other mismatched pairs, as shown in the Δ*T*_m column, it is clear that the geranyl-2-thiouridine has a stronger pairing preference for G compared to other bases. The *T*_m of duplexes containing ges2U-G pair is 10°C higher than the native ges2U-A duplex, with an increased Δ*G*⁰ of 2.6 kcal/mol (entry 5 and 6). Both the ges2U-C and ges2U-U duplexes have very similar *T*_m as the ges2U-A one. In comparison, the native duplexes with U-G and U-A pairs have similar stability and are much stronger than U-C and U-U paired ones, with a slightly lower *T*_m (2.9°C, entry 1 and 2) in the U-G duplex. In this sense, the introduction of geranyl group could increase the base pair specificity by stabilizing the ges2U-G pair.

CD analysis of geranylated RNA duplexes

We also checked the conformation of the geranylated RNA duplexes in the context of both ges2U-A and ges2U-G pairs by CD spectroscopy. As shown in Supplementary Figure S28, both geranylated RNA duplexes showed very similar conformation in comparison to their native counterparts. In the U-A duplexes spectrum (Supplementary Figure S28A) there is a strong positive peak around 270 nm, a relatively weak negative peak around 240 nm and a weak positive peak around 225 nm, all of which resemble the characteristic peaks of the normal A-form structure in solution (40,41). In comparison, the spectrum of U-G pair containing duplexes (Supplementary Figure S28B) shows that the geranylated duplex retains a slightly better A-form conformation than the native one based on a slight shift to 260 nm for the strong positive peak, and the absence of the negative peak at 240 nm and positive peak at 225 nm. The CD spectra of nucleic acids are affected by many factors including sequence, base stacking and conformation; given this, our data indicates that the geranyl modification does not cause gross perturbations in overall helical structures and folding but does have some impact on local conformations.

Molecular simulation of geranylated RNA duplexes and the ribosomal binding of geranylated tRNA anticodon loop

We further examined the structural features of geranyl-modified RNA duplexes containing both ges2U-A and

Table 1. Duplex stability and base pairing specificity of geranyl-2-thiouridine (ges2U) in the context of a 12mer RNA duplex 5'-GGACUXCUGCAG-3' & 3'-CCUGAYGACGUC-5' (X pairs with Y)

Entry	Base pairs		T_m (°C) ^a	ΔT_m (°C) ^b	$-\Delta G^\circ_{37}$ (kcal/mol)
	X	Y			
1	U	A	60.8		18.6
2	U	G	57.9	-2.9	16.9
3	U	C	48.3	-12.5	11.9
4	U	U	50.4	-10.4	13.4
5	ges2U	A	34.6		7.6
6	ges2U	G	44.6	+10.0	10.2
7	ges2U	C	34.6	0	7.5
8	ges2U	U	36.1	+1.5	7.9

^aThe T_m were measured in sodium phosphate (10 mM, pH 7.0) buffer containing 100 mM NaCl.

^b ΔT_m values are relative to the duplex with native U-A pair and ges2U-A pair respectively.

^cObtained by non-linear curve fitting using Meltwin 3.5.

ges2U-G pairs using molecular simulations. Consistent with our previous DNA work with geranyl modified uridines, most of the ges2U-A pairs are in weakly paired state with one or zero hydrogen bonds observed over the course of our simulation, although two different hydrogen-bonding modes were transiently observed with the N6 hydrogen of adenosine bonding to either N3 or O4 of ges2U (Figure 2A and B). When the base-pairing is weak, the modified base with its hydrophobic tail fluctuates rapidly and adopts a variety of random conformations, dramatically distorting the local helical structure. This loss of hydrogen bonding and distortion in the helical structure might be the major reason that the ges2U-A pair containing duplex in either A-form or B-form has a much lower T_m than their native counterparts. In contrast, when the complementary base is guanine, the geranyl-uracil ring is held in place by two stable hydrogen bonds (Figure 2C) causing the orientation of the hydrophobic tail to be more tractable and dictated by the conformation of the duplexes. As shown in Supplementary Figure S29, the geranyl modification breaks the base-pairing of the uridine/thymidine with adenosine as evidenced by the probability being highest for zero hydrogen bonds. However, with the guanine nucleotide, the geranylated group forms a stable base-pair in both the DNA and RNA, as seen in the dominant probability for two hydrogen bonds between the bases.

More interestingly, in the case of the B-form DNA, the geranyl group was observed to be mainly accommodated in the minor groove and could orient to both 5' and 3' directions (green and purple sticks in Figure 2D). In addition, the geranyl group has strong interactions with the hydrophobic pocket of the duplex, as indicated by the red colored region in Figure 2D. The surface of the DNA and the RNA is colored according to the amount of atomic contacts with the geranyl group (red = high; blue = low). In the A-form RNA duplex, where the hydrophobic pocket is dramatically diminished due to the presence of 2'-OH groups, the geranyl group was observed to point out of the wide and shallow minor groove of the helical structure and mainly point to the 3'-end of the geranyl strand with backbone interactions (Figure 2E). As a result, the overall duplex remains close to the native one. Figure 2F shows rPP', a plot of strand separation using the interstrand distances between the phos-

phate atoms along the duplex in both A-form and B-form contexts. As expected, the strand separation is higher for the A-form RNA as compared to the B-form DNA. The highlighted region shows the portion of the helices most affected by the presence of the geranylated-thio modification. In the case of the DNA, we observed a substantial difference in the separation between the strands (the two blue curves), indicating widening of the minor groove to accommodate the hydrophobic geranyl group. In contrast, the minor groove is not conducive for the binding of the hydrophobic tail in the RNA, causing minimal changes in the interstrand separation (the two red curves). These specific conformations might also explain the T_m difference between the ges2U-G and U-G in RNA duplex (~13°C) being higher than their DNA counterparts (~2°C) (26). The favorable binding of the geranyl group to the minor groove of the DNA keeps the duplex almost as stable as the unmodified DNA, which is consistent with the duplex stability enhancement of regular minor-groove binders (42), whereas the stability of the modified RNA drops significantly compared to its unmodified version.

To investigate the role of geranyl-2-thiouridine in anticodon-codon recognition of tRNAs, we further performed MD simulations of the ASL of tRNA^{Lys} containing the modifications at the position 34. The 2-thiouridine modification at the U34 position of tRNA^{Lys}_{UUU} is known to be important for its codon recognition (43), and the 2-thio-group has also been shown to exist in its geranyl form conjugating with other functionalities such as the methyl-amino-methyl (mnm) group at the position 5 of the uracil (17). Therefore, we studied the mnm-5-geranyl-2-thio modified (mnm⁵ges²U₃₄) and the unmodified versions of tRNA^{Lys}_{UUU} in the presence of its cognate and near cognate mRNA fragments at the A-site of ribosome using molecular simulations. Two structures of 16S ribosomal subunit of *T. thermophilus* containing the ASL of tRNA^{Lys}_{UUU} and the mRNA codon fragment in the decoding center were used as our starting models. The two structures differ in their codon sequences (1XMO has AAG while 1XMQ has AAA as the codon). As explained in the methods section, we used the crystal structures and mutated U34 base to mnm⁵ges²U₃₄ and performed all-atom MD simulations to compare the effect of ges2U modification on the U:A and U:G base pairs

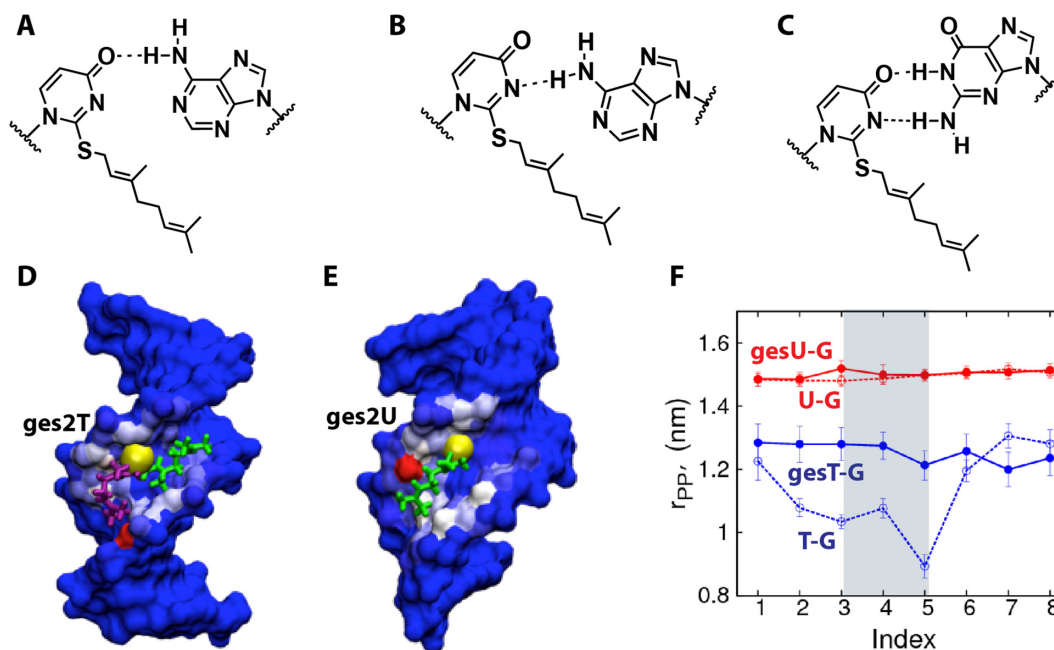


Figure 2. Molecular simulation studies of geranylated duplexes. (A and B) Proposed base pairing patterns between geranylated 2-thiouridine (ges2U) and A with one hydrogen bond between either O4 or N3 of ges2U and N6 of A. (C) Proposed base pairing patterns between geranylated 2-thiouridine (ges2U) and G with two stable hydrogen bonds. (D and E) Contact map of the geranyl group in the B-form DNA duplex and A-form RNA, respectively; red = high, blue = low. In B-form DNA, the geranyl group fits well into the minor groove and interacts with the hydrophobic pocket (red/white colored surface), pointing either to 5' or 3'-directions (green and purple sticks). In the RNA, the geranyl groups interact with the backbone oxygen atoms through their methyl groups. The dominant orientation of the geranyl group in shown as green sticks. The yellow spheres represent the sulfur atom. (F) The duplex separation plotted by the interstrand phosphate distances. For the B-form DNA, $r_{PP'}$ is calculated as the distance between the *i*th phosphate group from the 5' end of strand 1 and *i*-third phosphate group from the 3' end of strand 2. For the RNA, $r_{PP'}$ is calculated as the distance between the *i*th phosphate group from the 5' end of strand 1 and *i*-second phosphate group from the 3' end of strand 2. This choice corresponds to the phosphate group pairs that are closest to each other.

that are directly involved in the codon–anticodon interactions. Given the size of this system (~ 2 million atoms including solvent), a reversible binding simulation of the tRNAs to show differences in binding affinity is beyond the scope of our work. Therefore, we performed nanosecond timescale simulation of this system, which can specifically answer the following two questions: (i) when modified with the $mnm^5\text{ges}^2\text{U}_{34}$ group, can U34 adopt conformations to allow for stable Watson–Crick-like hydrogen bonds with the G codon (44) at the ribosomal binding site? (ii) If so, can the long hydrophobic geranyl group be accommodated at the A-site without disruption of the codon–anticodon and ribosomal interactions? We used our simulation data to monitor hydrogen bonding patterns between codon–anticodon bases and ribosomal bases, and other structural changes that occurred as a result of introducing the modified base in the ASL. The distance and angle cut off for formation of a H-bond were set at donor–acceptor (D–A) distance $\leq 3.3\text{\AA}$ and hydrogen–donor–acceptor (H–D–A) angle $\leq 30^\circ$.

As shown in Figure 3, our main result from these simulations is that the hypermodified U34 shows weak hydrogen bonding with A3 codon, whereas it maintains two Watson–Crick-like hydrogen bonds with the G3 codon, similar to the ones observed in the context of duplex. This behavior is opposite to the unmodified U34 base, which prefers A3 over G3. Specifically, for the first codon–anticodon pair (AAA–UUU), A3–U34 base pair forms two H-bonds that were dominantly observed through the course of the simulation.

However, introducing the $mnm^5\text{ges}^2\text{U}$ modification in the position 34 causes the A3–U34 base-pair to break, similar to their behavior in the duplex simulations. In the second case (AAG–UUU) with the wobble base pair G3–U34, primarily two H-bonds are maintained during the simulation. As discussed before, the introduction of $mnm^5\text{ges}^2\text{U}$ at position 34 changes the protonation state of atom N3 of U34. This change in protonation state promotes Watson–Crick-like hydrogen bonding pattern between the G–U base-pair, leading to a different set of D–H–A atoms (Supplementary Figure S30) and an increase in the probability of forming two steady H-bonds through most of the duration of the simulation. In addition, we find that the geranyl group can be accommodated at the binding site, where it has stacking/van der Waals interaction with the A4 codon and the residues G530 and C1054 of the ribosome, while the 5-*mnm* group interacts with the backbone phosphate group (Supplementary Figure S31). Furthermore, we also find that the ribosomal interactions, including A1492, A1493, G530 and C1054 with the mRNA and the tRNA are not disrupted during the course of the simulation (Supplementary Figure S32).

Therefore, the MD simulations collectively suggest possible binding modes for the geranyl-modified tRNAs at the wobble position, which might facilitate their selective codon discrimination. While our simulations show that the long hydrophobic geranyl group can possibly be accommodated at the binding site, a systematic free-energy calculation is needed to quantify the unfavorable free-energy associated

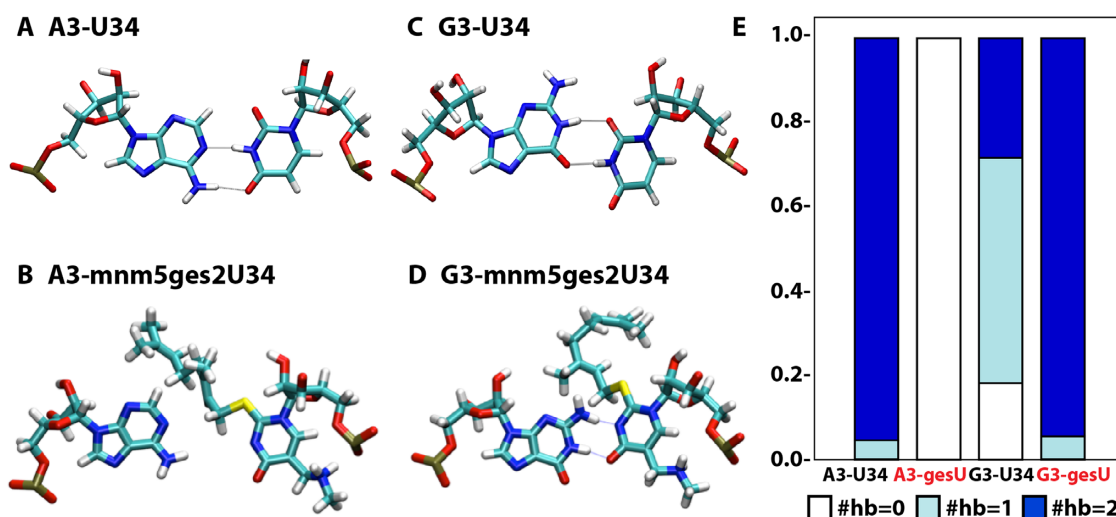


Figure 3. The (A–D) hydrogen bonding patterns for the the modified (mnm5ges2U₃₄) and unmodified (U₃₄) base with the A and G at the third position of the codon. (E) Comparison of the numbers of hydrogen bonds percentages [#hb (0,1,2)] for each of the base pairs during the simulation time.

with introducing the modification. It has been proposed that the geranyl-thio modification might be an intermediate in conversion of 2-thiouridine to 2-selenouridine (23,24), while it has also been shown to play a role in codon discrimination (17). The hydrophobic terpene group might also be specially evolved to fine-tune the efficiency and specificity of codon–anticodon interactions during the translation or play other important cellular functions. To further explore these hypotheses and to truly understand the biological role of this modification, we need to perform experimental binding studies of the tRNA to the ribosome and/or more extensive simulations (e.g. in the presence of a longer mRNA fragment with the 70S ribosome), as well as the complex structure studies, which are currently being pursued in our lab.

CONCLUSION

In this work we have synthesized geranylated RNA, studied their base pairing stability and specificity, and investigated their structural features in both duplexes and ribosome contexts. Our biophysical and molecular simulation studies collectively showed that this hydrophobic terpene group, a recently discovered natural RNA post-transcriptional modification, results in a much higher thermostability of U-G pair over the normal Watson–Crick U-A pair and other mispairs, indicating a strong base pairing specificity and discrimination between G and other nucleobases in both duplex structures and the codon–anticodon recognition in the presence of ribosome binding. The two hydrogen bonds formed between the geranylated 2-thiouridine and the guanosine base are dominantly stable in all the structural contexts we studied. This work provides the explanation to why there is better recognition of G-ending codes over the others by geranylated tRNAs. Besides enhancing the base pairing specificity, this bulky hydrophobic geranyl group has been speculated to play other functions, such as RNA localization and RNA transportation, as well as geranyl- and seleno-involved metabolic pathways. Further

detailed investigation of this unique natural modification in RNA will provide new insights into its current biological functions, and the potential evolutionary roles in lipid synthesis in the original stages of the RNA world.

SUPPLEMENTARY DATA

Supplementary Data are available at NAR Online.

ACKNOWLEDGEMENTS

We thank Leah Seebald and Prof. Daniele Fabris for the help in mass spec analysis. This work used computing resources from the Extreme Science and Engineering Discovery Environment (XSEDE).

FUNDING

University at Albany, State University of New York start-up fund (to J.S.); PhRMA Foundation (to A.C.); Simons Foundation Postdoctoral Fellowship [338863 to R.W.]. Funding for open access charge: University at Albany start-up fund. *Conflict of interest statement.* None declared.

REFERENCES

- Machnicka, M.A., Milanowska, K., Osman Oglou, O., Purta, E., Kurkowska, M., Olchowik, A., Januszewski, W., Kalinowski, S., Dunin-Horkawicz, S., Rother, K.M. *et al.* (2013) MODOMICS: a database of RNA modification pathways—2013 update. *Nucleic Acids Res.*, **41**, D262–D267.
- Cantara, W.A., Crain, P.F., Rozenski, J., McCloskey, J.A., Harris, K.A., Zhang, X., Vendeix, F.A., Fabris, D. and Agris, P.F. (2011) The RNA Modification Database, RNAMDB: 2011 update. *Nucleic Acids Res.*, **39**, D195–D201.
- Sample, P.J., Koreny, L., Paris, Z., Gaston, K.W., Rubio, M.A., Fleming, I.M., Hinger, S., Horakova, E., Limbach, P.A., Lukes, J. *et al.* (2015) A common tRNA modification at an unusual location: the discovery of wyosine biosynthesis in mitochondria. *Nucleic Acids Res.*, **43**, 4262–4273.
- Phizicky, E.M. and Hopper, A.K. (2010) tRNA biology charges to the front. *Genes Dev.*, **24**, 1832–1860.

5. Hopper, A.K. and Shaheen, H.H. (2008) A decade of surprises for tRNA nuclear-cytoplasmic dynamics. *Trends Cell Biol.*, **18**, 98–104.
6. Liu, N., Dai, Q., Zheng, G., He, C., Parisien, M. and Pan, T. (2015) N(6)-methyladenosine-dependent RNA structural switches regulate RNA-protein interactions. *Nature*, **518**, 560–564.
7. Joyce, G.F. (2002) The antiquity of RNA-based evolution. *Nature*, **418**, 214–221.
8. Iben, J.R. and Maraia, R.J. (2012) tRNAomics: tRNA gene copy number variation and codon use provide bioinformatic evidence of a new anticodon:codon wobble pair in a eukaryote. *RNA*, **18**, 1358–1372.
9. Graham, W.D., Barley-Maloney, L., Stark, C.J., Kaur, A., Stolarchuk, C., Sproat, B., Leszczynska, G., Malkiewicz, A., Safwat, N., Mucha, P. et al. (2011) Functional recognition of the modified human tRNA^{Lys3}(UUU) anticodon domain by HIV's nucleocapsid protein and a peptide mimic. *J. Mol. Biol.*, **410**, 698–715.
10. Bilbille, Y., Gustilo, E.M., Harris, K.A., Jones, C.N., Lusic, H., Kaiser, R.J., Delaney, M.O., Spremlji, L.L., Deiters, A. and Agris, P.F. (2011) The human mitochondrial tRNA^{Met}: structure/function relationship of a unique modification in the decoding of unconventional codons. *J. Mol. Biol.*, **406**, 257–274.
11. Sheppard, K., Akochy, P.M. and Soll, D. (2008) Assays for transfer RNA-dependent amino acid biosynthesis. *Methods*, **44**, 139–145.
12. Sanders, C.L., Lohr, K.J., Gambill, H.L., Curran, R.B. and Curran, J.F. (2008) Anticodon loop mutations perturb reading frame maintenance by the E site tRNA. *RNA*, **14**, 1874–1881.
13. Zhou, X.-L., Ruan, Z.-R., Wang, M., Fang, Z.-P., Wang, Y., Chen, Y., Liu, R.-J., Eriani, G. and Wang, E.-D. (2014) A minimalist mitochondrial threonyl-tRNA synthetase exhibits tRNA-isoacceptor specificity during proofreading. *Nucleic Acids Res.*, **42**, 13873–13886.
14. Hopper, A.K. (2013) Transfer RNA post-transcriptional processing, turnover, and subcellular dynamics in the yeast *Saccharomyces cerevisiae*. *Genetics*, **194**, 43–67.
15. Shepherd, J. and Ibbat, M. (2013) Direction of aminoacylated transfer RNAs into antibiotic synthesis and peptidoglycan-mediated antibiotic resistance. *FEBS Lett.*, **587**, 2895–2904.
16. Laxman, S., Sutter, B.M., Wu, X., Kumar, S., Guo, X., Trudgian, D.C., Mirzaei, H. and Tu, B.P. (2013) Sulfur amino acids regulate translational capacity and metabolic homeostasis through modulation of tRNA thiolation. *Cell*, **154**, 416–429.
17. Dumelin, C.E., Chen, Y., Leconte, A.M., Chen, Y.G. and Liu, D.R. (2012) Discovery and biological characterization of geranylated RNA in bacteria. *Nat. Chem. Biol.*, **8**, 913–919.
18. Veres, Z. and Stadtman, T.C. (1994) A purified selenophosphate-dependent enzyme from *Salmonella typhimurium* catalyzes the replacement of sulfur in 2-thiouridine residues in tRNAs with selenium. *Proc. Natl. Acad. Sci. U.S.A.*, **91**, 8092–8096.
19. Chen, P., Crain, P.F., Nasvall, S.J., Pomerantz, S.C. and Bjork, G.R. (2005) A “gain of function” mutation in a protein mediates production of novel modified nucleosides. *EMBO J.*, **24**, 1842–1851.
20. Wolfe, M.D., Ahmed, F., Lacourciere, G.M., Lahun, C.T., Stadtman, T.C. and Larson, T.J. (2004) Functional diversity of the rhodanese homology domain: the *Escherichia coli* ybbB gene encodes a selenophosphate-dependent tRNA 2-selenouridine synthase. *J. Biol. Chem.*, **279**, 1801–1809.
21. Ching, W.M., Wittwer, A.J., Tsai, L. and Stadtman, T.C. (1984) Distribution of two selenonucleosides among the selenium-containing tRNAs from *Methanococcus vannielii*. *Proc. Natl. Acad. Sci. U.S.A.*, **81**, 57–60.
22. Politino, M., Tsai, L., Veres, Z. and Stadtman, T.C. (1990) Biosynthesis of selenium-modified tRNAs in *Methanococcus vannielii*. *Proc. Natl. Acad. Sci. U.S.A.*, **87**, 6345–6348.
23. Bartos, P., Maciaszek, A., Rosinska, A., Sochacka, E. and Nawrot, B. (2014) Transformation of a wobble 2-thiouridine to 2-selenouridine via S-geranyl-2-thiouridine as a possible cellular pathway. *Bioorg. Chem.*, **56**, 49–53.
24. Jager, G., Chen, P. and Bjork, G.R. (2016) Transfer RNA bound to MnmH protein is enriched with geranylated tRNA—a possible intermediate in its selenation? *PLoS One*, **11**, e0153488.
25. Scott, A.I. (1997) How were porphyrins and lipids synthesized in the RNA world? *Tetrahedron Lett.*, **38**, 4961–4964.
26. Wang, R., Ranganathan, S.V., Basanta-Sanchez, M., Shen, F., Chen, A. and Sheng, J. (2015) Synthesis and base pairing studies of geranylated 2-thiothymidine, a natural variant of thymidine. *Chem. Commun.*, **51**, 16369–16372.
27. McDowell, J.A. and Turner, D.H. (1996) Investigation of the structural basis for thermodynamic stabilities of tandem GU mismatches: solution structure of (rGAGGUCUC)₂ by two-dimensional NMR and simulated annealing. *Biochemistry*, **35**, 14077–14089.
28. Cornell, W.D., Cieplak, P., Bayly, C.I., Gould, I.R., Merz, K.M., Ferguson, D.M., Spellmeyer, D.C., Fox, T., Caldwell, J.W. and Kollman, P.A. (1995) A second generation force field for the simulation of proteins, nucleic acids, and organic molecules. *J. Am. Chem. Soc.*, **117**, 5179–5197.
29. Dupradeau, F.Y., Pigache, A., Zaffran, T., Savineau, C., Lelong, R., Grivel, N., Lelong, D., Rosanski, W. and Cieplak, P. (2010) The R.E.D. tools: advances in RESP and ESP charge derivation and force field library building. *Phys. Chem. Chem. Phys.*, **12**, 7821–7839.
30. Cornell, W.D., Cieplak, P., Bayly, C.I. and Kollman, P.A. (1993) Application of RESP charges to calculate conformational energies, hydrogen bond energies, and free energies of solvation. *J. Am. Chem. Soc.*, **115**, 9620–9631.
31. Chen, A.A. and Garcia, A.E. (2013) High-resolution reversible folding of hyperstable RNA tetraloops using molecular dynamics simulations. *Proc. Natl. Acad. Sci. U.S.A.*, **110**, 16820–16825.
32. Chemical Computing Group Inc. (2016) Molecular operating environment (MOE).
33. Hess, B., Kutzner, C., van der Spoel, D. and Lindahl, E. (2008) GROMACS 4: algorithms for highly efficient, load-balanced, and scalable molecular simulation. *J. Chem. Theory Comput.*, **4**, 435–447.
34. Bussi, G., Donadio, D. and Parrinello, M. (2007) Canonical sampling through velocity rescaling. *J. Chem. Phys.*, **126**, 014101.
35. Berendsen, H.J.C., Postma, J.P.M., van Gunsteren, W.F., DiNola, A. and Haak, J.R. (1984) Molecular dynamics with coupling to an external bath. *J. Chem. Phys.*, **81**, 3684–3690.
36. Parrinello, M. and Rahman, A. (1981) Polymorphic transitions in single crystals: a new molecular dynamics method. *J. Appl. Phys.*, **52**, 7182–7190.
37. Darden, T., York, D. and Pedersen, L. (1993) Particle Mesh Ewald—an N·Log(N) method for Ewald sums in large systems. *J. Chem. Phys.*, **98**, 10089–10092.
38. Hess, B., Bekker, H., Berendsen, H.J.C. and Fraaije, J.G.E.M. (1997) LINCS: a linear constraint solver for molecular simulations. *J. Comput. Chem.*, **18**, 1463–1472.
39. Basanta-Sanchez, M., Temple, S., Ansari, S.A., D'Amico, A. and Agris, P.F. (2016) Attomole quantification and global profile of RNA modifications: Epitranscriptome of human neural stem cells. *Nucleic Acids Res.*, **44**, e26.
40. Hall, K.B. and McLaughlin, L.W. (1991) Thermodynamic and structural properties of pentamer DNA·DNA, RNA·RNA, and DNA·RNA duplexes of identical sequence. *Biochemistry*, **30**, 10606–10613.
41. Ratmeyer, L., Vinayak, R., Zhong, Y.Y., Zon, G. and Wilson, W.D. (1994) Sequence specific thermodynamic and structural properties for DNA·RNA duplexes. *Biochemistry*, **33**, 5298–5304.
42. Kumar, S., Reed, M.W., Gamper, H.B. Jr, Gorn, V.V., Lukhtanov, E.A., Foti, M., West, J., Meyer, R.B. Jr and Schweitzer, B.I. (1998) Solution structure of a highly stable DNA duplex conjugated to a minor groove binder. *Nucleic Acids Res.*, **26**, 831–838.
43. Murphy, F.V.T., Ramakrishnan, V., Malkiewicz, A. and Agris, P.F. (2004) The role of modifications in codon discrimination by tRNA(Lys)UUU. *Nat. Struct. Mol. Biol.*, **11**, 1186–1191.
44. Rozov, A., Demeshkina, N., Westhof, E., Yusupov, M. and Yusupova, G. (2015) Structural insights into the translational infidelity mechanism. *Nat. Commun.*, **6**, 7251.