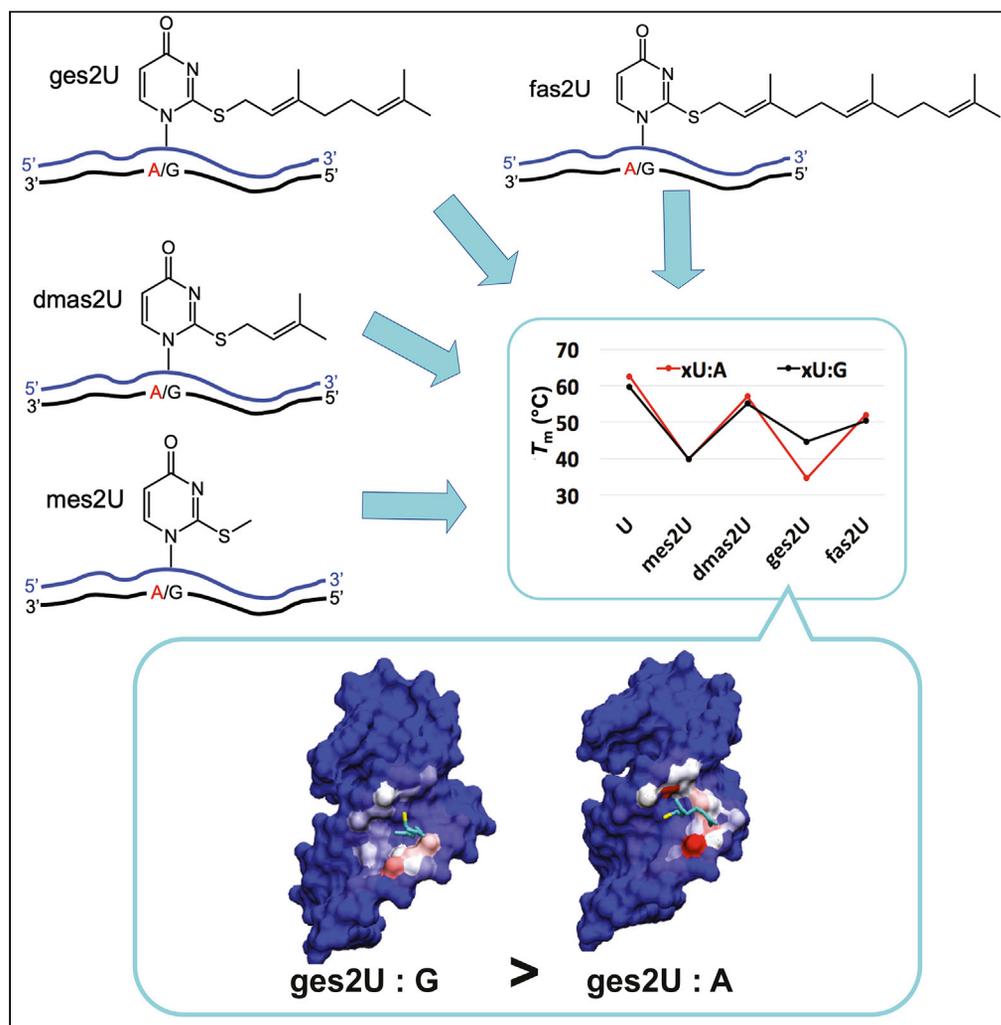


## Article

Terpene Chain Length Affects the Base Pairing Discrimination of *S*-geranyl-2-thiouridine in RNA Duplex

Phensinee Haruehanroengra, Sweta Vangaveti, Srivathsan V. Ranganathan, Song Mao, Max Daniel Su, Alan A. Chen, Jia Sheng

jsheng@albany.edu

## HIGHLIGHTS

Geranyl group on ges<sup>2</sup>U has perfect length to enhance binding to G over A in RNA

Other ges<sup>2</sup>U analogues has no pairing preference between G or A

MD shows the molecular dynamic behavior of different terpene chains in RNA duplex

## Article

Terpene Chain Length Affects the Base Pairing Discrimination of *S*-geranyl-2-thiouridine in RNA Duplex

Phensinee Haruehanroengra,<sup>1,2</sup> Sweta Vangaveti,<sup>2</sup> Srivathsan V. Ranganathan,<sup>2</sup> Song Mao,<sup>1,2</sup> Max Daniel Su,<sup>1,2</sup> Alan A. Chen,<sup>1,2</sup> and Jia Sheng<sup>1,2,3,\*</sup>

## SUMMARY

**Geranylation is a hydrophobic modification discovered in several bacteria tRNAs with the function of promoting codon bias during translation. However, why nature selects this C10-geranyl group remains a question. We conduct synthesis, UV-thermal denaturation, and molecular simulation studies in RNA duplexes and reveal possible reasons behind this natural selection. Among methyl-(C1), dimethylallyl-(C5), geranyl-(C10), and farnesyl-(C15) modified 2-thiouridines, only geranyl-group promotes U:G over U:A pair. Molecular simulation shows all the modified terpene groups point to the minor groove of RNA duplexes. The discrimination between U:G and U:A pairs derives from the difference in hydrogen bonding and interactions of the chain with the hydrophobic area in the minor groove. Geranyl group has perfect length to discriminate U:G and U:A pairs, whereas the others are either too long or too short to achieve the same behavior. This work indicates that geranyl group cannot be replaced by other terpene groups in promoting codon-specificity.**

## INTRODUCTION

Nucleic acids are naturally modified to diversify their structures and functions. There are more than 160 chemical modifications that have been reported in RNA systems. Among all the RNA species, tRNA contains the highest number (over 90) of modifications (Cantara et al., 2011; Machnicka et al., 2013). RNA modifications have been frontier research topics for their significance in exploring fundamental biochemistry and developing new therapeutics (Calore et al., 2015; Sobczak et al., 2003; Torres et al., 2014). Modifications on tRNA have been reported to be involved in many diseases including virus infections and cancers (Torres et al., 2014). They have been shown to regulate gene expression, fine-tune tRNA structure, enhance ribosomal binding, and affect the fidelity and efficiency of codon-anticodon recognition (Hou et al., 2015; Nachtergaele and He, 2017; Torres et al., 2014). The wobble position 34 or the first anti-codon position is frequently modified with a wide range of chemical groups that can affect the codon-anticodon interactions and fine-tune the translational efficiency and fidelity (Dumelin et al., 2012; Hou et al., 2015; Rozov et al., 2016; Urbonavicius et al., 2001). One of the common modification catalogs on uridine at wobble position is 2-thiouridine ( $s^2U$ , Figure 1(2)) derivatives including 5-methylaminomethyl-2-thiouridine ( $mnm^5s^2U$ ; Figure 1(3)) and 5-carboxymethylamino-methyl-2-thiouridine ( $cmnm^5s^2U$ ; Figure 1(4)). It has been shown that modifications on position 5 of uridine enhance the ribosomal binding and provide stronger codon recognition to both adenosine (A) and guanosine (G) (Hou et al., 2015; Rozov et al., 2016).

Thiolation at position 2 of uridine has been discovered on 16 out of 60 natural modified uridines (Cantara et al., 2011) and has been shown to increase the binding specificity of U to A while decreasing the recognition of G (Kumar and Davis, 1997; Larsen et al., 2015; Strobel and Cech, 1995). In addition, the sulfur atom at position 2 of uridine could be further replaced by the selenium atom, catalyzed by the selenouridine synthase (SelU, also called MnmH) (Veres and Stadtman, 1994). Interestingly, the same enzyme can also install geranyl group onto the sulfur atom at the position 2 using geranyl-pyrophosphate as the substrate (Figure 2) (Dumelin et al., 2012; Jager et al., 2016; Sierant et al., 2016; Wang et al., 2017). The geranyl group has been found in the wobble position ( $U_{34}$ ) of tRNAs specific for glutamate, glutamine, and lysine at the frequency of ~7% in many bacteria such as *Escherichia coli*, *Enterobacter aerogenes*, *Pseudomonas aeruginosa*, and *Salmonella enterica* var. *Typhimurium*. The geranylated-tRNA<sup>Lys</sup> has been reported to

<sup>1</sup>Department of Chemistry, University at Albany, State University of New York, 1400 Washington Avenue, Albany, NY 12222, USA

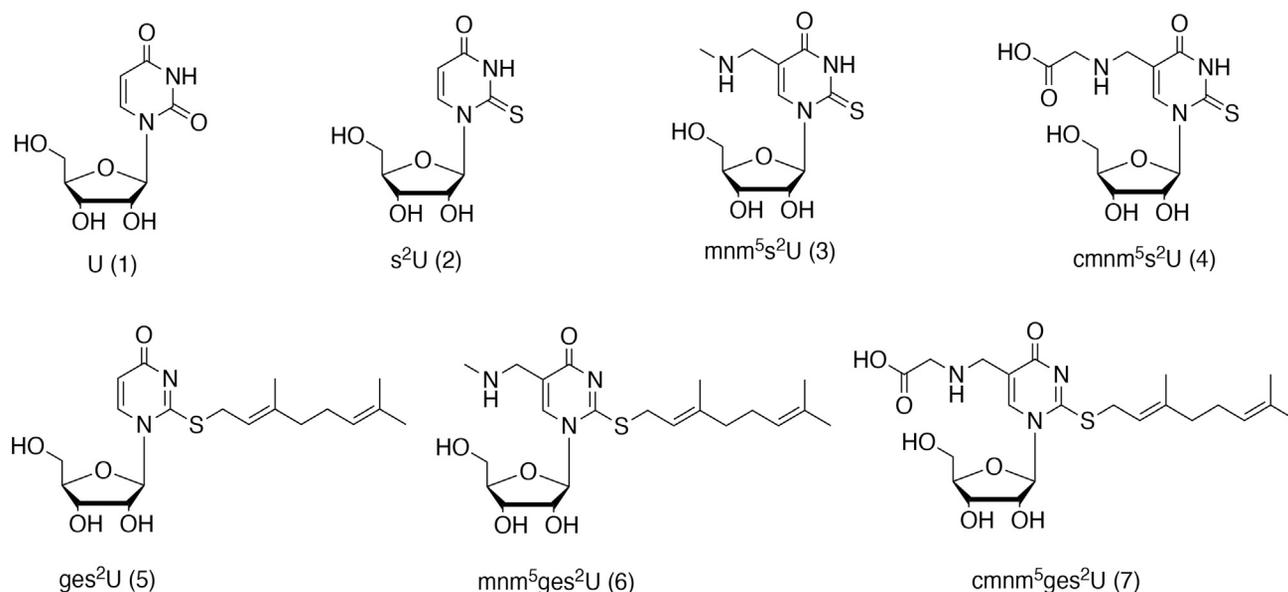
<sup>2</sup>The RNA Institute, University at Albany, State University of New York, Albany, NY 12222, USA

<sup>3</sup>Lead Contact

\*Correspondence: [jsheng@albany.edu](mailto:jsheng@albany.edu)

<https://doi.org/10.1016/j.isci.2020.101866>



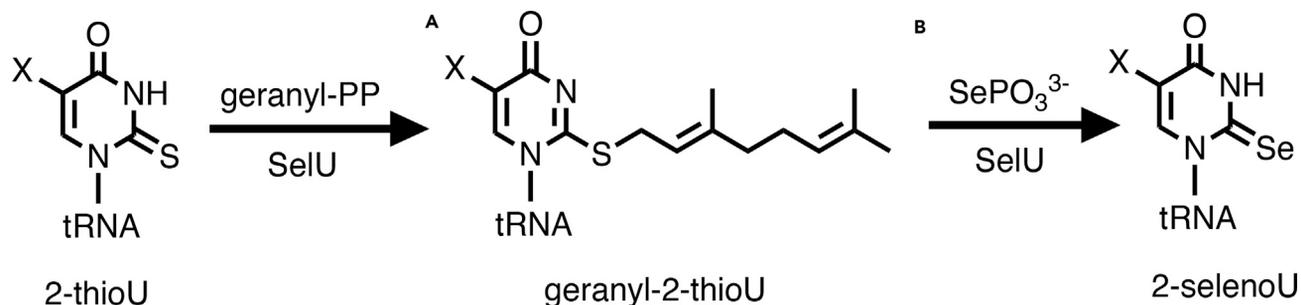


**Figure 1. Chemical Structure of Uridine Analogues**

Uridine (U, 1), 2-thiouridine ( $s^2U$ , 2), 5-methylaminomethyl-2-thiouridine ( $mnm^5s^2U$ , 3), 5-carboxymethylaminomethyl-2-thiouridine ( $cmnm^5s^2U$ , 4), geranylated 2-thiouridine ( $ges^2U$ , 5), 5-methylaminomethyl-S-geranyl-2-thiouridine ( $mnm^5ges^2U$ , 6), and 5-carboxymethylaminomethyl-S-geranyl-2-thiouridine ( $cmnm^5ges^2U$ , 7).

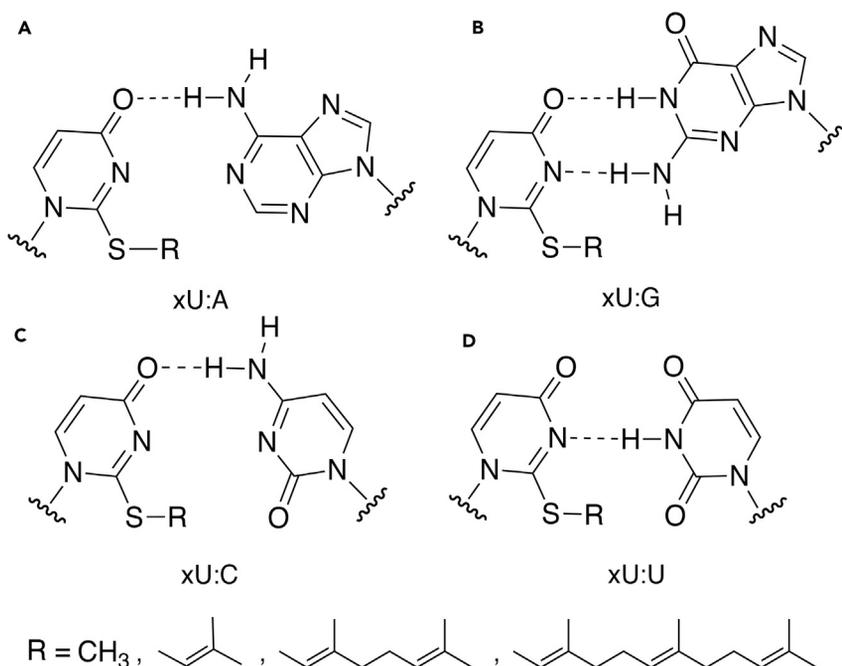
reduce  $-1$  frameshifting during the translation of *E. coli*, whereas geranylated-tRNA<sup>Glu</sup> promotes the codon bias of GAG to GAA (Dumelin et al., 2012). It is a fundamentally interesting question that how SelU conducts the bifunctions of both selenation and geranylation. It has been reported that the geranylation level was reduced as the level of selenium substrate exceeded 10 nM (Dumelin et al., 2012), which could be interpreted as geranyl-2-thiouridine being an intermediate of the transformation process from 2-thiouridine to 2-selenouridine (Figure 2) (Sierant et al., 2016, 2018). However, the geranylated uridine may play roles in certain biological processes such as tRNA localization, translational regulation, and cellular stress response, in addition to merely being the intermediate of selenouridine. We previously synthesized the geranyl-modified DNA and RNA oligonucleotides and studied their base pairing stability and specificity. The geranyl group was found to promote the T/U:G pairing over the others in both DNA ( $ges^2T$ ) (Wang et al., 2015) and RNA ( $ges^2U$ ) duplexes (Wang et al., 2016).

The addition of terpene group on sulfur atom at position 2 of uridine generates a new H-bonding pattern by switching the N3 from an H-bond donor (HD) to an acceptor (HA), resulting in the gain of second H-bond when pairing with guanosine (Figure 3B) and the loss of an H-bond when pairing with adenosine (Figure 3A).



**Figure 2. Proposed Catalytic Scheme of Selenouridine Synthase (SelU)**

Sulfur atom of 2-thiouridine (2-thioU) can be functionalized by SelU in the presence of geranyl pyrophosphate, resulting in the geranyl-2-thiouridine (A), which can be further converted into 2-selenouridine (B) with selenophosphate. The X group at position 5 represents either the mnm- or cmnm- shown in the compound 6 and 7 of Figure 1.

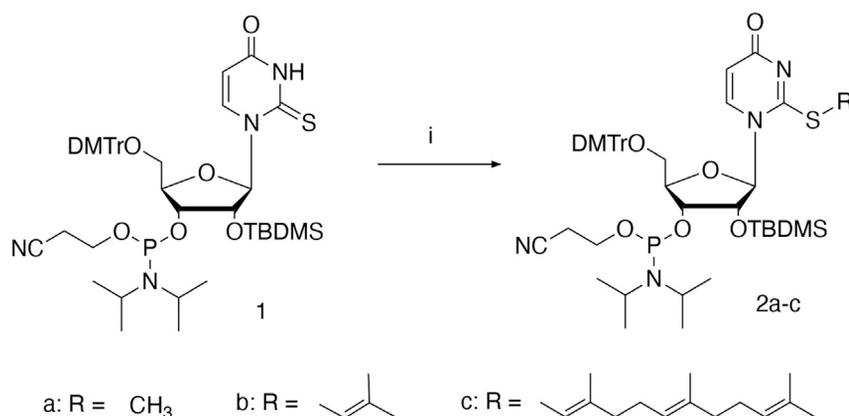


**Figure 3. Proposed Base Pairing Patterns**

The geranyl-2-thiouridine analogues (xU) with different length of carbon chains (R-group) pair with A, G, C, and U, respectively (A–D).

However, one would expect that irrespective of the length of the hydrophobic terpene group, the addition of just a methyl group at the same position (the sulfur atom of 2-thiouridine) would result in the same base pairing pattern, leading to the question, why does nature modify tRNA only with geranyl group when all the other possible terpene chains with different lengths are also available in cells? We have previously demonstrated the effect of terpene chain on the base pairing stability and specificity in the context of DNA duplexes containing geranyl-2-thiothymidine analogues (Haruehanroengra et al., 2017). In DNA-duplex, we found that at least a ten-carbon chain is necessary to achieve the base pairing specificity of xT:dG over xT:dA pairing. Both methyl and dimethylallyl groups at the same position showed no discrimination between these two pairs. Interestingly, farnesyl group showed similar specificity, but with slightly lower duplex stability compared with the geranyl one. Our molecular dynamic simulations showed that the xT:dA-containing duplexes are destabilized by the hydrophobic terpene groups. The longer the terpene modification, the more destabilization caused by the hydrophobic group by pushing the opposite nucleotides out of the duplex context, thereby weakening the stacking interactions and widening the DNA loop. In case of xT:dG-containing duplexes, however, both geranyl and farnesyl groups fit into the minor groove and compensate the overall stability loss of duplexes, whereas methyl and dimethylallyl groups are too short to interact with the hydrophobic area nor achieve the same effect.

Considering the structural differences between DNA and RNA duplexes, the geometries and interactions between the long terpene chains and RNA minor groove could be different from the DNA one. In this study, we applied the same set of terpene analogues, including methyl- (C1), dimethylallyl- (C5), geranyl- (C10), and farnesyl- (C15) modified 2-thiouridines into RNA duplexes, targeting to mimic the codon-anticodon interactions and further elucidate the significance of natural geranyl modification in promoting xU:G pair over xU:A in tRNA system. By annealing with the complementary RNA strands containing either canonical base pair or the mismatched non-canonical ones, we obtained the base pairing specificity and stability information of each modification via thermal denaturation studies. It is interesting to observe that the same hydrophobic terpene chain has different effects on duplex stability in the contexts of DNA and RNA. The geranyl group was shown to promote the pairing of ges<sup>2</sup>U:G over ges<sup>2</sup>U:A pair, which is consistent with DNA duplexes, whereas the base pairing discrimination between G and A in RNA context was not observed in farnesyl-modified RNA, unlike the case in DNA duplexes. We also performed molecular dynamic simulations to understand the molecular behavior of each hydrophobic terpene group in RNA duplexes



**Scheme 1. Synthesis of Geranyl-2-thiouridine Analogues Phosphoramidites**

(i) Methyl iodine, N,N-diisopropylethylamine (DIEA), methanol (MeOH), room temperature (a); dimethylallyl bromide, DIEA, MeOH, RT (b); farnesyl bromide, DIEA, MeOH, RT (c). The overall yields are above 90%.

containing xU:A or xU:G respectively and quantified their interactions and dynamics by employing the hydration patterns, time-series data of hydrogen bonding, base-stacking, and the locations of the modification.

## RESULTS

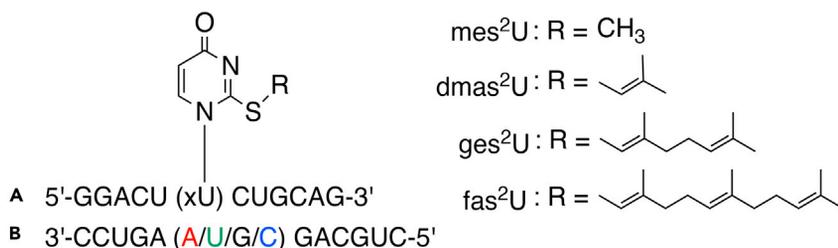
### Synthesis of S-geranyl-2-thiouridine Analogues and the RNA Oligonucleotides

As shown in [Scheme 1](#), the synthesis started from the commercially available 5'-DMTr-2'-TBDMS-2-thiouridine phosphoramidite. The terpene modification including methyl, dimethylallyl, and farnesyl-groups were functionalized at the 2-thio position on the uracil under basic condition to generate the key compound 2a-c in high yields. The <sup>31</sup>P-NMR spectra showed that the phosphoramidite group of the product 2a-c is quite stable under this reaction condition ([Figures S2, S5, and S8](#)). The modified phosphoramidites were further confirmed by <sup>1</sup>H-NMR and QTOF-MS ([Figures S1–S9](#)). The phosphoramidites were then incorporated into RNA oligonucleotides. The oligonucleotides were processed as described in [Transparent Methods](#) section and finally were verified by HR-MS. It has also been demonstrated previously that these terpene groups are stable with the standard reagents and conditions of solid-phase synthesis ([Haruehan-roengra et al., 2017](#)).

### Thermal Denaturation and Base Pairing Studies of RNA Duplexes Containing Modified Uridines

In this study, we used the 12mer RNA-duplex as the model, 5'-GGACU(xU)CUGCAG-3' and 5'-CUGCAG(Y)AGUCC-3', where X:Y pair is either canonical U:A or non-canonical U: C/G/U pairs ([Figure 4](#)), which is consistent to our previous study on geranyl-RNA ([Wang et al., 2016](#)). The UV-melting temperatures ( $T_m$ ) of xU-containing RNA duplexes were compared with the native counterparts ([Figure 5](#)). The corresponding melting temperatures and thermodynamic data are summarized in [Table 1](#). All the modifications at sulfur position on 2-thiouridine (xs<sup>2</sup>U) result in the decrease of  $T_m$  values, indicating that the addition of terpene groups on 2-thio position reduced the overall base pairing stability of the duplex, consistent with what we observed in the DNA context. However, the effect of terpene chain length on base pairing specificity seems quite different. Overall, studying the UV melting temperatures of RNA duplexes containing these uridine analogues (mes<sup>2</sup>U, dmas<sup>2</sup>U, ges<sup>2</sup>U, and fas<sup>2</sup>U) revealed the importance of geranyl group on base pairing selection in RNA system.

As shown in [Table 1](#), in the mes<sup>2</sup>U-containing duplexes, the replacement of oxygen at position 2 by methylated sulfur dramatically decreased melting temperatures ( $T_m$ ) of duplexes containing mes<sup>2</sup>U:A, mes<sup>2</sup>U:U, mes<sup>2</sup>U:G, and mes<sup>2</sup>U:C when compared with native U:A by 22.8°C, 12.9°C, 19.7°C, and 9.7°C, corresponding to an increased  $\Delta G^\circ$  of 7.5, 4.9, 7.0, and 3.3 kcal/mol, respectively. Similarly, in the dmas<sup>2</sup>U-containing duplexes, the  $T_m$  decreased by 5.5°C, 4.5°C, 6.0°C, and 5.1°C, with the increased  $\Delta G^\circ$  of 2.5, 2.0, 3.8, and 2.1 kcal/mol, respectively. For the farnesyl modification, the  $T_m$  of fas<sup>2</sup>U duplexes were lower than its native counterparts by 10.6°C, 11.2°C, 9.3°C, and 9.2°C, respectively, corresponding to an increased  $\Delta G^\circ$  of 5.7,



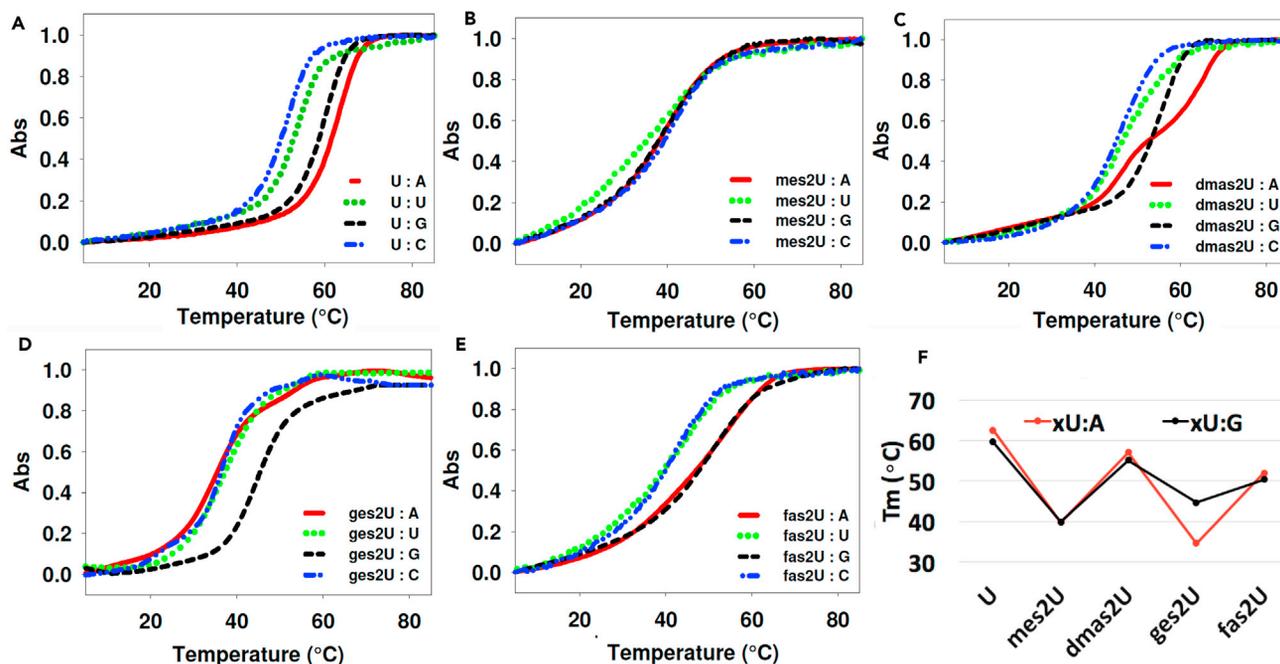
**Figure 4. Modified RNA Strand Containing S-geranyl-2-thiouridine Derivatives xU**

(A) The strand containing S-methyl-2-thiouridine (mes<sup>2</sup>U), S-dimethylallyl-2-thiouridine (dmas<sup>2</sup>U), S-geranyl-2-thiouridine (ges<sup>2</sup>U), and S-farnesyl-2-thiouridine (fas<sup>2</sup>U).

(B) The RNA complementary strand for thermal denaturation study.

4.7, 5.6, and 3.3 kcal/mol. If we compare the relative  $T_m$  of both U:A and U:G pairs containing duplexes with different lengths of modification, as shown in Figure 5F, while the RNA duplexes were stabilized by dimethylallyl group comparing to the methyl one, the stability dropped again when the terpene chain grew longer to C-10 geranyl group and recovered back with a C-15 farnesyl modification. It is also worthwhile to mention that the melting of dmas<sup>2</sup>U:A duplex showed a two-stage curve, indicating some specific interactions that this group may have with RNA duplexes to form certain stable transition intermediate structures during the duplex dissociation with heating.

We also reported the melting temperature of internal 2-thiouridine pairing with A/U/G/C (Figure S15). Our results agree with the earlier reports that 2-thiouridine stabilizes the pairing with A and U while slightly destabilizing the pairing with G (Kumar and Davis, 1997; Larsen et al., 2015), suggesting that the alteration of base pairing specificity observed earlier attributes to the addition of terpene groups on 2-thiouridine. The wobble position on tRNA is less discriminatory to achieve the degeneracy of the genetic code; however, it is naturally modified with the most diversity of chemical groups, which are known to regulate the translation efficiency. For example, s<sup>2</sup>U is commonly found on the wobble position of the anti-codon ending with U. It helps stabilizing the pairing of adenosine and destabilizing the pairing of guanosine, providing the codon



**Figure 5. Normalized UV-Melting Curves of RNA Duplexes**

(A–E) The duplex [5'-GGACU(xU)CUGCAG-3'] and [5'-CUGCAG(Y)AGUCC-3'] contain (xU) pairing with Y, with the xU representing (A) native uridine, (B) S-methyl-2-thiouridine (mes<sup>2</sup>U), (C) S-dimethylallyl-2-thiouridine (dmas<sup>2</sup>U), (D) S-geranyl-2-thiouridine (ges<sup>2</sup>U), and (E) S-farnesyl-2-thiouridine (fas<sup>2</sup>U).

(F) Summary of the melting temperatures. The curves represent the average data from four ramps of heating and cooling process.

| Entry | Base Pairs           |   | $T_m$ (°C) <sup>a</sup> | $\Delta T_m$ (°C) <sup>b</sup> | $-\Delta G_{37}^{\circ}$ (kcal/mol) <sup>c</sup> | $-\Delta H_{37}^{\circ}$ (kcal/mol) <sup>c</sup> | $-\Delta S_{37}^{\circ}$ (cal/K*mol) <sup>c</sup> |
|-------|----------------------|---|-------------------------|--------------------------------|--|--|---|
|       | X                    | Y |                         |                                |  |  |   |
| 1     | U                    | A | 62.5                    |                                | 16.6   | 107.6  | 293.3   |
| 2     | U                    | U | 53.3                    | -9.2                           | 14.0   | 112.8  | 318.6   |
| 3     | U                    | G | 59.6                    | -3.0                           | 16.0   | 111.8  | 308.9   |
| 4     | U                    | C | 50.9                    | -11.7                          | 12.6   | 100.3  | 282.8   |
| 5     | mes <sup>2</sup> U   | A | 39.7                    |                                | 9.1  | 48.8   | 127.8   |
| 6     | mes <sup>2</sup> U   | U | 40.4                    | 0.7                            | 9.1  | 45.8   | 118.4   |
| 7     | mes <sup>2</sup> U   | G | 39.9                    | 0.1                            | 9.1  | 41.6   | 104.5   |
| 8     | mes <sup>2</sup> U   | C | 41.2                    | 0.7                            | 9.4  | 58.8   | 159.5   |
| 9     | dmass <sup>2</sup> U | A | 57.0 <sup>c</sup>       |                                | 14.1   | 78.3   | 207.1   |
| 10    | dmass <sup>2</sup> U | U | 47.3                    | -9.8                           | 10.2   | 57.4   | 152.0   |
| 11    | dmass <sup>2</sup> U | G | 55.1                    | -1.9                           | 14.1   | 102.6  | 285.5   |
| 12    | dmass <sup>2</sup> U | C | 45.8                    | -11.2                          | 10.5   | 66.5   | 180.4   |
| 13    | ges <sup>2</sup> U   | A | 34.5                    |                                | 7.6  | 52.5   | 144.6   |
| 14    | ges <sup>2</sup> U   | U | 35.9                    | 1.4                            | 7.95   | 44.1   | 116.5   |
| 15    | ges <sup>2</sup> U   | G | 44.2                    | 9.7                            | 10.1   | 93.6   | 269.1   |
| 16    | ges <sup>2</sup> U   | C | 37.3                    | 2.9                            | 8.2  | 100.8  | 298.7   |
| 17    | fas <sup>2</sup> U   | A | 51.9                    |                                | 10.9   | 50.7   | 128.1   |
| 18    | fas <sup>2</sup> U   | U | 42.2                    | -9.8                           | 9.3  | 44.6   | 114.0   |
| 19    | fas <sup>2</sup> U   | G | 50.3                    | -1.6                           | 10.4   | 53.4   | 138.5   |
| 20    | fas <sup>2</sup> U   | C | 41.6                    | -10.3                          | 9.3  | 54.2   | 144.6   |

**Table 1. RNA Duplex Stability and Base Pairing Specificity**

The duplex contains methyl-2thiouridine (mes<sup>2</sup>U), dimethylallyl-2-thiouridine (dmass<sup>2</sup>U), and geranyl-2-thiouridine (ges<sup>2</sup>U) and farnesyl-2-thiouridine (fas<sup>2</sup>U), respectively in the context of 12nt-RNA: 5'-GGACU(xU)CUGCAG-3' and [5'-CUGCAG(Y)AGUCC-3'].

<sup>a</sup>The  $T_m$  were measured in 10 mM sodium phosphate buffer containing 100 mM NaCl (pH 7.0) and repeated four times.

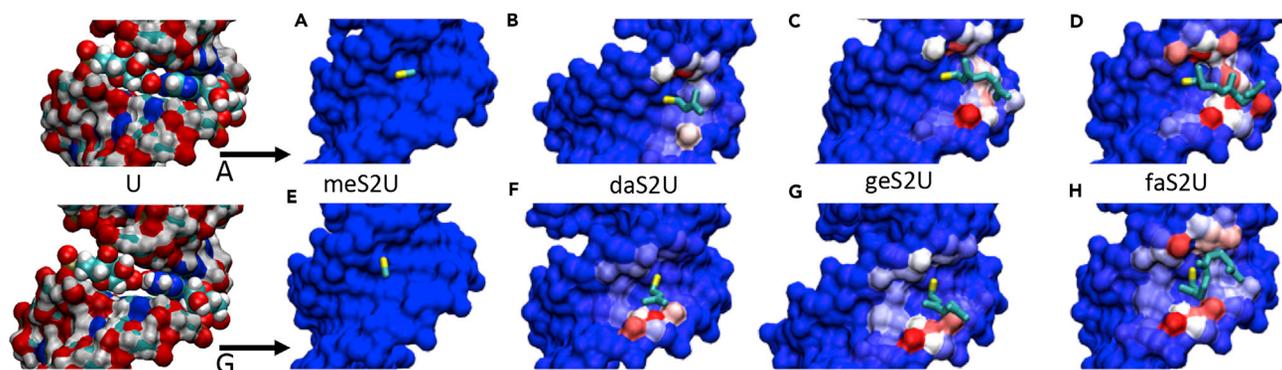
<sup>b</sup> $\Delta T_m$  values are relative to the duplex with xU-A pair of the same modification (x).

<sup>c</sup> $T_m$  was estimated from the average of two-stage melting temperatures.

bias during the translation process. Based on our study, the base pairing specificity can be switched to the opposite trend with the addition of geranyl group on sulfur atom of 2-thiouridine. In addition, we confirm that the terpene length of the geranyl is crucial for base pairing discrimination of ges<sup>2</sup>U:G over ges<sup>2</sup>U:A because other terpene groups (methyl, dimethylallyl, and farnesyl) could not provide such discrimination despite of the similarity of H-bond pattern. It is fascinating how nature selectively decorates the wobble position with various types of modifications as the tools to fine-tune the base recognition during translation. Moreover, the base pairing specificity studies of RNA with artificial groups such as methyl, dimethylallyl, geranyl, and farnesyl on 2-thiouridine can be applied to the development of DNA-/RNA-based therapeutics such as antisense and siRNA oligonucleotides (Reviews, 2007).

### Molecular Simulation

The molecular dynamics simulation studies were subsequently performed to study more mechanistic insights into these 2-thiouridine-bonded terpene chains with various lengths in RNA duplexes with either U:A or U:G pairs. Figure 6 shows the surface maps of the RNA colored based on their contacts with the modifications, thereby identifying their likely location in duplex context. Both of the canonical and modified bases are firmly held in place through stacking and base pairing interactions, and the hydrophobic



**Figure 6. Surface Representation of the RNA Duplexes**

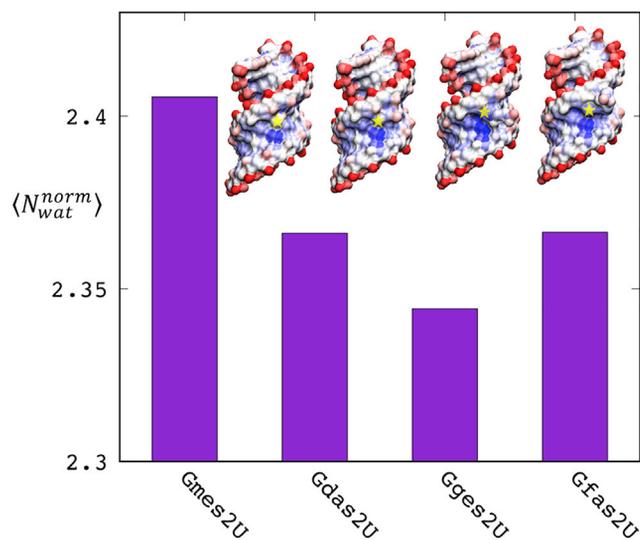
The strands are colored according to average number of contacts between the modification and the rest of the 12nt-RNA (low to high: blue to red) containing xU:A (A–D) and xU:G (E–H). The sulfur atom is shown in yellow and the hydrophobic chain of the modifications is shown in green.

chain behaves largely similarly when the opposite base is adenine or guanine. Unlike the DNA, the minor groove of the RNA is more hydrophilic due to the larger solvent accessible surface area and the presence of the polar 2' hydroxyl group, which is not accessible to the hydrophobic chain of the modification. Therefore, the modification is largely seen to interact with the backbone, including C5' and C4' atoms of the sugar. Although the contact with the duplex increases with chain length as expected, it is interesting to note that the modification in the A-paired system shows preferential interaction to the 3' side of the modified base, while being more promiscuous in the G-paired system (both the 5' and 3' sides).

Considering the importance of hydration (Siegfried et al., 2010) in the interaction pattern, we quantified the hydration change of RNA in the presence of different modifications by calculating the average number of water molecules in the first hydration shell (<4 Angstroms) of the RNA heavy atoms. The inset in Figure 7 shows the RNA heavy atoms colored with respect to the observed hydration behavior. As expected, the hydrophilic backbone shows stronger hydration (red to white) than the major and minor grooves (white to blue). Note that the major groove is not visible in this orientation. Interestingly, there is significant dehydration in the minor groove proximal to the local of the modification (highlighted by the stars), as shown by the distinct blue patch, which increases with the hydrophobic chain length, but saturates with the geranyl modification. To quantify this change, we calculated the average hydration for the whole RNA by averaging between all the heavy atoms. As expected, the hydration is non-monotonic and is the least for the geranyl containing RNA. This observation is also consistent with the preferential minor groove binding of the geranyl group driven by hydrophobicity.

We next calculated the average number of hydrogen bonds between uridine and the modified uridines with the opposite adenine and guanine nucleobases (Figure 8). Although the guanine maintains roughly two hydrogen bonds with uridine and the modified derivatives, which is expected from Figure 3, the bonding between the modified uridines and adenosine is observed to be very weak, with the number of hydrogen bonds dropping significantly to roughly 0.5 H-bond across the modifications. Interestingly, although the overall H-bonding behavior is similar between RNA and DNA duplexes, the stacking interactions are stronger in RNA. Figure 9 shows the probability of stacking distances between the adenine and guanine with their 5' and 3' neighbors. Similar to the DNA systems, in the case of guanine pairing, with the addition of the modification, the stacking interactions are strengthened toward the 5' side as evidenced by the decreased stacking distance and remains strengthened with subsequent increase in the chain length. The distances are unchanged for guanine in the 3' side. However, in the case of adenine, unlike DNA, the stacking interactions are preserved in all the cases, even in those where the hydrogen bonding interactions are significantly weakened. Even though the stacking distances increase for the A paired systems with both 5' and 3' bases, they still remain within stacking distance. These results indicate that, although the hydrogen bonding with adenine is weakened by the modifications, the stacking interactions is largely preserved, which is a stark difference compared with the DNA systems.

In addition to the conformational preference of the modification in the context of the duplex, we quantified its dynamics from the time-series data of the location of the modification. Specifically, after fitting the trajectory to the reference structure (by minimizing RMSD), we generated the time-series data for the location



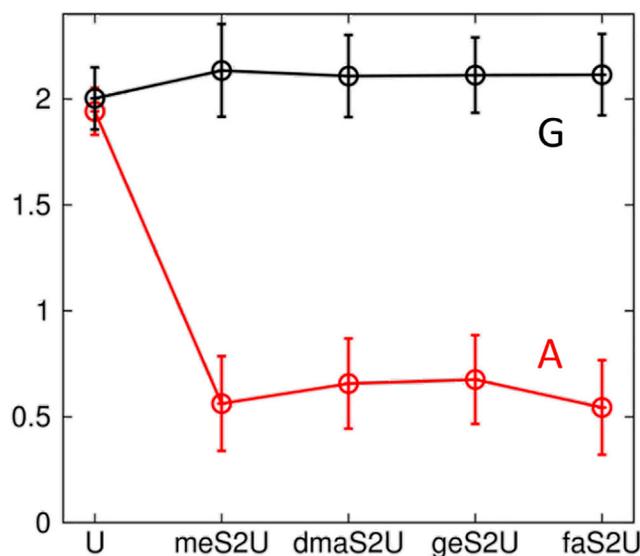
**Figure 7. Overall Hydration Index**

Quantified by average number of water molecules in the first hydration shell (<math><4</math> Angstroms) of the heavy atoms in the RNA. The inset shows the RNA duplex in surface form and colored by hydration index (from blue to white to red).

of the terminal terpene group, with the exception of the methyl modification, for which the whole methyl group was used. The mass center of the constituting carbon atoms was used to define the location of the terpene group. The time-series data are then used to calculate the auto-correlation function (ACF), using the following formula (Equation 1), where the  $\vec{x}$  vector denotes the time-dependent xyz coordinates of the modification, and  $\rho(t)$  is the time-series average.

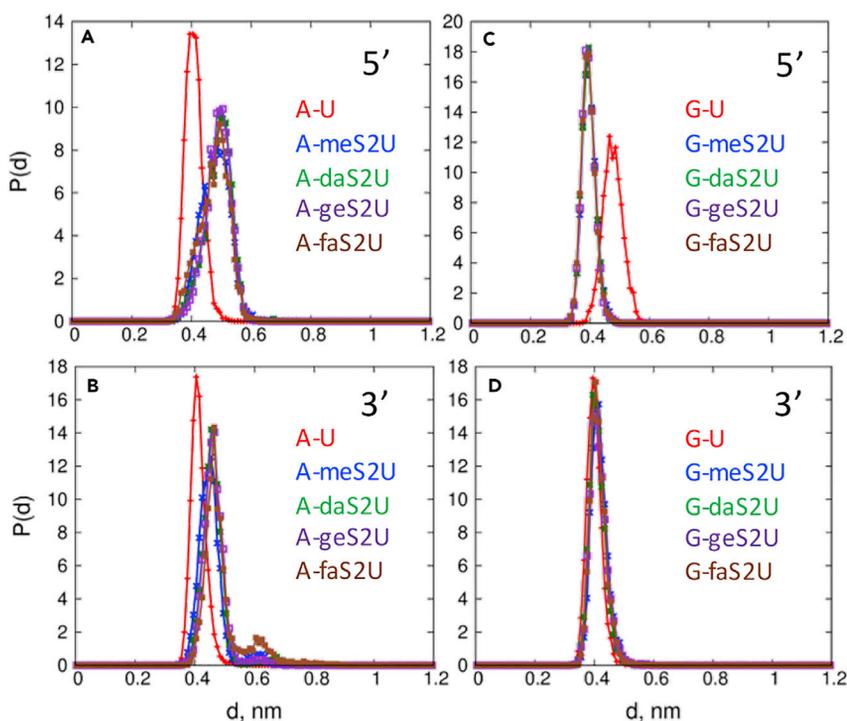
$$\rho(t) = \frac{\langle (\vec{x}(t) - \bar{x}) \cdot (\vec{x}(0) - \bar{x}) \rangle}{\langle |\vec{x}_i - \bar{x}|^2 \rangle} \quad (\text{Equation 1})$$

The ACF for all the systems with modifications are shown in Figure 10. For the A-paired systems, the dynamics of the modification is as reflected in the ACF follows a monotonic trend: faster to slower with the increase in the size of the modification. This trend is likely due to the increase in the preferential interaction



**Figure 8. Average Number of Hydrogen Bonds**

From the pairs of  $Xs^2U$  to G (black) or A (red) when X is terpene modification methyl (meS<sup>2</sup>U), dimethylallyl (dmaS<sup>2</sup>U), geranyl (geS<sup>2</sup>U), and farnesyl (faS<sup>2</sup>U).



**Figure 9. Probability Distributions of Stacking Distances between Adenine or Guanine and the Neighboring Base**

Defined by the geometric centers of the heavy atoms constituting the ring of the nucleobases.

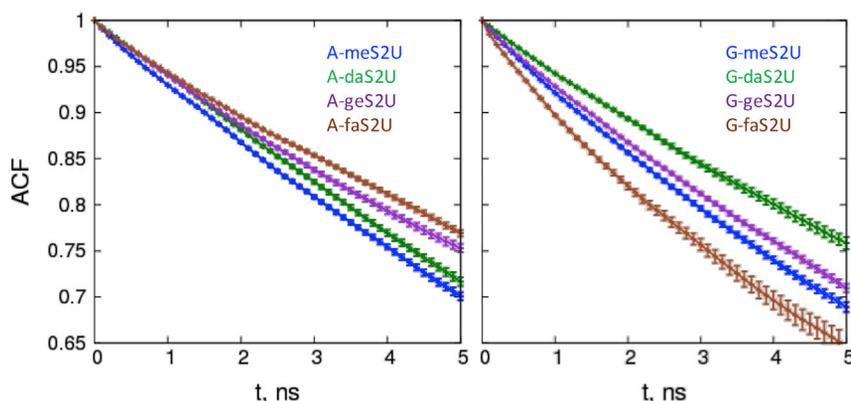
(A and B) Show the distributions for Adenine with the 5' and 3' neighbors, respectively.

(C and D) Show the same for Guanine.

of the modification to the 3' side with the increased chain length. However, for the G-paired systems, there is a reversal in this overall trend as seen in a non-monotonic behavior. The dimethylallyl modification possesses the longest correlation time, indicating a possible stronger preference for that group to interact with the duplex. With further increase in the modification chain length, the correlation time decreases, indicating faster dynamics. Although the contact maps show increase in interactions with chain length, the ACF curves for the G-paired systems indicate faster dynamics for larger chains, suggesting multiple modes (5' and 3' orientations) of interactions for the longer chains.

## DISCUSSION

Why does nature use the geranyl modification in the anti-codon stem-loop of the tRNA? Although this is a truly challenging question and needs more thorough biological investigations, from our experimental and simulation studies, we find some interesting difference in structure and stability between the various modifications even though they are expected to form the same base pairing pattern. We analyzed the effect of the terpene chain length on the base pairing discrimination especially between U:A and U:G pairs. It turns out that geranyl is the only group that can promote the binding difference between G and A within these terpene modifications. In our previous DNA duplex studies, we observed that both the geranylated and farnesylated 2-thiothymidines prefer to pair with G over A because the long terpene groups can further stabilize the DNA-duplexes by binding to the minor groove in DNA duplexes when they pair with G (Haruehanroengra et al., 2017). In the meantime, the fas<sup>2</sup>T:A pair was destabilized by the farnesyl group, generating the distinct discrimination between fas<sup>2</sup>T base pairing with G and A. However, this is not the case in RNA duplex. Although farnesyl group in fas<sup>2</sup>U:A could stabilize the overall RNA duplexes as shown in the increasing of melting temperature when compared with geranyl modification (Table 1, entry 13 and 17), this 15-carbon chain does not result in the base pairing discrimination between A and G in RNA system. The global shape of DNA or RNA duplexes should be the key factor in the behavioral change of the hydrophobic terpene groups. RNA duplexes are generally wider in diameter and shorter in length than DNA duplexes, providing the stronger van der Waals stacking force between nucleobases. Moreover, the addition of 2'-hydroxyl group of RNA provides a more hydrophilic environment in the minor groove of RNA duplexes.



**Figure 10. Autocorrelation Functions of the Position Vectors of the Modifications**

The position vector is defined as follows: methyl carbon for methyl  $\text{meS}^2\text{U}$  and the mass center of the terpene carbons for the other modifications.

Analyzing the molecular behavior of terpene chain provides the explanation to the stability and specificity studies. It is clear that the hydrophobic modifications were pointed into the minor groove for both A and G as the paired base albeit with different orientations. In the case of *S*-methyl-2-thiouridine-containing RNA duplexes, there is no significant interaction of methyl group in the minor groove (Figures 6A and 6E) due to its small size, suggesting the similar stability of  $\text{meS}^2\text{U}:\text{A}$  and  $\text{meS}^2\text{U}:\text{G}$  RNA duplexes. The increasing stability of RNA duplexes when the hydrophobic chain length increased from 1C to 5C (dimethylallyl group) is consistent with the enhanced interaction of the modification in the contact maps shown in Figures 6B and 6F. With an increase in the length of hydrophobic chains, as in the dimethylallyl group, there is enhanced interaction with the 4' and 5' carbons. The hydrophobic and van der Waals interactions between the dimethylallyl group and those hydrophobic regions of duplexes can enhance the overall stability of the RNA duplexes, consistent with the increase of the melting temperature when the terpene chain grew from 1C (methyl) to 5C (dimethylallyl). With further increase in chain length, there is a non-monotonic trend in the melting temperature, in contrast to the contact maps, which suggests stronger interaction for the geranyl group over the dimethylallyl group in both A-paired and G-paired systems. However, the stability of the farnesyl modification over the geranyl group is consistent with the enhanced interaction of the hydrophobic chain. The decrease in stability for the duplexes with geranyl group, especially for the A-paired system, might be a quirk of that chain length. We find that intrachain backbone contacts start to appear when the chain length increases from 5C to 10C, which might detract from the duplex-stabilizing interchain contacts. However, such intrachain contacts can coexist with the interchain ones with the longer chain, as in the case of the farnesyl group (15C). In addition, the interaction modes of the geranyl and farnesyl modifications are also different in the A-paired and G-paired systems. Although the hydrophobic chains interact preferentially with the 3' side in the A-paired system, the interaction is more promiscuous in the G-paired system. This behavior is also consistent with our hydration pattern studies and the faster dynamics exhibited by the modified groups in the G-paired systems. Therefore, we speculate that the promiscuous interaction of the geranyl group in combination with the balance of intra- and interchain contacts render the maximal differentiating capability in its pairing with A and G.

It is important to note that  $\text{dmas}^2\text{U}:\text{A}$  pair showed a two-states binding. At this point we could not provide a rational explanation of how the dimethylallyl imparts an intermediate state to the duplex. However, geranyl group is chemically less reactive compared with dimethylallyl that contains the allylic proton, which could be one of the reasons why geranyl group was evolutionary selected to promote wobble pairing to G. Collectively, geranyl group seems to be the right "sweet point" for introducing the pairing to G over the canonical base pairing A in  $\text{ges}^2\text{U}$ -RNA duplexes context. This fundamental structural study of terpene chain on the effect of base pairing specificity and stability provides not only the understanding of the naturally existed modification but also alternative modifications for future RNA therapeutic development.

### Limitations of the Study

The effect of terpene chain length observed in this study is in the context of RNA duplexes where the internal modified uridine is located at the center of an artificial 12-nt RNA strand. This study does not address

the effects of these geranyl-analogues in different RNA sequences. In nature, geranyl group presents at the wobble position of the tRNA anticodon stem loop where the interaction situation between the modified uridine and mRNA is different. The presence of surrounding ribosomal binding further complicates this interaction mode. Therefore, it will be very interesting to check the overall effect of geranyl analogues in the natural context, as well as to explore the impacts of these modifications on the translation efficiency, which could lead to potential therapeutic development. In addition, the molecular simulation (MD) studies had been performed with a lack of existing structural confirmation. Crystal or NMR structures will shed light on how these different length of terpene chains actually interact with the RNA duplexes and explain the two transitions state that has been observed in the case of dmas<sup>2</sup>U pairing to A.

## Conclusion

The systematic study on the effect of chain length of geranyl analogues on 2-thiouridine in the context of RNA duplexes revealed that geranyl group has the proper terpene chain length (C10) that promotes the binding of geranylated 2-thiouridine to guanosine (ges<sup>2</sup>U:G) over adenosine (ges<sup>2</sup>U:A) in RNA duplexes context. The hydrophobic modifications on 2-thiouridine theoretically altered the hydrogen bond pattern in the similar way despite of the length of terpene chains. Thermal denaturation study, however, revealed that methyl (C1), dimethylallyl (C5) and farnesyl (C15) showed no discrimination between G and A base pairing but geranyl (C10) group promoted the U:G over U:A pair. Interestingly, dimethylallyl presented two-states binding with adenosine pair (dmas<sup>2</sup>U:A) causing uncertainty on the base pairing preference. Therefore, among the four different chain lengths from C1 to C15, only geranyl group provided reliable base pairing preference of G over A. The molecular dynamic simulation provided detailed molecular behavior of the hydrophobic modification on 2-thiouridine. In RNA context, when the modified uridine pairs to G, two H-bonds were formed, whereas theoretical 0.5 H-bonds were formed when pairing with A. The modified uridines still align in the RNA duplex because of the base-stacking force. However, the difference in the orientation of the modified uracil base affected the possible interaction of the hydrophobic chain on the sulfur atom to the hydrophobic area of the minor groove. Collectively, this study indicates the importance of decorating the geranyl group on tRNA might stem from the length of the terpene chain to perform its function. It is significant for nature to selectively use geranyl modification to fine-tune and promote the base pairing specificity of geranyl-uridine to G over A and to achieve the biological significance such as reducing the -1 frameshifting. Moreover, geranylation in RNA could be a relic of the RNA world where this modification may have enhanced the structural and functional diversity of RNA prior to the emergence of proteins. More investigation is underway to understand if geranyl-modification on tRNA performs other significant biological functions in processes such as ribosomal binding, amino acylation, cellular stress responses, and RNA translocation.

## Resource Availability

### Lead Contact

Further information and requests for resources and reagents should be directed to and will be fulfilled by the Lead Contact, Jia Sheng ([jsheng@albany.edu](mailto:jsheng@albany.edu)).

### Materials Availability

There are no unique materials available.

### Data and Code Availability

The original NMR and MS spectra, HPLC profile, T<sub>m</sub> curves are available in the [Supplemental Information](#). The forcefield parameters generated for the modifications, as described in the [Methods](#) section, are provided in GROMACS format. Simulation data was generated and stored at the HPC hardware housed at the Data Center in UAlbany and are available for sharing upon request to the corresponding author.

## METHODS

All methods can be found in the accompanying [Transparent Methods supplemental file](#).

## SUPPLEMENTAL INFORMATION

Supplemental Information can be found online at <https://doi.org/10.1016/j.isci.2020.101866>.

## ACKNOWLEDGMENTS

This work was supported by The National Science Foundation (MCB-1715234 to J.S. and MRI grant 1726724) and University at Albany, State University of New York. A.A.C acknowledges support from NIH grant R35GM133469. We thank Cen Chen, Dr. Zhen Huang, Dr. Leah Seebald, Dr. Maksim Royzen, and Dr. Daniele Fabris for their help in mass spectrometry analysis. This work used computing resources from the Extreme Science and Engineering Discovery Environment (XSEDE).

## AUTHOR CONTRIBUTIONS

Conceptualization, J.S.; Methodology, P.H., S.V., S.V.R., and A.A.C.; Investigation, P.H., S.M., and M.D.S.; Writing—Original Draft, P.H., S.V.R., and J.S.; Writing—Review & Editing, P.H. and J.S.; Funding Acquisition, J.S.; Resources, J.S., S.V.R., and A.A.C.; Supervision, J.S.

## DECLARATION OF INTERESTS

All authors declare no conflict of interest.

Received: July 2, 2020

Revised: October 16, 2020

Accepted: November 20, 2020

Published: December 18, 2020

## REFERENCES

- Calore, M., De Windt, L.J., and Rampazzo, A. (2015). Genetics meets epigenetics: genetic variants that modulate noncoding RNA in cardiovascular diseases. *J. Mol. Cell. Cardiol.* 89 (Pt A), 27–34.
- 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.
- 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.
- Haruehanroengra, P., Vangaveti, S., Ranganathan, S.V., Wang, R., Chen, A., and Sheng, J. (2017). Nature's selection of geranyl group as a tRNA modification: the effects of chain length on base-pairing specificity. *ACS Chem. Biol.* 12, 1504–1513.
- Hou, Y.M., Gamper, H., and Yang, W. (2015). Post-transcriptional modifications to tRNA—a response to the genetic code degeneracy. *RNA* 21, 642–644.
- 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.
- Kumar, R.K., and Davis, D.R. (1997). Synthesis and studies on the effect of 2-thiouridine and 4-thiouridine on sugar conformation and RNA duplex stability. *Nucleic Acids Res.* 25, 1272–1280.
- Larsen, A.T., Fahrenbach, A.C., Sheng, J., Pian, J., and Szostak, J.W. (2015). Thermodynamic insights into 2-thiouridine-enhanced RNA hybridization. *Nucleic Acids Res.* 43, 7675–7687.
- 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.
- Nachtergaele, S., and He, C. (2017). The emerging biology of RNA post-transcriptional modifications. *RNA Biol.* 14, 156–163.
- Reviews, A. (2007). Annual review of biophysics and biomolecular structure. *Annu. Rev. Biophys. Biomol. Struct.* 36, 1–505.
- Rozov, A., Demeshkina, N., Khusainov, I., Westhof, E., Yusupov, M., and Yusupova, G. (2016). Novel base-pairing interactions at the tRNA wobble position crucial for accurate reading of the genetic code. *Nat. Commun.* 7, 10457.
- Siegfried, N., Kierzek, R., and Bevilacqua, P. (2010). Role of unsatisfied hydrogen bond acceptors in RNA energetics and specificity. *J. Am. Chem. Soc.* 132, 5342–5344.
- Sierant, M., Leszczynska, G., Sadowska, K., Komar, P., Radzikowska, C.E., Sochacka, E., and Nawrot, B. (2018). Escherichia coli tRNA 2-selenouridine synthase (SelU) converts S2U-RNA to Se2U-RNA via S-geranylated-intermediate. *FEBS Lett.* 592, 2248–2258.
- Sierant, M., Leszczynska, G., Sadowska, K., Dziergowska, A., Rozanski, M., Sochacka, E., and Nawrot, B. (2016). S-Geranyl-2-thiouridine wobble nucleosides of bacterial tRNAs; chemical and enzymatic synthesis of S-geranylated-RNAs and their physicochemical characterization. *Nucleic Acids Res.* 44, 10986–10998.
- Sobczak, K., de Mezer, M., Michlewski, G., Krol, J., and Krzyzosiak, W.J. (2003). RNA structure of trinucleotide repeats associated with human neurological diseases. *Nucleic Acids Res.* 31, 5469–5482.
- Strobel, S.A., and Cech, T.R. (1995). Minor groove recognition of the conserved G.U pair at the Tetrahymena ribozyme reaction site. *Science* 267, 675–679.
- Torres, A.G., Batlle, E., and Ribas de Pouplana, L. (2014). Role of tRNA modifications in human diseases. *Trends Mol. Med.* 20, 306–314.
- Urbonavicius, J., Qian, Q., Durand, J.M., Hagervall, T.G., and Bjork, G.R. (2001). Improvement of reading frame maintenance is a common function for several tRNA modifications. *EMBO J.* 20, 4863–4873.
- 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.
- Wang, R., Ranganathan, S.V., Basanta-Sanchez, M., Shen, F., Chen, A., and Sheng, J. (2015). Synthesis and base pairing studies of geranylated 2-thiouridine, a natural variant of thymidine. *Chem. Commun. (Camb.)* 51, 16369–16372.
- Wang, R., Vangaveti, S., Ranganathan, S.V., Basanta-Sanchez, M., Haruehanroengra, P., Chen, A., and Sheng, J. (2016). Synthesis, base pairing and structure studies of geranylated RNA. *Nucleic Acids Res.* 44, 6036–6045.
- Wang, R., Haruehanroengra, P., and Sheng, J. (2017). Synthesis of geranyl-2-thiouridine-modified RNA. *Curr. Protoc. Nucleic Acid Chem.* 68, 4.72.1–4.72.13.

**iScience, Volume 23**

**Supplemental Information**

**Terpene Chain Length Affects  
the Base Pairing Discrimination  
of S-geranyl-2-thiouridine in RNA Duplex**

**Phensinee Haruehanroengra, Sweta Vangaveti, Srivathsan V. Ranganathan, Song Mao, Max Daniel Su, Alan A. Chen, and Jia Sheng**

## Supporting Information

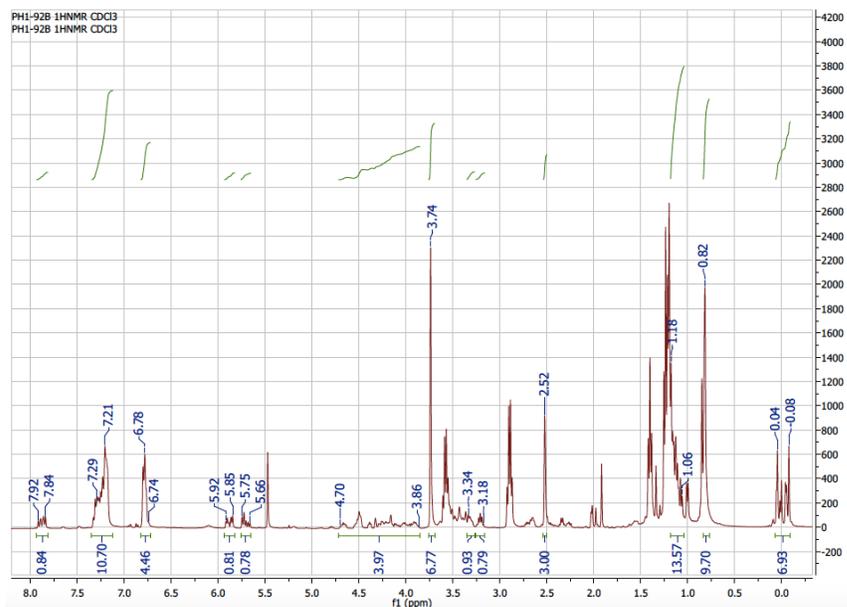


Figure S1.  $^1\text{H}$  NMR spectra of compound **2**. Related to Scheme **1**.

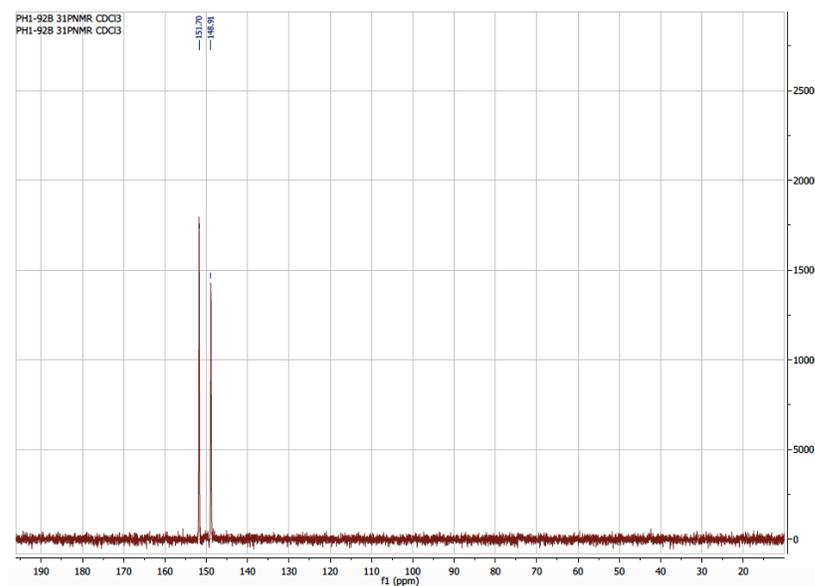


Figure S2.  $^{31}\text{P}$  NMR spectra of compound **2**. Related to Scheme **1**.

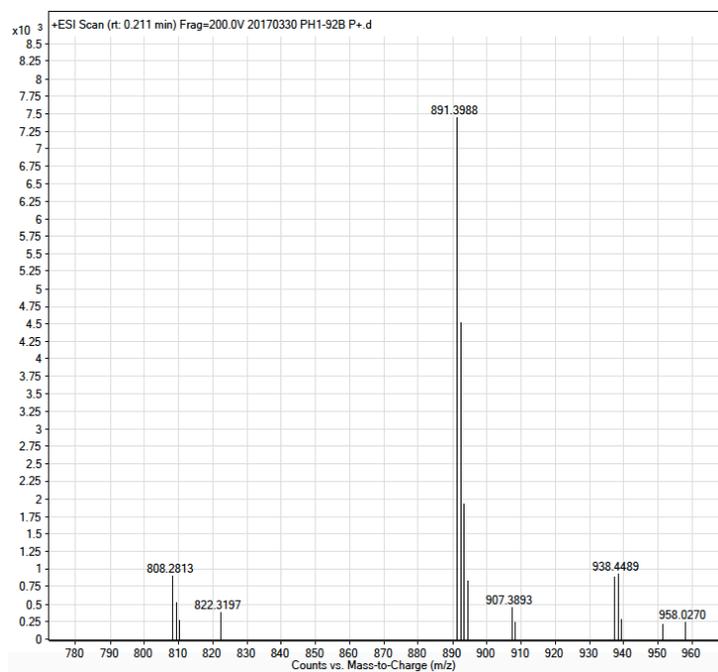


Figure S3. QTOF-MS spectra of compound 2. Related to Scheme 1.

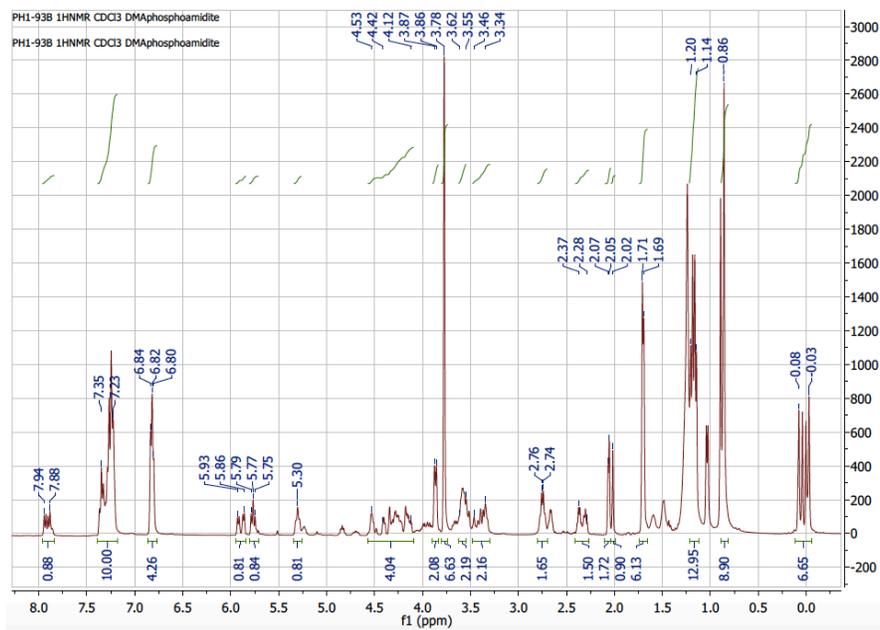


Figure S4. <sup>1</sup>H NMR spectra of compound 3. Related to Scheme 1.

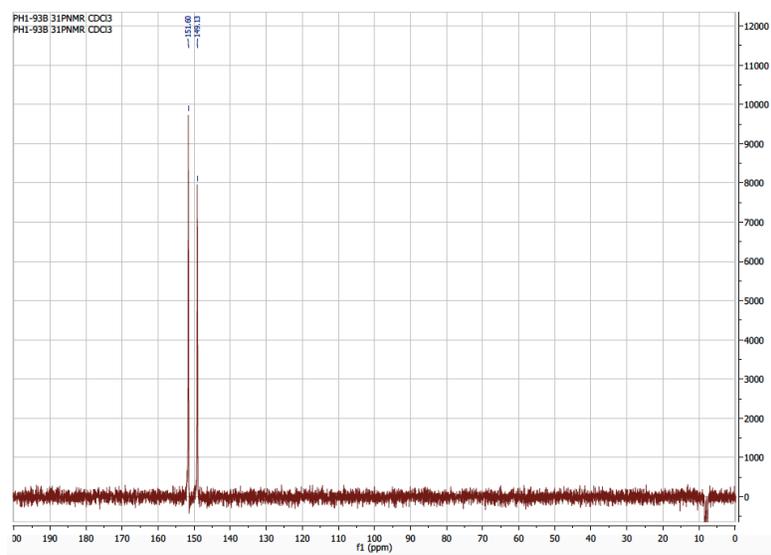


Figure S5.  $^{31}\text{P}$  NMR spectra of compound **3**. Related to Scheme 1.

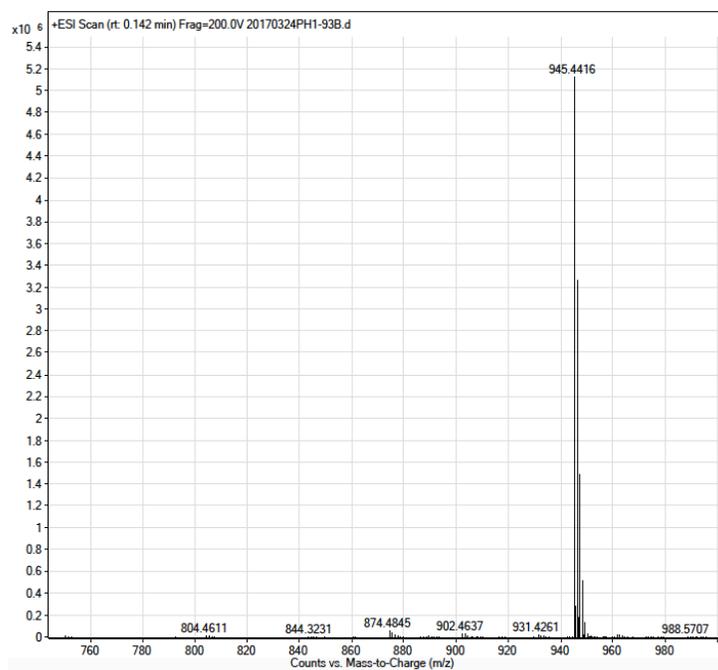


Figure S6. QTOF-MS spectra of compound **3**. Related to Scheme 1.

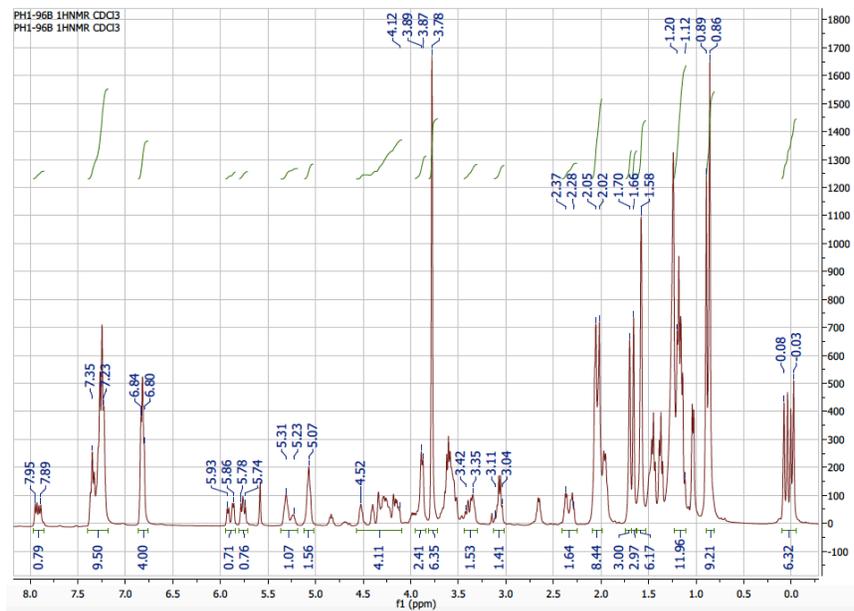


Figure S7.  $^1\text{H}$  NMR spectra of compound **4**. Related to Scheme 1.

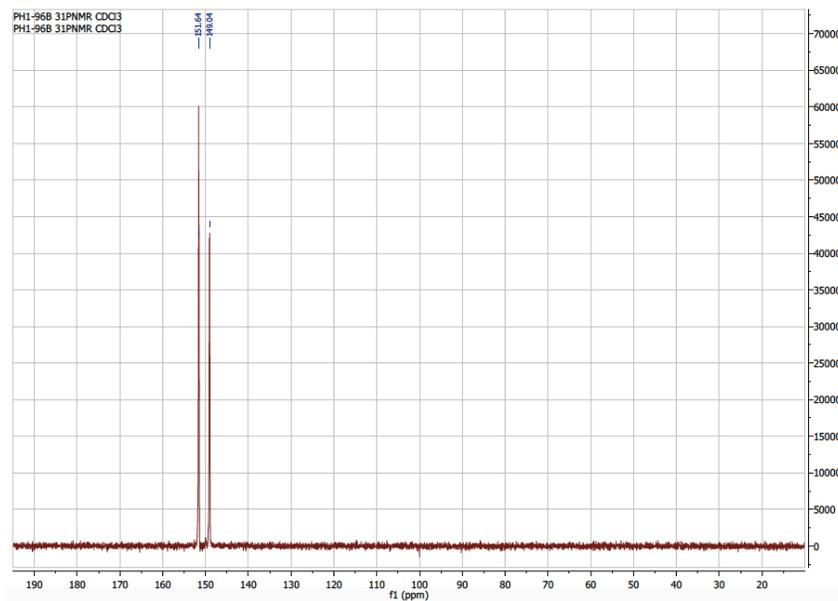


Figure S8.  $^{31}\text{P}$  NMR spectra of compound **4**. Related to Scheme 1.

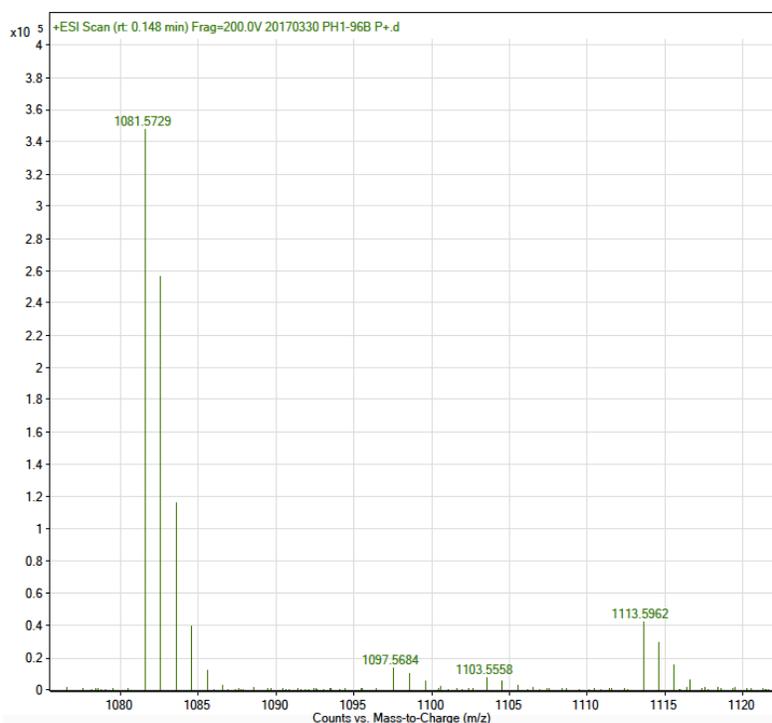


Figure S9. QTOF-MS spectra of compound **4**. Related to Scheme 1.

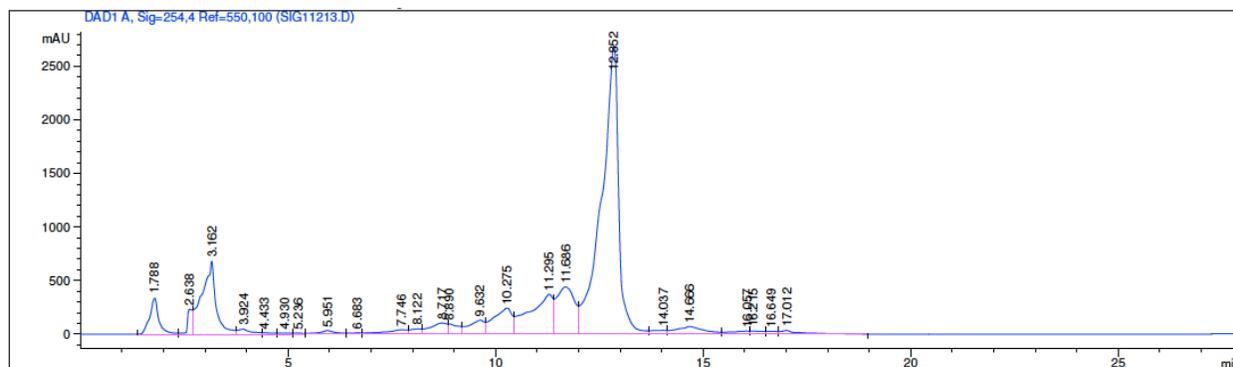


Figure S10. HPLC chromatogram of the crude RNA strand 5'-GGACU(mes<sup>2</sup>U)CUGCAG-3'. The profile was collected using Dionex DNASwift™ ion-exchange column (5x150 mm). Buffer A was DI-water and buffer B was 2M ammonium acetate buffer (pH 7.8). The RNA was eluted with a linear gradient of 0% to 70% buffer B over 20 min. Related to Figure 4.

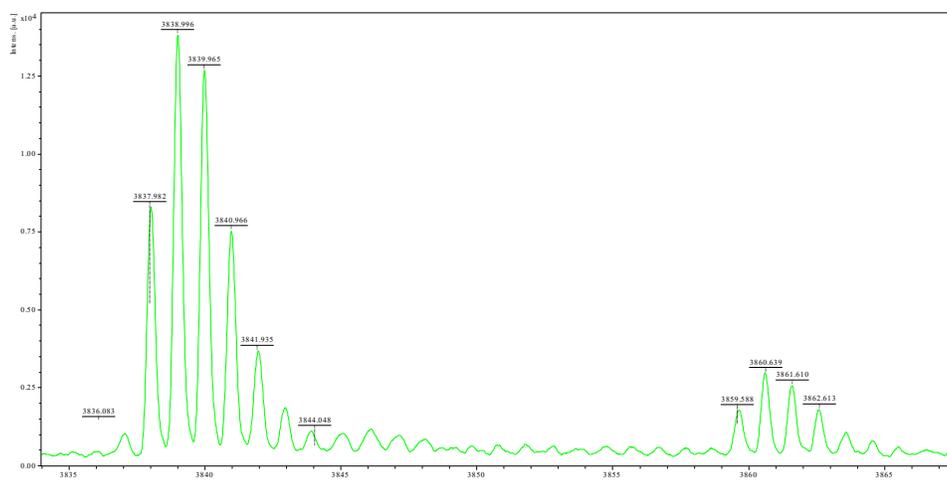


Figure S11. MALDI-MS spectra of 5'-GGACU(mes<sup>2</sup>U)CUGCAG-3' (observed mass: 3841.935, calculated mass: 3841.435). Related to Figure 4.

PH2-103-4\_P3 #1 RT: 0.01 AV: 1 NL: 1.13E4  
T: FTMS - p NSI sid=30.00 Full ms [400.00-1900.00]

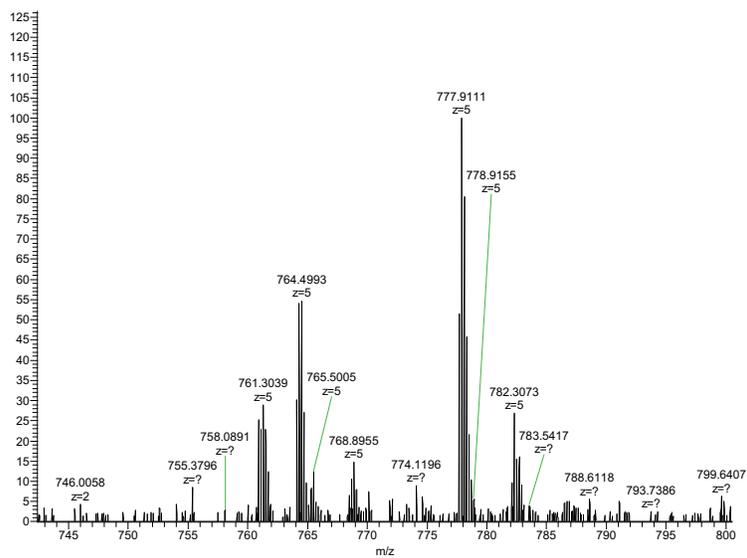


Figure S12. Orbitrap-ESI-MS spectra of 5'-GGACU(dmas<sup>2</sup>U)CUGCAG-3' (deconvoluted mass: 3894.555, calculated mass: 3895.527). Related to Figure 4.

PH2-103-7-6h\_P2 #1 RT: 0.01 AV: 1 NL: 6.60E3  
T: FTMS - p NSI sid=30.00 Full ms [400.00-1900.00]

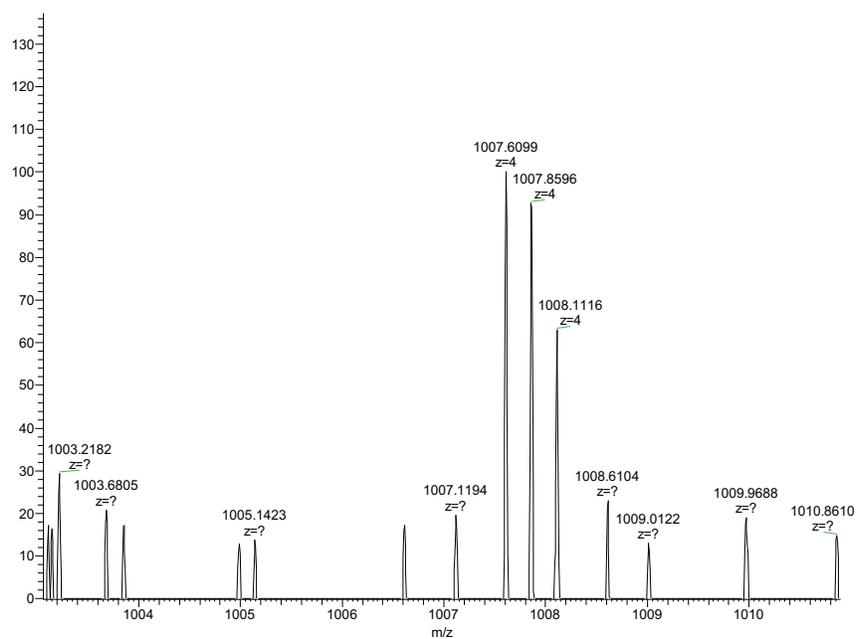


Figure S13. Orbitrap-ESI-MS spectra of 5'-GGACU(fas<sup>2</sup>U)CUGCAG-3' (deconvoluted mass: 4034.439, calculated mass: 4031.765). Related to Figure 4.

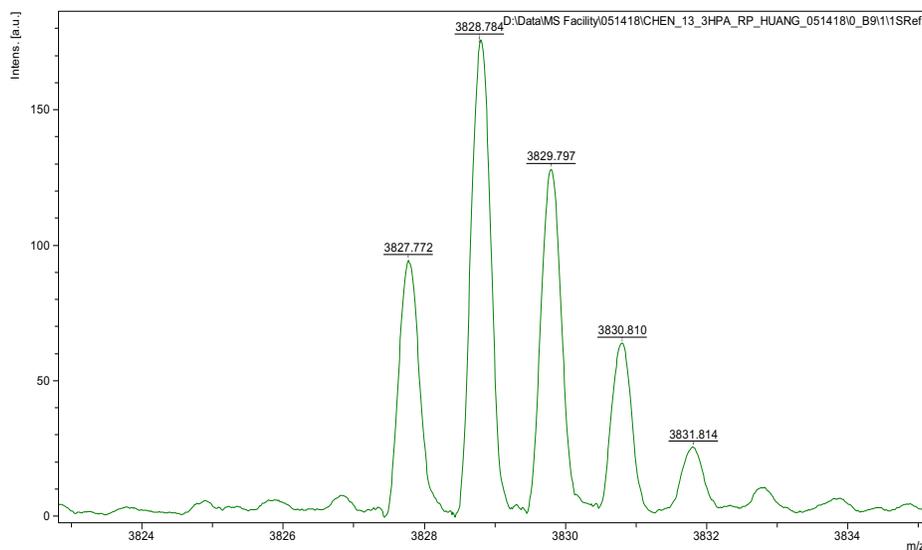


Figure S14. MALDI-MS spectra of 5'-GGACU(s<sub>2</sub>U)CUGCAG-3' (observed mass: 3828.784, calculated mass 3828.448). Related to Figure 4.

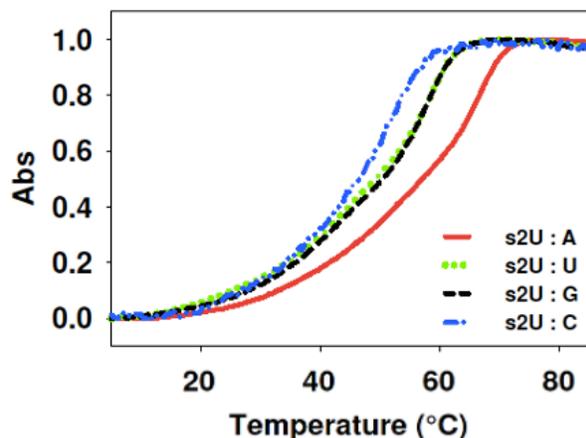


Figure S15. A normalized melting temperature curve ( $T_m$ ) of RNA duplexes, [5'-GGACU(s2U)CUGCAG-3'] and [5'-CUGCAG(Y)AGUCC-3'] where (xU) base pairs with Y. The (s2U) represents 2-thiouridine(A) and Y is A/U/G/C. Related to Figure 5.

Table S1. The ESI-HS-MS of RNA oligonucleotide containing geranyl analog. Related to Figure 4.

| Entry | Modification  | Sequence 5'→3'                   | Observed MS | Calculated MS |
|-------|---------------|----------------------------------|-------------|---------------|
| 1     | methyl        | GGACU(mes <sup>2</sup> U)CUGCAG  | 3841.935    | 3841.435      |
| 2     | dimethylallyl | GGACU(dmas <sup>2</sup> U)CUGCAG | 3894.555    | 3895.527      |
| 3     | farnesyl      | GGACU(fas <sup>2</sup> U)CUGCAG  | 4034.439    | 4031.765      |
| 4     | 2-thiouridine | GGACU(s <sup>2</sup> U)CUGCAG    | 3828.784    | 3828.448      |

Table S2. RNA duplex stability and base pairing specificity of 2-thiouridine (s2U) in the context of 12nt-RNA: 5'-GGACU(s<sup>2</sup>U)CUGCAG-3'] complementary with [5'-CUGCAG(A/U/G/C)AGUCC-3']. Related to Figure 5.

| Entry | Base pairs |   | $T_m$ (°C) <sup>a</sup> | $\Delta T_m$ (°C) <sup>b</sup> | $-\Delta G_{37}^\circ$ (kcal/mol) <sup>c</sup> | $-\Delta H_{37}^\circ$ (kcal/mol) <sup>c</sup> | $-\Delta S_{37}^\circ$ (cal/K*mol) <sup>c</sup> |
|-------|------------|---|-------------------------|--------------------------------|--|--|---|
|       | X          | Y |                         |                                |  |  |   |
| 1     | s2U        | A | 63.49                   |                                | 15.5   | 91.14  | 243.89  |
| 2     | s2U        | U | 55.73                   | -7.76                          | 11.15  | 50.31  | 126.28  |
| 3     | s2U        | G | 55.69                   | -7.80                          | 11.21  | 50.54  | 126.78  |
| 4     | s2U        | C | 48.65                   | -14.84                         | 10.14  | 51.1   | 131.75  |

## TRANSPARENT METHODS

### Materials and general procedures of synthesis

#### Synthesis and Base Pairing Studies of hydrophobic modified S2U-RNA

Anhydrous solvents were used and redistilled using standard procedures. All solid reagents were dried under a high vacuum line prior to use. Air sensitive reactions were carried out under argon protection. RNase-free water, tips and tubes were used for RNA purification. The 2-thio-5'-DMTr-2'-TBDMS-3'-phosphoramidite uridine was customized from Chemgene. Analytical TLC plates pre-coated with silica gel F<sub>254</sub> (Dynamic Adsorbents) were used for monitoring reactions and visualized by short wavelength UV light (254 nm). Flash column chromatography was performed using silica gel (32-63  $\mu$ m). All <sup>1</sup>H and <sup>31</sup>P NMR spectra were recorded on a Bruker 400 spectrometer. Chemical shift values are in ppm. High-resolution MS were achieved by ESI-Orbitrap or ESI-QTOF mass spectrometry at University at Albany, SUNY.

#### Synthesis of methylated-2-thiouridine phosphoramidite

##### Methyl-2-thio-5'-DMTr-2'-TBDMS Uridine 3'phosphoramidite (**2a**)

Under Argon atmosphere, the 5'-DMTr-2'-TBDMS-2-thiouridine phosphoramidite (**1**) (41mg, 0.047 mmol) was dissolved in 3mL of methanol. The mixture of 5 eq of diisopropylethylamine and 3 eq of iodomethane was added into the starting material flask. The reaction was stirred at room temperature for overnight. The completion of the methylation was monitored by QTOF-MS. The reaction product was evaporated using rotary evaporator and further purification by column chromatography. The elution solvent is 100% dichloromethane yield 95% of Methyl-2-thio-5'-DMTr-2'-TBDMS uridine 3'phosphoramidite (**2a**).

<sup>1</sup>H NMR (400 MHz, CDCl<sub>3</sub>)  $\delta$  7.92-7.84 (m, 1H), 7.29-7.21 (m, 10H), 6.78-6.74 (m, 4H), 5.92-5.85 (m, 1H), 5.75-5.66 (m, 1H), 4.70-3.86 (m, 4H), 3.74 (s, 6H), 3.34-3.18 (m, 2H), 2.52 (s, 3H: S-Me), 1.18-1.06 (m, 12H), 0.82 (s, 9H), 0.04-0.08 (m, 6H)

<sup>31</sup>P NMR (100MHz, CDCl<sub>3</sub>)  $\delta$  151.70, 148.91

QTOF-MS [M+H]<sup>+</sup>: 891.3988 (calc. 891.3946).

#### Synthesis of dimethylallylated-2-thiouridine phosphoramidite

##### Dimethylallyl-2-thio-5'-DMTr-2'-TBDMS Uridine 3'phosphoramidite (**2b**)

Under Argon atmosphere, the 5'-DMTr-2'-TBDMS-2-thiouridine phosphoramidite (**1**) (40mg, 0.0684 mmol) was dissolved in 3mL of methanol. The mixture of 10 eq of diisopropylethylamine and 3 eq of dimethylallyl bromide was added into the starting material flask. The reaction was stirred at room temperature for overnight. The completion of the addition reaction was monitored by QTOF-MS. The reaction product was evaporated using rotary evaporator and further purification by column chromatography with elution solvent of 100% dichloromethane. Yield 95% of dimethylallyl-2-thio-5'-DMTr-2'-TBDMS Uridine 3'phosphoramidite (**2b**).

<sup>1</sup>H NMR (400 MHz, CDCl<sub>3</sub>) δ 7.94-7.88 (m, 1H), 7.35-7.23 (m, 10H), 6.84-6.80 (m, 4H), 5.93-5.86 (m, 1H), 5.79-5.75 (t, 1H), 5.30 (m, 1H), 4.53-4.12 (m, 4H), 3.87-3.86 (d, 2H), 3.78 (s, 6H), 3.62-3.55 (m, 2H), 3.46-3.34 (m, 2H), 2.76-2.74 (m, 2H), 2.37-2.28 (m, 1H), 2.07-2.05 (m, 2H), 2.02 (s, 1H), 1.71-1.69 (m, 6H), 1.20-1.14 (m, 11H), 0.86 (s, 9H), 0.08-0.03 (m, 6H)

<sup>31</sup>P NMR (100MHz, CDCl<sub>3</sub>) δ 151.60, 149.13

QTOF-MS [M+H]<sup>+</sup>: 945.4416 (calc. 945.4416)

### Synthesis of farnesylated-2thiouridine phosphoramidite

#### Farnesyl-2-thio-5'-DMTr-2'-TBDMS Uridine 3'phosphoramidite (**2c**)

Under Argon atmosphere, the 5'-DMTr-2'-TBDMS-2-thiouridine phosphoramidite (**1**) (40mg, 0.0684 mmol) was dissolved in 3mL of methanol. The mixture of 3 eq of diisopropylethylamine and 1.5 eq of farnesyl bromide was added into the starting material flask. The reaction was stirred at room temperature for overnight. The completion of the reaction was monitored by QTOF-MS. The reaction product was evaporated using rotary evaporator and further purification by column chromatography with elution solvent of 100% dichloromethane. Yield 96% of Farnesyl-2-thio-5'-DMTr-2'-TBDMS Uridine 3'phosphoramidite (**2c**).

<sup>1</sup>H NMR (400 MHz, CDCl<sub>3</sub>) δ 7.95-7.89 (m, 1H), 7.35-7.23 (m, 10H), 6.84-6.80 (m, 4H), 5.93-5.86 (m, 1H), 5.78-5.74 (t, 1H), 5.31-5.23 (m, 1H), 5.07 (s, 2H), 4.52-4.12 (m, 4H), 3.89-3.87 (d, 2H), 3.78 (s, 6H), 3.42-3.35 (m, 2H), 3.11-3.04 (m, 2H), 2.37-2.28 (m, 2H), 2.05-2.02 (d, 8H), 1.70 (s, 3H), 1.66 (s, 3H), 1.58 (s, 3H), 1.20-1.12 (m, 12H), 0.86 (s, 9H), 0.08-0.03 (m, 6H)

<sup>31</sup>P NMR (100MHz, CDCl<sub>3</sub>) δ 151.64, 149.04

QTOF-MS [M+H]<sup>+</sup>: 1081.5729 (calc. 1081.5668)

### Synthesis of RNA oligonucleotides

The RNA oligonucleotides were chemically synthesized at 1.0 μmole scale using an Oligo-800 DNA/RNA synthesizer. The modified-uridine phosphoramidites (Scheme 1, 2a-c) were dissolved in dichloromethane to a concentration of 0.07 M. The trichloroacetic acid (3%) in dichloromethane was used for detritylation. The 5-ethylthio-1H-tetrazole solution (0.25M) in acetonitrile was utilized for a 12 min-coupling. The unreacted 5'-OH was capped using CapA solution (80% tetrahydrofuran / 10% acetic anhydride / 10% 2,6-Lutidine) and CapB solution (16% N-methylimidazole in THF). The oxidation on phosphate backbone was carried out by iodine in pyridine, THF and water. All the reagents are oligo-synthesizer standard solution. The synthesis was performed on DMT-off mode. The modified-oligonucleotides were subsequently cleaved using concentrated ammonium solution at room temperature for 6 hours and the protecting groups on 2'-OH was removed by HF treatment. After ethanol precipitation, the RNA-pellet was resuspended in RNA-free water and desalted using Sep-Pak SPE from Waters.

### **HPLC analysis and purification**

The resulting oligonucleotides were then purified by HPLC using Dionex DNASwift™ ion-exchange column (5x150 mm). Buffer A was DI-water and buffer B was 2M ammonium acetate buffer (pH 7.8). The RNA was eluted with a linear gradient of 0% to 70% buffer B over 20 min. A typical HPLC profile of the crude RNA strands was shown in Figure S10. The collected fractions were lyophilized, desalted using Water Sep-Pac C18 columns and concentrated using a DNA speed vac concentrator. The purified oligonucleotides were confirmed by high resolution ESI mass-spectrometry analysis (Table S1 and Figure S11-14).

### **Thermal denaturation and base pairing studies of xU-containing RNA duplexes**

The base pairing stability was evaluated by the measurement of melting temperature ( $T_m$ ) between the purified oligonucleotide containing the modified nucleotide (xU) and its complementary strand with canonical (A) and other non-canonical base pairs (U/G/C). In this study, twelve nucleotides long duplexes were used as the working model, aligning with our previous studies on geranyl-RNA: [5'-GGACU(xU)CUGCAG-3'] and [5'-CUGCAG(Y)AGUCC-3'], where Y is the canonical and non-canonical base pairing with (xU). The RNA duplexes were annealed in sodium phosphate buffer (10 mM, pH 6.5) containing 100 mM NaCl. The solutions were heated at 85 °C for 3 min then cooled down to room temperature for over 2 hours and stored at 4 °C overnight prior the experiments. Data points for each of the denaturing curves were acquired at 260 nm by heating and cooling from 5 to 85 °C for two cycles (4 ramps) (increments of 0.5 °C at the rate of heating of 0.5 °C/min), using the Cary-300 UV-Visible spectrometer equipped with a temperature controller system. The normalized absorptions were plotted using SigmaPlot 12.0. Melting temperatures and thermodynamic parameters of each duplex strand were obtained by fitting the melting curves using the Meltwin software 3.5 (McDowell & Turner, 1996).

### **Simulation methods**

To study the modifications in the context of RNA duplexes in MD simulations, we developed AMBER (Cornell et al, 1995) type force-field parameters for the atoms of the modified nucleoside. For obtaining the partial charges on the atoms, we used the online RESP charge-fitting server, RED (Dupradeau et al, 2010). 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 (Cornell et al, 1993). AMBER99 force-field parameters were used for bonded interactions (Cornell et al, 1995), and AMBER99 parameters with Chen-Garcia corrections were used for LJ interactions (Chen & Garcia, 2013).

The unmodified RNA duplex was constructed in A-form using Nucleic Acid Builder (NAB) suite of AMBER 11 package. Using WebMo, the web interface of Q-chem software (Anna I. Krylov, 2012), we performed mutations such as, U to meS<sup>2</sup>U, daS<sup>2</sup>U, geS<sup>2</sup>U, and faS<sup>2</sup>U, and A to G, to get the two different RNA duplexes for MD studies.

Molecular dynamics simulations were performed using Gromacs-4.6.3 package (Hess et al, 2008). The simulation system included the RNA duplex in a solution of 1M NaCl solution in a 3D periodic box. The

initial box size was 6 x 6 x 6 nm<sup>3</sup> containing 152 Na<sup>+</sup> ions, 130 Cl<sup>-</sup> ions and 6600 water molecules. The system was subjected to energy minimization to prevent any overlap of atoms, followed by a 5 ns equilibration run. The equilibrated system is simulated for 0.5 microseconds in 5 parallel simulations of 100 ns each. The MD simulations incorporated leap-frog algorithm with a 2 fs time-step to integrate the equations of motion. The system was maintained at 300K and 1 bar, using the velocity rescaling thermostat (Bussi et al, 2007) and Parrinello-Rahman barostat (Berendsen, 1984), respectively. The long-ranged electrostatic interactions were calculated using particle mesh Ewald (PME) (Darden et al, 1993) algorithm with a real space cut-off of 1.2 nm. LJ interactions were also truncated at 1.2 nm. TIP3P model (Jorgensen et al, 1983) was used represent the water molecules, and LINCS (Hess et al, 1997) algorithm was used to constrain the motion of hydrogen atoms bonded to heavy atoms. Co-ordinates of the RNA molecule were stored every 1 ps for further analysis.

## SUPPLEMENTAL REFERENCES

- Anna I. Krylov, P. M. W. G. (2012) Q-Chem: an engine for innovation. *WIREs Comput. Mol. Sci.*, 3, 317-326.
- Berendsen, H. J. C. P., J. P. M.; van Gunsteren, W. F.; DiNola, A.; Haak, J. R. (1984) Molecular dynamics with coupling to an external bath. *J Chem Phys*, 81, 3684-3690.
- Bussi, G., Donadio, D. & Parrinello, M. (2007) Canonical sampling through velocity rescaling. *J Chem Phys*, 126(1), 014101.
- Chen, A. A. & Garcia, A. E. (2013) High-resolution reversible folding of hyperstable RNA tetraloops using molecular dynamics simulations. *Proc Natl Acad Sci U S A*, 110(42), 16820-16825.
- Cornell, W. D., Cieplak, P., Bayly, C. I. & Kollmann, P. A. (1993) Application of RESP charges to calculate conformational energies, hydrogen bond energies, and free energies of solvation. *J Am Chem Soc*, 115(21), 9620-9631.
- 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. & Kollman, P. A. (1995) A Second Generation Force Field for the Simulation of Proteins, Nucleic Acids, and Organic Molecules. *J Am Chem Soc*, 117(19), 5179-5197.
- Darden, T., York, D. & Pedersen, L. (1993) Particle Mesh Ewald - an N.Log(N) Method for Ewald Sums in Large Systems. *J Chem Phys*, 98(12), 10089-10092.
- Dupradeau, F. Y., Pigache, A., Zaffran, T., Savineau, C., Lelong, R., Grivel, N., Lelong, D., Rosanski, W. & 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(28), 7821-39.
- Hess, B., Bekker, H., Berendsen, H. J. C. & Fraaije, J. G. E. M. (1997) LINCS: A linear constraint solver for molecular simulations. *J Comput Chem*, 18(12), 1463-1472.
- Hess, B., Kutzner, C., van der Spoel, D. & Lindahl, E. (2008) GROMACS 4: Algorithms for Highly Efficient, Load-Balanced, and Scalable Molecular Simulation. *J Chem Theory Comput*, 4(3), 435-447.
- Jorgensen, W. L., Chandrasekhar, J., Madura, J. D., Impey, R. W. & Klein, M. L. (1983) Comparison of Simple Potential Functions for Simulating Liquid Water. *J Chem Phys*, 79(2), 926-935.
- McDowell, J. A. & Turner, D. H. (1996) Investigation of the structural basis for thermodynamic stabilities of tandem GU mismatches: solution structure of (rGAGGUCUC)<sub>2</sub> by two-dimensional NMR and simulated annealing. *Biochemistry*, 35(45), 14077-14089.