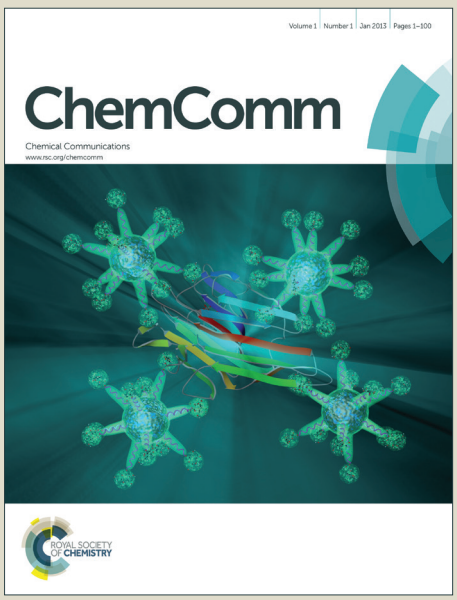
ChemComm

Accepted Manuscript



ROYAL SOCIETY OF CHEMISTRY

www.rsc.org/chemcomm

Journal Name

RSCPublishing

COMMUNICATION

Synthesis and Base Pairing Studies of Geranylated 2-thiothymidine, a Natural Variant of Thymidine

Cite this: DOI: 10.1039/xoxxoooooox

Rui Wang,[a,b] Srivathsan V. Ranganathan,[b] Maria Basanta-Sanchez,[b] Fusheng Shen,[a,b] Alan Chen[a,b] and Jia Sheng*[a,b]

Received ooth January 2012,
Accepted ooth January 2012

DOI: 10.1039/xoxxoooooox

www.rsc.org/

The synthesis and base pairing of DNA duplexes containing the geranylated 2-thiothymidine have been investigated. This naturally existing hydrophobic modification could grant better base pairing stability for T-G pair over normal T-A and other mismatched pairs in the duplex context. This study provides potential explanation for the different codon recognition preferences of the geranylated tRNAs.

DNA and RNA play essential roles in many biological processes by acting as both genetic information carriers and regulators/catalysts for numerous biochemical reactions. These functions are mainly achieved by the diverse and well-defined 3D structures formed by the normal Watson-Crick pairs and a variety of non-canonical base pairs, as well as other tertiary interactions.[1-4] In addition, natural DNA and RNA contain different chemical modifications to achieve structural and functional specificities and diversity. There are over 150 natural modifications that have been discovered so far in mRNA, rRNA, tRNA, non-coding RNA and DNA.[5, 6] Although their detailed functions have yet to be systematically explored, recent work suggests that some of these modifications have critical and previously unappreciated roles that might be closely related to many human diseases.[7-10] Therefore, studying the structures and nucleobase recognition properties of these chemical modifications is significant for further elucidation of RNA functions and development of new RNA-targeted therapeutics.

Within all the nucleic acid species, tRNAs contain the most numbers of modifications. The over 90 different chemical modifications are highly involved in several central biological processes including tRNA charging with specific amino acids, code-anticode recognition, as well as translation.[11-19] 2-thiouridine geranylation is a special modification that has been discovered very recently by Liu and coworkers in several bacteria including *E. coli, E. aerogenes, P. aeruginosa* and *S. Typhimurium*.[20] As shown in Fig.1 (compound 3, 4, 5), the large hydrophobic geranyl group is covalently attached to the sulphur atom of the 2-thiouridine derivatives by the enzyme SelU, the selenouridine synthase that can also replace sulphur with selenium in these substrates (Scheme. 1).[21-23]

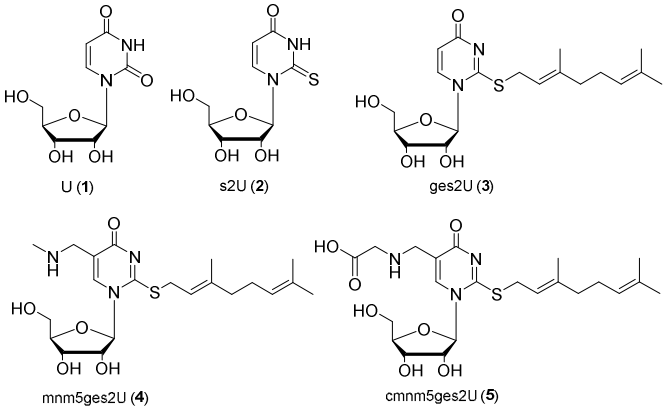
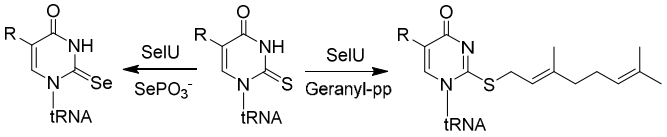


Fig. 1 Chemical structures of uridine (U, **1**), 2-thiouridine (s2U, **2**), geranylated 2-thiouridine (ges2U, **3**), geranylated 5-methylaminomethyl-2-thiouridine (mnm5ges2U, **4**), and geranylated 5-carboxylmethylaminomethyl-2-thiouridine (cmnm5ges2U, **5**).

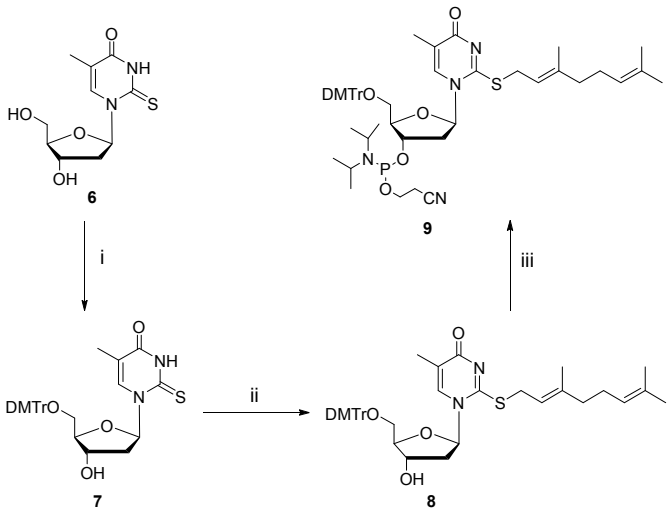


Scheme 1 The selenouridine synthase (SelU) have dual functions to replace sulphur with selenium in the presence of selenophosphate (left arrow) and to install geranyl groups to sulphur atom in the presence of geranyl pyrophosphate (right arrow). The R group represents the two groups mnm- or cmnm- in the position 5 of compound **4** and **5** in Fig. **1**.

Although it has been proposed that the S-geranyl-2-thiouridine could merely be an intermediate product in the transformation of 2-thiouridine to 2-selenouridine based on the fact that the geranylated

2-thiouridine could be chemically transferred to 2-selenouridine in the presence of the very reactive sodium hydroselenide (NaHSe),[24] it's an interesting question that why nature uses such a bulky hydrophobic group (the only one discovered so far) in hydrophilic RNA systems. Besides the speculation that such functionalities could be evolutionary fossils from ancient RNA-mediated lipid synthesis,[25] one expected effect would be that the base pairing patterns of geranyl-modified uridine will be different. Indeed, these geranylated residues are located in the first anticodon position (wobble position 34) of tRNAs specific for lysine, glutamine and glutamic acid, where the geranyl group plays biological functions by affecting the codon recognition and frameshifting during translation.[20] It has been observed that the geranylated $tRNA^{Glu}_{UUC}$ increases the codon recognition efficiency to GAG comparing to GAA,[20] implying the geranylated uridine has stronger base pairing with G over A in the end of codons. Therefore, studying the base pairing specificity of geranylated uridine analogs will be helpful to explore more detailed structures, functions and working mechanisms of this modification. Toward this goal, we report here the first synthesis of oligonucleotides containing geranylated thymidine, their duplex thermodynamic properties, as well as corroborating computational simulation studies.

As shown in scheme **2**, the selective tritylation of 5'-OH followed by the 2-thio-geranylation of the commercial available 2-thiothymidine generates the intermediate **8** in high yields. The phosphoramidite functionality was subsequently installed into the 3'-OH in order to make the geranylated thymidine phosphoramidite **9**, the building block for the oligonucleotide solid phase synthesis. It's noteworthy that this final product presents as a sticky gel and dissolves in dichloromethane much better than acetonitrile. Therefore, the dichloromethane solution was used for the following solid phase synthesis.



Scheme 2 Synthesis of S-geranyl-2-thiothymidine phosphoramidite building block for solid phase synthesis. (i) Dimethoxyltrityl chloride (DMTrCl), pyridine, r.t., 85%. (ii) Geranyl bromide, *N,N*-diisopropylethylamine (DIEA), MeOH, r.t., 92%. (iii) 2-Cyanoethyl *N, N*-diisopropylchlorophosphoramidite, *N,N*-diisopropylethylamine (DIEA), CH_2Cl_2, r.t., 72%.

This novel geranyl-thymidine phosphoramidite was subsequently incorporated into several DNA oligonucleotides using the standard protocol in a DNA synthesizer. The synthesized oligonucleotides were cleaved off the solid support and fully deprotected with AMA (ammonium hydroxide/methylamine) solution at 65 °C for 30 min, followed by the HPLC purification in both DMT on and off forms using a Zorbax SB-C18 column. It's surprising that the geranylated oligo has slightly shorter retention time than the native one (Fig. 2), indicating the highly hydrophobic geranyl group might have a special structure alignment in the oligonucleotides. The purified oligonucleotides were also confirmed by high-resolution mass-spectrometric analysis (Table S1 and Fig. S11-S13). These data also indicate that the geranyl group is stable under the iodine (20 mM, 90 s), strong acid (3% TCA) and strong base (AMA) treatments in different sequences. To demonstrate the presence of the geranyl modification, both native and modified DNA oligonucleotides underwent enzymatic hydrolysis with nuclease P1 and bacterial alkaline phosphatase. The nucleoside composition was determined by ultra-high performance liquid chromatography (UHPLC)-Tandem MS (MS/MS) and further compared to our synthesized standard geranyl thymidine. (Fig.S14 and Table S2, S3).

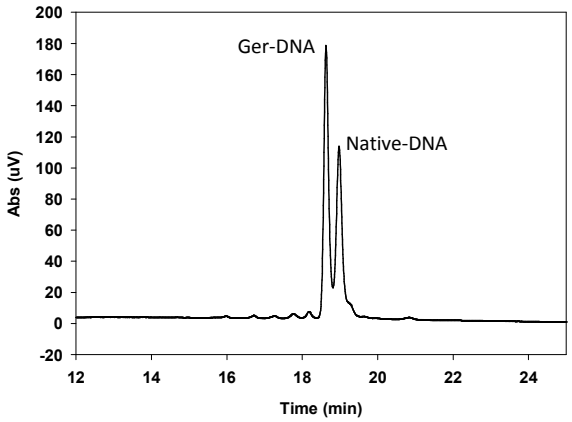


Fig. 2 HPLC co-injection profile of the purified native and geranyl-oligo: CTTCTXGTCCG (X represents either native T or ges2T).

With the geranylated oligonucleotides in hand, we carried out the thermal denaturation experiments of their duplexes in order to study the base pairing stability and selectivity. The UV-melting temperatures (T_m) curves of native and geranylated duplexes, 5'-CTTCT**X**GTCCG-3'•5'-CGGAC**Y**AGAAG-3', with matched and mismatched base pairs (X pairs with Y) are shown in Figure 3A and 3B; and the according data are summarized in Table 1. The geranyl group largely decreases the stability of duplex containing normal Watson-Crick pairs. One single geranylation decrease the T_m for 16 °C, corresponding to a reduced ΔG^0 of 4.3 kcal/mol (entry 1 and 5). The duplexes with mismatched base pairs, on the other hand, are not affected by geranylation as largely as the normal matched duplex. The T_m of duplexes containing T-G and T-T mismatches slightly decrease by 2.0 °C and 6.1 °C (entry 2 *vs* 6, and 4 vs 8), corresponding to a reduced ΔG^0 of 0.3 and 0.9 kcal/mol respectively; while the T-C mismatch increases by 0.9 °C, corresponding to a increased ΔG^0 of 0.4 kcal/mol (entry 3 *vs* 7). When directly comparing the Watson-Crick base pairs (T-A and ges2T-A) with their own other mismatched pairs, as shown in the ΔT_m column, it's clear that geranylated thymidine has stronger pairing preference with

G comparing to other bases. The T_m of duplexes containing ges2T-G pair is 5.9 °C higher than the native T-A duplex, with an increased ΔG^0 of 1.3 kcal/mol (entry 5 and 6). In comparison, the native duplex with T-G pair has a significantly lower stability (-8.1 °C) compared to the normal duplex (entry 1 and 2). Since it has been known that the 2-thio modification alone decreases the stability of DNA duplex containing 2-thioT-G pair by over 10 °C compared to the native T-G duplex,[26] the additional geranyl group should mainly contribute to the stability increase we observed in this study, although the sequences are different. These data collectively suggest that the geranyl modification decreases the normal T-A pair stability and improves the base pairing discrimination of T-G pair against T-A and other mismatched pairs. This might also explain the better recognition of the G-ending codons over the A-ending ones by geranylated tRNA.

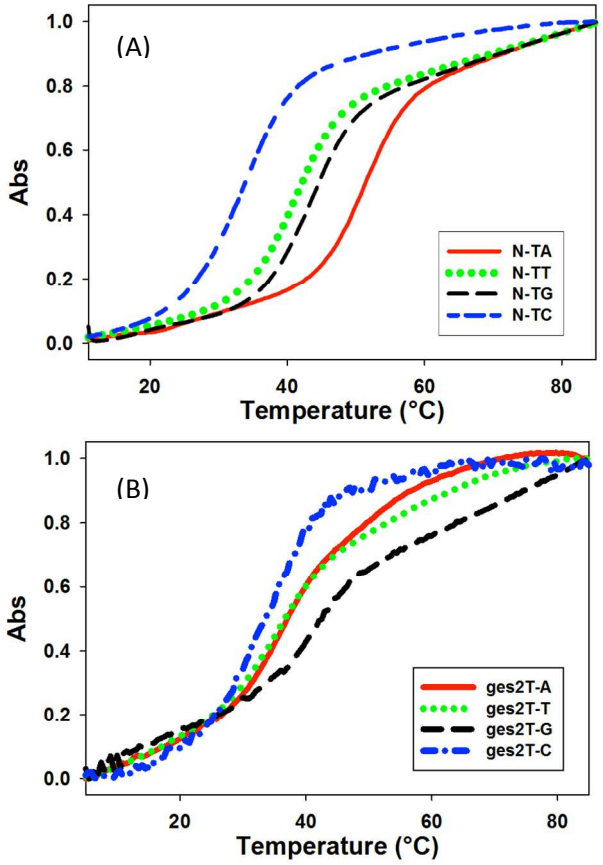


Fig. 3 Normalized UV-melting curves of DNA duplexes. (A) Native sequence (5'-CTTCT**T**GTCCG) pairs with matched and mismatched strands. (B) Geranylated sequence (5'-CTTCT**X**GTCCG) pairs with matched and mismatched sequences (X represents ges2T).

Table 1. Base pairing properties of geranyl thymidine (ges2T) in the context of the 11 base pair DNA duplex 5'-CTTCT**X**GTCCG-3'•5'-CGGAC**Y**AGAAG-3' (X pairs with Y).

Entry	Base pairs		T_m (°C)[a]	ΔT_m (°C)[b]	$-\Delta G°_{37}$[c] (kcal/mol)
	X	Y			
1	T	A	50.7	--	12.3
2	T	G	42.6	**-8.1**	9.6
3	T	C	32.5	-18.2	7.2
4	T	T	40.6	-10.1	8.7
5	ges2T	A	34.7	--	8.0
6	ges2T	G	40.6	**+5.9**	9.3
7	ges2T	C	33.4	-1.3	7.6
8	ges2T	T	34.5	-0.2	7.8

a: the 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 T-A pair and ges2T-A pair respectively.
c: Obtained by non-linear curve fitting using Meltwin 3.5.[27]

By analyzing the potential base pairing patterns (Fig. 4), this strong base pairing discrimination between U-G and U-A pair could be attributed to the changes of hydrogen bonding patterns by the geranyl group at the position 2, which switches the N3 to a hydrogen bonding acceptor from a donor. As a result, only guanosine that has two connected hydrogen bonding donors can form relatively stable two hydrogen bonds with the geranylated uridine or thymidine (Fig. 4B), while only one hydrogen bond can be theoretically formed in all the other three pairing patterns. In addition, the bulky hydrophobic geranyl group might further disrupt the hydrogen bonding formation and reduce the base pairing as well as the overall duplex stability, although it might experience increased base stacking.

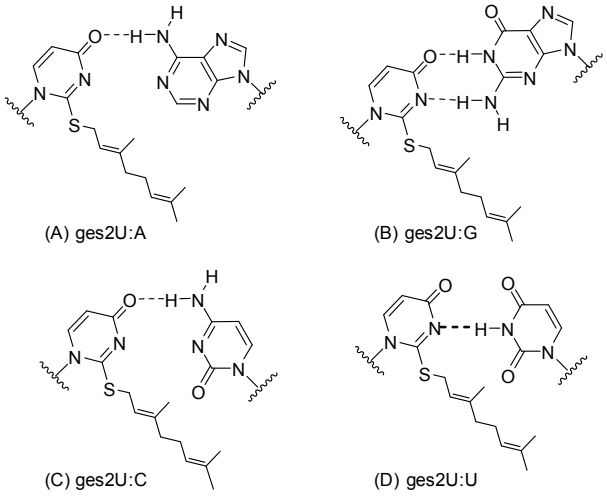


Fig. 4 Proposed base pairing and mispairing patterns of geranylated 2-thiouridine (ges2U) with A, G, C and U respectively.

To obtain more detailed structural insights into these geranylated base pairs, we developed AMBER-type molecular force-field parameters for this geranylated thymidine and carried out molecular dynamics (MD) simulations of the DNA duplexes containing ges2T-A and ges2T-G pairs respectively (see SI for the detailed methodology). The simulations show that most of the ges2T-A pairs are in an unbound state completely lacking hydrogen-bonding interactions, although two different hydrogen-bonding modes were transiently observed over the course of the simulation. As shown in Fig. 5A and 5B, the N6 hydrogen of adenosine can bond with both N3 and O4 of ges2T, while there is no hydrogen bond observed in Fig. 5C, which actually represents the most abundant status of ges2T-A interaction during the course of the MD simulation. The two hydrogen bond donor-acceptor distances fluctuate between 3 to 12 Å, mostly falling outside the normal hydrogen bonding distance of ~3.1 Å (Fig. 5E, red and green curves; and Fig. S15, left bars). However, the ges2T-G pair shows a much more stable interaction pattern with two normal hydrogen bonds present throughout the

simulation trajectory (Fig. 5D and Fig. 5E, blue and magenta curves; and Fig. S15, right bars). Although more detailed calculations are still needed to quantify the stacking effect of the geranyl groups with its neighbouring residues and the overall changes in free-energy that each effect could contribute, our study here confirms that the hydrogen bonding interaction between geranylated thymidine and the guanosine base are relatively stable and thus could significantly contribute to the better recognition of T-G pair over T-A pair.

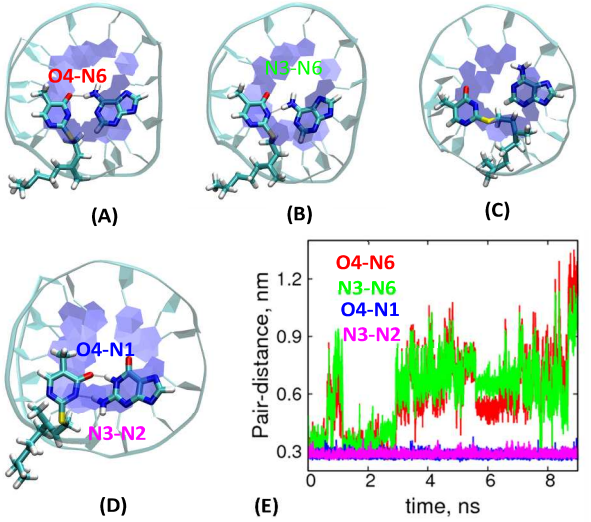


Fig. 5 Observed base pairing patterns of geranylated thymidine with adenosine and guanosine in a duplex context. (A) ges2T-A pair with one H-bond between O4 and N6; (B) ges2T-A pair with one H-bond between N3 and N6; (C) ges2T-A pair pattern with no hydrogen bond formation; (D) ges2T-G pair with two H-bonds between O4 and N1; N3 and N2. (E) Dynamic hydrogen bonding distances in ges2T-A (O4-N6, N3-N6) and ges2T-G (O4-N1, N3-N2) pairs along the simulation time.

In conclusion, we have synthesized the geranylated DNA oligonucleotides and studied their base pairing stability and specificity for the first time. This hydrophobic natural modification was found to make the T-G pair more stable than the normal Watson-Crick T-A pair by forming two relatively stable hydrogen bonds between the geranylated thymidine and guanosine base. This higher thermal stability of T-G pairing might provide a potential explanation to the better recognition of G-ending codes over the others by certain geranylated tRNAs. Besides diversifying the base pairing specificity, we speculate that this bulky hydrophobic geranyl group have more functions such as RNA localization, RNA transportation especially passing through the cell membranes, as well as some geranyl group involved metabolic pathways.[20] In addition, the other functionalities in position 5 of geranyl uridine may also play important roles together with the geranyl group. Furthermore, it's quite reasonable to speculate that such modifications are a molecular fossil from the RNA World, where RNA mediated the lipid synthesis. Therefore, further exploration of the structures and functions of this unusual hydrophobic geranyl modification will shed light on the new insights into the roles that these chemical modifications play in both current biological processes and the original stages of RNA World.

This work was supported by the start-up fund from SUNY Albany. Rui Wang is currently supported by postdoctoral fellowship from Simons Foundation (338863, R.W.). We are grateful to Prof. Daniele Fabris, Dr. Matteo Scalabrin, Prof. Qishan Lin, Prof. Maksim Royzen and Leah Seebald for their help in mass-spec analysis, Dr. Ken Halvorsen for his help in T_m study and Prof. Paul Agris for his comments.

Notes and references

[a] Department of Chemistry, [b] The RNA Institute, University at Albany, State University of New York, 1400 Washington Ave. Albany, NY, 12222. Email: jsheng@albany.edu; Tel: +1 518 437 4419

† Electronic Supplementary Information (ESI) available: Figures S1-S15 and details of experimental and simulation methods are provided. See DOI: 10.1039/c000000x/

1 M. J. Fedor, *Annu Rev Biophys*, 2009, **38**, 271.
2 J. A. Doudna and J. R. Lorsch, *Nat Struct Mol Biol*, 2005, **12**, 395.
3 A. Roth and R. R. Breaker, *Annu Rev Biochem*, 2009, **78**, 305.
4 J. Sheng, A. Larsen, B. D. Heuberger, J. C. Blain and J. W. Szostak, *J Am Chem Soc*, 2014, **136**, 13916.
5 M. A. Machnicka, K. Milanowska, O. Osman Oglou, E. Purta, M. Kurkowska, A. Olchowik, W. Januszewski, S. Kalinowski, S. Dunin-Horkawicz, K. M. Rother, M. Helm, J. M. Bujnicki and H. Grosjean, *Nucleic Acids Res*, 2013, **41**, D262.
6 W. A. Cantara, P. F. Crain, J. Rozenski, J. A. McCloskey, K. A. Harris, X. Zhang, F. A. Vendeix, D. Fabris and P. F. Agris, *Nucleic Acids Res*, 2011, **39**, D195.
7 P. J. Sample, L. Koreny, Z. Paris, K. W. Gaston, M. A. Rubio, I. M. Fleming, S. Hinger, E. Horakova, P. A. Limbach, J. Lukes and J. D. Alfonzo, *Nucleic Acids Res*, 2015, **43**, 4262.
8 E. M. Phizicky and A. K. Hopper, *Genes Dev*, 2010, **24**, 1832.
9 A. K. Hopper and H. H. Shaheen, *Trends Cell Biol*, 2008, **18**, 98.
10 N. Liu, Q. Dai, G. Zheng, C. He, M. Parisien and T. Pan, *Nature*, 2015, **518**, 560.
11 J. R. Iben and R. J. Maraia, *RNA*, 2012, **18**, 1358.
12 W. D. Graham, L. Barley-Maloney, C. J. Stark, A. Kaur, C. Stolarchuk, B. Sproat, G. Leszczynska, A. Malkiewicz, N. Safwat, P. Mucha, R. Guenther and P. F. Agris, *J Mol Biol*, 2011, **410**, 698.
13 Y. Bilbille, E. M. Gustilo, K. A. Harris, C. N. Jones, H. Lusic, R. J. Kaiser, M. O. Delaney, L. L. Spremulli, A. Deiters and P. F. Agris, *J Mol Biol*, 2011, **406**, 257.
14 K. Sheppard, P. M. Akochy and D. Soll, *Methods*, 2008, **44**, 139.
15 C. L. Sanders, K. J. Lohr, H. L. Gambill, R. B. Curran and J. F. Curran, *RNA*, 2008, **14**, 1874.
16 X.-L. Zhou, Z.-R. Ruan, M. Wang, Z.-P. Fang, Y. Wang, Y. Chen, R.-J. Liu, G. Eriani and E.-D. Wang, *Nucleic Acids Res*, 2014.
17 A. K. Hopper, *Genetics*, 2013, **194**, 43.
18 J. Shepherd and M. Ibba, *FEBS Lett*, 2013, **587**, 2895.
19 S. Laxman, B. M. Sutter, X. Wu, S. Kumar, X. Guo, D. C. Trudgian, H. Mirzaei and B. P. Tu, *Cell*, 2013, **154**, 416.
20 C. E. Dumelin, Y. Chen, A. M. Leconte, Y. G. Chen and D. R. Liu, *Nat Chem Biol*, 2012, **8**, 913.
21 Z. Veres and T. C. Stadtman, *Proc Natl Acad Sci U S A*, 1994, **91**, 8092.
22 P. Chen, P. F. Crain, S. J. Nasvall, S. C. Pomerantz and G. R. Bjork, *EMBO J*, 2005, **24**, 1842.
23 M. D. Wolfe, F. Ahmed, G. M. Lacourciere, C. T. Lauhon, T. C. Stadtman and T. J. Larson, *J Biol Chem*, 2004, **279**, 1801.
24 P. Bartos, A. Maciaszek, A. Rosinska, E. Sochacka and B. Nawrot, *Bioorg Chem*, 2014, **56**, 49.
25 A. I. Scott, *Tetrahedron Letters*, 1997, **38**, 4961.
26 H. O. Sintim and E. T. Kool, *J Am Chem Soc*, 2006, **128**, 396.
27 J. A. McDowell and D. H. Turner, *Biochemistry*, 1996, **35**, 14077