Published online 14 January 2017

Nucleic Acids Research, 2017, Vol. 45, No. 8 ***4825–4836***
doi: 10.1093/nar/gkw1347

C5-substituents of uridines and 2-thiouridines present at the wobble position of tRNA determine the formation of their keto-enol or zwitterionic forms - a factor important for accuracy of reading of guanosine at the 3′-end of the mRNA codons

Elzbieta Sochacka[1], Elzbieta Lodyga-Chruscinska[2], Justyna Pawlak[2], Marek Cypryk[3], Paulina Bartos[1], Katarzyna Ebenryter-Olbinska[1,4], Grazyna Leszczynska[1] and Barbara Nawrot[4,*]

[1]Institute of Organic Chemistry, Faculty of Chemistry, Lodz University of Technology, Zeromskiego 116, 90-924 Lodz, Poland, [2]Institute of General Food Chemistry, Faculty of Biotechnology and Food Sciences, Lodz University of Technology, Zeromskiego 116, 90-924 Lodz, Poland, [3]Department of Computer Modeling, Centre of Molecular and Macromolecular Studies, Polish Academy of Sciences, Sienkiewicza 112, 90-363 Lodz, Poland and [4]Department of Bioorganic Chemistry, Centre of Molecular and Macromolecular Studies, Polish Academy of Sciences, Sienkiewicza 112, 90-363 Lodz, Poland

Received May 9, 2016; Revised December 14, 2016; Editorial Decision December 20, 2016; Accepted December 30, 2016

ABSTRACT

Modified nucleosides present in the wobble position of the tRNA anticodons regulate protein translation through tuning the reading of mRNA codons. Among 40 of such nucleosides, there are modified uridines containing either a sulfur atom at the C2 position and/or a substituent at the C5 position of the nucleobase ring. It is already evidenced that tRNAs with 2-thiouridines at the wobble position preferentially read NNA codons, while the reading mode of the NNG codons by R5U/R5S2U-containing anticodons is still elusive. For a series of 18 modified uridines and 2-thiouridines, we determined the p*Ka* values and demonstrated that both modifying elements alter the electron density of the uracil ring and modulate the acidity of their N3H proton. In aqueous solutions at physiological pH the 2-thiouridines containing aminoalkyl C5-substituents are ionized in ca. 50%. The results, confirmed also by theoretical calculations, indicate that the preferential binding of the modified units bearing non-ionizable 5-substituents to guanosine in the NNG codons may obey the alternative C-G-like (Watson–Crick) mode, while binding of those bearing aminoalkyl C5-substituents (protonated under physiological conditions) and especially those with a sulfur atom at the C2 position, adopt a zwitterionic form and interact with guanosine via a 'new wobble' pattern.

INTRODUCTION

5-Substituted uridines and 2-thiouridines are post-transcriptionally modified nucleosides present in the position 34 (wobble or first position of the anticodon) in several transfer RNAs (tRNAs) in virtually all living organisms, from bacteria to human. According to the wobble hypothesis, their location is critical for the precise reading of genetic information (1). Some of them can recognize both A and G in the third position of the mRNA synonymous codons (Figure 1). The thermodynamic stability of the wobble base pair U-G is comparable to that of the Watson-Crick U-A base pair, although the RNA duplexes harboring this wobble base pair are thermally less stable than their Watson-Crick U-A counterparts (2–4). Replacement of the oxygen-2 of the uracil ring with a sulfur atom is observed for at least 10 uridines of the tRNAs specific for lysine, glutamic acid and glutamine (http://modomics.genesilico.pl (5), http://mods.rna.albany.edu (6)). The corresponding RNA duplexes containing a S2U-A base pair are thermodynamically more stable than those with an U-A base pair due to the preferential S2U C3'-*endo* sugar ring pucker, improved base stacking in the RNA chains and an enhanced overall A-type RNA duplex helical structure (4,7–17).

*To whom correspondence should be addressed. Tel: +48 42 6803248; Email: bnawrot@cbmm.lodz.pl

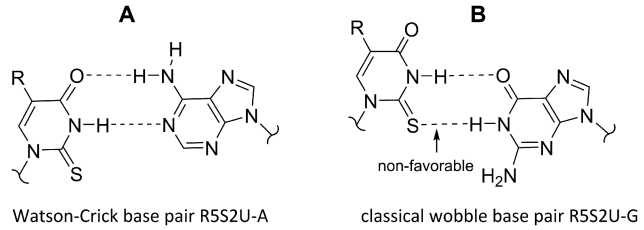
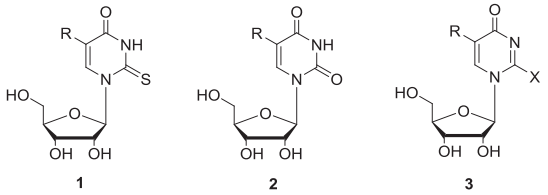


Figure 1. (**A**) A classical Watson-Crick R5S2U-A base pair, favorable; (**B**) a classical wobble R5S2U-G base pair, non-favorable.

Early data (14–17) suggested that 2-thiouridine is introduced into the wobble position of tRNA to enhance hybridization to adenosine in the NNA codons (where N is any nucleoside), whereas the wobble base pairing with guanosine in the NNG codons is restricted due to less effective hydrogen bonding between the N1H donor of guanine and the sulfur acceptor of 2-thiouracil (18,19) (Figure 1). However, the results of the subsequent biological studies contradicted this idea and suggested that the 3′-G-ending codons are well recognized by anticodons containing the 5-substituted 2-thiouridines (20,21). The most notable of these were the results demonstrating that anticodons with 5-methylaminomethyl-2-thiouridine (mnm5S2U) or 5-carboxymethylaminomethyl-2-thiouridine (cmnm5S2U) of the cytosolic tRNAs and those with 5-taurinomethyl-2-thiouridine (τm5S2U) of the mitochondrial tRNAs, all promote reading of both NNA and NNG codons. Other studies have found a similar tendency for the A and G recognition by the same 5-substituted, but not 2-thiolated, uridines (22). These results suggest that the substituent at the C5 position contributes to the electron density within the π electron system of a nucleobase (through its electron withdrawing/donating properties) and promotes the binding of the tautomers of 5-substituted uridines/2-thiouridines to the guanosine units.

Several of the 40 modified uridines/2-thiouridines found in the wobble position of tRNAs contain -O-R or -CH_2-R substituents at the C5 position of the uracil residue. The -O-R substituents (-OH, -OCH_3 and -OCH_2COOH, denoted ho, mo and cmo, respectively) are expected to increase the electron density of the uracil ring through a mesomeric effect originating from the overlapping of the p orbital of the oxygen atom with the π orbital of the uracil ring (23). The electron donating properties of the -CH_2-R substituents, e.g., -CH_3 (m) or -CH_2COOCH_3 (mcm), are weak and their contribution to the electron density of the pyrimidine ring is limited. However, the substituents containing aminoalkyl groups, e.g., -CH_2NHCH_3 (mnm) or -CH_2NHCH_2COOH (cmnm), significantly affect the electronic density of the nucleobases because their nitrogen atoms at a physiological pH (7.4) are substantially protonated (the p*K*a values of secondary amines exceed 9 units (24)). The protonated 5-aminoalkyl substituents exert strong electron-withdrawing properties and promote deprotonation of the N3H function.

Takai and Yokoyama suggested that mnm5S2U might recognize G in a non-canonical mode, in which the N3H function of the 2-thiouracil ring is ionized and the negative charge is localized at the sulfur atom (25). In this pre-structured ionic form, mnm5S2U may interact with the N1H and N2H donors of guanosine using either the N3 and anionic S2 acceptors (according to the Watson-Crick scheme), or the O4 and N3 acceptors (according to the wobble mode), the latter with the movement of the uridine unit toward the minor groove. Only recently, the mnm5S2U-guanosine base pair has been found in the crystal structure of the tRNA-mRNA complex bound to the 70S ribosome (26). The U34-G base pair found in the biological context has the latter geometry predicted by Takai and Yokoyama, that may be executed either by the keto-enol form of mnm5S2U or by its zwitterionic form.



Compound	R	Name	Substituent's abbreviation	X
a	H	-		H
b	CH_3	methyl	m	H
c	CH_2COOCH_3	methylcarboxymethyl	mcm	H
d	CH_2COOH	carboxymethyl	cm	H
e	CH_2CONH_2	carbamoylmethyl	ncm	H
f	CH_2NHCH_3	methylaminomethyl	mnm	H
g	CH_2NHCH_2COOH	carboxymethylamino methyl	cmnm	H
h	$CH_2NHCH_2CH_2SO_3H$	taurinomethyl	τm	H
i	OCH_3	methoxy	mo	H
j	H	-	meS	S-CH_3
k	H	-	geS	S-$C_{10}H_{17}$

Figure 2. Structures of the compounds used in the pH-potentiometric titration experiments.

Of note, crystallographic data obtained for codon-anticodon models in the ribosome context demonstrate that the keto-enol pre-structured forms of other 5-substituted uridines and 2-thiouridines may bind to the guanosine unit according to the C-G-like or the bifurcated model (27–30).

An abundance of the pre-structured form of a nucleoside in solution at a given pH is related to the p*K*a value of N3H in a nucleobase, which in turn depends on the electron withdrawing/donating properties of the substituent present at position C5. In the present study, we aimed to investigate an influence of the sulfur atom in position 2 and that of various substituents at position 5 on electronic properties of the modified uridines and to learn on their ability to read the guanosine unit at the 3′-end of the mRNA codons. Because the reported p*K*a values of the N3H group of 5-substituted 2-thiouridines and uridines (nucleosides **1** and **2**, respectively, Figure 2) had been previously obtained by different methods, their direct comparison was not meaningful. Additionally, some values were missing or were given as rough approximations. To this end, we prepared a series of compounds (Figure 2), which, for the first time, were used for the determination of p*K*a values in a series of uniform pH-potentiometric titration experiments. In the measurements, we also included 5-substituted 4-pyrimidinone nucleosides and *S*-alkylated derivatives of 2-thiouridine (**3**). Addition-

ally, the results were verified by theoretical DFT (density functional theory) calculations.

MATERIALS AND METHODS

All of the chemicals were Aldrich products of puriss grade.

Preparation of the 5-substituted 2-thiouridines 1, uridines 2 and 4-pyrimidinone nucleosides 3

All nucleosides used in experiments (Figure 2) are known compounds and were prepared in our laboratory according to reported procedures. The 2-thiouridines **1a-c,i** and the 5-substituted uridines **2b,c,i** were prepared by the *N*-glycosidic bond formation (13,16,31–38), while the nucleosides **1f-h** and **2f-h** were prepared by the introduction of a C5-substituent into the appropriate derivatives of 2-thiouridine/uridine according to published methods (39–42), in some cases using recently improved procedures (43,44). The nucleosides **1d,e** and **2d,e** were prepared from the appropriate 5-substituted precursors as described elsewhere (39,45). The 4-pyrimidinone ribonucleosides (R5H2U, **3a-f,i**) were prepared by the desulfuration of the parent 2-thionucleosides (46,47), while the derivatives **3j,k** were obtained by the *S*-methylation or *S*-geranylation of **1a** (48–50).

Potentiometric measurements

The acidity constants of the ligands (p*K*a) were determined by the pH-potentiometric titration of 2.0-ml samples. The concentration of the nucleoside in solution was 1×10^{-3} M. Measurements were carried out at 298 K and at a constant ionic strength of 0.1 M NaCl using a MOLSPIN pH meter (Molspin Ltd., Newcastle-upon-Tyne, UK) equipped with a digitally operated syringe (the Molspin DSI 0.250 ml) controlled by a computer. For the titrations, a carbonate-free NaOH solution of known concentration (0.1 M) was used and measurements were made using a Russel CMAWL/S7 semi-micro combined electrode, calibrated for hydrogen ion concentration using the method of Irving *et al*. (51). The accepted fit for the titration curves was always less than 0.01 ml. The number of experimental points was 100–150 for each titration curve. The titration points included in the evaluation could be reproduced within 0.005 pH units in the whole pH range examined (pH from 2 to 12). The protonation constants of the ligands were evaluated by performing iterative non-linear least squares fit of the potentiometric equilibrium curves through mass balance equations using the computer program SUPERQUAD (52). The sigma value (the root mean squared weighted residual) obtained after the refinement of the stability constants was 1, which suggested that the data were fitted within experimental error. The equilibrium constants reported in this work were obtained from a fitting performed using three titration curves simultaneously.

Computational methods

All quantum mechanical calculations were performed using the Gaussian 09 suite of programs (53). Geometries of the bases and base pair model systems were optimized using the hybrid B3LYP density functional (54) corrected for dispersion interactions using Grimme GD3 empirical term (55), with 6–31G(d) basis set in the gas phase and 6–31+G(d) basis set in aqueous solution. All stationary points were identified as stable minima by frequency calculations. The vibrational analysis provided thermal enthalpy and entropy corrections at 298 K within the rigid rotor/harmonic oscillator/ideal gas approximation (53). Thermochemical corrections were scaled by a factor of 0.98. More accurate electronic energies were obtained using the B3LYP functional, including the Grimme GD3 dispersion corrections (55), with the larger 6-311++G(3df,2p) basis set for the B3LYP/6-31G(d) (or B3LYP/6-31+G(d)) optimized geometries. Integration grid was set to ultrafine. All base pair interaction energies were corrected for the basis set superposition error (BSSE) using the counterpoise procedure (CP) of Boys and Bernardi (56). The BSSE's at B3LYP/6-311++G(3df,2p) level of theory were in the range of 0.27-0.45 kcal/mol for all complexes studied.

Geometries of nucleic bases and base pairs models in aqueous solution were optimized at the B3LYP-GD3/6-31+G(d) level within the Conductor-like Polarizable Continuum Model (CPCM) (57), assuming UFF cavities as implemented in Gaussian 09 (53). No restrains on geometries were applied. All minima were identified based on vibrational analysis, as above. The free energy differences between the tautomers of uracil derivatives were calculated using the simple thermodynamic cycle as $\Delta G_{(aq)}(T2\text{-}T1) = G_{(g)}(T2) + \Delta G_{(hydr)}(T2) - (G_{(g)}(T1) + \Delta G_{(hydr)}(T1))$, where T1, T2 - nucleic base tautomers, $G_{(g)}$ - Gibbs free energy of tautomer in the gas phase, $G_{(aq)} = G_{(g)} + \Delta G_{(hydr)}$- tautomer free energy in a water solution, $\Delta G_{(hydr)}$ - tautomer free energy of hydration (see Supplementary Figure S2) (58). Free energy of hydration ($\Delta G_{(hydr)}$) of nucleic base tautomers were estimated by a procedure implemented in Gaussian 09 at the CPCM-B3LYP/6-31+G(d) level for solution-optimized geometries. Atomic charges fitted to the electrostatic potential were calculated at the B3LYP/6-311++G(3df,2p) level according to the Merz-Singh-Kollman scheme (59). Electrostatic potential maps on the 0.002 au molecular electron density isosurfaces were plotted using Wavefunction Spartan'08 program (60). More methodological details are given in Supplementary Data.

RESULTS

Chemistry

The nucleoside analogs prepared for the studies are shown in Figure 2. A series of 5-substituted 2-thiouridines (R5S2U, **1a-h**) and the 2-thio analog (**1i**) of the naturally occurring mo5U (**2i**), as well as their 2-oxo congeners (R5U, **2a-i**) were prepared either by the *N*-glycosidic bond formation (the nucleosides with R = H, m, mcm or mo) (13,16,31–37) or by the introduction of a C5-substituent into the appropriate derivatives of 2-thiouridine/uridine (the nucleosides with R = mnm, cmnm, τm) (39–42) using reported (in some cases improved) procedures (43,44). The 4-pyrimidinone ribonucleosides (R5H2U, **3a-f,i**) were prepared by the desulfuration of the parent 2-thionucleosides

(46,47), while the derivatives **3j,k** were obtained by the *S*-alkylation of S2U (48–50). The synthetic procedures as well as the spectral and conformational characteristics of **1a**,**c-g**,**i**, **2a**,**c-g**,**i** and **3a**,**c-g**,**i** have been described in our recent paper (47).

Potentiometric titrations and p*K*a determination

The p*K*a values for the nucleosides (Table 1) were calculated from the respective pH-potentiometric titration curves (see Supplementary Data, Supplementary Figure S1) using an improved SUPERQUAD program (52). For the **d-h** series of compounds, the p*K*a values were determined for the additional carboxyl, carbamoyl, aminoalkyl and sulfonic groups that were present as part of the 5-substituents. Because of insufficient stability of **3g** and **3h**, their p*K*a values could not be determined.

As shown in Table 1, for the majority of the 2-thiouridines **1,** the p*K*a values of N3H were lower than that for the parent uridines **2** by 1 unit, with the exception of **1h**, whose p*K*a differed only by 0.4 unit from that of **2h**. The p*K*a values of 8.09 and 9.15 obtained for 2-thiouridine (**1a**) and uridine (**2a**), respectively, were in good agreement with the previously reported values of 8.05 (61) and 9.18 (62). Additionally, the recently reported p*K*a values 8.0 and 9.3 of these compounds, where the titrations were monitored by ultraviolet (UV) and nuclear magnetic resonance methods (7), confirm our results.

The ionization properties of the remaining nucleobases in both the 2-thiouridine and uridine series depended on the type of the C5 substituent. The p*K*a values for the dissociation of the N3H proton in **1f-h** and **2f-h** (bearing an mnm, cmnm or τm substituent) were lower than those for their corresponding parent, non-substituted units **1a** or **2a**. At the physiological pH, the aminoalkyl groups in **1f-h** and **2f-h** would be substantially protonated (p*K*a values of their aminoalkyl groups are >9) to become electron-withdrawing groups; thus, the acidities of the corresponding N3H hydrogen atoms should be higher. Accordingly, p*K*a values of 7.28, 7.36 and 7.10 were found for **1f**,**g** and **h**, respectively, compared to the p*K*a value of 8.45 for m5S2U (**1b**). The p*K*a values of the uridines **2f-h** were higher than that of their thio-analogs **1f-h**, but were lower than that of the aminoalkyl-free uridines **2a-e**. Interestingly, the p*K*a value of τm5U (**2h**) was significantly lower (7.51) than that of the remaining uridines and was close to the p*K*a values of the aminoalkyl-substituted 2-thiouridines.

The p*K*a values of the nucleosides **1i** and **2i** containing the -OCH_3 group were one unit lower than that for their 5-methyl-S2U and 5-methyl-U (**1b** and **2b**) congeners. This indicated that the -OMe substituent exerted electron-withdrawing effects because the postulated electron-donating properties would have decreased the N3H acidity and resulted in higher p*K*a values (23). Other investigated substituents were not critical for the ionization properties of the uracil and 2-thiouracil ribosides. As described earlier, the 5-methyl substitution of uridine lead to an increase in the p*K*a value, by ca. 0.4 unit (63). This effect was observed for m5S2U (**1b**) and m5U (**2b**), for which the respective p*K*a values were 8.45 and 9.54. Decreased acidity was found for **1d** and **2d** (the p*K*a values were 8.69 and 9.79, respectively; an increase by ca. 0.6 unit in comparison with the values for **1a** and **2a**, respectively) bearing a negatively charged carboxymethyl (-CH_2COO^-) side chain. This electron-donating effect was abolished by the conversion of the carboxymethyl substituent into neutral species such as methyl ester (**1c** and **2c**) or amide (**1e** and **2e**).

The p*K*a values for the conjugated acids (protonated at the N3 function) of 4-pyrimidinone nucleosides **3a-e** and **3i** ranged from 2.0 to 2.8, and the values were still lower for **3f** bearing the methylaminomethyl substituent at C5 and for **3j**, the *S*-methyl derivative of S2U (1.63 and 1.78, respectively). Due to their limited stability under the present experimental conditions, the p*K*a values for **3g** and **3h** could not be determined.

The next proton-releasing sites are the acidic groups present in the derivatives of the **d**, **g,** and **h** series of compounds. The p*K*a values of the -CH_2COOH group in **1d** and **2d** were virtually identical (3.74 and 3.80, respectively), while a higher value of 4.31 was observed for the 4-pyrimidinone nucleoside **3d** (Table 1) (64,65). The p*K*a values determined for cmnm5S2U (**1g**) and cmnm5U (**2g**) were slightly higher than those reported in the literature, likely reflecting the different conditions under which the measurements were made. Although the sulfonic acid residues in **1h** and **2h** were the most acidic, their p*K*a values (2.50 and 2.41, respectively) were higher than that for taurine itself (1.5) (66).

Assessment of ionized fractions of nucleosides 1 and 2

Based on the p*K*a values for the dissociation of the N3H proton, we calculated the fractions of ionized **1** and **2** under physiological pH (7.4) using the Henderson-Hasselbalch equation: p*K*a-pH = log [BH]/[B^-], where BH and B^- are the neutral and ionized (deprotonated) forms, respectively (67). The results showed (Table 2) that the 2-thionucleoside units bearing the substituents that contain a positively charged aminomethyl group, as in **1f-h**, preferentially exist in the N3-deprotonated (ionized) form. Thus, at physiological pH **1f-h** predominantly adopt the zwitterionic form (ZI), being positively charged at the aminoalkyl side chain and deprotonated at the N3-function. In the uridine series **2**, the most abundant zwitterion was found for 5-taurinomethyluridine **2h** (43%). 2-Thiouridines **1c** and **1e**, with R5 substituents incapable of protonation were ionized only in 18 and 24%, respectively, similarly to 2-thiouridine (**1a**, 17%). Interestingly, the ionized fraction of the mo5S2U thio-nucleoside (**1i**) was relatively high, at 42%, although only ca. 7% of its 2-oxo analog (mo5U) was ionized. The uridines **2a-e** exist predominantly in their canonical, uncharged forms.

Since the p*K*a values for the N3H functions in pyrimidine nucleosides are lower by ca. 0.4 unit than in the corresponding nucleotides (bearing a negative charge on the phosphate group (25)), we recalculated the content of the ionized fraction of the corresponding nucleotides using the measured p*K*a values increased by 0.4 unit (Table 2, data given in brackets).

Table 1. The p*K*a (SD = ± 0.01) values (determined at 25°C) for the dissociation of the N3H proton and for the deprotonation of other functional groups (underlined or double-underlined) present in the C5-side chains. For compound **3**, the proton-conjugated structure is given. The p*K*a values already reported in the literature are given in brackets. Nd – not determined

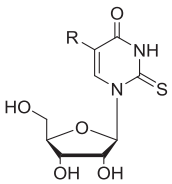
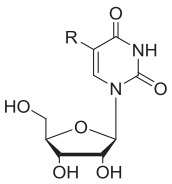
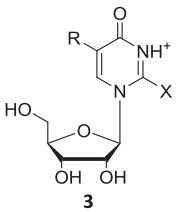
Compound	R (abbreviated name of the substituent)	Nucleoside			
		 **1**	 **2**	 **3**	
a	H	(N3) 8.09	(N3) 9.15	X=H	(N3) 2.07
b	CH_3 (m)	(N3) 8.45	(N3) 9.51 (9.7 (63))	H	(N3) 2.20
c	CH_2COOCH_3 (mcm)	(N3) 8.05	(N3) 9.15	H	(N3) 2.38
d	CH_2COOH (cm)	(N3) 8.69 (8.55 (64)) 3.74 (3.64 (64))	(N3) 9.79 (9.58 (61)) 3.80 (3.76 (61))	H	(N3) 2.80 4.31
e	CH_2CONH$_2$ (ncm)	(N3) 7.91 11.02	(N3) 9.10 11.68	H	(N3) 2.75 11.30
f	$CH_2NH_2^+CH_3$ (mnm)	9.51 (N3) 7.28	10.02 (N3) 8.15	H	8.66 (N3) 1.63
g	$CH_2NH_2^+CH_2$COOH (cmnm)	9.10 (8.93 (61)) (N3) 7.36 (7.28 (61)) 2.50 (2.33 (61))	10.13 (10.18 (65)) (N3) 8.24 (8.15 (65)) 3.05 (2.13 (65))	H	Nd
h	$CH_2NH_2^+(CH_2)_2SO_3H$ (τm)	9.04 (N3) 7.10 2.50	9.71 (N3) 7.51 2.41	H	Nd
i	OCH_3 (mo)	(N3) 7.56	(N3) 8.54	H	(N3) 2.10
j	H	-	-	S-CH_3	(N3) 1.78
k	H	-	-	S-$C_{10}H_{17}$	(N3) 2.01

Table 2. The fraction of nucleosides **1a-i** and **2a-i** with ionized N3H at pH 7.4, as calculated according to the Henderson–Hasselbalch equation

Cmpd.	**R**	Ionized nucleoside fraction [%]	
		for **1**	for **2**
a	H	17	2
b	CH_3	8	0.7
c	CH_2COOCH_3	18	1.7
d	CH_2COOH	7	0.4
e	CH_2CONH_2	24	2
f	CH_2NHCH_3	**57 (34)**	15
g	CH_2NHCH_2COOH	**52 (30)**	13
h	$CH_2NH(CH_2)_2SO_3H$	**67 (44)**	**43 (24)**
i	OCH_3	**42 (22)**	7

The p*K*a values of the nucleosides used are listed in Table 1.

Theoretical calculations

Gibbs free energies of tautomers of 1-methyl-uracil and 1-methyl-2-thiouracil and their C5-substituted derivatives. For Gibbs free energy computations, 1-methyl 5-substituted uracils (m1R5Ura) **4a,f,i** and 1-methyl 5-substituted 2-thiouracils (m1R5S2Ura) **5a,f,i** (Figure 3A) were used as models of nucleosides, in which the ribose moieties were 'reduced' to methyl groups to lower the computational cost. The Gibbs energies (G) of the diketo (K) and two keto-enol tautomeric forms (E2 and E4) were calculated for the gas phase and for an aqueous solution. In all cases, the K forms of the m1R5Ura and m1R5S2Ura series show the lowest

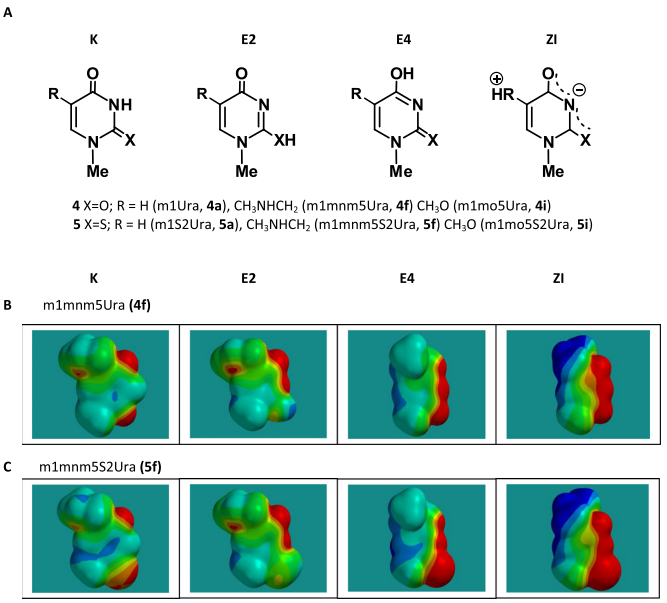


Figure 3. (**A**) The possible tautomeric diketo (K), and E2 and E4 keto-enol structures of 5-substituted 1-methyl uracils **4a,f,i** and their 2-thio analogs **5a,f,i**, for R = H (**a**); mnm (**f**) or mo (**i**) and zwitterionic forms (ZI) of **4f** and **5f** (R = mnm); (**B**) the electrostatic potential map of **4f** in K, E2, E4 and ZI form (**C**) the electrostatic potential map of **5f** in K, E2, E4 and ZI form; the blue and red colors indicate the most positive (electron-deficient) and the most negative (electron-rich) regions in the molecule, respectively.

free energy values, and were taken as the references (G_K). For the E2 and E4 tautomers of each compound, as well as for the corresponding zwitterions (ZI, Figure 3A), the ΔG_{rel} values were calculated using the expression: $\Delta G_{rel} = G\text{-}G_K$. The results thus obtained, presented in Table 3, confirm the earlier conclusions (obtained for **4a** and **5a**) on the higher relative stability of the diketo-tautomer compared to that of the E2 and E4 tautomers (68–70). Among the E2 and E4 tautomers, the former was the least stable (the highest values of ΔG_{rel}). Similar data were obtained for the remaining R5-substituted models **4f,i** and **5f,i** (R = mnm, mo). The corresponding relative electronic energies of tautomers at 25°C (298 K) are given in Supplementary Table S1.

Charge analysis of the zwitterionic forms of **4f** and **5f** revealed that the negative charge was delocalized over the electronegative centers O2/S2, N3 and O4 (Supplementary Figure S3). The ZI forms of **4f** and **5f** had very high ΔG_{rel} energy in the gas phase, but the calculations performed for the aqueous solutions showed that ZI were effectively stabilized by solvation and therefore are energetically much less demanding (Table 3). The dipole moments of zwitterionic structures are significantly larger than those in other tautomers (Supplementary Table S1). Therefore, the zwitterions are much more effectively stabilized by hydration. Taking into account that due to the deficiency of the continuum solvation method the free energies of solvation of these forms may be underestimated, their existence in solution may be more favored than it is suggested by the ΔG calculations. Certainly, the content of individual tautomers as well as the free energies of the protonated complex formation could change depending on the pH and temperature.

The other ionic forms likely to exist in solution in considerable concentrations are the protonated m1mnm5UraH$^+$ and m1mnm5S2UraH$^+$. Taking the Gibbs free energy of proton hydration $\Delta G_{hydr}(H^+) = -265.9$ kcal/mol (58), the Gibbs free energies of protonation of m1mnm5UraH$^+$ and m1mnm5S2UraH$^+$ in water were estimated as −1.9 and −0.4 kcal/mol, respectively.

The electrostatic potential maps for the K, E2, E4 and ZI forms of m1mnm5Ura **4f** and m1mnm5S2Ura **5f**, are shown in Figure 3B and C, respectively. The forms K have electron-rich regions in the vicinity to both O2/S2 and O4 atoms, while N3 is shielded by the hydrogen atom. In the forms E2 and E4 the electron-rich regions are O4…N and O2/S2…N, respectively. Notably, the zwitterionic forms of **4f** and **5f** clearly have different charge distribution. The electron-deficient region is located in the vicinity to the ammonium cation while the electron-rich region is dispersed over the O2/S2…N3…O4 edge. The electrostatic potential maps are consistent with atomic charge distributions of the corresponding tautomers (Supplementary Figure S3).

Stability of base pairs of tautomeric/zwitterionic forms of 5-substituted 1-methyl-uracil and 1-methyl-2-thiouracil with 9-methyl-guanine. The structure of nucleic acids is determined by several forces such as hydrogen bonding between nucleobases, aromatic π-stacking, base-backbone and backbone–backbone interactions as well as interactions with solvent molecules, metal ions, and other co-solutes. Among them, hydrogen-bonded base pairs substantially affect the overall structure and are decisive for thermal stability of nucleic acids (71). It should be mentioned that the stacking arrangement between consecutive bases in DNA and RNA/DNA double helices can enhance their hydrogen bonding ability, compared to the gas phase optimized complexes (72). In search of energetically favored structures displaying base pairing between the R5U/R5S2U nucleosides and guanosine, the enthalpies of hydrogen-bonded complex formation by the K, E2 and E4 tautomers of **4a,f,i** and **5a,f,i** as well as by the zwitterions of **4f** and **5f** shown in Figure 3, with 9-methyl-guanine (in the most stable keto form) were calculated (see Figure 4 and Table 4). The total interaction enthalpies at 25°C (ΔH^{298}) were calculated relative to the fully optimized bases. Geometries of the base pairs were optimized without any constrains, according to the standard approach (71). For a given base pair, ΔH^{298} was calculated according to the following equation:

$$\Delta H^{298} = H^{298}(\mathbf{U-G}) - (H^{298}(\mathbf{U}) + H^{298}(\mathbf{G})) + \mathrm{BSSE}$$

where H^{298}(**U–G**) is the enthalpy of the optimized **U–G** base pair, H^{298}(**U**) and H^{298}(**G**) are the enthalpies of the isolated and optimized **U** and **G** bases used in these studies, that is **U** = m1R5Ura/m1S2Ura and **G** = m9Gua in their most stable ('canonical') tautomeric forms. Thus, for the $\mathbf{U_K}$**-G** complex, the given ΔH value is simply the enthalpy of binding of the K tautomer of m1R5Ura/m1R5S2U with 9-methyl-guanine, while for the $\mathbf{U_{E2}}$**-G**, $\mathbf{U_{E4}}$**-G** and $\mathbf{U_{ZI}}$**-G** complexes, the given ΔH values include also the enthalpy of pre-structurization of the corresponding most stable K-tautomer into the higher energy E2, E4 or ZI forms. In some cases this procedure results in positive interaction enthalpy (when the tautomerization energy is higher than the hydrogen bonding energy). However, it allows the direct comparison of the stabilities of various complexes of a particular uracil/2-thiouracil derivatives with 9-methyl-guanine. The obtained ΔH^{298} values of these complexes are shown in Table 4. The ESP atomic charge distributions in the example base pairs of $\mathbf{UH^+_K}$**-G** and $\mathbf{U_{ZI(2,3)}}$**-G** for **4f** and **5f** are given in Supplementary Figure S4a and b, respectively.

The deformation enthalpy (which is the enthalpy required to adjust the isolated and relaxed bases to the geometry they adopt in the base pair) is ignored. However, for most base pairs the optimization led to the structures which are fairly close to planarity. Some non-standard pairings ($\mathbf{U_E}$**2G**, $\mathbf{U_{E4}}$**-G** and $\mathbf{U_{ZI(2,3)}}$**-G**) tended to adopt twisted geometries. The $\mathbf{U_{ZI(2,3)}}$**-G** forms are twisted by ca. 30° because of repulsive interactions between O4 of m1mnm5Ura and O6 of m9Gua. These base pairs showed also considerably reduced interaction enthalpies. Due to the base stacking and steric reasons, the base pairs in the duplexes are probably 'pushed' toward more planar conformations, which result in the additional reduction of interaction energy (73). Interestingly, for 2-thiouracil derivatives (**5a,f,i**) the base pairs of their E2 and E4 tautomers with m9Gua were twisted, due to non-planarity of the NH_2 group in 9-methyl-guanine, while the same complexes of uracil derivatives (**4a,f,i**) are almost perfectly planar (the geometries of tautomers and base pair complexes studied are available from the authors upon request). At this level of approximation, we did not study the effects of base stacking nor other

Table 3. The Gibbs free energies (ΔG_{rel}) for the lowest-energy E2 and E4 tautomers of m1Ura (**4a**), m1mnm5Ura (**4f**) and m1mo5Ura (**4i**) and their 2-thioanalogs (**5a,f,i**), as well as for the zwitterions of **4f** and **5f** at 25°C (298 K), in a gas phase and in water, relative to the energies of the respective references, the diketo forms (K) of 5-substituted 1-methyl uracils and 2-thiouracils

m1R5Ura/m1R5S2Ura	ΔG_{rel} in gas phase (kcal/mol)[a]					
	E2		E4		ZI[d]	
	O2 (**4**)	S2 (**5**)	O2(**4**)	S2(**5**)	O2(**4**)	S2(**5**)
R = H (**a**)	20.1 (19.2 (68))	17.0 (21.0 (69)), (16.3[c] (70))	11.7 (12.5 (68))	12.2 (12.5 (69)), (11.6[c] (70))		
R = CH_2NHCH_3 (**f**)	19.2	16.1	13.0	13.8	42.0	42.1
R = OCH_3 (**i**)	18.0	14.7	13.6	13.8		
	ΔG_{rel} in water (kcal/mol)[b]					
	O2	S2	O2	S2	O2	S2
R = H(**a**)	16.2 (16.0 (68))	14.4 (14.8 (69)), (18.6[c] (70))	10.8 (8.8 (68))	10.2 (10.6 (69)), (10.1[c](70))		
R = CH_2NHCH_3(**f**)	15.7	12.1	6.3	4.3	6.3	4.8
R = OCH_3 (**i**)	14.7	11.2	12.1	11.2		

[a]calculated with B3LYP-GD3/6-311++G(3df,2p)//B3LYP/6-31G(d).
[b]calculated with CPCM-B3LYP-GD3/6-311++G(3df,2p)//B3LYP/6-31+G(d).
[c]Electronic energy differences (ΔE) were reported; these values can slightly differ from the corresponding differences of Gibbs free energy.
[d]Notably, the continuum solvent model used in this study does not account for specific interactions, such as individual hydrogen bonds, between the solute and solvent molecules. For those instances, where such interactions are important (for example, in the case of ionic solutes), explicit solvent methods are better suited, where a cage of solvent molecules is constructed around the solute molecule. These methods can more accurately describe the solvent-solute interactions. However, construction of the realistic solvent cages would be difficult and time consuming.

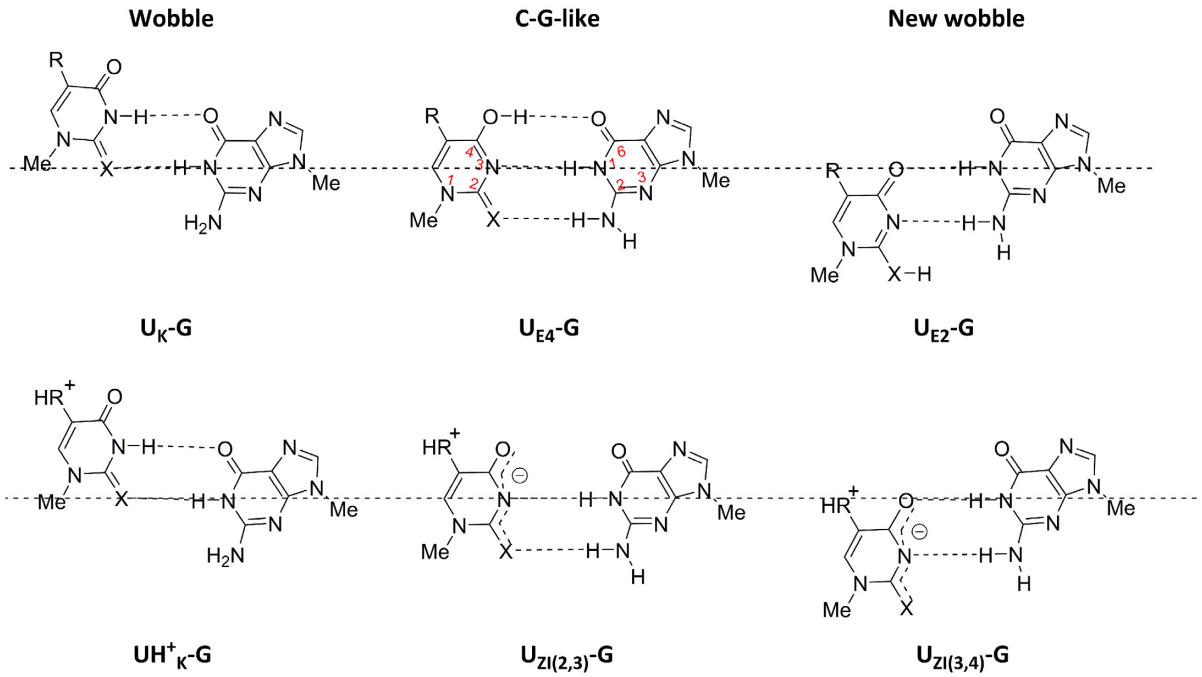


Figure 4. General scheme of the possible complexes between tautomers/zwitterions of m1R5Ura/m1R5S2Ura and m9Gua assembled according to wobble, Watson-Crick (C-G-like) and 'new wobble' base pairing mode. The numbering of the acceptor and donor ligands is shown in a Watson–Crick-like complex. X = O or S.

factors contributing to the structure and stability of nucleic acid duplexes.

The obtained data confirm that the typical U_K–G wobble base pairs with the stable keto form are thermodynamically the most favored among all investigated models. Evidently, replacement of the C2 oxygen atom in **4a,f,i** by a sulfur atom renders the S2…HN1 hydrogen bond weaker, as indicated by the finding that the corresponding enthalpies of formation of **U_K–G** complexes with 2-thiouracil derivatives **5a,f,i** are by 1.1–1.9 kcal/mol higher (less negative) (74–76). The protonated complexes of **4f** and **5f** (the **UH^+_K-G**-type) seem to be slightly stronger (the enthalpies are systematically more negative by ca. 0.3 and 0.5 kcal/mol for 2-oxo and 2-thiouracil, respectively) than the corresponding uncharged **U_K-G** complexes, because the positive charge in m1mnmH^+5Ura and m1mnmH^+5S2U is better stabilized in the complex. The formation of **U_{E2}-G** complexes by E2 tautomers of **4a** and **5a** lead to enthalpy values that were 12–17 kcal/mol higher than that for the reference wobble **U_K-G** complexes. This is the result of high energy of E2 tautomers (see Table 3) which is included in the overall energy of the **U_{E2}-G** complex formation. In contrast, the E4 tautomers, capable to form three hydrogen bonds with guanine (**U_{E4}-G** type) produce stable Watson-Crick C-G like

Table 4. Enthalpies of formation (in kcal/mol) for the complexes of guanine and modified uracil in water as calculated using the CPCM-B3LYP-GD3/6-311++G(3df,2p)//B3LYP/6-31+G(d) method

Base pair mode	Numbers of H-bond donor/acceptor atoms	ΔH^{298} (kcal/mol) of a base pair of m9Gua with m1R5Ura/m1R5S2Ura					
		m1Ura (**4a**)	m1S2Ura (**5a**)	m1mnm5Ura (**4f**)	m1mnm5S2Ura (**5f**)	m1mo5-Ura (**4i**)	m1mo5S2Ura (**5i**)
U_K–G	3-6/2-1	– 10.0	– 8.1	– 10.2	– 8.4	– 11.2	– 10.1
UH^+_K–G	3-6/2-1			– 10.5	– 8.9		
U_{E2}–G	4-1/3-2	7.2	4.1	5.4	2.8	5.7	2.6
U_{E4}–G	4-6/3-1/2-2	**– 7.8**	**– 6.6**	**– 7.6**	– 6.4	**– 8.0**	**– 7.7**
$U_{ZI(2,3)}$–G	3-1/2-2			– 5.3	– 5.6		
$U_{ZI(3,4)}$–G	4-1/3-2			– 5.9	**– 7.3**		

The schemes of the corresponding base pairs are shown in Figure 4 and the numbering of the H-bond acceptor and donor atoms is adapted from the UE4-G base pair. The enthalpies of the most plausible alternative structures of the U–G base pairs are given in bold.

complexes with enthalpy gain by ca. 2–3 kcal/mol lower than that of the **U_K-G** reference. Here, 2-thiouracils were slightly worse partners than uracils (by ca. 1.2 kcal/mol), except for 5-methoxy-substituted models where the difference between ΔH of **U_{E4}-G** complexes for **4i** and **5i** occurred negligible (0.3 kcal/mol). This is because the S…HN hydrogen bond is weaker than the O…HN one (18,19). The most notable were the 'new wobble' complexes of ZI forms of m1mnm5Ura (**4f**) and m1mnm5S2Ura (**5f**). Here, the enthalpy of formation of the **$U_{ZI(3,4)}$-G**-type complex for **5f**, was −7.3 kcal/mol, which was close to the ΔH value for the same pair in the protonated **UH^+_K-G**-type complex (ΔΔH = 1.6 kcal/mol), whereas for the **$U_{ZI(3,4)}$-G**-type complex for **4f** the enthalpy gap was bigger (−5.9 versus −10.5 kcal/mol). Notably, the enthalpy gains for the **$U_{ZI(2,3)}$-G**-type complexes of the zwitterionic **4f** and **5f** were found to be smaller.

DISCUSSION

Modified nucleosides present in the tRNA regulate protein translation in a highly dynamic manner (77–79). Forty of these nucleosides are modified uridines located in the wobble position of tRNA anticodon. Their function is to tune mRNA codons reading through enhanced specificity of interactions between the modified unit in the tRNA wobble position and the third letter of the codon. The modified U34 base pairings are supported by interactions created within ribosome, involving modifying elements like the hydrogen bonds with the U34 side-chain functions, π-stacking, sugar conformational effects and rearrangement of the donor–acceptor sites within the uracil ring. It has been suggested that the binding of hypermodified 5-substituted uridines/2-thiouridines to adenosine (recognition of the 5′-NNA-3′ codons by the Watson-Crick base pairing) or to guanosine (recognition of the 5′-NNG-3′ codons by the wobble type base pairing) is regulated by the presence of two modifying elements: (i) a sulfur atom at the C2 position and (ii) a substituent at the C5 position of uracil. The 2-thiouridines exhibit enhanced hybridization with adenosine while their pairings with guanosine are much weaker due to the less effective hydrogen bonding between the N1H donor of guanine and the sulfur acceptor of 2-thiouracil (18,19) (Figure 1). Moreover, accommodation of such the U34–G wobble base pair (employing the most stable keto tautomers of U and G) on the ribosome is disadvantageous, and would require a shift of modified U toward the major groove of the codon-anticodon mini-helix (Figure 5A). This displacement is not plausible due to the steric constraints of the 'wobble' site with the third codon letter fixed at the A-site of the ribosome (80–82). In contrast, when guanosine is located in the wobble position of the tRNA anticodon, the respective G34-U wobble base pair (Figure 5B) (non-isosteric to U34–G) (3,82) is well accommodated at the ribosome, with the nucleoside 34 shifted toward the minor groove. Thus, as confirmed by experimental data, the binding of R5U34 / R5S2U34 with G in the third codon position can be realized rather by alternative, pre-structured uracil-guanine base pairing employing the tautomeric or zwiterionic form of modified units (Figure 4). Formation of such pre-structured forms of modified nucleosides is determined by influence of their modified elements on the electronic density of the uracil ring that results in changes of the acidity of their N3H proton.

We used a series of modified/hypermodified uridines (**2**) and 2-thiouridines (**1**) and analyzed their p*K*a values (Table 1) and the relative abundance of their ionized fractions in aqueous solutions at the physiological pH (Table 2). For majority of 2-thiouridines **1**, the p*K*a values of the N3H function are found to be lower by one unit than that of the parent uridines **2**, indicating the key involvement of the sulfur atom on the increase of the N3H proton acidity (easier formation of an anionic form, facilitated tautomeric rearrangement). Furthermore, for **1f-h** and **2f-h**, with mnm, cmnm and τm C5-substituents, the p*K*a values of their aminoalkyl groups exceed 9. Thus, at physiological pH, these substituents are quantitatively protonated to become the electron-withdrawing groups and further facilitate the departure of the N3H proton. The results of our calculations (based on the measured p*K*a values) show that at pH 7.4, the fraction of the N3-ionized forms of the 5-aminoalkyl-modified 2-thiouridines **1f**, **1g** and **1h** was 55, 52 and 67%, respectively. Since the p*K*a values of N3H in uridines located in an RNA chain are higher by ca. 0.4 unit (25), the content of their ionized forms might be de-

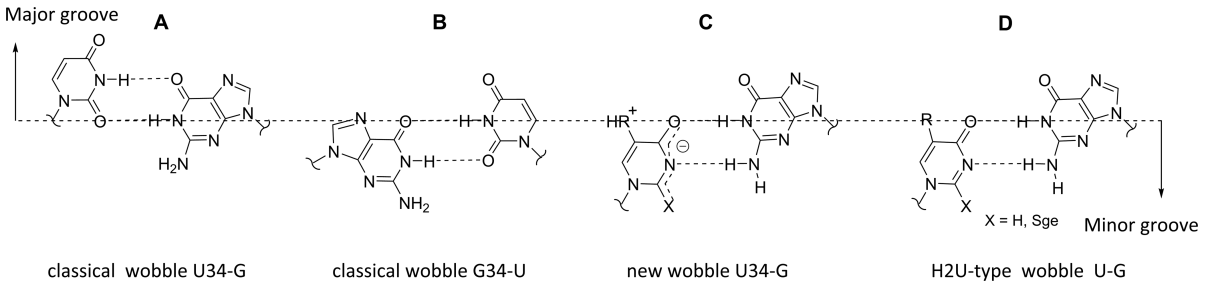


Figure 5. The spatial arrangement of the classical wobble U34–G (**A**) and G34-U (**B**), new wobble U34–G (**C**) (representing mnm5S2U–G base pair) and H2U–G or geS2U–G (R = (c) mnm) (**D**) in a mode analogous to the new wobble base pairs. The pyrimidine residue in (C) and 4-pyrimidinone residue in (D) are shifted toward the minor groove of the codon-anticodon mini-helix, allowing to accommodate at the ribosome in the G34 cavity. For (C) R = substituent with an aminoalkyl group (as in **1f-h**, Table 1), for (D) R = mnm or cmnm.

creased by ca. 20%, but is still significant (Table 2, data in brackets). Importantly, τm5U (**2h**) bearing the 2-oxo function was also substantially ionized (43%). Thus, the taurinomethyl side chain appears to be the main determinant of the electronic structure of **1h** and **2h** units in mitochondrial tRNAs. Incorrect maturation of the mt-tRNA often results in the absence of taurine modification in the U34/S2U34, leading to the impaired reading of the G unit at the 3′-end of the codon and ultimate development of mitochondrial disorders (78). These severe biological alterations might be related to the loss of ability of the modified uridine 34 to adopt the alternative pre-structured/tautomeric form allowing for the non-standard base pairing (29). This hypothesis, based on the contribution of the alternative tautomeric forms of U34 in efficient reading of guanosine in the third position of the codon, gets some support from our theoretical calculations (Table 4), in which we assessed the stability of selected model base pairs formed between various forms of m1R5Ura and m9Gua. Although the classical $\mathbf{U_K}$**-G** wobble base pairs (Figure 4 and Table 4) are the most stable among all investigated base pairs, their accommodation at the ribosome, as mentioned above, is not allowed due to the controlled size and shape of the wobble base pair cavity (compare Figure 5A and B) (3,26,80–82). Regardless the presence of modification at the position 5 of the heterocyclic ring, the E4 tautomers of the wobble R5U/R5S2U units offer the energetically advantageous base pairing with G ($\mathbf{U_{E4}}$**-G** type, Figure 4, Table 4, data in bold), possible for all kinds of the analyzed modifications. This happens due to a lower energy cost of the formation of E4 versus E2 (Table 3), further supported by creation of three hydrogen bonds in the C-G-like base pair. The exception is 5-methylaminomethyl substituted m1S2U (**5f**), for which the zwitterionic form is almost as stable as its E4 form (ΔG_{rel} in water = 4.8 versus 4.3 kcal/mol, respectively, Table 3) and this ZI form gives with m9Gua energetically more stable $\mathbf{U_{ZI(3,4)}}$**-G** base pair ($\Delta H = -7.3$ kcal/mol) than the E4 form ($\Delta H = -6.4$ kcal/mol for $\mathbf{U_{E4}}$**-G** base pair). These data indicate the highest stability of this $\mathbf{U_{ZI(3,4)}}$**-G** model base pair (corresponding to the mnm5S2U–G base pair) in a new wobble arrangement (called also 'reverse wobble' (26)), which was previously suggested (4,25,83), and recently found at the ribosome context (Supplementary Figure S5) (26).

Notably, in the pre-structured form of mnm5S2U (as in $\mathbf{U_{ZI(3,4)}}$**-G** base pair, Figure 5C) the nucleobase resembles a 4-pyrimidinone residue (Figure 5D). In our earlier studies on the oxidative desulfuration of 2-thiouridine, we analyzed such 4-pyrimidinone models at the nucleoside (H2U, Figure 5D, X = H, R = H) and RNA oligonucleotide (H2U-RNA) levels (84). With the help of UV-melting and DSC (differential scanning calorimetry) experiments, and theoretical modeling we have earlier demonstrated that the isolated H2U–G base pair is thermally as stable as the isolated U–G wobble base pair ($\Delta H = 10.0$ versus 9.5 kcal/mol) (4,84). The H2U-type uridines, analogous to those present in the $\mathbf{U_{ZI(3,4)}}$**-G** base pair, assure efficient binding (and recognition) of the complementary guanosines within the RNA duplex. The H2U model has identical donor/acceptor pattern as the *S*-geranyl-2-thiouridine (49) discovered *in vivo* (Figure 5D, X = geS, R = (c)mnm), whose binding mode with G is the same as that in H2U–G and in mnm5S2U–G new wobble base pairs (84–86). Moreover, translation efficiency of the glutamate 5′-GAG-3′ over 5′-GAA-3′ codon with geranylated tRNA was shown to be increased (49).

As proposed earlier (4,25,77–78,83), the ZI form of mnm5S2U interacting with a guanosine residue has been found recently in the crystal structure of *Escherichia coli* $tRNA^{Lys}$ complexed with 70S ribosome and a short mRNA fragment (26) (Supplementary Figure S5A). The presence of a 2-thio modification together with a C5-aminoalkyl substituent appears to be essential for the formation of this novel type wobble U34-G base pair. This conclusion is supported by other studies (30) demonstrating that the lack of a sulfur atom in mnm5U evidently changes its pairing with G. Thus, in the crystal of the anticodon stem-loop tRNA fragment bound to the 30S ribosomal subunit with a short mRNA fragment, the mnm5U-G base pair is proposed to be bifurcated (Supplementary Figure S5B), with two hydrogen bonds from N1H and N2H of G directed toward the oxygen O2 of mnm5U. Although both these base pairs (of mnm5U or mnm5S2U with G) differ only by the heteroatom at position 2, their geometries are different; mnm5U adopts rather the Watson-Crick C-G like geometry (as in $\mathbf{U_{E4}}$**-G**, Figure 4), while the pre-structured mnm5S2U preferentially interacts with G in the $\mathbf{U_{ZI(3,4)}}$**-G** mode and the pyrimidine unit is shifted toward the minor groove (Figures 4 and 5). Our theoretical calculations have shown that the $\mathbf{U_{E4}}$**-G** base pair for **4f** (an oxo-analog) is

more stable than of **5f** (a thio-analog) ($\Delta H = -7.6$ versus -6.4 kcal/mol), mostly because the H-bond with sulfur atom as a proton acceptor is weaker than the bond involving oxygen. In turn, in zwitterionic **5f** the charge distribution along the O4...N3...S2 edge is shifted toward O4, due to higher electronegativity of the oxygen atom compared to the sulfur atom (Supplementary Figure S3), while its electrostatic potential map shows potential distribution over bigger, more polarizable sulfur atom compared to O2 in the zwitterionic form of **4f** (Figure 3). Despite of this, the preferred mode of hydrogen bonding between ZI of **5f** and m9G is ascribed as $\mathbf{U_{ZI(3,4)}}$**-G** base pair and not $\mathbf{U_{ZI(2,3)}}$**-G**, for which the complex formation enthalpy is smaller ($\Delta H = -5.6$ versus -7.3 kcal/mol, Table 4).

Crystallographic studies of other examples of modified U34–G base pairs (identified in the crystal structures of tRNA/mRNA at the ribosomal environment) confirm their preferred C-G like alignment for mcm5SU (CH_2COOCH_3) (28), cmo (OCH_2COOH) (27) and τm5U ($CH_2NHCH_2CH_2SO_3H$) (29), modified units, except for the mentioned above mnm5S2U-G base pair found in the new wobble (reversed) mode (26). It should be pointed out that the resolution of these structures was not higher than 2.5 Å, so was not conclusive as to the conformation of the sugar ring in the modified units found in the wobble position. One may assume that the ribose in the E4 forms of uridine units (bound according to the C-G like mode) will adopt the C3'-*endo* sugar ring puckering, similar to their diketo (47) and 4-*O*-methyl-uridine forms (87). However, the zwitterionic forms in $\mathbf{U_{ZI(3,4)}}$**-G** base pairs will probably adopt preferentially the C2'-*endo* sugar ring conformation. Our earlier studies have shown that both sugar rings of 4-pyrimidinone models, that is of H2U and geS2U indeed in solution adopt C2'-*endo* puckering to a higher extent that their parent 2-thiouridine (47,88). Therefore, it seems that the 'wobble cavity' at the ribosome exhibits some spatial tolerance for accommodation of diverse U34–G base pairs.

CONCLUSIONS

In the present study, we analyzed the importance of the electron density/ionization features of wobble uridines to understand their biological properties. Our results suggest that the ionization features of the modified uridines U34 critical for the precise reading of genetic information are determined by the electronic character of the C5-substituents and by the presence of sulfur at C2 position. Our data offer an explanation for the biological and crystallographic observations regarding the presence of pre-structured forms of modified uridines and their contribution to the recognition of purines at the 3′-ends of the NNA and NNG codons.

SUPPLEMENTARY DATA

Supplementary Data are available at NAR Online.

ACKNOWLEDGEMENTS

Quantum chemical DFT calculations were supported by the PL-Grid Infrastructure. The authors thank Dr Piotr Guga for critical comments.

FUNDING

National Science Centre of Poland [UMO-2011/03/B/ST5/02669 2012–2015, UMO-2014/13/B/ST5/03979 2015–2018 to B.N., E.S.]; CMMS PAS Statutory funds; Lodz University of Technology, Poland. Funding for open access charge: National Science Centre of Poland UMO-2014/13/B/ST5/03979 2015–2018 to B.N., E.S.].

Conflict of interest statement. None declared.

REFERENCES

1. Crick,F.H.C. (1966) Codon-anticodon pairing: the wobble hypothesis. *J. Mol. Biol.*, **19**, 548–555.
2. Vendeix,F.A.P., Munoz,A.M. and Agris,P.F. (2009) Free energy calculation of modified base-pair formation in explicit solvent: A predictive model. *RNA*, **15**, 2278–2287.
3. Westhof,E. (2014) Isostericity and tautomerism of base pairs in nucleic acids. *FEBS Lett.* **588**, 2464–2469.
4. Sochacka,E., Szczepanowski,R.H., Cypryk,M., Sobczak,M., Janicka,M., Kraszewska,K., Bartos,P., Chwialkowska,A. and Nawrot,B. (2015) 2-Thiouracil deprived of thiocarbonyl function preferentially base pairs with guanine rather than adenine in RNA and DNA duplexes. *Nucleic Acids Res.*, **43**, 2499–2512.
5. Machnicka,M.A., Milanowska,K., Oglou,O.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.
6. Cantara,W.A., Crain,P.F., Rozenski,J., McCloskey,J.A., Harris,K.A., Zhang,X., Vendeix,F.A.P., Fabris,D. and Agris,P.F. (2011) The RNA modification database, RNAMDB: 2011 update. *Nucleic Acids Res.*, **39**, D195–D201.
7. 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.
8. Sheng,J., Larsen,A., Heuberger,B.D., Blain,J.C. and Szostak,J.W. (2014) Crystal structure studies of RNA duplexes containing s(2)U:A and s(2)U:U base pairs. *J. Am. Chem. Soc.*, **136**, 13916–13924.
9. Okamoto,I., Seio,K. and Sekine,M. (2006) Incorporation of 2′-O-methyl-2-thiouridine into oligoribonucleotides induced stable A-form structure. *Chem. Lett.*, **35**, 136–137.
10. Shohda,K., Okamoto,I., Wada,T., Seio,K. and Sekine,M. (2000) Synthesis and properties of 2′-O-methyl-2-thiouridine and oligoribonucleotides containing 2′-O-methyl-2-thiouridine. *Bioorg. Med. Chem. Lett.*, **10**, 1795–1798.
11. Sundaram,M., Durant,P.C. and Davis,D.R. (2000) Hypermodified nucleosides in the anticodon of $tRNA^{Lys}$ stabilize a canonical U-turn structure. *Biochemistry*, **39**, 12575–12584.
12. Testa,S.M., Disney,M.D., Turner,D.H. and Kierzek,R. (1999) Thermodynamics of RNA-RNA duplexes with 2- or 4-thiouridines: implications for antisense design and targeting a group I intron. *Biochemistry*, **38**, 16655–16662.
13. 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.
14. Davis,D.R. (1998) Biophysical and conformational properties of modified nucleosides in RNA (nucleic magnetic resonance studies). In: Grosjean,H and Benne,R (eds). *Modification and Editing of RNA*. ASM Press, Washington, pp. 85–102.
15. Agris,P.F., Sierzputowska-Gracz,H., Smith,W., Malkiewicz,A., Sochacka,E. and Nawrot,B. (1992) Thiolation of uridine carbon-2 restricts the motional dynamics of the transfer RNA wobble position nucleoside. *J. Am. Chem. Soc.*, **114**, 2652–2656.
16. Sierzputowska-Gracz,H., Sochacka,E., Malkiewicz,A., Kuo,K., Gehrke,C.W. and Agris,P.F. (1987) Chemistry and structure of modified uridines in the anticodon, wobble position of transfer RNA are determined by thiolation. *J. Am. Chem. Soc.*, **109**, 7171–7177.
17. Yamamoto,Y., Yokoyama,S., Miyazawa,T., Watanabe,K. and and Higuchi,S. (1983) NMR analyses on the molecular mechanism of the

conformational rigidity of 2-thioribothymidine, a modified nucleoside in extreme thermophile tRNAs. *FEBS Lett.*, **157**, 95–99.
18. Donohue,J. (1969) On N-H…S hydrogen bonds. *J. Mol. Biol.*, **45**, 231–235.
19. Wood,P.A., Pidcock,E. and Allen,F.H.(2008) Interaction geometries and energies of hydrogen bonds to C=O and C=S acceptors: a comparative study. *Acta Crystallogr. B.*, **64**, 491–496.
20. Krüger,M.K., Pedersen,S., Hagervall,T.G. and Sørensen,M.A. (1998) The modification of the wobble base of tRNAGlu modulates the translation rate of glutamic acid codons in vivo. *J. Mol. Biol.*, **284**, 621–631.
21. Hagervall,T.G., Pomerantz,S.C. and McCloskey,J.A. (1998) Reduced misreading of asparagine codons by Escherichia coli tRNALys with hypermodified derivatives of 5-methylaminomethyl-2-thiouridine in the wobble position. *J. Mol. Biol.*, **284**, 33–42.
22. Grosjean,H., de Crécy-Lagard,V. and Marck,C. (2010) Deciphering synonymous codons in the three domains of life: co-evolution with specific tRNA modification enzymes. *FEBS Lett.*, **584**, 252–264.
23. Egert,E., Lindner,H.J., Hillen,W. and Boehm,M.C. (1980) Influence of substituents at the 5 position on the structure of uridine. *J. Am. Chem. Soc.*, **102**, 3707–3713.
24. Clayden,J., Greeves,N., Warren,S. and Wothers,P. (2001) *Organic Chemistry*. Oxford University Press, Oxford, UK.
25. Takai,K. and Yokoyama,S. (2003) Roles of 5-substituents of tRNA wobble uridines in the recognition of purine-ending codons. *Nucleic Acids Res.*, **31**, 6383–6391.
26. 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 genetic code. *Nat. Commun*, **7**, 10457–10466.
27. Weixlbaumer,A., Murphy,F.V. IV, Dziergowska,A., Malkiewicz,A., Vendeix,F.A.P., Agris,P.F. and Ramakrishnan,V. (2007). Mechanism for expanding the decoding capacity of transfer RNAs by modification of uridines. *Nat. Struct. Mol. Biol.*, **14**, 498–502.
28. Vendeix,F.A.P., Murphy,F.V. IV, Cantara,W.A., Leszczyńska,G., Gustilo,E.M., Sproat,B., Malkiewicz,A. and Agris,P.F. (2012) Human tRNA(Lys3)(UUU) is pre-structured by natural modifications for cognate and wobble codon binding through keto-enol tautomerism. *J. Mol. Biol.*, **416**, 467–485.
29. Kurata,S., Weixlbaumer,A., Ohtsuki,T., Shimazaki,T., Wada,T., Kirino,Y., Takai,K., Watanabe,K., Ramakrishnan,V. and Suzuki,T. (2008) Modified uridines with C5-methylene substituents at the first position of the tRNA anticodon stabilize U.G wobble pairing during decoding. *J. Biol. Chem.* **283**, 18801–18811.
30. Murphy,F.V., 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.
31. Niedballa,U. and Vorbrüggen,H. (1974) Synthesis of nucleosides. 9. General synthesis of N-glycosides. I. Synthesis of pyrimidine nucleosides. *J. Org. Chem.*, **39**, 3654–3660.
32. Nawrot,B. and Sochacka,E. (2009) Preparation of short interfering RNA containing the modified nucleosides 2-thiouridine, pseudouridine, or dihydrouridine. *Curr. Protoc. Nucleic Acid Chem.*, Chapter 16, Unit 16.2.
33. Vorbrüggen,H. and Strehlke,P. (1969) Synthesis of nucleosides. II. Synthesis of Hall's methyl 2-thiouridine-5-acetate from yeast-tRNA. *Angew. Chem. Int. Ed. Engl.*, **8**, 977–978.
34. Baczynskyj,L., Biemann,K., Fleysher,M.H. and Hall,R.H. (1969) Synthesis of 2-thio-5-carboxymethyluridine methyl ester: a component of transfer RNA. *Can. J. Biochem.*, **47**, 1202–1203.
35. Malkiewicz,A. and Nawrot,B. (1987) The 2-thio analogs of tRNA components derived from 5-hydroxyuridine. *Z. Naturforsch.*, **42b**, 355–359.
36. Niedballa,U. and Vorbrüggen,H. (1974) A general synthesis of N-glycosides. IV. Synthesis of nucleosides of hydroxy and mercapto N-heterocycles. *J. Org. Chem.*, **25**, 3668–3671.
37. Niedballa,U. and Vorbrüggen,H. (1974) Synthesis of nucleosides. 10. General synthesis of N-glycosides. II. Synthesis of 6-methyluridines. *J. Org. Chem.*, **25**, 3660–3663.
38. Vorbrüggen,H. and Ruh-Pohlenz,C. (2000) Synthesis of nucleosides. In: Paquette,LA (ed.), *Organic Reactions*, J. Wiley and Sons, New York, Vol. **55**, pp. 1–741.
39. Ikeda,K., Tanaka,S. and Mizuno,Y. (1975) Syntheses of potential antimetabolites. XX. Syntheses of 5-carbomethoxymethyl- and 5-methylaminomethyl-2-thiouridine (the first letters of some anticodons) and closely related nucleosides from uridine. *Chem. Pharm. Bull.*, **23**, 2958–2964.
40. Reese,C.B. and Sanghvi,Y. (1984) The synthesis of 5-carboxymethylaminomethyluridine and 5-carboxymethylaminomethyl-2-thiouridine. *J. Chem. Soc. Chem. Commun.*, **1**, 62–63.
41. Malkiewicz,A., Sochacka,E., Ahmed,A.F. and Sayed,Y.S. (1983) The modified nucleosides from the 'wobble position' of tRNAs. The synthesis of 5-carboxymethylaminomethyluridine and 5-carboxymethylaminomethyl-2-thiouridine. *Tetrahedron Lett.*, **24**, 5395–5398.
42. Ogata,T., Shimazaki,T., Umemoto,T., Kurata,S., Ohtsuki,T., Suzuki,T. and Wada,T. (2009) Chemical synthesis and properties of 5-taurinomethyluridine and 5-taurinomethyl-2-thiouridine. *J. Org. Chem.*, **74**, 2585–2588.
43. Leszczynska,G., Pieta,J., Wozniak,K. and Malkiewicz,A. (2014) Site-selected incorporation of 5-carboxymethylaminomethyl-(2-thio)uridine into RNA sequences by phosphoramidite chemistry. *Org. Biomol. Chem.*, **12**, 1052–1056.
44. Bartosik,K. and Leszczynska,G. (2015) Synthesis of various substituted 5-methyluridines (xm5U) and 2-thiouridines (xm5s2U) via nucleophilic substitution of 5-pivaloyloxymethyluridine/2-thiouridine. *Tetrahedron Lett.*, **56**, 6593–6597.
45. Fissekis,J.D. and Sweet,F. (1970) Synthesis of 5-carboxymethyluridine. A nucleoside from transfer ribonucleic acid. *Biochemistry*, **9**, 3136–3142.
46. Sochacka,E. and Fratczak,I. (2004) Efficient desulphurization of 2-thiopyrimidine nucleosides to corresponding 4-pyrimidinone analogues using trans-2-phenylsulfonyl-3-phenyloxaziridine. *Tetrahedron Lett.*, **45**, 6729–6731.
47. Bartos,P., Ebenryter-Olbinska,K., Sochacka,E. and Nawrot,B. (2015) The influence of the C5 substituent on the 2-thiouridine desulfuration pathway and the conformational analysis of the resulting 4-pyrimidinone products. *Bioorg. Med. Chem.*, **23**, 5587–5594.
48. El-Tayeb,A., Qi,A., Nicholas,R.A. and Muller,C.E. (2011) Structural modifications of UMP, UDP, and UTP leading to subtype-selective agonists for P2Y2, P2Y4 and P2Y6 receptors. *J. Med. Chem.*, **54**, 2878–2890.
49. 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.*, **11**, 913–919.
50. 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.
51. Irving,H., Miles,M.G. and Pettit,L.D. (1967) A study of some problems in determining the stoichiometric proton dissociation constants of complexes by potentiometric titrations using a glass electrode. *Anal. Chim. Acta*, **38**, 475–481.
52. Gans,P., Vacca,A. and Sabatini,A. (1985) SUPERQUAD: An improved general program for computation of formation constants from potentiometric data. *J. Chem. Soc. Dalton Trans.*, 1195–1200.
53. Frisch,M.J., Trucks,G.W., Schlegel,H.B., Scuseria,G.E., Robb,M.A., Cheeseman,J.R., Scalmani,G., Barone,V., Mennucci,B., Petersson,G.A. *et al.* (2009) *Gaussian 09, Revision D.01*, Gaussian, Inc., Wallingford CT.
54. Becke,A.D. (1993) Density-functional thermochemistry. III. The role of exact exchange. *J. Chem.Phys.*, **98**, 5648–5652.
55. Grimme,S., Antony,J., Ehrlich,S. and Krieg,H. (2010) A consistent and accurate ab initio parameterization of density functional dispersion correction (DFT-D) for the 94 elements H-Pu. *J. Chem. Phys.*, **132**, 154104–154119.
56. Boys,S.F. and Bernardi,F. (1970) The calculation of small molecular interactions by the differences of separate total energies. Some procedures with reduced errors. *Mol. Phys.*, **19**, 553–566.
57. Cossi,M., Rega,N., Scalmani,G. and Barone,V. (2003) Energies, structures, and electronic properties of molecules in solution with the C-PCM solvation model. *J. Comput. Chem.*, **24**, 669–681.
58. Ho,J. and Coote,M.L. (2010) A universal approach for continuum solvent pKa calculations: are we there yet? *Theor. Chem. Acc.* **125**, 3–21.

59. Singh,U.C. and Kollman,P.A. (1984) An approach to computing electrostatic charges for molecules. *J. Comput. Chem.*, **5**, 129–145.
60. *Spartan'08*. Wavefunction, Inc., Irvine, CA.
61. Kowalik-Jankowska,T., Várnagy,K., Świątek-Kozłowska,J., Jon,A., Sóvágó,I., Sochacka,E., Malkiewicz,A., Spychała,J. and Kozłowski,H. (1997) Role of sulfur site in metal binding to thiopurine and thiopyrimidine nucleosides. *J. Inorg. Biochem.*, **65**, 257–262.
62. Knobloch,B., Linert,W. and Sigel,H. (2005) Metal ion-binding properties of (N3)-deprotonated uridine, thymidine, and related pyrimidine nucleosides in aqueous solution. *PNAS*, **102**, 7459–7464.
63. Saenger,W. (1984) *Principles of Nucleic Acid Structure*, Springer-Verlag, New York, NY.
64. Kowalik-Jankowska,T., Kozłowski,H., Savago,I., Nawrot,B., Sochacka,E. and Malkiewicz,A. (1994) Metal binding ability of hypermodified nucleosides. II Copper (II) complexes of uridine and 2-thiouridine derivatives. *J. Inorg. Biochem.*, **53**, 49–56.
65. Horie,N., Yamaizumi,Z., Kuchino,Y., Takai,K., Goldman,E., Miyazawa,T., Nishimura,S. and Yokoyama,S. (1999) Modified nucleosides in the first positions of the anticodons of $tRNA_4$Leu and $tRNA_5$Leu from Escherichia coli. *Biochemistry*, **38**, 207–217.
66. Dawson,R.M.C., Elliott,D.C., Elliott,W.H. and Jones,K.M. (1986) In: *Data for Biochemical Research*, 3rd edn., Oxford University Press, New York, NY, pp. 28.
67. de Levie,R. (2003) The Henderson-Hasselbalch Equation: Its History and Limitations. *J. Chem. Educ.*, **80**, 146.
68. Rejnek,J., Hanus,M., Kabelac,M., Ryjacek,F. and Hobza,P. (2005) Correlated ab initio study of nucleic acid bases and their tautomers in the gas phase, in a microhydrated environment and in aqueous solution. Part 4. Uracil and thymine. *Phys. Chem. Chem. Phys.*, **7**, 2006–2017.
69. Babu,N.S. (2013) Theoretical study of stability, tautomerism, equilibrium constants (pKT) of 2-thiouracil in gas phase and different solvents (water and acetonitrile) by the density functional theory method. *Am. Chem. Sci. J.*, **3**, 137–150.
70. Yekeler,H. (2000) Ab initio study on tautomerism of 2-thiouracil in the gas phase and in solution. *J. Comput. Aid. Mol. Des.*, **14**, 243–250.
71. Šponer,J., Jurečka,P. and Hobza,P. (2004) Accurate interaction energies of hydrogen-bonded nucleic acid base pairs. *J. Am. Chem. Soc.* **126**, 10142–10151.
72. Mignon,P., Loverix,S., Steyaert,J. and Geerlings,P. (2005) Influence of the π–π interaction on the hydrogen bonding capacity of stacked DNA/RNA bases. *Nucleic Acids Res.* **33**, 1779–1789.
73. Mizuno,H. and Sundaralingam,M. (1978) Stacking of Crick wobble pair and Watson-Crick pair: stability rules of G-U pairs at ends of helical stems in tRNAs and the relation to codon-anticodon wobble interaction. *Nucleic Acids Res.* **15**, 4451–4461.
74. Yokoyama,S. and Nishimura,S. (1995) Modified nucleosides and codon recognition. In: Söll,D and RajBhandary,U (eds.), *tRNA: Structure, Biosynthesis, and Function*. American Society for Microbiology, Washington, DC, pp. 207–223.
75. Curran,J.F. (1998) Modified nucleosides in translation. In: Grosjean,H and Benne,R (eds.), *Modification and Editing of RNA*. ASM Press, Washington, DC, pp. 493–516.
76. Watanabe,K. (2007) Role of modified nucleosides in the translation function of tRNAs from extreme thermophilic bacteria and animal mitochondria. *Bull. Chem. Soc. Japan*, **80**, 1253–1267.
77. Nawrot,B., Sochacka,E. and Düchler,M. (2011) tRNA structural and functional changes induced by oxidative stress. *Cell Mol. Life Sci.*, **68**, 4023–4032.
78. Düchler,M., Leszczynska,G., Sochacka,E. and Nawrot,B. (2016) Nucleoside modifications in the regulation of gene expression: focus on tRNA. *Cell Mol. Life Sci.*, **73**, 3075–3095.
79. Shigi,N. (2014) Biosynthesis and functions of sulfur modifications in tRNA. *Front. Genet.*, **5**, 67.
80. Demeshkina,N., Jenner,L., Westhof,E., Yusupov,M. and Yusupova,G. (2012) A new understanding of the decoding principle on the ribosome. *Nature*, **484**, 256–259.
81. Demeshkina,N., Jenner,L., Westhof,E., Yusupov,M. and Yusupova,G. (2013) New structural insights into the decoding mechanism: translation infidelity via a G-U pair with Watson-Crick geometry. *FEBS Lett.*, **587**, 1848–1857.
82. Westhof,E., Yusupov,M. and Yusupova,G. (2014) Recognition of Watson-Crick base pairs: constraints and limits due to geometric selection and tautomerism. *F1000Prime Rep*. **6**, 19.
83. Takai,K. (2006) Classification of the possible pairs between the first anticodon and the third codon positions based on a simple model assuming two geometries with which the pairing effectively potentiates the decoding complex. *J. Theor. Biol.*, **242**, 564–580.
84. Sochacka,E., Kraszewska,K., Sochacki,M., Sobczak,M., Janicka,M. and Nawrot,B. (2011) The 2-thiouridine unit in the RNA strand is desulphured predominantly to 4-pyrimidinone nucleoside under in vitro oxidative stress conditions. *Chem. Commun.*, **47**, 4914–4916.
85. 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.
86. 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.
87. Hruska,F.E. and Blonski,W.J.P. (1982) A proton and carbon-13 nuclear magnetic resonance study of nucleosides with methylated pyrimidine bases Quick View Other Sources. *Can. J. Chem.*, **60**, 3026–3032.
88. Leszczynska,G., Sadowska,K., Bartos,P., Nawrot,B. and Sochacka,E. (2016) S-geranylated 2-thiouridines of bacterial tRNAs: chemical synthesis and physicochemical properties. *Eur. J. Org. Chem.*, 3482–3485.