Expanding functionality of RNA: synthesis and properties of RNA containing imidazole modified tandem G–U wobble base pairs†

Eriks Rozners,* Romualdas Smicius and Chika Uchiyama

Received (in Cambridge, UK) 29th July 2005, Accepted 27th September 2005
First published as an Advance Article on the web 20th October 2005
DOI: 10.1039/b510846b

Imidazole modification at C-5 of uridine that is part of tandem G–U wobble base pairs causes slight reduction of thermal stability ($\Delta\Delta G^0_{310} < 0.4$ kcal mol^{-1}) and relatively small change in hydration of short RNA helices.

The ability of RNA to catalyze chemical reactions is truly remarkable. However, compared to proteins RNA is an inherently much less efficient catalyst.[1] Whereas proteins are built with twenty amino acids having a variety of functional groups, RNA is built with only four nucleosides with limited functionality. Most notably, RNA lacks positively charged functional groups, which are important for catalysis in proteins. Several laboratories have designed modified nucleoside triphosphates (mostly 5-substituted uridines) to incorporate amino[2,3] and imidazole[3,4] groups in RNA and DNA *via* polymerase-catalyzed reactions and artificial *in vitro* selection techniques. In several instances, the activity of the selected nucleic acid catalysts for RNA cleavage[4a] and amide synthesis[4b] reactions critically depended on the presence of the imidazole-modified uridines. The limitation of such an approach is that the polymerase replaces every uridine with the modified nucleoside, which may have an undesirable effect on folding of a large nucleic acid.

Due to rapid advances in understanding of nucleic acid structure and folding, an alternative approach of rational engineering of nucleic acid functionality at selected sites becomes a conceivable and attractive goal. However, the nucleoside heterocyclic bases are buried deep inside the canonical double helix. To reach the surface of the helix, the functional groups attached to the heterocycles of Watson–Crick bases pairs require relatively long linkers[2–4] that may not be ideal for rational structure engineering. In contrast, non-canonical base pairs (*e.g.*, G–U wobble in Fig. 1) tend to distort the regular helices, exposing heterocyclic bases. Such motifs have certain functional groups, which are usually involved in the Watson–Crick base pairing, available in the expanded grooves. Therefore, we propose that non-canonical base pairs are ideal for engineering of functional groups for catalysis and molecular recognition. Herein, we report the synthesis and properties of RNA containing imidazole attached directly to 5-position of uracyl heterocycles of tandem G–U wobble base pairs (Fig. 1). Although a variety of functional groups have been conjugated with heterocycles of Watson–Crick base pairs,[2–5] chemical modifications of non-canonical base pairs are virtually unexplored.

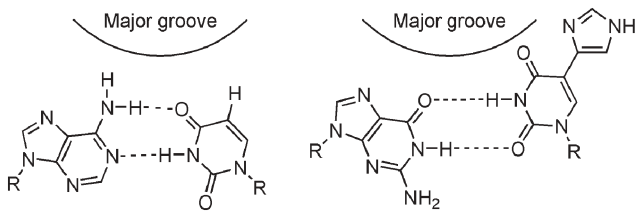
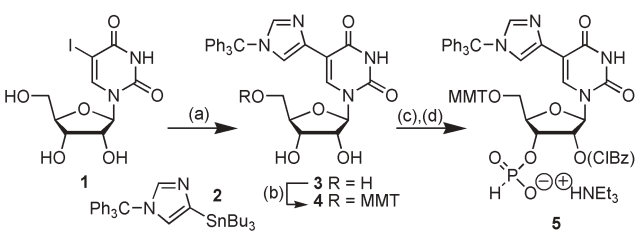


Fig. 1 Compared to an A–U base pair, 5-imidazole of a G–U wobble pair is shifted out in the major groove.

Department of Chemistry and Chemical Biology, Northeastern University, Boston, Massachusetts, 02115, USA.
E-mail: e.rozners@neu.edu; Fax: 1-617-373-8795; Tel: 1-617-373-5826
† Electronic supplementary information (ESI) available: Experimental procedures, detailed results of CD, thermal melting, and osmotic stressing experiments, and ^{1}H and ^{13}C NMR data. See DOI: 10.1039/b510846b

The modified uridine was prepared using a palladium catalyzed coupling of 5-iodouridine **1** and 4-tributylstannyl imidazole **2** (Scheme 1).[5,6] Both $(Ph_3P)_2PdCl_2$ and $(Ph_3P)_4Pd$ were good catalysts with the latter giving slightly higher yields and cleaner coupling. In contrast to similar literature precedents,[5a] efficient coupling was achieved without silver(I) oxide additive. For the protection of imidazole we chose the trityl group. In 3% dichloroacetic acid, which is used to remove the 5′-O-methoxytrityl group during the RNA solid-phase synthesis, only traces of deprotected **3** could be observed after 30 min (TLC). On the other hand, in 2% trifluoroacetic acid the half-life of the N-trityl protection in **3** was less than 15 min (TLC). Thus, we expected that the N-trityl group would be stable during the solid phase synthesis and could be removed using a stronger acid at the end of the synthesis.

The 5′-OH of the 5-imidazolyl uridine **3** was protected with the 4-monomethoxytrityl group (MMT). Selective benzoylation of the 2′-OH in **4** was immediately followed by phosphonylation of the 3′-OH in the same reaction mixture following our previously published procedures.[7,8] Isomerically pure 2′-*O*-(2-chlorobenzoyl)-3′-*O*-(H-phosphonate) **5** was obtained after silica gel column chromatography.



Scheme 1 Synthesis of H-phosphonate **5**. *Reagents and conditions*: (a) $(Ph_3P)_4Pd$, **2**, dioxane, 80 °C, 46 h, 74%; (b) 4-methoxytrityl chloride, pyridine, rt, overnight, 72%; (c) 2-chlorobenzoyl chloride, pyridine/CH_2Cl_2 (1:9), −78 °C, 45 min; (d) imidazole, PCl_3, NEt_3, CH_2Cl_2, −78 °C, 1 h, 63% (two steps).

To test our hypothesis that the imidazole heterocycle can be engineered in the major groove of the non-canonical tandem G–U wobble base pairs (Fig. 1) we prepared the self complementary sequences **6i–9i** (Table 1) using standard H-phosphonate oligoribonucleotide procedures[7–9] and the imidazole modified monomer **5**. The average coupling yields (typically >90% per step) were similar for **5** and the unmodified H-phosphonates. Deprotection of the imidazole was achieved by a prolonged treatment with 3% trifluoroacetic acid in the presence of triethylsilane. The overall yields, HPLC traces, and purity of the modified oligoribonucleotides were similar to native RNA prepared in our laboratory; the chemical composition of imidazole modified RNA was confirmed by MALDI-TOF mass spectrometry (for experimental details, see ESI†).

We chose **6–9** as the model compounds because Turner and co-workers have previously shown that these sequences display well-organized structure by NMR and a well-defined two-state melting behavior by UV spectroscopy.[10] In general, the 5′-UG motifs are thermally more stable than the 5′-GU motifs.[10] The NMR structure of **7** shows that the 5′-CGUG motif has only one hydrogen bond per G·U base pair.[10a] Consistent with previous studies,[10] the UV thermal melting of **6–9** and **6i–9i** gave good quality two-state melting curves. The results (Table 1) revealed that the imidazole modifications in **6i**, **7i**, and **9i** caused only a slight destabilization of 0.5–0.8 kcal mol^{-1} (Δt_m = −0.9 to −1.7 °C, per modification). An exception was **8i** where the imidazole modifications resulted in a 2.2 kcal mol^{-1} loss of thermal stability (Δt_m = −4.6 °C, per modification). These results were somewhat unexpected, because a structurally similar imidazole modification in DNA was reported to increase the stability of RNA–DNA heteroduplexes (Δt_m = +0.7 °C, per modification).[5a] We have previously observed that chemical modifications cause remarkably different responses in RNA *vs.* DNA.[8] To gain more insight into the effect of the imidazole modification on the tandem G–U motifs, we probed hydration of **6–9** using osmotic stress.

The osmotic stress is a thermodynamic method to evaluate the trends in hydration associated with biologically relevant equilibria from the dependence of the equilibrium constant on osmotic pressure (water activity).[11] For nucleic acids,[12] osmotic stress monitors the depression of the melting temperature upon decreasing of the water activity and calculates the number of water molecules (Table 1, Δn_W) associated with the double helix and released from the single strands upon melting:

$$\Delta n_W = (-\Delta H/R)[d(T_m^{-1})/d(\ln a_W)] \quad (1)$$

where $-\Delta H$ is the enthalpy determined from the width at the half-height of differentiated melting curves,[13] R is the universal gas constant (1.986 cal mol^{-1} K), and $d(T_m^{-1})/d(\ln a_W)$ is the slope of the plot of reciprocal melting temperature *vs.* the logarithm of water activity. Water activity is modulated by addition of low molecular weight co-solutes as osmotic stressors: ethylene glycol, glycerol, and acetamide (for details, see ESI†).

Spink and Chaires used osmotic stress to probe the hydration of long DNA molecules and found that about four water molecules per base pair were released upon melting of long DNA duplexes (*E. coli* DNA and poly(dA)–poly(dT)).[12] Recently, we demonstrated that osmotic stress can also detect relative changes of hydration caused by subtle structural differences in short DNA, RNA and 2′-OMe oligonucleotides.[14] Although the absolute numbers of water molecules obtained using different osmotic stressors is still under debate,[15] osmotic stress generally gives reliable relative estimates of the trends in hydration for a series of similar compounds.

Consistent with the previous studies,[12,14] we observed a good correlation of Δn_W for ethylene glycol and glycerol, whereas acetamide gave somewhat higher Δn_W numbers (Table 1). The differences in Δn_W numbers obtained with different osmotic stressors may be in part because of specific solute–oligonucleotide interactions.[15] However, comparisons within each series of osmotic stressors were consistent with the general trends discussed bellow. The errors for Δn_W were estimated using standard deviations of $-\Delta H^0$ and alternative graph fitting as described in the ESI.† Although these conservative error estimates are somewhat large, the consistency of the data among several osmotic stressors suggests that the differences are meaningful and allows for the following qualitative discussion of the trends.

The non-covalent interactions that determine the conformation and stability of RNA structure are not completely understood. NMR structures obtained by Turner and co-workers[10] show that **7–9** form similar helices, where the only significant departure from the A-type geometry is a slight overtwisting 5′ of the G and displacement of the G–U wobble pairs toward the major groove (Fig. 1). The different sequence dependent thermal stabilities appear to be due to subtle differences in electrostatic stacking interactions and hydrogen bonding. On the other hand, the importance of hydration for the conformation and stability of the G–U wobble pairs is little studied. Crystal structures of oligoribonucleotides show that the G–U wobble pairs are extensively hydrated in both grooves.[16] Analysis of the hydration pattern in the major groove,[16] suggests that the 5-imidazole

Table 1 Thermal melting and hydration of oligoribonucleotides having tandem G–U base pairs

	Sequence	t_m [a]/°C	$-\Delta G^0_{310}$ [a]/ kcal mol^{-1}	$-\Delta H^{0a}$/ kcal mol^{-1}	$-\Delta S^{0a}$/ kcal mol^{-1} K^{-1}	Ethylene glycol Δn_W [b]	Glycerol Δn_W [b]	Acetamide Δn_W [b]
6	GGC**UG**GCC	51.8 ± 0.1	12.1 ± 0.1	84.5 ± 2.3	245 ± 5	30.4 ± 2.2	27.2 ± 4.0	56.7 ± 2.7
6i	GGC**U^{Im}G**GCC	49.6 ± 0.2	11.3 ± 0.2	80.5 ± 3.0	236 ± 10	27.2 ± 2.2	21.4 ± 3.7	43.0 ± 3.2
7	GGC**GU**GCC	42.9 ± 0.2	9.5 ± 0.2	75.1 ± 2.7	220 ± 9	26.7 ± 3.7	31.1 ± 4.8	58.2 ± 3.4
7i	GGC**GU^{Im}**GCC	41.1 ± 0.3	9.0 ± 0.2	67.2 ± 4.7	197 ± 15	29.4 ± 4.7	31.6 ± 5.8	56.3 ± 5.5
8	GAG**UG**CUC	39.3 ± 0.3	8.6 ± 0.2	75.2 ± 3.4	233 ± 14	27.7 ± 4.1	24.6 ± 6.1	51.5 ± 5.0
8i	GAG**U^{Im}G**CUC	30.1 ± 0.2	6.4 ± 0.3	69.3 ± 4.6	238 ± 17	20.1 ± 3.4	23.7 ± 4.1	33.1 ± 4.4
9	GAG**GU**CUC	32.6 ± 0.3	6.8 ± 0.1	78.7 ± 2.8	267 ± 13	22.4 ± 4.6	20.3 ± 5.0	52.1 ± 4.8
9i	GAG**GU^{Im}**CUC	29.2 ± 0.2	6.0 ± 0.1	74.3 ± 2.6	235 ± 10	27.2 ± 4.9	21.6 ± 6.3	70.4 ± 4.4

[a] Oligonucleotide (2 μM) in 10 mM sodium cacodylate, 0.1 mM EDTA, and 300 mM NaCl. Results ± standard deviations. [b] Results ± error estimates.

modification may sterically exclude at least one water molecule out of the five "in-plane" hydration sites (see discussion and Fig. S2 in ESI†).

In general, the changes in hydration caused by the imidazole modification were relatively small. For the relatively more stable 5′-UG motifs **6** and **8**, the imidazole modification resulted in a loss of several water molecules per G–U base pair. For the 5′-GU motifs, imidazole modification caused either small changes (**7** to **7i**) or increased the hydration (**9** to **9i**). It is conceivable that more extensive hydration may contribute to higher thermal stability and conformational rigidity of 5′-UG compared to 5′-GU motifs. Interestingly, the most significant loss in thermal stability (**8** to **8i**) was also accompanied by the overall largest decrease in Δn_W. However, the relationship between t_m and Δn_W was not entirely consistent across the Table 1. More structural data will have to be obtained on hydration of G–U pairs in different sequence contexts to probe the role of hydration in conformation and thermal stability of these motifs.

To gain more insight into the effect of imidazole modification on tandem G–U wobble base pairs we compared the CD spectra of **6–8** with **6i–8i** (Fig. S3 in ESI†). Most notably, imidazole modifications that resulted in loss of hydration in **6** and **8** (Table 1) also caused significant changes in the CD spectra of these sequences. For **7** and **7i**, where imidazole caused the smallest changes in t_m, ΔG and Δn_W (Table 1), the CD spectra were also similar. Thus, our results suggest that the imidazole modification may fit best in the structurally more flexible G–U motifs, such as 5′-CGUG (**7**), which has only one hydrogen bond per G·U base pair according to the NMR structure.[10a] However, more studies are certainly needed to confirm the generality of this observation.

In summary, our results show that 5-imidazole modification of a G–U wobble pair is well accepted in most RNA sequence contexts, except **8**. The loss of thermal stability in **6i**, **7i** and **9i** is relatively small (per modification $\Delta\Delta G^0_{310} < 0.4$ kcal mol^{-1}). The imidazole modification also causes a small but significant rearrangement of the hydration of RNA in **6i**, **7i** and **9i**. Sequence **8i** represents an exception where imidazole causes a more significant loss of thermal stability and hydration. These results suggest that hydration effects are important to consider when designing chemically modified nucleic acids. It is possible that the loss of thermal stability frequently observed for non-polar modifications[17] is actually due to poor hydration rather than incompatible conformation or steric hindrance of the backbone. It is conceivable that the imidazole modification (which is likely to be partially protonated due to pK_a in vicinity of 7) uses its hydrogen bond donor/acceptor sites to rearrange the major groove water structure of G–U base pairs without disrupting the overall hydration of the duplex. In conclusion, G–U wobble pairs (and perhaps other non-canonical base pairs as well) can be used to rationally engineer imidazole incorporation at selected sites in RNA. This may open new avenues for design of more active nucleic acid catalysts (ribozymes) and receptors (aptamers) for biomedical and industrial applications.

We thank Northeastern University, The Petroleum Research Fund, administered by the American Chemical Society (37599-AC1) and NSF-NATO (postdoctoral fellowship DGE-0209488 to R. S.) for support of this research. We thank Heather Brodkin, James Glick, Dr Norman Chiu and Dr Paul Vouros for MALDI-TOF MS analysis and Dr James Manning for help with CD spectroscopy.

Notes and references

1 G. J. Narlikar and D. Herschlag, *Annu. Rev. Biochem.*, 1997, **66**, 19.
2 (*a*) A. Roychowdhury, H. Illangkoon, C. L. Hendrickson and S. A. Benner, *Org. Lett.*, 2004, **6**, 489; (*b*) T. R. Battersby, D. N. Ang, P. Burgstaller, S. C. Jurczyk, M. T. Bowser, D. D. Buchanan, R. T. Kennedy and S. A. Benner, *J. Am. Chem. Soc.*, 1999, **121**, 9781; (*c*) N. K. Vaish, R. Larralde, A. W. Fraley, J. W. Szostak and L. W. McLaughlin, *Biochemistry*, 2003, **42**, 8842; (*d*) N. K. Vaish, A. W. Fraley, J. W. Szostak and L. W. McLaughlin, *Nucleic Acids Res.*, 2000, **28**, 3316.
3 (*a*) S. E. Lee, A. Sidorov, H. Gourlain, N. Mignet, S. J. Thorpe, J. A. Brazier, M. J. Dickman, D. P. Hornby, J. A. Grasby and D. M. Williams, *Nucleic Acids Res.*, 2001, **29**, 1565; (*b*) T. Gourlain, A. Sidorov, N. Mignet, S. J. Thorpe, S. E. Lee, J. A. Grasby and D. M. Williams, *Nucleic Acids Res.*, 2001, **29**, 1898; (*c*) L. Lermer, Y. Roupioz, R. Ting and D. M. Perrin, *J. Am. Chem. Soc.*, 2002, **124**, 9960; (*d*) D. M. Perrin, T. Garestier and C. Helene, *J. Am. Chem. Soc.*, 2001, **123**, 1556; (*e*) J. D. Vaught, T. Dewey and B. E. Eaton, *J. Am. Chem. Soc.*, 2004, **126**, 11231; (*f*) K. Sakthivel and C. F. Barbas III, *Angew. Chem., Int. Ed.*, 1998, **37**, 2872.
4 (*a*) S. W. Santoro, G. F. Joyce, K. Sakthivel, S. Gramatikova and C. F. Barbas III, *J. Am. Chem. Soc.*, 2000, **122**, 2433; (*b*) T. W. Wiegand, R. C. Janssen and B. E. Eaton, *Chem. Biol.*, 1997, **4**, 675.
5 (*a*) A. J. Gutierrez, T. J. Terhorst, M. D. Matteucci and B. C. Froehler, *J. Am. Chem. Soc.*, 1994, **116**, 5540; (*b*) for a recent review on palladium assisted synthesis of nucleosides, see: L. A. Agrofoglio, I. Gillaizeau and Y. Saito, *Chem. Rev.*, 2003, **103**, 1875.
6 M. C. Jetter and A. B. Reitz, *Synthesis*, 1998, 829.
7 E. Rozners, R. Renhofa, M. Petrova, Y. Popelis, V. Kumpins and E. Bizdena, *Nucleosides Nucleotides*, 1992, **11**, 1579.
8 (*a*) E. Rozners, D. Katkevica, E. Bizdena and R. Strömberg, *J. Am. Chem. Soc.*, 2003, **125**, 12125; (*b*) E. Rozners and R. Strömberg, *J. Org. Chem.*, 1997, **62**, 1846.
9 (*a*) E. Westman, S. Sigurdsson, J. Stawinski and R. Strömberg, *Nucleic Acids Symp. Ser.*, 1994, **31**, 25; (*b*) J. Stawinski and R. Strömberg, in *Oligonucleotide Synthesis, Methods and Applications, Methods in Molecular Biology*, ed. P. Herdewijn, Humana Press, Totowa, NJ, 2005, vol. 288, ch. 6, pp. 81–100.
10 (*a*) X. Chen, J. A. McDowell, R. Kierzek, T. R. Krugh and D. H. Turner, *Biochemistry*, 2000, **39**, 8970; (*b*) J. A. McDowell and D. H. Turner, *Biochemistry*, 1996, **35**, 14077; (*c*) L. He, R. Kierzek, J. SantaLucia, Jr., A. E. Walter and D. H. Turner, *Biochemistry*, 1991, **30**, 11124.
11 (*a*) V. A. Parsegian, R. P. Rand and D. C. Rau, *Methods Enzymol.*, 1995, **259**, 43; (*b*) C. R. Robinson and S. G. Sligar, *Methods Enzymol.*, 1995, **259**, 395.
12 C. H. Spink and J. B. Chaires, *Biochemistry*, 1999, **38**, 496.
13 K. J. Breslauer, *Methods Enzymol.*, 1995, **259**, 221.
14 E. Rozners and J. Moulder, *Nucleic Acids Res.*, 2004, **32**, 248.
15 E. S. Courtenay, M. W. Capp, C. F. Anderson and M. T. Record, Jr., *Biochemistry*, 2000, **39**, 4455.
16 (*a*) P. Auffinger and E. Westhof, *J. Biomol. Struct. Dyn.*, 1998, **16**, 693; (*b*) R. Biswas, M. C. Wahl, C. Ban and M. Sundaralingam, *J. Mol. Biol.*, 1997, **267**, 1149; (*c*) S. R. Holbrook, C. Cheong, I. Tinoco, Jr. and S. H. Kim, *Nature*, 1991, **353**, 579.
17 S. M. Freier and K.-H. Altmann, *Nucleic Acids Res.*, 1997, **25**, 4429.

