J|A|C|S
COMMUNICATIONS

Published on Web 01/17/2007

Highly Selective Recognition of Cytosine over Uracil and Adenine by a Guanine Analogue, 2-*N*-Acetyl-3-deazaguanine, in 2′-*O*-Methyl-RNA/RNA and DNA Duplexes

Kohji Seio,[†,§] Takeshi Sasami,[‡,§] Akihiro Ohkubo,[‡,§] Kaori Ando,[¶] and Mitsuo Sekine*[,‡,§]

Frontier Collaborative Research Center and Department of Life Science, Tokyo Institute of Technology, and CREST, JST, 4259 Nagatsuta, Yokohama, Japan 226-8501, and College of Education, University of the Ryukyus, Okinawa 903-0213, Japan

Received July 7, 2006; E-mail: msekine@bio.titech.ac.jp

The selective hybridization of DNA and RNA to their complementary targets is the most essential property to develop nucleic acid-based technologies such as gene detection and antisense drugs. Other than the canonical Watson−Crick base pairs, however, there are a number of noncanonical base pairing motifs, such as mismatch base pairs, which hamper the selective hybridization by the Watson−Crick base pairs.

Among the four nucleobases, guanine most frequently forms mismatched base pairs. In DNA, the high stabilities of G−G[1], G−A[2], and G−T[3] mismatches are quantitatively evaluated. Similarly, the stability of the mismatched base pairs containing guanine, such as a G−U wobble base pair, was also reported in RNA.[4] Therefore, the development of guanine analogues that do not form such stable mismatched base pairs, especially G−U, is of great importance to improve the selectivity of the hybridization.

In this communication, we describe a new modified base, 2-*N*-acetyl-3-deazaguanine (a^2c^3G), and the synthesis of 2′-*O*-methyl-RNA incorporating a^2c^3G. We demonstrate that a^2c^3G is the first example of a guanine analogue that recognizes cytosine more selectively than guanine while keeping the stability of the Watson−Crick base pair essentially unchanged. We chose 2′-*O*-methyl-RNA considering its usefulness as a class of artificial RNA for antisense drug and hybridization probe targeting RNAs.[5]

First, the conformation properties of a^2c^3G were studied by ab initio calculations at the MP2/6-31G**//HF/6-31G** level. The geometries of the 9-methyl-a^2c^3G were optimized, and the geometry optimization was confirmed by the frequency analyses. The calculations revealed the presence of two planer (*Z*)-*N*-acetyl rotamers and one (*E*)-*N*-acetyl rotamer. The (*Z*)-rotamers in the "open" and "closed" forms are stabilized by the intramolecular hydrogen bonds between the carbonyl oxygen of the acetyl group and CH(3) and NH(1), respectively (Figure 1). The (*E*)-rotamer having no intramolecular hydrogen bond is less stable than the (*Z*)-rotamers (Figure S1). Hydrogen bonds between a carbonyl oxygen and a base aromatic proton are already reported.[6,7] Although the closed form which cannot form base pairs was the most stable, the acetyl group can rotate to give the open form because the activation energy, E_a, accompanying the rotation to the open form is only +4.2 kcal/mol (see Figures 1 and S1).

Next, the base pairing energies of a^2c^3G were calculated for the a^2c^3G−cytosine and −uracil base pairs with the BSSE correction,[8,9] and these energies were compared with those of the base pairs containing guanine. The results are summarized in Figure 1, Table S1, and Figure S2. In these calculations, the open form was chosen for a^2c^3G, and the wobble geometry (Figures S1 and S3) was chosen for a^2c^3G−U pair. The results indicated that the base pairing energy

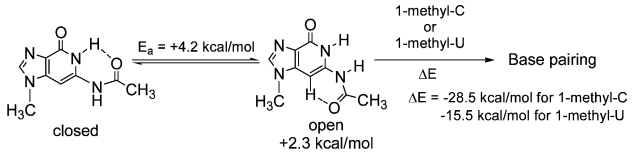

Figure 1. The rotamer structures and the energies accompanying the conformation change and the base pairing. The E_a and the energy of the open form are the relative values from the energy of the closed form.

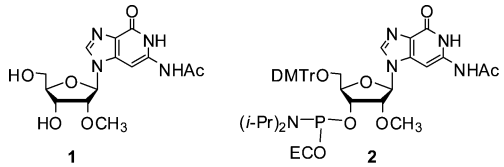

Figure 2. The structure of 2′-*O*-methyl-a^2c^3G **1** and its phosphoramidite **2**.

($\Delta E = -28.5$ kcal/mol) of the Watson−Crick-type a^2c^3G (open)−C base pair was larger than that (−24.5 kcal/mol) of the G−C pair, probably because of the more polar properties of the N−H bond. The base pairing energy (−15.5 kcal/mol) of a^2c^3G−U is also slightly increased from that (−13.1 kcal/mol) of G−U. Even in consideration of the energy cost of 2.3 kcal/mol to convert the closed form to the open form, the a^2c^3G−C base pair was still more stable by 1.7 kcal/mol than a G−C base pair. Similarly, the a^2c^3G−U base pair became slightly more stable by 0.1 kcal/mol than the G−U pair. As the result, the energy difference between a^2c^3G−C and a^2c^3G−U is greater than that between G−C and G−U by 1.6 kcal/mol (Table S1). These calculations suggested that a^2c^3G could have higher base discrimination ability than guanine as far as the wobble base pair with uracil was concerned. We also calculated the ΔG values of the formation of the unmodified base pairs G−C and G−U and the modified base pairs a^2c^3G−C and a^2c^3G−U from the frequency analyses, as shown in Figure S2. These results also supported the above conclusion.On the basis of these theoretical considerations, we synthesized the phosphoramidite derivative **2** of 2′-*O*-methyl-2-*N*-acetyl-3-deazaguanosine (**1**) starting from 3-deazaguanosine.[10,11] The detailed synthetic scheme and synthetic procedures of these compounds are described as a part of Supporting Information.

In order to clarify the hybridization and the base discrimination properties of a^2c^3G, 2′-*O*-methyl-RNAs (5′-CGGCXAGGAG-3′: X = a^2c^3G or G) were synthesized by use of **2** according to the standard RNA synthesis protocol. The protecting groups of the canonical bases, acetyl for cytosine, phenoxyacetyl for adenine, and 4-isopropylphenoxyacetyl for guanine, were removed by treatment with aqueous ammonia at ambient temperature. During this deprotection, the acetyl group of a^2c^3G was proven to be quite stable. The hybridization of the 2′-*O*-methyl-RNA to the RNA targets r(5′-CUCCUYGCCG-3′: Y = C, G, A, and U) was studied by measuring the melting temperature (T_m).

[†] Frontier Collaborative Research Center, Tokyo Institute of Technology.
[‡] Department of Life Science, Tokyo Institute of Technology.
[§] CREST, JST.
[¶] University of the Ryukyus.

10.1021/ja064612g CCC: $37.00

Table 1. T_m (°C) of 5′-CGGCXAGGAG-3′/3′-r(GCCGYUCCUC)-5′[a]

Y =	C	U	A	G
X = a^2c^3G	70.1	54.9	50.8	56.8
G	70.9	61.8	50.3	53.7
ΔT_m[b]	−0.8	−6.9	+0.5	+3.1

[a] The underline represents 2′-*O*-methyl-RNA. [b] $\Delta T_m = T_m$ of $a^2c^3G - T_m$ of G.

Table 2. T_m (°C) of 5′-CGGCXAGGAG-3′/3′-d(GCCGYTCCTC)-5′[a]

Y =	C	T	A	G
X = a^2c^3G	59.5	46.1	37.8	41.6
G	58.7	50.4	40.4	41.0
ΔT_m[b]	+0.8	−4.3	−2.6	+0.6

[a] The underline represents 2′-*O*-methyl-RNA. [b] $\Delta T_m = T_m$ of $a^2c^3G - T_m$ of G.

Table 3. T_m (°C) of 5′-CGG**CXAG**GAG-3′/3′-r(GCC**GAGC**CUC)-5′ or 3′-d(GCCGAGCCTC)-5′[a]

	5′-XA-3′/3′-r(AG)-5′	5′-XA-3′/3′-d(AG)-5′
X = a^2c^3G	43.8	28.4
G	49.8	35.7
ΔT_m[b]	−6.0	−7.3

[a] The underline represents 2′-*O*-methyl-RNA. The sites in boldface indicate the tetrad including a tandem G−A mismatch. [b] $\Delta T_m = T_m$ of $a^2c^3G - T_m$ of G.

As shown in Table 1, the Watson−Crick base pair between a^2c^3G and C was as strong as that between guanine and C ($\Delta T_m = -0.8$ °C). In contrast, the wobble mismatch with U was destabilized significantly by the modification of G to a^2c^3G ($\Delta T_m = -6.9$ °C). Thus the selectivity between C and U became higher. Although the a^2c^3G−A and a^2c^3G−G mismatch pairs were slightly stabilized by this modification, the differences between the T_m values of these mismatches and that of the a^2c^3G−C pair were still large enough, more than 13 °C, to achieve selective hybridization.

A similar improvement of the base recognition ability was also observed when the 2′-*O*-methyl-RNA was hybridized to DNA targets, d(CTCCTYGCCG: Y = C, G, A, and T). As shown in Table 2, the a^2c^3G−C base pair was as stable as the G−C base pair ($\Delta T_m = +0.8$ °C), and the a^2c^3G−T base pair was much less stable than the G−T wobble base pair ($\Delta T_m = -4.3$ °C). In the case of a duplex with the DNA target, stabilization of the mismatch pair with G was less significant ($\Delta T_m = +0.6$ °C) than in the duplex with RNA target.

These results indicated that a^2c^3G was superior to guanine as a component of hybridization probes that recognize cytosine strongly and selectively in this sequence.

Finally, we examined the effect of modification of the guanine moiety in tandem G−A mismatches in 5′-CXAG-3′/3′-GAGC-5′ (X = G or a^2c^3G) sequences. It is well-known from the structural studies that the tandem G−A mismatches can be stabilized in two different base pairing modes; one is a sheared-type, and the other is a face-to-face mode (Figure S3).

In the 5′-CGAG-3′/3′-GAGC-5′ sequence, the sheared-type base pairing is proven to be predominant.[12] Because a^2c^3G lacks a nitrogen atom at position 3, which is essential to form a sheared GA mismatch, the sheared-type base pairing of a^2c^3G and A must be destabilized in this nucleotide sequence. Shown in Table 3 are the T_m values of the 2′-*O*-methyl-RNA/RNA and 2′-*O*-methyl-RNA/DNA duplexes containing a 5′-CXAG-3′/3′-GAGC-5′ sequence. Apparently, the incorporation of a^2c^3G in place of guanine significantly decreased the T_m value of the tandem mismatches both in the 2′-*O*-methyl-RNA/RNA and in the 2′-*O*-methyl-RNA/DNA by 6.0 and 7.3 °C, respectively. Because the full match duplex and the single mismatch with A were not so much destabilized by a^2c^3G as shown in Table 1, the large destabilization incurred from the tandem GA mismatches by introduction of a^2c^3G in place of G indicated the formation of a sheared-type base pair even in the duplexes containing 2′-*O*-methyl-RNA and the repulsion between the methine of a^2c^3G at position 3 and the amino group of adenine. We also observed similar destabilization of the 5′-CXAG-3′/3′-GAGC-5′-type tandem G−A mismatches by introduction of 3-deazaguanine (c^3G).[13] In addition, Seela and co-workers reported the destabilization of a sheared-type G−A pair in a hammerhead ribozyme by c^3G.[14] Therefore, the destabilization of this type of tandem mismatches seemed to be the general properties of 3-deazaguanine nucleosides.

In conclusion, we have developed a new nucleobase, 2-*N*-acetyl-3-deazaguanine, which showed highly selective base recognition by destabilizing the wobble base pair with uracil. The higher selectivity might be attributed to the relatively weak a^2c^3G−U pair in comparison to the a^2c^3G−C pair as clarified by the ab initio calculations. Moreover, incorporation of a^2c^3G destabilized a tandem GA mismatch in a certain sequence probably because, due to the lack of the nitrogen atom, the sheared-type GA pairing was avoided. It should be emphasized, however, that the presence of the open conformation in the duplexes should be determined by use of ^{1}H NMR to clarify the base recognition mechanisms of a^2c^3G unambiguously. In addition, it is also important to consider interactions other than hydrogen bonds, such as stacking interactions on the bases of the 3D structure.

We previously reported 2′-*O*-methyl-2-thiouridine[15] as a uridine analogue that can stabilize the Watson−Crick base pair but destabilizes the wobble-type pairing with G. By the combinatorial use of 2-thiouracil and a^2c^3G in 2′-*O*-methyl-RNA, we might develop new hybridization probes or antisense molecules having higher selectivity to the target RNAs.

Acknowledgment. This work was financially supported by a grant in aid from CREST (JST) and by a grant of the Genome Network Project from Ministry of Education. This study was also supported by Industrial Technology Research Grant Program in ’05 from NEDO.

Supporting Information Available: Details of the experimental procedures, and the spectroscopic data of compound **2** and all the synthetic intermediates. The results of the ab initio calculations in detail. This material is available free of charge via the Internet at http://pubs.acs.org.

References

(1) Peyret, N.; Seneviratne, P. A.; Allawi, H. T.; SantaLucia, J., Jr. *Biochemistry* **1999**, *38*, 3468−1377.
(2) Allawi, H. T.; SantaLucia, J., Jr. *Biochemistry* **1998**, *37*, 2170−2179.
(3) Allawi, H. T.; SantaLucia, J., Jr. *Biochemistry* **1997**, *36*, 10581−10594.
(4) (a) Moroueh, M.; Chow, C. S. *Nucleic Acids Res.* **1999**, *27*, *1118*−1125. (b) Bevilacqua, J. M.; Bevilacqua, P. V. *Biochemistry* **1998**, *37*, 15877−15884. (c) Zhu, J.; Wartell, R. M. *Biochemistry* **1997**, *36*, 15326−15335.
(5) Majlessi, M.; Nelson, N. C.; Becker, M. M. *Nucleic Acids Res.* **1998**, *26*, 2224−2229.
(6) Miyata, K.; Kobori, A.; Tamamushi, R.; Ohkubo, A.; Taguchi, H.; Seio, K.; Sekine, M. *Eur. J. Org. Chem.* **2006**, 3626−3637.
(7) Thomas, G.; Gordon, J.; Rogg, H. *J. Biol. Chem.* **1978**, *253*, 1101−1105.
(8) Sponer, J.; Leszczynski, J.; Hobza, P. *J. Biomol. Struct. Dyn.* **1996**, *14*, 117−135.
(9) Kawahara, S.; Uchimaru, T.; Taira, K.; Sekine, M. *J. Phys. Chem. A* **2002**, *106*, 3207−3212.
(10) Cook, P. D.; Rousseau, R. J.; Mian, A. M.; Dea, P.; Meyer, R. B., Jr.; Robins, R. K. *J. Am. Chem. Soc.* **1975**, *97*, 2916−2917.
(11) Cook, P. D.; Rousseau, R. J.; Mian, A. M.; Dea, P.; Meyer, R. B., Jr.; Robins, R. K. *J. Am. Chem. Soc.* **1976**, *98*, 1492−1498.
(12) SantaLucia, J., Jr.; Turner, D. H. *Biochemistry* **1993**, *32*, 12612−12623.
(13) Seio, K.; Sasami, T.; Tawarada, R.; Sekine, M. *Nucleic Acids. Res.* **2006**, *34*, 4324−4334.
(14) Seela, F.; Debelak, H.; Andrews, L.; Beigelman, L. *Helv. Chim. Acta* **2003**, *86*, 2726−2739.
(15) Shohda, K.; Okamoto, I.; Wada, T.; Seio, K.; Sekine, M. *Bioorg. Med. Chem. Lett.* **2000**, *10*, 1795−1798.

JA064612G