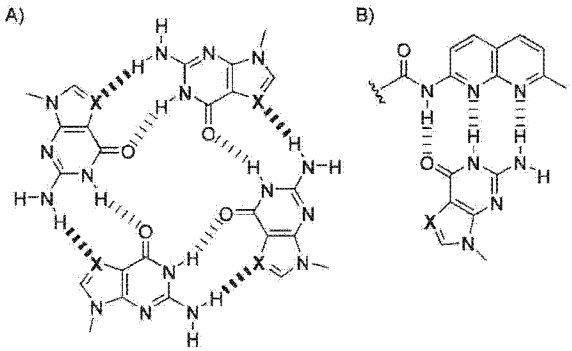
deazaguanine (Z), that is the N7 of G was replaced with CH. As shown in Scheme 1A, the G-quadruplex structure consists of two kinds of hydrogen bond; one is between the N1-H and C6 carbonyl oxygen, and the other is between the hydrogen in




**Scheme 1.** A) Structure of the G-quadruplex with X = N. 7-Deazaguanine (X = CH) cannot form a quadruplex due to the lack of hydrogen bonding to the 2-amino group (shown with a bold dash). B) Hydrogen bonding of *N*-acyl-2-amino-1,8-naphthyridine in **NT** to guanine (X = N) and 7-deazaguanine (X = CH).

the N2 amino group and N7. A loss of the $NH_2$···N7 hydrogen bonds from substituting G with Z and a potential steric repulsion between $NH_2$ and C7-H in a hypothetical Z-quadruplex suggests that formation of a quadruplex structure of telo15Z would be unlikely. In fact, ESI-TOF MS showed that telo15Z did not produce a dimer under the conditions given in Figure 1C. However, a 1:1 complex of telo15Z and NT similar to that of telo15 and NT was confirmed by the ion peak at *m/z* 1406.21 (Figure 1D). The fact that the guanine N7 is essential for the formation of a telo15 dimer, but not for the NT-bound complex, strongly suggested that NT binds to the Watson–Crick face of G and Z (Scheme 1B). Isothermal titration calorimetry[29] revealed that the binding constant of NT to telo15 was $5.71 \times 10^6\ M^{-1}$, with the titration curve fitting well to a 1:1 binding isotherm (see Figure S1 in the Supporting Information).

The structural changes of telo15 upon NT binding suggested by ESI-TOF MS measurements were supported by the absorbance changes at 260 and 295 nm (Figure 2). The melting of duplex to a single strand was monitored at 260 nm, whereas the melting of the quadruplex was monitored at 295 nm.[30] Since the G-quadruplex was stabilized by potassium cations, the measurements were carried out in 500 mM KCl. The absorption of telo15 at 295 nm decreased with increasing temperature to give a melting temperature of the G-quadruplex at 39.6 °C. In marked contrast, no apparent $T_m$ of telo15Z was observed at 295 nm due to its inability to form a G-quadruplex (Figure 2A). Neither telo15 nor telo15Z showed apparent absorbance changes at 260 nm (data not shown). In the presence of NT, melting of NT-bound telo15 and telo15Z was clearly observed at 260 nm with $T_m$ values of 62.8 and 71.3 °C, respectively (Figure 2B). The structural changes of telo15 and telo15Z were also clearly shown by CD spectra, with strong induced CD (Figure 3). The CD spectral change for telo15 while titrating with NT showed isodichroic points that indicated the single

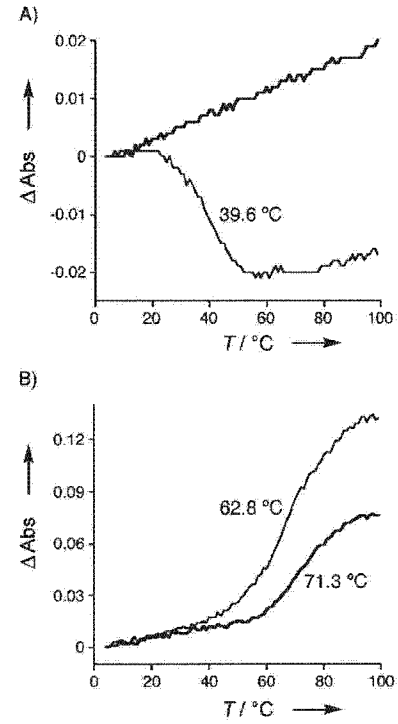



**Figure 2.** Thermal denaturation profiles of telo15 (thin line) and telo15Z (bold line; 5 μM) obtained in the absence and presence of NT (30 μM) in 500 mM KCl and 10 mM sodium cacodylate buffer (pH 7.0) from 4 to 98 °C with a heating rate of 1 °C $min^{-1}$. A) Absorbance changes at 295 nm in the absence of NT. B) Absorbance changes at 260 nm in the presence of NT. ΔAbs was obtained from the change in absorption of the solution before (4 °C) and after heating. The absorption change of NT was further subtracted from the data. The $T_m$ calculated by the median method is shown in the figure.

transition between the two structural states. The isodichroic points were not clear for the structural change for telo15Z, because telo15Z does not have a particular struture in the absence of NT, as shown by the $T_m$ measurements. The data from ESI-TOF MS, UV melting, and CD spectra indicated that 1) NT did not bind to the G-quadruplex form of telo15, 2) NT binding did not involve hydrogen bonds to G-N7, and 3) NT-bound structures had spectroscopic characteristics that resembled those of dsDNA; this suggested a hairpin formation of telo15 upon NT binding.

The possible hairpin form of telo15 involves a cluster of G–G mismatches in the stem region. To clarify the binding sites of NT among three G–G mismatches (GGa, GGb, and GGc), UV melting profiles of six mutant oligomers of telo15 were measured in the absence and presence of NT (Figure 4). The $T_m$ of telo15 in 100 mM NaCl was too low to be detected, but was 71.2 °C in the presence of NT, that is, 8.4 °C higher than the $T_m$ in 500 mM KCl. Mutant sequences telo15a and telo15b, in which GGa and GGb were replaced with an A–T base pair, did not show a measurable $T_m$, as for telo15. In the presence of NT, the $T_m$ of telo15a was 75.0 °C, that is, 3.8 °C higher than that of telo15; this showed that substitution of GGa had negligible effects on NT binding. In marked contrast, the $T_m$ of telo15b was 43.5 °C, that is, 27.7 °C lower than that of telo15. Likewise, the

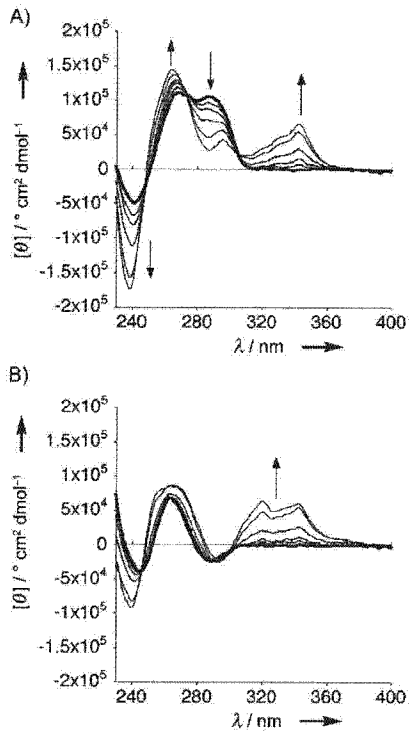



**Figure 3.** CD spectra of telo15 and telo15Z (5 μM) in the presence (thin line) and absence (bold line) of NT. The CD measurements were carried out with A) telo15 and B) telo15Z in 500 mM KCl and 10 mM sodium cacodylate buffer (pH 7.0) while titrating with NT (0, 0.5, 1, 2, 4, 6 μM) at 25 °C.

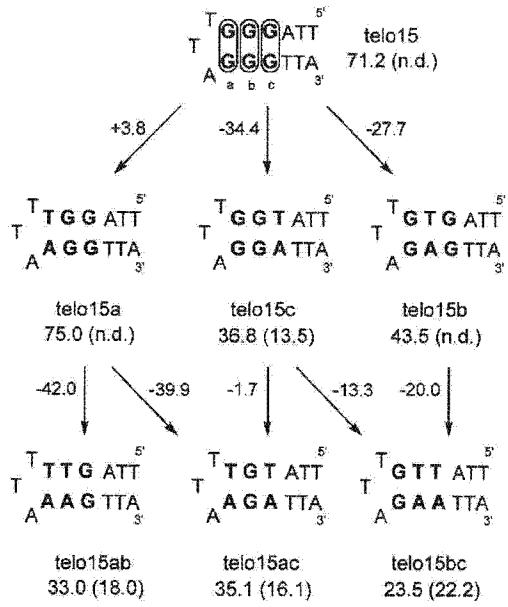



**Figure 4.** Possible hairpin structures of telo15 mutants (5 μM) and their $T_m$ (°C) in the presence of NT (30 μM) in 100 mM NaCl and 10 mM sodium cacodylate (pH 7.0). UV melting profiles were measured at 260 nm. The $T_m$ of the mutants is shown in parenthesis, n.d. = not detected. The numbers beside the arrows indicate the increase (+) or decrease (−) in $T_m$ between the two mutants.

$T_m$ of telo15c was 36.8 °C, a difference of 34.4 °C between telo15 and telo15c in the presence of NT. The subsequent substitution of GGb in telo15a to give telo15ab, and GGc in telo15a to give telo15ac, resulted in decreases of $T_m$ by 42.0 and 39.9 °C, respectively. The substitution of GGc in telo15b and GGb in telo15c, both giving telo15bc, also decreased the $T_m$ by 20.0 °C and 13.3 °C, respectively. Telo15bc was not stabilized by NT at all, as judged by the 1.3 °C increase in $T_m$ in the presence of NT. The substitution of GGa in telo15b to give telo15ab led to a $T_m$ decrease of 10.5 °C. Furthermore, the substitution of GGa in telo15c to give telo15ac had almost no effect. A large decrease in $T_m$ when a cluster of GGb and GGc was modified indicated that both GGb and GGc were essential for the strong and simultaneous stabilization of, and hairpin formation in, telo15 by NT, whereas GGa was only very weakly effective on NT binding. The weak binding of NT to GGa was most likely because GGa was directly neighboring the TTA hairpin loop. Binding of NT to GGa would cause strain in the hairpin loop and was not energetically favorable.

The data described here suggested that NT induced a hairpin structure in telo15 with a 1:1 stoichiometry of binding. The novel mode of NT binding to the human telomere sequence described in these studies would be useful for the design of next-generation molecules that would show more potent binding to the sequence. Human telomere repeats form parallel and a mixed parallel/antiparallel conformers according to the presence of potassium and sodium cations, respectively.[11–15] Furthermore, the stability of the G-quadruplex structures is likely to depend on the repeat length. It is particularly informative to evaluate whether NT binds to long stretches of human telomere repeats under physiological conditions and can cause inhibition of the telomerase function. Finally, strong stabilization of two consecutive G–G mismatches by NT implies that NT might bind to other biologically important G-rich sequences, such as a promoter region[31] and a CGG trinucleotide repeat[32] by hairpin formation.

## Experimental Section

**Preparation of NT:** A 25% glutaraldehyde solution saturated with NaCl was extracted with $CHCl_3$. The organic phase was evaporated to dryness to give glutaraldehyde as a viscous liquid. Glutaraldehyde (12 mg, 0.12 mmol) and $NaBH_3CN$ (18.7 mg, 0.30 mmol) were added to a solution of ND (110 mg, 0.25 mmol) in MeOH adjusted to pH 5 with acetic acid. The mixture was stirred at room temperature for 10 min, then poured into $CHCl_3$ and washed with brine. The organic layer was dried over $Na_2SO_4$ and evaporated in vacuo. The crude residue was purified by preparative thin-layer chromatography (TLC) to produce NT (34.5 mg, 15%) as a white solid. $^1H$ NMR ($CD_3OD$, 400 MHz) $\delta = 8.10$ (d, $J = 8.8$ Hz, 4H), 7.83 (d, $J = 8.4$ Hz, 4H), 7.83 (d, $J = 8.8$ Hz, 4H), 7.13 (d, $J = 8.4$ Hz, 4H), 2.79 (t, $J = 6.0$ Hz, 8H), 2.59 (s, 12H), 2.62–2.56 (8H), 2.53 (t, $J = 7.2$ Hz, 4H), 1.67 (m, 4H), 1.44 (m, 2H); $^{13}C$ NMR ($CD_3OD$, 100 MHz) $\delta = 174.1$, 163.7, 155.0, 154.9, 139.6, 138.2, 122.4, 119.4, 115.5, 55.2, 50.4, 35.5, 27.4, 26.5, 25.0; FABMS (NBA), *m/e* 955 $[M+H]^+$; HRMS calcd for $C_{53}H_{59}O_4N_{14}$: 955.4844 $[M+H]^+$; found: 955.4841.

**General procedure for ESI-TOF MS analysis of the binding complex:** DNA duplex solutions (20 μM) in 50% methanol containing

$NH_4OAc$ (100 mM) were measured with various ligands in the negative mode of ESI-TOF MS (JEOL AccuTOF JMS-T100N). Nitrogen was used as the desolvation gas, as well as a nebulizer. Conditions for ESI-TOF MS were as follows: orifice 1 temperature and desolvation room temperature were not controlled, needle voltage −2100 V, ring-lens voltage −5 V, orifice 1 voltage −45 V, orifice 2 voltage −6 V.

**Thermal denaturation profiles**: UV melting experiments of **telo15** and its mutants were carried out with DNA oligomer (5 μM) in sodium cacodylate (10 mM, pH 7.0) containing NaCl (100 mM). The absorbance of the sample was monitored at 260 or 295 nm from 4 °C to 98 °C with a heating rate of 1 °C $min^{-1}$ in the absence and presence of various concentrations of ligands. $T_m$ values were calculated by using the median method. Measurement of the G-quadruplex of **telo15** was carried out in KCl (500 mM) instead of NaCl (100 mM). The change in absorbance was obtained from the following equation:

$$\Delta Abs = Abs_{temp} - Abs_{4\,^\circ C} - (NT\,Abs_{temp} - NT\,Abs_{4\,^\circ C})$$

Here $Abs_{temp}$ is the absorption of **NT-telo15** at a specific temperature, and $NTAbs_{temp}$ is the absorption of **NT** at that temperature.

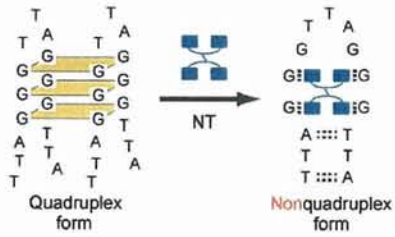
**Keywords:** inhibitors · ligand effects · mismatches · telomere repeats

[1] C. W. Greider, *Annu. Rev. Biochem.* **1996**, *65*, 337–365.
[2] E. H. Blackburn, *Nature* **1991**, *350*, 569–573.
[3] C. B. Harley, A. B. Futcher, C. W. Greider, *Nature* **1990**, *345*, 458–460.
[4] N. D. Hastie, M. Dempster, M. G. Dunlop, A. M. Thompson, D. K. Green, R. C. Allshire, *Nature* **1990**, *346*, 866–868.
[5] S. E. Holt, J. W. Shay, *J. Cell. Physiol.* **1999**, *180*, 10–18.
[6] J. L. Mergny, C. Hélène, *Nat. Med.* **1998**, *4*, 1366–1367.
[7] S. Neidle, G. Parkinson, *Nat. Rev. Drug Discovery* **2002**, *1*, 383–393.
[8] L. H. Hurley, *Nat. Rev. Cancer* **2002**, *2*, 188–200.
[9] S. M. Kerwin, *Curr. Pharm. Des.* **2000**, *6*, 441–478.
[10] S. Neidle, G. N. Parkinson, *Curr. Opin. Struct. Biol.* **2003**, *13*, 275–283.
[11] G. N. Parkinson, M. P. H. Lee, S. Neidle, *Nature* **2002**, *417*, 876–880.
[12] Y. Wang, D. J. Patel, *Structure* **1993**, *1*, 263–282.
[13] Y. Xu, Y. Noguchi, H. Sugiyama, *Bioorg. Med. Chem.* **2006**, *14*, 5584–5591.
[14] A. Ambrus, D. Chen, J. Dai, T. Bialis, R. A. Jones, D. Yang, *Nucleic Acids Res.* **2006**, *34*, 2723–2735.
[15] K. N. Luu, A. T. Phan, V. Kuryavyi, L. Lacroix, D. J. Patel, *J. Am. Chem. Soc.* **2006**, *128*, 9963–9970.
[16] J. F. Riou, L. Guittat, P. Mailliet, A. Laoui, E. Renou, O. Petitgenet, F. Megnin-Chanet, C. Hélène, J. L. Mergny, *Proc. Natl. Acad. Sci. USA* **2002**, *99*, 2672–2677.
[17] D. F. Shi, R. T. Wheelhouse, D. Sun, L. H. Hurley, *J. Med. Chem.* **2001**, *44*, 4509–4523.
[18] K. Shin-ya, K. Wierzba, K. Matsuo, T. Ohtani, Y. Yamada, K. Furihata, Y. Hayakawa, H. Seto, *J. Am. Chem. Soc.* **2001**, *123*, 1262–1263.
[19] W. Duan, A. Rangan, H. Vankayalapati, M. Y. Kim, Q. Zeng, D. Sun, H. Han, O. Y. Fedoroff, D. Nishioka, S. Y. Rha, E. Izbicka, D. D. Von Hoff, L. H. Hurley, *Mol. Cancer Ther.* **2001**, *1*, 103–120.
[20] S. M. Haider, G. N. Parkinson, S. Neidle, *J. Mol. Biol.* **2003**, *326*, 117–125.
[21] A. M. Burger, F. P. Dai, C. M. Schultes, A. P. Reszka, M. J. Moore, J. A. Double, S. Neidle, *Cancer Res.* **2005**, *65*, 1489–1496.
[22] R. A. Heald, C. Modi, J. C. Cookson, I. Hutchinson, C. A. Laughton, S. M. Gowan, L. R. Kelland, M. F. G. Stevens, *J. Med. Chem.* **2002**, *45*, 590–597.
[23] M. J. B. Moore, F. Cuenca, M. Searcey, S. Neidle, *Org. Biomol. Chem.* **2006**, *4*, 3479–3488.
[24] A. T. Phan, V. Kuryavyi, H. Y. Gaw, D. J. Patel, *Nat. Chem. Biol.* **2005**, *3*, 167–173.
[25] K. Nakatani, S. Sando, I. Saito, *Nat. Biotechnol.* **2001**, *19*, 51–55.
[26] K. Nakatani, S. Sando, H. Kumasawa, J. Kikuchi, I. Saito, *J. Am. Chem. Soc.* **2001**, *123*, 12650–12657.
[27] K. Nakatani, S. Hagihara, S. Sando, S. Sakamoto, K. Yamaguchi, C. Maesawa, I. Saito, *J. Am. Chem. Soc.* **2003**, *125*, 662–666.
[28] J. L. Beck, M. L. Colgrave, S. F. Ralph, M. M. Sheil, *Mass Spectrom. Rev.* **2001**, *20*, 61–87.
[29] A. N. Lane, T. C. Jenkins, *Q. Rev. Biophys.* **2000**, *33*, 255–306.
[30] J. L. Mergny, A. T. Phan, L. Lacroix, *FEBS Lett.* **1998**, *435*, 74–78.
[31] J. Seenisamy, E. M. Rezler, T. J. Powell, D. Tye, V. Gokhale, C. S. Joshi, A. Siddiqui-Jain, L. H. Hurley, *J. Am. Chem. Soc.* **2004**, *126*, 8702–8709.
[32] P. J. Hagerman, R. J. Hagerman, *Am. J. Hum. Genet.* **2004**, *74*, 805–816.

Received: December 22, 2006
Published online on ■■ ■■, 2007

**The single-stranded region** of the d-(TTAGGG) repeat is known to form a G-quadruplex structure in vitro. We report here a novel naphthyridine tetramer (NT) ligand that induces a nonquadruplex structure in the single-stranded region of the human telomeric sequence.




*Y. Goto, S. Hagihara, M. Hagihara, K. Nakatani**

■■–■■

**Small-Molecule Binding to the Nonquadruplex Form of the Human Telomeric Sequence**

Nucleic Acids Symposium Series No. 50 87–88
doi:10.1093/nass/nrl043

# Control of DNA hybridization by photoswitchable mismatch binding ligands

**Chikara Dohno, Shin-nosuke Uno and Kazuhiko Nakatani**

The Institute of Scientific and Industrial Research, Osaka University, 8-1 Mihogaoka, Ibaraki 567-0047, Japan

## ABSTRACT

**We herein demonstrate that mismatch binding ligands (MBL) can function as a molecular glue which brings two single stranded DNA (ssDNA) together to form the double stranded DNA (dsDNA). Incorporation of a photoisomerizable azobenzene linkage provides further ability of reversibly controlling duplex stability with light.**

## INTRODUCTION

Sequence-specific hybridization property of DNA constitutes the indispensable basis for the essential functions of DNA. For regulation of biological functions and construction of DNA-based materials, controlling DNA hybridization by external stimuli has become an increasingly important. Studies toward controlling or modulating the DNA hybridization with chemically modified oligonucleotides have been reported previously.[1-5] We here describe an approach to turn on the duplex formation by a small synthetic ligand from two natural non-modified ssDNA. In addition, reversible photo-control of the duplex stability has been explored by the introduction of photo-isomerizable moiety into MBL.

## RESULTS AND DISCUSSION

In switching the state of DNA hybridization, large stability difference between ssDNA-MBL and dsDNA-MBL complex provides an ideal switching threshold. Recent studies on the binding of MBL to the trinucleotide repeats revealed a novel mode of ligand binding to the mismatched DNA duplex, and this finding motivated us to develop a molecular glue of DNA.[6-8] The naphthyridine carbamate dimer (**NC**) selectively binds to the 5'-CGG-3'/5'-CGG-3' sequence (CGG/CGG) involving a G-G mismatch flanked by two C-G base pairs with a stoichiometry of **NC**:DNA to be 2:1 (Figure 1).[7] The **NC** binding to the CGG/CGG sequence induced two cytosines to be out of the π-stack. Since the unpaired cytosines are not directly engaged in the formation of **NC**-DNA complex, we anticipated that the flipped out cytosine could be substituted with other nucleotide base, such as thymine. The 5'-TGG-3'/5'-TGG-3' (TGG/TGG) sequence consists of three contiguous T-G, G-G, and G-T mismatches, and therefore, the dsDNA formation is energetically

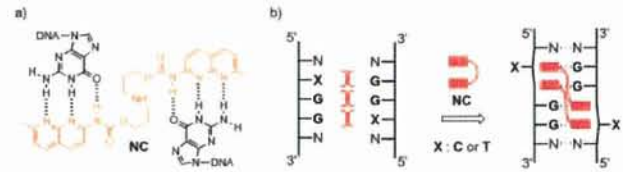



**Fig. 1** (a) Complementary hydrogen bonding between **NC** and two guanine bases. (b) A schematic illustration of the **NC** binding to the XGG/XGG sequence. **NC** induces the DNA hybridization.

unfavourable. These ssDNA, which do not spontaneously hybridize with each other, could be adhered by the **NC**-binding to the TGG/TGG sequence.[8] In fact, 11-mer 5'-(CCTT TGG TCAG)-3'/5'-(CTGA TGG AAGG)-3' were present as a ssDNA at room temperature due to the three contiguous mismatches. In marked contrast, melting temperature ($T_m$) of the 58.8 °C was observed in the presence of 100 μM **NC**, suggesting the formation of **NC**-bound TGG/TGG duplex.

MBL can act as a molecular glue which brings two ssDNA together to form the dsDNA. However, MBL does not have an ability to unfold the once formed dsDNA, and thus the induction of dsDNA formation by the MBL binding is fundamentally unidirectional. In order to control the DNA hybridization reversibly, we functionalized the MBL to be a photo-responsive DNA binder. **NC** is constituted of a base recognizing naphthyridine moiety and a flexible linker connecting the two naphthyridines. The linker is replaced by a photochromic azobenzene moiety for the design of a photo-switchable MBL, **$NC_2Az$** (Fig. 2). Azobenzene undergoes a reversible photoisomerization, which changes the relative orientations and positions of the naphthyridine units in **$NC_2Az$**, and thus modulates the binding ability.

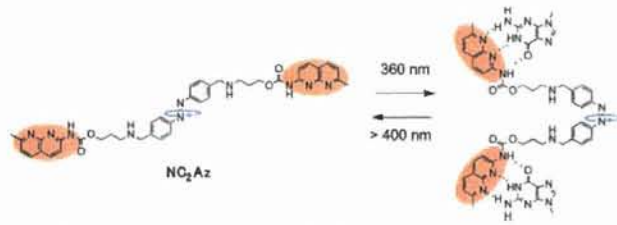



**Fig. 2** Photoisomerization of **$NC_2Az$**. UV light induces *trans* to *cis* isomerization of azobenzene moiety of **$NC_2Az$**, whereas visible light results in *cis* to *trans* isomerization.

Before photoirradiation, **$NC_2Az$** exists in the pure *trans*-form. Photoirradiation at a wavelength of 360 nm led to the *trans* to *cis* isomerization of **$NC_2Az$**. The *cis*/*trans* ratio after the irradiation was determined by HPLC detected at the isosbestic point. The photostationary state reached after 5 min irradiation consisted of a nearly equimolar mixture of *cis*- and *trans*-**$NC_2Az$**. The equilibrium mixtures reverted to the pure *trans* form of **$NC_2Az$** upon irradiation of visible light at 430 nm.

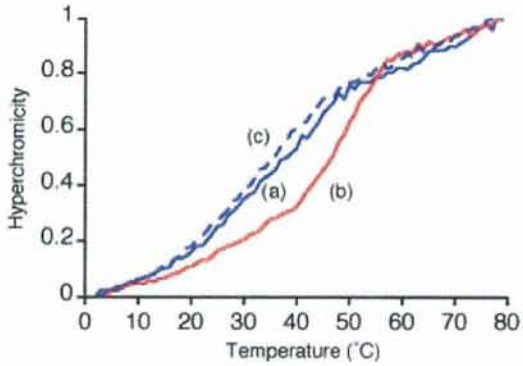



**Fig. 3** Thermal melting curves of GG-mismatch containing DNA duplex (4.5 $\mu$M) in the presence of **$NC_2Az$** (18.2 $\mu$M). The absorbance at 260 nm was measured in 10 mM Na·cacodylate buffer (pH 7.0) containing 0.1 M NaCl. Temperature was increased from 2 to 80 °C at a rate of 1°C/min. $T_m$ was calculated with the median method. (a) Without photoirradiation. (b) After photoirradiation at 360 nm for 5min. (c) (b) followed by irradiation at 430 nm for 5 min.

The **$NC_2Az$** binding to the mismatched DNA was evaluated by measuring $T_m$ of the **$NC_2Az$**-DNA complex. $T_m$ measurements were carried out using 11-mer DNA duplex 5'-(CTAA CGG AATG)-3'/5'-(CATT CGG TTAG)-3' containing a CGG/CGG mismatch sequence in the presence of **$NC_2Az$**. Without photoirradiation, **$NC_2Az$** exists in the *trans*-form, and the $T_m$ of 32.7°C was obtained (Figure 3a). Exposure of **$NC_2Az$** to 360 nm light increased the $T_m$ value by 15.2 °C ($T_m$ = 48.0°C, Figure 3b). Interestingly, the percentage of *cis*-**$NC_2Az$** (75%) was higher than that in the absence of the mismatched DNA. These results indicate *cis*-form of **$NC_2Az$** has higher binding ability to the CGG/CGG sequence. This is most likely because the nonplanar *cis*-azobenzne linkage in **$NC_2Az$** allows two naphthyridine units to be placed in the appropriate position for binding (Fig. 4). Thus, *trans* to *cis* photoisomerization of **$NC_2Az$** facilitated the formation of GG-mismatch-containing DNA duplex. Importantly, this duplex stabilization by **$NC_2Az$** was fully reversible. Irradiation of 430 nm light induced *cis* to *trans* transformation, and destabilized the duplex back to the normal mismatch-containing DNA (Fig. 3c).

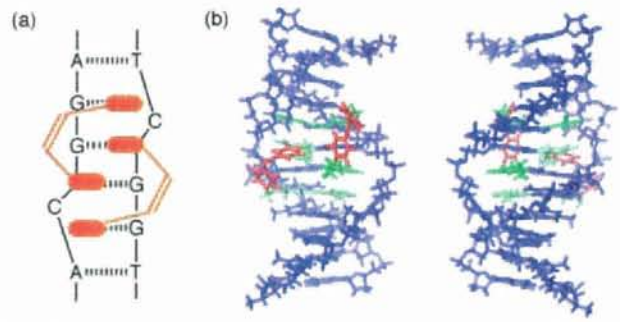



**Fig. 4** (a) Schematic representation of the binding of *cis*-**$NC_2Az$** to CGG/CGG sequence. (b) Molecular modelling of the *cis*-**$NC_2Az$**-CGG/CGG complex. The model structure was optimized by use of the AMBER* force field in water with MacroModel Version 9.1.

## CONCLUSION

MBL, such as **NC**, promotes DNA hybridization between the mismatch-containing DNAs by the formation of stable MBL-DNA complex. The introduction of a photochromic azobenzene linkage into MBL permits reversible control of duplex stability of the mismatch-containing DNA. The molecules described here represent a new class of compounds that function as a molecular glue not only in DNA hybridization but also in modulating the DNA secondary structures.

## REFERENCES

1. Asanuma, H., Ito, T., Yoshida, T., Liang, X., Komiyama, M. (1999) *Angew. Chem. Int. Ed.*, **38**, 2393-2395.
2. Asanuma, H., Liang, X., Yoshida, T., Yamazawa, A., Komiyama, M. (2000) *Angew. Chem. Int. Ed.*, **39**, 1316-1318.
3. Liu, Y., Sen, D. (2004) *J. Mol. Biol.*, **341**, 887-892
4. Ghosn, B., Haselton, F. R., Gee, K. R., Monroe, W. T. (2005) *Photochem. Photobiol.*, **81**, 953-959.
5. Hamad-Schifferli, K., Schwartz, J. J., Santos, A. T., Zhang, S. G., Jacobson, J. M. (2002) *Nature*, **415**, 152-155.
6. Nakatani, K., Hagihara, S., Goto, Y., Kobori. A., Hagihara, M., Hayashi, G., Kyo, M., Nomura, M., Mishima, M., Kojima, C. (2005) *Nat. Chem. Biol.*, **1**, 39-43.
7. Peng, T., Nakatani, K. (2005) *Angew. Chem. Int. Ed.*, **44**, 7280-7283.
8. Peng, T., Dohno, C., Nakatani, K. *Angew. Chem. Int. Ed.*, in press.

Nucleic Acids Symposium Series No. 50 147–148
doi:10.1093/nass/nrl073

# Inhibition of DNA replication by a d(CAG) repeat binding ligand

**Masaki Hagihara and Kazuhiko Nakatani**

The Institute of Scientific and Industrial Research, Osaka University, Mihogaoka 8-1, Ibaraki, Osaka 567-0047, Japan

## ABSTRACT

**Trinucleotide repeat expansions in genomic DNA are the molecular basis of a number of genetic diseases. The $(CAG)_n$, $(CTG)_n$, and $(CGG)_n$ repeats share the common sequence CXG, and inherently produce a hairpin structure involving X–X mismatch base pairs flanked by two G–C base pairs. The chance of a strand slippage leading to the repeat expansion is considered to increase with increasing stability of the hairpin form. Here we show our synthetic ligand naphthyridine–azaquinolone (NA) stabilized a hairpin form of the $(CAG)_n$ repeat and inhibited the polymerase-mediated DNA synthesis.**

## INTRODUCTION

The expansion of (CAG)n trinucleotide repeats in genomic DNA is related to the pathogenesis of Huntington's disease. The CAG repeat range is normally 9–35, whereas 38 or more repeats are found in the diseased state[1,2,3]. Although the mechanism of repeat expansion remains unclear, it is believed to involve strand slippage during DNA synthesis mediated by the formation of an alternative hairpin structure. The hairpin form of (CAG)n repeats involves the intramolecular pairing of CAG/CAG triads, the central A–A mismatch base pairs being flanked by two G–C base pairs. Thus ligands that bind to the CAG/CAG triad also are expected to bind to the hairpin form of the (CAG)n repeats. We have previously reported that a naphthyridine–azaquinolone (**NA**) ligand binds with high affinity to the CAG/CAG triad and we have recently shown the solution structure of **NA**–CAG/CAG complex by NMR spectroscopy[4]. NMR analysis revealed that two **NA** molecules intercalate into DNA helix, with the 2-amino-1,8-naphthyridine moiety presenting complementary hydrogen bonding to guanine, and 8-azaquinolone moiety presenting fully complementary hydrogen bonding to adenine. Furthermore binding of two **NA** molecules causes extrusion of two cytosines to the outside of the DNA helix (Figure 1).

In order to evaluate whether **NA** induced hairpin structure in CAG repeats can show the biological roles in DNA replication processes, we performed the thermal stability of **NA**-induced hairpin structure with different (CAG)n repeat length. The polymerase stop assay clearly showed that the hairpin formation by **NA**-binding to the (CAG)n repeat sequence in the template inhibited DNA polymerase to synthesize DNA strands.

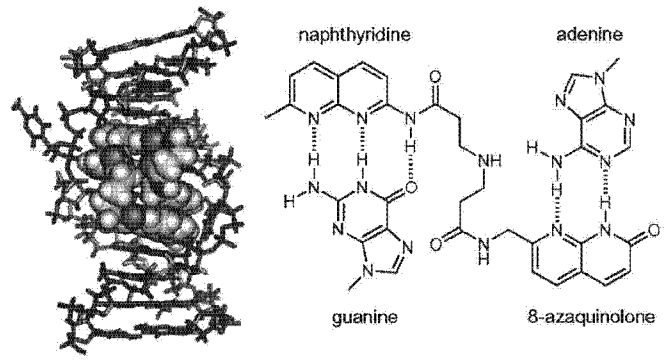



**Figure 1** (a) Structure of **NA**–CAG/CAG complex. (b) The naphthyridine chromophore makes a complemental hydrogen bonding surface to guanine, whereas 8-azaquinolone is complementally to adenine.

## RESULTS AND DISCUSSION

Strong **NA** binding to the CAG/CAG triad induced the formation of the **NA**-bound hairpin form in (CAG)n repeats DNA. The stabilization of hairpin structure was assessed by UV thermal denaturation studies. The thermal stability of $(CAG)_{10}$, ten repeats of CAG trinucleotides, was strongly enhanced by 24.7 °C in the presence of 20 μM **NA**.

Polymerase stop assay[5] was performed on the template containing ten repeat units of CAG sequence to test whether the hairpin structure induced by **NA** binding can interfere with the elongation of the (CAG)n repeat sequence. In the absence of **NA**, the 20-mer primer 5'-d(TAATACGACTCACTATAGGG)-3' that hybridized to the 3' end of the (CAG)n containing template was fully elongated by Klenow polymerase. As increasing the **NA** concentration, the intensity of the posed bands increased with a concomitant decrease of the fully elongated products. At 20 $\mu$M of **NA**, DNA synthesis by Klenow polymerase was almost stopped at the first CAG site (Figure 2). These results are consistent with a notion that inhibition of DNA polymerase was attributed to induced hairpin structure in CAG repeat DNA, but not to inhibition of Klenow DNA polymerase itself. These results showed that there was a good correlation between the thermal stability of induced hairpin structure and the inhibitory effects of DNA synthesis.

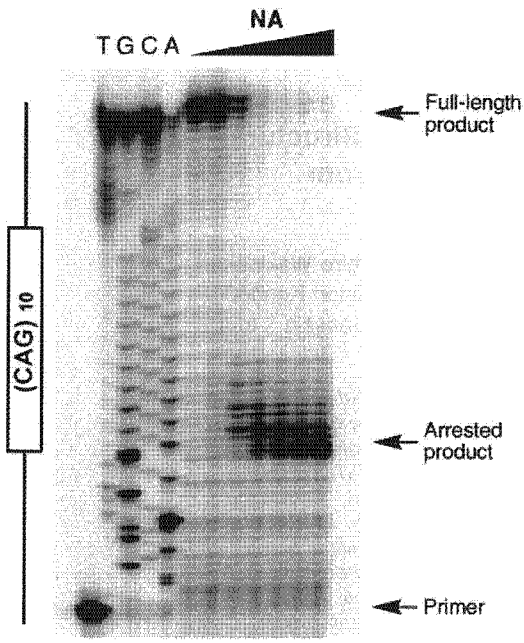



**Figure 2** Concentration-dependent inhibition of Klenow DNA polymerase-mediated DNA synthesis with **NA** (0, 5, 10, 15, 20, 25, 30 μM) on a DNA template containing the $(CAG)_{10}$ sequence. The lane markers T, G, C, and A indicate the bases on the template strand. Arrows indicate the positions of full-length DNA synthesis product, the arrested product, and the free primer.

## CONCLUSION

In conclusion, we found that **NA** leads to enhanced arrest of DNA synthesis by stabilizing an intrastrand hairpin structure formed by the (CAG)n repeat sequences. **NA** would be an important molecular probe for the repeat expansion mechanism and also effective therapeutic agents for genetic diseases.

## REFERENCES

1. Wells, R. D., Warren, S. T. (1998) *Genetic Instabilities and Hereditary Neurological Disease. Academic Press, San Diego.*
2. Stine, O. C., Pleasant, N., Franz, M. L., Abbott, M. H., Folstein, S. E., Ross, C. A. (1993) *Hum. Mol. Genet.*, **2**, 1547-1549.
3. Ashley, C. T., Warren, S. T. (1995) *Annu. Rev. Genet.*, **29**, 703-728.
4. Nakatai, K., Hagihara, S., Goto. Y., Kobori, A., Hagihara, M., Hayashi, G., Kyo, M., Nomura, M., Mishima, M., Kojima, C. (2005) *Nature Chemical Biology*, **1**, 39-43.
5. Han, H., Hurley, L. H., Salazar, M. (1999) *Nucleic Acids Res.*, **27**, 537-542.

J|A|C|S
ARTICLES
Published on Web 01/03/2004

# Nature of the Chemical Bond Formed with the Structural Metal Ion at the A9/G10.1 Motif Derived from Hammerhead Ribozymes

Yoshiyuki Tanaka,*,†,‡ Yasuhiro Kasai,† Shunsuke Mochizuki,§ Akihiro Wakisaka,§ Eugene H. Morita,||,⊥ Chojiro Kojima,# Atsushi Toyozawa,‡ Yoshinori Kondo,‡ Masumi Taki,†,▽ Yasuomi Takagi,† Atsushi Inoue,▽ Kazuhiko Yamasaki,+ and Kazunari Taira*,†,▽

*Contribution from the Gene Function Research Center, National Institute of Advanced Industrial Science and Technology (AIST), Central 4, 1-1-1 Higashi, Tsukuba Science City 305-8562, Japan, Graduate School of Pharmaceutical Sciences, Tohoku University, Aobayama, Sendai, Miyagi 980-8578, Japan, Environmental Molecular Science Group, AIST, Tsukuba, Ibaraki 305-8569, Japan, Integrated Center for Science, Ehime University, Matsuyama, Ehime 790-8566, Japan, Venture Business Laboratory, Ehime University, Matsuyama, Ehime 790-8577, Japan, Graduate School of Biological Science, Nara Institute of Science and Technology, Ikoma, Nara 630-0101, Japan, Department of Chemistry and Biotechnology, School of Engineering, The University of Tokyo, Hongo, Tokyo 113-8656, Japan, and Age Dimension Research Center, AIST, Tsukuba Science City 305-8566, Japan*

Received June 22, 2003; E-mail: taira@chembio.t.u-tokyo.ac.jp; tanaka@mail.pharm.tohoku.ac.jp

***Abstract:*** We have studied the interaction between metal ions and the metal ion-binding motif in hammerhead ribozymes, as well as the functions of the metal ion at the motif, with heteronuclear NMR spectroscopy. In this study, we employed model RNA systems which mimic the metal ion-binding motif and the altered motif. In $Co(NH_3)_6$(III) titrations, we observed large $^1H$ and $^{31}P$ chemical shift perturbations for the motif and found that outer-sphere complexation of $Co(NH_3)_6$(III) is possible for this motif. From the reinvestigation of our previous $^{15}N$ chemical shift data for Cd(II) binding, in comparison with those of organometallic compounds, we conclude that Cd(II) can form an inner-sphere complex with the nucleobase in the motif. Therefore, the A9/G10.1 site was found to accept both inner-sphere and outer-sphere complexations. The Mg(II) titration for a slightly different motif from the A9/G10.1 site (G10.1−C11.1 to A10.1−U11.1) revealed that its affinity to Mg(II) was drastically reduced, although the ribozyme with this altered motif is known to retain enzymatic activities. This observation suggests that the metal ion at these motifs is not a catalytic center of hammerhead ribozymes.

## Introduction

Biologically active RNA molecules utilize metal ions to fold into specific conformation or to form a catalytic center. In the case of hammerhead ribozymes, there is a metal ion-binding motif (the A9/G10.1 site) in its conserved core region (Figure 1).[1−6] To understand the mechanism of hammerhead ribozymes, it is important to analyze the interaction between metal ions and the consensus sequences of hammerhead ribozymes.[7−16]

† Gene Function Research Center, AIST.
‡ Tohoku University.
§ Environmental Molecular Science Group, AIST.
|| Integrated Center for Science, Ehime University.
⊥ Venture Business Laboratory, Ehime University.
# Nara Institute of Science and Technology.
▽ The University of Tokyo.
+ Age Dimension Research Center, AIST.

(1) Pley, H. W.; Flaherty, K. M.; McKay, D. B. *Nature* **1994**, *372*, 68−74.
(2) Scott, W. G.; Murray, J. B.; Arnold, J. R.; Stoddard, B. L.; Klug, A. *Science* **1996**, *274*, 2065−2069.
(3) Murray, J. B.; Terwey, D. P.; Maloney, L.; Karpeisky, A.; Usman, N.; Beigelman, L.; Scott, W. G. *Cell* **1998**, *92*, 665−673.
(4) Tanaka, Y.; Morita, E. H.; Hayashi, H.; Kasai, Y.; Tanaka T.; Taira, K. *J. Am. Chem. Soc.* **2000**, *122*, 11303−11310.
(5) Tanaka, Y.; Kojima, C.; Morita, E. H.; Kasai, Y.; Yamasaki, K.; Ono, A.; Kainosho, M.; Taira, K. *J. Am. Chem. Soc.* **2002**, *124*, 4595−4601.
(6) Suzumura, K.; Yoshinari, K.; Tanaka, Y.; Takagi, Y.; Kasai, Y.; Warashina, M.; Kuwabara, T.; Orita, M.; Taira, K. *J. Am. Chem. Soc.* **2002**, *124*, 8230−8236.
(7) Ruffner, D. E.; Stormo, G. D.; Uhlenbeck, O. C. *Biochemistry* **1990**, *29*, 10695−10702.
(8) (a) Peracchi, A.; Beigelman, L.; Usman, N.; Herschlag, D. *Proc. Natl. Acad. Sci. U.S.A.* **1996**, *93*, 11522−11527. (b) Peracchi, A.; Beigelman, L.; Scott, E. C.; Uhlenbeck, O. C.; Herschlag, D. *J. Biol. Chem.* **1997**, *272*, 26822−26826. (c) Wang, S.; Karbstein, K.; Peracchi, A.; Beigelman, L.; Herschlag, D. *Biochemistry* **1999**, *38*, 14363−14378.
(9) (a) Tuschl, T.; Gohlke, C.; Jovin, T. M.; Westhof, E.; Eckstein, F. *Science*, **1994**, *266*, 785−789. (b) Bassi, G. S.; Mollegaard, N. E.; Murchie, A. I.; von Kitzing, E.; Lilley, D. M. J. *Nat. Struct. Biol.* **1995**, *2*, 45−55. (c) Sigurdsson, S. T.; Tuschl, T.; Eckstein, F. *RNA*, **1995**, *1*, 575−583. (d) Bassi, G. S.; Murchie, A. I.; Lilley, D. M. J. *RNA* **1996**, *2*, 756−768. (e) Bassi, G. S.; Murchie, A. I.; Walter, F.; Clegg, R. M.; Lilley, D. M. J. *EMBO J.* **1997**, *16*, 7481−7489.
(10) Hampel, K. J.; Burke, J. M. *Biochemistry* **2003**, *42*, 4421−4429.
(11) Scott, E. C.; Uhlenbeck, O. C. *Nucleic Acids Res.* **1999**, *27*, 479−484.
(12) Yoshinari, K.; Taira, K. *Nucleic Acids Res.* **2000**, *28*, 1730−1742.
(13) Nakamatsu, Y.; Warashina, M.; Kuwabara, T.; Tanaka, Y.; Yoshinari, K.; Taira, K. *Genes Cells* **2000**, *5*, 603−612.
(14) Murray, J. B.; Scott, W. G. *J. Mol. Biol.* **2000**, *296*, 33−41.
(15) Takagi, Y.; Warashina, M.; Stec, W. J.; Yoshinari, K.; Taira, K. *Nucleic Acids Res.* **2001**, *29*, 1815−1834.


10.1021/ja036826t CCC: $27.50 

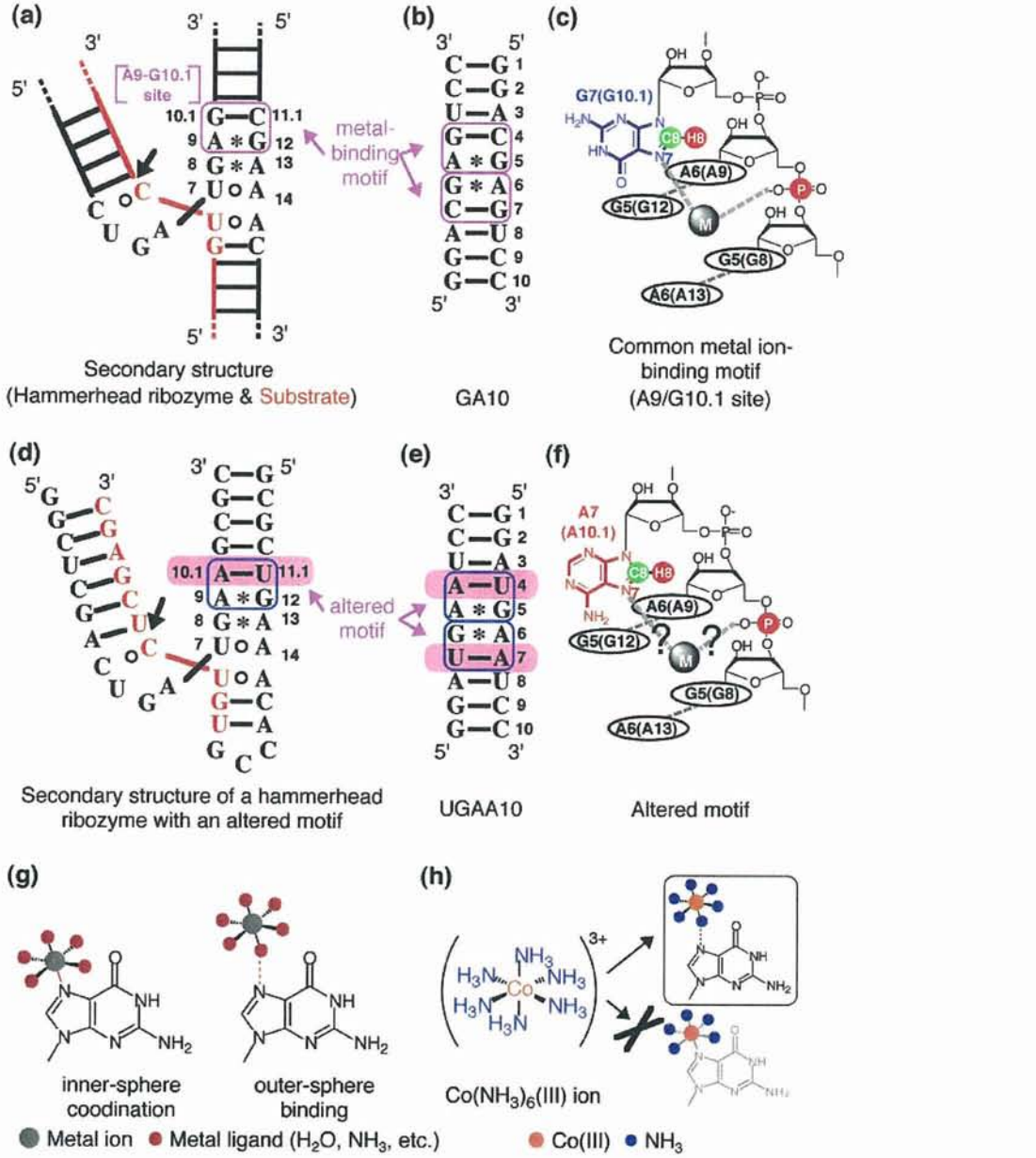



***Figure 1.*** Sequences and metals. (a) Common sequences and a secondary structure of hammerhead ribozymes. (b) GA10 with numbering systems. (c) Schematic representation of the common metal ion-binding motif (the A9/G10.1 site) of hammerhead ribozymes. The residue numbers in GA10 are shown, and the residue numbers in hammerhead ribozymes are also shown in parentheses. (d) Sequences and the secondary structure of the hammerhead ribozyme with A10.1−U11.1, studied by Ruffner et al.[7] (e) UGAA10: A GA10 analogue with an altered motif. (f) Schematic representation of the altered motif. In (a), (b), (d), and (e), the common metal-binding motif (the A9/G10.1 site) and the altered motif are surrounded by magenta and blue lines, respectively. The substituted sequences are highlighted with magenta background. In (a) and (d), enzyme and substrate strands are shown in black and red, respectively, and the cleavage site is indicated by a black arrow. In (c) and (f), a metal ion is shown in a gray circle with a character "M" and is linked to binding sites via dashed lines. The tandem G−A pairs are indicated by black dashed lines with residue names. Watson−Crick base pairs, non-Watson−Crick base pairs, and sheared-type G−A pairs are indicated in bars, open circles, and asterisks. The nuclei monitored by 1D $^{31}$P NMR spectra and 2D natural-abundance $^{1}$H−$^{13}$C HSQC spectra are highlighted (P, red; C8, green; H8, red). In (g), metal ion recognition modes are illustrated (inner-sphere coordination and outer-sphere binding [hydrogen bonding through metal ligands]). In (h), $Co(NH_3)_6$(III) binding modes are illustrated. In the case of $Co(NH_3)_6$(III), an inner-sphere coordination of the Co(III) center to nucleobases is inhibited; however, outer-sphere binding is allowed because of exchange-inert metal ligands ($NH_3$).

Specifically, it is becoming important whether the metal ions at the A9/G10.1 site are catalytic metals (a catalytic center) or structural metals (structural constituents), since this metal ion is the most frequently observed metal ion throughout the crystal structures of hammerhead ribozymes.[1−6,8,11−16] It is also known that this metal ion-binding motif and very similar motifs appear in ribosomal RNAs.[17] Therefore, it is obviously important to study the interaction between metal ions and the metal ion-binding motif.

Recently, studies on the metal−biomacromolecule interaction became one of the topics in biology, and there have been several

(16) O'rear, J. L.; Wang, S.; Feig, A. L.; Beigelman, L.; Uhlenbeck, O. C.; Herschlag, D. *RNA* **2001**, *7*, 537−545.

(17) Gautheret, D.; Konings, D.; Gutell, R. R. *J. Mol. Biol.* **1994**, *242*, 1−8.

*Table 1.* Chemical Shift Perturbations upon Additions of Metal Ions[a]

| sequences | metals | residue | C8/ppm | H8/ppm | ref. |
|---|---|---|---|---|---|
| r(GGACGAGUCC)$_2$ [GA10] | $MgCl_2$ | G7 | +0.5 (2.0 equiv) | +0.12 (9.0 equiv) | 4 |
| r(GGACGAGUCC)$_2$ [GA10] | $CdCl_2$ | G7 | +2.3 (5.0 equiv) | +0.38 (9.0 equiv) | 5 and this work |
| r(GGACGAGUCC)$_2$ [GA10] | $Co(NH_3)_6Cl_3$ | G7 | +0.1 (2.0 equiv) | +0.25 (3.0 equiv) | this work |
| r(GGACGAGUCC)$_2$ [GA10] | $NaClO_4$ | G7 | +0.1 (230 mM)[b] | +0.12 (800 mM)[b] | this work |
| d(ATGGGTACCCAT)$_2$ [12mer] | $ZnCl_2$ | G4 | +2.5 (8.0 equiv) | +0.20 (8.0 equiv) | 42 |
| d(ATGGGTACCCAT)$_2$ [12mer] | $ZnCl_2$ | G3 | +1.5 (8.0 equiv) | +0.05 (8.0 equiv) | 42 |
| d(TGGT) | Pt(en)$Cl_2$[c] | G3 | +1.1 (1.0 equiv) | +1.04 (1.0 equiv) | 43 |
| d(TGGT) | Pt(en)$Cl_2$[c] | G2 | +0.2 (1.0 equiv) | +0.26 (1.0 equiv) | 43 |

[a] Chemical shift perturbations were listed in ppm with molar equivalencies to duplexes (GA10 and 12mer) or a single strand of d(TGGT). Plus values indicate lower field shift. [b] Chemical shift perturbations from the basal solutions are listed. [c] Pt(en)$Cl_2$ (en = ethylenediamine) forms covalent complex d(TGGT)·Pt(en) with the two successive guanosines of d(TGGT).[43,44]

studies on this topic using solution NMR techniques. In the case of nucleic acids, there have been studies on the interaction between metal ions and nucleic acids.[4−6,18−23] DeRose and coworkers demonstrated that in hammerhead ribozymes there is at least one nitrogen atom which is coordinated to Mn(II), with the EPR spectrum measurements of Mn(II), and they suggested the Mn(II) coordination site on the ribozyme.[24−26] The solution NMR technique provides us with a more rigid answer for a metal ion-binding site in the RNA molecule.[4−6,18−23] Despite those extensive works, there is no direct evidence for what kind of chemical bonds were formed between metal ions and nucleobases in ribozymes. In the case of organometallic complexes for organic synthesis, it has been demonstrated that metal ligands dramatically influence reactivities, catalytic activities, and regio- and stereoselectivities of chemical reactions.[27] Thus, for understanding the mechanism of the action of hammerhead ribozymes, it is important to know not only what ligands are on the metal center, but also what kinds of chemical bonds are formed between the metal center and ligands. In our previous studies using heteronuclear multidimensional NMR spectroscopy, we used GA10 as a model RNA oligomer mimicking the metal ion-binding motif of hammerhead ribozymes (Figure 1) and found that Cd(II) specifically bound to the N7 nuclei of the G7 residue [N7(G7)].[4,5] However, we could not definitely conclude whether the binding of Cd(II) to N7(G7) was direct coordination or not.[4,5] This is because we could not detect the scalar coupling between $^{113}$Cd(II) and $^{15}$N7(G7),[5] as were the cases for other nucleoside−metal ion systems.[28,29] For further information, it is interesting to introduce several kinds of metal ion sources with different characters, such as $Co(NH_3)_6Cl_3$, $NaClO_4$, $MnCl_2$, $EuCl_3$, and $TbCl_3$.

The $Co(NH_3)_6$(III) ion is known as a model for a hexahydrated Mg(II) ion, since the ionic radius of $Co(NH_3)_6$(III), including $NH_3$ ligands, is very close to that of hexahydrated magnesium ions $[Mg(H_2O)_6]$.[30] In addition, $Co(NH_3)_6$(III) ions are ligated with exchange-inert $NH_3$ ligands, and then nucleobases cannot be directly coordinated to the Co(III) center of this complex (Figure 1g,h).[30] Therefore, it can be used for a control metal ion that cannot form an inner-sphere coordination.[30] Sodium ion is used to monitor the degree of ionic strength-dependent chemical shift changes. The Mn(II) ion is used to demonstrate whether metal ions are surely located near the nucleobase, since its paramagnetic property causes signal broadening around the binding site in a distance-dependent manner.[18−20] The Eu(III) and Tb(III) ions were also used as a fluorescent probe. Using these data as well as those of Cd(II) and Mg(II), we tried to reveal what kinds of chemical bonds are formed between a metal ion and the metal ion-binding motif in hammerhead ribozymes. The role of the metal ion at this motif was also investigated using a nucleotide-substituted RNA. Ruffner et al. indicated that the hammerhead ribozyme with the A10.1−U11.1 pair retained 30% cleavage activity relative to that with a common G10.1−C11.1 pair.[7] If the metal ion at the A9/G10.1 site is a catalytic metal of hammerhead ribozymes, the altered motif with A−U pair at the 10.1 and 11.1 positions should capture a metal ion. To clarify this point as well as the sequence requirements for the metal ion recognition, we performed titration experiments of the altered motif with Mg(II), the most probable cofactor in physiological conditions.[31]

## Results

**Titration Experiments with $^1$H and $^{31}$P NMR Spectra.** In previous studies, we have demonstrated that divalent cations, such as Mg(II) and Cd(II), were able to bind to the metal ion-binding motif through the N7 nucleus of G7 in GA10 using $^1$H, $^{13}$C, and $^{15}$N chemical shift perturbation (Tables 1 and 2).[4,5] In those studies, large chemical shift perturbations were observed for N7, C8, and H8 resonances of the G7 residues (−19.6, 2.3, and 0.38 ppm, respectively) upon the addition of Cd(II) ions (Tables 1 and 2).[4,5] Therefore, it was suggested that Cd(II) and Mg(II) ions directly coordinated to N7(G7). To clarify this situation in detail, we employed $Co(NH_3)_6$(III) ions, since exchange-inert $NH_3$ ligands prevent an inner-sphere coordination of the Co(III) center to nucleobases. If the direct coordination of a metal center is a prerequisite for the metal ion binding to the motif, $Co(NH_3)_6$(III) ions cannot bind to the motif. Contrary

(18) Allain, F. H.-T.; Varani, G. *Nucleic Acids Res.* **1995**, *23*, 341−350.
(19) Hansen, M. R.; Simorre, J. P.; Hanson, P.; Mokler, V.; Bellon, L.; Beigelman, L.; Pardi, A. *RNA* **1999**, *5*, 1099−1104.
(20) Rüdisser, S.; Tinoco, I., Jr. *J. Mol. Biol.* **2000**, *295*, 1211−1223.
(21) Butcher, S. E.; Allain, F. H.-T.; Feigon, J. *Biochemistry* **2000**, *39*, 2174−2182.
(22) Schmitz, M.; Tinoco, I., Jr. *RNA* **2000**, *6*, 1212−1225.
(23) Maderia, M.; Hunsicker, L. M.; DeRose, V. J. *Biochemistry* **2000**, *39*, 12113−12120.
(24) (a) Morrissey, S. R.; Horton, T. E.; Grant, C. V.; Hoogstraten, C. G.; Britt, R. D.; DeRose, V. J. *J. Am. Chem. Soc.* **1999**, *121*, 9215−9218. (b) Morrissey, S. R.; Horton, T. E.; DeRose, V. J. *J. Am. Chem. Soc.* **2000**, *122*, 3473−3481.
(25) Hoogstraten, C. G.; Grant, C. V.; Horton, T. E.; DeRose, V. J.; Britt, R. D. *J. Am. Chem. Soc.* **2002**, *124*, 834−842.
(26) Hunsicker, L. M.; DeRose, V. J. *J. Inorg. Biochem.* **2000**, *80*, 271−281.
(27) (a) Hegedus L. S. *Transition Metals in the Synthesis of Complex Organic Molecules*, 2nd ed.; University Science Books: Sausalito, CA, 1999. (b) Imahori, T.; Uchiyama, M.; Sakamoto, T.; Kondo, Y. *Chem. Commun.* **2001**, *23*, 2450−2451. (c) Uchiyama, M.; Miyoshi, T.; Kajihara, Y.; Sakamoto, T.; Otani, Y.; Ohwada, T.; Kondo, Y. *J. Am. Chem. Soc.* **2002**, *124*, 8514−8515.
(28) Buchanan, G. W.; Stothers, J. B. *Can. J. Chem.* **1982**, *60*, 787−791.
(29) Buchanan, G. W.; Bell, M. J. *Can. J. Chem.* **1983**, *61*, 2445−2448.
(30) Cowan, J. A. *J. Inorg. Biochem.* **1993**, *49*, 171−175.
(31) Zhou, J.-M.; Zhou, D.-M.; Takagi, Y.; Kasai, Y.; Inoue, A.; Baba, T.; Taira, K. *Nucleic Acids Res.* **2002**, *30*, 2374−2382.

*Table 2.* Summary of $^{15}N$ NMR Data

| metal ligands (ref.) | [ligand]/mM | metals | residue | $\Delta^{15}N^a$/ppm | $^1J_{N-M(II)}{}^b$/Hz |
|---|---|---|---|---|---|
| r(GGACGAGUCC)$_2$ [GA10] (5) | 0.45 | $CdCl_2$ | G7 | −19.6 (6.0 equiv) | not detected |
| guanosine monomer (28) | 500 | $Zn(NO_3)_2$ | n.a.[c] | −20.1 (1.0 equiv) | n.a. |
| guanosine monomer (28) | 500 | $Hg(NO_3)_2$ | n.a. | −20.5 (1.0 equiv) | not reported |
| inosine monomer (29) | 1000 | $Zn(NO_3)_2$ | n.a. | −15.2 (0.7 equiv) | n.a. |
| inosine monomer (29) | 1000 | $Hg(NO_3)_2$ | n.a. | −4.8 (0.75 equiv) | not detected |
| 1-methylimidazole[d] [MI] (33) | 240 | $Cd(NO_3)_2$ | n.a. | −28.7[d] (0.2 equiv) | not reported |
| 1-methylimidazole[d] [MI] (33) | 350 | $Zn(NO_3)_2$ | n.a. | −46.0[d] (0.22 equiv) | n.a. |
| $\beta$-lactamase (35) | 0.8 | $CdCl_2$ | His[e] | n.r.[f] | ~78−216 |
| EDTA (36) | 400 | $CdCl_2$ | n.a. | +3.3 (1.9 equiv) | 81 |

[a] Chemical shift perturbations of N7(guanosine), N3(1-methylimidazole), and tertiary amine nitrogens (EDTA) from the unmetalated state. Minus and plus values indicate higher and lower field shifts, respectively. [b] $J$-coupling between metalated nitrogens ($^{15}N$) and metal ions ($I = 1/2$), such as $^{113}Cd$ and $^{199}Hg$. Only absolute values are indicated. [c] Not applicable. [d] In the equilibrium system of 1-methylimidazole [MI] and metal ions [M(II)], the $^{15}N$ chemical shift of N3 reflects an average value of those from free, protonated, and metal complexes (M(II)-MI, M(II)-2·MI, M(II)-3·MI, M(II)-4·MI, and so on). Therefore, chemical shift values of 1-methylimidazole are average values of these species. [e] Four of seven histidines were metalated, but residue numbers in the amino acid sequence were not reported. [f] Not reported. However, in $\beta$-lactamase-Cd(II) system, ranges of $^{15}N$ chemical shifts of metalated and unmetalated residues were overlapped, and $^{15}N$ chemical shift perturbations due to a metalation were estimated to be small. For the interpretations of the $^{15}N$ NMR chemical shift, Alei and co-workers presented interesting data on the $^{15}N$ chemical shift of 1-methylimidazole.[33] Chemical structure of the five-membered ring of the guanosine base is the same as 1-methylimidazole, and N7 of guanosine corresponds to N3 of 1-methylimidazole (Figure 7). They indicated that the $^{15}N$ chemical shift of N3, a metal ion-coordination site of 1-methylimidazole, was perturbed toward high field (~15−30 ppm) upon the complexation with Cd(II) in $H_2O$.[33] In this system, it seems difficult for an imidazole ring to form stable outer-sphere complexes with Cd(II) (the outer-sphere binding via water molecules of Cd(II) ligands), since 1-methylimidazole is a monodentate ligand. In other words, the observed complex is thought to be a coordination compound, and the higher field shift of the N3 resonance of 1-methylimidazole is due to an inner-sphere coordination of the N3 atom to Cd(II). Therefore, when the imidazole ring nitrogen coordinates to transition metal ions, such as Cd(II) and Zn(II), the $^{15}N$ chemical shift can be perturbed toward high field.

to this, if $Co(NH_3)_6$(III) ions can bind to the motif, it is deduced that this metal ion-binding motif allows the metal ion binding through the metal ligands, including water molecules. To estimate ionic strength-dependent chemical shift changes, titration experiments with $NaClO_4$ were also performed. On the basis of the above concepts, we performed titration experiments with $Co(NH_3)_6Cl_3$ and $NaClO_4$, using one-dimensional (1D) $^1H$ and $^{31}P$ NMR spectra (Figure 2), as well as those of $CdCl_2$ (an example for tight binding).

Contrary to our early prediction, titration data with 1D $^1H$ and $^{31}P$ NMR spectra indicated $Co(NH_3)_6$(III) ion binding to GA10, since large chemical shift perturbations were observed for the resonances in the motif (the H8 resonance of the G7 residue [H8(G7)], phosphorus of the A6 residue [P(A6)], and the neighboring residues in GA10) (Figure 2). In contrast to the titrations for $Co(NH_3)_6$(III) and Cd(II), sodium ion titration curves of $^1H$ and $^{31}P$ resonances were almost linearly increased against the concentration of $NaClO_4$, except for those at the 5′ end (Figure 2). It was also found that ionic strength-dependent chemical shift perturbations for proton resonances were small (the maximum perturbation: 0.0002 ppm per 1 molar equivalent $NaClO_4$ for the H6 proton of the C4 residue) and negligible for the titration data of Cd(II) and $Co(NH_3)_6$(III). It was found that metal ion binding to the motif was not a simple electrostatic interaction, and sodium ions do not bind to the motif in a site-specific manner. This is consistent with the result by Herschlag's group that monovalent lithium ions did not bind to the motif during the catalytic reaction of hammerhead ribozymes.[16] Therefore, the chemical shift perturbations by Cd(II) and $Co(NH_3)_6$(III) titrations were due to their site-specific binding to GA10.

**Intermolecular NOEs between GA10 and $Co(NH_3)_6$(III).** We tried to exclude the possibility that the chemical shift perturbations for $Co(NH_3)_6$(III) titrations were due to the secondary effect from $Co(NH_3)_6$(III) binding to somewhere outside the motif. For this purpose, we measured two-dimensional (2D) $^1H-^1H$ NOESY spectra (90% $H_2O$, 10% $D_2O$). In the spectrum, several intermolecular NOEs were observed between protons of $Co(NH_3)_6$(III) ions and base protons of GA10 (Figure 3). Especially between the $Co(NH_3)_6$(III) ions and H8(G7) was one of the strongest intermolecular NOEs, and other NOEs were observed for the base protons and amino protons around the A9/G10.1 site (Figure 3a). In the upper panel of Figure 3a, possible NOE counterparts for $Co(NH_3)_6$(III) are indicated. A strong intermolecular NOE was also observed between the imino proton of the G7 residue and the $NH_3$ ligands (Figure 3b). It was surely demonstrated that the $Co(NH_3)_6$(III) ion bound to the A9/G10.1 site selectively.

**$^1H-^{13}C$ HSQC Spectra of GA10.** We recorded 2D $^1H-^{13}C$ HSQC spectra with or without metal salts such as $Co(NH_3)_6Cl_3$, $CdCl_2$, and $NaClO_4$ (Figure 4). As previously shown,[4,5] Cd(II) ions significantly perturbed the C8(G7) resonance toward low field (Figure 4a,b). In contrast, $Co(NH_3)_6Cl_3$ and $NaClO_4$ only perturbed the C8(G7) resonance within an error range of absolute chemical shift values (Figure 4c−f). From the titration and NOE data for the GA10−$Co(NH_3)_6$(III) system, $Co(NH_3)_6$(III) is thought to bind to N7(G7) via $NH_3$ ligand-mediated hydrogen bonding (outer-sphere binding); however, this outer-sphere binding did not affect the chemical shift of C8(G7). In other words, the lower field shift of C8(G7) upon the addition of Cd(II) is not explainable without an inner-sphere coordination of a metal ion to N7(G7). It was also found that sodium ions did not affect the chemical shifts of C8 atoms of all residues significantly.

**Effects of a Paramagnetic Metal Ion.** To exclude the possibility that the chemical shift perturbations for C8(G7) upon addition of Cd(II) and Mg(II) merely reflect the conformational changes due to their binding to allosteric sites, we have monitored signal broadening due to paramagnetic Mn(II) ion binding. This technique indicates that Mn(II) ions surely locate near the site where the resonances were broadened since the signal broadening occurs in a distance-dependent manner.[18−20] It is also known that Mn(II) and Cd(II) ions belong to relatively soft acids, according to the hard and soft acids and bases (HSAB) rule, and kinetic and thermodynamic behavior of Mn(II) is similar to that of Cd(II) in the catalytic reaction of ham-

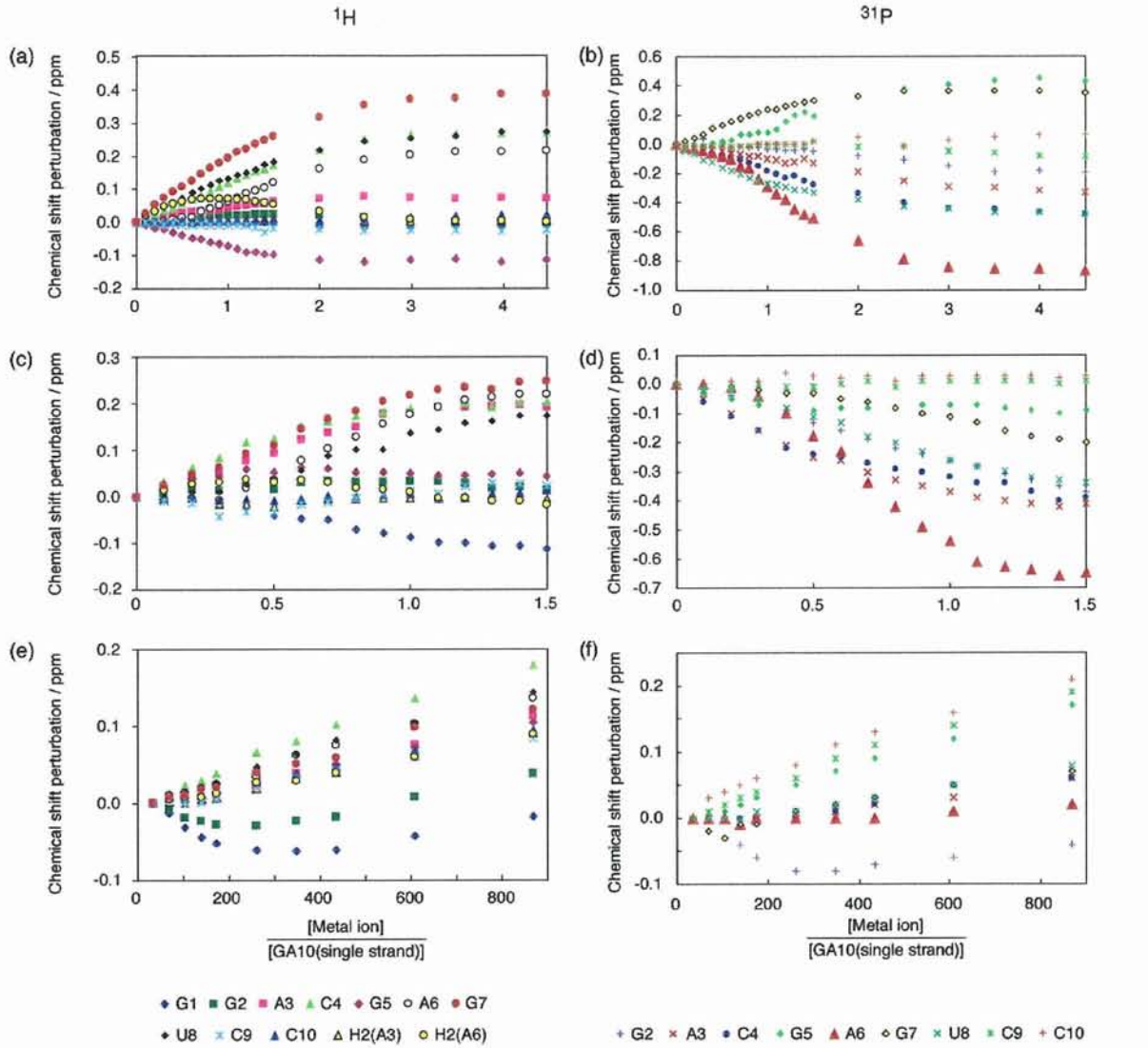



***Figure 2.*** Chemical shift perturbations for base protons [H8(purine), H6(pyrimidine), and H2(adenosine)] against a molar ratio ([metal ion]/[GA10(single strand)]). In (a), (c), and (e), those of $^{1}H$ resonances were plotted for Cd(II), $Co(NH_3)_6$(III), and sodium ions, respectively. In (b), (d), and (f), those of $^{31}P$ resonances were plotted for Cd(II), $Co(NH_3)_6$(III), and sodium ions, respectively. For Cd(II) titrations, various concentrations of $CdCl_2$ were added to the solution containing 1.5 mM GA10(single strand) and 40 mM $NaClO_4$. For $Co(NH_3)_6$(III) titrations, various concentrations of $Co(NH_3)_6Cl_3$ were added to the solution containing 2.0 mM GA10(single strand) and 40 mM $NaClO_4$. For sodium titrations, various concentrations of additional $NaClO_4$ were added to the solution containing 1.1 mM GA10(single strand) and 40 mM $NaClO_4$. Typical 1D $^{1}H$, $^{31}P$ NMR spectra were recorded on Bruker DMX 750 spectrometer at 313 K. For $^{1}H$ NMR spectra, spectral width was 10 000 Hz digitized into 16 384 complex points; 128 scans were averaged. For $^{31}P$ NMR spectra, spectral width was 10 000 Hz digitized into 16384 complex points; 1024 scans were averaged. We confirmed that resonance assignments were consistent with other spectra, such as 2D $^{1}H-^{1}H$ NOESY, $^{1}H-^{13}C$ HSQC, and $^{1}H-^{31}P$ HMQC NOESY spectra.

merhead ribozymes.[8−12,15,16] Accordingly, Mn(II) is regarded as a model of Cd(II). From this experiment, broadening of H8-(G7) was evident (Figure 5). It was demonstrated that Mn(II) was surely located near H8(G7). Therefore, Cd(II) is also thought to be located near N7(G7), and the chemical shift perturbation of H8, C8, and N7 of G7 should be due to the metal ion binding to N7(G7).

**Sequence Requirements for the Metal Ion Binding.** We have also tested whether the G7 residue (G10.1 in hammerhead ribozymes) can be replaced by adenosine without a loss of the affinities for divalent cations. In the metal ion-binding motif of GA10, N7(G7) was shown to be a Cd(II)-binding site, whereas O6(G7) was not.[5] Although there is one example where O6-(G10.1) is utilized as a metal ligand in the crystal structure of the hammerhead ribozyme−Co(II) complex,[3] it is not a definitive requirement for other systems.[1,2] Therefore, we thought that the guanosine residue at the G10.1 position is able to be substituted by an adenosine residue. Then, we synthesized the RNA oligomer, UGAA10:rGGAUGAAUCC, and measured the NMR spectra. It had been demonstrated that UGAA10 forms a duplex with sheared type G−A pairs in tandem at the middle of the duplex, and the UGAA sequence took a similar structure to that of r(GGGCUGAAGCCU)$_2$ by Heus et al.[32] It is also known that the hammerhead ribozyme with A10.1−U11.1

(32) (a) Tanaka, Y.; Kasai, Y.; Morita, E. H.; Kojima, C.; Toyozawa, A.; Yamasaki, K.; Taira, K.; Kondo Y. *Nucleic Acids Res.* **2003**, (Suppl. 3), 45−46. (b) Heus, H. A.; Wijmenga, S. S.; Hoppe, H.; Hilbers, C. W. *J. Mol. Biol.* **1997**, *271*, 147−158.

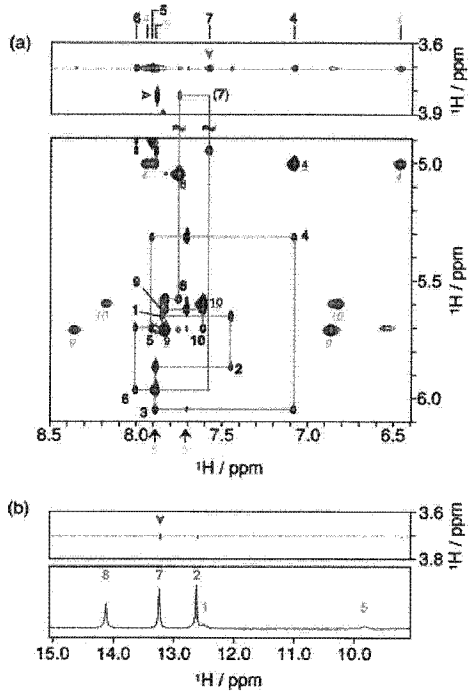



*Figure 3.* 2D $^1$H−$^1$H NOESY spectrum in the presence of 1 molar equivalent of $Co(NH_3)_6Cl_3$ to the motif. The solution contained 3.4 mM GA10, 40 mM $NaClO_4$, and 3.4 mM $Co(NH_3)_6Cl_3$. The chemical shift of $NH_3$ ligands of $Co(NH_3)_6$(III) are indicated by red dashed lines. (a) NOE cross-peaks from $Co(NH_3)_6$(III) to GA10 (upper panel) and sequential NOE between base protons and anomeric protons (upper and lower panels) are presented. These intraresidue NOE cross-peaks were labeled with their residue number. The underlined numbers and green italic numbers are intraresidue H5−H6 cross-peaks and H5 amino protons cross-peaks for the corresponding residue number, respectively. Below the lower panel of (a), black arrows indicate the presence of H2 resonances on the corresponding chemical shift, with the residue numbers (orange numbers). NOE counterparts for $Co(NH_3)_6$(III) are also indicated above the upper panel of (a), and that of H8(G7) is indicated with a red arrowhead. We assigned the resonance at 3.81 ppm as H1′ of G7, since it is well-known that the H1′ resonances of the guanosine in the C−G pair, which is adjacent to a sheared type G−A pair in the same arrangement as the A9/G10.1 site, always resonated at an extraordinarily high field.[41] Katahira et al. assigned $^1$H resonances of the same sequence as GA10 in the absence of $Co(NH_3)_6$(III) (H1′(G7) at 4.14 ppm for a solution containing 1.8 mM GA10, 20 mM sodium-phosphate buffer pH 7.0, and 150 mM NaCl, at 306 K).[41b] Heus and Pardi assigned $^1$H resonances of the RNA oligomer r(GGGC*GCAA*GCCUUAU), including *GCAA* loop with a *C*−*G* closing base pair, adjacent to a sheared type G−A pair in the loop (H1′(G9) at approximately 3.7 ppm for a solution containing 1.9 mM RNA, 10 mM sodium phosphate buffer pH 6.8, and 100 mM NaCl at 298 K).[41a] It should be noted that in the absence of $Co(NH_3)_6$(III), H1′-(G7) of GA10 resonated at 3.99 ppm (1.2 mM GA10, 40 mM NaCl, pH 6 at 313 K).[39] Therefore, the chemical shift of H1′(G7) is significantly dependent on the measurement conditions, and slightly different chemical shifts for Katahira's conditions and ours are probably attributed to the differences in conditions. We have also confirmed this assignment by monitoring the NOE cross-peak (an open arrowhead) with H2(A6) (7.89 ppm), as Katahira et al. showed.[41b] This NOE cross-peak was observed in all 2D $^1$H−$^1$H NOESY spectra in $D_2O$ in the presence and absence of $CdCl_2$ and $Co(NH_3)_6Cl_3$.[39] In (b), 1D $^1$H NMR spectrum of an imino proton region (lower panel) and NOE cross-peaks between $Co(NH_3)_6$(III) and imino protons are presented. In the lower panel of (b), residue numbers of imino protons are indicated. In the upper panel of (b), the cross-peak between imino proton of G7 and $Co(NH_3)_6$(III) is pointed to with a red arrowhead. In the NOESY spectrum, weak NOE cross-peaks from $Co(NH_3)_6$(III) were observed for nonspecific sites at this threshold. However, this phenomena might be a general feature of this experiment, since nonspecific NOEs from $Co(NH_3)_6$(III) were observed in other studies as well.[21,22] A 2D $^1$H−$^1$H NOESY spectrum was recorded on Bruker DMX750 at 293 K; 4096*2048 complex points for a spectral width of 15 700*15 000 Hz.

instead of G10.1−C11.1 retains 30% cleavage activity, relative to the hammerhead ribozyme with the common G10.1−C11.1 pair.[7] This altered sequence motif is the same as that in UGAA10. Therefore, if the metal ion at A9/G10.1 and the corresponding site is a catalytic metal, this altered motif in UGAA10 should also capture divalent cations. However, chemical shift perturbations for UGAA10 were much less than those observed for GA10 upon the addition of 1 molar equivalent of $MgCl_2$ to the metal ion-binding motif (Figure 6). The maximum chemical shift perturbation of base protons in UGAA10 was observed for H8(G2) (0.03 ppm), the distal site from the putative metal ion-binding site (A7) (Figure 6). In addition, this perturbation is much smaller than those due to the specific binding of metal ions to GA10 (0.38 ppm for H8(G7) in Cd(II) titrations). Therefore, it is plausible that the chemical shift perturbations of the proton resonances in UGAA10 merely reflect changes in surrounding environments and cannot be attributed to specific binding of Mg(II). It is also known that nonspecific binding of divalent cations to backbone phosphates as counterions can occur at this concentration.[5] Therefore, the affinity between the motif in UGAA10 and Mg(II) is comparable to or lower than that between backbone phosphates and counterions of divalent cations. Accordingly, we can conclude that this altered motif with an adenosine residue at 10.1 position was not sufficient to capture divalent cations. From this result, in conjunction with sequence requirements for the cleavage reaction of hammerhead ribozymes,[7] it was suggested that the role of the metal ion at the A9/G10.1 site is not a catalytic metal, but a structural metal for the folding of hammerhead ribozymes.

## Discussion

**Investigations for $Co(NH_3)_6$(III) Binding.** Intermolecular NOEs between $Co(NH_3)_6$(III) ions and the motif surely indicated that the $Co(NH_3)_6$(III) ion binds to the metal ion-binding motif (the A9/G10.1 site) (Figure 3). In addition, the binding isotherms of $^1$H resonances for $Co(NH_3)_6$(III) titrations indicated biphasic (two-step) transitions, which means that two $Co(NH_3)_6$(III) ions were captured by GA10 at the metal ion-binding motif (Figure 2). This phenomenon is the same as in the GA10−Cd(II) system in which the binding isotherms are biphasic, and two Cd(II) ions bind to a GA10 duplex.[5] It should be noted that $Co(NH_3)_6$(III) titration curves for H8(A6) and H2(A6) significantly deviated from a monophasic transition curve and cannot be explained without biphasic transitions, as was the case for GA10−Cd(II) system.[5] It was also found that the overall binding isotherms of $^1$H- and $^{31}$P resonances for Cd(II) and $Co(NH_3)_6$(III) titrations were similar to each other. The resonances for the specific sites (H8(G7) and P(A6)) were especially similar to each other, although degrees of perturbations for the above resonances and titration curves for some resonances were different to some extent (Figure 2).

These data indicate that the binding sites of both $Co(NH_3)_6$-(III) and Cd(II) ions are similar or the same. Unfortunately, we could not completely exclude the possibility that $Co(NH_3)_6$(III) bound to O6(G7) or neighboring sites except for N7(G7). However, it is plausible that one of the binding sites for $Co(NH_3)_6$(III) ions is N7(G7) because of many similarities of the binding isotherms for both $Co(NH_3)_6$(III) and Cd(II) ions at this moment. Therefore, $Co(NH_3)_6$(III) is thought to be able to bind to the motif through the ligand-mediated hydrogen

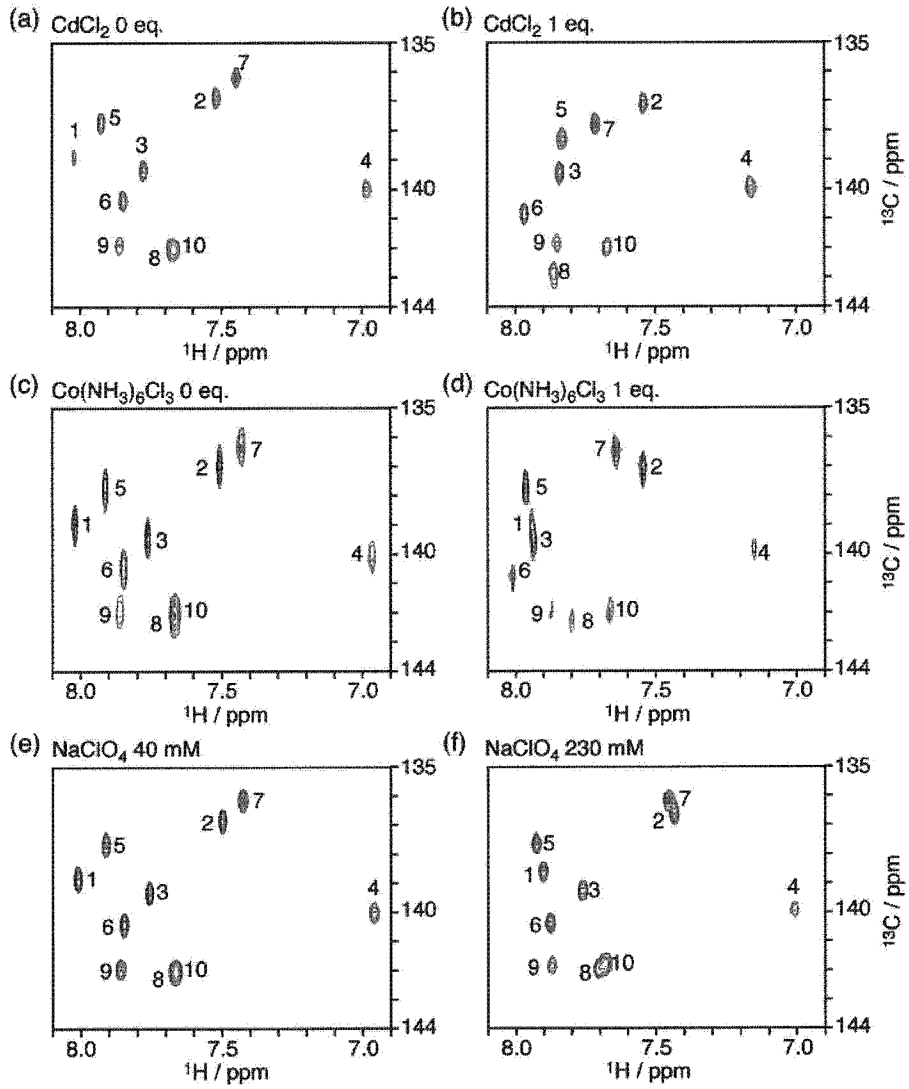



*Figure 4.* 2D $^{1}H-^{13}C$ HSQC spectra with or without metal ions. Panels (a) and (b) indicate spectra of basal solution (2.4 mM GA10(single strand), 40 mM $NaClO_4$) and that in the presence of $CdCl_2$ (2.4 mM GA10(single strand), 40 mM $NaClO_4$, and 2.4 mM $CdCl_2$). Panels (c) and (d) indicate spectra of basal solution (2.0 mM GA10(single strand), 40 mM $NaClO_4$) and that in the presence of $Co(NH_3)_6Cl_3$ (2.0 mM GA10(single strand), 40 mM $NaClO_4$, and 2.0 mM $Co(NH_3)_6Cl_3$). Panels (e) and (f) indicate spectra of basal solution (3.5 mM GA10(single strand), 40 mM $NaClO_4$) and that in the presence of excess $NaClO_4$ (3.5 mM GA10(single strand) and 230 mM $NaClO_4$). Intraresidue cross-peaks between H8 and C8 nuclei are labeled with their residue numbers, and that between H8(G7) and C8(G7) are indicated by red circles. In (a), (b), (e), and (f), spectra were recorded on Bruker DMX750 at 313 K; 1024*128 real points for a spectral width of 7003*5659 Hz. In (c) and (d), the point resolution was reduced to 1024*64 real points for the same spectral width since spectra were broadened upon the addition of 1 molar equivalent of $Co(NH_3)_6Cl_3$.

bonding (outer-sphere binding) (Figures 2−4). On the other hand, crystal structures indicate that several metal ions, such as Mn(II) and Co(II), can bind to the motif via an inner-sphere coordination to N7(G10.1) of hammerhead ribozymes.[1,3] Accordingly, the A9/G10.1 site is a motif which allows two kinds of binding modes, i.e., an inner-sphere coordination[1,3] and outer-sphere binding via metal ligands.

More importantly, the chemical shift of C8(G7) was not affected by the $Co(NH_3)_6$(III) ion binding (Figure 4c,d and Table 1), which means that hydrogen bonding between $NH_3$ ligands and the N7(G7) did not perturb the resonance of C8(G7) or other carbon resonances. In contrast, chemical shift values of both C8(G7) and H8(G7) were selectively perturbed upon the addition of Cd(II) and Mg(II) (Figure 4a,b and Table 1),[4] which can potentially coordinate to N7(G7). Together with the above data, simultaneous large lower field shift of C8 and H8 resonances of G7 cannot be explained without an inner-sphere coordination of a metal ion to N7(G7), although the resonance of C8(G7) for $Co(NH_3)_6$(III) titrations might be unchanged due to some cancellation by chance.

**Reinvestigations of $^{15}N$ NMR Data.** In our previous publication, we observed a higher field shift by 19.6 ppm for the $^{15}N$ resonance of N7(G7) without a detectable *J*-coupling between $^{15}N$ and $^{113}Cd$(II) (Table 2).[5] The same phenomena were observed for the cases of guanosine−$^{199}Hg$(II) complex in dimethyl sulfoxide (DMSO) solution (Table 2).[28] In those studies, it was thought that guanosine and inosine were the ligands for the metal ions, and these nucleosides directly coordinated to the metal ions.[28,29] However, there has been no direct evidence that indicates the formation of the coordination

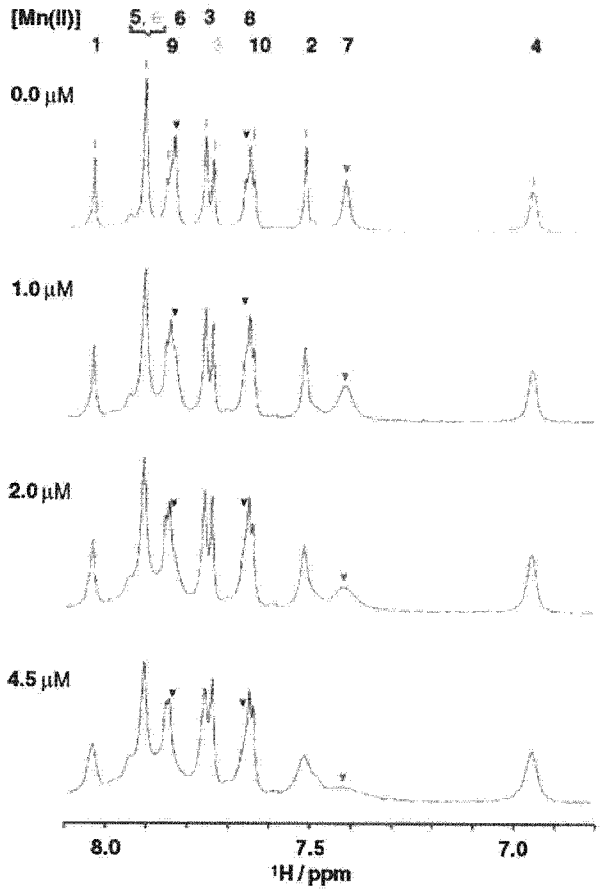



*Figure 5.* 1D $^1$H NMR spectra with increasing $MnCl_2$ concentration. Concentration of $MnCl_2$ for each spectrum is indicated on the left of each spectrum. Signal broadening was observed for H8(G7) (red arrowhead). Signal intensities of H8(A6) and H6(U8) (black arrowheads) seemed to be decreased as well. Assignments of base protons are indicated with their residue numbers, and small orange characters indicate H2 resonances of the corresponding residue.

compound for the above systems. We then tried to detect what kind of complex is formed between Cd(II) and a guanosine monomer in DMSO by means of a specially designed mass spectrometer using an electrospray interface.[34] In this mass spectrometry, we can detect charged clusters that are present beforehand in solution.[34b] For example, even solvated ions such as lithium ion–methanol and cesium ion–methanol clusters were detected with this system.[34b]

In the DMSO solution of Cd(II) and guanosine (the Cd(II)–guanosine–DMSO system), we detected several complexes of three components (Cd(II), guanosine, and DMSO). The major complexes of Cd(II)–guanosine–DMSO were 1:1:3 (*M/Z*: 315.4), 1:1:4 (*M/Z*: 354.4), and 1:1:5 (*M/Z*: 393.4) complexes, which may possibly indicate Cd(II) complexes with the ligands in tetrahedral, square pyramidal, and octahedral geometries, respectively (unpublished data). Although other ternary complexes with different molar ratios were detected, their populations were much less than the 1:1:3, 1:1:4, and 1:1:5 complexes. It was indicated that guanosine directly coordinates to a soft

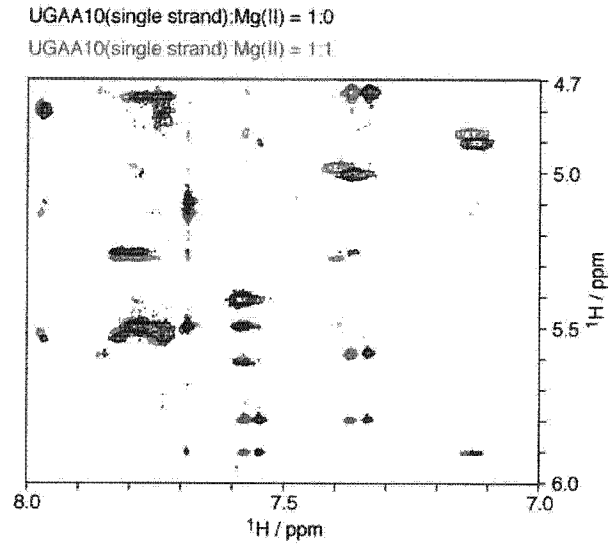



*Figure 6.* 2D $^1$H–$^1$H NOESY spectra of UGAA10. Black spectrum: 2 mM UGAA10(single strand) without $MgCl_2$ (UGAA10(single strand): $MgCl_2$ = 1:0 mol/mol). Red spectrum: 2.0 mM UGAA10(single strand) with 2.0 mM $MgCl_2$ (UGAA10(single strand): $MgCl_2$ = 1:1). Spectra were recorded on Bruker DMX500 at 300 K; 4096*2048 complex points for a spectral width of 4006*4006 Hz.

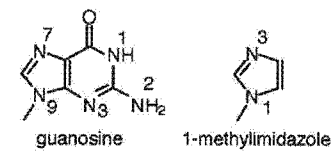



*Figure 7.* Chemical structures of guanine and 1-methylimidazole with numbering systems.

Lewis acid, Cd(II), in DMSO. Accordingly, the higher field shift (approximately 20 ppm) of the N7 resonance of the guanosine–Hg(II) or –Zn(II) in DMSO[28,29] is thought to be due to a direct coordination of guanosine to Hg(II) and Zn(II), since these metal ions belong to soft Lewis acids. In analogy, the higher field shift (19.6 ppm) of the N7(G7) in GA10 (Table 2)[5] indicates that the G7 residue directly coordinates to Cd(II) through N7. The lack of *J*-coupling between $^{113}$Cd(II) and $^{15}$N7(G7) is likely due to the fast exchange of the association–dissociation process of the Cd(II) binding to N7(G7).

On the other hand, in the cases for $\beta$-lactamase–Cd(II)[35] and ethylenediaminetetraacetic acid (EDTA)–Cd(II)[36] systems, *J*-coupling between $^{113}$Cd and $^{15}$N ($^1J_{N-Cd}$) were observed with only a slight chemical shift perturbation (Table 2). Although metalation sites of $\beta$-lactamase are imidazole rings of histidines and several imidazole rings surely coordinated to Cd(II), metalation did not affect $^{15}$N chemical shifts of imidazole rings very much. This is quite different from Cd(II)–GA10 and Hg(II)–guanosine systems (no detectable $^1J_{N-M(II)}$ and large chemical shift perturbations). Unfortunately, we could not conclude the origins of the discrepancies at this moment. However, it was found that there are at least two patterns of appearance for $^{15}$N NMR spectra which indicate an inner-sphere coordination of nitrogen to Cd(II). Therefore, caution must be

(33) Alei, M., Jr.; Morgan, L. O.; Wageman, W. E. *Inorg. Chem.* **1981**, *20*, 940–942.
(34) (a) Wakisaka, A.; Watanabe, Y. *J. Phys. Chem. B* **2002**, *106*, 899–901. (b) Mochizuki, S.; Wakisaka, A. *J. Phys. Chem. A* **2002**, *106*, 5095–5100.
(35) Damblon, C.; Prosperi, C.; Lian, L. Y.; Barsukov, I.; Soto, R. P.; Galleni, M.; Frere, J. M.; Roberts, G. C. K. *J. Am. Chem. Soc.* **1999**, *121*, 11575–11576.
(36) Hagen, R.; Warren, J. P.; Hunter, D. H.; Robarts, J. D. *J. Am. Chem. Soc.* **1973**, *95*, 5712–5716.

taken for the interpretation of $^{15}N$ NMR data when we determine the sort of chemical bonds from $^{15}N$ NMR data. For more definitive results, solid-state NMR with crystal powder of Cd(II)−guanosine complex might provide the definitive answer for the chemical bond formed between Cd(II) and N7 of guanosine, since in the crystal lattice association−dissociation dynamics is suppressed. In such a system, the Cd(II)−N7(G) bond can be determined by crystallography, and intrinsic *J*-coupling between $^{113}Cd(II)-^{15}N7(G)$ can be derived from solid-state NMR, and then $^{15}N$ and $^{13}C$ chemical shifts can be confirmed as well.

**Roles of the Metal Ion at the A9/G10.1 Site.** We have demonstrated that the altered motif in UGAA10, with an adenosine residue instead of a guanosine, was not sufficient to capture divalent cations from 2D $^1H-^1H$ NOESY spectra of UGAA10 (Figure 6). We had shown that nonspecific binding of metal ions was observed in the titration curves of GA10 at "mM" order, as well.[5] Thus, the altered motif is no longer thought to be a metal ion-binding motif. Interestingly, Sigel et al. indicated that N7 of adenosine is a much less efficient acceptor of metal ions than that of guanosine by means of potentiometric titrations at the nucleoside level.[37] They also demonstrated that basicities of nitrogen atoms were well-correlated with affinities of the nitrogen atoms to divalent cations, i.e., the order of affinities of N7 to divalent cations is guanosine > adenosine and the order of the basicities of N7 is guanosine > adenosine.[37] Therefore, we think that this effect is one of the major reasons for the reduced affinity of the altered motif to Mg(II). For more detailed physicochemical characterizations of this altered motif, titration experiments with $Co(NH_3)_6$(III) and Cd(II) are also required.

From the aspect of the roles of metal ions at the A9/G10.1 site, the above titration data of UGAA10 indicate a more important thing. Kinetic data by Ruffner et al. indicated that hammerhead ribozymes with the A10.1−U11.1 base pair (the altered motif) still possess 30% catalytic activities relative to those with the common sequence which includes the G10.1−C11.1 base pair.[7] Furthermore, in the sequence of the wild-type hammerhead ribozyme from the newt, A10.1−U11.1 is included instead of a G10.1−C11.1.[38] These data indicate that G10.1−C11.1 is not a strict requirement for the reaction catalyzed by hammerhead ribozymes, and G10.1−C11.1 can be replaced with A10.1−U11.1 without a loss of functions, although this base pair substitution significantly reduces a metal ion-binding ability of the motif (Figure 6). Therefore, if we assume that the metal ion at the A9/G10.1 site is a catalytic metal, reduced (almost loss of) metal ion-binding ability of UGAA10 is inconsistent with this assumption. Accordingly, it was suggested that the metal ion at the A9/G10.1 site is not a catalytic metal, although there might be a possibility that other conserved sequences might support metal ion binding to the altered motif upon conformational changes. Eckstein's group and Lilley's group indicated that hammerhead ribozymes changed their conformation in a divalent cation concentration-dependent manner.[9] The metal ion at the A9/G10.1 motif might be one of the structural cofactors for the formation of the γ-shape of hammerhead ribozymes.

(37) Kapinos, L. E.; Holy, A.; Günter, J.; Sigel, H. *Inorg. Chem.* **2001**, *40*, 2500−2508.
(38) (a) Forster, A. C.; Symons, R. H. *Cell* **1987**, *50*, 9−16. (b) Pabon-Pena, L. M.; Zhang, Y.; Epstein, L. M. *Mol. Cell Biol.* **1991**, *11*, 6109−6115.

## Experimental Section

**Sample Preparations.** RNA oligomers were chemically synthesized, purified, and quantitated as described previously.[5,39] Synthesized RNA oligomers are the following:

GA10:r(GGACGAGUCC)

UGAA10:r(GGAUGAAUCC)

The basal solutions for NMR measurements contained 40 mM $NaClO_4$ and various concentrations of metal salts ($CdCl_2$, $Co(NH_3)_6Cl_3$, $MnCl_2$, and $NaClO_4$). To avoid precipitation of the metal−buffer complex and the resulting pH changes, no buffer was added to the solution used in the NMR measurements.[4,5,39] Instead, the pH of each solution was adjusted to 6 with the direct titration as described previously.[4,5,39] All titration experiments were performed at 313 K, unless stated in legends to the figures. Other spectral parameters are presented in legends to the figures.

**Fluorescence Spectroscopy.** We measured fluorescence spectroscopy using Tb(III) and Eu(III) as fluorescence probes. It is expected that fluorescence-detected circular dichroism (FDCD) can be observed if the above ions are recognized by the motif via an inner-sphere coordination, since a chirality is generated around a fluorescent metal center in such a case. In the GA10−Tb(III) system, a canonical-sensitized emission was observed (data not shown), and we found that Tb(III) was able to bind to GA10. However, a sensitized emission was not observed for the GA10−Eu(III) system (data not shown). Unfortunately, in such a case, the sensitized emission of Tb(III) might be arisen from a weak dipole−dipole (through-space) interaction,[40] which means that energy transfer can occur even if a nucleobase does not coordinate to Tb(III). Therefore, we could not get critical evidence for an inner-sphere coordination of metal ions to the motif (the A9/G10.1 site) from fluorescence spectroscopy and stopped further experiments with fluorescence spectroscopy, although the sensitized emission from the GA10−Tb(III) complexes is still interesting. Solution conditions for fluorescence spectroscopy were 1 μM GA10, 50 mM PIPES pH 6.5, 50 mM NaCl, and various concentrations (0.6, 2.6, and 4.6 μM) of $TbCl_3$ and $EuCl_3$. Spectra were recorded on a JASCO spectrofluorometer FP-750.

**Acknowledgment.** We thank Dr. Kodama at BERI, Osaka, for the critical discussion of fluorescence spectroscopy. Y.T. was supported by a Grant-in-Aid for Young Scientists (B) (15750136) and by a Grant-in-Aid for Scientific Research on Priority Areas (15025211) from the Ministry of Education, Culture, Sports, Science and Technology, Japan.

JA036826T

(39) Kasai, Y.; Tanaka, Y.; Morita, E. H.; Tanaka, Y.; Taira, K. *Nucleic Acids Res.* **2001**, (Suppl. 1), 81−82.
(40) Horrocks, W. D., Jr. *Methods Enzymol.* **1993**, *226*, 495−538.
(41) (a) Heus, H. A.; Pardi, A. *Science* **1991**, *253*, 191−194. (b) Katahira, M.; Kanagawa, M.; Sato, H.; Uesugi, S.; Fujii, S.; Kohno, T.; Maeda, T. *Nucleic Acids Res.* **1994**, *22*, 2752−2759.
(42) Jia, X.; Zon, G.; Marzilli, L. G. *Inorg. Chem.* **1991**, *30*, 228−239.
(43) Mukundan, S., Jr.; Xu, Y.; Zon, G.; Marzilli, L. G. *J. Am. Chem. Soc.* **1991**, *113*, 3021−3027.
(44) (a) Sherman, S. E.; Gibson, D.; Wang, A. H.-J.; Lippard, S. J. *Science* **1985**, *230*, 412−417. (b) Sherman, S. E.; Lippard, S. J. *Chem. Rev.* **1987**, *87*, 1153−1181. (c) Sundquist, W. I.; Lippard, S. J. *Coord. Chem. Rev.* **1990**, *100*, 293−322.

THE JOURNAL OF BIOLOGICAL CHEMISTRY

Vol. 279, No. 6, Issue of February 6, pp. 4760–4767, 2004
*Printed in U.S.A.*

# Structure of the Ubiquitin-interacting Motif of S5a Bound to the Ubiquitin-like Domain of HR23B*

Received for publication, August 26, 2003, and in revised form, October 6, 2003
Published, JBC Papers in Press, October 29, 2003, DOI 10.1074/jbc.M309448200

**Kenichiro Fujiwara‡, Takeshi Tenno‡§, Kaoru Sugasawa¶, Jun-Goo Jee‡‖, Izuru Ohki**, Chojiro Kojima‡‡, Hidehito Tochio‡, Hidekazu Hiroaki‡, Fumio Hanaoka¶§§, and Masahiro Shirakawa‡‖¶¶**

*From the ‡Graduate School of Integrated Science, Yokohama City University, 1-7-29 Suehiro, Tsurumi, Yokohama, Kanagawa 230-0045, the §Graduate School of Science and Engineering, Ehime University, 3 Bunkyo, Matsuyama, Ehime 790-8577, the ¶Cellular Physiology Laboratory, Discovery Research Institute, RIKEN, 2-1 Hirosawa, Wako, Saitama 351-0198, the ‖Genomic Sciences Center, RIKEN, 1-7-22 Suehiro, Tsurumi, Yokohama, Kanagawa 230-0045, the **Department of Structural Biology, Biomolecular Engineering Research Institute, 6-2-3 Furuedai, Suita, Osaka 565-0874, the ‡‡Graduate School of Biological Sciences, Nara Institute of Science and Technology, 8916-5 Takayama, Ikoma, Nara 630-0101, and the §§Graduate School of Frontier Biosciences, Osaka University, 1-3 Yamada-oka, Suita, Osaka 565-0871, Japan*

**Ubiquitination, a modification in which single or multiple ubiquitin molecules are attached to a protein, serves signaling functions that control several cellular processes. The ubiquitination signal is recognized by downstream effectors, many of which carry a ubiquitin-interacting motif (UIM). Such interactions can be modulated by regulators carrying a ubiquitin-like (UbL) domain, which binds UIM by mimicking ubiquitination. Of them, HR23B regulates the proteasomal targeting of ubiquitinated substrates, DNA repair factors, and other proteins. Here we report the structure of the UIM of the proteasome subunit S5a bound to the UbL domain of HR23B. The UbL domain presents one hydrophobic and two polar contact sites for interaction with UIM. The residues in these contact sites are well conserved in ubiquitin, but ubiquitin also presents a histidine at the interface. The pH-dependent protonation of this residue interferes with the access of ubiquitin to the UIM and the ubiquitin-associated domain (UBA), and its mutation to a smaller residue increases the affinity of ubiquitin for UIM.**

Chains or single molecules of ubiquitin can be attached to cellular proteins, giving rise to poly- or monoubiquitination, respectively, which mediates distinct cellular signals (1–3). Polyubiquitination mainly targets proteins for degradation by the proteasome and can modulate diverse biological processes, such as cell cycle progression, apoptosis, antigen presentation, and stress response; by contrast, (monoubiquitination can act as both a endocytic sorting signal in vesicular transport and a key regulator of transcription, replication, and DNA repair (1–5). The versatility of ubiquitination signals is mediated by different downstream regulatory factors, many of which carry a UIM,[1] a ubiquitin-associated domain (UBA), or both (6–8). UIM was first identified in the proteasome subunit S5a and occurs in many proteins involved in rapid protein degradation or protein trafficking, such as endocytosis (Fig. 1*b*) (7).

So far, 27 UIM-containing proteins have been identified in humans (9). The S5a UIM requires a string of at least four $Lys^{48}$-linked ubiquitins for efficient binding (10). S5a of higher eukaryotes has two UIMs, located near each other in the C terminus. The C-terminal UIM by itself binds polyubiquitin chains as efficiently as the full-length S5a protein, whereas the N-terminal UIM alone binds polyubiquitin chains less efficiently (10). Nevertheless, a sequence alignment by Young *et al.* (10) revealed that S5a of *Saccharomyces cerevisiae* (Rpn10p) has only one UIM sequence that corresponds to the N-terminal UIM of higher eukaryotes (Fig. 1*b*). Recently, it has been shown that *S. cerevisiae* S5a contributes to the targeting of a subset of ubiquitinated substrates to the proteasome, but it is not the sole factor involved in polyubiquitin recognition (11, 12). Subunit S6′ of mammalian proteasome was also shown to recognize the polyubiquitin degradation signal (13). In contrast to S5a UIMs, those of the endocytic factors Hrs, Eps15, Eps15R, and Vps27 are proposed to bind monoubiquitin tags (8, 14, 15). The affinity between these UIMs and monoubiquitin seems low; the dissociation constant ($K_d$) for Hrs binding to monoubiquitin is 230–300 $\mu$M (16, 17).

Interactions of ubiquitin tags with UIM or UBA of downstream effectors are modulated by proteins that bear UbL domains, some of which are thought to bind UIM by mimicking ubiquitination. Similar to ubiquitin-like modifier proteins such as NEDD8, UbL domains share high sequence homology with ubiquitin (typically 24–33%) and thus adopt ubiquitin folds, as shown by the structure determination of the UbLs of PLIC-2, Parkin, and HR23B (18–20, 30).

Of the UbL-containing proteins, HR23B is the best characterized and has been shown to target the excision repair factor XPC/Rad4 and the endoplasmic reticulum-associated deglycosylation enzyme Png1p to the proteasome through interaction of its UbL with the proteasome (21–26). More recently, Rad23 (the *S. cerevisiae* homolog of HR23B) was proposed to deliver a variety of ubiquitinated cellular proteins to proteasome (27).

* This work was supported by grants from the Japanese Ministry of Education, Science, Sports and Culture (to M. S. and H. T.). The costs of publication of this article were defrayed in part by the payment of page charges. This article must therefore be hereby marked "*advertisement*" in accordance with 18 U.S.C. Section 1734 solely to indicate this fact.

*The atomic coordinates and structure factors (code 1UEL) have been deposited in the Protein Data Bank, Research Collaboratory for Structural Bioinformatics, Rutgers University, New Brunswick, NJ (http://www.rcsb.org/).*

¶¶ To whom correspondence should be addressed. Tel.: 81-45-508-7213; Fax: 81-45-508-7361; E-mail: shirakawa@tsurumi.yokohama-cu.ac.jp.

[1] The abbreviations used are: UIM, ubiquitin-interacting motif; UbL, ubiquitin like; HR23B, human homolog of RAD23 B; UBA, ubiquitin-associated domain; NOE, nuclear Overhauser effect; NOESY, NOE spectroscopy; GST, glutathione *S*-transferase.

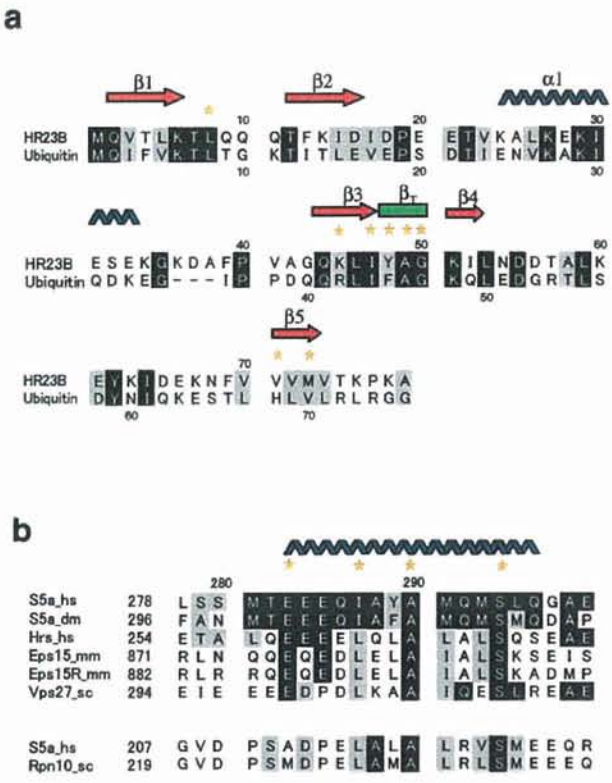



FIG. 1. **Sequence alignment of UbL and UIM.** *a*, sequences of human HR23B UbL and human ubiquitin. *b*, sequence alignment of the C-terminal UIM of human proteasome subunit S5a with UIMs of endocytic factors that have been either shown or implied to bind a ubiquitin tag (*top panel*) and of the N-terminal UIMs of human and *S. cerevisiae* S5a (Rpn10p) (*bottom panel*). The sequences are from human (*hs*), mouse (*mm*), *Drosophila melanogaster* (*dm*), and *S. cerevisiae* (*sc*). In *a* and *b*, identical residues are highlighted in *black*, and homologous residues are highlighted in *gray*. Secondary structural elements of HR23B UbL and S5a UIM are indicated. The residues shown to be important for the complex formation through structural or mutational analyses are marked with *asterisks*.

The interaction between the proteasome and the UbL domain of HR23B is essential for all of these functions. In higher eukaryotes, the C-terminal UIM of S5a has been proposed to be the receptor site for the UbL domain, to which it binds efficiently and with high affinity ($K_d$ = 3.4 μM; this study, see "Results" and "Discussion"); by contrast, the N-terminal UIM does not bind to UbL (22). In yeast, however, S5a does not have a sequence corresponding to this C-terminal UIM (10) and does not bind to Rad23 (11). Instead, the proteasome subunit Rpn1 serves as its receptor. These observations may suggest that the HR23B receptor sites in the proteasome differ between yeast and higher eukaryotes.

The interaction between ubiquitin/UbL and UIM has been studied by mutagenesis and NMR chemical shift perturbation experiments (10, 16, 18, 19, 28–30). These experiments all suggest that the conserved hydrophobic patch composed of the side chains of $Leu^8$, $Ile^{44}$, and $Val^{70}$ of ubiquitin serves as the binding site for S5a UIM or the proteasome.

To understand the structural basis for this recognition, we have solved the structure of the C-terminal UIM of human proteasome subunit S5a in complex with the UbL domain of human HR23B by solution NMR. A comparison of the UIM binding surface of UbL with the corresponding region of ubiquitin, coupled with extensive mutagenesis studies, indicates that the residues forming these interfaces are well conserved between UbL and ubiquitin, except for $His^{68}$ of ubiquitin. This residue can regulate the access of the UIM or UBA to ubiquitin in a pH-dependent manner. The present structural and mutational data enable us to revise the consensus sequence of UIM.

## EXPERIMENTAL PROCEDURES

*Preparation of Proteins and Peptides*—We expressed the N-terminal 87 residues of human HR23B, referred to as UbL, and residues 263–307 of S5a, referred to as UIM, as His-tagged proteins in *Escherichia coli* BL21(DE3). Labeled proteins were obtained by growing *E. coli* in synthetic media containing $^{15}NH_4Cl$ and $^{15}NH_4Cl/^{13}C$-glucose, respectively, and were purified chromatographically. UbL has an extra HHHHHH sequence at its C terminus, whereas UIM has an extra GSH sequence at its N terminus after cleavage by thrombin. The UIM-UbL complex was formed by titrating UIM into UbL, using changes in the amide group resonances of UbL in the $^{15}N,^1H$ HSQC (heteronuclear single quantum correlation) spectrum to indicate 1:1 stoichiometry. The UIM-UbL complex was in slow exchange on the chemical shift timescale. Samples for NMR measurements typically comprised 1.0 mM UIM-UbL complex in 20 mM potassium phosphate buffer (pH 6.8) and 5 mM KCl. Various combination of isotopically labeled UIM-UbL complex were used to obtain chemical shift assignments and distance restraints.

*NMR Spectroscopy*—NMR spectra were acquired at 303 K using a Bruker DRX500 or DRX800 spectrometer. For assigning the $^1H$, $^{15}N$, and $^{13}C$ resonances, we carried out a series of three-dimensional triple-resonance experiments (31). The stereospecific assignment of the methyl groups of the leucines and valines was achieved using 15% fractionally $^{13}C$-labeled protein. Distance restraints were derived from $^{15}N$ and $^{13}C$ resolved three- and four-dimensional NOESY experiments with a mixing time of 100 ms. Intermolecular restraints were derived from the four-dimensional $^{13}C,^{15}N$-edited NOESY spectrum of a complex formed from $^{13}C$-labeled UbL and $^{15}N$-labeled UIM with a mixing time of 150 ms and three-dimensional $^{13}C$-filtered-$^{13}C$-edited NOESY spectra of a complex formed from one unlabeled component and one $^{13}C,^{15}N$-labeled component with a mixing time of 100 ms. For torsion angle ($\phi$, $\psi$) restraints, the backbone vicinal coupling constants ($^3J_{HN,H\alpha}$) were determined by means of HNHA experiment (31) and a data base search procedure based on backbone chemical shifts, using the program TALOS (32). pH titration of the unique histidine of ubiquitin, $His^{68}$, was carried out by measuring one-dimensional heteronuclear multiple-bond $^1H,^{15}N$ correlation spectra of $^{15}N$-labeled human ubiquitin at 303 K.

*Structure Calculations*—For structure determination of the UIM-UbL complex, 100 structures were initially calculated by a simulated annealing procedure in CNS (33) and further refined by AMBER 7 (34). All interproton distance restraints were derived from unambiguously assigned NOE cross-peaks. The upper limits of NOE restraints were calibrated according to their intensities using ARIA protocols. For the final steps of the calculations, the restraints were included for 25 slowly exchanging backbone amides (2.8–3.4 Å (N–O) and 1.8–2.4 Å (H–O)). In total, 1,630 meaningful NOE restraints (639 intraresidual, 372 sequential, 211 medium range, 305 intramolecular long range, and 103 intermolecular) and 50 dihedral angle restraints were used.

Refinement using AMBER 7 consisted of 20-ps molecular dynamics, followed by 1000 steps of energy minimization. To approximate solvent interaction, a generalized Born model was used (35). Force constants were 20 kcal $mol^{-1}$ $Å^{-1}$ for distance restraints and 150 kcal $mol^{-1}$ $rad^{-2}$ for dihedral angle restraints. Of refined 100 structures, the best 20 structures were selected and analyzed using MOLMOL (36), AQUA, and PROCHECK-NMR (37) software. No NOE was violated by more than 0.3 Å, and no torsion restraint was violated by more than 4°. The hydrogen bond between UIM $Ser^{294}$ and UbL $Gly^{50}$ was assumed on the basis of both mutagenesis of UIM $Ser^{294}$ and observation of the bond in 98 of the 100 calculated structures.

*Mutational Analyses*—Mutant constructs were prepared with the GeneEditor *in vitro* site-directed mutagenesis system (Promega). The binding of S5a UIM to ubiquitin, HR23B UbL, and their mutants was determined by surface plasmon resonance using a BIACORE-X instrument. The equilibrium binding affinity of GST-UIM or its mutant immobilized on a CM5 chip was analyzed by monitoring the change in response units as a function of ubiquitin, UbL, or their mutant concentrations ranging from 0.1 to 500 μM at a flow rate of 20 μl/min in buffer containing 10 mM HEPES, pH 7.4, 150 mM NaCl, 3 mM EDTA, and 0.005% Surfactant P20. The binding of the UBA of yeast Dsk2p to ubiquitin was determined in the same way using a BIACORE-X instrument. A GST fusion protein that contains Dsk2p residues 328–373 (38) was immobilized onto a CM5 chip.

The effect of mutations on the affinity of S5a UIM for tetraubiquitin was examined by a surface plasmon resonance-based competition as-

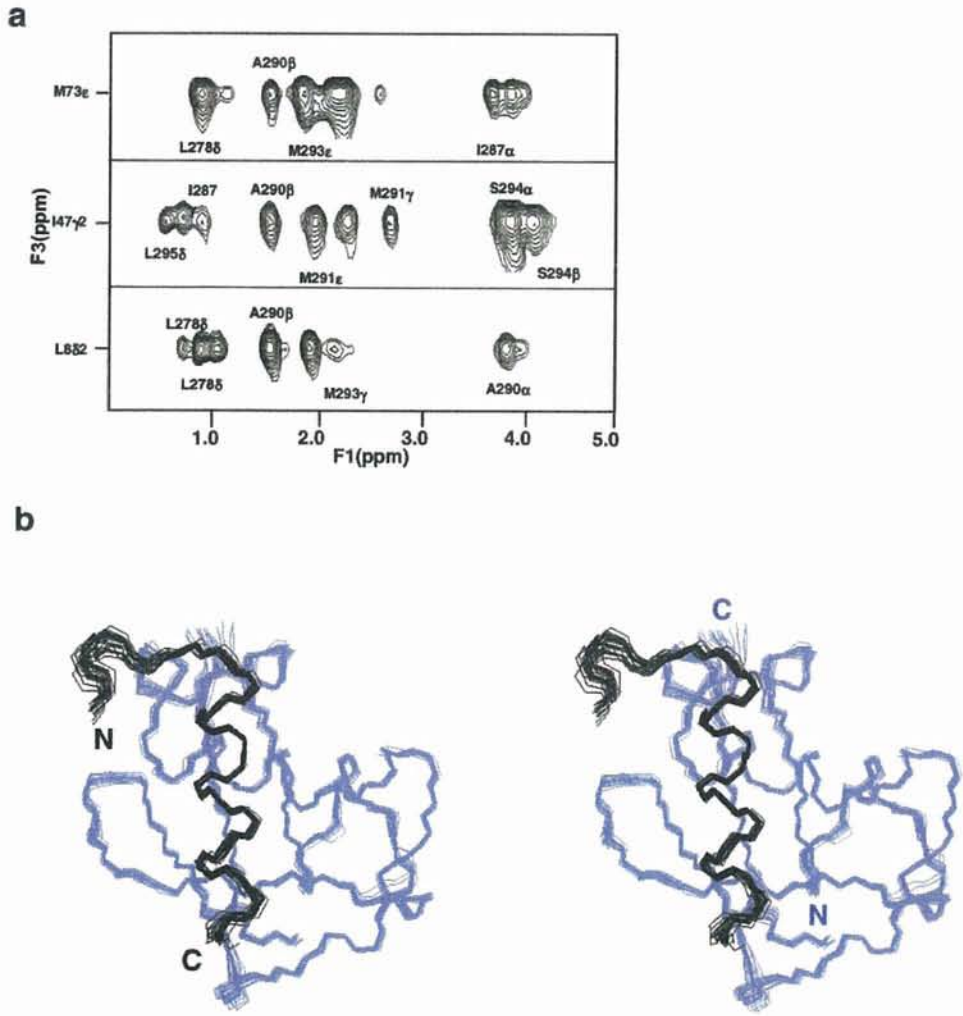



FIG. 2. **Structure determination of the UIM-UbL complex.** *a*, selected region of three-dimensional $^{13}C$-filtered-$^{13}C$-edited NOESY spectra of a complex formed from $^{13}C$,$^{15}N$-labeled UbL and unlabeled UIM with a mixing time of 100 ms, depicting intermolecular NOEs. *b*, stereo view of the best fit superposition of the 20 final structures of the complex between UIM (*black*) and UbL (*blue*). Residues 263–277 and 297–307 of UIM are omitted from all figures for clarity.

say. The resonance curves were measured for 0.1 μM tetraubiquitin binding to immobilized GST-UIM in the presence of 40 μM His-tagged wild type UIM or its point mutants. A buffer containing 10 mM HEPES, pH 7.4, 150 mM NaCl, 3 mM EDTA, and 0.005% Surfactant P20 was used at flow rate of 20 μl/min. The effect of ubiquitin mutation of $His^{68}$ to valine (H68V) for its binding affinity to the UIM of S5a was also analyzed by GST pull-down assay. 12 μg of GST-UIM was incubated with 3 μg of wild type or mutant ubiquitin in 100 μl of 20 mM HEPES (pH 7.9), 100 mM NaCl, 20% glycerol, 0.1% Nonidet P-40, and 200 μg/ml bovine serum albumin. The bound proteins were analyzed by GSH-mediated pull-down assay coupled with SDS-polyacrylamide gel electrophoresis and silver staining.

## RESULTS

*Structure Determination*—The structure of the UIM of human proteasome subunit S5a in complex with the UbL domain of human HR23B was determined from a total of 1,730 NMR-derived restraints (Fig. 2*a* and Table I). The structure of residues 1–75 of UbL and that of the UIM residues 278–296 of S5a are well defined (Fig. 2*b*). By contrast, the terminal residues of UIM have no contact with UbL and are therefore disordered. In the complex, UbL folds into an $\alpha/\beta$ structure comprising a layer of five-stranded twisted β-sheet, backed by a long helix (Fig. 3*a*). It closely resembles the structures of ubiquitin (Ref. 39; Protein Data Bank code 1d3z), as indicated by the root mean square deviation of 0.741 Å over 30 Cα coordinates for residues in the regions of secondary structure elements. The UIM-bound UbL of HR23B has a structure similar to that of the unliganded form (30). Therefore, the UbL domain of human HR23B shares the same fold as members of the ubiquitin-like modifier family and other ubiquitin-like domains.

UIM adopts a hook-like conformation, consisting of an N-terminal loop (residues 278–282) followed by an α helix (residues 283–296), which are flanked on either side by N- and C-terminal unstructured regions (residues 263–277 and 297–307). The helix dominates the binding interface, fitting snugly along strands β3 and β5 of UbL, whereas the N-terminal loop runs alongside the loop that connects helix α1 and strand β3 of UbL, mediating minor interfacial contacts. The helical portion of the UbL-bound UIM of S5a, residues 284–296, adopts a nearly identical fold to the corresponding part of a free UIM of yeast Vps27p (residues 303–315), whose crystal structure was solved as an antiparallel four-helix bundle consisting of four molecules (40).

*Recognition of UIM by UbL*—UbL has three principle contact sites with the helix of UIM that mediate a central hydrophobic and two polar interactions. The hydrophobic contact site is defined by the outward facing residues $Leu^8$, $Ile^{47}$, $Val^{71}$, and