博士論文番号：0381028

# Novel NMR methods developed for the structural analysis of biological macromolecules

野村　誠
奈良先端科学技術大学院大学
バイオサイエンス研究科　生体高分子構造学講座
（児嶋　長次郎　助教授）

平成１８年　１月３０日提出

# General Introduction

In the field of structural biology, the nuclear magnetic resonance (NMR) is one of the most powerful tools to obtain the 3D structure at an atomic resolution. NMR is categorized as a spectroscopic method and two properties, an energy level and a coherence transfer, are essential. The chemical shifts represent energy levels that are used for identification of each atom in a target molecule. With recent progress in the $^{15}N$ and $^{13}C$ stable isotope labeling techniques, the resonance assignments of the larger molecules can be acquired. The J-couplings transfer the coherences through covalent bonds. This coherence transfer is used for reconstructing the chemical structure from the assigned resonances. However the conformations of the biological polymers are characterized by a large amount of non-covalent interactions such as hydrophobic interactions and hydrogen bonds. In the viewpoint of the determination of the 3D structures of the biological polymers, the energy transfer through the space is the most important factor. The nuclear Overhauser effect (NOE) can transfer the coherences through the space up to 5 Å. At the initial stage of the structure determination, the random coil structure is generated (Figure 1 (left)). After adding NOE distance constraints (Figure 1 (right), lines colored in red), the protein formed specific conformation. It means the 3D structure is determined. 5 Å is

shorter than the diameter of proteins, 30 Å and more. Thus NOE information is not suitable for the quick determination of grovel fold of globular proteins and bending of rod-like molecules such as DNA.

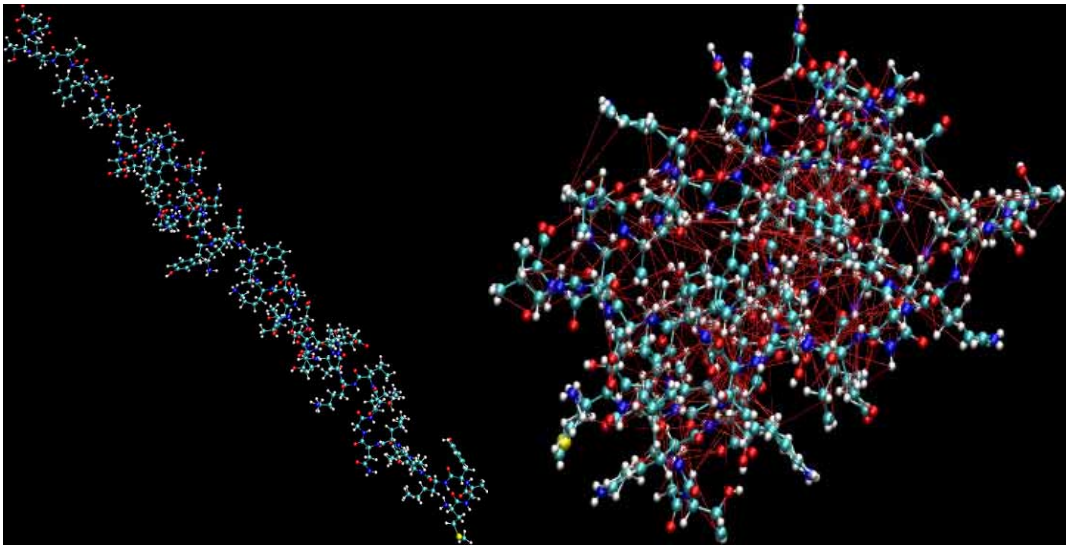


Figure 1. The random coil structure (left) and the determined NMR structure (right). The red lines indicate the NOE distance constraints up to 5 Å.

In my first study (I. Paramagnetic metal system, page 10), I have developed a novel method for obtaining long-range distance information. The paramagnetic metal, $Cu^{2+}$, has lone-pair electron in d-orbital. This electron spin causes the relaxations of the nuclear spins. This relaxation effect is detected as the line broadening of the NMR signal (Figure 2).

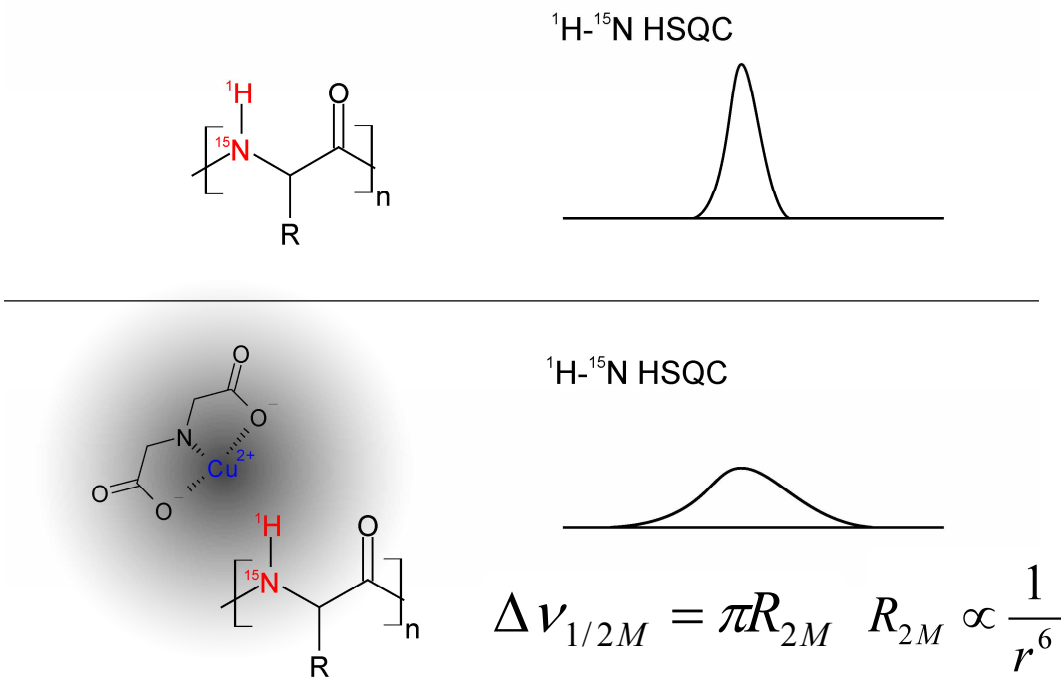



Figure 2. The paramagnetic relaxation effect (PRE). The top figure represents the $^1$H-$^{15}$N HSQC signal without the paramagnetic metal. In the presence of the paramagnetic metal, the $^1$H-$^{15}$N HSQC signal becomes broader depending on the distance between the electron of the paramagnetic metal and $^1$H nucleus. $\Delta\upsilon_{1/2M}$ and $R_{2M}$ are the half-height line width and the transverse relaxation rate induced by the paramagnetic metal, respectively. r is distance between the electron of the paramagnetic metal and $^1$H nucleus.

The relaxation rate (or line broadening) induced by the paramagnetic metal depends on the distance between the electron and the $^1$H nucleus. This paramagnetic relaxation effect (PRE) has longer distance information up to 20 Å than NOE (up to 5 Å). The novel method is consisted of the localization technique of the paramagnetic metal and the estimation of the relaxation rates.

In my second study (II. Drug DNA system, page 33), the solution structure of the drug-DNA complex was examined. The determined 3D structure indicated that two DNA bases were flipped out. However it was difficult to study the nature of the flipping out bases using the conventional NOE data (Figure 3).

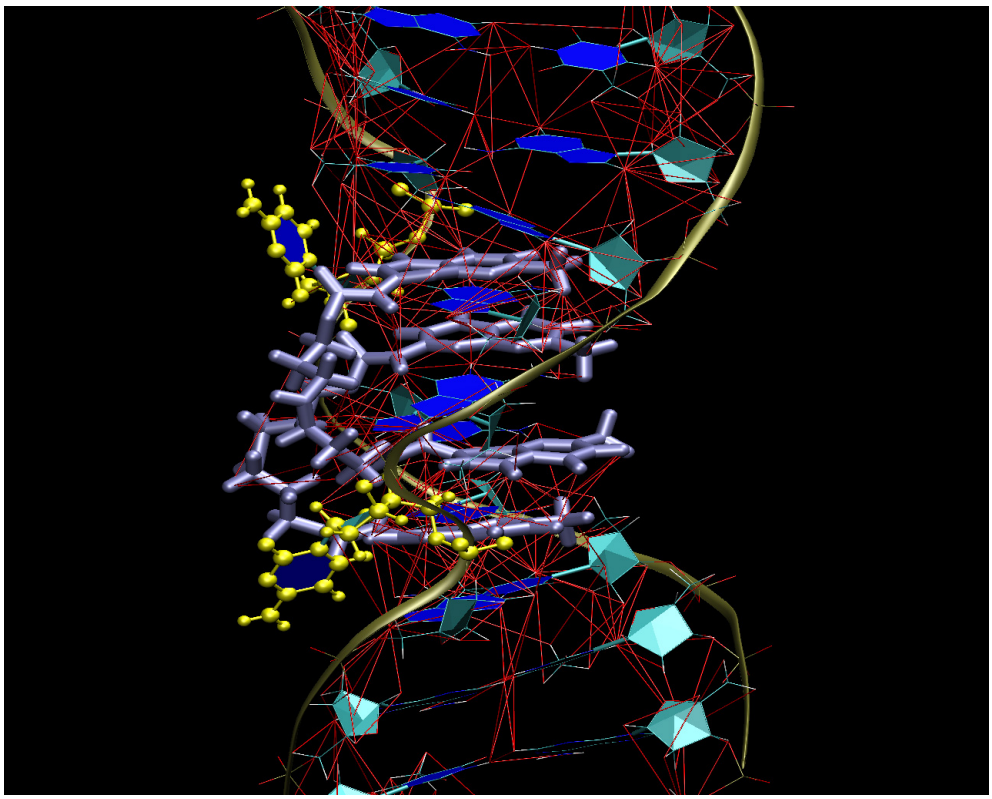
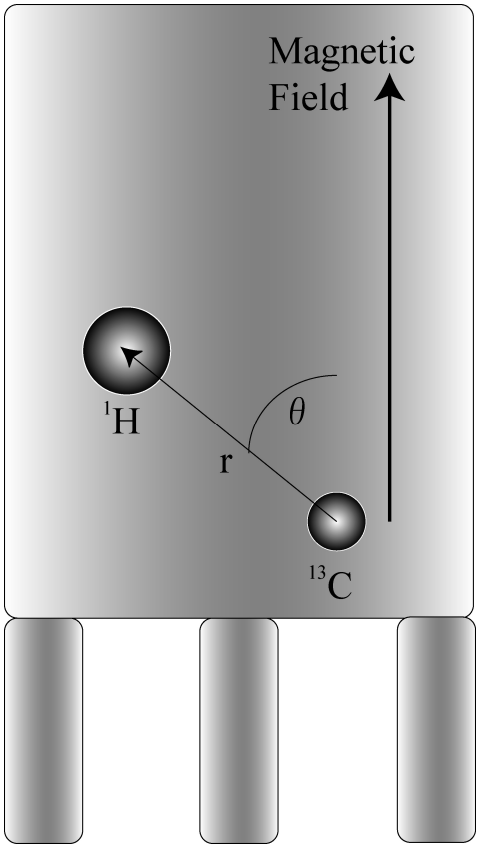


Figure 3. The NMR structure of the drug-DNA complex. Two gray molecules are drugs. Two bases colored in yellow represents the flipping out cytosine bases. The NOE distance information (up to 5 Å) used for the structure determination is represented as the red line. There is no NOE constraint around two flipping out bases.

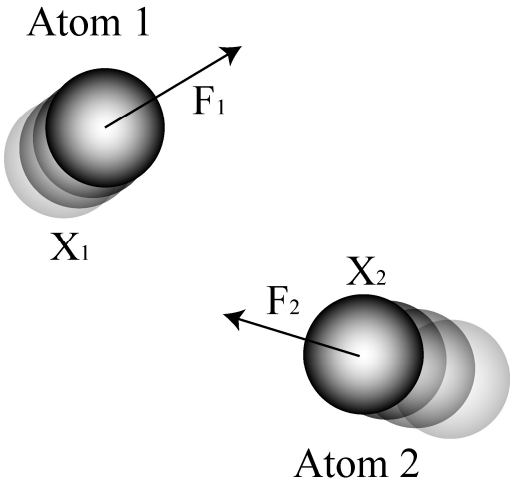
The long-range distance information, the residual dipolar coupling (**RDC**), was used for studying these flipping out bases. This **RDC** information provides the angles between

the $^{1}$H-$^{1}$H or $^{1}$H-$^{13}$C vector and the magnetic field (Figure 4). For studying the dynamics nature and the interaction site of these flipping out bases, molecular dynamics (**MD**) simulation was also carried out (Figure 5).




$$RDC \propto 3\cos^2\theta - 1$$

Figure 4. The concept of the residual dipolar coupling (**RDC**). The **RDC** gives the angle between the NMR magnetic field and the vector between two nuclei (ex. $^{1}$H and $^{13}$C) that is a kind of the long-range distance information.




$$m\frac{d^2x_1}{dt^2} = F_1$$

$$m\frac{d^2x_2}{dt^2} = F_2$$

$$F_1 = -\frac{\partial V(x_1, x_2)}{\partial x_1}$$

$$F_2 = -\frac{\partial V(x_1, x_2)}{\partial x_2}$$

Figure 5. The concept of the molecular dynamics (**MD**) simulation. This simulation is based on solving the Newton equations of all the atoms in the system. The force field (represented as $F_1$ and $F_2$) is the key factor for the realistic simulation.

Here I developed the novel method for the protein to obtain the long-range distance information using the paramagnetic metal. I also applied the **RDC** long-range distance method for studying the flipping out DNA bases of drug-DNA complex.

## Study 1: paramagnetic metal system

### Introduction

In the field of biological NMR, long-range distance constraints are required for large-size multi-domain proteins and high-throughput structure determinations. Paramagnetic metals have useful properties, such as paramagnetic relaxation enhancement and pseudo-contact shift, both of which affect the NMR signals over long-range distances. For metal proteins and metal–drug–DNA complexes, these paramagnetic effects have been successfully utilized as long-range constraints in structure determinations [1–7].

Kay and co-workers [8,9] developed a method to evaluate the distance between the backbone amide proton and $Cu^{2+}$ using the proton transverse relaxation rate enhanced by the paramagnetic effect for non-metal proteins. The method relies on localizing the paramagnetic ion to a modified target protein such as possessing a specially designed ATCUN motif tag [8,9]. Dvoretsky et al. [10] used the amide proton longitudinal relaxation rates for $Mn^{2+}$ that was localized by mutating an amino acid of the target protein to cysteine followed by covalent modification of this residue using thiol-reactive EDTA. Both methods rely on mutation of the target protein to localize the paramagnetic metal, a procedure that may result in structural changes of the target protein. $Cu^{2+}$-iminodiacetic acid ($Cu^{2+}$–IDA), a metal chelate complex, is an established reagent used in the area of

immobilized metal ion affinity chromatography (IMAC) [11]. Immobilized $Cu^{2+}$–IDA on a solid support specifically binds to histidine residues (Kd = $10^{-5}$–$10^{-4}$ M) [11–14], and has been used to purify various protein and DNA molecules [11,15]. Other applications involving the use of $Cu^{2+}$–IDA IMAC have dealt with investigations concerning the structure–function relationship of proteins, which have functional histidine residues [16,17]. Although there are many applications, the precise localization of $Cu^{2+}$–IDA on the protein surface has not been well characterized [11,12].

Here, the localization of $Cu^{2+}$–IDA on three different protein surfaces was examined using chemical shift changes, line broadenings and $^{1}H$ transverse relaxation rate enhancements induced by the paramagnetic effects of $Cu^{2+}$. The proteins examined were ubiquitin, which has one histidine residue (His68) on the surface, and three other proteins that contain a C-terminal poly histidine tag, the ubiquitin-like domain of human HR23B (HR23B UbL), the DNA binding domain of human interferon regulatory factor 4 (IRF4 DBD), and the seven trans-membrane retinal protein, *pharaonis* halorhodopsin (*p*hR). The potential acquisition of long-range distance constraints was also examined.

## Materials and methods

2.1. Sample preparation

Recombinant human ubiquitin was expressed in Escherichia coli BL21(DE3) as a GST-fusion protein. Following the addition of isopropyl β-D-thiogalactoside, protein expression was induced for 8 h at 30 ˚C in minimal medium containing $^{15}$N-labeled ammonium chloride. Cells were suspended in 50 mM Tris–HCl (pH 8.0), 100 mM KCl and 1 mM Pefabloc SC (Roche). The suspension was lysed by sonication, ultra-centrifuged, and the supernatant was loaded onto a glutathione–Sepharose 4B column (Amersham). Elution was performed with 50 mM Tris–HCl (pH 8.0), 100 mM KCl and 30 mM glutathione. The GST–ubiquitin fraction was loaded onto a Superdex 26/60 75 pg (Amersham) gel filtration column equilibrated with 50 mM phosphate (pH 6.5), 50 mM KCl and 1 mM Pefabloc SC. The eluted GST–ubiquitin fraction was then cleaved with PreScission protease (Amersham) and loaded onto a gel filtration column. The final NMR buffer consisted of 50 mM phosphate (pH 6.7) and 200 mM KCl. The $^{1}$H-$^{15}$N ubiquitin sample and its resonance assignments were gifted by Toshitatsu Kobayashi. The expression plasmid of the GST-ubiquitin was gifted by Dr. Toshiyuki Kohno.

The expression and purification of HR23B UbL and IRF4 DBD was performed as previously described [18]. The expression plasmid of HR23B UbL and its resonance

assignments were gifted by Drs. Kenichiro Fujiwara, Takeshi Tenno and Masahiro Shirakawa. The expression plasmid of IRF4 DBD was gifted by Drs. Kazuo Yamamoto and Toshifumi Matsuyama. The $^{1}H$-$^{15}N$ IRF4 DBD sample was gifted in part by Itsuko Ishizaki. The $^{1}H$-$^{15}N$ *p*hR sample and its resonance assignments were gifted by Hideyasu Okuda.

2.2. NMR spectroscopy

The stock solution of $Cu^{2+}$–IDA consisted of 12.5 mM $CuCl_2$, 13.8 mM IDA, 50 mM phosphate and 200 mM KCl (pH 2.2). Following the addition of this stock solution to 0.1 or 1 mM ubiquitin, 0.1 mM HR23B UbL, and 0.5 mM IRF4 DBD, the pH of the sample solution was adjusted to pH 6.7. The ratio of $Cu^{2+}$–IDA/protein was 0.4 for each unless otherwise noted.

All NMR data were obtained on a Bruker DRX800 spectrometer operating at 800 MHz with a triple-resonance probe head equipped with a XYZ-gradient unit. The $^{1}H$–$^{15}N$ HSQC spectra were collected at 293 K, with 2048 (t2) times 150 (t1) complex points and spectral widths of 14 kHz ($^{1}H$) and 1.6 kHz ($^{15}N$) with carrier positions at 4.7 and 121.8 ppm, respectively. A $\pi/2$ shifted sine-bell window function was applied before zero-filling and Fourier transformation. The processing was performed using the NMRPipe package [19], while chemical shifts, linewidths and peak volumes were measured using Sparky software [20]. The transverse relaxation rates were obtained using a $^{1}H$–$^{15}N$ HSQC based spin echo

experiment with a shaped refocusing pulse which was successfully used to obtain long-range distance information with the ATCUN $Cu^{2+}$ system [8]. The relaxation delay times were 8.5, 11, 13.5, 16, 18.5, 21, 23.5, 28.5, 31, 33.5, 38.5, and 48.5 ms for $Cu^{2+}$–IDA–ubiquitin, and 8.5, 11, 13.5, 16, 18.5, 21, 23.5, 28.5, 33.5, 38.5, 48.5, and 68.5 ms for free ubiquitin.

2.3. Data analyses

Linewidth and chemical shift change. Linewidths were measured for 1 mM ubiquitin samples with and without 0.4 mM $Cu^{2+}$–IDA. The differences in linewidth were then calculated. When the $Cu^{2+}$–IDA–ubiquitin peak was not found compared to free ubiquitin (S/N < 5), the difference in linewidth was set to the maximum value of 50 Hz. Chemical shift differences of samples with and without $Cu^{2+}$–IDA were calculated and averaged for three independent experiments. The differences in linewidth and chemical shift values were mapped to the PDB (1UBI).

Paramagnetic $^{1}H$ relaxation enhancements. Three independent relaxation rate measurements of the 1 mM ubiquitin sample complexed with 0.4 mM $Cu^{2+}$–IDA were performed. The differences between the transverse relaxation rates with and without $Cu^{2+}$–IDA were calculated and averaged. The mean values of three experiments were fitted to the relation $R_{2M} = a/r^6 + b$ by non-linear fitting using Mathematica 4.2 (Wolfram) with

error bar dependent weights, where $R_{2M}$ is the transverse relaxation rate enhanced by the paramagnetic effects, r is the distance between the target proton and the relaxation center, and a and b are constants. The amide proton position was obtained from the PDB (1UBI) for the residue with the generalized order parameter $S^2 > 0.7$ [21] and the relaxation center position (assumed as $Cu^{2+}$ atom position) was optimized.

## Results

3.1 Localization and interaction of $Cu^{2+}$–IDA with ubiquitin

$^{1}H$–$^{15}N$ HSQC experiments were performed in an effort to evaluate the interaction of $Cu^{2+}$–IDA with ubiquitin. Titration experiments using $Cu^{2+}$–IDA with 0.1 and 1 mM ubiquitin solutions resulted in dose-dependent and site-specific line broadenings and shifts. The localization of $Cu^{2+}$–IDA on the protein was not due to the presence of ionic interactions, but due to some type of coordination to the protein [16,22], since the addition of 200 mM KCl had no effect. The linewidth differences between $Cu^{2+}$–IDA–ubiquitin and free ubiquitin were mapped to the 3D structure of ubiquitin as shown in Fig. 1 (left), and strongly suggest that $Cu^{2+}$–IDA bound to ubiquitin around His68. The chemical shift differences up to 30 ppb (25 Hz) were also mapped (Fig. 1, right), supporting the presence of a specific binding site close to His68. At pH 5.0, these line broadenings and shifts were

not observed even though the $Cu^{2+}$–IDA is stable at this pH [16,22]. This result is consistent with the notion that the imidazole nitrogen of histidine (pKa = 6.04) acts as the potential electron donor for $Cu^{2+}$–IDA [16].

$Cu^{2+}$ has four coordination sites. IDA can coordinate to three of these, while the protein coordinates to the remaining site [23]. Other metals such as $Fe^{3+}$, $Co^{2+}$, and $Ni^{2+}$ can have four coordination sites, which could be chelated by IDA. $^{1}H$–$^{15}N$ HSQC spectra for these paramagnetic metal–IDA–ubiquitin complexes showed no significant line broadenings and chemical shift differences for $^{1}H$, however small chemical shift differences (less than one-tenth of those for $Cu^{2+}$) were observed for $^{15}N$ around His68. This result is consistent with the notion that the localization of the paramagnetic metal–IDA complex is around His68.

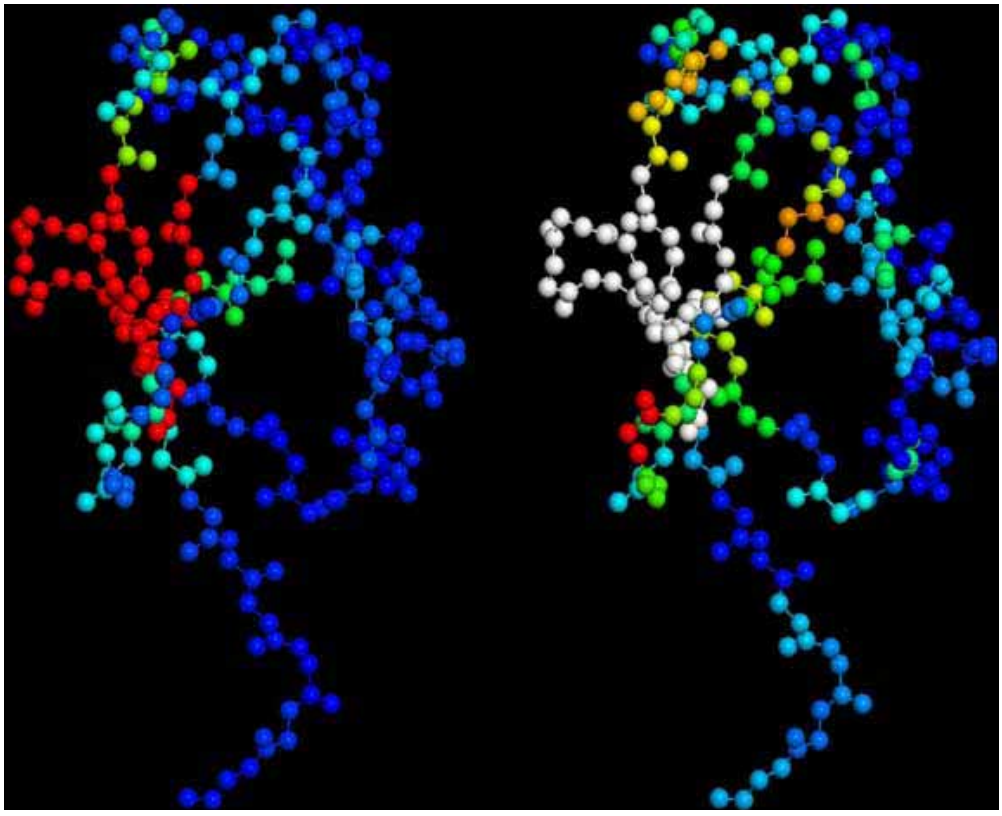


Fig. 1. Line broadenings (left) and peak shifts (right) induced by the $Cu^{2+}$–IDA complex mapped on the ubiquitin structure. With increasing broadenings and shifts, the color changes from blue to red. Maximum, medium and minimum values (50, 25, and 0 Hz for broadenings, and 30, 15, and 0 ppb for shifts) are colored in red, green and blue, respectively. Intermediate colors such as yellow (between green and red) are given gradually. In the peak shifts (right), white represents disappeared peaks. These pictures were drawn by RASMOL 2.6 [29].

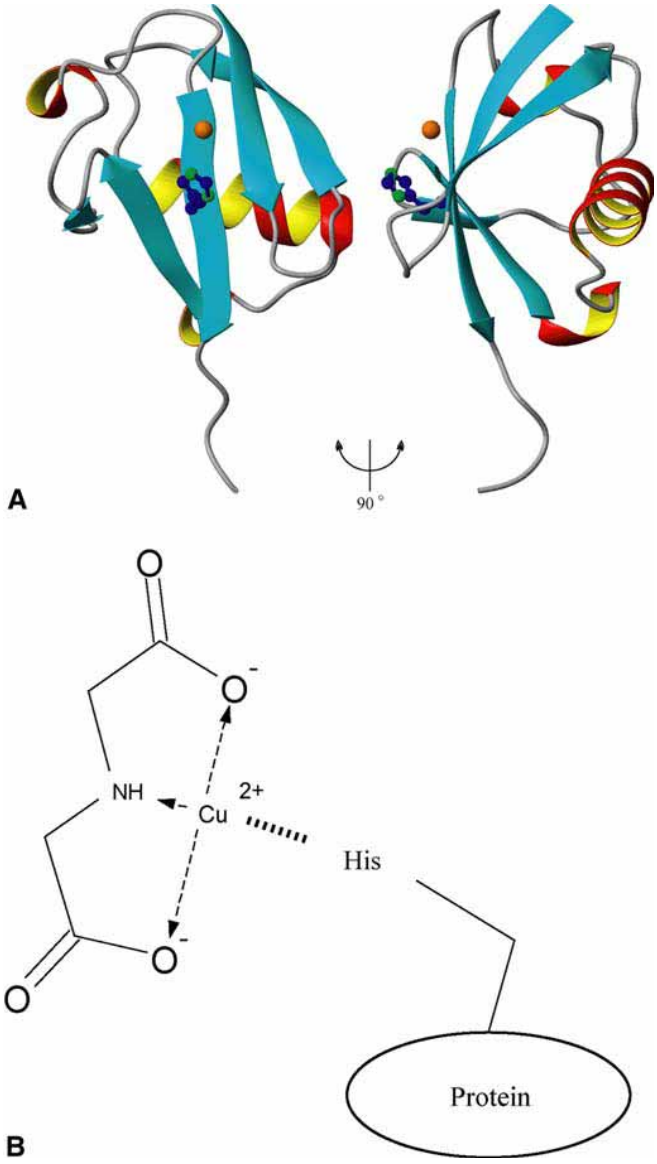



Fig. 2. The ribbon structure of ubiquitin with $Cu^{2+}$ (A), and the chemical structure of IDA and its coordination to $Cu^{2+}$ (B). Side chain of His68 is drawn as a ball-stick model with the two nitrogen atoms colored in green. $Cu^{2+}$ is shown as the gold sphere. The atomic coordinate of $Cu^{2+}$ is optimized, minimizing the calculated and observed proton transverse relaxation rates. The ribbon pictures were prepared by MOLMOL 2k.1 [30].

3.2. Paramagnetic relaxation enhancement analysis

Transverse relaxation rates were measured using spin echo experiments based on $^{1}H–^{15}N$ HSQC spectra in order to obtain long-range distance information. Distances from $Cu^{2+}$ to each amide proton were calculated using the crystal structure (PDB ID, 1UBI) and transverse relaxation rates, and were designated as the calculated and experimental distances, respectively. Using 49 out of 71 amide proton relaxation rates, the $Cu^{2+}$ position was optimized using the Mathematica software to minimize the difference between those distances. The distances from the optimized $Cu^{2+}$ position to $N^{\delta}$ and $N^{\varepsilon}$ of the His68 imidazole were 5.6 ± 0.6 and 4.5 ± 1.0 Å, respectively (Fig. 2). The B-factors for $N^{\delta}$and $N^{\varepsilon}$ were 27 and 28 $Å^2$, respectively, and much larger than those for the backbone atoms of His68 (2–6 $Å^2$). This indicates that the imidazole ring is flexible and that those coordinates are less reliable. Additionally, the delocalization of the paramagnetic electron might modulate the relaxation center position [2]. Thus, it is not clear from the distance measurements if His68 does indeed bind the $Cu^{2+}$–IDA.

The correlation plot between the calculated and experimental distances is shown in Fig. 3. The correlation coefficient R was 0.75. The observed transverse relaxation enhancement was up to 20 Å and the mean error from the calculated values was approximately 3.6 Å. This result highlights the potential in acquiring long-range distance information using the $Cu^{2+}$–IDA complex, although the precision was not very high. Relaxation experiments

were repeated on samples containing a change in $Cu^{2+}$–IDA concentration from 0.4 to 0.2 or 0.8 mM. The relaxation data determined were fitted to the theoretical equation and were well converged. The converged coordinates of $Cu^{2+}$ were quite similar to each other for the three data sets (RMSD = 1.3 Å). The experimental distances were correlated with the calculated ones, where R = 0.67 and 0.80 for the 0.2 and 0.8 mM samples, respectively.

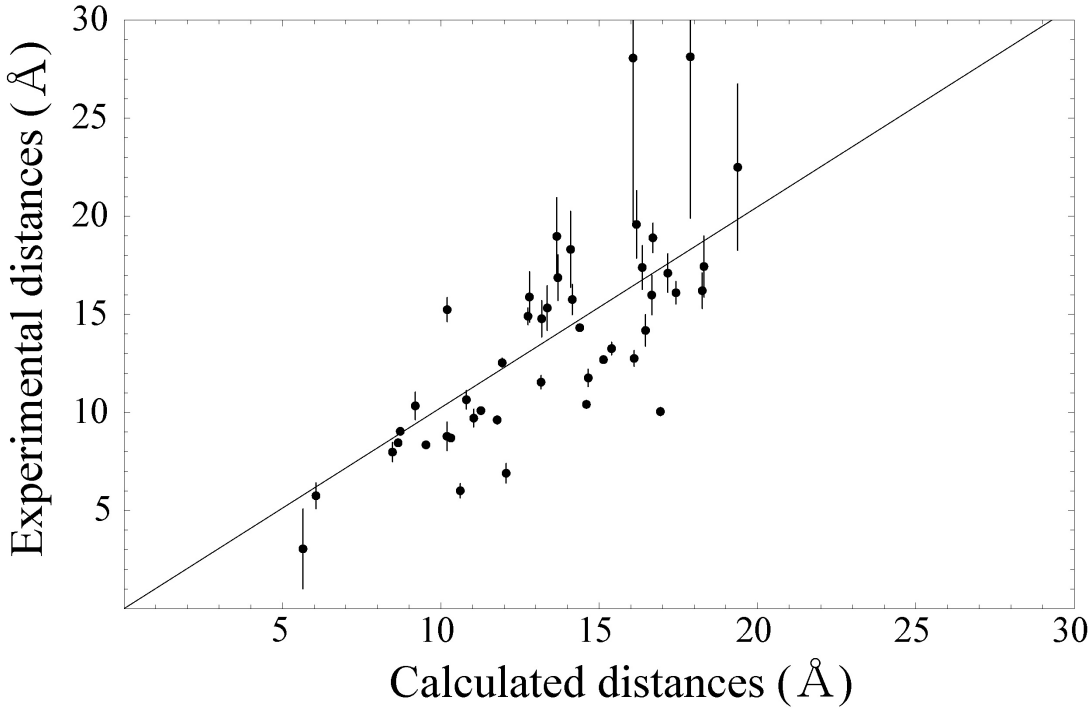



Fig. 3. Correlation plot between calculated and experimental distances from $Cu^{2+}$ to each backbone amide proton of ubiquitin. Calculated distances were taken from the crystal structure. The experimental distances were obtained from the paramagnetic enhancement of the proton transverse relaxation rates.

3.3 Electron paramagnetic resonance

The binding between ubiquitin and $Cu^{2+}$–IDA was evaluated further by EPR. EPR spectra were recorded by a JEOL JES-TE300 X-band spectrometer at 77 K using 2.5 mM

$Cu^{2+}$–IDA. The spectra significantly changed in the presence and absence of ubiquitin ($g_{\|}$ = 2.27 and 2.32, $A_{\|}$ = 15.8 and 15.9, and $g_{\perp}$ = 2:06 and 2.07, respectively). In the ubiquitin titration series, $Cu^{2+}$–IDA spectra were explained by the weighted sum of only two spectra, i.e., free and 1:1 complex. EPR splitting patterns of $Cu^{2+}$–IDA complexed with a histidine were quite similar to ubiquitin complex. These data strongly suggest that $Cu^{2+}$–IDA coordinates to ubiquitin at one specific site, His68. The localization of $Cu^{2+}$–IDA on two other proteins was investigated to reveal whether $Cu^{2+}$–IDA specifically binds to the surface histidine residue.

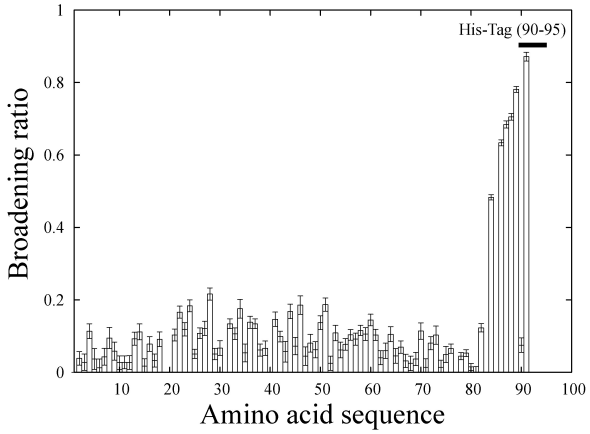
3.4 Application to 6x histidine tag

The $^{1}H$–$^{15}N$ HSQC spectra of HR23B UbL and IRF4 DBD with and without $Cu^{2+}$–IDA clearly demonstrated the site-specific line broadenings and shifts around the histidine residues. Broadening ratios for all proteins examined here are given in Fig. 4. Distance information was not elucidated for HR23B UbL and IRF4 DBD, since the coordinates were not available for the free proteins. A cysteine residue is the potential binding site of $Cu^{2+}$–IDA, but not examined in this study because of two difficulties. First, $Cu^{2+}$–IDA complex cannot be used with dithiothreitol (DTT) since a $Cu^{1+}$-DTT complex is formed [24]. Second, a free cysteine residue decreases protein stability, and the mixture of the

$Cu^{2+}$–IDA and a cysteine forms aggregates. Thus, *N*-ethylmaleimide was used to chemically modify the surface cysteine residue of IRF4 DBD in my experiments.

3.5 Application to the membrane protein

Membrane proteins are usually solubilized within a detergent micelle and it is difficult to construct an expression system. Distance information in general is insufficient, and any specific modification appears daunting, to say the least. I applied the $Cu^{2+}$–IDA system to a membrane protein solubilized within a detergent, and the $^{1}H$–$^{15}N$ HSQC spectrum showed site-specific line broadenings and shifts (Fig. 5-7). These results indicate the $Cu^{2+}$–IDA system is potentially applicable in the structural study of membrane proteins (Fig. 8).



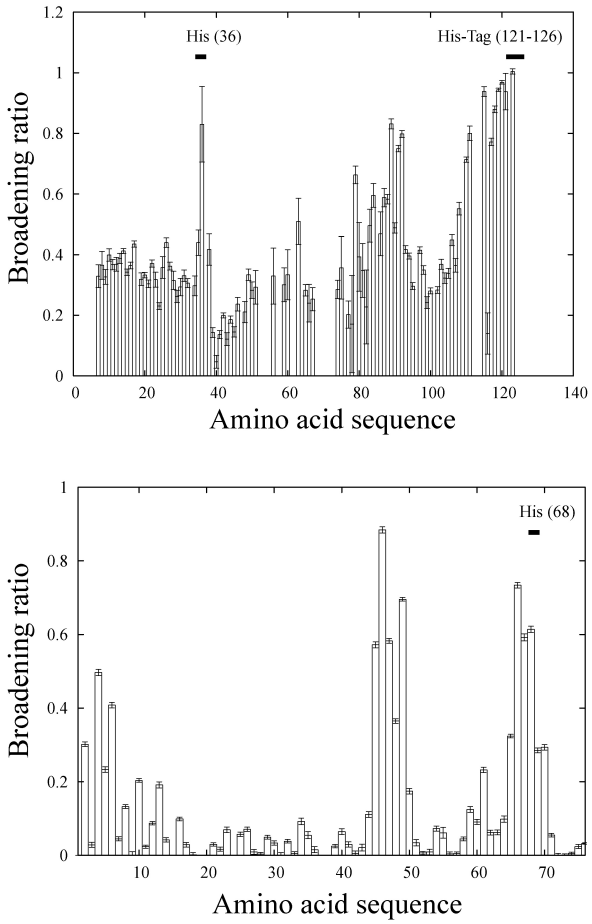



Fig. 4. The broadening ratios of HR23B UbL (A), IRF4 DBD (B), and ubiquitin (C) were plotted against the amino acid sequence. The broadening ratio is defined as $\sqrt{(I-I_0)^2}/\sqrt{I_0{}^2}$, where $I$ and $I_0$ represent the peak height with and without $Cu^{2+}$–IDA (1 equiv.), respectively. The 79th to 92th residues of IRF4 DBD were spatially close to His36 in the modeled structure.

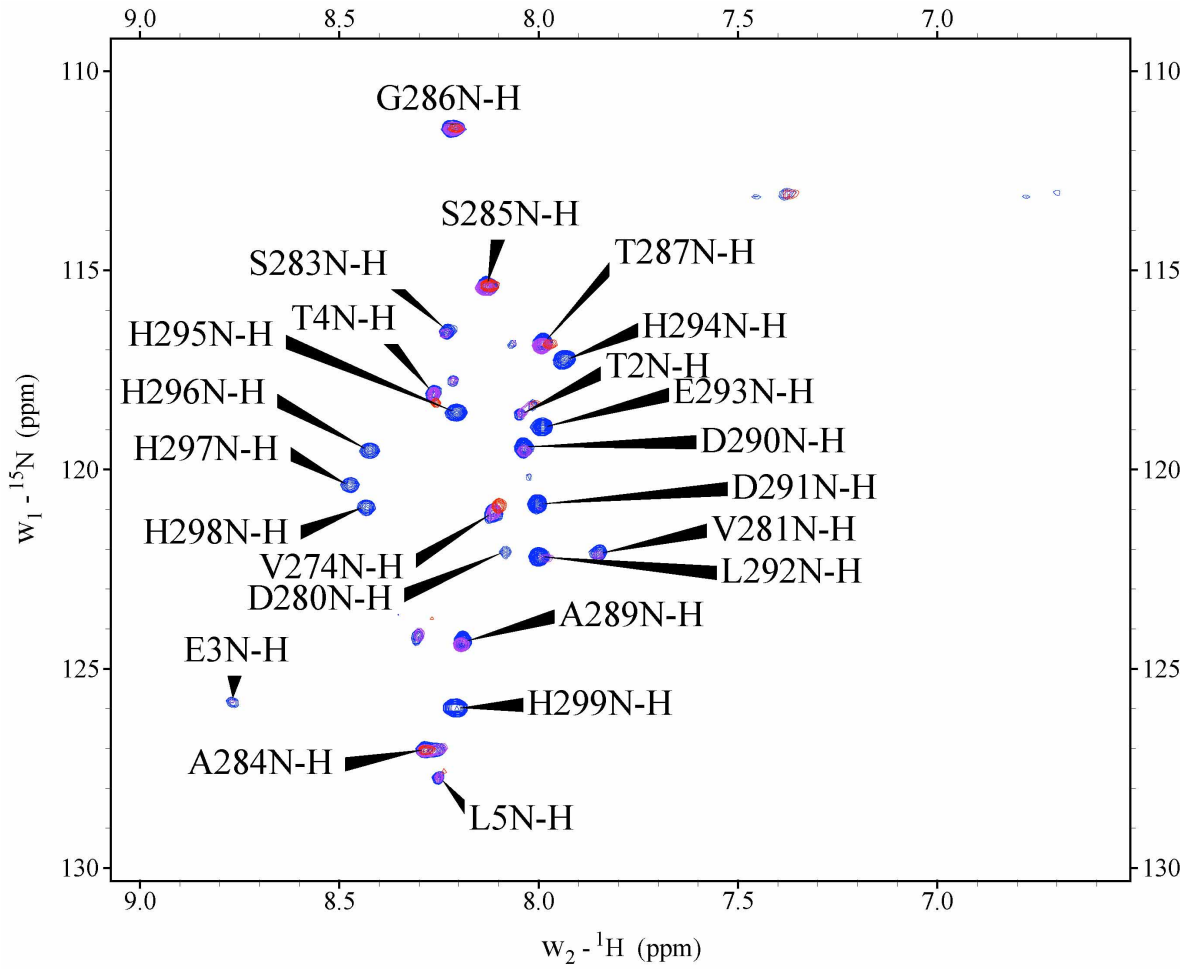



Fig. 5. $^{1}H$-$^{15}N$ HSQC spectra of *p*hR with the $Cu^{2+}$-IDA complex recorded on a Bruker DRX800 spectrometer with 0.4 mM protein, 10 mM citric acid (pH 5.0), 100 mM KCl and 0.4% Octyl-ß-D-thioglucoside at 298 K. Resonance assignments given in the Figure were derived from HNCACB, HN(CO)CACB, HNCO, and HN(CA)CO experiments. The ratios of $Cu^{2+}$-IDA to *p*hR are 0:1 (blue), 1:1 (purple), and 7:1 (red).

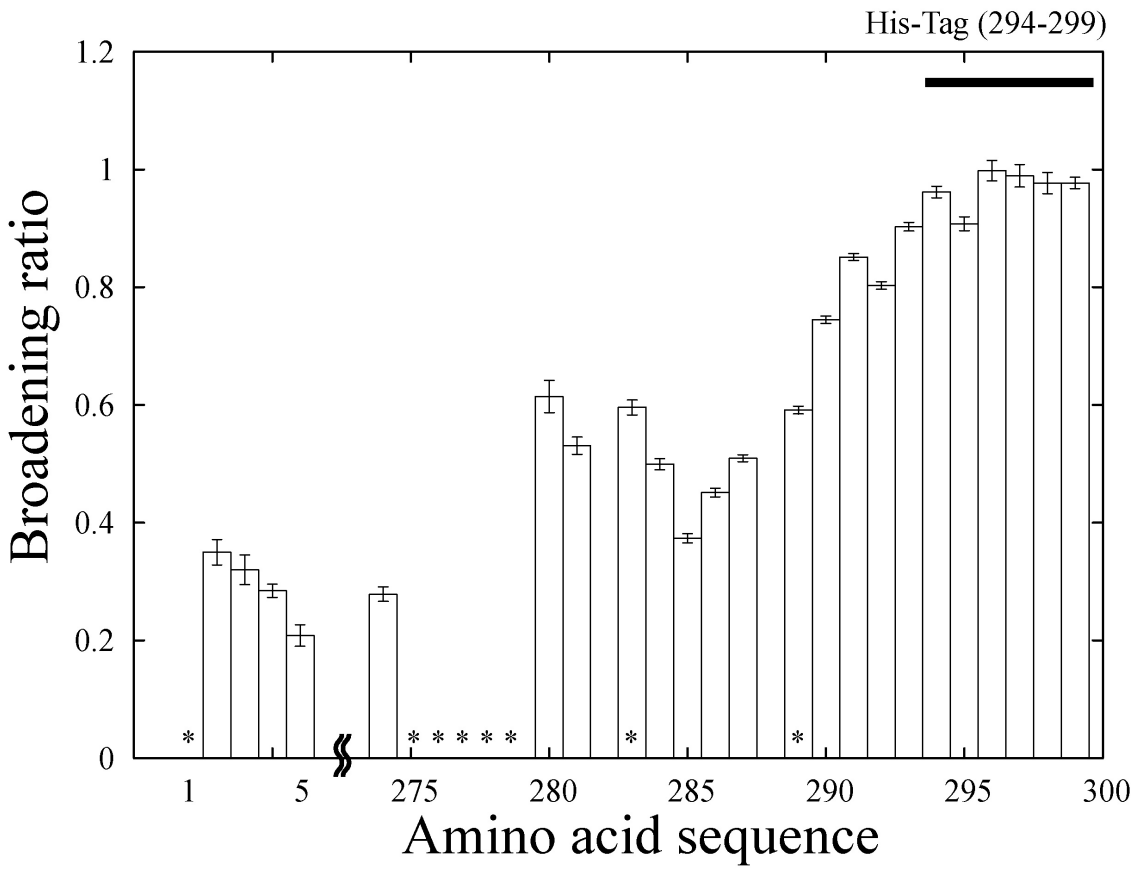



Fig. 6. The broadening ratios for *p*hR. The broadening ratio was calculated using $\sqrt{(I-I_0)^2}/\sqrt{I_0^2}$, where $I$ and $I_0$ represent the peak height with and without excess $Cu^{2+}$-IDA (7 equivalents), respectively (Figure 5). Asterisks represent the non-assigned peaks. Broadening ratios are higher for amino acids closely located to the His-tag region.

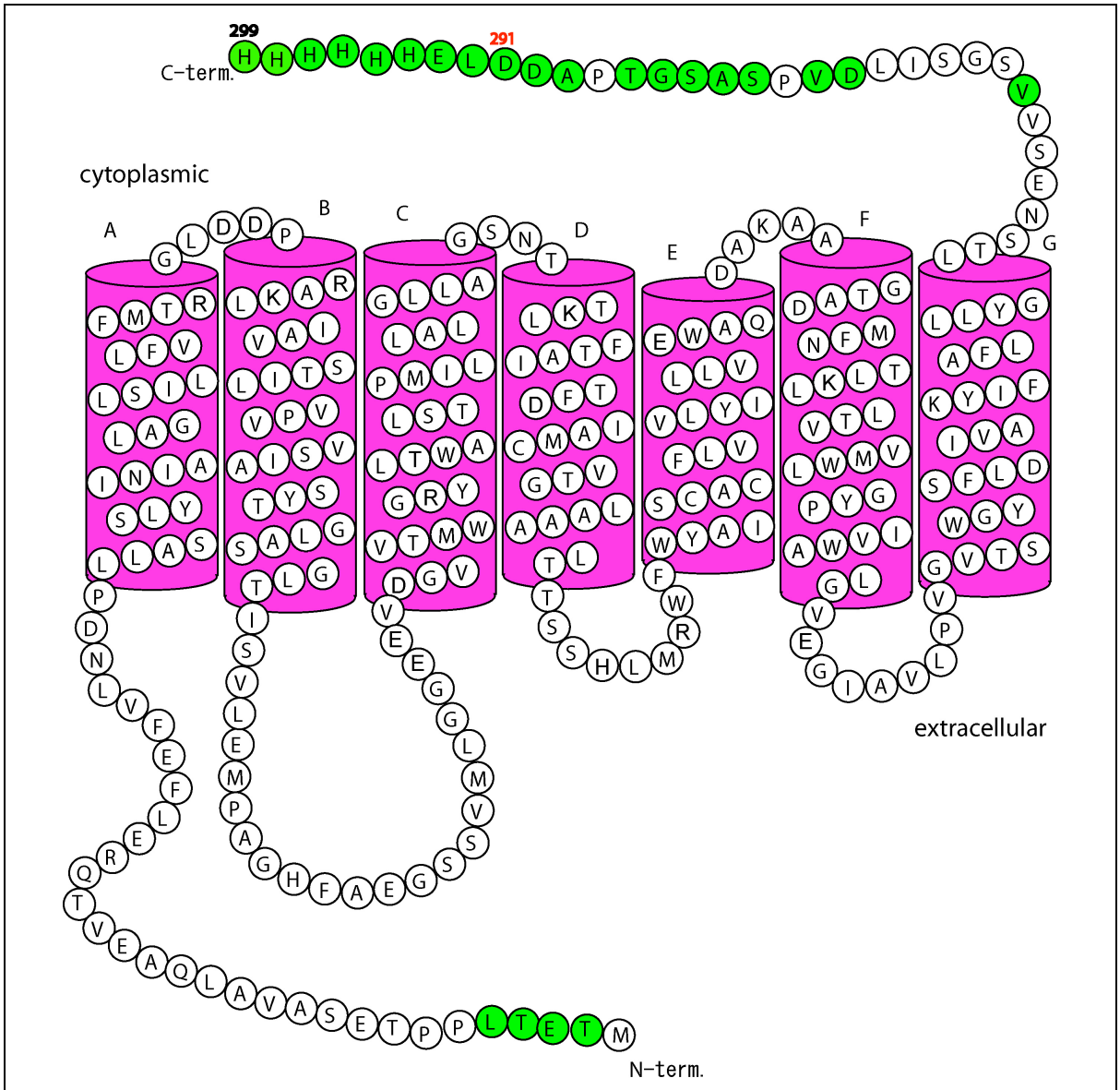



Fig. 7. Schematic drawing of the amino acid sequence and trans-membrane region of *p*hR. Circles colored in green represent the assigned amino acids on $^{1}H$-$^{15}N$ HSQC spectra shown in Figure 5.

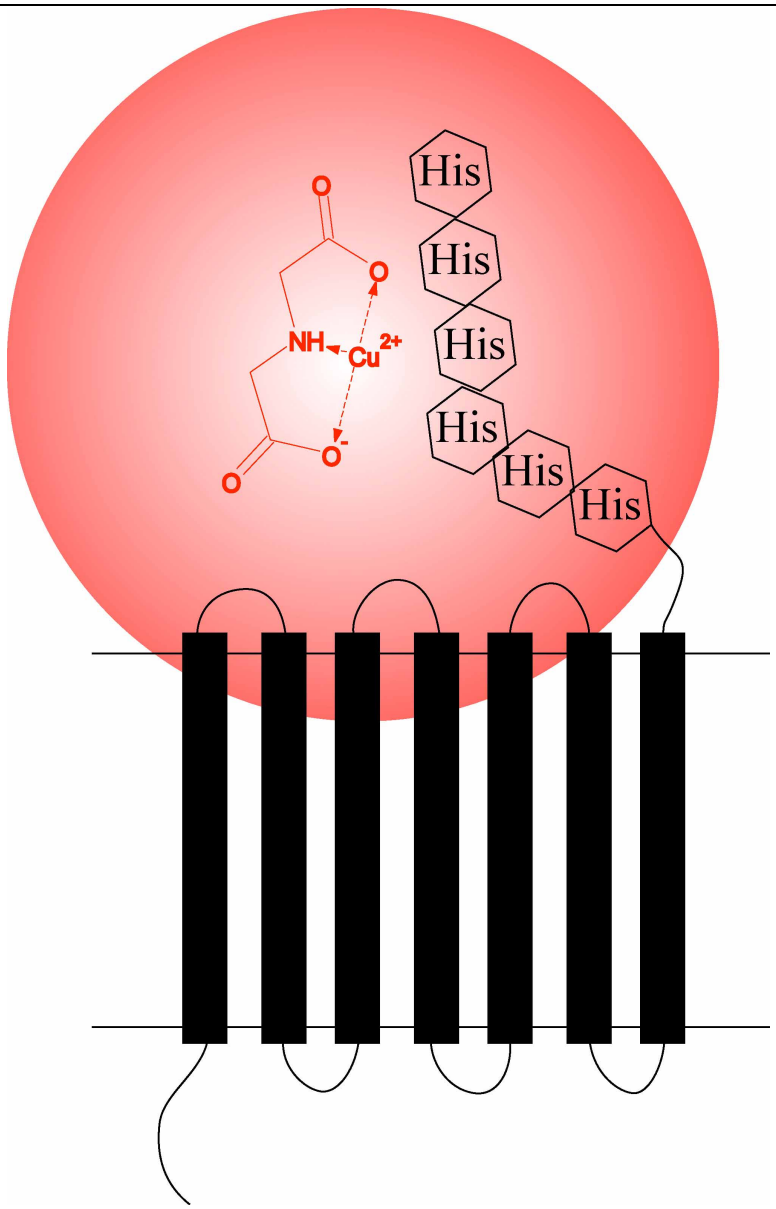



Fig. 8. Potential application of the $Cu^{2+}$-IDA system to the investigation of membrane proteins. Signal broadenings were observed up to 20 Å from $Cu^{2+}$, and therefore the $Cu^{2+}$-IDA system could distinguish amino acids located on the upper side of the membrane. Typical trans-membrane region is longer than 30 Å.

## Discussion

Other paramagnetic metal chelate complexes, such as $Mn^{2+}$–NTA and $Gd^{3+}$–EDTA, did not bind specifically to the histidine residues for three proteins possessing a poly hisitidine tag;

HR23B UbL, IRF4 DBD and the S5a UbL binding region. These results are consistent with hard-soft-acid-base (HSAB) theory [25–27].

Characteristic properties of the $Cu^{2+}$–IDA complex, that includes highly specific localization to the histidine residue and the capability of obtaining distance information, offer clear advantages in the analysis of macromolecular intermolecular interactions such as protein–protein, protein–DNA and protein–ligand interactions. Hansen et al. [28] employed the paramagnetic longitudinal proton relaxation enhancement of a metal protein to detect transient intermolecular protein–protein interactions. Mal et al. [9] utilized the ATCUN $Cu^{2+}$ system to detect protein–protein interactions. Therefore, transient interactions showing few intermolecular NOEs can be the subject of investigation using $Cu^{2+}$–IDA provided that a surface histidine residue is located near the interface. Another potential application of the $Cu^{2+}$–IDA complex can be in the investigation of membrane proteins.

In conclusion, all examined paramagnetic metal chelate complexes were not localized to a specific site on the protein surface with exception of $Cu^{2+}$–IDA that was specifically localized to His68 on the ubiquitin surface. This feature of $Cu^{2+}$–IDA was successfully utilized to elucidate long-range distance information from proton transverse relaxation enhancements. $Cu^{2+}$–IDA could also be used in conjunction with poly histidine tagged proteins, but it may be difficult to obtain precise distance constraints for proteins

possessing many histidine residues. Application of the $Cu^{2+}$–IDA method to the investigation of surface histidine residues is a technique worthy of consideration, since surface histidine residues are frequently found in many proteins and can play important roles in many biochemical reactions and interactions.

## Acknowledgements

I thank H. Okuda for his assistance, Prof. J.L. Markley and Prof. C. Griesinger for their critical comments, and Prof. S. Suzuki for his helpful comments on EPR analyses. This work was supported in part by Grants-in-Aid for 21st Century COE Research and Scientific Research, a Grant for the National Project on Protein Structural and Functional Analyses from MEXT (the Japanese Ministry of Education, Culture, Sports, Science and Technology), a Kaneko Narita Research Grant of Protein Research Foundation and a Grant from the Yamada Science Foundation.

## References

[1] Bertini, I., Donaire, A., Jimenez, B., Luchinat, C., Parigi, G., Piccioli, M. and Poggi, L. (2001) J. Biomol. NMR 21, 85–98.

[2] Bertini, I., Luchinat, C. and Parigi, G. (2001) Solution NMR of Paramagnetic Molecules. Elsevier, Amsterdam.

[3] Bertini, I., Luchinat, C. and Piccioli, M. (2001) Methods Enzymol. 339, 314–340.

[4] Gochin, M. (1997) J. Am. Chem. Soc. 119, 3377–3378.

[5] Gochin, M. (1998) J. Biomol. NMR 12, 243–257.

[6] Kikuchi, J., Iwahara, J., Kigawa, T., Murakami, Y., Okazaki, T. and Yokoyama, S. (2002) J. Biomol. NMR 22, 333–347.

[7] Arnesano, F., Banci, L., Bertini, I., Felli, I.C., Luchinat, C. and Thompsett, A.R. (2003) J. Am. Chem. Soc. 125, 7200–7208.

[8] Donaldson, L.W., Skrynnikov, N.R., Choy, W.Y., Muhandiram, D.R., Sarkar, B., Forman-Kay, J.D. and Kay, L.E. (2001) J. Am. Chem. Soc. 123, 9843–9847.

[9] Mal, T.K., Ikura, M. and Kay, L.E. (2002) J. Am. Chem. Soc. 124, 14002–14003.

[10] Dvoretsky, A., Gaponenko, V. and Rosevear, P.R. (2002) FEBS Lett. 528, 189–192.

[11] Gaberc-Porekar, V. and Menart, V. (2001) J. Biochem. Biophys. Methods 49, 335–360.

[12] Yip, T.T., Nakagawa, Y. and Porath, J. (1989) Anal. Biochem. 183, 159–171.

[13] Hutchens, T.W. and Yip, T.T. (1990) Anal. Biochem. 191, 160–168.

[14] Mrabet, N.T. (1992) Biochemistry 31, 2690–2702.

[15] Murphy, J.C., Jewell, D.L., White, K.I., Fox, G.E. and Willson, R.C. (2003) Biotechnol. Progr. 19, 982–986.

[16] Berna, P.P., Mrabet, N.T., Van Beeumen, J., Devreese, B., Porath, J. and Vijayalakshmi, M.A. (1997) Biochemistry 36, 6896–6905.

[17] Jiang, K.Y., Pitiot, O., Anissimova, M., Adenier, H. and Vijayalakshmi, M.A. (1999) Biochim. Biophys. Acta 1433, 198–209.

[18] Fujiwara, K., Tenno, T., Sugasawa, K., Jee, J.G., Ohki, I., Kojima, C., Tochio, H., Hiroaki, H., Hanaoka, F. and Shirakawa, M. (2004) J. Biol. Chem. 279, 4760–4767.

[19] Delaglio, F., Grzesiek, S., Vuister, G.W., Zhu, G., Pfeifer, J. and Bax, A. (1995) J. Biomol. NMR 6, 277–293.

[20] Goddard, T.D. and Kneller, D.G. (1999) Sparky 3. University of California, San Francisco.

[21] Schneider, D.M., Dellwo, M.J. and Wand, A.J. (1992) Biochemistry 31, 3645–3652.

[22] Sulkowski, E. (1987) in: Protein Purification: Micro to Macro (Burgess, R.R., Ed.), pp. 149–162, Alan R. Liss Inc, New York.

[23] Hochuli, E., Dobeli, H. and Schacher, A. (1987) J. Chromatogr. 411, 177–184.

[24] Kr zel, A., Lesniak, W., Jezowska-Bojczuk, M., Mlynarz, P., Brasun, J., Kozlowski, H. and Bal, W. (2001) J. Inorg. Biochem. 84, 77–88.

[25] Aime, S., D'Amelio, N., Fragai, M., Lee, Y.M., Luchinat, C., Terreno, E. and Valensin, G. (2002) J. Biol. Inorg. Chem. 7, 617–622.

[26] Kemple, M.D., Ray, B.D., Lipkowitz, K.B., Prendergast, F.G. and Nageswara Rao, B.D. (1988) J. Am. Chem. Soc. 110, 8275–8787.

[27] Petros, A.M., Mueller, L. and Kopple, K.D. (1990) Biochemistry 29, 10041–10048.

[28] Hansen, D.F., Hass, M.A., Christensen, H.M., Ulstrup, J. and Led, J.J. (2003) J. Am. Chem. Soc. 125, 6858–6859.

[29] Sayle, R.A. and Milner-White, E.J. (1995) Trends Biochem. Sci. 20, 374.

[30] Koradi, R., Billeter, M. and Wuthrich, K. (1996) J. Mol. Graph. 14, 51–55, 29–32.

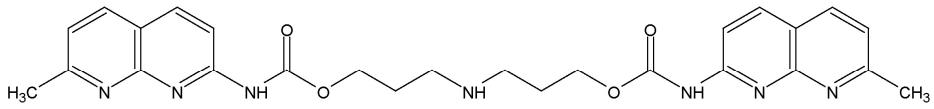
## Study 2: Drug DNA system

## Introduction

The trinucleotide diseases caused by $(CXG)_n$ repeats have characteristic mechanisms of pathogenesis. For example, fragile X syndrome is caused by the $(CGG)_n$ repeats in the FMR1 gene. The number of normal repeat length, n, is 6-60. When n goes over specific threshold, 200, the CpG island is methylated and going to be pathogenesis[1-3]. Thus, it is important to develop the counting technology of the repeat length n. The naphtyridine carbamate dimer (**NC**) was developed as the CGG/CGG triad recognition drug (Figure 1) [4]. By the CSI-TOF MS experiments, the stoichiometry of **NC** to CGG/CGG was 2:1. The chemical cleavage experiment showed two cytosines in the CGG/CGG triad cleavaged by hydroxylamine in the complex condition. This result implied the disruption of the G-C base pairing. As the previous study of the 3D structure of drug-CXG/CXG triad, we have recently determined the naphthyridine azaquinolone (**NA**) – CAG/CAG complex [5]. However, it is not known the recognition mechanisms of **NC** to the CGG/CGG triad.

Here, I aimed to understand the recognition mechanisms of **NC** to the CGG/CGG triad by the determination of the 3D structure.

(A)



(B)

5'-$C^1$ $T^2$ $A^3$ $A^4$ $\underline{C^5\ G^6\ G^7}$ $A^8$ $A^9$ $T^{10}$ $G^{11}$-3'

3'-$G^{22}$ $A^{21}$ $T^{20}$ $T^{19}$ $\underline{G^{18}\ G^{17}\ C^{16}}$ $T^{15}$ $T^{14}$ $A^{13}$ $C^{12}$-5'

***Figure 1.*** Molecular structure of naphtyridine carbamate dimer (**NC**) (A). The DNA sequence and numbering schemes used in this study (B).

## Materials and Methods

*Sample Preparation.* The oligomers 5'-d(CTAACGGAATG)-3' and 5'-d(CATTCGGTTAG)-3' were purchased from Fasmac Co. Ltd (Figure 1). These oligomers were dissolved in 112.5 µl of 20 mM sodium phosphate buffer with 0.1 mM NaCl (pH 6.8) . The oligomers were miexed in a 1:1 stoichiometry and annealed over night. This duplex were dialyzed against 25 ml of 20 mM sodium phosphate buffer with 1M NaCl (pH 6.8) to remove anionic impurities at three times using a membrane with a 3.5 kDa molecular weight cutoff, then dialyzed against 25 ml of 20 mM sodium phosphate with 0.1mM NaCl (pH 6.8). 25 µl $D_2O$ was added after dialyze. The final concentration of the duplex was 2.5 mM checked by UV absorbance at 260 nm.

The **NC** was synthesized as previously reported [4]. The 3.35 mg of **NC** was dissolved in 50µl of 20 mM sodium phosphate buffer with 0.1 mM NaCl (pH 6.8). The final

concentration of this stock solution was 125 mM, checked by UV absorbance at 321 and 332 nm.

*NMR Measurements.* $^{1}$H and $^{31}$P measurements were carried out in 250 μl Shigemi microcells. Experiments were carried out on Bruker Avance-800 spectrometer, operating at 800 MHz $^{1}$H frequency, and equipped with triple-resonace, three-axis pulse field gradient probeheads. Natural abundance $^{13}$C measurements were carried out in 250 μl Shigemi microcells. Experiments were carried out on Bruker DMX-500 spectrometer, operating at 500 MHz $^{1}$H frequency, and equipped with triple-resonance, three-axis pulse field gradient probeheads and cryogenic-probeheads.

The titration experiments of **NC** to 2.5 mM DNA duplex were carried out at 283 K with monitoring the region of the imino protons of DNA and the peptide HN proton signals of **NC** (9.5 – 14.0 ppm) of 1D $^{1}$H NMR spectra using the 1-1 echo water suppression sequence.

2D NOESY spectra were recorded with 4096 and 1024 complex points with a spectral width of 16026 and 16003 Hz and zero-filled to 8192 and 4096 in $t_2$ and $t_1$, respectively, with the mixing time of 30, 100 and 200 ms at 283 K. The recycle delay of NOESY spectrum with the mixing time 200 ms was set to 7 s, and this spectrum was used to obtain

the distance restraints for the structural calculations. The DQF-COSY, TOCSY and natural abundance $^{1}$H-$^{13}$C HSQC spectra were also recorded to conform the assignments. $^{1}$H-$^{31}$P HSQC spectra were recorded by optimizing $\tau_a$ = 16 ms as previously reported [6].

*Assignment and Restraints.* The 2D NOE intensities at the mixing time of 200 ms were determined by the fitting method using a Gaussian function in Sparky [7]. The resonances of metyl protons of thymidines and **NC** were assigned by intra-residue H6-Me and **NC** aromatic-Me cross peaks. The stereospecific assignment of H2'/H2" was based on the intensity difference between H1'-H2' cross-peaks and H1'-H2" cross-peaks of the NOESY spectrum at the mixing time 30 ms. The H5 and H6 resonances of cytosine and the chromophore proton resonances of **NC** were confirmed by DQF-COSY and TOCSY spectra. The H5' and H5" resonances were assigned from the residual NOEs and confirmed by H5'-H5" geminal cross-peaks to distinguish the H4' resonances. The resonances of metylene protons of **NC** were assigned by the cross-peaks with peptide protons which formed hydrogen bond to guanosine bases and the geminal cross-peaks. The resonances of $^{31}$P were assigned by the cross-peaks with H3' and H4' in the $^{1}$H-$^{31}$P HSQC spectra. The $^{1}$H-$^{13}$C HSQC spectra were also used to confirm the assignments.

For obtaining distance restraints, complete relaxation matrix analysis with the program

MARDIGRAS was used. The errors were generated by RANDMARDI mode with 100 iterations [8]. The correlation time was set to 5.14 ns from the average value of 5.03 ns from CSA-DD and 5.25 ns from T1/T2 determined for a 10 mer DNA duplex at 283 K [9,10]. The initial structure was prepared by the uniformly 5 Å distance restraints. Final distances with their errors were determined by averaging the upper and lower bonds and converted to the XPLOR/CNS format.

*Structure Calculation.* The initial structure of **NC** was optimized and parameterized by DFT calculation (STO-3G) by Gaussian. Structure calculations were performed by CNS 1.1 [11]. In the high temperature stage, the structures were heated to 20000 K in the course of 32000 steps of 15 ps. During this calculation, the force constants were 0.1 kcal $mol^{-1}$ $Å^{-4}$ for the van der Waals repulsion term, 60 kcal $mol^{-1}$ $Å^{-2}$ for experimental distance restraints, 5 kcal $mol^{-1}$ $rad^{-2}$ for dihedral restraints. In the first torsion slow-cooling stage, the structures heated to 20000 K and then cooled over the course of 16000 steps of 5 ps and 25 K of temperature step. During this calculation, the force constants were 1.0 kcal $mol^{-1}$ $Å^{-4}$ for the van der Waals repulsion term, 60 kcal $mol^{-1}$ $Å^{-2}$ for experimental distance restraints, 100 kcal $mol^{-1}$ $rad^{-2}$ for dihedral restraints. In the second cartesian slow-cooling stage, the structures heated to 2000 K and then cooled over the course of 24000 steps of 2 ps and 15

K of temperature step. During this calculation, the force constants were 1.0 to 4.0 kcal $mol^{-1}$ $Å^{-4}$ for the van der Waals repulsion term, 60 kcal $mol^{-1}$ $Å^{-2}$ for experimental distance restraints, 100 kcal $mol^{-1}$ $rad^{-2}$ for dihedral restraints. The structures were minimized during 10 cycles of 200 steps with experimental distance restraints and dihedral force constants of 30 kcal $mol^{-1}$ $Å^{-2}$ and 200 kcal $mol^{-1}$ $rad^{-2}$. The 10 lowest energy accepted structures were averaged and refined by XPLOR-NIH [12] with the RDCs data of the DNA bases and the **NC** chromophores. In the Cartesian coordinate dynamics, the force constats of the **RDC**s was the 0.01 to 0.5 kcal $mol^{-1}$ $Hz^{-2}$.

*RDC Measurements.* Pf1 phage (ASLA Labs) [13] was dissolved 250 μl $H_2O$ at a concentration 20 mg/ml. This solution was dialyzed against 25 ml of $H_2O$ at two times, then against 25 ml of $D_2O$. The alignment was conformed by quadrupolar splittings of $^2H$ NMR signals (12 Hz at 500 Mhz $^1H$ frequency). The **NC**-DNA complex was freeze-dried and then dissolved in $D_2O$. After repeating three times, the powder of the **NC**-DNA complex was dissolved in 250μl of 20 mg/ml Pf1 phage solution.

For $^1D_{CH}$ of the DNA bases and the **NC** chromophores, natural abundance $^1H$-$^{13}C$ IPAP HSQC spectra of carriers of 113 and 148 ppm were recorded with 512 and 64 complex points in a spectral width of 5482 and 5031 Hz and zero-filled to 2048 and 256 in $t_1$ and $t_2$,

respectively, at 283 K with and without Pf1 phage. The peak lists were obtained by SPARKY [7]. The $^{1}D_{CH}$ values were obtained by subtraction $^{1}D_{CH}$ + $^{1}J_{CH}$ with Pf1 phage from $^{1}J_{CH}$ without Pf1 phage.

For $D_{HH}$ of the cytosine bases and **NC** chromophores, DQF-COSY spectra were recorded with 8192 and 512 complex points in a spectral width of 8013 and 8002 Hz and zero-filled to 8192 and 1024 in $t_1$ and $t_2$, respectively, at 283 K with and without Pf1 phage. The diagonal peaks were suppressed using diagShift.M of NMRPipe modules [14,15]. The $D_{HH}$ + $^{3}J_{HH}$ with Pf1 phage and $^{3}J_{HH}$ without Pf1 phage values were obtained by amplitude-constrained multiplet evaluation (ACME) [15,16]. The amplitudes for constrain were obtained following protocol. The $D_{HH}$ + $^{3}J_{HH}$ or $^{3}J_{HH}$ values from H5-H6 of terminal residues, C1 and C12, were estimated by SPARKY in $t_1$. The obtained $D_{HH}$ + $^{3}J_{HH}$ or $^{3}J_{HH}$ values were used as constant values in constrain amplitude values as variable in ACME. The calculated amplitudes were used as constraint. The $D_{HH}$ values were obtained by subtraction $D_{HH}$ + $^{3}J_{HH}$ with Pf1 phage from $^{3}J_{HH}$ without Pf1 phage. In the case of unobserved peaks, the splitting values were set to zero.

*MD simulations.* All calculations were carried out using the AMBER8.0 program [17]. The Cornell force field (PARM99) and the force field which generated by Antechamber

modules with AM1-BBC charge method were used for the DNA and **NC**, respectively. The NMR structure was surrounded by a periodic box of size 47.7 Å × 64.6 Å × 48.7 Å with the 4062 TIP3P water molecules and 19 $Na^+$ ions for neutralization by LEAP module. After minimization with 1500 steps and 20 ps equilibration with 10 kcal positional restraints on the **NC**-DNA complex, the 2 ns simulation of 1,000,000 steps with a time step of 2 fs was carried out with constant 1 atm pressue, the Langevin dynamics for temperature control at 300 K, 10 Å cut off of the Lennard-Jones interactions, the particle mesh Ewald method (PME) for the electrostatic interactions by Sander module with SHAKE algorithm [18]. The trajectory was analyzed by Ptraj module and VMD software.

## Results

*Stoichiometry.* The titration experiment to 2.5 mM 5'-CTAACGGAATG-3'/5'-CATTCGGTTAG-3' duplex was carried out with monitoring the imino proton region. The peak shifts were slow-exchange manner (Figure 2). The peaks at 10.5-12.0 ppm appeared during the titration course. These peaks came from H1 imino proton of G6, G7, G17, G18 in the CGG/CGG triad and from amide proton of **NC**. The stoichimetry was 2:1 (**NC**:CGG/CGG). There is no intermediate state such as the 1:1 stoichiometry.

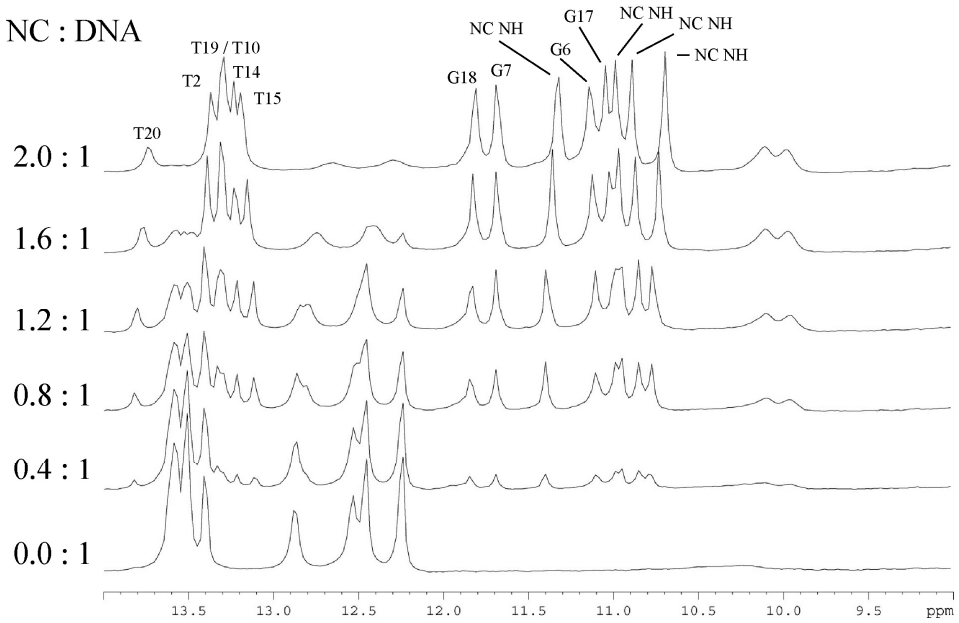



***Figure 2.*** The proton NMR spectra during a titration with **NC**.

*NMR assignment.* NOESY spectra at the mixing time 30, 200 and 300 ms were recorded. The assignment were confirmed by TOCSY, DQF-COSY, natural abundance $^{1}H$-$^{13}C$ HSQC spectra and $^{1}H$-$^{31}P$ HSQC spectra. At the stem regions of C1-A4, A8-G11, C12-T15 and T19-G22, the base-sugar cross peaks of sequential H6/H8-H1' and H6/8-H2' /H2" were observed (Figure 3(A)). At the CGG/CGG triad region, the inter-residue cross peaks H6/H8-H1' and H6/H8-H2'/H2" were not observed. The cross peaks between aromatic protons of **NC** and H8 of G were observed. These cross peaks suggested the base-chromophore stacking. The H5'/H5" protons in the CGG/CGG triad had cross peaks with the aromatic protons of **NC**. The methylene protons of one **NC** had the cross peaks with methylene protons of another **NC**. It suggested that two methylene linkers from two

NC were close in space. The $^{31}P$ chemical shifts were dispersed to 0.2-1.8 ppm in contrast with 1.0-1.6 ppm without **NC** (Figure 3(B)). The characteristic downfield and upfield shifts were occurred around the CGG/CGG triad.

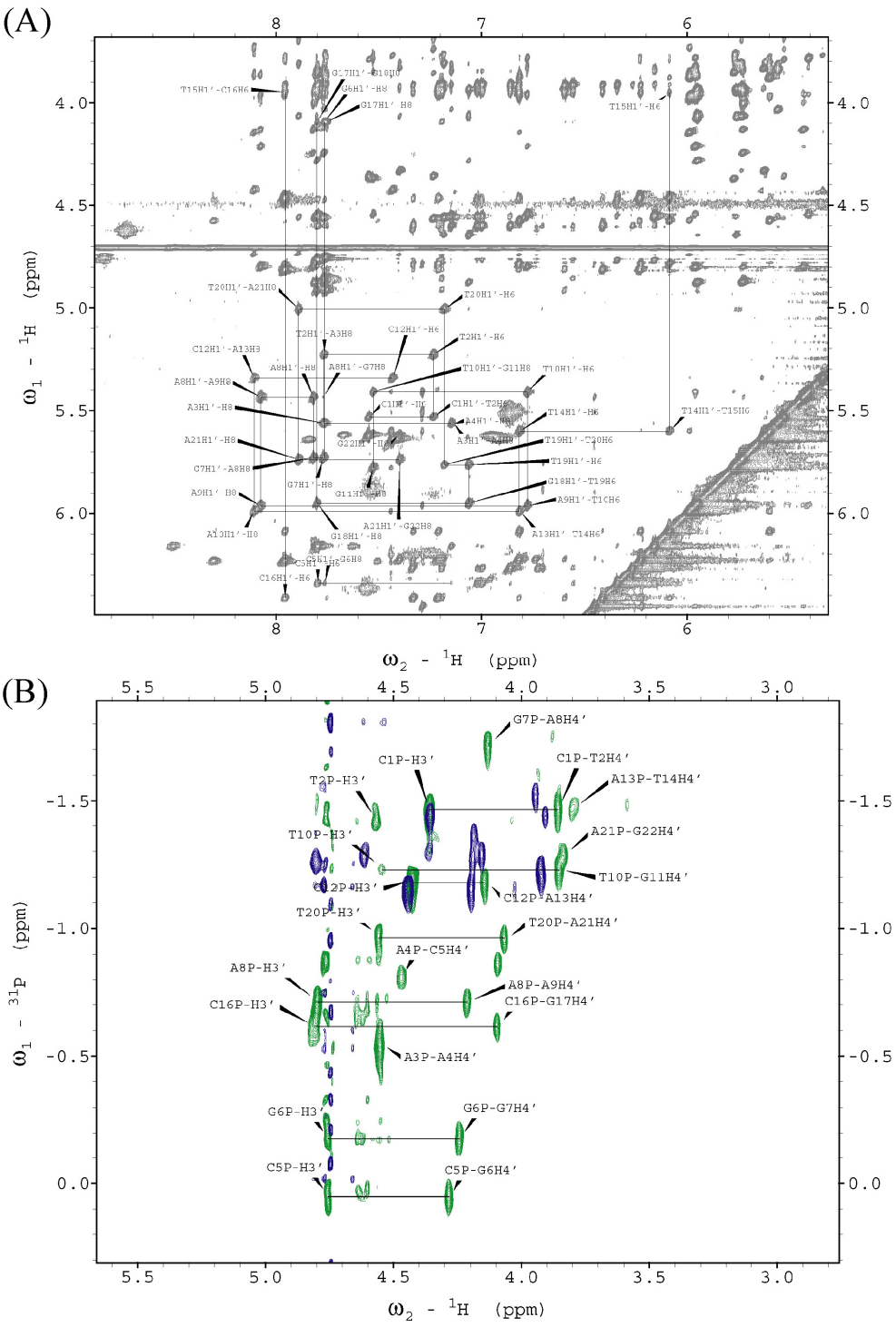



***Figure 3.*** The NOESY spectrum of the 2:1 **NC**:DNA complex is shown. Sequential connectivities of H6/8-H1' are indicated by solid lines.

*Structure determination.* The distance restraints were generated by RANDMARDI with NOESY spectrum at the mixing time 200 ms and 7 s recycle delay. The hydrogen bonding and planarity constraints were empirically introduced. The dihedral angle restraints were standard B-form at the stem site and nothing at and near the CGG/CGG triad. Total restraints were 533 NOEs, 74 dihedral angles and 21 planarity. All the restraints of C5 and C16 were not included because of the ambiguity. The structure calculation was carried out by CNS [11]. The refinement with the RDCs of the DNA bases and the **NC** chromophores were carried out by XPLOR-NIH [12]. As a result of the determined structure, two **NC** intercalated to the CGG/CGG triad (Figure 4(A-B)). Four naphthyridine chromophores of two **NC** formed hydrogen bonding with the four guanosine bases in the CGG/CGG triad (Figure 4(C)). The four methyl groups of the naphtyridine chromophores of two **NC** formed hydrophobic packing along the long axis. The two cytsines in the CGG/CGG triad were flipping out and forming the C-Bulges. The phosphodiester backbones were unwinding at the CGG/CGG triad.

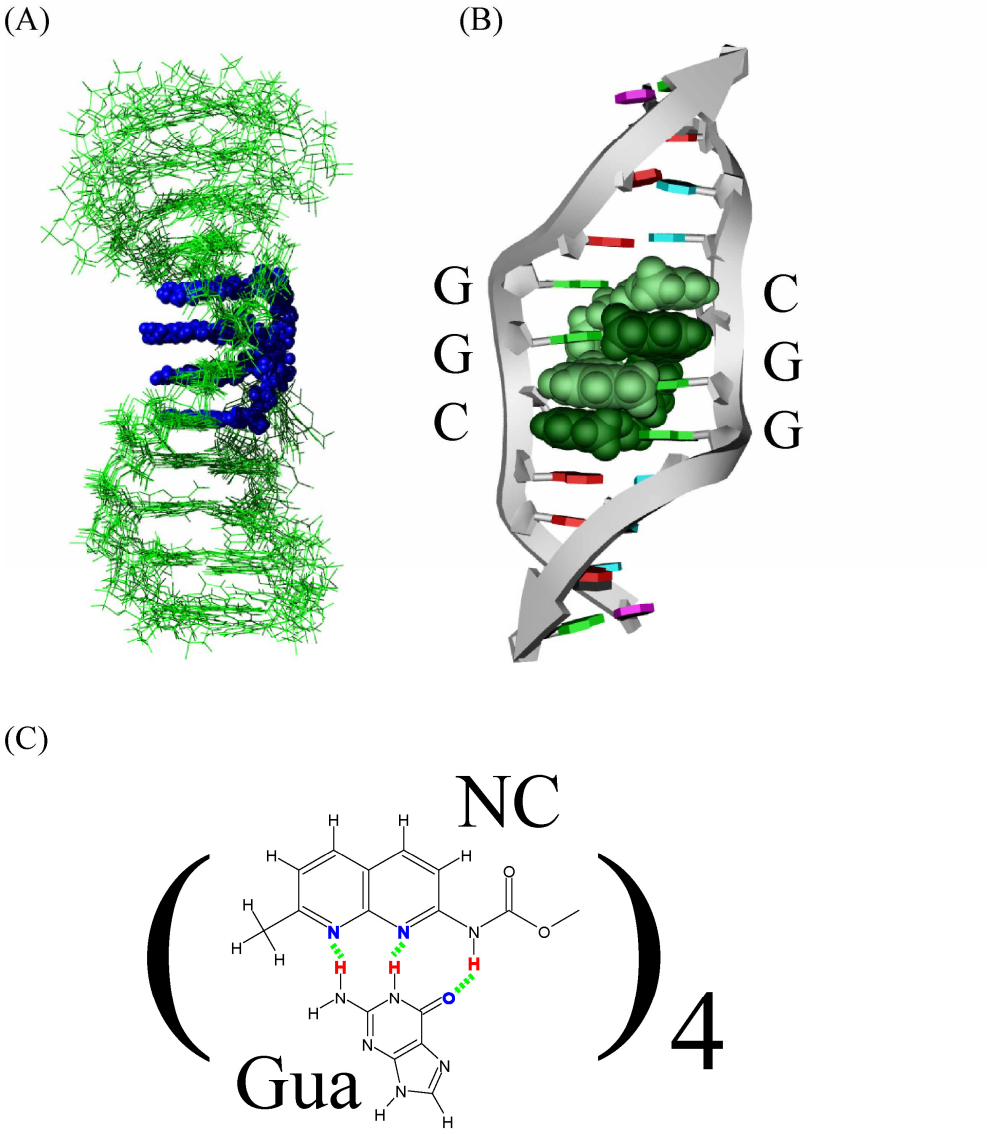



***Figure 4.*** The 10 lowest-energy structures of the 2:1 **NC**:DNA complexes are shown (A). Blue spheres indicate two **NC**s. The average structures from (A) with 90 degree rotation is shown (B). Four **NC** chromophores form hydrogen bonding with four guanosine bases as shown in (C).

*RDC experiment.* The NOE based structure of the **NC**-CGG/CGG complex gave less information about the flipping out of C5 and C16. To solve this problem, I have measured residual dipole coupling (**RDC**) of the DNA bases and the naphtyridine chromophores of

**NC**. The **RDC** values are ruled by the relative direction between the $^{1}$H-$^{13}$C or $^{1}$H-$^{1}$H vectors and the molecular alignment tensor. To obtain the $^{1}$H-$^{13}$C **RDC** values, natural abundance $^{1}$H-$^{13}$C IPAP HSQC spectra were recorded with and without pf1 phage as the alignment media. Two **RDC** values from C5-H5 of the flipping out C5 and C16 showed the negative values in contrast to the others from stacked in bases and chromophores showed the positive one (Figure 5(A) and Figure 6(A)). To obtain the $^{1}$H-$^{1}$H **RDC** values, DQF-COSY spectra were also recorded with and without pf1 phage. Two **RDC** values from H6-H5 of the flipping out C5 and C16 were disappeared after introducing pf1 phage in spite of the other peaks which include terminal two C1 and C12 were observed. It was due to the cancellation of the J values by the negative **RDC** values because the individual H6 and H5 peaks were observable in the $^{1}$H-$^{13}$C HSQC spectra at the same condition (Figure 5(B) and Figure 6(B)). These negative **RDC** values of the flipping out C5 and C16 showed that the directions of these bases were different to the other bases and the chromophores.

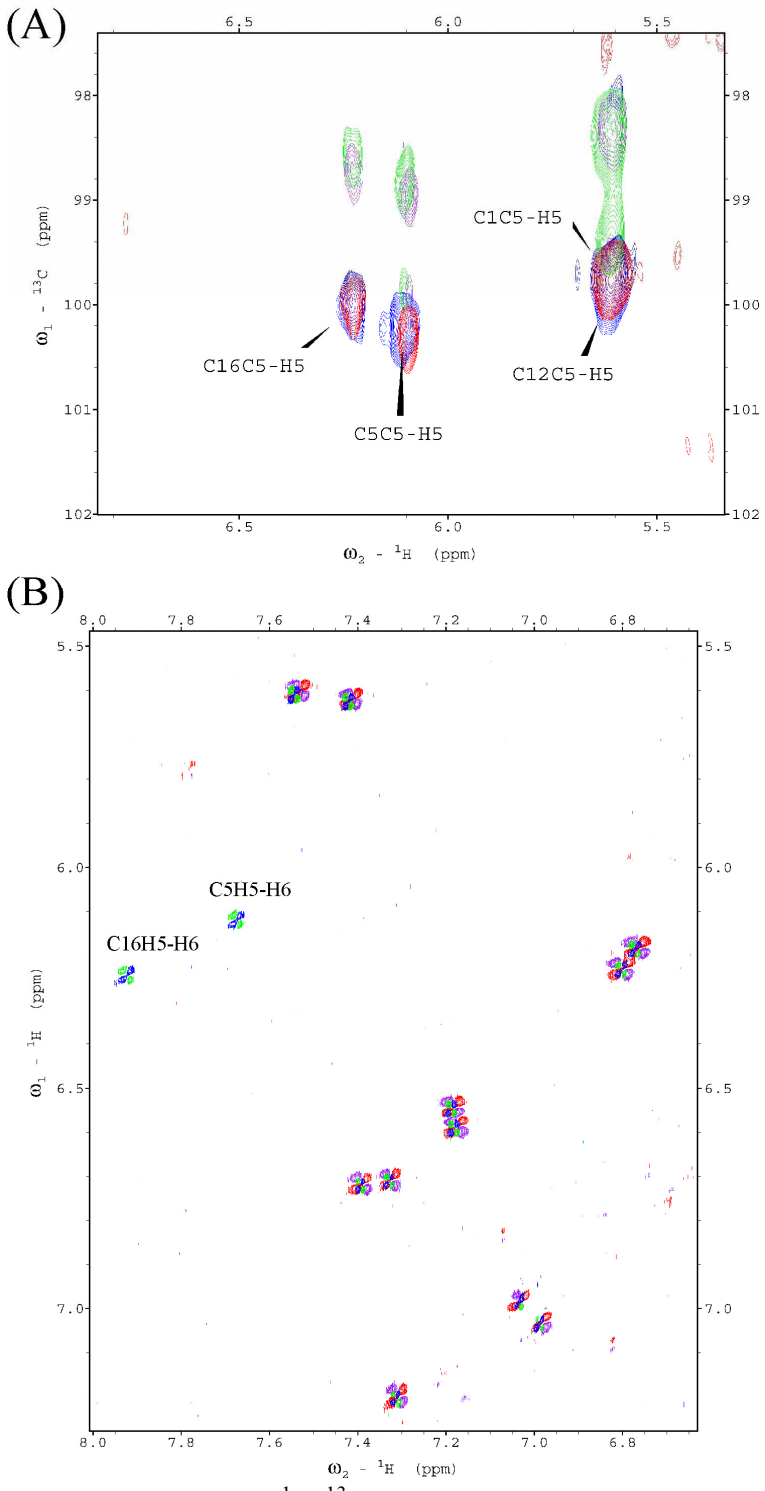



***Figure 5.*** The natural abundance $^{1}H$-$^{13}C$ IPAP HSQC spectra are shown (A). The DQF-COSY spectra are shown (B). The peaks colored in blue and green represent without pf1 phage and the peaks colored in red and purple represent with pf1 phage in both (A) and (B).

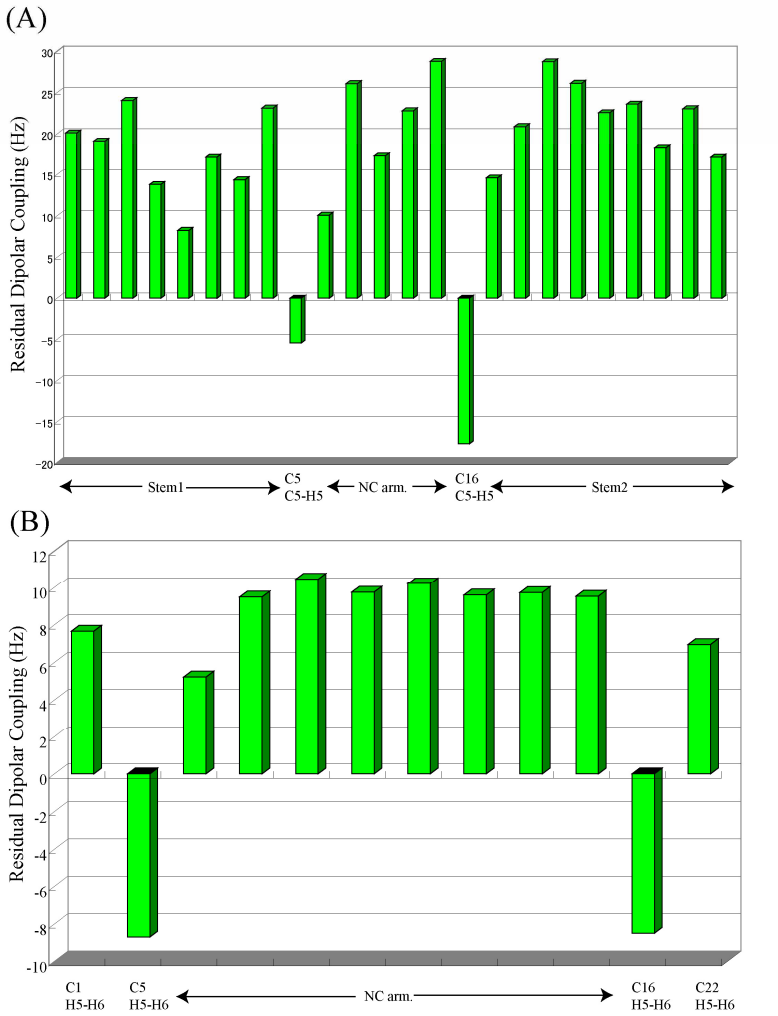



***Figure 6.*** The $^{1}D_{CH}$ and $D_{HH}$ values are represented in (A) and (B), respectively. Stem1 represents the B-form regions consist of C1-A4 and T19-G22. Stem2 represents the B-form regions consist of A8-G11 and C12-T15. NC arm. represents aromatic protons and carbons of the **NC** chromophores.

*MD simulation.* The **RDC** experiments showed the flipping out C5 and C16 had different

character compared with the other residues. The 2-ns molecular dynamics (**MD**) simulation of the **NC**-CGG complex was carried out. The NMR structure was used as the initial structure and the **MD** calculation was stable during the course of 2-ns (Figure 7(A)). Although there is no specific conformation for two flipping out C5 and C16 bases in the CGG/CGG triad in the initial structure, C5 and C16 immediately interact with the surface consist of the chromophores and alkyl linkers of **NC** until 400 ps (Figure 7(B)). This hydrophobic interaction basically preserved during 2-ns with several motion. The χangles of C5 and C16 kept 60 and -60 degree, respectively, in the most simulation course (Figure 7(B)). The decrease ofχangles represented the sliding on the hydrophobic surface. The obvious disruption of the hydrophobic interaction was observed once in C16 at 800 ps (Figure 7(B) asterisk), represented as the increase of theχangle.

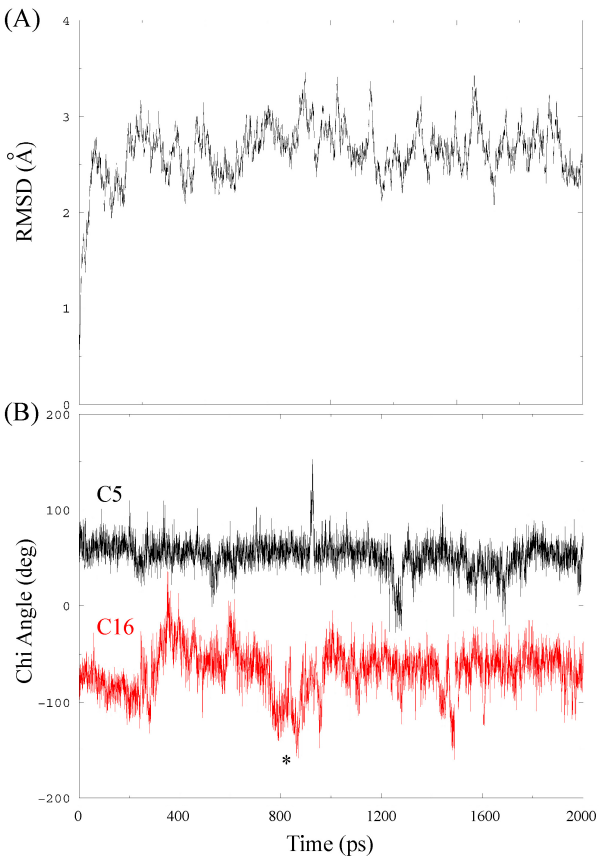



***Figure 7.*** The root-mean-squared deviation of the backbones of the 2:1 **NC**:DNA complex is shown in (A). The time dependence of the χangles of C5 and C16 is shown in (B).

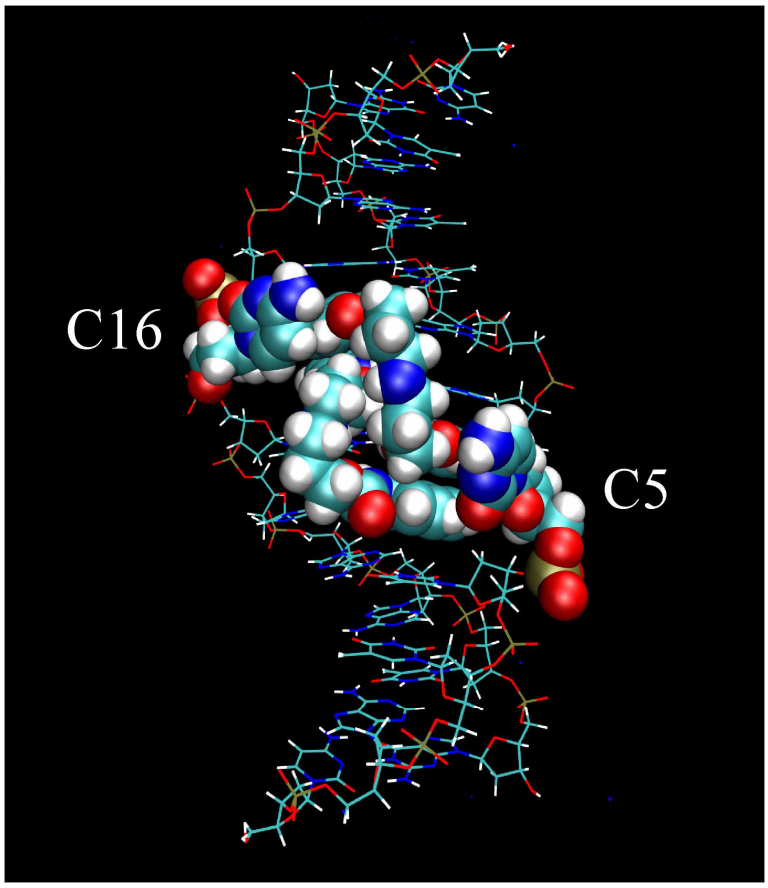



***Figure 8***. The average structure of the 2:1 **NC**:DNA complex during 2-ns **MD** simulation. The CPK models represent C5, C16 and two **NC** molecules.

## Discussion

*Structural feature.* The titration experiment showed the binding of **NC** to the CGG/CGG triad was the slow-exchange manner and observed the free state and the 1:2 complex state. It is consistent with the previous CSI-TOFF MASS study which showed no existence of a 1:1 complex [4]. These data may indicate that the binding mechanism of **NC** to the CGG/CGG triad is the co-operative manner in spite of the possibility of the 1:1 complex

based on the possible hydrogen bonding pattern. As a result of this study, the hydrophobic interaction among 4 methyls of **NC** and between **NC** and flipping out C5 and C16 also contribute stability of the 1:2 complex. The alkyl linkers of two **NC** also stabilize each other by the hydrophobic manner. It might be difficult to form these hydrophobic interactions in the supposed 1:1 complex.

*Specificity.* **NC** has much higher specificity to the CGG/CGG triad comparing with the reverse triad GGC/GGC. I also examined this reverse triad. The increase in $T_m$ of the GGC/GCC triad with **NC** was 6.7 ˚C, significantly lower to the CGG/CGG triad, 23.1 ˚C [4]. The titration experiment to 2.5 mM 5'-CTAAGGCAATG-3'/5'-CATTGGCTTAG-3' duplex was also carried out with monitoring the imino proton signals. During the titration, the broad peaks were obtained. The NOESY spectrum of the **NC**-GGC/GGC complex was also recorded, however the broad spectrum was obtained and it was impossible to father analysis. In addition, the free GGC/GCC triad was different from the free CGG/CGG triad. The additional peaks were observed at 9.5-11.0 ppm. The presence of these peaks suggested the additional hydrogen bonding were formed for the free GGC/GGC triad. This may contribute the extra stability of the free state of GGC/GGC, the instability of the **NC**-GGC/GGC complex. This result suggested that the intrinsic nature of the CGG/CGG

triad also contribute to the stability of the **NC**-CGG/CGG complex. The **MD** simulation showed the slight differences of the stabilities of the hydrophobic interactions for C5 and C16. The base of C5 also interacted with the base of A4 near H8 proton during the simulation. In the NOESY spectra, H6 of T15 resonance (6.1 ppm) showed high field shift comparing with the other bases (6.7-8.1 ppm). This asymmetric phenomenon at A4:T15 may also play the specific recognition to the CGG/CGG triad rather than the reverse GGC/GGC triad.

*Comparison.* In comparison with the **NA**-CAG complex previously determined (Figure 9(D)) [4], these two complexes have the similar basic recognition mechanisms based on the intercalation and hydrogen bonding. However there is the obvious difference that, in the **NC**-CGG/CGG complex (Figure 9(A)), the center guanosine was recognized by the naphtyridine chromophores which have the bulky methyl groups in the **NC**-CGG complex, in contrast to the only oxygen at the same position in the **NA**-CAG complex. In comparison with two 3D structures, the width between two strands at the intercalation site of the **NC**-CGG complex was narrower than the one of the **NA**-CAG complex. It is due to the hydrophobic packings of the methyl groups described above.

The two cytosines in the CGG/CGG triad were flipping out. The **NA**-CAG complex also

showed the flipping out of two cytsines in the CAG/CAG triad. Another difference between **NC** for the CGG/CGG triad and **NA** for the CAG/CAG triad is the length of the alkyl linker. In **NC**, the length of the alkyl linker is 9 which consist of 6 carbons, 1 nitrogen and 2 oxygens. In **NA**, the length of the alkyl linker is 5 which consist of 4 carbons, 1 nitrogen. Napththyridine dimer (**ND**) developed as the G.G mismatch recognition drug has the same linker design to **NA** one. I found that this **ND** also recognizes the CGG/CGG triad in spite of the low stability compared with **NC**. Thus, it is possible to access the contribution of the linker length by the comparison between the 3D structures of the **NC**-CGG/CGG complex and the **ND**-CGG/CGG complex. Followings the protocol applied to the **NC**-CGG complex, **ND** bound to the CGG/CGG triad with the 2:1 (**ND**:CGG/CGG) stoichiometry. The 3D structure was well converged (Figure 9(C)). The **RDC** experiment and **MD** simulation were also done and showed similar characters such as negative **RDC** values and the hydrophobic interactions of the flipping out cytosines. In comparison with the **NC**-CGG/CGG complex, there are two major differences. For both the **NC**-CGG/CGG and the **ND**-CGG/CGG complexes, the phosphodiester backbones were unwinding induced by the intercalation (Figure 9(C)). However the degree of the unwinding in the **NC**-CGG/CGG complex was greater than the **ND** (Figure 9(C) arrows). This feature was conformed by $^{1}H$-$^{31}P$ HSQC spectra. The downfield shifts of the phosphorous chemical shifts induced by

the **NC**-CGG/CGG complex were greater than the **ND**. Another feature was the loose packing of the methyl groups of the naphtyridine chromophores in the **ND**-CGG/CGG complex. This loose packing is unfavorable in the viewpoint of enthalpy. Thus, the short linker in **ND** restricts the complete methyl packing of the naphtyridine chromophores. This loose packing causes the instability of the **ND**-CGG/CGG complex compared with the **NC**-CGG/CGG complex. In the case of the **NA**-CAG complex, there is no need to pack the center chromophores because of the lack of the methyl groups at the same position. This is the reason why the length of the linker of **NA** is enough.

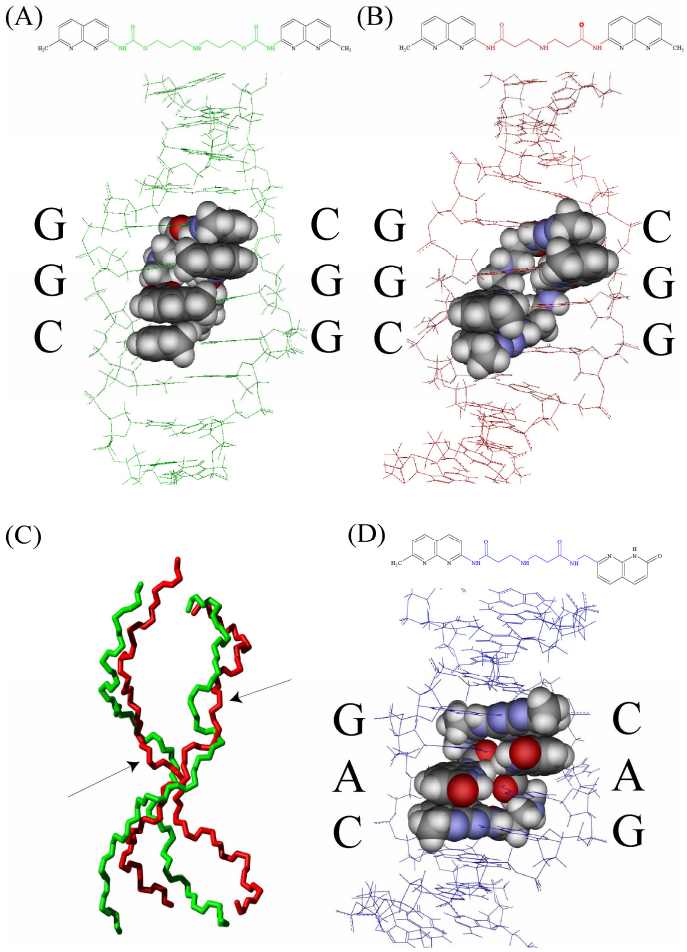



***Figure 9.*** The comparison with the CXG recognition drug series. The naphtyridine carbamate dimer (**NC**) with the CGG/CGG triad is shown in (A). The naphtyridine dimer (**ND**) with the CGG/CGG triad is shown in (B). The backbone comparison between the **NC**:DNA complex colored in green and the **ND**:DNA complex colored in red is shown in (C). The naphytiridine azaquinolone with the CAG/CAG triad is shown in (D).

*Cytosine flipping out.* The basic recognition mechanisms of **NC** was the enthalpy driven binding. In the viewpoint of the enthalpy, it seems that the two flipping out cytosines are unfavorable. As a result of the NMR structure determination, two cytosines, C5 and C16, were flipping out. It is consistent with the previous study that these two cytosine were cleavaged by hydroxylamine after the formation of the complex with **NC** [4]. In NOESY

spectrum, the inter-residual cross peaks for C5 and C16 were not observed much and were ambiguous. **RDC** data indicate that these flipping out cytosines were not stacking with the DNA bases and the **NC** chromophore, not flexible and had some comformation. If these bases were flexible, the **RDC** values are zero by the averaging of the random thermal motion. **MD** simulation showed that these two flipping out cytosines also contributed to the stabilization by the hydrophobic interaction. The two motion modes were observed. The bases of C5 and C16 sided on the hydrophobic surface. The base of C16 once leaved out the hydrophobic surface at 800 ps. These motions in the ns-ps time scale may cause the few ambiguous NOEs in the NOESY spectra.

*Conclusion.* In conclusion, I have determined the 3D structures of the **NC**-CGG complex by NMR. It revealed that the recognition mechanisms based on intercalation and hydrogen bonding. The methyl groups packing of the naphtyridine chromophores of **NC** contribute the stabilization. Two cytosines in the CGG/CGG triad were flipping out. As the result of the **RDC** experiments and **MD** simulation, these flipping out cytosines also stabilize the complex by the hydrophobic interaction with the hydrophobic surface consist of the alkyl linkers and the naphthyridine chromophores of **NC**. In the comparison with the study of the shorter linker complex, the **ND**-CGG/CGG The longer linker complex, **NC**-CGG/CGG is

stabilized by the methyl-group packing of the drug and the unwinding of the phosphordiester backbones.

## Acknowledgements

I thank Dr. T. Peng for discussion, Dr. Misima for the information of the **NA**-CAG/CAG complex and critical comments about NMR. This work was supported in part by Grants-in-Aid for 21st Century COE Research and Scientific Research from MEXT (the Japanese Ministry of Education, Culture, Sports, Science and Technology).

## References


[1] Pearson, C.E., Edamura, K.N. and Cleary, J.D. (2005) Nat. Rev. Genet. 6, 729-742.

[2] Gatchel, J.R. and Zoghbi, H.Y. (2005) Nat. Rev. Genet. 6, 743-755.

[3] Prospero, N.A.D. and Fischbeck, K.H. (2005) Nat. Rev. Genet. 6, 756-765.

[4] Peng, T. and Nakatani, K. (2005) Angew Chem Int Ed Engl 44, 7280-3.

[5] Nakatani, K. et al. (2005) Nat. Chem. Biol. 1, 39-43.

[6] Ono, A., Makita, T. and Tate, S. (1996) Magnetic Resonance in Chemistry 34, 40-46.

[7] Goddard, T.D. and Kneller, D.G., University of California, San Francisco.

[8] Liu, H., Spielmann, H.P., Ulyanov, N.B., Wemmer, D.E. and James, T.L. (1995) J Biomol NMR 6, 390-402.

[9] Kojima, C., Ono, A., Kainosho, M. and James, T.L. (1998) J Magn Reson 135, 310-33.

[10] Kojima, C., Ono, A., Kainosho, M. and James, T.L. (1999) J Magn Reson 136, 169-75.

[11] Brunger, A.T. et al. (1998) Acta Crystallogr D Biol Crystallogr 54 ( Pt 5), 905-21.

[12] Schwieters, C.D., Kuszewski, J.J., Tjandra, N. and Clore, G.M. (2003) J Magn Reson 160, 65-73.

[13] Hansen, M.R., Hansen, P. and Pardi, A. (1998) J Am Chem Soc 120, 11210-11211.

[14] Delaglio, F., Grzesiek, S., Vuister, G.W., Zhu, G., Pfeifer, J. and Bax, A. (1995) J Biomol NMR 6, 277-93.

[15] Delaglio, F., Wu, Z. and Bax, A. (2001) J Magn Reson 149, 276-81.

[16] Wu, Z., Delaglio, F., Tjandra, N., Zhurkin, V.B. and Bax, A. (2003) J Biomol NMR 26, 297-315.

[17] Pearlman, D.A. et al. (1995) Comput. Phys. Commun. 91, 1-41.

[18] Jorgensen, W.L. (1981) J Am Chem Soc 103, 335.