Studies on the Structure and Properties of the Non-canonical Base-pair C-U

A Doctoral Thesis
Submitted to Graduate School of Science
Osaka University

1998

Yoshiyuki TANAKA

CONTENTS

General Introduction

The aim of this thesis

The purpose of this thesis is to determine the structure of a non-canonical base-pair, cytidine-uridine (C-U) and to find the required conditions for the formation of such unstable base-pairs. It is also a goal of this study to establish the methods for the studies on unstable base-pairing.

Structure and physicochemical properties of nucleic acids

The nucleic acids can be classified into two groups by the chemical structure; one is a ribonucleic acid (RNA), the polymer of the ribonucleotides, which has a 2'-hydroxy group in the sugar ring, the other is deoxyribonucleic acid (DNA), the polymer of the deoxyribonucleotides, which does not have a 2'-hydroxy group in the sugar ring (Figure G.1). Five kinds of bases are mainly used at the base position of the nucleotides as presented in Figure G.2. They can make specific base-pairs, that is, the pairing through hydrogen bonds as drawn in Figure G.3 (Kyogoku *et al.*, 1966; Kyogoku *et al.*, 1967a,b; Thomas & Kyogoku, 1967; Rosenberg *et al.*, 1976; Seeman *et al.*, 1976). This types of the base-pair are most frequently observed in nature and called the Watson-Crick type base-pairs. In the polymeric structure, two polynucleotide strands bind to each other through the base-pair and make an anti-parallel double stranded helix so that a pseudo dyad axis is always presented between stacked base-pair (Watson & Crick, 1953; Rosenberg *et al.*, 1976; Seeman *et al.*, 1976). In cases of self-complementary base sequences, there exist a real dyad axis. There exists conformational polymorphism of the double helical structure, and they are called B-, A- and A'-forms (Figure G.4). An example of the complementary sequence and its three dimensional structure of each conformer is presented in Figure G.4. In the case of the genomic DNA, a complementary double stranded structure is maintained, but the most of RNA are expressed as single strands so that they must take various structures. The secondary structure motifs in RNA are presented in Figure G.5. The variety of motifs is closely related to the formation of the mismatches including the non-Watson-Crick base-pair, the target of this thesis. The mismatch means an irregular combination of bases different from the Watson-Crick base-pairs thus non-Watson-Crick base-pair means the mismatch which can form a certain base-pair.

It is well known that all of the living things utilize nucleic acids as genes, translational machines and so on. Especially, RNA has varieties of activities such as protein synthesis through the ribosomal RNA and self-splicing activities of introns in messenger RNA. These activities are indispensable for living things and known to be dependent on the structures. As mentioned above, the most of RNAs are present as single strands. Therefore, they include non-canonical base-pairs in the double helical region. That is the non-Watson-Crick type base-pairs or the mismatches. Accordingly, it is important to study their structures and characters in order to know the secondary and tertiary structure of RNAs. However available structural and nearest-neighbor (thermodynamic) parameters are limited to relatively stable ones like G-U, G-A and U-U (He *et al.*, 1991; SantaLucia *et al.*, 1991; Betzel *et al.*, 1994;

Katahira *et al.*, 1994; Allain *et al.*, 1995; Baeyens *et al.*, 1995; Jeffrey *et al.*, 1995; Baeyens *et al.*, 1996; Lietzke *et al.*, 1996; Wang *et al.*, 1996; Biswas *et al.*, 1997; Biswas & Sundaralingam, 1997). Unstable pairs such as C-U and C-C are not well studied, because the conformation of the mismatch site in the poly or oligonucleotides is irregular and the two state model analysis could not be applicable to the thermal denaturation of oligonucleotides. X-ray crystallography is a good way to determine the precise three dimensional structure of biological macromolecules. The solution NMR is good ways not only to determine the three dimensional structure in solution but also to monitor the environment of each site of the macromolecule. By using both methods, the structure and properties of unstable mismatches can be investigated.

X-ray crystallography and solution NMR as tools for analyzing the structures of biological macromolecules.

At this moment, X-ray crystallography is the most reliable method for determining the three dimensional structure of biological macromolecules. However most of the activities of biomolecules are expressed in the solution phase, therefore solution NMR is an indispensable method for the study of their properties in solution. As far as oligonucleotides, sometimes, their crystal structure does not represent the major conformer in the solution phase. It is better to use both methods to determine the three dimensional structure. To study the solution equilibrium or the other physicochemical properties in solution, the solution NMR is quite a powerful tool. The resolution of the signal by solution NMR is superior to the other spectroscopic methods for applications to biological macromolecules and thus the information of each site can be derived directly. Recent development of the site specific labeling techniques allowed us to pick up only the desired signals of the macromolecule.

Determination of the crystal structure of macromolecules

The structure factor from the X-ray diffraction data, $F(h,k,l)$, and the electron density at a certain coordinate, $\rho(x,y,z)$, in a unit cell are related by the equation below

$$\rho(x,y,z) = \frac{1}{V}\sum_h\sum_k\sum_l F(h,k,l)e^{-2\pi i(hx+ky+lz)} \tag{G.1}$$

$$= \frac{1}{V}\sum_h\sum_k\sum_l |F(h,k,l)|e^{-2\pi i(hx+ky+lz-\phi(h,k,l))} \tag{G.2}$$

$$F(h,k,l) = \int_v \rho(x,y,z)e^{2\pi i(hx+ky+lz)}dv \tag{G.3}$$

where x, y and z are the coordinates in a unit cell, h, k and l are the indice of the diffraction, V is the volume of the asymmetric unit, and ϕ is the phase of the diffraction. Although ϕs can not be known beforehand (phase problem), for biological macromolecules, ϕs are derived by means of the multiple isomorphous replacement (MIR), the molecular replacement and so on. Once phases are determined, the electron density can be calculated and approximate coordinates can be derived. To determine more precise coordinates, refinement is required. The refinement of the coordinates are achieved by decreasing the crystallographic residual.

In general, R-factor is used to monitor the extent of refinements.

$$\text{"R-factor "} = \frac{\Sigma\|F_O| - |F_C\|}{\Sigma|F_O|} \tag{G.4}$$

In the case of the low molecular weight organic compound less than approximately 1000, diffraction data can usually be recorded over the 1.0 Å resolution range. They can be refined by diffraction data only with the least square method. In the case of biological macromolecules, however, its crystal does not diffract to 1.0 Å in general, because of the high mosaicity of the crystal. Therefore, diffraction data are not enough to determine the coordinates. To compensate the missing data and keep the chemical structure correctly, restraints on covalent bonds such as bond length, bond angle, chirality and planarity of the aromatic group are introduced. To avoid steric hindrance, van der Waals potential is also included. As a result, the target function of the least square is formulated as below

$$E_{\text{total}} = E_{\text{X-ray}} + E_{\text{bond-length}} + E_{\text{bond-angle}} + E_{\text{chiral}} + E_{\text{planar}} + E_{\text{vdw}} \tag{G.5}$$

where $E_{\text{X-ray}}$ is the pseudo potential of crystallographic residual, $E_{\text{bond-length}}$, $E_{\text{bond-angle}}$, E_{chiral}, E_{planar} and E_{vdw} are constraints for a bond-length, bond-angle, chirality, planarity and van der Waals potential.

Structure calculation by using NMR data

Structure calculation by NMR is carried out in three steps. At first, most of the proton signals is assigned, secondly, NOE (nuclear Overhauser effect) cross peaks must be quantitated from two-dimensional NOESY (nuclear Overhauser enhancement spectroscopy) spectrum; finally, NOE based structure calculation is performed.

Assignment of proton signals and the structure calculation are performed by means of the NOE based method. In the NOESY spectrum, NOE cross peak can be observed only when two protons are located in short distance, less than 5.0 Å in practice. For protons A and B of a macromolecule, the cross relaxation rate, k, is given by

$$k = (1/2)\gamma^4 (h/2\pi)^2 \tau_c r^{-6} \tag{G.6}$$

where γ is magnetogyric ratio of proton, h is Plank constant, τ_c is the overall correlation time and r is the distance between protons A and B. When the distance is greater than 5.0 Å, k is approximately zero. NOE can not be observed.

The signal assignment of nucleic acids is achieved by the NOE based sequential assignment. In the double helical RNA, H6 in pyrimidine base or H8 in purine base give intra-residue and inter-residue NOE cross peaks to the protons in the sugar moiety, H1', H2" and H3' as shown in Figure G.6 because their distances are less than 5.0 Å. A part of the signals is assigned by COSY (correlation spectroscopy) spectrum which uses scalar coupling between protons.

NOE cross peak volume is proportional to the cross relaxation rate, k.

$$I = k * t_m \qquad (\sum^{N} k_n * t_m << 1,\ \frac{t_m}{T_1^A} << 1) \tag{G.7}$$

where T_1^A is the time constant of the longitudinal relaxation of the proton A and t_m is the

mixing time of NOESY spectra. Therefore, the relation between NOE cross peak volume, I, and the distance, r is given by

$$I = (1/2) t_m \gamma^4 (h/2\pi)^2 \tau_c r^{-6} \quad \text{(G.8)}$$

In the equation (G.8), only τc is unknown. The overall correlation time, τc can calculated as follows. The NOE data include the data on the protons whose distance is fixed and already known. Therefore, τc can be calculated from its peak volume and distance. Once τc is derived, distances can be calculated from the equation (G.8). This equation (G.8) assumes no spin diffusion which causes an indirect NOE cross peak. Therefore the relationship between the NOE peak volume and distance are not strictly restrained. To avoid overestimation, one should calculated the distances from the NOE cross peak volumes with relatively large tolerances. In practice, the NOE cross peaks are classified into about three groups by their peak volumes and converted to the distance restraints by comparing with a certain cross peak as an internal standard like the cross peak of H5-H6 in a pyrimidine base ring whose distance is 2.49 Å. Each group is translated to a specific upper limit of the distance including an appropriate tolerance. As a result, most of the cross peaks have large tolerances.

Then the structure calculation is performed to satisfy the obtained distance constraints. As is the case of NOE based structure calculation of macromolecules, number of NOE distance constraints are not enough to determine the coordinate of each atom. Therefore, stereo-chemical restraints are introduced. NOE based structure calculation of a protein molecule is performed by using the equation (G.9) as a target function

$$E_{\text{total}} = E_{\text{NOE}} + E_{\text{bond-length}} + E_{\text{bond-angle}} + E_{\text{chiral}} + E_{\text{planar}} + E_{\text{vdw}} \quad \text{(G.9)}$$

where E_{NOE} is the NMR derived distance constraints.

However for nucleic acids, the proton density is lower than that of a protein. Therefore, additional constraints for base-pairing are introduced. They consist of the constraints of the hydrogen-bonding distances and the planarity of the base-pair, $E_{\text{base-pair}}$. The target function of the least square of nucleic acids is given by

$$E_{\text{total}} = E_{\text{NOE}} + E_{\text{bond-length}} + E_{\text{bond-angle}} + E_{\text{chiral}} + E_{\text{planar}} + E_{\text{vdw}} + E_{\text{base-pair}} \quad \text{(G.10)}$$

Preparation of stable isotope labeled nucleic acids and its advantage for the structure analysis with NMR.

There are two ways for the preparation of stable isotope labeled RNA. One is the enzymatic method using labeled nucleoside triphosphates as starting materials. The other is the synthetic method using the phosphoramidite form of the labeled nucleosides as the materials.

The enzymatic method is convenient for preparing uniformly labeled oligonucleotides, that is, all of the nitrogen and carbon atoms are labeled throughout the sequence. The enzymatic method consists of five steps. 1) The culture of the bacteria in the medium with $^{15}NH_4Cl$, ^{13}C-labeled glucose or both as a nitrogen and carbon sources. 2) Fractionation of

RNA from bacterial cells. 3) Cleavage of RNA to nucleotides by a nuclease. Fractionation of each kind of nucleotide if necessary. 4) Exchanging nucleoside to nucleoside triphosphates (NTP) by kinases. 5) Oligonucleotide synthesis by T7 RNA polymerase using NTP as substrates. As is obvious from this protocol, labeling at a single nucleotide is very difficult. Only nucleotide specific labeling can be achieved.

On the contrary to this, position specific labeling can be achieved easily by chemical synthesis. Another advantage of the synthetic method is point labeling where a specific atom is labeled in a nucleotide can be done such as labeling only at the amino nitrogen of a cytidine residue in an oligonucleotide. Once the phosphoramidite form of a nucleotide can be prepared, the synthesis of the oligonucleotide is achieved by the solid phase phosphoramidite method. The solid phase phosphoramidite method is summarized in Figure G.7. The synthesis of the cytidine labeled at amino-nitrogen used in this study is described in Chapter 2.

Advantages of the point labeling in the structure analyses with NMR is as following. Unambiguous signal assignment can be done. Because only the wanted signal is observed without disturbing by many other signals of a macromolecule. When two or more signals are observed even if a single site is labeled, one can immediately the existence of multiple conformations. For the study of the physicochemical properties and structures of macromolecules, point labeling is a powerful tool.

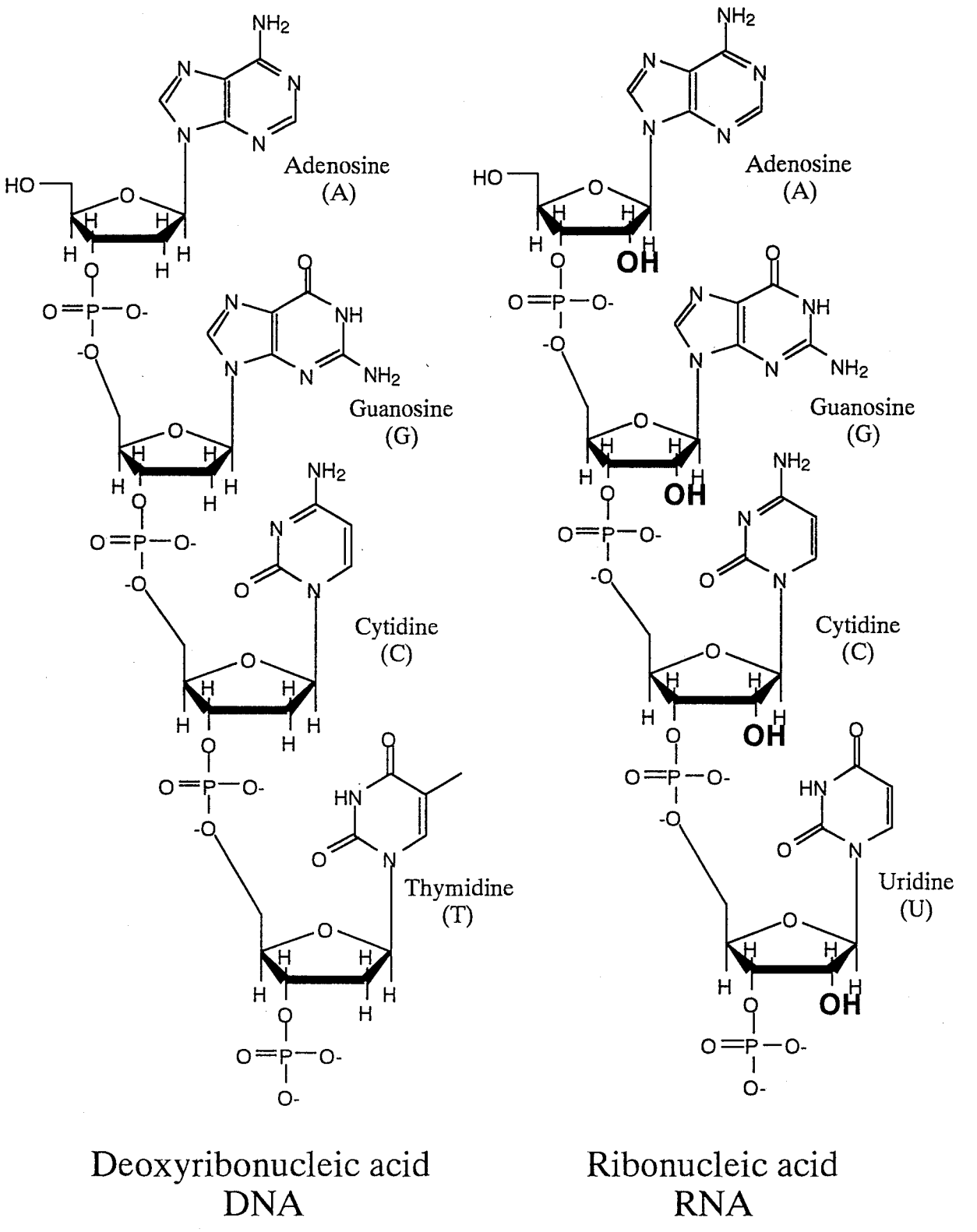


Figure G.1 RNA is a polymer of ribonucleotides. DNA is a polymer of deoxyribonucleotides. In DNA, thymidine is used instead of uridine.

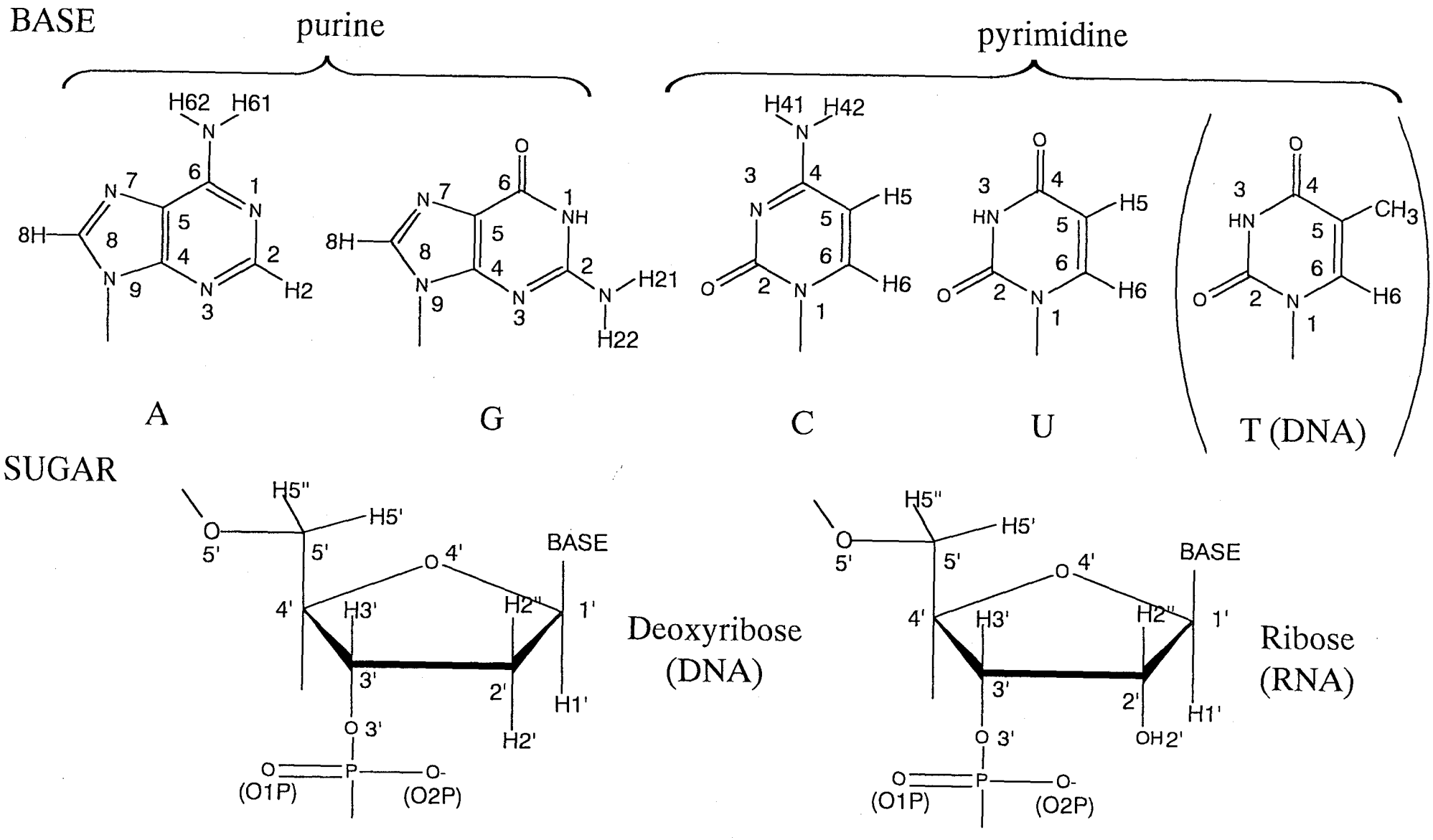
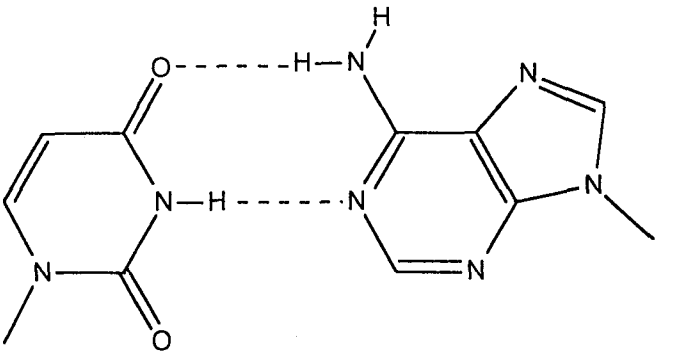
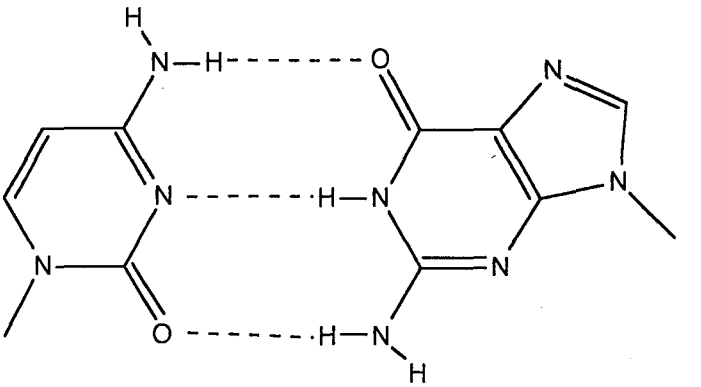


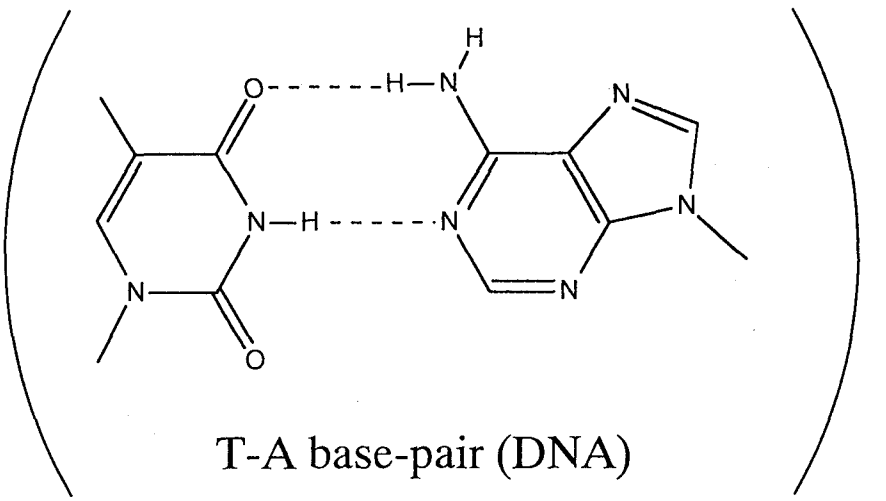
Figure G.2 Definition of the numbering system.



U-A base-pair



C-G base-pair



T-A base-pair (DNA)

Figure G.3 Configuration of the Watson-Crick base-pairs. Only the protons which participate in hydrogen bonds are dipicted with other heavy atoms. The broken lines represent hydrogen bonds.

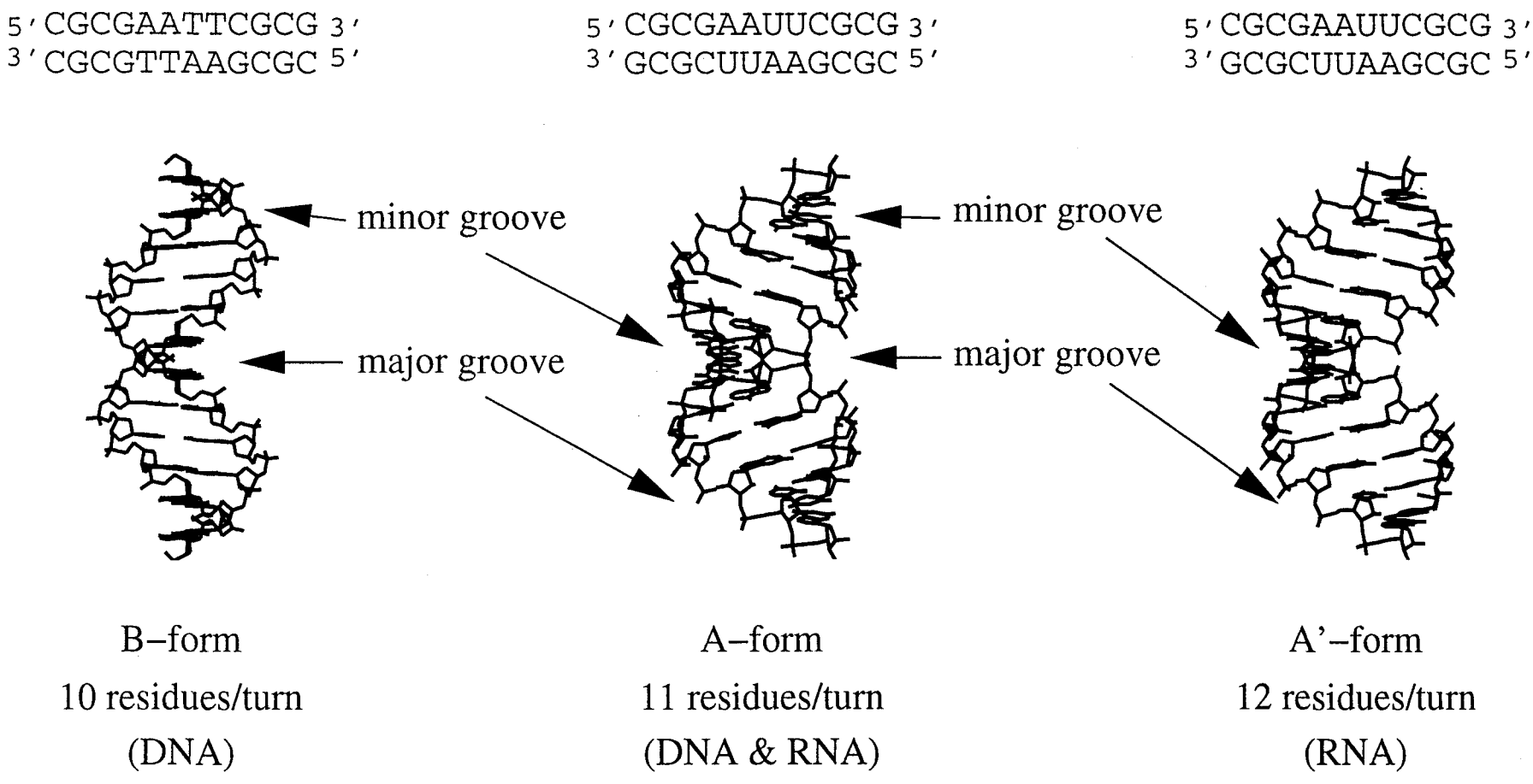


Figure G.4 Conformational polymorphism of the nucleic acid duplexes. All the sequences contains twelve nucleotides in each strand. In the B–form, uridines are replaced by tymidines, because only DNA can adopt the B–form conformer.

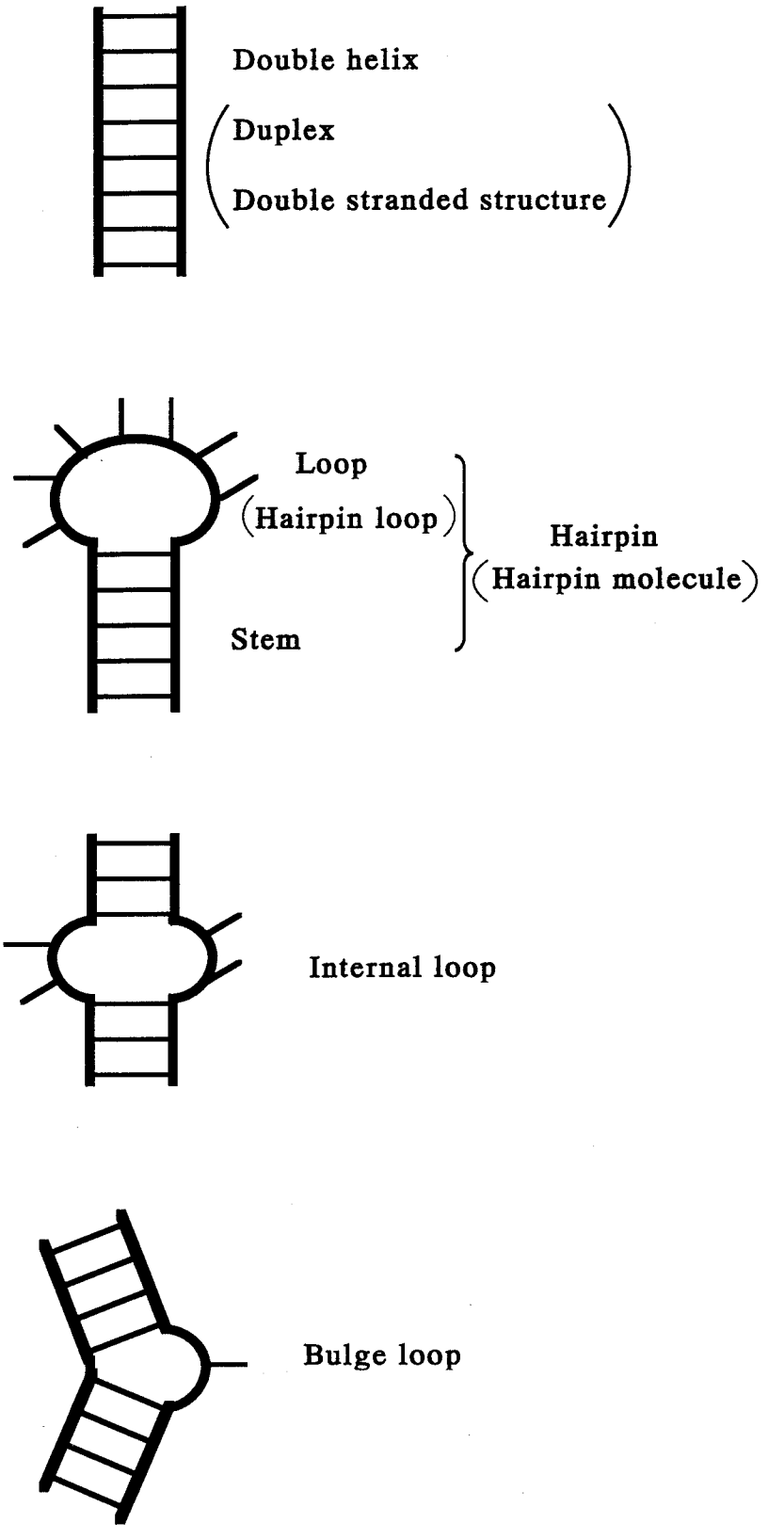


Figure G.5 Secondary structure motifs of RNA are presented. Different naming are also given.

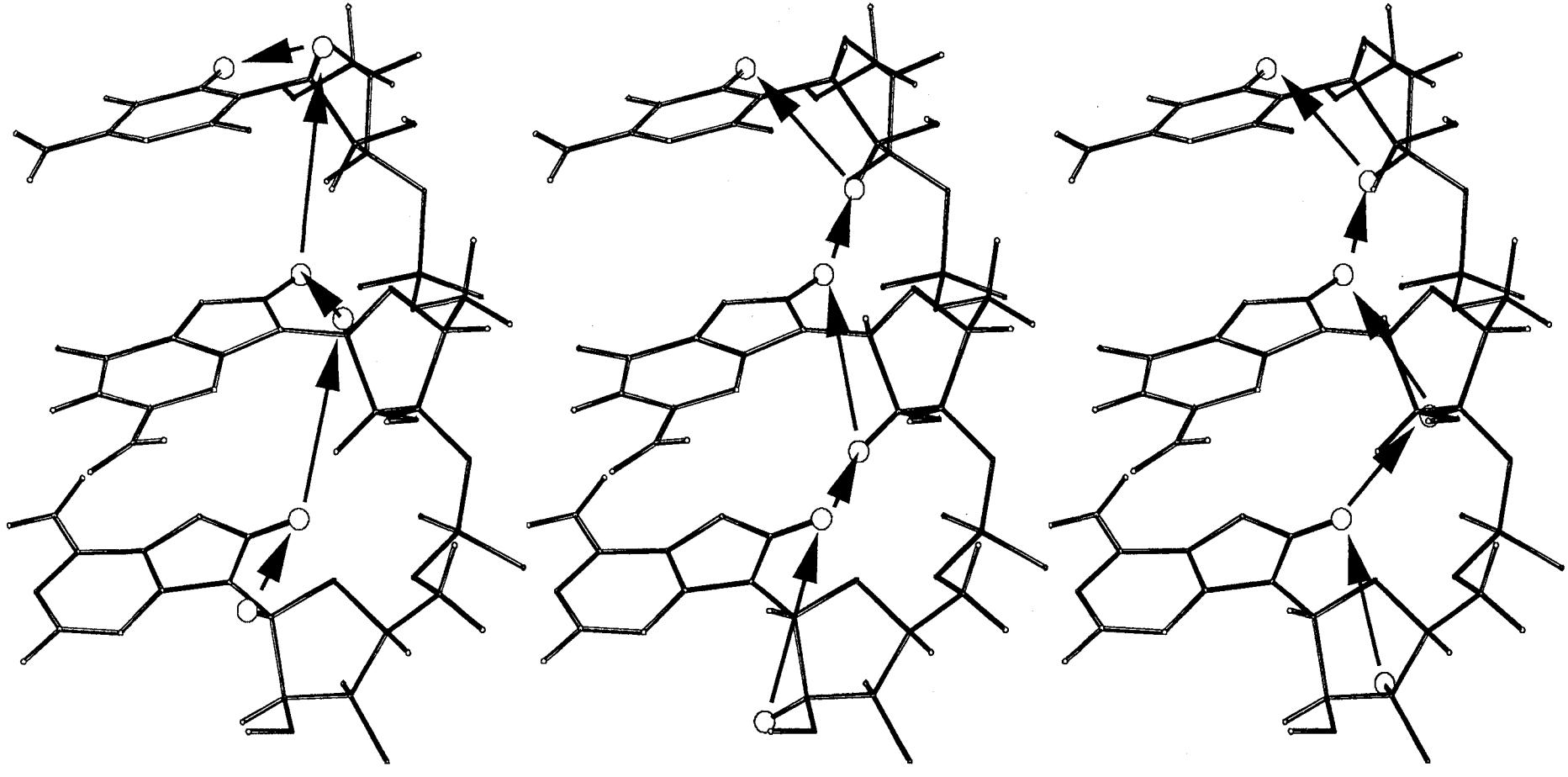


Figure G.6 Assignment procedure based on sequential NOEs. Open circles represent the protons which can be assigned from NOEs. H6 and H8 are the protons in purine and pyrimidine bases. H1', H2'', H3' are the protons in the sugar rings. The left, center and right figures show the sequencial NOEs for H1', H2'' and H3', respectively.

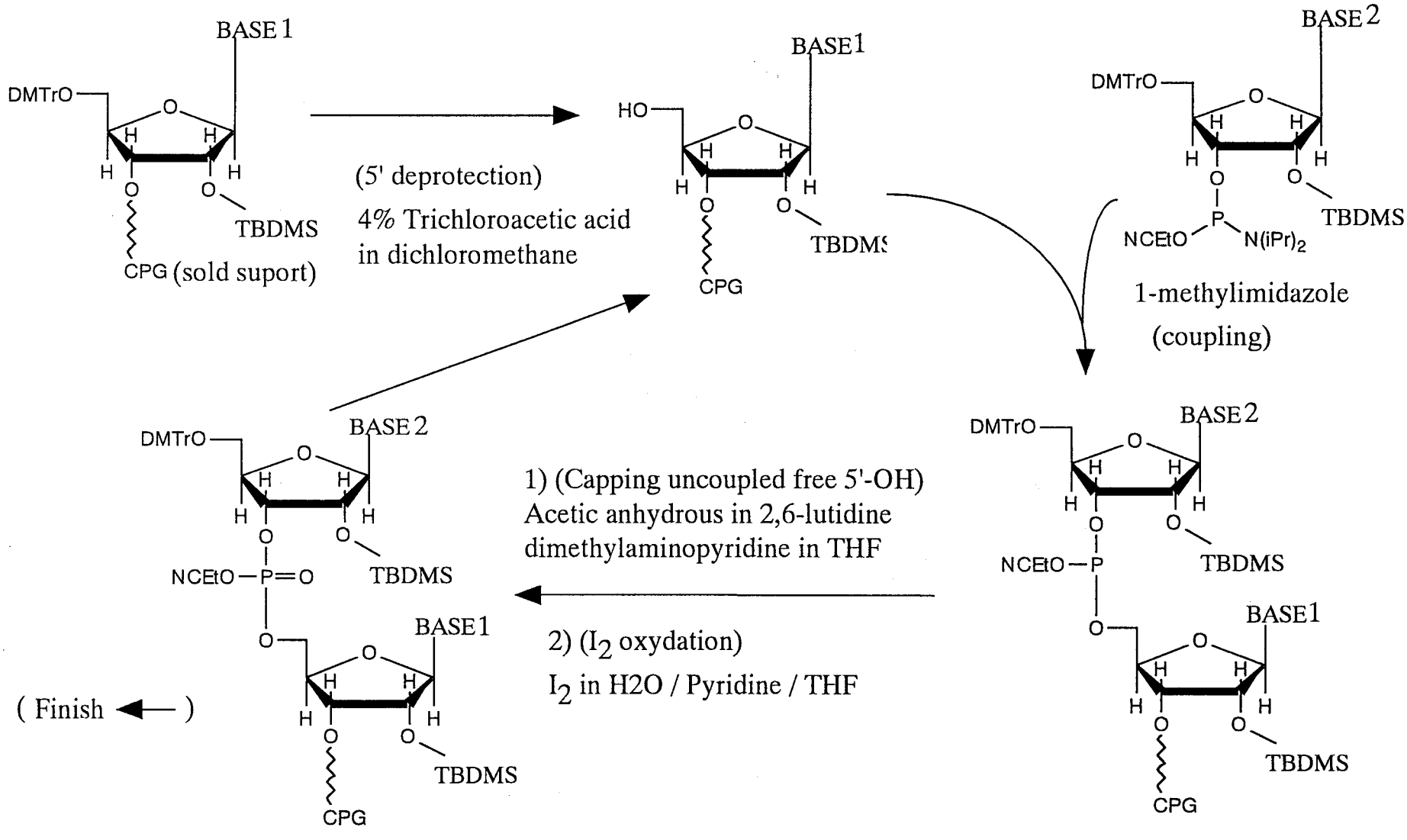


Figure G.7 The synthesis cycle of the solid-phase phosphoramidite method. CPG represents the solid support of controled-pore-glass resin. DMTr, dimethoxytrityl group, is a protecting group for the 5'-hydroxy group. TBDMS, *tert*-butyldimethlsillyl group, is a protecting group for the 2'-hydroxy group.

Chapter 1

The crystal structure of the C-U pair in r(UGAGCUUCGGCUC) and studies on the polymorphism of global helicity of the RNA duplexes

Abstract

The crystal structure of the tridecanucleotide, r(UGAGCUUCGGCUC), was solved using diffraction data up to 1.8Å resolution. Two molecules of the tridecamer, related by the crystallographic two-fold axis, form a right-handed double helix. The double helix includes four successive non-Watson-Crick base-pairs in the middle of the helix and a 5' uridine residue at the dangling end. Each double helix is piled up along the crystallographic c-axis to form a pseudo infinite double helix through the intermolecular U-U pair. In the major groove of successive G-U and C-U pairs, there is a hydrogen bond network of water molecules; the water molecules and phosphate-oxygen are linked together via hydrogen bonds in a cyclic manner. In this paper, we also studied structural polymorphism of an RNA double helix, using the single crystal structures of RNA oligomers previously reported as well. On detailed analysis of structure parameters, we found how to classify A-RNA and A'-RNA, the major conformers of an RNA double helix. When the mean major groove widths were plotted against mean inclination angles, two clearly separated clusters appeared. Each cluster contained canonical A-RNA and A'-RNA, respectively, and the tridecamer belongs to the A'-RNA group. Although the RNA double helix has been thought to be uniform and rigid recently, this finding implies that the conformation of the RNA double helix is not restricted in A-RNA and may not be so rigid as to prevent any fluctuation at all.

Introduction

In this chapter, the crystal structure of the tridecamer which includes C-U pairs, the target of this thesis, in the non-Watson-Crick base-pairs tract is described. In the former part, overall structure of the tridecamer and the local geometry around the non-Watson-Crick base-pairs including a hydration pattern are described. In the latter half of this chapter, classification of conformation of double helical RNA is discribed as well.

The oligomers including the sequence CUUCGG are known to form a thermally extraordinarily stable hairpin loop structure in solution (Tuerk *et al.*,1988; Sakata *et al.*, 1990; Cheong *et al.*, 1990; Antao & Tinoco, 1992). To investigate the relationship between the thermal stability and three-dimensional structure of the "UUCG" loop, many attempts have been carried out to crystallize the oligomers which contain this loop sequence (Fujii *et al.*, 1991; Holbrook *et al.*, 1991; Fujii *et al.*, 1992; Baeyens *et al.*, 1994; Cruse *et al.*, 1994). However all of these attempts including our case resulted in a formation of a double-stranded helical structure with internal loop instead of a hairpin loop structure. It was confirmed by NMR that this hairpin loop structure really exists in solution (Sakata *et al.*, 1990; Cheong *et al.*, 1990; Varani *et al.*, 1991; Allain and Varani, 1995b; Allain and Varani, 1997). The crystal structures, however, revealed that such a long internal loop by four bases can forms non-Watson-Crick base-pair: Two G-U and two C-U pairs in the internal loop maintain a right-handed double helical conformation and also pyrimidine-pyrimidine base-pair such as C-U can exist. Although the arrangement of the G-U and C-U pairs are similar to those which were solved previously (Holbrook *et al.*, 1991; Cruse *et al.*, 1994), it is worthy to reinvestigate their detailed geometries and hydration patterns because of different surrounding environments such as unit cell dimensions.

Numerous attempts have been concentrated to clarify the conformational polymorphism of double helical polynucleotides by using the X-ray fiber diffraction technique (Watson & Crick, 1953; Fuller *et al.*, 1965; Tomita & Rich, 1964; Arnott *et al.*, 1972) and solid state ^{31}P NMR of RNA fiber (Shindo *et al.*, 1985). In respect of RNA, it was found that there exists two major right-handed conformers; one is an A-RNA which has eleven nucleotides in one helical pitch, the other is an A'-RNA which has twelve nucleotides in a helical pitch (Arnott *et al.*, 1972). Another characteristic conformational feature of A'-RNA is to have a significantly wider major groove and lower value in inclination angle than A-RNA. However, little attention has been given to A'-RNA in spite of having a different feature compared with that of A-RNA. The reason is probably due to the facts shown below. First, the root mean square deviation (rmsd) value between A-RNA and A'-RNA is not so large at the level of one turn double helix, so A'-RNA could be interpreted as just a variation of A-RNA. Secondly, most of the structures of the RNA oligomers obtained from single crystals (Dock-Bregeon *et al.*, 1988; Dock-Bregeon *et al.*, 1989; Schindelin *et al.*, 1995; Baeyens *et al.*, 1995) showed characteristic conformational features of A-RNA but not those of A'-RNA. However, once we consider the RNA-RNA or the RNA-protein interaction, it is very important whether the

major groove is wide or narrow, and whether something else but an A-RNA conformer exists or not. It is a good chance to compare the structures of RNA double helix from a view point of conformational polymorphism, because of increasing crystal structure data of the RNA oligomers.

Materials and Methods

Crystallization of r(UGAGCUUCGGCUC)

The sample for crystallization was synthesized by solid-phase phosphoramidite method and purified as described by Sakata *et al.* (1990). The numbering of the bases and base-pair scheme in crystal is presented in Figure 1.1. Crystals were grown in the sitting drop vessel by vapor diffusion technique at 10℃ from a solution of 0.5mM RNA tridecamer, r(UGAGCUUCGGCUC), 6.7mM sodium cacodylate buffer (pH6.5), 1.3mM sodium phosphate, 13.3mM NaCl, 3.3mM $BaCl_2$, 3.7mM spermidine·3HCl, 0.8% polyethylene glycol #400 (PEG#400) against 4% PEG#400 as a reservoir. Crystals belong to the monoclinic space group, C2 with one tridecamer per asymmetric unit and the cell dimensions : a=38.49Å, b=32.30Å, c=38.76Å, β=117.56°.

Data collection

Single well formed crystal (1.0×1.0×0.2mm) was mounted and sealed in a glass capillary and intensity data up to 1.8Å resolution were collected. Number of the observed reflections and the reflections above 1 σ(F) is 3556 and 2742, respectively. Completeness is 88% and 68%, respectively. Diffraction intensities were recorded in $2\theta/\varpi$ scan mode with scan speed of 6°/min. and scan width of 1.2° on a Rigaku AFC5R diffractometer using graphite monochromated Cu Kα radiation (λ=1.5418Å). The unit cell dimensions were refined by 2θ values of 20 reflections in the range of 16°<θ<22°.

Structure determination

The crystal structure was solved by the combination of the molecular replacement and the isomorphous replacement method, using a single stranded A-form dodecamer : r(GAGCUUCGGCUC) where uridine(U1) at 5' end was omitted at first stage because of its flexible conformation. The high peaks with 3.0Å interval along w-axis in Patterson map suggests that the bases are stacked to each other, and the oligomer and the one related by crystallographic 2-fold axis form a double helix with its helical axis approximately parallel to the c-axis. From this assumption, two polarities are acceptable (Figure 1.2(a)(b)). One is the polarity where the minor groove of the middle part of the double helix was located in the minus region of b-axis (Figure 1.2(a)), the other is the opposite direction (Figure 1.2(b)). Then we performed the rigid body refinement to these two models without any rotation or translation search. After a few stage of rigid body refinement using 679 reflections in the

range of 10.0 to 3.0Å above 2σ(F), direction (a) and direction (b) showed a crystallographic residual, R, of 33% and 44%, respectively. So the direction (a) was selected as a correct orientation because of a significant lower value of R-factor. Then we calculated 2|Fo|-|Fc| map of whole molecule, and it was confirmed that the model from the direction (a) is well fitted to the electron density map. Further refinements were performed to the model of direction (a) and the resolution was gradually extended to 1.8Å with a refinement of B-factor except for U1. Even at this stage, 2|Fo|-|Fc| map indicated the disordered electron density around U1. The disordered U1 residues around the crystallographic two-fold axis are incorporated into pseudo continuous double helix. Then we assumed that the disordered U1 structure is caused by an asymmetric base-pairs around the crystallographic two-fold axis with the occupancy of each conformer of 50%. Then the models of disordered U1 satisfying the electron density were searched, and then U1 residue was refined by simulated annealing with the other part fixed, followed by a water molecules incorporation. We named the U1 conformer in the major groove side as U1a and the other as U1b. Final structure contains one RNA oligomer and 49 water molecules per asymmetric unit. The final R-factor is 16.7% for 2720 reflections from 8.0 to 1.8Å with $|F|>1\sigma(F)$. From the Luzzati plot (Luzzati, 1952), the estimated error in atomic coordinates of final model is 0.15 - 0.20Å. Deviations from a standard value of bond length and bond angle are 0.015Å and 3.4°, respectively. All structure calculation was done by X-PLOR version 3.1 (Brünger, 1992) and QUANTA96 at the Research Center for Protein Engineering for a graphics.

Calculation of helical parameters and torsion angles

Helical parameters and major groove widths are calculated by commercially purchased program Excel97 with some modifications. At first, target double helices were fitted with a canonical A-form RNA whose helical axis was parallel to the z-axis by X-PLOR, then we regarded z-axis as the helical axis. We followed the nomenclature defined by Diekmann *et al.*, 1989. But we add "*" after the parameter names (e.g., inclination*) to distinguish the parameter calculated by another program because of the simplification of the calculation method. The definitions of the all structure parameters are presented in Figure 1.3. Before the calculation of helical parameters, the center of base-pair (m) was defined as the mid-point of C8 and C6 atoms for a purine-pyrimidine base-pair or that of C6 and C6 atoms for pyrimidine-pyrimidine base-pair and the long axis vector ($\boldsymbol{l}$) of base-pair was defined as the vector from C8 to C6 atoms for a purine-pyrimidine base-pair or that from C6 to C6 atoms for pyrimidine-pyrimidine base-pair. X-displacement* is a distance between helical axis and (m): the mid-point of base-pair (Figure 1.3(a)). Inclination* is the angle between x-y plane and ($\boldsymbol{l}$): the long axis vector of base-pair (Figure 1.3(b)). Twist* is the angle between projections of ($\boldsymbol{l}$)s of the neighboring base-pair to the x-y plane (Figure 1.3(c)). Rise* is the difference in the z-coordinates of (m)s of the neighboring base-pairs (Figure 1.3(d)). Before the calculation of the major groove width and minor groove width, we defined P-P major and P-P minor in Figure 1.3(e).

The trans zigzag line and character "P" represents backbone and phosphorous atom, respectively. In A-RNA or A'-RNA, the nearest phosphorous atoms passing over the major groove and the minor groove are linked by solid line and their linear distances are P-P major and P-P minor, respectively. In B-DNA, P-P major and P-P minor are represented in the dashed lines. The major groove width and minor groove width were derived by subtraction 5.8Å, the sum of the van der Waals radii of two phosphorous atoms, from P-P major and P-P minor, respectively. The mean values and the standard deviations of each structure parameter are calculated in terms of each oligomer (Table 1.3(a)). The coordinates used for the calculation of the structure parameters except for the tridecamer of this work are taken from PDB; where ID number is 1RNA for $r(UUAUAUAUAUAUAA)_2$, 255D for $r(GGACUUCGGUCC)_2$, 205D for $r(GGACUUUGGUCC)_2$, 1SDR for r(UAAGGAGGUGAU) • r(AUCACCUCCUUA) (Dock-Bregeon *et al.*, 1989; Holbrook *et al.*, 1991; Baeyens *et al.*, 1995; Schindelins *et al.*, 1995). Torsion angles were calculated by X-PLOR as listed in Table 1.2.

Cluster analysis of the RNA oligomers

The mean structure parameters of all the oligomers were averaged in terms of the structure parameter then the standard deviation of each structure parameter was also calculated to standardize each structure parameter. The standardized structure parameters of the oligomers are listed in Table 1.3(b). Then pair-wise Euclidean distances between the oligomers were calculated using the standardized structure parameters (Table 1.4.). They were clusterized by means of the nearest neighbor method. Schematic diagram of the cluster analysis is presented in Figure 1.9.

Results

Overall structure

Crystal structure of the tridecamer, r(UGAGCUUCGGCUC), contains one strand of the tridecamer and 49 water molecules per asymmetric unit. The 5' terminal uridine residue takes two conformations with the occupancy of 0.5, because of the statistical disorder around the crystallographic two-fold axis. The sequence with the number of the bases and the interactions between the symmetrically related molecules are presented in Figure 1.1. The final structure gives an R-factor of 16.7% for 2720 reflections from 8.0 to 1.8Å with $|F|>1\sigma(F)$. From the Luzzati plot (Luzzati, 1952), the estimated error in atomic coordinates is 0.15 - 0.20Å. Deviations from a standard value of bond length and bond angle are 0.015Å and 3.4°, respectively.

Two strands of the tridecamer related by the crystallographic two-fold axis form a double stranded structure (Figures 1.1 & 1.4), as previously reported (Holbrook *et al.*, 1991; Fujii *et al.*, 1992; Cruse *et al.*, 1994), instead of a monomeric hairpin loop structure. All through the double helix, right-handed structure is kept not only in Watson-Crick base-pair portion but also in non-Watson-Crick base-pair portion (Figures 1.4). Each double helices are

piled up along the crystallographic c-axis to make a pseudo continuous double helix through the U-U pair (Figure 1.1). At the middle region of the helix, four successive non-Watson-Crick base-pairs (two G-U and two C-U pairs) are formed.

Temperature factors are listed in Table 1.1. It is interesting that there is cooperativities between the temperature factors of the base-paired residues, for example G2 : C13, A3 : U12 and so on, although any restraints of the B-factor between base-pair were not introduced during the course of refinement. However B-factor is so sensitive to the stacking interaction and packing interaction that we could not determine whether this cooperativity is an artifact or the result of base-pairing.

Molecular geometries of four successive non-Watson-Crick base-pairs

The 2|Fo|-|Fc| electron density map of the G-U pair is presented with its wire model in Figure 1.5(a). As seen in tRNAs and RNA oligomers (Holbrook *et al.*, 1991; Cruse *et al.*, 1994; Biswas *et al.*, 1997; Biswas and Sundaralingam, 1997), the G-U pair shows a wobble-type geometry and the electron density map clearly shows the existence of a water molecule (W1) which bridges the G-U pair in the major groove with hydrogen bonds to O6 of G9* (2.8Å) and to O4 of U6 (2.8Å) where "*" indicates the symmetry related molecule. This kinds of water molecules were often observed in the G-T pairs introduced in the A-, B- and Z-form DNA in crystal (Kneale *et al.*, 1985; Hunter *et al.*, 1986; Hunter *et al.*, 1987; Ho *et al.*, 1985). In RNA, the examples of the G-U bridging water molecules in the major groove were reported recently (Biswas *et al.*, 1997; Biswas and Sundaralingam, 1997). In the minor groove, there is also a water molecule (W2) which makes hydrogen bonds between O2' of U6 (2.7Å) and N2 of G9* (2.8Å). These two water molecules occupy the hydrogen bonding sites which lack the counter part of hydrogen bonds, and then they make the G-U pair stable enough to maintain the wobble-type pair.

The 2|Fo|-|Fc| electron density map of the C-U pair with its wire model (Figure 1.5(b)) indicates that the geometry of the C-U pair is the same as that found in cases of dodecamer and nonamer (Holbrook *et al.*, 1991; Cruse *et al.*, 1994), and in addition to the direct hydrogen bond between O4 of U7 and N4 of C8* (2.8Å), a water molecule (W3) forms two hydrogen bonds to N3 of U7 and N3 of C8* with distances of 2.7Å, and 2.9Å, respectively. As described in Figure 1.5(b), three angles concerning to the hydrogen bonds of W3 are in proper geometry (128°, 116°, 118°) in contrast to rather deviated angles of the direct hydrogen bond of the C-U pair (137°, 157°). Another remarkable feature is that the C1'-C1' distance of the C-U pair (11.9Å) becomes 1.2Å longer than that of the Watson-Crick base-pair (10.7Å) in order to hold this water molecule (W3) in proper geometry. Interestingly such kind of water molecules is commonly observed among all the C-U pairs (Holbrook *et al.*, 1991; Cruse *et al.*, 1994). All of these data suggests that the hydrogen bonds of W3 are strong enough to make the C-U pair open to the minor groove as to fix W3 in a proper geometry, and probably compensates the loss of one direct hydrogen bond between bases. As a result,

U7 and C8* must rotate by 20° and 17°, respectively, around the normal vectors to the base plane, toward the minor groove. This rotation also induces the slide of the C-U pair by 1.1Å from the normal position of the base-pair in A-form, and this slide makes the major groove deeper.

Hydration patterns around the G-U and C-U pairs

The hydration of non-Watson-Crick base-pair region was presented in Figure 1.6. The large open circles represent the water molecules and hydrogen bonds are drawn as broken lines. Inter-atomic distances less than 3.2Å are defined as hydrogen bonds. All the hydrogen bonding sites of bases faced to the major groove are fully occupied by water molecules and they are located approximately in the base-pair plane of G-U and C-U. On the other hand, hydrogen bonds between the water molecules in each base-pair plane are observed between W1 and W6 at G9*-U6 / C8*-U7 step (Figure 1.6(a)(b)) and between W6 and W8 at C8*-U7 / U7*-C8 step (Figure 1.6(c)(d)) only. Even between adjacent water molecules which are out of hydrogen-bonding distance, the electrostatic and van der Waals interactions seem to remain. Interestingly, the existence of W1 enables these water molecules to make a cyclic hydrogen bond network including the phosphate oxygen of C8* in the major groove of the G9*-U6 / C8*-U7 step, indicated with thick broken lines in Figure 1.6(a)(b). It might be true that W1 is an important molecule not only in the G-U pair but also in the hydrogen bond network at the G9*-U6 / C8*-U7 step.

At the C8*-U7 / U7*-C8 step, a pseudo cyclic hydrogen bond network is observed because the hydrogen bond circle is interrupted between W8 and W9. However, the electrostatic and van der Waals interactions seem to remain as mentioned above. The tandem C-U pairs are also stabilized by this hydrogen bond network in the major groove, but there is no hydrogen bond between W3 and the other water molecules.

In the cyclic and pseudo cyclic hydrogen bond networks, the phosphate oxygens of G9 and C8 are linked to N7 of G9 via three hydrogen bonds through two water molecules, and the phosphate oxygens of C8 and U7 are to N4 of C8 as well. This kind of hydrogen bond linkages can be seen in some other base-pair steps of this structure and are strictly maintained in the crystal structure of r(CCCCGGGG) (Egli *et al.*, 1996) as well. This observation suggests that this hydrogen bond linkage can reflect the common hydration pattern in the major groove. However, in the center of the major groove, that is, the deepest part of the groove, the hydration pattern of the tridecamer is not the same as r(CCCCGGGG). It is somewhat unclear whether these variations of the hydration pattern are sequence or conformational dependent.

Geometry of the terminal uridine residue U1

On the structure refinement process of the tridecamer, the electron density around U1 residue was always broad and not improved. It may be caused by statistical disorder which

arises from non-symmetrical base-pairing with 50% occupancy around the crystallographic two-fold axis. As a possible disordered model explainable for the electron density, the anti-anti (Figure 1.7(a)) and anti-syn models (Figure 1.7(b)) were considerable. In the case of the anti-anti model, the both pairing U1 residues are in the anti-conformation around the χ angle with single hydrogen bond, and in the case of the anti-syn model, one residue is in the anti- and the other residue is in the syn-conformation with two hydrogen bonds. We named the conformer in the major groove side as U1a and the other as U1b. To confirm which model is correct, the omit map of U1b was calculated and the result supported that the anti-anti model is better and more plausible than the anti-syn model. 2|Fo|-|Fc| electron density map of the U-U pair was presented in Figure 1.7(c). It is consistent that this type of the U-U pair was also observed in RNA hexamer (Wahl *et al.*, 1996), and in general, the syn-conformation is not preferred for pyrimidine nucleotides because of steric hindrance between O2 and sugar ring.

Crystallization condition and an equilibrium between hairpin and duplex conformers

We searched thoroughly possible crystallization conditions for the formation of the hairpin structure. We employed sodium cacodylate (pH6.5) as a buffer, sodium chloride as a monovalent cation, magnesium chloride or barium chloride as a divalent cation, polyethylene glycol #400 (PEG#400) or 2-methyl-2,4-pentanediol (MPD) as a precipitant under different conditions. We also used spermine or spermidine because the polyamine is responsible for the condensation of DNA in vivo and is often used for crystallization. Under an different condition using spermine and MPD, a plate-like crystal with the same cell dimension containing the tridecanucleotide as a double-stranded structure but not a hairpin loop structure.

Even though several NMR and thermal denaturation studies indicated the presence of a stable hairpin structure (Tuerk *et al.*,1988; Sakata *et al.*, 1990; Cheong *et al.*, 1990; Antao, V. P. & Tinoco, I. Jr., 1992), there is so far no report on the crystal structure of an RNA hairpin molecule using short oligonucleotides. At first, the reason why the "UUCG" hairpin loop crystallized in a double stranded form was thought that an equilibrium between the hairpin and double stranded structure was shifted to the double stranded form because of high concentrations of the oligomer and salt. However Kanyo *et al.*, (1996) showed by fluorine NMR that even in the crystallization condition of the dodecamer (Holbrook *et al.*, 1991; Baeyens *et al.*, 1994), the hairpin molecule was the main component in the solution phase. So we have to consider the other reason from the solution equilibrium and will discuss it later.

Conformational polymorphism

Overall structure of this double helix was compared with those of the A-RNA and A'-RNA conformers (Arnott *et al.*, 1972: 11 residues per turn for A-form; 12 residue per turn for A'-form). The root mean square deviations (rmsd) of the coordinates of a tridecamer duplex from A-RNA, and from A'-RNA are 1.59Å and 1.12Å, respectively, and the conformation

of the tridecamer is rather similar to A'-RNA. The non-Watson-Crick base-pair portion also shows a small distortion from A-RNA and A'-RNA. Then we calculated the backbone torsion angles for the comparison with those of A- and A'-RNA (Table 1.2). The mean torsion angles of the tridecamer are slightly different from the values of both A-RNA and A'-RNA, but similar to those of the double helical region of $tRNA^{Phe}$ calculated by Kitamura *et al.* (1984). The regular backbone conformations are conserved but exceptionally large distortions are found in ε and ζ of U6, β of U7, β and γ of G9 and β of G10 which are caused by local movements of the bases themselves or adjacent bases by introduction of non-Watson-Crick base-pairs.

To clarify the conformational properties in detail, we listed in Table 1.3(a) the mean helical parameters and the major and minor groove widths of the tridecamer together with those of A-RNA, A'-RNA, B-DNA and the single crystal structures of the oligonucleotides in previous works (255D in PDB code: Holbrook *et al.*, 1991; 1RNA: Dock-Bregeon *et al.*, 1988; 1SDR: Schindelin *et al.*, 1995; 205D: Baeyens *et al.*, 1995). All the coordinates are taken from PDB and their names represent the PDB code. 1SDR contains two double helices in an asymmetric unit, and they were named as 1SDR1 and 1SDR2, respectively. Their sequences are written in Materials and Methods. We selected the previously reported oligomers which were long enough to calculate the major groove width and have little standard deviations of the structure parameters. Basically, we follow the definitions and the nomenclatures of nucleic acid structure parameters by Diekmann *et al.* (1989), but "*" mark means that the calculation method was simplified. The structure parameters listed in Table 1.3(a)(b) are a minimum set of the variables which reconstruct the global conformation. There are two remarkable values. 1) Inclinations* of tridecamer and 255D are much smaller than the others. 2) the major groove width of tridecamer and 255D are significantly wider than the others. When the Inclinations* are plotted against the major groove width, two clearly separated clusters appeared (Figure 1.8). Interestingly, each cluster contains a canonical A-RNA and A'-RNA, respectively. The standard deviations of each point along both axes are much smaller than the distance between the clusters. Together with these data, we concluded that these two clusters must be assigned to an A-RNA group and an A'-RNA group. This implies that the A'-RNA conformation is confirmed in single crystal structures, that is, the tridecamer and 255D belong to the A'-RNA conformer. On the other hand, 1RNA, 1SDR1, 1SDR2 and 205D belong to the A-RNA conformer. The good separation between the two clusters suggests that there might exist energy barrier between A-RNA and A'-RNA, and it will be discussed later.

Cluster analysis

We performed the cluster analysis to reconfirm that the conformation of these oligomers can be divided into two groups. In the cluster analysis in general, the samples are classified based on similarity in valuables which describe the structural feature of the samples. As

mentioned previously, the mean structure parameters in Table 1.3(a) are the valuables which define the global conformation so that the cluster analysis can be performed using these parameters directly. However, they include distances and angles and also their values are dispersed. If the values in Table 1.3(a) are used for the cluster analysis, the result might be affected only by a certain parameter which has the largest value. It is better for these values to be standardized in terms of each structure parameter before the cluster analysis. The standardized structure parameters for the cluster analysis are listed in Table 1.3(b). Then Euclidean distances based on the standardized structure parameters between all the combinations of the oligomers, A-RNA, A'-RNA and B-DNA were calculated (Table 1.4). The oligomers, A-RNA, A'-RNA and B-DNA were clusterized by means of the nearest neighbor method (Figure 1.9). B-DNA and all RNA group are separated with quite a long distance of 5.66. Markedly, the A-RNA and A'-RNA groups in Figure 1.8 belonged to the different clusters by this method as well, and separated at the distance of 1.68. On the other hand, the longest distance in each cluster is 0.86 and 0.74, respectively, which are much shorter distances than that between A-RNA and A'-RNA groups. The difference in the distances of the inter- and intra-groups is of significance.

Discussion

The structure of the tridecamer reveals that the pyrimidine-pyrimidine pair such as the C-U pair can exist. Studies on stability of the non-Watson-Crick base-pairs are indispensable for the prediction of the secondary and tertiary structure of RNA. From this reason, thermodynamic studies of mismatches in double stranded RNAs were performed (SantaLucia *et al.*, 1991; He *et al.*, 1991; Wu *et al.*, 1995). The thermodynamic parameters and molecular structure of the G-U pair in solution were well studied (He, *et al.*, 1991; Wu *et al.*, 1995; Allain & Varani, 1995b). However, there was only thermodynamic parameters for the C-U pair (SantaLucia, *et al.*, 1991). Although the existence of the C-U pair was predicted in *Eschelichia coli* 23S ribosomal RNA (Gutell *et al.*, 1994) and the C-T mismatch pair was claimed to be stacked in a DNA duplex in solution (Boulard *et al.*, 1997), it is still unclear whether C and U can form a base-pair in solution as in crystal. The solution structure of the C-U pair is the next target to be solved. Furthermore, there is no thermodynamic parameters of asymmetrical tandem mismatches and they should be also determined. Without these parameters, there still remains ambiguity for the secondary structure prediction of a large RNA molecule. The geometry and hydration pattern of successive non-Watson-Crick base-pair in this study are important for the investigation of the meaning of the thermodynamic parameters and for the understanding of the general principle of RNA folding.

The crystal structure of the P4-P6 domain of the *Tetrahymena thermophila* intron (Cate *et al.*, 1996a,b) indicates that the main interaction between two helices located side by side in parallel is mediated by the minor groove contact. In the crystal structure of the tridecamer, two helices related by the crystallographic two-fold screw axis interact through

the minor groove as well. The packing interaction of the tridecamer might reflect a general principles of the duplex-duplex interaction. The reason why hairpin molecules using short oligonucleotides crystallize in a double stranded form (Holbrook *et al.*, 1991; Fujii *et al.*, 1992; Cruse *et al.*, 1994; Baeyens *et al.*, 1996) except for the Z-DNA hairpin (Chattopadhyaya *et al.*, 1988) can be explained from the point of the minor groove contact. The minor groove surface area of the self-associated double helix is larger than that of the hairpin molecule so that the self-associated double helices can make specific contacts with another molecule through the minor groove, while that of the hairpin molecule is so small that the hairpin molecules can not contact well enough to make the crystal lattice. To obtain the crystal containing hairpin molecules, the oligomer with a longer Watson-Crick base-paired stem should be used. If we want to obtain the structure of the UUCG loop, it may be better to introduce this loop into a much larger molecule such as the GAAA loop in the hammerhead ribozyme (Pley *et al.*, 1994; Scott *et al.*, 1995; Scott *et al.*, 1996).

In an early stage of RNA study, the structure of the RNA molecule was thought to be more flexible than that of DNA molecule, because the most of RNA molecules are expressed as single strands and take various structures. But in the last decade, its structure came to regard as more rigid than that of DNA molecule although the RNA molecule can take various structures (Puglisi *et al.*, 1992; Pley *et al.*, 1994; Scott *et al.*, 1995; Aboul-ela *et al.*, 1995; Puglisi *et al.*, 1995; Cai *et al.*, 1996; Jiang *et al.*, 1996; Cate *et al.*, 1996a,b; Aboul-ela *et al.*, 1996; Battiste *et al.*, 1996; Ye *et al.*, 1996; Scott *et al.*, 1996). However, if the RNA molecule is just a rigid molecule, it seems to be inconsistent with various activities of the RNA molecule such as the intron excision, the ribonuclease activity of RNase P and the protein synthesis through the ribosomal RNA. The existence of A'-RNA conformer not only expands the conformational space but also could introduce the flexibility again through the conformational exchange between A-RNA and A'-RNA, even if each conformer is rigid. There is a possibility that biologically active RNA molecules might utilize the flexibility which arises from the conformational exchange between A-RNA and A'-RNA conformers.

There still exists the problem whether A-RNA and A'-RNA are independent conformers like A- and B-forms of DNA. To speculate this problem, dynamics data of RNA fiber by solid-state ^{31}P NMR spectra can be used. Solid-state ^{31}P NMR spectrum of RNA fiber shows that A- or A'-form does not display any conformational change and heterogeneity (Shindo *et al.*, 1985). The conformational conservation of A- or A'-form is much more evident than the heterogeneous B-form DNA fiber as revealed by the solid-state ^{31}P NMR spectra (Shindo & Zimmerman, 1980; Shindo *et al.*, 1984; Shindo *et al.*, 1985; Fujiwara and Shindo, 1985). These may suggest that in solid state, A-RNA or A'-RNA can not be converted to the other conformer easily and there may exist a energy barrier but the matter is still obscure. In respect of solution phase, it is much obscurer because of the paucity of the data. Anyway this problem also must be studied. Then, factors which induce A'-RNA conformation must be explained. In the case of the present tridecamer, the possible explanation is that an A'-RNA

conformer which has twelve residues per turn is suitable for the tridecanucleotides to make pseudo infinite helices in crystal because an intermolecular U-U base-pair is flexible and seems not to affect the helical phase. The energy levels of both conformers might not be so different and there is a possibility where A'-RNA conformer could be induced and observed in natural RNA.

We retrieved the biologically occurring frequency of A'-RNA. In the crystal structure of the P4-P6 domain of *Tetrahymena thermophila* intron (1GID: Cate *et al.*, 1996a,b from PDB), we could not find any A'-RNA conformation. We searched the A'-RNA conformation in the crystal structure of the hammerhead ribozyme (1HMH: Pley *et al.*, 1994; 1MME: Scott *et al.*, 1995 from PDB) as well, but the helices are so short that we could not measure the major groove width to judge the conformation. Therefore, we calculated the rmsd of the coordinates between stem I, the longest double helix, of the hammerhead ribozyme and A-RNA as well as A'-RNA. The rmsd values between stem I and A'-RNA are smaller than those between stem I and A-RNA in four out of five crystallographically unique molecules in 1HMH and 1MME. The smallest rmsd value between stem I and A'-RNA is 1.03Å which indicates that stem I of this molecule is similar to A'-RNA. In the crystal structure of hammerhead ribozyme, stem I is located so close to stem II that can contact each other. Stem I may avoid the steric hindrance between stem I and II, by taking A'-RNA conformation. This also suggests that A'-RNA conformation could exist in other natural RNA, and the double helical region in a large RNA molecule may select suitable conformation according to environment.

The conformation of A'-RNA is different from A-RNA in the helical phase and the major groove width which affect both RNA-RNA and RNA-protein interaction. Markedly, in the case of Tat-TAR, rev-RNA aptamer and rev-RRE (Puglisi *et al.*, 1995; Ye *et al.*, 1996; Battiste *et al.*, 1996), Tat and rev peptides bind to the major groove of target RNAs and widen the major groove upon their binding. It is obvious that the A-RNA conformer does not allow these peptides to accommodate in its quite narrow major groove. Although the authors of the above three papers explained that the widening of the major groove was caused by the flexible conformation around the bulge loop, now there arises another possibility, that is, the global conformational change of the double helix to the A'-RNA conformer. The A'-RNA conformation which has the wide major groove must be taken into consideration for the RNA-protein interaction.

Although the cluster analysis is one of the general statistical methods, it has not been applied to the classification of the RNA conformations. The use of rmsd of coordinates for the classification of the RNA conformation in either A-RNA or A'-RNA brought an ambiguous result, while the cluster analysis gave quite a clear result for this purpose. The major advantages of the cluster analysis to the method of rmsd can be summarized in the following two points. 1) The double-stranded oligomers which are different in length and sequence can be compared directly. 2) The Euclidean distance used in the cluster analysis sensitively reflects the

conformational difference because it is calculated from the structure parameters. For example, the rmsd of the coordinates from the present tridecamer to those of A-RNA and that to A'-RNA are 1.59Å and 1.12Å, respectively, while the distances of the cluster analysis from the tridecamer to A-RNA and that to A'-RNA are 2.12 and 0.74, respectively. The difference is evident.

Table 1.1. Summary of B-factors.

	phosphate	ribose	base
U1a		25.06	22.93
U1b		25.09	23.32
G2	27.26	21.10	14.76
A3	20.60	16.80	11.26
G4	17.05	12.87	9.75
C5	14.46	12.62	10.88
U6	12.03	10.68	11.47
U7	10.05	8.65	8.58
C8	10.42	11.83	11.38
G9	15.08	13.25	10.53
G10	14.82	11.86	9.60
C11	11.64	11.57	8.50
U12	14.35	13.35	10.28
C13	17.36	16.36	14.19

Because of the disordered structure, U1 has two conformers, U1a and U1b.

Table 1.2. Torsion angles of the tridecamer, A-RNA, A'-RNA and $tRNA^{Phe}$.

	α	β	γ	δ	ε	ζ	χ
U1a			62.4	78.5	208.2	290.5	194.5
U1b			54.9	81.7	218.1	**308.9**	193.7
G2(U1a)	294.8	167.5	67.5	77.7	204.4	286.5	189.8
G2(U1b)	292.2						
A3	288.8	176.8	58.4	78.1	205.4	286.5	203.9
G4	282.2	178.7	58.6	81.9	210.1	292.8	195.9
C5	286.3	162.4	74.4	77.3	204.8	286.1	195.0
U6	292.7	165.3	54.4	77.6	**187.8**	**264.4**	193.3
U7	282.1	**187.0**	60.2	76.6	207.3	286.6	196.5
C8	290.4	171.4	56.8	76.3	207.4	292.4	197.5
G9	286.3	**159.9**	**78.8**	80.5	207.5	285.5	178.7
G10	284.0	**184.7**	54.8	77.6	206.2	286.2	187.2
C11	292.5	176.4	59.6	80.1	207.8	284.6	201.8
U12	293.9	165.2	61.0	77.4	214.2	285.8	205.3
C13	283.4	162.0	68.7	75.4	201.4		201.4
ave.(13mer)	288.4	171.5	62.2	78.3	206.5	287.4	195.3
SD(13mer)	4.62	9.15	7.50	1.98	6.80	9.46	7.02
A-RNA	294	186	49	95	202	294	205.0 / 202.0
A'-RNA	297	189	45	83	196	297	200
ave.($tRNA^{Phe}$)	272	186	68	79	206	287	196
SD($tRNA^{Phe}$)	32	34	21	15	29	23	19

Because of the disordered structure of U1, γ, δ, ε, ζ and χ of U1, and α of G2 have two values.

ave.(13mer) and SD(13mer) mean average and standard deviation of each torsion angle of the tridecamer in this work.

ave.($tRNA^{Phe}$) and SD($tRNA^{Phe}$) mean average and standard deviation of each torsion angle of the double helical region of $tRNA^{Phe}$.

Table 1.3. Structure parameters of each oligomer.

(a)	major groove width	minor groove width	Inclination*	Twist*	Rise*	X-displacement*
13mer	**8.5±0.51(6)**	10.0±0.92(9)	**10.2±1.78(12)**	30.8±4.62(11)	3.0±0.24(11)	4.4±0.67(12)
255D	7.5±1.01(6)	10.0±0.97(9)	11.0±2.37(12)	31.1±4.47(11)	2.9±0.28(11)	4.4±0.72(12)
1RNA	3.7±1.60(6)	10.2±0.48(11)	19.3±4.89(14)	33.4±3.08(13)	2.7±0.37(13)	4.3±0.62(14)
1SDR1	2.9±0.65(4)	10.3±0.46(9)	19.9±2.17(12)	33.5±3.92(11)	2.6±0.21(11)	4.4±0.32(12)
1SDR2	3.0±0.31(4)	10.3±0.32(9)	20.4±2.22(12)	33.3±3.90(11)	2.6±0.16(11)	4.3±0.23(12)
205D	4.3±0.62(4)	9.8±0.54(9)	15.2±2.84(12)	33.3±8.51(11)	2.8±0.45(11)	4.3±0.41(12)
A'-RNA	8.1	10.9	10.1	30.0	3.0	4.5
A-RNA	4.3	11.0	16.0	32.7	2.8	4.4
B-DNA	11.4	5.7	-4.6	36.0	3.4	-0.2

(b)	major groove width	minor groove width	Inclination*	Twist*	Rise*	X-displacement*
13mer	0.83	0.12	-0.37	-1.06	0.45	0.33
255D	0.50	0.15	-0.26	-0.86	0.17	0.39
1RNA	-0.75	0.26	0.81	0.40	-0.80	0.31
1SDR1	-1.02	0.30	0.88	0.48	-0.87	0.34
1SDR2	-0.99	0.30	0.94	0.35	-1.10	0.30
205D	-0.57	-0.01	0.27	0.32	-0.38	0.28
A'-RNA	0.71	0.70	-0.38	-1.48	0.57	0.40
A-RNA	-0.56	0.76	0.38	0.03	-0.19	0.33
B-DNA	1.83	-2.58	-2.26	1.83	2.16	-2.66

The definitions of the parameter names are given in Figure 8.

(a) Table of the mean structure parameters with their standard deviations and numbers of variables.

(b) Table of the standardized structure parameters for cluster analysis.

Table 1.4. Pair-wised Euclidian distances between the oligomers, A-RNA, A'-RNA and B-DNA by the standardized structure parameters.

	255D	1RNA	1SDR1	1SDR2	205D	A'-RNA	A-RNA	B-DNA
13mer	**0.49**	2.75	3.02	3.07	2.23	**0.74**	2.12	**5.66**
255D		2.29	2.55	2.60	1.78	0.94	**1.68**	5.81
1RNA			**0.30**	0.41	**0.76**	3.02	0.99	6.62
1SDR1				**0.27**	0.97	3.26	1.16	6.80
1SDR2					1.11	3.30	1.28	6.93
205D						2.59	**0.86**	6.01
A'-RNA							2.24	6.20
A-RNA								6.44

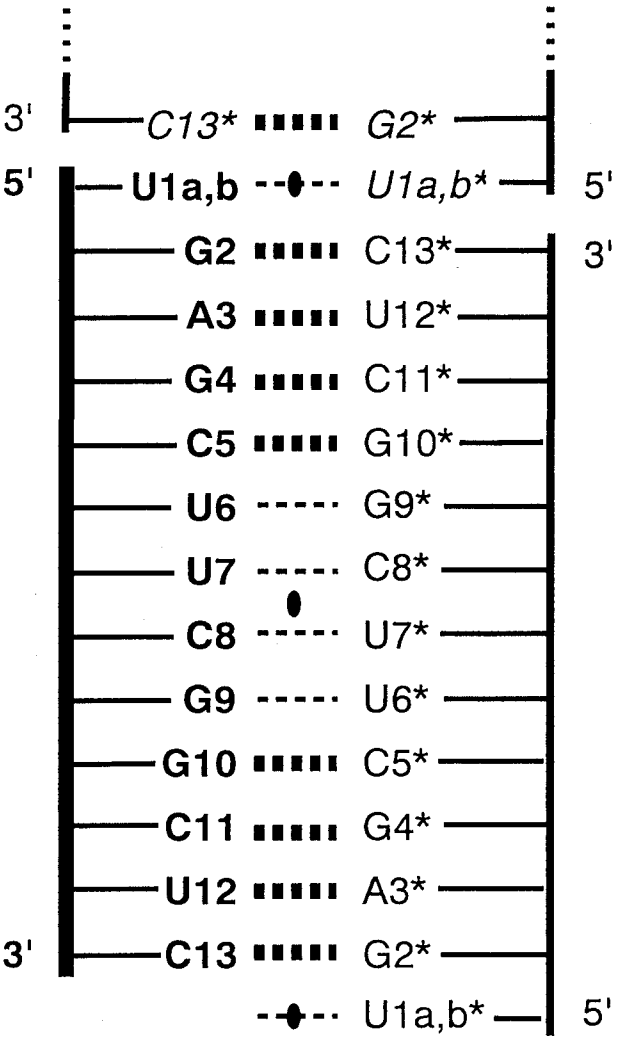


Figure 1.1. The numbering of the bases and the base-pair scheme are shown. The crystallographically independent molecule is presented in boldface and symmetrically related molecules are presented in normal type with asterisks. The helix piled up along the c-axis is also presented in italics. Because of the disordered structure, U1 has two conformers, U1a and U1b, which are denoted as U1a,b for short. The bold and thin broken lines represent Watson-Crick base-pairs and non-Watson-Crick base-pairs, respectively. The black ovals represent the crystallographic two-fold axis.

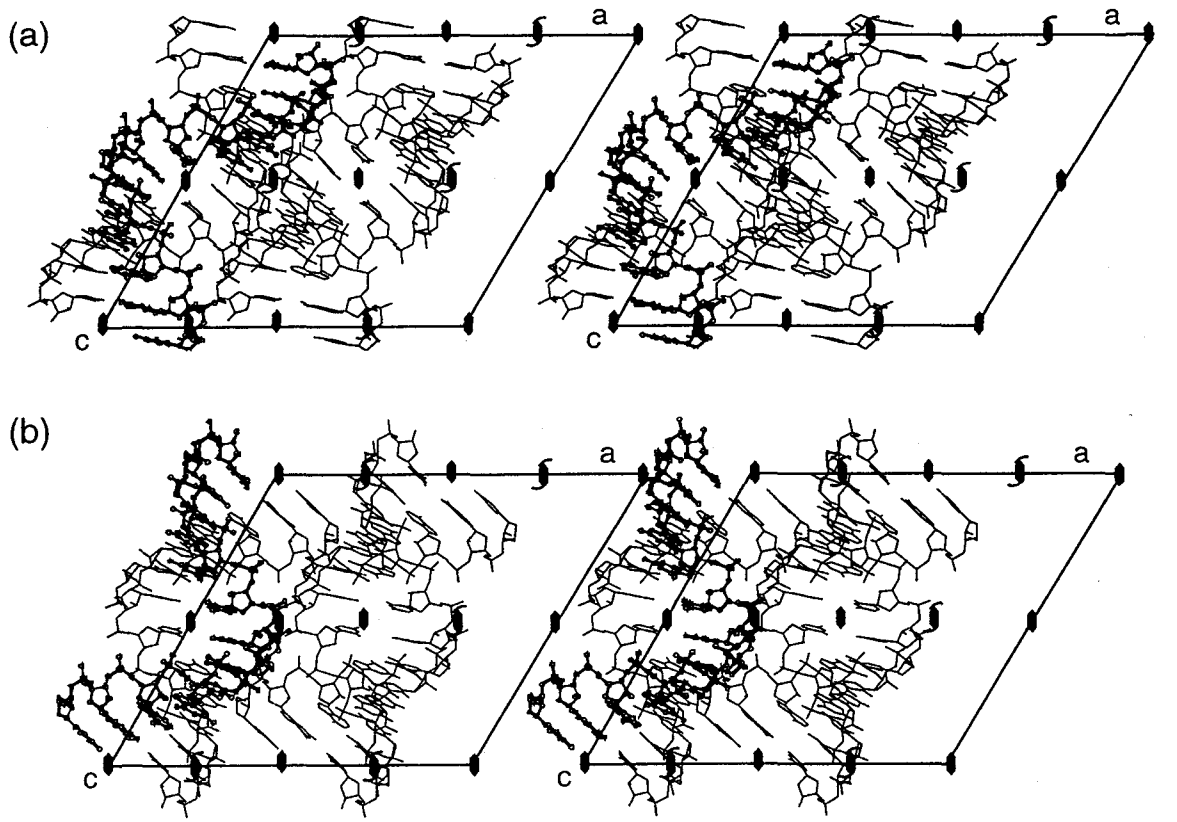


Figure 1.2. Two possible orientations which satisfy the crystallographic symmetry. All oligomers in a unit cell are included in this figure and a crystallographically unique molecule is presented as a ball–and–stick model. The a–axis and c–axis of the unit cells are indicated as a parallelogram in the figure, and the b–axis is perpendicular as to the paper. The crystallographic two–fold axes and two–fold–screw axes are also shown. Each figure is viewed from the plus region to the minus region of the b–axis.

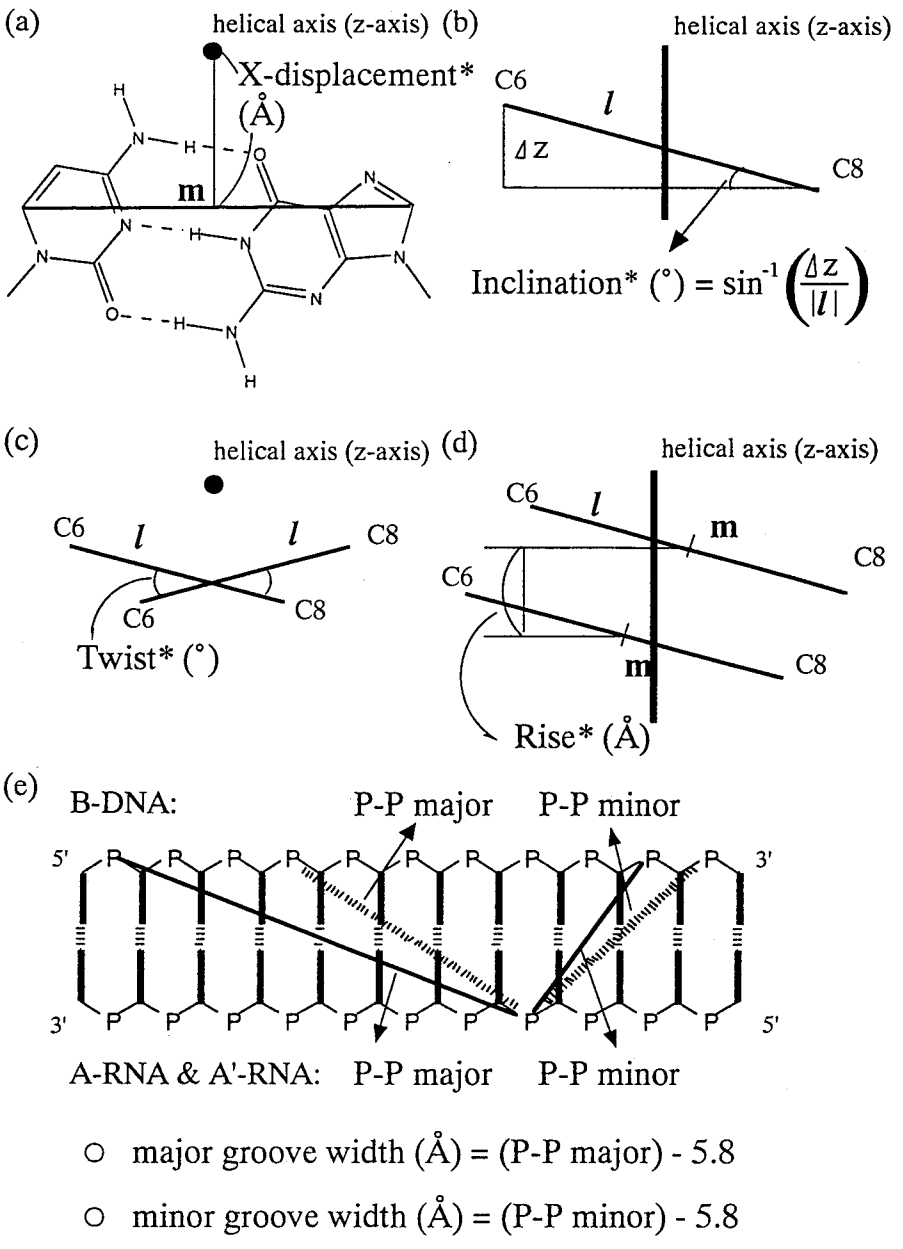


Figure 1.3. The definitions of the structure parameters are shown. "**m**" represents the mid-point between the C8 and C6coordinates in the case of a purine-pyrimidine base-pair, and that between the C6 and C6 coordinates in the case of apyrimidine-pyrimidine base-pair. "*l*" represents the long axis vector between C8 and C6 in the case of a purine-pyrimidinebase-pair, and that between C6 and C6 in the case of a pyrimidine-pyrimidine base-pair. z is the difference in the z-coordinates of C8 and C6 for a purine-pyrimidine base-pair, and those of C6 and C6 for a pyrimidine-pyrimidine base-pair.The filled black circles in (a) and (c), and the vertical thick lines in (b) and (d) indicate the helical axis. (a) X-displacement* is the distance between **m** and the helical axis (viewed down the helical axis). (b) Inclination* is the anglebetween the vector, *l*, and the x-y plane (viewed perpendicular to the helical axis). (c) Twist* is the angle between theprojections of vectors *l*s to the x-y plane of the neighboring base-pairs (viewed down the helical axis). (d) Rise* is thedifference in the z-coordinates of **m**s in neighboring base-pairs (viewed perpendicular to the helical axis). (e) P-P majorand P-P minor was defined before calculation of the major groove width and minor groove width. The trans zigzag line and"P" represent the backbone and phosphorus atoms of the double helix, respectively. In A-RNA or A'-RNA, the nearestphosphorus atoms through the major groove and the minor groove are linked by a solid line, and their linear distances areP-P major and P-P minor, respectively. In B-DNA, they are represented by dashed lines. The major groove width andminor groove width were obtained by subtracting 5.8Å from P-P major and P-P minor, respectively.

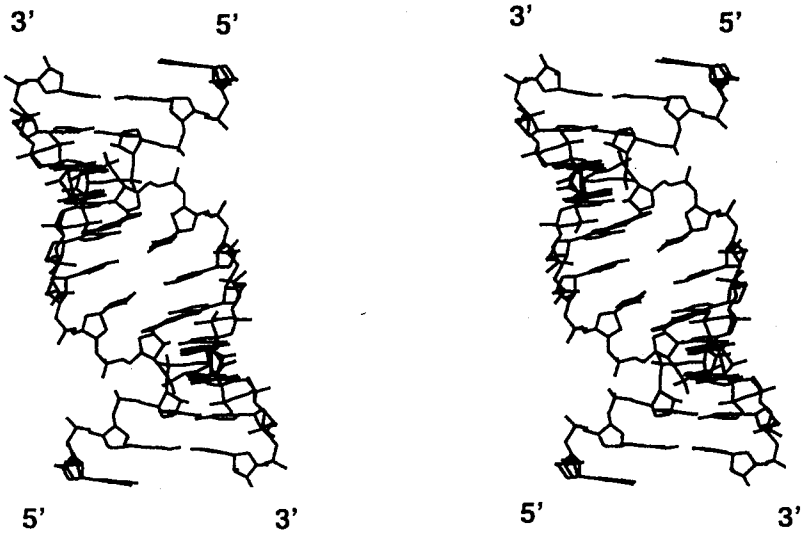


Figure 1.4. Stereo view of the double–stranded form of the tridecamers.

(a)

(b)

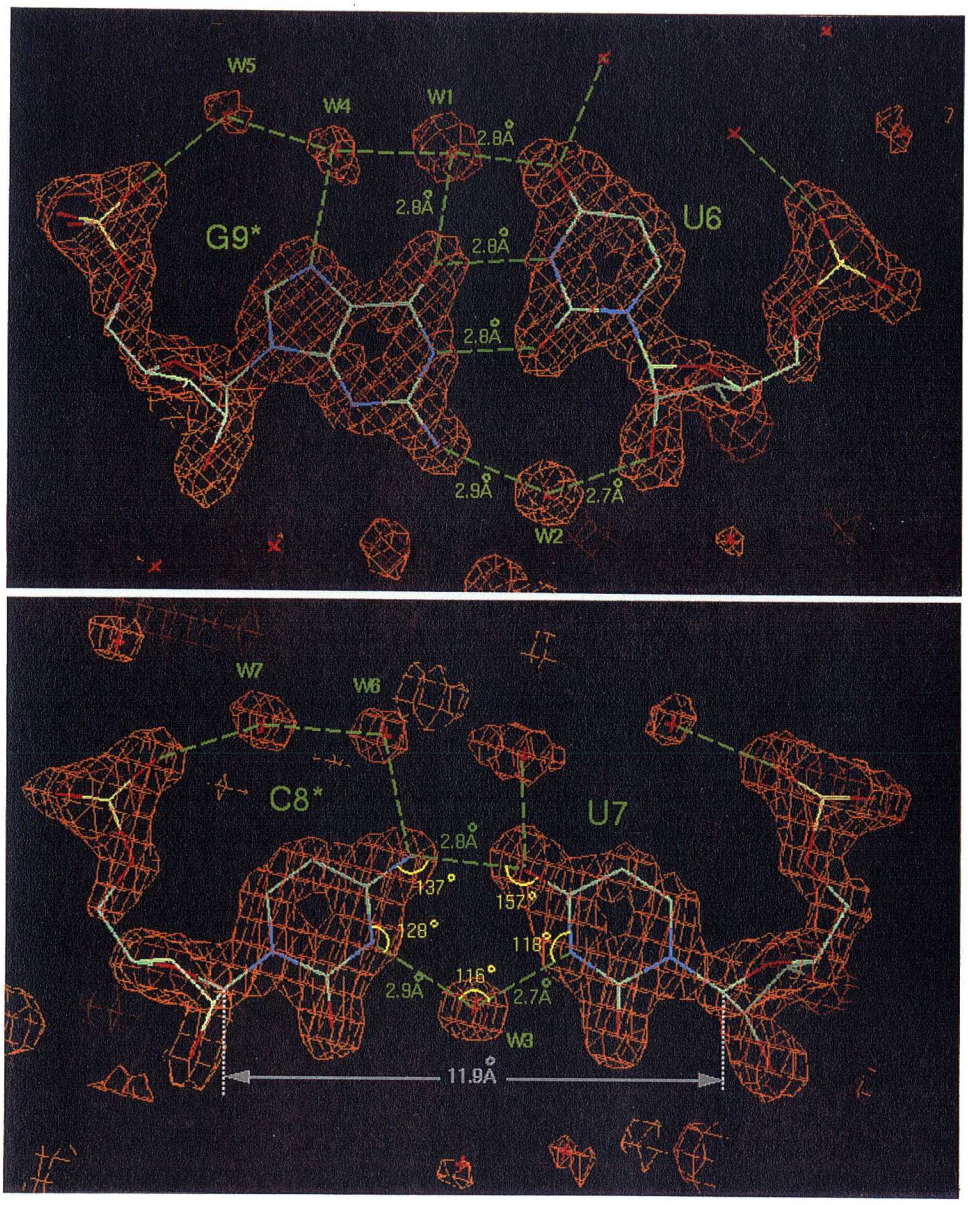


Figure 1.5. (a) 2|Fo|-|Fc| map of the G-U pair drawn at the level of 1.2 σ with a wire model. (b) 2|Fo|-|Fc| map of the C-U pair drawn at the level of 1.2 σ with a wire model. The names of the residues used here are given in Figure 1. The light green, red, blue and yellow in the wire model represent carbon, oxygen, nitrogen and phosphorus atoms, respectively. The red asterisks represent the oxygens of water molecules. The names of the water molecules which are discussed in the text are also presented. The green broken lines represent the hydrogen bond, with their distances. The hydrogen-bonding-related angles and the C1'-C1' distance of the C-U pair are presented as yellow and grey characters, respectively.

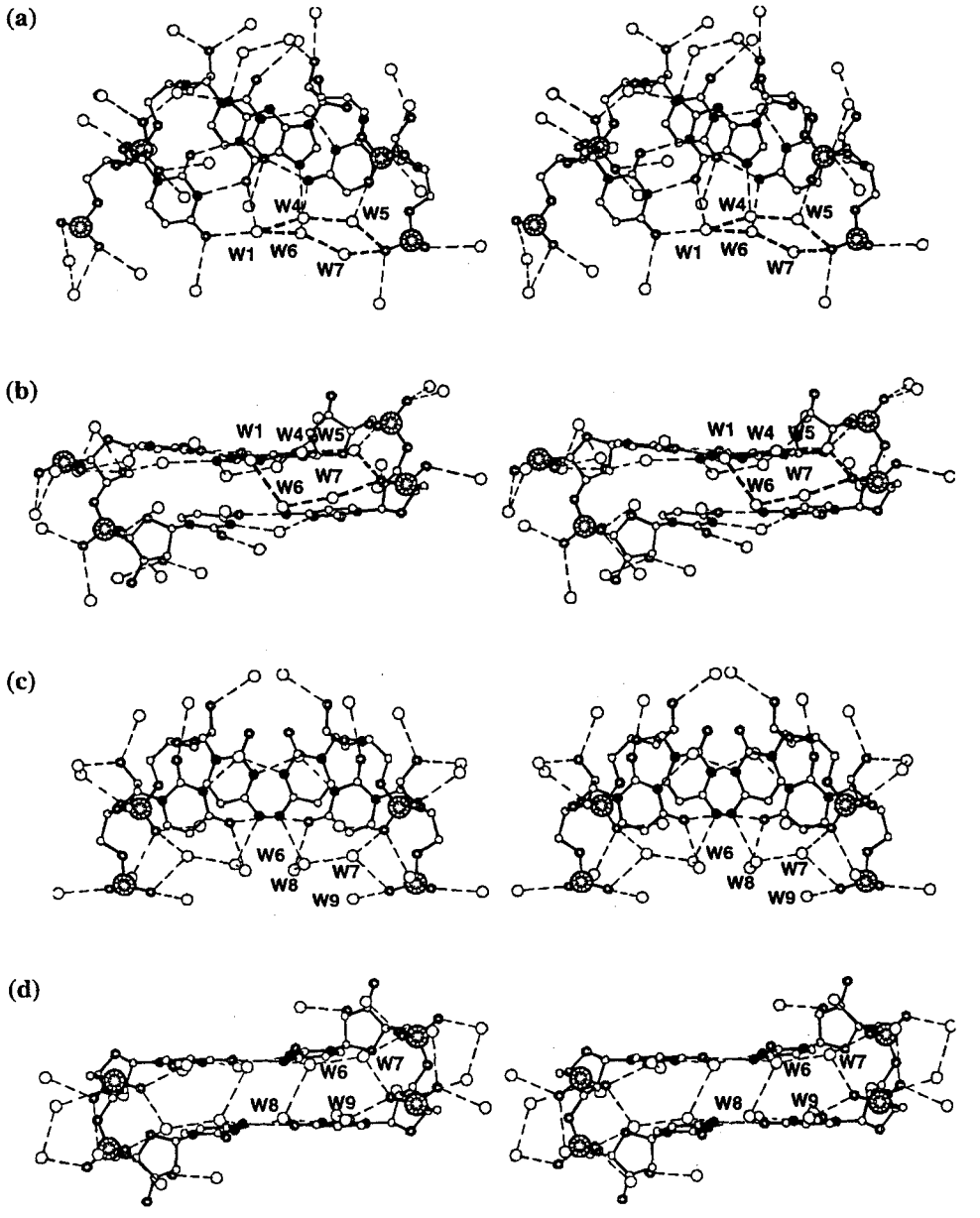


Figure 1.6. The hydrogen bond networks around the G9*-U6 / C8*-U7 step and the C8*-U7 / U7*-C8 step are shown. The sparked, double, shaded, small open and large open circles represent phosphorus, oxygen, nitrogen, and carbon atoms in RNA, and oxygen atoms in water molecules, respectively. The broken lines represent the hydrogen bonds with distances of less than 3.2Å. The bold broken lines represent the cyclic hydrogen bond network, with the names of the water molecules. (a) A stereo diagram of the G9*-U6 / C8*-U7 step viewed down the helical axis. (b) A stereo diagram of the G9*-U6 / C8*-U7 step viewed into the major groove. (c) A stereo diagram of the C8*-U7 / U7*-C8 step viewed down the helical axis. (d) A stereo diagram of the C8*-U7 / U7*-C8 step viewed into the major groove.

(a)

(b)

(c)

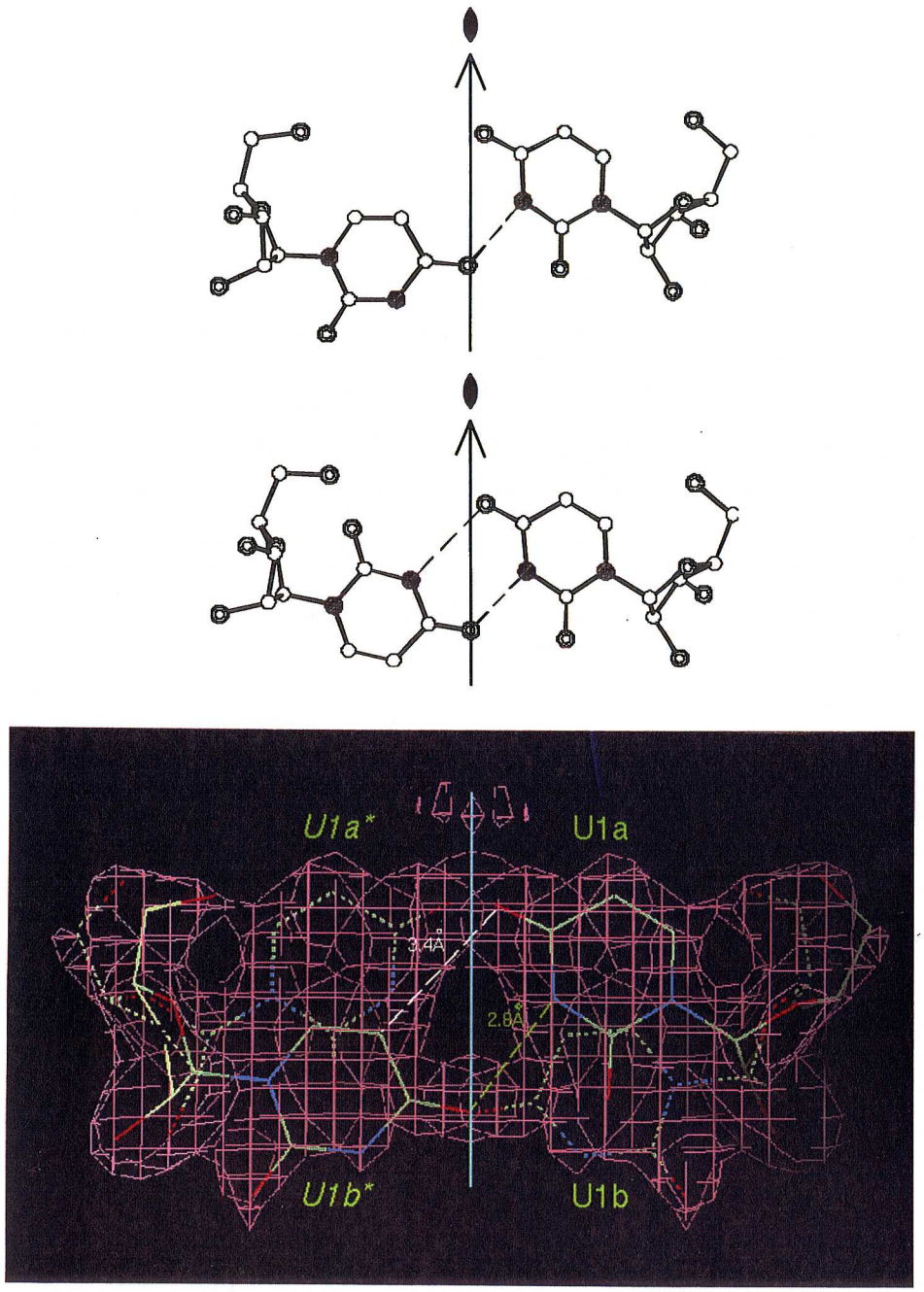


Figure 1.7. (a) The anti–anti model is drawn with a hydrogen bond as a broken line. The solid line indicates the crystallographic two–fold axis. (b) The anti–syn model is drawn with two hydrogen bonds as broken lines. The solid line indicates the crystallographic two–fold axis. (c) A 2|Fo|–|Fc| map of U1a and U1b drawn at the level of 0.8σ with a wire model. The names of the residues used here are the same as in Figure 1.1. The broken wire model represents the half of the disordered molecule which is related by the crystallographic two–fold axis. The light blue line represents the crystallographic two–fold axis. The green broken line represents the hydrogen bond, with its distance. The distance between C5 of *U1b** and O4 of U1a is 3.4 Å (white broken line)

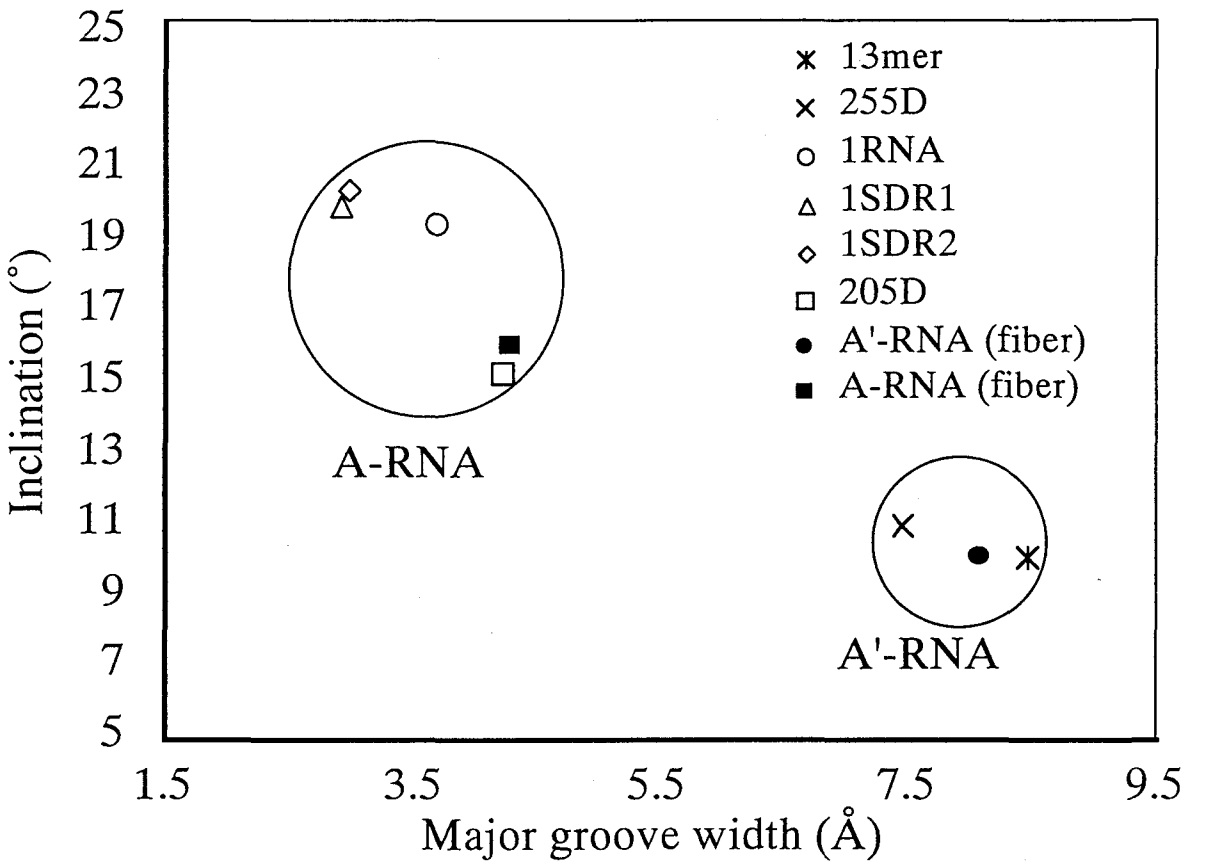


Figure 1.8. The mean inclination* value of each oligomer duplex isplotted against the value of the mean major groove width of each oligomerduplex. Each point represents the corresponding oligomer duplex in thekey. The coordinates and names of the oligomers were taken from PDBand its code. 1SDR includes two duplexes in an asymmetric unit, whichare named 1SDR1 and 1SDR2, respectively. See Materials and Methodsfor the names and sequences of the oligomers. A-RNA (fiber) and A'-RNA(fiber) represent the structures obtained from fiber diffraction data.

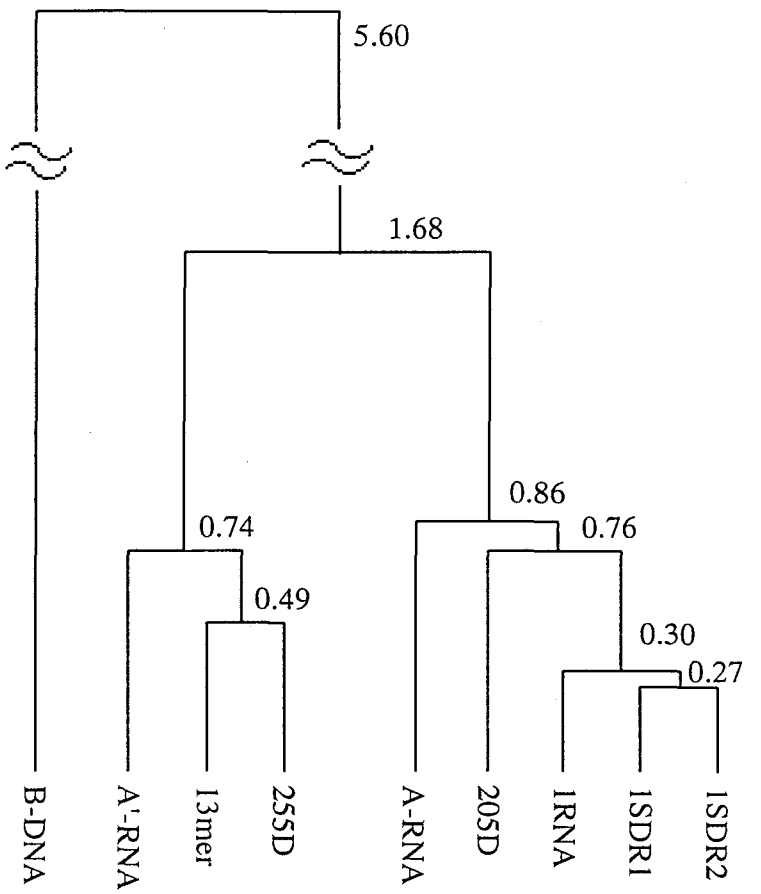


Figure 1.9. Schematic diagram of the cluster analysis is drawn. Oligomers are linked by the nearest neighbor method at the Euclidean distances, which are also listed in Table 4 in boldface. See Materials and Methods for the names and sequences of the oligomers.

Chapter 2

NOE based structure calculation of the C-U pair in solution

Abstract

The structure of the C-U mismatch pair in double stranded RNA oligomers, r(CGACUCAGG) • r(CCUGCGUCG), in solution was studied. The NOE based signal-assignment allowed to assign most of the proton signals, 164 out of 196, with the combination of COSY and ^{1}H-^{31}P hetero COSY spectra. These signals include three 2'-hydroxy protons which are rarely observed. On the other hand, the imino proton signal of the C-U pair was not observed at all. This indicates that the imino proton of the C-U pair is exposed to the bulk water. From the NOE based structure calculation, the RNA oligomers formed a double stranded helical structure with the C-U pair. The χ angles of the cytidine and uridine in the C-U pair were in the anti-conformation. All the results suggested that the configuration of the C-U pair is the same as that in crystal.

Introduction

In the previous chapter, the existence of the C-U pair was demonstrated in crystal. In this chapter, it is studied whether the C-U pair can exist in solution. To prove this, NOE based structure calculation was performed. As mentioned in General Introduction, it is no easy to determine the structure of thermally unstable base-pairs because there may be conformational distributions. The problems accompanying with the structure calculation also will be discussed. As well as the structure calculation, sample preparation of labeled RNA is also described. To study the structure of the C-U pair, sequences were newly designed. This is because the tridecamer in Chapter 1 forms monomeric hairpin loop structure in solution without the C-U pair. Since the C-U pair makes a double helix unstable, a double helix (rCU9) include many C-G pairs was designed to make the double helix as stable as possible.

It is very important to study the structures and characters of the mismatches in solution in order to know the secondary and tertiary structure of RNAs. In *Eschelichia coli* 23S ribosomal RNA the existence of the C-U pair was predicted, and putative a single C-U pair site exists (Gutell *et al.*, 1994). However data are limited to rather stable mismatches such as G-U and G-A which are the non-Watson-Crick base-pairs with two hydrogen bonds. The G-U mismatch pair which is frequently observed in tRNAs and other RNAs is stable enough to maintain a double helix (He *et al.*, 1991). The G-A mismatch pair also makes a stable base-pair in the analogous sequence to the hammerhead ribozyme (Katahira *et al.*,1994). Both of these mismatches form specific base-pairs in crystal (Fujii *et al.*, 1992; Pley *et al.*,1994) and solution (Katahira *et al.*, 1994, Allain & Varani, 1995a,b). On the other hand, thermodynamic studies showed that the pyrimidine-pyrimidine type mismatch such as C-U makes the double helix unstable except for U-U when a certain base-pair comes to the next (SantaLucia *et al.*, 1991; Wu *et al.*, 1995). In spite of the fact, the C-U pair is formed in a specific manner in crystal as described in the previous chapter and the other papers (Holbrook *et al.*, 1991; Cruse *et al.*, 1994). The C-U pair in crystal has a single direct hydrogen bond between bases and seems not to be so stable. It is interesting to see whether such a C-U pair is formed in solution or not.

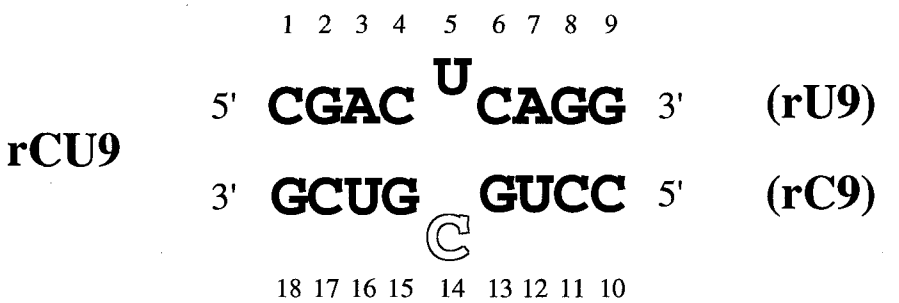
Materials and Methods

Synthesis of the phosphoramidite form of labeled nucleosides

Synthetic route of phosphoramidite form of the cytosine labeled at the amino-nitrogen is drawn in Figure 2.1. $TIPDSCl_2$ and $(Bz)_2O$ represent 1,3-dichloro-1,1,3,3-tetraisopropyldisiloxane and benzoylanhydrous, respectively. The final product "9" is the phosphoramidite form of cytidine for the material of the following oligomer synthesis by the solid phase phosphoramidite method. It is the same as that commercially available except for the 2'-hydroxy-protecting group. It has a tetrahydropyranyl (THP) group on the 2'-hydroxy group instead of the *tert*-butyldimethylsilyl (TBDMS) group. The purity of the final product "9" was judged by ^{31}P NMR.

Synthesis of oligonucleotides

The concepts of the sequence design is as follows. 1) The C-U pair should be located in the middle of the double helix. If the C-U pair is located at the end of the helix or near the end, it may fluctuate. 2) To make the double helix as stable as possible, high ratio of the C-G pair is desired. 3) To maximize the possibility of the formation of the C-U pair, both side of the C-U pair should be the C-G pair. 4) Each sequence should not form a stable self-associate double helix. 5) Other stable combination of the base-pairing should not exist. As a result, the sequences below were designed.



Scheme 2.1

Oligomers, rU9 and rC9, shown above were synthesized by the solid phase phosphoramidite method with leaving the 5'-dimethoxytrityl (DMTr) groups on the 5'-hydroxy groups by a DNA synthesizer Model 380B (ABI). The labeled cytidine was incorporated at the 14th base indicated by the outlined character (Scheme 2.1). For the other residues, the phosphoramidite forms of the non-labeled nucleotides (PerSeptive biosystems) with 2'-TBDMS group as a 2'-hydroxy-protecting group were used. Their base-protecting groups were deprotected in the solution containing aqueous ammonia and ethanol at the ratio of 3:1 at 55 °C for 16 hours in a screw vial. The solvent was evaporated and 1M tetrabutylammonium fluoride in tetrahydrofuran was added to the residuals. They were incubated at 37 °C for 16 hours in a screw vial to remove the 2'-protecting group, 2'-TBDMS group. The fraction of the oligomers with 5'-DMTr groups were collected by a reversed phase column C18 (Waters). Solutions containing the samples were evaporated, and aqueous HCl solution at pH 2.0 was added to the residuals to deprotect the 5'-DMTr groups at room temperature for 15 minutes. In the case of the oligomer with the labeled cytidine, the treatment of HCl solution pH 2.0 was done for 24 hours to remove 2'-THP group as well. They were neutralized and purified again by reversed phase column COSMOSIL (nacalai tesque). Then they were adsorbed on an anion exchange column Mono-Q (Pharmacia) and eluted by 2.0M NaCl for the counter-ion exchange. The excess amount of NaCl was desalted by gel filtration G3000PW (TOSO). They were quantitated from UV absorbance at 260 nm after nuclease P1 digestion.

Measurement of NMR

The solutions for NMR measurements contain 2 mM duplex oligomer, 20 mM sodium phosphate buffer at pH 7.0 and 50 mM NaCl in 90% H_2O and 10% D_2O. When we measured the sample in D_2O, we evaporated H_2O from the sample then dissolved it in the same volume of D_2O without adjustment of pD. The NOESY spectra in H_2O at 0 °C for obtaining NOE data necessary for the structure calculation and signal assignments were measured in 1024*512 column points for the spectral width of 13998.16*13987.53 Hz, 16 scans for each column point on a 600 MHz spectrometer, JEOL alpha600. The NOESY spectra in 90%, H_2O and 10% D_2O at 30 °C for the structure calculation and signal assignments were measured in 1024*512 column points for the spectral width of 14367.82*12002.66 Hz, 16 scans for each column point on a 600 MHz spectrometer, Bruker DRX600. The NOESY spectra in D_2O at 10 °C for the structure calculation and signal assignments were measured in 1024*512 column points for the spectral width of 14367.82*12002.66 Hz, 16 scans for each column point on a JEOL alpha600. All of them were measured for the mixing times of 50, 100 and 200 msec. The COSY spectra were measured at 25 °C in 1024*256 column points for the spectral width of 5000*5000 Hz, with 8 scans on a 500 MHz spectrometer, Bruker ARX500. The NOESY spectra at every 10 °C from 0 °C to 70 °C were taken in 1024*256 column points for the spectral width of 5000*5000 Hz, with 4 scans per column point and mixing time 50 msec on a JEOL alpha600. ^{1}H-^{31}P hetero COSY spectrum (Sklenár *et al.*, 1986)was measured at 27 °C in 512*256 column points for the spectral width of 7042.25*2024.6 Hz, with 32 scans on a Bruker ARX500. HSQC spectrum was measured at 10 °C in 512*20 column points for the spectral width of 4006.4*709.56 Hz, with 32 scans on a Bruker DMX500. Proton chemical shifts were calibrated by the equation below (Hartel *et al.*, 1982).

$$\delta(H_2O) = 7.83 - \frac{T}{96.9} \tag{2.1}$$

where $\delta(H_2O)$ is the chemical shift (ppm) of the water protons and T is a temperature (K).

Assignment of proton signals and classification of NOE intensities

Assignment of proton signals were done mainly by use of the NOE connectivities including the sequential NOEs, as explained in General Introduction. H5-H6 proton signals were determined by the COSY spectrum, but proton signals in the sugar rings were not determined because coupling constants were so small that the cross peaks between them could not be observed. Peak picking and the measurement of peak heights from the NOESY spectra was performed by a program NMRPipe/PIPP system (Garrett, 1991; Delaglio, 1995). For the measurements of the cross-peak heights between non-exchangeable protons, the NOESY spectrum at 30 °C, mixing time of 200 msec. was used. Then, for those of the cross-peak heights for exchangeable protons, the NOESY spectrum at 0 °C, mixing time of 200 msec. was used.

They were classified into three groups by their peak heights, strong, medium and weak. In the case of the cross-peaks between the non-exchangeable protons, an internal standard which separates strong and medium peaks was a mean half height of the cross peaks

between H5 and H6 of pyrimidine bases whose distance is fixed at 2.49 Å. That which separates medium and weak peaks was the height of the cross peak between H2 and H1' of A3 whose distance is approximately 5.0 Å.

In the case of the cross peaks between the non-exchangeable protons and the exchangeable ones, an internal standard peak is that of H5-H42 in cytidine at the eleventh residue (C11) whose distance is fixed at 2.48 Å. Its peak height was defined as I(n-e). The threshold between strong and medium peaks is 0.9*I(n-e), and that between medium and weak peaks is 0.35*I(n-e). In the case of the cross peaks between the exchangeable protons, internal standard peak is that of H1-H41 in the G-C pair between G9 and C10 whose distance is fixed at 2.48 Å. Its peak height was defined as I(e-e). The threshold between strong and medium peaks is 1.0*I(e-e), and that between medium and weak peaks is 0.65*I(e-e). Some peaks from the other NOESY spectra were included in the weakest group because the same threshold can not be applied. These cross peaks were converted to the distance constraints where strong, medium and weak peaks correspond to 0~3.5 Å, 0~5.0 Å and 0~6.5 Å distance constraints, respectively. 450 distance constraints were included in the NOE constraints table. Constraints to keep the configurations of Watson-Crick base-pair were also included. They consisted of distance constraints for hydrogen bonds, those which prevent base-pair sharing and base-pair planarities. In respect of the C-U pair, this kinds of constraints were not introduced at all. All of the constraints employed here were summarized in Table 2.1.

NOE based structure calculation

In principle, structure calculations were performed as to satisfy the experimental constraints and geometries around covalent bond (covalent geometries for short) as described in General Introduction. The equation (G.10) was used as a target function. All structure calculations were done by X-POLR ver. 3.1 (Brünger, 1992) based on the NOE constraints and constraints for the base-pairing. The calculation was composed of three steps (Figure 2.2). 1) The initial coordinate was generated from two independent helical single strands then molecular dynamics at 350K without the experimental constraints was performed to those strands to generate random structure. 2) Simulated annealing was performed to a random structure to get a roughly folded structure. In this step, experimental constraints such as NOE distance constraints and constraints for base-pairing were introduced. The base-pair constraints are composed of distance constraints for hydrogen bonds, those of base-pair sharing resistance and base-pair planarities For van der Waals potential, that of square-well type generating only the repulsive force was employed. 3) Refinement was performed using Lennard-Jones potential as a van der Waals potential with larger van der Waals radii of the atoms in the bases than default values. Otherwise, significant short contacts between bases remain. At the same time, the weighting factor of NOE distance constraints was set 1000 times bigger than step 2 to make the structure which strictly satisfies the NOE distance constraints. Step 1 and step 2 were circulated for 500 cycles and 22 structures were accepted at this moment. After

refinement step, energies of above 22 structures and mean rmsd values from the averaged coordinates as functions of the numbers of averaged structures were plotted (Figure 2.8(a), (b)) Between the eighth and ninth structures of the energy order, rmsd values were significant different. The structures which have larger energy values than the ninth structure were rejected, and eight structures were acceptable. These structure calculations and selection protocols were the same as those described by Allain and Varani (1997).

Results

Synthesis of the phosphoramidite form of cytidine and oligoribonucleotides

Synthetic route of the phosphoramidite form of the cytosine labeled at the amino-nitrogen is drawn in Figure 2.1. The yield of each step was presented in Figure 2.1 and the yield of the final product "9" from non-labeled uridine as a start compound is 28%. In the reaction of the final product "9", some aducts form the phosphoramidite reagent which inhibit the oligomer synthesis are also produced. It was purified by a silica-gel column and divided into five fractions in the order of the elution time from the column. The purity of each fraction was checked by ^{31}P NMR (Figure 2.3). Two peaks around 150 ppm correspond to product "9" because there exists chiral center around the phosphorus atom. The peaks around 10 ppm are the aducts which inhibit the following oligonucleotides synthesis. The fractions 1 to 3 were merged together for the use of the following oligonucleotide synthesis by the solid phase phosphoramidite method. Then 580 mg of the final product "9" was obtained. It is enough to perform the solid phase synthesis of oligonucleotides.

The oligomers of each strand was synthesized by the solid phase phosphoramidite method. They were purified as described in "Materials and Methods" as to become a single peak on the chart of HPLC by the analytical grade reversed phase column.

Assignment of the proton signals

In the sequential NOE region between base protons of H6 or H8 and H1' in the sugar ring of the NOESY spectrum in D_2O at 10 °C, sequential NOEs were completely observed through the sequence of each strand (Figure 2.4). They were indicated by solid lines (from the first to 9th residues) and broken lines (from the 10th to 18th residues). The sequential NOEs were observed between base protons and sugar protons H2" or H3' as well. These results strongly suggested the existence of a double stranded structure. In the NOESY spectrum in H_2O at 0 °C, there existed strong intra-residual cross peaks between H5 and the amino-protons in the cytidine bases. These amino-protons also gave cross peaks to the imino-protons of guanosines which were the counterpart of the base-pair. Then residual two imino proton signals in this region were assigned to those of two uridines in U-A pair because they showed cross peaks with the H2 protons of adenosines. There are eight peaks corresponding to the number of the Watson-Crick base-pairs. NOE cross peaks between the imino protons were also consistent with the combination of base-pairs in scheme 1. The expected NOE connectivities

for the double helical structure are presented in Figure 2.5 and all of these NOEs were observed except for those from the amino proton signals of guanosines which were not observed because of the exchange between water hydrogen atoms. The assignment of partial H1' and the imino protons of G2, G8, U12, G13, G15 and U16 were confirmed by the procedure as described in the paper by Heus & Pardi (1991) (Figure 2.5). This method also assume the double helical structure, and the double helical structure of rCU9 was also confirmed.

All of the cross peaks except for several weak ones were assigned consistently. H2" proton signals were confirmed by the cross peaks between H1' and H2" on the NOESY spectra at different temperatures (0 to 30 °C) because H1' and H2" protons are only three bonds apart from each other and always gives cross peaks at any temperatures. H3' proton signals were confirmed by ^{1}H-^{31}P hetero COSY spectrum. In the assignment of H4', H5' and H5" protons, proton signals which showed strong intra-residual cross peaks with H1' were assigned to H4' protons and residual peaks to H5' and H5". These three types of protons were not confirmed because of no other characteristic cross peaks. As a result, most of the non-exchangeable protons were assigned except for a part of H5' or H5" protons. Among the exchangeable protons, the imino proton of U5 at the C-U pair site, the amino protons of G2, G9, G13 and G18 were not observed. Interestingly, a part of the H2' protons of the 2'-hydroxy group were observed. They were confirmed by the NOE cross peaks with H1', H2", H3' of the same residue and base protons of the neighboring residues of the 3'-side. We confirmed that these signals were not those of the amino protons of guanosines which have not been assigned yet. Although there were several signals which seemed to be 2'-hydroxy proton signals, they were not assigned to them. Because there still remained the possibility that those signals arose from the amino protons of guanosines.

For an unambiguous assignment of the amino protons in the C-U pair, ^{15}N-HSQC spectrum was used. Because only the cytidine in the C-U pair was ^{15}N-labeled. It is presented in Figure 2.6 with the NOESY spectrum. Two peaks which correspond to two amino protons were observed on the same chemical shift of the nitrogen which means two protons were located on the same nitrogen atom. In the amino proton region of the NOESY spectrum, cross peaks of intra-residual amino protons in the C-U pair were also observed. Because of very fast exchange with the hydrogen of water molecule, these cross peaks could be observed only when the mixing time was short (50 msec. and 100 msec.). Interestingly, their chemical shifts were greatly shifted to upfield when compared with those of the C-G pairs. This will be discussed in Chapter 3.

The sequential NOEs suggested that U5 and C14 at the C-U site were incorporated in the double helix. As mentioned above, neither imino proton signal of uridine in the C-U pair nor its exchange peak with the hydrogen atoms of water molecules was observed. The imino proton of uridine could be exposed to bulk water. These results are consistent with the crystal structure of the C-U pair in Chapter 1.

NOE distance constraints

For the structure calculation, NOE peak volumes or heights must be quantitate at first. Then the cross peak heights were quantitated by a program NMRPipe/PIPP system (Garrett, 1991; Delaglio, 1995). As described in General Introduction, the NOE intensity correlates with the corresponding distance. When the three dimensional structure of the sample is assumed beforehand, the predicted structure is confirmed by comparing the NOE intensities with the corresponding distances. In the case of double helical RNAs, two major conformations of A- and A'- forms is known as mentioned in Chpater 1. Unfortunately, they are indistinguishable from each other by any NMR data. Therefore, A-form double helical structure of rCU9 was generated as the structure of rCU9. The correlation between the observed peak heights and the predicted distances from the A-form structure were examined to confirm whether double stranded A-form structure was plausible for that of rCU9. From the equation G.7, the NOE peak volume is proportional to r^{-6} but its width is not affected by the peak height, thus, NOE peak height is proportional to r^{-6}. In Figure 2.7, the sixth roots of NOE peak heights ((NOE peak height)$^{1/6}$) are plotted to the reciprocal of the predicted distances (1/r) (Figure 2.7). This scattergram shows approximately linear correlation between NOE peak heights and predicted distances. It was confirmed that the assumption of the structure of rCU9 was almost plausible. However, their strict correlation of the equation G.7 was not maintained in the following two points: the fitting line did not pass over the origin and data were considerably deviated from the fitting line. For the reasons of the first point, the equation G.7 work out only when the mixing time of the NOESY spectrum is short enough. In this case, mixing time of 200 msec is too long. It is known that the evolution of NOE depend on the residual magnetization which is not transferred by the NOE. This effect makes the slope of the fitting line smaller as presented in Figure 2.7. For the second problem, it is mainly due to the inaccuracy of the quantitation of peak heights from severe peak overlaps. Secondly, in the equation G.7, no spin diffusion which causes the loss of NOE intensities of strong peaks is considered. This also affects the first problem and makes the slope of the fitting line smaller as well. Anyway, it should be noted that the correlation between the NOE peak heights and the distance does exist. Since their correlation is not strict for the macromolecules in general, NOEs were classified into approximately three groups. Then, appropriate upper limits of the distance constraints are given to each category in practice.

To determine the appropriate thresholds which divide the cross peaks into three groups can be derived by using the scattergram of Figure 2.7. The H5-H6 cross peak heights were often used for the internal standard. However, if the mean value of the H5-H6 cross peak heights is selected as the lower limit of "strong" peaks, the number of peaks in the category "strong" will become too small. Then the half value of the mean of the H5-H6 cross peak heights is selected as the lower limit of "strong" peaks. According to this threshold, the maximum inter-proton distance in the category "strong" of Figure 2.7 is approximately 3.5 Å.

Therefore, the maximum inter-proton distances for the category "strong" was defined as 3.5 Å. As the lower limit of the medium strength peaks, the peak height of the cross peak between H2 and H1' of A7 was employed, because it seems to have no spin diffusion pathway and be the longest distance (4.7 to 5.3 Å) among all the cross peaks due to the direct dipole-dipole interaction. From these threshold, there are some cross peaks in the shaded area of Figure 2.7 which violate the upper limit of the distance constraints, but their violations are small and seem to reflect the sequence dependent local geometry. They do not matter. In the case of the cross peaks from the exchangeable protons, the classifications of NOE distance constraints were checked by the scattergrams as mentioned above. In the calculation of nucleic acids structures, constraints to keep the Watson-Crick base-pair geometry are introduced in general. It is plausible to introduce such kind of constraints because imino protons of the Watson-Crick base-pair was observed and they gave cross peaks to the protons of the counterpart of base-paired residues: those of guanosines to the amino protons of cytidines and those of uridines to the amino protons and H2s of adenosines. For the C-U pair, however, this kinds of constraints were not introduced at all. All of the experimental constraints are summarized in Table 2.1. Structure calculations were performed as to satisfy these experimental constraints and covalent geometries.

NOE based structure calculation

The correlation between NOE peak heights and the distances above suggests that the structure of rCU9 is a double helix. However, simulated annealing method was employed to avoid local minimum structure. To eliminate the record of the A-form helical structure from the initial ones, single stranded A-form helical structure were randomized by the molecular dynamics without experimental constraints before the simulated annealing. From 500-time-repeated structure calculations, 22 structures were accepted out of 500 initial structures. Then the refinements were performed and eight structures were accepted from 22 structures. Energies of 22 structures after refinements are presented in Figure 2.8(a). They are aligned in the order of total energies. It must be noted that these energy terms are not realistic energies, but just target functions of the least square for the structure calculation. Total energy is a summation of all the energy terms. Bond, angle, vdw represent the energies of bond length, bond angle and Lennard-Jones potential. Improper includes pseudo energies of chirality and planarity of aromatic rings. NOE and plan represent pseudo energies of the NOE distance constraints and base-pair planarity. The constraints for the hydrogen bonding and the sharing resistance are basically the distance constraints, thus, they were included in the NOE distance constraints. In Figure 2.8(b), mean rmsd values from the averaged coordinates are plotted as a function of the number of averaged structures. They were calculated from the structure with the lowest total energy. Total energies showed almost flat until the fourteenth structure. However rmsd plots showed large fluctuations at the fifth and ninth structures. These fluctuations means that the last structure included for averaging has different structural feature from the former ones.

In the case of fifth structure, rmsd values are decreased again after it and the structures from the fifth to eighth ones are accepted. However, after the ninth structure rmsd values are basically increasing and they were eliminated from the acceptable structures. Then, accepted eight structures were fitted to each other as presented in Figure 2.9. Their mean standard deviations of the coordinates from those of the averaged structures of the eight calculated structures was 1.77 Å. Covalent geometries of the final structures were checked. Their deviations from the standard bond lengths and bond angles were 0.00675 Å and 1.25°, respectively. They were summarized in Table 2.3.

All of the eight structures were converged to the right-handed double helical structures. The end of the helices were deviated each other because the some calculated helices show kinks around the C-U site. Due to the lack of the NOE data of the minor groove side at the C-U site, this side tends to open.

One calculated structure of rCU9 is presented in Figure 2.10(a), where highlighted lines indicate the C-U pair. In Figure 2.10(b), the superposition of the C-U pair is presented. In Figure 2.10(c), two examples of the C-U pair are chosen from the eight calculated structures. One is an example which form a hydrogen bond between the amino group of cytidine and the keto-oxygen of uridine (Figure 2.10(c) lower), the other is the one which can not form any hydrogen bond (Figure 2.10(c) upper). Through the eight calculated structures, the C-U pair was stacked between the neighboring base-pairs in the double helix. There is no evidence for the presence of other conformer where the C and U base rings are flipped-out from the double helical structure. It is very interesting that such a thermally unstable pair like C-U prefers the stacked conformer at the room temperature and the flipped-out conformer could not be detected.

The structures of the C-U site were not converged very much. This is because the precision of the structure determination by NMR is limited because of large tolerances of distance constraints. However, all the χ angles of U5s and C14s in the eight calculated structures showed the anti-conformation. The χ angle around the glycosyl bond (the bond between the sugar ring and the base) is one of the important structure parameters which determine the relative orientation of the base from the sugar ring. In the case of the anti-conformation of pyrimidine bases, C2 atom in the base moiety is located at the far side for O4' of the sugar ring. This result is consistent that the crystal structure of the C-U pair showed anti-conformation and the pyrimidine nucleosides strongly prefer anti-conformation. The mean value of the distance between the amino nitrogen of C14 and the keto-oxygen of U5 (N-O distance) is larger than the hydrogen-bonding distance and the mean value of the distance between C1' of C14 and U5 (C1'-C1' distance) is 10.1 Å with the standard deviation of 0.39 Å. This value is different from that in the crystal structure of C-U pair. However, the lower structure in Figure 2.10(c) is similar to the one in crystal (Figure 1.5(b)): N-O distance of the lower structure in Figure 2.10(c) is 3.2 Å (2.8 Å in crystal) where a hydrogen bond can be formed, its C1'-C1' distance is 11.4 Å (11.9 Å in crystal). To examine whether the

geometry of the crystal structure satisfies the NOE distance constraints, structure calculation was performed with additional constraints for the formation of the C-U pair as in crystal. The result showed that the crystal structure of the C-U pair was also acceptable for the NOE constraints. The absence of the imino proton signal of U5 at the C-U site also suggests that the C-U in solution adopts the same conformation as in crystal. The further analysis of the C-U pair will be studied in Chapter 3.

Discussions

Until the crystal structures of the C-U pairs were solved, the pyrimidine-pyrimidine base-pair was not known to exist. Since the size of the pyrimidine base rings are small, the backbones of the pyrimidine-pyrimidine base-pair was said to be located closer than those of the Watson-Crick base-pair. It was thought that this kind of distortion makes a double helix unstable. However, the crystal structure showed that the C-U pair could be incorporated into a double helix with small distortions. Rather, the C1'-C1' distance of the C-U pair was longer than that of the Watson-Crick base-pair. At the moment, the detailed structure of the C-U pair in solution was not determined. However, it was found that at the room temperature or lower the C-U in solution prefers the stacked conformer, and their χ angles were in the anti-conformation. This finding is also interesting, because The C and U bases in the mismatch site have no strong affinity each other and their base-pair seems to be unstable as an isolated base-pair in solution. In conclusion, the C-U pair in RNA molecules do not disrupt the continuity of a double helix. This is a very important point to predict the secondary structure of large RNA molecules. The putative C-U pair site in *Eschelichia coli* 23S ribosomal RNA is expected to form a base-pair. To clarify this, cooperativity of the denaturation process also must be studied. If there is a cooperativity of the denaturation process to the other part of a double helix as is the case of the Watson-Crick base-pair, the probability of the existence of the C-U pair will be increased. The stacked conformation of the C-U pair is consistent that the C-T pair in DNA was claimed to be stacked in a duplex DNA in solution (Boulard *et al.*, 1997). (In DNA, T (thymidine) is used instead of U (uridine) but the conformation of the double helix is B-form usually.) The remaining problem is to clarify the precise configuration of the C-U pair in solution. For this purpose, the presence of hydrogen bond between the C-U pair should be examined.

It was difficult to define the precise structure of the C-U pair in solution, because the imino proton signal of U5 in the C-U pair was not observed. In the Watson-Crick base-pair region of rCU9, there are many NOE cross peaks among the imino protons as claimed by Heus and Pardi, (1991). The distance constraints on the exchangeable protons are efficient for the convergence of the structure (Allain & Varani, 1997). In fact, the region where the Watson-Crick base-pairs are formed was well converged. It was impossible to determine the fine structure of the base-pairs whose imino protons were not observable by using the NOE distance constraints with large tolerances.This kind of difficulties existed even in the case of

the U-U pair whose imino proton signals could be obseved (Wang *et al.*, 1996). There are two ways to overcome this problem. 1) Back calculation refinement using a full relaxation matrix assuming a single correlation time. 2) Incorporation of the other experimental constraints such as constraints on the hydrogen bond, torsion angles and so on. The first way was examined by Kojima (Kojima, 1995). The NOE data derived from different mixing times led to different structures. Their inclination values of the base-pairs were systematically changed. This means that an assumption of a single correlation time may not be correct. Therefore, an accuracy of the structure derived from this method is questionable. However, it should be noted that the structure derived from all the NOE data of different mixing times gave a plausible structure (Kojima, 1995). This might suggest that the systematic errors in NOE data can be canceled out when NOE data from a certain range of the mixing times are used. At this moment, this method must be performed carefully and some improvement is required. The safer way is the second one where the other experimental data must be incorporated. From this requirement, the existence of the hydrogen bond must be studied.

Table 2.1. Number of experimental constraints

(a) NOE distance constraints

NOE intensity	strong (3.5Å)	medium (5.0Å)	weak (6.5Å)	total
E ——— E	22	23	9	54
E ——— N	5	15	72	92
N ——— N	50	205	49	304
total	77	243	130	450

(b) Constraints for Watson-Crick base-pairs

hydrogen-bond	22
sharing resistance	32
base-pair planarity	8 base-pairs

(c) Internal standards for NOE distance constraints

lower limit of "strong" category for E - E	$I_{[H1\text{-}H41(G9\text{-}C10)]}$
lower limit of "medium" category for E - E	$0.65*I_{[H1\text{-}H41(G9\text{-}C10)]}$
lower limit of "strong" category for E - N	$0.9*I_{[H5\text{-}H42(C11)]}$
lower limit of "medium" category for E - N	$0.35*I_{[H5\text{-}H42(C11)]}$
lower limit of "strong" category for N - N	$I_{[H2\text{-}H1'(A3)]}$
lower limit of "medium" category for N - N	$0.5*I_{[H5\text{-}H6(pyrimidines)]}$

"E" and "N" represent exchangable and non-exchangable protons, respectively. $I_{[H1\text{-}H41(G9\text{-}C10)]}$, $I_{[H5\text{-}H42(C11)]}$ and $I_{[H2\text{-}H1'(A3)]}$ represent heights of cross peaks between H1 of G9 and H41 of C10, between H5 and H42 of C11 and between H2 and H1' of A3, respectively. $I_{[H5\text{-}H6(pyrimidines)]}$ represents a mean peak height between H5 and H6 of all the pyrimidine bases. The upper limit of the distance constraint for each category are presented. For the "strong", "medium" and "weak" peaks, their distances are 3.5 Å, 5.0 Å and 6.5 Å, respectively.

Table 2.2 Assignment of the proton resonances of rCU9

	H1/H3	NH1	NH2	H2'	H5/H2	H6/H8	H1'	H2''	H3'	H4'	H5'/H5''	
C1	-	8.23	6.99		5.92	7.95	5.53	4.48	4.49	4.25	3.86	3.96
G2	12.43				-	7.73	5.69	4.64	4.68	4.47	4.15	
A3	-	8.33	6.66		7.73	7.87	5.92	4.50	4.56	4.47	4.13	
C4	-	8.34	7.04	6.28	5.15	7.27	5.33	4.03	4.27	4.31	3.98	4.42
U5		-	-		5.28	7.54	5.49	4.47	4.21	4.34	3.99	4.41
C6	12.35	8.35	7.18		5.75	7.90	5.47	4.39	4.57	4.41	4.06	4.34
A7	-	7.90	6.26		7.03	8.05	5.86	4.65	4.73	4.40	4.10	4.51
G8	13.15	8.41	6.03		-	6.99	5.42	4.33	4.28	4.36	3.97	
G9	13.31				-	7.16	5.74	3.96	4.12	4.34	3.93	4.38
C10	-	8.43	7.18		5.97	8.08	5.47	4.41	4.43	4.28	3.87	4.01
C11	-	8.51	7.04		5.57	7.95	5.55	4.36	4.54	4.41	4.08	
U12	13.51	-	-		5.40	7.81	5.47	4.49	4.54	4.35	4.08	4.41
G13	12.35			6.45	-	7.68	5.69	4.04	4.61	4.38	4.08	
C14	-	6.51 *	6.11*		5.00	7.47	5.59	4.48	4.30	4.28	4.42	
G15	13.21	8.31	6.23		-	7.35	5.59	4.49	4.29	4.39	4.02	4.24
U16	14.56	-	-		4.98	7.79	5.48	4.42	4.44	4.36	4.02	
C17	-	8.34	6.93	6.78	5.57	7.73	5.54	4.26	4.42	4.34	4.02	
G18	12.81				-	7.57	5.73	4.03	4.22	4.15	3.98	4.38

Horizontal bars mean that there is no corresponding proton. The brack boxes mean they could not be assigned. For the assignment of non-exchangeable protons, NOESY at 25 °C was employed. For exchangeable protons, NOESY at 0 °C was used. Their chemical shifts were calibrated as described in "Materials and Methods". For H5' and H5" protons, stereo specific assignments were not carried out. Asterisks for the amino protons means that specific assignment was not carried out.

Table 2.3 Quality of the structure

(a) Rmsd values from the standard values of bond lenghs and bond angles, and those from NOE distance constraints for each structure.

	bond length(Å)	bond angle(°)	NOE(Å)
coodinate1	6.60×10^{-3}	1.20	5.53×10^{-4}
coodinate2	6.73×10^{-3}	1.23	5.68×10^{-4}
coodinate3	6.67×10^{-3}	1.23	5.58×10^{-4}
coodinate4	6.87×10^{-3}	1.26	5.77×10^{-4}
coodinate5	6.71×10^{-3}	1.24	5.55×10^{-4}
coodinate6	6.78×10^{-3}	1.24	5.56×10^{-4}
coodinate7	6.82×10^{-3}	1.29	5.80×10^{-4}
coodinate8	6.86×10^{-3}	1.28	5.46×10^{-4}
average	6.75×10^{-3}	1.25	5.62×10^{-4}
SD	9.6×10^{-5}	0.028	1.21×10^{-5}

(b) Summary of the structure analysis

rmsd from the mean structure	1.77Å
number of accepted structures (number of initial structures)	8(500)
number of violations for NOE (> 0.01Å)	0

(a) Average means mean rmsd value of all the coordinates. SD means standard deviation from mean value. Accepted coordinates are numbered sequencially in the order of energy values. (b) If the distance from the calculated structure is longer than the upper limit of the NOE distance constraints by 0.01 Å, it is regarded as violations for NOE (> 0.01Å).

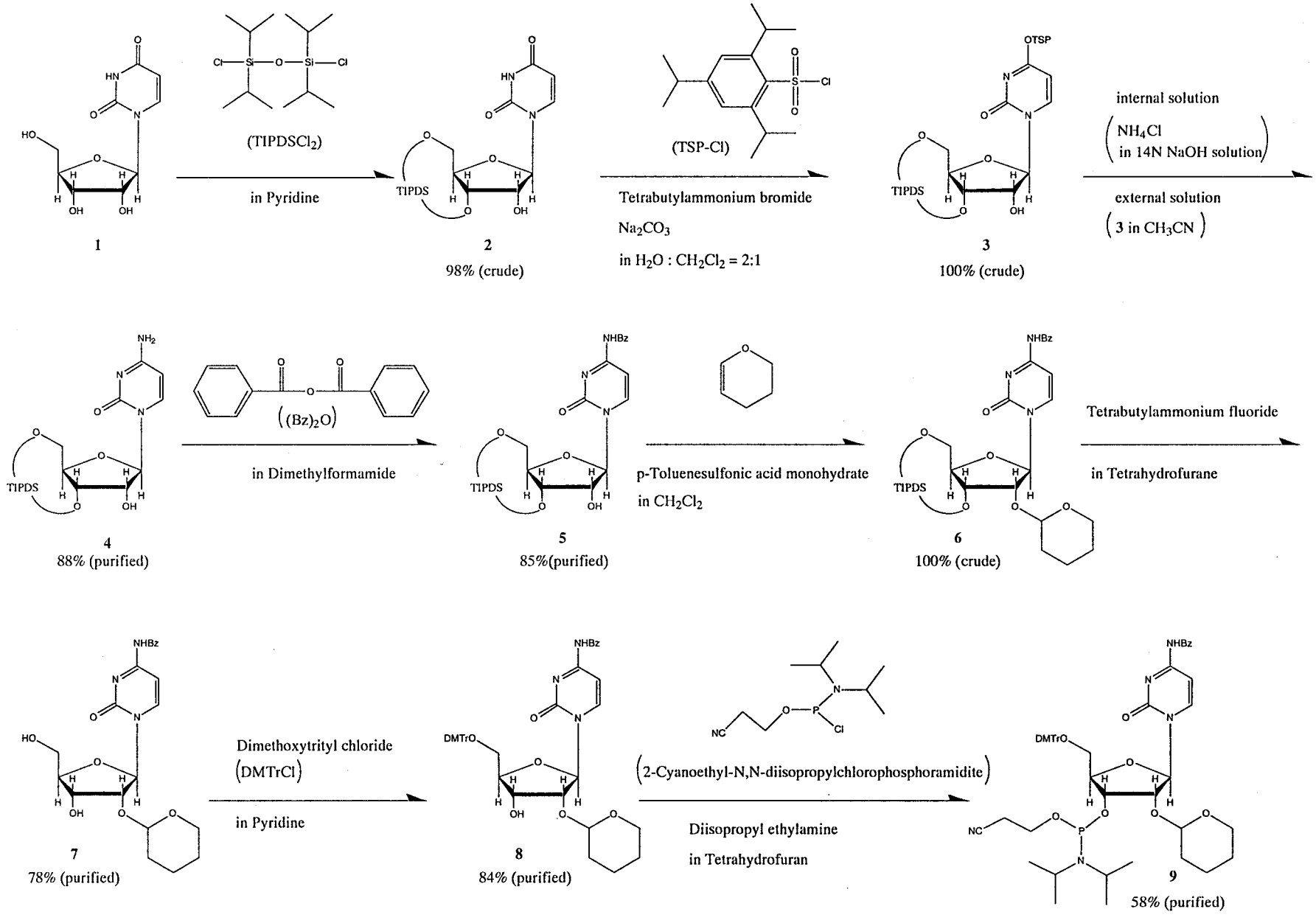


Figure 2.1 Route of the synthesisof amidite form of the cytidine.

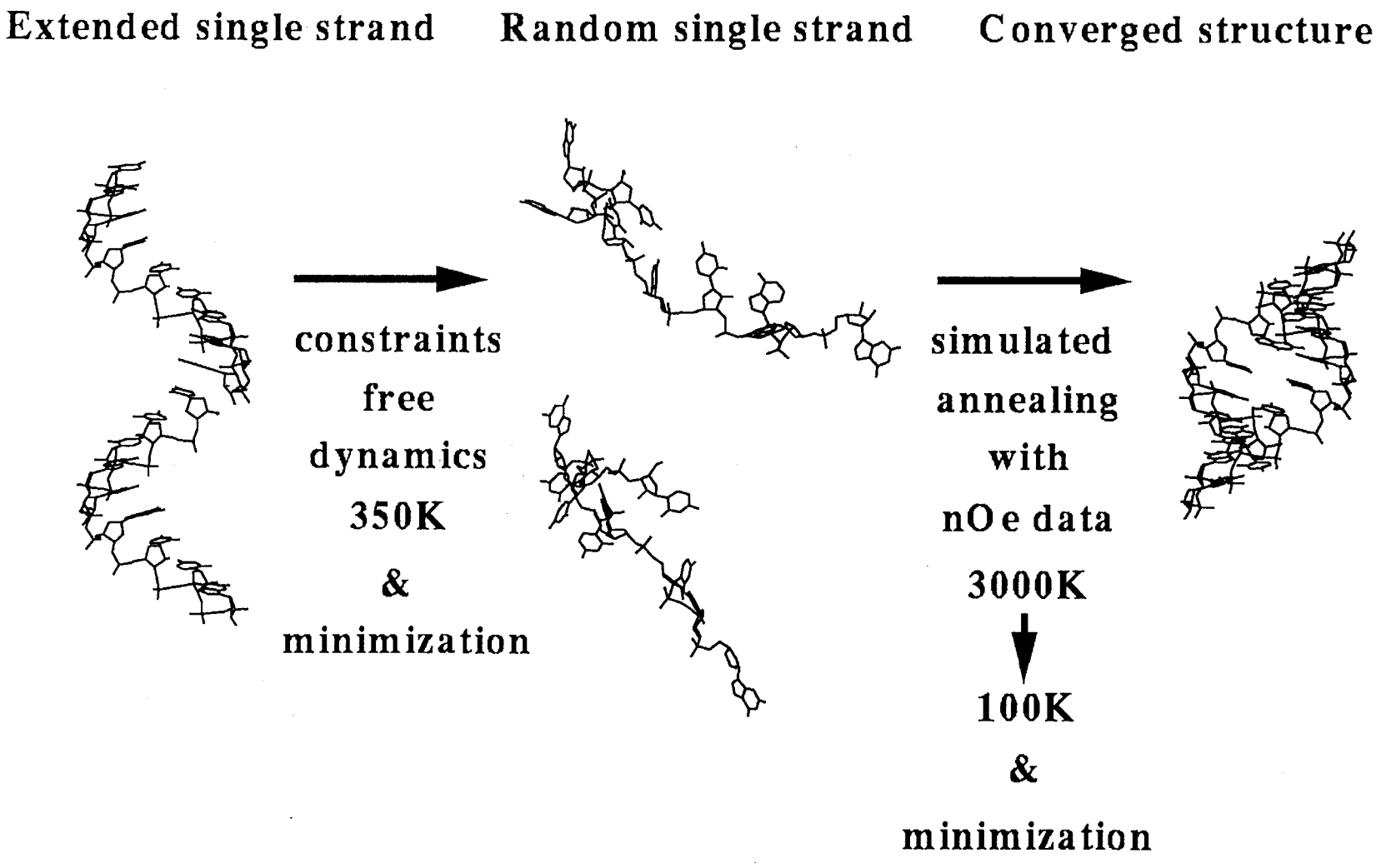


Figure 2.2. Protocol for the structure calculation.

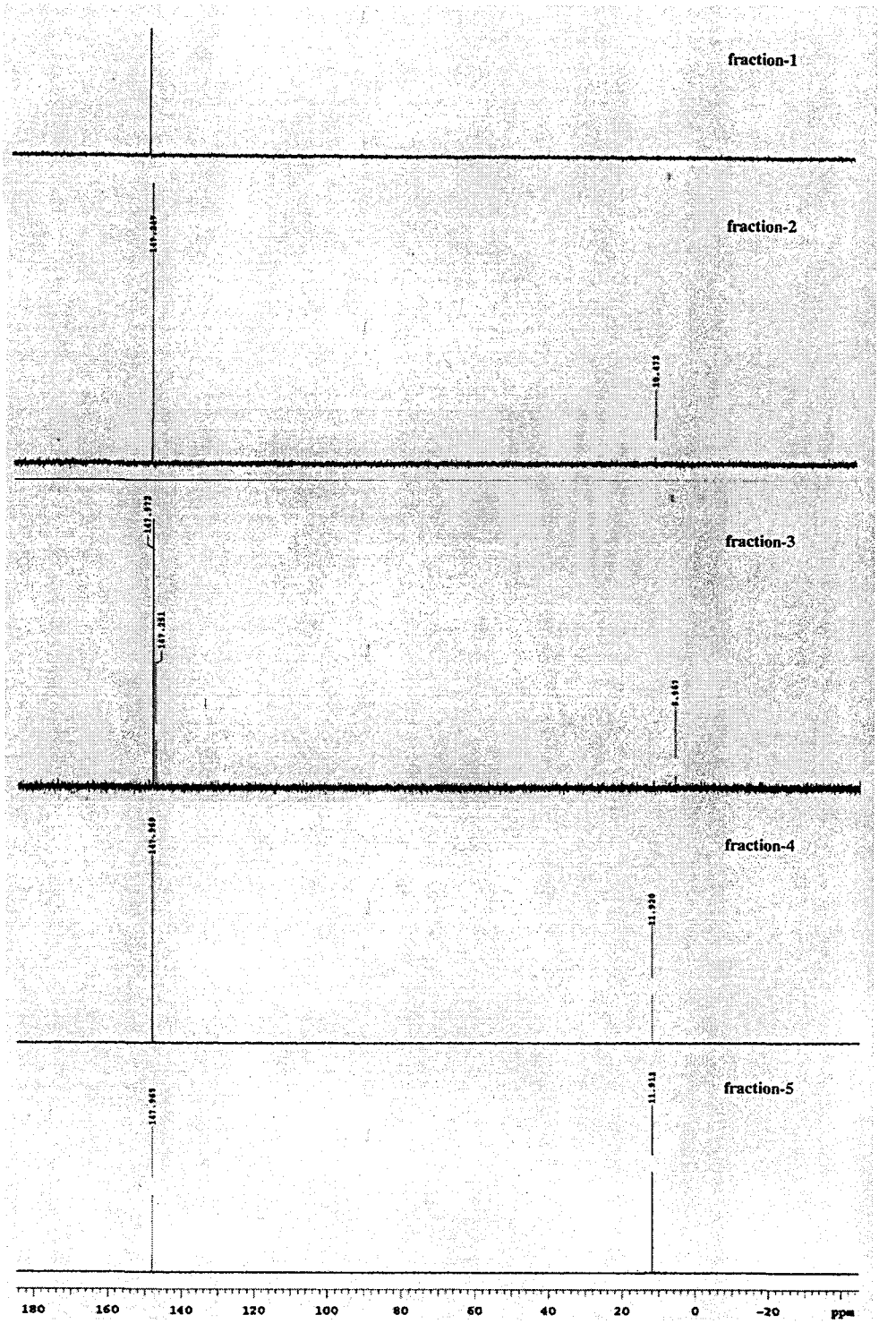


Figure 2.3. ^{31}P-NMR spectrum of each fraction of the silica-gel column. Spectra are presented in the order of elution time.The peaks around 150 ppm arise from the products, and those around 10 ppm come from the byproducts.

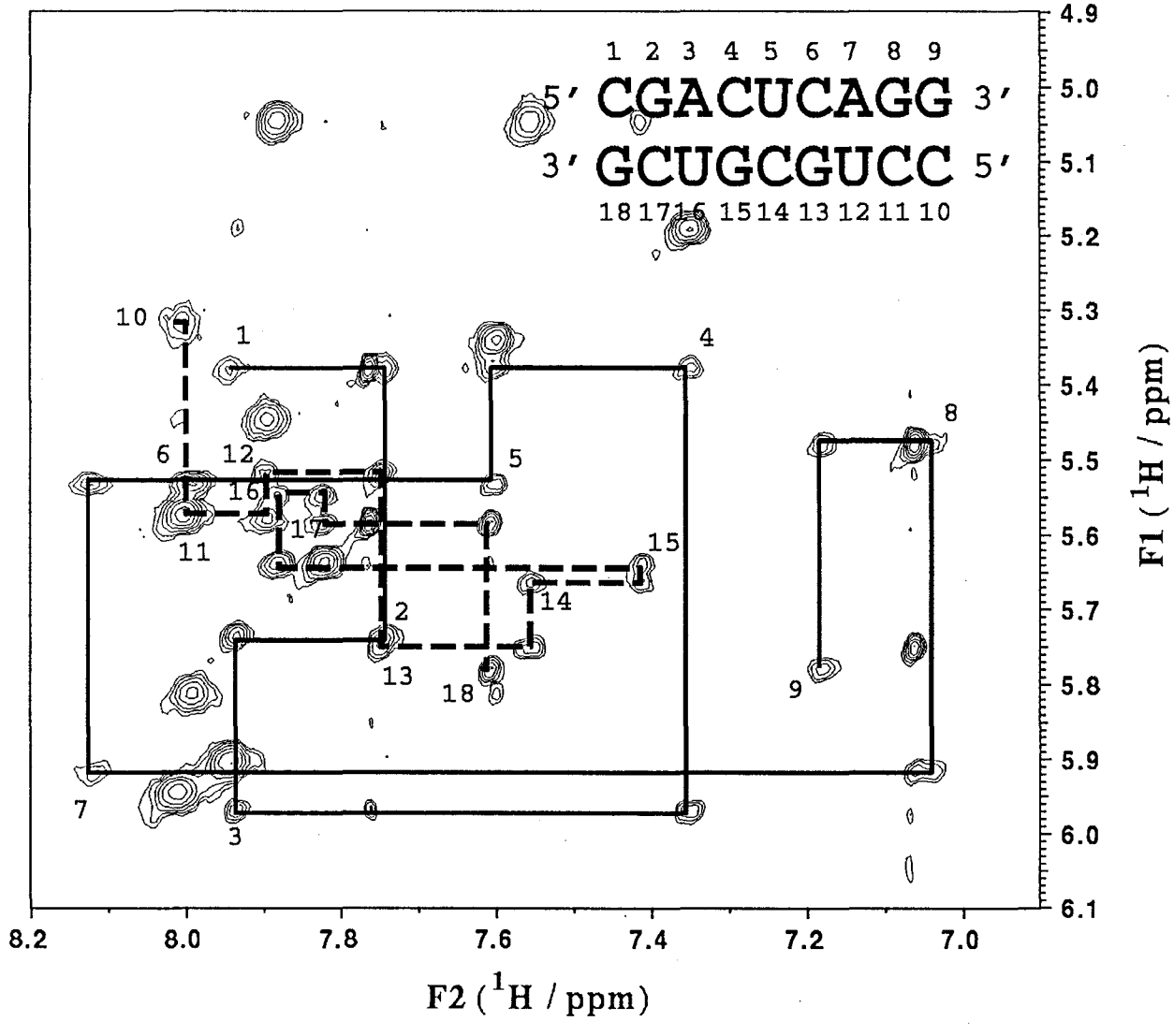
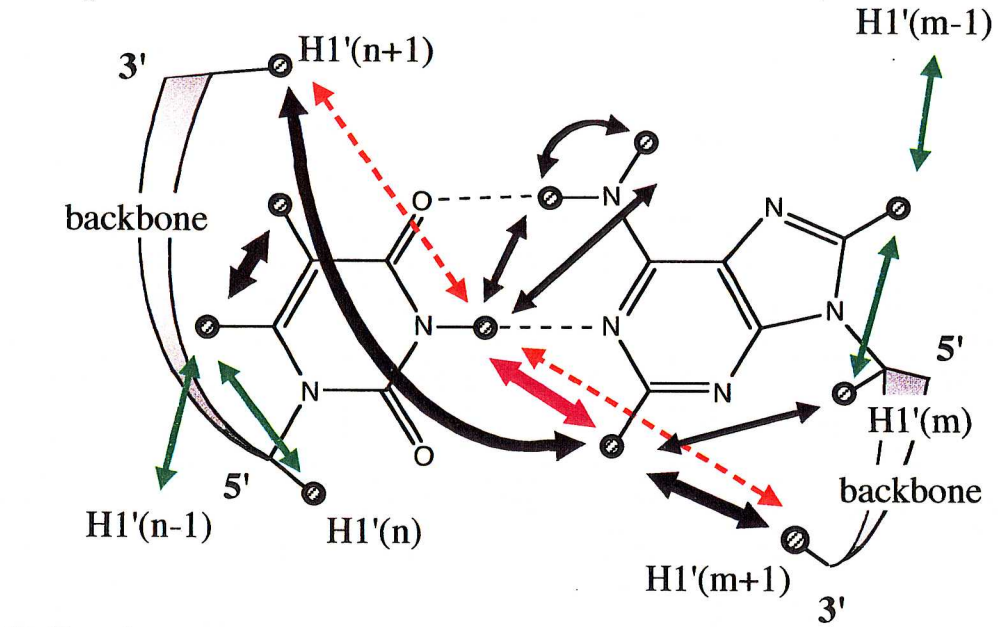


Figure 2.4. Sequential NOE (H1'-H6,8) region of 2D-NOESY spectra. Those of upper and lower strands are indicated as solid and broken lines, respectively.

U-A pair



C-G pair

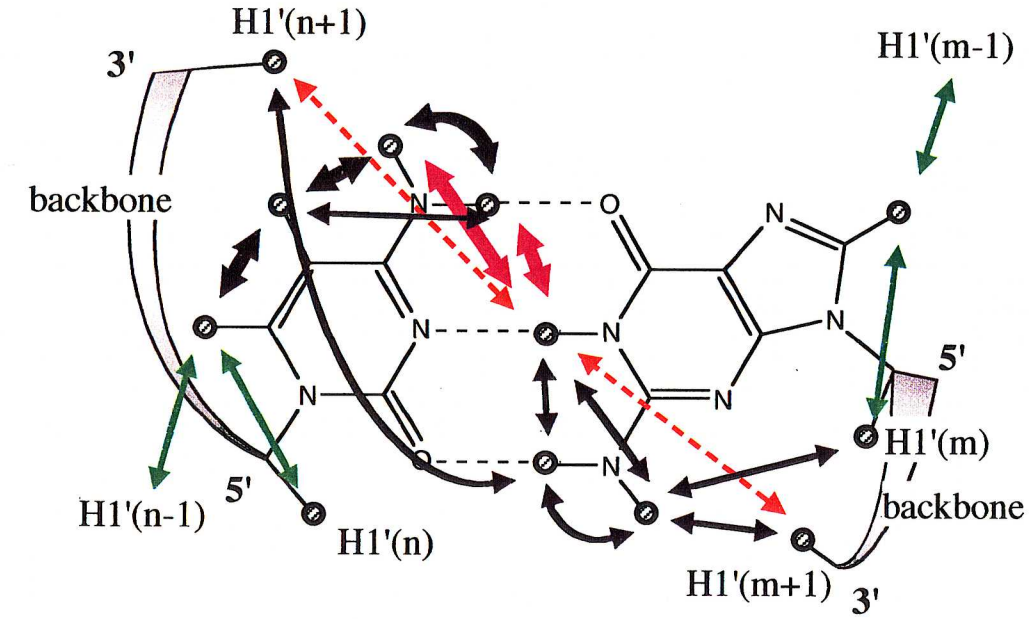


Figure 2.5 Expected NOE connectivities are presented with arrows. Protons are presented in shaded circles and hydrogen bonds are in black thin broken lines. The thickness of arrows represents the strength of NOEs. Sequential NOEs are presented in green arrows. Essential NOEs to prove the presence of base-pairing are presented in magenta arrows. NOEs through the diffusion proposed by Heus and Pardi (1991) are presented in red broken lines. H1'(n), H1'(n-1) and H1'(n+1) represents intra-residual H1' proton, H1' proton of the neighboring residue at 5' side and H1' proton of the neighboring residue at 3' side, respectively.

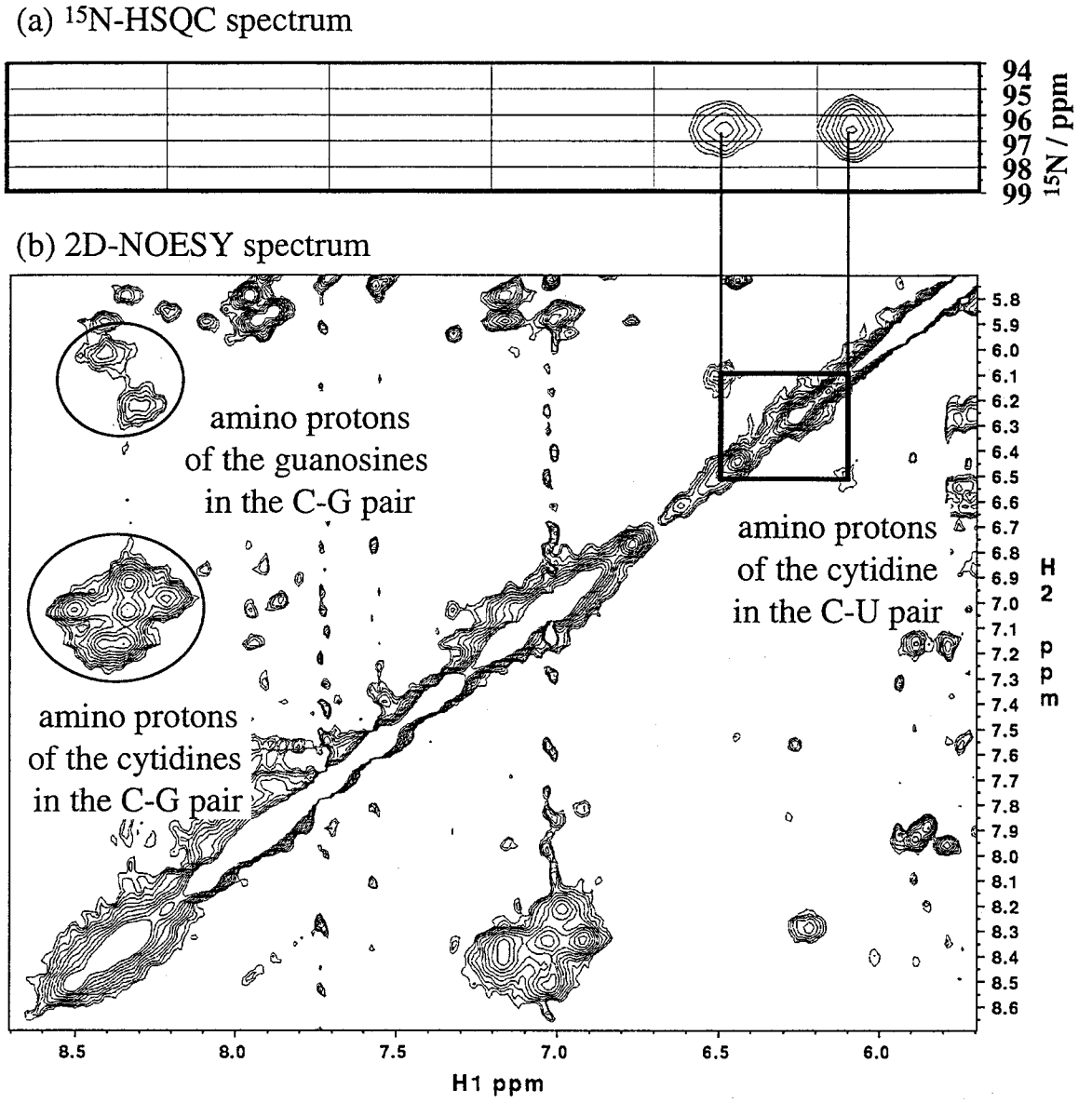


Figure 2.6. Assignment of the amino protons of the cytidine in the C-U pair. (a) ^{1}H-^{15}N-HSQC spectrum and (b) 2D-NOESY spectrum of the amino proton region. Intra-residual cross peak of the amino protons in the C-U pair are connected by bold lines. Their signals in both spectra are connected by thin lines. The Amino protons in guanosines and cytidine in the C-G pair are surrounded by obals.

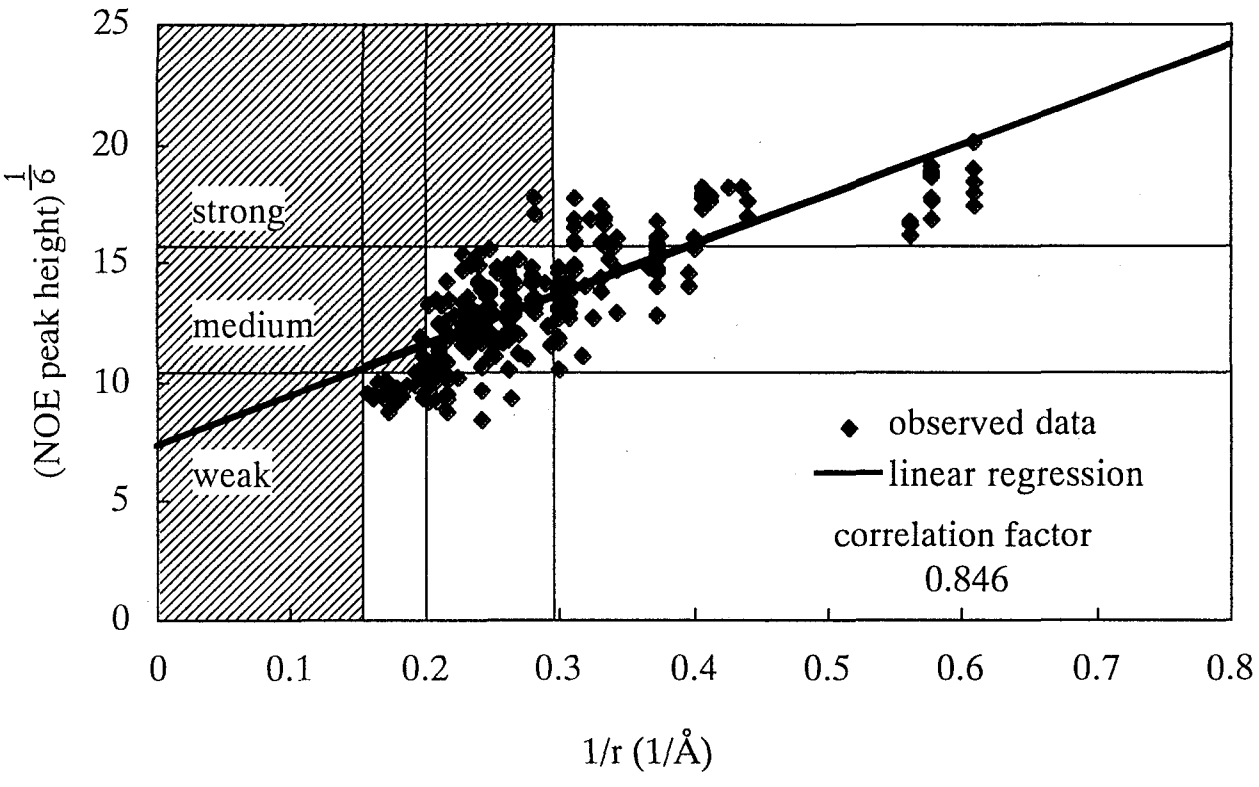


Figure 2.7 Correlation between NOE and distance. Their correlation factor is 0.846, and best-fit line is presented as a bold line. Upper horizontal line is the threshold between strong and medium peaks. Lower horizontal line is the threshold between medium and weak peaks. Vertical lines are reciprocals of corresponding distance constraints for strong (3.5 Å), medium (5.0 Å) and weak (6.5 Å) peaks, respectively. The peaks in the shaded area are out of the maximum distances the categories.

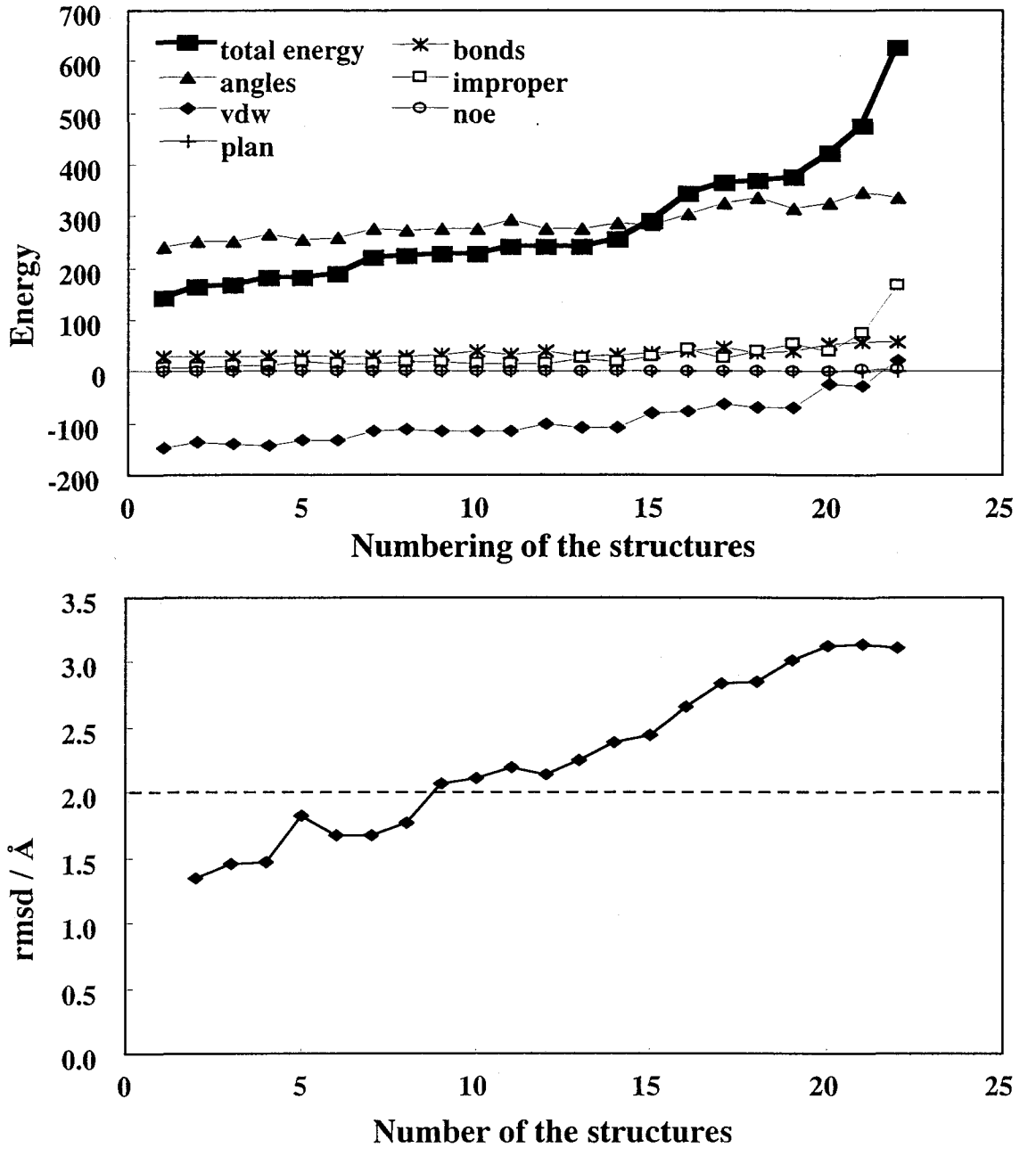


Figure 2.8 (a) Energy value of each structure as a target function of the least square is plotted. Total energy is a summation of all the energies. Bond, angle, vdw represent the energies of bond length, bond angle and Leonard-Jones potential. Improper includes pseudo energies of chirality and planarity of aromatic rings. NOE and plan represent pseudo energies of NOE distance constraints and base-pair planarity. (b) Mean rmsd values from the averaged coordinates are plotted as a function of the number of averaged structures. Coordinates were averaged from the structure which is the lowest total energy. In the case of "5", 5 lowest energy structures were averaged and mean rmsd value from its averaged coordinate was calculated. Broken line repersents the upper limit of acceptance.

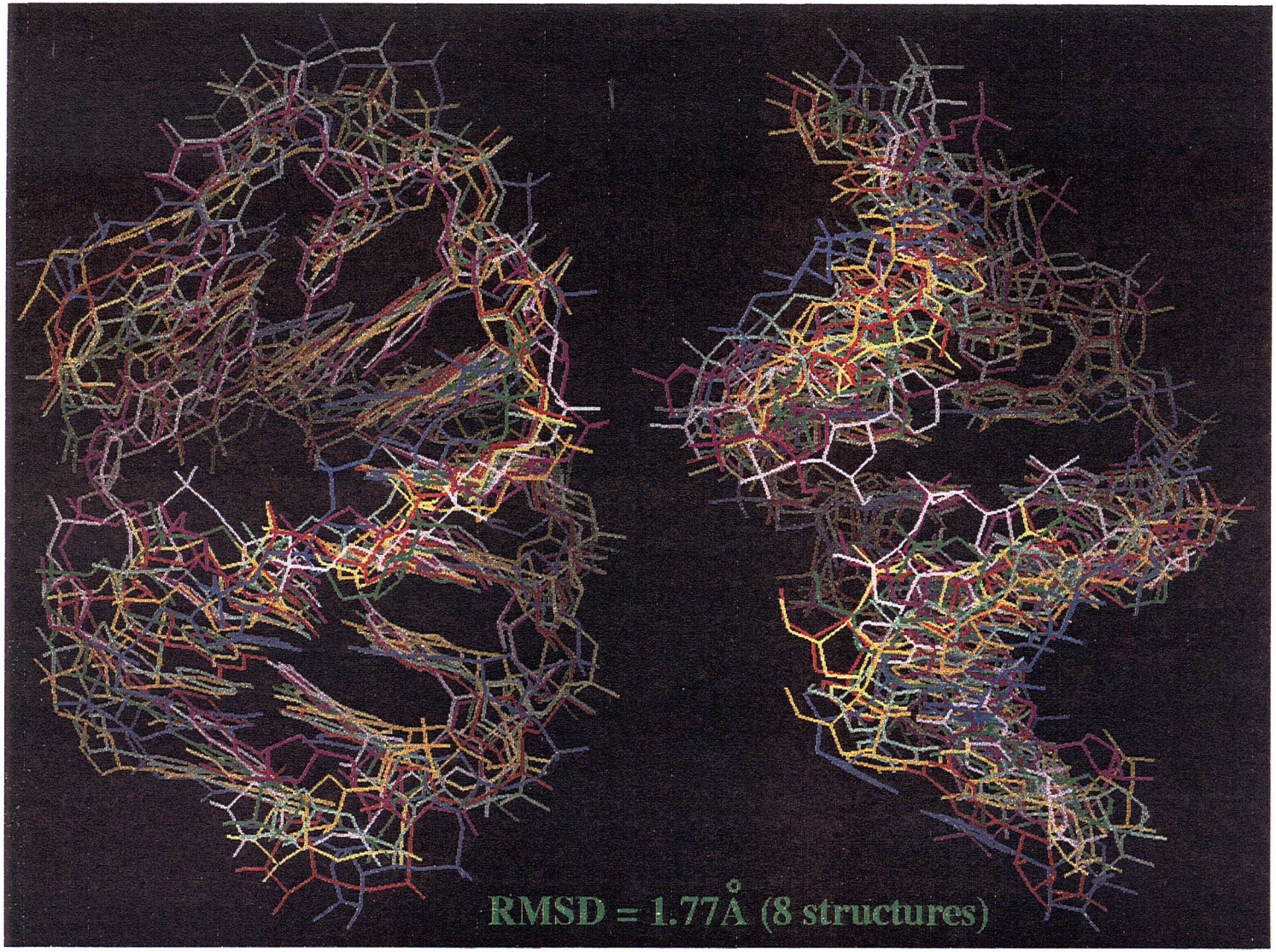


Figure 2.9 Superposition of eight structures are presented (right) View from the major groove (left)

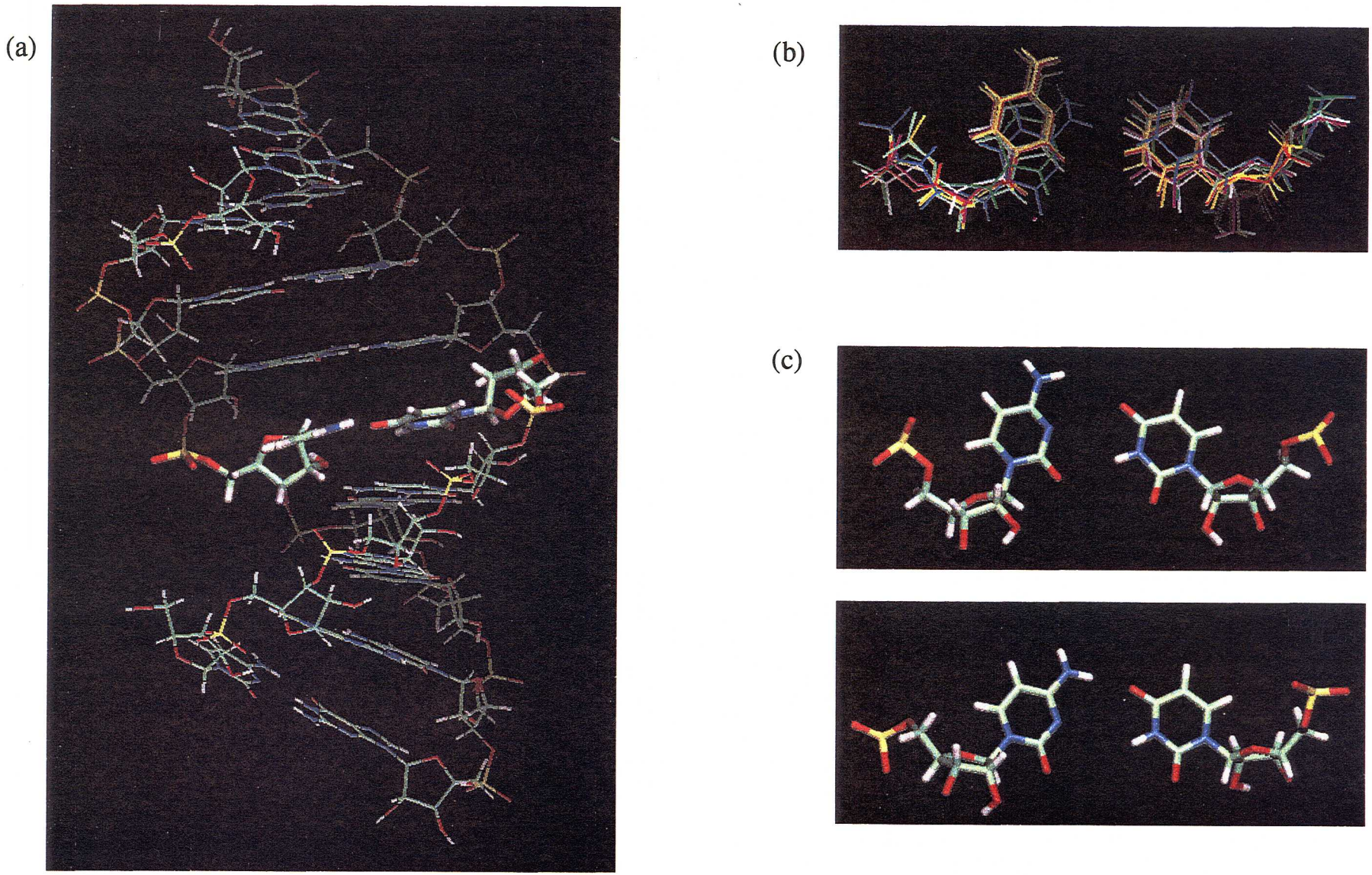


Figure 2.10 (a) One of the eight calculated structure of rCU9. The C-U pair is indicated as bold lines. (b) Superposition of eight structures of the C-U pair. (c) Two arrangements of the C-U pair. (upper) An example which can not form a hydrogen bond. (lower) An example which can form a hydrogen bond.

Chapter 3

Physicochemical properties of the C-U pair in solution

Abstract

Thermal denaturation process of the oligonucleotide containing the C-U pair was studied by the temperature profiles of the amino nitrogen chemical shifts and CD spectra. The temperature profile of the amino nitrogen chemical shifts showed a sigmoidal curve in the transition from a duplex to single strands. This profile implies the existence of the hydrogen bond of the amino group in the C-U pair, in comparison with the temperature profiles of the duplex with the C-G pair and cytidine itself. Furthermore, transition curve of the chemical shift of the amino nitrogen in the C-U pair was also explainable by the thermodynamic parameters obtained from the thermal denaturation experiment with CD spectra. This implies that the dissociation process of the C-U pair is highly cooperative with the other part of the duplex. The duplex containing a single C-U pair could be approximated by the two-state model between a duplex and single strands through concentrations from 2 mM to approximately 1 μM. This is the same as the denaturation process of the Watson-Crick base-paired duplex. Finally, the arrangement of the C-U pair was reinvestigated. The NOE peak heights were better correlated with the distances of the same arrangement as in crystal. All the data strongly suggest the presence of the C-U pair in the double helical oligoribonucleotides in solution.

Introduction

In the previous chapter, it was found that the C-U pair was stacked in the double helix at room temperature. In this chapter, it was studied whether the C-U pair is still stacked in the double helix or flipped-out at higher temperature, and whether any hydrogen bond is really formed between the amino nitrogen of cytidine and the keto-oxygen of uridine.

To study the process of the thermal denaturation of the oligonucleotide, the temperature profiles of the circular dichloism (CD) spectrum and ultraviolet (UV) absorption spectrum are often used. These methods are suitable to know the global equilibrium between duplexes and denatured single strands. On the contrary, the temperature profile of the NMR chemical shifts was suitable to observe the equilibrium of each residue of an oligonucleotide. There were several experiments on the behaviors of the non-exchangeable protons (Pardi *et al.*, 1981; Williams *et al.*, 1989), and exchangeable protons (Pardi *et al.*, 1981; Leroy *et al.*, 1988) of DNA. The non-exchangeable protons are observable after denaturation, therefore the whole process of denaturation is followed. However, the exchangeable protons which are directly involved for the formation of hydrogen bonds are not observed at higher temperatures because of the fast exchange with the hydrogen atoms of water molecules. To overcome this problem, the temperature profile of the ^{15}N chemical shifts was used. ^{15}N chemical shift is known to give a sigmoidal transition to upfield as the hydrogen bond was broken (Gao & Jones, 1987; Goswami *et al.*, 1993; Gaffney *et al.*, 1993; Gaffney *et al.*, 1995; Zhang *et al.*, 1997). ^{15}N chemical shift is sensitive to the existence of protons on it. Therefore the pKa value of each nitrogen atom was also determined by ^{15}N chemical shift (Wang *et al.*, 1991; Büchner *et al.*, 1978).

The patterns how the signals look like can be classified in three cases, by the exchange rate (k_{ex}) between the duplexes and denatured single strands,

$$k_{ex} >> 2\pi\Delta\upsilon \tag{3.1}$$

$$k_{ex} \cong 2\pi\Delta\upsilon \tag{3.2}$$

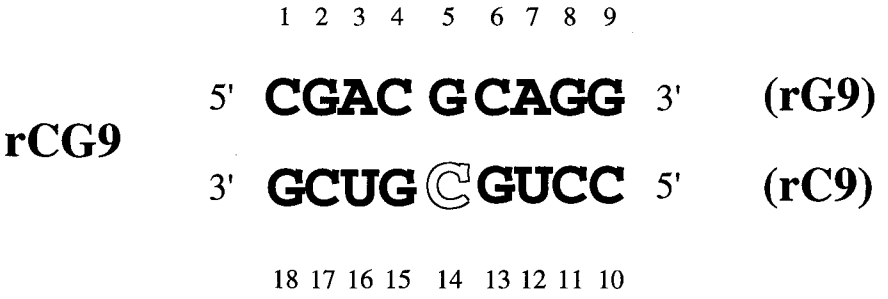
$$k_{ex} << 2\pi\Delta\upsilon \tag{3.3}$$

where $\Delta\upsilon$ is the chemical shift difference in Hz between duplexes and denatured single strands. In the case of 3.1, the temperature profile of chemical shifts gives a sigmoidal curve. In the case of 3.2, signals are broaden out during the transition. In the case of 3.3, two signals which come from the duplexes and denatured single strands are separately observed. In any case, approximate temperature of transition can be known.

To know the dissociation process of the C-U pair, the temperature profile of ^{15}N-chemical shift is a suitable method. By comparing its dissociation process with that of the other part of the helix, cooperativity of the dissociation of the C-U pair can be examined.

Materials and Methods

Synthesis of a control RNA oligomer



Scheme 3.1

Oligomer, rG9 in scheme 3.1 was prepared as described in Chapter 2 for a control duplex, rCG9.

Measurements of NMR spectra

The samples used here were rCU9, rCG9, rC9 and cytidine monomer. They were dissolved in the solution containing 20 mM sodium phosphate buffer at pH 7.0 and 50 mM NaCl in H_2O. Concentrations were 2 mM for the oligomers, 100 mM and 400 mM for the cytidine monomer.

Thermal denaturation experiments on non-exchangeable protons were performed from 0 to 70 °C by the NOESY spectra and from 50 °C to 70 °C by the COSY spectra by changing every 5 °C. Each spectrum was taken in 1024*256 column points for the spectral width of 5000*5000 Hz, with 4 scans per column point and mixing time 50 msec on a JEOL alpha600. ^{15}N one dimensional (1D) spectra were measured at various temperatures for 1024 points with the spectral width of 100.67 Hz, on a Bruker ARX500. HSQC spectra were measured at various temperatures in 512*20 column points for the spectral width of 4006.4*709.56 Hz on a Bruker DMX500. The pulse sequence of the improved HSQC spectrum were described by Mulder *et al.* (Mulder *et al.*, 1996). ^1H one dimensional spectra were measured at various temperatures for 1024 points with the spectral width of 5000 Hz, on a Bruker DMX500. Calibration of the proton chemical shifts was referred to the chemical shift of water hydrogen, $\delta(H_2O)$, as described in "Materials and Methods" of Chapter 2. The absolute frequency, ν_0^H, of a proton at 0 ppm was calculated by the equation below,

$$\nu_0^H = \nu_{center}^H - 500.13 * \delta(H_2O) \tag{3.4}$$

where ν_{center}^H is the frequency of observation center when the water-hydrogen signal is at the center of the spectrum. The chemical shift of nitrogen was calibrated by the equation below (Live *et al.*, 1984; Bax & Subramanian, 1986),

$$\nu_0^N = \nu_0^H * \left(\frac{\gamma_N}{\gamma_H}\right) \tag{3.5}$$

where ν_0^N is the absolute frequency of nitrogen at 0 ppm, γ_N and γ_H are the gyromagnetic ratios for nitrogen and proton, respectively. Then the absolute frequency of each signal was converted to a chemical shift by using ν_0^N.

Thermal denaturation experiments with CD spectra

Temperature profiles of rCU9 and rCG9 with CD spectra were measured in the same buffer and salt condition as NMR measurements. The concentration of the rCG9 was 125 μM and those of rCU9 were 250, 125, 67.5, 31.3, 15.6, 7.81, 3.91, 1.95 and 0.98 μM. The solutions containing from 250 to 15.6 μM oligomer were measured in a sample cell of 1 mm path length, and then the solutions containing from 15.6 to 0.98 μM oligomer were measured in a sample cell of 10 mm path length. CD spectra were measured at 22 temperatures between 6 and 69 °C for rCU9, and 20 temperatures between 6 and 84 °C for rCG9. Each spectrum was acquired from 320 to 200 nm in each 1 nm for the response of 1 second at the scan speed of 50 nm/min with 7 scans. The least square fittings to the temperature profiles of CD spectra were performed by means of the global analysis method (Kodama, 1997; Sugeta, 1981). In the global analysis method, following equations are used,

$$\Delta G = \Delta H(T) - T\Delta S(T) \tag{3.6}$$

$$\Delta H(T) = \Delta H(T_0) + \int_{T_0}^{T} \Delta C_P dT \tag{3.7}$$

$$\Delta S(T) = \Delta S(T_0) + \int_{T_0}^{T} \frac{\Delta C_P}{T} dT \tag{3.8}$$

where T and T_0 are temperature and that of the standard state, respectively, ΔG, ΔH, ΔS and ΔC_P are the molar Gibbs's free energy, molar enthalpy, molar entropy and molar heat capacity under the constant pressure, respectively. In the case of non-self-complementary double helix, the equilibrium is given by

$$[S_A] + [S_B] \rightleftharpoons [D] \tag{3.9}$$

$$K(T) = \frac{2\alpha}{(1-\alpha)^2 C} \tag{3.10}$$

where $[S_A]$, $[S_B]$ and [D] are concentrations of free strand-A, free strand-B and double helix, respectively, K(T) is a binding constant, C is a total single-strand concentration, and α is a molar fraction of the double helix. The relationship between the CD spectrum and the molar fraction is given by

$$\Delta\varepsilon(T) = \Delta\varepsilon_{S_A}(T) + (1-\alpha)\Delta\varepsilon_{S_B}(T) + \alpha\Delta\varepsilon_D(T) \tag{3.11}$$

$$\Delta\varepsilon_{S_A}(T) = \Delta\varepsilon_{S_A}(T_0) + a_A T \tag{3.12}$$

$$\Delta\varepsilon_{S_B}(T) = \Delta\varepsilon_{S_B}(T_0) + a_B T \tag{3.13}$$

$$\Delta\varepsilon_D(T) = \Delta\varepsilon_D(T_0) + a_D T \tag{3.14}$$

where $\Delta\varepsilon(T)$, $\Delta\varepsilon_{S_A}(T)$, $\Delta\varepsilon_{S_B}(T)$ and $\Delta\varepsilon_D(T)$ are observed molar ellipticity, molar ellipticity of strand-A, molar ellipticity of strand-B and molar ellipticity of double helix at a certain wavelength, respectively, $\Delta\varepsilon_{S_A}(T_0)$, $\Delta\varepsilon_{S_B}(T_0)$ and $\Delta\varepsilon_D(T_0)$ are the molar ellipticities of strand-A, strand-B and double helix at the temperature of T_0, a_A, a_B, and a_D are the slopes of the plots of the molar ellipticities of strand-A, strand-B and double helix against temperature. Each $\Delta\varepsilon(T)$ is an observed molar ellipticity at each temperature, and, $\Delta H(T_0)$, $\Delta S(T_0)$, $\Delta C_P(T_0)$, $\Delta\varepsilon_{S_A}(T_0)$, $\Delta\varepsilon_{S_B}(T_0)$, $\Delta\varepsilon_D(T_0)$ a_A, a_B, and a_D are variables for the least square fitting.

The least square calculations were performed to the 3D temperature profiles of all the concentrations, simultaneously. Because of the computational limit, data on the wavelengths

of every 5 nm were picked up.

Results

Thermal denaturation process of rCU9 monitored with NMR

Temperature profile of the ^{15}N chemical shift at amino nitrogen of C14 is presented in Figure 3.1(a) and those of the non-exchangeable protons, H5s and H6s, are presented in Figures 3.1(b) and (c). The predicted fraction of the double helix from the thermal denaturation experiment with CD spectra is presented in Figures 3.1(d). The detailed description of the CD experiment is described in the next section. Temperature profile of the ^{15}N chemical shift showed sigmoidal transition, but those of the non-exchangeable protons did not. However, both temperature profiles showed that the transition from double helix to denatured single strands seems to occur between 40 and 60 °C. In the case of the non-exchangeable protons, exchange rate (k_{ex}) between duplexes and single strands seemed to be relatively slow against the chemical shift difference ($\Delta\nu$) and existed between the states shown by equation 3.2 and 3.3. This is because the signals from all the residues in the double helix and the partial residues in the denatured single strands could be seen in the same COSY spectrum at 50 °C. Although the mid-point of the transition (*Tm*) determined by the non-exchangeable protons is obscure, the transition of each site seemed to occur simultaneously. Unfortunately, the signals from the denatured single strands could not be assigned. *Tm* values from the ^{15}N chemical shift and non-exchangeable protons seemed to be the same. The predicted fraction of the double helix from the thermal denaturation experiment with CD spectra described later also showed that the transition occurred approximately between 40 and 60 °C, which is consistent with the temperature profile of the ^{15}N chemical shift. All the data suggest that the C-U pair is dissociated cooperatively with the other base-pairs.

Surprisingly, the *Tm* values determined by the ^{15}N chemical shift is about 50 °C, which is as stable as the A-U-base-pair-rich double helix of this size. The *Tm* value is known to increase, as the ratio of the G-C pair increases. Therefore, this high *Tm* value is mainly due to the high ratio of the C-G pairs in rCU9. This high *Tm* value, however, can not be explained without strong stacking interaction between the C-U pair and the neighboring base-pairs.

Thermal denaturation experiment with CD spectra.

The denaturation process of rCU9 examined with NMR looked cooperative dissociation of all the base-pairs. To confirm this, the denaturation process of the duplex must be monitored through the wide range of the concentrations. For this purpose, CD spectrum is suitable because the wide range of the concentrations of the samples can be measured by using different path-length cells. The CD spectra of rCU9 at each temperature (3-dimensional (3D) temperature profile) were measured at various concentrations. Their concentrations were logarithmically reduced from 250 μM to 0.98 μM by the factor of 1/2. The 3D temperature profile at 31 μM, 1/8 concentration of the initial concentration, is presented in Figure 3.2(a).

In the figure, one of the horizontal axes is wavelength, the other horizontal axis is temperature, and the vertical axis is signal intensity of the CD spectrum. As the temperature was raised, the CD spectra were also changed due to the denaturation of the double helix to single strands. The observed spectrum at each temperature is the summation of those from the species in solution. In Figure 3.2(b), temperature profiles of the CD signal intensities at every 10 nm from 300 to 210 nm taken from the 3D temperature profile are also depicted. In Figure 3.3, the temperature profile (270 nm) at each concentration is presented. As the concentration decreased, the *Tm* value decreased.

The least square fittings were performed to the 3D profiles. At first, in the least square fitting, the two-state equilibrium between a duplex and denatured single strands with ΔC_P of zero was assumed and then the two-state equilibrium with a constant ΔC_P was examined. The results of the least square fitting to all the 3D profiles are summarized in Table3.1. Calculated parameters for rCG9 are also given as a reference of all-Watson-Crick base-paired double helix. *Tm* value of rCU9 is less than that of rCG9 by 19 °C. Although *Tm* value of the double helix decreased significantly by substitution of the C-G pair to the C-U pair, its value still keeps the level of the least stable double helices of this size. Observed data and calculated temperature profiles in the case of ΔC_P of zero are presented in Figure 3.4. The calculated thermodynamic parameters well accounted for the 3D temperature profiles at all concentrations (Figure 3.4). This implies that the two-state model was applicable to analyze the thermal denaturation process of rCU9, that is, the denaturation process of rCU9 is highly cooperative. It was demonstrated that the cooperative denaturation observed with NMR spectra was true. However, systematic errors between the data from different cells of different path-lengths could not be eliminated at present study. Therefore, it is hard to estimate whether ΔC_P was correctly determined or not by this method. Anyway, it should be noted that the thermal denaturation process of the double helix including a single C-U pair is highly cooperative. In the case 1 of Table 3.1 for both rCU9 and rCG9, these parameters were consistent with those from the calorimetry. Denaturation process of rCU9 described in the previous section was simulated by using these thermodynamic parameters of rCU9 in the case 1 of Table 3.1.

Behavior of the amino nitrogen chemical shifts

To study the existence of the hydrogen bond between the amino-nitrogen of cytidine and the keto-oxygen of uridine, temperature profiles of the chemical shifts of the amino-nitrogens in rCU9, rCG9, rC9 and cytidine monomer were compared (Figure 3.5). Because of the double strand formation of rC9 at low temperature, ^{15}N chemical shift of cytidine monomer in natural abundance was used as a control of a hydrogen-bond-free amino nitrogen (magenta triangles). However, at higher temperature than 44 °C, rC9 existed as a denatured single strand, and its chemical shift value is also useful as a reference of non-hydrogen-bonding. Observed chemical shift (rC9) was plotted as green squares with a linear line from the least square fitting to the data. A broken line means the putative chemical shift of the amino-nitrogen

in C14 as a single strand. The chemical shift of the amino-nitrogen in cytidine monomer was different from that in rC9. This is because the concentration of each sample is different: rC9, 2 mM; cytidine monomer, 400 mM. It is also said that pyrimidine bases such as uracil and cytosine tend to stack themselves at higher concentration (Nagata et al., 1965; Varghese & Wang, 1968). Therefore, high concentration of cytidine monomers could promote the association and then its chemical shift could be shifted to upfield by the ring current shift. It is important to note that its temperature profile can be approximated by a linear line. Thus the intrinsic chemical shift of the amino-nitrogen in rC9 at lower temperature became to be predictable. This kind of linear correlation between the temperature and the chemical shift of the nitrogen was observed in the case of the chemical shift of exo-cyclic amino group of [2-^{15}N]- O^6-methyl-2'-deoxy-guanosine (O^6MeG) monomer (Gaffney et al., 1993). Temperature profile of the amino-nitrogen of C14 in rCG9 (blue circles) showed a linear change at temperatures lower than 50 °C. This is a reference example whose amino-nitrogen is included in the hydrogen bonding in the Watson-Crick base-pair.

The temperature profile of the amino-nitrogen of C14 in rCU9 is presented in red diamonds with a simulated line by using thermodynamic parameters with the CD spectra in the case 1 of Table 3.1. It showed sigmoidal curve which is similar to that of the amino-nitrogen in a Watson-Crick-base-paired guanosine in a RNA octamer (Zhang et al., 1997). It is known that an upfield-shifted sigmoidal transition curve is concerned with the hydrogen bond breaking of the Watson-Crick base-pair (Zhang et al., 1997; Gao & Jones, 1987; Gaffney et al., 1993). The downfield shift of the amino nitrogen atom is expected when the proton attached to the nitrogen atom is involved in the hydrogen bond (Haushalter *et al.*, 1996). Thus, the upfield-shifted sigmoidal transition curve at higher temperatures means the breakage of the hydrogen bond.

On the other hand, the amino-nitrogen of the guanosine in the G-U pair showed the downfield-shifted sigmoidal transition as the G-U pair was broken (Zhang et al., 1997). Zhang *et al.* claimed that its downfield shift was due to the deshielding effect caused by altered stacking (Zhang et al., 1997). If the transition curve of the amino-nitrogen of C14 in rCU9 is due to the alteration of stacking, it means that the amino-nitrogen is sandwiched between the other bases at higher temperatures. This assumption is inconsistent with the character of denatured single strands.

There still remains the other explanation for the upfield-shifted sigmoidal transition. This is a planar ring current shift which means a ring current shift by the pairing base in the same base-pair plane. Planar ring current shift is known to cause a downfield shift the same as hydrogen bonding. However, no sigmoidal transition was observed for the nitrogen atom which was not involved in the hydrogen bond but was located near the pairing base in the case of N1 of O^6MeG (Gaffney et al., 1993). This implies that the planar ring current shift can be neglected. Accordingly, the upfield-shifted sigmoidal transition of the amino nitrogen in the C-U pair must be due to the breakage of the hydrogen bonding.

When the chemical shift values of rCU9, rCG9 and rC9 blow 40 °C are compared, their relations are always

$$\delta_{^{15}N}(rCG9) > \delta_{^{15}N}(rCU9) > \delta_{^{15}N}(rC9) \quad (3.15)$$

where $\delta_{^{15}N}(rCG9)$ is the chemical shift value of the amino-nitrogen of C14 in rCG9, $\delta_{^{15}N}(rCU9)$ is the chemical shift value of the amino-nitrogen of C14 in rCU9 and $\delta_{^{15}N}(rC9)$ is the chemical shift value of the amino-nitrogen of C14 in rC9. This relation seems to reflect the strength of the hydrogen bond. Because the electron density around the amino-nitrogen becomes lower when the attached proton forms hydrogen bond, shielding effect decreases, and then its chemical shift value becomes larger, that is, the downfield shift is observed when compared with that of the non-hydrogen-bonded amino-nitrogen.

Chemical shifts of the amino-protons

HSQC spectra of the amino groups of the C-U pair in rCU9, the C-G pair in rCG9 and cytidine monomer at 10 °C are presented in Figure 3.6. In the case of rCG9, the hydrogen-bonded amino-proton of C14 showed significantly large downfield shift when compared with that of cytidine monomer. The chemical shift value of the other amino-proton of C14 has similar value to that of the corresponding amino-proton in cytidine monomer. On the contrary, both amino-protons of C14 in rCU9 showed significantly large upfield shift from those of cytidine monomer. In general, a hydrogen-bonded proton shows large downfield shift. It looks inconsistent with the data of the ^{15}N chemical shift. Here we have to remind the crystal structure of the C-U pair in Chapter 1 (Figure 1.5). There are two structural features which affect the chemical shifts of the amino-protons. 1) The angles on the hydrogen bond between the amino-nitrogen of cytidine and keto-oxygen of the uridine were deviated from the proper geometry. 2) The C-U pair was significantly slided toward the minor groove side. If the solution structure is the same as that in crystal, the following phenomena are expected. The first feature makes hydrogen bond weaker. It does not cause the large downfield shift in contrast with the large downfield shift by a strong hydrogen bond. The second feature locates the amino group of the C-U pair between the rings of the neighboring bases. Therefore, the chemical shifts of the amino protons are expected to be shifted to upfield significantly. Accordingly, the crystal structure of the C-U pair well account for chemical shift values of the amino protons. This also suggests that the structure of the C-U pair is similar to that in crystal.

Temperature profiles of the amino-protons of rCU9, rCG9 and cytidine monomer are described (Figure 3.7). In the case of rCG9, the chemical shift of two amino proton signals were almost flat until 42 °C and then both of them shifted to that of water hydrogen at higher temperature than 42 °C. They were observed independently until their signals disappeared. In the case of the cytidine monomer, two amino proton signals were converged at 39 °C. Chemical shift difference between two signals became smaller gradually as the temperature was raised. This implies that the exchange rate between two amino-protons increases gradually.

In the case of rCU9, the chemical shift of the two amino-proton signals almost flat below 21 °C. Suddenly, the averaging of the amino-proton signals started at 21 °C and they converged at 36 °C. These imply that the exchange rate of the two amino-protons in the C-U pair does not change until 21 °C but it is accelerated after 21 °C. This type of acute transition was observed in the case of the mixture of urea and acetate in dimethylformamide. In the mixture, hydrogen bonds exit between urea and acetate (Haushalter *et al.*, 1996). Therefore, such an acute transition of the chemical shifts on the amino protons in urea must be due to hydrogen bonding. In the absence of acetate, the temperature profiles of their chemical shifts were gradually averaged (Haushalter *et al.*, 1996). This is the same as the case of a cytidine monomer. The temperature profile of the amino proton chemical shifts also indicate the existence of a hydrogen bond.

Arrangement of the C-U pair in solution.

As the existence of the hydrogen bond was proven, it is worth reinvestigating the arrangement of the C-U pair. The data on the C-U pair in solution are summarized in three points.

1) The C-U pair is stacked in the double helix and their χ angles are in anti-conformation.

2) The imino proton signal of U5 in the C-U pair can not be observed at all.

3) The amino nitrogen in the C-U pair is incorporated in the hydrogen bonding.

From the absence of the imino proton signal of U5, it is thought to be exposed to bulk water. Possible candidate for the counterpart of the hydrogen bonding of the amino group in C14 is either O4 of U5 and O2 of the U5. Accordingly, there are two possible arrangement of the C-U pair for satisfying these data (Figure 3.8(a),(b)). At first, structures of the two arrangements were calculated with additional constraints to fix the arrangement of the C-U pair. Both arrangements satisfy NOE constraints because of large tolerances. Then correlations between the distances and NOE peak heights for these two arrangements were examined as described in Figure 2.4 of Chapter 2. The correlations for the arrangements shown in Figure 3.8 (a) and (b) are given in Figure 3.9 (a) and (b), respectively. In those figures, only the NOE constraints concerning with U5 and C14 were picked up. There should be a linear correlation between sixth root of the NOE peak height ((the NOE peak height)$^{1/6}$) and the distance from equation G.5. The best-fit lines for the data and correlation factors for (a) and (b) are shown in Figure 3.9. The correlation factor of (a) is a little smaller than that of (b). In Figure 3.9 (b), the points enclosed by ovals showed large deviation. They correspond to the cross peaks of G13H2"-C14H6, G15H8-C14H6, U5H2"-U5H6 and U5H2"-C6H5. The fact implies that for the arrangement of the C-U pair, (a) is more plausible. Accordingly, the averaged structure of the C-U pair is the same as that in crystal.

Discussion

Here we reinvestigate the chemical shift values of the amino protons and nitrogen. For the origin of the upfield-shifted sigmoidal transition of the nitrogen atom, ring current shift from the neighboring base-pairs was denied. However, the absolute value of its chemical shift could be affected by a ring current shift when the C-U pair is stacked in the double helix. The chemical shift difference between the hydrogen-bonded and non-hydrogen-bonded nitrogens of the C-G pair is at least 3 ppm, and that of proton is approximately 1 ppm. On the contrary to this, the maximum ring current shift observed for the non-exchangeable base proton was about 1 ppm. Basically, the ring current shift equally affects the chemical shifts of both protons and nitrogen atoms in ppm order. Therefore, the maximum ring current shift of nitrogens is also expected to be 1 ppm. Accordingly, their orders of the effects on their chemical shifts are as follows.

nitrogen: hydrogen bonding > ring current shift (3.16)

proton: hydrogen bonding ≅ ring current shift (3.17)

In the case of the chemical shift values of nitrogen atoms, the hydrogen bonding effect is superior to the ring current shift in general. However, in the case of those of protons, the ring current shift could come over the effect of the hydrogen bonding. In the case of the C-U pair, a ring current shift may occur to it. If it is right, the intrinsic chemical shift value is closer to that of the C-G pair. In conclusion, the chemical shift value of nitrogen atoms is a better probe of the hydrogen bonding than that of protons

Thermal denaturation experiments showed highly cooperative dissociation of the C-U pair to the other part of the double helix. This property is the same as those of Watson-Crick base-pairs. These results suggested that the two-state model is applicable to the thermal denaturation of the oligoribonucleotide containing a single C-U pair. Accordingly, the concept of the nearest neighbor parameters as thermodynamic parameters also seems to be applicable to the C-U pair. A single C-U site seems to be incorporated in the double helix. From the results, the single C-U mismatch site in *Escherichia coli* 23S ribosomal RNA can be stacked in the double helix. For the secondary structure prediction, a single C-U mismatch site is acceptable for double helices. Once the nearest neighbor parameters for the C-U pair is derived, the secondary structure prediction will be more accurate. Instead of these facts, the occurrence of the C-U pair in biological system is low. Substitution of Watson-Crick base-pairs to the C-U pair makes double helices weaker. Thermally unstable double helices are easy to be cleaved by nucleases. Therefore, the C-U pair was eliminated from double helices through the evolution of RNA molecules in biological system

It should be noted that the best-fit model of the C-U pair to the NOE data is the same as that in crystal. This indicates that the arrangement seen in the crystal structure is not an artifact which arose from the crystal packing and the flexibility of the internal loop of four successive non-Watson-Crick base-pairs. Remarkably, this arrangement is conformationally acceptable for a single C-U mismatch site in duplexes as well.

To determine the precise structure of the C-U pair, the back calculation method could be applicable because of the high correlation between distances and NOE peak heights. By the method more plausible structure of the C-U pair will be determined.

Table 3.1. Thermodynamic parameters obtained by the least square fitting to the temperature profiles of CD spectra for rCU9 and rCG9.

	rCU9		rCG9	
	case1	case2	case1	case2
ΔH (kJ mol^{-1})	−362 ± 2.1	−302 ± 6.8	−348 ± 2.7	−7 ± 5.0
ΔS (J K^{-1}mol^{-1})	−1059 ± 6.7	−860 ± 22	−946 ± 8.2	−120 ± 16
ΔC_P (J K^{-1}mol^{-1})	0.0	−4500 ± 520	0.0	−8900 ± 180
rmsd	1.34	1.32	0.11	0.09
Tm at 125 μM (°C)	45		64	

In case1, ΔC_P was assumed to be zero. In case2, ΔC_P was assumed to be constant and thermodynamic parameters are values at 300K. Root mean square deviations (rmsd) are given under the thermodynamic parameters. For rCG9, parameters were derived from the temperature profile only at 125 μM. Tm values at 125 μM are given in the bottm line.

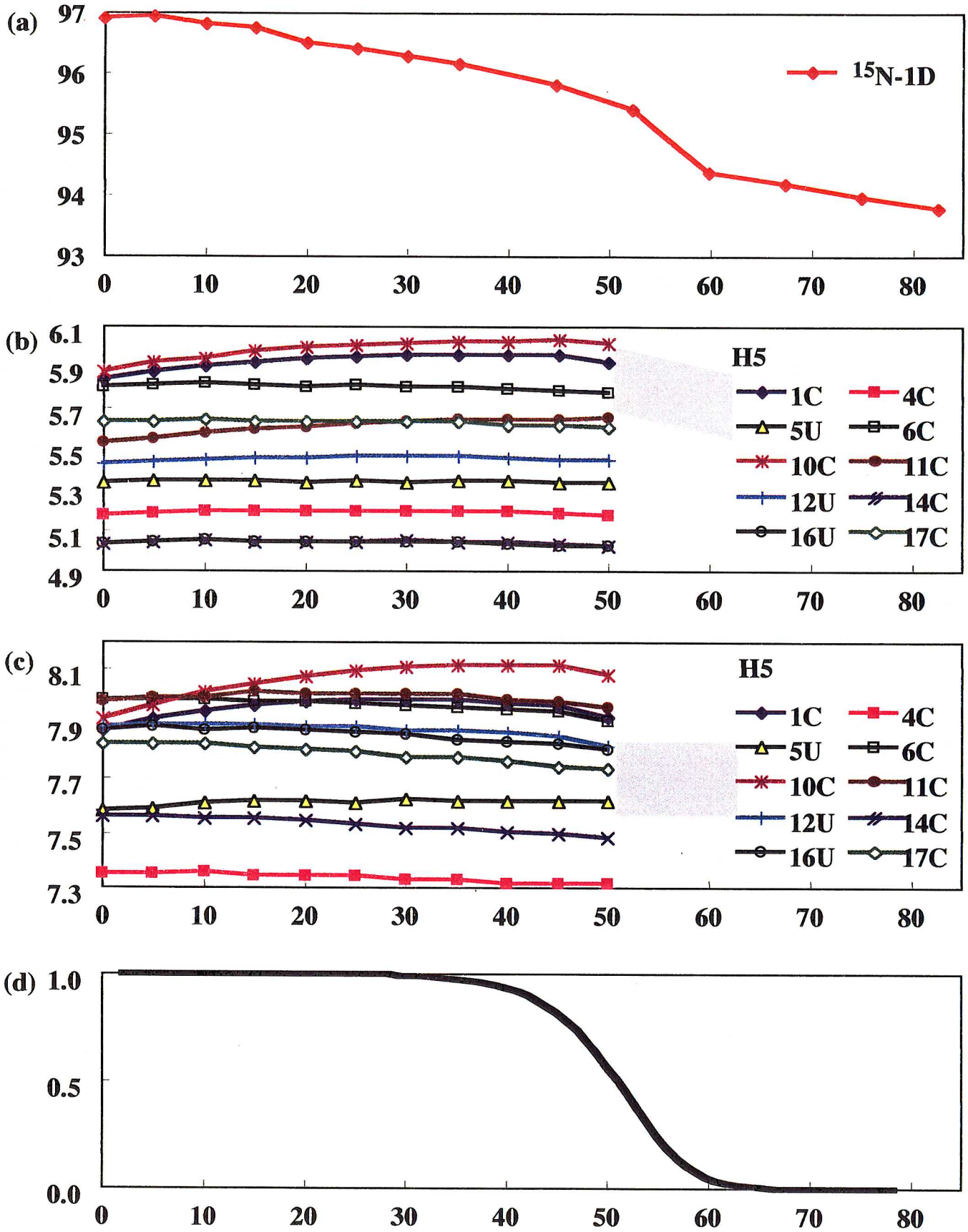


Figure 3.1 (a) Temperature profile of the amino nitrogen chemical shift of C14. (b) Temperature profiles of protons (H5) chemical shifts of rCU9. The area where signals from denatured strands were observed is shaded. (c) Temperature profiles of protons (H6) chemical shifts of rCU9. The area where signals from denatured strands were observed is shaded. (d) Fractions of the double helix predicted from temperature profiles of CD spectra are plotted to temperature.

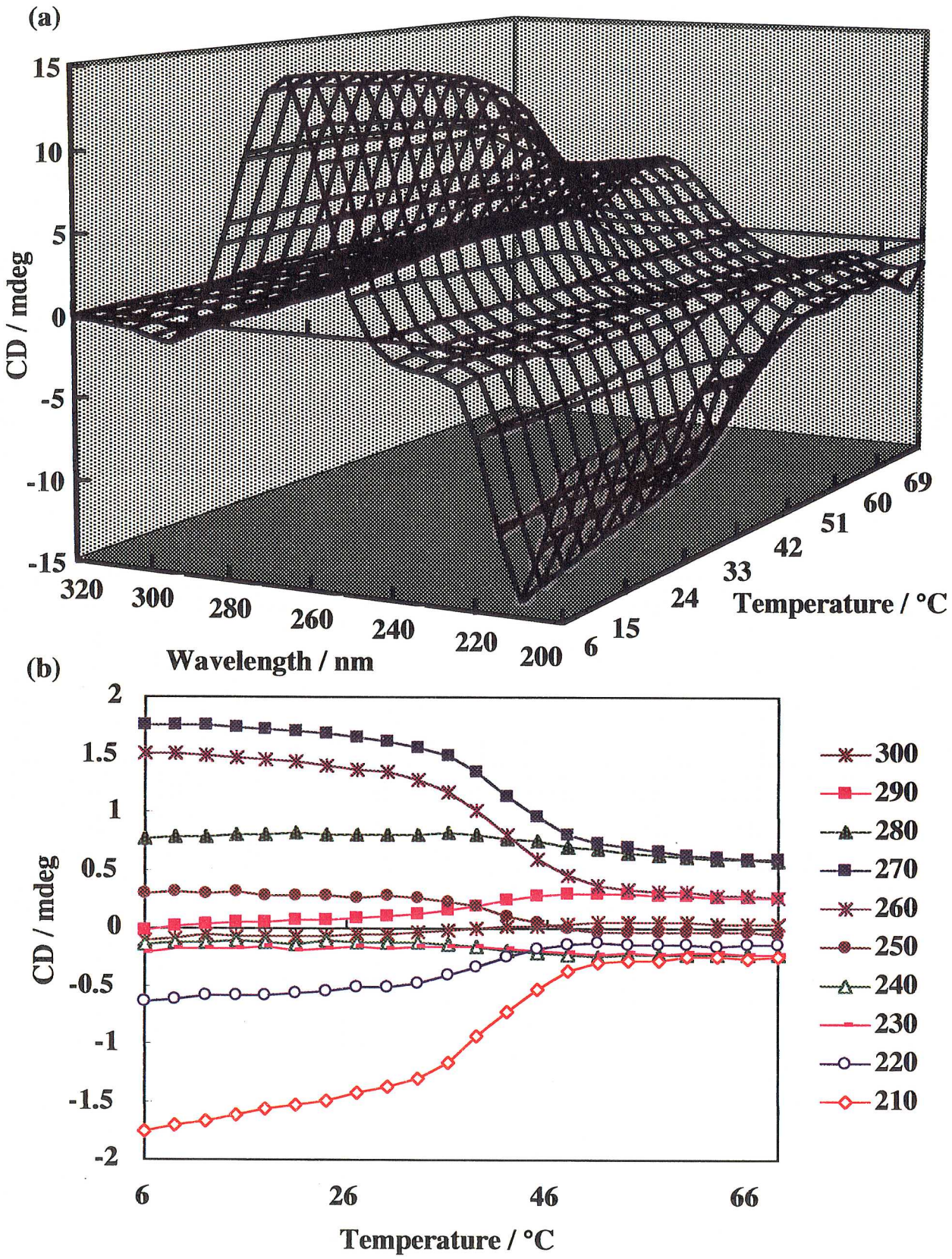


Figure 3.2 (a) 3-dimensional (3D) temperature profile of CD spectra. One of horizontal axes is wavelength and the other horizontal axis is temperature. Vertical axis is a signal intensities of CD spectra. (b) Temperature profiles of CD signal intensities are presented from 210 to 300 nm by every 10 nm.

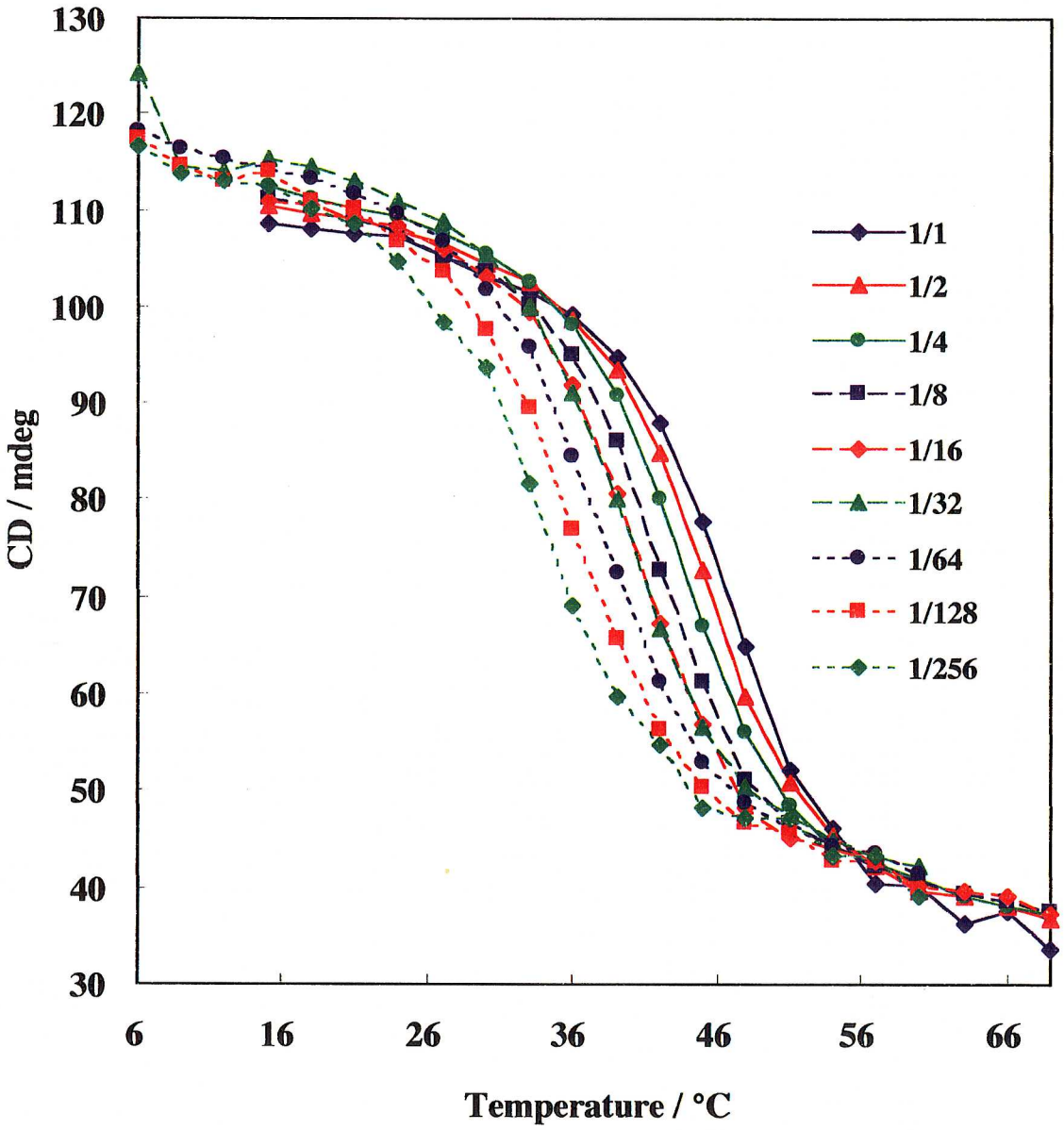


Figure 3.3 Temperature profiles of CD spectra for different concentrations at 270 nm.

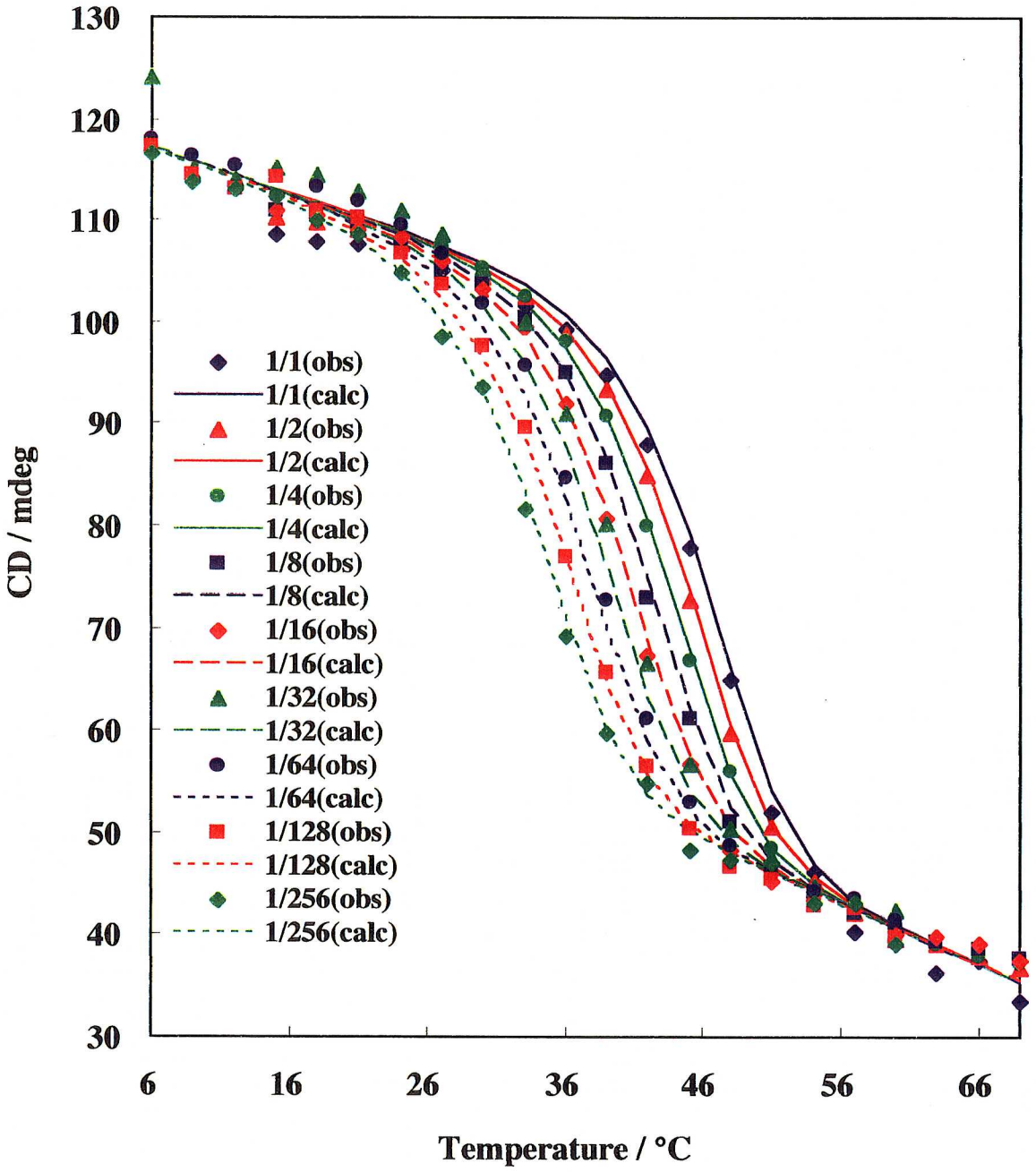


Figure 3.4 Least square fitting for the temperature profiles of the CD spectra. Observed data given by several kinds of shaped points were taken at 270 nm fro 9 different concentrations. The relative concentrations to the initial one are indicated in the figurre. "obs" means observed datum and "calc" means calculated profile.

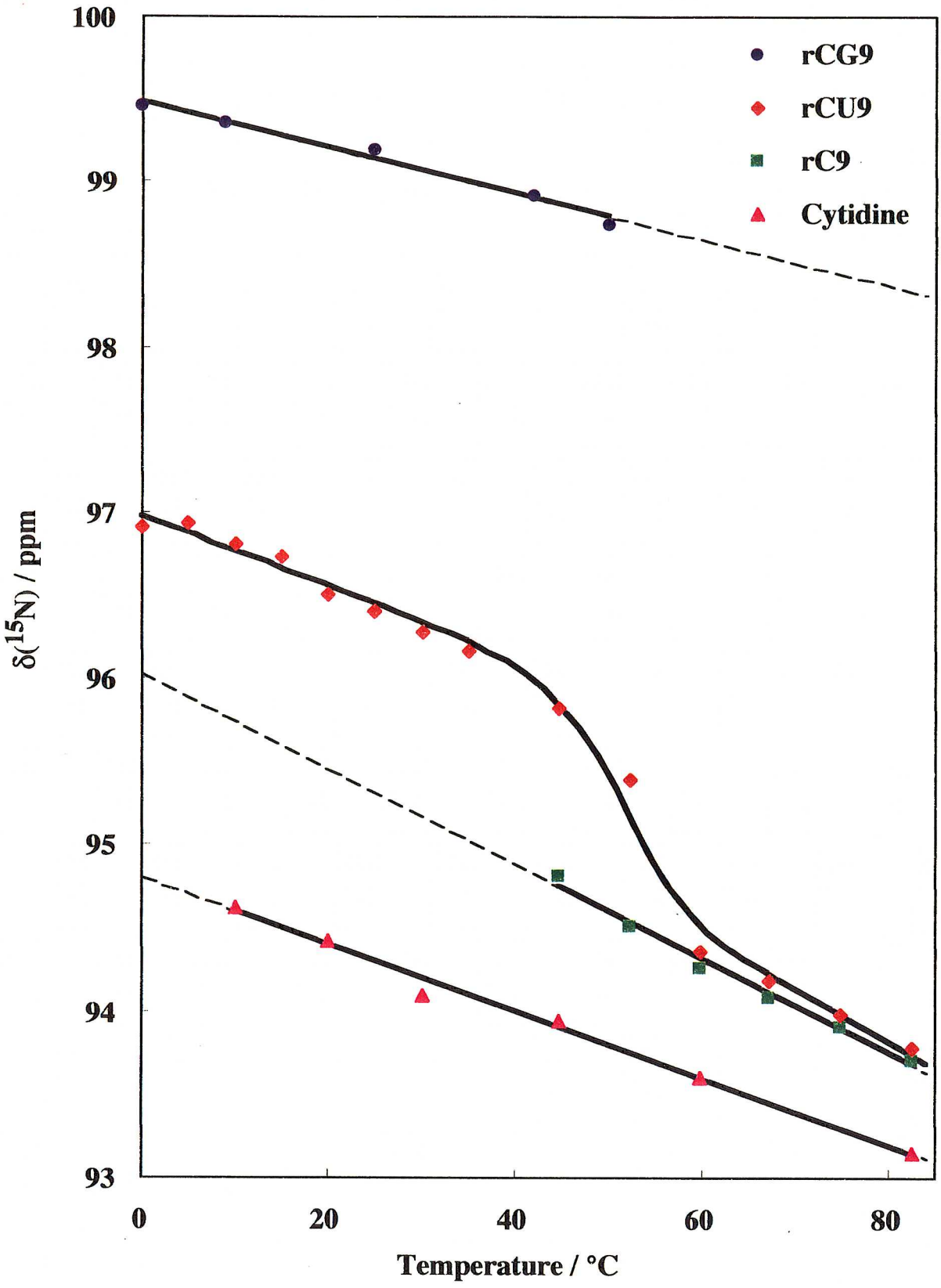


Figure 3.5 Temperature profiles of the amino nitrogen chemical shifts. Those of rCG9, rCU9, rC9 and cytidine monomer are presented as blue circle, red diamond, green square and magenta triangle, respectively. The least square fitting was performed and the results are presented in a thick lines. The broken lines represent predicted temperature profiles which are extrapolated from the least square fitting lines.

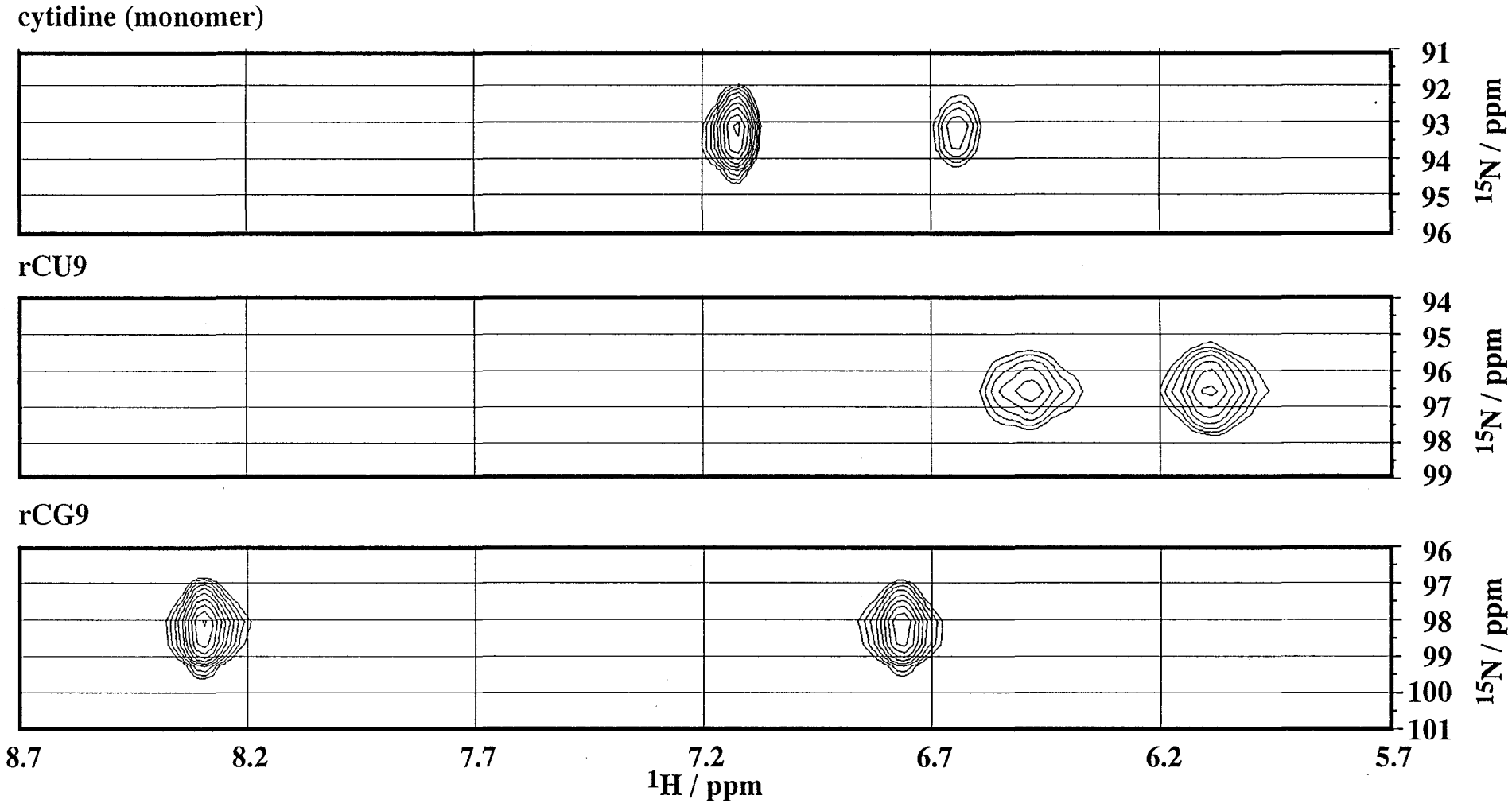


Figure 3.6 ^{1}H-^{15}N-HSQC spectra of the amino groups of cytidine monomer, rCU9 and rCG9.

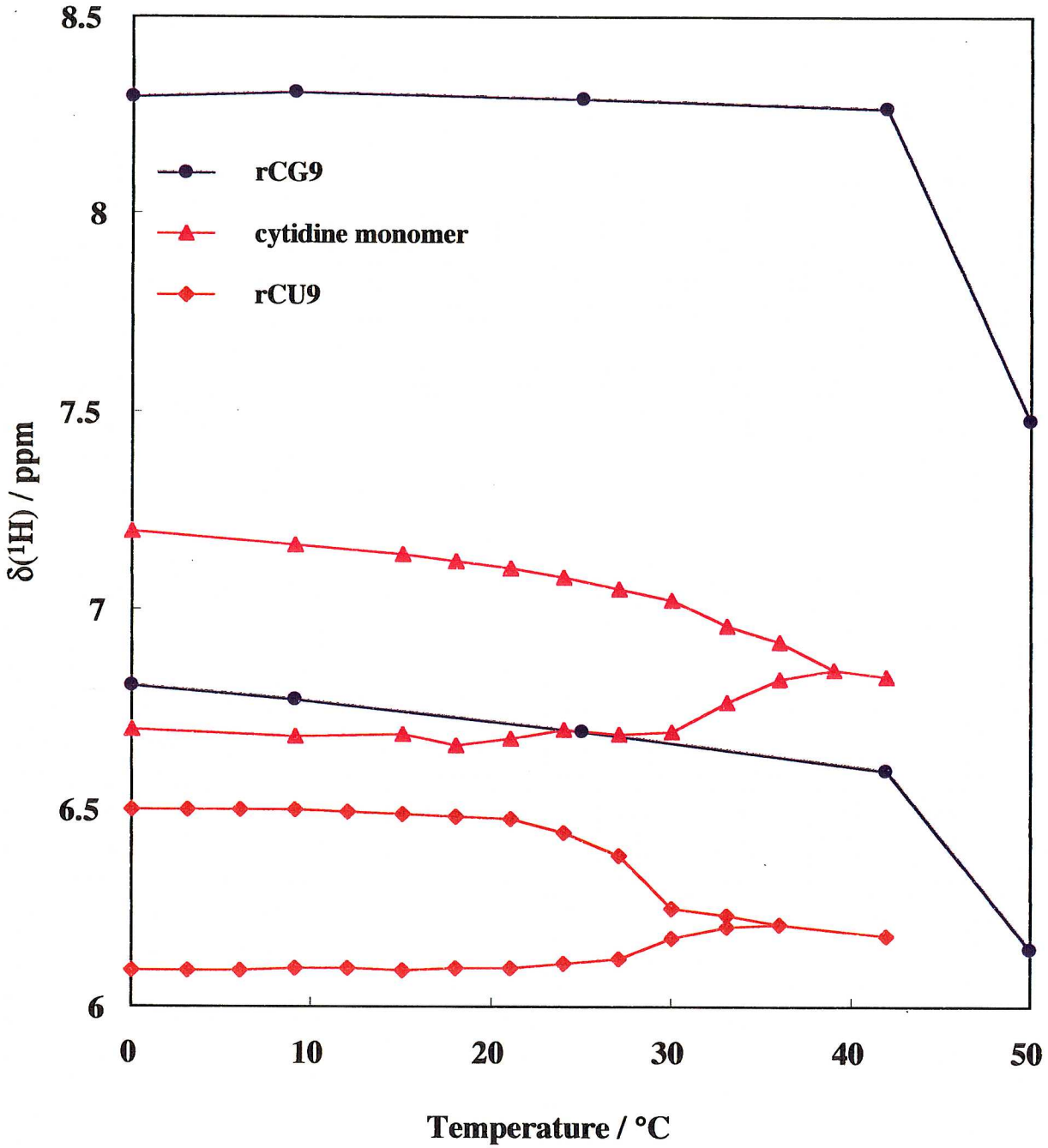
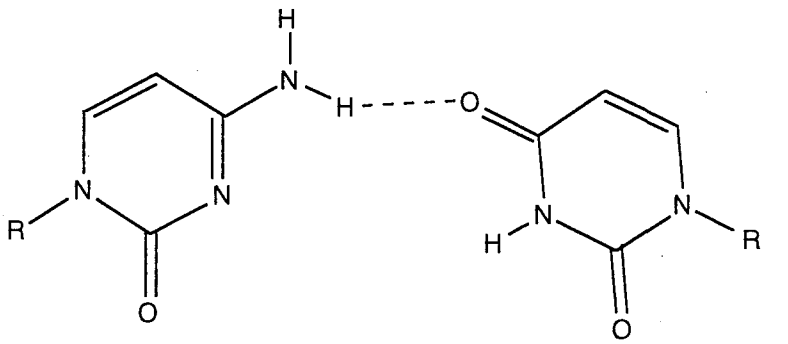


Figure 3.7 Temerature profile of the amino protons of cytidine monomer (magenta triangles), rCU9 (red diamonds) and rCG9 (blue circles). Chemical shifts were derived from ^{1}H-1D spectra for cytidine monomer and from HSQC spectra for rCU9 and rCG9.

(a) **arrangement-(a)**



(b) **arrangement-(b)**

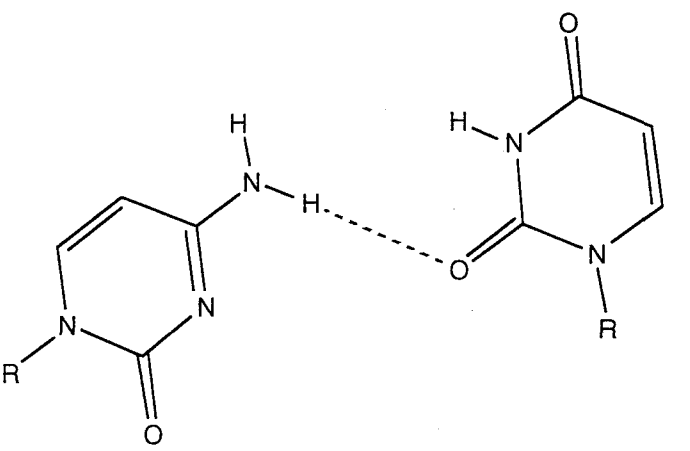


Figure 3.8 Two possible arrangements of the C-U pair are presented. (a) The arrangement which has hydrogen bond between the amino group of cytidine and the keto-oxygen (O4) of uridine. This is the same configuration as in crystal. (b) The arrangement which has hydrogen bond between the amino group of cytidine and the keto-oxygen (O2) of uridine.

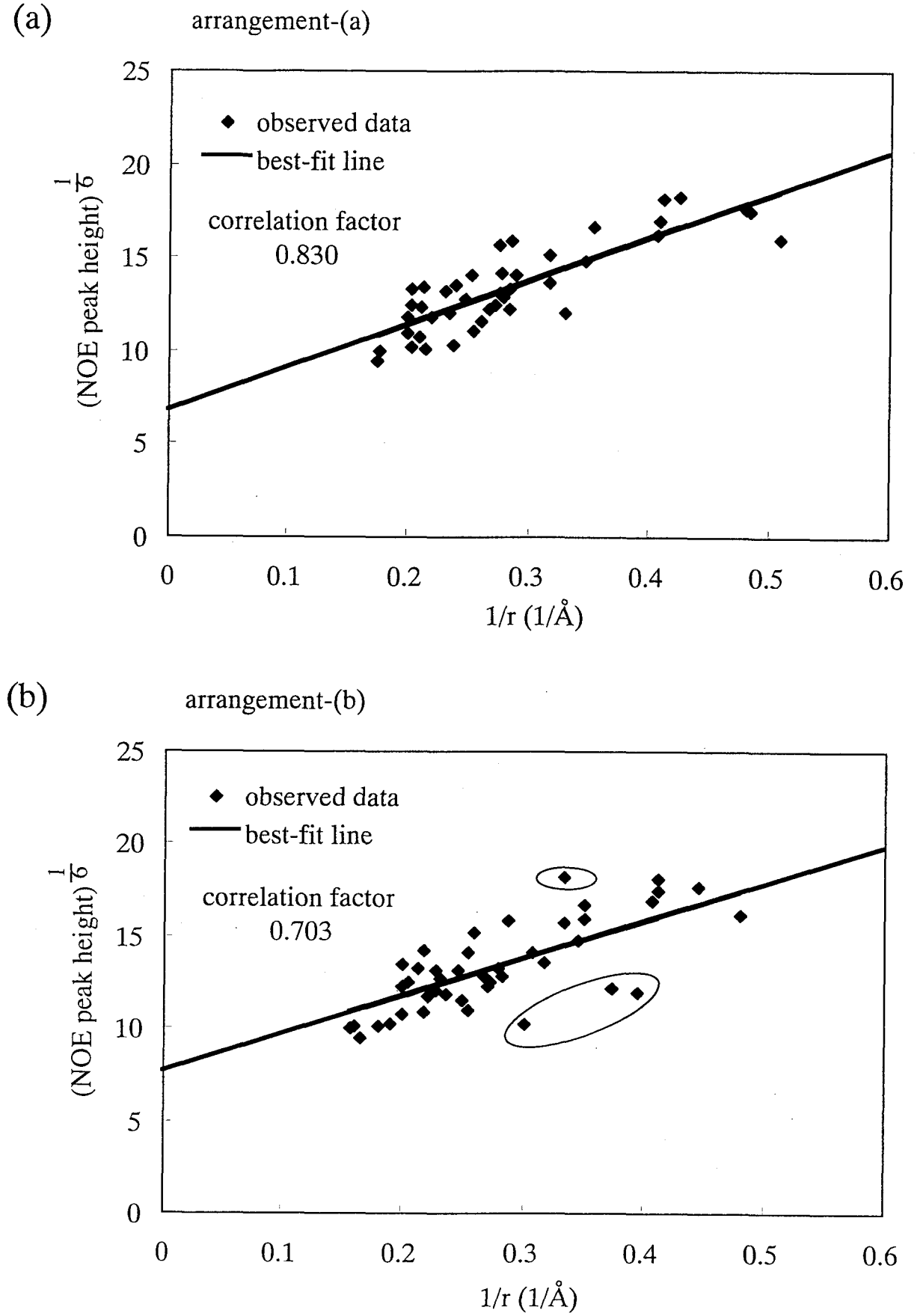


Figure 3.9 Scatter diagrams between the sixth roots of NOE peak height and distances for arrangements-(a) and -(b). The least square fitting lines are presinted in the figures. Correlation factors are also presented.

Conclusion

In Chapter 1, the crystal structure of oligoribonucleotide, r(UGAGCUUCGGCUC), is presented. First of all, it was demonstrated that the pyrimidine-pyrimidine base-pair such as the C-U pair exists. It was not expected before its crystal structure was solved. This is because introduction of the pyrimidine-pyrimidine mismatch makes a duplex more unstable in general. Therefore, it had seemed that the pyrimidine-pyrimidine mismatch does not form a base-pair in a specific manner. However, all the crystal structures of the C-U pair ever solved were adopted in the same arrangement. It was also found that hydrogen bond network through the water molecules is important for its stabilization. Secondly, the global conformation of the tridecamer was found to adopt in the A'-RNA conformer. This is the first observation of A'-RNA conformation in a single crystal. The method to classify the conformation of RNA double helix was also established.

In Chapter 2, NOE based solution structure calculation with NMR was performed to clarify the structure of the C-U pair in solution. It revealed that nonamer duplex RNA including the C-U pair, rCU9, forms a right-handed double helix, and the C-U pair was stacked in the double helix with the χ angles in the anti-conformation. Since the imino proton signal was completely absent, it is suggested that the imino proton is exposed to bulk water. This result is consistent if we assume that the arrangement of the C-U pair in duplex in solution is the same as that in crystal.

In Chapter 3, physicochemical properties of the C-U pair were studied with CD spectrum and NMR. Firstly, the dissociation process of the C-U pair was highly cooperative to the other part of the duplex such as that of the Watson-Crick base-pairs. This implies that the nearest neighbor parameters on the thermodynamics can be determined for the C-U pair. From the view-point of the secondary and tertiary structure prediction, it has become clear that a single C-U mismatch site does not interrupt the continuity of the double helix. Secondly, the existence of the hydrogen bond between the C-U pair was demonstrated with ^{15}N-NMR. Strictly saying, the amino group of the C-U pair is involved in the hydrogen bond. Thirdly, from the correlation between NOE peak heights and the distances, the best-fit model for the NOE distance constraints was the same structure as in crystal. This indicates that crystal structure of the C-U pair reflected the nature of the C-U pair in solution.

Finally, all the methods employed here are applicable to the C-U pair as well. The findings extremely widen the use of above techniques. The structures and properties of other mismatches will be solved in the future.

References

Aboul-ela, F., Karn, J. & Varani, G. (1995). The structure of the human immunodeficiency virus type-I TAR RNA reveals principles of RNA recognition by Tat protein. *J. Mol. Biol.* **253**, 313-332.

Aboul-ela, F., Karn, J. & Varani, G. (1996). Structure of HIV-1 RNA in the absence of ligands reveals conformation of the tridecanucleotide bulge. *Nucl. Acids Res.* **24**, 3974-3981.

Allain, F. H.-T. & Varani, G. (1995a) Divalent metal ion binding to a conserved wobble pair defining the upstream site of cleavage of group I self-splicing intron. *Nucl. Acids Res.* **23,** 341-350.

Allain, F. H.-T. & Varani, G. (1995b). Structure of the P1 Helix from Group I self-splicing introns *J. Mol. Biol.* **250**, 333-353.

Allain, F. H.-T. & Varani, G. (1997). How accurately and precisely can RNA structure be determined by NMR? *J. Mol. Biol.* **267**, 338-351.

Antao, V. P. & Tinoco, I., Jr. (1992). Thermodynamic parameters for loop formation in RNA and DNA hairpin tetraloops. *Nucl. Acids Res.* **20**, 819-824.

Arnott, S., Hukins, D. W. L. & Dover, S. D. (1972). Optimized parameters for RNA double-helices. *Biochem. Biophys. Res. Commun.* **48**, 1392-1399.

Baeyens, K. J., Jancarik, J. & Holbrook, S. R. (1994). Use of low-molecular-weight polyethylene glycol in the crystallization of RNA oligomers. *Acta Crystallog. sect. D,* **50**, 764-767.

Baeyens, K. J., De Bondt, H. L. & Holbrook, S. R. (1995). Structure of an RNA double helix including uracil-uracil base pairs in an internal loop. *Nature Struct. Biol.* **2,** 56-62.

Baeyens, K. J., De Bondt, H. L., Pardi, A. & Holbrook, S. R. (1996). A curved RNA helix incorporating an internal loop with G•A and A•A non Watson-Crick base pairing. *Proc. Natl. Acad. Sci. USA*, **93**, 12851-12855.

Battiste, J. L., Mao, H., Rao, N. S., Tan, R., Muhandiram D. R., Kay, L. E., Frankel, A. D. & Williamson, J. R. (1996). Alpha helix-RNA major groove recognition in an HIV-1 rev peptide-RRE RNA complex. *Science*, **273**, 1547-1551.

Bax, A. & Subramanian, S. (1986). Sensitivity enhanced two-dimensional heteronuclear shift correlation NMR spectroscopy. *J. Magn. Reson.* **67**, 565-569.

Betzel, Ch., Lorenz, S., Fürste, J. P., Bald, R., Zhang, M., Schneider, Th. R., Wilson, K. S. & Erdmann, V. A. (1994). Crystal structure of domain A of Thermus flavus 5S rRNA and the contribution of water molecules to its structure. *FEBS Letter*, **351**, 159-164.

Biswas, R., Wahl, M. C., Ban, C. & Sundaralingam, M. (1997). Crystal structure of an alternating octamer r(GUAUGUA)dC with adjacent G•U wobble pairs. *J. Mol. Biol.* **267**, 1149-1156.

Biswas, R. & Sundaralingam, M. (1997). Crystal structure of r(GUAUGUA)dC with tandem G•U/U•G wobble pairs with strand slippage. *J. Mol. Biol.* **270**, 511-519.

Boulard, Y., Cognet, J. A. H. & Fazakerley, G. V. (1997). Solution structure as a function of pH of two central mismatches, C•T and C•C, in the 29 to 39 K-*ras* gene sequence, by nuclear magnetic resonance and molecular dynamics. *J. Mol. Biol.* **268**, 331-347.

Brünger, A. T. (1992). X-PLOR version 3.1 manual, Yale University Press, New Haven, CT.

Büchner, P., Maurer, W. & Rüterjans, H. (1978). Nitrogen-15 nuclear magnetic resonance spectroscopy of ^{15}N labeled nucleotides. *J. Mag. Res.* **29**, 45-63.

Cai, Z. & Tinoco I., Jr. (1996). Solution structure of loop A from hairpin ribozyme from tobacco ringspot virus satellite. *Biochemistry*, **35**, 6026-6036.

Cate, J. H., Gooding, A. R., Podell, E., Zhou, K., Golden, B. L., Kundrot, C. E., Cech, T. R. & Doudna, J. A. (1996a). Crystal structure of a group I ribozyme domain: principle of RNA packing. *Science*, **273**, 1678-1685.

Cate, J. H., Gooding, A. R., Podell, E., Zhou, K., Golden, B. L., Szewczak, A. A., Kundrot,

C. E., Cech, T. R. & Doudna, J. A. (1996b). RNA tertiary structure mediation by adenosine platforms. *Science*, **273**, 1696-1699.

Chattopadhyaya, R., Ikuta, S., Grzeskowiak, K. & Dickerson, R. E. (1988). X-ray structure of a DNA hairpin molecule. *Nature,* **334**, 175-179.

Cheong, C., Varani, G. & Tinoco, I., Jr. (1990). Solution structure of an unusually stable RNA hairpin. *Nature,* **346**, 680-682.

Cruse, W. B. T., Saludjian, P., Biala, E., Strazewski, P., Prangé T. & Kennard, O. (1994). Structure of a mispaired RNA double helix at 1.6-Å resolution and implication of RNA secondary structure. *Proc. Natl. Acad. Sci. USA*, **91**, 4160-4164.

Delaglio, F., Grezesiek, S., Vuister, G. W., Zhu, G., Pfeifer, J. & Bax, A. (1995). NMRPipe: A multidimensional spectral processing system based on UNIX pipes. *J. Biomol. NMR*, **6**, 277-293.

Diekmann, S., Dickerson, R. E., Bansal, M., Calladine, C. R., Hunter W. N., Kennard, O., Lavery, R., Nelson, H. C. M., Olson, W. K., Saenger. W., Shakked, Z., Sklenar, H., Sounpasis, D. M., Tung, C.-S., von Kitzing, E., Wang, A. H.-J. & Zhurkin, V. B. (1989). Definitions and nomenclature of nucleic acid structure parameters. *EMBO J.* **8**, 1-4, *J. Biomol. Struct. Dynam.* **6**, 627-634, *Nucl. Acids Res.* **17**, 1797-1803, *J. Mol. Biol.* **205**, 787-791.

Dock-Bregeon, A. C., Chevrier, B., Podjarny, A., Moras, D., de Baer, J. S., Gough, G. R., Gilham, P. T. & Johnson, J. E. (1988). High resolution structure of the RNA duplex $[U(U\text{-}A)_6A]_2$. *Nature*, **335**, 375-378.

Dock-Bregeon, A. C., Chevrier, B., Podjarny, A., Johnson, J., de Baer, J. S., Gough, G. R., Gilham, P. T. & Moras, D. (1989). High resolution structure of the RNA duplex $[U(U\text{-}A)_6A]_2$. *J. Mol. Biol.* **209**, 459-474.

Egli, M., Portmann, S. & Usman, N. (1996). RNA hydration: A detailed look. *Biochemistry*, **35**, 8489-8494.

Fujii, S., Tanaka, Y., Tomita, K.-I., Sakata, T., Hiroaki, H., Tanaka, T., & Uesugi, S. (1991). Molecular structure of extraordinarily stable RNA : model building and X-ray analysis. *Nucl. Acids Symp. Ser.* **25**, 181-182.

Fujii, S., Tanaka, Y., Uesugi, S., Tanaka, T., Sakata, T. & Hiroaki, H. (1992). Conformational features of the four successive non-Watson-Crick base pairs in RNA duplex. *Nucl. Acids Symp. Ser.* **27**, 63-64.

Fujiwara, T. & Shindo, H. (1985). Phosphorus-31 nuclear magnetic resonance of highly oriented DNA fibers. 2. Molecular motions in hydrated DNA. *Biochemistry*,**24**, 896-902.

Fuller, W., Wilkins, M. H. F., Wilson, H. R., Hamilton, L. D. & Arnott, S. (1965). The molecular configuration of deoxyribonucleic acid IV. X-ray diffraction study of A form. *J. Mol. Biol.* **12**, 60-80.

Gaffney, B. L., Goswami, B. & Jones, R. A. (1993). Nitrogen-15-labeled oligodeoxynucleotides. 7. Use of ^{15}N NMR to probe H-bonding in an O^6MeG•C base pair. *J. Am. Chem. Soc.* **115**, 12607-12608.

Gaffney, B. L., Kung, P.-P., Wang, C. & Jones, R. A. (1995). Nitrogen-15-labeled oligodeoxynucleotides. 8. Use of ^{15}N NMR to probe Hoogsteen hydrogen bonding at guanine and adenine N7 atoms of a DNA triplex. *J. Am. Chem. Soc.* **117**, 12281-12283.

Gao, X. & Jones, R. A. (1987). Nitrogen-15-labeled oligodeoxynucleotides. Characterization by ^{15}N NMR of d[CGTACG] containing $^{15}N^6$- or $^{15}N^1$-labeled deoxyadenosine. *J. Am. Chem. Soc.* **109**, 3169-3171.

Garrett, D. S., Powers, R., Gronenborn, A. M. & Clore, G. M. (1991). A common sense approach to peak picking in two-, three-, and four-dimensional spectra using automatic computer analysis of contour diagrams. *J. Magn. Reson.* **95**, 214-220.

Goswami, B., Gaffney, B. L. & Jones, R. A. (1993). Nitrogen-15-labeled oligodeoxynucleotides.

5. Use of ^{15}N NMR to probe H-bonding in an O^6MeG•T base pair. *J. Am. Chem. Soc.* **115**, 3832-3833.

Gutell, R. R., Larsen, N. and Woese, C. R. (1994). Lessons from an evolving rRNA: 16S and 23S rRNA structures from a comparative perspective. *Microbiol. Rev.* **58**, 10-26.

Hartel, A. J., Lankhorst, P. P. & Altona, C. (1982). Thermodynamics of stacking and of self-association of the dinucleoside monophosphate m62A-U from proton NMR chemical shifts: differential concentration temperature profile method. *Eur. J. Biochem.* **129**, 343-357.

Haushalter, K. A., Lau, J. & Roberts, J. D. (1996). An NMR investigation of the rates of rotation about the C-N bonds in urea and thiourea. *J. Am. Chem. Soc.* **118**, 8891-8896.

He, L., Kierzek, R., SantaLucia, J., Jr., Walter, A. E. and Turner, D. H. (1991). Nearest-neighbor parameter for G•U mismatches: (5'GU3')•(3'UG5') is destabilizing in the contexts (CGUG)•(GUGC), (UGUA)•(AUGU), and (AGUU)•(UUGA) but stabilizing in (GGUC)•(CUGG). *Biochemistry*, **30**, 11124-11132.

Heus, H. A. & Pardi, A. (1991). Novel ^1H nuclear magnetic resonance assignment procedure for RNA duplex. *J. Am. Chem. Soc.*, **113**, 4360-4361.

Ho, P. S., Frederick, C. A., Quigley, G. J., van der Marel, G. A., van Boom, J. H., Wang, A. H.-J. & Rich, A. (1985). G•T wobble base-pairing in Z-DNA at 1.0 Å atomic resolution: the crystal structure of d(CGCGTG). *EMBO J.* **4**, 3617-3623.

Holbrook, S. R., Cheong, C., Tinoco, I., Jr. & Kim, S.-H. (1991). Crystal structure of an RNA double helix incorporating a track of non-Watson-Crick base pairs. *Nature*, **353**, 579-581.

Hunter, W. N., Kneale, G., Brown, T., Rabinovich, D. & Kennard, O. (1986). Refined crystal structure of an octanucleotide duplex with G•T mismatched base-pairs. *J. Mol. Biol.* **190**, 605-618.

Hunter, W. N., Brown, T., Kneale, G., Anand, N. N., Rabinovich, D. & Kennard, O. (1987). The structure of Guanosine-Thymidine mismatches in B-DNA at 2.5-Å resolution. *J. Biol. Chem.* **262**, 9962-9970.

Jiang, F., Kumar, R. A., Jones, R. A. & Patel, D. J. (1996). Structural basis of RNA folding and recognition in an AMP-RNA aptamer complex. *Nature*, **382**, 183-186.

Kanyo, J. E., Duhamel, J. & Lu, P. (1996). Secondary structure of the r(CUUCGG) tetraloop. *Nucl. Acids Res.* **24**, 4015-4022.

Katahira, M., Kanagawa, M., Sato, H., Uesugi, S., Fujii, S., Kohno, T. & Maeda, T. (1994) Formation of sheared G:A base pairs in an RNA duplex modeled after ribozymes, as revealed by NMR. *Nucl. Acids Res.*, **22**, 2752-2759.

Kitamura, K., Mizuno, H., Amisaki, T. & Tomita, K. (1984). Local oscillatory motion of RNA helix derived from linear relationships of backbone torsion angles. *Biopolymers*, **23**, 1169-1184.

Kneale, G., Brown, T., Kennard, O. & Rabinovich, D. (1985). G•T base-pair in a DNA helix: The crystal structure of d(G-G-G-G-T-C-C-C). *J. Mol. Biol.* **186**, 805-814.

Kodama, T. (1997). Doctoral thesis, Graduate school of science, Osaka university.

Kojima, C. (1995) Doctoral thesis, Graduate school of science, Osaka university.

Kyogoku, Y., Lord, R. C. & Rich, A. (1966). Hydrogen bonding specificity of Nucleic Acid purines and pyrimidines in solution. *Science*, **154**, 518-520.

Kyogoku, Y., Lord, R. C. & Rich, A. (1967a). An infrared study of hydrogen bonding between adenine and uracil derivatives in chloroform solution. *J. Am. Chem. Soc.* **89**, 496-504.

Kyogoku, Y., Lord, R. C. & Rich, A. (1967b). The defect of substituents on the hydrogen bonding adenine and uracil derivatives. *Proc. Natl. Acad. Sci. USA*, **57**, 250-257.

Leroy, J. L., Kochoyan, M., Huynh-Dinh, T. & Guéron, M. (1988). Characterization of base-pair opening in deoxynucleotide duplexes using catalyzed exchange of the imino

proton. *J. Mol. Biol.* **200**, 223-238.
Lietzke, S. E., Barnes, C. L., Berglund, J. A. & Kundrot, C. E. (1996). The structure of an RNA dodecamer shows how tandem U-U base pairs increase the range of stable RNA structures and the diversity of recognition sites. *Structure*, **4**, 917-930.
Live, D. H., Davis, D. G., Agosta, W. C. & Cowburn, D. (1984). Long range hydrogen bond mediated effects in peptides: ^{15}N NMR study of gramicidin S in water and organic solvents. *J. Am. Chem. Soc.* **106,** 1939-1941.
Luzzati, P. V. (1952). Traitment statistique des erreurs dans la determination des structures cristallines. *Acta Crystallog.* **5**, 802-810.
Mulder, F. A. A., Spronk, C. A. E. M., Slijper, M., Kaptein, R. & Boelens, R. (1996). Improved HSQC experiments for the observation of exchange broadened signals. *J. Biomol. NMR*, **8**, 223-228.
Nagata, C., Imamura, A. Tagashira, Y. & Kodama, M. (1965). Quantum-mechanical study on the photodimerization of aromatic molecules. *J. Theoret. Biol.* **9**, 357-365.
Orbons, L. P. M., van der Marel, G. A., van Boom, J. H. & Altona, C. (1987). An NMR study of polymorphous behavior of the mismatched DNA octamer d(m^5C-G-m^5C-G-A-G-m^5C-G) in solution. The B-duplex and hairpin forms. *Eur. J. Biochem.* **170**, 225-239.
Pardi, A., Martin, F. H. & Tinoco, I., Jr. (1981) Comparative study of ribonucleotide, deoxyribonucleotide, and hybrid oligonucleotide helices by nuclear magnetic resonance. *Biochemistry*, **20**, 3986-3996.
Pley, H. W., Flaherty, K. M. & Mckey, D. B. (1994). Three-dimensional structure of a hammerhead ribozyme. *Nature*, **372**, 68-74.
Puglisi, J. D., Tan, R., Calnan, B. J., Frankel, A. D. & Williamson, J. R. (1992). Conformation of the TAR RNA-arginine complex by NMR spectroscopy. *Science*, **257**, 76-80.
Puglisi, J. D., Chen, L., Blanchard, S. & Frankel, A. D. (1995). Solution structure of a bovine immunodeficiency virus Tat-TAR peptide-RNA complex. *Science*, **270**, 1200-1203.
Rosenberg, J. M., Seeman, N. C., Day, R. O. & Rich, A. (1976). RNA double helical fragments at atomic resolution: II. The crystal structure of sodium guanidylyl-3',5'-cytidine nonahydrate. *J. Mol. Biol.* **104**, 145-167.
Sakata, T., Hiroaki, H., Oda, Y., Tanaka, T., Ikehara, M. & Uesugi, S. (1990). Studies on the structure and stabilizing factor of the CUUCGG hairpin RNA using chemically synthesized oligonucleotides. *Nucl. Acids Res.* **18**, 3831-3839.
SantaLucia, J., Jr., Kierzek, R. and Turner, D. H. (1991). Stabilities of consecutive A•C, C•C, G•G, U•C and U•U mismatches in RNA internal loops: evidence for stable hydrogen-bonded U•U and C•C+ pairs. *Biochemistry*, **30**, 8241-8251.
Schindelin, H., Zhang, M., Bald, R., Fürste, J.-P., Erdmann, V. A. & Heinemann, U. (1995). Crystal structure of an RNA dodecamer containing the *Escherichia coli* Shine-Dalgarno sequence. *J. Mol. Biol.* **249**, 595-603.
Scott, W. G., Finch, J. T. & Klug, A. (1995). The crystal structure of an all-RNA hammerhead ribozyme: a proposal mechanism for RNA catalytic cleavage. *Cell*, **81**, 991-1002.
Scott, W. G., Murray, J. B., Arnold, J. R. P., Sttoddard, B. L. & Klug, A. (1996). Capturing the structure of a catalytic RNA intermediate: the hammerhead ribozyme. *Science*, **274**, 2065-2069.
Seeman, N. C., Rosenberg, J. M., Suddath, F. L., Kim, J. J. P. & Rich, A. (1976). RNA double helical fragments at atomic resolution: I. The crystal and molecular structure of sodium adenylyl-3',5'-uridine hexahydrate. *J. Mol. Biol.* **104**, 109-144.
Shindo, H. & Zimmerman, S. B. (1980). Sequence dependent variations in the backbone geometry of a synthetic DNA fibre. *Nature*, **283**, 690-691.
Shindo, H., Fujiwara, T., Akutsu, H., Matsumoto, U. & Shimizu, M. (1984). Multiple conformations coexist and interchange in natural B DNA fibers. *J. Mol. Biol.* **174**,

221-229.
Shindo, H., Fujiwara, T., Akutsu, H., Matsumoto, U. & Kyogoku, Y. (1985). Phosphorus-31 nuclear magnetic resonance of highly oriented DNA fibers. 1. Static geometry of DNA double helices. *Biochemistry*, **24**, 887-895.
Sklenár, V., Miyashiro, H., Zon, G., Miles, H. T. & Bax, A. (1986). Assignment of the ^{31}P and ^{1}H resonances in oligonucleotides by two-dimensional NMR spectroscopy. *FEBS Letter*, **208**, 94-98.
Sugeta, H. (1981). Spectrometric determination of formation constants and estimation of molar absorption spectra of individual components in chemical equilibria. Infrared study of intermolecular hydrogen bonding of 2-aminopyridine. *Bull. Chem. Soc. Jpn.* **54**, 3706-3710.
Thomas, G. J. & Kyogoku, Y. (1967). Hypochromism accompanying purine-pyrimidine association interaction. *J. Am. Chem. Soc.* **89**, 4170-4175.
Tomita, K.-I. & Rich, A. (1964). X-ray diffraction investigations of complementary RNA. *Nature*, **201**, 1160-1163.
Tuerk, C., Gauss, P., Thermes, C., Groebe, D. R., Gayle, M., Guild, N., Stormo, G., d'Aubenton-Carafa, Y., Uhlenbeck, O. C., Tinoco, I., Jr., Brody, E. N. & Gold, L. (1988). CUUCGG hairpins: Extraordinarily stable RNA secondary structures associated with various biochemical processes. *Proc. Natl. Acad. Sci. USA*, **85**, 1364-1368.
Varani, G., Cheong, C. & Tinoco, I., Jr. (1991). Structure of unusually stable RNA hairpin. *Biochemistry*, **30**, 3280-3289.
Varghese, A. J. & Wang, S. Y. (1968). Thymine-thymine adduct as a photoproduct of thymine. *Science*, **160**, 186-187.
Wahl, M. C., Rao, S. T. & Sundaralingam, M. (1996). The structure of r(UUCGCG) has a 5'-UU-overhang exhibiting Hoogsteen-like *trans* U•U base pairs. *Nature Struct. Biol.* **3**, 24-31.
Wang, C., Gao, H., Gaffney, B. L. & Jones, R. A. (1991). Nitrogen-15-labeled oligodeoxynucleotides. 3. Protonation of the adenine N1 in the A•C and A•G mispair of the duplexes $\{d[CG(^{15}N^{1})AGAATTCCCG]\}_2$ and $\{d[CGGGAATTC(^{15}N^{1})ACG]\}_2$. *J. Am. Chem. Soc.* **113**, 5486-5488.
Wang, Y.-X., Huang, S. & Draper, R. E. (1996). Structure of a U•U pair within a conserved ribosomal RNA hairpin. *Nucl. Acids Res.* **24**, 2666-2672.
Watson J. D. & Crick F. H. C. (1953). A structure for deoxyribose nucleic acid. *Nature*, **171**, 737-738.
Williams, A. P., Longfellow, C. E., Freier, S. M., Kierzek, R. & Turner, D. H. (1989) Laser temperature-jump, spectroscopic, and thermodynamic study of salt effects on duplex formation by dGCATGC. *Biochemistry*, **28**, 4283-4291.
Wu, M., McDowell, J. A. & Turner D. H. (1995). A periodic table of symmetric tandem mismatches in RNA. *Biochemistry*, **34**, 3204-3211.
Ye, X., Gorin, A., Ellington, A. D. & Patel, D. J. (1996). Deep penetration of an alpha-helix into a widened RNA major groove in the HIV-1 rev peptide-RNA aptamer complex. *Nature Struct. Biol.* **3**, 1026-1033.
Zhang, X., Gaffney, B. L. & Jones, R. A. (1997). ^{15}N NMR of a specifically labeled RNA fragment containing intrahelical GU wobble pairs. *J. Am. Chem. Soc.* **119**, 6432-6433.

Acknowledgments

The present work has been performed under the direction of Professor Yoshimasa Kyogoku, Institute for Protein Research, Osaka University. The author would like to express his sincere gratitude to Professor Y. Kyogoku and Associate professor T. Yamazaki for his cordial guidances, profitable discussions, intimate encouragements, and especially, allowing the author to try the experiments which he wanted to do, throughout the course of his study.

The author would like to express his sincere gratitude to Emeritus Professor K. Tomita, Faculty of Pharmaceutical Science, Osaka University, and Professor S. Fujii, University of Shizuoka, for their cordial guidances, profitable discussions and intimate encouragements. The study on the crystallography was originally performed under their directions. The author would like to express his thanks to Professor S. Uesugi, Yokohama National University, Mr. T. Sakata, Osaka University, Dr. H. Hiroaki, Biomolecular Engineering Institute, for generous gifts of RNA oligomer, their cordial guidances, profitable discussions and intimate encouragements.

The author also wishes to express his cordial appreciation to Professor M. Kainosho, Associate Professor A. Ono, Tokyo Metropolitan University, for their cordial guidances, intimate encouragements and allowing the author to use their utilities. Thanks are also due to Dr. A. Ono and Mr. S. Miyashita, Tokyo Metropolitan University, for their critical advises on the synthesis of the samples.

Furthermore, the author wishes to express his cordial appreciation to Associate Professor H. Shindo and Dr. M. Shimizu, Tokyo College of Pharmacy, for the collaboration on the studies of the triplex, their cordial guidances and intimate encouragements.

During the course of his study, the author has received a number of fruitful discussions, helpful comments and hearty encouragements from many persons. The author also wishes to express his sincere thanks to, Associate professor H. Sugeta and Dr. T. Kodama, Institute for Protein Research, Osaka University, for their intimate encouragements and critical advises on the CD measurements. Thanks are also due to all the members of Division of Molecular Biophysics, Institute for Protein Research, especially, to Drs. T. Kodama, C. Kojima, K. Uegaki and M. Shimizu for the overnights discussions again and again, also to Mr. K. Yasuno, Mrs. I. Yokota-Uchiyama, Miss N. Oda and Mr. K. Teruya for the sample preparations. Special Thanks are due to Dr. T. Kodama and Mr. M. Sugahara, Osaka University, for mental supports when the author really need them.

Finally, the author wishes to sincere thanks to his family for the everything they did for him.

List of Publications

1. Fujii, S., Tanaka, Y., Tomita, K.-I., Sakata, T., Hiroaki, H., Tanaka, T., & Uesugi, S. "Molecular structure of extraordinarily stable RNA: model building and X-ray analysis." (1991). *Nucleic Acids Symposium Series* **25**, 181-182.

2. Fujii, S., Tanaka, Y., Uesugi, S., Tanaka, T., Sakata, T. & Hiroaki, H. "Conformational features of the four successive non-Watson-Crick base pairs in RNA duplex." (1992). *Nucleic Acids Symposium Series* **27**, 63-64.

3. Tanaka, Y. & Kyogoku, Y. "Topology of triple helical DNA." (1995). *Nucleic Acids Symposium Series*, **34**, 89-90.

4. Tanaka, Y., Yokota, I., Oda, N., Uegaki, K. & Kyogoku, Y. "Physicochemical properties of DNA oligomers in the regulatory region of the human interferon-b gene." (1996) *Nucleic Acids Symposium Series*, **35**, 159-160.

5. Tanaka, Y., Kojima, C., Yamazaki, T., Kyogoku, Y., Miyashita, S., Ono, A. & Kainosho, M. "Structure model and physicochemical properties of the C-U mismatch pair in the double stranded RNA in solution." (1997). *Nucleic Acids Symposium Series*, **37**, 271-272.

6. Tanaka, Y., Tomita, K.-I., Sakata, T., Hiroaki, H., Tanaka, T., Uesugi, S., Kyogoku, Y. & Fujii, S. "RNA conformational polymorphism: A'-RNA conformation in the crystal structure of r(UGAGCUUCGGCUC)." (1997) submitted to *Journal of Molecular Biology.*

7. Tanaka, Y., Kojima, C., Yamazaki, T., Kodama, T., Miyashita, S., Ono, A., Kainosho, M. & Kyogoku, Y. "Solution structure and physicochemical properties of the C-U pair in the double stranded RNA." (1998) manuscript in preparation.