OUKA
The University of Osaka
Institutional Knowledge Archive

Title	Precise Structure Determination of a DNA Oligomer by NMR
Author(s)	児嶋, 長次郎
Citation	大阪大学, 1995, 博士論文
Version Type	VoR
URL	https://doi.org/10.11501/3100506
rights	
Note	

The University of Osaka Institutional Knowledge Archive : OUKA

https://ir.library.osaka-u.ac.jp/

The University of Osaka

Precise Structure Determination of a DNA Oligomer by NMR

A Doctoral Thesis
Submitted to Graduate School of Science
Osaka University

1995

Chojiro KOJIMA

Precise Structure Determination of a DNA Oligomer by NMR

A Doctoral Thesis
Submitted to Graduate School of Science
Osaka University

1995

Chojiro KOJIMA

CONTENTS

General Introduction 3

Part I New Developments in the Assignment Methods of DNA Signals

Chapter 1 15
Structure Analysis of a DNA 17mer by Use of Heteronuclear 2D NMR

Chapter 2 20
HAL Experiment, a Novel Filtering Method, and Its Application to DNA Assignment

Chapter 3 25
Filtering Methods for Selection of Singlet and Doublet Signals in NMR Spectra of DNA Oligomers

Chapter 4 32
HSQC Spectrum of the Amino Groups of Nucleic Acids

Part II Precise Structure Determination of a DNA Decamer

Chapter 5 37
NMR Rsym Factor as Analogy of the X-ray Rsym Factor, for Error Estimation of the NOESY Peak Volume

Chapter 6 40
Investigation of Validity of a Single Correlation Time Model as the First Order Approximation for the Estimation of the ^{1}H-^{1}H Distances in a DNA Decamer

Chapter 7 43
Precise Structure Determination of a DNA Oligomer by the Back Calculation Refinement against the NOESY Intensities with Different Mixing Times

Chapter 8 49
Simulation of the Precision and Accuracy of the NMR Structure of a DNA Oligomer by the Metric Matrix Distance Geometry Calculation

Chapter 9 53
NMR Analysis of the Sugar Conformation of a DNA Oligomer with (2'R)-[2'-^{2}H]- and (2'S)-[2'-^{2}H]-Labeling

General Discussion and Conclusion 56

References 59
Acknowledgments 61
List of Publications 62

General Introduction

Contents of this thesis

The contents of this thesis are summarized here.

General Introduction

The general introduction describes the aim of this thesis and the background for understanding this thesis, i.e., the necessity of the study of the DNA structure, the structural property of DNA, semi-quantitative analysis of the conformation of DNA by NMR, the metric matrix distance geometry (mmDG) construction of three dimensional coordinates from local structural informations, NMR structure determination of DNA, the restrained molecular dynamics (rMD) refinement, and the NMR back calculation refinement. In these sections, fundamental knowledges, concepts and technical terms are given with some explanations.

Part I New Developments in the Assignment Methods of DNA Signals

This part describes some assignment methods developed by the author for analyzing the NMR signals of DNA oligomers. In *chapter 1 "Structure Analysis of a DNA 17mer by Use of Heteronuclear 2D NMR"*, the improvement of the assignment strategy for relatively larger DNA oligomers (> 10 kDa) combined with heteronuclear 2D NMR and homonuclear 3D NMR techniques is presented. A DNA 17mer was analyzed in detail as an example. In *chapter 2 "HAL Experiment, a Novel Filtering Method, and Its Application to DNA Assignment"*, a novel NMR technique has been developed and named the HAL experiment. This technique was incorporated in the two dimensional NOESY detection pulse and applied for the analysis of a DNA dodecamer. In *chapter 3 "Filtering Methods for Selection of Singlet and Doublet Signals in NMR Spectra of DNA Oligomers"*, almost all the filtering methods for selection of singlet and doublet signals of the nucleic acids base protons are compared to find the best way for different purposes, such as for signal assignment and for the detection of the interaction of nucleic acids with drugs or proteins. Here, some methods are newly developed and compared with others previously reported. In *chapter 4 "HSQC Spectrum of the Amino Groups of Nucleic Acids"*, the HSQC spectrum of ^{15}N labeled exocyclic cytosine amino groups is obtained for the first time by the use of the spin lock water suppression. Based on the data, the dynamics about the amino group in DNA were investigated semi-quantitatively.

Part II Precise Structure Determination of a DNA Decamer

This part describes some methodological developments proposed by the author for the calculation of the structure of a DNA oligomer. The methods are also related to the precision of the obtained structure. In *chapter 5 "NMR Rsym Factor as Analogy of the X-ray Rsym Factor, for Error Estimation of the NOESY Peak Volume"*, NMR Rsym factor was proposed for the error estimation of the NOESY peak volume. This is a good indicator which sensitively depends on several error sources in the process of the volume determination. In *chapter 6 "Investigation of Validity of a Single Correlation Time Model as the First Order Approximation for the Estimation of the ^1H-^1H Distances in a DNA Decamer"*, the ^1H-^1H

distances in a DNA oligomer were calculated from the NOESY cross peak intensities. The full relaxation matrix approach was tested to obtain more accurate distances in multi spin systems. In this calculation, the isotropic rotational model was adopted for the overall motion of the DNA decamer. This approximation is fairly good for the limited NOESY mixing time range as the first order approximation. In *chapter 7 "Precise Structure Determination of a DNA Decamer by the Back Calculation Refinement against the NOESY Intensities with Different Mixing Times"*, the three dimensional coordinates of a DNA decamer were determined by the back calculation refinement against the NOESY intensities. The precision of the structure was maximized by simultaneous use of the four sets of NOESY intensities with four different mixing times. In *chapter 8 "Simulation of the Precision and Accuracy of the NMR Structure of a DNA Decamer by the Metric Matrix Distance Geometry Calculation"*, the precision and accuracy of the structure of a DNA decamer were checked by the simulation of the metric matrix distance geometry calculations. The simulation indicates the extent of over-determination of the structure of the DNA oligomer by NMR and the necessity of reliable indicators on the amount of structural informations. The author proposed some indicators. In *chapter 9 "NMR Analysis of the Sugar Conformation of a DNA Decamer with (2'R)-[2'-^{2}H]- and (2'S)-[2'-^{2}H]-Labeling"*, vicinal ^{1}H-^{1}H coupling constants were accurately measured by using the (2'R)-[2'-^{2}H]- and (2'S)-[2'-^{2}H]-labeled DNA decamers. It was found that for unlabeled sample, this type of the coupling constant has systematic errors which come from the cross term of the relaxation and scalar coupling.

General Discussion and Conclusion

Finally, the author will discuss about the possibility and limitation of nuclear magnetic resonance as a tool for determining the structure and dynamics of DNA oligomers in solution, and give conclusive evidence for the physicochemical nature of the DNA oligomers treated throughout this thesis.

Aim of this thesis

The purpose of this thesis is to develop techniques for the determination of detailed three dimensional structures of DNA oligomers to find out their physicochemical nature. A DNA oligomer is only a fragment of natural DNA molecule. However, the specific oligomer keeps the character of natural DNA, such as interaction with a specific protein. Therefore, this oligomer is a suitable model for understanding the nature of natural DNA molecule by using NMR spectroscopy.

The structure determination of the DNA double strands by calculation is usually underdetermined. In particular, the overall shape or bend of the molecule is not restrained. Thus, neither the distance geometry nor the ab initio simulated annealing method, which are routinely used for the protein structure calculation, will produce unique structures for DNA. The problem has been avoided by including additional restraints, such as repulsive distance restraints between appropriately chosen phosphate groups of DNA, or by starting the refinement process from proper ideal conformations. For example, A and B type conformations of DNA double strands were used as starting coordinates, and then the convergence of the refined coordinates could be assessed.

Why is the DNA structure not determined by standard methods? Dose neither the distance geometry nor the ab initio simulated annealing method really give unique structures? The questions motivated the author to develop techniques for the determination of detailed three dimensional structures of DNA oligomers in solution.

Necessity of the study of the DNA structure

Deoxyribonucleic acid (DNA) is a key material for the life of a living thing. It contains all informations to maintain one's life. Some of the informations are stored in it as chemical properties of nucleic acids themselves. About 40 years ago, the Watson-Crick model of the DNA structure was proposed (Watson and Crick, 1953). Their double stranded model is still valid and showed us the importance of the structure of DNA for the replication of genetic information. However, a recent progress in detailed structural studies of DNA oligomers shows that their local conformational deviations from model structures depend on their nucleotide sequences. This sequence dependency of the DNA conformation is one of the most essential factor which controls the interaction with proteins. In case of transcription, the expression of stored informations is mainly regulated by the interaction of DNA with several types of DNA binding proteins (for example, many sequence specific binding proteins, λCro, GAL4 and so on). Thus it becomes more and more important to study the conformation of DNA in detail for understanding a living system at the molecular level.

Structural features of DNA

DNA consists of four kinds of purines and pyrimidines bases (A, T, G and C), deoxyriboses and the phosphate groups connected by glycosyl and phosphodiester bonds (Fig. la). There are tremendous amount of possible conformations because there are eleven freedoms of internal rotations around bond axes per each residue (Fig. lb). However, based

on the crystallographic studies of nucleic acids, the well established conformations of the double stranded DNA helices are classified into three families; A, B and Z forms (for example Saenger, 1984), which are shown in Fig. 2. The conformation of DNA is mainly described by the rotation angle around the glycosyl bond (χ), sugar pucker (ν_0, ν_1, ν_2, ν_3 and ν_4) and the conformation around the phosphate group (α, β, γ, δ, ε and ζ). The definitions of some terms used for describing nucleic acids structures are given in Figs. 3 and 4. Additionally, d(AGTC) means that a deoxyadenosine (a deoxyribose connected with an adenine by a glycosyl bond), a deoxyguanosine, a deoxythymidine and a deoxycytidine are connected by phosphodiester bonds from 5' to 3' as shown in Fig. 1a. $d(AGTC)_2$ means that two d(AGTC) chains form a double stranded helix. Similarly d(AAGG) : d(CCTT) means a double stranded helix formed by d(AAGG) and d(CCTT) strands complementarily each other.

Semi-quantitative analysis of the conformation of DNA by NMR

The X-ray crystallographic study is a very powerful method to elucidate the conformation of a nucleic acid molecule. In fact, three representative conformations, A, B and Z forms, have been established by this method. From several experiments, it turns out that the conformation of a nucleic acid molecule in solution is different from that in crystal. For example, a DNA molecule which takes A or Z form in crystal, does B form in solution (for example, Saenger, 1984). So it is necessary to elucidate the conformation of DNA in solution.

As mentioned above, three representative conformations of nucleic acids have been established, but they are just ideal ones. The conformation around each nucleotide residue deviates from the ideal forms. The deviations are important, because they are recognized by DNA binding proteins and then specific interaction between them takes place. Different from the other spectroscopic techniques, NMR spectroscopy gives us the conformational information about each residue separately at an atomic resolution. Thus NMR is a very suitable method to study the sequence-dependent conformation of a DNA in solution.

By NMR spectroscopy, one can get conformational informations of DNA about inter-proton distances and dihedral angles around some bonds. Then three dimensional structure of DNA is constructed based on these informations. In practice these informations are given by two types of two dimensional NMR techniques, NOESY and COSY.

Fig. 5a shows a COSY pulse sequence (Jeener et al., 1979). For simplification, two spins, A and B, coupled with a three bond spin-spin coupling constant will be considered. After the first pulse, the magnetization of an A spin is aligned along a y axis. During a t1 period, the magnetization of an A spin is mixed with that of a B spin through spin-spin coupling. After the second pulse, this mixed magnetization is detected as that of a B spin labeled with an A spin. Thus the cross peak between the resonances of an A spin and a B spin appears in a two dimensional COSY spectrum. A cross peak occurs when coupling is effective, in other words, the coupling constant is not so small. The three bond coupling constant of H-C-C-H varies depending on its dihedral angle, thus appearance or disappearance of a cross peak in a COSY spectrum gives us the information about the dihedral angle.

Fig. 5b shows a NOESY pulse sequence (Jeener et al., 1979). Once again two spins system, A and B, will be considered, but it is not necessary to be coupled by spins. After the

second pulse, the magnetizations of A and B spins relax following the next equations,

$$dM_A / dt = - R_A M_A - R_{AB} M_B, \quad [1]$$

$$dM_B / dt = - R_{BA} M_A - R_B M_B, \quad [2]$$

where M_A and M_B are the deviations of z magnetization from equilibrium values, R_A and R_B are the spin lattice relaxation rates and R_{AB} and R_{BA} are the cross relaxation rates. The amount of M_A is modulated by the amount of M_B during a mixing period, and after the third pulse the magnetization of an A spin labeled with a B spin is detected. Thus the cross peak between the resonances of an A spin and a B spin appears in the two dimensional NOESY spectrum. When the relaxation is dominated by dipole-dipole interaction,

$$R_{AB} = k (6\tau_C / (1 + (2\omega\tau_C)^2) - \tau_C) / r_{AB}^6, \quad [3]$$

where r_{AB} is the distance between two spins, τ_C is a correlation time, ω is a resonance frequency, and k is a constant. When a mixing time τ_m is short, the intensity of the cross peak between A and B spins in a NOESY spectrum, I_{AB}, follows the next equation (Macura and Ernst, 1980),

$$I_{AB} = R_{AB} \tau_m. \quad [4]$$

Thus the ratio of the inter proton distances can be obtained by the following equation,

$$r_{AB} / r_{CD} = (I_{CD} / I_{AB})^{1/6}. \quad [5]$$

If the distance between C and D spins is fixed geometrically, we can get r_{AB} based on a NOESY spectrum by the above equation. In the case of nucleic acids, H2' - H2" (1.78 A), CH5 - CH6 (2.46 A) or TCH_3 - TH6 (2.80 A) could be used as reference distances.

Metric matrix distance geometry (mmDG) construction of three dimensional coordinates from local structural informations

Many types of structural informations (distances, J-coupling data, chemical cross-linking, neutron scattering, predicted secondary structures, etc.) can be conveniently expressed as intra- or intermolecular distances. The distance geometry formalism allows these distances to be assembled and to calculate three dimensional structures consistent with them. It will also uncover any inconsistency in the input data in the process of calculation.

The distance geometry calculation starts at translating bond lengths, bond angles, dihedral angles, improper angles, and van der Waals radii in the initial molecular structure into a (space) matrix of distance bounds between bonded atoms, atoms that are bonded to a common third atom, or atoms that are connected to each other through three bonds, using the equations in Crippen and Havel (1988). Experimental constraints can be added by the NOE and vicinal coupling constant information. The lists of the restraints are read, translated into

distance constraints, and entered into the bound matrix.

Input distances usually define only a small fraction of the interatomic distances in the system. However, they do imply upper and lower bounds on the other atoms to which they are connected by means of triangle inequalities (Crippen and Havel 1988). Determining these implied bounds is equivalent to a well studied computational problem, finding the shortest path between two points in a directed graph. The shortest path algorithm used in this implementation is that of Dijkstra (see Tarjan, 1983, for a review). It proceeds by calculating the shortest paths from one atom (the root) to all other atoms. The shortest path routine calculates both upper and lower bounds by keeping track of two halves of the graph, one that holds the upper bounds and one that holds the lower. By solving the shortest path problem for the upper bounds first, the program can then calculate the lower bounds by continuing the shortest path routine with the second half of the graph. This process is repeated, with each atom taking its turn as root.

Once upper and lower bounds are known for all interatomic distances in the system, a matrix of actual distances chosen between those bounds must be created in order to calculate coordinates. This is done for each pair of atoms in the system with a uniform distribution of random values. Since the initial upper and lower bounds matrices are consistent with all possible conformations of the molecule, interatomic distances chosen independently of each other are, in general, inconsistent with any particular conformation.

The process of the calculation of coordinates from the resulting actual distance matrix is called embedding. Briefly, the distance matrix is converted to a metric matrix of distances centered around their collective centroid, as described by Crippen and Havel (1988). If the three largest eigenvalues of this metric matrix are all positive, their corresponding eigenvectors contain the x, y, and z coordinates of all the atoms. If the three largest eigenvalues are not positive, the chosen distances are too inconsistent to be embedded.

The expected radius of gyration of the embedded coordinates can be calculated from the complete matrix of interatomic distances produced for the embedding. This expected radius is often larger or smaller than the radius of gyration calculated from the embedded coordinates. Therefore, a scaling factor equal to the ratio of the expected to actual radii of gyration is calculated. The embedded coordinates should be multiplied by this factor, because it makes subsequent regularization easier.

A motorization procedure to improve the consistency of the chosen distances is available. Briefly, the bounds matrix is smoothed for one root, the distance from the root atom to another atom is chosen, and the procedure is repeated, changing roots after the distances from the current root to all other atoms have been chosen. This procedure ensures that later choices of inter atomic distances are consistent with earlier ones. However, it requires considerably more CPU time than the actual embedding because the bounds matrix is frequently resmoothed. It also creates a sampling problem because the order in which distances are chosen has a great impact on which parts of the molecule explore their conformational space completely.

This ordered motorization protocol can be modified to improve the conformational sampling of the coordinates produced. Random motorization picks the root atoms in random order, so that the molecule's conformational freedom is not necessarily used up in just one

region. This modification in the distance geometry calculation gives equally good conformational sampling and reduces the CPU time requirements. In partial motorization, the bounds matrix is resmoothed after choosing distances from only a fraction of the root atoms, after which distances are chosen from the other atoms without resmoothing. This procedure works reliably, even with only four root atoms used in the retightening phase.

NMR structure determination of DNA

The determination and refinement of solution NMR structures is based on inter proton distance estimates, coupling constants measurements, and other information, such as hydrogen bonding patterns.

The structure determination of DNA double strands by calculation is usually underdetermined. In particular, the overall shape or bend of the molecule is a free parameter. Thus, neither distance geometry nor ab initio simulated annealing will produce unique structures. The problem can be avoided by including additional restraints, such as repulsive distance restraints between appropriately chosen phosphate groups in the case of DNA, or by starting the refinement process from ideal conformations. For example, A and B type conformations of DNA double strands could be obtained from other programs (QUANTA, AMBER, MIDAS and so on). These starting coordinates should be refined and then the convergence of the refined coordinates should be assessed.

Restrained molecular dynamics (rMD) refinement

The molecular dynamics is the computational simulation of behavior of the molecule based on the empirical potential energy (for example, Mccammon and Harvey, 1987). Iterative integration of the Newton's equation is carried out. Through simulation, the molecule escapes from a local minimum and reaches a global minimum of the energy by appropriate assessment of the temperature of the system. The structure obtained should not depend on its initial structure. In the restrained molecular dynamics, the geometrical information given by NMR spectroscopy such as inter proton distances and dihedral angles about some bonds are incorporated as pseudo potential energies. The added pseudo energies work to reduce the probable range of configurational space by formally increasing the fictitious energy of structures which do not conform to experimental results of NMR and lead the molecule to the potential minimum efficiently. Thus, a precise structure is constructed by means of NMR spectroscopy.

Back calculation refinement

This section describes how to back calculate the NOESY spectrum from the three dimensional structure of a molecule, and how to refine an initial model by directly minimizing the difference between the observed 2D NOE intensities and those calculated by the full relaxation matrix approach. Spin diffusion effects are fully accounted during the refinement. Derivatives with respect to atomic coordinates are analytically calculated (Yip and Case, 1989). An example of the refinement with this method was given by Nilges et al. (1991), (for further reading, see Ernst, Bodenhausen, and Wokaun, 1987; James et al., 1991; Keepers and James, 1984; Lipari and Szabo, 1982; Macura and Ernst, 1980; Nilges et al., 1991; Solomon,

1955).

The basis for the refinement is the calculation of the volume of a cross peak between spins i and j, I^C_{ij}, from the atomic coordinates by means of the relaxation matrix **R** (Macura and Ernst 1980; Ernst et al. 1987; Keepers and James 1984),

$$I^C_{ij} \propto [\exp(-\mathbf{R}\tau_m)]_{ij}, \qquad [6]$$

where τ_m is the mixing time. The **R** is a function of the transition rates Ω^{ij},

$$\mathbf{R}_{ij} = \begin{cases} \Omega_2^{ij} - \Omega_0^{ij} & \text{if} \quad i \neq j \\ \sum_{k \neq j} \Omega_0^{kj} + 2\Omega_1^{kj} + \Omega_2^{kj} + R_{leak} & \text{if} \quad i = j \end{cases}. \qquad [7]$$

R_{leak} describes the non-NOE magnetization losses from the lattice. Ω^{ij} are defined by spectral densities and dipolar coupling strengths as following (Solomon 1955),

$$\begin{aligned} \Omega_0^{ij} &= d_{ij} J(0) \\ \Omega_1^{ij} &= \frac{3}{2} d_{ij} J(\omega) \\ \Omega_2^{ij} &= 6 d_{ij} J(2\omega) \end{aligned} \qquad [8]$$

and

$$d_{ij} = \frac{\gamma^4 h^2}{40\pi r_{ij}^6} \qquad [9]$$

where γ is the gyromagnetic ratio of proton and r_{ij} is the distance between spins i and j. At present, only protons can be used in the refinement.

In the simplest model, it is assumed that a single isotropic correlation time τ_C is sufficient to describe the spectral densities $J(\omega)$ (Solomon, 1955),

$$J(\omega)_{ij} = \frac{\tau_c}{1 + (\omega\tau_c)^2}. \qquad [10]$$

Groups of protons whose resonances are degenerate the most of the methyl protons due to rapid motion are treated as roughly as follows in CORMA, version 1.5 (Keepers and James, 1984). Each such group is represented by one spin, whose intensity is scaled by the number of protons in the group, and the distance to the group is calculated as the $\langle r^{-3} \rangle^{-1/3}$ average over the protons in the group [Eq. 9]. In addition, a diagonal leakage rate is added to each group of protons.

The derivative of I^C_{ij} with respect to a coordinate μ (Eq. 12 of Yip and Case, 1989) can be written as

$$\nabla_\mu I^C_{ij} = \nabla_\mu [\exp(-\mathbf{R}\tau_m)]^{(ij)} = \text{Trace}[(\nabla_\mu \mathbf{R})\mathbf{L}\mathbf{F}^{(ij)}\mathbf{L}^T], \qquad [11]$$

where $F^{(ij)}$ is defined as

$$F_{ru}^{(ij)} \equiv \begin{cases} -L_{ir}L_{uj}^{T}\dfrac{\exp(-\lambda_r\tau)-\exp(-\lambda_u\tau)}{\tau(\lambda_r-\lambda_u)} & \text{if} \quad r \neq u \\ L_{ir}L_{rj}^{T}\exp(-\lambda_r\tau) & \text{else} \end{cases} \qquad [12]$$

L and **Λ** are the matrices of eigenvectors and eigenvalues of **R**, respectively. λ_r is the rth eigenvalue of the relaxation matrix, **R**.

Once the NOESY spectrum and its gradient are calculated, the relaxation energy $E_{relaxation}$ can be expressed as a function of the difference between (the functions of) observed and calculated intensities, and analytic derivatives with respect to atomic coordinates can be readily obtained by using the chain rule

$$E_{relaxation} = W_N \sum_{Spectra} \sum_{i=1}^{N_S} \omega_i \cdot \text{well}(I_i^C, k_S I_i^O, \Delta_i, n)^m \qquad [13]$$

where W_N is the energy constant for the relaxation term, I^C_i and I^O_i are respectively the calculated and observed intensities, Δ_i is an error estimate for I^O_i, ω_i is a weight factor, k_S is the calibration factor for each spectrum, and N_S is the number of cross peaks in each spectrum. The function well(a, b, Δ, n) is defined as the absolute value of the difference between the nth powers of a and b, where b has an error estimate Δ,

$$\text{well}(a,b,\Delta,n) \equiv \begin{cases} (b-\Delta)^n - a^n & a^n < (B-\Delta)^n \\ 0 & (B-\Delta)^n < a^n < (B+\Delta)^n \\ a^n - (b+\Delta)^n & a^n > (B+\Delta)^n \end{cases} \qquad [14]$$

The individual error estimates Δ_i reflect the errors in the peak volumes, usually subjective estimates, especially due to noise and spectral overlap. At present, the error estimates are also used to ascertain if a measured intensity is to be used in the determination of the overall calibration factor.

Values for the exponents of n = 1 / 2 and m = 2 (Eqs. 13 and 14) correspond to the refinement of the residual in X-ray crystallography. These values tend to put a high weight on large intensities, resulting in worse fit of intensity for which calculated value is too small. Following a suggestion by James et al. (1991), we use m = 2 and n = 1 / 6. The discontinuity of the gradient may lead to instability during the refinement.

In addition to the overall weight W_N, individual weights ω_i can be applied to each term in the sum in Eqs. 13 and 16, e.g., in order to increase the relative weight of the small intensities. (It should be noted that this is achieved already by setting n = 6 in Eq. 13.) The scheme $\omega_i = (1 / I^O_i)^{m/2}$ corresponds to a common weighting scheme used in crystallography if experimental σ values are unreliable or unavailable. In the NMR case, however, there is no theoretical justification for this weighting scheme. (In crystallography, the statistical error σ of an intensity measurement I^O_i is $(I^O_i)^{1/2}$.) The weights are scaled such that $\min(\omega_i) = 1$.

The calibration factor between observed and back calculated intensities is determined simply as the ratio of the sums of all calculated and observed intensities,

$$k_S = \frac{\sum_{i=1}^{N_S} I_i^C}{\sum_{i=1}^{N_S} I_i^O}. \qquad [15]$$

This ratio can be determined separately for each spectrum, or overall for all data points. Volumes that are not very reliable (i.e., they have a large error estimate) can be excluded from the ratio. The calibration factor can be updated automatically at every step. For technical reasons, this does not allow the calculation of the exact derivatives at present, so the calibration should not be updated automatically at every step during conjugate gradient minimization. During annealing, the effects due to the error in the gradient are negligible.

As a measure of the fit of the refined coordinates to the NOESY data, a generalized R value is used:

$$R^n = \frac{\sum_{Spectra} \sum_{i=1}^{N_S} \omega_i \cdot well(I_i^C, k_S I_i^O, \Delta_i, n)}{\sum_{Spectra} \sum_{i=1}^{N_S} \omega_i \cdot (k_S I_i^O)^n} \qquad [16]$$

For $n = 1 / 2$, $\Delta_i = 0$, and $\omega_i = 1$, R^n corresponds to the R value commonly used in crystallography (Stout and Jensen 1989). For $n = 6$, R^n corresponds to the value suggested by James et al. (1991).

^1H-^1H distance calculation by the NOESY back calculation (MARDIGRAS)

MARDIGRAS is an NOE derived distance refinement program, and must therefore be used in conjunction with other computational techniques from generating and refining geometrically consistent molecular structures from distance information. The two computations particularly suited to this purpose are distance geometry and restrained molecular dynamics (X-PLOR, AMBER and so on). The theoretical back ground is described in the preceding section. The primary assumptions behind MARDIGRAS are (1) that the overall tumbling is isotropic and (2) that, aside from small local motions such as the methyl rotation, the molecule is rigid and the inter proton distance can be described by a single number.

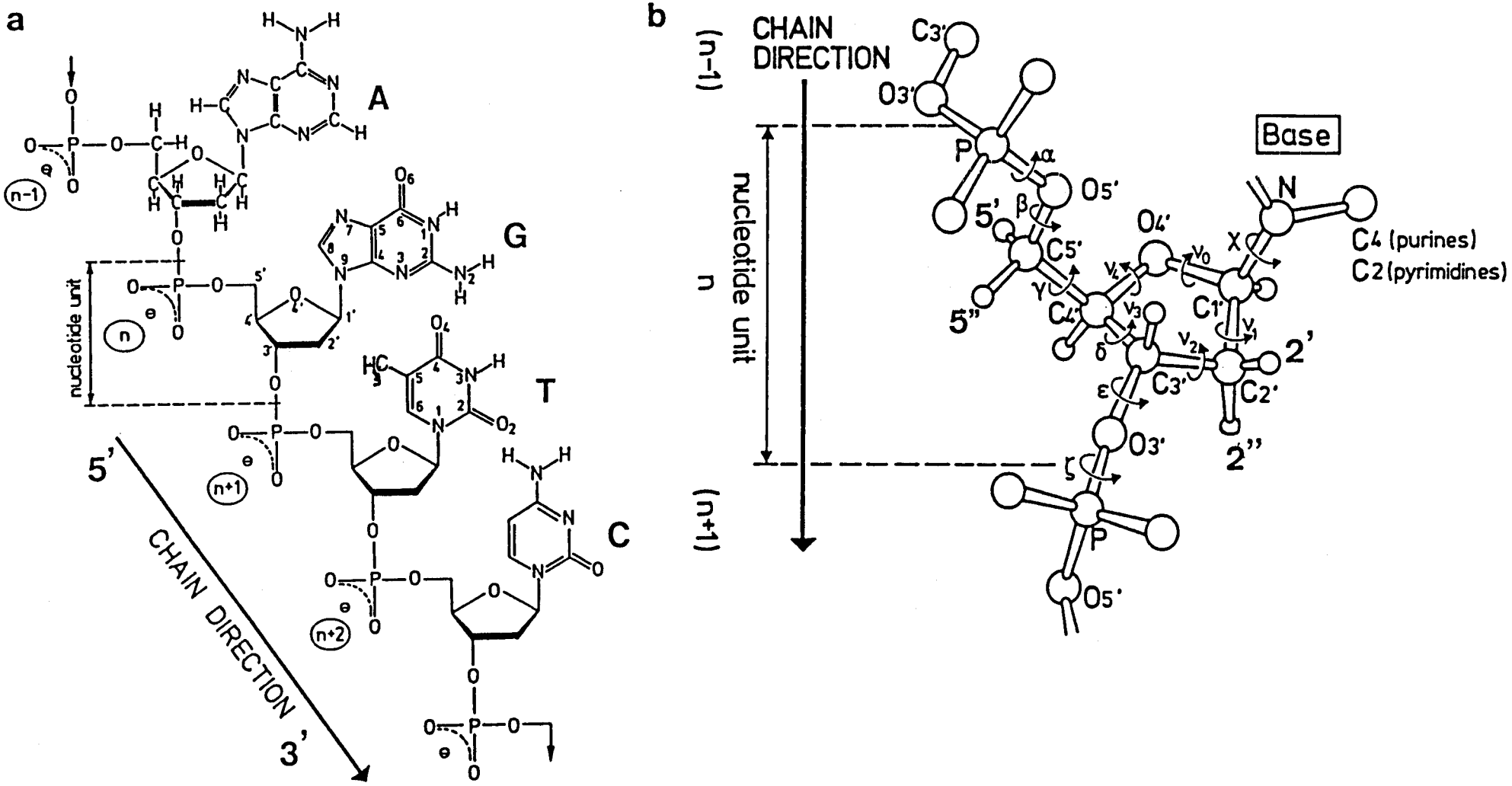


Figure 1 (a) The composition of DNA. A; Adenine, G; Guanine (purine bases), T; Thymine, C; Cytosine (pyrimidine bases). (b) Atomic numbering scheme and definition of dihedral angles of DNA. Here a deoxyribose is shown. When H2" is replaced by a hydroxy group, the sugar is termed as a ribose.

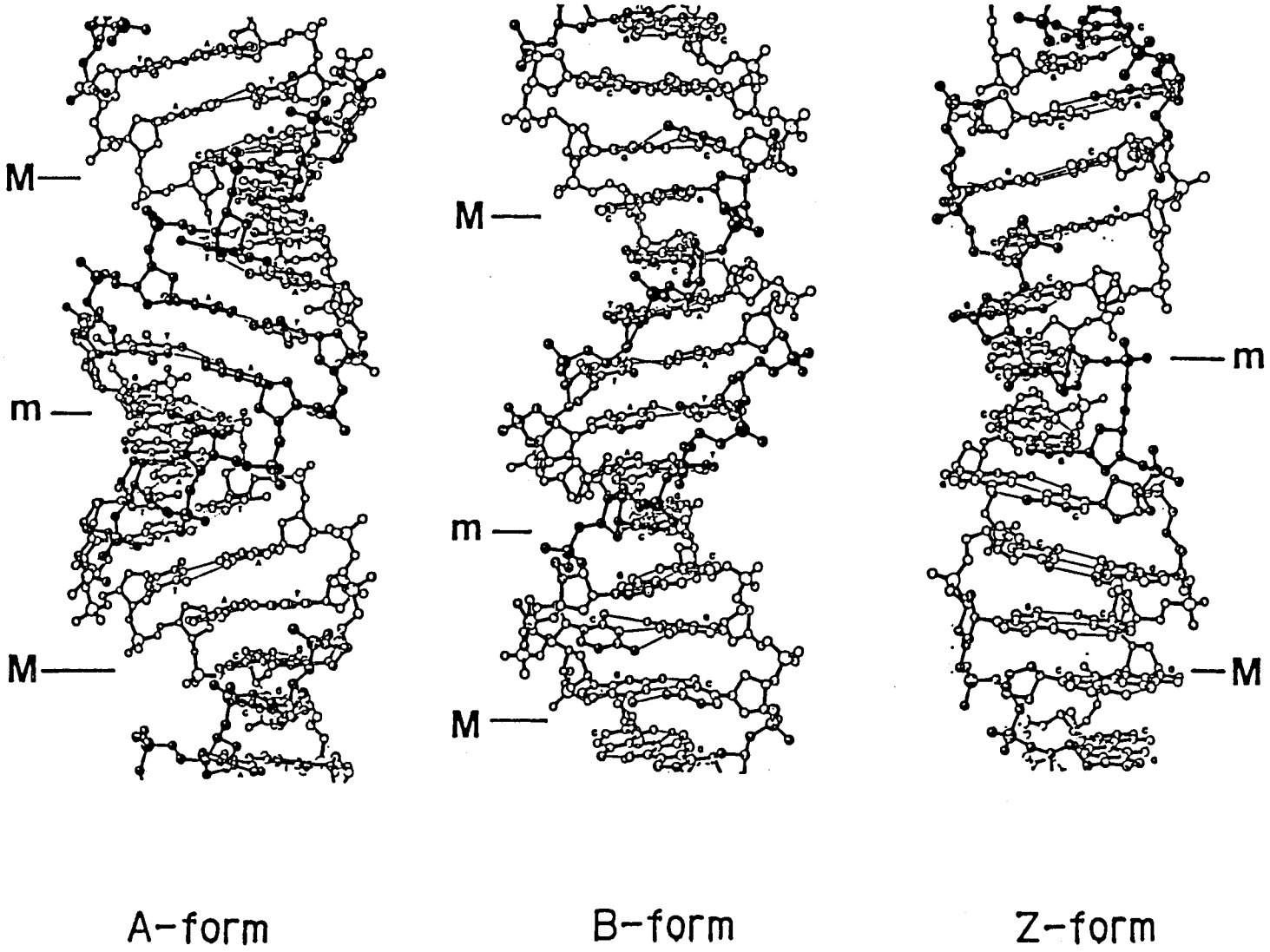


Figure 2 Three typical conformations of DNA; A, B and Z forms. M and m indicate the major and minor grooves, respectively.

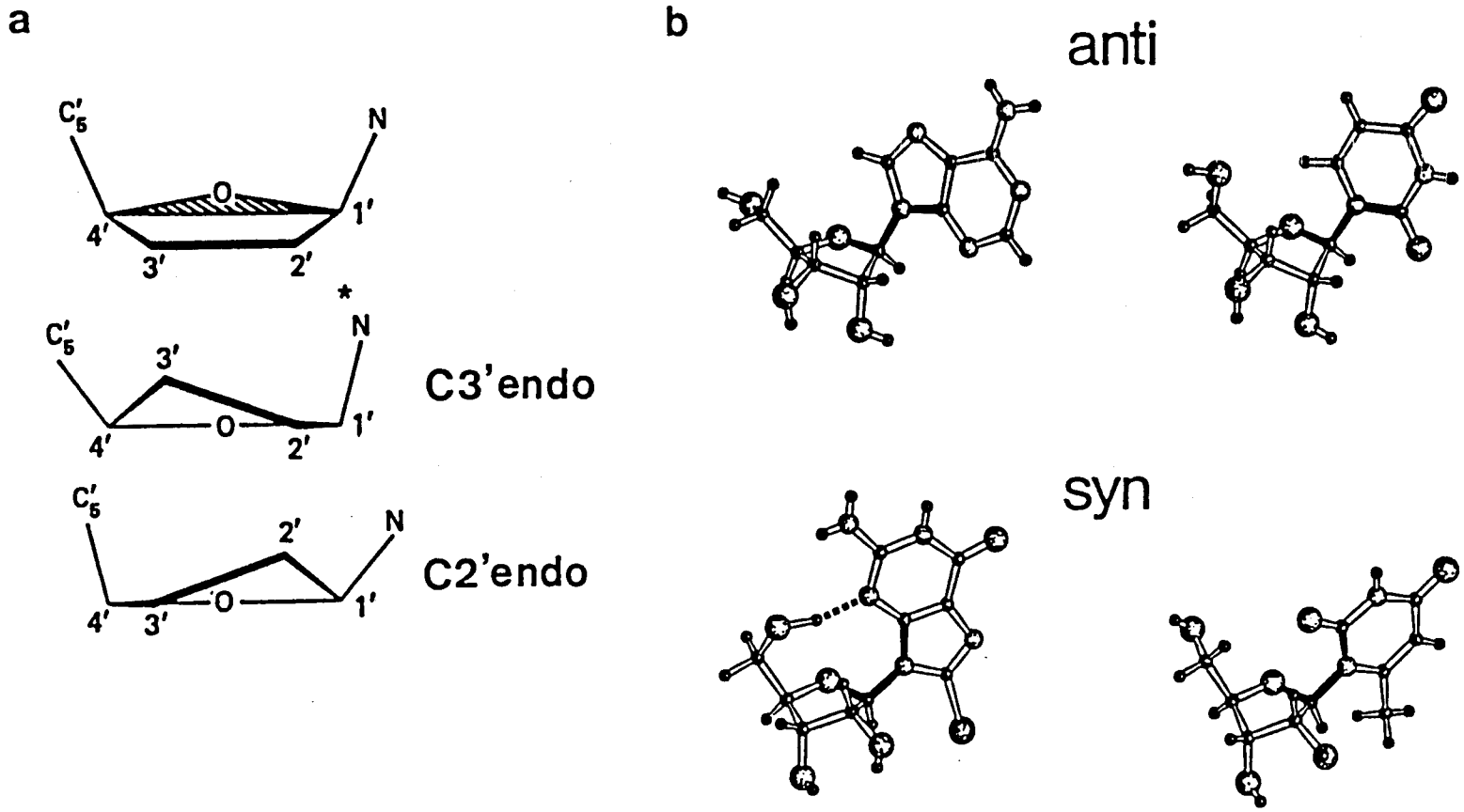


Figure 3 (a) Definition of sugar puckering; C3' endo and C2' endo. (b) Definition of anti and syn about a glycosyl bond. Shadowed bonds are noticed.

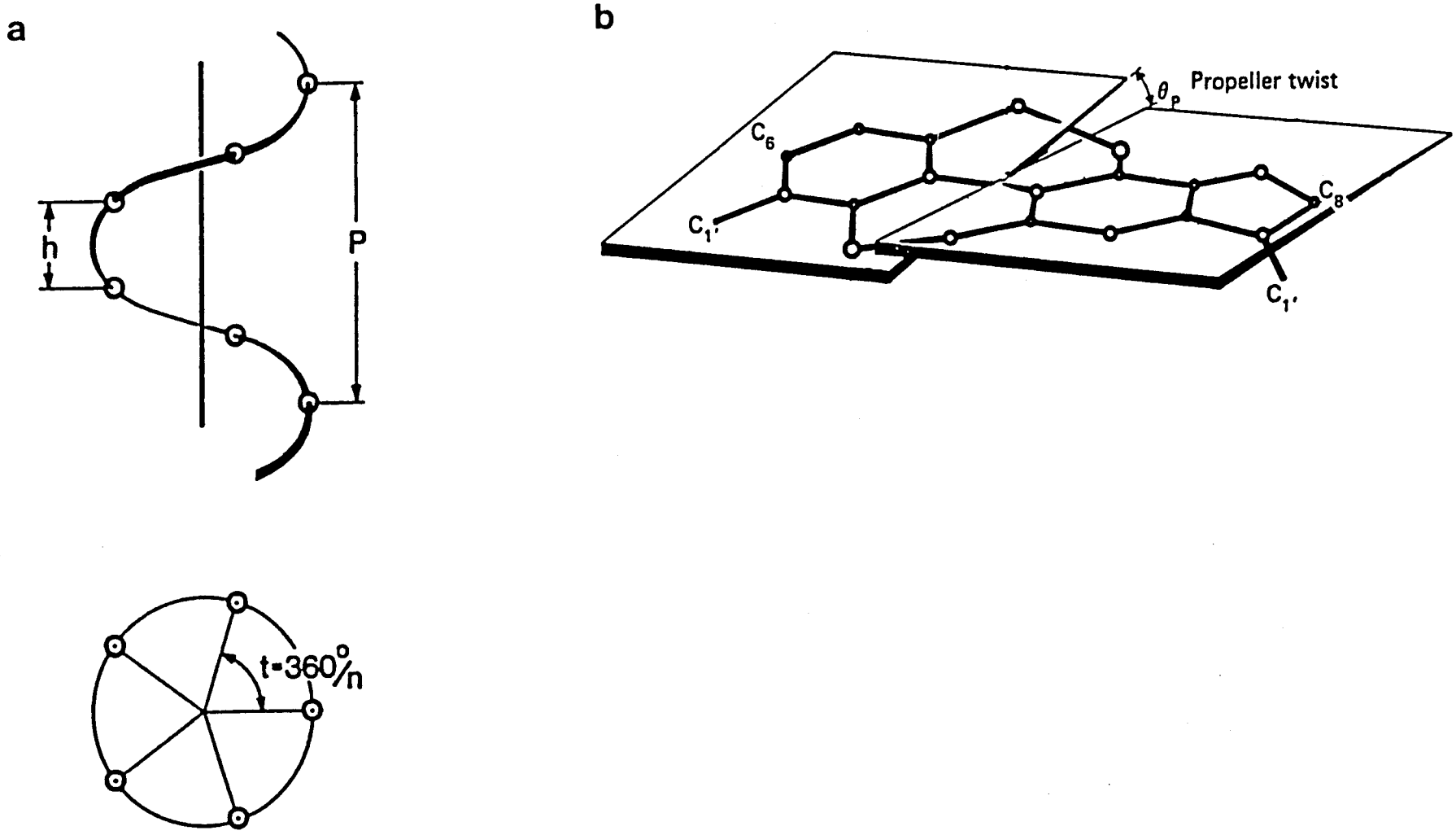


Figure 4 (a) Definition of helical parameters pitch P, axial rise per residue h, and twist t, shown for a right-handed helix with n = 5 residue per turn. (b) Definition of propeller twist.

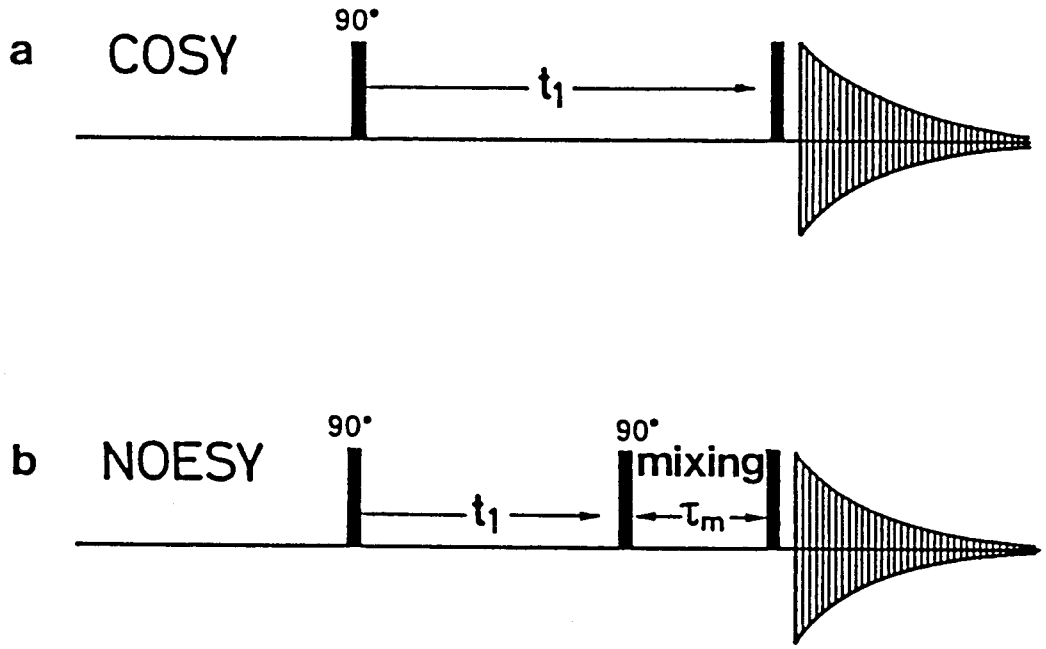


Figure 5 Pulse sequences in two dimensional NMR. (a) COSY (b) NOESY.

Part I

New Developments in the Assignment Methods of DNA Signals

Chapter 1

Structure Analysis of a DNA 17 mer by Use of Heteronuclear 2D NMR

ABSTRACT

DNA oligomer d(CGGAAGACTCTCCTCCG) : d(CGGAGGAGAGTCTTCCG) named UAS_G (17 mer, M.W. = 11000) was studied by ^1H NMR and heteronuclear two dimensional (2D) NMR. All the labile protons and half of the non-exchangeable protons were assigned by the use of conventional ^1H 2D experiments including NOESY with the 1-1 echo excitation for water suppression. Signal degeneracy in the sugar proton region made it difficult to make assignments of the remaining half of the non-exchangeable protons of the oligomer in ^1H 2D spectra. Here we report a new protocol using $^1H/^{13}C$ and $^1H/^{31}P$ HSQC spectra combined with homonuclear three dimensional NOESY-TOCSY. By this protocol, most of the proton resonances of the oligomer have been assigned, and it turned out that the whole conformation of the oligomer takes a B form like structure with a local conformational deviation from the standard, i.e., unwinding around the 8th cytosine at the center of 17 base pairs.

INTRODUCTION

A DNA oligomer duplex d(CGGAAGACTCTCCTCCG) : d(CGGAGGAGAGTCTTC CG) named UAS_G (17 mer, M.W. = 11000) has an identical base sequence to the UAS_G sequence in the GAL1-GAL10 promoter of *Saccharomyces cerevisiae* which is a binding site of the GAL4 protein. The UV-CD study showed that the overall structure of the oligonucleotide is in the B form (Hanssen et al., 1991). Here, we report the complete proton resonance assignment of the 17mer DNA and a local conformational deviation of the molecule detected by means of ^1H and heteronuclear NMR. This is the first report of the detailed NMR analysis of a oligonucleotide over 16 mer.

MATERIALS AND METHODS

Two strands of the UAS_G oligonucleotide 17 mers were chemically synthesized by a solid phase automated DNA synthesizer (ABI 381B). The sequences of the DNA oligomers were 5'-CGGAAGACTCTCCTCCG-3' (UAS_G+), 5'-CGGAAGACTCTCCTCCG -3' (UAS_G-). These oligomers were deblocked and purified by a reverse phase C18 HPLC column twice, first dimetoxy-triphenyl-methyl-protection-group-on (Tr-on) and second Tr-off, with the acetonitrile (0.1M triethylamine acetate) gradient. The counter ion was changed to Na^+ by an anion-exchange FPLC column, followed by passing a gel filtration HPLC column to remove extra salts. The UAS_G+ and UAS_G- strands were mixed at 1 : 1 ratio which was estimated by the nuclease P1 hydrolysis method, then the solution was lyophilized and dissolved in a solution with 10 mM sodium phosphate buffer and 150 mM NaCl. The samples for NMR measurements comprised 3 mM or 6 mM DNA at neutral pH.

All the homonuclear ^1H 2D spectra were recorded on JEOL GX500 and GSX500 spectrometers, and all the heteronuclear experiments were done with Bruker AM500 and AMX500 spectrometers. The NOESY spectrum in H_2O was observed by the replacement of the third single detection pulse with the 1-1 echo detection pulse for solvent suppression (Sklenar et al., 1987). Homonuclear three dimensional NOESY-TOCSY spectrum was recorded with the parameters previously reported (see Piotto et al., 1991). In the $^1H/^{13}C$ heteronuclear single quantum correlation (HSQC) experiment, the carrier frequency for $\omega 1$ (^{13}C) was set at 140 ppm (the base carbon resonance region), or at 70 ppm (the sugar carbon resonance region). Total measuring time for each experiment was 48 hours. In the $^1H/^{31}P$ HSQC spectrum, the carrier frequency for $\omega 1$ (^{31}P) was set at -1 ppm (the middle of the phosphate buffer resonance region and the phosphate resonance region of the DNA oligomer).

RESULTS AND DISCUSSION

Assignments of labile proton signals

All the imino proton signals appeared in the 12~14 ppm range of chemical shifts. Then we supposed that all base pairs of the UAS_G 17 mer DNA duplex were in the Watson-Crick type and sequentially stacked on each other. The validity of this assumption was empirically justified for the double stranded DNA whose sequence was possible to form the Watson-Crick base pairs. Under this assumption, all labile proton resonances were assigned selfconsistently by measuring the NOESY spectrum in H_2O. In the NOESY spectrum, many cross peaks among imino protons, adenine C2 protons and cytosine amino protons were apparently observed.

Assignments of non-labile proton signals

Because of signal degeneracy, only a part of non-exchangeable proton resonances could be assigned by standard NOESY and DQF-COSY. The incompleteness of the assignment leads to ambiguity of the analysis. In order to make complete assignments, the author constructed a new assignment protocol as follows (Fig. 1).

New protocol for the assignment of non-labile proton signals of larger DNA oligomers. The reduction of the ambiguity of the assignment was achieved by three additional experiments, i.e., ^1H three-dimensional NOESY-TOCSY (Piotto et al., 1991), $^1H/^{13}C$ heteronuclear single quantum correlation (HSQC) and $^1H/^{31}P$ HSQC experiments. Then, standard experiments such as NOESY, DQF-COSY and TOCSY were combined with additional ones in the following four steps. In the first step, the base proton resonances were identified by $^1H/^{13}C$ HSQC spectrum. The second step was sequential assignments, in which the base proton resonances and the sugar 1', 2' and 2" ones were connected by the ^1H-^1H distance correlated spectra, NOESY and 3D NOESY-TOCSY, using the identified resonances in the first step. Third, the assignments of intra nucleotide sugar proton signals were achieved by the ^1H-^1H bond and/or distance correlated spectra, DQF-COSY, TOCSY, NOESY, 3D NOESY-TOCSY. These second and third operations were checked complementarily, step by step. Finally the chemical shifts of the sugar 3', 4', 5' and 5" proton resonances were confirmed by $^1H/^{31}P$ HSQC, if possible. Since the chemical shift dispersion of the phosphorus nuclei of double stranded DNA oligomers was narrow, this step was difficult to perform. The phosphorus chemical shift reflects the conformation around the phosphate group, i.e., two dihedral angles α and ζ, the O-P-O bond angle and the P-O bond length. Then, this step was useful not only for assignment but also for detection of the locally perturbed conformation around the phosphate groups.

Identification of the base proton resonances by $^1H/^{13}C$ HSQC. All the chemical shifts of the base proton resonances were identified by the $^1H/^{13}C$ HSQC spectrum at natural abundance. They were not used for the sequence specific assignment directly, but peak separation increased with their coupled ^{13}C chemical shifts (Leupin et al., 1987; Ashcroft et al., 1991). Although these experiments clearly identified most of the base proton resonances following sequence, this method was not sufficient for the identification of the sugar proton resonances because of overlapping of both the ^1H and ^{13}C chemical shifts. They should be

used for the identification of atom species, i.e., which resonance was from 1', 2', 3', 4' and 5'. This information was helpful because the 4' and 5'/5" proton resonances were almost completely overlapping.

Sequential connectivities by the ^1H-^1H correlation spectra. By using the identified proton chemical shifts at step 1, NOESY, DQF-COSY, TOCSY and homonuclear three-dimensional (3D) NOESY-TOCSY (Piotto et al., 1991) spectra were measured to identify the sequential connectivity such as those between base protons and sugar 1', 2', 2" protons. First, the cross peaks between cytosine 5H and 6H protons were easily identified in the DQF-COSY and TOCSY spectra. The inter residue sequential distance connectivities between the base protons and the sugar 1' protons of the adjacent nucleotide were obtained in the same NOESY spectrum. However, the overlapping of weak sequential cross peaks with the intense intra base cytosine 5H/6H ones made it difficult to trace sequential connectivities in conventional 2D experiments. This problem often happens in case of relatively larger G/C rich DNA oligomers. The author solved this problems by using both the $^1H/^{13}C$ HSQC and 3D NOESY-TOCSY experiments. In the former experiment, the ^{13}C dimension was used to separate the crowded one dimensional base proton spectrum, and also to homogenize each peak intensity. The latter introduced the TOCSY transfer between the intra base cytosine 5H and 6H resonances, and/or among the sugar proton resonances, as an additional axis to separate the crowded NOESY two dimensional spectrum.

Identification of the other sugar proton resonances. All the other sugar protons were identified by ^1H 3D NOESY-TOCSY, NOESY, TOCSY, DQF-COSY and NOESY. DQF-COSY spectra did not help much because of serious signal cancellation caused by signal overlapping. NOESY spectrum was most useful because of the connectivities from well separated resonance regions, such as base or sugar 1', 2' and 2" proton ones, to the other sugar proton region. Almost all the sugar proton resonances, except the 5' and 5" proton resonances, were assigned. Moreover, stereospecific assignments were also achieved for 2' and 2" proton resonances. This step was crossly checked by the previous sequential assignment.

Confirmation of the assignments by correlation with ^{31}P resonances. The identifications of the chemical shifts of the sugar 3', 4', 5' and 5" proton resonances were confirmed by $^1H/^{\beta 1}P$ HSQC. Unfortunately, the chemical shift dispersion of the phosphorus resonances was very narrow, so the identification was not clear. Their narrow dispersion means that all the local conformations around the phosphate group are similar to each other.

Conformation Analysis

Following the above strategy, the author could assign almost all the proton resonances of UAS_G and analyze the conformation of the oligonucleotide. The distance information derived from the NOESY experiments and the J information derived from the DQF-COSY experiments identified that the overall structure belongs to the B form family. The result is consistent with the previous CD work (Hanssen et al., 1991), in which the overall conformation of this molecule was supposed to be in the B form. However, when the author looked at the NMR data carefully, he noticed unusual chemical shifts around the 8th cytosine, i.e., at the polypurine/polypyrimidine junction of this 17 mer DNA. Not only the unusual chemical

shifts, the sugar puckering and some distance correlation could not be accounted for the standard B form conformation. The resonances of the 8th cytosine amino protons showed somewhat unusual chemical shifts (Fig. 2). From some NOESY distance informations, we supposed that the local conformation of the junction might be a little unwound from the helix of the B form DNA. These conformational properties of the polypurine/polypyrimidine junction was pointed out previously, as known as Calladine rule (Saenger, 1984).

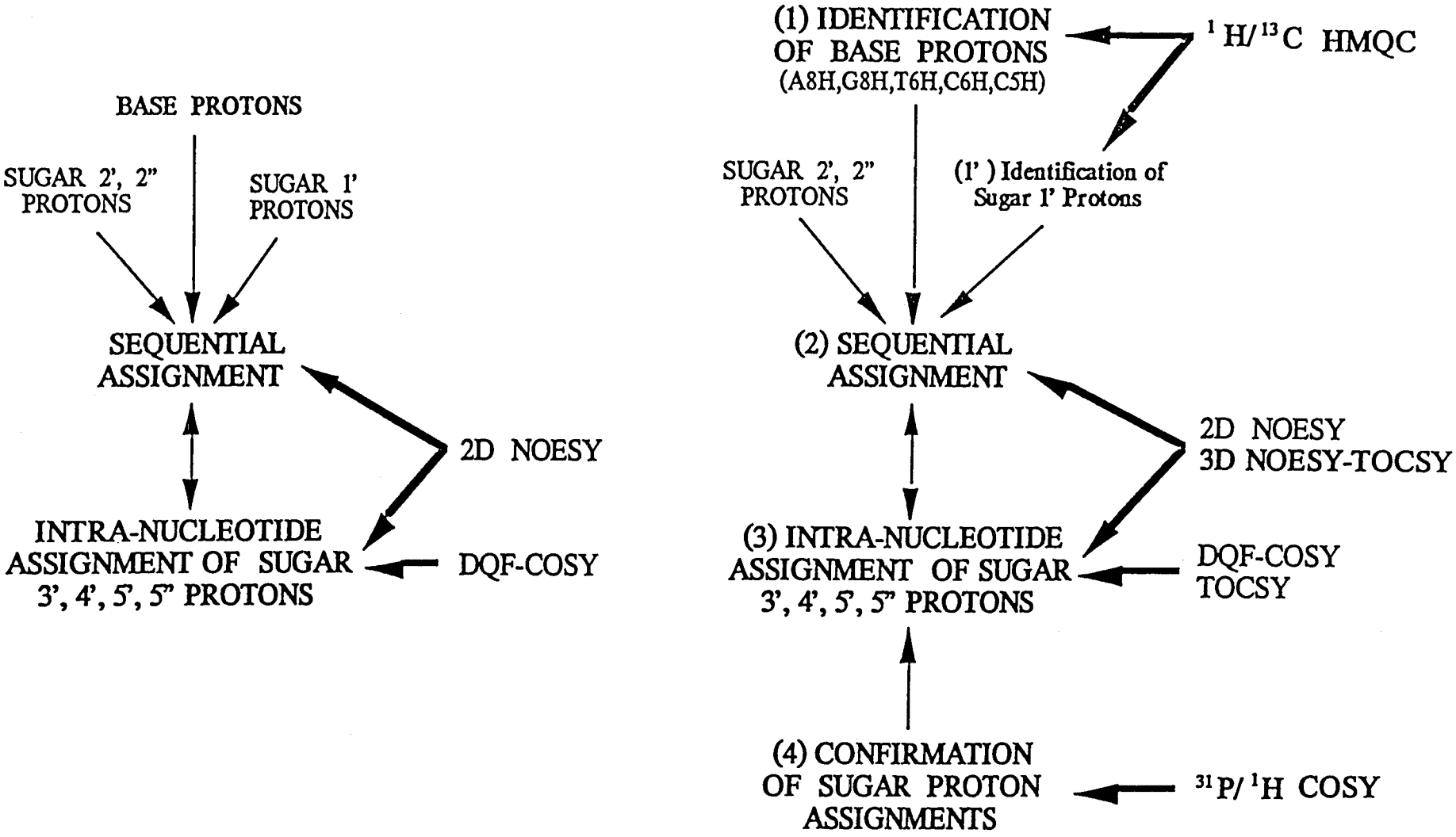


Figure 1. The schematic presentation of the previous (left) and the present (right) protocol for the assignment of the non-labile proton signals of larger DNA oligomers.

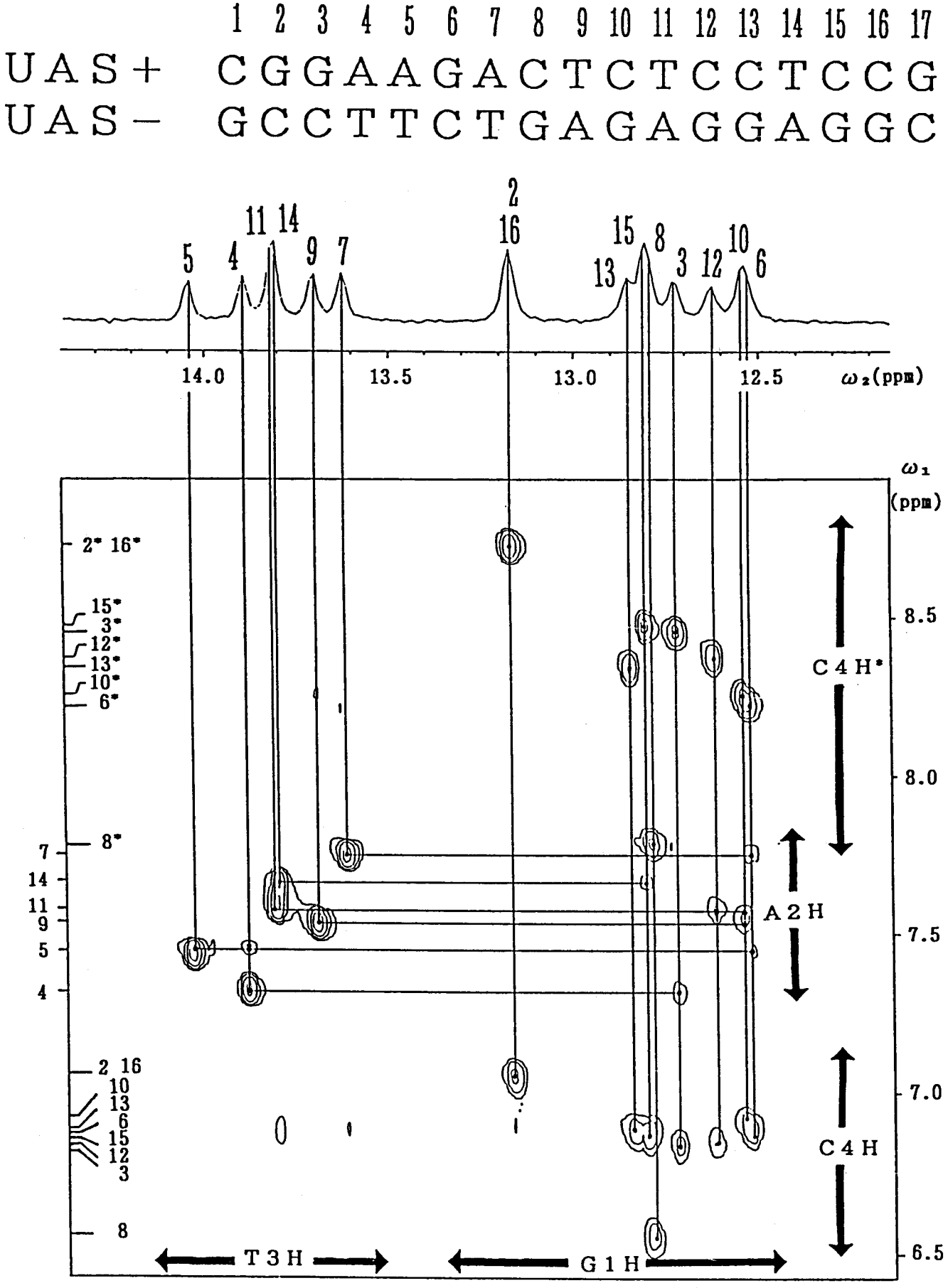

1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17
UAS+ CGGAAGACTCTCCTCCG
UAS− GCCTTCTGAGAGGAGGC
5 4 11 14 9 7 2 16 13 15 8 3 12 10 6
14.0 13.5 13.0 12.5 ω2(ppm)
ω1 (ppm)
2* 16*
15* 3* 12* 13* 10* 6*
8*
7 14 11 9 5 4
2 16 10 13 6 15 12 3
8
8.5 8.0 7.5 7.0 6.5
C4H*
A2H
C4H
T3H
G1H

Figure 2. The NOESY spectrum of the 17mer DNA oligomer named UAS_G in H_2O at 288 K. The water signal reduction was performed by the 1-1 excitation pulse. The sequence is given in the upper of the figure. The numbers in the chart indicate the residue numbers in the sequence. T3H, G1H, C4H and A2H mean the spectral regions of the thymine imino, guanine imino, cytosine amino and adenine 2H proton resonances, respectively. * shows the hydrogen bonded amino proton resonances.

Chapter 2

HAL Experiment, A Novel Filtering Method and Its Application to DNA Assignment

ABSTRACT

The solution structures of small DNA oligomers have been studied by using conventional ^{1}HNMR. A little larger DNA molecules (M.W. > 10000) were hardly studied due to difficulties such as signal overlapping. Here we propose a novel filtering technique to resolve the overlapping and simplify the DNA spectrum. This technique consists of the **Ha**hn echo and spin locking, and was named HAL. The HAL experiment selects singlet peaks (A2H, A8H, G8H and T5H) from mixed doublet ones (C6H), and is applicable to the NOESY detection pulse. This NOESY-HAL experiment gives the spectrum which has clearly separated cross peaks between all proton peaks (ω1) and the filtered peaks (ω2). The proposed techniques, HAL and NOESY-HAL, were applied to the signal assignment of a DNA oligomer d(CGCGAATTCGCG)$_2$.

INTRODUCTION

The structures of small DNA oligomers have been well characterized by ^1H NMR (Hosur et al., 1988; van de Ven and Hilbers, 1988; Wuthrich, 1986). The homonuclear 2D spectra of short oligonucleotides exhibit separated cross peaks, and accurate and efficient assignments can be easily made by conventional procedures (Hosur et al., 1988; van de Ven and Hilbers, 1988; Wuthrich, 1986). However, the demand for the study of the structures of a little larger DNA molecules cannot be satisfied solely through sequential assignments by means of NOESY, COSY and TOCSY spectra. Recent studies of a DNA triplex clearly illustrated the difficulties (Sklenar and Feigon, 1988; Sklenar and Feigon, 1990). However, Sklenar and Feigon overcame them partly by introducing the HOENOE experiment (Sklenar and Feigon, 1988). This excellent experimental technique allows simplification of the proton NMR spectrum of the base proton region (7.0 ~ 8.2 ppm) and thus results in easy spectrum analysis. In this chapter, the author propose another new experiment, HAL (**Ha**hn echo and spin **l**ocking), for simplifying the complicated spectrum of the DNA base proton region. This experiment can pick up singlet peaks in the base proton region, while HOENOE picks up only doublets, and thus HAL has some advantage over HOENOE.

PULSE SCHEME

The HAL experiment is schematically shown in Fig. 1A. It consists of two parts, i.e., Hahn echo (Hahn, 1950; Carr and Purcell, 1954) and spin locking "pulse building" blocks. In the former, which is a well known experimental block, the chemical shift term is refocused, but the scalar coupling term is evolved. In the latter, the unwanted magnetization due to the coupling term is dephased through B1 inhomogeneity. In Fig. 1B, a special case of a HAL experiment is shown, where the spin locking block is replaced by a β (= $\pi/4 + n\pi/2$) pulse. This experimental block can transfer the unwanted magnetization to an unobservable one. A two-dimensional experiment can also be constructed from this HAL one in combination with NOESY, as shown in Fig. 1C. This 2D experiment is useful for confirming signal assignments.

The behavior of the magnetization in a HAL experiment can be explained by means of product operator formalism (Sorensen et al., 1983). The features of HAL experiments are most easily characterized by considering the case where both isolated- and two-spin systems coexist. In an isolated-spin system the magnetization is not evolved during the $\Delta-\pi-\Delta$ period, but in a two-spin system, correlated as to J, it is evolved, as follows.

$$I_{1y} \xrightarrow{\pi J_{12}(2\Delta)2I_{1z}I_{2z}} I_{1y} \cos \pi J_{12}(2\Delta) - 2I_{1x}I_{2z} \sin \pi J_{12}(2\Delta), \qquad [1]$$

where I_1 and I_2 are the scalar coupled nuclei, and J_{12} is their coupling constant. As the delay, Δ, is adjusted to $(4J_{12})^{-1}$, the first term on the right side disappears after 2Δ, only the second term remaining. For this reason, the magnetization originating in the two-spin system is

written as a single product operator form, $-2I_{1x}I_{2z}$. Spin locking is applied at the next stage. The spin locking pulse dephases the magnetization which is evolved through a spin locking Hamiltonian, $\mathbf{H}^{(SL)}$. The evolution in the course of the spin locking process can be rationalized in terms of so-called *collective* spin modes (Ernst et al., 1987; Braunschweiler and Ernst, 1983). The modes originating in the isolated-spin system correspond to the sum of and difference of single-spin operators, $(1/2)(I_{1y} + I_{2y})$ and $(1/2)(I_{1y} - I_{2y})$, respectively, and in the two-spin system to the sum and difference of two-spin operators, $(I_{1z}I_{2x} + I_{1x}I_{2z})$ and $(I_{1z}I_{2x} - I_{1x}I_{2z})$. However, the difference terms of both the single- and two-spin operators are not formed by the Hahn echo block of a HAL experiment, because all of the chemical shift terms are refocused, as described above. The commutation relations between the sum terms and the Hamiltonian, $\mathbf{H}^{(SL)}$, are

$$[\,\mathbf{H}^{(SL)}_{y}\,,\,(1/2)(I_{1y} + I_{2y})\,] = 0, \qquad [2]$$

$$[\,\mathbf{H}^{(SL)}_{y}\,,\,(I_{1z}I_{2x} + I_{1x}I_{2z})\,] = i\,(I_{1z}I_{2z} - I_{1x}I_{2x}), \qquad [3]$$

and $$\mathbf{H}^{(SL)}_{y} = \beta I_{1y} + \beta I_{2y} = 2\beta\,(1/2)(I_{1y} + I_{2y}), \qquad [4]$$

where β is the spin locking pulse angle. The phase of the spin locking pulse was set to that of the magnetization of the isolated-spin system. Eq.[3] leads immediately to the following time evolution,

$$(I_{1z}I_{2x} + I_{1x}I_{2z}) \xrightarrow{\beta I_{1y} + \beta I_{2y}} (I_{1z}I_{2x} + I_{1x}I_{2z})\cos 2\beta + (I_{1z}I_{2z} - I_{1x}I_{2x})\sin 2\beta, \qquad [5]$$

while the sum term of single-spin operators remains invariant, as indicated in eq.[2]. Therefore, in the isolated-spin system the magnetization is not dephased by a spin locking pulse, but in the two-spin system it is. This is the principle of the signal selection used in the HAL experiment shown in Fig. 1A. In the special case of the HAL experiment shown in Fig. 1B, the spin locking pulse angle β is set to $(\pi/4 + n\pi/2)$ in eq.[5], the cosine term vanishes and the sine term remains. Therefore, the magnetization is unobservable in the two-spin system.

These phenomena are notable because the spin locking Hamiltonian used here has clearly an opposite property to the isotropic mixing Hamiltonian used for the spin locking in TOCSY (Ernst et al., 1987; Braunschweiler and Ernst, 1983). Then, the replacement of the spin locking pulse with an isotropic mixing pulse causes HAL to lose the property of dephasing of the magnetization in the two-spin system. So the spin locking pulse in Fig. 1A should be a purging one, i.e., a long pulse with a single phase.

RESULTS

The HAL experiment has been applied to the spectral analysis of a standard DNA duplex whose sequence is d(CGCGAATTCGCG)2. For the DNA oligomer, adenine H2 and H8, thymine H6 and C5H3, and guanine H8 belong to the isolated spin system, and the

cytosine H5 and H6 protons are classified into the two-spin system. In Fig. 2A, an ordinal base proton spectrum of the DNA oligomer is shown. All the resonances had already been assigned (Hare et al., 1983) and the cytosine H6 signals are indicated by arrows in Fig. 2A. The other peaks arise from the AH2, AH8, TH6 and GH8 protons. The spectrum in Fig. 2B was recorded using the HAL pulse sequence in Fig. 1A under the same conditions as in the Fig. 2A experiment. The scalar coupling constant, J, between cytosine H5 and H6, is approximately equal to 7.5 Hz, and the delay, Δ, was set to 33 msec. The spin locking pulse angle was 5000° (~ 500 μsec) with the 160 kHz B1 field, and no attenuation was applied for this pulse. On comparison of the spectrum in Fig. 2A with that in Fig. 2B, it can be clearly seen that the cytosine H6 signals disappear and the other signals remain on application of the HAL pulse sequence.

The intensity of the doublet signal of cytosine H6 changes, depending on both the delay time, Δ, and the spin locking pulse angle. When the delay time cannot be adjusted to $(4J_{12})^{-1}$, the cosine term of eq.[1] is formed and not dephased by the spin locking pulse. This situation is found in other regions of the spectrum of DNA such as the sugar proton region, because the adjusted delay time differs from the $(4J_{ij})^{-1}$ values of sugar proton multiplets.

On the other hand, eq.[3] explains how the spin locking pulse angle, β, affects the intensity of the cytosine H6 resonance. When β is set at $(\pi/4 + n\pi/2)$, the magnetization transferred to the multiquantum term is not observed in the detection period. This is the case of the special version of the HAL experiment shown in Fig. 1B. The spin locking pulse angle dependency of the signal intensity is given in Fig. 3A. As the number, n, increases, the doublet signals begin to appear due to pulse angle imperfection. The angle, β, should be adjusted exactly, because twice β affects the doublet signal intensity.

For easy abolition of doublet signals, therefore, utilization of the dephasing property of the spin locking is more practical. This is the case of the pure HAL experiment in Fig. 1A. The spin locking time dependency of the signal intensity is shown in Fig. 3B. As can be seen, a longer pulse causes signal dephasing because of the B1 inhomogeneity. For complete dephasing, the spin locking pulse angle should be much greater than 2500°, which is roughly equal to 250 μsec with the 160 kHz B1 field.

The effect of relaxation during the echo delay, Δ, should be taken into account. As the DNA oligomer gets larger, this effect becomes more important, viz. all or partial signal relaxation during the delay, Δ, makes a HAL experiment infeasible. However, the sample used here is a DNA dodecamer (M.W.~8000) and gives clearly separated signals. At least we can say that the HAL experiment is useful for the confirmation of the signal assignments of a DNA dodecamer.

Two-dimensional experiments were performed using the pulse sequences of conventional NOESY and NOESY-HAL in Fig. 1C, and the results were compared. Fig. 4A shows the conventional NOESY spectrum and Fig. 4B the spectrum obtained in the NOESY-HAL experiment. Both spectra are for the region of the sugar 2' 2" proton (1.8 ~ 3.0 ppm) in $\omega 1$ and of the base proton (7.0 ~ 8.2 ppm) in $\omega 2$. The arrows indicate the positions of cytosine H6 signals. Clearly selected singlet peaks are seen in Fig. 4B with a high S/N ratio. The residual weak peaks for the H6 positions are due to imperfect signal cancelling. The

pulse sequence in Fig. 1C was constructed by replacement of the NOESY detection pulse with the HAL pulse sequence. The replacement of the NOESY preparation pulse with the HAL one gives a spectrum with a little worse S/N ratio (data not shown).

DISCUSSION

The concept of the HAL experiment is a counterpart of the HOENOE one. In other words, the HAL experiment picks up singlet peaks in the base proton region, and only the doublet cytosine H6 signals vanish, while the HOENOE experiment selects only the doublet cytosine H6 signals, the other signals vanishing. Recently, Piotto and Gorenstein reported that the homonuclear 3D NOESY-TOCSY experiment was useful for the spectral analysis of a GG mismatched DNA oligomer (Piotto and Gorenstein, 1991). The information provided by their 3D experiment might be the best among all the published homonuclear spectra. However, analysis through homonuclear 3D experiments is not easy in general and takes a lot of time. Moreover, the point resolution is too low to recognize each peak in the crowded region. The HAL and HOENOE experiments have advantages over the 3D NOESY-TOCSY one in the simplification and resolution of complicated spectra.

The HAL experiment requires the adjustment of some parameters, that is, the 90° pulse width, delay, Δ ($=1/4J_{12}$), and spin locking pulse angle. In the simplest case where we know the value, J_{12}, and the spin locking pulse angle is set at 5000° (~ 500 μsec) with the 160 kHz B1 field, we only adjust the 90° pulse width. Significantly, the sensitivity of the HAL experiment is better than that of the HOENOE one (data not shown). These advantages, the easy parameter adjustment and the good sensitivity, should be taken into account when considering an application.

In conclusion, by using the HAL experiment we can easily pick up singlet peaks mixed with doublet peaks. This is applicable to the NOESY detection pulse and useful for the confirmation of signal assignments. In this note we demonstrated that the experiment is applicable to the spectral analysis of a DNA oligomer and selects singlet peaks in the base proton region. In addition, other applications are conceivable. For example, the titration of the DNA signals on the addition of drugs and proteins may be easily performed. Also, this type of experiment is applicable for the peak separation of RNA oligomers and organic compounds which have isolated-spin systems mixed with two-spin systems.

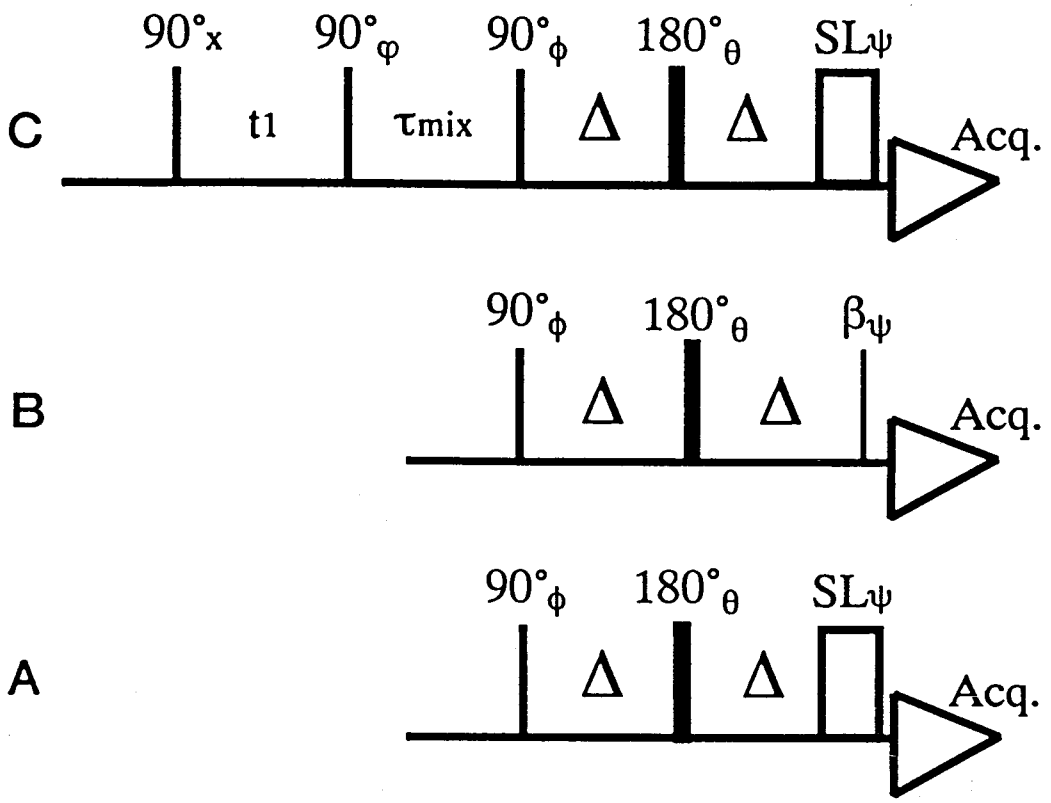


FIG. 1. (A) The HAL pulse sequence. The delay time, Δ, is set to $1/(4J_{12})$. SL is the spin-locking pulse in full power, 500 μs being typically used. (B) The pulse sequence of a special case of a HAL experiment, where β is the flip angle of the last pulse and set to $(\pi/4 + n\pi/2)$; n = 0 is preferred. The delay time, Δ, is the same as that in the experiment in A. (C) The pulse sequence of a two-dimensional version of the HAL experiment. This gives a NOESY spectrum edited in the ω_2 dimension by the detection of singlet signals. The parameters, the delay time, Δ, and the spin-locking time are the same as those in the experiment in A. The phase cycling used for A and B was $\phi = x; \theta = 2(x),2(-x),2(-y),2(y); \psi = y,-y$; Acq. $= 4(y),4(-y)$, and for C, $\varphi = 8(x),8(-x); \phi = x; \theta = 2(x),2(-x),2(-y),2(y); \psi = y,-y$; Acq. $= 4(y),8(-y),4(y)$.

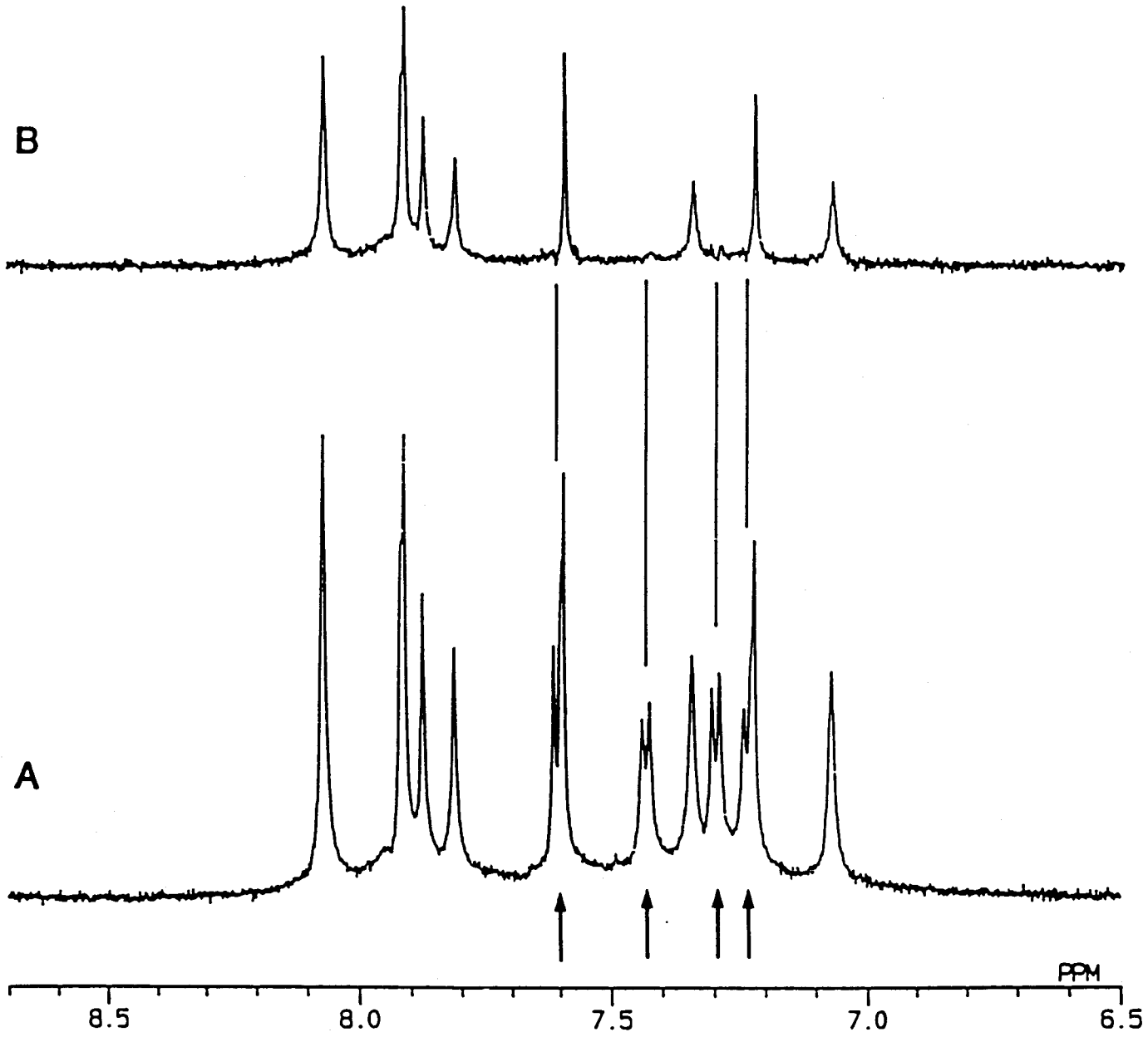


FIG. 2. (A) Conventional single-pulse one-dimensional spectrum. The spectral region corresponds to the DNA base-proton region and contains the following signals: AH2, AH8, GH8, TH6, and CH6. The arrows indicate the resonance positions of CH6. (B) HAL experiment spectrum. The parameters used to obtain this spectrum are described in the text. These two spectra were recorded with a Jeol GSX500 NMR spectrometer under the same sample conditions; i.e., a DNA sample was dissolved in 400 μl D_2O, to 1 m*M*, pH 7, in 20 m*M* phosphate buffer and 50 m*M* NaCl.

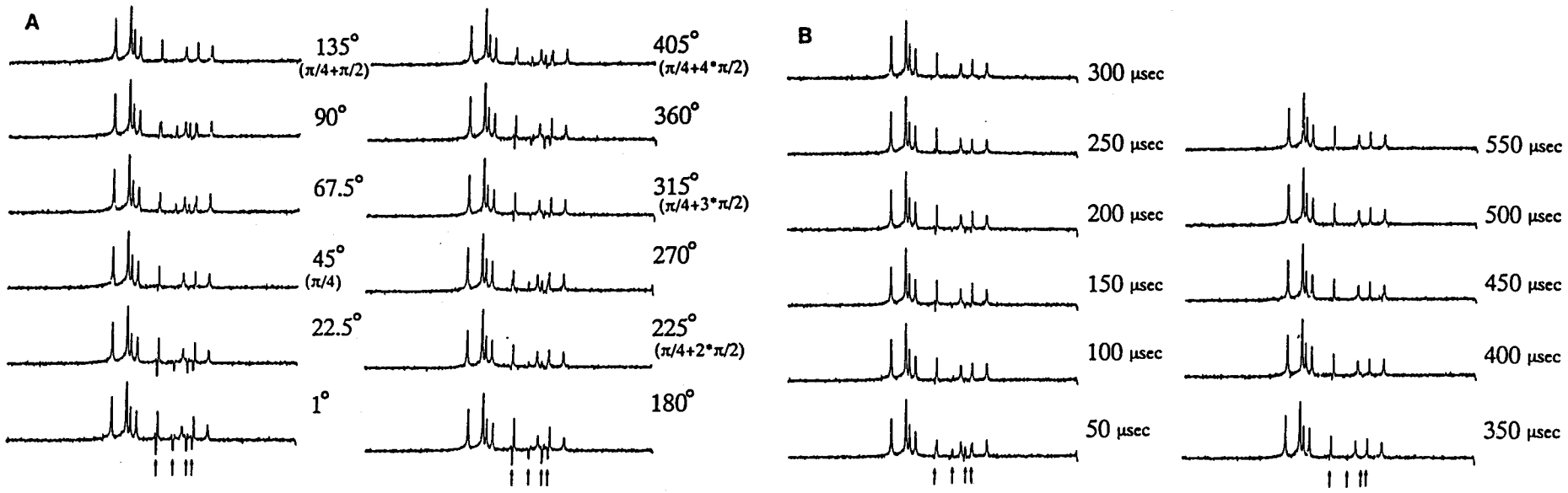


FIG. 3. (A) The spin-locking-pulse-angle dependence of the doublet signals. The numbers beside the spectra are the flip angles of the spin-locking pulse given in degrees. The angles in parentheses are given in radians. (B) The spin-locking time dependence of the doublet signals. In both A and B, the arrows indicate the resonance positions of cytosine H6, described as doublets, and the spectral region is that of the base proton. All spectra were recorded under the same sample conditions as those given in the legend to Fig. 2, and the delay time, Δ, was also the same for both series of experiments. The strength of the spin-locking pulse field was set to that of the hard-pulse field, the 160 kHz B_1 field.

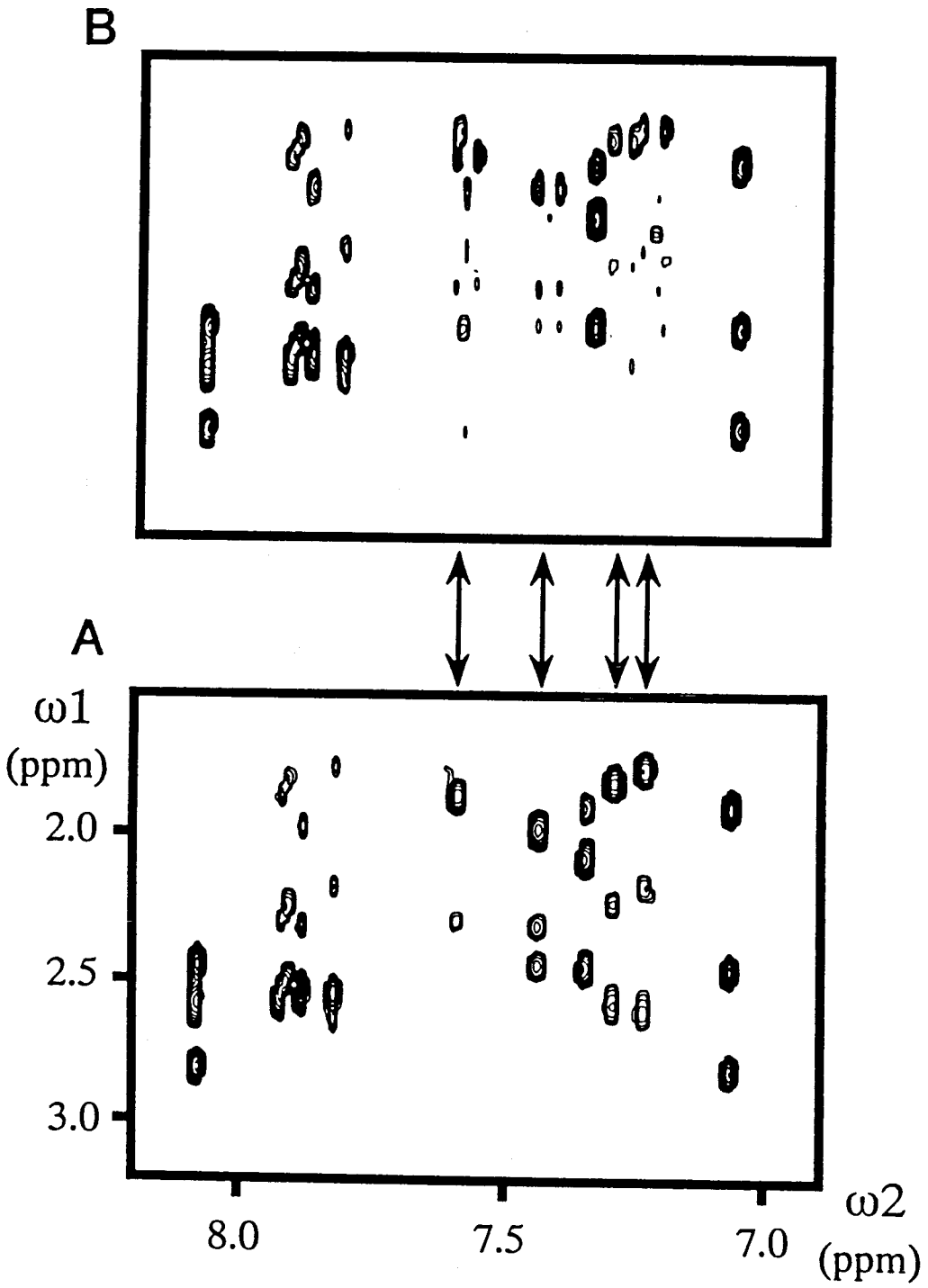


FIG. 4. (A) Conventional NOESY spectrum and (B) NOESY-HAL spectrum derived with the pulse sequence given in Fig. 1C. The arrows indicate the resonance positions of cytosine H6 doublets; ω_1 (1.8 ~ 3.0 ppm) is the frequency region of the sugar 2′, 2″ protons and ω_2 (7.0 ~ 8.2 ppm) is that of base protons. The sample conditions are the same as those given in the legend to Fig. 2 except for the solution concentration and volume, 3 mM in 500 μl.

Chapter 3

Filtering Methods for Selection of Singlet and Doublet Signals in NMR Spectra of DNA Oligomers

ABSTRACT

The base proton (purine 6H and pyrimidine 8H) resonances are key signals for the assignment of the proton resonances of DNA oligomers. They are classified into two groups, i.e., cytosine 6H signals observed as *doublets* and the other base proton signals observed as *singlets*. Here we propose some experiments for distinguishing the cytosine 6H signals from the others. Moreover, the ability of signal selection and the sensitivity as to signal detection were compared for each experiment, and the optimum conditions for spectral measurements were surveyed. Some of the experiments were employed as the NOESY detection pulse. Previously proposed experiments, such as HOENOE and HAL, were also compared with them.

INTRODUCTION

Proton resonance assignments for DNA oligomers are mainly performed by using NOE connectivity (Wuthrich, 1986; Hosur et al., 1988; Van de Ven and Hilbers, 1988; Feigon et al., 1992). The through-bond connectivity obtained by COSY or TOCSY, however, can only be used for limited purposes, such as simple identification of the sugar protons and the base protons. It should be noted that no sequential connectivity to neighboring nucleotide residues can be obtained from this through-bond connectivity, and thus ambiguity remains as to the sequential assignments for nucleic acid oligomers. Such aspects of assignments were previously reviewed and explained in detail (Wuthrich, 1986; Hosur et al., 1988; Van de Ven and Hilbers, 1988; Feigon et al., 1992).

Recently, Sklenar and Feigon (1990) introduced a novel experiment, HOENOE, which reduced the ambiguity of the assignments by picking cytosine 6H doublets up from the base proton region (7~8.5 ppm) in NOESY. Though several experiments for the detection of multiplets were proposed previously, none of them, except HOENOE, was designed for the detection of doublet signals of nucleic acids. For example, experiments involving the use of double quantum transitions, i.e., DQF-COSY, 2Q spectroscopy and DQ-NOESY (Van de Ven et al., 1985), were applied for nucleic acids, but were mainly employed for the assignment of the complex sugar proton resonances, not for the selection of *doublets* in the base proton region (7~8.5 ppm). Similar selected and simplified information to that in the case of the HOENOE experiment can be obtained by means of other types of experiments. For example, the HAL experiment proposed by us picks up *singlets* instead of *doublets* (Kojima and Kyogoku, 1993). These two experiments allow simplification of the base proton region in the NMR spectra of DNA oligomers and thus result in easy spectrum analyses.

Here we compared several one-dimensional filtering experiments, including newly proposed ones. The procedures for the selection are classified into three categories, *doublet* selection, *singlet* selection, and detection of both but with opposite signs. For simplicity, we call cytosine 6H signals *doublets* and the other base proton (thymine 6H, guanine 8H and adenine 8H and 2H) signals *singlets*. Furthermore, we extend some of the new techniques to NOESY type two-dimensional experiments, and compare the abilities of *singlet* and *doublet* selection in practice.

THEORETICAL BACKGROUND

Classification of experiments

In Fig. 1, the pulse trains of seven experiments are given. SGP (A) is the conventional single pulse experiment given as a reference for the other six experiments (B~G). Two experiments (B and C) were designed for the selection of *singlets*. (B) is the HAL experiment proposed previously by us (Kojima and Kyogoku, 1993), HAL being an abbreviation of HAhn-echo and spin Locking. SL in the figure indicates a spin locking pulse, but a purging

pulse is actually used. (C) is the SQF experiment now newly proposed. SQF comprises single quantum filtering achieved by phase cycling. Both experiments (B) and (C) have the same delay time, Δ, which is set to 1/4J for the separation of *singlets* and *doublets*. A more detailed explanation was given in the previous chapter.

Three experiments, (D) ~ (F), were designed for the detection of *doublets*. (D) is a DQF experiment regarded as one-dimensional DQF-COSY (Piantini et al., 1982; Rance et al., 1983; Shaka and Freeman, 1983; Muller et al., 1986). (E) is the HOE experiment, that is a part of the HOENOE experiment (Sklenar and Feigon, 1990). (F) is the HARD experiment newly proposed here and regarded as a one-dimensional 2Q experiment (Bax et al., 1981; Sorensen et al., 1983). HARD is an abbreviation of **HA**hn-echo **R**efocused **D**QF. These three experiments will be discussed in more detail later.

(G) is a HAHN experiment (Hahn, 1950). This has a delay time, Δ, which is set to 1/4J. *Singlets* and *doublets* are observed with opposite signs, positive and negative.

Optimized parameters for the doublet selection experiments

For singlet detection experiments, there is no parameter which should be adjusted to get the best efficiency. This is also true for the HAHN experiment. On the other hand, the experiments designed for *doublet* detection should be optimized by searching for the best efficiency as to detection.

A scheme of HOE is shown in Fig. 1E. In this experiment, cytosine 5H resonances (5.3~6 ppm) are selectively exited and followed by means of the isotropic mixing pulse. The total signal transfer efficiency, F, of the HOE experiment relative to that of SGP was calculated as the product of two factors. As pointed out previously (Braunschweiler and Ernst, 1983), one is the conversion factor under the isotropic mixing conditions and the other is the relaxation factor. Thus, the total efficiency, F, is described by the following equation,

$$F = \sum_i \left[\frac{1}{2}(1 - \cos(2\pi Ji \cdot \delta 1)) \cdot \exp(-\frac{\delta 1}{T1\rho i}) \right]$$
$$= \sum_i \left[\sin^2(\pi Ji \cdot \delta 1) \cdot \exp(-\frac{\delta 1}{T1\rho i}) \right] \qquad [1]$$

where d1 is the mixing time, J the coupling constant of a *doublet*, and T1ρ the relaxation time in the rotating frame. The subscript, "i", denotes the specified cytosine 6H resonance. In the case of an aqueous solution of a DNA oligomer, the value of T1ρ is nearly equal to that of T2 (Freeman, 1988), and all of the cytosine 6H resonances have almost constant T2 and ^3J values, which are given as T2av and Jav. Then, the total signal transfer efficiency is approximated with the following equation,

$$F(\delta 1, T2av) \cong \sin^2(\pi Jav \cdot \delta 1) \cdot \exp(-\frac{\delta 1}{T2av}). \qquad [2]$$

Eq. 2 means that the total signal transfer efficiency is a simple function of two time parameters, i.e., the mixing time, δ1, and T2av. As the averaged coupling constant, Jav, is fixed at 7.5Hz

for DNA oligomer duplexes, the total signal transfer efficiency was calculated to be as plotted in Fig. 2A. The ridge line in Fig. 2A indicates the condition that gives the maximum efficiency with a certain T2av value. Accordingly, the delay time, d1, is uniquely determined by the T2av value of an individual sample, viz., the δ1 value is the function of the T2av value under the efficiency maximized conditions. The averaged transverse relaxation time, T2av, is about 120 msec for a DNA dodecamer (Lane et al., 1991), so the d1 value was determined to be 59 msec from Fig. 2A.

For the DQF experiment in Fig. 1D, the delay time, δ0, should be shorter. The total signal transfer efficiency, F, depends on the delay time, δ0, and individual transverse relaxation times, T2i. In practice, the T2i values are replaced by the averaged T2av one, as mentioned above. These parameters are related through the following equation,

$$F(\delta 0, T2av) \cong \sin(\pi Jav \cdot 2\delta 0) \cdot \exp(-\frac{2\delta 0}{T2av}). \qquad [3]$$

The HARD experiment, shown in Fig. 1F, also exhibits similar relations among the following three parameters; the delay time, δ2, the averaged transverse relaxation time, T2av, and the total signal transfer efficiency, F.

$$F(\delta 2, T2av) \cong \sin^2(\pi Jav \cdot 2\delta 2) \cdot \exp(-\frac{4\delta 2}{T2av}). \qquad [4]$$

The optimum conditions for both the DQF and HARD experiments were searched for in similar ways to as for HOE. Three-dimensional graphical expressions for the total efficiency are displayed in Fig. 2B for DQF and in Fig. 2C for HARD, using 7.5 Hz as the Jav value in the calculations with Eqs. 3 and 4.

The optimized delay time, total delay time, and total efficiency were calculated by using Jav (7.5 Hz) and T2av (120 msec) for three *doublet* selection experiments, HOE, DQF and HARD, and are given in Table 1. The total delay time means the sum of the optimized delay times in each experiment (Fig. 1), and total efficiency, F, means the relative intensity compared to that of SGP.

Two-dimensional experiments

The experiments shown in Fig. 1 can be extended to two-dimensional ones by combining them with the NOESY experiment. As mentioned above, the HOE experiment has already been combined with NOESY as HOENOE (Sklenar and Feigon, 1990). Similarly, the NOESY detection pulse is replaced by HAL, SQF or DQF. These two-dimensional experiments are named NOESY-HAL, NOESY-SQF and NOESY-DQF, respectively. The HARD and HAHN experiments may be combined with NOESY, but these experiments require longer delay times and thus may result in lower S/N ratios. Actually, the calculated efficiency of the HARD experiment is worse than those of the other *doublet* selection experiments, DQF and HOE, as shown in Table 1.

The replacement of the detection pulse in NOESY with HAL, SQF or DQF gives the

ω2-filtered NOESY spectrum. The ω1-filtered spectra are also obtained by replacing the preparation pulse in NOESY with them, just like in the case of HOENOE. Since all the experiments in Fig. 1, except for HOE and HARD, do not give in-phase absorption spectra in the spectral region of the sugar protons, NOESY type experiments combined with them do not give cross peaks with the sugar proton signals on ω1 filtering. Moreover, the point resolution of the ω2 axis is higher than that of the ω1 one, and thus ω2-filtering is preferable for the analysis of the base proton resonances.

EXPERIMENTS

The experimental conditions, except for the pulse conditions, were the same for all the experiments. The NMR measurements were performed with a JEOL GSX500 spectrometer operating at 500 MHz and the temperature was set at 303K.

The material used for these experiments was a double-stranded DNA oligomer, $d(CGCGAATTCGCG)_2$. The solution comprised 3 mM duplex in 500 ml buffer, 20 mM sodium phosphate and 50 mM sodium chloride, at neutral pH.

For the HAL, SQF, NOESY-HAL and NOESY-SQF experiments, the delay time was set to 1/4J, 34.5 msec. The DQF, HOE, HARD and NOESY-DQF experiments were performed using the delay times listed in Table 1. The 90° pulse width was 10 msec for high power excitation, 23.5 msec for MLEV17 mixing, and 1 msec for selective excitation. The spin lock pulse width was 500 msec at the high power level in the HAL experiment. For all spectra, 128 hyper complex points for t1 and 1024 complex points for t2 were recorded, and 128 scans for each t1 increment were used.

RESULTS AND DISCUSSION

One-dimensional experiments

The spectra in Fig. 3 were respectively obtained by performing the experiments given in Fig. 1. All the signals in these spectra were assigned previously (Hare et al., 1983; Lane et al., 1991). The relative signal intensities of the spectra in Fig. 3, compared to SGP, are listed in Table 2.

Selectivity: The ability of signal selection in each experiment can be checked by referring to Fig. 3. As to *singlet* detection experiments, HAL and SQF showed slightly incomplete selection of the signals. Both spectra contained remaining small doublet peaks. These experiments were performed with a 500 ml sample volume. When the sample volume was reduced to 400 ml, these small peaks disappeared. Thus, it can be said that they reflect B1 inhomogeneity. Consequently, the signal selection should be influenced by the imperfection of the 90° and 180° pulses. To overcome this problem, the optimum sample volume should be searched for. Another solution of this problem is the use of the gradient enhancement

technique (Hurd et al., 1992) or the composite pulse technique (Ernst et al., 1987; Freeman, 1988), or simply a micro cell.

In the cases of the *doublet* detection experiments, the four cytosine 6H signals at 7.25, 7.32, 7.46 and 7.62 ppm were clearly selected in all experiments. The two weak signals at 7.10 ppm and 7.36 ppm represent the *singlets* from thymine 6H. The HOE experiment shows relatively worse selection because the ROE (rotating frame NOE) part in HOE is difficult to adjust. In particular, the peak at 8.09 ppm exhibits negative intensity.

Sensitivity: The calculated sensitivities are given in Table 1 as efficiency, and the experimental ones are listed in Table 2. It should be noted that the optimized parameters in Table 1, the theoretically expected values, are close to those experimentally obtained from the one-dimensional NMR spectra in Fig. 4.

As judged on comparison among the *doublet* selection experiments, the DQF experiment should exhibit the best sensitivity. The theoretically expected sensitivity of the HOE experiment was better than that of the HARD experiment, but the observed sensitivity was the worst. This fact may be explained in two ways. One is the lack of some factors in the theoretical calculation. For example, the relaxation during the selective excitation and refocusing periods, and the efficiency of selective excitation were not taken into account. The other is the experimental difficulty, namely, the determination of the 90° pulse for the selective excitation without signal dephasing, and the complete adjustment of the refocusing delay time. Imperfection of the former condition causes lower sensitivity and the latter introduces signal decay. So, the experimentally obtained sensitivity of HOE may not be the best obtained under given conditions.

Total evaluation for practical use: For practical use of these experiments, comparison in view of convenience is important. In this sense, the HOE experiment has problems. It includes soft and semi-soft pulses, which should be adjusted properly, as previously mentioned. For the use of these experiments as one-dimensional NMR, HARD and HAHN experiments show better performance, and are useful for such cases as the titration of DNA-drug or DNA-protein systems. For the assignment of signals, the DQF, HAL and SQF experiments may be chosen because of their high sensitivity.

NOESY type 2D experiments

The NOESY-DQF spectra are presented in Fig. 4. The displayed region is the cross section between the sugar 2', 2" proton region (1.5~3.2 ppm) and the base proton region (6.9~8.2 ppm). All the peaks of these spectra were assigned previously (Hare et al., 1983; Lane et al., 1991). The arrows in the figures indicate the positions of cytosine 6H signals.

The difference between the two spectra, (A) and (B), arises from the condition of the zero-order phase value in the ω2 dimension; 0° for (A) and 90° for (B). The peak shape in Fig. 4A represents the anti-phase *doublet* in ω2. This shape is favorable for the identification of *doublet* peaks, but unfavorable for the assignment of signals in a complicated pattern. The experiment giving the spectrum in Fig. 4B was designed to overcome this disadvantage. The peak appears as a dispersive anti-phase doublet, though it looks like a single peak at low resolution (Pelczer et al., 1991). The sensitivity of the NOESY-DQF experiment is so high

that the spectrum in Fig. 4B retains a high S/N ratio. Such NOESY-DQF gives cross peaks between the unfiltered resonances in ω1 and the filtered resonances in ω2, cytosine 6H. However, NOESY-DQF has another advantage. The DQF part selects not only the cytosine 6H resonances but also the evolved multi-quantum coherence from the thymine 6H resonance coupled with the methyl protons. This coherence gives a cross peak between the intra-base protons in the thymine ring, i.e., the methyl and 6H protons. Such cross peaks are hardly observed even in COSY because of the small coupling constant between them.

Comparison of available information: NOESY-DQF and HOENOE experiments are classified as *doublet* selection ones. For the selection of the base proton resonances, there is no difference between them except for the peak shape; anti-phase for NOESY-DQF and in-phase for HOENOE. In other regions, like the sugar proton one, HOENOE also gives in-phase peaks, but NOESY-DQF does not give simple anti-phase peaks. This means that HOENOE gives the same information as conventional NOESY does in the sugar proton region. The 3D NOESY-TOCSY experiment (Piotto and Gorenstein, 1991) provides more information on peak connectivity than conventional NOESY. This 3D experiment covers the information from NOESY-DQF, but does not allow clear signal selection because of its still crowded and overlapped spectrum. Another experiment, DQ-NOESY (Van de Ven et al., 1985), is similar to NOESY-DQF. DQ-NOESY is a kind of ω1-filtered NOESY. This means that the chemical shift term in ω1 evolves under the double quantum coherence. Though this experiment was designed to separate the sugar proton resonances, the selection in the base proton region is better than that in the case of NOESY-DQF. As DQ-NOESY needs a refocusing delay like HARD in Fig. 1F, its sensitivity is lower than that of NOESY-DQF.

CONCLUSION

We compared several one-dimensional filtering experiments for *doublet* (cytosine 6H signal) and *singlet* (other base proton signals) selection. Some of them are newly proposed here and the others were previously reported. The ability of signal selection is so high for each experiment that no problem is found at this point except in the cases of signal cancellation of opposite signals. So far as sensitivity is concerned, the DQF and SQF experiments showed better performance. Thus, like the previously reported NOESY-HAL, they are applicable to the NOESY detection pulse, maintaining a high S/N ratio, and giving NOESY-DQF and NOESY-SQF. Moreover, they have so high selectivity that they can be practically used for the assignment of nucleic acid signals and the confirmation of the signal assignments in crowded signal regions.

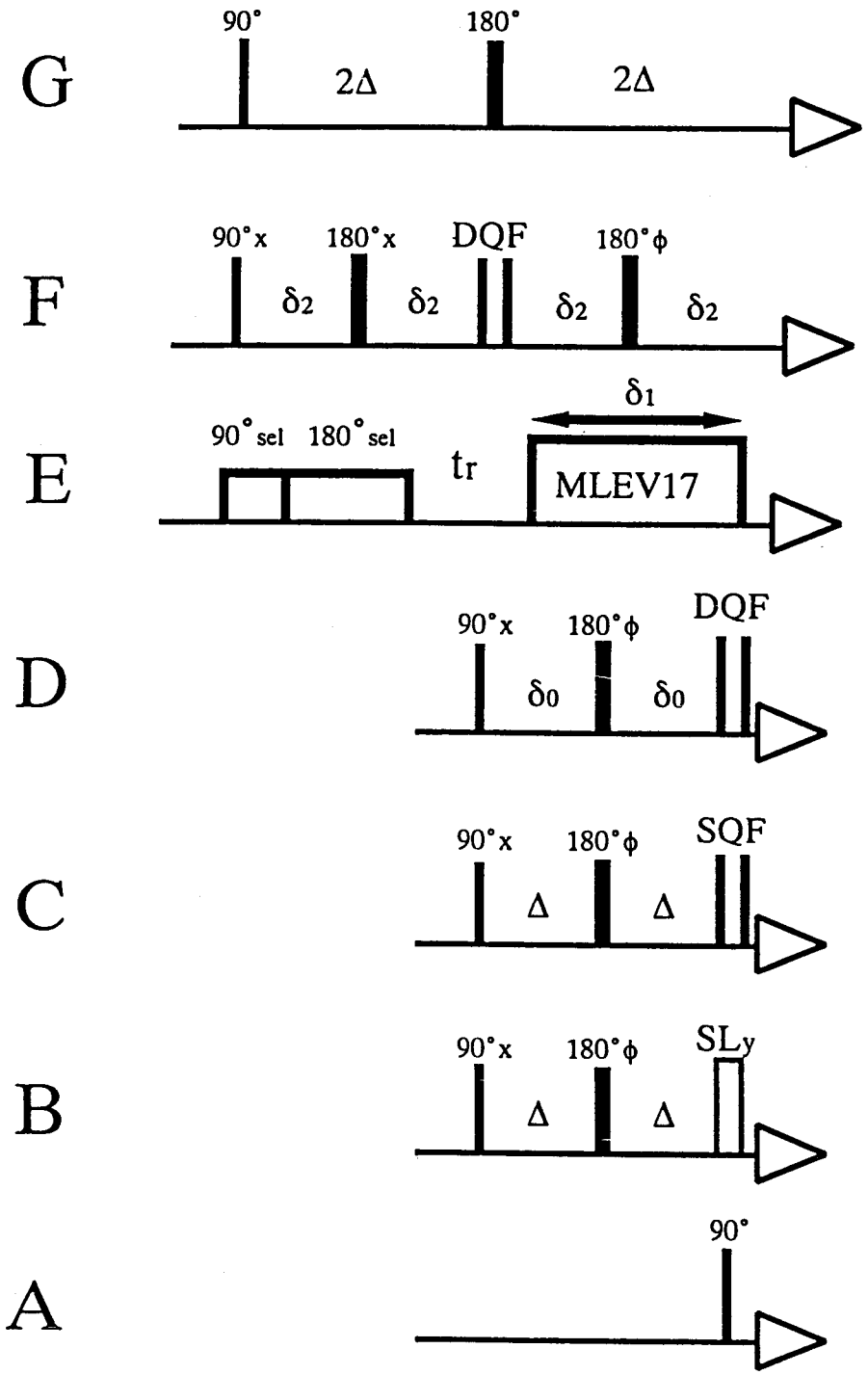


Fig. 1. (A–G) Pulse sequences used for several types of 1D experiments. The experiments are the following: (A) SGP; (B) HAL; (C) SQF; (D) DQF; (E) HOE; (F) HARD; and (G) HAHN. All the sequences were designed for specific purposes, i.e.: (A) reference; (B,C) singlet signal selection; (D–F) doublet signal selection; and (G) both singlet and doublet signal detection but with opposite signs. Except for the conventional experiment (A) and the echo experiment (G), the phase cyclings were as follows: ϕ = x, SQF (first pulse phase = 4(x),4(−x); second one = x,y,−y,−x,−x,−y,y,x), Acq. = x,y,−y,−x for (C); ϕ = x, DQF (first pulse phase = 4(x),4(−x), second one = x,y,−y,−x,−x,−y,y,x), Acq. = x,−y,y,−x for (D); DQF (first pulse phase = 4(x),4(−x), second one = x,y,−y,−x,−x,−y,y,x), ϕ = x, Acq. = x,−y,y,−x for (F). In (E) 'sel' indicates the selective excitation pulse. Phase cycles for (B) and (E) were given previously (Sklenář and Feigon, 1990; Kojima and Kyogoku, 1993).

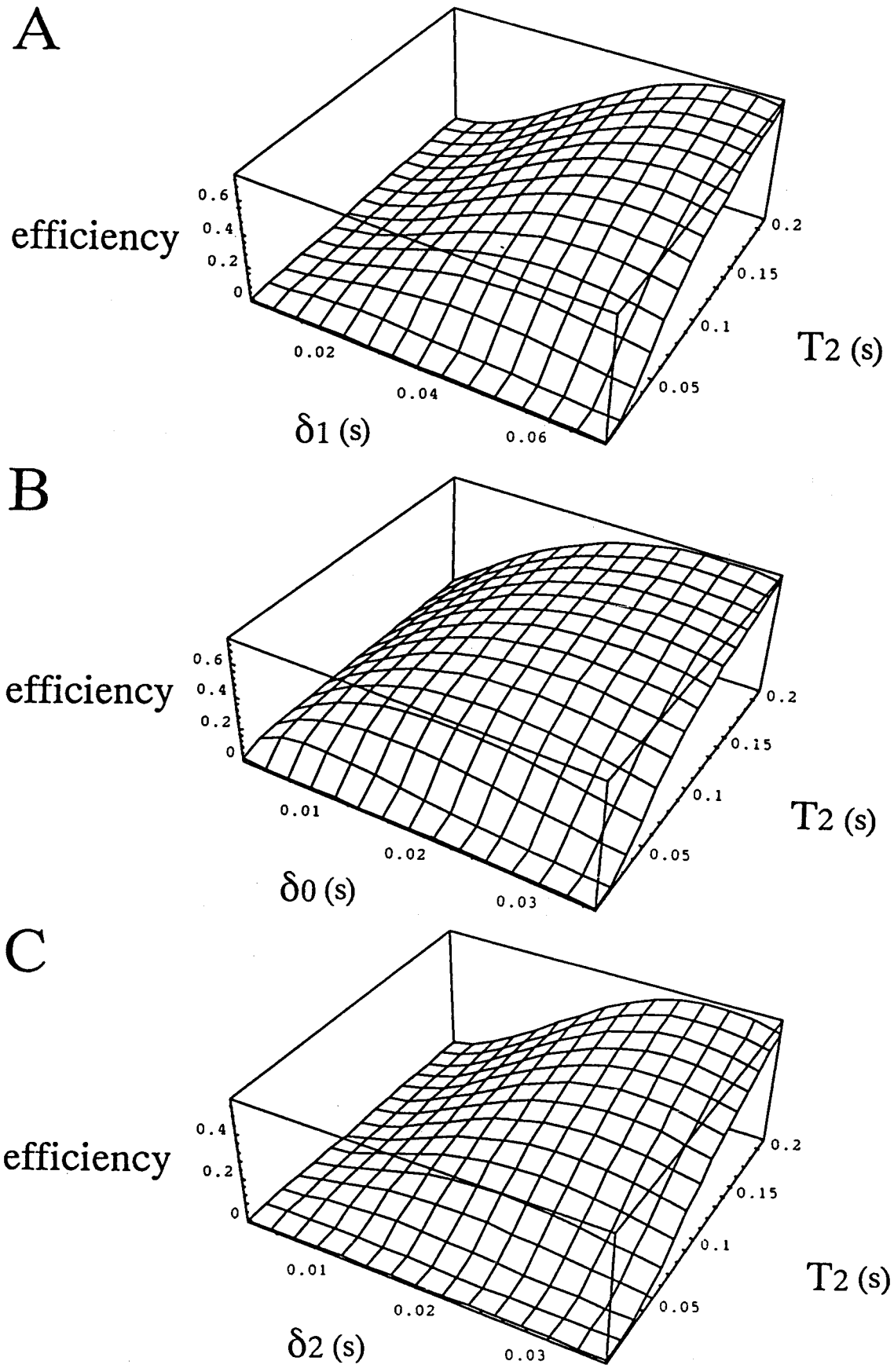


Fig. 2. (A) Three-dimensional graphic display of the calculated signal transfer efficiency for the HOE experiment as a function of the averaged T2 value, $T2_{av}$, and the mixing time, δ_1. (B) DQF and (C) HARD: the parameters are the same as those in (A), except for the delay time, δ_0 and δ_2. These graphics show the optimum measurement conditions with various T2 values.

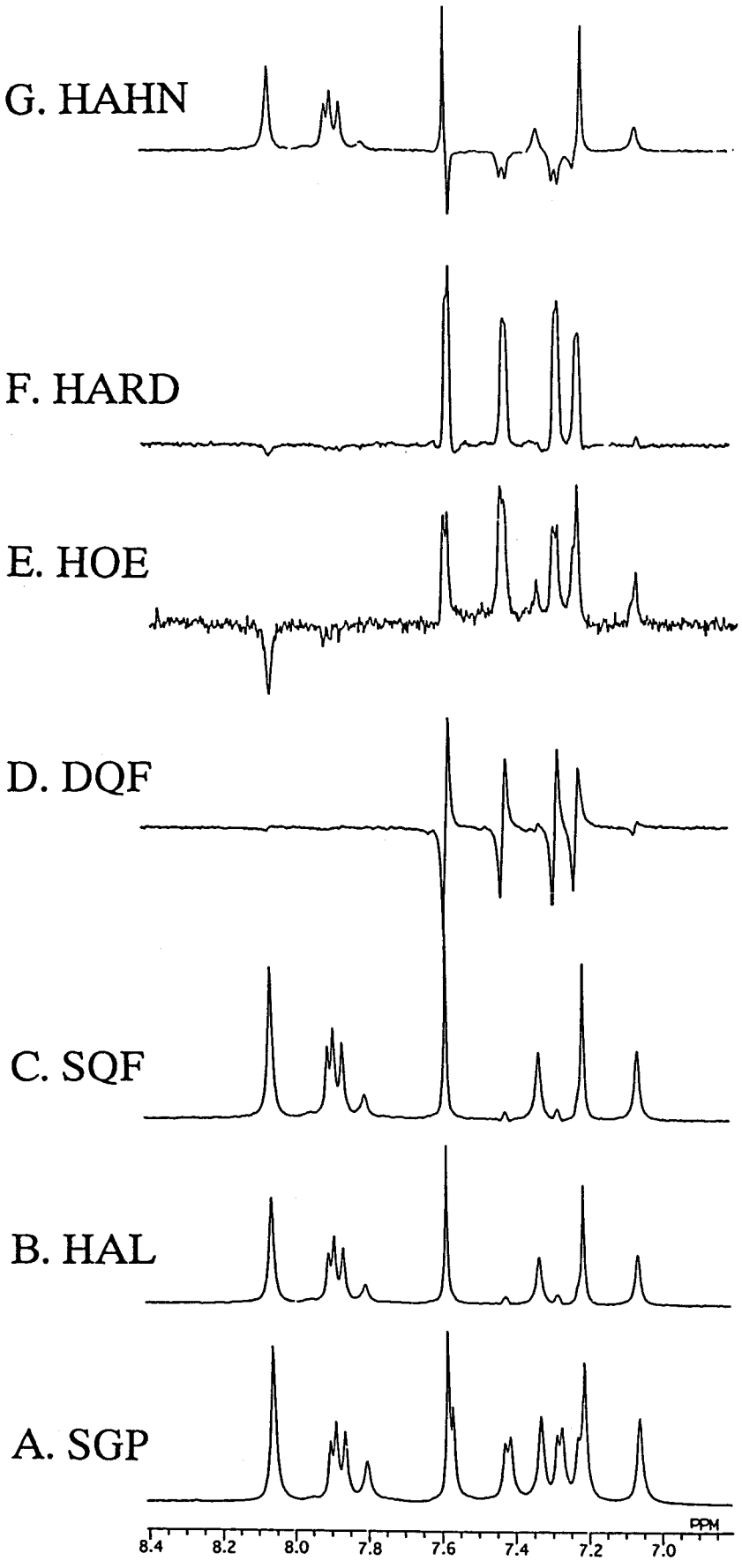


Fig. 3. (A–G) One-dimensional spectra obtained with the pulse sequences shown in Fig. 1. The experiments are identical to those in Fig. 1. In all the spectra, the base proton region is shown (6.9–8.4 ppm).

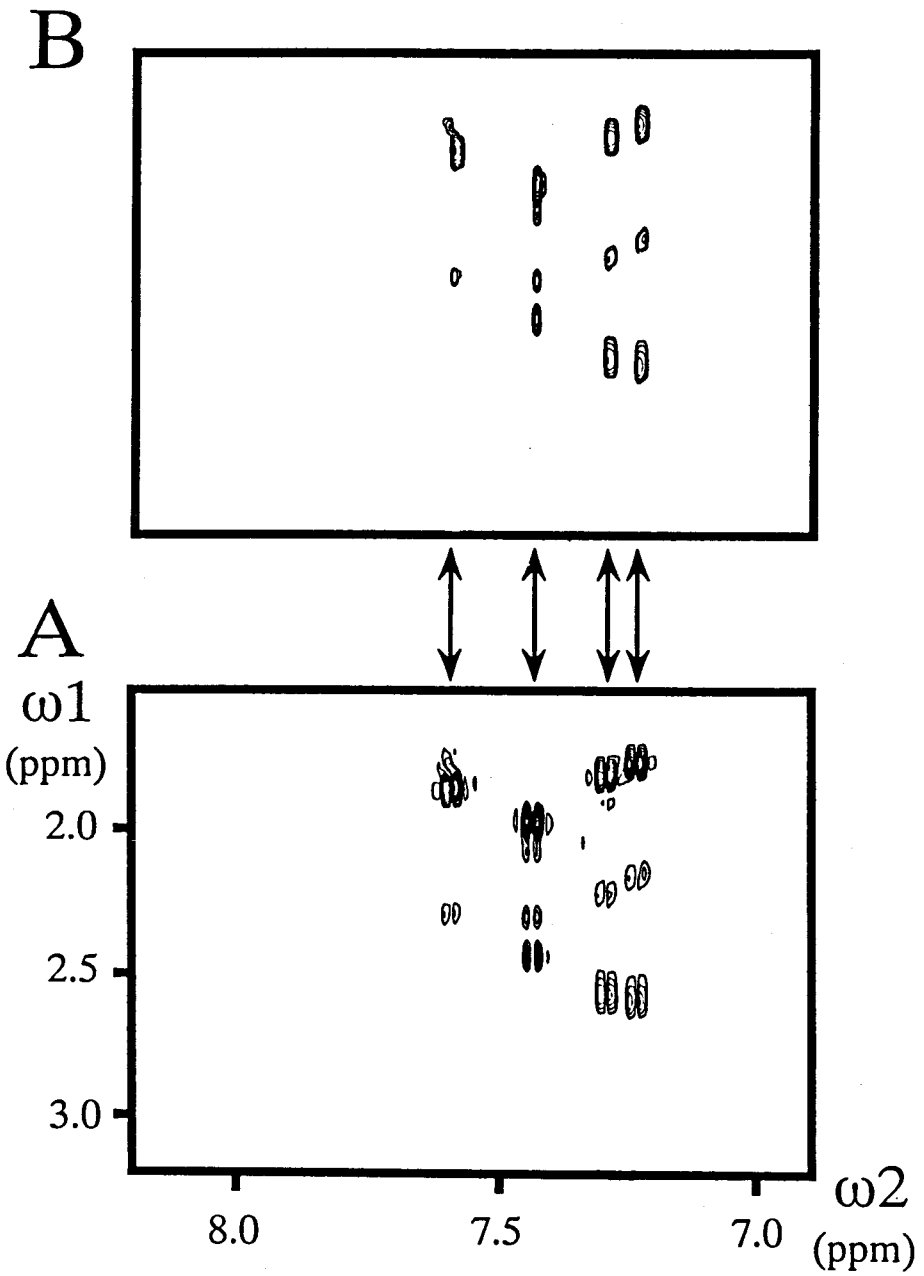


Fig. 4. NOESY-DQF spectra with a different zero-order phase value for the ω2 dimension; (A) 0°; and (B) 90°. ω1 (1.8–3.0 ppm) is the frequency region of the sugar 2',2" protons and ω2 (7.0–8.2 ppm) that of the base protons. The arrows indicate the resonance positions of cytosine H6 doublets (Hare et al., 1983; Lane et al., 1991).

Chapter 4

HSQC Spectrum of the Amino Groups of Nucleic Acids

ABSTRACT

There are some problems for a quantitative signal analysis of the amino group signals of nucleic acids. Here, we develop a simple method for the quantitative observation of the amino group signals of DNA oligomers, i.e., the combination of the ^{15}N labeling at the exocyclic amino groups of the cytosine residues and the water suppressed HSQC by spin locking with a relatively long recycle delay. By this method, the amino proton and nitrogen resonance positions were clearly identified.

INTRODUCTION

Dynamic characters of nucleic acids such as the base pair opening have been analyzed by measuring the exchange rates of the imino protons at the msec time scale (Gueron et al., 1987; Leroy et al., 1988; Leroy et al., 1991). Although the amino proton signals have higher potentiality for characterization of the base pair opening motion and for the detection of the inter molecular hydrogen bonding on the drug and/or protein binding, only a few works have been reported for the aqueous condition (Teitelbaum and Englander, 1975; McConnell, 1986). This fact was mainly caused by two reasons. One is the lack of the method for the quantitative observation of the amino group signals, because they are usually overlapping with the base proton signals and exchanging rapidly with water. Second, an extra motion of the amino group known as the rotation around the amino bond (Shoup et al., 1972; Raszka and Kaplan, 1972) has not yet been well analyzed.

The "water flip back" methods (Nagayama et al., 1990; Spera et al., 1991; Grzesiek and Bax, 1993; Stonehouse et al., 1994; Key et al., 1994) have enabled us to minimize the sensitivity loss of signals due to the rapid exchange with water. Among them the simplest technique is HMQC by the 1-1 pulse (Nagayama et al., 1990; Spera et al., 1991). Other techniques are more complicated ones, i.e., the selective pulse and/or the pulsed field gradient, which require extra instrumentations. The observed line widths by the HMQC method are generally wider than that by HSQC (Norwood et al., 1990), but there is no conventional HSQC method for the observation of rapidly exchanging protons without extra instrumental options. Here the author propose the HSQC method with water suppression.

MATERIALS AND METHODS

Preparation of the exocyclic amino ^{15}N labeled cytosine and its oligomer

Amino-^{15}N-deoxycytosine was synthesized from 4-(1,2,4-triazolyl)-deoxyuridine by the reaction with 3M $^{15}NH_4OH$ (Hodge et al., 1991). It was connected to the phosphoamidite derivative, and then following after blocking the amino group, oligomers were synthesized by a solid phase automated DNA synthesizer. The sequences of the DNA oligomers were 5'-CGGAAGACTCTCCTCCG-3' (UAS_G+), 5'-CGGAAGACTCTCCTCCG-3' (UAS_G-) and 5'-AAGGCCTT-3' (M8). The labeled cytosines are indicated by bold letters. The oligomers with the exocyclic amino ^{15}N labeled cytosines were obtained at relatively high yield.

The oligomers were deblocked and purified by a reverse phase C18 HPLC column twice, first on the dimetoxy triphenyl methyl protection group and second off, with the acetonitrile (0.1M triethylamine acetate) gradient. The counter ion was changed to Na^+ by an anion exchange FPLC column, following a gel filtration HPLC column to remove extra salts. The UAS_G+ and UAS_G- strands were mixed at 1 : 1 ratio, which was estimated by P1 nuclease hydrolysis, and then lyophilized and dissolved in a 20 mM sodium phosphate buffer with 50 mM NaCl. The sample for NMR measurements comprised 1 mM DNA at neutral

pH.

NMR Spectroscopy

All the NMR spectra were recorded at 288 K on a Bruker AMX500 or ARX500 spectrometer equipped with a 5 mm broadband heteronuclear reverse detection prove. The HMQC spectrum was taken by the 1-1 water flip back pulse (Nagayama et al., 1990; Spera et al., 1991) and the HSQC spectrum was observed by the spin lock water dephasing pulse (Messerle et al., 1989). The pulse repetition delay was 3 sec for both HMQC and HSQC experiments. The peak area intensities were obtained by wave fitting in one dimensional HSQC spectrum. The experimental conditions were the same to those used in the two dimensional NMR except the pulse repetition delay time.

RESULTS AND DISCUSSION

The HMQC spectrum was taken with the 1-1 water flip back pulse, and the excitation frequency center was set at the amino proton resonance frequency. The HSQC spectrum was taken with the 2 msec spin lock water dephasing pulse. In Fig. 1, both the HMQC and HSQC spectra of a 17mer DNA fragment named UAS_G are shown. It is noted that the HSQC spectrum of the nucleic acid amino proton resonances has been hardly obtained with the shorter pulse repetition delay time due to rapid exchange with water, but, it will be easily measured with the delay time longer than 3 sec such as at present condition. The linewidths of the amino nitrogen resonances in the HSQC spectrum were sharper than those in HMQC. Relative intensities of the peaks were not constant in HMQC depending on the excitation profile of the 1-1 pulse. At this point, HSQC is preferred, i.e., it is obviously steady.

The peak intensity depends on the pulse repetition delay time. In Fig. 2, the relative peak intensities of the hydrogen bonded and non-bonded amino protons of a 8 mer DNA named M8 were plotted against the pulse repetition delay time. The frequency difference between these two amino proton resonances was 900 Hz, and the differences between water resonance and those of the hydrogen bonded and non-bonded protons were 1700 and 800 Hz, respectively. The both amino proton resonances were separately observed, but their linewidths were broader than those of the imino protons. The line shapes of the two amino proton resonances were similar to each other. Their results indicates that the exchange rate between the two amino protons was slower than 900 Hz. However, the lower limits of the exchange rate between the two protons could not be determined from these data.

The delay time dependency shown in Fig. 2 was approximately simulated by the pseudo saturation recovery model with the relaxation time constant, 1.58 sec. This result has two important meanings. One, at least 5 sec is necessary for the pulse repetition delay time for quantitative analysis. That is, although three seconds is enough to record a spectrum, that is non-quantitative one. The result was obtained from a short self complemental DNA 8 mer, so the pulse repetition time value will be generally useful for all the Watson-Crick type base pairing duplexes. The following changes might introduce a little optimization of the repetition

delay time, such as pH, ion strength, temperature and so on. Two, there was no difference in the signal intensity between the hydrogen bonded and non-bonded amino protons. It means that their relaxation time constants were the same, and then the relaxation rate was faster than 0.6 ($\sim 1.58^{-1}$) Hz.

From the consideration of line shape, we could get the upper limit of the exchange rate between the two amino proton resonances (< 900 Hz). Then the lower limit was also estimated from the pseudo saturation recovery experiments (> 0.6 Hz). So, its rate was limited from 0.6 to 900 Hz, the same range of the exchange rate between the imino and water protons (~ several Hz). The exact exchange rate between the two amino protons should be reconsidered after combining with the exchange rate with solvent water protons.

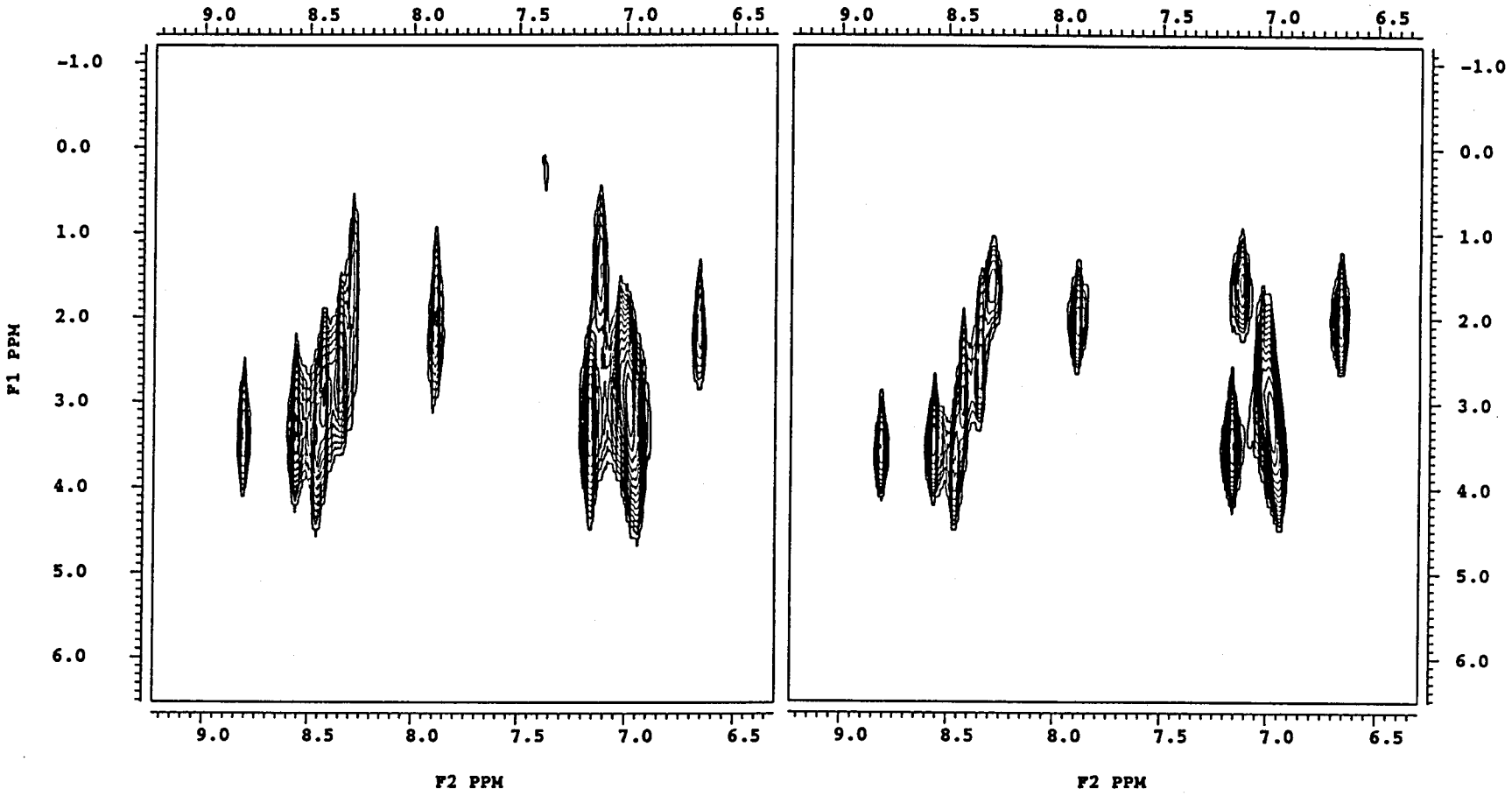


Figure 1. The HMQC (left) and HSQC (right) spectra of the 17mer DNA fragment named UAS_G with the pulse repetition delay 3 sec in H_2O at 288 K. The water signal reduction was performed by the 1-1 excitation pulse for HMQC and spin locking for HSQC. No modification for signal intensity was done.

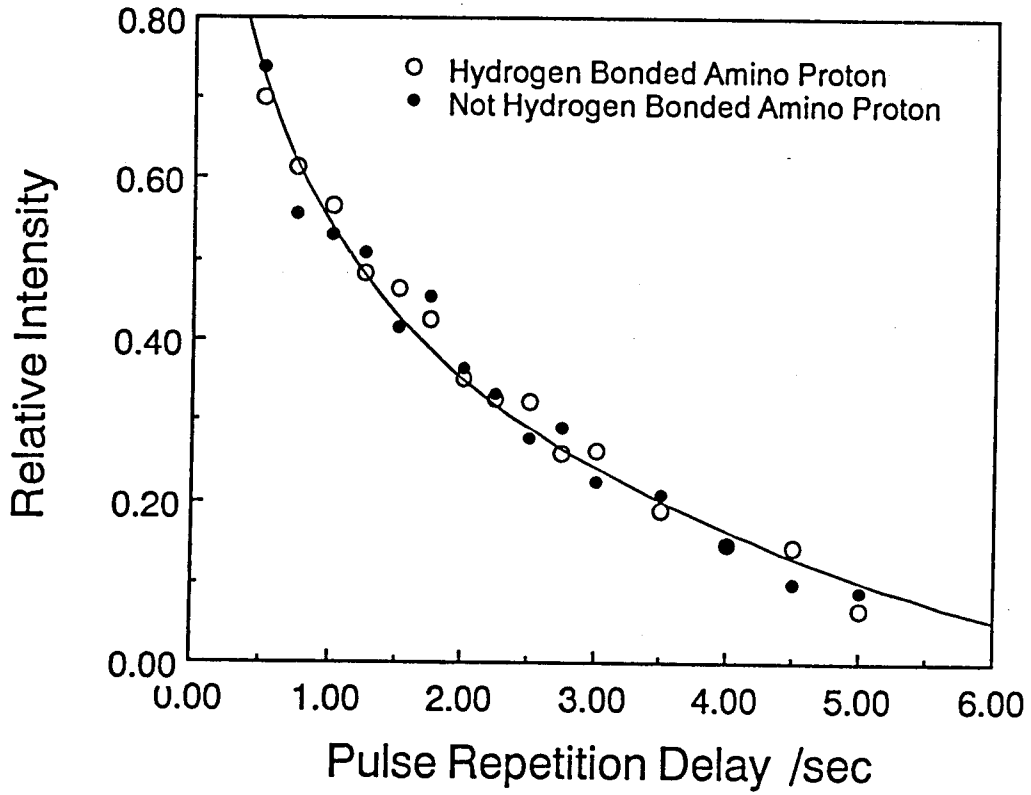


Figure 2. The relative signal intensities of the two amino protons were plotted as a function of the pulse repetition delay time for the 8 mer DNA fragment named M8. These intensities were obtained from the one dimensional HSQC experiment by fitting integration of the peaks. Open circles in the plot indicate the intensity of the amino proton signal and closed circles for non-bonded one. The line drew the best fitted decay curve obtained by the pseudo saturation transfer approximation for each amino proton.

Part II

Precise Structure Determination of a DNA Decamer

Chapter 5

NMR Rsym Factor as Analogy of the X-ray Rsym Factor, for Error Estimation of the NOESY Peak Volume

ABSTRACT

NMR Rsym factor, as an analogy of the X-ray Rsym factor, was introduced for the purpose of evaluating the reliability of the obtained NOESY peak volume values. This is calculated by the following equation,

$$\text{Rsym (NMR)} = \Sigma \mid I - \langle I \rangle \mid / \Sigma I$$

where I is an observed NOESY peak volume and <I> is an averaged volume of those of the two symmetric cross peaks. If the Rsym value is small, the reliability of peak volume value is high by this definition. NMR Rsym factor showed strong dependency on several error sources in the process of the determination of the NOESY peak volumes. We conclude that the NMR Rsym factor is a good indicator of the reliability of the obtained NOESY peak volumes.

INTRODUCTION

NOE values are necessary for the NMR structure determination using the direct NOE refinement calculation and even if for the semi-quantitative use of the distance geometry (DG) and simulated annealing (SA) calculations. Therefore, an indicator, which reflects the reliability of the obtained NOESY peak volumes depending on several factors, is necessary for the estimation of the reliability of all the calculated structures. Here we introduced the NMR Rsym factor for the error estimation of the 2D NOESY peak volumes.

DEFINITION

NMR Rsym factor is calculated by the following equation,

$$\text{Rsym (NMR)} = \Sigma \mid I - \langle I \rangle \mid / \Sigma I \qquad [1]$$

where I is an observed NOESY peak volume and <I> is an averaged volume of those of two symmetric peaks appear in the NOESY spectrum. If the Rsym value is small, the reliability of peak volume values is high by this definition. This NMR Rsym factor clearly showed strong dependencies on several factors, especially on the pulse repetition delay time. The peak volumes of the cross peaks at the counter positions in a NOESY spectrum should be equal (Solomon, 1954; Macura and Ernst, 1980), if the magnetization completely relaxes during the repetition delay between each scans. So, we could define NMR Rsym factor as analogy of the X-ray Rsym factor for checking reliability of the peak volume value.

APPLICATION AND DISCUSSION

To get an exact experimental NOESY peak volume, we have to estimate errors in data acquisition and processing parameters, i.e., pulse sequence, point resolution in acquisition and processing, window function, threshold, integration method and so on, by using the NMR Rsym factors.

Firstly, three different NOESY experiments were tested to estimate errors which arose from the zero quantum coherence. However, this factor did not produce obvious errors, i.e., less than 5% in the NMR Rsym factor. This independency is consistent with a previous report estimated by computational simulations (Wang et al., 1992).

NMR Rsym factor depended strongly on integration parameters, such as thresholds and integration regions. The fitting procedure of the peak volume integration of the NMR2 software package (New Methods Research, Inc.) gave small NMR Rsym values, but the intensity summation procedure in a specified region did not. By the summation method, the observed volumes depended strongly on the selection of the region and the threshold of the

display. The fitting procedure of the integration is preferred from the view of the NMR Rsym factor value. This is consistent with the previous report (Weiss and Ferretti, 1983; Nibedita et al., 1993; Brown and Huestis, 1994).

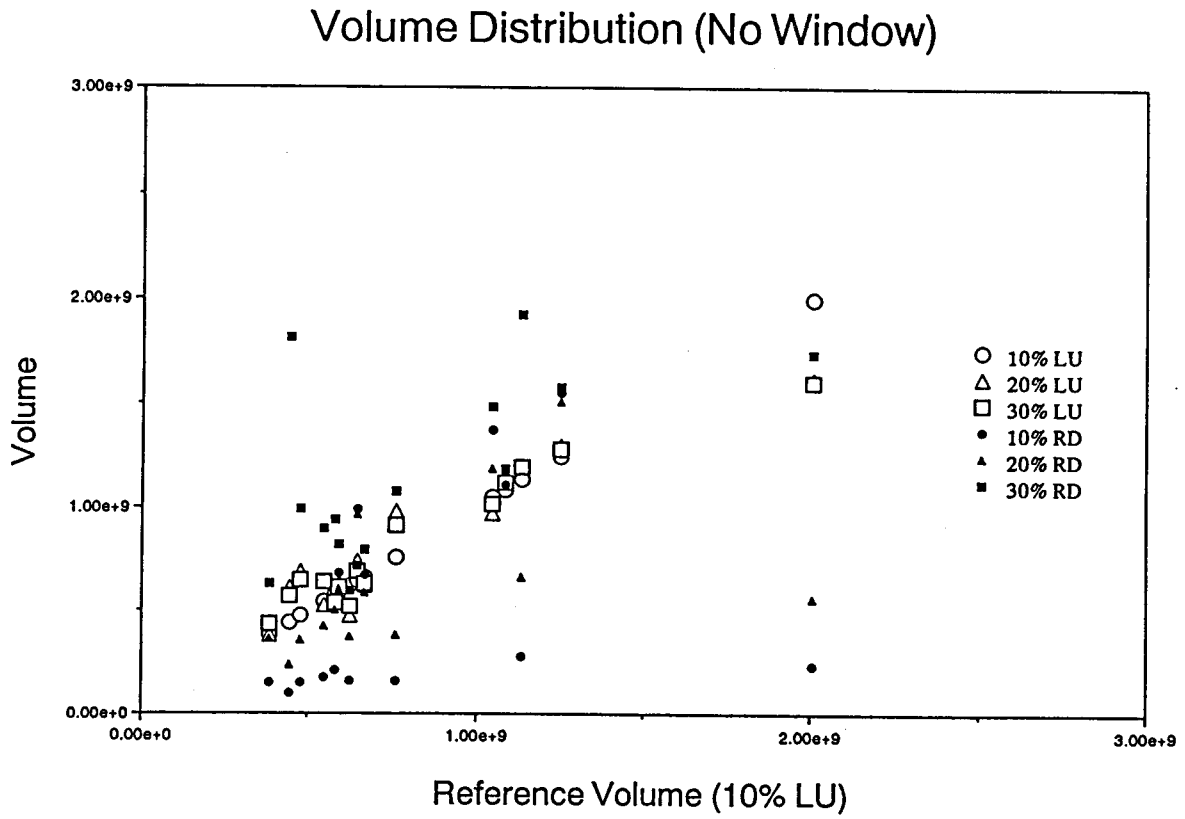
Not only the absolute peak volume but the NMR Rsym factor depended most strongly on the first point intensity of FID, which was affected by a window function (Fig. 1). From the results, the reduction of absolute intensities which seemed as an increase of errors, caused the decrease of the NMR Rsym factor value which result in as an increase of the precision. However, it should be understood that, the NMR Rsym factor did not reflect errors, but only systematic errors were detected for the integrated absolute intensities. This is important because the window function affects the peak separation in the crowded spectral region.

NMR Rsym factor depended strongly on the pulse repetition delay time only when it is too short. The theoretically recommended repetition delay, five times of the longitudinal relaxation time T1, is too long to record signals with high S/N ratio. Too short repetition delay leads to uncomfortable and not analytical relaxation. Here, this unexpected relaxation is detected by NMR Rsym factor through the localized and unequilibrated magnetization (local T1 value difference among individual protons). Then, this characterized property was useful to determine the minimum repetition delay time. For example, at least 2.5 times of T1 as the repetition delay is necessary for the data collection of the DNA decamer.

Finally, we recommend to give attention to the point resolution dependency of the NMR Rsym factor. If a certain NOESY cross peak came from different spin systems in both dimensions, the point resolution in the acquisition should be equal or nearly equal in both dimensions. The difference of the point resolution in the acquisition and processing introduces an additional apparent linewidth dependency. In that case it is necessary to consider the window function dependency of the NMR Rsym factor.

In this chapter, NMR Rsym factor as analogy of the X-ray Rsym factor, was introduced for the purpose of evaluating the reliability of the obtained NOESY peak volume values. The NMR Rsym factor showed the strong dependency on several error sources arose in the process of the determination of NOESY peak volumes. We conclude that the NMR Rsym factor is a good indicator of the reliability of the obtained NOESY peak volume.

Figure 1


Volume Distribution (No Window)
Volume
3.00e+9
2.00e+9
1.00e+9
0.00e+0
0.00e+0
1.00e+9
2.00e+9
3.00e+9
10% LU
20% LU
30% LU
10% RD
20% RD
30% RD
Reference Volume (10% LU)

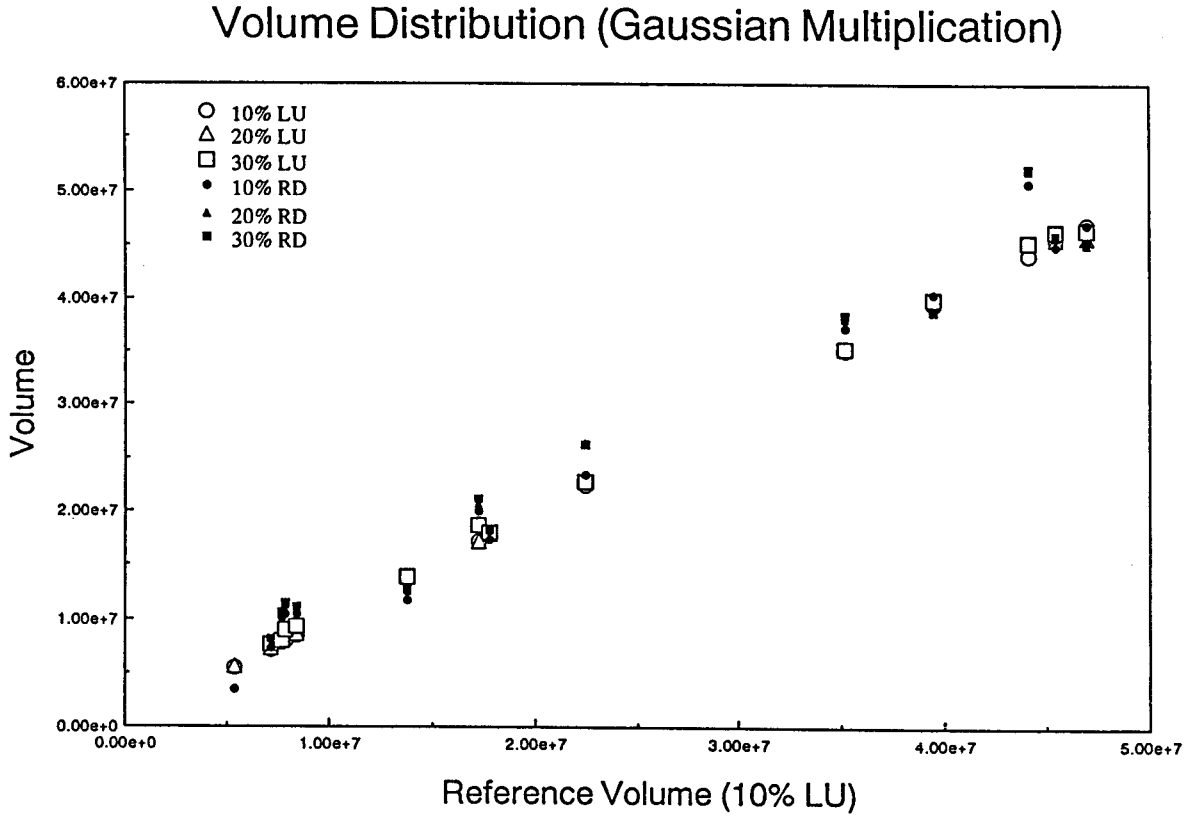

Volume Distribution (Gaussian Multiplication)
Volume
6.00e+7
5.00e+7
4.00e+7
3.00e+7
2.00e+7
1.00e+7
0.00e+0
0.00e+0
1.00e+7
2.00e+7
3.00e+7
4.00e+7
5.00e+7
10% LU
20% LU
30% LU
10% RD
20% RD
30% RD
Reference Volume (10% LU)

Figure 1. Effect of the window function to the relative volume intensity. The volume intensities of some DNA peaks measured in different threshold value spectra are plotted against those of the corresponding peaks appeared in the left upper (LU) area of the p-type NOESY spectrum with 10 % threshold. Upper: without any window function, Lower: with the Gaussian function window. Percentage: the threshold level against the maximum peak intensity; LU: left upper region peak in the p-type NOESY; RD: right lower region peak in the p-type NOESY. The Rsym factor values calculated for the peaks shown in the upper plot are 0.317, 0.176 and 0.162 for 10, 20 and 30 % threshold, respectively. The total Rsym factor calculated for the three threshold levels is 0.249. The Rsym factors calculated for the lower plot are 0.040, 0.042 and 0.041 for 10, 20 and 30 % threshold, respectively, and the total Rsym factor is 0.043.

Chapter 6

Investigation of the Validity of the Single Correlation Time Model as the First Order Approximation for the Estimation of the ^1H-^1H Distances in a DNA Decamer

ABSTRACT

The ^1H-^1H distances in a DNA oligomer were calculated from the NOESY cross peak intensity sets with eight different mixing times. The full relaxation matrix approach was tested to obtain more accurate distances in multi spin systems. In this calculation, the isotropic rotational model was adopted as a motional model of a DNA decamer. In spite of the wide dispersion in correlation of the ^1H-^1H distances calculated from the different intensity sets with different mixing times, this approximation was fairly good as the first order approximation. However, because of the existence of some error sources, too short mixing times (30 and 60 msec) caused shorter ^1H-^1H calculated distances, and too long mixing times (250 and 300 msec) brought an ambiguous estimation of distances. Therefore the NOESY mixing time should be limited in the range of 90 to 200 msec for the full relaxation matrix calculation in the case of the DNA decamer.

INTRODUCTION

The ^1H-^1H distance informations are necessary for the NMR structure determination using the distance geometry (DG) and simulated annealing (SA) methods as the input experimental data. In the case of the ^1H rich molecules like proteins, the number of the input distances is most critical, but the precision of the distance information is not so seriously required for the structure calculation. The number of the ^1H-^1H distances depends on the density of protons, so the precision of the ^1H-^1H distances is required for the low ^1H density molecules like nucleic acids. Here the ^1H-^1H distances of a DNA decamer, $d(GCATTAATGC)_2$, were calculated from the NOESY cross peak intensities by the use of the full relaxation matrix calculation with the MARDIGRAS software (Borgias and James, 1990). In the process of this calculation, we investigated the validity of the single correlation time model for the motion of the DNA decamer.

METHODS

All NOESY spectra were obtained on a Bruker ARX500 spectrometer (500 MHz) and the temperature was set to 303 K. The NOESY spectra were recorded using eight mixing times, 30, 60, 90, 120, 150, 200, 250 and 300 msec, and the pulse repetition delay time was 5 sec. Each mixing time was changed in the range of 10 msec to suppress the cross peaks which arose from the zero quantum coherence transfer. For all the spectra, 256 hyper complex points for t1 and 1024 complex points for t2 were recorded, and 32 scans for each t1 increment were performed. Then, the Gaussian windows were applied to both frequency axes, followed by zero filling to 1024 real points. NOEs were obtained from the volume integration of the NOESY cross peaks using the fitting procedure by the NMR2 software (New Method Research Inc.). The errors of the peak volumes were estimated below 5 % in the NMR Rsym factor (chapter 5).

RESULTS AND DISCUSSION

In Fig. 1, all the ^1H-^1H distances, calculated from the seven intensity sets of NOESY recorded with seven different mixing times, were plotted against that of 120 msec. The plotted distances calculated for the short mixing times (30 and 60 msec) were shown in the upper of Fig. 1, for the middle mixing times (90, 150 and 200 msec) in the middle of the figure, and for the long times (250 and 300 msec) in the lower. In spite of the wide dispersion of correlation of each ^1H-^1H distances calculated from the different intensity sets with different mixing times, the middle plot of Fig. 1 for the middle mixing times showed a clear correlation. This result should be understood that the isotropic rotational model was fairly good as the first order approximation of the molecular motion in the middle mixing time range.

However, too short mixing times (30 and 60 msec) caused shorter ^1H-^1H calculated distances as shown in the upper plot of Fig. 1, and too long mixing times (250 and 300 msec) introduced an ambiguous estimation of distances as shown in the lower of Fig. 1. So, the distances calculated from both of the too short and too long mixing times were deviated and not reliable. This result might be understood as a failure of the single correlation time model, because the relaxation matrix should be self consistent with all the mixing time range. In addition other error sources are conceivable. First, the assignment of the 5' and 5" proton resonances which occupied 23 % of the non-labile proton resonances was not completed. Moreover many of heavily overlapping NOESY cross peaks were not taken into account. This incompleteness of the input data for the relaxation matrix might lead to the leakage of the calculated magnetization as shown in the lower plot of Fig. 1 for the long mixing times. Second, only the precision of the estimated peak volume values was taken care of in our experiment, that is, the accuracy was not taken into account. In the case of the low S/N ratio spectra (30 and 60 msec mixing NOESY), the volume values of smaller peaks, which have relatively bigger errors, might be extremely neglected. This caused severe lack of the input data, and introduces systematic errors for weighted intense peaks (upper of Fig. 1). It was noted that the failure of the single correlation time model could not be observed clearly in our study.

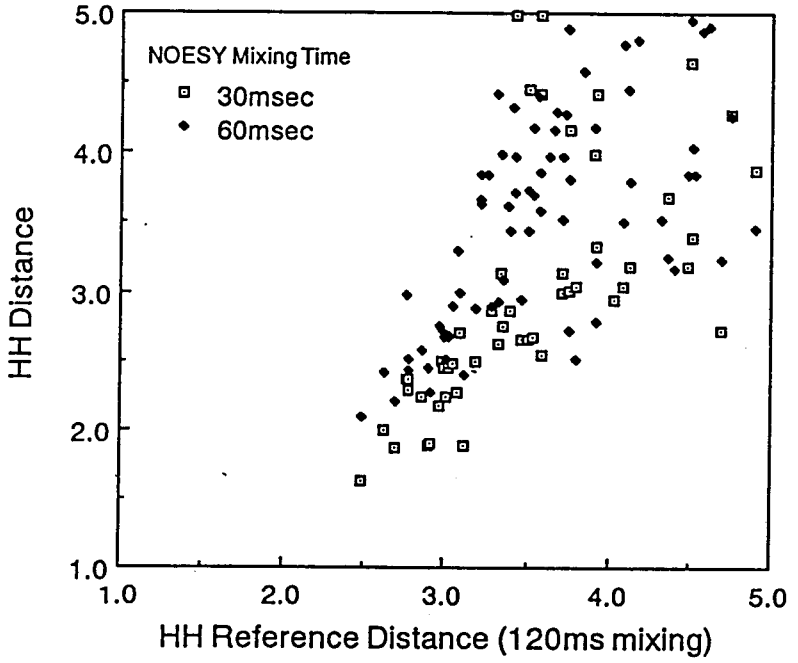
By searching the minimum of the R^{-6} factor (James et al., 1991) value in the wide range of the correlation time below 30 nsec, the optimal correlation time was 14 nsec. This was somewhat larger than the estimated value, 2 nsec, for the DNA fragment of this size by other methods (for a review, see Lane, 1993). However, the presence of a single minimum supports the single correlation time model. A larger value of the optimized correlation time would be understood as a proof of the presence of internal local motions.

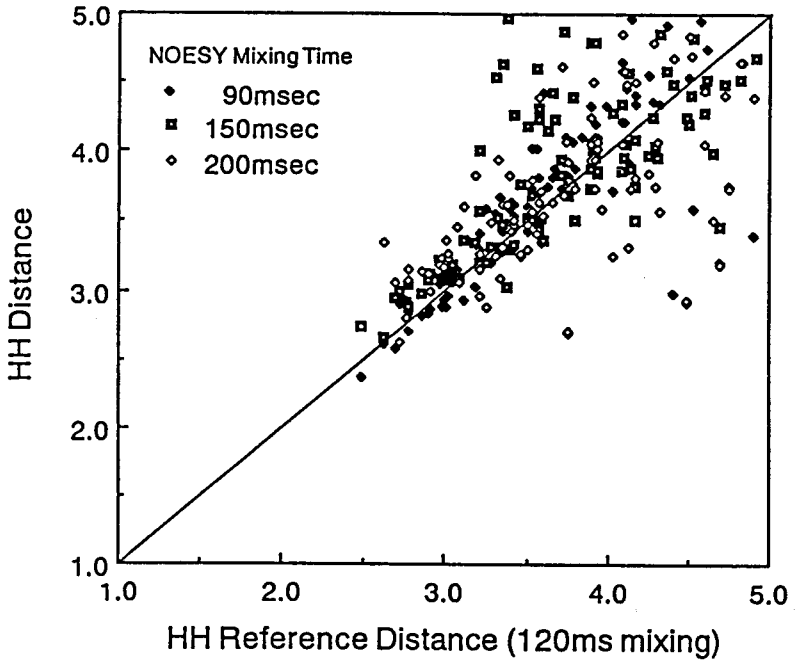
The validity of the single correlation time approximation was checked by changing the NOESY mixing time. We could not record obvious discrepancy throughout the mixing time range from 30 to 300 msec. However the mixing time should be limited in the range of 90 to 200 msec for the full relaxation matrix calculation of the DNA decamer, because of other error sources. The error sources would cause some deviation of the ^1H-^1H distances even in the limited mixing time range, and the wide dispersion shown in the middle plot of Fig. 1 should be partially attributable to the errors.

CONCLUSION

The middle plot of Fig. 1 showed a clear correlation of ^1H-^1H distances calculated from different intensity sets with different mixing times for the middle mixing times (90 ~ 200 msec) despite wide dispersion. Moreover, the presence of the local internal motion was suggested. However, we can conclude that the isotropic rotational model is fairly good as the first order approximation of the molecular motion of the DNA decamer at least.

Figure 1


NOESY Mixing Time
30msec
60msec
HH Distance
HH Reference Distance (120ms mixing)
1.0
2.0
3.0
4.0
5.0


NOESY Mixing Time
90msec
150msec
200msec
HH Distance
HH Reference Distance (120ms mixing)
1.0
2.0
3.0
4.0
5.0

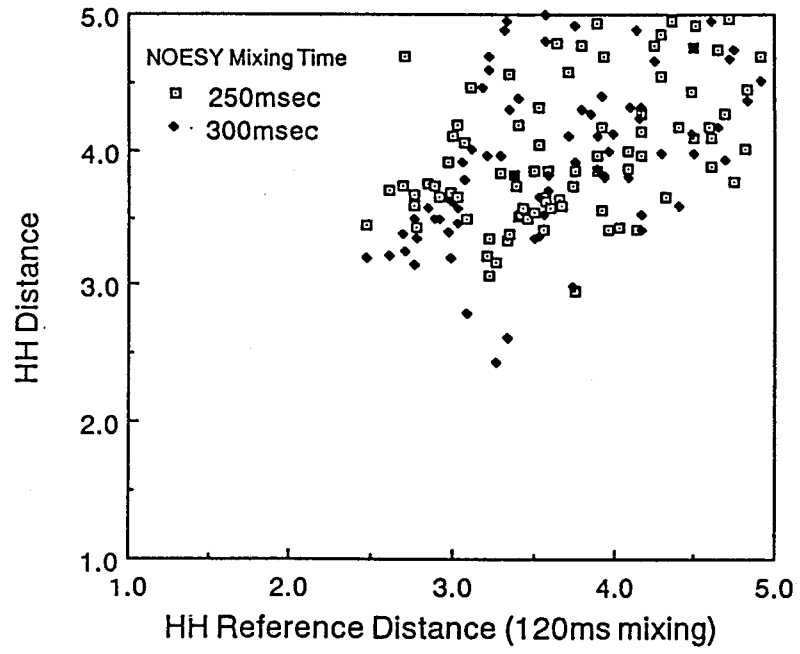

NOESY Mixing Time
250msec
300msec
HH Distance
HH Reference Distance (120ms mixing)
1.0
2.0
3.0
4.0
5.0

Figure 1. The plots of the ^{1}H-^{1}H distances obtained for a proper mixing times against the distances obtained at a references mixing time, 120 msec. Upper, for the short mixing times (30 and 60 msec); middle, for the middle mixing times (90, 150 and 200 msec); and lower, for the long mixing times (250 and 300 msec). The ^{1}H-^{1}H distances were obtained from a full relaxation matrix calculation using the MARDIGRAS software with a single correlation time approximation. The eight intensity sets of the zero quantum coherence suppressed NOESY spectra, recorded with eight different mixing times, were used for that calculation. The errors of the peak intensities were estimated below 5 % in the NMR Rsym factor (chapter 5).

Chapter 7

Precise Structure Determination of a DNA Decamer by the Back Calculation Refinement against NOESY Intensities with Different Mixing Times

ABSTRACT

Three dimensional coordinates of a DNA decamer, d(GCATTAATGC)$_2$, were obtained by the back calculation refinement against NOESY cross peak intensities with four different mixing times, 90, 120, 150 and 200 msec. The precision of the structure, calculated with the combined NOESY data for all the four different mixing times together, was greater than those of the structures independently calculated for each mixing time. The combined data calculation was performed by using a single relaxation matrix with a certain correlation time, four sets of the NOESY cross peak intensities and the mixing times. This protocol can be regarded as pseudo fitting to the NOESY build up curves of each peak, and thus leads to a precise structure.

INTRODUCTION

The structure of nucleic acid double strands is usually undetermined precisely from NMR informations. In particular, for the overall shape or bend of the molecule there is no boundary information from NMR. Thus, it has been believed that neither the distance geometry nor ab initio simulated annealing methods produce unique structures. Up to the present, this problem have been avoided by including additional restraints, such as repulsive distance restraints between appropriately chosen phosphate groups, or by starting the refinement process from proper ideal conformations. When relatively many restraints are available, the latter procedure is generally adopted, that is, typical A and B form coordinates of nucleic acids are used for the start of refinement.

Recent methodological developments of the final structure refinement by the use of the relaxation matrix could result a coordinate group of alike conformers more precisely. Especially, direct NOE refinement based on the relaxation matrix method of the Yip and Case algorithm (Yip and Case, 1989; General Introduction) is useful, because this can be used for directly minimizing the convergence of coordinates in the final refinement step. For example, a rough coordinate group, obtained by the distance geometry or the ab initio simulated annealing procedure, causes unrealistically averaged ones which define bond lengths and bond angles deviated from ideal ones. In that case, the empirical energy of the averaged structure with the distorted coordinates have been minimized. However, this energy minimized averaged structure obtained by this procedure does not represent the averaged conformer with the averaged input data. So the final refinement, such as direct NOE refinement, is necessary to obtain a set of more realistic averaged coordinates as a substitutive procedure.

Many double stranded DNA oligomers with a variety of different base sequences are well studied by NMR. A self complementarily base paired decamer, $d(GCATTAATGC)_2$, was also previously analyzed by NMR (Chazin et al., 1986). The report told us that this decamer formed a duplex and the overall conformation was similar to the typical B form by a semi-quantitative analysis. Moreover, two DNA decamers with a little different sequences, $d(CGATTAATCG)_2$ and $d(CCATTAATGG)_2$, were studied by X-ray diffraction in detail (Quintana et al., 1992; Goodsell et al., 1994). They also reported that the conformation of the decamers resembled a typical B form.

In this chapter, we employed the direct NOE refinement method at the final step for the determination of the three dimensional coordinates of a DNA decamer, $d(GCATTAATGC)_2$. The input NOESY cross peak intensity sets were obtained under four different mixing times. The precision of the structure which was calculated by the combined use of the NOE data for four different mixing times, was greater than those independently calculated by the data for each mixing time. This combined use protocol can be regarded as pseudo-fitting to the NOESY build up curves of each peak, and thus leads to a precise structure.

MATERIALS AND METHODS

Sample preparation

DNA oligonucleotide 5'-GCATTAATGC-3' was chemically synthesized by a solid phase automated DNA synthesizer (ABI 381B). These oligomers were deblocked and purified by a reverse phase C18 HPLC column twice, first on the dimetoxy triphenyl methyl protection group (Tr-on) and second, Tr-off, with the acetonitrile (0.1M triethylamine acetate) gradient. The counter ion was changed to Na^+ by an anion exchange FPLC column, following a gel filtration HPLC column to remove extra salts. The quantity was estimated by the nuclease P1 hydrolysis method, then lyophilized and dissolved in a 20 mM sodium phosphate buffer with 50 mM NaCl. The samples for NMR measurements comprised 6 mM DNA at neutral pH.

NMR experiments and estimation of the NOE intensity

All ^1H spectra were obtained on a Bruker ARX500 spectrometer (500 MHz) and the temperature was set to 303 K. NOESY spectra were recorded using four mixing times, 90, 120, 150 and 200 msec, and the pulse repetition time was 5 sec. The mixing time delay was changed in the range of 10 msec to suppress the cross peaks which come from the zero quantum coherence transfer. For all the spectra, 256 hyper complex points for t1 and 1024 complex points for t2 were recorded, and 32 scans for each t1 increment were performed. Then, the Gaussian window was applied to both dimensions, following zero filling to 1024 real points for both dimensions. NOEs were obtained from the volume integration of the NOESY cross peaks using the fitting procedure of the NMR2 software package (New Method Research Inc.). The errors of the peak volumes were estimated below 5 % in precision.

Coordinate calculation

The three dimensional coordinates were calculated by the direct NOE refinement in the X-PLOR software (Brunger, 1992). The initial coordinates of the DNA decamer were derived from the QUANTA software (Molecular Simulations, Inc.) as those of an ideal B form. The systematic errors of the calculation which arose from the single use of the initial coordinates, were removed by the random initial velocity in the molecular dynamics used in the direct NOE refinement of the X-PLOR software. Input informations were NOE peak volumes, their error range 5 %, some NOE observation conditions and relaxation parameters. We employed the isotropic rotational model as the first approximation of the molecular motion. The single correlation time was postulated to 2 nsec. All the methyl protons were treated as unresolved ones, and the rotational motion was assumed to be enough fast for the R^{-3} averaging approximation of dipolar relaxation (Trop, 1980).

Two different strategies were taken for evaluating the effectiveness of the use of several NOESY peak volume intensity sets. The First strategy is that the four intensity sets with different mixing times were assumed to form a single input matrix for the calculation based on the combined data sets, and processed in a single relaxation matrix. The second method is that the four intensity sets were assumed to form four independent input data, and processed in each of the four relaxation matrices, separately. Results of the two procedures of calculations were compared by the pairwise root mean square deviation (RMSD) values. The

first one gave a coordinates set RMSD, but the second one did four different coordinates sets RMSDs. Average of each four coordinates sets gave a new coordinates set, then the new and pairwise RMSD were obtained throughout the second procedure.

RESULTS AND DISCUSSION

Estimation of NOE Peak Volume

NOEs were obtained from the volume intensities of the NOESY cross peaks. To get the exact experimental peak volumes, we estimated errors arose from the stages of acquisition and processing, i.e., pulse sequence, pulse repetition delay, point resolution in acquisition and processing, window function, threshold, integration method and so on. The precision of the peak volume values was checked by the following evidence. The peak volumes of the cross peaks at counter positions in a NOESY spectrum should be equal, if the magnetization completely relaxes during the repetition delay between each scans (Solomon, 1954; Macura and Ernst, 1980). The threshold level of the integration should not affect the relative peak volume values, in other words, the integrated values should not depend on the spectral S/N ratio. Using these indicators, we checked the precision of the peak volume intensities by plotting the volume values of one of two symmetric peaks against the other one at some different threshold levels.

The precision of each volume depended most strongly on the integration parameters, such as threshold and integration area. The fitting procedure of the peak volume integration in the NMR2 software package showed high precision of the values, but the intensity summation procedure in a specified area did not. The summation method strongly depended on the selection of the area and the threshold of the display, and, moreover, showed low precision. The fitting procedure of the integration should be preferred as suggested by many reports (Weiss and Ferretti, 1983; Nibedita et al., 1993; Brown and Huestis, 1994).

Three different NOESY experiments were tested to estimate the errors of peak volumes which were caused through the zero quantum coherence. However, this factor was found not to produce obvious errors. It suggests that other factors such as S/N ratio and processing parameters cause bigger errors. Not only the absolute peak volume intensities but their precision depended most strongly on the first point intensity of FID. From the result, the reduction of the absolute intensities which seemed as an increase of errors caused the decrease of the NMR Rsym factor value which seemed as an increase of precision. However, it should be understood that the precision did not reflect errors, but only the systematic errors were detected in the absolute intensities. This is important because the window function affects the peak separation in the crowded spectral region. Precision also depended strongly on the pulse repetition delay time when it is too short. The theoretically recommended repetition delay, five times of the longitudinal relaxation time T1, is too long to record signals with high S/N ratio. Too short repetition delay led uncomfortable and non-analytical relaxation. Here, this unexpected relaxation was detected as the deviation from the symmetric character of the NOESY peaks through the localized and unequilibrated magnetization (local T1 value difference

among protons). Then, this characterized property was useful to determine the minimum repetition delay time. For example, 2.5 times of T1 as the repetition delay is necessary at least for the DNA decamer sample.

Structure Determination

The postulated single correlation time, 2 nsec, was taken from the previous reports on dynamic laser light scattering, rotational depolarization of fluorescence, phosphorus, carbon or deuterium NMR relaxation analyses and so on for the same size DNA oligomers (Lane, 1993). If internal motions do not affect the total isotropic rotational correlation time, this postulated value can be effectively used for the full relaxation matrix calculation.

In Fig. 1, five calculated coordinates sets are presented as five superimposes of five structures for each set. All the coordinates were obtained from the same conditions, that is, the combined input of four data sets, 90 msec mixing time data set, 120 msec, 150 msec and 200 msec, from left to right in Fig. 1. Pairwise RMSD values were listed in Table 1, with the root mean square (RMS) differences from the calculated coordinates for the combined NOE data given in the left end of Fig. 1. The coordinates presented in Fig. 2a as a stereo view were the same to the most left one in Fig. 1. The coordinates in Fig. 2b were obtained from the average of the four coordinates sets for different mixing times shown in Fig. 1 except for the left end one. As the precision, the pairwise RMSD value were 0.15 and 0.50, for Fig. 2a and Fig. 2b respectively.

The RMSD values in Table 1 told us two important facts. One, the pairwise RMSD value of the combined data one (shown as ALL in Table 1), was smallest among the five sets. Two, the RMS difference values from the combined data coordinates, showed a dependency on the mixing times. The first result can be understood as that the combined use of the data sets was most effective for obtaining the precise structure. The second result might be understood as that the combined use method caused different weights for the input data, which might be produced from systematic errors of peak volume intensities. However, the RMSD values of the obtained coordinates were bigger than each pairwise RMSD value. So, the second fact had better to be treated as an indicator for the improvement, and should not be regarded as the failure of this method.

The pairwise RMSD value of the three dimensional coordinates obtained by this calculation, strongly reflects the combination of calculations, the cut off distance of electrostatic interaction, the charge and radii of atoms, the weight of each potential, the annealing schedule and so on. Thus, the RMSD value should be understood as not the indicator of the precision of the structure, but as one of the indicators of the reliability of the obtained averaged coordinates.

The structure of nucleic acid is naturally flexible and difficult to determine three dimensional coordinates in solution. Therefore, it is important to solve the so-called multi conformer problem for structure determination. Previously reported three dimensional structures of DNA oligomers by NMR (see Wijmenga et al., 1993 for review) were incomplete in this sense (Lane, 1990). Under the multi conformer condition, the structure is not described by a single coordinate group of alike conformers (Schmitz et al., 1993; Kojima et al., 1993).

However, all the methods for determination of the solution conformer, NMR, CD and so on, clearly show that the population of minor conformers is relatively low. Thus, the main conformation should be determined by NMR. On the other hand, from the view point of the NOE time scale, the multi-conformations at different time scales might be averaged out, and may be seen as an averaged single conformer. For the determination of the true averaged conformation, the internal local motions should be taken into account in detail.

CONCLUSION

The three dimensional coordinates of a DNA decamer, d(GCATTAATGC)$_2$, were obtained by the back calculation refinement against NOESY cross peak intensities with four different mixing times, 90, 120, 150 and 200 msec. The precision of the structure which was calculated by using the combined data for four different mixing times was better than those independently calculated for each mixing time. The combined use method was the most effective procedure to utilize the four NOESY peak volume intensity sets, and useful to minimize the empirical energy arouse from the distorted averaged coordinates under the averaged experimental restraints. This protocol can be regarded as the pseudo fitting to the NOESY build up curves of each peak, and thus leads to the more precise and realistic averaged structure.

Table 1. Pairwise RMSDs and RMS differences among calculated coordinates shown in Figure 1

	ALL	90	120	150	200
Pairwise RMSD	0.15	0.32	0.36	0.42	0.47
RMS Difference*	target	1.92	1.67	1.56	1.37

* RMS difference from the target coordinates.

Figure 1

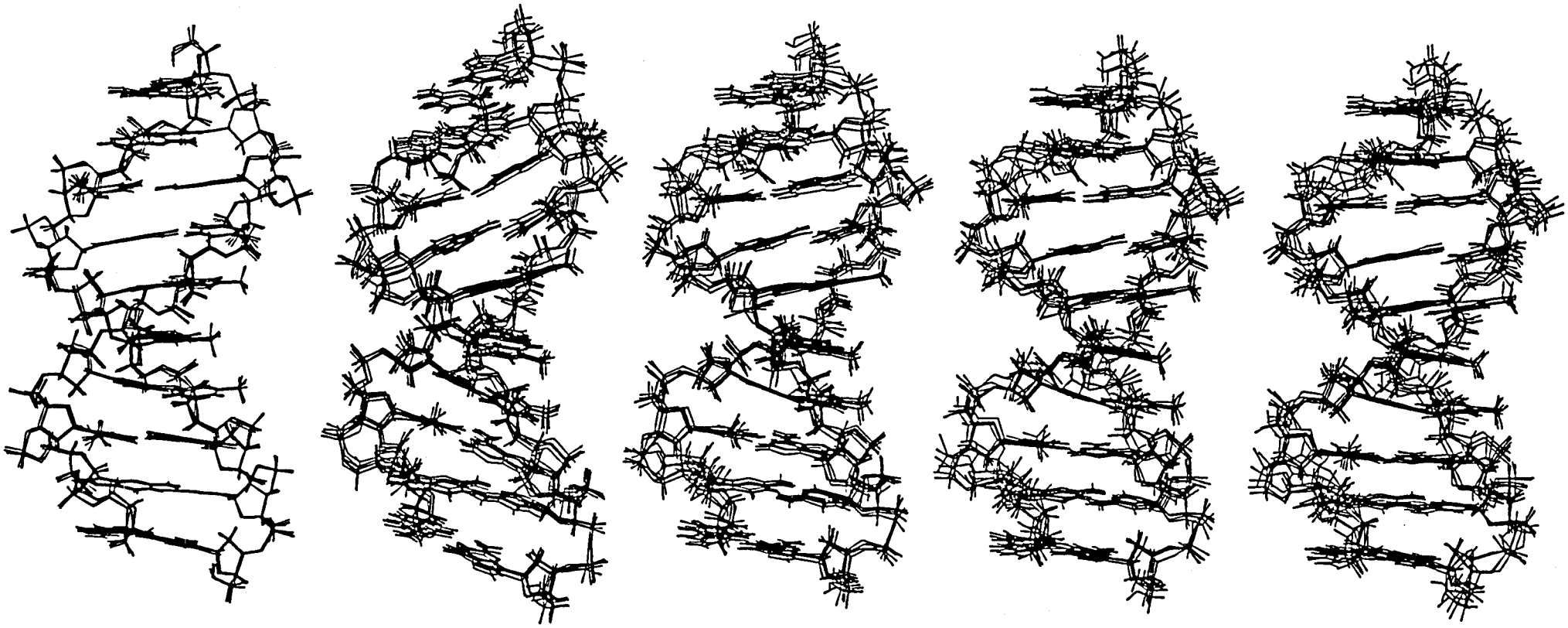


Figure 1. Five calculated coordinates sets are presented as five superimposes of five structures for each set. All the coordinates were calculated by the direct NOE refinement in X-PLOR software (Brunger, 1992). The left end set was obtained by the combined use of the four NOESY peak volume intensity sets for four different mixing times, 90, 120, 150 and 200 msec, assuming a single input matrix for the calculation, and processed in a single relaxation matrix. For the rest ones, four intensity sets were treated as four independent input data, and each of four relaxation matrix was solved separately. Each input data set for each coordinates set, from left to right in Figure 1, was obtained with four different mixing times, 90, 120, 150 and 200 msec, respectively. Pairwise RMSD values are listed in Table 1, with the root mean square (RMS) difference values from those of the combined data coordinates given in the left end of Figure 1.

Figure 2a

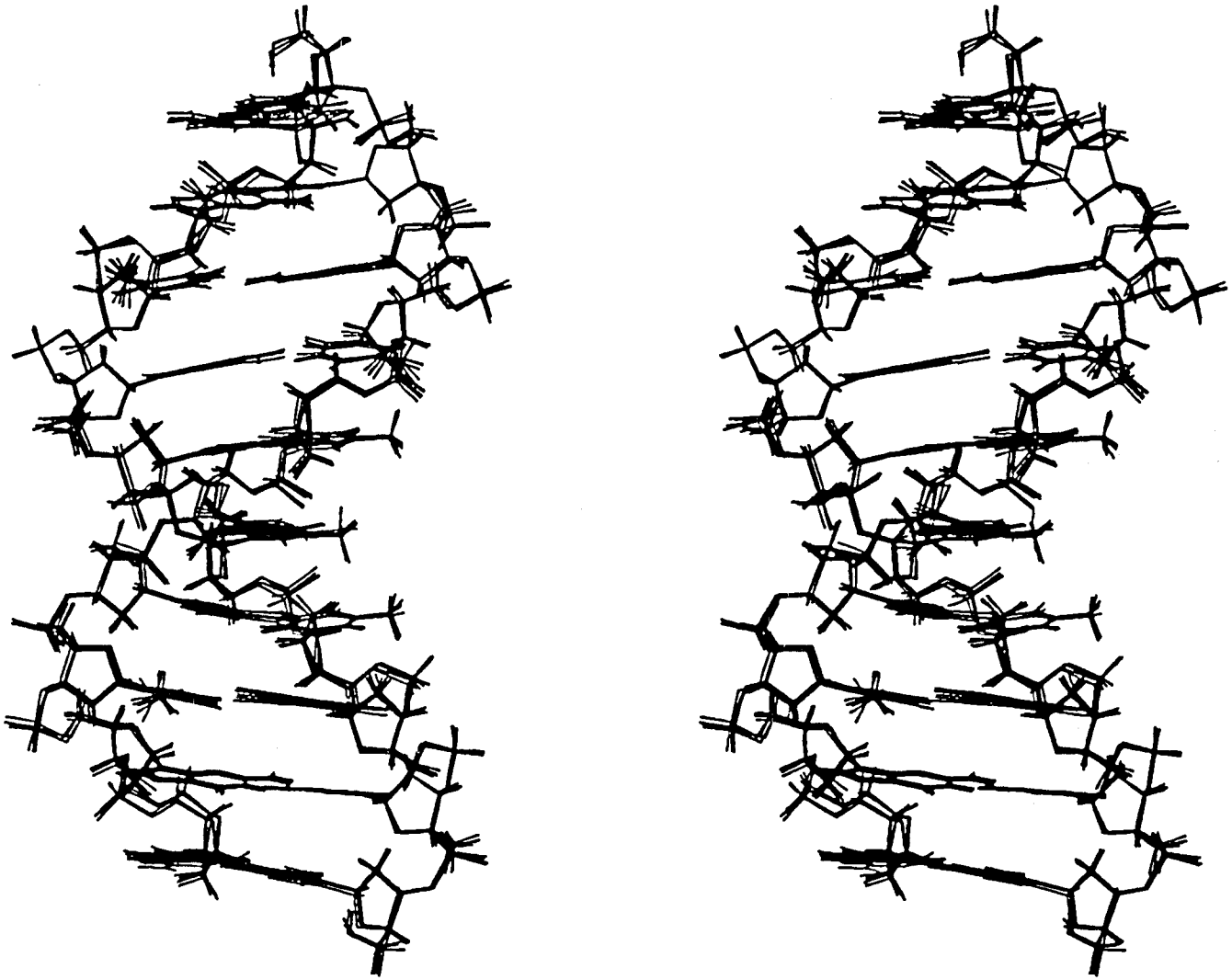


Figure 2b

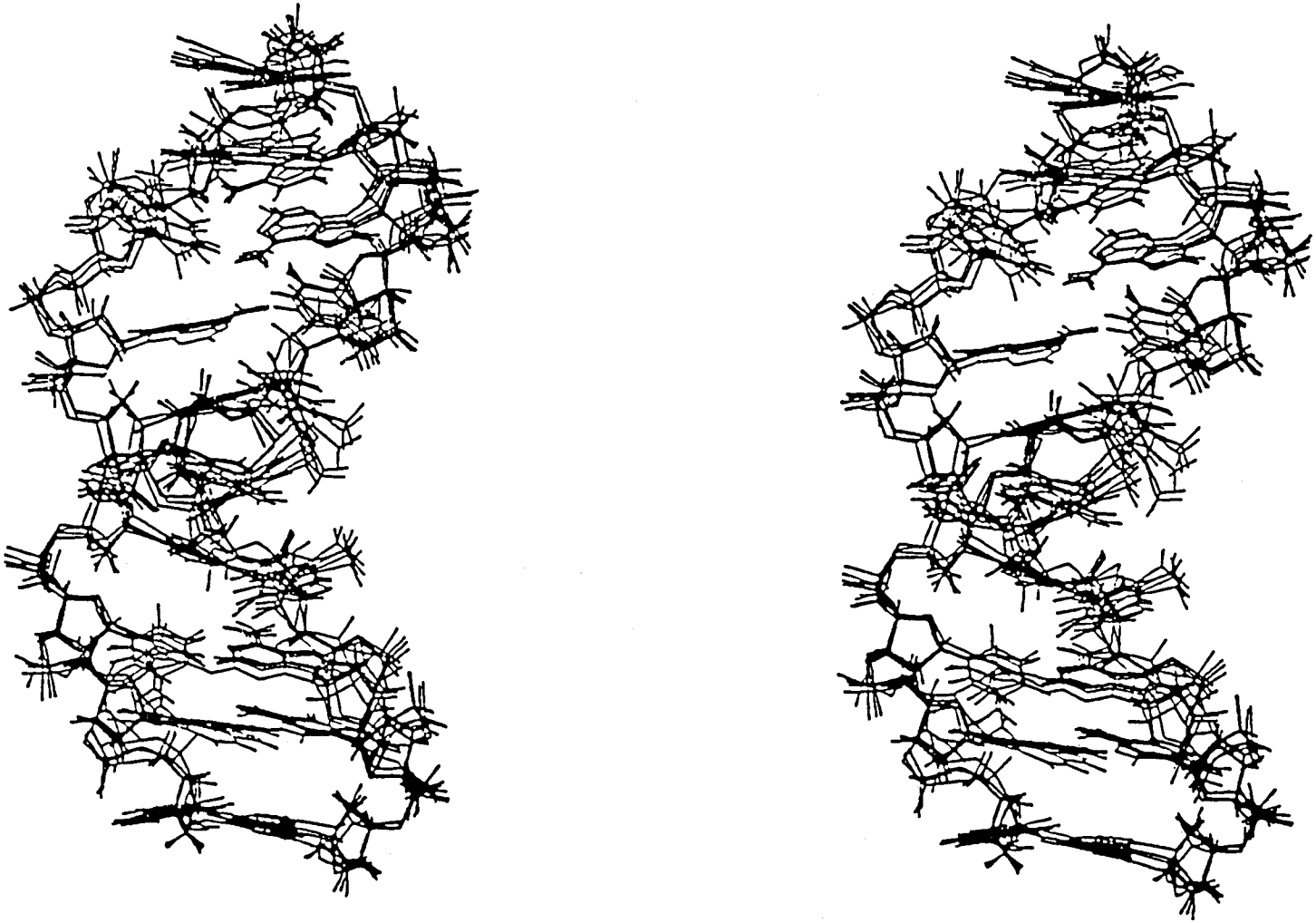


Figure 2. The stereo view of the superimpose presentation of calculated coordinates. (a) Same coordinates of the left end ones of Figure 1. (b) The coordinates obtained by averaging four coordinates sets shown in Figure 1. The pairwise RMSD values were 0.15 (for (a)) and 0.50 (for (b)), respectively.

Chapter 8

Simulation of the Precision and Accuracy of the NMR Structure of a DNA Decamer by the Metric Matrix Distance Geometry Calculation

ABSTRACT

The accuracy and precision of a DNA decamer structure were investigated by the use of metric matrix distance geometry simulations. At the first step, the ^{1}H-^{1}H distances were taken from the three dimensional coordinates of a single conformer determined experimentally and regarded as a real structure. At the next step, the ensemble of the three dimensional coordinates of probable structures were calculated from the previously adopted distances by using the metric matrix distance geometry method. The input distance data varied in the range between the upper and lower limits. This range factor was important for the definition of the accuracy and precision of the structure determined by experimental NOEs. The simulation indicates the extent of over-determination of the structure of the DNA oligomer by NMR and the necessity of reliable indicators on the amount of structural informations. The author proposed some indicators.

INTRODUCTION

The structure calculated by NMR data of a DNA double strand is usually not well determined. In particular, there is no bound conditions for the overall shape or bend of the molecule. Then, it is believed that neither the distance geometry nor the ab initio simulated annealing method will produce unique structures. A previous paper (Pardi et al., 1988) reported that even with exact distances, it is not possible to uniquely define the conformation of DNA oligomer from the distance data available from NMR experiments. In their calculations, the data of many sugar protons which are usually unassigned in NMR experiments lacked. We reproduced the simulation of Pardi et al. (1988) including these proton data and resulted in a different conclusion.

METHODS

Since the accuracy of the structure is not known until the real structure is obtained, a simulation was performed as following. First, 1205 ^{1}H-^{1}H distances below 5 Å were taken from the three dimensional coordinates of the single conformer which was determined experimentally (chapter 7) and regarded as a real structure. Second, the ensemble of the three dimensional coordinates of a probable structure was calculated from the previously adopted distances by using the metric matrix distance geometry method (Havel and Wuthrich, 1984; Crippen and Havel, 1988). The poor sampling problem (Metzler et al., 1989) for the calculation was solved by the full randomized motorization (Kuszewski et al., 1992). The coordinates obtained by the distance geometry embedding were regularized by minimizing an effective energy term that represented all upper and lower bounds in the distance geometry matrix. The input distance data varied in precision depending on the range between the upper and lower limits.

RESULTS AND DISCUSSION

The left end coordinates set in Fig. 1 was obtained experimentally by using the direct NOE refinement procedure (chapter 7), regarded as a real structure. The input data for that calculation were 340 NOE peak volume values. The other four different coordinates sets in Fig. 1 were calculated by simulation with four input distances sets different in precision, i.e., the precision range was taken as the standard deviation of the left end sets, 10 % of distances, 0.5 Å and 1 Å for distances, respectively, from the second left to right. The RMSD values between pairwise coordinates sets and the RMS difference values from the assumed real structure were listed in Table 1.

The accuracy and precision of the structure were checked by the RMSD value, i.e., pairwise RMSD value as an indicator of precision and the RMS difference value as that of

accuracy. Both the precision and accuracy of the structure clearly depend on the quality of the input data, i.e., the increase of the quality of the input data leads to more precise and accurate structures. The pairwise RMSD value for the most well determined one was 0.47 for all atoms, and should be assumed as precise.

The pairwise RMSD values in Table 1 of the experimentally obtained coordinates and the most well determined ones from the simulation were 0.15 and 0.47, respectively. The fact indicates the extent of over-determination of the structure of the DNA oligomer by NMR. That is, the RMSD value of the experimentally obtained one is strongly influenced by the empirical energy minimization. So, this RMSD value 0.15 should be understood that the empirical energy minimization helps the averaging among the calculated coordinates derived from only the NMR data. It is also shown that the necessity of reliable indicators on the amount of structural informations or on the accuracy and precision of the determined structure. Here the author proposes some indicators, i.e., the completeness, $R_{accuracy}$ and $R_{precision}$ factors. In the following paragraphs, the definition and physical meaning of these indicators will be explained.

The completeness is defined by the ratio of the amount of the input information to that of the ideally available one as following.

$$\text{completeness} = N_{inp} / N_{all}, \qquad [1]$$

where N_{inp} is the number of input distances or peak volumes and N_{all} is the number of all the ^1H-^1H distances below 5 Å expected from the three dimensional coordinates experimentally determined. For the selection of the distances below 5 Å, each ^1H-^1H distance will be averaged among all distances that are calculated from coordinates. It is necessary to calculate together its standard deviation for the evaluation of other indicators, $R_{accuracy}$ and $R_{precision}$ factors. The coordinates were simulated from the previously adopted distances with upper and lower range given as its standard deviations by using the metric matrix distance geometry method. The $R_{accuracy}$ and $R_{precision}$ factors are defined as the RMS difference value of the simulated coordinates from the experimentally determined ones and the pairwise RMSD value among the simulated ones, respectively.

The completeness, $R_{accuracy}$ and $R_{precision}$ factors are representing the completeness of the input data, the reliability of the coordinates obtained by NMR and the convergence limit of the coordinates, respectively. The completeness factor can be calculated from the number of distances derived from a well determined coordinates set, whose pairwise RMSD convergence value should be less than 0.5 Å. In the present case the RMSD value of the experimentally determined ones is 0.15, thus the completeness value was calculated as 0.28 for the ratio of 340 to 1205, the number of the obtained to avairable distances, respectively. This value is low because of no input data for exchangeable and 5'/5" protons. A more quantitative procedure is necessary to include the data of exchangeable protons, and to separately observe the 5'/5" proton signals. The $R_{accuracy}$ and $R_{precision}$ factors for the best simulated coordinates were 0.59 and 0.47, respectively. These values indicate that the experimentally determined coordinates were precise but a little distorted ones including small systematic errors. This result would

come form relatively small number input data for the structure determination or might come from the rough single correlation time approximation of the relaxation matrix treatment in the step of coordinate determination.

The pairwise RMSD value, 0.15, for the experimentally determined coordinates should be understood as the precision of the averaged coordinates. The author proposed to call this kind RMSD of the precision of the averaged coordinates, the $R_{precision}(av)$ factor. This value should be distinguished from the $R_{precision}$ factor, and this difference indicates the extent of over-determination of the structure as discussed above.

In the previous report (Pardi et al., 1988), accuracy and precision were checked on the helical parameters of DNA double strands. However, the author believes that both the accuracy and precision of the structures can not be checked by the particular structural parameters, but should be evaluated by the completeness, $R_{accuracy}$, $R_{precision}$ and $R_{precision}(av)$ factors. The particular structural parameters might not directly reflect input informations and could not be applied for the case of a irregular conformation, such as a triple helix.

Here, the accuracy and precision of a DNA decamer structure were investigated by the use of the metric matrix distance geometry simulations. The quality of the input distance information is important to define the accuracy and precision of the calculated structure. The accuracy and precision of the structure derived from NMR data should be checked by the completeness, $R_{accuracy}$, $R_{precision}$ and $R_{precision}(av)$ factors.

Table 1. Pairwise RMSDs and RMS differences among simulated coordinates shown in Figure 1

	Target	SD	10%	0.5Å	1Å
Pairwise RMSD	0.15[a]	0.47[b]	1.10	1.22	1.36
RMS Difference*	target	0.59[c]	1.46	1.58	1.95

* RMS difference from the target coordinates.

[a] Rprecision(av), [b] Rprecision, [c] Raccuracy.

Figure 1

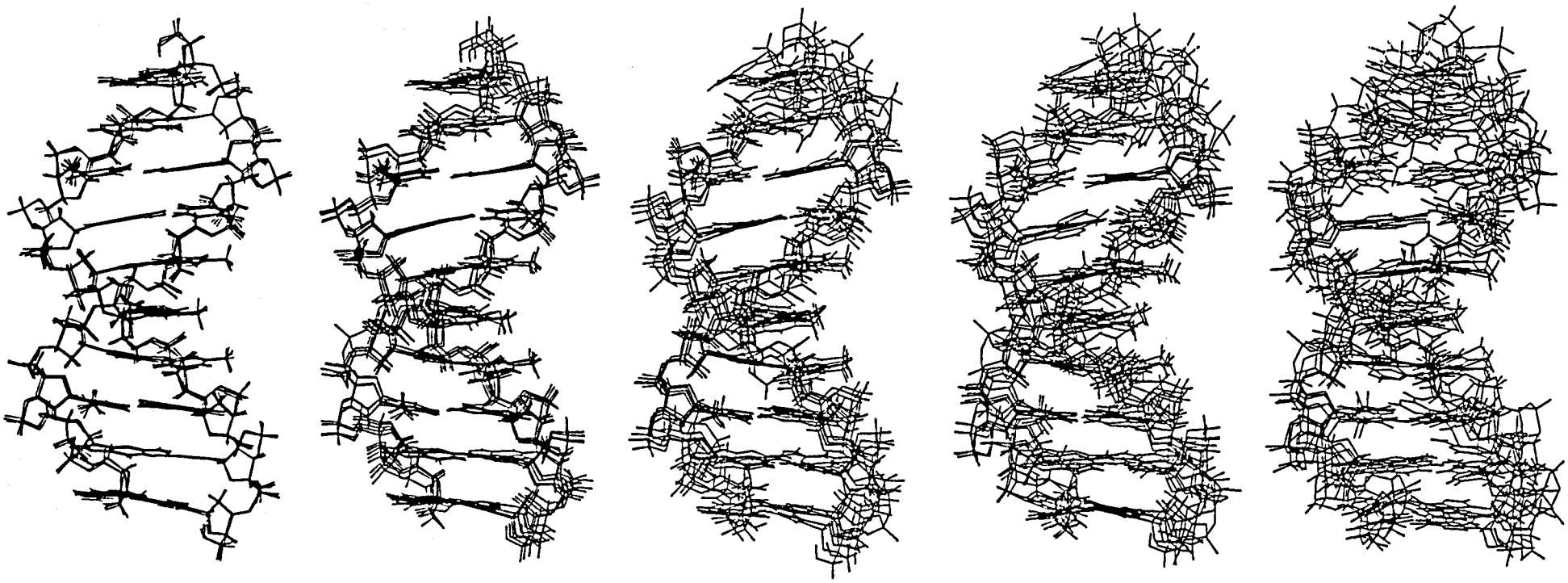


Figure 1. Left end coordinates set was obtained experimentally by using the direct NOE refinement procedure (chapter 7) which is regarded as a real structure in this chapter. The other four different coordinates sets were calculated by the simulation with four different input distances sets, i.e., the most precise coordinates set in the standard deviation range of the left end ones, 10 % of distances, 0.5 Å and 1 Å for distances, from the second left to right respectively. The RMSD values among coordinates sets and RMS difference values from the assumed real structure were listed in Table 1.

Chapter 9

NMR Analysis of the Sugar Conformation of a DNA Decamer by (2'R)-[2'-^{2}H]- and (2'S)-[2'-^{2}H]-Labeling

ABSTRACT

The highly stereoselective (2'R)-[2'-^{2}H]- and (2'S)-[2'-^{2}H]- labeled DNA decamer were synthesized for the structural or dynamical study of the DNA decamer. These stereoselective [2'-^{2}H]-labeling made possible explicit stereospecific assignments and exact determination of the vicinal coupling constants, $^3J_{HH}$. Some of these values are a little larger than those previously obtained. Moreover, the longitudinal relaxation times of the 1' proton resonances increased for both ^{2}H labeled compounds.

INTRODUCTION

Previously determined three dimensional structures of DNA by NMR were incomplete in spite of many efforts for rationalization, because of little treatment of the internal and global motions such as twisting, bending, breathing and so on. For the precise structure determination, it is necessary to describe these motions by conformational changes at several time scales. James and coworkers studied the sugar puckering of the DNA oligomer by using the $^3J_{HH}$ value, and described its motion as the population of the conformations (Schmitz et al., 1991). However, these studies require the exact determination of the $^3J_{HH}$ value and explicit stereospecific assignments of the sugar 2'R and 2'S proton resonances. For this purpose, the most reliable procedure is the stereospecific deuterium labeling at 2' position (Kawashima et al., 1993; Kawashima et al., 1994). Here, the (2'R)-[2'-^{2}H]- and (2'S)-[2'-^{2}H]-labeled DNA decamers were synthesized, and by using them, the explicit stereospecific assignments and the exact determination of the $^3J_{HH}$ value were achieved.

MATERIALS

The (2'R)-[2'-^{2}H]- and (2'S)-[2'-^{2}H]- 2'-deoxyribonucleotides were synthesized by the previously reported procedure (Kawashima et al., 1993; Kawashima et al., 1994). The (2'R)-[2'-^{2}H]- and (2'S)-[2'-^{2}H]- labeled DNA decamers were synthesized by a DNA synthesizer. The 2'-deuterium labeling is so highly stereoselective that the explicit stereospecific assignments of the sugar 2'R / 2'S proton resonances were easily achieved.

RESULTS AND DISCUSSION

Two-dimensional DQF-COSY spectra of the non-labeled, (2'R)-[2'-^{2}H]-labeled, and (2'S)-[2'-^{2}H]-labeled decamers were shown, in Fig. 1a., in Fig. 1b and in Fig. 1c, respectively. These spectra showed some expected or unexpected results caused by the deuterium labeling at 2' position. First, the spin system were simplified and apparent linewidths were also reduced. The actual longitudinal relaxation time which was determined by the inversion recovery method was lengthened about 1.5 times for the 1' proton. Second, some cross peaks which were not observed for the non-labeled compound could be observed for the deuterium labeled ones, such as the cross peaks between 3' and 2'R proton resonances. Third, phase distortions of some cross peaks which were induced from the strong coupling condition were corrected to pure absorption line shapes.

This deuterium labeling also introduced the exact determination of the $^3J_{HH}$ constants between the sugar 1' and 2'S or 2'R protons, shown in Tables 1 and 2. For the well separated 1' proton resonances (half of all), we could determine the $^3J_{HH}$ constants in the one dimensional spectrum with precision of 0.01 Hz. For the remaining ones indicated by asterisks in Tables 1

and 2, the DQF-COSY spectrum helped the determination of the vicinal coupling constants. Direct reading of the ^3J values from the DQF-COSY spectrum caused some larger coupling constants from errors (Neuhaus et al., 1985), and thus two additional procedures were performed, i.e., fitting and a simple simulation method with precision of 0.5 Hz. In the fitting, some one dimensional slices of the DQF-COSY spectrum about individual cross peaks were optimized by the combination of the positive and negative absorption lorentzian (FELIX, Biosym Technologies). The simulation was performed by employing the peak center values of the anti phase dispersive line shapes (Kim and Prestegard, 1989). The E.COSY spectrum (Griesinger et al., 1985; Griesinger and Ernst, 1987) was taken for measuring coupling constants of the non-labeled DNA decamer. Some serious overlapping inhibited the determination of several ^3J values, however 70 % of the coupling values between the 1' and 2'R or 2'S protons were determined. They are also shown in Tables 1 and 2.

Comparing the determined ^3J values, almost all the values of the non-labeled decamer derived from the E.COSY experiment are smaller than those of ^2H labeled compounds. The coupling constants determined from E.COSY should give larger value by the cancellation of the positive and negative peak (Neuhaus et al., 1985). Therefore present result clearly showed that the coupling constants derived from the non-labeled and ^2H labeled decamers were different. In fact, some $^3J_{H1'H2'}$ values of the ^2H labeled compound were bigger than the maximum values estimated from the previous empirical Karplus type parameters (Rinkel and Altona, 1987). The difference might be explained by two ways. First, the ^2H labeled compound might have different parameters for the Karplus type vicinal coupling constants. This will be proved by the collection of the many data sets of the J values which consist of the 1' / 2'R (2'S), 2'R (2'S) / 3' and 3' / 4' proton-proton vicinal couplings for the ^2H labeled oligomers. Only the present data can not check the validity. Second, the interference between J coupling and cross relaxation might be observed (Harbison, 1993). This will be more likely for the DNA oligomer systems because the 2'R and 2'S protons are closely located (1.75 Å). These two ways of the explanation should be proved before the use of the observed J coupling constants for the determination of the sugar puckering.

Table 1. The $^{3}J_{H1'H2'}$ values determined for non-labeled and 2'R-^{2}H-labeled DNA decamers

Method	1G	2C	3A	4T	5T	6A	7A	8T	9G	10C
1D	11.38	11.6*	11.70	11.7*	11.79	12.44	11.04	10.6*	11.9*	9.09
DQF-COSY										
direct	11.2	11.9	11.9	11.2	12.1	12.5	11.9	11.4	11.4	9.3
fitting	10.4	11.3	11.4	10.4	11.2	11.9	10.9	10.4	10.8	8.5
simulate	10.9	11.6	11.6	10.8	11.7	12.1	11.4	10.9	11.1	9.0
E.COSY	9.3	9.8	9.8	9.0	11.0		10.3	9.3	9.5	

Table 2. The $^{3}J_{H1'H2''}$ values determined for non-labeled and 2'S-^{2}H-labeled DNA decamers

Method	1G	2C	3A	4T	5T	6A	7A	8T	9G	10C
1D	7.9*	6.48	6.74	8.1*	7.6*	7.9*	6.94	7.8*	7.0*	8.59
DQF-COSY										
direct	8.1	8.4	8.4	9.0	8.7	8.6	9.0	9.2	8.1	9.0
fitting	7.2	7.3	7.0	7.6	5.2	6.8	7.0	8.0	7.2	8.1
simulate	7.3	7.7	7.6	8.2	8.0	7.7	7.9	8.4	7.3	8.3
E.COSY			6.7	6.8		6.8	6.9	6.8	6.3	

Figure 1a

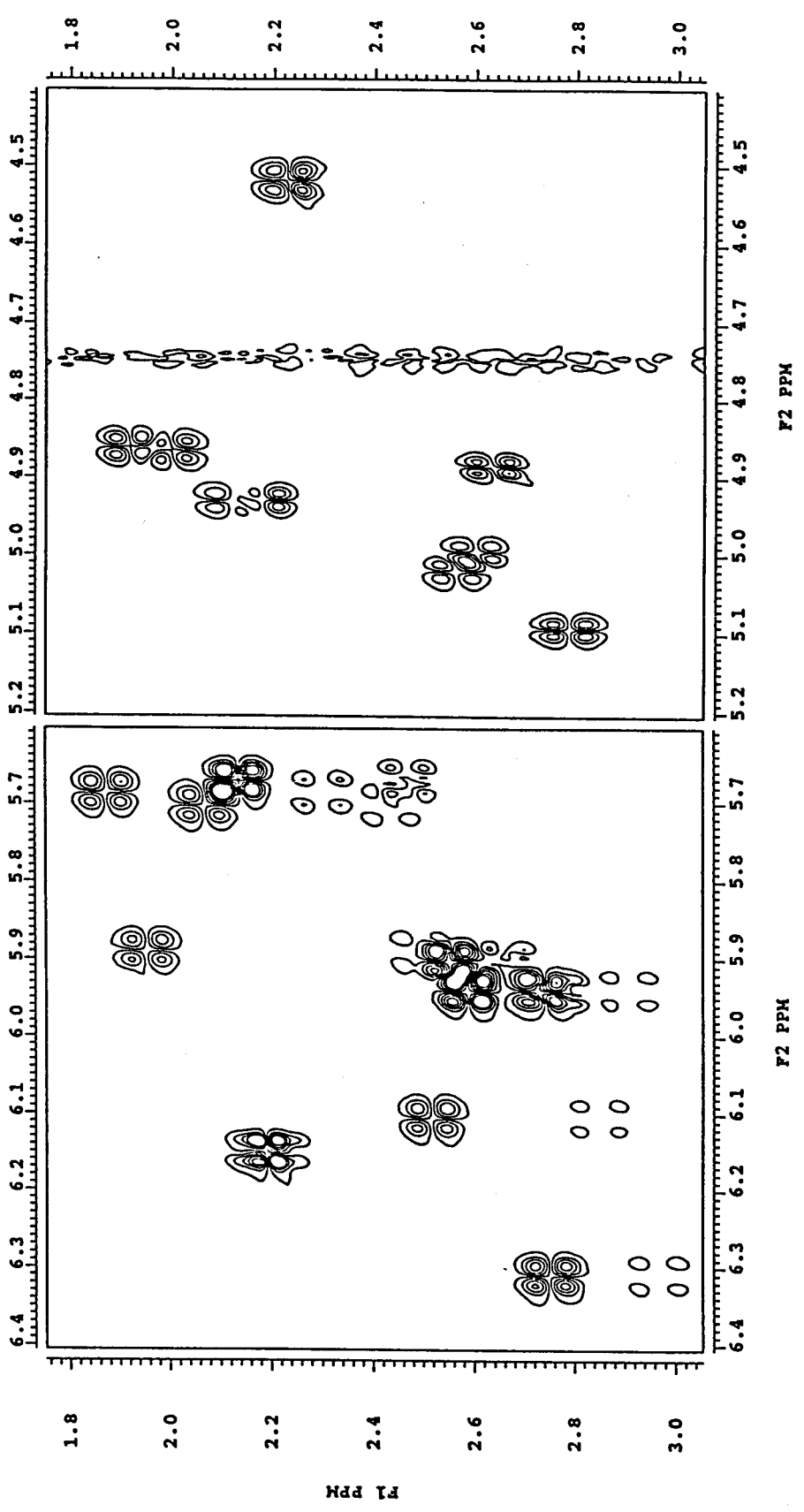


Figure 1b

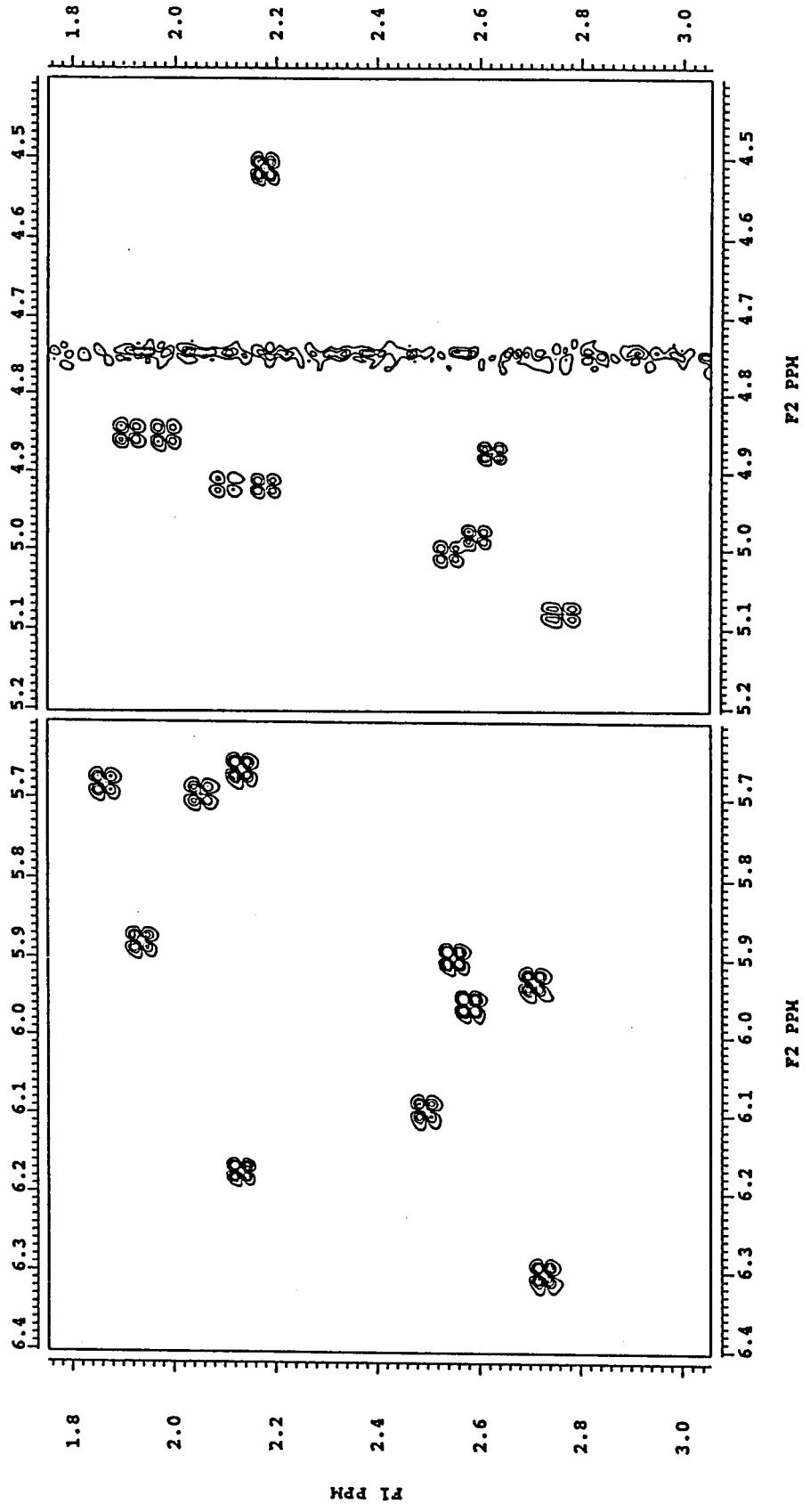

F1 PPM
F2 PPM
F2 PPM
1.8
2.0
2.2
2.4
2.6
2.8
3.0
4.5
4.6
4.7
4.8
4.9
5.0
5.1
5.2
5.7
5.8
5.9
6.0
6.1
6.2
6.3
6.4

Figure 1c

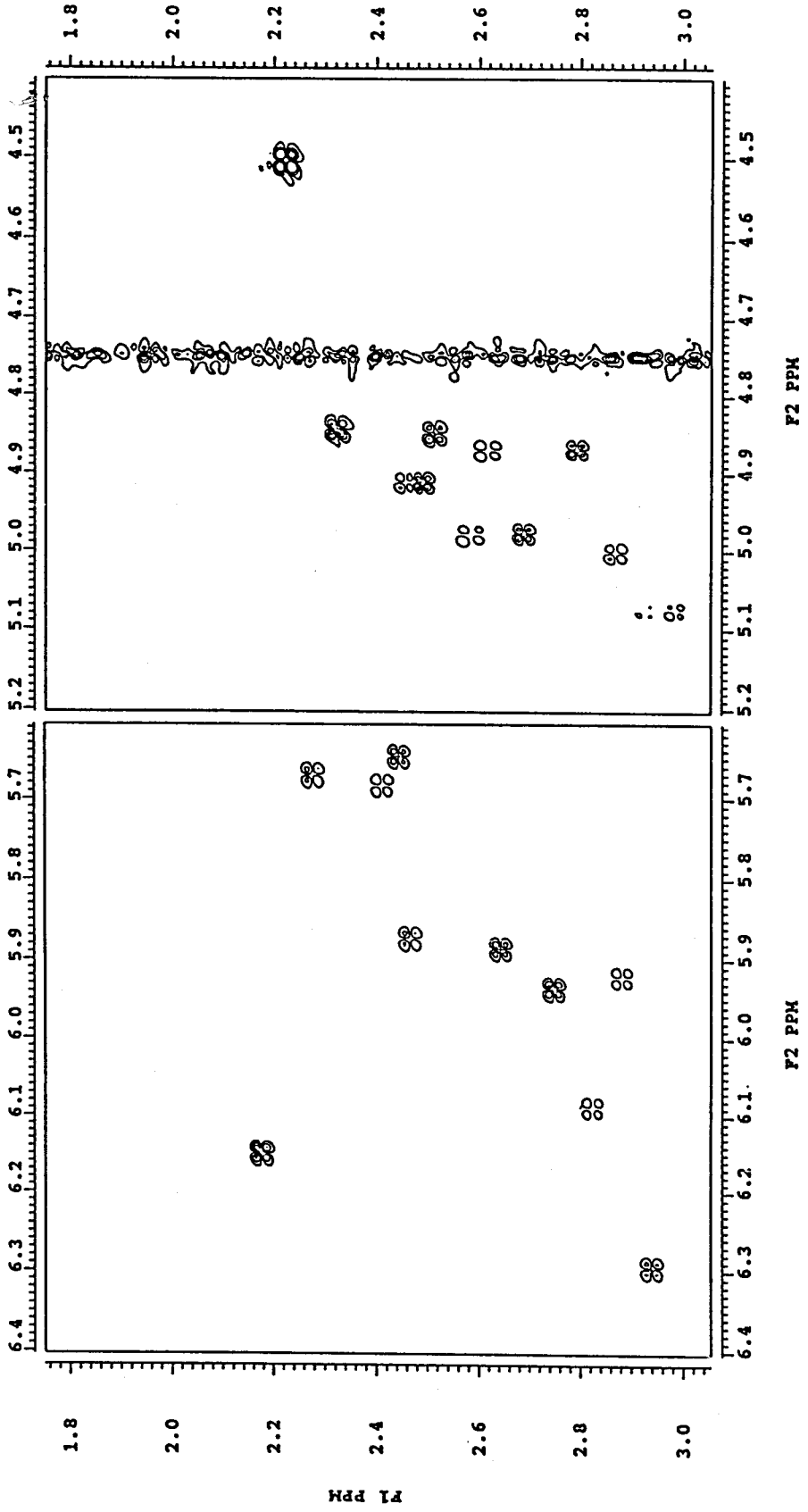


Figure 1. Two dimensional DQF-COSY spectra of (a) the non-labeled, (b) the (2'R)-[2'-^{2}H]-labeled and (c) the (2'S)-[2'-^{2}H]-labeled DNA decamers. The spectra were recorded on a Bruker ARX 500 spectrometer with 512 hyper complex points for t1 and 512 complex points for t2, multiplying the π / 4 shifted sine bell window functions for both dimensions, and zero filled to 1024 real points for both dimensions. The sample solution contained 3 mM of each duplex with the 20 mM phosphate buffer and 50 mM NaCl at neutral pH.

General Discussion and Conclusion

Finally, the author discusses here the possibility and limitation of the nuclear magnetic resonance technique as a tool for determining the structure and dynamics of DNA oligomers in solution, and gives the conclusive evidence for the structural feature of the DNA oligomers treated throughout this thesis.

Structure determination of DNA oligomers by NMR

The structure of nucleic acid is difficult to determine the structure in solution by NMR from several reasons. The most important and inherent reason is the low proton density. In the case of a protein, its density is nearly 70 protons per one thousand of molecular mass, but for DNA, 35 protons. So the number of the ^{1}H-^{1}H distances is extremely small for the DNA molecule compared with the protein which is now well studied by NMR. The construction of the three dimensional coordinates by NMR is governed by the number of the input ^{1}H-^{1}H distances below 5 Å. This density makes the limitation of NMR for the structure construction of DNA (see chapter 8).

All the other difficulties, the necessity of precise NOE and ^{3}J values, the consideration of the strong coupling, narrow chemical shift dispersion and short T2 of the sugar protons, and so on., can be overcome. The most of the above difficulties have been discussed in this thesis. For the measurement of the precise values of NOEs and vicinal coupling constants, more technical improvements and careful treatments of the data were required (see chapters 5 ~ 7, 9). Although the problems of the strong coupling and short T2 were not treated extensively in this thesis, the ^{2}H labeling will be helpful to solve the difficulty (see chapter 9). The problem of the narrow chemical shift dispersion could be partly overcome by picking up specific resonance peaks (chapters 1 ~ 3), but the essential improvement may be achieved by other techniques, for example, the use of ^{13}C and ^{15}N labeling. In fact, these stable isotope labeling works are now in progress in the groups of Osaka and Tokyo Metropolitan Universities, and the wide chemical shift dispersion becomes available by the additional ^{13}C dimension.

It is also important to solve the so-called multi conformer problem for the structure determination. Previously reported three dimensional structures of DNA oligomers by NMR (see Wijmenga et al., 1993 for review) were incomplete in this sense (Lane, 1990). Under the multi conformer condition, the structure can not be described by a single coordinate group of alike conformers (Schmitz et al., 1993; Kojima et al., 1993). However, all the methods for the determination of the solution conformer, NMR, CD and so on, clearly show that the population of minor conformers is relatively low. Thus, the main conformation should be determined by NMR neglecting the minor conformers. Then, what is a problem? If a DNA oligomer takes only a single rigid conformation, there is no problem, but it is unlikely. At least, the flexibility of the molecule should be investigated by several dynamics studies.

Dynamics study of DNA oligomers revealed in NMR

For proper understanding of the measurable NMR data, the knowledges about motions or dynamics are required, i.e., the mode of motion, the time order or life time, populations (if definable) and so on. In this thesis, these types of data were not included enough, but the

contribution of dynamics of DNA oligomers was partly discussed in the refined structural studies (chapters 4, 6 and 7). The precise analyses of the structure generally require the information about dynamics. For example, the distances between protons exchanging with water can be determined only when the exchange rate with water and the molecular rotational motion rate are known.

The author believes that one of the most characteristic property of NMR is the dependency of the specific dynamics at the local site. The global rotational motion can be observed by many methods, viz., fluorescence depolarization, dynamic laser light scattering, dielectric relaxation, and so on. However the site specific local motions may not be observed by any other method except for NMR.

Structural feature of DNA oligomers

The structure of the DNA oligomer was determined uniquely by NMR as empirical energy minimized coordinates or averaged coordinates (chapter 7). The distance informations of the maximum quality gave coordinates with a single alike conformer by the metric matrix distance geometry (mmDG) simulation (chapter 8). This result shows that the mmDG simulation should be employed to indicate the reliability of the obtained coordinates.

However, the structure of the DNA oligomer is not rigid (chapters 4 and 6). The ^{2}H labeled DNA oligomers were used to analyze it in detail (chapter 9). Although the dynamics were not directly analyzed in this thesis, the beginning of the next step for understanding the dynamical behavior of DNA oligomers was shown.

Closing remarks

I love NMR. Don't you have a try?

REFERENCES

Arnott,S. and Hukins,D.W. (1972). *Biochem.Biophys.Res.Commun.*, **47**, 1504-1509.

Arnott,S. and Selsing,E. (1974). *J.Mol.Biol.*, **88**, 509-521.

Arnott,S., Chandrasekaran,R., Hall,I.H. and Puigjaner,L.C. (1983). *Nucleic Acids Res.*, **11**, 4141-4155.

Ashcroft,J., Live,D.H., Patel,D.J. and Cowburn,D. (1991). *Biopolymers* **31**, 45-55.

Bax, A., Freeman, R., Frenkiel, T.A. and Levitt, M.H. (1981). *J. Magn. Reson.* **43**, 478-483.

Braunshweiler,L. and ERNST,R.R. (1983). *J. Magn. Reson.* **53**, 521.

Brunger,A.T. (1992). "X-PLOR Version 3.1 manual", Yale University Press.

Carr,H.Y. and Purcell,E.M. (1954). *Phys. Rev.* **94**, 630.

Calladine,C.R. (1982). *J.Mol.Biol.* **161**, 343.

Chazin,W.J., Wuthrich,K., Hyberts,S., Rance,M., Denny,W.A. and Leupin,W. (1986). *J. Mol. Biol.*, **190**, 439-453.

Crippen,G. and Havel,T. (1988). "Distance Geometry and Molecular Conformation", Research Studies Press: Taunton, Somerset, England.

Ernst,R.E., Bodenhausen,G. and Wokaun,A. (1987). "Principles of Nuclear Magnetic Resonance in One and Two Dimensions", Clarendon, Oxford.

Farrar,T.C. and Becker,E.D. (1971). "Pulse and Fourier Transform NMR", Academic Press.

Feigon, J., Sklenar, V., Wang, E., Gilbert, D. E., Macaya, R. F. and Schultze, P. (1992). *Method. Enzymol.* **211**, 235-253.

Freeman, R. (1988). "A Handbook of Nuclear Magnetic Resonance", Longman, London.

Giniger,E., Varnum,M.S. and Ptashne,M. (1985). *Cell* **40**, 767.

Goodsell,D.S., Grzeskowiak,M.K-. and Dickerson,R.E. (1994). *J.Mol.Biol.* **239**, 79-96.

Griesinger,C., Sorensen,O.W. and Ernst,R.R. (1985). *J.Am.Chem.Soc.* **107**, 6394-.

Griesinger,C. and Ernst,R.R. (1987). *J.Magn.Reson.* **75**, 261-.

Grzesiek,S. and Bax,A. (1993). *J.Am.Chem.Soc.* **115**, 12593-12594.

Gueron,M., Kochoyan,M. and Leroy,J.L. (1987). *Nature* **328**, 89-92.

Hahn, E.L. (1950). *Phys. Rev.*, **80**, 580-594.

Hanssen,A., Hoy,M.V. and Kodadek,T. (1991). *Biochem. Biophys. Res. Commun.*, **175**, 492-499.

Harbison,G.S. (1993). *J.Am.Chem.Soc.* **115**, 3026-3027.

Hare,D.R., Wemmer,D.E., Chou,S.-H., Drobny,G. and Reid,B.R. (1983). *J.Mol.Biol.* **171**, 319-336.

Havel, T.F. and Wuthrich, K. (1984). *Bull.Math.Bio.* **46**, 673-698.

Hodge,R.P., Brush,C.K., Harris,C.M. and Harris,T.M. (1991). *J.Org.Chem.* **56**, 1553-1564.

Hore,P.J. (1989). *Methods in Enzymology* **176**, 64.

Hosure,R.V., Govil,G. and Miles,H.T. (1988). *Magn.Reson.Chem.* **26**, 927-944.

Hurd, R.E., John, B.K., Webb, P. and Plant, D. (1992). *J. Magn. Reson.* **99**, 632-637.

James,T.L., Gochin,M., Kerwood,D.J., Schmitz,U. and Thomas,P.D. (1991). "Computational Aspects of the Study of Biological Macromolecules by NMR" (J.C. Hoch, ed.), Plenum Press, New York.

Jeener,J., Meier,B.H., Backman,P. and Ernst,R.R. (1979). *J.Chem.Phys.*, **71**, 4546-4553.

Kaptein,R., Zuiderweg,E.R.P., Scheek,R.M., Boelens,R. and van Gunsteren,W.F. (1985). *J.Mol.Biol.* **182**, 179-182.

Kawashima,E., Sekine,T., Miyahara,M., Aoyama,Y., Hirao,I., Kainosho,M., Kyogoku,Y. and Ishido,Y. (1992). *Nucleic Acids Res., Sym. Ser.*, **27**, 81-82.

Kawashima,E., Aoyama,Y., Sekine,T., Nakamura,E., Kainosho,M., Kyogoku,Y. and Ishido,Y. (1993). *Tetrahedron letters* **34**, 1317-1320.

Kawashima,E., Toyama,K., Sekine,T., Ohshima,K., Kainosho,M., Ono,A., Kyogoku,Y., Kojima,C. and Ishido,Y. (1994). *Nucleic Acids Research, Sym.Ser.* **31**, 41-42.

Kawashima,E., Sekine,T., Miyahara,M., Aoyama,Y., Kainosho,M., Ono,A., Kyogoku,Y., Kojima,C. and Ishido,Y. (1994). *Nucleic Acids Research, Sym.Ser.* **31**, 43-44.

Kay,L.E., Xu,G.Y. and Yamazaki,T. (1994). *J.Magn.Reson.Ser.A* **109**, 129-133.

Keepers,J.W. and James,T.L. (1984). *J.Magn.Reon.* **57**, 404-426.

Kessler,H., Gehrke,M. and Griesinger,C. (1988). *Angew.Chem.Int.Ed.Engl.* **27**, 490.

Kim,Y. and Prestegard,J.H. (1989). *J.Magn.Reson.* **84**, 9-13.

Kochoyan,M., Leroy,J.L. and Gueron,M. (1987). *J.Mol.Biol.* **196**, 599.

Kojima,C., Ohkubo,T. and Kyogoku,Y. (1991). *Nucleic Acids Res., Sym. Ser.*, **25**, 177-178.

Kojima,C., Kyogoku,Y., Ishido,Y, Kawashima,E., Sekine,T., Hirao,I. and Kainosho,M. (1993). *Nucleic Acids Res., Sym.Ser.* **29**, 185-186.

Kojima,C. and Kyogoku,Y. (1993). *J. Magn. Reson.Ser.B* **102**, 214-217.

Kojima,C. and Kyogoku,Y. (1994). *J.Biomol.NMR* **4**, 181-191.

Kojima,C. and Kyogoku,Y. (1994). *Nucleic Acids Research, Sym.Ser.* **31**, 171-172.

Kuszewski,J., Nilges,M. and Brunger,A.T. (1992). *J.Biomol.NMR* **2**, 33-56.

Lane,A.N. (1990). *Biochim. Biophys. Acta* **1049**, 189-204.

Lane, A.N., Jenkins, T.C., Brown, T. and Neidle, S. (1991). *Biochemistry* **30**, 1372-1385.

Leroy,J.L., Kochoyan,M., Huynh-Dinh,T. and Gueron,M. (1988). *J.Mol.Biol.* **200**, 223-238.
Leroy,J.L., Gao,X., Gueron,M. and Patel,D.J. (1991). *Biochemistry* **30**, 5653-5661.
Leupin,W., Wagner,G., Denny,W. and Wuthrich,K. (1987). *Nucl.Acids Res.* **15**, 267-275.
Lipari,G. and Szabo,A. (1982). *J.Am.Chem.Soc.* **104**, 4546-4559.
Macura,S. and Ernst,R.R. (1980). *Molecular Physics*, **41**, 95-117.
Marion,D., Kay,L.E., Spark,S.W., Torchia,D.A. and Bax,A. (1989). *J.Am.Chem.Soc.* **111**, 1515.
Mccammon,J.A. and Harvey,S.C. (1987). "Dynamics of Proteins and Nucleic Acids", Cambridge University Press.
McConnell,B. (1986). *J.Biomol.Struct.Dynam.* **4**, 419-436.
Messerle,B.A., Wider,G., Otting,G., Weber,C. and Wuthrich,K. (1989). *J.Magn.Reson.* **85**, 608-613.
Muller,N., Ernst,R.R. and Wuthrich,K. (1986). *J.Am.Chem.Soc.* **108**, 6482-6492.
Nagayama,K., Yamazaki,T., Yoshida,M., Kanaya,S. and Nakamura,H. (1990). *J.Biochem.* **108**, 149-152.
Neuhaus,D., Wagner,G., Vasak,M., Kagi,J.H.R. and Wuthrich,K. (1985). *Eur.J.Biochem.* **151**, 257-
Neuhaus,D. and Williamson,M.P. (1989). "The nuclear overhauser effect in structural and conformational analysis", VCH, New York.
Nibedita,R., Ajay Kumar,R., Majumdar,A., Hosur,R.V. and Govil,G. (1993). *Biochemistry* **32**, 9053-9064.
Nilges,M., Clore,G.M. and Gronenborn,A.M. (1988). *FEBS Lett.* **229**, 317-324.
Nilges,M., Clore,G.M. and Gronenborn,A.M. (1988). *FEBS Lett.* **239**, 129-136.
Nilges,M., Habazettl,J., Brunger,A.T. and Holak,T.A. (1991). *J.Mol.Biol.* **219**, 499-510.
Norwood,T.J., Boyd,J., Heritage,J.E., Soffe,N. and Campbell,I.D. (1990). *J.Magn.Reson.* **87**, 488-501.
Oschkinat,H., Griesinger,P.J., Kraulis,P.J., Sorensen,R.R., Ernst,R.E., Gronenborn,A.M. and Clore,G.M. (1988). *Nature* **332**, 374.
Pardi,A., Walker,H., Rapoport,G., Wider,G. and Wuthrich,K. (1983). *J.Am.Chem.Soc.* **105**, 1652.
Pardi,A., Hare,D.R. and Wang,C. (1988). *Proc.Natl.Acad.Sci.USA* **85**, 8785-8789.
Pelczer,I., Bishop,K.D., Levy,G.C. and Borer,P.N. (1991). *J.Magn.Reson.* **91**, 604-606.
Piantini,U., Sorensen,O.W. and Ernst,R.R. (1982). *J.Am.Chem.Soc.*, **104**, 6800-6801.
Piotto,M.E. and Gorenstein,D.G. (1991). *J.Am.Chem.Soc.* **113**, 1438-1440.
Quignard,E., Teoule,R., GuY,A. and Fazakerley,V. (1985). *Nucl.Acids Res.* **13**, 7829.
Quintana,J.R., Grzeskowiak,K., Yanagi,K. and Dickerson,R.E. (1992). *J.Mol.Biol.* **225**, 379-395.
Rance,M., Sorensen,O.W., Bodenhausen,G., Wagner,G., Ernst,R.R. and Wuthrich,K. (1983). *Biochem.Biophys.Res.Commun.* **117**, 479-485.
Raszka,M. and Kaplan,N.O. (1972). *Proc.Nat.Acad.Sci.USA* **69**, 2025-2029.
Rinkel,L.J. and Altona,C. (1987). *J.Biomol.Struct.Dyns.* **4**, 621-649.
Saenger,W. (1984). "Principles of Nucleic Acid Structure", Springer-Verlag New York Inc.
Schmitz,U., Pearlman,D.A. and James,T.L. (1991). *J.Mol.Biol.*, **221**, 271-292.
Schmitz,U., Nikolai,B.U., Kumar,A. and James,T.L. (1993). *J.Mol.Biol.* **234**, 373-389.
Shaka,A.J. and Freeman,R. (1983). *J.Magn.Reson.* **51**, 169-173.
Shoup,R.R., Miles,H.T. and Becker,E.D. (1972). *J.Chem.Phys.* **76**, 64-70.
Sklenar,V., Brooks,B.R., Zon,G. and Bax,A. (1987). *FEBS Lett.* **216**, 249-252.
Sklenar,V. and Feigon,J. (1988). *J.Am.Chem.Soc.* **112**, 5644.
Sklenar,V. and Feigon,J. (1990). *Nature* **345**, 836.
Sklenar,V. and Feigon,J. (1990). *J.Am.Chem.Soc.* **112**, 5644-5645.
Solomon,I. (1955). *Phys.Rev.* **99**, 559-565.
Sorensen,O.W., Levitt,M.H. and Ernst,R.R. (1983). *J.Magn.Reson.* **55**, 104-113.
Sorensen,O.W., Eich,G.W., Levitt,M.H., Bodenhausen,G. and Ernst,R.R. (1983). *Prog.Nucl.Magn.Reson.Spectrosc.* **16**, 163.
Spera,S., Ikura,M. and Bax,A. (1991). *J.Biomol.NMR* **1**, 155-165.
Stonehouse,J., Shaw,G.L., Keeler,J. and Laue,E.D. (1994). *J.Magn.Reson.Ser A* **107**, 178-184.
Stout,G.H. and Jensen,L.H. (1989). "X-ray Structure Determination", Wiley, New York.
Struhl,K. (1989). *Annu.Rev.Biochem.* **58**, 1051.
Tarjan,R. (1983). "Data Structures and Network Algorithms", Society for Industrial and Applied Mathematics, Philadelphia.
Teitelbaum,H. and Englander,S.W. (1975). *J.Mol.Biol.* **92**, 79-92.
van de Ven,F.J.M., Haasnoot,C.A.G. and Hilbers,C.W. (1985). *J.Magn.Reson.*, **61**, 181-187.
van de Ven,F.J.M. and Hilbers,C.W. (1988). *Eur J.Biochem.* **178**, 1-38.
Vuister,G.W., Boelens,R., Padilla,A., Kleywegt,G.J. and Kaptein,R. (1990). *Biochemistry* **29**, 1829.
Wang,K.Y., Borer,P.N., Levy,G.C. and Pelczer,I. (1992). *J.Magn.Reson.* **96**, 165-170.
Watson,J.D. and Crich,F.H.C. (1953). *Nature* **171**, 737-738.
Wijmenga,S.S., Mooren,M.M.W. and Hilbers,C.W. (1993). In Roberts,G.C.K. (ed.), "NMR of Macromolecules - A Practical Approach". IRL Press, Oxford, 217-288.
Wuthrich,K. (1986). "NMR of Proteins and Nucleic Acids", Wiley, New York.
Yip, P. and Case, D. A. (1989). *J.Magn.Reson.* **83**, 643-648.
Ziji,P.C.M.V. and Moonen,C.T.W. (1990). *J.Magn.Reson.* **87**,18.

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 Prof. Y. Kyogoku for his cordial guidances, giving me good experimental environment, profitable discussion and intimate encouragement throughout the course of my study in spite of my irregular way of life.

The author also wishes to express his sincere thanks to Mr. Takashi Kodama, Institute for Protein Research, Osaka University, for teaching me physicochemical concepts of biochemistry and salvaging my broken mental life through continuing discussions and comments spending his precious time with his friendship. Grateful thanks are due to some friends and lovers for helping the completeness of my graduate student life.

The author wishes to express his cordial appreciation to all colleagues, Drs Akira Ono, Etsuko Kawashima, Takeshi Sekine, Keizo Toyama, Ichiro Hirao, Toshio Shida, Tadayasu Ohkubo, Katsuhiko Inoue, Shigenari Iwai, Tomoko Nishizaki, Yasuo Komatsu, Nobuaki Suzuki, Yasuharu Serikawa, Yoshiyuki Tanaka, Hidehito Tochio, Akemi Kawamura and Takashi Kodama, for their generous supply of synthetic oligonucleotides and letting me use their facilities, their enthusiastic discussions or comments with encouragements. During the course of the study the author has received a number of fruitful discussions, helpful comments and hearty encouragements from several persons, Drs Toshio Yamazaki, Shigenori Nishimura, Shuichi Hiroaki, Ken-ichi Uchida, Masato Katahira, Shin-ichi Tate, Toshimichi Fujiwara, Hideo Akutsu, Rieko Ishima, Istvan Pelczer, Markus R. Walchli, Ritsuko Sakaguchi-Katahira, Kimiko Yamamoto-Nakao, Nobuaki Nemoto, Takuya Yoshida, Takeyuki Tanaka, Hiroshi Matsuo, Haruhiko Tamaoki, Takayuki Odahara, Kazuo Yamamoto, Masato Shimizu, Koichi Uegaki, Yuji Kobayashi, Masahiro Shirakawa and Hiromu Sugeta. The author wishes to express his cordial appreciation to all (previous and now) the member of Division of Molecular Biophysics, Institute for Protein Research, Osaka University. His acknowledgments are also due to Professor Masatsune Kainosho, Tokyo Metropolitan University, Professor Yoshiharu Ishido, Tokyo College of Pharmacy, Professor Eiko Ohtsuka, Hokkaido University, and all the member of their groups. This work was supported financially by the Japan Society for the Promotion of Science for Young Scientists.

Finally, the author wishes to express his sincere gratitude to his parents for everything they did for their son with endless love and to his family for their unfailing understanding and affectionate encouragements.

Chojiro KOJIMA

Spring, 1995

List of Publications

1. Kojima,C., Ohkubo,T. and Kyogoku,Y. "Structure analyses of DNA oligomers by use of heteronuclear 2D NMR", (1991). *Nucleic Acids Research, Sym.Ser.*, **25**, 177-178.

2. Shirakawa,M., Matsuo,H., Serikawa,Y., Suzuki,N., Kojima,C., Ohkubo,T. and Kyogoku,Y. "Conformational changes of proteins and DNA induced by their interactions", (1992). *Molecular Structure and Life,* 129-140.

3. Kojima,C. and Kyogoku,Y. "HAL experiment, a novel filtering method and its application to DNA assignment", (1993). *J.Magn.Reson.Ser.B* **102**, 214-217.

4. Kojima,C., Kyogoku,Y., Ishido,Y., Kawashima,E., Sekine,T., Hirao,I. and Kainosho,M. "Precise analyses of DNA structure by NMR", (1993). *Nucleic Acids Research, Sym.Ser.* **29**, 185-186.

5. Kojima,C. and Kyogoku,Y. "Filtering methods for selection of singlet and doublet signals in NMR spectra of DNA oligomers", (1994). *J.Biomol.NMR* **4**, 181-191.

6. Kojima,C. and Kyogoku,Y. "Precise structure determination of DNA oligomer by NMR", (1994). *Nucleic Acids Research, Sym.Ser.* **31**, 171-172.

7. Kawashima,E., Toyama,K., Sekine,T., Ohshima,K., Kainosho,M., Ono,A., Kyogoku,Y., Kojima,C. and Ishido,Y. "Novel synthesis of [5'-^{2}H]-2'-deoxyribonucleoside derivatives and synthesis of their oligonucleotide", (1994). *Nucleic Acids Research, Sym.Ser.* **31**, 41-42.

8. Kawashima,E., Sekine,T., Miyahara,M., Aoyama,Y., Kainosho,M., Ono,A., Kyogoku,Y., Kojima,C. and Ishido,Y. "Highly diastereoselective synthesis of (2'S)-[2'-^{2}H]-2'-deoxyribonucleosides and synthesis of their oligonucleotides", (1994). *Nucleic Acids Research, Sym.Ser.* **31**, 43-44.

9. Kyogoku,Y., Kojima,C., Lee,S.J., Suzuki,N., Matsuo,H. and Shirakawa,M. (1995). *Method. Enzymol.*, in press.

10. Shida,T., Kojima,C., Ono,A., Kainosho,M. and Kyogoku,Y. "Conformation change of 31 kDa Holliday Junction revealed by NMR spectroscopy with sequence selective thymine sugar ^{13}C labeling", in preparation.

The rest of the content of this thesis will be published in the following articles and communications.

11. Kojima,C. and Kyogoku,Y. "NMR Rsym factor as analogy of the X-ray Rsym factor, for the error estimation of the NOESY peak volume", in preparation.

12. Kojima,C. and Kyogoku,Y. "Investigation of validity of a single correlation time model as the first order approximation for the estimation of the ^{1}H-^{1}H distances in a DNA decamer system", in preparation.

13. Kojima,C. and Kyogoku,Y. "Precise structure determination of a DNA decamer by the back calculation refinement against the NOESY intensities with different mixing times", in preparation.

14. Kojima,C. Kawashima,E., Sekine,T., Miyahara,M., Aoyama,Y., Ono,A., Kainosho,M., Ishido,Y. and Kyogoku,Y. "NMR analysis of the sugar conformation of a DNA decamer with (2'R)-[2'-^{2}H]- and (2'S)-[2'-^{2}H]-labeling", in preparation.

15. Kojima,C. and Kyogoku,Y. "Conformational analysis of the polypurine / polypyrimisine junction of a DNA 17 mer by NMR", in preparation.

16. Kojima,C. and Kyogoku,Y. "Simulation of the precision and accuracy of the NMR structure of a DNA decamer by the metric matrix distance geometry calculation", in preparation.

17. Kojima,C., Kawashima,E., Serikawa,Y., Sekine,T., Ono,A., Kainosho,M., Ishido,Y. and Kyogoku,Y. "HSQC spectrum of the amino groups of nucleic acids as a tool for the observation of the protein-DNA interaction", in preparation.