^{1}H Nuclear Magnetic Resonance Study on Equilibrium between Two Four-Stranded Solution Conformations of Short d(C_nT)†

Kenji Kanaori,‡ Atsunori Maeda,‡ Hideyuki Kanehara,§ Kunihiko Tajima,‡ and Keisuke Makino*,§

Department of Polymer Science and Engineering, Kyoto Institute of Technology, Matsugasaki, Sakyo-ku, Kyoto 606, Japan, and Institute of Advanced Energy, Kyoto University, Gokasho, Uji 611, Japan

Received March 3, 1998; Revised Manuscript Received June 23, 1998

ABSTRACT: NMR analysis of d(C_4T) showed the slow exchange between two distinct tetramers (each fully symmetric) in solution. For one tetramer, NOE cross-peak patterns characteristic of an i-motif structure (H1′−H1′ and H6−H1′/H1′−H6) were observed between C1 and T5, indicating that this tetramer takes a completely intercalated conformation where the T5 residue is stacked on the C1·C1$^+$ pair of the other duplex (*S*-form). The other was found to be a tetramer in which one of the duplexes is shifted by one nucleotide unit (*R*-form), resulting in nonstacking 3′ end thymidine residues and an equal number of stacked C·C$^+$ pairs to that of the *S*-form. The same spectral features were observed for d(C_3T) but neither for d(TC_3) nor d(TC_4), indicative of the critical role of the position of the thymidine residue in the tetrad isomerization. From NMR denaturation profiles, the *S*-forms were found to be more stable than the *R*-forms, and the linear relationship between the logarithm of the equilibrium constant (K = [tetramer]/[single]^4) and the inverse of temperature ($1/T$) was confirmed for both forms, indicating conformity to the two-state transition model. Both enthalpy and entropy values of the formation of the *S*-form from four single strands were more negative than those of the *R*-form. The enthalpy term should contribute to the stabilization of the *S*-forms at low temperatures. The difference of the free energy values [$\Delta G°$(*S*-form) − $\Delta G°$(*R*-form)] was found to be −2.1 and −2.7 kJ·mol^{-1} at 20 °C for d(C_4T) and d(C_3T), respectively, explaining the higher stability of the *S*-forms. With increasing temperature, these two topologies were found to comparably exist at equilibrium in solution with slow exchange via dissociation to the single strands. A biological role of this topological isomerization is also suggested.

For almost 3 decades, it has been known that nucleic acids containing stretches of cytidine residues can form parallel strands held by cytosine-protonated cytosine base pairs (C·C$^+$) (Figure 1a) (*1*−*4*). Recently, NMR analysis carried out for d(TC_5) at pH 4.3 showed unusual interresidue sugar−sugar NOE cross-peaks between H1′−H1′, which are not detected for a conventional DNA duplex (*5, 6*). This unique structure of such C-rich oligodeoxynucleotides has been identified as a four-stranded complex, namely, i-motif (Figure 1b). In this structure, two base-paired parallel-stranded duplexes are associated, their base pairs are fully intercalated, and the relative orientation of the duplexes is antiparallel. Crystallographic data on d(C_4) and d(C_3T) also confirmed the four-stranded folding (*7, 8*). Not only an intermolecularly but also an intramolecularly folded i-motif was reported for several oligodeoxynucleotides which has CCC-repeat sequences (*9*−*13*).

The previous NMR investigations on the structure of d(TC_n) (n = 2, 3, and 5) showed that the four single strands forming the i-motif structure are equivalent on the NMR time scale (*5, 6, 11*). For example, the number of the cytidine imino peaks is equal to that of the consecutive cytidine residues (n), the total area of the cytidine imino peaks is 2n per tetramer, and a single line can be observed for the imino, H6, and methyl signal of T1 although the tetramer molecule contains four thymidine residues. The same spectral features were observed for d(C_4TC_4), d(5mCCT) (5mC, 5-methylcytidine),[1] and so on (*6, 11*).

Recently, it has been reported for some C-rich oligodeoxynucleotides that isomeric conformations of the i-motif exist in solution. For instance, d(T5mCC) forms two tetramers in comparable proportions in solution (*11*). The existence of at least two intramolecular i-motif isomers was also reported for CCC-repeat sequences, d[(CCCTAA)$_3$CCC] (*10*) and d(CCCTCTCCTTTTTTCCCTCTCC) (*14*). The recent study on d(5mCCTCC) showed two intercalated nonequivalent symmetrical duplexes (*15*). Because of these findings, the conformational heterogeneity of i-motif structures has attracted a great deal of considerable attention.

Furthermore, regarding basic properties of the formation of isomeric i-motif conformations, it has been suggested for d(C_n) (n = 2−5) (*10, 11*) that the intercalation manner of the i-motif gives two possible arrangements of fully intercalated tetramers, with either 3′ or 5′ end cytidines being located outside stacked C·C$^+$ pairs. However, because of poor resolution of the NMR spectra or of a small fraction of

† This work was supported by a Grant-in-Aid for Scientific Research (08458176 and 09780545) from the Ministry of Education, Science and Culture of Japan.

‡ Kyoto Institute of Technology.

§ Kyoto University.

[1] Abbreviations: 5mC, 5-methylcytidine; CD, circular dichroism; RPLC, reversed-phase liquid chromatography; 1D, one-dimensional; 2D, two-dimensional; DQF-COSY, double-quantum-filtered correlated spectroscopy; HOHAHA, homonuclear Hartmann−Hahn spectroscopy; NOESY, nuclear Overhauser enhancement spectroscopy.

S0006-2960(98)00492-9 CCC: $15.00
Published on Web 08/21/1998

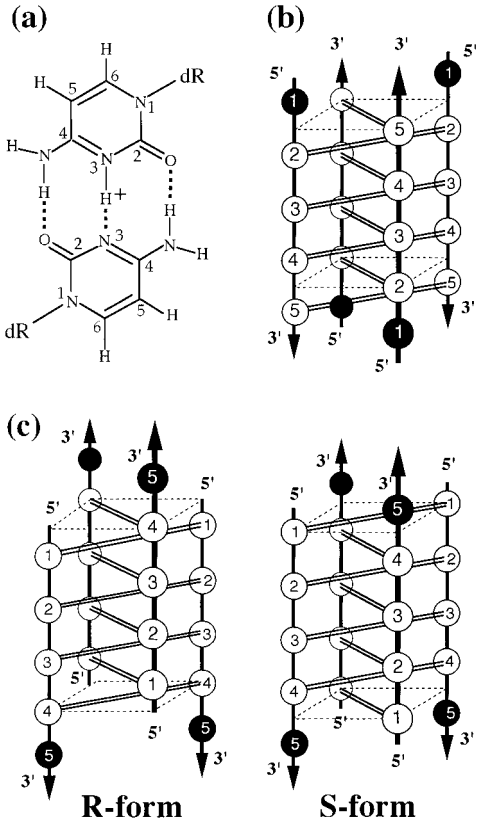

FIGURE 1: (a) Chemical structure of the C·C$^+$ base pair. dR indicates a deoxyribose moiety. Schematic drawings of NMR-observable i-motif topologies of (b) d(TC$_4$) and (c) d(C$_4$T): *R*-form (left) and *S*-form (right). Cytidine and thymidine residues are symbolized by open and closed circles, respectively. Eight C·C$^+$ base pairs are formed in the parallel duplexes for all the topologies, indicated by double lines between open circles.

the total amount, the equilibrium between the two arrangements has not been investigated in detail, and the obvious energetic difference has not been reported at all.

In the present study, therefore, we explored short natural type oligodeoxycytidines, d(C$_n$T) and d(TC$_n$) ($n = 3$ and 4), by NMR spectroscopy to see if isomeric i-motif conformations can exist for these oligomers. Based on the results obtained, it was evident for the first time that the unmodified C-repeat oligodeoxynucleotides containing thymidine residue at the 3′ terminus, d(C$_n$T), can take two distinct tetrad conformations whose intercalation topologies are different from each other (Figure 1c), and thermodynamic properties of the two i-motif structures were also obtained to discuss the factors governing the occurrence of the isomeric topologies.

MATERIALS AND METHODS

Sample Preparation. DNA oligomers, d(C$_4$T), d(C$_3$T), d(TC$_3$), and d(TC$_4$), were purchased from Cruachem (Kyoto, Japan) or Greiner Japan (Kanagawa, Japan) and purified by reversed-phase liquid chromatography (RPLC) as previously described (*16*). The purity of the oligomers was checked by RPLC and ^{1}H NMR. The strand concentrations used in the following measurements were estimated using the extinction coefficients at 260 nm which were calculated by the nearest-neighbor method (*17*).

CD Spectroscopy. CD spectra were measured over 200−350 nm on a J-720 spectropolarimeter (Japan Spectroscopic Co., Japan) using jacketed quartz cells with optical path lengths of 1 and 0.1 cm. The CD-cell temperature was regulated by circulating ethylene glycol/water mixture in the jacket whose temperature was controlled by an RTE-100 thermostat (NESLAB, Japan). After each temperature change, samples were allowed to stand for 20 min for the next measurements. The samples were premelted at 80 °C prior to the measurements to destroy secondary structure and then allowed to thermally equilibrate.

Gel Filtration Chromatography. Gel filtration chromatography was performed using a CCPM-II pumping system coupled with an SC-8020 system controller and a UV-8020 detector (Tosoh, Japan). The samples were annealed by heating the sample at 80 °C and cooled slowly down to the room temperature. The column used was TSK-G3000SW (10 μm, 8 × 600 mm). Other chromatographic conditions are as follows: sample size, 30 μL (100 μM); eluent, 20 mM acetate buffer (pH 4.5) containing 100 mM NaCl; flow rate, 0.6 mL/min; detection, 280 nm; temperature, ambient.

NMR Measurements. NMR samples were prepared in 20 mM acetate buffer (pH 4.5) containing 100 mM NaCl in D_2O. The strand concentration was 10 mM unless otherwise stated. The sample in an NMR tube was premelted at 80 °C prior to the measurements and then stored at various temperatures. Attainment of the equilibrium among single- and four-stranded components was confirmed by measuring the time course of one-dimensional (1D) ^{1}H NMR spectra. All the NMR spectra were recorded on a Bruker ARX-500 spectrometer. For resonance assignments, DQF-COSY (*18*, *19*), HOHAHA (mixing time 40 ms) (*20*, *21*), and NOESY (mixing times 200 and 350 ms) (*22*, *23*) were performed according to the time-proportional phase incrementation method (*24*). For the accurate evaluation of NOE cross-peak intensities, a NOESY spectrum was acquired by applying a short mixing time of 50 ms. The two-dimensional (2D) NMR measurements were carried out at 10 °C. Exchangeable protons were partly assigned in 90% H_2O/10% D_2O by jump-and-return NOESY (*25*) at 5 °C. Chemical shifts were referred to internal sodium 3-(trimethylsilyl)propionate-*2,2,3,3-d$_4$*.

Analysis of Thermodynamic Parameters. Concentrations of single- and four-stranded components were determined by integrating corresponding peak areas of 1D spectra measured with 45° pulse (5 μs), long delays of 10 s, and 32 scans. The equilibrium constant K (=[tetramer]/[single]^4) can be written as $K = \theta/\{4C_0^3(1-\theta)^4\}$ (θ is the fraction of folded oligodeoxynucleotide, and C_0 is the total strand concentration). K was determined at each temperature from the denaturation profiles in the θ range of 0.1−0.9. If the dissociation of the tetramer to the single strands conforms to the two-state transition model, linearity (correlation coefficient ≥ 0.998) should be obtained between the logarithm of K and the inverse of temperature. Enthalpy and entropy values were determined by a slope and an y-intercept of the plot of ln(K) and 1/T on the basis of the equation: $\ln(K) = -(\Delta H/R)\cdot(1/T) + (\Delta S/R)$.

RESULTS

Evidence for Formation of i-Motif Structure by CD Spectroscopy and Gel Filtration Chromatography. Specific Cotton effect appearing in CD spectra of C-rich oligode-

oxynucleotide is informative of protonated and nonprotonated cytosine base pairing (*26*, *27*), and an enlarged positive peak with red shift and a concomitant negative peak are indicative of the i-motif structure (*9*). When this structure collapses, this positive peak is lowered with blue shift and simultaneously the negative peak disappears.

First, pH-dependent CD spectroscopy was carried out for d(C_4T) (45 μM) at 0 °C. This sample showed both the 288-nm positive and 265-nm negative peaks in the pH range of 4.0−5.0, while at higher (pH > 5.0) and lower (pH < 4.0) pHs, the positive peak decreased with blue shift to 282 nm and the concomitant negative peak disappeared (Supporting Information). Temperature-dependent CD spectroscopy was performed for d(C_4T) at pH 4.5 in the temperature range of 0−60 °C. Both the positive and negative peaks appeared below 10 °C and became smaller as the temperature was elevated. Thermal stability was concentration-dependent, which allows us to confirm the intermolecular interaction (data not shown).

To confirm whether tetramers were formed, gel filtration chromatography of d(C_4T) was carried out. As references, d(T), d(T_2), d(T_4), d(T_{12}), d(T_{20}), d(T_{30}), and d(T_{50}) were taken. The chromatogram of d(C_4T) gave two peaks. The second eluted peak was assigned to the single strand by the calibration curve obtained by using the reference molecules. The observation of only one peak for multimers (the first eluted peak) indicates that there is no significant difference in the molecular size between the major and minor components. The molecular size of the d(C_4T) tetramer determined by the calibration plot was equivalent to d(T_{12}). This is probably because the i-motif does not take an extended structure like d(T_n) oligomers but a compact packed structure with the slower elution.

Temperature Dependence of 1D ^{1}H NMR Spectra of d(C_4T). At varied temperatures, ^{1}H NMR measurements were carried out for d(C_4T) at the strand concentration of 10 mM at pH 4.5, and the resulting spectra are depicted in Figure 2. The left parts (7.40−8.10 ppm) of the figures are regions for H6 protons, and the right parts (1.74−1.94 ppm) are those for T5 CH_3 protons. Each 1D spectrum was collected after the attainment of equilibration. In the temperature range of 20−55 °C, three peaks were observed in the T5 CH_3 region (e.g., 1.81, 1.83, and 1.88 ppm at 30 °C), and the signal at 1.88 ppm was assigned to T5 CH_3 of the disordered single strand, because at high temperatures (>55 °C), only this peak appeared and the others disappeared, and also because the chemical shift of this peak was independent of the temperature change. Below 20 °C, the single-stranded component was scarcely observed.

With decreasing temperature, the T5 CH_3 peaks at 1.81 and 1.83 ppm were shifted upfield, such as 1.77 and 1.81 ppm at 15 °C, respectively. The signal intensity of these peaks showed different temperature dependence. With decreasing temperature, the intensity of the former peak increased up to 35 °C and then decreased, while that of the latter showed a constant increase. Analogous temperature dependency was also clearly observed for some peaks in the H6 region. As shown shortly, detailed analysis of 2D NMR spectra gave the conclusion that all the cross-peaks fell into two complete sets. So far here, therefore, these two peak sets seem to be attributed to some ordered structures generated by inter- or intramolecular interaction.

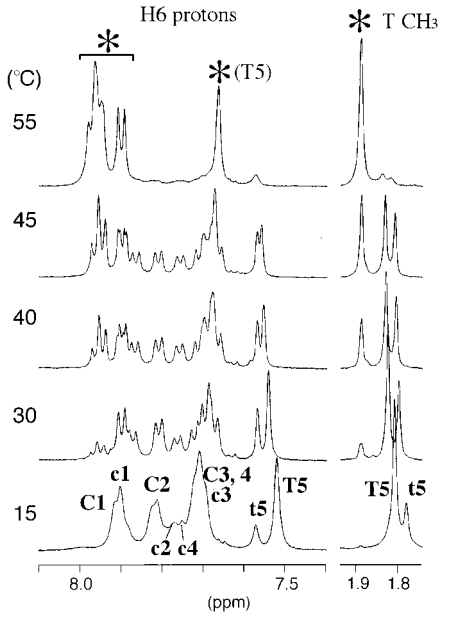

FIGURE 2: Temperature dependence of 1D ^{1}H NMR spectra of H6 and T5 CH_3 proton region of d(C_4T) in D_2O. Conditions: strand concentration, 10 mM; pH, 4.5. Capital letters (C1−T5) and small letters (c1−t5) represent residues of the *S*-form and *R*-form, respectively. Asterisks indicate peaks of the disordered single strand.

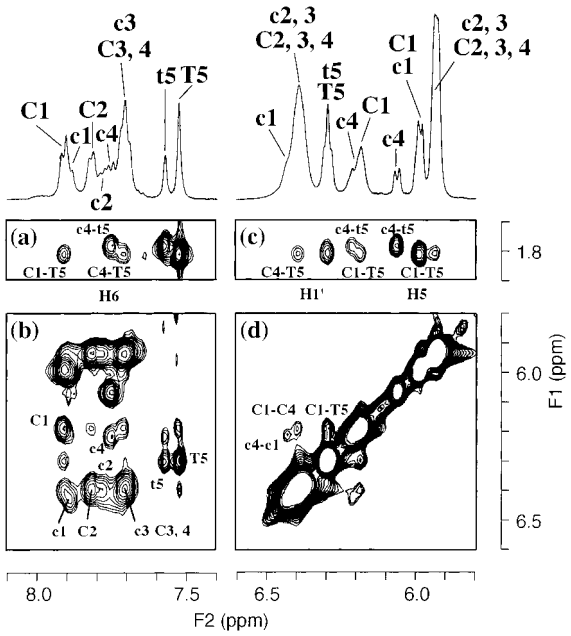

FIGURE 3: NOESY spectra of (a) T5 CH_3−H6, (b) H1′/H5−H6, (c) T5 CH_3−H1′/H5, and (d) H1′/H5−H1′/H5 regions of d(C_4T). Capital letters (C1−T5) and small letters (c1−t5) represent residues of the *S*-form and *R*-form, respectively. Intraresidue NOE connectivities between H1′ and H6 in panel b are indicated by residue names, and interresidue NOEs are indicated by hyphenated residue names in panels a, c, and d. Mixing time of spectra, a−c was 350 ms, and that of spectrum d was 50 ms.

Resonance Assignment of d(C_4T). To assign the signals of d(C_4T), 2D ^{1}H NMR spectra (DQF-COSY, HOHAHA, and NOESY) were measured at 10 °C where no peaks were observed for a single-stranded d(C_4T). The NOESY spectra obtained are shown in Figure 3. Peaks of T5 CH_3 at 1.81 ppm and H6 at 7.52 ppm were connected by a NOE cross-peak as well as those of T5 CH_3 at 1.77 ppm and H6 at 7.57 ppm (Figure 3a). Hence, both peaks at 7.52 and 7.57 ppm

were assigned to T5 H6 protons. By analysis of the NOESY and HOHAHA spectra, most of the deoxyribose protons of T5 were found to show two distinct peaks, although some signals were completely overlapped, for example, H1′ signals at 6.29 ppm. The imino proton of T5 showed two peaks at 11.21 and 11.26 ppm, both of which were also connected by NOE cross-peaks with T5 CH_3 signals at 1.81 and 1.77 ppm, respectively (data not shown). Thus all the protons of T5 residue could also be classified into two groups.

For the cytidine residues, the resonance assignment of the H6, H5, and H1′ protons is shown in the 1D spectra of Figure 3. The assignment of C1 deoxyribose and base protons started with the most upfield-shifted C1 H5′/H5" protons, and was unambiguously accomplished by using DQF-COSY, HOHAHA, and NOESY spectra. For the protons of the C4 residue, the assignment was successfully carried out on the basis of the cross-peaks with the T5 CH_3 protons (Figure 3a,c). For C2 and C3 protons, their exact chemical shift values were not completely defined because of severe signal overlapping; however, their approximate positions could be estimated as shown in the 1D spectrum of Figure 3. Consequently, all the protons were found to exhibit two distinct peaks, and thus, these signals could be classified into the two peak sets.

The intensities of the well-resolved T5 H6 and CH_3 signals were good markers for the proportion between these two peak sets. One peak set represented by the T5 CH_3 peak at 1.81 ppm is, hereafter, referred to the major set, while the other by the T5 CH_3 peak at 1.77 ppm is the minor set. At 10 °C, the peak ratio (major/minor) was about 4 for both of the T5 H6 and CH_3 signals. For all other nicely separated peaks, the same ratio was obtained. The ratios were changed as a function of temperature as shown in Figure 2, suggesting the equilibrium between the two peak sets. Another remarkable finding is that no NOE cross-peak was observed between the major and minor peak sets at all, indicative of the slow chemical exchange of these two groups on the NMR time scale.

Evidence of Two Distinct Tetramers of d(C_4T). In the NOESY spectra, features typical of the i-motif such as H1′−H1′ (Figure 3d) and amino−H2′/H2" cross-peaks (Figure 3S in Supporting Information) were observed for both peak sets. Additional supporting data for the i-motif formation were obtained by the chemical shifts of the internal amino and imino protons of cytidine residues. In both peak sets, the internal amino protons appeared in the low-field side of 9.0−9.3 ppm and the imino protons in the range of 15.4−15.8 ppm (data not shown), indicating that all the cytidine residues are part of hemiprotonated base pairs (*5*, *6*, *11*). These NOE cross-peaks and chemical shifts indicate that both of the peak sets arise from the i-motif structures. Accordingly, there exists possible assumptions that the two peak sets arise from slowly exchanging two i-motif conformers or a single nonsymmetrical i-motif.

If the latter is the case, some NOE cross-peaks should be observed between the major peaks and the minor peaks. As stated before, however, no single NOE cross-peak was observed between the two peak sets even with a long mixing time of 350 ms. These facts can rule out the single nonsymmetrical i-motif interpretation. Moreover, the proportions of the corresponding peaks between the major and minor components were continuously altered by the temperature change, as shown in Figure 2. Therefore, all the present NMR results support the assumption of two distinct tetramers (each fully symmetric) which exchange slowly on the NMR time scale.

Major Tetramer Conformation of d(C_4T). For the major set, the NOESY cross-peak pattern showed interresidue H1′−H1′ cross-peaks (Figure 3d) and interresidue H1′−H6 cross-peaks occurring in both directions (5′ to 3′ and 3′ to 5′), so-called 'rectangular' pattern, between T5 and C1 residues (Figure 3b). The former was observed even in the NOESY spectrum with a short mixing time of 50 ms, and the latter was seen only in the NOESY spectra obtained with long mixing times. This trend was reported previously (*5*). Although all the H1′−H1′ cross-peaks were not separately observed on account of overlapping of the H1′ and H6 protons of C2, C3, and C4 residues, the topology of the major component would be the fully stacked, 'standard' i-motif topology, hereafter called *S*-form (right in Figure 1c). In the *S*-form, the T5·T5 base pair is stacked on the C1·C1 pair of the other duplex, which is indicated by the cross-peaks between the C1 amino protons and the T5 CH_3 protons (Figure 3S in Supporting Information). Weak sequential NOE cross-peaks were observed for the d(C_nT) *S*-form in the NOESY spectrum obtained with a long mixing time of 350 ms (Figure 3b), such as C1 H1′−C2 H6, C4 H1′−T5 H6. Such a weak sequential cross-peak was also previously reported for d(TC_5) (*5*). These sequential cross-peaks should be partly due to spin diffusion effects because these cross-peaks were not observed in the NOESY spectra obtained with a shorter mixing time (data not shown).

Minor Tetramer Conformation of d(C_4T). Compared to the results for the major set, for the minor set, the H1′−H1′ cross-peak and rectangular pattern of NOESY cross-peaks were not observed between T5 and any cytidine residues. It was also found that the internal and external amino protons of the C4 residue of the minor component did not show any cross-peaks with the H2′/H2" protons and the T CH_3 protons (Figure 3S in Supporting Information). On the contrary, sequential NOE cross-peaks connectivities between C4 and T5 were observed as is seen in conventional DNA duplexes, such as C4 H1′−T5 H6 and C4 H6/H5-T5 CH_3 (Figure 3a−c). The chemical shifts of H5, H6, and H1′ protons of C2 and C3 were almost identical between the major and minor peak sets, while those of the C4 residue of the minor component differed from those of the major component. These facts indicate that the minor component consists of the two antiparallel duplexes, one of which is shifted to 3′ direction from the position of the *S*-form and the other by one-base sliding where the thymidine residue is located on the periphery of the stacked $C{\cdot}C^+$ pairs (left in Figure 1c). Hereafter, this unusual 'rare' structure will be called *R*-form. This unusual i-motif topology has been previously proposed for one of the two tetramers of d(T5mCC) (*11*).

NMR Study on Other Analogous Sequences. Similar 1D and 2D NMR measurements were performed for d(C_3T), d(TC_3), and d(TC_4). A NOESY spectrum of d(C_3T) manifested that d(C_3T) takes two i-motif topologies similar to those of d(C_4T), *S*-form (the major peak set) and *R*-form (the minor peak set). (Note that the NOESY spectrum and temperature dependence of 1D ^{1}H NMR spectra of d(C_3T) are supplied in Supporting Information.) Although chemical shifts of each cytidine H6 proton were severely overlapped

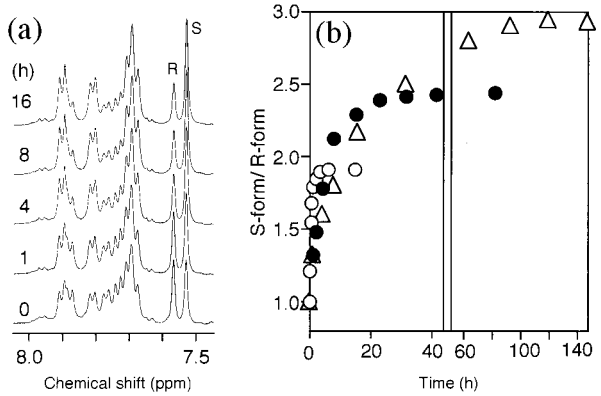

FIGURE 4: (a) Change of 1D ^{1}H NMR spectra of aromatic region of d(C_4T) which was left at 20 °C after annealing the sample. (b) Time dependence of the ratios (*S*-form/*R*-form) at various temperatures: 15 (triangle), 20 (closed circle), and 25 (open circle) °C. The first spectrum (0 h) was obtained 5 min after the annealing at 80 °C. The proportions of the *S*-form to the *R*-form were determined by comparing the areas of T5 H6 signals labeled by "S" (*S*-form) and "R" (*R*-form).

between the major and minor sets, T4 H6 and methyl protons were separated sufficiently for determining tetrad topologies and their populations. Chemical shifts of the thymidine base protons resembled to those of d(C_4T). The T4 H6 proton of the major component was shifted upfield while the methyl signal downfield, compared to those of the minor component.

For d(TC_3) and d(TC_4), only the *S*-form was observed (data not shown), which is consistent with the previously reported results on d(TC_n) (*6*). In this topology, the T1·T1 base pair is stacked on the 3′ end C·C$^+$ pair of the other duplex (Figure 1b). Thus it was concluded that the *R*-form was not observed for d(TC_n) by NMR spectroscopy although d(C_nT) takes the two different i-motif topologies. The equilibrium between the *S*-form and *R*-form is considered to exist for d(TC_n) whose chemical structure is almost similar to that of d(C_nT), and thus the amount of the d(TC_n) *R*-form would be too small to be observed by NMR spectroscopy.

Equilibration Time of S- and R-forms. The equilibrium between the *S*- and *R*-forms of d(C_4T) took a long time at low temperatures. In the present study, therefore, we measured temperature dependence of the signal intensity below 30 °C where the amount of the single-stranded component is small (<5%): Note that a conversion between each tetramer and single strands dominates equilibrium, as shown shortly. The sample was heated at 80 °C and immediately cooled to the specific temperature, and then 1D ^{1}H NMR spectra were obtained as a function of time until the equilibrium was attained. The spectrum obtained immediately after the quenching procedure exhibited the 1:1 population for the two sets; then gradually the population ratio of the major to the minor became equilibrated according to the given temperature (Figure 4a). The equilibration, for instance, took about 5 h at 25 °C and 30 h at 20 °C. Below 20 °C, the conversion between the two topologies was very slow, and the establishment of the equilibrium at 15 °C needed more than 3 days in this experimental procedure; the final ratio of *S*-form/*R*-form was 2.9 (open triangle in Figure 4b). On the other hand, for d(C_3T), the equilibrium between the *S*- and *R*-forms was reached more rapidly, probably because of the existence of an NMR-observable amount of the single-stranded component even below 10 °C.

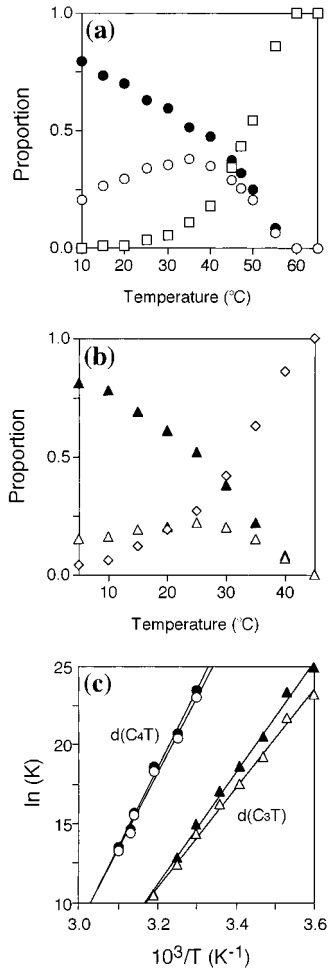

FIGURE 5: Denaturation profiles of (a) d(C_4T) and (b) d(C_3T) and (c) plots of ln(*K*) versus 1/*T* for d(C_4T) and d(C_3T) tetramers. Closed and open marks indicate the *S*-form and *R*-form tetramer fractions of d(C_4T) (circles) and d(C_3T) (triangles), respectively. Square and diamond marks indicate the single-stranded fractions of d(C_4T) and d(C_3T), respectively.

Thermodynamic Parameters of S- and R-Forms. To estimate the i-motif stability of d(C_nT), denaturation profiles of the *S*- and *R*-forms were obtained for d(C_4T) and d(C_3T) by plotting the change of the peak areas in the 1D NMR spectra against various temperatures (Figure 5). The peak areas of T5 CH_3 and H6 signals were used as markers for estimating the populations of the *S*- and *R*-forms. Apparent melting temperatures (T_m) of d(C_4T) and d(C_3T) were determined by summing the two tetrad components and comparing the sum with the fraction of the single strand. The T_m values obtained were 49 and 32 °C for d(C_4T) and d(C_3T) at the concentration of 10 mM, respectively. Denaturation profiles of d(TC_4) and d(TC_3) were also examined in the same condition: 10 mM strand concentration, 20 mM acetate buffer, 100 mM NaCl, pH 4.5. The T_m values obtained for d(TC_4) and d(TC_3) were 55 and 43 °C, respectively. The latter value was almost identical to that of the previously reported value (*6*). These results demonstrate that d(TC_n) forms a more stable tetramer than d(C_nT). All the T_m values were concentration-dependent, indicating the intermolecular i-motif formation (data not shown).

With a decrease in temperature, the *S*-form topology was found to become more favorable for both d(C_4T) and d(C_3T). For instance, the proportions of the *S*-form to the *R*-form of

d(C_4T) and d(C_3T) were about 4 and 8 at 10 °C, respectively. The reason for this preference of the *S*-form was studied thermodynamically. To obtain thermodynamic parameters, ln(*K*) was plotted against 1/*T* for both the tetramers of d(C_4T) and d(C_3T). The plots yielded straight lines for all four tetramers (all the fitting coefficients ≥ 0.998), indicating the two-state transition model between each tetramer and the single strands: Note that previously, the two-state transition model has been adopted successfully for explaining the denaturation profile of the i-motif (*9, 10, 14*). According to the method described in Materials and Methods, $\Delta H°$ and $\Delta S°$ values were determined for both topologies of the two oligomers. On the basis of the two-state transition model, both $\Delta H°$ and $\Delta S°$ values of the formation of the *S*-form from four single strands were found to be more negative than those of the *R*-form: For d(C_4T), $\Delta\Delta H° = \Delta H°$(*S*-form) $- \Delta H°$(*R*-form) $= -11$ kJ·mol^{-1} and $\Delta\Delta S° = \Delta S°$(*S*-form) $- \Delta S°$(*R*-form) $= -30$ kJ·K^{-1}·mol^{-1}, and for d(C_3T), $\Delta\Delta H° = -30$ kJ·mol^{-1} and $\Delta\Delta S° = -93$ kJ·K^{-1}·mol^{-1} with an error of about 10%. Thus it is deduced that the enthalpy term contributes to the stabilization of the *S*-forms at low temperatures, while as temperature increases, the entropy term takes part in the stabilization of the *R*-forms. The differences of the free energy between the major and minor topologies [$\Delta G°$(*S*-form) $- \Delta G°$(*R*-form)] were also estimated to be -2.1 and -2.7 kJ·mol^{-1} at 20 °C for d(C_4T) and d(C_3T), respectively. This result indicates that the energetic difference between the *S*- and *R*-forms of d(C_4T) was smaller than that of d(C_3T) and that the increase in the number of consecutive cytidine residues results in the increase in the proportion of the *R*-form.

DISCUSSION

From the present NMR study on d(C_4T) and d(C_3T), it can be pointed out that two i-motif structures, *S*-form and *R*-form (Figure 1c), exist at equilibrium in the solution of d(C_nT). In 1993, the i-motif was first discovered in the structure of d(TC_5) in an acidic solution (*5*), and it was demonstrated that this unique structure is built on symmetrical hemiprotonated base pairs formed by corresponding residues of parallel strands. Then i-motif formation has also been confirmed by NMR and X-ray crystallography for many C-rich oligodeoxynucleotides (*6*−*8, 11, 13, 25, 28*). Throughout such a number of elaborate investigations, almost all the i-motif structures determined so far were found to form a single structure of a fully intercalated i-motif as represented in Figure 1b or 1c (right), and the exception was only observed for C-modified oligonucleotides such as d(T5mCC) that takes a 1:1 mixture of the *S*-form and *R*-form in solution (*11*). This coexistence of the two topologies was ascribed by the authors to steric hindrance between two methylated base pairs in the *S*-form of d(T5mCC). However, such isomeric conformations have not been demonstrated for naturally occurring C-rich oligodeoxynucleotides. The present study on d(C_nT), therefore, demonstrates the first discovery of the equilibrium between the two i-motif topologies of the wild-type oligodeoxycytidine stretches, indicating that the position of the thymidine residues is highly responsible for the occurrence of the two topologies.

We should think next why the equilibrium between the two i-motif topologies can be observable by NMR for d(C_nT) but not for d(TC_n). Before going into the detailed discussion, it should be noted that the NMR-observable d(TC_n) i-motif has been already assigned to the *S*-form in the Results section. It is reasonable that d(TC_n) should also exist in the equilibrium between the two topological structures. If this is the case for d(TC_n), the fact that we observed only the *S*-form indicates that the energy difference between the two topologies of d(TC_n) should be large, much larger than the values for d(C_nT). On the other hand, for d(C_nT), the stability is rather comparable between the S- and *R*-forms since the free energy differences between both the forms were estimated to be as small as -2.1 and -2.7 kJ·mol^{-1} at 20 °C for d(C_4T) and d(C_3T). It is likely that this small energy difference resulted in the observation of the isomeric i-motifs. This small free energy difference between the two i-motif topologies may be caused by the destabilization of the *S*-form, based on our current observation that the T_m value of the d(C_nT) tetramers was much lower than that of the NMR-observable d(TC_n) tetramer, by about 11 and 6 °C for $n = 3$ and 4, respectively.

Generally the i-motif stability of d(TC_n) or d(C_nT) can be attributed (1) to hydrogen bonding between cytosine and protonated cytosine, (2) to base stacking between C·C$^+$ pairs or between C·C$^+$ and T·T pairs, and (3) to intermolecular van der Waals contacts between the deoxyribose moieties of the antiparallel strands in the narrow groove of the i-motif. The number of the hydrogen bondings formed in the C·C$^+$ pairs is 6*n* for both the d(C_nT) and d(TC_n) tetramers. The numbers of the base stacking C·C$^+$−C·C$^+$ and C·C$^+$−T·T are ($2n - 1$) and 2, respectively, for both the d(C_nT) and d(TC_n) *S*-forms. However, for the d(TC_n) tetramer, not only the T1·T1 base pair can stack on the C(n+1)·C(n+1) base pair, but also the deoxyribose moiety of T1 could make some van der Waals contacts with that of 3′ C(n+1), as suggested by the interresidue NOEs of H1′−H1′, H1′−H2″, and H1′−H4′ (*5, 11*). These intermolecular van der Waals contacts are considered to make the i-motif structure highly stable (*5*). On the other hand, the van der Waals contacts between deoxyribose moieties of T5 and C1 in the d(C_nT) *S*-form may be weaker than those of T1 and C5 in the d(TC_n) *S*-form, because of the following observations. The chemical shifts of T5 H2′/H2″ protons were degenerated for both the d(C_nT) tetramers, but those of T1 H2′/H2″ protons were separated for the d(TC_n) tetramer. In addition, no NOE peak was observed between C1 H1′ and T5 H2″, although the NOE peaks between C1 H1′ and T5 H1′/H4′ were observed. These observations indicate that the deoxyribose moiety of T5 weakly contacts with that of C1 in the opposite strand. This assumption is supported by the reported i-motif structures (*5, 11*), which suggest that the 3′ terminal deoxyribose moiety of T5 is supposedly sticking out and located outside the T5·T5 base pair. The small extent of the van der Waals contacts in the *S*-form of d(C_nT) may be the reason that the d(C_nT) *S*-form is less stable than that of d(TC_n). Thus mutual geometry of the deoxyribose moieties in the narrow groove of the i-motif may affect the stability of the i-motif.

Next, we discuss here the reason that the *S*-form is more stable than the *R*-form. The *R*-form of d(C_nT) possesses the same number of hydrogen bonds and base stacking between C·C$^+$ pairs as the *S*-form (Figure 1c); however, in the *R*-form, the T(n+1)·T(n+1) base pairs do not stack on the C·C$^+$ pair of the other duplex, hardly contributing to the stabilization.

This unstacking of the 3′ end T·T pair is probably reflected by the less negative enthalpy values, compared to that of the *S*-form. Another reason for the observed enthalpic destabilization of the *R*-form of d(C_nT) may be the absence of the interstrand van der Waals contacts between C1 and T5. However, the fairly strong sequential NOEs between C(*n*) and T(*n*+1) of the *R*-form of d(C_nT) indicate that the thymidine residue, T(*n*+1), does not move freely in solution but interacts with the C(*n*) residue of the same strand. Recently, it was reported that hydrophobic pairs such as T·T placed at the end of DNA duplexes considerably stabilize the duplexes (*29*). The T(*n*+1)·T(*n*+1) base pairs may stabilize the *R*-form of d(C_nT) and, thus, partly contribute to the small free energy difference between the *S*- and *R*-forms of d(C_nT).

The i-motif conformation of d(C_3T) has been studied by X-ray crystallography (*8*). The coexistence of the *R*-form was not reported in the paper, and the crystal structure was the *S*-form whose C1·$C1^+$ base pair is intercalated between T4·T4 and C3·$C3^+$ base pairs. This result is probably due to favorable packing effects of the *S*-form in the crystal. NMR measurement was also carried out for the d(C_3T) tetramer but did not suggest the coexistence of the two i-motif topologies (*15*). We think that this is possibly because the NMR measurement was conducted under conditions different from ours. For example, a mixed organic solvent was used and temperature as low as −30 °C. An analogous sequence, d(5mCCT), in which T is attached to the 5′ terminus did not show the coexistence of the i-motif topologies (*11*). This may be due to steric hindrance between stacked 5mC·$5mC^+$ pairs in the *R*-form. Consequently no research has yet been able to discuss the role of the attached thymidine residue for the topological diversity. In the present study, we have found for the first time that the isomerization of simple unmodified C-repeats was characterized in relation to the position of the thymidine residue. This trend may be useful for predicting the topology of longer C-rich oligonucleotides. It has been already accepted that the i-motif formation can be detected by various methods such as NMR, CD, UV, Raman spectroscopy, gel filtration chromatography, and so on (*6*, *9*−*11*, *14*, *30*). However, the coexistence of the two i-motif topologies in solution could not be detected by gel filtration chromatography or CD spectroscopy in the present study, although these measurements showed the formation of the i-motif. Except for NMR spectroscopy, therefore, all the methods are presumably insensitive to the difference or the equilibrium between the i-motif topologies. However, even by conventional NMR spectroscopy, it will be difficult to define a precise stacking order of C·C^+ base pairs of the i-motif for long C-rich oligodeoxynucleotides such as d(GATCTTCCCCCCGGAA) (*31*) on account of severely overlapping peaks. In such a case, the trend we have found here may allow us to predict a preferable i-motif topology, because the location of the thymidine residue seems to dominate the occurrence of the two topologies.

In the present study, we have revealed the conformational isomerization of the i-motif structure by precise NMR spectroscopy. This structural i-motif diversity has relevant biological implications in connection with the previous findings. Typical previous results are as follows: It has already been demonstrated in a number of research papers that i-motif possibly exists in cellular systems under physiological conditions. A fragment of the C-rich strand of telomeres (vertebrates and *Tetrahymena*) containing four oligodeoxycytidine stretches has been reported to form an i-motif by intramolecular folding at neutral pH (*10*, *32*). In addition to this basic knowledge, the formation of a double hairpin with a T:G:G:T tetrad followed by an i-motif has been reported recently for C-rich oligodeoxynucleotides, part of a DNA box in human centromeric α satellite targeted by the centromere protein B (*33*), suggesting the involvement of the i-motif in some functioning complexes. Although a protein which specifically recognizes the i-motif structure has not been identified so far, it has been reported that a HeLa nuclear protein binds to the single-stranded d$(CCCTAA)_n$ telomeric motif (*34*) and that the C-rich strand of telomeres may fold in vivo forming an i-motif stabilized by proteins and participating in recombination, meiotic chromosome pairing, or other genomic processes (*10*, *25*, *28*, *32*, *35*). Above-mentioned results on the intramolecular i-motif are of interest because it is conceivable from them that the i-motif structure not only works as rigid structural elements but also may regulate biological functions together with the G-tetrad formation in the counterstrand and related proteins. The structural diversity of the i-motif we have found in the present study may be added to such structures. Since it has been suggested previously that some C-rich oligodeoxycytidines take a mixture of the intramolecular i-motif isomers (*10*, *36*), it is probable that isomerization of the i-motif structures occurs in C-repeats in the telomere. The ratio of the i-motif isomers, if they are generated, could be varied upon change of pH, temperature, and ionic strength and possibly by interaction with proteins. If such equilibrium between the isomeric i-motif takes place in vivo, it may function, for instance, as a switch for the interaction of the i-motif with proteins, G-tetrads, etc., thus further regulating and modulating the function in the replication and meiotic recombination processes.

SUPPORTING INFORMATION AVAILABLE

CD spectra obtained for d(C_4T) as a function of (a) pH and of (b) temperature (Figure 1S), gel filtration chromatogram of d(C_4T) and a plot of the log of the molecular weight versus the elution volume (Figure 2S), NOESY spectrum of the amino protons and CH_3/H2′/H2" proton region of d(C_4T) in 90% H_2O/10% D_2O solution (Figure 3S), NOESY spectra of d(C_3T) (mixing time 350 ms) in D_2O (Figure 4S), and temperature dependence of 1D ^{1}H NMR spectra of H6 and thymidine methyl proton regions of d(C_3T) in D_2O (Figure 5S) (5 pages). Ordering information is given on any current masthead page.

REFERENCES

1. Langridge, R., and Rich, A. (1963) *Nature (London) 198*, 725−728.
2. Hartman, J. K. A., and Rich, A. (1965) *J. Am. Chem. Soc. 87*, 2033−2039.
3. Akinrimisi, E. O., Sander, C., and Ts'o, P. O. P. (1963) *Biochemistry 2*, 340−344.
4. Inman, R. B. (1964) *J. Mol. Biol. 9*, 624−637.
5. Gehring, K., Leroy, J. L., and Guéron, M. (1993) *Nature 363*, 561−565.
6. Leroy, J. L., Gehring, K., Kettani, A., and Guéron, M. (1993) *Biochemistry 32*, 6019−6031.

7. Chen, L., Cai, L., Zhang, X., and Rich, A. (1994) *Biochemistry 33*, 13540−13546.
8. Kang, C. H., Berger, I., Lockshin, C., Ratliff, R., Moyzis, R., and Rich, A. (1994) *Proc. Natl. Acad. Sci. U.S.A. 91*, 11636−11640.
9. Manzini, G., Yathindra, N., and Xodo, L. E. (1994) *Nucleic Acids Res. 22*, 4634−4640.
10. Leroy, J. L., Guéron, M., Mergny, J. L., and Hélène, C. (1994) *Nucleic Acids Res. 22*, 1600−1606.
11. Leroy, J. L., and Guéron, M. (1995) *Structure 3*, 101−120.
12. Rohozinski, J., Hancock, J. M., and Keniry, M. A. (1994) *Nucleic Acids Res. 22*, 4653−4659.
13. Han, X., Leroy, J. L., and Guéron, M. (1998) *J. Mol. Biol. 278*, 949−965.
14. Mergny, J. L., Lacroix, L., Han, X., Leroy, J. L., and Hélène, C. (1995) *J. Am. Chem. Soc. 117*, 8887−8898.
15. Nonin, S., and Leroy, J.-L. (1996) *J. Mol. Biol. 261*, 399−414.
16. Kanehara, H., Mizuguchi, M., Tajima, K., Kanaori, K., and Makino, K. (1997) *Biochemistry 36*, 1790−1797.
17. Fasman, G. D. (1975) *Handbook of Biochemistry and Molecular Biology*, 3rd ed., Vol. 1, CRC Press, Cleveland, OH.
18. Piantini, U., Sørensen, O. W., and Ernst, R. R. (1982) *J. Am. Chem. Soc. 104*, 6800−1680.
19. Rance, M., Sørensen, O. W., Bodenhausen, G., Wagner, G., Ernst, R. R., and Wüthrich, K. (1983) *Biochem. Biophys. Res. Commun. 117*, 479−485.
20. Braunschweiler, L., and Ernst, R. R. (1983) *J. Magn. Reson. 53*, 521−528.
21. Davis, D. G., and Bax, A. (1985) *J. Am. Chem. Soc. 107*, 2820−2821.
22. Jeener, J., Meier, B. H., Bachmann, P., and Ernst, R. R. (1979) *J. Chem. Phys. 71*, 4546−4553.
23. Macura, S., Huang, Y., Suter, D., and Ernst, R. R. (1981) *J. Magn. Reson. 43*, 259−281.
24. Marion, D., and Wüthrich, K. (1983) *Biochem. Biophys. Res. Commun. 113*, 967−974.
25. Kang, C. H., Berger, I., Lockshin, C., Ratliff, R., Moyzis, R., and Rich, A. (1995) *Proc. Natl. Acad. Sci. U.S.A. 92*, 3874−3878.
26. Edwards, E. L., Ratliff, R. L., and Gray, D. M. (1988) *Biochemistry 27*, 5166−5174.
27. Edwards, E. L., Patrick, M. H., Ratliff, R. L., and Gray, D. M. (1990) *Biochemistry 29*, 828−836.
28. Berger, I., Kang, C., Fredian, A., Ratliff, R., Moyzis, R., and Rich, A. (1995) *Nature Struct. Biol. 2*, 416−425.
29. Schweitzer, B. A., and Kool, E. T. (1995) *J. Am. Chem. Soc. 117*, 1863−1872.
30. Benevides, J. M., Kang, C. H., and Thomas, G. J., Jr. (1996) *Biochemistry 35*, 5747−5755.
31. Singh, S., Patel, P. K., and Hosur, R. V. (1997) *Biochemistry 36*, 13214−13222.
32. Ahmed, S., Kintanar, A., and Henderson, E. (1994) *Nature Struct. Biol. 1*, 83−88.
33. Gallego, J., Chou, S.-H., and Reid, B. R. (1997) *J. Mol. Biol. 273*, 840−856.
34. Marsich, E., Piccini, A., Xodo, L. E., and Manzini, G. (1996) *Nucleic Acids Res. 24*, 4029−4033.
35. Nonin, S., Phan, A. T., and Leroy, J. L. (1997) *Structure 5*, 1231−1246.
36. Ahmed, S., and Henderson, E. (1992) *Nucleic Acids Res. 20*, 507−511.

BI980492G