NMR studies of pH-dependent conformational polymorphism of alternating (C-T)_n sequences

T.N.Jaishree and Andrew H.-J.Wang*

Biophysics Division and Department of Cell and Structural Biology, University of Illinois at Urbana-Champaign, Urbana, IL 61801, USA

Received March 29, 1993; Revised and Accepted July 1, 1993

ABSTRACT

Alternating (C-T)_n sequences are involved in the H-DNA structure associated with (GA)_n.(CT)_n sequences. Low pH values facilitate H-DNA formation. We have undertaken a detailed analysis of the structural consequences of the (C-T)_n sequence as a function of pH. The structures of three DNA oligonucleotides, d(CT)₄, d(TC)₄ and d(TC)₁₅, have been studied by NMR. We found that their conformations are polymorphic and pH dependent. There are at least three major conformational species: an antiparallel-stranded (APS) duplex with entirely C:T base pairs at pH 7, an antiparallel-stranded (APS) duplex with entirely C⁺:T base pairs at pH 3, and a possible parallel-stranded (PS) duplex with C⁺:C and T:T base pairs near pH 5. In the intermediate pH range, the APS duplex may have varying numbers of C⁺:T and C:T base pairs, and there may be a fast exchange going on between APS duplex species involving these two kinds of base pairs. However, the transition between the APS and PS duplexes is slow. Structural refinement of the two octamers, d(TC)₄ and d(CT)₄, at pH = 6.9 and pH = 3 using 2D-NOE data suggests that the molecules are likely in the duplex form at 5°C. We lack evidence that the structure at pH 3 is a PS structure with T nucleotides residing in the exterior of the helix. Titration of the longer oligonucleotide, d(TC)₁₅, showed a prominent pK_a of ~6, approaching the value of 7.0 obtained from the titration of poly-(dC).

INTRODUCTION

Alternating (C-T)_n sequences are frequently found to participate in the structure of triple-stranded DNA or H-DNA (1–4). The (CT)_n sequence uses a Hoogsteen base pairing scheme to bind in the major groove of the (GA)_n.(TC)_n duplex (5–7). Low pH facilitates the triplex formation, due to the requirement of the protonated C in the C⁺.G.C base pair. However, protonation of the (C-T)_n sequence may favour other structures. Gray et al. (8–10) proposed that at low pH poly(dC-dT) forms an unusual duplex structure in which the cytosines are hemi-protonated, and form continuously-stacked C⁺:C base pairs with the thymidine

nucleotides extruding into the solvent region. An NMR study (11) of a short DNA hexamer d(CT)₃ at pH=3 supported the model of Gray et al. The key supporting evidence was the absence of NOE crosspeaks between the aromatic protons from adjacent pyrimidines, i.e., CH5/CH6 to TH6 and T-methyl (11).

Recently there has been a resurgence of interest in the study of pH-dependent behavior of nucleic acids, in part due to the possible biological significance of H-DNA (1–4) and the potential application of antigene DNA/RNA via the triple helix (12). Several surprising results have appeared. We have shown that DNA sequences associated with 5'-CGA have a strong tendency to form parallel-stranded homo-base paired double helix under low pH conditions (13). For example, the heptamer d(CGACGAC) forms a remarkably stable PS-DNA duplex at pH 6.0 (14). The structure is stabilized considerably by the hemi-protonated C:C⁺ base pair and the extensive *inter-strand* stacking of G:G base pair over A:A base pair.

Poly-(dC) has been shown to have a very high pK_a and its structure has been proposed to be either a parallel duplex (15) or a single helix (16). An NMR study of d(A₅C₅) at pH 3 indicated that the C₅ segment involved C:C⁺ base pairs in a parallel helix (17). This type of C:C⁺ base pair has been observed in the crystal structure of protonated d(CpG) (18,19). More recently, on the basis of NMR data Gehring et al. (20,21) have proposed a very unusual tetra-stranded fully intercalated helical structure for an oligo-(dC)-containing the molecule, d(TCCCCC). Those results clearly showed that DNA sequences containing a significant portion of oligo-C may have complex pH-dependent conformations.

We have undertaken a detailed analysis of the structural consequences of the alternating C-T sequence under the influence of pH. We have synthesized three DNA oligonucleotides, d(CT)₄, d(TC)₄ and d(TC)₁₅ and analyzed their structures by NMR. Our results revealed an interesting pH-dependent conformational behavior associated with these molecules.

MATERIALS AND METHODS

The oligonucleotides were synthesized by an Applied Biosystems DNA synthesizer and purified by Sepharose G50 column chromatography. The purity of the preparation was judged to

* To whom correspondence should be addressed

be greater than 95% by reverse phase (C4 column) HPLC and NMR analyses. Samples for ^1H NMR analysis were prepared in D_2O as described earlier (22). The solution (final volume of 0.5 ml) contained different concentrations of the molecules (2.5 mM duplex of $\text{d}(\text{TC})_4$ and $\text{d}(\text{CT})_4$ and 1 mM duplex of $\text{d}(\text{TC})_{15}$ with 150 mM NaCl). The pH of this unbuffered sample was monitored and adjusted at each lyophilization step and the reported pH values were measured after the collection of NMR data. For the pH measurements in D_2O solutions, the pH reading in D_2O was calibrated ($\text{pH} = \text{pD} - 0.4$) according to the method of Glasoe et al. (23). All reported pH measurements were made at room temperature.

The ^1H NMR spectra were recorded on a GE GN500 500 MHz spectrometer. The chemical shifts (in ppm) are referenced to the HDO peak which is calibrated to 2,2-dimethyl-2-silapentane-5-sulfonate (DSS) at different temperatures. The 2D-NOESY spectra were collected by the method of States et al. (24) with a composite pulse NOESY sequence. The spectra were recorded with 512 t_1 complex blocks of 2048 complex points each (in the t_2 dimension) and averaged for 16 scans per block. The total recycle time for the NOESY experiment (mixing time 200 ms) was 2.76 s. The 2D data sets were processed with the program FELIX (25) on Silicon Graphics workstations using the truncation apodization described earlier (22) and an exponential multiplication at 5 Hz was used. The intensities of the NOE crosspeaks were measured by the program MYLOR (22). There were ~600 NOE crosspeaks (many of them are overlapped), considered to be above the noise level for the octamers and they were used in the refinement. An iterative spectral-driven procedure SPEDREF (22) using the full-matrix relaxation theory was used for the refinement. Models incorporating various C:T base pairs were prepared using QUANTA (26) and X-PLOR (27). The exchangeable proton 1D spectrum was collected using the selective excitation 1331 pulses

Figure 1. Different base pairing schemes among cytosine (C), N3-protonated cytosine (C^+) and thymine (T). Depending on the relative orientation (cis or trans) of the glycosyl bonds, the resulting duplex may be anti-parallel (APS) or parallel (PS).

to excite resonances ranging from 9 ppm to 17 ppm as described by Hore et al. (28). Hence the peak integrals are distorted.

There are a number of possible base pairing schemes involving cytosine and thymine (29). Some of the relevant ones are shown in Figure 1. Note that when the two glycosyl bonds of a base pair are on the same side of the base pair (i.e., cis configuration), the helix is *usually* anti-parallel. When they are on the opposite side (i.e., trans configuration), the helix is *usually* parallel.

RESULTS AND DISCUSSION

pH-dependent behavior of alternating (C-T) sequences

The pH-dependent behavior of the alternating (C-T) sequences was studied by carefully titrating the solution of $\text{d}(\text{TC})_4$ (2.5 mM duplex) and $\text{d}(\text{TC})_{15}$ (1 mM duplex) with HCl. The titration curves (Figure 2) clearly show that well-defined pK_a values exist for both molecules. A sharp pK_a at 5.9 can be seen for $\text{d}(\text{TC})_{15}$, indicating a cooperative absorption of protons by the 30-mer DNA. This pK_a value is close to that of the poly(dT-dC) ($\text{pK}_a = 5.8$) studied by Brown et al. (9).

The possible structures associated with the alternating $\text{d}(\text{T-C})$ sequences were first analyzed by 1D-NMR. The non-exchangeable proton 1D spectra of $\text{d}(\text{TC})_4$ and $\text{d}(\text{TC})_{15}$ at different pHs at 5°C are shown in Figure 3A and 3B,

Figure 2. pH-titration curves of $\text{d}(\text{TC})_4$ and $\text{d}(\text{TC})_{15}$. The pH titration curves were analyzed by fitting to the Nernst equation using different pK_a and n values which are depicted on the respective figures.

Figure 3. One dimensional non-exchangeable proton 500 MHz NMR spectra at different pHs of (A), $d(TC)_4$ and (B), $d(TC)_{15}$.

respectively. It is clear that there are several distinct species depending on the pH. At neutral pH (6.5–6.9) and acidic pH (2.8–3.0), the spectra are well resolved. The corresponding resonances from the acid pH form generally have slightly more downfield chemical shifts than those of the neutral pH form. However, at intermediate pH (3.8–5.0), a new set of broad resonances appears and co-exists with those from the extremes of the pH ranges. For example, new resonances appear at 1.76 ppm, slightly more upfield than the T-methyl resonance for both the acid and neutral pH forms of $d(TC)_4$ (Figure 3A). Therefore this new species with broad resonances is in slow exchange with the acid and neutral pH forms. In fact, the spectrum has become severely broadened for $d(TC)_{15}$ (Figure 3B) suggesting that there are likely to be multiple structures associated with $d(TC)_{15}$ near the pK_a .

Temperature dependent studies were carried out to probe the stability of the structure associated with the various species. Figure 4 compares the 1D-spectra of $d(TC)_4$ and $d(CT)_4$ at different temperatures at pH 3.86. Notice that at 5°C there is a significantly higher proportion of the new species (e.g., resonance at 1.76 ppm) for the $d(TC)_4$ sequence than for the $d(CT)_4$ sequence. By raising the temperature to 25°C, the resonances from the new species have mostly disappeared. Further increasing the temperature only sharpens the peaks, but their chemical shifts remain largely unchanged. It is likely that the molecule is mostly single stranded at temperatures above 35°C.

Structural analyses by 2D-NOESY

What are the structures of the $(T-C)_n$ sequence at different pHs? We have carried out detailed 2D NMR analyses of $d(TC)_4$ and $d(CT)_4$ at three different pHs (3.0, 3.86 and 6.9) to attempt to sort out the answers. Figure 5 shows the complete 2D-NOESY spectra at 5°C. It can be seen that many crosspeaks are clearly observed. The assignment of the resonances was straightforward using standard sequential assignment procedure (30). With the exception of terminal residues, the repeat resonances from different sequence locations are not resolved due to overlap. Consequently, resonances have been assigned to types of

Figure 4. Temperature dependent studies of $d(TC)_4$ and $d(CT)_4$ at pH 3.86. Note that some broad resonances (e.g., at 1.76 ppm and 7.0 ppm) disappeared before 25°C.

nucleotides (C or T), instead of to a specific residue. Figure 6 shows some of the expanded regions of the 2D-NOESY spectra in which the key NOE crosspeaks are indicated. Most importantly, we observed a clear crosspeak between the CH6 (and CH5) and the TMe protons in both the pH 3.0 and pH 6.9

Figure 5. The complete experimental two-dimensional non-exchangeable proton NOESY spectra of $d(TC)_4$ (top two panels), and $d(CT)_4$ (bottom panel) at two pHs.

spectra. The intensity of this crosspeak (relative to other crosspeaks) is comparable to that observed in the CT sequence of the B-DNA decamer $d(CCAGGCCTGG)$ (data not shown) or the TC sequence of the B-DNA dodecamer $d(CGCGAATTCGCG)$ (data not shown). These results suggest that for both $d(TC)_4$ and $d(CT)_4$ the cytosine and thymine bases are stacked so that the NOE between CH6 and T-ME protons are observed ($<4 \text{ \AA}$). In the intermediate pH range, the intensity of this crosspeak decreases due to exchange among different conformations.

What are the possible structures? Under neutral pH conditions, the cytosine base is not protonated. Two base pairing schemes involving at least two hydrogen bonds are possible, a C:T pair (APS orientation) or a C:C pair (PS orientation) (Figure 1). We chose an anti-parallel APS duplex with full C:T pairs as the working model at pH 6.9 (model-I of Scheme 1). Scheme 1 includes only blunt-ended duplex models with at least two hydrogen bonds in the base pairs. The rotational correlation time τ_c ($\sim 2 \text{ nsec}$) of the two octamers and the spin-lattice relaxation time T_1 ($\sim 1-2.2 \text{ sec}$) of their protons suggest that the molecules are not aggregated (e.g., not concatenated forms due to sticky ends). While a symmetrical multistranded structure like a tetraplex (20) may exist, the low T_m of the molecule indicates otherwise.

Model I was refined using SPEDREF (21) to an NMR R-factor of 17%. Sections of the simulated spectrum are shown in Figure 6. However, it should be pointed out that due to the serious overlap of the resonances (because of the repeating nature of the sequence), the refinement is less definitive in providing a discrimination of the various models (e.g., Model-I and Model-V with C:C and T:T base pairs of Scheme 1). The refined Model-I has all the sugars in the C2'-endo conformation with average pseudorotation angles of 140° and average χ of -70° .

Under acidic conditions (pH=3.0), all cytosine bases are likely

Figure 6. Expanded regions of the experimental (top panels) and simulated (bottom panels) 2D-NOESY spectra of $d(TC)_4$. Key crosspeaks between the aromatic and H2'/H2'' are indicated. The crosspeaks between CH6 and T-methyl protons (arrow) are clearly visible, supporting the argument that the C and T bases are stacked.

Scheme 1.

to be protonated. In this case, (C⁺:T pair) a cis (APS) orientation is possible. A model incorporating fully protonated C⁺:T pairs (Model-III of Scheme 1) was constructed and subjected to refinement, resulting in an R-factor of 18%. Sections of the simulated spectrum are shown in Figure 6. The other possibility involves C⁺:C⁺ and T:T base pairs in the trans (PS) orientation (Model VII).

Under intermediate pH conditions, some of the cytosines are protonated. This may have several consequences. A hemi-protonated C⁺:C base pair in the trans (PS) orientation is now possible (Figure 1). This will force the thymine to form T:T base pairs which have two possible configurations. This PS helix (Model-IV of Scheme 1) is substantially different from the neutral (Model-I) and acid (Model-III) APS helices. Indeed, new resonances are observed, most notably those at the 1.76 ppm (T-methyl) and at 7.0 ppm (T-H6). This suggests that the PS Model-IV is in slow equilibrium with the APS Model-I and Model-III.

It is possible that other models exist. For example, some C:T pairs may now be protonated resulting in a partially protonated APS helix (e.g., Model-II or Model VI). The transition in going from Model-I to Model-III via Model-II is expected to have a low energy barrier, i.e., a rapid exchange among them. This was observed, as the common resonances associated with Model-I and Model-III remain sharp and their chemical shifts (except those of the H5 and H6 protons of cytosine bases) change very little.

Based on the existing data, we propose that the pH-dependent behavior throughout the pH range of the alternating (TC)_n sequence involves a complex equilibrium. Some steps in the equilibrium network are slow (e.g., between I and IV), while others are fast (e.g., between I and II). The multiple species in equilibrium at intermediate pH (3.8–5.0) make a quantitative structural analysis difficult. For example, the spectra of d(TC)₁₅ in this pH range are extremely broad (Figure 3B), reaffirming the suggestion of multiple species in equilibrium. In fact, we can not rule out a PS helix with continuously stacked C⁺:C base pairs and the thymidine nucleotides extruding into the solvent

Figure 7. Exchangeable proton spectra of d(TC)₄ at two pHs (top two spectra) and d(TC)₁₅ at pH 4.7 (bottom spectrum). The spectra were collected using the 1331 selective excitation technique (28) at 5°C.

region as suggested by Gray et al. (8–10). However, one thing that is clear is that an unusual PS helix can not be the *sole* structure at pH ~5 or the structure at pH=3 (11). Also, the equilibrium seen in the NMR data and pH titration can be accounted for only if the conformation proceeds from Model-I at neutral pH and goes to Model-III at pH 3.0. Our data suggests that the d(CT)_n polymer behaves the same way at least in a qualitative sense.

Exchangeable protons

Additional useful information may be obtained by studying the exchangeable protons. Figure 7 presents the 1D spectra of d(TC)₄ at two pH conditions (top two spectra). It can be seen that at pH 6.9 none of the exchangeable protons (T-imino and C-amino) was observed. This indicates that their exchange rate with bulk water was too fast. At pH 3.0, three new sets of resonances at 8.5, 9.2 and 11.3 ppm appeared. Their plausible assignments are the CNH6(external), CNH6(internal) amino and TNH3 imino protons respectively. The chemical shifts of the TNH3 imino protons are consistent with that proposed in other studies (31). Their exchange rate with bulk water was significantly slowed down by the acidic conditions, rendering them observable. However, despite extensive attempts, no resonance near 15 ppm, arising from a C⁺:T base pair could be detected. It is possible that the imino proton of the protonated C is still labile and exchanges rapidly with water proton.

Under intermediate pH conditions, the exchangeable proton spectrum of d(TC)₁₅ is extremely broad. Two major broad peaks, centered at ~9.9 ppm and ~11.5 ppm, were found. They may be tentatively assigned to C-amino and T-imino protons respectively. They are somewhat more down field than the C-amino and T-imino protons of the low pH form. These data are

again consistent with the notion that there are several conformational species in equilibrium at pH ~4–5.

CONCLUSION

Polypyrimidine repeats like (TC)_n have been shown to play biologically important roles in replication (32), recombination (33), transcription (34), and chromatin organization (35). Some d(TC)_n binding proteins have been isolated in different cell lines (36,37). Hence these polypyrimidine tracts are biologically relevant sequences and knowledge about their structure would be useful in understanding how they function.

Our studies here demonstrate that a simple repeating (C-T) sequence has a complex pH-dependent conformational behavior. Under neutral (pH 7) and acidic (pH 3) conditions, the (C-T)_n molecule probably adopts an anti-parallel helical structure with C:T and C⁺:T base pairs respectively. Significant NOE crosspeaks exist between adjacent cytosine and thymine residues in both forms, supporting well-defined structures (i.e., not single strand). The low pH structure is *not* consistent with the model in which continuously stacked C⁺:C base pairs are formed and the thymidine nucleotides extrude into the solvent region, as Sarma et al. have proposed (11). Under intermediate pH conditions, multiple conformational species coexist with different rates of exchange. While the model of Gray et al. (8–10) may be among the multiple species, it is not a major species. Whether the novel tetra-stranded structure associated with oligo-(dC) molecules, as proposed by Gehring et al. (20,21), exists for alternating (C-T) sequence remains to be tested. To obtain an unambiguous answer on the possible structures associated with (C-T)_n sequence, we have synthesized two oligonucleotides, d(CT(m⁵C)TCU(m⁵C)UCT) and d(CGACT(m⁵C)TCT), incorporating modified pyrimidine bases to disrupt the monotonic dinucleotide repeat. These two molecules should provide 2D-NMR spectra with significantly improved chemical shift dispersion which will permit a more definitive measurement of NOE crosspeak intensities. Because of the 5'-CGA motif (14), the latter molecule is expected to adopt a PS structure under slightly low pH conditions. NMR studies are underway.

ACKNOWLEDGMENTS

This work was supported by grants from NIH (GM-41612 and CA-52506) to AHJW. We thank Dr Howard Robinson for assistance and discussion.

REFERENCES

- Htun, H. & Dahlberg, J. E. (1988) *Science* **241**, 1791–1796.
- Johnston, B. H. (1988) *Science* **241**, 1800–1804.
- Pulleyblank, D. E., Haniford, D. B. & Morgan, A. R. (1985) *Cell* **42**, 271–180.
- Moser, H. E. & Dervan, P. B. (1987) *Science* **238**, 645–650.
- de los Santos, C., Rosen, M. & Patel, D. (1989) *Biochemistry* **28**, 7282–7289.
- Radhakrishnan, I., Gao, X., de los Santos, C., Live, D. & Patel, D. (1991) *Biochemistry* **30**, 9022–9030.
- Sklenar, V. & Feigon, J. (1990) *Nature* **345**, 836–838.
- Gray, M. D. & Vaughan, M. (1980) *Nucleic Acids Res.* **8**, 3695–3707.
- Brown, D. M., Gray, M. D., & Patrick, M. H. (1985) *Biochemistry* **24**, 1676–1683.
- Gray, M. D., Ratliff, R. L., Antao, V. P. & Gray, C. W. (1987) *J. Biomol. Struct. & Dyn.* **2**, 147–166.
- Sarma, M. H., Gupta, G. & Sarma, R. H. (1986) *FEBS Lett.* **205**, 223–229.
- Oligodeoxynucleotides—Antisense Inhibitors of Gene Expression* (1989) Edited by Cohen, J. S., CRC Press Inc., Boca Baton, FL.
- Robinson, H., van der Marel, G. A., van Boom, J. H. & Wang, A. H.-J. (1992) *Biochemistry* **31**, 10510–10517.
- Robinson, H. & Wang, A. H.-J. (1993) *Proc. Natl. Acad. Sci.* **90**, 5224–5228.
- Langridge, R. & Rich, A. (1963) *Nature* **198**, 725–728.
- Arnott, S., Chandrasekaran, R. & Leslie, A. G. W. (1976) *J. Mol. Biol.* **106**, 735–748.
- Luo, J., Sarma, M. H., Yuan, R. & Sarma, R. H. (1992) *FEBS Lett.* **306**, 223–228.
- Coll, M., Solans, X., Font-Altaba, M. & Subirana, J. A. (1987) *J. Biomol. Struct. Dyn.* **4**, 797–811.
- Cruse, W. B. T., Egert, E., Kennard, O., Sala, G. B., Salisbury, S. A. & Viswanitra, M. A. (1983) *Biochemistry* **22**, 1833–1839.
- Gehring, K., Leroy, J.-L. & Gueron, M. (1993) *Nature* **363**, 561–565.
- Leroy, J.-L., Gehring, K., Kettani, A. & Gueron, M. (1993) *Biochemistry* **32**, 6019–6031.
- Robinson, H. & Wang, A. H.-J. (1992) *Biochemistry* **31**, 3524–3533.
- Glase, P. K. & Long, F. A. (1960) *J. Phys. Chem.* **64**, 188–190.
- States, D. J., Haberkorn, R. A., & Ruben, D. J. (1982) *J. Magn. Reson.* **48**, 286–292.
- FELIX, Version 1.1*, (1992) Hare Research, Woodinville, WA.
- QUANTA, Version 3.2.3*, (1991) Polygen Corporation, 200 Fifth Avenue, Waltham, MA 02254.
- Brunger, A. T. (1992) *X-PLOR, Version 3.0*, The Howard Hughes Medical Institute and Yale University, New Haven, CT.
- Hore, P. J. (1983) *J. Mag. Reson.* **55**, 283–300.
- Il'yachova, I. A., Lysov, Y. P., Chernyi, A. A., Shchylolkina, A. K., Gottikh, B. P. & Florentiev, V. A. (1990) *J. Biomol. Struct. Dyn.* **7**, 879–897.
- Hare, D. R., Wemmer, D. E., Chou, S. H., Drobný, G. & Reid, B. R. (1983) *J. Mol. Biol.* **171**, 319–336.
- He, L., Kierzek, R., SantaLucia Jr, J., Walter, A. E. & Turner, D. H. (1991) *Biochemistry* **30**, 11124–11132.
- Caddle, M. S., Lussier, R. H. & Heintz, N. H. (1990) *J. Mol. Biol.* **211**, 19–33.
- Weinreb, A., Collier, D. A., Birshstein, B. K. & Wells, R. D. (1990) *J. Biol. Chem.* **265**, 1352–1359.
- Manor, H., Rao, B. S. & Martin, R. G. (1988) *J. Mol. Evol.* **27**, 96–101.
- Opstelten, R. J. G., Clement, J. M. E. & Wanka, F. (1989) *Chromosoma* **98**, 422–427.
- Gilmour, D. S., Thomas, G. H. & Elgin, S. C. R. (1989) *Science* **245**, 1487–1490.
- Yee, H. A., Wong, A. K. C., van de Sande, J. H. & Rattner, J. B. (1991) *Nucleic Acids Res.* **19**, 949–953.