ELSEVIER

FEBS Letters 337 (1994) 139–144

FEBS 13511

FEBS LETTERS

Conformations of the alternating $(C\text{-}T)_n$ sequence under neutral and low pH

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

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

Received 13 October 1993; revised version received 29 November 1993

Abstract

The structures of the $(C\text{-}T)_n$ sequence at two different pHs have been analyzed by 500 MHz 2D-NMR using a modified DNA decamer d(CT[m^{5}C]TCU[m^{5}C]UCT) as a model system. The chemical modifications serve to perturb the monotonous C-T repeat, and consequently to yield a better chemical shift dispersion. The results reinforce our earlier suggestion that there are three major pH-depedent conformational species: two antiparallel-stranded (APS) duplexes at pH 7 and pH 3, and a different structure near pH 5. Structural refinement of the decamer duplexes at pH = 7.5 and pH = 2.9 using 2D-NOE data suggests that the C : T or C$^+$: T base pairs are continuously stacked. Exchangeable proton NMR spectra at pH 7.5 and pH 2.9 are consistent with C : T or C$^+$: T base pairing schemes in which a water molecule bridges the two bases.

Key words: Unusual DNA conformation; Modified DNA base; Nuclear Overhauser effect

1. Introduction

Polypyrimidine sequences are frequently found in DNA and they play important biological roles such as in replication [1], recombination [2], transcription [3], and chromatin organization [4]. Also, repetitive $(GA_n \cdot (TC)_n$ sequences are known to readily form triple-stranded DNA or H-DNA especially under low pH conditions [5–8]. Interestingly, a number of proteins have been found to bind quite specifically to the pyrimidine $(C\text{-}T)_n$ sequence [9,10]. These observations raise a tantalizing prospect that these sequences alone (without their complementary strand) may have well-defined structures.

As part of a program to study unusual nucleic acid structures, we have recently analyzed by NMR the structural consequences of the alternating C-T sequence under the influence of pH, using $d(CT)_4$, $d(TC)_4$ and $d(TC)_{15}$ as model systems [11]. Our results revealed an interesting pH-dependent conformational behavior associated with these molecules. Under neutral (pH 7.5) and acidic (pH 2.9) conditions, the $(C\text{-}T)_n$ molecule likely adopts anti-parallel duplex structures with C : T and C$^+$: T base pairs, respectively. Under intermediate pH conditions, multiple conformational species coexist with different rates of exchange.

However, due to the monotonous repeating (C-T) sequence it was not possible to have an unambiguous measurement of individual NOE crosspeak intensities, thereby preventing a definitive three dimensional structural analysis. To circumvent this problem, we have synthesized the modified DNA decamer d(CT[m^{5}C]TCU-[m^{5}C]-UCT) and subjected it to 2D-NMR analyses. The chemical modifications at the C^5 position of the 5-methylcytosine (replacing cytosine) and uracil (replacing thymine) do not perturb the base pairing capability. Here we show that the modified decamer yielded a significantly better chemical shift dispersion, allowing a better resolution of the NOE crosspeaks. The structural analysis was carried out on the two decamer duplexes at pH 2.9 and 7.5 by the NOE-constrained refinement procedure SPEDREF [12].

2. Materials and methods

The oligonucleotide was synthesized by an Applied Biosystems DNA synthesizer using the normal and the modified pyrimidine nucleoside phosphoramidites as building blocks. The deblocked oligonucleotide was purified by Sepharose G50 column chromatography. The purity of the preparation was judged to be greater than 95% by reverse phase (C4 column) HPLC and NMR analyses. Samples for ^{1}H NMR analysis were prepared in D_2O as described earlier [12]. The concentration of the decamer was 2.5 mM duplex in 150 mM NaCl and the solution was titrated to desirable pHs as described earlier [11]. The 2D ^{1}H NMR spectra were recorded on a Varian VXR500 500 MHz spectrometer equipped with a Nalorac inverse detection probe. The 1D ^{1}H NMR spectra were recorded on a GE GN500 500 MHz spectrometer. The proton 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 162 MHz ^{31}P NMR spectra were collected on a Varian Unity400 spectrometer, with phosphoric acid as an external reference. Models incorporating various C : T base pairs were built

*Corresponding author. Fax: (1) (217) 244 3181.

SSDI 0014-5793(93)E1428-O

using QUANTA [13] and X-PLOR [14]. The m^5C and the U residues were built in using existing force fields for the substituted atom types. The exchangeable proton 1D spectra were collected using the selective excitation $\bar{1}33\bar{1}$ pulses to excite resonances ranging from 9 ppm to 17 ppm as described by Hore et al. [15]. Two-dimensional NOESY spectra were collected by the method described in the legend of Fig. 2.

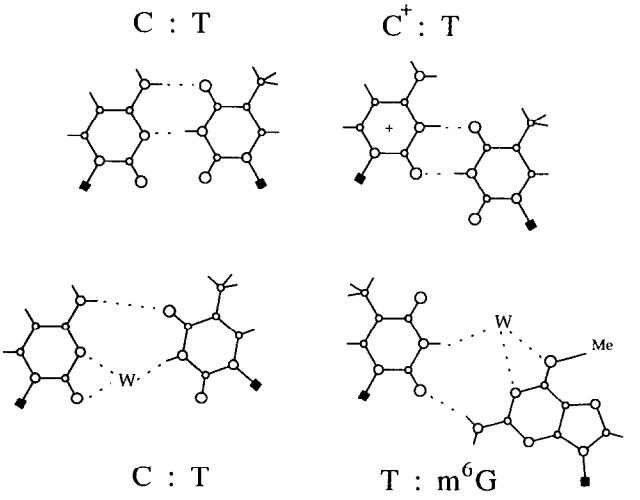
3. Results and discussion

The pH-dependent structures associated with the decamer $d(C_1T_2[m^5C_3]T_4C_5U_6[m^5C_7]U_8C_9T_{10})$ were first probed by 1D-NMR. The relevant C-T base pairs are shown in Scheme 1. The non-exchangeable proton 1D spectra at different pHs at 5°C are shown in Fig. 1A. At neutral pH (6.5–7.5) and acidic pH (2.8–3.0), the spectra are reasonably well resolved. More importantly, the chemical shift dispersion is definitely improved as compared to that of the unmodified $d(C\text{-}T)_4$ [11]. The assignment of the resonances has been done by 2D-NOESY spectra described later. The corresponding resonances from the acid pH form generally have more downfield chemical shifts than those of the neutral pH form. At intermediate pH (3.8–5.0), a new set of broad resonances (e.g., ~1.76 ppm) appears and co-exists with those from the extremes of the pH ranges. We have not been able to convert the molecule into this new species completely, consequently no further structural analysis was possible.

We have carried out detailed 2D NMR analyses of the decamer at pH 2.9 and 7.5. Inspection of the complete 2D-NOESY spectra at 5°C revealed many crosspeaks (Fig. 2A), suggesting well-defined structures. The assignment of the resonances was straightforward using standard sequential assignment procedures [16]. Fig. 2B shows some of the expanded regions of the 2D-NOESY spectra in which several key NOE crosspeaks are indicated.

Even though there is still some overlap between the resonances of sugar protons, the resonances of the aromatic base protons and the methyl protons are fairly well resolved and the spectra at both pHs show the internucleotide connectivities characteristic of a continuously stacked B-DNA-like helical structure. We observed clear crosspeaks between the T/C/U H^6 (and C/U H^5) and the T/m^5C Me protons in both the pH 2.9 and pH 7.5 spectra. For example, the crosspeaks between T_2H^6-m^5C_3Me, U_6H^6-m^5C_7Me, T_4Me-$m^5C_3H^6$ and C_1H^6-T_2Me are visible in both spectra. These results suggest that in the decamer at both pHs the C/m^5C and T/U pyrimidine bases are stacked so that the NOE crosspeaks betweenthe T/C/U H^6 and T/m^5C Me protons are observed (~4 Å).

At neutral pH, our previous analysis of $d(C\text{-}T)_4$ and



Scheme 1. Different C:T base pairing schemes in the cis orientation among cytosine (C), N^3-protonated cytosine (C^+) and thymine (T). Water-mediated base pairs are shown for the C:T and T:m^6G mismatches.

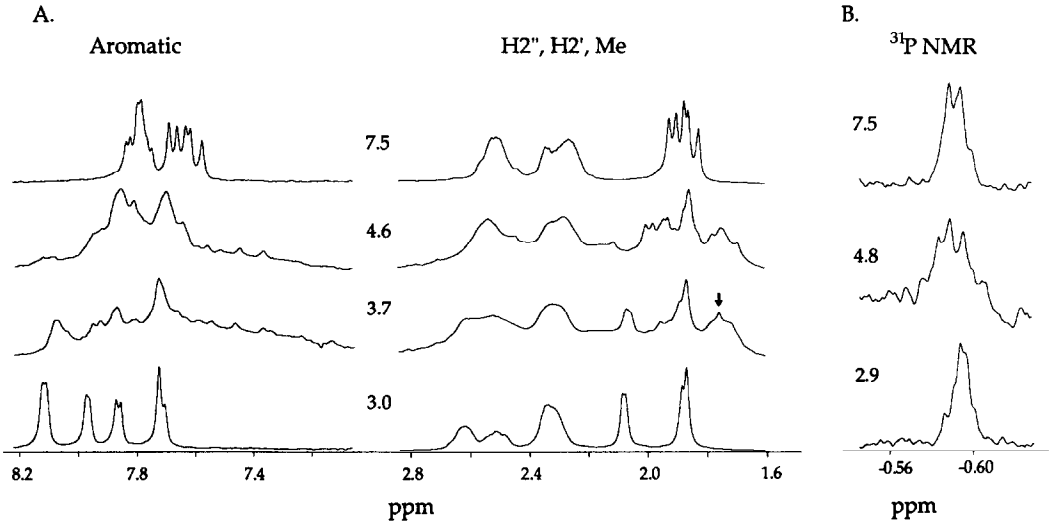


Fig. 1. One dimensional non-exchangeable proton 500 MHz NMR spectra of DNA decamer d(CT[m^5C]TCU[m^5C]UCT) at pH 2.9, 4.6 and 7.5. The 1D 162 MHz ^{31}P NMR spectra of the decamer at 3 pHs are also shown.

d(T-C)$_4$ suggested that an anti-parallel APS duplex with entirely C:T pairs (see Scheme 1) is a plausible model. The C:T base pairing scheme may involve two hydrogen bonds, with the HN^3 imino proton of T bonded to the N^3 of the unprotonated C. We have recorded the 1D exchangeable proton NMR spectra of the modified decamer (Fig. 3A) and d(C-T)$_{15}$ 30-mer (Fig. 3B) at two pHs. It can be seen that at pH 7.5, the T/U imino proton resonance has disappeared. In contrast, the amino protons of C/m^{5}C are clearly visible with the chemical shifts centered around 6.7 ppm and 7.5 ppm. Non-hydrogen-bonded C amino protons have been observed by others at ~6.5 ppm [17,18] and so the peaks at 7.5 ppm are probably from hydrogen-bonded C amino protons. This observation is consistent with the formation of a C-T base pair. Interestingly, the T/U imino protons showed up strongly as a broad singlet ($\Delta\omega$ = 55 Hz) at 11.4 ppm in the spectra of both molecules at pH 2.9 (top panels, Fig. 3). Non-hydrogen-bonded T imino protons have been seen by Kalnik et al. at ~8.85 ppm [19]. The chemical shift of 11.4 ppm may indicate hydrogen bonding. Indeed, SantaLucia et al. [20] observed such a resonance in C$^+$: U base pair at 11.4 ppm in r(GCGCUCGC) at pH 5.0. The C/m^{5}C amino proton resonances move significantly downfield to center around 8.5 ppm and 9.3 ppm, respectively. The downfield shift is consistent with the protonation of the C/m^{5}C bases. However, we could not detect any signal of the imino proton from the protonated C/m^{5}C in the region near 15 ppm, as observed in the C$^+$: C base pairs [21–23]. It is possible that the C$^+$ imino resonance in the C$^+$: T base pair should be more upfield due to it being hydrogen bonded to T-O^4 oxygen. However, we do not see a separate peak different from 11.4 ppm for this proton. Though unlikely, it may be that the T imino and C$^+$ imino protons show very similar chemical shifts (11.4 ppm), therefore only a singlet was observed.

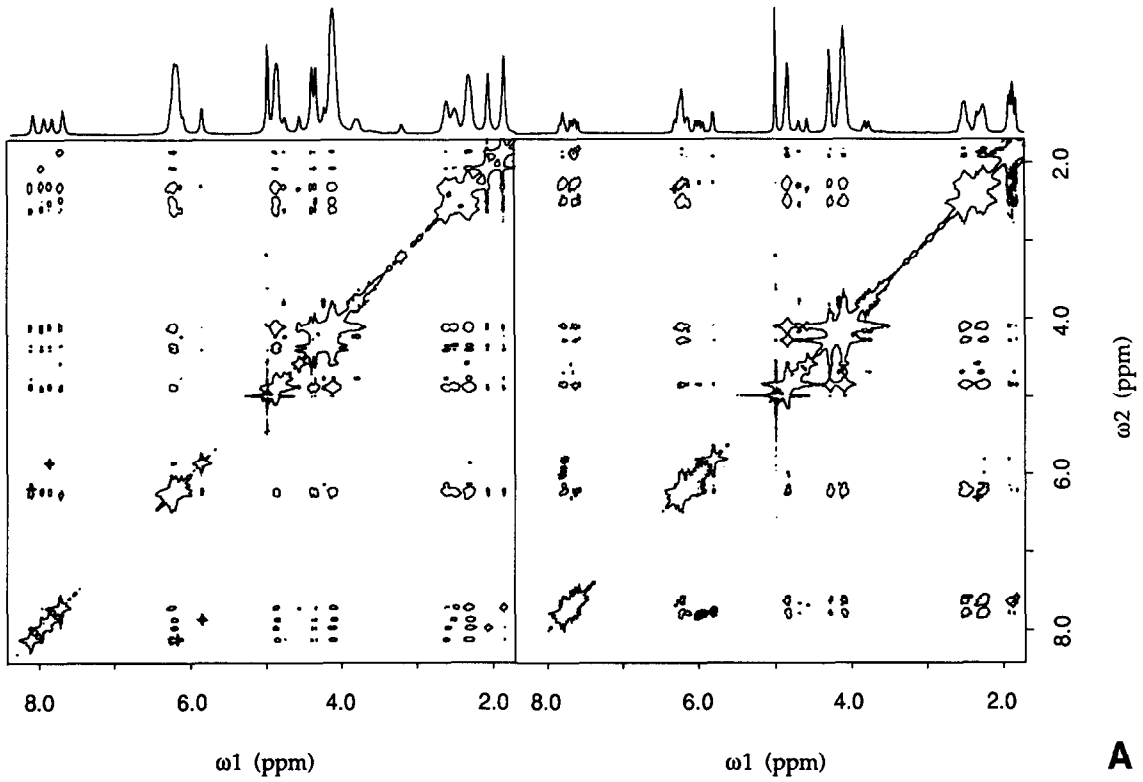
We interpret the results as follows. At neutral pH, the exchange rate of the T/U imino protons with bulk water is fast. This happens likely because the C:T base pair with two hydrogen bonds makes the two O^2 keto oxygens too close (3.1 Å). A water molecule is able to help stabilize the base pair by bridging the two bases as shown in Scheme 1. This type of water-mediated C:U base pair has been observed in the crystal structure of r(CGACUUCGGUCC)$_2$ [24]. That decamer crystallized as a duplex, instead of a hairpin, with two C:U mismatches. A bridging water was clearly visible bonded simultaneously to N^3 of C and N^3 of U. Our exchangeable proton NMR data augment previous NMR data of d(CGCGAATTCTCG) by Patel et al. [25]. In fact, a similar TM : m^{6}G base pair with one strong hydrogen bond and one weak hydrogen bond was proposed by Kalnik et al. [19]. The same structural reason, i.e. the close van der Waals clash of the two keto oxygens (O^4 of T and O^6 of m^{6}G) causing the base pair to open up, may be operative here. This contention was further confirmed by the analysis of an oligonucleotide using a ^{15}N-labeled (at the N^3 of thymine) T opposing an m^{6}G [26]. It is therefore plausible that the (C-T)$_n$ sequence adopts an antiparallel duplex conformation incorporating water-mediated C : T base pairs.

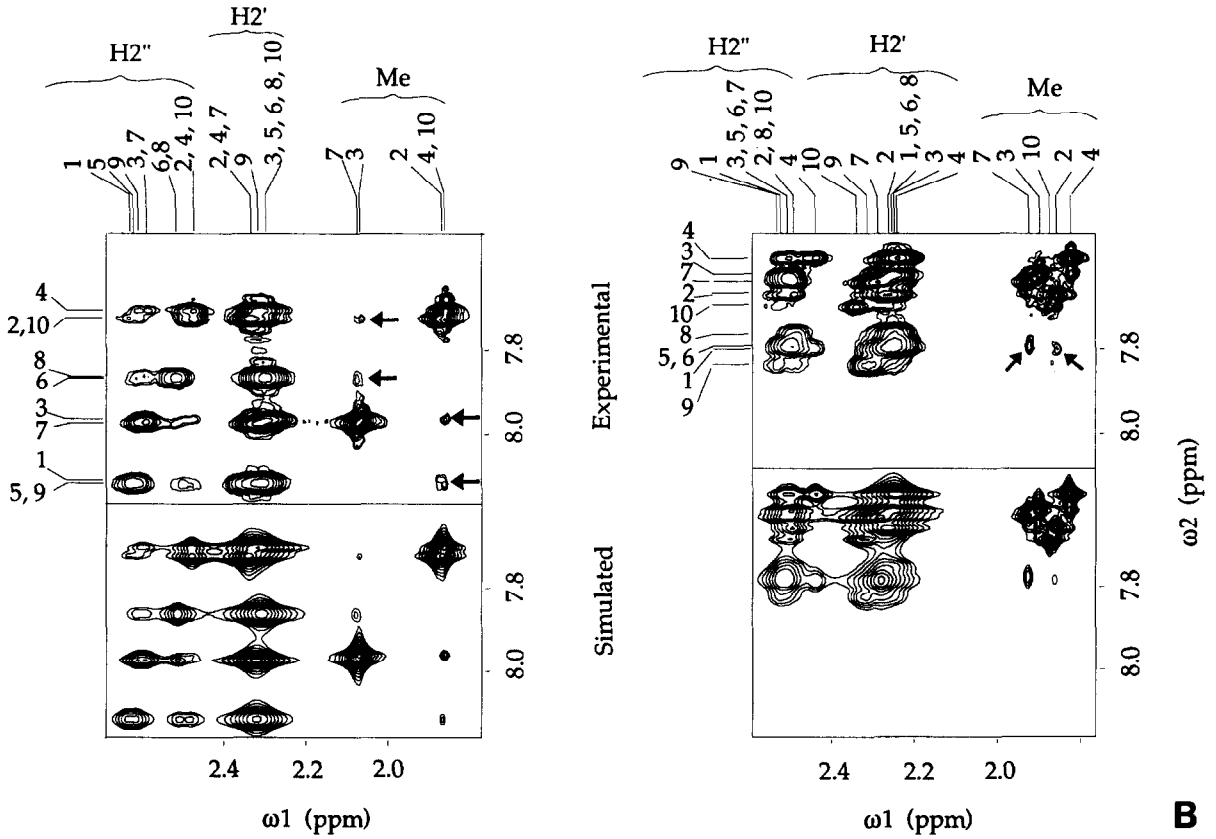
At pH 3.0, the exchange rate with bulk water of all exchangeable protons, except possibly the imino proton of the protonated C$^+$, was significantly retarded by the acidic conditions. The absence of the C$^+$ imino proton resonance implies that it is significantly more labile than the imino proton of T and still exchanges rapidly with water protons.

Models of the decamer duplexes for the low pH and neutral pH conditions were constructed, using the C$^+$: T and C : T base pairs (without the bridging water), respectively. Both models were refined using SPEDREF [12] to NMR R-factors of 17% and 18%, respectively. Sections of the simulated spectra are shown in Fig. 2B. Since some overlaps of the resonances (because of the repeating nature of the sequence) still exist, the refinement is less definitive. Nonetheless the refinements provide some clue about the conformations since base stacking information is fairly clear. Fig. 4 shows a refined model of the low pH form. The refined model has most of the sugars in the C2′-*endo* conformation. However, the base stacking patterns of the CpT steps and the TpC steps are different causing the C/m^{5}C residues and T/U residues to have different conformations. The difference between the CpT and the TpC stacking patterns is more evident in case of the neutral pH structure. This observation is also consistent with our 1D ^{31}P NMR data which shows two different chemical shifts at neutral pH (Fig. 1B). The average pseudorotation (P) angles and the glycosyl χ for C and T are (P = 130°, χ = −50°) and (P = 140°, χ = −75°), respectively. These values may not be very reliable due to the overlap in the chemical shifts of the sugar protons. The difference between the roll angles for the alternate TpC and the CpT step stacking patterns is ~15° in the neutral pH structure.

Under intermediate pH conditions, some of the cytosines are protonated. This may cause hemi-protonated C$^+$: C base pairs in the trans (PS) orientation, thus forcing the thymines to form T : T base pairs. This structure is substantially different from the neutral pH and acidic pH APS helices, as evident from the new resonances near 1.76 ppm (T-methyl) and at 7.0 ppm (T-H^6). This suggests that the new structure is in slow exchange with the APS models on the NMR time scale.

We have attempted to trap the structure associated with (C-T)$_n$ at intermediate pH by synthesizing d(CGACT[m^{5}C]TCT) which may form an unique PS duplex on the following premise. We have shown recently that DNA sequences containing 5′-CGA have a strong propensity to form a parallel-stranded homo-base


8.0
6.0
4.0
2.0
ω1 (ppm)
ω2 (ppm)
A


H2"
H2'
Me
Experimental
Simulated
2.4
2.2
2.0
ω1 (ppm)
ω2 (ppm)
B

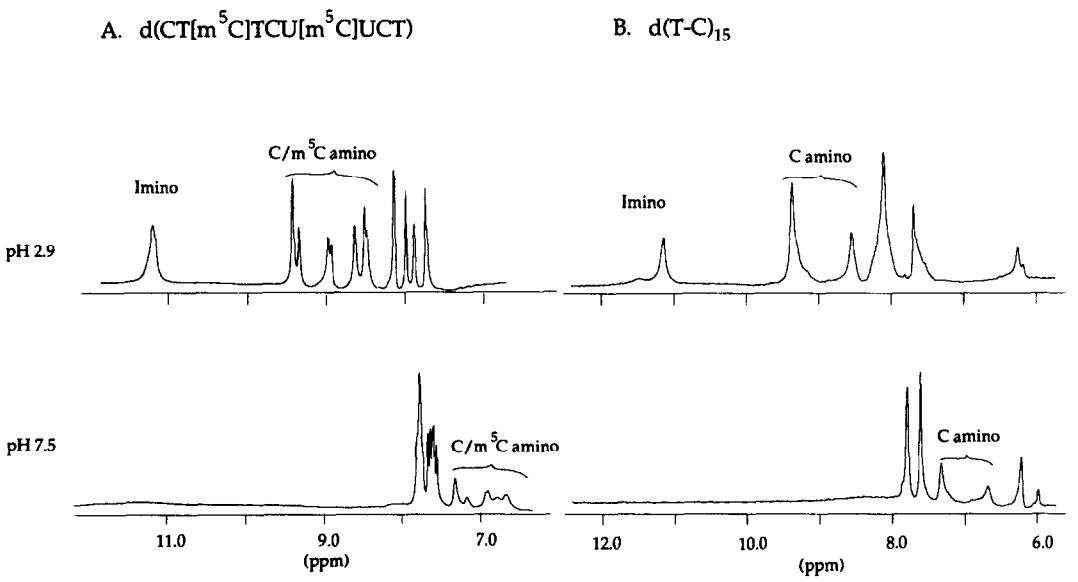


Fig. 3. Exchangeable proton spectra of (A) d(CT[m^{5}C]TCU[m^{5}C]UCT) and (B) d(TC)$_{15}$ at two pHs. The spectra were collected using the $1\bar{3}3\bar{1}$ selective excitation technique [15] at 5°C.

paired double helix under low pH conditions [23,27]. For example, the heptamer d(CGACGAC) forms a remarkably stable PS-DNA duplex at pH 6.0 [23]. The structure is stabilized considerably by the hemi-protonated C : C$^+$ base pair and the extensive *inter-strand* stacking of the G : G base pair over the A:A base pair. Because of the 5′-CGA motif in the d(CGACT[m^{5}C]TCT) sequence, the molecule is expected to adopt a PS structure under low pH conditions [23]. Surprisingly this 9-mer DNA forms a very stable hairpin structure (unpublished data). This has caused the molecule to have multiple conformations even at low pH, as evident by the broad NMR spectrum, rendering the structural analysis of the low pH form impossible.

4. Conclusion

Our results show that the alternating (T-C)$_n$ sequence, incorporated with many modified bases, has a complex pH-dependent conformational behavior, as that in the unmodified sequence. Our present work provided further information on the dynamic property of the base pair. The water mediated interaction appears to exist in both the C : T and C$^+$: T base pairs. Under neutral pH conditions, the (C-T)$_n$ sequence forms a duplex structure similar to that shown in Fig. 4, but the base pairs readily open up to accommodate the bridging water. Whether the (C-T) binding proteins recognize the (C-T)$_n$ sequence in its duplex form or single-stranded form remains to be determined.

Acknowledgments: This work was supported by grants from NIH (GM-41612 and CA-52506) to AHJW. The Varian VXR500 spectrometer was supported by NIH shared instrumentation Grant 1S10RR06243. We thank Dr. H. Robinson and Dr. V. Mainz for their assistance.

←

Fig. 2. (A) Complete 2D-NOESY spectra of d(C$_1$T$_2$[m^{5}C]$_3$T$_4$C$_5$U$_6$[m^{5}C]$_7$U$_8$C$_9$T$_{10}$) at pH 2.9 and 7.5. Many NOE crosspeaks are observed, suggesting well-defined structures under the experimental conditions. (B) Expanded regions of the experimental (top panels) and simulated (bottom panels) 2D-NOESY spectra. Key crosspeaks between the aromatic and H$^{2'}$/H$^{2''}$ are indicated. The numbers correspond to the residue numbers in the sequence from the 5′ side. The crosspeaks between H^6 and methyl protons (arrow) are clearly visible, supporting the argument that the C/m^{5}C and T/U bases are stacked. The 2D-NOESY spectra were collected by the method of States et al. [28] 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 4.14 s. The 2D data sets were processed with the program FELIX [29] on Silicon Graphics workstations using the truncation apodization function described earlier [12] and an exponential multiplication at 5 Hz. The intensities of the NOE crosspeaks were measured by the program MYLOR [12]. There were ~ 900 NOE crosspeaks (many are still overlapped), considered to be above the noise level and they were used in the refinement. An iterative spectral-driven procedure SPEDREF [12] using the full-matrix relaxation theory was used for the refinement.

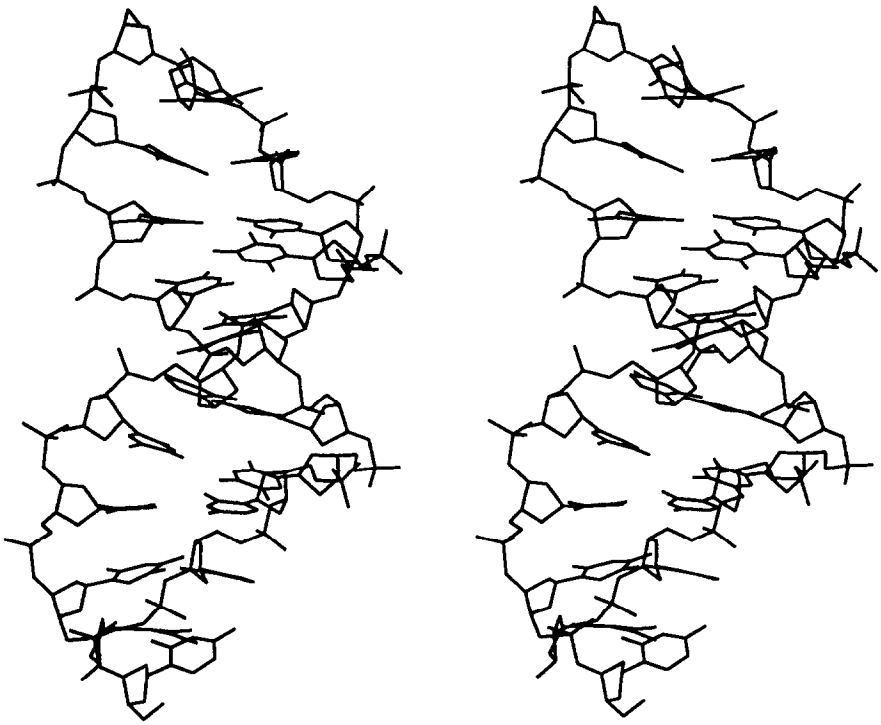


Fig. 4. NOE-refined model of the (C-T) decamer at pH 2.9.

References

[1] Caddle, M.S., Lussier, R.H. and Heintz, N.H. (1990) J. Mol. Biol. 211, 19–33.
[2] Weinreb, A., Collier, D.A., Birshtein, B.K. and Wells, R.D. (1990) J. Biol. Chem. 265, 1352–1359.
[3] Manor, H., Rao, B.S. and Martin, R.G. (1988) J. Mol. Evol. 27, 96–101.
[4] Opstelten, R.J.G., Clement, J.M.E. and Wanka, F. (1989) Chromosoma 98, 422–427.
[5] Htun, H. and Dahlberg, J.E. (1988) Science 241, 1791–1796.
[6] Johnston, B.H. (1988) Science 241, 1800–1804.
[7] Pulleyblank, D.E., Haniford, D.B. and Morgan, A.R. (1985) Cell 42, 271–280.
[8] Moser, H.E. and Dervan, P.B. (1987) Science 238, 645–650.
[9] Gilmour, D.S., Thomas, G.H. and Elgin, S.C.R. (1989) Science 245, 1487–1490.
[10] Yee, H.A., Wong, A.K.C., van de Sande, J.H. and Rattner, J.B. (1991) Nucleic Acids Res. 19, 949–953.
[11] Jaishree, T.N. and Wang, A.H.-J. (1993) Nucleic Acids Res. 21, 3839–3844.
[12] Robinson, H. and Wang, A.H.-J. (1992) Biochemistry 31, 3524–3533.
[13] QUANTA, Version 3.2.3, (1991) Polygen Corporation, 200 Fifth Avenue, Waltham, MA 02254.
[14] Brunger, A.T. (1992) X-PLOR, Version 3.0, The Howard Hughes Medical Institute and Yale University, New Haven, CT.
[15] Hore, P.J. (1983) J. Mag. Reson. 55, 283–300.
[16] Hare, D.R., Wemmer, D.E., Chou, S.H., Drobny, G. and Reid, B.R. (1983) J. Mol. Biol. 171, 319–336.
[17] Kalnik, M.W., Li, B.F.L., Swann, P.F. and Patel, D.J. (1989) Biochemistry 28, 6182–6192.
[18] Carbonnaux, C., van der Marel, G.A., van Boom, J.H., Guschbauer, W. and Fazakerley, G.V. (1991) Biochemistry 30, 5449–5458.
[19] Kalnik, M.W., Li, B.F.L., Swann, P.F. and Patel, D.J. (1989) Biochemistry 28, 6170–6181.
[20] SantaLucia, Jr., J., Kierzek, R. and Turner, D.H. (1991) Biochemistry 30, 8242–8251.
[21] Gehring, K., Leroy, J.-L. and Gueron, M. (1993) Nature 363, 561–565.
[22] Luo, J., Sarma, M.H., Yuan, R. and Sarma, R.H. (1992) FEBS Lett. 306, 223–228.
[23] Robinson, H. and Wang, A.H.-J. (1993) Proc. Natl. Acad. Sci. USA 90, 5224–5228
[24] Holbrook, S.R., Cheong, C., Tinoco, I. and Kim, S.-H. (1991) Nature 353, 579–581.
[25] Patel, D.J., Kozlowski, S.A., Ikuta, S. and Itakura, K. (1984) Fed. Proc. Fed. Am. Soc. Exp. Biol. 43, 2663–2670.
[26] Goswami, B., Gaffney, B.L. and Jones, R.A. (1993) J. Am. Chem. Soc. 115, 3832–3833.
[27] Robinson, H., van der Marel, G.A., van Boom, J.H. and Wang, A.H.-J. (1992) Biochemistry 31, 10510–10517.
[28] States, D.J., Haberkorn, R.A. and Ruben, D.J. (1982) J. Magn. Reson. 48, 286–292.
[29] FELIX, Version 1.1, (1992) Hare Research, Woodinville, WA.