Volume 205, number 2 FEBS 3988 September 1986

A cytosine · cytosine base paired parallel DNA double helix with thymine · thymine bulges

M.H. Sarma, G. Gupta and R.H. Sarma

Institute of Biomolecular Stereodynamics, State University of New York at Albany, Albany, NY 12222, USA

Received 4 July 1986

500 MHz ^{1}H NMR studies using 2D-NOESY indicate that the oligonucleotide d(CTCTCT) at low pH forms a parallel double helix with cytosine · cytosine base pairs and thymine · thymine bulges. This unusual structure may explain the hypersensitivity of S_1 nuclease at low pH towards supercoiled plasmids containing $d(CT)_n$ inserts.

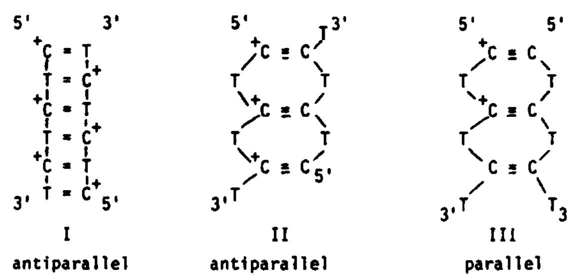
^{1}H-NMR 2D-NOESY Oligonucleotide structure

1. INTRODUCTION

There is a revolution taking place in DNA structural studies. Single-crystal data have demonstrated the totally unexpected presence of Hoogsteen G^+-C pairs when the oligonucleotides are crystallized under normal pH conditions [1]. A · G pairs have been known to be accommodated within a double helix without major structural distortion ([2–4], see, e.g. fig.5 in [2]). In all these structures the double helix with the mismatch forms an anti-parallel helix. In this context, it is of great interest to note that an unusual anti-parallel helix with C · C mismatches and T · T bulges has been advocated as a theoretical possibility to rationalize the hypersensitivity of S_1 nuclease at low pH towards certain supercoiled plasmids containing $d(CT)_n$ inserts [5]. The advent of 2D-NMR spectroscopy enables one to determine experimentally whether such unusual structures are true.

2. EVIDENCE THAT d(CTCTCT) UNDER LOW pH FORMS A BASE-PAIRED DOUBLE HELIX

The oligonucleotide d(CTCTCT) was synthesized by the phosphotriester method. Under low pH, this molecule can in principle take up the following three structural motifs, hereafter referred as helix I, II and III.



Note that the Cs are hemiprotonated under low pH. The hydrogen-bonding patterns involved with $C^+ \cdot C$ and $C^+ \cdot T$ are illustrated in fig.1C as an inset. In fig.1A we show the low-field region of the 500 MHz ^{1}H NMR spectrum of d(CTCTCT) at pH 3.0 in H_2O. The resonances marked * arise from hydrogen-bonded pyrimidine · pyrimidine pairs and these resonances disappeared when the spectrum was taken in H_2O at pH 7.0. The data clearly reveal that lowering the pH causes the pyrimidines to form hydrogen-bonded base pairs. The resonance at 11.10 ppm originates from the N3H imino proton; those at 9.24 and 8.4 ppm must

Published by Elsevier Science Publishers B.V. (Biomedical Division)
00145793/86/$3.50

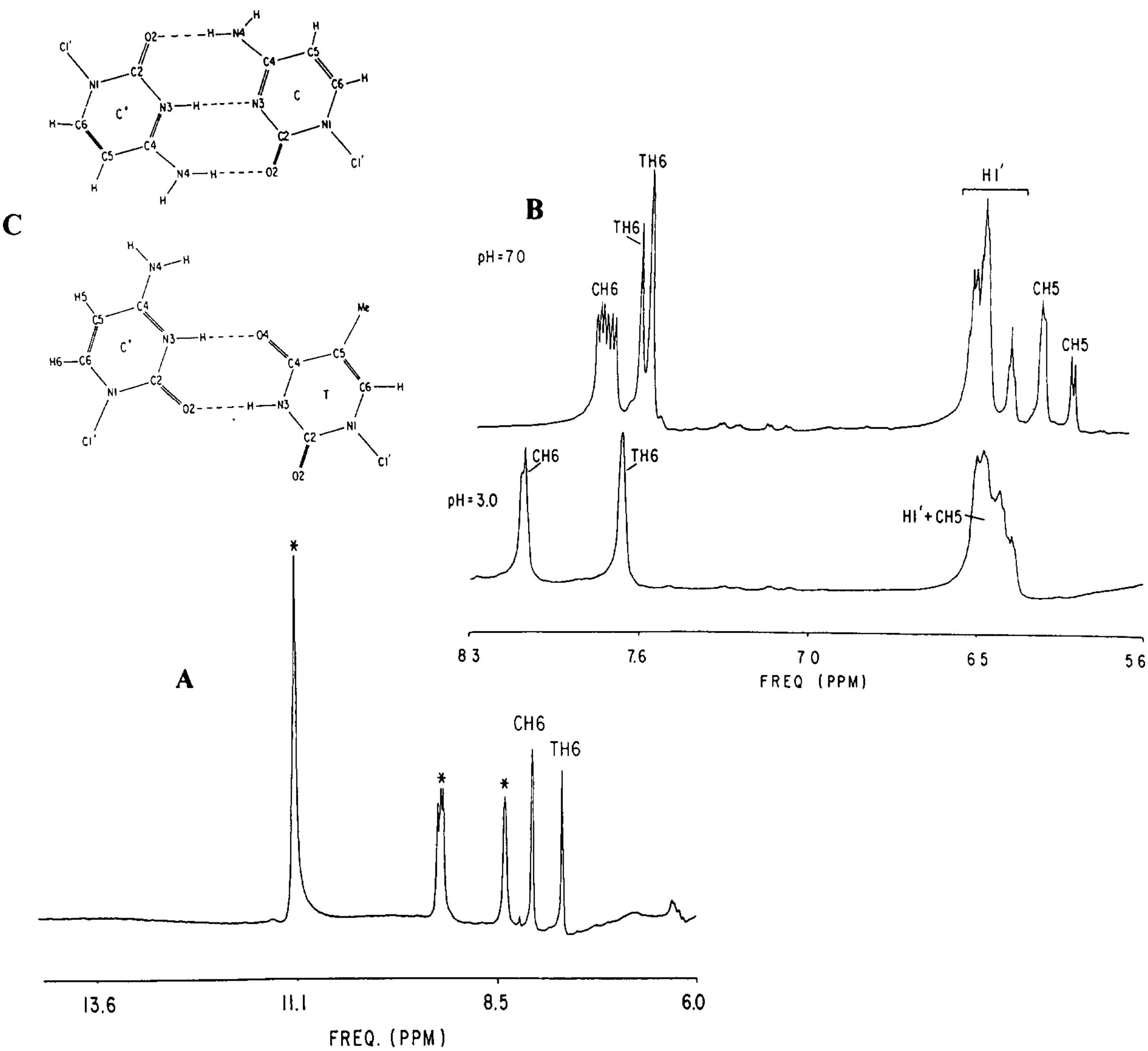


Fig.1. (A) 500 MHz ^{1}H NMR spectrum (low-field region) of d(CTCTCT) in H_2O at pH 3.0. At pH 7.0, the signals marked * disappear, suggesting that at low pH they originate from pyrimidine-pyrimidine hydrogen-bonded base pairs (see text for details). (B) 500 MHz ^{1}H NMR spectra (low field) of d(CTCTCT) at pH 7 (top) and pH 3 (bottom). For complete spectrum, see top of fig.2. (C) Possible pyrimidine·pyrimidine hydrogen-bonded pairing schemes in d(CTCTCT).

originate from internal and/or external amino proton (a total of six participating protons) hydrogen bonds [6,7]. It should be mentioned that the areas and intensities of the resonances cannot be used to estimate the relative number of protons because of the use of a time-shared long pulse sequence to suppress water and that the positions are different from those expected for normal WC pairs because of the protonation of C. In fig.1B we show the low-field region of the D_2O spectra of the hexamer at pH 7.0 and 3.0 (for the complete spectrum at pH 3.0 see top and side of fig.2). Dramatic differences exist between the spectra at pH 7.0 and 3.0. At pH 7.0 when the hexamer is single-stranded, the

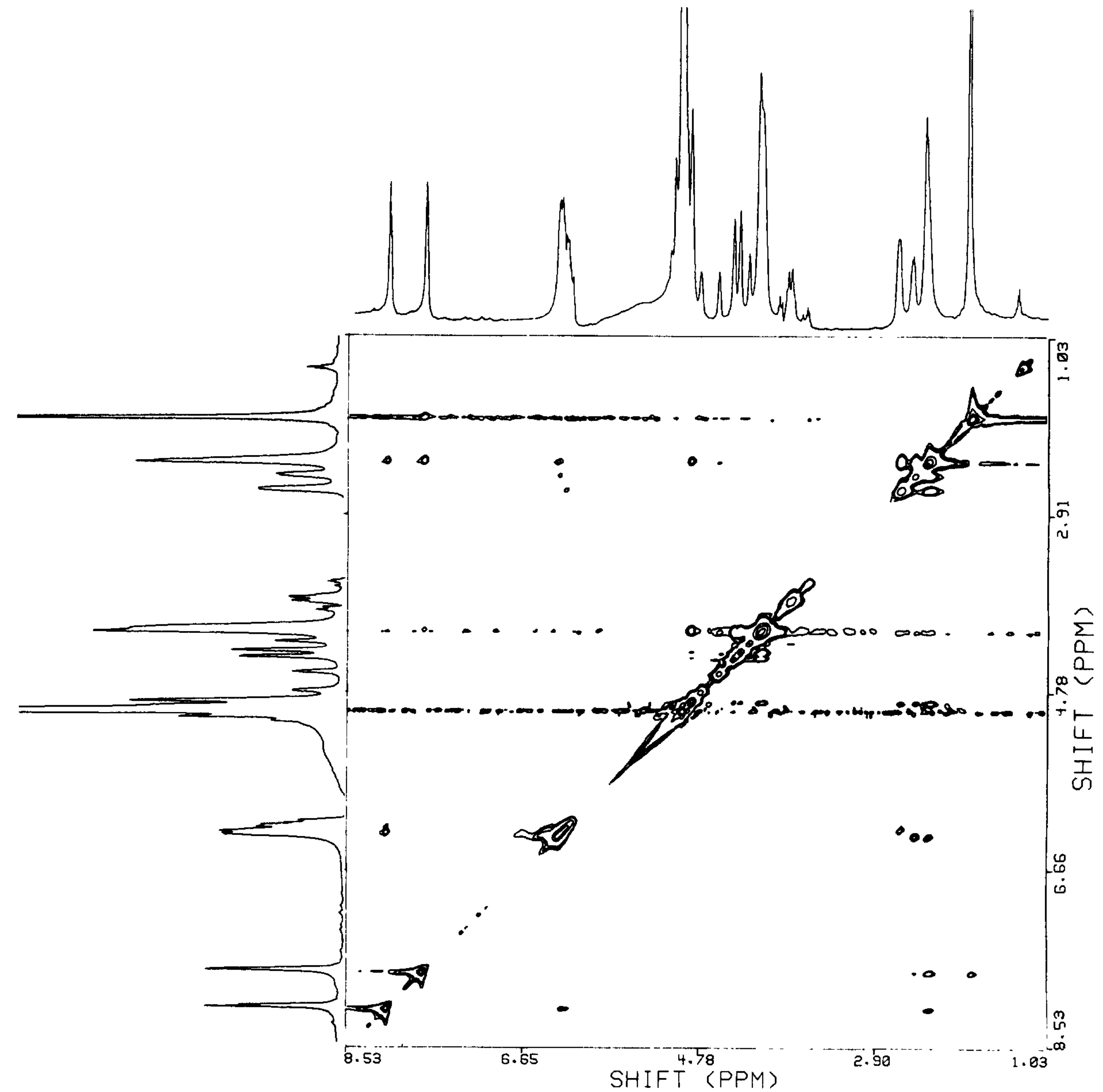


Fig.2. 2D-NOESY spectrum of d(CTCTCT)$_2$ in D_2O at 5°C at pH 3. The spectrum was recorded in the pure absorption mode [14] using a pulse-sequence (90°-t_1-90°-τ_m-90°Acq)$_n$. 512 free induction decays of 1024 data points each were acquired. The phase cycling scheme of States et al. [14] was used and 64 experiments were performed for each t_1 value using 1 s of relaxation delay. HDO was presaturated. The time domain data were processed on a VAX computer; free induction decays were weighed with an exponential multiplication factor of 4 Hz before Fourier transformation and phase correction. Because of the use of the phase cycling scheme of States et al. [14], the cross-peaks in this figure and the 1D projections in fig.3 can be interpreted quantitatively.

resonances are sharper and even the couplings are visible; at pH 3.0, the formation of the double helix and base pairing cause the resonances to shift, broaden and merge.

3. EVIDENCE THAT d(CTCTCT) UNDER LOW pH DOES NOT FORM HELIX I WITH $C^+ \cdot T$ PAIRS, BUT HELIXES II/III WITH $C^+ \cdot C$ PAIRS

The data presented so far do not enable one to distinguish among the three possible helixes I, II and III. In helixes II and III, the $C^+ \cdot C$ pairs are stacked on the top of each other while all Ts bulge out, i.e. there is no spatial interaction between neighboring C and Ts. The situation is completely different in helix I because here neighboring C and T are stacked on top of each other. Thus, one could ascertain the presence or absence of helixes I, II/III by examining the spatial interactions between C and T in d(CTCTCT) at low pH. The complete 2D-NOESY spectrum of the oligomer at low pH is shown in fig.2. A few 1D projections from the NOESY spectrum (fig.2) are shown in fig.3. Of particular interest is projection A which shows the cross-peaks between TCH3 and other protons. Note that the only cross peak that is observed for TCH3 is to the one located at TH6 (because TH6 and TCH3 in the same base are ~3 Å apart). If helix I were present in solution, we should have seen a cross-peak between TCH3 and neighboring CH6 which is a persistent characteristic of a duplex in which next-neighbors are stacked [8,9]. The data in projection A of fig.3 are corroborated by projection C which shows the cross-peaks of CH6 with other protons and here again one sees no cross-peaks between CH6 and TCH3. Lack of NOE interaction between C and T enables one to rule out $C^+ \cdot T$ paired helix I and provides support for the anti-parallel/parallel $C^+ \cdot C$ paired helixes II and III in which Ts bulge out.

4. DISTINCTION BETWEEN $C^+ \cdot C$ PAIRED ANTI-PARALLEL AND PARALLEL HELIXES

The primary difference between helixes II and III lies in the chain direction: II is anti-parallel, III is parallel. Stereochemical considerations reveal that in the formation of $C^+ \cdot C$ paired duplexes, for the parallel arrangement (III), both the Cs could have glycosyl torsion in the *anti* domain; for the anti-parallel arrangement (II) one C in the pair is *anti* while the other is *syn*. The primary NOE pattern in the 2D NOESY spectrum (fig.2) and the relevant 1D projections (fig.3) involving H6-H1′

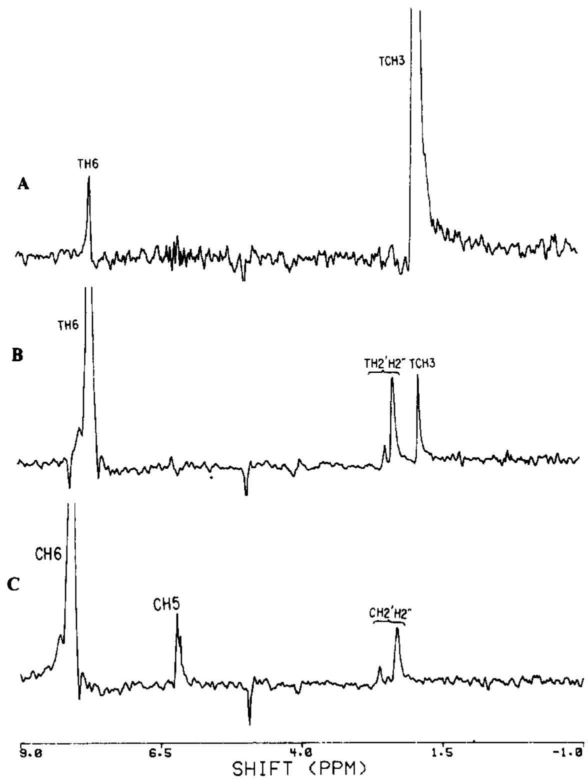


Fig.3. 1D projections from the 2D-NOESY experiment in fig.2. (A) TCH3 shows a strong NOE at TH6, and none at CH6/CH5. This suggests that C and T - next-neighbours in sequence - are far away from each other (>4.5 Å) in space. (B) TH6 shows strong NOE to TH2′/H2″ (and obviously to TCH3). This along with no transfer to TH1′ and TH3′ clearly indicates that the Ts are in the (C-2′ *endo, anti*) conformation. (C) CH6 shows strong NOE to CH2′/H2″ (and obviously to CH5). This along with no transfer to CH1′ and CH3′ clearly indicates that Cs are in the (C-2-*endo, anti*) conformation. Also, the absence of NOE at TCH3 from CH6 indicates that Ts bulge out as was independently demonstrated in panel A. Assignment: By comparing the 2D-COSY spectra of d(CTCTCT) in the duplex (pH 3, 5°C) and that in the single strand (pH 7, 20°C) and by comparing NOESY for the duplex at 300 and 500 ms mixing times, it was possible to assign the H1′, H2′, H2″, H3′, resonances within the observed envelopes as well as the CH6/CH5 pair and TCH3/TH6 pair in the spectrum. A discussion of them is beyond the scope of this paper. The COSY and the NOESY (at 500 ms mixing time) data, along with chemical shift values of H1′, H2′, H2″, H3′ and base protons are available from the authors upon request.

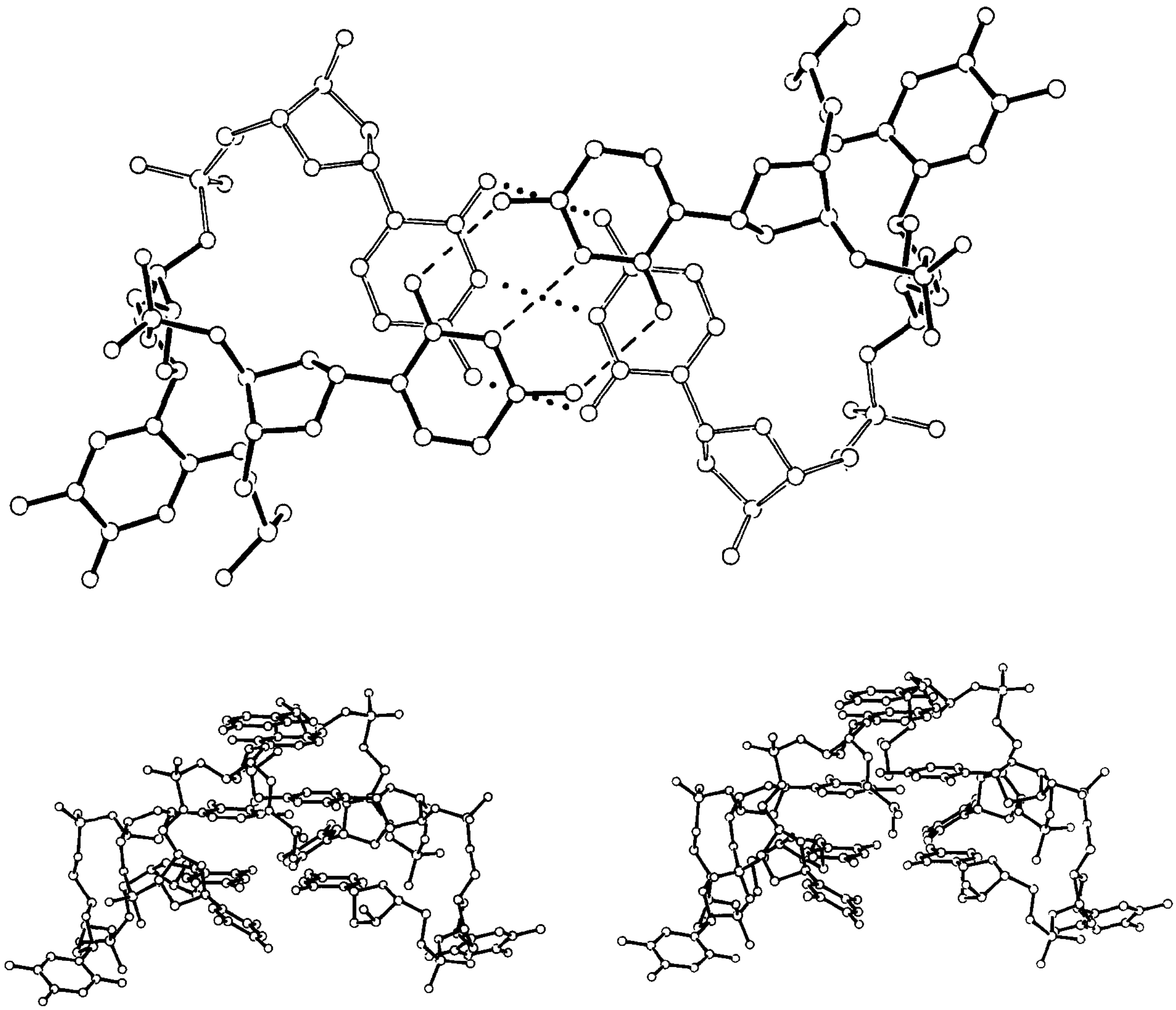
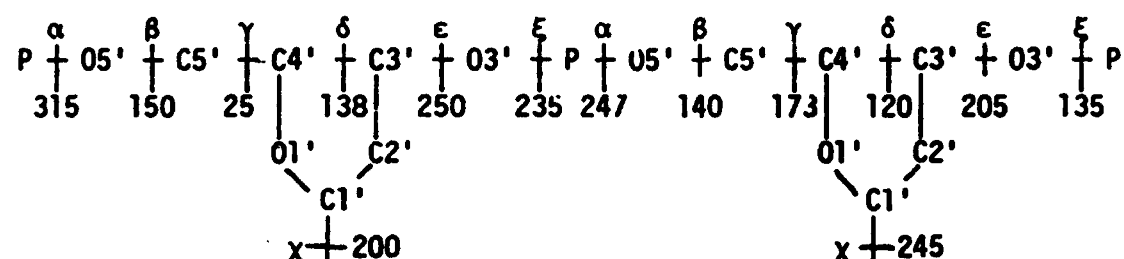


Fig.4. Structural description of the $C^+ \cdot C$ base paired parallel DNA helix in which Ts bulge out (i.e. helix III, see text). Top, view down the helix axis; bottom, stereo view perpendicular to the helix axis. The segment shown is Cl-T2-C3 which is conformationally identical to C3-T4-C5. The bonding (– – –) connects the top base pair; (···) connects the bottom base pair. Even though sugar pucker and glycosyl torsion of C and T fall in the (C-2′ *endo, anti*) domain, the magnitudes of the torsion angles employed for these residues are not identical. The employed values are diagramatically illustrated below. However, all employed torsion angles are well within stereochemically allowed regions. While the model proposed here is by no means the only possible structure for C^+-C paired parallel helix in which Ts bulge out, it is certainly one that is stereochemically satisfactory and agrees with the NMR data.



can be used to distinguish between *syn* and *anti* glycosidic torsions. In a *syn* conformation CH6/TH6–CH1′/TH1′ distances are about 2.2 Å and one expects very strong NOEs at CH1′(TH1′ from CH6/TH6. Figs. 2 and 3 clearly demonstrate that NOEs from CH6/TH6 to CH1′/TH1′ are extremely weak or not present at all, ruling out the *syn* conformation for both C and T. The evidence presented so far leads to the conclusion that d(CTCTCT) at low pH populates as a double helix in which $C^+ \cdot C$ are base paired, Ts bulge out, the glycosyl torsion of both C and T are *anti* and the chain direction is parallel.

5. STRUCTURAL DETAILS AND COMPUTER MODELLING

It can be seen from the primary spectrum at low pH (fig.1) that all the 3 CH6 belonging to 3Cs do overlap in frequency; so do all the TH6 belonging to 3 Ts. Hence, it is not possible to characterize the nucleotide geometries of the 3 Cs/Ts individually. However, one could determine whether all the 3 Cs/Ts belong to the (C3′-*endo, anti*) or (C2′-*endo, anti*) conformation: for the (C3′-*endo, anti*) conformation, cross-peaks from CH6/TH6 will appear at H3′, H2′/H2″ regions, while for the (C2′-*endo, anti*) conformation cross-peaks from CH6/TH6 will appear only in the H2′/H2″ region (and not in the H3′ region) [10,11]. From the projections in fig.3B,C it is clearly seen that there are no cross-peaks between CH6/TH6 and H3′, but strong cross-peaks are observed in the H2′/H2″ region. Therefore, we conclude that both C and T in the duplex of d(CTCTCT) at low pH adopt the (C2′-*endo, anti*) conformation.

Molecular models were computer generated using the structural information provided by the NMR data and by using stereochemical knowledge available for the backbone torsion angles. Two views are displayed in fig.4. There are several lines of independent evidence that support the structure we have proposed. In our structure there is a physical overlap between adjacent $C^+ \cdot C$ pairs which might provide the stereochemical basis for the formation of photo-adducts between two Cs in poly(dC-dT) at low pH [12]. The oligomer at low pH showed a CD spectrum (now shown) strikingly similar to that of poly(dC-dT) at low pH [12], suggesting that the structure of poly(dC-dT) at low pH may be very close to that proposed in fig.4. In fact, Brown et al. [12] have suggested a parallel double-helical model for poly(dC-dT) at low pH based upon CD, photo-chemical and model building studies. Even though the physico-chemical methods employed by Brown et al. [12] cannot provide any direct and reliable information about chain direction or about details of the nucleotide geometry, it is gratifying to note that the present 2D-NMR studies clearly show that their intuition about these from modelling studies was essentially correct. In the complex between r(CpA) and proflavin, single-crystal data [13] revealed the presence of a parallel structure in which $C^+ \cdot C$ and $A \cdot A$. were base paired. Even though this is a drug driven structure for a dinucleotide of RNA, at least it indicates, along with the present study, the plethora of structural motifs the double helix can assume. The present discovery of the existence of d(CTCTCT) at low pH as a parallel helix with $C^+ \cdot C$ pairs and $T \cdot T$ bulges opens up great possibilities for future research in the continuing story of DNA such as how proteins and other ligands recognize this unusual double helix. For example, preliminary studies indicate that proflavin can intercalate into this helix and we have also been able to observe NOEs between proflavin and the DNA protons.

ACKNOWLEDGEMENTS

The authors thank M.D. Frank-Kamenetskii and V.I. Ivanov for extensive discussion of the problem addressed in this paper, and for suggesting to undertake a detailed study of the solution structure of d(CTCTCT) at low pH. This research is supported by National Institutes of Health (GM29787) and National Foundation for Cancer Research. The high-field NMR experiments were performed at the NMR Facility for Biomolecular Research located at the F. Bitter National Magnet Laboratory, M.I.T. The NMR Facility is supported by grant no. RR00995 from the Division of Research Resources of the NIH and by the National Science Foundation under contract no. C-670.

REFERENCES

[1] Alexander Rich, Plenary Lecture (1985) 43rd

Pittsburgh Diffraction Conference, Nov. 6–8.
[2] Kennard, O. (1985) J. Biomol. Struct. Dyn. 3, 205–226.
[3] Patel, D.J., Kozlowski, S.A., Ikuta, S. and Itakura, K. (1984) Fed. Proc. 43, 2663–2672.
[4] Kan, L.S., Chandrasegaran, S., Pulford, S.M. and Miller, P.S. (1983) Proc. Natl. Acad. Sci. USA 80, 4263–4267.
[5] Lyamichev, V.I., Mirkin, S.M. and Frank-Kamenetskii, M.O. (1985) J. Biomol. Struct. Dyn. 2, 327–338.
[6] Patel, D.J. (1976) Biopolymers 15, 533–558.
[7] Patel, D.J. (1977) Biopolymers 16, 1635–1656.
[8] Gupta, G., Sarma, M.H. and Sarma, R.H. (1986) Int. J. Quant. Chem. QBS 12, 183–199.
[9] Sarma, M.H., Dhingra, M.M., Gupta, G. and Sarma, R.H. (1985) Biochem. Biophys. Res. Commun. 131, 269–276.
[10] Gupta, G., Sarma, M.H. and Sarma, R.H. (1985) J. Mol. Biol. 186, 463–469.
[11] Sarma, M.H., Gupta, G. and Sarma, R.H. (1986) Biochemistry, in press.
[12] Brown, D.M., Grey, D.M., Patrick, M.H. and Ratliff, R.L. (1985) Biochemistry 24, 1676–1683.
[13] Westhof, E., Rao, S.T. and Sunderlingam, M. (1980) J. Mol. Biol. 142, 331–361.
[14] States, D.J., Haberkorn, R.H. and Ruben, D.J. (1982) J. Mag. Reson. 48, 286–291.