¹H NMR study of the solution structure of the self-complementary dodecanucleotide d(TGCA)₃

W.C.Stevens, D.H.Huang¹, R.D.Wells and N.R.Krishna*

Departments of Biochemistry and ¹Medicine and the Comprehensive Cancer Center, Schools of Medicine and Dentistry, University of Alabama at Birmingham, Birmingham, AL 35294, USA

Received April 1, 1988; Accepted June 4, 1988

ABSTRACT:

The deoxyoligonucleotide d(TGCA)₃ is a candidate for exhibiting unusual conformations. Its ¹H NMR spectrum under low salt conditions has been obtained at 400 MHz and assigned using two-dimensional NMR techniques. The sugar puckers and glycosidic torsions have been determined by inspecting the relative intensities of the intranucleotide NOEs and COSY crosspeaks. At low electrolyte concentration (100 mM NaCl) the molecule exists as a right-handed duplex with twelve Watson-Crick base-pairs and deoxyribose moieties assuming the O1'-endo to C1'-exo pucker.

INTRODUCTION

DNA exhibits considerable diversity in conformation (1-3). Of the different conformers of DNA, the most profoundly different is left-handed Z-DNA, which has recently been shown to exist in vivo (4) and is likely to be involved in gene regulation. The Z conformation is generally found in sequences with a regular alternation of purines and pyrimidines, heretofore either mostly or completely cytosine and guanine. There are exceptions, however, and some base-pair (bp) sequences with interruptions in pu-py alternation stabilize in the left-handed form in negatively supercoiled circular DNA (5-7) or under suitable solvent conditions in linear duplexes (8).

It has recently been found (9) that when the self-complementary d(TGCA)₃ sequence is cloned in a supercoiled plasmid it can, under appropriate conditions, assume the left-handed conformation even though it has only 50% GC content. The sequence, being both a direct and an inverted repeat, adopts a cruciform structure under different conditions of superhelical stress (9).

We are, therefore, interested in the conformational repertoire of the TGCA repeat, independent of all effects induced by plasmid supercoiling. Anticipating broad linewidths and troublesome degeneracies in the NMR spectrum of the twenty bp oligonucleotide, we chose instead the dodecamer d(TGCA)₃ for our NMR analysis. This paper describes the nucleotide-specific ¹H NMR assignments and conformational features under low salt conditions.

MATERIALS AND METHODS

Approximately 175 A_{260} units of d(TGCATGCATGCA) were purchased from the Institute of Molecular Biophysics at Florida State University. The oligomer had been prepared on a commercial synthesizer and purified by anion-exchange chromatography. It was dissolved in 0.35 mL of 5.0 mM sodium phosphate buffer with 0.10 M NaCl and annealed. This solution was repeatedly lyophilized against D_2O and finally redissolved in 99.995% D_2O . Measurements in H_2O contained 10% D_2O for lock.

NMR spectra were obtained on a Bruker WH-400 (1H at 400.13 MHz) equipped with an Aspect 3000 computer. Temperature was maintained at 298 K and chemical shifts are referenced to the residual HDO signal, which was assigned 4.80 ppm. Except for the experiments in H_2O , the carrier was placed on the residual solvent signal. NMR data were processed on a MicroVAX II computer with the FTNMR software provided by Dr. Dennis Hare.

The phase-sensitive NOESY experiment was performed using the method of States, Haberkorn and Ruben (10). A mixing time of 300 ms was sufficient to obtain the longer range NOEs necessary for completing the assignment process. 128 transients were collected for each 1K spectrum and 219 points were collected in the t_1 dimension, zero-filled to provide a 1K x 1K matrix. Before Fourier transformation, the data set was multiplied by an unshifted sine bell window in the t_2 dimension and by a shifted (15°) sine bell in the t_1 dimension. The data were symmetrized to suppress t_1 noise and other artifacts.

The COSY data were obtained in the usual way with 256 1K FIDs of 64 transients each. The data were zero-filled for a 1K x 1K matrix which was multiplied in both dimensions by unshifted sine bell windows, Fourier transformed and symmetrized.

The NOE difference spectra were obtained using the 1- τ -1 hard pulse sequence (11,12) with the carrier located approximately midway between the imino and aromatic regions of the spectrum.

SEQUENTIAL RESONANCE ASSIGNMENTS

The application of 2D-NMR spectroscopy for resonance assignment and for conformational characterization of DNA oligomers is well documented in the literature (13-20). The combined use of NMR spectroscopy with restrained molecular dynamics simulations (21) or with distance geometry calculations (22,23) appears to be a promising approach in defining the detailed conformational properties of DNA oligomers in the solution phase. In making resonance assignments, it is often necessary to perform several different kinds of two-

dimensional NMR experiments, but sometimes it is possible to do the entire analysis using just one NOESY experiment of suitably long mixing time. That is the case here, though other experiments were checked for confirmation.

The method for making sequential, nucleotide-specific proton resonance assignments has been described in detail elsewhere (16-20) and so is not belabored here. It bears emphasis, though, that in cases where there is resonance overlap in the base and H1' region, the intra- and internucleotide NOEs used can sometimes provide a temptingly consistent assignment set that is wrong. In the case of this molecule, weak NOEs from cytosine H5s to neighboring sugars, rarely mentioned in the literature, allowed us to select one set of assignments over an alternative. So, especially in cases where degeneracy is a problem, every speck of "noise" in the NOESY map merits examination.

Figure 1 shows the one-dimensional NMR spectrum of the dodecamer. The exchangeable proton spectrum has been slightly resolution-enhanced and the residual solvent signal was suppressed by presaturation. The exchangeable proton signals (inset) were obtained as described above.

Figure 1 - The one-dimensional ^1H NMR spectrum of d(TGCA)_{12} . The residual solvent signal was removed with presaturation and the data were slightly resolution-enhanced. The imino region (inset) was baseline corrected.

Figure 2 - Region of the NOESY spectrum showing the connectivities between the base H6/H8 and the sugar H1'.

Figure 2 shows the NOE crosspeaks from the base H6 or H8 protons to the ribose H1's. Also included are the three cytosine H5-H6 signals and, lying beneath them, three NOEs from the cytosine H5s to the H8s of the preceding guanines.

If degeneracy were not a problem, each row and each column would contain two crosspeaks: an intra- and an internucleotide NOE. We should have twelve columns and twelve rows, but here there are only nine of each. Along the top-most row at 7.09 ppm, we see ambiguity arising from thymine H6 degeneracy and, along the rightmost column at 5.63 ppm, cytidine H1' degeneracy introduces ambiguity. One row toward the top contains the NOEs from three different pyrimidines.

A starting point is required and comes from elsewhere on the NOESY map (figure 3): Of the three adenine H8s, which are recognizable by their chemical shifts, only two display NOEs to thymine methyls. Since these must come from

Figure 3 - Region of the NOESY spectrum showing the intra- and internucleotide NOEs between base H6/H8 and sugar H2', H2". Some of the connectivities are shown for illustration. At right are the NOEs from thymine methyls.

the succeeding thymines on the same strand, it follows that the adenine H8 without the NOE must be A₁₂.

Proceeding upward from the internucleotide NOE to the H1' of C₁₁, one must pass through the intranucleotide NOE of G₆ to get to the first of the degenerate or unresolved crosspeaks. The intranucleotide NOE of C₁₁ is coincident with the internucleotide NOE from C₇(H6) to G₆(H1'). Also overlapping are the inter- and intranucleotide NOEs of G₁₀, and the intranucleotide NOEs of T₈ and T₉.

With these degeneracies in mind, one should be able to follow the assignment pathway shown in Figure 2, albeit with some difficulty. Notable are the weak NOE from C₁₁(H6) to G₁₀(H1') and the absent NOE from G₂(H8) to T₁(H1'). It should be remarked here that there are three or more different ways to make assignments that are consistent within this region of the NOESY. One of the incorrect ways even survives the crosscheck in the base-to-2',2" region. The assignments reported here, though, are the only ones that provide a perfectly consistent interpretation of all the NOESY data.

The sequential assignment process can also be performed in the same fashion in the region where the base protons show NOEs to the 2" and 2' protons. Part

of the sequential assignment pathway is shown in figure 3 for purposes of illustration.

The 2'' and 2' assignments reported in Table 1 were confirmed by their NOEs to H1'. The 3' protons were usually determined from their NOEs to H2'', H2' and confirmed by their NOEs to H1' (the region itself being checked against a relay-COSY). The H4' assignments came from their strong NOEs to H1' and were confirmed by H3' to H4' NOEs and COSY crosspeaks. The 5' and 5'' protons could not be assigned unambiguously. The adenine H2 signals were tentatively assigned from weak interbase NOEs and confirmed by their NOEs with the imino protons.

Weak NOE crosspeaks from the cytosine H5s to sugar protons are shown in figure 4. The three signals to the left are internucleotide NOEs to the 2'' protons of the preceding guanosines and the three to the right are intranucleotide NOEs to 2' protons. As mentioned earlier, the presence of these weak NOEs enabled us to select one set of assignments (see Table 1) from two alternative but otherwise equally consistent sets.

Table 1: ¹H chemical shifts for the dodecanucleotide d(TGCA)₃ in 0.10 M NaCl, 5.0 mM phosphate, pH=7, T=298 K.^a

	Nuc	H1'	H2''	H2'	H3'	H4'	H6/8	H5/H2/Me
1	T	5.93	2.26	1.88	4.65	4.06	7.34	1.54
2	G	6.01	2.84	2.74	4.98	4.40	8.04	----
3	C	5.62	2.45	2.13	4.86	4.23	7.41	5.43
4	A	6.27	2.95	2.69	5.04	4.44	8.32	7.67
5	T	5.72	2.39	1.99	4.86	4.12	7.08	1.44
6	G	5.82	2.68	2.57	4.95	4.33	7.79	----
7	C	5.62	2.41	2.07	4.83	4.18	7.33	5.31
8	A	6.22	2.92	2.65	5.00	4.40	8.27	7.57
9	T	5.74	2.33	1.93	4.86	4.12	7.08	1.42
10	G	5.75	2.58	2.58	4.96	4.35	7.82	----
11	C	5.82	2.31	1.96	4.76	4.10	7.33	5.41
12	A	6.31	2.47	2.62	4.67	4.17	8.21	7.68

^a In ppm, referenced to the residual HDO signal which was assigned 4.80 ppm.

GENERAL CONFORMATIONAL FEATURES

The handedness of the dodecamer is readily inferred from the strong inter-base NOEs from the guanosine H8s to the cytidine H5s and from the adenosine H8s to the thymidine methyls (21). The bases are oriented *anti* to the sugars as we observe that, in NOESYs with shorter mixing times, the H6/H8 to H1' NOEs fall off rapidly in intensity while the H6/H8 to H2' NOEs remain strong. For left-handed Z-DNA, in which the purines assume a *syn* orientation, strong NOE contacts between the purine H8 and the sugar H1' are expected (24). Finally, Figure 3 shows that the intranucleotide NOEs from H6/H8 are strong and that the stronger of the internucleotide NOEs is to H2'' of the 5'-flanking sugar, which is normal for B-DNA. In A-DNA, which can be ruled out here, a base H6/H8 proton is closest to the H2' of the preceding nucleotide and quite some distance from its own (25,26).

Figure 4 - The regions as marked of the NOESY spectrum.

Figure 5 - The regions as marked of the COSY spectrum.

The puckers of the sugar moieties are probably best determined from a simple visual inspection of the COSY spectrum (27). Figure 5 shows the $1'-2',2''$ and $3'-2',2''$ couplings. The $2''$ protons exhibit stronger NOEs to the $1'$ protons and are distinguished on that basis. They lie at lower field except in the cases of G_{10} , where $2'$ and $2''$ are degenerate, and A_{12} , where $2''$ is upfield of $2'$. The assignments of the A_{12} protons are confirmed by a stronger NOE between $2'$ and $3'$ and by the appearance of the COSY crosspeaks.

We observe that the relative magnitudes of the COSY crosspeaks are nearly the same for $1'-2'$ and $1'-2''$ coupling and that these intensities are comparable to the $3'-4'$ couplings. Additionally, crosspeaks for $3'-2''$ coupling are either very weak or absent, except in the case of A_{12} . For all but the terminal base pair, the sugar puckers must lie between $O1'$ -endo and $C1'$ -exo (27).

On the basis of slightly weaker 3'-4' coupling, one can place the purine nucleotides, except for A₁₂, closer to C1'-exo and the pyrimidine nucleotides, except for T₁, closer to O1'-endo conformation. With its significant 2"-3' and large 3'-4' couplings, A₁₂ seems to have the lowest pseudorotation angle of all, probably almost exactly O1'-endo.

Nilges *et al.* (21) have used NMR and restrained molecular dynamics to study the structure of the self-complementary hexamer d(GCATGC), which our dodecamer contains. On the basis of the relative magnitudes of the intranucleotide sugar-base NOEs, they report the sugar puckers for the hexamer to lie in a somewhat broad range from O1'-endo to C2'-endo. The C2'-endo conformation (for which $J_{3,4} \approx 0$) is outside the range of the sugar puckers in our dodecamer, whose nucleotides each exhibit a substantial 3'-4' COSY crosspeak. It is possible that the narrower range of puckers we observe is a consequence of our oligomer's larger size.

Finally, owing to the self-complementary nature of this oligonucleotide, there are six unique basepairs that should give rise to five imino signals -one does not generally expect to see the imino proton resonance from the terminal AT base pair in aqueous solution even near its freezing point since it will be very broad due to fraying of the duplex. We observe the melting behavior expected for a duplex in spectra recorded in ¹H₂O (28).

At 298 K there are four signals present, the sharp peak at 12.48 ppm belonging to the most interior GC pair, the two most downfield peaks belonging to the interior AT pairs, and the broad signal at 12.61 ppm coming from the GC pair third from the end. As temperature is raised, the signals disappear in the expected order, from the ends inward, the AT pairs at 13.53 and 13.41 ppm melting together. Upon lowering temperature to below 280 K, the outermost GC pair makes its appearance slightly downfield of the other GC pairs. These assignments based on temperature dependence are in agreement with assignments based on NOEs between the imino and nonexchangeable base protons. NOE difference spectra consist primarily of intra- and inter-base-pair NOEs proceeding from adenosine H2 and indicate a normal right-handed duplex with base pairing of Watson-Crick type (29). The absence of any imino signals in the 10.5-11.5 ppm range indicates that, at the salt and strand concentrations used in this study, the abundance of hairpin conformations is not significant (30).

The assignments reported in this paper provide a basis for further spectroscopic characterization of the conformational variability of d(TGCA), under different solvent conditions and ionic strengths. We have here established that the oligomer exists as a conventional B-helix when the ambient salt concentra-

tion is low and we are undertaking studies of the B-Z transition in order to contribute, ultimately, to understanding the role of base-pair sequence in the stabilization of alternative conformations in DNA.

Recent spectrophotometric studies (31) on the binding of actinomycin D with dodecamers containing the d(TGCA) sequence show that the drug exhibits stronger affinity for the d(TGCA) sequence than for other sequences studied. Further, the drug binds to the d(TGCA)₃ sequence with a 3:1 stoichiometry (31). Our investigation also serves as a basis for an NMR spectroscopic characterization of the complexes of the dodecamer with the antitumor antibiotic, actinomycin D.

ACKNOWLEDGEMENTS

Support of this work by the U.S. Public Health Service Grants CA-13148 and GM-30822 and by the grants DMB-8502666 and DMB-8607785 from the National Science Foundation is gratefully acknowledged. The oligonucleotide synthesizer at FSU was purchased with funds from the NSF grant DMB-8516946. We thank Dr. Dennis Hare for providing his FTNMR software and we wish to acknowledge the efforts of Mr. Hank Henricks in preparing the oligonucleotide.

* To whom correspondence should be addressed.

REFERENCES

1. Wells, R.D. and Harvey, S.C., Eds. (1988) Unusual DNA Structures Springer-Verlag, New York, NY
2. Wells, R.D. (1988) J. Biol. Chem. 263, 1095-1098.
3. Rich, A., Nordheim, A., and Wang, A.H.-J. (1984) Ann. Rev. Biochem. 53, 791-846.
4. Jaworski, A., Hsieh, W.-T., Blaho, J.A., Larson, J.E., and Wells, R.D. (1987) Science 238, 773-777.
5. McLean, M.J., Blaho, J.A., Kilpatrick, M.W., and Well, R.D. (1986) Proc. Natl. Acad. Sci. USA 83, 5884-5888.
6. Kilpatrick, M.W., Klysik, J., Singleton, C.K., Zarling, D.A., Jovin, T.M., Hanau, L.H., Erlanger, B.F. and Wells, R.D. (1984) J. Biol. Chem. 259, 7268-7274.
7. Nordheim, A., Lafer, E.M., Peck, L.J., Wang, J.C., Stollar, B.D., and Rich, A. (1982) Cell 31, 309-318.
8. Feigon, J., Wang, A.H.-J., van der Marel, G.A., van Boom, J.H., and Rich, A. (1985) Science 230, 82-84.
9. McLean, M.J., and Wells, R.D. (1988) J. Biol. Chem., in press
10. States, D.J., Haberkorn, R.A., and Ruben, D.J. (1982) J. Magn. Reson. 48, 286-292.
11. Clore, G.M., Kimber, B.J., Gronenborn, A.M. (1983) J. Magn. Reson. 54, 170-173.
12. Kime, M.J., and Moore, P.B. (1983) Biochemistry 22, 2615-2622.
13. Feigon, J., Wright, J.M., Leupin, W., Denny, W.A., and Kearns, D.R. (1982) J. Am. Chem. Soc. 104, 5540-5541
14. Pardi, A., Walker, R., Rapoport, H., Wider, G., and Wuthrich, K. (1983) J. Am. Chem. Soc. 105, 1652-1653.

15. Wuthrich, K. (1986) NMR of Proteins and Nucleic Acids, John Wiley & Sons, New York, NY
16. Hare, D.R., Wemmer, D.E., Chou, S.-H., Drobny, G., and Reid, B.R. (1983) J. Mol. Biol. **171**, 319-336.
17. Scheek, R.M., Boelens, R., Russo, N., van Boom, J.H., and Kaptein, R. (1984) Biochemistry **23**, 1371-1376.
18. Feigon, J., Leupin, W., Denny, W.A., and Kearns, D.R. (1983) Biochemistry **22**, 5943-5951.
19. Weiss, M.A., Patel, D.J., Sauer, R.T., and Karplus, M. (1984) Proc. Natl. Acad. Sci. USA **81**, 130-134.
20. Frechet, D., Cheng, D.M., Kan, L.-s., and Ts'o, P.O.P. (1983) Biochemistry **22**, 5194-5200.
21. Nilges, M., Clore, G.M., Gronenborn, A.M., Brunger, A.T., Karplus, M., and Nilsson, L. (1987) Biochemistry **26**, 3718-3733.
22. Hare, B.R. and Reid, B.R. (1986) Biochemistry **25**, 5341-5350.
23. Patel, D.J., Shapiro, L., and Hare, D.R. (1988) in Unusual DNA Structures, Wells, R.D. and Harvey, S.C., Eds., Springer-Verlag, New York, N.Y., pp. 115-161.
24. Patel, D.J., Kozlowski, S.A., Nordheim, A., and Rich, A. (1982) Proc. Natl. Acad. Sci. USA **79**, 1413-1417.
25. Reid, D.G., Salisbury, S.A., Bellard, S., Shakked, Z., and Williams, D.H. (1983) Biochemistry **22**, 2019-2023.
26. Patel, D.J., Shapiro, L., and Hare, D. (1986) Biopolymers **25**, 693-706.
27. Chary, K.V.R., Hosur, R.V., Govil, G., Zu-kun, T., and Miles, H.T. (1987) Biochemistry **26**, 1315-1322.
28. Boelens, R., Scheek, R.M., Dijkstra, K., and Kaptein, R. (1985) J. Magn. Res. **62**, 378-386.
29. Assa-Munt, N., Granot, J., Behling, R.W., and Kearns, D.R. (1984) Biochemistry **23**, 944-955.
30. Orbons, L.P.M., van der Marel, G.A., van Boom, J.H., and Altona, C. (1987) J. Biomol. Struct. Dyns. **4**, 939-963 and refs therein.
31. Chen, F.-M. (1988) Biochemistry **27**, 1843-1848.