Vol. 119, No. 2, 1984 BIOCHEMICAL AND BIOPHYSICAL RESEARCH COMMUNICATIONS
March 15, 1984 Pages 663-670

COMPLETE ASSIGNMENT OF THE NON-EXCHANGEABLE PROTON NMR RESONANCES OF $[d-(GGAATTCC)]_2$ USING TWO-DIMENSIONAL NUCLEAR OVERHAUSER EFFECT SPECTRA

Michelle S. Broido*, Gerald Zon‡, and Thomas L. James†

Department of Pharmaceutical Chemistry, School of Pharmacy,
University of California, San Francisco, CA 94143

‡National Center for Drugs and Biologics,
Food and Drug Administration, Bethesda, MD 20205

Received February 2, 1984

Summary: Proton two-dimensional nuclear Overhauser effect (2D NOE) spectra have been obtained for $[d-(GGAATTCC)]_2$ at a series of mixing times. By performing the experiment at 500 MHz and obtaining pure absorption phase spectra, sufficient resolution was achieved to be able to assign all seventy of the non-exchangeable proton resonances in the self-complementary octamer duplex, including the H_3' resonances which overlap with the solvent peak.

Introduction: ^1H nuclear magnetic resonance (NMR) has been shown to be a very useful tool for the study of conformation and dynamics of nucleic acid oligomers. Unfortunately, the small chemical shift dispersion found in many regions of the spectrum, most notably for the different sugar proton resonances, makes spectral assignment very difficult. Thus, most of the ^1H NMR studies of nucleic acids to date have focused on the base protons and the anomeric protons. In order to extract the very useful information inherent in the resonance behavior of the remainder of the sugar protons, it is imperative that methods be developed by which the entire ^1H NMR spectrum can be assigned.

The problem of small chemical shift dispersion becomes readily apparent when one considers that for a given nucleic acid there are essentially only

*Present address: Department of Chemistry, 695 Park Avenue, Hunter College, New York, N.Y. 10021.

†Recipient of Research Career Development Award AM00291 from the National Institutes of Health. Address correspondence to this author at the University of California, San Francisco

0006-291X/84 $1.50

four different mononucleotides of which the polymer is composed and that the sugar ring found in any given mononucleotide is chemically equivalent to that found in any other mononucleotide. The primary reasons for magnetic inequivalence of the base protons in identical mononucleotides are the ring current shifts induced by neighboring bases; magnetic inequivalence of the sugar rings in the different nucleotides arises from the different shielding effects of the heterocyclic bases and from changes in helical parameters due to base sequence effects on conformation. Within a given molecule, these variations are sufficiently small that there are only small changes in chemical shift for a given type of sugar proton. This small chemical shift dispersion prevents the use of standard decoupling and nuclear Overhauser effect measurements of sugar protons, since the required selectivity of irradiation for such experiments is not achievable.

Using two-dimensional NMR (2D NMR) techniques, it is possible to overcome many of the difficulties mentioned above. Indeed 2D NMR studies on some oligonucleotides have been carried out permitting partial assignment of the proton resonances.[1-4] Even better resolution is afforded by the pure-absorption two-dimensional nuclear Overhauser effect (2D NOE) experiment.[5] We have used this experiment to assign all of the non-exchangeable ^1H resonances in the self-complementary DNA duplex $[d-(5'GGAATTCC3')]_2$. Repeating the experiment at several mixing times provided both a cross-check on assignments and yielded valuable information on the magnitudes of internuclear separations. (This latter aspect will be discussed elsewhere.) The cross-peak intensity in a 2D NOE experiment is a function of the inverse sixth power of the internuclear separations, the molecular motions, and the experimental parameter τ_m, the mixing time. The τ_m dependence can be useful for assignment purposes: if, for example, one proton is dipolar-coupled to two others but more strongly to one, and if the resonances of these latter two protons are at very similar frequencies, then the cross-peak intensity of the more strongly coupled proton will be more pronounced at shorter τ_m. Thus, comparison of 2D NMR spectra at two (or

more) different mixing times in a spectral region in which there are many overlapping resonances enables assignments to be made or checked, provided one has some knowledge of structural constraints (vide infra).

Materials and Methods: DNA Synthesis. The self-complementary octamer [d-(5'GGAATTCC3')]$_2$ was synthesized as follows. Deoxynucleotide N,N-diisopropylaminomethoxyphosphines,[6] having 5'-dimethoxytrityl (DMT) and either N-benzoyl (for dA and dC) or N-isobutyryl (for dG) blocking groups, were purchased from Applied Biosystems (Foster City, CA) as powders that were ≥ 95% pure by ^{31}P NMR. Commercially available (Applied Biosystems) silica, derivatized[7] with 5'-DMT-N-benzoyl-dC, was used as the solid support (1 μmol dC, 25 mg), which was carried out with an Applied Biosystems Model 380A DNA synthesizer. The synthesis cycle[7] consisted of detritylation (100s) with 3% Cl_3CO_2H in CH_2Cl_2, tetrazole-catalyzed coupling (180s) of the phosphoramidite (20-fold molar excess), capping (120s) of 5'-HO groups with acetic anhydride in tetrahydrofuran that contained 2,6-lutidine and 4-dimethylaminopyridine as catalyst, and oxidation (30s) of the resultant phosphite linkage with I_2-H_2O in tetrahydrofuran that contained 2,6-lutidine. The total time for these steps and intermediary washes with either CH_3CN or CH_3NO_2 was 18 min, and coupling yields were generally ≥ 94% based on a colorimetric assay for DMT cation at 498 nm. O-Demethylation (30 min) with PhSH-Et_3N in p-dioxane (1:2:2 v/v) was carried out automatically after 7 cycles of coupling, and oligodeoxynucleotides were then removed from the silica support with concentrated NH_4OH solution (2 ml), which was delivered to the column in 8 portions over a 1 h period. The resultant ammoniacal solution was heated at 55°C overnight, and the crude product that was obtained by evaporation to dryness was purified by HPLC using a μBondapak C_{18} column (7.8 x 30 cm, Waters Assoc., Milford, MA) with 0.1 M triethylammonium acetate, pH 7.8 and an initial linear gradient of CH_3CN (20-30% over 10 min) at a flow rate of 4 ml/min. The major 5'-DMT-bearing component that eluted at 12 min was concentrated in vacuo, detritylated with 3% aqueous HOAc for 10 min, and then extracted with EtOAc to afford the final product, which was > 95% pure by HPLC analysis as described above, except for the use of a linear gradient of 5-20% CH_3CN over 15 min (elution time 11 min). The size and homogeneity of this octamer was confirmed by enzymatically labeling the 5' end with ^{32}P and comparison with oligo-dT standards using a 20% polyacrylamide sequencing gel[8] as well as by ^{31}P NMR. The final product obtained from 12 syntheses was pooled and lyophilized twice from D_2O to give 11.3 mg of material.

The resulting lyophilized sample was redissolved in ethanol, passed through several "Sep-Pak" cartridges (Waters Associates, part no. 51910), dried under a nitrogen stream, dissolved in 0.4 ml of pH 7 deuterium oxide solution 0.180 M NaCl, 0.100 M Na_3PO_4, 0.2 mM EGTA, to a final concentration (single-strand) of ~ 4.5 mM. This sample was lyophilized several times and finally dissolved in 100% D_2O (Aldrich). The NMR sample tube used was a Wilmad 528PP tube.

NMR Spectra. All ^{1}H NMR spectra were run on the Nicolet NM500, equipped with Nicolet 1280 computer, at the University of California, Davis Nuclear Magnetic Resonance Facility. Pure-absorption 2D NOE spectra were run with the carrier frequency at the residual HDO peak; this peak was suppressed using a phase-cycled soft pulse during the post-acquisition delays. The 90° pulses were typically 12 μsec and the post-acquisition delays were 12 sec. Spectra were recorded with a sweepwidth of ±3000 Hz, 4K points, and typically 400 spectra in the t_1 dimension, 32 acquisitions signal-averaged per spectrum, giving a processed data matrix 1K x 1K (zero-filled in the t_1 dimension). The temperature was 20°C.

Results and Discussion: The 500 MHz ^1H NMR spectrum of the duplex $[d-(GGAATTCC)]_2$[9] is shown in Figure 1. Indicated in the figure are the ranges of chemical shift for each type of proton. Although the base protons (region A) are reasonably well-resolved, the sugar protons, particularly in region D which contains 24 protons, are strongly overlapping. Several of the 3' protons (region C) are buried under the residual solvent peak and clearly cannot be assigned by conventional means.

The base protons of the duplex in pH 5.05 solution were assigned by Patel and Canuel.[10] Our assignments at pH 7 are consistent with these assignments and were confirmed by T_1 measurements, J-couplings of the cystosine H_5 and H_6, and internal consistency with other assignments through dipolar connectivity.

Figure 2 shows the 500 MHz 2D NOE (pure absorption) spectrum of the duplex, taken with a 250 msec mixing time. For the purposes of assignment, it was initially assumed that the octamer duplex used in this study was of

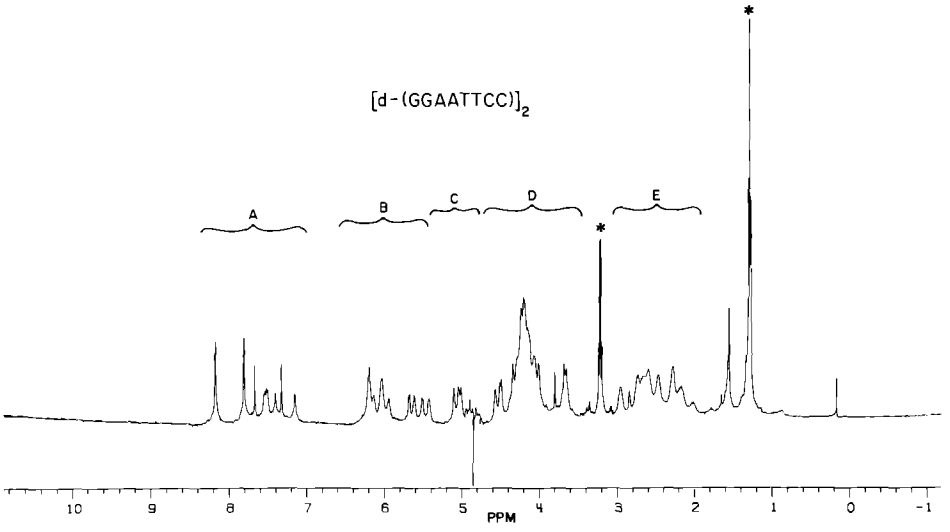

Figure 1. 500 MHz spectrum of $[d-(5'GGAATTCC3')]_2$; chemical shifts are relative to HOD at 20°C being 4.84 ppm downfield from TSP. Indicated spectral regions are:
A. base protons: A-H_8, H_2, G-H_8, C-H_6, T-H_6; 8.2-7.2 ppm
B. anomeric protons: ($H_{1'}$) and C-H_5; 6.2-5.4 ppm
C. $H_{3'}$: 5.1-4.56 ppm.
D. $H_{4'}$, $H_{5'}$, $H_{5''}$: 4.50-3.6 ppm
E. $H_{2'}$, $H_{2''}$: 3.0-2.0 ppm
F. T-CH_3: 1.6-1.3 ppm
The asterisks indicate residual Et_3NH^+ resonances.

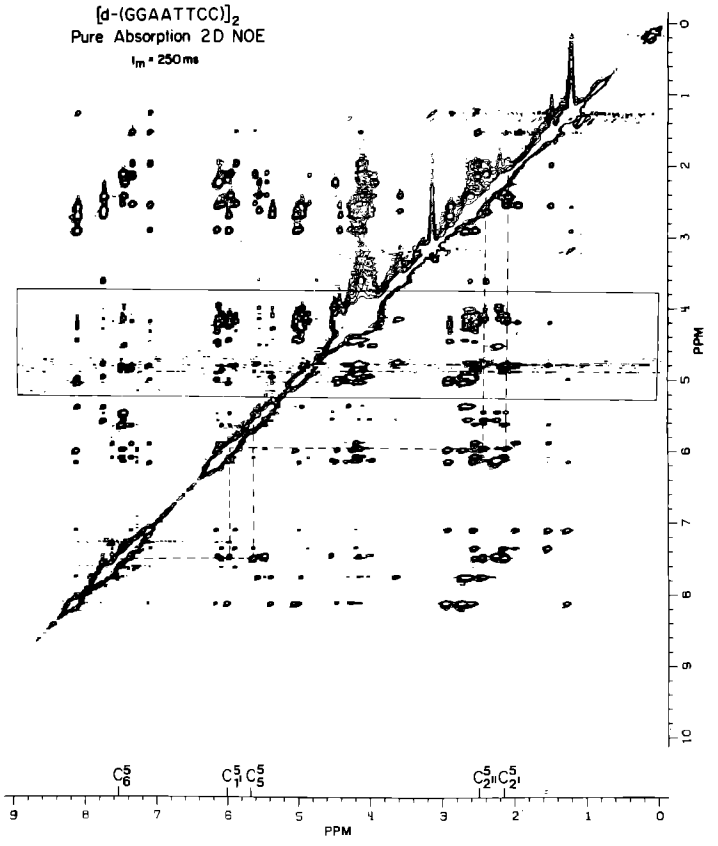

Figure 2. 500 MHz pure-absorption 2D NOE spectrum of $[d-(5'GGAATTCC3')]_2$; τ_m=250 msec; 20°C. The vertical lines indicate the base proton resonance positions. The dashed lines indicate the connectivities of the C^5-H_6, H_5, $H_{1'}$, $H_{2'}$ and $H_{2''}$ protons. The boxed region is shown expanded in Figure 3.

the B-DNA family of structures although not necessarily B-DNA itself. Any lack of validity to this assumption would readily become apparent in the dipolar connectivities since some internuclear distances vary considerably between different conformational families, e.g., A, B, or Z. In a B-DNA structure, a given base proton will undergo dipolar interaction with its own 1', 2', and 2'' protons, as well as with these same protons on the sugar ring of the 5' neighboring nucleotide. These interactions are greater for the base to its own $H_{2'}$ than to its own $H_{2''}$, and greater to its 5' neighboring $H_{2''}$ than to its neighboring $H_{2'}$.[4] As demonstrated by Scheek *et al.*,[2] one can trace the dipolar connectivities from base proton to $H_{1'}$ to

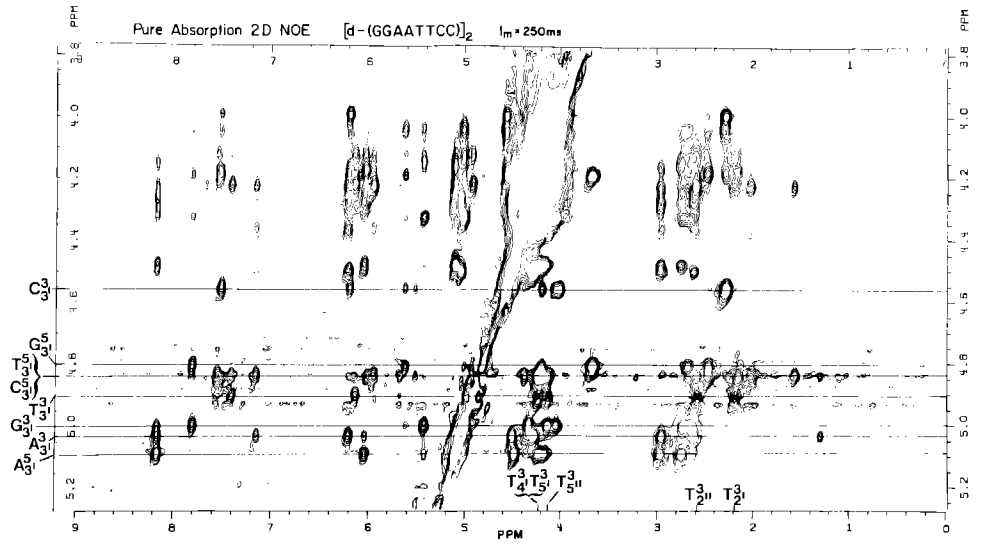

Figure 3. See caption to Figure 2. The horizontal lines indicate the resonance frequencies of the 3' protons. The Xs indicate the cross-peaks arising from the interaction of the T^3-H_3, with the T^3-H_4', H_5', H_5'', H_2' and H_2''.

H_2' and H_2'' this way. In Figure 2 the dashed lines indicate examples of such connectivities.

A spectrum such as Figure 2 has frequency resolution too coarse to enable identification of many of the H_3', H_4', H_5' and H_5'' resonances. Looking at an expanded region of the spectrum, as in Figure 3, distinct resonance frequencies are identifiable. In Figure 3, the vertical expansion in the frequency range around HDO (region C in Figure 1) separates the cross-peaks arising from H_3' interactions with other protons. Looking at region C of Figure 1, the individual resonance frequencies cannot be identified, in part due to suppression of the HDO peak. Overlap from axial and cross-peaks in Figures 2 and 3 arising from region D of the spectrum prohibits establishing axial to cross-peak connectivities as in Figure 2; however, the 4', 5' and 5'' resonances can be assigned after assignment of the 2', 2'' and 3' resonances. For example, in Figure 3 there are distinct cross-peaks at 4.22 ppm, 4.13 ppm, 2.58 ppm and 2.19 ppm, connected to the resonance at 4.91 ppm. The peaks at 2.19 and 2.58 ppm have been assigned to the 2' and 2'' protons, respectively, of T^3, so the resonance at 4.91 ppm is

from T^3-H_3'. Therefore, the resonance at 4.22 ppm is from T^3-H_4' and that at 4.13 ppm is from T_3-H_5' or H_5''. Examination of the same spectral region in the spectrum obtained with a 100 msec mixing time shows that the intensity of the cross-peak at 4.22 ppm is greater than that at 4.13 ppm. Furthermore, examination of a B-DNA model shows that, within a given sugar ring, H_3' is significantly closer to H_5'' than to H_5' and that H_5' is relatively close to the H_1' of the 5' neighboring sugar ring. In the 100 msec-mixing time spectrum, there is a distinct cross-peak between the resonance at 5.93 ppm (T^5-H_1',) with that at 4.22 ppm. Thus, T^3-H_4' and H_5' resonate at 4.22 ppm and T^3-H_5'' at 4.13 ppm.

Using similar reasoning, we have been able to assign all of the non-exchangeable resonances of the octamer, with some ambiguity remaining in some of the assignments of the H_2' versus the H_2'' and of the H_5' versus the H_5''. These assignments are found in Table 1.

Table 1. Chemical Shifts[a] of the Non-Exchangeable Protons in $[d\text{-}(5'GGAATTCC3')]_2$

	H_8	H_6	H_5	H_2[b]	CH_3	H_1'	H_2'[c]	H_2''[c]	H_3'	H_4'	H_5'[c]	H_5''[c]
G^5	7.80					5.62	(2.68	2.47)	4.80	4.19	3.66	3.66
G^3	7.80					5.42	(2.74	2.65)	5.00	4.33	4.13	4.05
A^5	8.16			(7.31 7.66)		6.04	(2.96	2.74)	5.09	4.48	(4.22	4.16)
A^3	8.16			(7.31 7.66)		6.20	(2.96	2.61)	5.03	4.50	4.33	4.28
T^5		7.20			1.29	5.93	2.02	2.58	4.48	4.37	4.22	4.14
T^3		7.40			1.58	6.12	2.19	2.58	4.91	4.22	4.22	4.13
C^5		7.54	5.68			6.02	2.16	2.47	4.84	4.37	(4.22	4.14)
C^3		7.50	5.51			6.18	2.28	2.28	4.56	4.19	4.05	4.01

[a] ± 0.02 ppm; chemical shifts relative to HDO resonating at 4.84 ppm downfield from TSP.

[b] It was not possible to determine which of the H_2 resonances is A^5-H_2 and which A^3-H_2.

[c] In some cases it was not possible to distinguish between H_2' and H_2'' resonances or between H_5' and H_5'' resonances. Ambiguities are apparent from the parentheses.

As other workers have shown, the 2D NOE experiment is very useful for assigning resonances of nucleic acid base protons, anomeric protons, and the 2' and 2'' protons. We have extended its use to the assignment of the remainder of the sugar protons, including resonances overlapping with residual solvent peak (the $H_{3'}$ resonances) and the very dense spectral region containing the $H_{4'}$, $H_{5'}$ and $H_{5''}$ resonances. This has resulted in the assignment of all 70 non-exchangeable protons of a self-complementary octamer, the largest oligonucleotide yet assigned by any means.

Acknowledgments: We wish to thank Dr. Vladimir J. Basus for setting up the pure-absorption 2D NOE sequence on the UC Davis NMR instrument and Dr. Gerald B. Matson for assistance in obtaining spectra. We also thank Dr. William Egan, Dr. Frederick T. Gates and Ms. Kathleen A. Gallo for aid in preparation of the octamer. Financial support was provided by National Institutes of Health grants GM25018 and CA27343.

References

1. Feigon, J., Wright, J.M., Leupin, W., Denny, W.A., and Kearns, D.R. (1982) J. Am. Chem. Soc. **104**, 5207-5216.
2. Scheek, R.M., Russo, N., Boelens, R., Kaptein, R., and van Boom, J.H. (1983) J. Am. Chem. Soc. **105**, 2914-2916.
3. Feigon, J., Leupin, W., Denny, W.A., and Kearns, D.R. (1983) Biochemistry **22**, 5943-5951.
4. Haasnoot, C.A.G., Westerink, H.P., van der Marel, G.A., and van Boom, J.H. (1983) J. Biomolec. Struc. Dynamics **1**, 131-149.
5. States, D.J., Haberkorn, R.A., and Ruben, D.J. (1982) J. Magn. Reson. **48**, 286-292.
6. McBride, L.J., and Caruthers, M.H. (1983) Tetrahedron Lett. **24**, 245-248.
7. Caruthers, M.H., Beaucage, S.L., Becker, C., Efcavitch, W., Fisher, E.F., Galluppi, G., Goldman, R., deHaseth, P., Martin, F., Mattuecci, M., and Stabinsky, Y (1982) In Setlow, J. and Hollaender, A., (Eds.), Genetic Engineering, Vol. **4**, Plenum Press, New York, pp. 1-17.
8. Maxam, A.M., and Gilbert, W. (1980) Methods Enzymol **65**, 499-560.
9. The mononucleotides in d-(5'GGAATTCC3') are identified as $G^5G^3A^5A^3T^5T^3C^5C^3$ in the text.
10. Patel, D.J., and Canuel, L.L. (1979) Eur. J. Biochem. **96**, 267-276.