Williams, D. H. (1983) *Biochemistry 22*, 2019-2025.
Rinkel, L. J., & Altona, C. (1987) *J. Biomol. Struct. Dyn. 4*, 621-649.
Scheek, R. M., Russo, N., Boelens, R., Kaptein, R., & Van Boom, J. H. (1983) *J. Am. Chem. Soc. 105*, 2914-2916.
Scheek, R. M., Boelens, R., Russo, N., Van Boom, J. H., & Kaptein, R. (1984) *Biochemistry 23*, 1371-1376.
Sheth, A., Ravikumar, M., Hosur, R. V., Govil, G., Tan, Z-k., & Miles, H. T. (1987a) *Biochem. Biophys. Res. Commun. 144*, 26-34.
Sheth, A., Ravikumar, M., Hosur, R. V., Govil, G., Tan, Z-k., Roy, K. B., & Miles, H. T. (1987b) *Biopolymers 26*, 1301-1313.
Tan, Z-k., Ikuta, S., Huang, T., Dagaiczyk, A., & Itakura, K. (1982) *Cold Spring Harbor Symp. Quant. Biol. 47*, 383-391.
Widmer, H., & Wüthrich, K. (1987) *J. Magn. Reson. 74*, 316-336.
Wüthrich, K. (1986) *NMR of Proteins and Nucleic Acids*, Wiley, New York.
Wynants, C., & Van Binst, G. (1984) *Biopolymers 23*, 1799-1804.

NMR Studies of DNA $(R+)_n\cdot(Y-)_n\cdot(Y+)_n$ Triple Helices in Solution: Imino and Amino Proton Markers of T·A·T and C·G·C^+ Base-Triple Formation†

Carlos de los Santos, Mark Rosen, and Dinshaw Patel*
Department of Biochemistry and Molecular Biophysics, College of Physicians and Surgeons, Columbia University, New York, New York 10032

Received April 14, 1989; Revised Manuscript Received May 23, 1989

ABSTRACT: High-resolution exchangeable proton two-dimensional NMR spectra have been recorded on 11-mer DNA triple helices containing one oligopurine $(R)_n$ and two oligopyrimidine $(Y)_n$ strands at acidic pH and elevated temperatures. Our two-dimensional nuclear Overhauser effect studies have focused on an 11-mer triplex where the third oligopyrimidine strand is parallel to the oligopurine strand. The observed distance connectivities establish that the third oligopyrimidine strand resides in the major groove with the triplex stabilized through formation of T·A·T and C·G·C^+ base triples. The T·A·T base triple can be monitored by imino protons of the thymidines involved in Watson–Crick (13.65–14.25 ppm) and Hoogsteen (12.9–13.55 ppm) pairing, as well as the amino protons of adenosine (7.4–7.7 ppm). The amino protons of the protonated (8.5–10.0 ppm) and unprotonated (6.5–8.3 ppm) cytidines in the C·G·C^+ base triple provide distinct markers as do the imino protons of the guanosine (12.6–13.3 ppm) and the protonated cytidine (14.5–16.0 ppm). The upfield chemical shift of the adenosine H8 protons (7.1–7.3 ppm) establishes that the oligopurine strand adopts an A-helical base stacking conformation in the 11-mer triplex. These results demonstrate that oligonucleotide triple helices can be readily monitored by NMR at the individual base-triple level with distinct markers differentiating between Watson–Crick and Hoogsteen pairing. Excellent exchangeable proton spectra have also been recorded for $(R+)_n\cdot(Y-)_n\cdot(Y+)_n$ 7-mer triple helices with the shorter length permitting spectra to be recorded at ambient temperature. Our NMR studies open up the possibility of studying triple helices containing errors and lesions, as well as the interaction of ligands with this DNA structural motif in aqueous solution.

There is currently great interest in the chemistry and biology of triple-stranded DNA helices and their potential role in the regulation of the eukaryotic genome [reviewed in Wells et al. (1988), Htun and Dahlberg (1989) and Broitman and Fresco (1989)]. The early characterization of the poly(U·A·U) triple helix where the third strand was an oligopyrimidine (Felsenfeld et al., 1957) has been complemented by the more recent characterization of the poly(A·U·A) triple helix where the third strand is an oligopurine (Broitman et al., 1987). Related studies established the sequence requirements (Riley et al., 1966; Morgan & Wells, 1968) and the role of pH (Lee et al., 1984) for triple-helix formation.

Much of the recent interest in triple-helix formation involving (deoxypurine)$_n$·(deoxypyrimidine)$_n$ sequences [designated $(R)_n\cdot(Y)_n$] results from the observation of such tracts at recombination hot spots and upstream from eukaryotic genes [reviewed in Wells et al. (1988)]. The detection of S1 nuclease hypersensitivity in active chicken globin chromatin (Larson & Weintraub, 1982) and its fine mapping to $(R)_n\cdot(Y)_n$ stretches (Nickol & Felsenfield, 1983; Schon et al., 1983) focused much attention on the unusual DNA structure at such sites. The observed hyperreactivity was explained through $(R+)_n\cdot(Y-)_n\cdot(Y+)_n$ triple-helix formation (the signs indicate the directionality of the strands) with the remaining single-stranded $(R-)_n$ segment sensitive to single-stranded nucleases (Christophe et al., 1985; Lyamichev et al., 1986). Triple-helix formation in $(R)_n\cdot(Y)_n$ mirror repeats is favored by both superhelical stress and low pH (Mirkin et al., 1987) and can be readily monitored by two-dimensional gel electrophoresis (Mirkin et al., 1987; Collier et al., 1988) and chemical probes (Hanvey et al., 1988a,b; Voloshin et al., 1988; Htun & Dahlberg, 1988; Johnston, 1988). Finally, chemical footprinting studies on $(G)_n\cdot(C)_n$ sequences in supercoiled plasmid DNA suggest a switch from a $(G+)_n\cdot(C-)_n\cdot(C+)_n$ triple helix in the absence of Mg ions to a $(G+)_n\cdot(C-)_n\cdot(G-)_n$ triple helix

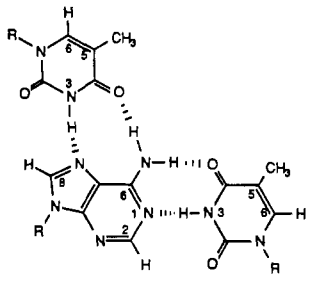
† The research was supported by NIH Grant GM34504. C.d.l.S. is supported, in part, by a fellowship from the CONICET, Republic of Argentina. The NMR spectrometers were purchased from funds donated by the Robert Woods Johnson Trust toward setting up an NMR Center in the Basic Medical Sciences at Columbia University.

0006-2960/89/0428-7282$01.50/0

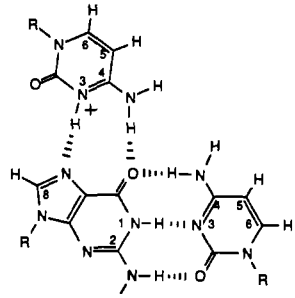
in the presence of Mg ions (Kohwi & Kohwi-Shigemetsu, 1988).

There is also interest in developing sequence-specific nucleic acid binding ligands that recognize 1–2 turns of the double helix. The simplest approach is through triple-helix formation following addition of a third strand that complements the sequence in the duplex. Such probes where the added strand is $(Y)_n$ (Moser & Dervan, 1987; Prasenth et al., 1988) or possibly $(R)_n$ (Cooney, et al., 1988) can then be used to cleave DNA in a sequence-specific manner or to investigate the selective blockage of gene expression.

The current structural view of $(R+)_n\cdot(Y-)_n\cdot(Y+)_n$ triple helices is based on an interpretation of X-ray fiber diffraction data (Arnott & Selsing, 1974), molecular modeling with optimized stereochemistry (Arnott et al., 1976), and chemical footprinting (Moser & Dervan, 1987; Hanvey et al., 1988a,b; Voloshin et al., 1988; Htun & Dahlberg, 1988; Johnston, 1988) experiments. This model predicts that the nucleic acid triple helix adopts an A-form structure with $(Y+)_n$ located in the major groove of the $(R+)_n\cdot(Y-)_n$ duplex and parallel to $(R+)_n$. Further, the model proposed that the $(Y-)_n$ and $(Y+)_n$ strands pair with the $(R+)_n$ strand through Watson–Crick and Hoogsteen pairing, respectively, stabilized by T·A·T **1** and C·G·C$^+$ **2** base triples.



1

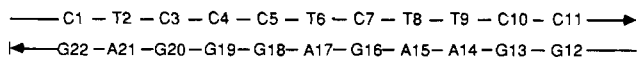


2

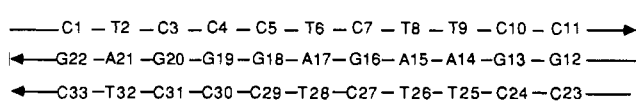
Two-dimensional proton NMR methods have been successfully applied to characterize structural and dynamics features of DNA oligomer duplexes at the individual base pair level in aqueous solution [reviewed in Reid (1987), Patel et al. (1987), and van de Ven and Hilbers (1988)]. The present paper extends this approach to studies of structural features in triple-helical DNA at the oligomer level. Our goal is to identify NMR markers that differentiate between Hoogsteen and Watson–Crick pairing in T·A·T and C·G·C$^+$ base triples and characterize the global conformation of the triple helix in aqueous solution. Exchangeable proton NMR studies are reported on the $(R+)_{11}\cdot(Y-)_{11}$ 11-mer duplex **3** and the $(R+)_{11}\cdot(Y-)_{11}\cdot(Y+)_{11}$ 11-mer triplex **4** at acidic pH and 41 °C. The generality of the observation is established by further studies on a $(R+)_7\cdot(Y-)_7\cdot(Y+)_7$ 7-mer triplex at acidic pH and 30 °C.

Experimental Procedures

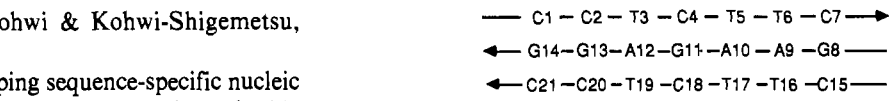
The sequence and numbering system of the $(R+)_n\cdot(Y-)_n$ 11-mer duplex **3**, the $(R+)_n\cdot(Y-)_n\cdot(Y+)_n$ 11-mer triplex **4**, and the $(R+)_n\cdot(Y-)_n\cdot(Y+)_n$ 7-mer triplex **5** are



3



4



5

Oligonucleotide Synthesis. The deoxyoligonucleotide strands were synthesized on a Beckman System 1 plus automated DNA synthesizer using a solid-phase cyanoethyl phosphoramidite method. The crude 5′-dimethoxytritylated oligonucleotides were isolated by treatment of the support with concentrated aqueous ammonia for 46 h at room temperature. The products were purified by reverse-phase HPLC in two stages followed by desalting on a Sephadex G-25 column. The oligonucleotides were finally converted to sodium form by using a Dowex sodium form cation-exchange resin column.

Duplex Preparation. A 1:1 stoichiometry of the 11-mer duplex **3** was achieved at room temperature by monitoring the thymidine methyl resonances during the gradual addition of the purine $(R+)_n$ to the pyrimidine $(Y-)_n$ strand dissolved in phosphate buffer, pH 6.8. The stoichiometry ratio was found to be 100 A_{260} units of the $(Y-)_n$ strand to 142 A_{260} units of the $(R+)_n$ strand to yield 238 A_{260} units of the 11-mer duplex **3**.

Triple-Helix Preparation. The triple helices were formed by adding the third pyrimidine $(Y+)_n$ strand dissolved in phosphate buffer, pH 6.8, to a solution of the duplex dissolved in phosphate buffer, pH 6.8, at the UV stoichiometric ratio found for the formation of the duplex. The triple helices were dialyzed two times against a 0.1 mM EDTA solution for 2 h and once against deionized water for 1 h. The pH was adjusted to 5 with HCl before the NMR spectra were recorded.

NMR Experiments. Proton NMR spectra were recorded by using quadrature detection on Bruker AM 500 and Bruker AM 400 spectrometers. Two-dimensional phase-sensitive (States et al., 1982) NOESY spectra (mixing time 120 ms) were recorded with a jump and return pulse sequence (Plateau et al., 1982) for the reading pulse. The carrier frequency was centered on the H_2O signal, and the maximum excitation was centered at 11 ppm. The time domain data sets consisted of 1024 complex points in the t_2 dimension and 257 increments in the t_1 dimension. The free induction decays (FIDs) were apodized with a 90°-shifted sine bell function zeroed to the 1024th point in the t_2 dimension and to the 256th point in the t_1 dimension before Fourier transformation. The spectra in the t_2 dimension were base-line corrected with a fifth-order polynomial base-line fitting routine supplied by Dr. Arthur Pardi (unpublished program) after Fourier transformation. Two-dimensional data sets were processed with FTNMR software (Dr. Dennis Hare, unpublished program) on VAX 11-780 and micro VAX-II computers. Two-dimensional spectra were symmetrized prior to plotting on a HP 7475 plotter.

Results

11-mer Duplex **3**. The proton NMR spectrum (8–16 ppm) of the 11-mer duplex **3** in H_2O buffer, pH 6.8 at 15 °C, is plotted in Figure 1A. The base-paired imino protons resonate between 12.0 and 14.5 ppm with better resolution for the four thymidine imino protons (13.5–14.5 ppm) relative to the seven guanosine imino protons (12.3–13.5 ppm). The imino protons have been assigned following analysis of the cross peaks in the contour plot (5.5–14.5 ppm in both dimensions) of the two-dimensional NOESY spectrum (mixing time 120 ms) of the 11-mer duplex in H_2O, pH 6.8 at 15 °C (Figure 2A). The NOE cross peaks in expanded NOESY contour plots of the

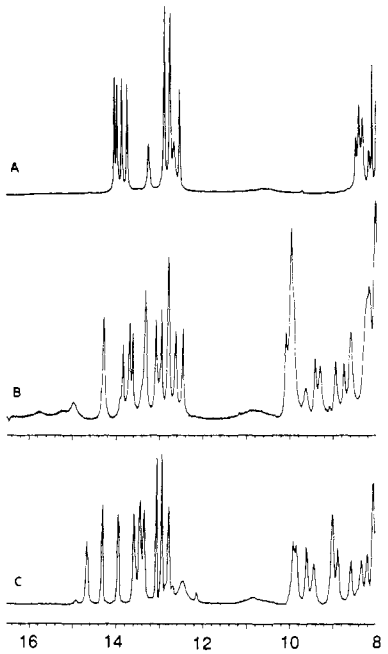


FIGURE 1: 500-MHz exchangeable proton spectra (8.0–16.0 ppm) of (A) the 11-mer duplex **3**, pH 6.8, 15 °C, (B) the 11-mer triplex **4**, pH 5.0, 41 °C, and (C) the 7-mer triplex **5**, pH 5.0, 30 °C in H_2O solution. The purine H8 protons in the purine-rich strand were deuterated by heating to high temperature at basic pH prior to annealment with the pyrimidine-rich strands for the 7-mer triplex **5** in (C).

Table I: Proton Chemical Shifts in the 11-mer Duplex **3** in H_2O Buffer, pH 6.8 at 15 °C

	A·T pair		G·C pair	
	T–H3	A–H2	G–H1	C–H4
C1·G22			13.27	7.80, 7.04
T2·A21	13.99	7.80		
C3·G20			12.78	8.43, 6.88
C4·G19			12.90	8.44, 6.76
C5·G18			12.90	8.36, 6.88
T6·A17	13.76	7.56		
C7·G16			12.55	8.35, 7.00
T8·A15	14.05	7.42		
T9·A14	13.88	7.33		
C10·G13			12.78	8.51, 6.98
C11·G12			12.67	8.20, 7.05

symmetrical imino proton region (Figure 3A) and the region connecting imino protons with amino and base protons (Figure 3B) permit correlations between protons within a base pair and between adjacent base pairs as listed in the caption to Figure 3. The assignments of the thymidine and guanosine imino protons, the cytidine and adenosine amino protons, and adenosine H2 protons in the 11-mer duplex **3** are given in Table I. The observed NOEs between the thymidine imino and adenosine H2 protons (peaks I–M, Figure 3B) establish Watson–Crick A·T pairing, and those between guanosine imino and hydrogen-bonded cytidine amino protons (peaks N–S, Figure 3B) establish Watson–Crick G·C pairing in the 11-mer duplex **3**. Note that the hydrogen-bonded and exposed amino protons of cytidine are separated by ~1.5 ppm (Table I).

11-mer Triplex **4**. The proton NMR spectrum (8–16 ppm) of the 11-mer triplex **4** in H_2O, pH 5.0 at 41 °C, is plotted in Figure 1B. The imino protons exhibit broad resonances between 14.5 and 16.0 ppm and narrow resonances between 12.0 and 14.5 ppm. In addition, narrow exchangeable resonances are detected between 8.5 and 10.5 ppm in the 11-mer triplex **4** (Figure 1B) which are not detected in the 11-mer duplex **3** (Figure 1A).

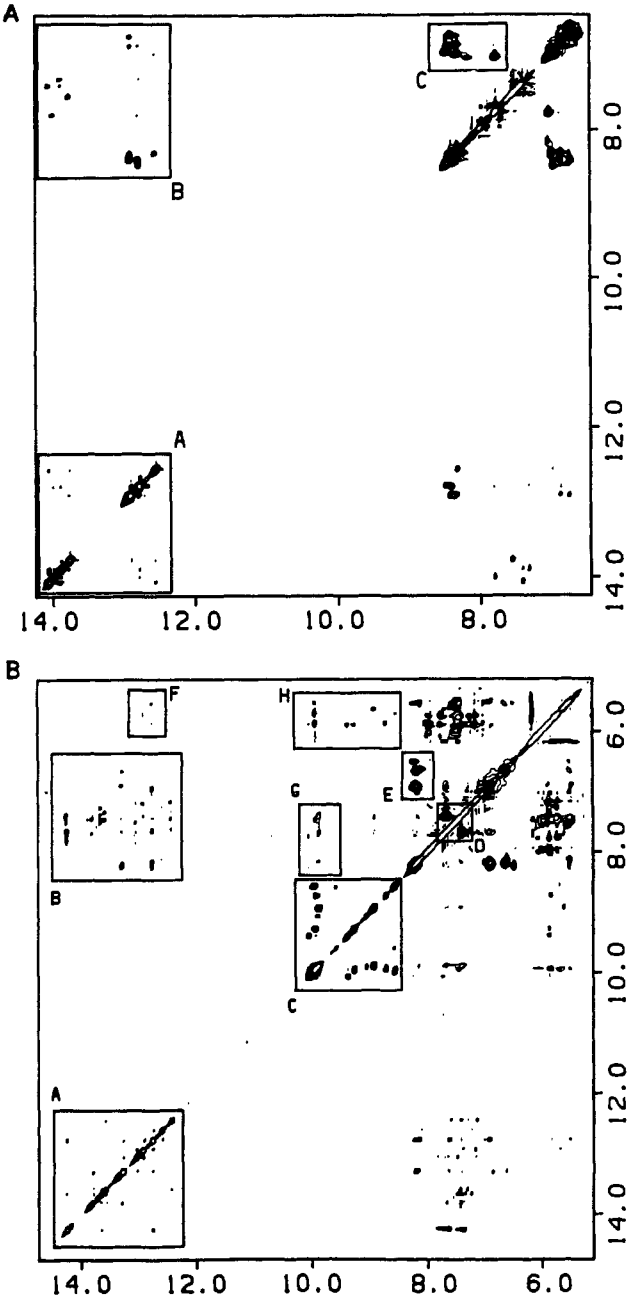


FIGURE 2: (A) 500-MHz NOESY contour plot (120-ms mixing time) of the 11-mer duplex **3** in H_2O, pH 6.8, 15 °C and (B) 400-MHz NOESY contour plot (120-ms mixing time) of 11-mer triplex **4**, pH 5.0, 41 °C, in H_2O solution.

The contour plot (5.5–14.5 ppm in both dimensions) of the NOESY spectrum (120-ms mixing time) of the 11-mer triplex **4** in H_2O, pH 5.0 at 41 °C, is plotted in Figure 2B. We detect NOE cross peaks between exchangeable imino protons (boxed region A, Figure 2B), between exchangeable amino protons on the same (boxed regions C–E, Figure 2B) and adjacent bases (boxed region G, Figure 2B), and between imino protons and amino and nonexchangeable protons (boxed regions B and F, Figure 2B).

Watson–Crick Pairing in T·A·T Triples. The observed NOEs between the imino proton of T and the H2 and hydrogen-bonded amino protons of A permit characterization of Watson–Crick A·T pairing in T·A·T triples **1** in the 11-mer triplex **4**. Thus, the thymidine imino proton at 13.79 ppm exhibits an NOE to the adenosine H2 proton at 7.41 ppm (peak E, Figure 4A) and a weaker NOE to the adenosine hydrogen-bonded amino proton at 7.67 ppm (peak F, Figure

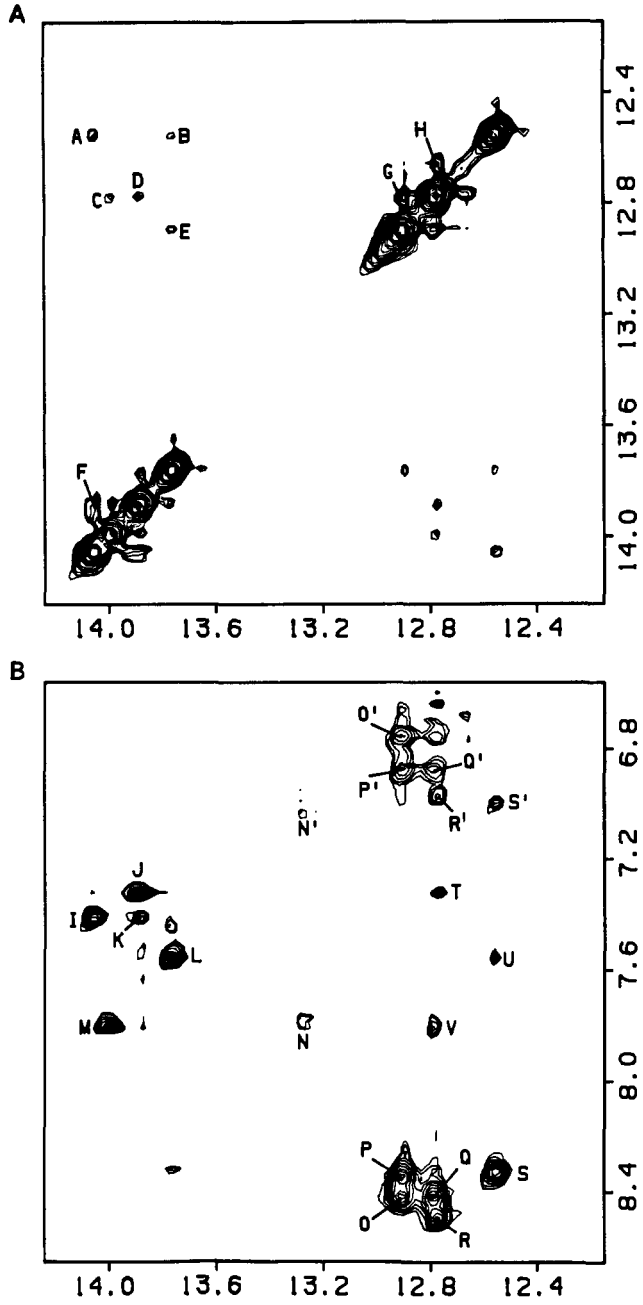


FIGURE 3: (A) Expanded region of the NOESY plot corresponding to boxed region A in Figure 2A establishing distance connectivities among imino protons in the 11-mer duplex **3**. The NOE cross peaks A–H are assigned as follows: (A) T8(imino)–G16(imino); (B) T6-(imino)–G16(imino); (C) T2(imino)–G20(imino); (D) T9(imino)–G13(imino); (E) T6(imino)–G18(imino); (F) T8(imino)–T9(imino); (G) G19(imino)–G20(imino); (H) G12(imino)–G13(imino). (B) Expanded region of the NOESY plot corresponding to boxed region B in Figure 2A establishing distance connectivities among imino protons and the base and amino protons in the 11-mer duplex **3**. The NOE cross peaks I–V are assigned as follows: (I) T8(imino)–A15-(H2); (J) T9(imino)–A14(H2); (K) T9(imino)–A15(H2); (L) T6-(imino)–A17(H2); (M) T2(imino)–A21(H2); (N, N′) G22(imino)–C1(H4,h/e); (O, O′) G19(imino)–C4(H4,h/e); (P, P′) G18-(imino)–C5(H4,h/e); (Q, Q′) G20(imino)–C3(H4,h/e); (R, R′) G13(imino)–C10(H4,h/e); (S, S′) G16(imino)–C7(H4,h/e); (T) G13(imino)–A14(H2); (U) G16(imino)–A17(H2); (V) G20(imino)–A21(H2). The symbols h and e stand for hydrogen-bonded and exposed cytidine amino protons, respectively, involved in Watson–Crick G·C pairing.

4A) across the A·T pair in a Watson–Crick alignment. The adenosine H2 protons can be readily identified by their characteristic long spin–lattice relaxation times in inversion recovery experiments. This approach permits assignment of

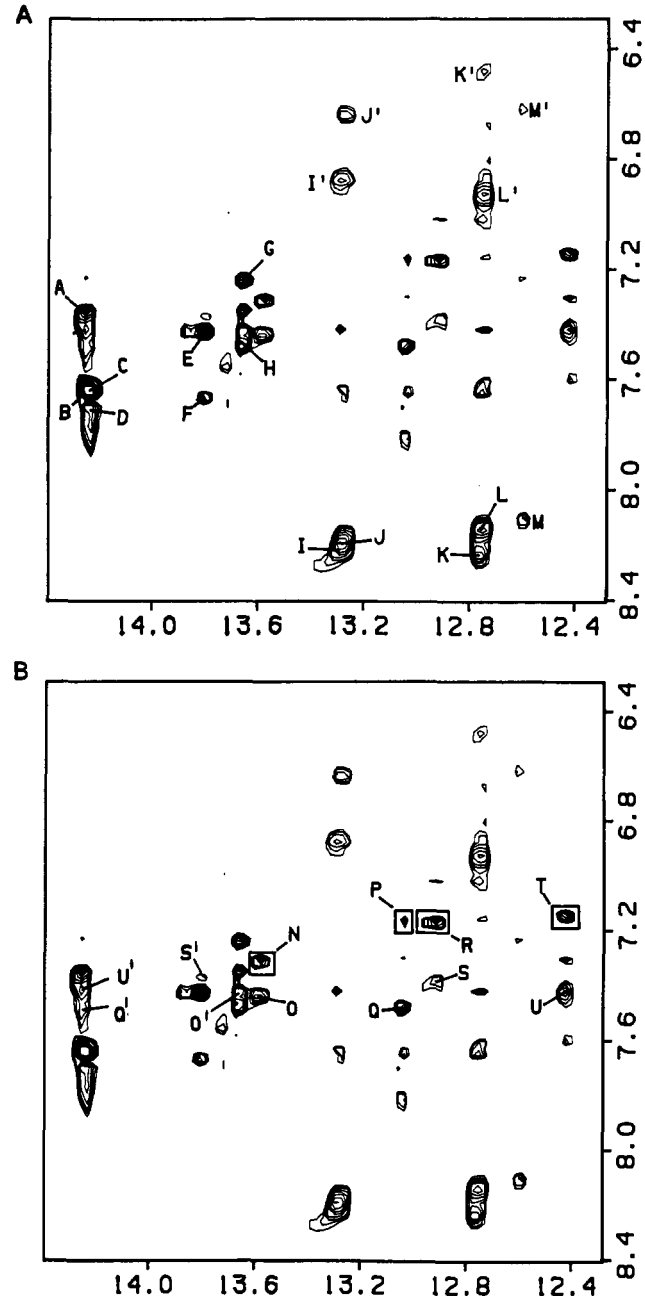


FIGURE 4: Expanded duplicate NOESY contour plots (120-ms mixing time) of the 11-mer triplex **4**, pH 5.0, 41 °C, in H_2O solution establishing distance connectivities between the 12.2–14.4 ppm imino proton region and the 6.4–8.4 ppm amino and aromatic proton region. (A) Assignments of the NOE cross peaks across Watson–Crick A·T pairs in T·A·T base triples and NOE cross peaks across Watson–Crick G·C pairs in C·G·C^+ base triples. The cross peaks A–M are assigned as follows: (A) T8(*imino*)–A15(H2); (B) T8(*imino*)–A15(H6,W–C); (C) T2(imino)–A21(H2); (D) T2(imino)–A21(H6,W–C); (E) T6-(imino)–A17(H2); (F) T6(imino)–A17(H6,W–C); (G) T9(imino)–A14(H2); (H) T9(imino)–A14(H6,W–C); (I, I′) G18(imino)–C5(H4,h/e); (J, J′) G20(imino)–C3(H4,h/e); (K, K′) G19(imino)–C4(H4,h/e); (L, L′) G16(imino)–C7(H4,h/e); (M, M′) G13-(imino)–C10(H4,h/e). The symbol W–C stands for adenosine NH_2 protons directed toward Watson–Crick edges. (B) Assignments of the NOE cross peaks across Hoogsteen A·T pairs in T·A·T triples. The boxed cross peaks N, P, R, and T correspond to H8 protons that disappear in samples of the 11-mer triplex in which all purine H8 protons are deuterated. The cross peaks N–U are assigned as follows: (N) T25(imino)–A14(H8); (O) T25(imino)–A14(H6,H); (P) T32-(imino)–A21(H8); (Q) T32(imino)–A21(H6,H); (R) T28(imino)–A17(H8); (S) T28(imino)–A17(H6,H); (T) T26(imino)–A15(H8); (U) T26(imino)–A15(H6,H); (O′) T9(imino)–A14(H6,H); (Q′) T2(imino)–A21(H6,H); (S′) T6(imino)–A17(H6,H); (U′) T8(imino)–A15(H6,H). The symbol H stands for the adenosine NH_2 proton directed toward the Hoogsteen edge.

Table II: Proton Chemical Shifts in the 11-mer Triplex **4** in H_2O, pH 5.0 at 41 °C

Watson–Crick A·T Pairing in T·A·T Base Triple			
T–H3	A–H2	A–H6	assignment
14.25	7.35	7.60	A15·T8
14.23	7.64	7.71	A21·T2
13.79	7.41	7.67	A17·T6
13.64	7.23	7.47	A14·T9
Hoogsteen A·T Pairing in T·A·T Base Triple			
T–H3	A–H8	A–H6	assignment
13.56	7.31	7.43	A14·T25
13.03	7.17	7.48	A21·T32
12.90	7.17	7.37	A17·T28
12.42	7.14	7.42	A15·T26
Watson–Crick G·C Pairing in C·G·C$^+$ Base Triple			
G–H1	C–H4		assignment
13.29	8.21, 6.87		G18·C5
13.27	8.19, 6.64		G20·C3
12.76	8.24, 6.48		G19·C4
12.74	8.14, 6.93		G16·C7
12.59	8.11, 6.62		G13·C10
Hoogsteen G·C Pairing in C·G·C$^+$ Base Triple			
C$^+$–H3	C$^+$–H4		assignment
	10.06, 9.37		
	9.98, 9.04		
	9.87, 8.91		
	9.94, 9.25		
	9.94, 8.70		
	9.94, 8.55		
	9.58, 8.57		

the 14.25, 14.23, 13.79, and 13.64 ppm resonances to thymidine imino protons in Watson–Crick A·T alignment in T·A·T triples in the 11-mer triplex **4**, and these chemical shifts are summarized in Table II.

Hoogsteen Pairing in T·A·T Triples. The observed NOEs between the imino proton of T and the H8 and hydrogen-bonded amino protons of A permit characterization of Hoogsteen A·T pairing in T·A·T base triples **1** in the 11-mer triplex **4**. Thus, the thymidine imino proton at 12.42 ppm exhibits an NOE to the adenosine H8 proton at 7.14 ppm (peak T, Figure 4B) and an NOE to the adenosine hydrogen-bonded amino proton at 7.42 ppm (peak U, Figure 4B) across the A·T pair in a Hoogsteen alignment. The adenosine H8 proton can be readily identified by selectively deuterating all purine H8 protons in the purine-rich strand by heat treatment at basic pH prior to generation of the triple helix. The NOE cross peaks between the thymidine imino protons and the adenosine H8 protons across the Hoogsteen A·T pair in the T·A·T triple (boxed peaks N, P, R, and T, Figure 4B) are not detected in the 11-mer triplex **4** generated with all purine H8 protons deuterated in the purine-rich strand. This approach permits assignment of the 13.56, 13.03, 12.90, and 12.42 ppm resonances to the thymidine imino protons in Hoogsteen A·T alignment in T·A·T base triples in the 11-mer triplex **4**, and these chemical shifts are summarized in Table II.

One adenosine amino proton is involved in hydrogen bonding in the Watson–Crick A·T alignment while the other adenosine amino proton is involved in hydrogen bonding in the Hoogsteen A·T alignment in the T·A·T base triple **1**. We note that the adenosine amino protons exhibit similar chemical shifts since the cross peaks between them (boxed region D, Figure 2B) are close to the diagonal in the 11-mer triplex **4**.

Watson–Crick Pairing in C·G·C$^+$ Triples. The observed NOEs between the imino proton of G and the amino protons of C permit characterization of Watson–Crick pairing in the C·G·C$^+$ triple **2** in the 11-mer triplex **4**. Thus, the guanosine imino proton at 12.76 ppm exhibits NOEs to the hydrogen-bonded (peak K, Figure 4A) and exposed (peak K′, Figure 4A) amino protons of cytidine across the G·C pair in a Watson–Crick alignment. The nonterminal guanosine imino protons involved in Watson–Crick alignment in the C·G·C$^+$ triple resonate between 12.59 and 13.29 ppm in the 11-mer triplex **4** as summarized in Table II. We detect NOEs between the hydrogen-bonded (8.1–8.25 ppm) and exposed (6.6–6.95 ppm) amino protons of cytidine (box E, Figure 2B) involved in Watson–Crick alignment in the C·G·C$^+$ triple in the 11-mer triplex **4**.

Hoogsteen Pairing in C·G·C$^+$ Triples. The proposed Hoogsteen G·C$^+$ pairing in C·G·C$^+$ base triple **2** is stabilized by two donor hydrogen bonds involving the imino and amino protons of protonated cytidine. Protonation of the cytidine ring should result in downfield shifts of its exchangeable and nonexchangeable protons. This permits assignment of the exchangeable proton resonances between 8.5 and 10.1 ppm to the amino protons of protonated cytidines in C·G·C$^+$ base triples in the 11-mer triplex **4** (Figure 1B). Strong NOEs are detected between the hydrogen-bonded (9.55–10.05 ppm) and exposed (8.55–9.30 ppm) protonated cytidine amino protons involved in G·C$^+$ Hoogsteen alignment (boxed region C, Figure 2B, and peaks H–N, Figure 5B) in the C·G·C$^+$ base triples in the 11-mer triplex **4**. The imino protons of the protonated cytidines in the C·G·C$^+$ base triples resonate as broad peaks between 14.5 and 16.0 ppm (Figure 1B).

It should be noted that the protonated cytidine imino protons (14.5–16.0 ppm) and amino protons (8.5–10.1 ppm) originating in the C·G·C$^+$ base triple in the 11-mer triplex **4** (Figure 1B) are not detected in the corresponding 11-mer duplex (Figure 1A).

Sequence-Specific Assignments in 11-mer Triplex **4**. Some progress has been made toward sequence-specific assignment of the imino protons in the 11-mer triplex **4**. The observation of NOEs from the Watson–Crick and Hoogsteen thymidine imino protons to the same adenosine amino protons permits a correlation of the thymidines in each T·A·T base triple in the 11-mer triplex **4**. Thus, the 7.42 ppm Hoogsteen adenosine amino proton exhibits NOEs to the 12.42 ppm Hoogsteen thymidine imino proton (peak U, Figure 4B) and the 14.25 ppm Watson–Crick thymidine imino proton (peak U′, Figure 4B). In a similar manner, NOE cross peaks O and O′, Q and Q′, and S and S′ in Figure 4B correlate the thymidines in the remaining three T·A·T triples in the 11-mer triplex **4**. There is only one set of adjacent T·A·T triples (T8·A15·T26 and T9·A14·T25), and the observed NOEs between Watson–Crick thymidine imino protons at 13.64 and 14.25 ppm (peak G, Figure 5A) and between Hoogsteen thymidine imino protons at 12.42 and 13.56 ppm (peak A, Figure 5A) identified the (T8–T9)·(A14–A15)·(T25–T26) step in the 11-mer triplex **4**. The 14.25 ppm Watson–Crick thymidine imino proton exhibits an NOE to the 12.74 guanosine imino proton (peak B, Figure 5A), which in turn exhibits an NOE to the 12.90 ppm Hoogsteen thymidine imino proton (peak D, Figure 5A). On the basis of the sequence of the 11-mer triplex **4**, these sets of connectivities identify the (T6–C7–T8–T9)·(A14–A15–G16–A17)·(T25–T26–C27–T28) steps in the 11-mer triplex **4**. The imino and amino proton chemical shift assignments for this segment of the 11-mer triplex **4** are summarized in Table II. We also detect an NOE between the 13.79 ppm Watson–Crick thymidine imino proton of T6 and the 13.29 ppm guanosine imino proton (peak F, Figure 5A) that must be assigned to G18 on the basis of adjacent T6·A17·T28 and C5·G18·C29 triples.

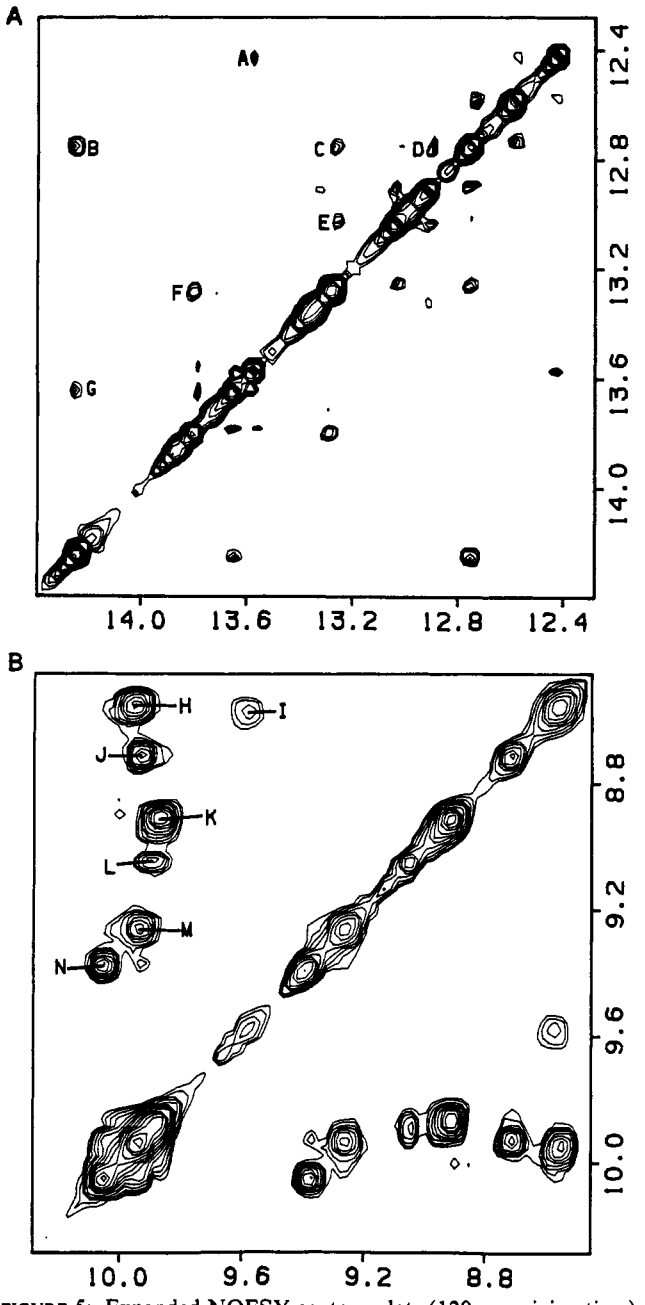


FIGURE 5: Expanded NOESY contour plots (120-ms mixing time) of the 11-mer triplex **4**, pH 5.0, 41 °C, in H_2O solution. (A) NOE cross peaks in the symmetrical 12.0–14.5 ppm imino proton region. The cross peaks A–F are assigned as follows: (A) T25(imino)–T26(imino); (B) T8(imino)–G16(imino); (C) G19(imino)–G20(imino); (D) G16(imino)–T28(imino); (E) G20(imino)–T32(imino); (F) T6-(imino)–G18(imino); (G) T8(imino)–T9(imino). (B) NOE cross peaks in the symmetrical 8.5–10.1 ppm protonated cytidine amino proton region.

This leaves only thymidine T32, and its Hoogsteen imino proton at 13.03 ppm exhibits an NOE to the 13.27 ppm guanosine imino proton (peak E, Figure 5A), which in turn exhibits an NOE to the 12.76 ppm guanosine imino proton (peak C, Figure 5A). These sets of NOEs identify the (T2–C3–C4)·(G19–G20–A21)·(C30–C31–T32) segment of the 11-mer triplex **4**, and the corresponding imino and amino proton assignments are also listed in Table II. This completes the assignment of the nonterminal thymidine imino, guanosine imino, adenosine amino, and cytidine amino protons in the 11-mer triplex **4**. By contrast, the protonated cytidine imino resonances are too broad (Figure 1B), while the protonated cytidine amino protons that give well-resolved NOE cross peaks between hydrogen-bonded and exposed partners (Figure 5B) cannot be unambiguously assigned to specific positions in the 11-mer triplex **4** at this time (Table II).

7-mer Triplex **5**. The proton NMR spectrum (8–16 ppm) of the 7-mer triplex **5** in H_2O, pH 5.0 at 30 °C, is plotted in Figure 1C. The purine H8 protons in the purine-rich strand are deuterated by heating at basic pH prior to annealment with the pyrimidine-rich strands to generate the spectrum of the 7-mer triplex recorded in Figure 1C. The imino proton region for nonterminal base pairs (12.3–14.4 ppm) and the protonated cytidine amino proton region (8.2–10.0 ppm) are unusually well resolved in the 7-mer triplex **5** spectrum at 30 °C. An internal protonated cytidine imino proton is detected as a narrow resonance at 14.66 ppm in the 7-mer triplex **5** at 30 °C and exhibits strong NOEs to its protonated cytidine amino protons at 9.93 and 9.00 ppm.

These data establish that the shorter 7-mer triplex **5** can be studied at ambient temperature (30 °C) (Figure 1C), while narrow resonances are observed for the longer 11-mer triplex **4** only at elevated temperature (41 °C) (Figure 1B).

Discussion

Triple-Helix Formation. The 11-mer triplex **4** was generated by addition of equimolar amounts of the three strands. It became apparent by monitoring the thymidine CH_3 resonances that the triplex was in equilibrium with its 11-mer duplex **3** and single strand and that this equilibrium shifted toward the triplex when the temperature was lowered. The exchangeable imino proton spectrum of the 11-mer triplex **4** at 41 °C (Figure 1B) did not exhibit resonances from the small amount of 11-mer duplex **3** in equilibrium since the imino proton spectra of the latter were exchanged out at this temperature.

The formation of the 11-mer triplex **4** containing a third oligopyrimidine strand parallel to the oligopurine strand requires acidic pH conditions. Our studies were undertaken at pH 5.0, conditions that favored protonation of cytidines in the third oligopyrimidine strand to form the C·G·C^+ base triple. The triple helix dissociates to its duplex and strand components when the pH is raised toward neutrality. The studies on the 11-mer triplex also had to be undertaken at elevated temperatures (41 °C) since the spectra broadened significantly at lower temperatures.

Exchangeable Proton Spectra. The exchangeable proton spectra of the 11-mer triplex **4** in H_2O, pH 5.0 at 41 °C, are unusually well resolved in the imino proton region downfield from 12 ppm (Figure 1B). These imino resonances can be assigned by class, permitting a differentiation between thymidine imino protons involved in Watson–Crick and Hoogsteen pairing in the T·A·T base triple. Similarly, separate markers are available for the guanosine imino and protonated cytidine imino protons in the C·G·C^+ base triple. Further, the amino protons of protonated and unprotonated cytidines are resolved from each other in the C·G·C^+ base triple, providing markers for cytidines involved in both Watson–Crick and Hoogsteen pairing. These results establish that the hydrogen-bonding properties of triple helices consisting of one helical turn can be readily monitored by the exchangable imino and amino proton markers in NMR spectra.

T·A·T Base Triple. The exchangeable protons in the T·A·T triple include the imino protons of thymidine on the two oligopyrimidine strands and the amino protons of adenosine on the oligopurine strand. The experimental data establish that the imino protons of thymidine involved in Watson–Crick pairing (13.65–14.25 ppm) resonate to low field of the imino protons of thymidine involved in Hoogsteen pairing

(12.9–13.55 ppm) for the 11-mer triplex **4** (Table II). Since both imino protons are involved in ring NH to ring N hydrogen bonds in the T·A·T base triple **1**, the results suggest a slightly stronger hydrogen bond in the Watson–Crick pair than in the Hoogsteen pair.

The adenosine amino protons are separated by ~1.5 ppm in Watson–Crick A·T pairs in the 11-mer duplex **3**, reflecting the chemical shift difference between hydrogen-bonded and exposed amino protons. By contrast, the adenosine amino protons are separated by <0.3 ppm in the 11-mer triplex **4** (boxed region D, Figure 4), reflecting the fact that both amino protons are involved in hydrogen bonding to carbonyl groups in the T·A·T base triple **1**. The adenosine amino proton involved in Watson–Crick pairing resonates to low field of its counterpart involved in Hoogsteen pairing in the T·A·T base triple **1**.

C·G·C⁺ Base Triple. The exchangeable protons in the C·G·C⁺ base triple include the imino and amino protons of guanosine, the imino and amino protons of protonated cytidine, and the amino protons of unprotonated cytidine.

The guanosine imino protons in the 11-mer triplex **4** resonate between 12.6 and 13.3 ppm (Figure 1B), which is in the same chemical shift range as guanosine imino protons in the 11-mer duplex **3** (Figure 1A).

The imino protons of the protonated cytidine in the 11-mer triple **4** are broad at 41 °C and resonate between 14.5 and 16.0 ppm. Their low-field chemical shift and fast exchange rates reflect their protonation state.

The amino protons of protonated cytidine resonate ~1.5 ppm downfield from the amino protons of unprotonated cytidine in the 11-mer triplex **4** (Table II), thus clearly distinguishing the former cytidines involved in Hoogsteen pairing from the latter cytidines involved in Watson–Crick pairing. The chemical shift difference between the hydrogen-bonded and exposed protons of the protonated cytidine is ~1.0 ppm, while that between the hydrogen-bonded and exposed protons of the unprotonated cytidine is ~1.5 ppm.

There are NOE cross peaks between the guanosine imino protons and exchangeable protons centered at ~6.0 ppm (boxed region F, Figure 2B) that most likely originate in the averaged amino resonances of guanosine.

Helix Type. The H8 protons of adenosine resonate between 7.1 and 7.3 ppm in the 11-mer triplex **4** (see boxed peaks in Figure 4B). These chemical shifts are upfield from the chemical shifts of adenosine H8 protons in the 11-mer duplex **3** that are centered about ~8.0 ppm. Ring current calculations (Arter & Schmidt, 1976; Giessner-Prettre & Pullman, 1970) establish that purine H8 protons in oligopurine stretches resonate approximately 1 ppm to high field in A-form helices relative to B-form helices. These differences reflect the different stacking geometries in A-form and B-form DNA. The observed upfield shifts of the adenosine H8 protons in the 11-mer triplex **4** are consistent with formation of an A-form base stacking structure for the oligopurine strand in the 11-mer triplex **4** in solution.

A comparison of the chemical shifts of the individual imino protons involved in Watson–Crick A·T and G·C pairing in thee 11-mer duplex **3** (Table I) with those in the 11-mer triplex **4** (Table II) establishes that both downfield and upfield shifts are detected on triplex formation. These shifts in turn must reflect ring current contributions associated with a B-DNA helix in the 11-mer duplex **3** and an A-DNA helix in the 11-mer triplex **4**.

Future Prospects. The above NMR studies on $(R+)_n \cdot (Y-)_n \cdot (Y+)_n$ triple helices can be extended in several directions. This study has focused on oligopyrimidine third strands at acidic pH with T·A·T and C·G·C⁺ base-triple formation. Recent binding studies using affinity columns suggest formation of T·A·A base triples involving A·A Hoogsteen pairing and formation of C·G·G base triples involving G·G Hoogsteen pairing (Letai et al., 1988). We propose to investigate third strands containing both pyrimidines and purines that satisfy the proposed binding code for base-triple formation (Letai et al., 1988).

It has been established that Mg ions induce a switch from C·G·C⁺ base triples to C·G·G base triples at $(G)_n \cdot (C)_n$ stretches in supercoiled plasmid DNA (Kohwi & Kohwi-Shigematsu, 1988). By contrast, the proton NMR spectra of the $(R+)_n \cdot (Y-)_n \cdot (Y+)_n$ 11-mer triplex **4** and 7-mer triplex **5** were essentially unchanged on addition of a few millimolar Mg ions. It is not clear what role Mg ions will play, if any, at the high concentrations of triplex and their associated monovalent counterions used in the NMR studies.

The above application of two-dimensional NMR to triplexes establishes that spectral resolution is available at the individual base-triple level with a range of markers monitoring Watson–Crick and Hoogsteen pairing. The available resolution should permit investigation of triplexes containing helical errors (mismatches, bulges) and lesions (abasic, exocyclic, and O-alkylation sites) as well as stimulate studies of the interactions of ligands with the unoccupied minor groove of this DNA structural motif.

References

Arnott, S., & Selsing, E. (1974) *J. Mol. Biol. 88*, 509.

Arnott, S., Bond, P. J., Selsing, E., & Smith, P. J. C. (1976) *Nucleic Acids Res. 3*, 2459–2470.

Arter, D. B., & Schmidt, P. G. (1976) *Nucleic Acids Res. 3*, 1437–1447.

Broitman, S. L., & Fresco, J. R. (1989) *Prog. Nucleic Acids Res. Mol. Biol.* (in press).

Broitman, S. L., Im, D. D., & Fresco, J. R. (1987) *Proc. Natl. Acad. Sci. U.S.A. 84*, 5120–5124.

Christophe, D., Cabrer, B., Bacolla, A., Targovnik, H., Pohl, V., & Vassart, G. (1985) *Nucleic Acids Res. 13*, 5127–5144.

Collier, D. A., Griffin, J. A., & Wells, R. D. (1988) *J. Biol. Chem. 263*, 7397–7405.

Cooney, M., Czernuszewicz, G., Postel, E. H., Flint, S. J., & Hogan, M. E. (1988) *Science 241*, 456–459.

Felsenfeld, G., Davies, D. R., & Rich, A. (1957) *J. Am. Chem. Soc. 79*, 2023–2024.

Giessner-Prettre, C., & Pullman, B. (1970) *J. Theor. Biol. 27*, 87–95.

Hanvey, J. C., Klysik, J., & Wells, R. D. (1988a) *J. Biol. Chem. 263*, 7386–7396.

Hanvey, J. C., Shimizu, M., & Wells, R. D. (1988b) *Proc. Natl. Acad. Sci. U.S.A. 85*, 6292–6296.

Htun, H., & Dahlberg, J. E. (1988) *Science 241*, 1791–1796.

Htun, H., & Dahlberg, J. E. (1989) *Science 243*, 1571–1576.

Johnston, B. H. (1988) *Science 241*, 1800–1804.

Kohwi, Y., & Kohwi-Shigematsu, T. (1988) *Proc. Natl. Acad. Sci. U.S.A. 85*, 3781–3785.

Larsen, A., & Weintraub, H. (1982) *Cell 29*, 609–622.

Lee, J. S., Woodworth, M. L., Latimer, L. J., & Morgan, A. R. (1984) *Nucleic Acids Res. 12*, 6603–6614.

Letai, A. G., Palladino, M. A., Fromm, E., Rizzo, V., & Fusco, J. R. (1988) *Biochemistry 27*, 9108–9112.

Lyamichev, V. I., Mirkin, S. M., & Frank-Kamenetskii, M. D. (1986) *J. Biol. Mol. Struct. Dyn. 3*, 667–669.

Mirkin, S. N., Lyamichev, V. I., Drushlyak, K. N., Dobrynin, V. N., Fillipov, S. A., & Frank-Kamenetskii, M. D. (1987) *Nature 330*, 495–497.
Morgan, A. R., & Wells, R. D. (1968) *J. Mol. Biol. 37*, 63–80.
Moser, H. E., & Dervan, P. B. (1987) *Science 238*, 645–650.
Nickol, J. M., & Felsenfeld, G. (1983) *Cell 35*, 467–477.
Patel, D. J., Shapiro, L., & Hare, D. (1987) *Q. Rev. Biophys. 20*, 35–112.
Plateau, P., & Gueron, M. (1982) *J. Am. Chem. Soc. 104*, 7310–7311.
Praseuth, D., Perrouault, L., Le Doan, T., Chassignol, M., Thuong, N., & Helene, C. (1988) *Proc. Natl. Acad. Sci. U.S.A. 85*, 1349–1353.
Reid, B. R. (1987) *Q. Rev. Biophys. 20*, 1–34.
Riley, M., Maling, B., & Chamberlin, M. J. (1966) *J. Mol. Biol. 20*, 359–389.
Schon, E., Evans, T., Welsh, J., & Efstratiadis, A. (1983) *Cell 35*, 837–848.
States, D. J., Haberkorn, R. A., & Ruben, D. J. (1982) *J. Magn. Reson. 48*, 286–292.
Van de Ven, F. J., & Hilbers, C. W. (1988) *Eur. J. Biochem. 178*, 1–38.
Voloshin, O. N., Mirkin, S. M., Lyamichev, V. I., Belotserkovskii, B. P., & Frank-Kamenetskii, M. D. (1988) *Nature 333*, 475–476.
Wells, R. D., Collier, D. A., Harvey, J. C., Shimizu, M., & Wohlrab, F. (1988) *FASEB J. 2*, 2939–2949.

Crystallographic Studies of the Mechanism of Xylose Isomerase†

Gregory K. Farber,‡ Arthur Glasfeld,‡ Gérard Tiraby,§ Dagmar Ringe,‡ and Gregory A. Petsko*,‡

Department of Chemistry, Massachusetts Institute of Technology, Cambridge, Massachusetts 02139, and Laboratoire de Microbiologie et Génétique Appliquées du CNRS, Université Paul Sabatier, F-31062 Toulouse Cedex, France

Received February 28, 1989; Revised Manuscript Received May 19, 1989

ABSTRACT: The mechanism of xylose isomerase (EC 5.3.1.5) has been studied with X-ray crystallography. Four refined crystal structures are reported at 3-Å resolution: native enzyme, enzyme + glucose, enzyme + glucose + Mg^{2+}, and enzyme + glucose + Mn^{2+}. One of these structures (E·G·Mg) was determined in a crystal mounted in a flow cell. The other structures were equilibrium experiments carried out by soaking crystals in substrate containing solution. These structures and other studies suggest that, contrary to expectation, xylose isomerase may not use the generally expected base-catalyzed enolization mechanism. A mechanism involving a hydride shift is consistent with the structures presented here and warrants further investigation. Additional evidence in support of a hydride shift comes from comparing xylose isomerase with triosephosphate isomerase which is known to catalyze an analogous reaction via an enediol intermediate. Evidence is presented that suggests that aldose–ketose isomerases can be divided into two groups. Phospho sugar isomerases generally do not require a metal ion for activity and show exchange of substrate protons with solvent. In contrast, simple sugar isomerases all require a metal ion and show very low solvent exchange. These observations are rationalized on the basis of the need for stereospecific sugar binding.

Using X-ray crystallography, we have embarked on a mechanistic study of xylose isomerase (EC 5.3.1.5) with the goal of uncovering the structural basis of its catalysis of the interconversion of xylose and xylulose. Extensive use of xylose isomerase (XyI) is made in industry for the production of high-fructose corn syrup by taking advantage of its fortuitous glucose isomerase activity (Layman, 1986). The enzyme may also prove useful in ethanol production from xylan-containing raw materials (Wilhelm & Hollenberg, 1984). Despite its industrial importance, relatively little is known about the mechanism of XyI.

Much of what has been assumed about the mechanism of xylose isomerase has been inferred by comparison with triosephosphate isomerase (TIM), which catalyzes a chemically analogous reaction, the isomerization of D-glyceraldehyde 3-phosphate to dihydroxyacetone phosphate, via an enediol intermediate (Albery & Knowles, 1976; Rose et al., 1969; Rose, 1981). These comparisons may be misleading. It has been shown that the true substrates for the reaction catalyzed by XyI are α-glucose (Schray & Rose, 1971) and α-fructose (Makkee et al., 1984). Therefore, before catalyzing isomerization, XyI must open the pyranose ring. Another difference between the two reactions is the phosphate group. TIM and other phospho sugar isomerases can use the phosphate group to orient and bind the sugar. The substrates for simple sugar isomerases have no equivalent chemical handle.

There are significant structural similarities between TIM and XyI. Both enzymes are eight-stranded α/β-barrels (Banner et al., 1975; Carrell et al., 1984; Farber et al., 1987), but XyI contains an additional C-terminal domain which is involved in intermolecular contacts. Despite the structural similarity and the analogous chemical transformations catalyzed by the two enzymes, some details of the kinetic behavior show marked difference. TIM requires no cofactor for catalysis and operates at the diffusion-controlled limit (Blacklow et al., 1988). In contrast, XyI requires divalent metal cations, usually Mg^{2+} or Mn^{2+}, for activity and has a k_{cat}/K_m 5 orders of magnitude lower than that of TIM (Suekane et al., 1978). In addition, TIM catalyzes the exchange of protons from

† This work was supported by NIH Grant GM26788 to G.A.P. and D.R.
‡ Massachusetts Institute of Technology.
§ Université Paul Sabatier.

0006-2960/89/0428-7289$01.50/0