Unusual DNA Conformation at Low pH Revealed by NMR: Parallel-Stranded DNA Duplex with Homo Base Pairs†

Howard Robinson,‡ Gijs A. van der Marel,§ Jacques H. van Boom,§ and Andrew H.-J. Wang*,‡

Biophysics Division and Department of Cell and Structural Biology, University of Illinois at Urbana-Champaign, Urbana, Illinois 61801, and Gorlaeus Laboratories, Leiden State University, Leiden 2300RA, The Netherlands

Received May 14, 1992; Revised Manuscript Received August 24, 1992

ABSTRACT: We have investigated the conformational potentials of several DNA oligonucleotides containing sequences related to 5′-CGA in neutral pH and low pH (<5.0) conditions. One-dimensional proton NMR spectra show that d(CGATCG), d(TCGATCGA), and d(CGATCGATCG) exhibit new sets of resonances at low pH (~3.8–4.4), when compared to those from the neutral pH samples. The low pH form and the neutral pH form are in slow equilibrium. Analyses of the data suggest that these sequences under low pH conditions adopt structures distinct from B-DNA. Two-dimensional nuclear Overhauser effect spectroscopy (2D NOESY) data from the DNA hexamer d(CGATCG) of the neutral and low pH samples were used to analyze their respective structures in solution. An iterative NOE spectral-driven refinement procedure, SPEDREF [Robinson, H., & Wang, A. H.-J. (1992) *Biochemistry 31*, 3524–3533], was used to show that the neutral pH structure is close to canonical B-DNA. In contrast, analysis of the low pH form using the 2D NOESY data suggests that its structure is consistent with a right-handed *parallel-stranded* (PS) double helix with symmetrical non-Watson–Crick (C^+:C, G:G, A:A, T:T) homo base pairs. Supporting evidence for the PS helix includes the asymmetric inversion–recovery relaxation times associated with the two ends of the helix. The structure is favored by the 5′-CGA sequence in which the cytosines provide the C^+:C pairing for the nucleation step and the GpA step is significantly stabilized by the *inter*strand G–A stacking interactions. This new and unusual structure is different from any of the parallel-stranded structures previously described, and it may have biological roles, such as in DNA recombination or in RNA structures.

The polymorphic behavior of DNA conformation is now a well-established phenomenon. It depends not only on the nucleotide sequence but also on extrinsic factors such as counterions, humidity, or pH. One of the most extensively studied examples is the left-handed Z-DNA structure, which is highly favored by the alternating dG–dC[1] sequence (Wang et al., 1979; Rich et al., 1984). In addition, the equilibrium between the B-DNA and Z-DNA is influenced by the above-mentioned factors. For example, high concentration of salt (NaCl, $MgCl_2$) converts poly(dG–dC) from the B form to the Z form (Wang et al., 1979). Similarly, low pH conditions also favor the Z-DNA conformation for poly(dG–dC) (Chen, 1984).

Another interesting form of DNA structure is the triple-stranded H-DNA whose formation is more readily observed in a G-rich region, such as $(dG)_n$·$(dC)_n$, in solution at low pH (<5) (Htun & Dahlberg, 1988; Johnston, 1988; Pulleyblank et al., 1985). The presumed model of this structure involves a protonated oligo(dC^+) strand folding back to the major groove side of the double helix and forming the triple (dC)·(dG)·$d(C^+)$ base pairs using a Hoogsteen pairing geometry between the dG and the protonated dC^+ (Moser & Dervan, 1987). The ease with which this triplex structure can form is associated with the moderately high pK_a value of cytosine N^3 in DNA (Inman, 1964).

When the N^3 of cytosine is protonated, it can no longer pair to guanine in the normal Watson–Crick geometry. This has an interesting consequence of potentiating other possible base pairing schemes, including the Hoogsteen geometry with two hydrogen bonds (HN^3 of the protonated C^+ to N^7 of G and HN^4 of C^+ to O^6 of G). Another possibility is the self-pairing between a protonated C^+ and a neutral C with three hydrogen bonds. Several double-stranded helical structures involving this type of C^+:C base pairs, e.g., poly(dC)·poly(dC^+), have been studied by optical spectroscopic methods (Antao et al., 1986, 1990; Edwards et al., 1990).

We are interested in exploring stable alternative DNA structures in "normal" DNA sequences under the influence of low pH. Many DNA oligonucleotides have been used in the crystallization experiments under low pH conditions with the hope of obtaining crystals of protonated DNA molecules. We discovered that certain DNA oligomers, including d(CGATCG), can only be crystallized at low pH conditions. These observations have prompted us to analyze their structures in solution by NMR spectroscopy. We found that the DNA oligonucleotides d(CGATCG), d(TCGATCGA), and d(CGATCGATCG), which have sequences related to 5′-CGA, adopt a structure different from B-DNA at pH below 5.5. A plausible model, consistent with the observed 2D NOE data, is one with parallel-stranded double helices in which all base pairs are of the non-Watson–Crick self-pairing type, i.e., A with A, T with T, G with G, and finally C with C^+. In this paper, we present evidence that supports this novel DNA structure and discuss its possible biological relevance.

† This work was supported by NIH Grants GM-41612 and CA-52506 to A.H.-J.W. This work was presented at the 32nd ENC meeting (poster no. 71) in St. Louis (1991).

* Corresponding author.

‡ University of Illinois at Urbana-Champaign.

§ Leiden State University.

[1] Abbreviations: A, T, G, and C, adenine, thymine, guanine, and cytosine or their respective nucleotides; NMR, nuclear magnetic resonance; 2D NOESY, two-dimensional nuclear Overhauser enhancement spectroscopy; LPH, low pH conformer; PSts and PSte, parallel-stranded structure with shortened and extended T:T base pairs.

0006-2960/92/0431-10510$03.00/0

MATERIALS AND METHODS

The oligonucleotides were synthesized either according to a procedure published earlier (van der Marel et al., 1981) or on the Applied Biosystems DNA synthesizer. Samples for ^{1}H NMR analysis were prepared in D_2O as described earlier (Robinson & Wang, 1992). The pH 7.0 preparation contained 0.9 mM duplex d(CGATCG) with 150 mM NaCl and 50 mM sodium phosphate buffer (pH 7.0) in the 0.5-mL sample. The pH 4.2 preparation contained 1.9 mM duplex d(CGATCG) and 200 mM NaCl in the 0.5-mL sample. The pH of this unbuffered solution was monitored and adjusted at each lyophilization step and after the collection of NMR data. For the pH measurements in D_2O solutions, the pH reading in D_2O was calibrated against titrations of pH standards in D_2O on the same electrode. All reported pH measurments were made at 25 °C.

The ^{1}H NMR spectra were recorded on a GE GN-500-MHz spectrometer. The chemical shifts (in ppm) are referenced to the HDO peak which is calibrated to 2,2-dimethyl-2-silapentane-5-sulfonate (DSS) at different temperatures. The 2D NOESY were collected by the States et al. (1982) method with a composite pulse NOESY sequence. The spectra were recorded with 512 t_1 complex blocks of 2048 complex points each (in the t_2 dimension) and averaged for 16 scans per block. The total recycle time for the 200-ms NOESY experiment with the pH 4.2 sample was 7.6 s. The 2D data sets were processed with the program FELIX version 1.1 (Hare Research, Woodinville, WA) using the Silicon Graphics workstations using the truncation apodization described earlier (Robinson & Wang, 1992) and exponential multiplication at 4 Hz. An iterative procedure was used to refine the model derived from the two-dimensional NOESY data. The detailed discussion of the refinement procedure has been described elsewhere (Robinson & Wang, 1992). This procedure adjusts a model so as to minimize differences between the observed 2D NOE spectrum and a spectrum simulated from the refining model. The simulation uses a full matrix relaxation where the cross-relaxation rate constants were derived form isotropic rigid rotation spectral density functions. The molecular correlation time for these spectral density functions were optimized on the basis of matching the cytosine H5–H6 distances and also by analysis of the aggregate of all the derived distance errors (see Supplementary Material, Figures 2SA and 2SB). A 7-ns correlation time was picked for the refinement of the pH 4.2 sample of d(CGATCG) at 5 °C. The X-PLOR (Brünger, 1990) NOE strength scalar was set to 20, and the SPEDREF refinements were carried out to 40 cycles for each model.

RESULTS

pH-Dependent Structural Transition. One-dimensional ^{1}H NMR spectra of the aromatic and methyl regions of d(CGATCG) at various pH's are shown in Figure 1. At pH 7.0, the hexamer is in the right-handed B-DNA conformation, which is confirmed by the structural refinement as described elsewhere (Robinson & Wang, 1992). When the pH of the solution is gradually titrated to the lower values, an additional set of resonances starts to emerge. These new resonances are in slow exchange with the corresponding resonances of the B-DNA conformer. Interestingly, the resonances associated with the low pH form (abbreviated LPH conformer) have significantly broader line widths than those of the B-DNA form at neutral pH. For example, the line width at half-height of the LPH conformer proton G2H^8 is 19.2 Hz and, for the neutral B-DNA conformer, 7.9 Hz.

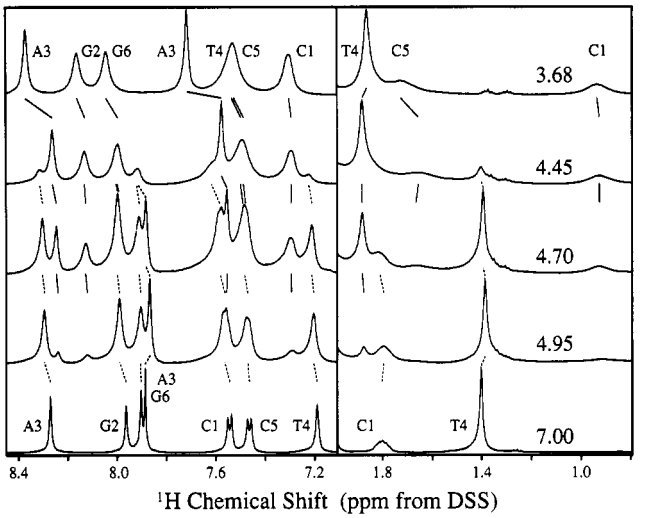

FIGURE 1: One-dimensional NMR spectra of nonexchangeable DNA aromatic and methyl protons of d(CGATCG) in D_2O at the indicated pH values at 10 °C. There are two distinct spectra associated with neutral pH and low pH solutions, suggesting that two DNA conformations are in slow equilibrium. This is evident in the spectrum at pH 4.70 where the two conformational populations are about equal. The individual resonances are identified and traced by solid lines for the low pH conformer and dashed lines for the neutral pH conformer. Note that the resonances associated with the A3 residue are substantially more downfield at pH 3.68.

The changes in the chemical shifts of the two forms are quite dramatic. For example, the thymine methyl resonance moved downfield by 242 Hz from 1.417 ppm (at pH 7.0) to 1.90 ppm (at pH 4.2). (Chemical shifts and line widths are listed in Table IS in the Supplementary Material.) Similarly, the C1H$^{2'}$ resonance moves upfield by 477 Hz from 1.909 to 0.956 ppm. When the pH is further lowered below pH 4.2 on to 3.68, there are additioinal large changes in chemical shifts for some of the resonances. In particular, the resonances associated with the adenine A3 base moves further downfield (from 7.60 to 7.72 ppm for the A3H^2 proton and from 8.14 to 8.38 ppm for the A3H^8 proton), suggesting that this adenine may become protonated at the N^1 position. Smaller changes are seen for some of the protons of the remaining nucleotides. In contrast to the conformational change associated with the pK_a around 4.7, where slow exchange is indicated by the presence of two distinct lines for each resonance, the protonation at lower pH (with a pK_a about 3.5) is associated with conformations that are in very rapid exchange (see Figure 1).

The pK_a of the LPH conformational transition (4.7) is very close to the pK_a of the cytosine base in DNA (Inman, 1964). It is therefore reasonable to assume that the cytosines are first protonated as the pH is lowered. But how many C's are protonated, and is the LPH conformer a duplex? We have carefully titrated a 5.3 mM duplex solution of the hexamer with HCl. The titration curve is shown in Figure 2, and the shape of the curve is consistent with there being *two* protons absorbed per B-DNA dimer, resulting in the new LPH conformation that is itself a homodimer. The sample at pH 4.2 contains about 96% doubly-protonated LPH hexamer duplex. The ^{1}H NMR spectra show that this new LPH homodimer is symmetric, since only one resonance is observed for each pair of sequence-related protons (the same holds for the B-DNA conformer). Thus, a multistranded species like the triplex structure is not possible, as the latter is not consistant with the 2-fold symmetry observed in the NMR spectra. The pK_a of this titration is 4.72, consistent with the NMR titration study. The shape of the curve also shows that additional protons are picked up by the molecule below pH 4.2, which

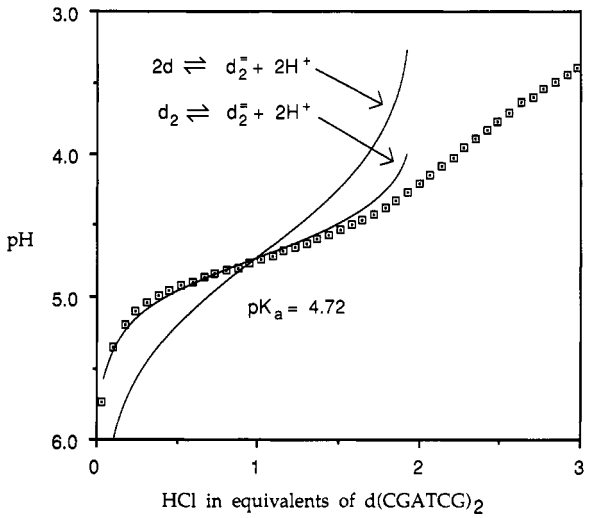

FIGURE 2: Titration of the DNA hexamer d(CGATCG) in H_2O. The experimental data are shown as squares. The data agree with the prediction of a model with two protons being picked up by a dimer molecule (duplex) to produce a new dimer, but not with the model where the product is two single-stranded molecules.

is consistent with the NMR data suggesting that adenine is protonated with a pK_a at about 3.5.

Structural Analysis. We have examined the solution structure of the LPH conformer by a quantitative analysis of 2D NOE spectrum by the procedure SPEDREF (Robinson & Wang, 1992). Figures 3 and 4 show sections of the observed 2D NOE spectrum with lines to indicate the connectivity of cross peaks which support the assignment (exchange cross peaks at a slightly higher pH also support this assignment, data not shown). The patterns of cross peaks are very similar to those that would be expected, if the backbone of the LPH form were in a B-DNA conformation. For this reason, we first tried refining a B-DNA model against this observed data. The lower halves of Figures 3 and 4 show sections of the simulated spectrum from this refinement. The SPEDREF procedure has produced a model (Figure 5) whose simulated spectrum is very close to the observed spectrum. The *R* factor between the integrals of both spectra is 16.7%, which is indicative of a very good match.

Despite the quantitatively good fit between the simulated spectrum for the refined B-DNA like model, there is evidence from both the 1D NMR titrations (Figure 1) and 2D NOE spectra that show that the gross characteristics of this B-DNA model are not correct. First, we have observed that two protons have been absorbed and that the most likely site of protonation is at cytosines' N^3. A proton at this location would conflict with the hydrogen-bonding pattern of Watson–Crick base pairing between cytosine and guanine. Second, the spectra clearly show that the LPH conformation is in slow exchange with the unprotonated B-DNA conformation, thus the LPH conformation must be grossly different from B-DNA. Third, the prominent cross peak between A3H^2 and G2H^8 cannot be accounted for by any moderate adjustment of the B-DNA model by the SPEDREF procedure.

The observed protonation and the slow exchange between the B-DNA and LPH conformers could be explained by protonation of the N^3's of two of the four cytosines and flipping of their associated guanines (to syn conformation) to form Hoogsteen base pairing. Inspection of the NOE's between both G2H^8 and G6H^8 and their H$^{2'}$ and H$^{2''}$s (see Figure 4) clearly indicates that neither of these guanines can be in syn conformations as these NOE's are all very large. In addition, we have constructed models with syn conformations for both

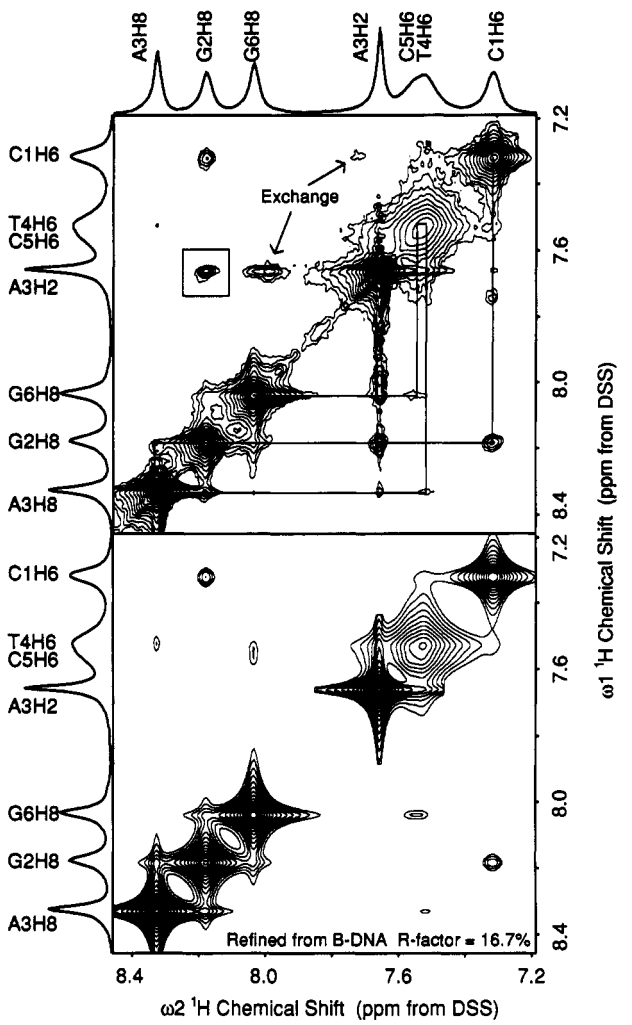

FIGURE 3: Aromatic to aromatic section of the phase-sensitive 2D NOESY spectra of the d(CGATCG) duplex at pH 4.2. The mixing time was 200 ms, and the temperature was 5 °C. The top panel shows the experimental NOESY cross peaks between the aromatic protons (6.2-8.6 ppm). The lower panel is the simulated NOE spectra of the same region based on the refined B-DNA model (*R* factor = 16.7%). Note that the strong experimental NOE cross peak between the A3H^2 and G2H^8 protons is missing in the simulated spectrum of the refined B-DNA. The intrastrand distance between these two protons is 8.5 Å.

guanines, and both were pulled back toward anti conformations as expected by the size of the cross peaks of their H^8 protons to their own H$^{2'}$ and H$^{2''}$ (spectra and structural information from these refinements are in the Supplementary Material).

The observed NOE cross peaks in Figures 3 and 4 are indicative of the local backbone conformation being close to B-DNA. But there seems no way that the cross peak between A3H^2 and G2H^8 can be explained via a model with a B-DNA-line conformation. This difficulty arises because the error in one cross peak (A3H^2 to G2H^8) has been sacrificed by the refinement minimizer due to the overwhelming consensus from the other cross peaks. The contradiction is that the LPH cross peaks support a conformation that is locally very similar to B-DNA; however, the gross structure must be entirely different. We suggest that the LPH spectrum arises from a parallel arrangement of the two identical DNA strands, rather than the antiparallel pairing of B-DNA.

Support for a parallel strand motif comes from measurements of the relaxation of each spin via inversion–recovery, shown in Figure 6 for both the B-DNA and LPH conformers. At pH 7, where the B-DNA conformer predominates, spins at both ends relax much faster than spins near the middle. The

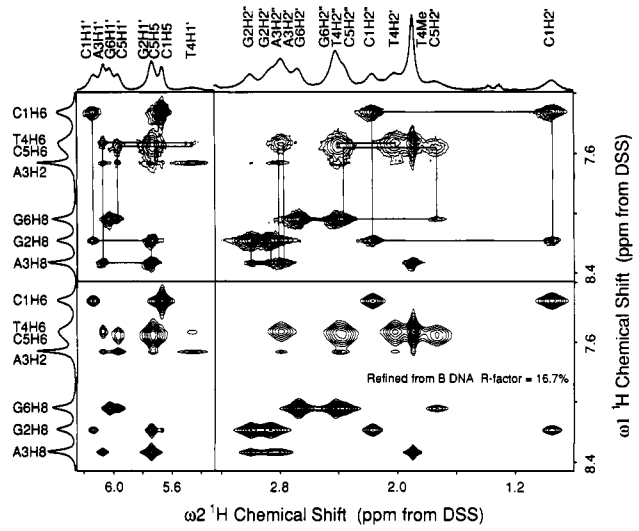

FIGURE 4: Portions of the phase-sensitive 2D NOESY spectra of the d(CGATCG) duplex at pH 4.2 which provides key information in the assignments of resonances, the glycosyl conformation, and sugar puckers. The top panels are the experimental NOESY cross peaks between the aromatic protons (6.2–8.6 ppm) and H1′ (5.5–6.2 ppm), and the $H^{2'}$, $H^{2''}$ (4.7–5.2 ppm) and $H^{1'}$ (5.5–6.2 ppm). The lower panels are the simulated NOE spectra of the same regions based on the refined B-DNA model (R factor = 16.7%).

simplified rotational model of a static structure with rigid isotropic tumbling time predicts that the ends would relax slower than the middle. This is expected because there are fewer neighbors for the spins near the ends such that contributions to dipolar-facilitated relaxations are reduced. The observation for the B-DNA conformer clearly shows that this consideration is vastly overshadowed by other relaxation mechanisms. The most important is probably the increased degree of internal motion of B-DNA associated with the ends due to reduced constriction of the molecule relative to the middle (Figure 6, lower panel).

The observed relaxation of the LPH conformer, however, is slowed at one end of the molecule and sped up at the other, with the faster side similar to the B-DNA conformer in their relaxation rates (Figure 6B). It would be impossible to have the base partners (i.e., C1 and G6) behaving so differently if the base paring were antiparallel. This is strongly suggestive of a parallel, rather than antiparallel, base-paired structure for the LPH conformer. At the 5′ end of the molecule, the molecule behaves more like a rigid structure where fewer neighbors at the end contribute to the relaxation, resulting in a longer relaxation time. At the 3′ end, the relaxation pathways seem to be dominated via increased internal motion.

There are seven possible ways for homo-base-pairing of a parallel strand scheme as diagrammed in Figure 7. For two cytosines, a protonation at N^3 of one cytosine would allow for a hemiprotonated C^+:C base pair with a very strong hydrogen-bonded network and explain why only two protons are absorbed per dimer. For adenine, guanine, and thymidine, there are two conformations each that provide for a hydrogen-bonded network. We have labeled these possibilities "shortened" or "extended", on the basis of the distances between their $C^{1'}$ carbons. Since the local structure of the LPH conformation is similar to that of B-DNA, we constructed models by starting with one strand of a B-DNA $d(CGATCG)_2$ duplex and rotating it to rejoin the other strand in a parallel manner by using the program MIDAS (Ferrin et al., 1988) on a Silicon Graphics IRIS computer. This model was then idealized and energy-minimized with the constraint of the desirable homo-base-pairing schemes and with the imposition of an exact 2-fold symmetry along the helix axis by using the program NUCLSQ (Hendrickson & Konnert, 1979; Westhof et al., 1985). This process removes all the bad van der Waals contacts in the model. There are two possible ways each for the A:A, G:G, and T:T self-pairs. In the PSts model, four different homo-base-pairings [C^+:C, A:A (short), G:G (short), and T:T (short)] out of the the seven possible homo-base-pairings are used. This model was subjected to the SPEDREF NOE-constrained refinement, resulting in an NMR R factor of 16.5%. Sections of the simulated spectrum for the refined PSts model are shown in Figure 8. A related PSte model, which differs from the PSts model (ts stands for T shortened) only in the T:T base pair with a extended scheme, has been refined to an NMR R factor of 16.7%.

Two Parallel-Stranded Duplexes. The refined structure PSts model, shown in Figure 9, possesses a molecular 2-fold axis coincided with the helix axis, in contrast to the case in B-DNA where the 2-fold axis is perpendicular to the helix axis. The two grooves of the double helix are identical. All nucleotides have an anti glycosyl conformation. In this model, there are very good base stackings as evident in Figure 9. The double helix seems to segregate into two halves, with the bases stacking from the top to the bottom of the helix. The sugar–phosphate backbone makes a sharp bend at the phosphate linkage between the ApT step due to the stretched GpA step. From Figure 9, it can be seen that the homo base pairs have large dihedral angles between the two bases. For example, the two G:G base pairs have large propeller twist angles (~20°). This is consistent with the observations found in other structures of shortened G:G base pair (Coll et al., 1990). A very interesting consequence of the PS helix is that there is a very substantial base stacking between G2 from one strand and A3 from the *opposite* strand. This strong stacking interaction, more clearly seen in Figure 10, apparently provides the necessary stabilizing force to hold the PS helix together. The structural parameters (including the backbone torsion angles, sugar puckers, and glycosyl torsion angles) and the atomic coordinates of both the refined PSts and PSte models are listed in Tables 2S and 3S in the Supplementary Material.

The unique stacking between G and A makes the $A3H^2$ and $G2H^8$ protons have a close contact (3.6 Å) which gives rise a strong NOE cross peak. In the *anti*parallel-stranded B-DNA, the two protons are quite far apart, either in the case of the intrastranded G to A distance (8.67 Å) or in the case of the interstranded G to A distance (11.6 Å). With a symmetric homodimer, an ambiguity exists in that both partners in the homodimer have identical chemical shifts, such that cross peaks may be interpreted as arising from either intra- or interstrand NOE's. Since there is no information in the observation to attribute a cross peak intensity to either the inter- or intrastrand NOE, SPEDREF does not presuppose this missing information and interprets each NOE on the basis of the refining model.

It should be pointed out that there are multiplicities in symmetric homo-base-pairings. Therefore, it is possible that multiple conformations may coexist in equilibrium for the structure of the LPH conformer. This is suggested by the broad resonances in the spectra of the LPH form. One alternative model is to replace the T:T (shortened) base pair in the PSts model with the T:T (extended) base pair in the PSte model. Notice that the transition between those two types of T:T base pairs involves only the sliding of the two bases. The structural consequence of the T vs T sliding appears to have minimal effects on the PS helix. The apparent dynamics associated with this exchange do cause difficulty in

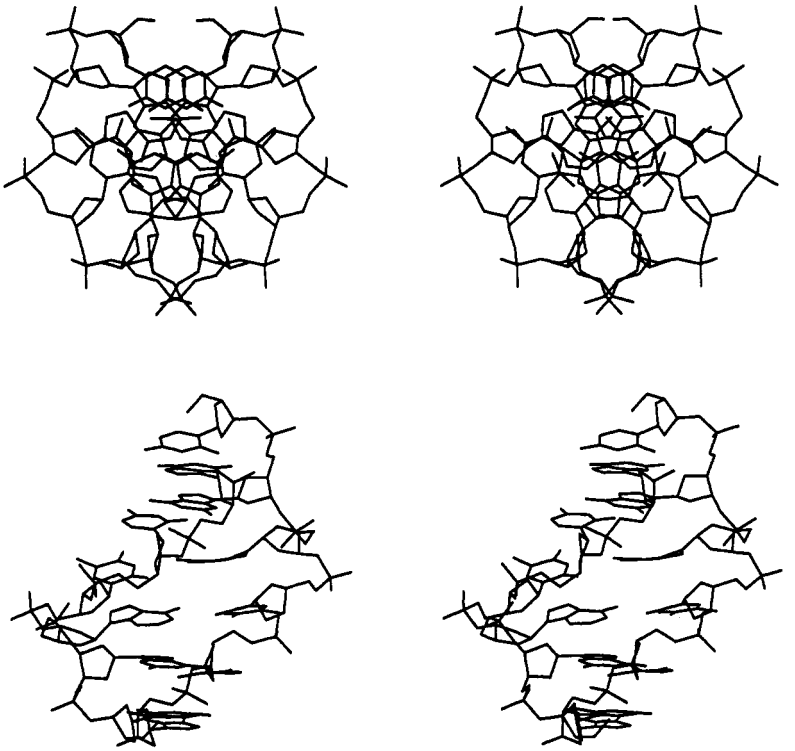

FIGURE 5: Stereoview of the skeletal model of one of the B-DNA like refined models seen in top and side views. All protons have been omitted for clarity.

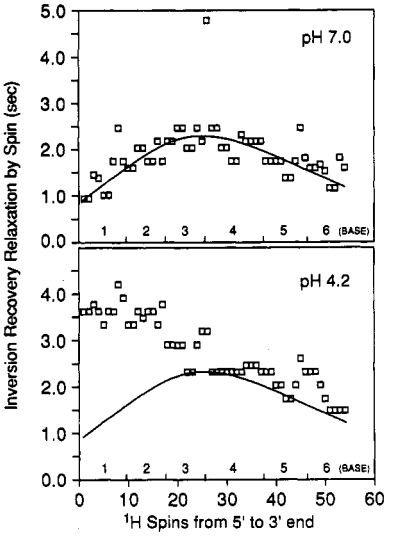

FIGURE 6: Inversion–recovery relaxation time of all spins in the neutral pH form and the low pH form of d(CGATCG). In the neutral pH form, all the resonances fall in the range of 1.0–2.5 s (except for A3H2) and have a symmetrical relation between the two ends where the relaxation is faster than the middle. This is consistent with B-DNA, which has a 2-fold symmetry in the middle of the duplex (between the two A–T base pairs) perpendicular to the helix axis. In contrast, the spins in the LPH conformer have a very different distribution of their relaxation times in going from one end of the molecule to the other end. The data suggest that the duplex has a 2-fold symmetry along the helix axis.

the application of our current relaxation theory in the SPEDREF refinement procedure. Thus, a high accuracy refinement of the low pH structure is precluded. Although the PSts model refines to a slightly lower R factor than the PSte model (16.5% versus 16.7%), the refinement procedure pulls the T4 of the PSts model part way toward the PSte model. This is approximately what would be expected if the two were in dynamic exchange. That is, the best single model would be an average.

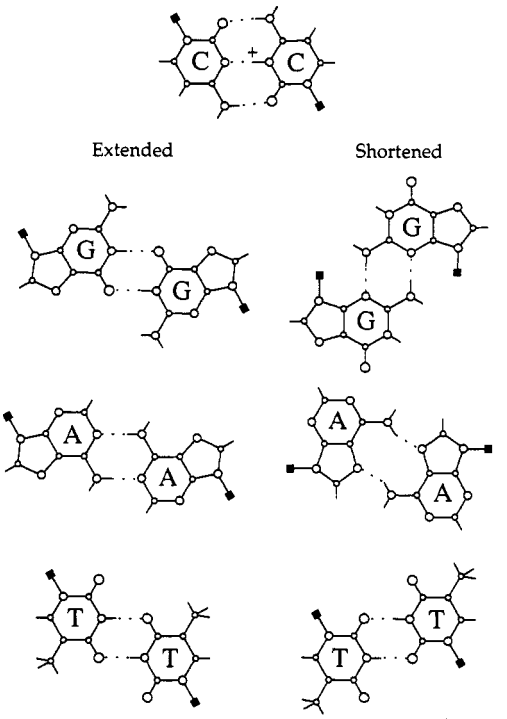

FIGURE 7: Seven different symmetrical homo base pairs. The C^+:C base pair is hemiprotonated with the proton shared by the two N^3 sites of the cytosines. There are two possible pairing schemes each for the other bases (G:G, A:A, and T:T). On the left side, the $C^{1'}$–$C^{1'}$ distances in the base pairs are longer than the corresponding pairings on the right side.

The observation that the resonances associated with T are considerably broader than the resonances of the preceding three residues is consistent with the conclusion of exchange for T. It is also evident that all resonances observed at pH 4.2 are broader than the corresponding resonances for the B-DNA conformer at pH 7. The reason for this broadening

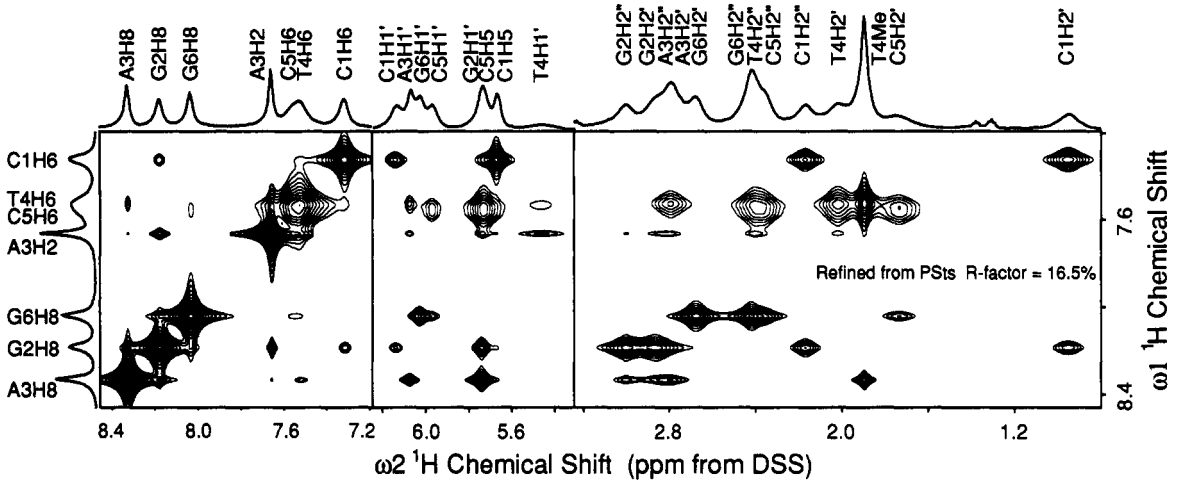

FIGURE 8: Simulated sections of the NOESY spectrum for the PSts refined model (R factor = 16.5%). The NOE cross peak between the $A3H^2$ and $G2H^8$ protons is now consistent with the observed spectrum shown in Figure 3 (upper section).

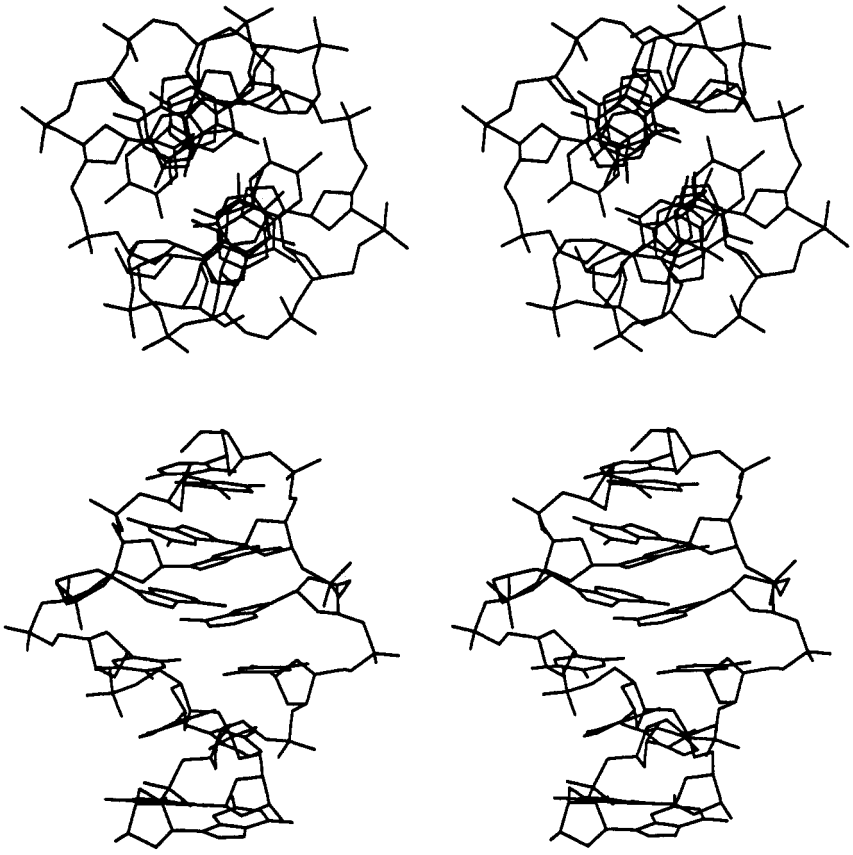

FIGURE 9: Stereoview of the skeletal model of the PSts refined model seen in top and side views. All protons have been omitted for clarity.

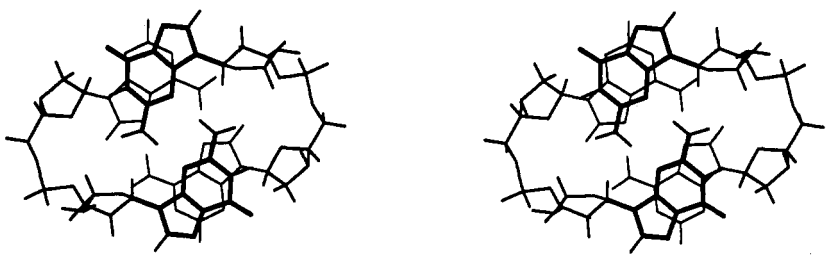

FIGURE 10: Stereoview of the stacking of the G2 and A3 base pairs for the PSts refined model showing the close distance between the interstrand guanine H^8 and adenine H^2 protons (3.56 Å). Also note the stacking arrangement where the overlap of aromatic ring systems is interstrand.

is partly due to conformation exchange of the type discussed above for the T residue, but it also is due to the exchange between the LPH conformers and the B-DNA conformers. This exchange is directly observable in the upper section of Figure 3. The exchange cross peak between the LPH (~96%) and B-DNA (~4%) conformers for the $A3H^2$ resonance are clearly observable due to the large differential chemical shifts and the relative sharpness of these resonances.

In principle, imino proton spectra in H_2O should provide additional information that would support the models that we present here. We believe that the novelty of the non-Watson–Crick parallel base pairings affects the assumptions about chemical shifts for the imino and amino protons when compared to B-DNA. Thus, many ambiguities are present in this data which we are at present attempting to resolve with studies of related sequences (data not shown).

Other DNA Sequences. We have examined several other DNA oligonucleotides to see whether they exhibit similar pH dependent structural changes. Two other DNA oligomers, d(TCGATCGA) and d(CGATCGATCG), show a similar behavior, but not the sequences d(CGTACG), d(GCGTACGC), d(GCGATCGC), d(CGCTAGCG), and d(CGCGATCGCG) as judged by their NMR spectra (data not shown). These results suggest that sequences with repeats of 5′-CGA have the propensity to adopt the low pH conformation. This observation is consistent with the structural property of the PS models discussed above. The 5′-cytosine provides the protonation site under low pH conditions for the nucleation of the PS structure. The extensive interstrand G–A stacking interactions provide the necessary key stabilizing forces, so that the LPH conformer is favored over the B-DNA at low pH. It is possible that a repeating sequence of 5′-CGA, e.g., 5′-CGACGACGACGA, may form an even more stable structure, as the competing structure is now a single-strand form. Experiments are in progress to test this. It is worth noting that another type of strong interstrand G–A stacking interaction has been seen in an oligonucleotide containing G–A mismatch base pairs (Li et al., 1991).

In addition to the three molecules [d(CGATCG), d(TCGATCGA), and d(CGATCGATCG)] that have been shown to exist as two pH-dependent structures, an earlier NMR study of the tetramer d(TCGA) has suggested a similar behavior, though no explicit structure has been proposed in the latter case, except implying that Hoogsteen base pairing is possible (Topping et al., 1988). That study had observed a number of similar properties as in the present study. For example, Topping et al. (1988) noticed that the low pH form is quite stable at 5 °C.

The new structure of the low pH form represents one of the stable alternative DNA structures that have been found to be in slow exchange (on the NMR time scale) with the right-handed B-DNA under suitable conditions in solution. The best known example is the left-handed Z-DNA associated with alternating dG–dC sequence which has been clearly shown by NMR studies to be in slow equilibrium with B-DNA in 35% methanol solution (Feigon et al., 1984, 1985).

DISCUSSION

The structure of DNA is strongly influenced by the pH of the solution. In particular, DNA polymers containing long stretches of cytosines adopt unusual structures at low pH. Poly(C) is believed to exist as a parallel-stranded duplex due to the protonation of the cytosine bases in DNA which allows C^+:C base pairs to form (Hartman & Rich, 1965). Other dC-rich polymers have been shown to have similar properties (Antao et al., 1986, 1990; Edwards et al., 1990). An NMR study has shown that the DNA oligomer d(CTCTCTCT) formed a duplex structure with C^+:C base pairs, but the authors concluded that the thymine bases are extruded out of the helix at low pH conditions, as NOE's characteristic of C–T stacking could not be observed (Sarma et al., 1986). In our study of the LPH conformer of duplex d(CGATCG), there is clear evidence of stacking with indication of dynamic exchange for the T:T base pair.

So far, this phenomenon of pH-dependent conformation appears to occur only in DNA molecules which are abundant in cytosines. The observation that sequences associated with d(CGA) adopt parallel-stranded DNA structure raises the issue of the generality of parallel-stranded nucleic acid structures. In fact, they are not so uncommon. As mentioned earlier, in H-DNA triple helix, one of the pyrimidine strands pairs to the purine strand in a parallel direction using Hoogsteen base pairing (Moser & Dervan, 1988). Nucleic acids containing α-nucleotides have been postulated to form a parallel duplex structure with the complementary strand of normal β-nucleic acid (Thuong et al., 1987; van Genderen et al., 1990). Alternatively, van der Sande et al. (1988) have designed DNA molecules with special sequence complementarity so that they form parallel duplex with Watson–Crick base pairs. Finally, model system of the tetrastranded structure of the poly(dG) molecule has been shown to contain parallel-stranded structures with guanine–guanine base pairs (Kang et al., 1992).

The model we propose here is completely different from those mentioned above. Our PS duplex models contain non-Watson–Crick symmetric homo-base-pairings which have been observed in a number of crystal structures of nucleosides and nucleotides. For example, d(CpG) crystallized as a mini-parallel-stranded duplex with C^+:C and G:G (Coll et al., 1987; Cruse et al., 1983). A detailed conformational comparison between the two d(CpG) fragments in the hexamer d(CGATCG) PS model (Table IISa) and the d(CpG) crystal structure (Coll et al., 1987) shows remarkable agreement. No torsional angle deviates by more than 15° between the two structures. Finally, in the complex of d(CpA)–proflavine crystal structure, proflavine is intercalated in the parallel-stranded minihelix with C^+:C and A:A pairs (Westhof & Sundaralingam, 1980).

Additional studies have shown that other self-pairings are not an uncommon phenomenon. In addition to the examples described above, A:A (Rich et al., 1961), T:T (van Genderen et al., 1990; Camerman et al., 1990), and G:G (Coll et al., 1990) symmetric pairings have also been seen. Therefore, it would not be surprising that some nucleic acid sequences adopt such symmetric homopairing, perhaps facilitated by the protonation of cytosines, which would result in parallel helical structures.

Whether such parallel structures are relevant in biological systems remains to be determined. It is interesting to note that when two B-DNA duplexes of homologous sequences are brought next to each other (as in DNA recombination) forming an intermediate structure, there are two pairs of parallel strands in such structure. It is also likely that such structures may be found in highly folded RNA molecules. In transfer RNA structures, segments of RNA backbone are running in a parallel orientation. For example, the TψC loop is paired to the D loop with two base pairs (G18–ψ55 and G19–C20), and the two backbones are parallel (Kim et al., 1974; Robertus et al., 1974). Finally, the uniqueness of such structures may be recognized by drugs or proteins. We have crystallized the 2:1 complex of actinomycin D and d(CGATCGATCG) from a solution at pH 3.2 but not from solution at higher pH's (unpublished results). Since this decamer has the unique structure at low pH presented in this paper, work is in progress to determine whether the same structure also exists in the drug–DNA complex.

ACKNOWLEDGMENT

We thank Dr. Y.-C. Liaw for his assistance in producing some starting models in the early stage of the project.

Technical help from Z. Gan and V. Mainz of the School of Chemical Sciences Molecular Spectroscopy Lab is gratefully acknowledged.

SUPPLEMENTARY MATERIAL AVAILABLE

Tables IS, IIS, and IIIS listing chemical shifts and line widths, torsion angles, and atomic coordinates (for the refined PSte and PSts models); Figure 1S, giving 1D temperature spectra; Figure 2S, giving a determination of correlation time; and Figures 3S, 4S, and 5S, giving simulated spectra and models for various refined structures (41 pages). Ordering information is given on any current masthead page.

REFERENCES

Antao, V. P., Gray, C. W., Gray, D. M., & Ratliff, R. L. (1986) *Nucleic Acids Res. 14*, 10091–10112.

Antao, V. P., Ratliff, R. L., & Gray, D. M. (1990) *Nucleic Acids Res. 18*, 4111–4122.

Brünger, A. T. (1990) *X-PLOR, Version 3.0*, The Howard Hughes Medical Institute and Yale University, New Haven, CT.

Camerman, N., Mastropaolo, D., & Camerman, A. (1990) *Proc. Natl. Acad. Sci. U.S.A. 87*, 3534–3537.

Chen, F.-M. (1984) *Biochemistry 23*, 6159–6165.

Coll, M., Solans, X., Font-Altaba, M., & Subirana, J. A. (1987) *J. Biomol. Struct. Dyn. 4*, 797–811.

Coll, M., Sherman, S. E., Gibson, D., Lippard, S. J. & Wang, A. H.-J. (1990) *J. Biomol. Struct. Dyn. 8*, 315–330.

Cruse, W. B. T., Egert, E., Kennard, O., Sala, G. B., Salibury, S. A., & Viswamitra, M. A. (1983) *Biochemistry 22*, 1833–1839.

Edwards, E. L., Patrick, M. H., Ratliff, R. L., & Gray, D. M. (1990) *Biochemistry 29*, 828–836.

Feigon, J., Wang, A. H.-J., van der Marel, G. A., van Boom, J. H., & Rich, A. (1984) *Nucleic Acids Res. 12*, 1243–1263.

Feigon, J., Wang, A. H.-J., van der Marel, G. A., van Boom J. H., & Rich, A. (1985) *Science 230*, 82–84.

Ferrin, T. E., Huang, C. C., Jarvis, L. E., & Langridge, R. (1988) *J. Mol. Graphics 6*, 13–27.

Hartman, K. A., & Rich, A. (1965) *J. Am. Chem. Soc. 87*, 2033–2035.

Hendrickson, W. A., & Konnert, J. H. (1979) in *Biomolecular Structure, Conformation, Function and Evolution* (Srinvasan, R., Ed.) pp 43–57, Pergamon, Oxford.

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

Inman, R. B. (1964) *J. Mol. Biol. 9*, 624–637.

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

Kang, C. H., Zhang, X., Ratliff, R., Moyzis, R., & Rich A. (1992) *Nature 356*, 126–131.

Kim, S. H., Suddath, F. L., Quigley, G. J., McPherson, A., Sussman, J. L., Wang, A. H.-J., Seeman, N. C., & Rich, A. (1974) *Science 185*, 435–440.

Li, Y., Zon, G., & Wilson, W. D. (1991) *Proc. Natl. Acad. Sci. U.S.A. 88*, 26–30.

Moser, H. E., & Dervan, P. B. (1987) *Science 238*, 645–650.

Pulleyblank, D. E., Haniford, D. B., & Morgan, A. R. (1985) *Cell 42*, 271–280.

Rich, A., Davies, D. R., Crick, F. H. C., & Watson, J. D. (1961) *J. Mol. Biol. 3*, 71–86.

Rich, A., Nordheim, A., & Wang, A. H.-J. (1984) *Annu. Rev. Biochem. 53*, 791–846.

Robertus, J. D., Ladner, J. E., Finch, J. T., Rhodes, D., Brown, R. S., Clark, B. F. C., & Klug, A. (1974) *Nature 250*, 546–551.

Robinson, H., & Wang, A. H.-J. (1992) *Biochemistry 31*, 3524–3533.

Sarma, M. H., Gupta, G., & Sarma, R. H. (1986) *FEBS Lett. 205*, 223–229.

States, D. J., Haberkorn, R. A., & Ruben, D. J. (1982) *J. Magn. Reson. 48*, 286–292.

Thuong, N. T., Asseline, U., Roig, V., Takasugi, M., & Helene, C. (1987) *Proc. Natl. Acad. Sci. U.S.A. 84*, 5129–5133.

Topping, R. J., Stone, M. P., Brush, C. K., & Harris, T. M. (1988) *Biochemistry 27*, 7216–7222.

van der Marel, G. A., van Boeckel, C. A. A., Willie, G., & van Boom, J. H. (1981) *Tetrahedron Lett. 22*, 3887–3888.

van der Sande, J. H., et al. (1988) *Science 241*, 551–557.

van Genderen, M. H. P., Hilbers, M. P., Koole, L. H., & Buck, H. M. (1990) *Biochemistry 29*, 7838–7845.

Wang, A. H.-J., Quigley, G. J., Kolpak, F. J., Crawford, J. L., van Boom J. H., van der Marel, G. A., & Rich, A. (1979) *Nature 282*, 680–686.

Westhof, E., & Sundaralingam, M. (1980) *Proc. Natl. Acad. Sci. U.S.A. 77*, 1852–1856.

Westhof, E., Dumas, P., & Moras, D. (1985) *J. Mol. Biol. 184*, 119–145.