States, D. J., Haberkorn, R. A., & Ruben, D. J. (1982) *J. Magn. Reson. 48*, 286.
Wang, A. H.-J., Quigley, G. J., Kolpak, F. J., van der Marel, G., van Boom, J. H., & Rich, A. (1981) *Science 211*, 171–174.
Wang, A. H.-J., Fujii, S., van Boom, J. H., van der Marel, G. A., van Boeckel, S. A. A., & Rich, A. (1982) *Nature 299*, 601–604.
Wang, Y., Thomas, G. A., & Peticolas, W. L. (1987) *J. Biomol. Struct. Dyn. 5*, 249–274.
Wemmer, D. E., Chou, S. H., & Reid, B. R. (1984a) *J. Mol. Biol. 180*, 41–60.
Wemmer, D. E., Chou, S. H., Hare, D. R., & Reid, B. R. (1984b) *Biochemistry 23*, 2262–2268.
Zimmerman, S. B., & Pheiffer, B. H. (1981) *Proc. Natl. Acad. Sci. U.S.A. 78*, 78–82.

The Z–Z Junction: The Boundary between Two Out-of-Phase Z-DNA Regions†

Brian H. Johnston,*,‡,∥ Gary J. Quigley,§,∥ Michael J. Ellison,∥,⊥ and Alexander Rich∥

Molecular Biology Department, SRI International, 333 Ravenswood Avenue, Menlo Park, California 94025, Department of Chemistry, Hunter College, City University of New York, New York, New York 10021, Department of Biochemistry, University of Alberta, Edmonton, Alberta, Canada T6A2H7, and Department of Biology, Massachusetts Institute of Technology, Cambridge, Massachusetts 02139

Received December 19, 1990; Revised Manuscript Received February 25, 1991

ABSTRACT: The boundary between two segments of Z-DNA that differ in the phase of their syn–anti alternation about the glycosidic bond is termed a Z–Z junction. Using chemical probes and two-dimensional gel electrophoresis, we examined a Z–Z junction consisting of the sequence d[$(CG)_8C(CG)_8$] inserted into a plasmid and used energy minimization techniques to devise a three-dimensional model that is consistent with the available data. We show that both alternating CG segments undergo the B–Z transiton together to form a Z–Z junction. The junction is very compact, displaying a distinctive reactivity signature at the two base pairs at the junction. In particular, the 5′ cytosine of the CC dinucleotide at the junction is hyperreactive toward hydroxylamine, and the two guanines of the GG dinucleotide on the complementary strand are less reactive toward diethyl pyrocarbonate than are the surrounding Z-DNA guanines. Statistical mechanical treatment of the 2-D gel data yields a ΔG for forming the Z–Z junction equal to 3.5 kcal, significantly less than the cost of a B–Z junction and approximately equal to the cost of a base out of alternation (i.e., a Z-DNA pyrimidine in the syn conformation). The computer-generated model shows little distortion of the Z helix outside of the central two base pairs, and the energy of the structure and the steric accessibility of the reactive groups are consistent with the data.

Z-DNA is a left-handed conformation of the double helix, characterized by a dinucleotide repeat in which anti and syn conformations of the bases alternate in succession along the chain [for review, see Rich et al. (1984)]. Because purine residues adopt the syn conformation more easily than do pyrimidine residues, Z-DNA forms most readily in sequences having alternations of purines and pyrimidines (APP sequences). Those APP sequences that most easily form Z-DNA, alternating cytosine and guanine residues [$d(CG)_n$], have been intensively studied, but long sequences of this type have not been found in natural DNA. On the other hand, long stretches of $d(CA)_n$·$d(GT)_n$ ($n \approx 15$–30) are found scattered throughout the eukaryotic genome, about once every 50–100 kb, as a form of middle repetitive DNA (Hamada et al., 1982). In addition, long APP tracts having mixed sequence have identified in many sites, including regulatory regions, in various genomes. All the above sequences can be shown to form Z-DNA in vitro under negative superhelical tension (Rich et al. 1984). [Of the simple APP sequences, only $d(AT)_n$ appears not to readily form Z-DNA (Ellison et al., 1986).] Although the biological role of Z-DNA is not well understood, there is evidence that Z-DNA forms in vivo in *Escherichia coli* (Haniford & Pulleyblank, 1983; Jawarsky et al., 1987; Rahmouni & Wells, 1989) as well as in eukaryotic nuclei capable of carrying out replication and transcription (Wittig et al., 1989).

In addition to strictly alternating purine–pyrimidine sequences, natural DNA contains many examples of alternating sequences in which the phase of alternation changes in the midst of the sequence. If such a sequence is in the Z conformation, a Z–Z junction will exist at the change in phase. For example, the sequence 5′-$(CG)_n(GC)_n$-3′ could form Z-DNA in each of the segments within parentheses but their phasing would be different. Thus, if the cytosines are maintained in the anti conformation and the guanines in the syn conformation, the sequence of conformations at the center of the sequence (...CGCGGCGC...) on one strand would be anti-syn-anti-syn-syn-anti-syn-anti (syn-anti-syn-anti-anti-syn-anti-syn on the other strand). It is the two adjacent syn

†Supported by NIH Postdoctoral Fellowship 3F32GM09031 and NIH Grant 1R29GM41423 to B.H.J.; NIH Grant GM41359, PSC-CUNY Grant 668255 and NIH Research Centers in Minority Institutions Award RR-03037 to G.J.Q.; and grants from the NIH, the NSF, the Office of Naval Research, NASA, and the American Cancer Society to A.R. M.J.E. was supported by a postdoctoral fellowship from the Medical Research Council of Canada. This work was first presented at the Fifth Conversation in Biomolecular Stereodynamics, Albany, NY, 1987.

* To whom correspondence should be addressed at SRI International.
‡ SRI International.
∥ Massachusetts Institute of Technology.
§ City University of New York.
⊥ University of Alberta.

0006-2960/91/0430-5257$02.50/0

(anti) conformations that would represent the Z–Z junction. (Even if not all the cytosines and guanines are respectively anti and syn, a junction consisting of adjacent syn and anti conformations will still occur.) Evidence for the existence of Z–Z junctions was first obtained through a study of chemical reactivity, which allowed one to determine which residues are in the anti conformation and which are in the syn conformation (Johnston & Rich, 1985). This was possible because the reactivity of a base can depend not only on the type of base but also on its conformation. It was found that the pattern of reactivity of adjacent purines at a Z–Z junction was different from that of the same bases within a block of Z-DNA, and this suggested that the Z–Z junction had a characteristic conformaation.

A survey of the nucleotide sequences found in regulatory regions of genes reveals many instances of APP sequences containing phase shifts representing candidate Z–Z junctions. The presence of these sequences raises questions as to whether Z–Z junctions exist in vivo and whether they might be specific recognition elements for regulatory proteins. Although some additional work has been carried out on Z–Z junctions (Mirkin et al., 1987; McClean & Wells, 1988; McClean et al., 1988), more information is needed concerning the nature of these junctions, specifically, the size of the Z–Z junction, the energy cost of forming a Z–Z junction, and its structure. Using chemical modification studies, we show here that the Z–Z junction is a very compact structure in which the changes in conformation occur within two residues. Using two-dimensional gels, we show that the energy of Z–Z junction formation is considerably less than the energy of B–Z junction formation. Finally, by juxtaposing two out-of-phase Z-DNA segments and examining the possible joining structures, we have been able to construct a plausible Z–Z junction model, the conformation of which has been refined by energy-minimization calculations. The conformation and energy of this model are in agreement with experimental data.

Materials and Methods

Plasmid Construction and Manipulations. Two complementary DNA 35-mers, 5′-d[GATC$(CG)_8$G$(CG)_8$]-3′ and 5′-d[GATC$(CG)_8$C$(CG)_8$]-3′, were synthesized by the phosphoramidite method. These two single-stranded oligodeoxynucleotides were annealed to each other and cloned into the *Bam*HI site of pBR322. A plasmid, pZ-GG-Z, containing the correct insert sequence as determined by the method of Maxam and Gilbert (1980) was isolated and used in all subsequent experiments. An identical plasmid, but containing three successive guanosine residues interrupting the alternating CG sequence instead of two, called pZ1 in Ellison et al. (1985) and pZ-GGG-Z here, was contructed in a similar fashion. Toposiomers of pZ-GG-Z, pZ-GGG-Z, pLP32 (Peck & Wang, 1982), and pBR322, ranging from the relaxed configuration to a linking difference (α) of approximately −30 turns, were prepared and electrophoresed on two-dimensional agarose gels as described previously (Peck & Wang, 1983; Ellison et al., 1985).

Chemical Reactivity Measurements. Chemical modification reactions on linearized plasmids or plasmids at native superhelical density were performed as previously described (Johnston & Rich, 1985). The chemically modified plasmids were digested with *Taq*I. To visualize the strand containing the --CC-- sequence, the 5′ ends were labeled by dephosphorylating the fragment with calf intestinal phosphatase (Boehringer-Mannheim), followed by treatment with polynucleotide kinase (New England Biolabs) and [γ-^{32}P]ATP (New England Nuclear). The opposite strand was visualized by labeling at the 3′ end with use of the Klenow fragment of DNA polymerase I (New England Biolabs) and [α-^{32}P]dCTP. Subsequent digestion with *Sph*I yielded fragments labeled only at the end containing the insert, and these fragments were gel-purified, cleaved at modified sites by treatment with piperidine at 90 °C, electrophoresed on a sequencing gel, and autoradiographed.

Computer Modeling Studies. Construction of a model for the structure of the Z–Z junction of pZ-GG-Z was approached as follows. Initial coordinates were taken from the d(CpGpCpGpCpG) structure (Wang et al., 1979) that was the basis for the Z_I and Z_{II} DNA models (Wang et al., 1981) as well as most subsequent Z-related structures (Crawford et al., 1980; Fujii et al., 1985; Ho et al., 1985). The third base pair was excised, leaving the phosphates on both sides. The initial two base pairs (the floating portion) were translated by 3.7 Å along the helix axis toward the remainder of the molecule (the fixed portion). The floating portion was then rotated and translated in the plane perpendicular to the helix axis to optimize stacking between the two fragments in a manner consistent with the need to link the adjacent residues. At this point the redundant phosphate groups from the two portions, which were partially overlapping, were reduced to single linking phosphates by selecting reasonable initial positions.

The program AMBER (Weiner & Kollman, 1981) was then utilized to idealize and evaluate this starting model, since we have extensive experience with it on both Z-DNA and other DNA and drug-DNA complexes (Kozelka et al., 1985, 1986, 1987). The general procedure consists of first idealizing out the major distortions, then relaxing the immediate vicinity of the Z–Z junction, and finally performing a global minimization. The steps were carried out by the use of variably weighted quadratic constraints on the positions of the various atoms relative to their initial positions in the model. For stage one, all the bases, the riboses not attached to the linking phosphate, and the remote phosphates were highly constrained (100 kcal/Å^2), the adjacent riboses were constrained at an intermediate value (10 kcal/Å^2), and the phosphate groups were unconstrained. This yielded a sterically plausible structure after minimization. The next step consisted of reducing the intermediate constrained atoms to 1 kcal/Å^2 and the constraints on the adjacent bases and phosphates to 10 kcal/Å^2. After several additional cycles of lowering and dropping constraints, a final model was arrived at by minimizing with all constraints to the original model removed. Globally minimized models of the original $(CG)_3$ (Wang et al., 1979) and a 13-mer of B-DNA containing all nearest-neighbor stacking combinations built from standard coordinates for idealized B-DNA (Arnott et al., 1976) were used for comparison.

Results

Chemical Reactivity

Z–Z Junctions. To see whether the putative Z–Z junction exhibits any unusual chemical reactivity that might provide a signature for its presence and perhaps some insight into its structure, the reactivity of plasmid pZ-GG-Z with diethyl pyrocarbonate (DEP) or hydroxylamine ($HONH_2$) was examined as described in Materials and Methods. DEP is hyperreactive with purines that are in the syn conformation in Z-DNA (Johnston & Rich, 1985; Herr, 1985), and $HONH_2$ is hyperreactive at B–Z junctions (Johnston & Rich, 1985) and single-stranded regions (Rubin & Schmid, 1980; Johnston, 1988a). An autoradiogram of the resulting sequencing gel is

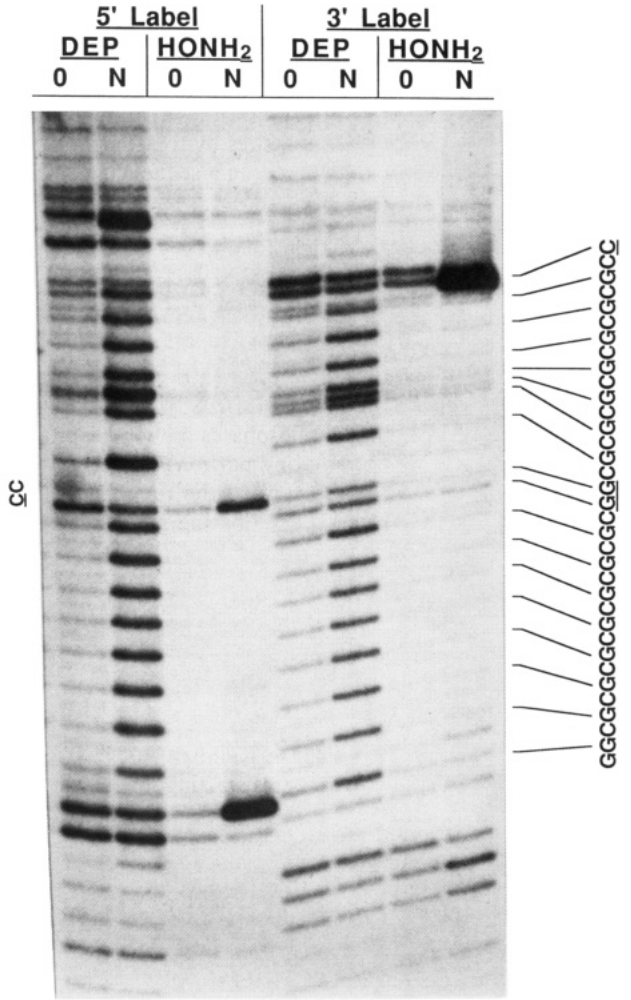

FIGURE 1: Reactivity to diethyl pyrocarbonate (DEP) and hydroxylamine (HA) of plasmid pZ-GG-Z when relaxed (0) or at native bacterial supercoiling (N), displayed on a 20% polyacrylamide sequencing gel containing 1.5% bisacrylamide and 8.3 M urea. The left-hand four lanes were 5′-end-labeled and the right-hand four lanes were 3′-labeled, all at the same *Taq*I site; hence, for both strands, bands of the same mobility correspond to approximately the same position in the sequence. The reactive C and the less reactive GG residues of the Z–Z junction sequence d(CC)·d(GG) are underlined; the hyperreactive cytosines at the B–Z junctions are the darkest bands at the ends of the CG insert in the supercoiled lanes. 0, relaxed; N, native bacterial superhelical density.

shown in Figure 1. The plasmid was either relaxed (0) or at native bacterial superhelical density (N; $\sigma \approx -0.055$) at the time of modification. End labeling was at either the 5′ or the 3′ position as shown, to display the reactivity of each strand. In both strands a marked hyperreactivity to diethyl pyrocarbonate of the CG insert is apparent at native supercoiling (second and sixth lanes), consistent with the appearance of Z-DNA as inferred from the two-dimensional gels (see below). In the middle of the CG insert are the two adjacent guanines where the phase of purine–pyrimidine alternation changes. In contrast to the guanines contained within the alternating stretches, these junction guanines exhibit only slight hyperreactivity (sixth lane, 3′ label, DEP, N). The complementary bases on the opposite strand are a pair of adjacent cytosines. The 5′ cytosine of this cytosine pair is strongly hyperreactive to hydroxylamine when supercoiled, whereas the cytosine next to it shows no reactivity (fourth lane, 5′ label, HA, N).

B–Z Junctions. Even more reactive is the first C of the sequence 5′-GGATCCGCG..., which forms one boundary of the CG insert (fourth lane, bottom). Previous work (Johnston & Rich, 1985) has shown that, in supercoiled pLP32 (the analogous plasmid without the discontinuity in the CG alternation), the same sequence is also hyperreactive to hydroxylamine, and this was identified as a B–Z junction. [pLP32 and pZ-GG-Z were run in adjacent lanes to confirm that the same cytosine is reactive in both plasmids (data not shown).] On the opposite strand, labeled at the 3′ end, there is a very prominent hyperreactivity to hydroxylamine in the same sequence at the opposite end of the insert, where the other B–Z junction is expected (last lane, top). Hyperractivity is seen only on one strand at each B–Z junction because there are no cytosines in the complementary strands at the junction sites. (However, faint reactivity can be seen at cytosines 2–3 bases away on either side of the B–Z junction in the last lane, suggesting some mobility at this junction.) The irregular spacing of the guanine bands above the Z–Z junction is commonly seen with alternating CG sequences and is attributed to the formation of secondary structure during electrophoresis. Interestingly, the pattern for pZ-GG-Z is markedly different from that for pLP32 [see Figures 4 and 5 in Johnston and Rich (1985)], despite the fact that there is only one base change between the two sequences. Apparently the secondary structures that can be formed by the two sequences are markedly different.

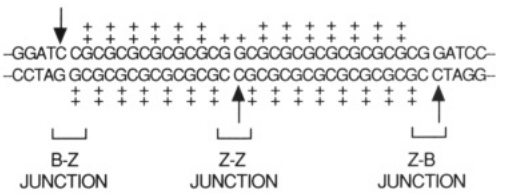

FIGURE 2: Summary of reactivity data. Hyperreactive bases under conditions of negative supercoiling are designated by + for DEP hyperreactivity and arrows for hydroxylamine hyperreactivity. B–Z and Z–Z junctions are as indicated.

Treatment of an unmodified sample with piperidine resulted in no visible bands at this exposure (not shown). The reason for the moderate supercoiling-independent hyperreactivity visible in all DEP lanes at C and T residues at the positions of the B–Z and Z–Z junctions is unknown; in other experiments these bands were much fainter or absent. In any case, the lack of dependence upon supercoiling means that reactivity at these sites is not related to the presence of Z-DNA or a Z–Z junction. The chemical reactivity data are summarized in Figure 2.

Unwinding Analysis

The conversion of a sequence from B-DNA to Z-DNA in a negatively supercoiled plasmid results in a decrease in the torsional strain, and hence the writhe, of the plasmid. If the transition takes place within the range of superhelical density in which topoisomers are resolved on agarose gels, there is a consequent change in the plasmid's electrophoretic mobility. In such a case, an analysis using two-dimensional gel electrophoresis can determine the extent of unwinding and the superhelical density required for the transition, and from those parameters the energy required for the transition can be derived (Peck & Wang, 1982). Results of 2-D gels on pZ-GG-Z and pZ-GGG-Z are shown in Figure 3. The left-hand series of topoisomers is pBR322, loaded in a separate well as a control. No unwinding transition is apparent in the range of topoisomers resolved by the gel for this plasmid. The right-hand series of topoisomers is a mixture of pZ-GG-Z and pZ-GGG-Z. The latter is similar to pZ-GG-Z except that the alternating CG sequence of the insert is interrupted by three consecutive guanines (top strand) instead of two as in pZ-

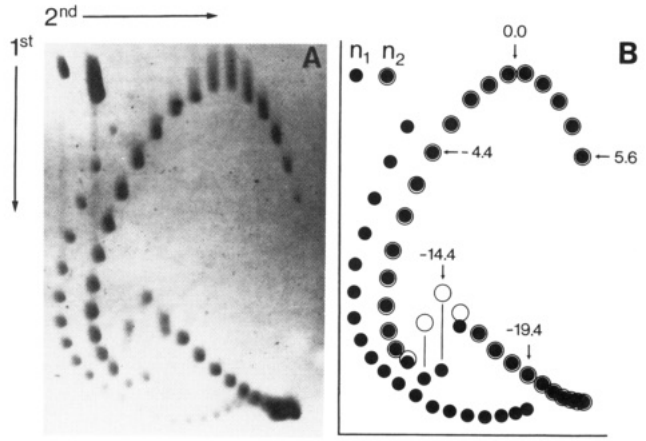

FIGURE 3: Analysis of the unwinding transition using two-dimensional gel electrophoresis. (A) Photographic negative of an ethidium bromide stained gel, with directions of electrophoresis as shown. (B) Interpretation; n_1 and n_2 are nicked circles. Three mixtures of topoisomers were loaded: pBR322 alone in one well (left-hand series of topoisomers and n_1), and a mixture of pZ-GG-Z (● in panel B) and pZ-GGG-Z topisomers (○) in a second well (right-hand series and n_2). Topoisomer numbers were assigned as shown, on the basis of the interpolated determination of the position of a_0 ("0.0"), the idealized, completely relaxed topoisomer.

GG-Z. Hence, there is no change of phase between the two alternating CG sequences and therefore no Z–Z junction.

It can be seen that topoisomers of pZ-GG-Z and pZ-GGG-Z are superimposed for values of linking difference that exceed −16.4 in magnitude. Their mobilities indicate that the entire insert has converted to Z-DNA and that therefore the structural perturbation at the Z–Z boundary does not greatly alter the left-handed character of the sequence. However, the midpoint of the B–Z transition for pZ-GG-Z occurs at a slightly higher superhelical density than that for pZ-GGG-Z, indicating that a Z–Z junction is energetically more unfavorable than a base pair out of alternation for sequences consisting solely of C and G. A quantitative analaysis (Ellison et al., 1985, 1986) of these results based on the statistical–mechanical model of Peck and Wang (1983) yields an unfavorable free energy of 3.5 kcal/mol per junction as the "cost" of this Z–Z junction, compared to 1.2 kcal for the single GC base pair disrupting the alteration of pZ-GGG-Z (Ellison et al., 1985). These numbers assume that the energetic cost of each B–Z junction is 5.0 kcal/mol per junction and the average cost of converting a CG base pair from the B form to the Z form is 0.33 kcal/mol per bp. Those values were derived with use of pLP32 (Peck & Wang, 1983), which is identical with pZ-GG-Z and pZ-GGG-Z except that it contains 32 bp of purely alternating CG.

Modeling of the Z–Z Junction

To evaluate the structural feasibility of a Z–Z junction, we undertook the computer-aided construction of a model for the Z–Z junction sequence of pZ-GG-Z. The design goal was to build a Z–Z junction in which the DNA on either side maintained the left-handed twist and other characteristics of Z-DNA as close as possible to the junction within the constraints of having a conformationally reasonable structure with an energy consistent with that calculated from the unwinding measurements.

Figure 4 is a stick drawing (in stereo) of the model. Both visual examination and energy analysis indicate that the residues one base pair removed from the junction are virtually unaffected by the junction and even the bases on either side of the junction stack in an unaltered manner on their next adjacent bases as seen in Figure 5a,c. The stacking at the junction itself is unique. The effective twist is about −24° compared to an average of −30° for Z-DNA (−60° for the two-base-pair repeat) and +36° for B-DNA. (Such a difference in twist cannot be distinguished by the 2-D gel analysis presented in Figure 3.) This stacking arrangement results in an increase in the accessibility of the hyperreactive C4 and/or C6 positions of Cyt3 relative to Cyt4 and a reduction in accessibility of the hyporeactive N7 positions of Gua7 and Gua8 compared to Gua2 and Gua5, in agreement with the results shown in Figure 1 (see Discussion).

It is difficult to compare the stacking energy of the Z–Z junction, shown in Figure 5b, to that of Z-DNA itself since Z–DNA has two distinct stacking arrangements: CG base pair over GC base pair (CpG stacking) and GC base pair over CG base pair (GpC stacking). The Z–Z junction consists of the CG base pair over a CG base pair (CpC/GpG stacking), which is not directly comparable to either stacking arrangement in Z-DNA. Using the relationships of the various energies in the standard B-DNA and Z-DNA models, we estimate the stacking and conformational energy cost of inserting a Z–Z junction to be about 4 kcal/mol. This compares favorably with the 3.5 kcal derived from the 2-D gel analysis (see above).

The pCpCp linkage is shown in Figure 6. In Z-DNA the ribose of the pCp linkage is directed opposite to the chain

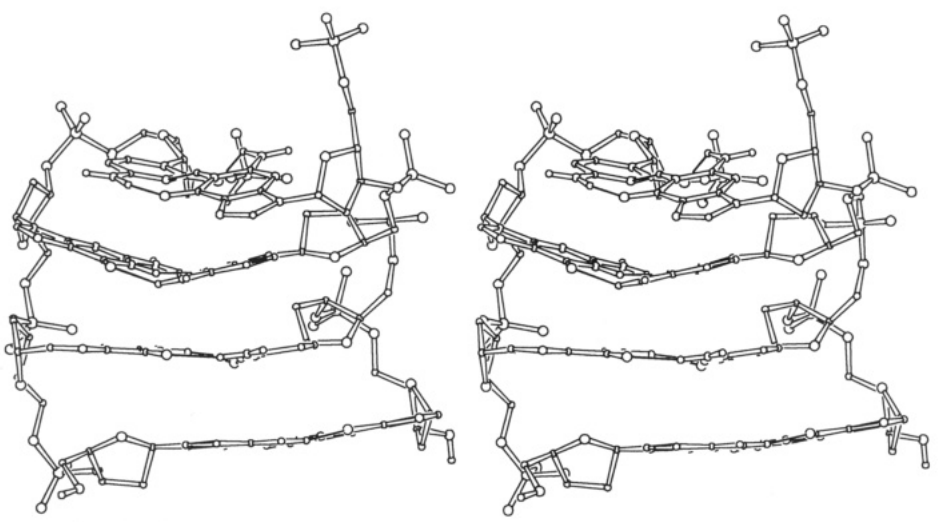

FIGURE 4: A stereo ORTEP drawing of the modeled Z–Z junction, d(pGpCpCpG)·d(CpGpGpCp) region.

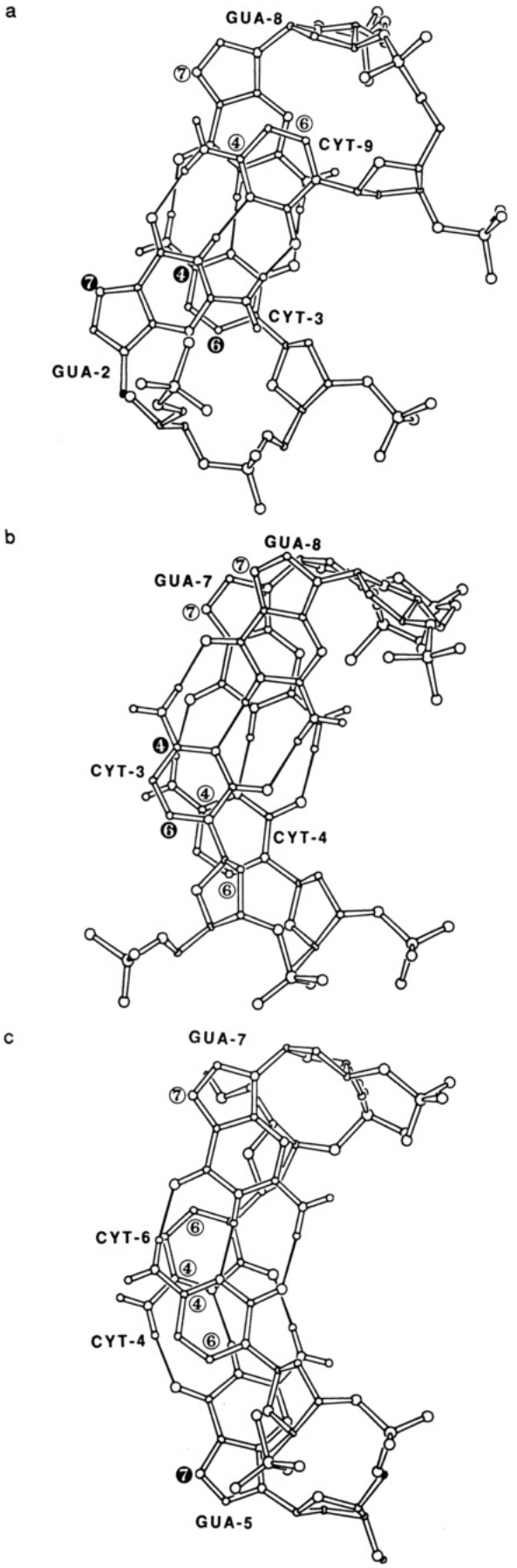

FIGURE 5: ORTEP drawings of the base stacking associated with the Z–Z junction. (a) The stacking of the adjacent GpC dinucleotide, G2·C9 over C3·G8. (b) The stacking of the Z–Z junction, C3·G8 over C4·G7. (c) The stacking of the adjacent CpG dinucleotide, C4·G7 over G5·C6.

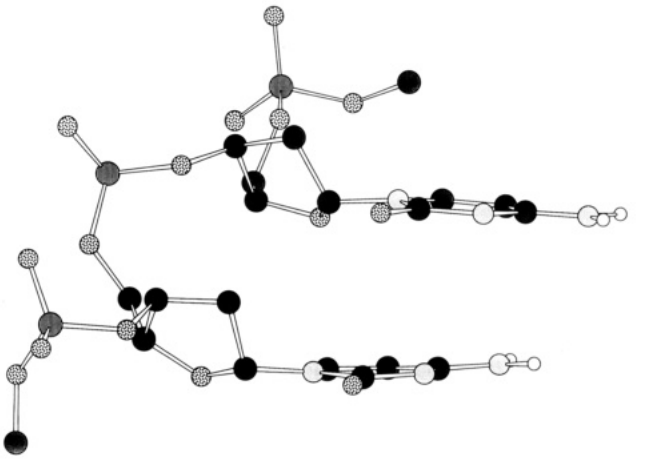

FIGURE 6: A ball-and-stick drawing of the pCpCp portion of the Z–Z junction.

advance, so that the two adjacent phosphate groups have to pass over the ribose to connect with the adjacent guanosines (Wang et al., 1979, 1981). This results in a taut backbone, which in part accounts for the apparent lack of flexibility of Z-DNA and the observation that Z-DNA does not easily permit intercalation. The pCpCp linkage at the Z–Z junction requires two consecutive crossovers, and we were concerned about whether a model would be excessively strained as a result. However, as can be seen, tilting of the ribose rings from their usual orientation parallel to the helix axis allows the linkage to form with relatively little distortion and with little expenditure of energy.

Discussion

The results of the chemical probing experiments indicate that the Z–Z junction is well-defined and highly localized. Only one of the two adjacent cytosines that mark the change in purine–pyrimidine alternation is hyperreactive to hydroxylamine when supercoiled, and no nearby cytosines on the opposite strand exhibit reactivity. Hydroxylamine reacts principally with cytosines that either are single-stranded (Rubin & Schmid, 1980; Johnston & Rich, 1985; Johnston, 1988a) or are in a double helix that is distorted in such a way as to allow out-of-plane nucleophilic attack at the C4 and/or C6 positions; for example, a B–Z junction (Johnston & Rich, 1985). In the case of a Z–Z junction, the type of distortion recognized by hydroxylamine is confined to the (GG)·(CC) sequence. Changes in DEP reactivity are also confined to that dinucleotide pair: all guanines within the two contiguous Z-DNA blocks are equally hyperreactive except for the two guanines within the (GG)·(CC) sequence, which are less reactive. Our proposed model is consistent with these results; it requires no substantial distortion of the Z structure beyond the two base pairs directly at the junction.

We previously reported the tentative identification of a Z–Z junction in highly supercoiled pLP32 at the position of the bar in the sequence 5′-GCGCG*G*ATCC*A*G*G*|AC*G*G*G*TGTG-3′, partly on the basis of a change in the phase of the alternating pattern of purine reactivity to DEP (to which underlined bases are hyperreactive) and dimethyl sulfate (DMS, to which bases in italics are hyperreactive). The dinucleotide (GA) at the putative Z–Z junction exhibited less hyperractivity to DEP when supercoiled than did the purines in the adjacent Z-DNA regions; this was also seen in pZ-GG-Z. On the opposite strand, the criterion of cytosine hyperreactivity was not applicable here, as the 5′ pyrimidine at the Z–Z junction was a thymine. However, the cytosine on the 3′ side of the junction

was not hyperreactive to hydroxylamine, as is the case in pZ-GG-Z. Interestingly, the above sequence from pLP32 can accommodate both a Z–Z junction and a number of bases out of alternation (including the DMS-reactive bases indicated above in italics) when the level of supercoiling is substantial.

The sequence $(GT)_6(CA)_6$ has been reported to show reactivity toward OsO_4 at the T of the 5′-TC-3′ dinucleotide when supercoiled (McClean & Wells, 1988). Both NH_2OH and OsO_4 react with ring carbons at the 4-, 5-, or 6-position of pyrimidines at B–Z junctions (Johnston & Rich, 1985), and the apparent accessibility of the 5′ pyrimidine at Z–Z junctions is a feature that both reagents appear to recognize as well. However, the OsO_4 reactivity at $(GT)_6(CA)_6$ was taken as evidence that the reactive T was in the syn conformation rather than the anti conformation as in our model. Although we cannot rule out that possibility, the syn conformation is unfavorable for pyrimidines (Haschemeyer & Rich, 1967), and in cases of out-of-alternation thymines, the syn conformation is not particularly reactive to OsO_4 (Johnston, 1988b).

The energetic analysis based on unwinding data from 2-D gels yields a value of 3.5 for the cost of a Z–Z junction in an otherwise alternating CG sequence, compared to the analogous completely alternating sequence. A similar analysis by Mirkin et al. (1987) of a slightly different sequence yielded a value for the energetic cost of a (CC)·(GG)-type Z–Z junction of 4 kcal. By way of comparison, the cost of the base pair out of alternation in pZ-GGG-Z is 2.4 kcal; the cost of an out of alternation AT base pair within an alternating CG sequence (CGCGAGCGCG) is 3.4 kcal (Ellison et al., 1985); and the cost of forming a B–Z junction in pLP32 (and presumably pZ-GG-Z) is 5.0 kcal. Thus, to form Z-DNA between two *Bam*HI sites, one must overcome a barrier of over 10 kcal, the cost of two B–Z junctions, plus the energy cost of converting the sequence from B to Z. If there is a phase change in the sequence, the extra expenditure of 3.5 kcal is rather modest by comparison.

Whereas 2-D gels can provide energetic data on conformational transitions that are difficult to obtain by other means, they can resolve only transitions that result in substantial unwinding (i.e., transitions involving long sequences) and that occur at relatively modest superhelical densities. Because of these limitations, for many natural sequences chemical reactivity appears to be the best diagnostic tool for the identification and precise localization of Z-DNA. This is especially the case with B–Z and Z–Z junctions, for which two-demensional gels can provide only inferential data.

The Z–Z junction model that we have presented is consistent with the distinct chemical reactivity of the two adjacent cytosine bases. In Figure 5 the reactive carbons (C4 and/or C6) of cytosines are indicated by circled numbers; white numbers within black circles are for the reactive atoms (whether C4 or C6 is uncertain), and black numbers within white circles are for the unreactive atoms. A careful examination of these atoms in the three steps around the junction (Figures 5, parts a, b, and c) reveals that only C6 of the reactive Cyt3 is relatively accessible to out-of-plane attack, whereas in the unreactive Cyt4 both C4 and C6 are blocked by atoms above and below. In addition, the N7 positions that react with DEP are somewhat more protected in the two guanine residues at the junction (Gua7 and Gua8) than at other guanines (Gua2 and Gua5), again consistent with the observed reactivity.

However, consistency does not prove the model, and the question remains as to whether and/or to what degree the model is correct. The first consideration is whether other conformationally distinct models could have been constructed. Examination of both the initial model building and the final model reveal that for the CpC side the structure is conformationally quite limited and somewhat strained as is. The GpG linkage has rather more flexibility, but it appears unlikely that any other conformation would provide nearly as good a stacking arrangement, subject to the constraints of the CpC linkage.

The second consideration is whether the model obtained is the best obtainable subject to the starting model. In building the starting model, some attention was paid to the possibilities of interactions that would trap the structure in a particular and unfavorable local minimum. Similar checks were made as the phosphate groups were being idealized and also at the final structure. As a result, we believe the structure we obtained is likely to represent the minimum energy state for the starting model.

It is possible that the starting model was inadequate, that the true structure is significantly different from that proposed, and that the Z-DNA structure is more perturbed beyond the immediate Z–Z junction. However, any such longer range perturbations would have to be subtle enough that they would neither affect reactivity to DEP and hydroxylamine nor greatly alter the left-handed character of the Z-DNA as determined from the 2-D gel analysis. Moreover, the ability of the canonical Z-DNA structure to accommodate with relatively little perturbation a wide range of sequences, including some having bases out of alternation (Wang et al., 1985; Feigon et al., 1985; Johnston & Rich, 1985; Johnston, 1988b), attests to the robustness of the structure and argues in favor of the proposed model.

The model has been obtained by use of an energy-minimization procedure containing a number of substantial approximations. In particular, the whole counterion–solvent environment is approximated by the use of large-radius, free-floating unit charge counterions and an assumed distance-dependent dielectric constant of $4R_{ij}$ (R_{ij} being the distance between atoms i and j). The true structure with real water molecules and counterions is likely to exhibit small variations from that proposed, much as is observed in B- and Z-DNA between the crystal structures and the energy-minimized structures.

The model-building described here demonstrates that a Z–Z junction with reasonable geometry and minimum perturbation to the adjacent Z regions can be constructed at a modest energy cost and with substantial unwinding. If a different structure is correct, it must be even more energetically economical.

If regions of Z-DNA exist and play a role in functioning biological systems, then the range of sequences in which Z-DNA-like structures can be formed is substantially increased by the ability of DNA to accommodate a Z–Z junction without substantial energy cost. A key point is that these structures are "Z-DNA-like" in that they provide on the order of 60–65° of unwinding of the DNA per base pair, just as do regular Z-DNA or Z-DNA with out-of-alternation base pairs. On the other hand, the Z–Z junction structure is distinct in several ways. The first and most obvious is that the d(CpC)·d(GpG) step of our sequence is clearly distinct from either the CpG or GpC steps that constitute the adjacent alternating regions of Z-DNA, so that a protein that recognized either or both types of Z-DNA dinucleotide steps would be unlikely to accept a Z–Z junction and vice versa. More broadly, a Z-DNA binding protein could distinguish an extended Z-DNA segment or two closely separated Z-DNA segments from segments containing a Z–Z junction since in Z-DNA the repeat is two

residues and a Z-Z junction would create a relative displacement of around 3.7 Å along the axis and 30° around the axis (a half repeat for each).

It is interesting to note that a protein could recognize a sequence forming a Z-Z junction by simply sensing alterations in the backbone geometry. By contrast, sequence recognition for B-DNA must in general involve direct contact with the bases since calculations indicate that the sequence dependence of the backbone is at a low enough energy level to be unreliable for recognition.

References

Arnott, S., Campbell-Smith, P. J., & Chandrasekaran, R. (1976) in *Handbook of Biochemistry and Molecular Biology, Nucleic Acids* (Fasman, G. D., Ed.) Vol. 2, pp 411–422, CRC Press, Cleveland, OH.

Crawford, J. L., Kolpak, F. J., Wang, A. H.-J., Quigley, G. J., van der Marel, G., Rich, A., & van Boom, J. H. (1980) *Proc. Natl. Acad. Sci. U.S.A. 77*, 4016–4020.

Ellison, M. J., Kelleher, R. J., Wang, A. H.-J., Habener, J. F., & Rich, A. (1985) *Proc. Natl. Acad. Sci. U.S.A. 82*, 8320-8324.

Ellison, M. J., Feigon, J., Kelleher, R. J., Wang, A. H.-J., Habener, J. F., & Rich, A. (1986) *Biochemistry 25*, 3648-3655.

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

Fujii, S., Wang, A. H.-J., Quigley, G. J., Westerink, H., van der Marel, G., van Boom, J. H., & Rich, A. (1985) *Biopolymers 24*, 243–250.

Hamada, H., Petrino, M. G., & Kakunaga, T. (1982) *Proc. Naatl. Acad. Sci. U.S.A. 79*, 6465–6469.

Haniford, D. B., & Pulleyblank, D. E. (1983) *J. Biomol. Struct. Dyn. 1*, 593–609.

Haschemeyer, A. E. V., & Rich, A. (1967) *J. Mol. Biol. 27*, 369–384.

Herr, W. (1985) *Proc. Natl. Acad. Sci. U.S.A. 82*, 8009–8013.

Ho, P. S., Frederick, C. A., Quigley, G. J., van der Marel, G. A., van Boom, J. H., Wang, A. H.-J., & Rich, A. (1985) *EMBO J. 4*, 3617–3624.

Jawarsky, A. J., Hsieh, W.-T., Blaho, J. A., Larson, J. E., & Wells, R. D. (1987) *Science 774*, 773–777.

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

Johnston, B. H. (1988b) *J. Biomol. Struct. Dyn. 6*, 153–166.

Johnston, B. H., & Rich, A. (1985) *Cell 42*, 713–724.

Kollman, P., Weiner, P., Quigley, G., & Wang, A. (1982) *Biopolymers 21*, 1945–1970.

Kozelka, J., Petski, G. A., Lippard, S. J., & Quigley, G. J. (1985) *J. Am. Chem. Soc. 107*, 4079–4081.

Kozelka, J., Archer, S., Petsko, G. A., Lippard, S. J., & Quigley, G. J. (1987) *Biopolymers 26*, 1245–1272.

Maxam, A. M., & Gilbert, W. (1980) *Methods Enzymol. 65*, 499–560.

McClean, J. J., & Wells, R. W. (1988) *J. Biol. Chem. 263*, 7370–7377.

McClean, M. J., Lee, J. W., & Wells, R. W. (1988) *J. Biol. Chem. 263*, 7378–7385.

Mirkin, S. M., Lyamichev, V. I., Kumarev, V. P., Kobzev, V. F., Nosikov, V. V., & Vologodskii, A. V. (1987) *J. Biomol. Struct. Dyn. 5*, 79–88.

Peck, L., & Wang, J. (1982) *Proc. Natl. Acad. Sci. U.S.A. 79*, 4560–4564.

Peck, L., & Wang, J. (1983) *Proc. Natl. Acad. Sci. U.S.A. 80*, 6206–6210.

Rahmouni, A. R., & Wells, R. D. (1989) *Science 246*, 358–363.

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

Rubin, C. M., & Schmid, C. W. (1980) *Nucleic Acids Res. 8*, 4613–4619.

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

Wang, A. H.-J., Quigley, G. J., Kolpak, F. J., van der Marel, G., Rich, A., van Boom, J. H. (1981) *Science 211*, 171–176.

Wang, A. H.-J., Gessner, R. A., van der Marel, G. A., van Boom, F. H., & Rich, A. (1985) *Proc. Natl. Acad. Sci. U.S.A. 82*, 3611.

Weiner, P., & Kollman, P. (1981) *J. Comput. Chem. 2*, 287–303.

Wittig, B., Dorbic, T., & Rich, A. (1989) *J. Cell Biol. 108*, 755–764.