Articles

Structure and Thermodynamics of Nonalternating C·G Base Pairs in Z-DNA: The 1.3-Å Crystal Structure of the Asymmetric Hexanucleotide d(m^{5}CGGGm^{5}CG)·d(m^{5}CGCCm^{5}CG)[†,‡]

Gary P. Schroth, Todd F. Kagawa, and P. Shing Ho*

Department of Biochemistry and Biophysics, 2011 Agriculture and Life Sciences Building, Oregon State University, Corvallis, Oregon 97331

Received July 6, 1993; Revised Manuscript Received September 21, 1993[⊗]

ABSTRACT: We have solved the single-crystal X-ray structure of the complementary hexanucleotides d(m^5-CGGGm^{5}CG) and d(m^{5}CGCCm^{5}CG). The hexamer duplex was crystallized as Z-DNA, but contains a single C·G base pair that does not follow the alternating pyrimidine/purine rule for Z-DNA formation. This is the first crystal structure which serves to illustrate the structural consequences of placing a cytosine in the sterically disfavored *syn* conformation. In addition, since these sequences are not self-complementary, the individual strands of this asymmetric hexamer are unique in sequence and therefore distinguishable in the crystal lattice. Nevertheless, the crystal of this duplex is isomorphous with other Z-DNA hexamer structures. The asymmetry of this hexamer sequence required that the structure be solved using two unique models, which are distinguished by the orientation of hexanucleotides in the crystal lattice. In one model (the GG model) the cytosine in the *syn* conformation is packed against the terminal guanine base of a symmetry-related hexamer, while in the alternative model (the CC model) this cytosine sits exposed in a solvent channel of the lattice. We find that neither model alone can completely account for the observed electron densities. The two models ultimately were refined together. A composite structure consisting of 65% GG model and 35% CC model refined to an *R*-factor of 19.3%, which was significantly lower than refinements using either model alone. A detailed analysis of these two structures shows that, in spite of the out-of-alternation C·G base pair, the features characteristic of Z-DNA have been maintained. Both models, however, show significant local structural adjustments to accommodate the single cytosine base which is forced to adopt the *syn* conformation in each hexamer. In general, it appears that in order to relieve the energetically unfavorable steric contacts between the cytosine base in the *syn* conformation and the deoxyribose sugar, the base is forced into a highly buckled conformation, and that this large buckle in turn alters the conformation of neighboring residues. This unusual conformation also significantly weakens base-stacking interactions between the cytosine in *syn* and the adjacent residues in the helix and affects the exposure of the bases to solvent. We conclude that this crystal structure provides a molecular rationale for why nonalternating bases are energetically disfavored in Z-DNA.

The inversion of canonical right-handed B-DNA to the higher energy, left-handed Z-DNA conformation is one of the more dramatic structural transitions in biology. Currently Z-DNA is undoubtedly the best understood of all non-B-DNA conformations, due in large part to the many high-resolution molecular structures obtained from X-ray diffraction studies on single crystals of synthetic oligonucleotides (Rich et al., 1984; Jovin et al., 1987). Single-crystal X-ray diffraction studies of alternating pyrimidine/purine (APP)[1] oligonucleotides have defined Z-DNA as a left-handed DNA helix, containing standard Watson–Crick base pairs, with a helical twist of –12 bp/turn (Wang et al., 1979; Gessner et al., 1989). The general structure of Z-DNA is further characterized by alternating conformations of the pyrimidine and purine nucleotides in which (i) the pyrimidine nucleotides adopt the *anti* glycosidic conformation, usually with a C2′-*endo* sugar pucker, and (ii) the purine nucleotides adopt the *syn* glycosidic conformation, usually with a C3′-*endo* sugar pucker (Wang et al., 1979; Gessner et al., 1989). Theoretical analyses have suggested that pyrimidine nucleotides have a preference for the *anti* conformation, whereas for purines both the *anti* and *syn* conformations are energetically equivalent (Haschmeyer & Rich, 1967). For these reasons it is thought that the alternating *anti*/*syn* nucleotide conformations, coupled with the alternating C3′-*endo* / C2′-*endo* ribose sugar puckers, explain the strong preference for Z-DNA to form in alternating pyrimidine/purine sequences.

The standard Z-DNA conformation is based mainly on the crystal structure of the hexamer sequence d(CGCGCG), which when crystallized in 1979 first demonstrated the propensity for Z-DNA to form readily in APP sequences (Wang et al., 1979). The molecular details of other sequences of Z-DNA

[†] This study was supported by the American Cancer Society (Grant NP-740). G.P.S. is supported by an American Cancer Society Postdoctoral Fellowship, and P.S.H. is a recipient of a Junior Faculty Research Development Award from the American Cancer Society.

[‡] Crystallographic coordinates and structure factors have been submitted to the Brookhaven Protein Data Base under File Names 145D (coordinate data) and 145DSF (structure factors) and the Nucleic Acids Data Base under File Name ZDFB37 (coordinate data).

[⊗] Abstract published in *Advance ACS Abstracts,* November 15, 1993.

[1] Abbreviations: APP, alternating pyrimidine/purine; MPD, 2-methyl-2,4-pentanediol.

0006-2960/93/0432-13381$04.00/0

have been studied from the very high resolution (typically approaching 1.0 Å) structures of derivative hexanucleotides. With a single exception, all of the previously published crystal structure studies of Z-DNA have been limited to strictly alternating pyrimidine/purine sequences, usually variations of the original d(CGCGCG) sequence [summarized in Dickerson (1992)]. The structure of the hexanucleotide d(m^{5}CGATm^{5}CG) serves as the only example of an oligonucleotide crystal structure which places pyrimidines in the *syn* conformation (Wang et al., 1985). However, in all of the Z-DNA crystal structures previously solved, guanines have only been observed in the *syn* conformation, with cytosines adopting only the *anti* conformation. Thus the accumulated crystallographic data says little about the propensity of a large majority of other non-APP DNA sequences to adopt the left-handed conformation.

The energetics for Z-DNA formation for a broader range of sequences have been derived from electrophoretic studies of supercoil-induced B- to Z-DNA transitions (Peck & Wang, 1983). These studies have shown that negative supercoiling is absolutely required to facilitate a B–Z transition within APP sequences in topologically constrained DNA molecules and, additionally, that there is a clear hierarchy in the ability of various APP dinucleotide pairs to form Z-DNA: d(CG) > d(CA)-d(TG) > d(TA) in terms of stability as Z-DNA [summarized in Ho et al. (1986)]. The energetics of the B–Z transition can also be altered by modifying substituents of the DNA bases; for instance, methylation of cytosines at the C5 position dramatically lowers the B–Z transition energy of the d(CG) dinucleotide (Zacharias et al., 1988). Variations on the APP rules for Z-DNA formation, in which some of the bases are "out-of-alternation", have also been studied electrophoretically, giving results which have shown that Z-DNA can accommodate both C·G and A·T base pairs which place pyrimidines in the unfavorable *syn* conformation. Ellison et al. (1985, 1986) quantitated the B–Z transition energy of the d(GG)·d(CC) dinucleotide [i.e., a single G to C transversion within a d(CG) dinucleotide] and found it to be equal to the transition energy of an "in-alternation" d(TA) dinucleotide. Furthermore, Wells and his colleagues have shown that Z-DNA can propagate across up to four consecutive thymine (or adenine) nucleotides, provided that they are flanked by stretches of alternating d(CG) sequences (McLean et al., 1986). This growing body of work on the ability of non-APP sequences to form Z-DNA has been important in expanding the possibilities for the types of sequences which can form Z-DNA under the stress of negative supercoiling. We have recently shown that potential Z-DNA-forming regions in naturally occurring genomic sequences are not limited only to strictly APP sequences, and in fact many of the highest potential Z-DNA-forming sequences we identified in the human genome contain out-of-alternation bases (Schroth et al., 1992).

In this paper we describe the crystal structure of the non-APP hexanucleotide d(m^{5}CGGGm^{5}CG) with its complementary strand d(m^{5}CGCCm^{5}CG). This structure represents the first oligonucleotide crystal structure in which a cytosine nucleotide is shown to adopt the unusual *syn* conformation. In solving the structure of this non-self-complementary oligonucleotide, we have had to overcome problems associated with the crystallographic uniqueness of each strand in this hexamer (the two strands of the DNA duplex are not identical). Thus two separate models were derived for the hexanucleotide which account for the unique positions of each strand within the structure. The features that were found to be common between the two models could thereby reasonably be assigned to the intrinsic properties of the out-of-alternation C·G base pairs in Z-DNA. Other features appear to be associated either with crystal packing effects or with refinement differences related to the occupancies of the two competing models. When taken into account, these effects allow us to deconvolute the structural features that are intrinsic to out-of-alternation base pairs in Z-DNA from local features which are associated with the different packing of each of the DNA duplexes into the crystal lattice. Furthermore, the solution conditions under which this sequence crystallizes, coupled with thermodynamic analysis of the molecular structure of this hexanucleotide, provides some additional insight into the ability of non-APP sequences to adopt the Z-DNA conformation.

EXPERIMENTAL PROCEDURES

The two complementary hexanucleotides used in these studies, d(m^{5}CGGGm^{5}CG) and d(m^{5}CGCCm^{5}CG), were synthesized on an Applied Biosystems DNA synthesizer in the Central Services Laboratory of the Center for Gene Research and Biotechnology at Oregon State University. Byproducts of the oligonucleotide syntheses were removed by Sephadex G-10 column chromatography. Judging from the coupling efficiencies of the synthesis, the purity of the oligonucleotides was estimated to be better than 95%. Equimolar ratios of the two oligonucleotides were mixed and crystallized at room temperature using the vapor diffusion method, equilibrating against 10% 2-methyl-2,4-pentanediol (MPD) as the precipitating agent in the reservoir. The solution conditions for crystallizing this particular hexanucleotide as Z-DNA were predicted using a variation on the semiempirical method of Ho et al. (1991). For this crystallization we take into account the 2.4 kcal/mol energetic penalty associated with a single G to C transversion in Z-DNA as determined by Ellison et al. (1985) (for more details, see the Results section). The initial conditions used for crystallizing this sequence were 0.8 mM hexanucleotide (single strands); 14 mM sodium cacodylate, pH 7.0; 1.4% MPD; and magnesium chloride concentrations ranging from 4 to 32 mM. Additional crystallization setups were performed using the same initial conditions, with the addition of 2.8 mM spermidine to each sample. Within 10 days, crystals formed under almost all of the described conditions. However, the highest quality crystals were obtained when the initial magnesium chloride concentration was 21 mM and the initial spermidine concentration was 2.8 mM.

X-ray diffraction data was collected at room temperature from a single hexagonal rod-shaped crystal, with approximate dimensions of 0.3 × 0.3 × 0.6 mm. Data for this crystal was collected on a Siemens rotating-anode diffractometer to a resolution of better than 1.3 Å. The crystal was determined to be in the space group $P2_12_12_1$, typical of the lattice of all other Z-DNA hexamer crystals, with unit-cell dimensions of a = 17.865 Å, b = 30.822 Å, and c = 44.797 Å.

The structure of the hexanucleotide in this crystal was determined by molecular replacement. The initial models for the refinement were built using the Insight II software package from BioSym, Inc., by modifying the 1.3-Å-resolution Z-DNA structure of d(m^{5}CGm^{5}CGm^{5}CG) (Fujii et al., 1982) as follows: (i) the C5 methyl groups at the cytosines in positions 3 and 9 in the hexamer duplex were removed, and (ii) the two internal d(C·G) base pairs (3 and 9; 4 and 10) were arranged such that each strand had either two consecutive cytosines or two consecutive guanines. Because the hexanucleotides used in these experiments are not self-complementary, and therefore

Table I: Refinement Data for d(m^{5}CGGGm^{5}CG)·d(m^{5}CGCCm^{5}CG)

resolution (Å)	8.0–1.3
reflections used, $F_o > 1.5\sigma(F_o)$	5029
solvent molecules	48
R-values	
combined GG/CC models	19.3%
GG model only	24.3%
CC model only	27.0%

resolution range (Å)	no. of reflections	% completeness	*R*-value[a] shell	*R*-value[a] accumulative
8.0–2.57	847	97.6	14.95	14.95
2.57–2.05	778	94.2	18.31	16.07
2.05–1.80	728	89.8	19.90	16.82
1.80–1.64	686	86.4	18.97	17.11
1.64–1.52	629	77.4	23.03	17.63
1.52–1.43	483	62.4	26.95	18.16
1.43–1.36	458	58.0	30.89	18.75
1.36–1.30	420	53.1	32.87	19.31

[a] The *R*-values given are for the combined GG and CC models, refined at an occupancy of 65% and 35%, respectively.

are not crystallographically identical, there are potentially two crystal packing configurations that the hexamer can assume in the $P2_12_12_1$ space group. We therefore used two different models when solving this structure. The two models can be distinguished from each other by the numbering of the nucleotide residues of each strand in the hexamer duplex. The nucleotide residues in a hexamer duplex are numbered C1 through G6 in the 5′ to 3′ direction on one strand and C7 through G12 in the 5′ to 3′ direction on the complementary strand. In one model, an out-of-alternation G·C base pair is placed at the 3 and 10 positions (the "GG model"), while the alternative model (the "CC model") places an out-of-alternation C·G base pair the the 4 and 9 positions. The key features of the two models are described fully in the Results section.

Models for the hexamer duplex in the two packing orientations were first refined separately using both simulated annealing and standard conventional positional methods of the X-PLOR refinement package (Brunger, 1992). The data used in these refinements fall within the resolution range of 8.0–1.3 Å and include 6973 total reflections, of which 5029 were observed above $1.5\sigma(F)$. The separately refined GG and CC model structures contained 48 solvent molecules each and had *R*-factors of 24.3% and 27.0%, respectively. The detailed information about the refinement, including the percent completeness of the data and the *R*-factor at each resolution shell, is given in Table I. To determine the extent that the molecular mechanics calculations used in X-PLOR induce their own perturbations to the conformation of the DNA, the structures were also refined using the Hendrickson–Konnert least-squares refinement method (Hendrickson & Konnert, 1979).

$2(F_o - F_c)$ difference maps were separately calculated for the GG and CC models (shown in Figure 2). The sections of this difference map for the 3·10 and 4·9 base pairs show that there was significant residual electron density at each of these asymmetric base-pair steps. This indicates that neither model alone could fully fit the diffraction data, but that the best model would likely be a composite structure in which the hexanucleotide is packed in both orientations (see Results for a full description).

The X-PLOR package was also used to estimate the contribution of each packing model to this compostie structure. For this "occupancy" refinement, the two separately refined hexamer duplex models were combined into a single model, and the occupancies of the two models were optimized. Each model was defined as one of two possible alternate segments of the combined model. The occupancies of the model were refined using only the positions N1, C2, C6, and O6 base atoms of the guanine residues and the N4 base atom of the cytosine residues for the 4·9 and 3·10 base pairs of each model. These atoms represent the atomic positions which are of greatest difference between the models of the two packing orientations, as observed in the $2(F_o - F_c)$ difference maps (Figure 2), and were therefore the best criteria for distinguishing between the two models. The refined occupancies of these atoms were assigned to the remaining atoms of both models, and simulated annealing and conventional positional refinement routines were continued. The final occupancy-refined model contained 65% contribution from the GG structure, with a 35% contribution from the CC structure (explained further in Results). The final occupancy-refined structure contained 48 solvent molecules and refined to an *R*-factor of 19.3%.

Conformation angles and base-pair morphologies for the refined GG and CC structures were kindly calculated for us by Marla S. Babcock of Rutgers University (Babcock & Olson, 1993). Additional conformational calculations were also performed on 10 isomorphous APP Z-DNA structures whose coordinates were deposited in the Nucleic Acids Database [NDB; see Berman et al. (1992)]. The results of these calculations on regularly alternating pyrimidine/purine structures were used to generate average values for Z-DNA base-pair morphology in other crystallized hexamers (as shown in Table III). Calculations of van der Waals energies were performed using the Insight II software from BioSym, Inc. The conformations of model structures were optimized to local energy minima using the AMBER force field of the DISCOVER molecular simulation program from BioSym, Inc. Hydration free energies were calculated as in our previous work (Kagawa et al., 1989, 1993; Ho et al., 1991). The atomic coordinates and structure factors have been deposited in the Brookhaven Crystallographic Data Base and the Nucleic Acids Data Base.

RESULTS

Crystallization of d(m^{5}CGGGm^{5}CG)·d(m^{5}CGCCm^{5}CG). We have previously shown that there is a strong relationship between the inherent stability of a hexanucleotide sequence as Z-DNA and the cationic conditions required to form diffraction-quality single crystals of that sequence. Using this relationship, a method was developed to predict the crystallization conditions of hexamer duplexes as Z-DNA (Ho et al., 1991), based on the ability of solvation free energy (SFE) calculations on sequences as B- and Z-DNA to predict the relative stabilities of the two conformations (Kagawa et al., 1989). A correlation was observed between the ability of alternating pyrimidine/purine hexanucleotide sequences to adopt the Z-conformation and the cation concentrations [as defined by the cationic strength (CS) in the crystallization setups] reported to yield diffraction-quality crystals of these sequences as Z-DNA (Ho et al., 1991). A further correlation was found between the hydration free energies of Z-DNA-forming hexamers and the experimentally determined supercoil-induced B–Z transition energies (Ho et al., 1991). Thus, a semiempirical relationship exists between theoretical solvation free energy calculations, experimentally determined B–Z transition energies, and the cationic strength required to crystallize a hexanucleotide as Z-DNA (Ho et al., 1991;

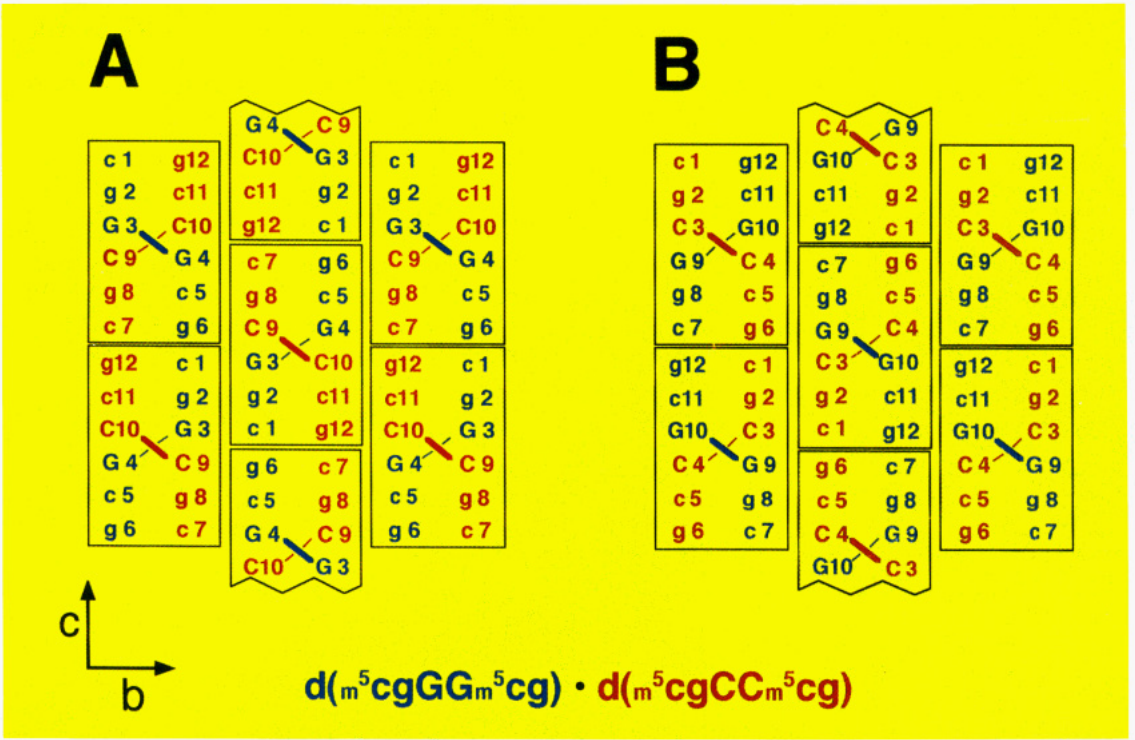

FIGURE 1: Crystal-packing diagrams of the two possible orientations available to the hexamer duplex in the $P2_12_12_1$ space group. The asymmetric unit in crystals of hexamers of Z-DNA in this space group is a single hexamer duplex. Each hexamer duplex, each of which is highlighted with a box, defines one half-turn of Z-DNA and comprises two strands of non-self-complementary hexanucleotides of DNA. The duplexes stack with their helical axes aligned parallel to the crystallographic *c*-axis. Adjacent stacked duplexes are related by a 2-fold screw transformation along the helical axis, generating essentially continuous helices of Z-DNA in the crystal lattice. The hexamers are additionally related by a second 2-fold screw transformation in the plane of the C5·G8 base pair, parallel to the crystallographic *b*-axis. The nucleotide residues in a hexamer duplex are numbered C1 through G6 in the 5′ to 3′ direction on one strand and C7 through G12 in the 5′ to 3′ direction of the complementary strand. In both packing diagrams, the d(m^{5}CGGGm^{5}CG) hexanucleotides are colored blue and the d(m^{5}CGCCm^{5}CG) hexanucleotides are colored red. Panels A and B are views down the crystallographic *a*-axis and show the packing diagrams for the GG model and the CC model, respectively. In the GG model shown in panel A, the d(m^{5}CGGGm^{5}CG) hexanucleotide (blue strand) is assigned residue numbers 1–6 and the d(m^{5}CGCCm^{5}CG) hexanucleotide (red strand) is assigned residue numbers 7–12. In the CC model shown in panel B, the d(m^{5}CGCCm^{5}CG) hexanucleotide is assigned residue numbers 1–6 and the d(m^{5}CGGGm^{5}CG) hexanucleotide is assigned residue numbers 7–12. The two models differ in their residue assignments for each strand in the asymmetric unit, and therefore the out-of-alternation base pair in the GG model occurs at the G3·C10 base pair versus the C4·G9 base pair in the CC model.

Kagawa et al., 1993). These relationships have been used previously by us to predict the crystallization conditions of the novel hexanucleotides d(m^{5}CGUAm^{5}CG) and d(CICGCG) as Z-DNA (Kagawa et al., 1989; Zhou & Ho, 1990; Ho et al., 1991).

In this paper, we have further extended these correlations to the crystallization of d(m^{5}CGGGm^{5}CG)·d(m^{5}CGCCm^{5}CG), which contains a single non-APP base in each strand (previously we have only dealt with strictly APP sequences). We indeed do find that these relationships apply also for non-APP sequences, at least in the case of a single out-of-alternation C·G base pair. According to the studies of Ellison et al. (1985, 1986), the dinucleotides d(TA) and d(GG)·d(CC) have identical B–Z transition energies of 2.4 kcal/mol of dinucleotide. The similarity between the B- to Z-DNA transition free energies of the out-of-alternation d(CC)·d(GG) and the in-alternation d(TA) dinucleotides thus predicts that the CS required to crystallize the sequence d(m^{5}CGGGm^{5}CG)·d(m^{5}CGCCm^{5}CG) should be comparable to that required previously to crystallize the similar sequence d(m^{5}CGTAm^{5}CG). We found in this work that a CS of 0.70 M is required for the crystallization of the d(m^{5}CGGGm^{5}CG)·d(m^{5}CGCCm^{5}CG) hexamer duplex, while Wang and co-workers (1984) needed a CS of 0.56 M for the crystallization of the d(m^{5}CGTAm^{5}CG) hexamer. These cationic strengths represent very similar final conditions for forming diffraction-quality Z-DNA crystals, even though these two hexanucleotides were grown from different starting conditions and in different laboratories. This suggests that our method for predicting crystallization conditions may be applicable to other non-APP oligonucleotides as Z-DNA.

Packing and Structure of the Asymmetric Hexamer Duplex in the Crystal Lattice. The asymmetric hexamer duplex d(m^{5}CGGGm^{5}CG)·d(m^{5}CGCCm^{5}CG) crystallized as Z-DNA in the $P2_12_12_1$ space group and is therefore isomorphous with previously crystallized self-complementary hexamers of Z-DNA (Wang et al., 1979; Gessner et al., 1989; Dickerson, 1992). The asymmetric unit in crystals of Z-DNA hexamers is a single hexamer duplex which defines one half-turn of Z-DNA. In the crystal, hexamer duplexes are stacked end-to-end along the crystallographic *c*-axis to form essentially continuous helices of Z-DNA. Unlike self-complementary hexamer duplexes, the individual strands of the current asymmetric hexamer duplex are unique in sequence and therefore distinguishable in the crystal structure. It is therefore possible to differentiate the two potential packing orientations available to the hexamer duplex in the crystal lattice. The crystal of our asymmetric sequence in fact can be thought of as composed of two potential orientations of the hexanucleotides in the lattice. This can best be represented as two distinct models for the asymmetric unit. To completely describe the structure of this crystal, we must therefore determine the contribution of each model to the overall lattice packing.

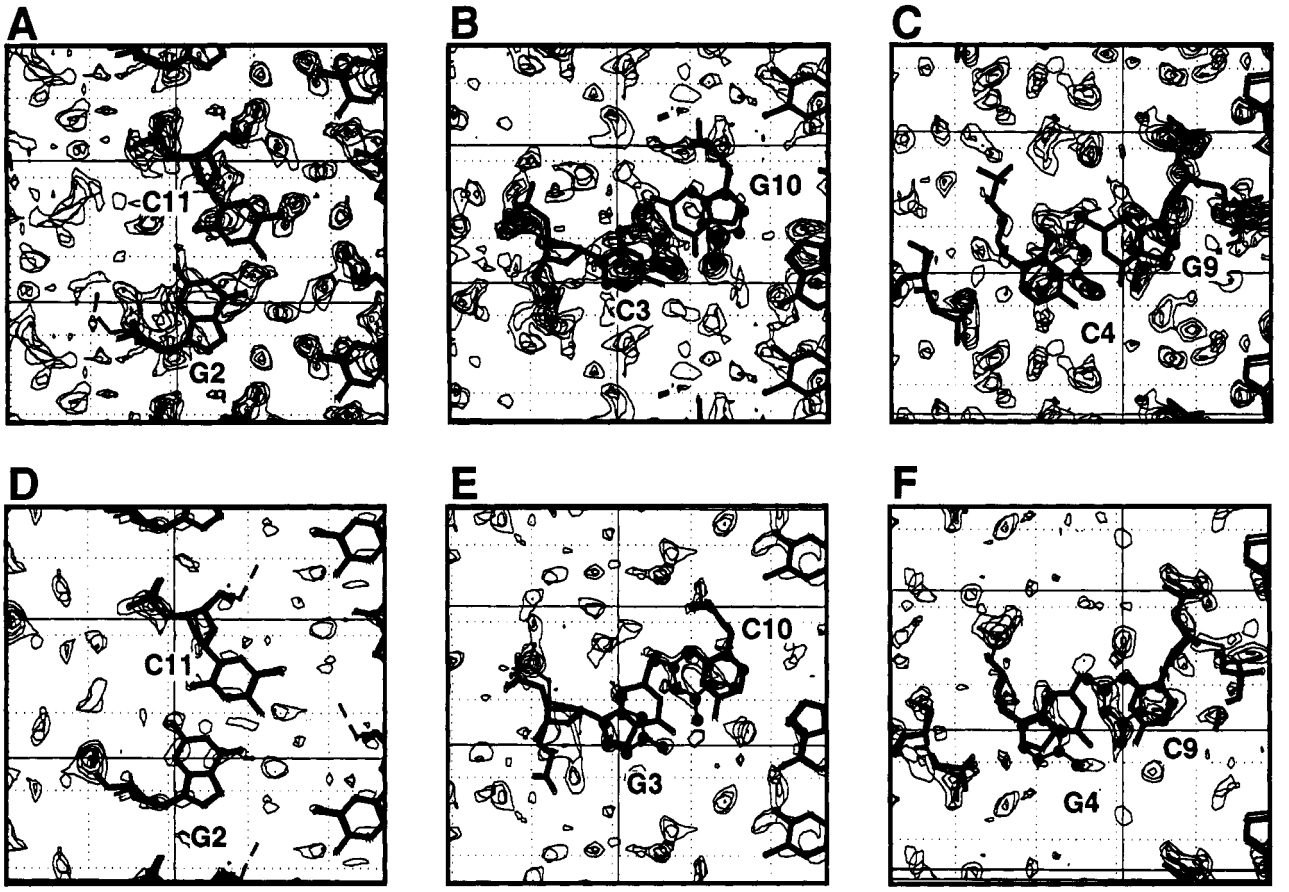

FIGURE 2: $2(F_o - F_c)$ difference electron density map sections comparing selected base pairs of the GG and CC models. Panels A, B, and C show electron density map sections of the G2·m^{5}C11, C3·G10, and C4·G9 base pairs, respectively of the CC model. Panels D, E, and F show electron density map sections of the G2·m^{5}C11, G3·C10, and G4·C9 base pairs, respectively, of the GG model. Each electron density map section represents a 2.5-Å-thick slice taken along the crystallographic *c*-axis, which is parallel to the helical axes of the hexamers in the crystal. The hexamer duplex which was used to calculate the difference map is drawn with thick bonds. The atoms of the alternative packing model (that was not used to calculate the maps) are drawn as balls and sticks using thin bonds.

The differences between the models can be better appreciated when the manner in which the asymmetric units pack within the crystal lattice is considered. The Z-DNA crystal lattice stacks the hexanucleotide duplexes end-to-end to generate essentially continuous helical strands of Z-DNA. The hexamer duplex units of this non-self-complementary sequence may stack in either of two orientations in the crystal lattice, both of which maintain the stacking interactions within the crystalline helical arrays. Figure 1 shows crystal packing diagrams of the two possible orientations the hexamer duplex can assume in the $P2_12_12_1$ space group. Each hexamer duplex, represented as boxes in the figure, comprises two strands of complementary hexanucleotides of DNA. In this figure, the d(m^{5}CGGGm^{5}CG) hexanucleotides are shown in blue, while the d(m^{5}CGCCm^{5}CG) hexanucleotides are colored red. The differences between the two models can be illustrated by considering the numbering scheme of the antiparallel double helix. In the first model, dubbed the GG model (Figure 1A), residue numbers 1–6 are assigned to the d(m^{5}CGGGm^{5}CG) hexanucleotide in the 5′ to 3′ direction, and residue numbers 7–12 are assigned to the complementary d(m^{5}CGCCm^{5}CG) hexanucleotide in the 5′ to 3′ direction. Thus, in this GG model, the GG steps occur at residues 3 and 4, while the CC steps are at residues 9 and 10. The G3·C10 positions represent the base pair that is out-of-alternation. In the opposing CC model (Figure 1B), the residue assignments for the strands are reversed, where the d(m^{5}CGCCm^{5}CG) hexanucleotide is now assigned residue numbers 1–6 in the 5′ to 3′ direction and the d(m^{5}CGGGm^{5}CG) hexanucleotide is assigned residue numbers 7–12 in the 5′ to 3′ direction. With these alternate residue assignments for the hexanucleotides in the CC model, the CC steps now occur at residues 3 and 4 and the GG steps are at residues 9 and 10. The GG model lattice exposes guanines 3 and 4 to an open solvent channel within the crystal lattice while cytosines 9 and 10 are packed tightly against d(m^{5}CG) base pairs 1·12 and 2·11 of the symmetry-related hexanucleotide. The alternative orientation of the hexamer duplexes in the CC model reverses these crystal-packing interactions. The base pair out-of-alternation for the CC model is C4·G9. Both packing models are self-consistent with the symmetry of the $P2_12_12_1$ space group, and therefore have identical stacking interactions, and maintain the directions of the two antiparallel strands that make up the continuous Z-DNA helices. Thus, the CC and GG models show that two discrete and distinguishable orientations of the asymmetric hexamer duplex can be accommodated in a single crystal lattice.

Figure 2 compares selected $2(F_o - F_c)$ difference electron density map sections of base pairs G2·C11 (panels A and D), 3·10 (panels B and E), and 4·9 (panels C and F) from refinement of the diffraction data using the CC model versus the GG model. The sections of the G2·C11 base pair in the CC and GG models shown in panels A and D, respectively, compare the refined positions of a base pair common to both models. As expected, the atoms of these common base pairs refine to identical positions in both models. The difference map calculated using the CC model (panels A, B, and C) shows more background noise compared to that calculated

using the GG model, reflecting a better fit of this latter model to the data (R = 27.0% for the CC model, versus 24.3% for the GG model). The relatively low level of residual difference density at the G2·C11 base pair in these map sections shows that both models are consistent with the observed data in the regions that are common between the models and are characteristic of the remaining common C1·G12, C5·G8, and G6·C7 base pairs.

The greatest difference between the models of the two packing orientations occurs at the central 3·10 (panels B and E) and 4·9 base pairs (panels C and F). In these regions, interchanging of the strands between the two models is apparent in the reversal of the positions of the bases of the central C·G base pairs. A comparison of these $2(F_o-F_c)$ map sections shows that the density contours associated with the base atoms are not unreasonable in appearance for both models. However, the presence of residual density in the spaces between the paired bases indicates that neither model alone can account for all the electron density in these variable regions of the crystal. The residual density observed in panels B and C of Figure 2 (calculated with the CC model) clearly shows that the additional atoms of a purine base are present where the model places a pyrimidine base, and vice versa. Similarly, the map calculated using the GG model shows that atoms associated with the alternative CC model are present, although to a lesser extent. Additional evidence for the two packing models for this structure can be seen at the phosphates of the C3 (Figure 2B) and G3 bases (Figure 2E). The difference maps at these positions clearly show residual densities that locate the slightly displaced phosphates of the alternative packing models. A comparison to the map sections of the G2·C11 base pair which is common to both models shows that the quality of the data should give a very clean fit of the model to the calculated electron density maps for base pairs that are identical to both the CC and the GG model.

Thus, the $2(F_o - F_c)$ difference maps indicate that the asymmetric duplex does not pack exclusively in either of the two orientations shown in Figure 1, but distributes between both packing orientations with a mixture of helices in the orientations shown. Although the duplexes in this mixture will still form a continuous Z-DNA helix, they would no longer be related by the crystallographic symmetry of the $P2_12_12_1$ space group. The distribution of hexamer duplexes between the two packing orientations in the crystal ultimately results from differences in the crystal-packing interactions a hexamer duplex experiences in each orientation. These interactions include both end-to-end base-stacking interactions between hexamer duplex units and interactions between the major-groove surfaces of the duplex units of different helices (side-on interactions). Therefore, the crystal structure of this asymmetric hexamer duplex allows the examination of how the DNA structure is influenced by DNA–DNA interactions in the crystal lattice.

Occupancy Refinement and Final Positional Refinement of the GG and CC Models. Initially, separate refinements of the GG and CC models to the crystallographic data were carried out using the simulated annealing and conventional positional refinement routines of X-PLOR (Brunger, 1992). The final R-factor of 24.3% for the GG model versus an R-factor of 27.0% for the CC model suggested that the two arrangements of hexamers do not contribute equally to the lattice of this crystal. Since each of the separately refined models contained identical numbers of solvent molecules, the differences in the R-factors indicate that the structure in the GG model can better account for the calculated electron densities. This is shown in Figure 2, in which the GG model much more accurately accounts for the electron density at the central two base pairs than does the CC model (compare panels E and F to panels B and C).

To quantitate the contribution of each model (as measured by the percent occupancy of each model in the crystal) to the overall structure of the crystal, we refined the structure using a single composite model constructed from the separately refined models of the two packing orientations. We used the occupancy refinement algorithm in X-PLOR to determine the contribution of these models to the overall crystal packing (Brunger, 1992). The atoms of the four common flanking base pairs, along with the atoms of the backbone atoms of the central two base pairs, refined to very similar positions in the composite model. The atoms that are unique to each model are the base atoms of the two central base pairs. The atomic positions unique to each orientation allow the greatest distinction between the two packing models (as seen in the difference map in Figure 2) and therefore would be the most reliable measure of the occupancy of each model.

Although the central base pairs of the two packing models differ in their orientation, a number of atoms overlap in this region between the two models. Comparing the density sections of the 4·9 base pair of the two models (Figure 2C,F), 7 of the 11 atoms of the guanine very nearly overlap with 7 of the 8 cytosine base atoms. This leaves 5 atoms per base pair in the central dinucleotides (or a total of 20 atomic coordinates between the two models) that have unique and distinguishable positions between the two packing orientations. The unique atoms used in the occupancy refinement were the N1, C2, C6, and O6 atoms of the guanine bases and the N4 nitrogen of the cytosine bases at the 4·9 and 3·10 base pairs. The contributions of the two orientations to the crystal lattice were determined after final refinement to give a distribution of 65% GG model and 35% CC model in this crystal. The electron density maps in Figure 3 were calculated from the distribution of orientations seen in the composite models. This $2F_o - F_c$ map shows that, at this resolution, the calculated electron densities fit all the atoms that are common and that are unique to the two packing models.

Using the Boltzmann relationship for two distinct states, the 65:35 distribution of the hexamer duplex between the two packing orientations in the crystal translates into a difference in interaction energy of approximately –0.4 kcal/mol of hexamer in favor of the GG packing orientation. The common flanking d(m^{5}CG) dinucleotides would not be expected to favor one orientation versus the other (although, as shown in Figure 3, they do not have identical conformations). The distribution of packing orientations in the crystal is therefore likely to be dependent upon interactions which occur at the central two base pairs with base pairs of adjacent, symmetry-related hexanucleotides.

Molecular Structure of d(m^{5}CGCCm^{5}CG)·d(m^{5}CGGGm^{5}CG). With the arrangement of the asymmetric unit in the crystal determined, we can now start to analyze the structures of the two contributing models to try to develop a molecular rationale for why nonalternating bases are disfavored in Z-DNA. By comparing the structures of the two models, we should be able to differentiate between structural effects that are inherent to having a single out-of-alternation base pair in Z-DNA versus the structural effects induced by packing forces in the crystal lattice and/or the refinement process. We observe that overall both the GG and the CC model have features typical of a Z-DNA hexanucleotide structure; however, both models differ slightly, but signifi-

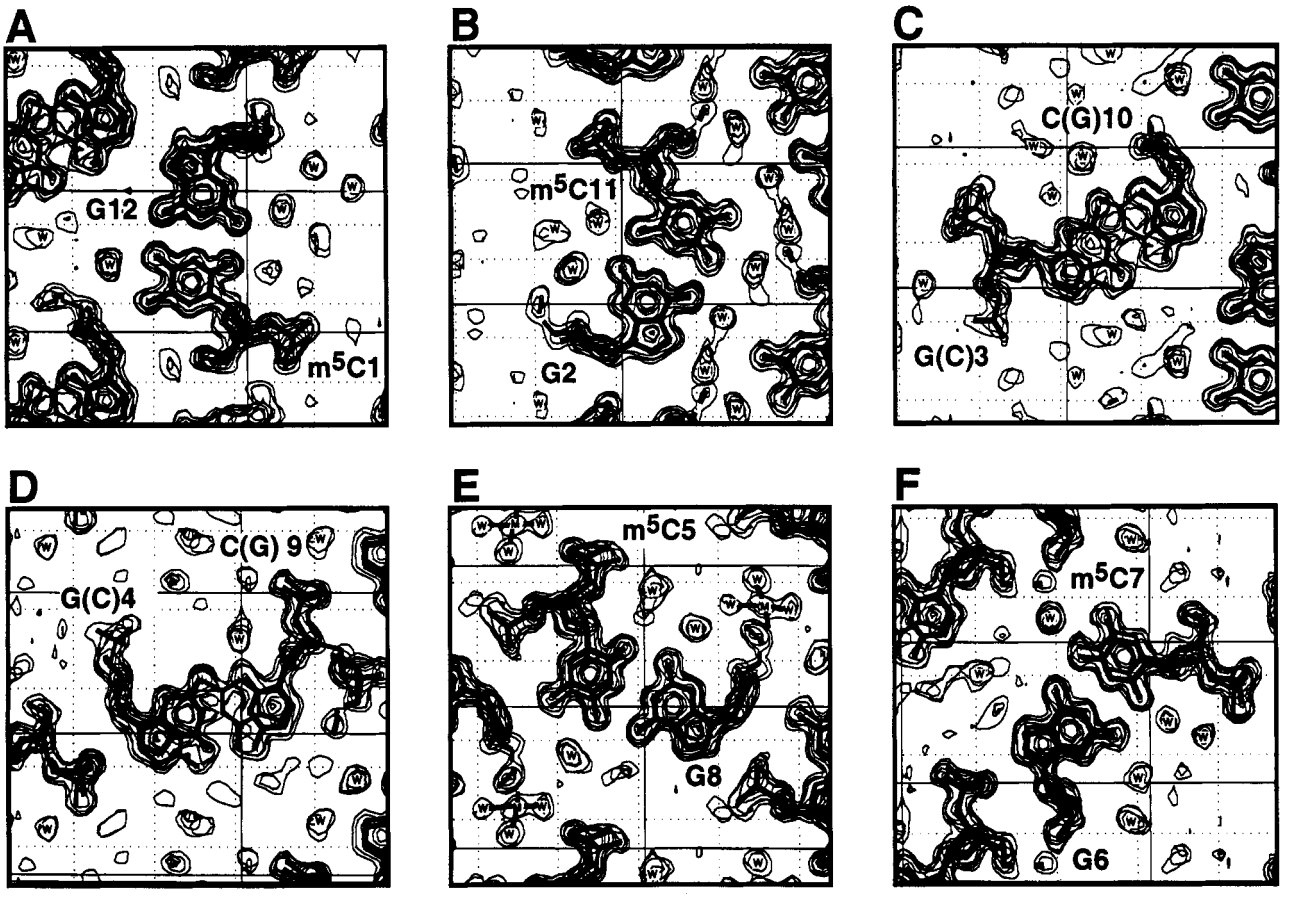

FIGURE 3: $2F_o - F_c$ electron density map calculated using the occupancy-refined model (65% GG and 35% CC model). Panel A shows the electron density of the m^{5}C1·G12 base pair; panel B, the G2·m^{5}C11 base pair; panel C, the 3·10 base pairs positions; panel D, the 4·9 base pair positions; panel E shows m^{5}C5·G8; panel F shows G7·m^{5}C6. The GG model is drawn with thick bonds, and the CC model is drawn with thin bonds.

cantly, at each of the bases that are out-of-alternation. We also find that the two hexamer models have other structural differences which are most likely due to a combination of differences between crystal packing of the GG versus the CC model, coupled with effects associated with the refinement process.

Both of the refined structures are in the left-handed Z-DNA conformation and maintain all of the characteristic structural features of Z-DNA, including alternating *anti/syn* glycosidic bond angles, alternating C2′/C3′-*endo* sugar puckers, and alternating patterns of base pair morphologies such as buckle and twist. In Table II we compare the standard conformational angles, as well as the sugar pseudorotation angles, for the GG and CC models. In this table we also compare both of these out-of-alternation structures to the 0.9-Å structure for the prototypical Z-DNA-forming sequence d(CGCGCG), a structure which serves as the highest resolution example of an isomorphous, in-alternation Z-DNA hexamer (Gessner et al., 1989). In both of our models the base pairs that are out-of-alternation are able to assume the general alternating *anti/syn* conformation, as seen in the in-alternation Z-DNA hexamer d(CGCGCG). For instance, in both models the out-of-alternation cytosine is shown to readily adopt a *syn* conformation, with very standard χ angles of 248.3° for C10 of the GG model and 251.1° for C4 of the CC model (see Table II). Not surprisingly, the out-of-alternation guanine residues also readily assume the *anti* conformation in these structures, even though they are at positions of the hexamers usually occupied by cytosines in Z-DNA. All of the χ values observed in the GG and CC models are well within the range of alternating *anti* and *syn* conformations seen in the structure of d(CGCGCG), and therefore the structure of Z-DNA across nonalternating sequences appears to be mostly dictated by the alternation of *syn* and *anti* glycosidic conformations, rather than the preference of each base for a given conformation. The conservation of the alternating *anti/syn* conformations was also observed in the crystal structure of d(m^{5}CGATm^{5}CG) at the out-of-alternation adenine and thymine residues of each strand (Wang et al., 1985).

Although the GG and CC models are both clearly in standard Z-DNA conformations, the two crystallographically unique models are not identical. One interesting difference is that only the GG model, and not the CC model, has the Z_{II} backbone conformation at residues 4 and 5. This is a feature that has been noted in most other isomorphous hexamer Z-DNA structures, and it is thought to be an important packing feature of Z-DNA hexamers in the crystal lattice (Schneider et al., 1992a). This difference supports the notion that the CC model is interacting differently with its neighboring hexanucleotides compared to the crystal packing of the GG model, an idea which is also supported by several other structural variations between the two models. In Table II we have highlighted all of the angles which are the most variable between the GG model, the CC model, and the standard Z-DNA structure of d(CGCGCG). In this comparison only the CC model has conformational angles which are significantly different from the other two models. These unusual conformational angles are generally consistent with a different crystal-packing environment for the CC model compared to either the GG model or the d(CGCGCG) hexamers. The

Table II: Comparison of Conformational Angles[a]

	residue[b]	angle							
		α	β	γ	δ	ϵ	ζ	X	P
(1)	m^{5}Cyt	n.d.	n.d.	59.4	145.5	264.8	76.0	37.4	161.8
		n.d.	n.d.	*179.3*	154.4	260.9	76.2	30.9	168.8
		n.d.	n.d.	(47.5)	(144.3)	(267.4)	(80.8)	(31.6)	(154.3)
(2)	Gua	72.2	184.4	175.6	105.9	231.6	311.1	245.3	29.6
		67.7	189.1	181.4	92.3	236.7	295.1	255.1	24.0
		(63.4)	(186.4)	(174.1)	(94.0)	(239.6)	(294.0)	(245.8)	(30.5)
(3)	Gua or Cyt	185.7	222.7	64.1	158.9	264.6	64.7	29.2	167.7
		209.3	241.3	47.8	131.5	269.3	80.3	22.4	140.4
		(216.0)	(234.6)	(50.2)	(152.1)	(257.8)	(75.6)	(22.2)	(148.1)
(4)	Gua or Cyt	78.7	188.5	177.8	94.3	182.6	70.2	238.7	18.4
		56.2	188.9	179.8	91.6	*243.7*	303.2	251.1	21.6
		(70.1)	(187.9)	(177.5)	(95.1)	(181.2)	(65.0)	(240.4)	(30.3)
(5)	m^{5}Cyt	172.7	150.8	44.5	138.7	265.3	69.1	29.0	144.5
		207.6	213.4	58.9	144.5	247.0	81.7	20.9	151.4
		(169.3)	(167.0)	(42.7)	(142.0)	(266.6)	(74.3)	(35.4)	(154.7)
(6)	Gua	93.3	174.8	177.6	143.8	n.d.	n.d.	265.9	160.1
		74.4	184.8	190.4	153.6	n.d.	n.d.	255.6	181.5
		(73.9)	(178.0)	(179.6)	(148.7)	n.d.	n.d.	(256.2)	(161.8)
(7)	m^{5}Cyt	n.d.	n.d.	54.9	141.5	265.6	73.7	30.8	144.5
		n.d.	n.d.	*298.7*	149.1	257.9	78.2	23.0	158.8
		n.d.	n.d.	(53.1)	(146.8)	(270.7)	(77.6)	(36.6)	(161.0)
(8)	Gua	70.3	188.9	172.5	104.7	253.5	295.2	247.2	28.2
		74.8	190.8	179.1	92.4	241.3	282.8	257.5	32.0
		(61.3)	(186.9)	(174.6)	(95.2)	(244.2)	(286.3)	(251.8)	(29.3)
(9)	Cyt or Gua	204.9	210.7	59.1	147.5	257.9	69.6	30.5	162.8
		189.0	244.9	63.8	130.3	275.4	68.2	25.3	142.2
		(220.3)	(225.6)	(54.7)	(148.9)	(261.9)	(79.9)	(23.6)	(154.3)
(10)	Cyt or Gua	74.6	187.7	181.3	100.8	242.9	301.4	248.3	19.0
		70.4	180.1	177.1	93.5	232.6	295.7	254.4	22.7
		(64.2)	(178.8)	(179.0)	(102.7)	(247.6)	(290.2)	(250.7)	(18.4)
(11)	m^{5}Cyt	190.9	229.0	59.1	132.9	266.0	73.0	26.4	143.2
		199.4	240.6	59.0	140.9	260.6	67.3	22.2	153.7
		(212.4)	(236.3)	(50.1)	(143.3)	(262.5)	(71.8)	(21.5)	(146.9)
(12)	Gua	75.1	184.5	187.5	153.0	n.d.	n.d.	266.2	170.9
		79.4	185.8	183.9	*103.2*	n.d.	n.d.	246.6	*56.9*
		(78.4)	(184.0)	(186.0)	(148.7)	n.d.	n.d.	(260.2)	(167.5)

[a] The conformation angles of three different structures are compared: (1) The top value is calculated from the GG model. (2) The middle value is calculated from the CC model. (3) The bottom value (in parentheses) is calculated from the Gessner et al. (1989) Mg^{2+}/spermine structure of d(CGCGCG). [b] Residues in the three hexamers are numbered 1–6 and 7–12 on each strand. The 3/4 and 9/10 positions are different for each these three hexamers: (1) For the GG model these positions are GG and CC. (2) For the CC model they are CC and GG. (3) For d(CGCGCG) they are CG and CG.

difference map in Figure 2 shows that, at least for the base at position 4, the positions of the phosphate backbone are significantly different between the two models. However, some of the details derived from the CC model may not be as reliable due to the lower contribution of this model relative to the GG model to the overall structure of the crystal.

The most interesting observation of these two structures is that the out-of-alternation base pair is significantly distorted and that this distortion has structural effects on both base pairs adjacent to the cytosine in the *syn* conformation. Figure 4 shows two views of the out-of-alternation 3·10 base pair, along with each of the flanking base pairs, from the GG model. From the side view (top two structures), one can see that the cytosine in the *syn* conformation (labeled "C10" in Figure 4) is not planar with its complementary base and, in fact, is in a highly buckled conformation. This buckling is further quantitated in Table III, where we have listed some angles to describe the base-pair morphology of these hexamers. For each type of angle in Table III we compare the GG and CC models to average values, with standard deviations, which were calculated from 10 other APP Z-DNA crystal structures deposited in the Nucleic Acids Database (NDB; Berman et al., 1992).

Base-pair buckle is a characteristic feature of Z-DNA, and it generally alternates between low and high values along a Z-DNA hexamer structure. This alternating buckle pattern holds true for both the GG and the CC model; however, in both of these structures, the out-of-alternation C·G base pairs (base pair 3·10 in the GG model and base pair 4·9 in the CC model) have larger than average buckle angles. The buckle at the 3·10 base pair of the GG model is −13.4°, while the average for other Z-DNA structures at this position of the hexamer is −5.2° ± 1.7°. This shows that the out-of-alternation base pair in the GG model has been forced to adopt a very unusual conformation (over 4SD away from average Z-DNA). According to Wang et al. (1985), in the crystal structure of d(m^{5}CGATm^{5}CG) both of the internal base pairs also have large buckle angles, indicating that this may be a feature common to out-of-alternation base pairs in Z-DNA. It seems that the cytosine in the *syn* position is the major determinant of the increased buckle seen in these structures, because we find that it, much more than the guanine, has moved out of the plane of the 3·10 base pair (shown clearly in Figure 3). In addition to the large buckle, Table III also shows that the 3·10 base pair of the GG model is distorted in other ways, as it also has unusually large values

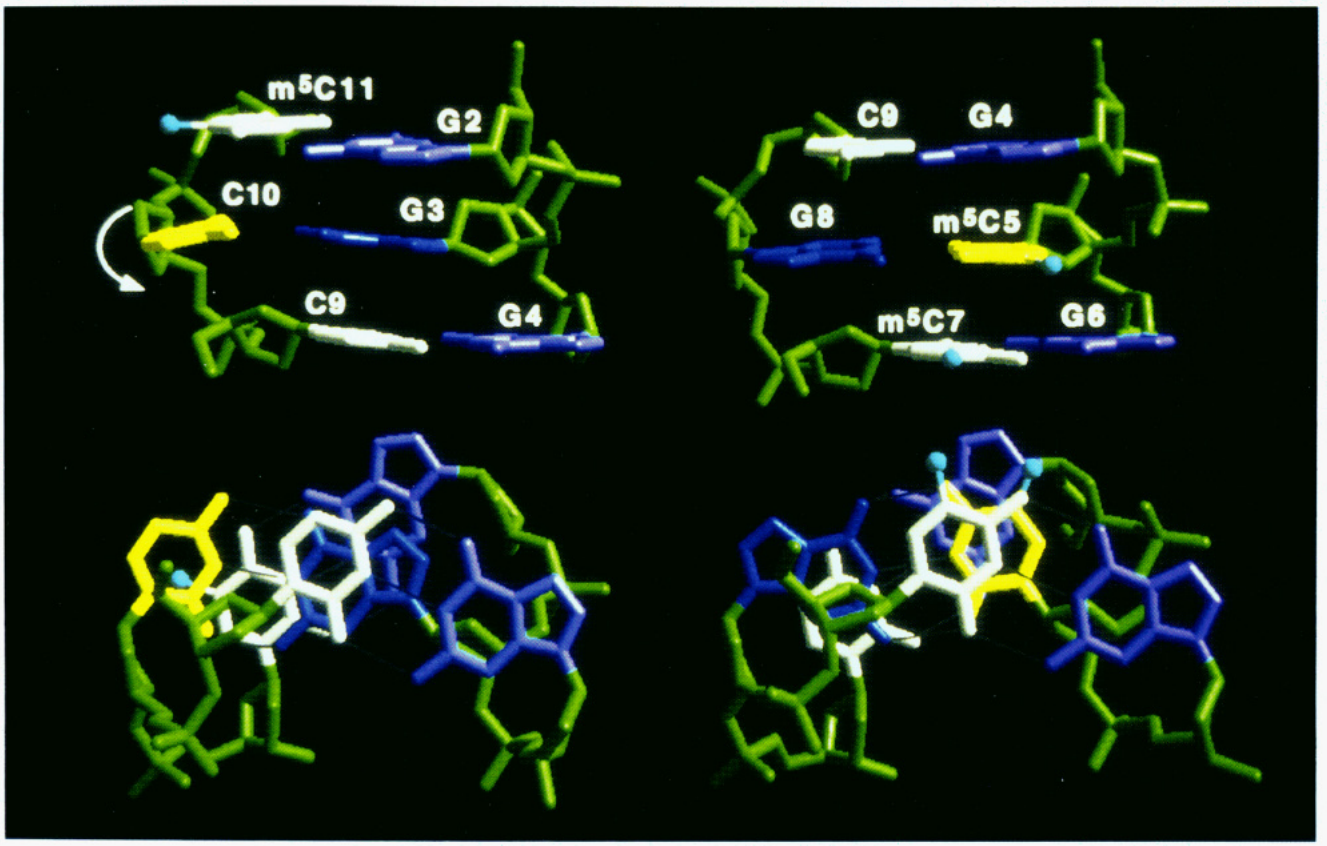

FIGURE 4: Structure of C·G out-of-alternation and in-alternation base pairs. Three d(C·G) base pairs (G2·m^{5}C11, G3·C10, and G4·C9) that place the C10 nucleotide in the disfavored *syn* conformation (left two structures) are compared to three base pairs (G4·C9, m^{5}C5·G8, and G6·m^{5}C7) that follow the all purines in *syn* and all pyrimidines in *anti* conformation rule that favors Z-DNA formation (right two structures). All bases are taken from the current structure. The central base pairs of each three-base-pair stack are highlighted by coloring the cytosine base yellow and the guanine base blue. The cytosine bases of the flanking base pairs are white, and the flanking guanine bases are violet. The backbone atoms are green, and methyl groups are shown as cyan balls off the C5 position of the cytosines. The two top figures are views perpendicular to the helical axis, while the two bottom figures are views down the stack and along the helical axis. The arrow indicates the rotation of the glycosidic bond of the C10 base resulting in the large base-pair buckle of this cytosine in the *syn* conformation.

for base-pair angle (over 3SD larger than average) and propeller twist (2SD away from average). Finally, the unusual conformation of the 3·10 base pair in the GG model causes perturbations of the two adjacent base pairs. Both the 2·11 and the 4·9 base pair of the GG model also have altered base-pair morphologies (see italicized values in Table III), as a result of their proximity to the cytosine in the *syn* conformation.

The bottom two structures in Figure 4 show views of the base stacking in the GG model along the crystallographic *c*-axis near the out-of-alternation 3·10 base pair. Most strikingly we find that the C10 cytosine in the *syn* conformation is almost completely unstacked from between the adjacent 2·11 and 4·9 base pairs. As a comparison, on the right side of Figure 4 we show the normal stacking seen at an analogous position of in-alternation base pairs of the current structure. This lack of strong base stacking, which appears to be a direct result of putting a cytosine base in the *syn* conformation, is probably a major determinant in the lowered stability of non-APP sequences. This unstacking also probably gives the cytosine base the flexibility to buckle and twist in the manner described above and given in Table III. In conclusion, the significant molecular differences between this crystal structure and other regularly repeating APP crystal structures, such as d(CGCGCG), can easily justify the differences in thermodynamic stability of the d(GG)·d(CC) dinucleotide compared to the d(CG) dinucleotide (Ellison et al., 1985).

A number of solvent molecules can be clearly identified in the $2F_o - F_c$ electron density map of the composite structure (Figure 3). It is clear from this figure that, for each of the common in-alternation base pairs (panels A, B, E, and F), there are two well-defined water molecules in the minor-groove crevice that are within hydrogen-bond distance of the N3 amino groups of the guanine and the O2 keto oxygens of the cytosine bases. This is similar to the set of waters that constitute the spine of waters generally found in the minor grooves of both B- and Z-DNA oligonucleotides (Schneider et al., 1992b). This is not the case, however, for the cytosine in the *syn* conformation at the central base pairs (panels C and D). For this base (the C10 position of the GG model in Figure 3C and the C4 position of the CC model of Figure 3D), there are no solvent molecules located within 4 Å of the O2 keto oxygen of the cytosine. This suggests that, for both packing models, the base pair that is out-of-alternation disrupts the continuous spine of water in the minor-groove crevice.

For nearly all crystal structures of APP hexanucleotides in Z-DNA, there has been a hexaaquamagnesium cluster bridging the pyrimidine bases of the central base pairs at the major-groove surface (Gessner et al., 1989). This was not observed in the current structure. The electron density map (Figure 3C,D) shows only a diffuse set of contours which do not define any well-ordered solvent structure at the major-groove surface at the 3·10 and 4·9 base pairs. It is possible either that this magnesium cluster is too disordered to be defined in the crystal structure or that the base-pair stacking in this structure displaces the out-of-alternation cytosine so that the geometry of hydrogen-bond acceptors does not allow

Table III: Comparison of Selected Base-Pair Morphologies[a]

base pair	feature (deg)			
	buckle	prop. twist	open	angle
1·12	−2.5	−2.7	5.1	3.6
	−7.3	2.5	5.8	7.7
	−4.8(3.7)	−0.5(2.7)	4.1(1.0)	5.8(3.3)
2·11	*1.5*	4.7	*1.2*	4.9
	3.7	−3.8	*7.7*	5.3
	6.5(2.1)	0.5(3.7)	3.6(0.8)	7.5(2.2)
3·10	*−13.4*	*2.1*	2.4	*13.5*
	−7.0	−2.9	4.7	7.5
	−5.2(1.7)	−3.5(2.8)	5.2(2.9)	6.8(2.0)
4·9	10.0	*4.1*	2.5	10.8
	10.1	−2.6	4.7	10.4
	6.1(4.2)	−1.1(2.2)	3.0(1.5)	7.4(2.5)
5·8	−2.8	0.5	3.1	2.8
	0.2	*−7.1*	6.0	7.1
	−2.9(2.7)	−1.4(1.0)	4.3(1.6)	3.5(2.5)
6·7	6.0	−1.7	1.8	6.2
	6.2	−0.3	5.5	6.2
	1.7(3.5)	1.6(2.5)	4.8(2.9)	4.6(1.7)

[a] The values (in degrees) for base-pair buckle, propeller twist, base-pair opening and base-pair angle are given, as calculated using the method of Babcock and Olson (1992). The three values given are that for the GG model (top), that for the CC model (middle), and an average value for 10 other APP Z-DNA structures (bottom). The standard deviations for the average values are given in parentheses. The values which are at least 2SD away from the average Z-DNA values are italicized.

this cluster to bind (Figure 4). The only evidence for a cation cluster is located at the phosphate of guanine G8. This is identical to the clusters observed at this position in nearly all Z-DNA crystal structures (Gessner et al., 1989), and thus is not a feature unique to the out-of-alternation conformation.

Thermodynamic Stability of Pyrimidine Bases in the syn Conformation in Z-DNA. We have analyzed the structural features of the current sequence, d(m^{5}CGCCm^{5}CG)·d(m^{5}CGGGm^{5}CG), and compared these results to those of previously solved Z-DNA hexanucleotide structures. The most relevant comparisons are to analogous thermodynamic analyses for the structure of d(m^{5}CGm^{5}CGm^{5}CG), which has all the pyrimidine bases in the *anti* conformation, and of d(m^{5}CGATm^{5}CG), the only other oligonucleotide sequence which has been crystallized having pyrimidine bases in the *syn* conformation.

The commonly stated dogma is that pyrimidines are sterically inhibited from adopting the *syn* conformation (Haschmeyer & Rich, 1967). In terms of the overall internal energies of the Z-DNA duplex, this should present itself as less favorable (more positive) van der Waals energies for C4 of the CC model and for C10 of the GG model for this structure. To test this, we compared the van der Waals energies of the cytosines in *syn* for each of these models to cytosines in the standard *anti* conformation and guanines in the *syn* conformation [as in the sequence d(m^{5}CGm^{5}CGm^{5}CG)]. These results (summarized in Table IV) show that indeed the overall van der Waals energies of the GG model are higher than those of d(m^{5}CGm^{5}CGm^{5}CG). However, the base pair that is out-of-alternation (the 3·10 base pair in the case of the GG model) is overall negative. Furthermore, the overall van der Waals contribution to the internal energies of the CC model for the out-of-alternation structure is lower than that of the in-alternation d(m^{5}CGm^{5}CGm^{5}CG) sequence. Does this suggest that in fact pyrimidines are not sterically inhibited from adopting the *syn* conformation? The interpretation of these

Table IV: Comparison of Steric Energies of Base Pairs in the Structure d(m^{5}CGm^{5}CGm^{5}CG)[a] and in the Two Models (The GG and CC Models) of the Current Structure[b,c]

base pair	m^{5}CGm^{5}CGm^{5}CG	GG model	CC model
1·12	16.8	19.1	17.5
2·11	3.7	16.3	10.6
3·10	6.9	*−1.3*	3.2
4·9	2.3	4.8	*−0.3*
5·8	5.9	−5.0	7.4
6·7	18.4	21.7	13.1

[a] All the bases in this model follow the alternating pyrimidine/purine rule for Z-DNA formation. [b] One CG base pair is placed out-of-alternation in these models. [c] The van der Waals energies of the base pairs containing the out-of-alternation CG base pair are shown in italics. In the GG model, this is the G3·C10 base pair, while in the CC model, this is the C4·G9 base pair.

results must be tempered by the fact that the current structure was refined using X-PLOR, which utilizes an energy optimization algorithm during refinement, whereas the d(m^{5}CGm^{5}CGm^{5}CG) structure was not. Nonetheless, the steric interactions of the base pair that neighbors the cytosine in *syn* (specifically, the C2·G11 base pair of the GG model and the G5·C8 base pair of the CC model) are significantly more positive than those of the analogous base pairs in the alternating structure.

In order to resolve this apparent contradiction, a model of d(m^{5}CGCCm^{5}CG)·d(m^{5}CGGGm^{5}CG) was built by replacing the G4·C9 base pair of the d(m^{5}CGm^{5}CGm^{5}CG) structure with the inverted C4·G9 base pair and demethylating the C3 cytosine base. In forcing a cytosine base to adopt the *syn* conformation, in the context of the standard Z-DNA structure, we find that, indeed, there is a steric collision between the pyrimidine base and the deoxyribose sugar of that residue. The resulting van der Waals energy term for the C4·G9 position of this model is 3.5 kcal/mol higher than for the in-alternation G·C base pair that would normally be at this position. The higher energy in this model is primarily associated with a collision between the O2 oxygen of the pyrimidine base and the hydrogen at the C3′ position of the deoxyribose sugar ring. The van der Waals energy for the full hexamer is increased by 5.6 kcal/mol, most of which is centered at the C4 base. The remaining 2 kcal/mol of energy is equally distributed between the adjacent C3·G10 and C5·G8 base pairs, primarily resulting from loss of favorable contacts between the C4·G9 base pair and its 5′ and 3′ nearest neighbors.

To determine how the structure might adjust to accommodate these steric problems, we minimized the energy of the central C3·G10 and C4·G9 base pairs of this model structure, holding the two 5′ flanking and the two 3′ flanking base pairs fixed to simulate the strong crystal-packing forces that exist at these positions. The structure was optimized to an overall energy gradient of 1 kcal/mol/step. Additional minimization of the structure did not result in any significant alterations to the structure. The final optimized structure showed that, indeed, the out-of-alternation C4 cytosine residue is distorted to relieve the steric collision of the original model. This distortion is manifested as a large buckling of the cytosine base of the C4·G9 base pair (−14° for the base pair), similar to that observed in the current d(m^{5}CGCCm^{5}CG)·d(m^{5}CGGGm^{5}CG) crystal structure. Even though the in-alternation C3·G10 was also free to relax during the molecular simulation, no significant distortions were observed for this base pair. These results strongly suggest that the out-of-alternation cytosine in Z-DNA will buckle to relieve the strain of placing this pyrimidine base in the *syn* conformation.

We may now ask, what is the root of the observed buckling of the cytosine base when placed in the *syn* conformation? One may suspect that the features of the out-of-alternation base pairs in the crystal may be artifacts associated with the energy minimization algorithm of the X-PLOR molecular refinement package. However, nearly identical structures for both the GG and the CC model were obtained using the Hendrickson–Konnert least-squares refinement method (Hendrickson & Konnert, 1979), which does not employ molecular simulation to restrain the models. This indicates that the structural perturbations observed are indeed directly associated with the conformation of the molecules in the crystal. To study the molecular basis for the base-pair buckling, we constructed a model for the sequence d(CGCCCG)·d(CGGGCG), where the cytosines are not methylated. This structure was energy minimized as before. The resulting model showed that the out-of-alternation C4·G9 base pair is only slightly buckled, but instead slides significantly toward the major-groove surface of Z-DNA. Thus, there is a different mechanism which the cytosine base can exploit to adopt the *syn* conformation in Z-DNA, and that is to slide the base away from the ribose sugar ring. This demonstrates that the methyl group of the 5′ adjacent cytosine base, which stacks directly below the cytosine in *syn*, is responsible for the buckling of this base. Although the sliding mechanism does not play a role in relieving the stress of the out-of-alternation pyrimidine bases in the current crystal structure, a similar displacement toward the major-groove surface of thymine bases that are out-of-alternation was observed in the crystal structure of the sequence d(m^{5}CGATm^{5}CG) (Wang et al., 1985).

The observed distortions to the model structure and within the current crystal structure of the cytosine in *syn* are shown to relieve up to 5 kcal/mol of steric energy. The experimentally determined $\Delta G_{T(B-Z)}$ value of 2.4 kcal/mol of dinucleotide for a d(CC)·d(GG) dinucleotide nonetheless indicates that a pyrimidine in the *syn* conformation is an unfavorable conformation. If we assert that the distortion associated with the cytosine in the *syn* conformation results in an exposure of the base in the major-groove surface, then the higher energy may be manifested in this distorted conformation. One possible consequence of this type of distortion would be to expose a larger surface of this base to water. Thus, we could reason that the unfavorable energy associated with a cytosine in the *syn* conformation of Z-DNA would appear as an increase in the hydrophobicity of the exposed surface of the DNA. To test this, we have calculated the solvation free energy (SFE) for both models of the current crystal structure and compared these to the SFEs of the standard alternating pyrimidine/purine crystal structures and the out-of-alternation d(m^{5}CGATm^{5}CG) structure. These results are listed in Table V, along with the calculated SFEs for model B-DNA structures of these same sequences to allow an estimate of the hydration effects on the relative stabilities of B- and Z-DNA in each case. It is clear from this analysis that placing one or more pyrimidine bases out-of-alternation increases the hydrophobicity of Z-DNA and thus contributes to the destabilizing effect of placing cytosine and thymine bases in the *syn* conformation. At the atomic level, this was observed as a disruption of the spine of waters in the minor-groove surface at the out-of-alternation base pairs, and perhaps is related to the lack of ordered solvent structure at the major-groove surface of these base pairs (Figure 3).

Table V: Comparison of the Z- Minus B-DNA Solvent Free Energies ($\Delta\Delta G_{H(Z-B)}$) and the Calculated B- to Z-DNA Transition Free Energies ($\Delta G_{T(B-Z)}$) of CG Dinucleotides That Are In-Alternation versus Out-of-Alternation[a]

	crystal structure		
	d(m^{5}CGm^{5}CGm^{5}CG)	d(m^{5}CGGGm^{5}CG)	d(m^{5}CGATm^{5}CG)
$\Delta\Delta G_{H(Z-B)}$	−0.3	*0.2*	2.4
$\Delta G_{T(B-Z)}$	−0.2	*1.2*	7.4

[a] Values are averaged across the four internal base pairs of the d(m^{5}CGm^{5}CGm^{5}CG), the current d(m^{5}CGGGm^{5}GG)·d(m^{5}CGCCm^{5}CG) (values shown in italics), and the d(m^{5}CGATm^{5}CG) single-crystal structures. Free energies are in kcal/mol/dinucleotide. The transition free energies ($\Delta G_{T(B-Z)}$) were estimated from the relationship $\Delta G_{T(B-Z)} = 1.41\Delta\Delta G_{H(Z-B)} + 0.65$, as derived in Kagawa et al. (1993). For comparison, the values for the same base pairs of a model structure, d(m^{5}CGCGm^{5}CG), which has the same mix of methylated and unmethylated d(CG) base pairs as the current structure, are −0.23 for $\Delta\Delta G_{H(Z-B)}$ and −0.06 for the estimated $\Delta G_{T(B-Z)}$.

DISCUSSION

We present the 1.3-Å crystal structure of the sequence d(m^{5}CGGGm^{5}CG)·d(m^{5}CGCCm^{5}CG) in the left-handed Z-conformation. This sequence was designed to study the effects of placing a single cytosine base out-of-alternation (in the *syn* conformation) within the context of the left-handed Z-DNA structure. Since this is not a self-complementary sequence, the two strands of the DNA duplex are structurally unique and can thus pack in two opposite orientations within the lattice. Solving the structure of this hexanucleotide, therefore, required that we distinguish between the two orientations and resulted in two distinct models (the GG model and the CC model; see Figures 1–3) for the sequence. The two models are distinguished by the position of the out-of-alternation C·G base pair within the sequence and within the crystal lattice. The GG model places this cytosine at the G3·C10 base pair, with the C10 cytosine in *syn* and packed tightly against the G12 of an adjacent symmetry-related hexamer. The CC model places this same base at the C4·G9 base pair position, which exposes the cytosine to an open solvent channel. The crystal lattice can only be completely described by a combination of both models, with the GG model representing the arrangement of ~65% of the asymmetric unit within the crystal, while 35% of the hexamers are arranged according to the CC model. This translates into approximately −0.4 kcal/mol of difference in crystal-packing free energy for the two arrangements. We are currently varying the sequence and crystallization conditions to affect this distribution of packing arrangements in order to study the thermodynamics of crystal packing in DNA crystals.

By comparing the two structures, we were able to distinguish between structural features that differed between the two models (those which are associated with crystal-packing effects) and features that were shared between the models (those which are inherent to placing a single cytosine base in the *syn* conformation). The most striking features common to both models involve buckling of the out-of-alternation cytosine base in the 5′ direction of the hexamer (see Table III) and relative unstacking of this base from between adjacent bases (see Figure 4). These distortions serve to relieve the steric problems resulting from placing a pyrimidine base in the *syn* conformation (Haschmeyer & Rich, 1967). The buckling is directly associated with the methyl group of a 3′ adjacent cytosine within the sequence and is primarily observed as an increase in the van der Waals energy of the structure. Surprisingly, the cytosine base that is forced to adopt the unfavorable conformation does not, however, show the largest van der Waals energy. The out-of-alternation base pair in fact has the lowest energy in both models derived from the single-crystal data (Table IV). The steric problems appear to be distributed through the DNA hexamer and focus

primarily at the base pair that is directly adjacent in the 3′ direction, with some additional destabilization of the 5′ adjacent base pair. Thus a single cytosine placed in the *syn* conformation is both buckled and unstacked, which in the process destabilizes the two nearest neighboring base pairs. These results suggest that if two separate C·G base pairs are placed out-of-alternation in Z-DNA, we would necessarily distort two base pairs and destabilize a total of four additional base pairs, two adjacent to each out-of-alternation C·G. If, however, the base pairs out-of-alternation are directly adjacent to each other, we would still have two distorted bases, but perhaps only two adjacent base pairs would need to be destabilized. This suggests that contiguous bases out-of-alternation may be less destabilizing than isolated base pairs that are out-of-alternation in Z-DNA. Indeed, if enthalpic interactions are the primary determinants of the stability of sequences in Z-DNA, the energies calculated for the d(m^5-CGATm^{5}CG) crystal structure support this conclusion.

A competing mechanism in which the base slides toward the major-groove surface to relieve this strain was not observed in the current crystal structure. Both the sliding and the buckling distortions were, however, observed for the two thymine bases in *syn* in the crystal structure of the sequence d(m^{5}CGATm^{5}CG). The additional slide of the bases is responsible for the additional exposure of the hydrophobic bases to solvent in this latter structure and, thus, may be responsible for the additional instability of out-of-alternation d(AT) dinucleotides as Z-DNA.

ACKNOWLEDGMENT

We would like to thank Dr. Charles Campana of Siemens for collecting X-ray data sets for us and Dr. Marla S. Babcock of Rutgers University for calculating the conformational angles shown in Tables II and III. We would also like to thank Dr. Charles Robert, Blaine Mooers, and the rest of the Ho lab group for many helpful discussions throughout this work.

REFERENCES

Babcock, M. S., & Olson, W. K. (1993) in *Computation of Biomolecular Structures: Achievements, Problems, and Perspectives* (Soumpasis, D. M., & Jovin, T. M., Eds.) pp 65–85, Springer-Verlag, Heidelberg.

Berman, H. R., Olson, W. K., Beveridge, D. L., Westbrook, J., Gelbin, A., Demeny, T., Hsieh, S.-H., Srinivasan, A. R., & Schneider, B. (1992) *Biophys. J. 63*, 751–759.

Brunger, A. T. (1992) *X-PLOR: A system for X-ray crystallography and NMR*, Version 3.1, Yale University Press, New Haven, CT.

Connally, M. L. (1983) *Science 221*, 709–713.

Dickerson, R. E. (1992) *Methods Enzymol. 211*, 67–111.

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

Fujii, S., Wang, A. H.-J., van der Marel, G., van Boom, J. H., & Rich, A. (1982) *Nucleic Acids Res. 10*, 7879–7892.

Gessner, R. V., Frederick, C. A., Quigley, G. J., Rich, A., & Wang, A. H.-J. (1989) *J. Biol. Chem. 264*, 7921–7935.

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

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

Ho, P. S., Ellison, M. J., Quigley, G. J., & Rich, A. (1986) *EMBO J. 5*, 2737–2744.

Ho, P. S., Kagawa, T. F., Tseng, K., Schroth, G. P., & Zhou, G. (1991) *Science 254*, 1003–1006.

Jovin, T. M., Soumpasis, D. M., & McIntosh, L. P. (1987) *Annu. Rev. Phys. Chem. 38*, 521–560.

Kagawa, T. F., Stoddard, D., Zhou, G., & Ho, P. S. (1989) *Biochemistry 28*, 6642–6651.

Kagawa, T. F., Tseng, K., Howell, M. L., & Ho, P. S. (1993) *Nucleic Acids Res.* (in press).

McLean, M. J., Blaho, J. A., Kilpatrick, M. W., & Wells, R. D. (1986) *Proc. Natl. Acad. Sci. U.S.A. 83*, 5884–5888.

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

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

Schneider, B., Ginell, S. L., Jones, R., Gaffney, B., & Berman, H. R. (1992a) *Biochemistry 31*, 9622–9628.

Schneider, B.,Cohen, D., & Berman, H. M. (1992b) *Biopolymers 32*, 725–750.

Schroth, G. P., Chou, P.-J., & Ho, P. S. (1992) *J. Biol. Chem. 267*, 11846–11855.

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

Wang, A. H.-J., Hakoshima, T., van der Marel, G., van Boom, J. H., & Rich, A. (1984) *Cell 37*, 321–331.

Wang, A. H.-J., Gessner, R. V., van der Marel, G. A., van Boom, J. H., & Rich, A. (1985) *Proc. Natl. Acad. Sci. U.S.A. 82*, 3611–3615.

Wells, R. D. (1988) *J. Biol. Chem. 263*, 1095–1098.

Zacharias, W., O'Connor, T. R., & Larson, J. E. (1988) *Biochemistry 27*, 2970–2978.

Zhou, G., & Ho, P. S. (1990) *Biochemistry 29*, 7229–7236.