The Structures and Relative Stabilities of d(G·G) Reverse Hoogsteen, d(G·T) Reverse Wobble, and d(G·C) Reverse Watson-Crick Base-pairs in DNA Crystals

Blaine H. M. Mooers, Brandt F. Eichman and P. Shing Ho*

Department of Biochemistry and Biophysics, ALS 2011 Oregon State University Corvallis, OR 97331, USA

We have solved the structures of the homoduplex d(Gm^5CGCGCG)_2, and the heteroduplexes d(GCGCGCG)/d(TCGCGCG) and d(GCGCGCG)/d(CCGCGCG). The structures form six base-pairs of identical Z-DNA duplexes with single nucleotides overhanging at the 5'-ends. The overhanging nucleotide from one strand remains stacked and sandwiched between the blunt-ends of two adjacent Z-DNA duplexes, while the overhanging base of the opposing strand is extra-helical. The stacked and the extra-helical bases from adjacent duplexes pair to form a distorted d(G·G) reverse Hoogsteen base-pair in the d(Gm^5CGCGCG)_2 homoduplex, and d(G·T) reverse wobble and d(G·C) reverse Watson-Crick base-pairs in the d(GCGCGCG)/d(TCGCGCG) and d(GCGCGCG)/d(CCGCGCG) heteroduplexes, respectively. Interestingly, only the d(G·T) and d(G·C) base-pairs were observed in the heteroduplexes, suggesting that both the d(G·T) reverse wobble and d(G·C) reverse Watson-Crick base-pairs are more stable in this crystal environment than the d(G·G) reverse Hoogsteen base-pair. To estimate the relative stability of the three types of reverse base-pairs, crystals were grown using various mixtures of sequences and their strand compositions analyzed by mass spectrometry. The d(G·C) reverse Watson-Crick base-pair was estimated to be more stable by ~1.5 kcal/mol and the d(G·T) reverse wobble base-pair more stable by ~0.5 kcal/mol than the d(G·G) reverse Hoogsteen base-pair. The step during crystallization responsible for discriminating between the strands in the crystal is highly cooperative, suggesting that it occurs during the initial nucleating event of crystal growth.

Keywords: reverse base-pairs; DNA structure; crystallography; nucleic acid stability

*Corresponding author

Abbreviations used: RRE, Rev responsive element.

Introduction

The proper pairing of nucleotide bases ensures fidelity in replication and transcription of the genetic information in a cell. The pairing of guanine with cytosine and adenine with thymine in what is now known as standard Watson-Crick base-pairing forms the basis for the structure of antiparallel DNA and RNA duplexes. Unusual base-pairs, however, also play important roles in the transmission of genetic information. Here, we study the structures of reversed base-pairs formed by nucleotides that overhang at the 5'-end of the homoduplex d(Gm^5CGCGCG)_2, and of the d(GCGCGCG)/d(TCGCGCG) and d(GCGCGCG)/d(CCGCGCG) heteroduplexes. A distorted d(G·G) reverse Hoogsteen base-pair is formed in the homoduplex, while d(G·T) reverse wobble and d(G·C) reverse Watson-Crick base-pairs of the type observed in RNA structures form in the respective heteroduplexes.

The two strands of most DNA and RNA structures are oriented antiparallel to each other. In DNA duplexes, Watson-Crick-type base-pairs are the predominant interactions that hold the two strands together. When bases are mismatched in DNA, unusual base-pairing can occur, including wobble base-pairs between G and T and Hoogsteen-type base-pairs between two purine nucleotides (Figure 1). These are less stable than
standard Watson-Crick-type base-pairs. In RNA structures, unusual base-pairing is more prevalent, and has been observed to stabilize the complex tertiary structures of large polynucleotides such as tRNA (Quigley & Rich, 1976), hammerhead ribozymes (Pley et al., 1994), and the self-splicing group I intron from Tetrahymena thermophila (Cate et al., 1996).

Watson-Crick, wobble, and Hoogsteen-type base-pairs all have reverse analogues in which one base is completely inverted. Here, we refer to “reverse” base-pairs as the pairing of the nucleotide bases in which the glycosidic bonds are oriented essentially antiparallel to each other (Figure 1). The three base-pairs that we study here are asymmetric in the same manner that the normal pairings are asymmetric. The Hoogsteen and reverse Hoogsteen d(G·G) pairs match the Watson-Crick face of one purine with the Hoogsteen face of the other. The reverse analogues of the d(G·T) wobble and d(G·C) Watson-Crick base-pairs match the two respective Watson-Crick faces of the bases. In contrast, truly symmetric reverse base-pairs, with two identical bases related by a dyad axis perpendicular to the base-pair plane, have been used in the design of synthetic parallel-stranded DNA oligomers (Rippe et al., 1992; Robinson et al., 1994). These are interesting structures, although their biological relevance has yet to be determined.

In large RNA structures, however, loops that fold into local secondary and tertiary structures often require the formation of unusual base-pairs, including reverse base-pairs even if the strands are in antiparallel orientations. For example, in the crystal structure of yeast tRNA^{Phe}, a reverse Watson-Crick base-pair at (G15·C48) links the a region of the D arm to the variable V loop, and a reverse Hoogsteen base-pair at (G22·m7G46) links the D arm to the variable V loop (Kim, 1978). In a second example, a reverse Hoogsteen-type base-pair forms between G7 and G11 at the base of a GNRA structural motif in an RNA aptamer designed to recognize and bind ATP (Jiang et al., 1996). Finally, in the NMR solution structure of the hairpin formed by r(GGAC(UUCG)GUCC), a r(G·U) reverse wobble base-pair stabilizes the base of a two nucleotide loop (Varani et al., 1991). This hairpin structure is thought to occur frequently in ribosomal and messenger RNAs. Thus, non-Watson-Crick base-pairs are important for the proper folding of RNA molecules into the compact tertiary structure of their functional forms.
Non-Watson-Crick base-pairs are also important for RNA-protein recognition. Genetic and biochemical studies have shown that protein binding sites in RNA are often associated with important non-Watson-Crick base-pairs (Allmang et al., 1994; Ibba et al., 1996). In addition, protein binding of RNA loops can induce the formation of non-Watson-Crick base-pairs. For example, the HIV Rev peptide binding to the Rev responsive element in the env gene of HIV is associated with the formation of two homopurine base-pairs (Battiste et al., 1994; 1996).

The infrequent occurrence of reverse base-pairs makes it difficult to study the intrinsic stability associated with specific structures. We present here the atomic resolution structures of three different reverse base-pairs formed by nucleotides that overhang the 5'-ends of DNA duplexes. By studying the structures of these base-pairs and their effects on duplex formation during crystallization, we have estimated the stability of the d(G-C) reverse Watson-Crick and d(G-T) reverse wobble base-pairs relative to the distorted d(G-G) reverse Hoogsteen base-pair.

Results

We have solved the structures of the heptanucleotide duplexes d(GCGCGCG)\textsubscript{2} d(G\textsubscript{m5}CGCGCG)\textsubscript{2} (where m\textsubscript{5}C is cytosine methylated at the C5 carbon of the base), d(GCGCGCG)/d(TCGCGCG) and d(GCGCGCG)/d(CCGCGCG). In all four structures, the six underlined nucleotides pair to form left-handed Z-DNA d(CGCGCG) duplexes. The nucleotides within these duplexes are numbered 1 to 7 for each nucleotide of the common d(GCGCGCG) strand, and 8 to 14 for the opposing d(NCGCGCG) strand (where N is either G, C, or T). A single nucleotide (G1 of d(GCGCGCG) and N8 of d(NCGCGCG)) is left overhanging each of the 5'-ends of the duplexes. These overhangs pair with overhangs from adjacent duplexes in the crystal lattice to form three different reverse base-pairs. The overhanging dG nucleotides of the homoduplexes d(GCGCGCG)\textsubscript{2} and d(G\textsubscript{m5}CGCGCG)\textsubscript{2} form nearly identical reverse Hoogsteen-type d(G-G) base-pairs (rhGG: Figure 2(a)). However, only the structure of the methylated sequence will be discussed here; it provided a more reliable structure, as was evident from the final R-factors of the refined structures. The duplexes of d(GCGCGCG)/d(ITCGCGCG) form reverse wobble d(G-T) base-pairs (rwGT) (Figure 2(c)), while the duplexes of d(GCGCGCG)/d(CCGCGCG) form reverse Watson-Crick d(G-C) base-pairs (rwcGC) (Figure 2(b)). Thus, in all the structures, the overhanging nucleotide G1 remains stacked against the Z-DNA duplex, while N8 is extra-helical (Figure 3). In the remainder of this section, we will first discuss the duplex structures and crystal lattice interactions that are common to all the sequences, followed by a more detailed description of the structure for each type of base-pairing.

Z-DNA duplex structure

The six bases at the 3'-end of each sequence form standard Watson-Crick d(C-G) base-pairs. The resulting structure is a left-handed duplex that is nearly identical to the Z-DNA structure of d(CG)\textsubscript{3} (Wang et al., 1979, 1981). The base conformations alternate between ant\textsubscript{i} for the dC nucleotides (\(\chi \approx -150^\circ\)) and syn for the dG nucleotides (\(\chi \approx +60^\circ\)) along each strand of the duplex. The sugar conformations alternate between C2'\textsubscript{-endo} for the dC and C3'-endo for the dG nucleotides. The C3'-endo sugar facilitates formation of the syn conformation by the purine bases. The 3'-terminal dG nucleotide of each strand, although in the syn conformation, adopts a C2'-endo sugar conformation, as was observed in all the standard Z-DNA structures of hexanucleotides (Ho & Mooers, 1997).

Figure 2. Electron density omit maps of: (a) reverse Hoogsteen-like d(G-G) (rhGG); (b) reverse Watson-Crick d(G-C) (rwcGC); and (c) reverse wobble d(G-T) (rwGT) base-pairs. Shown are \(F_0 - F_c\) maps in which the overhanging bases at the 5'-ends were excluded from the phasing information used to calculate the structure factors. The hydrogen bonds that link the two bases of each base-pair and the common guanosine in syn to its deoxyribose sugar are indicated by the broken lines, along with the distances for each hydrogen bond.
Thus, the duplexes have the general features of Z-DNA as defined by the prototype structure of d(CG)$_3$.

The helical parameters of the duplex structures are compared to that of d(CGCGCG) in greater detail in Table 1. The helical twist alternates on average between $-9.1^\circ$ for the d(CpG) step and $-47.7^\circ$ for the d(GpC) step for all three structures, to give an average of $-28.4^\circ$ per base-pair. This is nearly identical to the $-30.1^\circ$ twist per base-pair observed in the structure d(CG)$_3$. When comparing the individual structures, the helical twists between the base-pairs of the homoduplex are identical (to within one standard deviation) to those of the analogous base steps in standard Z-DNA. The two heteroduplexes, however, are slightly less left-handed at the d(GpC) dinucleotide steps than observed in d(CG)$_3$. This distortion in the heteroduplex structures is associated with perturbations required to pair the base of the extra-helical nucleotides with the stacked guanosine residue G1 of an adjacent duplex.

The other obvious difference is the large buckle of the base-pairs at the ends of the duplex regions in all of the current structures. The Z-DNA structure of d(CG)$_3$ is very rigid, and the base-pairs very planar (showing very little propeller twisting or buckling). The high buckle at the ends is again likely associated with distortions required to form the unusual base-pairs here. All other helical parameters, including the rise at each base step and the propeller twist at each base-pair are identical to those in the structure of d(CG)$_3$.

The solvent structures at the major groove surface and the minor groove crevice of the Z-DNA duplexes are for the most part nearly identical to that of the magnesium form of d(CG)$_3$ (Gessner et al., 1989). For most base-pairs, two water molecules bridge the N4 amino groups at the major groove surface of adjacent cytidine residues in each d(CpG) dinucleotide step. Differences in the pattern of water interactions at this surface result from disruptions caused by the binding of a Co(NH$_3$)$_6^{3+}$ complex at the central d(CpG) dinucleotide. The amino ligands from this complex form hydrogen bonds to the O4 oxygen atom and N7 nitrogen atom of G5, while two amino groups hydrogen bond to the phosphate oxygen atom of the phosphodiester linking C4 to G5. This phosphate oxygen atom is also hydrogen-bonded to a water ligand of a hydrated magnesium complex. Water molecules from this Mg(H$_2$O)$_6^{2+}$ complex also form hydrogen bonds to the backbones of two adjacent duplexes and to the N7 of G12 on one of these duplexes. Thus, the hydrated magnesium
Table 1. Comparison of the helical parameters for base steps and base-pairs of the (CG)$_3$ Z-DNA regions

<table>
<thead>
<tr>
<th></th>
<th>(CG)$_3$</th>
<th>rhGG</th>
<th>rwcGC</th>
<th>rwGT</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Twist (degree/base step)</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>C2G3/C13G14</td>
<td>−9.2</td>
<td>−7.7</td>
<td>−6.5</td>
<td>−7.4</td>
</tr>
<tr>
<td>G3C4/G12C13</td>
<td>−48.9</td>
<td>−47.7</td>
<td>−47.5</td>
<td>−47.5</td>
</tr>
<tr>
<td>C4G5/C11G12</td>
<td>−9.4</td>
<td>−10.4</td>
<td>−10.5</td>
<td>−10.3</td>
</tr>
<tr>
<td>G5C6/G10C11</td>
<td>−50.8</td>
<td>−48.5</td>
<td>−48.0</td>
<td>−47.1</td>
</tr>
<tr>
<td>C6G7/C9G10</td>
<td>−12.2</td>
<td>−9.1</td>
<td>−9.8</td>
<td>−10.6</td>
</tr>
<tr>
<td>Average for CG steps</td>
<td>−10.3 ± 1.7</td>
<td>−9.1 ± 1.3</td>
<td>−8.9 ± 2.1</td>
<td>−9.4 ± 1.8</td>
</tr>
<tr>
<td>Average for GC steps</td>
<td>−49.9 ± 1.3</td>
<td>−48.1 ± 0.6</td>
<td>−47.8 ± 0.4</td>
<td>−47.3 ± 0.3</td>
</tr>
<tr>
<td><strong>Rise (Å/base step)</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>C2G3/C13G14</td>
<td>3.8</td>
<td>3.4</td>
<td>3.4</td>
<td>3.4</td>
</tr>
<tr>
<td>G3C4/G12C13</td>
<td>3.6</td>
<td>3.7</td>
<td>3.7</td>
<td>3.6</td>
</tr>
<tr>
<td>C4G5/C11G12</td>
<td>3.9</td>
<td>4.0</td>
<td>3.9</td>
<td>4.1</td>
</tr>
<tr>
<td>G5C6/G10C11</td>
<td>3.5</td>
<td>3.6</td>
<td>3.6</td>
<td>3.7</td>
</tr>
<tr>
<td>C6G7/C9G10</td>
<td>4.1</td>
<td>3.8</td>
<td>3.9</td>
<td>3.8</td>
</tr>
<tr>
<td>Average for CG steps</td>
<td>3.9 ± 0.2</td>
<td>3.7 ± 0.3</td>
<td>3.7 ± 0.3</td>
<td>3.8 ± 0.3</td>
</tr>
<tr>
<td>Average for GC steps</td>
<td>3.6 ± 0.1</td>
<td>3.7 ± 0.1</td>
<td>3.7 ± 0.1</td>
<td>3.7 ± 0.1</td>
</tr>
<tr>
<td><strong>Propeller twist (degree/base-pair)</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>C2-G14</td>
<td>1.1</td>
<td>−1.2</td>
<td>6.3</td>
<td>−0.7</td>
</tr>
<tr>
<td>G3-C13</td>
<td>3.2</td>
<td>1.3</td>
<td>0.1</td>
<td>2.2</td>
</tr>
<tr>
<td>C4-G12</td>
<td>−0.9</td>
<td>−0.3</td>
<td>−0.7</td>
<td>−0.9</td>
</tr>
<tr>
<td>G5-C11</td>
<td>−1.5</td>
<td>−0.3</td>
<td>−0.9</td>
<td>1.2</td>
</tr>
<tr>
<td>C6-G10</td>
<td>0.5</td>
<td>−0.3</td>
<td>2.2</td>
<td>−3.6</td>
</tr>
<tr>
<td>G7-C9</td>
<td>2.7</td>
<td>3.8</td>
<td>0.1</td>
<td>5.4</td>
</tr>
<tr>
<td><strong>Buckle (degree/base-pair)</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>C2-G14</td>
<td>1.9</td>
<td>10.2</td>
<td>10.0</td>
<td>10.4</td>
</tr>
<tr>
<td>G3-C13</td>
<td>−3.5</td>
<td>−6.1</td>
<td>−3.5</td>
<td>−5.9</td>
</tr>
<tr>
<td>C4-G12</td>
<td>3.0</td>
<td>0.6</td>
<td>0.0</td>
<td>0.1</td>
</tr>
<tr>
<td>G5-C11</td>
<td>−2.4</td>
<td>−1.7</td>
<td>−2.1</td>
<td>−0.7</td>
</tr>
<tr>
<td>C6-G10</td>
<td>2.0</td>
<td>3.3</td>
<td>3.8</td>
<td>8.2</td>
</tr>
<tr>
<td>G7-C9</td>
<td>0.0</td>
<td>−4.7</td>
<td>−1.3</td>
<td>−8.3</td>
</tr>
</tbody>
</table>

Values of twist, rise, propeller twist, and buckle are shown for each of the three structures (rhGG, rwcGC, and rwGT) described in the text and are compared with the 1.0 Å crystal structure of (dCG)$_3$ containing only MgCl$_2$ (Gessner et al., 1989). Morphologies of the 5’-overhanging ends are not shown. Each structure was evaluated using the program NASTE (Nucleic Acid STructure Evaluation), which utilizes a global helix axis to determine each parameter. The rhGG, rwcGC, and rwGT structures were analyzed with the 5’-overhanging removed and with the remaining (dCG)$_3$ duplex juxtaposed to the reference (dCG)$_3$ structure. Residues 2 to 7 and 9 to 14 in the rhGG, rwcGC, and rwGT duplexes, respectively, correspond to residues 1 to 6 and 7 to 12 in (dCG)$_3$.

links together three duplexes and appears to be important in stabilizing the crystal lattice.

In the minor groove, the four central (C-G) base-pairs show a spine of interconnected water molecules. This spine is formed by two water molecules lying nearly in the plane of each base-pair. The disruption of this spine at the terminal base-pairs is associated with the large buckling and the unusual stacking of the paired overhanging nucleotides. The DNA structure and the solvent structure around the six standard Watson-Crick d(G-C) base-pairs, therefore, are very similar to that of standard Z-DNA, with some variations that are specific for the crystal lattice interactions.

Crystal lattice interactions

The duplexes are aligned end-to-end along the crystallographic c-axis, similar to the stacking of Z-DNA hexanucleotides in this space group (Figure 3). In the crystal lattice of standard Z-DNA hexanucleotides, the duplexes stack end-to-end to form quasi-continuous columns along the c-axis. Each adjacent column is staggered by two base-pairs along this axis. In the current heptamer structures, however, the adjacent Z-DNA duplex regions are all exactly aligned. This can be envisioned as a series of discrete stacked sheets of Z-DNA.

The most noticeable feature of the crystal lattice, however, is that in these heptanucleotide sequences, there is a single base overhanging each 5’-end. The structures of the overhanging nucleotides are not identical, even in the homoduplex d(GCGCGCGG)$_2$. In all cases, one of the overhanging guanosine nucleotide residues sits stacked against the duplex, while the other nucleotide residue (whether it is guanosine, cytidine, or thymidine) is extra-helical, extending out and away from the duplex. The extra-helical base pairs with the stacked guanosine residue of an adjacent duplex within each layer, effectively interlinking the Z-DNA duplexes. The extra-helical base also serves to fill the gap between two stacked duplexes. Thus, the lattice consists of layers of duplexes in which each duplex is linked to two adjacent duplexes by pairing the bases that overhang the 5’-ends.

The pairing of extra-helical bases has previously been observed in the crystal structure of the Z-DNA hexamer of d(CCGCGG) (Malinina et al., 1994). In this case, the bases at both ends flip out and form Watson-Crick base-pairs between adjacent duplexes, leaving only four standard base-
pairs as Z-DNA in the center of the hexamer structure. In the lattice of the current heptanucleotides, the reverse base-pairs bring adjacent helices so close together that a direct helix to helix hydrogen bond forms between O2P of G3 and the O3’ of G7 of the duplexes. This hydrogen bond is analogous to the short (2.63 Å) interhelical hydrogen bond observed between the O1P of G2 of one hexamer duplex and the O5’ of G12 in an adjacent hexamer in the crystal structure of the Z-DNA hexamer d(CG)$_3$. (Wang et al., 1979, 1981).

The third base step in each strand of these heptanucleotide structures (G3/C4 and G10/C11) is in the unusual $Z_{II}$ conformation, which has been attributed to crystal packing effects in the crystal structure of d(CG)$_3$. (Wang et al., 1979, 1981). When a base step is in the $Z_{II}$ conformation, the intervening phosphate is rotated outward about 1 Å away from the minor groove. In all three heptamer crystal structures, the phosphates of nucleotides C4 and C11 are not directly hydrogen-bonded to a neighboring duplex or a metal complex, as was observed in the base steps that adopt the $Z_{II}$ conformation in the crystal structures of Z-DNA hexamers and decamers (Gessner et al., 1985; Brennan et al., 1986). Thus, while the $Z_{II}$ conformation in the heptamers may be caused by crystal packing, the lattice interactions causing this distortion remain unclear.

**Structure of the helical stacked guanosine nucleotide**

In all cases, the guanosine residue that remains stacked against the Z-DNA duplex is in the syn conformation, extending the alternating anti-syn pattern of nucleotides from the duplex to include this overhanging nucleotide. The syn conformation of this nucleotide is stabilized by a hydrogen bond between the O5’ oxygen atom of the terminal hydroxyl group to the N2 amino and N3 nitrogen atoms of the guanine base (Figure 2). The Watson-Crick edge of the guanine is subsequently oriented to allow pairing with the intervening base of the extra-helical overhanging nucleotide of an adjacent duplex. Since the interduplex base-pairs bring the 5’-nucleotides of adjacent duplexes together within these layers, the strands held together in this manner are necessarily parallel with each other. These base-pairs are therefore the reverse type, with the d(G·G) overhangs forming reverse Hoogsteen-type base-pairs, the d(G·C) overhangs forming reverse Watson-Crick base-pairs, and d(G·T) overhangs forming reverse wobble base-pairs (Figures 1 and 2).

**Structure of the reverse Hoogsteen d(G·G) base-pair**

The Watson-Crick edge of the stacked guanine faces the Hoogsteen edge of the extra-helical guanosine nucleotide of an adjacent duplex to form a reverse Hoogsteen-type d(G·G) base-pair (rhGG, Figure 2(a)). In this case, although the stacked guanosine is in syn, the extra-helical guanosine base adopts the anti conformation. This is analogous to the two mismatched G(anti)-G(syn) Hoogsteen base-pairs in the structure of d(CGCGAATTCGCC) (Skelly et al., 1993). The hydrogen bonds that hold the base-pair together are shown in Figure 2(a). The N1 to N7 distance is in the range of hydrogen-bond donor to acceptor distances observed in Watson-Crick base-pairs. The N2 to O6 distance, however, is significantly longer than that expected for a standard hydrogen bond. This is a result of the shift of both guanine residues within the plane of the bases. This shift is noticeably when the rhGG base-pair is superimposed on a rhGG base-pair that is part of the r[GGGG] triplet in yeast tRNA$^{	ext{Phe}}$ (Westhof et al., 1988; Figure 4), and is likely the result of an additional hydrogen bond formed between the N2 amino nitrogen of the extra-helical base and the O2P oxygen of cytidine 9 of a third duplex. The base planes of the two guanine residues are almost exactly coplanar as a consequence of being sandwiched between two Watson-Crick base-pairs of the stacked Z-DNA duplexes (Figure 5a). In contrast, the Hoogsteen d(G·G) base-pairs in the antiparallel duplex of d(CGCGAATTCGCC) are propeller twisted about their long axes (Skelly et al., 1993). In the current structure, several water molecules link each overhanging guanine base to neighboring DNA atoms. There is, however, no water molecule that directly bridges the two guanine residues in this rhGG base-pair. The structure of this base-pair is distorted by the crystal lattice and therefore may not represent the type of rhGG base-pair expected.
to form in RNA structures. We will still refer to this as the rhGG base-pair, however, because it does conform to the geometry of this type of base-pair (Figure 1) and, as we will see later, will be a useful reference structure for comparing the stabilities of the rwcGC and rwGT base-pairs.

Structure of the reverse Watson-Crick d(G·C) base-pair

A reverse Watson-Crick base-pair is formed by pairing the stacked guanosine with the extra-helical cytidine of a d(CCGCGCG) strand, with the Watson-Crick edges of their bases facing each other. In this case, the cytidine in anti pairs with the guanosine in syn. In contrast to the normal d(G·C) base-pairs, the resulting rwcGC base-pair is held together by only two hydrogen bonds (from the N1 and N2 of the guanine to O2 and N3 atoms, respectively, of the cytosine base. In addition, a single water molecule was observed to bridge the guanine N2 to the cytosine N4 nitrogen atom, which may provide additional stability to this base-pairing (Figure 5b). When taken together, the reverse Watson-Crick base-pair appeared to have the greatest number of base-base and base-water-base hydrogen bonds of the three reverse base-pairs presented here.

Structure of the reverse wobble d(G·T) base-pair

A reverse wobble base-pair is formed by pairing the stacked guanosine with the extra-helical thymidine of the d(TCGCGCG) strand, with their Watson-Crick edges of their bases facing each other. In this case, both nucleotides are in the syn-configuration. The resulting rwGT base-pair is held together by two hydrogen bonds (from the N1 and O6 atoms of the guanine to the O4 and N3 atoms, respectively, of the thymine base; Figure 3c). In several crystal structures of DNA duplexes containing wobble G·T mismatches, a water molecule or a hydrated magnesium cation bridges the guanine O6 and the thymine O4 (Ho et al., 1985; Hunter et al., 1986). No analogous solvent interaction was observed in the current rwGT base-pair. In fact, very few water molecules were observed around this base-pair (Figure 5(c)). This may be the result of slight positional disorder within the base-pair plane. The average temperature factors for this base-pair are ~twofold higher than the remainder of the DNA, and 20% higher than that of the rhGG and rwcGC base-pairs.

The deoxyribose O5 oxygen atom of the thymidine forms a hydrogen bond to the O2 atom oxygen in the base of cytidine C9 from an adjacent stacked duplex. This intermolecular hydrogen bond apparently stabilizes the syn-conformation of the thymidine nucleotide. In comparison, the deoxyribose O5 of the anti cytidine in the rwGC base-pair does not show this same hydrogen-bonding interaction. As in the case of the rhGG base-pair, the bases in the rwGT base-pair are nearly coplanar. This is similar to the normal wobble d(G·T) base-pair observed in the crystal structure of a Z-DNA hexamer (Ho et al., 1985). On the other hand, the normal wobble d(G·T) and d(G·U) base-pairs in several A-DNA crystal structures (Kneale et al., 1985; Hunter et al., 1986; Vojtechovsky et al., 1985).
Structure and Stability of Reverse Base-pairs

Relative stability of reverse base-pairs

We observed only the rwGT base-pairs in the crystals of the heterogeneous duplexes formed by mixing the sequence d(GCGCGCG) with d(TCGCGCG). To confirm this observation from the crystal structure, the single crystal was redissolved and the DNA strand composition analyzed by MALDI mass spectrometry (Figure 6). The mass spectrum showed that, within experimental error, the crystal was composed of near equal ratios of the two strands. Mass spectra recorded from the four remaining crystals in the crystallization set-up were identical with that of the mounted crystal, indicating that this was not unique to the crystal that we had originally studied. We would expect that mixing these sequences would result in a 1:2:1 ratio of d(GCGCGCG)$_2$, d(GCGCGCG)/d(TCGCGCG), and d(TCGCGCG)$_2$ duplexes in solution. This mixture of homo- and heteroduplexes should also be observed when d(GCGCGCG) is paired with d(CCGCGCG). Again, only the rwGC pairing was observed. Thus, the crystal lattice discriminates between the reverse base-pairs that are formed by the overhanging nucleotides, favoring both rwGT and rwGC over rhGG base-pairs.

In order to gain additional insight into the mechanism for this discrimination and the free energy differences between the rhGG versus either rwGT or rwGC, we studied the crystallization of these duplexes in solutions containing increasing ratios of the d(GCGCGCG) strand (G$_S$) relative to either the d(TCGCGCG) or d(CCGCGCG) strands, redissolved the DNA in the crystals, and quantified the strand composition within the crystals by mass spectrometry. The mass spectra showed equal quantities of d(GCGCGCG) and d(TCGCGCG) when the two strands were added at a 1:1 ratio, but showed only the d(GCGCGCG) strand at ratios $\geq$ 2:1. For the d(CCGCGCG) containing crystals, equal proportions of both strands persisted to a ratio of 3:1. At a 4:1 ratio of d(GCGCGCG) added to d(CCGCGCG), however, the spectrum showed predominantly (>90%) the d(GCGCGCG) strand. Thus, the crystals convert from the heteroduplexes to the homoduplex of d(GCGCGCG)$_2$ as the ratio of d(GCGCGCG) added was increased. The order of stability for the reverse base-pairs can thus be defined as rwcGC > rwGT > rhGG. This was confirmed in an experiment where crystals were grown with all three sequences added in equal proportions. In this case, the mass spectrum of the dissolved crystals showed only the d(GCGCGCG) and d(CCGCGCG) strands, indicating that this was the preferred pairing of the DNA.

To estimate the stability of the rwGC and rwGT base-pairs relative to the rhGG base-pair, we derived a thermodynamic model to simulate these titration results. The sharp transition from hetero- to homoduplexes in the crystals suggests that discrimination between the various reverse base-pairs occurs at a highly cooperative step.

Table 2. Backbone and glycosidic torsion angles and sugar puckers of 5'-overhanging nucleotides

<table>
<thead>
<tr>
<th></th>
<th>rhGG</th>
<th>G8</th>
<th>G1</th>
<th>rwGC</th>
<th>C8</th>
<th>G1</th>
<th>rwGT</th>
<th>T8</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\gamma$ (C5'-C4')</td>
<td>28.01</td>
<td>28.01</td>
<td>28.01</td>
<td>28.01</td>
<td>28.01</td>
<td>28.01</td>
<td>28.01</td>
<td>28.01</td>
</tr>
<tr>
<td>$\delta$ (C4'-C3')</td>
<td>175.43</td>
<td>175.43</td>
<td>175.43</td>
<td>175.43</td>
<td>175.43</td>
<td>175.43</td>
<td>175.43</td>
<td>175.43</td>
</tr>
<tr>
<td>$\epsilon$ (C3'-O3')</td>
<td>-155.20</td>
<td>-155.20</td>
<td>-155.20</td>
<td>-155.20</td>
<td>-155.20</td>
<td>-155.20</td>
<td>-155.20</td>
<td>-155.20</td>
</tr>
<tr>
<td>$\chi$ (C1'-N)</td>
<td>75.88</td>
<td>75.88</td>
<td>75.88</td>
<td>75.88</td>
<td>75.88</td>
<td>75.88</td>
<td>75.88</td>
<td>75.88</td>
</tr>
<tr>
<td>Sugar pucker</td>
<td>C2'-endo</td>
<td>C2'-endo</td>
<td>C2'-endo</td>
<td>C2'-endo</td>
<td>C2'-endo</td>
<td>C2'-endo</td>
<td>C2'-endo</td>
<td>C2'-endo</td>
</tr>
</tbody>
</table>

| Torsion angles (degrees) shown above were calculated using the program Xfit (MacRee, 1992). The conformation of the base relative to the deoxyribose ring is denoted beside $\chi$ angles as either S (syn) or A (anti). |
during crystallization. This is most likely during the nucleating event that initiates the formation of the crystals. In this model, we consider only two different duplexes that can crystallize, the homoduplex (GG) formed by the d(GCGCGCG) strands in solution (G₃) and the heteroduplex (GY) formed by G₃ and Y₃ in solution (where Y represents either the d(GCGCGCG) or d(TCGCGCG) strand). The homoduplexes (YY) are not considered in the model because these have not been observed to crystallize in these studies. Qualitatively, we can think of this mechanism as one in which the initiation step of crystallization is the formation of a lattice in a solution consisting of either the homo- or heteroduplexes of the DNA. Once formed, this lattice allows the addition of duplexes that are identical with those already in the lattice, excluding all others. Thus, discrimination between base-pairs with those already in the lattice, excluding all other species that contribute. Thus, the ratio \( \frac{[GG]}{[GY]} \) is given by equation (3), where \( K_{GGL} \) and \( K_{GGL} \) are the equilibrium constants for formation of the nucleating lattices:

\[
\frac{[GG]}{[GY]} = \frac{[GGS]}{[GY]} = \frac{K_{GGL}}{K_{GGL}} = \frac{[GGS]}{[GY]} \tag{3}
\]

The ratios of the homo- and heteroduplexes in solution are dependent on the ratios of the single-strands added to the solutions (equations (4) to (6)):

\[
2G \leftrightarrow GG, \quad K_{GGS} = [GGS]/[G] \tag{4}
\]

\[
G + Y \leftrightarrow GY, \quad K_{GYS} = [GYS]/[G][Y] \tag{5}
\]

\[
\frac{[GGS]}{[GYS]} = \frac{K_{GGS}[G]^2}{[GYS][G][Y]} \tag{6}
\]

Since the base-pairing and structure of the resulting duplex regions are identical between the homo- and heteroduplexes, we can assume that \( K_{GGS} = K_{GYS} \). Thus, \( \frac{[G]}{[Y]} = \frac{[GGS]_o}{[GYS]_o} \). During the slow nucleation steps, the DNA is predominantly in solution, so that \( \frac{[G]}{[Y]} \) can be assumed to be the ratio of the two strands added to solution. The relative probabilities for initiating the GG₃ and GY₃ lattices are related to the strand compositions of the crystallization setup and to the difference in free energy between the lattices of the two types of duplexes (\( \Delta \Delta G^\circ \)), as in equation (7):

\[
\frac{[GG_{L}]}{[GY_{L}]} = \frac{[GGS]}{[GY]} = \frac{K_{GGL}}{K_{GGL}} = \frac{[GGS]}{[GY]} \tag{7}
\]

The mass spectrometry analyses provide the compositions of the crystals as the types of single-strands and not of DNA duplexes. Thus, the observed quantities are the ratios of the individual strands that are associated with each type of duplex that is potentially found in each crystal. The GG₃ species contributes two strands of d(GCGCGCG) while GY₃ contributes one such strand. The observed quantity of d(GCGCGCG) in the lattice (\( q_{GSS} \)) is thus 2 GG₃ + GY₃. The amount of Y strand observed from the mass spectra (\( q_{YSS} \)) is \( q_{GY} \), since this is the only species that contributes. Thus, the ratio \( \frac{q_{GSS}}{q_{YSS}} = 2[q_{GSS}/[GY_{L}]] + 1 \). The data we obtained showed the complete conversion from the hetero- to the homoduplex; we therefore represent the observed data as the fraction of \( q_{YSS} \) from 0.5 to 0.0. Finally, this can be related to the ratio of the two strands in solution by equation (8):

\[
f_Y = 0.5/\left(\left([G]/[Y]\right)e^{-\Delta \Delta G^\circ/RT} + 1\right) \tag{8}
\]

Using equation (8), we can simulate titration curves for \( [G]/[Y] \) from 1:1 to 5:1 and values for \( n = 1 \) to 16, in which the transition from the heteroduplex \( (f_Y = 0.5) \) to the homoduplex \( (f_Y = 0) \) occurs at \( [G]/[Y] \approx 3.5 \) (Figure 7a). The simulated curve at \( n = 16 \) reproduces the sharp transition observed between the two lattice types (assuming a 10% accuracy in defining \( f_Y \) from the mass spectrometry data). This suggests that the discrimination between rwcGC and rhGG base-pairs in the crystal lattice occurs when a minimum of 16 duplexes (the contents of four complete unit cells) associate to form an initiation complex. The assembly of four complete unit cells in the lattice, therefore, appears to be the defining step for the composition of the crystals. The assembly of a minimum of four unit cells produces an environment in which all the possible intermolecular interactions (within the unit cell and between unit cells) are represented.

With the value of \( n = 16 \), or four unit cells defining the minimum size of the initiation complex, we can now estimate the free energy differences between rhGG and the other two types of base-pairs. In this initiation complex, there is one interduplex base-pair interaction within each unit cell and one between each unit cell. For the most compact assembly, there are a total of four base-pairs within and four between the four unit cells, yielding a total of eight identical base-pairing interactions. If we now consider the total difference in free energy of interaction as the sum of eight individual interactions (\( \Delta \Delta G^\circ / \text{int} \)), the titration curves can be simulated (equation (8); Figure 7b). This assumes that only the differences in the free energy for pair-
Figure 7. Comparison of the fractions of the d(YCGCGCG) strand \( f_Y \) (Y observed quantity of dYCGCGCG) and \( G \) = quantity of dCCCGCGG in the crystal) as the ratio of the d(YCGCGCG) strand relative to the d(YCGCGCG) strand \([G_s]/[Y_s]\) is increased from 1:1 to 4:1 in the crystallization setups, as observed in the crystals and calculated using the model and equation (8) in the text. (a) The fraction of the strand dCCCGCGG \( f_G \) in the crystal of the heteroduplex dYCGCGCG/dCCCGCGG was determined at each \([G_s]/[Y_s]\) ratio by MALDI mass spectrometry (filled circles, with errors approximated at 10%). The simulated curves were calculated with the number of duplexes in the initiation complex \((n)\) set at 1 to 16 in equation (8), and \( \Delta \Delta G^\circ \) set to a value that places the midpoint of the transition at \([G_s]/[Y_s]=3.5\) for each curve. (b) The fraction of the d(YCGCGCG) strand in crystals of the heteroduplexes dYCCCGCGG/d(YCGCGCG) is compared for \( Y = \) cytidine (filled circles) and \( Y = \) thymidine (open boxes). The values for \( f_Y \) were calculated using equation (8), with \( n=16 \) and the difference in free energy between the rhGG and the rwcGC and rwGT reverse base-pairs normalized to each interaction expected within and between four unit cells \( \Delta \Delta G^\circ /\text{interaction} \) set at 0.0, 0.5, 1.0, and 1.5 kcal/mol.

Discussion

We present here the structures of reverse base-pairs formed by pairing the 5'-overhanging nucleotide of d(Gm*CCGCGG) with either the 5'-guanosine of d(Gm*CCGCGG) to form a d(G-C) reverse Hoogsteen, the 5'-thymidine of d(TCCCGG) to form a d(G-T) reverse wobble, or the 5'-cytidine of d(CCCCGGG) to form a d(G-C) reverse Watson-Crick base-pair (Figure 2). The common guanosine nucleotide is stacked against the standard Z-DNA and is relatively inflexible in its conformation (Figure 3). In all three structures, this guanosine is in the syn conformation, extending the alternating anti-syn character of the Z-DNA duplex.

The extra-helical nucleotide, which distinguishes each type of base-pair, adopts either the anti conformation (in rhGG and rwcGC) or the syn conformation (in rwGT). This is likely determined by the requirements for orienting the base so that the proper face is presented to the stacked guanine to form the base-pair. For example, the thymine of the rwGT base-pair must lie towards the major groove of the guanine in order to properly pair its N3 and O4 atoms with the O6 and N1 of the guanine base (Figure 2(c)). This pushes the pyrimidine base closer to the duplex in the lattice and thus requires that the nucleotide adopt the more compact syn conformation (Figure 5(c)). The cytosine and guanine bases of the rwcGC and rhGG base-pairs, on the other hand, are pulled towards the minor groove, and thus can adopt the more extended anti conformation (Figure 2(a) and (b)). Indeed, the rhGG is an extreme case, where the extra-helical guanosine residue is extended to the point where it now has only a single hydrogen bond within the base-pair. A second hydrogen-bonding interaction occurs between the N2 amino nitrogen atom of this extended guanosine and the O2P oxygen atom of cytidine 9 in a third duplex (Figure 5a). This feature of the purine-purine base-pair is not observed in either of the base-pairs formed with the pyrimidine residues.

The structure of the rwcGC base-pair is identical with that of the analogous base-pairs observed in tRNA (Westhof et al., 1988). The root-mean-square (rms) deviation between the atoms of the bases in this structure and that of the tRNA \( ^{\text{Phe}} \) is 0.160 Å. In comparison, the average rms deviation of the d(G-C) Watson-Crick base-pairs within the structure of rwcGC is 0.124 Å, and relative to the d(G-C) Watson-Crick base pair in the structure of B-DNA at 2.5 Å resolution is 0.184 Å. Likewise, the bases of the rwGT base-pair are similar to the G-U reverse wobble base-pair observed in the solution structure of the unusually stable RNA hairpin formed by the sequence r(GGAC(UUCC))GUC (Varani et al., 1991).

In all of these crystal structures, only one well-defined type of reverse base-pair was observed, even though we attempted to solve the structures with all possible combinations of base-pairs. It is easy to rationalize the common stacked guanosine nucleotide, since this is an extension of the highly invariant Z-DNA duplex. It is less obvious, however, why the extra-helical base of the heteroduplexes should always be the pyrimidine, since the pairing two guanine residues of the homoduplex can obviously be accommodated by the same lattice. Mass spectrometry analysis of these heterodu-
plex showed that crystals grown with equal ratios of each strand are composed only of the heteroduplexes. In addition, when looking at the population of crystals, all the crystals analyzed in this way had the same composition. We concluded that the difference must result from the greater stability of the rwGC and rwGT base-pairs over the rhGG in this environment.

Since these base-pairs result from crystal lattice interactions, we were able to estimate the relative stability of each type of base-pair within nearly identical environments by analyzing the compositions of crystals grown with various ratios of the parent strand (d(GCGCGCG)) with the paired pyrimidine strand (d(CGGCGCG) or d(TCGCGCG)). In the case of the rwGC structure, the transition from all heteroduplexes to all homoduplexes in the crystals was highly cooperative. The simplest model for this transition defines the discriminating step as the initial nucleation event of crystal growth. This is not the only interpretation of the cooperativity. However, it seems reasonable that once a particular lattice type is established, the growth of the crystal can proceed only by extending this same lattice. The minimum cooperativity coefficient that fits the data requires that 16 duplexes be involved in the nucleation of the crystal. This is the content of four complete unit cells, suggesting that this is the minimum size of a regular lattice. In this model, all the interactions between molecules within the unit cell and between unit cells are established within this minimum lattice. Thus, once this initial lattice is formed, the structure of the pockets in which the extra-helical bases must fit become defined. These pockets then allow only a single type of duplex to add to the crystal lattice.

The free energies determined here are specific for the base-pairing and lattice interactions observed in these crystals, with the rwGC > rwGT > rhGG base-pairs. All of these form two hydrogen bonds within the lattice, either directly to the stacked guanine, as in the rwGC and rwGT structures, or one with the stacked guanine and one with the backbone of a third duplex as in the rhGG structure. With both the atomic resolution structures and the thermodynamic data in hand, we can ask what accounts for the differences in stability. The rwGC base-pair places the pyrimidine nucleotide in the anti conformation. In addition, water molecules in the plane of the base-pairs accommodate the unfulfilled hydrogen-bonding groups of the bases and thus may contribute to the overall stability of the rwGC basepairing.

The relative instability of the rwGT base-pair likely results from the disfavored syn conformation adopted by this pyrimidine base (Haschemeyer & Rich, 1967; Neumann et al., 1979). Furthermore, no water molecule was observed bridging the guanine and thymine bases, as has been observed in the structures of “normal” d(G·T) wobble base-pairs (Ho et al., 1985; Kneale et al. 1985; Hunter et al., 1986).

The relative instability of the rhGG base-pair is likely associated with the sliding of the extra-helical guanine away from the stacked guanine, leaving only a single distorted hydrogen bond between the two bases. Still, one would expect the hydrogen bond interactions between the N2 amino nitrogen atom and the phosphate group of a third duplex (there is one direct hydrogen bond and a second water molecule-mediated interaction) to compensate for the hydrogen bond lost between the two bases. However, the coordination of three duplexes to form all the observed hydrogen bond interactions may not occur during the nucleation event. It is not clear whether a more standard rhGG base-pair forms at these initial stages, which is then distorted by subsequent lattice interactions. To test these possibilities, we crystallized a reverse d(A·A) base-pair, but its structure is significantly different from the d(G·G) so that they are not comparable at this time. The structure and thermodynamics of a d(G·I) base-pair in this crystal system should resolve this problem. Thus, although the rhGG base-pair observed here is highly distorted by the crystal lattice, and thus may not represent the structure expected in RNA structures, it does serve as a reference for comparing the stability of rwGC and rwGT base-pairs.

Materials and Methods

Synthesis, purification, and crystallization

The seven-base oligonucleotides d(Gm5CGCGCG), d(GCGCGCG), d(TCGCGCG), and d(CCGCGCG) were synthesized using phosphoramidite chemistry on an Applied Biosystems DNA synthesizer in the Center for Gene Research and Biotechnology at Oregon State University. Size exclusion chromatography on a Sephadex G-25 column was used to remove salts, blocking groups, and prematurely terminated oligonucleotides. The oligonucleotides were lyophilized, redissolved in 30 mM sodium cacodylate buffer (pH 7.0), and used for crystallization without further purification. The oligonucleotides d(GCGCGCG) and d(Gm5CGCGCG) produce homoduplexes with two G overhangs. The oligonucleotides d(GCGCGCG) and d(TCGCGCG) were mixed in an equimolar ratio to yield duplexes with G and T overhangs. Likewise, d(GCGCGCG) and d(CCGCGCG) were similarly mixed to produce duplexes with C and G overhangs.

Crystals of the duplexes were grown at room temperature by vapor diffusion in sitting drop setups. All sequences crystallized from initial solutions containing 0.5 mM DNA (single-strands), 50 mM sodium cacodylate (pH 7.0), 1 mM MgCl2, 2.5 mM cobalt hexammine (Aldrich), and 5% (v/v) 2-methyl-2,4-pentanediol (MPD), equilibrated against a reservoir of 17% MPD. Blocky, amber-colored plates appeared within one week and reached dimensions of up to 0.4 mm × 0.4 mm × 0.1 mm within two weeks.

806 Structure and Stability of Reverse Base-pairs
Table 3. Diffraction data from the crystal structures of the sequences d(Gm5CGCGCG)₂, d(GCGCGCG)/d(CCGCGCG), and d(GCGCGCG)/d(TCGCGCG) which crystallized in the space group P2₁2₁2₁. The crystals are represented by their overhangs as rWG, rwGC, and rwGT, respectively.

<table>
<thead>
<tr>
<th>Unit cell dimensions (Å)</th>
<th>rhGG</th>
<th>rwGC</th>
<th>rwGT</th>
</tr>
</thead>
<tbody>
<tr>
<td>a</td>
<td>20.34</td>
<td>20.32</td>
<td>20.28</td>
</tr>
<tr>
<td>b</td>
<td>29.62</td>
<td>29.54</td>
<td>29.41</td>
</tr>
<tr>
<td>c</td>
<td>51.93</td>
<td>51.84</td>
<td>51.89</td>
</tr>
<tr>
<td>Measured reflections</td>
<td>28,961</td>
<td>20,832</td>
<td>5,784</td>
</tr>
<tr>
<td>Unique reflections</td>
<td>5,324</td>
<td>5,340</td>
<td>2,601</td>
</tr>
<tr>
<td>Resolution range (Å)</td>
<td>29.60±1.46</td>
<td>14.20±1.48</td>
<td>14.00±1.88</td>
</tr>
<tr>
<td>R_presp(I) (%)</td>
<td>8.6</td>
<td>8.3</td>
<td>7.2</td>
</tr>
</tbody>
</table>

* R_presp(I) = 100 x (Σ|I(I-1)-(Ihkl)|)/n where I is the integrated intensity of a reflection, (I) is the average of all observations of the reflection and its symmetry equivalents, and n is the number of unique reflections. All positive, non-zero reflections were merged.

X-ray diffraction data collection

X-ray diffraction data for the crystals were collected at room temperature using a Siemens P4 diffractometer with a Siemens Hi-STAR area detector (Cu-Kα radiation from a sealed tube source). The raw data were integrated and scaled using the software package SAINT (Siemens, Inc.). All crystals were isomorphous (in the space group P2₁2₁2₁, with nearly identical unit cell dimensions (Table 3)), and diffracted to high resolution (1.68 to 1.9 Å).

Structure solution and refinement

The structure of d(Gm5CGCGCG) was solved first using features of the diffraction data to construct an appropriate model for molecular replacement. The dimensions of the unit cell suggested that the heptamer was in the Z-DNA form and aligned along the crystallographic c-axis. The space group was the same as that of most previously crystallized Z-DNA hexamers, and the a and b unit cell axial lengths were very similar to those of the archetypal Z-DNA hexamer d(CGCGCG) (Wang et al., 1979, 1981). The length of the c axis (~52 Å) could accommodate 14 base-pairs with a helical rise of 3.7 Å, suggesting that the helical axes of two stacked heptamers were aligned parallel to the c-axis. This was confirmed by the Patterson map, which showed base-pair cross vectors spaced 3.7 Å apart along the c-axis.

The alternating purine-pyrimidine heptamer sequence could pair to form d(Gm5CGCGCG) hexamer duplexes with guanosine residues overhanging the 3'-ends or d(m5CGCGCG) hexamer duplexes with guanosine residues overhanging the 5'-ends. This latter case seemed more likely. Sequences of the type d(CGCGCG) typically crystallize as Z-DNA, while d(Gm5CGCGCG) crystallize as A-DNA (Mooers et al., 1995). This is consistent with studies by Quadrifoglio et al., (1984) showing that short alternating d(GC) sequences, but not alternating d(GC) sequences form Z-DNA in solution. Both possibilities, however, were tested. Models of both types of structures were constructed using the program InsightII (Biosym/MSI Corp.) with standard helical parameters for Z-DNA. The initial 3'-overhang model was generated by removing the cytidine nucleotide at the 5'-end of the duplex structure of d(CGCGCGCG), while the 5'-overhang model was constructed by removing the nucleotide at the 3'-end of the duplex structure of d(GCGCGCGCG).

For each initial model, the best orientation and positioning of the heptamer in the unit cell was located using the rotation and translation search functions of the program AMoRe (Navaza, 1994). The R-values for the best initial solutions were 44.7% for the 5'-overhang model and 47.7% for the 3'-overhang model. Subsequent refinement of these models demonstrated that the 5'-overhang structure was correct. The model was refined to an R-value of 41.6% using the rigid body and rigid parts (with the bases, deoxyribose, and phosphate treated as independent groups) refinement functions in X-PLOR (Brünger, 1992) using a new parameter file for the DNA (Parkinson et al., 1996) and data from 8 to 3.5 Å. After simulated annealing (with a starting temperature of 3000 K) the R-factor was reduced to 30.6% for data from 8 to 2.2 Å.

The actual conformations of the overhanging nucleotides were determined from electron density maps calculated using only the phasing information from the six base-pairs of the d(CGCGCG) duplex. Difference maps generated using XtalView (MacRee, 1992) showed that only one of the 5'-terminal guanine bases was stacked. The other overhanging nucleotide was flipped out and extended so that it base-paired with the stacked overhanging guanine of a neighboring duplex (Figure 2). The refinement converged to a final R of 20.7% (R_mwu = 27.8%) at 1.68 Å resolution, with 49 water molecules added, including one cobalt hexa-ammine and one hydrated magnesium complex. The coordinate error of less than 0.2 Å was estimated from a Luzzati plot (Luzzati, 1952).

The final structures of d(CGCGCGCG) and d(Gm5CGCGCG) were identical in all respects. The structures of the heteroduplexes of d(CGCGCGCG)/d(TCGCGCG) and d(CGCGCGCG)/d(CGCGCGCG) were solved in a similar fashion, using the d(CGCGCG) duplex region of the d(CGCGCGCG) structure as the starting model, and defining the conformations of the overhanging bases from difference maps. The statistics for the refined structures of d(Gm5CGCGCG) and d(CGCGCGCG)/d(TCGCGCG), and d(CGCGCGCG)/d(CGCGCGCG) are listed in Table 4.

In the crystallization solution of the heteroduplexes, homoduplexes of d(CGCGCG) may have been present. As an independent check on the composition of the crystals of the heteroduplexes, four large crystals of the duplex d(CGCGCGCG)/d(TCGCGCG) were isolated, carefully washed with cold crystallization solution lack-
Table 4. Refinement results for the crystal structures of \(d(GmCGCGCG)_2\) (rhGG), \(d(GCGGC\bar{C}CGCGCG)/d(CCGCGCG)\) (rwGC), and \(d(GCGGC\bar{C}CGCGCG)/d(TCGCGCG)\) (rwGT)

<table>
<thead>
<tr>
<th></th>
<th>rhGG</th>
<th>rwGC</th>
<th>rwGT</th>
</tr>
</thead>
<tbody>
<tr>
<td>(R)-working (%)</td>
<td>20.7</td>
<td>20.9</td>
<td>19.1</td>
</tr>
<tr>
<td>(R)-free (%)</td>
<td>28.7</td>
<td>27.2</td>
<td>28.6</td>
</tr>
<tr>
<td>Resolution range (Å)</td>
<td>8–1.68</td>
<td>8–1.80</td>
<td>8–1.90</td>
</tr>
<tr>
<td>Data completeness (%)</td>
<td>82.2</td>
<td>84.5</td>
<td>79.3</td>
</tr>
<tr>
<td>Number of reflections</td>
<td>3538</td>
<td>3127</td>
<td>2230</td>
</tr>
<tr>
<td>No. of non-hydrogen DNA atoms</td>
<td>286</td>
<td>281</td>
<td>283</td>
</tr>
<tr>
<td>No. of water molecules</td>
<td>49</td>
<td>64</td>
<td>47</td>
</tr>
<tr>
<td>Av. B-factors (Å(^2))</td>
<td>15.9</td>
<td>13.3</td>
<td>17.5</td>
</tr>
<tr>
<td>DNA atoms</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Water atoms</td>
<td>31.1</td>
<td>31.5</td>
<td>29.6</td>
</tr>
<tr>
<td>r.m.s. deviation from ideality</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Bond lengths (Å)</td>
<td>0.007</td>
<td>0.008</td>
<td>0.008</td>
</tr>
<tr>
<td>Bond angles (degrees)</td>
<td>1.477</td>
<td>1.304</td>
<td>1.311</td>
</tr>
</tbody>
</table>

*Exclusive of the reflections sequestered in the test set to calculate \(R\)\(_{free}\). Refinements were made with a three sigma on \(F\) cutoff on each dataset.

Each structure was refined with a cobalt hexa-ammine and a hexa-aquomagnesium complex in addition to these water molecules.

**Crystallization and mass spectrometry analyses of crystals grown with different strand compositions**

To estimate the relative stability of the different duplex pairings, crystals were grown from solutions in which \(d(GCGCGCG)\) was mixed with either \(d(CCGCGCG)\) or \(d(TCGCGCG)\) in molar ratios of 1:1, 2:1, 3:1, and 4:1, with the total concentration held constant at 1.4 mM. The crystallization solutions contained the identical buffers, salts, and precipitants as those that yielded the original crystals. After two weeks, crystals were isolated from the setups, washed several times with deionized water, then dissolved into deionized water for analysis by mass spectrometry.

Dialyzed DNA from the dissolved crystals was analyzed by matrix-assisted laser desorption/ionization (MALDI) mass spectrometry using a custom-built time-of-flight instrument, as previously described (Jensen et al., 1993). All samples were analyzed with a matrix of 10 mg/ml of 2,4,6-trihydroxyacetophenone (Aldrich) in a 50 mM diammonium hydrogen citrate/50% (v/v) acetonitrile solution. For each mass analysis, 0.5 μl of DNA analyte was mixed in a 1:2 ratio with the matrix solution and 0.5 μl of this mixture was placed on the sample stage. At the first sign of crystal formation (generally 10 to 15 seconds after deposition when viewed with a stereo microscope), the droplet was gently wiped with a lab tissue, leaving a seed layer of crystallites on the surface of the sample stage. Another 0.5 μl of the analyte/matrix mixture was then deposited on top of the seed layer and then gently rinsed with cold (4 °C) Millipore-filtered water. Each mass spectrum was recorded as the sum of 30 consecutive spectra, each produced by a single pulse of 355 nm photons from a Nd:YAG laser (Spectra Physics). Mass spectra were calibrated using ion-signals from the matrix.

**Acknowledgments**

This work has been supported by grants from the National Science Foundation (MCB 9304467), the National Institutes of Health (R5GM54538A) and the Environmental Health Sciences Center (EHSC) at Oregon State University (NIEHS ES00210). We thank Dr Victor Hsu for sharing his computer resources, Dr Herman Schreuder of Marion Merrell Dow Research Institute for his program \(NASTE\) (Nucleic Acid Structure Evaluation), a program developed in this laboratory for analysis of the helical parameters in \(Z\)-DNA structures. The final coordinates and structure factors for the structures of \(d(GmCGCGCG)_2\) \(d(GCGCGCG)/d(CCGCGCG)\), and \(d(GCGCGCG)/d(TCGCGCG)\) have been deposited in the Nucleic Acid Data Base (Berman et al., 1992). Their reference codes are ZDGB55, ZDG054, and ZDG056, respectively.

**References**


*Edited by I. Tinoco*

(Received 13 February 1997; received in revised form 1 April 1997; accepted 2 April 1997)