Solution Structure of a Pyrimidine·Purine·Pyrimidine Triplex Containing the Sequence-specific Intercalating Non-natural Base D₃

Edmond Wang¹, Karl M. Koshlap¹, Paul Gillespie², Peter B. Dervan² and Juli Feigon¹*

¹Department of Chemistry and Biochemistry and Molecular Biology Institute University of California Los Angeles, CA 90095, USA
²Division of Chemistry and Chemical Engineering California Institute of Technology, Pasadena CA 91125, USA

We have used NMR spectroscopy to study a pyrimidine-purine-pyrimidine DNA triplex containing a non-natural base, 1-(2-deoxy-β-D-ribofuranosyl)-4-(3-benzamido)phenylimidazole (D₃), in the third strand. The D₃ base has been previously shown to specifically recognize T·A and C·G base-pairs via intercalation on the 3' side (with respect to the purine strand) of the target base pair, instead of forming sequence-specific hydrogen bonds. ¹H resonance assignments have been made for the D₃ base and most of the non-loop portion of the triplex. The solution structure of the triplex was calculated using restrained molecular dynamics and complete relaxation matrix refinement. The duplex portion of the triplex has an overall helical structure that is more similar to B-DNA than to A-DNA. The three aromatic rings of the D₃ base stack on the bases of all three strands and mimic a triplet. The conformation of the D₃ base and its sequence specificity are discussed.

Introduction

Canonical nucleic acid triplexes are formed by the binding of a single strand in the major groove of a homopurine-homopyrimidine duplex (Felsenfeld et al., 1957; Sun & Helene, 1993; Moser & Dervan, 1987; Wells et al., 1988). Two general motifs for triplexes exist, parallel and anti-parallel, which are defined by their third strand orientation relative to the homopurine strand in the Watson–Crick paired target duplex. In parallel triplexes, the third strand is homopyrimidine and binds parallel to the homopurine strand of the duplex. A protonated C in the third strand binds to a G in the duplex via Hoogsteen base-pairing forming a C·G·C triplet. Similarly, a T in the third strand recognizes an A in the duplex forming a T·A·T triplet (de los Santos et al., 1989; Le Doan et al., 1987; Moser & Dervan, 1987; Rajagopal & Feigon, 1989a,b). In anti-parallel triplexes, the third strand is homopurine or a mixture of purines and thymines, and binds anti-parallel to the homopurine strand of the duplex. The G·C and A·T base-pairs are recognized by a G and A (or T), respectively, forming G·G·C and A·A·T (or T·A·T) triplets via reverse Hoogsteen base-pairing (Beal & Dervan, 1991; Broitman et al., 1987; Lipsett, 1964; Radhakrishnan et al., 1991).

The possibility of targeting genes with high sequence specificity makes triplexes potentially versatile biochemical tools. Triplexes have been used for gene regulation (Cooney et al., 1988; Degols et al., 1994; Grigoriev et al., 1992; Maher, 1992), DNA isolation and detection (Ji et al., 1994; Olivas & Maher, 1994) and DNA modification (Havre et al., 1993; Luebke & Dervan, 1991). While most DNA-binding proteins are limited to recognition of a few base-pairs of DNA, triplex formation can easily encompass sequences longer than ten base-pairs with a mismatch discrimination comparable to that of duplex formation (Greenberg & Dervan, 1995; Mergny et al., 1991; Roberts & Crothers, 1991). Triplexes could potentially be designed to recognize...
a single site in an entire genome (Strobel & Dervan, 1991; Strobel et al., 1991). Unfortunately, recognition by the third strand is limited to only two of the four possible base-pairs. To some extent a T-A base-pair can be recognized by a G (Griffin & Dervan, 1989) and possibly a C (Belotserkovskii et al., 1990), and a C-G base-pair can be recognized by a T (Yoon et al., 1992), but triplexes containing these non-canonical triplets are less stable. Also, since a C and T already recognize a G-C and A-T base-pair, respectively, sequence specificity is lost when employing such alternative triplets.

The non-natural deoxyribonucleoside, 1-(2-deoxy-β-D-ribofuranosyl)-4-(3-benzamido)phenylimidazole (D₃) (Figure 1(a)) was designed to recognize a C-G base-pair by forming specific hydrogen bonds (Griffin et al., 1992). However, chemical footprinting studies found that when the D₃ base is incorporated into a pyrimidine oligonucleotide it recognizes both T-A and C-G base-pairs (Griffin et al., 1992). Moreover, recognition was found to be strongly dependent on the identity of the 3' neighboring triplet (i.e. 3' with respect to the purine strand) but only weakly dependent on the identity of the 5' neighboring triplet (Griffin et al., 1992; Kiesling et al., 1992). In order to investigate the binding mode of the D₃ base targeted to a Watson–Crick T-A base-pair, the oligonucleotide with the sequence 5'-AGATAGAACCCTCTATTCTATATCTCT-3' (DTA triplex) was synthesized (the bases in the two loops are underlined and the D₃ base and the target T-A base-pair are in bold). This was designed to be similar to sequences that had previously been shown to fold into stable intramolecular triplets (Macaya et al., 1991c, 1992b; Sklenář & Feigon, 1990; Wang et al., 1992). A preliminary study using one- and two-dimensional NMR spectroscopy demonstrated that this D₃-containing oligonucleotide forms an intramolecular triplex, with a CCCC and a TATA loop connecting the Watson–Crick and the Hoogsteen-paired strands, respectively (Koshlap et al., 1993) (Figure 1(b)). This study also revealed that instead of hydrogen-bonding to the Watson–Crick base-pair, the D₃ base intercalates between its target T-A base-pair and the adjacent 3' T-A-T triplet (i.e 3' with respect to the purine strand). Here we present the ¹H NMR assignments and solution structure of the DTA triplex. The structure reveals that intercalation of the D₃ base is readily accommodated and that the D₃ base mimics a triplet. The over-all helical structure is more similar to B-DNA than to A-DNA. The effect of the D₃ base intercalation in a triplex structure is discussed, and the results are compared with those from fiber diffraction and NMR-derived triplex structures.

### Results

1H NOESY spectrum of exchangeable resonances of DTA

The imino region of a NOESY spectrum of the DTA triplex in water at 1°C is shown in Figure 2. This spectrum exhibits features typical of previously characterized intramolecular triplexes, e.g. imino proton resonances corresponding to both Watson–Crick and Hoogsteen hydrogen-bonded base-pairs are observed (Sklenář & Feigon, 1990) (Figure 2(a)). Crosspeaks from the downfield-shifted amino protons, characteristic of protonated cytosine bases, are also observed (Figure 2(b)). Assignment of the spectrum followed protocols previously developed in our laboratory (Macaya et al., 1991, 1992b; Rajagopal & Feigon, 1989a; Sklenář & Feigon, 1990) and are given in Table 1. Analysis of the NOESY spectrum confirms that the DTA oligonucleotide forms an intramolecular triplex. In Figure 2(a), the Watson–Crick imino-imino sequential connectivities are indicated above the diagonal and the Hoogsteen imino-imino sequential connectivities are indicated below the diagonal. There is an unexpected break in the sequential connectivities between T4 and T16. No imino-imino crosspeaks are observed to C’26 due to chemical exchange with water during the mixing time. However, the C’26 imino was readily identified by the crosspeaks to its own amino resonances (Figure 2(b)). T27 and T25 were identified on the basis of some of the intra-triplet NOE crosspeaks characteristic of T-A-T triplets, e.g. both the Watson–Crick and the Hoogsteen T imino protons have crosspeaks to the central A amino resonances (Sklenář & Feigon, 1990). Of particular interest are the crosspeaks observed between imino and D₃ protons. As is indicated in Figure 2(b), the T16 imino proton of the DTA triplex has a crosspeak to the D₃ H3/5C resonance. Crosspeaks are also observed between the T29 imino resonance and the D₃ H5B and H6B protons.
Regions of a two-dimensional $^2$H$_2$O NOESY spectrum of the DT A triplex at 25°C are shown in Figures 3 and 4. Except for the D$_3$ base, assignment of the DT A triplex could largely be done as previously described for intramolecular triplexes using NOESY, PCOSY, HOHAHA and HOENOE spectra (Macaya et al., 1992b). Assignments were obtained for all of the base, H1', H2',2" and H3' resonances in the triplex, except for the nucleotides in the TATA loop and several of the nucleotides in the two triplets at the 5' end (with respect to the purine strand) of the triplex. The assignments are given in Table 1.

The assignments of the D$_3$ base protons were obtained by analysis of the data from PCOSY, NOESY and $^1$H-$^1$C HMQC NMR experiments. In the PCOSY spectrum (Marion & Bax, 1988; Mueller, 1987) of the triplex (data not shown), four unusual crosspeaks are observed in the H1'-H1' region, which are absent in the PCOSY spectra of other triplexes. These crosspeaks form two spin systems of three protons each. In the NOESY spectrum, the same crosspeaks are observed with additional crosspeaks between the outer protons within each spin system (Figure 3). These two sets of spins must be attributed to protons of rings B and C of the D$_3$ base. Protons H2A and H5A were identified by their aromatic-aromatic sequential connectivity to T29H6 (Figure 3, peaks a and b), and confirmed by a weak crosspeak between the two. In the aromatic-methyl region of the NOESY, there is a proton that has a crosspeak to only T4Me and was identified as H2B (Figure 4, peak e). This, as well as other D$_3$ proton resonance assignments, was supported by a $^1$H-$^1$C HMQC spectrum (Bax et al., 1983; data not shown).

If the relative orientation of rings A and B of D$_3$28 were opposite from what is depicted in Figure 1(a), one would also expect to see an NOE crosspeak between protons H2B and H5A, which is not observed. Likewise, this alternative orientation would cause any NOE crosspeaks between H5A and both H6B and H5B to be either very weak or non-existent. Thus, the lack of any NOE crosspeaks other than that of T4Me to H2B, as well as strong crosspeaks between both the central and outer protons of one of the spin systems mentioned above and the resonance identified as either H5A or H2A distinguishes ring B from C, as well as protons H6B from H4B, and H5A from H2A.

The remaining three spin system must be assigned to the C ring. Although the C ring is a five-spin system, resonances for H2C and H6C and for H3C and H5C are degenerate. This is apparently due to motional averaging (ring flips). H3/5C was assigned from the PCOSY spectrum as the central spin in the three-spin system, H4C and H2/6C are distinguished on the basis of their NOESY cross-peaks, i.e. crosspeaks between H4C to both A17H8 and T16H1', and H2/6C to A17H2 (Figure 3, peaks t, m and s).
Table 1. DTA triplex chemical shifts

<table>
<thead>
<tr>
<th>Residue</th>
<th>Imino</th>
<th>Aminos</th>
<th>H6,H8</th>
<th>H2,H5,Me</th>
<th>H1'</th>
<th>H2'</th>
<th>H2''</th>
<th>H3'</th>
</tr>
</thead>
<tbody>
<tr>
<td>A1</td>
<td>—</td>
<td>6.99, 7.73</td>
<td>7.73</td>
<td>—</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>G2</td>
<td>12.69</td>
<td>—</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>A3</td>
<td></td>
<td>7.63, 7.71</td>
<td>7.43</td>
<td>7.56</td>
<td>5.92</td>
<td>1.89</td>
<td>2.51</td>
<td>4.76</td>
</tr>
<tr>
<td>T4</td>
<td>13.34</td>
<td></td>
<td>6.32</td>
<td>0.34</td>
<td>5.61</td>
<td>1.81</td>
<td>2.15</td>
<td>4.70</td>
</tr>
<tr>
<td>A5</td>
<td></td>
<td>7.08, 7.47</td>
<td>7.67</td>
<td>7.14</td>
<td>6.05</td>
<td>2.59</td>
<td>2.90</td>
<td>4.86</td>
</tr>
<tr>
<td>G6</td>
<td>12.46</td>
<td>—</td>
<td>7.20</td>
<td>—</td>
<td>5.76</td>
<td>2.38</td>
<td>2.87</td>
<td>4.85</td>
</tr>
<tr>
<td>A7</td>
<td></td>
<td>7.52, 7.71</td>
<td>7.37</td>
<td>7.35</td>
<td>5.84</td>
<td>2.44</td>
<td>2.72</td>
<td>4.88</td>
</tr>
<tr>
<td>A8</td>
<td></td>
<td>7.52, 7.72</td>
<td>7.77</td>
<td>7.94</td>
<td>6.04</td>
<td>2.44</td>
<td>2.57</td>
<td>4.77</td>
</tr>
<tr>
<td>C9</td>
<td></td>
<td>7.41</td>
<td></td>
<td>5.50</td>
<td>5.75</td>
<td>1.97</td>
<td>2.32</td>
<td>4.66</td>
</tr>
<tr>
<td>C10</td>
<td></td>
<td>7.73</td>
<td></td>
<td>5.97</td>
<td>6.04</td>
<td>2.01</td>
<td>2.36</td>
<td>4.50</td>
</tr>
<tr>
<td>C11</td>
<td></td>
<td>7.55</td>
<td></td>
<td>5.59</td>
<td>5.73</td>
<td>1.88</td>
<td>2.34</td>
<td>4.66</td>
</tr>
<tr>
<td>C12</td>
<td></td>
<td>7.94</td>
<td></td>
<td>6.12</td>
<td>6.21</td>
<td>2.41</td>
<td>2.67</td>
<td>4.73</td>
</tr>
<tr>
<td>T13</td>
<td>14.89</td>
<td>—</td>
<td>7.66</td>
<td>1.79</td>
<td>6.09</td>
<td>2.26</td>
<td>2.71</td>
<td>4.85</td>
</tr>
<tr>
<td>T14</td>
<td>14.09</td>
<td>—</td>
<td>7.53</td>
<td>1.65</td>
<td>6.07</td>
<td>2.28</td>
<td>2.61</td>
<td>4.90</td>
</tr>
<tr>
<td>C15</td>
<td></td>
<td>7.12, 8.18</td>
<td>7.55</td>
<td>5.53</td>
<td>5.92</td>
<td>2.02</td>
<td>2.45</td>
<td>4.74</td>
</tr>
<tr>
<td>T16</td>
<td>13.12</td>
<td>—</td>
<td>7.37</td>
<td>1.63</td>
<td>6.02</td>
<td>2.36</td>
<td>2.51</td>
<td>4.98</td>
</tr>
<tr>
<td>A17</td>
<td></td>
<td>6.45, 6.99</td>
<td>8.30</td>
<td>7.55</td>
<td>5.99</td>
<td>2.65</td>
<td>2.80</td>
<td>4.90</td>
</tr>
<tr>
<td>T18</td>
<td>13.73</td>
<td>—</td>
<td>7.22</td>
<td>1.29</td>
<td>5.92</td>
<td>2.18</td>
<td>2.50</td>
<td>4.78</td>
</tr>
<tr>
<td>C19</td>
<td></td>
<td>6.89, 8.17</td>
<td>7.49</td>
<td>5.45</td>
<td>5.91</td>
<td>2.01</td>
<td>2.46</td>
<td>4.60</td>
</tr>
<tr>
<td>T20</td>
<td>14.45</td>
<td>—</td>
<td></td>
<td>1.53</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>T21</td>
<td></td>
<td>—</td>
<td></td>
<td>1.71</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>A22</td>
<td></td>
<td>—</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>T23</td>
<td></td>
<td>—</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>A24</td>
<td></td>
<td>—</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>T25</td>
<td>13.98</td>
<td>—</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>T27</td>
<td>12.81</td>
<td>9.17, 10.04</td>
<td>7.92</td>
<td>5.91</td>
<td>6.09</td>
<td>2.18</td>
<td>2.66</td>
<td>4.57</td>
</tr>
<tr>
<td>C28</td>
<td>13.07</td>
<td>—</td>
<td>7.65</td>
<td>1.74</td>
<td>6.34</td>
<td>2.33</td>
<td>2.53</td>
<td>5.04</td>
</tr>
<tr>
<td>T29</td>
<td>13.34</td>
<td>—</td>
<td>7.31</td>
<td>1.64</td>
<td>6.16</td>
<td>2.26</td>
<td>2.59</td>
<td>4.92</td>
</tr>
<tr>
<td>C30</td>
<td>13.37</td>
<td>9.38, 9.76</td>
<td>7.86</td>
<td>5.71</td>
<td>6.12</td>
<td>2.20</td>
<td>2.67</td>
<td>4.73</td>
</tr>
<tr>
<td>T31</td>
<td>13.37</td>
<td>—</td>
<td>7.70</td>
<td>1.70</td>
<td>6.18</td>
<td>2.27</td>
<td>2.69</td>
<td>4.87</td>
</tr>
<tr>
<td>T32</td>
<td>12.51</td>
<td>—</td>
<td>7.53</td>
<td>1.59</td>
<td>6.33</td>
<td>2.25</td>
<td>2.30</td>
<td>4.55</td>
</tr>
</tbody>
</table>

D28 aromatic resonances

<p>| | | | | | | | | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>H2A</td>
<td>8.22</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>H2B</td>
<td>7.17</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>H2/C</td>
<td>7.08</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>H5A</td>
<td>7.45</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>H4B</td>
<td>6.22</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>H3/C</td>
<td>6.46</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>H5B</td>
<td>6.55</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>H4C</td>
<td>6.16</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>H6B</td>
<td>6.88</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Chemical shifts for the exchangeable and non-exchangeable resonances are at 1°C and 25°C, respectively. Not all resonances in the loops and in the two 5'-terminal triplets could be assigned due to chemical exchange broadening.

* Not applicable.

Evidence for D3 intercalation

Numerous NOE crosspeaks are observed between protons from all three aromatic rings of the D3 base and other nucleotides in the DTA triplex. In all cases, the observed NOE crosspeaks are to protons of either the T29-A5-T16 triplet or the T4-A17 base-pair. No crosspeaks are seen between the D3 base protons and those of the T27-A3-T18 triplet. Specifically, crosspeaks a through c involve the A ring of D328, crosspeaks d through h involve the B ring and crosspeaks i through m involve the C ring. Just as in the H1', H5-aromatic region, each ring primarily has NOEs to a different strand of the triplex. The A ring has crosspeaks to only T29 (third strand), the B ring has crosspeaks to only T4 (Watson–Crick purine strand) and the C ring has crosspeaks to only T16 (Watson–Crick pyrimidine strand). The one exception is the T4Me, which has crosspeaks to both A and B rings. All the T4Me and the T4H6 resonances have large upfield shifts compared to the usual triplex TMe and TH6 chemical shifts, as is illustrated by the position of their crosspeak (Figure 4, peak n); for comparison, the crosspeak
between T18Me and T18H6 is also labeled (Figure 4, peak o).

Just as in the imino-imino region of the H2O NOESY spectrum, some disruptions in the sequential connectivities are observed. No base-H1' or base-H2',H2'' sequential connectivities are observed between D328 and T27, between T4 and A5, or between A17 and T16 (Figure 3).

NOE restraints

A total of 728 distance restraints were used for the refinement steps prior to relaxation matrix refinement. Of these, 62 are hydrogen-bond restraints, 408 are derived from H2O NOESY spectra, and 126 are repulsive restraints derived from predicted H2O NOESY crosspeak volumes. The distribution of distance restraints by residue is shown in Figure 5(a). In Figure 5(b), the distance restraints are classified as intra-residue, sequential and inter-residue (involving residues other than neighboring residues) restraints. For relaxation matrix refinement, the H2O and repulsive restraints were converted to volume restraints at all five NOESY mixing times yielding 2670 volume restraints. The remaining 194 water and hydrogen-bond distance restraints were left as distance restraints.

Torsion angle restraints

Torsion angle restraints for the sugar puckerings and δ backbone angles were calculated from H1' to H2',H2'' crosspeaks in a H2O HeteroCOSY spectrum (data not shown). The residues that had sufficiently resolved crosspeaks to obtain reliable coupling constants using CHEOPS are indicated in Figure 5. The peak fits had correlation coefficients that ranged from 90 to 96%.

To restrain the ε backbone angle, a 31P-1H H2O heteroCOSY was collected (data not shown) from which the J3P,H3' couplings were determined to be no greater than 14 Hz, indicating that the ε dihedral angles were not in the +gauche conformation (Blommers et al., 1991; Roongta et al., 1990). All residues were restrained to exclude the +gauche conformation for the ε angle.

We are aware that there has been some recent
debate concerning the validity of fitting coupling constants to a two-state sugar pucker model (Harbison, 1993; Zhu et al., 1994). However, most of the measurable sugar puckers were almost exclusively S-type, with the exception of T27, which is predominantly N-type.

**DTA triplex structure**

Figure 6 shows several views of the refined DTA triplex structures. The seven lowest energy structures are shown superimposed in Figure 6(a). The average RMS deviations from ideal geometry for the bonds, angles and impropers are 0.008 Å, 1.592° and 0.276°, respectively, indicating that these structures have very good covalent geometry. The loops have few restraints (Figure 5) and adopt many different conformations (not shown). The terminal triplets are strongly perturbed by the conformation of the loops, which artificially increases the variability of these triplets. In addition, the two triplets at the 5' end of the triplex (with respect to the purine strand) are only partially assigned (Table 1) leading to many fewer restraints and greater conformational freedom. For these reasons, only the atoms in this "core" of the triplex (triplets 3 to 7) were used to calculate the superposition of structures. These effects are also apparent in the RMSD of the individual residues (Figure 7). The loops and the terminal triplets have a larger RMSD compared to the core of the triplex. The average pairwise RMSD and the average $R_{1/6}$-factor (Brünger, 1992) for the core of the DTA triplex are 1.3 (±0.19) Å and 3.9 (±0.1)% respectively. The number of NOE distance violations greater than 0.2 Å varied from one to three per structure and the number of dihedral angle violations greater than 5° varied from four to seven per structure.

The backbone angles, glycosidic angles, sugar puckers and angles specific to the D$_3$ base are shown in Figure 8 for all seven lowest energy structures. Most torsion angles fall into the standard range of conformations (Saenger, 1984). However, several angles consistently adopt unusual conformations in many of the refined structures: the glycosidic angle, $\chi$, in residues A5, A17 and T27, prefers the high-anti conformation; C19, C+26 and T27 all have N-type sugar puckers; and the T27 $b$ angle prefers the + gauche conformation.

The helix parameters are shown in Figure 9 as a function of residue or base step. The X-displacement is very uniform with good precision, although the precision decreases towards the 5' end (with respect to the purine strand) of the triplex due to substantially fewer restraints on the two 5'-terminal triplets. The axial rise is also fairly uniform except for the T4-A5 base step, where the rise more than doubles as a result of the D$_3$ intercalation. The helical twist exhibits the greatest stepwise variation of all the helix parameters and also has the lowest
precision, indicating that it is the least well-defined helical parameter. Again, the 5' end of the triplex displays the most variation.

The conformation of the D₃ base is best illustrated by the superposition of the D₃ base from the seven lowest energy structures (Figure 6(c)). The C ring and peptide group are clearly not as restricted in their conformation as the rest of the D₃ base. The major conformation is depicted schematically in Figure 1(a).

The stacking of the D₃28 base on the neighboring T₁6·A₅·T₂₉ triplet and T₄·A₁₇ base-pair is illustrated in Figure 10. The sequence specificity of D₃ intercalation is also addressed, where the T₄·A₁₇ base-pair has been isosterically replaced with the three other possible Watson–Crick base-pairs, and the T₁₆·A₅·T₂₉ triplet has been isosterically replaced with a C·G·C triplet. The D₃ base stacks with greater overlap and the peptide group is more centered over the polar groups between the bases, when the target base-pair is a T·A or C·G base-pair instead of a A·T or G·C base-pair. The D₃ base stacking is illustrated from a different view in Figure 6(b), highlighting its similarity to a triplet.

Discussion

Intramolecular triplex formation

The data presented here demonstrate that the DTA sequence forms a stable intramolecular triplex. The NOESY spectrum of the exchangeable resonances (Figure 2) strongly resembles those of other triplexes (Sklenár & Feigon, 1990). Fifteen sharp, well-defined resonances are observed in the spectral region where hydrogen-bonded imino protons are found. These have been assigned to the imino protons of the eight Watson–Crick base-pairs and the T₁₆·A₅·T₂₉ triplet. Further evidence for triplex formation is indicated by the presence of C' amino resonances, which are characteristically shifted downfield. With the exceptions discussed above, sequential imino-imino NOE connectivities may be traced, as is seen in Figure 2(a).
Figure 6. Solution structures of the DTA triplex. The Watson–Crick purine strand is yellow, the Watson–Crick pyrimidine strand is red and the third strand is blue. (a) Stereo view of the seven lowest energy structures looking into the groove formed by the Watson–Crick pyrimidine strand and the third strand. The D₃ base is green. Loop residues are not shown. (b) Similar view as in (a) but of a single structure to illustrate how the D₃ base mimics a triplet. Rings A, B and C of the D₃ base are blue, yellow and red, respectively, and the base of the D₃ base is shown as filled sticks. (c) View of the D₃ base of the seven lowest energy structures looking into the major groove of the duplex portion of the triplex. The rings are colored as in (b). The superpositions involved all heavy atoms shown in each Figure except for the terminal triplets, the C₂6·G₂·C₁₉ triplet and the D₃·28 peptidyl group and C-ring.
Helix morphology

The DTA triplex forms a regular helix that, in many ways, is more similar to B-DNA than A-DNA. This can be seen in the axial rise, which shows little variation except at the T4-A17-A5-T16 base step, where the rise more than doubles to accommodate D3 intercalation (Figure 9(a)). If the intercalation step is ignored, the average rise is 3.2 \( \text{Å} \), which is significantly closer to the rise in B-DNA (3.4 \( \text{Å} \)) than in A-DNA (2.6 \( \text{Å} \)).

In contrast to the axial rise, the helical twist of the DTA triplex is less regular (Figure 9(b)) and resembles A-DNA more than B-DNA. The helical twist displays greater variation because only the \( \epsilon \) and \( \delta \) backbone angles are restrained. In addition, the initial base step (with respect to the purine strand) is especially variable for several reasons:

1. it is the terminal triplet that tends to be less stable;
2. the terminal triplet has few resonance assignments (Table 1) and, thus, few restraints;
3. even though T20 and T25 are connected by a loop, the loop is unassignable and therefore unrestrained during refinement, so it adopts a wide range of conformations, which strongly perturbs the conformation of the initial triplet. A similar effect is seen for the second base step, but to a lesser extent.

The intercalation site also displays substantial variability, which is related to the multiple conformations of the C ring of the D3 base (discussed below). If we ignore the A1-T20-G2-C19 terminal step and the T4-A17-A5-T16 intercalation step, then the average twist is 31°, which, unlike the axial rise, is more similar to the twist in A-DNA (33°) than in B-DNA (36°).

The X-displacement of the Watson–Crick base-
Figure 9. Helical parameters as a function of base step or residue, averaged over the seven lowest energy structures. Error bars indicate the standard deviation. The average value for the DTA triplex along with the values for A- and B-DNA are indicated. (a) Axial rise. The average value was calculated omitting the T4·A17-A5·T16 base step. (b) Helical twist. The average value was calculated omitting the A1·T20-G2·C19 and T4·A17-A5·T16 base steps. (c) X-displacement.

The sugar puckers of the DTA triplex are more similar to those in B-DNA than in A-DNA. With the exception of three residues in the third strand, the sugars in the core of the DTA triplex prefer an S-type pucker (Figure 8).

Comparison with other triplex structures

Taken together, the helix parameters indicate that the DTA triplex has a morphology more like B-DNA than A-DNA. This is in contrast to early fiber diffraction studies on poly(dT)·poly(dA)·poly(dT) which concluded that the helical morphology was similar to that of A-form DNA (Arnott & Selsing, 1974). More recently, fiber diffraction studies on two different sequences have suggested that the diffraction patterns are more characteristic of B-form DNA than A-form DNA (Liu et al., 1994). In addition, NMR-derived structures of intramolecular pyrimidine-purine-pyrimidine triplexes containing a non-canonical G·T·A triplet at a single site (GTA triplex) (Radhakrishnan & Patel, 1994b; Wang et al., 1992) and a non-canonical T·C·G triplet at a single site (TCG triplex) (Radhakrishnan & Patel, 1994a) have been solved. A comparison of the helix parameters is given in Table 2. In general, the triplexes have similar helix parameters. The greatest difference is found in the X-displacement; the Arnott fiber diffraction structure has a large X-displacement, while the DTA, GTA and TCG triplexes all have significantly smaller X-displacements. As a result, the Arnott fiber diffraction structure has a significantly deeper major groove than the solution structures. Similarly, the DTA triplex has the shallowest major groove because it has the smallest X-displacement of all the triplex structures. An earlier NOE-derived model structure of a triplex, similar in sequence to the DTA triplex but containing only canonical T·A·T and C+·G·C triplets, also has a similar rise (3.1 Å), twist (32°) and X-displacement (−2.2 Å) (Macaya et al., 1992b).

It is worth noting that the NMR-derived triplex structures were refined by different methodologies. The GTA and TCG structures were both refined from A- and B-DNA starting models. In contrast, the DTA structure presented here was refined using distance geometry, which samples a larger area of conformational space (and results in a larger RMSD between structures). Although these structures all have similar helical parameters, they do have significant differences. It is unclear to what extent the refinement method contributes to these differences.

The sugars in triplexes were originally thought to be C3′-endo because Arnott’s fiber diffraction structure was very similar to A-DNA (Arnott & Selsing, 1974). However, NMR studies of intramolecular triplexes have shown that the sugars are predominantly S-type (Macaya et al., 1991, 1992a,b; Radhakrishnan & Patel, 1992). The DTA triplex and the GTA and TCG triplexes (Radhakrishnan & Patel, 1994a,b) also have predominantly S-type sugar puckers. C2′-endo sugar
puckers are also claimed in the more recent fiber diffraction results (Liu et al., 1994).

In summary, the general morphology of the triplexes falls into a distinct family with the following characteristics: (1) an axial rise very similar to that in B-DNA; (2) a low helical twist that is similar to that in A-DNA; (3) an X-displacement that is intermediate between that in A and B-DNA, but generally closer to B-DNA; and (4) a predominance of S-type sugar puckers as in B-DNA. These characteristics suggest that triplexes resemble B-DNA more than A-DNA.

D₃ intercalation

As previously reported, D₃ intercalates on the 3' side (relative to the purine strand) of its target base-pair rather than forming sequence-specific hydrogen-bond interactions (Koshlap et al., 1993). Disruptions in the sequential connectivities in the imino region (Figure 2(a)) and the aromatic to H1’ region (Figure 3) of NOESY spectra indicated intercalation (Koshlap et al., 1993). The intercalation explains the strong sequence dependence on the 3’ neighboring triplet and the relative insensitivity to the 5’ neighboring triplet observed by Kiessling et al., (1992).

The D₃ base is readily accommodated into the triplex, although the intercalation perturbs the helix in several ways. In comparison to our previous studies on a similar triplex that does not contain the D₃ base (Macaya et al., 1992b; Sklenař & Feigon, 1990), the resonances for the T25·A1·T20 and the C+26·G2·C19 triplets and the T A T A loop residues are much broader, even to the point of being unobservable, indicating that intercalation of the D₃ base destabilizes the 5’ end of the triplex. Another effect of D₃ intercalation is an increase in the axial rise with a concomitant unwinding of the helix at the intercalation site (Figure 9(b)). In fact, the low

Table 2. Helix parameters

<table>
<thead>
<tr>
<th>Structure*</th>
<th>Rise (Å)</th>
<th>Twist (°)</th>
<th>X-disp (Å)</th>
<th>Sugar pucker</th>
</tr>
</thead>
<tbody>
<tr>
<td>DTA</td>
<td>3.2ᵇ</td>
<td>31⁴</td>
<td>-1.4</td>
<td>Generally S-type</td>
</tr>
<tr>
<td>GTA</td>
<td>3.4</td>
<td>31</td>
<td>-1.9</td>
<td>Generally S-type</td>
</tr>
<tr>
<td>TCG</td>
<td>3.4</td>
<td>32</td>
<td>-2.1</td>
<td>Generally S-type</td>
</tr>
<tr>
<td>Liu (TAT)</td>
<td>3.3</td>
<td>28</td>
<td>N/Aᵇ</td>
<td>S-type⁹</td>
</tr>
<tr>
<td>Liu (hetero)</td>
<td>3.2</td>
<td>30</td>
<td>N/Aᵇ</td>
<td>S-type⁹</td>
</tr>
<tr>
<td>Arnott</td>
<td>3.3</td>
<td>30</td>
<td>-3.6</td>
<td>Assumed N-type</td>
</tr>
<tr>
<td>B-DNA</td>
<td>3.4</td>
<td>36</td>
<td>-0.7</td>
<td>S-type</td>
</tr>
<tr>
<td>A-DNA</td>
<td>2.6</td>
<td>33</td>
<td>-5.4</td>
<td>N-type</td>
</tr>
</tbody>
</table>

*DTA refers to the DTA triplex structure. GTA and TCG refer to the two NMR-derived triplex structures containing a non-canonical triplet (Radhakrishnan & Patel, 1994a,b). Liu (TAT) and Liu (hetero) refer to the two DNA sequences used in fiber diffraction studies (Liu et al., 1994). Arnott refers to the original fiber diffraction structure (Arnott & Selsing, 1974). B-DNA and A-DNA refer to the standard DNA duplex conformations (Saenger, 1984).

ᵇ The axial rise for the DTA triplex is calculated omitting the T4·A17·G2·T16 base step.
⁴ The helical twist for the DTA triplex is calculated omitting the A1·T20·G2·C19 and T4·A17·A5·T16 base steps.
ᵇ N/A, not available.
⁹ The sugar conformation is claimed by the authors, but no evidence is given (Liu et al., 1994).
twist observed at the A3-T4 step is even smaller if one considers that the D3 base acts like another base, so that the A3-T4 base step is really a dual A3-D3:28-T4 base step. Similar effects are seen in intercalating drug–DNA complexes (Addess et al., 1992; Gao & Patel, 1988; Gilbert et al., 1989).

Other effects of D3 intercalation are manifested in unusual torsion angles and sugar puckers (Figure 8). Backbone restraints were only applied to some of the δ angles and loosely applied to the ε angles. As a result, most backbone angles are only indirectly determined and the precision of the backbone angles varies from well defined to almost random (Figure 8). Nevertheless, most angles are predominantly within normal ranges (Saenger, 1984). In several of the structures, the T27 β angle is in the +gauche range instead of the usual trans conformation. The +gauche conformation places T27 in a more lateral position (with respect to the global helical axis) from C‘26. Also, the T27 χ angle is very close to the high-anti conformation, which effectively rotates the T27 sugar and phosphate closer to C‘26. The sugar puckers of T27 is also unusual in that it adopts an N-type conformation while most other residues are S-type. This N-type sugar pucker creates a large T27H6 to T27H3' crosspeak in the NOE spectrum (Figure 3, peak v) and a “reversed” coupling pattern in the T27H1' to T27H2',2” crosspeaks in the PCOSY spectrum (not shown). These structural changes are a consequence of the D328 base skipping over the T4·A17 base-pair to intercalate, thereby pulling T27 out of plane with A3-D328-T4 base step. Similar effects are seen in so that the A3-T4 base step is really a dual A3-D3:28-T4 base step. There are other unusual torsion angles in the loop residues and the terminal triplets, but they are probably not due to D3 intercalation. The loop residues do not necessarily conform to standard A- or B-DNA-like angles, and the terminal triplets are only partially assigned (Table 1), have many fewer restraints (Figure 5), and are strongly perturbed by the conformation of the loops.

D3 base conformation

The conformation of the D3 base is important in determining the interactions between the D3 base and its neighboring bases. Figure 6(c) shows the superposition of the D3 base from the seven lowest energy structures. The orientation of rings A and B are well defined by their numerous NOE interactions with the T4, A5, T16, A17 and T29 sugars and with each other (Figures 3 and 4). The χ angle adopts an anti conformation (Figure 8; where the N1A, C2A and H2A atoms are analogous to the N9, C8 and H8 atoms in a purine) that fixes the orientation of the A ring parallel to its neighboring bases and points the B ring into the major groove. In this configuration, the position of rings A and B are similar to those of a third strand pyrimidine and its Hoogsteen base-paired residue in the Watson–Crick purine strand, respectively.

The θ1 torsion angle (Figure 1(a)), which connects rings A and B, determines the rotational position of the B ring and, thus, the position of the peptide group and the C ring. In all seven structures, the θ1 torsion angle positions the B ring to be co-planar with the A ring and points the peptide group and C ring towards the Watson–Crick pyrimidine strand (Figures 6(c) and 8). Neither the rings nor the peptide group were restrained to be co-planar. This configuration places rings B and C in a similar position to a Watson–Crick base-pair.

The θ2 torsion angle adopts several orientations (Figures 6(c) and 8) but predominantly favors the orientation with the carbonyl group pointing out of the major groove. The θ2 torsion angle affects the position of the C ring, but only by ~2 Å at most. In theory, the conformation of the peptide group could be determined by NOE interactions involving the amide proton, but the amide resonance was not observed due to rapid chemical exchange with water. This also suggests that the amide proton is not involved in a hydrogen bond. Thus, the orientation of the peptide group is only indirectly determined by restraints to rings B and C.

The θ3 torsion angle is the most variable and least well defined in the D3 base (Figure 8). The chemical shift degeneracy of the C ring protons indicates that the chemical shifts are motionally averaged and that the C ring is rotating fast on the NMR time scale. Interestingly, significant conformational variation of the C ring is also observed in the refined structures (Figures 6(c) and 8). However, since the NOE
restraints used during the refinement are also degenerate, it is unclear whether the variability of the C ring seen in the calculated structures reflects actual rotation of the C ring or if it simply reflects the degeneracy of the restraints. The large standard deviation in the axial rise at the T4-A5 base step (Figure 9(a)) is a direct consequence of the conformational variability of the C ring. Even if the rise is halved (to account for the D3 base), the rise is still large (~4.3 Å) due to those structures in which the C ring is in a perpendicular (unstacked) conformation.

Stacking of the D3 base

As previously reported, the stacking of the D3 base causes upfield ring current shifts of ~1 ppm for the T4H6 and T4Me relative to other TH6-TMe crosspeaks (Figure 4 and Table 1) (Koshlap et al., 1993). This effect can be rationalized by examining Figure 10(a), where the stacking of the D3 base in relationship to the neighboring T4-A17 base-pair and T29-A5-T16 triplet is shown. The B ring (and partially the A ring) stacks directly over the T4Me and T4H6, which induces an upfield shift.

The most prominent feature of the stacking is the relative disorder of the D3 base position. Compared to the neighboring base-pair and triplet, the position of the D3 base is less well defined. In fact, the D3 base has the largest RMSD of any residue in the core of the triplex and is also the only residue (other than the loops) that has a greater RMSD of its base than its backbone (Figure 7). The major contribution to the RMSD comes from the peptide group and the C ring, as already discussed. The D3 base generally has some overlap with the T29-A5-T16 triplet and substantial overlap with the T4-A17 base-pair (Figure 10(a) and (e)). This may explain why the stability of D3-containing triplexes is strongly affected by the sequence of the adjacent base-pair, but are less affected by the sequence of the adjacent triplet (Kiessling et al., 1992).

An interesting result is that the D3 base mimics a triplet with each ring assuming the position of a base in each strand (Figure 6(b)). From the 3' end of the triplex (with respect to the purine strand), the first four triplets form a regular stacking pattern with the D3 base mimicking a fifth triplet. Then there is large unwinding of the helix between the D3 base and the adjacent T4-A17 base-pair (Figure 9(b)), so the mimicry is not perfect. After the T4-A17 base-pair, the helix continues with a regular twist again (excluding the A1-G2 step, which is ill defined). As discussed above, the unwinding at the T4-A5 step is more severe than is indicated in Figure 9(b), since the step is really two steps of T4-D328 and D328-A5. In fact, rings B and C (which mimic a Watson-Crick base-pair) have almost 0° twist relative to the T4-A17 base-pair. The unwinding is expected because the helical rise must increase to accommodate the D3 intercalation.

The D3 base superimposes on a triplet best when the angles θ1 and θ2 are both in the conformation shown in Figure 1(a), which is also the major conformation of the seven lowest energy structures. In such a superposition, the glycosidic bond of the D3 base is rotated away from the purine strand relative to a canonical triplet. This is precisely the effect that the N-type sugar pucker of T27 has on D28, as discussed above. These observations suggest that it may be possible to design a base that confers greater stability to the triplex by designing it to mimic a triplet more closely.

Sequence specificity

The sequence specificity of D3 intercalation appears to be derived primarily from stacking interactions between the D3 base and its neighboring bases. Although stacking interactions are difficult to determine, the D3 base does have slightly more overlap with T·A and C·G base-pairs (good targets for the D3 base) than it does with A·T and G·C base-pairs (poor targets for the D3 base; Figure 10). Sequence-specific hydrogen bond formation cannot be unambiguously ruled out, although the three chemical groups of the D3 base that can potentially form hydrogen bonds (the amide proton, the carbonyl oxygen and the N3A imidazole nitrogen) do not show evidence of hydrogen bonding. The amide proton is exchanging rapidly, which is inconsistent with hydrogen bond formation. The imidazole N3A nitrogen faces outward into the major groove with no nearby hydrogen bond acceptors (Figure 10). In addition, the only available acceptors for the peptide amide proton are the A5 and A17N1, the A5N7 and the T4O2, all of which are out of plane with the peptide group and could form only weak hydrogen bonds. Similarly, the only hydrogen bond donors near the peptide carbonyl group are the A5 and A17 amino groups and the T4 and T16 imino protons, which would all yield weak hydrogen bonds.

Electrostatic interactions may also be important, although the exact nature of the interaction is not clear. For example, in the major groove, the peptide polar groups are near several polar groups of the T·A base-pair (Figure 10(a)). However, in the major groove of a G·C base-pair (a poor target for the D3 base) the same polar groups are present in a similar arrangement (Figure 10(d)). Similarly, the major groove of a C·G and on A·T base-pair resemble one another, although a C·G base-pair is a good target for the D3 base and an A·T base-pair is a poor target. In an effort to elucidate further the determinants of sequence specificity, we are currently solving the solution structure of the D3 base targeted to a C·G base-pair.

D3 intercalation strongly prefers a neighboring T·A·T triplet over a C·G·C triplet. Comparison of the D3 base stacked over a T·A·T and a C·G·C triplet does not present any obvious reason for this preference (Figure 10(e) and (f)). However, the preference for a neighboring T·A·T triplet may be more general than just for the D3 base, since the canonical triplets also prefer neighboring T·A·T
triplets over C·G·C triplets (Kiessling et al., 1992). This preference may be an effect of the positive charge in a C·G·C triplet.

Intercalation occurs on the 3’ side of the target base-pair instead of the 5’ side. This is due to the right-handedness of the triplex. If one observes the helix in Figure 6(b), the third strand proceeds 5’ to 3’ right-handedness of the triplex. If one observes the triple helical charge in a C +·G·C triplet.

Conclusions

In summary, a pyrimidine-purine-pyrimidine triplex containing a D3 base forms a stable structure resembling other triplex structures, despite having to accommodate D3 intercalation. The stability is partially due to the ability of the D3 base to imitate a triplex, though imperfectly. This suggests that greater stability may be achieved by designing a base to more closely mimic a triplex. Pyrimidine-purine-pyrimidine triplets seem to have several features in common, including an axial rise of \( \approx 3.2 \) to \( 3.4 \) Å, a low helical twist of \( \approx 30 \) to \( 32^\circ \), an X-displacement of \( \approx -1 \) to \( -3 \) Å, and a predominance of S-type sugar puckers. Since triplets conform to neither A- nor B-DNA exclusively, they might best be considered a new form of DNA, T-form DNA, a separate, distinct family of structures.

Materials and Methods

Sample preparation

The phosphoramidite derivative of the artificial nucleoside, D3, was prepared as described (Griffin et al., 1992). The DNA oligonucleotides were synthesized on an ABI 380B DNA synthesizer. The oligonucleotides were deprotected with concentrated ammonia at 55°C for 24 hours, and subsequently desalted on a Sephadex G10 column. FPLC purification (D2F on) was performed using a gradient of 0 to 40% acetonitrile/100 mM triethylammonium acetate over 40 minutes. Detritylation was accomplished with 80% acetic acid for 30 minutes. Final FPLC purification (D2F-off) used a gradient of 0 to 27% acetonitrile/100 mM triethylammonium acetate over 30 minutes. To remove the triethylammonium acetate, the sample was run over a BioRad AG 50W-X8 ion-exchange column, followed by chromatography on a Sephadex G15 column. The appropriate fractions were then pooled and dried by lyophilization. The final conditions of the NMR samples were 100 mM NaCl, 5 mM MgCl2, and \( \approx 2 \) mM DNA in 400 \( \mu \)l of either 99.996\% \( ^2 \)H2O or 90\% \( ^2 \)H2O/10\% \( ^2 \)H2O; the pH was adjusted to 5.2 with NaOH/HCl.

NMR spectroscopy

All NMR experiments were performed on a General Electric GN500 spectrometer. Phase-sensitive nuclear Overhauser effect (NOESY) spectra (Kumar et al., 1980; Macura & Ernst, 1980) were obtained employing the method of States et al. (1982). For spectra in \( ^2 \)H2O, the residual \( ^2 \)H2O peak was suppressed by irradiation during the recycle delay. Water suppression for samples in \( ^2 \)H2O was achieved by using a 1H-spin echo read pulse (11Echo; Sklenák & Bax, 1987). The carrier was centered at the water resonance, and the imino and amino resonances were maximized by setting t = 50 ms. Further elimination of the residual water was accomplished during processing using a time-domain convolution routine (Marion et al., 1989).

For NOE restraints involving non-exchangeable resonances, a set of \( ^2 \)H2O NOESY spectra was collected at 25°C with mixing times of 40, 80, 120, 160 and 200 ms without removing the sample from the magnet. All NOESY spectra in the series were collected and processed identically except for the mixing time. The acquisition parameters were a sweep width of 5000 Hz in both t2 and t1, 350 points in t2, 350 points in t1, 32 scans per t1 block, and a recycle delay of two seconds. The t2 apodization was a 450 point, 45° phase shifted, squared sinebell with a first-order polynomial baseline flattening and the t1 apodization was a 350 point, 45° phase shifted, squared sinebell with a second-order polynomial baseline flattening. The resulting datasets were zero filled to 4096 by 4096 points in t2 and t1, 350 points in t1, 32 scans per t1 block, and a recycle delay of 1.8 seconds. Using the technique of Marion & Bax (1988), a one-dimensional reference spectrum was collected to subtract the dispersive component of the diagonal. The same parameters as the PCOSY were used except that 1356 scans were collected and the acquisition time was doubled. To insure that the total relaxation delay and \( ^2 \)H2O pre-saturation between scans was the same in both the reference and PCOSY spectrum, an additional delay equal to the acquisition time was added to the end of the PCOSY pulse sequence. The t2 apodization was a 700 point, 35° phase shifted, squared sinebell; and the t1 apodization was a 550 point, 45° phase shifted, squared sinebell. Selective regions of the spectrum were processed and zero filled to a resolution of 0.82 point/Hz for crosspeak pattern fitting.
NOE restraints

Crosspeak volumes from the $^1$H-$^3$P heteroCOSY spectrum were fit using CHEOPS (Macaya et al., 1992a; Schultze & Feigon, unpublished program). The dihedral angles $\nu_1$ and $\nu_2$ were determined from these coupling constants by fitting them to a two-state model for the sugar pucker with the program PSEUROT (de Leeuw & Schultze, 1983) and using the dihedral angle. The dihedrals were restrained with an error of $\pm 5^\circ$. These two dihedral restraints fix the backbone torsion angle $\delta$. Twelve residues were resolved enough to determine coupling constants, leading to 24 dihedral angle restraints. The $\epsilon$ dihedrals were derived from a $^1$H-$^2$H heteroCOSY spectrum from which it was determined that no $^2$H$_{25}$ couplings were greater than 14 Hz, which excludes the +gauche conformation for the $\epsilon$ dihedral angle. Therefore, the $\epsilon$ angle was restrained to $-120 (\pm 10)^\circ$ for all 31 phosphate groups. Twenty three residues from the core of the triplex and the CCC loop had their $\chi$ angle loosely restrained to the anti conformation ($-120 (\pm 90)^\circ$). The $\chi$ angles for G2, the T25-A1-T20 triplet, and the TAT loop had not been determined, since the conformation of their glycosidic bond could not be determined from their base to H1' crosspeaks. The $\chi$ angle for D328 was also not restrained so as not to bias the conformation of the D3 base. To maintain proper chirality, an additional dihedral angle restraint was applied to each chiral center, C1’, C3’ and C4’, and to each prochiral center, C2’, C5’ and P.

Structure determination

Structure calculations were performed on Digital Equipment Corporation Alpha workstations using XPLOR V3.1 (Brünger, 1992). The structure determination is composed of an initial refinement, which essentially follows the protocols outlined in the XPLOR V3.1 manual, with only minor changes, followed by an iterative, full relaxation matrix refinement. Iterative relaxation matrix refinement involves back-calculating the NOE volumes from the structures produced by the initial XPLOR protocols and refining these back calculated volumes against the observed volumes (Yip & Case, 1989). After one round of relaxation matrix refinement, those crosspeaks that are observed in the back-calculated spectra but are clearly not observed in the NOESY spectra are entered as repulsive restraints and the structures are re-calculated. Several rounds of this procedure are done until no further repulsive terms are found.

The XPLOR protocols essentially consist of distance geometry embedding of the structure consistent with the NOE distance and dihedral angle restraints, regularizaton of the structure, further refinement using simulated annealing (SA) with a simplified VDW energy term and a final refinement with a long SA protocol and a full VDW energy term. In addition, we supplemented these protocols with a final relaxation matrix refinement. The following changes to the protocols were made: (1) distance geometry sub-embedding (dg_sub_embed.inp), which used atoms N1A, C1B, N and C1C of the D3 base for sub-embedding (Figure 1(a)); 100 structures were calculated; (2) regularization (dgsa.inp), which decreased the time step to 2.5 fs to prevent wildly oscillating structures during dynamics, but increased the number of steps to 2000 to maintain the same total simulation time; 100 structures were calculated; (3) refinement (refine.inp), which decreased the time step to 2.0 fs and increased the number of steps to 5000 during dynamics; 100 structures were calculated; (4) full VDW refinement (refine_gentle.inp), which increased the temperature during dynamics to 500 K. Included hydro-
gen bond energy term and used the planarity restraint to keep the individual bases and the rings of the D₃ base flat; the 40 lowest energy structures from the previous stage of refinement were refined; and (5) relaxation matrix refinement, in which volumes from all five NOESY mixing times were used. The non-exchangeable and repulsive NOE distance restraints were replaced with the corresponding volume integrals. The exchangeable (and hydrogen bond) NOEs were left as distance restraints. The NOE energy scale was increased to 100. Isotropic tumbling was assumed with a correlation time of 5 ns. The NOE integrals were scaled to the same four crosspeaks as in the distance refinement. The relaxation energy parameters were set to cutoff = all, value = 6.0, potential = well, eexp = 1/6, eexp = 2, expo = 2, weight = 0.5. This is essentially a Vol¹/³ weighting of the integrals (Brünger, 1992; James et al., 1991). All other parameters are the same as in the full VDW refinement. The overall relaxation matrix refinement protocol is: (1) 100 steps of energy minimization with relaxation energy calibration turned on for the relaxation energy term; (2) 1000 steps of dynamics with a time step of 1.0 fs at 300 K with temperature bath coupling and auto calibration turned on for the relaxation energy term; (3) the dynamics trajectory is averaged over the last half of the dynamics run taken every 10 steps to obtain an average structure; and (4) 200 steps of energy minimization done in two 100-step phases with relaxation matrix calibration done after each phase. The 20 lowest energy structures from the previous stage of refinement were refined.

Structure analysis

The program CURVES (Lavery & Sklenar, 1988) was used to calculate the helical parameters in all the structures. Only the Watson–Crick duplex portion of the triplex was used in these calculations. XPLOR V3.1 was used to measure all the backbone angles and sugar puckers.

Acknowledgements

This work was supported by NIH grant R01 GM37254 to J.F., and an Office of Naval Research grant to P.B.D. The coordinates for the structures have been deposited with the Brookhaven Protein Data Base (accession no. 1WAN).

References


Solution Structure of a D₃-containing DNA Triplex

*Edited by P. E. Wright*

(Received 21 November 1995; received in revised form 17 January 1996; accepted 22 January 1996)