Global structure of a DNA three-way junction by solution NMR: towards prediction of 3H fold

Bin Wu, Frederic Girard, Bernd van Buuren¹, Jürgen Schleucher², Marco Tessari and Sybren Wijmenga*  

Department of Physical Chemistry/Biophysical Chemistry, University of Nijmegen, Toernooiveld 1 6225 ED Nijmegen, The Netherlands, ¹Unilever Research and Development, Olivier van Noortlaan 120, 3133AT Vlaardingen, The Netherlands and ²Department of Medical Biochemistry and Biophysics, Umeå University, S-90187 Umeå, Sweden

Received April 12, 2004; Revised and Accepted May 19, 2004

ABSTRACT

Three-way junctions (3H) are the simplest and most commonly occurring branched nucleic acids. They consist of three double helical arms (A to C), connected at the junction point, with or without a number of unpaired bases in one or more of the three different strands. Three-way junctions with two unpaired bases in one strand (3HS₂) have a high tendency to adopt either of two alternative stacked conformations in which two of the three arms A, B and C are coaxially stacked, i.e. A/B-stacked or A/C-stacked. Empirical stacking rules, which successfully predict for DNA 3HS₂ A/B-stacking preference from sequence, have been extended to A/C-stacked conformations. Three novel DNA 3HS₂ sequences were designed to test the validity of these extended stacking rules and their conformational behavior was studied by solution NMR. All three show the predicted A/C-stacking preference even in the absence of multivalent cations. The stacking preference for both classes of DNA 3HS₂ can thus be predicted from sequence. The high-resolution NMR solution structure for one of the stacked 3HS₂ is also reported. It shows a well-defined local and global structure defined by an extensive set of classical NMR restraints and residual dipolar couplings. Analysis of its global conformation and that of other representatives of the 3H family, shows that the relative orientations of the stacked and non-stacked arms, are restricted to narrow regions of conformational space, which can be understood from geometric considerations. Together, these findings open up the possibility of full prediction of 3HS₂ conformation (stacking and global fold) directly from sequence.

INTRODUCTION

Three-way junctions [3H according to IUPAC nomenclature (1)] are the simplest and most commonly occurring branched nucleic acids. They consist of three double helical arms, which are connected at the junction point (2,3). They can be classified according to the number of unpaired bases at the junction point, i.e. 3HSₙ indicates n unpaired bases in one strand. 2HSₙ indicates n and m unpaired bases in two different strands, whereas HSₙHSₚ indicates n, m and p unpaired bases in three different strands (1,2). The 3H shown in Figure 1A has, at the junction point, one strand with two unpaired bases and can be classified consequently as a 3HS₂. An alternative classification scheme, which uniquely defines 3Hs in terms of their base sequence arrangement around the junction, was devised by Altona (4,5). For further definition of relevant terms, refer to Figure 1 and its legend.

Three-way junctions are present in the RNA as structural and functional motifs (2,3), where they play important roles in cellular processes, such as translation (6) or splicing (7). In DNA, 3Hs arise during recombination involving phages (8,9). DNA 3Hs have also been proposed to occur in the expansions of triplet repeats found in genetically unstable genomic DNA associated with human diseases, such as for instance Huntington’s (10–13). Finally, DNA 3Hs are very attractive building blocks for the design of nanoscale assemblies because of their branched structural nature and inducible conformational versatility (14–16).

Because of this broad variety of functional activities, there is a wide interest in the structural studies of 3Hs. The fully base-paired 3Hs (3HS₀) only form the open unstacked conformation (Figure 1A, middle; disregarding the two X residues) (2). In contrast, 3HSₙ are expected to be able to adopt—apart from the open conformation (Figure 1A, middle)—also two alternative stacked conformations in which two arms are coaxially stacked, designated as A/B-stacked or conformer I (Figure 1A, left) and A/C-stacked or conformer II (Figure 1A, right) (2,17,18). The B/C-stacked form is expected to be less stable, because of the bulge at the junction in the continuous strand between the two coaxially stacked helices (2,17,18). Biophysical studies have provided considerable insight into the relative stabilities of these 3HSₙ conformations (2,19,20).

First, the stacked conformations were found at least in some cases to require multivalent cations (18,20). This is not surprising as nucleic acids are highly negatively charged. At low salt concentration—leading to a lowering of the charge neutralization—electrostatic repulsions between phosphate...
groups at the junction would tend to keep the structure extended and thus, favor the open conformation (2). Second, the sequence around the junction plays a critical role. 3HS\(_n\) (\(n > 0\)) conformations have been found to exhibit a high tendency for coaxial stacking, and in most cases, one conformer is predominant (2,19). Coaxial helix-helix stacking is a common structural feature of nucleic acid junctions and provides additional stabilization (2). However, the relative stacking interaction strength between the base pairs at the junction motif does not fully account for the observed stacking preference (21). Interestingly, it was also found that the nature of the penultimate base pair plays an important role in the

Figure 1. Schematic and definitions of the DNA three-way junction with two unpaired bases in one strand [3HS\(_2\) according to IUBMB definition (1)] and illustration of empirical stacking rules. (A) Notation and definitions according to the one proposed by Altona (4,5). The 3HS\(_2\) is presented in its open form as shown (center) with the arrowheads pointing to the 3' terminal and the base pairs symbolized by a ladder motif. The strands are numbered clockwise from 1 to 3, with strand 2 containing the two unpaired bases X-X. The double helical arms are clockwise designated as A, B and C (circled capital letters). The two unpaired bases X-X in strand 2 serve as spacers that make coaxial stacking possible. The open form (center) can then fold either into the A/B-stacked conformation (conformer I) on the left or into an A/C-stacked conformation (conformer II) on the right. Conformers I (A/B-stacked) and II (A/C-stacked) can be classified as parallel or anti-parallel depending on the relative orientation of the stacked and non-stacked arms (1,2). Residues important for conformer selection (19) are highlighted in gray: loop (X-X), junction (N1-N4), and arm A penultimate position (‘*’ and ‘#’, blocked). To predict stacking from sequence, two (extended) empirical stacking rules have been proposed [(19), see text]: (i) ‘Pyrimidine’ rule: pyrimidine (Py) at position ‘#’ favors conformer I (A/B-stacked), whereas Py at position ‘*’ favors conformer II (A/C-stacked). (ii) Extended ‘loop’ rule: a stable 5'-N\(_1\)/XXN\(_2\) quasi-hairpin loop (‘/’ indicates break in the backbone) favors conformer I (A/B-stacked), whereas a stable 5'-N\(_1\)/XXN\(_2\) quasi-hairpin loop favors conformer II (A/C-stacked). (B) Application of the extended stacking rules to TWJ11. For conformer II (A/C-stacked), the residue at the penultimate position of arm A, in the crossover strand [position ‘*’ (A) right], is a pyrimidine (C\(_{16}\)). And, the pseudo-hairpin loop formed is then 5'-C\(_3\)G\(_{27}\)T\(_{28}\)G\(_{29}\)/C\(_{18}\)-3', i.e. a stable CTT/G loop. Moreover, the conformer I (A/B-stacked) is disfavored as the pyrimidine rule is not respected (G\(_9\) is wrongly positioned) and the quasi-hairpin loop formed has the sequence G/TTC (5'-G\(_7\)/T\(_{28}\)T\(_{29}\)C\(_{30}\)-3', i.e. unstable. Consequently, TWJ11 is predicted to fold into the A/C-stacked conformation (conformer II). (C) The three novel sequences designed to test the extension of the 3HS\(_2\) stacking rules. TWJ10 and TWJ12 are, like TWJ11, designed to fold into an A/C-stacked conformation according to the folding rules (crucial positions indicated in gray-blocked and gray).
conformational selection process (21). Moreover, in the coaxially stacked conformation of a 3HS2, the junction loop can fold like a hairpin loop (19) (for exact definition see below). Inspired by these observations, van Buuren et al. (19) derived, from the reported stacking behavior of 24 published 3HS2, two empirical rules that aim to predict 3HS2 stacking preference from its sequence.

The two rules are illustrated in Figure 1A together with the strand and arm definitions. The first rule—the ‘pyrimidine rule’—states that in a stacked conformation, it is energetically advantageous when, in arm A, a pyrimidine (either C or T) is located in the crossover strand at the penultimate position (Figure 1A, motif * or #). The second rule—the ‘loop rule’—states that a thermodynamically stable (low-free energy) quasi-hairpin loop in the A/B-stacking mode will favor A/B-stacking whereas a thermodynamically unstable one will not. The physical reason for the importance of the penultimate base pair and thus for the ‘pyrimidine rule’ is still not well understood. On the other hand, the ‘loop rule’ is well understood from the observation (19) that in the A/B-stacked 3HS2 conformation the junction loop folds into a quasi-tetraloop; a quasi-tetraloop is defined as a tetraloop in which the first and second loop residues are not covalently attached. The extended loop rule does not make a statement about junction loop stability for A/C-stacking conformers. Interestingly, some 3HS2, within this structure set, do adopt an A/C-stacked conformation, but only when multivalent cations are present. This raises the question whether there would be a fundamental difference in stability between A/B- and A/C-stacked conformers. For several examples, the junction sequence is such that both rules favor an A/B-stacked conformation. The A/B-stacked conformation was then indeed found to be highly preferred even in the absence of multivalent cations (17,19). On the other hand, the loop rule does not make a statement about junction loop stability for A/C-stacking conformers. To achieve full A/C-stacking, all three sequences were chosen such that both the pyrimidine—and the loop rule favor an A/C-stacked conformation, i.e. a pyrimidine in the * location in arm and a PyNN/Pu-loop in the A/C-stacked conformation (Figure 1). As compared to TWJ11, which folds into an A/B-stacked conformation (19), in TWJ11 only the penultimate C–G base pair in arm A was switched (‘pyrimidine’ rule) and in arms B and C, the C–G base pairs ultimate to the junction (‘loop’ rule) (gray base pairs in TWJ11, Figure 1B, middle). To test whether indeed the penultimate position in arm B does not affect folding, the Pu–Py positions in the arm B penultimate base pair were switched going from TWJ10 to TWJ11. TWJ12 as TWJ11, except that in arm C penultimate and junction base pairs have been exchanged to test that stacking stability across junction (A/C-stacked: C30:G7//G17:C8 in TWJ11 versus T30:A7/G17:C8 in TWJ12, Figure 1B and C) does not have a major effect on folding. We report their stacking preference and also one of their high-resolution NMR solution structure (TWJ11, Figure 1C, middle). Its local structure—including the junction loop conformation—is well defined by an extensive set of classical NMR restraints [nuclear Overhauser effect (NOEs) and J-couplings]. Its global structure—relative arms orientation—is well defined by employing residual dipolar couplings restraints. The global structure of TWJ11 is analyzed and compared with that of other published 3H structures.

MATERIALS AND METHODS

DNA samples

DNA oligomers, such as TWJ10, TWJ11 and TWJ12 (Figure 1C) were chemically synthesized on solid phase (5 μmol scale column), high-performance liquid chromatography-purified and concentrate by re-precipitation in ethanol (DNA Technology A/S, oligo@dna-technology.dk, Aarhus, Denmark). They were dialyzed subsequently in 3 kDa Centricon centrifuge tubes against dDQ-water and freeze-dried. For each oligonucleotide, two NMR samples were prepared: one in 100% D2O and one in 95:5 H2O/D2O. The DNA samples were re-suspended in the desired solvent—containing 0.1 mM EDTA and 50 mM NaCl—to a final concentration of 1.5 mM. The pH was adjusted to 6.5 with 1 M DCI. DSS was used as a chemical shift internal reference. An extra D2O sample (0.1 mM EDTA, 50 mM NaCl, pH 6.5) was prepared in anisotropic conditions by adding a filamentous Pf1 bacteriophage solution to a TWJ11 D2O sample (24,25) to a final
concentration of 15 mg ml$^{-1}$. Solution of Pf1 in D$_2$O was purchased from ASLA Ltd (ASLA, Riga, Latvia). The NMR sample phage concentration was derived by monitoring the solvent deuterium splitting (26).

**NMR spectroscopy**

All NMR experiments were recorded on a Bruker spectrometer operating at a proton frequency of 600 MHz and equipped with a Cryoprobe$^{	ext{R}}$. Homonuclear two-dimensional (2D)-NMR experiments were acquired in phase-sensitive mode using TPPI (27). 2D-NOE spectroscopy (NOESY) in D$_2$O were recorded at 25°C (mixing delay: 30, 50, 80,120, 200 and 500 ms) using a spectral width of 5952 Hz in both dimensions, with 4096 and 600 complex points in the direct ($t_2$) and indirect dimensions ($t_1$). In H$_2$O, a 2D-NOESY spectra with mixing time of 250 ms was recorded at 5°C with a spectral width of 12 626 Hz in both dimensions and with 4096 and 400 complex points along $t_2$ and $t_1$ dimensions, respectively; water suppression was achieved by using Watergate (28). 2D TOCSY spectra were acquired using a DIPSI type spin-lock of 68 ms in length (29). In addition, standard (H,H) DQF-COSY and (31P,H)HMBC were recorded (30). Residual dipolar couplings (RDC) were collected from the Pf1 sample (30 C 1.0, 5H5, 1.5C 6H6, 1.5C 8H8, 22 to 37 Hz and have values range from 22 to 37 Hz and have experimental error of 1.5 Hz.

**NMR restraints**

All spectra were processed with XWINNMR (Bruker) and analyzed using the Triad software program (Sybyl package of Tripos Inc.). From the build-up series of D$_2$O 2D-NOESY spectra, a total of 509 non-exchangeable $^1$H$^-$-$^1$H distances were derived from the NOE intensities, using the full relaxation matrix algorithm implemented in Mardigras (32). Distance derived from the NOE intensities, using the full relaxation matrix algorithm implemented in Mardigras (32). Distance derivation was left floating but its axial and rhombic components were allowed to re-fold under the influence of the NOE and dihedral restraints. During the initial temperature increase from 300 to 20 000 K, the NOE force constant was gradually increased from 5 to 150 kcal Å$^{-2}$ mol$^{-1}$, whereas the dihedral term was kept small at 5 kcal deg$^{-2}$ mol$^{-1}$. These values were maintained during the subsequent cooling to 1000 K. The final cooling period to 300 K and subsequent short equilibration step was carried out via Verlet restrained molecular dynamics. During these steps, the force constants were scaled to their final values of 50 kcal Å$^{-2}$ mol$^{-1}$ and 200 kcal deg$^{-2}$ mol$^{-1}$, respectively. The protocol concludes with 1000 steps of Powell minimization.

(i) The TAD protocol, essentially similar to the one originally published (35), was slightly modified as described previously (19,36,37). Briefly, 100 extended structures were allowed to re-fold under the influence of the NOE and dihedral restraints. During the initial temperature increase from 300 to 20 000 K, the NOE force constant was gradually increased from 5 to 150 kcal Å$^{-2}$ mol$^{-1}$, whereas the dihedral term was kept small at 5 kcal deg$^{-2}$ mol$^{-1}$. These values were maintained during the subsequent cooling to 1000 K. The final cooling period to 300 K and subsequent short equilibration step was carried out via Verlet restrained molecular dynamics. During these steps, the force constants were scaled to their final values of 50 kcal Å$^{-2}$ mol$^{-1}$ and 200 kcal deg$^{-2}$ mol$^{-1}$, respectively. The protocol concludes with 1000 steps of Powell minimization.

(ii) The first 30 lowest energy structures (NOE-based) were considered for RDC refinement. Only RDCs of C-H vectors from well-defined B-type helical regions were selected as restraints for the RDC refinement (i.e. 46 RDC among 70, Table 2), to prevent scaling problems of RDCs due to internal dynamics of residues outside canonical helix elements. The individual geometry of each arm, defined by residues 9–16 (arm A), 19–26 (arm B), 1–6/31–36 (arm C), was kept fixed, whereas the junction and junction-loop residues (7/8/17/30 and 27/28/29/18, respectively) were allowed to re-arrange during the RDC refinement protocol. Thus, only the relative arm orientations were affected by the additional RDC-restraints and flexibility around the branch point allowed the reorientation of the arms to occur. This approach has strong likeness to the global refinement strategy applied by Pardi et al. [e.g. (38) and Blackledge et al. (39)]. This strategy was applied in a TAD-based protocol similar to the one described above (vide infra), except that the dipolar coupling energy term was turned on. The time steps were set to 15 ps. Initially, the temperature was raised to 4000 K and the system let to equilibrate during 4000 steps with force constants of 5 kcal Å$^{-2}$ mol$^{-1}$ (NOE), 2 kcal deg$^{-2}$ mol$^{-1}$ (Dihedral) and 0.001 kcal Hz$^{-2}$ mol$^{-1}$ (RDC). The force constants for NOE and RDC were then—over 20 000 steps—raised to their final values of 50 kcal Å$^{-2}$ mol$^{-1}$ and 0.2 kcal Hz$^{-2}$ mol$^{-1}$. After an equilibration period of another 4000 steps, the system was cooled to 300 K in 20 000 steps. The protocol ends then with 1000 steps of Powell minimization.

During the RDC refinement the alignment tensor’s orientation was left floating but its axial and rhombic components were kept fixed. The axial component of the alignment tensor $A_{ax} = A_{ax} = -1/2(A_{xx} + A_{yy})$ and its rhombicity $R = A_{xx}/A_{ax}$ with $A_{ax} = A_{xx} - A_{yy}$, were found to be equal to 19 Hz and 0.6, respectively, as determined from the distribution of RDC values (31,40); these values were separately confirmed via a grid search approach (38) (58).
Structure analysis
Structural statistics were derived for the 12 lowest energy structures. Restrictions violations were derived with Xplor 3.851 (35), root mean square deviations (RMSDs) were calculated with Molmol 2.6 (41), and DNA helical parameters were analyzed with Curves 5.3 (32,42,43). In addition, RMSDs between observed and back-calculated chemical shifts in the final structure ensemble were calculated with Nuchemics (34,44). The Euler angles defining the global 3H conformation were calculated using an ‘in house’ Matlab script (Matlab 6.1, the Math Works Inc.). Structures were displayed using either ViewerLite (Accelrys Inc.) or VMD (45). The 12 lowest energy structures derived from the RDC-refinement have been deposited with the Protein Data Bank (accession code 1SNJ).

RESULTS AND DISCUSSION
To verify experimentally that the 3HS2 stacking rules (19) can be extended to A/C-type stacking (vide infra), three new 3HS2 sequences were designed (TWJ10, TWJ11 and TWJ12, Figure 1C) and their solution structures investigated. To the H1 and 13C resonance assignments were obtained using standard methods (30). Resonance assignments of each of the three 3HS2 sequences are listed in the supplementary materials section (S1, S2 and S3, respectively). The complete sequential walk between H6/8 and H1/H5 resonances is also presented for all three sequences (S4, S5 and S6). Here, we detail the NMR evidence collected and used to define TWJ11 solution conformation, and further extend to TWJ10 and TWJ12.

TWJ11 folds into a three-way junction
Characteristic NOE connections in the 2D-NOESY spectra (in H2O and D2O), combined with complete 1H and 13C resonance assignment, provided direct evidence that TWJ11 indeed folds into a 3H. The H2O NOESY of TWJ11 showed three thymine imino resonances (T1, T6 and T25; Figure 2C). Each of them displays NOE contacts to two neighboring guanine residues and shows an intense NOE crosspeak to the H2 of its base-paired adenine (data not shown). Thus, three AT base pairs are formed and sandwiched between GC base pairs, as expected from the folded sequence (Figure 1B). Moreover, in the D2O NOESY at 25°C, each of the three H2 resonances is correlated to the H1’ of the paired thymidine (Figure 2A). The H1’ resonance assignment follows from the complete sequential assignment of the non-exchangeable protons (S5). The sequence-specific assignment of the D2O NOESY further shows that two 5’-CTTG-3’ hairpin loops are formed and that they cap the two helical arms. Their characteristic NOE patterns and 1H chemical shifts demonstrate that they both adopt the stable H2-type tetraloop fold, i.e. the tetraloop fold in which the second loop residue is folded into the minor groove and the third is stacked onto the base pair closing the loop (23,46–48). In conclusion, TWJ11 folds into a stable three-way junction, in which arms B and C are both capped with an H2-type 5’-CTTG-3’ hairpin loop.

TWJ11 adopts an A/C-stacked conformation
The A/C-stacking preference is supported by imino-imino NOEs and 17 characteristic B-DNA type inter-residue NOEs of non-exchangeable protons at the branch point (S7). Nine of them define the base stacking in the continuous strand (G7/C9), eight determine the base stacking in the crossover region (G17/C19). The sequential C9H2 to G17H7, C30H2 to G17H7 and C38H2 to G17H7 connectivities defining the A/C-stacking are illustrated in Figure 2A and B. No NOE contacts between the arm B and the A/C-stacked arm were found.

The junction loop of TWJ11 folds into a quasi H2-type CTTG loop
Interestingly, apart from the 5’-CTTG-3’ loops capping the arms A and B, an extra H2-type loop fold is observed namely for the junction loop (5’-C27T28T29/G18-3’). Its presence is evidenced by the highly characteristic NOE patterns and chemical shifts displayed by the residues involved (46,47). The latter is demonstrated by comparison of the conformational shifts (44) of the junction loop residues with respect to those of the classical H2-type CTTG loop (Table 1). As can be seen, the conformational shifts have a particularly small deviation (RMSD = 0.12 ppm) and correlate nearly perfectly (R2 = 0.98). In the standard CTTG loops—capping the arms—these parameters are even better. The slightly higher deviation and lower correlation, observed for the junction loop, is likely to be related to the break in the phosphate backbone (5’-CTTG-3’ versus 5’-CTTG-3’). Therefore, based on all these observations, the junction loop (5’-C27T28T29/G18-3’) could be safely assumed to fold as an H2-type hairpin loop, called quasi H2-type hairpin loop here, because of the break in the backbone.

TWJ10 and TWJ12 also adopt an A/C-stacked conformation with a quasi H2-type CTTG junction loop
The NMR spectra of sequences TWJ10 and TWJ12 were assigned and analyzed using the same methods as for TWJ11 (S1, S3, S4 and S6). Similar to TWJ11, the sequences TWJ10 and TWJ12 adopt an A/C-stacked conformation: NOE

<table>
<thead>
<tr>
<th>Residue</th>
<th>H6</th>
<th>H1’</th>
<th>CH3</th>
<th>H6</th>
<th>H1’</th>
<th>CH3</th>
</tr>
</thead>
<tbody>
<tr>
<td>Loop 1</td>
<td>7.87</td>
<td>6.37</td>
<td>2.01</td>
<td>7.94</td>
<td>6.43</td>
<td>2.04</td>
</tr>
<tr>
<td>Loop 2</td>
<td>7.56</td>
<td>5.44</td>
<td>1.33</td>
<td>7.59</td>
<td>5.47</td>
<td>1.33</td>
</tr>
<tr>
<td>Loop 3</td>
<td>7.53</td>
<td>5.49</td>
<td>1.34</td>
<td>7.53</td>
<td>5.43</td>
<td>1.28</td>
</tr>
</tbody>
</table>

*TWJ1 1H chemical shifts (19) standard CTTG tetra loops taken as reference.

1RMSD(TWJ11_loops) = (∑(y(TWJ11_loop2)−y(TWJ11_loop1))2)1/2

<table>
<thead>
<tr>
<th>Loop</th>
<th>TWJ1</th>
<th>TWJ1*</th>
<th>RMSTD</th>
<th>R2c</th>
</tr>
</thead>
<tbody>
<tr>
<td>H6</td>
<td>7.87</td>
<td>6.37</td>
<td>2.01</td>
<td>7.94</td>
</tr>
<tr>
<td>H1’</td>
<td>7.56</td>
<td>5.44</td>
<td>1.33</td>
<td>7.59</td>
</tr>
<tr>
<td>CH3</td>
<td>7.53</td>
<td>5.49</td>
<td>1.34</td>
<td>7.53</td>
</tr>
</tbody>
</table>

The latter is demonstrated by comparison of the conformational shifts (44) of the junction loop residues with respect to those of the classical H2-type CTTG loop. As can be seen, the conformational shifts have a particularly small deviation (RMSD = 0.12 ppm) and correlate nearly perfectly (R2 = 0.98). In the standard CTTG loops—capping the arms—these parameters are even better. The slightly higher deviation and lower correlation, observed for the junction loop, is likely to be related to the break in the phosphate backbone (5’-CTTG-3’ versus 5’-CTTG-3’). Therefore, based on all these observations, the junction loop (5’-C27T28T29/G18-3’) could be safely assumed to fold as an H2-type hairpin loop, called quasi H2-type hairpin loop here, because of the break in the backbone.

Table 1. 1H Chemical shift comparison of CTTG loops

*RMSTD(TWJ11_loops) = (∑(y(TWJ11_loop2)−y(TWJ11_loop1))2)1/2

<table>
<thead>
<tr>
<th>Loop</th>
<th>TWJ1</th>
<th>TWJ1*</th>
<th>RMSTD</th>
<th>R2c</th>
</tr>
</thead>
<tbody>
<tr>
<td>H6</td>
<td>7.87</td>
<td>6.37</td>
<td>2.01</td>
<td>7.94</td>
</tr>
<tr>
<td>H1’</td>
<td>7.56</td>
<td>5.44</td>
<td>1.33</td>
<td>7.59</td>
</tr>
<tr>
<td>CH3</td>
<td>7.53</td>
<td>5.49</td>
<td>1.34</td>
<td>7.53</td>
</tr>
</tbody>
</table>

1RMSD: linear correlation coefficient of chemical shifts of TWJ11 tetra loops versus reference TWJ1 tetra loop (cf above footnote b).
contacts among residues G7, C8, G17 and C30 for TWJ10 and among A7, C8, G17 and T30 for TWJ12 are observed. In addition, the junction loop folds in both cases—as for TWJ11—into a quasi H2-type hairpin loop (5′-CTT/G-3′), as follows from the characteristic chemical shifts and NOE contacts of the loop residues. In conclusion, TWJ10 and TWJ12 into an A/C-stacked conformation as predicted on the basis of the two folding rules (Figure 1). Also, the switch in the penultimate position of the Pu:Py base-pairing in arm B (TWJ10 versus TWJ11; Figure 1B and C) neither affects stacking preference nor weakens base-pairing across the junction (C30:G7//G17:C8 in TWJ11 versus T30:A7//G17:C8 in TWJ12; Figure 1B and C).

**Structure calculations and statistics**

Despite the complexity of the spectra, a relatively large number of conventional NMR restraints (NOE, torsion angles; see Materials and Methods) were collected for TWJ11 (av. 25 restraints per residue, Table 2). These restraints were applied with conservative error bounds (30) in the structure calculation protocol based on Torsion Angle Dynamics (35) (see Materials and Methods). Briefly, an ensemble of 100 extended starting structures were allowed to re-fold. From the resulting set, the 30 lowest energy structures were selected for further analysis. This set displayed good local structure statistics with no NOE violation >0.5 Å and no torsion angle violation >10° (Table 3). The overlay of the 12 lowest energy structures shows that TWJ11 has a well-defined local structure (Figure 3A and B, left).

<table>
<thead>
<tr>
<th>Table 2. NMR-derived restraints for TWJ11</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Restraints</strong></td>
</tr>
<tr>
<td>-------------------------------------------</td>
</tr>
<tr>
<td>Distances</td>
</tr>
<tr>
<td>Inter-residues</td>
</tr>
<tr>
<td>Intra-residues</td>
</tr>
<tr>
<td>Hydrogen bonds</td>
</tr>
<tr>
<td>Av. per residue</td>
</tr>
<tr>
<td>Dihedrals</td>
</tr>
<tr>
<td>Glycosidic</td>
</tr>
<tr>
<td>Sugar puckering</td>
</tr>
<tr>
<td>Backbone</td>
</tr>
<tr>
<td>Av. per residue</td>
</tr>
<tr>
<td>RDC</td>
</tr>
<tr>
<td>Detailed</td>
</tr>
<tr>
<td>Av. per residue</td>
</tr>
<tr>
<td>All</td>
</tr>
</tbody>
</table>

*Arm A runs from residue 8 to 17, arm B from 19 to 26, arm C from 1 to 7 and from 30 to 36, and junction loop encompasses residues 18, 27, 28 and 29. Numbers in parentheses represent the dipolar couplings effectively used during RDC-refinement.

A, B and C have a low RMSD on heavy atoms of 0.56, 0.50 and 1.16 Å, respectively (Table 3). Figure 3A shows that the A/C stacking is well determined by the 23 NOE restraints (57) defining the junction (RMSD = 0.73 ± 0.27 Å, residue 6–9, 16, 17, 30, 31). Finally, the pseudo-tetraloop C27T28T29/
Table 3. Structural Statistics for TWJ11

<table>
<thead>
<tr>
<th>Structural elements</th>
<th>Arm A</th>
<th>Arm B</th>
<th>Arm C</th>
<th>Junction loop</th>
</tr>
</thead>
<tbody>
<tr>
<td>Distance &gt; 0.3 Å</td>
<td>1.3 (± 0.5)</td>
<td>2.2 (± 0.6)</td>
<td>5.8 (± 0.9)</td>
<td>0.4 (± 0.9)</td>
</tr>
<tr>
<td>Dihedral angle &gt; 5.0 deg</td>
<td>1.6 (± 0.8)</td>
<td>2.1 (± 0.5)</td>
<td>5.5 (± 1.0)</td>
<td>0.1 (± 0.3)</td>
</tr>
<tr>
<td>RDC &gt; 2.0 Hz</td>
<td>3.2 (± 0.6)</td>
<td>3.7 (± 0.5)</td>
<td>2.3 (± 0.8)</td>
<td>—</td>
</tr>
<tr>
<td></td>
<td>3.6 (± 0.9)</td>
<td>3.4 (± 0.5)</td>
<td>2.0 (± 0.7)</td>
<td>—</td>
</tr>
<tr>
<td></td>
<td>9.6 (± 0.9)</td>
<td>9.7 (± 0.7)</td>
<td>20.6 (± 1.4)</td>
<td>—</td>
</tr>
<tr>
<td></td>
<td>3.9 (± 1.5)</td>
<td>1.9 (± 0.7)</td>
<td>5.3 (± 1.4)</td>
<td>—</td>
</tr>
<tr>
<td>Heavy-atoms RMSD (Å)</td>
<td>0.56 (± 0.24)</td>
<td>0.50 (± 0.33)</td>
<td>1.16 (± 0.46)</td>
<td>0.78 (± 0.22)</td>
</tr>
<tr>
<td></td>
<td>0.51 (± 0.21)</td>
<td>0.52 (± 0.31)</td>
<td>1.00 (± 0.38)</td>
<td>0.79 (± 0.30)</td>
</tr>
<tr>
<td>Helical parameters</td>
<td></td>
<td></td>
<td></td>
<td>A/C helix</td>
</tr>
<tr>
<td>Rise (Å)</td>
<td>4.4 (± 0.3)</td>
<td>3.9 (± 0.2)</td>
<td>4.5 (± 0.2)</td>
<td>—</td>
</tr>
<tr>
<td>Twist (deg)</td>
<td>28 (± 2)</td>
<td>30 (± 2)</td>
<td>30 (± 2)</td>
<td>—</td>
</tr>
<tr>
<td>Energy</td>
<td>Overall</td>
<td>NOE</td>
<td>Dihedral</td>
<td>RDC</td>
</tr>
<tr>
<td></td>
<td>942 (± 26)</td>
<td>167 (± 11)</td>
<td>30 (± 3)</td>
<td>—</td>
</tr>
<tr>
<td></td>
<td>980 (± 33)</td>
<td>173 (± 9)</td>
<td>37 (± 4)</td>
<td>31 (± 9)</td>
</tr>
</tbody>
</table>

*Arm A runs from residue 8 to 17, arm B from 19 to 26, arm C from 1 to 7 and from 30 to 36, and junction loop encompasses residues 18, 27, 28 and 29.

*Upper values refers to calculation without RDC refinement while the lower value to refers to RDC-refinement calculation. The presented values are averaged over the number of residues involved in the structural element.

*No distance violation >0.5 Å was observed.

*No dihedral angle >10° was observed.

*For the experimental RDC restraints, no violation >±27 Hz before RDC-refinement while no violation >±6 Hz after RDC-refinement.

*Overall energy is in Kcal.mol⁻¹, NOE energy term is in kcal. Å⁻¹ mol⁻¹, dihedral energy term is in kcal deg⁻¹ mol⁻¹, and the dipolar coupling energy term is in Kcal Hz⁻¹ mol⁻¹.

G18 (Figure 4A), is relatively well-defined by 12 NOE restraints (S7) (RMSD = 0.78 Å, Table 3).

Because of the absence of NOE contacts between arm B and the A/C-stacked helices—lack of long-range restraints—the relative orientation of the arms is poorly defined as can be seen in Figure 3 (A, B and C, left) and is evidenced by the high RMSD of 4.7 Å for the whole molecule. Fortunately, RDC have recently been shown to be a powerful tool to define the global fold of large biomolecules (31,38,39,49–51). A total of 70 13C–1H RDCs was collected for a TWJ11 sample dissolved in a solution of Pf1 phages (24,26) and measured at natural abundance using cryoprobe technology (i.e. ~2 RDCs per residue; Table 2). This RDC set was subsequently applied as additional restraints in further TAD type structure calculations using the locally defined 30 NOE-based structures as starting structures (see Materials and Methods). To concentrate on the improvement of the global structure definition, the conformations of the arms were kept fixed but not their relative orientation. The alignment tensor orientation was left floating, but the values of its axial and rhombic components (Ax and Rh, respectively) are kept fixed; the latter were determined from the RDC distribution and confirmed through a grid search (S8, see Materials and Methods). From the resulting globally refined ensemble, the 12 lowest energy structures were selected for further analysis.

The final ensemble of RDC-refined structures shows a well-defined global conformation (Figure 3A–C, right). The overall RMSD improved dramatically from 4.5 Å, for the NOE-based structure ensemble, to 2.03 Å for the RDC-refined ensemble (Table 3). The quality of the final structure set is evidenced further by the structural statistics of other parameters in Table 3. The RMSD of the back-calculated versus experimental RDC, improved dramatically from 9.2 Hz for the NOE-based structures to 1.6 Hz after global refinement (close to the experimental error on the RDC, Table 3).
Also, the correlation coefficient increases from a low value, 0.59, to close to perfect correlation, 0.99 (Table 3, S9). Furthermore, no RDC violation $>6$ Hz was observed in the final ensemble. At the same time, the structural characteristics—defining local structure—were not compromised by the additional RDC restraints. For instance, the stacking between arms A and C is affected by both RDC and the NOE contacts between junction base pairs (arms A and C are free to adjust in the structure calculation protocol). The number of NOE and torsion angle violations remained low with no NOE violation $>0.5$ Å and no torsion angle violation $>10^\circ$. Finally, the overall energy as well as the NOE and dihedral energy terms did not significantly increase as a result of the extra dipolar coupling energy term (Table 3).

**Mild but significant perturbations of the helical properties of the A/C helix at the junction penultimate position**

Although the backbone continuity is disrupted between residues G$_{17}$ and C$_{30}$ (Figure 4A), the A/C-stacked helix largely conforms to B-helix geometry (Figures 3 and 4A). To investigate in detail the A/C-helix geometry, helical parameters were extracted with Curves from the final structure ensemble. Moreover, to probe experimentally potential helix perturbations, chemical shifts and RDC patterns were analyzed (see Materials and Methods).

From Curves, the helical rise was found to be relatively high on average (4.4 ± 0.1 Å) as compared with the canonical B-DNA value of 3.4 Å (Table 3, S10). However, Curves sometimes overestimate the helical rise in places where other helical deviations occur, such as roll and tilt (52). Considering the twist angle values, the largest deviation is observed around the closing loop base pair in arm A as expected (19$^\circ$/C14) (S10). Around the junction, the twist angles display the pattern, 32$^\circ$/C14, 35$^\circ$/C14 and 32$^\circ$/C14, for G$_7$:C$_{30}$/A$_6$:T$_{31}$, G$_6$:C$_{17}$/G$_5$:C$_{30}$ and G$_5$:C$_{16}$/C$_5$:G$_{17}$, respectively (S10). On average, the A/C-helix twist angle ($T_{\text{ave}}$), was found to be 30$^\circ$, thus 6$^\circ$ smaller than in a regular B-helix. In conclusion, Curves analysis shows relatively mild deviations from B-helix values, considering either the overall A/C helix or the junction.

To probe and experimentally confirm such relatively mild deviations, residual dipolar couplings have been shown to be particularly suitable and rather sensitive. For example, the C$_1$H$_1'$ RDC form so-called dipolar waves (53,54), which within a regular helix, show a simple dependence on the twist angle ($T$) and base pair number ($n$):

$$D_{\text{helix}}^{\text{C1H1'}} = A + B \cos(2Tn + \phi)$$

Where $A$ and $B$ are factors dependent on the orientation of the DNA helix with respect to the alignment tensor and $\phi$ is a phase factor. $D_{\text{helix}}^{\text{C1H1'}}$ was plotted versus base pair number for $T = 24^\circ$, 30$^\circ$ or 36$^\circ$, together with the experimental C$_1$H$_1'$ RDC (Figure 4C). The best-fit uniform $T$-value ($T_{\text{opt}}$) for the A/C helix is 30$^\circ$. This value equals the average twist in the helical chain. Additionally, the twist was plotted versus the base-pair number in the A/C helix for $T = 24^\circ$, 30$^\circ$ or 36$^\circ$, together with the experimental C$_1$H$_1'$ RDC (Figure 4C). The best-fit uniform $T$-value ($T_{\text{opt}}$) for the A/C helix is 30$^\circ$. This value equals the average twist in the helical chain.
the ensemble of structures, $T_{ave}$ (Table 3). Because $T_{opt}$ is derived directly from the experimental RDC data, the close similarity of the two values gives additional credence to the quality of the derived structures. And thus, that the helical twist values (and other helix parameters) in the final ensemble, indeed faithfully reflect reality.

Chemical shifts are sensitive parameters particularly to probe local geometry, and thus are especially suited to reveal slight deviations from canonical B-helix conformation. In Figure 4D, chemical shift deviations from regular B-helix ($\Delta\delta_{obs,helix}$, calculated for aromatic-1H), have been plotted as a function of residue along the A/C-stacked helix. It is striking to observe that the most dramatic deviations occur at and around residue G9 rather than being localized exactly at the junction site (G2, C8, G17 and C30). A closer examination of the C12H12 RDC dipolar wave pattern (Figure 4C) also pinpoint residue G9, as affected by higher deviation from regular helical behavior.

In conclusion, we observe that regular helical properties are roughly conserved along the A/C-stacked helix, and more remarkably throughout the junction site. Nevertheless, experimental data (chemical shifts and RDC) designate G9 position as the most perturbed around the branch point. This is particularly striking as, together with C16, it forms the junction penultimate base pair (Figure 1B), source of the empirical ‘pyrimidine rule’.

The junction loop is a stable quasi H2-type hairpin loop
The 5’-CTT/G-3’ junction loop indeed adopts a quasi H2-type hairpin loop as seen in the final structure set (Figure 4A and B) as was already predicted from the chemical shifts and NOE patterns (vide infra). Residue T29 stacks onto the closing base pair C27:G18, whereas T28 folds back into arm B minor groove (Figure 4A and B). Despite the break in the backbone continuity between T29 and G18, the loop conformation is closely similar to a classical 5’-CTTG-3’ tetraloop (Figure 4B). This close correspondence in conformation combined with the well-known thermodynamic stabilities of regular hairpin loops (22,23), well explains the importance of the ‘extended loop rule’ in the prediction of 3H stacking preference.

Geometry restricts the 3H global conformational space: towards global fold prediction from sequence
We have demonstrated that the pyrimidine—and extended loop rules well predict DNA 3HS2 stacking preference from sequence. A second important aspect is their global conformation, i.e. the relative orientation of the stacked and non-stacked arms. If this can be predicted as well, a nearly complete picture of 3HS2 structure would be obtained. By analyzing the global structure of TWJ11 and representatives of other 3H classes, we attempt to find clues that would further explain the global conformation selection process.

The global conformation of a stacked 3H is fully defined by the relative orientation of the non-stacked arm and stacked arms and thus by three parameters, the three Euler angles defining their relative orientation. To perform this analysis, we define a right-handed reference frame, attached to the stacked helix and with its z-axis along the helix axis and in a similar fashion a second frame attached to the non-stacked arm (for exact definitions, see Figure 5A and B). The three Euler angles then correspond to the polar angles of the helix axis of the non-stacked arm with respect to the reference frame attached to the stacked helix ($\theta$, $\phi$, and to rotation around its own helix axis ($\omega$). These frames are adapted to the discussed 3H cases to prevent any bias arising from the difference in primary sequence. For the analysis, we considered structures from three main classes of 3H (Figure 5C): an A/C-stacked DNA 3HS2 (TWJ11), an A/B-stacked DNA 3HS2 (TWJ11) (19) and the A/C-stacked RNA 3H (HH) studied by Scott et al. (55).

The angle values ($\phi$, $\theta$, $\omega$) obtained for each of the considered structures are: 10 $\pm$ 1°, 42 $\pm$ 14°, $-19 \pm 10^\circ$ (TWJ 11, A/C-stack); 23°, 114°, 51° (TWJ1, A/B-stack); $-8^\circ$, 140°, $-62^\circ$ (HH, A/C-stack).

A first observation is that for all the 3Hs, the non-stacked arm moves essentially within the XOZ plan (Figure 5A and C), i.e. $\phi \sim 0^\circ$, whatever the inter-helix angle $\theta$ or rotation angle $\omega$. For comparison, in case of 4H, $\phi$ is $\sim 90^\circ$ (51,56). This is easily explained by the fact that for a 4H, $\phi \sim 0^\circ$ would cause clashes between the two stacked helices for inter-helix angles larger than zero ($\theta > 90^\circ$).

A second striking observation is made when comparing—in TWJ11 and HH—the relative orientations of the non-stacked arms (Figure 5C, left and right). The conformational analysis reveals that an increase in $\theta$ leads to a decrease in $\omega$, according to $\omega = -A\theta + B$ (A > 0, Figure 6). This anti-correlation can be explained in geometric terms. Pulling TWJ11 arm B downwards (increasing $\theta$; Figure 5C, left), leads to steric incompatibilities, i.e. the lower strand of arm B (including the junction loop) comes then into close contact with arm C (Figure 5C, left). To prevent this clash, a rotation around the arm B axis (increase of $\mid \omega \mid$) is required. Moreover, the sign of the rotation is important. As shown elsewhere (51,56–62), a positive $\omega$ rotation leads to intermingling or steric clashes of the two strands in the non-stacked arm (near the junction), whereas a negative $\omega$ rotation does not. In other words, a positive $\omega$ rotation is restricted to small angles, whereas a negative $\omega$ rotation is not. This, and the more-or-less fixed length of the backbone connecting stacked and non-stacked arm at the junction explains, the observed negative $\omega$ rotation upon an increase in $\theta$ in Figure 6. Based on similar considerations, the same linear dependence should apply for an A/B-stacked 3H (Figure 6, A/B stacking), with the difference that there is a phase shift as compared with the previous example ($\omega = -A\theta + B + C$; A > 0). This phase shift is due to the fact that in A/B-stacked helix, the junction loop is located in the upper strand (Figure 5C, middle), whereas it is located in the lower strand in an A/C-stacked helix (Figure 5C, left).

Finally, comparing the A/B- and A/C-stacked 3HS2 (Figure 5C, left and middle), an evident correlation is revealed between the junction loop position and the direction of the non-stacked arm. For TWJ11, the non-stacked arm points upward ($\theta < 90^\circ$), whereas the junction loop points downward; for TWJ1, the non-stacked arm points down ($\theta > 90^\circ$), whereas the junction loop points up. This correlation can be explained on steric grounds. Conformations with non-stacked arm and junction loop on the same side, i.e. up/up or down/down, are prohibited by lack of space for the loop in between the non-stacked and stacked arms.

In conclusion, the analysis shows that the global conformational space allowed for the 3HS2 considered is strongly restricted: $\phi \sim 0^\circ$, $\omega = -A\theta + B + C$, A > 0 $\sim 0.42,
Figure 5. Analysis of the global conformation of TWJ11 and other 3H classes. (A) Schematic presentation of the global conformation of a 3H, defined as the orientation of the non-stacked arm with respect to the stacked arms—i.e. the orientation of an axis frame attached to the non-stacked arm with respect to a reference axis frame attached to the stacked arms. The global conformation can then be defined by the polar angles \( \theta \) and \( \phi \) of the helix axis of the non-stacked arm and by the angle \( \omega \) describing rotation around of this helix axis. To allow for an unbiased comparison of the global conformation of the different classes of 3Hs shown in (C)—the reference frame and that of the non-stacked arm—are defined as follows. Reference frame: Z stacked helix axis; Y = Z \( \times \) \( \mathbf{v}_1 \); X = Y \( \times \) Z; // means parallel; \( \times \) means cross-product; \( \mathbf{v}_1 \) is the \( C_1 \)–\( C_1' \) vector in the lower junction base pair in (C) and pointing from the continuous to the crossover strand, e.g. \( C_8 \)–\( C_1 \) in TWJ11. Non-stacked arm frame: Z \( \times \) non-stacked arm helix axis, Y = Z \( \times \) \( \mathbf{v}_2 \); X = Y \( \times \) Z; \( \mathbf{v}_2 \) is the \( C_1' \)–\( C_1' \) vector of the first base pair from the junction in the non-stacked arm and pointing from the crossover strand coming from the upper arm in (C) toward the crossover strand coming from the lower arm (i.e. from strand 1 to 2 in A/C-stacking mode and from strand 2 to 3 in A/B-stacking mode), e.g. \( C_1 \)–\( C_1' \) to \( C_1' \)–\( C_1' \) in TWJ11. (B) Stereoview of TWJ11: inside the A/Chelix (arm A in red, arm C in green) and B arm (in blue) are displayed the axis frames used to define the respective helices orientation. For both axis systems, Z is in blue, Y in green and X in red. (C) Comparison of the orientation of the non-stacked arm of TWJ11, TWJ1 (19) and the hammerhead ribozyme HH (55). All structures are presented with their junction minor groove facing the viewer. TWJ11 has A/C-stacked conformation with an anti-parallel orientation, TWJ1 has A/B-stacked conformation anti-parallel, and HH has an A/C-stacked conformation parallel.

CONCLUSIONS

A new high-resolution 3H structure has been described based on the solution NMR data (NOEs, J-couplings as well as RDCs). It (TWJ11) adopts an A/C-stacked anti-parallel conformation, with arm B tilted away by 42 ± 14° from the main A/C-helix axis. The junction loop folds into pseudo tetraloop with an H2-type conformation. Two TWJ11 mutants (TWJ10, TWJ12) show the same A/C-stacking mode and junction loop fold. These stacking preferences can now be predicted from sequence for both classes of DNA 3HS2. Analysis of the global conformation of TWJ11 and other representatives of the 3H family shows that the relative orientations of the stacked and non-stacked arms are restricted to relatively narrow regions of conformational space. These restrictions follow from geometric considerations given the stacking preference. Together with the stacking rules extended to the complete DNA 3HS2 family, this constitutes an important step towards full prediction of 3HS2 conformation (stacking and global fold) directly from sequence.

SUPPLEMENTARY MATERIAL

Supplementary Material is available at NAR Online. Tables S1–S3 (chemical shifts of TWJ10, TWJ11 and TWJ12), Figures S4–S6 (sequential walk in NOESY in TWJ10,
ACKNOWLEDGEMENTS

This work was supported by grants from Chemistry section of the Dutch Science Organization (S.W.), the Kempe Minne Foundation (B.V.B.), the Swedish National Research Council (S.W.) and the Bioteknik Medel Umeå University (S.W.).

REFERENCES


