

# Nucleic acid folding determined by mesoscale modeling and NMR spectroscopy: solution structure of d(GCGAAAGC)

*Guillaume P. H. Santini<sup>†</sup>, Jean A. H. Cognet<sup>†</sup>, Duanxiang Xu<sup>‡</sup>, Kiran K. Singarapu<sup>‡</sup> and Catherine  
Hervé du Penhoat<sup>‡,§\*</sup>.*

<sup>†</sup>Laboratoire de Biophysique Moléculaire, Cellulaire et Tissulaire, UMR 7033 CNRS, Université Pierre  
et Marie Curie Paris 6, Genopole Campus 1, 5 rue Henri Desbruères, Évry 91030, France

<sup>‡</sup>Department of Chemistry, The State University of New York at Buffalo, Buffalo, New York 14260,  
USA

<sup>§</sup>Laboratoire de Biophysique Moléculaire, Cellulaire et Tissulaire, UMR 7033 CNRS, Université Paris  
13, 74 rue Marcel Cachin, 93017 Bobigny, France

Guillaume P. H. Santini, [santini@closun.snv.jussieu.fr](mailto:santini@closun.snv.jussieu.fr) ; Jean A. H. Cognet, [jean.cognet@upmc.fr](mailto:jean.cognet@upmc.fr) ;  
Duanxiang Xu, [dxu@micronmr.com](mailto:dxu@micronmr.com) ; Kiran K. Singarapu, [singarapu@wisc.edu](mailto:singarapu@wisc.edu) ; Catherine Hervé du  
Penhoat, [penhoat8@aol.com](mailto:penhoat8@aol.com) .

**RECEIVED DATE (to be automatically inserted after your manuscript is accepted if required  
according to the journal that you are submitting your paper to)**

Mesoscale modeling and NMR spectroscopy of d(GCGAAAGC)

CORRESPONDING AUTHOR FOOTNOTE \*Corresponding author. Tel: +33 1 4838 7391; Fax: +33 1

**ABSTRACT.** Determination of DNA solution structure is a difficult task even with the high-sensitivity method used here based on simulated annealing with 35 restraints/residue (cryoprobe 750 MHz NMR). The conformations of both the phosphodiester linkages and the dinucleotide segment encompassing the sharp turn in single-stranded DNA are often underdetermined. To obtain higher quality structures of a DNA GNRA loop, 5'-d(GCGAAAGC)-3', we have used a mesoscopic molecular modeling approach, called Biopolymer Chain Elasticity (BCE), to provide reference conformations. By construction, these models are the least deformed hairpin loop conformation derived from canonical B-DNA at the nucleotide level. We have further explored this molecular conformation at the torsion angle level with AMBER molecular mechanics using different possible ( $\epsilon, \zeta$ ) constraints to interpret the  $^{31}\text{P}$  NMR data. This combined approach yields a more accurate molecular conformation, compatible with all the NMR data, than each method taken separately, NMR/DYANA or BCE/AMBER. In agreement with the principle of minimal deformation of the backbone, the hairpin motif is stabilized by maximal base-stacking interactions on both the 5'- and 3'-sides and by a sheared G·A mismatch base pair between the first and last loop nucleotides. The sharp turn is located between the third and fourth loop nucleotides and only two torsion angles  $\beta(6)$  and  $\gamma(6)$  deviate strongly with respect to canonical B-DNA structure. Two other torsion angle pairs  $\epsilon(3), \zeta(3)$  and  $\epsilon(5), \zeta(5)$  exhibit the newly recognized stable conformation  $\mathbf{B}_{\text{II}\zeta^+}$  ( $-70^\circ, 140^\circ$ ). This combined approach has proven to be useful for the interpretation of an unusual  $^{31}\text{P}$  chemical shift in the 5'-d(GCGAAAGC)-3' hairpin.

**KEYWORDS.** DNA hairpin, phosphorus chemical shifts, Biopolymer Chain Elasticity, molecular dynamics

## Introduction

To shed light on macromolecular recognition of RNA and DNA molecules, numerous studies of their structure, thermodynamic stability, and folding have been undertaken.<sup>1,2,3</sup> Much effort has been devoted to predicting the tertiary structure from sequence based on elementary local structural motifs such as hairpins and frequently observed tertiary interactions.<sup>4</sup> For DNA, the recent evolution of gene medicine from an experimental technology into a viable strategy for developing therapeutics to treat human disorders<sup>5</sup> has further revived the interest in predicting and/or solving high-quality DNA structures at atomic resolution.

One of the most important nucleic acid building blocks, the ‘hairpin motif’, is composed of a helical stem capped by a 3- or 4-residue loop connecting the two strands forming the helix. In particular, the recurrent RNA GNRA (N = A, C, G, T or U; R = A or G) and UNCG tetraloops<sup>1</sup> and the DNA tri-loops associated with genetic diseases<sup>6</sup> have been studied comprehensively. The thermal stability of hairpins varies greatly<sup>7-9</sup> as a function of (i) the length and the sequence of the single-stranded loop region, (ii) the number of base pairs in the stem region, and (iii) the type of base pair that closes the stem and precedes the loop region. Furthermore, some hairpins are capable of adopting more than one stable conformation, and inter-conversion between these forms may constitute a conformational switch for cellular regulation.<sup>10,11</sup> Hairpin structures have been classified<sup>10,12</sup> based on a qualitative description of stacking interactions, base-pairing in both the stem and loop regions, as well as additional hydrogen bonds.

GNRA and UNCG tetraloop structures are remarkably stable and analogous DNA and RNA hairpins have very high melting temperatures: d(-GAAA-) ( $T_m = 76.5$  °C in 0.1 M NaCl),<sup>13</sup> r(-GAAA-) ( $T_m = 61$  °C in 0.01 M NaCl),<sup>14</sup> d(-TTCG-) (60.4 °C in 1 M NaCl) and r(-UUCG-) (76.5 °C in 1 M NaCl).<sup>15</sup> While the X-ray structure of an RNA GAAA tetraloop<sup>16</sup> is among the best resolved hairpin structures, a solution structure has not yet been reported for a DNA GAAA tetraloop. This tetraloop occurs at the

replication origin of bacteriophage G4 single-stranded DNA<sup>17,18</sup> and is thought to play a role in the initiation mechanism. To date, the closest DNA loop sequence for which a hairpin structure has been determined<sup>19</sup> is the GTTA tetraloop ( $T_m = 69.5$  °C for d(GCGTTAGC) in 0.1 M NaCl).<sup>13</sup>

Solution structures of nucleic acids are currently determined with standard NMR methods.<sup>22,23,24</sup> The local conformation of the sugar-phosphate backbone is described by six torsion angles ( $\alpha$ ,  $\beta$ ,  $\gamma$ ,  $\delta$ ,  $\epsilon$ ,  $\zeta$ ) while the orientation about the bond between the sugar and the base is defined by only one ( $\chi$ ).<sup>25</sup> Simple relationships between the experimental data, such as homo- and heteronuclear scalar coupling constants, linewidths and nuclear Overhauser effects (NOEs), define the conformation of the  $\beta$ ,  $\gamma$ ,  $\delta$ ,  $\epsilon$ , and  $\chi$  torsions while <sup>31</sup>P chemical shifts are related to the conformation about the  $\alpha$  and  $\zeta$  dihedral angles.<sup>26</sup> Recent theoretical studies<sup>27,28</sup> have confirmed that the phosphorus in the  $g^-/t$  ( $\alpha = -60^\circ$ ,  $\zeta = 180^\circ$ ) conformation is more deshielded than the one in the  $g^-/g^-$  ( $\alpha$ ,  $\zeta = -60^\circ$ ) conformation. <sup>31</sup>P chemical shifts have been used to determine  $B_I$  ( $\epsilon = t$ ,  $\zeta = g^-$ ) and  $B_{II}$  ( $\epsilon = g^-$ ,  $\zeta = t$ ) populations in DNA duplexes.<sup>29</sup> However, as other factors influence <sup>31</sup>P chemical shifts,<sup>24</sup> only very loose constraints ( $\pm 120^\circ$ ) have been applied to  $\alpha$  and  $\zeta$  and only in the case of the  $B_I$  conformation. As a result, these latter torsions are often poorly defined in NMR studies.

Molecular mechanics has not provided torsion angle restraints for either  $\alpha$  or  $\zeta$  as *in vacuo* analysis of staggered nucleic acid conformations has not indicated large variations in their energy.<sup>30</sup> From the onset, it was recognized that molecular modeling would be required to determine the various conformational families that contribute to the NMR-defined time-averaged structure of tetraloops. To obtain higher resolution conformation we chose a mesoscopic molecular modeling approach capable of producing the least deformed conformation from B-DNA.

Recently, we have shown that the published conformations of several DNA tri- or tetraloops, and RNA tetraloops adopt a simple global folding. For these complex structural motifs, the trajectory of the sugar-phosphate chain was shown to follow the folding of a flexible rod on the scale of several monomers.<sup>20,21</sup> This line, also called the elastic line, is the trajectory with the least deformation energy of

a continuous flexible thin rod computed with the theory of elasticity.<sup>20</sup> In this mesoscopic approach called biopolymer chain elasticity (BCE), and for DNA tri- or tetraloops, the sugar-phosphate chain is initially taken as B-DNA, and is transported with minimal deformation onto the elastic line.<sup>20</sup> In addition, the backbone trajectory is represented as a 1-D line in 3-D space around which nucleotides can be rotated (with angles  $\Omega_i$  and  $\chi_i$ , *vide infra*) independently. As a result, this approach reproduces the reversal in the direction of the phosphodiester backbone characteristic of hairpins in such a manner that only a few torsion angles of the hairpin structure vary with respect to canonical B-DNA values. It has also been successfully applied to reproduce the positions and orientations of the bases in these DNA tri- and tetraloops hairpins in agreement with both NMR derived distances<sup>20</sup> and Protein Data Bank (PDB) structures.<sup>21</sup> For these two reasons, pertaining to the global folding of the chain and to the local position and orientation of the bases, the BCE approach is by definition and by construction the simplest one for generating hairpin loop conformations from B-DNA with the least chain deformation.

Here we present a general combined approach for evaluating nucleic acid structure according to the scheme in Figure 1. NMR data for the hairpin formed by the DNA strand 5'-d(G<sub>1</sub>C<sub>2</sub>G<sub>3</sub>A<sub>4</sub>A<sub>5</sub>A<sub>6</sub>G<sub>7</sub>C<sub>8</sub>)-3' were obtained with <sup>1</sup>H, <sup>31</sup>P and natural abundance <sup>13</sup>C NMR spectroscopy. Distance and torsion angle restraints were obtained from line-widths, cross-peak patterns and NOESY cross-peak volumes according to literature protocols.<sup>22,23</sup> As the scope and limitations of deriving restraints for the  $\alpha$  and  $\zeta$  torsions from <sup>31</sup>P chemical shifts have not been established in the case of hairpins, all possible interpretations of the <sup>31</sup>P NMR data were systematically explored using the DYANA simulated annealing protocol.<sup>31</sup> These calculations yielded five NMR-derived ensembles, **1DYANA(a-e)** that differed only in the conformation of two ( $\epsilon, \zeta$ ) torsion angle pairs.

In parallel, the least deformed theoretical structures, **2BCEopt()**, were obtained by construction with the BCE approach. These conformations were modified to explore different global folding positions (*e.g.* stacking of A5 on A4 or not) suggested by NMR-derived data. A reference structure, **2BCEopt\_nmr**, was obtained representing the least-deformed B-DNA conformer that took into account

these NMR-derived folding observations. To resolve the ambiguity regarding the two ( $\epsilon, \zeta$ ) torsion angle pairs the **2BCEopt\_nmr** was systematically modified by energy-minimization with AMBER to give **3MIN(a-r)**. This provided an energy-based exploration of the ( $\epsilon, \zeta$ ) pairs while maintaining the global fold of **2BCEopt\_nmr**. Finally, the stability and dynamics of the BCE conformer that best reproduces all the NMR data was tested with AMBER molecular dynamics trajectories. This protocol affords an atomic resolution structure and a description of the dynamics of the 5'-d(GCGAAAGC)-3' DNA hairpin.

## Material and Methods

**Sample preparation.** 5'-d(GCGAAAGC)-3' was purchased from Trilink Inc and purified on a Sephadex G10 gel-filtration column. Half of the purified DNA was lyophilized twice from D<sub>2</sub>O buffer (pH 6.8, 10 mM Na-phosphate, 50 mM NaCl, 5  $\mu$ M EDTA, 2.5  $\mu$ M NaN<sub>3</sub>), dissolved in 500  $\mu$ L D<sub>2</sub>O and placed in a 5 mm tube yielding a 4 mM hairpin solution. The remainder was dissolved in 400  $\mu$ L of the same buffer in 90% H<sub>2</sub>O/10% D<sub>2</sub>O. 250  $\mu$ L were placed in a 5 mm Shigemi tube to yield a 4.5 mM sample.

**NMR spectroscopy.** NMR measurements (<sup>13</sup>C and <sup>31</sup>P, indirect chemical shift referencing to DSS) were performed at 25 °C using a Varian INOVA750 NMR spectrometer with a standard triple resonance probe unless stated otherwise. 1D 'jump-return' spectra<sup>33</sup> were acquired between 1 and 25 °C ( $\tau = 46$  and 105  $\mu$ s giving the maximum signal intensity in the imino and amino regions, respectively).

The following 2D <sup>1</sup>H-spectra were recorded in D<sub>2</sub>O: *2QF-COSY*<sup>34</sup> ( $t_{1\max} = 155$  ms,  $t_{2\max} = 155$  ms, total measurement time 14 h); *clean* [<sup>1</sup>H; <sup>1</sup>H]-*TOCSY*<sup>35</sup> (64 ms, 121 ms, 1.4 h;  $\tau_m = 38$  ms, 5 °C); [<sup>1</sup>H; <sup>1</sup>H]-*NOESY*<sup>36</sup> (73 ms, 147 ms, 3 h; five experiments with  $\tau_m = 50, 100, 150, 200,$  and 400 ms, respectively, recycle time 5 s; cryoprobe, 128 ms, 128 ms, 21 h; three experiments with  $\tau_m = 30, 40,$  and 60 ms, respectively, recycle time 6 s). The parameters for the 2D <sup>1</sup>H-spectrum measured in H<sub>2</sub>O were as

follows: ‘jump-return’ [ $^1\text{H}$ ;  $^1\text{H}$ ]-NOESY (68 ms, 315 ms, 6 h;  $\tau_m = 50, 100, 150,$  and 200 ms, respectively, recycle time 5 s,  $\tau = 105 \mu\text{s}$ , 1 °C; 68 ms, 68 ms, 6 h;  $\tau_m = 50, 100, 150$  ms, recycle time 5 s,  $\tau = 46 \mu\text{s}$ , 1 °C; cryoprobe, 68 ms, 68 ms, 21h;  $\tau_m = 60$  ms, recycle time 5 s, 5 °C,  $\tau = 40 \mu\text{s}$ ).

Finally, the following heteronuclear 2D experiments recorded in  $\text{D}_2\text{O}$ : [ $^{31}\text{P}$ ,  $^1\text{H}$ ]-COSY<sup>37,38</sup> (400 MHz, 171 ms, 244 ms, 11 h); [ $^{13}\text{C}$ ,  $^1\text{H}$ ]-HSQC<sup>39</sup> (51 ms, 128 ms, 12 h). The spectra were processed and analyzed using the programs PROSA<sup>40</sup> and XEASY<sup>41</sup>, respectively.

**Inter-proton distance restraints.** The buildup of the H2’/H2’’ NOESY cross-peak volumes at 25 °C was monitored (data corresponding to the shortest inter-proton distance expected, < 1.8 Å) and the maximum occurred in the spectrum with a 200-ms mixing time. Accordingly, the cross-peak volumes in the NOESY experiment acquired at 25 °C with the 100-ms mixing time were preferred for the DYANA calculations to limit spin diffusion. To take into account the increase in solvent viscosity at lower temperature that results in an increase in cross-relaxation rates, a shorter 50-ms mixing time was used at 1°C. 206  $^1\text{H}$ - $^1\text{H}$  modified upper limit restraints (upls) were derived from the 750 MHz NOESY cross-peak volumes using the isolated pair approximation.

Lower limit distance restraints (lols) supplementing those derived from the van der Waals radii were established for 144 well-resolved cross peaks of the NOESY spectrum recorded in  $\text{D}_2\text{O}$  according to the relation lower-limit or lol =  $0.6 * \text{upl} + 0.5 \text{ \AA}$ , where only lower limits shorter than 3.5 Å were retained.<sup>42</sup> All aromatic and sugar protons exhibited some clear cross-peaks in the NOESY spectrum with the shortest mixing time indicating that the absence of cross-peaks could not be simply due to line broadening as a result of exchange processes. Therefore, the absence of cross-peaks in the NOESY spectra for 36 pairs of spins with isolated signals was translated into additional lower limit restraints of 3.5 Å. Lol restraints were applied in two DYANA simulations, **1DYANA(b)** and **1DYANA(e)**.

Conservative hydrogen bond restraints were included in the structure calculations for the G·C pairs: a upl of 2.25 Å for the acceptor distance, and of 3.25 Å for the donor distance, respectively. A type XI side-by-side sheared G·A base pair<sup>25</sup> has been demonstrated for both mismatched duplex and hairpin

forms of these DNA sequences<sup>12,42</sup> as well as for their RNA counterparts (RNA GNRA<sup>43,44</sup>). Thus, restraints for a sheared G·A base pair<sup>25</sup> (H22(G3)/N7(A6) and H62(A6)/N3(G3) hydrogen bonds, as above) were also included.

**Torsion angle restraints.** Sugar ( $\delta$ ), base ( $\chi$ ), and backbone ( $\beta$ ,  $\gamma$ ,  $\epsilon$ ) torsion angle restraints were deduced from the NMR data (linewidths, NOESY cross-peak volumes, DQCOSY and [<sup>31</sup>P, <sup>1</sup>H]-COSY cross-peak fine structure and intensities) according to standard procedures.<sup>22,23</sup> To take into account the different possible interpretations of <sup>31</sup>P chemical shifts ( $\delta_p$ ), the values of the  $\alpha$  and  $\zeta$  torsion angles were restricted as follows: calculations without  $\alpha$  or  $\zeta$  restraints, **IDYANA(a)** without lols, and **IDYANA(b)** with lols; canonical B-DNA control calculations with loose  $\alpha$  and  $\zeta$  constraints ( $0^\circ \pm 120^\circ$ ) for  $\delta(^{31}\text{P})$  in the B<sub>1</sub> range, **IDYANA(c)** without lols;<sup>24,45,46</sup> calculations with loose  $\alpha$  and  $\zeta$  constraints for  $\delta(^{31}\text{P})$  in both the B<sub>1</sub> range ( $0^\circ \pm 120^\circ$ ) and in the downfield-shifted range ( $180^\circ \pm 40^\circ$ ), **IDYANA(d)** without lols, or **IDYANA(e)** with lols. The first and second simulations were based on minimal and maximal NOE data only whereas <sup>31</sup>P restraints were progressively introduced in the remaining simulations. The last simulation relied on all the experimentally observed NOE and <sup>31</sup>P data.

**DYANA structure calculations.** All upls and the preceding dihedral angle constraints were translated into total dihedral angle constraints using the FOUND module<sup>47</sup> of DYANA, which performs a grid search for allowed conformations in the space spanned by the nine torsion angles describing a dinucleotide segment. Stereospecific assignments for all of the H2' and H2'' protons with non-degenerate chemical shifts could be deduced from the relative values of the vicinal coupling constants with H1' and the stronger NOEs between H1' and H2'' compared to the H1'/H2' NOEs. The GLOMSA module<sup>48</sup> of the DYANA program corroborated these H2'/H2'' assignments and provided the stereospecific assignments of the H5'/H5'' methylene protons of A4, A5, and G7. All simulations included constraints to close the sugar rings (C4'-O4': 1.41 Å, C4'-C1': 2.40 Å, C5'-O4': 2.39 Å, H4'-O4': 2.12 Å).

**BCE structure calculations.** Folding DNA or RNA hairpin loops of 3-4 nucleotides with the BCE

approach can be summarized as a three-step procedure.<sup>20,21</sup> Each step corresponds to modeling GAAA molecular conformation on a different scale.

*Step 1.* This global deformation step takes place on the mesoscopic scale of the loop and is sketched in Figure 2a-(1-2). Single-stranded B-DNA in the stem or in the loop is generated along helical lines (Figure 2a-1). B-DNA helices<sup>49</sup> are simple solutions of the theory of elasticity of thin rods, and can therefore be taken as elastic lines. The trajectory of the elastic line of a given length associated with a tetra-loop in tri-dimensional space is uniquely determined and computed for the geometry of end conditions imposed by the B-DNA helices as illustrated in Figure 2a-(1-2).<sup>20</sup> It is the trajectory of least energy of deformation. Transportation of the whole loop chain onto the elastic line is described elsewhere.<sup>20</sup> The final conformation obtained after these transformations is the molecular model, “BCEori”, shown in Figure 2a-2.

*Step 2.* This deformation step takes place on the scale of the nucleotide. Two rotation angles,  $\Omega_i$  and glycosidic torsion angle  $\chi_i$ , were sufficient to orient each nucleotide  $i$  in the loop<sup>21</sup> with respect to helical B-DNA while the stem nucleotides are unchanged ( $\Omega = 0, \Delta\chi = 0$ ). As shown in Figure 2b for G3, A4, A5 and A6, the attachment of block of atoms to the elastic line provides a convenient setup to independently rotate each block of atoms about the tangent to the elastic line with an angle,  $\Omega$ . These simple rotations ( $\Omega_i, \chi_i$ ) had been searched to match the Cartesian coordinates of the GTTA hairpin given in PDB file 1ac7.<sup>19,21</sup> Both GAAA and GTTA possess a G.A base pair with the second loop nucleotide stacked onto it. This is why only the third loop nucleotide ( $\Omega_5, \chi_5$ ) values had to be searched to take into account the NMR data for the GAAA hairpin. Three positions of A5 ( $\Omega_5$ ) were explored: in the major groove ( $\sim 70^\circ$ ), stacked ( $90^\circ$ ) and un-stacked conformers ( $>90^\circ$ ). This step yields BCE molecular models, “BCEopt”, and the best model BCEopt\_nmr is shown in Figure 2a-3.

*Step 3.* This modeling step takes place on the atomic scale. In folding steps 1 & 2, individual blocks are translated and rotated without deformation. However the chemical bonds and bond angles of the main atoms of the sugar-phosphate backbone (O5', C5', C4', C3', O3', P) between individual blocks are

modified. We observed that these chemical bonds and bond angles are alternately extended when located outside the region of curvature or compressed when located on the concave side.<sup>21</sup> This is why each molecular structure is briefly energy-refined by molecular mechanics to restore backbone bond lengths and bond angles to values close to their canonical values.<sup>20,21</sup> BCEopt\_nmr is close to the global minimum and the energy refined molecular model, “BCEopt\_nmr\_min” is the closest local minimum and is little modified when compared to BCEopt\_nmr. The non canonical NMR-defined  $\beta(6)$  and  $\gamma(6)$  torsions values were introduced at this stage. Additional torsion angle constraints were included to explore the conformations of the  $(\epsilon_i, \zeta_i)$  pairs as follows.

**AMBER energy refinement.** The **1DYANA(a-e)** and the BCE molecular model of the 5'-d(GCGAAAGC)-3' hairpin (**2BCEopt\_nmr**) were energy refined with AMBER<sup>50,51</sup> (**1DYANA(a-e)\_min** and **2BCEopt\_nmr\_min**) as explained in previous work<sup>52</sup> under torsion angle restraints until the r.m.s. energy gradient was less than 0.05 kcal/(mol.Å). The force constant was set equal to 900 kcal/(mol.rad<sup>2</sup>). In order to maintain the *C2'-endo* conformation of all sugar puckers, torsion angles,  $\delta_{i,0}$ , (C5'-C4'-C3'-O3') were forced to nominal value,  $\delta_{i,0} = 144^\circ$ .<sup>25</sup> All restrained structures were subsequently relaxed without restraints.

As will be discussed below, structural studies of hairpins<sup>7,19,42</sup> have almost invariably pointed to non canonical values of  $\epsilon$  and  $\zeta$  at the sharp turn that do not correspond to either of the  $B_{I_r}$  or  $B_{II_r}$  local conformations found in helical crystal structures<sup>53</sup>, (cf. Figure 3). The  $\zeta$  value proposed for hairpins is almost invariably in the  $g^+$  to  $^{+}ac$  range. To extend the  $B_I$  and  $B_{II}$  notation to include hairpins we have found it useful to introduce the following conformations of the  $(\epsilon, \zeta)$  pair:  $B_{I_c}$  ( $-178^\circ, -104^\circ$ ),  $B_{II_c}$  ( $-70^\circ, -140^\circ$ ),  $B_{I_{\zeta^+}}$  as ( $-180^\circ, 100^\circ$ ) and  $B_{II_{\zeta^+}}$  ( $-70^\circ, 140^\circ$ ). Some combinations of adjacent torsion angles such as  $g^-/g^+$  or a  $g^+/g^-$  did not have to be considered because they generate well known steric hindrance.<sup>55</sup>

The  $(\epsilon, \zeta)$  torsions of the **2BCEopt\_nmr** structure were systematically modified with AMBER<sup>50,51</sup> to determine the favorable conformations of the G3pA4 (no restraints,  $B_{I_c}$ ,  $B_{II_c}$  and  $B_{II_{\zeta^+}}$ ) and the A5pA6 (no restraints,  $B_{I_c}$ ,  $B_{II_c}$  and  $B_{I_{\zeta^+}}$ ) steps. This procedure afforded the energy-minimized structures

### **3MIN(a-r).**

**AMBER molecular dynamics trajectories.** Molecular dynamics simulations were performed using the AMBER 8 package and the PARM94 force field.<sup>51</sup> This force field has been the most successful and the most used for nucleic acids. It is the basis of all subsequent AMBER force fields and has been extended to include larger sets of organic molecules. The 5'-d(GCGAAAGC)-3' hairpin was placed in a box that contained 2350 TIP3P water molecules (corresponding to a 12-Å hydration shell), 10 K<sup>+</sup> and 3 Cl<sup>-</sup> ions (corresponding to a concentration close to 0.25 M of added KCl). Target temperature and pressure were set at 298 K and 1 atm, respectively. The simulation protocols and positions of the ions were identical to those described by Auffinger *et al.*<sup>56,57</sup> Thus, the particle mesh Ewald (PME) summation method was used for the treatment of long-range electrostatic interactions. The chosen charge grid spacing is close to 1 Å and a cubic interpolation scheme was used. A cut-off of 9 Å for the van der Waals interactions and the Berendsen coupling scheme with a time constant of 0.4 ps were used. The standard PME parameters defined by AMBER led to an average Ewald error of 0.0001. Each trajectory was run with a 2-fs integration time step by using SHAKE bond constraints.

The initial structure for the MD trajectory corresponded to the conformer **3MIN(o)** obtained with the modified set of  $\Omega_i$  and  $\Delta\chi_i$  values established for the GAAA hairpin. The final MD trajectories were implemented without restraints with the exception of a weak restraint (2.5 kcal mol<sup>-1</sup> rad<sup>-2</sup>) on one  $\delta$  torsion angle (A4). The equilibrium phase lasted 400 ps after which 10 ns MD trajectories were generated. Molecular structures were recorded every 0.5 ps for analysis. The MD runs are presented in detail in the Results Section.

## **Results**

**Resonance assignments.** The spin systems were identified using the 2D 2QF-COSY, TOCSY, [<sup>13</sup>C, <sup>1</sup>H]-HSQC, and [<sup>31</sup>P, <sup>1</sup>H]-COSY experiments.<sup>22</sup> At 25°C, only the signals of the 5'-d(GCGAAAGC)-3'

hairpin are detected in the 1D and 2D spectra in D<sub>2</sub>O. The <sup>1</sup>H chemical shifts (69) of all non-exchangeable protons, the <sup>13</sup>C (42) and the <sup>31</sup>P (7) chemical shifts (Figure 4b) were determined and only the H2 of A4, A5 and A6 could not be unequivocally assigned due to superposition of aromatic proton signals. All the sequential <sup>1</sup>H assignments were obtained using standard protocols based on the observation of <sup>1</sup>H-<sup>1</sup>H NOEs and the detection of H3'(i-1)/P(i) and H5',H5'',H4'(i)/P(i) correlations. An expansion of the region of the NOESY spectrum ( $\tau_m$  100 ms) containing the H6,8( $\omega_2$ )/H1'( $\omega_1$ ) intra-residue and sequential cross-peaks is given in Figure 4a.

The H1 (G1, G3, G7) and H41/H42 (C2, C8) chemical shifts were obtained from the exchangeable proton 1D and NOESY spectra recorded at 1°C in a 90/10 mixture of H<sub>2</sub>O/D<sub>2</sub>O. Several other unassigned resonances were detected in the 6.0-13.0 ppm region at this temperature and could be classified according to their intensity (either full intensity or roughly one-third the intensity of the hairpin signals). The vast majority of the signals at 1° C could be accounted for by a mixture of major (hairpin, 75 %) and minor species (duplex, 25 %).

The chemical shifts of the 5'-d(GCGAAAGC)-3' hairpin have been deposited in the Biological Magnetic Resonance Bank (accession number 15898) whereas the partial chemical shift data of the minor species are collected in Table S1.

**Structures features derived from <sup>1</sup>H, <sup>13</sup>C and <sup>31</sup>P chemical shifts.** The NMR data of the 5'-d(GCGAAAGC)-3' sequence at 25°C reproduce the very characteristic signature of 5'-d(..PyGN<sub>n</sub>APu..)-3' hairpins:<sup>19,42</sup> (i) the pyrimidine H2' signal is shifted to high field (1.585 ppm), (ii) the H2'(G3) signal resonates at lower field than H2''(G3) (2.733 and 2.629 ppm, respectively), which is an inversion of the usual chemical shift order, and (iii) the H4' signal of the second loop nucleotide is shifted to high field (3.496 ppm).

The chemical shifts of the H1(G1) (extrapolated to 13.00 ppm at 25 °C) and H1(G7) (13.05 at 25 °C) exchangeable protons were characteristic of imino protons engaged in hydrogen bonds and could be extrapolated to N1-N3 distances of 3 Å.<sup>58</sup> The imino signal of G3 (10.7 ppm) was in the chemical shift

range of NH protons not engaged in hydrogen bonds and its chemical shift was analogous to those reported for imino protons of sheared G·A pairs.<sup>24,46</sup>

All the non terminal C3' (78.35-79.70 ppm) and C5' (67.28-68.73 ppm) sugar carbon signals of the 5'-d(GCGAAAGC)-3' tetra-loop hairpin resonated in the low field regions typical of sugars with C2'-*endo* pucker. Correlation of CP-MAS <sup>13</sup>C chemical shifts with sugar pucker had shown that C2'-*endo* conformers exhibit significant downfield shifts ( $\Delta\delta$  between 5 and 8 ppm as observed here) of the C3' and C5' deoxyribose signals compared to C3'-*endo* conformers in crystalline deoxyribonucleosides and deoxyribonucleotides.<sup>59,60</sup>

The <sup>31</sup>P chemical shifts of the d(GCGAAAGC)-3' hairpin varied from -4.18 (P8) to -4.95 (P3) ppm with the exception of the P(4) signal which showed a marked downfield shift ( $\delta_p$  -3.81 ppm), Figure 4b. <sup>31</sup>P chemical shifts in the range from -4.0 to -5.0 ppm are found in phosphates in regular A- or B-form structures where both  $\zeta$  and  $\alpha$  are *g*<sup>-</sup>.<sup>24,26,29</sup> Such data have frequently been used to exclude the *t* domain of both  $\zeta$  and  $\alpha$  torsion angles as ( $\zeta(g^-), \alpha(t)$ ) and ( $\zeta(t), \alpha(g^-)$ ) conformations are associated with downfield phosphorus chemical shifts.<sup>24</sup> To evaluate the possible interpretations of downfield <sup>31</sup>P chemical shifts and to explore the consistency of the NMR data at the G3pA4 step of d(GCGAAAGC)-3' we explored different subsets of the NMR-derived information with DYANA as outlined in the experimental section.

**Sugar puckering and torsion angles derived from 2D NMR.** The relative values of the vicinal homonuclear coupling constants of the deoxyribose protons were estimated from the intensity of the cross-peaks in the phase-sensitive DQ-COSY spectrum (Table S2). The H1'/H2' cross-peaks were more intense than the H1'/H2'' ones and multiple phase changes in the H1'/H2'' cross-peaks as compared to a single change in phase for the H1'/H2' cross-peaks corroborated that <sup>3</sup>J<sub>1,2''</sub> was smaller than <sup>3</sup>J<sub>1,2'</sub>. This limits all deoxyribose pseudorotational phase angles to 90°-190° with the exception of G7 (which presented degenerate chemical shifts for the 2'/2'' methylene protons) and the 3'-terminal nucleotide C8 (which was suspected to be undergoing some conformational averaging of its deoxyribose). The very

weak H2''/H3' and H3'/H4' (slightly stronger than the H2''/H3' ones in the case of G1, C2, G3, and A5) cross-peaks confirmed that the deoxyribose rings of the first six nucleotides adopted *C1'-exo* or *C2'-endo* sugar pucker. Inspection of the intra-sugar inter-proton distances derived from the NOESY cross-peak volumes revealed the order  $d(1'2'') < d(1'2') \sim d(1'4') \leq d(2''4')$  and is consistent with the sugar puckering inferred from the DQ-COSY spectrum.<sup>22</sup> The *anti* orientation of the  $\chi$  torsion angle of all the nucleotides could be determined from the intranucleotide base proton NOEs with the sugar protons ( $d(\text{H6}/8,1') \ll d(\text{H6}/8,\text{H2}'') < d(\text{H6}/8,2')$ ).

Initially, the constraints for the  $\beta$ ,  $\gamma$  and  $\epsilon$  torsion angles were obtained from the sum of the H5'/H5'', H4', and H3' coupling constants in the [<sup>31</sup>P, <sup>1</sup>H]-COSY spectrum, ( $\Sigma J_{\text{H5}'}$  and  $\Sigma J_{\text{H5}''}$ ,  $\Sigma J_{\text{H4}'}$ , and  $\Sigma J_{\text{H3}'}$ , respectively, Table S3). In particular, it could be deduced that the <sup>3</sup>J<sub>H,P</sub> coupling constants of the 5'/5'' methylene protons of A4, A5 and G7 were less than 10 Hz (i.e.  $\Sigma J_{\text{H5}'}$  and  $\Sigma J_{\text{H5}''}$  both less than 22 Hz) indicating that the  $\beta$  torsion fell within the 105°-255° range. H4'(n)/P(n) cross-peaks were detected for residues C2, G3, A4 (moderate), A5 (very weak) and G7 (moderate) indicating that the four consecutive bonds in the H4'-C4'-C5'-O5'-P backbone lie in the same plane forming a W-shaped conformation.<sup>61</sup> This is the case in B-DNA where  $\beta$  and  $\gamma$  adopt *t* and *g*<sup>+</sup> orientations, respectively. Finally,  $\Sigma J_{\text{H3}'}$  were < 13 Hz for the G3, A4, and G7 H3'(i-1)/P(i) crosspeaks in the [<sup>31</sup>P, <sup>1</sup>H]-COSY spectrum indicating  $\epsilon$  values between 125° and 355°. The H3'(A4)/P(A5) correlations were missing (<sup>3</sup>J<sub>H,P</sub> < 1 Hz) restricting  $\epsilon(4)$  to the 140°-160° range (the 310°-330° range is excluded for steric reasons).<sup>30</sup>

All these excluded torsion angle ranges for  $\alpha$ ,  $\beta$ ,  $\gamma$ ,  $\delta$ ,  $\epsilon$ ,  $\zeta$  and  $\chi$  are summarized in Figure 5 by red wedges whereas the allowed torsion angle domains are depicted by white wedges (overlaid by a black line representing the MD trajectories as discussed below). These NMR-defined torsion angle restraints, combined with upls from intranucleotide and sequential NOEs, were then used for grid searches over the dinucleotide conformational spaces, using the FOUND module of DYANA. As has been shown earlier<sup>62</sup> for a chimeric hybrid duplex, this protocol leads to additional restraints for many torsion angles (Figure 5, excluded ranges in orange).

**DYANA simulations.** The first two sets of DYANA simulations, **1DYANA(a)** (without lols) and **1DYANA(b)** (with lols) did not include any  $^{31}\text{P}$ -derived  $\alpha$  or  $\zeta$  restraints. Due to spectral overlap of P(2) and P(6) (Figure 4b), the cross-peaks that characterize the  $\beta(2)$  and  $\beta(6)$  torsion angles in the  $[\text{}^{31}\text{P}, \text{}^1\text{H}]$ -*COSY* spectrum were not initially interpreted in terms of torsion angle constraints. The results of these first calculations (*i.e.* the  $g^-$  and  $t$  values of  $\beta(6)$  and  $\gamma(6)$ ) made it possible to simulate the 1D spectrum of the  $-\text{P}(6)\text{-H5}'(6)\text{-H5}''(6)\text{-H4}'(6)-$  spin system taking into account the strong coupling of geminal H5' and H5''. The excellent fit between the 1D intensity profile and the P(6) cross-peaks in the  $[\text{}^{31}\text{P}, \text{}^1\text{H}]$ -*COSY* corroborated the unusual values of  $\beta(6)$  and  $\gamma(6)$  calculated in **1DYANA(a,b)**. In the last three simulations,  $\alpha$  and/or  $\zeta$  restraints were introduced progressively to interpret the  $^{31}\text{P}$  chemical shifts as explained in Materials and Methods. The torsion angle data for the **1DYANA(a-e)** ensembles have been collected in Table 1 (noncanonical values in bold).

The 26 upls and 10 torsion angle constraints per residue obtained by NMR were sufficient to define 47 out of the 51 torsion angles values that describe the 5'-d(GCGAAAGC)-3' hairpin structure. Indeed, the average spread of these values was  $\theta_i \pm 13^\circ$  for the five **1DYANA(a-e)** ensembles and included 39 very well-defined ( $\theta_i \pm 10^\circ$ ) and 8 moderately well-defined torsions ( $\theta_i \pm 24^\circ$ ). However, the  $(\epsilon, \zeta)$  pairs of the G3pA4 and A5pA6 steps ( $\theta_i \pm 54^\circ$ ) varied according to the different interpretations of the  $^{31}\text{P}$  chemical shifts:  $\text{B}_I$  or  $\text{B}_{II\zeta^+}$  conformation for the former linkage and  $\text{B}_{II}$ ,  $\text{B}_I$  or  $\text{B}_{II\zeta^+}$  for the latter one. The  $\text{B}_{II\zeta^+}$  conformation could be compatible with the  $^{31}\text{P}$  chemical shift of either P(4) (with  $\zeta(3)$  closer to  $t$  than to  $ac^+$ ) or P(6) (with  $\zeta(5)$  closer to  $ac^+$ ) whereas the  $\text{B}_I$  (**1DYANA(a,c,d)**) and  $\text{B}_{II}$  (**1DYANA(a,b)**) conformations at the G3pA4 and A5pA6 steps, respectively, are not compatible with the downfield shift of P(4) and the upfield shift of P(6). The best agreement between the average torsion angles of the DYANA ensembles and the NMR restraints is observed for the **1DYANA(e)**.

Separate simulations showed that restraining  $\alpha(4)$  to the  $t$  conformation ( $180 \pm 40^\circ$ ) as an alternate explanation for the downfield shift of P(4) was incompatible with the NMR data. Under these conditions, simulations led to numerous distance violations and strong torsion angle violations ( $>15^\circ$ ) of

$\alpha(4)$ .

The statistics for all structure calculations, **1DYANA(a-e)**, were very similar (target functions all  $< 0.35 \text{ \AA}^2$  and RMSD for the backbone heavy atoms  $< 0.6 \pm 0.2 \text{ \AA}$ ) and those of the **1DYANA(e)** ensemble have been collected in Table 2. The pairwise RMSD of all heavy atoms of all nucleotides was  $0.3 \pm 0.1 \text{ \AA}$  and the residual target function was  $0.3 \text{ \AA}^2$ . All of the simulations converged reasonably well ( $> 20$  structures) except **1DYANA(d)** that only contained 18 converged structures. The  $\alpha$  and  $\gamma$  dihedral angles, that are  $g^- (-62 \pm 15^\circ)$  and  $g^+ (48 \pm 11^\circ)$  in B-DNA, tend towards the *cis* range ( $-30^\circ$  to  $+30^\circ$ ) in some of the DYANA ensembles, particularly in the simulations without lols (five in **1DYANA(a)**, four in **1DYANA(d)**, and the most-deformed,  $\alpha(6) -14^\circ$ , in **1DYANA(c)**). These anomalies disappear upon minimization with AMBER. In the case of the **1DYANA(e)** ensemble, minimization also leads to  $149^\circ$ - and  $129^\circ$ -values of the average  $\zeta$  torsions of the G3pA4 and A5pA6 steps, respectively, in excellent agreement with the  $^{31}\text{P}$  chemical shifts. A systematic comparison between the NMR data calculated for the **1DYANA(a-e)** structures (upls, lols, torsion angles, and high-field-shifted proton chemical shifts<sup>64</sup> as described in the following paragraph) and the experimental data has been established with R-factors<sup>65,66</sup> that have been defined in the Supplementary Materials (S4). The corresponding values for all the structures described in this paper are collected in Table S5.

Ring current effects that account for the characteristic high-field shifts of H2'(C2) ( $\delta$  1.58 ppm compared to the average H2' value of 2.48 ppm) and H4'(A4) ( $\delta$  3.49 ppm compared to the average H4' value of 4.30 ppm) provide an independent experimental probe of the five DYANA ensembles. These parameters were calculated for the **1DYANA(a-e)** ensembles from the corresponding Cartesian coordinates using the program NUCHEMICS.<sup>64</sup> Small high-field H2'(C2) and H4'(A4) shifts were predicted for all of the NMR ensembles (Table S5) but the best overall agreement was obtained for **1DYANA(e)\_min**, Table 3. The shielding effects on the H2'(C2) and H4'(A4) spins can be attributed to their locations below the five-membered G3 ring and above the six-membered A6 ring, respectively.

In conclusion, complete interpretation of all the  $^{31}\text{P}$  chemical shifts in terms of torsion angle restraints

for the d(GCGAAAGC)-3' sequence leads to a converged structural ensemble, **1DYANAe**. As presented in Figure 1 and below, the internal consistency of these conformations were explored at the mesoscopic level with the BCE approach, and their fine structural ambiguities at the torsion angle levels were resolved with AMBER.

**BCE/AMBER structures.** The  $(\Omega_i, \Delta\chi_i)$  pairs that define the orientation of each nucleotide of the GTTA hairpin (1ac7)<sup>19</sup> with respect to the elastic line in the BCE approach had been obtained previously<sup>21</sup> from the NMR-defined coordinates. As illustrated in Figure 2b for the GAAA hairpin, these parameters were slightly modified to enhance the favorable stacking interactions on the 5'-side revealed by NMR as follows.  $\chi_3$  of G3 was increased from 14 to 25° to facilitate the G·A base pairing, and the values  $(\Omega_5, \chi_5)$  are changed from (102.8°, 26.6°), which oriented the third loop nucleotide of GTTA in the solvent, to (90°, -25°) to allow stacking of A5 onto A4 in GAAA. In practice this is the major mesoscopic change. The complete list of  $(\Omega_i, \chi_i)$  values leading to the conformation **2BCEopt\_nmr** is given in Figure 2b. Energy minimization *in vacuo* to restore bond lengths and valence angles yielded a minimized conformation (-124.4 kcal mol<sup>-1</sup>) close to the **2BCEopt\_nmr** shown in Figure 2a-3. This construction/optimization process can be followed by calculating R-factors that measure the agreement between the experimental NMR data and those of the model structures (**BCEopt()**) at each step as indicated in Table 3. The NMR data predicted for the BCE models improve significantly in the initial stages of the optimization. As expected, the torsion angles values ( $R_2$ ) do not show good agreement with the NMR constraints, since they remain as close as possible to B-DNA canonical torsion angles values during the complete optimization process described in Figure 2b, and assessed in Table 3 (2nd section). In particular, the  $(\epsilon, \zeta)$  pairs of the G3pA4 step can not explain the downfield shift of the G3 phosphorous.

To resolve the fine structural ambiguities explained above, systematic exploration of the  $\epsilon$  and  $\zeta$  torsions of the G3pA4 and A5pA6 steps ( $B_I, B_{II}, B_{I\zeta+}$  and  $B_{II\zeta+}$  as outlined in the Methods section) with AMBER minimization afforded eighteen BCE conformers **3MIN(a-r)** in the energy range -124.4 to -

110.1 kcal mol<sup>-1</sup>. All but four of the **3MIN(a-r)** conformers contained B<sub>I $\zeta$ +</sub> or B<sub>II $\zeta$ +</sub> orientations of the ( $\epsilon,\zeta$ ) pairs of some of the loop nucleotides (G3pA4 and/or A4pA5 and/or A5pA6 steps) whereas the four conformers with classical B<sub>I</sub> or B<sub>II</sub> orientations of all the nucleotides exhibited higher energy (–114.0 to –110.1 kcal mol<sup>-1</sup>). Significant modifications of several torsion angles of the starting structure, **2BCEopt\_nmr** (that were inconsistent with the NMR data), were observed in all of the minimized structures (explaining the relatively large R<sub>2</sub> values) with the exception of the **3MIN(o)** conformer, which remained stable. This latter structure (B<sub>II $\zeta$ +</sub> orientations for both the G3pA4 and A5pA6 steps) was able to reproduce all the available NMR data (as can be seen from the R-factors in Table 3) with an *in vacuo* energy, –117.6 kcal mol<sup>-1</sup>. The agreement between the experimental data and those calculated for the other **3MIN(a-r)** structures was poorer and the corresponding R-factors<sup>65,66</sup> are also given in the Supplementary Materials.

The stereoscopic view of the **3MIN(o)** BCE structure (21<sup>st</sup> model, PDB 2k71) after minimization with AMBER (red) and the superposition of the 20 **1DYANA(e)** (first 20 models, PDB 2k71) structures that best reproduce the NMR constraints after minimization with AMBER (blue) and is given in Figure 6.

**Structures from the molecular dynamics trajectories.** We performed many detailed state-of-the-art molecular simulations in explicit solvent and salt to test the starting conformations proposed by the DYANA procedure alone, the BCE/AMBER procedure alone, or by both together. Initial conformations produced by DYANA alone fitted well the NMR data, but were usually characterized by a high overall energy due to distorted backbone or base-pairs, and some incompatible torsion angle values. Those from the BCE/AMBER procedure alone had lower energies and less base or torsion angle distortions, but did not fit as well the NMR data, particularly detailed torsion angle values. Both types of structures yielded fair molecular dynamics simulations where the overall hairpin conformation was preserved. However, close inspection revealed some dynamical structural instability such as A4/A5 de-stacking, A5 base rotation from *anti* to *syn*, and different backbone torsion angle discrepancies such as conformation changes from C2'-*endo* sugar pucker to C3'-*endo*, or from B<sub>I</sub> to B<sub>II</sub> for several nucleotides.

Different strategies were attempted to obtain structurally stable molecular dynamics over more than 500 ps. The first consisted in constraining poorly-behaved torsion angles with sufficiently low force constraints to allow for atomic motions. In doing so, we observed two distinct phenomena:

(i) a so-called “butterfly effect”, well-known in chaos phenomena, *i.e.* a sensitive dependence on initial conditions. Two MD runs would diverge sharply after a long time ( $\sim 1000$  ps or 500000 integration steps), *e.g.* with the rotation, or not, of the A5 base from *anti* to *syn*, upon introducing a small constraint on  $\delta$  (A4).

(ii) the former phenomenon is clearly related to the second observation, where the introduction of a single small constraint on a backbone torsion angle in the loop (2-3.5 kcal/(mol.rad<sup>2</sup>) is sufficient to modify one or a series of several torsion angles distant by one or several nucleotides.

Combining NMR/DYANA and BCE/AMBER approaches lead to conformation **3MIN(o)**, with minimal backbone deformation from standard B-DNA, and good G·A base pairing as provided by BCE/AMBER, together with the modified torsion angles at the loop sharp turn and with well-stacked A4/A5 bases as indicated by NMR. As shown in Figure 5, these features were retained during the 2500-ps production period of the trajectory initiated with **3MIN(o)**. Comparison of the torsion angle ranges excluded by NMR (red and orange) and the MD trajectories (black) in Figure 5 shows that the latter simulations visit virtually only the experimentally-allowed conformations (white).

## Discussion

**Comparison of the <sup>31</sup>P chemical shifts with DNA hairpin folding.** The <sup>31</sup>P chemical shifts of published DNA triloops<sup>42,67-73</sup> and tetraloops<sup>19,75-77</sup> were compiled to probe the relation between these data and DNA hairpin folding (Tables S6 and S7). The fragment under scrutiny was restricted to the loop nucleotides and the adjacent base pair in the stem (N<sub>5</sub>- L<sub>1</sub>L<sub>2</sub>L<sub>3</sub>L<sub>4</sub>-N<sub>3</sub>, for tetraloops and N<sub>5</sub>- L<sub>1</sub>L<sub>2</sub>L<sub>3</sub>-N<sub>3</sub>, for triloops). The average value of the <sup>31</sup>P chemical shifts ( $\delta_p^{av}$ ) and the half width of the interval

$(\frac{1}{2}\Delta\delta_p^{\max})$  excluding an eventual outlier ( $|\Delta\delta| > 0.7$  ppm) were established. The chemical shift and position in the sequence of any outlier ( $\delta_{\text{outlier}}$ ) were also collected in the Tables. Finally, the overall chemical shift pattern was defined by the most upfield or downfield group of phosphate signals.

For 5'-d(GCGAAAGC)-3' these data are as follows:  $\delta_p^{\text{av}}$  is  $-4.71 \pm 0.24$  ppm, the chemical shift of the second loop nucleotide is an outlier ( $\delta_{L_2} -3.81$  ppm), and the  $L_1$  and  $N_3$  phosphates resonate at highest field ( $\delta_p = -4.95$  and  $-4.87$  ppm, respectively). Except for  $L_2$ , the observed average chemical shift and half interval values are very reminiscent of the those of several well-defined DNA triloops ( $\delta_p^{\text{av}}$  of CAAAG-,<sup>67</sup> CGAAG-,<sup>42</sup> and TGCAA-<sup>70</sup> are  $-4.65 \pm 0.40$ ,  $-4.82 \pm 0.24$ , and  $-4.71 \pm 0.31$  ppm, respectively) and in all cases the  $L_1$  and  $N_3$  phosphates *at the junction of the sheared pair in the loop and the stem* resonate at highest field ( $\delta_p < -4.85$  ppm).

All these chemical shifts are related to similar backbone conformations. The sharp turns in the latter DNA triloops are located between the penultimate and the last loop nucleotides as in the present study. Moreover, most of the torsion angles that adopt noncanonical conformation to produce the change in the direction of the phosphate backbone of these triloops ( $\epsilon, \zeta(L_2)$   $^-ac/t, ^+ac$  and  $\alpha, \gamma(L_3)$   $g^+, t$ ) adopt similar conformations in the GAAA tetraloop ( $\epsilon, \zeta(L_3)$   $^-ac, ^+ac$  and  $\beta, \gamma(L_4)$   $g^-, t$ ). The ( $\epsilon, \zeta$ ) conformation of the  $L_2$  nucleotide in these triloops is in the  $B_{I\zeta^+}$  to  $B_{II\zeta^+}$  range ( $g^-$  to  $t$ ,  $g^+$  to  $t$ ). The common feature of all these DNA triloop hairpins and of the GAAA tetraloop is maximal stacking at the stem-loop interface and in the loop.

Many other DNA hairpins such as CGAGG-,<sup>67</sup> CGATG-,<sup>69</sup> CTTTG-,<sup>71</sup> and GTTTC-<sup>73</sup> triloops and the CGTTAG,<sup>19,75</sup> ATTTAT,<sup>75,76</sup> CTTTGG,<sup>77</sup> and GTTTTC<sup>73</sup> tetraloops are characterized by much lower  $\delta_p^{\text{av}}$  values ( $-4.15 \pm 0.15$ ,  $-4.27 \pm 0.13$ ,  $-4.09 \pm 0.36$ ,  $-4.01 \pm 0.33$ ,  $-3.90 \pm 0.42$ ,  $-3.38 \pm 0.52$ ,  $-4.07 \pm 0.26$ ,  $-4.08 \pm 0.31$ , and  $-4.26 \pm 0.29$  ppm, respectively) and present very different  $^{31}\text{P}$  chemical shift profiles. Most of the structures proposed for this group of hairpins contain nucleotides with bases that fold into the minor and/or major grooves and the corresponding  $\alpha$  torsion angle is often in the  $t$  conformation. The vast majority of the signals of such phosphates are shifted to low field. The  $\delta_p^{\text{av}}$

values of this group are analogous to those of DNA duplexes ( $-3.88 \pm 0.31$  and  $-3.93 \pm 0.32$  ppm;<sup>29</sup>  $-4.16 \pm 0.32$  to  $-4.28 \pm 0.25$  ppm<sup>26</sup>). Finally, two other triloops,  $-\text{CACAG}-$ <sup>68</sup> and  $-\text{TATCA}-$ ,<sup>69</sup> present high  $\delta_{\text{p}^{\text{av}}}$  values ( $-4.51 \pm 0.30$  and  $-4.62 \pm 0.16$  ppm) but low-field shifted outliers that are related to phosphate conformation in the stem ( $\delta_{\text{N3}'}$ ,  $-3.28$  and  $-3.75$  ppm, respectively).

The much greater dispersion of RNA hairpin <sup>31</sup>P chemical shifts ( $\Delta\delta$ s of the CUUG,<sup>78</sup> GAAA,<sup>79</sup> and UUCG<sup>80</sup> hairpins are 3.60, 3.10 and 2.63 ppm, respectively) has been a major impediment to exploiting <sup>31</sup>P chemical shifts in structural studies of nucleic acids.<sup>24,78</sup> This survey points to a strong correlation between the <sup>31</sup>P chemical shift pattern and the mesoscopic conformation of the loop nucleotides in DNA hairpins with high-field  $\delta_{\text{p}^{\text{av}}}$  values.

**The  $B_{\text{II}\zeta+}$  conformation observed in related hairpins.** The average values of the  $(\epsilon, \zeta)$  torsion angle pairs of the 2.5-ns MD trajectory are shown in Figure 5. All but those of the G3pA4 and A5pA6 steps are located in the 95% confidence interval depicted by the  $B_{\text{Ir}}$  ellipse in Figure 3. These two outliers are also outside the 95% confidence interval depicted by the  $B_{\text{Ir}}$  ellipse. For comparative purposes the  $(\epsilon, \zeta)$  torsion angle pairs of recent helical crystal structures<sup>54</sup> have been superimposed on the  $(\epsilon, \zeta)$  plot in Figure 3. The majority of all these structures correspond to favorable staggered conformations (white and green) on the  $(\epsilon, \zeta)$  map. Two regions can be distinguished that englobe the  $(\epsilon, \zeta)$  torsion angle pairs of the sharp turns of hairpin structures. The  $B_{\text{I}\zeta+}$  one includes those of the DNA GTTA tetraloop, (PDB 1ac7),<sup>19</sup> and the AAA (PDB 1xue),<sup>67</sup> the GCA (PDB 1bjh),<sup>70</sup> and the GAA (PDB 1pqt)<sup>42</sup> triloops. The  $B_{\text{II}\zeta+}$  one includes the  $(\epsilon, \zeta)$  torsion angle pairs of the G3pA4 and A5pA6 steps of the GAAA tetraloop (PDB 2k71) and is located at the limit of the  $B_{\text{Ir}}$  region.

**Comparison to solid-state structure.** In the course of the present work, analysis of the low temperature spectra indicated the presence of a second species below 5 °C and the corresponding chemical shifts pointed to a classical mismatched duplex with sheared G·A base pairs. Crystallographic data recently reported for 5'-d(GCGAAAGC)-3'<sup>81</sup> demonstrated the presence of a base-intercalated anti-parallel duplex in the solid state corroborating the existence of double-stranded DNA GAAA sequences

such as the one detected at low temperature.

The DNA GAAA hairpin is extraordinarily stable and yet two species co-exist at low temperatures, the hairpin and a mismatched duplex. It suggests that both conformations are very stable, and that this particular sequence could behave as a molecular switch.

**The combined approach NMR/DYANA and BCE/AMBER, with MD.** Solving the fine structure of a DNA or RNA molecule by NMR is a difficult task for several reasons. Very stable triloops and tetraloops hairpins adopt very compact structures with very finely adjusted base-pairing and stacked conformations. Use of information from NMR and derived from the DYANA ensembles, was followed by the exploration of the ambiguous torsion angle conformations with model structures **3MIN(a-r)**. This was made possible because the BCE construction/optimization process (**BCEopt()** structures) presented in Figure 2b and assessed in Table 3 is remarkable for three main reasons. Firstly, nucleotides can be rotated at will about the elastic line representing the sugar-phosphate backbone to set the nucleotide in any given conformation, *e.g.* stacked, in one of the grooves, or in the solvent. In the resulting conformations, all nucleotides are well positioned in space so as to reproduce all NMR distance constraints through a very small set of independent rotation angles ( $\Omega_i, \chi_i$ ). We believe that this is the first molecular modeling approach that can achieve such independent rotations in a hairpin loop. It simultaneously provides insight both into the molecular conformations and into their fit to NMR data.

Furthermore, these constructions are endowed with two essential advantages. They possess the least deformed sugar-phosphate backbones and therefore the least modified B-DNA torsion angles. They can be chosen as the reference state, from which departures are studied. In addition they are practically free from steric hindrance, which is very remarkable for such compact conformations. These two features are mandatory for systematic investigation of the different possible combination of torsion angles by energy minimization with AMBER, *i.e.* a meaningful exploration of small differences between different conformations without being hampered by high energy steric hindrance.

As a result, we have been able to identify the single combination of ( $\epsilon, \zeta$ ) pairs in the loop region, and

therefore the B<sub>IIz+</sub> conformation of G3 and A5 structure that matched all NMR constraints (best R-factors in Table 3). It is to be noted that the NMR data (*i.e.* a slightly weaker H2''(A4)/H8(A5) NOE than expected) would be compatible with a minor population corresponding to partial de-stacking of A5 or enhanced internal motions in the loop nucleotides. In agreement with these dynamics, dihedral angle transitions accompanied by some protruding of the A5 base into the solvent were observed during the 7.5 ns following the 2.5-ns production period in the 10-ns MD trajectory described above.

As observed in the course of this work, modifying a few torsion angles may have severe unwanted consequences upon the entire backbone during energy refinement. Consequences upon the molecular dynamics are even more drastic and unpredictable since the change of a single torsion angle value to its correct observed value (*e.g.*  $\delta$  A4) may also modify base stacking in the loop. This situation is due to the fixed end conditions of the loop and steric hindrance in this region that generate complex relationships between the backbone torsion angles.

This dilemma was resolved with the combined simulated annealing (DYANA) and BCE/AMBER approach. We observe that the minimized unmodified *in vacuo* BCE conformation has the lowest energy. This conformation results from two different energy optimizations at two different scales: BCE at the mesoscopic scale of several nucleotides, and AMBER at the atomic scale. It is well known that AMBER force fields<sup>51</sup> and subsequent modifications perform well with double helical conformations.<sup>82</sup> This may be why the closest conformation to helical B-DNA has the lowest energy. Moreover we observe that, except for the restraint on  $\delta$  (A4), the force field performs very well over a long-term molecular dynamics (2500 ps). This suggests that energy minimization alone might not be sufficient to evaluate such tight conformations as hairpin loops. It appears that the two major changes in torsion angles with respect to B-DNA, namely the correct  $\beta(6)$  and  $\gamma(6)$  orientations, must be included in the simulations to provide sufficient dynamic stability and conformational freedom both at the level of atomic and overall backbone motions. The molecular conformations determined in this work should provide yet another case study to test new force field developments.

## Conclusions

The major family of conformers for the 5'-d(GCGAAAGC)-3' sequence has been determined by simulated annealing (NMR/DYANA) and molecular modeling (BCE/AMBER). Only two torsion angles  $\beta(6)$  and  $\gamma(6)$  ( $g^-$  and  $t$ , respectively), deviate significantly from B-DNA values while some averaging about the  $B_{II\zeta+}$  conformation is occurring at the G3pA4 and A5pA6 steps. A survey of the literature has revealed that this latter conformation or the  $B_{I\zeta+}$  one are regularly encountered in DNA hairpins.

Automatic comparison of NMR data calculated for BCE conformers with the experimental data (i.e. R-factors for upls, lols, torsion angles, and  $^1\text{H}$   $\Delta\delta$ s) during both the construction/optimization phase and the systematic exploration of conformations about the  $(\epsilon, \zeta)$  pairs of the G3pA4 and A5pA6 steps was unequivocal. Only the **3MIN(o)** structure was able to reproduce all the NMR data. Furthermore, AMBER minimization and MD trajectories indicated that it corresponded to both a local and global minimum. Thus, the downfield-shifted P4 signal (-3.81 ppm) is associated with the  $B_{II\zeta+}$  conformation ( $\zeta$ , 143° and 149° for the **1DYANAe** and **1DYANAe\_min** structures, respectively). With the exception of the P4 phosphate in the  $B_{II\zeta+}$  conformation, DNA loop nucleotides of stable hairpins with maximal stacking at the stem-loop interface and in the loop appear on the contrary to be characterized by a high-field shifted  $^{31}\text{P}$  signals ( $\delta_p^{\text{av}} < -4.6$  ppm).

Resolution of most DNA structures by NMR to-date has suffered from insufficient data. Although  $^{13}\text{C}$  and  $^{15}\text{N}$  labeling is the best approach for obtaining more abundant data, few studies have benefited from this technique because enzymatic incorporation of labeled dNTPs into DNA sequences is much more difficult than for RNA.<sup>83</sup> As has been shown in this work, unusual  $^1\text{H}$  and  $^{31}\text{P}$  chemical shifts can also be translated into additional NMR constraints that facilitate structure determination using NUCHEMICS<sup>64</sup> and systematic BCE exploration of  $(\alpha, \epsilon, \zeta)$  torsion angles, respectively.

In conclusion, this investigation has demonstrated that the BCE approach can generate least-deformed

conformation from B-DNA while automatically monitoring the fit between experimental NMR data and those calculated for the theoretical structures. The development of a plug-in based on the BCE approach would be a tremendous help in NMR structural studies of nucleic acids and this will be the focus of future work.

**Acknowledgment.** C.P. is indebted to the CNRS (France) for funding a sabbatical leave at the University of Buffalo and would also like to express her thanks to Prof. T. Szyperski for stimulating discussions and for access to the NMR facilities in his laboratory where this project was initiated.

**Supporting Information Available:** Partial  $^1\text{H}$  NMR chemical shifts of the low-temperature minor species. Relative intensities of the DQCOSY crosspeaks of the GAAA hairpin. Linewidths in the 1D  $^1\text{H}$  NMR spectrum, crosspeak widths and relative intensities in the 2D [ $^{31}\text{P}$ ,  $^1\text{H}$ ]-COSY spectrum of the GAAA hairpin. R-factor definitions and the corresponding values for all the structures described in this work. Survey of the  $^{31}\text{P}$  chemical shifts of selected DNA triloop sequences. Survey of the  $^{31}\text{P}$  chemical shifts of selected DNA and RNA tetraloop sequences. This information is available free of charge via the Internet at <http://pubs.acs.org>.

## Figure Captions

**Figure 1.** Scheme depicting a general strategy for resolving nucleic acid structure based on NMR data with two different approaches: simulated annealing (DYANA) and BCE construction. All conformers are further refined by AMBER minimization and molecular dynamics.

**Figure 2.** Schematic overview of the construction steps of the GAAA tetraloop hairpin conformation with the Biopolymer Chain Elasticity approach.<sup>20,21</sup> (a) Top: general construction viewed from the major groove, (a-1) generation of two single-stranded helical DNA chains (in red) and of a four-nucleotide

helical segment (in blue) about the vertical Oz axis (dashed line); the segment (blue ribbon) is considered as a continuous flexible thin rod and is bent into the capping elastic solution curve (yellow ribbon); this elastic line is computed such that tangent points and tangents at the loop helix extremities, match those of the two helices (red arrows); (a-2) Frenet trihedron (not shown) computed at fixed curvilinear abscissa of the helical and the elastic rods are used as local reference frames to express atomic coordinates relative to rods trajectories. By use of a simple differential geometry operation the helical segment is folded onto the elastic curve into BCEori loop (in orange). (a-3) BCEopt conformations (in red) are obtained by rotations of the nucleotides,  $i$ , about the tangent to the elastic line with angle,  $\Omega_i$ , and by rotations of the base,  $i$ , with angle,  $\Delta\chi_i$ , as detailed next. (b) Nucleotides and bases orientation parameters. Top: side and top view of each nucleotide in the loop before (orange) and after (red) rotations. Bottom: values,  $\Omega$ , plotted (bold black line) from 5' to 3' ends along curvilinear abscissa of the rod loop, and of base rotation increment,  $\Delta\chi_i$ , about the glycosidic bond (red boxes).

**Figure 3.** Plot of the  $\zeta$  versus  $\epsilon$  torsion angles overlaying the eclipsed (grey) and staggered (white) conformations of a dihedral fragment that represent the favored  $g^-$ ,  $t$ ,  $g^+$  regions. The  $B_{I_r}$  (blue) and  $B_{II_r}$  (red) ellipses encompass 95% of crystal structures.<sup>53</sup> The  $(\epsilon, \zeta)$  pairs of the  $B_{I_c}$  (c, constraint),  $B_{II_c}$ ,  $B_{I_{\zeta^+}}$  (positive  $\zeta$  values, symmetrical about the  $\zeta_{180^\circ}$  axis with respect to  $B_{I_c}$ ) and  $B_{II_{\zeta^+}}$  (positive  $\zeta$  values, symmetrical about the  $\zeta_{180^\circ}$  axis with respect to  $B_{II_c}$ ) constraints are given in red. All the south/east (blue stars) or north (pink squares) conformations in the PDB have been taken directly from ref. 54. The average values of the  $(\epsilon, \zeta)$  pairs in the MD trajectory have been represented with black dots highlighted in white and the positions of the  $(\epsilon, \zeta)$  pairs in the  $B_{I_{\zeta^+}}$  conformation of the AAA,<sup>67</sup> GAA,<sup>42</sup> GCA,<sup>70</sup> GTTA<sup>19</sup> and TTTG<sup>77</sup> hairpins have been indicated in blue letters.

**Figure 4.** (a) Spectral region of the 750 MHz [<sup>1</sup>H,<sup>1</sup>H]-NOESY spectrum (100-ms mixing time) of 5'-d(GCGAAAGC)-3' at 25 °C that contains the H1'(ω<sub>1</sub>)/H6,8(ω<sub>2</sub>) connectivities. The chemical shift positions of the base protons in individual nucleotides are given at the top, and connectivities are

indicated with arrows.  $^1\text{H}$ - $^1\text{H}$  upper distance limit constraints for sequential connectivities are indicated near the corresponding crosspeaks. (b) 160 MHz  $^{31}\text{P}$  NMR spectrum of a 4 mM solution of the 5'-d(GCGAAAGC)-3' hairpin in  $\text{D}_2\text{O}$  at 25 °C.

**Figure 5.** Superposition of NMR constraints with colored sectors and MD time series. In the dial frames, backbone and glycosidic torsion angle values of the six non-terminal nucleotides of 5'-d(GCGAAAGC)-3' increase clockwise with 0 at the top of the dials. Circular plots of dihedral angle constraints for the **1DYANA(e)** simulation and mean values are indicated by a blue radial line. Red sectors correspond to the excluded ranges based on the experimentally determined linewidths associated with characteristic NOEs<sup>22</sup> or the  $^{31}\text{P}$  chemical shifts; orange ones to those excluded by the FOUND module of DYANA using local NOE distance constraints. Average values over the DYANA conformations are given in blue and are indicated by a blue line. Time trajectories of the 2500-ps production period of free molecular dynamics with the single restraint  $\delta$  on A4 (144°, 2.5 kcal/(mol.rad<sup>2</sup>)). Time increases from the center to the circumference and the detailed trajectories are in black. Average values over the molecular dynamics are given in red and depicted with a red line. Starting conformation was the energy minimized BCE conformer **3 MIN(o)** obtained as the resulting conformation of the combined approach DYANA and BCE/AMBER described in scheme 1. Last row: in green, mean values and stdev of high resolution (< 1.9 Å) crystal structures of B-DNA with bimodal distributions,  $B_{\text{Ir}}$  and  $B_{\text{Irr}}$ <sup>53</sup> in red, average and stdev values from molecular dynamics simulations and molecular modeling of NMR derived data ( $B_{\text{Imd}}$ ).<sup>63</sup>

**Figure 6.** Stereoscopic view of the superposition of the 20 **1DYANA(e)** structures minimized with AMBER (blue) (first 20 models, PDB 2k71) that best reproduce the NMR constraints and the BCE structure **3 MIN(o)** minimized with AMBER (red) (21<sup>st</sup> model, PDB 2k71). The BCE elastic line is shown as a yellow ribbon.

**Table 1.** Torsion angles defining the 5'-d(GCGAAAGC)-3' structures obtained in the **1DYANA(a-e)**<sup>a</sup> simulations. Values in bold display significant deviations when compared to those of B-DNA.

Residue	$\alpha$	$\beta$	$\gamma$	$\delta$	$\epsilon$	$\zeta$	$\chi$
C2	-67 ± 6	-173 ± 5	34 ± 1	135 ± 1	-156 ± 2	<b>-136 ± 1</b>	-108 ± 0
	-48 ± 14	-175 ± 12	35 ± 4	148 ± 0	-176 ± 5	-104 ± 10	-114 ± 1
	-55 ± 8	-180 ± 5	33 ± 2	131 ± 1	-167 ± 1	-121 ± 0	-108 ± 0
	-51 ± 9	168 ± 10	33 ± 2	132 ± 1	-156 ± 4	-112 ± 6	-108 ± 0
	-52 ± 1	171 ± 0	37 ± 2	140 ± 0	-175 ± 2	-82 ± 1	-108 ± 0
G3	-27 ± 0	126 ± 1	60 ± 1	127 ± 0	180 ± 0	-120 ± 0	-110 ± 0
	-65 ± 36	175 ± 13	58 ± 24	149 ± 0	-165 ± 0	<b>-155 ± 0</b>	-95 ± 0
	-40 ± 0	141 ± 0	59 ± 0	127 ± 0	-178 ± 0	-116 ± 0	-99 ± 0
	-52 ± 6	146 ± 4	65 ± 3	131 ± 0	-175 ± 0	<b>-131 ± 0</b>	-101 ± 0
	-73 ± 14	-174 ± 3	43 ± 14	141 ± 0	<b>-72 ± 0</b>	<b>143 ± 0</b>	-99 ± 0
A4	-25 ± 0	-176 ± 1	21 ± 0	150 ± 0	173 ± 0	-111 ± 0	-99 ± 0
	-38 ± 0	141 ± 0	46 ± 0	150 ± 0	179 ± 0	-114 ± 0	-107 ± 0
	-47 ± 0	175 ± 0	48 ± 0	149 ± 0	174 ± 0	-106 ± 0	-98 ± 0
	-27 ± 0	169 ± 0	26 ± 0	140 ± 0	177 ± 0	-112 ± 1	-99 ± 0
	-75 ± 0	142 ± 0	29 ± 0	150 ± 0	-178 ± 0	-121 ± 0	-98 ± 0
A5	-25 ± 0	146 ± 0	35 ± 0	120 ± 0	<b>-81 ± 0</b>	<b>169 ± 1</b>	-124 ± 0
	-51 ± 0	160 ± 0	56 ± 0	129 ± 0	<b>-108 ± 0</b>	<b>169 ± 0</b>	-140 ± 0
	-45 ± 0	147 ± 0	65 ± 0	110 ± 0	-178 ± 0	-127 ± 0	-151 ± 0
	-29 ± 2	156 ± 1	30 ± 1	120 ± 0	<b>-64 ± 0</b>	<b>136 ± 0</b>	-145 ± 0
	-35 ± 0	159 ± 0	34 ± 0	126 ± 0	<b>-63 ± 0</b>	<b>133 ± 0</b>	-135 ± 0
A6	-46 ± 2	<b>-63 ± 0</b>	<b>-153 ± 0</b>	119 ± 1	177 ± 0	-100 ± 0	-115 ± 0
	-42 ± 0	<b>-64 ± 0</b>	<b>-154 ± 0</b>	142 ± 0	-176 ± 4	-99 ± 4	-126 ± 0
	-14 ± 0	<b>-48 ± 0</b>	<b>-138 ± 0</b>	134 ± 0	-175 ± 1	-83 ± 0	-111 ± 0
	-66 ± 0	<b>-73 ± 0</b>	<b>-140 ± 0</b>	143 ± 0	-166 ± 1	-101 ± 2	-116 ± 0
	-75 ± 0	<b>-65 ± 0</b>	<b>-139 ± 0</b>	147 ± 0	-167 ± 3	-101 ± 3	-124 ± 0
G7	-27 ± 2	159 ± 2	54 ± 1	110 ± 1	-179 ± 1	-98 ± 1	-134 ± 0
	-44 ± 7	170 ± 4	48 ± 6	148 ± 0	-174 ± 4	-121 ± 9	-106 ± 3
	-45 ± 0	155 ± 0	73 ± 0	106 ± 0	177 ± 0	-94 ± 0	-127 ± 0
	-36 ± 3	167 ± 5	38 ± 3	135 ± 6	-177 ± 6	-120 ± 12	-114 ± 3
	-42 ± 2	177 ± 3	29 ± 3	147 ± 0	177 ± 0	-110 ± 1	-101 ± 1
B-DNA <sup>b</sup>	-62 ± 15	176 ± 9	48 ± 11	128 ± 13	-176 ± 11	-95 ± 10	-102 ± 14
B-DNA <sup>c</sup>	-71 ± 2	177 ± 2	59 ± 1	132 ± 4	(-120)	(-180)	-117 ± 5

<sup>a</sup>**1DYANA(a)** - 208 upls and 69 dihedral constraints without <sup>31</sup>P-defined constraints on  $\alpha$  or  $\zeta$ ; **1DYANA(b)** - 202 upls, 246 lols, and 78 dihedral constraints without <sup>31</sup>P-defined constraints on  $\alpha$  or  $\zeta$ ; **1DYANA(c)** - 208 upls and 67 dihedral constraints including <sup>31</sup>P-defined constraints for B<sub>I</sub>-like  $\alpha$  and  $\zeta$  ( $0 \pm 120^\circ$ ); **1DYANA(d)** - 208 upls and 66 dihedral constraints including <sup>31</sup>P-defined constraints for B<sub>I</sub>-like  $\alpha$  and  $\zeta$  ( $0 \pm 120^\circ$ ) and  $\zeta$  constraints for the G3pA4 step ( $180 \pm 40^\circ$ ); **1DYANA(e)** - 206 upls and 246 lols and 73 dihedral constraints including <sup>31</sup>P-defined constraints for B<sub>I</sub>-like  $\alpha$  and  $\zeta$  ( $0 \pm 120^\circ$ ) and  $\zeta$  constraints for the G3pA4 step ( $180 \pm 40^\circ$ ). <sup>b</sup>High resolution (< 1.9 Å) crystal structures of B-DNA where bimodal distributions were described for  $\epsilon$  and  $\zeta$  (the second maximum in the distribution of the histograms of these torsion angles is given in parentheses below).<sup>53</sup> <sup>c</sup>From molecular dynamics simulations and molecular modeling of NMR-derived data.<sup>63</sup>

**Table 2** Statistics for the structure determination of d(GCGAAAGC) from the **1DYANA(e)** ensemble.

---

<b>Experimental constraints</b>	
Assigned NOE peaks	326
Dihedral angles constrained by J-coupling	40
Hydrogen bonds	8
<b>Input for the DYANA structure calculation</b>	
NOE upper distance limits	206
Number of distance constraints per residue	
Intraresidue	18
Sequential	8
Total	26
Hydrogen bond distance constraints	16
Ring closure distance constraints	40
Dihedral angle constraints	73
<b>Stereospecific assignments</b>	11/14
<b>Residual DYANA target function (<math>\text{\AA}^2</math>)</b>	0.3
<b>Residual constraint violations</b>	
NOE upper distance: number > 0.10 $\text{\AA}$	0
Dihedral angle: number > 1°	0
<b>Average rmsd values (<math>\text{\AA}</math>) and their standard deviations calculated relative to the mean coordinates for backbone heavy atoms of different nucleotide selections</b>	
1-8	$0.3 \pm 0.1$ ( $0.4 \pm 0.1$ ) <sup>a</sup>
2-7	$0.2 \pm 0.1$ ( $0.2 \pm 0.1$ ) <sup>a</sup>

---

<sup>a</sup> All heavy atoms.

**Table 3** Comparison of the R-factors of the theoretical structures obtained during BCE construction and optimization (**BCEopt()**) of the d(GCGAAAGC) hairpin with those of the best models obtained experimentally (**1DYANA(e)**) and by systematic exploration of the ( $\epsilon, \zeta$ ) pairs of the G3pA4 and A5pA6 steps (**3MIN(o)**) structures. The definitions of the R-factors<sup>65,66</sup> ( $R_1$ - $R_3$ ) and corresponding data for all structures and are available as supplementary materials (S4 and S5, respectively).

Molecule	$R_1$ (upl)	$R_1$ (lol)	$R_2$ (torsion angle)	$\Delta\delta(\text{H2}'(\text{C2}))^a$	$\Delta\delta(\text{H4}'(\text{A4}))^a$	$R_3$ ( $\Delta\delta$ )
<b>1DYANA(e)</b> <sup>b</sup>	0.097	0.0005	0.14	0.72	0.53	0.77
<b>1DYANA(e)_min</b> <sup>b</sup>	<b>0.093</b>	<b>0.00001</b>	<b>1.25</b>	<b>0.42</b>	<b>0.37</b>	<b>0.49</b>
<b>Helix</b> <sup>c</sup>	0.180	0	3.14	0.83	0.92	1.11
<b>BCEori</b> <sup>c</sup>	0.125	0.0031	2.62	1.22	0.58	1.08
<b>BCEopt(<u>G</u>)</b> ( $\Omega, \chi$ ) <sup>c</sup>	0.119	0.0024	2.62	1.01	0.58	0.97
<b>BCEopt(<u>G</u>)</b> ( $\Omega, \chi$ ) <sup>c</sup>	0.114	0.0024	2.73	0.48	0.54	0.65
<b>BCEopt(<u>G</u><u>A</u>)</b> ( $\Omega, \chi$ ) <sup>c</sup>	0.114	0.0006	2.90	0.24	0.35	0.38
<b>BCEopt(<u>G</u><u>A</u><u>A</u>)</b> ( $\Omega, \chi$ ) <sup>c</sup>	0.112	0.0006	2.94	0.16	0.59	0.50
<b>BCEopt(<u>G</u><u>A</u><u>A</u><u>A</u>)</b> ( $\Omega, \chi$ ) <sup>c</sup> <b>or 2BCEopt_nmr</b>	0.114	0.0006	3.34	0.15	0.70	0.58
<b>3MIN(all except o)</b> <sup>d</sup>	0.102- 0.134	0.0003-0.002	2.4-4.3	0.21-0.57	0.13-1.66	0.21-1.45
<b>3MIN(o)</b> <sup>d</sup>	<b>0.102</b>	<b>0</b>	<b>0.81</b>	<b>0.59</b>	<b>0.01</b>	<b>0.33</b>

<sup>a</sup>Absolute value of the difference between the experimental and theoretical chemical shifts calculated with NUCHEMICS.<sup>64</sup>

<sup>b</sup>The best experimental structure **1DYANA(e)** was based on the following constraints - 206 upls and 246 lols and 73 dihedral constraints including <sup>31</sup>P-defined constraints for B<sub>I</sub>-like  $\alpha$  and  $\zeta$  ( $0 \pm 120^\circ$ ) and  $\zeta$  constraints for the G3pA4 step ( $180 \pm 40^\circ$ ).

<sup>c</sup>Starting from the B-DNA helix conformation of Figure 2a-1, the BCE optimization protocol involves rotation of the loop nucleotides with two angles ( $\Omega, \chi$ ) as depicted in Figure 2b. The R-factors given above were established for each step of this process to demonstrate the improvement in the model structure during this optimization. The resulting **BCEopt()** structures are characterized by: (i) the loop nucleotide that has been optimized is underlined, and (ii) the rotation angles that have been optimized for this additional loop nucleotide. <sup>d</sup>The minimized BCE models **3MIN(a-r)** were obtained by systematic screening of constraints for the ( $\epsilon, \zeta$ ) pairs of the G3pA4 and A5pA6 steps. In the case of **3MIN(o)** the constraints were as follows: C2'-endo,  $\beta(6)$  and  $\gamma(6)$  as well as B<sub>IIc+</sub> for the G3pA4 step and no constraints for the A5pA6 step.

## References

- (1) Klosterman, P. S.; Hendrix, D. K.; Tamura, M.; Holbrook, S. R.; Brenner, S. E. *Nucleic Acids Res.* **2004**, *32*, 2342.
- (2) Lee, C. L.; Gutell, R. R. *J. Mol. Biol.*, **2004**, *344*, 1225.
- (3) Jaeger, L.; Westhof, E.; Leontis, N. B. *Nucleic Acids Res.*, **2001**, *29*, 455.
- (4) Lescoute, A.; Leontis, N. B.; Massire, C.; Westhof, E. *Nucleic Acids Res.* **2005**, *33*, 2395.
- (5) Patel, S. D.; Rhodes, D. G.; Burgess, D. J. *The AAPS Journal* **2005**, *7*, e61.
- (6) Sinden, R. R.; Potaman, V. N.; Oussatcheva, E. A.; Pearson, C. E.; Lyubchenko, Y. L.; Shlyakhtenko, L. S. *J. Biosci.* **2002**, *27*, 53.
- (7) Hilbers, C. W.; Heus, H. A.; van Dongen, M. J. P.; Wijmenga, S. S. *The hairpin elements of nucleic acid structure: DNA and RNA folding. In Nucleic Acids and Molecular Biology* Eckstein, F. & Tilley, D. M. J., eds, Springer-Verlag, Berlin Heidelberg, 1994, *8*, p.56.
- (8) Varani, G. *Ann. Rev. Biophys. Biomol. Structure* **1995**, *24*, 379.
- (9) Nakano, M.; Moody, E. M.; Liang, J.; Bevilacqua, P. C. *Biochemistry* **2002** *41*, 14281.
- (10) van Dongen, M. J. P.; Wijmenga, S. S.; van der Marel, G. A.; van Boom, J. H.; Hilbers, C. W. *J. Mol. Biol.* **1996**, *263*, 715.
- (11) Tucker, B. J.; Breaker, R. R. *Curr. Opin. Struct. Biol.* **2005**, *15*, 342.
- (12) Chou, S.-H.; Chin, K.-H.; Wang, A. H.-J. *Nucleic Acids Res.* **2003**, *31*, 2461.
- (13) Hirao, I.; Nishimura, Y.; Tagawa, Y.; Watanabe, K.; Muira, K.-I.; *Nucleic Acids Res.* **1992**, *20*, 3891.
- (14) Grajcar, L.; El Amri, C.; Ghomi, M.; Femandjian, S.; Huteau, V.; Mandel, R.; Lecomte, S.;

Baron, M-H. *Biopolymers* **2006**, 82, 6.

- (15) Antao, V. P.; Tinoco Jr., I. *Nucleic Acids Res.* **1992**, 20, 819.
- (16) Correll, C. C.; Swinger, K. *RNA* **2003**, 9, 355.
- (17) Godson, G. N.; Barrell, B. G.; Staden, R.; Fiddes, G. C. *Nature* **1978**, 276, 236.
- (18) Hirao, I.; Ishida, M.; Watanabe, K.; Miura, K. *Biochim. Biophys. Acta* **1990**, 1087, 199.
- (19) van Dongen, M. J. P.; Mooren, M. M. W.; Willems, E. F. A.; van der Marel, G. A.; van Boom, J. H.; Wijmenga, S. S.; Hilbers, C. W. *Nucleic Acids Res.* **1997**, 25, 1537.
- (20) Pakleza, C.; Cognet, J. A. H. *Nucleic Acids Res.* **2003**, 321, 1075.
- (21) Santini, G. P. H.; Pakleza, C.; Cognet, J. A. H. *Nucleic Acids Res.* **2003**, 321, 1086.
- (22) Wüthrich, K. *in NMR of Proteins and Nucleic Acids*, John Wiley & Sons, New York, 1986, p 203.
- (23) Kim, S.-G.; Lin, L.-J.; Reid, B. R. *Biochemistry* **1992**, 31, 3564.
- (24) Varani, G.; Aboul-ela, F.; Allain, F. H.-T. *Progr. NMR Spectrosc.* **1996**, 29, 51.
- (25) Saenger, W. *in Principles of Nucleic Acid Structure*, Springer-Verlag, Ed CR Cantor, New York 1984.
- (26) Gorenstein, D. G.; Schroeder, S. A.; Fu, J. M.; Metz, J. T.; Roongta, V.; Jones, C. R. *Biochemistry* **1988**, 27, 7223.
- (27) Přecechtělová, J.; Munzarová, M. L.; Novák, P.; Sklenář, V. *J. Phys. Chem. B* **2007**, 111, 2658.
- (28) Přecechtělová, J.; Padrta, P.; Munzarová, M. L.; Sklenář, V. *J. Phys. Chem. B* **2008**, 112, 3470.
- (29) Heddi, B.; Foloppe, N.; Bouchemal, N.; Hantz, E.; Hartmann, B. *J. Am. Chem. Soc.* **2006**, 128,

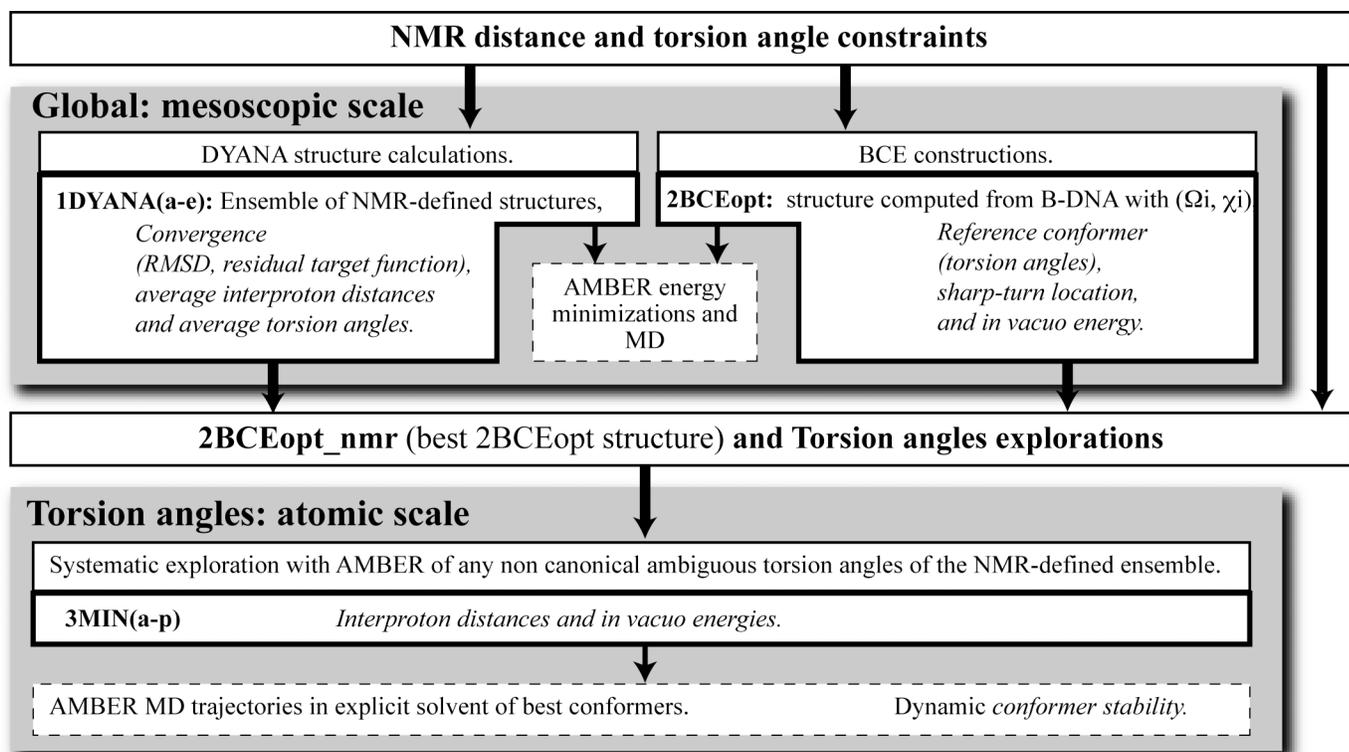
9170.

- (30) Keepers, J. W.; Kollman, P. A.; Weiner, P. K.; James, T. L. *PNAS* **1982**, *79*, 5537.
- (31) Güntert, P.; Mumenthaler, C.; Wüthrich, K. *J. Mol. Biol.* **1997**, *273*, 283.
- (32) Pearlman, D. A.; Case, D. A.; Caldwell, J. W.; Ross, W. R.; Cheatham, T. E.; DeBolt, S.; Ferguson, D.; Seibel, G.; Kollman, P. A. *Comp. Phys. Commun.* **1995**, *91*, 1.
- (33) Plateau, P.; Gueron, M. *J. Am. Chem. Soc.* **1982**, *104*, 7310.
- (34) Piantini, U.; Sorensen, O. W.; Ernst, R. R. *J. Am. Chem. Soc.* **1982**, *104*, 6800.
- (35) Griesinger, C.; Otting, G.; Wüthrich, K.; Ernst, R. R. *J. Am. Chem. Soc.* **1988**, *110*, 7870.
- (36) Kumar, A.; Ernst, R. R.; Wüthrich, K. *Biochem. Biophys. Res. Commun.* **1980**, *95*, 1.
- (37) Sklenář, V.; Bax, A. *J. Magn. Reson.* **1987**, *74*, 469.
- (38) Castagné, C.; Murphy, E. C.; Gronenborn, A. M.; Delepierre, M. *Eur. J. Biochem.* **2000**, *267*, 1223.
- (39) Bodenhausen, G.; Ruben, D. J. *Chem. Phys. Lett.* **1980**, *69*, 185.
- (40) Guntert, P.; Dotsch, V.; Wider, G.; Wüthrich, K. *J. Biomol. NMR* **1992**, *2*, 619.
- (41) Bartels, C.; Xia, T.; Billeter, M.; Güntert, P.; Wüthrich, K. *J. Biomol. NMR* **1995**, *6*, 1.
- (42) Padrta, P.; Stefl, R.; Kralik, L.; Zidek, L.; Sklenář, V. *J. Biomol. NMR* **2002**, *24*, 1.
- (43) Heus, H. A.; Pardi, A. *Science* **1991**, *253*, 191.
- (44) Jucker, F. M.; Heus, H. A.; Yip, P. F.; Moors, E. H. M.; Pardi, A. *J. Mol. Biol.* **1996**, *264*, 968.
- (45) Allain, F. H. T.; Varani, G. *J. Mol. Biol.* **1995**, *250*, 333.

- (46) Chou, S.-H.; Zhu, L.; Reid, B. R. *J. Mol. Biol.* **1997**, *267*, 1055.
- (47) Güntert, P.; Billeter, M.; Ohlenschläger, O.; Brown, L. R.; Wüthrich, K. *J. Biomol. NMR* **1998**, *12*, 543.
- (48) Güntert, P.; Braun, W.; Wüthrich, K. *J. Mol. Biol.* **1991**, *217*, 517.
- (49) Arnott, S.; Campbell-Smith, P.; Chandrasekaran, R. *Atomic coordinates and molecular conformations for DNA-DNA, RNA-RNA, and DNA-RNA helices. In Fasman, G.D. (ed.), CRC Handbook of Biochemistry and Molecular Biology*, CRC Press Inc., Cleveland, Ohio, USA, 1976, *2*, p. 411.
- (50) Case, D. A.; Cheatham, T. E.; Darden, T.; Gohlke, H.; Luo, R.; Merz, K. M.; Onufriev, A.; Simmerling, C.; Wang, B.; Woods, R. *J. Computat. Chem.* **2005**, *26*, 1668.
- (51) Cornell, W. D.; Cieplak, P.; Bayly, C. I.; Gould, I. R.; Merz, K. M.; Ferguson, D. M.; Spellmeyer, D. C.; Fox, T.; Caldwell, J. W.; Kollman, P. A. *J. Am. Chem. Soc.* **1995**, *117*, 5179.
- (52) Boulard, Y.; Cognet, J. A. H.; Gabarro-Arpa, J.; Le Bret, M.; Carbonnaux, C.; Fazakerley, G. V. *J. Mol. Biol.* **1995**, *246*, 194.
- (53) Schneider, B.; Neidle, S.; Berman, H. M. *Biopolymers* **1997**, *42*, 113.
- (54) Djuranovic, D.; Hartmann, B. *J. Biomol. Struct. Dyn.* **2003**, *20*, 771.
- (55) Flory, P. *in Statistical Mechanics of Chain Molecules*. Interscience. ISBN 0-470-26495-0. 1969  
Reissued 1989. ISBN 1-56990-019-1.
- (56) Auffinger, P.; Louise-May, S.; Westhof, E. *Biophys. J.* **1999**, *76*, 50.
- (57) Auffinger, P.; Vaiana, A. C. *in Handbook of RNA Biochemistry, Molecular dynamics of RNA systems*, Hartmann, R. K.; Bindereif, A.; Schön, A.; Westhof, E. Eds., Wiley WCH, Weinheim, 2005, p. 560.

- (58) Barfield, M.; Dingley, A. J.; Feigon, J.; Grzesiek, S. *J. Am. Chem. Soc.* **2001**, *123*, 4014.
- (59) Santos, R. A.; Tang, P.; Harbison, G. S. *Biochemistry* **1989**, *28*, 9372.
- (60) Ebrahimi, M.; Rossi, P.; Rogers, C.; Harbison, G. S. *J. Magn. Reson.* **2001**, *150*, 1.
- (61) Sarma, R. H.; Mynott, R. J.; Wood, D. J.; Hrusha, F. E. *J. Am. Chem. Soc.* **1973**, *95*, 6457.
- (62) Szyperski, T.; Götte, M.; Billeter, M.; Perola, E.; Cellai, L.; Heumann, H.; Wüthrich, K. *J. Biomol. NMR* **1999**, *13*, 343.
- (63) Cognet, J. A. H.; Boulard, Y.; Fazakerley, G. V. *J. Mol. Biol.* **1995**, *246*, 209.
- (64) Wijmenga, S. S.; Kruithof, M.; Hilbers, C. W. (1997) *J. Biomol. NMR* **10**, 337.
- (65) Gonzalez, C.; Rullmann, J. A. C.; Bonvin, A. M. J. J.; Boelens, R.; Kaptein, R. (1991) *J. Magn. Reson.* **91**, 659.
- (66) Koning, T. M. G., Boelens R., Van der Marel G. A., Van Boom J. H., Kaptein R. *Biochemistry* (1991) **30**, 3787.
- (67) Chou, S.-H.; Zhu, L.; Gao, Z.; Cheng, J.-W.; Reid, B. R. *J. Mol. Biol.* **1996**, *264*, 981.
- (68) Lamoureux, M.; Patard, L.; Hernandez, B.; Couesnon, T.; Santini, G. P. H.; Cognet, J. A. H.; Gouyette, C.; Cordier, C. *Spectrochimica Acta Part A* **2006**, *65*, 84.
- (69) Mauffret, O.; Amir-Aslani, A.; Maroun, R. G.; Monnot, M.; Lescot, E.; Femandjian, S. *J. Mol. Biol.* **1998**, *283*, 643.
- (70) Zhu, L.; Chou, S.-H.; Reid, B. R. *PNAS* **1996**, *93*, 12159.
- (71) Kuklenyik, Z.; Yao, S.; Marzilli, L.G. *Eur. J. Biochem.* **1996**, *236*, 960.
- (72) Chou, S.-H.; Tseng, Y.-Y.; Chu, B.-Y. *J. Biomol. NMR* **2000**, *17*, 1.

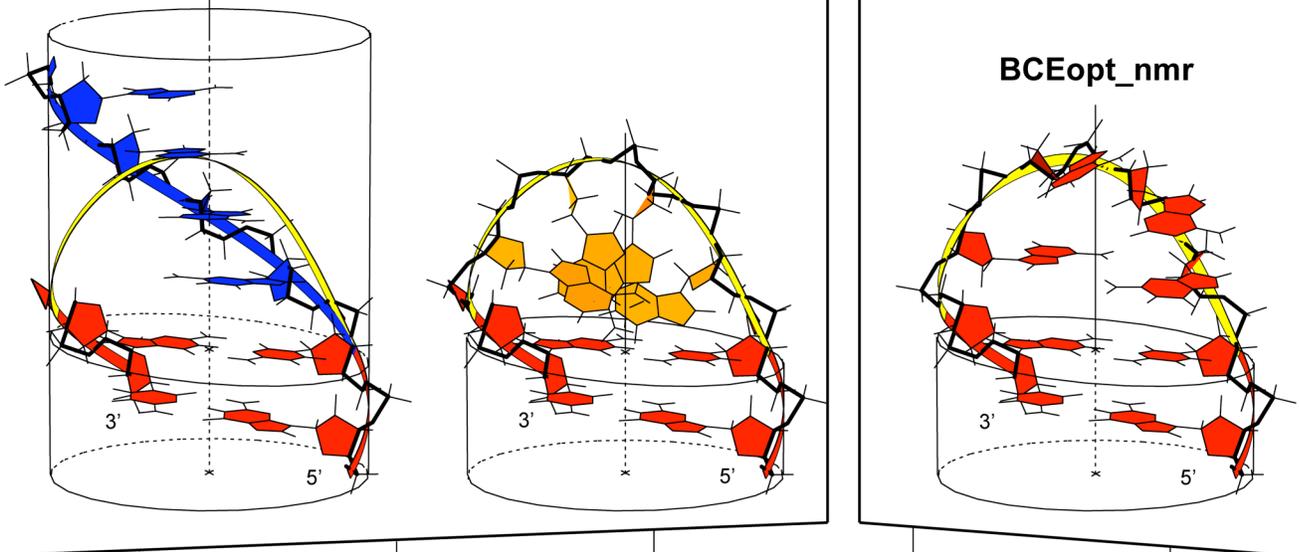
- (73) Baxter, S. M.; Greizerstein, M. B.; Kushlan, D. M.; Ashley, G. W. *Biochemistry* **1993**, *32*, 8702.
- (74) Chou, S.-H.; Chen, J.-W.; Fedoroff, O. Y.; Chuprina, V. P.; Reid, B. R. *J. Amer. Chem. Soc.* **1992**, *114*, 3114.
- (75) Hilbers, C. W.; Heus, H. A.; van Dongen, M. J. P.; Wijmenga, S. S. *Nucleic Acids and Molecular Biology* **1994**, *8*, 56.
- (76) Blommers, M. J. J.; van de Ven, F. J. M.; van der Marel, G. A.; Van Boom, J. H.; Hilbers, C. W. *Eur. J. Biochem.* **1991**, *201*, 33.
- (77) Chou, S.-H.; Chin, K.-H.; Chen, C. W. *J. Biomol. NMR* **2001**, *19*, 33.
- (78) Jucker, F. M.; Pardi, A. *Biochemistry* **1995**, *34*, 14416.
- (79) Legault, P.; Pardi, A. *J. Magn. Reson. B* **1994**, *103*, 82.
- (80) Furtig, B.; Richter, C.; Bermel, W.; Schwalbe, H. *J. Biomol. NMR* **2004**, *28*, 69.
- (81) Sunami, T.; Kondo, J.; Hirao, I.; Watanabe, K.; Miura, K.-I.; Takénaka, A. *Acta Cryst.* **2004**, *D60*, 90.
- (82) Perez, A.; Lankas, F.; Luque, J. F.; Orozco, M. *Nucleic Acids Res.* **2008**, *36*, 2379.
- (83) Werner, M. H.; Gupta, V.; Lambert, L. J.; Nagata, T. *Methods Enzymol.* **2001**, *338*, 283.



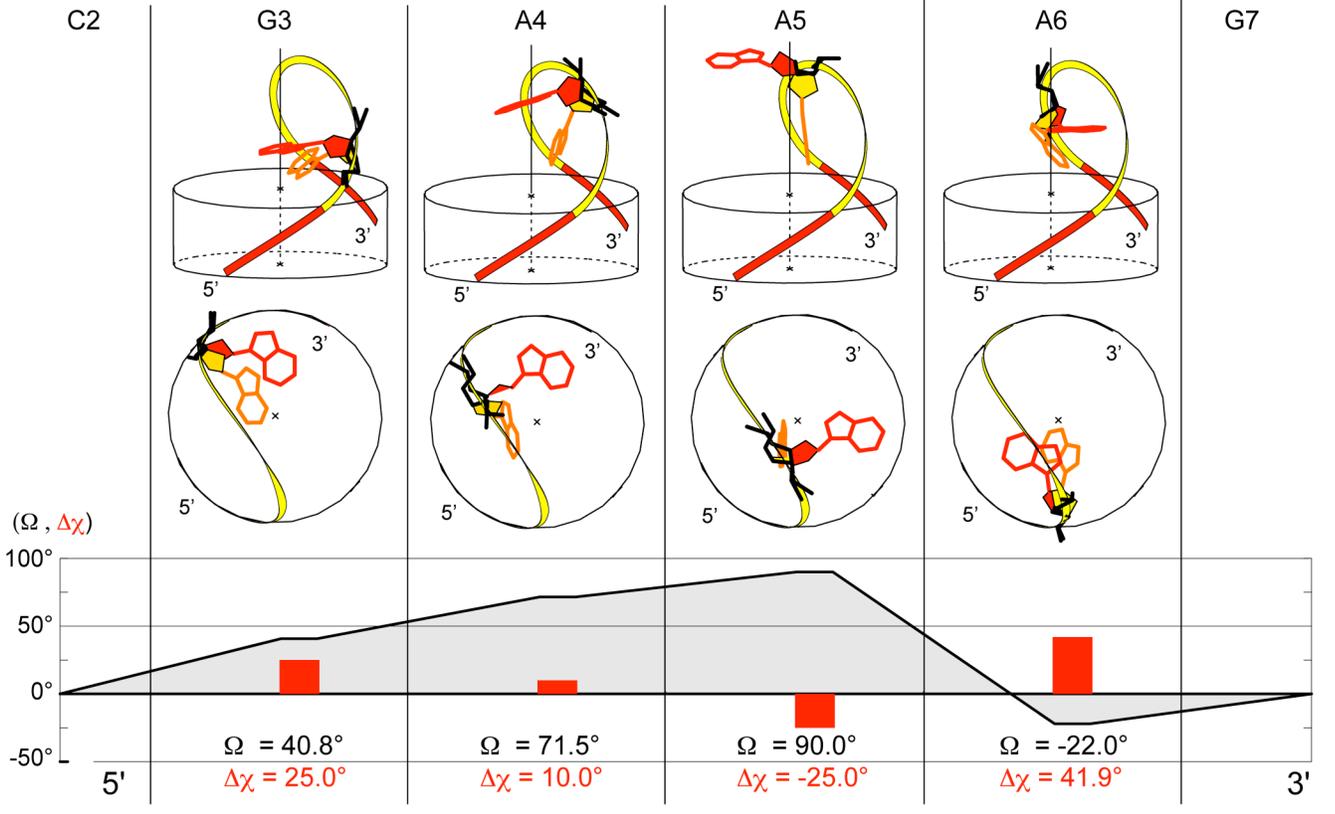
**Figure 1**

**(a) Construction of BCEopt**

(a-1) Helices  $\xrightarrow{\text{STEP 1}}$  (a-2) BCEori  $\xrightarrow{\text{STEP 2}}$  (a-3) BCEopt



**(b) BCEopt\_nmr nucleotide orientation parameters**



**Figure 2**

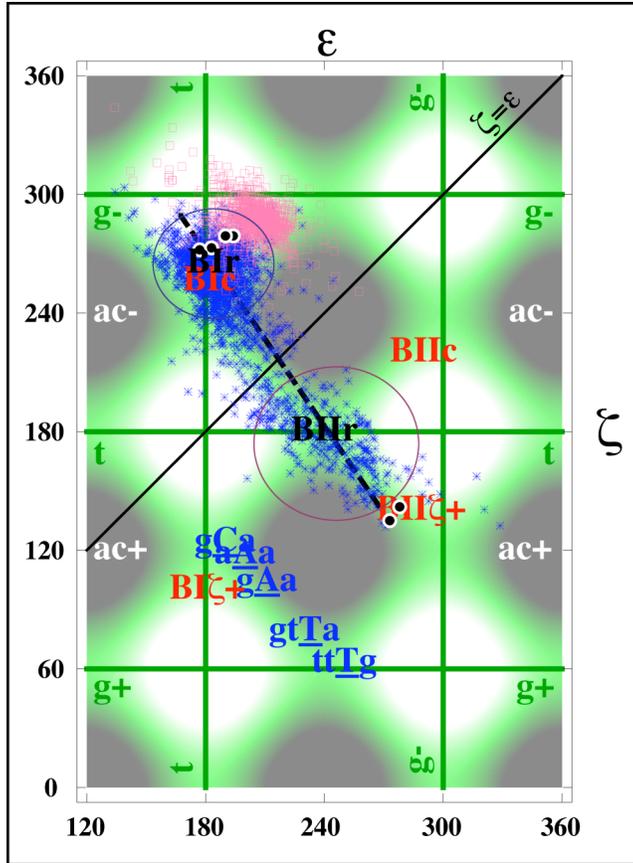
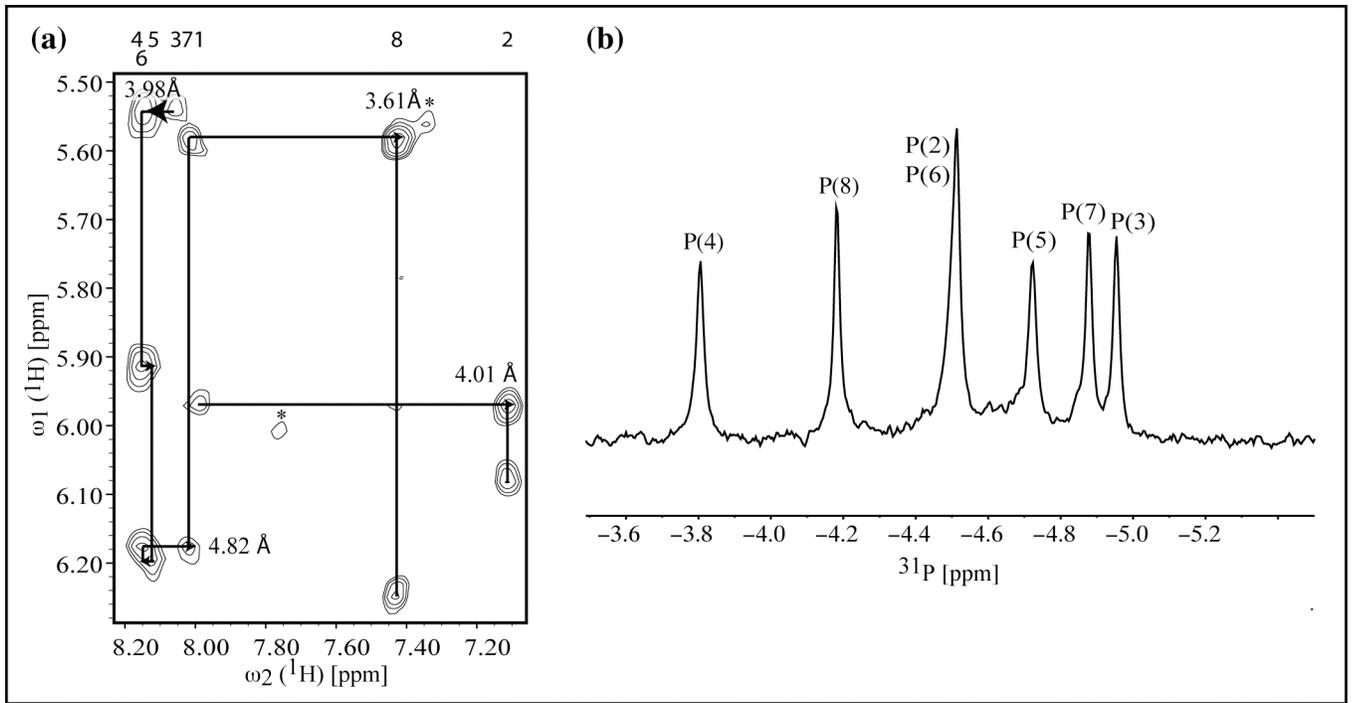


Figure 3



**Figure 4**

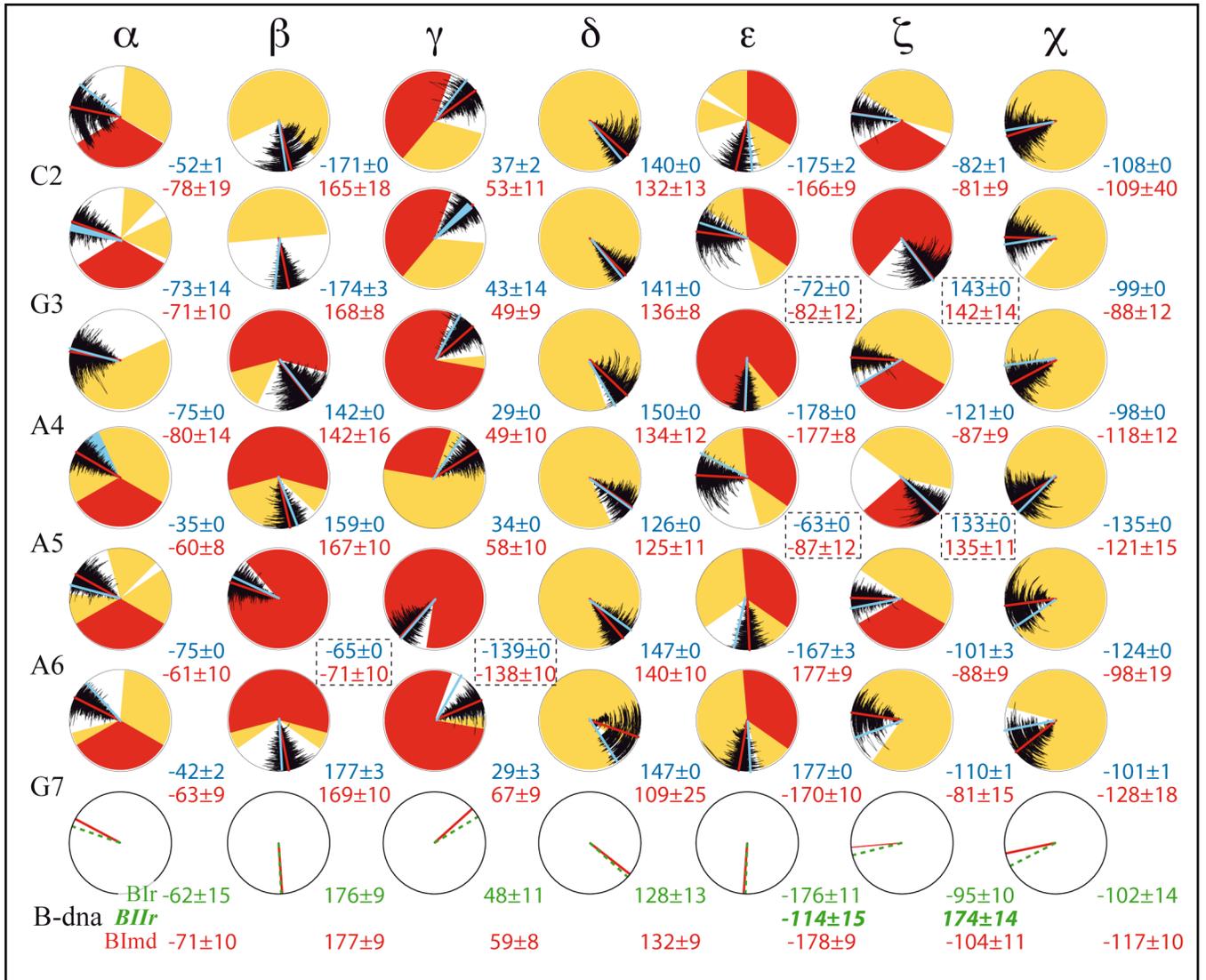
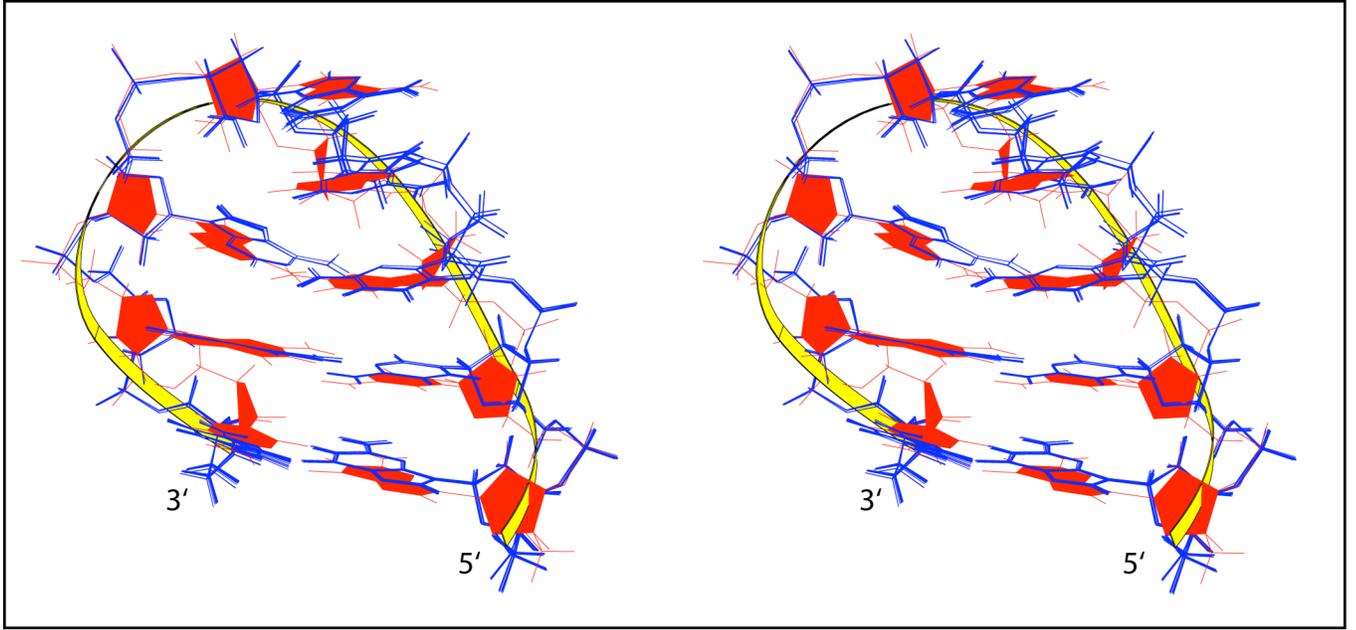


Figure 5



**Figure 6**