Article

Nucleic Acid Folding Determined by Mesoscale Modeling and NMR Spectroscopy: Solution Structure of d(GCGAAAGC)

Guillaume P. H. Santini, Jean A. H. Cognet, Duanxiang Xu, Kiran K. Singarapu, and Catherine Herve# du Penhoat


Downloaded from http://pubs.acs.org on May 7, 2009

More About This Article

Additional resources and features associated with this article are available within the HTML version:

• Supporting Information
• Access to high resolution figures
• Links to articles and content related to this article
• Copyright permission to reproduce figures and/or text from this article

View the Full Text HTML
Nucleic Acid Folding Determined by Mesoscale Modeling and NMR Spectroscopy: Solution Structure of d(GCGAAAGC)

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

Laboratoire de Biophysique Moléculaire, Cellulaire et Tissulaire, UMR 7033 CNRS, Université Pierre et Marie Curie Paris 6, Genopole Campus 1, 5 rue Henri Desbruières, Évry 91030, France; Department of Chemistry, The State University of New York at Buffalo, Buffalo, New York 14260; and Laboratoire de Biophysique Moléculaire, Cellulaire et Tissulaire, UMR 7033 CNRS, Université Paris 13, 74 rue Marcel Cachin, 93017 Bobigny, France

Received: November 16, 2008; Revised Manuscript Received: March 16, 2009

Introduction

To shed light on macromolecular recognition of RNA and DNA molecules, numerous studies of their structure, thermodynamic stability, and folding have been undertaken.1–3 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.4 For DNA, the recent evolution of gene medicine from an experimental technology into a viable strategy for developing therapeutics 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 tetraloops5 and the DNA triloops associated with genetic diseases6 have been studied comprehensively. The thermal stability of hairpins varies greatly7–9 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 interconversion between these forms may constitute a conformational switch for cellular regulation.10,11 Hairpin structures have been classified10,12 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-) (Tm = 76.5 °C in 0.1 M NaCl),13 r(-GAAA-) (Tm = 61 °C in 0.01 M NaCl),14 d(-TTCG-) (60.4 °C in 1 M NaCl), and r(-UUCCG-) (76.5 °C in 1 M NaCl). While the X-ray structure of an RNA GAAA tetraloop16 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 DNA17,18 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 determined19 is the GTTA tetraloop (Tm = 49.5 °C in 0.1 M NaCl).13 Solution structures of nucleic acids are currently determined with standard NMR methods.22–24 The local conformation of the sugar-phosphate backbone is described by six torsion angles (α, β, γ, δ, ε, ζ) while the orientation about the bond between the sugar and the base is defined by only one (χ).25 Simple relationships between the experimental data, such as homo-
heteronuclear scalar coupling constants, line widths, and nuclear Overhauser effects (NOEs), define the conformation of the β, χ, η, ε, and γ torsions while 31P chemical shifts are related to the conformation about the α and ζ dihedral angles. Recent theoretical studies\textsuperscript{26,28} have confirmed that the phosphorus in the g°/t (α = −60°, ζ = 180°) conformation is more deshielded than the one in the g°/g (α, ζ = −60°) conformation. 31P chemical shifts have been used to determine B1 (ε = t, ζ = g°) and B11 (ε = g°, ζ = t) populations in DNA duplexes.\textsuperscript{29} However, as other factors influence 31P chemical shifts,\textsuperscript{24} only very loose constraints (±120°) have been applied to α and ζ and only in the case of the B1 conformation. As a result, these latter torsions are often poorly defined in NMR studies.

Molecular mechanics has not provided torsion angle restraints for either α or ζ as in vacuo analysis of staggered nucleic acid conformations has not indicated large variations in their energy.\textsuperscript{30} 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.\textsuperscript{29,31} This line, also called the elastic line, is the trajectory with the least deformation energy of a continuous flexible thin rod as in vacuo analysis of staggered nucleic acid folding observations. To resolve the ambiguity regarding the B-DNA conformer that took into account these NMR-derived restraints (e.g., stacking of A5 on A4 or not) suggested by NMR-derived data. A reference structure, 2BCEopt(), 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 (ε,ζ) torsion angle pairs, the 2BCEopt() was systematically explored using the DYANA simulated annealing protocol.\textsuperscript{31} These calculations yielded five NMR-derived ensembles, 1DYANA(a-e) that differed only in the conformation of two (ε,ζ) 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(), 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 (ε,ζ) torsion angle pairs, the 2BCEopt() was systematically modified by energy minimization with AMBER to give 3MIN(a-r). This provided an energy-based exploration of the (ε,ζ) pairs while maintaining the global fold of 2BCEopt(). 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(GCGAAGC)-3’ DNA hairpin.

**Materials and Methods**

**Sample Preparation.** 5’-d(GCGAAGC)-3’ was purchased from Trilink Inc. and purified on a Sephadex G10 gel-filtration column. Half of the purified DNA was lyophilized twice from D2O buffer (pH 6.8, 10 mM Na phosphate, 50 mM NaCl, 5 μM EDTA, 2.5 μM NaN3), dissolved in 500 μL of D2O and placed in a 5 mm tube yielding a 4 mM hairpin solution. The
Remainder was dissolved in 400 μL of the same buffer in 90% H$_2$O/10% D$_2$O. 250 μL was placed in a 5 mm Shigemi tube to yield a 4.5 mM sample.

**NMR Spectroscopy.** NMR measurements (13C and 31P, 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 were acquired between 1 and 25 °C (τ = 46 and 105 μs giving the maximum signal intensity in the imino and amino regions, respectively).

The following 2D 1H spectra were recorded in D$_2$O: 2QF-COSY34 (τ$_{max}$ = 155 ms, τ$_{min}$ = 155 ms, total measurement time 14 h); clean [1H,1H]-TOCSY35 (64 ms, 121 ms, 1.4 h; τ$_{min}$ = 38 ms, 5 °C); [1H, 1H]-NOE36 (73 ms, 147 ms, 3 h; five experiments with τ$_{min}$ = 50, 100, 150, 200, and 400 ms, respectively, recycle time 5 s; cryoprobe, 128 ms, 128 ms, 21 h; three experiments with τ$_{min}$ = 30, 40, and 60 ms, respectively, recycle time 6 s). The parameters for the 2D 1H spectrum measured in H$_2$O were as follows: jump-return [1H,1H]-NOESY (68 ms, 315 ms, 6 h; τ$_{min}$ = 50, 100, 150, and 200 ms, respectively, recycle time 5 s, τ = 105 μs, 1 °C; 68 ms, 68 ms, 6 h; τ$_{min}$ = 50, 100, 150 ms, recycle time 5 s, τ = 46 μs, 1 °C; Cryoprobe, 68 ms, 68 ms, 21 h; τ$_{min}$ = 60 ms, recycle time 5 s, 5 °C, τ = 40 μs).

Finally, the following heteronuclear 2D experiments recorded in D$_2$O: [31P,1H]-COSY37,38 (400 MHz, 171 ms, 244 ms, 11 h, 0.5 Å; where only lower limits shorter than 3.5 Å were retained.42 All aromatic and sugar protons exhibited some clear 0.5 Å relaxation rates, a shorter 50 ms mixing time was used at 1 °C. Accordingly, the cross-peak volumes in (data corresponding to the shortest interproton distance expected, τ$_{min}$) for the B$_r$ range, DYANA(c) without lols;24,45,46 calculations with loose α and ζ constraints for δ(31P) in the both B$_{r}$ range (0° ± 120°) and in the downfield-shifted range (180° ± 40°), DYANA(d) without lols, or DYANA(e) with lols. The first and second simulations were based on minimal and maximal NOE data only whereas 31P restraints were progressively introduced in the remaining simulations. The last simulation relied on all the experimentally observed NOE and 31P data.

**DYANA Structure Calculations.** All upls and the preceding dihedral angle constraints were translated into total dihedral angle constraints using the FOUND module47 of DYANA, which performs a grid search for allowed conformations in the space spanned by the nine torsion angles describing a dinucleotide segment. Stereoelectronic assignments for all of the H2' and H2'' protons with nondegenerate 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 module48 of the DYANA program corroborated these H2'/H2'' assignments and provided the stereoelectronic 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.** For DNA or RNA hairpin loops of 3–4 nucleotides with the BCE approach can be summarized as a three-step procedure.20,21 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 helices49 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 tetraloop in tridimensional 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).20 It is the trajectory of least energy of deformation. Transportation of the whole loop chain onto the elastic line is described elsewhere.20 The final conformation obtained after these transformations is the molecular model, “BCE ori”, shown in Figure 2a-2.

**Step 2.** This deformation step takes place on the scale of the nucleotide. Two rotation angles, Ω, and glycosidic torsion angle ξ, were sufficient to orient each nucleotide i in the loop23 with respect to helical B-DNA while the stem nucleotides are unchanged (Ω = 0, ξ = 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, Ω. These simple rotations (Ω,ξ) had been searched to match the Cartesian coordinates of the GAAA hairpin given in PDB file 1ac7.19,21 Both GAAA and GATA possess a G·A base pair...
with the second loop nucleotide stacked onto it. This is why only the third loop nucleotide (Ω₅, /khi₅) values had to be searched to take into account the NMR data for the GAAA hairpin. Three positions of A₅ (Ω₅) were explored: in the major groove (∼70°), stacked (90°), and unstacked conformers (>90°). 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 and 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. 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.

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 β(6) and γ(6) torsions values were introduced at this stage. Additional torsion angle constraints were included to explore the conformations of the (ε,ζ) 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 AMBER20,51 (1DYANA(a-e)_min and 2BCEopt_nmr_min) as explained in previous work under torsion angle restraints until the rms energy gradient was less than 0.05 kcal/(mol·Å). The force constant was set equal to 900 kcal/(mol·rad²). In order to maintain the C₂′-endo conformation of all sugar puckers, torsion angles, δ₁₀, (C₅′-C₄′-C₃′-O₃′) were forced to nominal value, δ₁₀ = 144°. All restrained structures were subsequently relaxed without restraints.

As will be discussed below, structural studies of hairpins have almost invariably pointed to noncanonical values of ε and ζ at the sharp turn that do not correspond to either of the B_Ir or B_{IIr} local conformations found in helical crystal structures (cf. Figure 3). The ζ value proposed for hairpins is almost invariably in the g+ to g+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 (ε,ζ) pair: B_{Ik} (−178°, −104°),...
Mesoscale Modeling and NMR Spectra of d(GCGAAAGC)

Figure 3. Plot of the $\zeta$ versus $\epsilon$ torsion angles overlaying the eclipsed (gray) and staggered (white) conformations of a dihedral fragment that represent the favored $g^\ast$, $t$, and $g^\ast$ regions. The $B_{ik}$ (blue) and $B_{ik}$ (red) ellipses encompass 95% of crystal structures. The $(\epsilon, \zeta)$ pairs of the $B_{ik}$ (c, constraint), $B_{ik}$, and $B_{ik}$ (positive $\zeta$ values, symmetrical about the $\zeta_{180}$ axis with respect to $B_{ik}$) and $B_{ik}$ (positive $\zeta$ values, symmetrical about the $\zeta_{180}$ axis with respect to $B_{ik}$) 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 are represented with black dots highlighted in white, and the positions of the $(\epsilon, \zeta)$ pairs in the BI $\zeta_{180}$ conformation of the AAA, AAG, GAA, GCA, GGT, GGA, GTT, and TTQ hairpins are indicated in blue letters.

$B_{ik}$ ($-70^\circ$, $-140^\circ$), $B_{ik}$ as ($-180^\circ$, $100^\circ$), and $B_{ik}$ ($-70^\circ$, $140^\circ$). Some combinations of adjacent torsion angles such as $g^\ast/g^\ast$ or a $g^\ast/g^\ast$ did not have to be considered because they generate well-known steric hindrance. The $(\epsilon, \zeta)$ torsions of the 2BCeopt_nmr structure were systematically modified with AMBER to determine the favorable conformations of the G3pA4 (no restraints, BIIr, BIIc, and BIIc) and the A5pA6 (no restraints, BIIr, BIIc, and BIIc) steps. This procedure afforded the energy-minimized structures 3MIN(o-r).

AMBER Molecular Dynamics Trajectories. Molecular dynamics simulations were performed using the AMBER 8 package and the PARM94 force field. 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+ and 4 Cl− 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. 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 cutoff 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 $\Omega$, and $\Delta\chi_3$ 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−1 rad−2) 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, $[^{13}C, ^{1}H]$HSQC, and $[^{31}P, ^{1}H]$-COSY experiments. At 25 °C, only the signals of the 5′-d(GCGAAAGC)-3′ hairpin are detected in the 1D and 2D spectra in D2O. The 1H chemical shifts (69) of all nonexchangeable protons, the $^{13}C$ (42), and the $^{31}P$ (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 $^1H$ assignments were obtained using standard protocols based on the observation of $^1H$−$^1H$ 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 (t $\mu$s 100 ms) containing the H6,8(ω)/H1′(ω) intraresidue 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 H2O/D2O. 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 no. 15898) whereas the partial chemical shift data of the minor species are collected in Table S1.

Structures Features Derived from $^1H$, $^{13}C$, and $^{31}P$ Chemical Shifts. The NMR data of the 5′-d(GCGAAAGC)-3′ sequence at 25 °C reproduce the very characteristic signature of 5′-d(PyGNAAPu)-3′ hairpins (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. 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 G·A pairs.

All the nonterminal C3′ (78.35–79.70 ppm) and C5′ (67.28–68.73 ppm) sugar carbon signals of the 5′-d(GCGAAAGC)-3′ tetraloop hairpin resonated in the low-field regions typical of sugars with $\Delta\delta$ 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.

The $^{31}P$ 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 (Δδ = -3.81 ppm), Figure 4b. 31P 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 ζ and α are g-24,26,29. Such data have frequently been used to exclude the t domain of both ζ and α torsion angles as (ζ(2′),α(t)) and (ζ(t),α(g-2′)) conformations are associated with downfield phosphorus chemical shifts.24 To evaluate the possible interpretations of downfield 31P 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 3J1′,2′ was smaller than 3J1″,2″. 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 intrasugar interproton distances derived from the NOESY cross-peak volumes revealed the order d(1′2″) < d(1′2′) ~ d(1′4″) ~ d(2′2″) and is consistent with the sugar puckering inferred from the DQ-COSY spectrum.22 The anti orientation of the χ torsion angle of all the nucleotides could be determined from the intranucleotide base proton nOes with the sugar protons (d(H6/ 8,1′) < d(H6/8,H2″ < d(H6/8,2″)).

Initially, the constraints for the β, γ and ε torsion angles were obtained from the sum of the H5′/H5″, H4″, and H3′ coupling constants in the 31P,1H-COSY spectrum, (ΣJH5, ΣJH5, ΣJH4, and ΣJH3, respectively, Table S3). In particular, it could be deduced that the 3JH,H coupling constants of the 5′/5″ methylene protons of A4, A5, and G7 were less than 10 Hz (i.e., ΣJH5 and ΣJH5 both less than 22 Hz), indicating that the β torsion fell within the 105°–255° range. H4″(π)/P(π) 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.30 This is the case in B-DNA where β and γ adopt t and g+ orientations, respectively. Finally, ΣJH5 < 13 Hz for the H3, C2, and G4 and G7 H3′(i-1)/P(i) cross peaks in the [31P,1H]-COSY spectrum indicating ϵ values between 125° and 355°. The H3′(A4)/P(A5) correlations were missing (3JH,L < 1 Hz) restricting ϵ(4) to the 140°–160° range (the 310°–330° range is excluded for steric reasons).30

All these excluded torsion angle ranges for α, β, γ, δ, ε, ζ, and γ 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 earlier62 for a chimeric hybrid duplex, this protocol leads to additional constraints 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 3P-derived α or ζ restraints. Due to spectral overlap of P(2) and P(6) (Figure 4b), the cross peaks that characterize the β(2) and β(6) torsion angles in the [31P,1H]-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 β(6) and γ(6)) made it possible to simulate the 1D spectrum of the –P(6)–H5″–H5″–H4″–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 [31P,1H]-COSY corroborated the unusual values of β(6) and γ(6) calculated in 1DYANA(a,b).

In the last three simulations, α and/or ζ restraints were
obtained by NMR were sufficient to define 47 out of the 51 (noncanonical values in bold).

The best agreement between the average torsion angles of the G3pA4 and A5pA6 steps, respectively, are not compatible with the 31P chemical shifts64 as described in the following paragraph) and the statistics for all structure calculations, IDYANA(a-e), were very similar (target functions all <0.35 Å² and rmsd for the backbone heavy atoms <0.6 ± 0.2 Å) and those of the IDYANA(e) ensemble have been collected in Table 2. The pairwise rmsd of all heavy atoms of all nucleotides was 0.3 ± 0.1 Å and the residual target function was 0.3 Å². All of the simulations converged reasonably well (>20 structures) except IDYANA(d) that only contained 18 converged structures. The α and γ dihedral angles, that are g°- (-62 ± 15°) and g°+ (48 ± 11°) in B-DNA, tend toward the cis range (-30° to +30°) in some of the DYANA ensembles, particularly in the simulations without lols (five in IDYANA(a), four in IDYANA(d), and the most deformed, α(6) -14°, in IDYANA(c)). These anomalies disappear upon minimization with AMBER. In the case of the IDYANA(e) ensemble, minimization also leads to 149° and 129° values of the average γ torsions of the G3pA4 and A5pA6 steps, respectively, in excellent agreement with the 31P chemical shifts. A systematic comparison between the NMR data calculated for the IDYANA(a-e) structures (upls, lols, torsion angles, and high-field-shifted proton chemical shifts66 as described in the following paragraph) and the experimental data has been established with R factors65,66.

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 nonterminal 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 line widths associated with characteristic NOEs22 or the 31P chemical shifts, and 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 δ on A4 (144°, 2.5 kcal/mol·rad²). 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 B3 conformer 3 MIN(o) obtained as the resulting conformation of the combined approach DYANA and BCE/AMBER described in Figure 1. Last row: in green, mean values and stdev of the backbone heavy atoms of all nucleotides was 0.6 ± 0.2 Å and those of the IDYANA(e) ensemble have been collected in Table 2. The pairwise rmsd of all heavy atoms of all nucleotides was 0.3 ± 0.1 Å and the residual target function was 0.3 Å². All of the simulations converged reasonably well (>20 structures) except IDYANA(d) that only contained 18 converged structures. The α and γ dihedral angles, that are g°- (-62 ± 15°) and g°+ (48 ± 11°) in B-DNA, tend toward the cis range (-30° to +30°) in some of the DYANA ensembles, particularly in the simulations without lols (five in IDYANA(a), four in IDYANA(d), and the most deformed, α(6) -14°, in IDYANA(c)). These anomalies disappear upon minimization with AMBER. In the case of the IDYANA(e) ensemble, minimization also leads to 149° and 129° values of the average γ torsions of the G3pA4 and A5pA6 steps, respectively, in excellent agreement with the 31P chemical shifts. A systematic comparison between the NMR data calculated for the IDYANA(a-e) structures (upls, lols, torsion angles, and high-field-shifted proton chemical shifts66 as described in the following paragraph) and the experimental data has been established with R factors65,66.

The statistics for all structure calculations, IDYANA(a-e), were very similar (target functions all <0.35 Å² and rmsd for the backbone heavy atoms <0.6 ± 0.2 Å) and those of the IDYANA(e) ensemble have been collected in Table 2. The pairwise rmsd of all heavy atoms of all nucleotides was 0.3 ± 0.1 Å and the residual target function was 0.3 Å². All of the simulations converged reasonably well (>20 structures) except IDYANA(d) that only contained 18 converged structures. The α and γ dihedral angles, that are g°- (-62 ± 15°) and g°+ (48 ± 11°) in B-DNA, tend toward the cis range (-30° to +30°) in some of the DYANA ensembles, particularly in the simulations without lols (five in IDYANA(a), four in IDYANA(d), and the most deformed, α(6) -14°, in IDYANA(c)). These anomalies disappear upon minimization with AMBER. In the case of the IDYANA(e) ensemble, minimization also leads to 149° and 129° values of the average γ torsions of the G3pA4 and A5pA6 steps, respectively, in excellent agreement with the 31P chemical shifts. A systematic comparison between the NMR data calculated for the IDYANA(a-e) structures (upls, lols, torsion angles, and high-field-shifted proton chemical shifts66 as described in the following paragraph) and the experimental data has been established with R factors65,66.

Separated simulations showed that restraining α(4) to the t conformation (180 ± 40°) 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°) of α(4).
TABLE 1: Torsion Angles Defining the 5'-d(GCGAAAGC)-3' Structures Obtained in the 1DYANA(a-e) Simulations

<table>
<thead>
<tr>
<th>residue</th>
<th>α</th>
<th>β</th>
<th>γ</th>
<th>δ</th>
<th>ε</th>
<th>ζ</th>
<th>χ</th>
</tr>
</thead>
<tbody>
<tr>
<td>C2</td>
<td>-67 ± 6</td>
<td>-173 ± 5</td>
<td>34 ± 1</td>
<td>135 ± 1</td>
<td>-156 ± 2</td>
<td>-136 ± 1</td>
<td>-108 ± 0</td>
</tr>
<tr>
<td>G3</td>
<td>-27 ± 0</td>
<td>126 ± 1</td>
<td>60 ± 1</td>
<td>127 ± 0</td>
<td>180 ± 0</td>
<td>-120 ± 0</td>
<td>-110 ± 0</td>
</tr>
<tr>
<td>A4</td>
<td>-25 ± 0</td>
<td>-176 ± 1</td>
<td>21 ± 0</td>
<td>150 ± 0</td>
<td>173 ± 0</td>
<td>-111 ± 0</td>
<td>-99 ± 0</td>
</tr>
<tr>
<td>A5</td>
<td>-51 ± 0</td>
<td>160 ± 0</td>
<td>56 ± 0</td>
<td>129 ± 0</td>
<td>-108 ± 0</td>
<td>169 ± 1</td>
<td>-124 ± 0</td>
</tr>
<tr>
<td>A6</td>
<td>-46 ± 2</td>
<td>-63 ± 0</td>
<td>-153 ± 0</td>
<td>119 ± 1</td>
<td>177 ± 0</td>
<td>-100 ± 0</td>
<td>-115 ± 0</td>
</tr>
<tr>
<td>G7</td>
<td>-33 ± 7</td>
<td>159 ± 2</td>
<td>54 ± 1</td>
<td>110 ± 1</td>
<td>-179 ± 1</td>
<td>-98 ± 1</td>
<td>-134 ± 0</td>
</tr>
<tr>
<td>B-DNAa</td>
<td>-62 ± 15</td>
<td>176 ± 9</td>
<td>48 ± 11</td>
<td>128 ± 13</td>
<td>-176 ± 11</td>
<td>-95 ± 10</td>
<td>-102 ± 14</td>
</tr>
<tr>
<td>B-DNAb</td>
<td>-71 ± 2</td>
<td>177 ± 2</td>
<td>59 ± 1</td>
<td>132 ± 4</td>
<td>-178 ± 2</td>
<td>-104 ± 4</td>
<td>-117 ± 5</td>
</tr>
</tbody>
</table>

a Values in bold display significant deviations when compared to those of B-DNA. 1DYANA(a): 208 ups and 69 dihedral constraints without $^{31}$P-defined constraints on α or ζ. 1DYANA(b): 202 ups, 246 lols, and 78 dihedral constraints without $^{31}$P-defined constraints on α or ζ. 1DYANA(c): 208 ups and 67 dihedral constraints including $^{31}$P-defined constraints for B-like α and ζ (0 ± 120°). 1DYANA(d): 208 ups and 66 dihedral constraints including $^{31}$P-defined constraints for B-like α and ζ (0 ± 120°) and ζ constraints for the G3pA4 step (180 ± 40°). 1DYANA(e): 206 ups and 246 lols and 73 dihedral constraints including $^{31}$P-defined constraints for B-like α and ζ (0 ± 120°) and ζ constraints for the G3pA4 step (180 ± 40°). b High-resolution (~1.9 Å) crystal structures of B-DNA where bimodal distributions were described for ε and ζ (the second maximum in the distribution of the histograms of these torsion angles is given in parentheses below).35

that have been defined in the Supporting Information (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) (δ 1.58 ppm compared to the average H2' value of 2.48 ppm) and H4'(A4) (δ 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.66 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 $^{31}$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 $(\Omega_5, \Omega_3)$ pairs that define the orientation of each nucleotide of the GTTA hairpin (1ac7)35 with respect to the elastic line in the BCE approach had been obtained previously21 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 $(\Omega_5, \Omega_3)$ were 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 $(\Omega_5, \Omega_3)$ 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$^{-1}$) 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(t)) 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 constants, since they remain as close as possible to B-DNA.
TABLE 2: Statistics for the Structure Determination of d(GCGAAAGC) from the IDYANA(e) Ensemble

<table>
<thead>
<tr>
<th>Experimental constraints</th>
<th>326</th>
</tr>
</thead>
<tbody>
<tr>
<td>dihedral angles constrained by J-coupling</td>
<td>40</td>
</tr>
<tr>
<td>hydrogen bonds</td>
<td>8</td>
</tr>
<tr>
<td>Input for the DYANA structure calculation</td>
<td>206</td>
</tr>
<tr>
<td>NOE upper distance limits</td>
<td>206</td>
</tr>
<tr>
<td>no. of distance constraints per residue</td>
<td>18</td>
</tr>
<tr>
<td>intraresidue</td>
<td>8</td>
</tr>
<tr>
<td>sequential</td>
<td>26</td>
</tr>
<tr>
<td>total</td>
<td>16</td>
</tr>
<tr>
<td>hydrogen bond distance constraints</td>
<td>40</td>
</tr>
<tr>
<td>ring closure distance constraints</td>
<td>73</td>
</tr>
<tr>
<td>dihedral angle constraints</td>
<td>11/14</td>
</tr>
<tr>
<td>Stereospecific assignments</td>
<td>0.3</td>
</tr>
<tr>
<td>Residual DYANA target function (Å²)</td>
<td>0</td>
</tr>
<tr>
<td>Residual constraint violations</td>
<td>0</td>
</tr>
<tr>
<td>NOE upper distance: number &gt; 0.10 Å</td>
<td>0</td>
</tr>
<tr>
<td>dihedral angle: number &gt; 1°</td>
<td>0</td>
</tr>
<tr>
<td>average rmsd values (Å) and their standard deviations calculated relative to the mean coordinates for backbone heavy atoms of different nucleotide selections</td>
<td>0.3 ± 0.1 (0.4 ± 0.1)°</td>
</tr>
<tr>
<td>1–8</td>
<td>0.2 ± 0.1 (0.2 ± 0.1)°</td>
</tr>
</tbody>
</table>

All heavy atoms.

2BCE_opt_nmr (that were inconsistent with the NMR data), were observed in all of the minimized structures (excluding the relatively large R² values) with the exception of the 3MIN(o) conformer, which remained stable. This latter structure (B_{Hi} 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⁻¹. The agreement between the experimental data and those calculated for the other 3MIN(a-r) structures was poorer and the corresponding R factors were also given in the Supporting Information.

The stereoscopic view of the 3MIN(o) BCE structure (21st model, PDB 2k71) after minimization with AMBER (red) and the superposition of the 20 IDYANA(e) (first 20 models, PDB 2k71) structures that best reproduce the NMR constraints after minimization with AMBER (blue) 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 destacking, 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 B1 to B4 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 (~1000 ps or 500 000 integration steps), e.g., with the rotation, or not, of the A5 base from anti to syn, upon introducing a small constraint on δ (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·rad²) 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 3P Chemical Shifts with DNA Hairpin Folding. The 3P chemical shifts of published DNA triloops17,28,73 and tetraloops19,75–77 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_{3''}–L_{3}L_{3},L_{3}L_{3}N_{4'}) for tetraloops and N_{5'}–L_{5}L_{5}L_{5}L_{5}N_{5'} for triloops). The average value of the 3P chemical shifts (δP) and the half-width of the interval (1/2ΔδP) excluding an eventual outlier (ΔδP > 0.7 ppm) were established. The chemical shift and position in the sequence of any outlier (δ_{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.
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

<table>
<thead>
<tr>
<th>molecule</th>
<th>$R_1$ (upl)</th>
<th>$R_1$ (lol)</th>
<th>$R_2$ (torsion angle)</th>
<th>$\Delta(\delta(H^2(C2))$</th>
<th>$\Delta(\delta(H^4(A4))$</th>
<th>$R_3$ ($\Delta$)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1DYANA(e)$^\epsilon$</td>
<td>0.097</td>
<td>0.0005</td>
<td>0.14</td>
<td>0.72</td>
<td>0.53</td>
<td>0.77</td>
</tr>
<tr>
<td>1DYANA(e)_min$^\epsilon$</td>
<td>0.093</td>
<td>0.00001</td>
<td>1.25</td>
<td>0.42</td>
<td>0.37</td>
<td>0.49</td>
</tr>
<tr>
<td>helix$^\epsilon$</td>
<td>0.180</td>
<td>0</td>
<td>3.14</td>
<td>0.83</td>
<td>0.92</td>
<td>1.11</td>
</tr>
<tr>
<td>BCEopt$^d$</td>
<td>0.125</td>
<td>0.0031</td>
<td>2.62</td>
<td>1.22</td>
<td>0.58</td>
<td>1.08</td>
</tr>
<tr>
<td>BCEopt(G) ($\Omega$, $\chi$)$^d$</td>
<td>0.119</td>
<td>0.0024</td>
<td>2.62</td>
<td>1.01</td>
<td>0.58</td>
<td>0.97</td>
</tr>
<tr>
<td>BCEopt(GA) ($\Omega$, $\chi$)$^d$</td>
<td>0.114</td>
<td>0.0024</td>
<td>2.73</td>
<td>0.48</td>
<td>0.54</td>
<td>0.65</td>
</tr>
<tr>
<td>BCEopt(GAA) ($\Omega$, $\chi$)$^d$</td>
<td>0.114</td>
<td>0.0006</td>
<td>2.90</td>
<td>0.24</td>
<td>0.35</td>
<td>0.38</td>
</tr>
<tr>
<td>BCEopt(GAA) ($\Omega$, $\chi$)$^d$ or 2BCEopt_mnr</td>
<td>0.112</td>
<td>0.0006</td>
<td>2.94</td>
<td>0.16</td>
<td>0.59</td>
<td>0.50</td>
</tr>
<tr>
<td>3MIN (all except o)$^d$</td>
<td>0.102–0.134</td>
<td>0.0003–0.002</td>
<td>2.4–4.3</td>
<td>0.21–0.57</td>
<td>0.13–1.66</td>
<td>0.21–1.45</td>
</tr>
<tr>
<td>3MIN(o)$^d$</td>
<td>0.81</td>
<td>0.59</td>
<td>0.01</td>
<td>0.33</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

$^a$ The definitions of the $R$ factors$^{65,66}$ ($R_1$–$R_3$) and corresponding data for all structures and are available as Supporting Information (Tables S4 and S5, respectively). $^b$ Absolute value of the difference between the experimental and theoretical chemical shifts calculated with NUCHEMICS.$^{64}$ $^c$ The best experimental structure 1DYANA(e) was based on the following constraints: 206 upls and 246 lols and 73 dihedral constraints including 31P-defined constraints for BI-like $\alpha$ and $\zeta$ ($0 \pm 120^\circ$) and $\zeta$ constraints for the G3pA4 step ($180 \pm 40^\circ$). $^d$ 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$^d$ 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. $^e$ 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 BUI$^E$ for the G3pA4 step and no constraints for the A5pA6 step.

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 3MIN(o) minimized with AMBER (red) (21st model, PDB 2k71). The BCE elastic line is shown as a yellow ribbon.

For 5’-d(GCGAAAGC)-3’ these data are as follows: $\delta^\epsilon$ is $-4.71 \pm 0.24$ ppm, the chemical shift of the second loop nucleotide is an outlier ($\delta_{L2}$ $-3.81$ ppm), and the L1 and N1 phosphates resonate at highest field ($\delta_{L1}$ $-4.95$ and $-4.87$ ppm, respectively). Except for L2, the observed average chemical shift and half-interval values are very reminiscent of the those of several well-defined DNA triloops ($\delta^\epsilon$ of -CGAAAG,-CGAG,-CGAAG,-CGTATG,-CGTATG,-CGTGG,-TCTTGA,$^{71}$ and -CGTTC,$^{73}$ triloops and the CGTTAG,$^{19,75}$ ATTTAT,$^{75,76}$ ATTTTG,$^{77}$ and GTTTTC$^{77}$ tetraloops are characterized by much lower $\delta^\epsilon$ values ($-4.15 \pm 0.17$, $-4.27 \pm 0.13$, $-4.09 \pm 0.36$, $-4.01 \pm 0.33$, $-3.90 \pm 0.42$, $-3.38 \pm 0.52$, $-4.07 \pm 0.26$, $-4.08 \pm 0.31$, and $-4.26 \pm 0.29$ ppm, respectively) and present very different 31P 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 $\tau$ torsion angle is often in the t conformation. The vast majority of the signals of such phosphates are shifted to low field. The $\delta^\epsilon$ values of this group are analogous to those of DNA duplexes ($-3.88 \pm 0.31$ and $-3.93 \pm 0.32$ ppm,$^{29}$ $-4.16 \pm 0.32$ to $-4.28 \pm 0.25$ ppm$^{29}$). Finally, two other triloops, -CAGAC,$^{69}$ and -TATCA,$^{69}$ present high $\delta^\epsilon$ 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 ($\delta_{N3}^E$, $-3.28$ and $-3.75$ ppm, respectively).

The much greater dispersion of RNA hairpin 31P chemical shifts ($\Delta \delta$ of the CUUG,$^{78}$ GAAA,$^{79}$ and UUCG$^{80}$ hairpins are 3.60, 3.10, and 2.63 ppm, respectively) has been a major impediment to exploiting 31P chemical shifts in structural studies of nucleic acids.$^{24,28}$ This survey points to a strong correlation between the 31P chemical shift pattern and the mesoscopic conformation of the loop nucleotides in DNA hairpins with high-field $\delta^\epsilon$ values.
**B_{II}**+ 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 \(B_{II}\) ellipse in Figure 3. These two outliers are also outside the 95% confidence interval depicted by the \(B_{II}\) ellipse. For comparative purposes the \((\epsilon, \zeta)\) torsion angle pairs of recent helical crystal structures\(^{34}\) 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 \(B_{II}+\) one includes those of the DNA GTTA tetraloop (PDB 1ac7),\(^{19}\) and the AAA (PDB 1xue),\(^{82}\) the GCA (PDB 1bhh),\(^{30}\) and the GAA (PDB 1pq),\(^{42}\) triloops. The \(B_{II}+\) 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 \(B_{II}\) 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’ demonstrated the presence of a base-intercalated antiparallel 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 coexist 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.

**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(\(t\)) structures) presented in Figure 2b and assessed in Table 3 is remarkable for three main reasons. First, 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 (\(\Theta, \phi, \gamma\)). 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 \(B_{II}+\) 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 destacking 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\(^{31}\) and subsequent modifications perform well with double helical conformations.\(^{82}\) 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 \(\gamma(6)\) (\(\gamma\) and \(t\), respectively), deviate significantly from B-DNA values while some averaging about the \(B_{II}+\) conformation is occurring at the G3pA4 and A5pA6 steps. A survey of the literature has revealed that this latter conformation or the \(B_{II}+\) 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 \(\Delta H\)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 B12π+ conformation (ζ, 143° and 149° for the 1DYNAne and 1DYNAnemin structures, respectively). With the exception of the P4 phosphate in the B12π+ 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 31P signals (δP ~ −4.6 ppm).

Resolution of most DNA structures by NMR to date has suffered from insufficient data. Although 13C and 15N labeling using NUCHEMICS64 and systematic BCE exploration of the 31P chemical shifts of selected DNA and RNA tetraloop sequences. This material is available free of charge for all the structures described in this work. Survey of the 31P chemical shifts of selected DNA triloop sequences.

As has been shown in this work, unusual 1H and 31P chemical shifts can also be translated into additional NMR constraints that facilitate structure determination using NUCHEMICS64 and systematic BCE exploration of (α,ε,ζ) 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. Szymerski for stimulating discussions and for access to the NMR facilities in his laboratory where this project was initiated.

Supporting Information Available: Partial 1H NMR chemical shifts of the low-temperature minor species. Relative intensities of the DQCOSY crosspeaks of the GAAA hairpin. Linewidths in the 1D 1H NMR spectrum, cross-peak widths, and relative intensities in the 2D [13P,1H]-COSY spectrum of the GAAA hairpin. R factor definitions and the corresponding values for all the structures described in this work. Survey of the 31P chemical shifts of selected DNA triloop sequences. Survey of the 31P chemical shifts of selected DNA and RNA tetraloop sequences. This material is available free of charge via the Internet at http://pubs.acs.org.

References and Notes
