Available online at www.sciencedirect.com
SCIENCE DIRECT®
C. R. Chimie 9 (2006) 426–432
COMPTES RENDUS CHIMIE
http://france.elsevier.com/direct/CRAS2C/

NMR and molecular modelling studies of an RNA hairpin containing a G-rich hexaloop

Flore Joli [a,*], Nadia Bouchemal [a], Brigitte Hartmann [b], Edith Hantz [a]

[a] *Laboratoire BioMoCeTi, CNRS UMR 7033, UFR SMBH, université Paris-13, 74, rue Marcel-Cachin, 93017 Bobigny cedex, France*
[b] *Laboratoire de biochimie théorique, CNRS UPR 9080, Institut de biologie physico-chimique, 13, rue Pierre-et Marie-Curie, 75005 Paris, France*

Received 21 March 2005; accepted 26 May 2005

Available online 06 September 2005

Abstract

A preferential target of antisense oligonucleotides directed towards the *+1* region of the mRNA *PGY/MDR1* gene coding for the P-glycoprotein is an RNA hairpin containing a G-rich hexaloop. Here, the hairpin structure is studied by NMR and molecular modelling. The conformation of the 6-bp stem is close to a regular A double helix, in spite of the presence of a G–U wobble pair. In the $^{5'}$r(GGGAUG)$^{3'}$ loop, NMR experiments highlight numerous sequential internucleotide distances, *C2′-endo* sugars and marked up and down-field phosphorus chemical shifts. Various molecular dynamics in explicit solvent made with or without NMR constraints show that *C2′-endo* sugars and some unusual backbone conformations are intrinsic properties of the six-nucleotide loops. Finally, we present a preliminary three-dimensional structure of the hairpin in which the apical loop exhibits a *U*-turn like, delimiting two groups of stacked bases: GGGA in continuation of the stem, and UG, isolated from the stem but connected to the first group by hydrogen bonds. ***To cite this article: F. Joli et al., C. R. Chimie 9 (2006)***.

Résumé

Une cible préférentielle des oligonucléotides antisens, située au niveau de la région +1 de l'ARNm du gène PGY/MDR1 codant pour la P-glycoprotéine, est une tige boucle d'ARN contenant une hexaboucle riche en guanine. La structure de cette tige-boucle a été étudiée par RMN et modélisation moléculaire. La conformation de la tige de 6 bp est proche d'une double hélice A régulière, et ce malgré la présence d'une paire wobble G–U. Dans la boucle $^{5'}$r(GGGAUG)$^{3'}$, les expériences RMN permettent d'observer un nombre important de distances séquentielles internucléotidiques, des sucres majoritairement *C2′-endo* et des déplacements chimiques du phosphore inhabituels. Plusieurs dynamiques moléculaires en solvant explicite et réalisées avec ou sans contraintes RMN, montrent que les sucres en *C2′-endo* et les conformations particulières du squelette sont des propriétés intrinsèques des boucles à six nucléotides. Enfin, nous présentons une structure tridimensionnelle préliminaire de la tige-boucle, dans laquelle la boucle est repliée en un motif proche d'un *U-turn*, délimitant deux groupes de bases empilées : GGGA dans le

* Corresponding author.
E-mail addresses: fjoli@smbh.univ-paris13.fr (F. Joli), nbouchemal@smbh.univ-paris13.fr (N. Bouchemal), Brigitte.Hartmann@ibpc.fr (B. Hartmann), hantz@smbh.univ-paris13.fr (E. Hantz).

1631-0748/$ - see front matter © 2005 Académie des sciences. Published by Elsevier SAS. All rights reserved.
doi:10.1016/j.crci.2005.05.011

prolongement de la tige et UG isolé de la tige, mais reliée au premier groupe par des liaisons hydrogène. ***Pour citer cet article : F. Joli et al., C. R. Chimie 9 (2006)***.

Keywords: RNA; Hairpin; Hexaloop; NMR; Molecular dynamics

Mots-clés : ARN ; Épingle à cheveux ; Hexaboucle ; RMN ; Dynamique moléculaire

1. Introduction

RNA exhibits a large diversity of conformations and, among them, hairpin structures are frequently encountered. Hairpins are often involved in interaction with proteins and nucleic acids [1–3]. Thus, a stem–loop structure, belonging to the mRNA region of the *PGY/MDR1* gene, is a preferential target of different antisense oligonucleotides that suppress the transmembrane P-glycoprotein (P-gp) function [4]. This glycoprotein, encoded by the *mdr1* gene, acts as an energy-dependent drug-efflux pump and is involved in the control of the cellular drug accumulation. The overexpression of P-gp is directly linked to multidrug resistance, a major problem in cancer therapeutics. The purpose of the present study is to determine the structure of the antisense target RNA hairpin in order to understand the structural role played by this RNA motif in this type of strategy.

This stem–loop structure contains a six-base-pair stem comprising a G–U mismatch and the G-rich hexaloop $^{5'}$r(GGGAUG)$^{3'}$. The purine-rich tetraloops are well-known, in particular the most commonly occurring GNRA (N: any nucleotide, R: purine) loops [5,6]. In contrast, little is known about the solution structural properties of purine-rich 6-nucleotide loops. Recently, the website SCOR (http://scor.lbl.gov) has published a general survey of all the hexaloop structures resolved using RX and NMR data. Among the solution structures, only two present 5-purine bases in the loop (PDB codes 1bvj and 1r2p) [3,7]. These loops are adenine-rich loops: GUAAAA (1bvj) and UGAAAG (2rp2). In both cases, the loops are structured in two distinct stacked groups, the first one comprising the GU or the UG bases, and the second one containing to the adenine tracts.

In the present work, the solution conformation of the GGGAUG loop closed by the G–U mismatch containing stem is studied by NMR and molecular modelling in explicit solvent. Special attention is paid to the sugar and the backbone conformations. In particular, the hexaloop intrinsic properties of these two components are investigated. Based on numerous internucleotide distances, a preliminary 3D structure is proposed, in which the loop structure consists of two groups of bases (four on the 5′-side and two in the 3′-side) separated by stacking break and a sharp turn located between the A and the U bases. The global feature is inverted with regard to those observed for the A-rich hexaloops (two bases on the 5′-side and four bases in the 3′-side). In all cases, the purine tracts are composed by either adenines or guanines are well stacked together.

2. Materials and methods

2.1. NMR measurements

Unlabelled RNA strand was synthesised by RNATEC and 98% ^{15}N–^{13}C labelled RNA strand was synthesised by SILANTES. The two samples were purified by HPLC. The samples were dissolved in an aqueous solution (10 mM sodium phosphate buffer, 10^{-5} M EDTA, pH 6.9). Final concentrations were 1 mM for the unlabelled sample and 0.8 mM for the ^{15}N–^{13}C labelled sample. The solutions were heated to 80 °C and slowly cooled to room temperature.

NMR experiments were performed using a 500-MHz VARIAN Unity INOVA spectrometer. The experiments were carried out with a 5-mm gradient indirect detection probe and a 5-mm triple (^{1}H, ^{13}C, ^{15}N) detection probe. NMR data were processed on a Silicon Graphics workstation using the FELIX program (MSI program San Diego, CA, USA).

A set of NOESY experiments in H_2O buffer were performed at 5 °C in order to slow down the exchange of the imino and amino protons, with a JUMP AND RETURN gradient pulse water suppression [8]. Non-exchangeable base protons were correlated to nucle-

otide spin systems as completely as possible using standard NOESY, DQF-COSY with a ^{31}P decoupling pulse and TOCSY experiments. In order to assign the phosphorus chemical shifts through their correlation with the H3′ sugar proton, ^{31}P–^{1}H NMR spectra were performed at 25 and 35 °C: a proton observe heteronuclear COSY, a single quantum coherence (HSQC) spectroscopy and a HSQC-TOCSY [9,10]. These experiments were acquired at 25 and 35 °C to facilitate the assignment.

NOE build-up rates were calculated from NOESY experiments at 25 °C with 75, 150 and 200 ms mixing times and were compared with the build-up rate of the U5, U15 and U17 H5 to H6 NOE (2.45 Å). Interproton distances were estimated from these experiments using the distance extrapolation method [11].

The ^{13}C, ^{15}N-labelled sample has been used to achieve the assignment, taking advantage of the RNA-Pack experiments [12]. 3D CHHC-COSY [13] completed the spin-system assignment of sugar protons. Intranucleotide correlations of sugar and base resonances were achieved using a triple-resonance multiple-quantum TROSY-HCN experiment [14], G- or A-HNC-TOCSY-CH [15] and C- or U-HNCCCH [16,17] were used to correlate imino or amino protons to their corresponding base H8/H6/H2 protons.

2.2. Structure calculations

The model was built and minimised with the internal coordinate program JUMNA [18], using the AMBER98 force field [19]. We started from a first strand composed of the $^{5'}$r(GAGGUCGGGAUG)$^{3'}$ sequence and a second strand containing the $^{5'}$r(GAUCUC)$^{3'}$ sequence (Fig. 1). The six bases of the two strands are paired in a RNA double helix in A-form, while the loop part is in an extended conformation. Constrained distances (d) were applied progressively by steps of 0.2 Å between the 3′-G12 of the first strand and the G13-5′ of the second strand (initially $d \sim 20$ Å) until the single strand part was folded in loop ($d \sim 2$ Å). This incrementing of 0.2 Å ensured a good energetic convergence at each stage, together with a gradual smooth relaxation of the folding structure. This structure was used as starting point in the AMBER program.

Molecular dynamics were performed using AMBER 7.0 program [20] and the Parm99 force field [21]. The RNA hairpin was neutralised with Na^+ counter-ions (one Na^+ vs. one phosphate group) and explicitly solvated by a 12-Å water shell in all directions (~6000 TIP3P water molecules) in a truncated octahedral box. After 2250 cycles of energy minimisation, the minimised system was heated to 300 K, rescaling the velocities as necessary, and coupling to a heat bath using the Berendsen algorithm [22]. The simulations were then performed at constant temperature and pressure (NTP) using the Berendsen algorithm. Bond lengths involving hydrogen atoms were constrained using Shake algorithm [23], which enabled an integration time step of 2 fs. Long-range electrostatic interactions were treated using particle mesh Ewald (PME) approach [24,25] with a 9-Å direct space cutoff, a direct sum tolerance criteria of 10^{-5}, and a reciprocal space charge grid spacing of roughly 1 Å.

Analyses of RNA structures were carried out using CURVES [18,26].

3. Results and discussion

3.1. Experimental data

3.1.1. UV measurements

UV intensity measurements as a function of the temperature show only one transition and a melting tem-

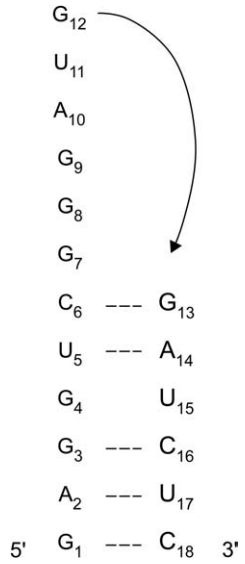

Fig. 1. Numbering of the 18-mer RNA hairpin nucleotides. The arrow indicates the constraint used to build the hairpin.

perature (T_m) of 58 °C for RNA concentrations from 10^{-5} to 10^{-4} mM. This single transition corresponds to the hairpin to random coil transition since the T_m is independent of concentration. Thus, NMR experiments, acquired at 25 °C, were made largely below the T_m value.

3.1.2. Hydrogen bonds

H_2O NOESY experiments allow exchangeable proton assignments. NOE cross-peaks show Watson–Crick base pairing in the stem, apart from the G–U mismatch that assumes the wobble conformation with hydrogen bonds between, on the one hand, the H–N3(U) and the O6(G) atoms, and, on the other hand, the O2(U) and the H–N1(G) atoms. No NOE cross-peaks belonging to the loop can be observed.

3.1.3. ^{31}P chemical shifts

A one-dimensional ^{31}P spectrum is shown in Fig. 2. The ^{31}P resonances cover a chemical shift range from –0.9 to 0.28 ppm, larger than for regular helices [27]. The most striking feature is the unusual chemical shifts of all the phosphates belonging to the loop region, towards either upfield (G7pG8) or down-field (all the phosphates between G8 and G13) regions. Nevertheless, in the $^{5'}$(GUC)$^{3'}$.$^{5'}$(GAU)$^{3'}$ part of the stem, comprising the G–U mismatch and the junction with the loop, some phosphate groups also appear out of the A double-helix range, well illustrated here by the chemical shifts of the $^{5'}$(GACG)$^{3'}$.$^{5'}$(UCUC)$^{3'}$ stem region. The structural interpretation of the unusual chemical shift seems trustworthy for B-DNA where only the correlated angles ζ and ε are variable [28]. However, in a loop, we can expect that α, β and γ backbone angles are not limited to solely one conformation. Theoretically, γ and β angle values can be estimated using the coupling constants $^3J_{H4'\text{-}H5'/H5''}$, $^4J_{H3'\text{-}H5'/H5''}$ and $^3J_{P\text{-}H5'/H5''}$, interpreted using Karplus relationship [29]. Unfortunately, these coupling constants are very difficult to obtain in our case, due to strong spectral overlaps of these proton resonances. So, in this context, the ^{31}P chemical shifts are not yet directly interpretable.

3.1.4. Sugar pucker

The DQF-COSY data are related to the sugar conformations (Fig. 3). 11 sugars are in *C3′-endo* conformation ($^3J_{H1'\text{–}H2'} < 3$ Hz, not seen in the spectrum): 10 belong to the stem, and the last corresponds to the G7 sugar. Around the mismatch, U5 is in *C2′-endo* conformation. In the loop, five sugars (the sixth being the G7 sugar) are clearly in the *C2′-endo* conformation, as the coupling constants are greater than 7 Hz.

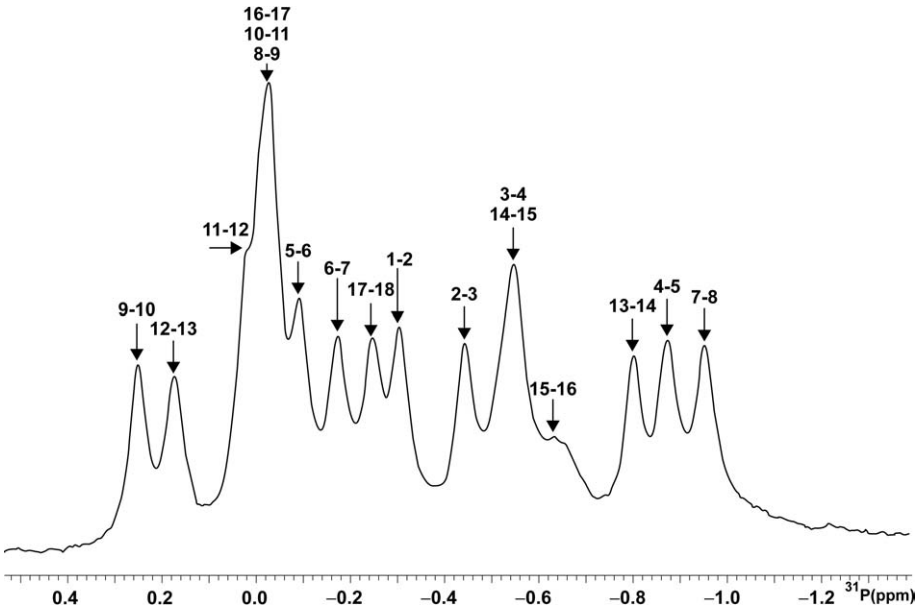

Fig. 2. 1D ^{31}P NMR spectrum at 202.43 MHz inD_2O at 25 °C (sw = 1100; nt = 512; external H_3PO_4 reference).

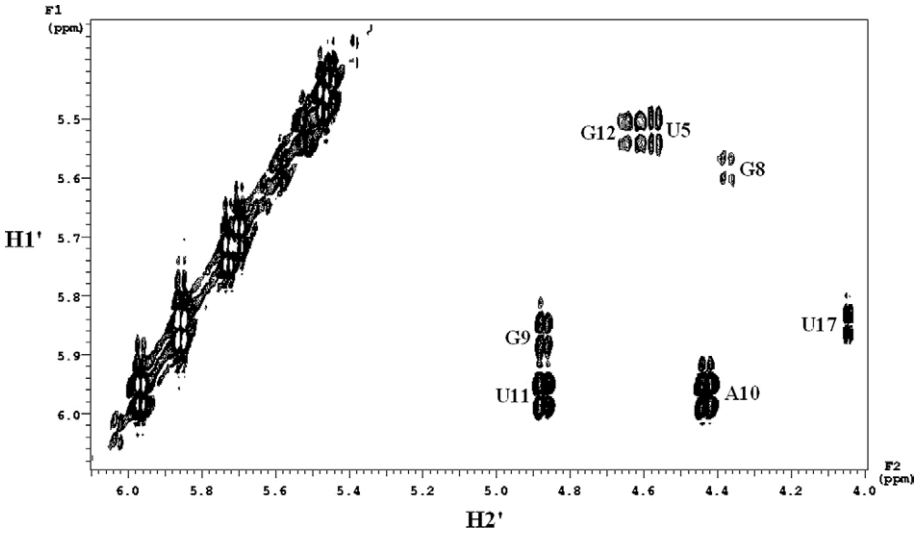

Fig. 3. Expanded phase-sensitive DQF-COSY spectrum in D_2O at 25 °C corresponding to the H1′ to H2′ region. The resolution was 1 Hz in F2 dimension.

3.1.5. NOE distances

A set of 156 intranucleotide distances and 66 internucleotide distances was collected and measured. In supplement to these quantitative distances, we add 24 qualitative internucleotide distances for which the corresponding peaks are (*i*) well seen but superimposed ($d < 5$ Å); (*ii*) too weak to be precisely measurable (5 Å $< d <$ 6.5 Å) and (*iii*) non-measurable ($d > 6.5$ Å). Therefore, the internucleotide data set that we will use later is composed of 47 and 43 distances for the stem and the loop, respectively.

Among the intranucleotides distances, the weak intensity of H6/H8–H1′ cross-peaks compared to H5–H6 reference cross-peaks indicates that, relative to the sugar moiety, all the bases adopt the *anti* conformation, comprising the loop bases. The internucleotide distances in the stem are typical of a RNA A-form. The double helix structure does not seem significantly distorted by the G4–U15 mismatch, although remarkably short H2′–H6/H8 and H6/H8-H6/H8 distances are observed for the G3pG4·U15pC16 step. In the stem–loop junctions and in the loop itself (C6–G13 part, corresponding to seven steps), a significant number of internucleotide distances is observed. All the sugar protons are correlated to the H6/H8 of the next base (distances less than 5 Å). The sequential base-base protons give measurable cross-peaks for five steps and thus two breaking-down points appear: between A10 and U11 and between G12 and G13. Thus, two stacking groups are drawn by the NOE cross-peaks. The first is composed of the C6-A10 bases and the second one of the U11 and G12 bases, isolated from the G13 stem base.

3.2. Modelling studies

3.2.1. Molecular dynamics without constraints

A trajectory of 12 ns without NMR constraints was performed and the MD and NMR distances were compared. In the stem, all the hydrogen bonds are stable, including those of the wobble G–U mismatch. The sugar conformations are in agreement with the NMR data, in particular the *C2′-endo* sugars found in the loop. Comparing the distances extracted from the MD (d_{MD}) and those extracted from the NOE cross-peaks ($d_{exp} \pm 10\%$, in order to take into account the experimental error), we find that the intranucleotide distances are respected in both the stem and the loop. The internucleotide distances are also in good agreement in the stem since the absolute value of ($d_{MD} - d_{exp}$) does not exceed 0.3 Å for 39 distances on a total of 47. The eight remaining discrepancies are all located between the wobble pair and the loop. In contrast, severe violations (between 1 and 2.4 Å) are observed in the loop around U11 and, in a less extend, around G7. In particular U11 stacks with both A10 and G12, while the experimental data clearly show a gap between U11 and A10. These discrepancies could originate in the simu-

lation duration that is not long enough for ensuring an extensive exploration of this complex conformational space. Finally, the backbone angles are consistent with an A-form in the stem whereas they undergo various transitions in the loop. Thus, we list nine different combinations for the $\varepsilon/\zeta/\alpha/\beta/\gamma$ series (from sugar to sugar).

3.2.2. Molecular dynamic with constraints

A second molecular dynamics of 5 ns was then performed under constraints. Only internucleotides distances (47 and 43 distances for the stem and the loop, respectively) were applied, since, on the one hand, intranucleotide distances and sugar conformations are spontaneously respected in the free MD and, on the other hand, the phosphate chemical shifts cannot be interpreted in terms of structural data (as discussed above in Section 3.1.3). The resulting average structure is shown in Fig. 4. The stem structure is close to the average structure obtained in the MD without constraints (RMSD 1.3 Å). It looks like a canonical A-form, with average values of –4.0 Å for the X-disp, 30° for the twist and 3.0 Å for the rise. Nevertheless, the mismatch induces some local perturbations: the base pair itself shows a shear value of –2.5 Å and the G3pG4.U15pC16 step is affected by a small twist of 18°. The internucleotide distances extracted from the NMR experiments are well-respected in the loop. The

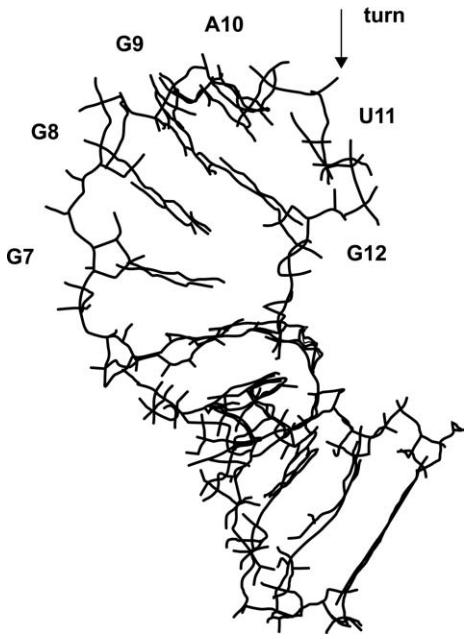

Fig. 4. A view of the average structure of the RNA hairpin.

C6, G7, G8, G9 and A10 bases form a first group as they are stacked. The A10 and U11 are coplanar, but not stacked. Finally, G12 is stacked with U11, remaining far from the G13 stem base. In regard to the free MD, introduction of internucleotide constraints are thus sufficient to change the relative position of U11 with its neighbours. Sugars remain stable in *C2′-endo* conformation, apart from the G7 sugar, in line with the NMR data. The backbone angles are less variable than in the free MD, and seven angle combinations on the nine detected in the free MD are explored. In particular, the sharp backbone turn observed between A10 and U11 stabilises the non-canonical conformers ε (*gauche–)* and ζ (*gauche+*). Finally, the stacking breaks and thus the fold, could be counterbalanced by stable hydrogen bonds observed between (*i*) N3(G9) and H–N1(G12), (*ii*) O4′(G9) and H–N2(G12), and (*iii*) H–N2(G9) and O6(G12).

3.2.3. Abasic loop dynamic

In order to investigate the backbone properties, a third molecular dynamics of 20 ns was performed on the RNA hairpin containing an abasic hexaloop. In absence of bases, the loop sugars undergo fast *C3′-endo* and *C2′-endo* transitions, and each sugar spends 30% of the time in *C2′-endo* conformation. Thus, the bases stabilise the *C2′-endo* conformers, but clearly take advantage of an intrinsic property of the hexaloop. The backbone angles explore a large conformational space as 14 combinations can be indexed. Nevertheless, some features appear common to our three MDs. The β angle is always in *trans* conformation and γ is in *gauche+* conformation in a large majority. The ε angle populates two domains: *trans* and *gauche–*. The most variable angles are ζ and α, which occupy the three domains: *gauche–*, *gauche+* and *trans*. Thus, in line with earlier crystallographic data analysis [30,31] made on different RNA structures, the existence of intrinsic torsion angle preferences is clear in a hexaloop. Nevertheless, further investigations will be essential to obtain a reliable relationship between these observations and the experimental ^{31}P chemical shifts.

4. Conclusion

By NMR and molecular dynamics in explicit solvent, we have determined the main structural charac-

teristics of an RNA stem capped by a G-rich hexaloop. The stem structures obtained by molecular dynamics with and without NMR constraints converge to the same A-type conformation. The wobble G–U mismatch moderately perturbs the overall conformation, despite sugars in *C2′-endo* conformations and unusual backbones values between the mismatch and the loop. For the hexaloop part of the hairpin, strongly ordered, we can establish the following facts. First, sugar puckers are found in a majority in *C2′-endo* conformations, probably to extend the strand with the help of numerous but limited unusual backbone angle conformations. This seems to be an intrinsic property for hexaloops, although further investigations are required to better describe the backbone behaviour. Second, the loop is stabilised by stacking interactions and hydrogen bonds. Thus, on the 5′-side, the four purine bases (G7, G8, G9 and A10) are stacked together and at the 3′-end of the loop, U11 stacks on G12, while remaining far from the stem. This fold is different but consistent with the two other available solution structures of A-rich hexaloops.

Acknowledgements

The laboratory of M. Delepierre (Institut Pasteur) is gratefully acknowledged for providing NMR facilities to perform the TROSY-HCN experiment used in this work. We thank L. Tuckerman for comments on the manuscript.

References

[1] P. Legault, J. Li, J. Mogridge, L.E. Kay, J. Greenblatt, Cell 93 (1998) 289.
[2] K.Y. Chang, I. Tinoco Jr., J. Mol. Biol. 269 (1997) 52.
[3] E.V. Puglisi, J.D. Puglisi, Nat. Struct. Biol. 5 (1998) 1033.
[4] E.V. Kostenko, R.S. Beabealashvilly, V.V. Vlassov, M.A. Zenkova, FEBS Lett. 475 (2000) 181.
[5] S.E. Butcher, T. Dieckmann, J. Feigon, J. Mol. Biol. 268 (1997) 348.
[6] H.A. Heus, A. Pardi, Science 253 (1991) 191.
[7] R.K. Sigel, D.G. Sashital, D.L. Abramovitz, A.G. Palmer, S.E. Butcher, A.M. Pyle, Nat. Struct. Mol. Biol. 11 (2004) 187.
[8] P. Plateau, M. Gueron, J. Am. Chem. Soc. 104 (1982) 7310.
[9] G. Bodenhausen, D. Ruben, Chem. Phys. Lett. 69 (1980) 185.
[10] T.J. Norwood, J. Boyd, J.E. Heritage, N. Soffe, I.D. Campbell, J. Magn. Reson. 87 (1990) 488.
[11] J.D. Baleja, B.D. Sykes, J. Magn. Reson. 91 (1991) 624.
[12] P.J. Lukavsky, J.D. Puglisi, Methods 25 (2001) 316.
[13] A. Pardi, E.P. Nikonowicz, J. Am. Chem. Soc. 114 (1992) 9202.
[14] B. Brutscher, J.P. Simorre, J. Biomol. NMR 21 (2001) 367.
[15] J.P. Simorre, G.R. Zimmermann, L. Mueller, A. Pardi, J. Biomol. NMR 7 (1996) 153.
[16] J.P. Simorre, G.R. Zimmermann, A. Pardi, B.T. Farmer 2nd, L. Mueller, J. Biomol. NMR 6 (1995) 427.
[17] J.P. Simorre, G.R. Zimmermann, L. Mueller, A. Pardi, J. Am. Chem. Soc. 118 (1996) 5316.
[18] R. Lavery, H. Sklenar, J. Biomol. Struct. Dyn. 6 (1988) 63.
[19] R. Lavery, K. Zakrzewska, H. Sklenar, Comput. Phys. Commun. 91 (1995) 135.
[20] D.A. Case, et al. University of California, San Francisco, CA, 2002.
[21] T.E. Cheatham 3rd, P. Cieplak, P.A. Kollman, J. Biomol. Struct. Dyn. 16 (1999) 845.
[22] H.J.C. Berendsen, J.P.M. Postma, W.F. Van Gunsteren, J. Chem. Phys. 81 (1984) 3684.
[23] J.P. Ryckaert, G. Ciccotti, H.J.C. Berendsen, J. Comput. Phys. 23 (1977) 327.
[24] T. Darden, D. York, L.J. Pedersen, J. Chem. Phys. 98 (1993) 10089.
[25] T.E. Cheatham 3rd, J.L. Miller, T. Fox, T. Darden, P.A. Kollman, J. Am. Chem. Soc. 117 (1995) 4193.
[26] R. Lavery, J. Biomol. Struct. Dyn. 3 (1988) 191.
[27] D.G. Gorenstein, in: Methods in Enzymology, Academic Press, 1992, pp. 254–286.
[28] C. Tisne, E. Hantz, B. Hartmann, M. Delepierre, J. Mol. Biol. 279 (1998) 127.
[29] P. Lankhorst, C. Haasnoot, C. Erkelens, C. Altona, J. Biomol. Struct. Dyn. 1 (1984) 1387.
[30] B. Schneider, Z. Moravek, H.M. Berman, Nucleic Acids Res. 32 (2004) 1666.
[31] L.J. Murray, W.B. Arendall 3rd, D.C. Richardson, J.S. Richardson, Proc. Natl Acad. Sci. USA 100 (2003) 13904.