DNA Folding

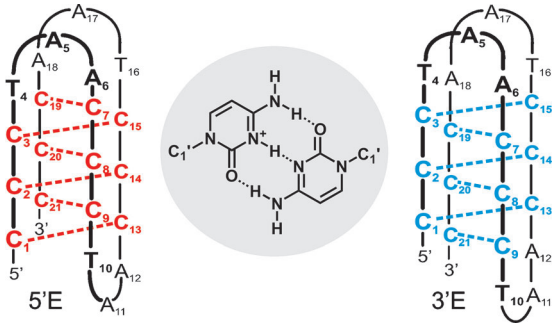
DOI: 10.1002/anie.201104938

Time-Resolved NMR Spectroscopic Studies of DNA i-Motif Folding Reveal Kinetic Partitioning**

*Anna Lena Lieblein, Janina Buck, Kai Schlepckow, Boris Fürtig, and Harald Schwalbe**

Dedicated to Professor Joachim Engels

In addition to the DNA double helix,[1] DNA oligonucleotides can adopt structures including G-quadruplexes[2] and i-motifs[3] that are stabilized by non-Watson–Crick base pairing. These structures consist of either guanosine- or cytidine-rich stretches interrupted by loop-forming nucleotides. Cytidine-rich i-motifs are formed at pH values of ≈6 and stabilize two strands with intercalated hemiprotonated C···H···C$^+$ base pairs (referred to as C·C$^+$; Scheme 1).[3,4] i-Motifs can differ in the number of base pairs, length, intercalation, and loop topology.[5] The biological function of i-motifs is under debate: proteins have been isolated that specifically bind to the C-enriched sequence of telomeric DNA.[6] In the insulin-linked polymorphic region (ILPR), nucleotides were identified to form both, i-motif and G-quadruplex structures leading to replication inhibition.[7] Furthermore, i-motif-based repression of transcription has been described in the c-*myc* promoter region.[8] Recently, i-motifs have been utilized as DNA nanodevices to monitor pH changes in vivo in *Drosophila* haemocytes[9] and in *Caenorhabditis elegans*.[10]



Scheme 1. Secondary structure of 21 nt long DNA i-motif with different intercalation topologies. Schematic representation of a C·C$^+$ base pair formed at pH 6.

Despite the potential importance of i-motifs in the regulation of gene expression and their use as probes monitoring spatiotemporal pH changes in live cells, little is known about the molecular mechanism of DNA i-motif formation. So far, i-motif folding has been investigated by surface plasmon resonance (SPR) and by quartz crystal microbalance with dissipation measurements (QCM-D).[11] FRET (Förster resonance energy transfer) studies revealed high reversibility of the pH-induced folding of i-motifs and multiphasic folding kinetics with folding and unfolding time constants in the order of minutes.[9]

Herein we investigate the structural changes and the kinetics of the pH-induced folding of a DNA i-motif. Static NMR spectroscopy experiments were performed to characterize the structure of a 21-nucleotide (nt) long i-motif (d(CCCTAA)$_3$CCC), a sequence found in vertebrate telomeres.[12] Four cytidine stretches are linked by TAA loop-forming nucleotides and an intramolecular i-motif that includes six C·C$^+$ base pairs is predicted to form. As shown in detail, certain i-motifs adopt an equilibrium of slowly interconverting conformers (Scheme 1). For the sequence investigated herein, two distinct conformations are populated at a ratio of 3:1.

We investigated the kinetics of folding of the i-motif initiated by a pH-jump from pH 9 to pH 6. Under these conditions, folding follows a kinetic partitioning mechanism, where two conformations form in the first step with a rate constant of the order of 2 min^{-1}. Subsequent refolding of the kinetically favored conformation to the thermodynamically more stable conformation is slow, with rate constants of the order of 10^{-3} min^{-1}. We propose that the two conformations differ in the intercalation topology of the C·C$^+$ base pairs. At equilibrium, the closing C·C$^+$ base pair can either be formed at the 5′-end of the C-rich strand (5′E) in the major conformation or at the 3′-end (3′E) in the minor conformation (Scheme 1). The slow conformational refolding can thus be described as a change in intercalation of C·C$^+$ base pairs.

The assignment of NMR resonance signals is a prerequisite for the analysis of folding kinetics at atomic resolution. Spectra of the DNA i-motif show five distinct imino proton resonances in the region of (15–16 ppm) characteristic for hemiprotonated C·C$^+$ base pairs (Figure 1C, top spectrum).[13] Theoretically, the NMR spectrum of six C·C$^+$ base pairs should result in six imino proton resonances.

To unambiguously assign the observed imino resonances, isotope-filtered NMR experiments on selectively labeled

[*] Dipl.-Chem. A. L. Lieblein, Dr. J. Buck, Dr. K. Schlepckow, Dr. B. Fürtig, Prof. Dr. H. Schwalbe
Institute for Organic Chemistry and Chemical Biology, Center of Biomolecular Magnetic Resonance
Johann Wolfgang Goethe-University Frankfurt/Main
Max-von-Laue-Strasse 7, 60438 Frankfurt (Germany)
E-mail: schwalbe@nmr.uni-frankfurt.de
Homepage: http://schwalbe.org.chemie.uni-frankfurt.de

[**] Work in the group of H.S. is supported by DFG and the state of Hessen (BMRZ). H.S. is member of the DFG-funded cluster of excellence: macromolecular complexes. We thank Elke Stirnal with help in HPLC purification of DNA samples, Dr. C. Richter, Dr. J. Rinnenthal, and Prof. Dr. A. Heckel for insightful discussions.

Supporting information for this article is available on the WWW under http://dx.doi.org/10.1002/anie.201104938.

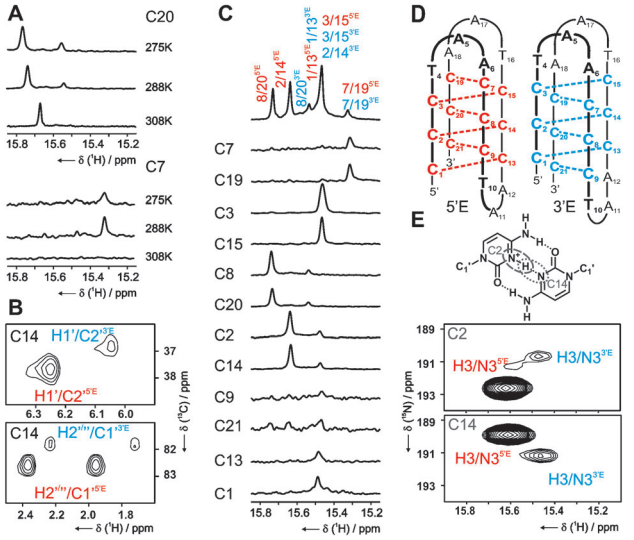


Figure 1. A) Imino proton region of the 1D ^{15}N-filtered spectra of nucleotide-selectively labeled DNA sequences C20 and C7 at different temperatures; B) HCCH-COSY[16] spectrum of the DNA with selectively ^{13}C-labeled C14, recorded at 600 MHz. Cross peaks of H1′/C2′ and H2′′′/C1′ for major (red) and minor (blue) conformation; C) Assignment of imino proton resonances using ^{15}N-filtered NMR experiments and nucleotide-selectively labeled DNA sequences. Top: 1D spectrum of the imino proton region with annotated base pairs (unlabeled DNA). Base pairs of the major and minor conformations are labeled and color-coded as indicated in the 1D spectrum. Bottom: 1D ^{15}N-filtered spectra of DNA sequences with DNAs containing only a single isotope-labeled cytidine. Spectra were recorded at 800 MHz with 128 scans for C2, C3, C8, C14, C15, and C20 and 1024 scans for C1, C7, C9, C13, C19, and C21. D) Secondary structures of the 21 nt DNA i-motif major (red, 5′E) and minor (blue, 3′E) conformations. E) Top: Schematic representation of a C·C$^+$ base pair of C2 and C14 sharing the same proton. Bottom: ^{15}N-HMQC-spectra of selectively labeled DNA sequences C2 and C14. Experimental conditions unless otherwise specified: 230–440 μM DNA, 25 mM phosphate buffer (pH 5.3), 800 MHz, 288 K.

DNA samples were performed, following a labeling strategy proposed by Dai et al.[17] Each cytidine nucleotide was 50% ^{15}N- and ^{13}C-labeled one at a time, resulting in 12 different labeled DNA sample for the resonance assignment of the imino protons. In each C·C$^+$ base pair, two cytidines share one proton and the proton chemical shifts are degenerate. By contrast, the N3 ^{15}N chemical shifts of the two cytidines are not degenerate as observed in ^{15}N-HMQC experiments (Figure 1E). Spectra of selectively labeled DNAs C2, C8, C14, and C20 show two signals with different intensities in line with the presence of a second conformation (Figure 1C). These findings are consistent with previous reports revealing conformational heterogeneity for i-motif structures.[17]

1D ^{15}N-filtered spectra displayed a second conformation for sequences C2, C4, C8, and C20 but not for the other nucleotides owing to signal overlap. In ^{13}C-HSQC spectra, however, a second set of signals is observed for all nucleotides (Supporting Information, Figure S1C). To further rule out that the additional signals (Figure 1B,C,E and S1C) are not due to NMR natural abundance signals, HCCH-COSY experiments were recorded for sequences with labeled C14 (Figure 1B and Figure S2) and C20 (Figure S2). In these experiments, we could unambiguously connect cross peaks of H1′ and H2′′′ for both conformations by magnetization transfer between ^{13}C-enriched carbon atoms C1′ and C2′ (Figure 1B). The possibility that the doubled peak set could arise from distorted symmetry effects can be excluded by comparison of ^{13}C correlation spectra of the selectively labeled samples, where we observed double peak sets with different intensities for all peaks.

To further support the existence of equilibrium between two different i-motif conformers, ^{15}N-filtered experiments were recorded at various temperatures. Data for two labeled sequences, C20 and C7, are shown in Figure 1A. C20 represents a base pair in the inner core of the i-motif in both the major and minor conformation, which we assigned to 5′E and 3′E, respectively. In the 5′E conformation, C7 is involved in base pairing next to the loops whereas it is in the inner core in the minor conformation 3′E. At 275 K, signals from C20 indicate a population ratio of 3:1. At 288 K, the population ratio increases to 4:1 and at 308 K, the minor conformation can no longer be observed. For C7, a single signal is observed which represents both the minor and major conformation because their chemical shifts are identical (Figure 1C). The intensity of the imino proton signal increases from 275 K to 288 K and disappears at 308 K. These observations are reversible. We conclude that an increase in temperature leads to destabilization of the minor conformer. The temperature-dependent measurements (Figure 1A and S3) are further supported by hydrogen-exchange experiments that report on individual base-pair stabilities.[14] The exchange rates show that base pairs for example, C8/C20 (Figure S4) are more stable in the major than in the minor conformation.

Since the conformational heterogeneity could also involve equilibrium between monomer and dimer conformations, we performed native polyacrylamide gel electrophoresis (PAGE) and circular dichroism (CD) melting experiments. Both experiments were conducted over a range of DNA concentrations. Native PAGE shows only a single band corresponding to monomeric species even at high DNA concentrations of 210 μM (Figure S5). In agreement with this observation, analysis of CD melting profiles revealed identical melting temperatures for samples within a DNA concentration range from 10 μM to 500 μM (Figure S6). We conclude that the stability of the i-motif is concentration-independent and both major and minor conformations are monomeric in the tested concentration range. Also pH- and salt-titrations showed no effect on the conformational equilibrium (Figure S7).

The formation of monomeric structures was further supported by NOE connectivity walks that are only compatible with an i-motif structure but not with a double-stranded DNA conformation (Figure S8). For the major conformation (5′E), the core intercalation could be confirmed by H1′–H1′ cross peaks across the narrow groove (Figure S1A/B). Detected NOE cross peaks between residues in the intercalation core and adjacent loop residues are furthermore indicative of an i-motif in a 5′E conformation. Due to the lower population and substantial signal overlap, a complete NOE connectivity walk could not be obtained for the minor

conformation (3′E). However, in ^{13}C-filtered NOESY experiment of, for example, C14 cross peaks of the proton to the spatially close nucleotide across the narrow groove C8(H1′) (major, 5′E) and C7(H1′) (minor, 3′E), can be observed (Figure S9). On the basis of CD and PAGE experiments and considering the absence of resonances from Watson–Crick base pairs in NMR spectra, we conclude that the minor conformation is monomeric. 1D ^{15}N-filtered experiments show similar base pairing for both conformations. Together with evidence from ^{13}C-filtered NOESY spectra, we propose that the minor population represents an i-motif structure adopting a 3′E conformation, in accordance with the results from kinetic investigation described below.

Utilizing a rapid-mixing device[18] we characterized the kinetics of the pH-induced i-motif folding and unfolding at atomic resolution using time-resolved NMR spectroscopy in situ. Unfolding of the DNA structure was initiated by a pH jump from pH 6 to 9 at 288 K by addition of NaOH solution; it was found to be fast with a decay rate of $k_{unfold} \approx 26\ min^{-1}$ (Figure S10/S11). i-Motif folding was induced from the non-base-paired state at pH 9 by injection of HCl solution to generate a pH value of 6. The build-up curves for imino proton signals involved in base pairs C8/C20$^{5'E}$ and C8/C20$^{3'E}$ (Figure 2) as well as C7/C19$^{5'E/3'E}$ and C2/C14$^{5'E}$ (Figure S12) were monitored.

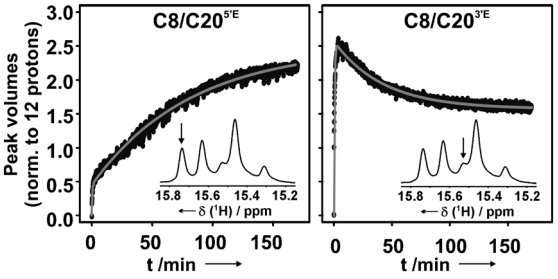


Figure 2. Kinetic traces of peak volumes of imino protons in base pairs C8/C20$^{5'E}$ and C8/C20$^{3'E}$ (the curves are normalized to 12 protons, which is the total number of imino protons in the 2 monomeric conformations). Fits are derived from differential equations (Supporting Information) and are shown as gray solid lines. Inset: Imino proton region with analyzed peak indicated by an arrow; 4 experiments before and 2044 experiments after the injection of HCl solution were recorded with an experimental time of 5 s per single kinetic point with 0.7 mM DNA at 800 MHz and 288 K.

Two distinct kinetic phases were observed for all the monitored signals: signals stemming from the major conformation C8/C20$^{5'E}$ increase continuously in intensity, while build-up curves from the minor conformation C8/C20$^{3'E}$ first increase and then decrease in intensity. The kinetics of folding for base pair C8/C20 can be monitored for the major and the minor conformation because their imino proton resonance signals are resolved. In both conformations, this base pair is located in the inner core of the i-motif and is well protected against solvent exchange. The intensities are not modulated by differential solvent exchange and therefore faithfully represent differences in the population of both conformations.

In the first folding step, both the major and minor conformations are formed with different rates. The rate constants for formation of the 3′E conformation are very similar ($k^{3'E} = 1.66 \pm 0.16\ min^{-1}$ and $k^{3'E} = 1.31 \pm 0.02\ min^{-1}$) for base pairs C8/C20$^{5'E}$ and C8/C20$^{3'E}$, respectively. The 3′E conformation forms faster that the 5′E conformation ($k^{5'E} = 0.21 \pm 0.02\ min^{-1}$ and $k^{5'E} = 0.89 \pm 0.02\ min^{-1}$ for base pairs C8/C20$^{5'E}$ and C8/C20$^{3'E}$, respectively). Hence, after initiation of folding, both 5′E and 3′E are formed rapidly. During the second slower transition, equilibrium is established with rates in the range of $10^{-3}\ min^{-1}$. Comparison of the transition between minor and major conformation reveals that $k_{eq}{}^{5'E}$ is larger than $k_{eq}{}^{3'E}$ indicating that the equilibrium is shifted towards the major conformation (5′E). All base pairs display the same kinetic parameters within experimental uncertainty (Table 1).

Table 1: Kinetic parameters obtained from the fit to the data shown in Figure 2 and Supporting Information Figure S12. Four spectra were recorded before and 2044 spectra after mixing with HCl solution.

Base pairs	$k^{5'E}$ [min^{-1}][a] $k_{eq}{}^{5'E}$ [10^{-3} min^{-1}][a]	$k^{3'E}$ [min^{-1}][a] $k_{eq}{}^{3'E}$[10^{-3} min^{-1}][a]
C8/C20$^{5'E}$	0.21 ± 0.02	1.66 ± 0.16
	7.50 ± 0.05	4.67 ± 0.07
C2/C14$^{5'E}$	0.79 ± 0.05	2.90 ± 0.18
	6.52 ± 0.07	4.56 ± 0.08
C8/C20$^{3'E}$	0.89 ± 0.02	1.31 ± 0.02
	14.93 ± 0.14	8.61 ± 0.10
C7/C19$^{5'E/3'E}$	1.57 ± 0.03	2.23 ± 0.04
	10.34 ± 0.08	2.91 ± 0.04

[a] The fitting function is given in the Supporting Information. The time resolution per kinetic data point was 5 s. Errors result from data fitting.

Taken together, we deduce the following i-motif folding pathway (Figure 3): starting from a non-base paired state, both conformations (5′E and 3′E) are formed first, albeit with different rates. The formation of the 3′E conformation is a factor of two faster than formation of the 5′E conformation. The slower phase of the kinetics then corresponds to the relaxation of the system towards equilibrium for which the 5′E conformation is favored. The slower second step is in line with a mechanism where the initially formed less-stable conformation has to unfold first (either fully or in a sequential way) to be able to refold to the more stable conformation. The refolding step requires a change of the intercalation topology.[19]

The results raise the question why formation of the less-stable 3′E conformation is kinetically favored. Inspection of the NMR structures of the i-motif with sequences d(5mCCT$_3$CCT$_3$ACCT$_3$CC) (Protein data bank (PDB) code: 1a83)[20] and d(CCCTAA5mCCCTAACCCUAA-CCCT) (PDB code: 1ELN)[21] reveals close proximity of two nucleobases on the double looped side of the intercalated stem. For the latter sequence, in the 5′E conformation, these positions are occupied by pyrimidine bases (T4 and T16) that can be easily accommodated and may be stabilized by formation of a T–T base pair. The C1′–C1′ distance in T–T base pairs at 10.4 Å is only marginally longer than in C·C$^+$

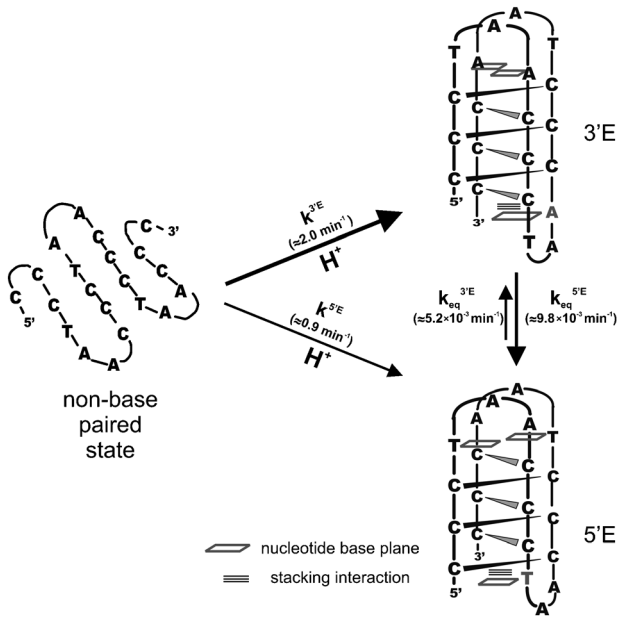


Figure 3. Model of the folding pathway of the DNA i-motif. After a pH jump from pH 9 to 6, major (5′E) and minor (3′E) conformations are formed. In a second transition, the minor conformation refolds towards the major i-motif structure until equilibrium is reached. Stacking interactions and nucleotide base planes are indicated in accordance with the known NMR structures (PDB code: 1a38[20] and 1ELN[21]).

base pairs (ca. 9.4 Å).[22] In contrast, in the 3′E conformation these positions are occupied by the sterically more demanding purine bases A6 and A18. We propose the differences in stabilities of these competing structural motifs to explain the higher stability of the 5′E conformation. The 3′E conformation forms faster due to stacking interactions on the single-loop side of the molecule: depending on the register of the intercalated stem, either the first or the last residue of the loop stacks on the first C·C$^+$ base pair on the single-loop side. In case of the 5′E conformation, a thymidine nucleobase stacks onto the stem, while an adenine nucleobase stacks onto the stem in the 3′E conformation.

The purine stacking interaction is energetically more favorable and leads to the initial stabilization of the 3′E conformation. Thus, the kinetically preferred 3′E conformation will refold into the thermodynamically more stable 5′E conformation in the second step of i-motif folding.

In conclusion, we investigated the intramolecular folding of a 21 nt long DNA found in telomeric DNA. The ground-state structure of this i-motif is conformationally heterogeneous. We utilized static and time-resolved NMR spectroscopy to dissect the folding kinetics of the i-motif. Detecting conformational heterogeneity within the folded state is the basis of the observed kinetic partitioning. The kinetically favored minor conformation is initially stabilized by stacking interactions and refolds into a major conformation that is stabilized by an extra T–T base pair. i-Motifs fold in the first step with a rate constant of 2 min^{-1} and refold to a kinetically favored conformation with rate constants of the order of 10^{-3} min^{-1}, while their unfolding is very rapid. Based on the slow kinetics, we conclude that i-motif regulation circuits are slow compared to those regulated by RNA elements including riboswitches[23] and RNA thermometers.[24] i-Motif binding proteins found in nuclear extracts[6] could therefore help stabilizing specific i-motif folds and act as folding chaperones to accelerate folding as previously shown for complex RNA molecules.[25] For applications of i-motifs as DNA nanomachines, our data provide structural restraints to improve response times and output efficiency in FRET-based applications for pH monitoring in live cells.[10]

Received: July 14, 2011
Revised: September 12, 2011
Published online: November 17, 2011

Keywords: DNA dynamics · DNA folding · i-motif · quadruplex structures · time-resolved NMR spectroscopy

[1] J. D. Watson, F. H. Crick, *Nature* **1953**, *171*, 737 – 738.
[2] M. Gellert, M. N. Lipsett, D. R. Davies, *Proc. Natl. Acad. Sci. USA* **1962**, *48*, 2013 – 2018.
[3] K. Gehring, J. L. Leroy, M. Gueron, *Nature* **1993**, *363*, 561 – 565.
[4] L. Chen, L. Cai, X. Zhang, A. Rich, *Biochemistry* **1994**, *33*, 13540 – 13546.
[5] M. Guéron, J. L. Leroy, *Curr. Opin. Struct. Biol.* **2000**, *10*, 326 – 331.
[6] E. Marsich, A. Piccini, L. E. Xodo, G. Manzini, *Nucleic Acids Res.* **1996**, *24*, 4029 – 4033.
[7] P. Catasti, X. Chen, L. L. Deaven, R. K. Moyzis, E. M. Bradbury, G. Gupta, *J. Mol. Biol.* **1997**, *272*, 369 – 382.
[8] J. Dai, E. Hatzakis, L. H. Hurley, D. Yang, *PLoS One* **2010**, *5*, e11647.
[9] S. Modi, G. S. M. D. Goswami, G. D. Gupta, S. Mayor, Y. Krishnan, *Nat. Nanotechnol.* **2009**, *4*, 325 – 330.
[10] S. Surana, J. M. Bhat, S. P. Koushika, Y. Krishnan, *Nat. Commun.* **2011**, DOI: 10.1038/ncomms1340.
[11] H. Xia, Y. Hou, T. Ngai, G. Zhang, *J. Phys. Chem. B* **2010**, *114*, 775 – 779.
[12] Y. Peng, X. Li, J. Ren, X. Qu, *Chem. Commun.* **2007**, 5176 – 5178.
[13] C. de Los Santos, M. Rosen, D. Patel, *Biochemistry* **1989**, *28*, 7282 – 7289.
[14] G. M. Dhavan, J. Lapham, S. Yang, D. M. Crothers, *J. Mol. Biol.* **1999**, *288*, 659 – 671.
[15] J. L. Leroy, M. Gueron, *Structure* **1995**, *3*, 101 – 120.
[16] H. Schwalbe, J. P. Marino, G. C. King, R. Wechselberger, W. Bermel, C. Griesinger, *J. Biomol. NMR* **1994**, *4*, 631 – 644.
[17] J. Dai, A. Ambrus, L. H. Hurley, D. Yang, *J. Am. Chem. Soc.* **2009**, *131*, 6102 – 6104.
[18] K. H. Mok, T. Nagashima, I. J. Day, J. A. Jones, C. J. Jones, C. M. Dobson, P. J. Hore, *J. Am. Chem. Soc.* **2003**, *125*, 12484 – 12492.
[19] B. Fürtig, P. Wenter, L. Reymond, C. Richter, S. Pitsch, H. Schwalbe, *J. Am. Chem. Soc.* **2007**, *129*, 16222 – 16229.
[20] X. Han, J. L. Leroy, M. Gueron, *J. Mol. Biol.* **1998**, *278*, 949 – 965.
[21] A. T. Phan, M. Gueron, J. L. Leroy, *J. Mol. Biol.* **2000**, *299*, 123 – 144.
[22] S. Nonin, J. L. Leroy, *J. Mol. Biol.* **1996**, *261*, 399 – 414.
[23] a) A. Haller, M. F. Souliere, R. Micura, *Acc. Chem. Res.* **2011**, DOI: 10.1021/ar200035g; b) L. Bastet, A. Dube, E. Masse, D. A. Lafontaine, *Mol. Microbiol.* **2011**, *80*, 1148 – 1154.
[24] J. Kortmann, S. Sczodrok, J. Rinnenthal, H. Schwalbe, F. Narberhaus, *Nucleic Acids Res.* **2011**, *39*, 2855 – 2868.
[25] L. Rajkowitsch, D. Chen, S. Stampfl, K. Semrad, C. Waldsich, O. Mayer, M. F. Jantsch, R. Konrat, U. Blasi, R. Schroeder, *RNA Biol.* **2007**, *4*, 118 – 130.

