Pergamon

Tetrahedron Vol. 50, No. 29, pp. 8711-8736, 1994

0040-4020/94 $7.00+0.00

0040-4020(94)00468-4

Remarkable Conformational Change Promoted by 3'-Ethylphosphate at the Branch-point of a Tetrameric Lariat-RNA Dictates Its Self-cleavage Reaction Modelling Some Catalytic RNAs (Ribozymes)

Bertrand Rousse, Christian Sund, Corine Glemarec, Anders Sandström, Peter Agback & Jyoti Chattopadhyaya*

Department of Bioorganic Chemistry, Box 581, Biomedical Center,
University of Uppsala, S-751 23 Uppsala, Sweden

*Abstract: Small synthetic lariat RNAs have been found to undergo site specific self-cleavage to give an acyclic branched-RNA with 2',3'-cyclic phosphate and a 5'-hydroxyl termini, which is reminiscent of the products formed in some catalytic RNAs. These lariat-RNAs are much smaller than the natural catalytic RNAs such as the hammerhead ribozyme ($k = \sim 1\ min^{-1}$ at 37 °C), and their rate of the self-cleavage is also much slower ($k = 0.25 \times 10^{-4}\ min^{-1}$ for lariat hexamer **18**, and $0.16 \times 10^{-3}\ min^{-1}$ for lariat heptamer **19** at 22 °C). We have shown that the trinucleotidyl loop in the tetrameric and pentameric lariat-RNAs (ref. 10) is completely stable whereas the tetranucleotidyl or pentanucleotidyl loop in the hexameric or heptameric lariat-RNA (ref. 10-13) does indeed have the required local and global conformation promoting the self-cleavage. It has been also shown that simple 2'→5' or 3'→5'-linked cyclic RNAs, **16** and **17**, respectively, are completely stable and their structures are considerably different from the self-cleaving lariat-RNAs such as **18** or **19**. In our search to explore the optimal structural requirement for the self-cleavage reaction of RNA, we have now synthesized **14** in which the branch-point adenosine has a 2'→5'-linked tetranucleotidyl loop and a 3'-ethylphosphate moiety mimicking the 3'-tail of the lariat-hexamer **18**. We here report that the unique 3'-ethylphosphate function at the branch-point in **14** is the key structural feature that orchestrates its self-cleavage reaction ($k = 0.15 \times 10^{-4}\ min^{-1}$ at 19 °C) compared to the stable 2'→5'-linked cyclic RNA **16** (see Fig. 1). We also report the detailed conformational features of the self-cleaving tetrameric lariat-RNA **14** by 500 MHz NMR spectroscopy and Molecular dynamics simulations in the aqueous environment. A comparative study of the temperature dependence of the N ⇄ S equilibrium for the lariat-tetramer **14** and the 2'→5'-linked cyclic tetramer **16** shows that the A^1 residue in **14** is in 92% S-type conformation at 20 °C, wheras it is only in 55% S in **16** with a 3'-hydroxyl group. This displacement of the N ⇄ S pseudorotational equilibrium toward the S geometry is due to the enhanced gauche effect of the 3'-OPO_3Et^- group at the branch-point adenosine in **14** compared to 3'-OH group in **16**. This 3'-OPO_3Et^- group promoted stabilization of the S geometry at the branch-point by $\Delta H \approx 4\ kcal.mol^{-1}$ in **14** is the conformational driving force promoting its unique self-cleavage reaction. The comparison of $\Delta H°$ and $\Delta S°$ of the N ⇄ S pseudorotational equilibria in **14** and **16** (see Table 5) clearly shows the remarkable effect of the 3'-ethylphosphate group in **14** in being able to dictate the conformational changes from the sugar moiety of the branch-point adenosine to the entire molecule (conformational transmission). Thus the S conformation in A^1, U^2 and C^6 sugar moieties is clearly thermodynamically more stabilized while it is considerably destabilized in G^3 owing to the 3'-ethylphosphate group in **14** compared to **16**. It is interesting to note that the magnitude of enthalpy and entropy for the North to South transition of the A^1 sugar in **14** is comparable to the enthalpy and entropy of transition between the A- and B-form of the lariat hexamer **18** (ref. 12). This self-cleaving tetrameric lariat-RNA **14** is the smallest RNA molecule hitherto known to undergo the self-cleavage reaction and hence it is the simplest model of the active cleavage site of the natural self-cleaving catalytic RNA.*

Most of the catalytically active natural RNA molecules are large and form complex tertiary structures[1-8]. In several cases, the catalytic activity includes the site specific self-cleavage of a phosphodiester bond to give a 2',3'-cyclic phosphate and a 5'-hydroxyl terminus[1,2]. This self-cleavage reaction occurs by a transesterification

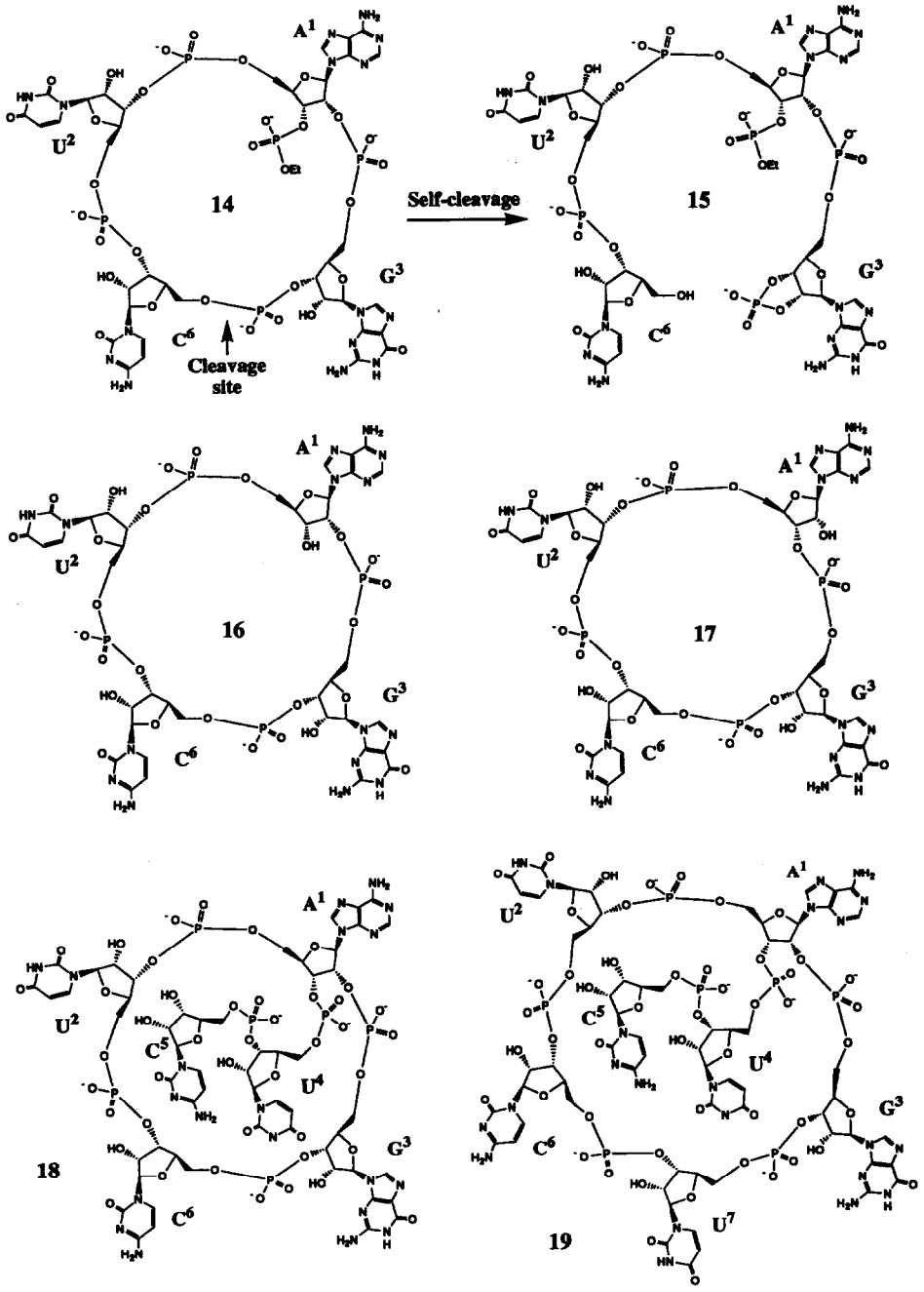
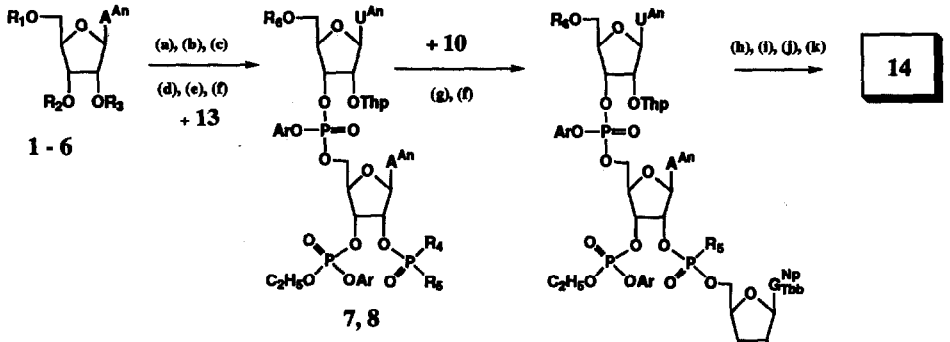


Figure 1: The various synthetic self-cleaving lariat-RNA model systems.

reaction which takes place by the direct attack of the 2'-hydroxyl group on the vicinal phosphate, and by in-line displacement of the 5'-oxygen. The reaction probably occurs via a trigonal bipyramidal phosphate that may be a transition state or an intermediate[9,11,12,39]. The detailed chemical mechanism and the conformational requirements for the cleavage reaction is however unknown. To understand the mechanism of the self-cleavage reaction and the local conformation necessary for the specificity of the cleavage reaction, it will be necessary to determine the 3D structure of a self-cleaving RNA. There are however two main problems which make the direct study of these self-cleaving systems difficult: (i) The rate of the cleavage reaction is very fast (*ca.* 1 min^{-1} at 37 °C), which means that the RNA undergoes self-cleavage reaction as soon as it is dissolved in the aqueous solution, (ii) The natural catalytic RNAs are very large, and therefore a full structure determination by NMR spectroscopy is difficult due to the spectral overlap. It should be also noted that the conformational studies on the cleavage site geometry carried out on the analogs of catalytic RNAs, in which the cleavage is blocked by substituting the 2'-OH with 2'-OMe, 2'-F or by 2'-H or by any other chemical alteration at the cleavage site, would give artificial results because of the obstruction of the natural gauche effect of 2'-OH group, the way it drives N⇄ S equilibrium and the way it involves itself through hydrogen bonding in the folding of RNA. We have recently found that some small lariat RNAs which mimic the sequence at and around the branch-point in Group II type splicing undergo site specific self-cleavage to give a 2',3'-cyclic phosphate and a 5'-hydroxyl termini[10-13]. These systems (Fig. 1) are much smaller than the natural catalytic RNAs such as the hammerhead ribozyme, and their rate of self-cleavage is also much slower: for lariat hexamer **18**: $k = 0.25 \times 10^{-4}$ min^{-1}, and for lariat heptamer **19**: $k = 0.16 \times 10^{-3}$ min^{-1} at 22 °C. For these reasons, we believe that our small lariat-RNA systems are good models to study the structural and conformational requirement for the self-cleavage reaction. We have thus shown that the trinucleotidyl loop in the tetrameric[10,13] and pentameric[10,13] lariat-RNA is completely stable whereas the tetranucleotidyl or the pentanucleotidyl loop in hexameric[11-13] or heptameric[11-13] lariat-RNA does indeed have the required local and global conformation that promotes the self-cleavage. In these works[10-13], we have also shown that simple 2'→5' or 3'→5'-linked cyclic RNAs, **16** and **17**, respectively, are completely stable and their structures are considerably different from the self-cleaving lariat-RNAs such as **18** or **19**. In our search to determine the optimal structural requirement for the self-cleavage reaction of RNA, we have now synthesized **14** which is an analogue of **16**. The simple lariat RNA **14** has the 3'-ethylphosphate function at the branch-point to mimic the 3'-tail of the lariat-hexamer **18**. We here report that the unique 3'-ethylphosphate function at the branch-point in **14** is the key structural feature that orchestrates the self-cleavage reaction of the lariat tetramer **14** compared to its parent 2'→5'-linked cyclic RNA **16** (Fig. 1). We also report the detailed conformational features of the self-cleaving tetrameric lariat-RNA **14** by 500 MHz NMR spectroscopy and Molecular dynamics simulation study in the aqueous environment, and compare some of these conformational features with those of the *stable* counterparts **16** and **17**, and *self-cleaving* lariat hexamer **18** and heptamer **19** which we have studied previously[11-13]. It may also be noted that the self-cleaving tetrameric lariat-RNA **14** is the smallest RNA molecule hitherto known to undergo self-cleavage reaction and hence this study provides the simplest model of the active cleavage site of the natural self-cleaving catalytic RNA.

(A) Synthesis of lariat (2'→5')-linked RNA tetramer-(3'-ethylphosphate) 14.

The lariat 2'→5'-linked tetramer-3'-ethylphosphate **14** was prepared essentially using the procedures developed in the phosphotriester chemistry[14,15] (see Scheme 1). The key building block for the synthesis of

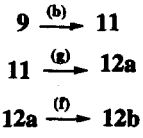


1. R_1 = CPTr, R_2 = H, R_3 = Px
2. R_1 = CPTr, R_2 = P(O)(OAr)($O^-Et_3NH^+$), R_3 = Px
3. R_1 = CPTr, R_2 = P(O)(OAr)($O^-Et_3NH^+$), R_3 = H
4. R_1 = CPTr, R_2 = P(O)(OAr)($O^-Et_3NH^+$), R_3 = P(O)(OCe)(O-(4-Npe))
5. R_1 = CPTr, R_2 = P(O)(OAr)(OC_2H_5), R_3 = P(O)(OCe)(O-(4-Npe))
6. R_1 = H, R_2 = P(O)(OAr)(OC_2H_5), R_3 = P(O)(OCe)(O-(4-Npe))

7. R_4 = OCe, R_5 = O-(4-Npe) ,R_6 = DMTr
8. $R_4 = O^-NH_4^+$, R_5 = O-(4-Npe) ,R_6 = DMTr

9. R_5 = O-(4-Npe) , R_6 = DMTr, R_7 = OCe
11. R_5 = O-(4-Npe) , R_6 = H, R_7 = OCe
12a. R_5 = O-(4-Npe) , R_6 = H, $R_7 = O^-Et_3NH^+$

12b. R_5 = O-(4-Npe)

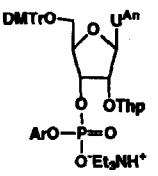


9 $\xrightarrow{(b)}$ 11

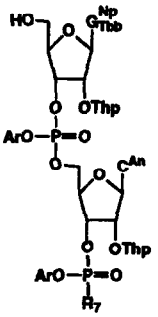
11 $\xrightarrow{(g)}$ 12a

12a $\xrightarrow{(f)}$ 12b

REAGENTS
(a): O-chlorophenylphosphoro-bis-(1,2,4 triazolide), pyridine, MeCN , RT;
(b) : TCA , 2% MeOH-CH_2Cl_2, 0°C;
(c) : (4-NpeO)(CeO)PN(i-Pr)$_2$/tetrazole, MeCN/DMF , RT;
(d) : Dry ethanol, MsCl, MeIm, pyridine , RT;
(e): Hydrazine hydrate, pyridine/AcOH (3:1) , 0°C;
(f): MSNT , pyridine , RT;
(g): Et_3N (20 eq), pyridine, RT;
(h): DBU (10 eq.), pyridine, RT;
(i): TMG /NBO , dioxane/H_2O (8;2) , RT;
(j): Conc. NH_3 (aq) , RT;
(k): 80% aq. AcOH , RT.



13



10

Abbreviations

A: adenine-9-yl, **An:** 4-anisoyl, **Ar:** 2-chlorophenyl, **C:** cytosine-1-yl, **Ce:** 2-cyanoethyl, **CPTr-Br:** 4,4',4"-tris(4,5-dichlorophthalimido)trityl bromide, **DBU:** 1,8-diazabicyclo[5.4.0]undec-7-ene, **DMTr:** 4,4'-dimethoxytrityl, **G:** guanine-9-yl, **MeIm:** N-methylimidazole, **Ms-Cl:** 1-mesitylene-sulfonylchloride, **MSNT:** 1-mesitylenesulfonyl-3-nitro-1,2,4-triazole, **NBO:** *syn*-4-nitro-benzaldoxime, **Np:** 2-nitrophenyl, **Npe:** 4-nitrophenylethyl, **Px:** pixyl or 9-phenylxanthen--9-yl, **Tbb:** *t*-butylbenzoyl, **TCA:** trichloroacetic acid, **Thp:** tetrahydropyran-2-yl, **TMG:** 1,1,3,3-tetramethylguanidine, **U:** uracil-1-yl.

Scheme 1

lariat RNA **14** was the appropriately protected 2',3'-vicinal phosphate[15] derivative of the branch-point adenosine **5** which was synthesized in five steps starting from 5'-*O*-(4,4',4"-tris(4,5-dichlorophthalimido)trityl)-6-*N*-anisoyl-2'-*O*-(9-phenylxanthen-9-yl) (pixyl)-adenosine (**1**) [**1** → **2** → **3** → **4** → **5**].

6-*N*-anisoyl-2'-*O*-pixyl-adenosine was tritylated with 4,4',4"-tris(4,5-dichlorophthalimido)trityl bromide (CPTr-Br)[16,17] in a regioselective manner to give compound **1** (65%), which was converted to the 3'-(2-chlorophenyl)phosphate derivative **2** (78%). The 2'-*O*-pixyl group was then removed from **2** under acidic condition (0.055 M trichloroacetic acid in 2% MeOH in CH_2Cl_2)[18] to give **3** (90%), which was then phosphorylated using (2-cyanoethoxy)-(2-(4-nitrophenyl)ethoxy)diisopropylaminophosphine[15] followed by an oxidation to give **4** (59%). The 3'-phosphodiester group in **4** was converted to the corresponding 3'-phosphotriester **5** (75%) using ethanol in presence of 1-mesitylenesulfonyl chloride (Ms-Cl) and 1-methylimidazole (MeIm)[19] as the activating agent. The 5'-*O*-CPTr group[16,17] was then removed from **5** by a brief treatment with a 1 M solution of hydrazine hydrate in pyridine-acetic acid (3:1, v/v) to afford **6** (77%). Compound **6** was then coupled to a uridine block (**13**) in presence of 1-mesitylenesulfonyl-(3-nitro-1,2,4-triazole) (MSNT)[14] to give the dimer **7** (50%). The 2'-cyanoethylphosphate group at the branch point in **7** was then converted to the corresponding phosphodiester function (20 eq of Et_3N) to give **8** (80%), which was condensed with the 5'-hydroxy block **10** in presence of MSNT to give **9** (79%). Treatment of **9** under the acidic condition, described for the conversion of **2** → **3**, gave the 5'-hydroxy tetramer **11** (50%). Finally the 3'-phosphotriester block **11** was converted to the phosphodiester **12a** (60%) using the same conditions described for the synthesis of **8**. The oligomer **12a** was intramolecularly cyclized[11,12,15,20] to give the protected tetrameric lariat RNA **12b**(63%) using MSNT (15 eq). Deprotection and purification of **12b** were carried out in the usual manner[21] to give **14** (30%).

(B) Characterization of lariat RNA tetramer 14 by NMR spectroscopy

All non-exchangeable protons and phosphorus resonances were assigned using 1D- and 2D-NMR experiments (^1H at 500 MHz). The H6 and H5 protons of uridine and cytidine are distinguished by their different coupling constants, 8.1 and 7.6 Hz respectively. The methyl protons of the ethyl-phosphate group as a triplet are easily observed in the 1D-spectrum (Fig. 2). The clean-TOCSY spectrum showed four relays indicating the presence of four pentose sugar residues consistent with the lariat structure shown in **14** (Fig. 1). These sugar residues were then connected with the aromatic protons of the respective nucleobases by the observed NOE connectivities using the ROESY spectrum (Fig. 7). Five phosphorous resonances were also observed for the five constituent phosphodiester residues in **14** (Fig. 2B). Finally, the assignments of the phosphate resonances, shown in Fig. 2B, were achieved by a ^1H-^{31}P correlation spectrum (Fig. 5) in which each phosphate residue was correlated to their respective 3' and 5' sugar protons (and to the ethyl group) on the basis of J-coupling: the phosphate resonance at δ1.51 was connected to H3'A^1 at δ4.74 and CH_2 of ethyl at δ3.99; the phosphate resonance at δ1.34 was connected to H3'U^2 at δ4.76 and to H5'/5"A^1 at δ4.39 and 4.11; the phosphate resonance at δ0.73 was connected to H3'C^6 at δ4.42 and to H5'/5"U^2 at δ4.09; the phosphate resonance at δ -0.15 was connected to H3'G^3 at δ4.57 and to H5'/5"C^6 at δ4.47 and 4.09; the phosphate resonance at δ -0.27 was connected to H2'A^1 at δ5.10 and to H5'/5"G^3 at δ4.29 and 4.10. That the intramolecular cyclization of the lariat-loop indeed has taken place between the uridine (U^2) and cytidine (C^6) (see the chemistry part for details) residues is evident from their common internucleotidyl phosphate resonance

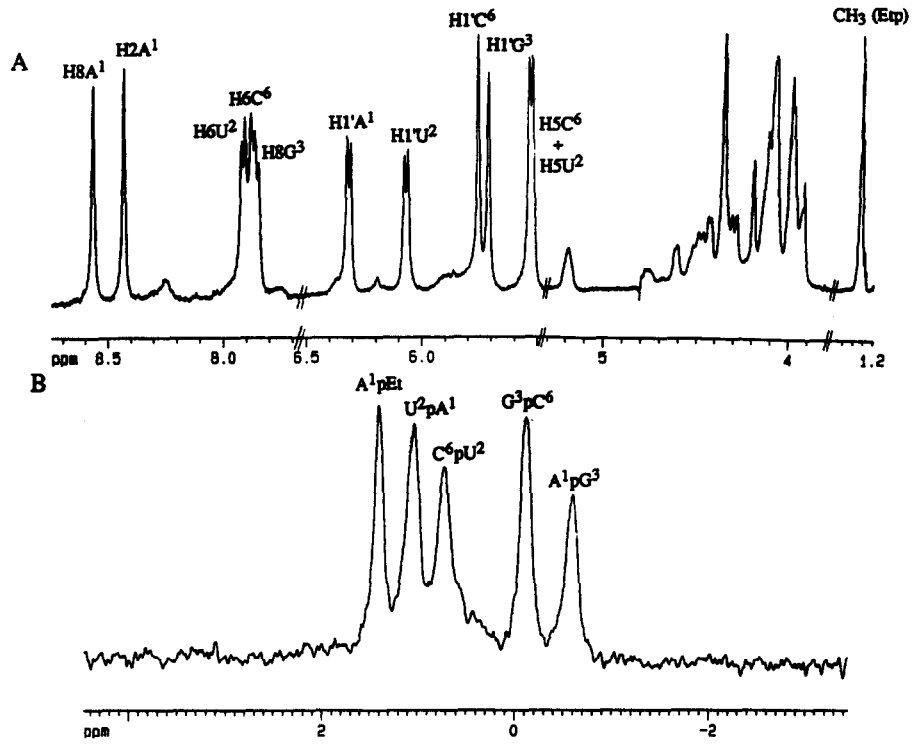


Figure 2: (A) 1D ^{1}H-NMR spectrum of lariat tetramer **14**, recorded at 15°C in $^{2}H_2O$ (reference: δ_{CH_3CN} = 2.0 ppm). (B) 1D ^{31}P-NMR spectrum of **14**, recorded at 15°C in $^{2}H_2O$ (reference: $\delta_{3',5'\text{-cAMP}}$ = 0.0 ppm).

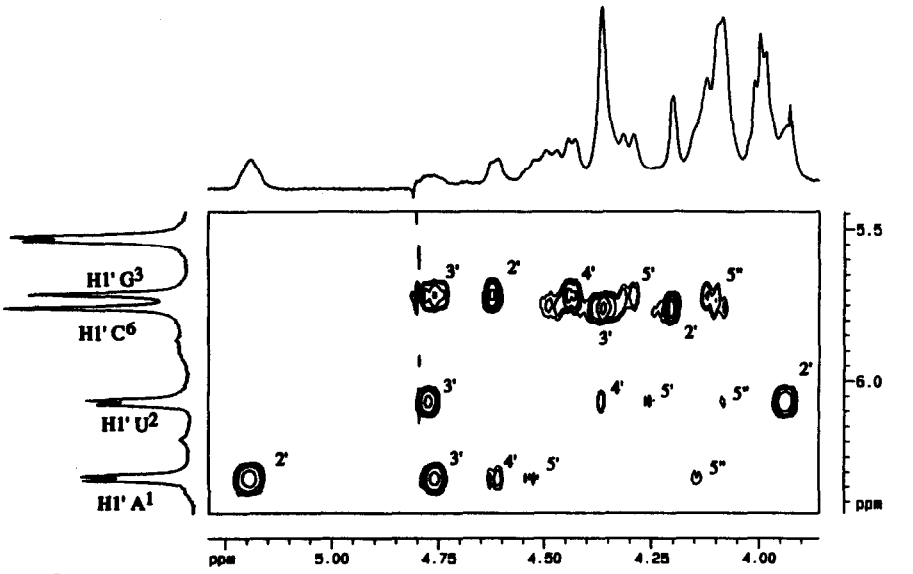


Figure 3: Clean-TOCSY spectrum of lariat tetramer **14**, recorded at 15°C in $^{2}H_2O$. The relays from the four different sugar residues of **14** are readily observed.

which absorbed at δ0.73 ppm showing connectivities with the H5'/5" protons of U^2 at δ4.09 ppm and the H3' proton of C^6 at δ4.42 ppm (Fig. 5). This proves that the lariat-loop has indeed been formed as shown in **14**.

(C) Self-cleavage reaction of lariat-RNA tetramer 14

At 19 °C, the ^{1}H- and ^{31}P-NMR spectra of the lariat tetramer **14** in aqueous solution show the gradual appearence of another distinct set of resonances which clearly account for the site-specific self-cleavage product **15**. We have noted that while the intensities of the resonances attributable to the self-cleavage product **15** increased as a function of time (Fig. 4), the intensities of the ^{1}H- and ^{31}P-resonances of lariat-RNA **14** decreased. One of the new ^{31}P-resonances that appeared as the self-cleavage reaction progressed is indeed very deshielded compared to the other phosphate resonances. This downfield phosphate absorbs at δ21 ppm (see Fig. 5), which is characteristic for the resonance of a 2',3'-cyclic phosphate[22,11-13]. In the 2D ^{1}H-^{31}P chemical shift correlation experiment, the ^{31}P-resonance at δ21 ppm shows cross peaks with both H2' and H3' protons of the G^3 nucleotide belonging to the new set of resonances for **15**. This phosphate did not show any correlation with H5' or H5" protons, while all other phosphate resonances between 0 and 1 ppm showed clear connectivities to both a H3' and H5'/H5" protons suggesting that the latter group resonances is indeed from internucleotidyl phosphates and the former is from 2',3'-cyclic phosphate (Fig. 5). This together with the measurement of $^3J_{H2'P}$ and $^3J_{H3'P}$ couplings of *ca.* 7 Hz for the G^3 of **15** proves that the site of phosphodiester self-cleavage is specific and occurs at the 3'-phosphate of G^3 in **14** to give a product **15** with a 2',3'-cyclic phosphate of G^3 and a 5'-hydroxyl termini at C^6. Furthermore, a detailed examination of a Clean-TOCSY spectra (Fig. 6) of the self-cleavage reaction mixture showed *only two distinct sets of sugar proton resonances* attributable to **14** and **15**. Note that the set of resonances from four sugar residues of **15** has been identified by ^{1}H-^{31}P correlation with the distinctly identifiable cyclic 2',3'-phosphate of G^3, and the second set has been identified from an independent Clean-TOCSY spectra of the pure starting material **14** (Fig. 3). The self-cleavage reaction of **14** at a concentration of 4 mM at pH 7 (19 °C) has been followed for more than 3 months by ^{1}H-NMR spectroscopy and the rate (k) was found to be 0.15×10^{-4} min^{-1}.

Table 1: ^{1}H-NMR chemical shifts (δ scale, referenced to CH_3CN set at 2.00 ppm) and ^{31}P-NMR chemical shifts (δ scale referenced to cAMP set at 0.00 ppm) for **14** and **15** at 292 K in ^2H_2O.

Compd.		A^1	U^2	G^3	C^6	Compd.		A^1	U^2	G^3	C^6
	H1'	6.32	6.07	5.69	5.77		H1'	6.13	5.59	5.84	5.74
	H2'	5.10	3.94	4.79	4.32		H2'	5.22	3.87	5.10	
	H3'	4.74	4.76	4.57	4.42		H3'	4.97	4.38	5.02	
	H4'	4.60	4.37	4.42	4.32		H4'	4.41	4.02	4.33	4.19
14	H5'	4.39	4.09	4.29	4.47	**15**	H5'	4.14	3.97	4.00	3.87
	H5"	4.11	4.09	4.10	4.09		H5"	3.97	3.87	4.00	3.75
	H8/H6	8.53	7.87	7.78	7.79		H8/H6	8.18	7.62	7.68	7.56
	H2/H5	8.20	5.51	/	5.51		H2/H5	7.94	5.74	/	5.85
	2'P	-0.27	/	/	/		2'P	0.11	/	21.0	/
	3'P	1.51	1.34	-0.15	0.73		3'P	1.34	0.86	21.0	0.60

(D) Conformation of the self-cleaving lariat-RNA 14

(i) Assignment

The ^{1}H-NMR resonances of **14** and **15** have been assigned from 2D Clean-TOCSY[23] (Figs. 3 and 6), DQF-COSY[24] and NOESY[25] and ROESY[25] experiments. The Clean-TOCSY and the DQF-COSY were used

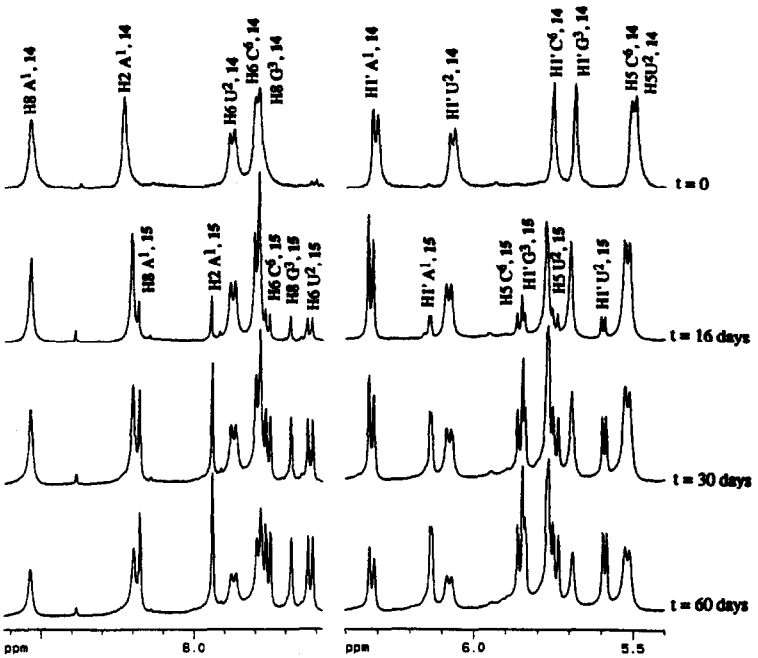


Figure 4: ^{1}H-NMR spectra of the aromatic and anomeric regions showing the progress of the self-cleavage reaction of the lariat tetramer **14** into **15** with a guanosine 2',3'-cyclic phosphate and a 5'-hydroxyl cytidine termini. The spectra were recorded at 19°C in $^{2}H_2O$ (See Table 1 for the chemical shifts and Table 2 for coupling constants of **14** and **15**).

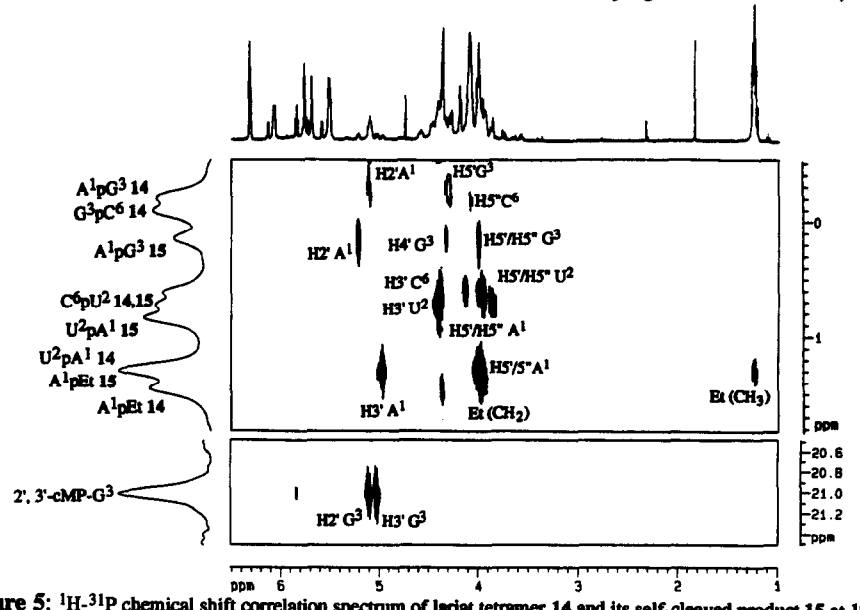


Figure 5: ^{1}H-^{31}P chemical shift correlation spectrum of lariat tetramer **14** and its self-cleaved product **15** at 19°C in $^{2}H_2O$. The 2',3'-cyclic phosphate of G^3 absorbs at 21 ppm and shows cross peaks with its H2' and H3'. The spectrum was cut along the ^{31}P axis between 1.5 and 20.5 ppm in order to show the internucleotidyl 3'→5' phosphates which all absorb between -1 and 1.5 ppm. Note that the H3'A^1, H3'U^2 and H3'G^3 crosspeaks of **14** could not be observed due to the water presaturation at 4.7 ppm.

to identify the J-coupling network for each sugar residue, whereas NOESY and ROESY were used to connect the nucleobase protons to their constituent sugar units. The ^{31}P- and ^{13}C-NMR resonances were assigned with the help of ^{1}H-^{31}P (Fig. 5) and ^{1}H-^{13}C chemical shift correlation experiments[26] at 19 °C.

(ii) Conformation of the riboses

The conformation of the sugar rings described as an equilibrium between the North and South conformation was calculated from the $^{3}J_{1'2'}$ coupling constant[27] in the temperature range 5 °C to 80 °C using the equation:

$$\%N = 100 * \frac{8.5 - J_{1'2'}}{8.2} \quad .... \quad \text{Eq. 1}$$

The limiting coupling constants of $J^{N}_{1'2'} = 0.3$ Hz and $J^{S}_{1'2'} = 8.5$ Hz for the pure N- and S-conformers were used since they have been observed in similar molecules[10] (Tables 2 and 4). Table 3 shows that at 19 °C, the A^1 and U^2 residues are in the South-type conformation, while the G^3 and C^6 residues are in the North-type conformation. The two-state N⇄ S equilibrium[40] for each nucleotide of **14** is very sensitive to a temperature increase as can be seen from the $^{3}J_{1'2'}$ and % N reported in Table 4. Thus, at 5°C, the A^1 and U^2 are in an almost pure S-type conformation, while at 80 °C, they show 60 and 40% respectively of S-type conformers. The G^3 and C^6 nucleotides exist in a pure North conformation at 5 °C, but at 80 °C, the C^6 residue shows 54% of North type sugar and the G^3 residue only 35%. Inspection of Table 4 where the temperature dependence of the N ⇄ S equilibrium[40] is reported for the cyclic tetramers **14**, **16** and **17** shows that the purity of the conformational states of the sugar residues of **14** are indeed much higher than those of **16** and **17**.

Table 2: ^{1}H-^{1}H, ^{1}H-^{31}P and ^{13}C-^{31}P coupling constants[c] for **14** and **15** at 292 K in $^{2}H_2O$.

	14				**15**			
	A^1	U^2	G^3	C^6	A^1	U^2	G^3	C^6
$J_{1'2'}$	7.7	8.2	0	1.7	3.4	6.4	3.3	
$J_{2'3'}$	5.5	3.9	4.4	5.5	5.5	4.4	6.6	
$J_{3'4'}$	4.4		8.0	7.7	6.6	2.2	4.3	
$J_{4'5'}$	Σ=5.0	Σ=3.9	1.7	a	Σ=3.3	Σ=4.4	4.0	
$J_{4'5''}$			2.1	a				
$J_{2'P2'}$	8.0	/	/	/				
$J_{2'P3'}$	2.1	1.8	/	a				
$J_{3'P3'}$	6.4	8.4	8.1	a				
$J_{5'P5'}$	2.0		2.8	a				
$J_{5''P5'}$	a		3.9	a				
$J_{C1'P2'}$	1.9	/	/	/	b	b	b	b
$J_{C3'P2'}$	a	/	/	/	b	b	b	b
$J_{C2'P3'}$	a	8.3	< 1	< 1	b	b	b	b
$J_{C4'P3'}$	1.5	Σ=8.6	7.0	a	b	b	b	b
$J_{C4'P5'}$	9.2		8.3	a	b	b	b	b

[a] Could not be measured due to spectral overlap. [b] Not measured. [c] Accuracy: ± 0.25 Hz for J_{HH} and J_{HP} and ± 0.5 Hz for J_{CP}.

Thus, while the A^1 residue in **14** is in 92% S-type conformation at 20°C, the A^1 residue in **16** where the 3'-ethylphosphate is replaced by a 3'-hydroxyl is only in 55% S conformation. This displacement of the N ⇄ S pseudorotational equilibrium toward the S geometry is due to the enhanced *gauche effect* of a 3'-OPO_3Et^-

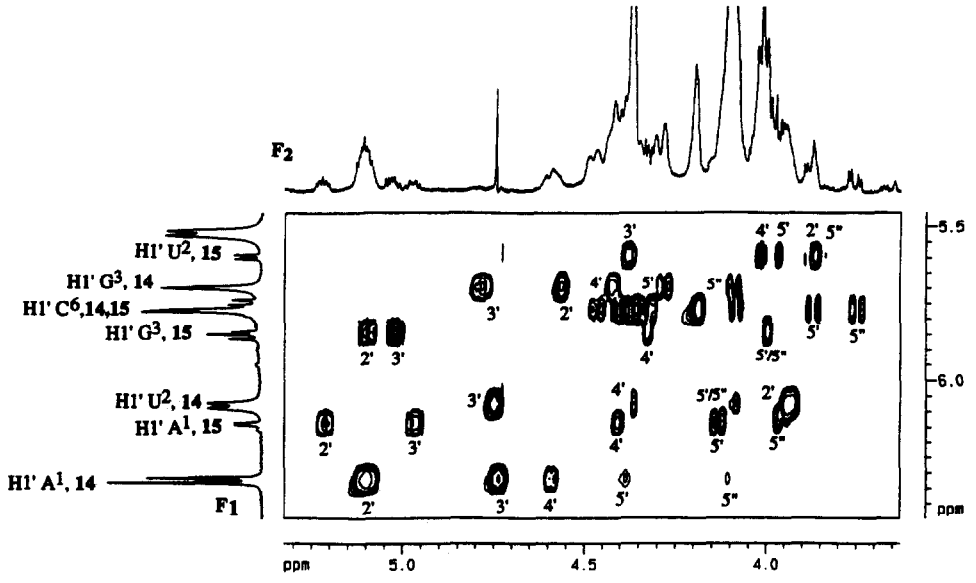


Figure 6: Clean-TOCSY spectrum of lariat tetramer **14** and its self-cleaved product **15** at 19°C in ^2H_2O. The F_1 axis represents the anomeric H1' region while the F_2 axis represents the H2' to H5'/H5" region.

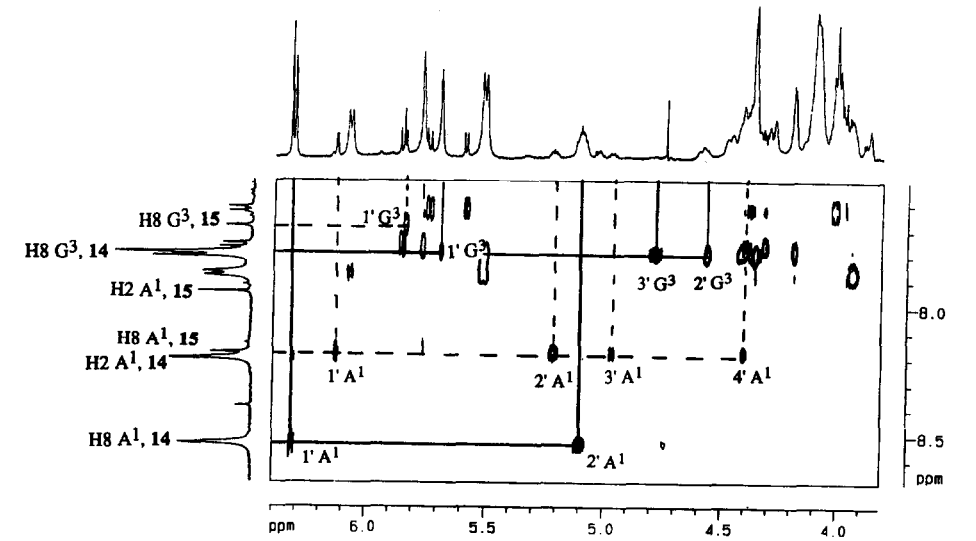


Figure 7: ROESY spectrum of lariat tetramer **14** and its self-cleaved product **15** at 19°C in ^2H_2O. Only the cross peaks for the purine nucleotides A^1 and G^3 are indicated. The H8 of A^1 in **14** shows a strong nOe with its H2' and a weak nOe with H1', while G^3 shows strong nOes with its H2' and H3'. Solid lines show the relays of the sugars of **14** while the dotted lines show those of the cleaved product **15**.

group at the branch-point A^1 residue in **14** over a 3'-OH group in **16**[28]. It can also be seen that the A^1 residue of **14** is very sensitive to an increase of temperature (96% S at 5 °C → 59% S at 80 °C), while in **16** and **17**, the A^1 residue changes only from 65% S at 5 °C to 46% S at 80 °C for **16** and from 92% S at 5 °C to 80% S at 80 °C for **17**. These data qualitatively indicate that the energy difference between the N and S sugar of A^1 in **14** is greater than the energy difference between the N and S sugar of A^1 in **16** and **17**. The populations of N and S conformers for **14**, **16** and **17** at the various temperatures given in Table 4 were used to calculate the enthalpy

Table 3: Conformational properties (Population of North, γ^+ and β^t rotamers, orientation of the glycosidic bond χ and preferred rotamer around the C3'-O3' bond (ε) for **14** at 292 K in ^2H_2O.

	A^1	U^2	G^3	C^6
%N	10	4	100	83
χ	anti	anti	anti	anti
$\%\gamma^+$	85	97	98	a
$\%\beta^t$	82	β^t	74	a
ε	ε^t (C2'-O2')	ε^-	ε^t	ε^t
	ε^- (C3'-O3')			

[a] Could not be determined

Table 4: $^3J_{1'2'}$ (Hz) and % N[b] at different temperatures for **14**, **16**, and **17** in ^2H_2O. The errors in %N are estimated to be ±3 % units and are based solely on the estimated error of ±0.25 Hz in coupling constants.

			278 K	283 K	293 K	303 K	313 K	323 K	333 K	343 K	353 K
14	A^1	$J_{1'2'}$	8.2	8.0	7.8	7.2	6.7	6.3	5.8	5.4	5.1
		% N	4	6	9	16	22	27	33	38	41
	U^2	$J_{1'2'}$	8.2	8.0	8.0	broad	broad	broad	3.8	3.6	3.6
		% N	4	6	6	-	-	-	57	60	60
	G^3	$J_{1'2'}$	0	0	0	broad	broad	broad	3.9	5.0	5.6
		% N	100	100	100	-	-	-	56	43	35
	C^6	$J_{1'2'}$	0	1.7	3.4	3.5	3.6	3.6	4.0	4.0	4.1
		% N	100	83	62	61	60	60	55	55	54
16	A^1	$J_{1'2'}$	5.6	5.4	4.8	4.8	4.5	4.4	4.3	4.2	4.1
		% N	35	38	45	45	49	50	51	52	54
	U^2	$J_{1'2'}$	7.7	7.6	7.5	7.4	7.3	7.3	7.1	7.1	6.9
		% N	10	11	12	13	15	15	17	17	20
	G^3	$J_{1'2'}$	<1	<1	1.2	1.6	1.8	2.5	3.0	3.6	3.9
		% N	>91	>91	89	84	82	73	67	60	56
	C^6	$J_{1'2'}$	<1	<1	<1	1.1	2.2	3.0	3.7	3.9	4.2
		% N	>91	>91	>91	90	77	67	59	56	52
17	A^1	$J_{1'2'}$	7.8	7.6	7.1	6.9	6.9	7.0	7.0	6.9	6.9
		% N	9	11	17	20	20	18	18	20	20
	U^2	$J_{1'2'}$	7.6	7.4	7.1	6.9	6.8	6.7	7.0	6.4	6.3
		% N	11	13	17	20	21	22	18	26	27
	G^3	$J_{1'2'}$	0	1.1	1.1	2.7	3.3	a	4.9	4.9	a
		% N	100	90	90	71	63	-	44	44	-
	C^6	$J_{1'2'}$	0	2.1	3.4	3.8	4.9	a	5.5	5.9	a
		% N	100	78	62	57	44	-	37	32	-

[a] Not measured due to spectral overlap, [b] %N = 100 * (8.5 - $J_{1'2'}$) / 8.2

($\Delta H°$) and the entropy ($\Delta S°$) of the two-state N⇄ S pseudorotational equilibria[40] through van't Hoff plots of $\ln \frac{X_S}{X_N}$ *vs.* $\frac{1}{T}$. The slopes of the straight lines gave the $\Delta H°$ values and the intercepts gave the $\Delta S°$ (Table 5, Figs. 8 and 9). For computational reasons, the mole fractions were set to 1% or 99% when they were calculated to be 0% or 100% by Eq. 1. Table 5 shows that the enthalpy and entropy of the N to S transition for A^1, U^2, G^3 and C^6 of **16** and **17** are similar whereas it is very different for **14** (Figs. 8 & 9). The comparison of $\Delta H°$ and $\Delta S°$ between **14** and **16** clearly shows (Figs. 8 and 9) the conformational effect of the 3'-ethylphosphate group at the branch-point adenosine in **14** has been indeed transmitted to the entire molecule. Thus the S conformation in A^1, U^2 and C^6 sugar moieties is clearly thermodynamically more stabilized while it is considerably destabilized in G^3. Clearly, the thermodynamic stabilization by $\Delta H \approx 4$ kcal.mol^{-1} of the S conformation of the branch-point sugar moiety by the 3'-phosphate group in **14** compared to 3'-OH group in **16** is due to the gauche effect[28] of the ethylphosphate moiety which is believed to be responsible for the drive of its self-cleavage reaction. Earlier, a comparison of thermodynamics of N⇄S equilibrium in 2'-deoxyadenosine and 3'-dAMP showed[28] a less pronounced 3'-phosphate gauche effect promoted stabilization of the S sugar conformation (only by $\Delta H \approx 0.3$ kcal.mol^{-1}) compared to 3'-OH. It is interesting to note that the magnitude of enthalpy and entropy for the North to South transition of the A^1 sugar in **14** is similar to the enthalpy and entropy of transition between the A- and B-form of the lariat hexamer[12] **18**. The similarity of **16** and **17** and the observed effect of 3'-ethylphosphate in **14** also suggest that the substitution of 2'-OH in **17** with a 2'-ethylphosphate group should bring about a conformational change that is comparable to that of **14**. Hence, it is likely that if the 2'-OH group of **17** is transformed to a 2'-phosphate, it should also undergo self-cleavage reaction.

Table 5: The $\Delta H°$ and $\Delta S°$ values (90% confidence) for the N ⇄ S equilibrium[40] in **14**, **16** &**17**.

Compound	Residue	$\Delta H°$ (kcal.mol^{-1})	$\Delta S°$ (cal.K^{-1}.mol^{-1})	$-T\Delta S°$[a] (kcal.mol^{-1})	$\Delta G°$ (kcal.mol^{-1})	Temp. range (K)	χ^2
14	A^1	-5.8 ±1.3	-15.9 ±3.8	4.7 ±1.1	-1.1 ±2.4	278 - 353	2.1
	U^2	-7.1 ±2.2	-21.3 ±6.4	6.3 ±1.9	-0.8 ±4.1	278 - 353	12.6
	G^3	12.3 ±2.8	36.2 ±8.2	-10.6 ±2.4	1.7 ±5.2	278 - 353	3.4
	C^6	1.2 ±0.8	3.2 ±2.5	-0.9 ±0.7	0.3 ±1.6	293 - 353	0.8
16	A^1	-1.8 ±0.5	-5.5 ±1.7	1.6 ±0.5	-0.2 ±1.1	278 - 353	2.9
	U^2	-1.9 ±1.2	-2.6 ±3.8	0.8 ±1.1	-1.2 ±2.3	278 - 353	0.2
	G^3	6.1 ±1.1	17.0 ±3.2	-5.0 ±0.9	1.2 ±2.0	278 - 353	1.5
	C^6	6.6 ±1.1	18.7 ±3.4	-5.5 ±1.0	1.1 ±2.1	278 - 353	7.5
17	A^1	-1.2 ±1.1	-0.7 ±3.6	0.2 ±1.0	-1.0 ±2.2	278 - 353	4.5
	U^2	-2.2 ±1.0	-4.3 ±3.0	1.3 ±0.9	-1.0 ±1.9	278 - 353	3.5
	G^3	7.3 ±1.2	22.1 ±3.9	-6.5 ±1.1	0.8 ±2.4	283 - 343	8.8
	C^6	5.0 ±1.1	16.2 ±3.4	-4.8 ±1.0	0.3 ±2.1	293 - 343	2.4

[a] The $-T\Delta S$ term is computed for 293 K.

(iii) Conformation about the glycosidic bond

The conformation about the glycosidic bond was determined from 2D NOESY and ROESY experiment at 19 °C with a mixing time of 600 ms. The H6 of the U^2 nucleotide shows a very strong nOe with its H2' indicating an *anti* orientation of the glycosidic bond and also confirming the S-type conformation of the sugar ring determined from the $J_{1'2'}$ coupling. The H6 of C^6 has a strong nOe with its H3' indicating also the *anti*

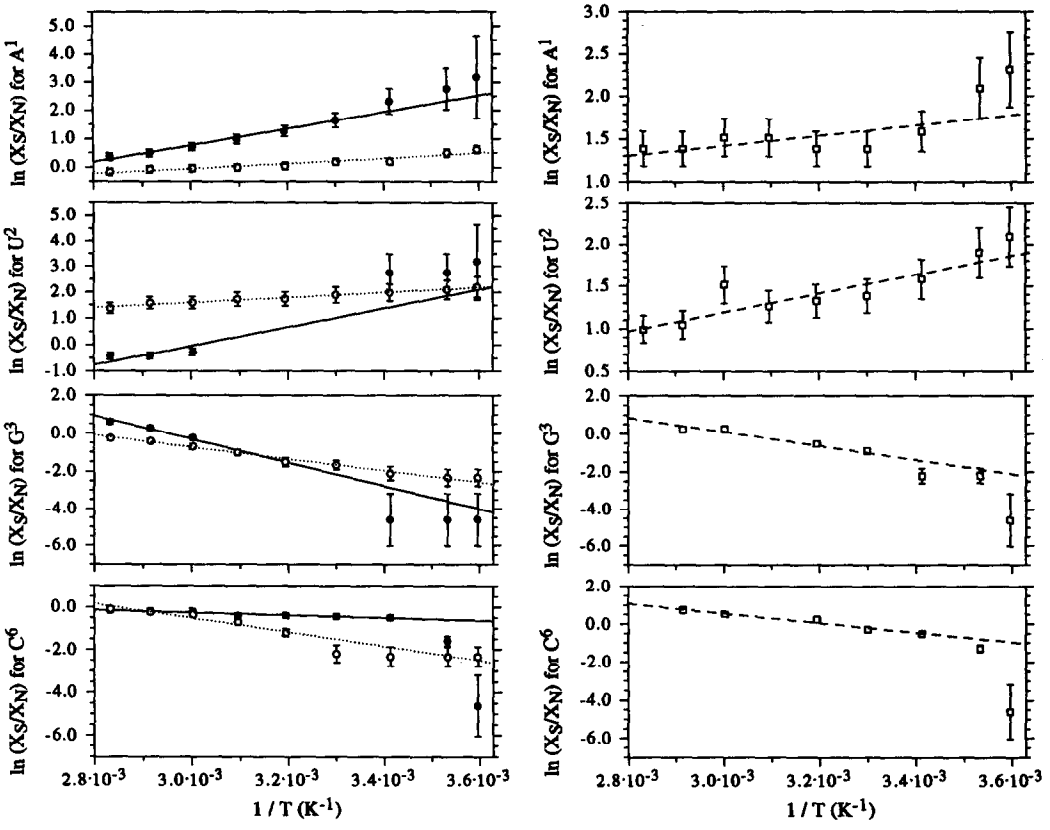


Figure 8: van't Hoff plots of $\ln\frac{X_S}{X_N}$ as a function of $\frac{1}{T}$ for each of the residues in **14** and **16** (left side panels) and **17** (right side panels). The mole fractions of N and S conformers were calculated using equation 1. ΔH° and ΔS° for the N $\rightleftarrows$ S equilibrium were calculated from the slope and from the intercept according to the relation: $\ln\frac{X_S}{X_N} = \frac{-\Delta H^\circ}{R} * \frac{1}{T} + \frac{\Delta S^\circ}{R}$. The absolute errors ($\varepsilon_{(\ln X_S/X_N)}$) in $\ln\frac{X_S}{X_N}$ were based only on the estimated error in $J_{1'2'}$ of ±0.25 Hz, which corresponds to an error (ε_{mf}) of ±0.0305 in mole fraction. $\varepsilon_{(\ln X_S/X_N)}$ was set to the larger of the two values computed from $\left|\ln\frac{X_S \pm \varepsilon_{mf}}{X_N \pm \varepsilon_{mf}} - \ln\frac{X_S}{X_N}\right|$. Data points from **14** are shown as filled circles (left set of graphs), data points from **16** are shown as open circles (left set of graphs), and data points from **17** are shown as open squares (right set of graphs). The chi-square fitted straight lines are solid for **14**, dotted for **16**, and dashed for **17**. The results are shown in Table 5. The inverted temperature scale (X-axis) is the same for all eight graphs.

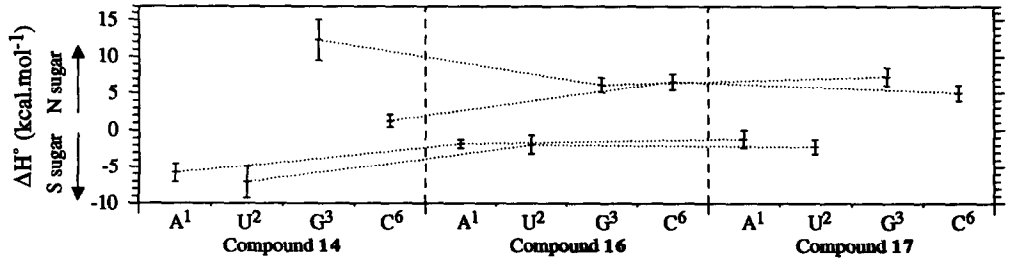


Figure 9: A comparison (dotted lines connecting identical sugar residues) of the calculated ΔH° for the N $\rightleftarrows$ S equilibrium for each of the sugar residues in **14**, **16**, and **17**. The error bars are for a 90% confidence level (see Table 5 and Fig. 8).

conformation and also supporting the North conformation determined from the $J_{1'2'}$ coupling. The H8 of A^1 shows a strong nOe cross peak with its H2' and smaller cross peak with its H1' and H3' which suggest a preference for the *anti* geometry. The H8 of G^3 has a strong nOe with its H3' and weaker nOes with H2' and H1' indicating a preference for the *anti* conformation.

(iv) Conformation of the phosphate backbone

The coupling constants J_{HH}, J_{HP} and J_{CP} necessary to define the conformation of the backbone torsional angles were measured at 19°C from 2D DQF-COSY spectra with and without phosphorus decoupling (J_{HH}, J_{HP}) and from a ^1H-^{31}P HSQC[29] chemical shift correlation experiment (J_{CP}, J_{HP}). The population of *gauche* $^+$ (γ^+) rotamer around the C4'-C5' bond has been estimated from the values of the $J_{4'5'}$ and $J_{4'5''}$ coupling constants obtained from 2D DQF-COSY spectra, using the formula[30]:

$$\%\gamma^+ = 100 * \frac{13.3 - J_{4'5'} - J_{4'5''}}{9.7} \quad .... \quad \text{Eq. 2}$$

and the population of *trans* conformer about the C5'-O5' bond (β^t) was calculated using the formula[31]:

$$\%\beta^t = 100 * \frac{25.5 - J_{H5'P} - J_{H5''P}}{20.5} \quad .... \quad \text{Eq. 3}$$

The $J_{H5'P}$ and $J_{H5''P}$ coupling constants were obtained from the subtraction of DQF-COSY cross peaks with and without phosphorus decoupling. From Table 3 it can be seen that the A^1, U^2 and G^3 nucleotides preferentially adopt the γ^+ and β^t conformation (> 70%) which is the conformation generally observed in oligonucleotides. Due to overlap of ^1H resonances, it was not possible to extract the $J_{4'5'}$, $J_{4'5''}$, $J_{5'P}$ and $J_{5''P}$ coupling constants for C^6 and hence its γ^+ and β^t populations were not calculated. The conformation around the C3'-O3' bond (ε) has been monitored by the $J_{C2'P3'}$ and $J_{C4'P3'}$ coupling constants which were measured from a ^1H-^{31}P HSQC chemical shift correlation experiment. A large $J_{C2'P3'}$ and small $J_{C4'P3'}$ indicates a high population of ε^- conformer while the reverse indicates a high population of ε^t conformer[31,32]. Thus, the data in Table 2 shows that the C3'-O3' bond of G^3 is strongly populated in ε^t conformation. It was not possible for the three other nucleotides to measure both the $J_{C2'P3'}$ and $J_{C4'P3'}$ coupling constants. For the A^1 residue, the $J_{C2'P3'}$ could not be measured. However, the small value of the $J_{C4'P3'}$ together with the observation of a long range $^4J_{H2'P3'}$ coupling of 2.1 Hz suggest that the A^1 residue adopts the ε^- geometry around the C3'-O3' bond. The long range $^4J_{H2'P3'}$ is only possible for a planar W-pathway created when the ribose is in the S-type conformation and the C3'-O3' bond in the ε^- conformation[32]. The $J_{C4'P3'}$ of the U^2 nucleotide could not be obtained alone. Instead, the sum $J_{C4'P3'} + J_{C4'P5'}$ of 8.6 Hz was measured. Since the U^2 has been found to be in the β^t conformation, this implies that the $J_{C4'P5'}$ coupling constants is large (ca. 8 Hz), and therefore the $J_{C4'P3'}$ must be very small. This together with the facts that the $J_{C2'P3'}$ is very large and that a $^4J_{H2'P3'}$ coupling is present suggest that the U^2 residue prefers the ε^- conformation. For the C^6 residue, only the $J_{C2'P3'}$ could be measured. The small value of < 1 Hz suggests a preference for the ε^t orientation around the C3'-O3' bond.

(E) Molecular dynamics simulation in water for 226 ps

Initial coordinates for the molecular dynamics simulation of **14** were taken from a 226 ps long MD-trajectory of our previously studied[12] lariat-RNA hexamer **18** (A-form). Coordinates of the hexameric

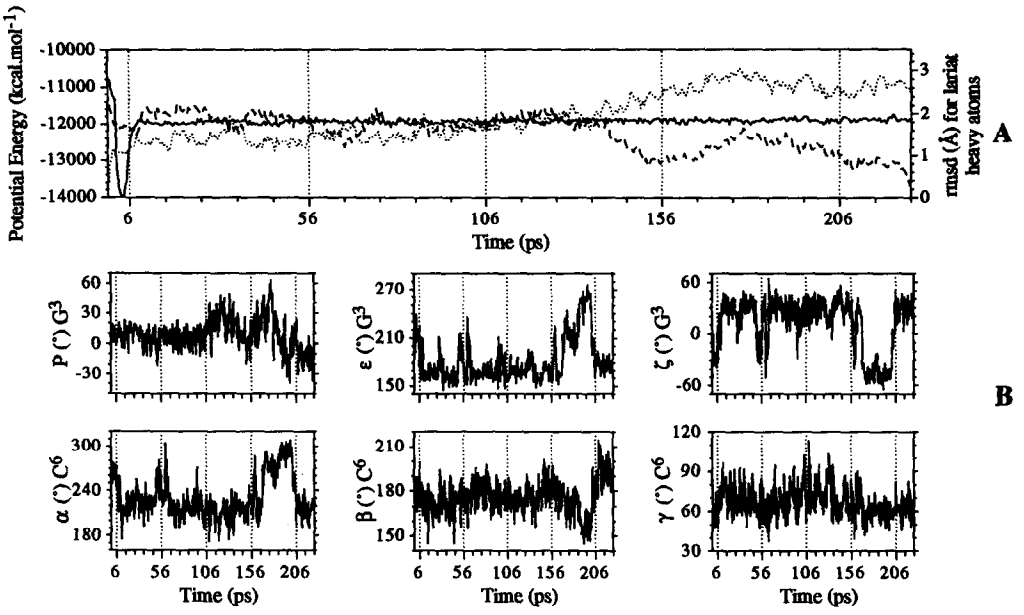


Figure 10: Panel A shows the evolution of potential energies over the 226 ps MD simulations in water (left scale, solid lines) for **14**. In the same graph is also shown the rmsd in Å (right scale, non-solid lines) of the heavy atoms of the lariat-RNA structure (water-oxygens and sodium ions were not included in rmsd calculations). The dotted line correspond to the rmsd relative to the starting structure obtained from the *in vacuo* minimization of the initial coordinates, the broken line correspond to the rmsd relative to the structures obtained at 226 ps. Panel B shows the variation of the phase angle (P), ε, ζ, α, β, and γ torsions of the G^3 and C^6 residues at the self-cleavage site during the entire MD simulation in water at 294 K.

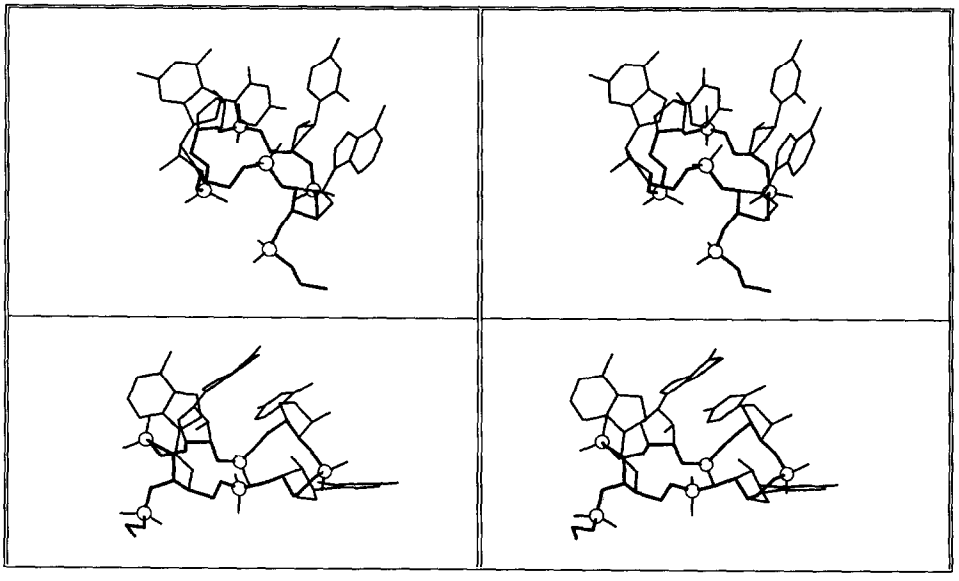


Figure 11: Stereoviews of the lariat RNA **14** from the snapshots at 96 ps (the last fully NMR constrained structure) and at 226 ps (the last unconstrained structure) from the MD simulation in water at 294 K. All water molecules, sodium ions, and hydrogen atoms have been omitted for clarity. All phosphorus atoms are shown as open circles and the sugar-phosphate backbone is shown with thick lines. Panel A: Tetramer **14** at 96 ps. Panel B: Tetramer **14** at 226 ps. See Figs. 10 and 13 for variation of sugar-phosphate backbone for structures for the entire trajectory.

lariat-RNA structure at 96 ps which was the last structure with full NMR restraints were chosen. The coordinates of the loop nucleotides, A^1, U^2, G^3, and C^6, were used directly and the coordinates for O5', C5', and C4' of U^4 of **18** were used for the ethyl residue in **14**. This structure was energy minimized *in vacuo*[33] and subsequently immersed in a periodic box of water with sodium cations[34]. The solvated system was then subjected to 226 ps of MD simulation. The MD simulation was first carried out with harmonic constraints which were derived from NMR-data (0-96 ps) and then completely without constraints (106-226 ps). Static torsional constraints corresponding to the major conformer as observed by NMR were used during the MD simulations. Constraints were only imposed on torsions where NMR data showed that the preference for one rotamer exceeded ~70%. Restraints on the ε torsion of C^6 and ε' (C2'-O2'-P-O5') torsion of A^1 were not imposed since only one $^3J_{CP}$ was available. In addition, the constraining harmonic potential wells have a flat bottom (Table 6) and the width of the flat bottom would for most torsions allow large torsional movements within the NMR-obeyed conformational hyperspace.

Table 6: The constrained torsions with the target dihedral angle and the allowed variation in degrees, and the force constant in kcal.rad^{-2}. The harmonic potential for the constraints have a flat bottom, its width given by the allowed variation, and the force constants for the regions outside the allowed variation is the same for both lower side and higher side

	G^3	C^6	U^2	A^1	Force constant
ν_0	9.6 ±8.0[a]	9.6 ±8.0[a]	-22.8 ±8.0[b]	-22.8 ±8.0[b]	150.0
ν_2	37.0 ±8.0[a]	37.0 ±8.0[a]	-34.8 ±8.0[b]	-34.8 ±8.0[b]	150.0
β	180.0 ±60	-	180.0 ±60	180.0 ±60	150.0
γ	60.0 ±40	-	60.0 ±40	60.0 ±40	150.0
ε	180.0 ±60	-	300.0 ±60	300.0 ±60	150.0
χ	180.0 ±90	180.0 ±90	180.0 ±90	180.0 ±90	50.0

[a] The center values of the ν_0 and ν_2 torsional constraints correspond to P = 3° and Ψ = 37°. With the allowed torsional variation the following ranges are obtained: 16°>P>-17° and 47° >Ψ>29°. [b] The center values of the ν_0 and ν_2 torsional constraints correspond to P = 160° and Ψ = 37°. With the allowed torsional variation the following ranges are obtained: 178°>P>138° and 47° >Ψ >28°.

Table 7: The atom types and partial atomic charges for the ethyl residue of **14**.

Atom name	AMBER atom type (all atom force field, parm91)	Charges used in AMBER	Charges from STO-3G
O5'	OS	-0.510	-0.510997
C5'	CT	0.142	0.141647
H5'1 / H5'2	HC	0.007	0.006984[a]
C4'	CT	-0.228	-0.228583
H4'1 / H4'2 / H4'3	HC	0.055	0.055479[a]

[a] average value of all methyl/methylene hydrogen atoms.

The AMBER[35] force field does not include all parameters for the ethylphosphate moiety, the "3'-tail" of **14**. The original AMBER phosphate residue (POM) was used to link the branchpoint adenosine, via its 3'-oxygen, with the ethyl residue. The original AMBER 'OS', 'CT', and 'HC' atom types were used for the atoms of the

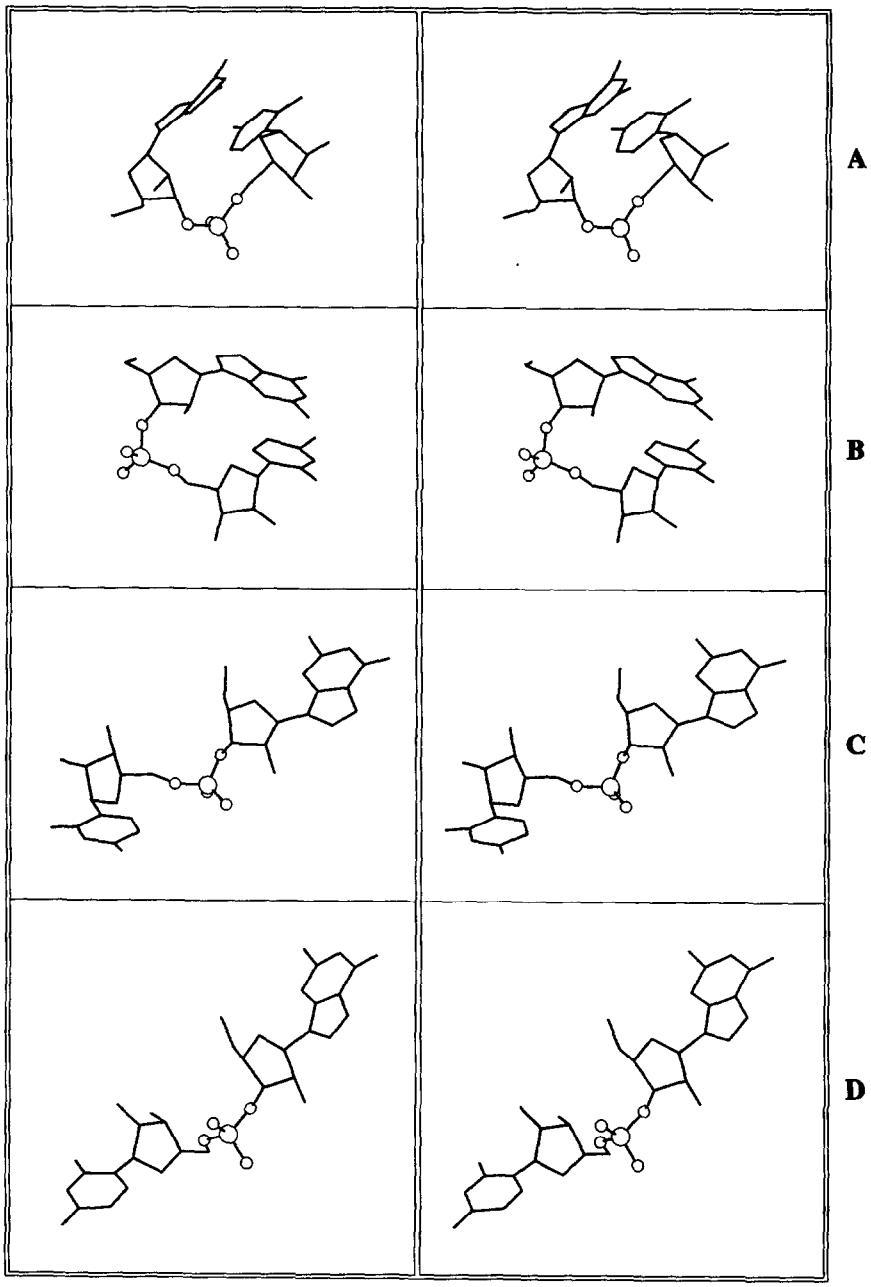


Figure 12: Stereoviews showing the conformation at the self-cleavage sites of the lariat tetramer **14**, hexamer **18** (A- and B-form), and heptamer **19** (G^3 and C^6/U^7) from the snapshots at 226 ps. All water molecules, sodium ions, and hydrogen atoms have been omitted for clarity. The O3'-PO_2-O5' fragments are shown as ball and stick models while the remainder is shown as wireframe. Structures of the lariat hexamer and lariat heptamer are taken from ref. 12. Panel A: Tetramer **14** (P of G^3 = -12°, ε = 168°, ζ = 33°, α = -137°, β = -169°, γ = 51° and P of C^6 = 18°). Panel B: Hexamer A-form (P of G^3 = 44°, ε = -158°, ζ = -58°, α = -80°, β = 171°, γ = 68° and P of C^6 = 53°). Panel C: Hexamer B-form (P of G^3 = 177°, ε = -151°, ζ = 78°, α = 85°, β = 178°, γ = 67° and P of C^6 = 37°). Panel D: Heptamer (P of G^3 = 123°, ε = -96°, ζ = 146°, α = 64°, β = -176°, γ = 54° and P of U^7 = 150°).

ethyl residue. No modifications of the force field parameters for these atom types were made. The partial charges for the atoms were calculated using the Gaussian-92[36] program: (i) The geometry of dimethylphosphate was optimized at the HF/3-21G level, the two C-O-P-O torsions were fixed at -60° however. The initial bond lengths and bond angles were taken from the AMBER force field's equilibrium values. (ii) The optimized dimethylphosphate was converted to ethyl methylphosphate by replacing the methyl proton being *trans* to the phosphorus atom with a methyl group. (iii) The geometry of the ethyl methylphosphate was optimized at the HF/3-21G level in two steps. In the first step, the two C-O-P-O torsions were fixed at -60°, in the second step all internal coordinates were optimized. (iv) The geometry of the fully optimized ethyl methylphosphate was then used for the calculation of the atom centered atomic partial charges. The charge calculation was done using the Merz-Kollman method included in the Gaussian-92 program. We chose to perform the charge calculations at the STO-3G level to obtain charges consistent with the charges of the original AMBER residues[37]. The atomic radii used in the atomic charge calculation were taken from the AMBER[35] all atom force field. The final atomic charges are presented in Table 7.

The NMR-data (Tables 2-4) indicated that the conformational preferences of the sugar rings within the lariat-loop were high (>70% preference for one conformer) and generally also for the χ, β, γ, and ε torsions. Evaluations of ^{1}H-^{1}H, ^{1}H-^{31}P and ^{13}C-^{31}P coupling constants related to β, γ, and ε torsions of C^6 were not conclusive and, hence these torsions were not restrained.

Noticeable changes in conformation of the lariat RNA structure occurred within ~10 picoseconds of MD simulation in water (Fig. 10). These changes are most probably due to the fact that the initial coordinates of **14** come from the lariat hexamer **18** containing an additional two nucleotides in the "3'-tail", and the interactions of the "3'-tail" through hydrogen bonding and stacking with the loop nucleotides is naturally not present in **14**. The conformation did not, however, change dramatically when the torsional constraints were released after the first ~100 ps of MD simulations. Local conformational changes took place; the α torsion of G^3 changed from *gauche*$^+$ to *trans* and *gauche*$^-$, accompanied by changes of the β, γ, and ζ torsions of G^3. The overall global conformation of the entire molecule remained essentially the same however. A similar conclusion can also be reached from a comparison of NMR-constrained (see Table 6) and unconstrained conformers in Fig. 11. The similarities between the constrained and the unconstrained part of the MD simulation (Fig. 13), in terms of preferred regions of torsional space, also indicate that this lariat-RNA does not easily undergo large conformational changes, at least not in the time scale used for the present MD simulations.

Similarities can also be seen among the torsions of the tetramer **14** and the A-form of the hexamer[12] **18** as well as between the B-form of the hexamer and the heptamer **19** (Figs. 12 and 13). Clearly, the relatively small size of the loops restrict its conformational freedom because of its constrained nature, and therefore, these lariat-RNA structures populate to a large extent the same regions of torsional space.

The more important structural features observed during the 226 ps MD simulation in water of **14** seem to be stacking interactions between G^3 and C^6, this was also observed for a lariat hexamer (A-form), and hydrogen bonding between one of the 4-amino protons of C^6 and one of the oxygens of the phosphate linking A^1 with G^3. The base stacking as well as the hydrogen bond are seen during large parts (0-~140 ps) of the MD simulation in water.

The lack of any major changes in the conformation of **14** upon releasing the NMR constraints indicate that the conformers generated in the MD simulation in water agree well with the NMR-derived structural features (Tables 3 and 6). The ensemble of conformers generated during the MD trajectory of 226 picoseconds

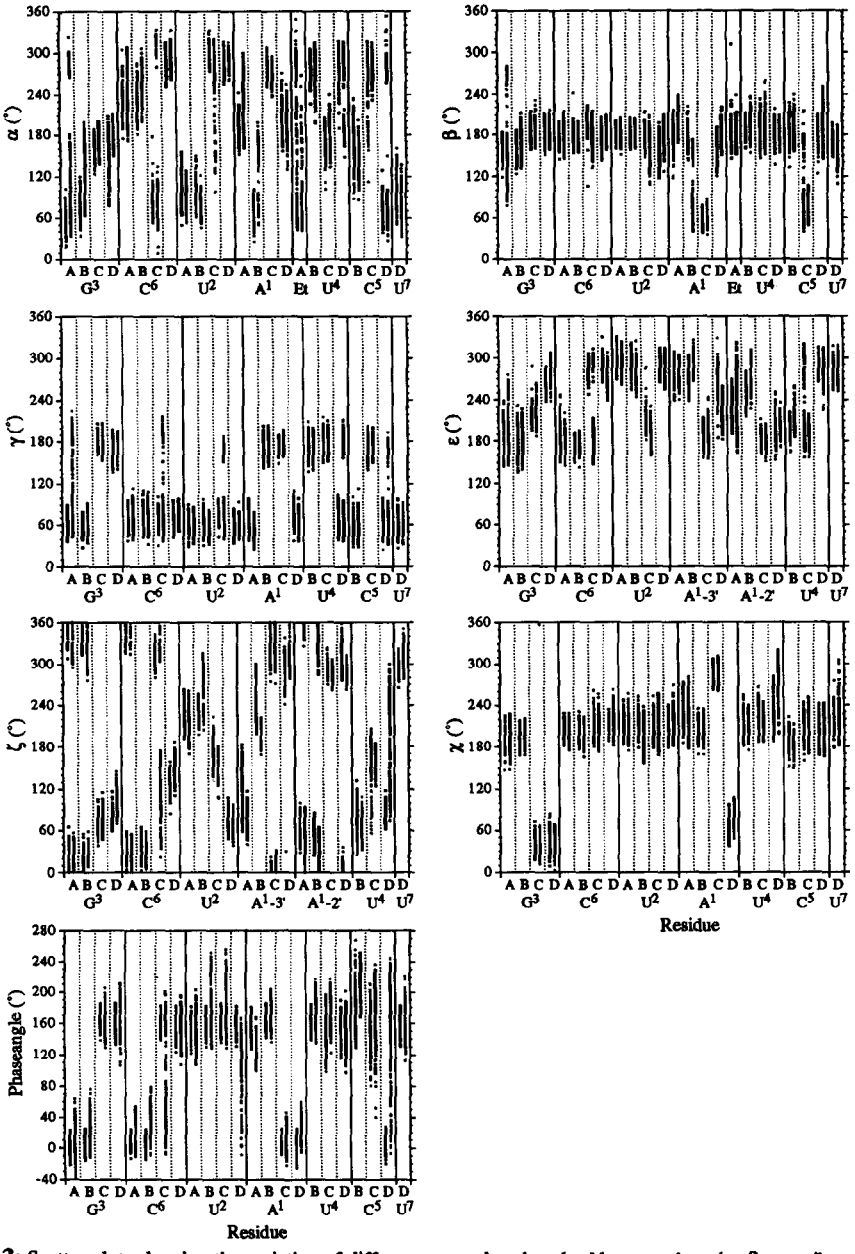


Figure 13: Scatter plots showing the variation of different sugar-phosphate backbone torsions (α, β, γ, ε, ζ, and χ torsions and phase angles (P)) for structures sampled every 0.5 ps from the entire 226 ps MD simulations in water in both NMR constrained (Table 6) and unconstrained modes. The rectangular boxes subtitled (A) represent variations of torsions for tetramer **14**; left lane in (A) represents NMR constrained part (0-96 ps) of the MD and the right lane in (A) represents the unconstrained part (96-226 ps) of the MD. The rectangular boxes subtitled (B) represent variations of torsions for the A-form of hexamer **18**: left lane in (B) represents NMR constrained part (0-96 ps) of the MD and the right lane in (B) represents the unconstrained part (96-226 ps) of the MD. The rectangular boxes subtitled (C) represent variations of torsions for the B-form of hexamer **18**: left lane in (C) represents NMR constrained part (0-96 ps) of the MD and the right lane in (C) represents the unconstrained part (96-226 ps) of the MD. The rectangular boxes subtitled (D) represent variations of torsions for the heptamer **19**: left lane in (D) represents NMR constrained part (0-96 ps) of the MD and the right lane in (D) represents the unconstrained part (96-226 ps) of the MD. Data for the lariat hexamer and lariat heptamer are taken from ref. 12.

are not artificially held in these conformations due to the NMR constraints, suggesting that these conformers (Fig. 11 for stereoviews at 96 and 226 picoseconds) can be considered as good representatives of the actual solution structures. The low number of constraints due to the limited number of experimental observables do not allow a more precise structure determination. Other low energy structures may also be in agreement with the present NMR data. These considerations suggest that our present protocol allows us to compute and visualize a representative subset of conformers for the lariat-RNA.

The conformation of the sugar-phosphate backbone at the site of self-cleavage between the G^3 and the C^6 residues is of particular interest. The 3'-phosphate of G^3 is attacked by the vicinal 2'-OH to give a trigonal bipyramidal phosphate as a transition state/intermediate which upon displacement of the 5'-oxygen (C^6) gives a 2',3'-cyclic phosphate (G^3) and a 5'-hydroxy terminus (C^6). In an earlier study of similar lariat structures[12], it was seen that the backbone torsions of the lariat hexamer **18** (B-form) and the lariat heptamer **19** were similar suggesting that the B-form of the lariat-hexamer and the lariat heptamer undergo similar local conformational transitions in order to reach the transition state/intermediate of the cleavage reaction. Lariat **14** on the other hand, shows similarities to the previously studied A-form of the lariat hexamer[11,12]. It is only the α and γ torsions of the A^1 residue which are significantly different between **14** (α^t / γ^+) and the A-form of **18** (α^+ / γ^t). Since these torsions are involved in a 'crankshaft' motion, the overall geometry is only little affected by these differences.

The average distances between the O2' of the G^3 residue and the phosphorus atom of the vicinal 3'-phosphate is 3.4 Å for **14** while it is 3.5, 3.3, and 4.0 Å for hexamer A-, and B-forms, and heptamer respectively[12] (Fig. 12 for the stereoviews of the cleavage-site). Geometrical considerations require that the ε-torsion of the G^3 residue is close to 120° for the O2'-P bond to form in the transesterification step and the ζ-torsion should be close to 180° to position the O5' of the C^6 residue for its in-line displacement by the incoming O2' nucleophile. The torsions in **14** would need to change from ε^t to $\varepsilon \approx 120°$ and $\zeta^{-/+}$ to $\zeta \approx 180°$ to obtain the local geometry necessary for self-cleavage. Initial studies[38] made in this laboratory on guanosine 2',3'-cyclic oxyphosphorane using *ab initio* methods (Gaussian-92, HF/3-21G basis set) show the following geometry for the trigonal bipyramidal intermediate: $\varepsilon = +110°$, $\zeta = -178°$, $\alpha = 66°$, $d_{O2'\text{-}3'P} = 1.8$ Å. Other studies[39] using *ab initio* calculations on cyclic oxyphosphoranes as models of the intermediates during cleavage of RNA, have shown that the in the ground state the α-torsion favours the *gauche*$^+$ rotamer.

Conclusion

A comparison of the conformational properties of the lariat tetramer **14** with the lariat hexamer **18**[12], lariat heptamer **19**[12] and cyclic tetramers **16** and **17** shows that the conformation of **14** is very similar to the conformation of the A-form of the lariat hexamer **18**. All nucleotides show pure ribose conformation (ca. 100% N or S) and have the glycosidic bond in the *anti* conformation. The cyclic tetramer **16** on the other hand exhibits two main conformational differences when compared to the lariat tetramer **14**. In **16**, the conformational purity of the A^1 riboses is not as high (45% N at 20 °C) as in **14** (8% N at 20 °C) and the G^3 residue is in the *syn* conformation while it is in the *anti* conformation in **14**. A comparison of the rate of self-cleavage of the lariat systems and cyclic tetramers shows that (i) The lariat heptamer which has a loop of five nucleotides cleaves six times faster than the lariat hexamer which has a loop of four nucleotides. On the other hand, a lariat pentamer with only three nucleotides in the loop does not undergo self-cleavage[10]. These

results indicate that the number of nucleotides in the lariat-loop is important and suggest therebye that a certain conformational flexibility is necessary for the self-cleavage reaction to take place. (ii) The lariat hexamer **18** which has a 3'-tail of two nucleotides cleaves only 1.7 times faster than the lariat tetramer **14** where the tail of nucleotides at the branch-point A^1 is substituted by a 3'-ethylphosphate residue ($t_{1/2}$ = 32 days at 19 °C). The cyclic tetramer **16** with a 3'-hydroxyl group was found to be completely stable for more than 3 months at 19 °C. These data suggest that the presence of nucleotides at the 3'-tail at the branch-point is not an absolute requirement for the self-cleavage reaction. Our observation that the lariat tetramer **14** with a 3'-ethylphosphate residue at the branch-point self-cleaves while the cyclic tetramer with a 3'-hydroxyl **16** does not is interesting. A comparative study of the temperature dependence of the N ⇄ S equilibrium for the lariat-tetramer **14** and the 2'→5'-linked cyclic tetramer **16** shows that the A^1 residue in **14** is in 92% S-type conformation at 20 °C, wheras in **16** with a 3'-hydroxyl group it is only in 55% S. This displacement of the N ⇄ S pseudorotational equilibrium toward the S geometry is due to the enhanced *gauche effect* of a 3'-OPO_3Et^- group in **14** compared to 3'-OH group in **16**. Clearly, this 3'-OPO_3Et^- group promoted stabilization of the S geometry at the branch-point by $\Delta H \approx 4$ kcal.mol^{-1} is the *conformational driving force* promoting the self-cleavage reaction. The comparison of $\Delta H°$ and $\Delta S°$ of the N⇄ S pseudorotational equilibria between **14** and **16** (Table 5) also shows the remarkable effect of the conformational transmission of the 3'-ethylphosphate group at the branch-point adenosine in **14** on the entire molecule. The comparable relative rates of self-cleavage of tetrameric lariat-RNA **14** ($k = 0.15 \times 10^{-4}$ min^{-1}) with that of hexameric lariat-RNA **18** ($k = 0.25 \times 10^{-4}$ min^{-1}) suggest that the bulky and polar 3'-ethylphosphate group mimics the phosphodiester function of a nucleotide residue stabilizing a certain set of the local sugar conformers and thereby changing the sugar-phosphate backbone globally. What is therefore clear is that the 3'-phosphodiester function at the branch-point is certainly more important structural feature than the the number or types of nucleotide bases at the the 3'-tail. It is possible that the cooperativity of various local sugar conformers and their transmission through the phosphodiester linkage stabilize or facilitate the formation of the intermediate or transition state which is required[11,12] for the self-cleavage reaction. Synthetic and conformational work is now in progress to investigate the importance of having a guanosine nucleotide at the cleavage site. Attempts are also being made to design and synthesize oligomeric molecules that will specifically bind to the target pathogen- or tumor-specific RNA and induce geometrical changes to its specific site(s) similar to the cleavage site geometry of the self-cleaving lariat-RNAs, causing the target RNA to self-cleave.

Experimental

(A) Synthesis.

^{1}H-NMR spectra were recorded in δ scale with Jeol FX 90 Q, Jeol JNM-GX 270 and Bruker AMX-500 spectrometers at 90, 270 and 500 MHz respectively, using TMS or H_2O (set at 4.7 ppm) as internal standards. ^{31}P-NMR spectra were recorded at 36 and 202 MHz in the same solvent using 85% phosphoric acid as external standard. Thin layer chromatography was carried out using pre-coated Merck silica gel F_{254} TLC preparative TLC or high performance (HPLC) plates in the following CH_2Cl_2-MeOH mixtures: (A) 95:5 (v/v), (B) 90:10 (v/v), (C) 85:15 (v/v). 4,4',4"-Tris(4,5-dichlorophthalimido)trityl bromide (CPTr-Br) was prepared as reported in the literature[16,17]. Dry pyridine was obtained by distillation from calcium hydride (CaH_2) and tosylchloride (TsCl). Acetonitrile (MeCN) was distilled from phosphorus pentoxide (P_2O_5) under argon. Dimethylformamide (DMF) was distilled from CaH_2. Triethylamine (Et_3N) and 1,8-diazabicyclo[5.4.0]undec-7-ene (DBU) were distilled from CaH_2 under argon. Column chromatography of all the protected

intermediates was carried out using Merck G 60 silica gel. DEAE-Sephadex A-25 from Pharmacia was used for the ion exchange chromatography. An LDC equipment with ConstaMetric Pump model III and Gradient Master was used for analytical HPLC chromatography. A Gilson equipment with Pump Model 303, Manometric Module Model 802C and Dynamic Mixer 811B connected to a Dynamax computer program for gradient control was used for semi-preparative reverse phase HPLC separations. 2'-*O*-Thp derivatives of all nucleosides used in this work were separated and used in a diastereomerically pure form which are designated as "Low R_f" or "High R_f" subsequently in the following experimental section. All reactions were carried out at room temperature, unless otherwise specified. Analytical HPLC and high pressure semi-preparative Spherisorb S50DS2 column chromatography were carried out using gradients of solution B (50% MeCN in 0.1 M triethylammonium acetate (TEAA)) and solution A (5% MeCN in 0.1 M TEAA).

5'-*O*-(CPTr)-2'-*O*-pixyl-6-*N*-(4-anisoyl)-adenosine (1) 6-*N*-anisoyl-2'-*O*-pixyl-adenosine (2 g, 3.04 mmol) was rendered anhydrous by co-evaporation with dry DMF (30 ml). To the DMF solution was added successively, 2,6-lutidine (636 µl, 5.47 mmol), CPTr-Br (5.81 g, 6.08 mmol) and $AgNO_3$ (1.03 g, 6.08 mmol). After being stirred vigorously for 4 h the mixture was diluted with CH_2Cl_2 and extracted with H_2O. Each aqueous layer was back extracted with CH_2Cl_2. The two CH_2Cl_2 extracts, free from DMF, were combined and filtered over anhydrous Na_2SO_4 before evaporation. The residue was co-evaporated with toluene and dissolved in benzene. After the mixture stood for 15 min, precipitated CPTr-OH was filtered and washed with benzene. The filtrate was evaporated till dryness and submitted to column chromatography (30-70% hexane / 1% pyridine / CH_2Cl_2) which afforded **1** (3.02 g, 65%), R_f: 0.75 (A). ^{1}H-NMR (90 MHz, $CDCl_3$ + 2,6-lutidine): 8.89 (br, 1H) NH; 8.37 (s, 1H) H8; 8.03-6.83 (m, 34H) H2, CPTr, pixyl, -CO*Ph*OCH$_3$; 6.57-6.29 (m, 2H) pixyl; 5.95 (d, $J_{1',2'}$ = 7.2 Hz, 1H) H1'; 5.07 (m, 1H) H2'; 4.23 (m, 1H) H4'; 3.88 (s, 3H) -COPhO*CH*$_3$; 3.54 (m, 1H) H3'; 3.39-3.23 (m, 2H) H5', H5".

5'-*O*-(CPTr)-2'-*O*-pixyl-6-*N*-(4-anisoyl)-adenosine-3'-triethylammonium-(2-chlorophenyl)phosphate (2). Compound **1** (1.5 g, 0.98 mmol) in dry pyridine (6.8 ml) was treated with 0.25 M acetonitrile solution of *o*-chlorophenyl bis-(1,2,4-triazolide) (15.6 ml, 3.92 mmol) for 4 h. Aqueous ammonium bicarbonate (NH_4HCO_3) work-up followed by silica gel chromatography (1-7% EtOH / 1% pyridine / CH_2Cl_2) afforded **2** (Et_3NH^+ salt, 1.39 g, 78%), R_f: 0.5 (C). ^{1}H-NMR (90 MHz, $CDCl_3$+2,6-lutidine): 8.23-6.53 (m, 40H) H8, H2, arom; 6.24-6.15 (m, 2H) H1', pixyl; 5.33 (m, 1H) H2'; 4.74 (m, 1H) H4'; 4.49 (m, 1H) H3'; 3.89 (s, 3H) -COPhO*CH*$_3$; 3.30 (m, 2H) H5', H5"; 2.90 (q, 6H) CH_2 of Et_3NH^+; 1.21 (t, 9H) CH_3 of Et_3NH^+. ^{31}P-NMR ($CDCl_3$+CD_3OD+2,6-lutidine): -5.76 ppm.

5'-O(CPTr)-6-*N*-(4-anisoyl)-adenosine-3'-triethylammonium-(2-chlorophenyl)phosphate (3). The 3'-phosphodiester block **2** (1.3 g, 0.70 mmol) was dissolved in 2% MeOH-CH_2Cl_2 solution (85 ml) and chilled to ~0°C in an ice bath. Trichloroacetic acid (TCA) (1.75 g, 10.70 mmol) was dissolved also in 2% MeOH-CH_2Cl_2 solution (85 ml) and chilled to ~0°C prior to pouring into the above solution of **2**. The final concentration of acid was 0.055 M. After stirring for 40 min at ~0°C the reaction was quenched by a small amount of pyridine, and then poured into 0.2 M NH_4HCO_3 solution which was saturated with sodium chloride (NaCl) and acidified with dry ice. This aqueous phase (pH ~6.5) was extracted with CH_2Cl_2 (3 x 50 ml), the organic phase was dried over $MgSO_4$ and evaporated. The residue was dissolved in a small amount of CH_2Cl_2 and precipitated into a diethyl ether-petroleum ether solution (2:1, v/v). After centrifugation the supernatant was decanted and the white pixyl free solid **3** (Et_3NH^+ salt, 990 mg, 90%) was dried in vacuum, R_f: 0.2 (C). ^{1}H-NMR (90 MHz, $CDCl_3$+CD_3OD): 8.54-6.83 (m, 28H) H8, H2, arom; 6.13 (m, 1H) H1'; 5.14 (m, 2H) H2', H3'; 4.38 (m, 1H) H4'; 3.85 (s, 3H) -COPhO*CH*$_3$; 3.35 (m, 2H) H5', H5"; 2.90 (q, 6H) CH_2 of Et_3NH^+; 1.21 (t, 9H) CH_3 of Et_3NH^+. ^{31}P-NMR ($CDCl_3$+CD_3OD): -6.94 ppm.

5'-*O*-(CPTr)-2'-[(2-cyanoethyl)-(2-(4-nitrophenyl)ethyl)phosphoryl]-6-*N*-(4-anisoyl)-adenosine-3'-triethylammonium-(2-chlorophenyl)phosphate(4). (2-cyanoethoxy)-(2-(4-nitrophenyl)ethoxy)-(diisopropyl amino)phosphine (2.3 g, 6.3 mmol) was weighed into a dry 50 ml round bottomed flask and dry DMF/MeCN (7:3, v/v) (11ml) was added under argon. Then dry and sublimed tetrazole (1.34 g, 18.9 mmol) was added under stirring, and it rapidly went into solution followed by a quick formation of a precipitate. After few minutes stirring, the compound **3** (990 mg, 0.63 mmol) previously dissolved in dry DMF/MeCN (7:3 v/v) (10 ml) was added to the suspension and the clear reaction solution was stirred for 45 min. A solution of 0.1 M I_2 in THF/pyridine/H_2O (7:2:1, v/v/v) (66 ml, 6.6 mmol) was added and the solution was stirred for 15 min, then poured into 0.1 M $Na_2S_2O_3$ / concentrated NH_4HCO_3 solution and extracted with CH_2Cl_2. The DMF free gum was obtained after toluene co-evaporation of the organic residue and purified by short column chromatography (1-6% EtOH / CH_2Cl_2) to finally give the 3'-phosphodiester-2'-phosphotriester block **4** (Et_3NH^+ salt, 690 mg, 59%), R_f: 0.40 (C). ^{1}H-NMR (90 MHz, $CDCl_3$): 8.54 (s, 1H) H8; 8.31-6.83 (m, 31H) arom, H2; 6.18 (m, 1H) H1'; 5.94 (m, 1H) H2'; 5.33 (m, 1H) H3'; 4.68-4.01 (m, 5H) H4', -O*CH*$_2$CH$_2$CN, -O*CH*$_2$CH$_2$PhNO$_2$; 3.86 (s, 3H) -COPhO*CH*$_3$; 3.41 (m, 2H) H5', H5"; 3.04 (m, 2H)-OCH$_2$*CH*$_2$PhNO$_2$; 2.90 (q, 6H) CH_2 of Et_3NH^+; 2.71

(m, 2H) -OCH_2CH_2CN; 1.21 (t, 9H) CH_3 of Et_3NH^+. ^{31}P-NMR ($CDCl_3$+CD_3OD): -4.26; -4.31; [-8.90 to 9.26] ppm.

5'-*O*-(CPTr)-2'-[(-cyanoethyl)-(2-(4-nitrophenyl)ethyl)phosphoryl]-6-*N*-(4-anisoyl)-adenosine-3'-ethyl-(2-chlorophenyl)phosphate (5). The compound **4** (690 mg, 0.373 mmol) was condensed with dry ethanol (438 µl, 7.47 mmol) in dry pyridine (1.7 ml) by addition of *N*-methylimidazole (MeIm) (446 µl, 5.62 mmol) and 1-mesitylenesulfonyl chloride (MsCl) (978.2 mg, 4.5 mmol). The reaction was stirred for 12 h followed by an aqueous NH_4HCO_3 work-up, the organic phase was dried and purified by silica gel chromatography (1-4% EtOH / CH_2Cl_2). The 2',3'-bisphosphotriester block **5** was obtained as a white powder, after co-evaporation with toluene and cyclohexane (494 mg, 75%), R_f: 0.8 (B). ^1H-NMR (90 MHz, $CDCl_3$): 8.59-6.94 (m, 32H) H8, H2, arom; 6.31-6.01 (m, 2H) H1', H2'; 5.68 (m, 1H) H3'; 4.62-4.01 (m, 7H) H4', -OCH_2CH_2CN, -$OCH_2CH_2PhNO_2$, -OCH_2CH_3; 3.88 (s, 3H) -$COPhOCH_3$; 3.66-3.48 (m, 2H) H5', H5"; 2.96 (m, 2H) -$OCH_2CH_2PhNO_2$; 2.71 (m, 2H) -OCH_2CH_2CN; 1.33 (t, 3H) -OCH_2CH_3. ^{31}P-NMR ($CDCl_3$ + CD_3OD): [-2.24 to -2.78], [-7.37 to -7.69] ppm.

2'-[(-cyanoethyl)-(2-(4-nitrophenyl)ethyl)phosphoryl]-6-*N*-(4-anisoyl)-adenosine-3'-ethyl-(2-chloro phenyl)phosphate (6). Compound **5** (494 mg, 0.282 mmol) was dissolved in a solution of 1 M hydrazine hydrate (409 µl, 8.46 mmol) in pyridine-acetic acid (3:1, v/v, 8.5 ml). The solution became rapidly colored with the precipitation of 4,5-dichlorophthalhydrazide. Acetic acid (17 ml) was added dropwise at 0°C, and the resulting solution stirred for 10 min at 20°C. The excess of hydrazine hydrate was then quenched by the addition of acetyl acetone (1.24 ml, 11.25 mmol). Aqueous NH_4HCO_3 work up followed by silica gel column chromatography (1-5% EtOH / CH_2Cl_2) afforded **6** (190 mg, 77%), R_f: 0.6 (B). ^1H-NMR (90 MHz, $CDCl_3$): 8.77 (s, 1H) H8; 8.14-6.96 (m, 13H) H2, arom; 6.17 (m, 1H) H1'; 5.76 (m, 1H) H2'; 5.36 (m, 1H) H3'; 4.63 (m, 1H) H4'; 4.43-3.90 (m, 9H) H4', H5', H5", -OCH_2CH_2CN, -$OCH_2CH_2PhNO_2$, -OCH_2CH_3; 3.90 (s, 3H) -$COPhOCH_3$; 2.90 (m, 2H) -$OCH_2CH_2PhNO_2$; 2.57 (m, 2H) -OCH_2CH_2CN; 1.32(m, 3H) -OCH_2CH_3. ^{31}P-NMR (90 MHz) ($CDCl_3$ + CD_3OD): [-2.58 to -3.18], [-7.10 to -7.34] ppm.

Dimer 7: The 5'-hydroxy block **6** (190 mg, 0.219 mmol) was condensed with 5'-*O*-DMTr-2'-*O*-Thp-3-*N*-(4-anisoyl)-uridine-3'-triethylammonium(2-chlorophenyl)phosphate (**13**) (*Low R_f*) (173 mg, 0.165 mmol) in dry pyridine (0.9 ml) in presence of 1-mesitylenesulfonyl-3-nitro-1,2,4-nitrotriazole (MSNT) (176 mg, 0.593 mmol). Aqueous NH_4HCO_3 work up and silica gel column chromatography (1-3% EtOH / 1% pyridine / CH_2Cl_2) afforded the dimer **7** (197 mg, 50%), R_f: 0.5 (B).^1H-NMR (270 MHz, $CDCl_3$): 9.00 (br, 1H) NH; 8.77, 8.72 (2xs, 1H) AH8; 8.20-6.80 (m, 35H) AH2, UH6, arom; 6.34-6.12 (m, 2H) AH1', UH1'; 5.66-3.94 (m, 16H) sugar protons, Thp, -OCH_2CH_2CN, -$OCH_2CH_2PhNO_2$, -OCH_2CH_3; 3.90, 3.87 (3xs, 6H) 2x-$COPhOCH_3$; 3.78 (2xs, 6H) 2x-OCH_3; 3.60-3.34 (m, 4H) UH5', UH5", Thp; 2.94 (m, 2H) -$OCH_2CH_2PhNO_2$, 2.60 (m, 2H) -OCH_2CH_2CN; 1.40-1.20 (m, 9H) Thp, -OCH_2CH_3. ^{31}P-NMR ($CDCl_3$ + CD_3OD): [-2.24 to -3.05] , [-7.30 to -8.25] ppm.

Dimer 8: The dimer **7** (197 mg, 0.109 mmol) was treated with Et_3N (303 µl, 2.18 mmol) in pyridine (1.21 ml) and stirred for 4 h. The reaction mixture was diluted with dry pyridine and evaporated and co-evaporated with toluene. Silica gel column chromatography afforded the dimer **8** (NH_4^+ salt, 153 mg, 80%), R_f: 0.7 (B).^1H-NMR (90 MHz, $CDCl_3$): 8.50-6.75 (m, 36H) AH8, AH2, UH6, arom; 6.32-6.07 (m, 2H) AH1', UH1'; 5.66 -3.95 (m, 16H) UH5, sugar protons, Thp, -OCH_2CH_3, -$OCH_2CH_2PhNO_2$; 3.86 (s, 6H) 2x-$COPhOCH_3$; 3.76 (s, 6H) 2x-OCH_3; 3.60-3.34 (m, 4H) UH5', UH5", Thp; 2.51 (m, 2H) -$OCH_2CH_2PhNO_2$; 1.40-1.20 (m, 9H) Thp, -OCH_2CH_3: ^{31}P-NMR ($CDCl_3$ + CD_3OD): -3.58, -7.51, -7.86, -8.37, -9.54 ppm.

Tetramer 9: A mixture of **8** (153 mg, 84.8 µmol) and the dimer **10** (148 mg, 97.5 µmol) was dissolved in dry pyridine (0.45 ml). MSNT (175 mg, 0.594 mmol) was added and the reaction mixture stirred under argon 3 h. An aqueous $NaHCO_3$ work up, followed by silica gel chromatography (1-4% EtOH / 1% pyridine/ CH_2Cl_2) afforded the tetramer **9** (248 mg, 79%) as a white powder after co-evaporation with toluene and cyclohexane, R_f: 0.6 (C).^1H-NMR (90 MHz, $CDCl_3$ + 2,6-lutidine): 8.63 (br, 1H) AH8; 8.30-6.64 (m, 58H) AH2, CH6, CH5, UH6, GH8, arom; 6.30-5.11 (m, 12H) anomeric protons, sugar protons, UH5; 4.89-4.08 (m, 20H) sugar protons, 3x Thp , -OCH_2CH_3, -$OCH_2CH_2PhNO_2$, -OCH_2CH_2CN, 3.85 -3.79 (2xs, 9H) 3x -$COPhOCH_3$, 3.76 (s, 6H) 2x-OCH_3, 3.70-2.65 (m, 12H) sugar protons, 3x Thp, -$OCH_2CH_2PhNO_2$, -OCH_2CH_2CN; 1.85-1.25 (m, 30H) 3x Thp, t-butyl, -OCH_2CH_3. ^{31}P-NMR ($CDCl_3$ + CD_3OD): -1.63, -1.73, -2.49, [-6.78 to -7.88] ppm.

Tetramer 11: The tetramer **9** (248 mg, 75 µmol) was detritylated using TCA (0.07 M). After stirring for 4h 45 min followed by an aqueous NH_4HCO_3 work up the organic phase was dried and purified by sillica gel chromatography (1-4% EtOH / CH_2Cl_2). The 5'-hydroxy-2'-phosphotriester block **11** was obtained as a white powder after co-evaporation with toluene and cyclohexane (110 mg, 50%), R_f: 0.35 (C). ^1H-NMR (90 MHz, $CDCl_3$): 8.63 (br, 1H) AH8; 8.30-6.75 (m, 45H) AH2, CH6, CH5, UH6, GH8, arom; 6.46-5.20 (m, 12H)

anomeric protons, sugar protons, UH5; 4.87-3.97 (m, 20H) sugar protons, 3 x Thp, -OCH_2CH_3, -O$CH_2CH_2PhNO_2$, -OCH_2CH_2CN, 3.86 -3.79 (2xs, 9H) 3x -COPhOCH_3; 3.86-2.73 (m, 12H) sugar protons, 3 x Thp, -OCH$_2CH_2$PhNO$_2$, -OCH$_2CH_2$CN; 1.85-1.25 (m, 30H) 3 x Thp, t-butyl, -OCH$_2CH_3$. ^{31}P-NMR ($CDCl_3$ + CD_3OD):-2.05, -2.29, -2.95, [-7.15 to -8.37] ppm

Tetramer 12a : The same procedure as described for the dimer **8** was applied to the tetramer **11**, however the purification was carried out using preparative TLC plates, pre-washed with 20% MeOH / CH_2Cl_2 and then run with 15% MeOH / CH_2Cl_2, which afforded the compound **12a** (Et_3NH^+ salt, 65 mg, 60%), R_f: 0.8 (B).^{1}H-NMR (270 MHz, $CDCl_3$): 8.63 (br, 1H) AH8; 8.30-6.75 (m, 45H) AH2, CH6, CH5, UH6, GH8, arom; 6.46-5.20 (m, 12H) anomeric protons, sugar protons, UH5; 5.00-3.97 (m, 16H) sugar protons, 3x Thp , -OCH_2CH_3, -O$CH_2CH_2PhNO_2$, 3.86 -3.79 (2xs, 9H) 3x -COPhOCH_3; 3.86-2.73 (m, 16H) sugar protons, 3x Thp, -OCH$_2CH_2$PhNO$_2$, CH_2 of Et_3NH^+; 1.85-1.25 (m, 39H) 3 x Thp, t-butyl, -OCH$_2CH_3$, CH_3 of Et_3NH^+. ^{31}P-NMR ($CDCl_3$ + CD_3OD):[-2.07 to -2.97], [-6.27 to -8.25]ppm.

Fully protected tetrameric lariat RNA 12b: The 5'-hydroxy-2'-phosphodiester block **12a** (65 mg, 22 µmol) was co-evaporated with dry pyridine and redissolved in dry pyridine (250 ml / mmol, 4 mM). Then MSNT (71 mg, 15 eq) was added to the reaction solution over a period of 17 h. Aqueous NH_4HCO_3 work up followed by silica gel column chromatography (1-4% EtOH in CH_2Cl_2) afforded the title compound **12b** as a white powder after co-evaporation with toluene and cyclohexane (40 mg, 63%), R_f: 0.75 (C). ^{1}H-NMR (90MHz, $CDCl_3$ + CD_3OD): 8.87-6.77 (m, 42H) arom, AH8, AH2, GH8, CH6, CH5, UH6; 6.69-4.15 (m, 32H) UH5, sugar protons, 3 x Thp ,-OCH_2CH_3, -O$CH_2CH_2PhNO_2$; 3.86 (m, 9H) 3 x -COPhOCH_3; 3.34-2.50 (m, 8H) 3x Thp, -OCH$_2CH_2$PhNO$_2$; 1.60-1.05 (m, 30H) 3x Thp, t-butyl, -OCH$_2CH_3$.^{31}P-NMR ($CDCl_3$+CD_3OD): -1.12. to -9.06 ppm.

Deprotection of lariat 12b to 14: Fully protected oligomer **12b** (40 mg, 13.9 µmol) was dissolved in dry pyridine (1.3 ml) followed by addition of 0.5 M DBU / pyridine solution (278 µl) and the reaction mixture was stirred for 4.5 h at 20°C. The reaction was quenched by addition of 1 M acetic acid-pyridine (139 µml). After 10 min stirring the solution was poured into H_2O and extracted with CH_2Cl_2. The organic phase was evaporated and co-evaporated with distilled dioxane and redissolved in dioxane/water (4.2 ml, 8:2, v/v); *syn*-4-nitrobenzaldoxime (115 mg, 695 µmol) and 1,1,3,3-tetramethyl guanidine (78 µl, 625 µmol) were then added. After stirring for 26 h at RT the solvents were removed by evaporation *in vacuo* and concentrated ammonia (20 ml, d = 0.9) was added. The reaction mixture was stirred for 7 days and was then evaporated and co-evaporated with distilled water. The residue was treated with 80% aq. acetic acid (15 ml) for 24 h. After evaporation and co-evaporation with distilled water the residue was dissolved in distilled water (20 ml) and extracted with diethyl ether (3 x 20 ml) and subsequently with CH_2Cl_2 (6 x 20 ml) and the aqueous phase was evaporated to dryness.

Purification of lariat-RNA 14: DEAE-Sephadex A-25 column chromatography was carried out by redissolving the residue in 0.001 M NH_4HCO_3 buffer and applied to a DEAE-Sephadex A-25 column (2 x 25 cm, HCO_3^-form) and eluted with a linear gradients 0.001 M - 0.30 M - 0.80 M of NH_4HCO_3 solution (500 ml/1000 ml/500 ml respectively; pH 7.5). The main symmetrical peak eluted between 0.5 and 0.6 M was collected (471 A_{260} units, 82%). Analytical HPLC (0-40% of solution B, over 30 min at a rate of 1 ml/min) revealed that this material contained approximately 60-70% of the total product. Batches of 5-7 mg of lyophilized material obtained from the middle and rear fractions were each dissolved in solution A at pH 7.0 (900-1000 µl). These were then taken in Eppendorf tubes, centrifuged and then injected in a semi-preparative Spherisorb S5ODS2 column (8 x 250 mm) equilibrated in solution A. Gradient elution with MeCN in 0.1M TEAA (0-5% solution B over 20 min, 1 ml/min) resolved the desired peak with base-line separation (detector was set at 254 nm). The purified material was collected, evaporated and lyophilized several times (~9 x 1 ml) until all TEAA salt was removed (monitored by ^{1}H-NMR) to give **14** (Et_3NH^+ salt, 15 mg, 30%). For preparation of NMR samples, the Et_3NH^+ salt of **14** was converted to the corresponding Na^+ salt by elution of appropriately sized batches with distilled water through a Dowex column (1 x 20 cm, Na^+ form) maintained at 2°C. The eluant was collected in a 50 ml round bottomed flask which was cooled in EtOH / dry ice bath. The frozen aqueous solution was lyophilized and the white residue was redissolved in a small amount of distilled water, swiftly transferred to a dry ice frozen 10 ml screw cap bottle, lyophilized and thereafter stored as a dry white solid.

(B) NMR Spectroscopy

The NMR samples were lyophilized from 99.8% ^2H_2O, dissolved in 0.5 ml of 99.9% ^2H_2O and transferred into 5 mm tubes. The sample concentration used for recording the NMR spectra were 4 mM. The NMR spectra

of **14** and **15** were recorded on a Bruker AMX-500 spectrometer operating at 500.13 MHz for proton, 202.4 MHz for phosphorus and 125.76 MHz for carbon. ^{1}H-NMR spectra were collected with 32K data points and zero filled to 64K. A trace of dry acetonitrile was added as an internal reference (set at 2.00 ppm) for ^{1}H-NMR chemical shifts measurements. ^{31}P-NMR spectra were acquired with 16K data points and zero filled to 32K. 3',5'-cyclic AMP was used as an external reference set at 0.00 ppm. The 2D spectra were recorded in pure-phase absorption mode with the time proportional incrementation method and with low power preirradiation of the residual HDO peak during the relaxation delay. The DQF-COSY with and without ^{31}P decoupling were acquired with 8192 complex data points in t_2 and 512 points in t_1. The data were zero filled to give a 4096 x 2048 point matrix, and a $\pi/4$ shifted sine-square bell window was applied in both directions before Fourier transformation. For each crosspeak, the corresponding row was extracted as a 1D spectrum, backtransformed and zero filled twice before Fourier transformation. The final digital resolution was 0.25 Hz/pt. The clean-TOCSY spectra were acquired with 256 spectra of 4K data points. The t_1 domain was zero filled to 1K and a $\pi/2$ shifted sinesquare window was applied in both dimensions before Fourier transformation. The MLEV-17 sequence was applied for mixing using an extra delay of 65 μs for compensation of nOe, total mixing time 300ms. Two different power levels were used for excitation and spinlock. The NOESY and ROESY spectra were acquired with 2048 complex data points in t_2 and 256 points in t_1. The data were zero filled to give a 2048 x 1024 point matrix, and a sine-square bell window was applied in both directions before Fourier transformation. For ROESY, a CW spinlock (2.5 kHz) was used for mixing. A mixing time of 600 ms was used. The ^{1}H-^{31}P correlation experiments were run in the absolute value mode. A delay of 83 ms was used. 256 experiments were recorded, for each experiment we recorded 32 scans of 2K data points. The spectral range used was 1500 Hz in the t_1 direction and 4200 Hz in t_2. The spectrum was zero filled to 512 data points in t_1 and a sinesquare ($\pi/2$) window was applied before Fourier transformation. The ^{1}H-^{13}C chemical shift correlation experiments (HSQC) were recorded with 1200 experiments of 2048 points and 32 scans. Heteronuclear decoupling was achieved by the GARP-1 sequence. The spectral width was 4200 Hz in the ^{1}H direction and 2515 Hz in the ^{13}C dimension. The digital resolution was 1.0 Hz/point in the t_2 direction and 0.5 Hz/point in the t_1 direction.

Acknowledgements

Authors thank Swedish Board for Technical Development and Swedish Natural Science Research Council for generous financial support. Authors also thank Wallenbergs Stiftelse, University of Uppsala and Swedish Research Council (FRN) for funds toward the purchase of 500 MHz NMR spectrometer.

References

(1) Kole, R.; Altman, S. *Proc. Natl. Acad. Sci. USA* **1979**, *76*, 3795.
(2) Foster, A. C.; Symons, R. H. *Cell* **1987**, *49*, 211.
(3) Uhlenbeck, O. C. *Nature* **1987**, *328*, 596.
(4) Buzayan, J. M.; Gerlach, W. L.; Bruening, G. *Nature* **1986**, *323*, 349.
(5) Hampel, A.; Tritz, R. *Biochemistry* **1989**, *28*, 4929.
(6) Belinksy, M. G.; Britton, E.; Dinter-Gottlieb, G. *FASEB J.* **1993**, *7*, 130.
(7) Sharmeen, L.; Kuo, M. Y. P.; Dinter-Gottlieb, G. *J. Virol.* **1988**, *62*, 685.
(8) Saville, B. J.; Collins, R. A. *Cell* **1990**, *61*, 685.
(9) Koizumi, M.; Ohtsuka, E. *Biochemistry* **1991**, *30*, 5145.
(10) Agback, P.; Sandström, A.; Yamakage, S.-I.; Sund, C.; Glemarec, C.; Chattopadhyaya, J. *J. Biochem. Biophys. Meth.* **1993**, *27*, 229.
(11) Agback, P.; Glemarec, C.; Yin, L.; Sandström, A.; Plavec, J.; Sund, C.; Yamakage, S.-I.; Viswanadham, G.; Rousse, B.; Puri, N.; Chattopadhyaya, J. *Tetrahedron Lett.* **1993**, *34*, 3929.
(12) Rousse, B.; Puri, N.; Viswanadham, G.; Agback, P.; Glemarec, C.; Sandström, A.; Sund, C.; Chattopadhyaya, J. *Tetrahedron* **1994**, *50*, 1777.
(13) Agback, P.; Glemarec, C.; Sandström, A.; Yin, L.; Plavec, J.; Sund, C.; Yamakage, S.-I.; Viswanadham, G.; Rousse, B.; Puri, N.; Chattopadhyaya, J. in *Structural Biology: State of the Art 1993, Proceedings of the 8th Conversations*; Sarma, R. H. Sarma, M. H. (Eds.); Adenine Press, New York, Vol. 1, p.293 (1994).
(14) Reese, C. B. *Tetrahedron* **1978**, *34*, 3143.
(15) Sund, C.; Agback, P.; Chattopadhyaya, J. *Tetrahedron* **1993**, *49*, 649.

(16) Sekine, M.; Hata, T. *J. Am. Chem. Soc.* **1984**, *106*, 5763.
(17) Sekine, M.; Hata, T. *J. Am. Chem. Soc.* **1986**, *108*, 4581.
(18) Sandström, A.; Kwiatkowski, M.; Chattopadhyaya, J. *Acta Chem. Scand.* **1985**, *B 39*, 273.
(19) Efimov, V.A. and Chakhmakhcheva, O.G *et al. Nucl. Acids Res.*, **1983**, *11*, 8369.
(20) Sund, C.; Agback, P.; Chattopadhyaya, J. *Tetrahedron*, **1991**, *47*, 9659. Bonora, G *et al.*; *Nucl. Acids Res.*, **1990**, *18*, 2661. van Boom, J *et al. Nature*, **1987**, *325*, 279 & *Nucl. Acids Res.* **1988**, *10*, 4607. Jones, R *et al. Nucleosides & Nucleotides*, **1985**, *4*, 377. Reese, C.B, *et al. Nucl. Acids Res.* **1989**, *10*, 8221. Piccially, G *et al. Tetrahedron Lett.* **1987**, *228*, 5727 & *Tetrahedron*, **1989**, *45*, 4523.
(21) Sund, C.; Földesi, A.; Yamakage, S.; Agback, P.; Chattopadhyaya, J. *Tetrahedron* **1991**, *47*, 6305.
(22) Gorenstein, D. G.; Findlay, J. B.; Momii, R. K.; Luxon, B. A.; Kar, D. *Biochemistry* **1976**, *15*, 3796.
(23) Griesinger, C.; Otting, G.; Wütrich, K.; Ernst, R. R. *J. Am. Chem. Soc.* **1988**, *110*, 7870.
(24) Derome, A. E.; Williamson, M. *J. Magn. Reson.* **1990**, *88*, 177.
(25) Morris, G. A. *Magn. Reson. Chem.* **1986**, *24*, 371.
(26) Bax, A.; Griffey, R. H.; Hawkins, B. L. *J. Magn. Reson.* **1983**, *55*, 301.
(27) de Leeuw, F. A. A. M.; Altona, C. *J. Chem. Soc. Perkin Trans. II* **1982**, 375.
(28) Plavec, J.; Thibadeau, C.; Viswanadham, G.; Sund, C.; Chattopadhyaya, J. *J. Chem. Soc. Chem. Comm.* **1994**, p.781. *J. Am. Chem. Soc.* **1994**, *116*, 4038.
(29) Schmieder, P.; Ippel, J. H.; van den Elst, H.; van der Marel, G. A.; van Boom, J. H.; Altona, C.; Kessler, H. *Nucl. Acids Res.* **1992**, *20*, 4747.
(30) Altona, C. *Recl. Trav. Chim. Pays-Bas* **1982**, *101*, 413.
(31) Lankhorst, P. P.; Haasnoot, C. A. G.; Erkelens, C.; Altona, C. *J. Biomol. Struct. Dynamics* **1984**, *1*, 1387.
(32) Lankhorst, P. P.; Haasnoot, C. A. G.; Erkelens, C.; Westerink, H. P.; van der Marel, G. A.; van Boom, J. H.; Altona, C. *Nucl. Acids Res.* **1985**, *13*, 927.
(33) The initial structure was subjected to conjugate gradient energy minimization in vacuum (until rms gradient < 10^{-4} kcal.mol^{-1}.Å^{-1} (= 3614 steps)) with the SANDER module of the AMBER 4.0 (ref. 35) with the supplied all atom force field, parm91 using NMR constraints (Tables 3 & 6). A distance dependent dielectric ($\varepsilon = 4r_{ij}$) and infinite cutoffs for nonbonded interactions was used.
(34) The structure of **14**, generated from the procedure described in ref. 33, was used as the starting conformer for the 226 ps MD in water using periodic boundary conditions and the modified AMBER force field (see ref. 10 and 12 for details). Five sodium ions were placed in identical manner as discussed in ref. 10 and 12. 64 Cubes of 216 TIP3P water molecules were added followed by the removal of water molecules being either too close (2.0 Å for oxygen, 1.8 Å for hydrogens) or too distant (7.5, 10.0, and 10.0 Å in X-, Y-, and Z-directions respectively) from the solute. The lariat tetramer structure was solvated by 1074 water molecules. Two initial energy minimizations were performed on the solvated structure using a protocol described in ref. 10 and 12. The final structure was then subjected to 226 ps of MD exactly under the same condition as described in ref. 12 except the final bath temperature at 294 K.
(35) Pearlman, D. A.; Case, D. A.; Caldwell, J. C.; Seibel, G. L.; Singh, U. C.; Weiner, P.; Kollman, P. A., AMBER 4.0 (University of California, San Fransisco, CA) **1991**
(36) Frisch, M. J.; Trucks, G. W.; Head-Gordon, M.; Gill, P. M. W.; Wong, M. W.; Foresman, J. B.; Johnson, B. G.; Schlegel, H. B.; Robb, M. A.; Replogle, E. S.; Gomperts, R.; Andres, J. L.; Raghavachari, K.; Binkley, J. S.; Gonzalez, C.; Martin, R. L.; Fox, D. J.; Defrees, D. J.; Baker, J.; Stewart, J. J. P.; Pople, J. A., Gaussian 92 Rev. A (Gaussian Inc., Pittsburgh, PA) **1992**
(37) Singh, U. C.; Kollman, P. A. *J. Comp. Chem.* **1984**, *5*, 129.
(38) Plavec, J.; Yang, L.; Chattopadhyaya, J. *unpublished data* **1993**,
(39) Taira, K.; Uebayasi, M.; Furukawa, K. *Nucl. Acids Res.* **1989**, *17*, 3699.
(40) Previous NMR studies have clearly shown the presence of two distinctly identifiable dynamically interconverting N and S conformations of some sugar moieties in B⇄ Z DNA[a,b] or A ⇄ Z RNA[c] or A-form ⇄ B-form lariat RNA[10-12] transformations as a result of the change of the salt or alcohol concentration in the buffer or as a result of change of NMR measurement temperature. These dynamically interconverting two-state N and S sugar pseudorotamers are characterized ($^3J_{1'2'}$ = ~0.5 Hz and ~8 Hz, respectively) by only two distinct sets of resonances owing to their different stereochemical environments. No evidence for the third state has been detected yet in solution. That is why we have considered only a dynamic two-state N⇄ S equilibrium. (a) Feigon, J *et al. Nucl. Acids Res.* **1984**, *12*, 1243. (b) Tran-Dinh, S *et al. Biochemistry* **1984**, *23*, 1362. (c) Tinoco, I *et al. Nucl. Acids Res.* **1986**, *14*, 1279; *Biopolymers* **1990**, *29*, 109.

(*Received in UK* 7 *April* 1994; *revised* 24 *May* 1994; *accepted* 27 *May* 1994)