Tetrahedron Vol. 43, No. 17, pp. 3997 to 4006, 1987
Printed in Great Britain.

0040-4020/87 $3.00+.00

A 270 MHz ^{1}H-NMR STUDY OF THREE "BRANCHED" RIBONUCLEOTIDES $A^{2'p5'U}_{3'p5'G}$, $A^{2'p5'C}_{3'p5'G}$ AND $A^{2'p5'G}_{3'p5'G}$: THE IMPORTANCE OF BEING BRANCHED WITH A 2'⟶5' PHOSPHODIESTER BOND IN THE PRE-mRNA PROCESSING REACTIONS (SPLICING).

J-M. VIAL, G. REMAUD, N. BALGOBIN AND J. CHATTOPADHYAYA*

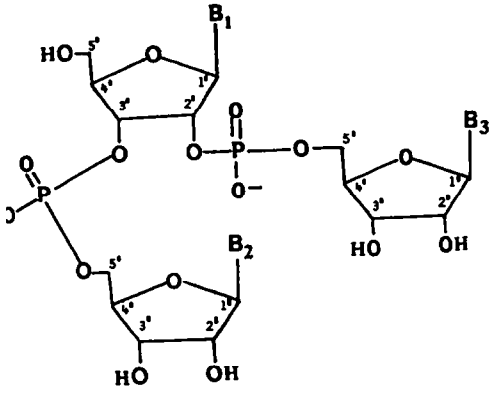
Department of Bioorganic Chemistry, Uppsala University, Biomedical Center,
Box 581, S-751 23 Uppsala, Sweden

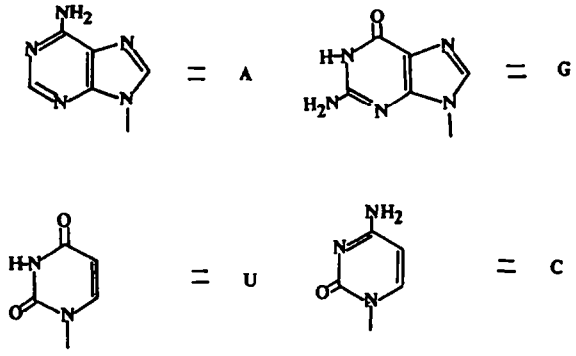
(*Received in UK* 12 *May* 1987)

Abstract: Conformational studies with three analogues 3, 4 and 5 of naturally-occurring branched ribonucleotides 1 and 2 have shown that they adopt a secondary structure which is overwhelmingly controlled by a stacking between adenine(2' → 5')nucleobase residues while the 3' → 5' linked guanine residue is apart. This observation along with the fact that all four nucleosides can form lariats in nuclear mRNA splicing, but it is guanosine as the 2' → 5' nucleotide that can drive the splicing reaction to completion, suggests a biological significance of the additional 2' → 5' phosphodiester bond formation in the splicing reaction. It is likely that the final implication of formation of such a 2' → 5' linked lariat is that it provides a pathway to assume a free energy minimum conformation through the 2' → 5' stacking, especially in the case of guanosine, to drive the splicing reaction to completion (stacking-driven-energy-pump).

A two-step mechanism[1-6] has been proposed for the precursor mRNA splicing in mammalian cells, although precise details of this unique chemical reaction have yet to be resolved. In the first step of the reaction, cleavage occurs at the 5' splice site to give an obligatory circular "branched" RNA (lariat) structure with intron attached to exon 2. In such lariat structures the 5'-phosphodiester residue from the 5'-end of the intron forms an additional 2' → 5' phosphodiester bond with an adenosine residue as the branch point, which is already 3' → 5' phosphodiester linked to a pyrimidine residue as in 1 and 2. In the second step, the scission of the covalent bond between the circular intron and the exon 2 takes place, followed by the ligation of exons 1 and 2. It has been now clearly established in case of nuclear mRNA splicing[7-10] that (1) the sequence requirement of branched structure formation, is not stringent, all nucleotides can serve as branch acceptor with different efficiency (A = C>U/G). Mutation of the branch acceptor A-nucleotide to U and G leads to the use of the A-nucleotide just one nucleotide upstream of the mutated position; (2) the structure of the branch is essential for the second step of the reaction, i.e., joining of the two exons. Branches at the A-nucleotide (wild type) and the C-nucelotide are permissive whereas U and G inhibit the second step of the splicing reaction; (3) G-nucleotide as the 2' → 5' nucleotide is absolutely important for the second step of the splicing reaction, however A, U and C are permissive just for the formation of branch. Naturally, we asked ourselves what is more important in the splicing reaction: the G residue as the 2' → 5' nucleobase or the 2' → 5' phosphodiester linkage or both? We have recently carried out a 270 MHz ^{1}H-NMR studies of two branched RNA structures in order to understand their structural make up that may have a special relevance in the

splicing reaction. We found[11] that the conformation of the branched trinucleotides 1 and 2 are very similar and is comprised of a strongly stacked state between 2'→ 5' linked adenine and guanine moieties while the 3'→ 5' linked pyrimidine residue is apart and adopts a coplanar state with the adenine part. In the present work we report our studies with the isomeric branched trinucleotides 3 and 4, in which 3' → 5' residue is guanine and the 2' → 5' residue is either a U or a C residue. This work shows, in conjunction with our previous studies, that the stacking between 2'→ 5' linked nucleobases is independent of nature of nucleobase in branched trinucleotides 1 to 4. This has been furthermore focussed in our present conformational studies with the branched trinucleotide 5 in which both 3' → 5' and 2' → 5' phosphodiester linked nucleobases are guanine residues and yet it is the 2' → 5' stacking that is overwhelmingly preferred to the 3'→ 5' stacking.





1 B^1 = A; B^2 = U; B^3 = G

2 B^1 = A; B^2 = C; B^3 = G

3 B^1 = A; B^2 = G; B^3 = U

4 B^1 = A; B^2 = G; B^3 = C

5 B^1 = A; B^2 = B^3 = G

Table 1. ^{1}H-NMR assignments[a] of branched trinucleotides 3, 4 and 5.

Compound	Fragment	H8	H2	H6	H5	H1'	H2'	H3'	H4'	H5'	H5''
3	pAp	8.24	7.96	-	-	6.02	5.06	4.79	4.40	3.66	
	pG	7.98	-	-	-	5.84	4.79	4.50	4.29	4.15	
	pU	-	-	7.32	5.60	5.54	3.93		3.76		3.47
4	pAp	8.23	7.89	-	-	6.03	5.08	4.80	4.40	3.66	
	pG	7.98	-	-	-	5.84	4.80	4.50	4.30	4.15	
	pC	-	-	7.29	5.70	5.50	3.86		3.78		3.51
5	pAp	8.16	7.78	-	-	6.07	5.17	4.83	4.29	3.67	
	3'pG	7.95	-	-	-	5.82	4.74	4.48	4.38	4.16	
	2'pG	7.62	-	-	-	5.55	4.38	4.17	3.91	3.84	3.55

[a] in ppm, CH_3CN as internal reference set at 2.00 ppm. T = 25 °C.

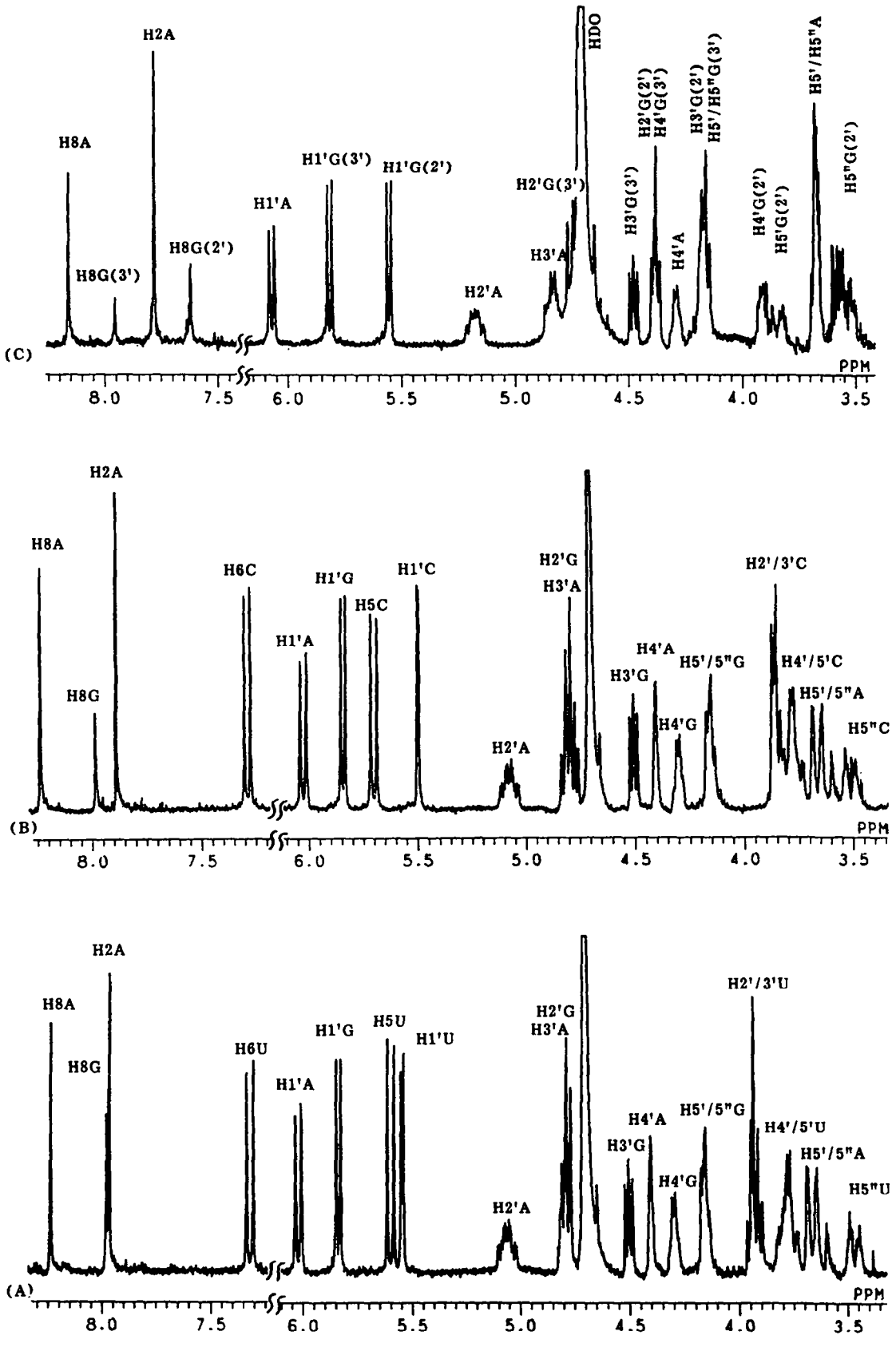


Fig.1: 270 MHz ^{1}H-NMR spectra of compound 3 (A), compound 4 (B), compound 5 (C) at 25 ^{o}C.

Assignment of resonance

All ^1H resonances of branched trinucleotides 3, 4 and 5 are assigned on the basis of incremental analysis by comparing the studies with the parent dimers: adenylyl(3'→ 5')guanine (A3'p5'G)[12], adenylyl(2' → 5')guanine (A2'p5'G)[11], adenylyl(2' → 5')uracil (A2'p5'U)[13] or adenylyl(2'→ 5')-cytidine (A2'p5'C)[12]. Such studies have shown the effect of the 2'-substituents, 5'-pU, 5'-pC or 5'-pG, in trimers 3, 4 and 5, respectively. The study of chemical shifts as a function of temperature (from 7 °C to 80 °C) have been also used for assignment purposes. Fig. 1 shows the 270 MHz ^1H-NMR spectra of trinucleotides 1, 2 and 3 in ^2H_2O. The assignment of all protons of trinucleotides have been corroborated by detailed 2D NMR studies using SECSY and/or COSY pulse sequences; a SECSY spectrum of the trinucleotide 3 is shown as an example (Fig. 2). Complete assignments of all ^1H resonances of all three trinucleotides, 3, 4 and 5, are shown in Table 1.

(i) Resonances of aromatic protons

H6 resonance of the pyrimidine moiety was assigned by its characteristic coupling with H5, appearing as a doublet of ca. 8 Hz for uracil and ca. 7.5 Hz for cytosine. A comparative study of aromatic protons of mononucleotides at high temperature with those of aromatic singlets from adenine and guanine residues of branched trinucleotides 3, 4 and 5, have shown that, in trinucleotides, they are shielded in the following order H8G>H2A>H8A as in A3'p5'G and A2'p5'G. For the trinucleotide 5, the resonances of 3'-substituent (3'-pG) are more downfield than the 2'-substituent (2'-pG). A comparison between A3'p5'G and A2'p5'G also leads to the same conclusion. It has been found, in general, that the resonances from the 2'-substituent are more shielded than those from the 3'-substituent; this observation is true irrespective of the nature of the nucleobase[11,14]. It is also known that the aromatic protons of the purines display different intensities (H2A>H8A>H8G) in ^2H_2O solution due to their different acidities, which was also used for the assignment purposes.

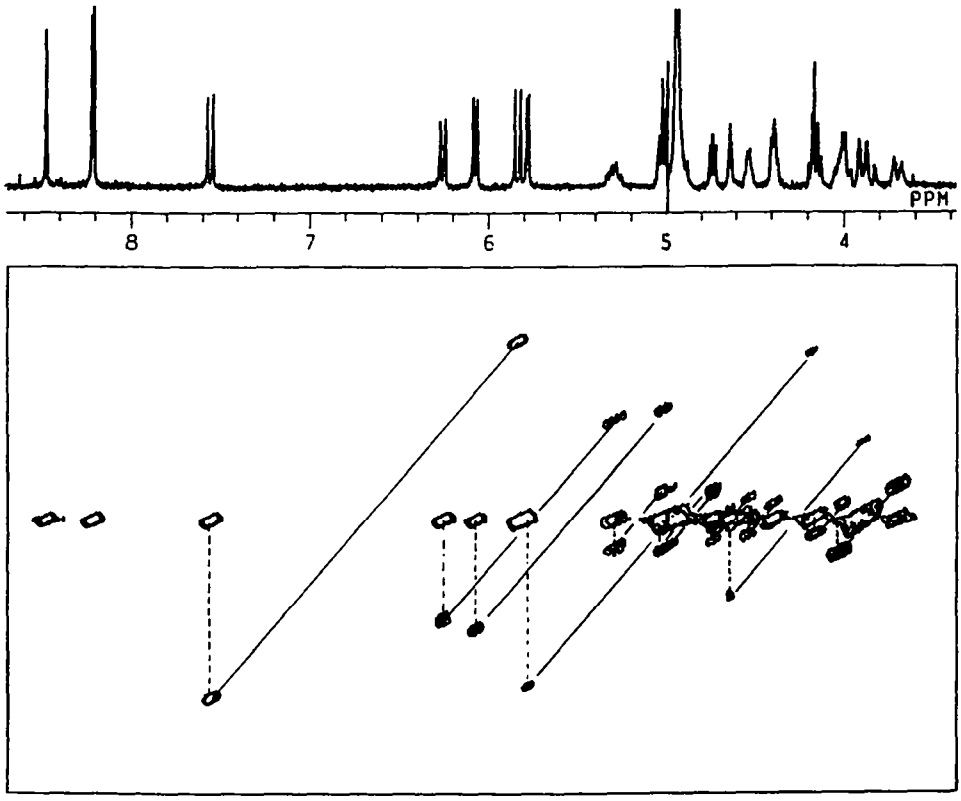


Fig.2: 270 MHz SECSY spectrum of compound 3 at 25 °C. The connection of J network is shown by the solid lines.

(ii) Resonances of sugar protons

The H1'A, in trinucleotides 3, 4 and 5, is the most downfield signal of anomeric protons as determined from the connectivities of H2'A and H3'A which can be conveniently assigned due to their downfield shifts and characteristic ^{31}P-^{1}H couplings. An incremental assignment of the trinucleotides by comparison with the dinucleotides, A3'p5'G, A2'p5'G, A2'p5'U and A2'p5'C, as discussed above, has shown that H1'G is more downfield than H1'U and H1'C in 3 and 4, respectively. Similarly, a comparison of spectra of A3'p5'G and A2'p5'G with that of the trinucleotide 5 has clearly established that the H1' of 3' → 5' linked guanosine residue is more downfield than the H1' of 2' → 5' linked guanosine residue. 2D NMR (SECSY and COSY) experiments were performed for each di- and trinucleotides for a complete assignment of sugar protons. This is illustrated through the SECSY spectrum of trinucleotide 3 shown in Fig. 2. Fig. 2 shows that the J network of all three pentose sugar moieties are clearly accessible by following the correlation spots from the anomeric part to the highfield part.

Finally, 2D $^{31}P/^{1}H$ NMR shift correlation experiments for trinucleotides 3, 4 and 5 confirmed previous assignments of ^{31}P resonances for compounds 1 and 2. For example, the fig. 3 shows the $^{31}P/^{1}H$ correlation spectrum for compound 3 in which it can be seen that the 2' → 5' phosphodiester-phosphate is more shielded and experiences spin couplings with H5'U and H5''U. It should also be noted both in our present and previous studies[11] that the 5' and 5'' protons of the 2'-substituent are well separated while the corresponding protons from the 3'-substituent are superimposed at 270 MHz irrespective of the nature of the nucleobase.

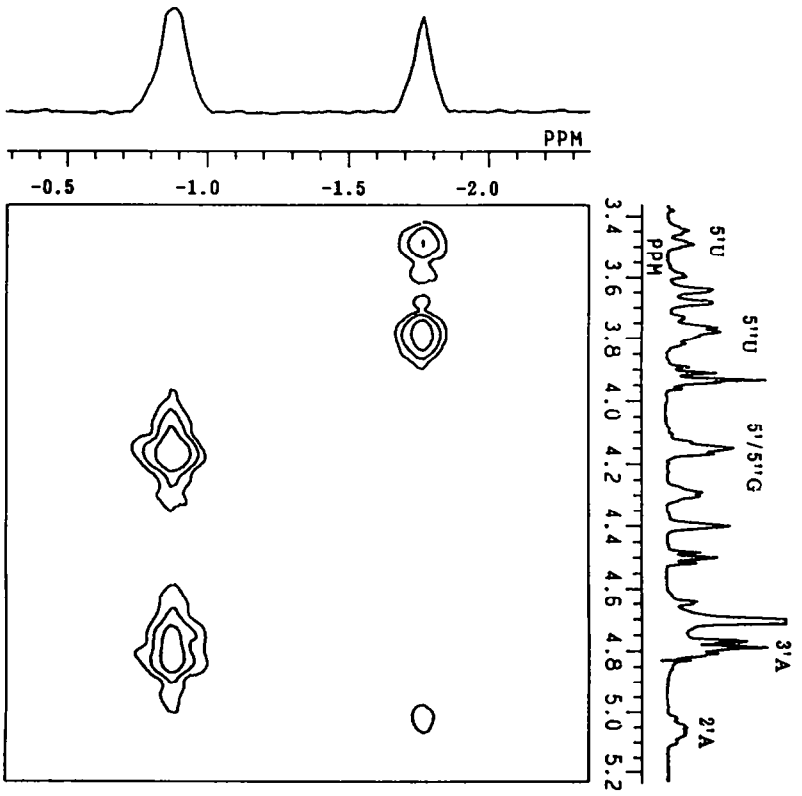


Fig.3: $^{31}P/^{1}H$ correlation spectrum of compound 3 at 25 °C. The projection is represented for the ^{31}P-NMR spectrum.

RESULTS AND DISCUSSION

The ring-current effect of the base, the diamagnetic anisotropy of the electron-rich group like C=O or P=O bonds and the electric field effects are the major factors that determine the chemical shifts in an oligonucleotide molecule[15]. In the stacked form, the aromatic (especially H2A, H5U, H5C) and anomeric protons experience an upfield shift due to the diamagnetic effect (ring current effect) of the neighbouring base. Therefore a study of the chemical shift of the aromatic and anomeric protons as a function of the temperature allows the detection of a stacking between two nucleobases under consideration. On the other hand H8A, H8G, H6U or H6C show a dependence on the glycosidic torsion angle. Temperature dependent chemical shifts of aromatic and anomeric protons from different nucleoside residues of trinucleotides 3, 4 and 5 are shown in Fig. 4 and are summarized in table 2.

Table 2. Temperature-dependent chemical shifts[a] of branched trinucleotides 3, 4 and 5.

Compound	Fragment	Proton	7°C	21°C	41°C	80°C
$A^{2'p5'U}_{3'p5'G}$		H8	8.222	8.237	8.237	8.222
(3)	pAp	H2	7.898	7.952	7.998	8.063
		H1'	5.949	6.006	6.047	6.089
		H6	7.278	7.313	7.343	7.397
	pU	H5	5.576	5.596	5.614	5.646
		H1'	5.508	5.537	5.566	5.616
		H8	7.952	7.974	7.982	7.967
	pG	H1'	5.843	5.839	5.835	5.826
$A^{2'p5'C}_{3'p5'G}$		H8	8.224	8.236	8.234	8.216
(4)	pAp	H2	7.818	7.874	7.925	8.006
		H1'	5.957	6.013	6.052	6.090
		H6	7.236	7.276	7.313	7.382
	pC	H5	5.678	5.698	5.715	5.752
		H1'	5.457	5.490	5.527	5.600
		H8	8.224	8.236	8.234	8.216
	pG	H1'	5.844	5.844	5.842	5.832
$A^{2'p5'G}_{3'p5'G}$		H8	8.162	8.162	8.161	8.166
(5)	pAp	H2	7.708	7.763	7.832	7.960
		H1'	6.045	6.066	6.083	6.109
		H8	7.606	7.618	7.629	7.658
	2'pG	H1'	5.539	5.550	5.565	5.604
		H8	7.957	7.958	7.952	7.927
	3'pG	H1'	5.819	5.818	5.816	5.812

[a] CH_3CN as internal reference set at 2.000 ppm.

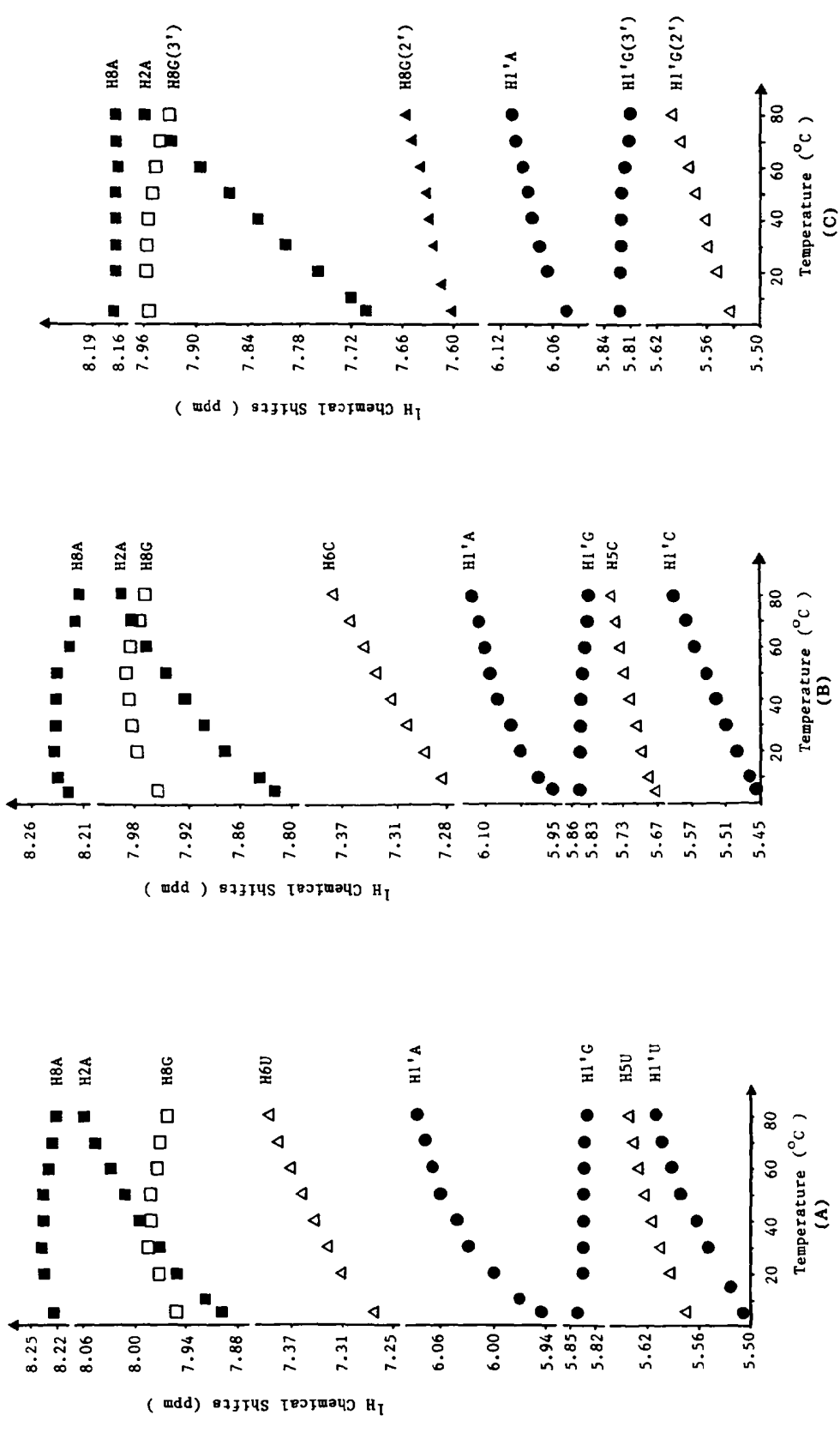


Fig.4: Temperature dependent variation of ^1H chemical shifts of aromatic and anomeric protons of (A) compound 3, (B) compound 4, (C) compound 5.

A comparison between temperature dependent chemical shifts in 3, 4 or 5 with those of A3'p5'G, A2'p5'U, A2'p5'C or A2'p5'G clearly shows that the behaviour of the branched trinucleotides is very similar to these of (2' → 5') phosphodiester linked dimers. Therefore a stacking between the adenine moiety and the base of the 2'-substituent seems to dictate the predominant conformational state of the branched trinucleotides. The H2A experiences an upfield shift upon a decrease of temperature from 80 ° to 7 °C of ca. 0.16 ppm in 3, ca. 0.19 ppm in 4 and ca. 0.25 ppm in 5, showing a varied degree of ring current effect from different 2' → 5' linked nucleobase (G> C >U). On the other hand, H1'G (3'-substituent) in 3 and 4 goes downfield upon a temperature change from 80 ° to 7 °C which is not found in A3'p5'G where H1'G goes upfield by ca. 0.08 ppm. This discrepancy can be rationalized by an absence of a stacking between A and G in A3'p5'G part of 3, 4 and 5. However, the temperature profile (Fig. 4) of H8G (3'-substituent) suggests that a certain amount of stacked form exists (upfield shift of H8G from ca. 35 °C to 7 °C) at low temperature.

An approximate population of pseudorotamers[16] of the sugar residue at various temperature have been estimated from the $J_{1',2'}$ coupling constants using the procedure described by Altona and his coworkers[17]. The percentage of N pseudorotamers at a temperature under consideration has been calculated by the following equation.

$$J^{Exp}_{1',2'} = X_N J^{N}_{1',2'} + (1-X_N) J^{S}_{1',2'}$$

where X_N is the mole fraction of the N-type conformers, $J^{N}_{1',2'}$, $J^{S}_{1',2'}$ and $J^{Exp}_{1',2'}$ represent the couplings between H1' and H2' for a pure N-type, S-type and for the compound under consideration, respectively. We have used general conditions[18] for pseudorotational analysis of riboses of 3'-guanosine residue ($J^{N}_{1',2'}$ = 1 Hz and $J^{S}_{1',2'}$ = 7.8 Hz). In compounds 3, 4 and 5, $J^{N}_{1',2'}$ = 1.1 Hz and $J^{S}_{1',2'}$ = 8.1 Hz have been used for the A(2' → 5')X part since Doornboss et al.[13] have found these values in 2'→ 5' linked dinucleotides. The results in the table 3 show that in 3 and 4 the adenosine residue is pure S-type and the pyrimidine residue is pure N-type while 3'pG moiety has a conformation of ca. 40% N. These features are closely comparable to those found in the dimer A2'p5'U[13] and A2'p5'C[12]. Similarly, the adopted conformation of the different sugar moieties in compound 5 is almost identical to those found in A2'p5'G[11]. This conformational analysis clearly confirms that the secondary structure of the branched trinucleotides 3, 4 and 5 closely mimic stacked states which are present in the corresponding (2'→ 5') dimers and the 3'pG is free apart. This is also corroborated by the fact that the conformational states of the sugar moieties in compounds 3, 4 and 5 are also almost similar to the corresponding adenine (2'→ 5') pyrimidine/guanine. It therefore emerges from this study that the secondary structures of these three branched

Table 3. Distribution of pseudorotamers[a] of three constituent sugar residues in branched trinucleotides 3, 4 and 5.

Compound	T°C	$J_{1'2'}$[b](pAp)	%N	$J_{1',2'}$[b](2'pX)[c]	%N	$J_{1',2'}$[b](3'pG)	%N
	7	7.6	7	2.2	84	5.1	40
3	21	7.5	8	2.3	83	5.1	40
	80	6.9	17	3.8	61	5.5	34
	7	7.8	4	0.7	100	5.0	41
4	21	7.6	7	0.9	100	5.6	32
	80	7.1	14	2.4	81	5.7	31
	7	6.8	18	4.9	45	5.7	31
5	21	6.5	23	5	44	5.6	32
	80	6.3	26	5.3	40	5.9	28

[a]see discussion part, [b]in Hz, estimated accuracy ±0.3 Hz, [c]X = U for 3, X = C for 4, X = G for 5.

trinucleotides 3, 4 and 5 closely resemble to those we found[11] for the naturally occurring branched trinucleotides, $A^{2'p5'G}_{3'p5'U}$ and $A^{2'p5'G}_{3'p5'C}$. Recent theoretical studies of the conformation of (2' → 5') polynucleotides[19] have shown a striking correlation between the glycosyl rotation and C2'-O2' torsion which are not present in (3' → 5') polynucleotides. A comparison of structural features between (2' → 5') and (3' → 5') polynucleotides also shows that C2'-O2' torsion and the attached 2'-phosphate, in the former, are distantly nearer to the heterocyclic base which limit the rotational freedom of the (2' → 5') chain quite significantly compared to the (3' → 5') chain. These observations are in complete agreement with our studies showing that the effect of the (3' → 5') nucleotide substituent on the (2' → 5') stacked state in the branched trinucleotides 1 to 5 is almost negligible. We anticipate that the exact 2' → 5' stacking energies in these branched trinucleotides 1 to 5 will depend upon the constituent 2' → 5' nucleobase.

Since our present study, in conjunction with our earlier observations[11], clearly show that the stacking between 2' → 5' linked nucleobases are overwhelmingly more preferred to the competing 3' → 5' linked bases in the branched trinucleotides, 1 to 5, it is, therefore, conceivable that the overall conformation of trinucleotide arise due to the free energy minimum state of the preferred 2' → 5' stacking. Recent studies on sequence requirements in nuclear mRNA splicing has shown, by detailed mutation experiments *in vivo*[7-10], that the formation of lariat is possible with all four nucleobases, but guanine as the 2' → 5' nucleotide is aboslutely necessary for the second step of splicing reaction (i.e., the scission of the 3'-exon and ligation of 5'- and 3'-exons). These studies furthermore indicate that the steric and precise energy requirements for the formation of 2' → 5' branched intron (lariat) may be quite independent of the scission of the 3'-exon and its subsequent ligation to the 5'-exon. It is conceivable that an intramolecular conformational transition of the 2' → 5' branched intron, with the 3'-exon attached, from a higher energy state (3' → 5' stacking) to a lower energy state (2' → 5' stacking) is necessary to drive the splicing reaction to completion (conformation-driven-energy-pump[11]). Since all four nucleosides can form lariats in mutation experiments *in vivo*[7-10], it is, therefore, probable that the formation of the 2' → 5' linked branched pre-mRNA with anyone of the four nucleosides, in general, constitute a pathway to adopt a free energy minimum state. It is guanine, however, as the 2' → 5' branching nucleotide perhaps gives an ideal free-energy minimum 2' → 5' stacked state which generates sufficient free energy of activation for the cleavage of the 3'-exon in the second step of splicing.

Work is now in progress to deduce the exact thermodynamic parameters of different stacking energies of dinucleotides and branched tri- and tetranucleotides by 500 MHz ^{1}H-NMR spectroscopy.

EXPERIMENTAL

NMR samples

Detailed synthetic strategy for the preparation of branched RNA 1 and 2 have been reported[20]. A similar approach[20] has been used for branched trinucleotides 3 and 4. Compound 5 has been prepared using a procedure developed by Ogilvie and his coworkers[14]. The samples were treated with Dowex Na^+ exchange resin and lyophilized two times from 99.8% ^2H_2O and then co-evaporated in 99.9% ^2H_2O. Finally the samples (ca. 5 mg) were dissolved in 0.6 ml of 99.9% ^2H_2O. The pH was found to be 6.5. A trace of dry acetonitrile was added to the samples as an internal reference (set at 2 ppm).

NMR spectroscopy

Spectra were obtained on a Jeol GX 270 spectrometer. 1D spectra were recorded on 16 K datapoints and FIDs were zero-filled to 32 K datapoints before Fourier transformation with a spectral range of 2000 Hz, the digital resolution was 0.12 Hz.

The SECSY spectrum was recorded by the basic pulse sequence[21]. 256 FIDs consisting of 1 K datapoints were recorded, a sine bell window was applied in both directions. Before Fourier transformation the matrix was zero-filled to 512 points in the F_1 direction.

The $^{31}P/^1H$ shift correlation was recording using the method described by A. Otter and coworkers[22]. The spectral range was 250 Hz for ^{31}P direction (F_2) and 2000 Hz for ^1H direction (F_1). The delay 1/2J and 1/4J were set at 58 ms and at 29 ms, respectively. A 512 x 256 data set was collected and zero-filled to 512 x 1024 complex points.

ACKNOWLEDGEMENTS

Authors thank Swedish Board for Technical Development (STU) and Swedish Natural Science Research Council (NFR) for generous financial support, Wallenbergstiftelsen for the funds for the purchase of 270 MHz NMR and Ingegärd Schiller for secretarial assistance.

REFERENCES

1. N. Hernandez and W. Keller, Cell, 35, 89 (1983).
2. P.J. Grabowski, R.A. Padgett and P.A. Sharp, Cell, 37, 415 (1984).
3. M.M. Konarska, P.J. Grabowski, R.A. Padgett and P.A. Sharp, Nature, 313, 552 (1985).
4. A. Krämer, W. Keller, B. Appel and R. Lührmann, Cell, 299 (1984).
5. B. Ruskin, a.R. Krainer, T. Maniatis and M.R. Green, Cell, 38, 317 (1984).
6. B. ruskin and M.R. Green, Science, 229, 135 (1985).
7. M. Aebi, H. Hornig, R.A. Padgett, J. Reiser and C. Weissmann, Cell, 47, 555 (1986).
8. H. Hornig, M. Aebi and C. Weissmann, Nature, 324, 589 (1986).
9. G.A. Freyer, J. Arenas, K.K. Perkins, H.M. Furneaux, L. Pick, B. Young, R.J. Roberts and J. Hurwitz, J. Biol. Chem., 262, 4267 (1987).
10. M. Aebi, H. Hornig and C. Weissmann, Cell, (in press).
11. G. Remaud, J-M. Vial, A. Nyilas, N. Balgobin and J. Chattopadhyaya, Tetrahedron, 43, 947 (1987).
12. P.O.P. Ts'O, N.S. Kondo, M.P. Schweizer and D.P. Hollis, Biochemistry, 8, 997 (1969).
13. J. Doornbos, J.A.J. Den Hartog, J.H. van Boom and C. Altona, Eur. J. Biochem., 116, 403 (1981).,
14. M.J. Damha, R.T. Pon and K.K. Ogilvie, Tetrahedron Lett., 26, 4839 (1986).
15. A. Pullman and B. Pullman, Quarterly Reviews of Biophysics, 14, 289 (1981).
16. C. Altona and M. Sandaralingam, J. Am. Chem. Soc., 94, 8205 (1972) and ibid, 95, 2333 (1973).
17. C.A.G. Haasnoot, F.A.A.M. de Leeuw and C. Altona, Tetrahedron, 36, 2783 (1980).
18. F.A.A. de Leeuw and C. Altona, J. Chem. Soc. Perkin II, 375 (1982).
19. A.R. Srinirasan and W.K. Olson, Nucleic Acid Res., 14, 5461 (1986).
20. J-M. Vial, N. Balgobin, G. Remaud, A. Nyilas and J. Chattopadhyaya, Nucleosides and Nucleotides 6 (1,2), 209 (1987).
21. G.A. Morris, Magn. Reson. Chem., 24, 371 (1986).
22. A. Otter, J.W. Lown and G. Kotovych, Magn. Reson. Chem., 24, 251 (1986).