Tetrahedron Vol. 44, No. 2, pp. 571 to 589, 1988
Printed in Great Britain.

0040-4020/88 $3.00 + .00

270 MHz ^{1}H-NMR STUDIES OF FOUR "BRANCHED" TETRARIBONUCLEOTIDES: A3′p5′A(2′p5′G)3′p5′U , A3′p5′A(2′p5′G)3′p5′C , U3′p5′A(2′p5′G)3′p5′U & U3′p5′A(2′p5′G)3′p5′C WHICH ARE FORMED AS THE LARIAT BRANCH-POINT IN THE PRE-mRNA PROCESSING REACTIONS (SPLICING).

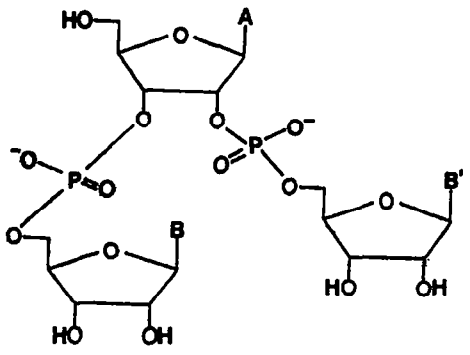
X-X. Zhou[1], A. Nyilas[1,2], G. Remaud[1] and J. Chattopadhyaya[1*]

[1]Department of Bioorganic Chemistry, Box 581, Biomedical Center, University of Uppsala, S-751 23 Uppsala, Sweden
[2]Center for Agricultural Biotechnology, 2101 Gödöllö, Hungary

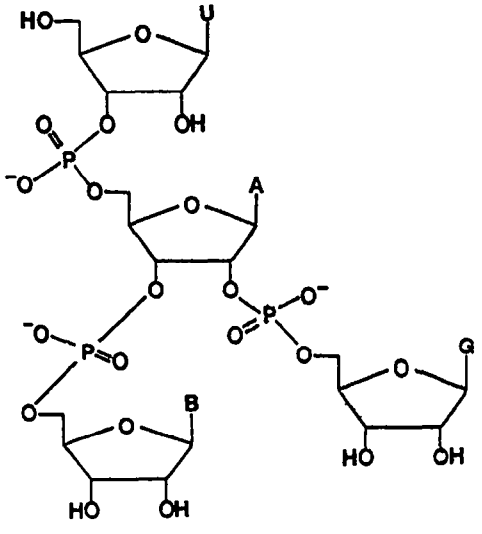
(*Received in UK 9 November* 1987)

Summary: ^{1}H-NMR conformational studies of four title branched ribonucleotides, formed as a local structure in the splicing intron, have been carried out by assigning ^{1}H resonances by 2D NMR such as triple Relay-COSY and ^{31}P/^{1}H correlation experiments and then observing the temperature-dependent (7 - 80 °C) (i) chemical shifts of aromatic and sugar protons, and (ii) shifts of N ⇌ S equilibrium of different sugar moieties. Detailed 2D NOE (NOESY) experiments have subsequently furnished informations regarding the spatial proximities of different protons in these molecules. These studies suggest a coherent conformational picture of the four title branched tetraribonucleotides: (a) the branch-point adenosine (A*) and its 3′-terminal pyrimidine residue are N-N stacked; (b) the A* is coplanar with its 2′-terminal guanosine residue; (c) the nucleotide (A or U) at the 5′-terminus of the A* is S-S stacked with the 2′-terminal guanosine residue and it is also S-N stacked to the A*; (d) the glycosidic bonds of the 3′-terminal pyrimidine residue (C or U) and 5′-terminal residue (A) are in *anti* and *syn* conformations respectively, whereas the A* and the 2′-terminal guanosine residues have approximately an equal population of *syn* and *anti* conformers. These conformational features of branched tetraribonucleotides are reminiscent of a distorted linear 3′→ 5′ linked single stranded A-RNA helix fragment.

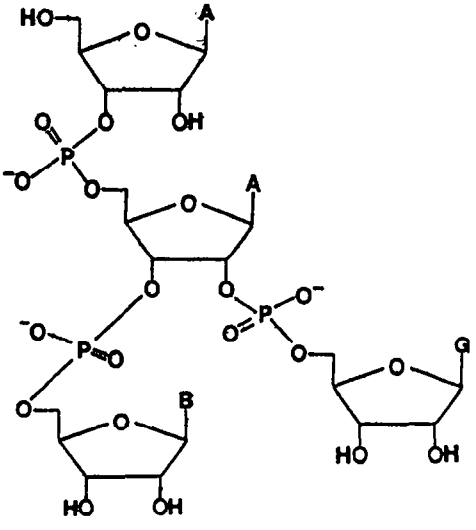
Sequences of nucleotides in the DNA encoding a protein in a eukaryotic cell is not arranged continously, the coding sequences (exons) are interrupted by large stretches of noncoding DNA (introns). After the RNA is transcribed from DNA by RNA polymerase, it is the specificity of the excision of correct intron and ligation of exons in a pre-mRNA processing reaction (splicing) that determines the specificity of amino acids in a protein and its subsequent biological activity[2]. Three different types of splicing reactions[3] have been so far identified: (a) group I type, (b) group II type (both are self-splicing reactions), and (c) nuclear mRNA precursor processes. In contrast to the group I type self-splicing reaction, the group II and nuclear mRNA introns produce branched RNAs (lariat) with the terminal guanosine residue linked through a 2′→ 5′ phosphodiester bond to a branch-point adenosine residue within the intron. Group I splicing is an intermolecular transesterification reaction requiring only guanosine as a cofactor, which binds and reacts in a catalytic pocket consisting of conserved RNA sequences within the intron (ribozyme) to self-splice, whereas the self-splicing of group II introns



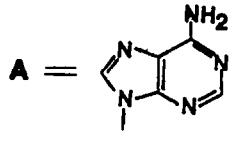
1 B = C, B′ = G
2 B = U, B′ = G
3 B = G, B′ = U
4 B = G, B′ = C
5 B = G, B′ = G

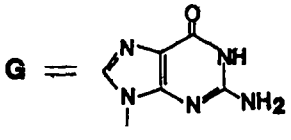


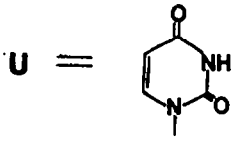
6 B = C
7 B = U

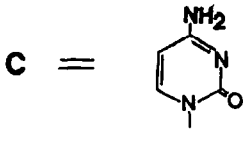


8 B = C
9 B = U









undergoes an intramolecular transesterification reaction by the nucleophilic attack of 2′-hydroxyl group of the adenosine at the branch-point, producing a lariat[3]. However, all three types of splicing reactions show striking parallels: (a) the first step is the cleavage at the 5′-splice site, (b) phosphate moiety at the splicing sites are conserved, (c) short conserved sequence elements for intron folding plays an important role in the alignment of splice junctions. These similarities strongly suggest that the splicing process of nuclear mRNA precursors is closely related to those of RNA-catalyzed self-splicing reactions[3]. Several questions are central[2,3] to the study of RNA splicing: how the sequence information in an intron specifies a particular set of phosphodiester bonds to be cleaved and ligated? Why branched RNAs are formed in group II and nuclear mRNA splicing but not in group I splicing? Does the free energy of activation for transesterification reactions in self-splicing reactions come from intron folding? How the 5′-exon recognizes the 3′-splice site for a nucleophilic attack by its 3′-hydroxyl group to ligate itself to a 3′-exon? What is the role of Mg^{2+} ion in holding and positioning the 3′-OH of the 5′-exon as a nucleophile to the rest of the molecule (intron - 3′-exon)? Why guanosine plays such an important role to drive such precise transesterificaion reactions in the process of splicing?

We have recently examined the conformations of two naturally-occurring branched trinucleotides[4,5] 1 and 2 and their analogues[6] 3, 4 and 5 in order to understand if there is any conformational constrain or preference in 1 and 2, over their analogues 3 - 5, which act as a positive signal for rapid and sepcific excision of introns and ligation of exons in group II and nuclear pre-mRNA splicing reactions. These studies have shown[5,6] that it is the 2′→ 5′ stacking that controls the overall conformation of branched trinucleotides 1 - 5 while the 3′→ 5′ linked nucleobase residue is nearly coplanar with the branch-point adenosine residue (A*). It also emerged through these studies[7] that the adenine(2′→ 5′)pyrimidine stacking in unnatural branched trinucleotides 3 and 4 was more pronounced (free energy minimum) than the adenine(2′→ 5′)guanine stacking in naturally-occurring branched trinucleotides 1 and 2. The latter study[7] indicated that the 2′→ 5′ phosphodiester linked branching was important to serve as a conformational distortion point by seriously altering the local geometry of a well-stacked helical A- or A′-RNA conformation. Such drastic local change in mRNA geometry at the 3′-splice junctions (ca. 10 base upstream) may itself serve as a signal for the later course of the reaction. The specific choice of a guanine residue over a pyrimidine as a nucleobase makes also the branching-site conformationally less rigid and reasonably flexible for a 2′→ 5′ phosphodiester linked branched molecule during the later course of the splicing reaction. At this point, we addressed to the conformational influence of an additional 5′-terminal nucleotide on the overall branched structures of 1 and 2. A comparison of consensus sequences at the branch-point revealed[8] that the 5′-terminal nucleotide adjacent to the 5′-end of the branch-point is a uridine (U) residue in a group II splicing intron while it is an adenosine (A) residue in nuclear mRNA splicing intron. We have therefore synthesized[9] four branched tetraribonucleotides, 6, 7, 8 and 9, with both U and A as the 5′-terminal nucleotide next to A*. We reasoned that an understanding of the effect of 5′-terminal nucleotide unit(s) on the conformation of branched trinucleotide core may allow us to assess how the sequence specificities in an intron dictate a particular set of phosphodiester bonds to be cleaved and ligated in a splicing reaction!

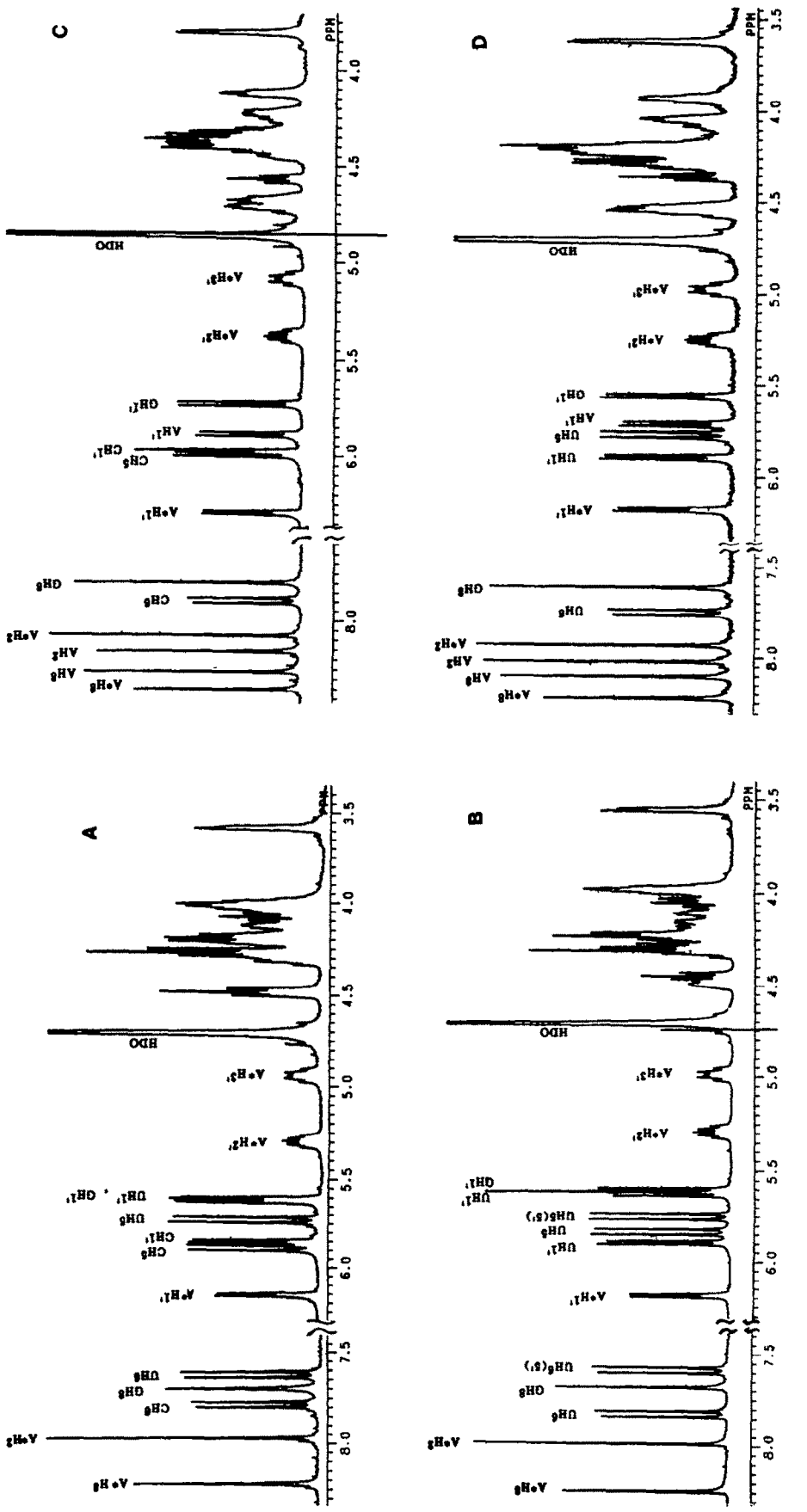


Fig. 1: 270 MHz ^{1}H-NMR spectra of branched ribonucleotides in ^2H_2O at 298 K.
(A) compound 6 ; (B) compound 7 ; (C) compound 8 ; (D) compound 9.

Assignment of resonances

Fig. 1 shows the 270 MHz ^{1}H-NMR spectra of branched tetranucleotides 6, 7, 8 and 9 in ^2H_2O. A complete assignment of all ^{1}H chemical shifts have not been possible because of strong overlaps of absorptions at the low frequency region at 270 MHz. Nevertheless, the aromatic and anomeric protons of all four tetranucleotides 6 - 9 could be unequivocally assigned on the basis of incremental analysis by comparison with either branched or linear (3′→ 5′) parent trinucleotides. Such studies have clearly shown the effect of the addition of either a 2′-substituent, 5′-pG, in the linear 3′→ 5′ trinucleotide or a 5′-substituent, 3′-Up and 3′-Ap, at the 5′-end of the branched trinucleotide core. We have previously shown[9] that the 2D double quantum filtered COSY spectra of the branched tetranucleotides have indeed led to their structure determinations unequivocally. We herein show the use of the triple Relay-COSY experiments in assignments of complex sugar-protons of oligoribonucleotides. Such triple Relay-COSY experiments are an extension of the double Relay-COSY proposed by Bax and coworker[10]. Thus, through three relays, subspectra can be obtained showing the J network from H1′ to H5′/5′′ (shown, for example, in fig. 2) The temperature dependent spectra, from ~5 °C to ~80 °C, have been also used for assignment purposes. The ^{1}H chemical shifts of aromatic and anomeric protons are shown in table 1.

(i) Resonances of aromatic proto...

We envisaged that the simplest case for the assignment of the proton absorptions amongst the tetranucleotides would be to start with compound 6 since it is constituted by four different nucleobases and, therefore, the aromatic protons of each of the constituent nucleobase would have their characteristic chemical shifts and separated from each other. Thus H6 and H5 of the pyrimidine moieties appear as a doublet of ca. 8 Hz for uracil and ca. 7.5 Hz for cytosine, these H5, H6 coupling constants are independent of any temperature change. A distinction between H5 and H6 of 3′-Up and 5′-pC in 6 has therefore been easily made without ambiguity. The assignments of the purine nucleobases in 6 have been achieved by comparing the chemical shifts of H8G, H8A and H2A of branched trimers 1 and 2 with those of H8G, H8A and H2A of branched tetramers 6 - 9. During our earlier works[5,6], we have found a general trend that the aromatic singlets of adenine at the branch-point and guanine residues in the branched trinucleotides are shielded in the following order: H8G > H2A > H8A. We have also found 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 confirming their respective assignments[5,6]. Such unambiguous assignments of aromatic protons in compound 6 have allowed us to observe changes in chemical shifts, in particular, due to substitution of nucleobases in tetranucleotides 7, 8 and 9. In addition, it has been found in the present work that the reported ^{1}H chemical shifts of branched trimers 1 - 5 are affected by the addition of a nucleobase at 5′-end of adenine at the branch point ($p5'A*^{2'p}_{3'p}$). This has been specially useful in the assignments of aromatic protons in compounds 8 and 9 where ambiguities arise due to the presence of two adenine residues. We have found that the H8 of the branch-point adenine, $p5'A*^{2'p}_{3'p}$, is the most deshielded signal while its H2 is normally shielded. Interestingly, the H8 and H2 of the 5′-terminal adenine residue, 3′-Ap, absorb between the H8 and H2 resonances of the branch-point adenine moiety which have been confirmed by 2D NOE experiments (vide infra).

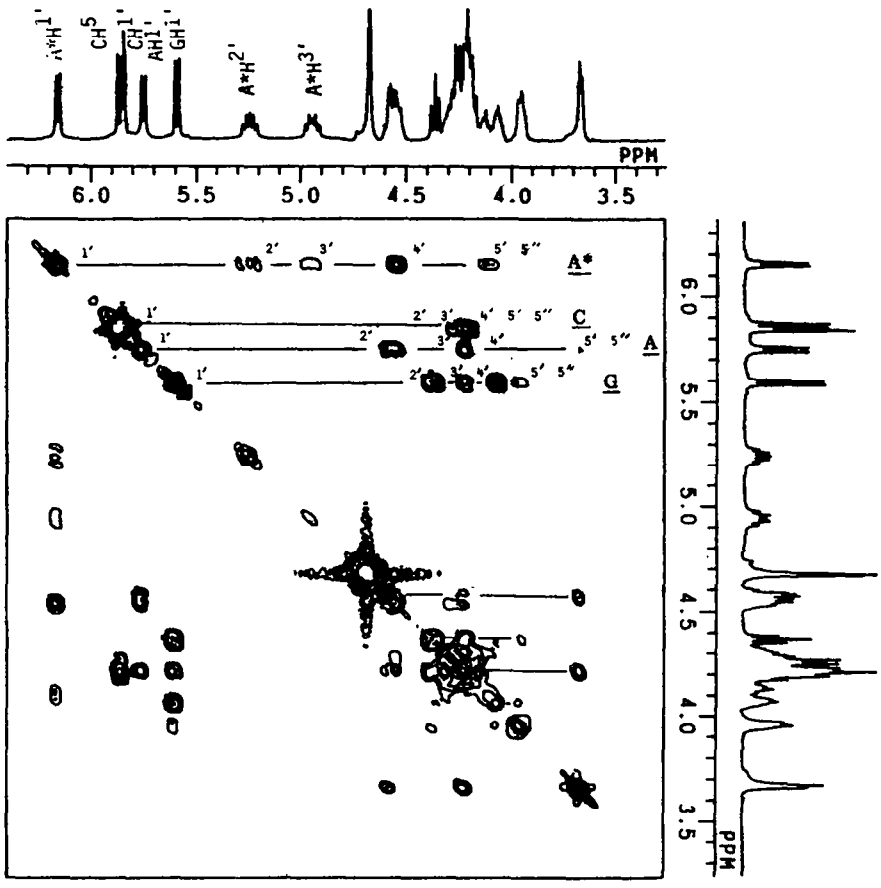


Fig.2: 270 MHz triple Relay-COSY spectrum of compound 8 at 30 °C. The J network of each sugar moiety is shown by solid lines.

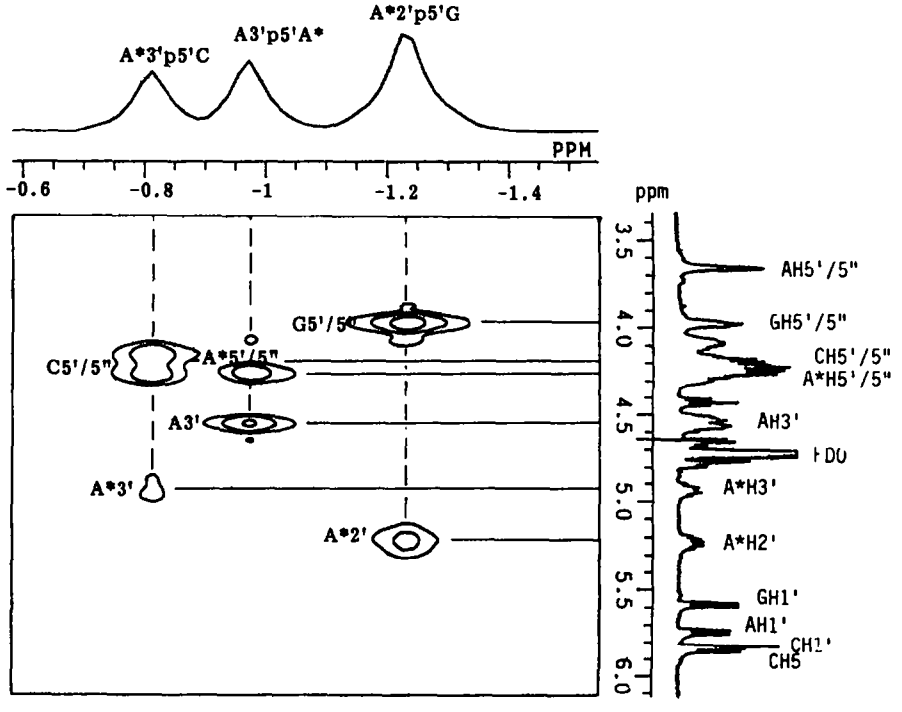


Fig.3: $^{31}P/^{1}H$ correlation spectrum of compound 8 at 25 °C. The projection is represented for the ^{31}P-NMR spectrum.

(ii) Resonances of the anomeric protons.

The H1'A* in branched tetranucleotides, 6, 7, 8 and 9, is the most downfield doublet amongst anomeric protons which has been confirmed by 2D experiments such as DQCOSY[9] (double quantum filtered COSY) and triple Relay-COSY. Fig. 2 shows for compound 8, as an example, that there is a connectivity between H1'A*, H2'A* and H3'A* which are conveniently assigned due to their downfield shifts and characteristic $^{31}P,^{1}H$ couplings. An incremental assignment of tetranucleotides (6, 7, 8 and 9) by comparison with those of the branched trinucleotides 1 and 2 has shown that the H1' of 3'→ 5' linked uridine residue at the branch-point is more deshielded than the anomeric protons of 5'-pG, 3'-Ap or 3'-Up in compounds 7 and 9. A characteristic feature of 3'→ 5' linked cytidine residue at the branch-point in compound 6 and 8, in comparison with the corresponding uridine derivatives (i.e. tetranucleotides 7 and 9), is that the H5 of cytosine moiety is more shielded than the H5 of the corresponding uracil residue. The doublets of cytosine or uracil moieties at times overlap with the anomeric protons but, as mentioned above, they are easily distinguished due to their characteristic coupling constants.

The 2'→ 5' phosphodiester phosphate is more shielded[5,6] and experiences spin-spin couplings with H5'/5'' of guanosine residue as shown in the characteristic $^{31}P/^{1}H$ correlation spectrum for compound 8, shown in fig. 3, as an example. A remarkable feature is that H5' and H5'' of guanosine have an almost identical chemical shift in tetramers 6 - 9. This is in sharp contrast to those found[5,6] in the branched trimers 1, 2, 3, 4 and 5, in which 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. This denotes that the 2'→ 5' phosphate backbone conformation of branched tetranucleotides are quite different from those found in branched trinucleotides.

RESULTS

(i) Effect of temperature on ^{1}H chemical shifts in branched tetranucleotides 6 - 9.

The study of the chemical shifts of the aromatic and anomeric protons as a function of the temperature is used as an indication of stackings between nucleobases. In a stacked form, the aromatic (specially H2A, H5U and H5C) and anomeric protons experience an upfield shift due to the ring-current effect of the neighbouring nucleobases[11]. On the other hand, H8A, H8G, H6U or H6C chemical shifts show a dependence on the glycosidic torsion angle. The variation of ^{1}H chemical shifts of anomeric and aromatic protons from 7 °- 80 °C in different nucleoside residues of branched tetranucleotides 6, 7, 8 and 9 are shown in fig. 4 and are summarized in table 1. Some common features emerge from these data: H8A* moves upfield for 6, 7, 8 and 9 when the temperature decreases from 80 ° to 7 °C. It may be recalled that in our earlier work[5] with the branched trinucleotides 1 and 2, we did not notice any change in the chemical shift of H8A*. Therefore, the observed shifts for H8A* in branched tetramers 6 - 9 are probably due to the nucleobase at the 5'-terminus of branch-point adenosine. H2A* is shifted by ca. 0.13 ppm (0.2 ppm in branched trimers), H1'A* by ca. 0.04 ppm in 6 and 8 and ca. 0.02 ppm in 7 and 9, establishing a stronger ring-current effect of C than U. H6U of 3'→ 5' linked pyrimidine in 7 and 9 experiences a downfield shift by ca. 0.09 ppm while in the branched trimers 1 and 2 the shift is only ca. 0.05 ppm over the same range of temperature.

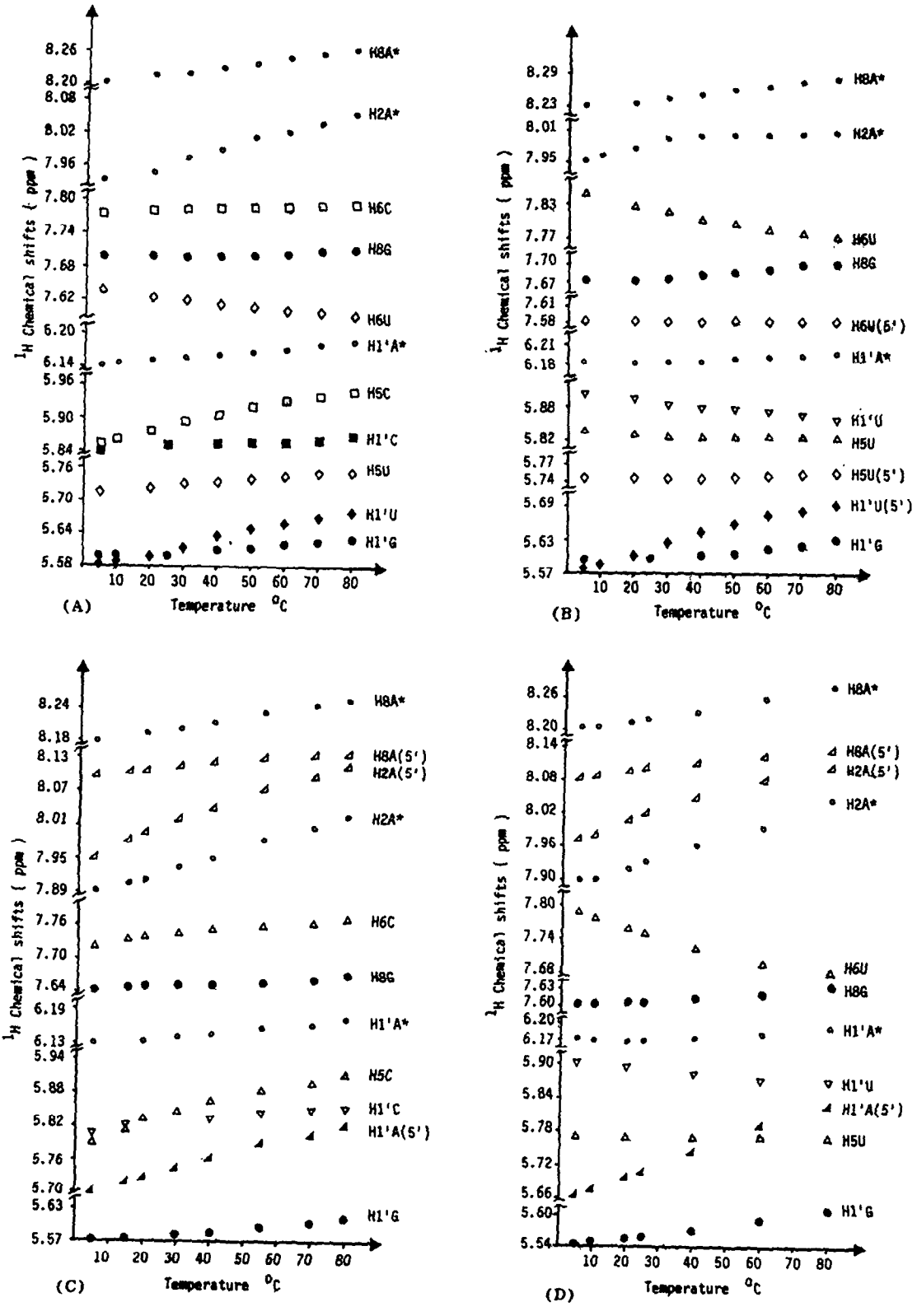


Fig. 4: Temperature dependent variation of ^{1}H chemical shifts of aromatic and anomeric protons of (A) compound 6 , (B) compound 7 , (C) compound 8 , (D) compound 9 .

Table 1: ^{1}H chemical shifts[a] for 6, 7, 8 and 9 as a function of the temperature.

Compound	Fragment	Proton	7 °C	20 °C	40 °C	80 °C
U3'p5'A*2'p5'G 3'p5'C (6)	3'-Up	H6	7.636	7.626	7.614	7.597
		H5	5.713	5.726	5.735	5.751
		H1'	5.585	5.615	5.637	5.633
	A*2'p 3'p	H2	7.936	7.958	7.991	8.060
		H8	8.202	8.213	8.230	8.266
		H1'	6.143	6.152	6.163	6.188
	5'-pC	H6	7.773	7.780	7.785	7.795
		H5	5.851	5.880	5.908	5.952
		H1'	5.843	5.849	5.855	5.870
	5'-pG	H8	7.698	7.697	7.699	7.716
		H1'	5.604	5.607	5.612	5.633
U3'p5'A*2'p5'G 3'p5'U (7)	3'-Up	H6	7.586	7.587	7.587	7.583
		H5	5.744	5.745	5.747	5.752
		H1'	5.577	5.608	5.645	5.690
	A*2'p 3'p	H2	7.955	7.979	8.013	8.079
		H8	8.231	8.238	8.252	8.281
		H1'	6.184	6.183	6.186	6.200
	5'-pU	H6	7.849	7.829	7.806	7.775
		H5	5.834	5.831	5.827	5.828
		H1'	5.905	5.894	5.880	5.861
	5'-pG	H8	7.679	7.679	6.684	7.708
		H1'	5.596	5.598	5.604	5.630

Compound	Fragment	Proton	7 °C	20 °C	40 °C	80 °C
A3'p5'A*2'p5'G 3'p5'C (8)	3'-Ap	H2	7.950	7.997	8.040	8.102
		H8	8.098	8.109	8.120	8.133
		H1'	5.701	5.727	5.763	5.825
	p5'A*2'p 3'p	H2	7.894	7.918	7.955	8.029
		H8	8.180	8.193	8.217	8.258
		H1'	6.132	6.137	6.148	6.174
	5'-pC	H6	7.727	7.739	7.752	7.771
		H5	5.795	5.828	5.863	5.919
		H1'	5.816	5.836	5.835	5.853
	5'-pG	H8	7.649	7.650	7.655	7.678
		H1'	5.571	5.577	5.588	5.615
A3'p5'A*2'p5'G 3'p5'U (9)	3'-Ap	H2	7.973	8.011	8.050	8.109
		H8	8.082	8.097	8.114	8.135
		H1'	5.665	5.700	5.748	5.826
	p5'A*2'p 3'p	H2	7.901	7.925	7.967	8.049
		H8	8.204	8.216	8.239	8.280
		H1'	6.170	6.168	6.173	6.194
	5'-pU	H6	7.785	7.760	7.733	7.687
		H5	5.771	5.770	5.771	5.770
		H1'	5.903	5.892	5.884	5.876
	5'-pG	H8	7.609	7.610	7.617	7.641
		H1'	5.548	5.557	5.572	5.613

[a] from CH_3CN set at 2.000 ppm. Concentration ~0.015 M in ^2H_2O.

On the other hand, H5U, H6U and H1'U of 7 and 9 are more deshielded compared to the corresponding absorptions in the branched trimer 2. The 3'→ 5' linked branch-point H5U (in 7 and 9) has a steady chemical shift over a temperature range of 7 °- 80 °C whereas H5C (in 6 and 8) moves upfield by ca. 0.12 ppm. In the corresponding branched trimers, they both move upfield by ca. 0.06 ppm[5]. Interestingly, H8G, in all four tetranucleotides, moves only ca. 0.03 ppm upfield over the temperature range studied as compared to its 0.05 ppm upfield shift in the corresponding branched trinucleotides. When 3'-Ap is at the 5'-position of A*, as in compound 8 and 9, the H2 and H8 move upfield by ca. 0.15 ppm and ca. 0.04 ppm, respectively. On the other hand, the H6U of the 5'-terminal uracil residue, in compound 6, goes downfield and it has a steady chemical shift in compound 7 over the temperature range 7 °- 80 °C. These features are closely comparable to those of linear (3'→ 5') linked triribonucleotides: AAPy and PyAPy[12].

(ii) Population of pseudorotamers (N ⇌ S) of sugar residues in branched tetra nucleotides.

An approximate population of pseudorotamers[13] of the sugar moieties at various temperatures have been estimated from the $J_{1',2'}$ coupling constants of each sugar residues of branched tetranucleotides 6 - 9 using the procedure described by Altona and coworkers[14]. The following equation has been employed in order to calculate the percentage of N pseudorotamers at a temperature under consideration:

$$J^{EXP}_{1',2'} = X_N J^N_{1',2'} + (1-X_N) J^S_{1',2'} .$$

X_N is the molar fraction of the N-type conformers. $J^N_{1',2'}$, $J^S_{1',2'}$ and $J^{EXP}_{1',2'}$ represent the coupling constant between the H1' and H2' protons for a pure N-type, S-type and for a given compound, respectively. Using the general conditions for pseudorotational analysis of ribose ($J^N_{1',2'}$ = 1 Hz and $J^S_{1',2'}$ = 7.8 Hz)[15], the percentages of N conformers have been calculated for each nucleoside and are shown in table 2. Thus it emerged that the branch-point p5'A*$^{2'p}_{3'p}$ adopts a predominant N conformation (ca. 65 % at low temperature). The 3'→ 5' linked branch-point pyrimidine residue (U in compounds 7 and 9 and C in compounds 6 and 8) show a varied degree of conformational equilibrium (N ⇌ S) upon temperature changes while the N ⇌ S equilibrium for the sugar residue of the 3'-terminal U in 7 and 9 remains virtually unchanged (45% N) from 7 ° to 80 °C. On the other hand, the pseudorotational equilibrium of the sugar residue of C in 6 and 7 shift from 63% N at low temperature to 50% N at a higher temperature. These features are closely comparable to those of A3'p5'U and A3'p5'C[16,17], which show that the C as a nucleobase participates in stacking more strongly than the U residue. Interestingly, the 5'-substituent, 3'-Up or 3'-Ap, at the branch-point prefers the S conformation (ca. 35 % N) and their N ⇌ S equilibrium remain unchanged at higher temperatures, whereas the G residue as the 2'-substituent at the branch-point prefers also an S conformation (ca. 45% N at lower temperatures to 34% N at higher temperatures). It may be recalled from our earlier study[5] that in branched trinucleotides 1 and 2, the branch-point adenosine-sugar is in predominant S conformation (ca. 70%) and the temperature change (5 °- 80 °C) had a very little effect on its N ⇌ S equilibrium. In a sharp contrast with our observations with trinucleotides 1 and 2, the branch-point adenosine-sugar in 6 - 9 show a temperature-dependent N ⇌ S equilibrium (ca. 75% N at 7 °C and ca. 40% N at 80 °C). This presumably shows that the conformation of tetranucleotides at the branch-point are much more flexible than

Table 2: Variation of $J_{1',2'}$[a] of 6, 7, 8 and 9 and the percentage[b] of N pseudo rotamer as a function of temperature[c].

Compound	T °C	$J_{1',2'}$ 3'-X_p[d]	% N	$J_{1',2'}$ p5'A*2'p 3'p	% N	$J_{1',2'}$ 5'-pG	% N	$J_{1',2'}$ 5'-pY[e]	% N
	7	5.4	35	2.8	74	5.0	41	3.5	63
6	20	5.2	38	3.5	63	5.2	38	3.9	57
	40	5.4	35	4.2	53	5.3	36	4.2	53
	80	5.3	36	5.0	41	5.5	34	4.3	51
	7	5.5	34	3.1	69	4.7	45	4.6	47
7	20	5.7	31	3.8	59	5.1	40	4.7	45
	40	5.5	34	4.6	47	5.4	35	4.4	50
	80	5.3	36	5.5	34	5.5	34	4.7	45
	7	5.1	40	3.3	66	4.8	44	3.5	63
8	20	5.2	38	3.9	57	5.0	41	3.2	64
	40	5.1	40	4.3	51	5.4	35	4.0	56
	80	5.1	40	5.4	35	5.5	34	4.4	50
	7	5.5	34	3.6	62	4.8	44	4.7	45
9	20	5.6	32	4.1	54	5.1	40	4.7	45
	40	5.1	40	4.8	44	5.2	38	4.9	43
	80	5.4	35	5.5	34	5.5	34	4.7	45

[a] in Hz, accuracy estimated at 0.3 Hz; [b] see discussion part; [c] concentration ~0.015 M in ^2H_2O; [d] X = U for 6 and 7, X = A for 8 and 9; [e] Y = C for 6 and 8, Y = U for 7 and 9.

those reported for the branched trinucleotides. Furthermore, our earlier work[6] with unnatural branched trinucleotide analogues 3 and 4 have shown an enhanced 2'→ 5' stacking than the corresponding natural trinucleotides, 1 and 2, which clearly reflect an S conformational lock of the branch-point adenosine sugar (7% N at 7 °C and 17% N at 80 °C). The preferential, temperature-dependent N conformation of the branch-point adenosine sugar in tetranucleotides 6 - 9, in conjunction with ca. 50% reduced shielding of H8G in a temperature-dependent chemical shift study, show a clear lack of 2'→ 5' stacking which was found to be a dominant feature in the secondary structures of the branched trinucleotides 1 - 5. The effect of 5'-terminal U or A residues in tetranucleotides in 6 - 9 has been displayed through the destablization of the 2'→ 5' stacked conformations of parent trinucleotides 1 and 2 such that the 3'→ 5' stacking between the branch-point A* and its 3'-terminal pyrimidine residue U or C (i.e. 3'→ 5' stacking, the stacking being stronger with 3'-terminal C than U) has now become a predominant conformational mode in the overall secondary structures of the tetranucleotides 6 - 9. Major N conformations have been found to exist for all constituent sugar residues in linear (3'→ 5') linked tetraribonucleotides, AACC[18], and in triribonucleotides, AAC[12] and UAU[12]. In these oligoribonucleotides, the conformational equilibrium (N ⇌ S) of all sugar residues varies from pure N at low temperatures to less N at higher temperatures, consistant with the destacking caused by an increasing temperature. In the present study, it is a similar N-N stacking equilibrium that pre-

dominates between the branch-point adenosine and its 3'-terminal pyrimidine moiety. The predominant S conformation and the shift of N ⇆ S equilibrium of 2'-terminal G residue (34% N at low temperatures and ca. 42% N at a higher temperature) in tetranucleotides 6 - 9 (such temperature variation of N ⇆ S equilibrium of the sugar residue of G in branched trinucleotides 1 and 2 also exists!), in conjunction with observations that (i) both H2 and H8 of the branch-point adenosine (A*) residue are deshielded by more than 0.15 ppm as compared to their parent branched trinucleotides, and (ii) the fact that the 5'-terminal nucleotide is in predominant S conformation (ca. 35% N which shows no temperature-dependence N ⇆ S equilibrium) lead us to conclude that there is a stacking between 5'-terminal nucleobase and the 2'-terminal G residue of A*. Appropriate NOE data (*vide infra*) has been subsequently obtained in support of our latter proposal. These NOE data have furthermore shown that the 5'-terminal residue is also partly stacked onto A*, which is perhaps a reminiscence of a distorted single stranded A-RNA helix fragment[19].

In the 3'→ 5' linked tetramer AACC[18] the sugar moiety of each nucleobase has a high percentage of N conformer. The linear trinucleotides AAC and UAU which have a similar sequence as in the branched tetramers 7 and 9, the N conformer is also largely preferred[12,20]. Therefore, the branched tetranucleotides 6 - 9 display an average conformation closely identical to the linear (3'→ 5') linked oligonucleotides, except for the 5'-terminal nucleotide which adopts predominently an S conformation.

(iii) 2D NOE studies with the branched tetranucleotides 6 - 9.

In the present work, 2D NOE studies have provided two types of data in our attempts to understand the preferred conformation of tetranucleotides: (i) determination of the spatial proximities between the nucleotide residues, and (ii) between the nucleobase and its own sugar unit in order to assess the preferred conformation around the glycosidic bond (*Syn* ⇆ *Anti*). Such studies have been used to confirm the presence of aforementioned stackings between nucleobases unequivocally. In order to delineate the strength of the NOE, each 2D NOE experiments have been performed with different mixing times (700 ms and 350 ms with 9 mM solutions of compounds 6 to 9). This procedure has allowed us a qualitative classification of all observed NOE as strong, medium and weak NOE. For example, the NOE between H5 and H6 in pyrimidine residue has been classified as strong while an NOE detected only at a mixing time of 700 ms but disappeared with a mixing time of 350 ms was attributed as a weak NOE. The NOE connectivities, shown in figs. 5 and 6, have been also confirmed at a lower concentration (5 mM) and also with a lower mixing time (100 ms) showing an absence of artifacts in the observed NOEs due to any spin diffusion and/or self-association.

For each branched tetranucleotides two zooms of the 2D NOESY matrix are shown illustrating the NOE connectivities in the aromatic region (Fig. 5a-d) and between the aromatic protons and the upfield protons originating from the sugar moieties (Fig. 6). An examination of 2D NOE data of all four tetranucleotides 6 - 9 show that these NOE data can be schematically divided into schemes 1 and 2. These schemes show spatial proximities of different aromatic and sugar protons in a nucleotide residue and also between internucleotide residues. The schemes 1 and 2 represent a conformational similarity between compounds 6 and 7 [both containing 5'-terminal U residue at the branch-point (A*)] and compounds 8 and 9 [both containing 5'-terminal A at the branch-point (A*)], respectively. Thus it appears

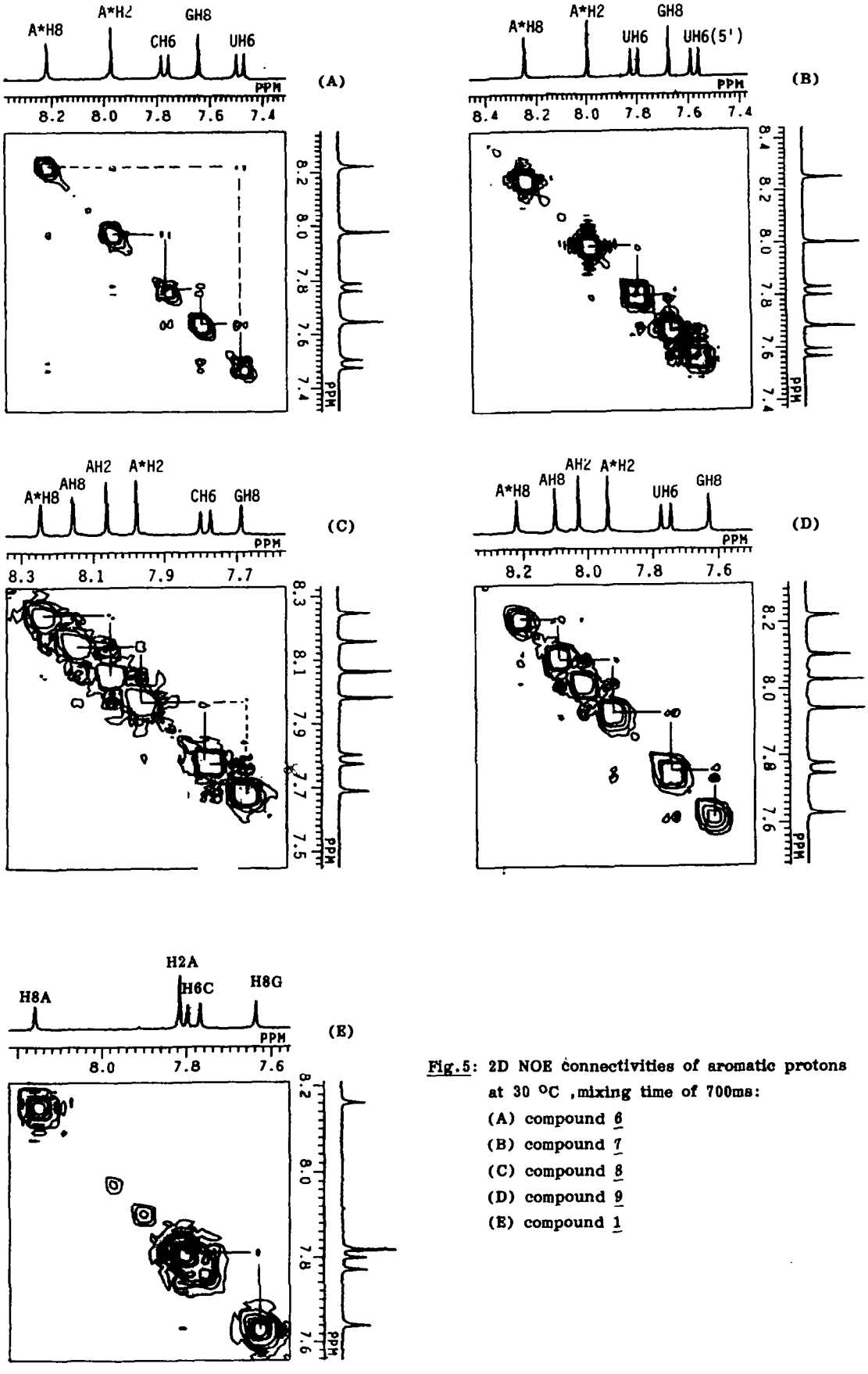


Fig.5: 2D NOE connectivities of aromatic protons at 30 °C, mixing time of 700ms:
(A) compound 6
(B) compound 7
(C) compound 8
(D) compound 9
(E) compound 1

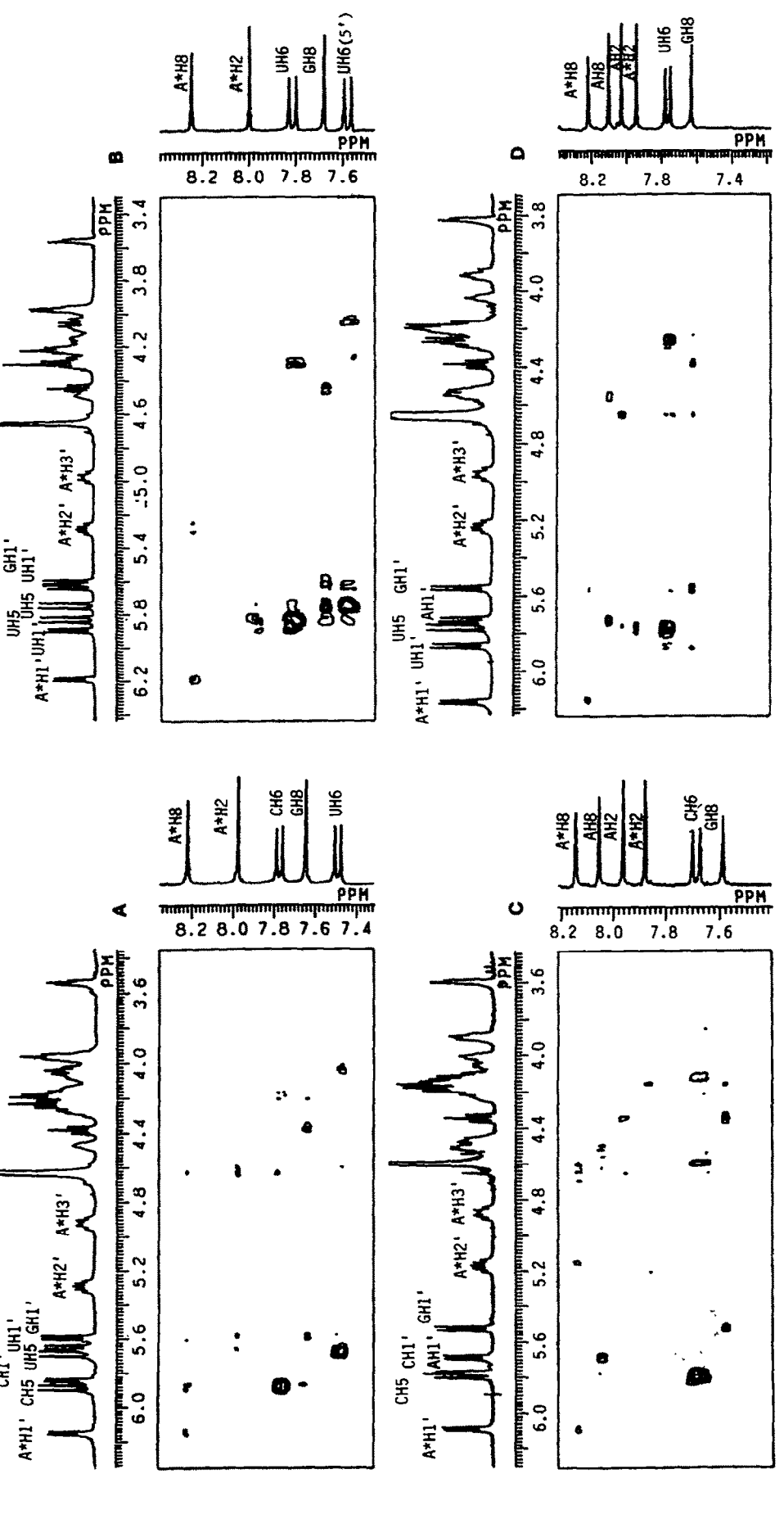


Fig.6: 2D NOE connectivities between aromatic protons and aliphatic protons at 30 ° C, mixing time of 700 ms; (A) compound 6 ,(B) compound 7 , (C) compound 8 , (D) compound 9 .

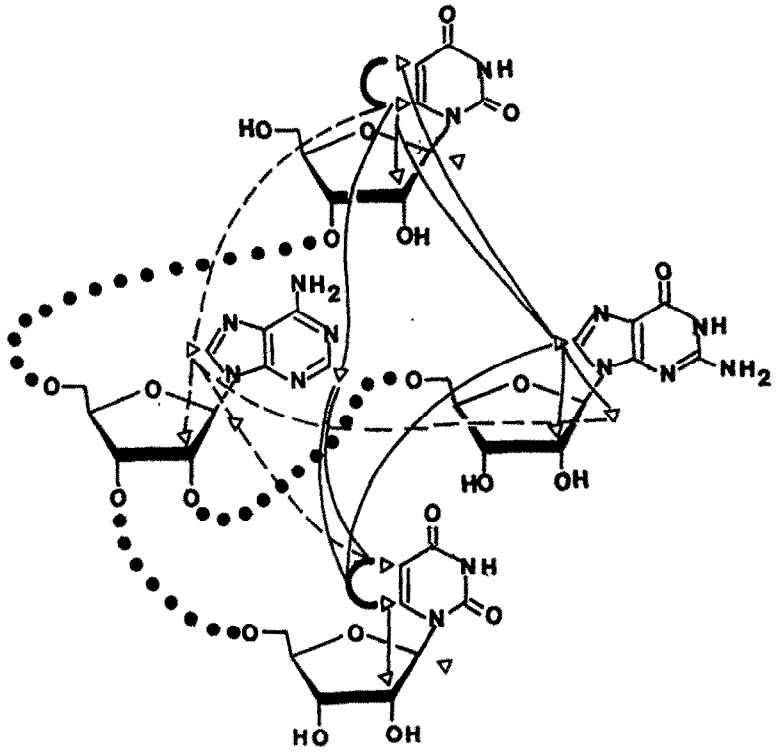
that H6 of the pyrimidine base is close to H2′ of its own sugar. The H8 of the 5′-terminal A (8 and 9) is close to H1′ of its own sugar. We therefore conclude that these two nucleotides adopt an anti and syn conformation respectively. It appears, for A* and G, that there is approximately an equal population of syn and anti conformers since H8 of A* or G has an NOE with its both H1′ and H2′. It is also clear that the H6 of the 3′→ 5′ linked pyrimidine at the branch-point (A*) has always NOEs with both H8G and H2A*. For the compounds 8 and 9, a full NOE connectivities between A and A* is present showing the spatial proximity of these two residues. A similar NOE network between U and A* in compounds 6 and 7 is much poorer, only a neat NOE is detected between H2A* and H5U. On the other hand, the spatial proximity between terminal U and G are reasonably close as seen through relevant NOE in fig. 5a and 5b.

A very important observation is the absence of NOE between the aromatic protons of A* and G suggesting an absence of stacking between these two nucleobases. However, a weak NOE has been detected between H8A* and H1′G which is most probably due to their spatial proximities owing to the coplanarity of these two nucleobases. On the other hand, it should be noted that in the branched trimer $A^{2'p5'G}_{3'p5'C}$ (1), for example, an NOE is detectable between H2A* and H8G (Fig. 5e) confirming the stacking between these two nucleobases.

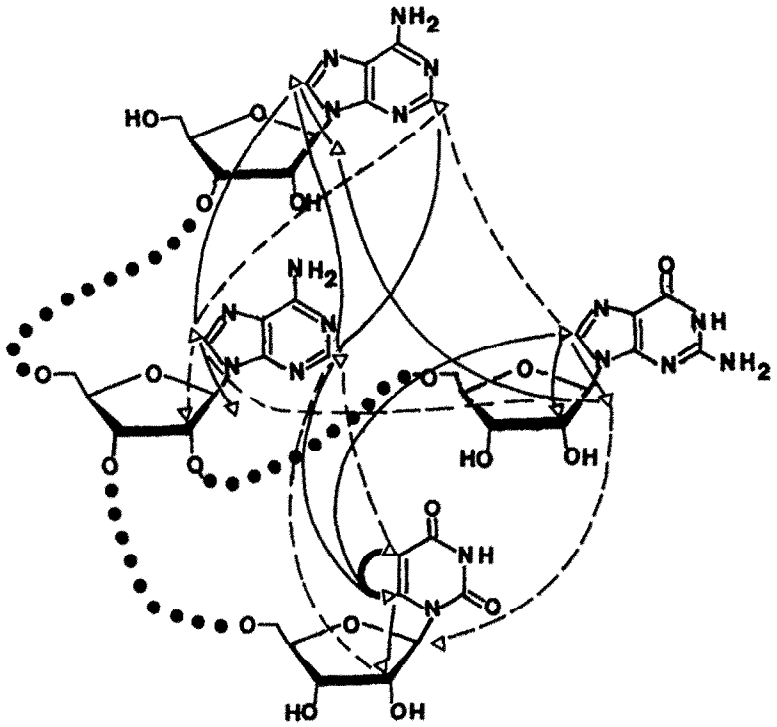
DISCUSSION

Previous studies[5,6] from this laboratory have shown that a 270 MHz ^{1}H-NMR analysis of branched trimers, 1 - 5, can indeed show the overall conformation of these unique compounds. ^{1}H-NMR at 500 MHz has confirmed our initial results by showing that the preferential 2′→ 5′ stacking in naturally-occurring branched trinucleotides 1 and 2 is only in near free energy minimum while in unnatural analogues 3 and 4, the 2′→ 5′ stacked conformation is indeed in a free energy minimum state[7]. The implication of this observation is that natural branched trinucleotides are more conformationally flexible than the unnatural counterpart. The predominant conformational feature of the strong 2′→ 5′ stacking between the branched point A* and the 2′-substituent in these trinucleotides 1 - 5 was associated with a high percentage of S conformer for the sugar moiety of A* while the 3′-substituent was coplanar with A*. These features are clearly absent in the branched tetranucleotides (6, 7, 8 and 9). The differences can be summarized as follows: (a) N conformation for A*, (b) different phosphate back bone in $A^{*2'p5'G}_{3'p}$ linkage, (c) a shielding of H8G reduced by half over a range of temperature 80 ^{o} to 7 ^{o}C, (d) at high temperature, a deshielding of H8A, H2A and H1′A and H8G is observed in compounds 6 - 9 as compared to A2′p5′G or $A^{2'p5'G}_{3'p5'U}$ and $A^{2'p5'G}_{3'p5'C}$ but the extent of these shieldings are close to $A^{2'p5'U}_{3'p5'G}$ or $A^{2'p5'C}_{3'p5'G}$, and (e) no NOE connectivity between A* and G.

Temperature-induced ^{1}H chemical shifts and the N ⇌ S equilibrium of pseudorotamers suggest that (a) the branch-point adenosine (A*) and its 3′-terminal pyrimidine residue are N-N stacked; (b) 5′-terminal nucleotide is considerably S-S stacked with the 2′-terminal guanosine residue at the branch-point; (c) 5′-terminal nucleotide is also partly S-N stacked onto the branch-point adenosine. These conformational features are accomodated by a stacked equilibrium of different natural N-N, S-N and S-S stacked states, reminniscent of a distorted linear (3′→ 5′) linked single stranded A-RNA helix fragment[19].



Scheme 1: NOE connectivities amongst different aromatic-aromatic and aromatic-aliphatic protons for compound 7; compound 6 also shows very similar NOE connectivities.



Scheme 2: NOE connectivities amongst different aromatic-aromatic and aromatic-aliphatic protons for compound 9; compound 8 also shows very similar NOE connectivities.

Legends for Schemes 1 & 2

Phosphodiester function: (● ● ● ● ● ●)

NOE : *strong* - (▬▬▬)

medium - (———)

weak - (— — — —)

A simplified picture of the secondary structures show that a stacking exist between neighbouring bases except between A* and G. The 5'-terminal nucleobase (A or U) stacks upon both A* and G in order to accomodate its predominant S conformation, the 5'-terminal A residue in 8 and 9 is, however, in the syn conformation across its glycosidic bond, which are in sharp contrast to a normal 3'→ 5' linked oligoribonucleotide structure. The nucleobase G is coplanar with A* and, therefore, cannot be stacked, but the paramagnetic effect of the aromatic ring of 5'-pG seems to deshield the protons of A* as in 3, 4, 5 in which 3'→ 5'pG was coplanar with A*[6].

Finally, on the basis of the 2D NOE connectivities, we have distinguished between the average conformations of 6, 7, in one hand, and of 8 and 9, on the other. Thus the NOE connectivities show a network which is more extended between the 5'-terminal 3'-Ap and A* than with the 5'-terminal 3'-Up and A*. This is rationalized by a preferred stacking between purine(3'→ 5')purine (A*) than pyrimidine(3'→5')purine (A*)[21], suggesting that the 5'-terminal 3'-Ap residue in tetranucleotides 8 and 9 stabilizes the short helix in such a way that is more similar to the usual single stranded A-RNA helix than a 5'-terminal 3'-Up in 6 and 7.

It may be noted that one other strong evidence, which has stemmed from the enzymatic work of Edmonds and her coworkers[22], that independently suggests that there is indeed a drastic difference between the conformations of a branched tetranucleotide, G3'p5'$A^{2'p5'G}_{3'p5'C}$, and a trinucleotide, $A^{2'p5'G}_{3'p5'C}$. She noted[23] that the branched core structures, such as 1 and 2, which lack a 5'-nucleotide unit can not be phosphorylated at the 5'-hydroxyl group with T4-polynucleotide kinase while G3'p5'$A^{2'p5'G}_{3'p5'C}$ could be easily labelled using [γ-^{32}P]ATP and T4-polynucleotide kinase. Such a difference in the enzymatic recognitions between a branched tetra- and a trinucleotide is best explained by the fact that the conformation of branched tetranucleotides resembles a distorted linear 3'→ 5' linked single stranded A-RNA helix fragment, which is easily recognized by the enzyme, whereas the predominant conformational feature of branched trinucleotide core is comprised of unnatural stacking between 2'→ 5' linked nucleobases and the 3'-pyrimidine residue is coplanar with the branch-point adenosine. Clearly, additional systematic conformational studies with additional ribonucleotide residues in all three termini of the branch-point A* (i.e., 2'→ 5', 3'→ 5' and 5'→ 3') of branched trinucleotides 1 and 2 are required in order to understand the biological implication of the lariat formation in group II and nuclear mRNA splicing reactions. Further work along this line is in progress in this laboratory.

EXPERIMENTAL

NMR samples

Detailed preparation of branched RNA 6 - 9 have been reported[9]. The samples were purified on DEAE Sephadex column using a linear gradient (0.0 M to 0.5 M) ammonium hydrogen carbonate buffer (pH 7.5, 800 ml each) and the ammonium counterion was replaced with Na^+ by passing the pure tetramers through a Dowex (Na^+ form) column using double-distilled water as an eluent and evaporated to dryness and then co-evaporated twice in 99.8% in ^2H_2O. Finally the samples (ca. 6 mg for NOE studies, ca. 12 mg for temperature dependence and ca. 15 mg for triple Relay-COSY) were dissolved in 0.5 ml of ^2H_2O "100 atom % D" (Aldrich). The pH was found to be ~7.5. A trace of dry acetonitrile was added to the samples as an internal reference (set at 2.000 ppm).

NMR spectroscopy

Spectra were obtained on a Jeol JNM-GX 270 spectrometer equipped with a 5 mm probe head. 1D spectra were recorded using 16 K data points and FIDs were zero-filled to 32 K before Fourier transformation without any broadening factor. With a spectral range of 2000 Hz, the digital resolution was 0.12 Hz.

The triple Relay-COSY has been edited from the double Relay COSY[10]. Phase cyclings (16 steps) were made to minimize the diagonal peaks. Since this sequence involves free precession periods with three consecutive transfers, the symmetry of the 2^n matrix is altered. Therefore, artefacts can arise (peaks on the other diagonal could be seen after symmetrisation). We suggest to put the carrier on one side of the spectrum. We found that the triple Relay-COSY gives the same informations as HOHAHA pulse sequence[24] but the parameters for the former are easier to settle which makes it very convenient for assigning oligoribonucleotides. The water peak was slightly irradiated continously except during the acquisition. The three mixing times were set at 120 ms. 256 FIDs consisting of 512 data points were recorded, a sine bell window was appllied in both directions. Before Fourier transformation the matrix was zero-filled to 512 x 1024 points.

The $^{31}P/^{1}H$ shift correlation was recorded using the method described by A. Otter and coworkers[25]. The spectral range was 500 Hz for ^{31}P direction (F_2) and 2000 Hz for ^{1}H direction (F_1). The delay 1/2J and 1/4J were set at 58 ms and at 29 ms, respectively. A 256 x 512 data set was collected and zero-filled to 512 x 1024 complex points.

The 2D NOE spectra were recorded by the basic pulse sequence[26] where the mixing time τm was systematically changed as t_1 increments in order to minimize the peaks arising from J contributions (see, for example, ref. 27). The water peak was continuously irradiated under low power except during the acquisition time t_2 and τm. 128 FIDs of 512 data points were recorded. Each FID was constituted of 208 scans and the pulse delay was 2s (20 to 24 h for the complete experiment). A sine bell apodization was appllied on a zero-filled 512 x 1024 matrix.

ACKNOWLEDGEMENTS

Authors thank Wallenbergstiftelsen for funds for the purchasse of a 270 MHz NMR spectrometer, Swedish Board for Technical Development and Swedish Natural Science Research Council for generous research grants and Mrs. Ingegärd Schiller for excellent secretarial assistance.

REFERENCES

1. B. Alberts, D. Bray, J. Lewis, M. Raff, K. Roberts, J.D. Watson, The Molecular Biology of the Cell, Garland Publishing, Inc., New York and London, 1983.
2. See for review of field until 1985, R.A. Padgett et al. [Annu. Rev. Biochem., 55, 1119 (1986)]; S.E. Leff, M.G. Rosenfeld, R.M. Evans, Ann. Rev. Biochem., 55, 1091 (1986).
3. T.R. Cech and B.L. Bass, Ann. Rev. Biochem., 55, 599 (1986); W. Keller, Cell, 39, 4223 (1984); T.R. Cech, Cell, 44, 207 (1986); P.A. Sharp, Science, 235, 766 (1987) and references therein; J.V. Price, J. Mol. Biol., 196, 217 (1987); H. Hornig, M. Aebi and C. Weissmann, Nature, 324, 589 (1986) and referances therein.
4. J-M. Vial, N. Balgobin, G. Remaud, A. Nyilas and J. Chattopadhyaya, Nucleosides & Nucleotides, 6, 209 (1987).
5. G. Remaud, J-M. Vial, A. Nyilas, N. Balgobin and J. Chattopadhyaya, Tetrahedron, 43, 947 (1987).
6. J-M. Vial, G. Remaud, N. Balgobin and J. Chattopadhyaya, Tetrahedron, 43, 3997 (1987).
7. L.H. Koole, N. Balgobin, H.M. Buck, W.H.A. Kuijpers, A. Nyilas, G. Remaud, J-M. Vial and J. Chattopadhyaya, J. Am. Chem. Soc., (submitted).
8. R. van der Veen, A.C. Arnberg, G. van der Horst, L. Bonen, H.F. Tabak and L.A. Grivell, Cell, 44 225 (1986).
9. X-X. Zhou, A. Nyilas, G. Remaud and J. Chattopadhyaya, Tetrahedron (in press).
10. A. Bax and G. Drobny, J. Magn. Reson., 61, 306 (1985).
11. A. Pullman and B. Pullman, Quarterly Reviews of Biophysics, 14, 289 (1981).
12. C-H. Lee and I. Tinoco Ir., Biophys Chem., 11, 283 (1980).
13. C. Altona and M. Sandaralingam, J. Am. Chem. Soc., 94, 8205 (1972) and ibid, 95, 2333 (1973).
14. C.A.G. Haasnoot, F.A.A.M. de Leeuw and C. Altona, Tetrahedron, 36, 2783 (1980).
15. F.A.A. de Leeuw and C. Altona, J. Chem. Soc. Perkin II, 375 (1982).

16. P.O.P. Ts'o, N.S. Kondo, M.P. Schweizer and D.P. Hollis, Biochemistry, 8, 997 (1969).
17. C-H. Lee, F.S. Ezra, N.S. Kondo, R.H. Sarma and S.S. Danyluk, Biochemistry, 15, 3627 (1976).
18. J. Doornbos, C.T.J. Wreesmann, J.H. van Boom and C. Altona, Eur. J. Biochem., 131, 571 (1983).
19. C. Altona, Recl. Trav. Chim. Pays-Bas, 101, 413 (1982).
20. C.H. Lee, Eur. J. Biochem., 137, 347 (1983) and idem, 137, 357 (1983).
21. W. Saenger in "Principles of Nucleic Acid Structure", Springer-Verlag, New York, pp. 116-158 (1984).
22. R. Kierzek, D.W. Kopp, M. Edmonds and M.H. Caruthers, Nucleic Acid Res., 14, 4751 (1986).
23. M. Edmonds, personal communication on July 7th, 1987.
24. D.G. Davis and A. Bax, J. Am. Chem. Soc., 107, 2820 (1985).
25. A. Otter, J.W. Lown and G. Kotovych, Magn. Reson. Chem., 24, 251 (1986).
26. G.A. Morris, Magn. Reson. Chem., 24, 371 (1986).
27. G. Wider, S. Macura, A. Kumar, R.R. Ernst and K. Wütrich, J. Magn. Reson., 56, 207 (1984).