Sequence-specific Solution Structures of the Four Isosequential Pairs of Single-stranded DNAs and RNAs

Subhrangsu Chatterjee[‡], Wimal Pathmasiri[‡] & Jyoti Chattopadhyaya*

Department of Bioorganic Chemistry, Box 581 ,Biomedical Center,

Uppsala University, SE-75123 Uppsala, Sweden.

Fax:+46-18554495 E-mail: jyoti@boc.uu.se

ABSTRACT

The role of the sequence-context in the self-organization of four single-stranded (ss) isosequential pairs of DNAs (**1** – **4**) and RNAs (**5** – **8**), [d/r-($^{5'}$C^{1}A^2**X**^{3}G^4**Y**^{5}A^{6}C^7): **X**^{3} = A or C, **Y**^{5} = A or C; sequence variations: 2^2 = 4], has been elucidated by NMR-constrained Molecular Dynamics (MD) simulations (2 ns). Following sequence-specific observations have been made from the solution NMR and the NMR constrained MD simulation study: (i) Analysis of the NOESY footprints, mainly $(H8/H6)_n$ to $(H1'$ and $H3')_{n-1}$ contacts, of ssDNAs (**1** - **4**) and ssRNAs (**5** – **8**) in the aqueous medium have shown that all ssDNAs (**1** - **4**) and ssRNAs (**5** - **8**) adopt right handed stacked helical structures in the NMR time scale. (ii) Intra-residual cross-peak intensities for the $H(8/6)_n$- $H(1'/2'/2''/H3')_n$ contacts in ssDNAs and ssRNAs are stronger at the 3'-ends in comparison with those at the 5'-ends, suggesting that the dynamics of the nucleobases at the 3'-end are more restricted, whereas those at the 5'-end are more flexible. (iii) This relative NMR found mobility is consistent with the final RMSd calculations of the final NMR-MD structures of ssDNAs and ssRNAs. They show that the 5'-end nucleobases have higher RMSd values compared to those at the 3'-end, except for the sequence d/r($^{5'}$C^{1}A^2**A**^{3}**G**^{4}**A**^{5}A^{6}C^7). (iv) Relative nOe intensities of inter-residual $H(8/6)_n$ - $H(1')_{n-1}$ and $H(8/6)_n$ - $H(3')_{n-1}$ contacts, as well as NMR observed fluctuations in the sugar conformations, for ssDNAs (**1** – **4**) and ssRNAs (**5** – **8**) show that no ssDNA or ssRNA adopts either a typical B-type DNA or A-type RNA form. (v) In the final NMR-MD structures all the [$H8/6N_{(n)} \leftrightarrow H1'N_{(n-1)}$/ $H3'N_{(n-1)}$, N = A, G, C] distances in different isosequential pairs of ssDNA (**1** – **4**) and ssRNA (**5** – **8**) change depending upon the sequence context of the single-stranded nucleic acids. Both in the deoxy and ribo series, it is the purine-rich sequences [d/r-($^{5'}$C^{1}A^2**A**^{3}**G**^{4}**A**^{5}A^{6}C^7) which form the most stable self-organized right-handed helical structures because of the favorable purine-purine stacking interactions. (vi) Stacking pattern at each of the dinucleotide steps show that the base-base nearest neighbor stacking interactions depend solely upon the sequence contexts of the respective ssDNAs (**1** – **4**) and ssRNAs (**5** – **8**). See pages 47 – 145 for **Supplementary Information** for detailed spectroscopic data.

Introduction

Single stranded DNA (ssDNA) and RNA (ssRNA) exist in eucaryotic and prokaryotic cells, and are known to have vital functions: While ssDNAs are the intermediates for replications during cell division, ssRNAs is produced from ssDNAs by transcription using RNA polymerase. The triplet code of ssRNA for each amino acid is then translated into proteins in the ribosomal

[‡] Contributed equally

machinery (*1*) through codon–anticodon recognition and interaction with proteins and many other cofactors. On the other hand, small interfering RNA (siRNA) (*2, 3*), and microRNA(*4*) are involved in the down-regulation of genes. The single-stranded 5'-RFN element(*5*) of mRNAs bind specific metabolites and respond by altering their shapes in biologically useful ways, demonstrating that aptamers are also present in the natural world(*6*). The ssRNA can also form different hairpins (*7*) and pseudoknot structures which are actively involved in the down-regulation of several genetic diseases (*8*). ssDNA is known to be involved in protein recognition in DNA-protein interactions (*9*), replication (*1*), and telomere recognition (*10, 11*).

Unpaired terminal nucleotides (dangling nucleotides) are present in different important RNA structures: The 3'-end of tRNA (*12*) is constituted with CCA trinucleotidic sequence which plays a key role in its structure and function. Codon-anticodon interaction between mRNA and tRNA can be stabilized by the dangling bases (*13*) for the synthesis of proteins in the ribosomal machinery (*1*). NMR (*14*) and UV (*15*) spectroscopy have already demonstrated that the duplex stabilization can be altered depending upon the sequence composition of the dangling ends (*13*) .

Secondary DNA and secondary/tertiary RNA structure play key roles in controlling many genetic diseases [expansion of the trinucleotide repeat C*N*G (*N* = A,G) sequences as in neurodegenerative disorders, Huntington's disease (*16*), Leukemia (*17*)]. Earlier investigators have shown that any genetic disorder can be expressed or silenced depending upon the alteration in sequence context of ssDNA and ssRNA. Recently, expansion of a dodecamer repeat, $d(CCCCGCCCCGCG)_n$ upstream of cystatin B gene has shown to be the most common mutation in Progressive Myoclonus Epilepsy (EPM1)(*8*).

ssRNAs can form various tertiary and secondary structures in sequence specific manner (*18*). These types of structure formation involves building of different foldings and scaffolds by specific complementary nucleobases (metal ions as cofactors) through inter or intra-molecular interactions, as in the sequence-specific cleavage activity for group I (*19-22*) and group II introns (*22-25*), RNase P RNA (*26*), HDV ribozyme (*27, 28*), hammerhead ribozyme(*29*), kissing hairpins (*30*), and unnatural allosteric ribozymes (*31*).

Therefore, understanding of the ssDNA/ssRNA structures is important for understanding their functions in the cell. ssDNA/ssRNA are devoid of any base-pairing. Intra and inter residual stacking (offset, face to face, or edge to face nearest-neighbor base-base interactions) (*32, 33*), non-covalent interactions (mediated by the effect of salts and hydration) and dispersion forces play a vital role in the self-organization of these structures.

Low resolution temperature-dependent NMR studies on di-, tri-, tetra-, and penta-nucleotides(*14a-h*), calorimetry(*34*), absorption spectroscopy(*15*) and optical rotatory dispersion spectroscopy(*35*) on ssDNA and ssRNA suggested that the base-base stacking interactions dictate the geometry and function of ssDNA and ssRNA. Differential scanning calorimetric study of

13mer DNA duplex melting showed that the ssDNA sequences (s_1 and s_2 in 13mer DNA duplex) are likely to posses considerable global order that can significantly influence driving forces associated with the duplex formation at 25°C (*34*).

Solution phase hairpin loop structure of RNA shows that sugar residues in the stem of hairpin RNA reside in the North-type (N) conformations(*36*). In contrast, sugar residues in ssRNA structures in loop exist in the South-type (S) conformation (*36-38*). Interestingly, sugars of the trinucleotide sequences in anticodon of tRNA remain in the North-type (*36*). This variability of sugar conformations in different single-stranded forms of RNA leads us to suggest that the structure and sequence context should play an important role in the functions of these nucleic acids, because we understand how the chemical nature of the aglycone (purine versus pyrimidine) dictates the conformational preference of the sugar-phosphate backbone (*39*).

Recently, we have reported that purine-rich hexameric isosequential ssDNA and ssRNA can form self-organized structures through the base-base stacking interactions(*40*). The NMR-constrained molecular dynamics (1.5 ns) derived geometries of the adenine-adenine overlaps at each dinucleotide step of the hexameric ssDNA and ssRNA, [d/r-($^{5'}$GAAAAC)], show that the relatively electron-rich imidazole stacks above the electron-deficient pyrimidine in the 5′ to 3′ direction in ssDNA, while, in contradistinction, the pyrimidine stacks above the imidazole in the 5′ to 3′ direction in ssRNA.

The pK_a measurement of the heptameric ssDNA and ssRNA sequences by ^{1}H NMR (600 MHz) showed (*41*) that the local changes in the microenvironment perturbs the pK_a values (pK_{a1}) of 1-guaninyl nucleobase in a sequence-specific manner(*42*). From the extent of the shift of the perturbed pK_{a2s} observed from the nearest-neighbors as a result of *N1* deprotonation of 1-guaninyl residue, we also could provide unequivocal evidence regarding the sequence-specific nature of the coupled nucleobases in the stacked system in ssDNA/ssRNA as a result of the internucleobase cross-talk (*41, 43-45*).

In a separate study, it was also shown through our pK_a study (*46*) that all the internucleotidic phosphates of ssRNAs have different microenvironments depending upon their sequence context, which result in to different sequence-specific alkaline hydrolysis rates, whereas the microenvironments of the internucleotidic phosphates in a ssDNA counterpart are very closely similar.

We now report, for the first time, the sequence-specific nature of the self-organization of the ssDNA and ssRNA structures in order to shed light on their respective reactivity. We have thus addressed the following questions in this work: (i) How the structural variability is altered depending upon the sequence context in the ssDNA and ssRNA? (ii) How does purine to pyrimidine substitution in the middle of the single strand change the stacking pattern at the fraying ends (5' or 3')? (iii) What is the relative stability of the isosequential ssDNA and ssRNA?

(iv) Is the relative mobility of the core and the 3' and 5'-ends similar? If not, which is more flexible, 3' or 5', in the ssDNA and ssRNA? (v) How are the phosphate backbone conformations in the isosequential ssDNA and ssRNA modulated depending upon the sequence context?

Experimental procedures

(A) NMR Experiments and resonance assignments

All NMR experiments were performed in Bruker DRX 600/ 500 MHz NMR at 278 K, 283 K, and 298K in D_2O and DSS was used as the internal standard. Structural assignments of single-stranded d($^{5'}C^1A^2\underline{\mathbf{X}}^3G^4\underline{\mathbf{Y}}^5A^6C^7$) ($\underline{\mathbf{X}}^3$ = A or C, $\underline{\mathbf{Y}}^5$ = A or C) (**1 - 4,** Figure 1) and r($^{5'}C^1A^2\underline{\mathbf{X}}^3G^4\underline{\mathbf{Y}}^5A^6C^7$) ($\underline{\mathbf{X}}^3$ = A or C, $\underline{\mathbf{Y}}^5$ = A or C) (**5 – 8,** Figure 1) are presented in Fig S1 – S40 in SI. Chemical shifts of aromatic protons are shown in Table S1 in the Supporting information (SI), whereas coupling constants of the anomeric protons are shown in Tables 1 and 2.

The proton resonances of all ssRNAs/ssDNAs have been assigned by ^{1}H nuclear Overhauser spectroscopy (NOESY), ^{1}H double-quantum-filtered correlation spectroscopy (DQF-COSY) with and without ^{31}P decoupling, TOCSY, and ^{31}P, ^{1}H correlation spectroscopy at 278, 283, and 298 K. For each free induction decay (FID) of NOESY, ^{1}H DQF-COSY, ^{31}P-decoupled ^{1}H DQF-COSY, and total correlation spectroscopy (TOCSY) spectra of 64-128 scans were recorded with a relaxation delay of 2 s (see SI for details). Four thousand complex data points were collected in the t_2 dimension, and 256-1024 experiments were run in the t_1 dimension. All NOESY spectra were recorded using a mixing time (τ_m) of 800 ms (see text).

(B) NMR constraints used for structure determination

Three types of NMR observed constraints have been used for NMR-constrained simulated annealing (SA) and molecular dynamics (MD), NMR-SA/MD, derived structures: (1) Distance constraints (Tables S11-14 in SI), (2) Sugar constraints (Tables 1 and 2), and (3) Phosphate backbone constraints (Table S2-S9).

(1) Distance constraints based on NOESY spectra

NOESY cross-peaks at 800 ms mixing time were used only as loose constraints by two-spin approximation, along with other dihedral constraints (see below), throughout the molecular dynamics simulation using a reference cross-peak intensity of H5-H6 of cytidine (2.54 Å). The cross-peaks were classified as strong (s, 1.8-3.0Å), medium (m, 3.5-5.0Å) and weak (w, 3.0-6.5Å). Strong peaks are considered to originate from direct dipole-dipole NOE transfer, while the constraints of the weak peaks allow for a large contribution from spin-diffusion during the relatively long mixing time, 800 ms. All inter and intra-residual cross-peaks used in the NMR-

SA/MD are shown in Tables S11 – S14 (for classification of intensities of nOe crosspeaks), and Figures S41 – S49 in SI (for all nOe footprints and connectivities of all eight ssDNAs and ssRNAs).

A total of 80 distance constraints for d($^{5'}C^1A^2A^3G^4A^5A^6C^7$), 63 distance constraints for d ($^{5'}C^1A^2A^3G^4C^5A^6C^7$), 77 distance constraints for d($^{5'}C^1A^2C^3G^4A^5A^6C^7$), 108 distance constraints for d ($^{5'}C^1A^2C^3G^4C^5A^6C^7$) are used from NMR experiments (Tables S11 – S14), along with other constraints (Tables 1-2, S2-S9 in SI) in the SA/MD protocol (Figure 5). Similarly, Total 55 distance constraints for r($^{5'}C^1A^2A^3G^4A^5A^6C^7$), 62 distance constraints for r($^{5'}C^1A^2A^3G^4C^5A^6C^7$), 58 distance constraints for r($^{5'}C^1A^2C^3G^4A^5A^6C^7$), and 60 distance constraints for r($^{5'}C^1A^2C^3G^4C^5A^6C^7$) have been introduced from NMR experiments in the SA/MD protocol (Table 3).

(2) Dihedral constraints

(i) Sugar Pucker constraints: The sums of the vicinal coupling constants between 1',2' and 1',2" ($\Sigma^3J_{H1'H2'/H2''}$) in ssDNAs (**1 - 4**) and vicinal coupling constants between 1',2' ($^3J_{H1'H2'}$) in ssRNAs (**5 - 8**) from the phosphorus decoupled DQF-COSY at 600 MHz were used to estimate the percentile population of 2'-*endo*-type (S) and 3'-*endo*-type (N) sugar conformations (*47*) of all nucleotide residues in both the heptameric ssDNAs and the ssRNAs (Tables 1 and 2). When the ssDNA residues were predominantly in the S-type conformation (%S ≥ 70), they were constrained to P = 150-210°, and for the ssRNA residues, which were found to be predominantly in N-type conformation (%N ≥ 70), they were constrained to P = 0-60°. The ssDNA and ssRNA residues with mixed conformations (%N/S ca 1:1) were constrained to a broader range encompassing both S-and N-type conformations, P = 0-210°. Total 45 sugar constraints were given on the basis of phase angles.

(ii) Backbone Constraints: Qualitative NMR data was used to assess the available conformational hyperspace of the backbone: The (n)P-(n)H4' correlation is only detectable when the four bonds in the H4'-C4'-C5'-O5'-P backbone are located in the same plane forming a W-shaped conformation. This is possible when the β and γ torsions are trans and gauche$^+$, respectively, which is the most common conformation (*48*) for both A/B-DNA or A-RNA. The presence of strong (n)P-(n)H4' cross-peaks (Figures S5, S10, S15, S20, S25, S30, S35, S40 in SI) for four isosequential pairs of ssDNA and ssRNA show that the β^t and γ^+ conformations are frequently populated. The $^3J_{H4'\,H5'/\,H5''}$ coupling is very sensitive to the γ torsion. With the γ torsion in gauche$^+$ conformation, no strong couplings between H4' and H5'/H5" (1.0-2.5 Hz) is expected, whereas both trans and gauche$^-$ will show strong coupling (>10 Hz) between the H4' and one of the H5's. No strong $^3J_{H4'H5'/\,H5''}$ could be found for any of the residues for any of the ssDNAs or

ssRNAs. It should however be noted that this region (3-4 ppm) is severely crowded especially for ssRNAs. If ε is in gauche$^-$ conformation, it should produce a detectable $^4J_{H2'P}$ coupling when the sugar is in S-type conformation (*47*). For ssDNAs, no such cross-peaks could be observed; thus the ε torsion is dominated by trans conformation since gauche$^+$ appears to be sterically forbidden (*47*). Since the sugar conformation of the ssRNA is mainly in 3'-*endo* conformation, only the sterically forbidden ε^+ was excluded. Backbone and dihedral torsion used in the NMR derived MD simulations for ssDNA and ssRNA are given in the Table S2 – S9 in SI. Total 19 backbone constraints were given (Table 3).

(C) Structure-building based on the NMR constraints using simulated annealing (SA) and Molecular Dynamics (MD) protocol

As stated above, the NMR constraints used for the structure building are based on (i) distance constraints, (ii) endocyclic sugar torsion constraints, and (iii) backbone constraints.

(1) **Initial structure building**

Canonical starting structures for the heptameric ssDNAs and ssRNAs were built with AMBER 7.0 using the nucgen option. The DNA heptamer was built from a less favored A-type antiparallel DNA duplex, from which the second strand was subsequently deleted. The RNA heptamer was built from the B-type antiparallel RNA duplex, where the second strand was deleted. The XLEAP module configured with the parm94 parameter set for AMBER 7.0 was used to create the topology file and final starting coordinates. The phosphate negative charges were neutralized by the addition of 6 Na^+ counter ions. The ssDNA was solvated in a periodic box with the dimensions 42 x 42 x 44 Å^3 filled with 4977 – 5088 TIP3P water molecules, surrounding the molecule by 8 Å in all dimensions. With the same procedure, the ssRNA was solvated in a 41 x 43 x 47 Å^3 box filled with 5463 – 5751 water molecules. All simulations were performed solvated in a box filled with water molecules with added sodium ions.

(2) **Minimizing the initial structures**

The system was first equilibrated in several steps. First the solute was restrained, while the water molecules were minimized for 1000 steps using the steepest descent minimization algorithm. This minimization step is then repeated once restraining only the heavy atoms in the solute. Second, a short MD simulation using 1 fs time steps was run on the system, once again with restrained solute, heating from 100 to 298 K during 3 ps and was then simulated for a total of 30 ps to allow the water molecules rearrange and relax. Another 30 ps of MD was run on the water molecules only, introducing long-range electrostatic interactions using particle mesh Ewald

summation. Finally, the whole system was minimized in five cycles, 1000 steps each, gradually releasing the restraints on the solute molecule.

(3) **Simulated Annealing (SA-MD)**

The resultant structures after minimization were then subjected for repeated cycles (30ps each) of heating followed by cooling steps (Figure 5). During the first 10 ps, the structures were heated from 100 to 400 K of MD at constant pressure. The MD was then run at 400 K for another 10 ps while gradually increasing the strength of the dihedral and sugar constraints (Table S2-S9) in the Supporting Information) to full strength (50 kcal mol^{-1} rad^{-2}), and softly introducing the NOE constraints to full strength (10 kcal mol^{-1} $Å^{-2}$), as shown in Figure 5. Subsequently, while cooling the system to 100 K during another 10 ps MD simulation the strength of NOE constraints were increased to 50 kcal mol^{-1} $Å^{-2}$, retaining the dihedral constraints at 50 kcal mol^{-1} rad^{-2}. This was then cycled to produce 30 structures in ssDNA except for ssDNA (**2**) where 80 cycles were performed. For all ssRNAs 70-80 cycles were performed in order to converge structures with low RMSd. After each full cycle (30 ps) of SA the RMSd value of the final structure at 100K was compared with all the final structures of the preceding SA steps in order to evaluate the convergence. When the ΔRMSd value was found to be within 0.5-1.6 Å the structures of ssDNA and ssRNA were considered to be converged (Figure 6). Sugar puckering (N/S-type) of the each sugar residues of final structures of the ssDNAs and ssRNAs at the end of each SA cycle at 100K were compared with the phase angles and puckering amplitude range calculated from the 1'2' and 1'2" coupling constants (Figure 2, Tables 1 and 2).

After the convergence of final structure at 100K, the final structure with lowest RMSd is taken as initial structure for MD simulation.

(4) **NMR-Constrained Molecular Dynamics Simulations**

All NMR constraints for MD simulations are identical to the NMR constraints used in the SA protocol except for the sugar pucker constraints, which had a broader range than those used in the SA protocol: P = 0 - 120° for N-type and 120 - 210° for S-type. Before the production run was started, the system was once more heated from 100 to 298 K during the first 3 ps of a 30 ps MD simulation. During the 3 ps of heating, the experimentally derived NMR distance and dihedral constrains were also scaled up from 0 to 20 kcal mol^{-1} $Å^{-2}$ and 0-2 kcal mol^{-1} rad^{-2}, respectively, and kept constant during the rest of the MD simulation. The production MD was then run using 1.5 fs time steps, 8.0 Å VdW interactions cutoff, and the particle mesh Ewald summation method with 1.0 dielectric constant for the long-range electrostatic interactions. The root-mean-square deviation (RMSd) (Figure S50) and potential energy profile of the trajectory was monitored

(Figure S51), and the simulation was stopped when both variables had reached equilibrium at 298 K. The MD simulations for both ssDNA (**1** - **4**) and ssRNA (**5** - **8**) were run for a total of 2 ns.

(5) **Structure sampling**

MD coordinates were dumped to the trajectory every 0.15 ps of the molecular modeling simulation. From the trajectory, water molecules were stripped off from the structures and RMSd and potential energy of the structures were then analyzed for confirmation of the convergence of the structures. Ten snapshots from the last 100 ps of this region of the simulation were stripped of water, and then saved as snapshots every 10 ps from 1.9 ns to the end of the simulation at 2 ns for both the ssDNA and ssRNA simulations (Figure 7). The average structure of the MD trajectory of this last 100 ps was also calculated based on one coordinate set per 0.15 ps and minimized for 2000 steps using the conjugate gradient method and the full NMR constraints were switched on to bring the most obvious averaging effects back to equilibrium (RMSD between the initial average and the minimized average was < 1Å).

(6) **Structure Analysis of the Backbone**.

The ptraj module for Amber 7.0 was used to extract the dihedrals of the sugar-phosphate backbone and the sugar phase angles from the average structure.

(7) **Structure Analysis of the Helical Parameters**.

The *X3*DNA program was used to calculate the helical parameters of the averaged single-stranded DNA and RNA.

Results and Discussion

(A) **Sugar Conformations of ssDNAs and ssRNAs as observed by NMR**

Sugar conformations of each nucleotide residue in ssDNAs and ssRNAs have been estimated from the ^{31}P-coupled and ^{31}P-decoupled DQF-COSY spectra at 298K. The $^{3}J_{HH}$ coupling constant analysis gave us the percentile North-type populations (Tables 1 and 2). Plots of percentile North-type population against the nucleotide number in the ssDNA or ssRNA sequence are shown in Figure 2. The summary of our observations (Figure 2) is as follows:

(i) The 5'-end is more flexible compared to the 3'-end for four isosequential pairs of ssDNA and ssRNA. *These suggest that the nucleobases at the 5'-end are relatively less stacked compared to those at the 3'-end (Figure 2).*

(ii) A close examination of the crystal structures for double-stranded DNAs and RNAs showed that the constituent sugar residues exist in S-type (34° > P > -1°)(*47*) and N-type (194° >

P > 137°)(*47*), respectively. Present work on the solution phase NMR-constrained SA/MD structure of purine-rich ssDNAs and ssRNAs has revealed that sugar residues in ssDNA and ssRNA exist also in S- and N-type, respectively, with some exceptions. These suggest that the percentile N/S-type sugar population in ssDNA/ssRNA can be used as one of the criteria, in conjunction with the nOe data, to evaluate whether a particular residue is stacked with the nearest neighbor or not.

(iii) In all of our ssDNAs (**1** - **4**), the sugar conformations (Tables 1 and 2) of most residues are S-type, except $\mathbf{A^2}$ (77% N-type), $\mathbf{A^3}$ (77% N-type), $\mathbf{G^4}$ (54% N-type) in d($^{\mathbf{5'}}C^1A^2A^3G^4C^5A^6C^7$) (**2**), $\mathbf{A^2}$ (68% N-type) and $\mathbf{G^4}$ (52% N-type) in d($^{\mathbf{5'}}C^1A^2C^3G^4A^5A^6C^7$) (**3**), $\mathbf{A^2}$ (54% N-type) and $\mathbf{A^3}$ (42% N-type) in d($^{\mathbf{5'}}C^1A^2A^3G^4A^5A^6C^7$) (**1**). Sugar residues of $\mathbf{A^2}$ and $\mathbf{A^3}$ in d($^{\mathbf{5'}}C^1A^2A^3G^4A^5A^6C^7$) (**1**) are found to be in ≈50:50 N/S population compared to those of the purine-rich hexameric ssDNA, d($^{\mathbf{5'}}G^1A^2A^3A^4A^5C^6$)(*40*), in which the A residues were predominantly in S-type conformations. These data show that the sugar pseudorotamer populations in these ssDNAs are sequence-dependent. The respective sugar conformations of d($^{\mathbf{5'}}C^1A^2A^3G^4A^5A^6C^7$) (**1**), d($^{\mathbf{5'}}C^1A^2A^3G^4C^5A^6C^7$) (**2**) and d($^{\mathbf{5'}}C^1A^2C^3G^4A^5A^6C^7$) (**3**) also suggest that they do not truly belong to A- or B- type conformation (Figure 3) (see discussion on nOe footprints in Section 2 as well as S49 Panel A – H in SI).

(iv) In ssRNAs, the sugar conformations of most residues are N-type except for $\mathbf{C^3}$ (54% N-type) and $\mathbf{G^4}$ (54% N-type) in r($^{\mathbf{5'}}C^1A^2C^3G^4A^5A^6C^7$) (**7**), $\mathbf{G^4}$ (45% N-type) in r($^{\mathbf{5'}}C^1A^2C^3G^4C^5A^6C^7$) (**8**), suggesting that those residues are weakly stacked, and those ssRNAs adopt non-A-RNA type structure *(see the nOe observations in Section 2, suggesting non-A-type conformation).*

(v) In all ssRNAs (**5** – **8** in Figure 1), the central G^4 residue is fully conserved, whereas its nearest-neighbors are substituted systematically either by A or C, one at a time, at 3'- or 5'-end of G^4. The 3'-terminal AC or 5'-teriminal CA residues are expected to be flexible. The effect of the nearest-neighbors on the G^4 residue in ssRNA can however be broadly divided in to two groups, basing on the N/S population of G^4 (Table 1 and 2), which is dictated by the central triplet core structure (underscored): (a) ≈70% N-type in r($^{\mathbf{5'}}C^1A^2\underline{A}^3\underline{G}^4\underline{A}^5A^6C^7$) and r($^{\mathbf{5'}}C^1A^2\underline{A}^3\underline{G}^4\underline{C}^5A^6C^7$), and (b) ≈50% N/S in r($^{\mathbf{5'}}C^1A^2\underline{C}^3\underline{G}^4\underline{A}^5A^6C^7$) and r($^{\mathbf{5'}}C^1A^2\underline{C}^3\underline{G}^4\underline{C}^5A^6C^7$). Similarly in ssDNA, we also see two different groups of nearest-neighbors (central triplet sequence is underscored) affecting the sugar conformations of $\mathbf{G^4}$: (a) ≈70% S-type in d($^{\mathbf{5'}}C^1A^2\underline{A}^3\underline{G}^4\underline{A}^5A^6C^7$) and d($^{\mathbf{5'}}C^1A^2\underline{C}^3\underline{G}^4\underline{C}^5A^6C^7$), and (b) ≈50% N/S in d($^{\mathbf{5'}}C^1A^2\underline{A}^3\underline{G}^4\underline{C}^5A^6C^7$) and d($^{\mathbf{5'}}C^1A^2\underline{C}^3\underline{G}^4\underline{A}^5A^6C^7$). *This shows that the nature of stacking is variable basing on whether it is ssRNA or ssDNA as well as their sequence-context (see Section 2 for the nOe evidence).*

(vi) It is interesting to note that when 1-cytosinyl moiety ($\mathbf{C^5}$) is the 3'-nearest neighbor to $\mathbf{G^4}$ in d-($^{\mathbf{5'}}C^1A^2A^3G^4\underline{C}^5A^6C^7$) the sugar puckering of $\mathbf{A^2}$/$\mathbf{A^3}$ residues at 5'-end exist in more N-type conformation (>76 %) in ssDNA, but all residues at the 3'-end ($\mathbf{C^5}$/$\mathbf{A^6}$/$\mathbf{C^7}$) adopt S-type

conformation (>90%) (Figure 3). On the contrary, when 1-cytosinyl (**C**^{3}) is the 5'-nearest neighbor to **G**^{4} in d($^{5'}$C^{1}A^{2}C^{3}G^{4}A^{5}A^{6}C^7) the sugar puckering of **C**^{1}/**C**^{3} at the 5'-end exist in more S-type conformation (> 85%) in comparison with that of d($^{5'}$C^{1}A^{2}A^{3}G^{4}C^{5}A^{6}C^7), whereas all residues (**A**^{5}/**A**^{6}/**C**^{7}) at 3'-end exist in predominantly S-type conformation (>70%) (Figure 3). *Therefore the effect of 1-cytosinyl as the 3'-nearest neighbor to G^4 changes the sugar population at the 5' -end of ssDNA, whereas 1-cytosinyl as the 3'-nearest neighbor to G^4 makes no difference in the sugar population at 3'-end.*

(vii) In ssRNAs, this influence of 1-cytosinyl at the 3'- versus 5'-nearest neighbor to **G**^{4} has been found to be just the reverse: With the 1-cytosinyl (**C**^{3}) as the 5'-nearest neighbor to **G**^{4} in r($^{5'}$C^{1}A^{2}C^{3}G^{4}A^{5}A^{6}C^7), the sugars of **C**^{3} and **G**^{4} exist in N/S-type mixed population. But when the **C**^{5} is the 3'-nearest neighbor to **G**^{4} in r($^{5'}$C^{1}A^{2}A^{3}G^{4}C^{5}A^{6}C^7) all residues exist predominantly in N-type conformations (>74%) (Figure 3*). Thus, the effect of C^3 or C^5 as the 3'- or 5'-nearest neighbor has different effect on G^4 , the C^3 as the 5'-nearest neighbor to G^4 makes it more flexible compared to C^5 as the 3'-nearest neighbor.*

(viii) Sugar residues of **A**^{2} and **A**^{3} in d($^{5'}$C^{1}A^{2}A^{3}G^{4}C^{5}A^{6}C^7) (**2**) are found to be predominantly in the N-type conformation which is never found in a typical B-type dsDNA structures. On the contrary, sugar residue for **G**^{4} in d($^{5'}$C^{1}A^{2}A^{3}G^{4}C^{5}A^{6}C^7) (**2**) are found to exist in N/S mixed (≈1:1) conformation. Interestingly, sugar residues of **A**^{2}, **A**^{3} and **G**^{4} show almost the identical conformational pattern in the isosequential pairs of ssDNA (**2**) and ssRNA (**6**) [Figure 3]. Nucleotides **C**^{1}, **C**^{5}, **A**^{6} and **C**^{7} show the B-type structure, **A**^{2}, **A**^{3} show A-type structure and **G**^{4} shows A-/B-mixed type structure in ssDNA (**2**). In contrast, only **G**^{4} shows mixed A-/B-type RNA structure in ssRNA (**6**).

(ix) Sugar residue of **A**^{2} in d($^{5'}$C^{1}A^{2}A^{3}G^{4}C^{5}A^{6}C^7) (**3**) is also found to remain in N-type conformation like the ribo-counterpart in isosequential ssRNA (**7**). In contradistinction, sugar residue of **G**^{4} exists in N-/S- mixed type conformation in both isosequential ssDNA (**3**) ssRNA (**7**) [Figure 3].

(x) But in pyrimidine-rich ssDNA (**4**) all sugar residues are found to be in S-type sugar conformations, whereas **G**^{4} in ssRNA (**8**) exist in mixed N/S-type population (≈1:1).

Thus the above features of the sugar conformation clearly showed the strong role of the sequence context in the self-organization of ssDNA vis-à-vis ssRNA.

(B) ***NOE Distances and the self-organization of the backbone***

In the B-type dsDNA and A-type dsRNA, the nucleobases are highly organized due to interstrand base-pairing as well as intra and inter-strand stacking interactions. These features signify that the degrees of freedom in the double-stranded forms of RNA and DNA are relatively less compared to the isosequential single-strand, because the latter has more tumbling and

internal motions due to absence of interstrand base-pairings and stackings. Hence the ssDNA and ssRNA structures can only be organized by intra-strand stacking as well as by hydration and salt effect. Thus, the ssDNA/ssRNA are expected to have lower energy barriers for internal motion involving the nucleobase itself, the pseudorotation of the sugar moieties as well as the sugar-phosphate backbone, resulting in overall increased dynamics.

(I) Internal motion and overall molecular tumbling: 2D NOESY experiments were performed with 800 ms mixing time because of two reasons: First, the size of the ssDNA and ssRNA oligomers studied are relatively small leading to shorter correlation times compared to more commonly studied larger duplex structures. This has a direct effect on nOe build up rates. The increased dynamics can also make any particular pair of protons spending more time in a specific conformation that either do contribute very strongly or do not contribute at all to the net NOE transfer, thereby leading to either over or underestimation of the corresponding distances. Second, the short correlation time of both the overall molecular tumbling and the internal motions will reduce the nOe transfer rate through the spectral density function, resulting in slow build-up rates for double-stranded DNA or RNA oligomers. As the degrees of freedom in ssDNA and ssRNA are more compared to their double stranded forms the mixing time is considerably higher to achieve the evolution of crosspeaks for all nOe connectivities involving intra (n) and inter (n-1) residual aromatic protons (H8/H6/H2) and sugar protons for all four isosequential pairs of ssDNAs and ssRNAs. The nOe cross-peaks evolved for the interproton contacts have an undefined contribution from the spin diffusion, and therefore they were categorized only qualitatively as strong (s), medium (m) and weak (w) according to their relative intensities with respect to H5-H6 proton nOe crosspeak of 1-cytosinyl residue as the internal reference [Table S11-S14 in for ssDNAs (**1** - **4**) and ssRNA (**5** - **8**)]. The strong nOe cross-peaks have the major contribution from the direct dipole-dipole nOe transfer, while the medium and weak peaks allow for a large contribution from spin-diffusion during the relatively long 800 ms mixing time. All 2D nOe experiments at 283K and 298K were recorded using 1mM ssDNA and ssRNA solutions where D_2O was used as solvent and DSS was used as the internal reference. 1mM concentration was maintained for all ssDNA and ssRNA samples to avoid the intermolecular self aggregation which can affect the ^{1}H chemical shifts.

(II) Experimental dihedral constraints: Constraint for sugar-phosphate backbone dihedrals (α, β, γ, δ, ε, ζ, χ) were taken from the crystal structures (*40, 47*) and ^{31}P-^{1}H NMR correlations (*e.g.* presence or absence of H2′ or H4′ to 3′ and/or 5′-phosphate for γ and ε torsion (*47*)), and DQF-COSY (^{31}P-coupled and ^{31}P-decoupled spectra for γ and ε torsion) experiments. It is noteworthy that all the distance and dihedral constraints (Tables S2 – S9, S11 – S14 in SI) in the SA/MD step were made from the nOe intensities and DQF-COSY and ^{31}P-^{1}H NMR correlation spectra to model the NMR refined structures of ssDNA and ssRNA.

It should be noted that inter-residual $(H8/H6)_n$ – $(H1'/H3')_{n-1}$ nOe crosspeak intensity depends on the sugar conformations of two nearest neighbor residues in ssDNA and ssRNA, *i.e*, if conformation of 5′-sugar is predominantly N- or S-type in ssRNA or ssDNA the nOe crosspeaks evolved for $(H8/H6)_n$ – $(H1'/H3')_{n-1}$ contacts are dependent on the distance between aromatic protons of 3′-nearest neighbor nucleobase (n) and the preceding (5′) sugar protons (n-1). If the conformation of the 5′-sugar (n-1) is of mixed N/S populations, intensities of nOe crosspeaks for $(H8/H6)_n$ – $(H1'/H3')_{n-1}$ contacts change dynamically depending on the percentile populations of the N or S type over the S or N type in ssDNA or ssRNA. The nOe intensity of $(H8/H6)_n$ – $(H1'/H3')_{n-1}$ contacts also depend upon χ orientation of the 3′-nearest neighbor (n) nucleobase.

(III) The nOe footprints: The nOe footprints for the four isosequential pairs of ssDNA and ssRNA (Table 4, Figure S49, panels A – H in SI for nOe footprints) lead to the following observations:

(i) The nOe footprints of dsDNA and dsRNA provide direct evidence for the right handed helical turn (*47, 49*) because they typically show the nOe connectivity between $H8_n/H6_n$ aromatic protons with the $(H1'/H2'/2''/H3')_{n-1}$ of the preceding sugar (n-1) (2.0 - 5.1 Å in the right handed A- / B-form). But in Z-type DNA no $(H8/H6)_n \rightarrow (H1'/ H2'/2''/ H3')_{n-1}$ nOe contacts are found because the distances are larger (6.9-8.0 Å). But $(H8/H6)_n \rightarrow (H5'/5'')_{n-1}$ nOe contacts are found in Z-type structures because inter-residual distances remain within 3.3 - 4.7 Å. In A and B type structures, $(H8/H6)_n \rightarrow (H5'/5'')_{n-1}$ contacts can not be found because inter-residual distances range from 6.2 -7.2 Å. No nOe contact for $(H8/H6)_n \rightarrow (H5'/5'')_{n-1}$ has been however found in the four pairs of isosequential ssDNA and ssRNA. *It suggests that none of the sequences shows the Z-type structure.*

(ii) Since strong and medium inter-residual nOe contacts for aromatic $(H8/H6)_n$-$(H1'/H3')_{n-1}$ have been found for all the four isosequential pairs of ssDNA and ssRNA structures (**1-4** for ssDNA and **5-8** for ssRNA) *we conclude they remain in the right handed forms in the NMR time scale*.

(iii) The structures of all ssRNA sequences **(5 - 8)** are not of A-type since $(H8/H6)_n$-$(H1')_{n-1}$ intensities of contacts have been found (Figure S49 Panels B, D, F, and H in SI) to be comparable with $(H8/H6)_n$- $(H3')_{n-1}$. Since Z-DNA structure with CG repeats has shown to have *Syn* (+) glycosyl torsion, one would expect nOe contacts between $(H2)_n$ and $(H2'/ H3')_n$, which has not been found with any of our ssDNAs or ssRNAs.

(iv) For the ssDNA sequences, [d($^{5'}C^1A^2A^3G^4A^5A^6C^7$) (**1**) / d($^{5'}C^1A^2A^3G^4C^5A^6C^7$) (**2**) / d($^{5'}C^1A^2C^3G^4A^5A^6C^7$) (**3**)], $(H8/H6)_n$- $(H1')_{n-1}$ nOe contacts were found (S49 Panel A, C, E, and G in SI)) to be relatively stronger than $(H8/H6)_n$ - $(H3')_{n-1}$, whereas in d($^{5'}C^1A^2C^3G^4C^5A^6C^7$) (**4**) intensities of $(H8/H6)_n$ - $(H1')_{n-1}$ nOe contacts were found to be comparable with that of $(H8/H6)_n$

- $(H3')_{n-1}$ suggesting that these structures do not show typical B-type conformation (Table 4 , Figure 4, Figure S49, panels A – H in SI for nOe footprints).

(v) The 3'-*endo* sugar conformations in ssRNA sequences are qualitatively identified as $(H8/H6)_n$- $(H2')_{n-1}$ connectivities have been found (Table S13 in SI). Maximum intensities for $(H8/H6)_n$- $(H2')_{n-1}$ contacts have been found for ssRNA [r($^{5'}C^1A^2A^3G^4A^5A^6C^7$) (**5**) /r($^{5'}C^1A^2A^3G^4C^5A^6C^7$) (**6**)(/r-($^{5'}C^1A^2C^3G^4C^5A^6C^7$) (**8**)], which suggest that sugar residues in these ssRNA sequences exist predominantly in N-type conformation. But in contradistinction, $(H8/H6)_n$- $(H2')_{n-1}$ connectivities show nOe crosspeaks of weak intensities in [r($^{5'}C^1A^2C^3G^4A^5A^6C^7$)] (**7**)**,** which suggests that the population of the N-type conformations for the sugar residues in [r($^{5'}C^1A^2C^3G^4A^5A^6C^7$)] (**7**) are relatively less (Table 4 , Figure 4, Figure S49, panels A – H in SI for nOe footprints).

(vi) It is noteworthy that the number of $(H8/H6)_n$ - $(H3')_{n-1}$ contacts in [d($^{5'}C^1A^2A^3G^4C^5A^6C^7$)] and [d($^{5'}C^1A^2C^3G^4A^5A^6C^7$)] (**2/3**) are less compared to those in [d($^{5'}C^1A^2A^3G^4A^5A^6C^7$)] and [d($^{5'}C^1A^2A^3G^4A^5A^6C^7$)] (**1/4**). The A^2/A^3 sugar residues in ssDNA (**2**) show ≈1:1 N/S population, whereas A^2/G^4 in ssDNA (**3**) show ≈70% N-type sugar populations. This suggests that ssDNAs, (**2**) and (**3**) adopt a non-B-type DNA structures compared to ssDNAs (**1**) and (**4**) (Figure S49, panels A-E in SI for nOe footprints).

(vii) Intra-residual nOe crosspeaks between aromatic (H8/H6) and sugar protons are found (Table S12 and S14 in SI) to be stronger for nucleotide residues at the 3'-end compared to those at the 5'-end in ssDNAs (**1-4**) and ssRNAs (**5 - 8**), Figure S49, panels A – H in SI for nOe footprints.

(C) **Evaluation of the stacking patterns from the nOe contacts.**

In Tables S11 and S13, the relative strength of inter-residual nOe contacts of ″n″ to ″n-1″ [H1′/H2′/H3′ of ″n-1″ at the 5′-end with the aromatic protons of ″n″ towards 3′-end] by pairwise comparison between isosequential ssDNA and ssRNA are shown as evidences supporting the nearest-neighbor stacking interaction (Figure S49, panels A – H in SI for nOe footprints). The main conclusions are as follows:

(i) The ssRNAs (**5**) and (**6**) are found to be more stacked compared to ssDNAs (**1**) and (**2**), whereas ssDNAs (**3**) and (**4**) are found to be more strongly stacked compared to their isosequential ssRNA counterparts (**7**) and (**8**), because more number of strong of inter-residual ″n″ to ″n-1″ nOe contacts are found (Tables S11 and S13 in SI) in the former group compared to the latter.

(ii) The relative strength of inter-residual nOe contacts of ″n″ to ″n-1″ within each group of ssDNAs (**1** – **4**) versus ssRNAs (**5** – **8**) also allows us to classify the ssRNAs and ssDNAs (Figure 1) on the basis of their relative stacking strength within the triplet core (shown in pink) :

d($^{5'}C^1A^2\underline{A}^3\underline{G}^4\underline{A}^5A^6C^7$) (**1**) > d($^{5'}C^1A^2\underline{C}^3\underline{G}^4\underline{C}^5A^6C^7$) (**4**) > d($^{5'}C^1A^2\underline{C}^3\underline{G}^4\underline{A}^5A^6C^7$) (**3**) > d($^{5'}C^1A^2\underline{A}^3\underline{G}^4\underline{C}^5A^6C^7$) (**2**). For ssRNAs, the stacking order is however r($^{5'}C^1A^2\underline{A}^3\underline{G}^4\underline{A}^5A^6C^7$) (**5**) > r($^{5'}C^1A^2\underline{A}^3\underline{G}^4\underline{C}^5A^6C^7$) (**6**) > r($^{5'}C^1A^2\underline{C}^3\underline{G}^4\underline{C}^5A^6C^7$) (**8**) > r($^{5'}C^1A^2\underline{C}^3\underline{G}^4\underline{A}^5A$ (**7**), which is consistent with the results of the NMR constrained simulated annealing and molecular dynamics studies (Section 4 and Experimental Section).

(D) ***NMR-Constrained Molecular Dynamics (NMR-MD) simulations for ssDNA and ssRNA.***

As the nucleotides in the 3' and 5' fraying-ends are more dynamic in nature compared to the other nucleotides inside the ssDNA and ssRNA the contribution of the fraying-end nucleotides to the total RMSd were excluded. Thus, the NMR–derived MD structures of ssDNA and ssRNA showed RMSd of 1-4 Å depending upon the sequence (see Figure S50 in SI). Potential energies of ssDNA and ssRNA (Figure S51 in SI) have been calculated throughout the 2.0 ns of constrained MD steps, and found to be constant with the time. It means that the structures in the final MD steps have very similar potential energy, and the energy barrier for interconversion also is very small.

From the NMR spectra for ssDNAs and ssRNAs it has been found that the inter-residual nOe contacts are more intense at the 3'-end compared to 5'. It suggests that the 3'-ends of ssDNAs (**1** – **4**) and ssRNAs (**5** - **8**) are more organized compared to those at the 5'-ends. From the RMSd calculations for nucleobases at the 5'-end (nt 1-3) and 3'-end (nt 5-7) in the MD steps, it is revealed that orientations of 5'-end nucleobases are indeed more distorted in comparison with those of the 3'-end nucleobases (Table 10).

Overlay of 10 structures, each harvested every 10 ps from the last 100 ps of the total of 2 ns NMR-constrained MD run for each ssDNA and ssRNA shows that they adopt different structural patterns depending upon their sequence context (see Figure 7). The average NMR-MD structures of ssDNAs and ssRNAs show that the dinucleotide stacking pattern inside the ssDNAs and ssRNAs differ from one sequence to another (Figure 8), and is solely dependent on the extent of enrichment of purine (A/G) versus pyrimidine (C) residues in those strands.

(E) ***Observations from the final 10 structures harvested from the last 100ps of 2ns MD run for ssDNAs and ssRNAs.***

(i) From the overlay of 10 final MD structures of ssDNA (**1** - **4**) and ssRNA (**5** - **8**), taken from last 100 ps of the total 2 ns MD run, it can be seen (Fig 7) that the 5'-ends show more mobility than that of the 3'-ends.

(ii) Purine-rich sequences in ssDNA and ssRNA are more organized than the pyrimidine rich sequences [except d($^{5'}C^1A^2C^3G^4C^5A^6C^7$) (**4**)].

(iii) From the RMSd analysis (Table 9) of the final 10 structures, it can be seen that the relative stabilities (pairwise) of the four pairs of isosequential heptameric ssDNAs (**1** – **4**) and ssRNA (**5** – **8**) are as follows: (**1**) > (**5**) , (**6**) > (**2**), (**3**) > (7), (**8**) > (**4**). The RMSd of 10 overlaid structures in Figure 7 for ssDNA are between 0.2 – 1.0 Å, whereas for ssRNAs the RMSds are between 0.1-0.8 Å (Table 9).

(iv) The RMSd of the MD simulated structures of B-DNA with respect to typical B-type dsDNA is found to be ≈3.9 Å, whereas the RMSd of the MD simulated structures of A-RNA structures with respect to typical A-type dsRNA is ≈2Å (*50*). Comparison of RMSd between the final SA-MD structures of ssDNAs (**1** – **4**) and ssRNAs (**5** – **8**) and their respective typical A- or B-type ssDNAs or ssRNA structures reveal that A to B transition in ssDNA or B to A-transition in ssRNA affect mostly the base residues. The RMSd values for sugar residues are found to be more than that of backbone in ssDNA and ssRNA (Table 5 and 6). Furthermore, comparison of RMSd values in Tables 5-8 suggests that the NMR/MD derived structures of all ssDNAs and ssRNAs do not adopt typical B-type ssDNA or A-type ssRNA.

(F) ***The base-base stacking patterns at each dinucleotide step in ssDNA and ssRNA from NMR-derived MD structures.***

(i) Stacking interactions between neighboring nucleobases have been found to be sequence-dependent in ssDNAs and ssRNAs (Fig 8).

(ii) In case of purine rich sequences, [d($^{5'}C^1A^2A^3G^4A^5A^6C^7$) (**1**), r($^{5'}C^1A^2A^3G^4A^5A^6C^7$) (**5**)] the purine-purine [A^2-A^3, A^3-G^4, G^4-A^5] stacking interactions in ssDNAs are different compared to those in the isosequential ssRNAs: In ssDNA [d($^{5'}C^1A^2A^3G^4A^5A^6C^7$) (**1**)] imdazole of a purine (A,G) stacks over the imdazole part of the preceding purine nucleobase (A,G) at 3'-end but the pyrimidine part partially stacks over the pyrimidine part of 3'-nearest neighbor purine (Fig 8): In ssRNA [r($^{5'}C^1A^2A^3G^4A^5A^6C^7$) (**5**)] two types of purine-purine stacking interactions have been observed. (a) Towards 5'-end, the pyrimidine moiety of $\mathbf{G^4}$ stacks over the pyrimidine of $\mathbf{A^3}$, again pyrimidine of $\mathbf{A^3}$ stacks over pyrimidine of $\mathbf{A^2}$. (b) In contrast, towards the 3'-end, pyrimidine of $\mathbf{G^4}$ stacks over the imidazole of $\mathbf{A^5}$, the pyrimidine of $\mathbf{A^5}$ stacks over imidazole of $\mathbf{A^6}$. Interestingly, earlier we have found that the imidazole of purine stacked over the pyrimidine of the nearest-neighbor purine towards the 3'-end in hexameric ssDNA (*40*) whereas, pyrimidine of purine stacked over the imidazole of another purine towards 3'-end in hexameric ssRNA. *This shows that the purine-purine stacking pattern in the purine rich sequences depends upon the sequence context* (Figure 8). (c) The central trinucleotidic stacking pattern in d($^{5'}C^1A^2\underline{A^3}\underline{G^4}\underline{A^5}A^6C^7$) (**1**) is imidazole ($A^3$)-imidazole ($G^4$)-imidazole ($A^5$). But in isosequential ssRNA it is pyrimidine (A^3)-pyrimidine (G^4)- imidazole (A^5). (d) The pyrimidine of 1-Cytosinyl stacks over the imidazole part of purine (A/G) in d($^{5'}C^1A^2C^3G^4C^5A^6C^7$) (**4**), and the pyrimidine

part of purine stacks over the pyrimidine (1-cytosinyl) from 5'- to 3'-end (Figure 8). (e) In the same isosequential ssRNA sequence r($^{5'}C^1A^2C^3G^4C^5A^6C^7$), pyrimidine-$C^3$ stacks over the imdazole-G^4, and pyrimidine-C^5 stacks over the imidazole-A^6. No pyrimidine (A/G)-pyrimidine (C) stacking has been observed. (f) In d($^{5'}C^1A^2C^3G^4A^5A^6C^7$) (**3**), it has found that purine-purine stacking patterns in dinucleotide steps differ depending upon the sequence context of the dinucleotides. Thus, imidazole (G^4) stacks partially over the imidazole-A^5. Similarly pyrimidine-G^4 stacks over the pyrimidine (A^5) partially. But imidazole (A^5) stacks fully over the imidazole of (A^6) whereas, pyrimidine part (A^5) stacks partially over the pyrimidine of (A^6). In contrast, purine-purine stacking patterens in r($^{5'}C^1A^2C^3G^4A^5A^6C^7$) (**7**) are found to be the same as those in ssRNA (**5**). (g) In d($^{5'}C^1A^2A^3G^4C^5A^6C^7$) (**2**), purine-purine stacking patterns are found to be dependent on the dinucleotide sequence context: Interestingly, imidazole of A^2 stacks over the imidazole of A^3 and pyrimidine A^2 part partially stacks over pyrimidine of A^3 [purine rich ssDNA (**1**) type]. On the contrary, imidazole-A^3 does not stack over the imidazole of G^4 but pyrimidine part (A^3) fully stacks over pyrimidine of G^4 [A^2 – G^4 purine-purine stacking in ssRNA (**5**)]. In r($^{5'}C^1A^2A^3G^4C^5A^6C^7$) (**6**), pyrimidine of A^2 stacks over the imidazole of A^3 and pyrimidine A^3 stacks over imidazole of G^4 [G^4 – A^6 , purine-purine stacking in ssRNA (**5**)].

(h) All the [H8/6$N_{(n)}$ ↔ H1'$N_{(n-1)}$/ H3'$N_{(n-1)}$, N = A, G, C] distances in different isosequential pairs of ssDNA (**1** – **4**) and ssRNA (**5** – **8**) change depending upon the sequence context of the single stranded nucleic acids (Figure 8).

(G) ***Conformational Analysis for final ssDNA and ssRNA structures obtained after 2ns of NMR/MD***

The NMR-MD derived structures of ssDNAs and ssRNAs show the following preferences of helical parameters and dihedral angles. This shows that the preferred conformational hyperspace for four isosequential pairs of ssDNA and ssRNAs are indeed different from and typical B-DNA and A-RNA duplex structures (*49*). Dihedral angles found in all the four pairs of isosequential ssDNAs and ssRNAs structures derived from NMR constrained MD show that they do not adopt typical B-type DNA or A-type RNA forms (Table 11 and Figure 9). Analysis of base step parameters (*49*) also shows that the ssDNAs (**1** - **4**)/ssRNAs (**5** – **8**) do not adopt typical native duplex-type structures (Figure S52 in SI).

α [$_{(n-1)}O_{5'}$-P- $O_{3'}$-$C_{5'}$]: For all ssDNA sequences it has been found that dihedral angle α spreads between 280°-300° (except A^2 in [d($^{5'}C^1A^2A^3G^4A^5A^6C^7$)]), wheras it is ≈319° in pure B-type double stranded DNA (Figure 9). Similarly, for all ssRNA sequences it has been found that α exist in between 240°-300°, (except α of A^5 in [r($^{5'}C^1A^2A^3G^4A^5A^6C^7$]), whereas it exists 292° in typical A-type double stranded RNA (Fig 9).

β [P - $O_{5'}$ - $C_{5'}$ – $C_{4'}$]: For all ssDNA sequences dihedral angle β has been found to remain between 140°-190° (β^{trans}), whereas it exists 136° in pure B-type double stranded DNA. Again, for all ssRNA sequences it has been found that dihedral angle β exists between 160°-190° (β^{trans}), [except β of A^3 in r($^{5'}C^1A^2A^3G^4A^5A^6C^7$), β of A^5 in r($^{5'}C^1A^2C^3G^4A^5A^6C^7$), β of C^5 in r($^{5'}C^1A^2C^3G^4C^5A^6C^7$)], whereas it remains 178° in pure A-type double stranded RNA (Figure 9).

γ [$O_{5'}$- $C_{5'}$ – $C_{4'}$- $C_{3'}$]: This has been found to be interesting that most of the dihedral angles γ in ssDNA and ssRNA remain 40°-60° and 45°-74° respectively which are almost matching with the γ values found in double stranded B-type DNA (38°) and A-type RNA (54°). These values of angles show that γ dihedral exist in preferably γ+ orientation in ssDNA and ssRNA as like as in dsDNA and dsRNA. γ^{trans} orientations have been found in A^2 in [d($^{5'}C^1A^2A^3G^4A^5A^6C^7$)] and C^1 residues in [d($^{5'}C^1A^2C^3G^4C^5A^6C^7$)] respectively. Similarly, γ^{trans} orientations have been found in C^1 and A^5 residues in [r($^{5'}C^1A^2A^3G^4A^5A^6C^7$)] (Figure 9).

δ [$C_{5'}$ – $C_{4'}$- $C_{3'}$-$O_{3'}$]: δ angle values for all ssDNA sequences have been found to exist within 120°-160° whereas δ value (Figure 9).was found to be 139° in B-type dsDNA [except A^2, A^4 in [d($^{5'}C^1A^2C^3G^4A^5A^6C^7$), A^6 in [r($^{5'}C^1A^2C^3G^4C^5A^6C^7$).]: δ for all ssRNA sequences have been found to exist within 75°-90° (except C^1 r($^{5'}C^1A^2A^3G^4C^5A^6C^7$), A^2 r($^{5'}C^1A^2A^3G^4A^5A^6C^7$), C^3 and G^4 in r($^{5'}C^1A^2C^3G^4A^5A^6C^7$) and G^4 in r($^{5'}C^1A^2C^3G^4C^5A^6C^7$) .

ε [$C_{4'}$ – $C_{3'}$- $C_{3'}$-P]: In both B-type and A-type duplexes ε_t is the favorable conformation. The ssDNA and ssRNA heptamers show ε_t conformation with a few exceptions: C^1 and G^4 in d($^{5'}C^1A^2A^3G^4A^5A^6C^7$), C^1, A^2, A^3, and G^4 in d($^{5'}C^1A^2A^3G^4C^5A^6C^7$), and C^1 in r($^{5'}C^1A^2A^3G^4C^5A^6C^7$) and A^2 r($^{5'}C^1A^2A^3G^4A^5A^6C^7$) where the residues exist in ε^- conformation (Figure 9).

ζ [$C_{3'}$-$O_{3'}$-P-$O_{5'(n+1)}$]: All residues in DNA and RNA duplexes are found to be existing in 203° and 289°. Similarly, ζ values in most of the residues in ssDNA and ssRNA with different sequence context have been found to be ranging from 250° - 300° and 270°- 300° respectively. Exceptions are found in ζ values of A^2, G^4 in d($^{5'}C^1A^2A^3G^4A^5A^6C^7$), and A^2, A^3, G^4 in d($^{5'}C^1A^2A^3G^4C^5A^6C^7$) in ssDNA. Similarly, ζ values of C^1 in r($^{5'}C^1A^2A^3G^4C^5A^6C^7$), A^2 in r($^{5'}C^1A^2A^3G^4A^5A^6C^7$), G^4 in r($^{5'}C^1A^2C^3G^4A^5A^6C^7$) and r($^{5'}C^1A^2C^3G^4C^5A^6C^7$) (Figure 9) are deflected from the ζ value in the typical A-type RNA.

χ [$O_{4'}$-$C_{1'}$-N_9-C_4 for purines and $O_{4'}$-$C_{1'}$-N_1-C_2 for pyrimidines)]: Residues in both A- and B-type duplexes are known to exist in the most favourable *anti* (χ = 180 ± 90°) conformations whereas *syn* (χ = 0 ± 90°) conformation exists in Z-type duplexes (in G in CG repeat duplexes). χ is found to remain in anti conformation in most of the residues of ssDNA and ssRNA

heptamers studied except A^2 and A^6 d($^{5'}C^1A^2A^3G^4C^5A^6C^7$), and A^5 in d($^{5'}C^1A^2A^3G^4A^5A^6C^7$), C^3 in d($^{5'}C^1A^2C^3G^4A^5A^6C^7$), C^1, A^2, C^3, G^4 in r($^{5'}C^1A^2C^3G^4A^5A^6C^7$), A^2 in r($^{5'}C^1A^2A^3G^4A^5A^6C^7$), C^1 in r($^{5'}C^1A^2A^3G^4C^5A^6C^7$) (Figure 9).

Conclusions

(1) In ssDNA and ssRNA the sugar conformations in each residue can fluctuate depending upon the sequence context.

(2) From the NMR spectra it has been found that the intra-residual cross peak intensities for the $H(8/6)_n$- $H(1'/2'/2''/H3')_n$ contacts in ssDNA and ssRNA to be stronger at 3'-ends in comparison with the 5'-ends which suggests that nucleobases at 5'-end are more dynamic in nature compared to those at the 3'-ends.

(3) RMSd calculations of final NMR-MD structures of ssDNA and ssRNA show that the 5'-end nucleobases have higher RMSd values compared to those at the 3'-end [except the sequence d($^{5'}C^1A^2A^3G^4A^5A^6C^7$) (**1**)]. It has been found to be interesting in the mixed sequences [d($^{5'}C^1A^2A^3G^4C^5A^6C^7$), r($^{5'}C^1A^2A^3G^4C^5A^6C^7$),] though the 5'-ends are more purine-rich compared to corresponding 3'-ends, the RMSd values of nucleobases at 5'-ends are more than those of the nucleobases at 3'-ends.

(4) Dinucleotide stacking pattern shows that the base-base nearest neighbor stacking interactions depend wholly upon the sequence contexts of ssDNAs (**1** – **4**) and ssRNAs (**5** – **8**).

(v) The central trinucleotidic stacking pattern in purine rich [d($^{5'}C^1A^2A^3G^4A^5A^6C^7$)] is *imidazole* ($\mathbf{A^3}$)- *imidazole* ($\mathbf{G^4}$)- *imidazole (*$\mathbf{A^5}$). But in isosequential ssRNA it is *pyrimidine* ($\mathbf{A^3}$)- *pyrimidine* ($\mathbf{G^4}$)- *imidazole* ($\mathbf{A^5}$).

(5) All the [$H8/6N_{(n)} \leftrightarrow H1'N_{(n-1)}$/ $H3'N_{(n-1)}$, N = A, G, C] distances in different isosequential pairs of ssDNA (**1** – **4**) and ssRNA (**5** – **8**) change depending upon the <u>sequence context</u> of the single stranded nucleic acids.

(6) The nOe footprints of four isosequential pairs of ssDNA and ssRNA suggest that they adopt right-handed helical forms, and do not adopt either pure B-type DNA or A-type RNA structures. The exact self-assembly of ssDNA or ssRNA to the right-handed helical form depend upon the sequence context. This means that in a long stretch of replicating ssDNA or in mRNA, there are different polymorphic structural motifs of the right-handed helical forms possible, and which may serve as recognition element for various ligand binding for specific biological functions.

Acknowledgement.

Generous financial support from the Swedish Natural Science Research Council (Vetenskapsrådet), the Swedish Foundation for Strategic Research (Stiftelsen för Strategisk Forskning) and the EU-FP6 funded RIGHT project (Project no. LSHB-CT-2004-005276) is gratefully acknowledged.

Supporting Information Available: Table S1. ^{1}H chemical shifts of aromatic protons for all ssDNA and ssRNA heptamers. **Tables S2-S9**: Constraints for Dihedrals and sugar phase angles for simulated annealing (SA) of all ssDNA and ssRNAs. **Table S10**: The oligomerization shift of all aromatic protons in all ssDNA and ssRNA heptamers with respect to monmers. **Tables S11-S14.** Inter- (n-1) and Intra- (n) residual nOe contacts for all ssDNAs and ssRNAs. **Tables S15 – S16.** Distances (Å) between aromatic (n) and sugar protons n and (n-1) of the cannonical A-RNA and B-DNA duplexes for comparison with the final NMR-SA-MD structures of ssRNA/ssDNA. **Tables S17 – S24.** Distances (Å) between aromatic (n) and sugar protons n and (n-1) of the final NMR-SA-MD structures of ssRNA/ssDNA. **Figures S1 - S49.** NOESY footprints, DQF-COSY (^{31}P coupled and decoupled), TOCSY, ^{1}H-^{31}P correlation spectra of all ssDNAs and ssRNAs. **Figure S50.** RMSd of fimal ssDNAs and ssRNAs. **Figure S51.** Plots of total potential energy during NMR-MD simulation for ssDNA and ssRNA. **Figure S52.** Plots of Sugar puckering (Phase angle, *P*) and helical parameters of the final NMR-SA-MD structures of ssRNA/ssDNA (Roll, Slide, Inclination, and Nearest neighbour base atom overlap). **Figure S53.** Plots (A - H) of mass weighted RMSD for (nt1-3) [red] and (nt5-7) [blue] nucleobase residues in ssDNAs (**1** - **4**) and ssRNAs (**5** – **8**) with Time (ps) shows that 5′-ends are more dynamic compared to 3′-ends in ssDNAs and ssRNAs. All RMSD for nucleobase residues were calculated referencing the final SA structures of ssDNAs (**1** - **4**) and ssRNAs (**5** – **8**) at 100 K.

References

(1) Lehninger, A. (1993) *Principles of Biochemistry*, Second Edition ed., Worth, New York, USA.

(2) Dykxhoorn, D. M., and Lieberman, J. (2006) Knocking down disease with siRNAs. *Cell 126*, 231-5.

(3) Dykxhoorn, D. M., and Lieberman, J. (2006) Running interference: prospects and obstacles to using small interfering RNAs as small molecule drugs. *Annu Rev Biomed Eng 8*, 377-402.

(4) Borenstein, E., and Ruppin, E. (2006) Direct evolution of genetic robustness in microRNA. *Proc Natl Acad Sci U S A 103*, 6593-8.

(5) Winkler, W. C., Cohen-Chalamish, S., and Breaker, R. R. (2002) An mRNA structure that controls gene expression by binding FMN. *Proc Natl Acad Sci U S A 99*, 15908-13.

(6) Weisberg, R. A., and Storz, G. (2002) Take your vitamins with a pinch of RNA. *Mol Cell 10*, 1266-8.

(7) Svoboda, P., and Cara, A. D. (2006) Hairpin RNA: a secondary structure of primary importance. *Cell Mol Life Sci 63*, 901-8.

(8) Pataskar, S. S., Dash, D., and Brahmachari, S. K. (2001) Progressive myoclonus epilepsy [EPM1] repeat d(CCCCGCCCCGCG)n forms folded hairpin structures at physiological pH. *J Biomol Struct Dyn 19*, 293-305.

(9) Yang, W., and Van Duyne, G. D. (2006) Protein-nucleic acid interactions: from A(rgonaute) to X(PF). *Curr Opin Struct Biol 16*, 1-4.

(10) Lue, N. F., and Xia, J. (1998) Species-specific and sequence-specific recognition of the dG-rich strand of telomeres by yeast telomerase. *Nucleic Acids Res 26*, 1495-502.

(11) Eldridge, A. M., Halsey, W. A., and Wuttke, D. S. (2006) Identification of the determinants for the specific recognition of single-strand telomeric DNA by Cdc13. *Biochemistry 45*, 871-9.

(12) Morl, M., and Marchfelder, A. (2001) The final cut. The importance of tRNA 3'-processing. *EMBO Rep 2*, 17-20.

(13) Bommarito, S., Peyret, N., and SantaLucia, J., Jr. (2000) Thermodynamic parameters for DNA sequences with dangling ends. *Nucleic Acids Res 28*, 1929-34.

(14) (a) Lee, C. H., Ezra, F. S., Kondo, N. S., Sarma, R. H., and Danyluk, S. S. (1976) low resolution - Conformational properties of dinucleoside monophosphates in solution: dipurines and dipyrimidines. *Biochemistry 15*, 3627-39. (b) Lee, C. H., and Tinoco, I., Jr. (1980) Conformation studies of 13 trinucleoside diphosphates by 360 MHz PMR spectroscopy. a bulged base conformation I. base protons and H1' protons. *Biophys Chem 11*, 283-94. (c) Ezra, F. S., Lee, C. H., Kondo, N. S., Danyluk, S. S., and Sarma, R. H. (1977) Conformational properties of purine-pyrimidine and pyrimidine-purine dinucleoside monophosphates. *Biochemistry 16*, 1977-87. (d) Doornbos, J., den Hartog, J. A., van Boom, J. H., and Altona, C. (1981) Conformational analysis of the nucleotides A2'-5'A, A2'-5'A2'-5'A and A2'-5'U from nuclear magnetic resonance and circular dichroism studies. *Eur J Biochem 116*, 403-12. (e) Stone, M. P., Johnson, D. L., and Borer, P. N. (1981) Unusual structures in single-stranded ribonucleic acid: proton nuclear magnetic resonance of AUCCA in deuterium oxide. *Biochemistry 20*, 3604-10. (f) van den Hoogen, Y. T., Treurniet, S. J., Roelen, H. C., de Vroom, E., van der Marel, G. A., van Boom, J. H., and Altona, C. (1988) Conformational analysis of the tetranucleotides m6(2)A-m6(2)A-U-m6(2)A(m6(2)A = N6-dimethyladenosine) and U-m6(2)A-U-m6(2)A and of the hybrid dA-r(U-A). A one- and two-dimensional NMR study. *Eur J Biochem 171*, 155-62. (g) Kondo, N. S., Holmes, H. M., Stempel, L. M., and Ts'o, O. P. (1970) Influence of the phosphodiester linkage (3'-5', 2'-5', and 5'-5') on the conformation of dinucleoside monophosphate. *Biochemistry 9*, 3479-98. (h) Ts'o, P. O., Kondo, N. S., Schweizer, M. P., and Hollis, D. P. (1969) Studies of the conformation and interaction in dinucleoside mono- and diphosphates by proton magnetic resonance. *Biochemistry 8*, 997-1029.

(15) Powell, J. T., Richards, E. G., and Gratzer, W. B. (1972) The nature of stacking equilibria in polynucleotides. *Biopolymers 11*, 235-50.

(16) Kovtun, I. V., Therneau, T. M., and McMurray, C. T. (2000) Gender of the embryo contributes to CAG instability in transgenic mice containing a Huntington's disease gene. *Hum Mol Genet 9*, 2767-75.

(17) Kantarjian, H. M., and Cortes, J. (2006) New strategies in chronic myeloid leukemia. *Int J Hematol 83*, 289-93.

(18) Bhattacharyya, D., and Bansal, M. (1994) Analysis of sequence dependent variations in secondary and tertiary structure of tRNA molecules. *J Biomol Struct Dyn 11*, 1251-75.

(19) Cech, T. R., Zaug, A. J., and Grabowski, P. J. (1981) In vitro splicing of the ribosomal RNA precursor of Tetrahymena: involvement of a guanosine nucleotide in the excision of the intervening sequence. *Cell 27*, 487-96.

(20) Cech, T. R. (1990) Self-splicing of group I introns. *Annu Rev Biochem 59*, 543-68.

(21) Shan, S. O., and Herschlag, D. (2000) An unconventional origin of metal-ion rescue and inhibition in the Tetrahymena group I ribozyme reaction. *Rna 6*, 795-813.

(22) Takagi, Y., Warashina, M., Stec, W. J., Yoshinari, K., and Taira, K. (2001) Recent advances in the elucidation of the mechanisms of action of ribozymes. *Nucleic Acids Res 29*, 1815-34.

(23) Sontheimer, E. J., Gordon, P. M., and Piccirilli, J. A. (1999) Metal ion catalysis during group II intron self-splicing: parallels with the spliceosome. *Genes Dev 13*, 1729-41.

(24) Podar, M., Chu, V. T., Pyle, A. M., and Perlman, P. S. (1998) Group II intron splicing in vivo by first-step hydrolysis. *Nature 391*, 915-8.

(25) Padgett, R. A., Podar, M., Boulanger, S. C., and Perlman, P. S. (1994) The stereochemical course of group II intron self-splicing. *Science 266*, 1685-8.

(26) Guerrier-Takada, C., Gardiner, K., Marsh, T., Pace, N., and Altman, S. (1983) The RNA moiety of ribonuclease P is the catalytic subunit of the enzyme. *Cell 35*, 849-57.

(27) Ferre-D'Amare, A. R., Zhou, K., and Doudna, J. A. (1998) Crystal structure of a hepatitis delta virus ribozyme. *Nature 395*, 567-74.

(28) Nakano, S., Chadalavada, D. M., and Bevilacqua, P. C. (2000) General acid-base catalysis in the mechanism of a hepatitis delta virus ribozyme. *Science 287*, 1493-7.

(29) Ohmichi, T., and Kool, E. T. (2000) The virtues of self-binding: high sequence specificity for RNA cleavage by self-processed hammerhead ribozymes. *Nucleic Acids Res 28*, 776-83.

(30) Comolli, L. R., Pelton, J. G., and Tinoco, I., Jr. (1998) Mapping of a protein-RNA kissing hairpin interface: Rom and Tar-Tar*. *Nucleic Acids Res 26*, 4688-95.

(31) Jaschke, A. (2001) Artificial ribozymes and deoxyribozymes. *Curr Opin Struct Biol 11*, 321-6.

(32) Hunter, C. A., and Sanders, J. K. M. (1990) The nature of p-p interactions. *Journal of the American Chemical Society 112*, 5525-34.

(33) Ribas, J., Cubero, E., Luque, F. J., and Orozco, M. (2002) Theoretical study of alkyl-pi and aryl-pi interactions. Reconciling theory and experiment. *J Org Chem 67*, 7057-65.

(34) Vesnaver, G., and Breslauer, K. J. (1991) The contribution of DNA single-stranded order to the thermodynamics of duplex formation. *Proc Natl Acad Sci U S A 88*, 3569-73.

(35) Warshaw, M. M., and Tinoco, I., Jr. (1965) Absorption and optical rotatory dispersion of six dinucleoside phosphates. *J Mol Biol 13*, 54-64.

(36) Puglisi, J. D., Wyatt, J. R., and Tinoco, I., Jr. (1990) Solution conformation of an RNA hairpin loop. *Biochemistry 29*, 4215-26.

(37) Auffinger, P., and Westhof, E. (1997) Rules governing the orientation of the 2'-hydroxyl group in RNA. *J Mol Biol 274*, 54-63.

(38) Varani, G., Cheong, C., and Tinoco, I., Jr. (1991) Structure of an unusually stable RNA hairpin. *Biochemistry 30*, 3280-9.

(39) Thibaudeau, C., and Chattopadhyaya, J. (1999) *Stereoelectronic Effects in Nucleosides and Nucleotides and their Structural Studies*, Uppsala University, Uppsala, Sweden.

(40) Isaksson, J., Acharya, S., Barman, J., Cheruku, P., and Chattopadhyaya, J. (2004) Single-stranded adenine-rich DNA and RNA retain structural characteristics of their respective double-stranded conformations and show directional differences in stacking pattern. *Biochemistry 43*, 15996-6010.

(41) Acharya, S., Barman, J., Cheruku, P., Chatterjee, S., Acharya, P., Isaksson, J., and Chattopadhyaya, J. (2004) Significant pKa perturbation of nucleobases is an intrinsic property of the sequence context in DNA and RNA. *J Am Chem Soc 126*, 8674-81.

(42) Acharya, P., Cheruku, P., Chatterjee, S., Acharya, S., and Chattopadhyaya, J. (2004) Measurement of nucleobase pKa values in model mononucleotides shows RNA-RNA duplexes to be more stable than DNA-DNA duplexes. *J Am Chem Soc 126*, 2862-9.

(43) Acharya, P., Acharya, S., Cheruku, P., Amirkhanov, N. V., Foldesi, A., and Chattopadhyaya, J. (2003) Cross-modulation of the pKa of nucleobases in a single-stranded hexameric-RNA due to tandem electrostatic nearest-neighbor interactions. *J Am Chem Soc 125*, 9948-61.

(44) Acharya, P., Acharya, S., Foldesi, A., and Chattopadhyaya, J. (2003) Tandem electrostatic effect from the first to the third aglycon in the trimeric RNA owing to the nearest-neighbor interaction. *J Am Chem Soc 125*, 2094-100.

(45) Acharya, S., Acharya, P., Foldesi, A., and Chattopadhyaya, J. (2002) Cross-modulation of physicochemical character of aglycones in dinucleoside (3'-->5') monophosphates by the nearest neighbor interaction in the stacked state. *J Am Chem Soc 124*, 13722-30.

(46) Barman, J., Acharya, S., Zhou, C., Chatterjee, S., Engstrom, A., and Chattopadhyaya, J. (2006) Non-identical electronic characters of the internucleotidic phosphates in RNA modulate the chemical reactivity of the phosphodiester bonds. *Org Biomol Chem 4*, 928-41.

(47) Saenger, W. (1983) in *Springer Advanced Texts in Chemistry* (Cantor, C., Ed.), Springer-Verlag, New York, USA.

(48) Rinkel, L. J., and Altona, C. (1987) Conformational analysis of the deoxyribofuranose ring in DNA by means of sums of proton-proton coupling constants: a graphical method. *J Biomol Struct Dyn 4*, 621-49.

(49) Blackburn, G., and Gait, M. (1996), Oxford University Press, New York.

(50) Cheatham, T. E., III, and Kollman, P. A. (1997) Molecular Dynamics Simulations Highlight the Structural Differences among DNA:DNA, RNA:RNA, and DNA:RNA Hybrid Duplexes. *Journal of the American Chemical Society 119*, 4805-4825.

Table 1: $^{3}J_{H1'2'}$ and (%) North-type conformations in each and every nucleotide residue in ssRNA

Sequences	C^1		A^2		$A^3/A^3/C^3/C^3$		G^4		$A^5/A^5/C^5/C^5$		A^6		C^7	
	$J_{1'2'}$	% N	$J_{1'2'}$	% N	$J_{1'2'}$	% N	$J_{1'2'}$	% N	$J_{1'2'}$	% N	$J_{1'2'}$	% N	$J_{1'2'}$	% N
r($^{5'}C^1A^2A^3G^4A^5A^6C^7$)	2.86	73	3.07	70	3.42	65	2.93	72	2.17	83	1.97	86	2.73	75
r($^{5'}C^1A^2A^3G^4C^5A^6C^7$)	2.93	72	2.17	83	2.73	75	3.28	67	2.17	83	2.52	78	2.86	73
r($^{5'}C^1A^2C^3G^4A^5A^6C^7$)	1.97	86	2.73	75	4.17	54	4.17	54	2.52	78	2.31	81	2.73	75
r($^{5'}C^1A^2C^3G^4C^5A^6C^7$)	2.66	76	2.52	78	3.07	70	4.80	45	3.21	68	2.52	78	2.86	73

Table 2: $\Sigma^{3}J_{H1'2',\,H1'2''}$ and (%) North-type conformations in each and every nucleotide residue in ssDNA

Sequences	C^1		A^2		$A^3/A^3/C^3/C^3$		G^4		$A^5/A^5/C^5/C^5$		A^6		C^7	
	Σ^3J	% N	Σ^3J	% N	Σ^3J	% N	Σ^3J	% N	Σ^3J	% N	Σ^3J	% N	Σ^3J	% N
d($^{5'}C^1A^2A^3G^4A^5A^6C^7$)	15.29	7	12.52	54	13.22	42	14.11	27	14.23	25	14.93	14	14.87	25
d($^{5'}C^1A^2A^3G^4C^5A^6C^7$)	15.52	3	11.16	77	11.16	77	12.51	54	15.64	1	15.11	10	15.17	9
d($^{5'}C^1A^2C^3G^4A^5A^6C^7$)	15.64	1	11.69	68	14.93	13	12.63	52	13.93	30	14.76	16	15.64	1
d($^{5'}C^1A^2C^3G^4C^5A^6C^7$)	14.64	18	14.93	13	15.64	1	14.46	21	15.64	1	15.64	1	14.82	15

Table 3:All distance and dihedral constraints used in the NMR-SA/MD and NMR-MD steps.

Heptamer Sequence	Distance Constraints	Dihedral Constraints		Total Constraints
		Phosphate Backbone	Sugar	
d($^{5'}C^1A^2A^3G^4A^5A^6C^7$) (**1**)	80	19	45	144
d($^{5'}C^1A^2A^3G^4C^5A^6C^7$) (**2**)	63	19	45	127
d($^{5'}C^1A^2C^3G^4A^5A^6C^7$) (**3**)	77	19	45	141
d($^{5'}C^1A^2C^3G^4C^5A^6C^7$) (**4**)	108	19	45	172
r($^{5'}C^1A^2A^3G^4A^5A^6C^7$) (**5**)	55	19	45	119
r($^{5'}C^1A^2A^3G^4C^5A^6C^7$) (**6**)	62	19	45	126
r($^{5'}C^1A^2C^3G^4A^5A^6C^7$) (**7**)	58	19	45	122
r($^{5'}C^1A^2C^3G^4C^5A^6C^7$) (**8**)	60	19	45	124

Table 4: Inter-residual 2D NOESY cross peak intensities for $H8_{(n)}/H6_{(n)}$- H1'(n-1) and $H8_{(n)}/H6_{(n)}$- H3'(n-1) for pairs of isosequential ssDNAs (**1-4**) and ssRNAs (**5-8**). For B-type strand intensity of $H8_{(n)}/H6_{(n)}$- H1'(n-1) **(a)** is stronger compared to that of $H8_{(n)}/H6_{(n)}$- H3'(n-1) **(b)**. For A-type strand intensity of $H8_{(n)}/H6_{(n)}$- H3'(n-1) **(b)** is stronger compared to that of $H8_{(n)}/H6_{(n)}$- H1'(n-1) **(a).**

nt Steps	nOe Contact	($^{5'}C^1A^2A^3G^4A^5A^6C^7$)		($^{5'}C^1A^2A^3G^4C^5A^6C^7$)		($^{5'}C^1A^2C^3G^4A^5A^6C^7$)		($^{5'}C^1A^2C^3G^4C^5A^6C^7$)	
		ssDNA(1)*	ssRNA (5)*	ssDNA (2) ⊕	ssRNA (6) ⊕	ssDNA (3) #	ssRNA (7)#	ssDNA (4)ϒ	ssRNA (8)ϒ
1 - 2	H6/H8 $_{(n)}$ -H1' $_{(n-1)}$	-	W(1a)	-	-	-	W (1a)	-	W (1a)
	H6/H8 $_{(n)}$ -H3' $_{(n-1)}$	-	S (1b)	-	-	-	S (1b)	-	S (1b)
2-3	H6/H8 $_{(n)}$ -H1' $_{(n-1)}$	S (1a)	M (2a)	S (1a)	M (1a)	M (1a)	M (2a)	M (1a)	M (2a)
	H6/H8 $_{(n)}$ -H3' $_{(n-1)}$	M (1b)	S (2b)	W (1b)	S (1b)	M (1b)	W (2b)	W (1b)	S (2b)
3-4	H6/H8 $_{(n)}$ -H1' $_{(n-1)}$	S (2a)	M (3a)	M (2a)	M (2a)	-	S (3a)	W (2a)	M (3a)
	H6/H8 $_{(n)}$ -H3' $_{(n-1)}$	M (2b)	S (3b)	W (2b)	M (2b)	-	W (3b)	M (2b)	S (3b)
4-5	H6/H8 $_{(n)}$ -H1' $_{(n-1)}$	S (3a)	M (4a)	S (3a)	S (3a)	M (2a)	S (4a)	S (3a)	S (4a)
	H6/H8 $_{(n)}$ -H3' $_{(n-1)}$	M (3b)	M (4a)	M (3b)	S (3b)	W (2b)	-	M (3b)	S (4b)
5-6	H6/H8 $_{(n)}$ -H1' $_{(n-1)}$	S (4a)	M (4a)	-	W(4a)	M (3a)	M (5a)	W (4a)	W (5a)
	H6/H8 $_{(n)}$ -H3' $_{(n-1)}$	M(4b)	S (4b)	W(4b)	S (4b)	M (3b)	S (5b)	M (4b)	S (5b)
6-7	H6/H8 $_{(n)}$ -H1' $_{(n-1)}$	S (5a)	M (5a)	-	W(5a)	M (4a)	M (6a)	W (5a)	S (6a)
	H6/H8 $_{(n)}$ -H3' $_{(n-1)}$	M(5b)	S (5b)	-	M(5b)	W (4b)	S (6b)	W (5b)	S (6b)

1a – 5a: signify $H8_{(n)}/H6_{(n)}$- H1'(n-1) nOe contact
1b – 5b: signify $H8_{(n)}/H6_{(n)}$- H3'(n-1) nOe contact
W – weak (3.5 – 6.5 Å), **M** - medium (3.5 – 5.0 Å), **S** - strong (1.8 – 3.0Å)
NOESY cross-peaks in Panel A and Panel E in Figure S51.
⊕ NOESY cross-peaks in Panel B and Panel F in Figure S51.
NOESY cross-peaks in Panel C and Panel G in Figure S51.

Table 5: Comparison of RMSd (Å) of NMR-SA/MD structures for ssDNA (**1 - 4**) with the staring A type ssDNA structures and the B-type ssDNA structures.

ssDNA (1 – 4)	RMSd (Å) with A-type DNA≠				RMSd (Å) with B-type DNA*			
	Backbone	Base	Sugar	Heavy Atom	Backbone	Base	Sugar	Heavy Atom
d($^{5'}C^1A^2A^3G^4A^5A^6C^7$) (**1**)	3.06	6.17	3.70	**5.10**	2.91	3.64	2.86	**2.90**
d($^{5'}C^1A^2A^3G^4C^5A^6C^7$) (**2**)	3.21	7.99	4.10	**6.34**	3.06	6.35	3.44	**3.00**
d($^{5'}C^1A^2C^3G^4A^5A^6C^7$) (**3**)	2.25	5.04	2.87	**3.90**	3.11	3.41	3.21	**3.10**
d($^{5'}C^1A^2C^3G^4C^5A^6C^7$) (**4**)	2.62	6.20	3.55	**4.40**	2.29	4.59	2.76	**3.00**

* B type DNA contains C2'-*endo*-C3'-*exo* sugars(*47*)
≠ A type DNA contains C3'-*endo*-C2'-*exo* sugars (*47*)

Table 6: Comparison of RMSd (Å) of NMR-SA/MD structures for ssRNA (**5 – 8**) with respect to their starting B-type ssRNA*/ A-type ssRNA≠ structures

ssRNA (5 – 8)	RMSd (Å) with A-type RNA≠				RMSd (Å) with B-type RNA*			
	Backbone	Base	Sugar	Heavy Atom	Backbone	Base	Sugar	Heavy Atom
r($^{5'}C^1A^2A^3G^4A^5A^6C^7$) (**5**)	1.53	4.39	2.18	**2.00**	1.19	1.99	1.20	**4.00**
r($^{5'}C^1A^2A^3G^4C^5A^6C^7$) (**6**)	2.71	5.57	3.08	**4.60**	2.46	4.08	2.85	**4.70**
r($^{5'}C^1A^2C^3G^4A^5A^6C^7$) (**7**)	3.14	7.40	3.76	**3.60**	2.97	4.61	3.34	**6.50**
r($^{5'}C^1A^2C^3G^4C^5A^6C^7$) (**8**)	3.99	6.85	4.20	**7.50**	3.81	8.75	4.75	**5.60**

* B type RNA contains C2'-*endo*-C3'-*exo* sugars. B-type RNA structures were constructed using Nucgen module in Amber 7.0 (for details see the experimental section).
≠ A type DNA contains C3'-*endo*-C2'-*exo* sugars (*47*)

Table 7: Comparison of RMSd (Å) of final NMR- MD structures (after 2ns MD) for ssDNA (**1 - 4**) with the A-type and the B-type ssDNA structures.

ssDNA (1 – 4)	RMSd (Å) with A-type DNA≠				RMSd (Å) with B-type DNA*			
	Backbone	Base	Sugar	Heavy Atom	Backbone	Base	Sugar	Heavy Atom
d($^{5'}C^1A^2A^3G^4A^5A^6C^7$) (**1**)	2.80	5.40	3.50	**4.30**	2.60	3.20	2.60	**2.60**
d($^{5'}C^1A^2A^3G^4C^5A^6C^7$) (**2**)	3.40	9.50	4.50	**8.40**	3.30	6.40	3.60	**5.30**
d($^{5'}C^1A^2C^3G^4A^5A^6C^7$) (**3**)	1.70	5.00	2.30	**4.50**	2.90	3.00	2.90	**2.80**
d($^{5'}C^1A^2C^3G^4C^5A^6C^7$) (**4**)	1.80	4.00	2.30	**3.50**	1.70	3.20	1.80	**2.70**

* B type DNA contains C2'-*endo*-C3'-*exo* sugars(*47*)
≠ A type DNA contains C3'-*endo*-C2'-*exo* sugars(*47*)

Table 8: Comparison of RMSd (Å) of final NMR- MD structures (after 2ns MD) for ssRNA (**1** - **4**) with the A-type and the B-type ssRNA structures.

ssRNA (5 – 8)	RMSd (Å) with A-type RNA≠				RMSd (Å) with B-type RNA*			
	Backbone	Base	Sugar	Heavy Atom	Backbone	Base	Sugar	Heavy Atom
r($^{5'}C^1A^2A^3G^4A^5A^6C^7$) (**5**)	1.50	2.80	1.80	**2.50**	1.60	3.40	2.00	**3.20**
r($^{5'}C^1A^2A^3G^4C^5A^6C^7$) (**6**)	2.40	3.50	2.90	**2.40**	2.70	5.20	3.20	**4.80**
r($^{5'}C^1A^2C^3G^4A^5A^6C^7$) (**7**)	1.50	5.50	2.10	**4.70**	1.20	3.50	1.60	**2.30**
r($^{5'}C^1A^2C^3G^4C^5A^6C^7$) (**8**)	2.00	5.40	2.70	**4.70**	2.00	2.80	2.40	**2.30**

* B-type RNA contains C2'-*endo*-C3'-*exo* sugars. B-type RNA structures were constructed using Nucgen module in Amber 7.0 (for details see the experimental section).
≠ A-type DNA contains C3'-*endo*-C2'-*exo* sugars (*47*)

Table 9: RMSd between the final converged 10 NMR constrained MD structures of ssDNA/ssRNA, taken from the last 100 ps of the total 2 ns MD run (initial structures for MD were taken from the final SA structure).

Sequence	RMSd (Å)	
	ssDNA (**1** – **4**)	ssRNA (**5** – **8**)
$^{5'}C^1A^2A^3G^4A^5A^6C^7$	**0.3**	**0.8**
$^{5'}C^1A^2A^3G^4C^5A^6C^7$	**0.6**	**0.1**
$^{5'}C^1A^2C^3G^4A^5A^6C^7$	**0.2**	**0.5**
$^{5'}C^1A^2C^3G^4C^5A^6C^7$	**1.0**	**0.4**

Table 10: Mass weighted averaged RMSd values for the 5′ and 3′ nucleobases' in last 100ps MD structures of ssDNA/ssRNA within the last 100 ps of the converged region.

Sequence	RMSd (Å)			
	ssDNA (**1** – **4**)		ssRNA (**5** – **8**)	
	5′(1-3 nt)	3′(5-7 nt)	5′(1-3 nt)	3′ (5-7 nt)
$^{5'}C^1A^2A^3G^4A^5A^6C^7$	**1.45***	**1.55***	**6.57**	**0.98**
$^{5'}C^1A^2A^3G^4C^5A^6C^7$	**5.20**	**1.31**	**2.39**	**1.04**
$^{5'}C^1A^2C^3G^4A^5A^6C^7$	**5.94**	**2.39**	**4.16**	**1.60**
$^{5'}C^1A^2C^3G^4C^5A^6C^7$	**4.25**	**1.78**	**3.97**	**1.46**

* there is no difference of RMSd values found for 5′ and 3′ nucleobases in ssDNA (1)

Table 11: Comparison of dihedral angles (ϕ) of ssDNA and ssRNA obtained from the final NMR-MD structures after 2ns MD run with dihedral angles in the typical B-type dsDNA and A type dsRNA duplex, Standard dihedral parameters were taken from (*49*).

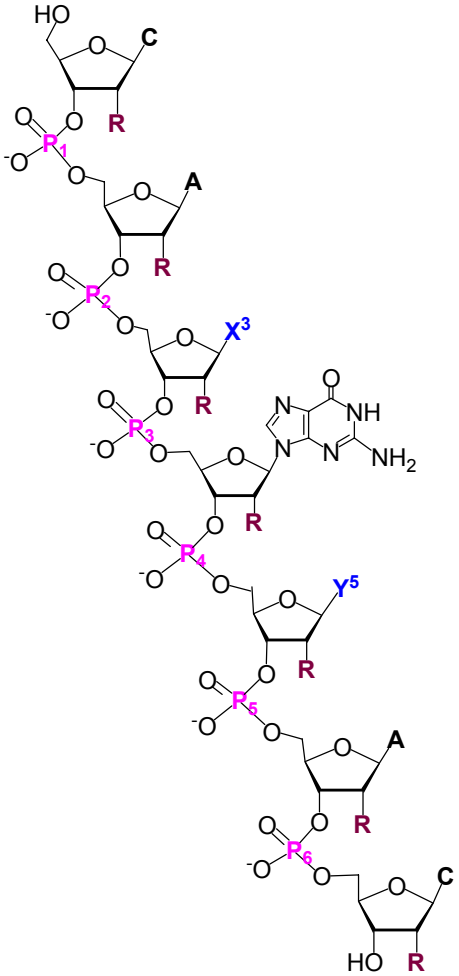
Sequence	α	β	γ	δ	ε	ζ	χ
A-RNA(Double Helix)	292	178	54	82	207	289	202
r($^{5'}C^1A^2A^3G^4A^5A^6C^7$)	137-288	138-188	47-187	73-147	188-279	136-300	196-280
r($^{5'}C^1A^2A^3G^4C^5A^6C^7$)	238-288	166-176	53-60	79-143	208-285	93-312	50-210
r($^{5'}C^1A^2C^3G^4A^5A^6C^7$)	283-293	143-178	57-63	75-146	184-205	195-289	223-293
r($^{5'}C^1A^2C^3G^4C^5A^6C^7$)	287-297	125-183	53-65	73-148	184-229	181-211	197-253
B-DNA(Double Helix)	319	136	38	139	227	203	258
d($^{5'}C^1A^2A^3G^4A^5A^6C^7$)	81-301	165-189	45-195	113-155	122-183	67-348	177-218
d($^{5'}C^1A^2A^3G^4C^5A^6C^7$)	281-289	126-175	42-54	134-148	207-280	140-295	227-298
d($^{5'}C^1A^2C^3G^4A^5A^6C^7$)	283-300	169-179	50-69	82-147	175-205	266-290	229-277
d($^{5'}C^1A^2C^3G^4C^5A^6C^7$)	290-298	166-179	49-188	90-144	180-188	271-281	223-257

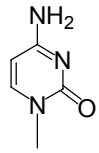
(**1**): R = H, $X^3 = Y^5 = A$
d($^{5'}C^1p_1A^2p_2A^3p_3\underline{G}^4p_4A^5p_5A^6p_6C^7$)
(**5**): R = OH, $X^3 = Y^5 = A$
r($^{5'}C^1p_1A^2p_2A^3p_3\underline{G}^4p_4A^5p_5A^6p_6C^7$)

(**2**): R = H, $X^3 = A$, $Y^5 = C$
d($^{5'}C^1p_1A^2p_2A^3p_3\underline{G}^4p_4C^5p_5A^6p_6C^7$)
(**6**): R = OH, $X^3 = A$, $Y^5 = C$
r($^{5'}C^1p_1A^2p_2A^3p_3\underline{G}^4p_4C^5p_5A^6p_6C^7$)

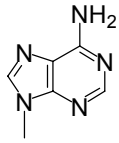
(**3**): R = H, $X^3 = C$, $Y^5 = A$
d($^{5'}C^1p_1A^2p_2C^3p_3\underline{G}^4p_4A^5p_5A^6p_6C^7$)
(**7**): R = OH, $X^3 = C$, $Y^5 = A$
r($^{5'}C^1p_1A^2p_2C^3p_3\underline{G}^4p_4A^5p_5A^6p_6C^7$)

(**4**): R = H, $X^3 = Y^5 = C$
d($^{5'}C^1p_1A^2p_2C^3p_3\underline{G}^4p_4C^5p_5A^6p_6C^7$)
(**8**): R = OH, $X^3 = Y^5 = C$
r($^{5'}C^1p_1A^2p_2C^3p_3\underline{G}^4p_4C^5p_5A^6p_6C^7$)

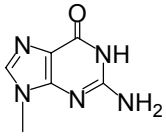




1-Cytosinyl (C)



9-Adeninyl (A)



9-Guaninyl (G)

Figure 1: Heptameric ssDNAs (**1** - **4**) and ssRNAs (**5** - **8**).

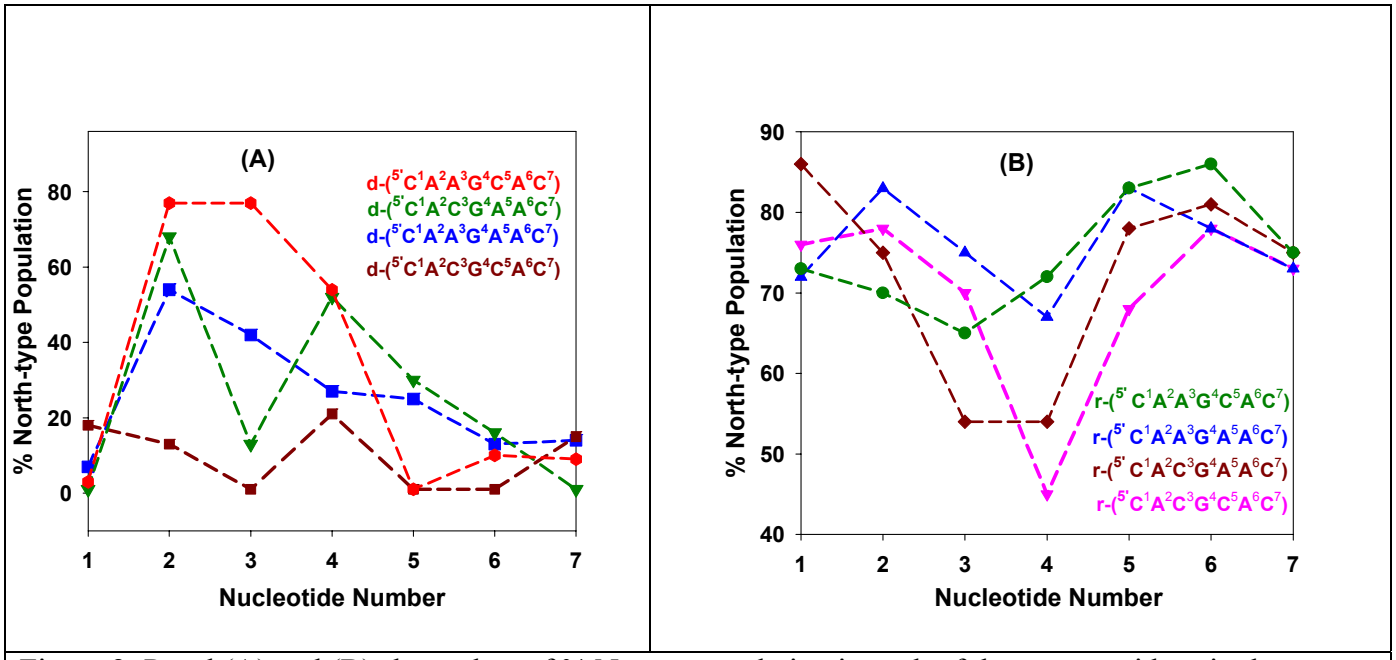
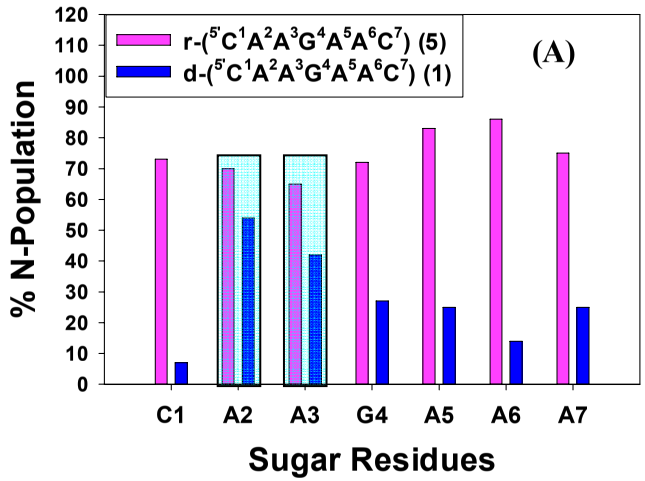
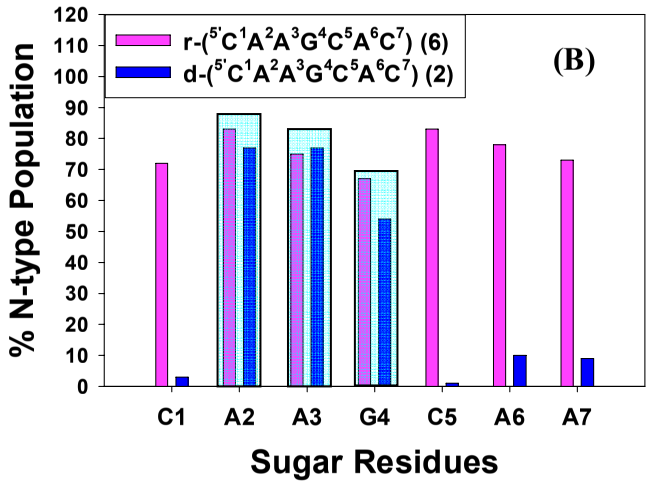
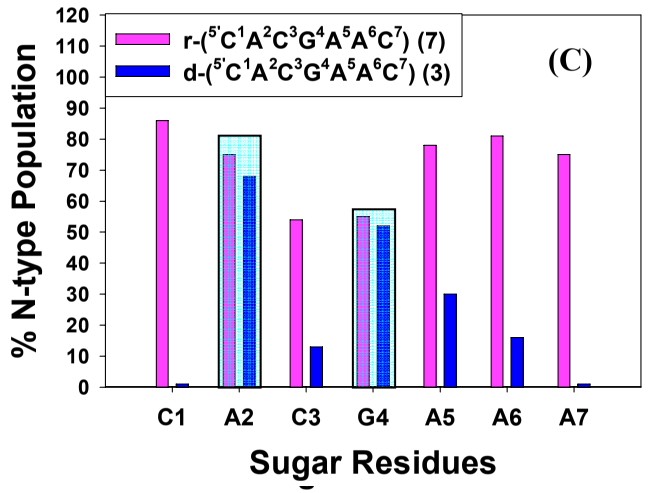


Figure 2: Panel (A) and (B) show plots of % N-type population in each of the sugar residues in the ssRNA/DNA *versus* the Nucleotide Number (1-7, shown as superscript in the sequence) from 5′→3′ end in the isosequential ssDNA and ssRNA. The (%) N-type population in each sugar residue was calculated from $^{3}J_{H1'2'}$ for RNA and $\Sigma^{3}J_{H1'2',\,H1'2''}$ for DNA (*40*).


r-(5'C1A2A3G4A5A6C7) (5)
d-(5'C1A2A3G4A5A6C7) (1)
(A)
% N-Population
120
110
100
90
80
70
60
50
40
30
20
10
0
C1
A2
A3
G4
A5
A6
A7
Sugar Residues


r-(5'C1A2A3G4C5A6C7) (6)
d-(5'C1A2A3G4C5A6C7) (2)
(B)
% N-type Population
120
110
100
90
80
70
60
50
40
30
20
10
0
C1
A2
A3
G4
C5
A6
A7
Sugar Residues



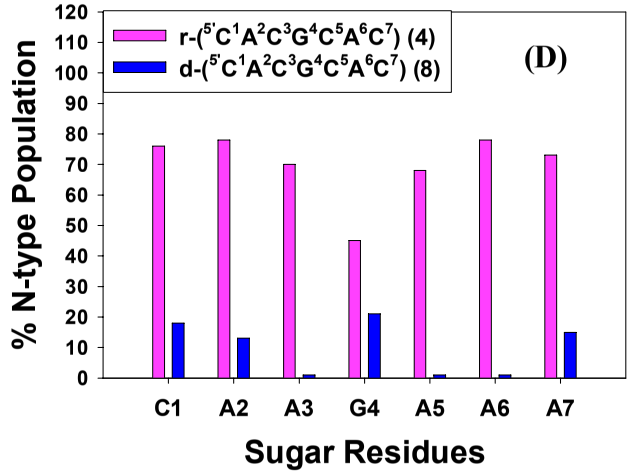
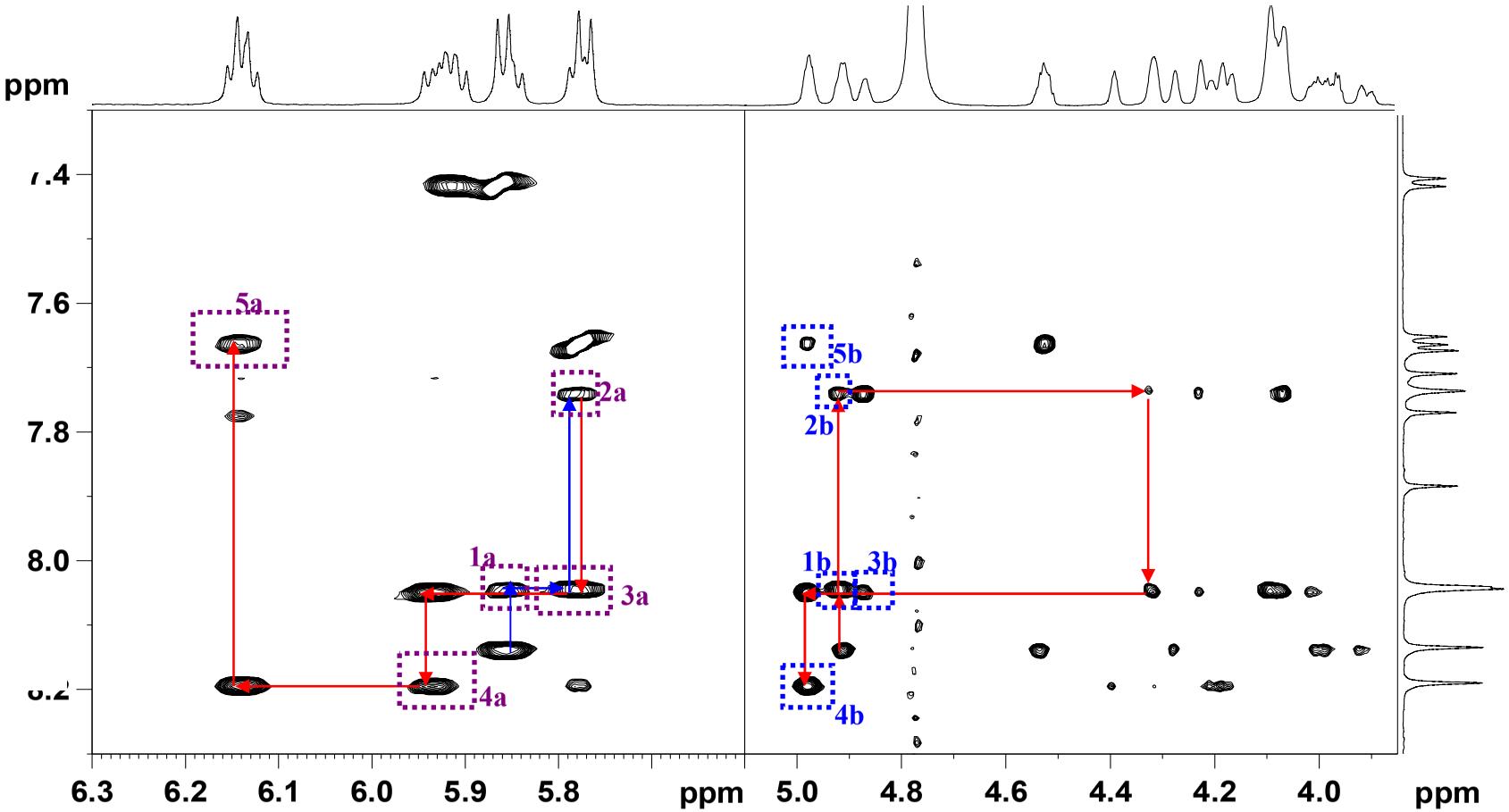


Figure 3: Panels (A) –(D) show sugar conformations for four pairs of isosequential ssDNA (**1** - **4**) and ssRNA (**5** - **8**) with different sequence context. Sugar residues of A^2, A^3 in d(**5'**$C^1A^2A^3G^4A^5A^6C^7$) (**1**), A^2, A^3 , G^4 in d(**5'**$C^1A^2A^3G^4C^5A^6C^7$) (**2**), A^2, G^4 in d(**5'**$C^1A^2C^3G^4A^5A^6C^7$) (**3**) show N-type and mixed N/S populations.


ppm
7.6
7.8
8.0
5a
2a
1a
3a
4a
5b
2b
1b
3b
4b
6.3
6.2
6.1
6.0
5.9
5.8
ppm
5.0
4.8
4.6
4.4
4.2
4.0
ppm

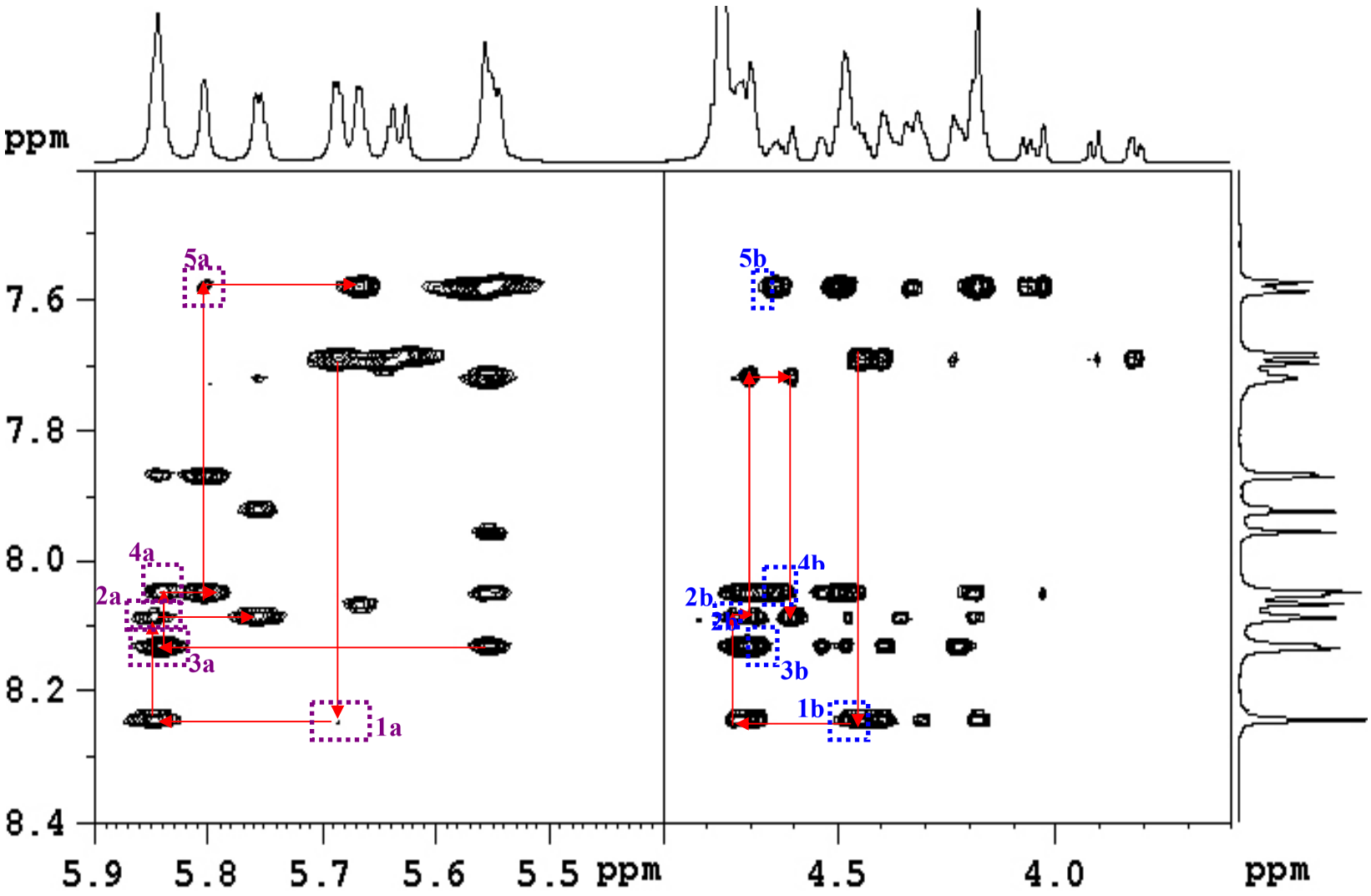


Figure 4: Panel (A) and (B) reperesnts 2D NOESY spectra showing the $(H8/H6)_n \rightarrow (H1')_{n-1}$ crosspeaks (**a**), and $(H8/H6)_n \rightarrow (H3')_{n-1}$ (**b**) crosspeaks for ssDNA, **d($^{5'}C^1A^2A^3G^4A^5A^6C^7$) (1).**and ssRNA, **r($^{5'}C^1A^2A^3G^4A^5A^6C^7$)** (Table 4) . It suggests that ssDNA, **d($^{5'}C^1A^2A^3G^4A^5A^6C^7$) (1)** exists in near B-type form and ssRNA exists in near A-type form. For other ssDNA and ssRNA see Figure S51 and Table S11 in SI.

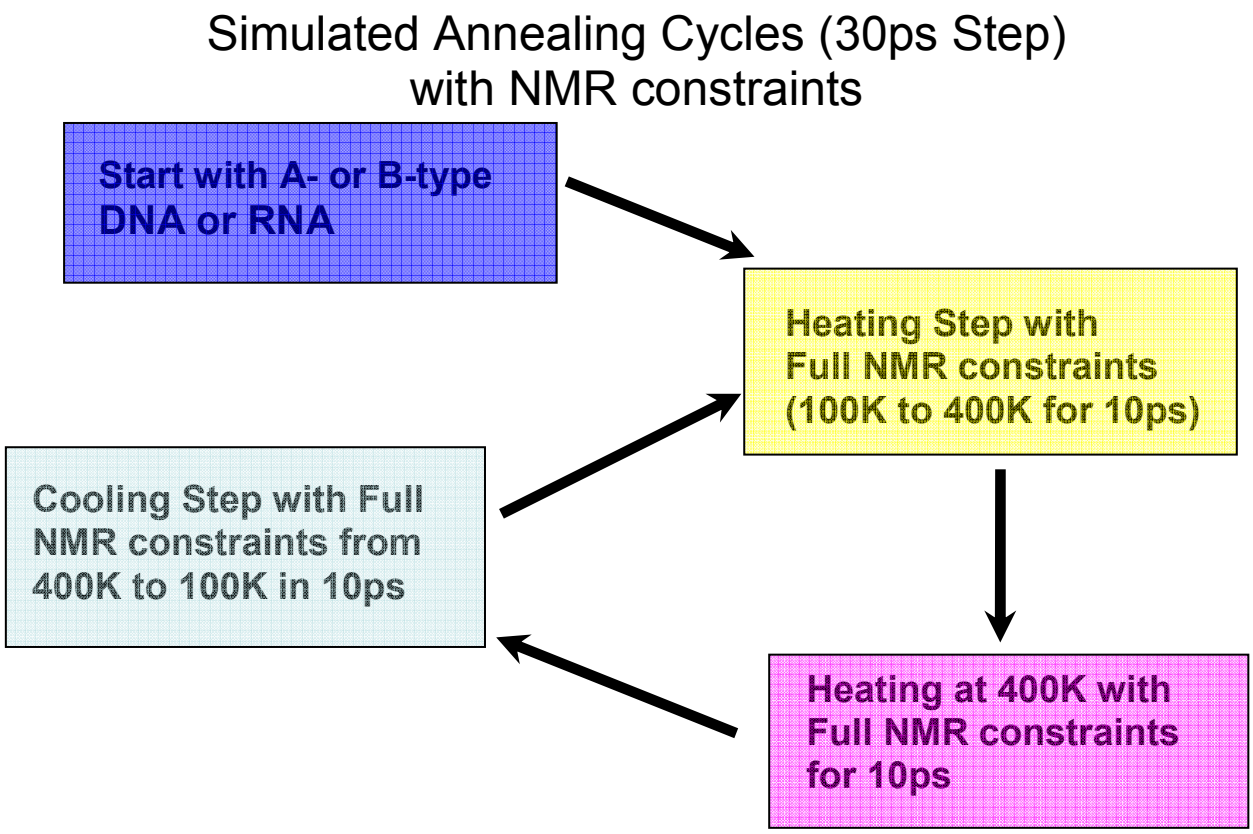
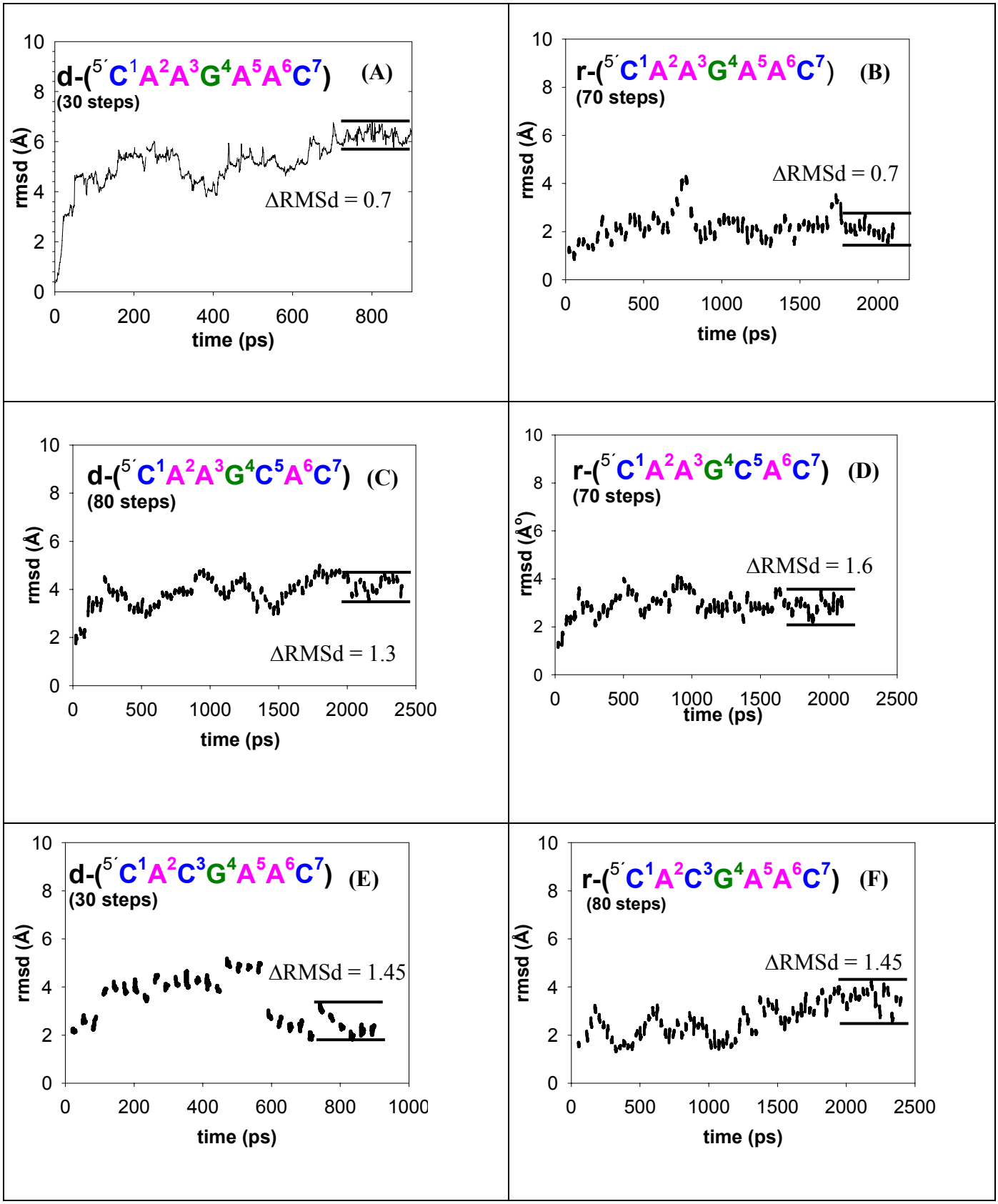
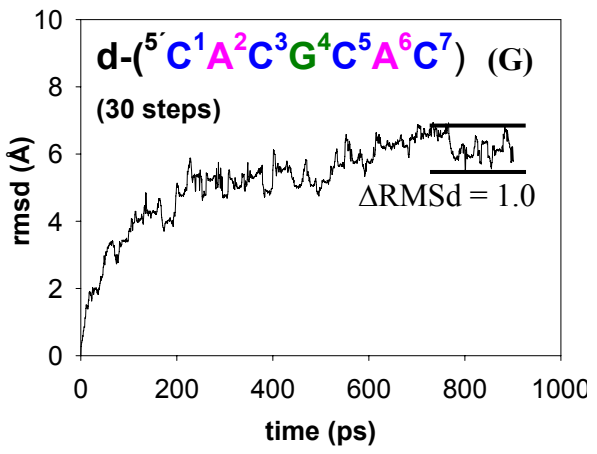


Figure 5: Simulated Annealing Ptotocol





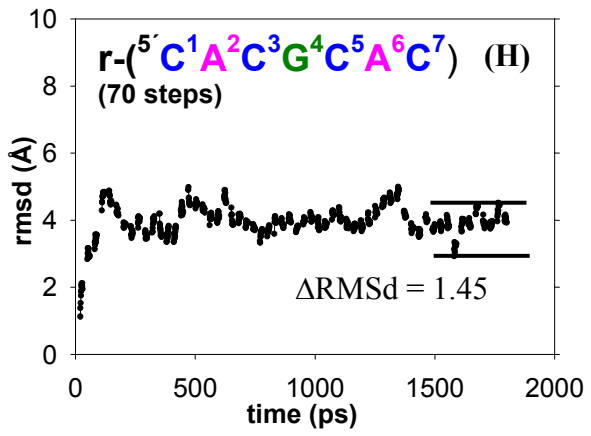
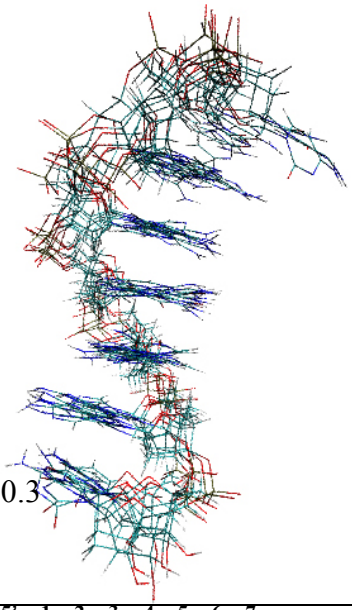
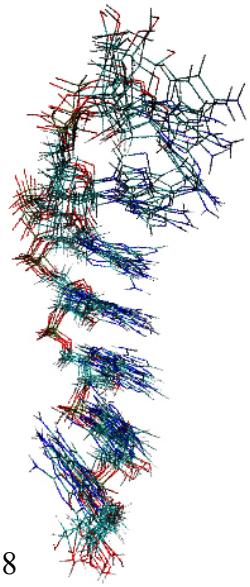
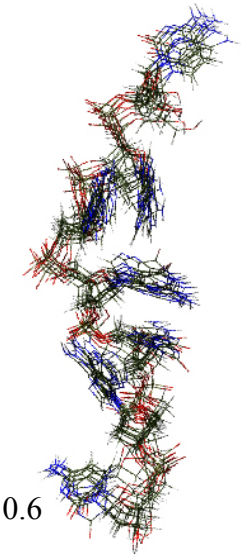
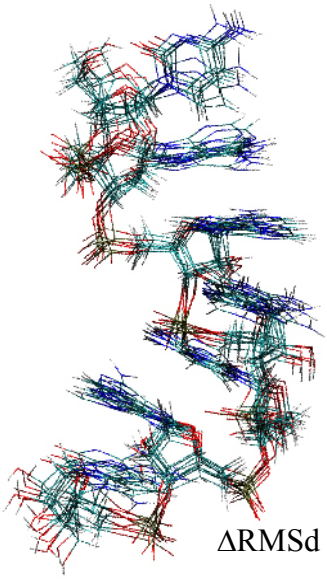
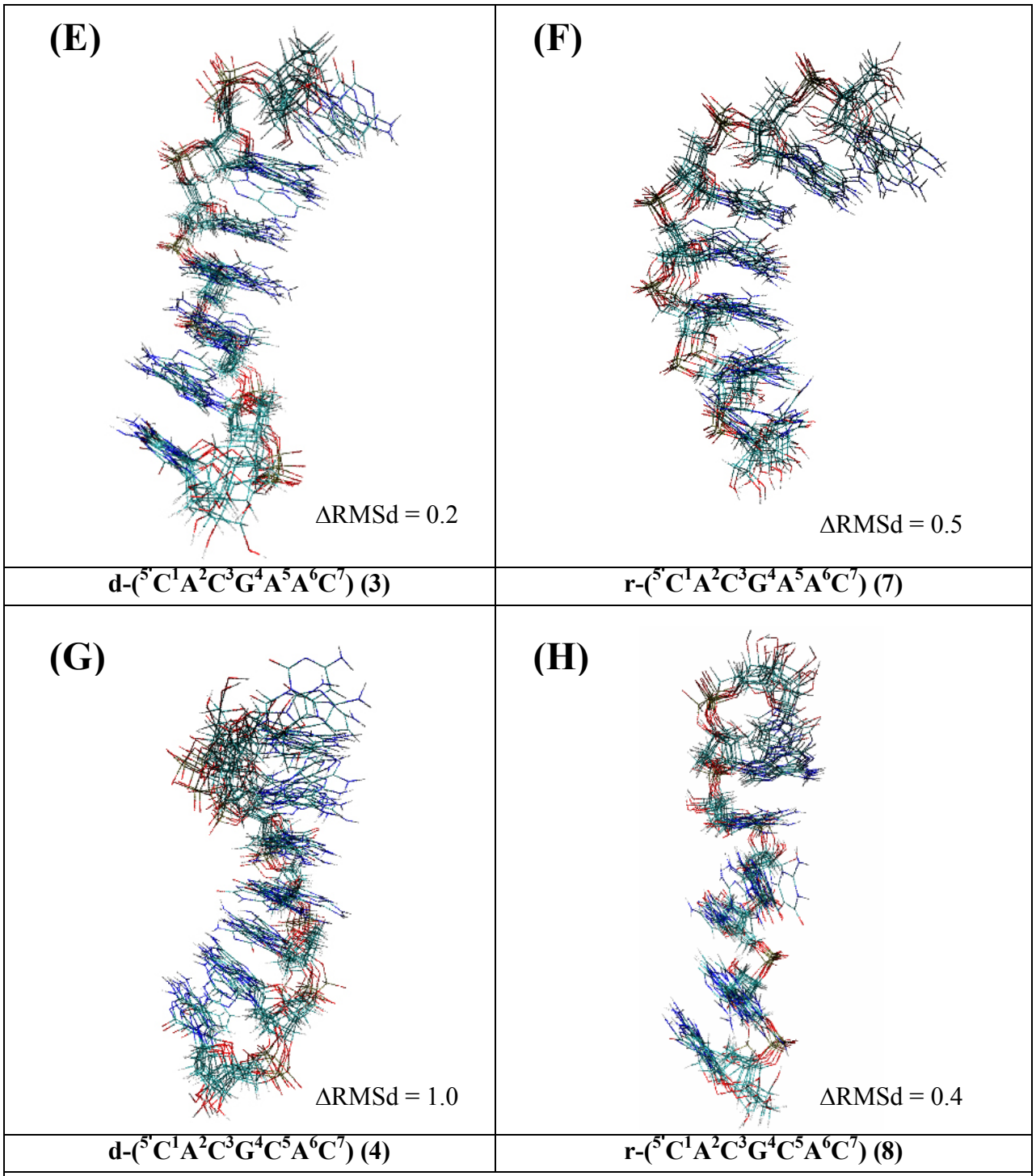


Figure 6: Panels (A)-(H) show RMSD analysis for Simulated Annealing (SA) steps (Figure 1) for the four isosequential ssDNA (1 – 4) /ssRNA (5 – 8) strands, referencing the initial minimized structures of them.

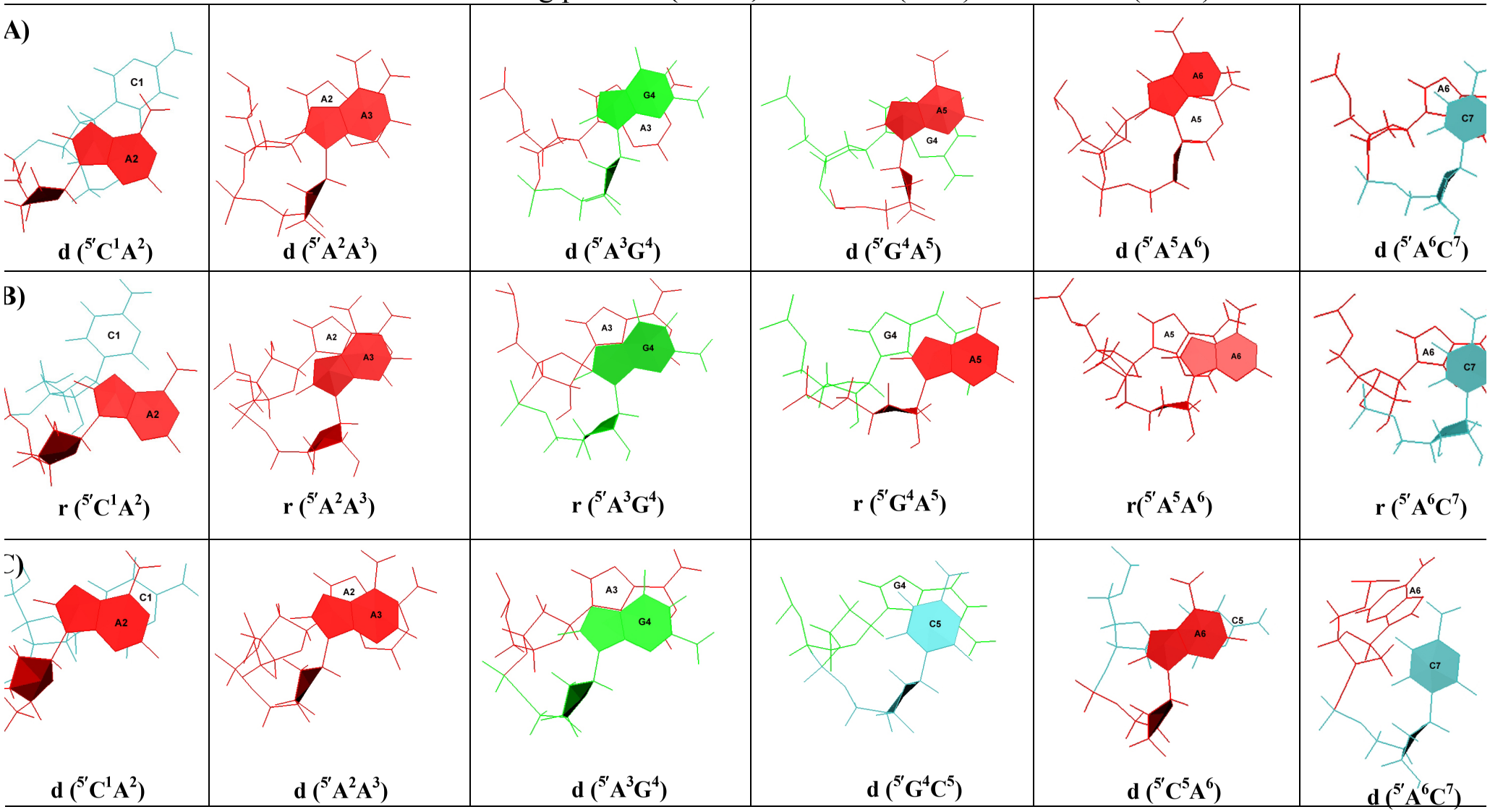
ssDNA	ssRNA
(A)  ΔRMSd = 0.3	**(B)**  ΔRMSd = 0.8
d-($^{5'}C^1A^2A^3G^4A^5A^6C^7$)	**r-($^{5'}C^1A^2A^3G^4A^5A^6C^7$)**
(C)  ΔRMSd = 0.6	**(D)**  ΔRMSd = 0.1
d-($^{5'}C^1A^2A^3G^4C^5A^6C^7$)	**r-($^{5'}C^1A^2A^3G^4C^5A^6C^7$)**



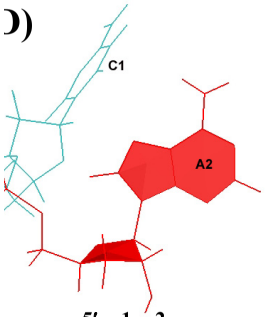
d-($^{5'}C^1A^2C^3G^4A^5A^6C^7$) (3)	**r-($^{5'}C^1A^2C^3G^4A^5A^6C^7$) (7)**
d-($^{5'}C^1A^2C^3G^4C^5A^6C^7$) (4)	**r-($^{5'}C^1A^2C^3G^4C^5A^6C^7$) (8)**

Figure 7: Panel (A)- (H) show overlay of 10 NMR derived SA-MD structures for ssDNAs (**1** – **4**) and ssRNAs (**5** – **8**). ΔRMSd represents the RMSd between 1st and the 10th SA-MD structures (Table 9).

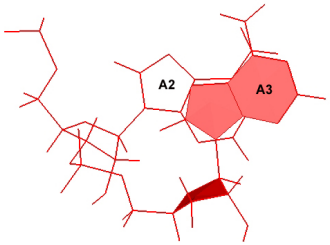
Dinucleotide Stacking patterns **(5′→ 3′)** in ssDNA (**1 - 2**) and ssRNA (**5 – 6**)



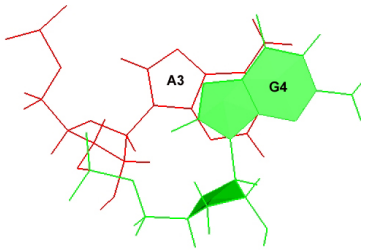
D)



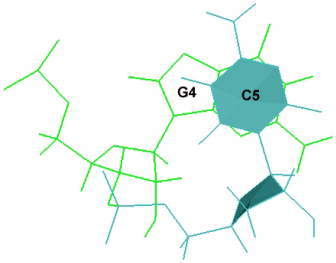
r ($^{5'}C^1A^2$)



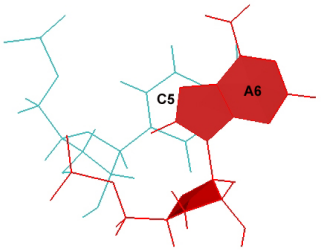
r ($^{5'}A^2A^3$)



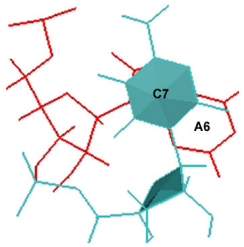
r ($^{5'}A^3G^4$)



r ($^{5'}G^4C^5$)

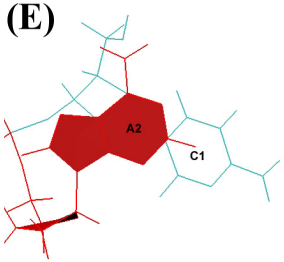


r($^{5'}C^5A^6$)

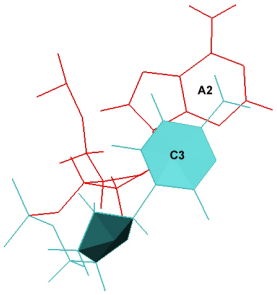


r ($^{5'}A^6C^7$)

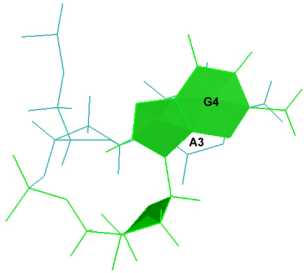
(E)



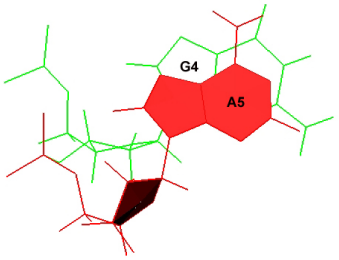
d ($^{5'}C^1A^2$)



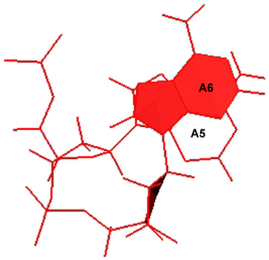
d ($^{5'}A^2C^3$)



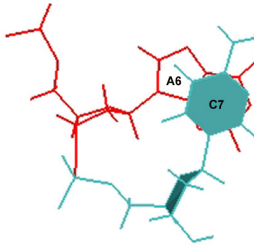
d ($^{5'}C^3G^4$)



d ($^{5'}G^4A^5$)

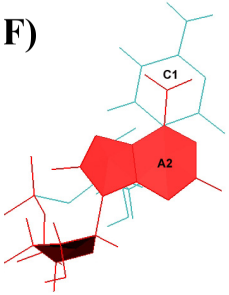


d ($^{5'}A^5A^6$)

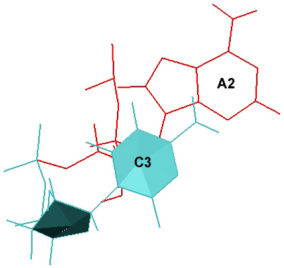


d ($^{5'}A^6C^7$)

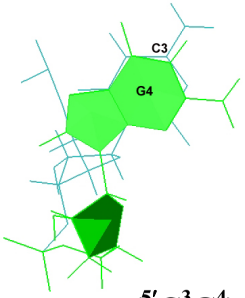
F)



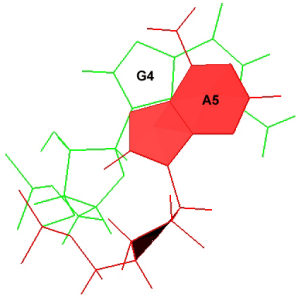
r ($^{5'}C^1A^2$)



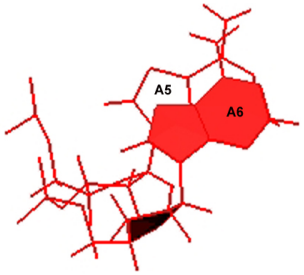
r($^{5'}A^2C^3$)



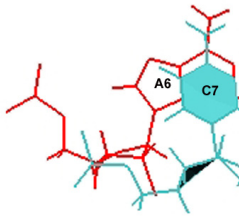
r ($^{5'}C^3G^4$)



r ($^{5'}G^4A^5$)



r ($^{5'}A^5A^6$)



r ($^{5'}A^6C^7$)

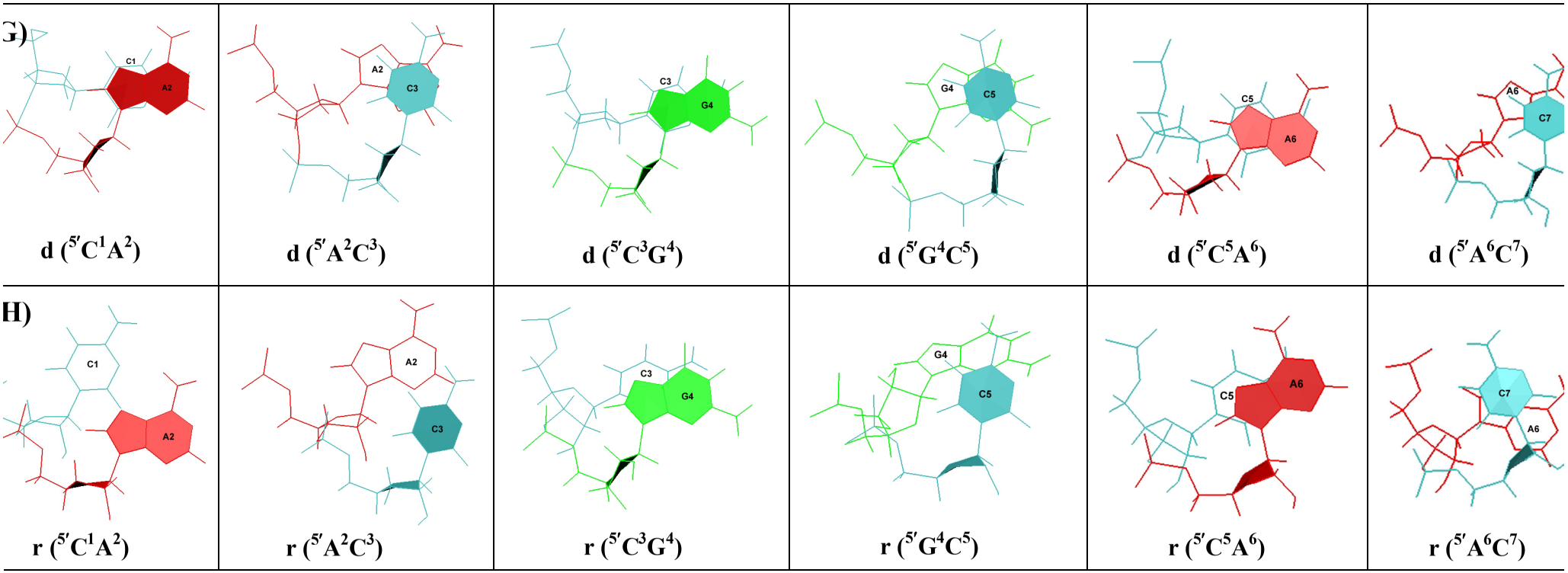
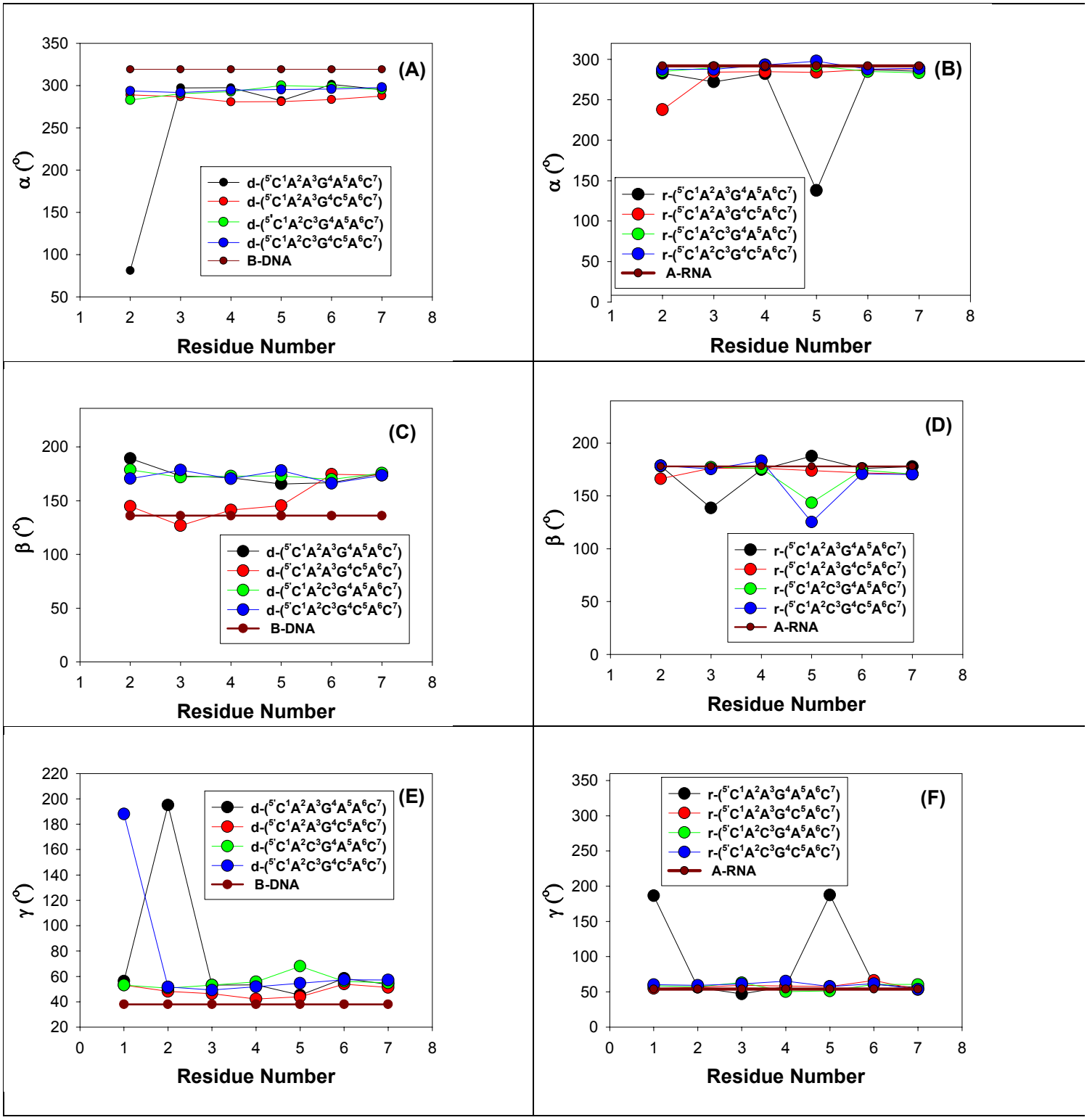
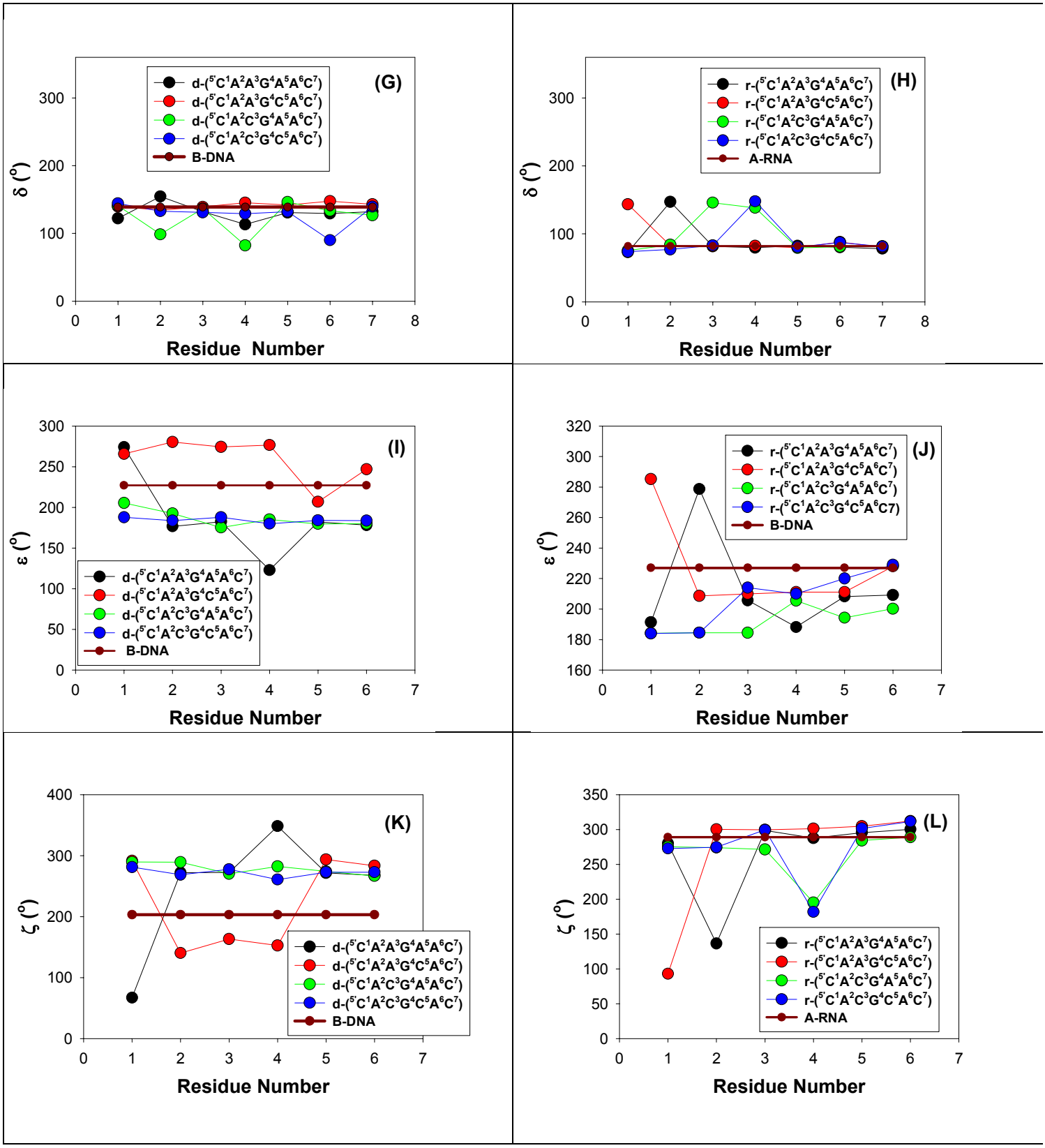
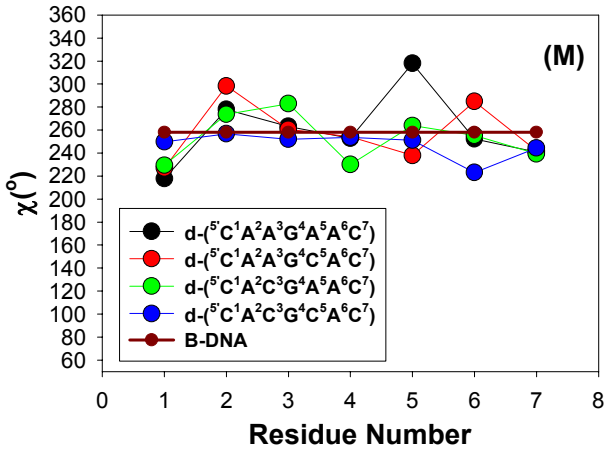


Figure 8. Panels (A)- (H) show dinucleotide base-base stacking patterns from (**5′→ 3′**) ends for ssDNAs (**1** - **4**) and ssRNA (**5** - **8**) to show the nearest neighbor stacking patterns.







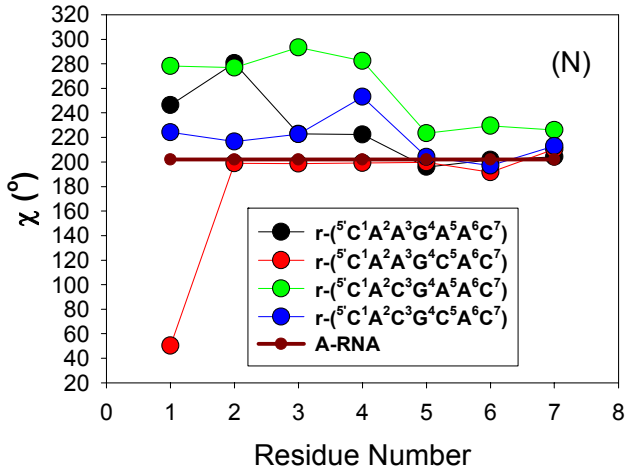
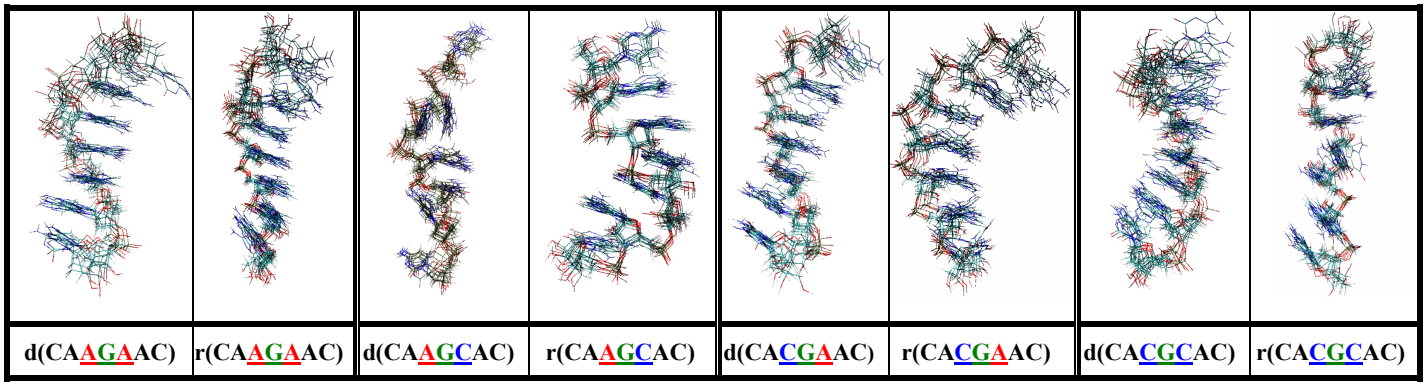


Figure 9: Panels (**A-P**) shows comparison of all dihedral angles (ϕ) for four isosequential pairs of ssDNA (**1** – **4**)and ssRNA (**5** – **8**) with the corresponding dihedral angles in dsDNA and dsRNA(*49*).

TOC Graphic

Sequence-specific Solution Structures of the Four Isosequential Pairs of Single-stranded DNAs and RNAs

Subhrangsu Chatterjee, Wimal Pathmasiri & Jyoti Chattopadhyaya*



Supplementary Information

Sequence-specific Solution Structures of the Four Isosequential Pairs of Single-stranded DNAs and RNAs

Subhrangsu Chatterjee, Wimal Pathmasiri &

*Jyoti Chattopadhyaya**

Department of Bioorganic Chemistry, Box 581 ,Biomedical Center,

Uppsala University, SE-75123 Uppsala, Sweden.

Fax:+46-18554495 E-mail: jyoti@boc.uu.se

Table of Contents

Table S6: Dihedrals Constraints (ϕ) and sugar phase angles (P) for simulated annealing (SA) of d($^{5'}C^1A^2C^3G^4A^5A^6C^7$) (**3**).

Table S7: Dihedrals Constraints (ϕ) and sugar phase angles (P) for simulated annealing (SA) of r($^{5'}C^1A^2C^3G^4A^5A^6C^7$) (**7**).

Table S8: Dihedrals Constraints (ϕ) and sugar phase angles (P) for simulated annealing (SA) of d($^{5'}C^1A^2C^3G^4C^5A^6C^7$) (**4**).

Table S9: Dihedrals Constraints (ϕ) and sugar phase angles (P) for simulated annealing (SA) of r($^{5'}C^1A^2C^3G^4C^5A^6C^7$) (**8**).

Table S10: The oligomerization shift estimated from ^{1}H chemical shift at the neutral (N) state at 298 K for aromatic protons of ssDNA 1 – 4 using appropriate monomeric analogues (Ref. 40) as well as that of ssRNA 5 - 8 using appropriate monomeric analogues (Ref. 40).

Table S11. Inter-residual (n-1) nOe contacts for ssDNA **1** – **4**.

Table S12. Intra-residual (n-1) nOe contacts for ssDNA **1** – **4**.

Table S13. Inter-residual (n) nOe contacts for ssRNA **5** – **8**.

Table S14. Intra-residual (n) nOe contacts for ssRNA **5** – **8**.

Table S15 – S16. Distances (Å) between aromatic (n) and sugar protons n, (n-1) of Cannonical A-RNA and B-DNA duplexes.

Table S17 – S24. Distances (Å) between aromatic (n) and sugar protons n, (n-1) of four isosequential ssDNA (**1** – **4**) and ssRNA (**5** – **8**) Heptamers.

Figure S1. NOESY footprint of d($^{5'}C^1A^2A^3G^4A^5A^6C^7$) (**1**) showing the connectivity of nucleotide residues.

Figure S2. The expanded ^{31}P coupled DQF-COSY spectra of H1'/H2'/H2"/H3' region for

d($^{5'}C^1A^2A^3G^4A^5A^6C^7$) (**1**) at 298 K. The region 6.3 – 1.4 ppm in F1 and region 6.3 – 4.5 ppm in F2 dimension of the d($^{5'}C^1A^2A^3G^4A^5A^6C^7$) (**1**) showing the spin connectivity.

Figure S3. The expanded ^{31}P coupled DQF-COSY spectra of H3'/H4'/H5'/H5" region for d($^{5'}C^1A^2A^3G^4A^5A^6C^7$) (**1**) at 298 K. The region 5.1 – 3.5 ppm in both F1 and F2 dimensions of the d-$^{5'}$($C^1A^2A^3G^4A^5A^6C^7$) (**1**) showing the spin connectivity.

Figure S3.1. The ^{31}P decoupled DQF-COSY spectrum of d($^{5'}C^1A^2A^3G^4A^5A^6C^7$) (**1**) at 298 K. For assignments see **S2** and **S3**.

Figure S4. Expanded TOCSY spectra of the H2'/H2"/H3'/H4'/H5'/H5" region (1.5 – 5.3 ppm in F1 direction) to anomeric (H1') region (5.1 – 6.25 ppm in F2 direction) for d($^{5'}C^1A^2A^3G^4A^5A^6C^7$) (**1**) at 298 K.

Figure S5. Expanded ^{31}P - ^{1}H correlation spectroscopy of ^{31}P region (-1.9 – -2.85 ppm in F2 direction) to H3'/H4'/H5'/H5" region (5.2 – 3.8 ppm in F1 direction) for d($^{5'}C^1p_1A^2p_2A^3p_3G^4p_4A^5p_5A^6p_6C^7$) (**1**) at 298 K.

Figure S6. NOESY footprint of r($^{5'}C^1A^2A^3G^4A^5A^6C^7$) (**5**) showing the connectivity of nucleotide residues.

Figure S7. The expanded ^{31}P coupled DQF-COSY spectra of the anomeric H1' region to the H2' for r($^{5'}C^1A^2A^3G^4A^5A^6C^7$) (**5**) at 298 K. The spin connectivity between anomeric H1' region (5.45 – 5.95 ppm in F2) and H2' region (4.0 – 4.9 ppm in F1) of r($^{5'}C^1A^2A^3G^4A^5A^6C^7$)(**5**) have been shown.

Figure S8. The expanded ^{31}P coupled DQF-COSY spectra of H2'/H3'/H4'/H5'/H5" region for r($^{5'}C^1A^2A^3G^4A^5A^6C^7$) (**5**) at 298 K. The region 4.8 – 3.7 ppm in both F1 and F2 dimensions of the r($^{5'}C^1A^2A^3G^4A^5A^6C^7$) (**5**) showing the spin connectivity.

Figure S8.1. The ^{31}P decoupled DQF-COSY spectrum of r($^{5'}C^1A^2A^3G^4A^5A^6C^7$) (**5**) at 298 K. For assignments see **S7** and **S8**.

Figure S9. Expanded TOCSY spectra of the H2'/H3'/H4'/H5'/H5" region (4.9 – 3.7 ppm in F1 direction) to anomeric (H1') region (5.5 – 5.9 ppm in F2 direction) for r($^{5'}C^1A^2A^3G^4A^5A^6C^7$) (**5**) at 298 K.

Figure S10. Expanded ^{31}P - ^{1}H correlation spectroscopy of ^{31}P region (-1.6 – -2.5 ppm in F2 direction) to H2'/H3'/H4'/H5'/H5" region (4.8 – 3.7 ppm in F1 direction) for r($^{5'}C^1p_1A^2p_2A^3p_3G^4p_4A^5p_5A^6p_6C^7$) (**5**) at 298 K.

Figure S11. NOESY footprint of d($^{5'}C^1A^2A^3G^4C^5A^6C^7$) (**2**) showing the connectivity of nucleotide residues.

Figure S12. The expanded ^{31}P coupled DQF-COSY spectra of H1'/H2'/H2"/H3' region for d($^{5'}C^1A^2A^3G^4C^5A^6C^7$) (**2**) at 298 K. The region 2.9 – 1.5 ppm in F1 and region 6.45 – 4.4 ppm in F2 dimension of the d($^{5'}C^1A^2A^3G^4C^5A^6C^7$) (**2**) showing the spin connectivity.

Figure S13. The expanded ^{31}P coupled DQF-COSY spectra of H3'/H4'/H5'/H5" region for d-$^{5'}$($C^1A^2A^3G^4C^5A^6C^7$) (**2**) at 298 K. The region 5.1 – 3.4 ppm in both F1 and F2 dimensions of the d($^{5'}C^1A^2A^3G^4A^5A^6C^7$) (**2**) showing the spin connectivity.

Figure S13.1. The ^{31}P decoupled DQF-COSY spectrum of d($^{5'}C^1A^2A^3G^4C^5A^6C^7$) (**2**) at 298 K. For assignments see **S12** and **S13**.

Figure S14. Expanded TOCSY spectra of the H2'/H2"/H3'/H4'/H5'/H5" region (1.0 – 5.1 ppm in F1 direction) to anomeric (H1') region (6.35 – 5.6 ppm in F2 direction) for d($^{5'}C^1A^2A^3G^4C^5A^6C^7$) (**2**) at 298 K.

Figure S15. Expanded ^{31}P - ^{1}H correlation spectroscopy of ^{31}P region (-1.8 – -3.0 ppm in F2 direction) to H3'/H4'/H5'/H5" region (5.1 – 3.85 ppm in F1 direction) for d($^{5'}C^1p_1A^2p_2A^3p_3G^4p_4C^5p_5A^6p_6C^7$) (**2**) at 298 K.

Figure S16. NOESY footprint of r($^{5'}C^1A^2A^3G^4C^5A^6C^7$) (**6**) showing the connectivity of nucleotide residues.

Figure S17. The expanded ^{31}P coupled DQF-COSY spectra of the anomeric H1' region to the H2' for r($^{5'}$C^{1}A^{2}A^{3}G^{4}C^{5}A^{6}C^{7}) (**6**) at 298 K. The spin connectivity between anomeric H1' region (5.4 – 6.1 ppm in F2) and H2' region (4.0 – 4.8 ppm in F1) of r($^{5'}$C^{1}A^{2}A^{3}G^{4}C^{5}A^{6}C^{7}) (**6**) have been shown.

Figure S18. The expanded ^{31}P coupled DQF-COSY spectra of H2'/H3'/H4'/H5'/H5" region for r($^{5'}$C^{1}A^{2}A^{3}G^{4}C^{5}A^{6}C^{7}) (**6**) at 298 K. The region 4.85 – 3.75 ppm in both F1 and F2 dimensions of the r($^{5'}$C^{1}A^{2}A^{3}G^{4}C^{5}A^{6}C^{7}) (**6**) showing the spin connectivity.

Figure S18.1. The ^{31}P decoupled DQF-COSY spectrum of r($^{5'}$C^{1}A^{2}A^{3}G^{4}C^{5}A^{6}C^{7}) (**6**) at 298 K. For assignments see **S17** and **S18**.

Figure S19. Expanded TOCSY spectra of the H2'/H3'/H4'/H5'/H5" region (4.85 – 3.6 ppm in F1 direction) to anomeric (H1') region (5.45 – 6.1 ppm in F2 direction) for r($^{5'}$C^{1}A^{2}A^{3}G^{4}C^{5}A^{6}C^{7}) (**6**) at 298 K.

Figure S20. Expanded ^{31}P - ^{1}H correlation spectroscopy of ^{31}P region (-1.55 – -2.6 ppmin F2 direction) to H2'/H3'/H4'/H5'/H5" region (4.7 – 4.0 ppm in F1 direction) for dr($^{5'}$C^{1}p$_{1}$A^{2}p$_{2}$A^{3}p$_{3}$G^{4}p$_{4}$C^{5}p$_{5}$A^{6}p$_{6}$C^{7}) (**6**) at 298 K.

Figure S21. NOESY footprint of d($^{5'}$C^{1}A^{2}C^{3}G^{4}A^{5}A^{6}C^{7}) (**3**) showing the connectivity of nucleotide residues.

Figure S22. The expanded ^{31}P coupled DQF-COSY spectra of H1'/H2'/H2"/H3' region for d($^{5'}$C^{1}A^{2}C^{3}G^{4}A^{5}A^{6}C^{7}) (**3**) at 298 K. The region 2.95 – 1.4 ppm in F1 and region 6.5 – 4.4 ppm in F2 dimension of the d($^{5'}$C^{1}A^{2}C^{3}G^{4}A^{5}A^{6}C^{7}) (**3**) showing the spin connectivity.

Figure S23. The expanded ^{31}P coupled DQF-COSY spectra of H3'/H4'/H5'/H5" region for d($^{5'}$C^{1}A^{2}C^{3}G^{4}A^{5}A^{6}C^{7}) (**3**) at 298 K. The region 5.1 – 3.5 ppm in both F1 and F2 dimensions of the d$^{'}$(^{5}C^{1}A^{2}C^{3}G^{4}A^{5}A^{6}C^{7}) (**3**) showing the spin connectivity.

Figure S23.1. The ^{31}P decoupled DQF-COSY spectrum of d($^{5'}$C^{1}A^{2}C^{3}G^{4}A^{5}A^{6}C^{7}) (**3**) at 298 K. For assignments see **S22** and **S23**.

Figure S24. Expanded TOCSY spectra of the H2'/H2"/H3'/H4'/H5'/H5" region (1.5 – 5.15 ppm in F1 direction) to anomeric (H1') region (6.35 – 5.4 ppm in F2 direction) for d($^{5'}C^1A^2C^3G^4A^5A^6C^7$) (**3**) at 298 K.

Figure S25. Expanded ^{31}P - ^{1}H correlation spectroscopy of ^{31}P region (-1.9 – -2.7 ppm in F2 direction) to H3'/H4'/H5'/H5" region (5.1 – 3.9 ppm in F1 direction) for d($^{5'}C^1p_1A^2p_2C^3p_3G^4p_4A^5p_5A^6p_6C^7$) (**3**) at 298 K.

Figure S26. NOESY footprint of r($^{5'}C^1A^2C^3G^4A^5A^6C^7$) (**7**) showing the connectivity of nucleotide residues.

Figure S27. The expanded ^{31}P coupled DQF-COSY spectra of the anomeric H1' region to the H2' for r($^{5'}C^1A^2C^3G^4A^5A^6C^7$) (**7**) at 298 K. The spin connectivity between anomeric H1' region (5.5 – 6.05 ppm in F2) and H2' region (3.9 – 4.9 ppm in F1) of r($^{5'}C^1A^2C^3G^4A^5A^6C^7$) (**7**) have been shown.

Figure S28. The expanded ^{31}P coupled DQF-COSY spectra of H2'/H3'/H4'/H5'/H5" region for r($^{5'}C^1A^2C^3G^4A^5A^6C^7$) (**7**) at 298 K. The region 4.85 – 3.7 ppm in both F1 and F2 dimensions of the r($^{5'}C^1A^2C^3G^4A^5A^6C^7$) (**7**) showing the spin connectivity.

Figure S28.1. The ^{31}P decoupled DQF-COSY spectrum of r($^{5'}C^1A^2C^3G^4A^5A^6C^7$) (**7**) at 298 K. For assignments see **S27** and **S28**.

Figure S29. Expanded TOCSY spectra of the H2'/H3'/H4'/H5'/H5" region (4.8 – 3.7 ppm in F1 direction) to anomeric (H1') region (5.55 – 6.0 ppm in F2 direction) for r($^{5'}C^1A^2C^3G^4A^5A^6C^7$) (**7**) at 298 K.

Figure S30. Expanded ^{31}P - ^{1}H correlation spectroscopy of ^{31}P region (-1.6 – -2.5 ppm in F2 direction) to H2'/H3'/H4'/H5'/H5" region (4.8 – 3.8 ppm in F1 direction) for r($^{5'}C^1p_1A^2p_2C^3p_3G^4p_4A^5p_5A^6p_6C^7$) (**7**) at 298 K.

Figure S31. NOESY footprint of d($^{5'}C^1A^2C^3G^4C^5A^6C^7$) (**4**) showing the connectivity of nucleotide residues.

Figure S32. The expanded ^{31}P coupled DQF-COSY spectra of H1'/H2'/H2"/H3' region for d($^{5'}C^1A^2C^3G^4C^5A^6C^7$) (**4**) at 298 K. The region 3.1 – 1.6 ppm in F1 and region 6.5 – 4.4 ppm in F2 dimension of the d($^{5'}C^1A^2C^3G^4C^5A^6C^7$) (**4**) showing the spin connectivity.

Figure S33. The expanded ^{31}P coupled DQF-COSY spectra of H3'/H4'/H5'/H5" region for d($^{5'}C^1A^2C^3G^4C^5A^6C^7$) (**4**) at 298 K. The region 5.2 – 3.5 ppm in both F1 and F2 dimensions of the d($^{5'}C^1A^2C^3G^4C^5A^6C^7$) (**4**) showing the spin connectivity.

Figure S33.1. The ^{31}P decoupled DQF-COSY spectrum of d($^{5'}C^1A^2C^3G^4C^5A^6C^7$) (**4**) at 298 K. For assignments see **S32** and **S33**.

Figure S34. Expanded TOCSY spectra of the H2'/H2"/H3'/H4'/H5'/H5" region (1.5 – 5.2 ppm in F1 direction) to anomeric (H1') region (6.4 – 5.9 ppm in F2 direction) for d($^{5'}C^1A^2C^3G^4C^5A^6C^7$) (**4**) at 298 K.

Figure S35. Expanded ^{31}P - ^{1}H correlation spectroscopy of ^{31}P region (-1.8 – -2.5 ppm in F2 direction) to H3'/H4'/H5'/H5" region (5.1 – 3.9 ppm in F1 direction) for d($^{5'}C^1p_1A^2p_2C^3p_3G^4p_4C^5p_5A^6p_6C^7$) (**4**) at 298 K.

Figure S36. NOESY footprint of r($^{5'}C^1A^2C^3G^4C^5A^6C^7$) (**8**) showing the connectivity of nucleotide residues.

Figure S37. The expanded ^{31}P coupled DQF-COSY spectra of the anomeric H1' region to the H2' for r($^{5'}C^1A^2C^3G^4C^5A^6C^7$) (**8**) at 298 K. The spin connectivity between anomeric H1' region (5.5 – 6.1 ppm in F2) and H2' region (4.0 – 4.9 ppm in F1) of r($^{5'}C^1A^2C^3G^4C^5A^6C^7$) (**8**) have been shown.

Figure S38. The expanded ^{31}P coupled DQF-COSY spectra of H2'/H3'/H4'/H5'/H5" region for r($^{5'}$C^{1}A^{2}C^{3}G^{4}C^{5}A^{6}C^{7}) (**8**) at 298 K. The region 4.9 – 3.7 ppm in both F1 and F2 dimensions of the r-$^{5'}$(C^{1}A^{2}C^{3}G^{4}C^{5}A^{6}C^{7}) (**8**) showing the spin connectivity.

Figure S38.1. The ^{31}P decoupled DQF-COSY spectrum of r($^{5'}$C^{1}A^{2}C^{3}G^{4}C^{5}A^{6}C^{7}) (**8**) at 298 K. For assignments see **S37** and **S38**.

Figure S39. Expanded TOCSY spectra of the H2'/H3'/H4'/H5'/H5" region (5.0 – 3.75 ppm in F1 direction) to anomeric (H1') region (5.4 – 6.1 ppm in F2 direction) for r($^{5'}$C^{1}A^{2}C^{3}G^{4}C^{5}A^{6}C^{7}) (**8**) at 298 K.

Figure S40. Expanded ^{31}P - ^{1}H correlation spectroscopy of ^{31}P region (-1.7 – -2.5 ppm in F2 direction) to H2'/H3'/H4'/H5'/H5" region (4.8 – 4.0 ppm in F1 direction) for r($^{5'}$C^{1}p$_{1}$A^{2}p$_{2}$C^{3}p$_{3}$G^{4}p$_{4}$C^{5}p$_{5}$A^{6}p$_{6}$C^{7}) (**8**) at 298 K.

Figure S41. Aromatic – anomeric and aromatic – sugar proton NOESY crosspeaks for d($^{5'}$C^{1}A^{2}A^{3}G^{4}A^{5}A^{6}C^{7}) (**1**).

Figure S42. Aromatic – anomeric and aromatic – sugar proton NOESY crosspeaks for r($^{5'}$C^{1}A^{2}A^{3}G^{4}A^{5}A^{6}C^{7}) (**5**).

Figure S43. Aromatic – anomeric and aromatic – sugar proton NOESY crosspeaks for d($^{5'}$C^{1}A^{2}A^{3}G^{4}C^{5}A^{6}C^{7}) (**2**).

Figure S44. Aromatic – anomeric and aromatic – sugar proton NOESY crosspeaks for r($^{5'}$C^{1}A^{2}A^{3}G^{4}C^{5}A^{6}C^{7}) (**6**).

Figure S45. Aromatic – anomeric and aromatic – sugar proton NOESY crosspeaks for d($^{5'}$C^{1}A^{2}C^{3}G^{4}A^{5}A^{6}C^{7}) (**3**).

Figure S46. Aromatic – anomeric and aromatic – sugar proton NOESY crosspeaks for r($^{5'}$C^{1}A^{2}C^{3}G^{4}A^{5}A^{6}C^{7}) (**7**).

Figure S47. Aromatic – anomeric and aromatic – sugar proton NOESY crosspeaks for d($^{5'}C^1A^2C^3G^4C^5A^6C^7$) (**4**).

Figure S48. Aromatic – anomeric and aromatic – sugar proton NOESY crosspeaks for r($^{5'}C^1A^2C^3G^4C^5A^6C^7$) (**8**).

Figure S49. Panels (A) - (D), show the nOe connectivites and cross peak intensities for H8/6$_{(n)}$ – H1'$_{(n-1)}$ [marked as (a)] and H8/6$_{(n)}$ – H3'$_{(n-1)}$ [marked as (b)] in ssDNAs **(1 – 4)**. Panels (E) - (H), show the nOe connectivites and cross peak intensities for H8/6$_{(n)}$ – H1'$_{(n-1)}$ [marked as (a)] and H8/6$_{(n)}$ – H3'$_{(n-1)}$ [marked as (b)] in ssRNAs (**5** – **8**) (see Table S11 for the detailed comparison of relative intensities of H8/H6$_{(n)}$ – H1'$_{(n-1)}$ and H8/H6$_{(n)}$ – H3'$_{(n-1)}$ crosspeaks).

Figure S50. Plots of mass weighted RMSD beetween the ssDNA and ssRNA molecular modelling trajectories and their correponding most stable SA stuctures at 100K.

Figure S51. Plots of total potential energy during NMR-MD simulation (all the constraints switched on) steps against the time (ps) for ssDNA and ssRNA.

Figure S52. Plots of Sugar puckering (Phase angle, *P*) and helical parameters (Roll, Slide, Inclination, and Nearest neighbour base atom overlap) for the four isosequential ssDNA and ssRNA averaged MD structures..

Figure 53. Plots (A- H) of mass weighted RMSD for (1-3) [red]and (5-7) [blue] nucleobase residues in ssDNAs (**1** - **4**) and ssRNAs (**5** – **8**) with Time (ps) shows that 5′-ends are more dynamic compared to 3′-ends in ssDNAs and ssRNAs. All RMSD for nucleobase residues were calculated referencing the final SA structures of ssDNAs (**1** - **4**) and ssRNAs (**5** – **8**) at 100 K.

Table S1: ^{1}H chemical shifts [δ_H, in ppm] at the neutral state at 298 K for ssDNA heptamers **1 – 4** as well as ssRNA heptamers **5 – 8**.

Compounds		^{1}H Chemical Shift [δ_H] at neutral state at 298 K			
		δ_{H8}	δ_{H2}	δ_{H6}	δ_{H5}
d(C^{1}A^{2}A^{3}G^{4}A^{5}A^{6}C^7) (1)	**dC**	-	-	7.414	5.860
	dA	8.134	7.886	-	-
	dA	8.044	7.676	-	-
	dG	7.735	-	-	-
	dA	8.040	7.711	-	-
	dA	8.190	7.771	-	-
	dC	-	-	7.660	5.774
r(C^{1}A^{2}A^{3}G^{4}A^{5}A^{6}C^7) (5)	**rC**	-	-	7.687	5.639
	rA	8.244	7.924	-	-
	rA	8.093	7.954	-	-
	rG	7.731	-	-	-
	rA	8.144	7.869	-	-
	rA	8.061	8.063	-	-
	rC	-	-	7.583	5.562
d(C^{1}A^{2}A^{3}G^{4}C^{5}A^{6}C^7) (2)	**dC**	-	-	7.421	5.867
	dA	8.145	7.914	-	-
	dA	8.090	7.746	-	-
	dG	7.830	-	-	-
	dC	-	-	7.397	5.644
	dA	8.318	8.048	-	-
	dC	-	-	7.720	5.831
r(C^{1}A^{2}A^{3}G^{4}C^{5}A^{6}C^7) (6)	**rC**	-	-	7.717	5.633
	rA	8.258	7.916	-	-
	rA	8.078	8.016	-	-
	rG	7.673	-	-	-
	rC	-	-	7.618	5.514
	rA	8.278	8.078	-	-
	rC	-	-	7.646	5.639

Compounds		^{1}H Chemical Shift [δ_H] at neutral state at 298 K			
		δ_{H8}	δ_{H2}	δ_{H6}	δ_{H5}
d(C^{1}A^{2}C^{3}G^{4}A^{5}A^{6}C^7) (3)	**dC**	-	-	7.476	5.874
	dA	8.337	8.105	-	-
	dC	-	-	7.425	5.715
	dG	7.766	-	-	-
	dA	8.086	7.790	-	-
	dA	8.219	7.806	-	-
	dC	-	-	7.674	5.791
r(C^{1}A^{2}C^{3}G^{4}A^{5}A^{6}C^7) (7)	**rC**	-	-	7.757	5.690
	rA	8.315	8.091	-	-
	rC	-	-	7.566	5.598
	rG	7.839	-	-	-
	rA	8.198	7.915	-	-
	rA	8.081	8.091	-	-
	rC	-	-	7.598	5.574
d(C^{1}A^{2}C^{3}G^{4}C^{5}A^{6}C^7) (4)	**dC**	-	-	7.470	5.876
	dA	8.341	8.055	-	-
	dC	-	-	7.516	5.774
	dG	7.917	-	-	-
	dC	-	-	7.449	5.715
	dA	8.336	8.085	-	-
	dC	-	-	7.729	5.845
r(C^{1}A^{2}C^{3}G^{4}C^{5}A^{6}C^7) (8)	**rC**	-	-	7.777	5.689
	rA	8.321	8.099	-	-
	rC	-	-	7.576	5.580
	rG	7.852	-	-	-
	rC	-	-	7.684	5.602
	rA	8.303	8.093	-	-
	rC	-	-	7.655	5.647

Table S2: Dihedrals Constraints (ϕ) and sugar phase angles (P) for simulated annealing (SA) of d-$C^1A^2A^3G^4A^5A^6C^7$ (**1**).

ϕ (deg)	C^1	A^2	A^3	G^4	A^5	A^6	C^7
α	-	-	-	-	-	-	-
β	-	140-220	140-220	140-220	140-220	140-220	140-220
γ	20-100	20-100	20-100	20-100	20-100	20-100	20-100
ε	140-220	140-220	140-220	140-220	140-220	140-220	-
ζ	-	-	-	-	-	-	-
P	150-210	0-210	0-210	150-210	150-210	150-210	150-210

Table S3: Dihedrals Constraints (ϕ) and sugar phase angles (P) for simulated annealing (SA) of r-$C^1A^2A^3G^4A^5A^6C^7$ (**5**).

ϕ (deg)	C^1	A^2	A^3	G^4	A^5	A^6	C^7
α	-	-	-	-	-	-	-
β	-	140-220	140-220	140-220	140-220	140-220	140-220
γ	20-100	20-100	20-100	20-100	20-100	20-100	20-100
ε	140-340	140-340	140-340	140-340	140-340	140-340	-
ζ	-	-	-	-	-	-	-
P	0-60	0-60	0-60	0-60	0-60	0-60	0-60

Table S4: Dihedrals Constraints (ϕ) and sugar phase angles (P) for simulated annealing (SA) of d-$C^1A^2A^3G^4C^5A^6C^7$ (**2**).

ϕ (deg)	C^1	A^2	A^3	G^4	C^5	A^6	C^7
α	-	-	-	-	-	-	-
β	-	140-220	140-220	140-220	140-220	140-220	140-220
γ	20-100	20-100	20-100	20-100	20-100	20-100	20-100
ε	140-220	140-220	140-220	140-220	140-220	140-220	-
ζ	-	-	-	-	-	-	-
P	150-210	0-60	0-60	0-210	150-210	150-210	150-210

Table S5: Dihedrals Constraints (ϕ) and sugar phase angles (P) for simulated annealing (SA) of r-$C^1A^2A^3G^4C^5A^6C^7$ (**6**).

ϕ (deg)	C^1	A^2	A^3	G^4	C^5	A^6	C^7
α	-	-	-	-	-	-	-
β	-	140-220	140-220	140-220	140-220	140-220	140-220
γ	20-100	20-100	20-100	20-100	20-100	20-100	20-100
ε	140-340	140-340	140-340	140-340	140-340	140-340	-
ζ	-	-	-	-	-	-	-
P	0-60	0-60	0-60	0-60	0-60	0-60	0-60

Table S6: Dihedrals Constraints (ϕ) and sugar phase angles (P) for simulated annealing (SA) of d-$C^1A^2C^3G^4A^5A^6C^7$ (**3**).

ϕ (deg)	C^1	A^2	C^3	G^4	A^5	A^6	C^7
α	-	-	-	-	-	-	-
β	-	140-220	140-220	140-220	140-220	140-220	140-220
γ	20-100	20-100	20-100	20-100	20-100	20-100	20-100
ε	140-220	140-220	140-220	140-220	140-220	140-220	-
ζ	-	-	-	-	-	-	-
P	150-210	0-60	150-210	0-210	150-210	150-210	150-210

Table S7: Dihedrals Constraints (ϕ) and sugar phase angles (P) for simulated annealing (SA) of r-$C^1A^2C^3G^4A^5A^6C^7$ (**7**).

ϕ (deg)	C^1	A^2	C^3	G^4	A^5	A^6	C^7
α	-	-	-	-	-	-	-
β	-	140-220	140-220	140-220	140-220	140-220	140-220
γ	20-100	20-100	20-100	20-100	20-100	20-100	20-100
ε	140-340	140-340	140-340	140-340	140-340	140-340	-
ζ	-	-	-	-	-	-	-
P	0-60	0-60	0-210	0-210	0-60	0-60	0-60

Table S8: Dihedrals Constraints (ϕ) and sugar phase angles (P) for simulated annealing (SA) of d-$C^1A^2C^3G^4C^5A^6C^7$ (**4**).

ϕ (deg)	C^1	A^2	C^3	G^4	C^5	A^6	C^7
α	-	-	-	-	-	-	-
β	-	140-220	140-220	140-220	140-220	140-220	140-220
γ	20-100	20-100	20-100	20-100	20-100	20-100	20-100
ε	140-220	140-220	140-220	140-220	140-220	140-220	-
ζ	-	-	-	-	-	-	-
P	120-210	120-210	120-210	120-210	120-210	120-210	120-210

Table S9: Dihedrals Constraints (ϕ) and sugar phase angles (P) for simulated annealing (SA) of r-$C^1A^2C^3G^4C^5A^6C^7$ (**8**).

ϕ (deg)	C^1	A^2	C^3	G^4	C^5	A^6	C^7
α	-	-	-	-	-	-	-
β	-	140-220	140-220	140-220	140-220	140-220	140-220
γ	20-100	20-100	20-100	20-100	20-100	20-100	20-100
ε	140-340	140-340	140-340	140-340	140-340	140-340	-
ζ	-	-	-	-	-	-	-
P	0-60	0-60	0-60	0-210	0-60	0-60	0-60

Table S10: The oligomerization shift estimated from ^{1}H chemical shift at the neutral (N) state at 298 K for aromatic protons of ssDNA **1** – **4** using appropriate monomeric analogues as well as that of ssRNA **5** - **8** using appropriate monomeric analogues.

Compounds	$\Delta\delta_{N\,(\underline{M}-\underline{O})}$ [a]						
	C^1	**A^2**	**A^3/C^3**	**G^4**	**A^5/C^5**	**A^6**	**C^7**
d($C^{1'}A^2A^3G^4A^5A^6C^{7'}$)(**1**) r($C^1A^{2'}A^3G^4A^5A^6C^7$)(**5**)	0.202 (H5dC) 0.422 (H6dC) *0.430* (H5rC) *0.161* (H6rC)	0.330 (H8dA) 0.390 (H2dA) *0.249* (H8rA) *0.360* (H2rA)-	0.420 (H8dA) 0.600 (H2dA) *0.400* (H8rA) *0.330* (H2rA)-	0.343 (H8dG) *0.366* (H8rG)	0.424 (H8dA) 0.565 (H2dA) *0.349* (H8rA) *0.415* (H2rA)-	0.274 (H8dA) 0.505 (H2dA) *0.432* (H8rA) *0.221* (H2rA)-	0.317 (H5dC) 0.265 (H6dC) *0.538* (H5rC) *0.317* (H6rC)
d($C^{5'}A^{5'}A^5G^4C^3A^{3'}C^{3'}$)(**2**) r($C^{5'}A^{5'}A^5G^4C^3A^{3'}C^{3'}$)(**6**)	0.195 (H5dC) 0.415 (H6dC) *0.436* (H5rC) ***0.131*** (H6rC)	0.319 (H8dA) 0.362 (H2dA) *0.235*(H8rA) *0.368* (H2rA)	0.374 (H8dA) 0.530(H2dA) *0.415* (H8rA) *0.268* (H2rA)	0.248 (H8dG) *0.424* (H8rG)	0.461 (H5dC) 0.534 (H6dC) *0.613* (H5rC) *0.303*(H6rC)	0.146 (H8dA) 0.228 (H2dA) *0.215* (H8rA) *0.206* (H2rA	0.260 (H5dC) 0.205 (H6dC) *0.461* (H5rC) *0.288* (H6rC)
d($C^{5'}A^{5'}C^5G^4A^3A^{3'}C^{3'}$)(**3**) r($C^{5'}A^{5'}C^5G^4A^3A^{3'}C^{3'}$)(**7**)	0.188 (H5dC) 0.360 (H6dC) *0.379* (H5rC) ***0.091*** (H6rC)	0.127 (H8dA) 0.171 (H2dA) *0.178* (H8rA) *0.193* (H2rA)	0.390(H5dC) 0.506 (H6dC) *0.529*(H5rC) *0.355* (H6rC)	0.312 (H8dG) *0.258* (H8rG)	0.378 (H8dA) 0.486 (H2dA) *0.295* (H8rA) *0.369* (H2rA)	0.245 (H8dA) 0.470 (H2dA) *0.412* (H8rA) *0.193* (H2rA	0.300 (H5dC) 0.251 (H6dC) *0.526* (H5rC) *0.336* (H6rC)
d($C^1A^2C^3G^4C^5A^6C^{7'}$)(**4**) r($C^1A^{2'}C^3G^4C^5A^6C^{7'}$)(**8**)	0.186 (H5dC) 0.366 (H6dC) *0.380* (H5rC) ***0.071*** (H6rC)	0.123 (H8dA) 0.221 (H2dA) *0.172* (H8rA) *0.185* (H2rA)	0.331 (H5dC) 0.415 (H6dC) *0.547* (H5rC) *0.345* (H6rC)	0.161 (H8dG) *0.245* (H8rG)	0.390 (H5dC) 0.482 (H6dC) *0.525* (H5rC) *0.237* (H6rC)	0.128 (H8dA) 0.191 (H2dA) *0.190* (H8rA) *0.191* (H2rA	0.246 (H5dC) 0.196 (H6dC) *0.453* (H5rC) *0.279* (H6rC)

[a] The chemical shift difference [$\Delta\delta_{N(\underline{M}-\underline{O})}$, in ppm] between the monomer ($\underline{M}$ = NpEt / EtpNpEt / EtpN, where N = G, A or C) and Oligomers ($\underline{O}$) at the neutral (N) state. $\Delta\delta_{N\,(\underline{M}-\underline{O})} > 0$ signifies shielding and $\Delta\delta_{N\,(\underline{M}-\underline{O})} < 0$ signifies deshielding.

Table S11: Inter-residual (n-1) nOe contacts for ssDNA (**1-4**).

Nucleotide #*	Aromatic Proton	DNA (n-1)															
		d($^{5'}C^1A^2A^3G^4A^5A^6C^7$)				d($^{5'}C^1A^2A^3G^4C^5A^6C^7$)				d($^{5'}C^1A^2C^3G^4A^5A^6C^7$)				d($^{5'}C^1A^2C^3G^4C^5A^6C^7$)			
		H1'	H2'	H2''	H3'	H1'	H2'	H2''	H3'	H1'	H2'	H2''	H3'	H1'	H2'	H2''	H3'
1	H8/6(n)																
2	H8/6(n)	w	m	m	-	w	m	-	-	-	w	w	-	w	m	m	w
3	H8/6(n)	s	w	w	w	s	-	-	-	w	m	s	m	m	m	s	m
4	H8/6(n)	s	m	m	m	m	m	s	m	w	-	m	w	m	m	m	m
5	H8/6(n)	s	m	m	m	s	s	-	s	s	m	m	w	s	s	s	m
6	H8/6(n)	s	m	m	m	w	m	m	-	s	m	m	-	w	m	m	m
7	H8/6(n)	s	m	m	m	m	-	-	-	s	m	w	w	w	m	m	w

*shown as superscripts for each sequence as 1-7. w = weak, m = Medium, s = strong

Table S12: Intra-residual (n) nOe contacts for ssDNA (**1-4**).

Nucleotide #*	Aromatic Proton	DNA (n)															
		d($^{5'}C^1A^2A^3G^4A^5A^6C^7$)				d($^{5'}C^1A^2A^3G^4C^5A^6C^7$)				d($^{5'}C^1A^2C^3G^4A^5A^6C^7$)				d($^{5'}C^1A^2C^3G^4C^5A^6C^7$)			
		H1'	H2'	H2''	H3'	H1'	H2'	H2''	H3'	H1'	H2'	H2''	H3'	H1'	H2'	H2''	H3'
1	H8/6(n)	s	-	s	-	s	-	m	m	s	m	s	m	s	m	s	m
2	H8/6(n)	s	s	s	m	s	s	s	m	s	s	s	m	s	m	m	m
3	H8/6(n)	s	s	s	s	s	s	s	m	s	s	s	m	s	m	s	m
4	H8/6(n)	s	s	s	m	s	-	s	m	s	s	m	m	s	s	s	m
5	H8/6(n)	s	s	s	s	s	m	s	m	s	s	s	s	s	m	s	s
6	H8/6(n)	s	s	m	s	s	s	s	m	s	s	s	s	s	m	m	m
7	H8/6(n)	s	s	s	m	s	m	-	m	s	-	m	m	s	w	s	m

*shown as superscripts for each sequence as 1-7. w = weak, m = Medium, s = strong

Table S13: Inter-residual (n-1) nOe contacts for ssRNA (**5-8**).

Nucleotide #*	Aromatic Proton	RNA (n-1)															
		r($^{5'}C^1A^2A^3G^4A^5A^6C^7$)				r($^{5'}C^1A^2A^3G^4C^5A^6C^7$)				r($^{5'}C^1A^2C^3G^4A^5A^6C^7$)				r($^{5'}C^1A^2C^3G^4C^5A^6C^7$)			
		H1'	H2'	H2''	H3'	H1'	H2'	H2''	H3'	H1'	H2'	H2''	H3'	H1'	H2'	H2''	H3'
1	H8/6(n)																
2	H8/6(n)	w	s	s	m	m	-	w	m	m	w	m	m	w	s	s	m
3	H8/6(n)	m	m	s	m	m	s	m	w	w	m	w	s	m	m	s	m
4	H8/6(n)	-	m	s	m	m	m	w	-	w	m	w	s	-	m	s	m
5	H8/6(n)	m	s	m	-	-	m	s	-		s	m	s	m	s	m	-
6	H8/6(n)	m	s	s	m	s	m	s	m		w	m	s	m	s	s	m
7	H8/6(n)	m	s	s	m	m	m	m	-		w	s	w	m	s	s	m

*shown as superscripts for each sequence as 1-7. w = weak, m = Medium, s = strong

Table S14: Intra-residual (n-1) nOe contacts for ssRNA (**5-8**).

Nucleotide #*	Aromatic Proton	RNA (n-1)															
		r($^{5'}C^1A^2A^3G^4A^5A^6C^7$)				r($^{5'}C^1A^2A^3G^4C^5A^6C^7$)				r($^{5'}C^1A^2C^3G^4A^5A^6C^7$)				r($^{5'}C^1A^2C^3G^4C^5A^6C^7$)			
		H1'	H2'	H3'	H1'	H1'	H2'	H3'	H1'	H1'	H2'	H3'	H1'	H1'	H2'	H3'	H1'
1	**H8/6(n)**	s	s	s	s	s	s	s	s	s	s	s	s	s	s	s	s
2	**H8/6(n)**	s	m	s	s	s	m	s	s	s	m	s	s	s	m	s	s
3	**H8/6(n)**	s	s	s	s	s	s	s	s	s	s	s	s	s	s	s	s
4	**H8/6(n)**	s	m	s	s	s	m	s	s	s	m	s	s	s	m	s	s
5	**H8/6(n)**	s	s	s	s	s	s	s	s	s	s	s	s	s	s	s	s
6	**H8/6(n)**	s	s	s	s	s	s	s	s	s	s	s	s	s	s	s	s
7	**H8/6(n)**	s	s	s	s	s	s	s	s	s	s	s	s	s	s	s	s

*shown as superscripts for each sequence as 1-**7**. w = weak, m = Medium, s = strong

Table S15: 5'-r ($G^1A^2A^3G^4A^5G^6A^7A^8G^9C^{10}$).r($G^1C^2U^3U^4C^5U^6C^7U^8U^9C^{10}$)-3'
Structure taken from NCBI structure database (Biochemistry 1998, 37, 73 – 80)
Distances (Å) between aromatic (n) and sugar protons n, (n-1) of a Cannonical A-RNA duplex

n	n				n-1			
	H1′	H2′	H2′′	H3′	H1′	H2′	H2′′	H3′
H6C1	3.72	2.86	3.93	2.21	-	-	-	-
H8A2	3.85	2.21	3.66	4.02	4.48	3.39	2.80	5.17
H6T3	3.67	2.27	3.71	3.83	3.53	3.44	2.13	4.79
H8G4	3.86	2.17	3.48	4.25	4.53	2.85	2.63	4.6
H8G5	3.83	2.37	3.83	4.08	3.08	3.63	2.21	4.91
H6C6	3.69	2.26	3.72	3.73	3.57	3.17	2.15	4.79
H6C7	3.71	2.09	3.53	3.69	4.12	3.22	2.35	4.85
H8A8	3.86	2.17	3.62	4.0	3.99	3.32	2.40	4.91
H6T9	3.66	2.38	3.82	3.74	3.56	3.36	2.26	4.94
H8G10	3.86	2.33	3.80	3.79	4.24	3.18	3.00	5.30

Table S16: Canonical B-DNA Duplex where this is one strand.
Pdb taken from NCBI structure database , Reference J. Mol. Biol (1998) 284,1453-1463
5′-d $C^1A^2T^3G^4G^5C^6C^7A^8T^9G^{10}$-3′. Distances (Å) between aromatic (n) and sugar protons n, (n-1) of a Cannonical B-DNA duplex

n	n				n-1			
	H1′	H2′	H2′′	H3′	H1′	H2′	H2′′	H3′
H6C1	3.72	2.86	3.93	2.21	-	-	-	-
H8A2	3.85	2.21	3.66	4.02	4.48	3.39	2.80	5.17
H6T3	3.67	2.27	3.71	3.83	3.53	3.44	2.13	4.79
H8G4	3.86	2.17	3.48	4.25	4.53	2.85	2.63	4.6
H8G5	3.83	2.37	3.83	4.08	3.08	3.63	2.21	4.91
H6C6	3.69	2.26	3.72	3.73	3.57	3.17	2.15	4.79
H6C7	3.71	2.09	3.53	3.69	4.12	3.22	2.35	4.85
H8A8	3.86	2.17	3.62	4.0	3.99	3.32	2.40	4.91
H6T9	3.66	2.38	3.82	3.74	3.56	3.36	2.26	4.94
H8G10	3.86	2.33	3.80	3.79	4.24	3.18	3.00	5.30

Table S17: Distances (Å) between aromatic (n) and sugar protons n, (n-1) of Single Stranded Deoxyribo Heptamer 5′-$C^1A^2A^3G^4A^5A^6C^7$-3′.

n	n				n-1			
	H1′	H2′	H2′′	H3′	H1′	H2′	H2′′	H3′
H6C1	3.72	2.67	4.0	4.43	-	-	-	-
H8A2	3.85	2.14	3.47	4.33	3.62	6.43	5.73	6.12
H8A3	3.89	2.25	3.42	4.52	3.94	3.48	2.48	4.99
H8G4	3.91	2.23	3.46	4.29	4.45	3.43	2.55	4.92
H8A5	2.46	3.61	4.22	5.35	4.95	3.08	2.81	4.52
H8A6	3.92	2.36	3.66	4.51	4.68	2.85	2.75	4.79
H6C7	3.75	2.25	3.60	4.30	3.60	3.44	2.26	4.95

Table S18: Distances (Å) between aromatic (n) and sugar protons n, (n-1) of Single Stranded ribo Heptamer 5′-$C^1A^2A^3G^4A^5A^6C^7$-3′

n	n				n-1			
	H1′	H2′	H2′′	H3′	H1′	H2′	H2′′	H3′
H6C1	3.77	2.62	-	2.24	-	-	-	-
H8A2	3.77	2.22	-	4.61	4.94	2.65	-	4.64
H8A3	3.88	3.64	-	2.45	2.89	5.14	-	4.61
H8G4	3.91	3.49	-	2.28	5.36	2.79	-	3.28
H8A5	3.75	4.16	-	2.92	4.62	2.50	-	4.09
H8A6	3.81	3.98	-	2.80	4.91	2.45	-	3.08
H6C7	3.68	3.72	-	2.56	4.98	2.51	-	3.06

Table S19: Distances (Å) between aromatic (n) and sugar protons n, (n-1) of Single Stranded Deoxyribo Heptamer 5′-$C^1A^2A^3G^4C^5A^6C^7$-3′

n	n				n-1			
	H1′	H2′	H2′′	H3′	H1′	H2′	H2′′	H3′
H6C1	3.73	2.47	3.88	4.26	-	-	-	-
H8A2	3.61	2.46	2.65	4.71	6.66	3.77	4.92	3.57
H8A3	3.88	2.29	3.37	4.62	3.12	4.61	2.87	4.71
H8G4	3.89	2.29	3.59	4.50	3.03	4.32	2.59	4.21
H6C5	3.75	2.28	3.66	4.32	2.68	4.47	2.86	4.31
H8A6	3.80	2.12	3.25	4.41	5.87	3.56	3.60	3.37
H6C7	3.73	2.26	3.72	4.10	5.26	3.15	3.21	2.21

Table S20: Distances (Å) between aromatic (n) and sugar protons n, (n-1) of Single Stranded ribo Heptamer 5′-$C^1A^2A^3G^4C^5A^6C^7$-3′

n	n				n-1			
	H1′	H2′	H2′′	H3′	H1′	H2′	H2′′	H3′
H6C1	2.26	4.21	-	5.68	-	-	-	-
H8A2	3.77	4.08	-	2.95	2.45	4.84	-	4.37
H8A3	3.78	4.07	-	2.92	4.73	2.30	-	2.71
H8G4	3.78	4.05	-	2.90	4.76	2.33	-	2.68
H6C5	3.66	3.82	-	2.60	4.84	2.41	-	2.85
H8A6	3.73	4.15	-	3.21	4.93	2.66	-	2.51
H6C7	3.69	3.63	-	2.46	5.40	3.27	-	2.25

Table S21: Distances (Å) between aromatic (n) and sugar protons n, (n-1) of Single Stranded Deoxyribo Heptamer 5′-$C^1A^2C^3G^4A^5A^6C^7$-3′

n	n				n-1			
	H1′	H2′	H2′′	H3′	H1′	H2′	H2′′	H3′
H6C1	3.74	2.43	3.84	4.26	-	-	-	-
H8A2	3.84	2.53	3.96	2.22	6.32	4.63	4.01	4.40
H6C3	3.65	2.12	2.83	4.47	4.92	2.93	4.47	4.69
H8G4	3.90	3.33	4.37	2.34	3.92	4.02	2.44	4.95
H8A5	3.89	2.18	3.55	4.33	5.56	2.89	4.15	4.17
H8A6	3.90	2.26	3.52	4.47	4.19	3.26	2.33	4.63
H6C7	3.75	2.26	3.66	4.18	3.70	3.53	2.29	4.97

Table S22: Distances (Å) between aromatic (n) and sugar protons n, (n-1) of Single Stranded ribo Heptamer 5′-$C^1A^2C^3G^4A^5A^6C^7$-3′

n	n				n-1			
	H1′	H2′	H2′′	H3′	H1′	H2′	H2′′	H3′
H6C1	3.70	2.22	-	2.17	-	-	-	-
H8A2	3.82	2.41	-	2.17	6.27	3.99	-	5.89
H6C3	3.56	2.06	-	4.42	5.31	3.17	-	5.17
H8G4	3.73	2.36	-	4.74	4.90	4.53	-	5.84
H8A5	3.90	3.56	-	2.43	3.04	5.04	-	4.92
H8A6	3.93	3.41	-	2.23	5.34	2.69	-	4.05
H6C7	3.76	3.23	-	2.18	5.16	2.53	-	3.83

Table S23: Distances (Å) between aromatic (n) and sugar protons n, (n-1) of Single Stranded Deoxyribo Heptamer 5′-$C^1A^2C^3G^4C^5A^6C^7$-3′

n	n				n-1			
	H1′	H2′	H2′′	H3′	H1′	H2′	H2′′	H3′
H6C1	3.76	2.22	3.46	4.46	-	-	-	-
H8A2	3.91	2.29	3.52	4.53	4.63	2.91	2.35	4.21
H6C3	3.76	2.18	3.40	4.40	3.96	3.55	2.49	5.12
H8G4	3.90	2.29	3.54	4.49	4.61	3.01	2.42	4.48
H6C5	3.76	2.16	3.49	4.24	3.72	3.78	2.41	5.10
H8A6	3.89	2.88	4.28	3.83	4.18	3.42	2.59	5.12
H6C7	3.75	2.21	3.61	4.22	4.34	2.54	2.39	4.28

Table S24. Distances (Å) between aromatic (n) and sugar protons n, (n-1) of Single Stranded ribo Heptamer 5′-$C^1A^2C^3G^4C^5A^6C^7$-3′

n	n				n-1			
	H1′	H2′	H2′′	H3′	H1′	H2′	H2′′	H3′
H6C1	3.74	3.25	-	2.23	-	-	-	-
H8A2	3.89	3.65	-	2.39	4.93	2.60	-	4.48
H6C3	3.74	3.36	-	2.31	4.91	2.65	-	4.59
H8G4	3.90	2.28	-	4.38	5.98	3.49	-	2.93
H6C5	3.67	3.75	-	2.63	2.88	4.87	-	4.55
H8A6	3.78	4.02	-	3.02	5.08	2.74	-	2.43
H6C7	3.70	3.55	-	2.39	5.56	3.42	-	2.35

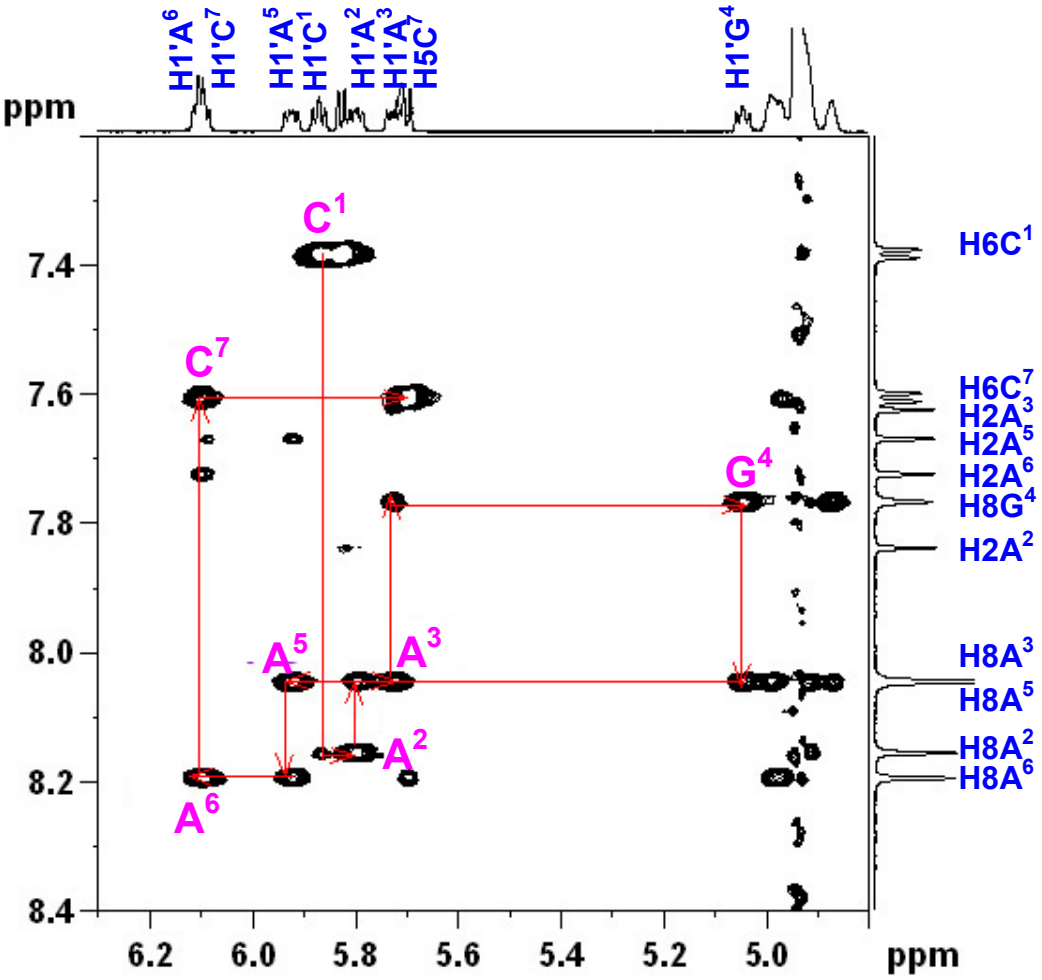


Fig S1. NOESY footprint of d($^{5'}C^1A^2A^3G^4A^5A^6C^7$) (**1**) showing the connectivity of nucleotide residues.

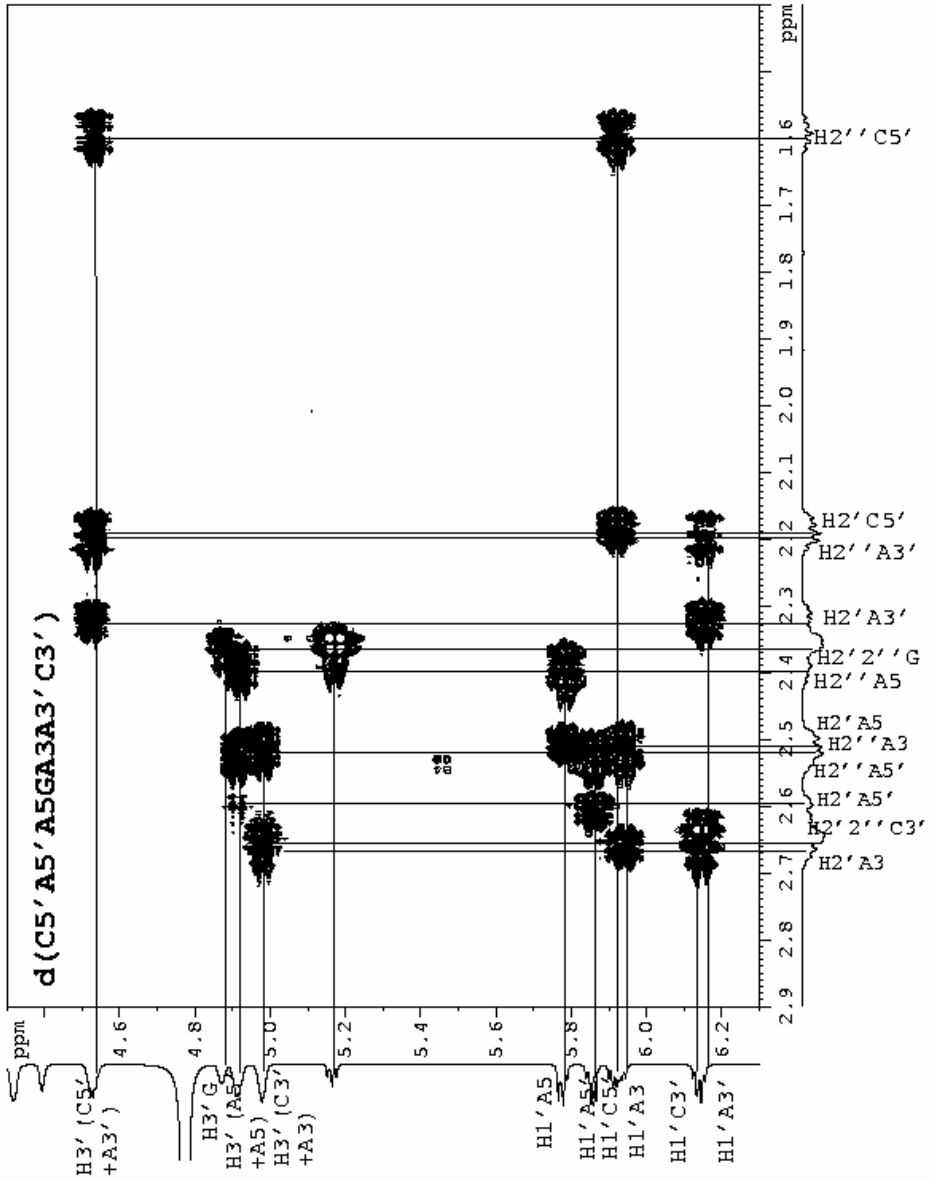


Figure S2. The expanded ^{31}P coupled DQF-COSY spectra of H1'/H2'/H2"/H3' region for d($^{5'}$C^{1}A^{2}A^{3}G^{4}A^{5}A^{6}C^{7}) (**1**) at 298 K. The region 6.3 – 1.4 ppm in F1 and region 6.3 – 4.5 ppm in F2 dimension of the d($^{5'}$C^{1}A^{2}A^{3}G^{4}A^{5}A^{6}C^{7}) (**1**) showing the spin connectivity.

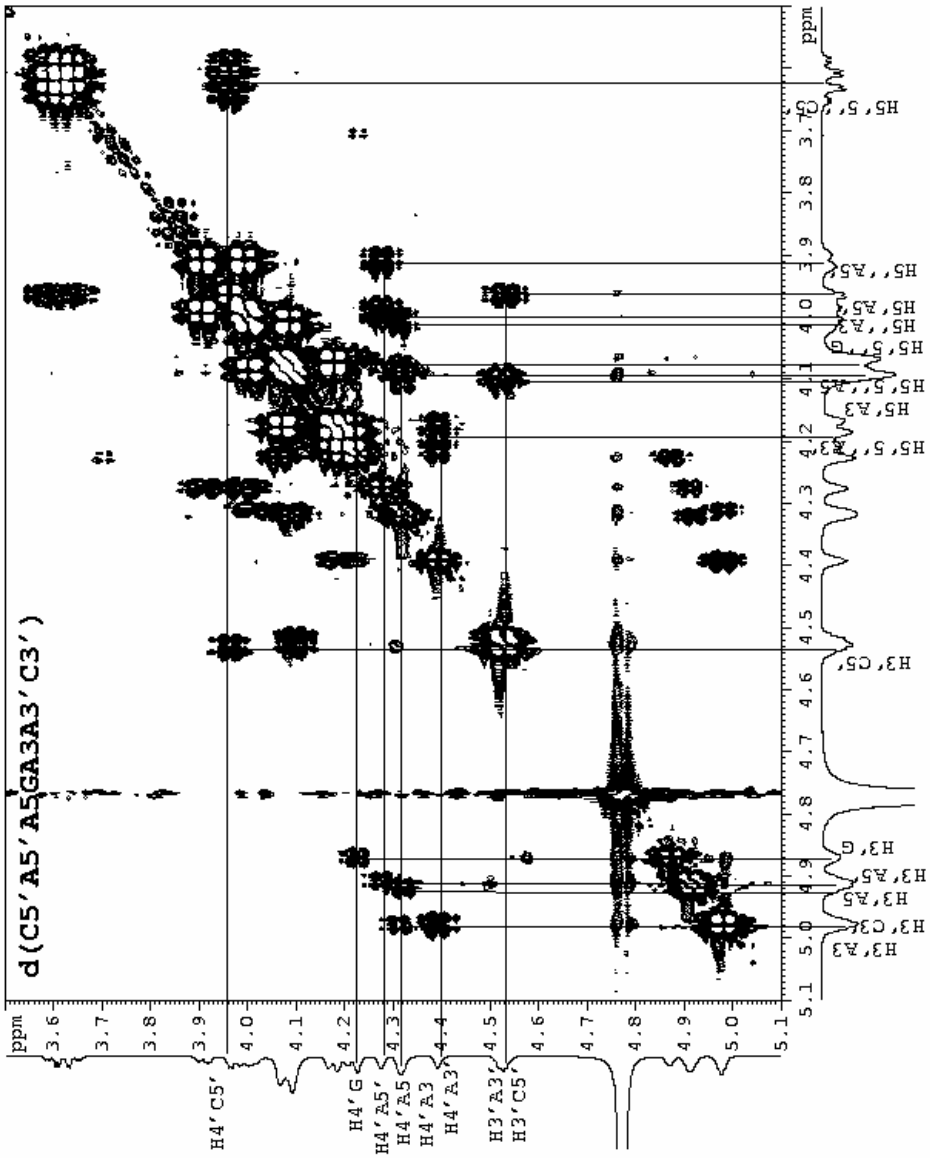


Figure S3. The expanded ^{31}P coupled DQF-COSY spectra of H3'/H4'/H5'/H5" region for d($^{5'}$C^{1}A^{2}A^{3}G^{4}A^{5}A^{6}C^{7}) (**1**) at 298 K. The region 5.1 – 3.5 ppm in both F1 and F2 dimensions of the d($^{5'}$C^{1}A^{2}A^{3}G^{4}A^{5}A^{6}C^{7}) (**1**) showing the spin connectivity.

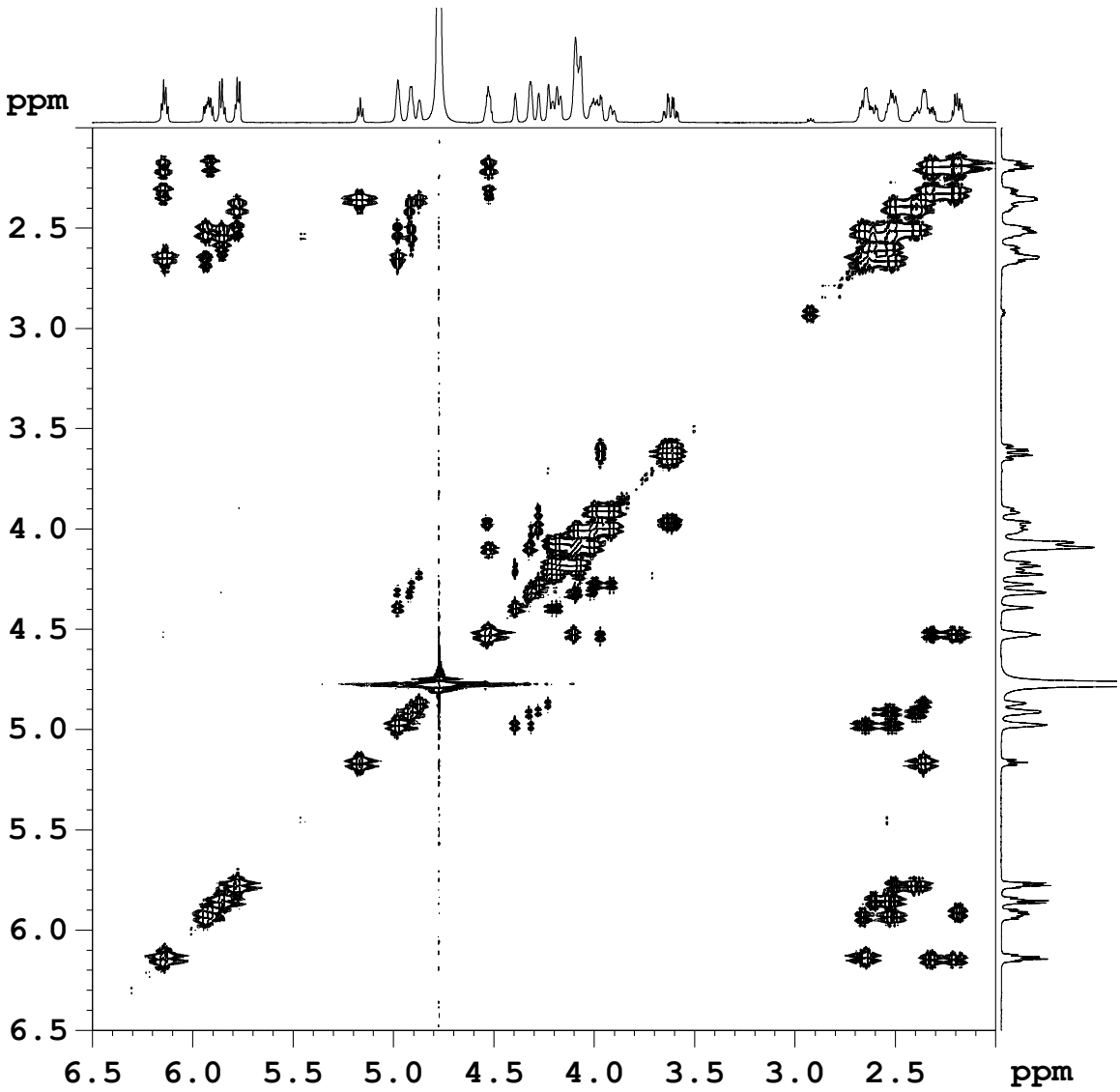


Figure S3.1. The ^{31}P decoupled DQF-COSY spectrum of d($^{5'}$C^{1}A^{2}A^{3}G^{4}A^{5}A^{6}C^{7}) (**1**) at 298 K. For assignments see **S2** and **S3**.

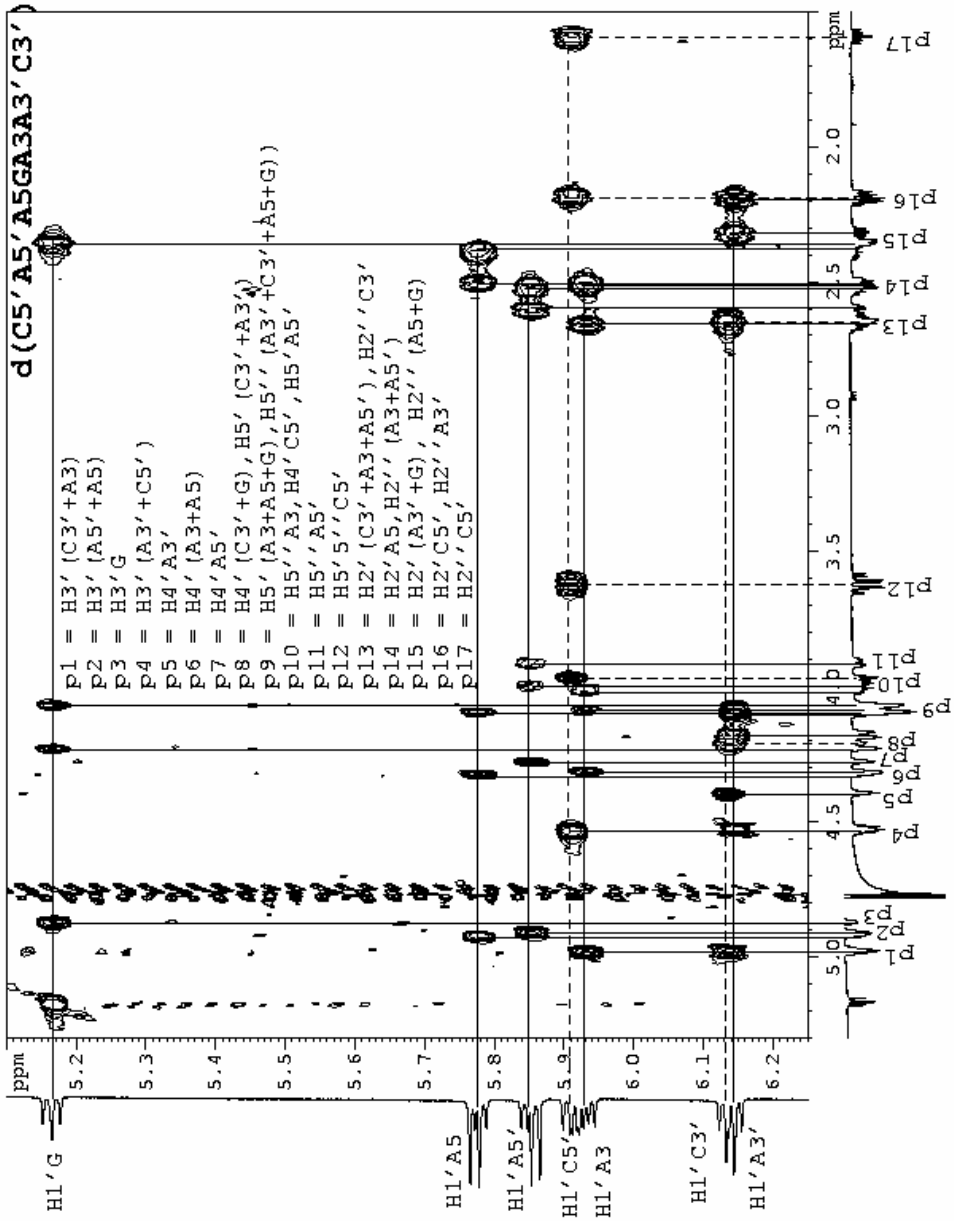


Figure S4. Expanded TOCSY spectra of the H2'/H2"/H3'/H4'/H5'/H5" region (1.5 – 5.3 ppm in F1 direction) to anomeric (H1') region (5.1 – 6.25 ppm in F2 direction) for d($^{5'}$C^{1}A^{2}A^{3}G^{4}A^{5}A^{6}C^{7}) (**1**) at 298 K.

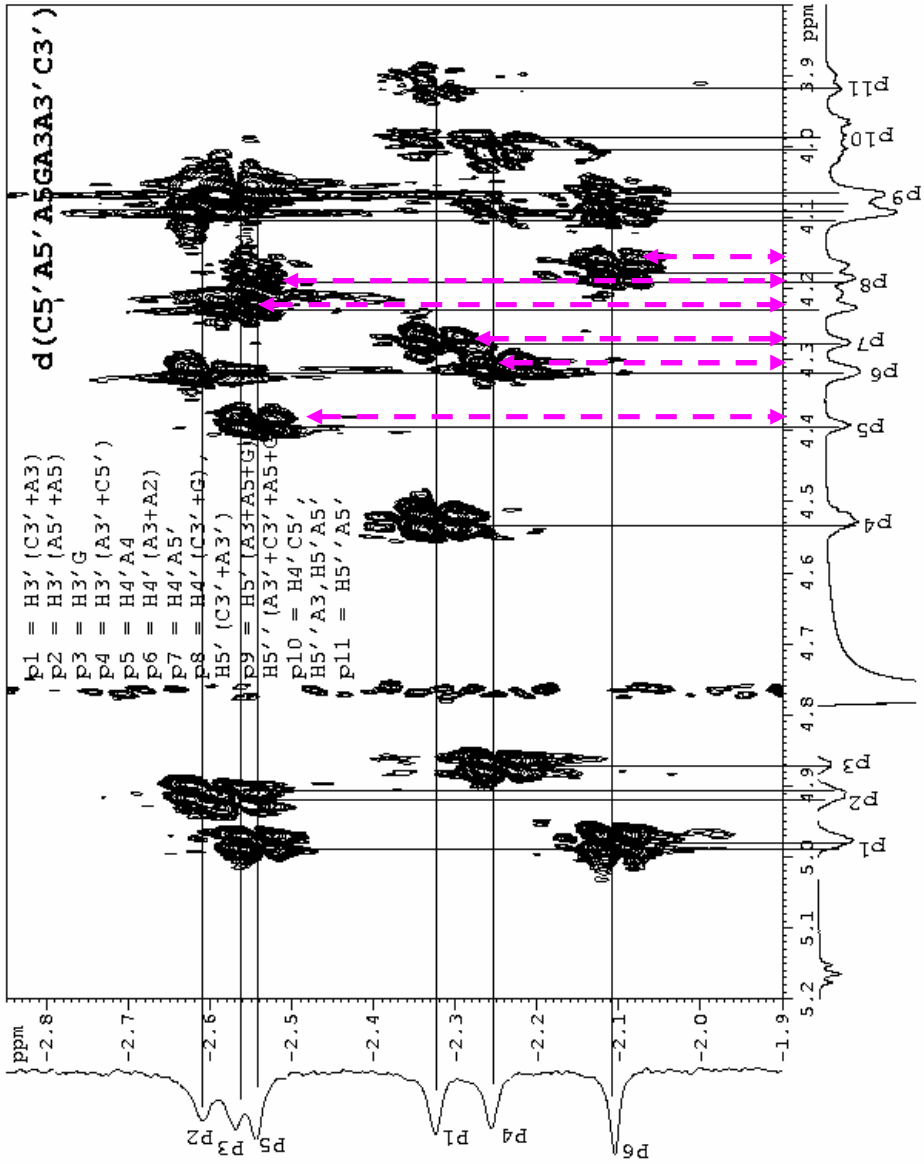


Figure S5. Expanded ^{31}P - ^{1}H correlation spectroscopy of ^{31}P region (-1.9 – -2.85 ppm in F2 direction) to H3'/H4'/H5'/H5" region (5.2 – 3.8 ppm in F1 direction) for d($^{5'}C^{1}p_{1}A^{2}p_{2}A^{3}p_{3}G^{4}p_{4}A^{5}p_{5}A^{6}p_{6}C^{7}$) (**1**) at 298 K. ← - - → H4' – P connectivity.

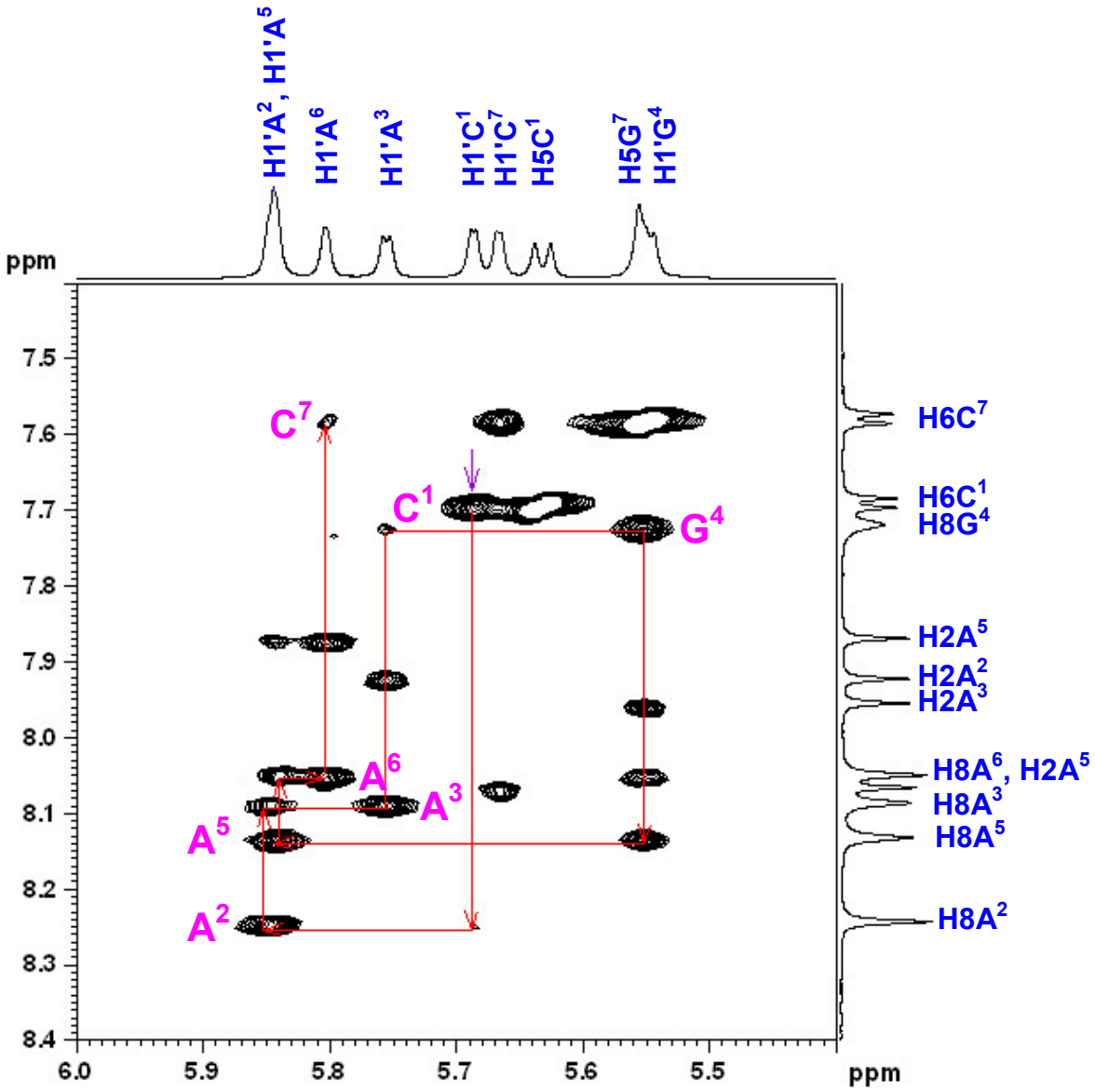


Figure S6. NOESY footprint of r($^{5'}C^1A^2A^3G^4A^5A^6C^7$) (**5**) showing the connectivity of nucleotide residues

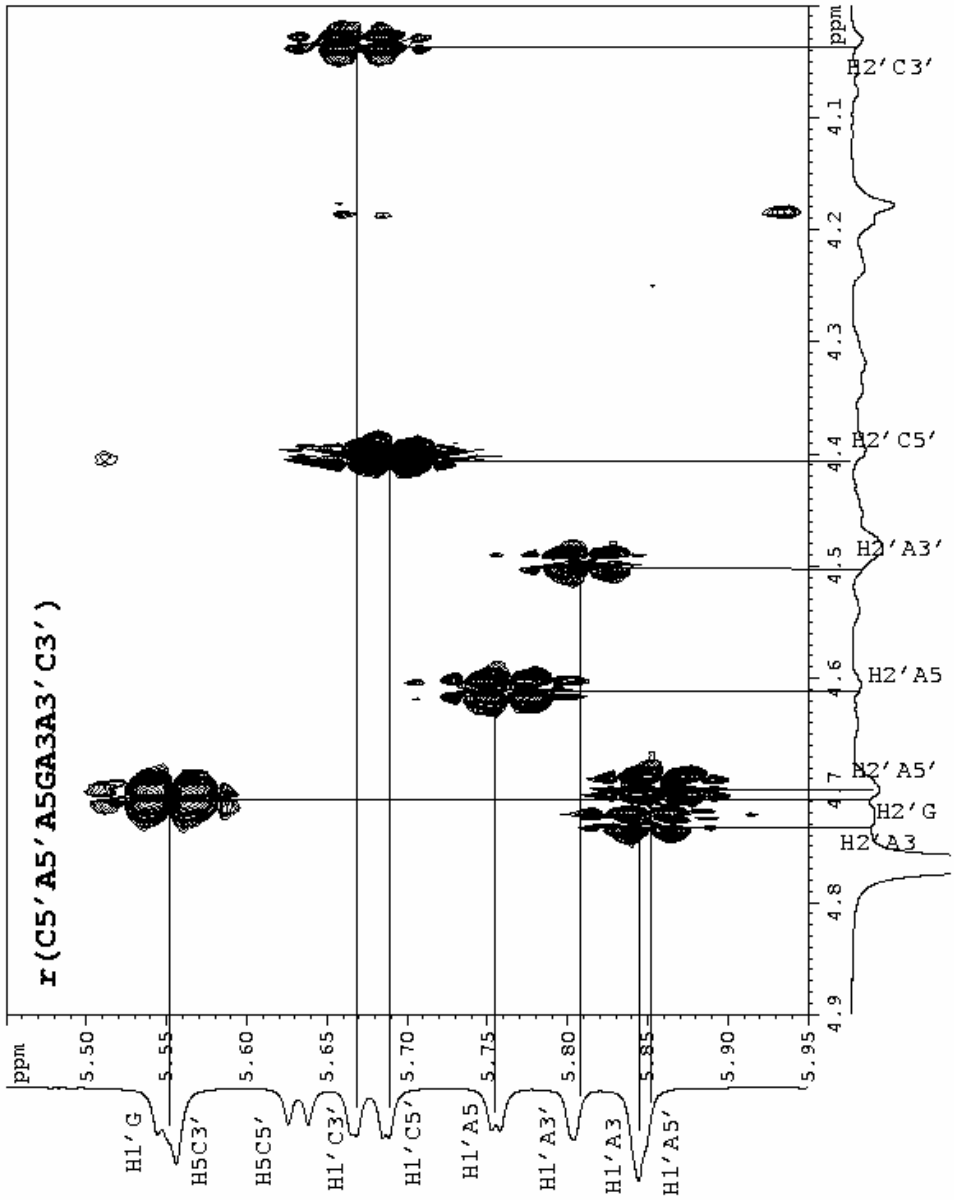


Figure S7. The expanded ^{31}P coupled DQF-COSY spectra of the anomeric H1' region to the H2' for r($^{5'}$C^{1}A^{2}A^{3}G^{4}A^{5}A^{6}C^{7}) (**5**) at 298 K. The spin connectivity between anomeric H1' region (5.45 – 5.95 ppm in F2) and H2' region (4.0 – 4.9 ppm in F1) of r($^{5'}$C^{1}A^{2}A^{3}G^{4}A^{5}A^{6}C^{7}) (**5**) have been shown.

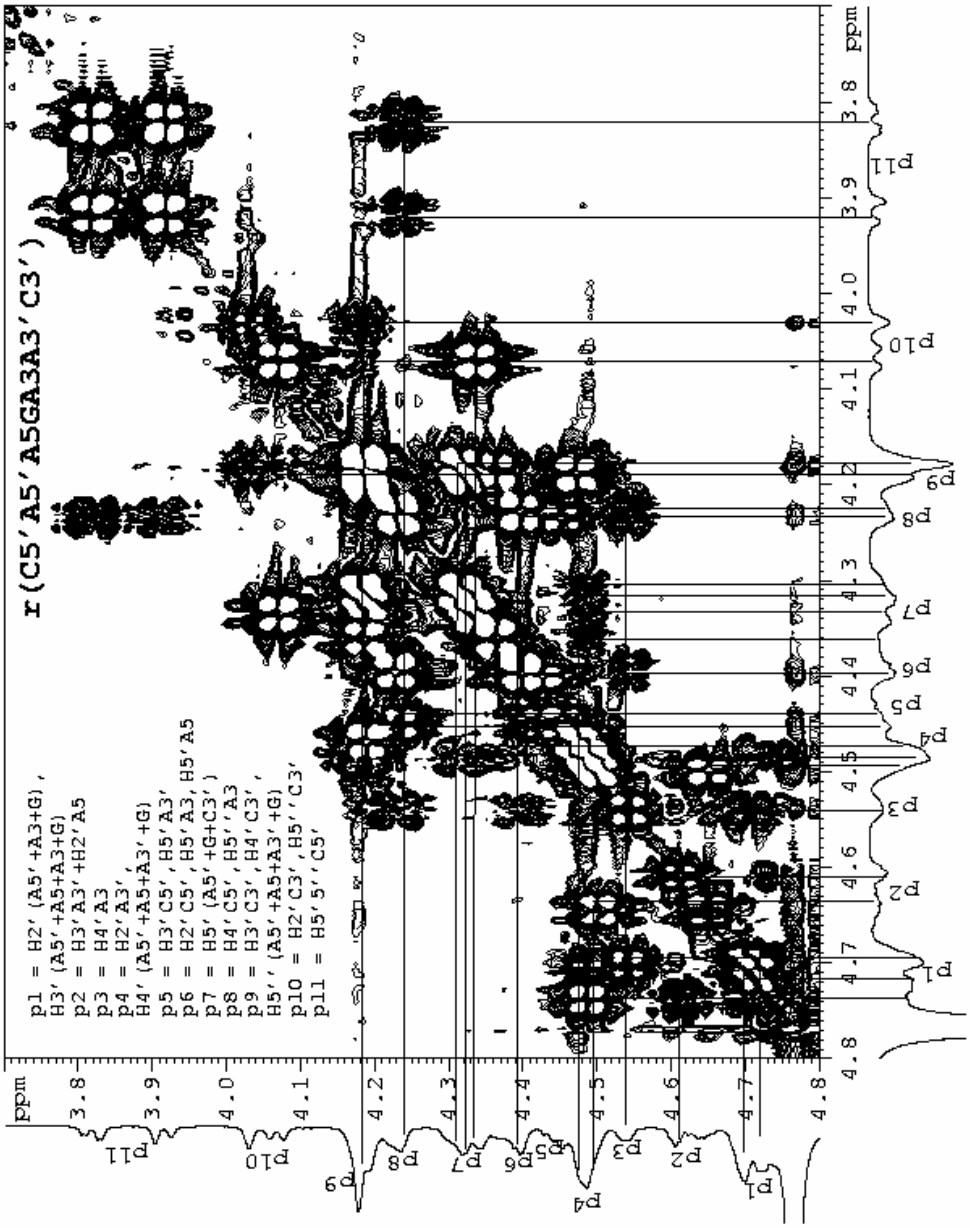


Figure S8. The expanded ^{31}P coupled DQF-COSY spectra of H2'/H3'/H4'/H5'/H5" region for r($^{5'}$C^{1}A^{2}A^{3}G^{4}A^{5}A^{6}C^{7}) (**5**) at 298 K. The region 4.8 – 3.7 ppm in both F1 and F2 dimensions of the r($^{5'}$C^{1}A^{2}A^{3}G^{4}A^{5}A^{6}C^{7}) (**5**) showing the spin connectivity.

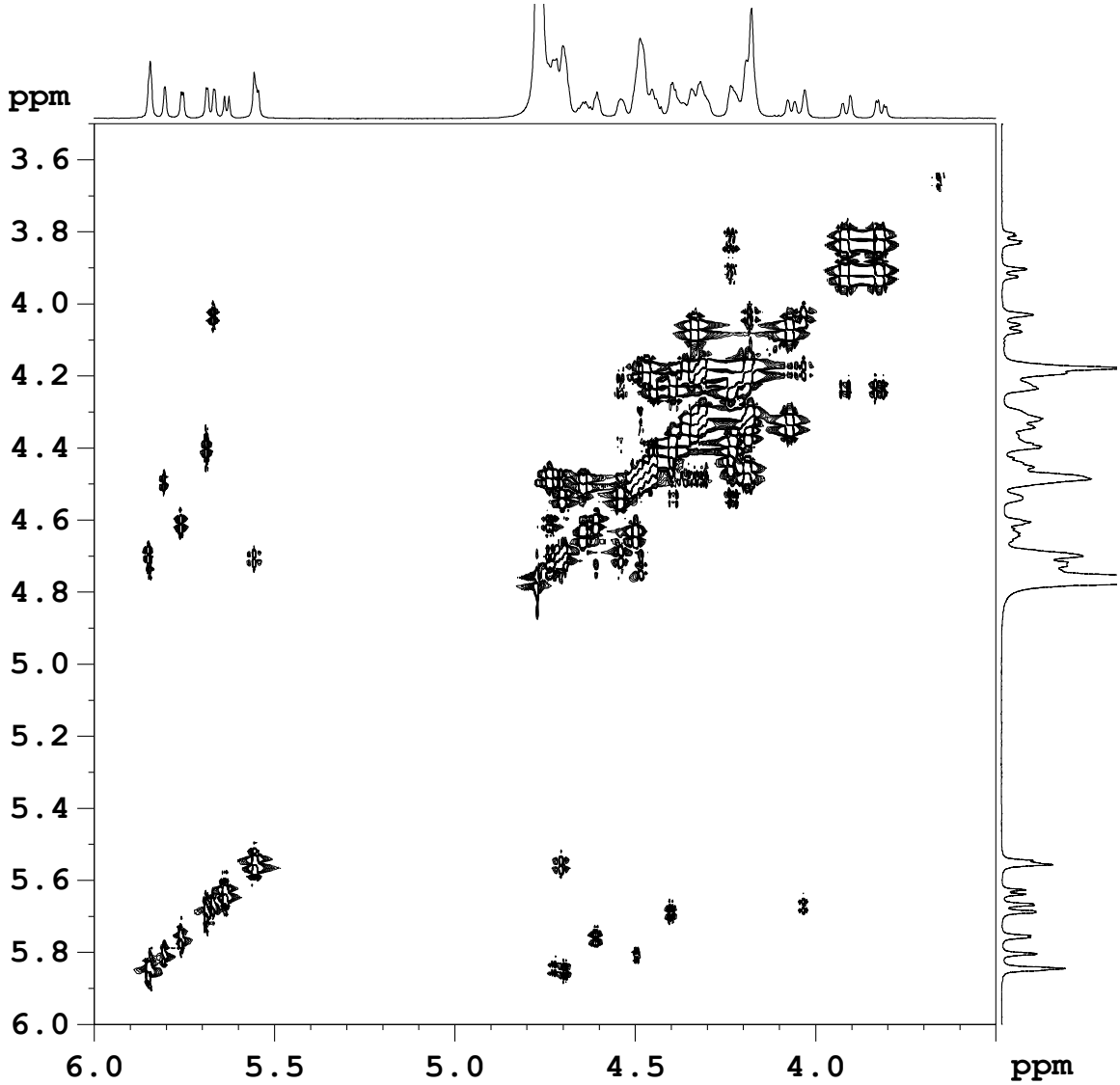


Figure S8.1. The ^{31}P decoupled DQF-COSY spectrum of r($^{5'}$C^{1}A^{2}A^{3}G^{4}A^{5}A^{6}C^{7}) (**5**) at 298 K. For assignments see S**7** and **S8**.

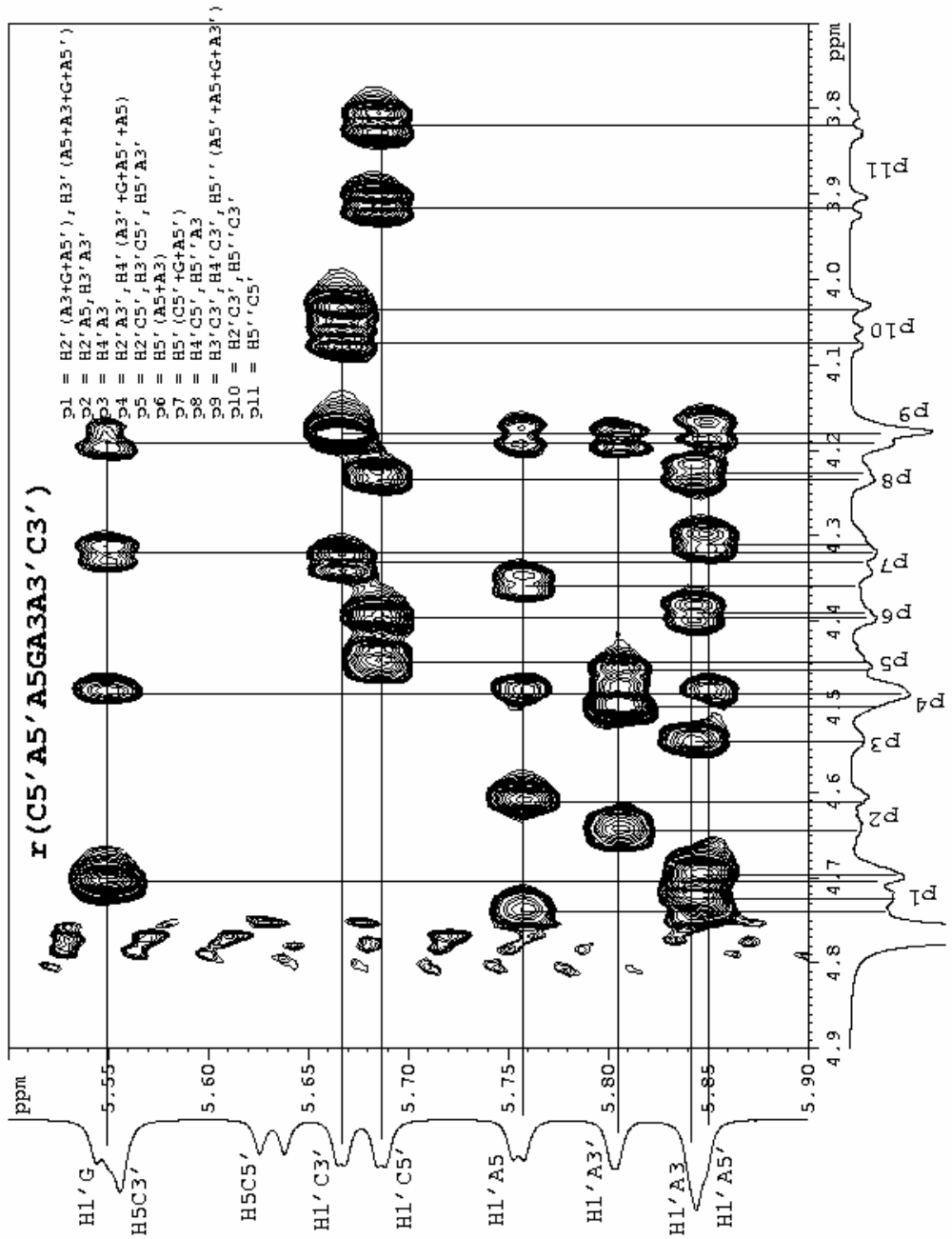


Figure S9. Expanded TOCSY spectra of the H2'/H3'/H4'/H5'/H5" region (4.9 – 3.7 ppm in F1 direction) to anomeric (H1') region (5.5 – 5.9 ppm in F2 direction) for r($^{5'}$C^{1}A^{2}A^{3}G^{4}A^{5}A^{6}C^{7}) (**5**) at 298 K.

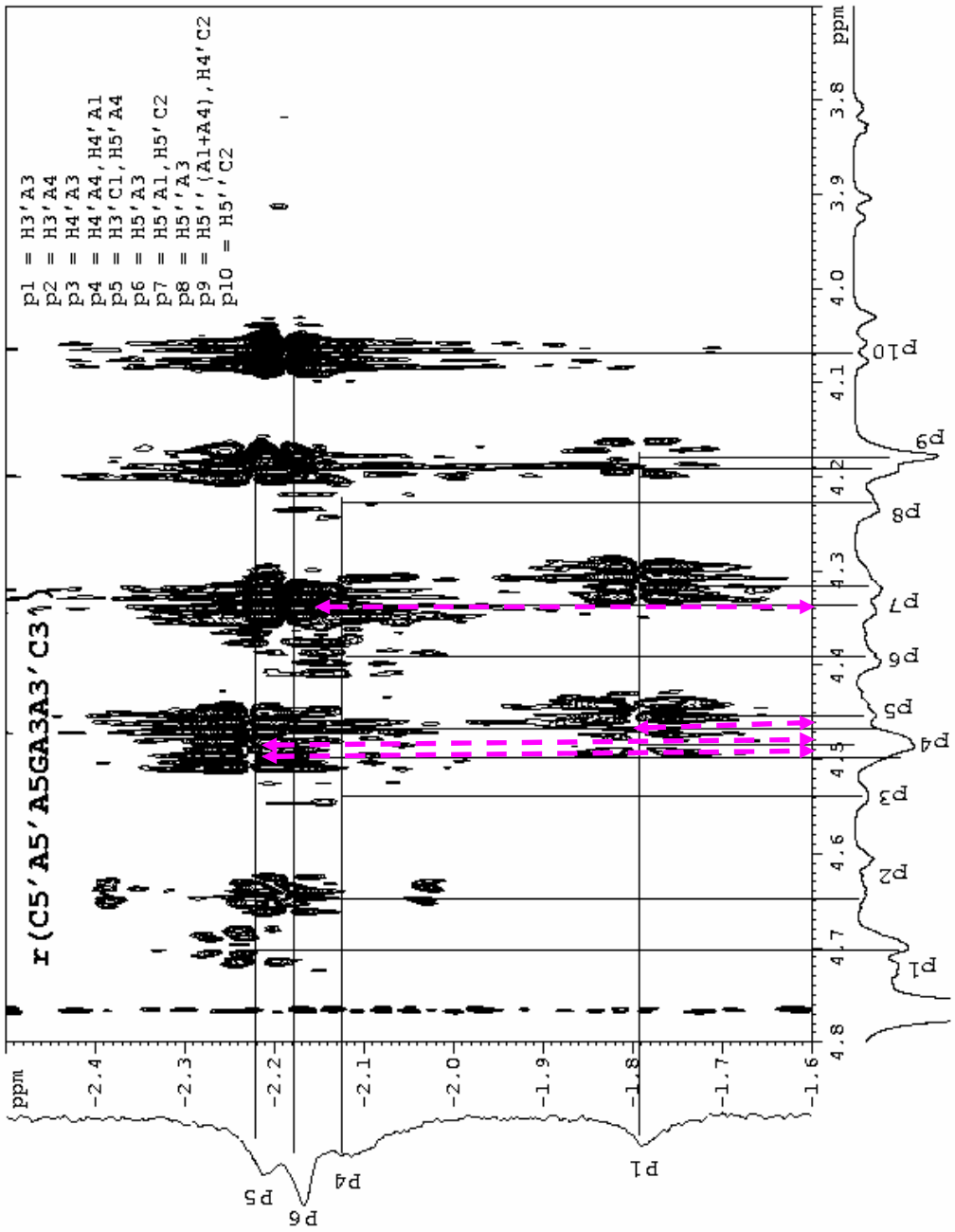


Figure S10. Expanded ^{31}P - ^{1}H correlation spectroscopy of ^{31}P region (-1.6 – -2.5 ppm in F2 direction) to H2'/H3'/H4'/H5'/H5" region (4.8 – 3.7 ppm in F1 direction) for r($^{5'}C^{1}p_{1}A^{2}p_{2}A^{3}p_{3}G^{4}p_{4}A^{5}p_{5}A^{6}p_{6}C^{7}$) (**5**) at 298 K. ← - - → H4' – P connectivity.

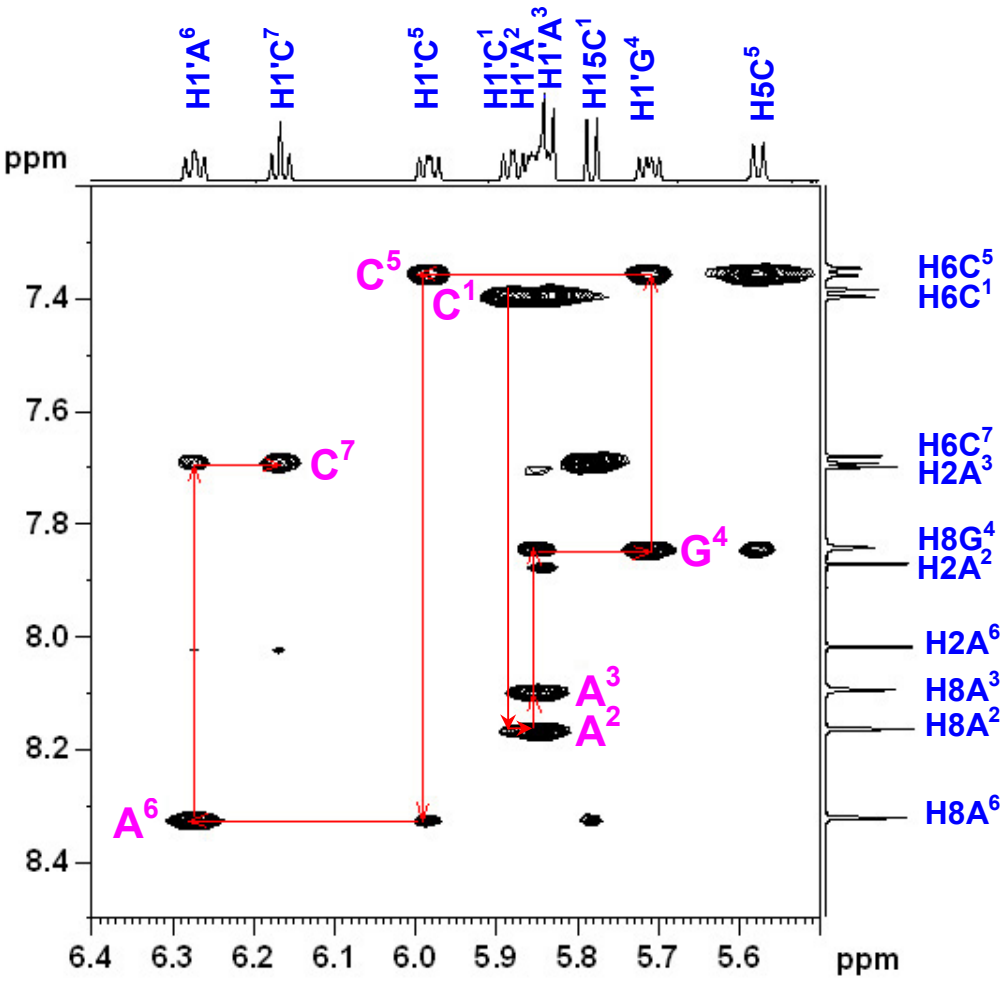


Figure S11. NOESY footprint of d($^{5'}$C^{1}A^{2}A^{3}G^{4}C^{5}A^{6}C^{7}) (**2**) showing the connectivity of nucleotide residues.

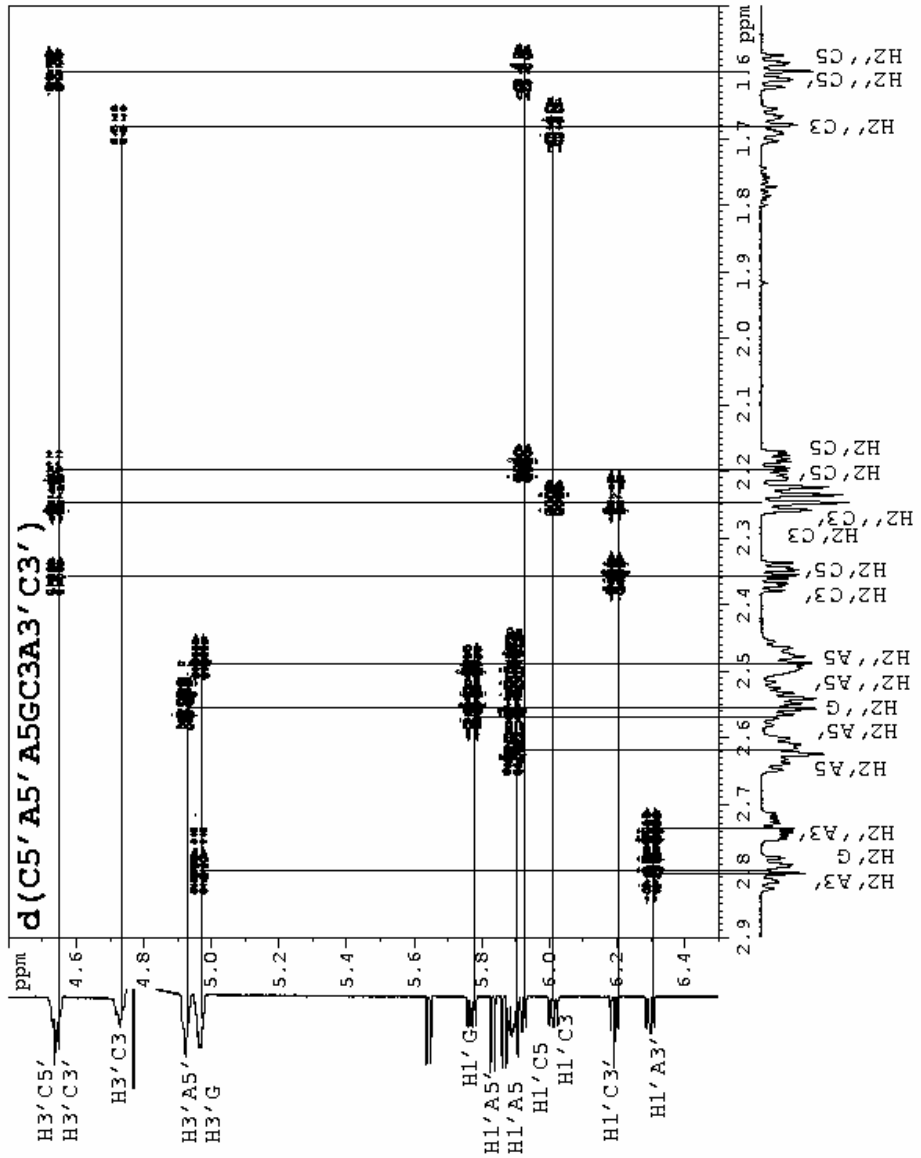


Figure S12. The expanded ^{31}P coupled DQF-COSY spectra of H1'/H2'/H2''/H3' region for d($^{5'}$C^{1}A^{2}A^{3}G^{4}C^{5}A^{6}C^{7}) (**2**) at 298 K. The region 2.9 – 1.5 ppm in F1 and region 6.45 – 4.4 ppm in F2 dimension of the d($^{5'}$C^{1}A^{2}A^{3}G^{4}C^{5}A^{6}C^{7}) (**2**) showing the spin connectivity.

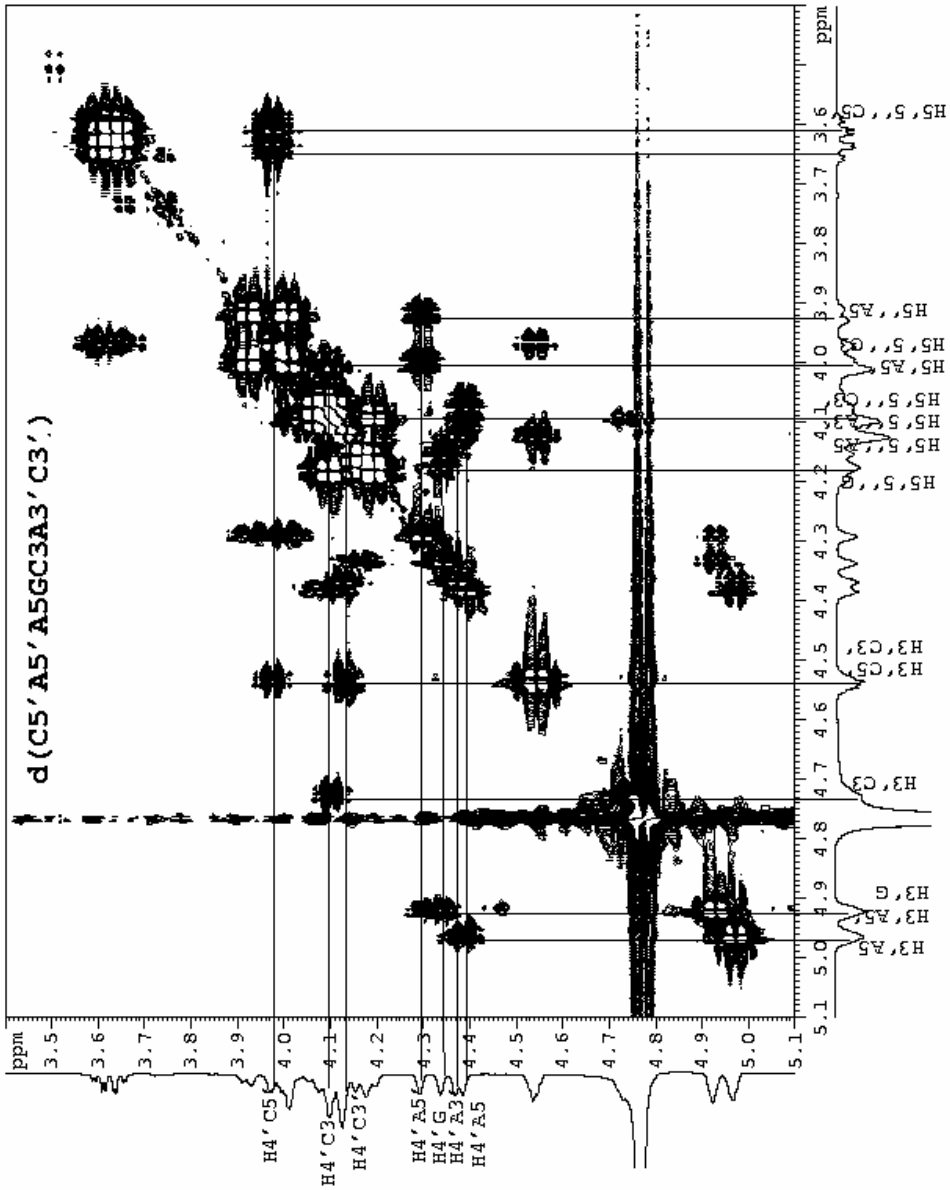


Figure S13. The expanded ^{31}P coupled DQF-COSY spectra of H3'/H4'/H5'/H5" region for d($^{5'}$C^{1}A^{2}A^{3}G^{4}C^{5}A^{6}C^{7}) (**2**) at 298 K. The region 5.1 – 3.4 ppm in both F1 and F2 dimensions of the d($^{5'}$C^{1}A^{2}A^{3}G^{4}C^{5}A^{6}C^{7}) (**2**) showing the spin connectivity.

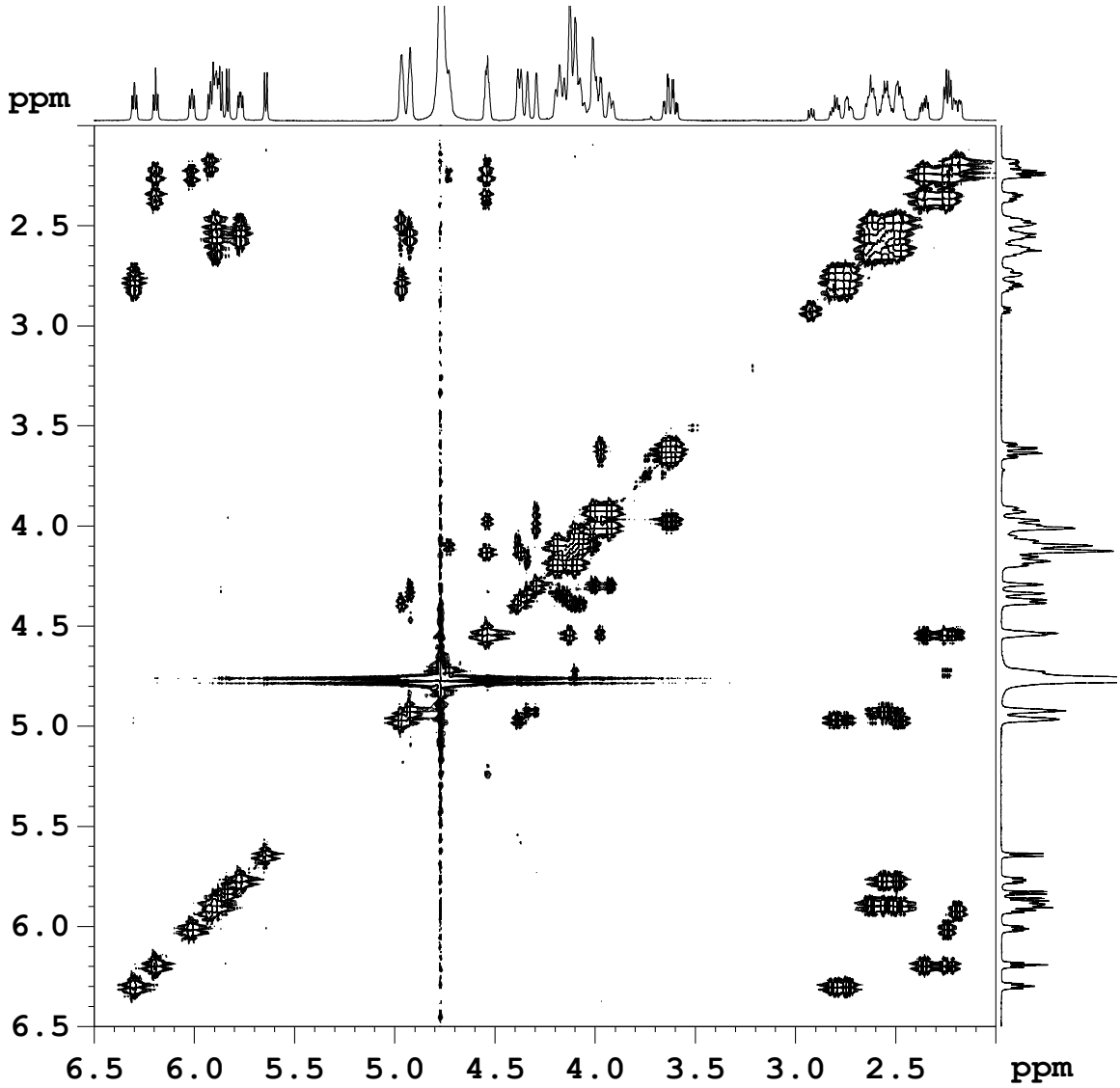


Figure S13.1. The ^{31}P decoupled DQF-COSY spectrum of d($^{5'}$C^{1}A^{2}A^{3}G^{4}C^{5}A^{6}C^{7}) (**2**) at 298 K. For assignments see **S12** and **S13**.

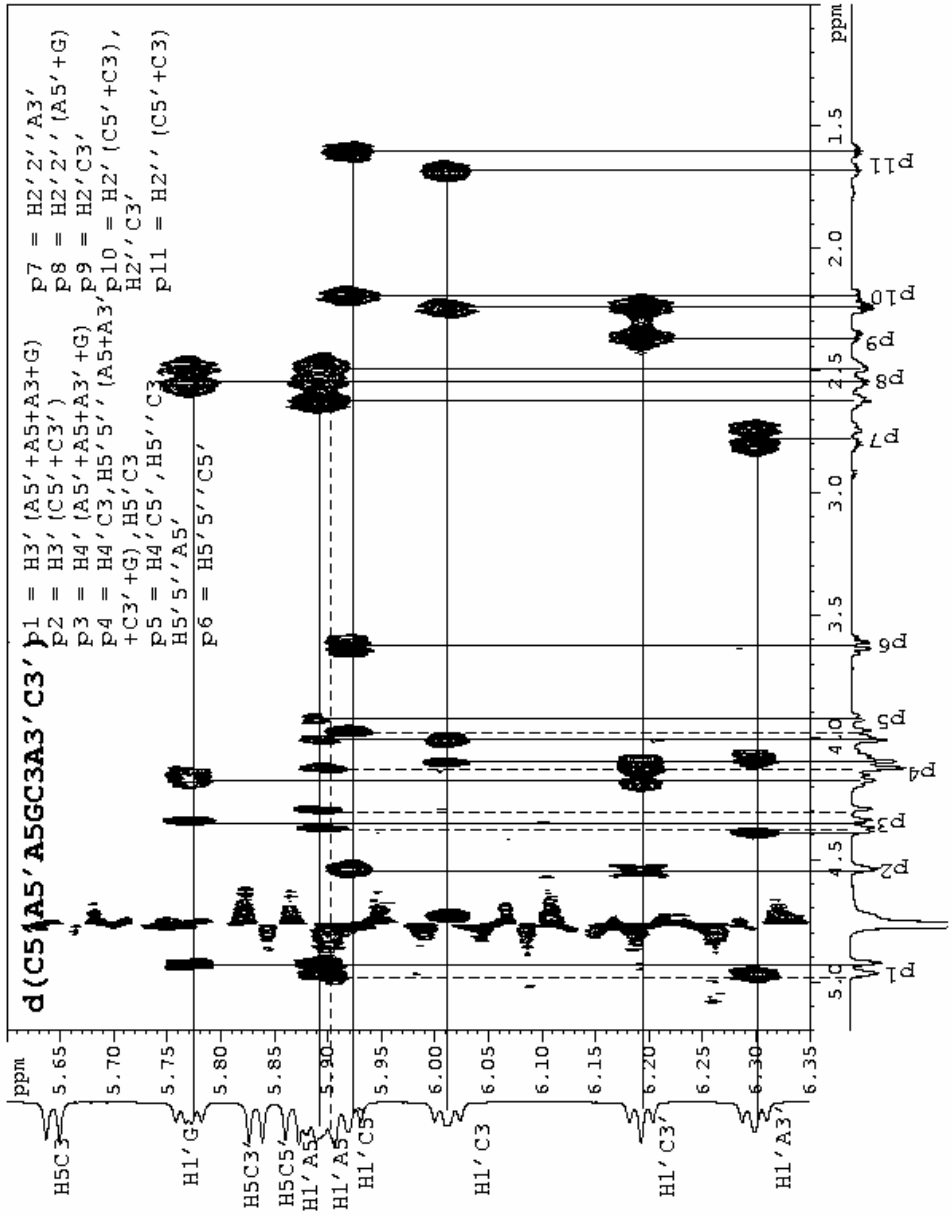


Figure S14. Expanded TOCSY spectra of the H2'/H2"/H3'/H4'/H5'/H5" region (1.0 – 5.1 ppm in F1 direction) to anomeric (H1') region (6.35 – 5.6 ppm in F2 direction) for d($^{5'}$C^{1}A^{2}A^{3}G^{4}C^{5}A^{6}C^{7}) (**2**) at 298 K.

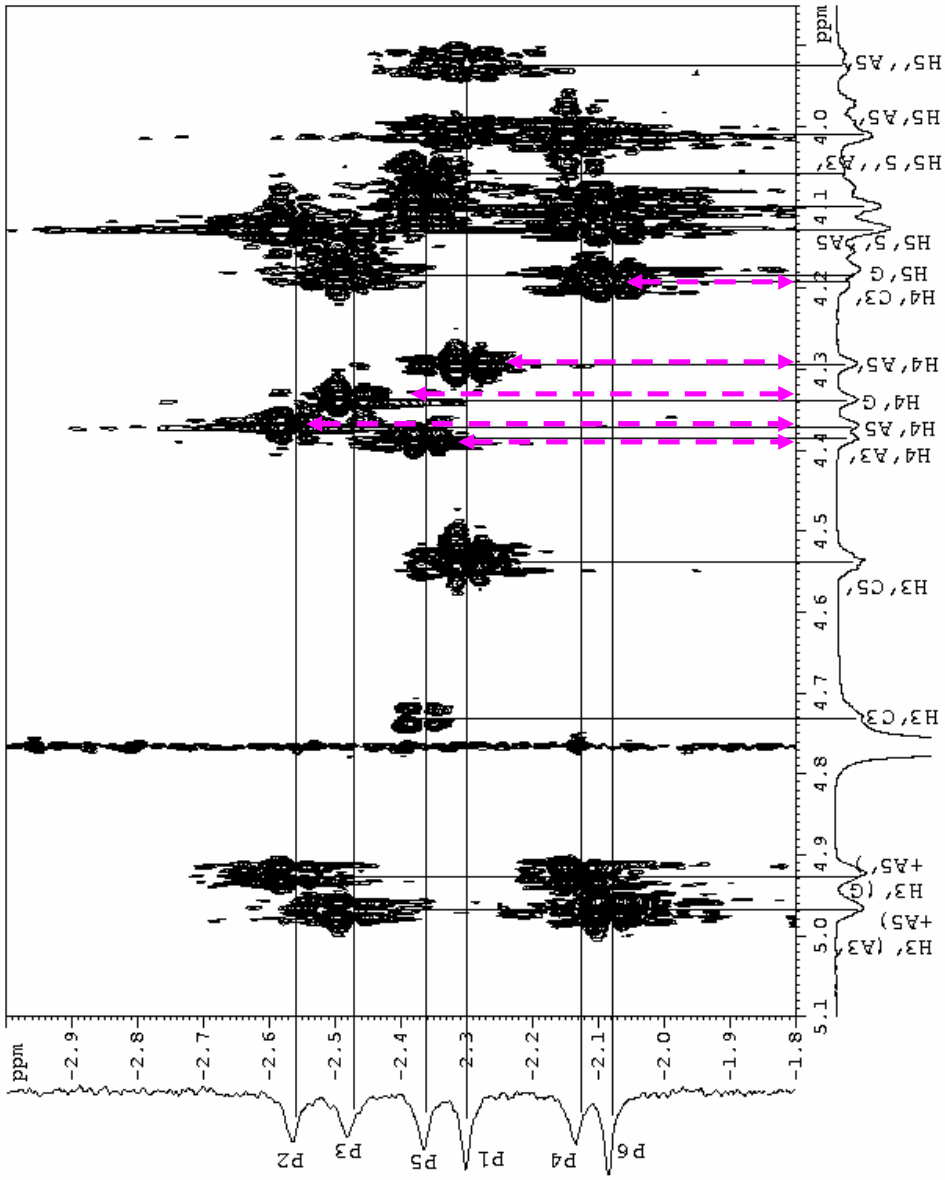


Figure S15. Expanded ^{31}P - ^{1}H correlation spectroscopy of ^{31}P region (-1.8 – -3.0 ppm in F2 direction) to H3'/H4'/H5'/H5" region (5.1 – 3.85 ppm in F1 direction) for d($^{5'}$C^{1}p$_{1}$A^{2}p$_{2}$A^{3}p$_{3}$G^{4}p$_{4}$C^{5}p$_{5}$A^{6}p$_{6}$C^{7}) (**2**) at 298 K. ← ▬ ▸ H4' – P connectivity.

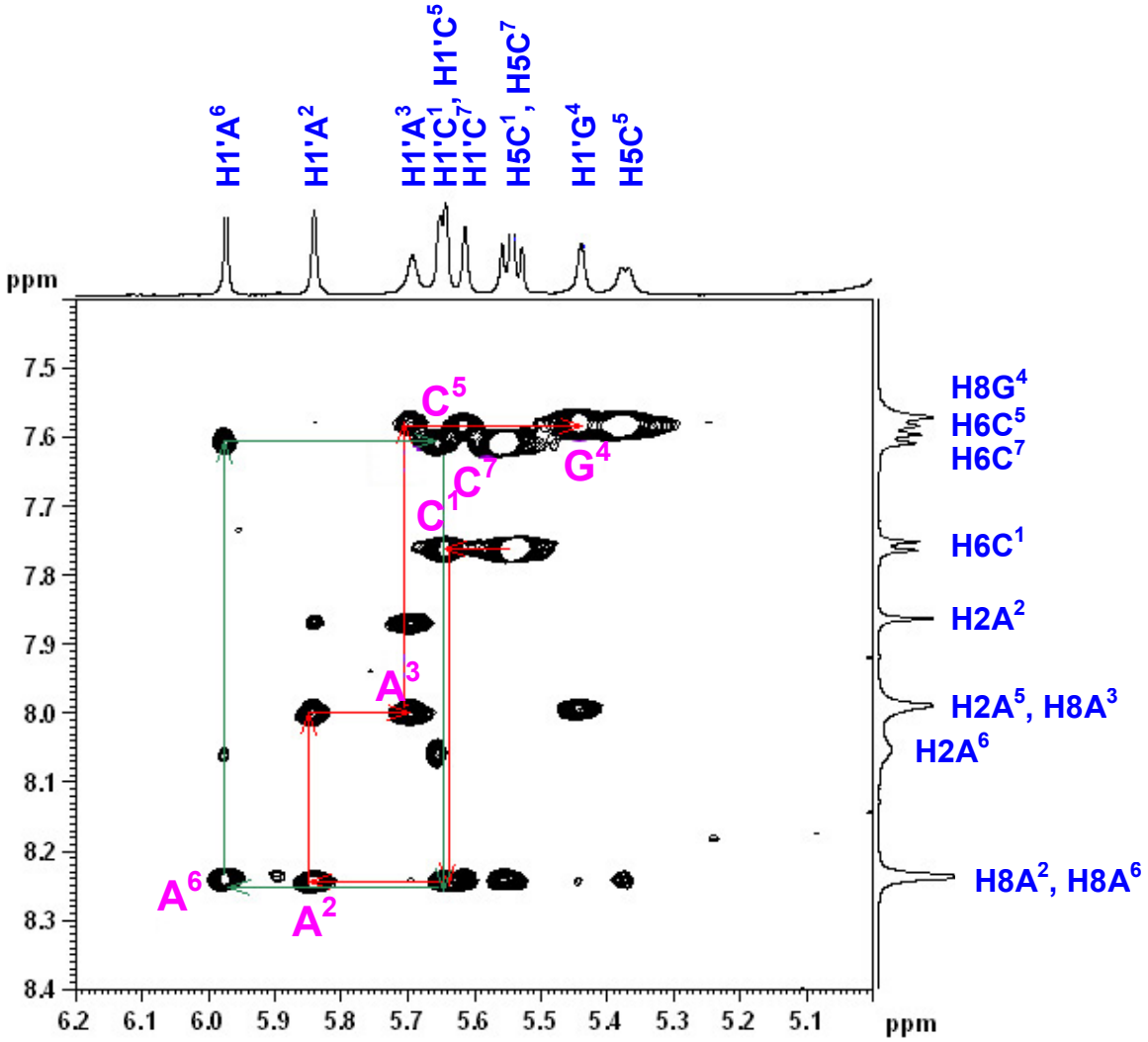


Figure S16. NOESY footprint of r($^{5'}$C^{1}A^{2}A^{3}G^{4}C^{5}A^{6}C^7) (**6**) showing the connectivity of nucleotide residues.

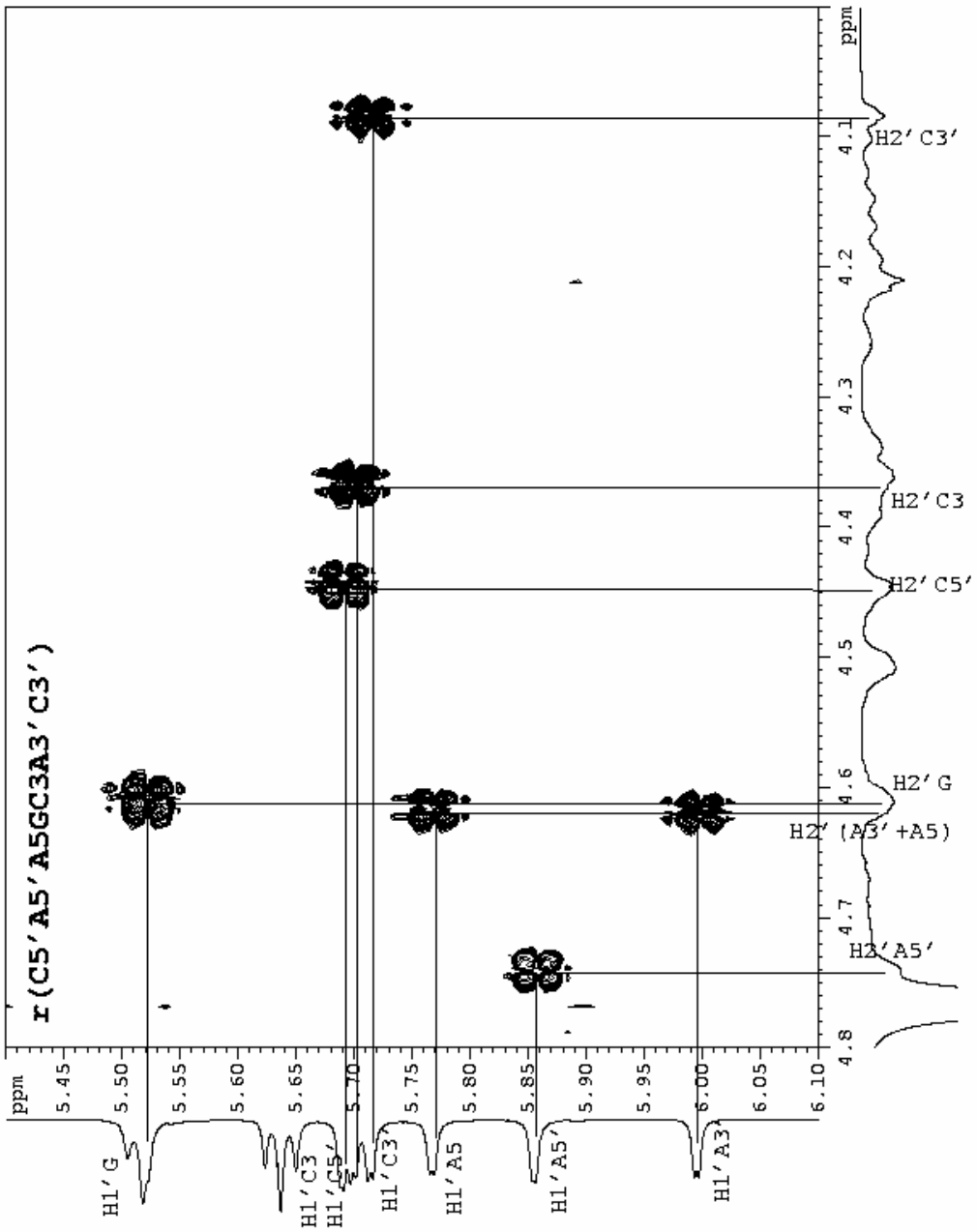


Figure S17. The expanded ^{31}P coupled DQF-COSY spectra of the anomeric H1' region to the H2' for r($^{5'}$C^{1}A^{2}A^{3}G^{4}C^{5}A^{6}C^{7}) (**6**) at 298 K. The spin connectivity between anomeric H1' region (5.4 – 6.1 ppm in F2) and H2' region (4.0 – 4.8 ppm in F1) of r($^{5'}$C^{1}A^{2}A^{3}G^{4}C^{5}A^{6}C^{7}) (**6**) have been shown.

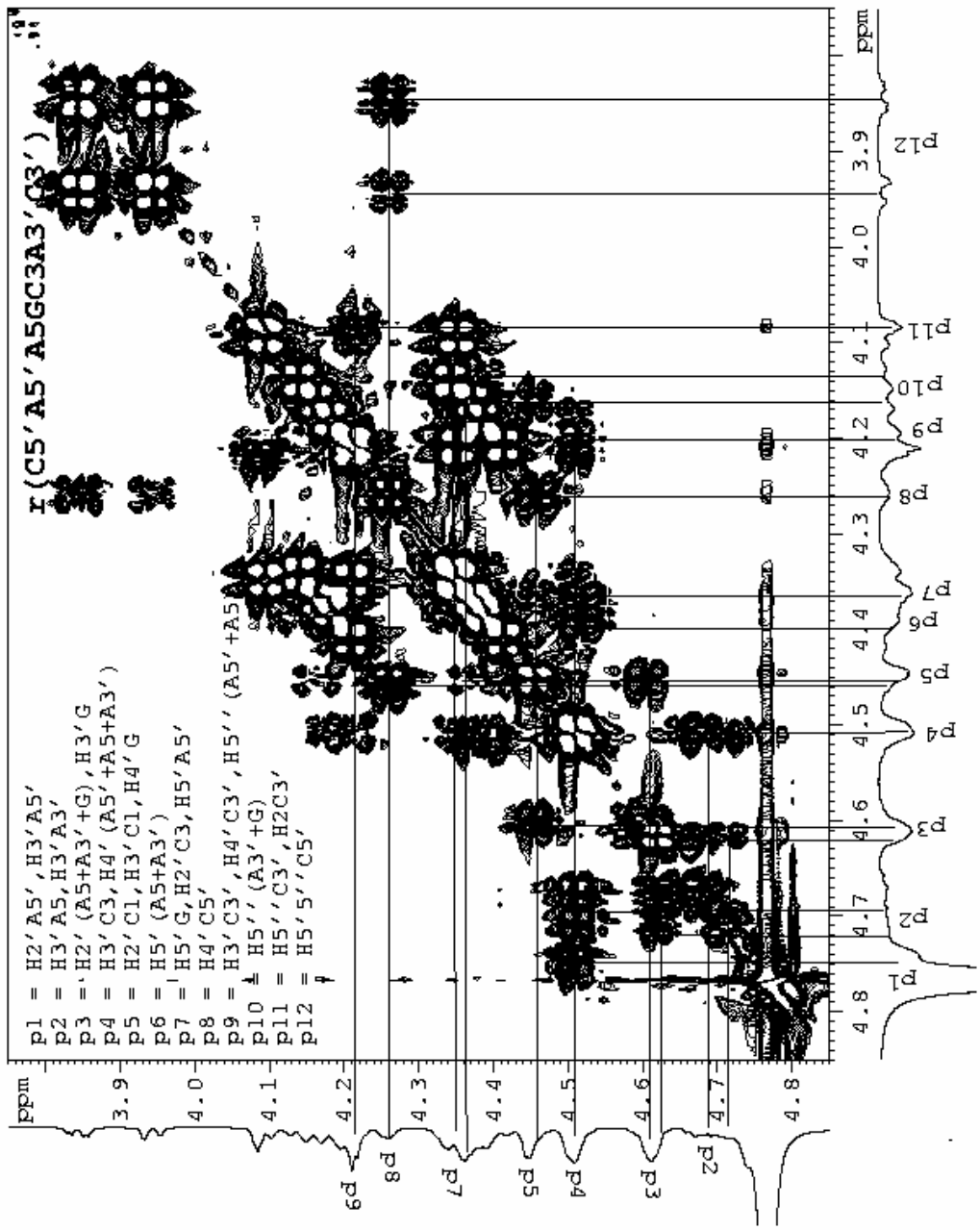


Figure S18. The expanded ^{31}P coupled DQF-COSY spectra of H2'/H3'/H4'/H5'/H5" region for r($^{5'}$C^{1}A^{2}A^{3}G^{4}C^{5}A^{6}C^{7}) (**6**) at 298 K. The region 4.85 – 3.75 ppm in both F1 and F2 dimensions of the r($^{5'}$C^{1}A^{2}A^{3}G^{4}C^{5}A^{6}C^{7}) (**6**) showing the spin connectivity.

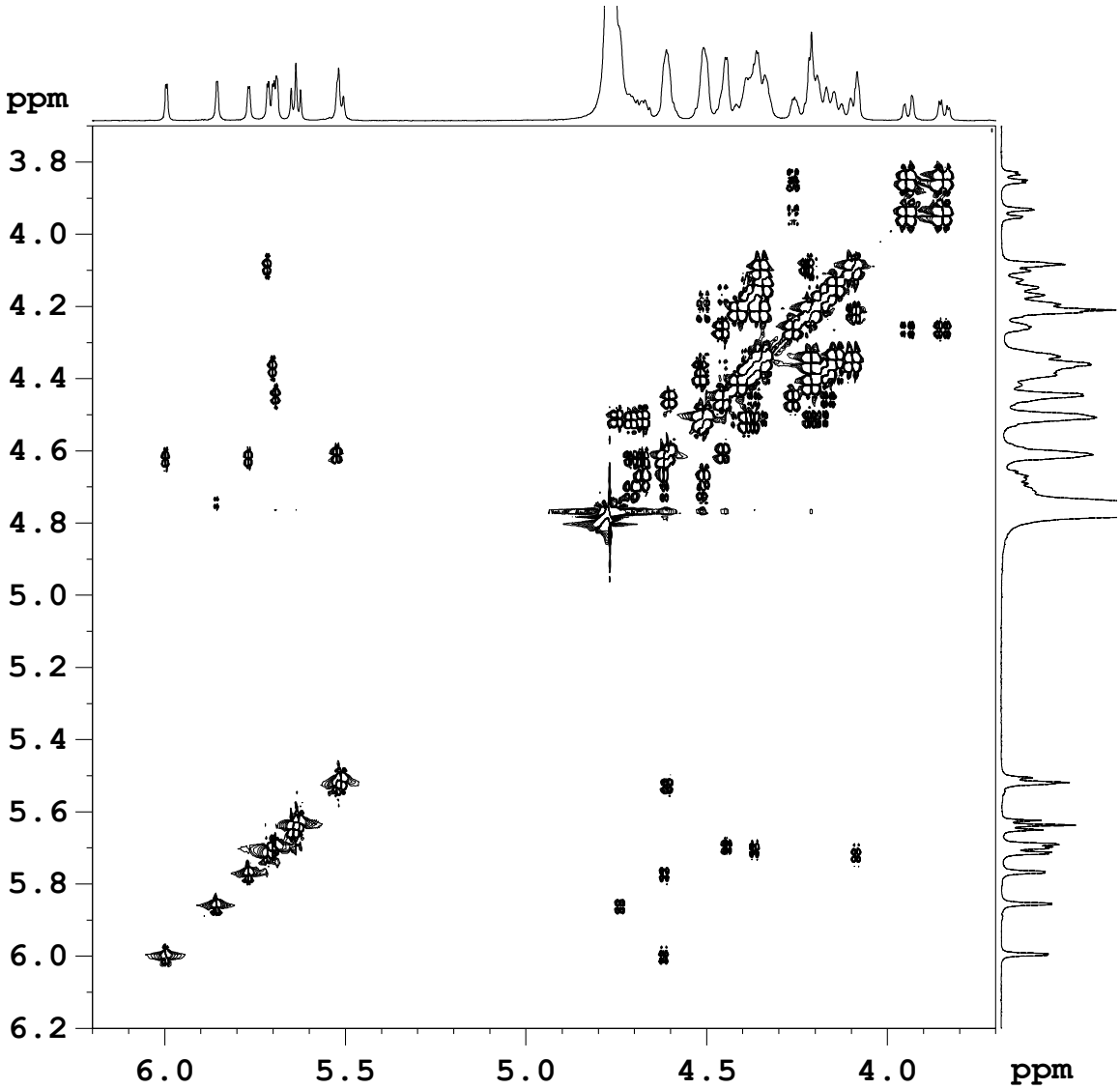


Figure S18.1. The ^{31}P decoupled DQF-COSY spectrum of r($^{5'}$C^{1}A^{2}A^{3}G^{4}C^{5}A^{6}C^{7}) (**6**) at 298 K. For assignments see **S17** and **S18**.

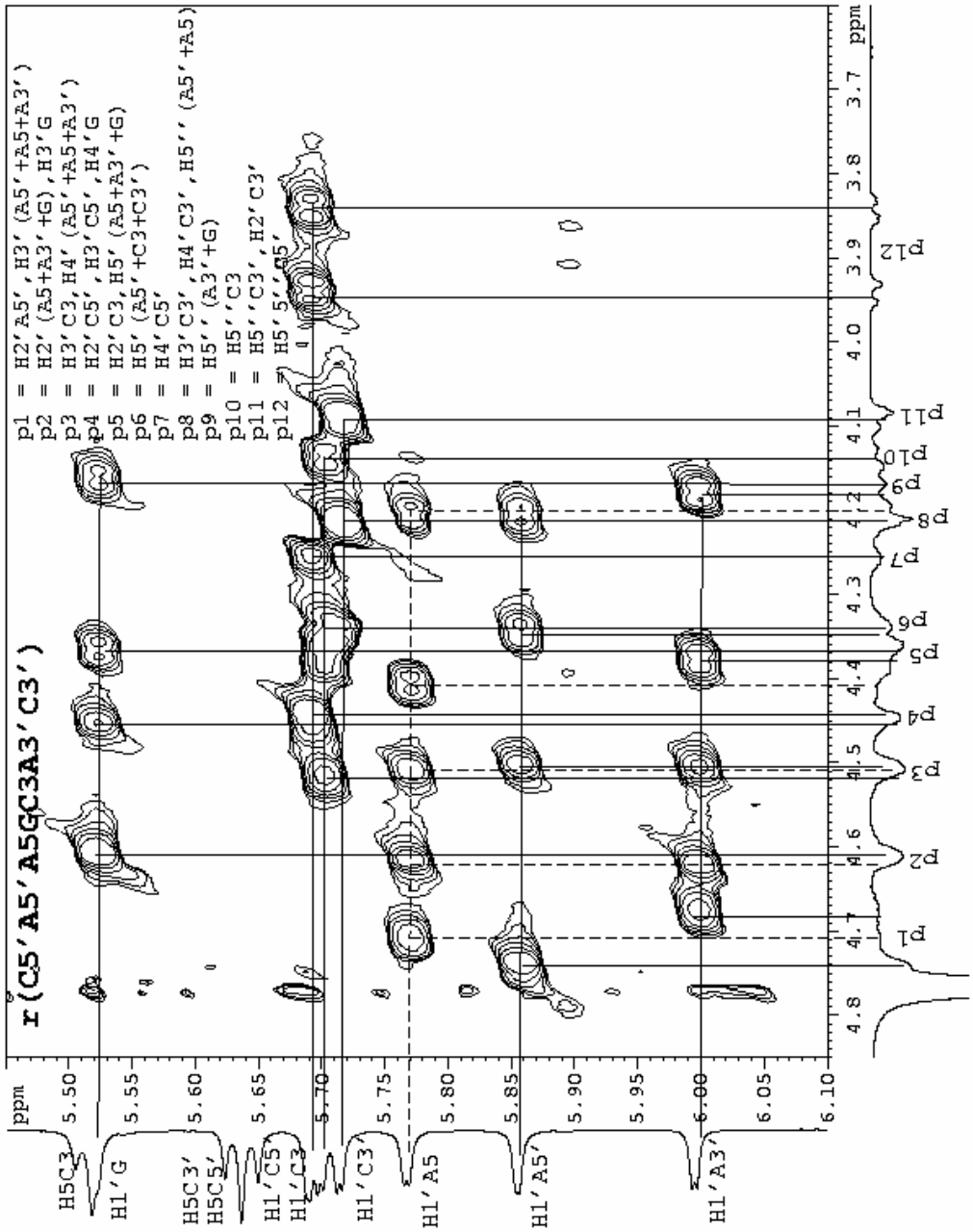


Figure S19. Expanded TOCSY spectra of the H2'/H3'/H4'/H5'/H5" region (4.85 – 3.6 ppm in F1 direction) to anomeric (H1') region (5.45 – 6.1 ppm in F2 direction) for r($^{5'}$C^{1}A^{2}A^{3}G^{4}C^{5}A^{6}C^{7}) (**6**) at 298 K.

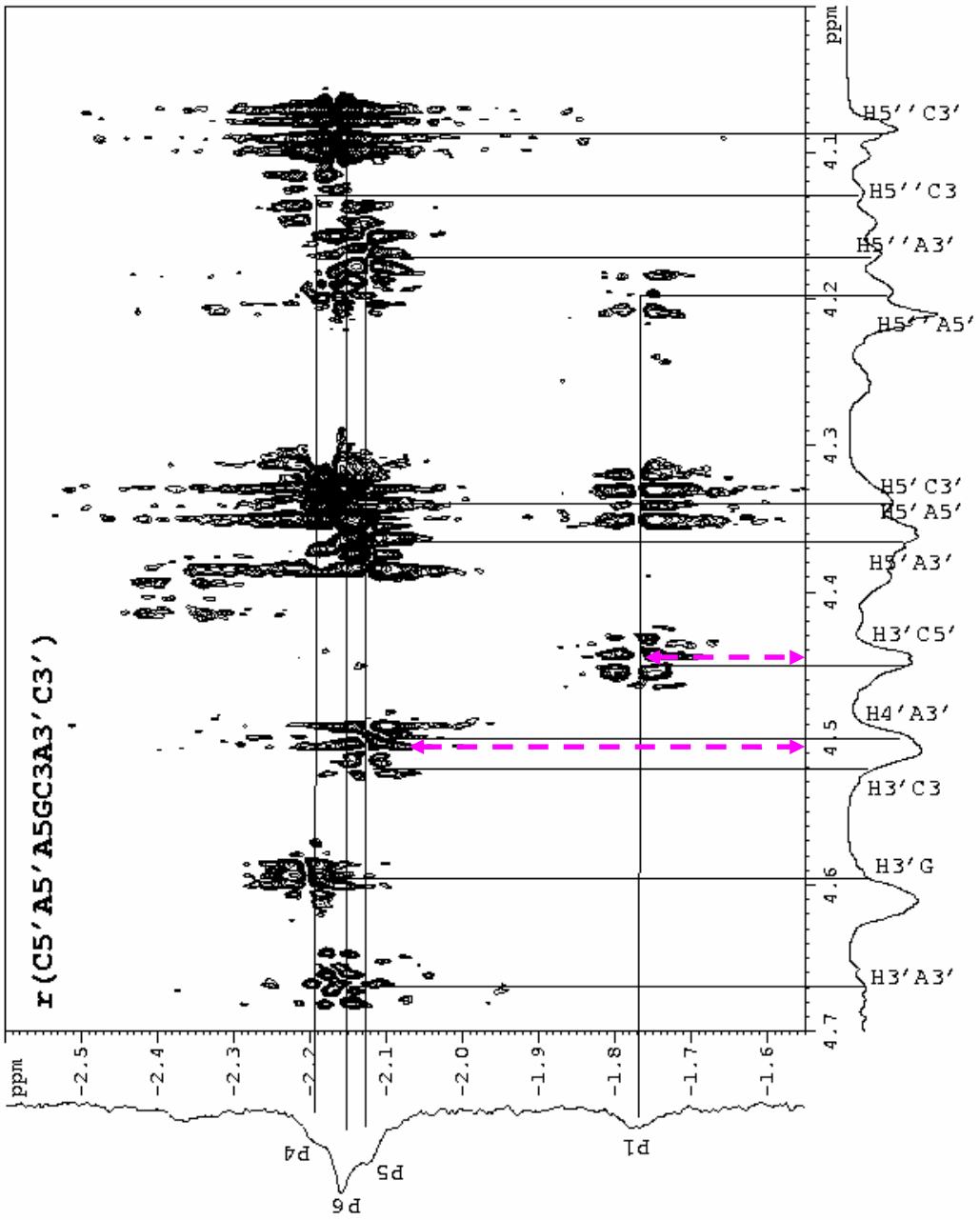


Figure S20. Expanded ^{31}P - ^{1}H correlation spectroscopy of ^{31}P region (-1.55 – -2.6 ppm in F2 direction) to H2'/H3'/H4'/H5'/H5" region (4.7 – 4.0 ppm in F1 direction) for r($^{5'}$C^{1}p$_{1}$A^{2}p$_{2}$A^{3}p$_{3}$G^{4}p$_{4}$C^{5}p$_{5}$A^{6}p$_{6}$C^{7}) (**6**) at 298 K. ← – – ► H4' – P connectivity.

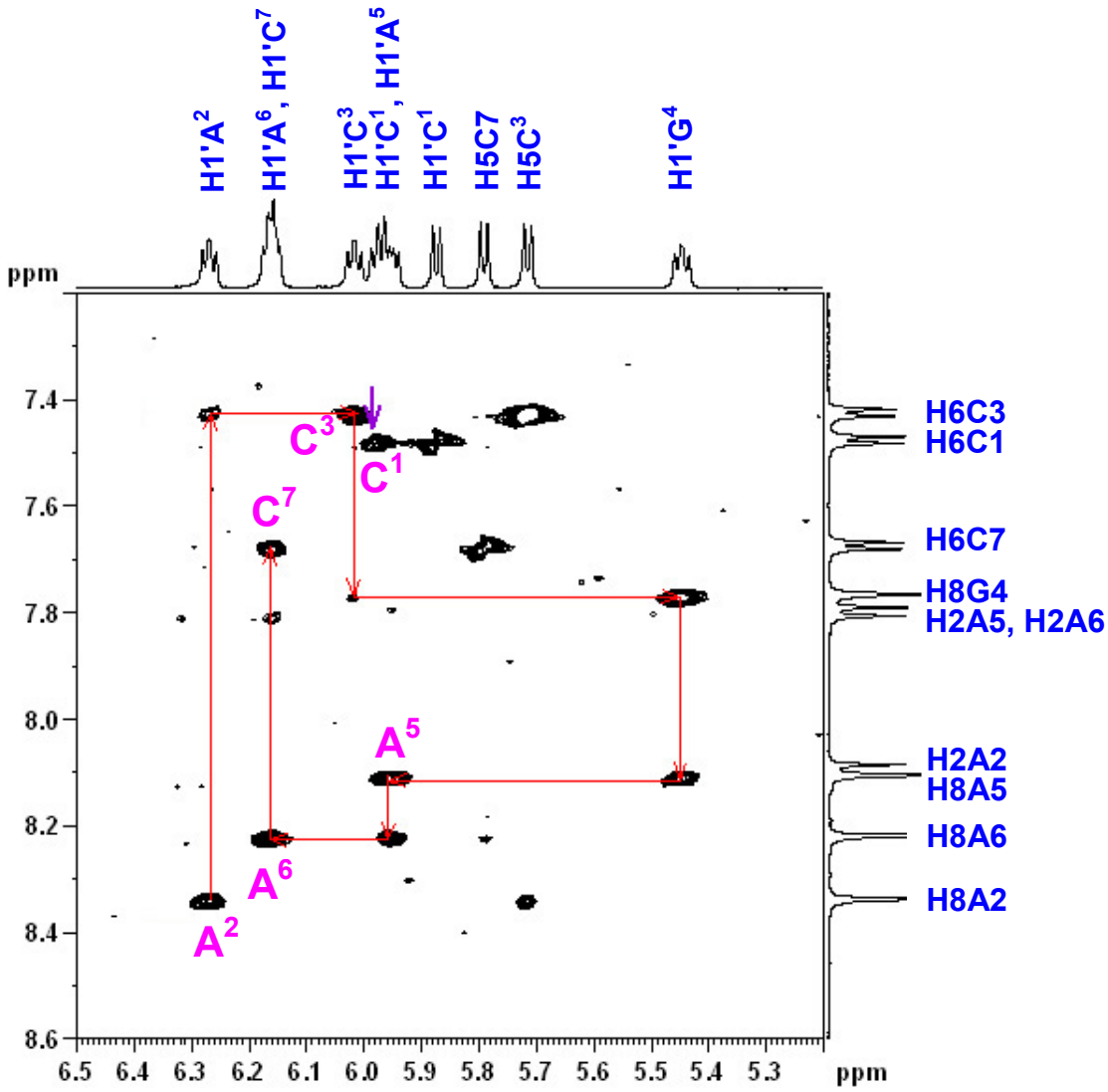


Figure S21. NOESY footprint of d($^{5'}C^1A^2C^3G^4A^5A^6C^7$) (**3**) showing the connectivity of nucleotide residues.

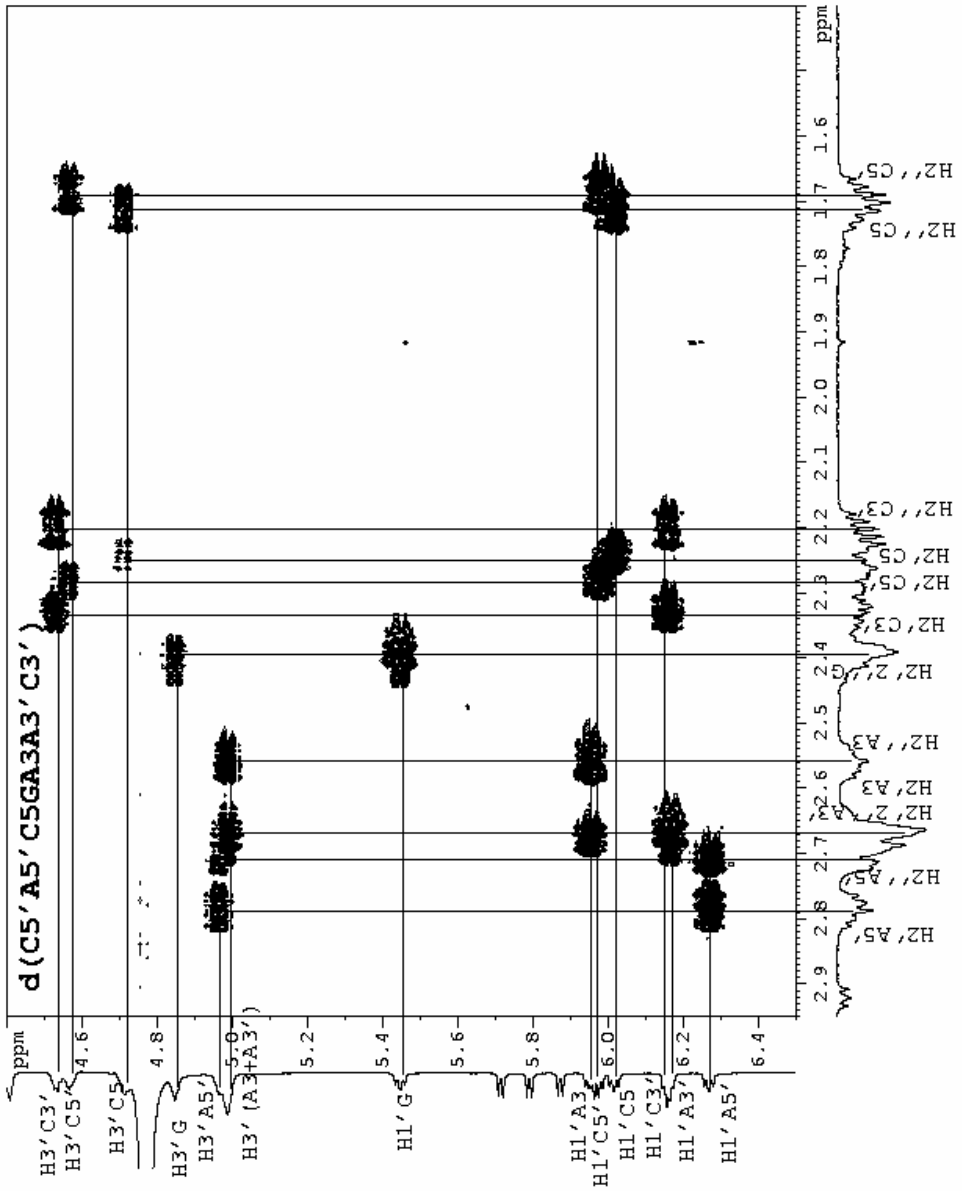


Figure S22. The expanded ^{31}P coupled DQF-COSY spectra of H1'/H2'/H2''/H3' region for d($^{5'}$C^{1}A^{2}C^{3}G^{4}A^{5}A^{6}C^{7}) (**3**) at 298 K. The region 2.95 – 1.4 ppm in F1 and region 6.5 – 4.4 ppm in F2 dimension of the d($^{5'}$C^{1}A^{2}C^{3}G^{4}A^{5}A^{6}C^{7}) (**3**) showing the spin connectivity.

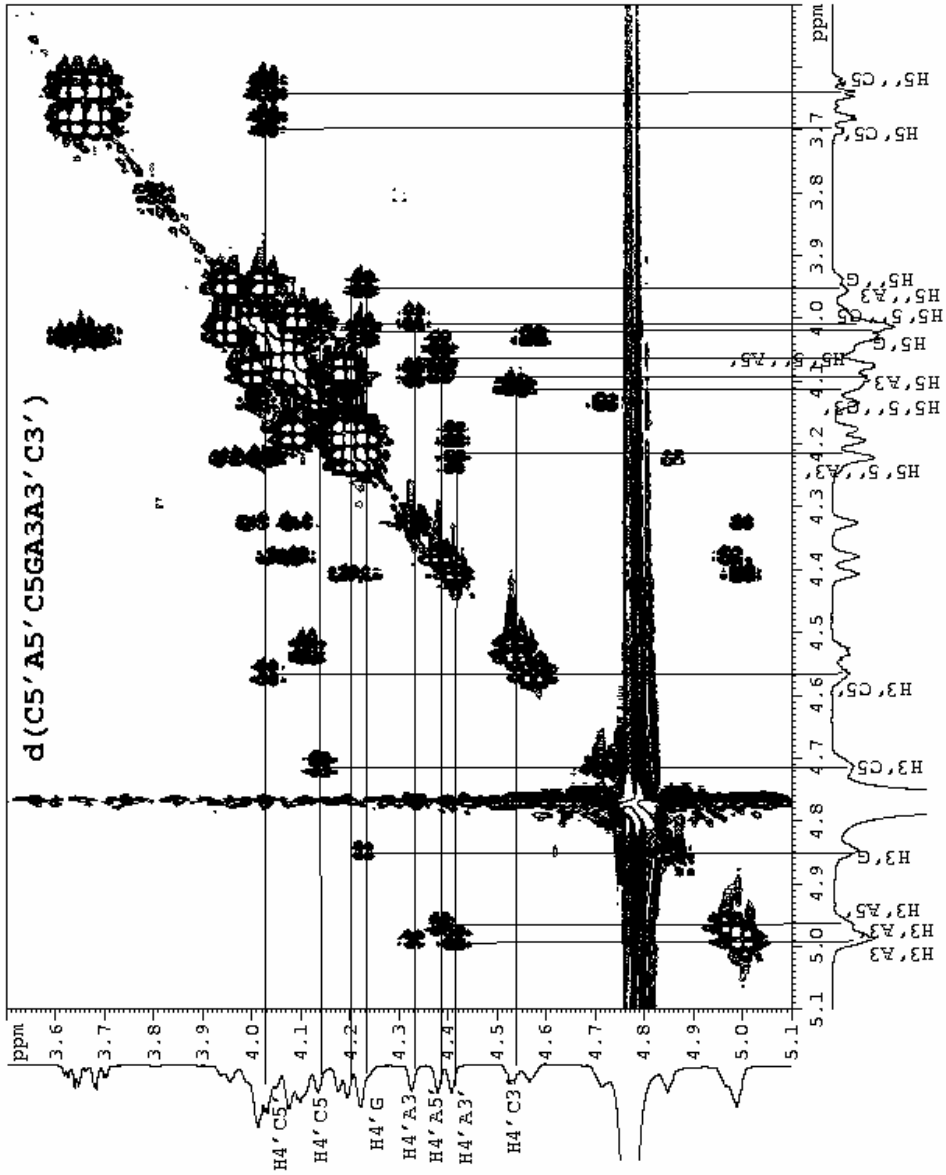


Figure S23. The expanded ^{31}P coupled DQF-COSY spectra of H3'/H4'/H5'/H5" region for d($^{5'}$C^{1}A^{2}C^{3}G^{4}A^{5}A^{6}C^{7}) (**3**) at 298 K. The region 5.1 – 3.5 ppm in both F1 and F2 dimensions of the d($^{5'}$C^{1}A^{2}C^{3}G^{4}A^{5}A^{6}C^{7}) (**3**) showing the spin connectivity.

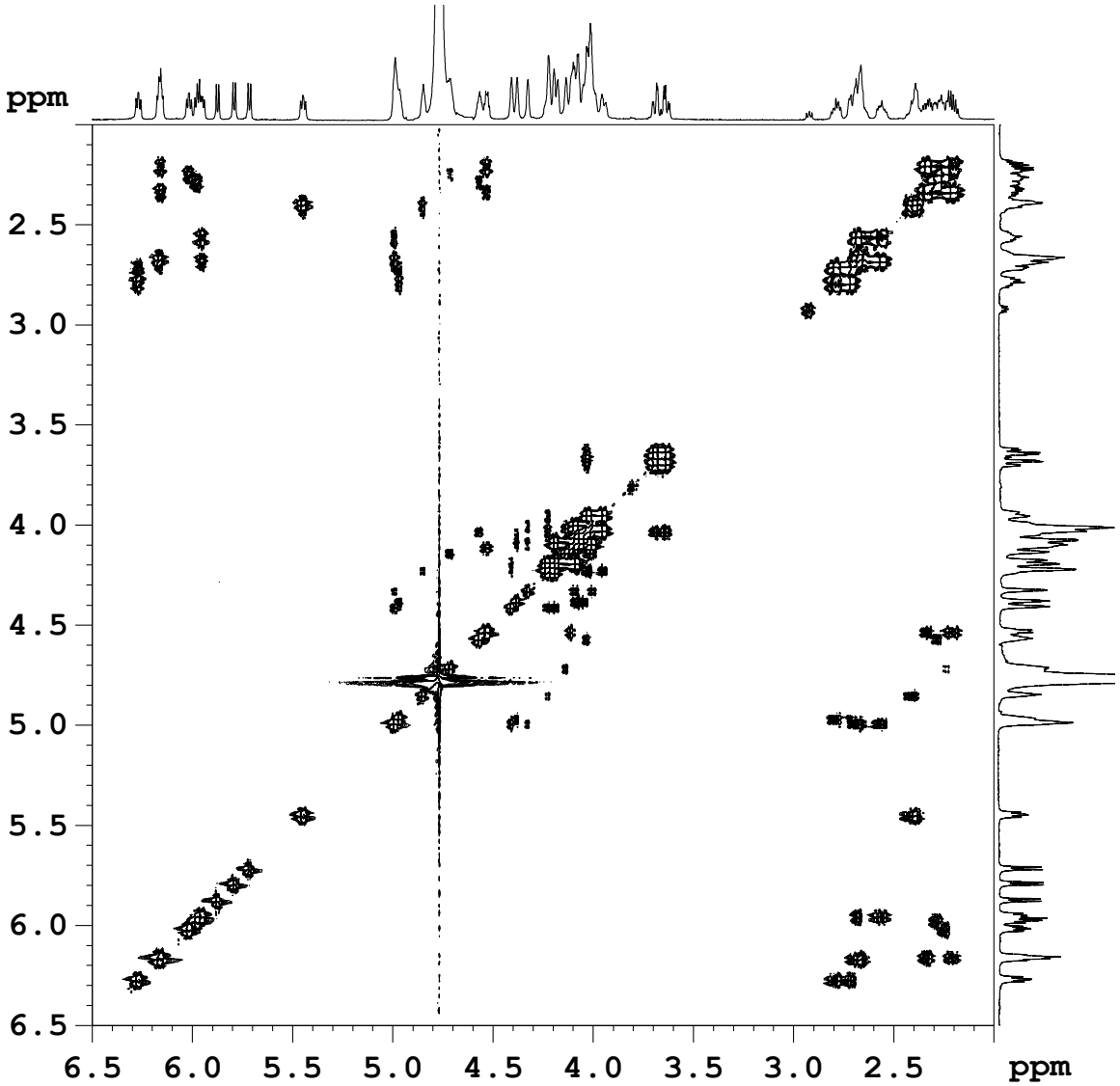


Figure S23.1. The ^{31}P decoupled DQF-COSY spectrum of d($^{5'}$C^{1}A^{2}C^{3}G^{4}A^{5}A^{6}C^{7}) (**3**) at 298 K. For assignments see **S22** and **S23**.

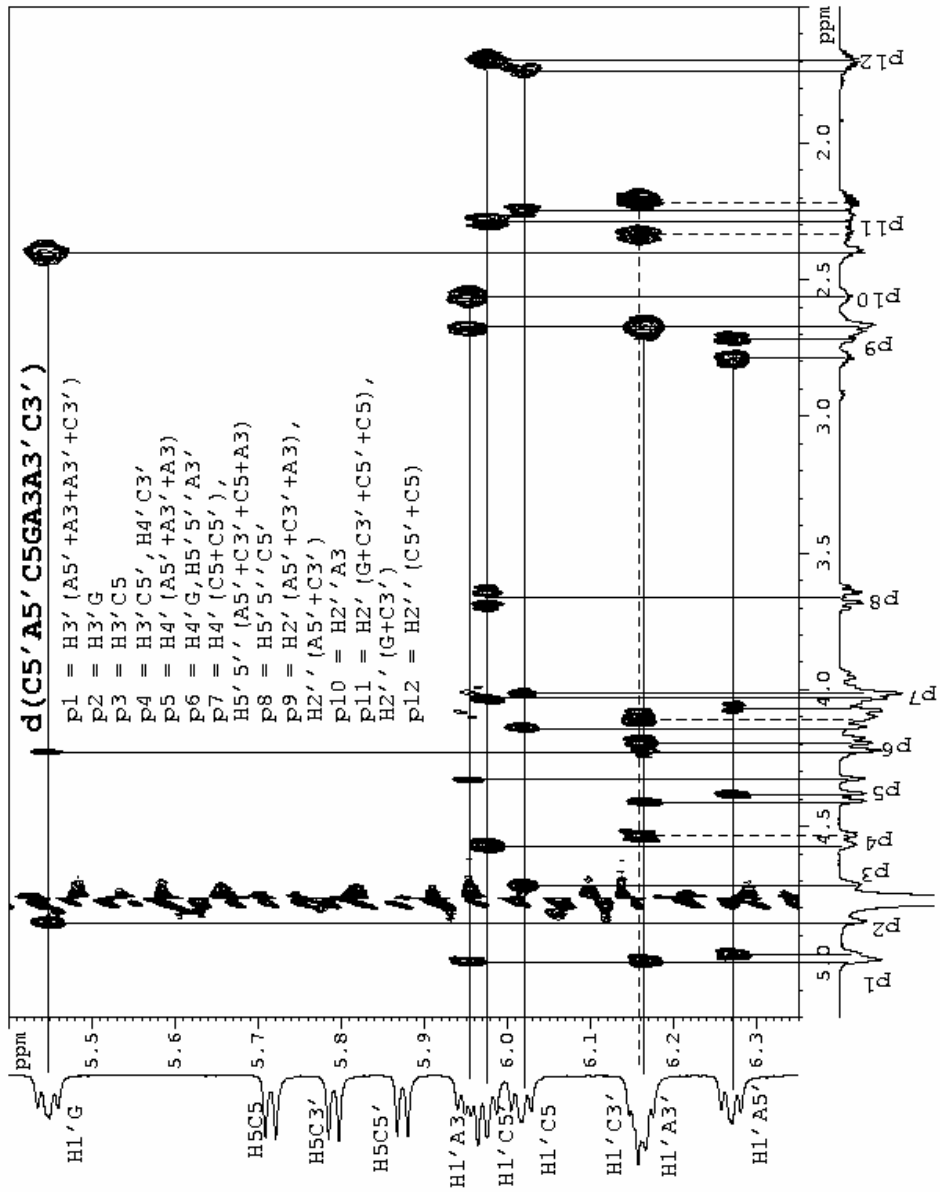


Figure S24. Expanded TOCSY spectra of the H2'/H2"/H3'/H4'/H5'/H5" region (1.5 – 5.15 ppm in F1 direction) to anomeric (H1') region (6.35 – 5.4 ppm in F2 direction) for d($^{5'}$C^{1}A^{2}C^{3}G^{4}A^{5}A^{6}C^{7}) (**3**) at 298 K.

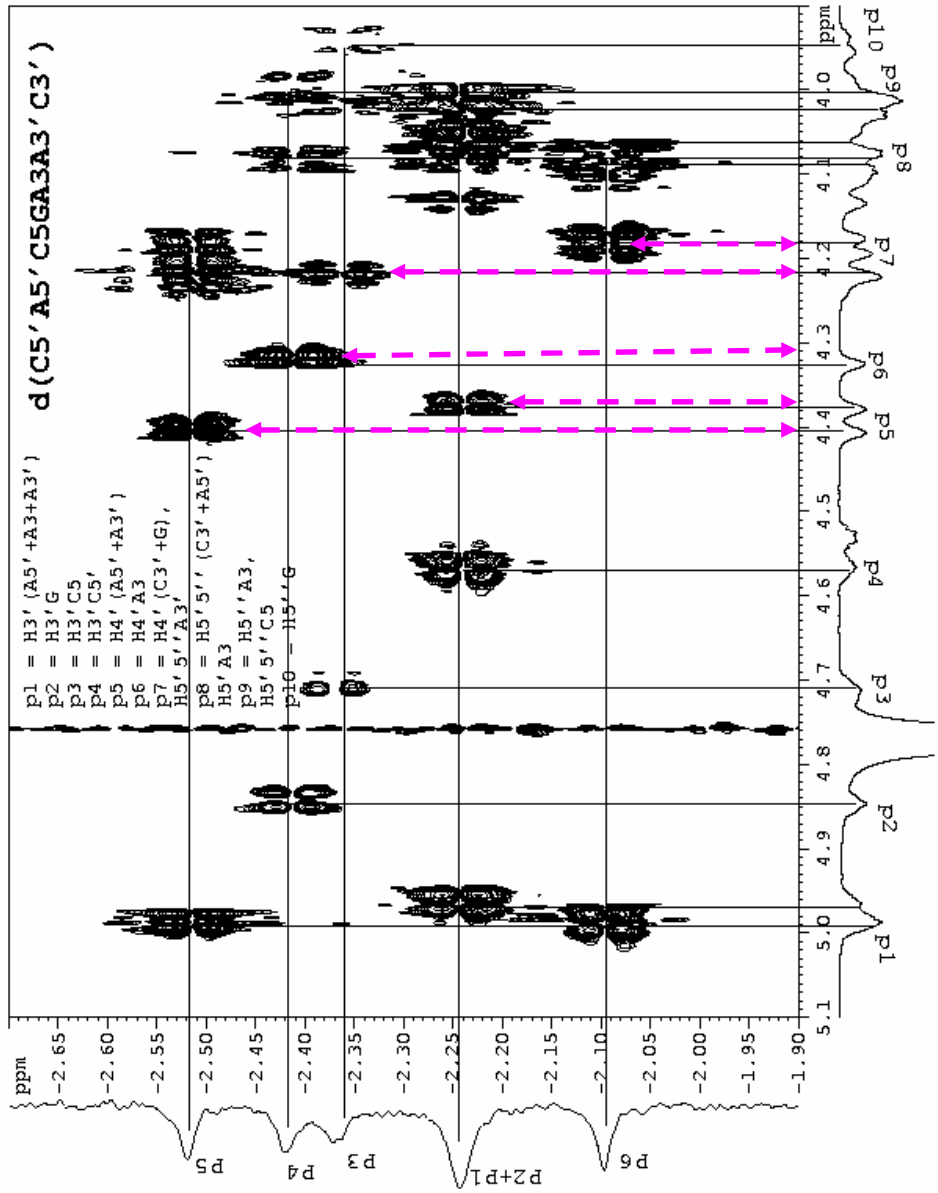


Figure S25. Expanded ^{31}P - ^{1}H correlation spectroscopy of ^{31}P region (-1.9 – -2.7 ppm in F2 direction) to H3'/H4'/H5'/H5" region (5.1 – 3.9 ppm in F1 direction) for d($^{5'}C^{1}p_{1}A^{2}p_{2}C^{3}p_{3}G^{4}p_{4}A^{5}p_{5}A^{6}p_{6}C^{7}$) (**3**) at 298 K. ← - - → H4' – P connectivity.

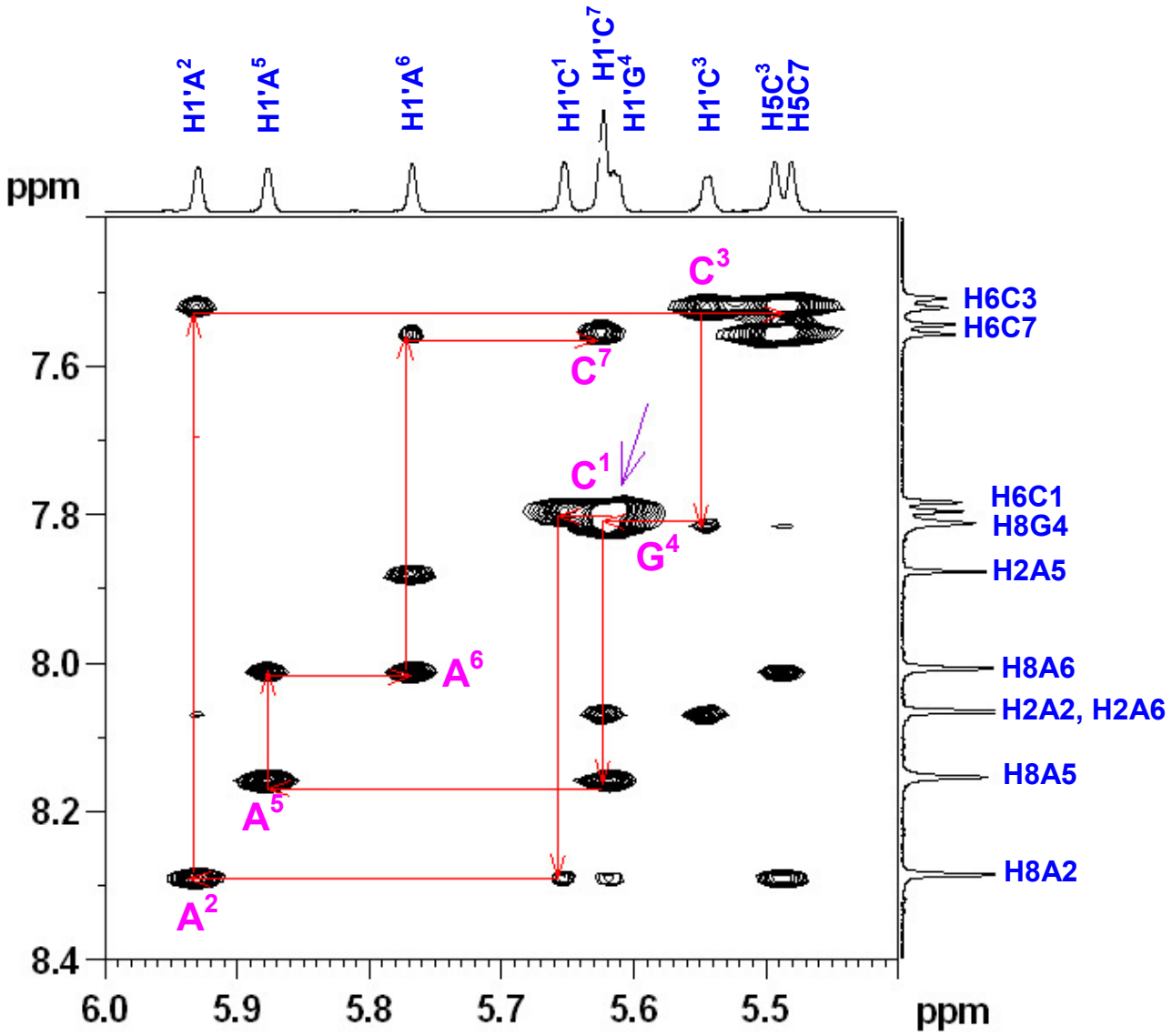


Figure S26. NOESY footprint of r($^{5'}$C^{1}A^{2}C^{3}G^{4}A^{5}A^{6}C^{7}) (**7**) showing the connectivity of nucleotide residues.

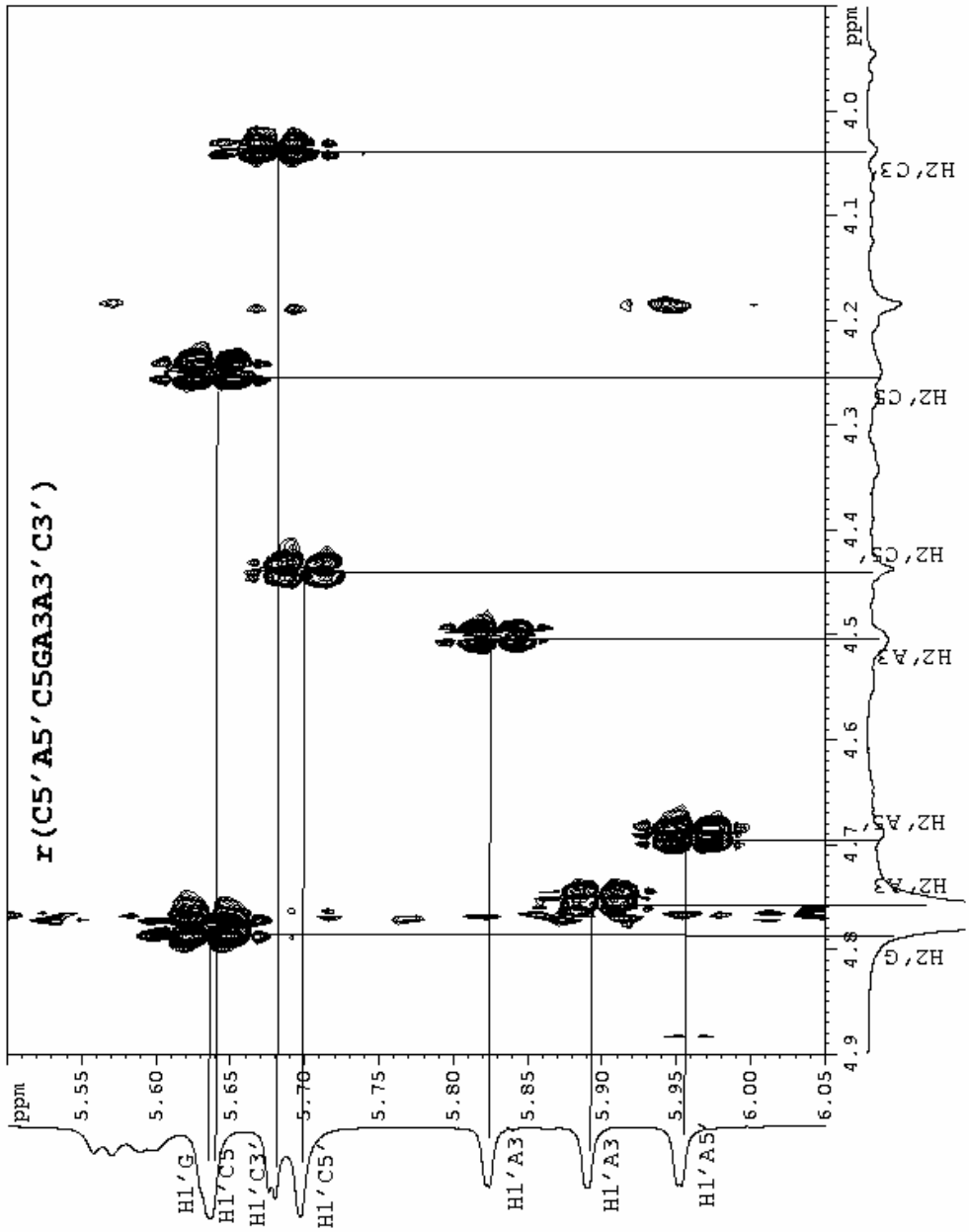


Figure S27. The expanded ^{31}P coupled DQF-COSY spectra of the anomeric H1' region to the H2' for r($^{5'}$C^{1}A^{2}C^{3}G^{4}A^{5}A^{6}C^{7}) (**7**) at 298 K. The spin connectivity between anomeric H1' region (5.5 – 6.05 ppm in F2) and H2' region (3.9 – 4.9 ppm in F1) of r($^{5'}$C^{1}A^{2}C^{3}G^{4}A^{5}A^{6}C^{7}) (**7**) have been shown.

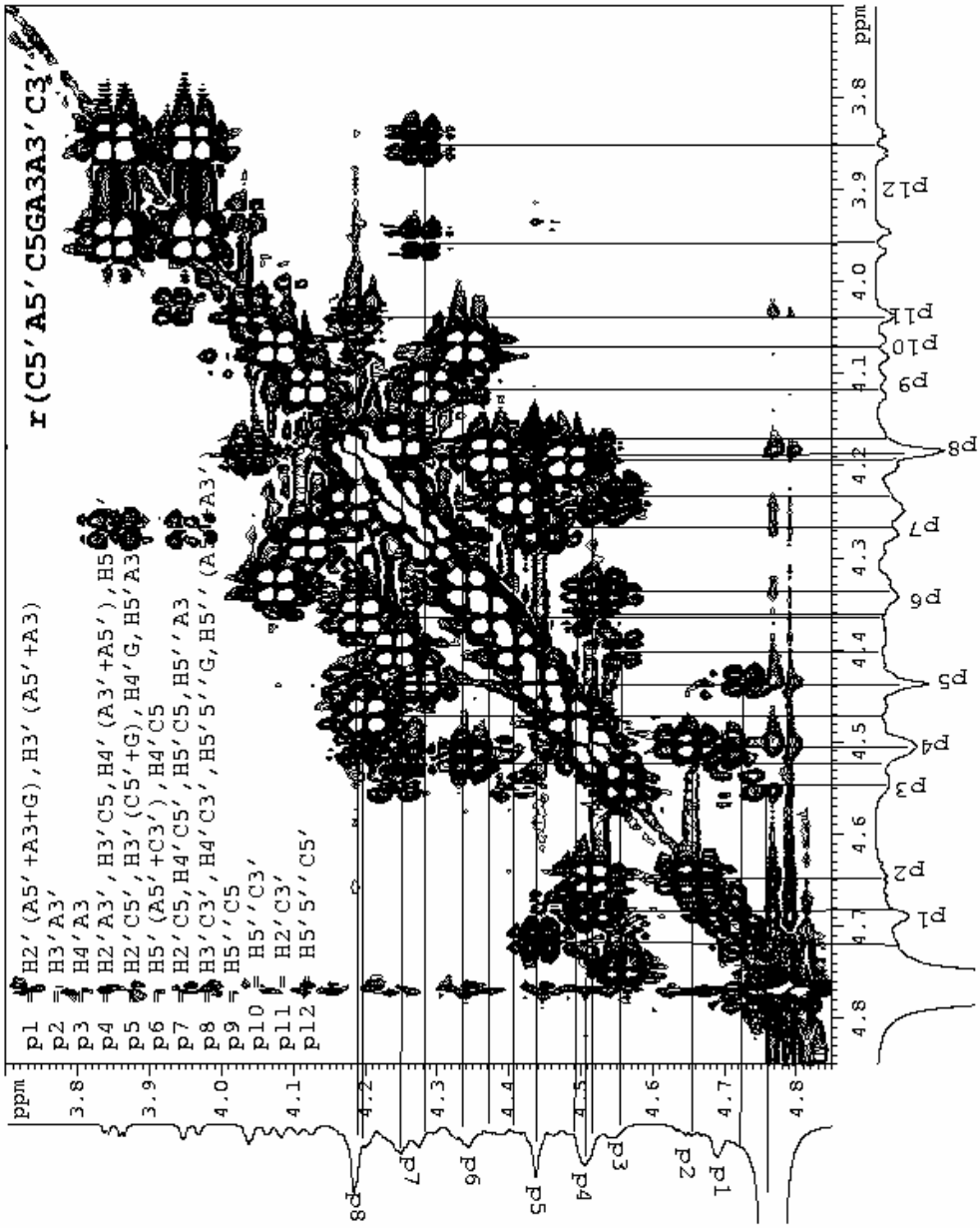


Figure S28. The expanded ^{31}P coupled DQF-COSY spectra of H2'/H3'/H4'/H5'/H5" region for r($^{5'}$C^{1}A^{2}C^{3}G^{4}A^{5}A^{6}C^{7}) (**7**) at 298 K. The region 4.85 – 3.7 ppm in both F1 and F2 dimensions of the r($^{5'}$C^{1}A^{2}C^{3}G^{4}A^{5}A^{6}C^{7}) (**7**) showing the spin connectivity.

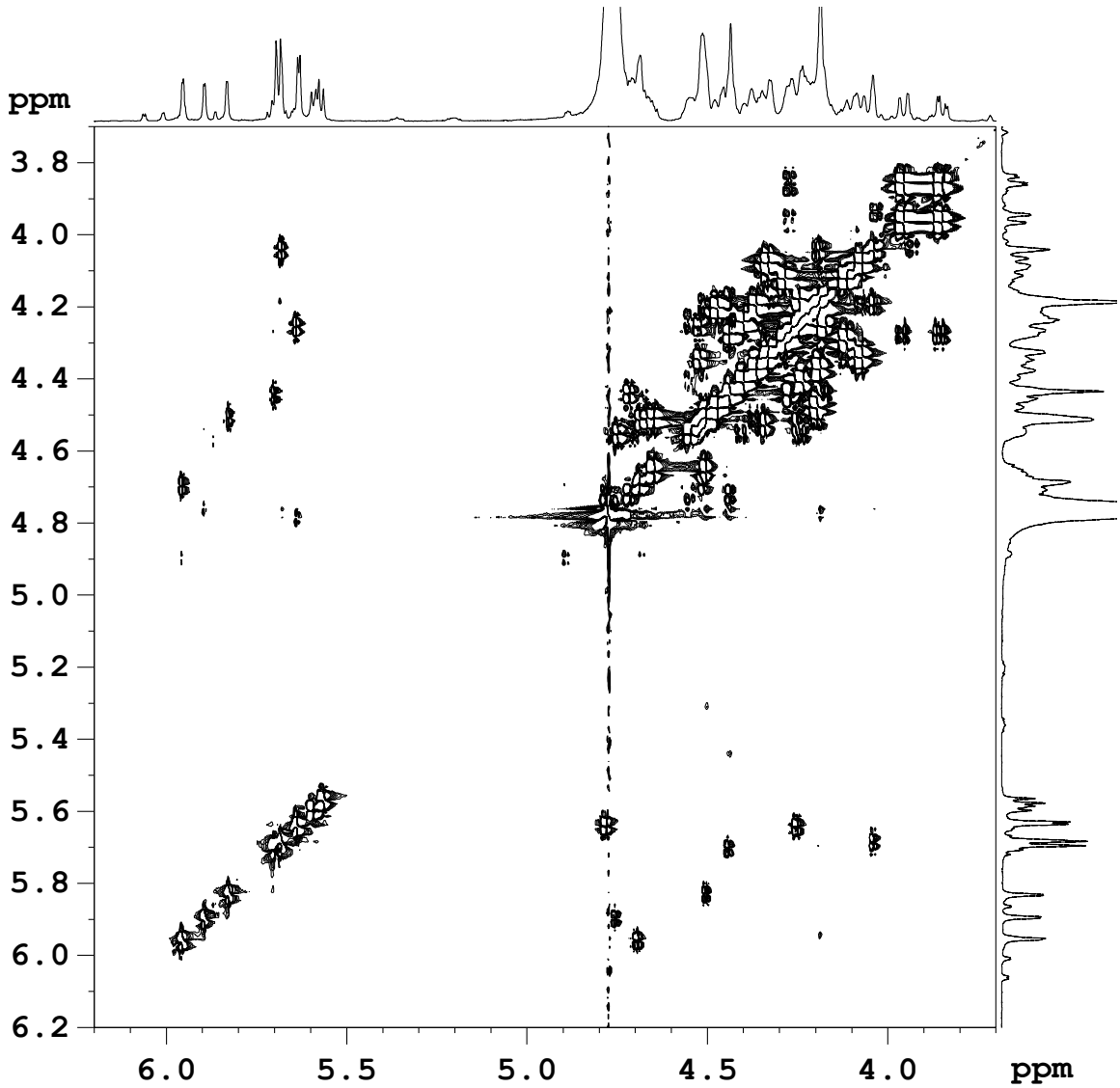


Figure S28.1. The ^{31}P decoupled DQF-COSY spectrum of r($^{5'}$C^{1}A^{2}C^{3}G^{4}A^{5}A^{6}C^{7}) (**7**) at 298 K. For assignments see **S27** and **S28**.

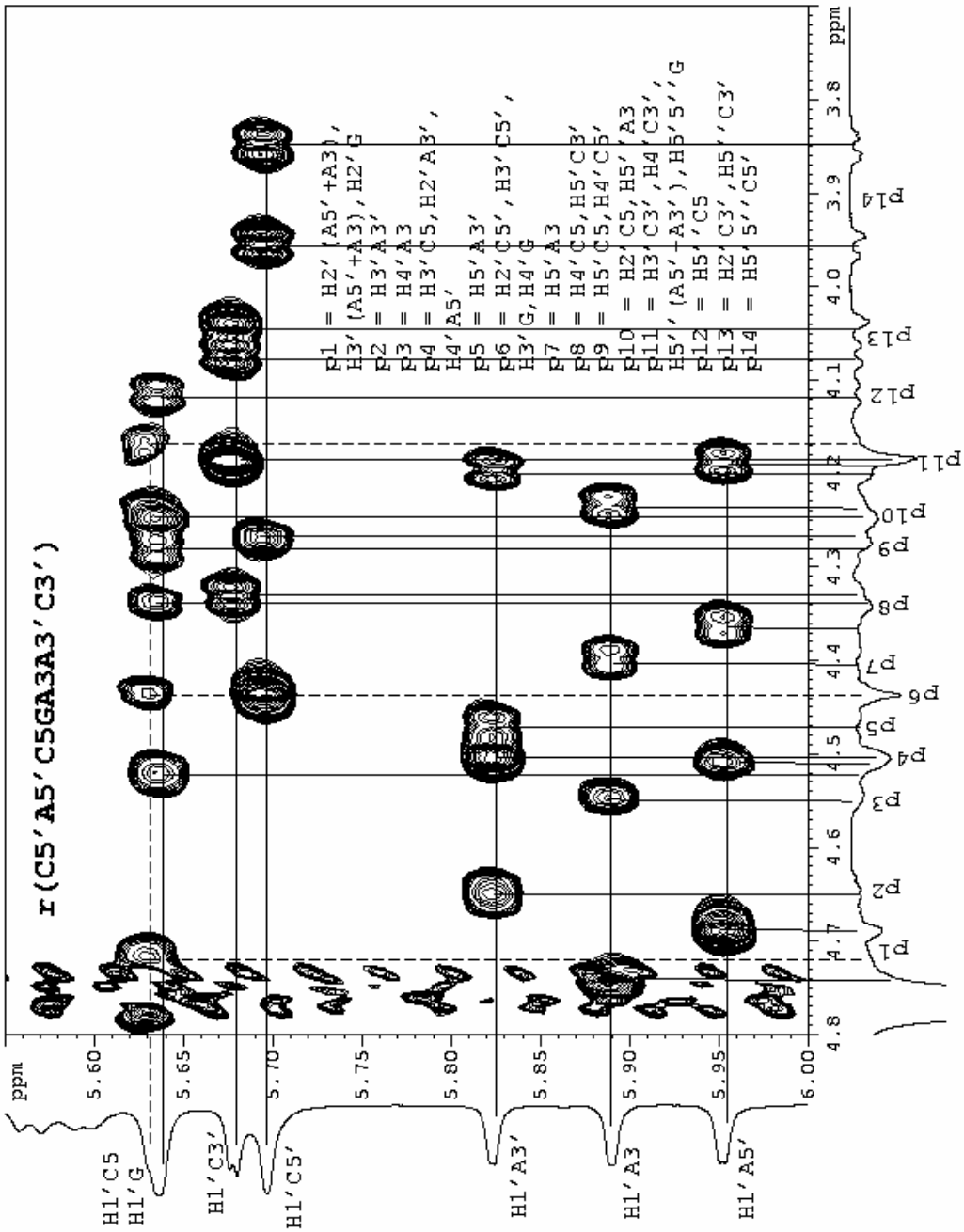


Figure S29. Expanded TOCSY spectra of the H2'/H3'/H4'/H5'/H5" region (4.8 – 3.7 ppm in F1 direction) to anomeric (H1') region (5.55 – 6.0 ppm in F2 direction) for r($^{5'}$C^{1}A^{2}C^{3}G^{4}A^{5}A^{6}C^{7}) (**7**) at 298 K.

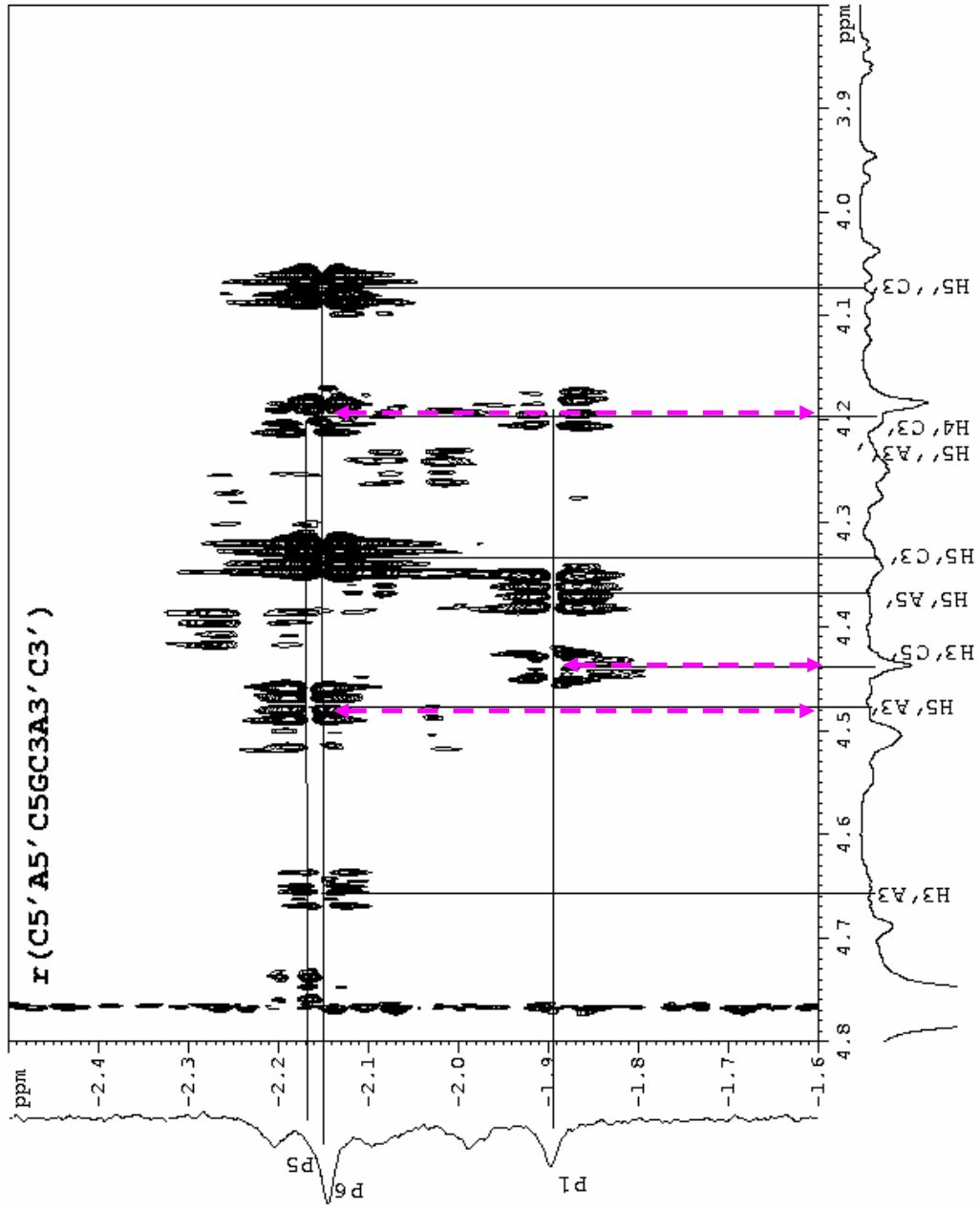


Figure S30. Expanded ^{31}P - ^{1}H correlation spectroscopy of ^{31}P region (-1.6 – -2.5 ppm in F2 direction) to H2'/H3'/H4'/H5'/H5" region (4.8 – 3.8 ppm in F1 direction) for r($^{5'}C^{1}p_{1}A^{2}p_{2}C^{3}p_{3}G^{4}p_{4}A^{5}p_{5}A^{6}p_{6}C^{7}$) (**7**) at 298 K. ← – ► H4' – P connectivity.

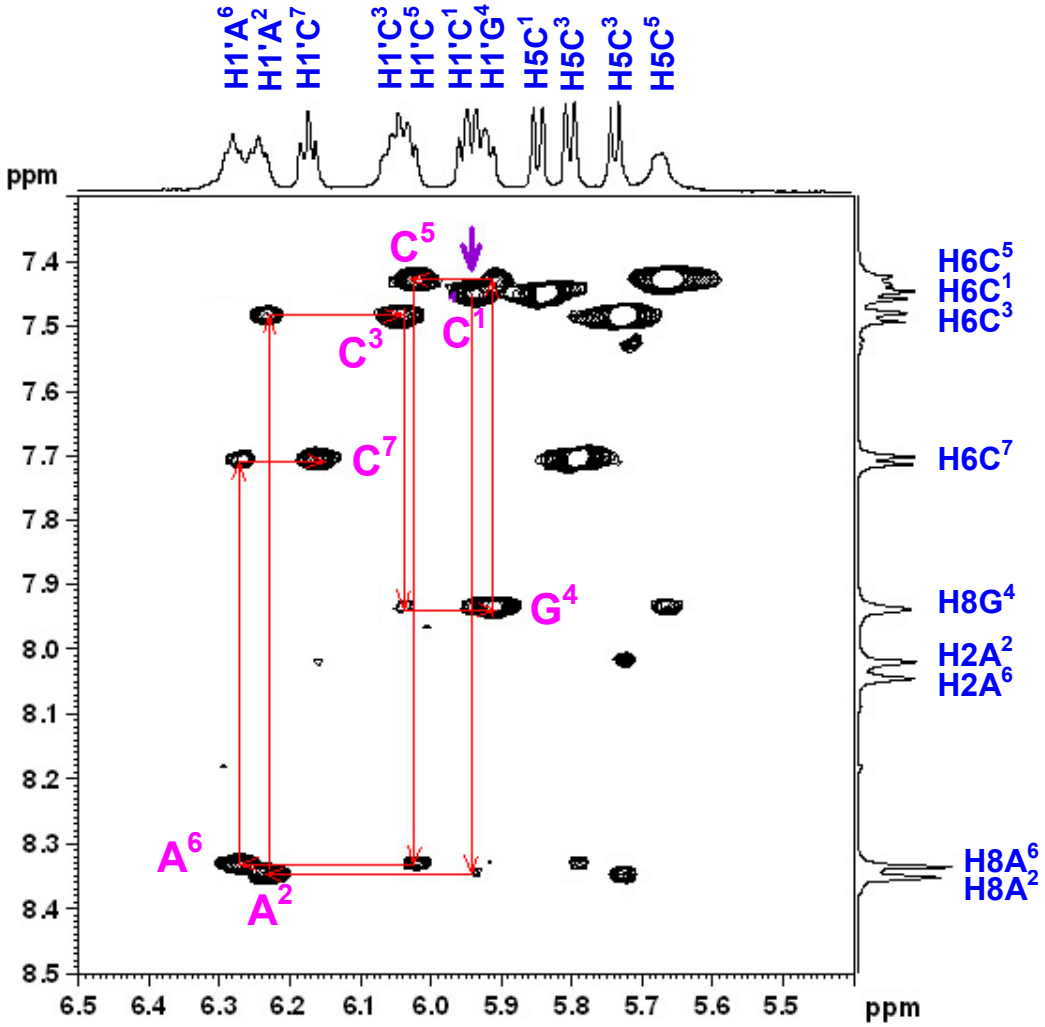


Figure S31. NOESY footprint of d($^{5'}C^1A^2C^3G^4C^5A^6C^7$) (**4**) showing the connectivity of nucleotide residues.

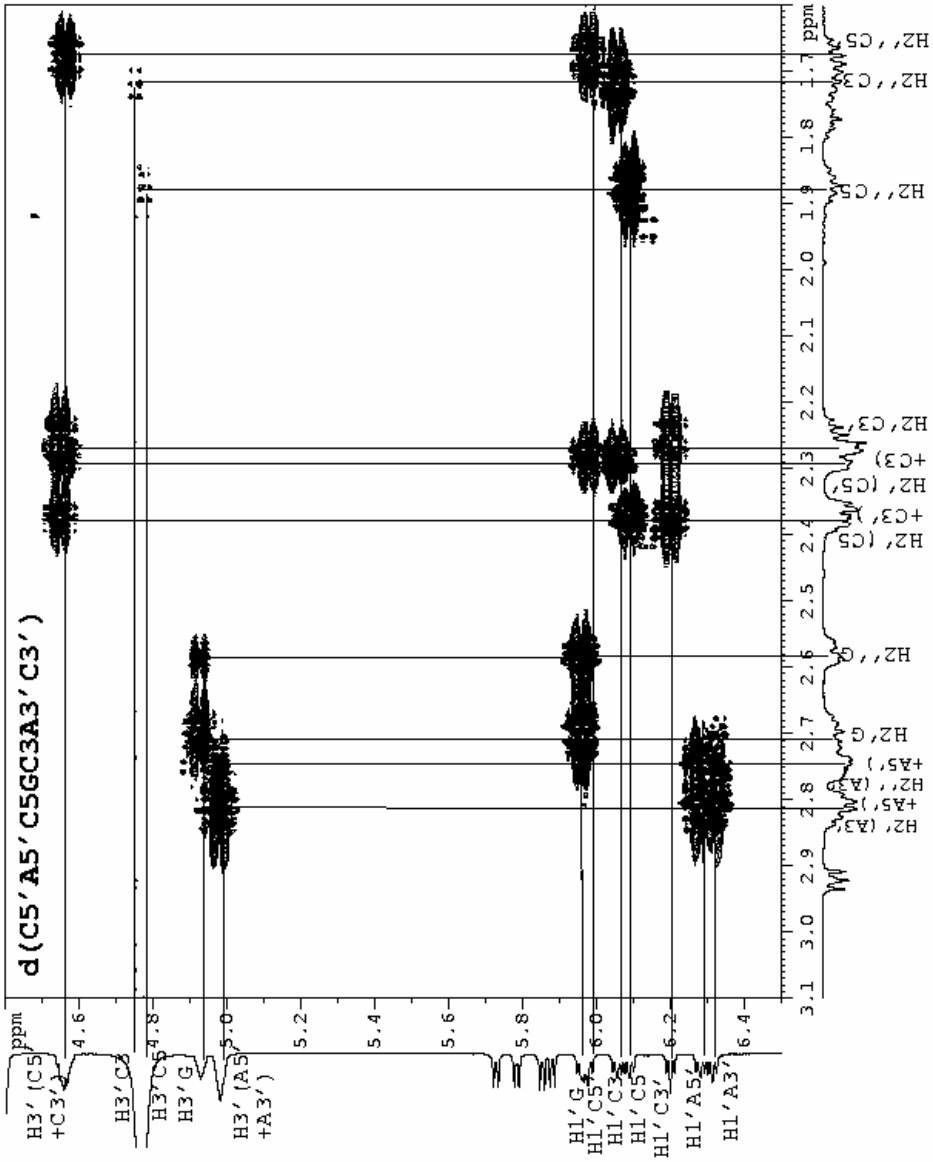


Figure S32. The expanded ^{31}P coupled DQF-COSY spectra of H1'/H2'/H2"/H3' region for d($^{5'}$C^{1}A^{2}C^{3}G^{4}C^{5}A^{6}C^{7}) (**4**) at 298 K. The region 3.1 – 1.6 ppm in F1 and region 6.5 – 4.4 ppm in F2 dimension of the d($^{5'}$C^{1}A^{2}C^{3}G^{4}C^{5}A^{6}C^{7}) (**4**) showing the spin connectivity.

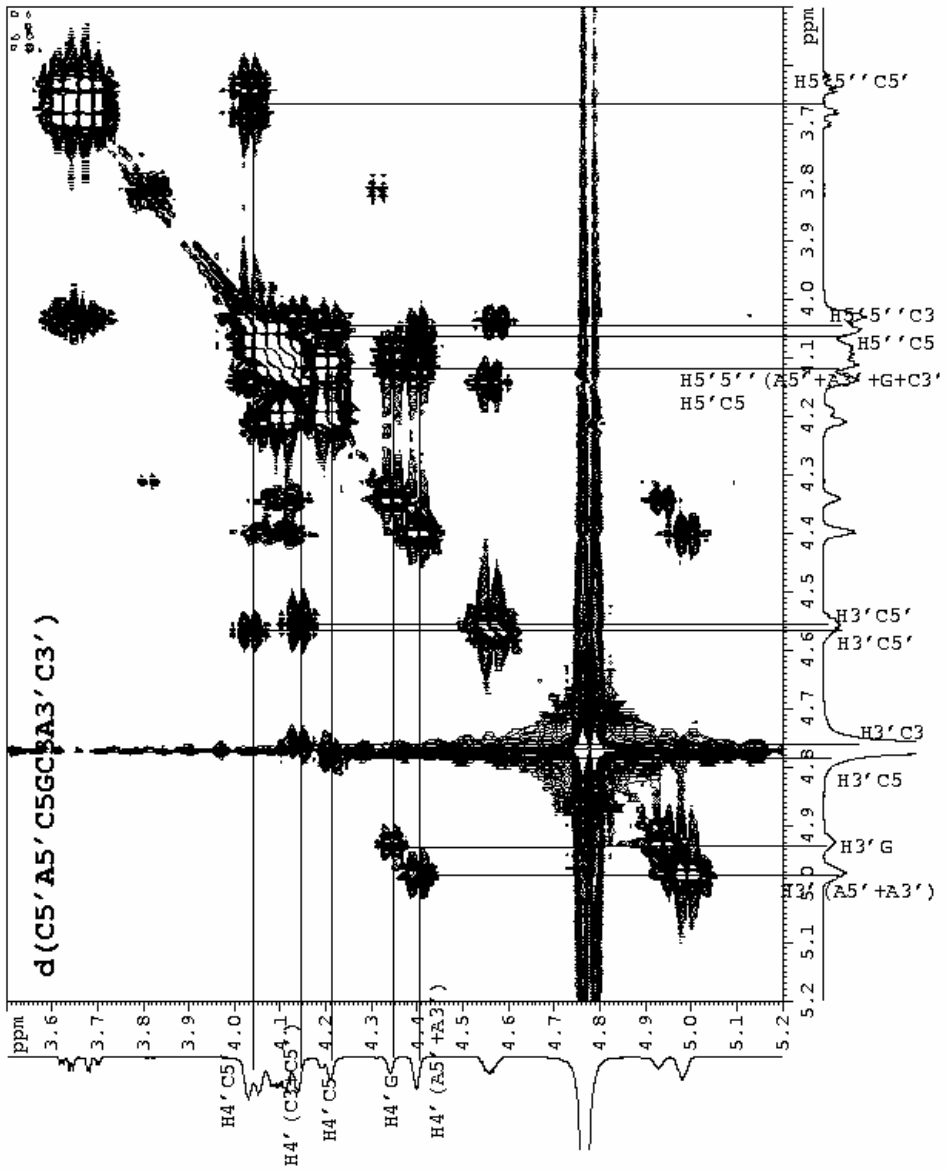


Figure S33. The expanded ^{31}P coupled DQF-COSY spectra of H3'/H4'/H5'/H5" region for d($^{5'}$C^{1}A^{2}C^{3}G^{4}C^{5}A^{6}C^{7}) (**4**) at 298 K. The region 5.2 – 3.5 ppm in both F1 and F2 dimensions of the d($^{5'}$C^{1}A^{2}C^{3}G^{4}C^{5}A^{6}C^{7}) (**4**) showing the spin connectivity.

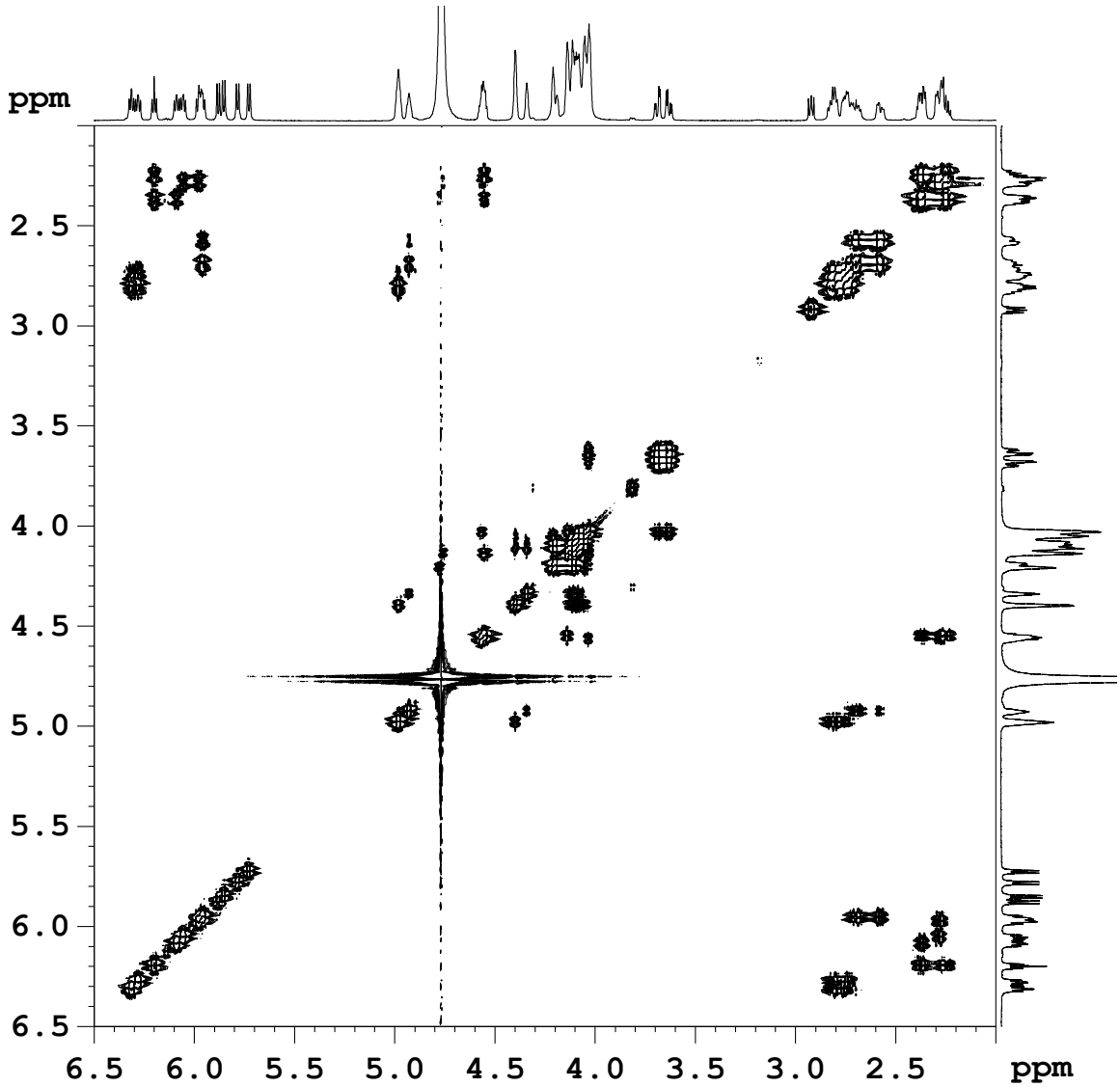


Figure S33.1. The ^{31}P decoupled DQF-COSY spectrum of d($^{5'}$C^{1}A^{2}C^{3}G^{4}C^{5}A^{6}C^{7}) (**4**) at 298 K. For assignments see **S32** and **S33**.

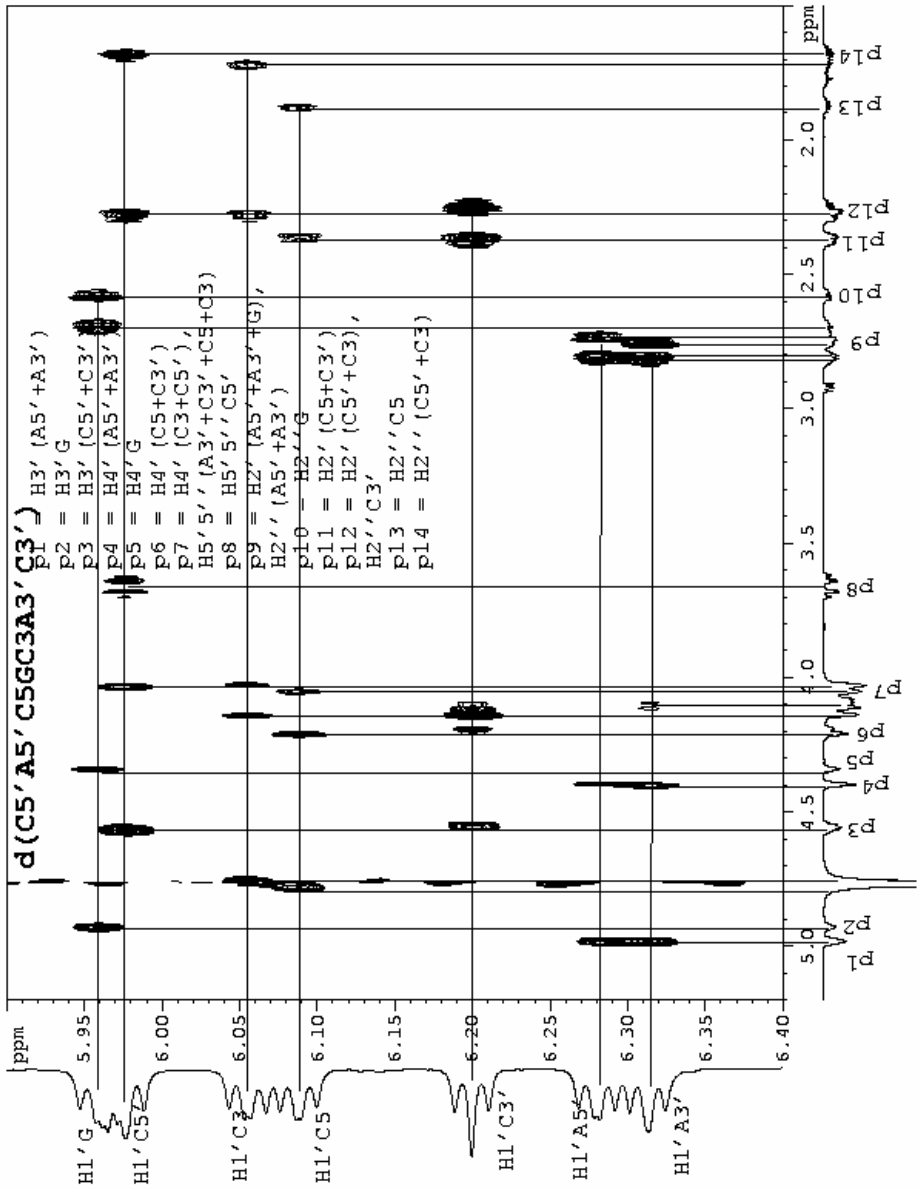


Figure S34. Expanded TOCSY spectra of the H2'/H2"/H3'/H4'/H5'/H5" region (1.5 – 5.2 ppm in F1 direction) to anomeric (H1') region (6.4 – 5.9 ppm in F2 direction) for d($^{5'}$C^{1}A^{2}C^{3}G^{4}C^{5}A^{6}C^{7}) (**4**) at 298 K.

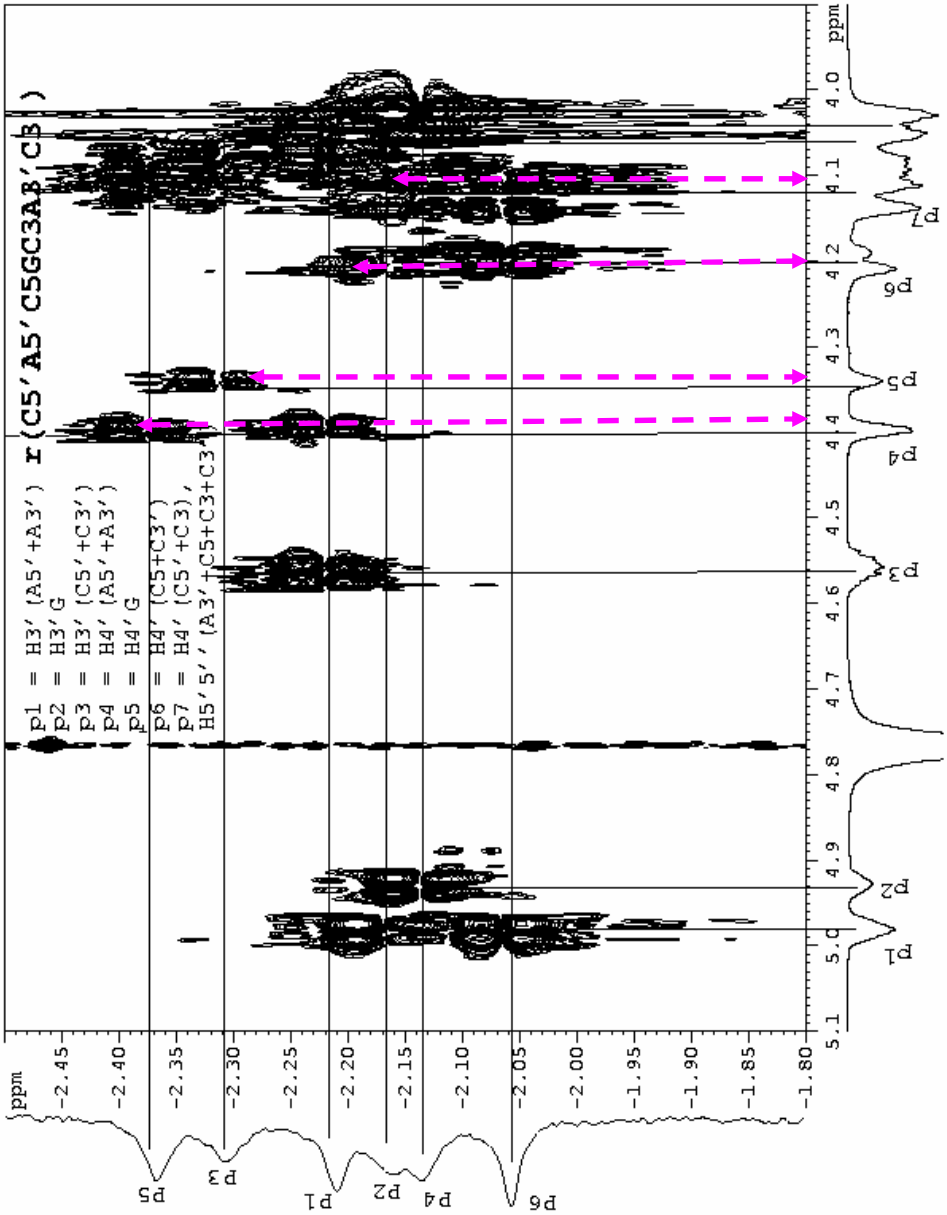


Figure S35. Expanded ^{31}P - ^{1}H correlation spectroscopy of ^{31}P region (-1.8 – -2.5 ppm in F2 direction) to H3'/H4'/H5'/H5" region (5.1 – 3.9 ppm in F1 direction) for d($^{5'}C^{1}p_{1}A^{2}p_{2}C^{3}p_{3}G^{4}p_{4}C^{5}p_{5}A^{6}p_{6}C^{7}$) (**4**) at 298 K. ← - - → H4' – P connectivity.

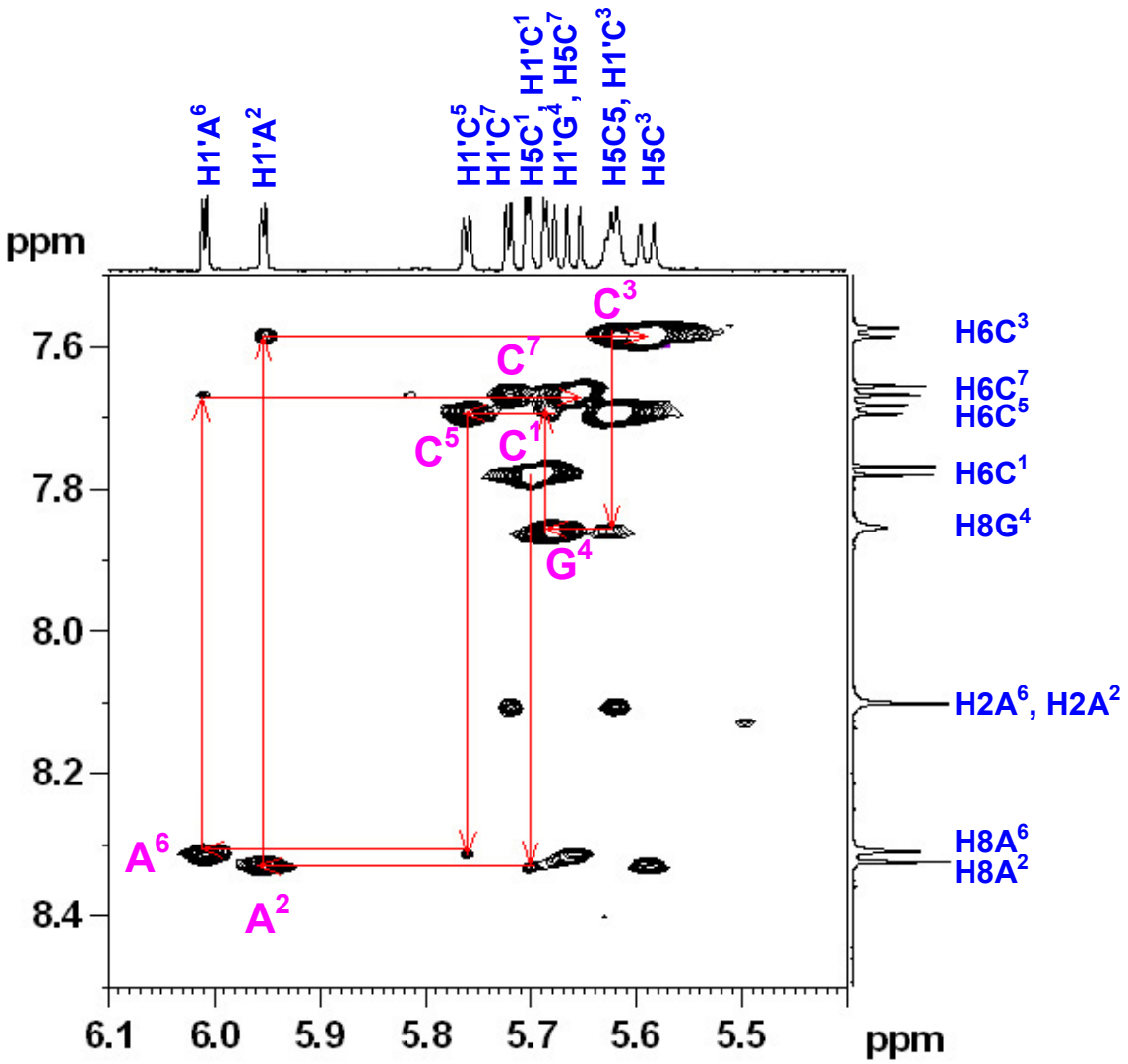


Figure S36. NOESY footprint of r($^{5'}$C^{1}A^{2}C^{3}G^{4}C^{5}A^{6}C^{7}) (**8**) showing the connectivity of nucleotide residues.

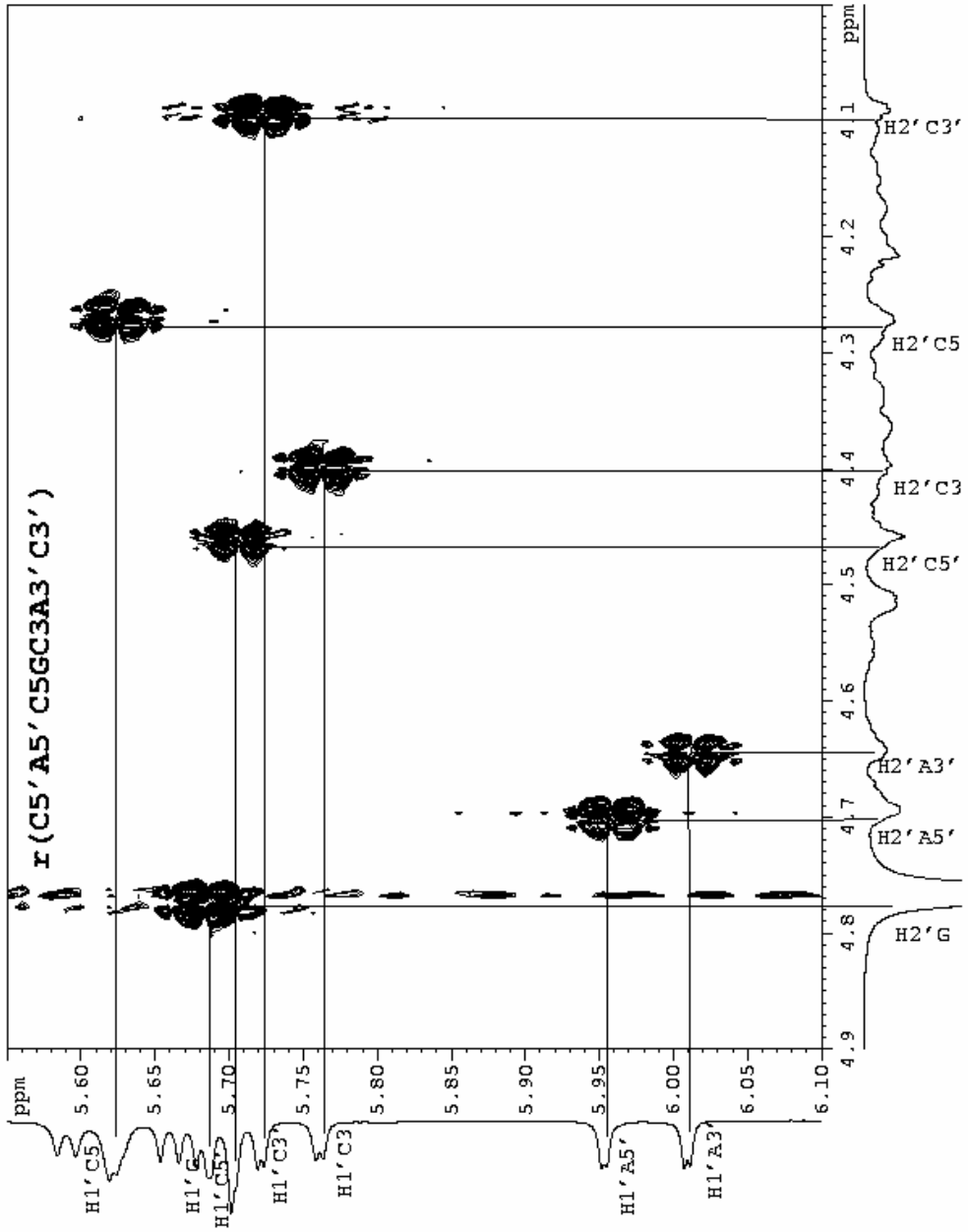


Figure S37. The expanded ^{31}P coupled DQF-COSY spectra of the anomeric H1' region to the H2' for r($^{5'}$C^{1}A^{2}C^{3}G^{4}C^{5}A^{6}C^{7}) (**8**) at 298 K. The spin connectivity between anomeric H1' region (5.5 – 6.1 ppm in F2) and H2' region (4.0 – 4.9 ppm in F1) of r($^{5'}$C^{1}A^{2}C^{3}G^{4}C^{5}A^{6}C^{7}) (**8**) have been shown.

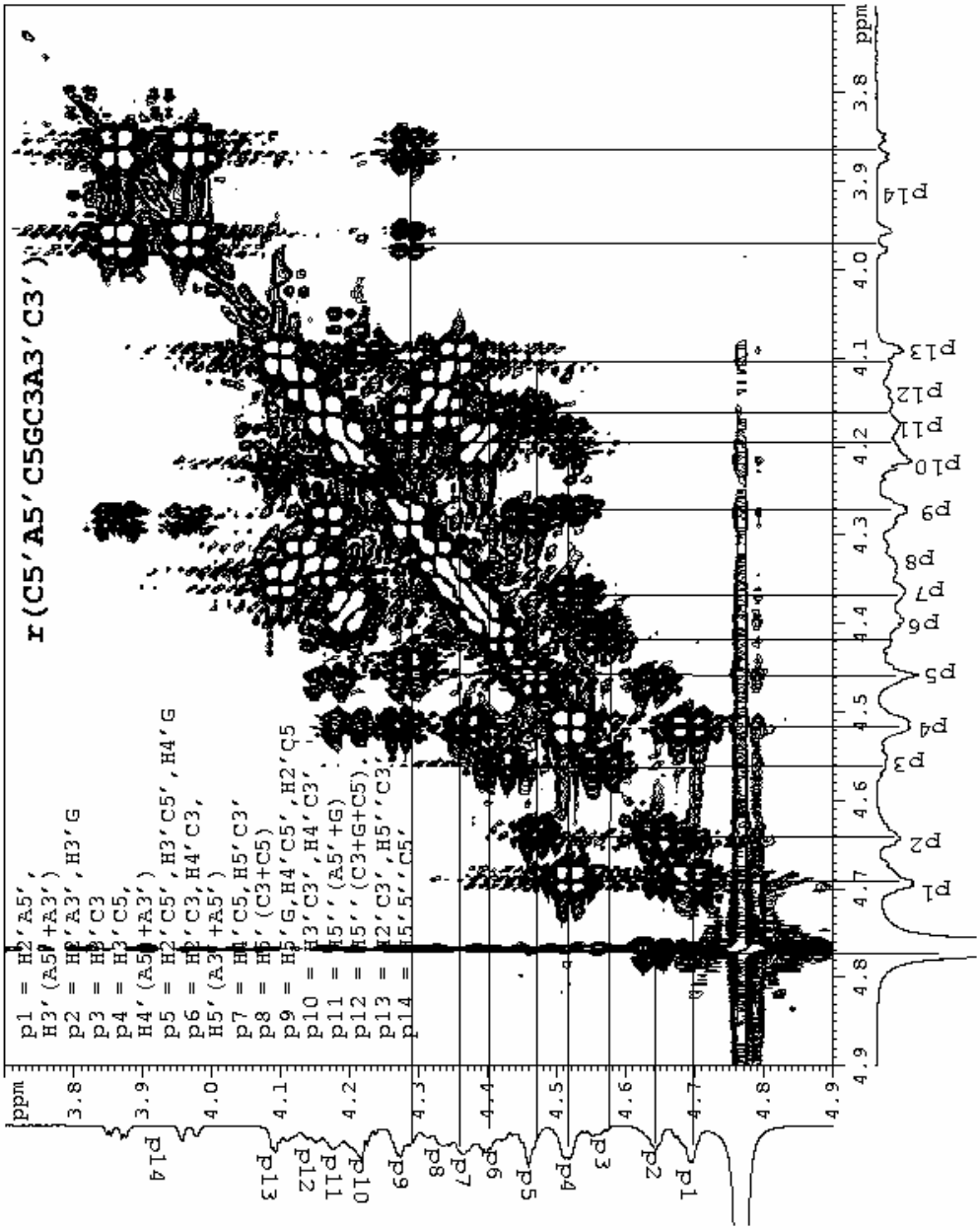


Figure S38. The expanded ^{31}P coupled DQF-COSY spectra of H2'/H3'/H4'/H5'/H5" region for r($^{5'}$C^{1}A^{2}C^{3}G^{4}C^{5}A^{6}C^{7}) (**8**) at 298 K. The region 4.9 – 3.7 ppm in both F1 and F2 dimensions of the r($^{5'}$C^{1}A^{2}C^{3}G^{4}C^{5}A^{6}C^{7}) (**8**) showing the spin connectivity.

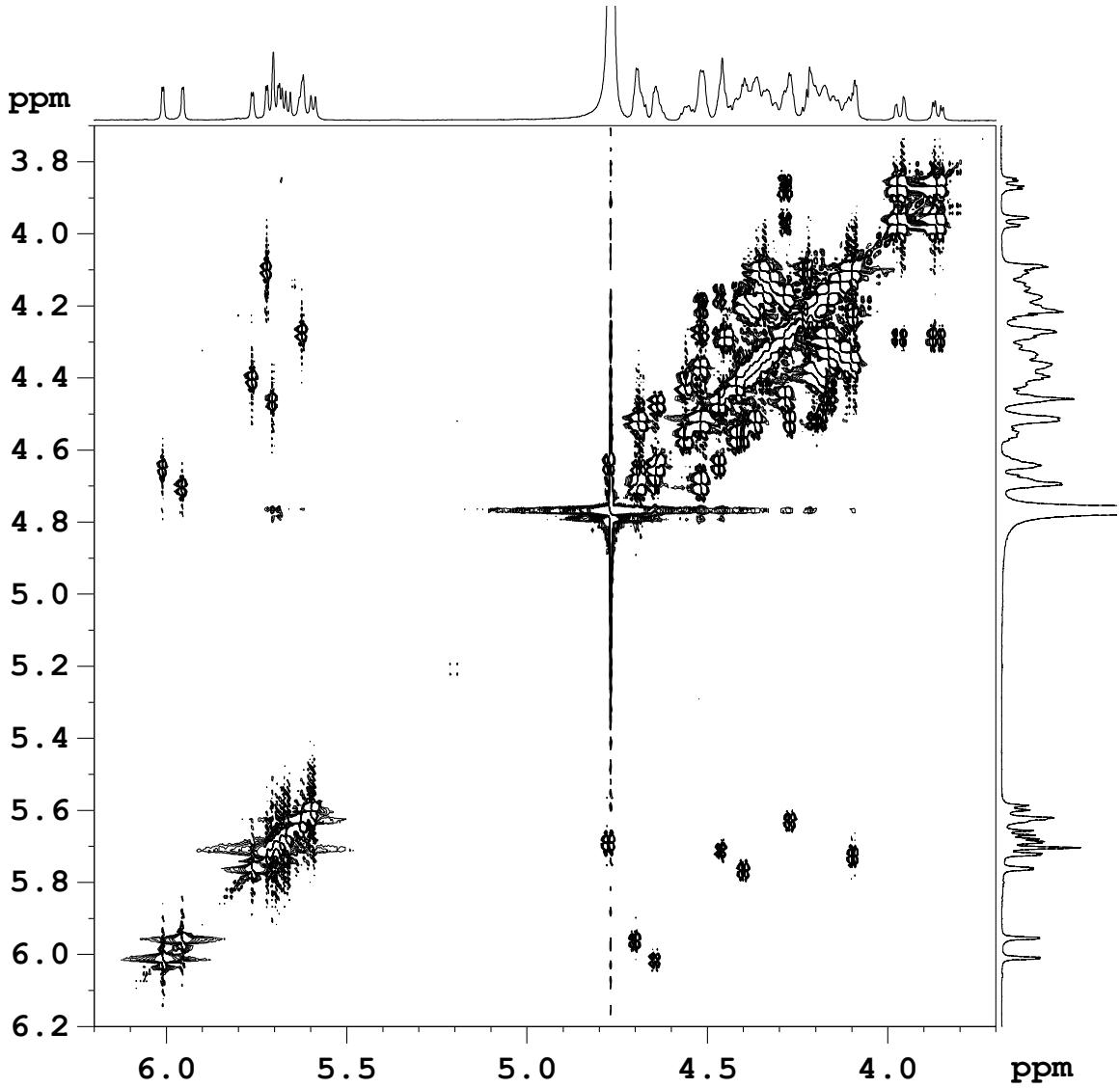


Figure S38.1. The ^{31}P decoupled DQF-COSY spectrum of r($^{5'}$C^{1}A^{2}C^{3}G^{4}C^{5}A^{6}C^{7}) (**8**) at 298 K. For assignments see **S37** and **S38**.

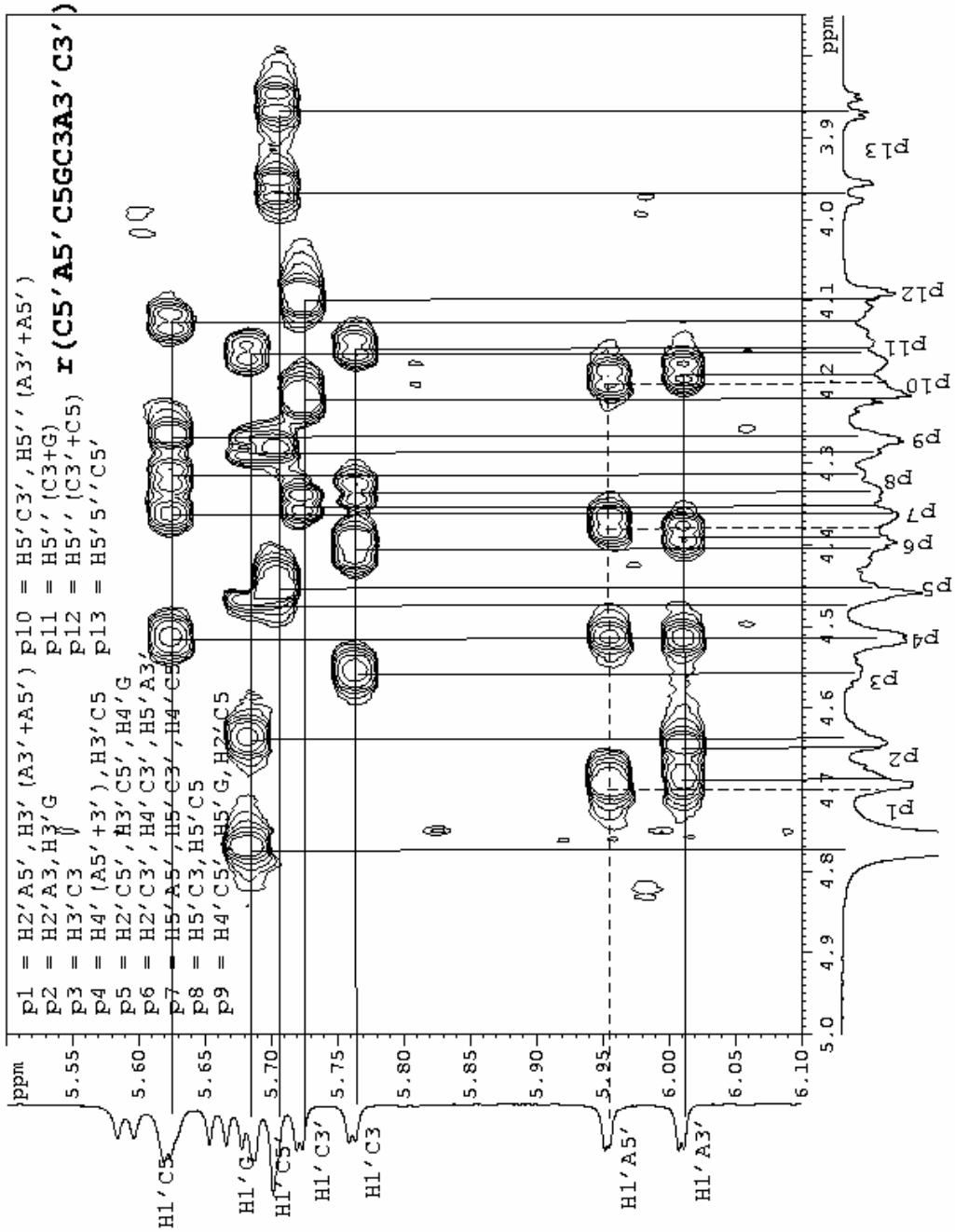


Figure S39. Expanded TOCSY spectra of the H2'/H3'/H4'/H5'/H5" region (5.0 – 3.75 ppm in F1 direction) to anomeric (H1') region (5.4 – 6.1 ppm in F2 direction) for r($^{5'}$C^{1}A^{2}C^{3}G^{4}C^{5}A^{6}C^{7}) (**8**) at 298 K.

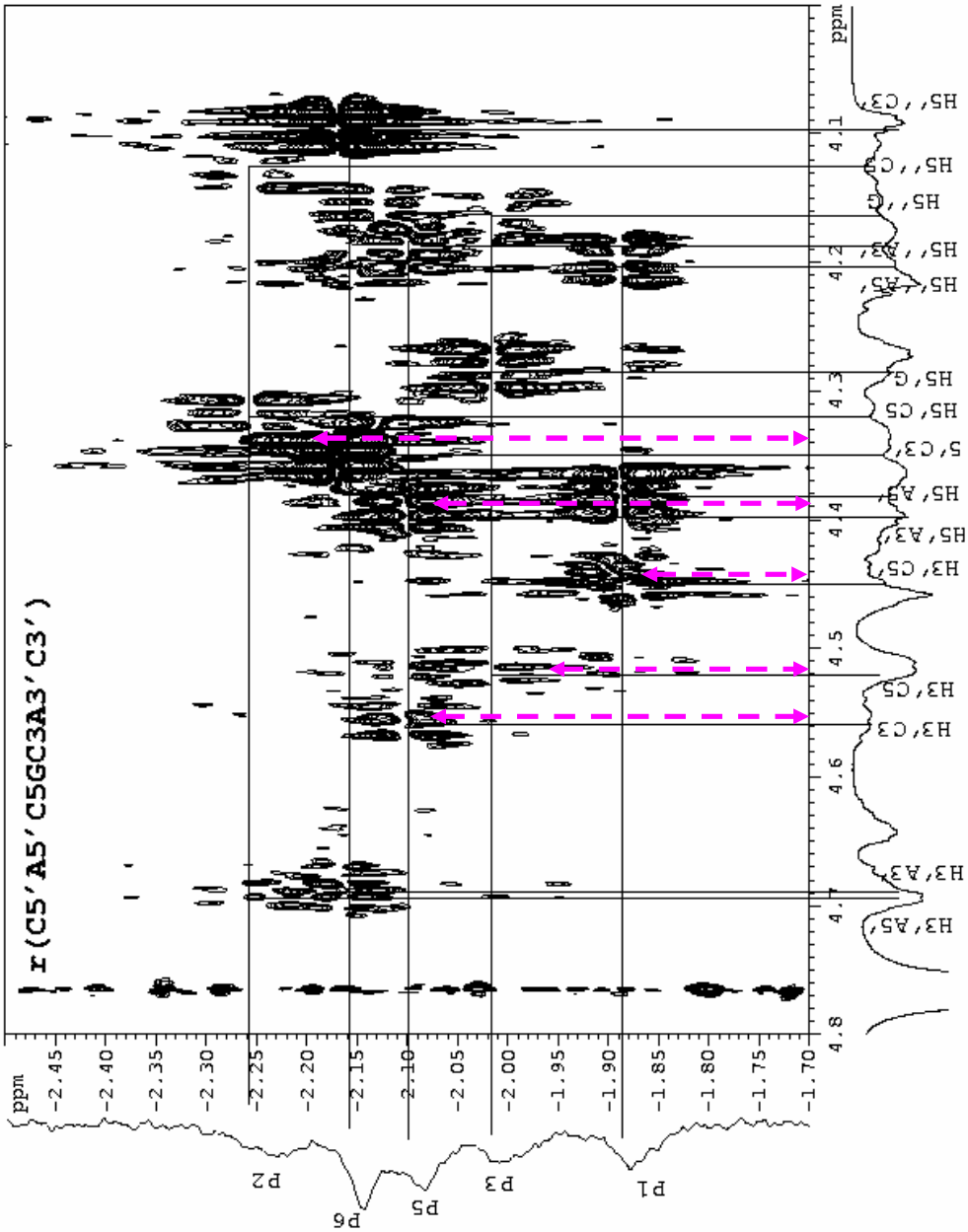


Figure S40. Expanded ^{31}P - ^{1}H correlation spectroscopy of ^{31}P region (-1.7 – -2.5 ppm in F2 direction) to H2'/H3'/H4'/H5'/H5" region (4.8 – 4.0 ppm in F1 direction) for r($^{5'}$C^{1}p$_{1}$A^{2}p$_{2}$C^{3}p$_{3}$G^{4}p$_{4}$C^{5}p$_{5}$A^{6}p$_{6}$C^{7}) (**8**) at 298 K. ← - - ► H4' – P connectivity.

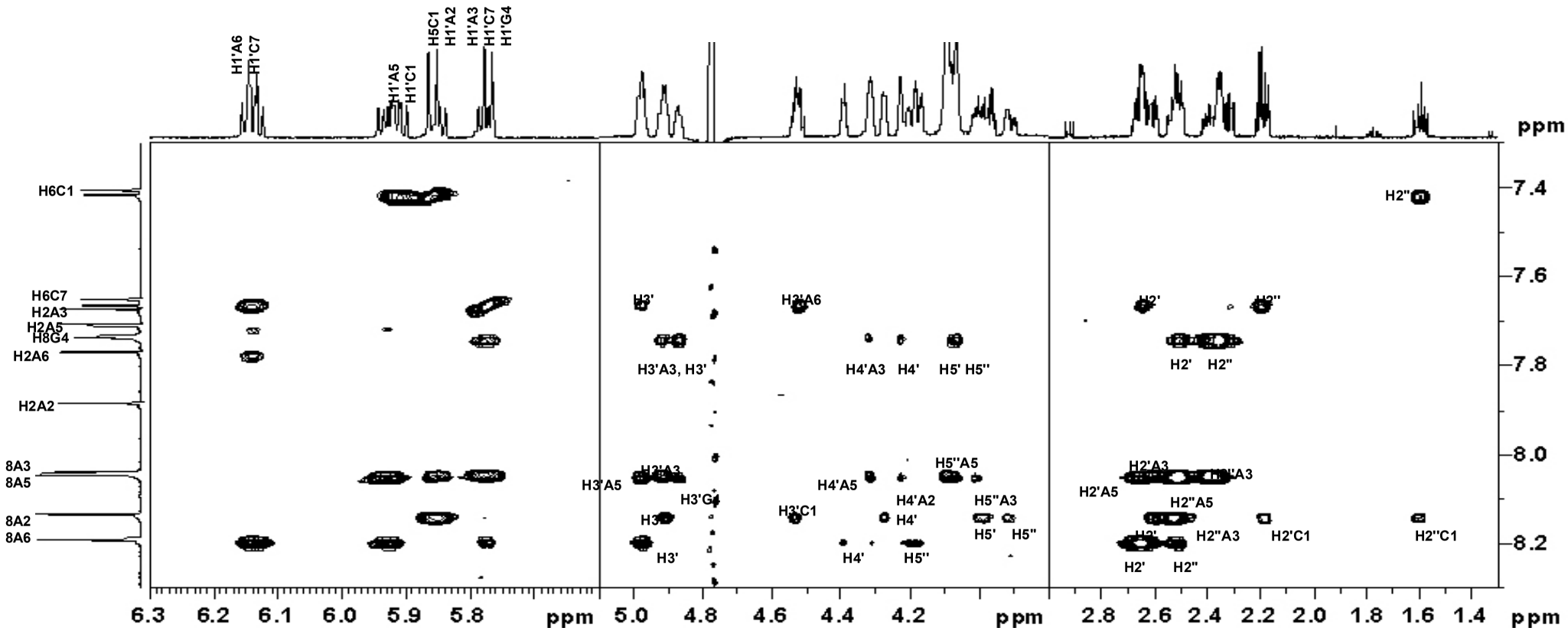


Figure S41. Aromatic – anomeric and aromatic – sugar proton NOESY crosspeaks for d($^{5'}C^1A^2A^3G^4A^5A^6C^7$) (**1**).

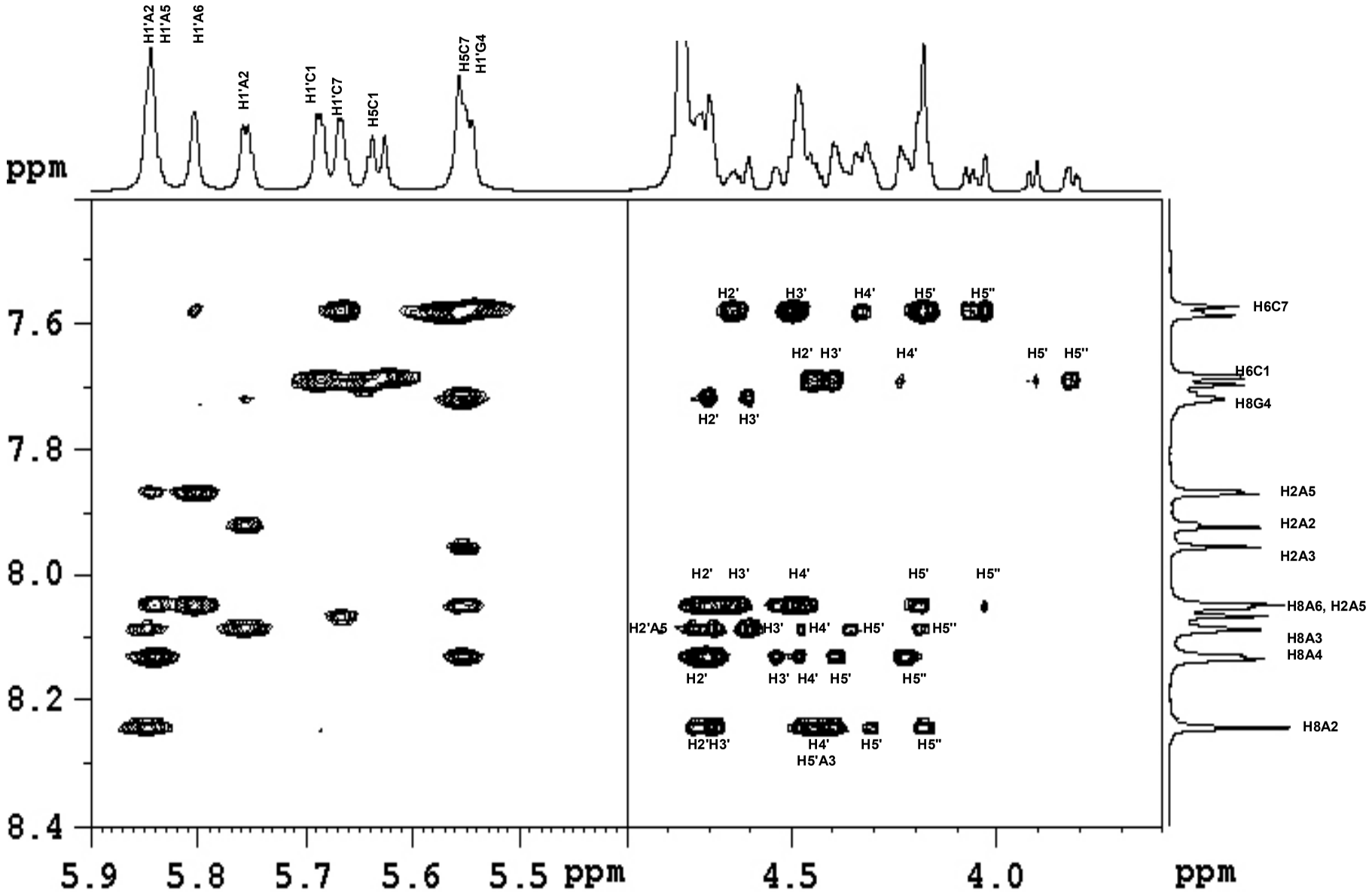


Figure S42. Aromatic – anomeric and aromatic – sugar proton NOESY crosspeaks for r($^{5'}$C^{1}A^{2}A^{3}G^{4}A^{5}A^{6}C^{7}) (**5**).

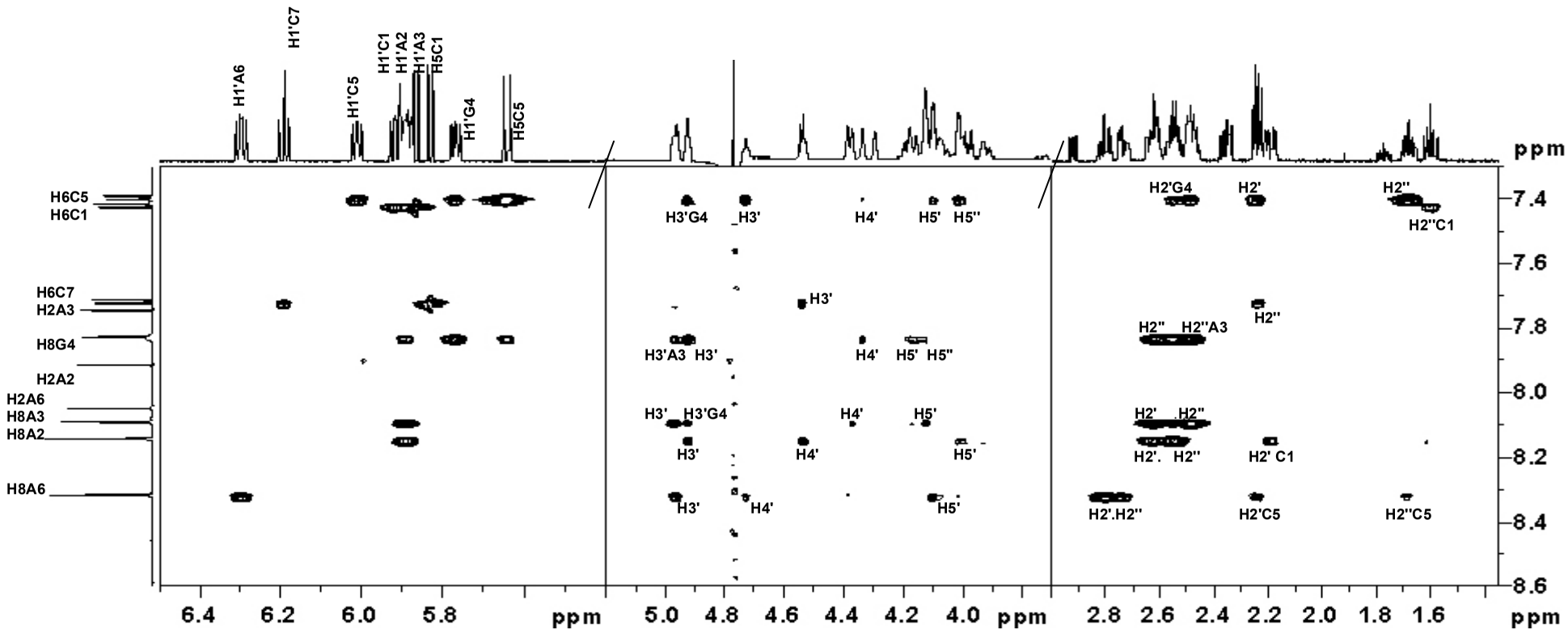


Figure S43. Aromatic – anomeric and aromatic – sugar proton NOESY crosspeaks for d($^{5'}$C^{1}A^{2}A^{3}G^{4}C^{5}A^{6}C^{7}) (**2**).

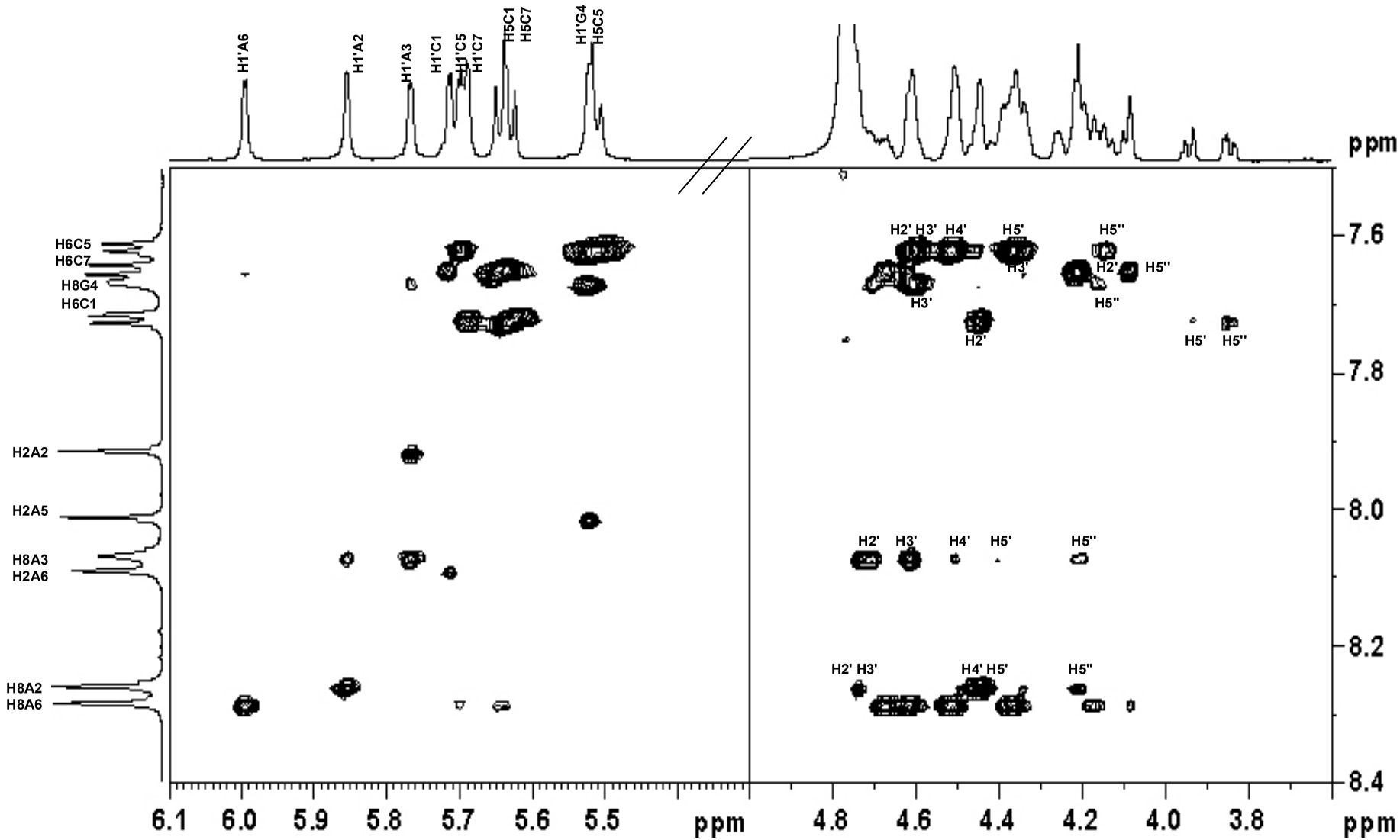


Figure S44. Aromatic – anomeric and aromatic – sugar proton NOESY crosspeaks for r($^{5'}$C^{1}A^{2}A^{3}G^{4}C^{5}A^{6}C^{7}) (**6**).

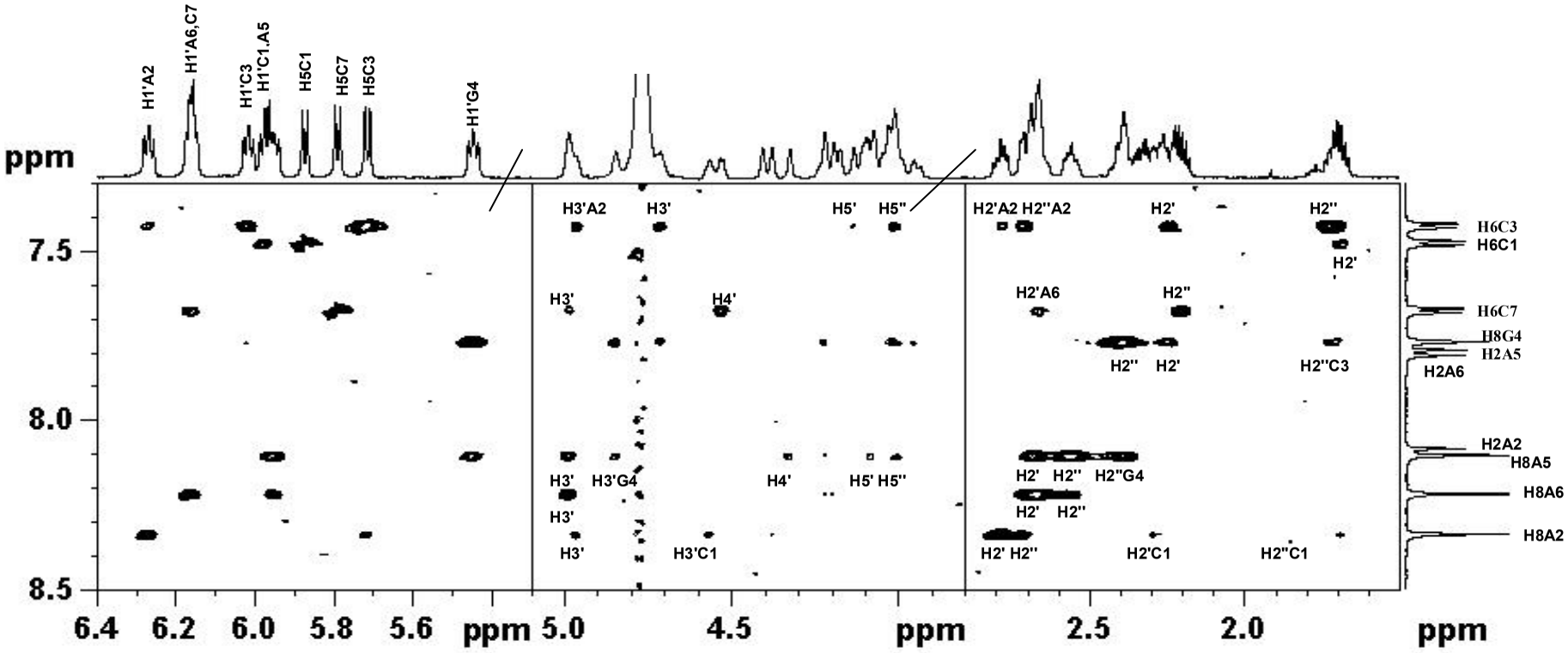


Figure S45. Aromatic – anomeric and aromatic – sugar proton NOESY crosspeaks for d($^{5'}C^1A^2C^3G^4A^5A^6C^7$) (**3**).

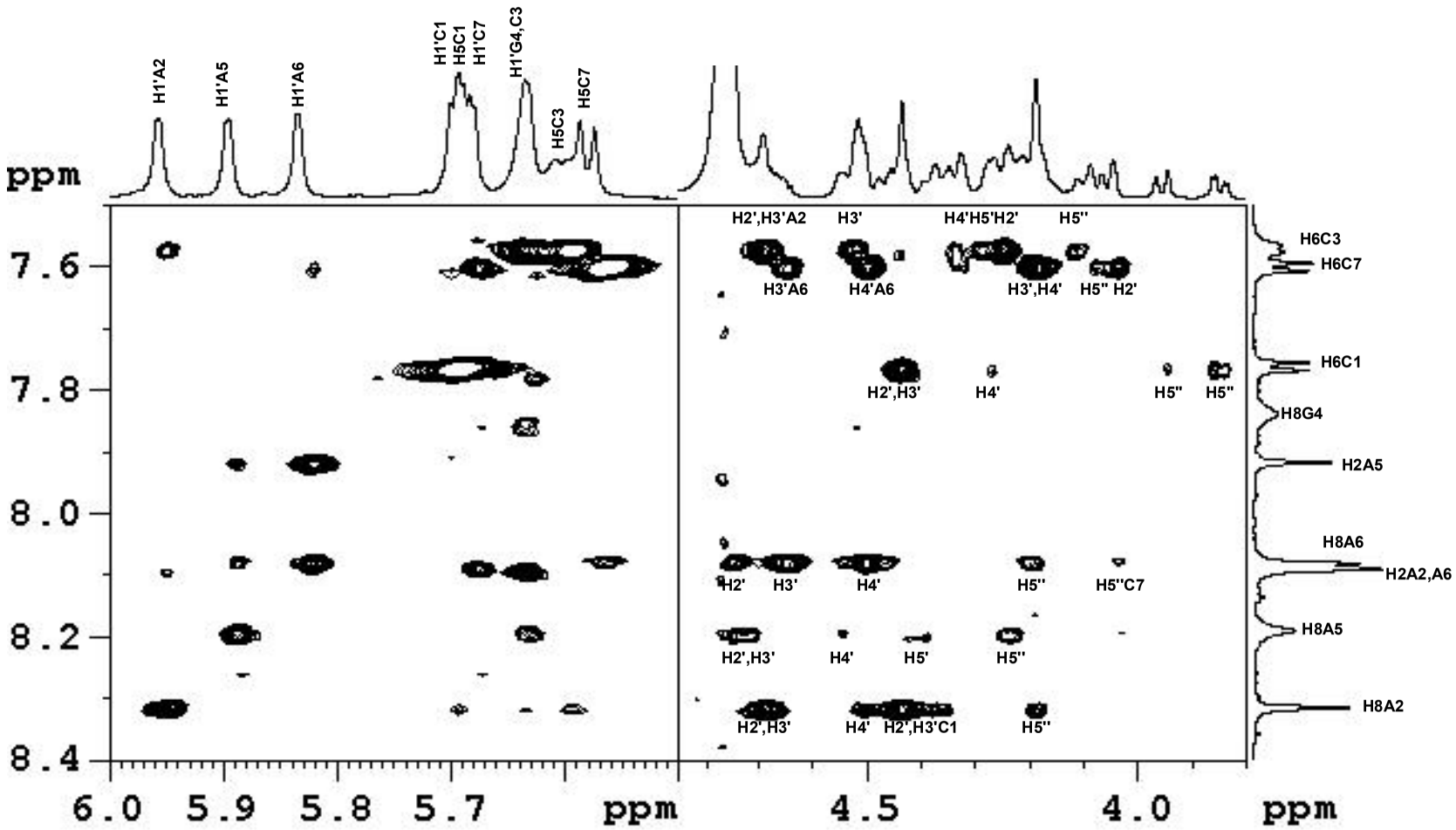


Figure S46. Aromatic – anomeric and aromatic – sugar proton NOESY crosspeaks for r($^{5'}$C^{1}A^{2}C^{3}G^{4}A^{5}A^{6}C^{7}) (**7**).

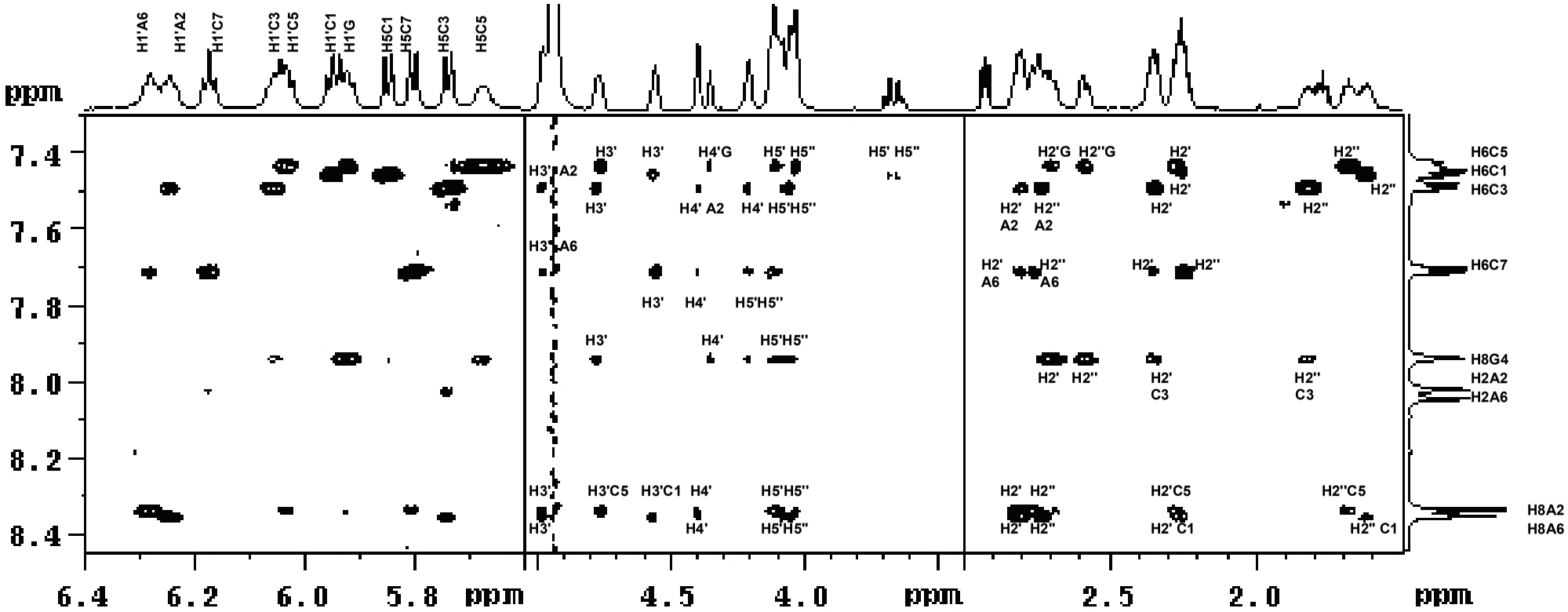


Figure S47. Aromatic – anomeric and aromatic – sugar proton NOESY crosspeaks for d($^{5'}$C^{1}A^{2}C^{3}G^{4}C^{5}A^{6}C^{7}) (**4**).

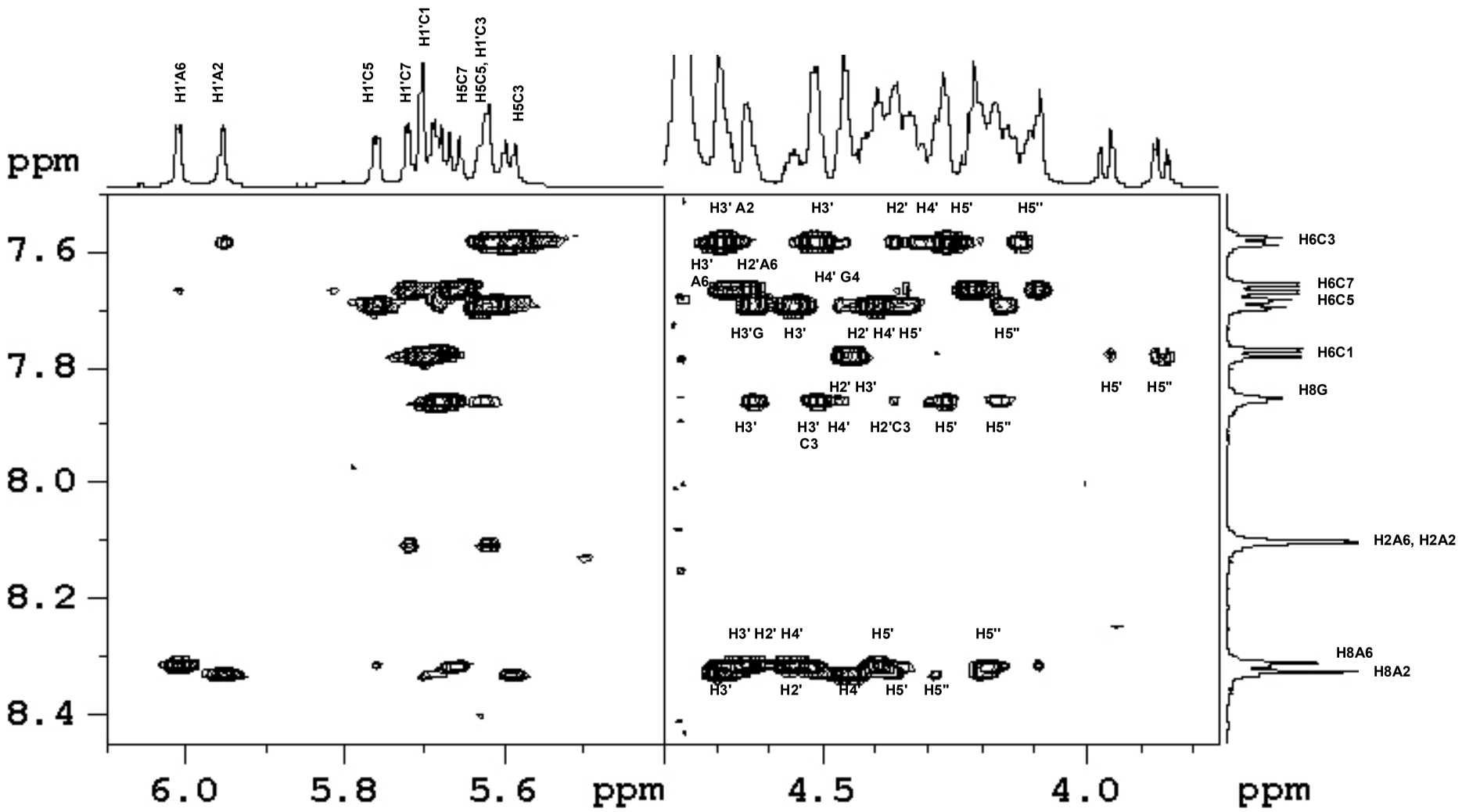


Figure S48. Aromatic – anomeric and aromatic – sugar proton NOESY crosspeaks for r($^{5'}$C^{1}A^{2}C^{3}G^{4}C^{5}A^{6}C^{7}) (**8**).

Figure S49. A.

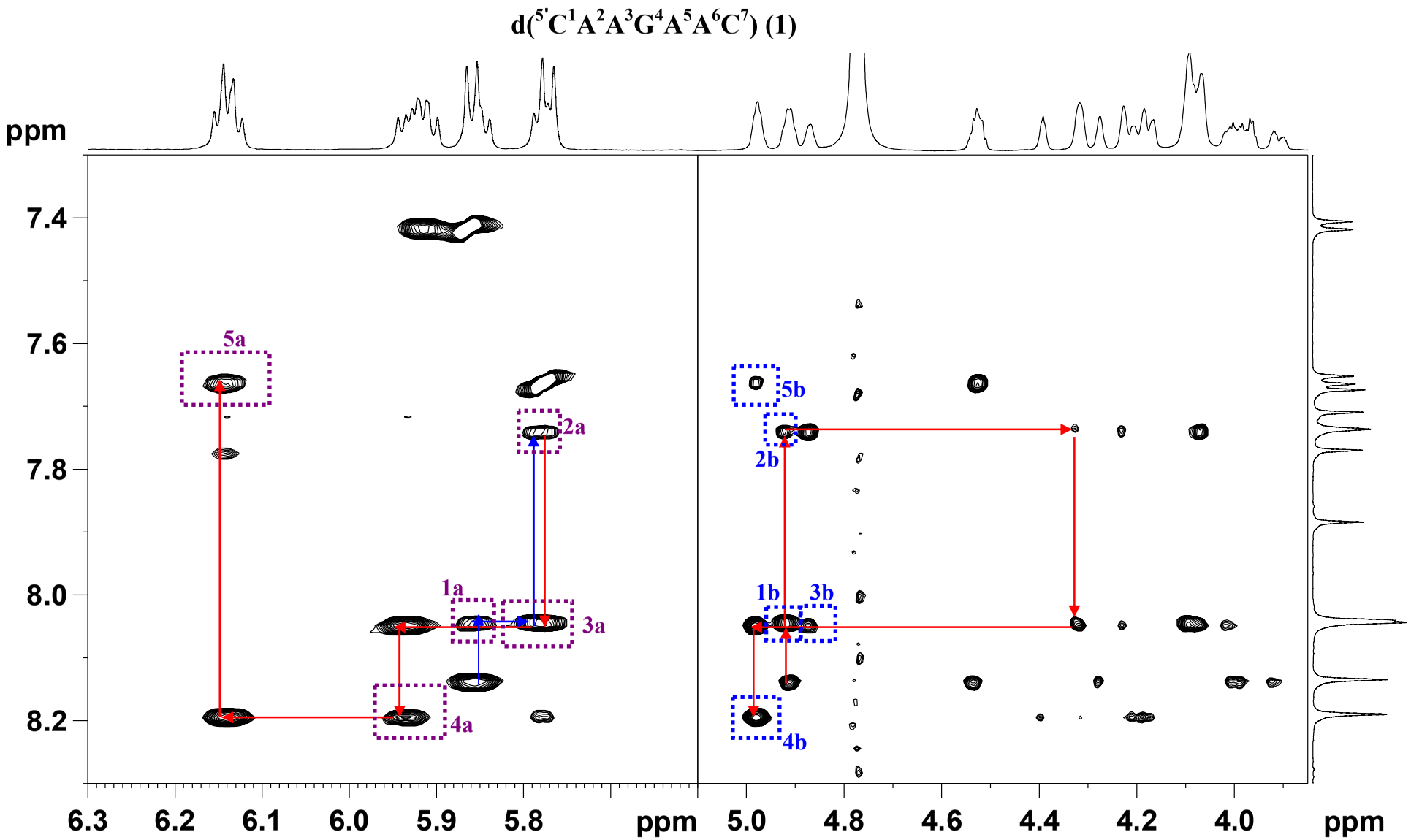


Figure S49. B.

r($^{5'}$C^{1}A^{2}A^{3}G^{4}A^{5}A^{6}C^{7}) (5)

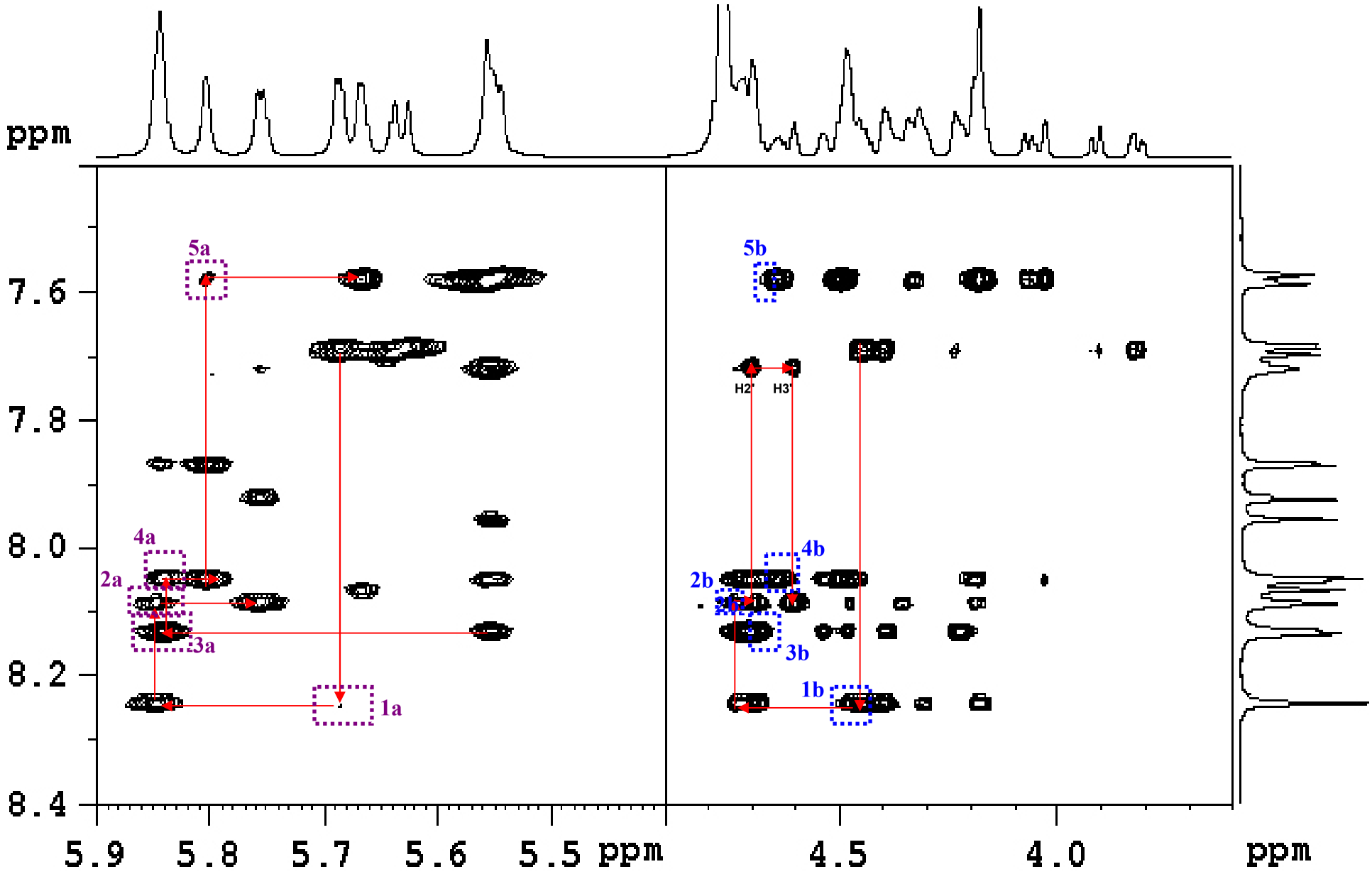


Figure S49. C.

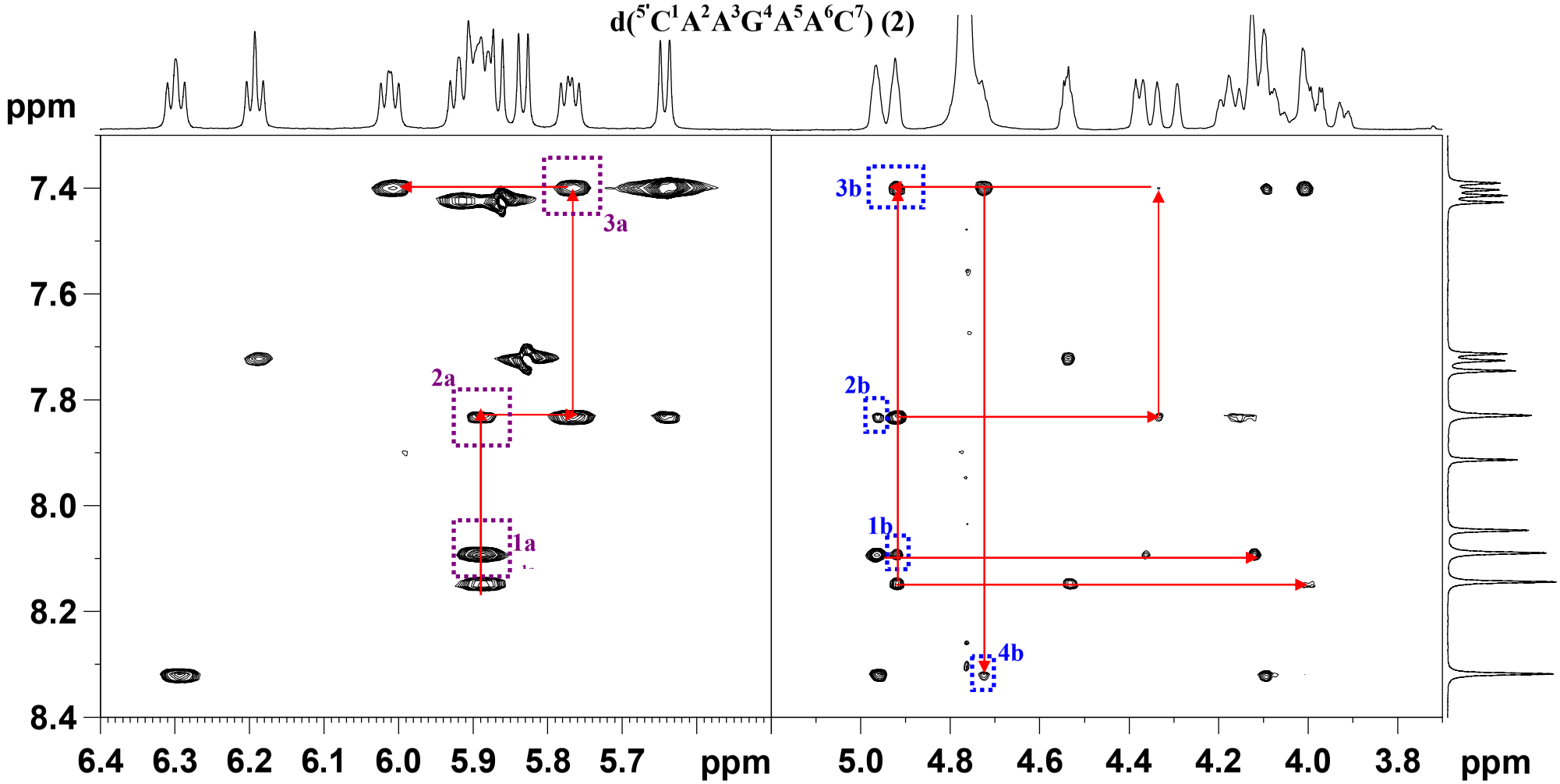


Figure S49. D.

r-($^{5'}C^1A^2A^3G^4C^5A^6C^7$) (6)

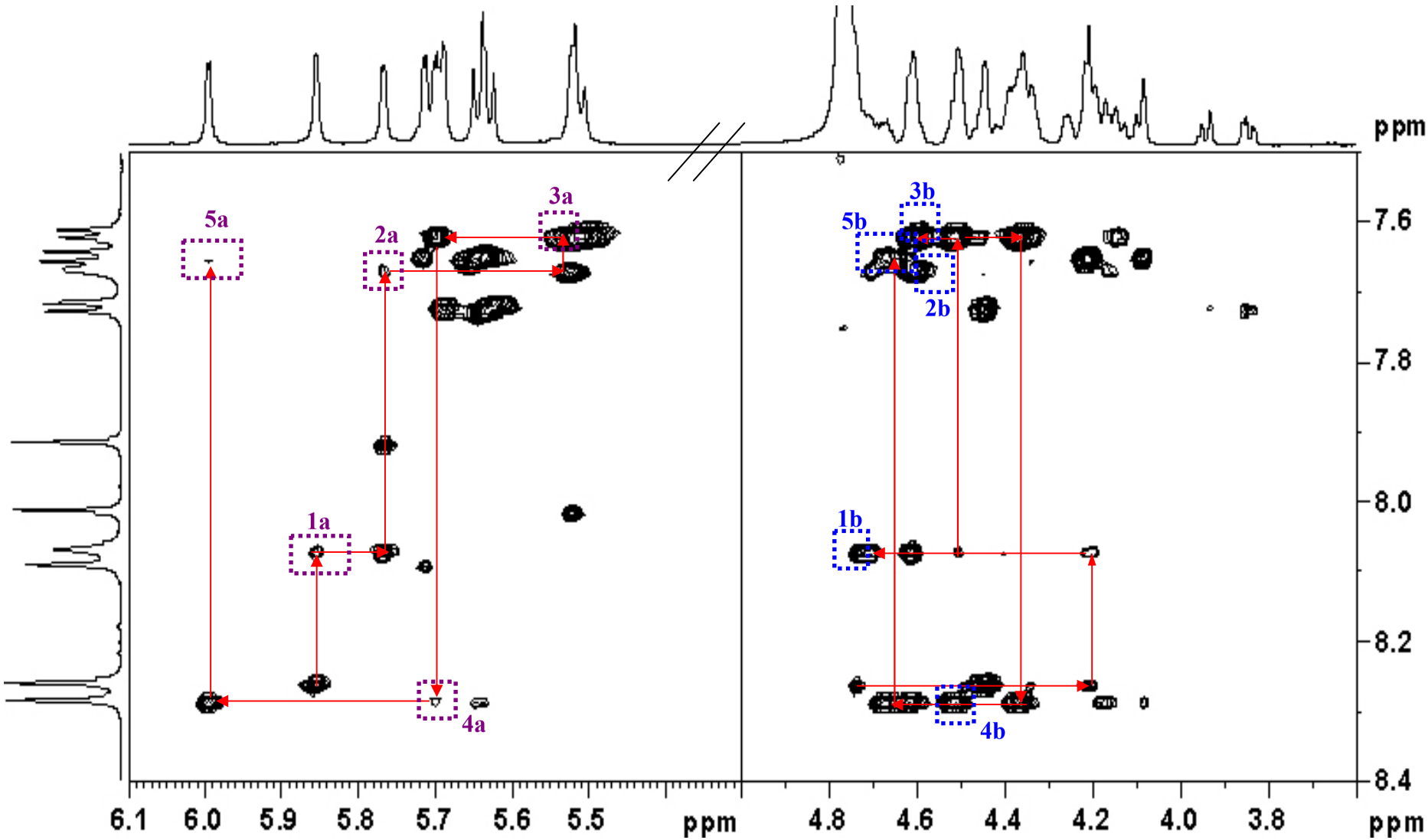


Figure S49. E.

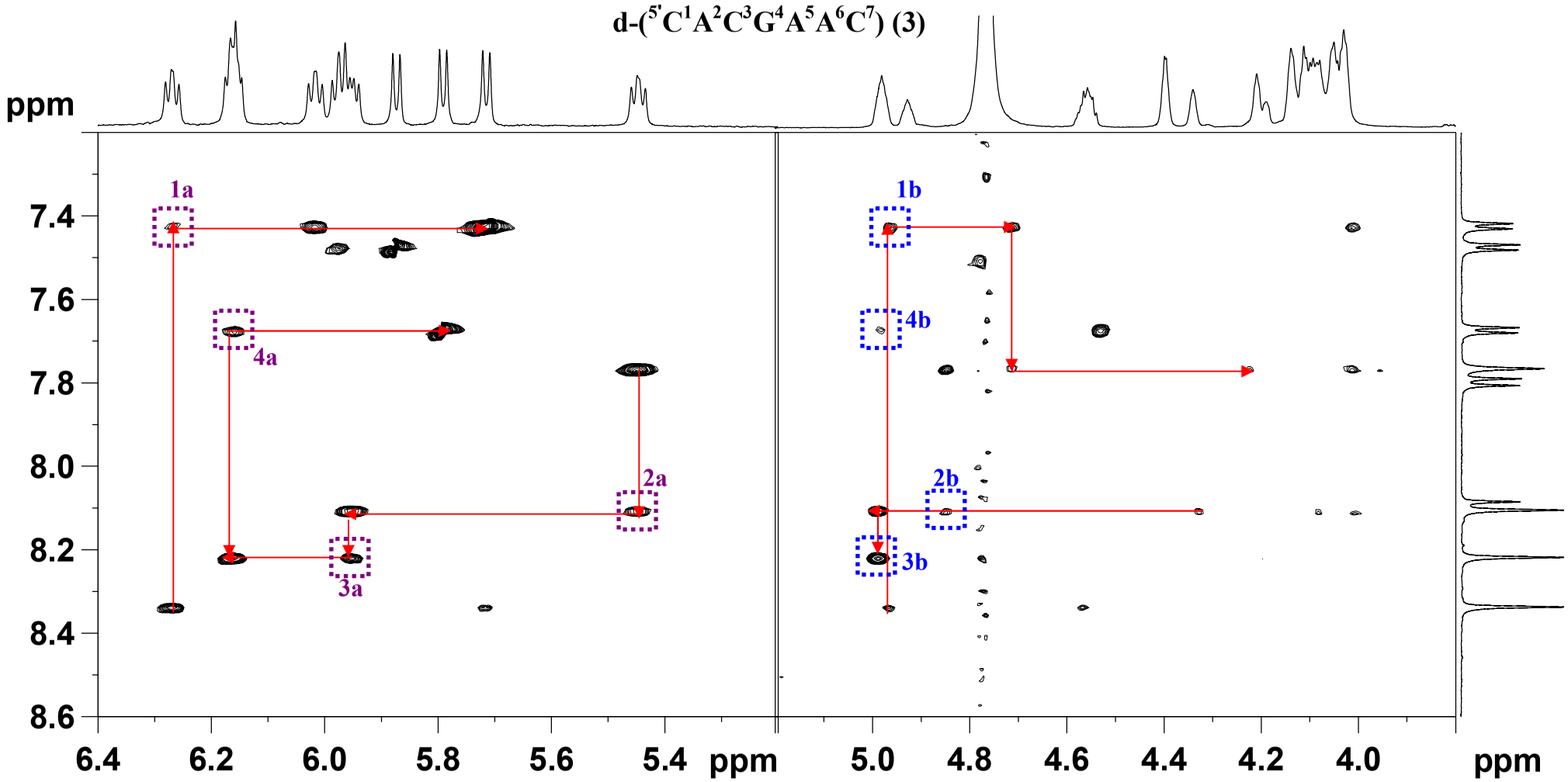


Figure S49. F.

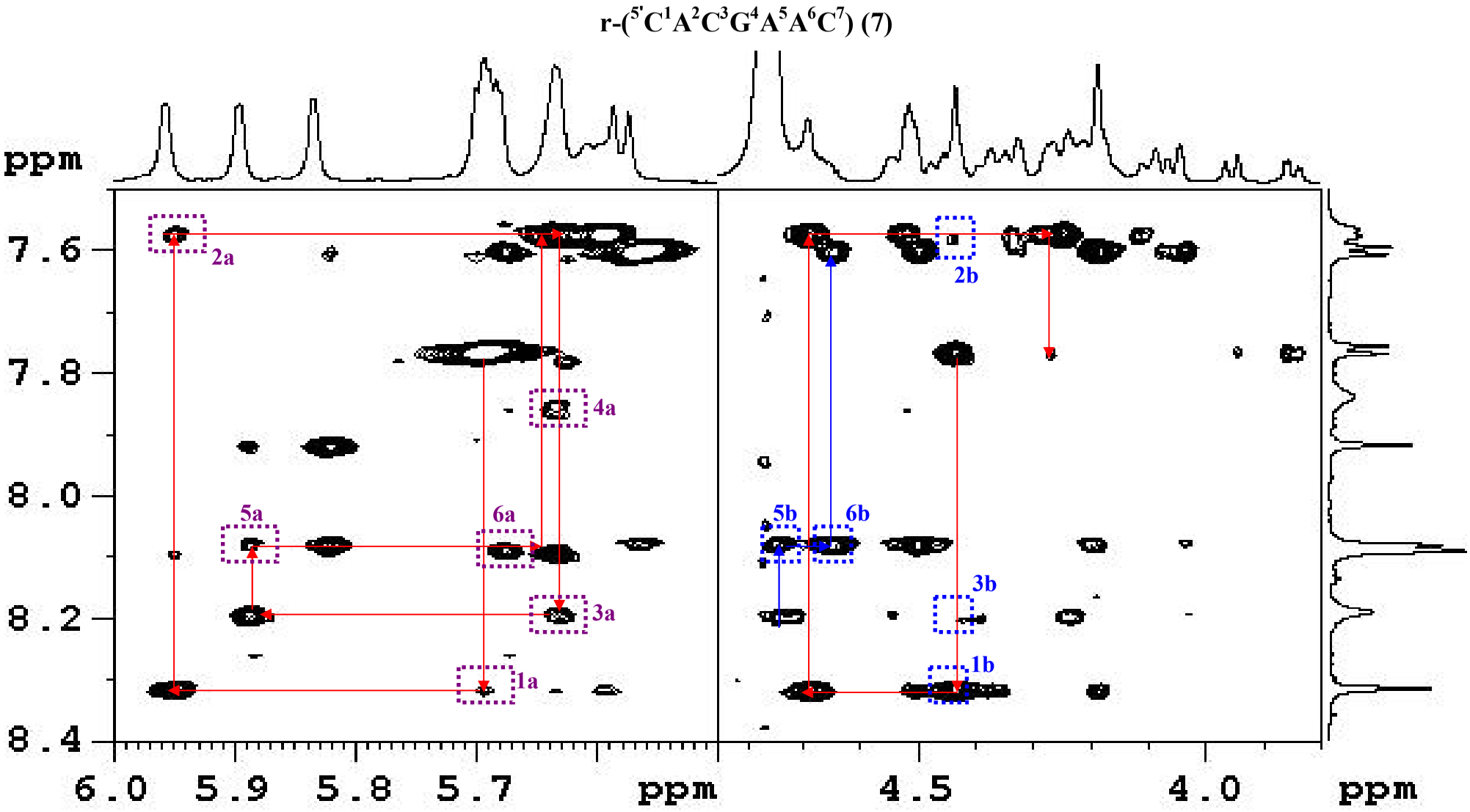


Figure S49. G.

d-($^{5'}C^1A^2C^3G^4C^5A^6C^7$) (4)

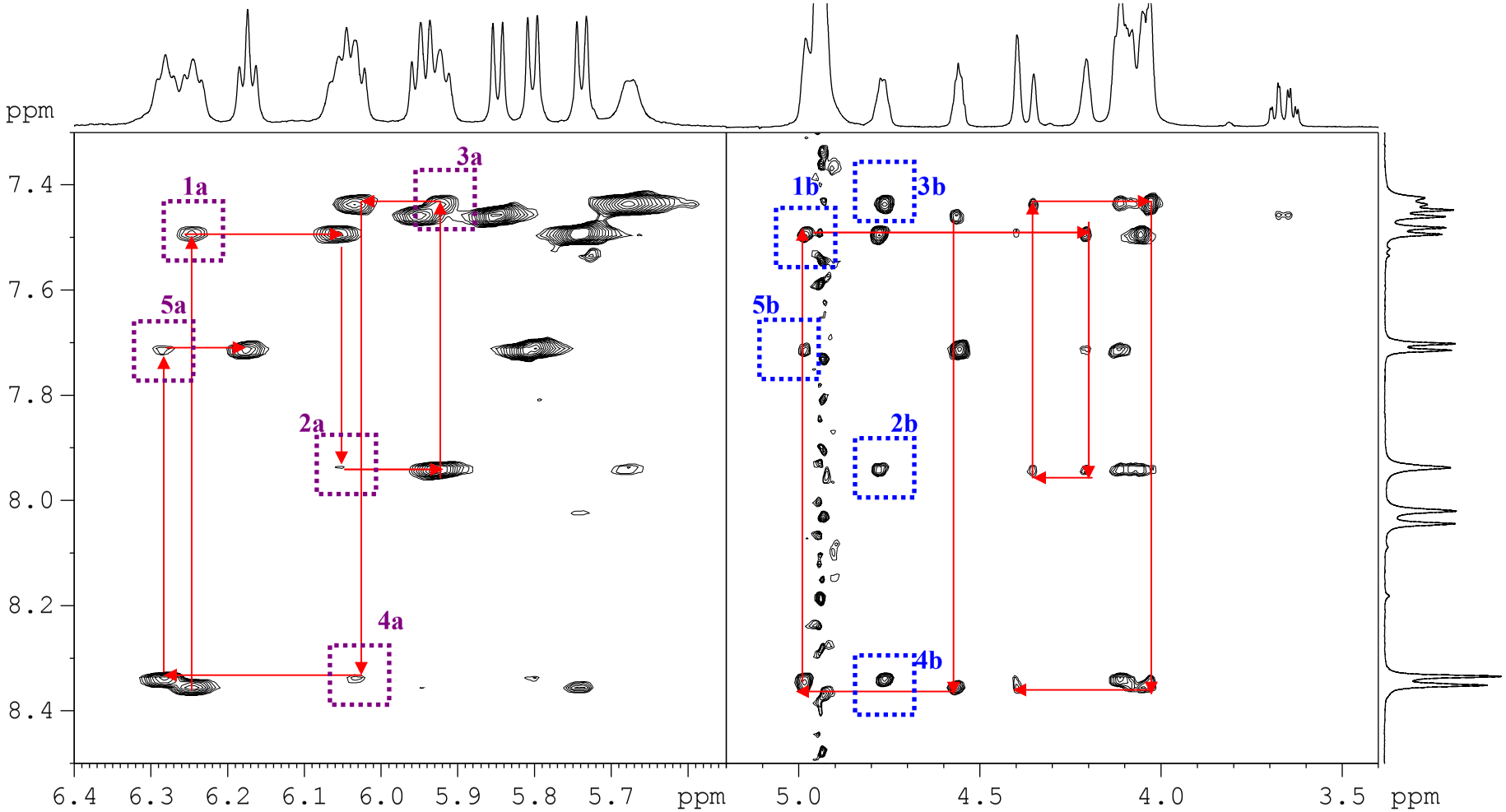


Figure S49. H.

r-($^{5'}$C^{1}A^{2}C^{3}G^{4}C^{5}A^{6}C^{7}) (8)

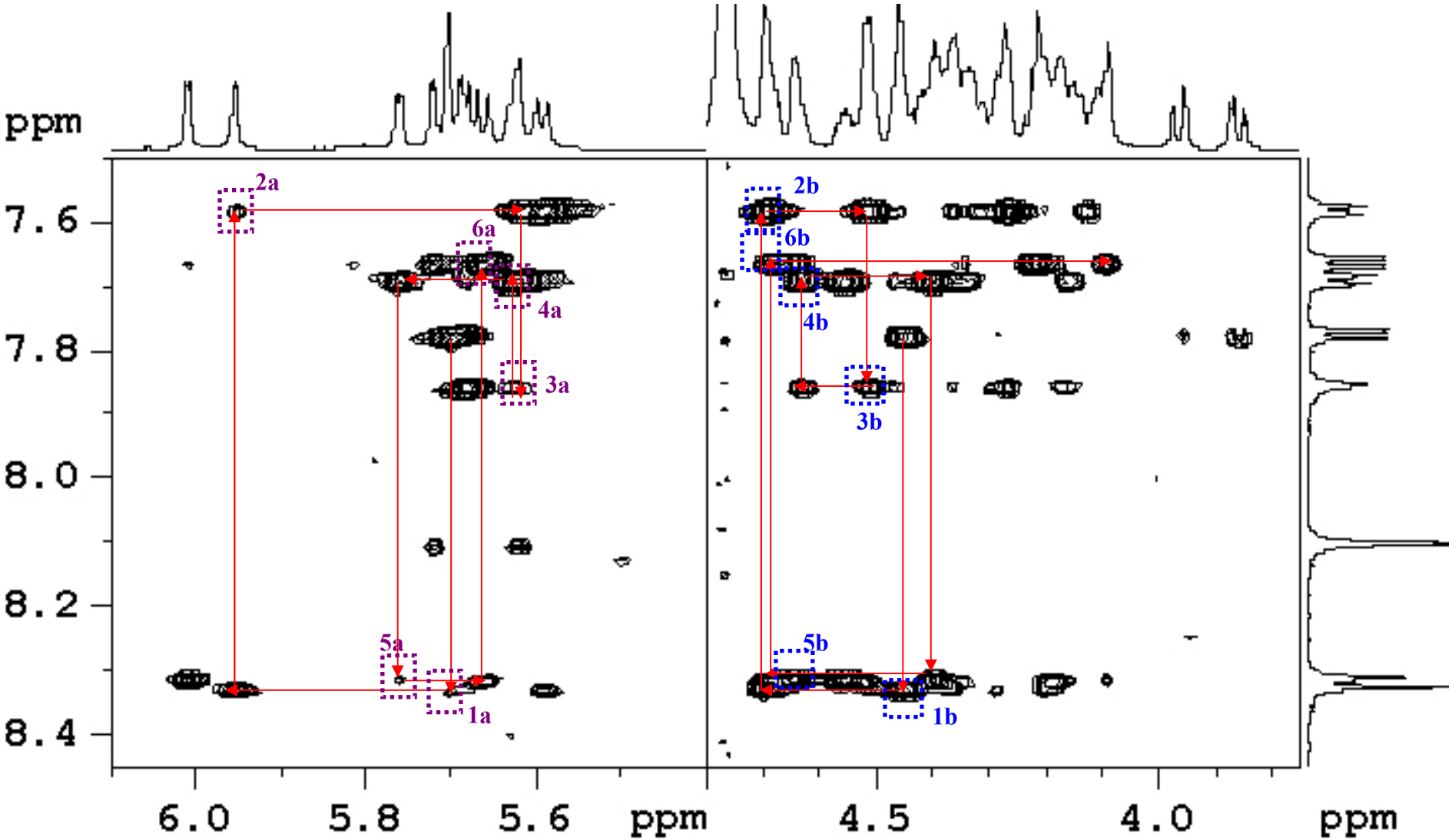
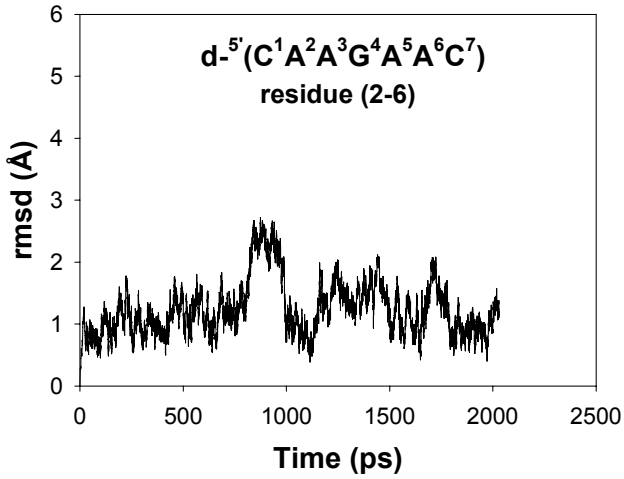
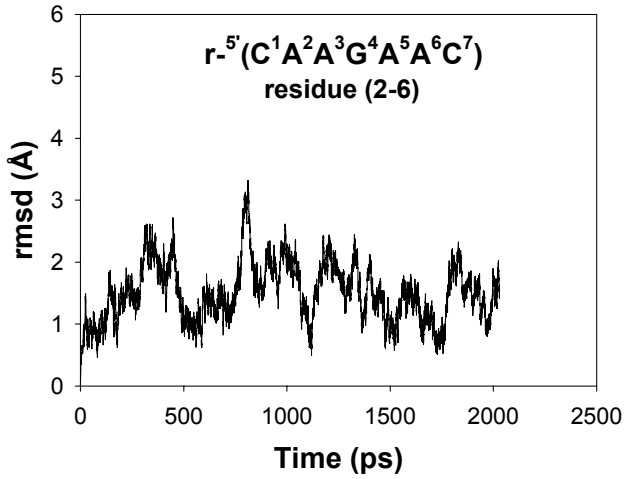
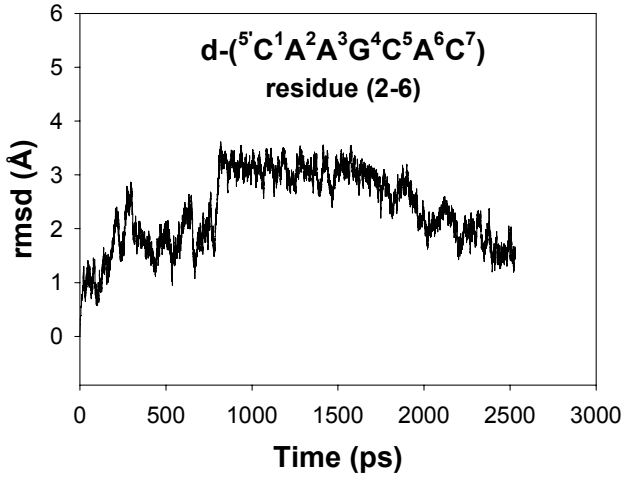
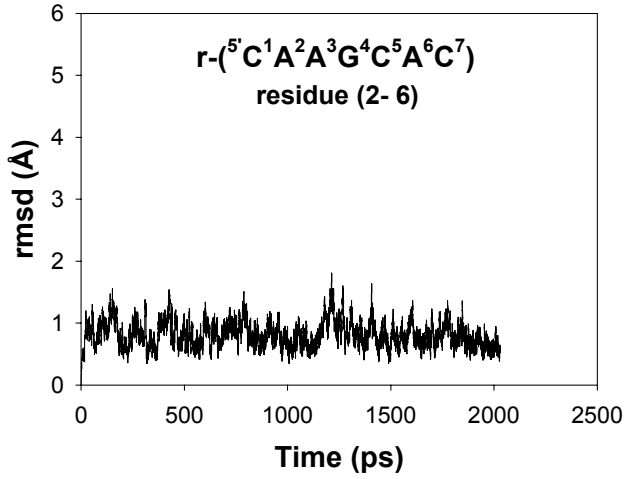


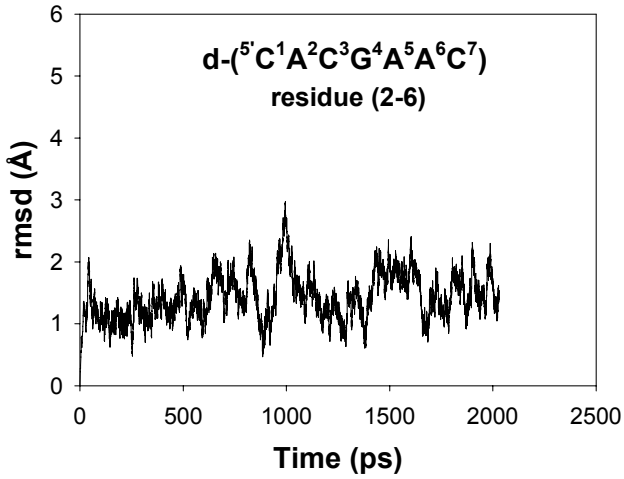
Figure S49. Panels (A) - (D), show the nOe connectivites and cross peak intensities for H8/6$_{(n)}$ – H1'$_{(n-1)}$ [marked as (a)] and H8/6$_{(n)}$ – H3'$_{(n-1)}$ [marked as (b)] in ssDNAs **(1 – 4)**. Panels (E) - (H), show the nOe connectivites and cross peak intensities for H8/6$_{(n)}$ – H1'$_{(n-1)}$ [marked as (a)] and H8/6$_{(n)}$ – H3'$_{(n-1)}$ [marked as (b)] in ssRNAs (**5** – **8**) (see Table S11 for the detailed comparison of relative intensities of H8/H6$_{(n)}$ – H1'$_{(n-1)}$ and H8/H6$_{(n)}$ – H3'$_{(n-1)}$ crosspeaks).

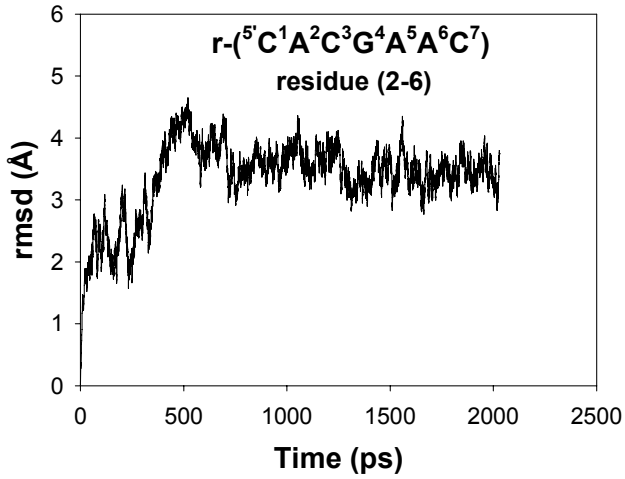












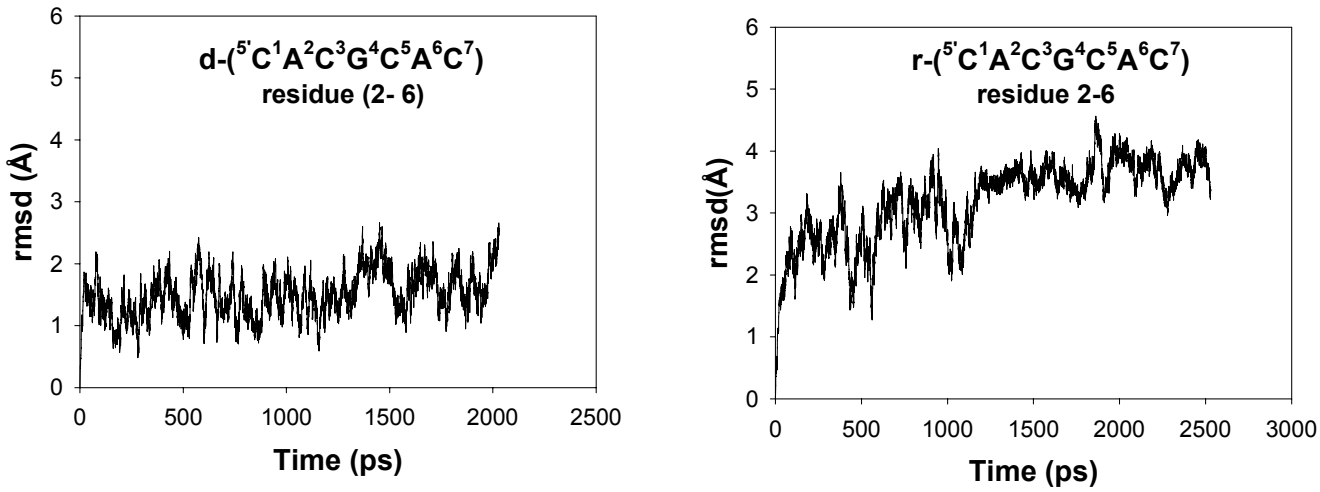
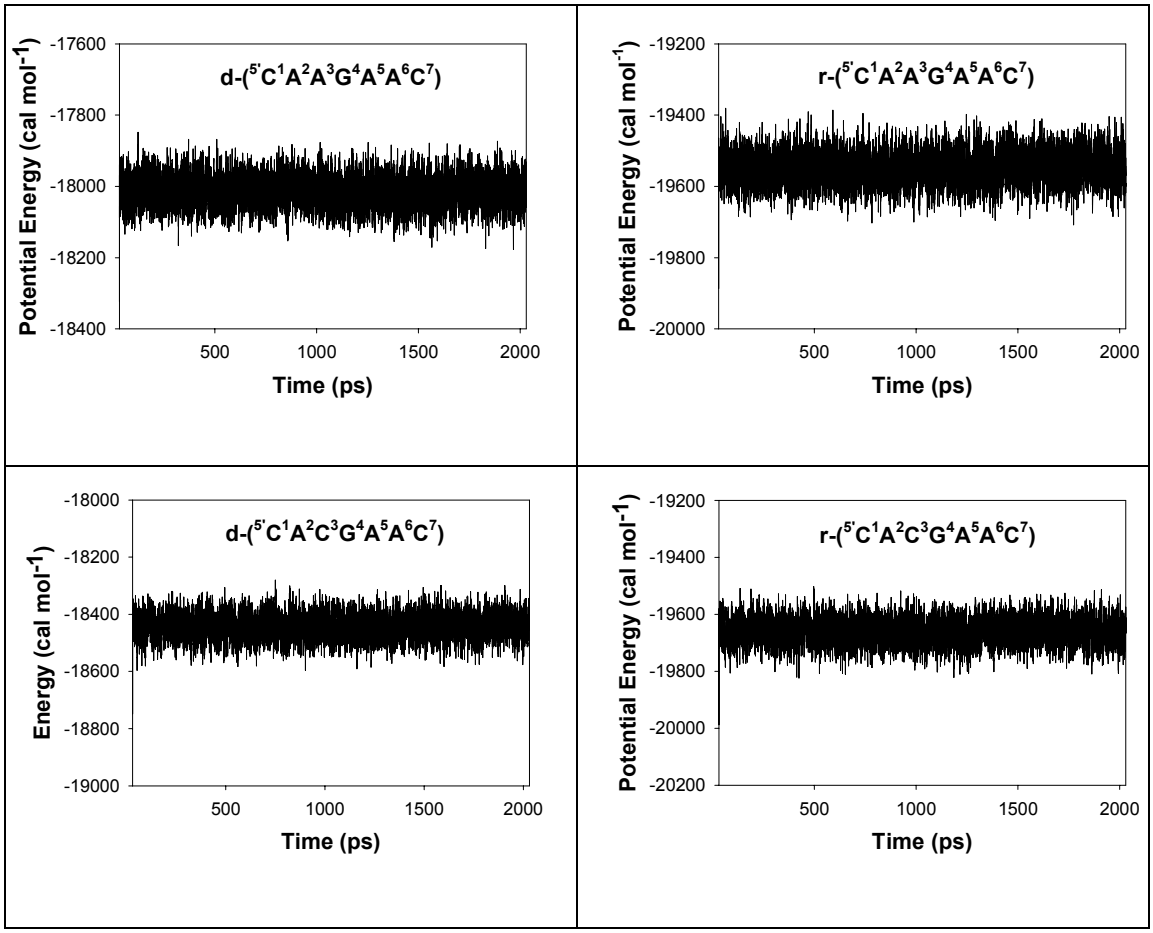


Figure S50. Plots of mass weighted RMSD beetween the ssDNA and ssRNA molecular modelling trajectories and their correponding most stable SA stuctures at 100K.



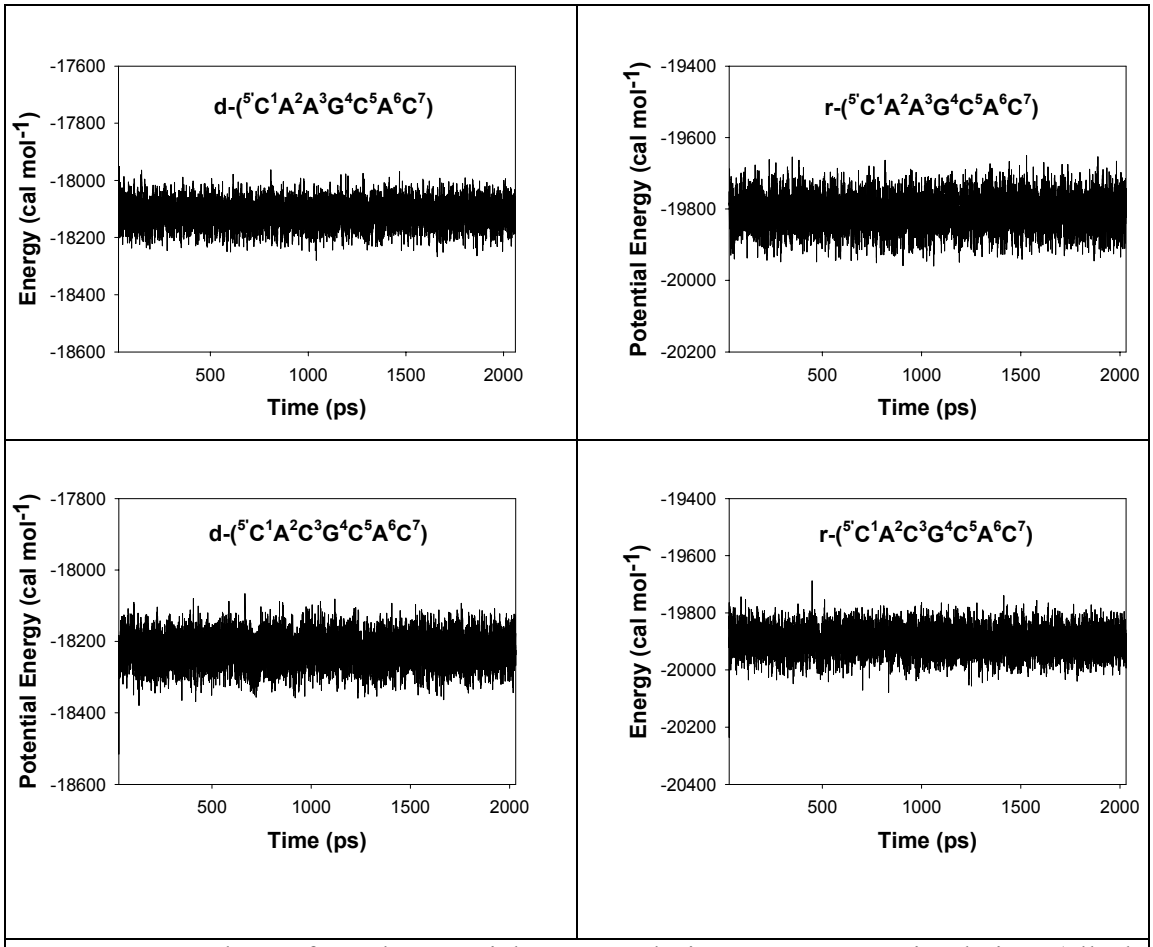
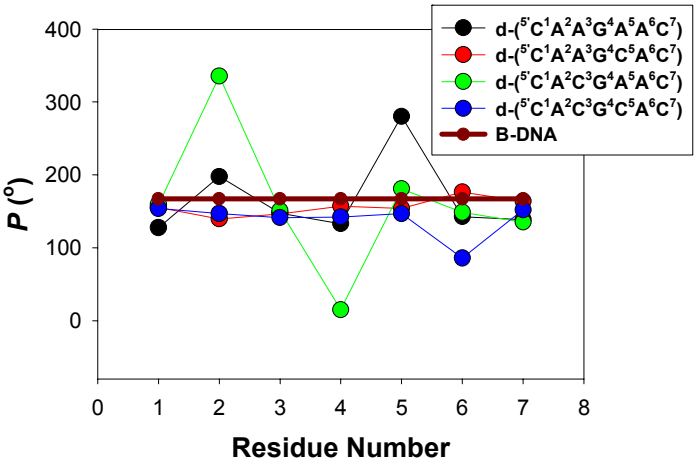
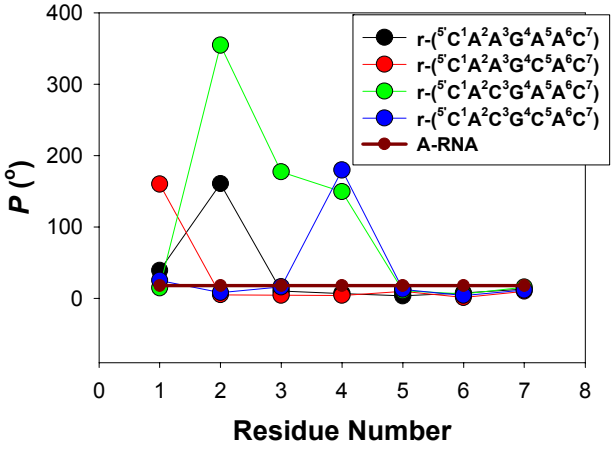
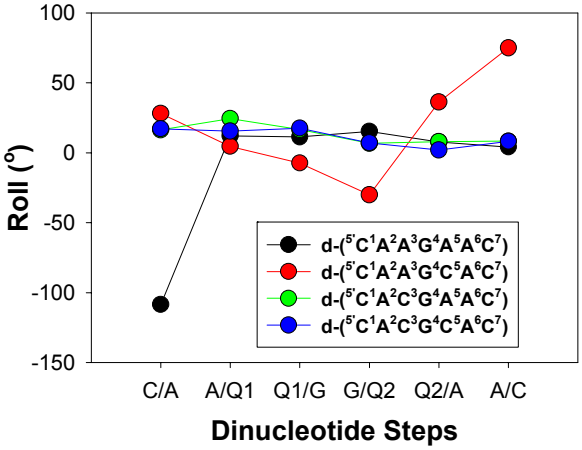
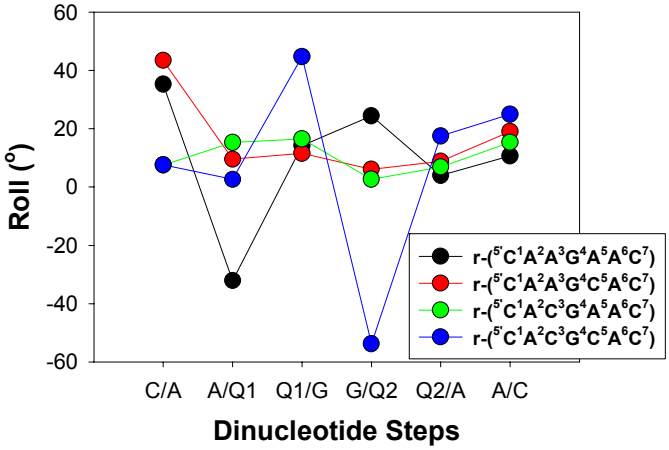


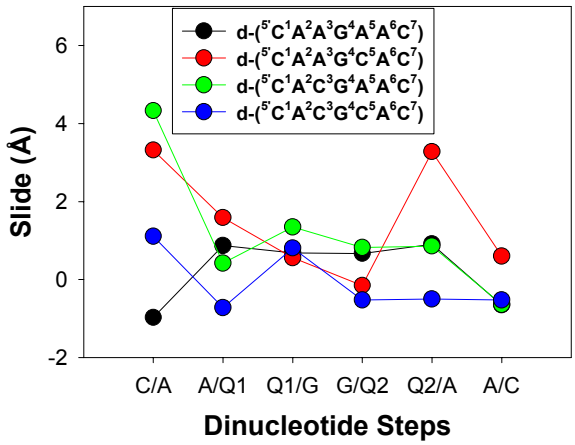
Figure S51. Plots of total potential energy during NMR-MD simulation (all the constraints switched on) steps against the time (ps) for ssDNA and ssRNA.

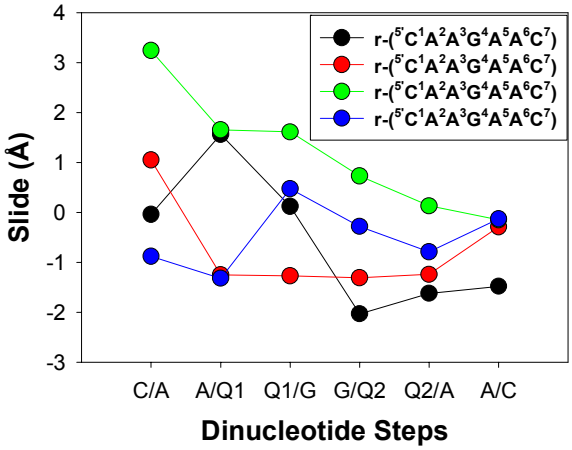












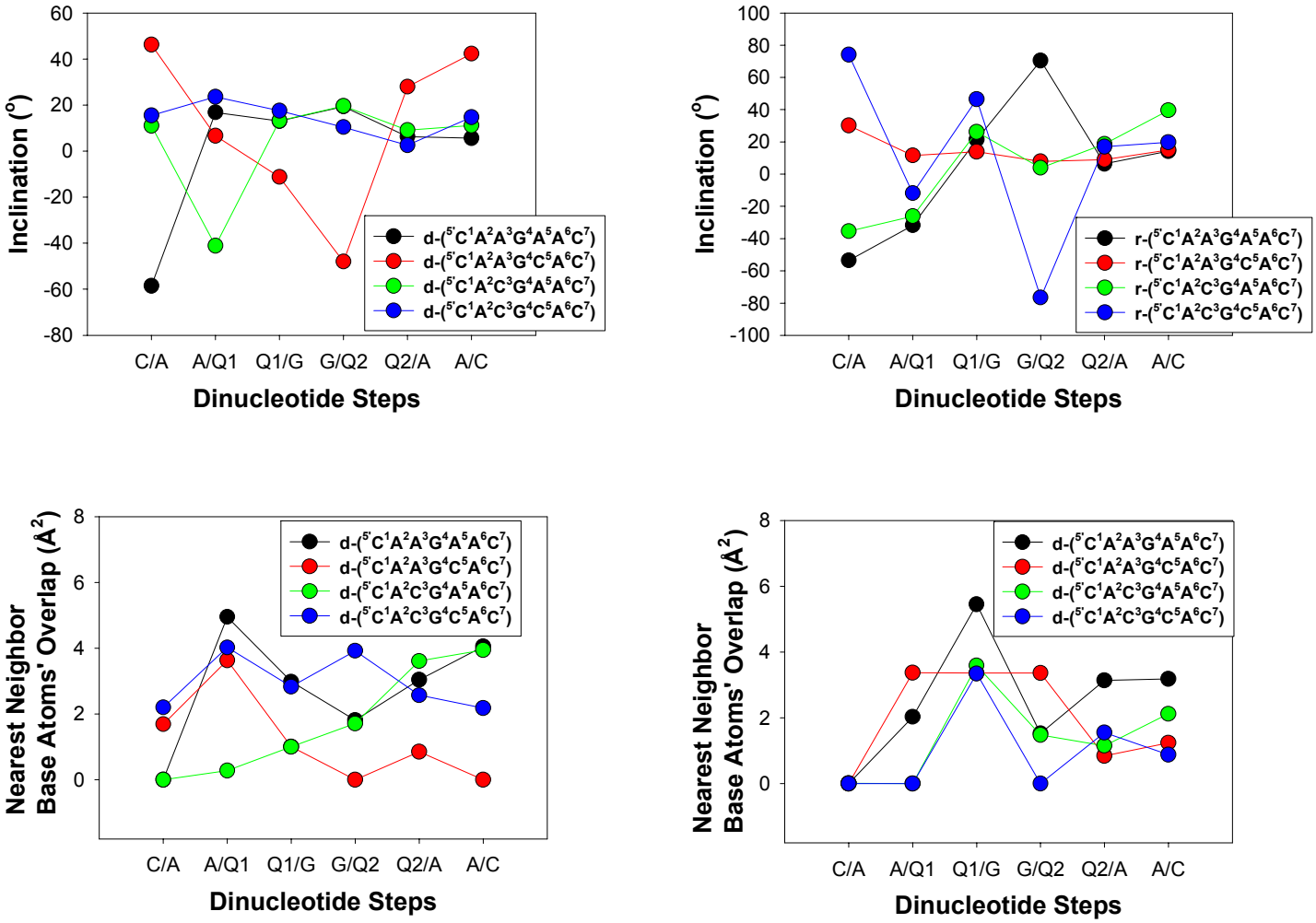
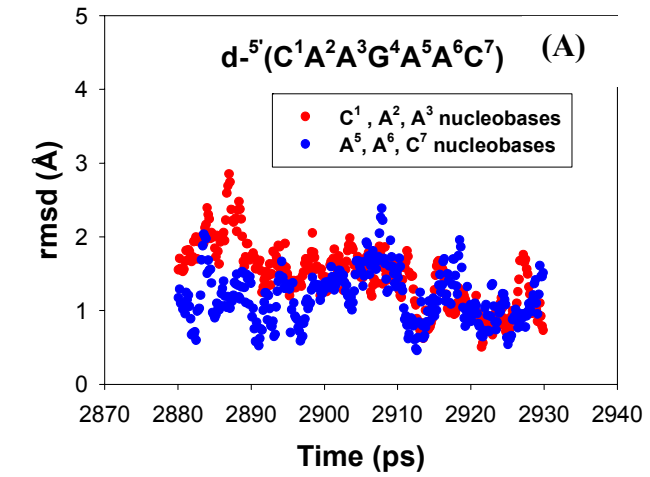
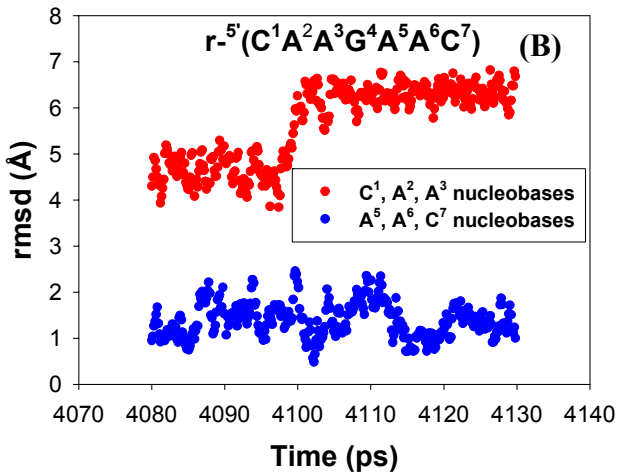
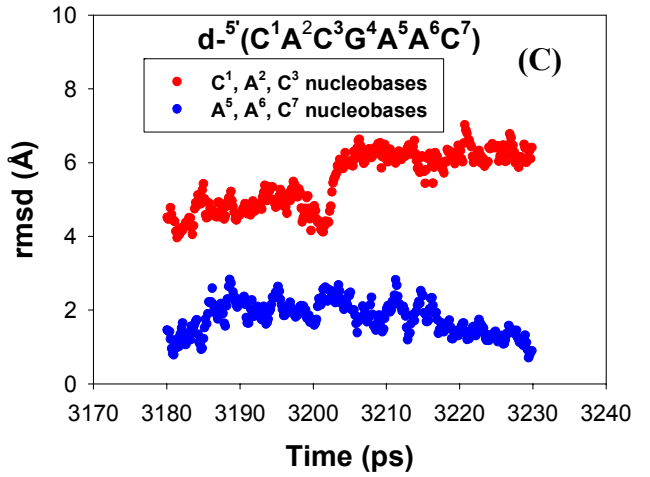
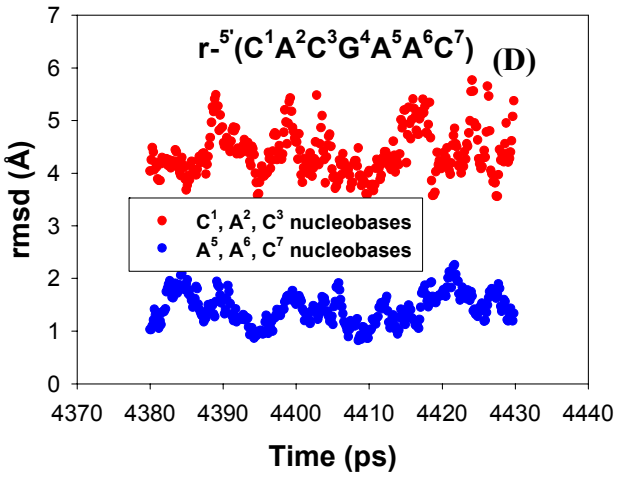


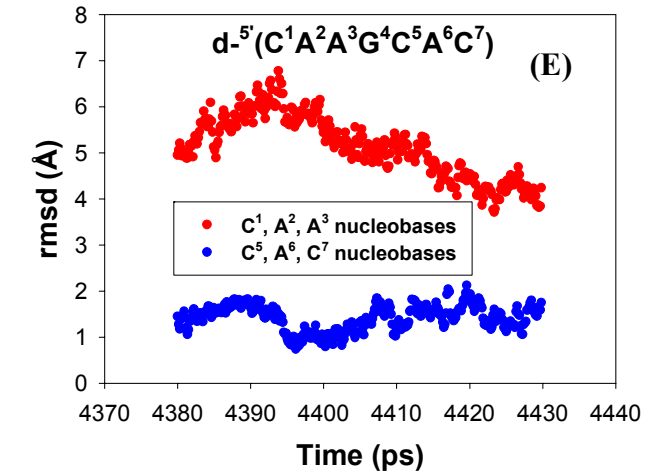
Figure S52. Plots of Sugar puckering (Phase angle, *P*) and helical parameters (Roll, Slide, Inclination, and Nearest neighbour base atom overlap) for the four isosequential ssDNA and ssRNA averaged MD structures.

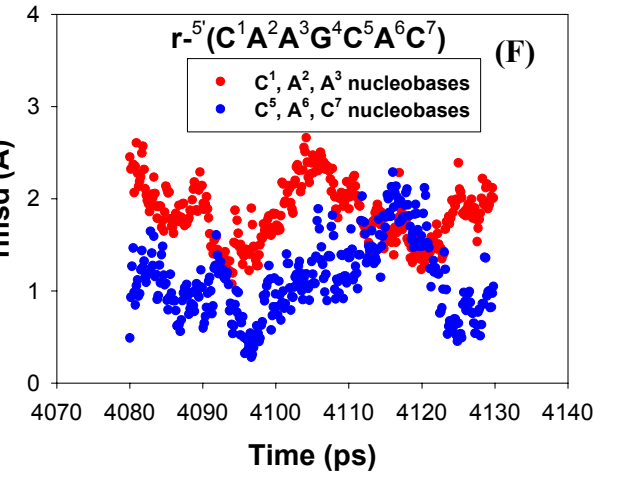












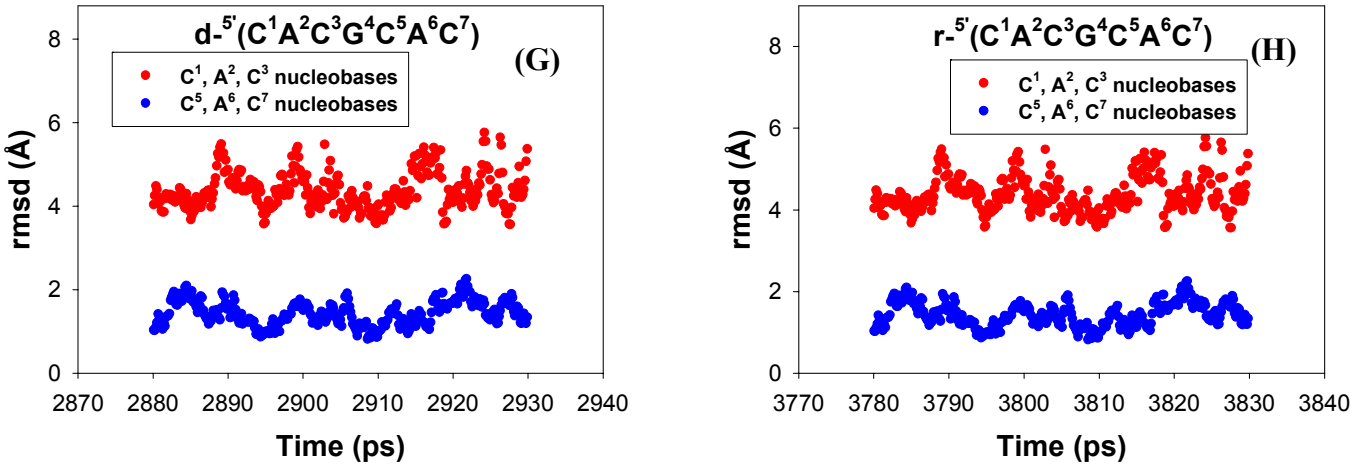


Figure 53. Plots (A- H) of mass weighted RMSD for (1-3) [red]and (5-7) [blue] nucleobase residues in ssDNAs (**1** - **4**) and ssRNAs (**5** – **8**) with Time (ps) shows that 5′-ends are more dynamic compared to 3′-ends in ssDNAs and ssRNAs. All RMSD for nucleobase residues were calculated referencing the final SA structures of ssDNAs (**1** - **4**) and ssRNAs (**5** – **8**) at 100 K.