H-Bond Activated Glycosylation of Nucleobases: Implications for Prebiotic Nucleoside Synthesis

Palwinder Singh*[a], Amrinder Singh[a], Jagroop Kaur[a], Wolgang Holzer[b]

[a] Department of Chemistry, Guru Nanak Dev University, Amritsar-143005. India

[b]Department of Drug and Natural Product Synthesis, University of Vienna, Althanstrasse, 14. A-1069, Wein, Austria

Electronic Supporting Information

Table S1. Composition of water used as reaction medium.

Physical parameter	Water sample – I (ultrapure de-ionized)*	Water sample – II (tap water)
pH	7.1±0.1	7.2±0.1
TDS (mg/L)	<0.1	400±10
TSS (mg/L)	NIL	NIL
Cond. (µs)	2.9±0.1	240
Hardness (mg/L)	15±1	270
Calcium (mg/L)	NIL	60±5
Magnesium (mg/L)	NIL	30±5
Iron (mg/L)	NIL	0.02±0.01
Chlorides (mg/L)	NIL	14±2

*purchased from Sigma-Aldrich

Materials. Formic acid, sodium hydroxide, acetonitrile, water, methanol, ethanol and other solvents were purchased from Sigma-Aldrich (St. Louis, MO). Adenine, adenosine, cytosine, uracil, guanine, D-ribose were purchased from Spectrochem Mumbai. Cytidine (Cat No. 03146, CAS No. 65-46-3, corresponding to β-cytidine) was purcahsed from Loba Chemie, Mumbai.

Mass Spectrometry. All mass spectra were recorded on Bruker MicroTOF QII mass spectrometer (Bruker Daltonik, Bremen, Germany). 50 µM Solution of nucleobase and D-ribose in ACN: H_2O (3:7) was injected through syringe with flow rate 180 µL/h set with KdScientific automated pump. Desolvation was performed with dry N_2 gas heated at 180 ^oC. Various parameters of mass spectrometer were optimized for maximum ion abundance of molecular ion peak. Typically, the capillary voltage was 4500 V and vacuum was maintained at 3-4x10^{-7} mbar.

For LC-MS, the Dionex Ultimate 3000 system was linked to mass spectrometer. Injection of 2 µL of the sample was given to C-18 column (Acclaim® 120 C18 5 µm 120 Å (4.6 x 250 mm) and methanol-water (2:8) was used as eluent. The flow rate was kept 0.2 ml and absorbance was read at 260 nm.

NMR spectra were recorded on Bruker 400 MHz NMR spectrometer. IR spectra were recorded on Perkin Elmer FTIR-C92035 Fourier Transform spectrometer. Optical rotation was recorded on AT-100 Atago automatic polarimeter. UV-vis spectra were recorded on BioteK UV-vis spectrophotometer.

General procedure for laboratory reactions. The reactants (nucleobase and sugar) were dissolved in C_2H_5OH:H_2O (1:9) (de-ionized water as well as tap water was used for different experiments) and heated at 60-70 ^oC for 8 days. At higher temperature, reaction mixture turns black and no product was isolated. Reaction was monitored with TLC and ESI-MS. After 8 days of the reaction, solvent was evaporated in vaucco and the solid residue was washed with ethanol. The solid mass was column chromatographed using methanol:chloroform (15:85) as eluent to obtain pure product **1** (15%) and **2** (12%).

Synthesis of adenosine 1. Adenine (1 g) and D-ribose (1.10 g) were dissolved in C_2H_5OH:H_2O (1:9) (25 ml) and heated at 60-70 ^oC for 8 days. Solvent was evaporated in vaucco and the solid residue was washed with ethanol. The solid mass was column chromatographed using methanol:chloroform (5:95) as eluent to obtain 290 mg pure adenosine **1** (15%). 700 mg adenine and 500 mg D-ribose were also recovered.

Synthesis of Cytidine 2. Cytosine (1 g) and D-ribose (1.32 g) were dissolved in C_2H_5OH:H_2O (1:9) (25 ml) and heated at 60-70 ^oC for 8 days. Solvent was evaporated in vaucco and the solid residue was washed with ethanol. The solid mass was column chromatographed using methanol:chloroform (10:90) as eluent to obtain 250 mg pure cytidine **2** (12%). 780 mg cytosine and 600 mg D-ribose were also recovered.

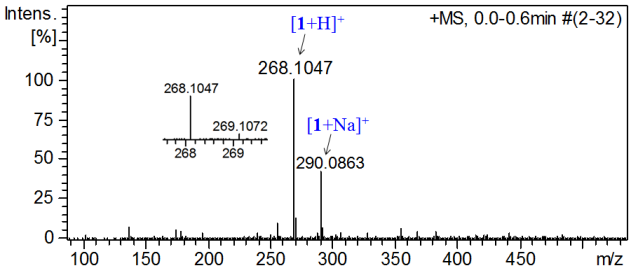


Figure S1. High resolution mass spectrum of product **1** indicating the formation of adenosine (calcd m/z 268.1040 $[M+H]^+$). Inset: expansion to show the isotopic pattern of molecular ion peak.

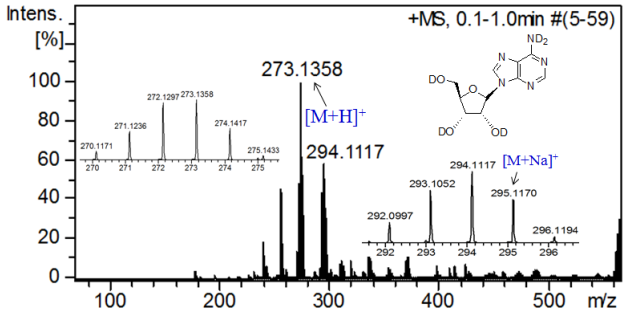


Figure S2. ESI-MS of adenosine **1** in D_2O-ACN (7:3). The peak at *m/z* 273.1358 corresponds to $[M+H]^+$ of adenosine after replacement of all the exchangeable Hs' with D (calcd *m/z* 273.1354, $[M+H]^+$) while *m/z* 274.1417 corresponds to $[M+D]^+$. Inset: deuterated adenosine and expansions of mass peaks with *m/z* 273 and 294.

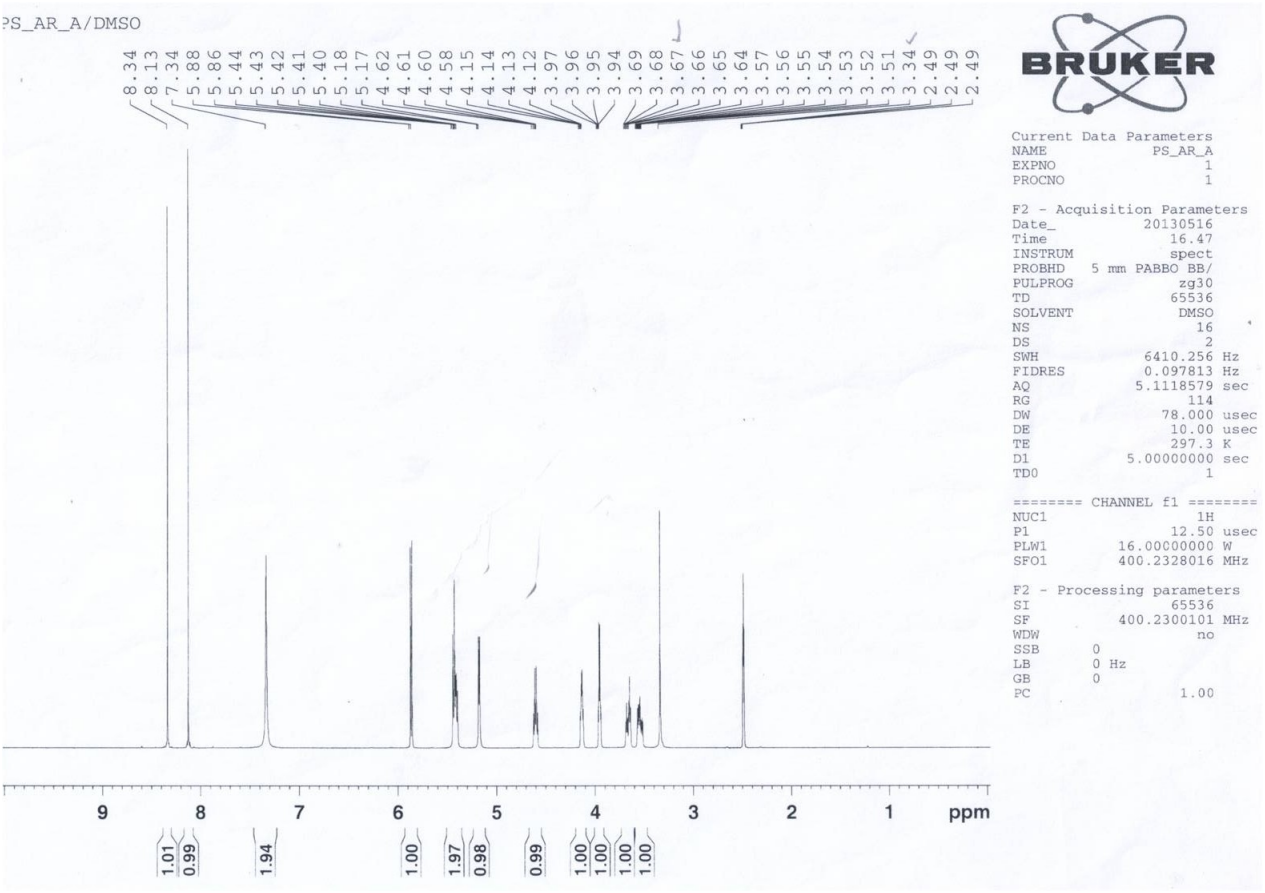


Figure S3. ^{1}H NMR spectrum of compound **1** in DMSO-d_6.

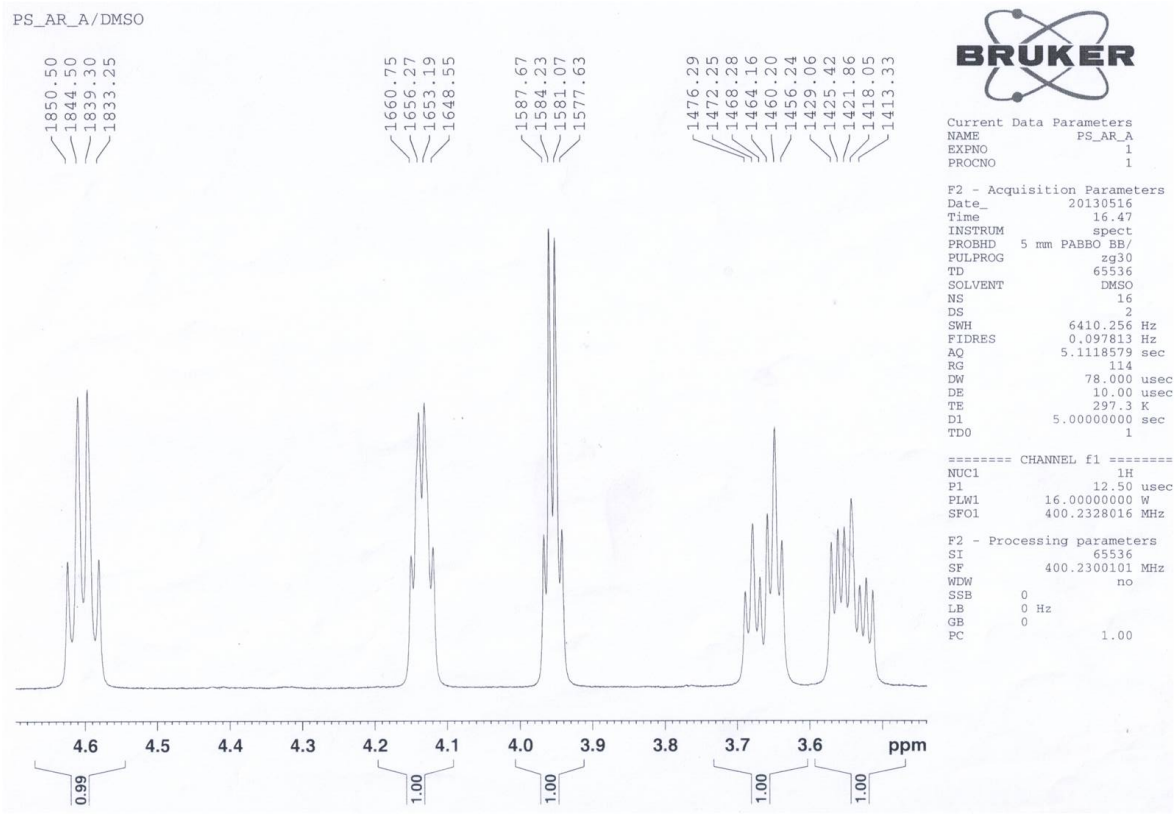


Figure S4. ^{1}H NMR spectrum of compound **1** showing expansion of region 3.5 to 4.6 ppm.

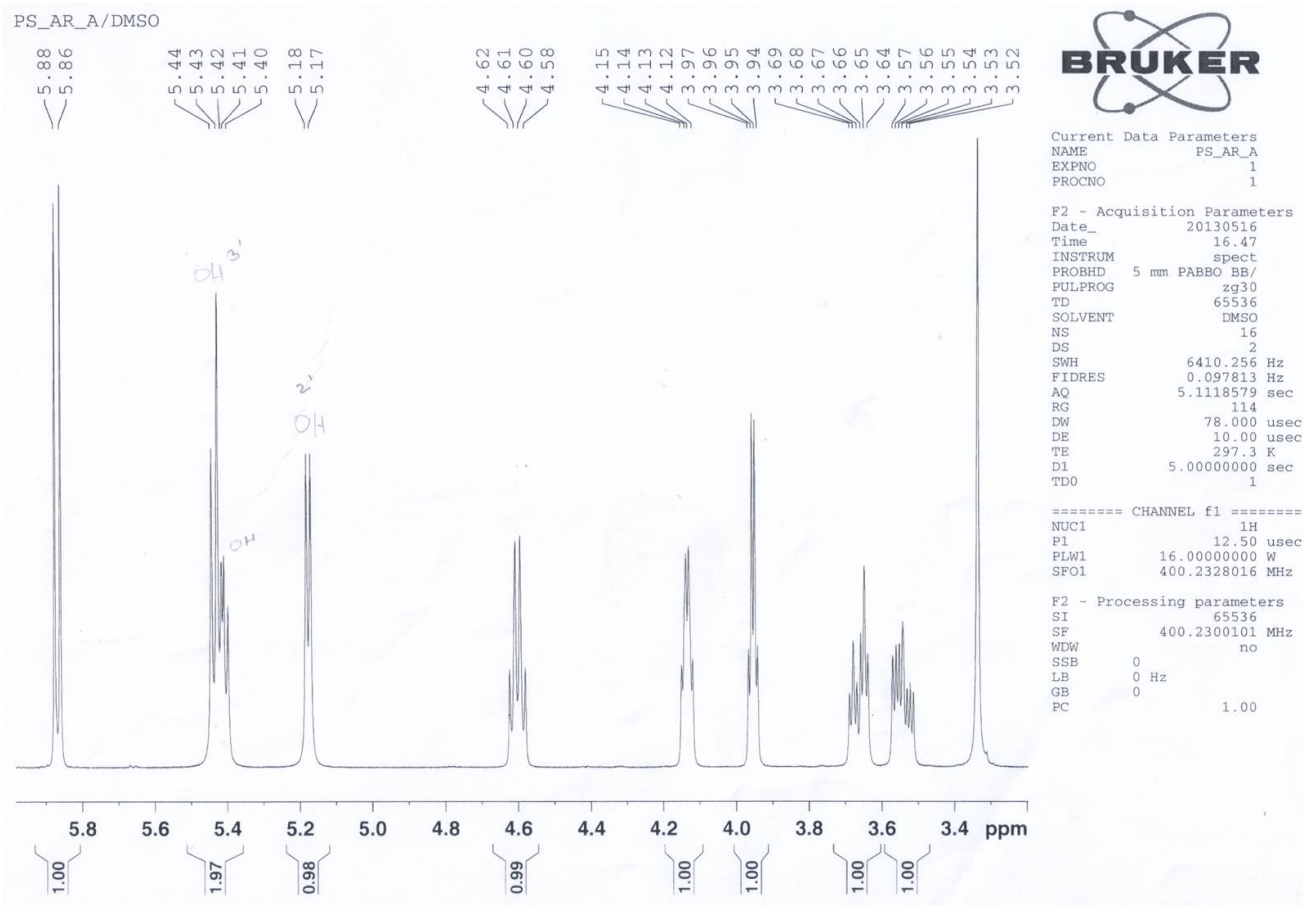


Figure S5. [1]H NMR spectrum of compound **1** showing expansion of region 3.4 to 5.8 ppm.

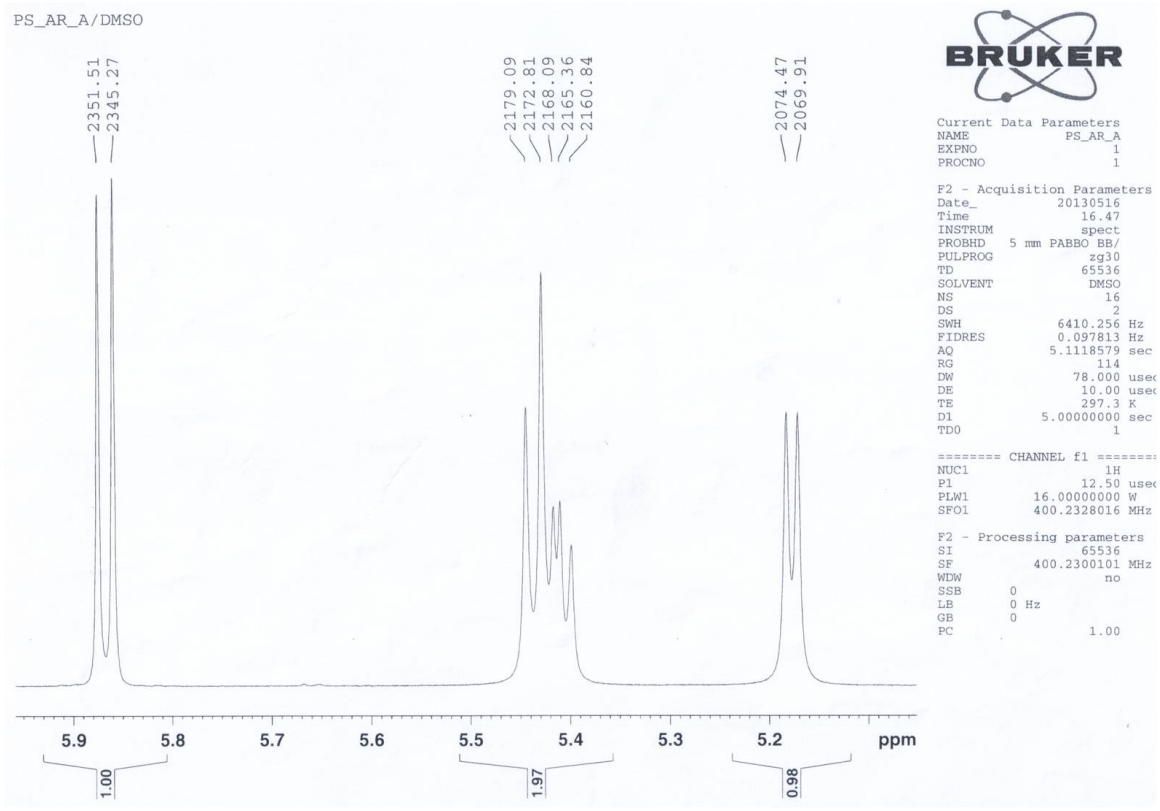


Figure S6. [1]H NMR spectrum of compound **1** showing expansion of region 5.2 to 6 ppm.

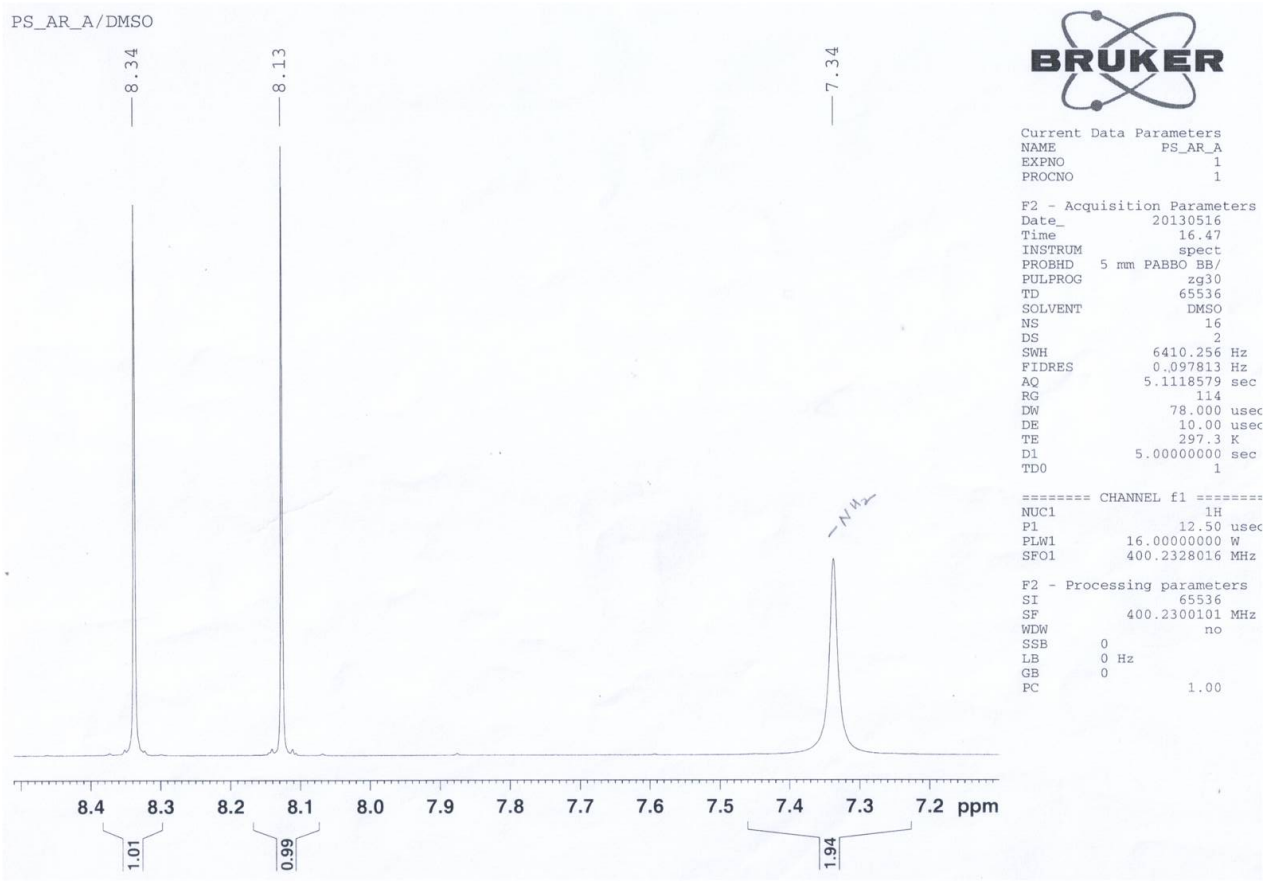


Figure S7. ^{1}H NMR spectrum of compound **1** showing expansion of region 7.2 to 8.4 ppm

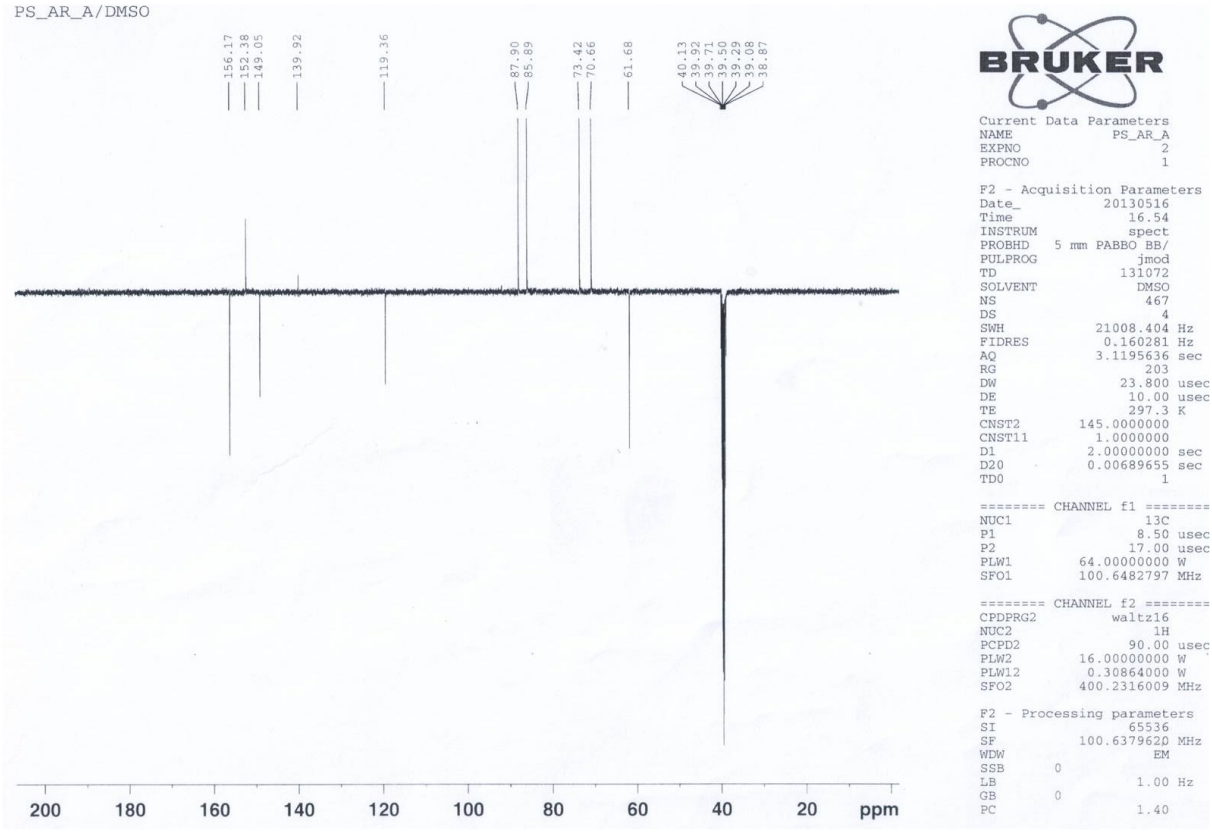


Figure S8. ^{13}C NMR spectrum of compound **1** in DMSO-d_6.

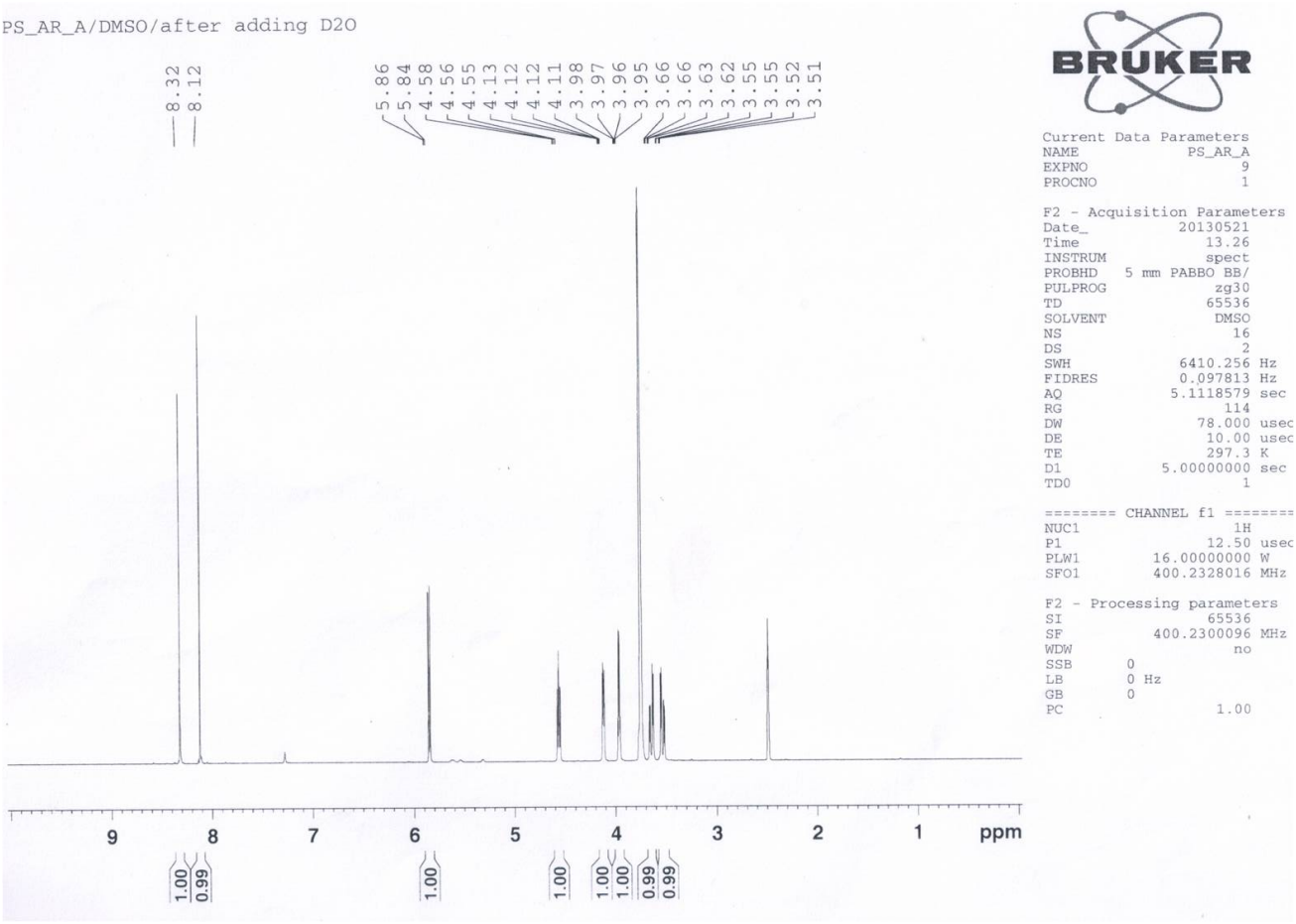


Figure S9. ^{1}H NMR spectrum of compound **1** after adding D_2O.

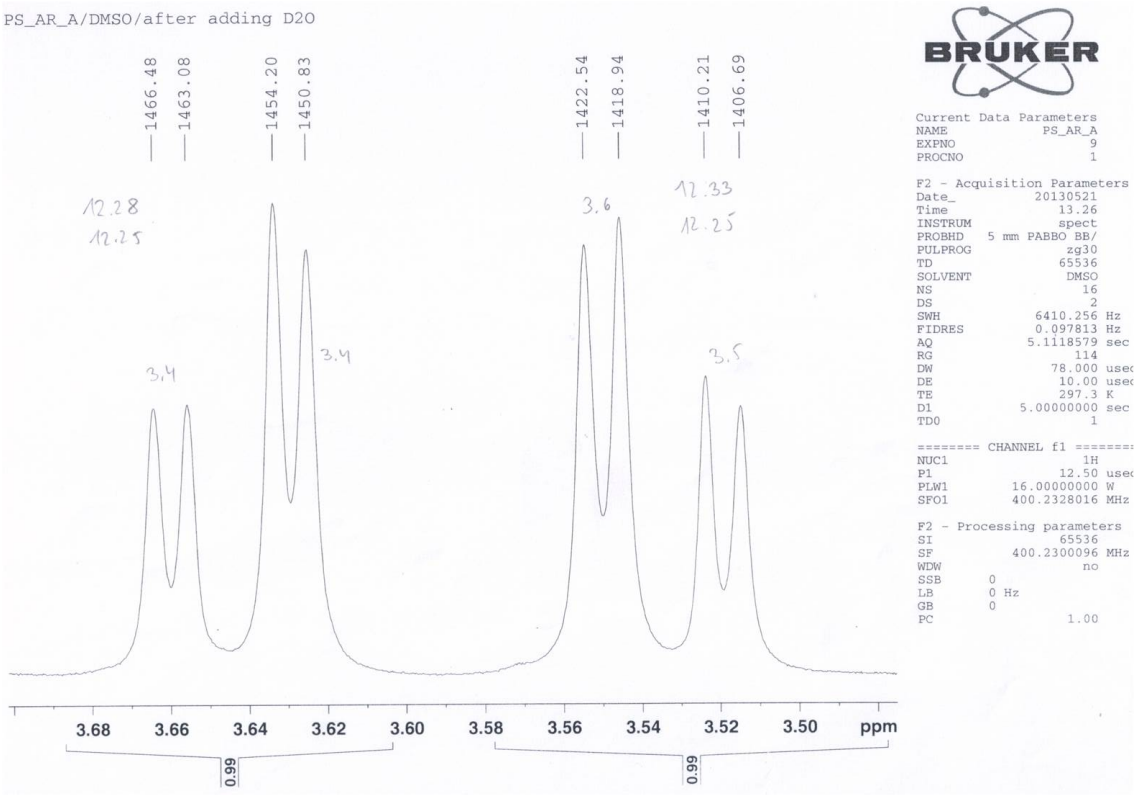


Figure S10. ^{1}H NMR spectrum of compound **1** after adding D_2O showing expansion of region 3.5 to 3.7 ppm.

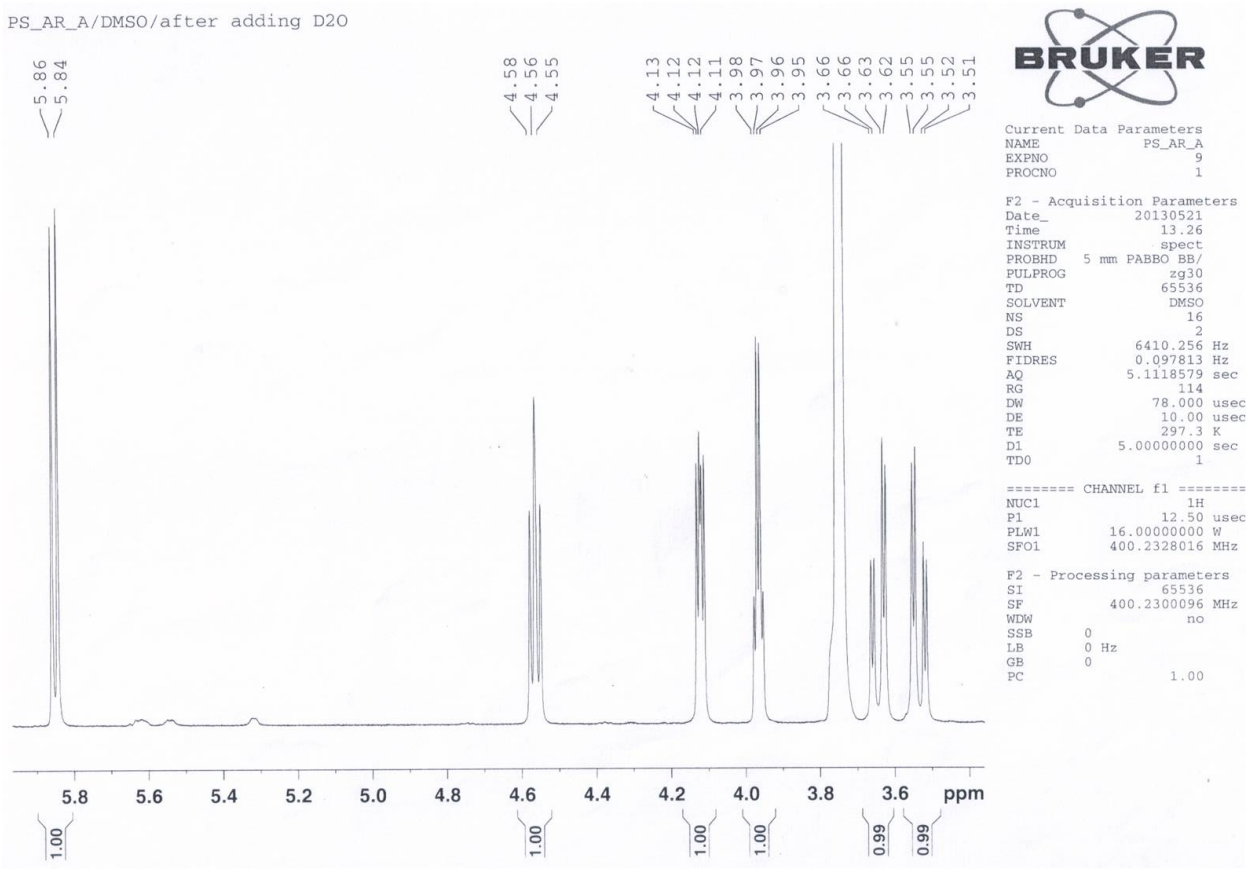


Figure S11. ^{1}H NMR spectrum of compound **1** after adding D_2O showing expansion of region 3.6 to 5.8 ppm.

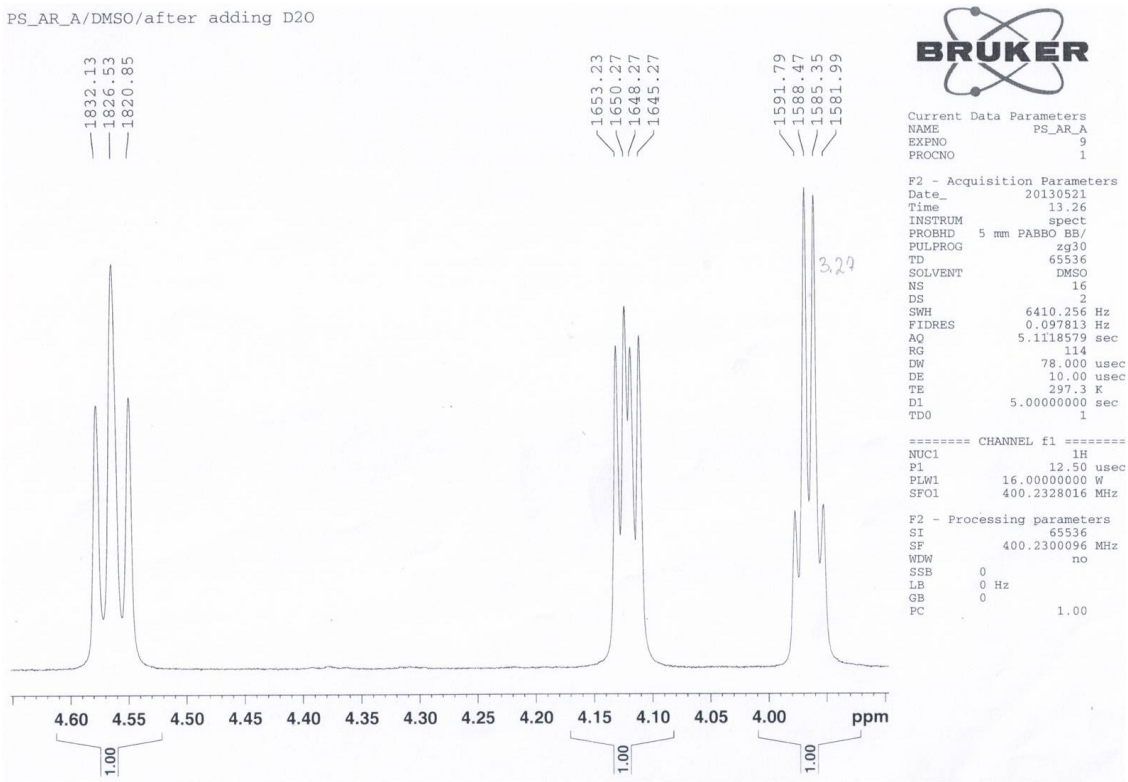


Figure S12. ^{1}H NMR spectrum of compound **1** after adding D_2O showing expansion of region 4 to 6 ppm.

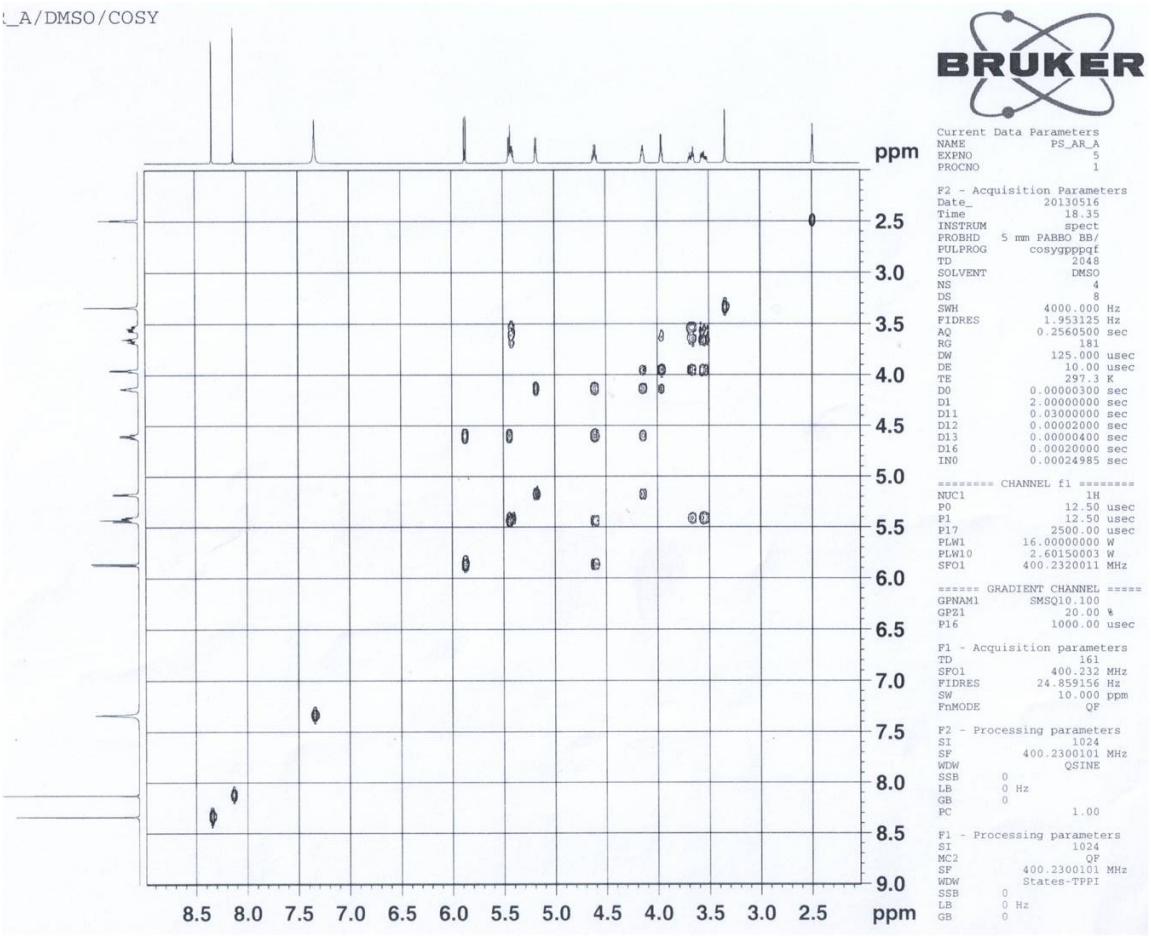


Figure S13. ^{1}H-^{1}H COSY NMR spectrum of compound **1**.

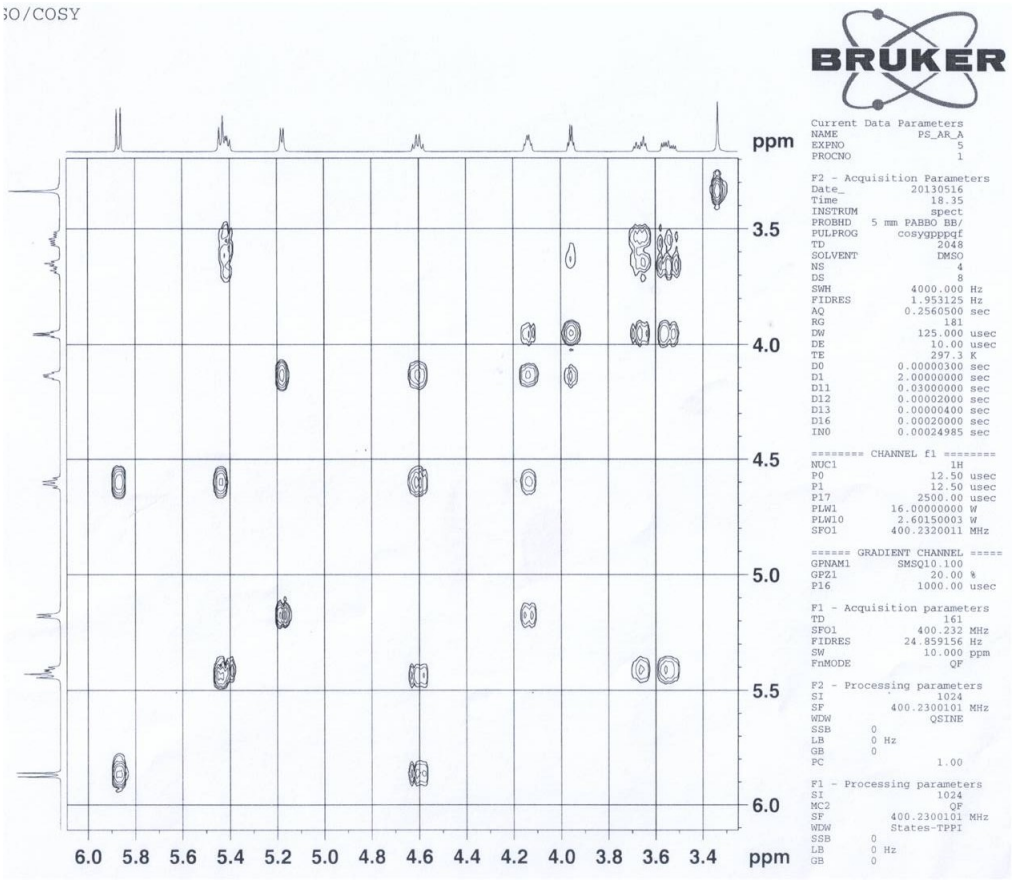


Figure S14. Expansion of spectrum of figure 12.

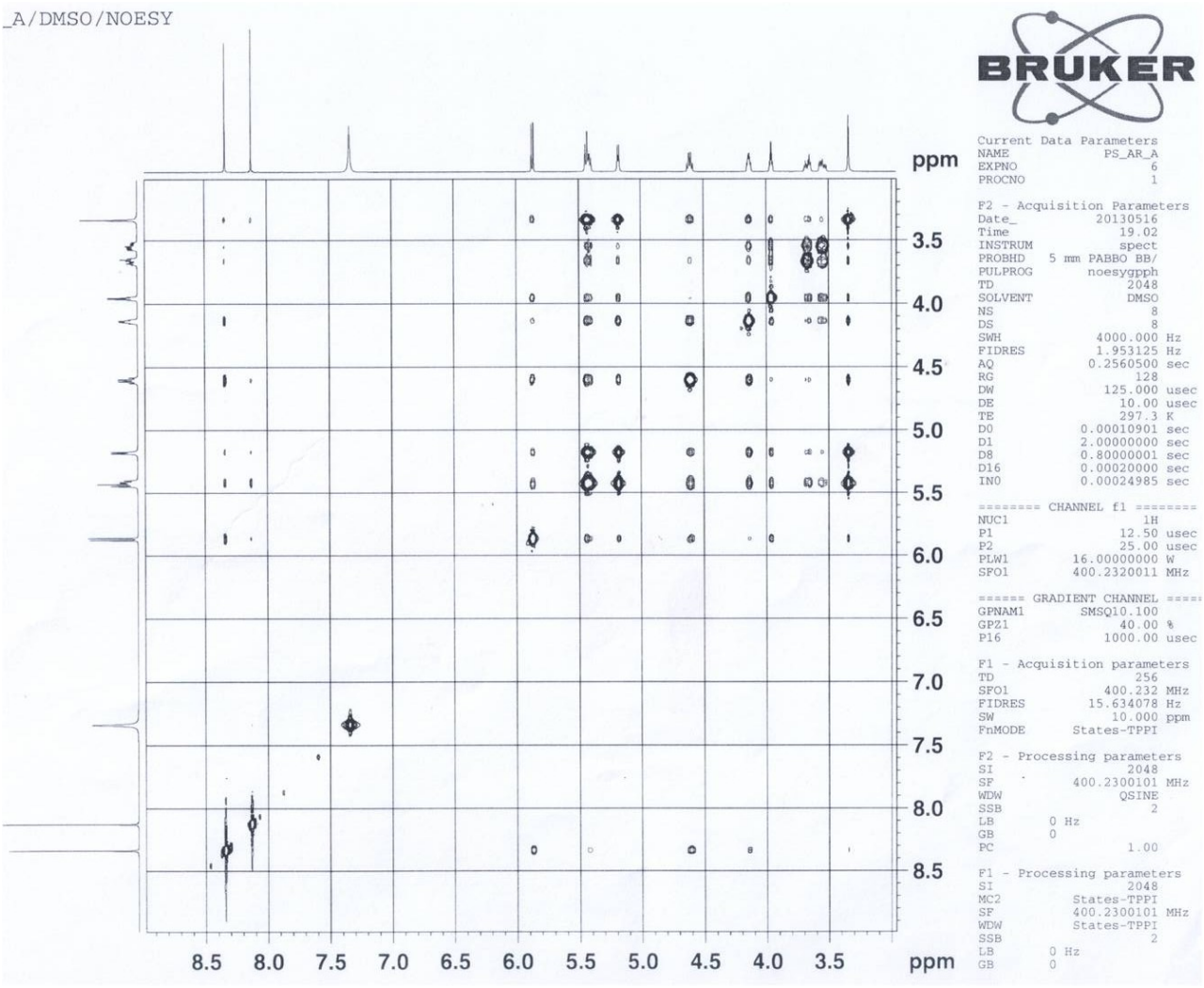


Figure S15. ^{1}H-^{1}H NOESY NMR spectrum of compound **1**.

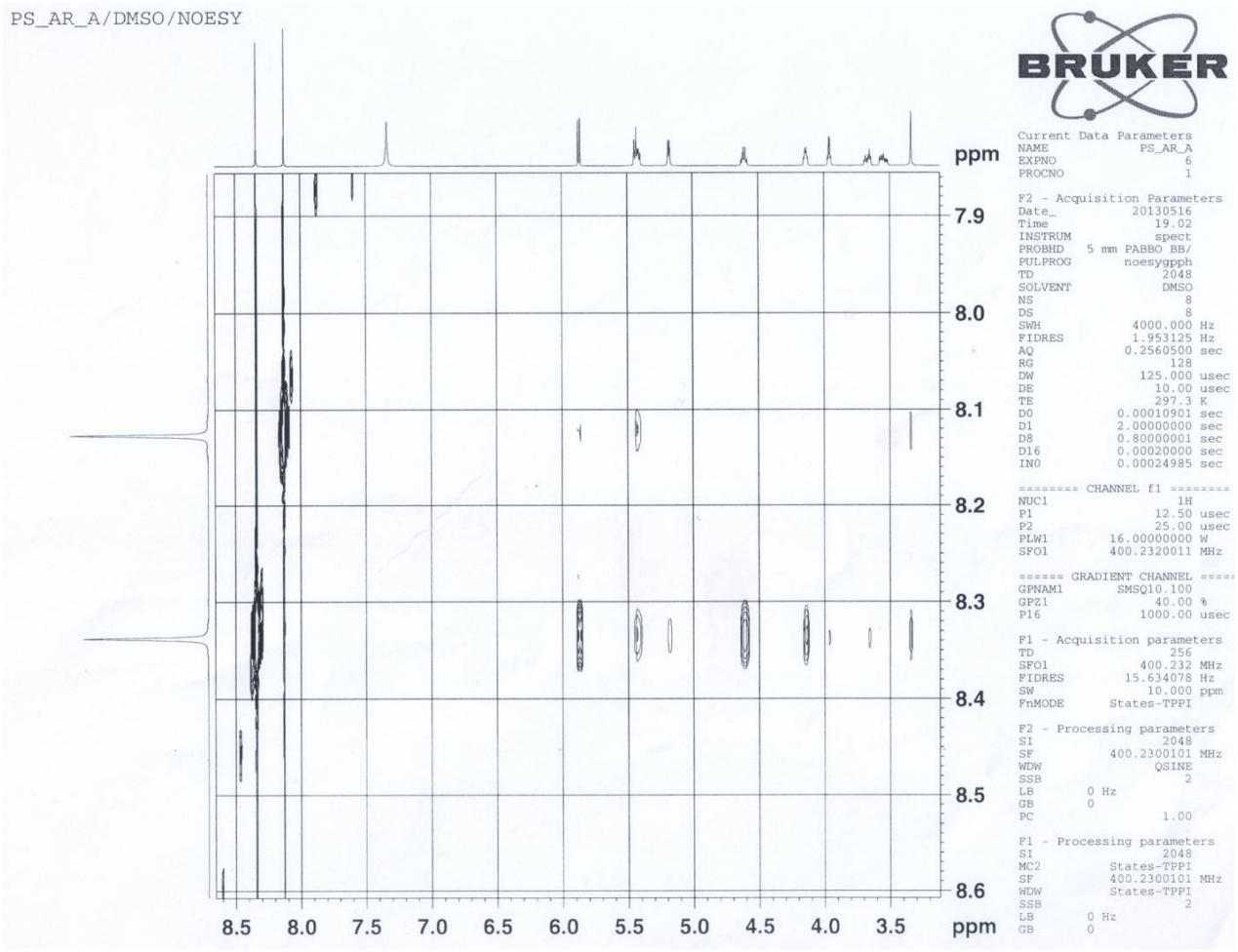


Figure S16. Expansion of spectrum of figure 14.

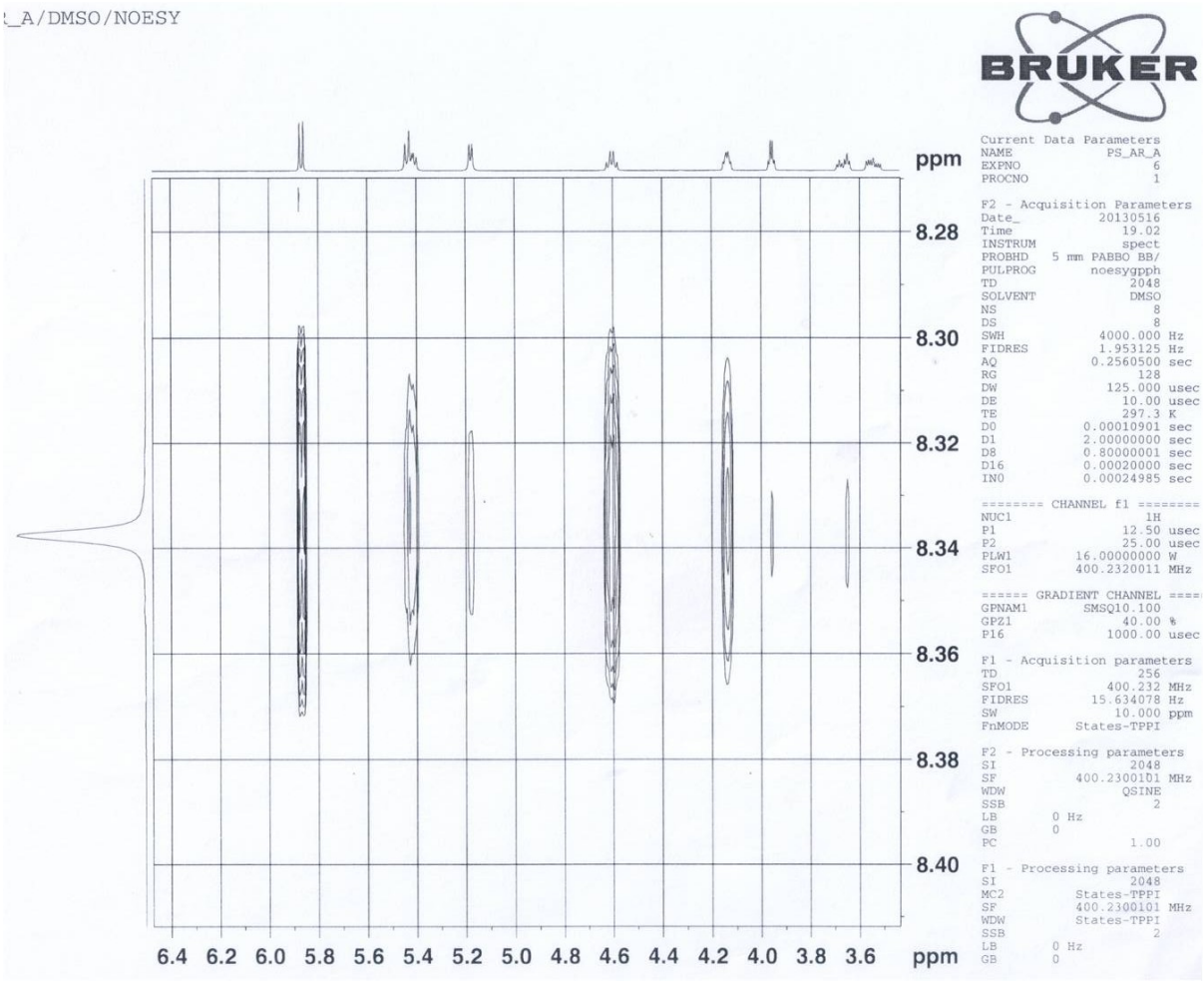


Figure S17. Expansion of spectrum of figure 14.

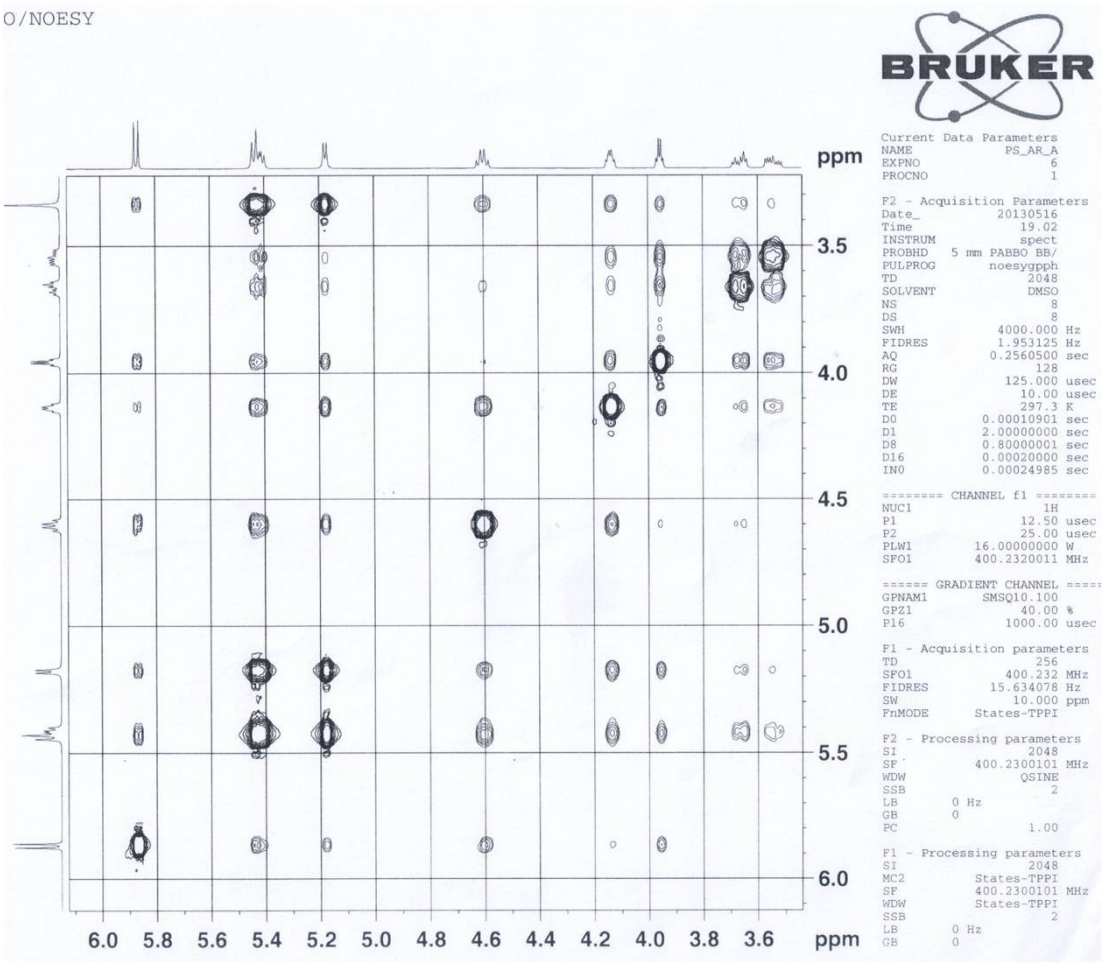


Figure S18. Expansion of spectrum of figure 14.

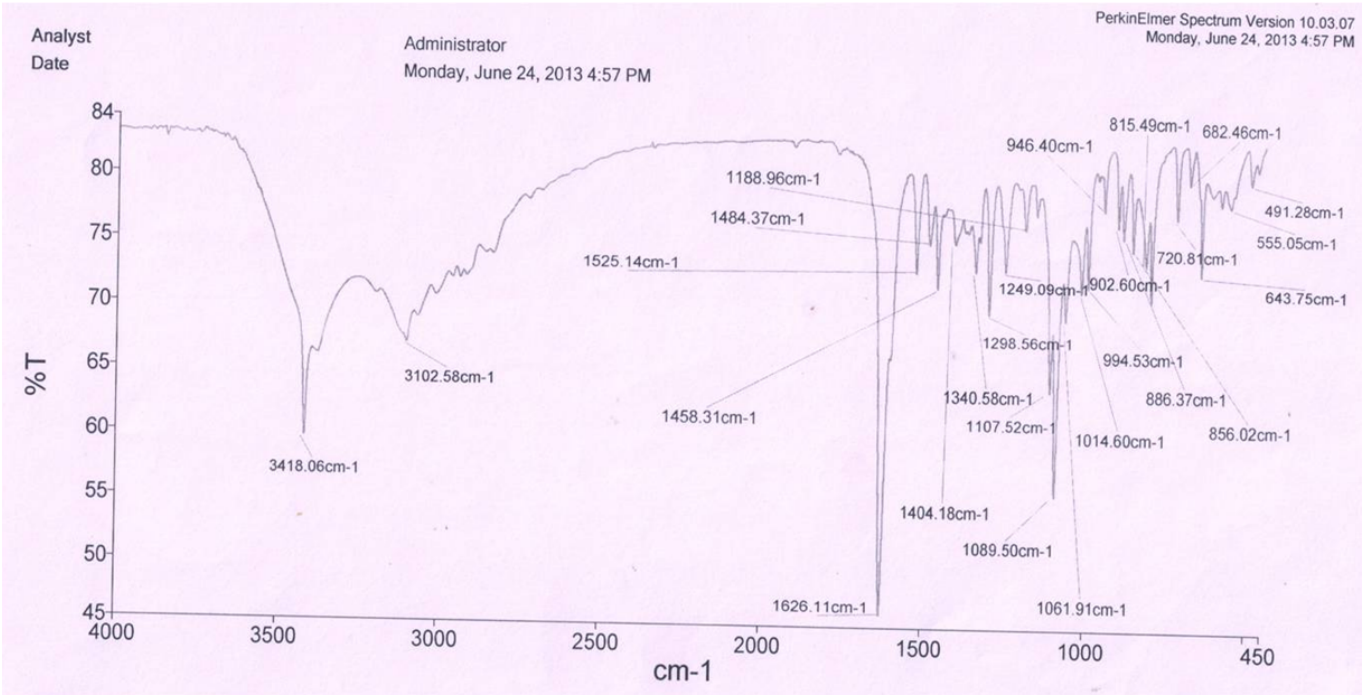


Figure S19. IR spectrum of compound **1**

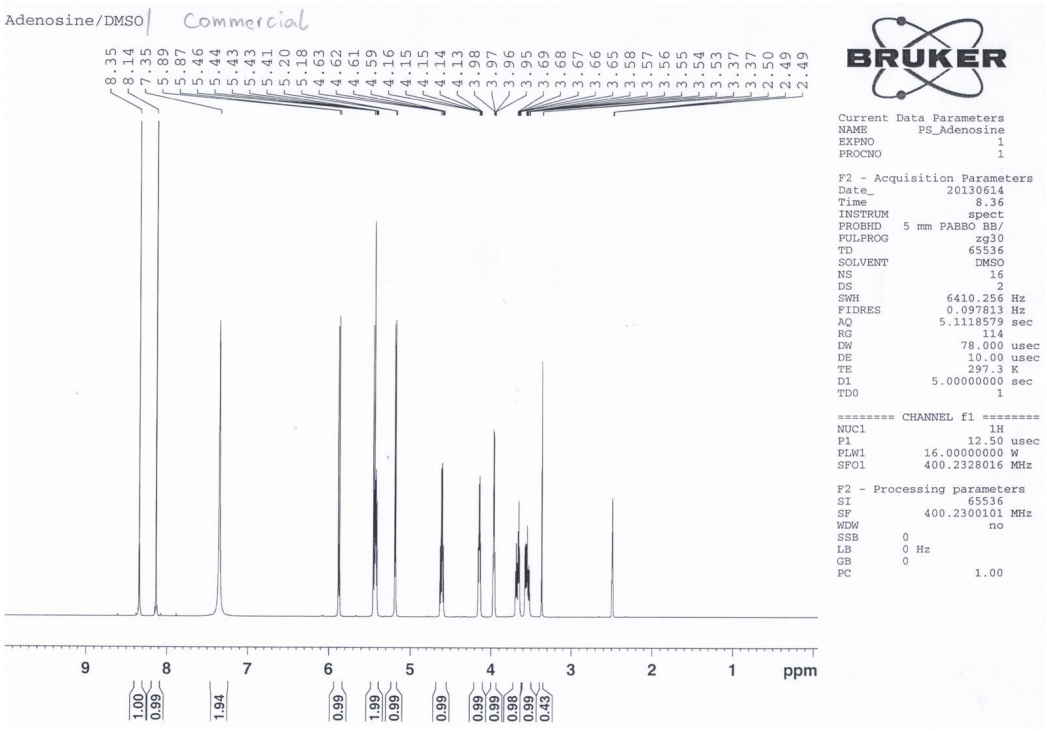


Figure S20. ^{1}H NMR spectrum of commercial sample of adenosine.

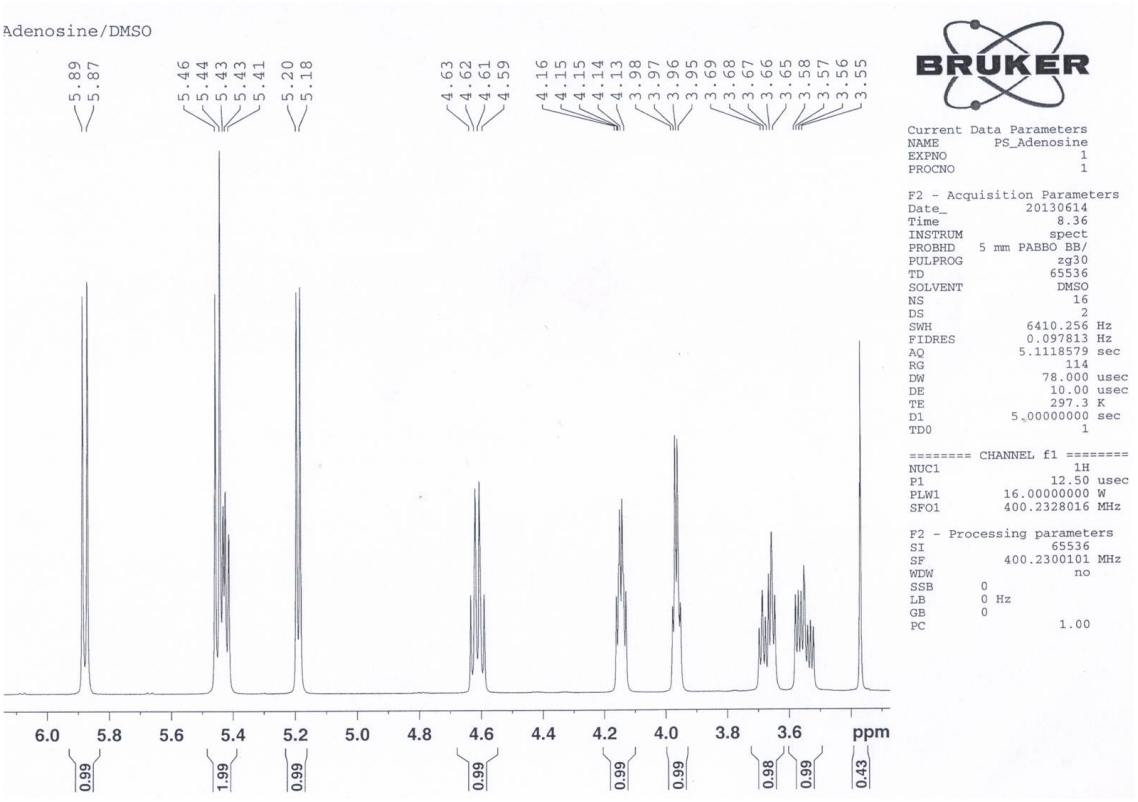


Figure S21. ^{1}H NMR spectrum of commercial sample of adenosine (expansion).

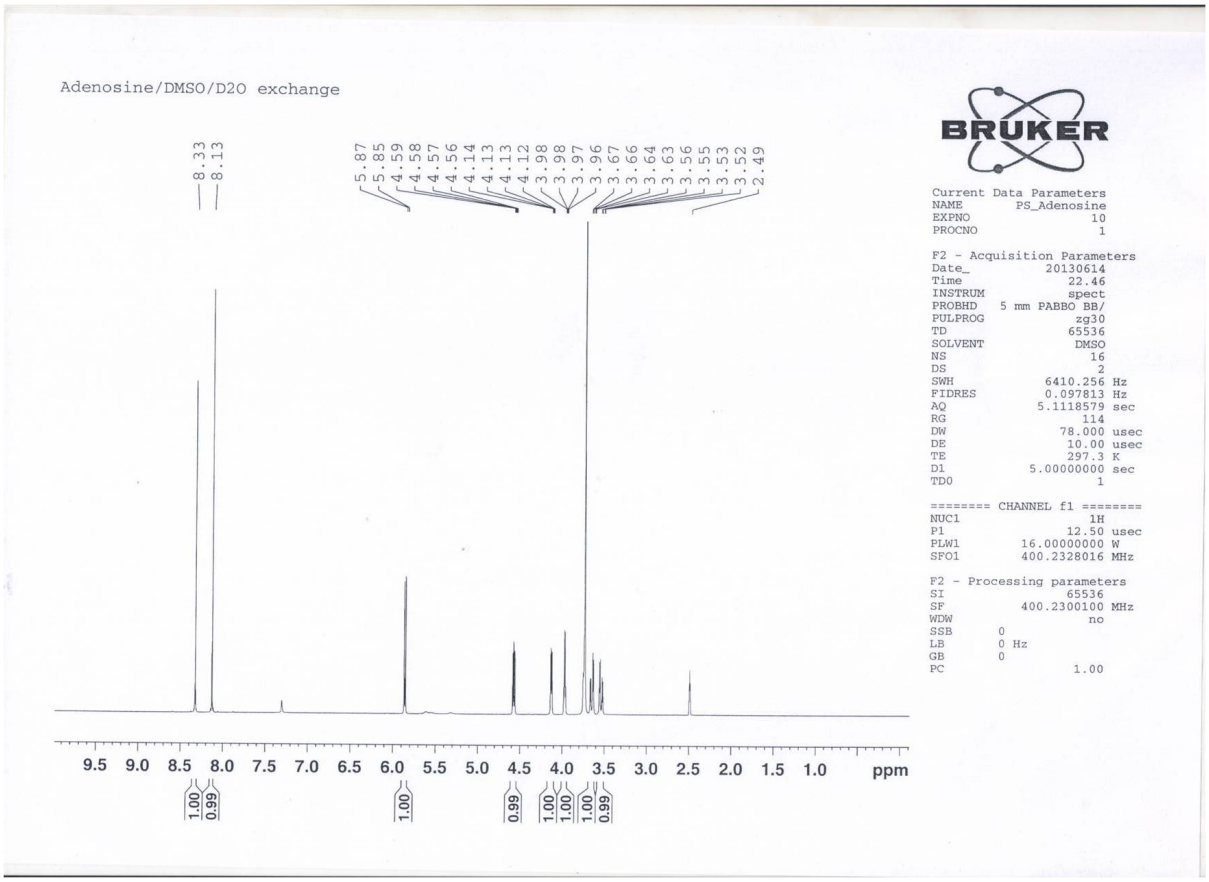


Figure S22. ^{1}H NMR spectrum of commercial sample of adenosine, D_2O exchange.

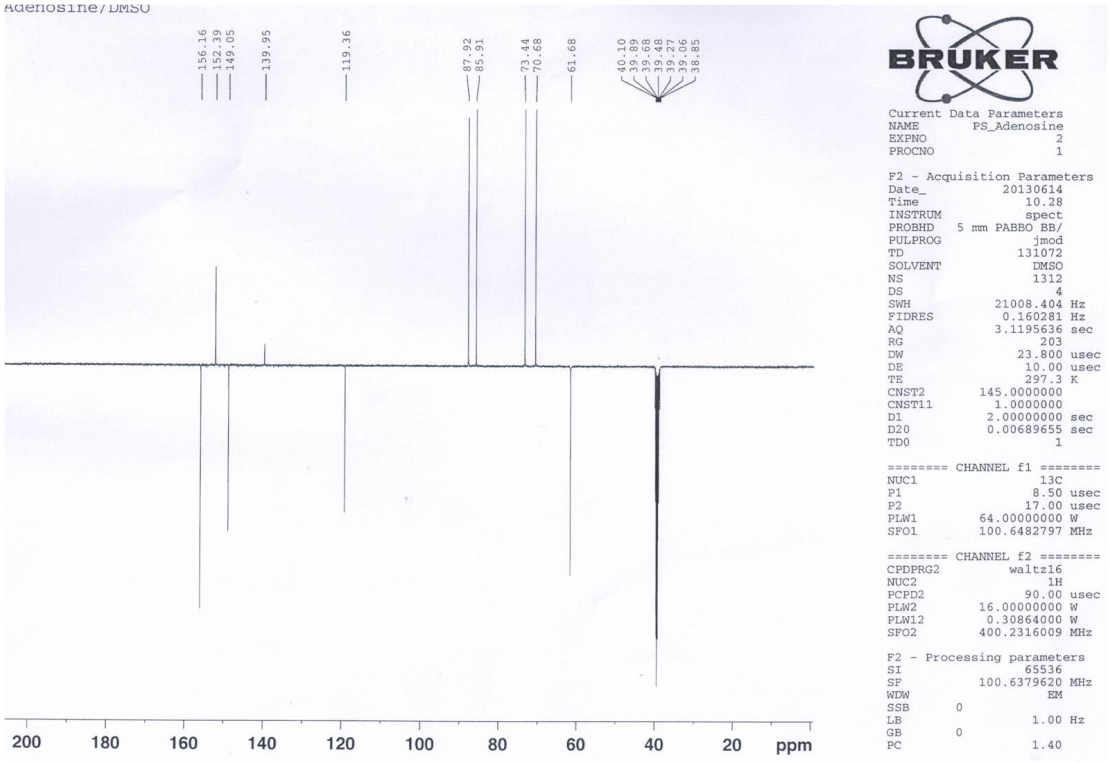


Figure S23. ^{13}C NMR spectrum of commercial sample of adenosine.

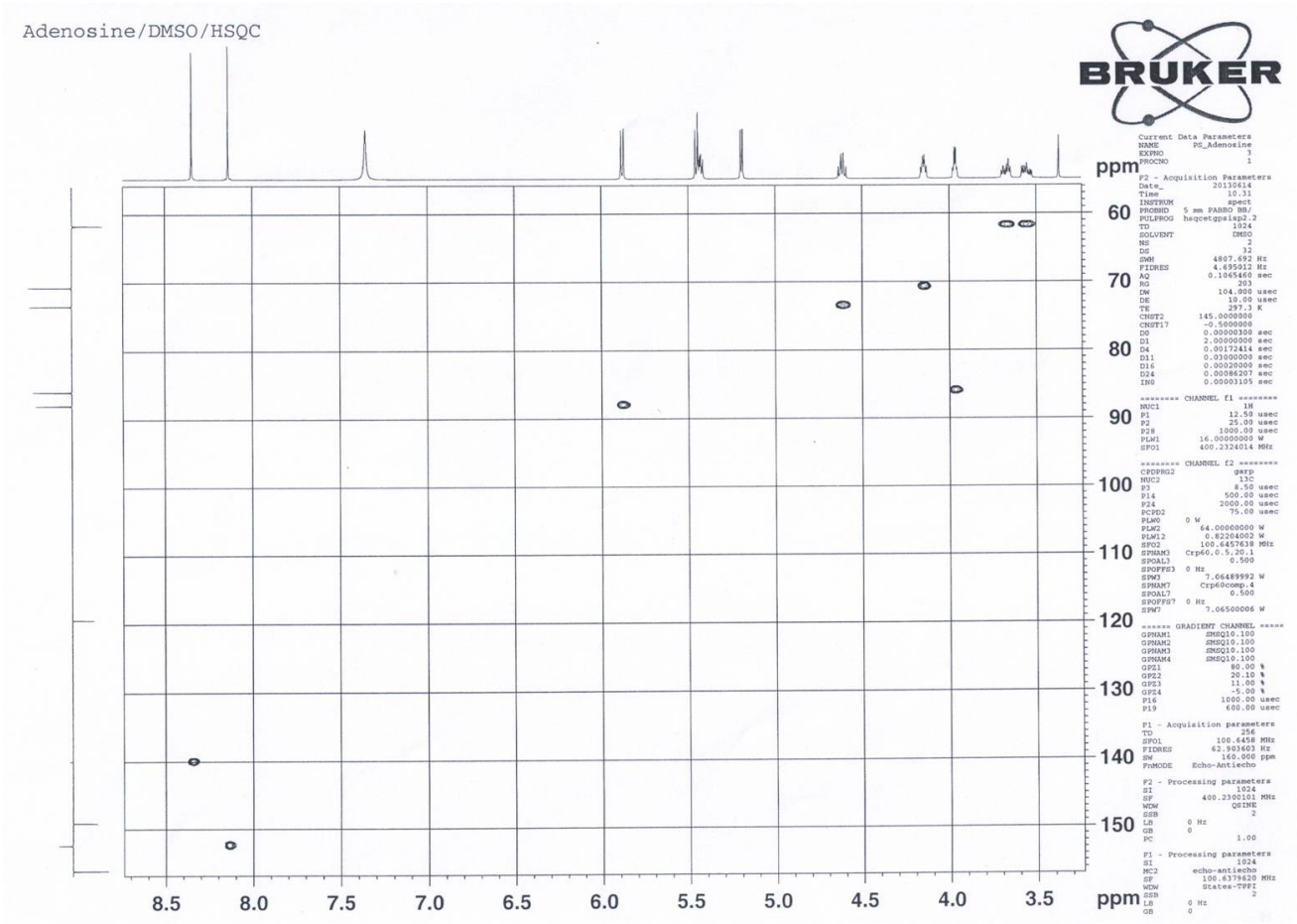


Figure S24. HSQC spectrum of commercial sample of adenosine.

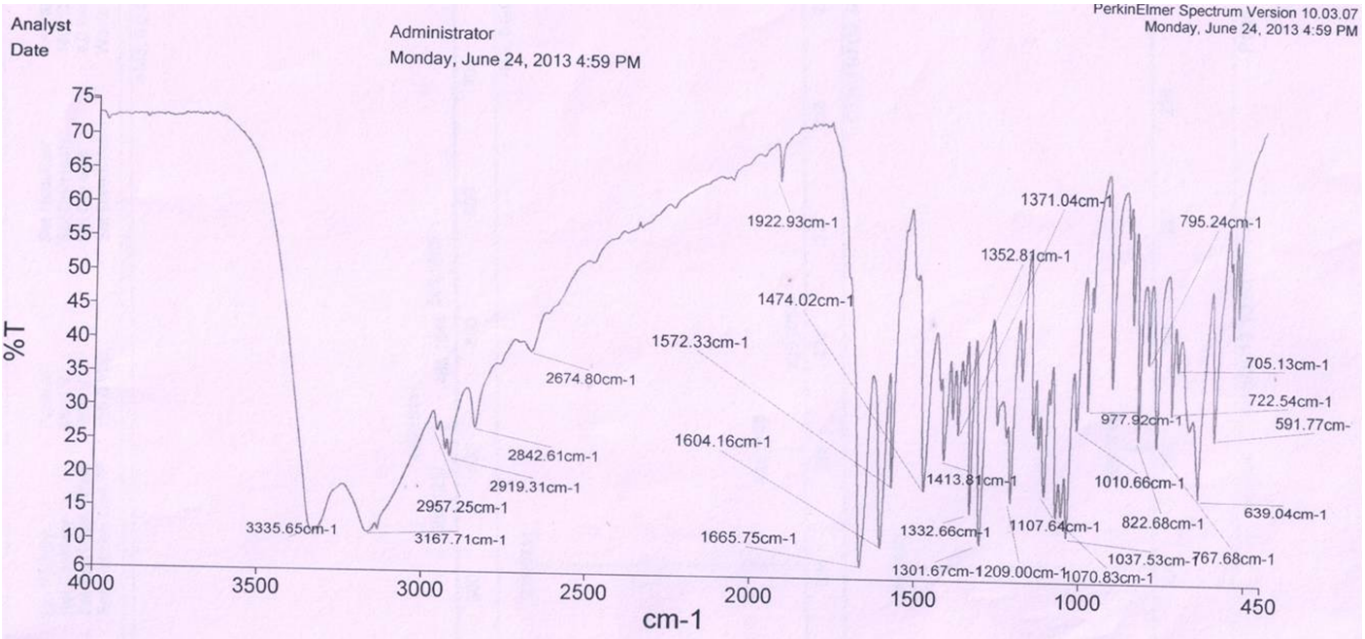


Figure S25. IR spectrum of commercial sample of adenosine

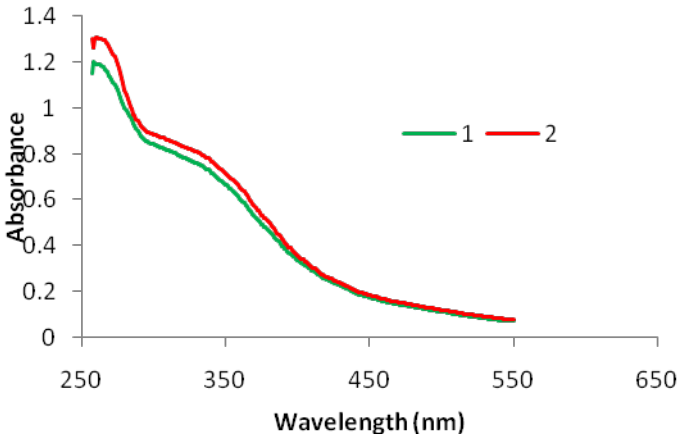


Figure S26. UV-vis spectrum of adenosine **1** (red trace) compared with commercial sample of adenosine (green trace).

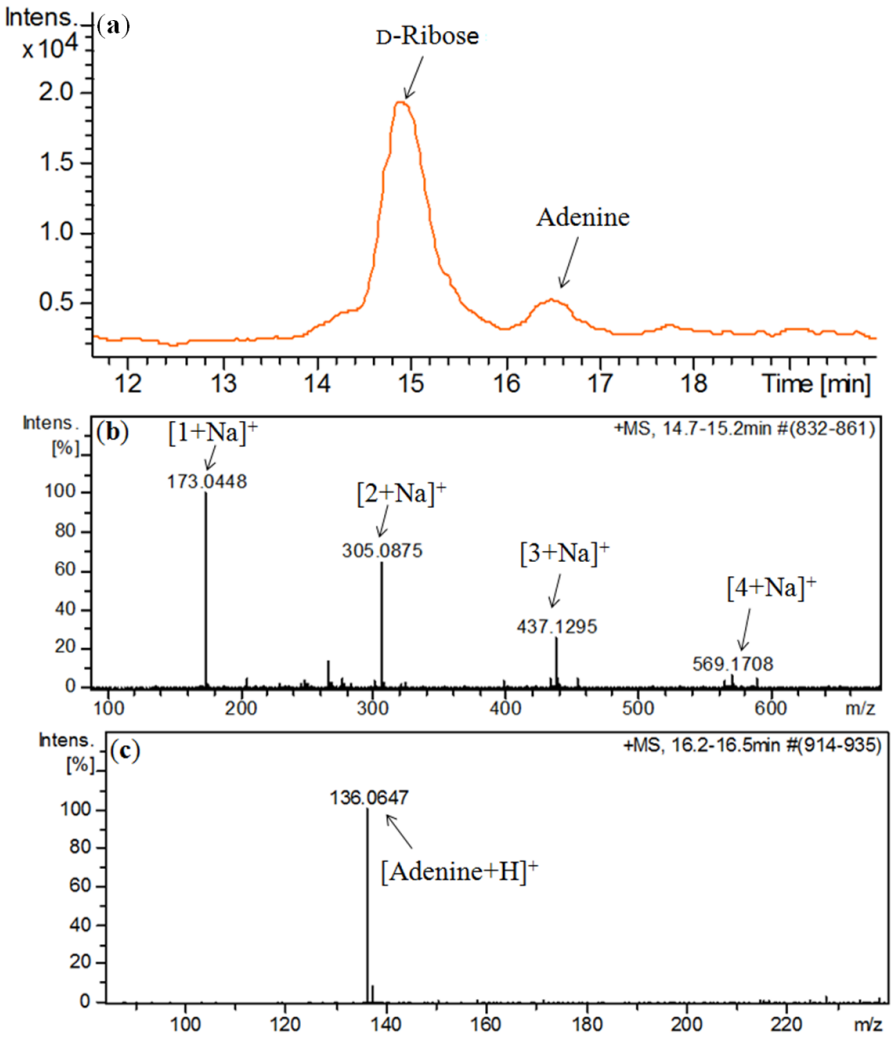


Figure S27. (a) LC chromatogram of the reaction mixture of scheme 1. The LC peak corresponding to adenosine (Figure 1a) is merged with the LC peak of D-ribose, (b) D-Ribose and its self condensation products (**2-4**, chart 1) , (c) unreacted adenine.

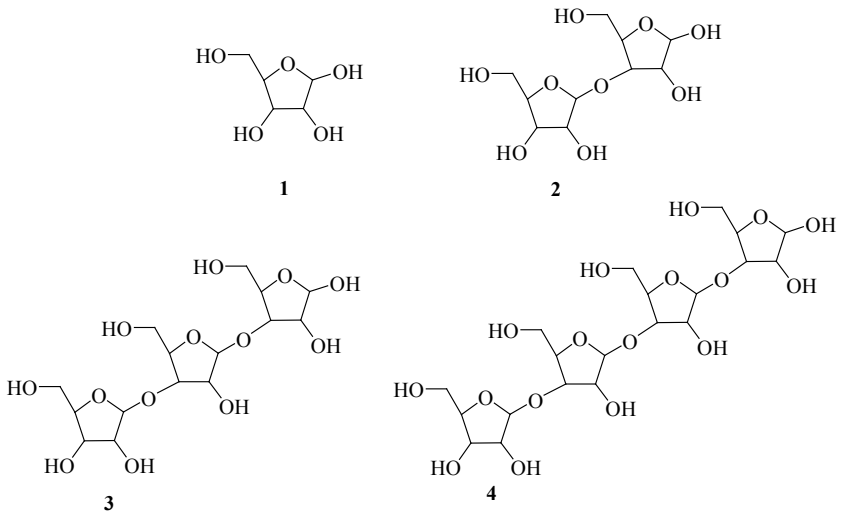


Chart S1. D-Ribose (**1**) and its self condensation products (**2-4**) formed during the reaction of scheme 2 (detected through LC-MS, Figure S1).

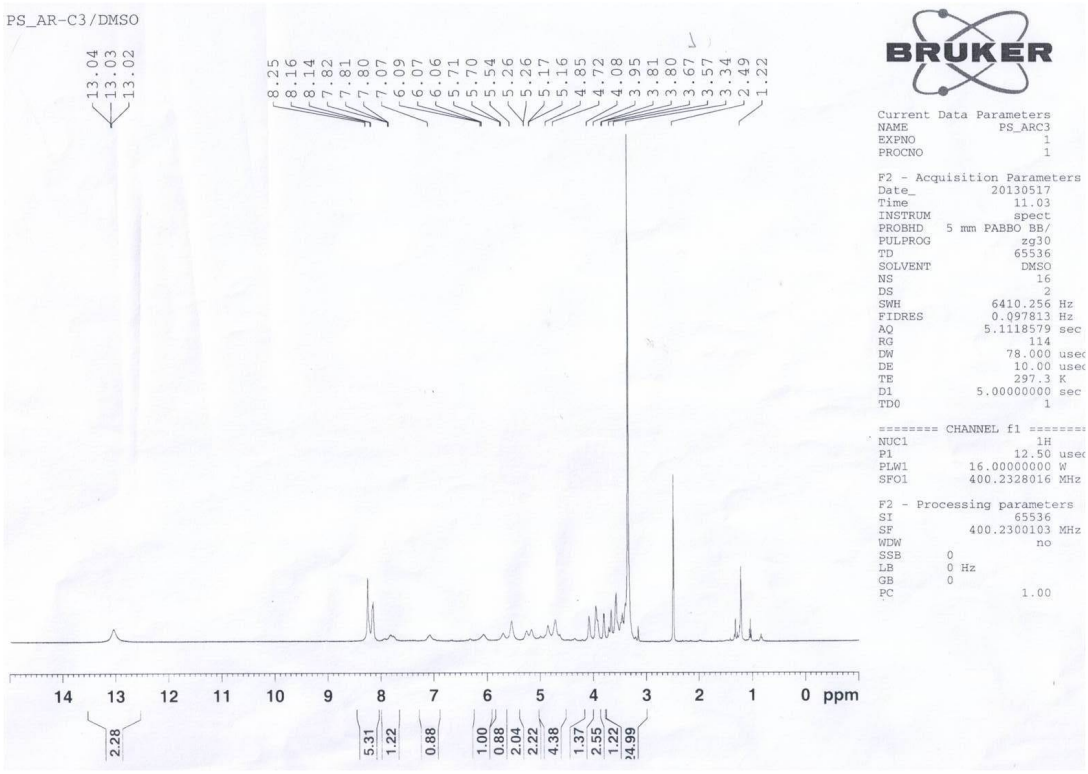


Figure S28. [1]H NMR spectrum of the product obtained by the reaction of adenine and D-ribose in tap water.

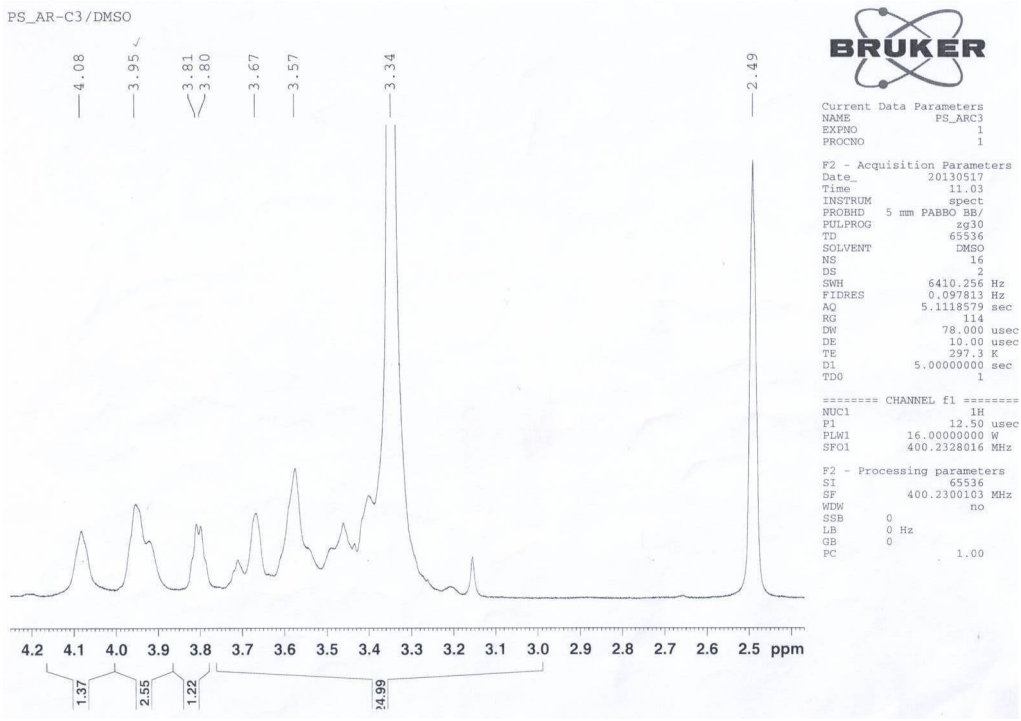


Figure S29. Expansion of [1]H NMR spectrum of the product obtained by the reaction of adenine and D-ribose in tap water.

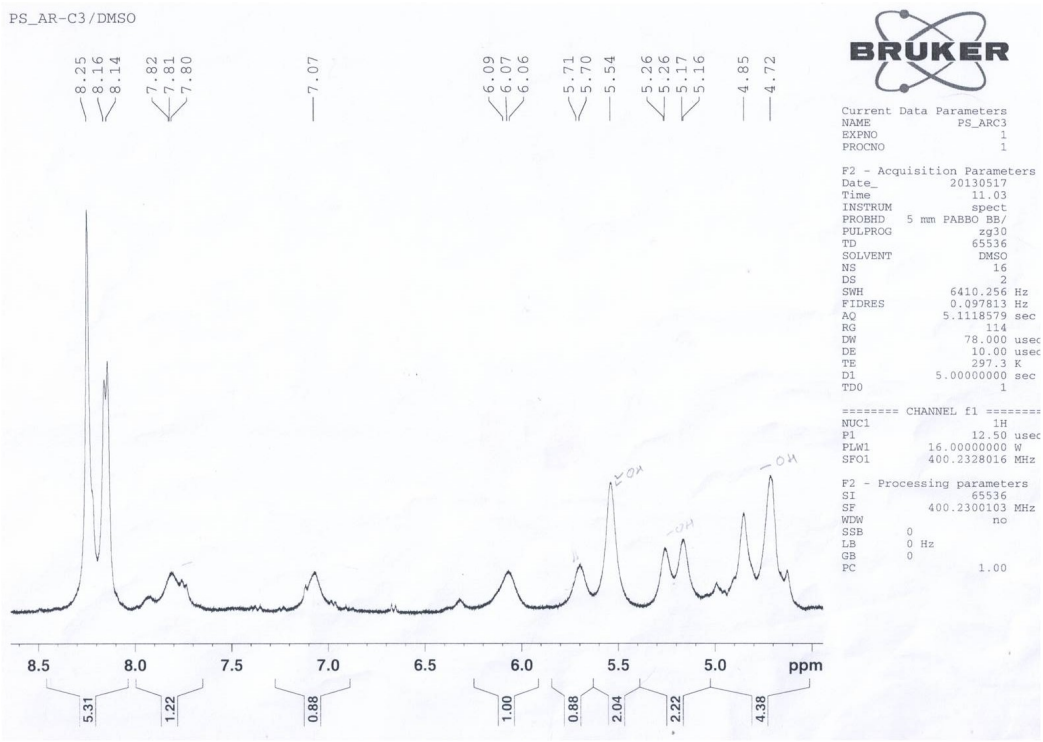


Figure S30. Expansion of ^{1}H NMR spectrum of the product obtained by the reaction of adenine and D-ribose in tap water.

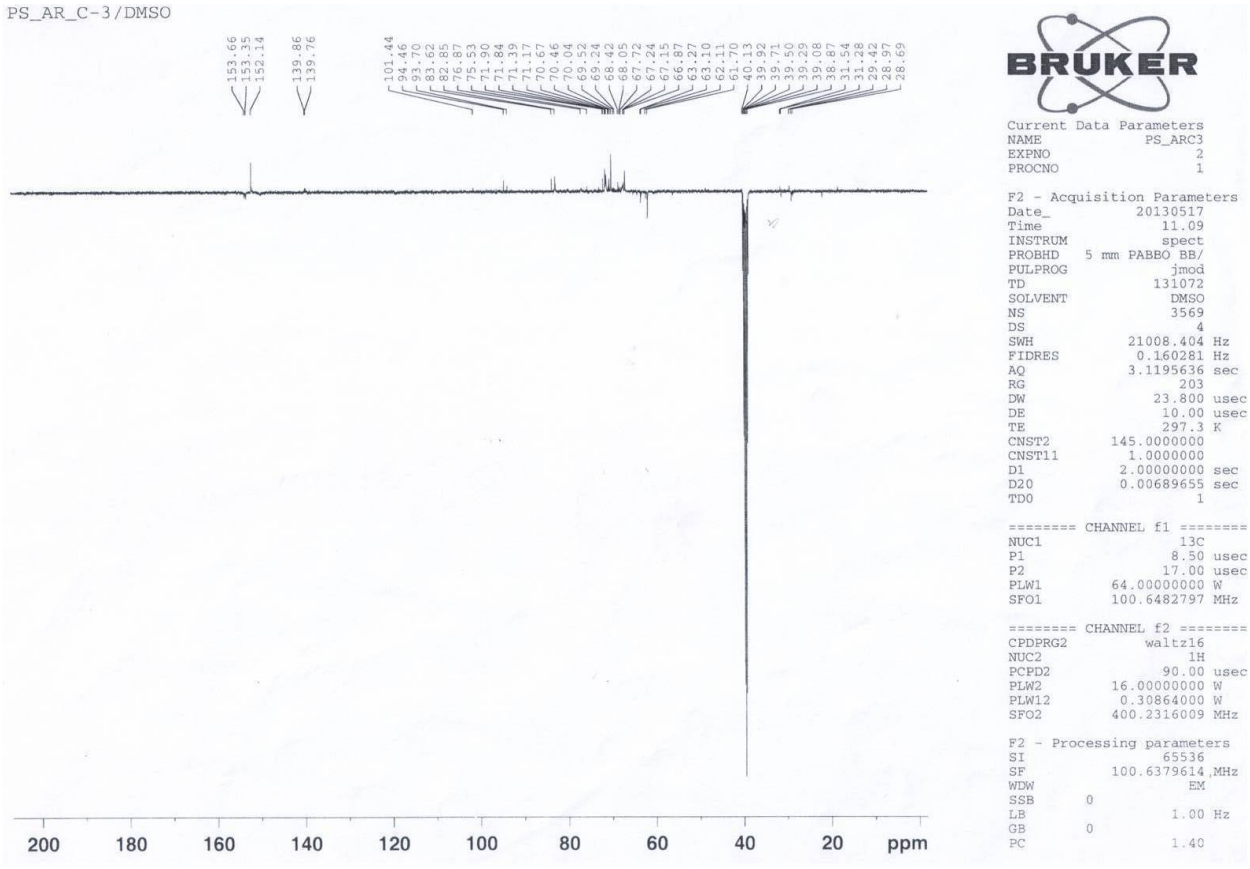


Figure S31. ^{13}C NMR spectrum of the product obtained by the reaction of adenine and D-ribose in tap water.

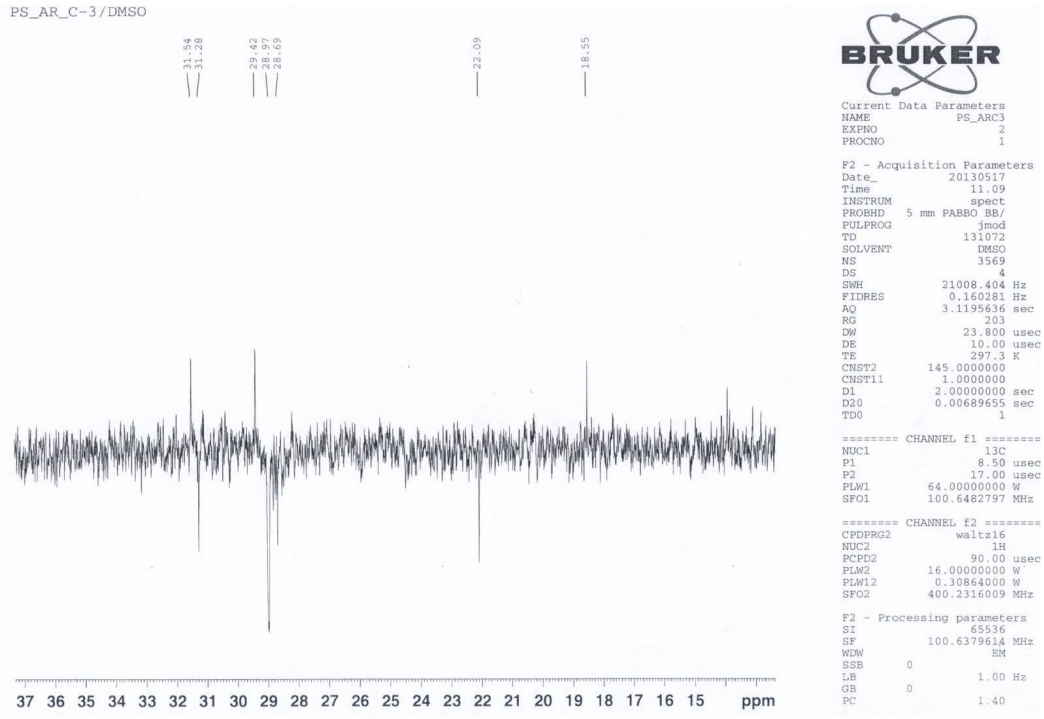


Figure S32. Expansion of ^{13}C NMR spectrum of the product obtained by the reaction of adenine and D-ribose in tap water.

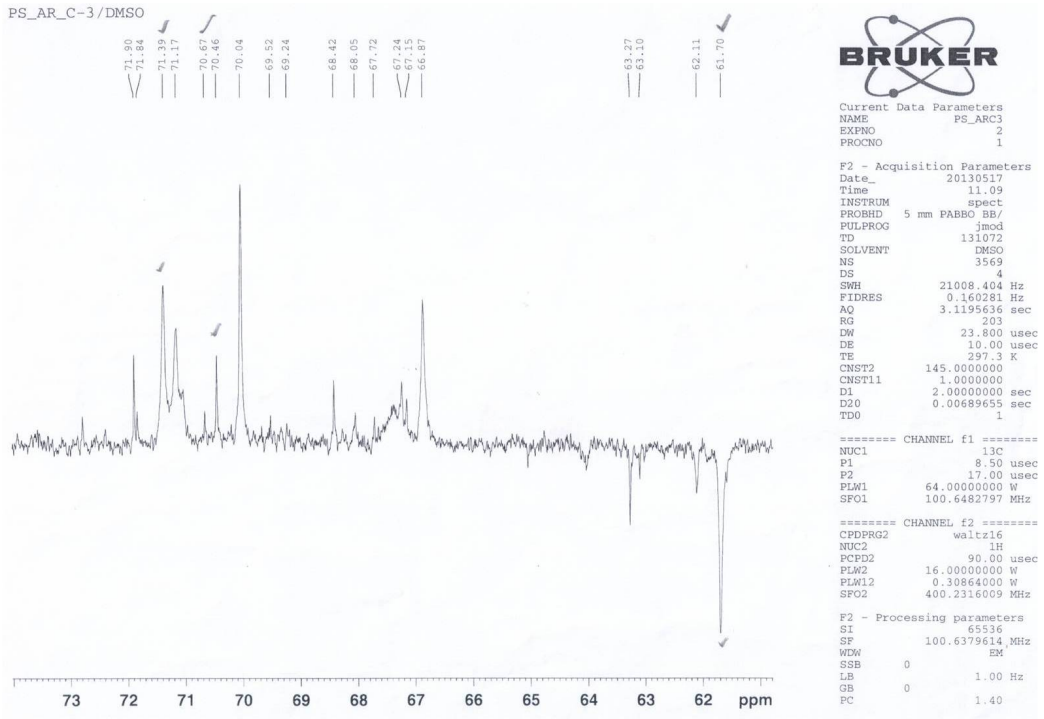


Figure S33. Expansion of ^{13}C NMR spectrum of the product obtained by the reaction of adenine and D-ribose in tap water.

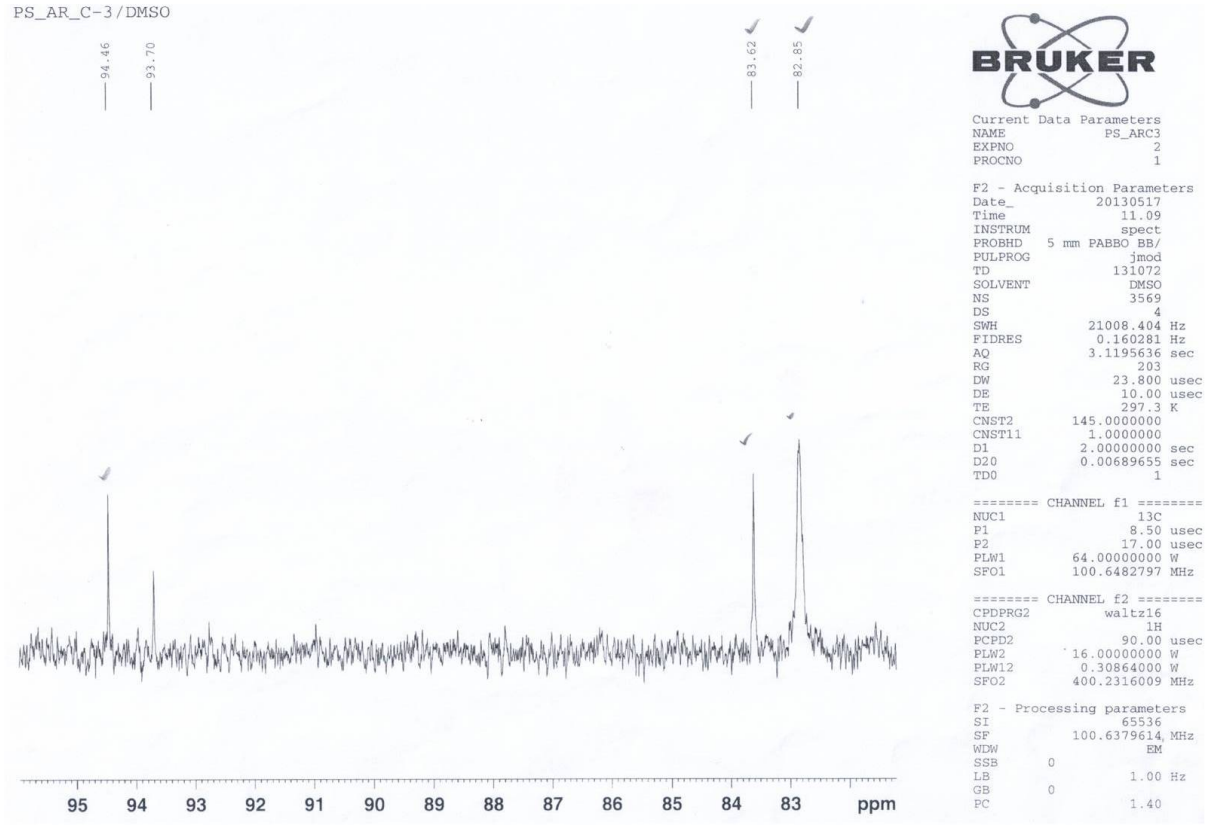


Figure S34. Expansion of ^{13}C NMR spectrum of the product obtained by the reaction of adenine and D-ribose in tap water.

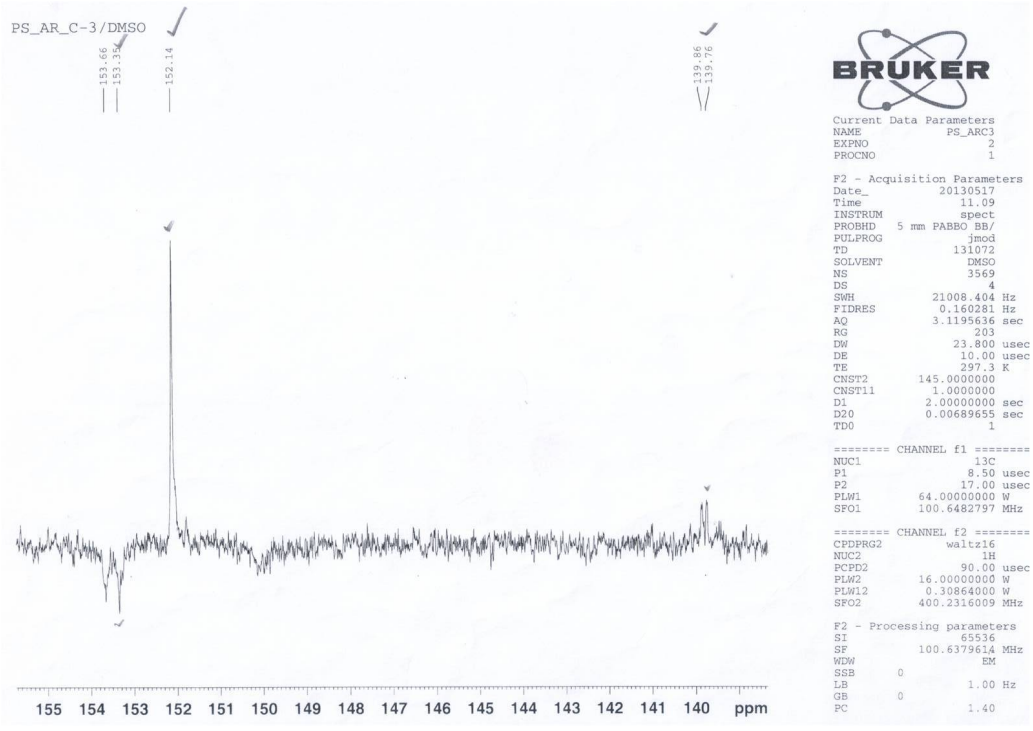


Figure S35. Expansion of ^{13}C NMR spectrometer of the product obtained by the reaction of adenine and D-ribose in tap water.

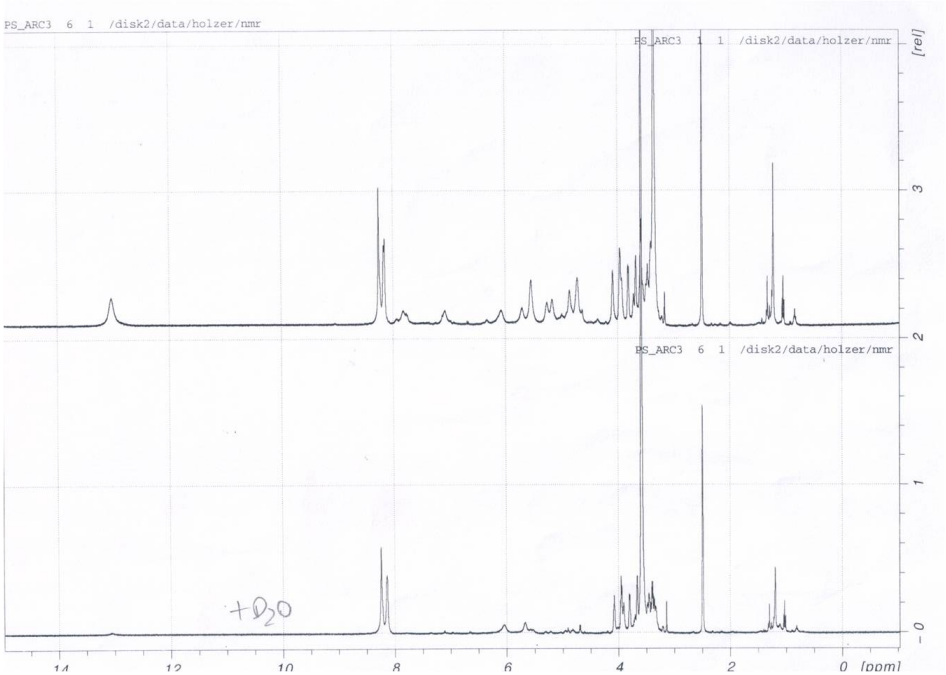


Figure S36. D_2O exchange ^1H NMR spectrum of the product obtained by the reaction of adenine and D-ribose in tap water.

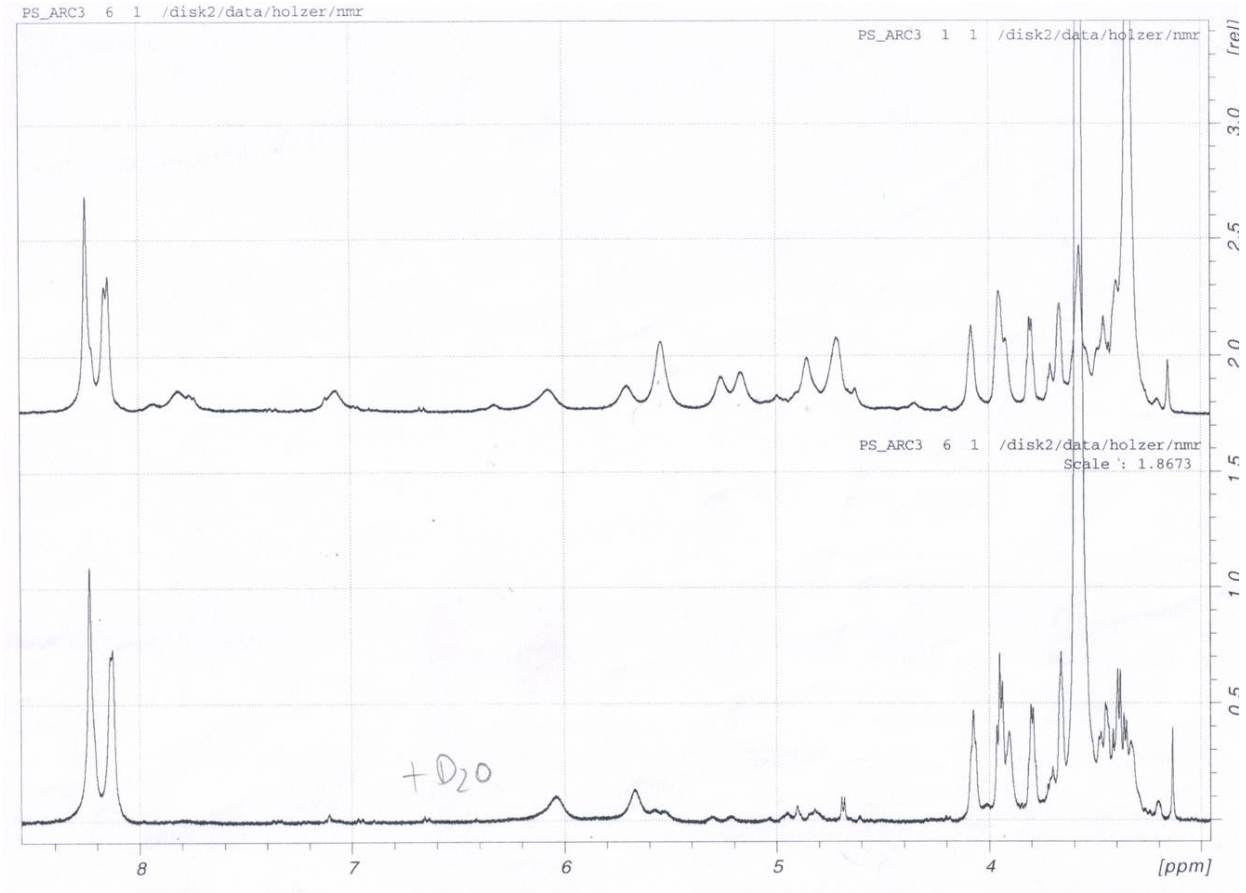


Figure S37. D_2O exchange ^1H NMR spectrum of the product obtained by the reaction of adenine and D-ribose in tap water.

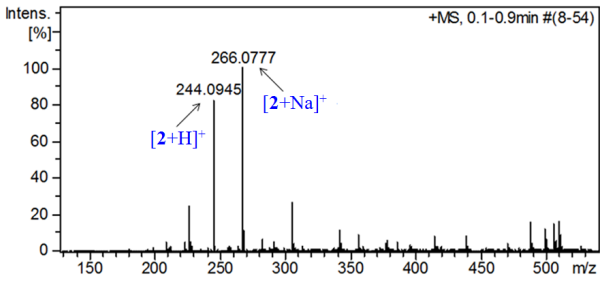


Figure S38. High resolution mass spectrum of cytidine **2** (scheme 3) (calcd *m/z* 244.0930, $[M+H]^+$).

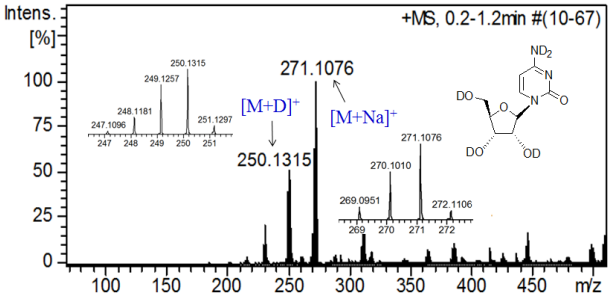


Figure S39. ESI-MS of cytidine 2 in D_2O-ACN (7:3). The peak at *m/z* 250.1315 corresponds to $[M+D]^+$ of cytidine after replacement of all the exchangeable Hs' with D (calcd *m/z* 250.1305, $[M+D]^+$) while *m/z* 271.1076 corresponds to $[M+Na]^+$. Inset: deuterated cytidine and expansions of mass peaks at *m/z* 250 and 271.

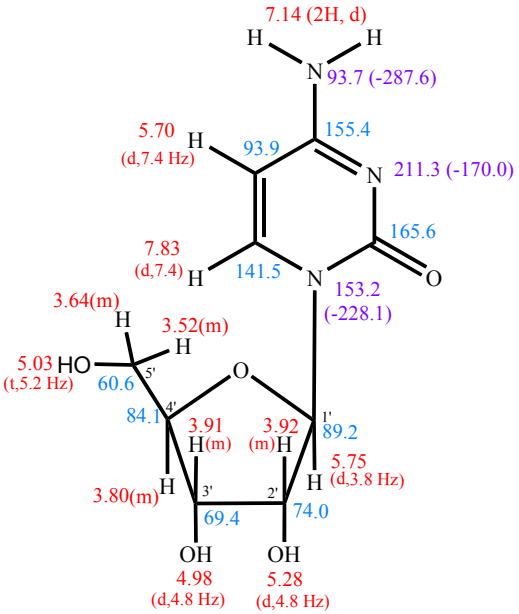


Figure S40. NMR assignments to H, C and N of compound **2**. Red: ^1H, DMSO: δ= 2.4900 ppm, Blue: ^{13}C, DMSO: δ= 39.5 ppm, Purple: ^{15}N, against $MeNO_2$ (0.00ppm) (in parenthesis); numbers without parenthesis are shifts against NH_3 liquid.

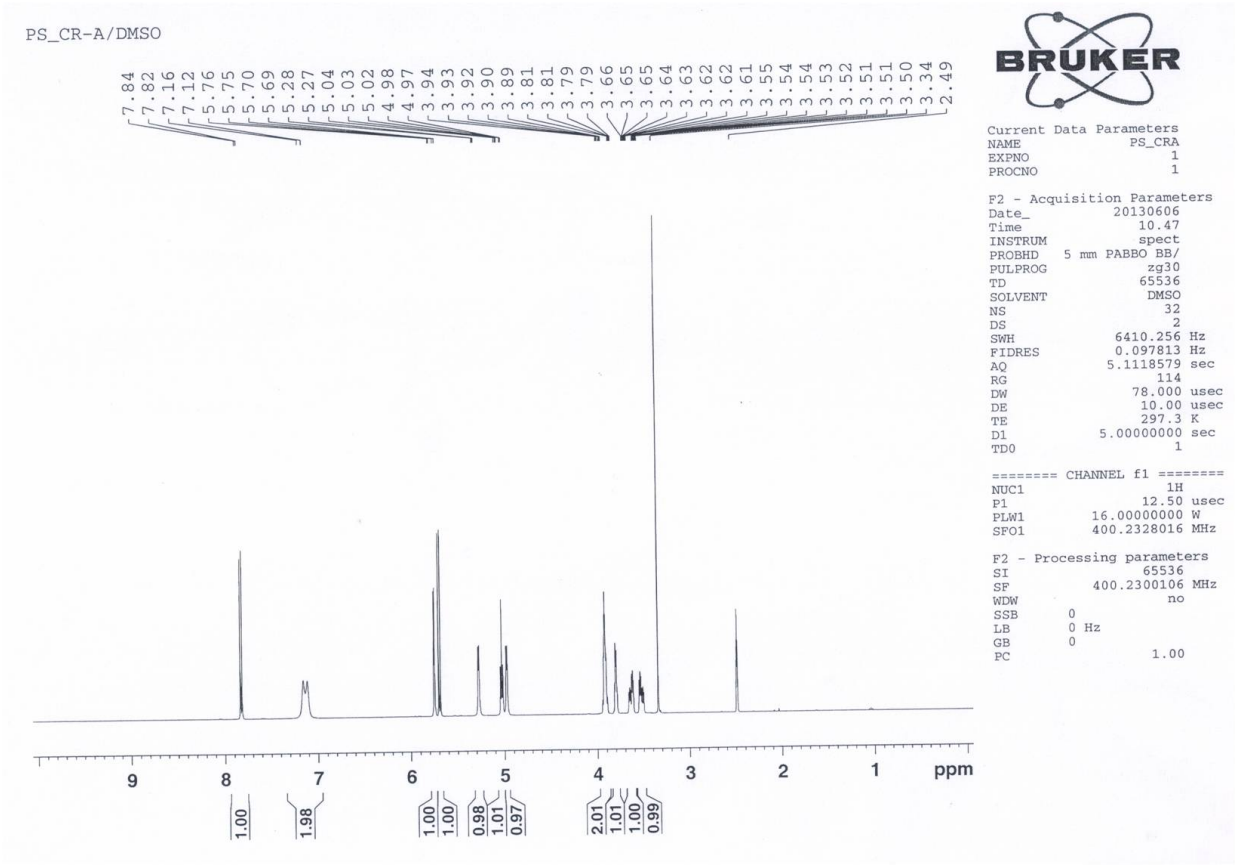


Figure S41. ^{1}H NMR spectrum of compound **2** in DMSO-d_6.

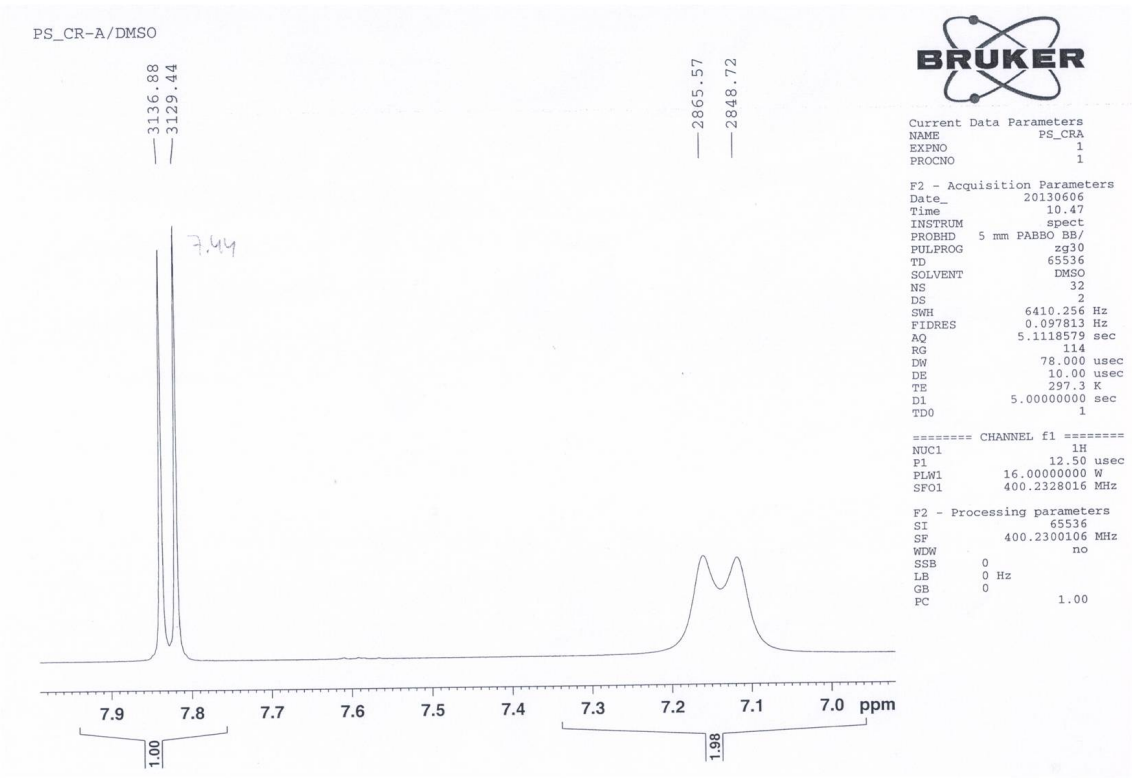


Figure S42. ^{1}H NMR spectrum of compound **2** (expansion).

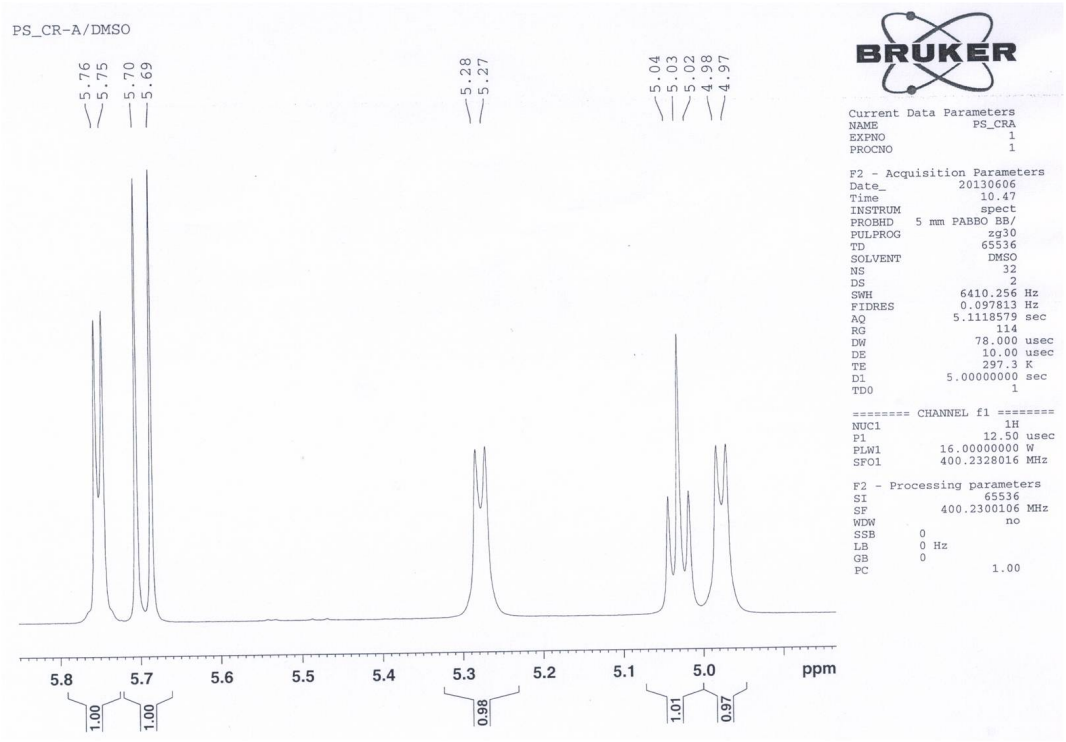


Figure S43. ^{1}H NMR spectrum of compound **2** (expansion).

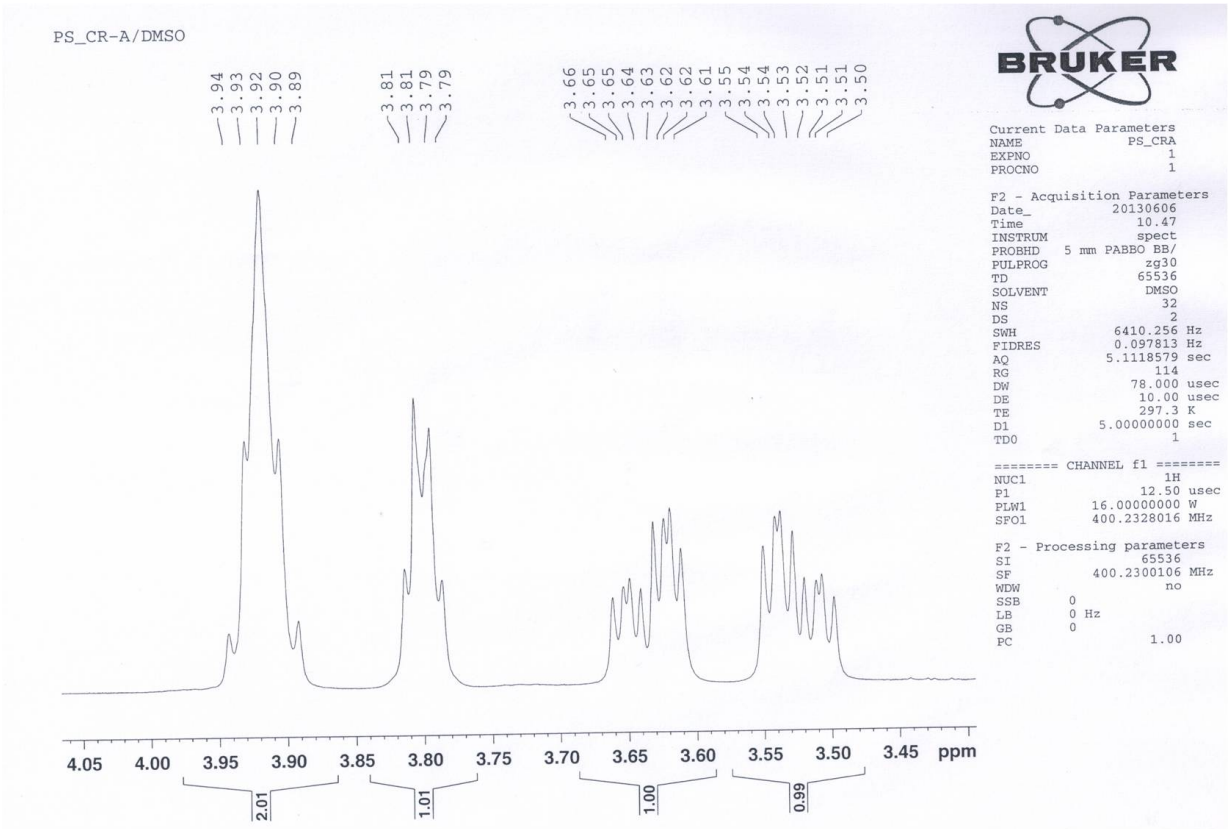


Figure S44. ^{1}H NMR spectrum of compound **2** (expansion).

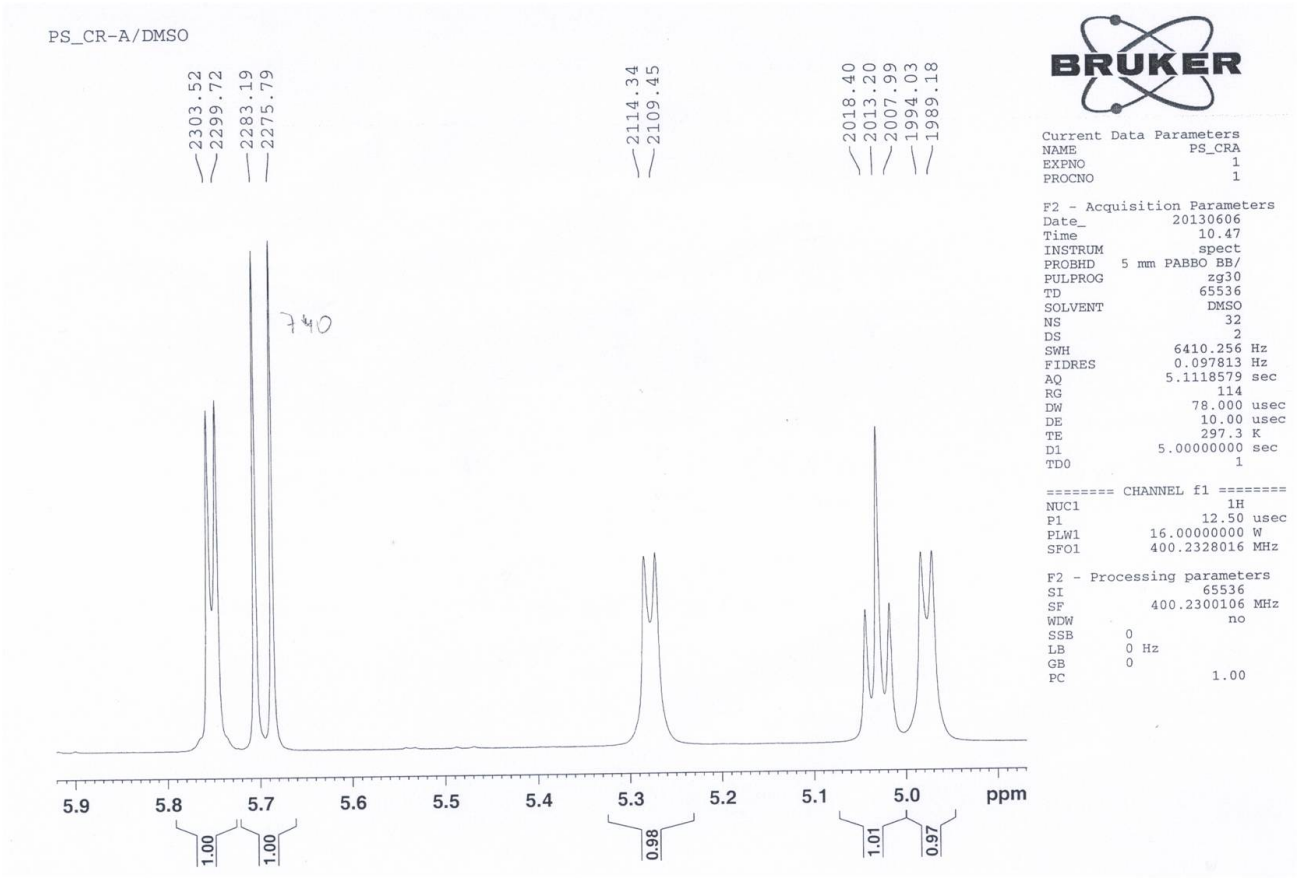


Figure S45. ^{1}H NMR spectrum of compound **2** (expansion).

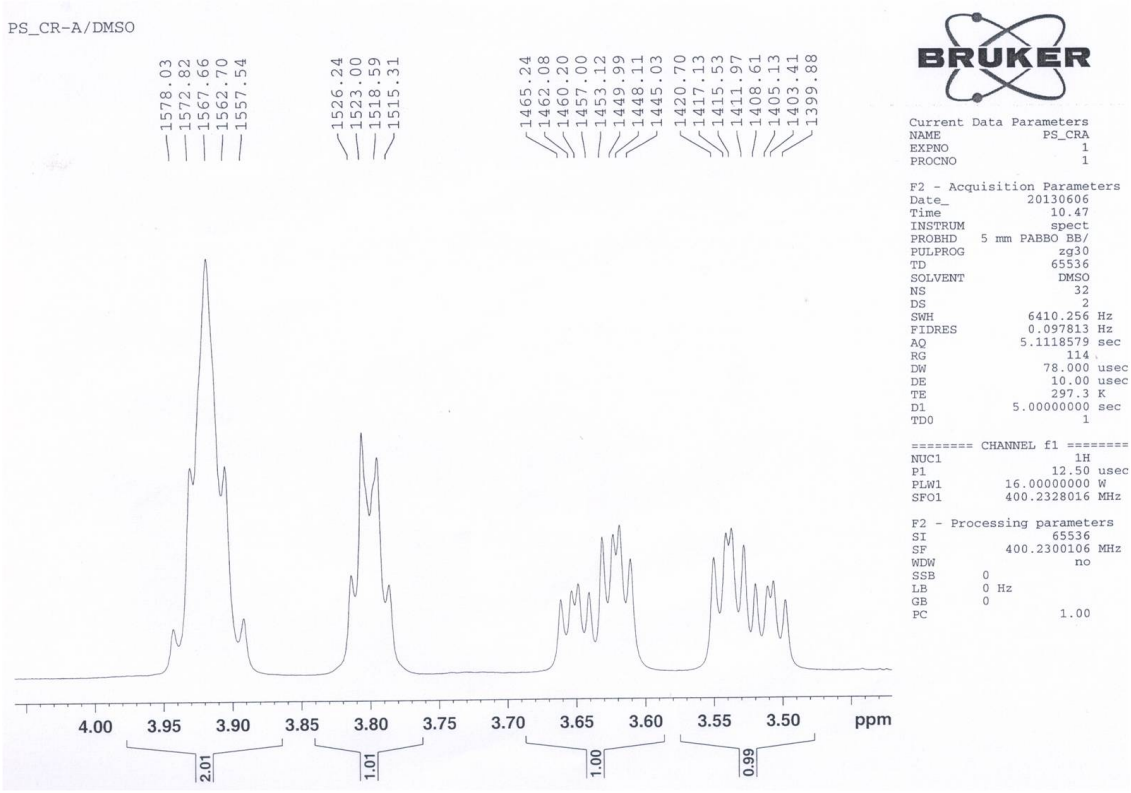


Figure S46. ^{1}H NMR spectrum of compound **2** (expansion).

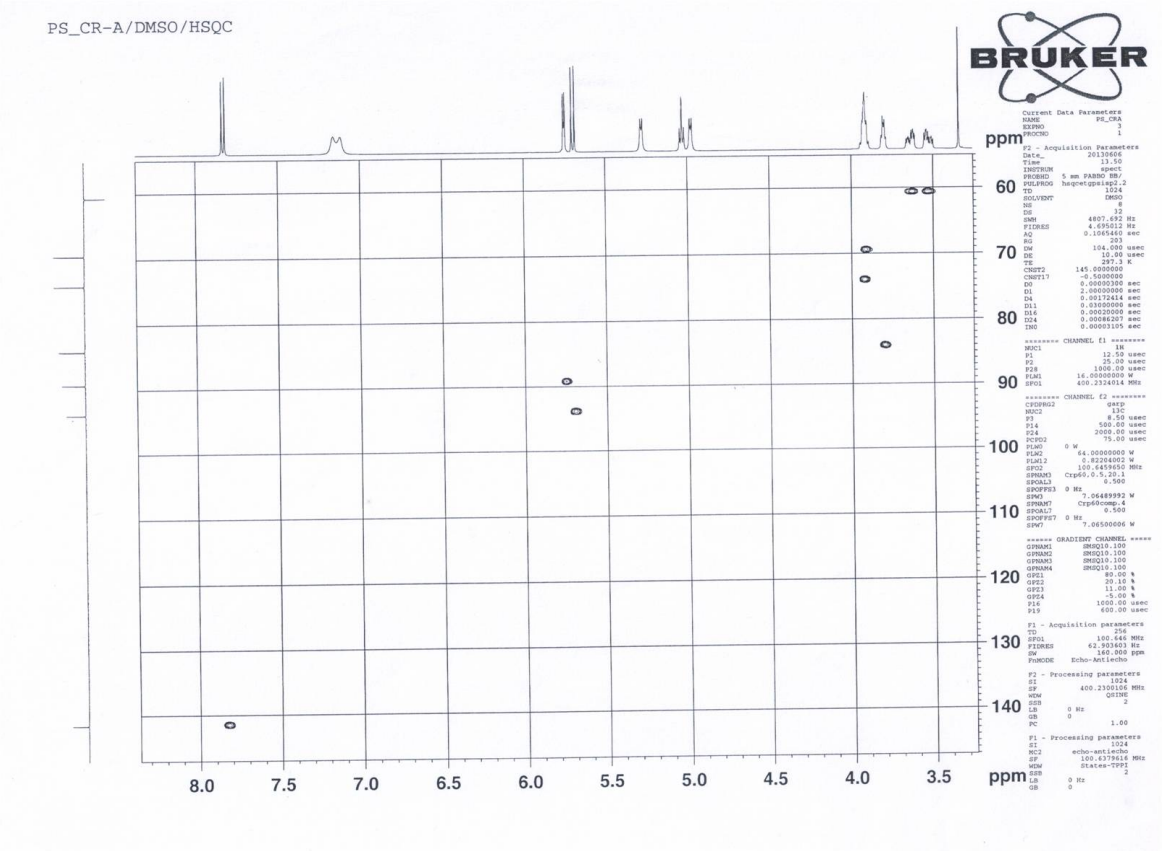


Figure S47. HSQC NMR spectrum of compound **2**.

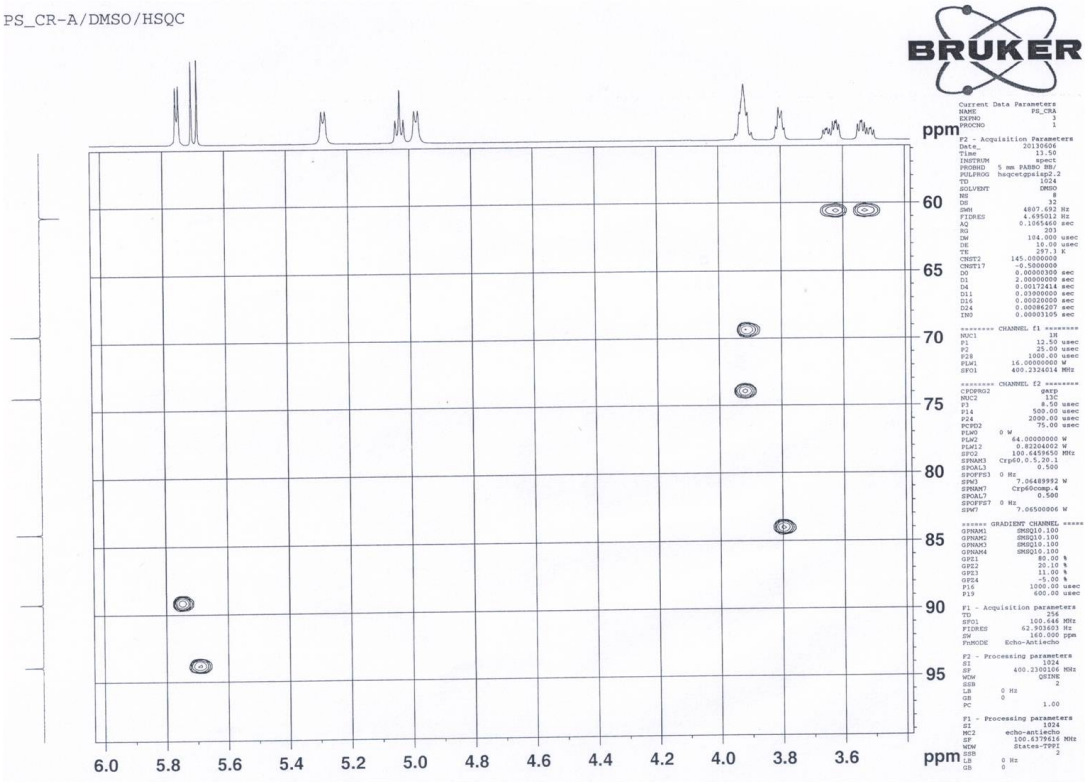


Figure S48. HSQC NMR spectrum of compound **2**.

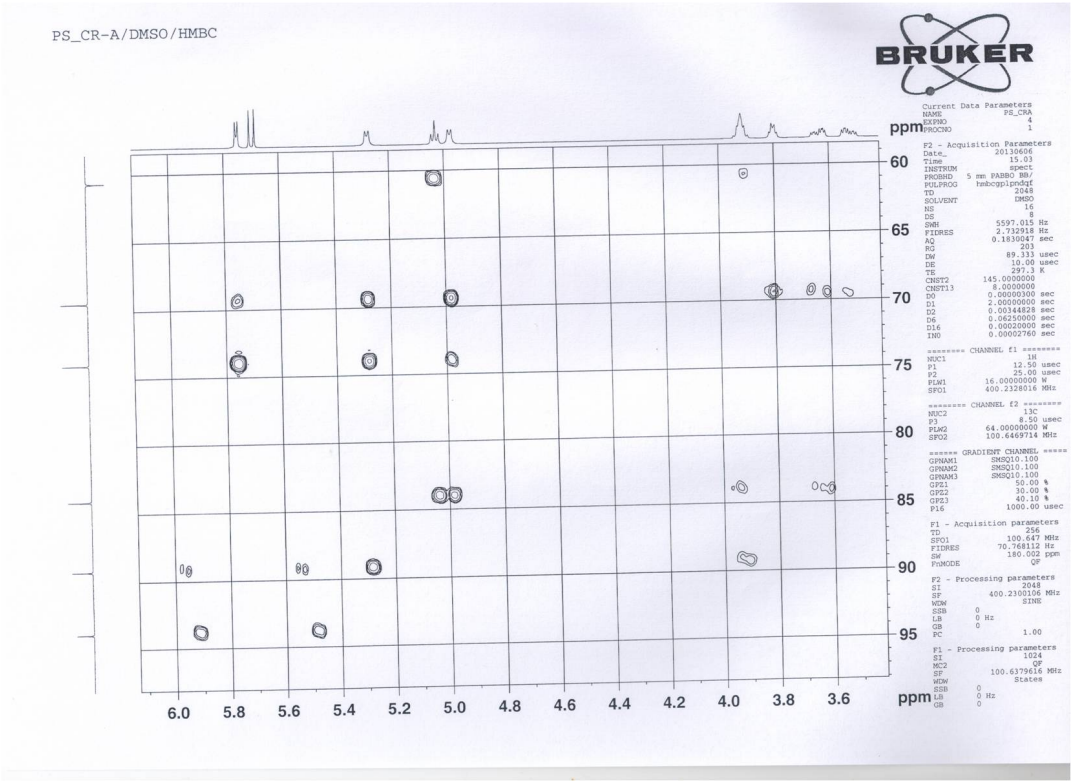


Figure S49. HMBC NMR spectrum of compound **2**.

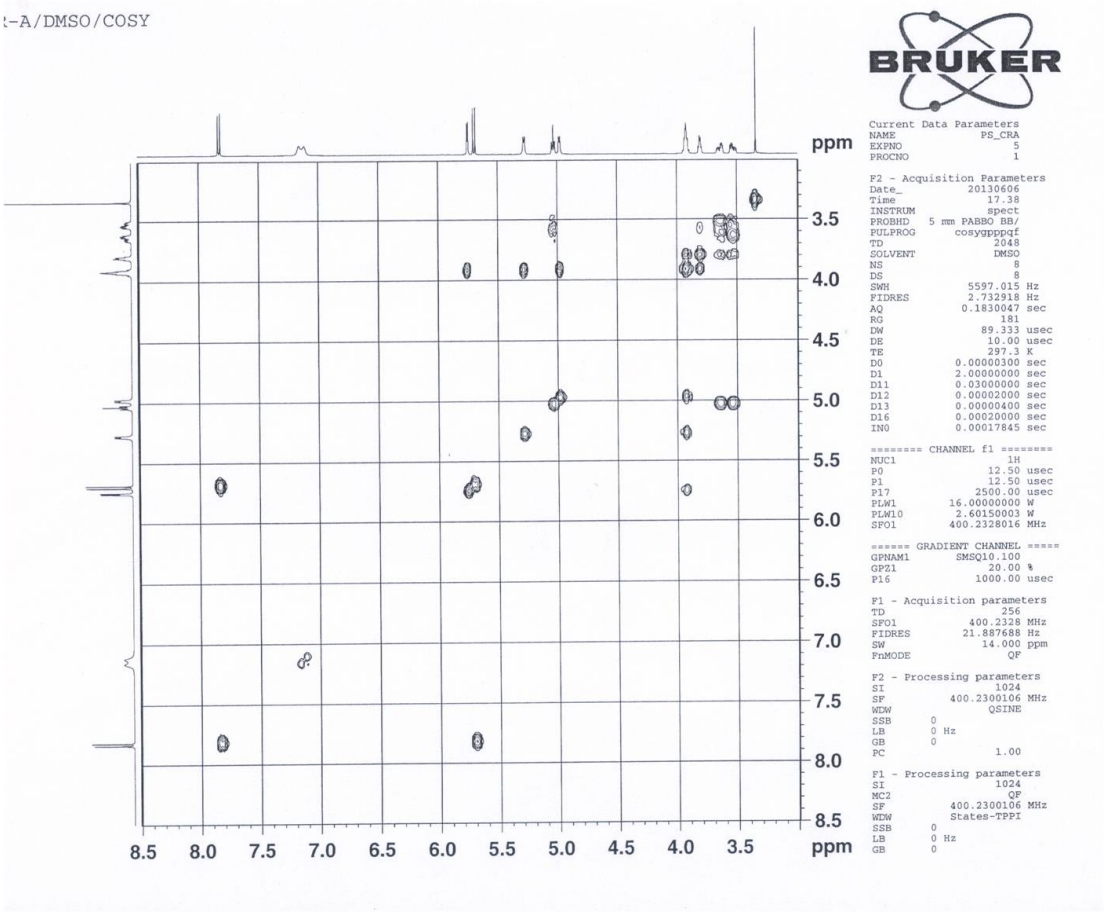


Figure S50. ^{1}H-^{1}H COSY NMR spectrum of compound **2**.

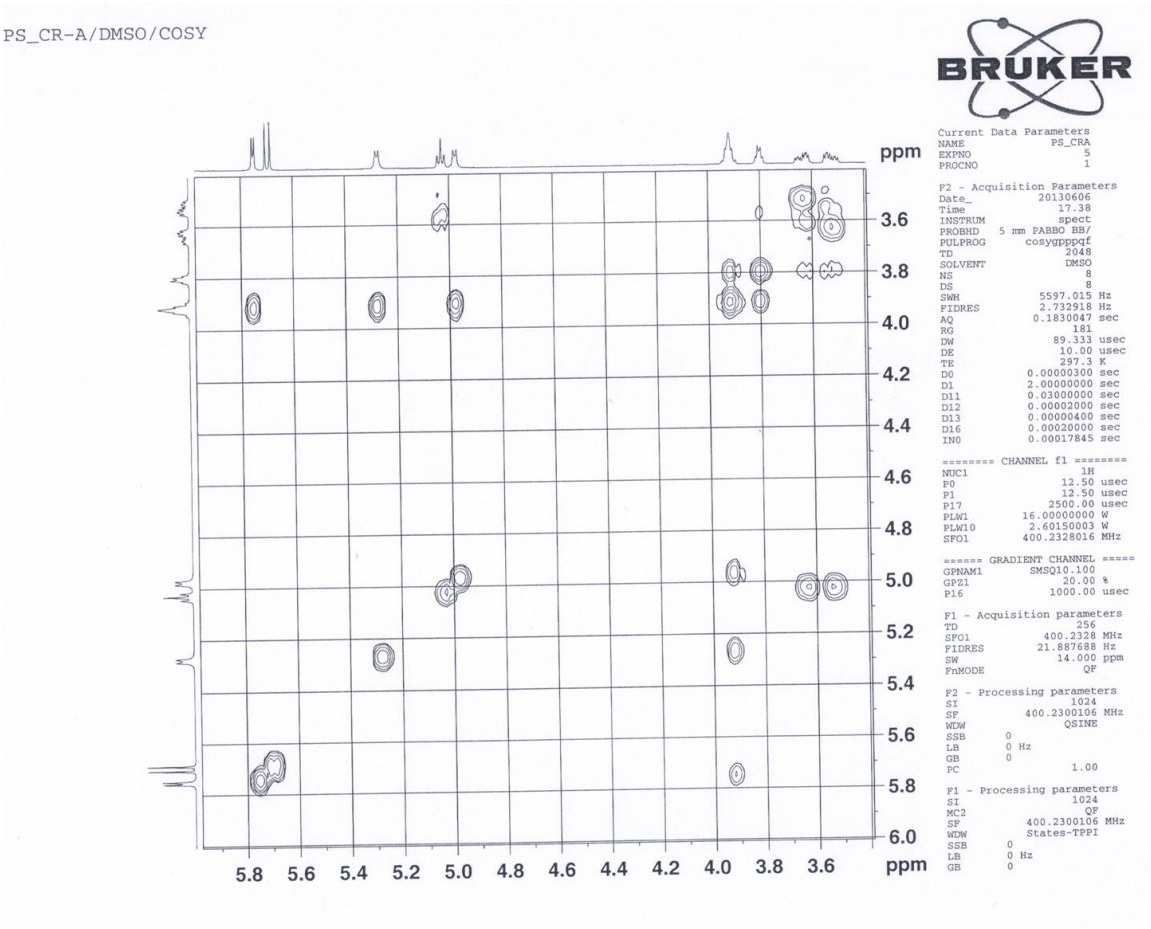


Figure S51. ^{1}H-^{1}H COSY NMR spectrum of compound **2** (expansion).

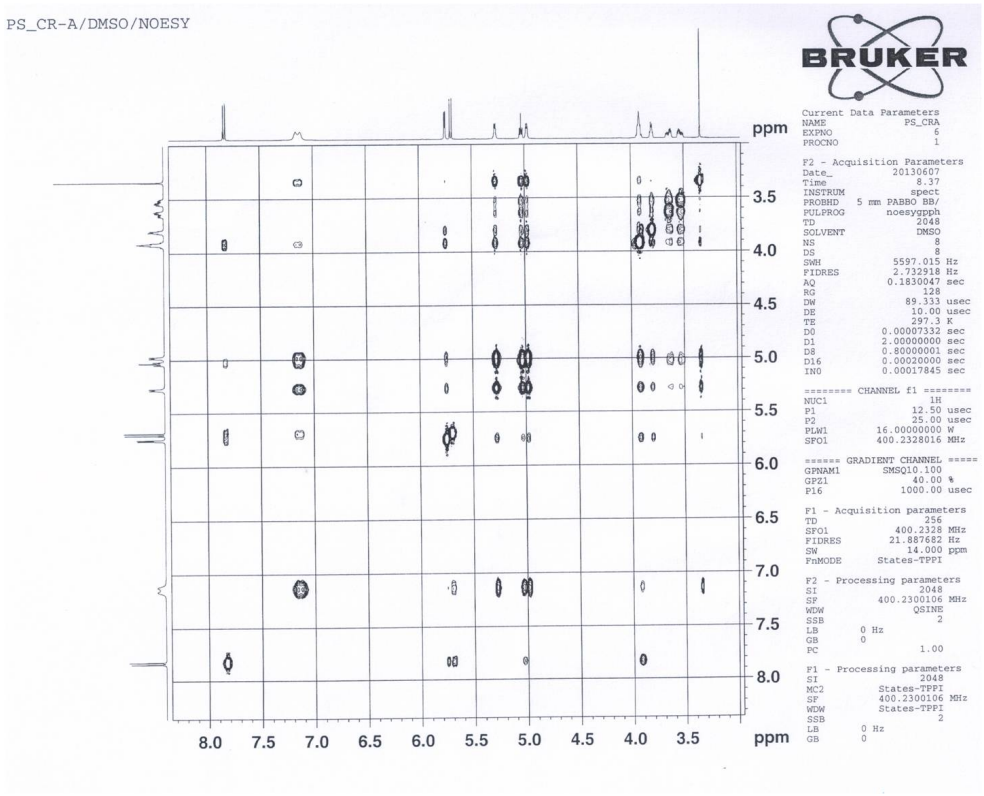


Figure S52. ^{1}H-^{1}H NOESY NMR spectrum of compound **2**.

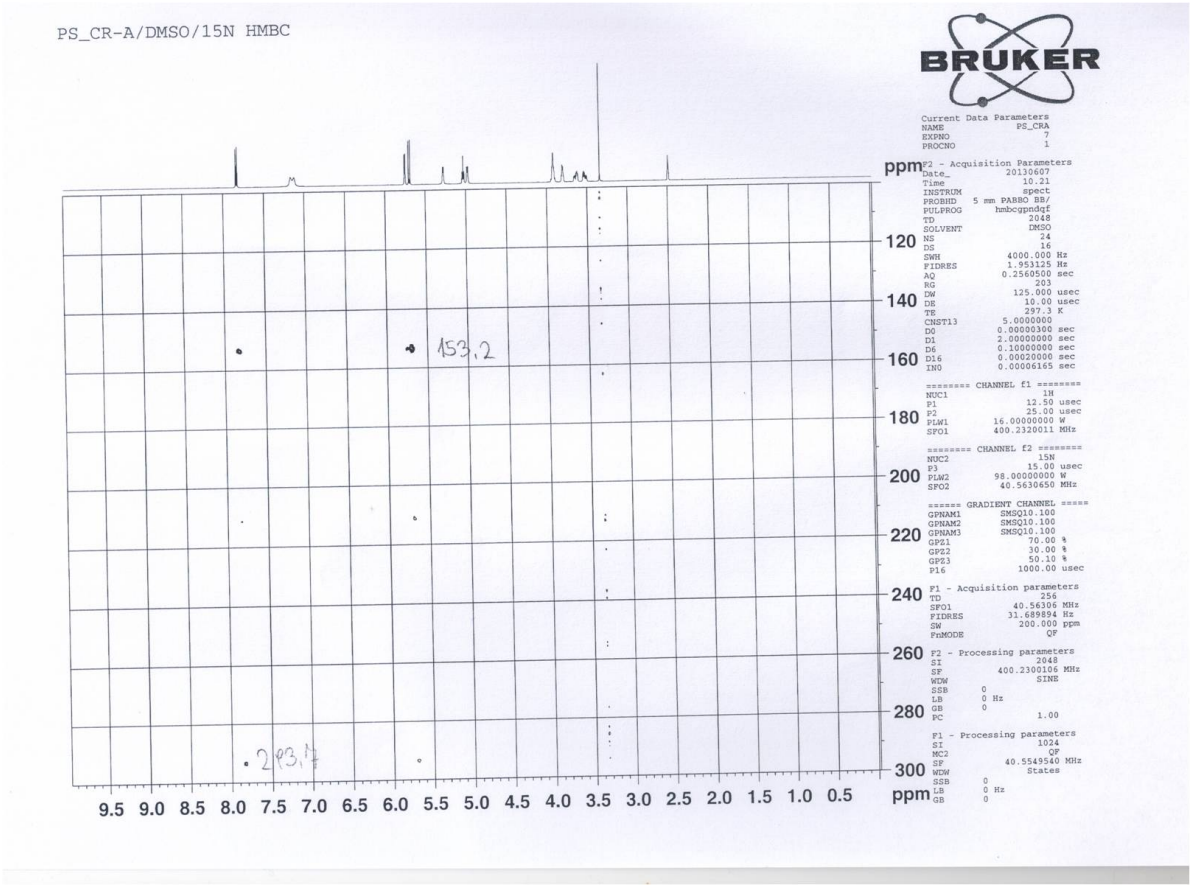


Figure S53. ^{15}N HMBC spectrum of compound **2**.

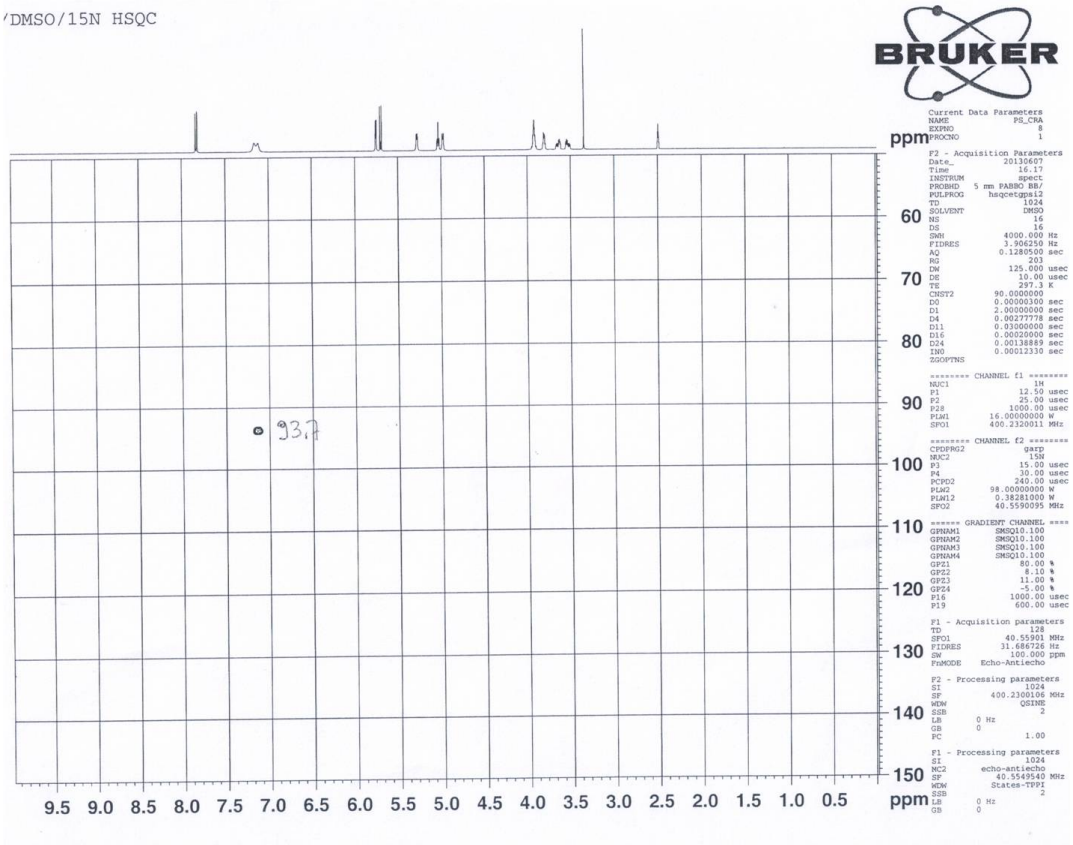


Figure S54. ^{15}N HSQC spectrum of compound **2**.

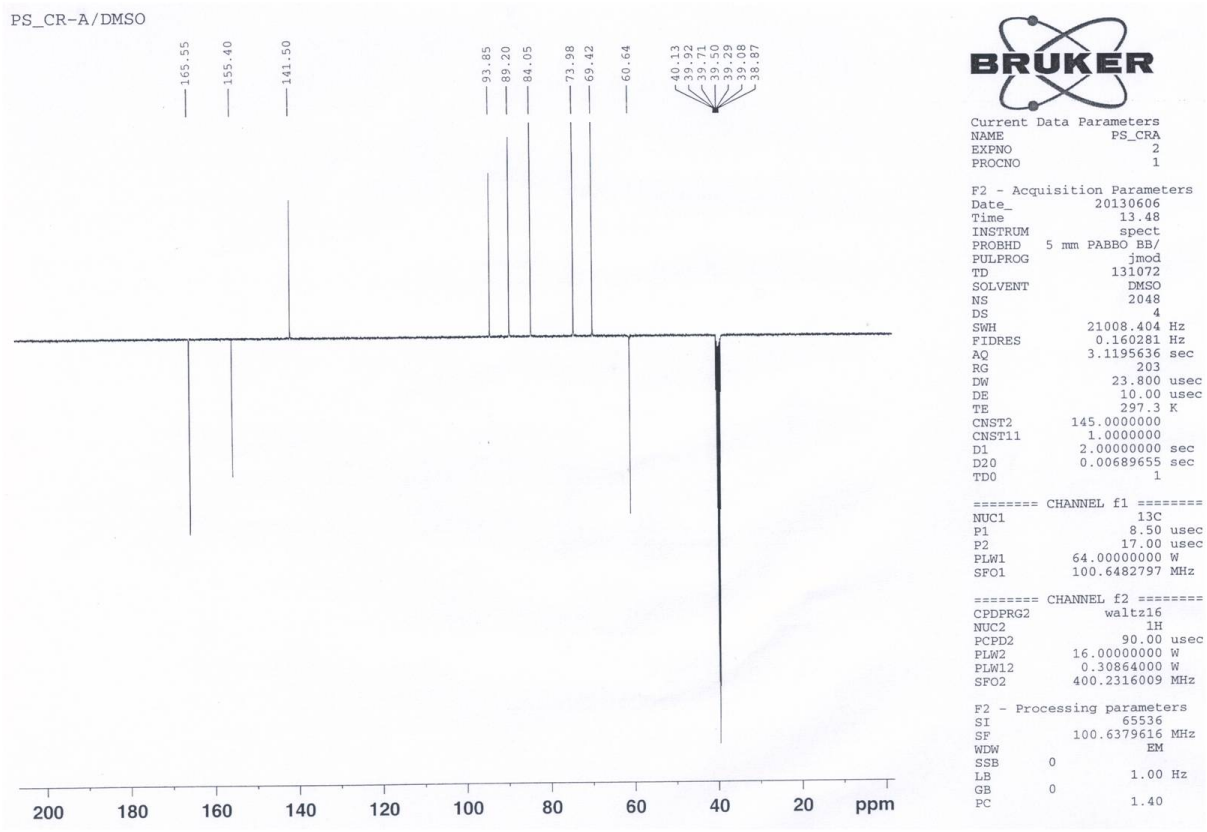


Figure S55. ^{13}C NMR spectrum of compound **2**.

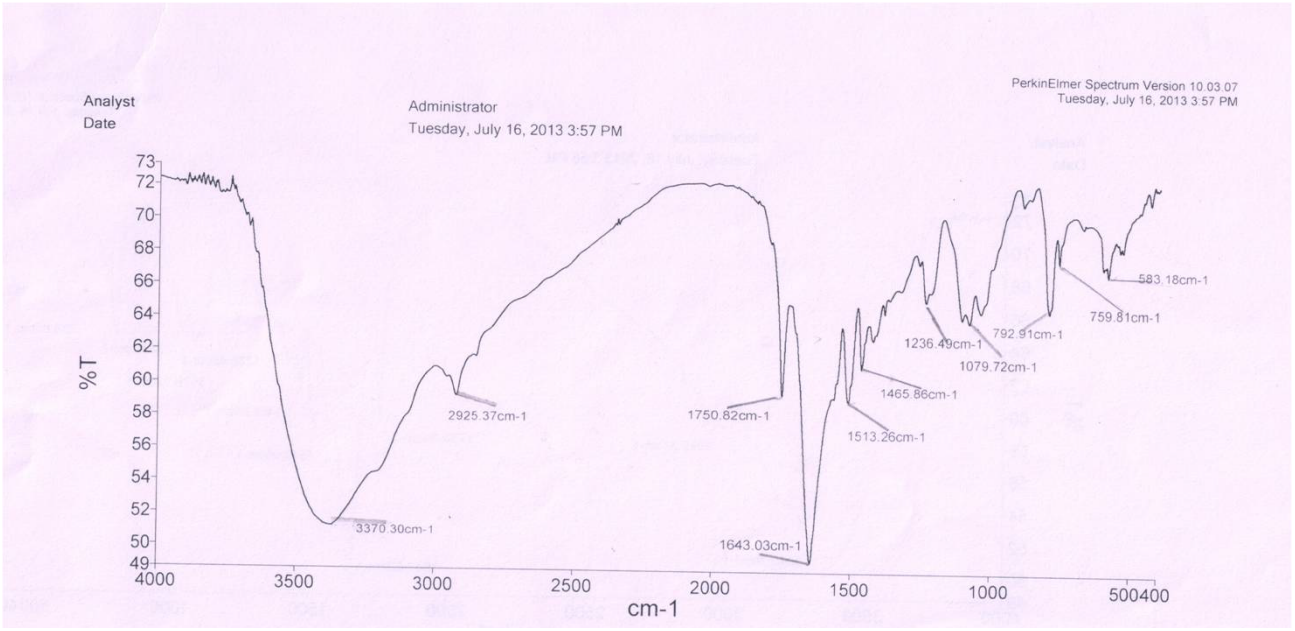


Figure S56. IR spectrum of compound **2**.

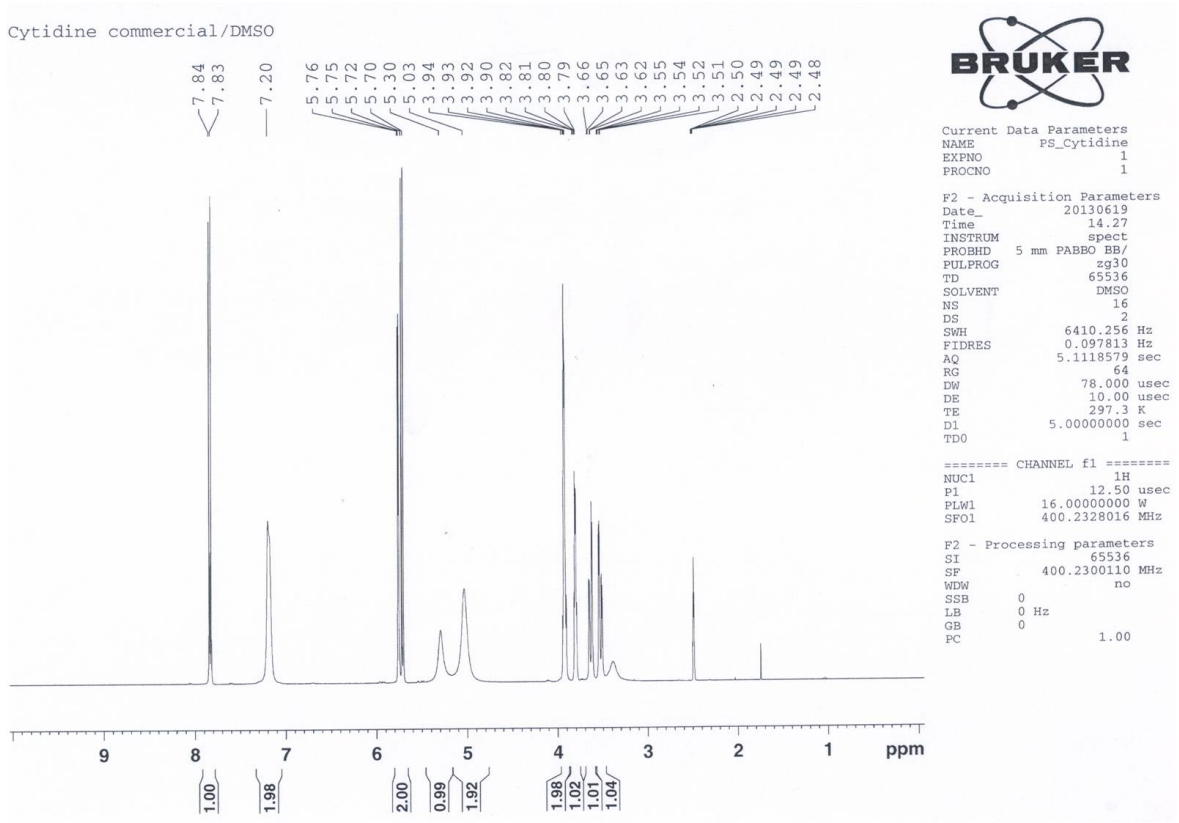


Figure S57. ^{1}H NMR spectrum of commercial sample of cytidine.

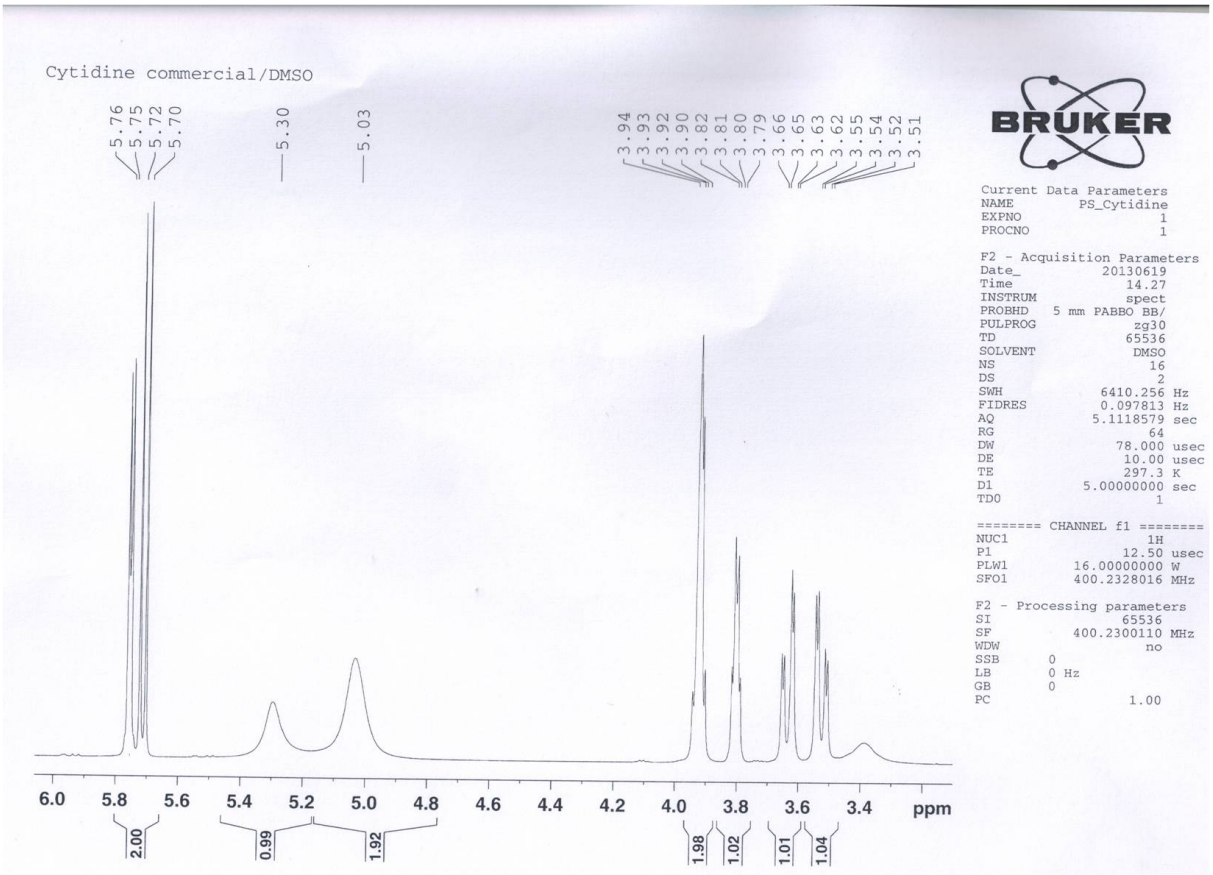


Figure S58. ^{1}H NMR spectrum of commercial sample of cytidine.

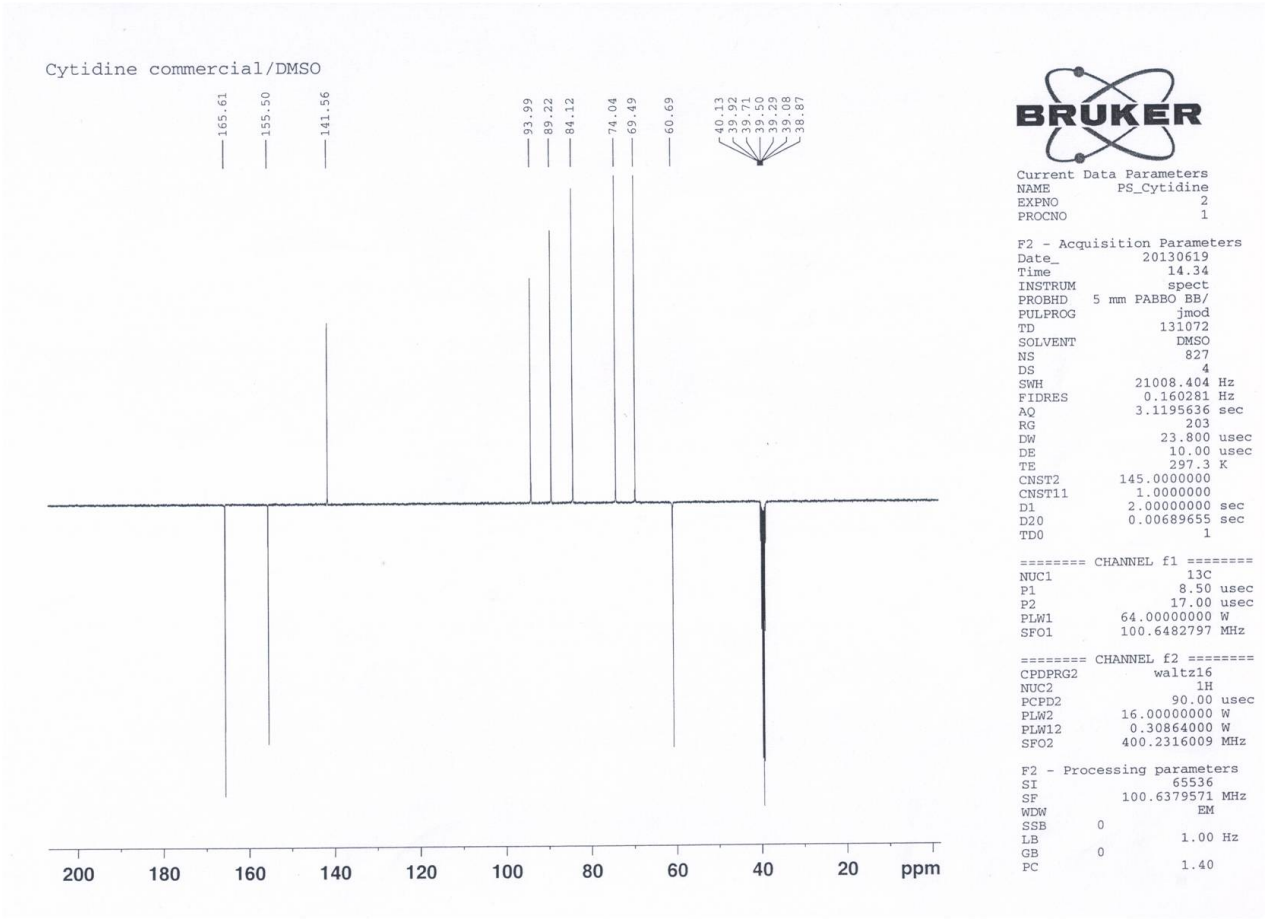


Figure S59. ^{13}C NMR spectrum of commercial sample of cytidine.

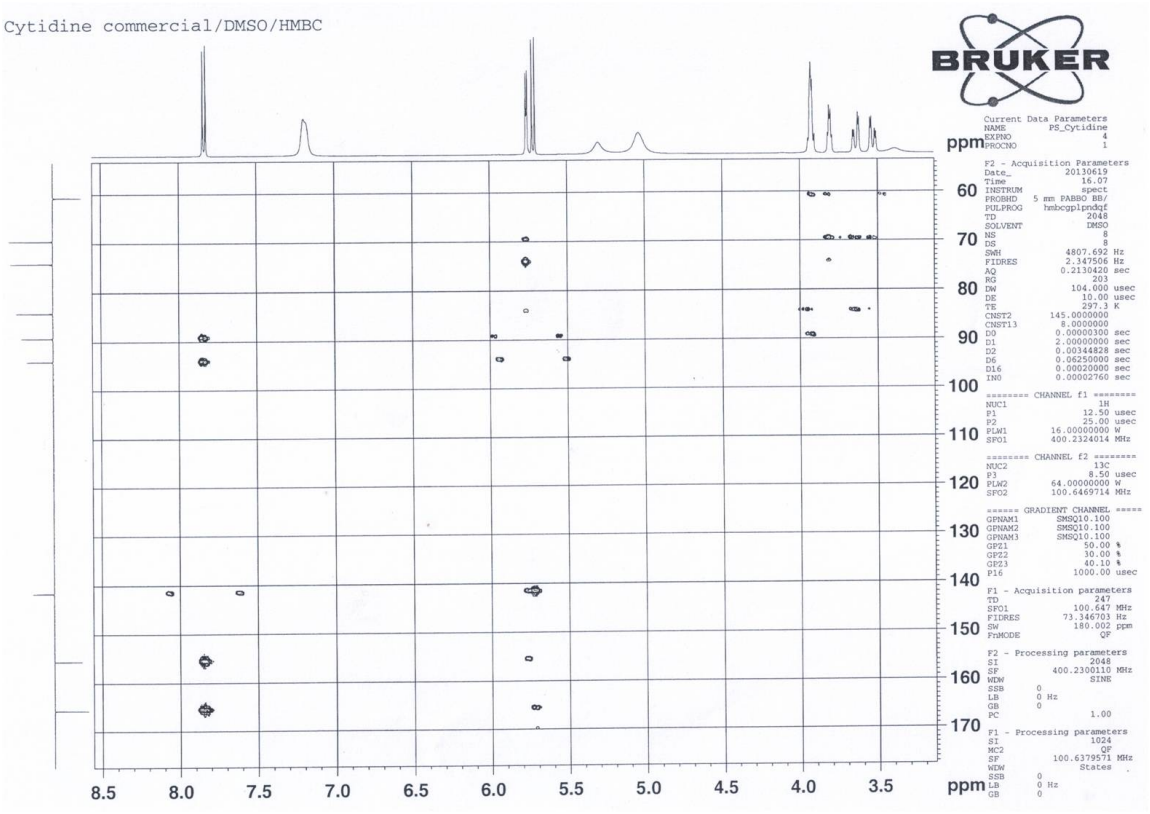


Figure S60. ^{1}H NMR spectrum of commercial sample of cytidine.

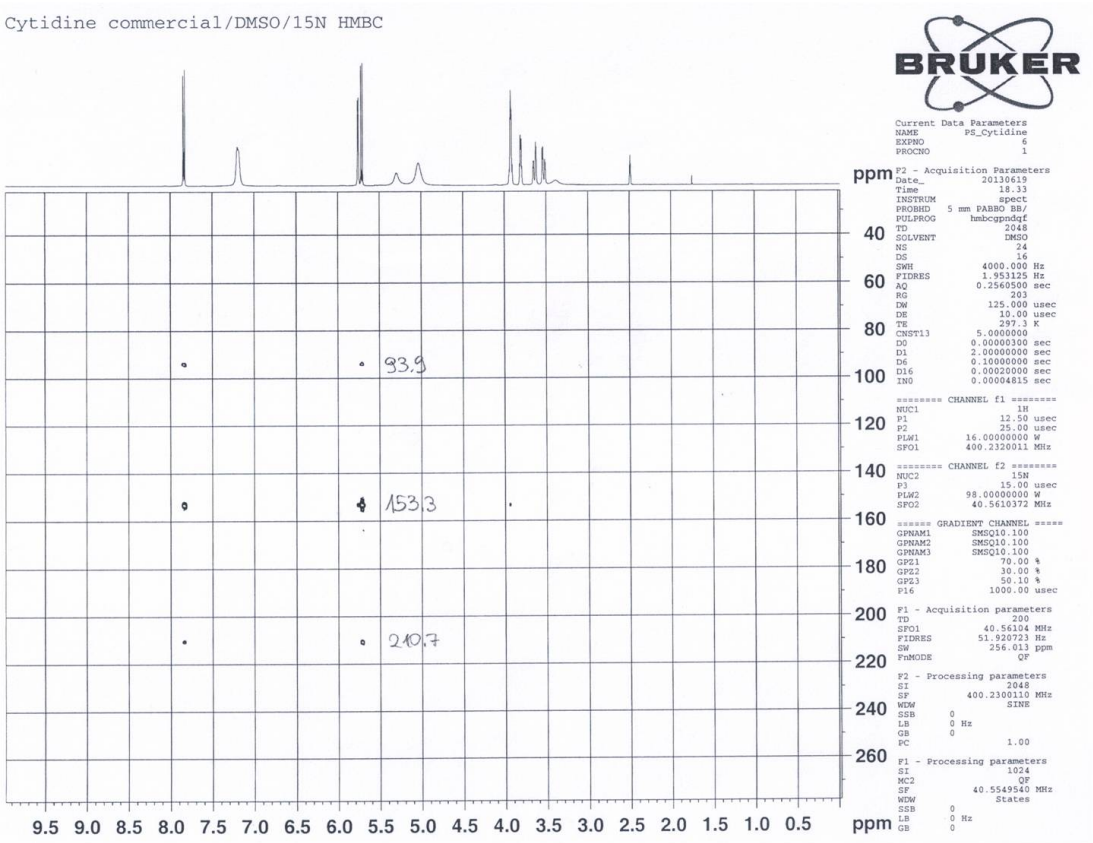


Figure S61. ^{15}N HMBC spectrum of commercial sample of cytidine.

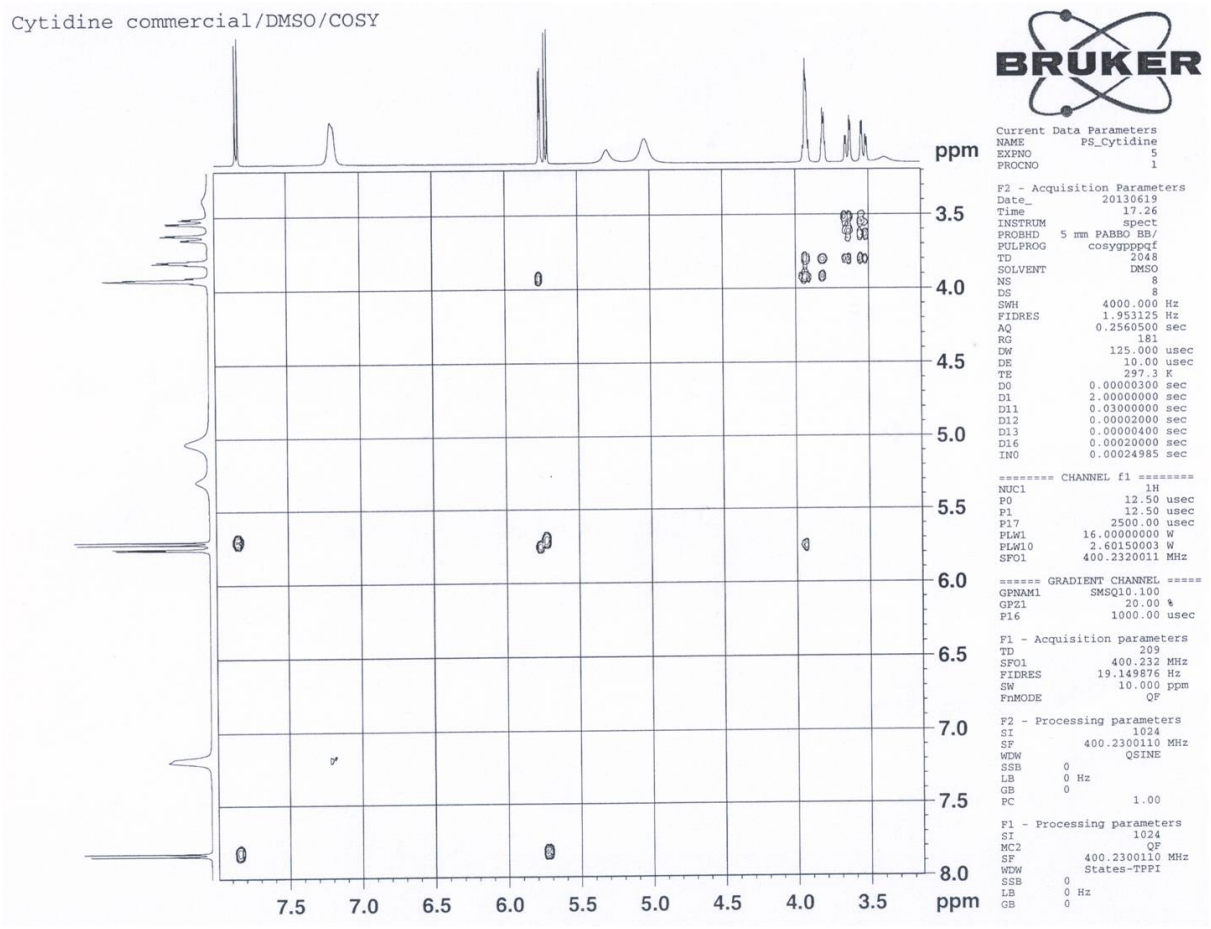


Figure S62. ^{1}H-^{1}H COSY NMR spectrum of commercial sample of cytidine.

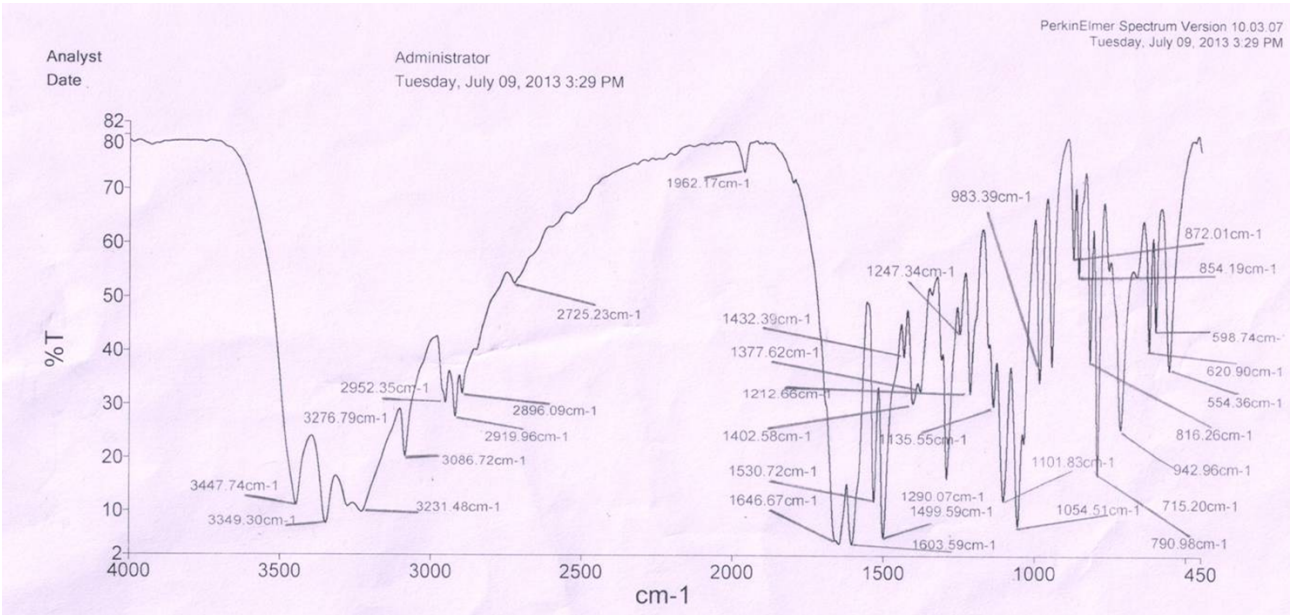
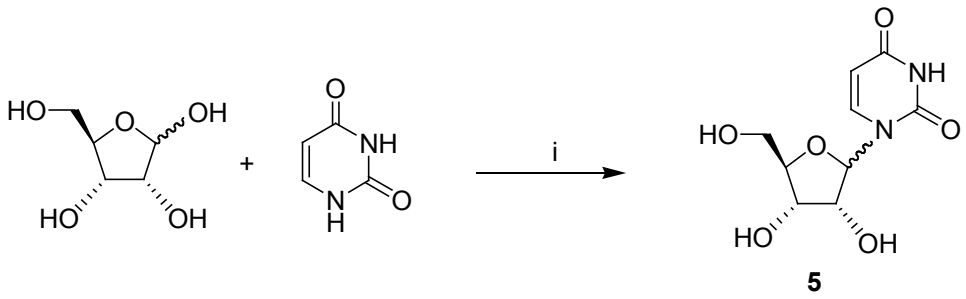


Figure S63. IR spectrum of commercial sample of cytidine.



Reaction conditions: i) Ethanol-water (1:9), 60-70 ^{o}C, 15 days

Scheme S1. Formation of uridine.

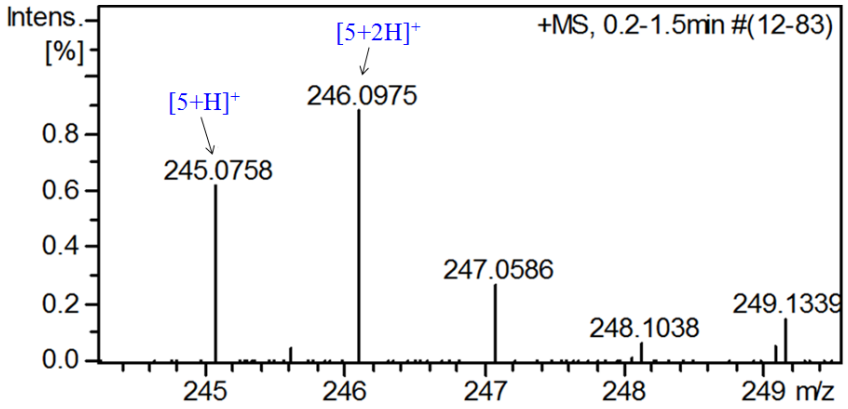


Figure S64. Mass spectrum of solution of uridine in H_2O:ACN (7:3).

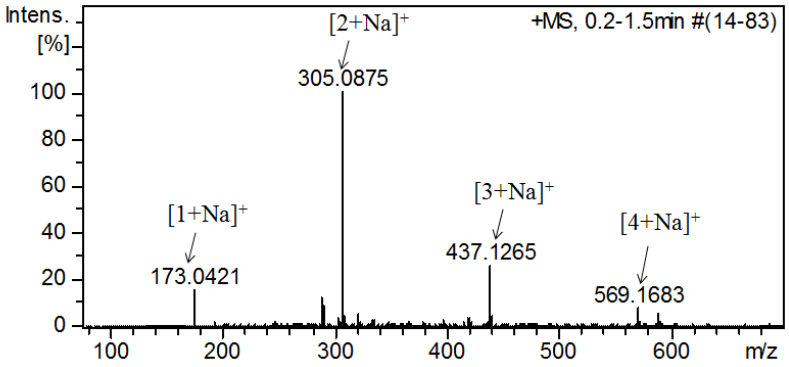


Figure S65. Complete mass spectrum of solution of uridine in H_2O:ACN (7:3) showing the formation of dimer, trimer, tetramer of D-ribose (Chart S1).

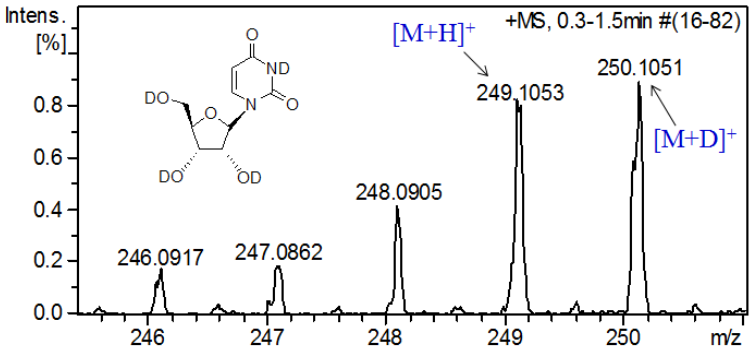
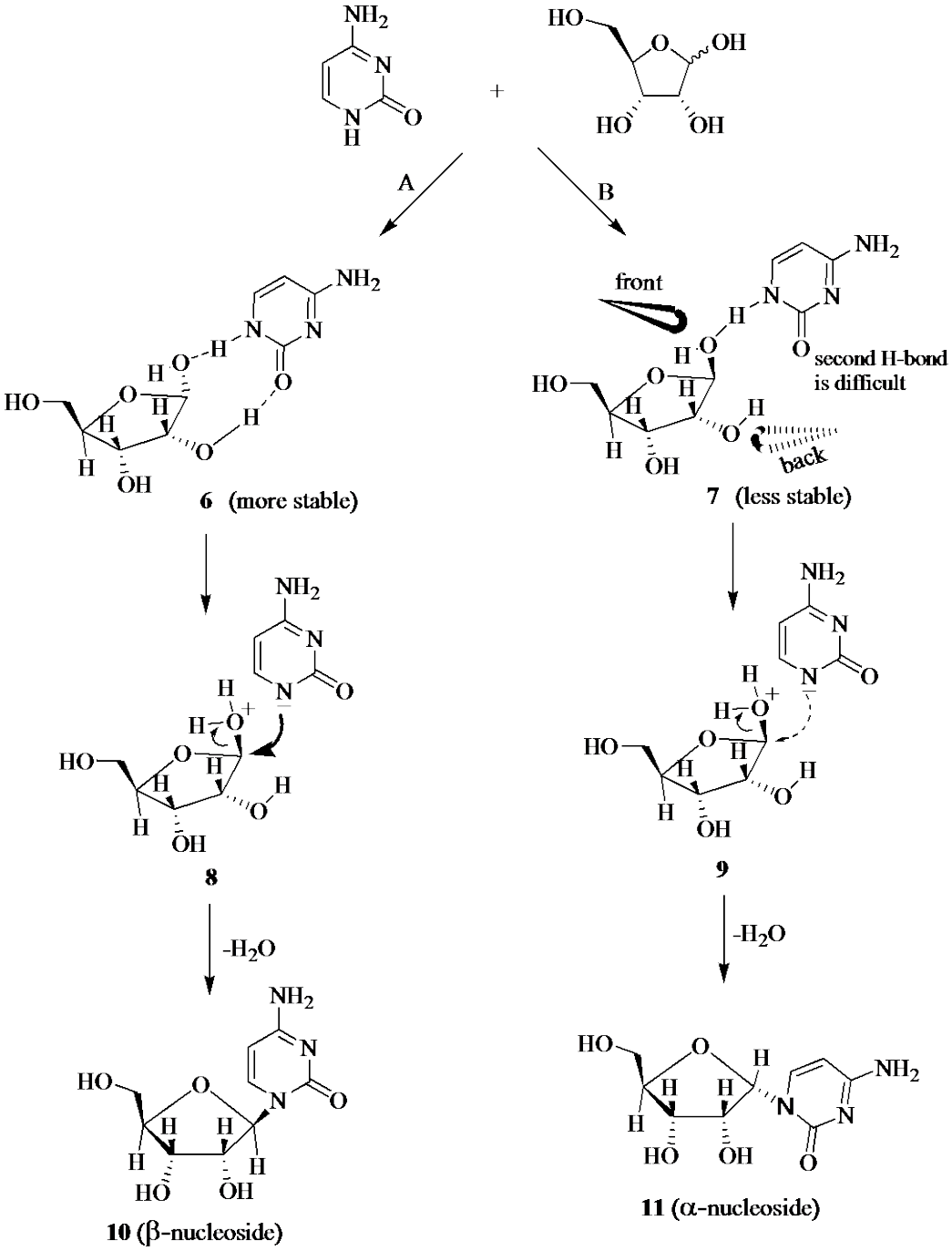


Figure S66. Mass spectrum of solution of uridine in D_2O:ACN (7:3).



Scheme S2. The plausible mechanism for glycosylation of cytosine.

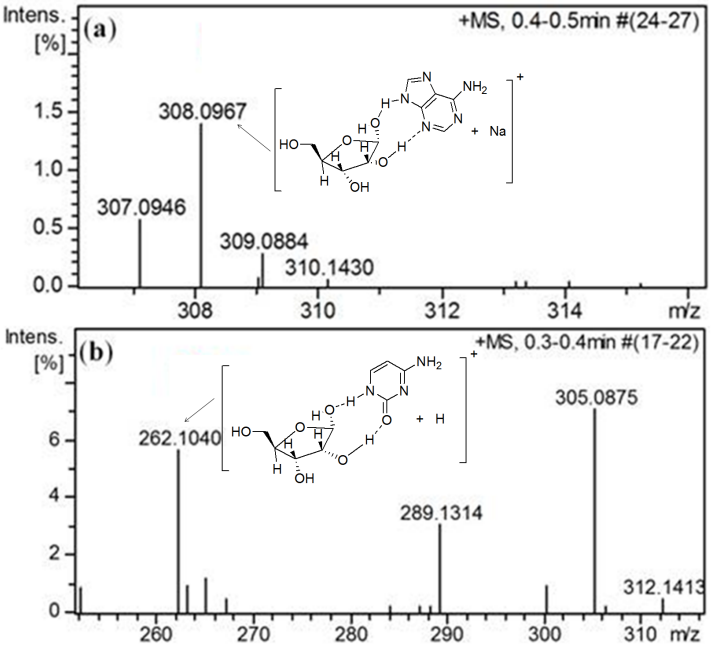


Figure S67. Mass spectrum showing formation of (a) adenine-ribose complex with *m/z* 308.0967 (calcd *m/z* 308.0965, $[M+Na]^+$). (b) cytosine-ribose complex with *m/z* 262.1040 (calcd *m/z* 262.1034, $[M+H]^+$).

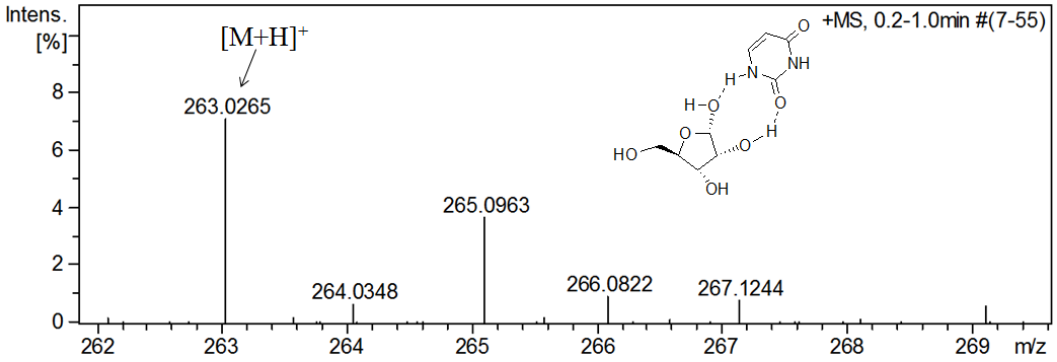


Figure S68. Mass spectrum of solution of uracil and ribose showing the formation of uracil-ribose H-bonded complex at *m/z* 263 ($[M+H]^+$). Structure is drawn on the basis of proposed mechanism for cytidine formation.

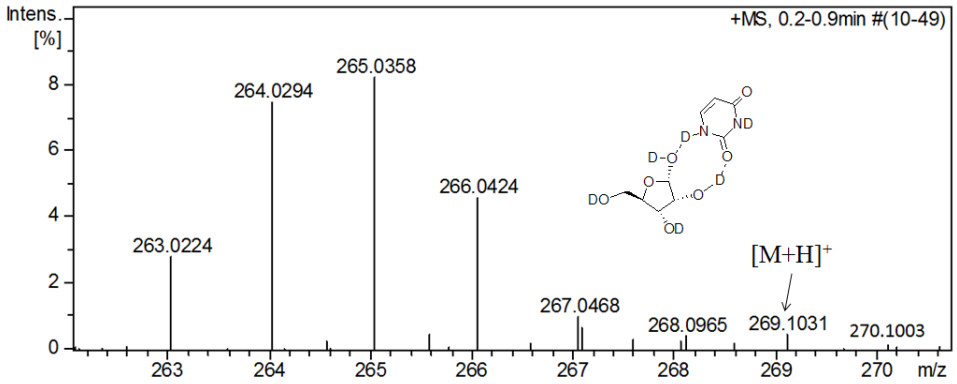


Figure S68-ii. Mass spectrum of solution of uracil and ribose in ACN:D_2O (3:7)

Mass spectra showing the formation of dimer/trimer of nucleobase and sugar

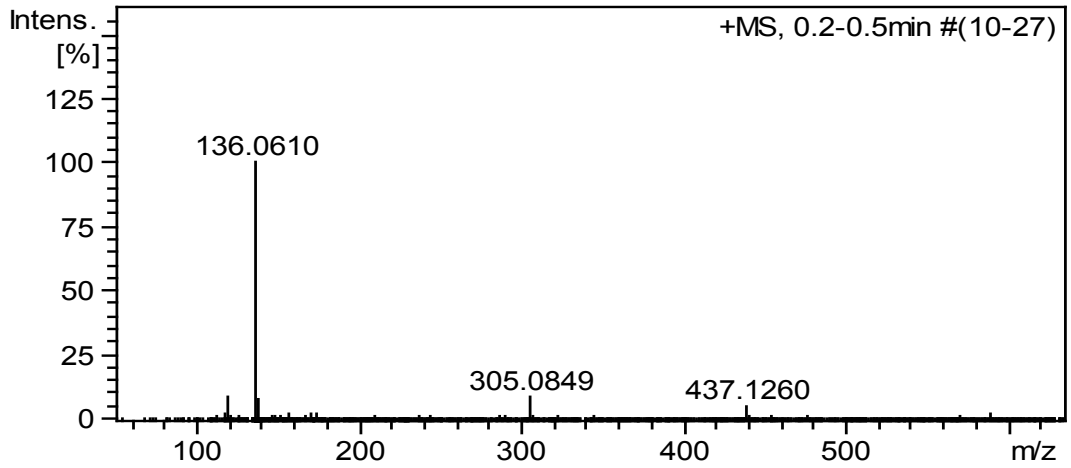


Figure S69. Mass spectrum of solution of adenine and ribose showing the formation of disaccharide and trisaccharide of D-ribose at *m/z* 305 and 437, respectively.

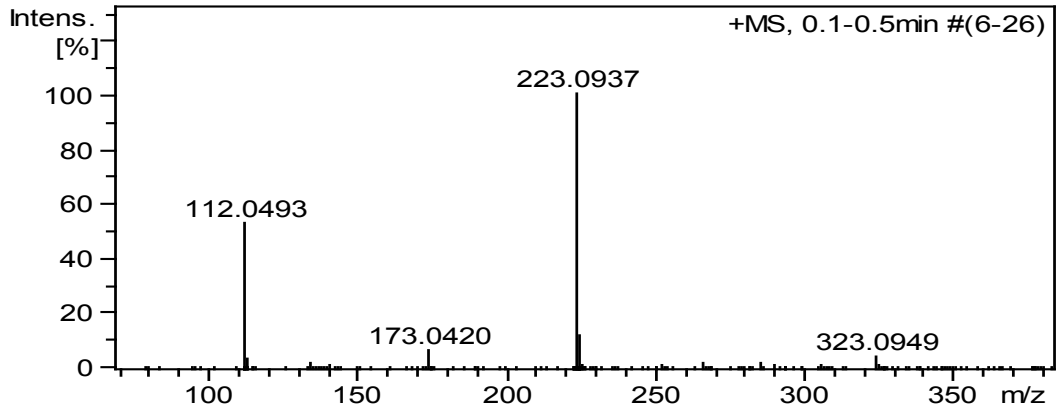


Figure S70. Mass spectrum of solution of cytosine and ribose showing the formation of dimers of cytosine and D-ribose at *m/z* 223 and 323, respectively.