Global Incorporation of Synthetic ATP Analogs Reveals Poly(A)-Dependent Differences in mRNA

Lu Zhou,[#[a, b]] Mengting Li,[#[a, c]] Junlin Wen, [#[a]] Chunlei Zhang,[[d]] Chenghe Xiong,[[a]] Wenjian Ma,[[c]] Ziwen Dai, *[[a]] Xuanjun Zhang, * [[b]] and Hui Mei*[[a]]

[a] Shenzhen Key Laboratory of Synthetic Genomics, Guangdong Provincial Key Laboratory of Synthetic Genomics, State Key Laboratory of Quantitative Synthetic Biology, Shenzhen Institute of Synthetic Biology, Shenzhen Institutes of Advanced Technology, Chinese Academy of Sciences, Shenzhen 518055, China

[b] Faculty of Health Sciences, University of Macau, Macau SAR 999078, China;

[c] College of Biotechnology, Tianjin University of Science and Technology, 300457, Tianjin, China

[d] Gingko Biotech Limited, Hong Kong SAR, 000000, China

[#] Contribute equally

* To whom correspondence should be addressed.

E-mail:

hui.mei@siat.ac.cn

Table of contents

Materials and methods

Materials and Instrumentation

Adenosine, t-butyl dimethylchlorosilane, imidazole, triethylamine trihydrofluoride, trimethyl phosphate, phosphorus oxitrichloride, tributyl ammonium pyrophosphate, acetyl chloride, propionyl chloride, butyryl chloride, valeryl chloride, hexanoyl chloride, octanoyl chloride, decanoyl chloride, benzoyl chloride, anhydrous sodium sulfate and tributylamine were purchased from Bide Pharmatech Co., Ltd. and used as received. Dichloromethane, ethyl acetate, methanol, tetrahydrofuran, acetonitrile, *N*, *N*-dimethylformaide and pyridinewere purchased from J&K Scientific Ltd. N1-methylpseudouridine triphosphate was purchased from TriLink BioTechnologies. The firefly luciferase-encoding pGL4.10[luc2] vector (Catalog number: E6551) was purchased from Promega. (Beijing) Biotech. Co., Ltd. ^{1}H, ^{13}C and ^{31}P NMR spectra were collected on Bruker Avance III400 MHz nuclear magnetic resonance spectrometer. High resolution mass spectra were obtained on a Bruker Autoflex Max Maldi-tof Mass spectrometer. Agarose gel was stained with Invitrogen SYBR Gold Nucleic Acid Stain and imaged with a Molecular Imager Gel Doc XR^{+} Imaging System.

Synthesis and characterization of chemically modified adenosine derivatives and their triphosphates

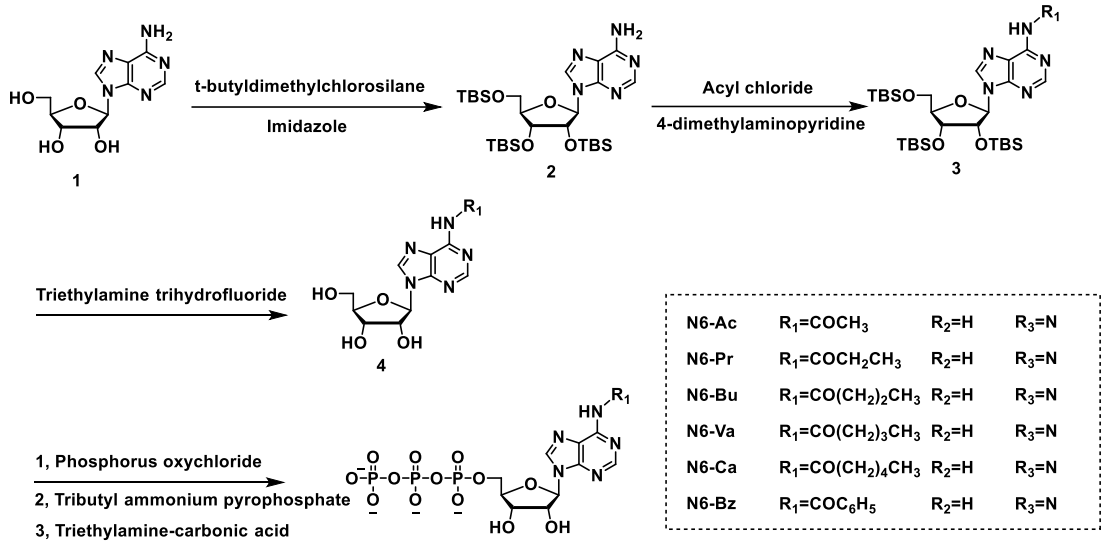


Figure S1. The synthetic scheme for the chemically modified adenosine triphosphate.

General experimental procedures

To a 250-mL round bottom flask equipped with a magnetic stir bar, Adenosine (8.0 g, 30 mmol, 1.0 equivalents [eq].) was dissolved in 150 mL of *N, N*-dimethylformamide, followed by the addition of t-butyl dimethyl chlorosilane (15.83 g, 105 mmol, 3.5 eq.) and imidazole (10.21 g, 150 mmol, 5.0 eq.). The reaction mixture was stirred at room temperature overnight. Next, the mixture was concentrated under vacuum, and redissolved in 100 mL of ethyl acetate, washed with water (100 mL) for 3 times. The organic phase was collected and dried over anhydrous Na_2SO_4, filtered and concentrated under reduced pressure. The crude product was isolated through a silica gel column chromatography (methanol:dichloromethane = 1:100) to afford **2** as white solid: yield 16.47g, (90%).

Compound **2** (3.05 g, 5 mmol, 1.0 eq.) was dissolved in 20 mL of anhydrous pyridine and then cooled to 0 °C, corresponding acyl chloride (6.0 mmol, 1.2eq.) was added dropwise. After that, the reaction mixture was stirred at room temperature for 4-6 h. The mixture was concentrated under vacuum, and redissolved in 20 mL of ethyl acetate, washed with 1 N HCl (3 × 20 mL) and washed with saturated aqueous. $NaHCO_3$ (20 mL). The organic phase was collected and dried over anhydrous Na_2SO_4, filtered and concentrated under reduced pressure. The crude product **3** was directly used for the next reaction without further purification.

Compound **3** (2 mmol, 1.0 eq.) was dissolved in 15 mL of tetrahydrofuran. Then, triethylamine trihydrofluoride (8 mmol, 4.0 eq.) was added, and the reaction mixture was stirred at room temperature overnight. After that, the reaction mixture was concentrated under reduced pressure and the product was isolated through a silica gel column chromatography (methanol:dichloromethane = 1:20 to 1:6) to given compound **4**. Next, compound **4** (0.1 mmol, 1.0 eq.) was co-evaporated with anhydrous pyridine (2 mL) for 3 times, and then, 1 mL of anhydrous trimethyl phosphate (stored·over 4·Å molecular sieves) was added under N_2 atmosphere. After the complete dissolution of reaction mixtures, the reaction was put in ice-water bath for 5 min and phosphorus oxychloride (15 μL, 0.15 mmol, 1.5 eq.) was added dropwise and the mixture was stirred in ice-water bath for 2.5 h. Afterward, anhydrous tributylamine (0.35 mL, 1.5 mmol, 15.0 eq., stored over potassium hydroxide) and tributylammonium pyrophosphate solution (0.5M in anhydrous *N,N*-dimethylformamide, 1.6 mL, 0.8 mmol, 8.0 eq.) were added and the reaction mixture was stirred at room temperature for another 0.5 h. The reaction was quenched by TEAB buffer (2.0 M, 5 mL) at 0 °C. After 10 min, the aqueous phase was washed with dichloromethane (10 mL) for 3 times, and purified by HPLC with C18 column and 0.05 M TEAB buffer and HPLC grade acetonitrile as the eluent. The collected solution was lyophilized to afford the products (**ac6A, pr6A, bu6A, va6A, ca6A, and bz6A**). The final products were confirmed by ^{1}H and ^{31}P NMR.

ac6A

^{1}H NMR (400 MHz, Deuterium Oxide) δ 8.72 (s, 1H), 8.60 (s, 1H), 6.16 (d, J = 5.6 Hz, 1H), 4.80 – 4.75 (m, 1H), 4.59 (s, 1H), 4.31 (s, 1H), 4.27 – 4.20 (m, 1H), 4.16 – 4.07 (m, 1H), 2.26 (s, 3H).

^{31}P NMR (162 MHz, Deuterium Oxide) δ -6.41 (d, J = 20.8 Hz), -11.38 (d, J = 20.5 Hz), -22.58 (t, J = 20.5 Hz).

pr6A

^{1}H NMR (400 MHz, Deuterium Oxide) δ 8.73 (s, 1H), 8.62 (s, 1H), 6.18 (d, J = 5.7 Hz, 1H), 4.79 (t, J = 5.5 Hz, 1H), 4.64 – 4.56 (m, 1H), 4.36 – 4.30 (m, 1H), 4.30 – 4.20 (m, 1H), 4.13 (dt, J = 11.6, 3.2 Hz, 1H), 2.57 (q, J = 7.5 Hz, 1H), 1.14 (s, 3H).

^{31}P NMR (162 MHz, Deuterium Oxide) δ -6.71 (d, J = 23.1 Hz), -11.36 (d, J = 20.3 Hz), -22.61 (t, J = 20.4 Hz).

bu6A

^{1}H NMR (400 MHz, Deuterium Oxide) δ 8.72 (s, 1H), 8.61 (s, 1H), 6.17 (d, J = 5.8 Hz, 1H), 4.82 – 4.77 (m, 1H), 4.60 (t, J = 4.3 Hz, 1H), 4.34 – 4.31 (m, 1H), 4.28 – 4.20 (m, 1H), 4.13 (dt, J = 11.7, 3.3 Hz, 1H), 2.51 (t, J = 7.4 Hz, 2H), 1.74 – 1.63 (m, 2H), 0.93

(t, J = 7.4 Hz, 3H).

^{31}P NMR (162 MHz, Deuterium Oxide) δ -6.39 (d, J = 21.0 Hz), -11.36 (d, J = 20.0 Hz), -22.56 (t, J = 20.5 Hz).

va6A

^{1}H NMR (400 MHz, Deuterium Oxide) δ 8.73 (s, 1H), 8.62 (s, 1H), 6.18 (d, *J* = 5.7 Hz, 1H), 4.79 (td, *J* = 5.7, 1.7 Hz, 1H), 4.63 – 4.58 (m, 1H), 4.33 (s, 1H), 4.28 – 4.21 (m, 1H), 4.17 – 4.09 (m, 1H), 2.55 (t, *J* = 7.5 Hz, 2H), 1.71 – 1.60 (m, 2H), 1.40 – 1.29 (m, 2H), 0.86 (t, *J* = 7.4 Hz, 3H).

^{31}P NMR (162 MHz, Deuterium Oxide) δ -6.52 (d, J = 20.7 Hz), -11.36 (d, J = 20.2 Hz), -22.57 (t, J = 20.5 Hz).

ca6A

^{1}H NMR (400 MHz, Deuterium Oxide) δ 8.73 (s, 1H), 8.62 (s, 1H), 6.17 (d, J = 5.8 Hz, 1H), 4.81 – 4.77 (m, 1H), 4.62 – 4.59 (m, 1H), 4.35 – 4.30 (m, 1H), 4.29 – 4.20 (m, 1H), 4.13 (dt, J = 11.8, 3.4 Hz, 1H), 2.54 (t, J = 7.5 Hz, 2H), 1.73 – 1.63 (m, 2H), 1.34 – 1.26 (m, 4H), 0.81 (t, J = 7.0 Hz, 3H).

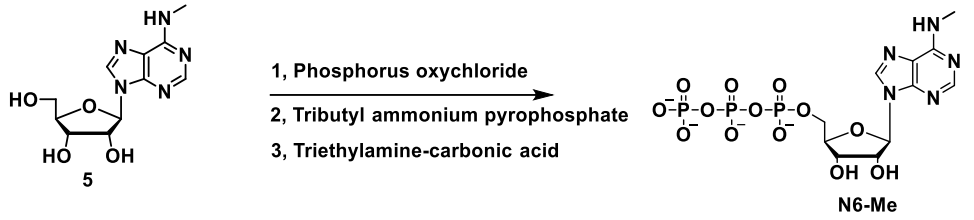
^{31}P NMR (162 MHz, Deuterium Oxide) δ -6.51 (d, J = 20.8 Hz), -11.35 (d, J = 20.3 Hz), -22.56 (t, J = 20.5 Hz).

bz6A

^{1}H NMR (400 MHz, Deuterium Oxide) δ 8.62 (s, 1H), 8.56 (s, 1H), 7.88 (d, J = 7.9 Hz, 2H), 7.51 (t, J = 6.8 Hz, 1H), 7.43 (t, J = 7.3 Hz, 1H), 6.12 (d, J = 5.7 Hz, 1H), 4.79 – 4.73 (m, 1H), 4.59 – 4.52 (m, 1H), 4.28 (s, 1H), 4.24 – 4.16 (m, 1H), 4.12 – 4.04 (m, 1H).

^{31}P NMR (162 MHz, Deuterium Oxide) δ -7.62 (d, J = 22.1 Hz), -11.40 (d, J = 20.2 Hz), -22.78 (t, J = 20.4 Hz).

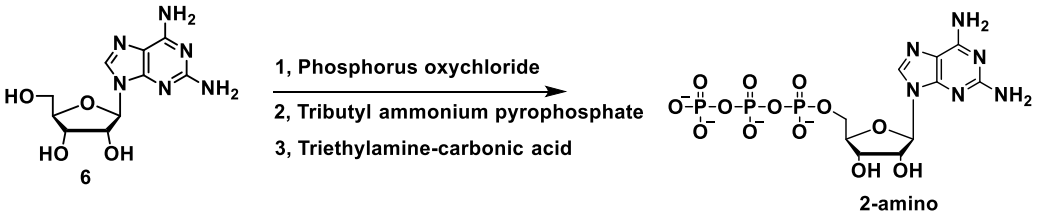
Synthesis of Non-acylated Triphosphates



Compound **m6A** was prepared following the synthetic procedure of molecule **ac6A**.

^{1}H NMR (400 MHz, Deuterium Oxide) δ 8.37 (s, 1H), 8.12 (s, 1H), 5.98 (d, J = 5.8 Hz, 1H), 4.70 – 4.68 (m, 1H), 4.51 (s, 1H), 4.24 (s, 1H), 4.20 – 4.12 (m, 1H), 4.07 – 4.00 (m, 1H), 2.96 (s, 3H).

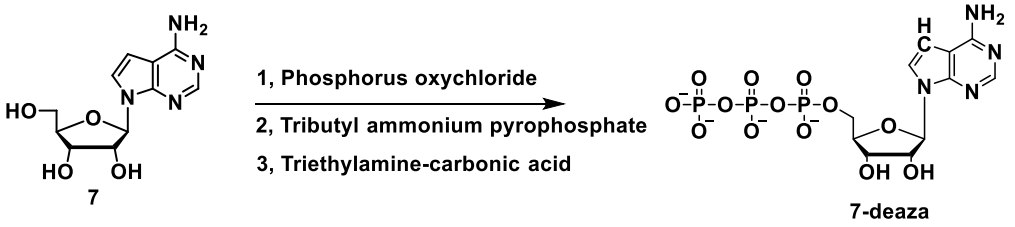
^{31}P NMR (162 MHz, Deuterium Oxide) δ -7.05 (d, J = 21.3 Hz), -11.39 (d, J = 20.2 Hz), -22.68 (t, J = 20.4 Hz).



Compound **am2A** was prepared following the synthetic procedure of molecule **ac6A**.

^{1}H NMR (400 MHz, Deuterium Oxide) *δ* 8.10 (s, 1H), 5.83 (d, *J* = 6.4 Hz, 1H), 4.68 – 4.64 (m, 1H), 4.50 – 4.45 (m, 1H), 4.24 (s, 1H), 4.19 – 4.12 (m, 1H), 4.10 – 4.02 (m, 1H).

^{31}P NMR (162 MHz, Deuterium Oxide) δ -6.40 (d, J = 20.8 Hz), -11.37 (d, J = 19.9 Hz), -22.59 (t, J = 20.5 Hz).



Compound **c7A** was prepared following the synthetic procedure of molecule **ac6A**.

^{1}H NMR (400 MHz, Deuterium Oxide) δ 8.05 (s, 1H), 7.51 (d, J = 3.6 Hz, 1H), 6.61 (d, J = 3.5 Hz, 1H), 6.15 (d, J = 6.9 Hz, 1H), 4.68 – 4.63 (m, 1H), 4.53 – 4.48 (m, 1H), 4.22 (s, 1H), 4.20 – 4.13 (m, 1H), 4.05 – 3.98 (m, 1H).

^{31}P NMR (162 MHz, Deuterium Oxide) δ -7.14 (d, J = 20.7 Hz), -11.41 (d, J = 20.0 Hz), -22.75 (t, J = 20.4 Hz).

mRNA synthesis and characterization

The DNA plasmids both with or without 110 nt poly(A) tail from VectorBuilder were linearized by BspQI (NEB, R0712S) according to the manufacturer's instructions, and the DNA plasmid pGL4.10 from Promega was PCR amplified, purified using HipureGel pure DNA mini kit (Magen, D211103). These linearized plasmids were characterized with agarose gel electrophoresis. mRNA constructs were synthesized by IVT using the T7 RNA polymerase (NEB, M0251L) with a reaction mixture containing linearized DNA template, 5 mM ATP, 5 mM UTP, 5 mM CTP, 1 mM GTP, 4 mM ARCA (NEB, S1411S), transcription buffer, and SUPERase-In RNase inhibitor (ThermoFisher Scientific, AM2694). Following transcription, DNase I (NEB, M0303S) was added to remove residual DNA, and the mRNA was purified using Monarch® Spin RNA Cleanup Kit (NEB, T2030L).

For polyadenylation, E. coli Poly(A) Polymerase (NEB, M0276S) was used in a reaction which containing untailed mRNA or short RNA, natural or base-modified ATP analogs, polymerase buffer, and SUPERase-In RNase inhibitor, incubated at 37 °C for 30 min. The Polyadenylated mRNA was then purified by ethanol precipitation method. The final mRNA was resuspended in RNase-free water, and the mRNA concentrations were quantified using the Qubit RNA assay (ThermoFisher Scientific, Q10210).

Luciferase Assay in Rabbit Reticulocyte Lysate

Protein yield was evaluated in a cell-free eukaryotic translation system using Rabbit Reticulocyte Lysate (RRL; Promega, L4960) according to the manufacturer's instructions. Briefly, lysate was combined with mRNA, an amino acid mixture lacking methionine, SUPERase-In RNase inhibitor, and incubated at 30 °C for 90 min. Translation efficiency was determined with the Luciferase Assay System (Promega, E1500) using a luminometer, and quantified as relative light units (RLU).

Luciferase Assay in Mammalian Cells

HeLa cells were maintained in DMEM culture media (ThermoFisher Scientific, 119951) containing 10% FBS and 1% penicillin–streptomycin (YEASEN, 60162ES76) in a 37 °C incubator with 5% CO_2. mRNA was transfected with Lipofectamine™ MessengerMAX (Thermo Fisher Scientific, LMRNA001) at equal mass following the manufacturer's protocol for 6 h, followed by an additional 24 h incubation. For luciferase assay, cells were lysed and firefly luciferase activity was measured using the Promega Rapid Detection Luciferase Assay System (Promega, E1500) at indicted time points.

mRNA stability quantified by RT–qPCR

Cells were transfected as described above. At the indicated time points (6 h and 24 h), cells were washed with ice-cold PBS and total RNA was isolated using MolPure® Flash CellTissue Total RNA Kit (YEASEN,19221ES50). RNA was reverse-transcribed using a mixed primer strategy (). qPCR was performed with SYBR Green chemistry on a real-time PCR instrument (e.g., QuantStudio), including melt-curve analysis. Primer sequences are listed in Table.1. Cq values were normalized to GAPDH, and relative mRNA abundance was calculated by the 2^–ΔΔCq method. Data represent mean ± SD from ≥3 biological replicates with technical duplicates.

Table.1 Primer sequences used for qPCR

Primer	Sequence (5'-3')
GAPDH-F	GAAGGTGAAGGTCGGAGTCAAC
GAPDH-R	CAGAGTTAAAAGCAGCCCTGGT
Luciferase-F	AGAGATACGCCCTGGTTCCT
Luciferase-R	TCCGATAAATAACGCGCCCA

Immunogenicity assay in RAW264.7 cells

RAW264.7 macrophages were maintained as described above. Cells in 48-well plates (2×10^4 cells/ well) were treated with R-848 (positive control), lipoMAX alone (mock, negative control), or complexed with mRNA for 24 h to harvest culture medium. The collected medium was subjected cytokine ELISA. The levels of IL-6 (LiankeBio, EK106) and TNF-α (LiankeBio, EK282) were measured in supernatants by ELISA. Cultures were performed in triplicate to quadruplicate and measured in duplicate. Absorbance at 450 nm (reference 570 nm) was measured using a microplate reader (e.g., BioTek Synergy). Cytokine concentrations were calculated from standard curves prepared by serial dilution of supplied standards.

Fluorescent Oligo(dT) Hybridization Assay

To confirm the successful polyadenylation of chemically base-modified ATP at the 3' end of mRNA, a hybridization assay was performed using Cy3-labeled 25 nt oligo dT (oligo dT-Cy3). The reaction mixture contained tailed mRNA and 50 pmol of oligo dT-Cy3 was incubated at room temperature for 30 min. The fluorescence signal was detected using a fluorescent gel imaging system (excitation: 550 nm, emission: 570 nm) to assess the hybridization efficiency. The appearance of fluorescence indicates successful hybridization, which indicated the presence of poly(A) tails at the 3' end of mRNA.

Supplementary figures

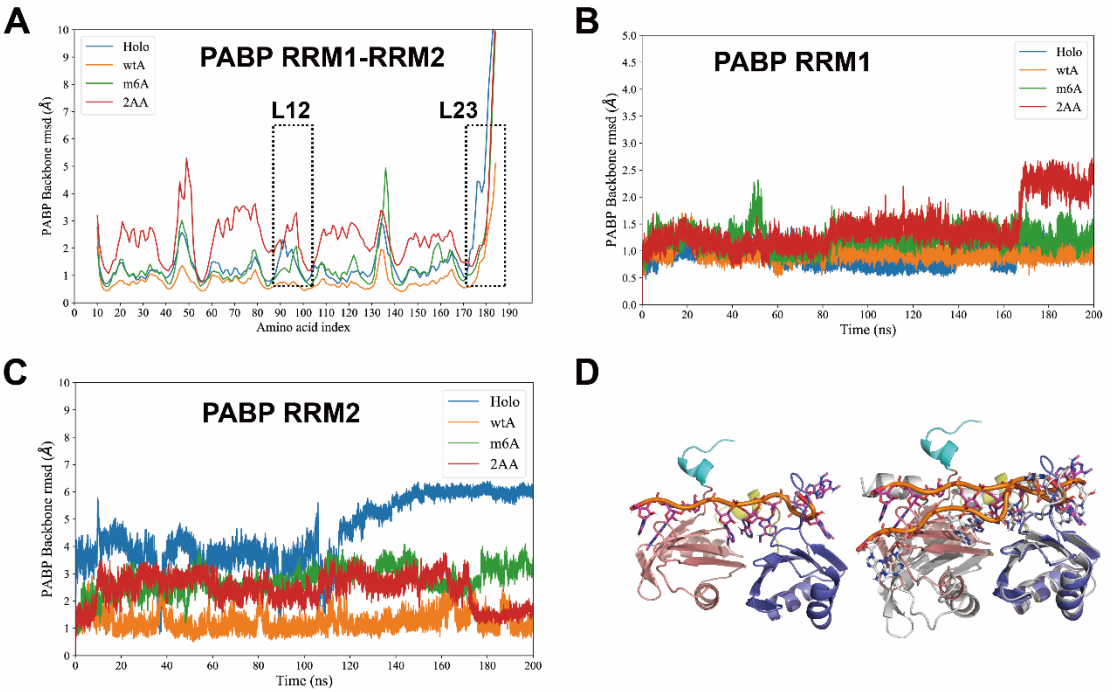


Figure S1. Molecular dynamics analysis of PABP RRM1-RRM2 in complex with natural and base-modified poly(A). (**A**) Per-residue backbone RMSD of PABP RRM1–RRM2 averaged over the simulation trajectory. Regions corresponding to the inter-domain linker (L12) and RRM2 C-terminal segment (L23) are highlighted. (**B**) Backbone RMSD of the isolated RRM1 domain over 200 ns simulation. (**C**) Backbone RMSD of the isolated RRM2 domain over 200 ns simulation. (**D**) Representative structures of the protein-polynucleotides PABP, poly(m6A) extracted from the last 10 ns trajectories of MD simulations. The left column is the protein/complex from MD simulations, while the right column shows the structural compassion to the crystal structure (PDB ID 4F02, colored in light gray) by superimposing the RRM1 domains. The protein is depicted as cartoon models with RRM1, L12 linker, RRM2 and L23 linker colored in tv-blue, yellow, pink and cyan, respectively. The backbone of polynucleotides is colored in orange with carbons of bases colored in purple.

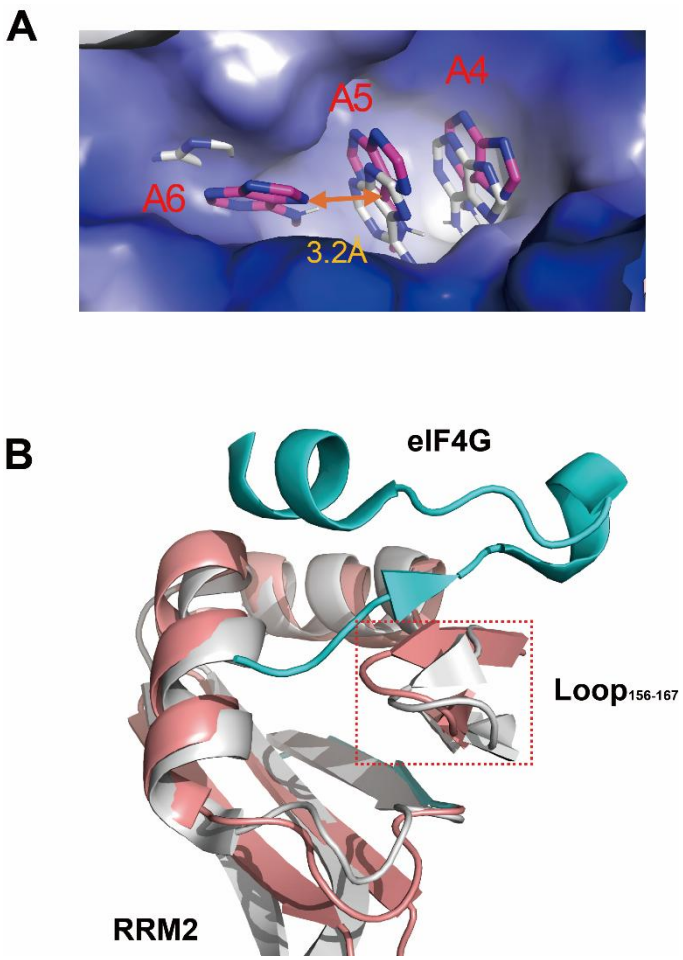


Figure S2. Atomic mechanisms of PABP disruption by adenine modifications. (**A**) Binding poses of the central three nucleobases (A4, A5 and A6) in complexes of Poly(m6A): PABP, Only the nucleobases were shown. (**B**) Structural changes of the eIF4G binding site of PABP. The RRM2 domains (colored in pink) from the Poly(m6A): PABP complex were superimposed to the cocrystal structure (PDB code 4F02) of Poly(A): PABP: eIF4G. The PABP in crystal structure were colored in gray and the eIF4G peptide was colored in cyan. The loop between residues E156 and K167 were indicated by a red box.

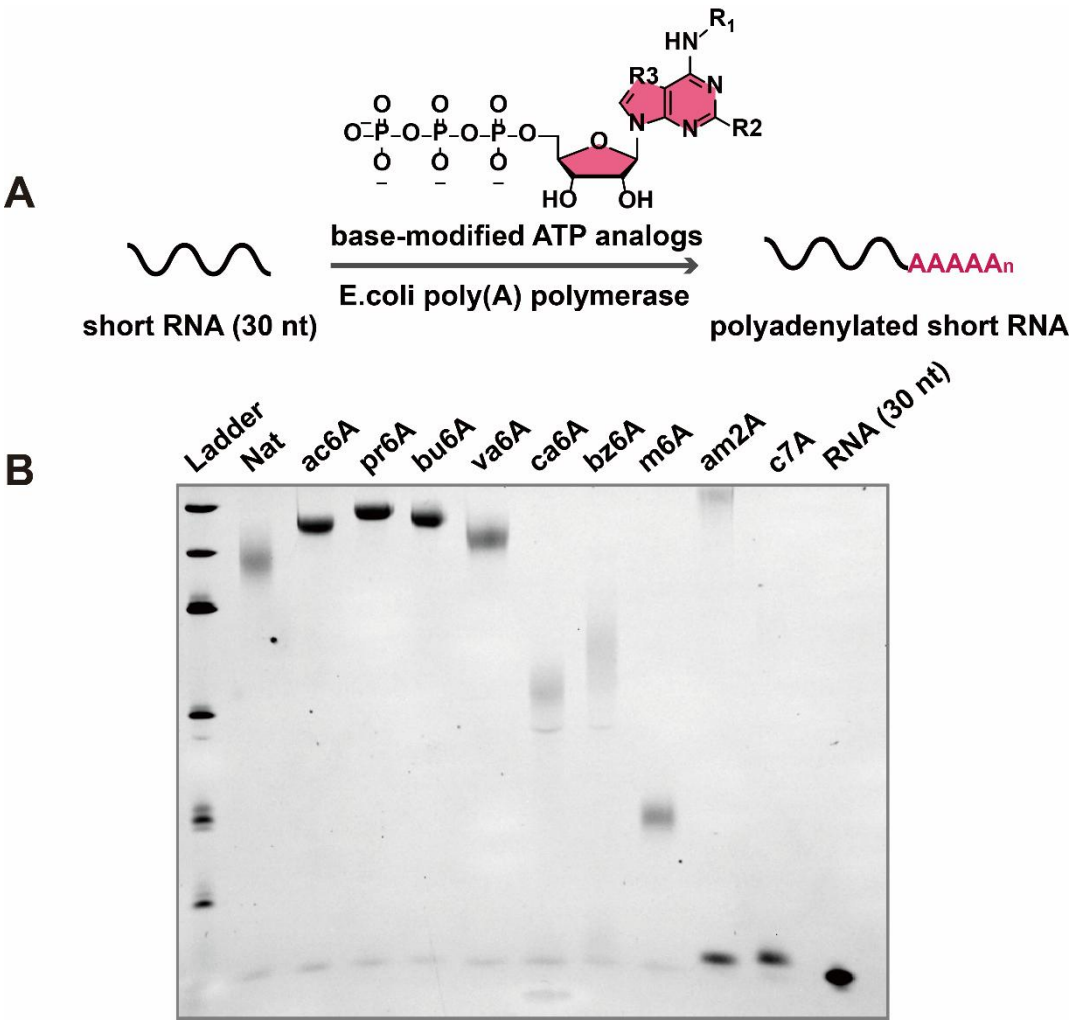


Figure S3. Enzymatic Synthesis of Polyadenylated Short RNAs using Base-Modified ATP Analogs. (**A**) Schematic illustration of the enzymatic polyadenylation process. A short 30 nt RNA primer was incubated with various base-modified ATP analogs in the presence of E. *coli* poly(A) Polymerase. This reaction extends the RNA at the 3' terminus to form modified poly(A) tails. (**B**) 10% denaturing PAGE analysis of polyadenylated products. The gel image demonstrates that most modified ATP analogs (ac6A, pr6A, bu6A, va6A, ca6A, bz6A, m6A and am2A) are successfully utilized by the polymerase as substrates, except c7A supported efficient tailing.

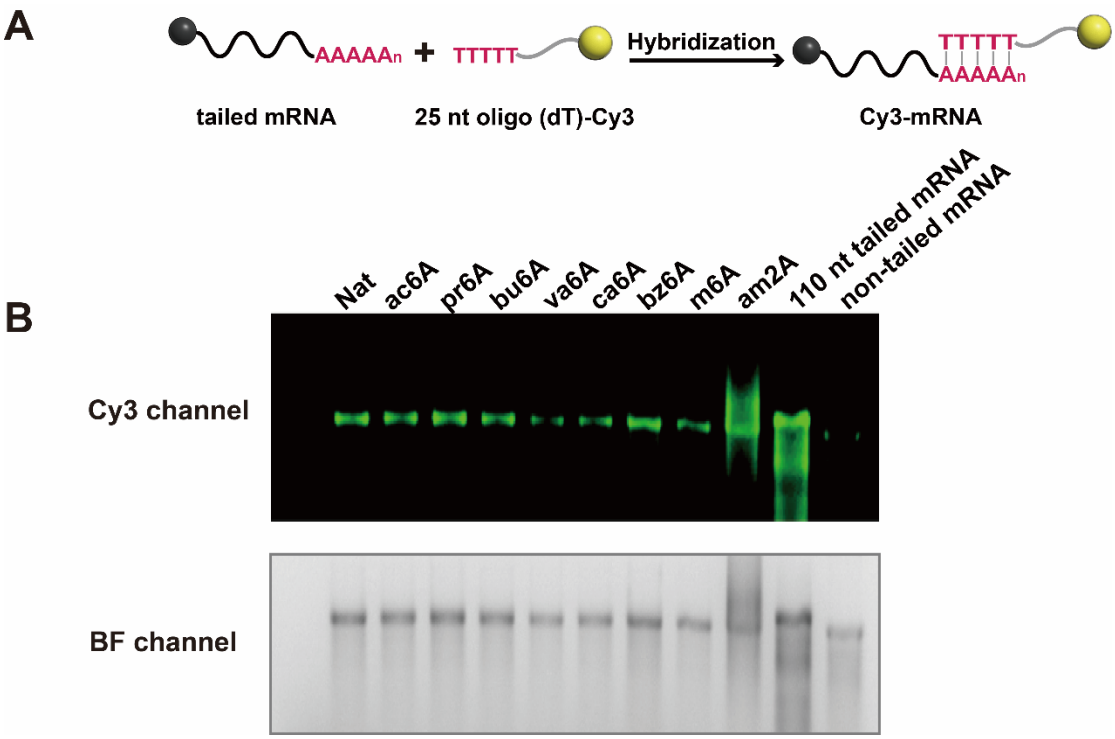
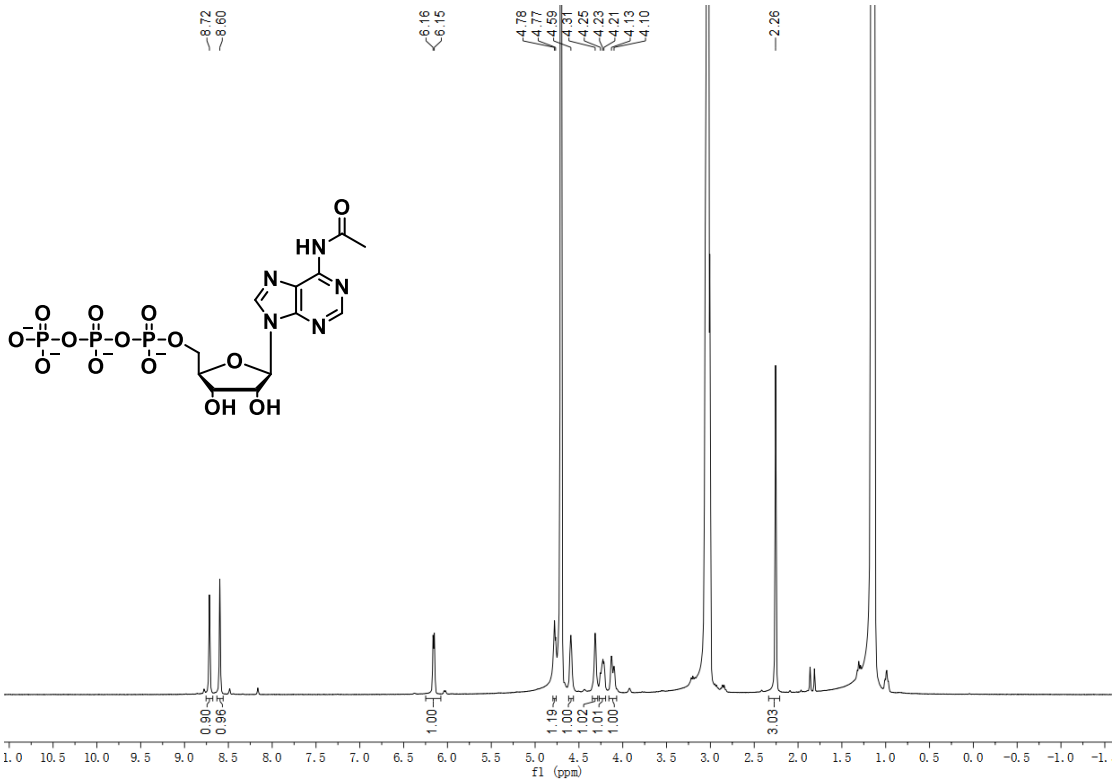


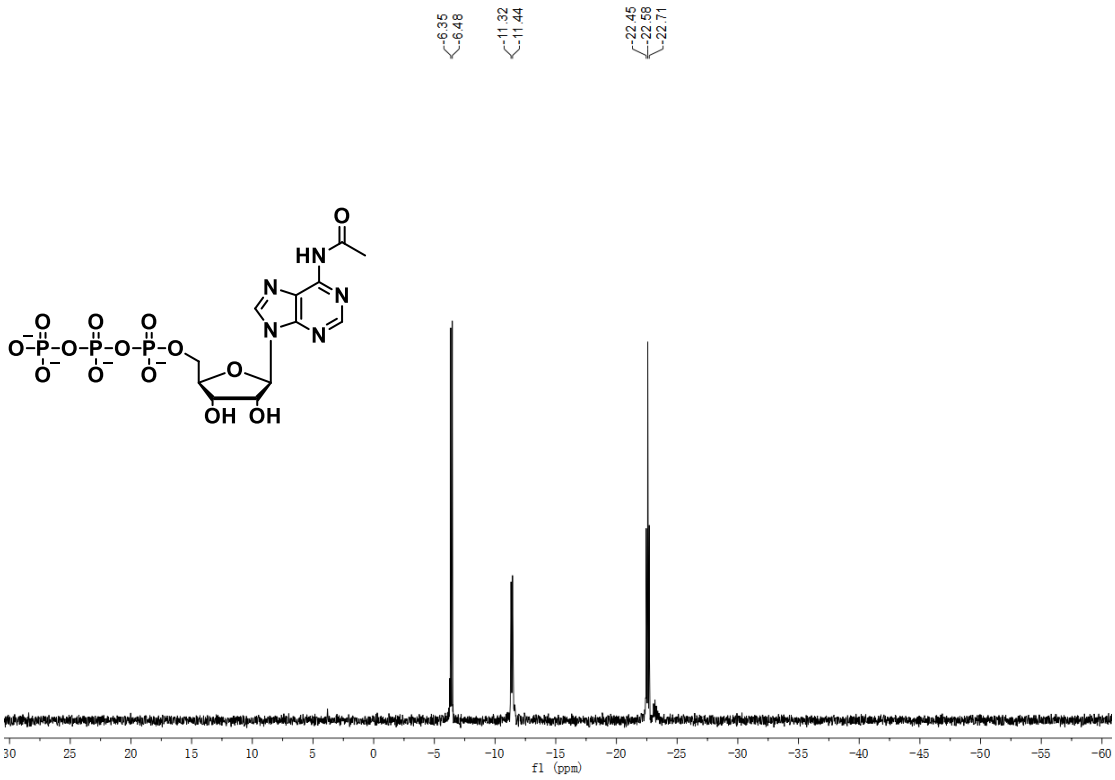
Figure S4. Validation of Poly(A) Tail Presence in Base-Modified mRNAs via Oligo(dT) Hybridization. (**A**) Schematic illustration of the hybridization assay. Capped and polyadenylated with base-modified ATP analogs mRNAs were hybridized with a Cy3-labeled 25 nt oligo(dT) probe. The probe specifically binds to the poly(A) tail, allowing for the visual confirmation of tail presence through fluorescence. (**B**) Gel electrophoresis analysis of hybridized mRNAs. Cy3 channel: The fluorescent bands confirm that all modified mRNAs (ac6A, pr6A, bu6A, va6A, ca6A, bz6A, m6A and am2A) successfully incorporated a poly(A) tail using E. *coli* poly(A) polymerase. BF (Bright Field) channel: Shows the total RNA loading for each sample.

NMR spectra

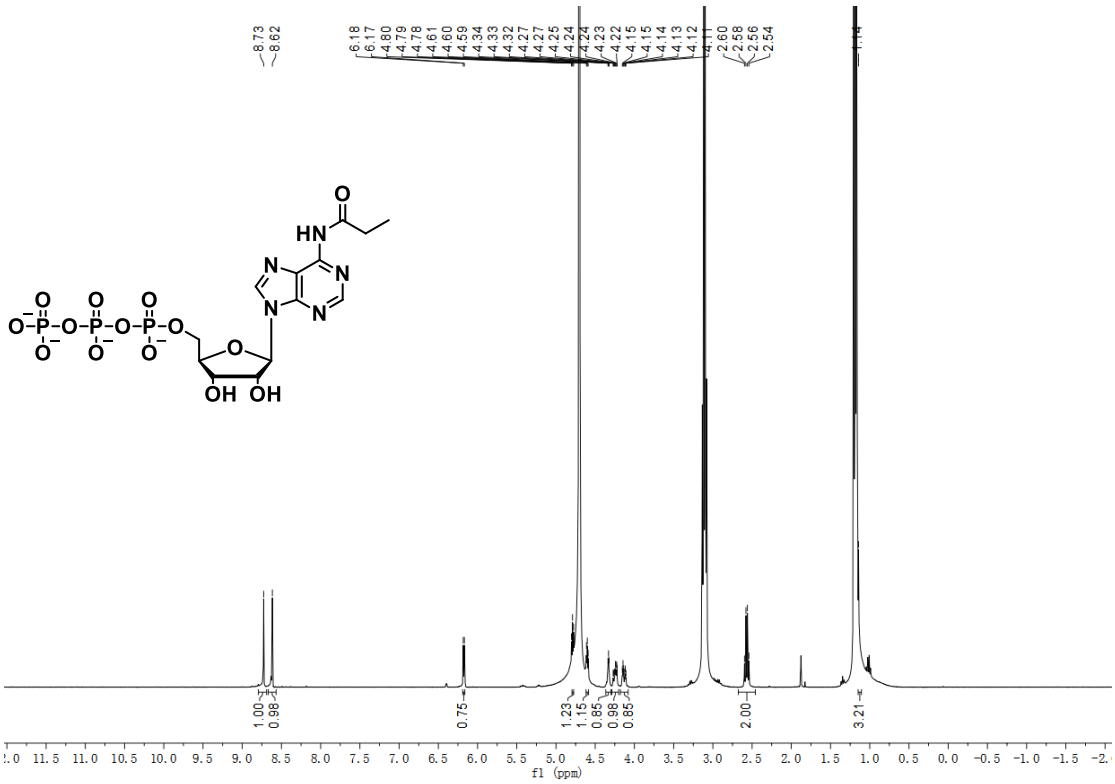
ac6A ^{1}H-NMR



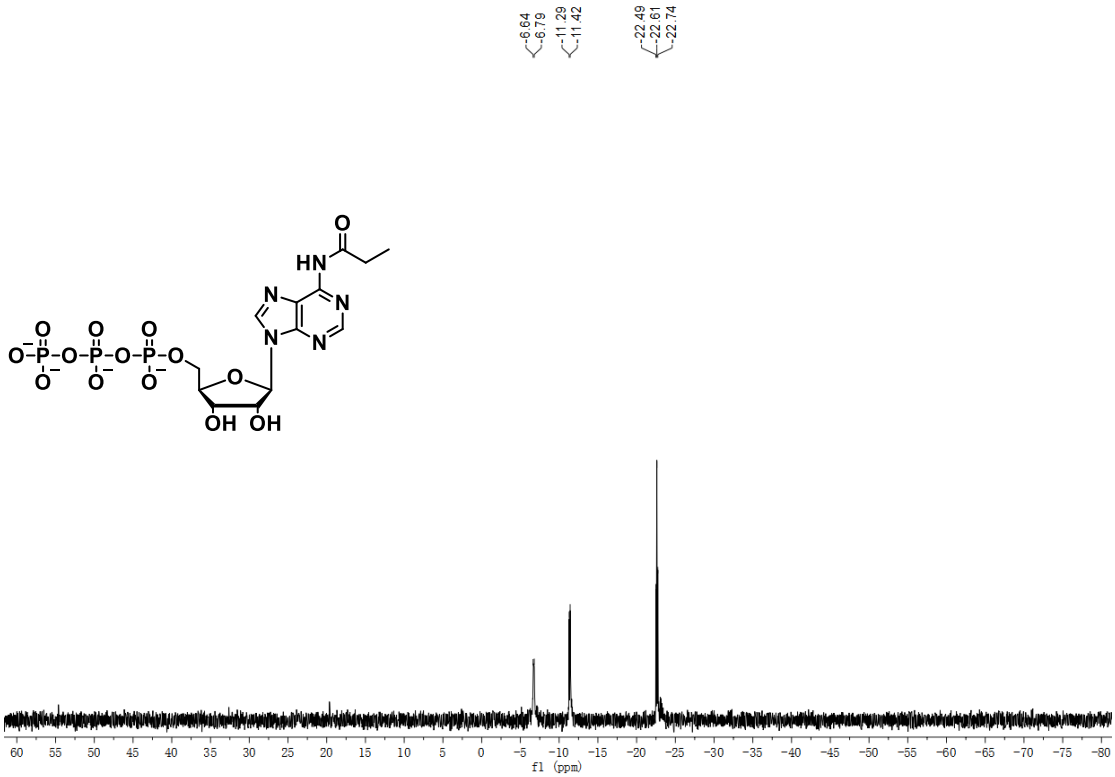
ac6A ^{31}P-NMR



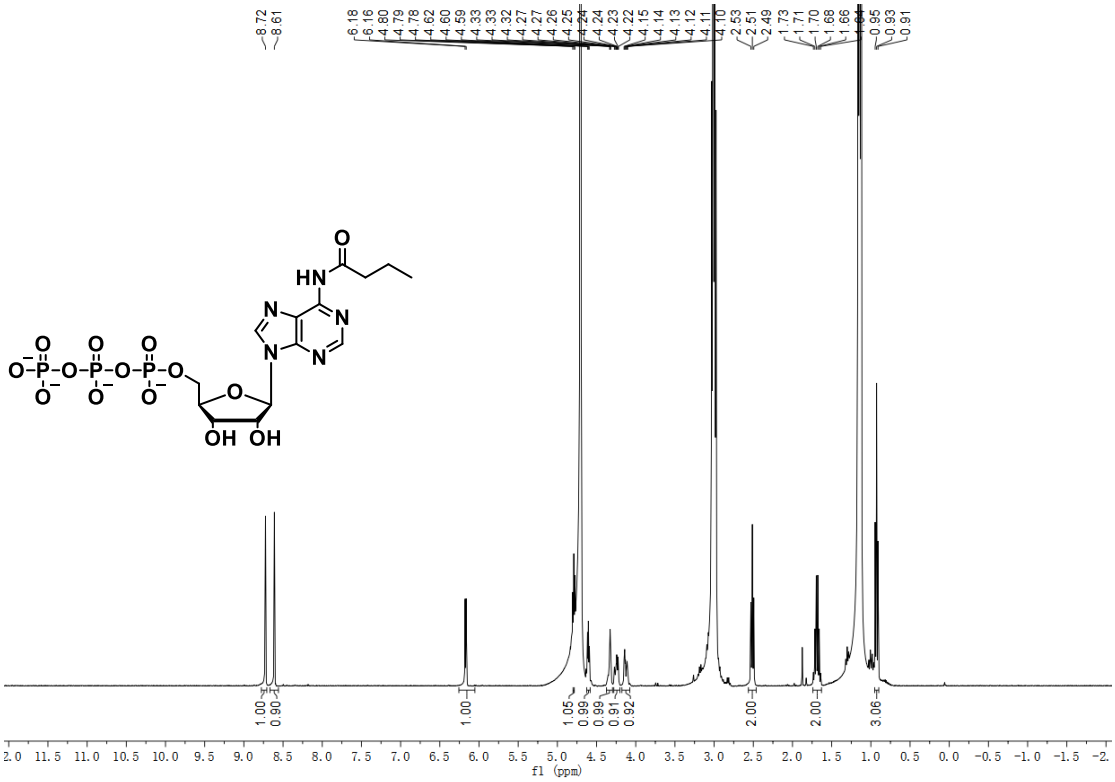
pr6A ^{1}H-NMR



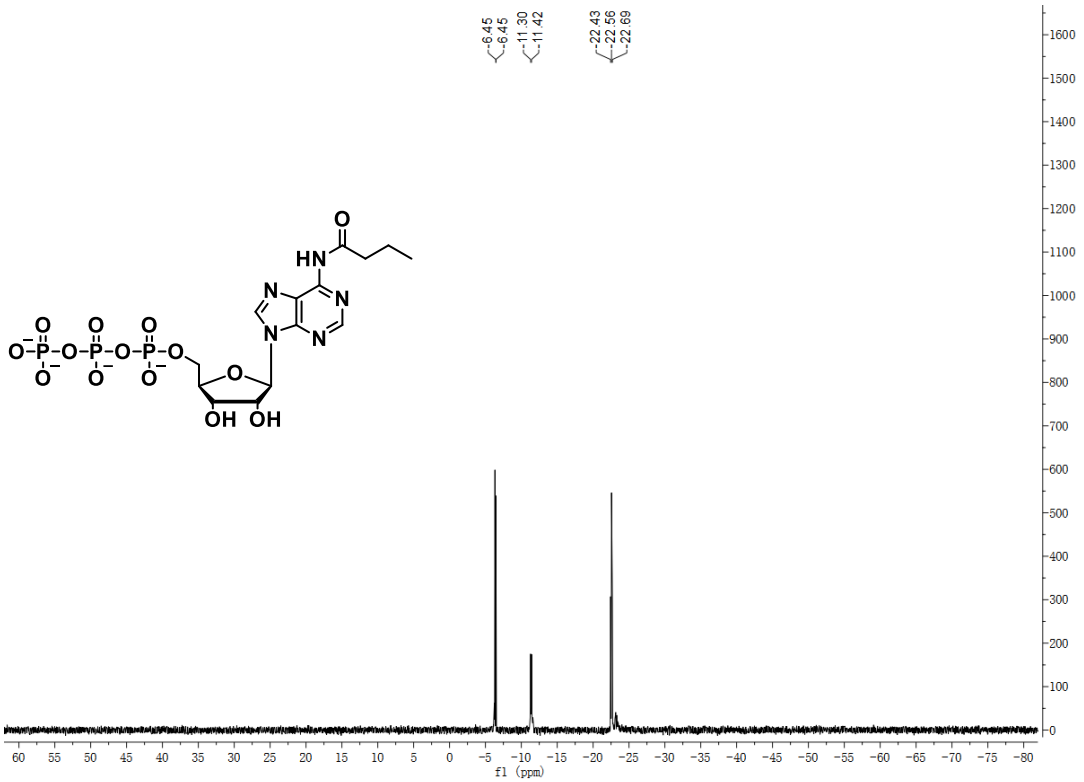
pr6A ^{31}P-NMR



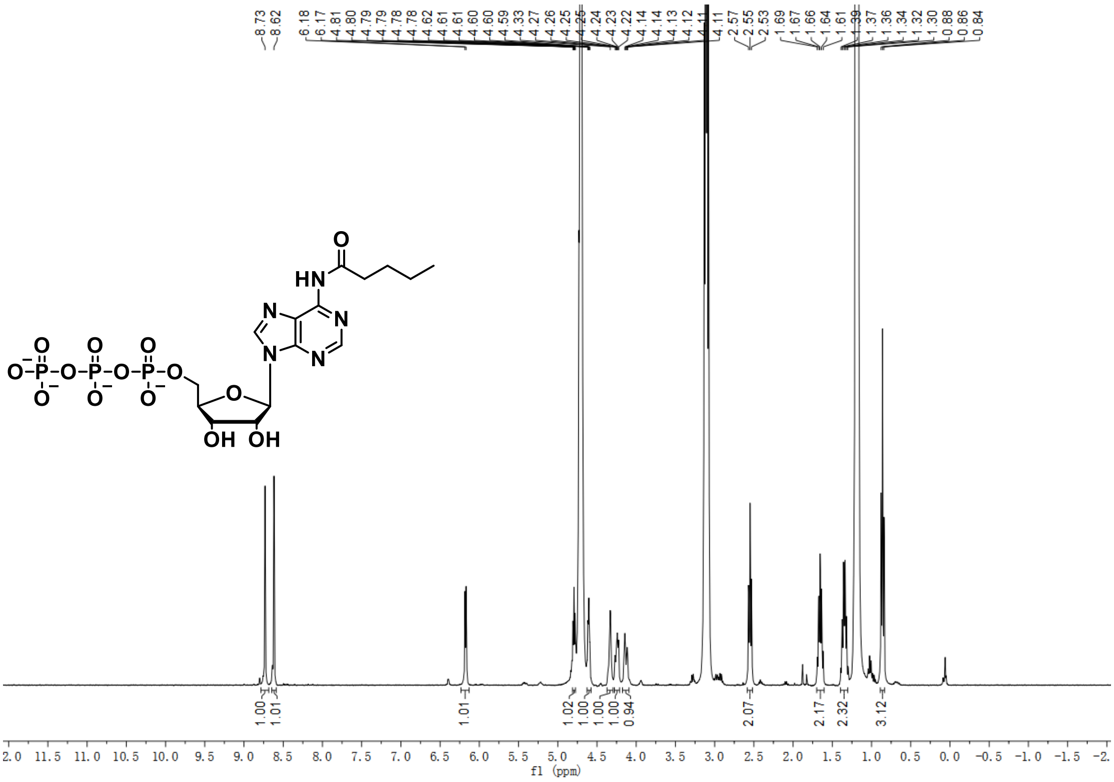
bu6A ^{1}H-NMR



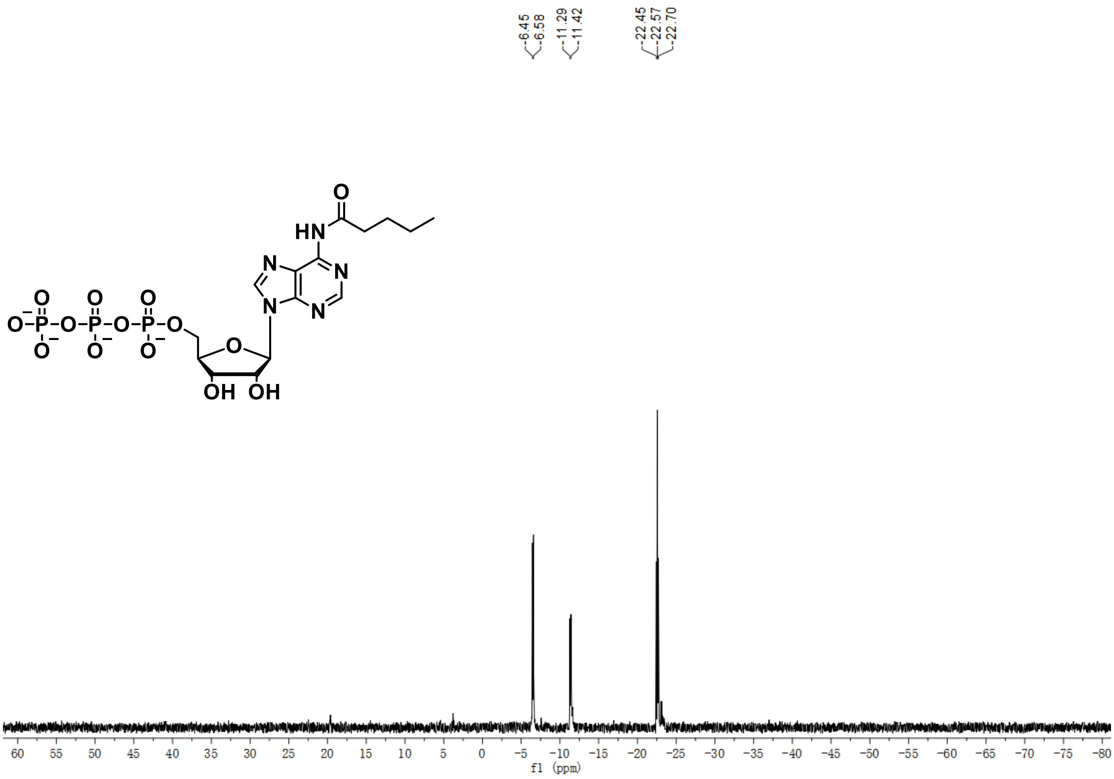
bu6A ^{31}P-NMR



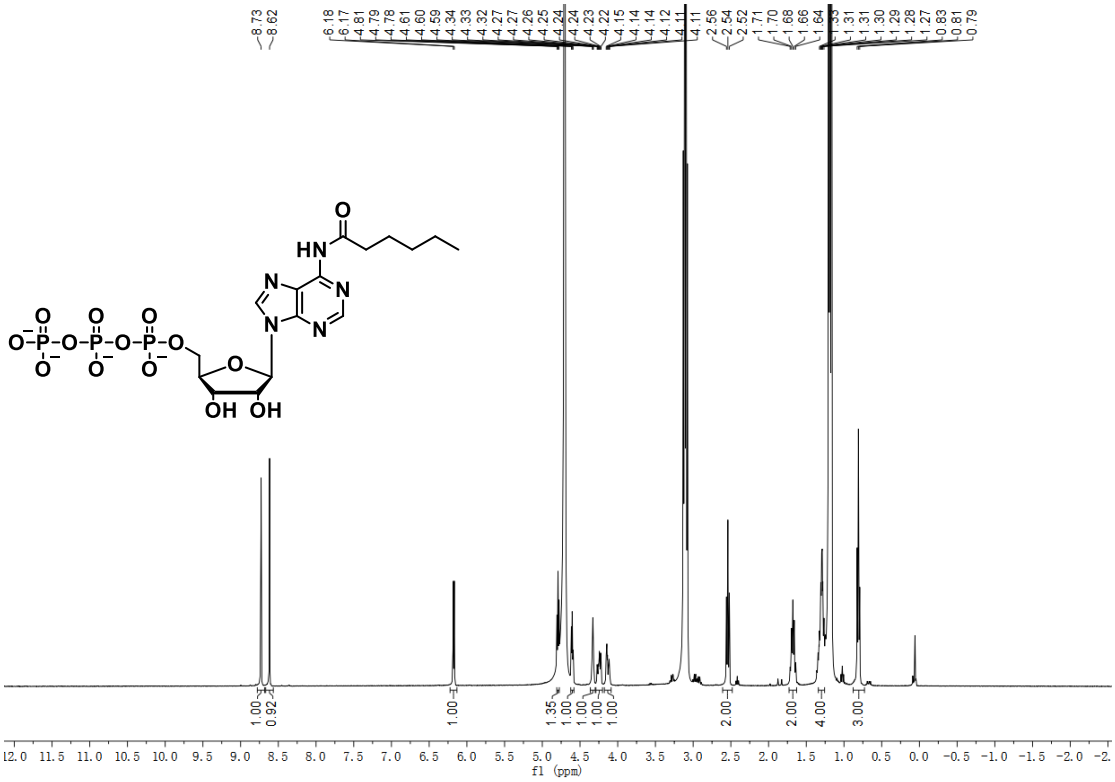
va6A ^{1}H-NMR



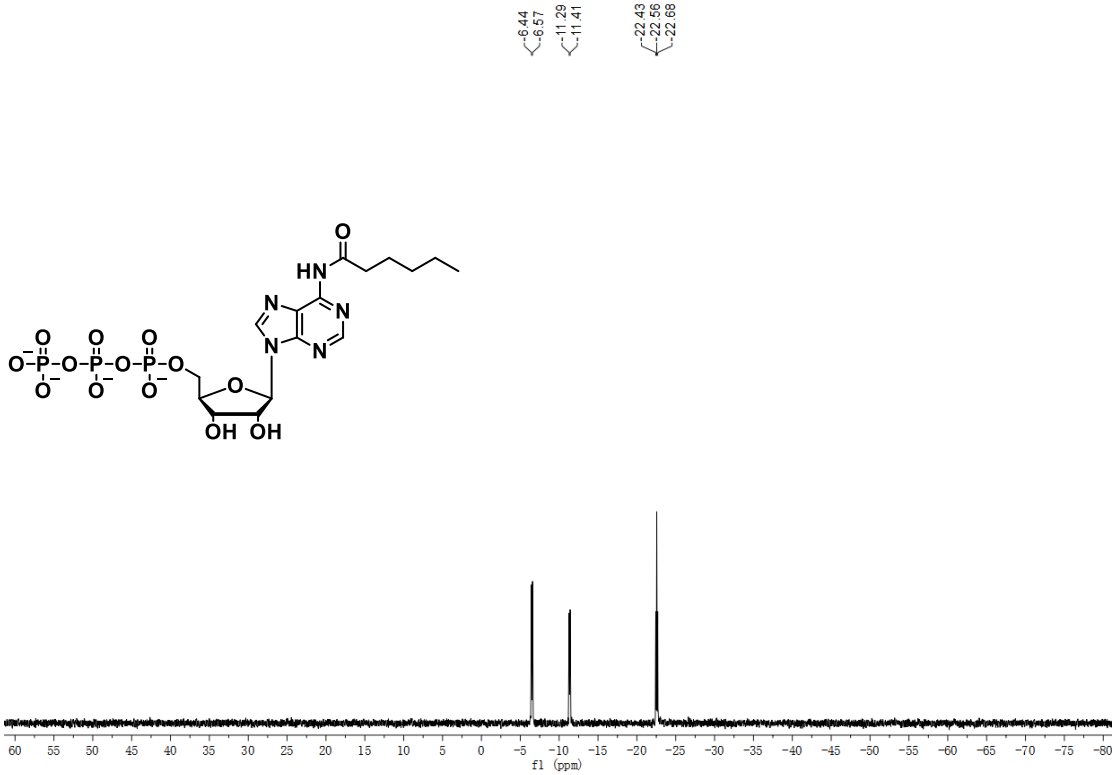
va6A ^{31}P-NMR



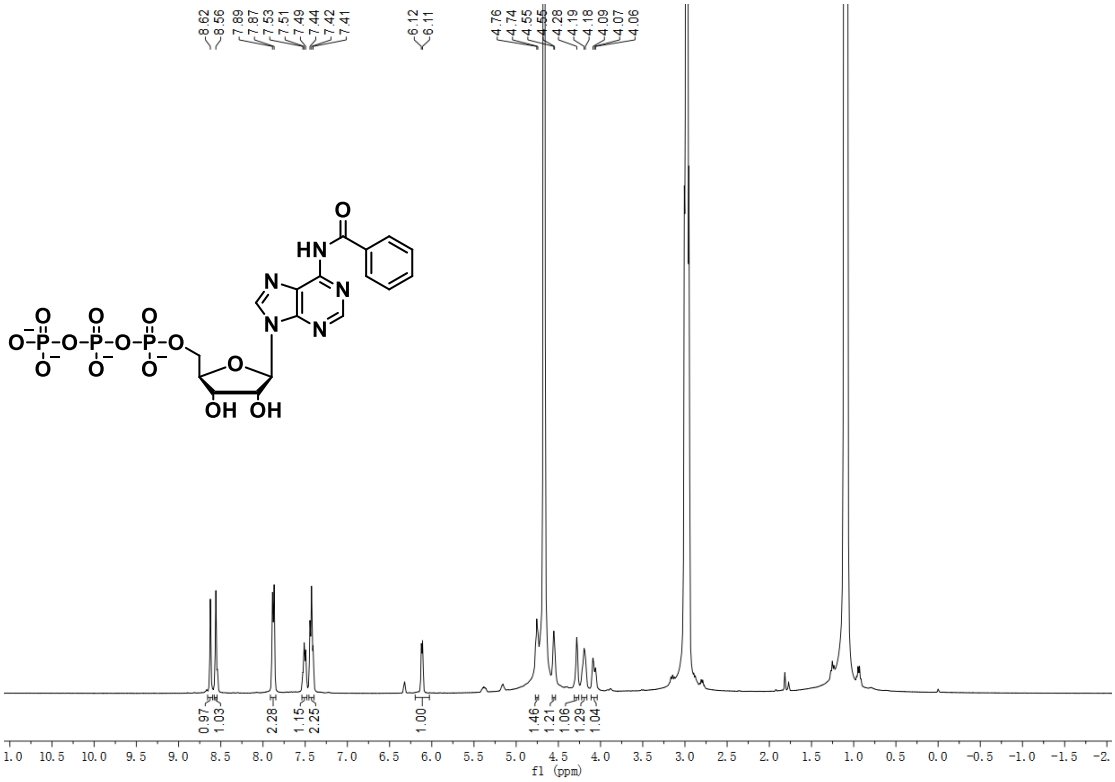
ca6A ^{1}H-NMR



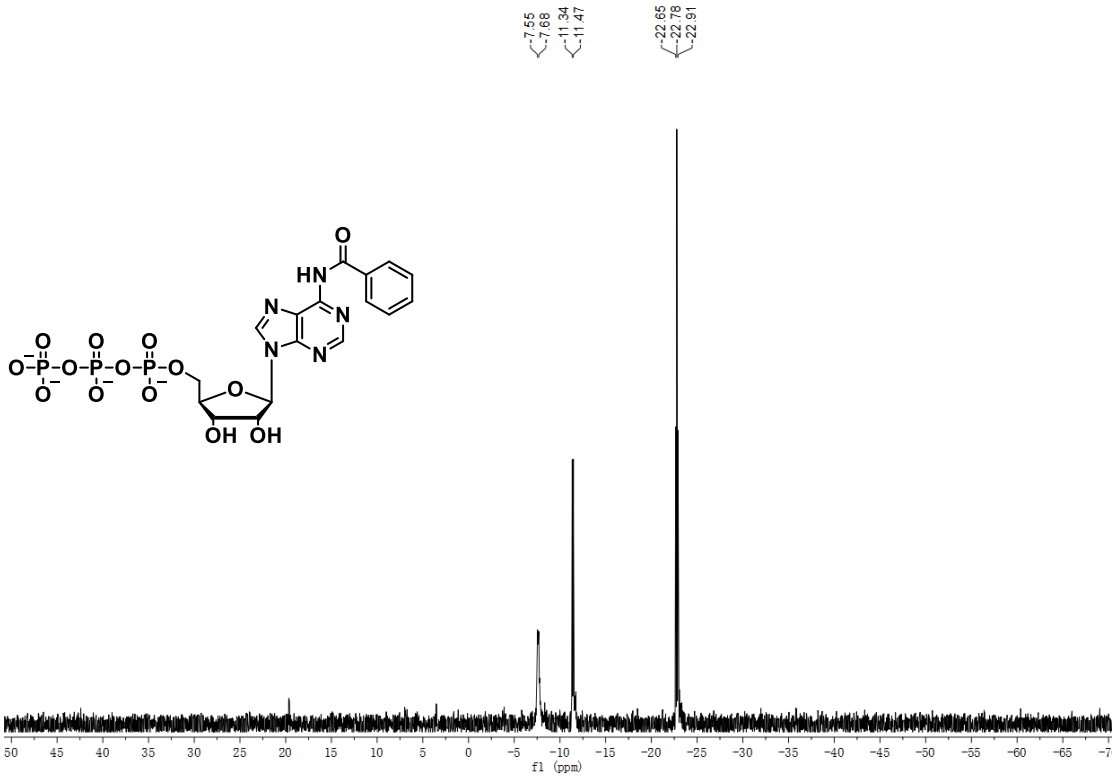
ca6A ^{31}P-NMR



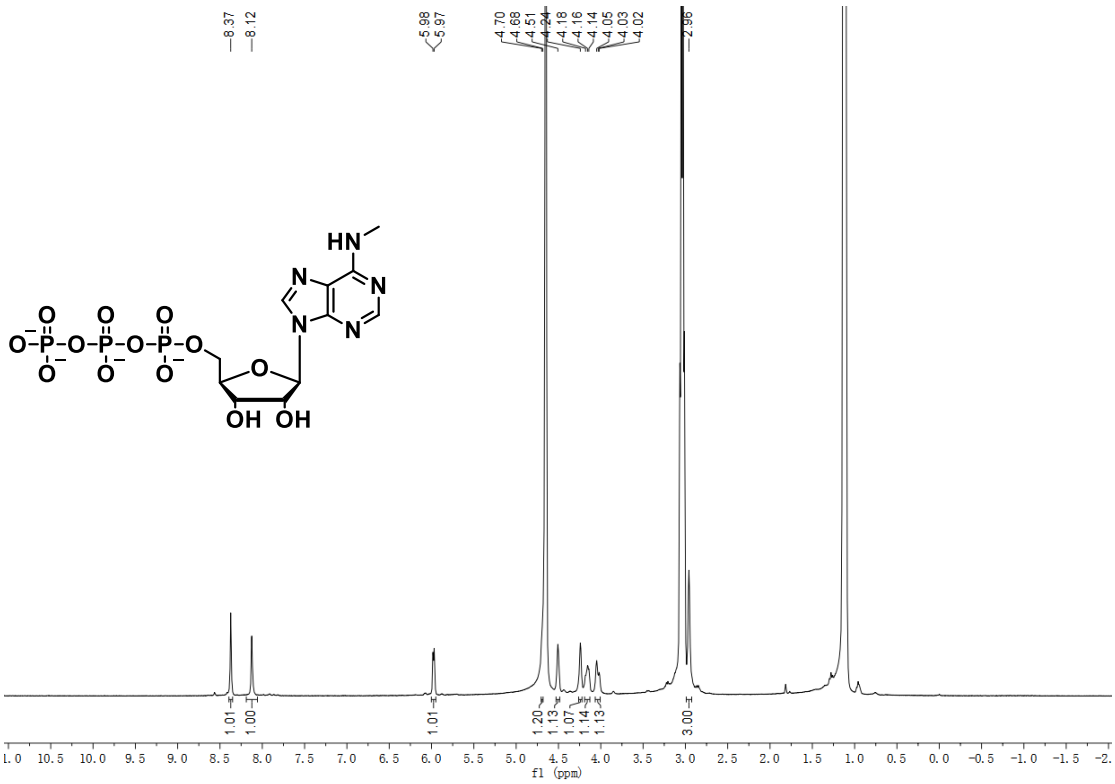
bz6A ^{1}H-NMR



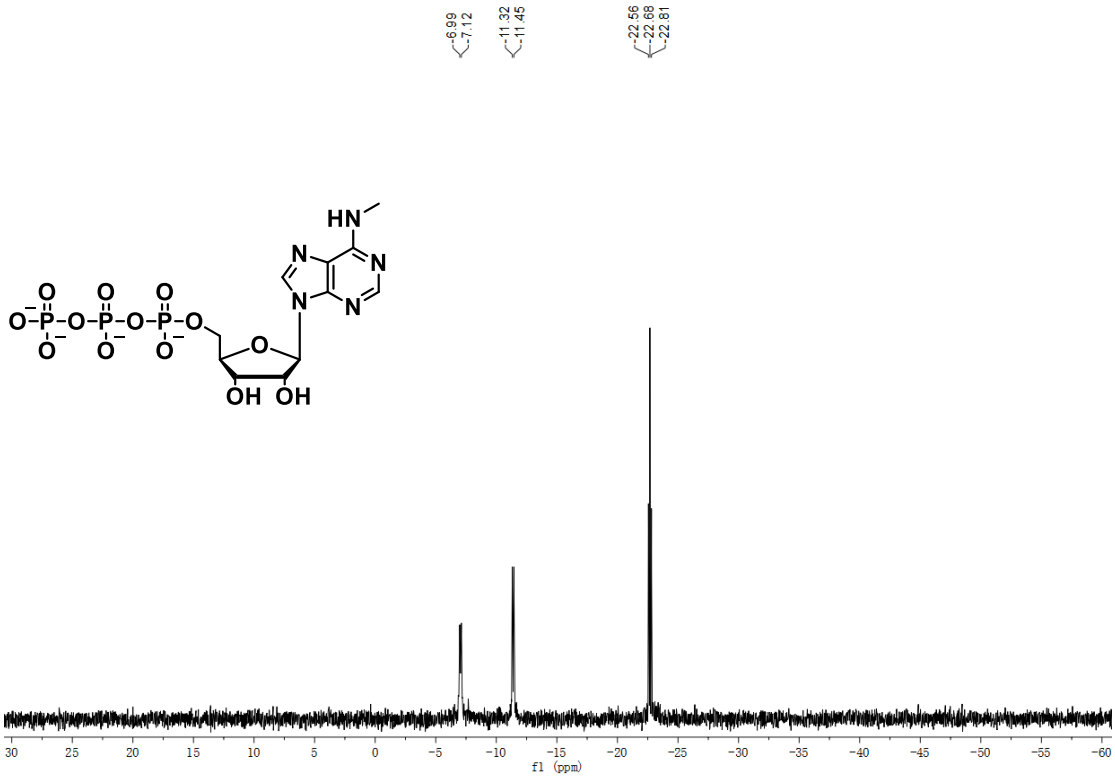
bz6A ^{31}P-NMR



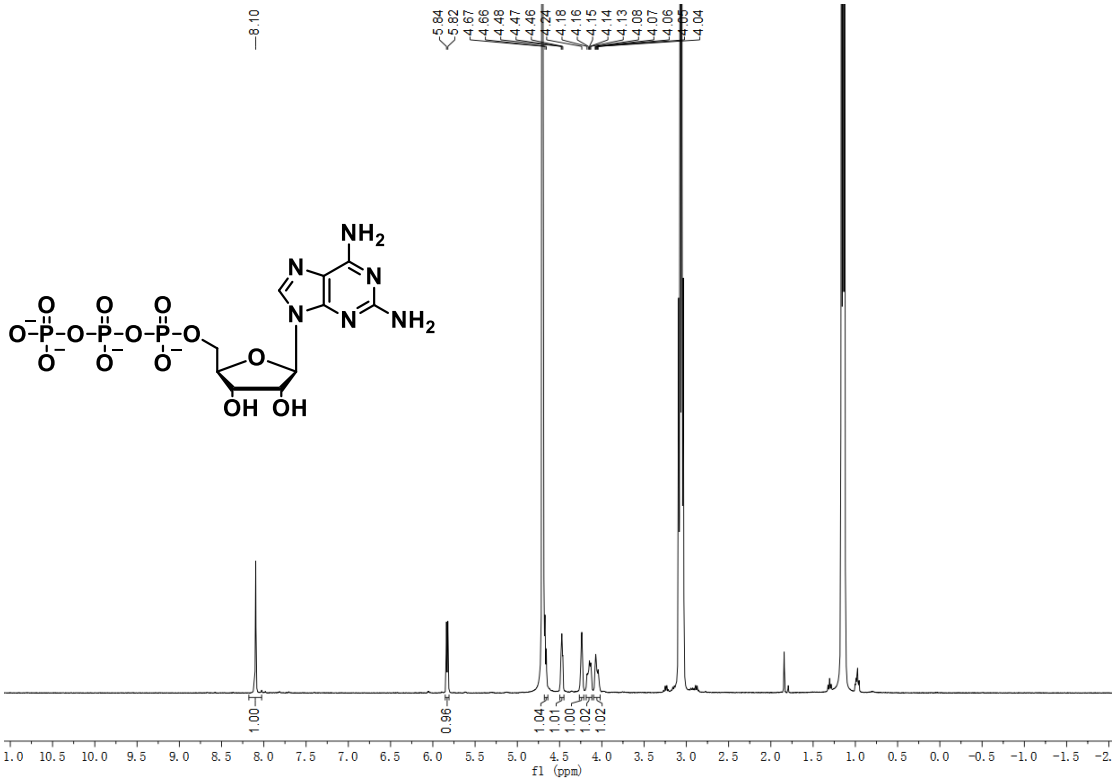
me6A ^{1}H-NMR



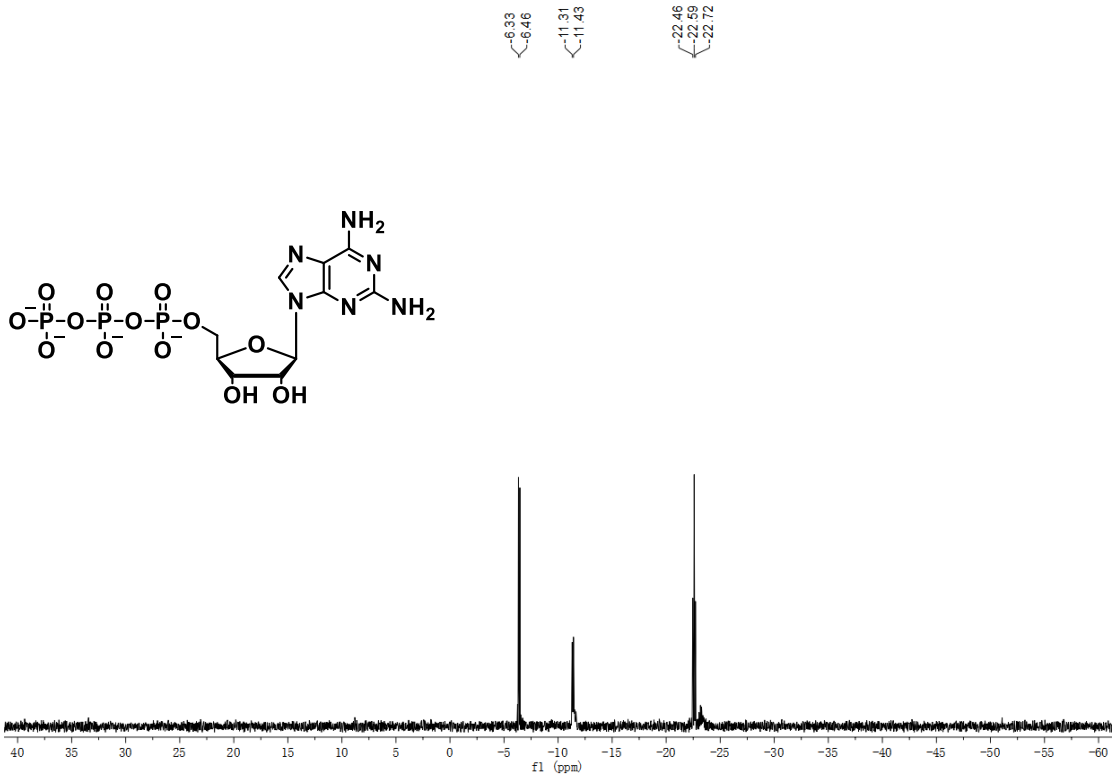
me6A ^{31}P-NMR



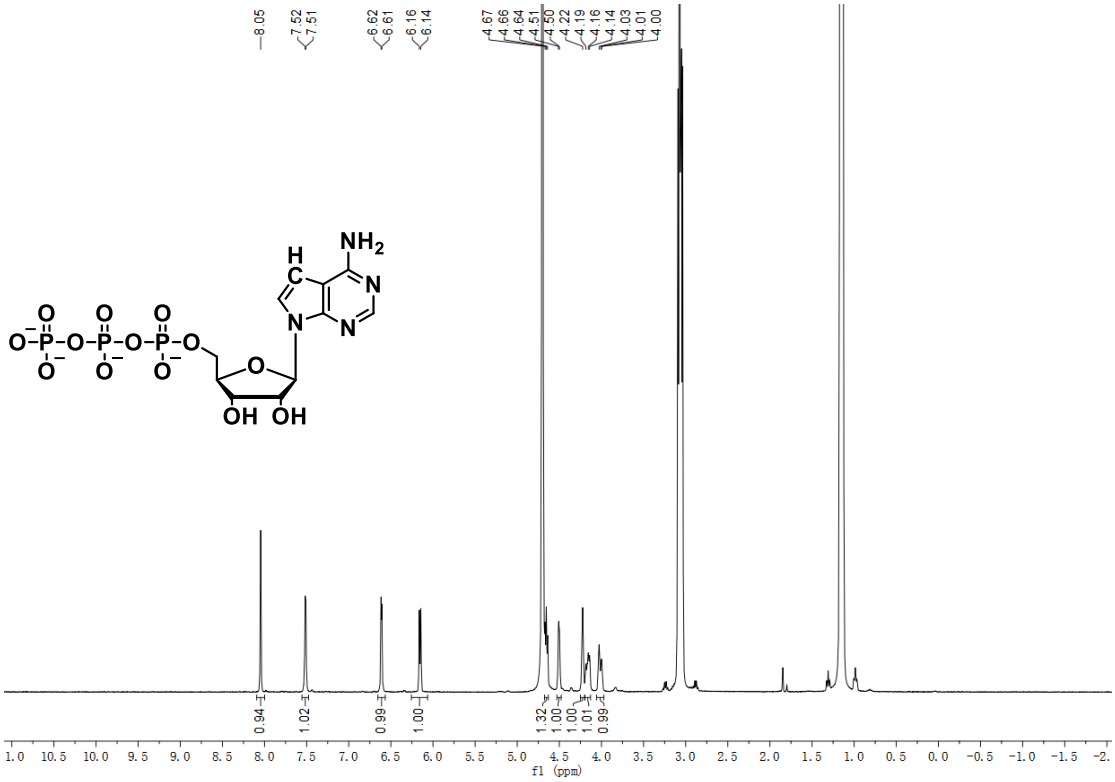
am2A ^{1}H-NMR



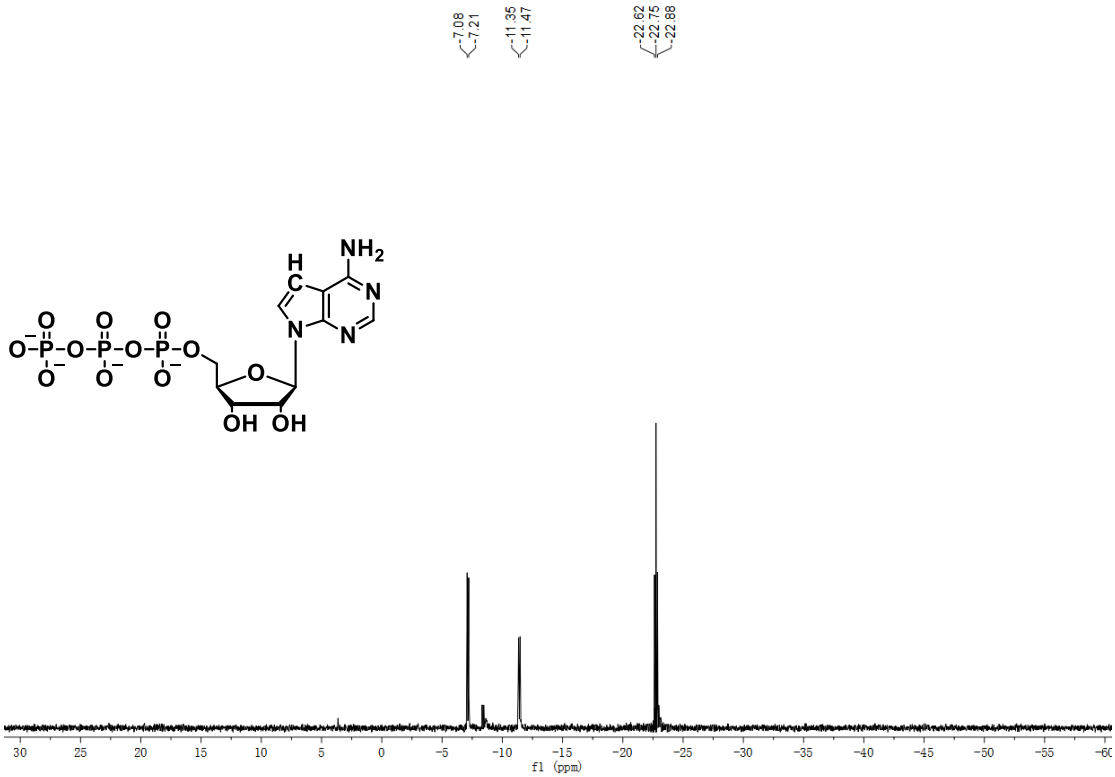
am2A ^{31}P-NMR



c7A ^{1}H-NMR

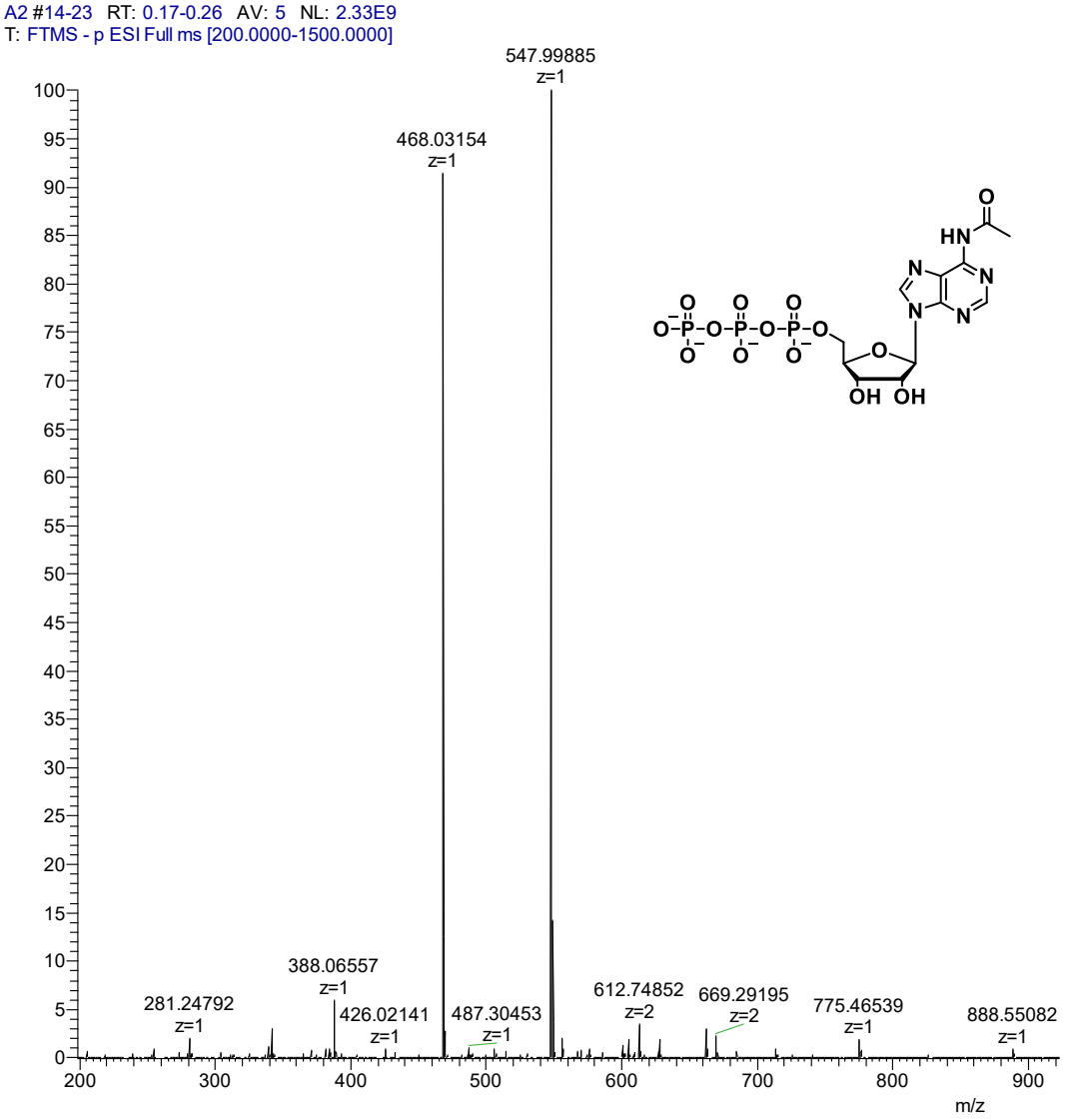


c7A ^{31}P-NMR

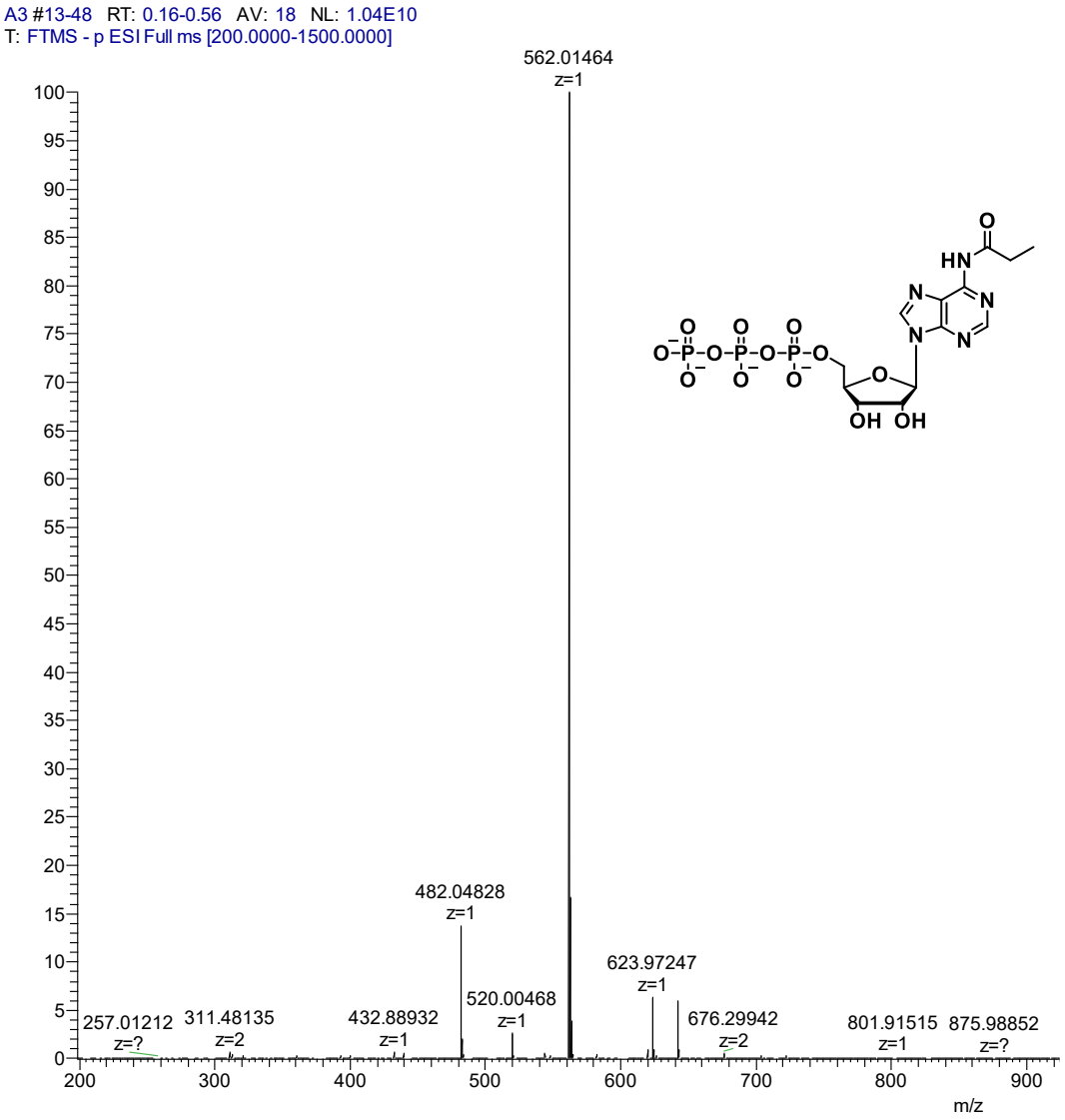


HRMS spectra

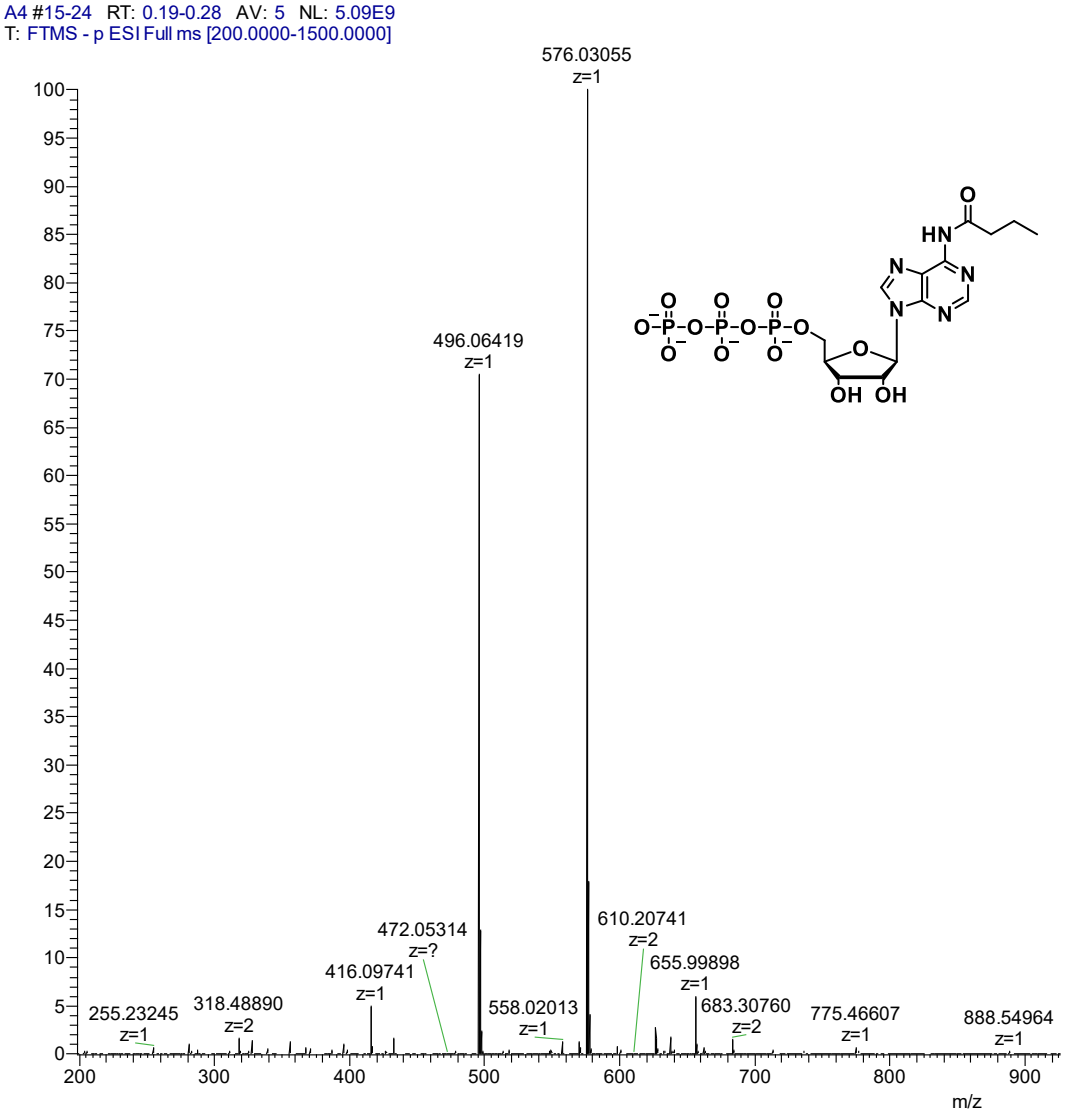
ac6A calcd for $C_{12}H_{17}N_5O_{14}P_3^-$ [M - H]$^-$ 547.9990, found:547.9989



pr6A calcd for $C_{13}H_{19}N_5O_{14}P_3^-$ [M - H]$^-$ 562.0141, found:562.0146

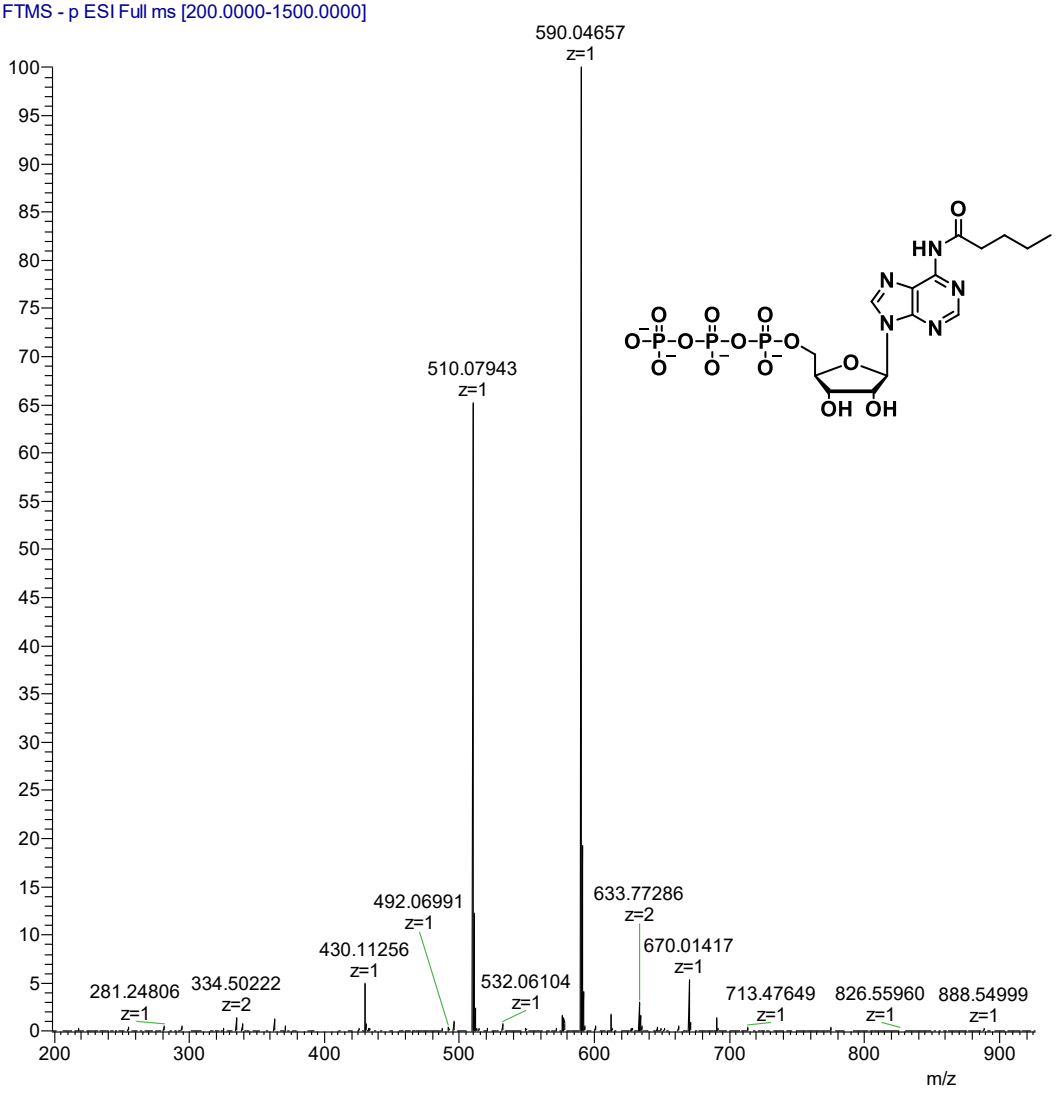


bu6A calcd for $C_{14}H_{21}N_5O_{14}P_3^-$ [M - H]$^-$ 576.0298, found:576.0306

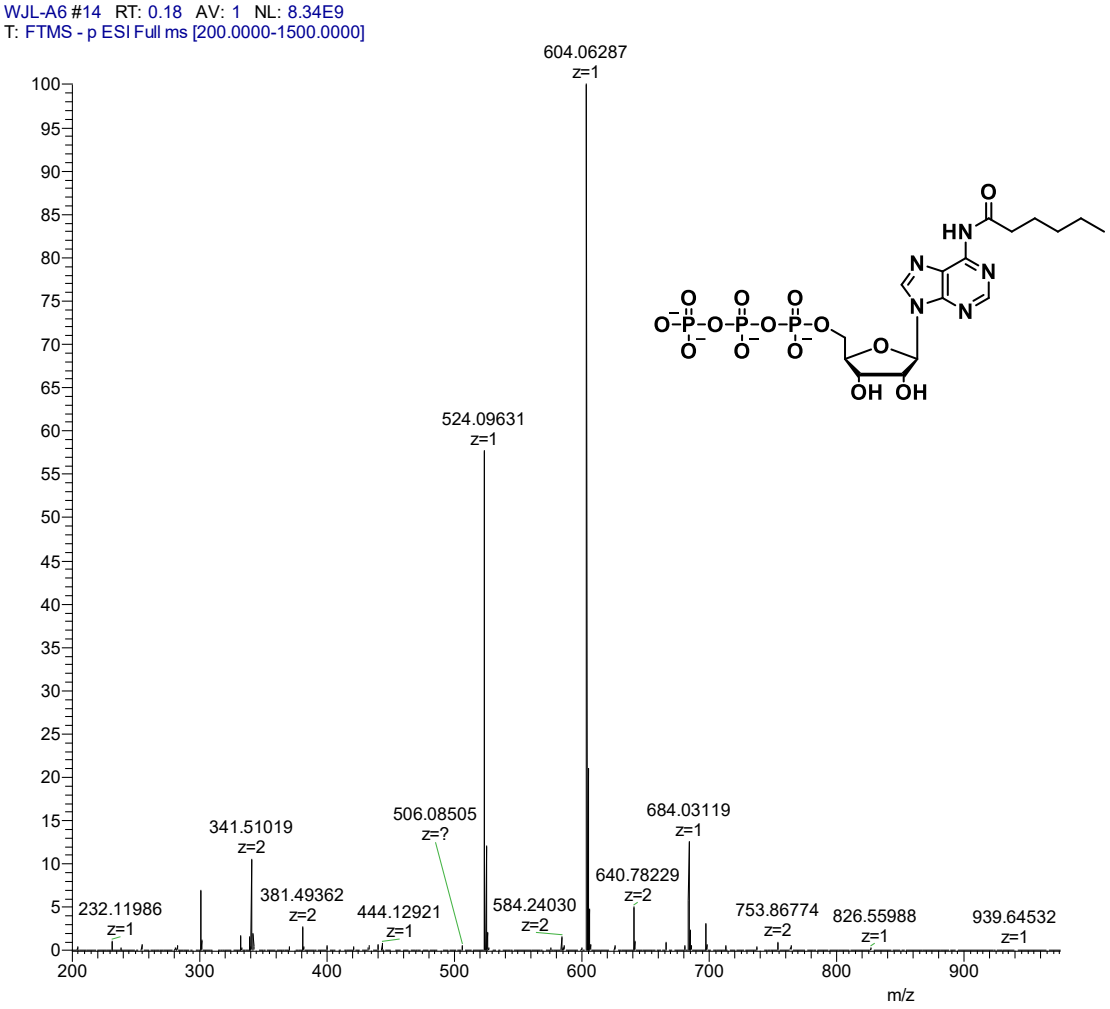


va6A calcd for $C_{15}H_{23}N_5O_{14}P_3^-$ [M - H]$^-$ 590.0454, found:590.0466

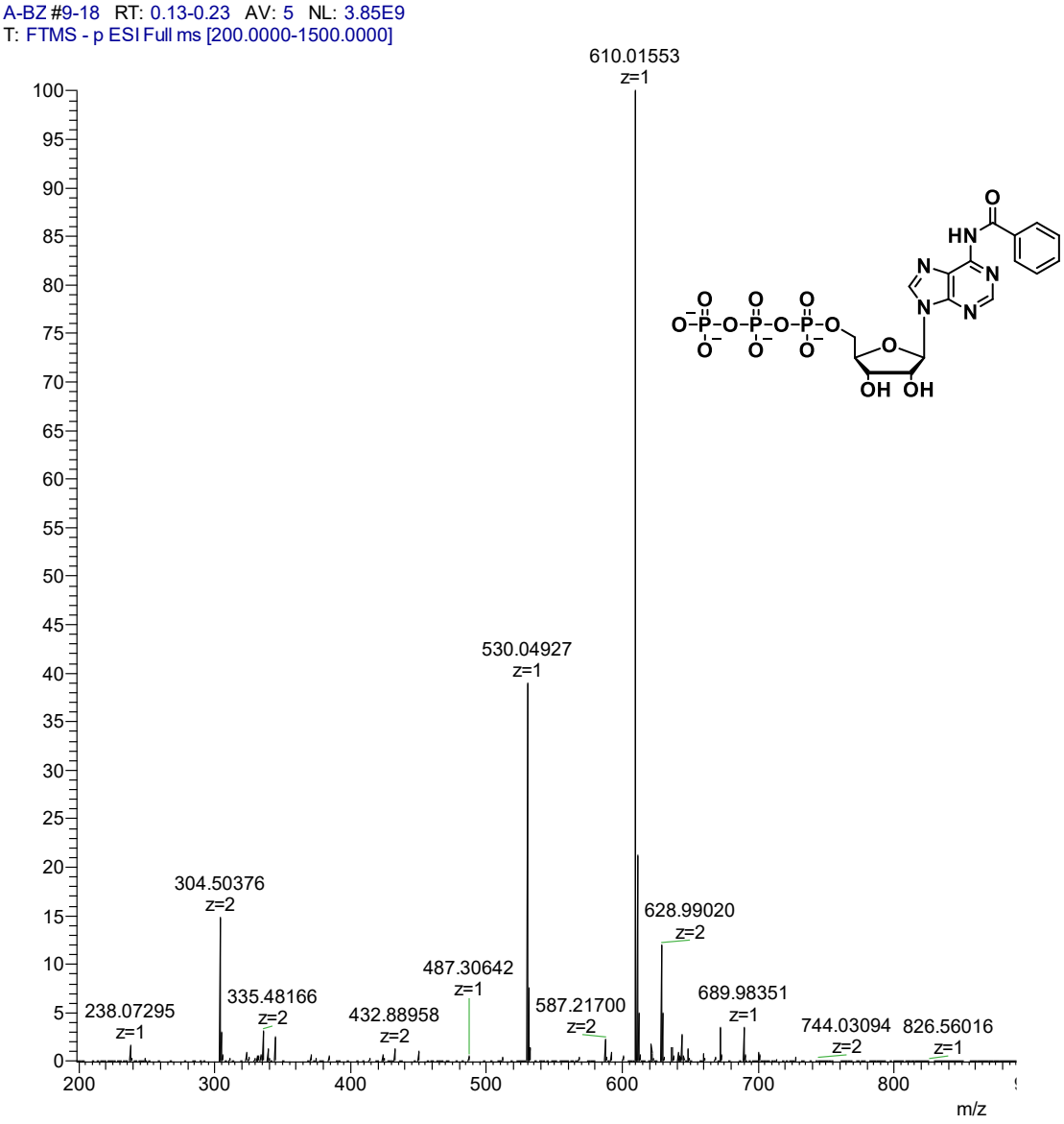
A5 #11-27 RT: 0.14-0.31 AV: 8 NL: 5.44E9
T: FTMS - p ESI Full ms [200.0000-1500.0000]



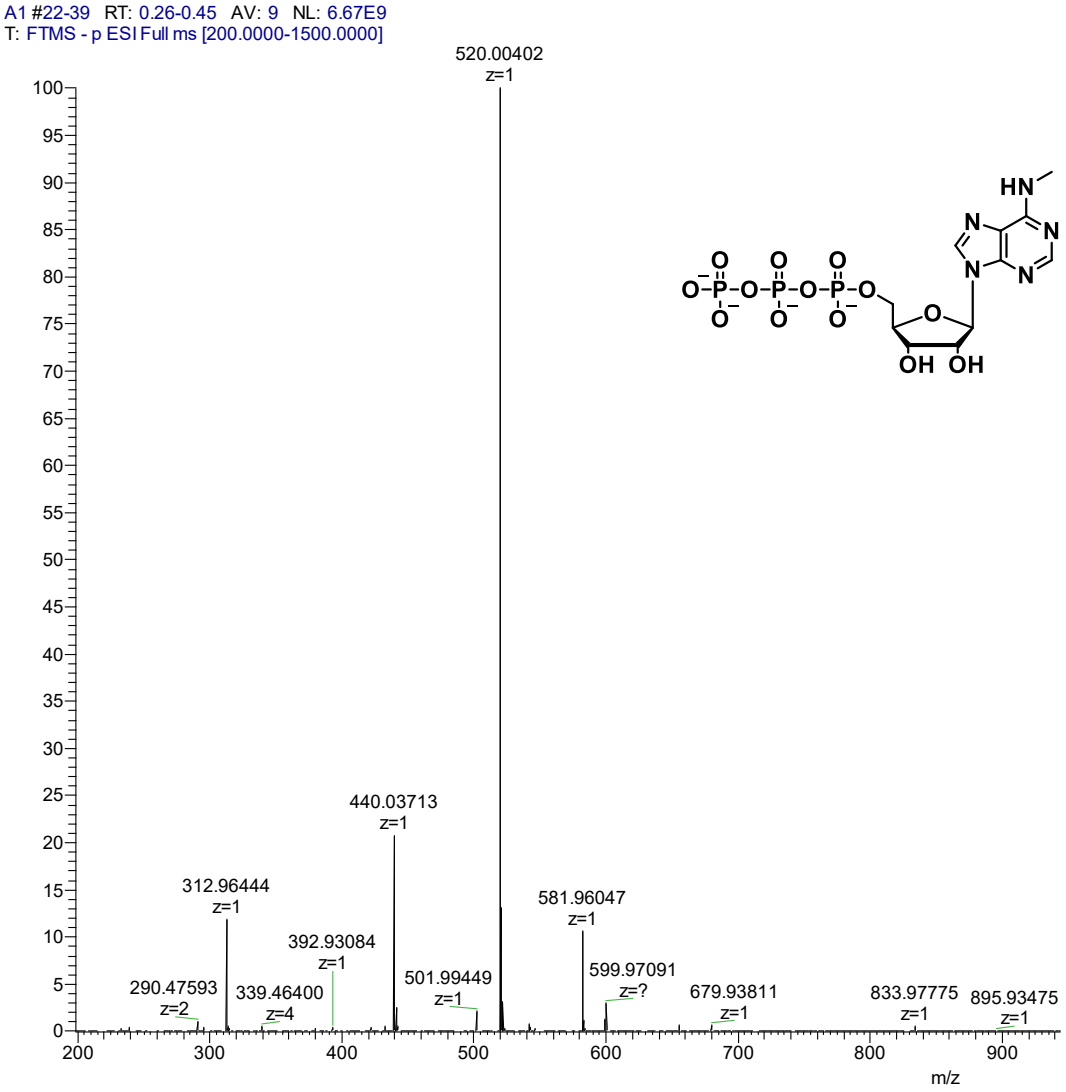
ca6A calcd for $C_{16}H_{25}N_5O_{14}P_3^-$ [M - H]$^-$ 604.0611, found:604.0629



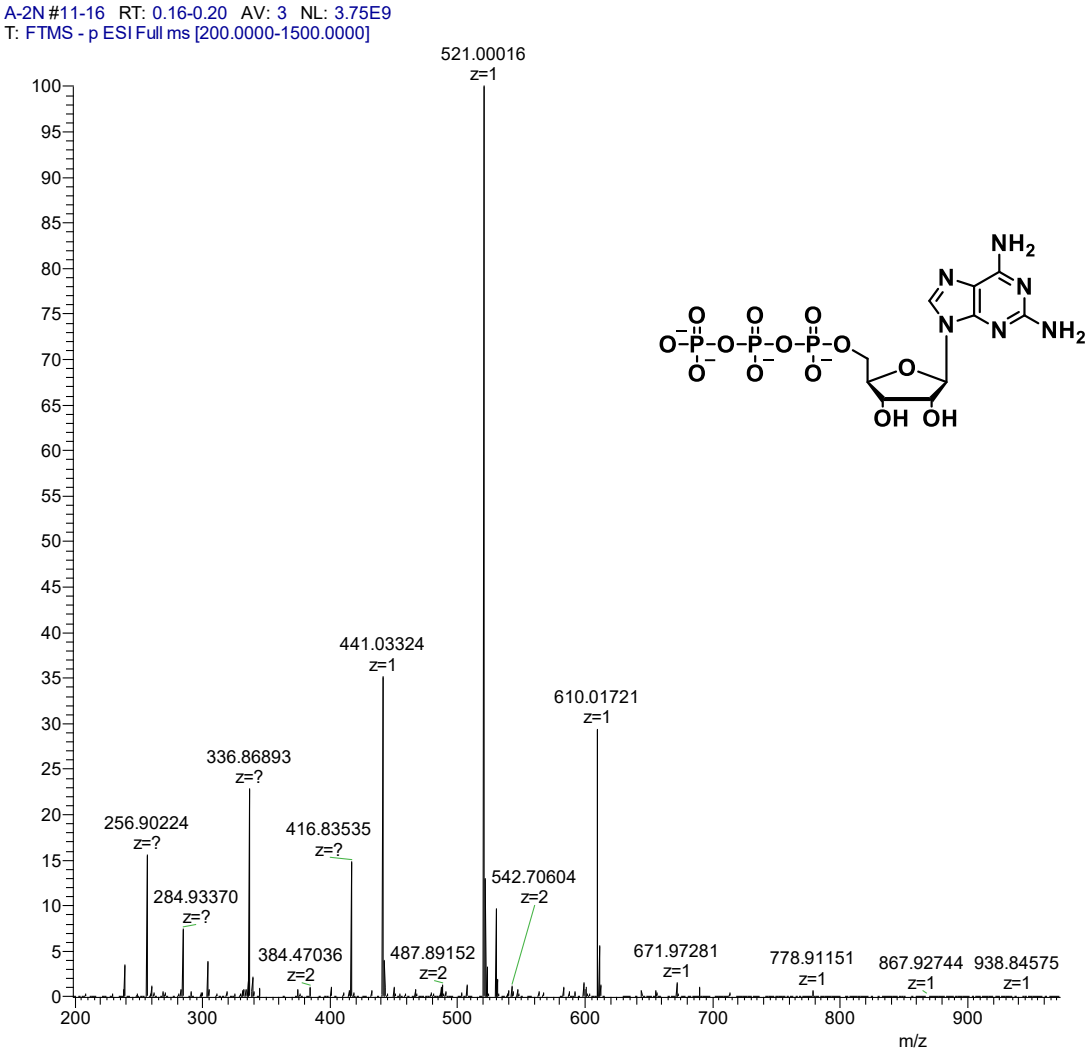
bz6A calcd for $C_{17}H_{19}N_5O_{14}P_3^-$ [M - H]$^-$ 610.0141, found:610.0156



me6A calcd for $C_{11}H_{17}N_5O_{13}P_3^-$ [M - H]$^-$ 520.0036, found:520.0040



am2A calcd for $C_{10}H_{16}N_6O_{13}P_3^-$ [M - H]$^-$ 520.9988, found:521.0002



c7A calcd for $C_{11}H_{16}N_4O_{13}P_3^-$ [M - H]$^-$ 504.9927, found:504.9942