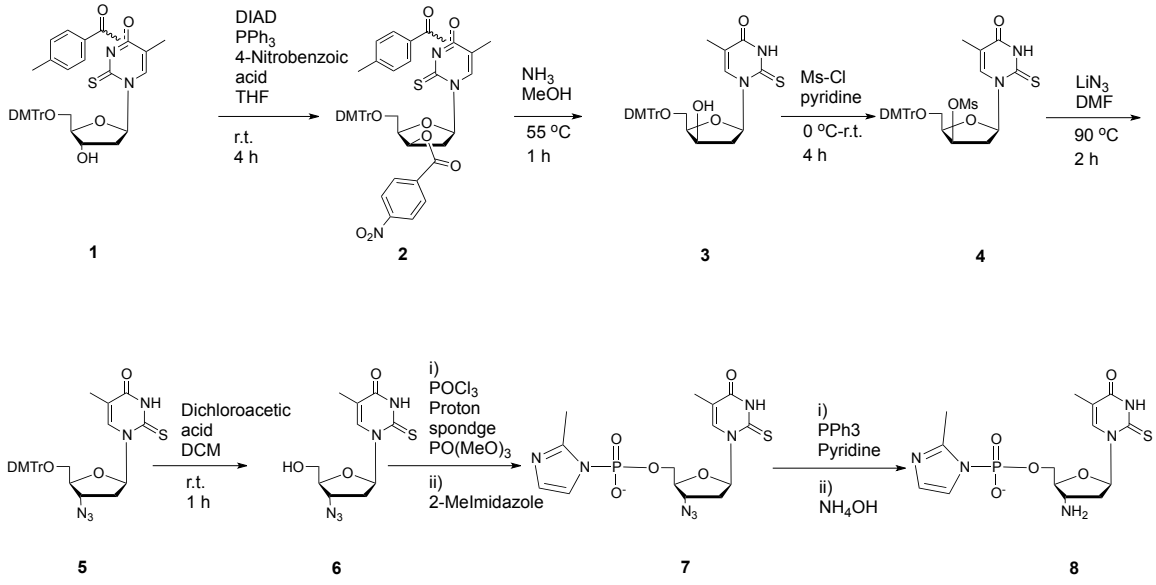
Supporting Information Appendix

For manuscript titled "Fast and accurate non-enzymatic copying of an RNA-like synthetic genetic polymer"

SI Text

General information for reagents and instrumentation. All solvents and reagents were reagent grade, purchased commercially, and used without further purification unless specified. All chemicals were purchased from Sigma-Aldrich unless otherwise indicated. Oligonucleotides used as primers or templates were synthesized on an Expedite nucleic acid synthesizer (Applied BioSystems) or purchased from IDT (Coralville, IA) unless otherwise indicated. All the Nuclear Magnetic Resonance (NMR) spectra were recorded on a Varian NMR spectrometer (Oxford AS-400). Chemical shifts are reported as parts per million (ppm) using tetramethylsilane (TMS) as internal standard or by reference to proton resonances resulting from incomplete deuteration of the NMR solvent. Data were reported as follows: (s = singlet, d = doublet, t = triplet, q = quartet, m = multiplet, dd = doublet of doublets, J = coupling constant in Hz, integration). Proton-decoupled ^{13}C NMR (100 MHz) spectra were reported in ppm from $CDCl_3$, CD_3OD, or DMSO-d6 (77.0, 49.0, or 39.5 ppm, respectively). Proton-decoupled ^{31}P NMR (161.8 MHz) spectra were reported in ppm using phosphate buffer as reference. Electrospray mass spectra were recorded on a Bruker Daltonics Esquire 6000 ESI-MS. LC-MS studies of oligonucleotides were carried out on Agilent 6520 Q-TOF LC/MS system. 5'-*O*-(dimethoxytrityl)-N^3/O^4-(toluoyl)-2-thiothymidine was purchased from Berry & Associates (Dexter, MI). The activated phosphor-2-methylimidazole nucleotide monomers were purified by reverse-phase preparative HPLC (Varian ProStar Preparative LC) on a Prep-C18 column (Varian Dynamax 250 × 21.4 mm) equilibrated with 25 mM triethylammonium bicarbonate, pH 8.0 and eluted with a linear acetonitrile gradient (0-60%).

Synthesis of 3′-amino-3′-deoxy-2-thiothymidine-5′-phosphor-2-methylimidazolide (3′-NH_2-2-MeImpddsT).



*1-[5'-O-(Dimethoxytrityl)-N^3/O^4-(toluoyl)-3'-O-(4-nitrobenzoyloxy)-β-$_D$-threo-pentofuranosyl]-2-thiothymine (Compound **2**)*

The preparation of compound **2** was adapted from a previously reported procedure (1-3), with minor modifications as follows.

To a solution of 5'-*O*-(dimethoxytrityl)-N^3/O^4-(toluoyl)-2-thiothymidine **1** (From Berry & Associates, Inc.) (100 mg; 0.147 mmol) in anhydrous THF (2.0 ml) were added triphenylphosphine (58 mg; 0.221 mmol) and diisopropyl azodicarboxylate (DIAD) (45 mg, 44 µl, 0.221 mmol) at room temperature. After 20 min, 4-nitrobenzoic acid (37 mg, 0.221 mmol) was added to the reaction mixture and the reaction mixture was stirred further for 4 h. The solvent was removed under vacuum and the residue was purified by flash column chromatography over silica gel using methanol-dichloromethane (1%-10%) as the eluent to afford **2** (103 mg, 85%) as a yellow foam. ^{1}H NMR δ (400 MHz, $CDCl_3$): 8.20 (d, *J* = 8.4 Hz, 1H), 7.80 (d, *J* = 8.0 Hz, 2H), 7.69 (s, 1H), 7.68-7.63 (m, 5H), 7.56-7.52 (m, 4H), 7.47-7.43 (m, 6H), 7.38 (dd, *J* = 7.6 Hz, 5.6 Hz, 1H), 6.84 (m, 1H), 6.74 (m, 2H), 5.79-5.76 (m, 1H), 4.60-4.53 (m, 1H), 3.75 (s, 6H), 3.68-3.62 (m, 1H), 3.57-3.53 (m, 1H), 2.99-2.93 (m, 1H), 2.46-2.44 (m, 1H), 2.40 (s, 3H), 2.03 (s, 3H); ESI-MS calcd for $C_{46}H_{41}N_3NaO_{10}S^+$ $[M+Na]^+$: 850.2, found: 850.0.

*1-[5'-O-(Dimethoxytrityl)-3'-OH-β-$_D$-threo-pentofuranosyl]-2-thiothymine (Compound **3**)*.

A suspension of **2** (146 mg, 0.177 mmol) in methanolic ammonia (10 ml) was stirred for 1 h at 55 ^{o}C. The homogeneous solution was concentrated under vacuum and the residue was purified by flash column chromatography over silica gel using methanol-chloroform (5%-10%) as the eluent to afford **3** (80 mg; 81% yield) as a white foam. ^{1}H NMR δ (400 MHz, $CDCl_3$): 7.73 (s, 1H), 7.67 (d, *J* = 6.8 Hz, 2H), 7.44 (d, *J* = 8.0 Hz, 2H), 7.33 (d, *J* = 8.0 Hz, 2H), 7.26 (dd, *J* = 6.4 Hz, 6.8 Hz, 1H), 7.18 (d, *J* = 7.6 Hz, 2H), 6.81 (d, *J* = 8.8 Hz, 4H), 6.73 (dd, *J* = 1.4 Hz, 6.8 Hz, 1H), 4.37-4.35 (m, 1H), 4.12-4.09 (m, 1H), 3.74 (s, 6H), 3.65-3.61 (m, 1H), 3.49-3.45 (m, 1H), 2.56-2.54 (m, 1H), 2.35 (s, 3H), 2.20-2.17 (m, 1H); ESI-HRMS calcd for $C_{31}H_{33}N_2O_6S^+$ $[M+H]^+$: 561.2054, found: 561.2060.

*1-[5'-O-(Dimethoxytrityl)-3'-O-mesyl-β-$_D$-threo-pentofuranosyl]-2-thiothymine (Compound **4**)*.

To a solution of **3** (80 mg; 0.143 mmol) in anhydrous pyridine (1.0 ml) were added MsCl (20 mg; 13 µl; 0.171 mmol) dropwise at room temperature. The reaction mixture was quenched with MeOH after stirring for 4 h and concentrated. The residue was dissolved in CH_2Cl_2 (100 ml) and washed with water (2×50 ml). The organic layer was dried (Na_2SO_4), concentrated under vacuum, and the residue was purified by flash column chromatography over silica gel using ethyl acetate-hexane (10%-50%) as the eluent to afford **4** (81mg; 89% yield) as a white foam. ^{1}H NMR δ (400 MHz, $CDCl_3$): 7.45 (s, 1H), 7.43 (d, *J* = 6.8 Hz, 2H), 7.35-7.29 (m, 5H), 7.25 (d, *J* = 6.4 Hz, 2H), 6.86 (d, *J* = 8.8 Hz, 4H), 6.75 (dd, *J* = 2.0 Hz, 5.6 Hz, 1H), 5.25-5.23 (m, 1H), 4.32-4.28 (m, 1H), 3.79 (s, 6H), 3.68-3.65 (m, 1H), 3.44-3.39 (m, 1H), 2.91-2.85 (m, 1H), 2.77 (s, 3H), 2.61-2.56 (m, 1H), 1.83 (s, 3H); ESI-HRMS calcd for $C_{32}H_{35}N_2O_8S_2^+$ $[M+H]^+$: 639.1829, found: 639.1832.

*3'-Azido-5'-O-(dimethoxytrityl)-3'-deoxy-2-thiothymidine (Compound **5**).*
To a solution of **4** (80 mg, 0.125 mmol) in anhydrous DMF (1.0 ml) was added lithium azide (31 mg, 0.626 mmol). The reaction mixture was heated at 90 °C until all starting materials were consumed. The solvent was then evaporated under vacuum and the residue was purified by flash column chromatography over silica gel using ethyl acetate-hexane (10%-60%) as the eluent to afford **5** (62 mg; 85% yield) as a white foam. ^{1}H NMR δ (400 MHz, $CDCl_3$): 7.85 (s, 1H), 7.64 (m, 1H), 7.51-7.42 (m, 1H), 7.38-7.31 (m, 2H), 7.31-7.19 (m, 5H), 6.83-6.80 (m, 4H), 6.79 (m, 1H), 4.33-4.29 (m, 1H), 3.99-3.94 (m, 1H), 3.76 (s, 6H), 3.61-3.57 (m, 1H), 3.33-3.29 (m, 1H), 2.68-2.60 (m, 1H), 2.46-2.39 (m, 1H), 1.46 (s, 3H); ESI-HRMS calcd for $C_{31}H_{31}N_5NaO_5S^+$ $[M+Na]^+$: 608.1938, found: 608.1949.

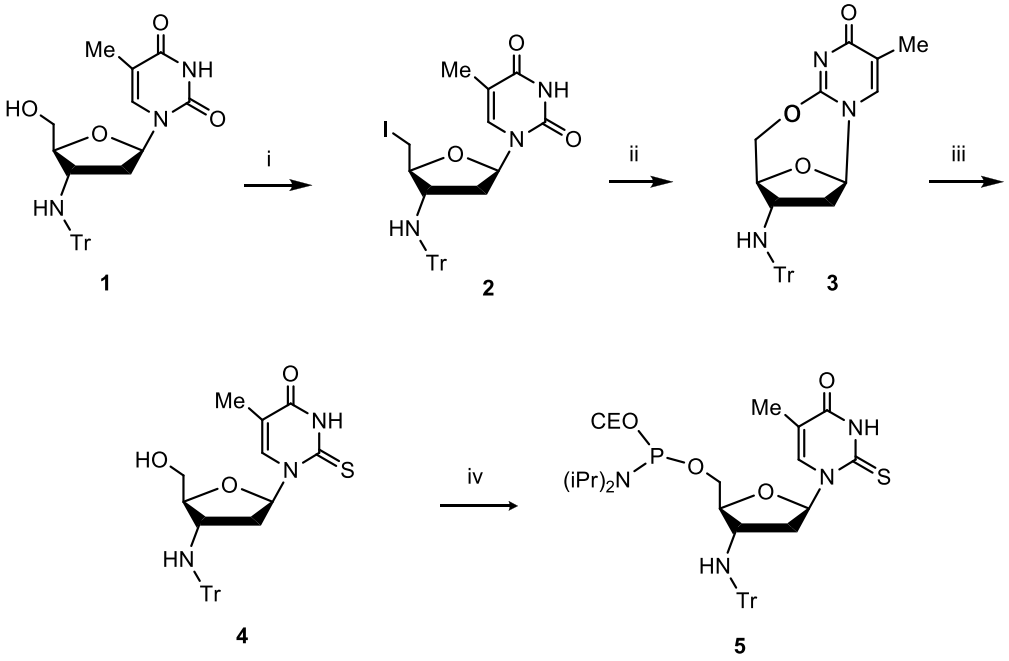
*3'-Azido-3'-deoxy-2-thiothymidine (Compound **6**).*
To a solution of **5** (62 mg; 0.106 mmol) in anhydrous dichloromethane (2.0 ml) was dropwise added 0.1 ml 5% dichloroacetic acid at room temperature in two portions. After stirring at room temperature for 30 min, the resulting red reaction mixture was concentrated under reduced pressure and partitioned between H_2O and $CHCl_3$. The organic layer was separated and dried over Na_2SO_4. After concentration, the residue was purified by flash column chromatography over silica gel methanol-chloroform (2%-10%) as the eluent to afford **6** (27 mg; 91% yield) as a white foam. ^{1}H NMR δ (400 MHz, CD_3OD): 8.17 (s, 1H), 6.91 (dd, J = 5.6 Hz, 6.0 Hz, 1H), 4.37-4.32 (m, 1H), 3.97-3.94 (m, 1H), 3.92-3.91 (m, 1H), 3.80-3.79 (m, 1H), 2.59-2.43 (m, 1H), 2.41-2.34 (m, 1H), 1.92 (s, 3H); ESI-HRMS calcd for $C_{10}H_{14}N_5O_3S^+$ $[M+H]^+$: 284.0812, found: 284.0819.

*3'-Azido-3'-deoxy-2-thiothymidine-5'-phosphor-2-methylimidazole (Compound **7**).*
Compound **6** (20 mg; 0.071 mmol) and proton sponge (18 mg; 0.085 mmol) were dried in a vacuum desiccator over P_2O_5 overnight before dissolving in trimethyl phosphate (1.0 ml). Then freshly distilled $POCl_3$ (8.0 μl; 0.29 mmol) was added dropwise at 0 °C. After stirring at 0°C for 1.0 h, 2-methylimidazole (30 mg; 0.353 mmol) was then added at 0 °C. After stirring for additional 4 h at room temperature, the reaction mixture was partitioned between H_2O and CH_2Cl_2. The crude product was further purified to afford **7** by reverse-phase preparative HPLC (Varian ProStar Preparative LC) on a Prep-C18 column (Varian Dynamax 250 × 21.4 mm) equilibrated with 25 mM triethylammonium bicarbonate, pH 8.0 and eluted with an acetonitrile linear gradient (0-60%). ^{1}H NMR δ (400 MHz, CD_3CN): 7.96 (s, 1H), 7.16 (s, 1H), 6.81(dd, J = 6.0 Hz, 6.4 Hz, 1H), 6.71 (s, 1H), 4.28-4.23 (m, 1H), 3.97-3.93 (m, 1H), 3.96-3.93 (m, 1H), 3.87 (m, 1H), 2.40-2.28 (m, 2H), 2.54 (s, 3H), 1.94 (s, 3H); ^{31}P NMR δ (168.1 MHz, CD_3CN): -10.67; ESI-MS calcd for $C_{14}H_{17}N_7O_5PS^-$ $[M^-]$: 426.08, found: 426.0.

*3'-Amino-3'-deoxy-2-thiothymidine-5'-phosphor-2-methylimidazolide (3'-NH_2-2-MeImpddsT **8**).*
To a solution of **7** (10 mg; 0.02 mmol) in a mixture solution of pyridine (1 ml) and 30% ammonium hydroxide solution (1.0 ml) was added triphenylphosphine (12 mg; 0.05 mmol) at room temperature. The reaction was then stirred for 5 h at room temperature. The resulting mixture was concentrated under vacuum and the residue was diluted with 1

ml of DMSO for $NaClO_4$ precipitation as previously described (4). The crude product was further purified by reverse-phase preparative HPLC as previously described to afford **8**. ^{1}H NMR δ (400 MHz, D_2O): 7.69 (s, 1H), 7.18 (s, 1H), 6.88 (dd, *J* = 4.8 Hz, 5.6 Hz, 1H), 6.83 (s, 1H), 4.16-4.12 (m, 1H), 4.08-3.95 (m, 2H), 3.66-3.61 (m, 1H), 2.45-2.34 (m, 1H), 2.46 (s, 3H), 1.87 (s, 3H); ^{31}P NMR δ (168.1 MHz, D_2O): -10.41; ESI-HRMS calcd for $C_{14}H_{19}N_5O_5PS^-$ $[M-H]^-$: 400.0850, found: 400.0862.

Efforts to synthesize 3′-NH-trityl-2-thiothymidine-5′-O-phosphoramidite.



Where: i - Ph_3P/I_2; ii - $AgNO_3$; iii - H_2S/gas; iv - CEO-P[N(iPr)$_2$]$_2$/Tetr/NH(iPr)$_2$

The 5′-O-phosphoramidite of 3′-NH-trityl-2-thiothymidine is the protected monomer required for the solid phase synthesis of 3′-NP-DNA oligomers containing 2-thio substituted T. Because the route that we employed to generate the 5′-phosphor-(2-methyl)-imidazolide is too inefficient to yield the amounts of the amidite required for solid phase synthesis, we sought a more efficient synthesis starting with 3′-NH-trityl-thymidine. We converted this to the 5′-iodo derivative, and then obtained the 2-5'-anhydro cyclic derivative in good yield. Unfortunately, all efforts to convert this compound to 3′-NH-trityl-2-thiothymidine by nucleophilic attack of attack of H_2S on the C2 position failed. We obtained a complex reaction mixture of at least five (detectable by TLC and RP HPLC) products, with no major reaction product in the mixture. The products were separated and isolated using high-resolution silica gel column chromatography. None of the isolated compounds contained the 2-thiothymine heterocyclic base (characteristic UV Abs. max. at ~280nm), as judged by UV spectral analysis, *c.f. with* commercially available sample of 2-thiothymidine. We are currently investigating alternative synthetic strategies in search of an efficient route to the desired phosphoramidite.

References:

1. Jain ML & Bruice TC (2006) Solid-phase synthesis of positively charged deoxynucleic guanidine (DNG) oligonucleotide incorporating 7-deazaguanine bases. *Bioorg Med Chem* 14(21):7333-7346.
2. Zhang S, Zhang N, Blain JC, & Szostak JW (2013) Synthesis of N3'-P5'-linked phosphoramidate DNA by nonenzymatic template-directed primer extension. *J Am Chem Soc* 135(2):924-932.
3. Eisenhuth R & Richert C (2009) Convenient syntheses of 3'-amino-2',3'-dideoxynucleosides, their 5'-monophosphates, and 3'-aminoterminal oligodeoxynucleotide primers. *J Org Chem* 74(1):26-37.
4. Zhang N, Zhang S, & Szostak JW (2012) Activated ribonucleotides undergo a sugar pucker switch upon binding to a single-stranded RNA template. *J Am Chem Soc* 134(8):3691-3694.

SI Figures

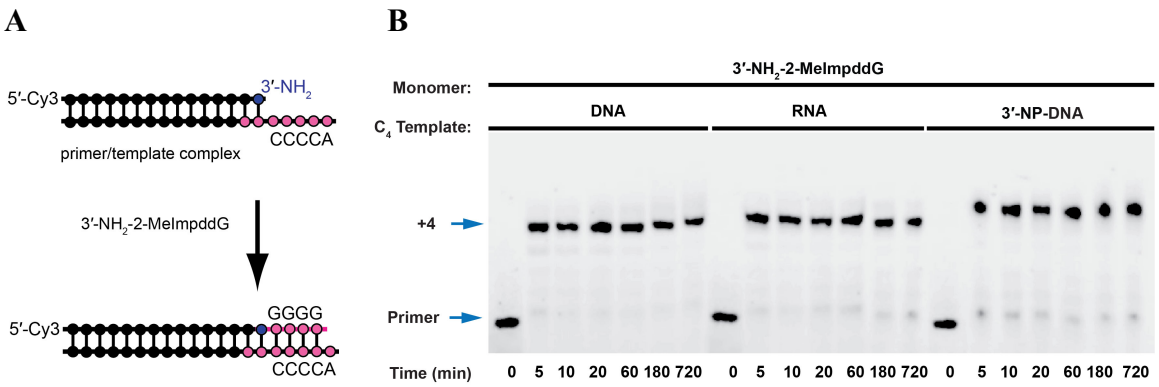


Fig. S1. Comparison of non-enzymatic primer extension reaction using 3′-NH_2-2-MeImpddG as a monomer on DNA, RNA and 3′-NP-DNA templates. (A) Reaction scheme for non-enzymatic primer extension reaction. Red segments indicate phosphoramidate bonds. (B) PAGE analysis of primer-extension products on indicated templates. Primer extension reactions were carried out as previously described, and the reaction was initiated by addition of 5.0 mM 3′-NH_2-2-MeImpddG. Arrows indicate primer and full-length product.

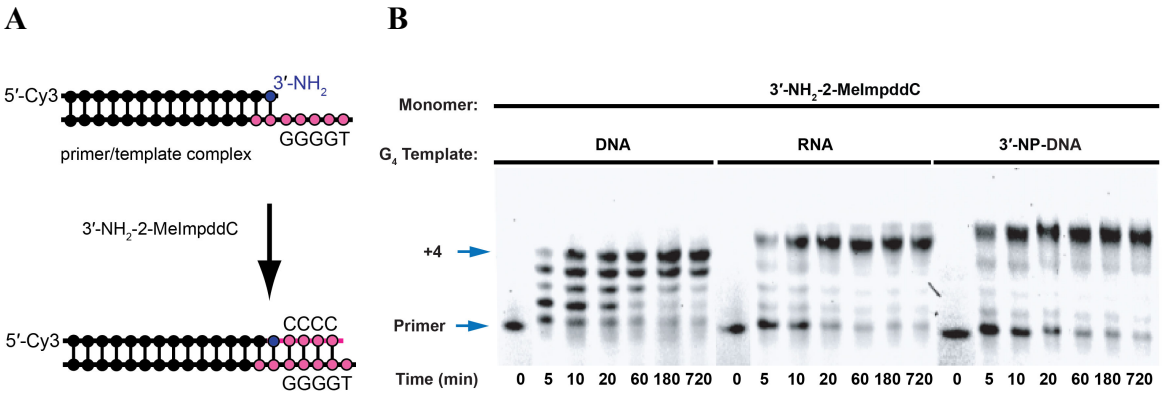


Fig. S2. Comparison of non-enzymatic primer extension reaction using 3′-NH_2-2-MeImpddC as a monomer on DNA, RNA and 3′-NP-DNA templates. (A) Reaction scheme for non-enzymatic primer extension reaction. Red segments indicate phosphoramidate bonds. (B) PAGE analysis of primer extension products on indicated templates. Primer extension reactions were carried out as previously described, and the reaction was initiated by addition of 5.0 mM 3′-NH_2-2-MeImpddC. Arrows indicate primer and full-length product.

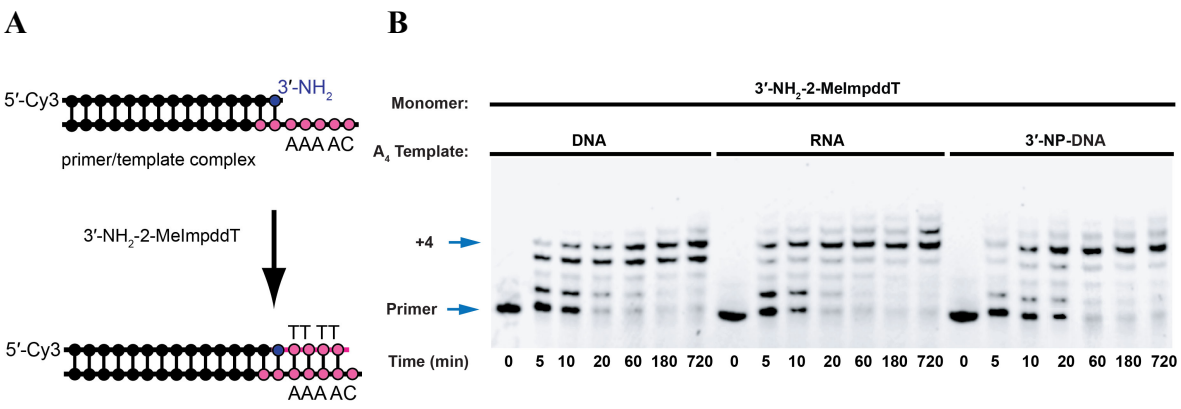


Fig. S3. Comparison of non-enzymatic primer extension reaction using 3′-NH_2-2-MeImpddT as a monomer on DNA, RNA and 3′-NP-DNA templates. (A) Reaction scheme for non-enzymatic primer extension reaction. Red segments indicate phosphoramidate bonds. (B) PAGE analysis of primer extension products on indicated templates. Primer extension reactions were carried out as previously described, and the reaction was initiated by addition of 10.0 mM 3′-NH_2-2-MeImpddT. Arrows indicate primer and full-length product.

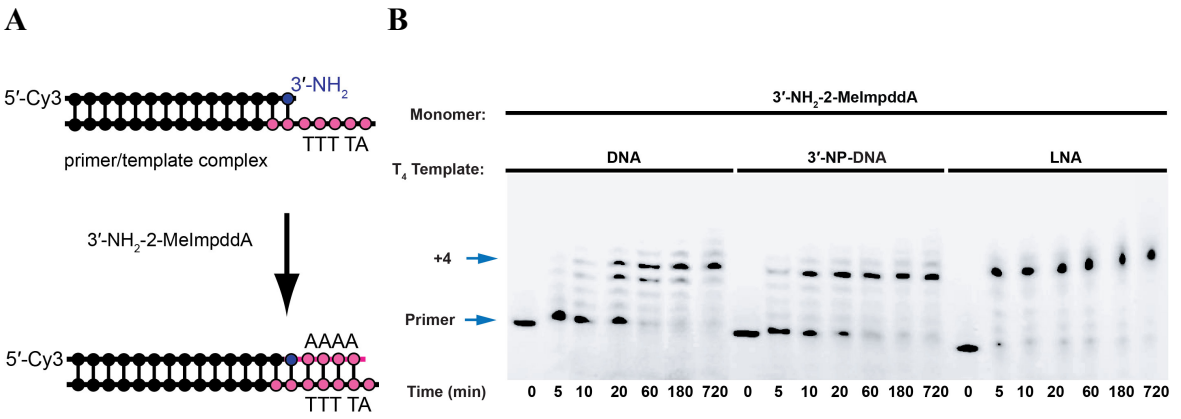
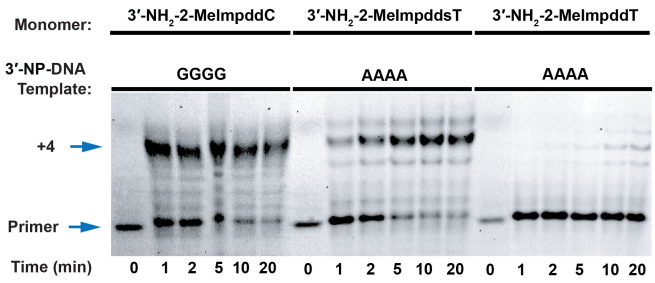


Fig. S4. Comparison of non-enzymatic primer extension reaction using 3′-NH_2-2-MeImpddA as a monomer on DNA, 3′-NP-DNA and LNA templates. (A) Reaction scheme for non-enzymatic primer extension reaction. Red segments indicate phosphoramidate bonds. (B) PAGE analysis of primer-extension products on indicated templates. Primer-extension reactions were carried out as previously described, and the reaction was initiated by addition of 10.0 mM 3′-NH_2-2-MeImpddA. Arrows indicate primer and full-length product.

A) Monomer concentration: 1.0 mM



B) Monomer concentration: 2.5 mM

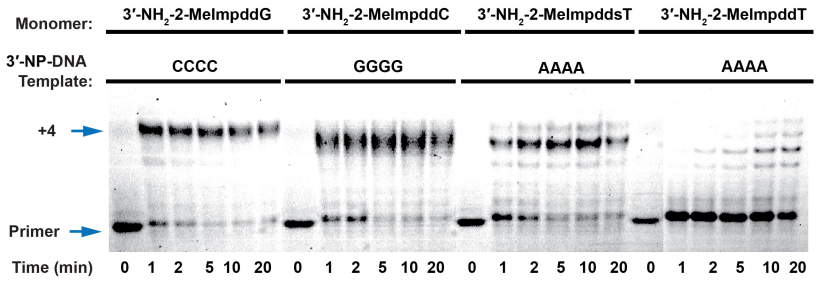


Fig. S5. Non-enzymatic primer-extension reactions on different 3′-NP-DNA templates with their complementary 3′-NH_2-2-MeImpddN monomers. PAGE analysis of primer-extension products on indicated templates. Primer-extension reactions were carried out as previously described, and the reaction was initiated by addition of 1.0 mM (A) and 2.5 mM (B) 3′-NH_2-2-MeImpddN as indicated. Arrows indicate primer and full-length product.

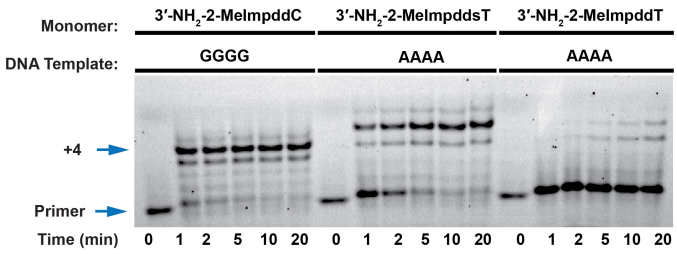
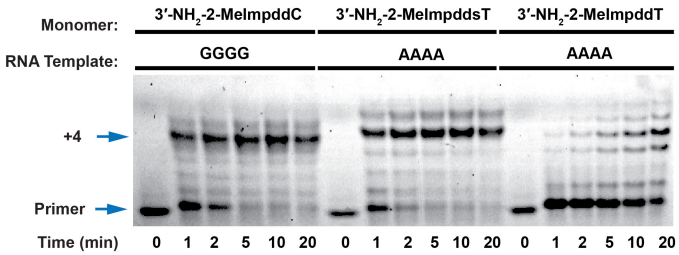
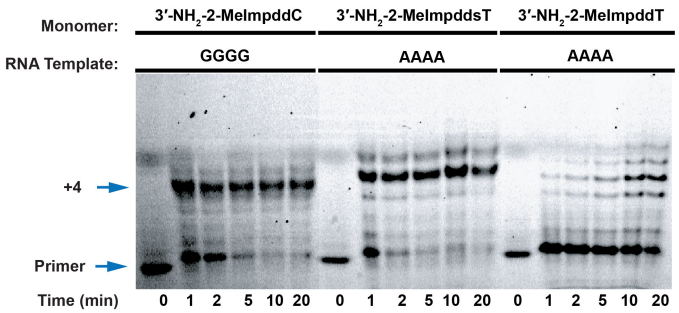


Fig. S6. Non-enzymatic primer-extension reactions on different DNA templates with their complementary 3′-NH_2-2-MeImpddN monomers. PAGE analysis of primer-extension products on indicated templates. Primer-extension reactions were carried out as previously described, and the reaction was initiated by addition of 1.0 mM 3′-NH_2-2-MeImpddN as indicated. Arrows indicate primer and full-length product.

A) Monomer concentration: 5.0 mM



B) Monomer concentration: 2.5 mM



C) Monomer concentration: 1.0 mM

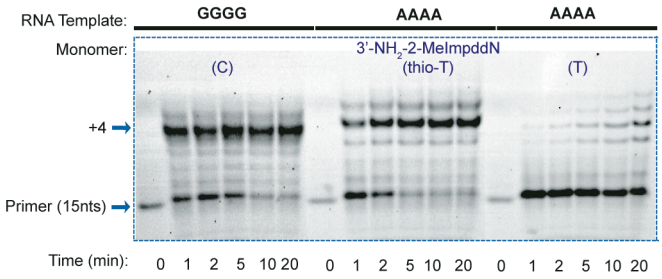


Fig. S7. Effect of monomer concentration on non-enzymatic primer extension reactions on RNA templates. PAGE analysis of primer extension products on indicated templates. Primer extension reactions were carried out as previously described, and the reaction was initiated by addition of 5.0 mM (A), 2.5 mM (B) and 1.0 mM (C) 3′-NH_2-2-MeImpddN as indicated. Arrows indicate primer and full-length product.

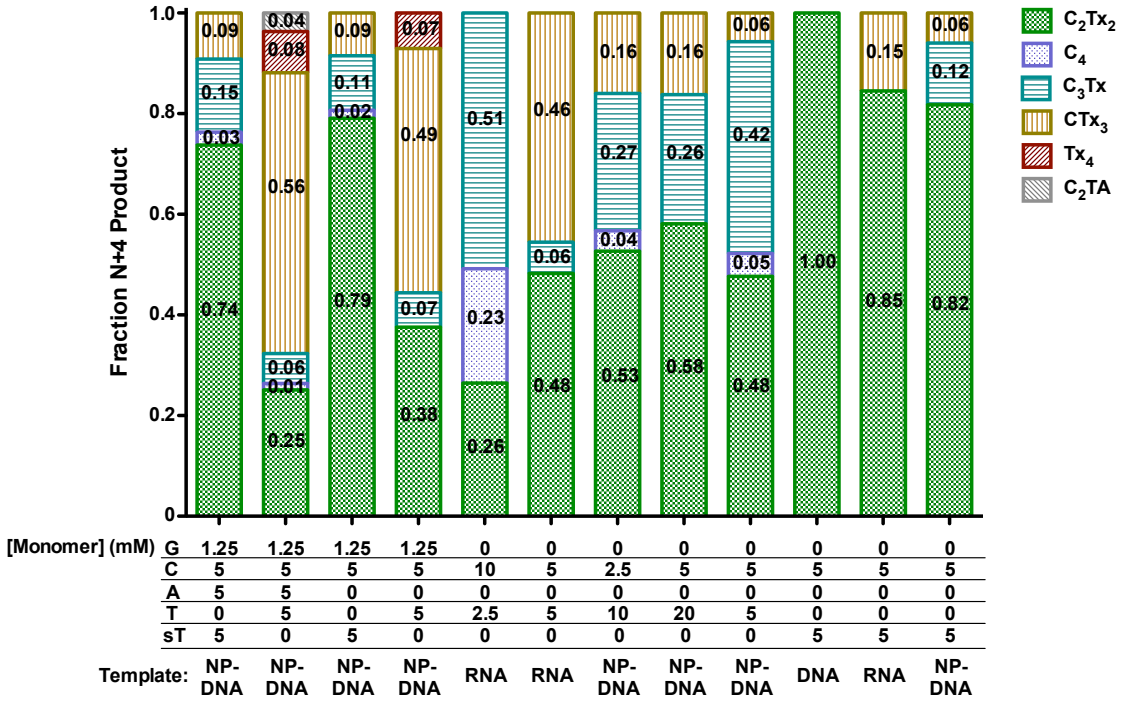


Fig. S8. Comparison of fidelity of full-length N+4 products from copying 5′-AGAG-3′ templates. NP-DNA: a chimeric DNA/3′-NP-DNA oligonucleotide 5′-CAGAGGACTATCGGC-3′ (3′-NP-DNA underlined).

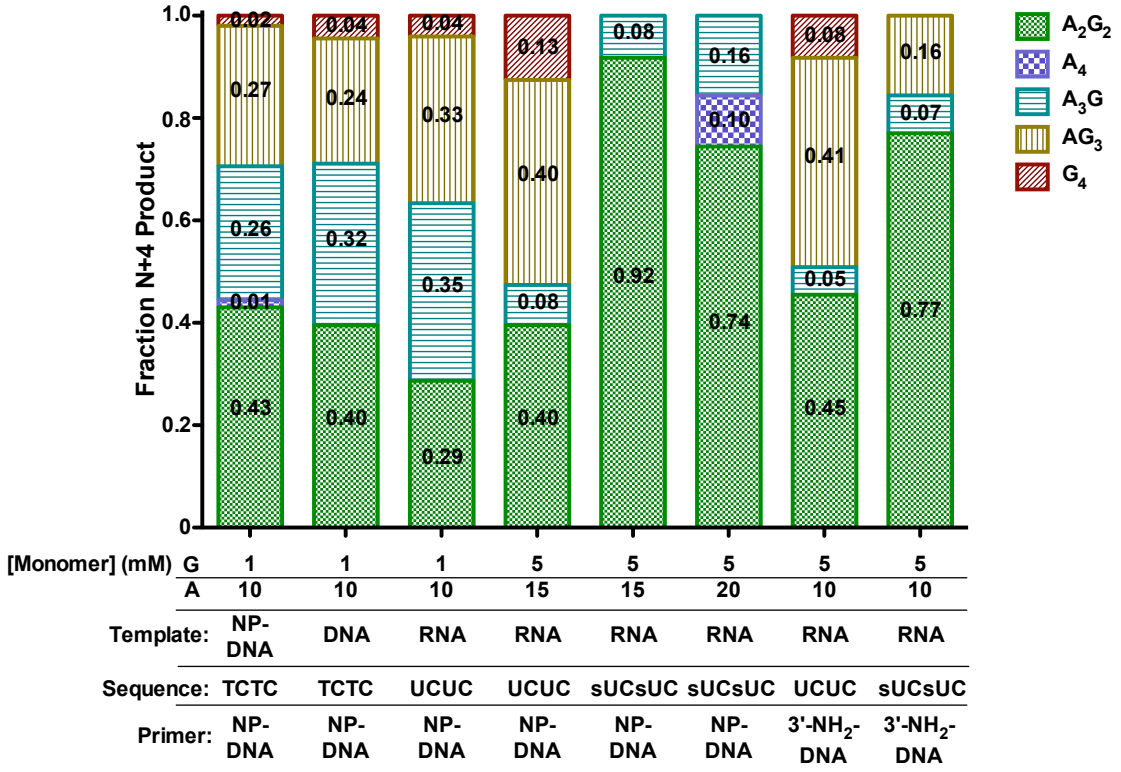


Fig. S9. Comparison on fidelity of full-length N+4 products from copying a template region of 5′-TCTC/UCUC/sUCsUC. 1) NP-DNA primer: an entirely 3′-NP-DNA oligonucleotide (linked by N3′-P5′-phosphoramidate bonds); 2) NP-DNA TCTC template: a chimeric DNA/3′-NP-DNA oligonucleotide 5′-ATCTCGACTATCGGC-3′ (3′-NP-DNA underlined).

SI Tables

Table S1. LC-MS sequencing of the crude product from non-enzymatic primer extension on a GAGA 3′-NP-DNA template.

Peak	Calculated monoisotopic mass	Observed monoisotopic mass	Error (ppm)	Observed mass difference between ladder oligos	Ladder sequence[a,b]	Ladder oligo length	Note
1	4230.9111	4230.8962	3.52	-	5′-GCCGATAGTCCsTCsT-3′	14	Full-length
2	3911.8719	3911.8594	3.20	319.0368	5′-GCCGATAGTCCsTC-3′	13	
3	3623.8096	3623.7942	4.24	288.0652	5′-GCCGATAGTCCsT-3′	12	
4	3304.7704	3304.7619	2.57	319.0323	5′-GCCGATAGTCC-3′	11	
5	3016.7080	3016.7002	2.60	288.0617	5′-GCCGATAGTC-3′	10	Primer

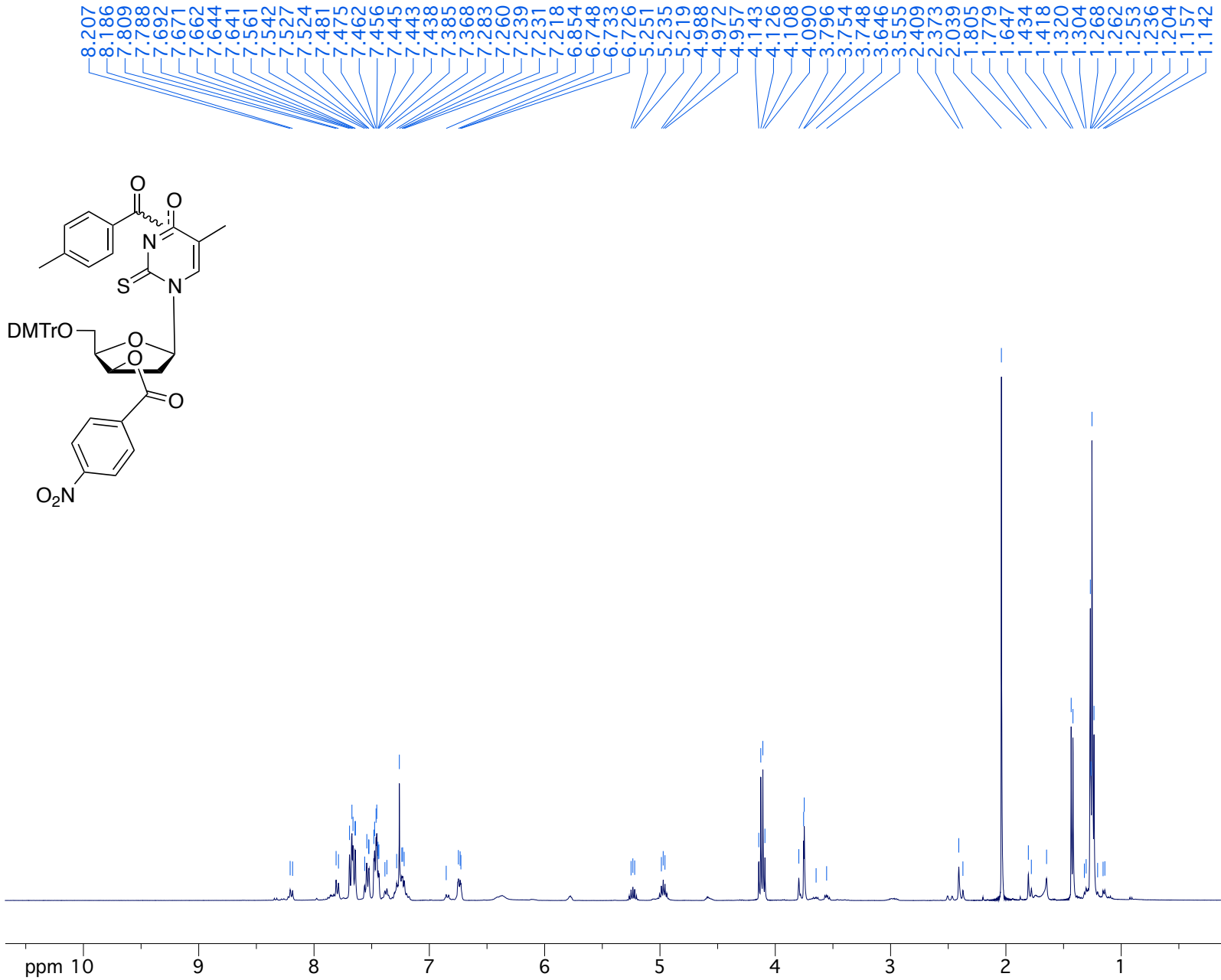
a. All sequences are linked with N3′-P5′ phosphoramidate bonds
b. 3′-end is a NH_2 group

Table S2. LC-MS Sequencing of the crude product from non-enzymatic primer extension on a 20-mer sUCsUC RNA template.

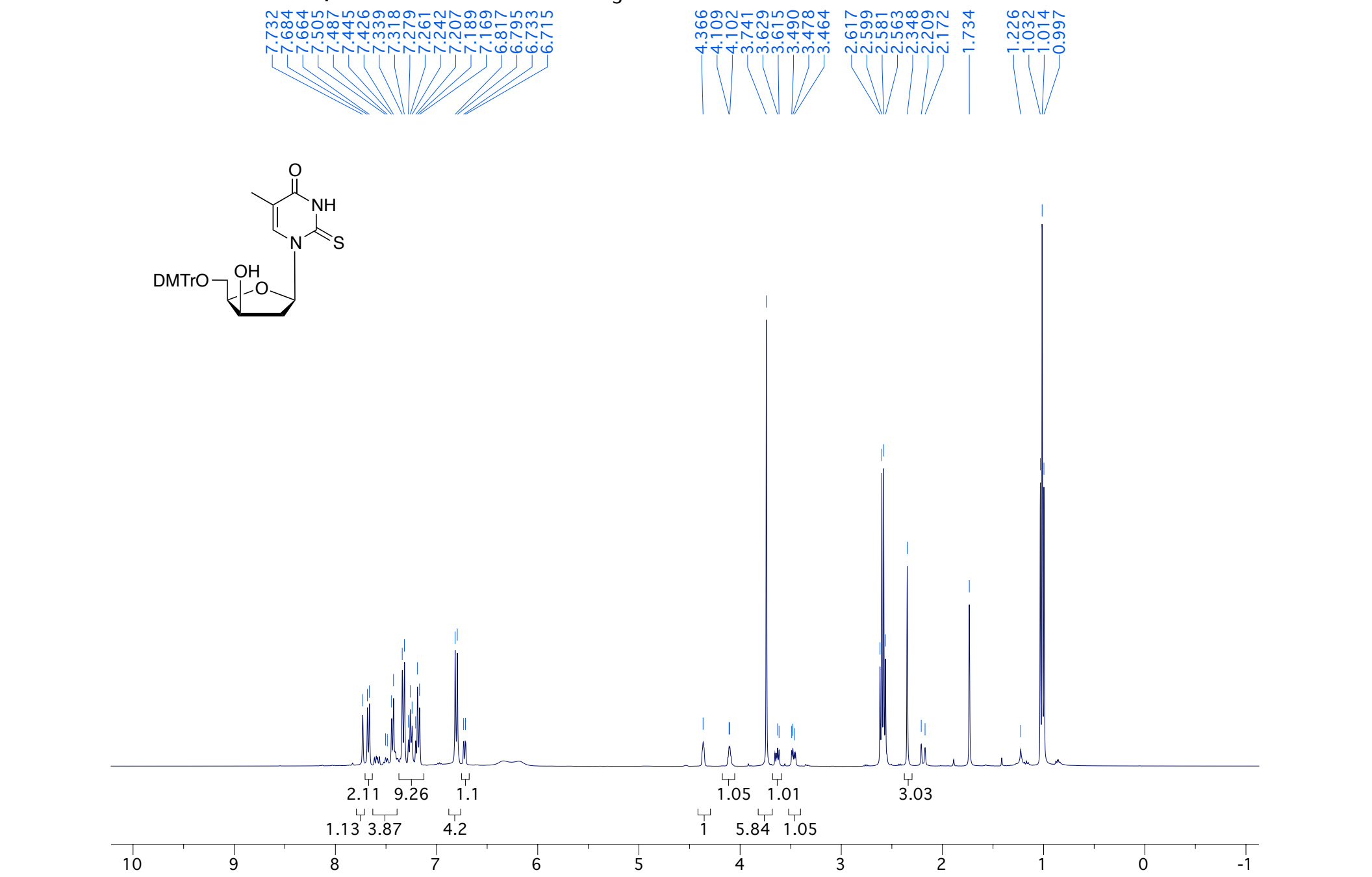
Peak	Calculated isotopic mass with 3 ^{13}C	Observed isotopic mass with 3 ^{13}C	Error (ppm)	Observed mass difference between ladder oligos	Ladder sequence[a,b]	Ladder oligo length	Note
1	6536.3229	6536.3393	-2.51	-	**N**-GAGA-3′	19	Full-length
2	6224.2493	6224.2364	2.07	312.1029	**N**-GAG-3′	18	
3	5896.1808	5896.1870	-1.05	328.0494	**N**-GA-3′	17	
4	5584.1072	5584.0953	2.13	312.0917	**N**-G-3′	16	
5	5256.0387	5256.0634	-4.70	328.0319	**N**	15	Primer

a. 3′-end of all oligos are NH_2-terminated
b. **N** = Primer (5′-TAMRA-GCG TAG ACT GAC TGG-3′)

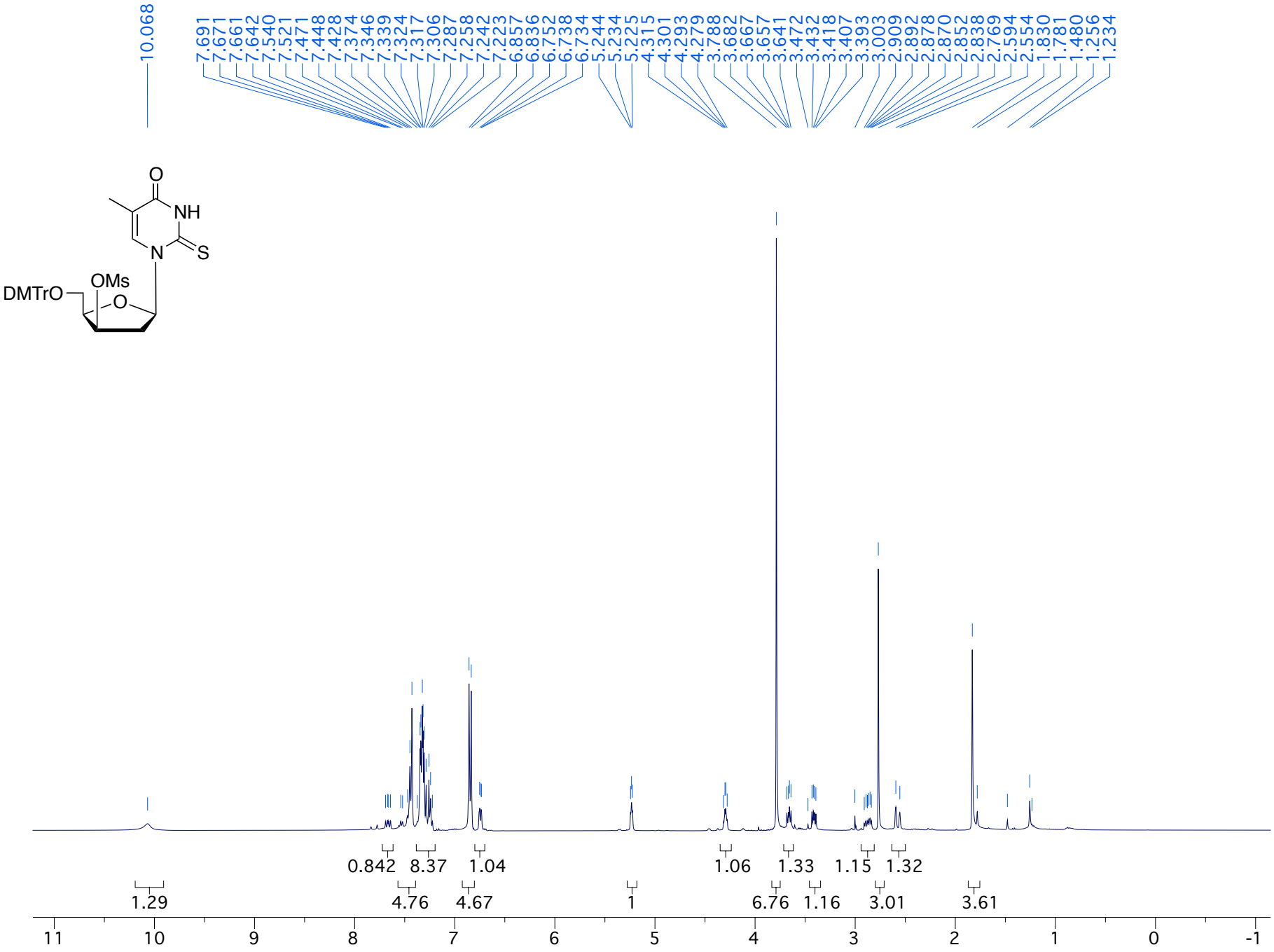
^{1}H-NMR for compound **2**, Solvent: $CDCl_3$



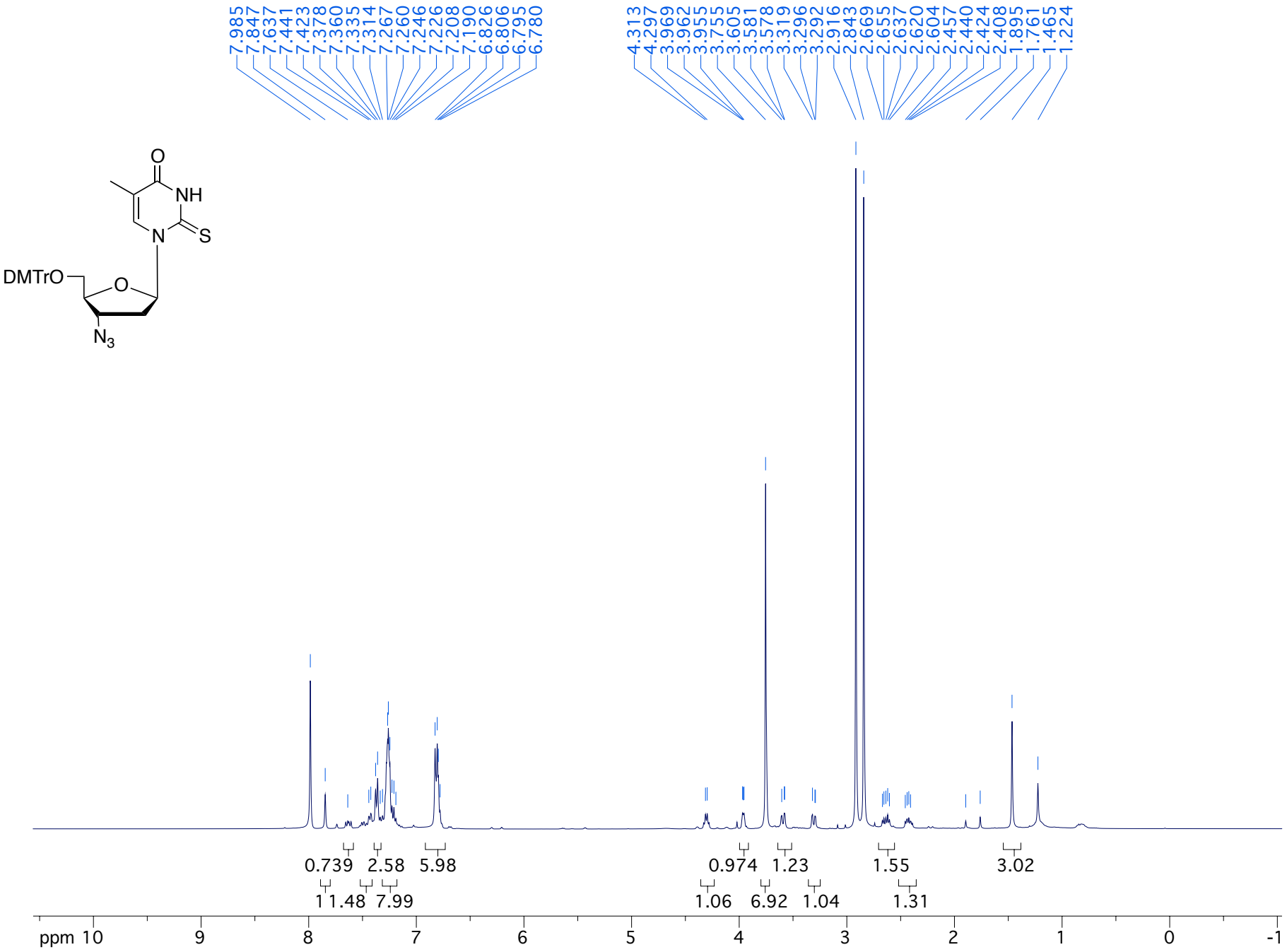
^{1}H-NMR for compound **3**, Solvent: $CDCl_3$



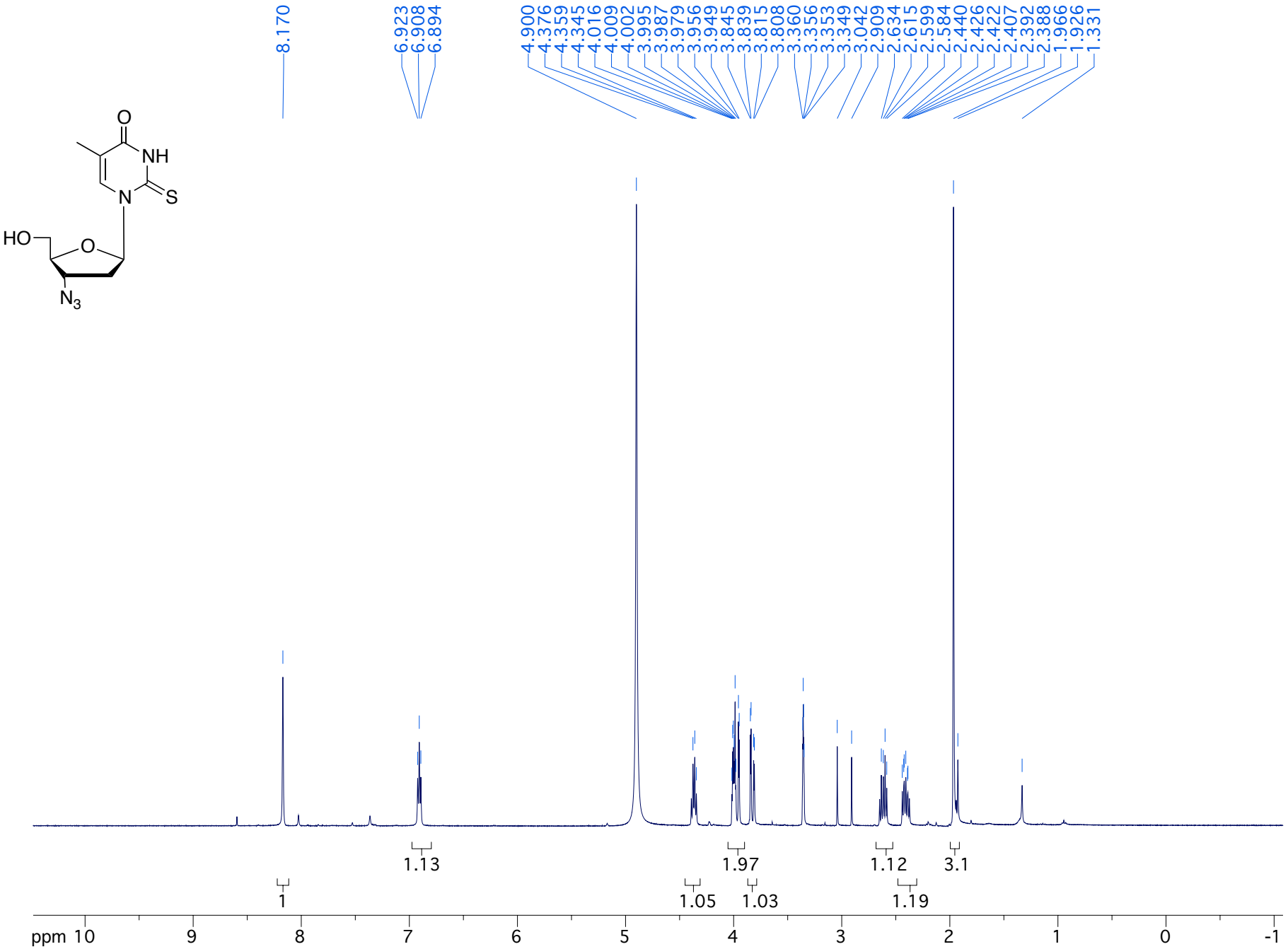
^{1}H-NMR for compound **4**, Solvent: $CDCl_3$



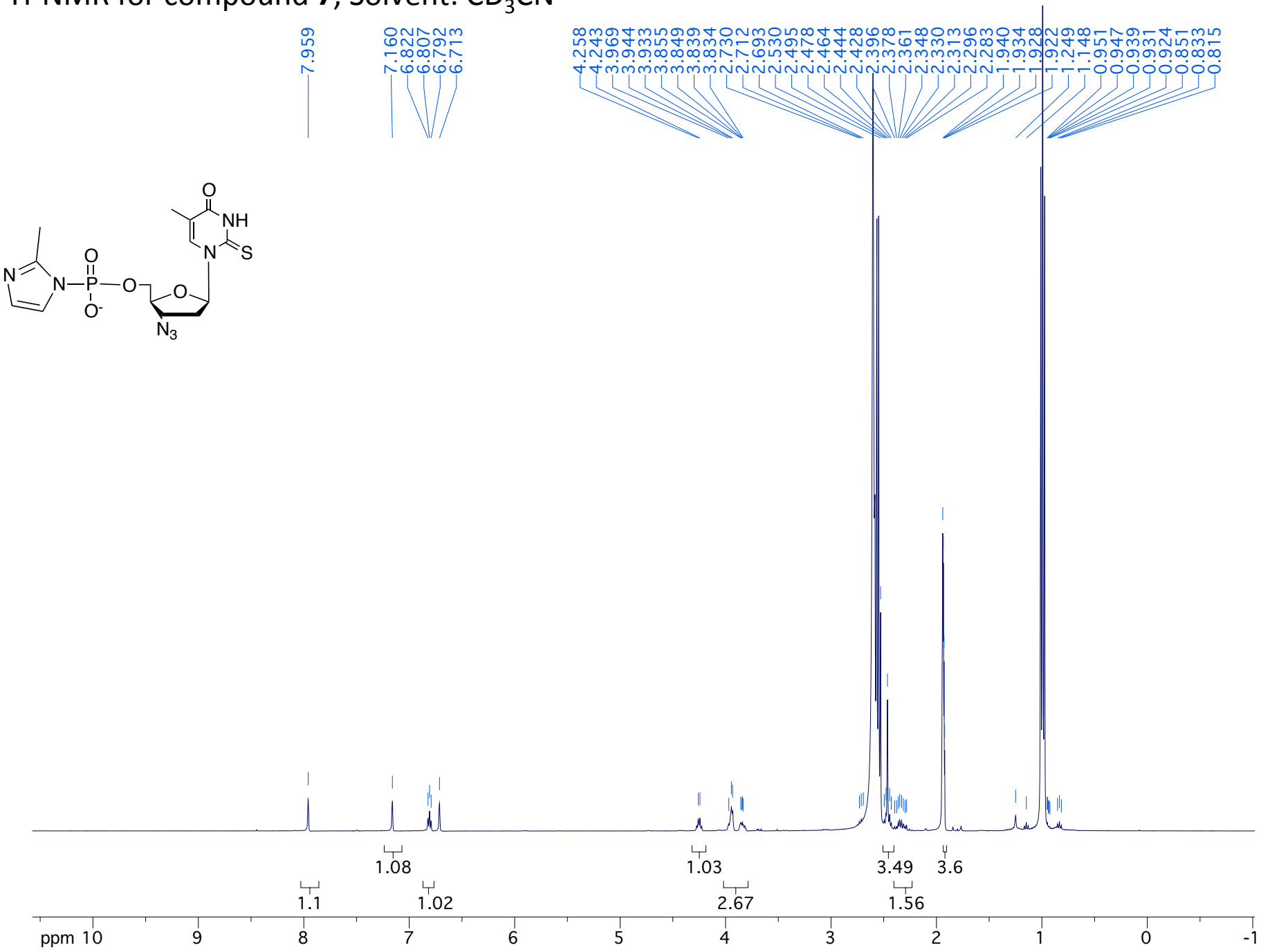
^{1}H-NMR for compound **5**, Solvent: $CDCl_3$



^{1}H-NMR for compound **6**, Solvent: CD_3OD



^{1}H-NMR for compound **7**, Solvent: CD_3CN



^{1}H-NMR for compound **8**, Solvent: D_2O