Synthesis of Glycine-containing Nucleotide Analog for Developing a Novel Chemo-enzymatic DNA Labeling Method

A thesis submitted to

Indian Institute of Science Education and Research Pune

in partial fulfilment of the requirements for the

BS-MS Dual Degree Programme

by

Roshni Rani Khamari

20161069

Indian Institute of Science Education and Research Pune

Dr. Homi Bhabha Road,

Pashan, Pune 411008, INDIA.

July 2021

Project Supervisor

Prof. Seergazhi G. Srivatsan

This thesis is dedicated to my parents.

Certificate

This is to certify that this dissertation entitled "Synthesis of glycine-containing nucleotide analog for developing a novel chemo-enzymatic DNA labeling method" towards the partial fulfilment of the BS-MS dual degree program at the Indian Institute of Science Education and Research, Pune represents study/work carried out by Roshni Rani Khamari at Indian Institute of Science Education and Research under the supervision of Dr. S. G. Srivatsan, Professor, Department of Chemistry, during the academic year 2020-2021.

Prof. S. G. Srivatsan

Committee:

Dr. S. G. Srivatsan

Dr. Amrita Hazra

Declaration

I hereby declare that the matter embodied in the report entitled "Synthesis of glycine-containing nucleotide analog for developing a novel chemo-enzymatic DNA labeling method" are the results of the work carried out by me at the Department of Chemistry, Indian Institute of Science Education and Research, Pune, under the supervision of Prof. S. G. Srivatsan and the same has not been submitted elsewhere for any other degree.

Roshni Rani Khamari

Roshni Rani Khamari
Date: 31st July, 2021

Acknowledgment

It is with great pleasure that I express my sincere gratitude to my project supervisor, Prof. S.G. Srivatsan, for trusting me to do an independent project in his lab at IISER Pune. He gave me a space to commit mistakes, rectify and relearn from them. I could do experimentation in a real sense, even in the restricted time available in the covid situation. I did learn how one has to first try, if failed, try again and rise in the experiments as well as in life. Apart from lab experience, one of the best things I learned from sir is; indeed, there will be hurdles coming one after another, but fighting to overcome those makes one successful in every aspect of life.

I would like to convey my gratitude to my TAC member Dr. Amrita Hazra for her valuable suggestions and for encouraging me to think about the future aspects of this project. I appreciate the efforts put by my former and present lab members Dr. Manisha, Dr. Ragini, Saddam, Akansha, Pulak, Sarupa, Swagatta, Amit, Indu, and Arindam, to help me out whenever I faced difficulties. I thank them for maintaining an enjoyable work environment in the lab.

I am forever indebted to Sandhya di without whose help I could have never progressed to this extend in synthesis. Her immense knowledge and willingness to teach, like a blessing for me. Her dedication to work, approach to any problem, positive attitude, and calmness of mind gave me great motivation. I am thankful to Jayashree di for giving me new ideas to experiment with whenever I was stuck.

I am forever thankful to my parents and grandparents for being my backbone and their constant love, care, and nurture. And last but not least, a shout out to my friends Ankit, Pallavi, Shreyash, Hitesh, Yasharth, KK, and Rahul for being stress relievers and constant supporters throughout IISER life.

Roshni Rani Khamari

Table of Contents

List of Figures

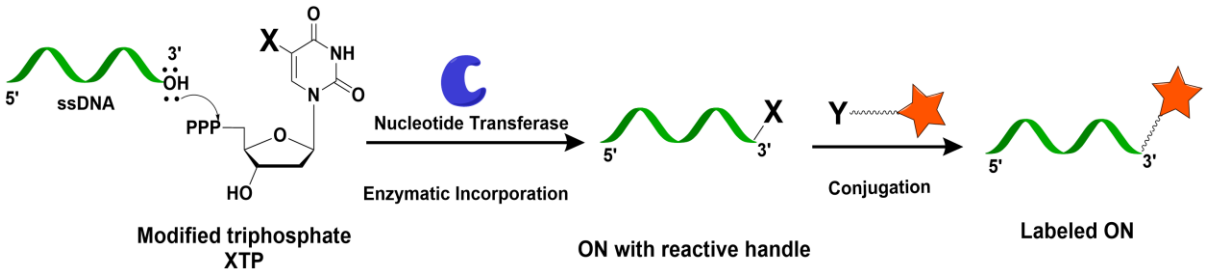
Abbreviation

ON	oligonucleotide
NMR	Nuclear Magnetic Resonance
EPR	Electron Paramagnetic Resonance
RNA	Ribonucleic Acid
DNA	Deoxyribonucleic Acid
ssDNA	single-stranded Deoxyribonucleic Acid
pH	potential of hydrogen
CuAAC	Copper-catalyzed azide–alkyne cycloaddition
SPAAC	Strain-promoted azide–alkyne cycloaddition
IEDDA	Inverse electron demand Diels–Alder reaction
E_L	latent electrophile
Nu_L	latent nucleophile
UTP	uridine triphosphate
TUTase	Terminal Uridylyl Transferase
TdT	Terminal deoxynucleotidyl transferase
TLC	Thin-Layer Chromatography
TBAF	Tetra-n-butylammonium fluoride
UV	Ultraviolet
TMS	tetramethylsilane
HRMS	High-resolution mass spectrometry
IR	Infrared
TEA	triethylamine
equiv.	equivalents
ESI	Electrospray ionization
TFA	trifluoroacetic acid
DMF	Dimethylformamide
MeOH	Methanol
DCM	Dichloromethane
EA	Ethyl Acetate
THF	tetrahydrofuran
TBDMS-Cl	*tert*-Butyldimethylsilyl chloride
Pd/C	palladium on carbon
DMSO- d_6	Dimethylsulfoxide-d_6
r.t:	room temperature
h.	hours

mmol	millimoles
mL	milliliters
mg	milligrams
g	grams
v/v	volume per volume

Abstract

Nucleic acid functionalizing technologies greatly aid in understanding the structure and functioning of nucleic acids. A chemoselective reaction has become one of the most powerful techniques for labeling nucleic acids in recent years. However, only a few qualify as bioorthogonal reactions that enable the labeling of nucleic acids *in vitro* and cellular environments. Therefore, the development of new nucleic acid labeling techniques remains a high priority. In this project, we intend to develop a chemoselective nucleic acid labeling technique by using glycine-modified nucleotide. The method generates an imine that reacts with an aldehyde primed with a tag to deliver an aminoalcohol. The formation of a latent electrophile followed by glycine residue-specific generation of nucleophilic intermediate is the key to attain chemoselectivity. The symmetric bis-aldehyde reagent can be used to install a variety of affinity tags and biophysical probes for potential application in the analysis of nucleic acids.



1.Introduction

The nucleic acids play a multifaceted role in the physiology of the cell. They store genetic information, regulate gene expression, and catalyze by interfacing with proteins, nucleic acids, and small molecules. They can expand their functional range by adopting a range of 3-dimensional structures and rapidly interconverting among them. The study of nucleic acid biogenesis, structure, dynamics, and structure-function relationships call for various biophysical and theoretical approaches. A great deal of biophysical exploration is based on fluorescence, nuclear magnetic resonance (NMR), electron paramagnetic resonance (EPR), X-ray crystallography, and microscopy methods.[1-5] Monomer of these universal biopolymers, i.e., nucleotides are inherently non-emissive, and labels do not exist in intrinsic form. Therefore, majority of nucleic acid study uses various labeling methods which allow access to the labeled molecule with a suitable reporter molecule. Reporter molecules can be either be bonded covalently to the phosphate backbone, sugar or nucleobase, or can be introduced noncovalently along the grooves.[6] The covalent labeling is achieved (a) chemically by solid-phase oligonucleotide (ON) synthesis, (b) enzymatically by transcription reactions and (c) chemo-enzymatically by using transferase enzymes. To synthesize ON in the solid phase, a nucleoside phosphoramidite is modified and incorporated site-specifically into the ON, using a combination of coupling, capping, oxidation and deprotection reactions. But the efficiency of incorporation decreases with increasing ON sequence length which leads to poor yields.[7] Though solid-phase synthesis is widely used but sometimes sensitive modifications cannot survive the harsh chemical manipulations and the overall process is laborious also. Alternatively, the ability of certain DNA and RNA polymerases to accept functionalized nucleotide substrates has allowed the incorporation of biophysical probes into nucleic acids. Further, as enzymatic incorporation involves mild reaction conditions, it can be used efficiently for sensitive modifications. Unlike the solid-phase synthesis method, the length of the ON sequence does not limit the efficacy of enzymatic incorporation. But the problems that come with enzymatic labeling techniques are (i) introducing modifications site-specifically is difficult and (ii) substrates containing bulky modifications are not well tolerated by the enzymes.[7] In order to overcome these challenges, bioorthogonal reactions were developed to modify ON via postsynthetic chemical reactions.

1.1. Bioorthogonal chemistry as a labeling tool

Chemical techniques that aid in selective modification of nucleic acids are beneficial because the vast complexity of cellular systems makes studying biomolecular structures in their native state a challenging undertaking. The basis of this technology is bioorthogonal chemical reactions (i.e., a reaction that does not interfere with the biological process), where substrates react rapidly and selectively in the presence of a plethora of functional groups such as electrophiles, nucleophiles, oxidants, reductants in an aqueous environment. More than two decades ago came the first report on Staudinger ligation and its application in cellular environment[8] by Saxon and Bertozzi; since then, the family of bioorthogonal chemistry has snowballed and found wide applications in a variety of fields. Requirements for a reaction to be considered bioorthogonal are:

- Selectivity between endogenous functional groups.
- Biological and chemical inertness of reactive partners and resulting covalent linkage.
- Rapid reaction kinetics required to maintain reactive partners concentrations minimal for *in vivo* and inside cell labeling.
- Biocompatibility of reaction considering the pH, temperature, and aqueous environment of the biological system.

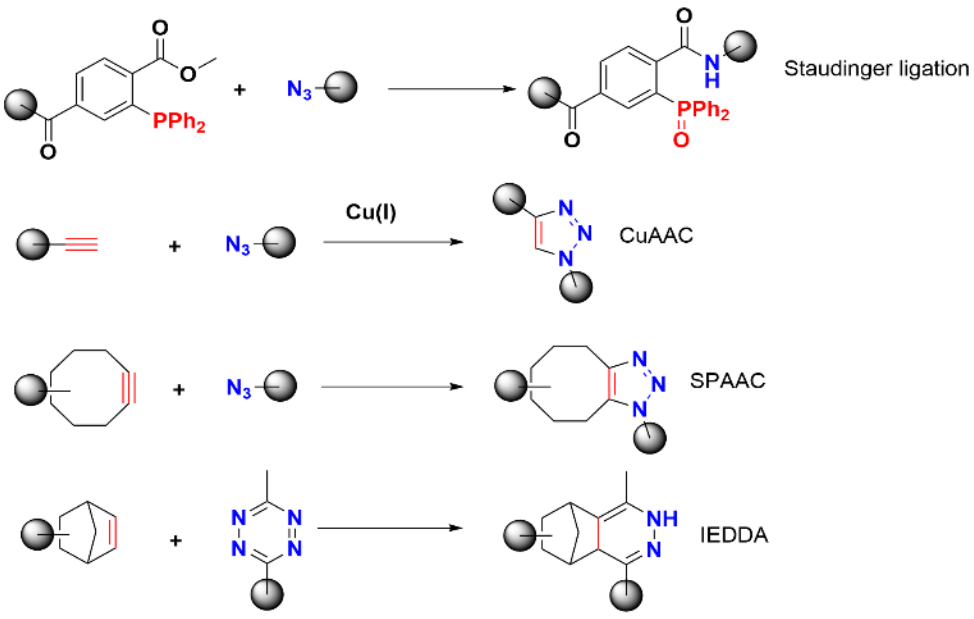


Figure 1. Commonly used bioorthogonal reactions for labeling nucleic acids.

Most commonly used bioorthogonal reactions in chemical biology includes Staudinger ligation, azide-alkyne cycloaddition catalyzed by copper (CuAAC), cycloaddition promoted by using strained alkyne substrates (SPAAC) and inverse-electron demand Diels-Alder (IEDDA) reactions[9] (Figure 1). The reagent-free Staudinger ligation uses the reaction between azide and phosphine moiety to produce an aza-ylide intermediate followed by hydrolysis to form a stable amide bond under water-tolerant conditions. The reaction offers great promise; however, instability of the phosphine substrates in the cellular environment and slow kinetics limits its range of applications.[7] CuAAC reaction developed independently by Sharpless[10] and Meldal[11] showed faster reaction kinetics, but cytotoxicity of copper was the primary concern for live cell applications. Copper-free reaction found by Bertozzi and coworkers[12], with up to 3 times faster reaction kinetics than the Staudinger ligation, is SPAAC reaction[9]. Similarly, the IEDDA reaction between highly strained trans-cyclooctene and electron-deficient tetrazines developed by Fox and co-workers has gained popularity for nucleic acid bioconjugation[13]. The main advantages of IEDDA reaction are its fast kinetics, specificity, biocompatibility, and reagent-free nature.

By utilizing the benefits of above mentioned bioorthogonal chemical reactions for postsynthetic ON modifications, limitations posed by solid-phase synthesis and enzymatic incorporation are partially negotiated and allows conjugation of various functional moieties to one modified ON.

1.2. Postsynthetic labeling

There are two general strategies for labeling ON, namely presynthetic and postsynthetic labeling (Figure 2). During presynthetic synthesis, isolated nucleotide monomers containing reporter molecules are constructed, which must be durable enough to withstand harsh chemical conditions during solid-phase synthesis, or modifications should be acceptable as substrate by enzymes and have the least steric limitations. The modified nucleotide with a label is incorporated either chemically or enzymatically before ON synthesis[14].

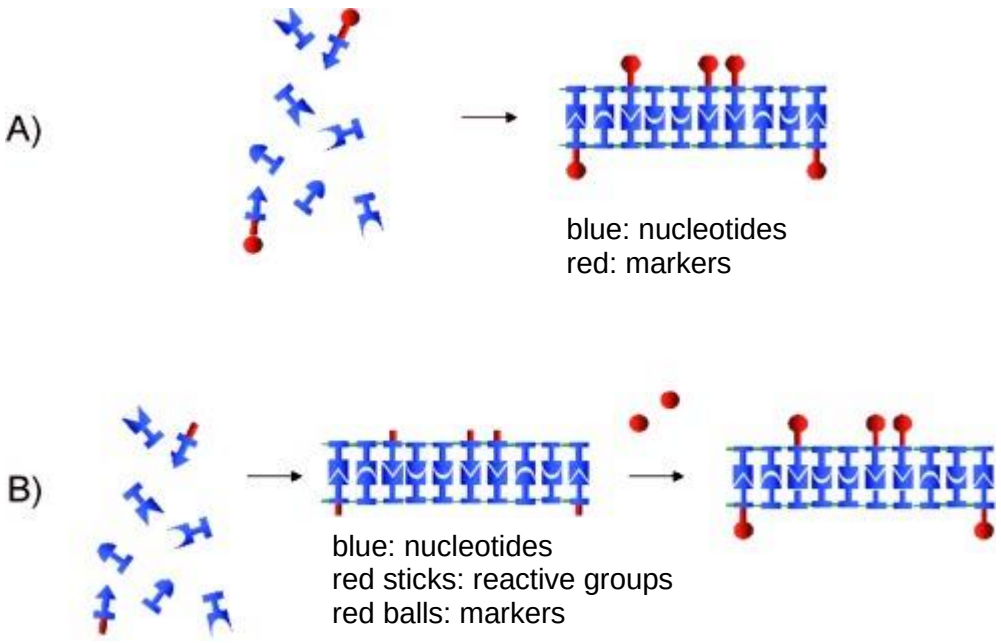


Figure 2. Schematic representation for (A) presynthetic and (B) postsynthetic labeling.[14]

The process of postsynthesis labeling begins with the introduction of nucleotides modified with reactive moiety during ON synthesis, which undergo chemoselective reaction with a reactive partner carrying the desired label after ON synthesis (Figure 3). This strategy can introduce reactive and sensitive moieties without much hindrance. In contrast to presynthetic method, a new label can be assigned without the synthesis of novel nucleotides. Because of this, postsynthetic labeling facilitates a high degree of modularity. However, difficulties in purifying incompletely labeled products from fully labeled constructs and low coupling yields can limit this approach.[14] The approach of combining the modularity of postsynthetic modification with the advantages of bioorthogonal chemical reactions can provide chemoselective, site-specific labeling under mild conditions avoiding the complicated purifications steps. Srivatsan and coworkers have already reported postsynthetic chemical method of functionalizing oligoribonucleotides with biophysical probes using Click, Staudinger[15], IEDDA[16,] and Suzuki cross-coupling reactions[17].

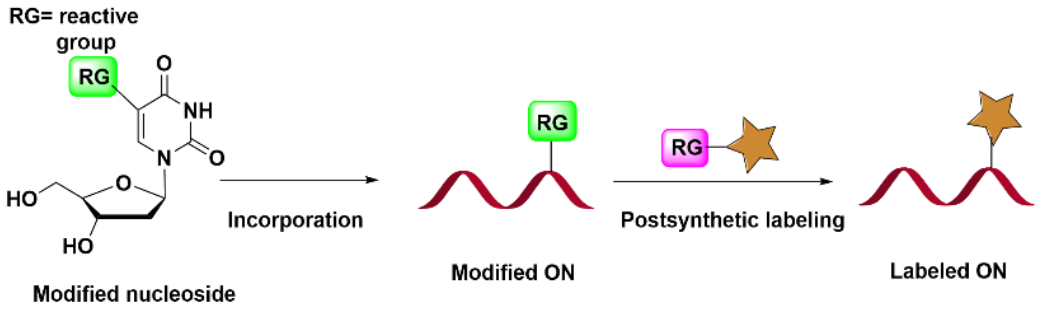


Figure 3. Schematic representation for postsynthetic labeling in nucleic acids.

1.3. Motivation

Recent work from Vishal Rai and coworkers has successfully labeled natural or engineered N-terminal glycine in proteins site specifically with excellent selectivity and efficiency.[18] It has been demonstrated that N-glycine is a unique target amidst other amino acids present in proteins. It provides a synthetic bioorthogonal aldehyde group tagged with a fluorophore, an affinity tag, and a ^{19}F NMR probe that can be used in the late-stage tagging process.

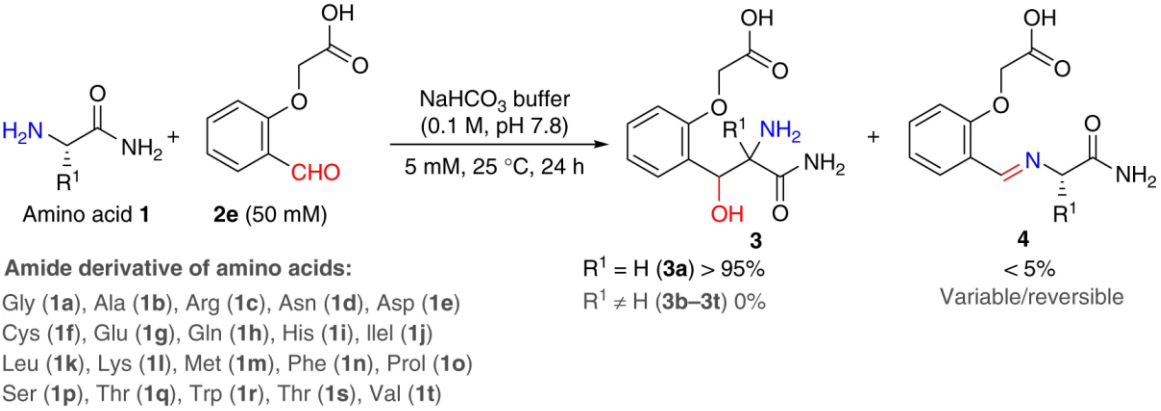


Figure 4. Selective formation of aminoalcohol with glycine, whereas other amino acids form an imine.[18]

Biomolecules are macromolecular substrates with multiple nucleophilic functional moieties. The functional group's ability for competition with other nucleophilic residues to react (chemoselectivity) and to compete with numerous copies of itself (site-selectivity) is therefore essential to achieve single-site labeling. To address the challenge of residue specificity and chemoselectivity simultaneously, Vishal Rai and coworkers hypothesized a multi-step transformation. It involves the initial generation of a latent electrophile (E_L) to address chemoselectivity, followed in a later stage by

generation of another reactive center, latent nucleophile (Nu_L), to provide residue specificity (Figure 5). By designing an imine with H-bond acceptors which facilitated proton shuttling, they were able to generate a stable Nu_L center. They carefully designed an ortho-methoxy benzaldehyde to avail the advantage of the constraint offered by the geometry of the aromatic ring. In the benzaldehyde designed, two oxygen centers were used as H-bond acceptors and one methoxy group ortho to aldehyde as one more H-bond acceptor. They have shown that for all proteinogenic amino acids except for glycine, the imine formed was inert towards any further transformations (Figure 4)**.** The researchers hypothesized that the H-bond acceptor oriented in an unfavorable manner was the cause of the high barrier to Nu_L formation in substituted amino acids.

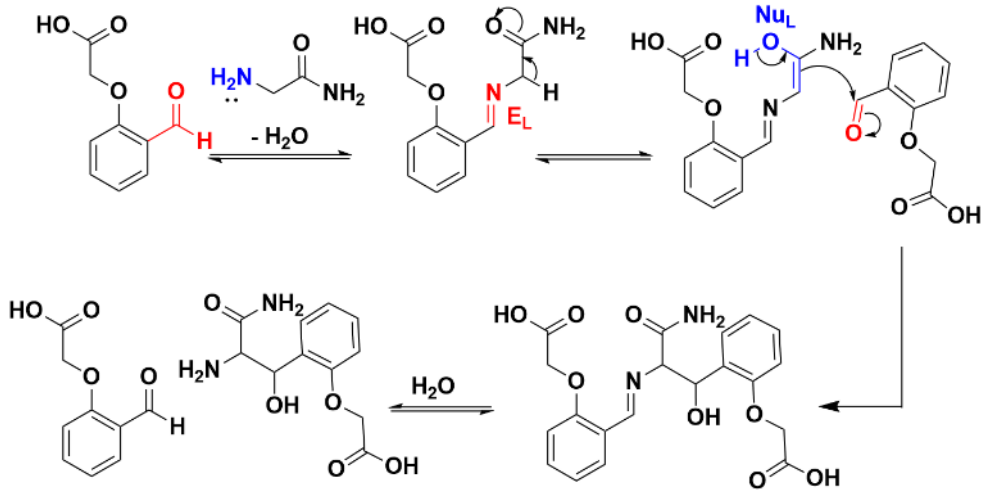


Figure 5. Plausible mechanism of formation aminoalcohol specific for glycine.

Proteins can be labeled for specific residues at single sites using the developed methodology. The reaction leads to the formation of aminoalcohol, which proceeds exclusively with N-terminus glycine in proteins. One of the notable features of this chemical technology is the absence of any competing reactions. Several tags can be installed on a single-site labeled protein using the symmetric bis-aldehyde reagent.

Inspired by this chemical methodology, we want to harness this chemoselective reaction in the context of nucleic acids. In this project, we plan to develop a modified nucleotide with glycine residue and use the aldehyde reagent with a label to achieve site-specific labeling in nucleic acids. The label can be any biophysical or imaging

probe, and we can build a library of probes to study the different aspects of the oligonucleotide sequence.

1.4. Project Objectives

Though there are numerous well-established ways to label nucleic acids, there is always a demand to develop a new, more efficient labeling technique to have diverse options. In this project, we aim to use the above reaction to label DNA ONs at the 3'-position chemo-enzymatically. Here, we want to synthesize a modified nucleotide with a chemically reactive glycine handle which will be incorporated into ONs enzymatically. The reactive glycine moiety will be reacted with an aldehyde group carrying a biotin label to form aminoalcohol (Figure 6). We envision that this methodology can provide new means to access functionalized DNA and RNA depending on the enzyme used. The advantage of this methodology is that it is reagent-free. The glycine-modified nucleotide can be inserted anywhere in the ON sequence using polymerases or nucleotide transferases, thereby allowing functionalization in different positions in the ON sequence. However, such a protocol is difficult in proteins where only N-terminal glycine can be functionalized by this methodology.

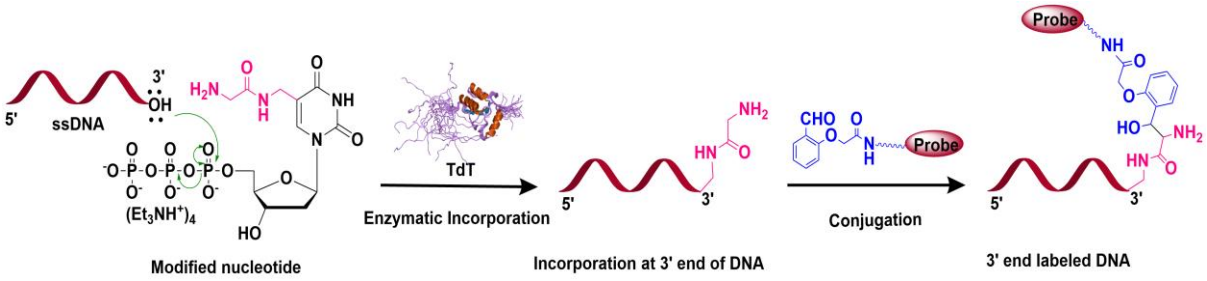


Figure 6. Enzymatic incorporation of modified nucleotide and bioconjugation of the reactive benzaldehyde containing a probe.

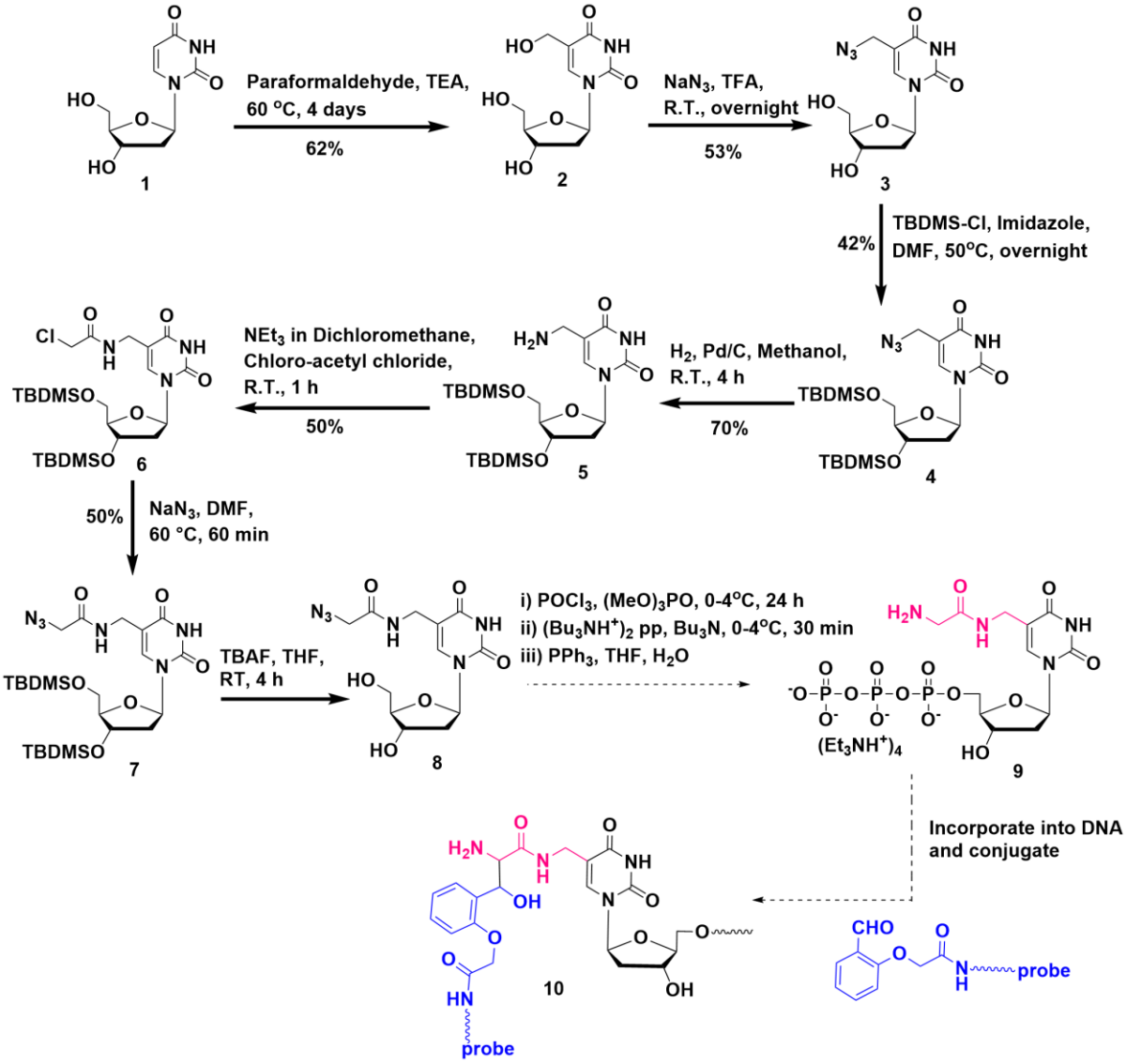
There has been recent work done by Srivatsan and coworkers, which suggests C5-modified uridine triphosphate (UTP) analogs are efficiently incorporated by enzyme, terminal uridylyltransferase (TUTase), site-specifically at the 3' end of RNA.[19] On a similar line in this project, we envision using template-independent terminal deoxynucleotidyl transferase (TdT) to achieve 3' end-labeling of single-stranded DNA. The exact position of modification is dictated by the information gathered from the structural consideration of the TdT enzyme. The modification introduced at the C-

5 position of pyrimidine is well tolerated by the transferase because of the void space around this position in the active site.[20]

Based on the considerations, we have planned the synthetic scheme to synthesize the glycine-modified analog (Scheme 1). Commercially available 2'-deoxyuridine (**1**) was used as the starting material. The primary precursor is an azide containing analog (**8**), which will be reduced by Staudinger reduction to form a reactive glycine handle. Glycine modified analog can be converted into triphosphate and can be incorporated into DNA ONs by using TdT or even by DNA polymerases. An advantage of this synthetic scheme is that one of the intermediates is an azide-modified nucleotide (**8**) which can be used for attaching cognate reactive partner having alkyne moiety using click chemistry. One scheme can generate two different types of nucleotides and can be used for clickable reactions as well as can form reversible aminoalcohol.

2. Results and Discussion

5-[(2-(azidoacetyl) amino)-N-methyl)]-2'-deoxyuridine (**8**) is the azide-containing analog which is the primary precursor of the required glycine-modified nucleotide. For the synthesis of **8,** we designed an appropriate synthetic scheme (Scheme 1).

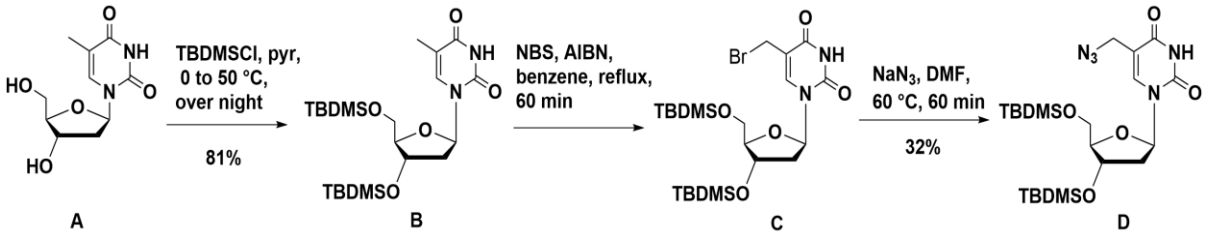


Scheme 1. Synthetic scheme for nucleotide synthesis and its bioconjugation with reactive partner.

Commercially available 2'-Deoxyuridine (**1**) and paraformaldehyde reacted to form **2** in 62% yield after four days[21]. The reaction between **1** and sodium azide in TFA as the solvent led to the conversion of hydroxymethyl group to azidomethyl with a 53% yield[21]. As mentioned in the referred paper, TFA was first neutralized using saturated sodium bicarbonate to pH=7 and then purified by column chromatography. But, TFA

is a strong acid (pK_a= 0.23), and sodium bicarbonate is a weak base; therefore, a lot of salt was precipitated after neutralization. The vast amount of salt caused difficulty in preparing the silica slurry for column chromatography and led to product loss. Next time to avoid this issue, I reduced the solvent under vacuum and co-evaporated multiple times using dichloromethane to remove the TFA present and purified using silica column chromatography without neutralization. But, even the slight presence of highly polar TFA led to a change in polarity of the solvent system and no separation was possible. Therefore, after reducing the solvent in vacuo, I neutralized the reaction mixture using saturated sodium bicarbonate unlike what is mentioned in the paper[21]. As a result, less salt was precipitated and was able to get proper separation in column chromatography. Using *tert*-butyldimethylsilyl chloride and imidazole as a base, free 3'–OH and 5'–OH of **3** were protected to yield **4.**

For the synthesis of **4** initially, a different route was followed[22] (Scheme 2). The starting material, 5-(methyl)-2'-deoxyuridine (**A**), was reacted with *tert*-butyldimethylsilyl chloride to obtain 3',5'-OH di-protected product **B**. 3',5'-di-O-(*tert*-butyldimethylsilyl)-2'-deoxyuridine (**B**) underwent bromination reaction to form 3',5'-di-O-(*tert*-butyldimethylsilyl)-5-(bromomethyl)-2'-deoxyuridine (**C**). The bromo-substituted product had no R_f difference from starting material **B**, so there was difficulty in estimating the extent of the reaction, and purification of the product formed was not possible. As bromination proceeds by radical mechanism, it led to the formation of many side products. 3',5'-di-O-(*tert*-butyldimethylsilyl)-5-(bromomethyl)-2'-deoxyuridine (**C**) was reacted with sodium azide to form 3',5'-di-O-(*tert*-butyldimethylsilyl)-5-(azidomethyl)-2'-deoxyuridine (**D**) by nucleophilic substitution reaction. The side products formed in bromination steps interfered with the substitution reaction leading to a low yield of the desired product. So, the initial few steps of the synthetic route were modified as mentioned in the synthetic scheme (Scheme1).



Scheme 2. Alternate route to synthesis 3',5'-di-O-(*tert*-butyldimethylsilyl)-5-(azidomethyl)-2'-deoxyuridine.

Azidomethyl moiety in **4** was reduced to aminomethyl using a catalytic amount of Pd on carbon to yield **5** in the hydrogen atmosphere. Amine group in **5** was coupled with chloroacetyl chloride by forming an amide bond with 50% yield to result in **6**. In the reaction mixture of **6**, several side products of low polarity were formed. The R_f difference between the desired product and side products was 0.2; still, proper separation in silica column chromatography was not possible because of differences in solubility and polarity of the side products in hexane: ethyl acetate solvent system. The side products had low solubility at the low polarity of the solvent system, and solubility was increasing with increasing the polarity. Therefore, on increasing the polarity, product and side products were eluting together. The solubility of the side products was observed to be more in dichloromethane by observing the distances run by the same spots in TLC in methanol/ dichrolomethane= 1/10 solvent system and ethyl acetate/ hexane= 1/2 solvent system. Later purification was tried using a three-solvent system, i.e., hexane: dichloromethane: ethyl acetate system, but the still complete separation was not achieved. Finally, the reaction mixture was dissolved in small amount of DCM and addition of n-pentane precipitated the desired product.

The chloride group in **6** was substituted with a nucleophile azide to result in **7** with 50% yield. **7** was stirred with tetra-n-butylammonium fluoride (TBAF) at room temperature resulting in TBDMS deprotection of both 3'–OH and 5'–OH of the deoxyribose sugar, giving **8**. The UV inactive tetra-n-butylammonium ions were removed by using Amberlite IR-120 (plus) Ion-exchange resin Na^+ form.

The formation of all the products was confirmed by HRMS and was characterized by ^1H, ^{13}C, DEPT NMR. For azide-containing products, IR data was also taken. All the spectra have been given in the annexure.

3. Conclusion and future plan

Synthesis of critical precursor molecule, 5-[(2-(azidoacetyl) amino)-N-methyl)]-2'-deoxyuridine (**8**) for the synthesis of nucleotide is done, and it is confirmed by HRMS data. Complete purification is to be done. The formation of desired compounds in every step was confirmed by HRMS and was characterized by ^1H and ^{13}C NMR.

Azide-containing compounds were also characterized using IR. NMR and IR spectra of all the compounds are given in the annexure.

Further steps include synthesis of glycine-modified 2'-deoxyuridine nucleotide analog. The modified nucleotide will be incorporated into a DNA oligonucleotide using template-independent terminal deoxynucleotidyl transferase. The incorporation can be controlled by varying the concentration of the enzyme or the nucleotide. After establishing bioconjugation with reactive counterpart, purification and characterization of the labeled DNA product will be done.

4. Experimental Section

4.1. Materials

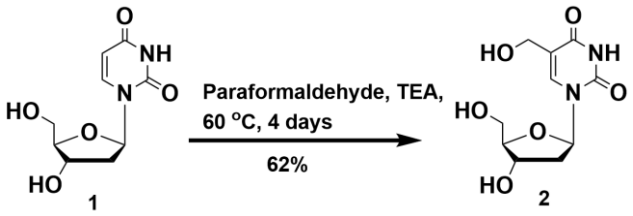
2'-Deoxyuridine was purchased from Alfa Aesar. Paraformaldehyde, sodium azide, imidazole, *tert*-Butyldimethylsilyl chloride, Palladium On carbon (10%), and Amberlite IR-120(plus) Ion-exchange resin Na^+ form was obtained from Sigma Aldrich. TBAF 1M solution in THF was purchased from Enamine. Chloroacetyl chloride was purchased from Spectrochem. TLC was used to monitor the progress of all reactions using precoated Merck silica plates (F254, 0.25 mm thickness). The TLC plates were examined under UV light or by staining with ninhydrin spray. All solvents were evaporated under reduced pressure on Heidolph rotatory evaporator below 40 ºC unless otherwise specified. Silica gel (100-200) mesh was used for column chromatography using the force flow of the solvent indicated.

4.2. Instrumentation

NMR was recorded in Bruker Avance III HD Ascend 400 MHz using TMS as an internal standard. Chemical shifts (δ) were expressed in ppm units downfield to TMS and coupling constants (*J*) in Hz and processed in Mnova NMR software from Mestrelab Research. Mass measurements were recorded using HRMS Water Synapt G2 high-definition mass spectrometer. IR spectra were recorded on Perkin-Elmer 1600 FT-IR spectrometers and were measured in cm^{-1}.

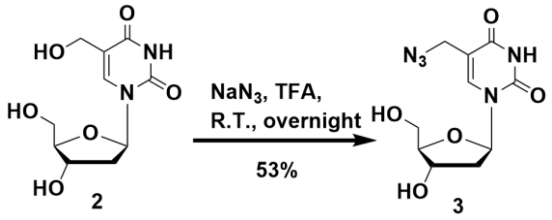
4.3. Synthesis

4.3.1. Synthesis of 5-Hydroxymethyl-2'-deoxyuridine (2):[21]



Commercially available 2'-Deoxyuridine (**1**) (10 g, 43.8 mmol) and paraformaldehyde (5.9 g, 197.2 mmol, 4.5 equiv.) were dissolved in 160 mL of 0.5 mol/L TEA aqueous solution. The reaction mixture was stirred at 60˚C for four days. Each day, the mixture was supplemented with paraformaldehyde (8.5 g, 254.8 mmol, 6.5 equiv.), TEA (2 mL), and water (20 mL). After completion of the reaction, the residue was purified using column chromatography (MeOH/ DCM = 1/10, v/v) to get the final compound as white solid (7 g, 62%). ^{1}H NMR (400 MHz, DMSO-d_6): δ (ppm): 11.31 (s, 1H), 7.72 (s,1H), 6.18 (t, *J* = 6.9 Hz, 1H), 5.25 (d, *J* = 4.2Hz, 1H), 4.93 (dt, *J* = 11.0, 5.3 Hz, 2H), 4.20 (q, *J* = 4.6 Hz, 1H), 4.13 (d, *J* = 4.7Hz, 2H), 3.77 (q, *J* = 4.0 Hz, 1H), 3.60–3.49 (m, 2H), 2.11–2.04 (m, 2H); ^{13}C NMR (101 MHz, DMSO-d_6) δ (ppm): 162.7, 150.4, 136.8, 114.3, 87.3, 83.9, 70.6, 61.5, 56, 39.5; HRMS (ESI): m/z for $C_{10}H_{15}N_2O_6$ $[M+H]^+$ calculated= 259.0930; found= 259.0923. Analytical data matches with the literature report.[21]

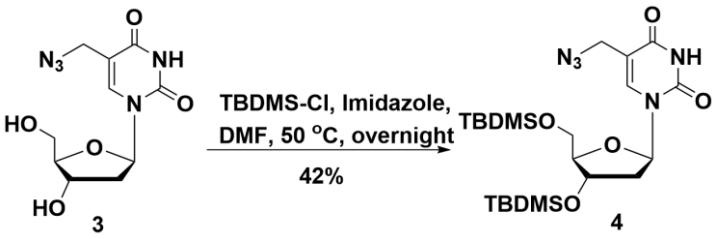
4.3.2. Synthesis of 5-Azidomethyl-2'-deoxyuridine (3):[21]



2 (7 g, 34.8 mmol) was dissolved in 460 mL of TFA, then NaN_3 (4.5 g, 69.7 mmol, 2 equiv.) was added slowly into it. At room temperature, the mixture was stirred overnight. The mixture was then concentrated under vacuum and then neutralized to pH=7 by using saturated $NaHCO_3$ solution. The crude was purified using column chromatography (MeOH/ DCM= 1/10, v/v) to get **3** as white solid (4 g, 53%). ^{1}H NMR (400 MHz, MeOH-d_4) δ (ppm): 8.11 (s, 1H), 6.23 (t, *J*= 6.6 Hz, 1H), 4.36 (dt, *J*= 6.5, 3.5 Hz, 1H), 4.04 (s, 2H), 3.89 (t, *J*= 3.3 Hz, 1H), 3.73 (dd, *J*= 16.6, 3.3 Hz, 2H), 2.29–2.14 (m, 2H), 1.26 (t, *J* = 7.3 Hz, 1H). ^{13}C NMR (101 MHz, MeOH-d_4) δ (ppm):

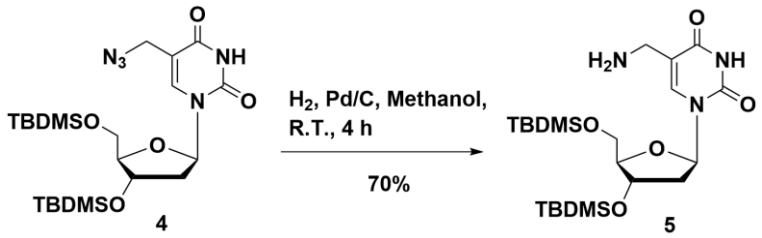
163., 119.5, 116.6, 110.4, 89, 86.7, 72.1, 62.7, 41.4. HRMS (ESI): m/z for $C_{10}H_{14}N_5O_5$ $[M+H]^+$ calculated= 284.0995; found = 284.1007. IR wavenumber (cm^{-1}): 3384, 2924, 2351, 2115, 1674, 1445, 1207, 1135, 854, 801, 712. Analytical data matches with the literature report.[21]

4.3.3. Synthesis of 3',5'-di-O-(*tert*-butyldimethylsilyl)-5-(azidomethyl)-2'-deoxyuridine (4):[22]



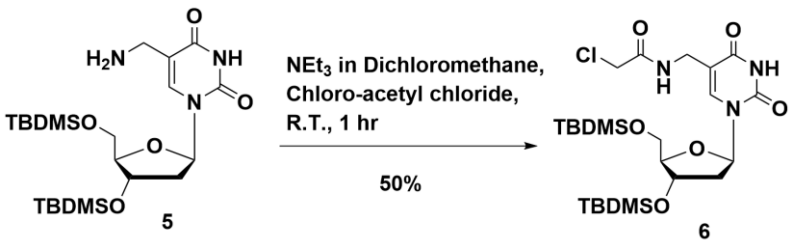
3 (4 g, 14.1 mmol) and imidazole (3.8 g, 56.5 mmol, 4 equiv.) were dissolved in 45 mL of dimethylformamide and chilled to 0 °C. To it we added TBDMS-Cl (4.7 g, 31 mmol, 2.2 equiv.) and the mixture was stirred at 50 °C overnight. The solution was diluted with ethyl acetate (50 mL) and washed with saturated NH_4Cl solution (3x 30 mL) and brine (2x 30 mL). The organic layer was dried using Na_2SO_4 and was concentrated. The crude was purified by column chromatography (EA/ hexane= 3/20, v/v) and compound **4** (3 g, 42%) was obtained as a white solid. ^{1}H NMR (400 MHz, Chloroform-*d*) δ (ppm): 9.08 (s, 1H), 7.71 (s, 1H), 6.30 (dd, *J* = 7.7, 5.8 Hz, 1H), 4.41–4.37 (m, 1H), 4.17–4.05 (m, 2H), 3.96 (d, *J* = 2.6 Hz, 1H), 3.86 (d, *J* = 8.6 Hz, 1H), 3.76 (d, *J* = 11.4 Hz, 1H), 2.34–2.27 (m, 1H), 2.03–1.96 (m, 1H), 0.92 (s, 9H), 0.89 (s, 9H), 0.11 (s, 6H), 0.08 (d, *J*= 3.1 Hz, 6H). ^{13}C NMR (101 MHz, Chloroform-*d*) δ (ppm): 162.8, 150, 138.7, 109.6, 88.2, 85.5, 72.3, 63.1, 47.3, 41.8, 26, 25.9, 18.6, 18.1, -4.5, -4.7, -5.3, -5.3. HRMS (ESI): m/z for $C_{22}H_{42}N_5O_5Si_2$ $[M+H]^+$ calculated= 512.2724; found= 512.2722.

4.3.4. Synthesis of 3',5'-di-O-(*tert*-butyldimethylsilyl)-5-(aminomethyl)-2'-deoxyuridine (5):[23]



4 (1.5 g, 2.9 mmol) was dissolved in methanol (24 mL) and to it 0.2 g Pd/C (5%) catalyst and 4 drops of acetic acid was added. At room temperature, the reaction mixture was stirred under hydrogen atmosphere for 4 hours. The catalyst was filtered through a celite bed after completion of the reaction and the filtrate was concentrated. The crude was then purified by column chromatography (MeOH/ DCM= 1/20, v/v) to obtain the desired product **5** as a white solid (1 g, 70%). ^{1}H NMR (400 MHz, Chloroform-*d*) δ (ppm): 7.58 (s, 1H), 6.32 (d, *J* = 5.8 Hz, 1H), 4.41–4.37 (m, 1H), 3.93 (s, 1H), 3.84 (d, *J* = 11.3 Hz, 1H), 3.75 (d, *J* = 11.3 Hz, 1H), 3.54 (s, 2H), 2.29–2.23 (m, 1H), 2.04–1.96 (m, 1H), 0.90 (d, *J* = 11.4 Hz, 18H), 0.09 (d, *J* = 10.0 Hz, 12H). ^{13}C NMR (101 MHz, Chloroform-*d*) δ (ppm): 163.6, 150.3, 136.2, 115.9, 88, 85.1, 72.4, 41.4, 40, 26, 25.9, 18.548, 18.1, -4.5, -4.7, -5.2, -5.3. HRMS (ESI): m/z for $C_{22}H_{44}N_3O_5Si_2$ $[M+H]^+$ calculated= 486.2819; found = 486.2822. The characterization data for **5** were in good agreement with those published in the literature.[23]

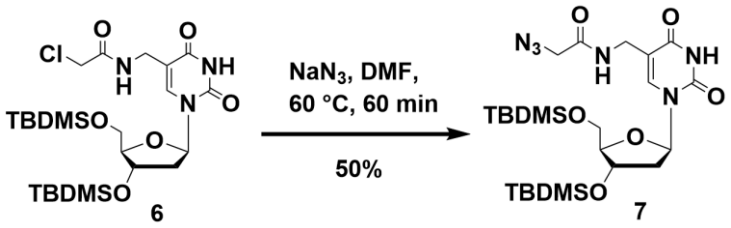
4.3.5. Synthesis of 3',5'-di-O-(*tert*-butyldimethylsilyl)-5-[(2-(chloroacetyl)amino)-N-methyl]-2'-deoxyuridine (6):[24]



5 (1.4 g, 2.9 mmol) and TEA (0.8 mL, 5.8 mmol, 2 equiv.) was dissolved in DCM (12 mL) and was chilled to 0 °C. To it solution of chloroacetylchloride (0.4 ml, 5.7 mmol, 2 equiv.) in DCM (10 mL) was added dropwise (5-10 min addition). The reaction mixture was stirred for 1 h at room temperature and then washed with water (3x 15

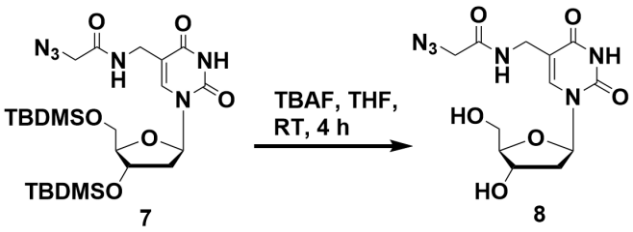
mL) and saturated $NaHCO_3$ solution (3x 15 mL). The organic layer was dried using Na_2SO_4 and was concentrated using vacuum. After purification using column chromatography (EA/ hexane= 1/5, v/v) and precipitation of the product in n-pentane, compound **6** (800 mg, 50%) was obtained as a white solid. ^{1}H NMR (400 MHz, Chloroform-*d*) δ (ppm): 9.12 (s, 1H), 7.74 (s, 1H), 7.40 (s, 1H), 6.27 (dd, *J* = 7.9, 5.7 Hz, 1H), 4.40 (dt, *J* = 5.5, 2.5 Hz, 1H), 4.15 (d, *J* = 5.9 Hz, 2H), 4.00 (s, 2H), 3.95 (q, *J* = 3.1 Hz, 1H), 3.79 (qd, *J* = 11.3, 3.4 Hz, 2H), 2.28 (ddd, *J* = 13.2, 5.8, 2.5 Hz, 1H), 2.02 (ddd, *J* = 13.5, 8.0, 5.9 Hz, 1H), 0.90 (d, *J* = 11.6 Hz, 18H), 0.10 (dd, *J* = 17.1, 2.2 Hz, 12H). ^{13}C NMR (101 MHz, Chloroform-*d*) δ (ppm): 166.1, 163.7, 150, 139, 110.4, 88.3, 85.6, 72.6, 63.3, 42.6, 41.4, 37.2, 26.1, 25.9, 18.6, 18.1, -4.5, -4.7, -5.2, -5.3. HRMS (ESI): m/z for $C_{24}H_{45}ClN_3O_6Si_2$ $[M+H]^+$ calculated= 562.2535; found = 562.2521.

4.3.6. Synthesis of 3',5'-di-O-(*tert*-butyldimethylsilyl)-5-[(2-(azidoacetyl)amino)-N-methyl)]-2'-deoxyuridine (7):[22]



6 (800 mg, 1.4 mmol) was put in 20 mL DMF, and NaN_3 (179.6 mg, 2.9 mmol, 2 equiv.) was added. The mixture was stirred for 60 min at 60 °C. Then it was diluted using EA (30 mL) and washed with water (2x30 mL) and brine (2x30 mL). The organic layer was dried using Na_2SO_4 and concentrated to give product **7** as white solid (400 mg, 50%). ^{1}H NMR (400 MHz, Chloroform-*d*) δ (ppm): 9.21 (s, 1H), 7.73 (s, 1H), 7.16 (s, 1H), 6.29–6.24 (m, 1H), 4.42–4.38 (m, 1H), 4.13 (d, *J* = 5.9 Hz, 2H), 3.95 (d, *J* = 2.6 Hz, 1H), 3.91 (s, 2H), 3.79 (dd, *J* = 9.7, 3.4 Hz, 2H), 2.28 (dd, *J* = 9.1, 4.1 Hz, 1H), 2.02 (t, *J* = 5.6 Hz, 1H), 0.90 (d, *J* = 11.3 Hz, 18H), 0.10 (d, *J* = 15.2 Hz, 12H). ^{13}C NMR (101 MHz, Chloroform-*d*) δ (ppm): 166.8, 163.808, 150.1, 139.1, 110.6, 88.3, 85.6, 72.6, 63.3, 52.7, 41.4, 36.8, 26.1, 25.9, 18.5, 18.1, -4.5, -4.7, -5.2, -5.3. HRMS (ESI): m/z for $C_{24}H_{45}N_6O_6Si_2$ $[M+H]^+$ calculated= 569.2939; found = 569.2953. IR wavenumber (cm^{-1}): 3397, 2947, 2853, 2107, 1708, 1491, 1248, 1038, 834, 780.

4.3.7. Synthesis of 5-[(2-(azidoacetyl)amino)-N-methyl)]-2'-deoxyuridine (8)[22]:



To ice-cold (0°C) solution of **7** (400 mg, 0.7 mmol) in THF (9 mL), tetra-n-butylammonium fluoride (TBAF) 1M solution in THF (0.6 mL, 2.1 mmol, 3 equiv.) was added slowly in inert atmosphere. The solution was stirred overnight allowing the mixture to warm to room temperature. The crude product was purified by column chromatography (MeOH/ DCM= 1/10, v/v) to get the compound a white solid. The residual tetrabutylammonium salts were removed by using Amberlite IR-120(plus) Ion-exchange resin Na^+ form. ^1H NMR (400 MHz, Deuterium Oxide) δ (ppm): 7.83 (s, 1H), 6.30 (t, *J* = 6.7 Hz, 1H), 4.47 (t, *J* = 4.0 Hz, 1H), 4.15 (s, 2H), 4.07 (s, 1H), 4.05 (s, 2H), 3.89 – 3.74 (m, 2H), 2.47 – 2.31 (m, 2H). ^{13}C NMR (101 MHz, Deuterium Oxide) δ (ppm): 170.53, 164.95, 139.22, 110.72, 86.77, 85.41, 70.45, 61.14, 51.83, 38.87, 35.95. HRMS (ESI): m/z for $C_{12}H_{16}N_6O_6Na$ $[M+Na]^+$ calculated= 363.1131; found = 363.1032.

5. References

1) Asseline, U. Development and applications of fluorescent oligonucleotides. *Curr. Org. Chem.* **2006**, *10*, 491–518.
2) Holbrook, S. R. Structural principles from large RNAs. *Annu. Rev. Biophys*. **2008**, *37*, 445–464.
3) Piton, N.; Mu, Y.; Stock, G.; Prisner, T.F.; Schiemann, O.; Engels, J. W. Base-specific spin-labeling of RNA for structure determination. *Nucleic Acids Res.* **2007**, *35*, 3128–3143.
4) Bardaro, M. F. J.; Varani, G. Examining the relationship between RNA function and motion using nuclear magnetic resonance. *Wiley Interdiscip. Rev. RNA* **2012**, *3*, 122–132.
5) Wachowius, F.; Höbartner, C. Chemical RNA modifications for studies of RNA structure and dynamics. *ChemBioChem* **2010**, *11*, 469–480.
6) Tanpure, A. A.; Pawar, M. G.; Srivatsan, S. G. Fluorescent nucleoside analogs: probes for investigating nucleic acid structure and function. *Isr. J. Chem.* **2013**, *53*, 366–378.
7) George, J. T.; Srivatsan, S. G. Posttranscriptional chemical labelling of RNA by using bioorthogonal chemistry. *Methods* **2017,** *120*, 28–38.
8) Saxon, E.; Bertozzi, C. R. Cell surface engineering by a modified staudinger reaction. *Science* **2000,** *287*, 2007–2010.
9) Rossin, R.; Robillard, M. S. Pretargeted imaging using bioorthogonal chemistry in mice. *Curr. Opin. Chem. Biol.* **2014**, *21*, 161–169.
10) Rostovtsev, V. V.; Green, L. G; Fokin, V. V.; Sharpless, K. B. A stepwise Huisgen cycloaddition process: copper(I)-catalyzed regioselective "ligation" of azides and terminal alkynes. *Angew. Chem. Int. Ed.* **2002,** *41*, 2596–2599.
11) Tornoe, C. W.; Christensen, C.; Meldal, M. Peptidotriazoles on solid phase: [1,2,3]-triazoles by regiospecific copper(I)-catalyzed 1,3-dipolar cycloadditions of terminal alkynes to azides. *J. Org. Chem.* **2002**, *67*, 3057–3064.
12) Agard, N. J.; Prescher, J. A.; Bertozzi, C. R. A strain-promoted [3 + 2] azide−alkyne cycloaddition for covalent modification of biomolecules in living systems. *J. Am. Chem. Soc.* **2004,** *126*, 15046–15047.
13) Wu, H.; Devaraj, N. K. Inverse Electron-Demand Diels–Alder bioorthogonal reactions. *Top. Curr. Chem.* **2015**, *374*, 109–130.

14) Gramlich, P. M. E.; Wirges, C. T.; Manetto, A.; Carell, T. Postsynthetic DNA modification through the copper-catalyzed azide–alkyne cycloaddition reaction. *Angew. Chem. Int. Ed.* **2008**, *47*, 8350–8358.

15) Rao, H.; Sawant, A. A.; Tanpure, A. A.; Srivatsan, S. G. Posttranscriptional chemical functionalization of azide-modified oligoribonucleotides by bioorthogonal click and Staudinger reactions. *Chem. Commun*. **2012**, *48*, 498–500.

16) George, J. T.; Srivatsan, S. G. Vinyluridine as a versatile chemoselective handle for the posttranscriptional chemical functionalization of RNA. *Bioconjugate Chem*. **2017**, *28*, 1529–1536.

17) Walunj, M. B.; Tanpure, A. A.; Srivatsan, S. G. Posttranscriptional labeling by using Suzuki-Miyaura cross-coupling generates functional RNA probes. *Nucl. Acid. Res.* **2018**, *46*, e65.

18) Purushottam, L., Adusumalli, S., Singh, U.; Unnikrishnan, V. B.; Rawale, D. G.; Gujrati, M.; Mishra, R. K.; Rai, V. Single-site glycine-specific labeling of proteins. *Nat. Commun*. **2019**, *10*, 2539.

19) George, J.; Azhar, M.; Aich, M.; Sinha, D.; Ambi, U.; Maiti, S.; Chakraborty, D.; Srivatsan, S. G. Terminal Uridylyl Transferase mediated site-directed access to clickable chromatin employing CRISPR-dCas9. *J. Am. Chem. Soc.* **2020**, *142*, 32, 13954–13965.

20) Sarac, I.; Hollenstein, M. Terminal deoxynucleotidyl transferase in synthesis and modification of nucleic acids. *ChemBioChem* **2019,** *20*, 860–871.

21) Xu, X.; Yan, S.; Hu, J.; Guo, P.; Wei, L.; Weng, X.; Zhou, X. One step to get 5-azidomethyl-2′-deoxyuridine from 5-hydroxymethyl-2′-deoxyuridine and detection of it through click reaction. *Tetrahedron* **2013**, *69*, 9870–9874.

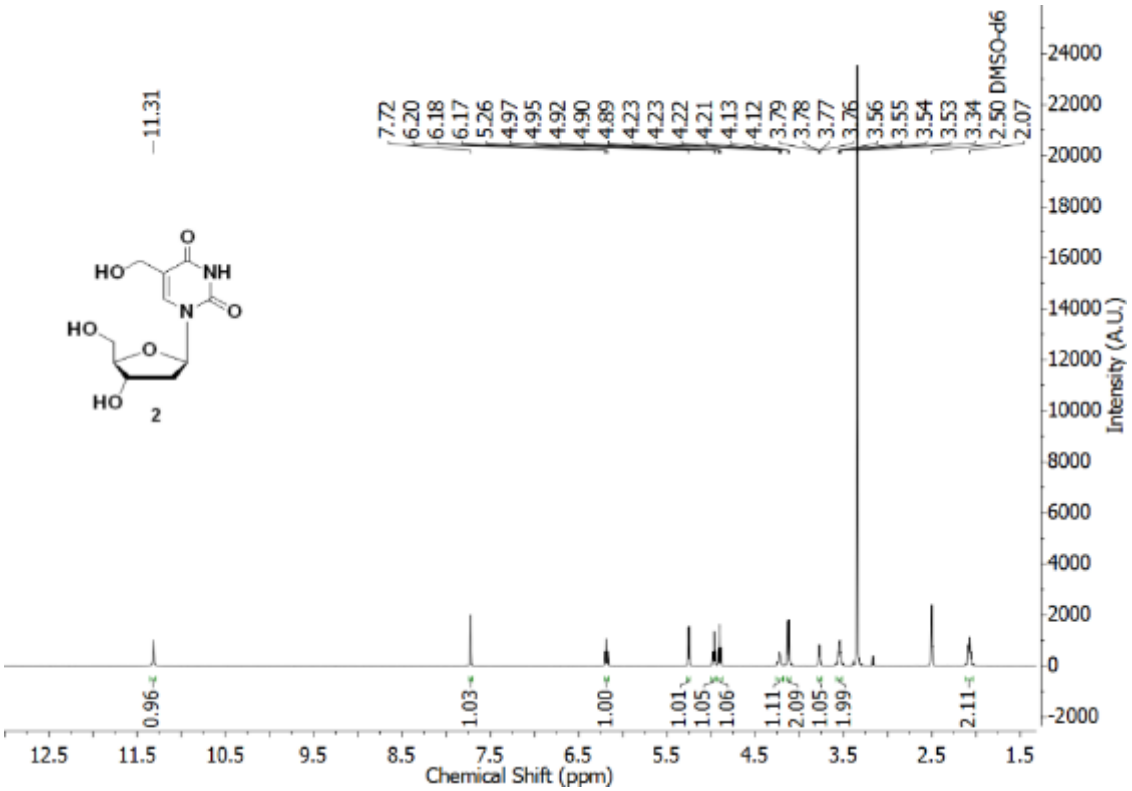
22) Neef, A. B.; Luedtke, N. W. An azide-modified nucleoside for metabolic labeling of DNA. *ChemBioChem* **2014**, *15*, 789–793.

23) Gubu, A.; Li, L.; Ning, Y.; Zhang, X.; Lee, S.; Feng, M.; Li, Q.; Lei, X.; Jo, K.; Tang, X. Bioorthogonal metabolic DNA labelling using vinyl thioether-modified thymidine and o-Quinolinone Quinone Methide. *Chem. Eur. J*. **2018**, *24*, 5895–5900.

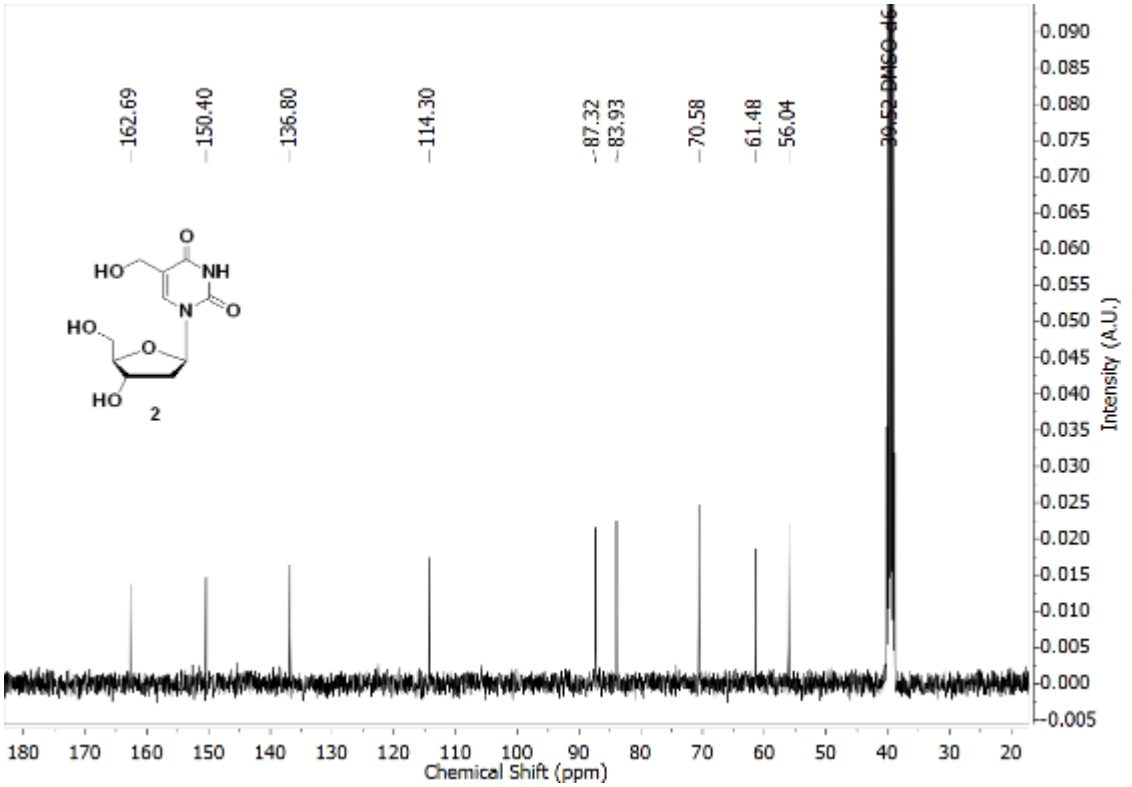
24) Kampf, A.; Pillar, C. J.; Woodford, W. J.; Mertes, M. P. Synthesis of 5-substituted 2'-deoxyuridines. *J. Med. Chem.* **1976**, *19*, 7, 909–915.

Appendix

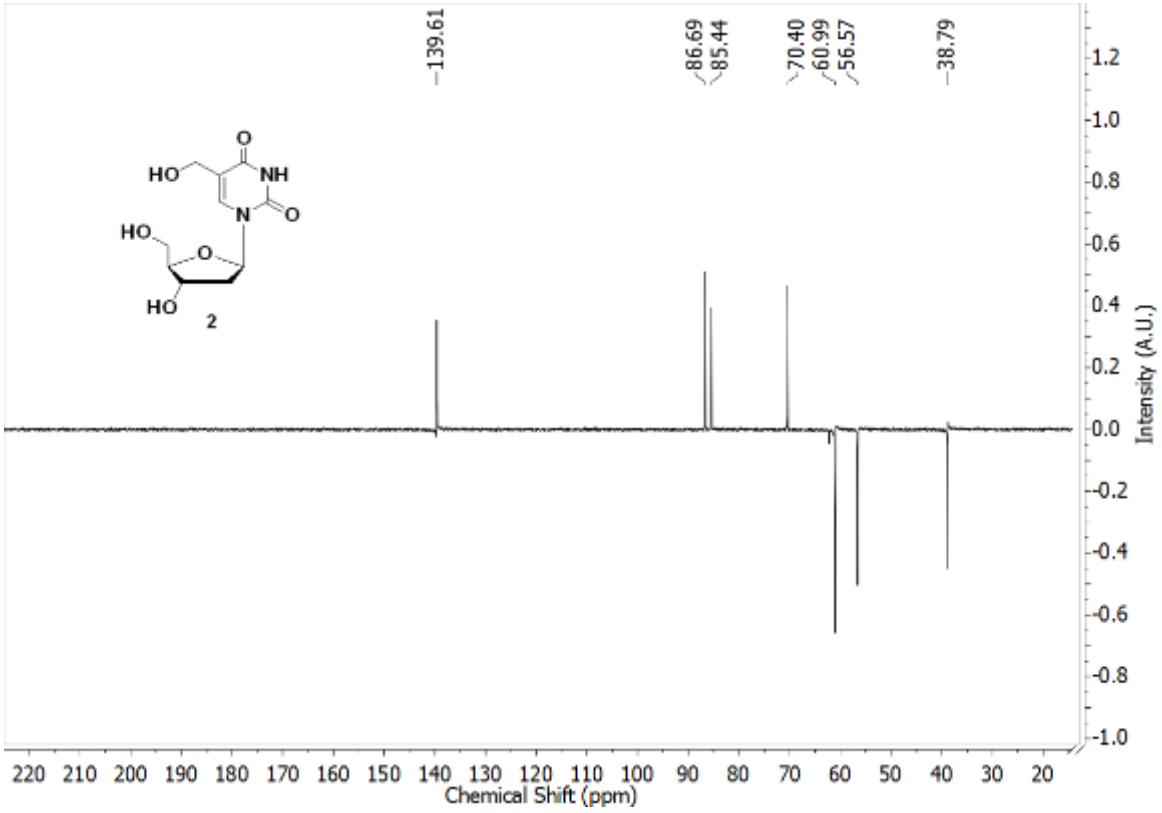
^{1}H NMR (400 MHz, DMSO-d_6) of **2**:



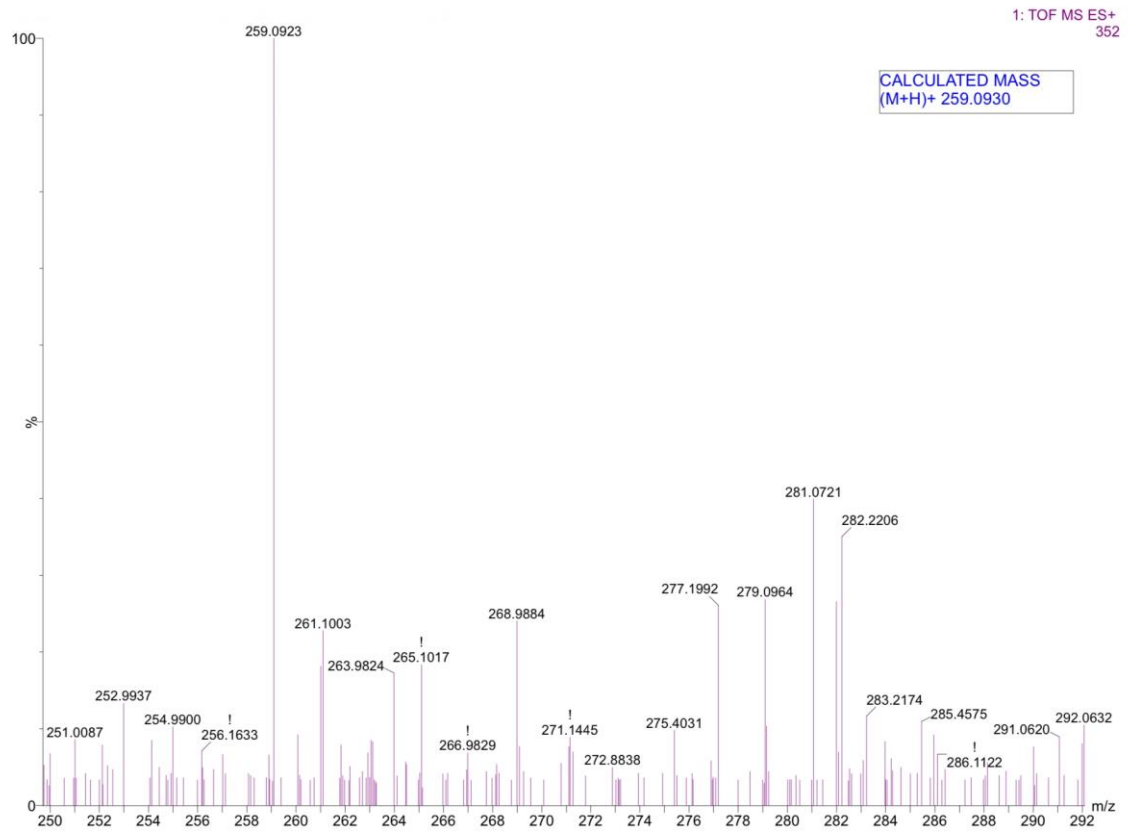
^{13}C NMR (101 MHz, DMSO-d_6) of **2**:



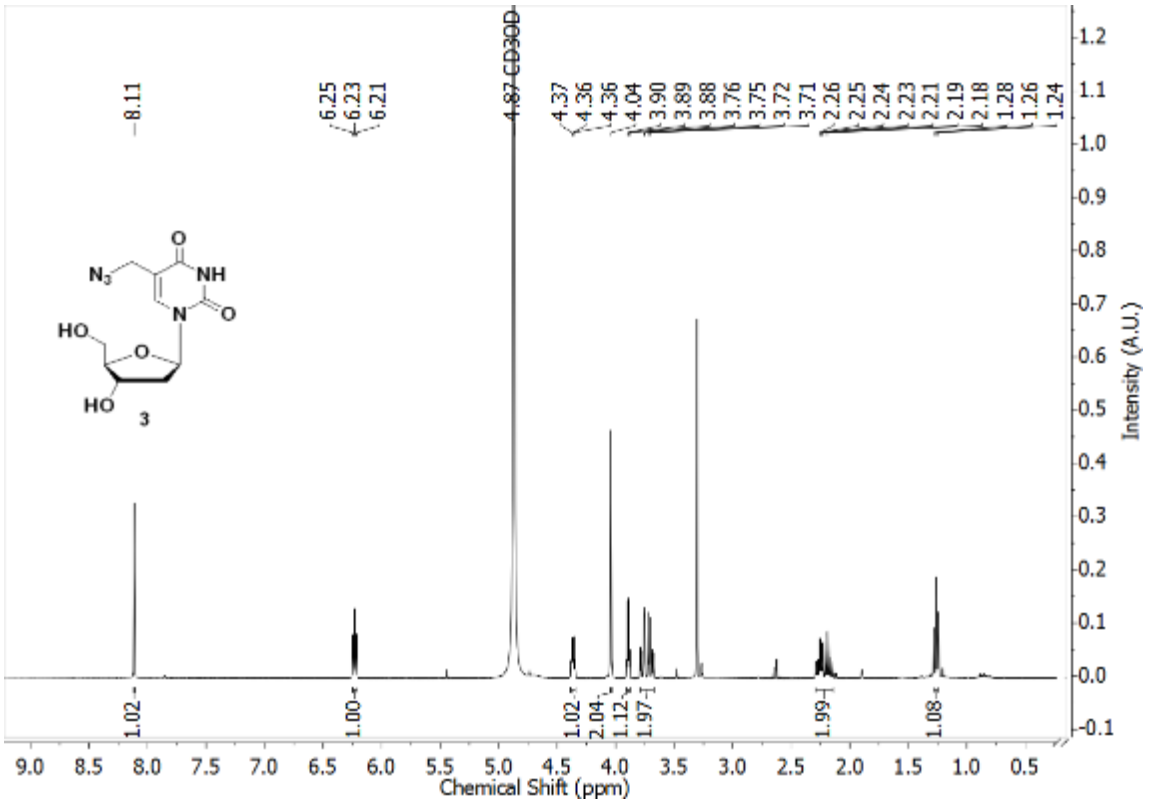
DEPT-135(101 MHz, DMSO-d_6) of **2**:



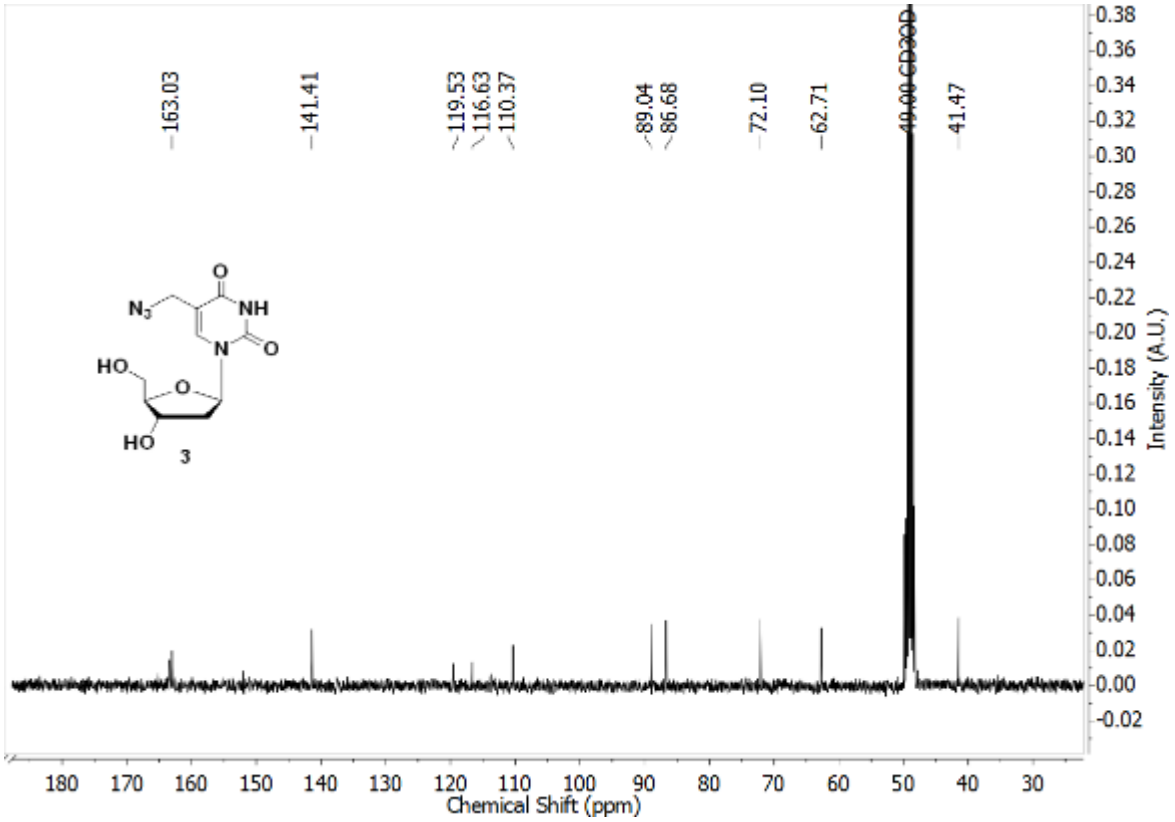
HRMS of **2**:



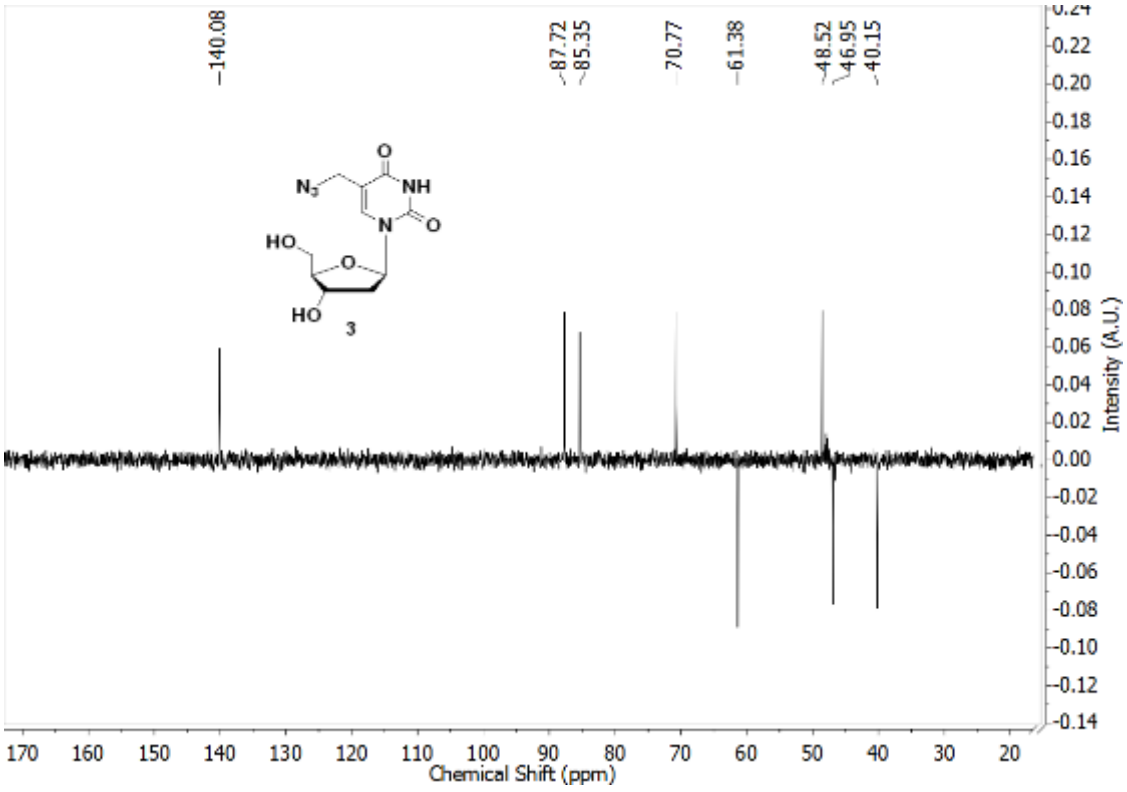
^{1}H NMR (400 MHz, Methanol-d_4) of **3**. Trace amount of methanol is present.



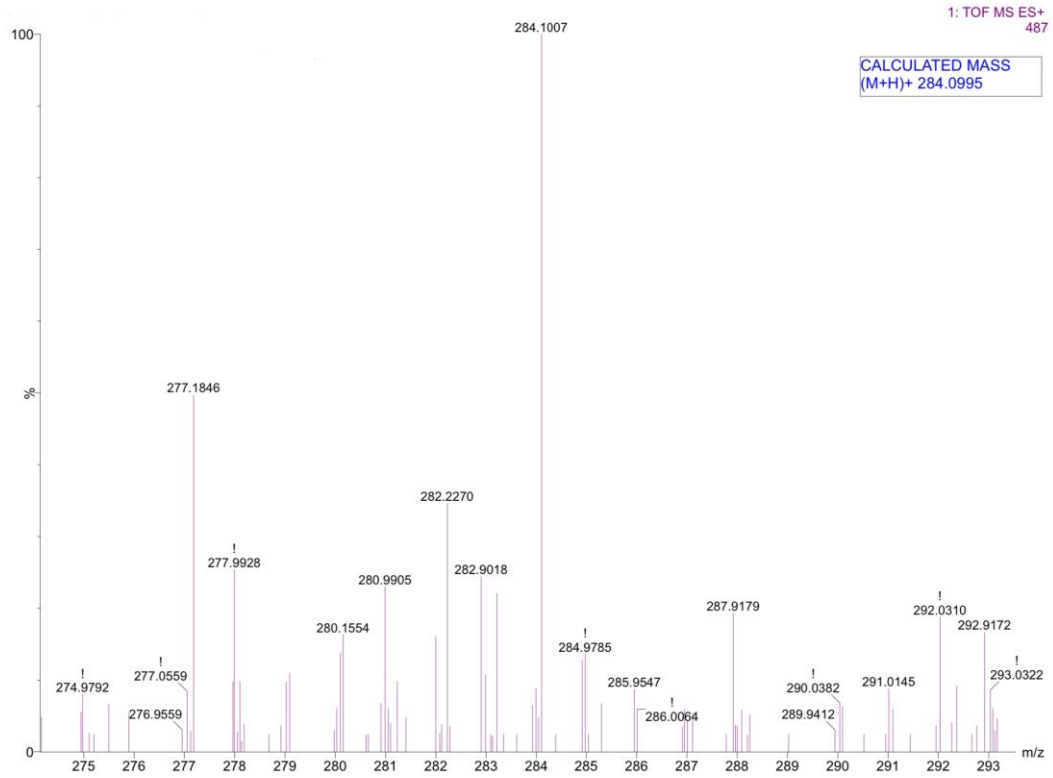
^{13}C NMR (101 MHz, Methanol-d_4) of **3**:



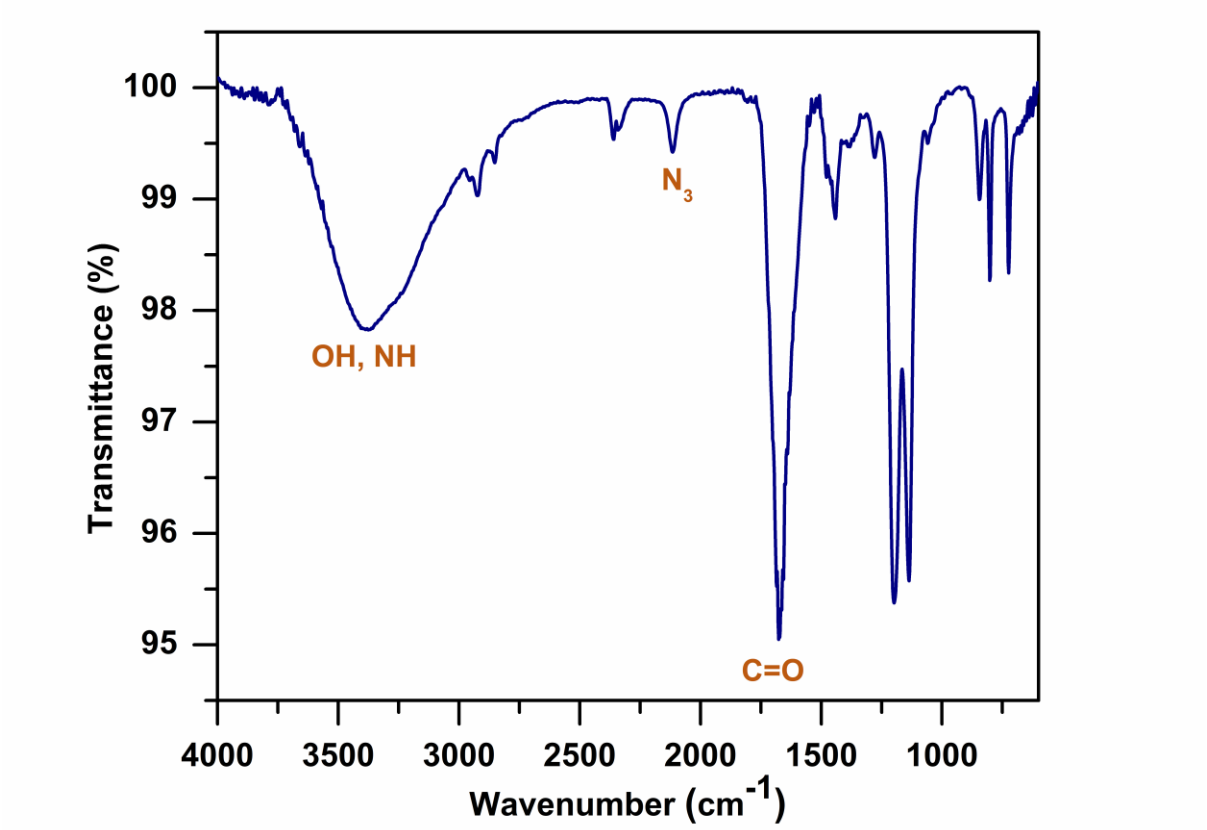
DEPT-135(101 MHz, Methanol- d_4) of **3**. Trace amount of methanol is present.



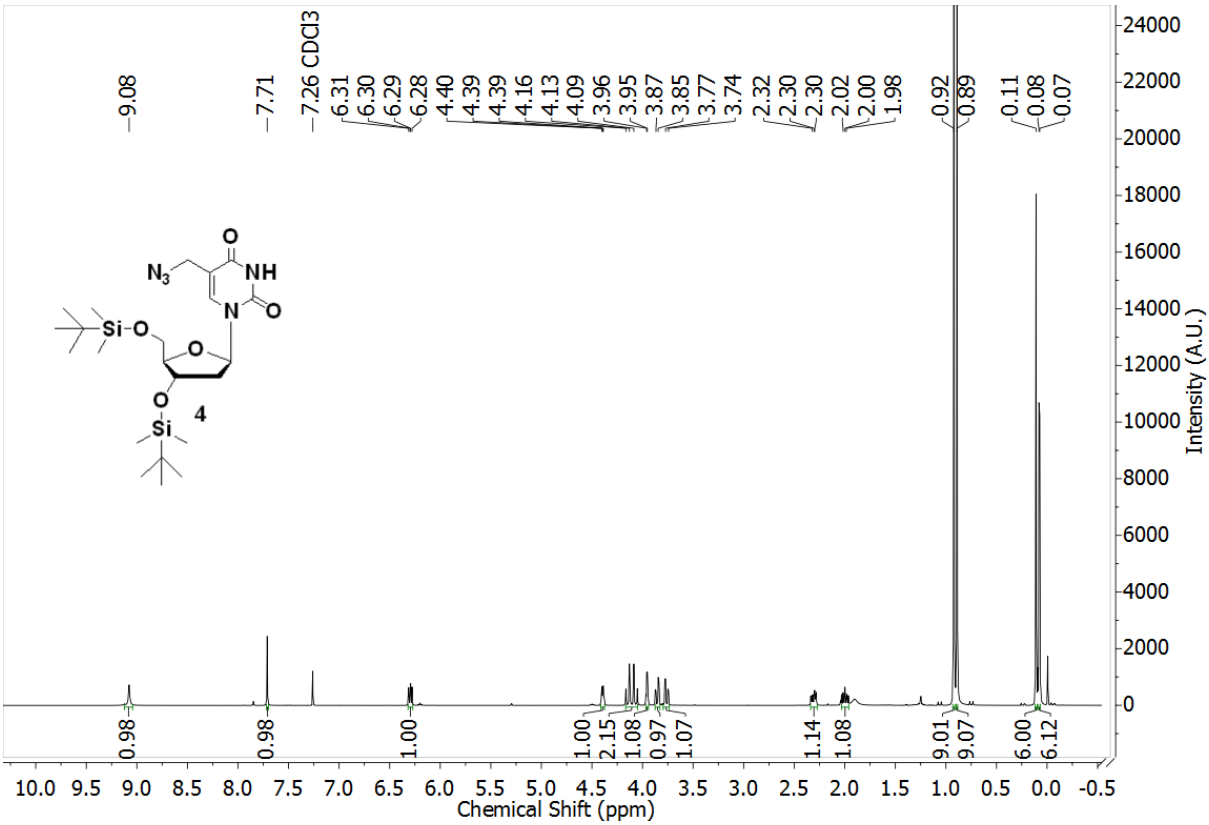
HRMS of **3**:



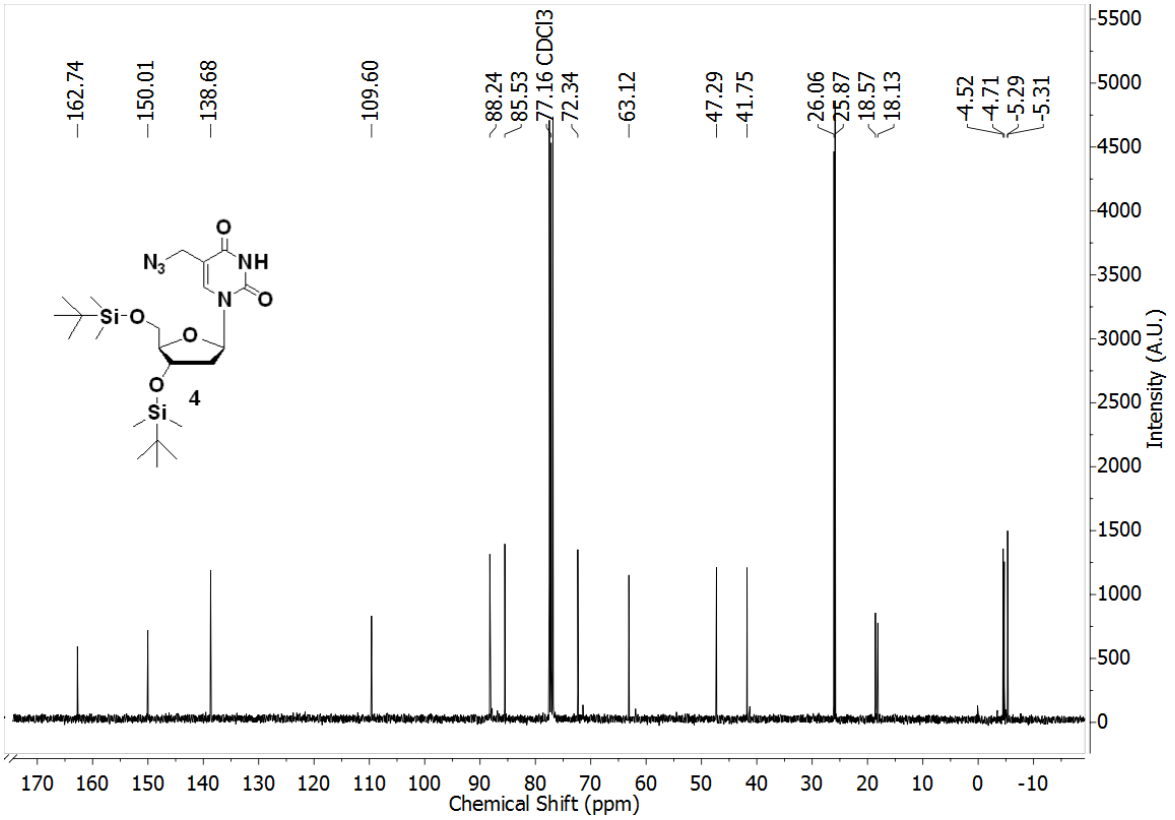
IR for **3**:



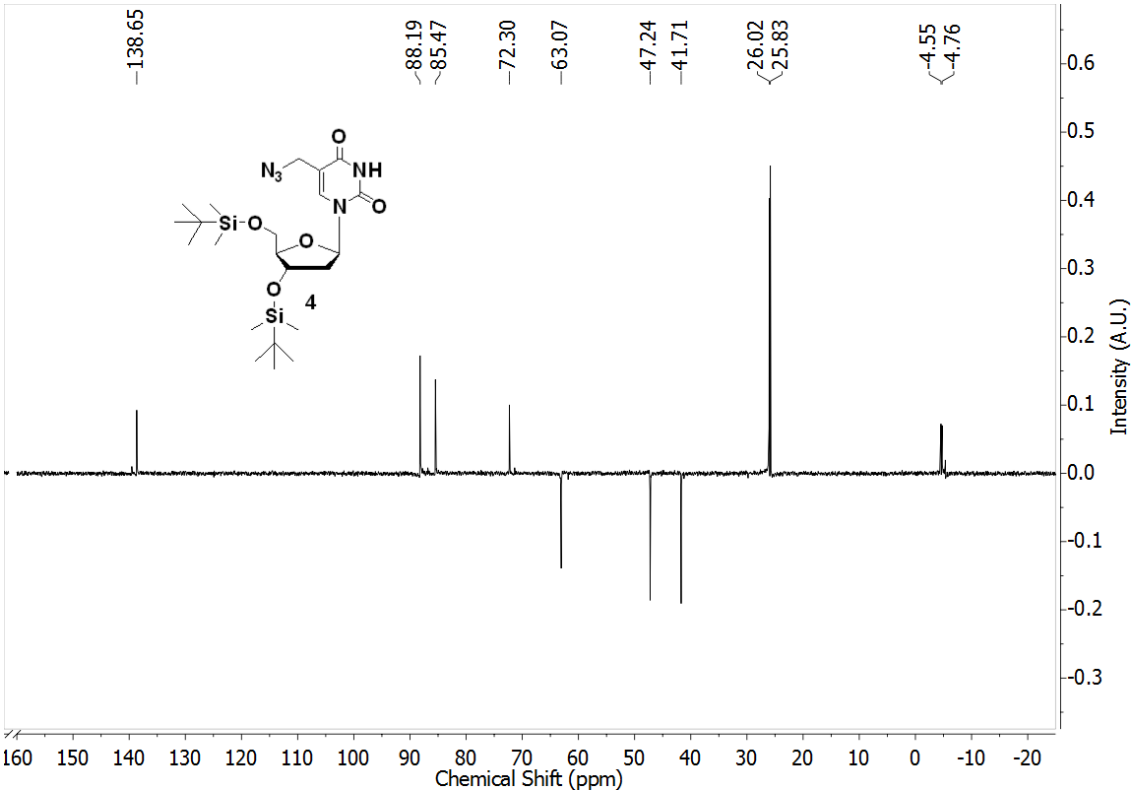
^{1}H NMR (400 MHz, Chloroform-*d*) of **4**:



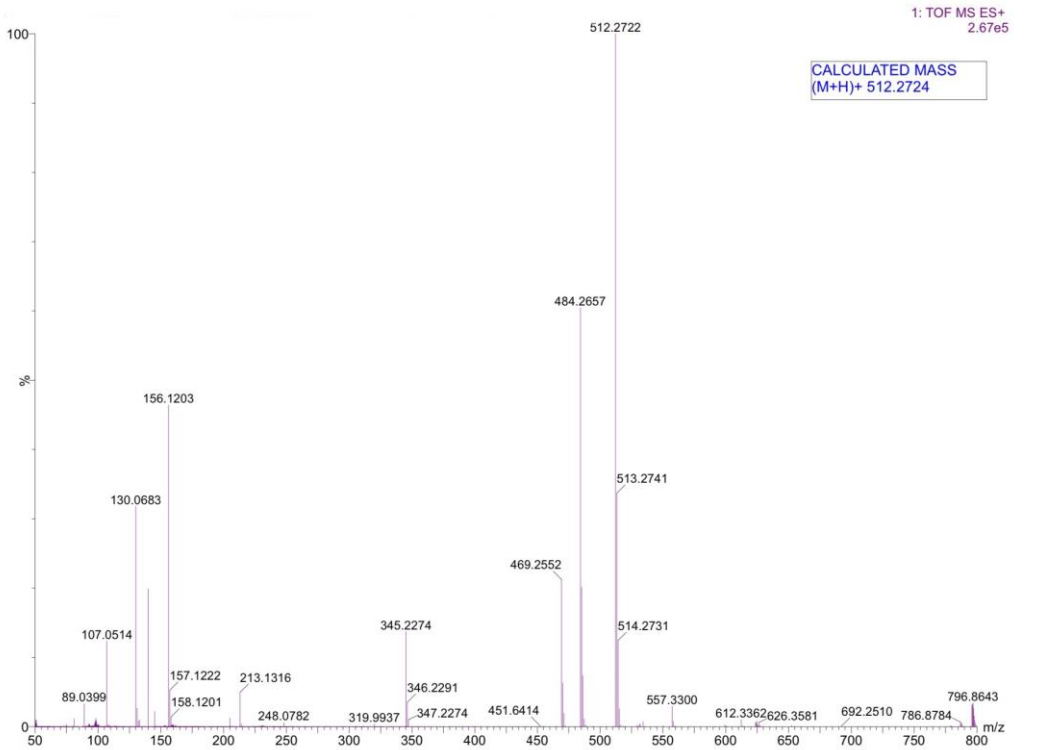
^{13}C NMR (101 MHz, Chloroform-*d*) of **4**:



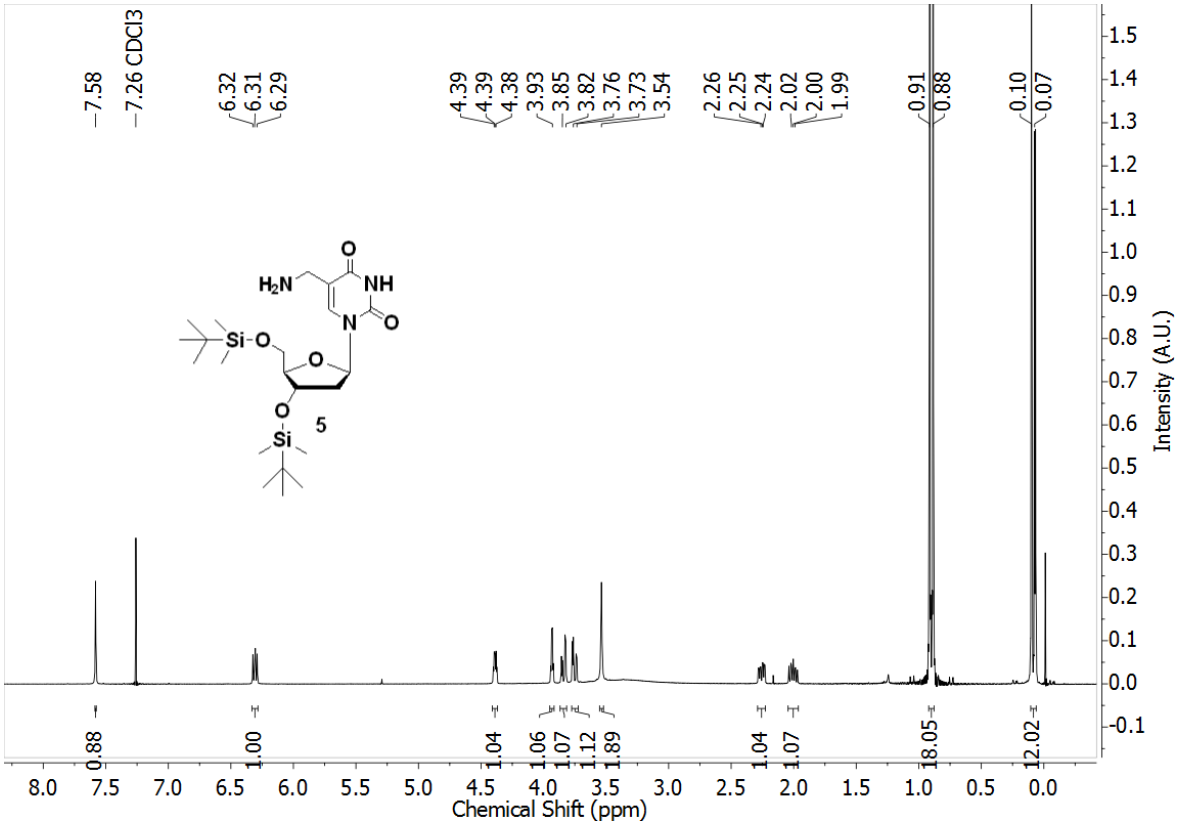
DEPT-135 (101 MHz, Chloroform-*d*) of **4**:



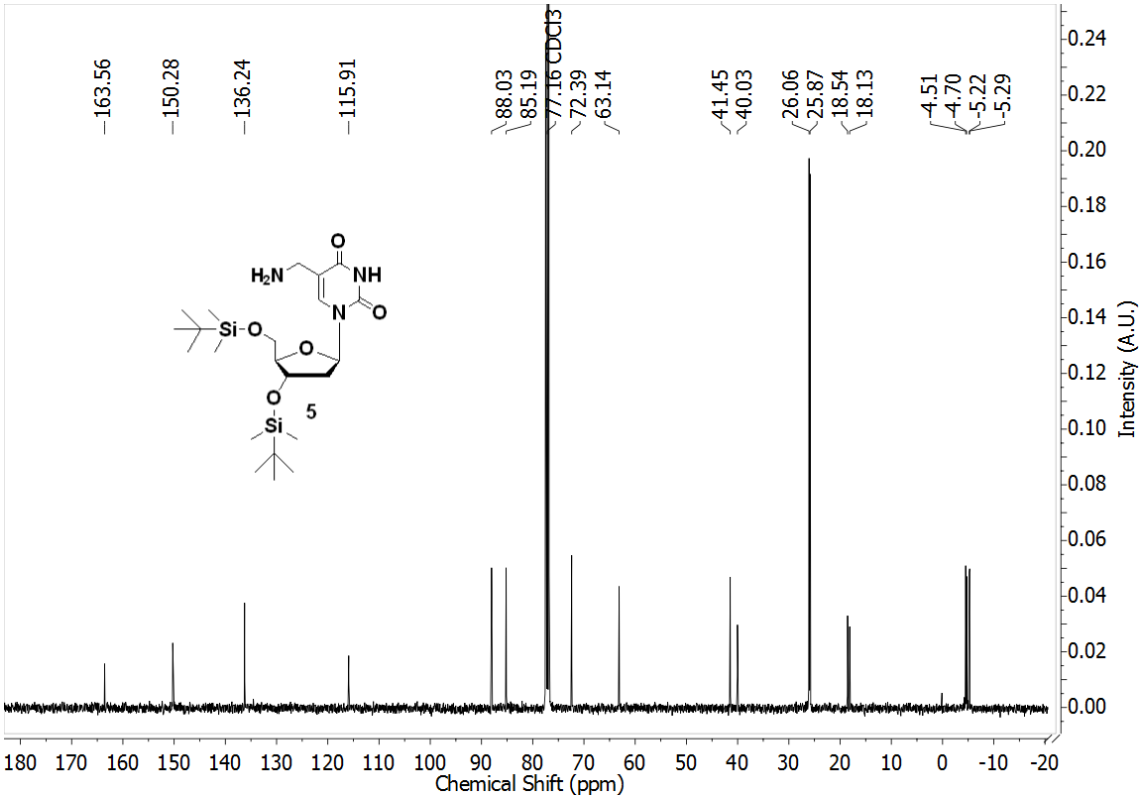
HRMS of **4**:



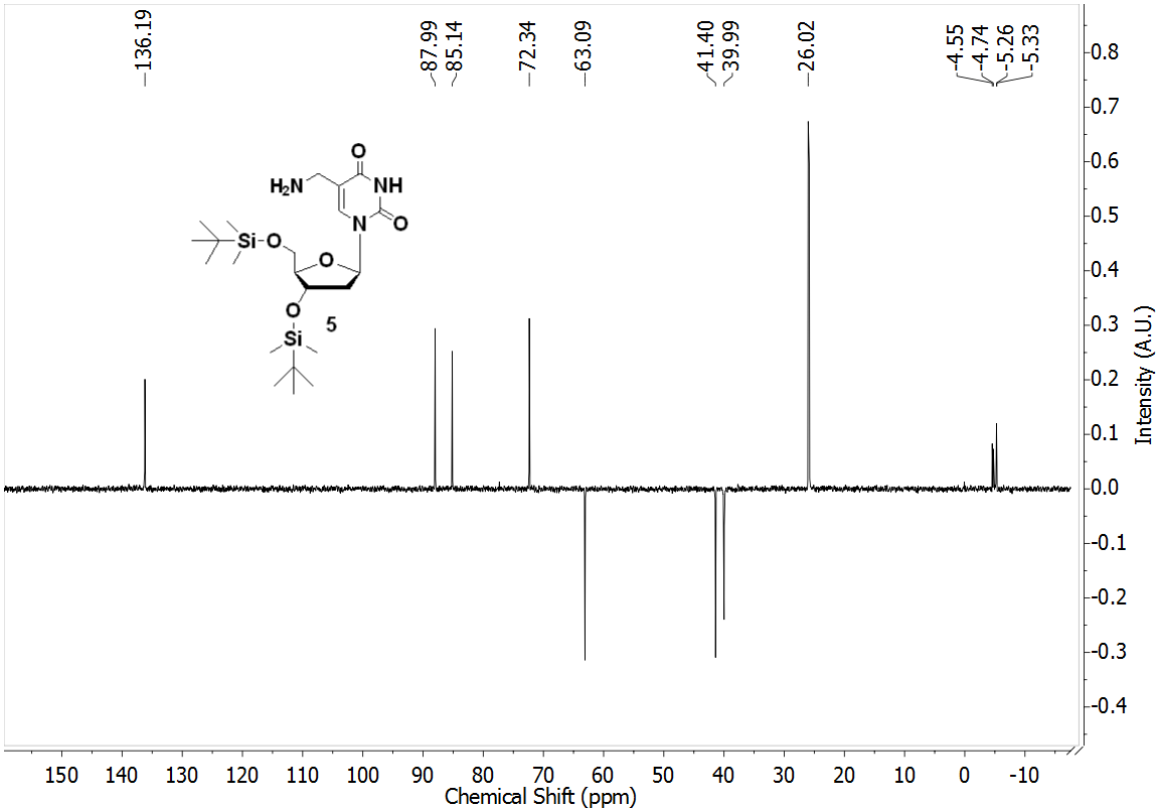
^{1}H NMR (400 MHz, Chloroform-*d*) of **5**:



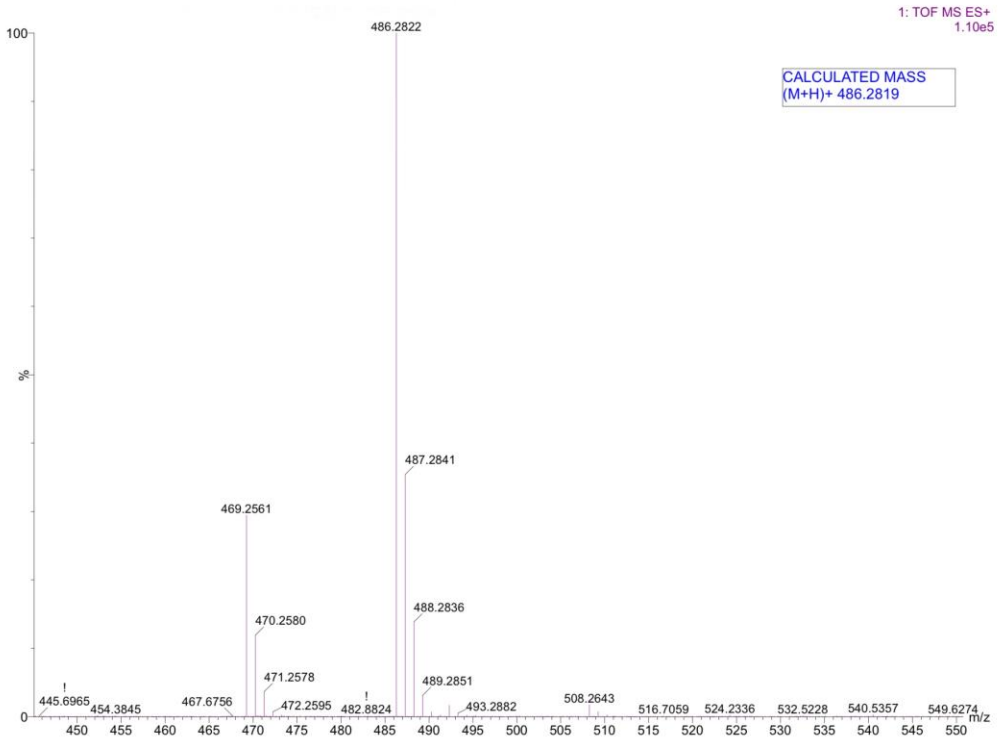
^{13}C NMR (101 MHz, Chloroform-*d*) of **5**:



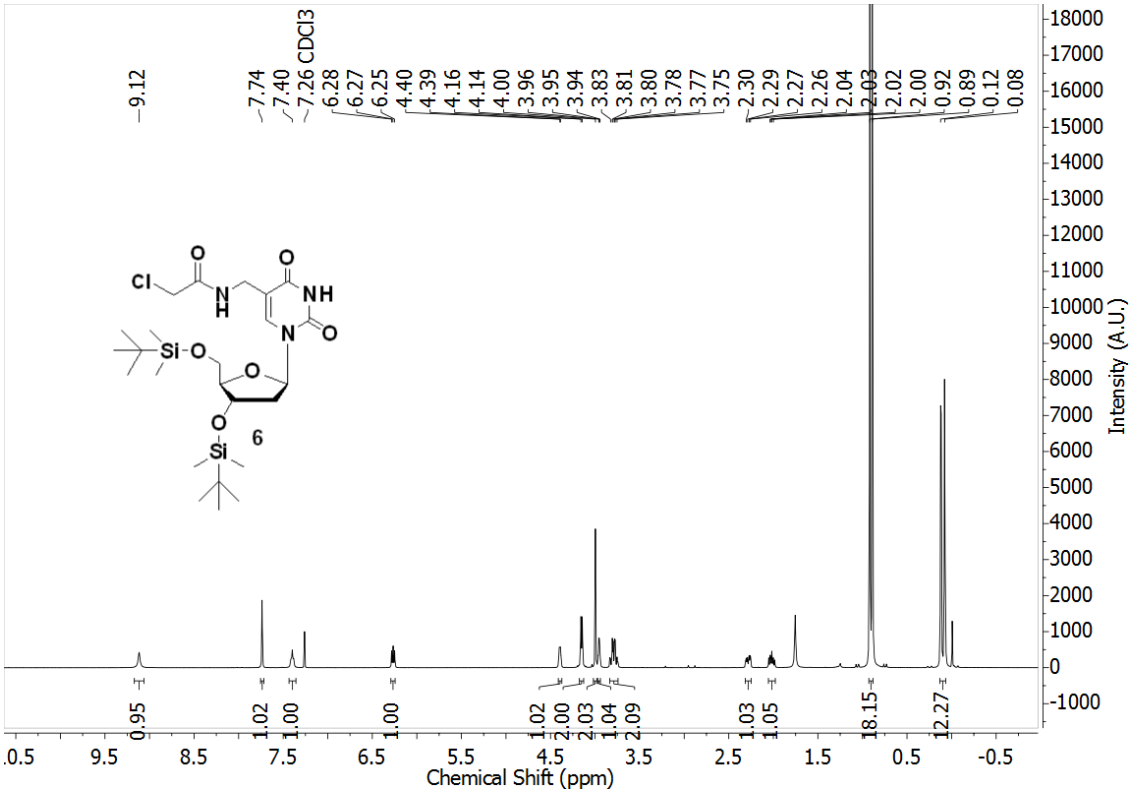
DEPT-135 (101 MHz, Chloroform-*d*) of **5**:



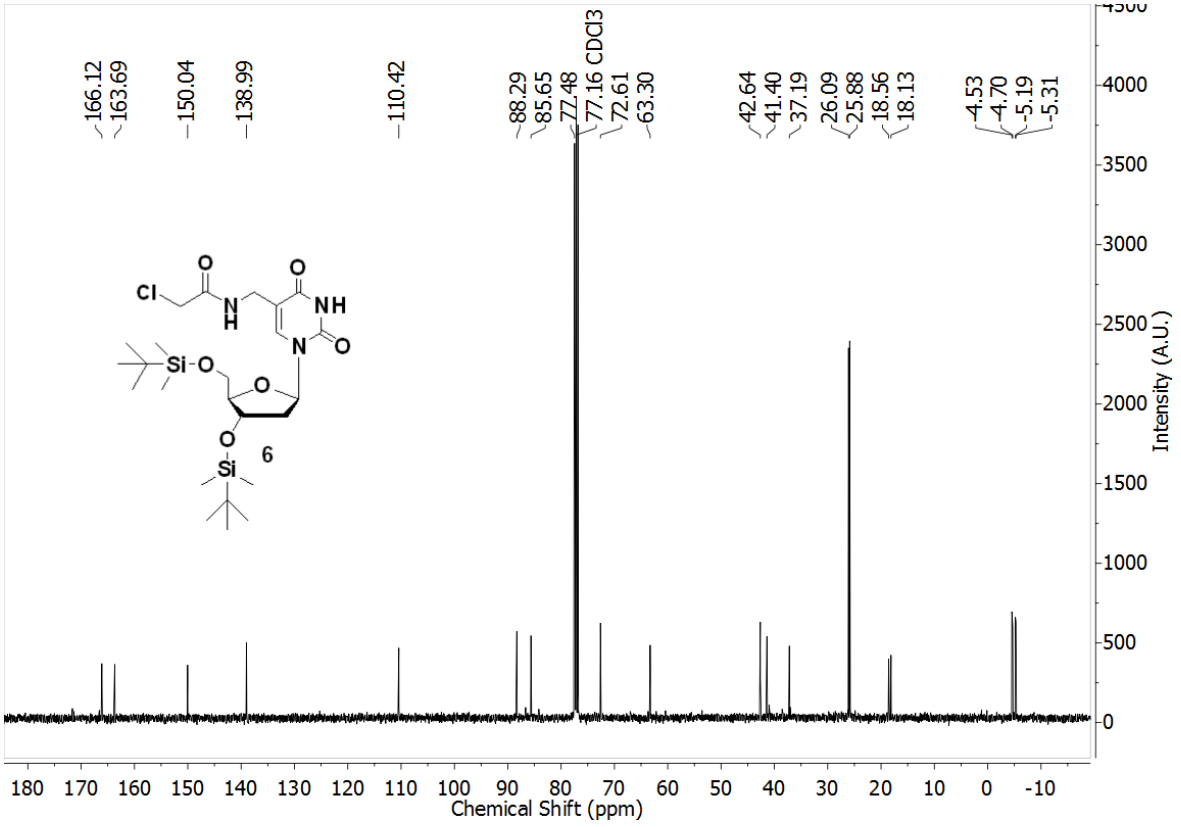
HRMS of **5**:



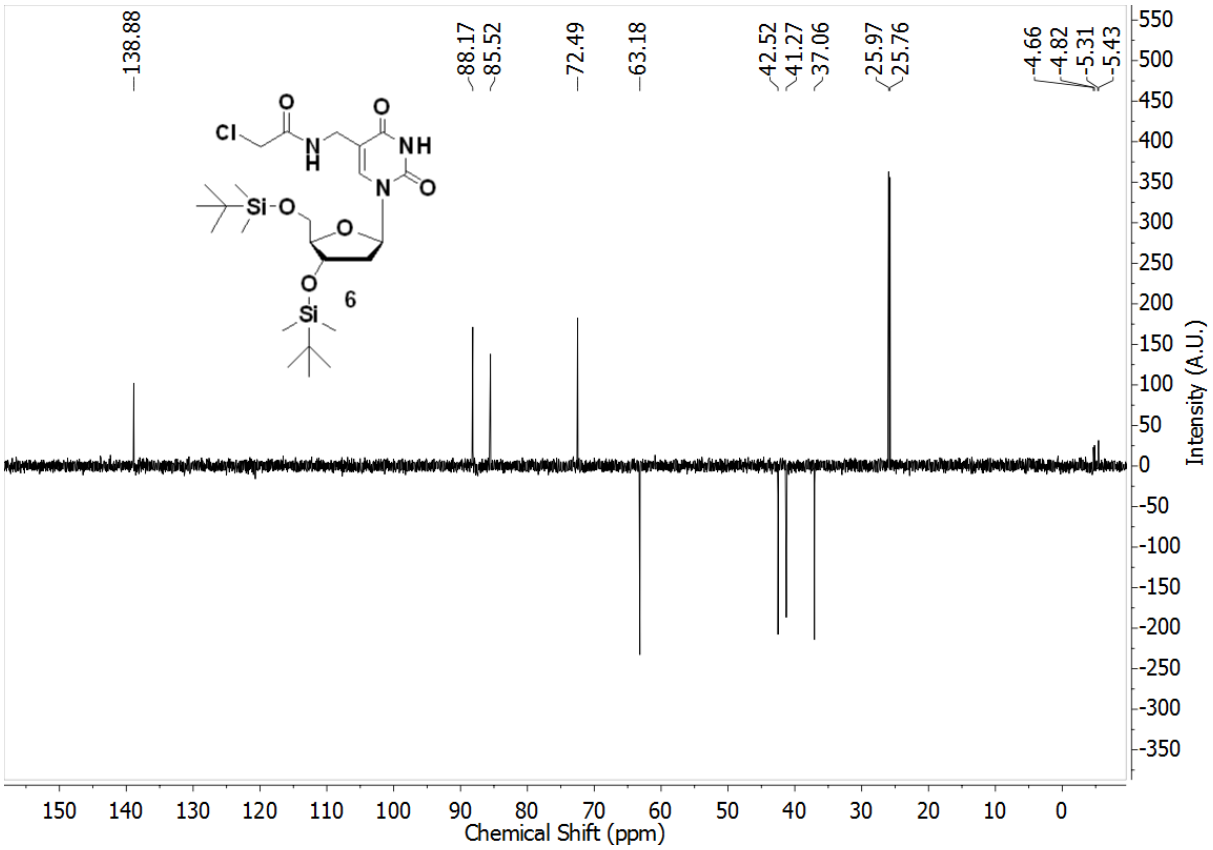
^{1}H NMR (400 MHz, Chloroform-*d*) of **6**. Trace amount of water is present.



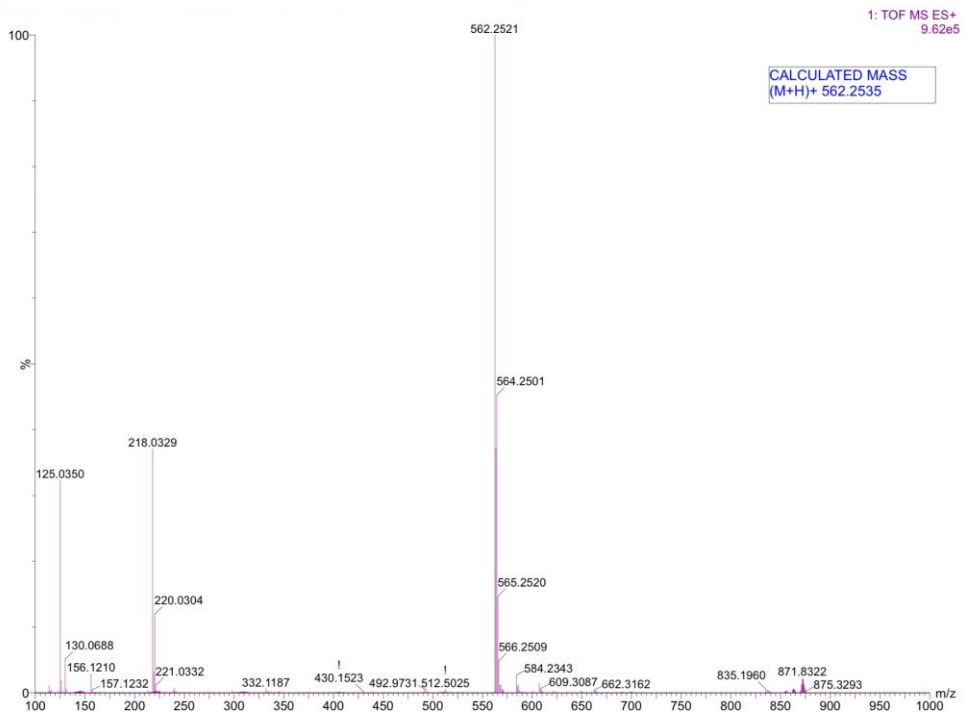
^{13}C NMR (101 MHz, Chloroform-*d*) of **6**:



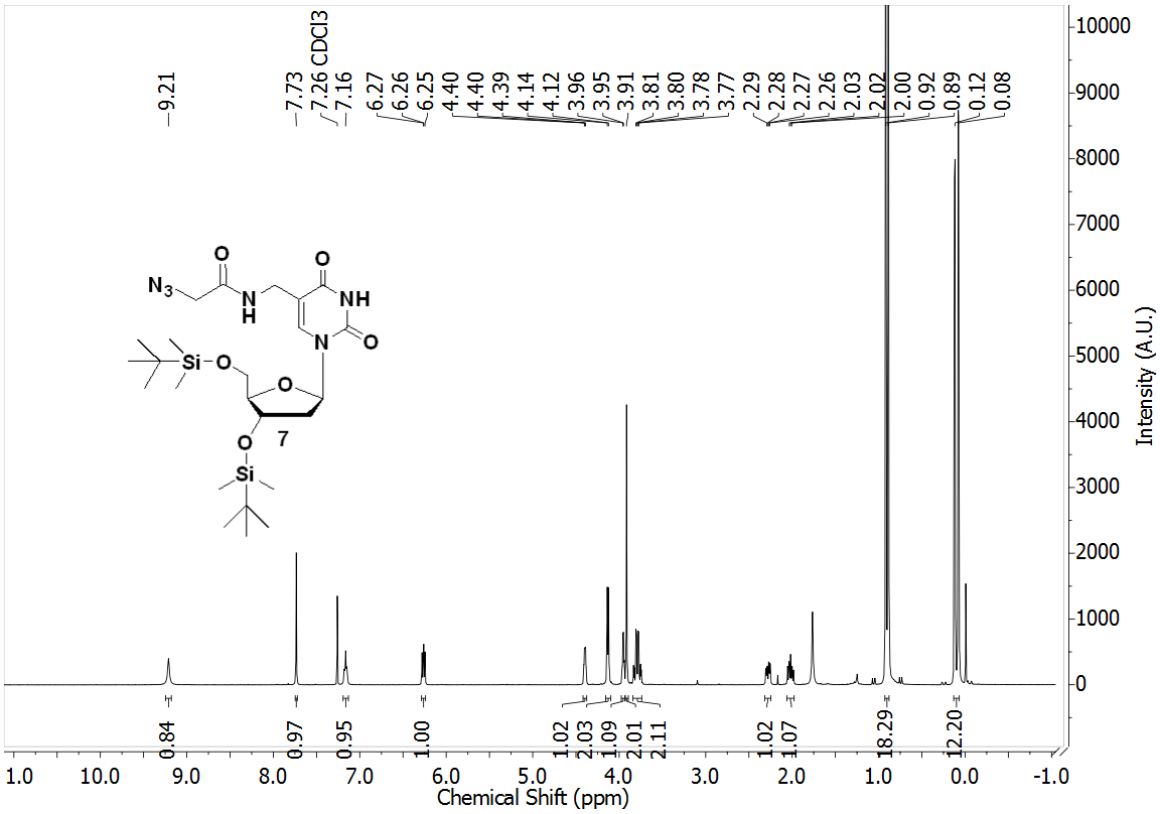
DEPT-135 (101 MHz, Chloroform-*d*) of **6**:



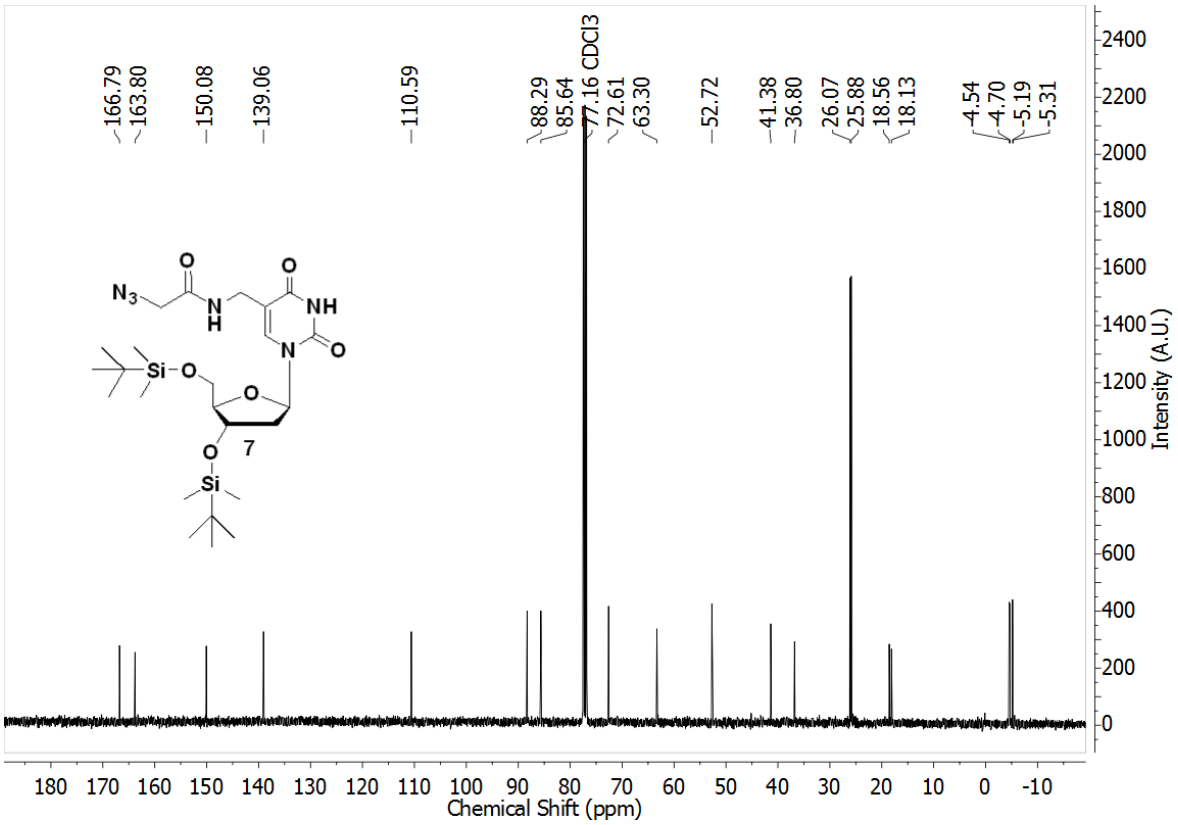
HRMS of **6**:



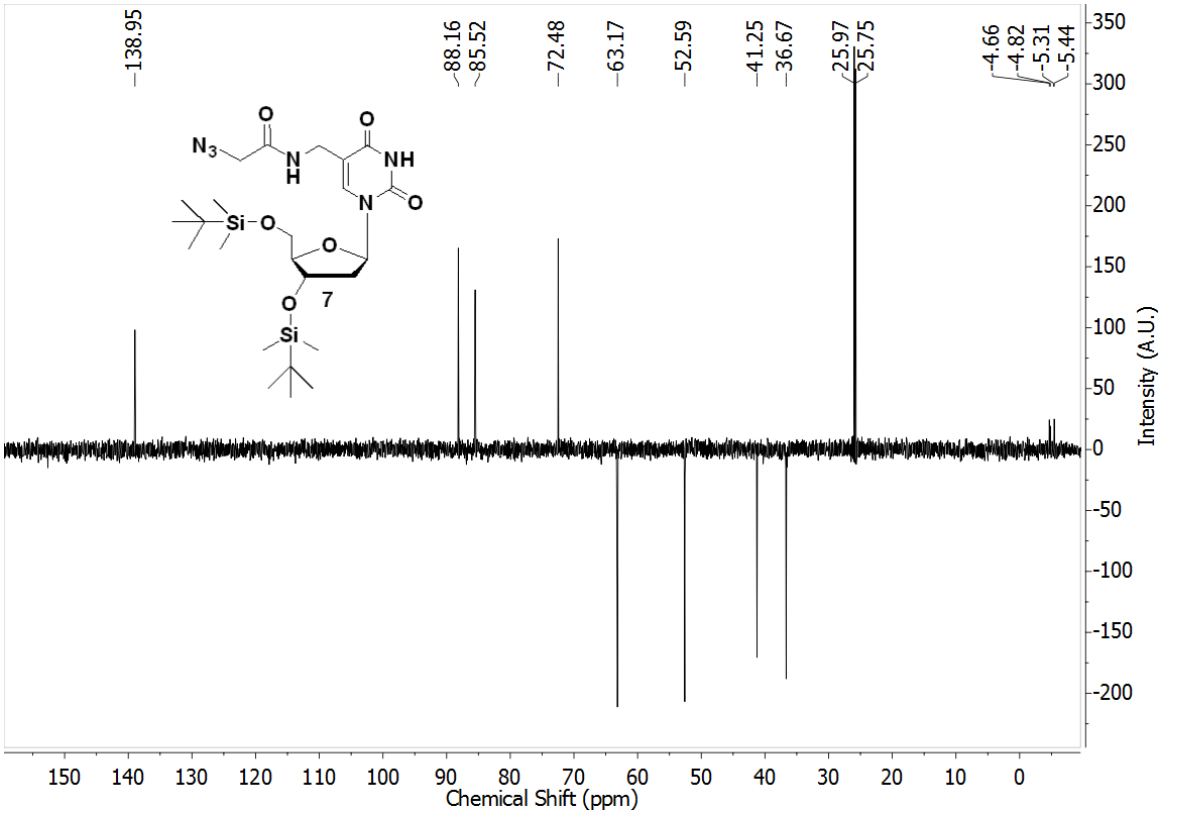
^{1}H NMR (400 MHz, Chloroform-*d*) of **7**. Trace amount of water is present.



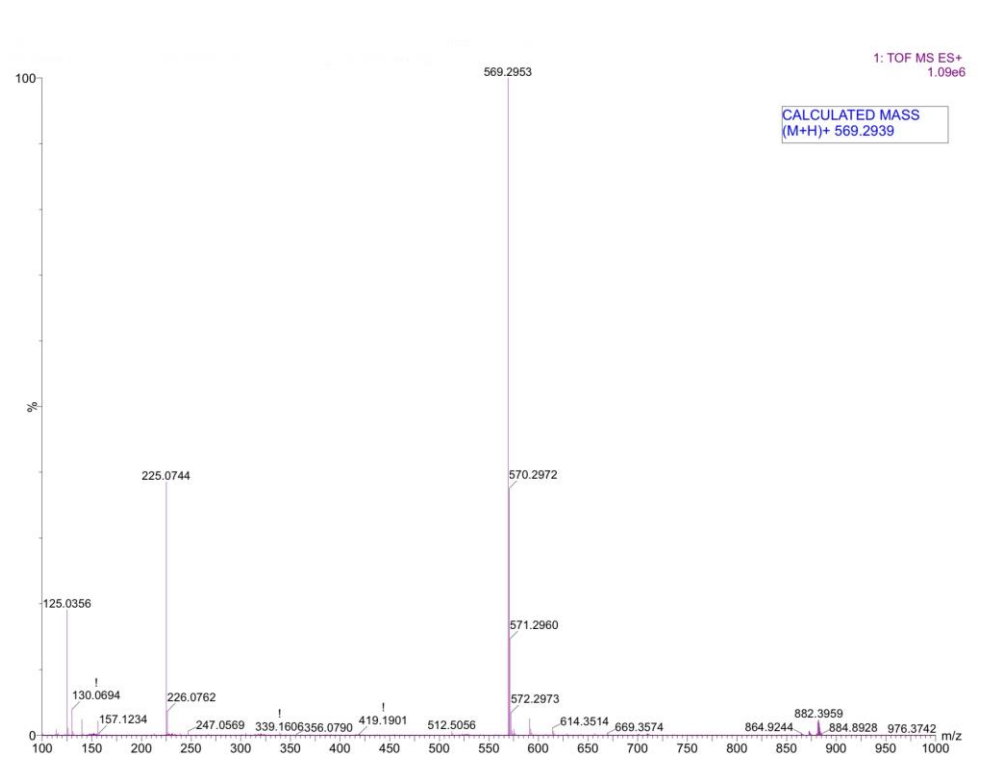
^{13}C NMR (101 MHz, Chloroform-*d*) of **7**:



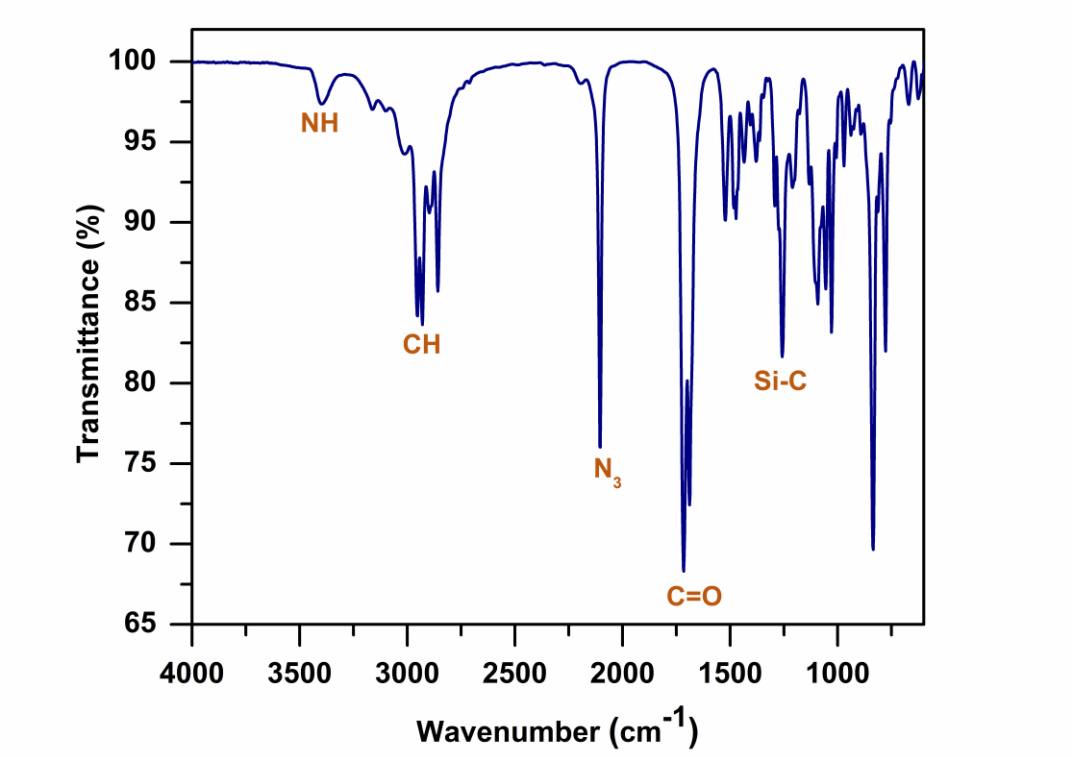
DEPT-135 (101 MHz, Chloroform-*d*) of **7**:



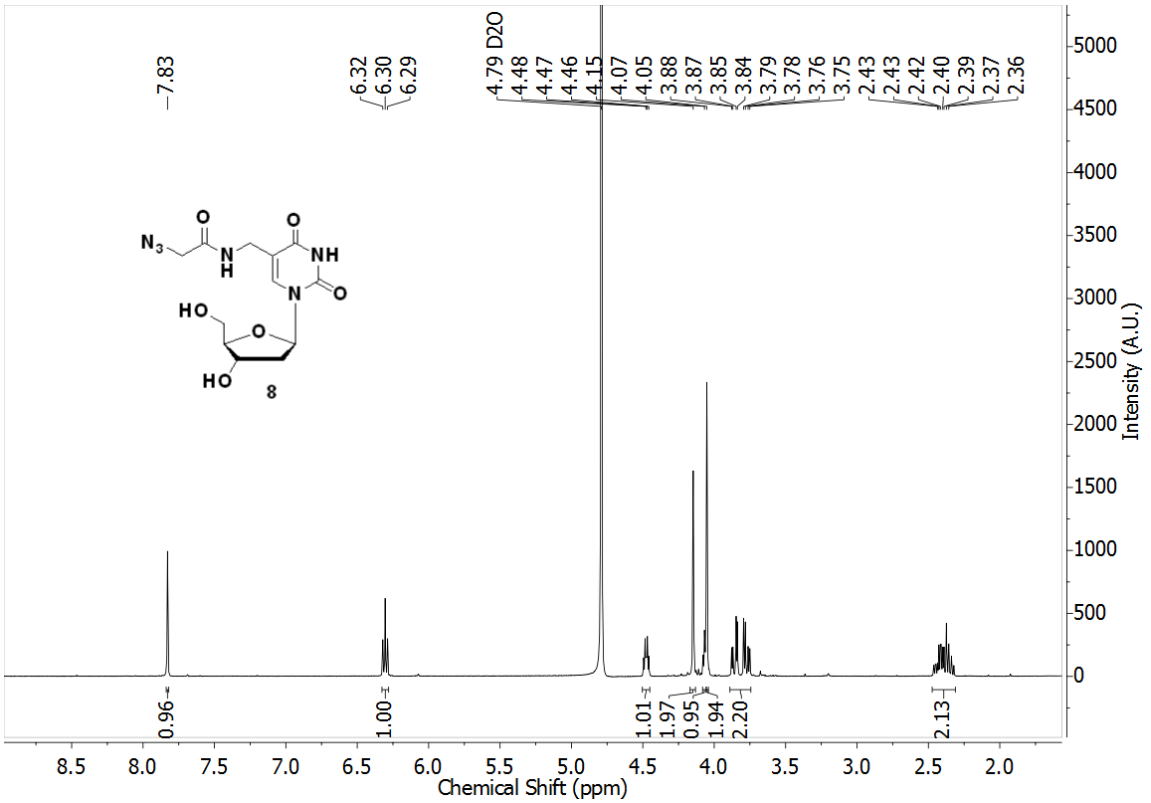
HRMS of **7**:



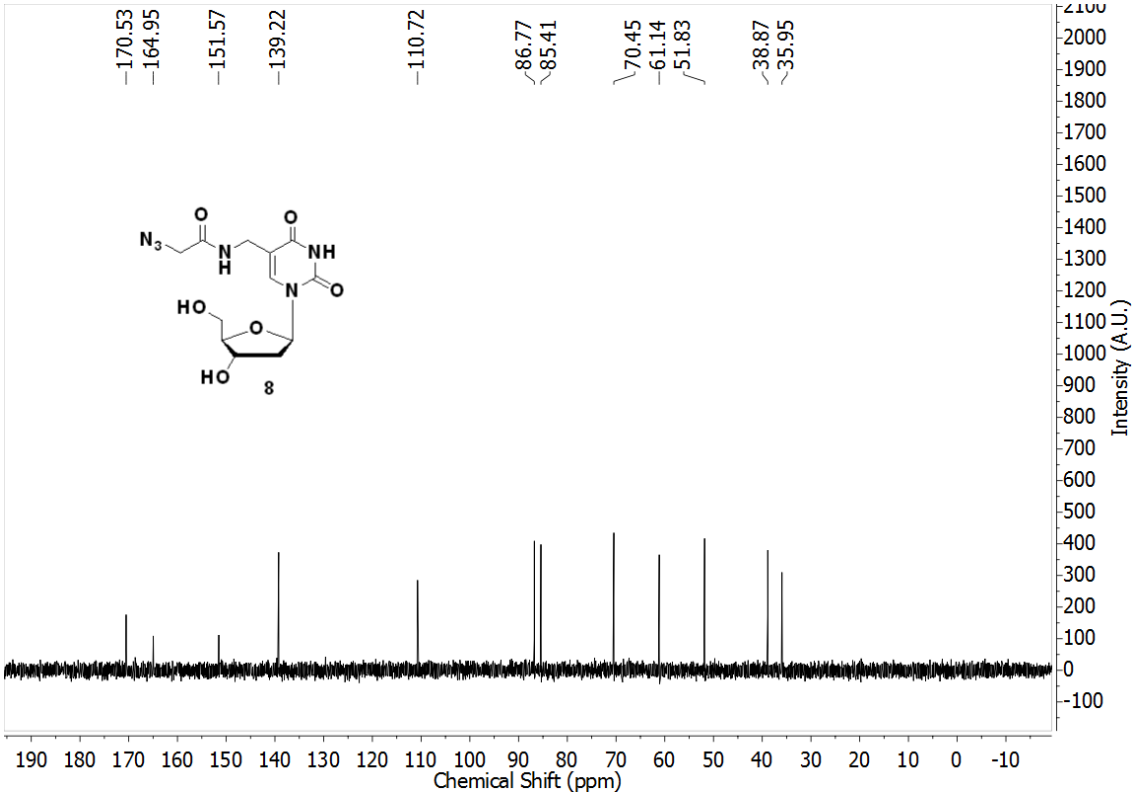
IR of **7**:



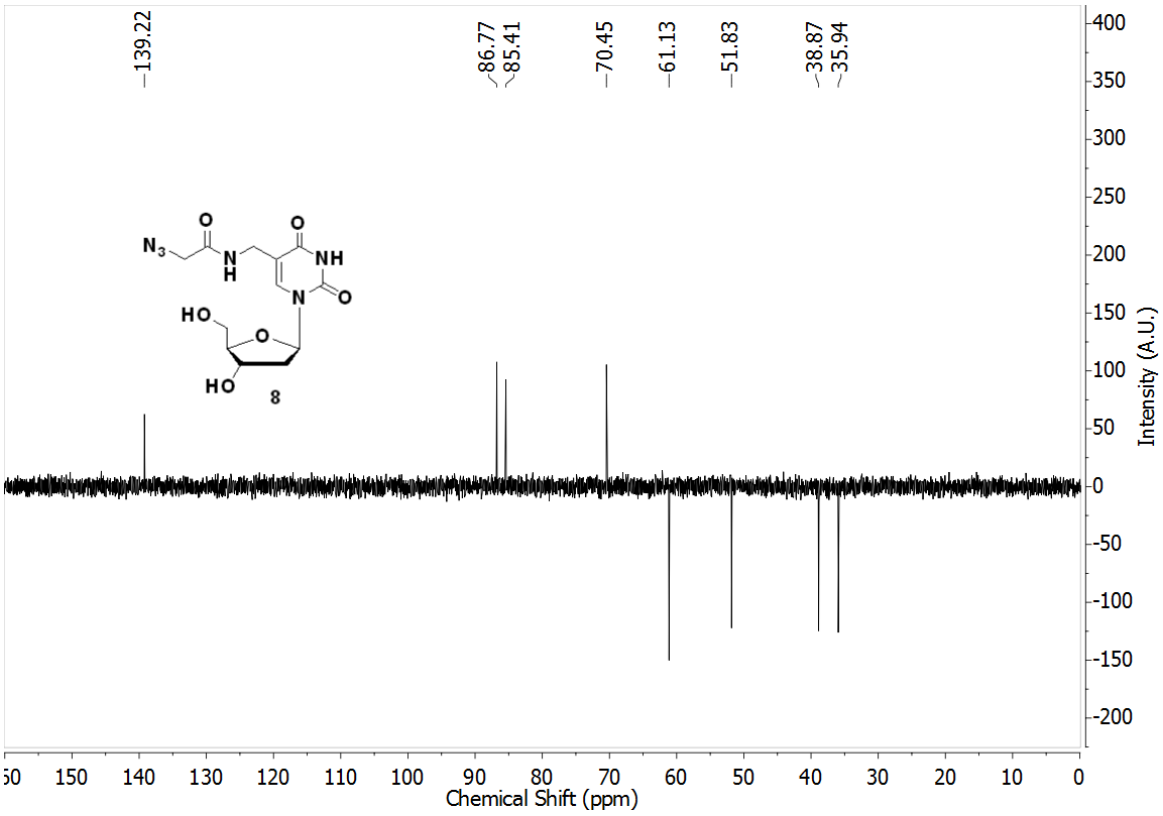
^{1}H NMR (400 MHz, Deuterium Oxide) of **8**:



^{13}C NMR (101 MHz, Deuterium Oxide) of **8**:



DEPT-135 (101 MHz, Deuterium Oxide) of **8**:



HRMS of **8**: