Design, Synthesis, and Evaluation of Cross-linked Single-stranded DNAs as Inhibitors of APOBEC3 Enzymes

A thesis presented in partial fulfilment of the requirements of the degree of

Doctor of Philosophy

in

Chemistry

at Massey University, Manawatū, New Zealand

HARIKRISHNAN MOHANA KURUP

2021

Dedicated to my family

Abstract

Drug resistance is a major problem associated with anti-cancer chemo- and immunotherapies. Recent advances in the understanding of resistance mechanisms revealed that APOBEC3 (A3) family enzymes contribute to the development of drug resistance in multiple cancers. A3 enzymes are polynucleotide cytidine deaminases that convert cytosine to uracil (C→U) in single-stranded DNA (ssDNA) and in this way protect humans against viruses and mobile retro-elements. On the other hand, cancer cells use A3s, especially A3A and A3B, to mutate human DNA, and thus by increasing rates of evolution, cancer cells escape adaptive immune responses and resist drugs. However, as A3A and A3B are non-essential for primary metabolism, their inhibition opens a strategy to augment existing anticancer therapies and suppress cancer evolution. It is known that ssDNA bound to both A3A and to a chimeric A3B ($A3B_{CTD}$) is not linear but adopts a distinctive U-shape, projecting the target cytosine into the active site. We hypothesized that locking DNA sequences into the observed U-shape may provide not only better substrates but also, appropriately modified, better inhibitors of A3 enzymes.

To test our hypothesis that pre-shaped ssDNA mimicking the U-shape observed in ssDNA-A3 complexes can provide a better binder to A3 enzymes, a Cu(I)-catalyzed azide-alkyne cycloaddition was used to create a cross-link between two modified nucleobases in ssDNA. The resultant cytosine-containing substrate, where the cytosine sits at the apex of the loop, was deaminated faster by A3B than a standard, linear substrate. The cross-linked ssDNA was converted into an A3B inhibitor by replacing the 2'-deoxycytidine in the preferred TCA substrate motif by 2'-deoxyzebularine (dZ) or 5-fluoro-2'-deoxyzebularine (5-FdZ), a known inhibitor of single nucleoside cytidine deaminases. This strategy yielded the first nanomolar A3B inhibitor (K_i = 100 ± 16 nM) and provides a platform for further development of modified ssDNAs as powerful A3 inhibitors.

We also synthesized three seven-membered ring-containing nucleosides as transition-state analogues of cytosine deamination and incorporated them in a ssDNA sequence to test the inhibition of A3 enzymes. In this work we successfully synthesized a nucleoside with a seven-membered ring-containing double bond (ddiazep) and two nucleosides having a seven-membered ring with hydroxyl groups (*R* and *S* isomers) as a nucleobase. However, the inhibition of A3 by these compounds was not as good as by a ssDNA containing dZ. Interestingly, there was a difference in the inhibition produced by seven-membered ring-containing *R* and *S* isomers. The inhibition data showed that the relative *S* stereochemistry was essential in the inhibition of A3.

Contributions

All the work in this thesis was completed by Harikrishnan Mohana Kurup

Except:

- Expression and purification of A3 enzymes were done by Dr. Fareeda M. Barzak and Dr. Stefan Harjes.
- ^{1}H NMR assays of terminal cross-links were performed by Dr. Fareeda M. Barzak
- ^{1}H NMR assays of cross-linked inhibitors using short hairpin oligos as substrate were performed and analyzed by global regression using Lambert's W function by Dr. Stefan Harjes.
- Characterization of substrates and inhibitors by non-linear regression analysis was done by Emeritus Professor Geoffrey B. Jameson.
- All the isothermal titration calorimetry (ITC) experiments were performed by Dr. Stefan Harjes.
- Nuclease resistance of the cross-linked oligo against snake venom phosphodiesterase was performed by Dr. Yongdong Su.
- X-ray crystal structure of the seven-membered ring-containing compound was determined by Dr. Subo Lee.
- All HRMS-ESI data were collected by Mr. David Lun.

Acknowledgements

Firstly, I would like to extend my sincere gratitude to my supervisory team, Assoc. Prof. Vyacheslav V. Filichev, Dr. Elena Harjes and Emeritus Prof. Geoffrey B. Jameson for choosing me to work on this research and the excellent and continuous support throughout the journey. This research is the outcome of their continuous advice, motivation, discussions, and fabulous support.

I would like to express my heartfelt thanks Dr. Maksim V. Kvach, who provided a great deal of theoretical and technical support in the lab for the first two years of my PhD. Also, a big thanks to our APOBEC and chemistry lab group members Dr. Stefan Harjes, Dr. Fareeda M. Barzak, Dr. Yongdong Su, Mr. Bruce Chilton for the long discussions, technical support, and ideas. Special thanks to Dr. Stefan Harjes and Dr. Fareeda M. Barzak for the expression and purification of the enzymes and the technical support provided for ITC and NMR assays.

A sincere thanks to SFS for paying the student fees and the support provided for the entire duration of my PhD. Thanks to Maurice Wilkins Centre for Molecular Biodiscovery and The Health Research Council of New Zealand for funding.

A heartfelt thanks to Dr. Maulik Mungalpara, Dr. Joel Cornelio, Mr. Shashank Tewari for the wonderful discussions. Thanks to my lunch buddies for the countless laughs and topics we discussed during lunch breaks.

Words have no power to express my deep sense of reverence to my mom and dad who have made innumerable sacrifices for my well-being. Their unconditional love & support, without any expectations, is truly amazing. Last but not the least, I would like to extend my heartfelt thanks to my wife, who continues to amaze me with her ability to manage little Aarav and me.

Publications relevant to this thesis

Publications relevant to this thesis

1. **Kurup, H. M**.; Kvach, M. V.; Harjes, S.; Barzak, F. M.; Jameson, G. B.; Harjes, E.; Filichev, V. V., Design, Synthesis, and Evaluation of a Cross-Linked Oligonucleotide as the First Nanomolar Inhibitor of APOBEC3 Enzymes. *Biochemistry*. (After positive referee comments, the article was withdrawn due the potential patentability of the research).

Provisional Patent filed on 26 Jan 2021 (Australia, ID 2021900164).

An international-type search was expedited through the Australian Patent Office with almost all claims being found novel and inventive in the opinion of the examiner. A patent strategy and landscape search has also been performed, suggesting that very limited prior art currently exists in the space providing a unique opportunity to build a robust and defensible IP portfolio around APOBEC3 inhibitors and methods of treatment.

Additional publications

1. Barzak, F. M.; Harjes, S.; Kvach, M. V.; **Kurup, H. M.**; Jameson, G. B.; Filichev, V. V.; Harjes, E., Selective inhibition of APOBEC3 enzymes by single-stranded DNAs containing 2'-deoxyzebularine. *Org. Biomol. Chem.* **2019,** *17* (43), 9435 - 9441.

2. Kvach, M. V.; Barzak, F. M.; Harjes, S.; Schares, H. A. M.; **Kurup, H. M.**; Jones, K. F.; Sutton, L.; Donahue, J.; D'Aquila, R. T.; Jameson, G. B.; Harki, D. A.; Krause, K. L.; Harjes, E.; Filichev, V. V., Differential Inhibition of APOBEC3 DNA-Mutator Isozymes by Fluoro- and Non-Fluoro-Substituted 2'-Deoxyzebularine Embedded in Single-Stranded DNA. *ChemBioChem* **2020,** *21* (7), 1028-1035.

3. Barzak, F. M.; Ryan, T. M.; Kvach, M. V.; **Kurup, H. M.**; Aihara, H.; Harris, R. S.; Filichev, V. V.; Harjes, E.; Jameson, G. B., Small-Angle X-ray Scattering Models of APOBEC3B Catalytic Domain in a Complex with a Single-Stranded DNA Inhibitor. *Viruses* **2021,** *13* (2), 290.

In the above publications, my contribution was the synthesis of oligonucleotide-based inhibitors used in the studies.

List of Abbreviations

Abbreviation	Definition
°C	Degrees Celsius
× g	Multiples of gravitational force
Δ*G*	Gibbs free energy change
Δ*H*	Enthalpy change
Δ*H*app	apparent heat change change
Δ*S*	Entropy change
NMR	Nuclear Magnetic Resonance
V_0	Initial rate of a reaction
[S]	Substrate concentration
[I]	Inhibitor concentration
5-FdZ	5-Fluoro-2'-deoxyzebularine
Å	Ångstrom (10^{-10} m)
A_{260}	Absorbance at 260 nm
A1	APOBEC1
A2	APOBEC2
A3(A-H)	APOBEC3A-H
$A3B_{CTD}$-AL1	$A3B_{CTD}$-QM-ΔL3-AL1swap
$A3B_{CTD}$*	$A3B_{CTD}$-QM-ΔL3-A3Aloop1swap-E255A
A4	APOBEC4
ACN	Acetonitrile
ACS	American Chemical Society
AID	Activation-induced cytosine deaminase
APOBEC	Apolipoprotein B mRNA-editing enzyme-catalytic polypeptide
Ara-C	Cytarabine
aq.	Aqueous
Boc	*tert*-Butoxycarbonyl
BRCA	Breast cancer gene
bs	Broad singlet
BSA	Bovine serum albumin
cDNA	Complementary DNA
CDA	Cytosine deaminase
COSY	Correlation spectroscopy

CPG	Controlled-porous glass
CTD	Carboxyl terminal domain
CsPB	Bacterial cold shock protein
C-terminal	Carboxyl terminal
CuAAC	Cu(I)-catalyzed azide-alkyne cycloaddition
Da	Dalton
DCM	Dichloromethane
DMAP	4-(Dimethylamino)pyridine
DMF	*N*,*N*-Dimethylformamide
DMSO	Dimethylsulfoxide
DMT	4,4′-Dimethoxytrityl
DNA	2'-Deoxyribonucleic acid
DSS	4,4-Dimethyl-4-silapentane-1-sulfonic acid
d	Doublet
dsDNA	Double-stranded DNA
dA	2'-Deoxyadenosine
dC	2'-Deoxycytidine
dG	2'-Deoxyguanosine
T	Thymidine
dU	2'-Deoxyuridine
dZ	2'-Deoxyzebularine
dd	Doublet of doublet
ddd	Doublet of doublet of doublet
[E]	Enzyme concentration
EDTA	Ethylenediaminetetraacetic acid
ER+	Estrogen receptor positive
HBV	Hepatitis B virus
HMBC	Heteronuclear multiple bond correlation spectroscopy
HPLC	High-performance liquid chromatography
HPV	Human papilloma virus
HRMS	High resolution mass spectrometry
Hz	Hertz
hCDA	Human cytosine deaminase
HSQC	Heteronuclear single quantum coherence
IR	Infrared
ITC	Isothermal titration calorimetry

K_a	Equilibrium binding association constant
k_{cat}	Catalytic rate constant
k_{cat}/K_m	Specificity constant (catalytic efficiency)
K_d	Equilibrium dissociation binding constant
K_i	Inhibition constant
K_m	Michaelis constant (apparent binding)
LCAA-CPG	Long chain aminoalkyl controlled-porous glass
m	Multiplet
min	Minute
mM	Millimolar
mRNA	Messenger RNA
NMR	Nuclear magnetic resonance
NOESY	Nuclear Overhauser effect spectroscopy
N-terminal	Amino terminal
NTD	N-terminal domain
OD	Optical density
Oligo	Oligonucleotide
PBMC	Peripheral blood mononuclear cells
PDB	Protein data bank
PS-Oligo	Phosphothioated oligonucleotide
ppm	Parts per million
RCM	Ring closure metathesis
R_f	Retention factor
ROS	Reactive oxygen species
RP-HPLS	Reverse-phase-HPLC
RNA	Ribonucleic acid
r.t.	Room temperature
[S]	Substrate concentration
s	Singlet
sA3G	Soluble-A3G
sec	Seconds
ssDNA	Single-stranded DNA
TBAF	Tetrabutylammonium fluoride
TBDPS	*tert*-Butyldiphenylsilyl
TBTA	tris[(1-Benzyl-1*H*-1,2,3-triazol-4-yl)methyl]amine
TCEP	(tris(2-Carboxyethyl)phosphine)

TEAA buffer	Triethylamine acetic acid buffer
THF	Tetrahydrofuran
THU	Tetrahydrouridine
TIC	Total ion current
TIPS	Tri*iso*propyl silyl
TLC	Thin layer chromatography
TMS	Trimethyl silane
TSP	3-(Trimethylsilyl)-2,2,3,3-tetradeuteropropionic acid
t	Triplet
tRNA	Transfer RNA
UDG	Uracil DNA glycosylase
UV	Ultraviolet
V_0	Initial rate of a reaction
Vif	Virion infectivity factor
V_{max}	Maximum rate of a reaction
Wt	Wild type

Amino acid names and abbreviations

Amino acid	3-letter abbreviation	1-letter abbreviation
Alanine	Ala	A
Arginine	Arg	R
Asparagine	Asn	N
Aspartic acid	Asp	D
Cysteine	Cys	C
Glutamic acid	Glu	E
Glutamine	Gln	Q
Glycine	Gly	G
Histidine	His	H
Isoleucine	Ile	I
Leucine	Leu	L
Lysine	Lys	K
Methionine	Met	M
Phenylalanine	Phe	F
Proline	Pro	P
Serine	Ser	S
Threonine	Thr	T
Tryptophan	Trp	W
Tyrosine	Tyr	Y
Valine	Val	V

Table of Contents

List of Figures

List of tables

List of schemes

Chapter 1. Introduction

Research and development investment in the field of cancer therapy by both academic and industrial concerns is enormous. One of the major reasons for this is the fact that there is no single agent which is a gold standard for any of the therapy regimens as of now. The main drawbacks of existing therapy methods are non-specificity, high cost of therapy, high risk-to-benefit ratio and most importantly development of drug resistance by tumors due to metastasis and mutagenesis.[1] At the beginning of the 21st century, little was known about the mechanisms of development of resistance by cancer cells. In most of the cases the chemotherapeutic agent which is active at the early periods of therapy fails later. This leads to recurrence of the disease and ultimately failure of therapy, leaving a huge health and economic burden on the patient and family.

Traditional cancer therapy methods include chemotherapy, radiotherapy, immunotherapy, combination therapy and surgery. To achieve therapeutic efficacy and to address the hurdles mentioned above it is necessary to focus on novel therapeutic combinations which would address them in an effective and specific manner. There is a rapid growth in understanding of the molecular mechanisms of cancer including resistance pathways.[2-6] The underlying downstream pathways inducing drug resistance are potential drug targets. Over the past decades combination therapy has been employed to minimize drug resistance but the overall cost of therapy and the side effects of the drug cocktail are still a question. Recent studies on many cancer cell-lines point towards an upregulation of a particular class of enzyme which mutate host genomic DNA and thereby drives cancer evolution allowing cancer cells to evade the immune system, develop drug resistance and metastasis.[7-9] These enzymes, known as APOBEC3, function by deaminating the cytosines in the single-stranded DNA (ssDNA) to uracil and thereby inducing mutations.

1.1. APOBEC enzyme superfamily

APOBEC3 enzymes are a sub-group of the larger Apolipoprotein B m-RNA editing enzyme catalytic polypeptide (APOBEC) gene family. They are zinc^{2+}-dependent cytosine deaminases that catalyze the conversion of cytosines to uracils (C-to-U) on a ssDNA (**Figure 1.1**).[10] The family consists of five major class of enzymes APOBEC1 (A1), APOBEC2 (A2), APOBEC3 (A3), APOBEC4 (A4), and activation-induced deaminases (AID). A1 has been shown to be involved in lipid metabolism and transport by modulating ApoB protein.[11-12] The functions of A2 and A4 are unknown, and both are catalytically inactive. A3 enzymes are arranged in tandem on human chromosome 22. So far 7 isoforms of A3 i.e., A3A, A3B,

A3C, A3D, A3F, A3G and A3H have been reported to be expressed in humans.[10] The first identified A3 enzyme, A3G, is the most studied because of its importance in restriction of human immunodeficiency virus-1 (HIV-1).[13] A3G was found to be expressed by human T cells and restricts the HIV genome by deaminating cytidines in the viral complementary DNAs (cDNA).[14-18] So far A3 enzymes are shown to be important in the context of inhibiting exogenous retroviruses and retrotransposons as a part of human innate immunity. A3s are also reported to restrict other DNA viruses including adenoviruses, hepatitis B virus (HBV), herpes, and human papilloma virus (HPV).[19-22] Recent research on transcriptomic profiling and genomic mutational analysis of SARS-CoV-2 has shown a high level of mutations producing UU pairs, characteristic of A3 induced mutations.[23-24] Even though A3s try to mutate and degrade the virus, the ability of the virus to counteract the effect of A3-induced mutations are one of the leading causes for the emergence of virulent strains.[24]

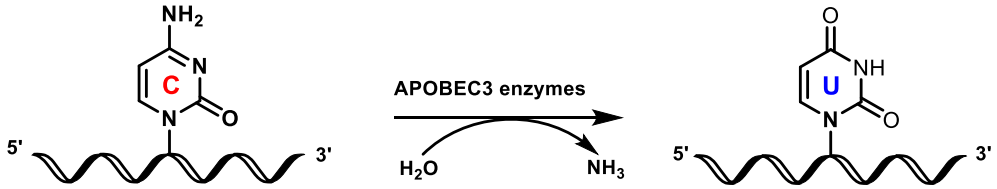


Figure 1.1. A3-catalyzed deamination of dC to dU in a ssDNA.

1.2. Expression and localization of different A3 enzymes

Because of the protective role of A3s, they are expressed in most of the tissues in human body and the level of expression can be tissue specific.[25-26] They are expressed in varying levels in immune cells including $CD4^+$, $CD8^+$ cells, B cells and myeloid cells. In a fresh peripheral blood mononuclear cell (PBMC) sample, A3C, A3D, A3F, A3G and A3H are highly expressed in B cells and T cells.[25] A3A is abundantly expressed in monocytes and A3B in B cells. All A3s are shown to be expressed in spleen.[26] Apart from immune cells, other tissues also express A3s in varying degrees. It is reported that A3A is over-expressed in lungs and adipose tissue and this may be attributed to the presence of large quantity of $CD14^+$ macrophages.[26] A3B, C, D and H are also expressed in lungs. Interestingly, A3B was found to be over-expressed in breast cancer tissues, which correlates with the role A3B plays, especially in breast cancer evolution.[27]

A3 enzymes have varying levels of localization in cells. A3D, F, G and H are cytoplasmic, A3A and A3C are widely distributed whereas A3B is predominantly nuclear.[28-29] The nuclear localization of A3B made it interesting to study in the background of the evolutionary aspects of cancer.

1.3. Structural insights of A3 enzymes

Significant efforts have been made in the past to study the structural features of different A3 enzymes mainly by NMR and crystallographic methods.[30-37] A3 enzymes exist either as a single domain (A3A, A3C, and A3H) or as a double domain (A3B, A3G and A3F). The double domain enzymes are characterized by the presence of a catalytically active C-terminal domain (CTD) and the N-terminal domain (NTD) is reported to be catalytically inactive.[15, 38] However, NTD is important in binding ssDNA substrates and enhancing the catalytic activity of the CTD.[39-40] All A3 enzymes are $zinc^{2+}$-dependent for their catalytic activity, which is like CDA enzymes and so the mechanism of deamination is believed to be similar for all of them. The conserved $zinc^{2+}$-binding residues are arranged adjacent to N-terminal end of two α helices within the CDA. The active site is conserved between CDA and A3 family. The defining feature of the APOBEC family is the presence of a hydrophobic core, six α helices, and five β strands (**Figure 1.2**).[33, 41] The substrate selectivity and difference in catalytic activity among the family members are largely contributed by the amino acid composition, length, and spatial arrangements of protein loops.[41] It has been shown that the parallel arrangement of β3-β4-β5 in A3s might be the key structural feature that distinguishes A3s from CDA.[42]

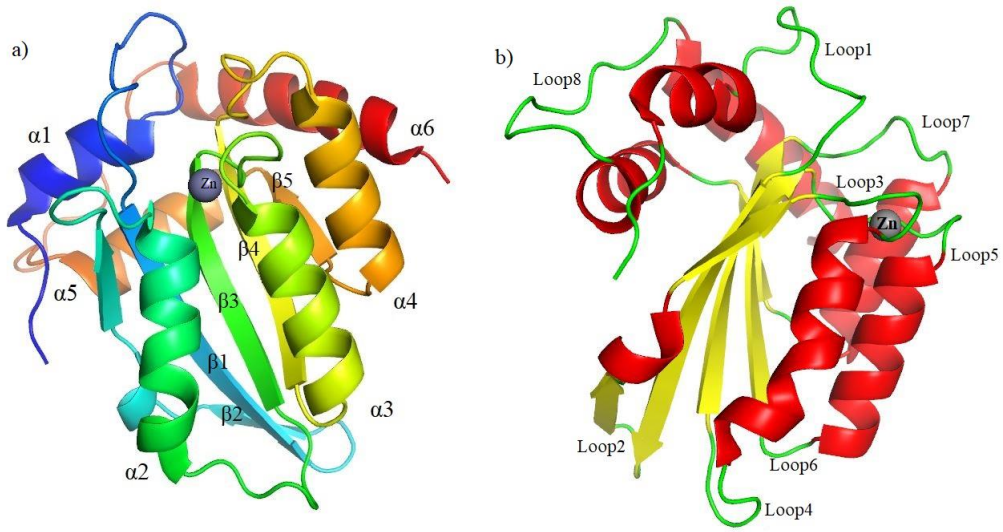


Figure 1.2. Crystal structure of $A3B_{CTD}$ PDB 5CQI ($A3B_{CTD}$-QM-ΔL3), showing different helices, strands, and loops.

a) The protein is color-coded in rainbow color, where the NTD is blue and CTD is red. b) α helices are shown in red, β strands are in yellow and the unstructured loops are shown in green. The structure is rendered in PyMol (2.5).

Several X-ray and NMR structures of A3G have been elucidated including individual CTDs and NTDs.[33-34, 43-46] All these structures share the same secondary structures including six α-helices and five β sheets with the Zn^{2+}-binding domain. It has been established that A3G

forms dimers and oligomers and that monomeric human A3G is more catalytically active compared to other forms.[47] Crystallization of full-length human A3G is challenging due to the protein's poor solubility. Due to the solubility issue, crystal structures of a soluble variant of rhesus macaque A3G (rmA3G) with limited catalytic activity and a soluble human A3G (sA3G) with a 2'-deoxycytidine dinucleotide bound near the catalytic site have been crystallized (**Figure 1.3**).[48-49] It has been shown that ssDNA binding regions are loops 1, 3, and 7 in CTD and loop 1 in NTD.[49] The dinucleotide in the crystal structure may have originated during the purification of the protein and was found to be important in the stability of the protein. As the dinucleotide is found away from the Zn^{2+}, it is not in a position to be deaminated.[49]

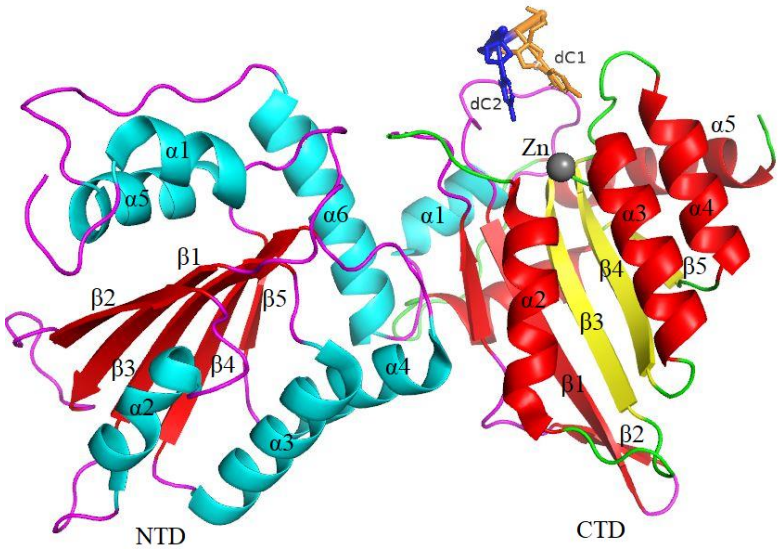


Figure 1.3. Crystal structure of a full length soluble human A3G (sA3G) bound to dinucleotide.

The structure is rendered in PyMol (2.5), PBD ID: 6WMB.

Based on the evolutionary and structural features of Zn^{2+} binding motif, A3 enzymes are divided into Z1, Z2 and Z3 phylogenetic sub-groups.[50-51] A3B, A3D, A3F, and A3G are double domain enzymes, in which A3B and A3G have Z2-Z1 domains and A3D and A3F contain Z2-Z2 domains (**Figure 1.4**).[52]

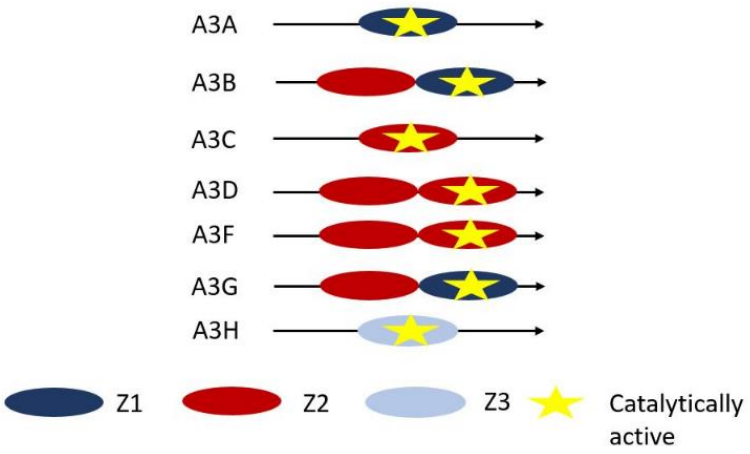


Figure 1.4. Schematic representation of structural domains of A3 enzymes.

1.3.1. Mechanism of A3-catalyzed deamination of cytosine to uracil in a ssDNA

Even though the mechanism of A3-catalyzed deamination is similar to that of CDA, A3 enzymes cannot deaminate individual cytidines and cytosines. It has been shown that A3s only deaminate cytidine if and only if it is incorporated into a ssDNA.[53-54] Upon binding of ssDNA to the enzyme, glutamate removes a proton from the water molecule which is coordinated by Zn^{2+}. The proton abstracted from water by Glu is transferred to N3 of cytosine to protonate it. The protonation makes the C4 of cytosine prone to nucleophilic attack by hydroxide forming a tetrahedral intermediate. Elimination of ammonia follows the collapse of the tetrahedral intermediate by protonation of $-NH_2$ facilitated by Glu and the uracil formed departs the active site (**Figure 1.5**).[29]

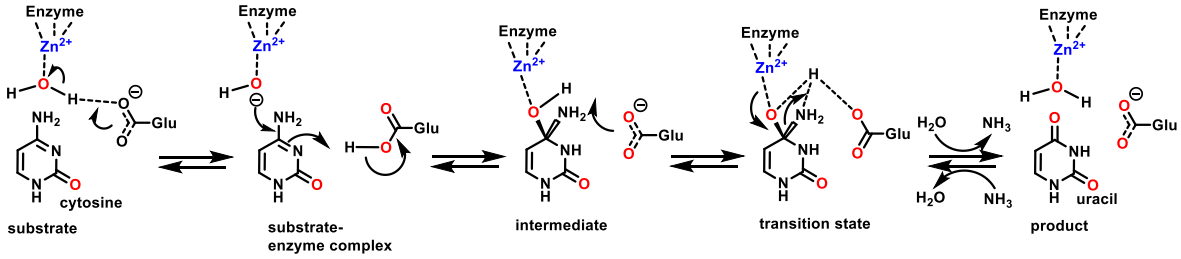


Figure 1.5. Proposed deamination mechanism of cytosine to uracil by A3 enzymes.

The mechanism of how ssDNA interacts with APOBECs is mediated by opening of the loops on the protein surface leading to the exposure of the catalytic site. This also facilitates interactions of positively charged and aromatic hydrophobic amino acid residues interacting with the negatively charged phosphate backbone and aromatic nucleic acid bases.[37]

1.3.2. Substrate preference of different A3 enzymes

A3 enzymes as discussed previously only deaminate dC in a ssDNA, but it is known that there is an intrinsic preference of different A3s towards different sequences of ssDNAs

containing dC.[55-57] The minimal length of ssDNA was found to be four nucleotides long according to a previous study.[58-59] The numbering convention used to denote the position of nucleotides in a ssDNA with reference to target dC is depicted in the **Figure 1.1** where, the target dC is at position 0, nucleotides 5' of the target dC are referred as -1, -2, etc. and those on 3'- of dC are given a positive number and referred as +1, +2, etc.

All A3 enzymes, except A3G, prefer a thymine immediately next to the target dC at the 5'-end (5'-T<u>C</u>), but A3G prefers cytosines immediately before the target cytosine (5'-C<u>C</u>). A3 enzymes have also shown to deaminate the target dC when it preceded by other nucleobases, but with reduced efficiency. Sequence preferences of different A3 enzymes are summarized in **Table 1.1**.

Table 1.1. Sequence preferences of different A3 enzymes.[52]

Family member	Sequence specificity, 5'-3'
AID	(A/T)(A/G)<u>C</u>
A1	(C/T)T<u>C</u>
A3A	(T/C)<u>C</u>
A3B	(T/C)<u>C</u>
A3C	(T/C/G)<u>C</u>
A3DE	(A/T)<u>C</u>
A3F	(T/G)<u>C</u>
A3G	(C)<u>C</u>
A3H	(T/C)<u>C</u>

Binding studies performed on A3A revealed the preference for pyrimidines at the -2 position regardless of the nucleotide at +1 and this preference was validated by real-time NMR deamination assays to show the importance of nucleotides at -2 and +1 positions relative to the target cytidine and the optimum length of the ssDNA for the effective deamination of the target cytidine.[60] A similar pattern of substrate preference and optimum length of sequence was observed for A3B, where the study elucidated the importance of adenosine at the +1 position along with various different combinations of nucleotides at the -1 and -2 positions by a real-time NMR deamination assay.[58]

1.3.3. Structural basis for ssDNA-APOBEC3 interactions

A3 enzymes were thought to be RNA deaminases when they were first studied, but later it became evident that A3 enzymes deaminate ssDNAs.[61] Interestingly there is no evidence of A3s deaminating double-stranded DNAs.

Previous research has indicated that a minimum of four to five nucleotides are required for A3s to effectively bind and deaminate the target 2'-deoxycytidine.[58, 60] Co-crystals of ssDNA with A3 proteins enlighten the structural preferences and how the nucleotide sequence alters the specificity among the protein family.[36, 62] It is evident that in co-crystals of a catalytically inactive A3A and ssDNA substrate (5'-AAAAAAAT<u>C</u>GGGAAA) the ssDNA adopts a U-shape upon binding to the enzyme. The target cytosine (underlined) and the nucleotide immediately next to it at the 5'-end (-1 position compared to the cytosine at 0 position) are anchored at the active site and the rest of the nucleotides are flipped away from the active site.[36, 62] The target dC is held in the active site projecting towards the catalytically conserved Zn^{2+} and glutamate facilitating deamination. It is also evident that dC and 5'-dT are stabilized by van der Waals forces and hydrogen bonding of the phosphate backbone by Tyr130 and Asn57 in loop 3. It has been also noted that the His29 also stacks with the +1 base (immediately next to the target cytidine at the 3'-end) and makes van der Waals interactions with the -2 nucleotide. The rest of the nucleobases stack above the +1 base as in double-stranded DNA (**Figure 1.6a** and **b**).[36]

Interestingly, the conformation of ssDNA in the co-crystal structure of a 7-mer ssDNA-inactive $A3B_{CTD}$-A3A loop 1 chimera ($A3B_{CTD}$* in which the loop 3 is truncated and loop 1 of $A3B_{CTD}$ is swapped with loop 1 of A3A) assumed a similar conformation as that in the inactive A3A. The T at the -1 position makes van der Waals interactions with Trp281 (loop 5), Tyr313 and Tyr315 (loop 7) making it sit in the hydrophobic pocket near the active site. It also makes important hydrogen-bond interactions with Asp314 and Tyr315 (loop 7). The target dC is hydrogen bonded to Tyr313 (loop 7), Ala254 (α2), Ser282 (loop 5), Thr214

(loop 1) positioning it deep into the active site, containing the catalytic glutamic acid (E255 in α2). It is clear from the crystal structures of both A3A and $A3B_{CTD}$* (**Figure 1.6**) that most of the interactions ssDNA makes with the protein are essentially identical.[36]

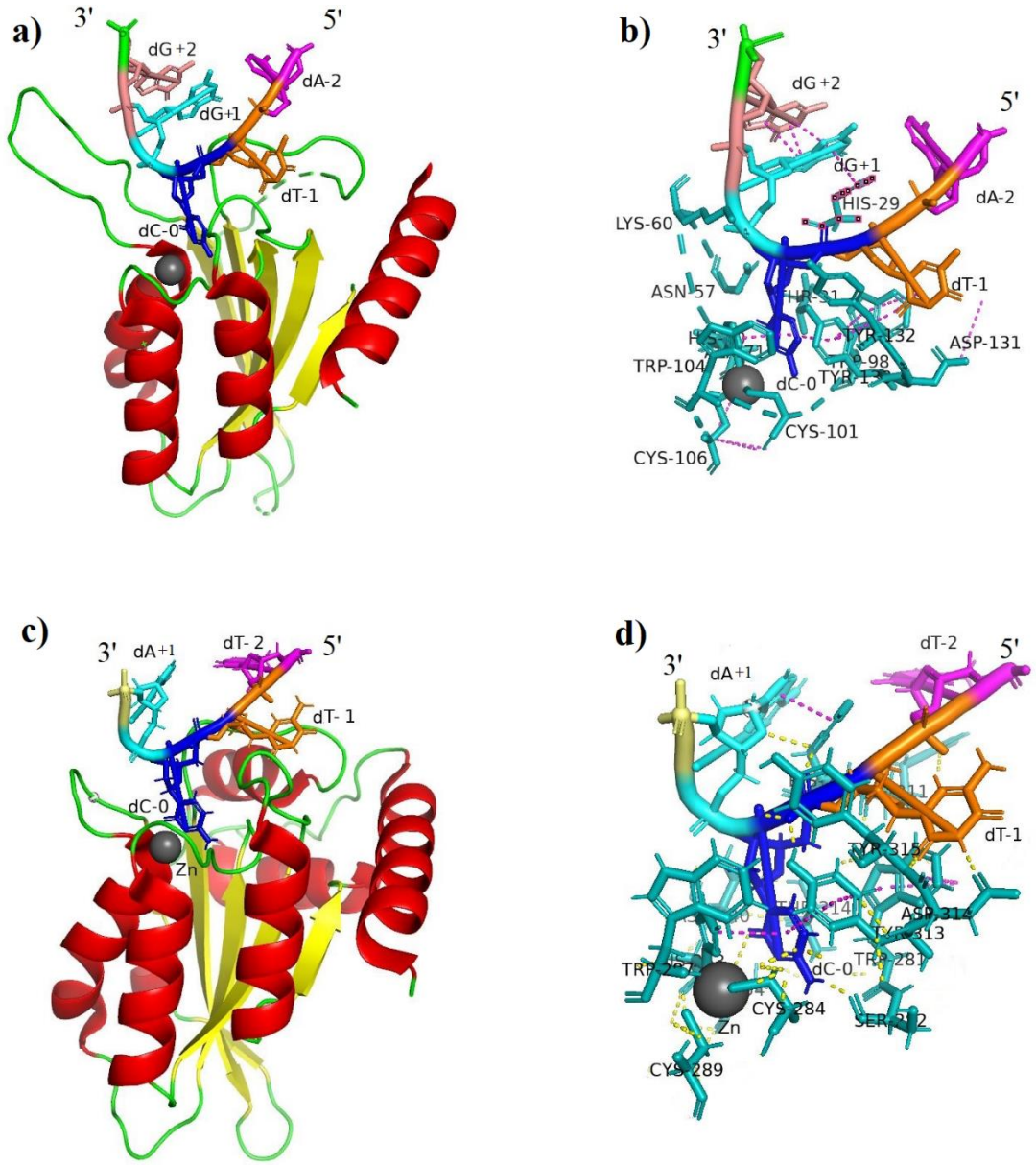


Figure 1.6. Crystal structures of human A3A and A3B-ctd bound to U-shaped ssDNA.

A) Crystal structure of of A3A–ssDNA complex showing target dC and dT-1, dA-2, dA+1 and dG+2 nucleotides, and the overall U-shaped conformation of the ssDNA. B) Nucleotide interactions with different amino acid residues in the protein (PDB ID: 5SWW). C) Crystal structure of of A3B-ctd–ssDNA complex showing target dC and dT-1, dT-2, and dA+1 nucleotides, and the overall U-shaped conformation of the ssDNA. D) Nucleotide interactions with different amino-acid residues in the protein (PDB ID: 5TD5). The structures are rendered in PyMol (2.5).

The co-crystal of catalytically inactive $A3G_{CTD}$ with a 9mer ssDNA (5'-AATCCCAAA) has recently been reported with the ssDNA adopting U-shape like A3A and $A3B_{CTD}$ (**Figure 1.7**).[63] The authors have reported important π-π interaction of T-3 with Trp211. The pyrimidine carbonyl group of C-2 forms a water-mediated hydrogen bond with Asp316 and Arg374. It also forms a hydrophobic interaction with Trp211. Due to this interaction, a cytosine is favoured at the -2 position in ssDNA. The carbonyl group of C-1 forms a hydrogen bond with Asp317, the N3 atom of this nucleobase forms a hydrogen bond with the amino proton of Gln318, and the amino group forms a hydrogen bond with the carboxyl group of Asp316.

The target cytosine (C0) has a π-stacking interaction with aromatic H257 as well as a T-shaped π–π interaction with Trp315. The amino group of C0, forms key hydrogen bond interactions with the carbonyl group of Ser286 and Glu259 through a water molecule coordinated by Zn^{2+}, which is the key to deamination by attacking the C4 of cytosine.

A+1 nucleobase forms π stacking against the His216. π–π stacking of histidine with a purine ring is stronger than with a pyrimidine ring, which explains the preference of purines in the +1 position. The 5'-phosphate group forms hydrogen bonds with His216 and Asn244.

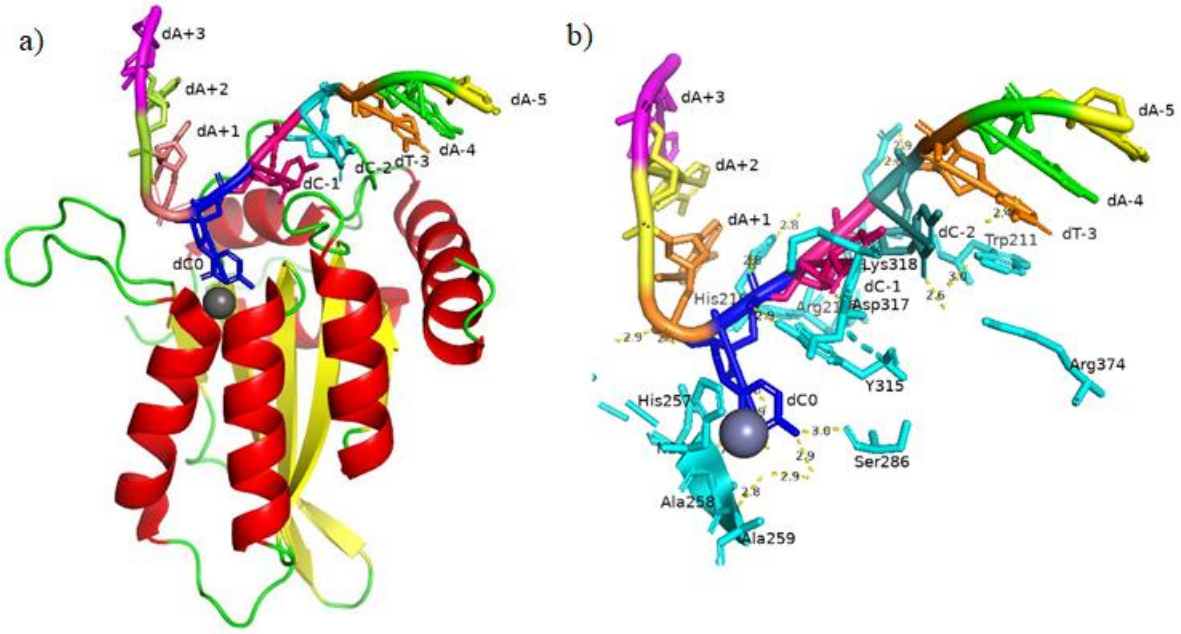


Figure 1.7. Crystal structure of $A3G_{CTD}$ bound to U-shaped 9mer ssDNA.

a) Crystal structure of of $A3G_{CTD}$–ssDNA complex showing target dC deep into the active site, and the overall U-shaped conformation of the ssDNA. b) Nucleotide interactions with different amino acid residues in the protein (PDB ID: 6BUX). The structure is rendered in PyMol (2.5).

1.4. Biological role of A3 enzymes

In normal cells, A3 enzymes function as a line of defence by combating a range of exogenous viruses and endogenous retroelements. However, overexpression and

upregulation of these enzymes can have detrimental effects, especially in cancer.[7-9, 27-28, 64-67]

1.4.1. A3s and their role in restricting viruses

In the early days of discovery, A3G was widely studied as a machinery for the defence against HIV-1. Other family members such as A3D, A3F, and A3H were also found to restrict HIV.[42, 68-69] A3 enzymes are known to restrict viruses by deamination-dependent and deamination-independent mechanisms.[70-72] Following virus infection, the viral genome undergoes reverse transcription with the help of the enzyme reverse transcriptase. Studies have shown that A3s bind to this enzyme to block the reverse transcription in a deaminase independent pathway and thereby inhibiting the viral replication.[70-73] In the deaminase-dependent pathway, the A3 enzymes catalyze deamination of C to U in viral cDNA resulting in scrambling of viral genome.[74-75] It is reported that all seven A3s can inhibit HIV-1 but the role of A3G is evident as it is primarily a cytoplasmic enzyme to combat the virus. However, HIV has evolved a mechanism, namely virion infectivity factor (Vif) protein, to bypass A3-mediated inhibition.[76-77] Vif inactivates A3s by ubiquitination leading to proteasomal degradation.[78-80] It is reported that in the absence of Vif, A3s mutate more than 10% of the HIV-1 genome during the reverse transcription phase.[17]

A3 enzymes are also involved in restricting a wide variety of DNA viruses such as HPV, HBV, and adenoviruses.[20-21, 69] Following infection, the DNA viral genome is transcribed into pre-genomic RNA, which is reverse transcribed into a cDNA and to a partial double-stranded DNA (dsDNA).[81] A3 enzymes are found to deaminate viral DNA during the replication cycle when the DNA becomes ssDNA. Like HIV-1, DNA viruses are also restricted by deaminase-independent mechanisms.[82]

1.4.2. Role of A3s in cancer evolution and drug resistance

A3 as a machinery in the innate immunity and defence mechanism works primarily by the mutagenic nature of the enzyme in C to U editing. Although A3s are powerful machinery working as part of the immune defence system, they are considered as double-edged swords due to their ability to mutate host genomic DNA. The C to U editing in T<u>C</u>A and T<u>C</u>T trinucleotide motif is copied to C to T transitions or sometimes the U bases can be removed by uracil DNA glycosylase enzyme (UDG), producing abasic sites in DNA which often result in C to T and C to G mutations.[83] Accumulation of these mutations over a period can lead to cancer initiation and evolution. Next-generation sequencing identified a range of clustered mutations suggesting A3-induced base substitutions in cancer genomes.[66] These clustered mutations were prevalent in certain cancers and were observed on the A3 preferred

TCA or TCT motifs. These mutations are at least partly a result of DNA strand break repair producing ssDNA which is an ideal substrate for A3 enzymes.[66, 84] Studies revealed by whole genome sequencing of different cancers including breast cancer, multiple myeloma, head, and neck cancer, etc. proved a high load of clustered mutation by A3s.[66, 84-86] The role of A3s as in mutation driver mechanism and mutation signatures associated with development of drug resistance leading to poor clinical outcomes has been extensively studied recently.[87-90] The upregulation of A3s predominantly by A3A and A3B was correlated with HPV infections, and therefore the high A3 mutational signature in HPV-induced cancers.[91-92]

Several studies revealed the specific role of A3B in a variety of cancers which is supported by the fact that it is predominantly nuclear.[27, 93] Research on different cancer cell lines including breast cancer cell lines showed mutations within the 5'-TC and 5'-TCA motifs to be attributed to A3B.[9, 27] A fluorescence-based DNA cytidine deaminase assay revealed the DNA editing activity and absence of it in the A3B knock-down experiments.[27] The relation between the expression of A3B and breast cancer evolution including development of drug resistance has been established by studying estrogen receptor-positive (ER+) groups of patients.[67, 87] The development of A3B-mediated resistance to tamoxifen therapy in ER+ breast cancer has been elucidated using xenograft experiments using MCF-7L cells injected into a group of A3B knock-down immune-deficient mice and an A3B-intact control group.[67] From each group, a random number of mice was given tamoxifen injections after 50 days of engraftment. The result revealed a suppression of tumor growth in tamoxifen-treated A3B knock-down mice, and the tamoxifen-treated A3B-intact mice developed resistance to the tamoxifen therapy and developed large tumors. This experiment thus proved the development of drug resistance to tamoxifen in ER+ breast cancer is driven by A3B.[67] Although A3B is the highly expressed enzyme among the A3 family in tumors, the most active deaminase is A3A. There is still a dispute over which of them is the major contributor to cancer evolution. A recent study revealed that a significant amount of deamination activity in breast cancer was induced by A3A.[94] Sequencing of human breast cancer (BRCA) tumors revealed that there is a high load of A3-induced mutation even though deletion of exon 5 of A3B is common in BRCA.[95] This suggested that the mutations are caused by other A3s and the recent literature proved that A3A is the major cause of mutations in BRCA which was determined by qRT-PCR on the expression of different A3 enzymes by analyzing a series of BRCA cell lines.[94] Therefore, A3A-driven mutations are predominant in BRCA tumors. It was also established that passenger hotspot mutations in hairpin loops are mostly driven by a high affinity of A3A towards these substrates, once again proving the importance of the U-shaped conformation of the ssDNA substrate in the enzyme.[96] This unusual conformation

of the ssDNA can thus be used for the rational development of inhibitors which can inhibit A3A and A3B. The inhibitors thus developed can be used as conjuvants along with an existing therapy to prolong the therapeutic benefit of the agent by halting cancer evolution and delaying the development of drug resistance.

1.5. Modulating the activities of A3 enzymes as a therapeutic target

Now that the clinical significance of A3 enzymes is well elucidated, exploring the potential of this knowledge either to inhibit these enzymes (therapy by hypomutation) or using the natural mutational activity of these enzymes (therapy by hypermutation) can be employed in anticancer and antiviral therapies.[97] Antiviral therapy, especially the ability of A3s to mutate and degrade HIV-1 in the absence of Vif, is one of the important aspects of therapy by hypermutation. One of the obvious strategies, therefore, would be the inhibition of Vif to keep the A3 enzymes, especially A3G, protected from Vif-mediated degradation.[98] Several small-molecule inhibitors were developed over the past years as inhibitors of Vif to combat HIV-1 but this methodology is still in its initial stages and requires more research.[99-101]

Even though HIV-1 can counteract the A3-mediated degradation through Vif, the virus effectively exploits residual A3G activity to mutate itself and thereby to counteract existing retroviral therapies.[78, 102] There is an increased signature of A3-catalyzed GG-AG and GA-AA mutations in HIV-1 patients suggesting the virus mutates itself by exploiting the sub-lethal levels of A3s in the event of HIV-1 infection.[103-104] It has been established that several resistance signatures against antiretroviral therapies are attributed to A3-catalyzed deamination.[105-107] So, it has been proposed that therapy by hypomutation, which is the direct inhibition of A3 enzymes, can be beneficial to counteract the capability of viruses to exploit A3s in developing drug resistance. Once this pathway is blocked, the combination of host immune defence mechanisms and antiretroviral therapy can be effectively combined to treat an infection.

As the involvement of A3A and A3B in genomic mutations, leading to development of cancer and drug resistance, was established over the past years, therapy by hypermutation seems to be one of the options but the high mutational load at lethal levels to kill a cancer cell seems problematic as it can induce secondary cancers. The strategy of therapy by hypomutation seems more viable if we can develop an inhibitor which can block the mutational effects of these enzymes. There might be an argument that the inhibition of these enzymes can have unwanted effects in the body as these enzymes are a part of human immune system. However, it has been established that A3B deletion is prevalent in Oceanic

populations (92.9%), and this shows the non-essential nature of this enzyme. The above study is a crucial evidence suggesting that the inhibition of A3B shall not affect the natural defence system of the body. A3A is not a part of the primary metabolism of the body and therefore can also be inhibited. As the drug resistance towards chemo-and immunotherapeutic agents in standard cancer therapy is partially attributed to A3s, effective inhibitors of these enzymes can thus be co-administered as a conjuvant with the established cancer therapeutic agents. This combination, therefore, will increase the efficacy of the therapy making the therapeutic agent work for a longer period of time.

1.6. Small molecule inhibitors of A3G

A3G-catalyzed deamination is exploited by HIV-1 to evade the human immune system and thus therapy by hypomutation can be beneficial to extend the efficacy of existing therapies. The small-molecule agent MN30, which is the first in class small molecule inhibitor for A3G, was identified by high throughput screening (**Figure 1.8**).[108]

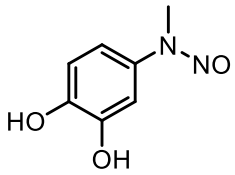


Figure 1.8. Chemical structure of first in class A3G inhibitor MN30.

It was shown by crystallographic and mass-spectrometric data that this inhibitor covalently binds to Cys321 of A3G in the CTD near the active site by auto-oxidation to an ortho-quinone form. Although this inhibitor binds near the active site, the substrate binding is unaffected. Even though the substrate binding is unaffected, MN30 bound to the enzyme causes a flip of Tyr315 and results in the substrate losing contact with Trp285, an important active-site residue.[108] This conformational change upon binding of this small-molecule inhibitor renders the enzyme catalytically inactive, despite binding of the DNA substrate.

Another small molecule inhibitor based on 4-amino-1,2,4-triazole-3-thiol scaffold (MN256.0105) was also reported (**Figure 1.9**). This inhibitor contains a thiol moiety that binds with free surface cysteine.[109] This work was based on the previous observation that MN30 binds to Cys321 to inhibit A3G.[108] One should note that Cys321 is absent in A3A and A3B, and thus these inhibitors are of little use for A3A and A3B.

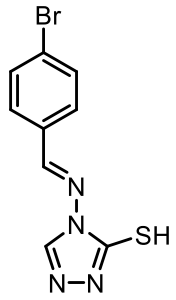


Figure 1.9. Chemical structure of MN256.01056.

1.7. CDA inhibitors

CDA is an important Zn^{2+}-dependent deaminase enzyme involved in the pyrimidine salvage pathway, where it catalyzes the deamination of cytidine and 2'-deoxycytidine to form uridine and 2'-deoxyuridine, respectively.[110] This enzyme catalyzes deamination by the formation of a high-affinity hydrated transition-state intermediate facilitated by Zn^{2+}-bound water followed by deamination. Even though cytidine and 2'-deoxycytidine are the natural substrates of CDA, it can deaminate other synthetic cytidine analogues. Some of these agents are clinically used in treating cancer and viral infections such as cytarabine and Decitabine (**Figure 1.10**).[111-113]

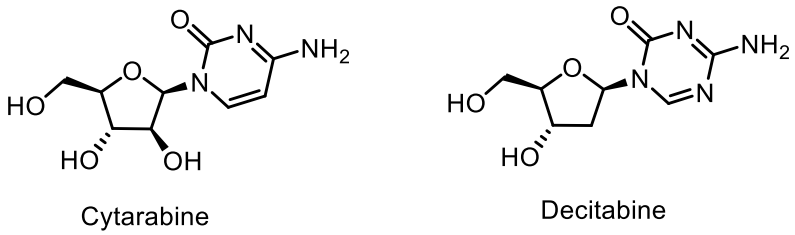


Figure 1.10. Chemical structures of synthetic cytidine analogues.

Deamination of these synthetic cytidine analogues by CDA therefore causes loss of therapeutic activity and leads to undesired side effects.[114] Another important observation is that CDA is often over-expressed in liver and spleen of males and the deamination of these drugs by CDA leads to less than 20 minutes of lifetime in the body.[115] Several efforts have been made by the scientific community to address the above-mentioned issues by developing CDA inhibitors so that they can be used as adjuvants with the existing therapy. These adjuvants increase the half-life of the original therapeutic agent (cytidine analogue) by inhibiting CDA mainly in the liver, and thereby benefiting the therapy.

Several CDA inhibitors have been described in the literature, of which zebularine is a classic example (Z).[116] A crystal structure of zebularine bound to CDA has revealed the formation of a hydrated species at the active site, which further confirms the mechanism how these inhibitors block the catalytic activity of CDA (**Figure 1.11**).[117]

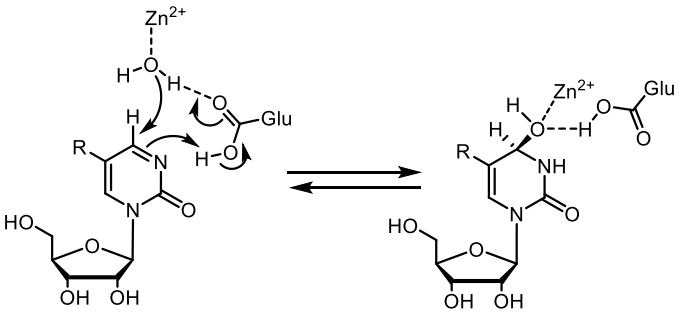


Figure 1.11. Zebularine (Z) acting as a transition state analogue upon binding to the active site of CDA enzymes.

Some of the CDA inhibitors described in the literature along with their inhibition constants against mouse kidney CDA are given below (**Figure 1.12**).[116, 118-123]

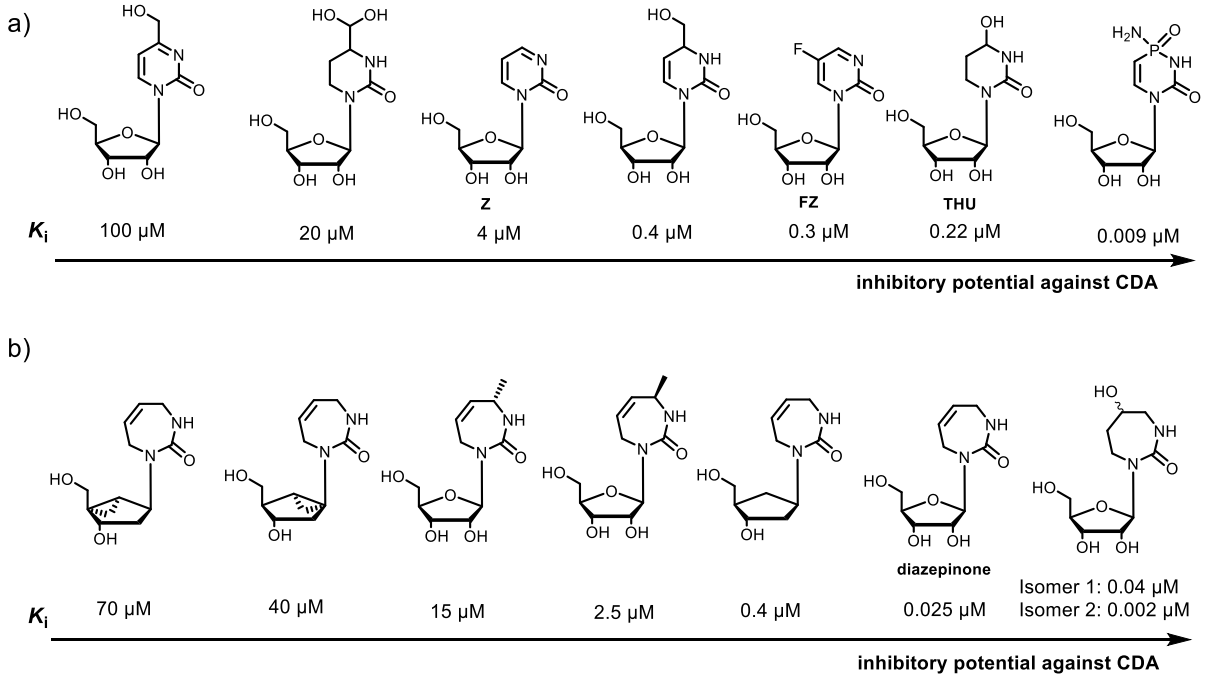


Figure 1.12. Chemical structures of CDA inhibitors reported in the literature and their inhibition constants (K_i) against mouse kidney CDA.

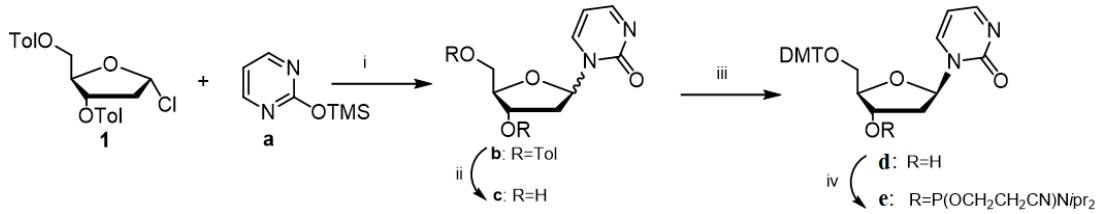
a) Compounds based on six-membered nucleobases. b) Compounds based on seven-membered nucleobases.

1.8. Oligonucleotide-based inhibitors of A3A, A3B, and A3G

As discussed earlier, deamination of cytosines in ssDNA by A3 enzymes is believed to be like that of CDA. Recently, oligonucleotides containing the known transition-state analogues of cytosine deamination were reported by our group as competitive inhibitors of A3 enzymes.[54, 124] These transition-state analogues, when incorporated in an A3 preferred sequence of ssDNA in place of the target cytosine (C0), are delivered into the active site of A3 enzymes resulting in inhibition.

1.8.1. Synthesis of 2'-deoxyzebularine (dZ) and 5-fluoro-2'-deoxyzebularine (5FdZ) phosphoramidites

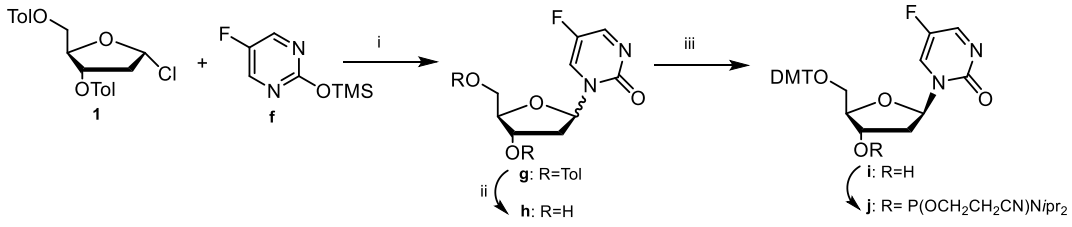
The synthesis of dZ and FdZ nucleosides and their phosphoramidite derivatives were reported previously by our group. [54, 124] The synthesis of dZ nucleoside and its phosphoramidite (**Scheme 1.1**) started with the coupling of a pre-silylated pyrimidine-2-one (**a**) on to the Hoffer's chloro sugar (**1**).[125] The coupling produced the protected nucleoside (**b**) with 12:88 α:β ratio. The toluoyl deprotection and selective protection of the 5'-hydroxyl group with 4,4'-dimethoxytrityl chloride produced the compound **d** in 54% yield. At this point the required β isomer was separated and was phosphitylated to yield the dZ phosphoramidite.



Scheme 1.1. Synthesis of dZ nucleoside and phosphoramidite derivative for incorporation into ssDNA.

Reagents and conditions: (i) $CHCl_3$, distillation, 10 min, 54%, 12:88 α:β (**b**); (ii) 28% aqueous ammonia, MeOH, 48 h (**c**); (iii) 4,4'-dimethoxytrityl chloride, pyridine, 0º to r.t., overnight, 54% (**d**); (iv) *N*,*N*-di*iso*propylamino-2-cyanoethoxychlorophosphine, Et_3N, CH_2Cl_2, 30 min, 88% (**e**).

For FdZ synthesis, the coupling reaction between Hoffer's chloro sugar (**1**) and a pre-silylated 5-fluoro-2-hydroxypyrimidine was performed in the presence of the Lewis acid $SnCl_4$ at -35º C. The mixture of anomers produced (1:9 α:β) were separated by silica gel column chromatography. The toluoyl groups on the pure β anomer were cleaved by using saturated ammonia in MeOH. Selective protection of the 5'-hydroxyl group in the compound **h** with 4,4'-dimethoxytrityl chloride followed by phosphitylation produced FdZ phosphoramidite in 52% yield over two steps (**Scheme 1.2**).



Scheme 1.2. Synthesis of FdZ nucleoside and phosphoramidite derivative for incorporation into ssDNA.

Reagents and conditions: (i) CH_2Cl_2, $SnCl_4$, -35° C, 10 min, 87%, 1:9 α:β (**g**); (ii) sat. ammonia in MeOH, 48 h (**h**); (iii) 4,4'-dimethoxytrityl chloride, pyridine, 0° to r.t., overnight, 60% (**i**); (iv) *N*,*N*-di*iso*propylamino-2-cyanoethoxychlorophosphine, Et_3N, CH_2Cl_2, 30 min, 89% (**j**).

Oligonucleotides were synthesized using standard phosphoramidite chemistry and the oligos were tested for inhibition of A3 enzymes using ^{1}H NMR assay. There was a clear difference in the inhibitory potential between the two oligos containing 2'-deoxyzebularine (dZ) and 5-fluoro-2'-deoxyzebularine (5FdZ): i.e., an oligo containing 5FdZ in a TCA motif (ATTTXATTT, where X= FdZ or dZ) exhibited an inhibition 3.5 times better than that of a dZ-containing oligo (**Figure 1.13**).

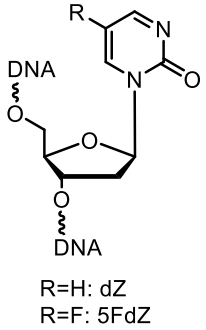


Figure 1.13. Transition-state analogues of cytosine deamination dZ or 5FdZ incorporated in ssDNA as inhibitors of A3 enzymes.

Inhibition of wild-type A3A was also demonstrated by a fluorescence-based activity assay.[124] Although the preferred motif for deamination of cytosine for A3A and A3B is TCA, these enzymes can also deaminate cytosine in ssDNA containing other nucleosides, but with lower efficiency.[31-32] Both A3A and A3B are also capable of deaminating dC in the CCC motif. A3G and A3A are known to deaminate dC at the 3′-end CCC at least 100 times and 5 times faster relative to other cytosines respectively. On the contrary, A3B deaminates cytosine at the 5′-end of CCC motif with at least 6 times faster initial rate of deamination for this dC in comparison to other cytosines.[31-32] These observations have led to the development of oligonucleotide-based inhibitors replacing dC by dZ at the 5′ end of the CCC motif (dZCC-oligo) as a selective inhibitor of A3B, and dZ at the 3' end of CCC (CCdZ-oligo) as an A3A or A3G inhibitor.[126]

1.9. Approach of this study

A clear relationship between the over-expression of A3 enzymes (mainly A3A and A3B), and evolution of cancer demands new therapeutic agents and treatment regimens. Therapy by hypomutation is clearly one way to address mutagenesis driven by A3 enzymes leading to improved treatment outcomes. It has been established that A3B is not an essential enzyme, and its deletion polymorphism is prevalent in some populations.[127] Therefore, the selective inhibition is not harmful to the normal defence mechanisms of the body. This makes A3B

inhibition a primary target to address mutagenesis in many cancers. Although there is still uncertainty around the long-term effects of A3 inhibition, we believe the non-essential nature of these enzymes in primary metabolism provides a window for therapeutic intervention.

1.10. Hypothesis of the study

Our previous observations of improved inhibition potential of a linear ssDNA by incorporation of a transition state analogue of cytidine deamination (5FdZ and dZ) was utilized as a starting point for this work.[54, 124] These studies provided us a methodology to deliver the transition-state analogue embedded in a ssDNA into the active site of A3 and block the enzyme. The above observations together with the available crystal structures of ssDNA bound to A3 enzymes inspired us to mimic the U-shaped conformation of ssDNA by synthesizing a cross-linked ssDNA. The cross-linked dC0-containing ssDNA could become a better substrate due to entropic and enthalpic advantage over the linear ssDNA substrates. Use of dZ or FdZ instead of dC0 in a cross-linked DNA can lead to a powerful A3 inhibitor. The inhibitors synthesized and characterized could potentially be used as a conjuvant in cancer chemotherapy.

1.11. Aims of this study

(i) To develop a cross-linking strategy for ssDNAs as substrates and inhibitors of A3 enzymes.

(ii) To synthesize and characterize the cross-linked ssDNAs as substrates and inhibitors of A3 enzymes in comparison to a standard linear inhibitor developed previously.

(iii) Finally, to synthesize and characterise new CDA inhibitors as transition-state analogues of cytosine deamination and test the inhibition of A3 enzymes when these modified nucleosides are incorporated into a preferred ssDNA sequence.

Chapter 2. Methodology and Methods

2.1. Methodology to develop cross-linked DNAs

To test our hypothesis, we decided to evaluate two different approaches for pre-shaping DNA using covalent cross-links between distant nucleotides in the sequence, i.e., terminal, and internal cross-linking strategies (**Figure 2.1**). For pre-shaping ssDNA (both strategies), we used copper(I)-catalyzed azide-alkyne cycloaddition (CuAAC), which is compatible with DNA synthesis and does not change the chemical structure of native nucleotides.[128-131] For the terminal cross-linking strategy, terminal alkyne and azide moieties were incorporated at 5' and 3'-ends of ssDNA, respectively. Commercially available alkyne **Y^{132}** (**Figure 2.1A**) was used and in-house synthesized azido-sugar **dR^{N3}** was attached to the solid support (controlled-pore glass, CPG). ssDNAs of different lengths were synthesized by standard DNA phosphoramidite chemistry. The second strategy used detailed structural information from the A3-ssDNA complexes (**Figure 1.6**).[36] Relative to the central 5'-TCA motif of the ssDNA, the positions of the alkyne and azide for the internal cross-linking strategy were selected based on the observation that 2'-deoxyadenosine at the +1 position immediately next to the target dC (position 0) and thymidine at the -2 position are close together and form hydrogen bond interactions with each other through a water molecule in the bound ssDNA in A3B$_{CTD}$ and A3A crystal structures.[36] This makes dA in the +1 position and dT in the -2 position suitable for cross-linking (**Figure 2.1C**).[36, 60, 65, 133] We employed in-house synthesized 2'-deoxyadenosine derivatives (**dA^{Y2}**, **dA^{Y4}**, **dA^{N3}**) or *N*-carbazolyl nucleoside (**dH^{E}**) to place them in position +1 and commercially available phosphoramidite of 5-ethynyl-2'-deoxyuridine[134] (**dU^{E}**, **Figure 2.1F**) or in-house synthesized phosphoramidites of azido-sugar **dR^{N3}** or modified 2'-deoxycytidine (**dC^{N3}**) to mimic dT in position -2 for the internal cross-linking strategy.

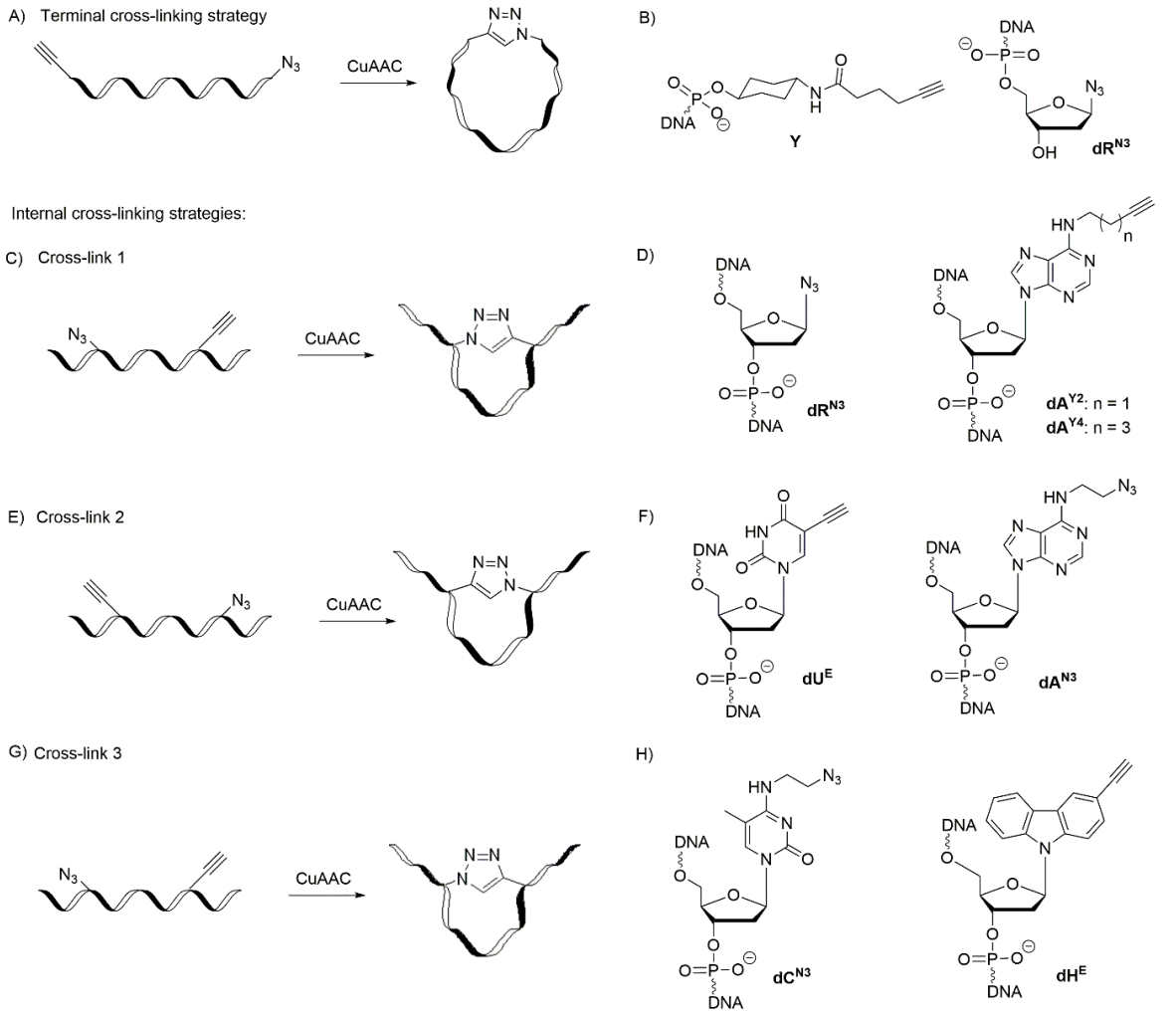


Figure 2.1. Different cross-linking strategies and nucleoside modifications used in this study.

Terminal (A, B) and internal (C-H) cross-linking strategies by Cu(I)-catalyzed azide-alkyne cycloaddition (CuAAC) to pre-shape a linear ssDNA as a substrate (dC-containing oligo) and as an inhibitor of A3B (dZ or FdZ-containing oligo). (B) Alkyne (Y) and azide (dR^{N3}) modifications used in the sequence for terminal cross-linking strategy. (C, E and G) Internal cross-linking strategies as well as alkyne and azide modifications used for creation of a 1,4-disubstituted 1,2,3-triazole cross-link (D, F, and H, respectively)

After the oligonucleotide synthesis, for terminal and internal cross-links, we employed CuAAC as a post-synthetic step, monitored by analytical reverse-phase HPLC and mass spectrometry. We also performed NMR experiments to prove the cross-linking reaction was successful by comparing a selected sequence before and after cross-linking. The cross-linked oligonucleotides (oligos) were then evaluated as A3 substrates and compared to a standard linear substrate by an NMR-based assay. Real-time NMR assays are advantageous because they are direct, utilizing only A3 enzymes and oligos in a suitable buffer system, unlike many fluorescence-based assays where a secondary enzyme i.e., uracil deoxyglycosidase (UDG) and a fluorescently modified oligo are used. In this study it is also questionable if the UDG can recognise cross-linked oligonucleotides, whereas in the ^{1}H NMR assay the aromatic proton (H-5) of the substrate dC and the product dU can be easily identified. The NMR-based assay directly yields the initial velocity of deamination of various ssDNA substrates,

including the modified ones, in the presence of A3 enzymes and consequently the Michaelis–Menten kinetic model is used to characterize substrates and inhibitors of A3. Therefore, currently, the NMR-based direct assay is best suited to study modified ssDNA-based substrates and inhibitors of A3. We can also monitor the disappearance of cytosine and/or the appearance of uracil as a function of time over the entire course of the reaction.

The A3 enzyme used in our experiments is the well-characterized and very active $A3B_{CTD}$-QM-ΔL3-AL1swap,[54] where loop 3 is deleted and loop 1 is replaced with the corresponding loop 1 from A3A in addition to four point mutations remote to the enzyme's active site. This is the same enzyme used in the crystal structure of the A3-DNA complex (except that for the X-ray structure the active-site Glu255 was mutated to Ala to prevent deamination of the substrates i.e., dC-containing ssDNA).[36] Our recent work on A3G has shown that inhibition of the C-terminal catalytic domain is mirrored by the inhibition of full-length enzyme, validating our approach to use catalytic C-terminal domains for inhibitor development.[126] The best ssDNA substrates were converted into A3 inhibitors by incorporating **FdZ** or **dZ** in place of the target dC. The evaluation and optimization strategy are summarized in **Figure 2.2**.

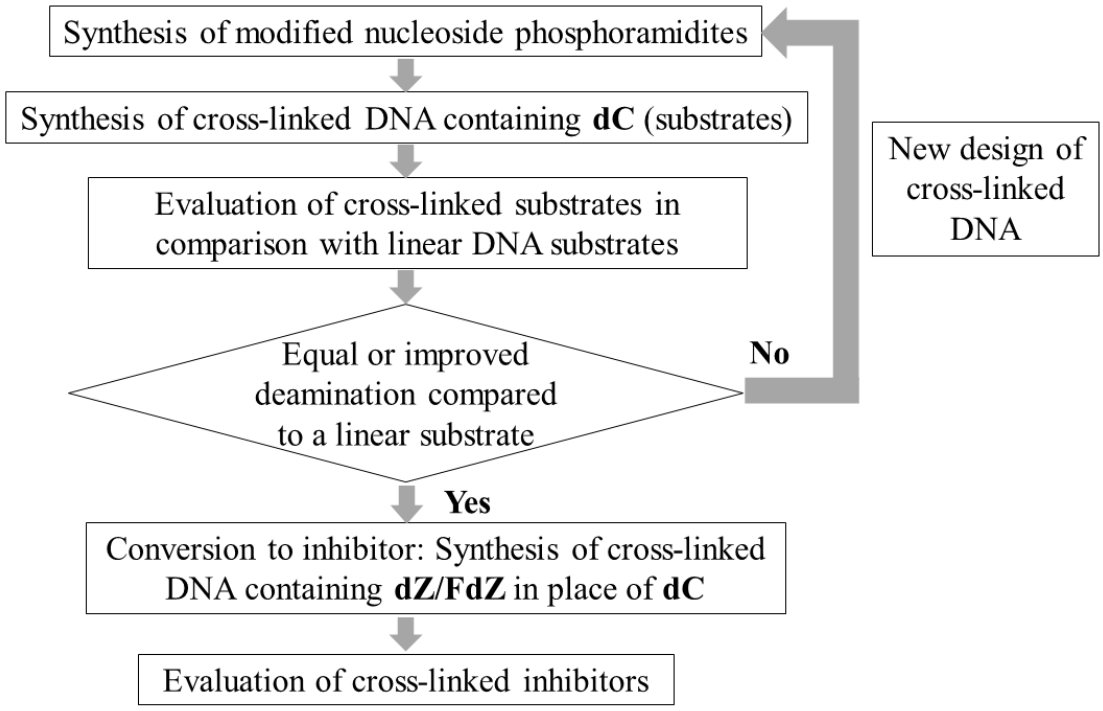


Figure 2.2. Methodology for the design and evaluation of cross-linked substrates and inhibitors.

We also attempted to characterize the thermodynamics of binding of selected cross-linked oligos with A3 enzymes in comparison with linear ssDNA by isothermal titration calorimetry (ITC). The inactive E72A mutant of A3A (the same as in A3A-ssDNA crystal

structure, PDB: 5SWW) was chosen to evaluate substrate binding and the active $A3B_{CTD}$-QM-ΔL3-AL1swap was chosen for inhibitor binding.

2.2. Synthesis of the modified nucleoside phosphoramidites

Our method of choice for the synthesis of preshaped ssDNAs relies on modified nucleoside precursors containing terminal alkynes and azides so that they can be cross-linked by CuAAC. For terminal cross-linking strategy, alkyne and azido-modified nucleosides must be incorporated at the termini of a ssDNA. The alkyne was purchased as a phosphoramidite (**Y**) from commercial sources as it was intended to be incorporated in the 5' end and the azidosugar (**dR^{N3}**) was synthesized in-house and was embedded in a controlled-porous glass (CPG) support so that it can be incorporated on the 3'-end of the ssDNA.

The second strategy employed an internal cross-link and relied on the synthesis of a series of azide-and alkyne-containing phosphoramidites. The design of these cross-links was based on the available information from the crystal structures showing the importance of the positions -2 and +1 relative to the target cytosine.[36, 60] The proximity of the nucleobases at these positions prompted us to cross-link them. These cross-links were also sub-divided as a) cross-link 1, b) cross-link 2, and c) cross-link 3 according to the chronology of their synthesis. The detailed descriptions of their synthesis and characterization are described in **Chapter 3**.

2.3. Synthesis of modified linear ssDNAs using automated DNA synthesizer

Invention of solid-phase chemical synthesis by Bruce Merrifield in the 1960s was a great boom to both peptide and oligonucleotide synthesis.[135] The most important advantages of solid phase synthesis are the a large excesses of solution phase reagents that can be used, and impurities or side products formed can be washed away during the synthesis and the process can be well automated. Materials used as solid support include controlled pore glass (CPG) or polystyrene. The phosphoramidite method of oligonucleotide synthesis (**Figure 2.3**) was a big success since the process is transferable to suit automatic synthesis.

The first nucleoside of the sequence is attached to the solid support and the synthesis starts by detritylation of the 5'-*O*-dimethoxytrityl group of the first nucleoside on the solid support by a strong acid such as trichloro- or dichloroacetic acid. Next is the critical step of coupling with 5'-*O*-dimethoxytrityl nucleoside phosphoramidite using 5-ethylthio-1*H*-tetrazole as an activator. The unreacted 5'-hydroxyl groups are then capped by acetic anhydride activated by *N*-methyl imidazole in pyridine as capping reagent. This step is important for the overall purity of the product. The capping step ensures that unreacted oligos

are not elongated further and thus prevents accumulation of n-1 sequences, where n is the length of an oligonucleotide. Finally, oxidation of the phosphotriester to phosphate completes the synthetic step followed by deprotection of the full-length oligonucleotide from the solid support and removal of other protecting groups on individual nucleotides and phosphates (cyanoethyl groups) under basic conditions. Purification of the full-length oligonucleotide can be done by reverse-phase HPLC or ion-exchange chromatographic separation.

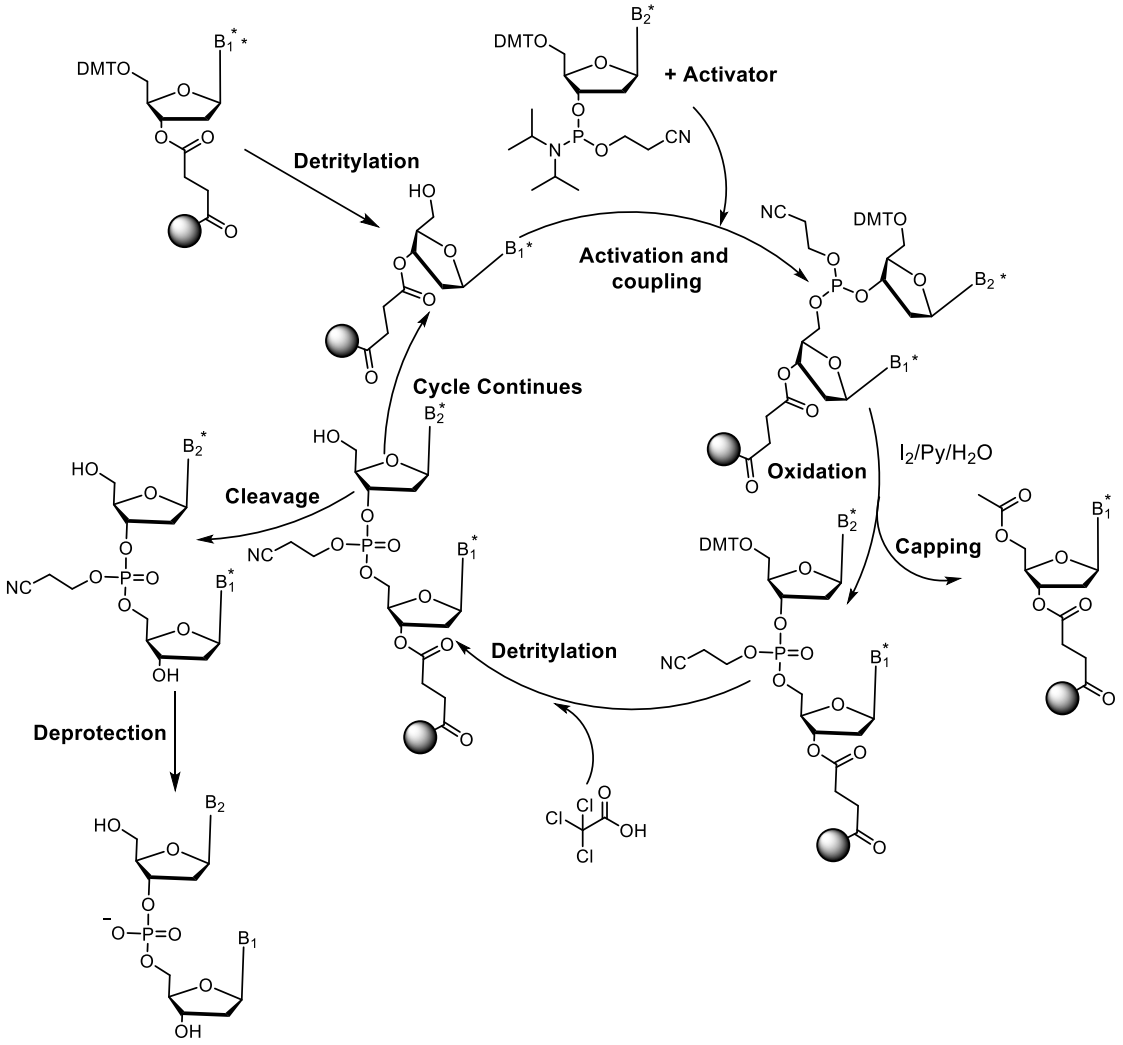


Figure 2.3. General steps and conditions of phosphoramidite method of oligonucleotide synthesis.

A series of oligonucleotides was synthesized with alkyne-and azido-modified nucleosides containing dC (as substrates) and dZ or FdZ (as inhibitors). We also synthesized a hairpin oligo with FdZ in the sequence to test the inhibition of A3 enzymes. The detailed procedures and lists of oligos are given in **Chapter 3**.

2.4. Cross-linking of linear ssDNAs by CuAAC

Copper(I)-catalyzed azide-alkyne cycloaddition (CuAAC) is a variation of the Huisgen 1,3-dipolar cycloaddition reaction producing di-substituted 1,2,3-triazoles from a terminal alkyne and an azide in the presence of copper(I). Copper-catalyzed cycloadditions produce 1,4-disubstituted products regioselectively.

The CuAAC reaction is well-suited for solid support synthesis and has been exploited very widely for peptide and oligonucleotide synthesis and in post-synthetic modifications.[136] One of the disadvantages of this methodology is the production of Cu(I) species in the reaction mixture, which produces reactive oxygen species (ROS) *in-vivo.*[137] Many efforts have been done in the past to stabilize the Cu(I) in solution by the use of ligands. Ligands (**Figure 2.4**) such as tris[(1-benzyl-1*H*-1,2,3-triazol-4-yl)methyl]amine (TBTA)[138] and water-soluble tris(3-hydroxypropyltriazolylmethyl)amine (THPTA)[139] are commonly employed in CuAAC. This reaction can be performed in aqueous solutions at room temperature with no alterations to the native nucleotides, which makes it an ideal choice for the modification of the DNA after synthesis (post-synthetic modification).

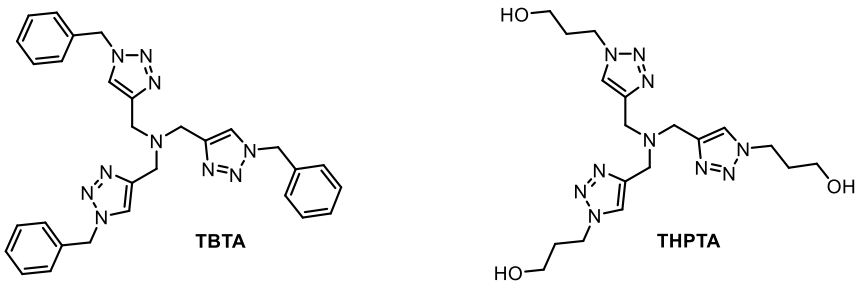


Figure 2.4. Chemical structures of TBTA and THPTA.

The versatility of CuAAC has made it possible to find applications in solid-phase peptide and oligonucleotide synthesis, nanotechnology, supramolecular chemistry, surface chemistry and biochemistry. Particularly with oligonucleotide synthesis, “click” ligation was very useful method to synthesize long oligonucleotides with high purity and stability.[129]

2.5. Evaluation of cross-linked ssDNAs by real-time ^{1}H NMR assay

The potential substrates and inhibitors synthesized were characterized in comparison with a standard linear 7-mer ssDNA by our previously reported real-time ^{1}H NMR assay. The NMR-based assay directly yields the initial velocity of deamination of various ssDNA substrates, including the modified ones, in the presence of A3 enzymes and consequently the Michaelis–Menten kinetic model can be used to characterize substrates and inhibitors of A3. Currently, the NMR-based direct assay is best suited to study modified ssDNA-based substrates and inhibitors of A3 as it utilizes only A3 enzymes and oligos in a suitable buffer

system, unlike many fluorescence-based assays where a secondary enzyme and a fluorescently modified oligo are used.

2.5.1. Kinetic characterization of active A3B$_{CTD}$ (A3B$_{CTD}$-QM-ΔL3-AL1swap)

Kinetic characterization of active A3B$_{CTD}$-QM-ΔL3-AL1swap was done using our previously described ^{1}H-NMR-based real-time deamination assay.[54, 124, 140] This assay monitors the real-time deamination of target dC in a 7-mer DNA oligo 5'-T$_4$**dC**AT to dU. Data acquisitions were done on a 700-MHz Bruker NMR spectrometer equipped with a 1.7 mm cryoprobe at 298 K. A series of ^{1}H NMR spectra was recorded of the substrate at various concentrations from 200 to 800 µM with 300 nM of A3B$_{CTD}$ -QM-ΔL3-AL1swap in buffer (pH 6.0) containing 50 mM sodium phosphate, 100 mM NaCl, 2.5 mM β-mercaptoethanol, 50 µM 3-(trimethylsilyl)-2,2,3,3-tetradeuteropropionic acid (TSP) and 10% D_2O. The H-5 proton doublet signal of the cytosine, which appears at 5.88 ppm (J = 7.7 Hz), was baselined and integrated (**Figure 2.5**). The signal of TSP at 0 ppm was used as an internal standard to determine the concentration of the substrate and converted to the product during the reaction, which appears at 5.71 ppm (J = 8.3 Hz) over a period of time.

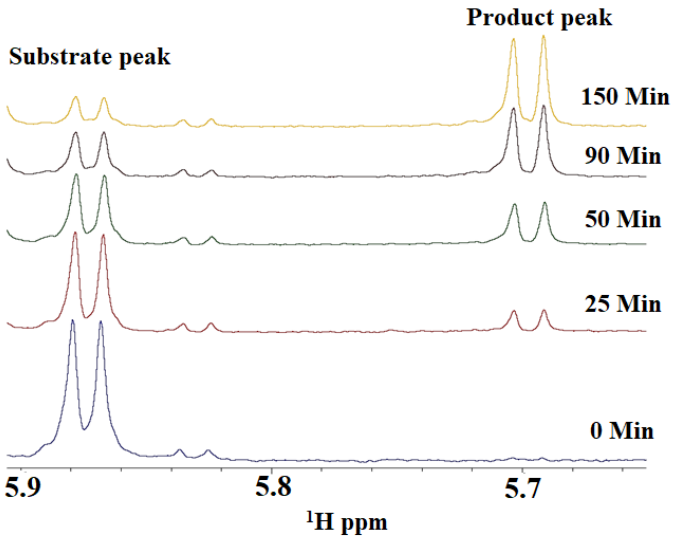


Figure 2.5. ^{1}H-NMR spectra for the substrate (5'-T$_4$dCAT) and its conversion to the product (5'-dT$_4$dUAT) over time.

The area of the integrated signal was converted to substrate concentration [S] and plotted versus time of the reaction (s) (**Figure 2.6**). Linear regression was used to fit the data to determine the initial speed of the reaction (V_0). The double reciprocal plot known as the Lineweaver-Burk plot is obtained by plotting $1/V_0$ against 1/[S] (**Figure 2.7**). The linear dependence of substrate concentration on the speed of deamination was then fitted with linear regression to determine Michaelis-Menten parameters K_m and k_{cat}.

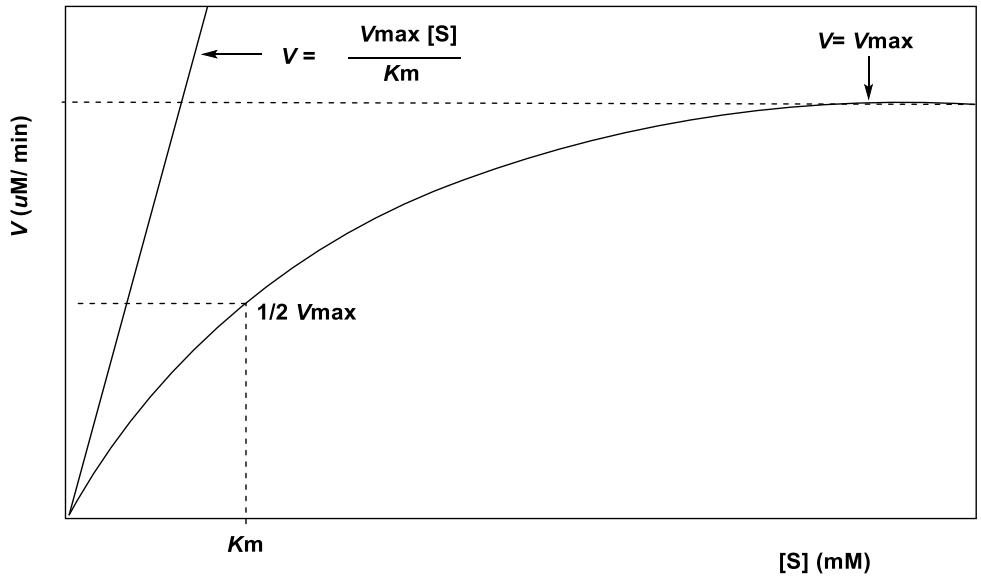


Figure 2.6. Graphical representation of Michaelis-Menten kinetics.

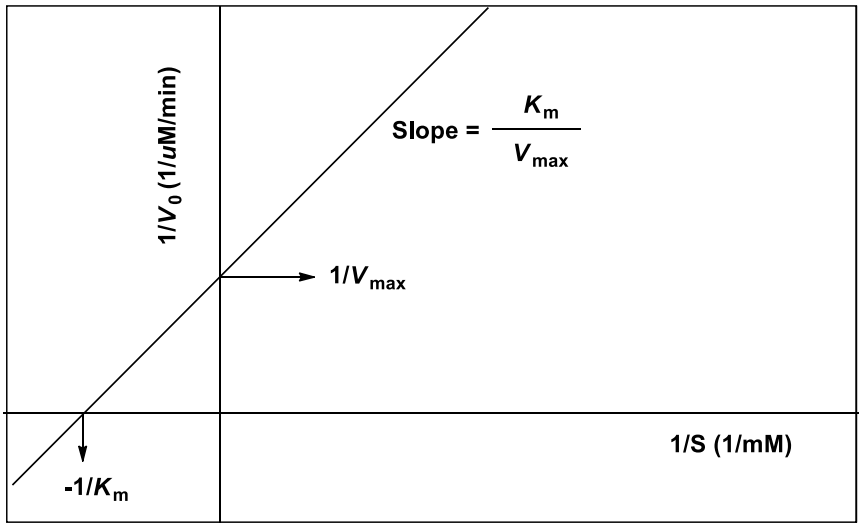


Figure 2.7. Double reciprocal plot or the Lineweaver-Burk plot.

According to the Michaelis-Menten equation (**Equation 1**), V_{max} is the maximum rate when an enzyme is saturated with a substrate, the Michaelis constant, K_m, is defined as the substrate concentration at which V is half of V_{max} and k_{cat} is the turnover number which is defined as the number of substrates converted to product per enzyme per second at V_{max}.

$$y = V_o = V_{max} \times \frac{[S]}{K_m + [S]} \quad \textbf{Equation 1}$$

$$V_{max} = k_{cat}[E] \quad \textbf{Equation 2}$$

$$y = V_o = k_{cat}[E] \times \frac{[S]}{K_m + [S]} \quad \textbf{Equation 3}$$

$$y = \frac{1}{V_o} = \frac{K_m + [S]}{k_{cat}[E][S]} \quad \textbf{Equation 4}$$

$$y = \frac{1}{V_o} = \frac{K_m}{k_{cat}[E]} \times \frac{1}{[S]} + \left(\frac{1}{k_{cat}[E]}\right) \quad \textbf{Equation 5}$$

where, V_0 = Initial rate of the reaction.

K_m = the Michaelis constant, describes the substrate's apparent binding affinity for the enzyme where [S] is at half of V_{max}.

V_{max} = maximum rate when an enzyme is saturated with substrate.

k_{cat} = Catalytic rate constant, describes the maximum number of substrate converted to product per enzyme per second.

[E] = Enzyme concentration.

[S] = Substrate concentration.

Equation 5 is now in the form of straight line (y = ax+b) where *y* represents the dependent variable ($1/V_0$) and *x* the independent variable (1/[S]). Using line fitting by linear regression, constants *a* and *b* can be derived, where *a* is the *y*-intercept and *b* is the slope of the line. Substituting values of *a* and *b* in **Equation 5**, k_{cat} and K_m can be derived. Uncertainties in K_m and k_{cat} were calculated using the error-propagation method based on uncertainties in *a* and *b* from linear regression obtained using LINEST function in Excel. The kinetic data were also fitted by non-linear least-squares using a global fit to provide values of kinetic parameters as a comparison to those obtained by the linearized Lineweaver-Burk plot.

2.5.2. Evaluation of substrate activity of cross-linked oligonucleotides

Substrate preference and deamination activity by $A3B_{CTD}$-QM-ΔL3-AL1swap on the synthesized cross-linked oligonucleotides with **dC** were evaluated using the NMR-based assay as described above. This assay used substrates and $A3B_{CTD}$-QM-ΔL3-AL1swap at known concentrations in a buffer (pH 6.0) containing 10 % deuterium oxide, 50 mM citrate-phosphate, 200 mM NaCl, 2 mM β-mercaptoethanol, 200 μM 4,4-dimethyl-4-silapentane-1-sulfonic acid (DSS). The H-5 proton doublet signals of the cytosine from different oligonucleotides were baselined and integrated. DSS peak at 0 ppm was used as an internal standard to determine the concentration of the substrate. The integrated signal area was converted to substrate concentration and plotted versus time of the reaction. The data were then fitted with linear regression to determine the initial speed of the reaction as described above.

2.5.3. Evaluation of cross-linked ssDNA inhibitors by real-time 1H NMR assay

We used our real-time ^{1}H NMR assay to compare the internally cross-linked **dZ/FdZ**-containing oligos and a hairpin containing FdZ with a linear DNA inhibitor containing **dZ/FdZ** (**dZ**-linear and **FdZ**-linear) that were characterized earlier.[54, 124] Residual activity

of A3B$_{CTD}$-QM-ΔL3-AL1swap on the unmodified oligo (5'-T$_4$**dC**AT) as a substrate in the presence of a known concentration of **dZ/FdZ**-containing inhibitors (linear and cross-linked) was measured using the NMR assay.

2.5.3.1. Evaluation of inhibitors using Dixon plot

A series of ^{1}H NMR spectra was recorded using a known concentration of the standard substrate oligo [S] with varying concentrations of inhibitors and a known concentration of A3B$_{CTD}$-QM-ΔL3-AL1swap in the buffer mentioned earlier. Initial rate of the reaction (V_0) was obtained in the same manner described in the section above. A Dixon plot was produced by plotting the inverse rate of the reaction ($1/V_0$) against the inhibitor concentration [I] assuming the competitive nature of our inhibitors. The Dixon plot is fitted with linear regression to produce a (y-intercept) and b (slope of line). Values of a, b, K_m and $k_{cat.}$ (obtained above) were substituted to get the inhibition constant (K_i) from regression fit using LINEST function of Excel; uncertainty of K_i was calculated by standard error propagation.

$$\boldsymbol{y = \frac{1}{V_o} = \frac{K_m}{K_i \times V_{max} \times [S]} \times [I] + \frac{K_m + [S]}{V_{max} \times [S]}} \qquad \textbf{Equation 6}$$

$$\boldsymbol{y = ax + b} \qquad \textbf{Equation 7}$$

$$\boldsymbol{a = \frac{K_m}{K_i \times V_{max} \times [S]} \qquad b = \frac{K_m + [S]}{V_{max} \times [S]}} \qquad \textbf{Equation 8}$$

$$\boldsymbol{V_{max} = k_{cat}[E]} \qquad \textbf{Equation 9}$$

$$\boldsymbol{K_i = \frac{K_m}{a \times k_{cat} \times [E] \times [S]}} \qquad \textbf{Equation 10}$$

$$\boldsymbol{[E] = \frac{K_m + [S]}{b \times k_{cat} \times [S]}} \qquad \textbf{Equation 11}$$

2.5.3.2. Evaluation of inhibitors using non-linear regression

In this method, non-linear regression analysis was done using XLSTAT add-on in Excel (Microsoft) performing a three-parameter fit (V_{max}, K_m and K_i) in the Michaelis-Menten enzyme kinetic equation for the competitive inhibitor. The non-linear regression analysis was performed by Emeritus Prof. Geoffrey B. Jameson and the details are given in **Chapter 3**.

2.5.3.3. Evaluation of inhibitors using Lambert's W function

Here we tested the inhibition of A3-catalyzed deamination on a dC-hairpin (**Figure 2.8**) by our best inhibitor in the series (done by Dr. Stefan Harjes). dC-Hairpin was selected because it was faster deaminated compared to the linear DNA, so much lower concentrations of enzyme than in previous assays (20 – 27 nM versus 200 nM) were used in the absence and presence of an inhibitor.

dC-hairpin

Figure 2.8. Structure of dC-hairpin used as substrate in evaluation of inhibitors using Lambert's W function

The kinetic data, measured over time until exhaustion of the substrate, were plotted, analyzed by global regression and fitted using the integrated form of the Michaelis-Menten equation (**Equation 12**).[141]

$$[S]_t = K_m W\left(\frac{[S]_0}{K_m}.e^{\frac{[S]_0 - V_{max}t}{K_m}}\right) \qquad \textbf{Equation 12}$$

where W is Lambert's W function, $[S]_t$ is the substrate concentration at specific time, $[S]_0$ is the initial substrate concentration, V_{max} and K_m are the Michaelis-Menten constants and t is the time. The two Michaelis-Menten constants, the initial substrate concentration and an offset which corrects for the integration baseline in the NMR spectra were fitted using Lambert's W function, which provides superior estimates of observed K_m and V_{max} than non-linear regression analysis of the initial rate data.

2.6. Binding studies of cross-linked oligos by isothermal titration calorimetry (ITC)

Isothermal titration calorimetry (ITC) was used to measure the affinity of substrates and inhibitors for A3 enzymes.[140, 142] One important aspect of ITC is that it is a universal technique, and the method is useful to measure accurately the reaction stoichiometry independent of affinity, even for the protein substrates where multiple interactions are possible to be studied by ITC.[143-144]

ITC experiments generally utilize a fixed concentration of the macromolecule in a specific volume and controlled quantities of the other reagent, mainly the substrate, under defined conditions of temperature and pressure. An accurate experiment can easily provide association constant K_a, apparent heat change ΔH_{app} and the stoichiometry n. ΔG obtained from the ITC thermodynamic experiment is directly related to K_d.[145] The basic equation for Gibbs free energy is,

$$\Delta G = \Delta H - T\Delta S \qquad \textbf{Equation 13}$$

So, the equilibrium dissociation constant K_d or inhibition constant K_i can be found directly from the ITC experiments.

$$\Delta G = -RT \ln K_a;\ K_d = 1/K_a \qquad \textbf{Equation 14}$$

Where, ΔH = enthalpy, ΔS = entropy, K_a = equilibrium binding association constant, and K_d = equilibrium dissociation binding constant.

The working principle of ITC relies on two differential cells which are maintained under adiabatic conditions. One cell, called the reference cell, contains only buffer and the other cell, called the sample cell, contains the macromolecule. The addition of a known concentration of the substrate produces heat due to macromolecule-substrate interactions, and the instrument measures the power that is used to compensate the heat event that occurred back to baseline as a function of time (**Figure 2.9**).

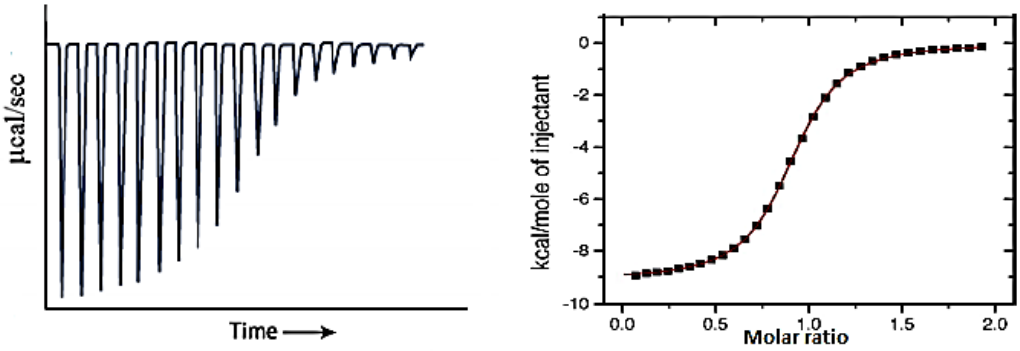


Figure 2.9. Raw ITC data and processed data for thermodynamic parameters.

We tested the thermodynamics of binding of our cross-linked oligos with A3 enzymes in comparison to the linear oligos using catalytically inactive A3A-E72A and the catalytically competent $A3B_{CTD}$-QM-ΔL3-AL1swap and compared this binding to that of linear oligos. For inhibitors, the catalytically active $A3B_{CTD}$-QM-ΔL3-AL1swap was used, as it has the active-site glutamic acid required for protonation of N3 in dZ/FdZ, which activates C4 to accept the nucleophilic H_2O coordinated to the Zn^{2+}, which converts dZ/FdZ into a transition-state analogue of cytidine deamination.[146]

In contrast, A3A-E72A was used for oligos containing dC so that the complications of deamination can be avoided. To prevent protein precipitation during the long timescale of the experiment it was necessary to use improved ITC buffers. ITC buffer 1: 50 mM MES, pH 6.0, 100 mM NaCl, 200 µM EDTA, 1 mM β-mercaptoethanol and ITC buffer 2: 50 mM Na^+/K^+ phosphate, pH 6.0, 50 mM NaCl, 50 mM choline acetate, 2.5 mM TCEP, 200 µM EDTA with 30 mg/mL BSA.

Analysis of ITC data was performed by software provided by the supplier of the MicroCal ITC200, where the raw data were fitted to a one-site binding model to derive parameters including enthalpy (ΔH), entropy (ΔS), equilibrium binding association constant (K_a), and stoichiometry (n). The equilibrium dissociation binding constant (K_d) can then be

derived as $K_d = 1/K_a$, from which the Gibbs free energy ($\Delta G^o = \Delta H^o - T\Delta S^o = -RT \ln (K_a/Q^o)$) can be calculated and combined with the measured ΔH^o to give ΔS^o, where T is temperature of 298 K. All ITC experiments were performed by Dr. Stefan Harjes.

2.7. Enzymes used in the current study

```
hA3A                MEASPASGPRHLMDPHIFTSNFNNG---IGRHKTYLCYEVERLDNGTSVKMDQHRGFLHN
A3A-E72A            MEASPASGPRHLMDPHIFTSNFNNG---IGRHKTYLCYEVERLDNGTSVKMDQHRGFLHN
hA3B_CTD            --------------PDTFTFNFNNDPLVLRRRQTYLCYEVERLDNGTWVLMDQHMGFLCN
A3B_CTDQMΔL3AL1     ------EILRYLMDPDTFTSNFNNG---IGRHKTYLCYEVERLDNGTSVKMDQHMGFLCN

hA3A                QAKNLLCGFYGRHAELRFLDLVPSLQLDPAQIYRVTWFISWSPCFSWGCAGEVRAFLQEN
A3A-E72A            QAKNLLCGFYGRHAALRFLDLVPSLQLDPAQIYRVTWFISWSPCFSWGCAGEVRAFLQEN
hA3B_CTD            EAKNLLCGFYGRHAELRFLDLVPSLQLDPAQIYRVTWFISWSPCFSWGCAGEVRAFLQEN
A3B_CTDQMΔL3AL1     E--------SGRHAELRFLDLVPSLQLDPAQIYRVTWFISWSPCFSWGCAGEVRAFLQEN

hA3A                THVRLRIFAARIYDY-DPLYKEALQMLRDAGAQVSIMTYDEFKHCWDTFVDHQGCPFQPW
A3A-E72A            THVRLRIFAARIYDY-DPLYKEALQMLRDAGAQVSIMTYDEFKHCWDTFVDHQGCPFQPW
hA3B_CTD            THVRLRIFAARIYDY-DPLYKEALQMLRDAGAQVSIMTYDEFEYCWDTFVYRQGCPFQPW
A3B_CTDQMΔL3AL1     THVRLRIKAARIYDY-DPLYKEALQMLRDAGAQVSIMTYDEFEYCWDTFVYRQGCPFQPW

hA3A                DGLDEHSQALSGRLRAILQNQGN
A3A-E72A            DGLDEHSQALSGRLRAILQNQGN
hA3B_CTD            DGLEEHSQALSGRLRAILQNQGN
A3B_CTDQMΔL3AL1     DGLEEHSQALSGRLRAILQ----
```

Figure 2.10. Amino acid sequences of human A3A (hA3A, $hA3B_{CTD}$ in comparison with the enzymes used in this work. $A3B_{CTD}$ is the C-terminal domain of A3B. $A3B_{CTD}$-QM-ΔL3-AL1swap (loop 1 swapped with A3A); A3A-E72A-active site glutamate mutated to alanine.

Chapter 3. Design, Synthesis, and Evaluation of Cross-linked ssDNAs as Inhibitors of APOBEC3 Enzymes

3.1. Introduction

In the last 15 years, alkyne- and azide-modified oligos have been widely used for the synthesis of a vast variety of bioconjugates by azide-alkyne 'click chemistry'. Usually, terminal alkynes can be introduced into oligos using phosphoramidite chemistry and several reagents are commercially available, such as the **Y** and **dU^{E}** modifications used in the current work. However, there are limited options for introduction of the azide functionality into oligos, especially in the middle of the DNA sequence. Usually, isolation of an azido-containing phosphoramidite and its use in the DNA synthesis are problematic[147-148] due to possible Staudinger reaction between the organic azide and a trivalent phosphorus in the phosphoramidite function. Indeed, the Staudinger reaction has been used recently for modification of a phosphate backbone.[149-154] Interestingly, a Staudinger reaction does not occur on the azide group in the support-bound oligonucleotide chain during subsequent elongation using nucleotide phosphoramidites.[155] This led to the development of azide-containing solid supports to produce oligos bearing organic azides at the 3'-end.[156-159] In 2014, Fomich *et al.*[160] demonstrated that azide-carrying phosphoramidites can be obtained and used in DNA synthesis to produce oligos having up to three organic azides in the middle of the sequence. This is achievable if the azide-carrying phosphoramidite is not concentrated and is used directly in automated DNA synthesis after aqueous work up and drying of the solution over sodium sulfate. These advances in the field were used in the current work to produce oligos containing an organic azide at the 3'-end as well as in the middle of the sequence using azide-modified 2-deoxyribose, dA and dC described below, respectively.

3.2. Results

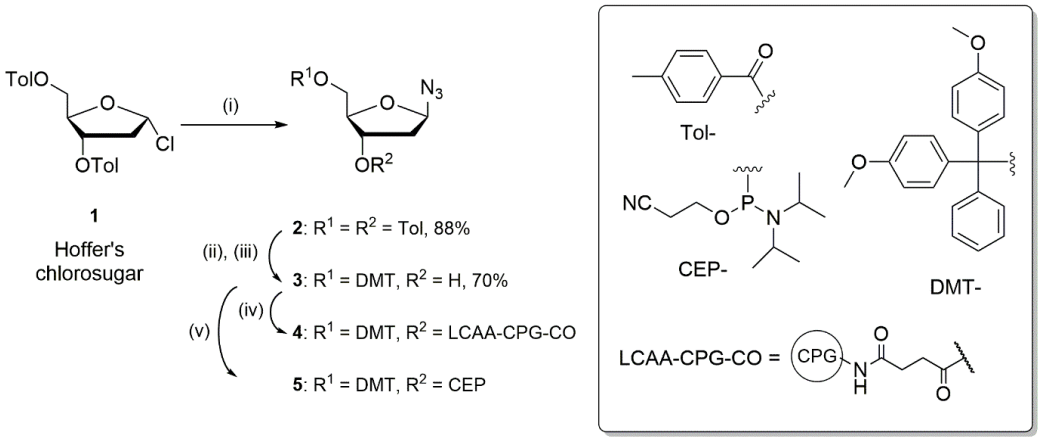
3.2.1. Synthesis of DNA precursors containing organic azides and alkynes for cross-links 1-3

The initial approach for making a cross-linked oligonucleotide was based on the available starting materials so that the U shape can be mimicked by creating a terminal cross-link. For this we employed a commercially available alkyne (**Y**) and synthesized an azido-sugar (**dR^{N3}**) embedded in CPG. For internal cross-linking, the positions of modified nucleosides were selected based on the knowledge from the crystal structure of A3A and A3B-ctd–ssDNA complexes. As described earlier in **Chapter 1**, A3A and A3B prefer a purine base at position +1. We selected 2'-deoxyadenosine as a purine base for the

modification in cross-link 1 and 2. The crystal structure also showed that the +1 nucleobase makes a water-mediated hydrogen bond with the -2 base which is a pyrimidine. Here we used azido-sugar (**dR^{N3}**), a commercially available 2'-deoxyuridine phosphoramidite (**dU^{E}**), and an azide-containing 5-methyl 2'-deoxycytidine (**dC^{N3}**). Phosphoramidite modifications were synthesized so that they can be easily incorporated in DNA sequences using standard automated DNA synthesis. The positions of the linker and the length of the linker were chosen based on the fact that in crystal structure the +1 and -2 bases are in close proximity to each other. The linkers containing azide and alkyne modifications were introduced so that once cross-linked, the linker should not interfere with the interactions seen between the oligo and the protein in the crystal structure. The artificial carbazole nucleobase (**dH^{E}**) in cross-link 3 was selected based on the observation that the nucleobase at position +1 engages in a stacking interaction with His29, and the nucleobases following that in 3'-position also stack. We envisioned that carbazole, being a highly hydrophobic aromatic moiety, at this position would result in better stacking interactions and would help to exclude water molecules from the active site, contributing to better binding of the oligo.

A method reported recently[161] described the synthesis of compound **2** from Hoffer's chlorosugar[125] **1** (**Scheme 3.1**) using a Lewis acid and trimethylsilyl azide, but a low β:α selectivity and moderate yields were reported. In contrast, we employed a phase-transfer protocol[162] with minor changes using NaN_3 and Bu_4NHSO_4 with vigorous stirring in saturated aqueous $NaHCO_3$ and chloroform followed by addition of Hoffer's chlorosugar **1**. The reaction was finished in 20 min. and after work-up resulted in almost pure azide **2** with a high yield (88%) and β:α ratio of 16:1 (6% of α-anomer). Subsequent recrystallization from EtOH slightly increased the β:α ratio to 19:1 (5% of the α-anomer). The β isomer was confirmed by NOESY NMR experiment (**Figure 3.1**). The spectrum shows interaction of H1 and H4 of compound **2** through space, and not between H1 and H3 confirming it as the β-anomer. This protocol was optimized to a multi-gram scale synthesis of azide **2** in our laboratory. The controlled-pore glass (CPG) conjugated to azido-sugar **4** was prepared from a 5'-dimethoxytrityl (DMT) protected azido-sugar **3** by a slight modification of the reported methods[157, 163] (**Scheme 3.1**).

The loading on CPG was found to be 39 μmol/g by UV absorption of the DMT cation released upon treatment of a sample of the modified support **4** with a 3% solution of dichloroacetic acid in dichloromethane. The steps involved in the synthesis of CPG conjugated azido-sugar **4** and the phosphoramidite **5** are described below.



Scheme 3.1. Synthesis of azido sugar phosphoramidite and CPG support.

Reagents and conditions: i) NaN_3, Bu_4NHSO_4, satd. $NaHCO_3$:$CHCl_3$ (1:1), rt, 20 min; ii) 30% aq. NH_3, MeOH, rt, 3 days; iii) 4,4'-dimethoxytrityl chloride (DMTCl), pyridine, 0 °C to rt, overnight; iv) long-chain aminoalkyl controlled-pore glass (LCAA-CPG-CO_2H), *N*-(3-dimethylaminopropyl)-*N'*-ethylcarbodiimide (EDC), Et_3N, 4-(dimethylamino)pyridine (DMAP), pyridine:DMF (1:1), rt, 3 days; v) *N*,*N*-di*iso*propylamino-2-cyanoethoxychlorophosphine (CEP-Cl), Et_3N, dry DCM, 0 °C.

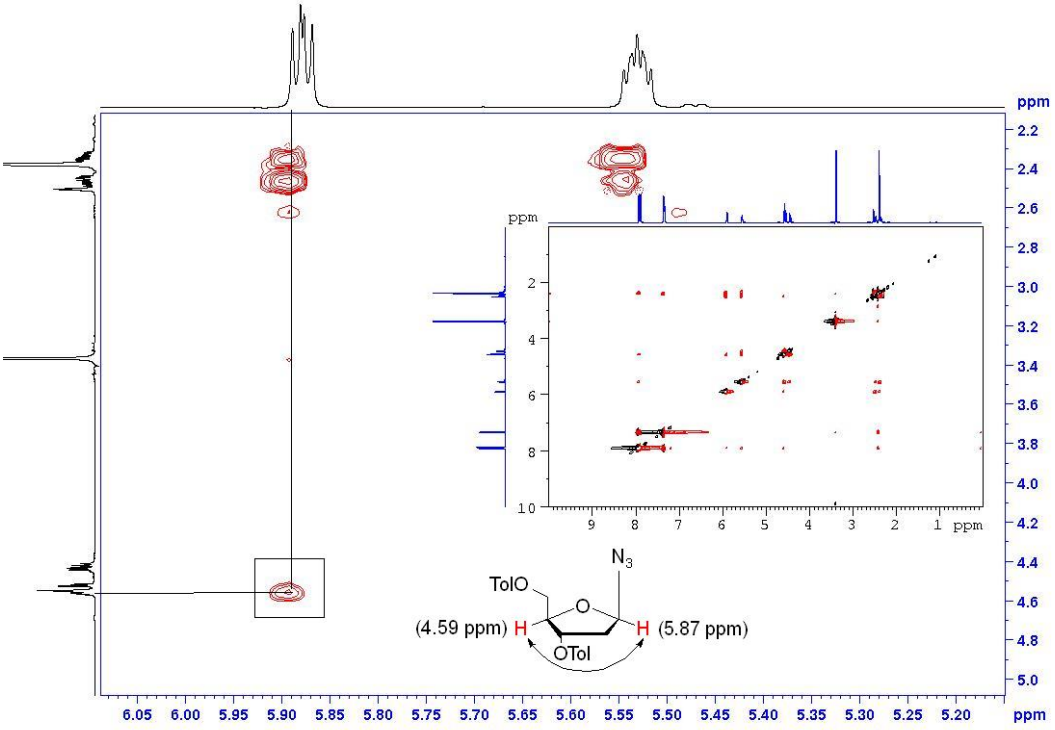
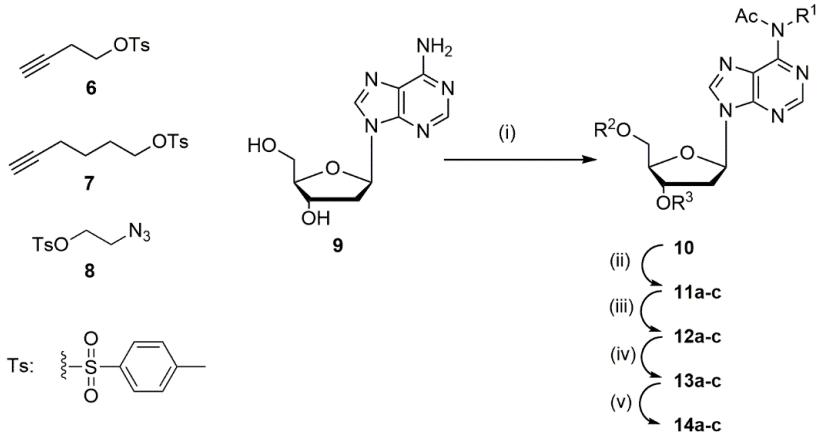


Figure 3.1.NOESY Spectrum (DMSO-d_6) of compound **2**.

The spectrum shows cross-peak between of 1 and 4-protons of compound **2** through space, confirming it as a β-anomer.

The second strategy relied on creation of internal cross-links 1 to 3 between azide- and alkyne-containing nucleotides. The synthesis of modified nucleosides is summarized in **Table 3.1** and **Scheme 3.2- Scheme 3.4.**



Scheme 3.2. Synthesis of azido and alkyne modified 2'-deoxyadenosine phosphoramidites.

Reagents and conditions: i) Acetic anhydride, pyridine, 60 °C; ii) compounds **6**, **7** or **8**, Cs_2CO_3, CH_3CN, 60 °C, overnight for the synthesis of **11a-c**, respectively; iii) Et_3N in MeOH/ H_2O, rt; iv) 4,4'-dimethoxytrityl chloride (DMTCl), pyridine, 0 °C→r.t., overnight; v) *N*,*N*-di*iso*propylamino-2-cyanoethoxychlorophosphine, Et_3N, CH_2Cl_2, rt. Compound **14c** was not isolated but used immediately in the DNA synthesis.

Table 3.1. Synthesis of modified 2'-deoxyadenosine phosphoramidites containing a terminal alkyne or an azide.

#	R^1	R^2	R^3	Yield, %
10	H	Ac	Ac	75
11a	$(CH_2)_2C{\equiv}CH$	Ac	Ac	80
11b	$(CH_2)_4C{\equiv}CH$	Ac	Ac	85
11c	$(CH_2)_2N_3$	Ac	Ac	80
12a	$(CH_2)_2C{\equiv}CH$	H	H	82
12b	$(CH_2)_4C{\equiv}CH$	H	H	82
12c	$(CH_2)_2N_3$	H	H	80
13a	$(CH_2)_2C{\equiv}CH$	DMT	H	89
13b	$(CH_2)_4C{\equiv}CH$	DMT	H	86
13c	$(CH_2)_2N_3$	DMT	H	89
14a	$(CH_2)_2C{\equiv}CH$	DMT	$P(N\textit{i}Pr)_2OCH_2CH_2CN$	92
14b	$(CH_2)_4C{\equiv}CH$	DMT	$P(N\textit{i}Pr)_2OCH_2CH_2CN$	90
14c	$(CH_2)_2N_3$	DMT	$P(N\textit{i}Pr)_2OCH_2CH_2CN$	-

Here, we decided to introduce azido- and alkyne-containing alkane linkers into the N^6 position of dA, due to proximity of an exocyclic amine of dA (at +1) and a 5-methyl group of thymidine (at -2) in the crystal structure. In the past, various methods employed for N^6 alkylation of dA usually relied upon alkylation in the N^1 position followed by subsequent Dimroth rearrangement to the N^6-alkylated product.[164-165] In contrast to previous reports, we observed a direct alkylation of the N^6 position in triacetylated dA (**7**) using tosylated alkyne and azide-containing alcohols (**6-8**)[166] and Cs_2CO_3 as the base. The alkylation at the N^6 position of dA was confirmed for compounds **11a-c** by $^1H,^{13}C$ heteronuclear multiple bond correlation (HMBC) NMR experiment in which three-bond correlations are detected. For example, in HMBC spectrum of compound **11c**, we observed that proton signals of the CH_2-group of the azidoalkyl linker at 4.31 ppm had a cross-peak not only with C^6 of adenine at 152.58 ppm but also with the carbonyl carbon of the acetyl group at 170.06 ppm (**Figure 3.2**).

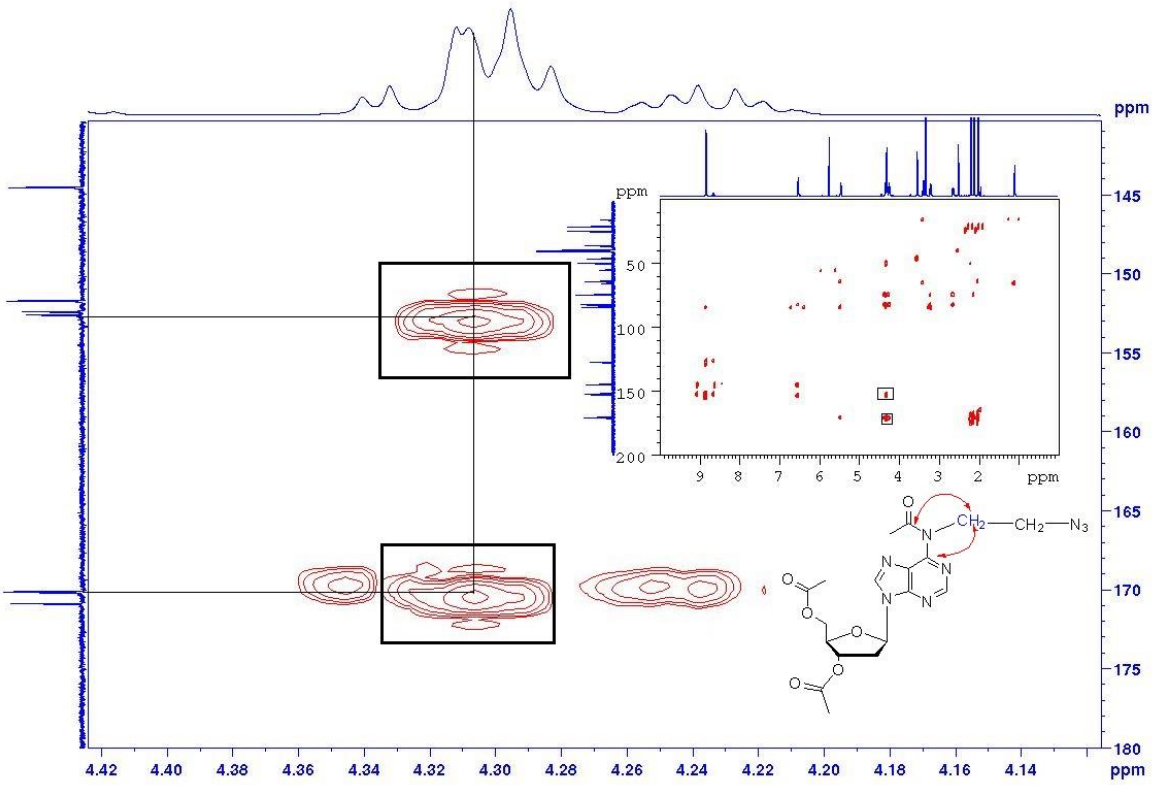


Figure 3.2. HMBC spectrum (DMSO-d_6) of the compound **11c**.

The enlarged portion of the spectrum shows the cross-peak between CH_2-group of the azidoalkyl linker with C^6 of adenine and with the carbonyl carbon of the acetyl group.

This is only possible to observe for N^6 and not for N^1 alkylated dA (three *versus* five bonds between hydrogen of CH_2 and carbon of C=O, respectively). Similar patterns of cross-peaks were observed for compounds **11a** and **11b**. After the selective deprotection of 3'- and 5'-acetyl groups, the 5'-hydroxy group of compounds **12a-c** was converted into the 4,4'-

dimethoxytrityl derivatives **13a-c** under standard conditions. Phosphoramidites **14a-b** for alkyne-containing dA were obtained using standard phosphitylation reaction with *N*,*N*-di*iso*propylamino-2-cyanoethoxychlorophosphine and Et_3N in dry CH_2Cl_2 followed by aqueous work-up and silica gel purification. However, this protocol must be modified for azide-containing dA. A pilot phosphitylation reaction was performed on compound **13c** using previously described conditions[160] in dry $CDCl_3$ producing the phosphoramidite **14c** with a purity of more than 70% as shown by ^1H and ^{31}P NMR spectra (**Figure 3.3a**).

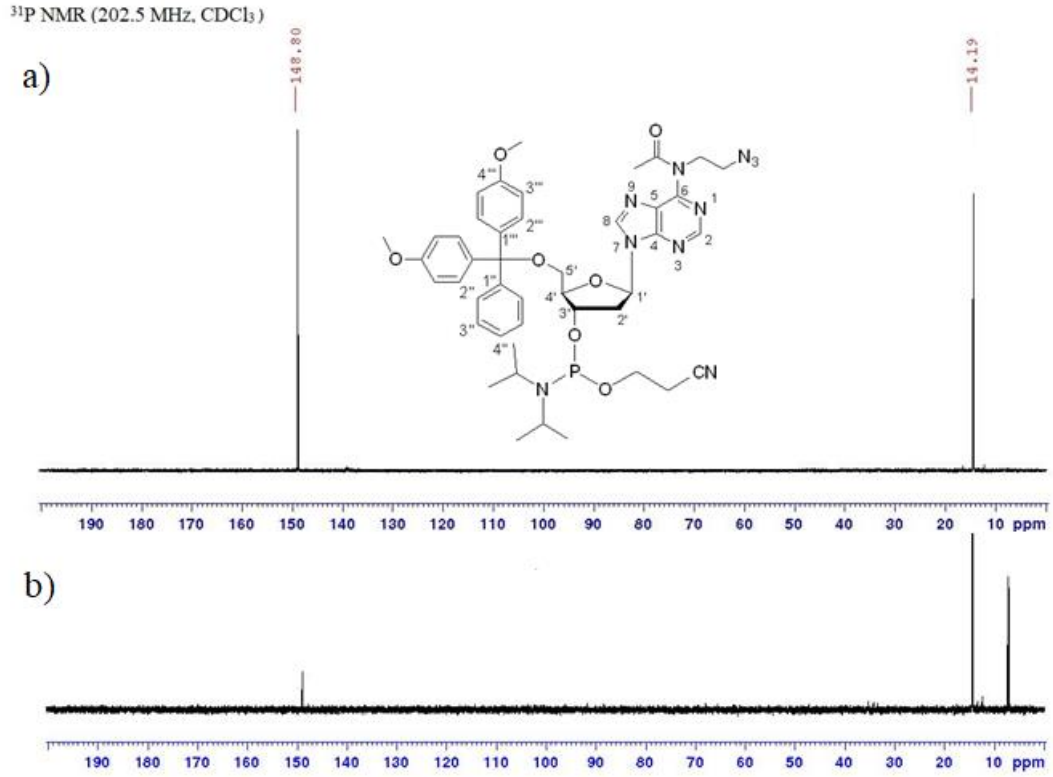
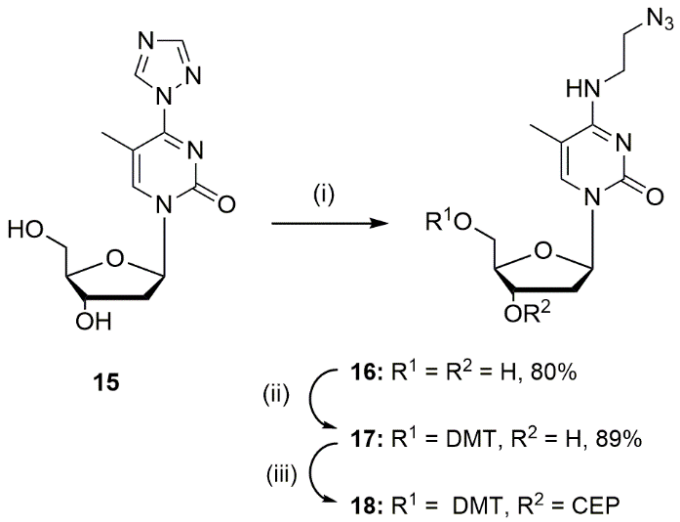


Figure 3.3. ^{31}P NMR spectra ($CDCl_3$) of compound **14c**, showing the degradation of the phosphoramidite over time.

a) ^{31}P NMR of **14c** immediately after the synthesis and b) ^{31}P NMR of the compound **14c** after 24 hr showing significant degradation as shown by the reduced intensity of the peak at 148.80 ppm and appearance of an additional peak at 7.3 ppm. The peak at 14.19 ppm is the degraded excess phosphitylation reagent in the reaction mixture.

This suggests that compound **14c** can be used for automated DNA synthesis, with a limitation that the freshly prepared phosphoramidite should be used immediately after its synthesis because the storage of this solution at room temperature for 24 hours results in significant degradation of compound **14c** as evidenced by NMR (**Figure 3.3b**). For automated DNA synthesis, the reaction was performed under the same conditions using dry dichloromethane as solvent to obtain phosphoramidite **14c**. A similar protocol was used for preparation of phosphoramidite of **dR^{N3}** (compound **5**, **Scheme 3.1**).

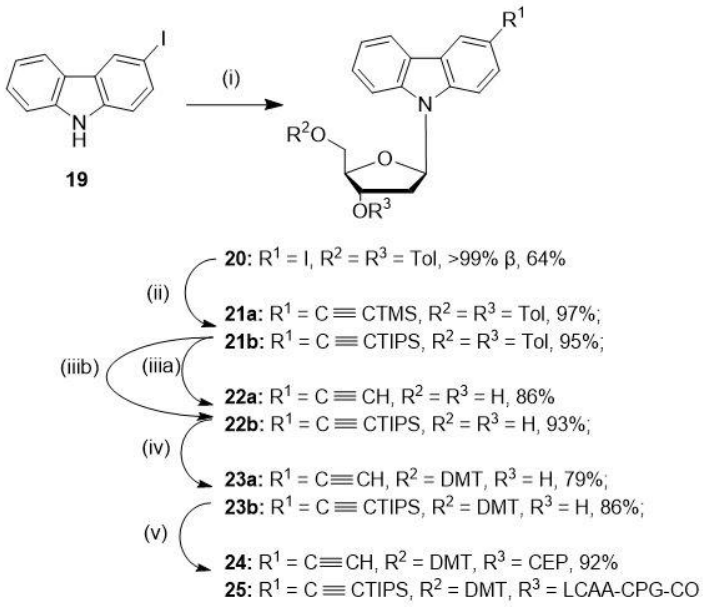
An azide-containing 5-methyl-2'-deoxycytidine **dC^{N3}** was prepared starting from 4-triazolothymidine **15**[167] and 2-azidoethylamine[168] (**Scheme 3.3**) followed by standard DMT-protection and phosphitylation using protocols for azide-containing phosphoramidites described above.



Scheme 3.3. Synthesis of 2'-deoxy-5-methylcytosine phosphoramidite containing a two-carbon linker with azide.

Reagents and conditions: i) 2-Azidoethylamine, 60 °C, 1 hr; ii) 4,4'-dimethoxytrityl chloride (DMTCl), pyridine, 0 °C→r.t., overnight; iii) *N*,*N*-di*iso*propylamino-2-cyanoethoxychlorophosphine (CEP-Cl), Et_3N, dry DCM, 0 °C. Compound **18** was not isolated but used immediately in the DNA synthesis.

Providing that A3A as well as A3B prefer a purine (dG or dA) in +1 position of the DNA sequence, we hypothesized that the use of a pseudo-nucleoside with a large hydrophobic residue instead of a nucleobase will provide a better binder to A3 enzymes. We decided to use a carbazole nucleoside **dH^{E}** because several carbazole derivatives have been successfully incorporated into DNA in the past,[169-173] some of which were used for modulation of thermal stability of DNA duplexes relying on noncovalent π-stacking interactions and/or intercalation with nucleic acids or with other polyaromatic compounds.[169-173] In addition, carbazole is a well-known scaffold in medicinal chemistry found also in natural products.[174-176] The synthesis of carbazole nucleoside **dH^{E}** started from 3-iodocarbazole **19** and Hoffer's chlorosugar **1** which coupled in the presence of K-^{t}BuO in 1,4-dioxane at room temperature (**Scheme 3.4**). The β isomer was confirmed by a NOESY NMR experiment. The spectrum shows interaction of 1'-and 4'-protons of the compound **20** through space, confirming it as a β-nucleoside (**Figure 3.4**).



Scheme 3.4. Synthesis of carbazole nucleosides containing unprotected and TIPS-protected terminal alkyne.

Reagents and conditions: i) Hoffer's chlorosugar, K-^{t}BuO, 1,4-dioxane, 60 ºC, 2 hr; ii) trimethylsilyl acetylene (TMS-acetylene) (for **a**) or tri*iso*propylsilyl acetylene (TIPS-acetylene) (for **b**), $Pd(PPh_3)_4$, CuI, Et_3N, dry DMF, 60 ºC, 2 hr; iiia) K_2CO_3, MeOH, rt, 2 hr (for **a**); iiib) 30% aq. NH_3, MeOH, rt, 3 days (for **b**); iv) 4,4'-dimethoxytrityl chloride (DMTCl), pyridine, 0 °C→r.t., overnight; v) for **24**: *N*,*N*-di*iso*propylamino-2-cyanoethoxychlorophosphine (CEP-Cl), Et_3N, dry DCM, 0 °C; for **25**: long-chain aminoalkyl controlled-pore glass (LCAA-CPG-CO_2H), *N*-(3-dimethylaminopropyl)-*N*′-ethylcarbodiimide (EDC), Et_3N, 4-(dimethylamino)pyridine (DMAP), pyridine:DMF (1:1), rt, 3 days.

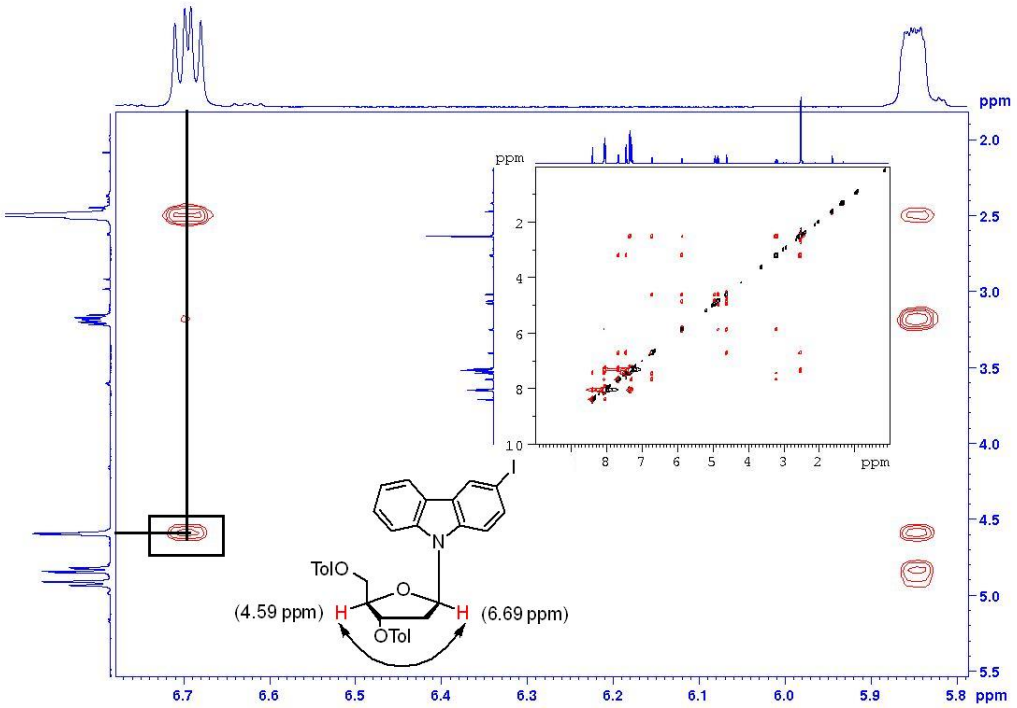


Figure 3.4. NOESY Spectrum ($CDCl_3$) of compound **20**.

The spectrum shows interaction of 1'-and 4'-protons of the compound **20** through space, confirming it as a β-nucleoside.

A subsequent Sonogashira reaction with trimethyl silyl (TMS) or tri*iso*propyl silyl (TIPS)-protected acetylene catalyzed by $Pd(PPh_3)_4$ and CuI resulted in compounds **21a-b**.

3.2.1.1. Mechanism of the Sonogashira reaction

The Sonogashira reaction is believed to be proceeding through two independent catalytic cycles *viz*, a palladium cycle and a copper cycle (**Figure 3.5**). The palladium cycle starts with oxidative addition of organo-halide or triflate (R^1-X or R^1-OTf) to the Pd^0 catalyst to produce a Pd^{II} and this is the rate limiting step of the reaction. This Pd^{II} species then connects with the copper cycle where a copper acetylide formed in the copper cycle gives species **iii** by a transmetallation to regenerate the copper catalyst. The final coupled product is formed by a *cis*/*trans* isomerization and reductive elimination to regenerate the catalyst.

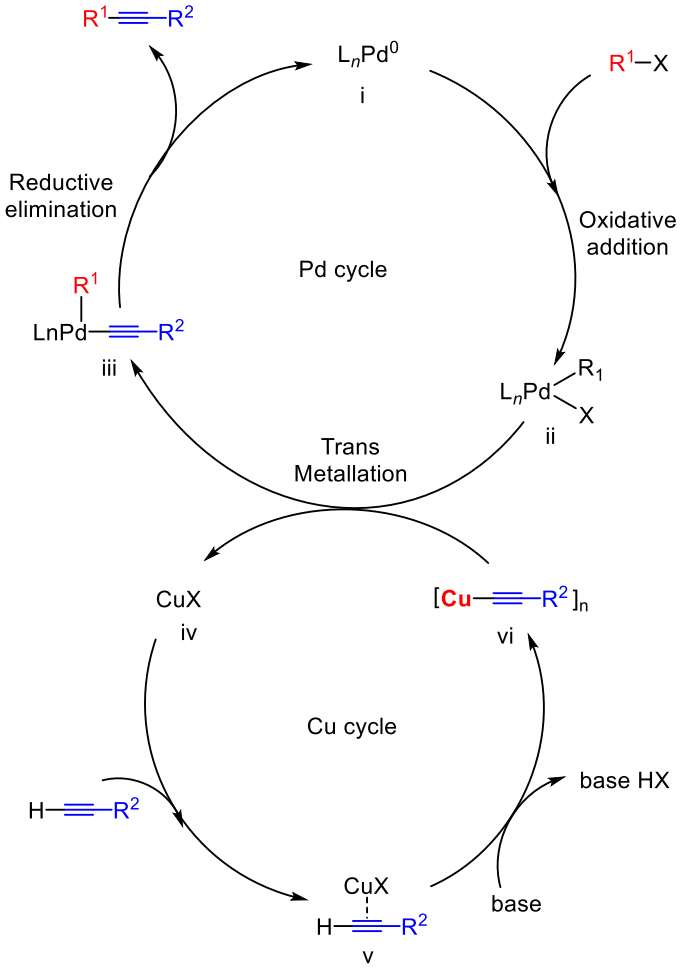


Figure 3.5. Mechanism of the palladium-catalyzed Sonogashira reaction.

Removal of 3',5'-*O*-toluoyl and TMS protecting groups in **21a** was accomplished by K_2CO_3 in MeOH and for deprotection of 3',5'-*O*-toluoyls in **21b** we used 30% aq. NH_3 in MeOH at rt for 3 days. Conversion into 5'-*O*-DMT derivatives followed by phosphitylation

to produce phosphoramidite **24**, or loading of TIPS-protected 3-ethynylcarbazole **23b** onto CPG-support providing compound **25** was performed using established protocols.

3.2.2. Synthesis of cross-linked oligonucleotides by CuAAC

Oligodeoxynucleotides (oligos) were synthesized by the phosphoramidite method on a MerMade-4 DNA/RNA synthesizer (BioAutomation) on a 5 μmol scale using the standard manufacturer's protocol for native nucleotides, whereas increased coupling time from 2 to 10 min was used for modified alkyne and azide phosphoramidites. To prevent Staudinger reaction between the azide and phosphoramidite present in one molecule, we did not purify products **5**, **14c** and **18** or concentrate them. Instead, these compounds were used directly for the oligonucleotide synthesis. The concentration of reagents **5**, **14c** and **18** in DCM was found to be 0.18, 0.20 and 0.14 M respectively, determined by UV absorption of the 4,4'-dimethoxytrityl cation released upon treatment of an aliquot of this sample with a 3% solution of dichloroacetic acid in dichloromethane. After the synthesis, oligos were cleaved from the solid support and deprotected at r.t, using conc. NH_4OH (for dC and dZ-containing oligos). This deprotection condition, however, led to significant degradation of FdZ-containing oligos so we used 10% Et_2NH in acetonitrile followed by ethylenediamine/toluene for the deprotection of FdZ-containing oligos.[124] The full-length oligos were purified by reverse-phase HPLC and characterized by electrospray ionization mass spectrometry (ESI-MS). We established that the protocol used for deprotection of FdZ-containing oligos leads to hydration of unprotected triple bonds in dU^E (**Figure 3.6**) and dH^E as determined by ESI-MS showing mass of isolated oligos being 18 units higher than expected (**Table 3.2** and **Figure 3.7**).

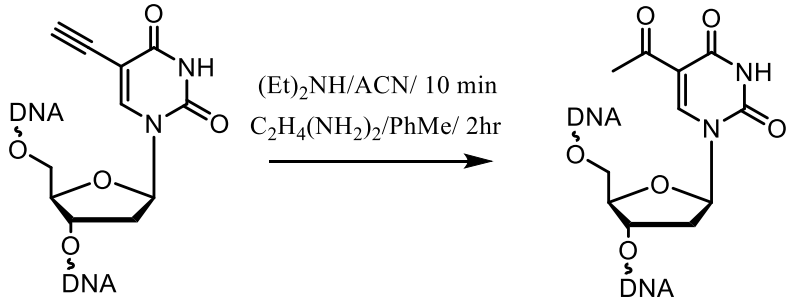


Figure 3.6. Hydration of the triple bond in dU^E using the FdZ oligo deprotection conditions producing an acetyl moiety.

Table 3.2. Sequence and the ESI-HRMS of an unprotected FdZ[C^{N3}(-2), (+1)] oligo under the protocol used for deprotection of FdZ-containing oligos leads to hydration of unprotected triple bonds in dH^E

Sequence, 5'→3'	Oligo abbreviation	ESI-MS [Da] found/calculated
$ATC^{N3}TFdZH^ETTT$	FdZ[C^{N3}(-2), (+1)]	2823.49/2805.53

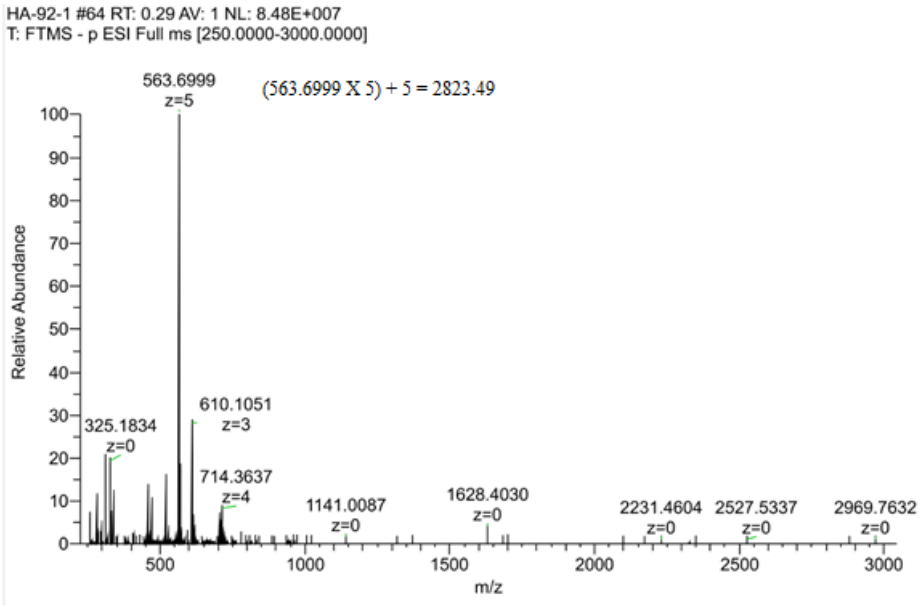


Figure 3.7. ESI-HRMS (negative mode) showing 18 mass units higher than what was expected for an oligo containing unprotected triple bond confirming the hydration leading to acetyl moiety.

Similar results have been reported for oligos containing 7-ethynyl-7-deazapurine, 5-ethynylpyrimidine nucleosides[177] and indole pseudonucleosides with alkynyl side chains.[178-179] To circumvent this problem we synthesized carbazole nucleoside **25** containing a TIPS-protecting group that is stable to DNA synthesis[177, 179] and during cleavage of dZ- as well as FdZ-containing oligos from the CPG support and nucleobase deprotection. However, standard removal of TIPS in FdZ-containing oligos using 1 M tetra-*n*-butylammonium fluoride (TBAF) in THF[177] led to a hydrated oligo showing 18 units higher mass by ESI-MS than calculated (**Table 3.3** and **Figure 3.8**).

Table 3.3. Sequence and the ESI-HRMS of an silyl protected 4-mer FdZ[C^{N3}(-2), (+1)] oligo under the protocol used for deprotection of TIPS protecting group leads to hydration followed by ring opening of FdZ.

Sequence, 5'→3'	Oligo abbreviation	ESI-MS [Da] found/calculated
C^{N3}TFdZH^{E}	4-mer FdZ[C^{N3}(-2), (+1)]	1293.27/1275.29

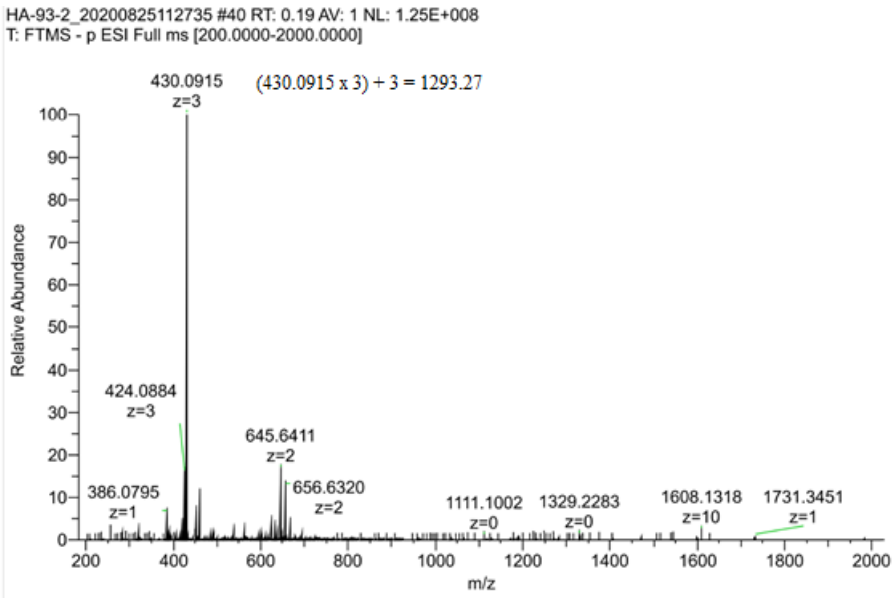


Figure 3.8. ESI-HRMS (negative mode) of an oligo containing hydrated FdZ showing 18 mass units higher than the expected.

^{1}H-NMR spectrum of this oligo revealed two strongly coupled doublets at 7.5 and 9.0 ppm which were previously assigned as three-bond ^{1}H-^{19}F coupling of an aldehyde **26** formed after OH-mediated degradation of 5-FdZ (**Figure 3.9**)[180]. Interestingly, we did not observe hydrated oligos after removal of TIPS for dZ-containing oligos, which is also in line with a previous report that the aldehyde formed after zebularine degradation is very reactive and cyclizes back to zebularine.[180]

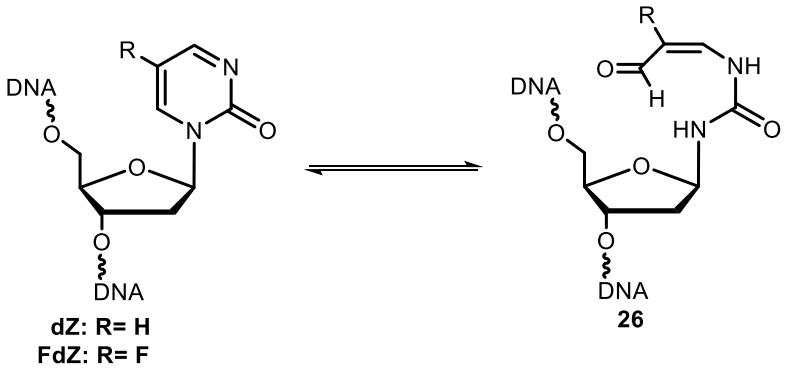


Figure 3.9. Hydration of dZ/FdZ-containing oligos upon treatment with 1 M TBAF in THF forming an aldehyde.

Strong nucleophiles can add to C4 and C6 of the substituted pyrimidine 2-ones facilitated by electron-withdrawing groups at N1 and C5. Conjugate addition of hydroxide at C6 leads to opening of the pyrimidine ring at C6 forming an aldehyde. In the case of FdZ the reaction is irreversible whereas dZ cyclizes back to the original form (**Figure 3.10**)

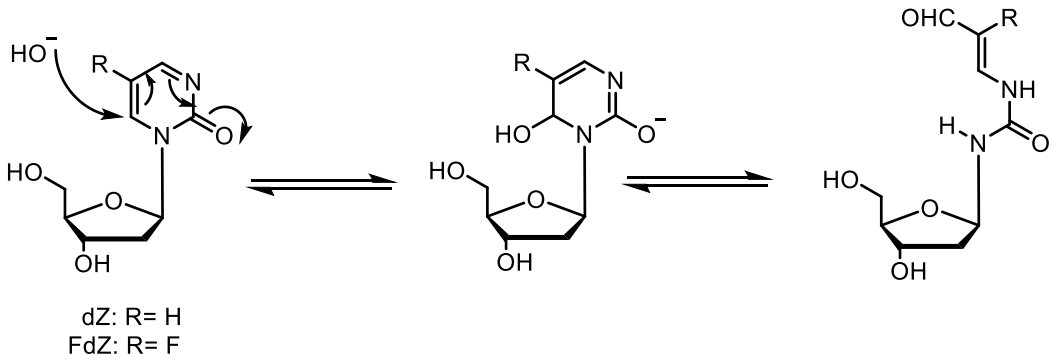


Figure 3.10. Hydration of FdZ/dZ nucleosides forming an aldehyde.

It also indicates that terminal triple bond in **dH^{E}** is not hydrated under conditions of TIPS-removal. To prevent HO^- mediated degradation of FdZ during TIPS deprotection, we included *o*-nitrophenol as a weak acid ($pK_a = 7.22$)[181] to a 1 M TBAF solution in THF, which led to successful cleavage of TIPS and isolation of desired FdZ-containing oligos. To summarize, dZ-containing oligos can be synthesized using modified nucleotides with unprotected terminal alkynes but deprotection of oligos using conc. NH_4OH should be performed at room temperature as the hydration of the triple bond was observed by previous studies at elevated temperatures.[177] Synthesis of FdZ-containing oligos requires assembly of oligos using nucleotides having TIPS-protected alkynes. After the synthesis, oligos on support are treated by 10% Et_2NH in acetonitrile for 10 min followed by incubation of the support in ethylenediamine/toluene mixture for 2 h at room temperature with subsequent release of TIPS-containing FdZ-oligo in water. Then at this stage or after oligo purification, TIPS-protecting group can be cleaved using 1 M TBAF solution in THF with addition of *o*-nitrophenol at 22 °C for 30 min. Detailed procedures are given in experimental section.

The purified linear oligos were cross-linked by CuAAC following the protocol implemented in our laboratory (**see section 3.5.8**). Progress of the reaction was monitored by reverse-phase HPLC. Gratifyingly, a new peak appeared in the chromatogram for all cross-linking experiments over a period of 16 hr, characterized by shorter retention times than those of the starting materials (**Table 3.4** and **Table 3.5**).

Table 3.4. List of linear oligos with azide/alkyne modifications as well as terminally cross-linked oligos having 1,4-disubstituted 1,2,3-triazole modification.

sequence, 5'→3'	Oligo abbreviation[a]	Linear oligos		Cross-linked (X) oligos	
		Retention time (min)	ESI-MS [Da] found/calculated	Retention time (min)	ESI-MS [Da] found/calculated
YTTTCATT**dR^{N3}**	dC-9mer	17.5	2552.47/2552.48	14.1	2552.47/2552.48
YTTCAT**dR^{N3}**	dC-7mer	18.0	1944.37/1944.37	13.6	1944.37/1944.37
YTCA**dR^{N3}**	dC-5mer	17.8	1336.29/1336.30	13.5	1336.29/1336.30

a) dC denotes substrate of A3 present at position 0 in the DNA sequence. The length of terminally cyclized oligos is described being 9, 7 or 5mer in relation to the length of the native DNA sequence having the same number of phosphate groups. In other words, modifications **Y** and **dR^{N3}** are counted as modified nucleotides. X at the end of abbreviation stands for a cross-linked oligo

Table 3.5. List of linear oligos with azide/alkyne modifications as well as internally cross-linked oligos having 1,4-disubstituted 1,2,3-triazole modification.

Oligo abbreviation[a]	Linear oligos		Cross-linked (X) oligos	
	Retention time(min)	ESI-MS [Da] found/calculated	Retention time(min)	ESI-MS [Da] found/calculated
Cross-link 1				
dC[R^{N3}(-3),A^{Y2}(+1)]	17.51	2646.46/2646.48	14.69	2646.46/2646.48
dC[R^{N3}(-2),A^{Y2}(+1)]	18.00	2646.46/2646.48	14.93	2646.47/2646.48
dC[R^{N3}(-1),A^{Y2}(+1)]	17.76	2646.46/2646.48	14.79	2646.47/2646.48
dC[R^{N3}(-2),A^{Y4}(+1)]	18.02	2674.48/2674.51	16.03	2674.48/2674.51
dZ[R^{N3}(-3),A^{Y2}(+1)]	17.51	2631.44/2631.47	14.81	2631.45/2631.47
dZ[R^{N3}(-2),A^{Y2}(+1)]	17.00	2631.44/2631.47	15.01	2631.45/2631.47
dZ[R^{N3}(-1),A^{Y2}(+1)]	17.76	2631.44/2631.47	15.02	2631.44/2631.47
dZ[R^{N3}(-2),A^{Y4}(+1)]	18.03	2659.47/2659.50	16.08	2659.47/2659.50

Oligo abbreviation[a]	Linear oligos		Cross-linked (X) oligos	
	Retention time (min)	ESI-MS [Da] found/calculated	Retention time (min)	ESI-MS [Da] found/calculated
Cross-link 2				
dC[U^{E}(-3),A^{N3}(+1)]	16.7	2757.46/2757.49	15.6	2757.47/2757.49
dC[U^{E}(-2),A^{N3}(+1)]	16.3	2757.46/2757.49	15.3	2757.46/2757.49
dZ[U^{E}(-3),A^{N3}(+1)]	16.9	2742.45/2742.48	15.7	2742.46/2742.48
dZ[U^{E}(-2),A^{N3}(+1)]	16.9	2742.45/2742.48	15.7	2742.45/2742.48
Cross-link 3				
dC[C^{N3}(-2),H^{E}(+1)]	18.08	2802.51/2802.54	16.52	2802.51/2802.54
dZ[C^{N3}(-2),H^{E}(+1)]	19.47	2787.51/2787.54	16.49	2787.51/2787.54
FdZ[C^{N3}(-2),H^{E}(+1)]	19.63	2805.49/2805.52	16.52	2805.49/2805.50
4-mer FdZ[C^{N3}(-2),H^{E}(+1)]	11.23	1275.27/1275.28	9.54	1275.27/1275.28
4-mer dZ[C^{N3}(-2),H^{E}(+1)]	11.20	1257.28/1257.29	9.59	1257.28/1257.29
3-mer dZ[C^{N3}(-2),H^{E}(+1)]	11.64	953.24/953.25	9.80	953.24/953.25

a) d stands for a DNA backbone; C, Z or FdZ denote substrate (dC) or inhibitors (dZ, FdZ) of A3 present in position 0 in the DNA sequence ATTTCATTT. The abbreviation of alkyne- and azido-containing nucleotides is presented in square brackets and the position of each modification relative to the position 0 being dC or **dZ/FdZ** is shown in brackets. X at the end of abbreviation refers to a cross-linked oligo containing 1,4-disubstituted 1,2,3-triazole. For example, abbreviation dC[R^{N3}(-3),A^{Y2}(+1)] indicates a linear cytosine-containing DNA sequence, in which dR^{N3} is present in position -3 instead of T and dA^{Y2} is in position +1 instead of dA of a parent dATTTCATTT sequence. The letter X at the end of its name, i.e. dC[R^{N3}(-3),A^{Y2}(+1)]X means that it is a cross-linked version of the sequence (**Figure 3.11**). Short oligos, 4-mer and 3-mer, are based on the shortened parent DNA sequence to 4-mer (dTTCA) and 3-mer (dTCA), which means that 4-mer FdZ[C^{N3}(-2),H^{E}(+1)]X is a cross-linked oligo obtained from a linear oligo dC^{N3} T FdZ dH^{E}.

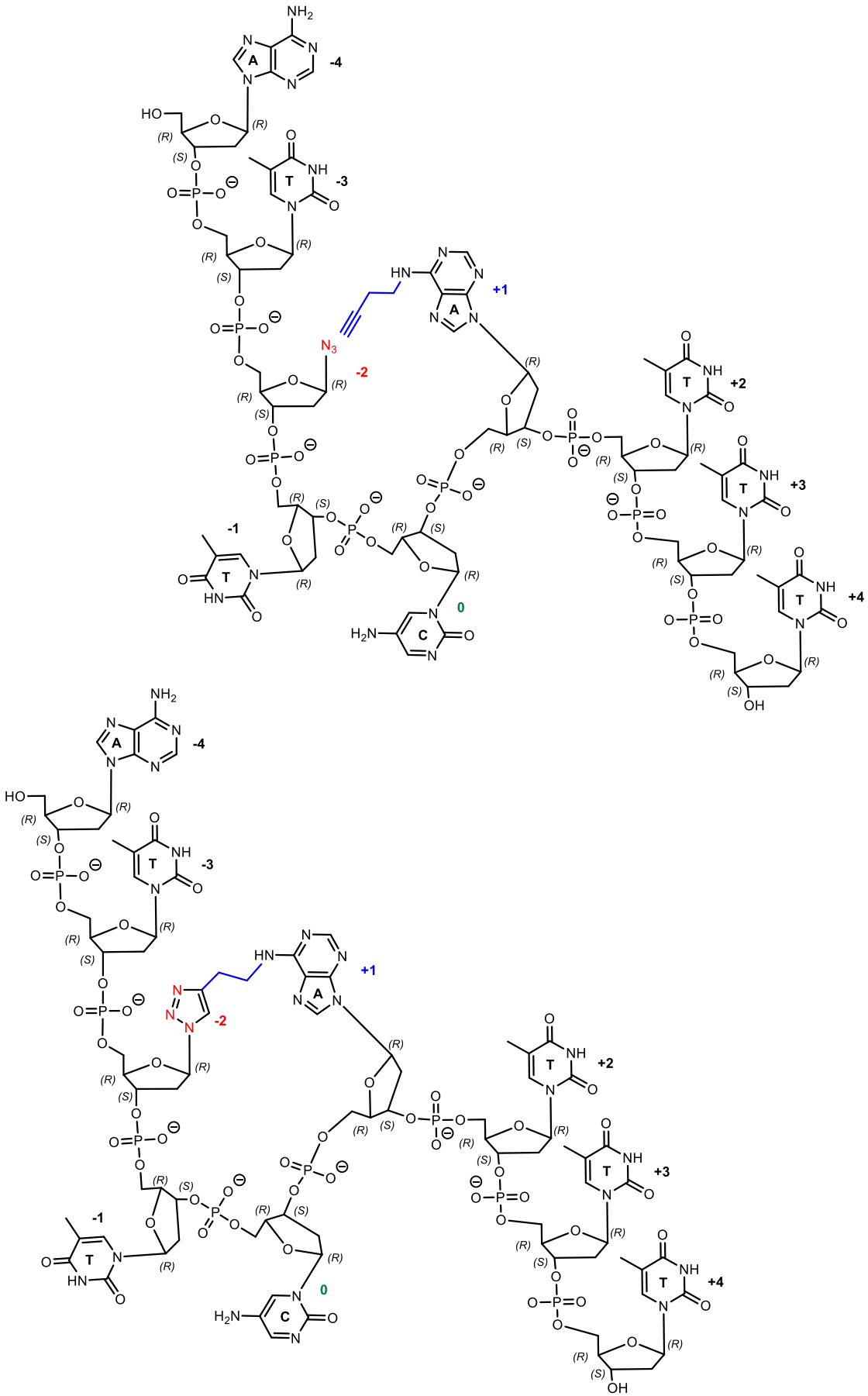


Figure 3.11. Schematic representation of linear dC[R^{N3}(-2),A^{Y2}(+1)] (top) and the cross-linked dC[R^{N3}(-2),A^{Y2}(+1)]X oligos (bottom).

These results agree with previously published properties of cyclized DNA, which were also prepared using CuAAC.[157] After separation by reverse-phase HPLC, the products were desalted and analyzed by ESI-MS showing, as expected, the same mass as for the starting materials, confirming the composition of monomeric cross-linked oligos (**Table 3.4** and **Table 3.5**). Abbreviation of individual oligos is explained in the footnotes to **Table 3.4** and **Table 3.5**.

3.2.3. Proof of cross-linking by RP-HPLC, HRMS and NMR experiments using an oligo synthesized for pilot studies

As the internal cross-linking strategy of oligos has rarely been performed in the past, we decided to perform a pilot study of an oligo containing an alkyne attached to 2'-deoxyadenosine (**dA^{Y2}**) at the +1 position and our azido sugar (**dR^{N3}**) at the -2 position creating cross-link 1, *i.e.* dC[R^{N3}(-2),A^{Y2}(+1)]X (**Figure 3.12**). We performed a trial cross-linking of the sequence using CuAAC, monitored the reaction using reverse phase HPLC, and isolated the product peak, which gave the same mass as that of the starting material, as the product and the starting material have the same atomic composition (**Figure 3.13** and **Figure 3.14**). While this was one of the proofs of cross-linking, we decided to use NMR experiments to provide further evidence of the successful cross-linking.

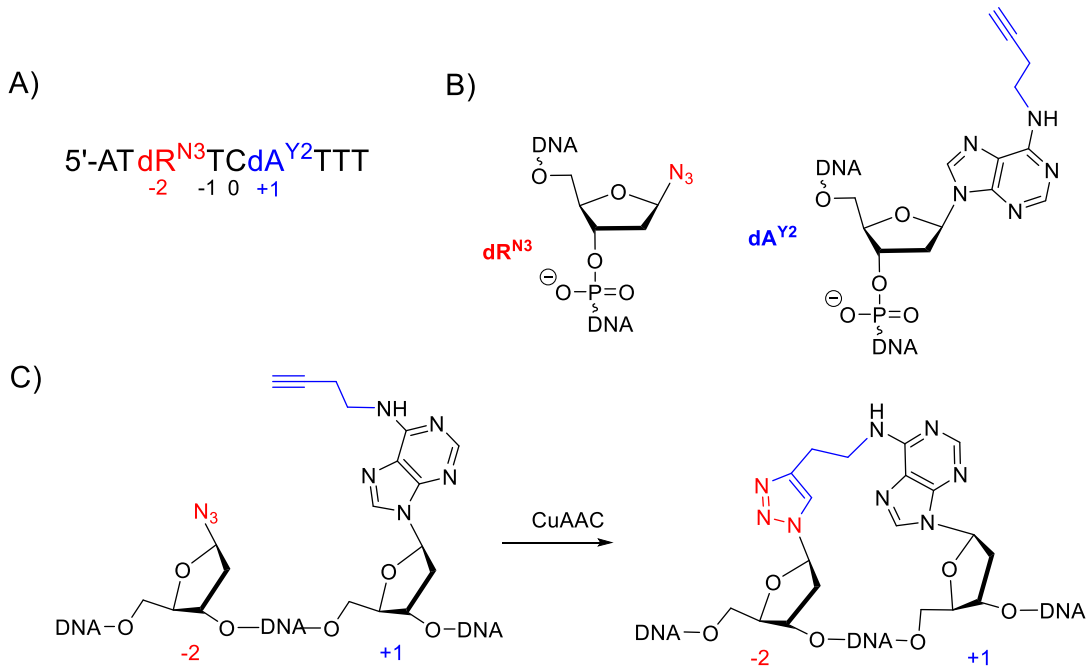


Figure 3.12. Structure of modified nucleosides and the sequence used for NMR experiments to prove creation of a cross-link in the oligo.

(A) Sequence of a linear oligo dC[R^{N3}(-2),A^{Y2}(+1)] used for internal cross-linking pilot study. (B) Chemical structure of modifications used in the sequence. (C) A representation of the cross-linking reaction using CuAAC.

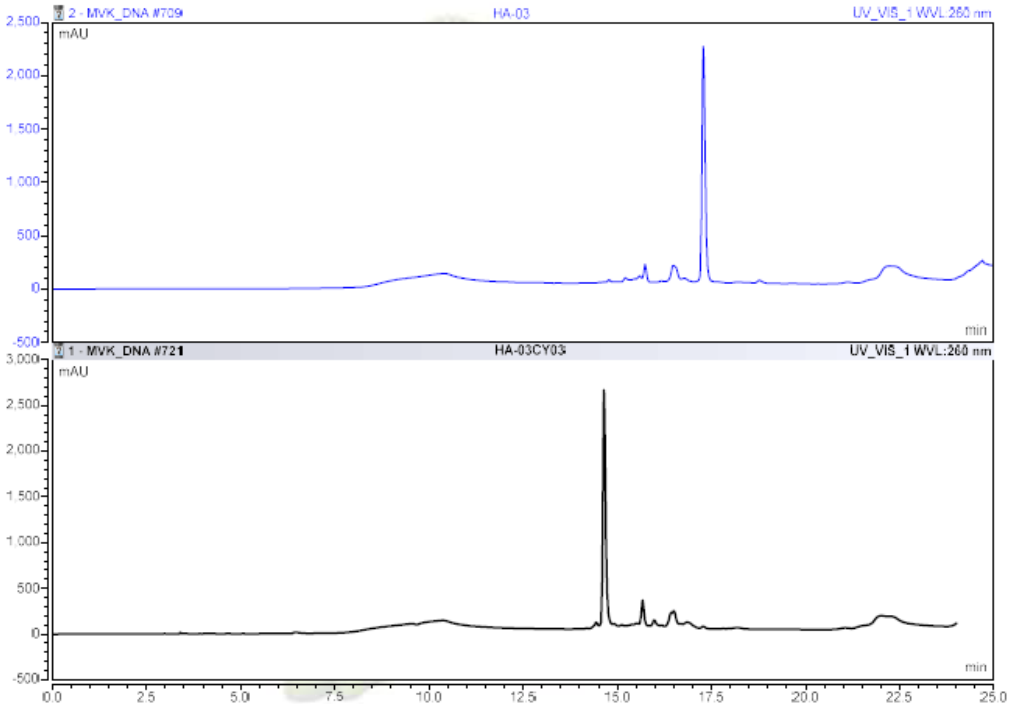


Figure 3.13. RP-HPLC profile showing a difference in retention time before and after cross-linking reaction.

RP-HPLC profiles (on 250/4.6 mm, C18 column) of oligo used in pilot study before (top) and after (bottom) CuAAC reaction showing the difference in retention times of the starting material and the cross-linked product.

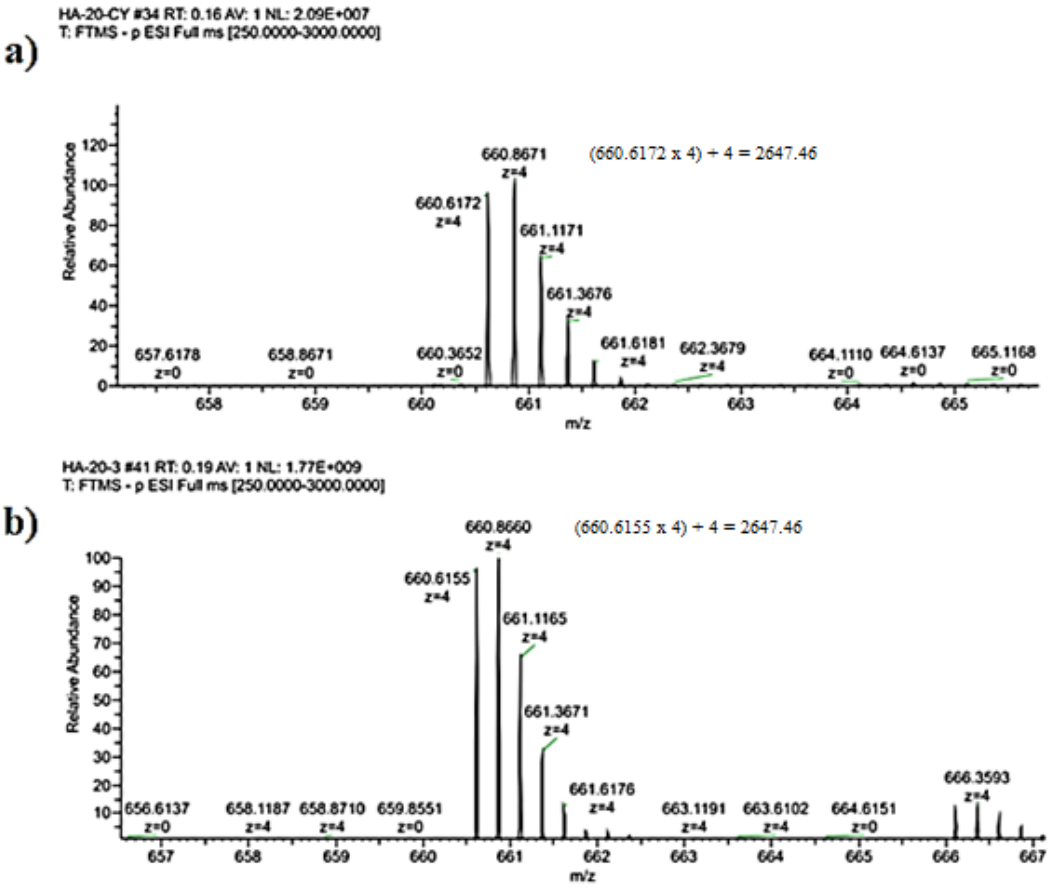


Figure 3.14. ESI-HRMS spectra of linear oligo (a) and cross-linked oligo (b) done as a part of the pilot study.

^{1}H NMR of linear oligo and cross-linked oligo showed an obvious difference in the overall chemical shifts (**Figure 3.15** and **Figure 3.16**).

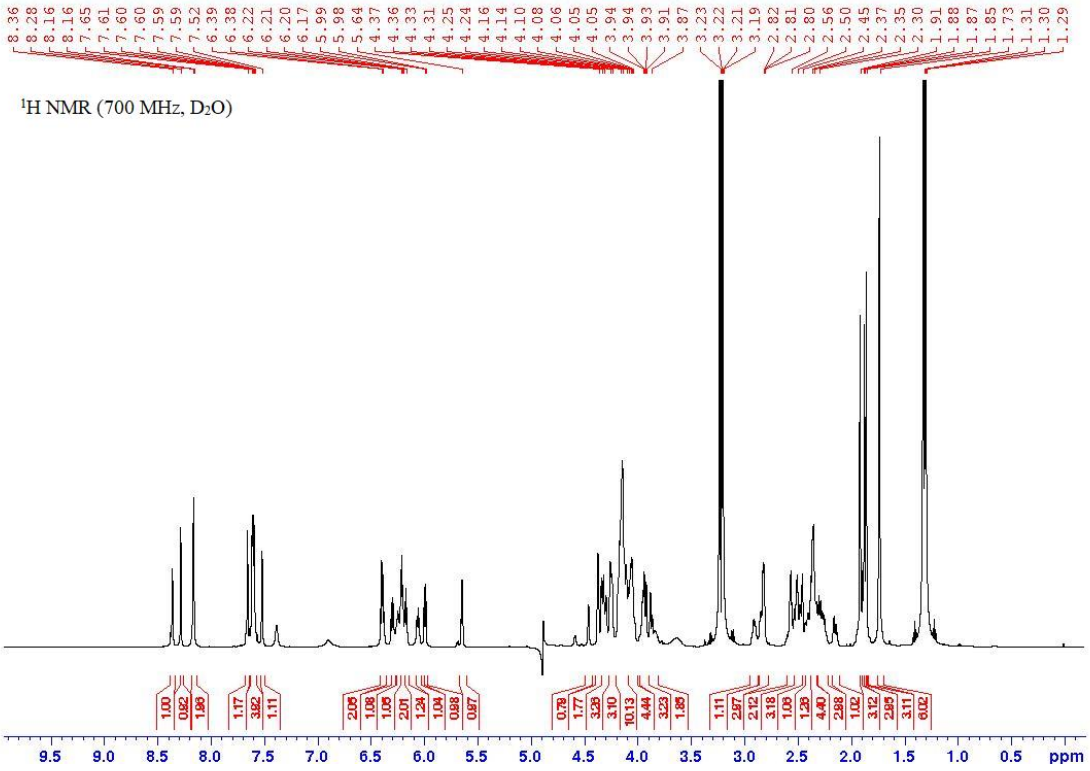


Figure 3.15. ^{1}H NMR spectrum of the linear oligonucleotide dC[R^{N3}(-2),A^{Y2}(+1)] used in internal cross-linking pilot study.

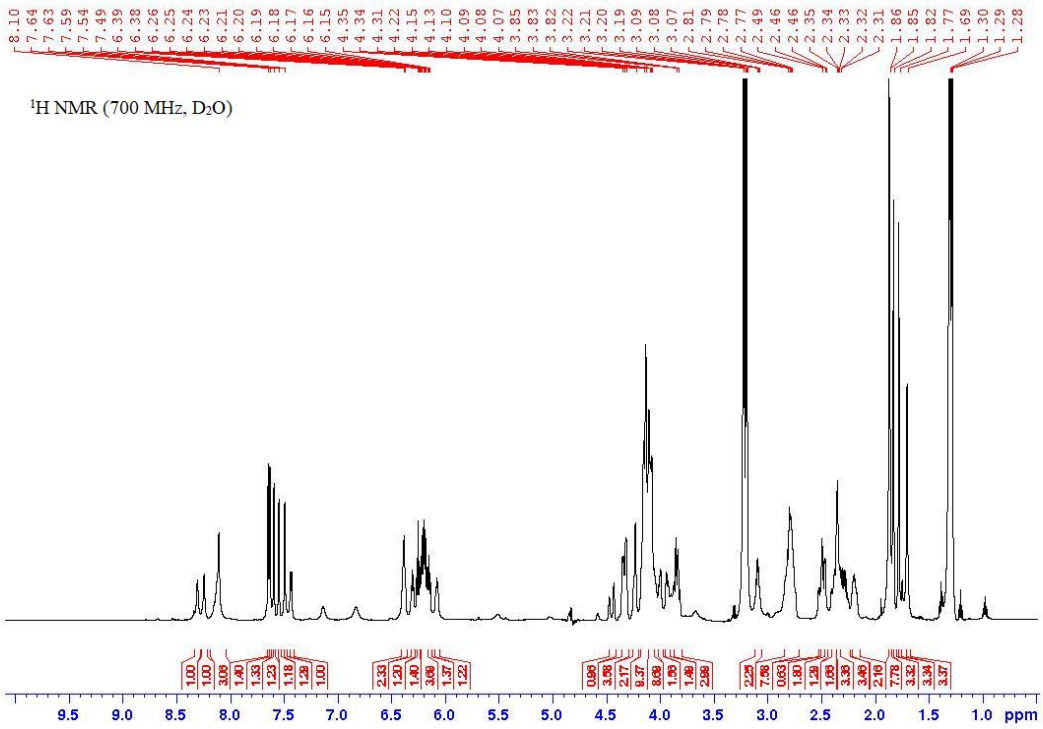


Figure 3.16. ^{1}H NMR spectrum of the cross-linked oligonucleotide dC[R^{N3}(-2),A^{Y2}(+1)]X used in internal cross-linking pilot study.

All the methyl peaks arising from the thymidines in the sequence were shifted (between 1.6 to 2.0 ppm) in the cross-linked oligo compared to the linear oligo (**Figure 3.17**), suggesting the formation of a new product (cross-linked oligo) as a result of CuAAC reaction. Thymidine is the only nucleotide in the sequence that has a methyl group, which can be easily identified in the 1H NMR by their characteristic chemical shifts.

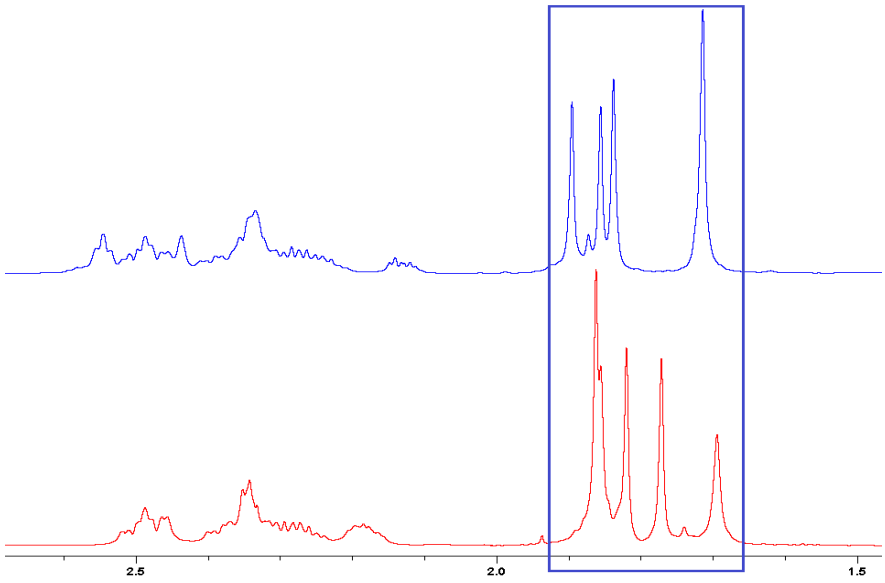


Figure 3.17. ^{1}H NMR chemical shifts of methyl peaks arising from five thymidines in linear (top) and cross-linked (bottom) oligos.

The anomeric proton of the azide 2'-deoxyribose in the linear oligo (triplet at 5.64 ppm, J = 5.60 Hz) and anomeric proton of the triazole nucleoside on the cross-linked oligo (6.16-6.18 ppm as a part of a multiplet) had different chemical shifts (**Figure 3.18**). Assignment of these peaks was done by 2D NMR experiments.

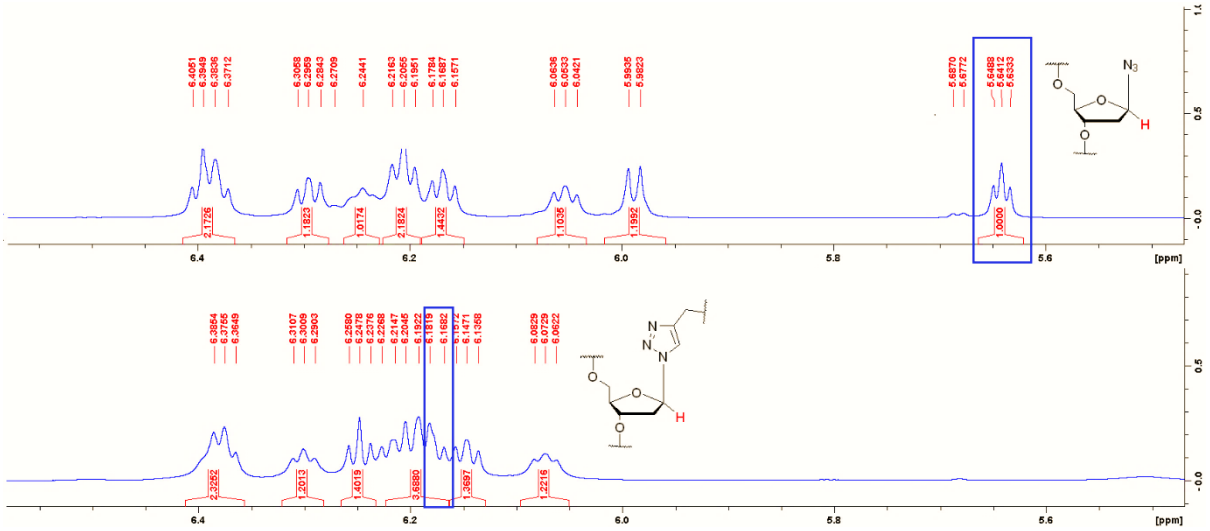


Figure 3.18. ^{1}H NMR spectra showing characteristic chemical shifts of anomeric protons of azido nucleotide (top) and triazole nucleotide after CuAAC (bottom).

In the linear oligonucleotide only anomeric proton at 5.64 ppm has no cross peaks with the aromatic region as there is no aromatic nucleobase in the azide nucleotide as shown by the HMBC and NOESY NMR experiments (**Figure 3.19** and **Figure 3.20**).

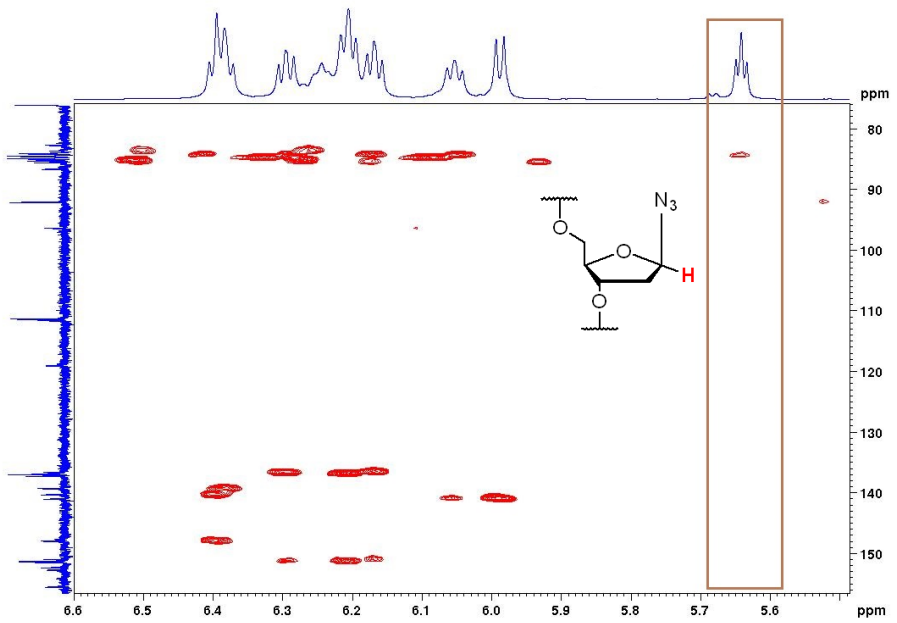


Figure 3.19. HMBC NMR spectrum showing no cross-peaks between the anomeric proton of azido nucleoside and aromatic carbons.

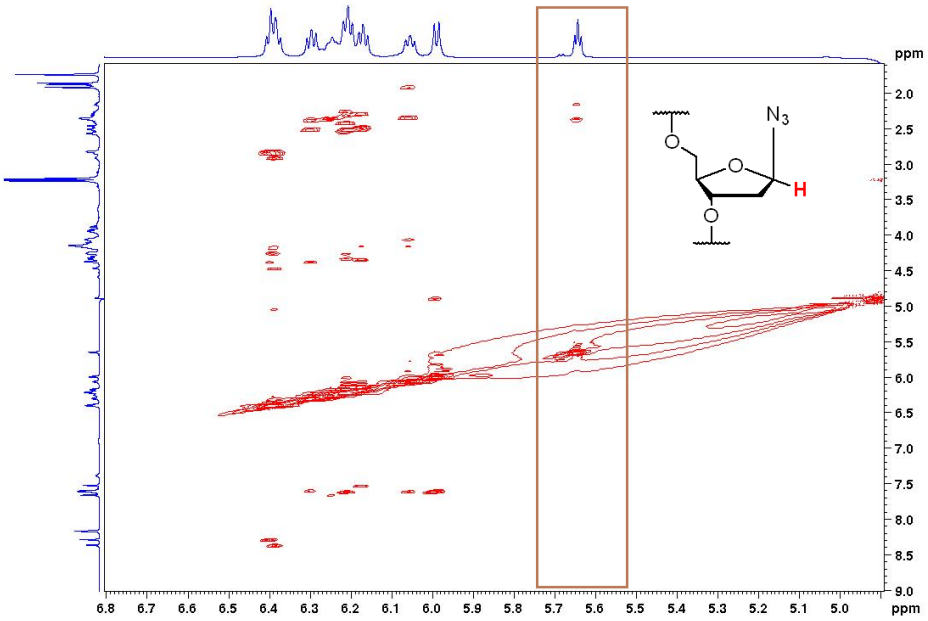


Figure 3.20. NOESY NMR spectrum showing no cross-peaks between anomeric proton of azido nucleotide and aromatic protons (see area in the box).

When it was cross-linked, there was an additional proton (8.13 ppm) in the aromatic region together with four aromatic protons of 2'-deoxyadenosines (**Figure 3.21**). This proton has a cross peak with the proton at 6.18 ppm in the NOESY NMR spectrum, which is the anomeric proton of the triazole nucleotide (**Figure 3.22)**

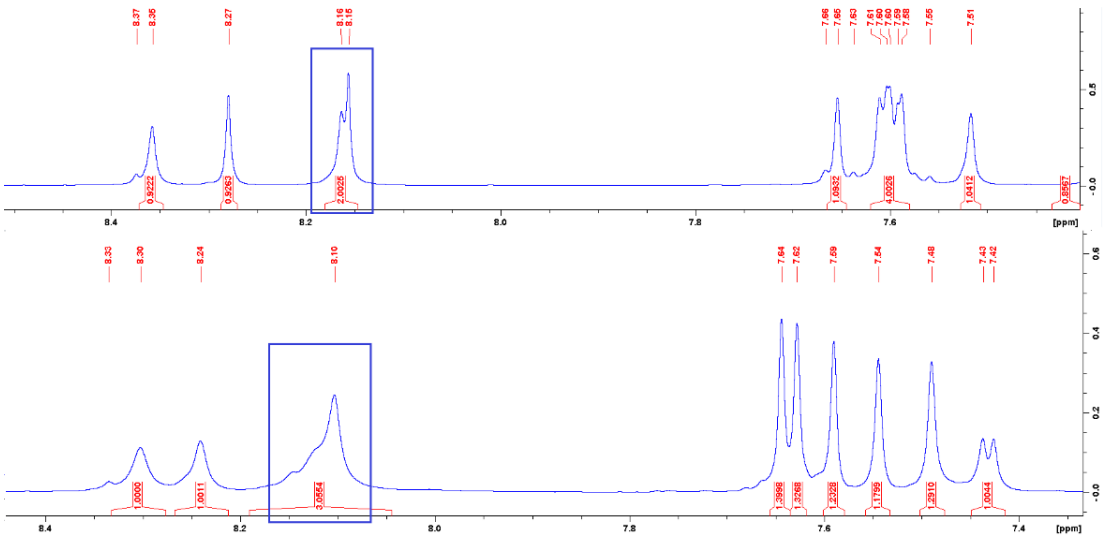


Figure 3.21. Comparison of the aromatic region of linear and cross-linked oligo in ^{1}H NMR.

^{1}H NMR spectra showing the aromatic protons without the triazole proton in the linear oligonucleotide (top) and with the triazole proton (bottom) in the cross-linked oligonucleotide. Area in the box shows appearance of an extra proton that belongs to the triazole in the cross-linked oligo, in comparison with the linear oligo (top).

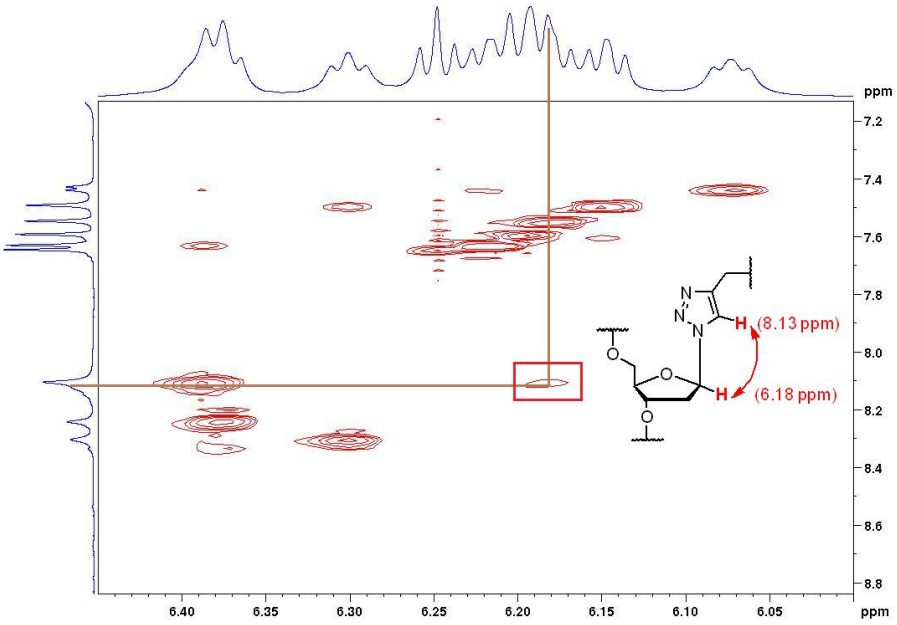


Figure 3.22. NOESY NMR spectrum showing cross-peaks of a triazole proton and the anomeric proton of the triazole nucleotide in the cross-linked oligo.

In addition, acetylenic proton (2.44 ppm) in the linear oligo was also assigned by HMBC NMR cross peak with the $\underline{C}H_2$ (18.30 ppm) immediately next to it (**Figure 3.23** and **Figure 3.24**). This characteristic peak was absent in the cross-linked oligo and provided additional evidence of the successful cross-linking by CuAAC.

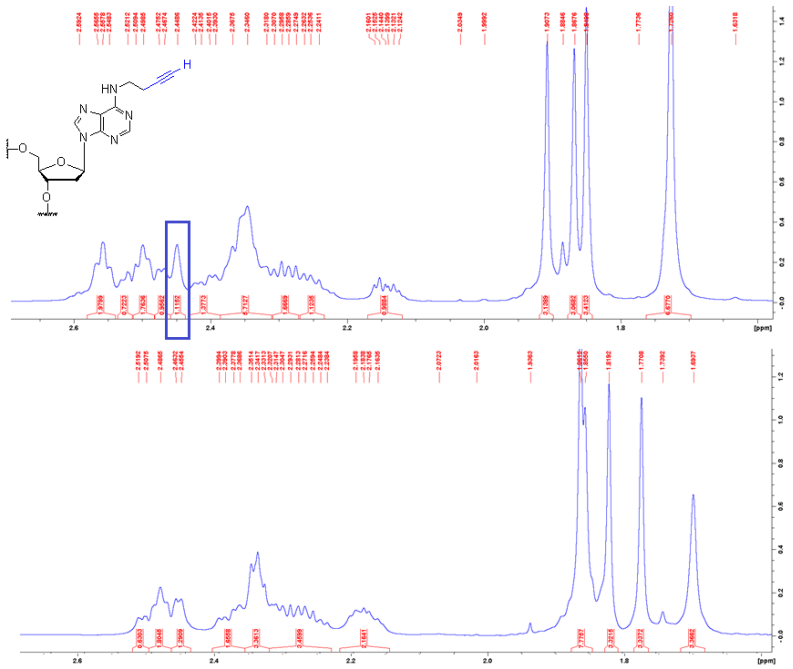


Figure 3.23. ^{1}H NMR showing acetylenic proton attached to 2'-deoxyadenosine nucleotide in the linear oligonucleotide (top) and the absence of it in the cross-linked oligo (bottom).

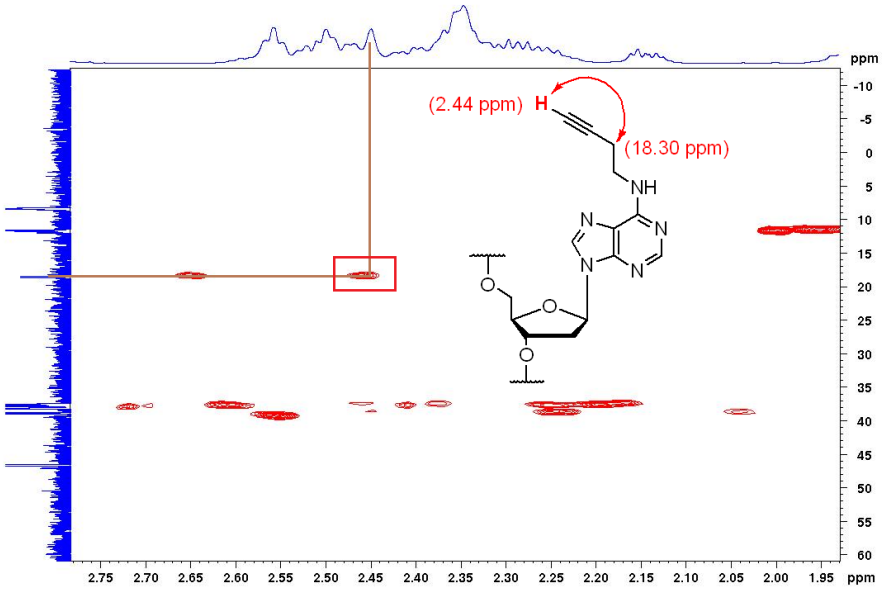


Figure 3.24. HMBC NMR spectrum showing cross-peaks of acetylenic proton attached to 2'-deoxyadenosine with the $\underline{C}H_2$ in the linear oligo.

The above NMR experiments together with reverse-phase HPLC retention time differences and HRMS establish the formation of the monomeric cross-linked oligo as a product of CuAAC reaction.

3.2.4. Design and synthesis of DNA hairpin as substrate and inhibitor of A3

It has been reported that A3A and $A3B_{CTD}$ bind to, and deaminate cytosine, in short 3-4 nucleotide loops of DNA hairpins,[96] which can be considered as natural mimics of cross-linked oligos. We synthesized a three-nucleotide loop DNA hairpin as a substrate (dC-hairpin, **Figure 3.25**) and as an inhibitor (FdZ-hairpin) and evaluated them together with our cross-linked oligos.

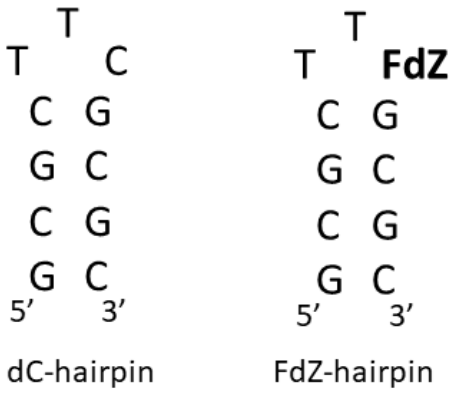


Figure 3.25. Schematic representation of dC and FdZ DNA hairpins.

The design of the hairpin was based on the fact that the loop consists of the required A3 recognition motif (5'-TTC) and a stem consisting of four G-C base pairing. The strong G-C base pairing greatly increases the thermal stability of the oligo to keep it in hairpin form in solution when compared to A-T base pairing. Also, it was observed that four G-C base pairs were necessary to have the hairpin to be stable. The oligo existing in the form of a hairpin was confirmed by ^{1}H-NMR experiment, where four G-C base-paired hydrogen-bonded protons were observed at 13.04-13.15 ppm (**Figure 3.26**). ITC experiments done by Dr. Stefan Harjes with various hairpin oligos proved that this oligo binds well to A3 enzymes.

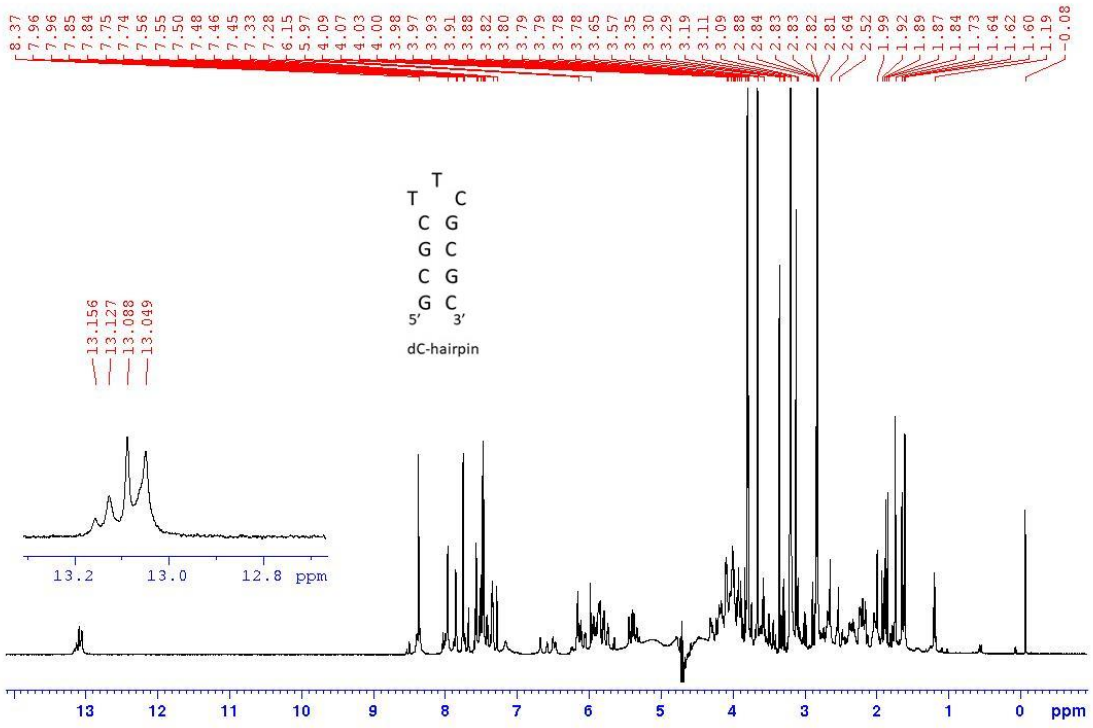


Figure 3.26. ^{1}H NMR (phosphate buffer pH 6.0 containing 20% D_2O) spectrum of the dC-hairpin showing the G-C base-paired hydrogen-bonded protons.

3.2.5. Evaluation of DNA substrates by NMR-based activity assay

Using the 9-mer oligonucleotide 5'-AT_3**dC**AT_3 as the standard substrate, we compared the deamination of both terminally and internally cross-linked oligos using our previously described real-time NMR assay.[54, 124, 126] For the terminally cross-linked oligos, we observed decreased rates of deamination from 9-mer to 5-mer. One should note that number of phosphates of the modified 9-mer sequence is the same as for the 9-mer DNA control (**Figure 3.27A**). This means that terminally cross-linked oligos are bad mimics of the DNA shape seen in the complex, although the flexibility associated with the longer oligos allows some deamination to occur. A decline in deamination rate was observed for oligos having

cross-link 1 relative to linear substrate (**Figure 3.27B**). One should note that a triazole ring formed after CuAAC in -1…-3 positions of the oligos was designed to mimic a pyrimidine nucleobase. However, cross-linked oligo dC[R^{N3}(-1),A^{Y2}(+1)]X was not recognised as an A3 substrate possibly due to poor mimicking of dT^{-1} by a triazole. The next cross-link 2 was designed based on modified nucleotides dU^{E} and dA^{N3}.

In marked contrast, a 1.5 times faster rate of deamination was observed for the internally cross-linked oligo dC[U^{E}(-2),A^{N3}(+1)]X compared to the standard, linear substrate (**Figure 3.27C**). The -2 and +1 positions were spatially close in the X-ray structure of the inactive $A3B_{CTD}$-QM-ΔL3-AL1swap(E255A)/DNA complex.[36] Moreover, an extremely slow rate of deamination was detected for dC[U^{E}(-3),A^{N3}(+1)]X oligo, which we attribute to a shape and conformation that hinders proper binding with the enzyme compared to the dC[U^{E}(-2),A^{N3}(+1)]X oligo. We did not create a cross-link between dA at position +1 and T at position -1 as the X-ray structure (pdb: 5TD5) shows a clear binding pocket for T in the -1 position, from which we infer that cross-linking in this position might disturb the DNA-protein interaction. A similar increase in deamination rate to cross-link 2 was observed for a dC-hairpin DNA, which was deaminated 1.6 times faster than the linear substrate (**Figure 3.27D**). However, a marked 2.3 times increase in deamination rate was detected for oligo dC[C^{N3}(-2),H^{E}(+1)]X having cross-link 3 connecting the -2 and +1 positions in the DNA sequence (**Figure 3.27D**). The difference between cross-link 2 and 3 is that in cross-link 2, there is a modified 2'-deoxyuridine nucleotide (**dU^{E}**) and at the position +1 there is a modified 2'-deoxyadenosine (**$dA^{Y2\ or\ 4}$**) present, whereas in cross-link 3 these positions are substituted by a modified 2'-deoxycytosine (**dC^{N3}**) and a carbazole-based nucleotide (**dH^{E}**) respectively. It is surprising that synthetic oligo dC[C^{N3}(-2),H^{E}(+1)]X was deaminated even faster than a native DNA hairpin, which may be attributed to additional interactions of carbazole nucleoside in cross-link 3 with amino acid side chains and associated water molecules.

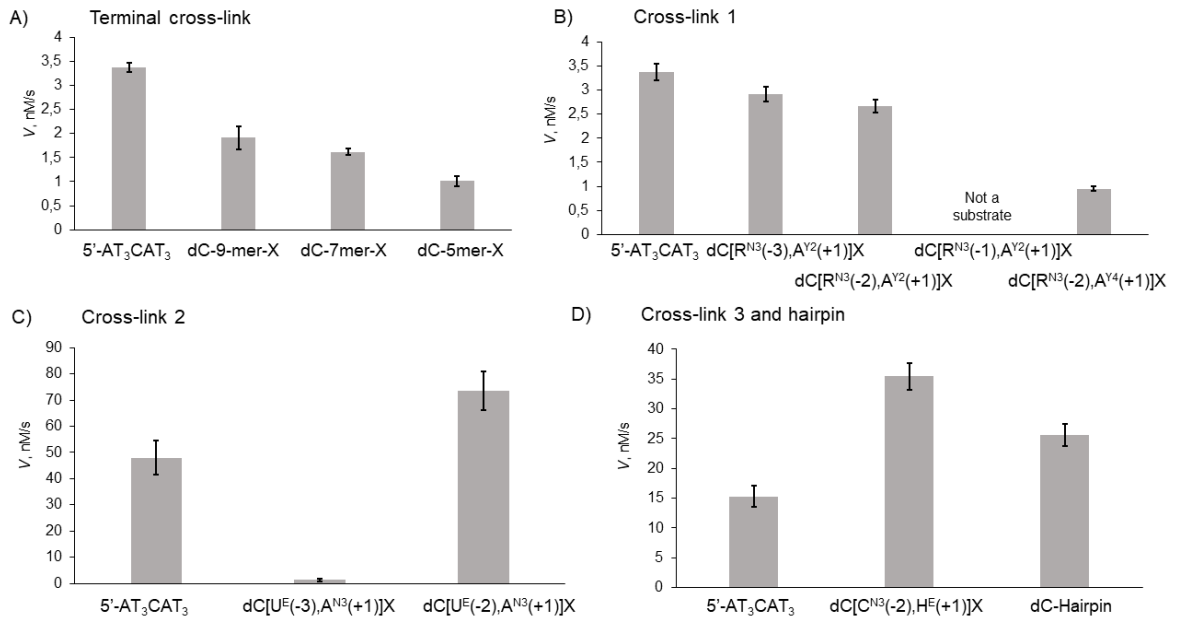


Figure 3.27. Initial rate of $A3B_{CTD}$-QM-ΔL3-AL1swap-catalyzed deamination of terminally and internally cross-linked oligos.

(A) Terminally and, (B-D) internally cross-linked oligos and a dC-hairpin in comparison with the linear substrate. Substrate concentrations are 100 μM in (A and B), 800 μM in (C) and 400 μM in (D). Enzyme concentrations are 50 nM in (A and B) and 300 nM in (C) and 100 nM in (D). Buffers are 50 mM citrate-phosphate (pH 5.5 for A-C) and 50 mM Na-phosphate (pH 6.0 for D), 200 mM NaCl, 2 mM β-mercaptoethanol, 200 μM 4,4-dimethyl-4-silapentane-1-sulfonic acid (DSS) and 10 % D_2O at 25 °C. Error bars are estimated standard deviations from triplicate measurements.

3.2.6. Evaluation of inhibitors by NMR assay

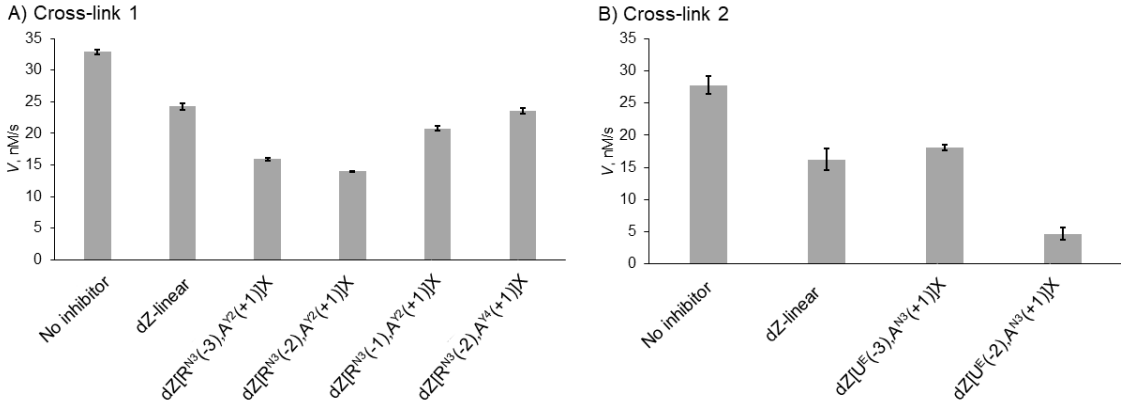
The increased speed of deamination of several cross-linked oligos prompted us to investigate the inhibition potential of these constructs by changing dC to dZ or FdZ. We used the NMR assay to compare the internally cross-linked dZ/FdZ-containing oligos with a linear DNA inhibitor containing dZ/FdZ (dZ-linear is 5'-AT_3**dZ**AT_3, FdZ-linear is 5'-AT_3**FdZ**AT_3), which was characterized earlier.[54, 124] Residual activity of $A3B_{CTD}$-QM-ΔL3-AL1swap on the unmodified oligo (5'-T_4**dC**AT) as a substrate in the presence of a known concentration of dZ/FdZ-containing inhibitors (linear and cross-linked) was measured using the NMR assay.

The results revealed that the dZ-containing oligo with an internal cross-link 1, oligos dZ[R^{N3}(-3),A^{Y2}(+1)]X and dZ[R^{N3}(-2),A^{Y2}(+1)]X, inhibited A3-catalyzed deamination to half the speed of an uninhibited A3 -catalyzed deamination reaction. In contrast, the oligos dZ[R^{N3}(-1),A^{Y2}(+1)]X and dZ[R^{N3}(-2),A^{Y4}(+1)]X derived from the slower deaminated substrates inhibited A3-catalyzed deamination similar to a linear dZ-containing oligo (dZ-linear). In cross-link 2, the dZ-containing oligo derived from the faster-deaminated substrate dZ[U^{E}(-2),A^{N3}(+1)]X inhibited the A3-catalyzed deamination by greater than an order of the magnitude more potently than our best linear dZ oligo (**Figure 3.28B**). On the other hand, dZ[U^{E}(-3),A^{N3}(+1)]X was found to be a much weaker inhibitor, in line with the earlier

observation that dC[U^{E}(-3),A^{N3}(+1)]X was a very poor substrate. Interestingly, its inhibitory potential is comparable with the linear oligo, probably because dZ interacts strongly with the active site of the enzyme.

The fastest-deaminated substrate dC[C^{N3}(-2),H^{E}(+1)]X based on cross-link 3 was converted to the most powerful A3 inhibitors studied here having dZ or FdZ nucleotides instead of dC (**Figure 3.28C**). It is notable that FdZ-containing oligo FdZ[C^{N3}(-2),H^{E}(+1)]X exhibited higher inhibitory potential than dZ-containing oligo dZ[C^{N3}(-2),H^{E}(+1)]X, which is in line with our previous observations[124] that FdZ-linear oligo is a more powerful inhibitor of $A3B_{CTD}$-QM-ΔL3-AL1swap than dZ-linear oligo (**Figure 3.28C**). It is also interesting that FdZ-hairpin DNA was not as powerful an inhibitor as cross-linked oligo FdZ[C^{N3}(-2),H^{E}(+1)]X, which also correlates with lower deamination rate for dC-hairpin in comparison with dC[C^{N3}(-2),H^{E}(+1)]X (**Figure 3.27D**).

Encouraged by these results we synthesized short 3-mer and 4-mer oligo inhibitors based on cross-link 3 that contained dZ or FdZ. The rationale behind the design of this short oligo was to create an inhibitor which is like a small-molecule inhibitor of A3. The strategy was to synthesize an oligo which makes the essential interactions with the A3 enzyme. The 4-mer FdZ[C^{N3}(-2),H^{E}(+1)]X was a less potent inhibitor than the cross-linked 9-mer oligo but its inhibitory potential was close to that for the FdZ-hairpin. A further decline in inhibition of $A3B_{CTD}$-QM-ΔL3-AL1swap-catalyzed C-to-U deamination was observed upon shortening of the DNA length to the 3-mer (-1,+1 link) and for the non-cross-linked oligos (linear oligos containing modified nucleotides before cross-linking).



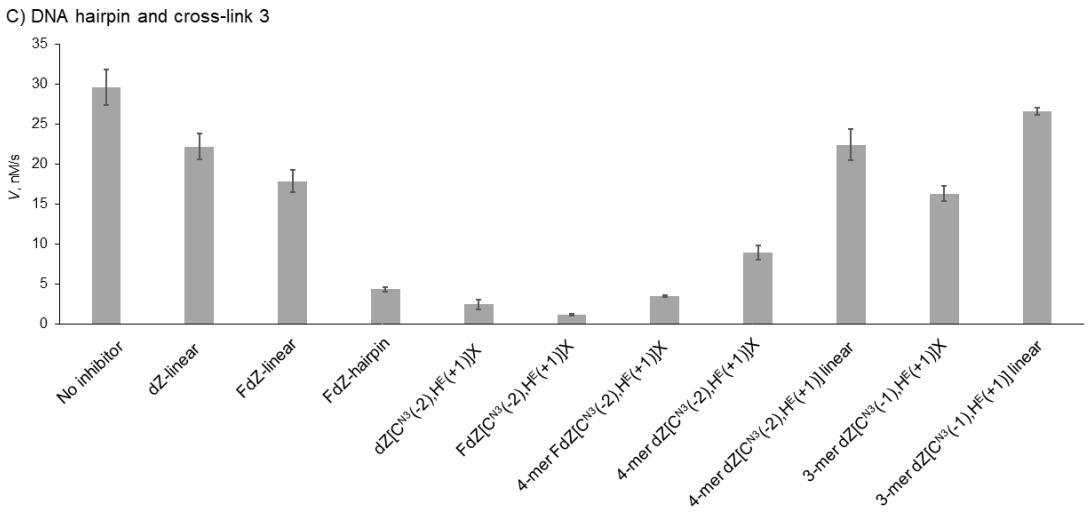


Figure 3.28. Inhibition of $A3B_{CTD}$-QM-ΔL3-AL1swap-catalyzed deamination of 5'-T_4CAT by dZ/FdZ-containing linear and cross-linked oligos.

Initial rate of deamination was measured under the following experimental conditions: 400 μM of 5'-T_4CAT, 8 μM (A, B) or 4 μM (C) of dZ/FdZ-containing oligos, 300 nM (A, B) or 200 nM (C) of $A3B_{CTD}$-QM-ΔL3-AL1swap in a 50 mM sodium phosphate buffer (pH 6.0) containing 100 mM NaCl, 2.5 mM β-mercaptoethanol, 50 μM 3-(trimethylsilyl)-2,2,3,3-tetradeuteropropionic acid (TSP) at 25 °C. Error bars are estimated standard deviations from triplicate measurements.

To determine the inhibition constants (K_i) of inhibitors, we first evaluated kinetic parameters of $A3B_{CTD}$-QM-ΔL3-AL1swap on two DNA substrates (**Table 3.6**), which were used to detect the residual deamination rate in the presence of inhibitors (**Table 3.7**).

Table 3.6. Kinetic parameters of $A3B_{CTD}$-QM-ΔL3-AL1swap on two DNA substrates used in this work

Oligo	Parameters	Linear analysis (Lineweaver-Burk plot)	Non-linear regression analysis
5'-AT_3dCAT_3	K_m, μM	200 ± 30	150 ± 30
	V_{max}, μM/s	0.014 ± 0.002	0.0126 ± 0.0008
	k_{cat}, s^{-1}	0.29 ± 0.04	0.26 ± 0.02
	k_{cat}/K_m, s^{-1} μM^{-1}	0.0014	0.0018
5'-T_4dCAT	K_m, μM	350 ± 70	290 ± 70
	V_{max}, μM/s	0.056 ± 0.008	0.051 ± 0.005
	k_{cat}, s^{-1}	0.19 ± 0.03	0.171 ± 0.015
	k_{cat}/K_m, s^{-1} μM^{-1}	0.0005	0.0006

Fitting the experimental data by non-linear least-squares using a global fit provided similar values of kinetic parameters to those obtained by the linearized Lineweaver-Burk plot (**Table 3.6**). Then, the inhibition constants for strongest cross-linked inhibitors were calculated by several methods assuming a competitive mode of inhibition[54] (**Table 3.7**) based on experiments in which the concentration of the individual inhibitor was varied, and the residual deamination initial rate was measured.

Table 3.7. The inhibition constants (K_i) for dZ/FdZ-containing oligo inhibitors of $A3B_{CTD}$-QM-ΔL3-AL1swap obtained by various calculation methods[a)]

Oligo name	Parameters	Lineweaver-Burk plot for the substrate/Dixon plot for the inhibitor	Non-linear regression analysis for both substrate and inhibitor
dZ-linear	K_i, μM	7.5 ± 1.7	7.0 ± 1.3
	V_{max}, μM/s		0.0126 ± 0.0008
FdZ-linear	K_i, μM	2.1 ± 0.8	4.4 ± 0.7
	V_{max}, μM/s		0.0126 ± 0.0008
Cross-link 2			
dZ[U^E(-2),A^{N3}(+1)]X	K_i, μM	0.66 ± 0.14	0.69 ± 0.14
	V_{max}, μM/s		0.052 ± 0.006
FdZ-hairpin	K_i, μM	0.28 ± 0.05	0.35 ± 0.06
	V_{max}, μM/s		0.052 ± 0.005
Cross-link 3			
FdZ[C^{N3}(-2),H^E(+1)]X	K_i, μM	0.075 ± 0.015	0.100 ± 0.016
	V_{max}, μM/s		0.051 ± 0.004
4-mer FdZ[C^{N3}(-2),H^E(+1)]X	K_i, μM	0.25 ± 0.05	0.28 ± 0.05
	V_{max}, μM/s		0.051 ± 0.005

a) dZ/FdZ-linear inhibitors were evaluated against 5'-AT_3**dC**AT_3 substrate using 50 nM of $A3B_{CTD}$-QM-ΔL3-AL1swap; all other inhibitors were evaluated against 5'-T_4**dC**AT substrate using 200 nM of $A3B_{CTD}$-QM-ΔL3-AL1swap.

All calculation methods revealed that cross-linked and hairpin dZ/FdZ oligos were more powerful inhibitors of $A3B_{CTD}$-QM-ΔL3-AL1swap than linear oligos. The K_i values calculated for FdZ[C^{N3}(-2),H^E(+1)]X were lowest among inhibitors and K_i of 0.100 ± 0.016

μM obtained using non-linear regression analysis is around 44 times lower than that for the linear **FdZ**-oligo (4.4 ± 0.7 μM).[124] The nanomolar inhibitory potential exhibited by our internally cross-linked oligos shows that properly pre-shaped, cross-linked DNA can become better inhibitors than the linear one and that the better substrate becomes the better inhibitor, which supports our original hypothesis.

To further characterize our best inhibitor FdZ[C^{N3}(-2),H^{E}(+1)]X we performed deamination of dC-hairpin (**Figure 3.29**), which is a faster deaminating substrate than linear DNA, using a much lower concentration of enzyme than in previous assays (20 – 27 nM versus 200 nM) in the absence and presence of inhibitor (100 nM). The kinetic data, measured over time until exhaustion of the substrate, were analyzed by global regression using Lambert's W function, which provides superior estimates of observed K_m and V_{max} than non-linear regression analysis of initial rate data.

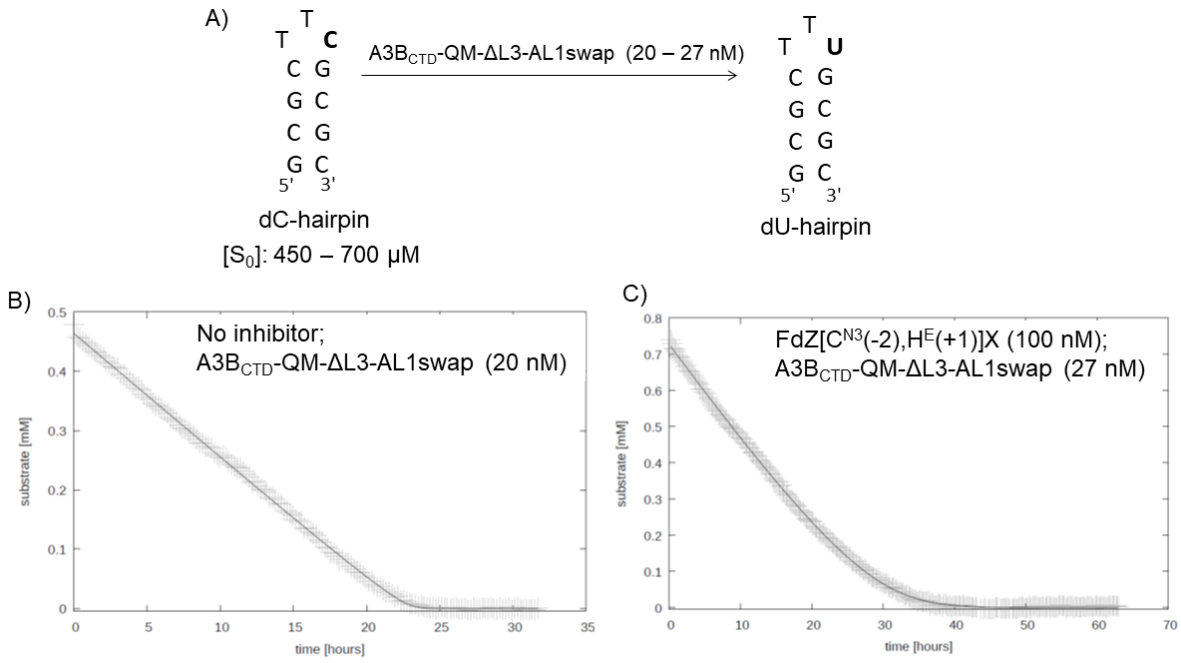


Figure 3.29. Inhibition of A3B$_{CTD}$-QM-ΔL3-AL1swap-catalyzed deamination of dC-hairpin by FdZ[C^{N3}(-2),H^{E}(+1)]X in the continuous NMR assay.

A) Schematic representation of A3B$_{CTD}$-QM-ΔL3-AL1swap-catalyzed deamination of dC-hairpin to dU-hairpin. B) dC-hairpin consumption over time in the absence of inhibitor; C) dC-hairpin consumption over time in the presence of FdZ[C^{N3}(-2),H^{E}(+1)]X (100 nM). Experimental results are shown by crosses and the global regression using Lambert's W function is shown by a solid line in B) and C). Deamination of dC-hairpin over time was measured under the following experimental conditions: 400 – 700 μM of dC-hairpin, 20 – 27 nM of A3B$_{CTD}$-QM-ΔL3-AL1swap in a 50 mM Na^+/K^+ phosphate buffer (pH 6.0) containing 100 mM NaF, 1 mM TCEP, 10% D_2O and 100 μM 3-(trimethylsilyl)-2,2,3,3-tetradeuteropropionic acid (TSP) at 25 °C.

Under the given experimental conditions, it is evident that the dC-hairpin is a better substrate for A3B$_{CTD}$-QM-ΔL3-AL1swap than the linear oligos 5'-AT_3**dC**AT_3 and 5'-

T_4**dC**AT because its K_m is in low μM range and k_{cat} of enzyme is similar to these substrates. The specificity constant defined by k_{cat}/K_m reinforces that the $A3B_{CTD}$-QM-ΔL3-AL1swap is ~ 25 times more efficient at deaminating the dC-hairpin than the linear 5'-AT_3**dC**AT_3. In the presence of an inhibitor (100 nM), V_{max} was unchanged (**Table 3.8**) which indicates that FdZ[C^{N3}(-2),H^{E}(+1)]X is a competitive inhibitor of $A3B_{CTD}$. The calculated inhibition constant (K_i) from the observed K_m values in the absence and in the presence of inhibitor was found to be in the low nM range (K_i = 6.8 ± 1.4 nM), which again shows that FdZ[C^{N3}(-2),H^{E}(+1)]X is a powerful inhibitor of $A3B_{CTD}$.

Table 3.8. Kinetic parameters of $A3B_{CTD}$-QM-ΔL3-AL1swap-catalyzed deamination of dC-hairpin in the absence and presence of FdZ[C^{N3}(-2),H^{E}(+1)]X obtained using Lambert's W function [a)]

Inhibitor	V_{max}, nM/s	k_{cat}, s^{-1}	K_m (observed), μM	K_i (calculated), nM	k_{cat}/K_m, s^{-1} μM^{-1}
No inhibitor	8.0 ± 0.4	0.296 ± 0.015	6.6 ± 0.7	-	0.045
FdZ[C^{N3}(-2),H^{E}(+1)]X (100 nM)	8.4 ± 0.4	0.311 ± 0.015	104 ± 5	6.8 ± 1.4	-

a) see experimental section for analysis of experimental data.

3.2.7. Binding of cross-linked oligos to A3 enzymes

We expected that the binding of inactivated $A3B_{CTD}$-QM-ΔL3-AL1swap to the substrate would be close to or even worse than K_m of 200 μM, which places substrate binding outside ITC applicability. The results of ITC experiments are presented in **Table 3.9**. One should emphasise that the composition of buffers used for NMR kinetic and ITC binding experiments were different, mainly because the proteins used in ITC experiments need to be at a higher concentration than for the kinetic assays (10-100 μM for ITC *versus* 20-300 nM in kinetic experiments). This required a search for buffer components that prevented protein precipitation during the ITC titrations. Our optimized ITC buffers were also different for A3A(E72A) and $A3B_{CTD}$-QM-ΔL3-AL1swap, which means that the K_d and K_i values cannot be directly compared. However, our aim was to compare the binding of different oligos to A3 to ascertain if general trends can be established.

Table 3.9. Dissociation constant of oligos from A3 enzymes obtained by isothermal titration calorimetry (ITC).[a]

Oligo	**Protein**	**ITC buffer[b]**	***K*d (μM)**
5'-AT_3**dC**AT_3	A3A-E72A	1	0.20 ± 0.06
5'-T_4**dC**AT	A3A-E72A	1	0.19 ± 0.04
dC-9mer-X	A3A-E72A	1	0.26 ± 0.06
dC-7mer-X	A3A-E72A	1	0.25 ± 0.06
dC-5mer-X	A3A-E72A	1	2.7 ± 0.6
Cross-link 1 dC[R^{N3}(-2),A^{Y2}(+1)]X	A3A-E72A	1	0.28 ± 0.06
Cross-link 1 dC[R^{N3}(-2),A^{Y4}(+1)]X	A3A-E72A	1	0.15 ± 0.06
5'-AT_3**dC**AT_3	A3A-E72A	2	0.12 ± 0.03
dC-hairpin	A3A-E72A	2	0.11 ± 0.03
Cross-link 1 dC[R^{N3}(-2),A^{Y2}(+1)]X		2	0.13 ± 0.04
Cross-link 1 dC[R^{N3}(-2),A^{Y4}(+1)]X	A3A-E72A	2	0.15 ± 0.03
Cross-link 3 dC[C^{N3}(-2),H^{E}(+1)]X	A3A-E72A	2	0.028 ± 0.006
5'-AT_3**dC**AT_3	A3A-E72A	2*	0.13 ± 0.03
Cross-link 2 dC[U^{E}(-2),A^{N3}(+1)]X	A3A-E72A	2*	0.027 ± 0.006
dZ-linear	$A3B_{CTD}$-QM-ΔL3-AL1swap	3	11.4 ± 2.3
Cross-link 2 dZ[U^{E}(-2),A^{N3}(+1)]X	$A3B_{CTD}$-QM-ΔL3-AL1swap	3	2.5 ± 0.5

[a] K_d is the dissociation constant for the ligand from A3. [b] ITC buffer 1: 50 mM MES, pH 6.0, 100 mM NaCl, 200 μM EDTA, 1 mM β-mercaptoethanol; ITC buffer 2: 50 mM Na^+/K^+ phosphate, pH 6.0, 50 mM NaCl, 50 mM choline acetate, 2.5 mM TCEP, 200 μM EDTA with 30 mg/mL bovine serum albumin (after preparation, buffer was frozen and defrosted before the experiments); ITC buffer 2* is freshly prepared buffer 2; ITC buffer 3: 50 mM Na^+/K^+ phosphate, pH 6.0, 200 mM trimethylamine *N*-oxide dihydrate, 2.5 mM tris(2-carboxyethyl)phosphine (TCEP), 200 μM EDTA.

The dissociation constants (K_d) obtained by ITC showed no gain in affinity by A3A-E72A for terminally cross-linked oligos in comparison with the standard linear oligos (5'-AT_3**dC**AT_3 and 5'-T_4**dC**AT). The dC-5mer-X had noticeably weaker binding to A3A-E72A, and a lower deamination rate as substrate to $A3B_{CTD}$-QM-ΔL3-AL1swap compared to its 7-mer and 9-mer analogues; this is partly consistent with previous observations on linear substrates of varying length[54, 60] but also with insufficient flexibility to achieve the right conformation, when a short oligo is cross-linked.

For oligos with cross-link 1 a slightly higher affinity was detected for the oligo with the four-carbon linker dC[R^{N3}(-2),A^{Y4}(+1)] in buffer 1 compared to the two-carbon linker dC[R^{N3}(-2),A^{Y2}(+1)] but in buffer 2 both four and two carbon linker oligos had affinities similar to that of the four-carbon linker oligo in buffer 1. It is worth mentioning that the substrates with cross-link 1 were not deaminated faster than the linear substrate by $A3B_{CTD}$-QM-ΔL3-AL1swap (**Table 3.9**), which was confirmed by a similar affinity of oligos with cross-link 1 by ITC.

The dC-hairpin and linear oligo 5'-AT_3**dC**AT_3 had similar affinity to A3A(E72A) in buffer 2 (pH 6.0). It has been reported that hairpins are preferred substrates for wild-type A3A at pH 7.0 indicating that buffer pH might have a prevalent role in discrimination between linear and hairpin DNA.

dC-Oligo with cross-link 3 showed four-times improved affinity to A3A(E72A) in comparison with linear DNA. This agrees with the improved deamination of this oligo by $A3B_{CTD}$-QM-ΔL3-AL1swap.

We noticed that the use of freshly prepared buffer 2 (marked as 2* in the **Table 3.9**) led to significantly improved stability of A3A(E72A) in solution. It did not affect the affinity of the linear oligo 5'-AT_3**dC**AT_3 to A3A(E72A) as K_d values in buffers 2 and 2* were similar (K_d = 0.12 ± 0.03 and 0.13 ± 0.03 μM, respectively). The oligo dC[U^{E}(-2),A^{N3}(+1)]X with internal cross-link 2, which gave high substrate activity, bound to A3A(E72A) four times more strongly (K_d = 0.027 ± 0.006 μM) in comparison with the linear substrate oligo in buffer 2* (5'-AT_3**dC**AT_3, K_d = 0.13 ± 0.03 μM). The high rate of deamination of this oligo to $A3B_{CTD}$-QM-ΔL3-AL1swap is backed by the increased binding affinity as shown by our ITC experiments.

Next, we evaluated the binding data of dZ-containing oligos towards the catalytically competent $A3B_{CTD}$-QM-ΔL3-AL1swap. The binding constant of inhibitor dZ[U^{E}(-2),A^{N3}(+1)]X (K_d = 2.5 ± 0.5 μM) was around 4.5 times higher than that of the dZ-containing linear oligo (K_d = 11.4 ± 2.3 μM), similar to the difference seen in substrates. This trend was also consistent with the trend observed in inhibition constants (K_i) determined by our real-time NMR assay. This effect is smaller in ITC measurements, most likely due to the

differences in experimental conditions, as the ITC experiments were conducted in a different buffer and in the presence of trimethylamine *N*-oxide (TMAO) to stabilize the enzyme.

3.2.8. Evaluation of nuclease stability of cross-linked oligo

Nuclease-stability of modified oligos was evaluated using snake venom phosphodiesterase (phosphodiesterase I, Sigma) and compared with degradation of the unmodified sequence 5'-T_4**dC**AT. Under the conditions used in this experiment, 5'-T_4**dC**AT and the linear 4-mer dZ oligo were completely degraded within 60 min (**Figure 3.30**), whereas the cross-linked 4-mer inhibitor FdZ[C^{N3}(-2),H^{E}(+1)]X showed significantly enhanced stability towards enzymatic digestion. These data demonstrate an additional benefit of the introduction of a cross-link into DNA-based inhibitors of $A3B_{CTD}$, meaning that cross-linked oligos will have extended life-times in biological media in comparison with linear oligos. The nuclease stability experiments were performed by Dr. Yong Dong Su.

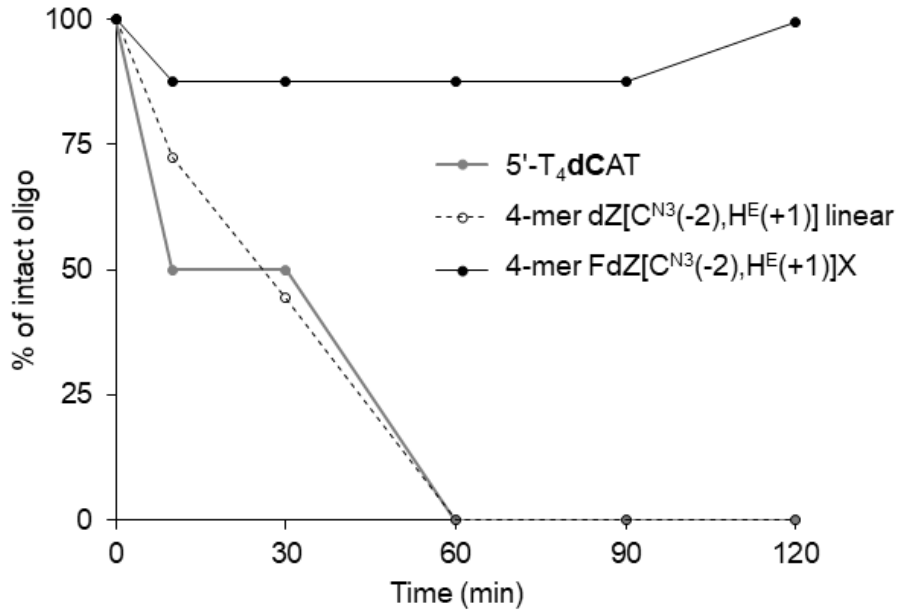


Figure 3.30. Evaluation of nuclease stability of cross-linked oligo.

Percentage of intact oligos determined after enzymatic digestion at 37 °C using snake venom phosphodiesterase (phosphodiesterase I, 0.016 units /L, from Sigma) in 120 min. Experiments were repeated at least twice with 6% error for each point (see experimental for details).

3.3. Discussion

We have previously developed the first selective A3 inhibitors[54, 124, 126] based on incorporation of cytosine-like 2'-deoxyzebularine[182-183] into short, linear ssDNA, achieving low micromolar inhibition constants (K_i). Recent structural studies[36, 62-63, 140] revealed that flexible ssDNA adopts a more rigid U-shaped loop upon binding to A3, projecting cytosine into the active-site pocket. One should note, however, that these studies were conducted on A3A, chimeric $A3B_{CTD}$ and modified $A3G_{CTD}$. Chimeric $A3B_{CTD}$ had four mutations distant from the active site, loop 3 was truncated and loop 1 of $A3B_{CTD}$ was swapped with loop 1 of

A3A, which constitute the main differences between amino acid sequences of A3A and $A3B_{CTD}$. This A3B construct has the substrate specificity of A3A, at least in the context of the CCC motif, which was used by us recently to develop A3B-specific ssDNA-based inhibitors.[126] There is no wild type $A3B_{CTD}$ crystal or NMR structure with ssDNA bound that would directly support a notion that DNA adopts a U-shape in complex with wild-type $A3B_{CTD}$ as well. However, A3A and, to the lesser extent, $A3B_{CTD}$ have been reported recently to deaminate cytosine in short 3-4 nucleotide loops of DNA hairpins.[96] This reassures that both A3A and $A3B_{CTD}$ accept U-shaped loops. Here, we demonstrates that by cross-linking dA at the +1 and T at the -2 positions, mimicking the close proximity observed in the crystal structure,[36] a better substrate and a better inhibitor can be obtained in comparison with linear ssDNA of the same length and nucleotide content. These results obtained in the NMR-based activity assay have been corroborated in ITC binding experiments. It is interesting that creating a covalent cross-link 2 between the same dA at the +1 position but with the T in the position -3 completely abolished substrate activity and did not improve inhibitory potential of dC- and dZ-containing oligos, respectively. The cytosine-containing terminally cyclized oligos were deaminated at a slower rate compared to the linear DNA substrate and for that reason were not converted to inhibitors. Different substrate activity and inhibitory potential of oligos based on cross-links 1-3 clearly indicated the importance of the chemical nature of the cross-link and how it is positioned in the DNA sequence.

The inhibitory potential was improved from 20-fold (**dZ**-linear) and 35-fold (**FdZ**-linear) to more than 1500-fold for FdZ[C^{N3}(-2),H^{E}(+1)]X (inhibitor potency, as assessed in the ratio of K_m/K_i). When the apparent inhibition constants (K_i) of oligos are compared with the K_m of the linear ssDNA substrate 5'-AT_3**dC**AT_3 (K_m = 150 μM, non-linear regression analysis). When the best inhibitor, FdZ[C^{N3}(-2),H^{E}(+1)]X, was evaluated against preferred substrate of $A3B_{CTD}$, dC-hairpin, a similar inhibition effect of ~ 1000-fold was observed.

Additionally, we expect that these cross-linked oligos by virtue of their unique shape will not interact strongly with other proteins and will not be able to form duplexes or other secondary structures with partially or fully complementary genomic sequences, potentially reducing toxicity and making these inhibitors of interest for clinical applications. Moreover, the cross-linked 4-mer oligo already showed enhanced stability towards enzymatic digestion by phosphodiesterase I, which indicates that this inhibitor should have extended lifetime in biological media and thus exhibit its inhibitory effect for a longer period of time.

The ability of such short 3-mer and 4-mer cross-linked dZ/FdZ-oligos to inhibit A3 can potentially lead to creation of small molecule-like A3 inhibitors targeting the active site of

the enzyme. The inhibitory potential of cross-linked oligos can be further improved by using more powerful inhibitors of C-to-U conversion than dZ and FdZ[184-185] providing that such molecules can be incorporated by chemical or enzymatic means into DNA.

The pre-shaping of oligos, using our 'click coupling' of alkyne and azide, may become a useful general synthetic strategy for oligonucleotide-based inhibitors, as U-shaped nucleic acids feature in complexes with several cellular proteins. So far, the U-shape has been observed in APOBEC-related tRNA adenosine deaminase TadA (PDB 2B3J)[186] as well as in unrelated proteins, such as telomere protein Pot1pc (PDB 4HIO)[184, 187-188] and bacterial cold-shock protein Bs-CspB (PDB 2ES2 and 3PF4).[189-190] More generally, the structural information can be harvested to design pre-shaped DNA/RNA inhibitors for any DNA/RNA-interacting protein, if a significant change of polynucleotide's shape is found upon complex formation.

3.4. Conclusions

Inhibition of APOBEC enzymes, especially A3A and A3B, is clinically relevant as these enzymes are involved in cancer mutagenesis and development of drug resistance. In this study we developed pre-shaped ssDNAs as potentially better substrates for A3 enzymes using a cross-linking strategy employing Cu(I)-catalyzed 'click chemistry' between alkyne and azide moieties embedded in a ssDNA fragment. This has led to the first nM inhibitors of $A3B_{CTD}$ by incorporating C-to-U inhibitor dZ or FdZ in place of the target dC in the cross-linked DNA fragments. This work provides a stepping-stone for further development of modified ssDNAs as potential A3 inhibitors and possibly as useful conjuvants to existing anti-viral and anti-cancer therapies.

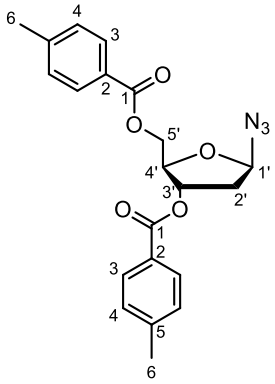
3.5. Experimental section

3.5.1. General information

All reactions were performed in oven-dried glassware under an atmosphere of dry argon or nitrogen unless otherwise noted. Moisture-sensitive reactions were carried out using standard syringe septum techniques and under an inert atmosphere of argon or nitrogen. All solvents and reagents were purified by standard techniques unless otherwise noted. Solvents for filtration, transfers, and chromatography were certified ACS grade. Evaporation of solvents was carried out under reduced pressure on a rotary vacuum evaporator below 40 ^{o}C. "Brine" refers to a saturated solution of sodium chloride in water. ^{1}H, ^{13}C, ^{31}P NMR spectra were recorded on Bruker 500- and 700-MHz spectrometers. Chemical shifts are reported in parts per million (ppm) downfield from tetramethylsilane. Spin multiplicities are described as s (singlet), bs (broad singlet), d (doublet), dd (double of doublets), dt (double of triplets), ddd (doublet of doublet of doublets), t (triplet), q (quartet), m (multiplet). Coupling constants are reported in Hertz (Hz). The assignments of proton and carbon signals were done using 2D homonuclear ^{1}H-^{1}H COSY, NOESY and heteronuclear ^{1}H-^{13}C HMQC or HSQC, and HMBC spectra. NMR spectra were processed in TopSpin. High-resolution electrospray mass spectra were recorded on a Thermo Fisher Scientific Q Exactive Focus Hybrid Quadrupole-Orbitrap mass spectrometer. Ions generated by ESI were detected in positive ion mode for small molecules and negative ion mode for oligonucleotides. Total ion count (TIC) was recorded in centroid mode over the *m*/*z* range of 100-3,000 and analyzed using Thermo Fisher Xcalibur Qual Browser. Analytical thin layer chromatography (TLC) was performed on MERCK precoated silica gel 60-F254 (0.5-mm) aluminum plates. Visualization of the spots on TLC plates was achieved either by exposure to UV light or by dipping the plates into aqueous $KMnO_4$ and heating with a heat gun. Silica gel column chromatography was performed using silica gel 60 (40–63 μm). Ratio of solvents for chromatography is provided in v/v. Oligonucleotide syntheses were carried out on a MerMade-4 DNA/RNA synthesizer (BioAutomation) on a 5 μmol scale using standard manufacturer's protocol for unmodified nucleotides.

3.5.2. Synthesis of modified azido CPG and phosphoramidite

3.5.2.1. Synthesis of 2-deoxy-3,5-di-*O*-(4-methylbenzoyl)-β-*D*-*erythro*-pentofuranosyl azide (2)[161]



To a stirring solution of sodium azide (97.5 g, 1.5 mol) and tetrabutylammonium hydrogen sulfate (101.7 g, 0.3 mmol) in saturated aqueous sodium bicarbonate solution (1 L), chloroform (1 L) was added, and the mixture was vigorously stirred for about 5 min until a milky emulsion was formed. Hoffer's chloro-sugar **1** (116.7 g, 0.3 mol) was rapidly added to the emulsion and the mixture was stirred 20 min. After disappearance of the starting material (R_f = 0.38, 30% EtOAc in hexane), the organic layer was washed with saturated sodium bicarbonate (1 L), water (2 × 1 L), dried over anhydrous sodium sulfate and filtered. The solvent was evaporated *in vacuo* and the product was recrystallized from ethanol (450 mL) to yield compound **2** (105 g, 88%, β:α ratio of 19:1, R_f = 0.30, 30% EtOAc in hexane).

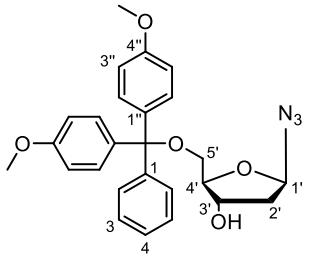
^{1}H NMR (500 MHz, DMSO-d_6): δ 7.94-7.84 (m, 4H, H-3); 7.36-7.29 (m, 4H, H-4); 5.87 (dd, 1H, J = 6.0 Hz, J = 4.2 Hz, H-1'); 5.56-5.50 (m, 1H, H-3'); 4.59-4.39 (m, 3H, H-4',5'); 2.49-2.42 (m, 1H, H-2'a); 2.375, 2.370 (2s, 6H, H-6); 2.37-2.30 (m, 1H, H-2'b).

^{13}C NMR (125.7 MHz, d_6-DMSO): δ 165.4; 165.3 (2C, C1); 144.0, 143.8 (2C, C5); 129.4, 129.3, 129.2, 129.2 (8C, C3,4); 126.7, 126.4 (2C, C2); 91.6 (C1'); 82.0 (C4'); 74.6 (C3'); 63.9 (C5'); 37.8 (C2'); 21.1, 21.1 (2C, C6).

IR ATR (cm^{-1}): 2943.74, 2126.16, 1715.46, 1611.78, 1509.61, 1443.95, 1407.97, 1378.51, 1271.93, 1238.71, 1176.04, 1114.15, 1101.23, 1083.75, 1059.76, 1035.31, 1019.42, 970.45, 939.79, 885.50, 844.42, 750.80, 691.50.

HRMS (ESI) *m*/*z*: [M+Na]$^+$ Calcd for $C_{21}H_{21}N_3O_5Na$ 418.1379; found 418.1372.

3.5.2.2. Synthesis of 2-deoxy-5-*O*-(4,4'-dimethoxytrityl)-β-D-*erythro*-pentofuranosyl azide (3)



Synthesis was performed by a similar procedure to that described earlier.[191] To a stirring solution of 2-deoxy-3,5-di-*O*-(4-methylbenzoyl)-β-D-*erythro*-pentofuranosyl azide **2** (3.96 g, 10 mmol) in 370 mL methanol was added 37 mL of 30% aq. ammonia solution and it stirred for 3 days. After the disappearance of the starting material, volatiles were removed by rotary vacuum evaporator and the residue was co-evaporated again with water to remove formed methyl toluate and then freeze dried from water to get the deprotected compound 2-deoxy-β-D-*erythro*-pentofuranosyl azide (1.65 g). This product was used without further purification to protect it with DMT. To a stirring solution of the deprotected azido sugar (1.65 g, 10 mmol) in dry pyridine (40 mL) at 0 °C 4,4'-dimethoxytrityl chloride (3.72 g, 11 mmol) was added and mixture was stirred at r.t. overnight. The reaction was quenched by the addition of water (10 mL) and volatiles were evaporated *in vacuo*. The residue was dissolved in 50 mL ethyl acetate and washed with brine (2 × 10 mL). The organic layer was dried over anhydrous sodium sulfate, filtered, and evaporated *in vacuo*. The crude product was purified by column chromatography over silica gel treated with 10% Et_3N in DCM and compound **3** was eluted with 5% EtOAc in DCM to afford the desired product **3** as a foam (3.13 g, 70%, R_f = 0.40, 10% MeOH in DCM).

^{1}H NMR (500 MHz, DMSO-d_6): δ 7.47-7.42 (m, 2H, H-2); 7.34-7.26 (m, 6H, H-3, 2''); 7.25-7.19 (m, 1H, H-4); 6.92-6.86 (m, 4H, H-3''); 5.68 (dd, 1H, J = 5.8 Hz, J = 3.1 Hz, H-1'); 5.25 (d, 1H, J = 5.0 Hz, OH); 4.20-4.12 (m, 1H, H-3'); 3.93-3.86 (m, 1H, H-4'); 3.73 (s, 6H, 2×CH_3); 3.15-3.02 (m, 2H, H-5'); 2.04-1.91 (m, 2H, H-2').

^{13}C NMR (125.7 MHz, DMSO-d_6): δ 158.0 (2C, C4''); 144.9 (C1); 135.7, 135.6 (2C, C1''); 129.7 (4C, C2''); 127.8 (2C, C3); 127.7 (2C, C2); 126.6 (C4); 113.2 (4C, C3''); 91.5 (C1'), 85.4 (C4'); 85.3 (*C*Ar$_3$); 69.9 (C3'); 63.9 (C5'); 55.0 (2C, *C*H$_3$); 40.4 (C2').

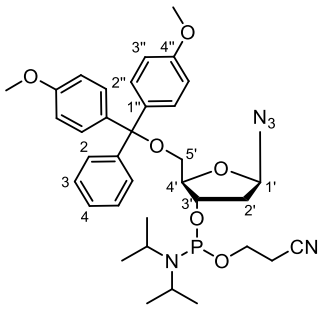
IR ATR (cm^{-1}): 2110.79, 1607.41, 1520.20, 1450.36, 1300.72, 1250.46, 1176.80, 1081.96, 1033.97, 829.42.

HRMS (ESI) *m/z*: [M+Na]+ Calcd for $C_{26}H_{27}N_3O_5Na$ 484.1848; found 484.1844.

3.5.2.3. Synthesis of modified azido-2-deoxyribose CPG (4)

Synthesis of the CPG support **4** was performed following the previously reported procedure.[163] LCAA-CPG (3.0 g) with a pore size of 500 Å and 120-200 mesh size obtained commercially (ChemGenes Corporation) was activated by 3% trifluoroacetic acid in DCM (30 mL) and kept overnight with gentle stirring. The slurry was filtered and washed with 9:1 triethylamine:di*iso*propylamine (50 mL), DCM and diethyl ether, then dried *in vacuo*. Activated LCAA-CPG was then treated with succinic anhydride (6.6 mmol, 0.66 g) and 4-dimethylaminopyridine (DMAP) (0.8 mmol, 0.1 g) in anhydrous pyridine (12 mL) and stirred gently at r.t. overnight. The slurry was filtered off and washed successively with pyridine, DCM, and ether, then dried *in vacuo*. This carboxylic derivatized CPG was then coupled with compound **3** (0.4 mmol, 0.184 g) in presence of 1-ethyl-3-(3-dimethylaminopropyl) carbodiimide (EDC base) (2.7 mmol, 0.48 mL), triethylamine (20 μL) and DMAP (0.8 mmol, 0.1 g) in 1:1 pyridine:DMF (36 mL). After 72 hours pentafluorophenol (1.5 mmol, 0.27 g) was added and mixture was kept overnight. The slurry was filtered and washed with DCM, incubated with a mixture of 9:1 pyridine:piperidine (10 mL) for 5 min followed by subsequent washings with DCM, acetonitrile, THF and again with DCM, then dried *in vacuo.* The CPG was then treated with the mixture composed of 2 mL of Cap A (acetic anhydride:2,6-lutidine:THF 1:1:8 v/v/v) and 2 mL of Cap B (1-methylimidazole:THF 4:21 v/v) for 2 hours, followed by washing with DCM, methanol, acetonitrile, THF and again with DCM and dried *in vacuo*. The load on the CPG **4** was determined to be 39 μmol/g based on UV absorption of DMT-cation at 504 nm (Ɛ = 76000 $L \cdot mol^{-1} \cdot cm^{-1}$) that was released from an aliquot of compound by treatment with 3% dichloroacetic acid in DCM.

3.5.3. Synthesis of 3-*O*-(*N*,*N*-di*iso*propylamino-2-cyanoethoxyphosphanyl)-2-deoxy-5-*O*-(4,4'-dimethoxytrityl)-β-D-*erythro*-pentofuranosyl azide (5)



To a stirring solution of azido sugar **3** (0.46 g, 0.99 mmol) in dry DCM (5 mL), under argon at rt were added Et_3N (0.19 mL) followed by 2-cyanoethyl *N*,*N*-di*iso*propyl chlorophosphoramidite (0.25 mL, 1 mmol). After the consumption of the starting material,

reaction mixture was washed with saturated $NaHCO_3$ solution (5 × 3 mL) followed by brine (5 mL). The organic layer was dried by passing through a column of anhydrous sodium sulfate. The concentration of azide phosphoramidite **5** in DCM solution (0.18 M) was calculated based on UV absorption of DMT-cation at 504 nm (Ɛ = 76,000 L/mol/cm) that was released from an aliquot of compound **5** by treatment with dichloroacetic acid in DCM.

R_f = 0.45, 10% acetone in DCM.

^{1}H NMR (500 MHz, DMSO-d_6) δ 7.52-7.47 (m, 2H, H-2); 7.41-7.37 (m, 3H, H-3 and H-4); 7.32-7.18 (m, 4H, H-2"), 6.87-6.82 (m, 4H, H-3"); 5.61 (dd, J = 5.77 Hz, J = 3.7 Hz, 1H, H-1'); 4.56-4.44 (m, 1H, H-3'); 4.22-4.15 (m, 2H, CH_2CH$_2$CN); 4.14-4.08 (m, 1H, H-4'); 3.79 (s), 3.78 (s) (6H, OCH$_3$); 3.60-3.50 (m, 2H, NCHCH$_3$); 3.32-3.21 (m, 2H, H-5'); 2.75-2.71 (m, 2H, 2.68-2.65 CH$_2$CH_2CN); 2.27-2.11 (m, 2H, H-2'a, 2'b); 1.18-1.13 (m, 9H), 1.04 (d, J =6.78 Hz) (12H, NCHCH_3).

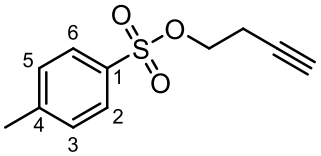
^{13}C NMR (125.7 MHz, DMSO-d_6) δ 158.60, 158.59, 158.46, 158.44 (2C, C4"); 144.79, 144.48 (C1); 135.69, 135.62, 135.59, 135.56 (2C, C1"); 130.14, 130.10, 130.07, 130.06 (4C, C2"); 128.25, 128.20, 128.14, 128.13, 127.93, 127.81 (4C, C2, 3); 126.96, 126.93, 127.76, 126.72 (C4); 116.92 (CN); 113.24, 113.11 (4C, C3"); 86.60, 86.15 (CAr3); 84.58, 84.53, 84.47, 84.42 (1C, C1'); 91.70, 91.69 (1C, C4'); 76.97, 76.93, 76.71, 76.66 (1C, C3'); 63.85, 63.56 (C5'); 60.22, 60.19, 60.18, 60.15 (1C, NCH$_2$CH$_2$CN); 55.25, 55.24, 55.23, 55.19 (2C, OCH$_3$); 43.27, 43.22, 43.17, 43.12 (2C, NCHCH$_3$); 39.70, 39.68, 39.55, 39.52 (C2'); 24.60, 24.57, 24.54, 24.51, 24.45, 24.42 (NCHCH$_3$); 20.12, 20.06, 19.92, 19.87, 19.83, 19.77 (CH$_2$CH$_2$CN).

^{31}P NMR (202.5 MHz, DMSO-d_6, ref. 85% H_3PO_4) δ 148.97, 148.59 in ~1:1 ratio.

IR ATR (cm^{-1}): $\bar{\upsilon}$ = 2972.88, 2113.95, 1608.30, 1509.48, 1383.09, 1265.46, 1179.73, 1035.37, 979.48, 832.56, 703.68.

3.5.4. Synthesis of modified 2'-deoxyadenosines

3.5.4.1. Synthesis of but-3-yn-1-yl 4-methylbenzenesulfonate (6)



But-3-yn-1-ol (2.0 g, 20.37 mmol) in DCM (50 mL) was cooled to 0 ^{o}C and then DMAP (34 mg, 0.27 mmol), tosyl chloride (6.52 g, 34.19 mmol) and Et_3N (4.8 mL, 34.19 mmol) were added. The reaction mixture was stirred for 30 min and then diluted with 50 mL more of DCM and washed with brine (50 × 2 mL). The organic layer was dried over anhydrous

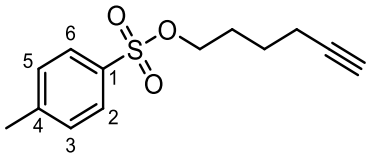
sodium sulfate, filtered, and concentrated *in vacuo*. The crude product was purified by column chromatography over silica gel (60-120 mesh) and eluted with hexane/EtOAc (7:3) to afford the desired product **6** (5.90 g, 92%, R_f = 0.45, 30% EtOAc in hexane).

^{1}H NMR (500 MHz, $CDCl_3$) δ 7.83 (d, J = 8.37 Hz, 2H, H2, H6); 7.38 (d, J = 8.44 Hz, 2H, H3, H5); 4.13 (t, J = 7.04 Hz, 2H, OC*H*$_2$CH$_2$CCH); 2.58 (ddd, J = 7.07, 7.03, 2.74 Hz, 2H, OCH$_2$C*H*$_2$CCH); 2.47 (s, 3H, ArCH$_3$); 1.99 (t, J = 2.67Hz, 1H, OCH$_2$CH$_2$CC*H*).

^{13}C NMR (125.7 MHz, $CDCl_3$) δ 145.03 (C1); 132.85 (C4); 129.92 (2C, C3, C5); 127.99 (2C, C2, C6); 78.39 (OCH$_2$CH$_2$<u>C</u>CH); 70.78 (OCH$_2$CH$_2$C<u>C</u>H); 67.44 (O<u>C</u>H$_2$CH$_2$CCH); 21.66 (Ar<u>C</u>H$_3$); 19.46 (OCH$_2$<u>C</u>H$_2$CCH).

HRMS (ESI) *m/z*: $[M+Na]^+$ Calcd for $C_{11}H_{12}NaO_3S$: 247.0399; Found: 247.0396.

3.5.4.2. Synthesis of hex-5-yn-1-yl 4-methylbenzenesulfonate (7)



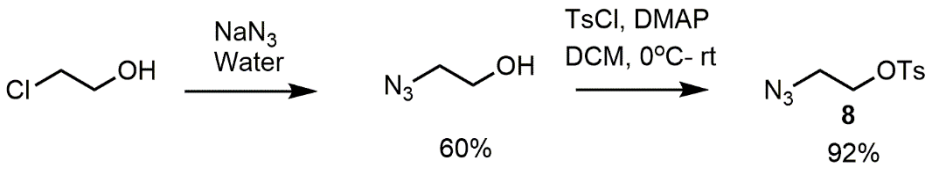
Compound **7** was synthesized using the same protocol as for compound **6** with 98% yield. R_f = 0.43, 20% EtOAc in hexane.

^{1}H NMR (500 MHz, $CDCl_3$) δ 7.82 (d, J =8.36 Hz, 2H, H2, H6); 7.37 (d, J = 8.34 Hz, 2H, H3, H5); 4.08 (t, J = 7.04 Hz, 2H, OC*H*$_2$CH$_2$CH$_2$CH$_2$CCH); 2.48 (s, 3H, ArCH$_3$); 2.19 (ddd, J = 6.97 Hz, 6.90 Hz, 2.59 Hz, 2H, CH$_2$CH$_2$CH$_2$C*H*$_2$CCH); 1.95 (t, J = 2.63 Hz, 1H, OCH$_2$CH$_2$CH$_2$CH$_2$CC*H*); 1.83-1.78 (m, 2H, OCH$_2$C*H*$_2$CH$_2$CH$_2$CCH); 1.61-1.55 (m, 2H, OCH$_2$CH$_2$C*H*$_2$CH$_2$CCH).

^{13}C NMR (125.7 MHz, $CDCl_3$) δ 144.77 (C1); 133.11 (C4); 129.86 (2C, C3 and C5); 127.89 (2C, C2 and C6); 83.40 (OCH$_2$CH$_2$CH$_2$CH$_2$<u>C</u>CH); 69.92 (O<u>C</u>H$_2$CH$_2$CH$_2$CH$_2$CCH); 68.97 (OCH$_2$CH$_2$CH$_2$CH$_2$C<u>C</u>H); 27.77 (OCH$_2$<u>C</u>H$_2$CH$_2$CH$_2$CCH); 24.23 (OCH$_2$CH$_2$<u>C</u>H$_2$CH$_2$CCH); 21.65 (Ph<u>C</u>H$_3$); 17.73 (CH$_2$CH$_2$CH$_2$<u>C</u>H$_2$CCH).

HRMS (ESI) *m/z*: $[M+Na]^+$ Calcd for $C_{13}H_{16}O_3S$ 275.0712; Found 275.0707.

3.5.4.3. Synthesis of 2-azidoethyl-4-methylbenzenesulfonate (8)



Scheme 3.5. Synthesis of 2-azidoethyl 4-tosylate (**8**)

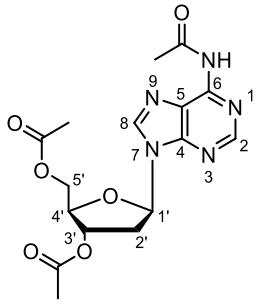
2-Azidoethan-1-ol was prepared according to a reported procedure[192] and converted to compound **8** using the same protocol as for compound **6**. Yield 92%.

^{1}H NMR (500 MHz, DMSO-d_6): δ 7.79 (d, J = 8.3 Hz, 2H); 7.34 (d, J = 8.1 Hz, 2H); 4.16 (t, 2H, J = 5.2 Hz, CH_2CH$_2$N$_3$); 3.48 (t, 2H, J = 5.2 Hz, CH$_2$$CH_2N_3$); 2.45 (s, 3H, Ar-C$H_3$).

^{13}C NMR (125.7 MHz, CDCl$_3$): δ 145.33 (1C, C1); 132.52 (1C, C4); 130.03 (2C, C3 and C5); 127.95 (2C, C2 and C6); 68.22 (1C, CH$_2$CH$_2$N$_3$); 49.58 (1C, CH$_2$$CH_2N_3$); 21.65 (1C, CH$_3$).

HRMS (ESI) m/z: [M+Na]$^+$ Calcd for C$_9$H$_{11}$N$_3$O$_3$SNa 264.0419; found 264.0413.

3.5.4.4. N^6,3',5'-Triacetyl-2'-deoxyadenosine (10)



N^6,3',5'-Triacetyl-2'-deoxyadenosine **10** was prepared similarly to a reported procedure with some modifications.[164] A mixture of 2'-deoxyadenosine **9** (1.0 g, 3.98 mmol), pyridine (7.5 mL) and Ac$_2$O (3.5 mL) was stirred at r.t. overnight. The resulting solution was heated to 60 ^{o}C for 6 h. After the disappearance of the starting material, monitored by TLC (R_f = 0.40, MeOH/DCM, 95:5), the reaction was cooled down and quenched with excess EtOH (20 mL). Volatiles were evaporated *in vacuo*. Traces of pyridine were co-evaporated with successive portions of EtOH and MeOH (20 mL each). The resultant oily liquid was diluted with EtOAc (100 mL) and washed with brine (2 × 20 mL). The organic layer was dried over anhydrous sodium sulfate, filtered, and evaporated *in vacuo.* The crude product was purified by column chromatography on silica gel, eluting with DCM/MeOH (9.5:0.5) to afford the desired product **10** as a white solid (1.15 g, 76%, R_f = 0.40, 5% MeOH in DCM).

^{1}H NMR (500 MHz, DMSO-d_6): δ 10.70 (s, 1H, NH); 8.663, 8.660 (2s, 2H, H-2, H-8); 6.48 (dd, 1H, $J_{1',2'}$ = 6.3, 7.8 Hz, H-1'); 5.44 (dt, 1H, J = 2.8, 6.5 Hz, H-3'); 4.32 (dd, 1H, $J_{4',5'a}$ = 4.2 Hz, $J_{5'a,5'b}$ = 11.0 Hz, H-5'a); 4.28 (ddd, 1H, $J_{3',4'}$ = 2.7 Hz, $J_{4',5'a}$ = 4.2 Hz, $J_{4',5'b}$ = 5.5 Hz, H-4'); 4.22 (dd, 1H, $J_{4',5'b=}$ 5.5 Hz, $J_{5'a,5'b}$ = 11.0 Hz, H-5'b); 3.20 (ddd, 1H, $J_{2'a,3'}$ = 2.8 Hz, $J_{2'a,1'=}$ 6.3 Hz, $J_{2'a,2'b}$ = 14.2 Hz, H-2'a); 2.60 (ddd, 1H, $J_{2'b,3'}$ = 6.6 Hz, $J_{2'b,1'}$ = 7.8 Hz, $J_{2'a,2'b}$ = 14.2 Hz, H-2'b); 2.26 (s, 3H, NHCOCH_3); 2.10, 2.00 (2s, 6H, OCOCH_3).

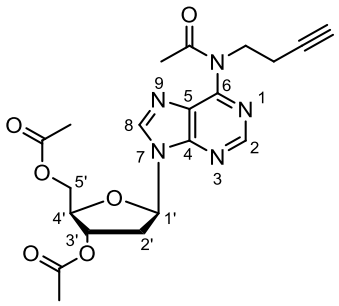
^{13}C NMR (125.7 MHz, DMSO-d_6): δ 170.1 (O<u>CO</u>CH$_3$); 170.0 (O<u>CO</u>CH$_3$); 168.8 (NH<u>CO</u>CH$_3$); 151.7 (C2); 151.5 (C4); 149.7 (C6); 142.9 (C8); 123.8 (C5); 83.7 (C1'); 81.8 (C4'); 74.2 (C3'); 63.5 (C5'); 35.3 (C2'); 24.3 (NHCO<u>CH$_3$</u>); 20.8 (OCO<u>CH$_3$</u>); 20.5 (OCO<u>CH$_3$</u>).

HRMS (ESI) *m/z*: [M+H]$^+$ Calcd for $C_{16}H_{20}N_5O_6$ 378.1414; found 378.1397.

3.5.4.5. General protocol for alkylation of N^6,3',5'-triacetyl-2'-deoxyadenosine (10)

To a stirring solution of N^6,3',5'-triacetyl-2'-deoxyadenosine (**10**, 1.0 g, 2.65 mmol) in acetonitrile (15 mL), Cs_2CO_3 (2.50 g, 7.67 mmol) and one of tosylated alcohols **6-8** (15.80 mmol) were added and mixture was heated to 60 °C. After the disappearance of the starting material on TLC the reaction mixture was diluted with EtOAc (50 mL) and washed with water (2 × 10 mL). The organic layer was dried over anhydrous sodium sulfate, filtered, and concentrated *in vacuo*. The crude product was purified by silica gel column chromatography using DCM/MeOH (9:1) to get desired compound as a sticky solid.

3.5.4.5.1. N^6-(But-3-yn-1-yl)-N^6,3',5'-triacetyl-2'-deoxyadenosine (11a)



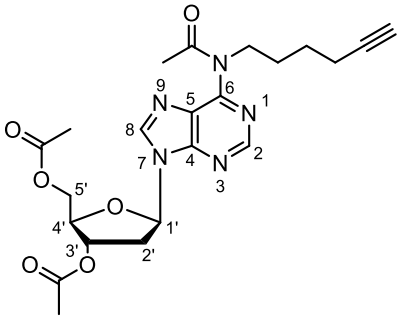
R_f = 0.37 in MeOH/DCM (1:9); yield: 0.64 g, 80%.

^{1}H NMR (500 MHz, DMSO-d_6) δ 8.83 (s, 1H, H-8); 8.82 (s, 1H, H-2); 6.53 (dd, 1H, $J_{1',2'a}$ = 6.8, $J_{1',2'b}$ = 7.0 Hz, H-1'); 5.46 (dd, $J_{3',2'b}$ = 3.28 Hz, $J_{3',4'}$ = 6.07Hz, 1H, H-3'); 4.35-4.31 (m, 2H, H-5'); 4.27-4.24 (m, 1H, H-4'); 4.21 (t, J = 7.45 Hz, 2H, NCH_2CH$_2$CCH); 3.20 (ddd, 1H, $J_{2'a,3'}$ = 2.8 Hz, $J_{2'a,1'=}$ 6.8 Hz, $J_{2'a,2'b}$ = 13.3 Hz, H-2'a); 2.63 (ddd, 1H, $J_{2'b,3'}$ = 3.3 Hz, $J_{2'b,1'}$ = 6.3 Hz, $J_{2'a,2'b}$ = 13.3Hz, H-2'b); 2.50 (app t, NCH$_2$*CH$_2$*CCH); 2.19 (s, 3H, NCOC*H$_3$*); 2.12 (s, 3H, OCOC*H$_3$*); 2.01 (s, 3H, OCOC*H$_3$*).

^{13}C NMR (125.7 MHz, DMSO-d_6) δ 171.12 (O<u>CO</u>CH$_3$); 170.60 (O<u>CO</u>CH$_3$); 170.53 (NH<u>CO</u>CH$_3$); 152.97 (C6); 152.88 (C4); 152.04 (C8); 144.83 (C2); 127.16 (C5); 84.38 (C1'); 82.35 (C4'); 81.71 (NCH$_2$CH$_2$<u>C</u>CH); 74.61 (C3'); 72.99 (NCH$_2$CH$_2$C<u>CH</u>); 63.97 (C5'); 45.94 (N<u>CH$_2$</u>CH$_2$CCH); 35.91 (C2'); 24.41 (NCO<u>CH$_3$</u>); 21.26 (OCO<u>CH3</u>); 20.99 (OCO<u>CH$_3$</u>); 18.47 (NCH$_2$<u>CH$_2$</u>CCH).

HRMS (ESI) *m/z*: [M +H]$^+$ Calcd for $C_{20}H_{24}N_5O_6$ 430.1721; Found 430.1713.

3.5.4.5.2. N^6-(Hex-5-yn-1-yl)-N^6,3',5'-triacetyl-2'-deoxyadenosine (11b)



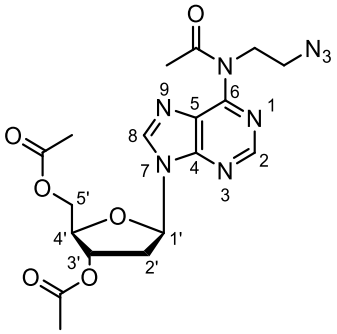
R_f = 0.45 in MeOH/DCM (1:9); yield: 1.03 g, 85%.

^{1}H NMR (500 MHz, DMSO-d_6) δ 8.86 (s, 1H, H-8); 8.83 (s, 1H, H-2); 6.53 (dd, 1H, $J_{1',2'a}$ = 6.9, $J_{1',2'b}$ = 7.1 Hz, H-1'); 5.46 (dd, $J_{3',2'b}$ = 6.07 Hz, $J_{3',4'}$ = 3.28 Hz, 1H, H-3'); 4.35-4.30 (m, 1H, H-5'a); 4.28-4.24 (m, 1H, H-4'); 4.23 (dd, $J_{5'b,5'a}$ = 11.23 Hz, $J_{5'b,4'}$ = 5.50 Hz, 1H, H-5'b); 4.11 (t, J = 7.45 Hz, 2H, NCH_2CH$_2$CH$_2$CH$_2$CCH); 3.21 (dd, $J_{2'a,2'b}$ = 2.8 Hz, $J_{2'a,1'}$ = 6.9 Hz, 1H, H-2'a); 2.64 (dd, 1H, $J_{2'b,2'a}$ = 6.34 Hz, $J_{2'b,3'}$ = 2.79 Hz, 1H, H-2'b); 2.12 (s, 3H, OCOCH_3); 2.10-2.09 (m, 2H, NCH$_2$CH$_2$CH$_2$CH_2CCH); 2.08 (s, 3H, OCOCH_3); 1.92 (s, 3H, NCOCH_3) 1.54 (q, J = 7.31 Hz, 2H, NCH$_2$CH_2CH$_2$CH$_2$CCH); 1.38 (q, J = 7.84 Hz, 2H, NCH$_2$CH$_2$CH_2CH$_2$CCH).

^{13}C NMR (125.7 MHz, DMSO-d_6) δ 170.82 (O<u>CO</u>CH$_3$); 170.56 (O<u>CO</u>CH$_3$); 170.51 (NH<u>CO</u>CH$_3$); 153.15 (C6); 153.01 (C4); 152.21 (C8); 144.95 (C2); 127.70 (C5); 84.54 (C1'); 84.42 (C4'); 82.37 (NCH$_2$CH$_2$CH$_2$CH$_2$<u>C</u>CH); 74.62 (C3'); 71.41 (C2'); 63.94 (C5'); 46.19 (N<u>CH$_2$</u>CH$_2$CH$_2$CH$_2$CCH); 35.92 (C2'); 27.68 (NCH$_2$<u>CH$_2$</u>CH$_2$CH$_2$CCH); 25.56 (NHCH$_2$CH$_2$<u>CH$_2$</u>CH$_2$CCH); 24.22 (NCO<u>CH$_3$</u>); 21.20 (OCO<u>CH$_3$</u>); 20.92 (OCO<u>CH$_3$</u>); 17.77 (NCH$_2$CH$_2$CH$_2$<u>CH$_2$</u>CCH).

HRMS (ESI) *m/z*: [M +Na]$^+$ Calcd for $C_{22}H_{27}N_5O_6Na$ 480.1859; Found 480.1853

3.5.4.5.3. N^6-(2-Azidoethyl)-N^6,3',5'-triacetyl-2'-deoxyadenosine (11c)



R_f = 0.38 in MeOH/DCM (1:9); yield: 0.94 g, 80%.

^{1}H NMR (500 MHz, DMSO-d_6): δ 8.84 (s, 1H, H-8); 8.82 (s, 1H, H-2); 6.52 (dd, 1H, $J_{1',2a'}$ = 6.7, $J_{1',2'b}$ 7.2 Hz, H-1'); 5.45 (dt, 1H, J = 2.3, 6.4 Hz, H-3'); 4.32 (dd, 1H, $J_{4',5'a}$ = 4.2 Hz, $J_{5'a,5'b}$ = 10.6 Hz, H-5'a); 4.31-4.28 (m, 3H, H5'b and N$CH_2$$CH_2$N3); 4.23 (ddd, 1H, $J_{3',4'}$ = 2.3 Hz, $J_{4',5'a}$ = 4.4 Hz, $J_{4',5'b}$ = 5.9 Hz, H-4'); 3.52 (t, J = 7.0 Hz, 2H, NCH$_2$CH_2N$_3$); 3.20 (ddd, 1H, $J_{2'a,3'}$ = 2.3 Hz, $J_{2'a,1'=}$ 6.7 Hz, $J_{2'a,2'b}$ = 14.3 Hz, H-2'a); 2.63 (ddd, 1H, $J_{2'b,3'}$ = 6.4 Hz, $J_{2'b,1'}$ = 7.2 Hz, $J_{2'a,2'b}$ = 14.3 Hz, H-2'b); 2.17 (s, 3H, NCOCH_3); 2.10 (s, 3H, OCOCH_3); 1.99 (s, 3H, OCOCH_3).

^{13}C NMR (125.7 MHz, DMSO-d_6): δ 170.8 (O<u>CO</u>CH$_3$); 170.1 (O<u>CO</u>CH$_3$); 170.0 (N<u>CO</u>CH$_3$); 152.6 (C6); 152.3 (C4); 151.6 (C8); 144.5 (C2); 126.7 (C5); 83.9 (C1'); 81.8 (C4'); 74.1 (C3'); 63.4 (C5'); 49.4 (NCH$_2$<u>CH</u>$_2$N$_3$); 45.6 (N<u>C</u>H$_2$CH$_2$N$_3$); 35.4 (C2'); 23.8 (NCO<u>CH</u>$_3$); 20.8 (OCO<u>CH</u>3); 20.5 (OCO<u>CH</u>$_3$).

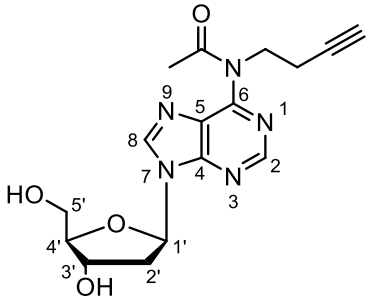
IR ATR (cm^{-1}): 3440.67, 2306.61, 2339.80, 2251.73, 2124.62, 1667.50, 1058.50, 1037.65, 1008.07, 823.02, 760.59, 667.93, 624.85.

HRMS (ESI) *m*/*z*: [M +H]$^+$ Calcd for $C_{18}H_{23}N_8O_6$ 447.1741; found 447.1737.

3.5.4.6. General procedure for removal of 3',5'-*O*-acetyl protecting groups

To a stirring solution of one of the compounds **11a-c** (2.0 mmol) in MeOH:water (10:10 mL) Et$_3$N (1.5 mL) was added at r.t. The reaction was monitored by TLC. After 15 min volatiles were evaporated *in vacuo*. The crude product was either triturated with ether providing the desired white solid for **12a-b** (85% and 82% yields) or purified by silica gel column chromatography using DCM/MeOH (7:3) to get desired compound **12c** as a semisolid (89% yield).

3.5.4.6.1. N^6-(But-3-yn-1-yl)-N^6-acetyl-2'-deoxyadenosine (12a)



R_f = 0.40 in MeOH/DCM (3:7).

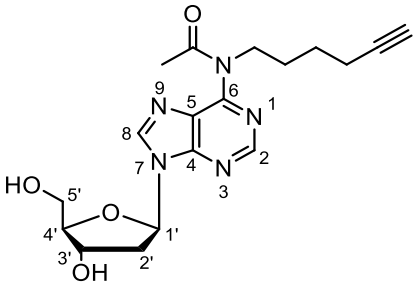
^{1}H NMR (500 MHz, DMSO-d_6) δ 8.83 (s, 1H, H-8); 8.81 (s, 1H, H-2); 6.48 (app t, J = 6.72 Hz, 1H, H-1'); 4.46-4.42 (m, 1H, H-3'); 4.18 (t, J = 7.40 Hz, 2H, NCH_2CH$_2$CCH); 3.91-3.89 (m, 1H, H-4'); 3.63 (dd, $J_{5'a,5'b}$ = 11.74 Hz, $J_{5'a,4'}$ = 4.69 Hz, 1H, H-5'a) 3.54 (dd, 1H, $J_{5'b,5'a}$ = 11.74 Hz, $J_{5'b,4'}$ = 4.76 Hz, 1H, H-5'b); 2.74 (ddd, $J_{2'a,2'b}$ = 11.23 Hz, $J_{2'a,3'}$ = 7.00 Hz,

$J_{2'a,1'}$ = 5.89 Hz, 1H, H-2'a); 2.69 (t, J = 2.63 Hz, 1H, NHCH$_2$CH$_2$CC*H*); 2.49-2.45 (m, 2H, NHCH$_2$C*H*$_2$CCH); 2.39-2.34 (m, 1H, H-2'b); 2.18 (s, 3H, NCOC*H*$_3$).

^{13}C NMR (125.7 MHz, DMSO-d_6) δ 171.11 (NCOCH$_3$); 152.93 (C2); 152.72 (C6); 151.92 (C4); 144.65 (C8); 127.03 (C5); 88.54 (C4'); 84.33 (C1'); 81.74 (NHCH$_2$CH$_2$CCH); 73.02 (NHCH$_2$CH$_2$CCH); 71.05 (C3'); 61.97 (C5'); 45.93 (NHCH$_2$CH$_2$CCH) 39.85 (C2'); 24.42 (NCOCH$_3$); 18.47 (NHCH$_2$CH$_2$CCH).

HRMS (ESI) *m/z*: [M +H]$^+$ Calcd for $C_{16}H_{20}N_5O_4$ 346.1510; Found 346.1505.

3.5.4.6.2. N^6-(Hex-5-yn-1-yl)-N^6-acetyl-2'-deoxyadenosine (12b)



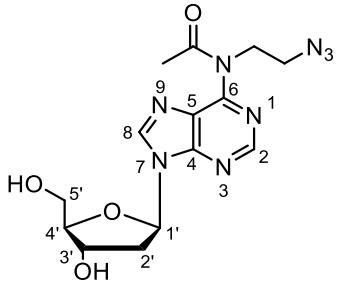
R_f = 0.48 in MeOH/DCM (3:7).

^{1}H NMR (500 MHz, DMSO-d_6) δ 8.84 (s, 1H, H-8); 8.83 (s, 1H, H-2); 6.49 (app t, J = 6.74 Hz, 1H, H-1'); 4.47-4.46 (m, 1H, H-3'); 4.10 (t, J = 7.22 Hz, 2H, NC*H*$_2$CH$_2$CH$_2$CH$_2$CCH); 3.92 (q, J = 4.38, 1H, H-4'); 3.65 (dd, $J_{5'a,5'b}$ = 11.82 Hz, $J_{5'a,4'}$ = 4.60 Hz, 1H, H-5'a) 3.56 (dd, 1H, $J_{5'b,5'a}$ =11.82 Hz, $J_{5'b,4'}$ = 4.60 Hz, 1H, H-5'b); 2.79 (ddd, $J_{2'a,2'b}$ = 13.83 Hz, $J_{2'a,3'}$ = 6.71 Hz, $J_{2'a,1'}$ = 6.49 Hz, 1H, H-2'a); 2.65 (t, J = 2.70 Hz, NCH$_2$CH$_2$CH$_2$CH$_2$CC*H*); 2.51-2.38 (m, 1H, H-2'b); 2.10 (s, 1H, NHCOC*H*$_3$); 2.09-2.07 (m, 2H, NCH$_2$CH$_2$CH$_2$C*H*$_2$CCH); 1.56-1.50 (m, 2H, NCH$_2$C*H*$_2$CH$_2$CH$_2$CCH); 1.41-1.35 (m, 2H, NCH$_2$CH$_2$C*H*$_2$CH$_2$CCH).

^{13}C NMR (125.7 MHz, DMSO-d_6) δ 170.82 (NCOCH$_3$); 152.98 (C6); 152.11 (C4); 144.83 (C8); 127.59 (C5); 88.54 (C4'); 84.63; 84.33 (C1'); 71.61 (NHCH$_2$CH$_2$CH$_2$CH$_2$CCH); 71.04 (C3'); 61.95 (C5'); 46.19 (NHCH$_2$CH$_2$CH$_2$CH$_2$CCH); 39.81 (C2'); 27.69 (NCH$_2$CH$_2$CH$_2$CH$_2$CCH); 25.58 (NHCH$_2$CH$_2$CH$_2$CH$_2$CCH); 24.25 (NCOCH$_3$); 17.77 (NCH$_2$CH$_2$CH$_2$CH$_2$CCH).

HRMS (ESI) *m/z*: [M +H]$^+$ Calcd for $C_{18}H_{24}N_5O_4$ 374.1823; Found 374.1817.

3.5.4.6.3. N^6-(2-Azidoethyl)-N^6-acetyl-2'-deoxyadenosine (12c)



R_f = 0.38 in MeOH/DCM (3:7).

^{1}H NMR (500 MHz, DMSO-d_6): δ 8.83 (s, 1H, H-8); 8.82 (s, 1H, H-2); 6.49 (app t, J = 6.7 Hz, 1H, H-1'); 5.43 (s, 1H, 3'-OH), 5.05 (s, 1H, 5'-OH), 4.45 (dt, 1H, J = 3.2, 5.6 Hz, H-3'); 4.29 (t, J = 6.1 Hz, 2H, NCH_2CH$_2$N$_3$); 3.90 (ddd, 1H, $J_{3',4'}$ = 2.3 Hz, $J_{4',5'b}$ = 4.6 Hz, $J_{4',5'a}$ = 5.9 Hz, H-4'); 3.63 (dd, 1H, $J_{4',5'a}$ = 5.9 Hz, $J_{5'a,5'b}$ = 11.7 Hz, H-5'a); 3.54 (dd, 1H, $J_{4',5'b}$= 4.6 Hz, $J_{5'a,5'b}$ = 11.7 Hz, H-5'b); 3.52 (t, 2H, J = 6.1 Hz, 2H, NCH$_2$CH_2N$_3$); 2.77 (ddd, 1H, $J_{2'a,3'}$ = 2.8 Hz, $J_{2'a,1'}$= 6.3 Hz, $J_{2'a,2'b}$ = 13.2 Hz, H-2'a); 2.37 (ddd, 1H, $J_{2'b,3'}$ = 6.2 Hz, $J_{2'b,1'}$ = 9.8 Hz, $J_{2'a,2'b}$ = 13.2 Hz, H-2'b); 2.16 (s, 3H, NCOCH_3).

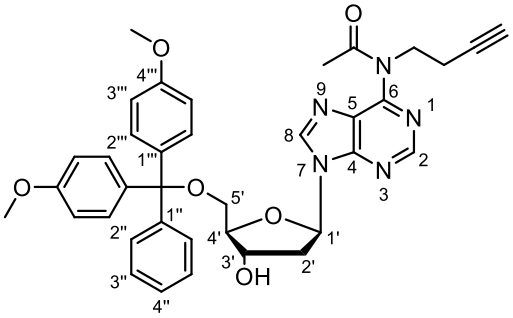
^{13}C NMR (125.7 MHz, DMSO-d_6): δ 170.8 (N<u>CO</u>CH$_3$); 152.5 (C2); 152.1 (C6); 151.5 (C4); 144.3 (C8); 126.6 (C5); 88.0 (C4'); 83.8 (C1'); 70.5 (C3'); 61.4 (C5'); 49.3 (NCH$_2$<u>CH$_2$</u>N$_3$): 45.6(N<u>C</u>H$_2$CH$_2$N$_3$), 39.8 (C2'); 23.8 (NCO<u>CH$_3$</u>).

HRMS (ESI) m/z: [M +H]$^+$ Calcd for $C_{14}H_{18}N_8O_4Na$ 385.1349; found 385.1342.

3.5.4.7. General protocol for 5'-*O*-4,4'-dimethoxytritylation of nucleosides

To a stirring solution of a nucleoside (1.8 mmol) in dry pyridine (5 mL), 4,4'-dimethoxytrityl chloride (0.67 g, 2.0 mmol) was added at 0 ^{o}C and the mixture was stirred at r.t. overnight under argon. After consumption of the starting nucleoside (TLC analysis), H_2O (1 mL) was added to quench the reaction. Solvents were evaporated *in vacuo* and the residue was dissolved in 50 mL DCM, washed with brine (2 × 10 mL), organic layer was dried over anhydrous sodium sulfate, filtered, and concentrated *in vacuo*. The crude product was purified by column chromatography over silica gel treated with 10% Et_3N in CH_2Cl_2 and eluted with 5% MeOH in DCM to afford the desired product as a foam.

3.5.4.7.1. 5'-*O*-(4,4'-Dimethoxytrityl)-N^6-(but-3-yn-1-yl)-N^6-acetyl-2'-deoxyadenosine (13a)



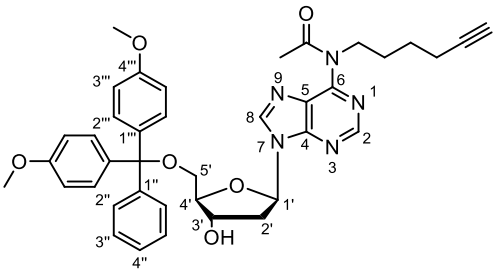
R_f = 0.45 (5% MeOH in DCM); yield: 0.68 g, 90%.

^{1}H NMR (500 MHz, DMSO-d_6) δ 8.73 (s, 1H, H-8); 8.71 (s, 1H, H-2); 7.31-7.30 (m, 2H,H-2"); 7.21-7.17 (m,7H, H-3", 4" and 2'"); 6.81- 6.76 (m, 4H, H-3'"); 6.51 (t, J = 6.33 Hz, 1H, H-1'); 5.42 (d, 1H, $J_{3',OH}$ = 4.63 Hz, H-O*H*) 4.51-4.49 (m, 1H, H-3'); 4.17 (t, J = 7.70 Hz, 2H, NC*H*$_2$CH$_2$CCH); 4.04-4.01 (m, 1H, H-4'); 3.70 (s, 3H,OC*H*$_3$); 3.71 (s, 3H, OC*H*$_3$); 3.23-3.16 (m, 2H, H-5'); 2.94-2.88 (m, 1H, 1H, H-2'a); 2.69 (t, J = 2.57 Hz, 1H, NCH$_2$CH$_2$CC*H*); 2.47-2.45 (m, 2H, NCH$_2$C*H*$_2$CCH); 2.41-2.39 (m, 1H, H-2'b); 2.12 (s, 3H, NCOC*H*$_3$).

^{13}C NMR (125.7 MHz, DMSO-d_6) δ 171.10 (N<u>CO</u>CH$_3$); 158.48, 158.46 (2C, C4'"); 152.89 (C6) 152.74 (C4); 151.85 (C2); 145.32 (C1"); 144.99 (C8); 136.06, 135.95 (2C, C1'"); 130.14 (4C, C2'"); 128.14 (2C, C3"); 128.11 (2C, C2"); 127.14 (C5); 127.05 (C4"); 113.53, 113.51 (4C, C3'"); 86.55 (C4'); 85.89 (<u>C</u>-Ar$_3$); 84.36 (C1'); 81.70 (NCH$_2$CH$_2$*C*CH); 72.99 (NCH$_2$CH$_2$<u>C</u>H); 70.99 (C3'); 64.56 (C5'); 55.47 (O<u>CH3</u>); 55.45 (O<u>CH3</u>); 45.98 (NH<u>C</u>H$_2$CH$_2$CCH); 38.96 (C2'); 24.36 (NCO<u>CH$_3$</u>); 18.47 (NHCH$_2$<u>CH$_2$</u>CCH).

HRMS (ESI) *m/z*: $[M+H]^+$ Calcd for $C_{37}H_{38}N_5O_6$ 648.2817; Found 648.2805.

3.5.4.7.2. 5'-*O*-(4,4'-Dimethoxytrityl)-N^6-(hex-5-yn-1-yl)-N^6-acetyl-2'-deoxyadenosine (13b)



R_f = 0.48 (5% MeOH in DCM); yield: 0.93 g, 86%.

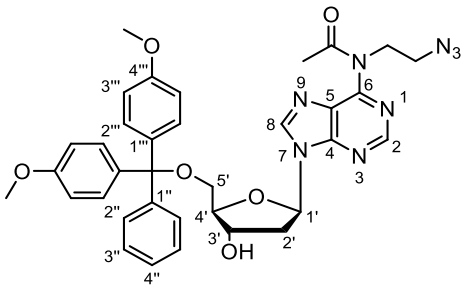
^{1}H NMR (500 MHz, DMSO-d_6) δ 8.77 (s, 1H, H-8); 8.73 (s, 1H, H-2); 7.32-7.29 (m, 2H, H-2"); 7.21-7.17 (m, 7H, H-3", 4" and 2'"); 6.81-6.78 (m, 4H, H-3'"); 6.51 (t, J = 6.56 Hz,

1H, H-1'); 5.44 (d, 1H, $J_{3',OH}$ = 4.63 Hz, H-O*H*) 4.55-4.53 (m, 1H, H-3'); 4.08 (t, *J* = 7.21 Hz, 2H, NC*H*$_2$CH$_2$CH$_2$CH$_2$CCH); 4.04-4.01 (m, 1H, H-4'); 3.72 (s, 6H, OC*H*$_3$); 3.20 (d, *J* = 4.98 Hz, 2H, H-5'); 2.98-2.93 (m, 1H, 1H, H-2'a); 2.64 (t, *J* = 2.66 Hz, 1H, NCH$_2$CH$_2$CH$_2$CH$_2$CC*H*); 2.45-2.40 (m, 1H, H-2'b); 2.01 (s, 3H, NCOC*H*$_3$); 2.04-2.00 (m, 2H, NCH$_2$CH$_2$CH$_2$C*H*$_2$CCH); 1.51-1.47 (m, 2H, NCH$_2$C*H*$_2$CH$_2$CH$_2$CCH); 1.37-1.31 (m, 2H, NCH$_2$CH$_2$C*H*$_2$CH$_2$CCH).

^{13}C NMR (125.7 MHz, DMSO-*d*$_6$) *δ* 170.79 (NCOCH$_3$); 158.47, 158.46 (2C, C4'''); 153.01, 152.96 (C1''); 152.07 (C2); 145.29 (C6); 145.15 (C8); 136.08, 135.94 (2C, C1'''); 130.14, 130.10 (4C, C2'''); 129.36 (C3''); 128.66 C3''); 128.12(C3'); 128.10(C4''); 127.69(C5); 127.03 (NCOCH$_3$); 125.77 (C4''); 113.51, 113.50 (4C, C3'''); 86.50 (C4'); 85.86 (C-Ar$_3$); 84.61 (NCH$_2$CH$_2$CH$_2$CH$_2$CCH); 84.32 (C1'); 71.58 (NCH$_2$CH$_2$CH$_2$CH$_2$CCH); 70.87; 64.42 (C5'); 55.45, 55.44 (2C, OCH3); 46.22 (NCH$_2$CH$_2$CH$_2$CH$_2$CCH); 38.89 (C2'); 27.69 (NCH$_2$CH$_2$CH$_2$CH$_2$CCH); 25.58 (NHCH$_2$CH$_2$CH$_2$CH$_2$CCH); 24.19; 17.74 (NH CH$_2$CH$_2$CH$_2$CH$_2$CCH).

HRMS (ESI) *m/z*: $[M+H]^+$ Calcd for $C_{39}H_{42}N_5O_6$ 676.3130; Found 676.3105.

3.5.4.7.3. 5'-*O*-(4,4'-Dimethoxytrityl)-*N*^{6}-(2-azidoethyl)-*N*^{6}-acetyl-2'-deoxyadenosine (13c)



R$_f$ = 0.40 (5% MeOH in DCM); yield: 0.81 g, 74%.

^{1}H NMR (500 MHz, DMSO-*d*$_6$): *δ* 8.75 (s, 1H, H-8); 8.72 (s, 1H, H-2); 7.32-7.30 (m, 2H, H-2''); 7.21-7.16 (m,6H, H-3'', 4'' and 2'''); 6.81-6.77 (m, 4H, H-3'''); 6.51 (dd, 1H, $J_{1',2'}$ = 5.9, 6.7 Hz, H-1'); 5.43 (d, 1H, $J_{3'OH}$ = 4.65 Hz, 3'-OH); 4.54-4.50 (m, 1H, H-3'); 4.27 (t, *J* = 6.17 Hz, 2H, NC*H*$_2$CH$_2$N$_3$); 4.09-4.02 (m, 1H, H-4'); 3.714 (s, 3H,OC*H*$_3$); 3.710 (s, 3H, OC*H*$_3$); 3.49 (t, *J* = 6.10 Hz, 1H, NCH$_2$C*H*$_2$N$_3$); 3.23-3.16 (m, 2H, H-5'); 2.95-2.90 (m, 1H, H-2'a); 2.43-2.38 (m, 1H, H-2'b); 2.11 (s, 3H, NCOC*H*$_3$).

^{13}C NMR (125.7 MHz, DMSO-*d*$_6$): *δ* 171.2 (NCOCH$_3$); 158.4, 158.4 (2C, C4'''); 152.9 (C6) 152.6 (C4); 151.9 (C2); 145.3 (C1''); 145.0 (C8); 136.0, 135.9 (2C, C1'''); 130.1 (4C, C2'''); 128.1 (2C, C3''); 128.1 (2C, C2''); 127.1 (C5); 127.0 (C4''); 113.5, 113.5 (4C, C3''');

86.5 (C4'); 85.8 (C-Ar$_3$); 84.3 (C1'); 70.9 (C3'); 64.5 (C5'); 55.4 (OCH3); 55.4 (OCH3); 49.8 (NCH_2CH$_2$N$_3$); 46.0 (NCH$_2$CH$_2$N$_3$); 38.9 (C2'); 24.2 (NCOCH$_3$).

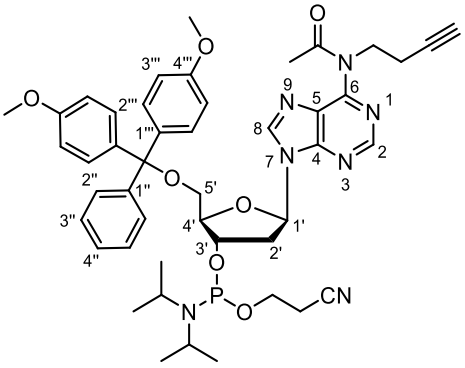
IR ATR (cm^{-1}): 3440.50, 2951.36, 2098.43, 1682.24, 1574.42, 1505.16, 1446.01, 1364.68, 1295.85, 1248.04, 1211.82, 1175.50, 1032.04, 826.80, 700.85, 579.51.

HRMS (ESI) *m/z*: $[M+H]^+$ Calcd for $C_{35}H_{37}N_8O_6$ 665.2836; found 665.2831.

3.5.4.8. General procedure for preparation of nucleoside phosphoramidites

To a stirring solution of 5'-*O*-DMT protected nucleoside (0.66 mmol) in dry DCM (10 mL) under argon at rt were added Et$_3$N (0.12 mL, 0.86 mmol) followed by 2-cyanoethyl *N*,*N*-di*iso*propyl chlorophosphoramidite (0.17 g, 0.71 mmol). After the disappearance of the starting material in 10 min the reaction mixture was washed with saturated sodium bicarbonate solution (5 × 3 mL) followed by brine (5 mL). The organic layer was dried by passing through anhydrous sodium sulfate column and the crude product was purified by column chromatography over silica gel (60-120 mesh) saturated with Et$_3$N (10%) and eluting with DCM/acetone (9:1) to give a white foam.

3.5.4.8.1. 5'-*O*-(4,4'-Dimethoxytrityl)-N^6-(but-3-yn-1-yl)-N^6-acetyl-3'-*O*-(*N*,*N*-di*iso*propylamino-2-cyanoethoxyphosphanyl)-2'-deoxyadenosine (14a)



R_f = 0.40 (10% acetone in DCM); yield: 0.45 g, 85%.

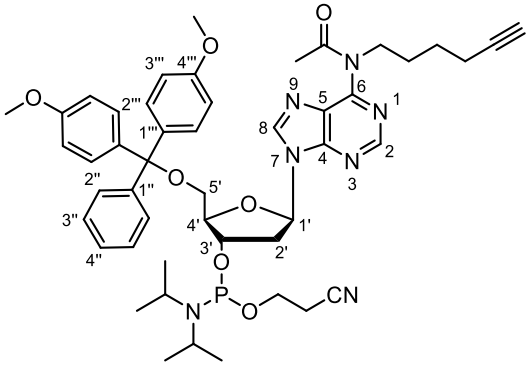
^{1}H NMR (500 MHz, DMSO-d_6) δ 8.74, 8.73 (2s, 1H, H-8); 8.70, 8.69 (2s, 1H, H-2); 7.32-7.29 (m, 2H, H-2''); 7.21-7.16 (m, 7H, H-3'', H-4''and H-2'''); 6.81-6.76 (m, 4H, H-3'''); 6.51 (q, *J* = 7.24 Hz, 6.69 Hz, 1H, H-1'); 4.85-4.79 (m, 1H, H-3'); 4.19-4.16 (m, 2H, NCH_2CH$_2$CCH), 4.15-4.13 (m, 1H, H-4'); 3.81-3.57 (m, 2H, CH_2CH$_2$CN); 3.71 (s), 3.70 (s) (6H, OCH$_3$); 3.68-3.63 (m, 1H, NCH_2CH$_2$CN); 3.60-3.50 (m, 2H, NCHCH$_3$); 3.30-3.20 (m, 2H, H-5'); 3.14-3.07 (m,1H, H-2'b); 2.76 (t, 1H, *J* = 5.83 Hz, 1H, NCH$_2$CH$_2$CCH); 2.68-2.65 (m, 2H, CH$_2$CH_2CN); 2.63-2.51 (m, 1H, H-2'a); 2.48-2.46 (m, 2H, NCH$_2$CH_2CCH); 2.11, 2.12 (2s, 3H, NCOCH_3); 1.14, 1.12 (2d, *J* = 6.8 Hz); 1.04 (d, *J* =6.78 Hz) (12H, NCHCH_3).

^{13}C NMR (125.7 MHz, DMSO-d_6) δ 170.84 (NCOCH$_3$); 158.25, 158.22 (2C, C4'''); 152.52 (C6); 152.47 (C4); 151.48 (1C, C2); 145.08, 145.04 (C8); 144.93, 144.92 (C1''); 135.66, 135.61, 135.54 (2C, C1'''); 129.86, 129.84 (4C, C2'''); 127.88, 127.82, 127.77 (4C, C2'', 3''); 126.99 (C5); 126.83 (C4''); 119.18, 119.00 (CN); 113.25, 113.23 (4C, C3'''); 85.76 (*C*-Ar3); 85.27, 85.24, 85.00, 84.96 (1C, C1'); 84.35, 84.26 (1C, C4'); 81.42 (NCH$_2$CH$_2$CCH); 73.40, 73.25; 72.71 (1C, C3'); 63.68, 63.56 (C5'); 58.74, 58.65, 58.59, 58.50 (1C, NCH$_2$CH$_2$CN); 55.20, 55.19 (2C, OCH$_3$); 45.70 (NHCH$_2$CH$_2$CCH); 42.88, 42.87, 42.79, 42.78 (2C, NCHCH$_3$); 37.61, 37.58, 37.39, 37.36 (C2'); 24.59, 24.55, 24.53, 24.50, 24.45, 24.38 (NCHCH$_3$); 24.08, 24.06 (1C, NCOCH$_3$); 20.10, 20.03, 19.98 (CH$_2$CH$_2$CN); 18.19 (NHCH$_2$CH$_2$CCH).

^{31}P NMR (202.5 MHz, DMSO-d_6, ref. 85% H$_3$PO$_4$) δ 147.83, 147.20 in ~1:1 ratio.

HRMS (ESI) *m/z*: $[M+H]^+$ Calcd for $C_{46}H_{55}N_7O_7P$ 848.3895; Found 848.3870.

3.5.4.8.2. 5'-*O*-(4,4'-Dimethoxytrityl)-N^6-(hex-5-yn-1-yl)-N^6-acetyl-3'-*O*-(*N*,*N*-di*iso*propylamino-2-cyanoethoxyphosphanyl)-2'-deoxyadenosine (14b)



R_f = 0.48 (10% acetone in DCM); yield: 0.7 g, 90%.

^{1}H NMR (500 MHz, DMSO-d_6) δ 8.76, 8.75 (2s, 1H, H-8); 8.74, 8.73 (2s, 1H, H-2); 7.30-7.29 (m, 2H, H-2''); 7.21-7.17 (m, 7H, H-3'', H-4'' and H-2'''); 6.82-6.77 (m, 4H, H-3'''); 6.53 (q, J = 7.17 Hz, 1H, H-1'); 4.90- 4.83 (m, 1H, H-3'); 4.19, 4.13 (2q, J = 4.68 Hz, 1H, H-4'); 4.08 (t, J = 6.83 Hz, 2H, NC*H*$_2$CH$_2$CH$_2$CH$_2$CCH); 3.72 (s), 3.71 (s) (6H, OCH$_3$); 3.69-3.63 (m, 2H, OC*H*$_2$CH$_2$CN); 3.58-3.50 (m, 2H, NC*H*CH$_3$); 3.22-3.12 (m, 2H, H-5'); 2.76 (t, 1H, J = 5.83 Hz, 1H, NCH$_2$CH$_2$CH$_2$CH$_2$CC*H*); 2.69-2.64 (m, 2H, OCH$_2$C*H*$_2$CN); 2.63-2.51 (m, 1H, H-2'a); 2.05, 2.04 (2s, 3H, NCOC*H*$_3$); 2.03-2.01 (m, 2H, NCH$_2$CH$_2$CH$_2$C*H*$_2$CCH); 1.50-1.46 (m, 2H, NCH$_2$C*H*$_2$CH$_2$CH$_2$CCH); 1.36-1.30 (m, 2H, NCH$_2$CH$_2$C*H*$_2$CH$_2$CCH); 1.15 (d, J = 6.8 Hz), 1.05 (d, J = 6.78 Hz) (12H, NCHC*H*$_3$).

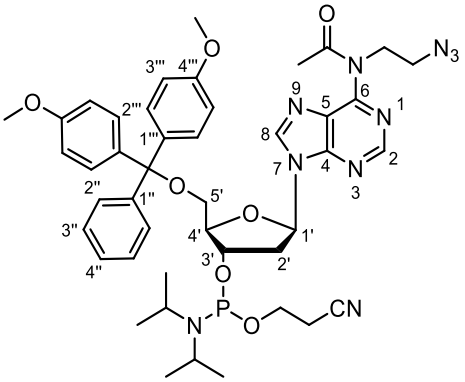
^{13}C NMR (125.7 MHz, DMSO-d_6) δ 170.77 (NCOCH$_3$); 158.51, 158.50 (2C, C4'''); 153.06 (C4), 152.80 (2C, C1''); 151.97 (C2); 145.50, 145.46 (C8); 145.18, 145.16 (C6);

135.95, 135.91, 135.88, 135.82 (2C, C1'''); 130.10, 130.07 (4C, C2'''); 128.12, 128.09, 128.04 (4C, C2'',3''); 127.79 (C5); 127.07 (C4''); 119.44, 119.25 (CN); 113.50 (4C, C3'''); 86.01, 86.00 (C1'); 85.47, 85.44 (C-Ar3); 84.65 ($NCH_2CH_2CH_2CH_2CCH$); 84.59, 84.50 (1C, C4'); 73.54; 73.39; 72.96; 72.83, 71.91 (C3'); 71.57, 63.82, 63.71 (1C, C5'); 59.01, 58.92, 58.86, 58.78 (1C, OCH_2CH_2CN); 55.46, 55.37 (2C, OCH_3); 46.22 ($NCH_2CH_2CH_2CH_2CCH$); 43.16, 43.14, 43.07, 43.04 (2C, $NCHCH_3$); 37.56 (C2'); 27.69 ($NCH_2CH_2CH_2CH_2CCH$), 27.63, 25.57 ($NHCH_2CH_2CH_2CH_2CCH$), 24.86, 24.80, 24.78, 24.72, 24.66 (4C, $NCHCH_3$); 24.20, 24.20 (1C, $NCOCH_3$); 20.37, 20.30, 20.24 (OCH_2CH_2CN); 17.75 ($NHCH_2CH_2CH_2CH_2CCH$).

^{31}P NMR (202.5 MHz, DMSO-d_6, ref. 85% H_3PO_4) δ 147.86, 147.25 in ~1:1 ratio.

HRMS (ESI) *m/z*: $[M+H]^+$ Calcd for $C_{48}H_{59}N_7O_7P$ 686.4208; Found 686.4181.

3.5.4.8.3. Synthesis of 5'-*O*-(4,4'-dimethoxytrityl)-N^6-(2-azidoethyl)-N^6-acetyl-3'-*O*-(*N*,*N*-diisopropylamino-2-cyanoethoxyphosphanyl)-2'-deoxyadenosine (14c)



To a stirring solution of compound **13c** (0.6 g, 0.90 mmol) in dry $CDCl_3$(5 mL), Et_3N (0.16 mL, 1 mmol) followed by 2-cyanoethyl-*N*,*N*-di*iso*propyl chlorophosphoramidite (0.27 mL, 1.08 mmol) were added under argon at 0 ^oC. After the consumption of the starting material, reaction mixture was washed with saturated $NaHCO_3$ solution (2 × 5 mL) followed by brine (5 mL). The organic layer was dried by passing through a column (12 × 1.5 cm) of anhydrous sodium sulfate. The solution of compound **14c** in $CDCl_3$ contains excess of phosphitylation reagent as shown by NMR (**Figure S69**). For automated DNA synthesis, the reaction was performed under the same conditions using dry DCM as a solvent. The concentration of azide phosphoramidite in DCM solution was calculated as 0.2 M based on UV absorption of DMT-cation at 504 nm (ε = 76000 $L \cdot mol^{-1} \cdot cm^{-1}$) that was released from an aliquot of compound **14c** by treatment with 3% dichloroacetic acid in DCM. R_f = 0.40 (10% acetone in DCM)

^{1}H NMR (500 MHz, $CDCl_3$): δ 8.72 (s, 1H, H-8); 8.27, 8.24 (2s, 1H, H-2); 7.39-7.37 (m, 2H, H-2''); 7.29-7.19 (m, 7H, H-3'', H-4'' and H-2'''); 6.80-6.77 (m, 4H, H-3'''); 6.54-6.51 (m, 1H, H-1'); 4.84-4.77 (m, 1H, H-3'); 4.40-4.37 (m, 2H, NC*H*$_2$CH$_2$N$_3$); 4.22-4.16 (m, 1H, H-4'); 4.15-4.09 (m, 2H, C*H*$_2$CH$_2$CN); 3.77 (2s, 6H, OCH_3); 3.60-3.50 (m, 2H, NC*H*CH_3); 3.50-3.46 (m, 2H, H-5'); 2.93 (ddd,1H, *J*= 6.6, 9.8, 13.4 Hz, H-2'b); 2.76-2.73 (m, 2H, CH$_2$C*H*$_2$CN); 2.64-2.61 (m, 1H, H-2'a); 2.58-2.54 (m, 2H, NCH$_2$C*H*$_2$N$_3$); 2.25, 2.24 (2s, 3H, NCOC*H*$_3$); 1.20-1.18 (m, 10H, NCHC*H*$_3$); 1.13 (d, 2H, NCHC*H*$_3$, *J* = 6.5 Hz).

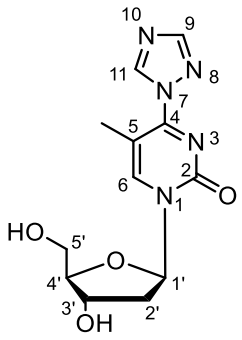
^{13}C NMR (125.7 MHz, $CDCl_3$): δ 171.6 (N<u>CO</u>CH_3); 158.5 (2C, C4'''); 153.0 (C6); 152.8 (C4); 152.7 (C6); 151.8 (1C, C2); 144.4 (C8); 142.5 (C1''); 135.56, 135.52, (2C, C1'''); 130.09, 130.06, 130.04, 130.02 (4C, C2'''); 128.1, 128.0, (4C, C2'', 3''); 127.8, 127.0, 126.99, 126.94, 126.91 (C4''); 117.58, 117.46 (CN); 116.9 (C5); 113.1 (4C, C3'''); 86.5 (C-Ar$_3$); 86.0, 85.8 (1C, C1'); 84.8, 84.7 (1C, C4'); 74.0, 73.9; 73.4, 73.3 (1C, C3'); 63.2, 63.1 (C5'); 58.3, 58.2, 58.16, 58.11 (1C, N<u>CH</u>$_2$CH$_2$CN); 55.2, 55.1 (2C, OCH_3); 49.98, 49.95 (1C, NH<u>CH</u>$_2$CH$_2$N$_3$); 46.1 (1C, 2'); 45.3, 45.2 (1C, NHCH$_2$<u>CH</u>$_2$N$_3$); 24.65, 24.61, 24.57, 24.51 (1C, NCO<u>CH</u>$_3$); 22.97, 22.95, 22.89, 22.87 (NCH<u>CH</u>$_3$); 20.46, 20.40, 20.2, 20.1, 20.05 (CH$_2$<u>CH</u>$_2$CN).

^{31}P NMR (202.5 MHz, DMSO-d_6, ref. 85% H_3PO_4) δ 148.8.

IR ATR (cm^{-1}): 3673.51, 2969.69, 2361.85, 2251.42, 2097.34, 1673.86, 1574.37, 1508.81, 1462.87, 1367.94, 1330.70, 1298.57, 1250.22, 1179.36, 1036.65, 911.31, 829.03, 732.85, 646.64, 558.98, 421.81.

3.5.5. Synthesis of 5'-*O*-(4,4'-dimethoxytrityl)-*N*^{4}-(2-azidoethyl)-3'-*O*-(*N*,*N*-di*iso*propylamino-2-cyanoethoxyphosphanyl)-2'-deoxy-5-methylcytidine

3.5.5.1. Synthesis of 4-(1*H*-1,2,4-triazol-1-yl)-2'-deoxy-5-methylcytidine (15)



Compound **15** was synthesized according to a reported procedure[193] (3.5 g, 58% yield).

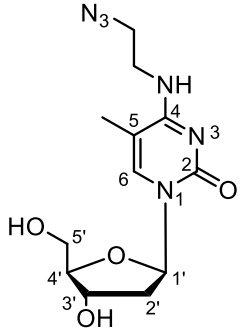
^{1}H NMR (500 MHz, DMSO-d_6) δ 9.30 (s, 1H, H-6); 8.61 (s, 1H, H-9); 8.37 (s, 1H, H-11); 6.11 (dd, 1H, $J_{1',2a'}$ = 5.6, $J_{1',2'b}$ = 7.3 Hz, H-1'); 5.31 (d, *J* = 3.8 Hz, 1H, 3'-OH); 5.21 (t, *J* = 5.3 Hz, 1H, 5'-OH); 4.27 (ddd, 1H, $J_{3',4'}$ = 1.9 Hz, $J_{3',2'a}$ = 3.6 Hz, $J_{3',2'b}$ = 5.6 Hz, H-3');

3.90 (ddd, 1H, $J_{3',4'}$ = 1.9 Hz, Hz $J_{4',5'b}$ = 3.5 Hz, H-4', $J_{4',5'a}$ = 3.7); 3.73 - 3.69 (m, 2H, H-5'a); 3.64 - 3.60 (m, 2H, H-5'b); 2.37 (m, 1H, H-2'a); 2.30 (s, 3H, CH_3); 2.14 (ddd, 1H, $J_{2'b,3'}$ = 5.6 Hz, $J_{2'b,1'=}$ 7.3 Hz, $J_{2'b,2'a}$ = 12.1 Hz, H-2'b).

^{13}C NMR (125.7 MHz, DMSO-d_6) δ 150.80 (1C, C4); 153.41 (1C, C11); 153.12 (1C, C2); 147.96 (1C. C9); 145.30 (1C, C6); 104.50 (1C, C5); 88.12 (1C, C4'); 86.89 (1C, C1'); 69.26 (1C, C3'); 60.44 (1C, C5'); 40.96 (1C, C2'); 16.19 (1C, CH_3).

HRMS (ESI) *m/z*: $[M+Na]^+$ Calcd for $C_{12}H_{15}O_4N_5Na$ 316.1022; Found 316.1013.

3.5.5.2. Synthesis of N^4-(2-azidoethyl)-2'-deoxy-5-methylcytidine (16)



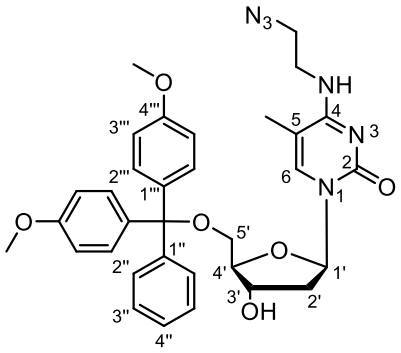
To a suspension of compound **15** (1.9 g, 6.5 mmol), in dry acetonitrile (20 mL) was added 2-azidoethylamine (2 g, 25 mmol) and the mixture was stirred at 50 ^{o}C for 30 min. After the consumption of the starting material **15**, the volatiles were evaporated *in vacuo*, and the crude product was purified by column chromatography over silica gel eluting with 0-30% MeOH in DCM to afford the desired product as an oil (1.6 g, 80% yield, R_f = 0.48, 20% MeOH in DCM).

^{1}H NMR (500 MHz, DMSO-d_6) δ 7.70 (s, 1H, H-6); 7.38 (t, J =4.9 Hz, 1H, $NHCH_2CH_2N_3$); 6.17 (dd, 1H, $J_{1',2a'}$ = 5.9, $J_{1',2'b}$ 7.3 Hz, H-1'); 5.21 (bs, 2H, 5'-OH, 3'-OH); 4.20 (ddd, 1H, $J_{3',4'}$ = 1.9 Hz, $J_{3',2'a}$ = 3.6 Hz, $J_{3',2'b}$ = 5.8 Hz, H-3'); 3.75 (ddd, 1H, $J_{3',4'}$ = 1.9 Hz, $J_{4',5'b}$ = 3.5 Hz, H-4', $J_{4',5'a}$ = 3.7 Hz); 3.59 (dd, 1H, $J_{4',5'a}$ = 3.7 Hz, $J_{5'a,5'b}$ = 11.8 Hz, H-5'a); 3.53 (dd, 1H, $J_{4',5'b}$ = 3.5 Hz, $J_{5'a,5'b}$ = 11.8 Hz, H-5'b); 3.52-3.50 (m, 2H, $NHCH_2CH_2N_3$); 3.47-3.45 (m, 2H, $NHCH_2CH_2N_3$); 2.08 (ddd, 1H, $J_{2'a,3'}$ = 3.6 Hz, $J_{2'a,1'}$ = 5.9 Hz, $J_{2'a,2'b}$ = 13.2 Hz, H-2'a); 1.97 (ddd, 1H, $J_{2'b,3'}$ = 5.8 Hz, $J_{2'b,1'}$ = 7.3 Hz, $J_{2'b,2'a}$ = 13.2 Hz, H-2'b); 1.87 (s, 3H, CH_3).

^{13}C NMR (125.7 MHz, DMSO-d_6) δ 162.93 (1C, C4); 154.87 (1C, C2); 137. 76 (1C, C6); 101.62 (1C, C5); 87.13 (1C, C4'); 84.63 (1C, C1'); 70.40 (1C, C3'); 61.35 (1C, C5'); 50.45 (1C, $NH\underline{CH_2}CH_2N_3$); 49.07 (1C, $NHCH_2\underline{CH_2}N_3$); 40.17 (1C, C2'); 13.06 (1C, CH_3).

HRMS (ESI) *m/z*: $[M+H]^+$ Calcd for $C_{12}H_{19}N_6O_4$ 311.1468; Found 311.1458.

3.5.5.3. Synthesis of 5'-*O*-(4,4'-dimethoxytrityl)-N^4-(2-azidoethyl)-2'-deoxy-5-methylcytidine (17)



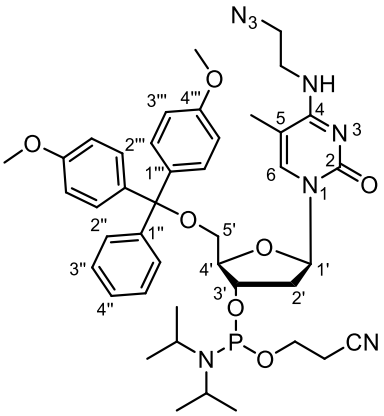
To a stirring solution of compound **16** (0.47 g, 1.5 mmol) in dry pyridine (10 mL), 4,4'-dimethoxytrityl chloride (0.62 g, 1.8 mmol) was added at 0 °C and the mixture was stirred at r.t. overnight under argon. After consumption of the starting nucleoside, water (1 mL) was added. Solvents were evaporated *in vacuo* and the residue was dissolved in 50 mL DCM, washed with brine (2 × 10 mL). Organic layer was dried over anhydrous sodium sulfate, filtered, and concentrated *in vacuo*. The crude product was purified by column chromatography over silica gel treated with 10% Et_3N in DCM and eluted with 20% MeOH in DCM to afford the desired product as a foam (0.83 g, 89% yield, R_f = 0.30, 10% MeOH in DCM).

^{1}H NMR (500 MHz, DMSO-d_6) δ 7.70 (s, 1H, H-6); 7.40-7.38 (m, 3H, H2", *NH*$CH_2CH_2N_3$); 7.31 (m 2H, H-3"); 7.27-7.22 (m, 5H, H-4" and H-2'"); 6.90- 6.6.89 (m, 4H, H-3'"); 6.23 (dd, 1H, $J_{1',2a'}$ = 5.9, $J_{1',2'b}$ 7.3 Hz, H-1'); 5.30 (d, J = 4.4 Hz, 1H, 3'-OH); 4.29 (ddd, 1H, $J_{3',4'}$ = 2.2 Hz, $J_{3',2'a}$ = 3.6 Hz, $J_{3',2'b}$ = 5.8 Hz, H-3'); 3.88 (ddd, 1H, $J_{3',4'}$ = 2.2 Hz, $J_{4',5'a}$ = 3.7 Hz $J_{4',5'b}$ = 3.5 Hz, H-4'); 3.73 (s, 6H, OCH_3); 3.51 (t, J =5.5 Hz, 2H, $NHCH_2$*CH_2*N_3); 3.48-3.45 (m, 2H, NH*CH_2*CH_2N_3); 3.23-3.16 (m, 2H, H5'); 2.20-2.16 (m, 1H, H-2'a); 2.15-2.08 (m, 1H, H-2'b); 1.51 (s, 3H, CH_3).

^{13}C NMR (125.7 MHz, DMSO-d_6) δ 162.93 (1C, C4); 158.17 (1C, C4'"); 154.76 (1C, C2); 144.71 (1C, C1"); 137.23 (1C, C6); 135.52 (1C, C1'"); 135.35 (1C, C1'"); 129.77 (2C, C2'"); 129.75 (2C, C2'"); 127.93 (2C, C3"); 127.73 (2C, C2"); 126.80 (1C, C4"); 113.27 (4C, C3'"); 101.83 (1C, C5); 85.84 (1C, C-Ar_3); 85.35 (1C, C4'); 84.51 (1C, C1'); 70.55 (1C, C3'); 63.69 (1C, C5'); 55.06 (2C, OCH_3); 49.06 (1C, $NHCH_2\underline{CH_2}N_3$); 40.48 (1C, C2'); 39.93 (1C, $NH\underline{CH_2}CH_2N_3$); 12.67 (1C, CH_3).

HRMS (ESI) *m/z*: $[M+Na]^+$ Calcd for $C_{33}H_{36}N_6O_6Na$ 635.2594; Found 635.2576.

3.5.5.4. Synthesis of 5'-*O*-(4,4'-dimethoxytrityl)-N^4-(2-azidoethyl)-3'-*O*-(*N*,*N*-di*iso*propylamino-2-cyanoethoxyphosphanyl)-2'-deoxy-5-methylcytidine (18)



Synthesis of compound **18** was performed in the same way as described for compound **14c**. The concentration of azide phosphoramidite in DCM solution was calculated to be 0.14 M based on UV absorption of DMT-cation at 504 nm (Ɛ = 76000 $L \cdot mol^{-1} \cdot cm^{-1}$) that was released from an aliquot of compound **18** by treatment with 3% dichloroacetic acid in DCM.

R_f = 0.45 (10% acetone in DCM).

^{1}H NMR (500 MHz, $CDCl_3$): δ 7.73, 7.67 (2s, 1H, H-6); 8.27, 8.24 (2s, 1H, H-2); 7.41-7.38 (m, 2H, H-2"); 7.30-7.19 (m, 6H, H-3", H-4" and H-2'"); 6.82-6.79 (m, 5H, H-3'" and H-2'"); 6.44-6.39 (m, 1H, H-1'); 5.42-5.38 (m, 1H, H-3'); 4.62-4.59 (m, 1H, H-4'); 4.12-4.07 (m, 2H, C*H*$_2$CHCN); 3.77, 3.76 (2s, 6H, OCH_3); 3.73-3.68 (m, 2H, NC*H*$_2$$CH_2N_3$); 3.56-3.46 (m, 5H, NC*H*CH_3, NCH_2C*H*$_2N_3$, H-5'a); 3.30-3.26 (m, 1H, H-5'b); 2.75- 2.72 (m, 1H, H-2'b); 2.59 (t, *J* = 6.5 Hz, 2H, CH_2C*H*$_2$CN); 2.29- 2.21 (m, 1H, H-2'a); 1.45, 1.42 (2s, CH_3); 1.14-1.12 (m, 12H, NCHC*H*$_3$).

^{13}C NMR (125.7 MHz, $CDCl_3$): δ 163.10 (1C, C4); 158.62 (2C, C4'"); 156.00 (C2); 144.47 (1C, C1"); 137.63 (C6); 135.47 (2C, C1'"); 130.21, 130.18 (2C, C2'"); 130.12 (2C, C2'"); 128.30 (2C, C2"); 128.21 (2C, C3"); 127.05, 127.01 (1C, C4"); 117.59, 117.39, 116.88 (CN); 113.18 (4C, C3'"); 102.04, 101.91 (1C, C5); 86.67, 86.65 (1C, C1'); 86.67, 85.65 (C-Ar3); 85.23, 85.20, 85.09, 85.05 (1C, C4'); 73.53, 73.39; 72.59, 72.45 (1C, C3'); 63.03, 62.60 (C5'); 58.36, 58.26, 58.21, 58.17 (1C, $\underline{CH_2}CH_2CN$); 55.22, 55.19 (2C, OCH_3); 50.47 (1C, $NH\underline{CH_2}CH_2N_3$); 40.85, 40.82, 40.68, 40.65 (1C, 2'); 40.29 (1C, $NHCH_2\underline{CH_2}N_3$); 22.94, 22.92, 22.86, 22.84 (2C, $NCH\underline{CH_3}$); 20.35, 20.29, 20.13, 20.07, 20.02 ($CH_2\underline{CH_2}CN$); 12.32, 12.30 (1C, CH_3).

^{31}P NMR (202.5 MHz, DMSO-d_6, ref. 85% H_3PO_4) δ 148.87, 148.32.

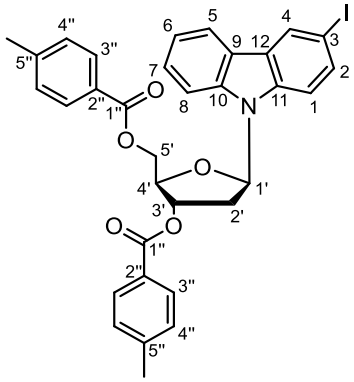
IR ATR (cm^{-1}): 3275.67, 2971.79, 2935.55, 2907.48, 2874.94, 2100.98, 1668.30, 1634.24, 1609.40, 1511. 06, 1464.34, 1253.62, 1180.73, 735.74.

3.5.6. Synthesis of modified carbazole nucleosides

3.5.6.1. Synthesis of 3-iodocarbazole (19)

3-Iodocarbazole was prepared according to a reported procedure.[194]

3.5.6.2. Synthesis of 9-[2-deoxy-3,5-di-*O*-(4-methylbenzoyl)-β-D-*erythro*-pentofuranosyl]-3-iodo-9*H*-carbazole (20)



To a solution of 3-iodocarbazole (3.2 g, 10 mmol) in 100 mL of 1,4-dioxane was added K-^{t}BuO (1.46 g, 13 mmol) followed by Hoffer's chlorosugar (4.19 g, 10.8 mmol) and the solution was stirred at rt for 2 hr. After the consumption of the starting materials as observed by TLC, the reaction mixture was diluted with EtOAc (100 mL) and washed with water. Organic layer was separated, dried over anhydrous sodium sulfate, filtered, and concentrated *in vacuo*. The crude product was purified by column chromatography over silica gel eluting with 0-10% EtOAc in hexane to afford the desired product as a foam (yield 4.5 g, 64%, with purity over 99% as β-nucleoside determined by ^{1}H NMR: R_f = 0.30, 10% EtOAc in hexane).

^{1}H NMR (500 MHz, $CDCl_3$) δ 8.38 (d, J =1.4 Hz, 1H, H-4); 8.06-8.00 (m, 5H, H-3", and H-6); 7.66-7.64 (m, 1H, H-8); 7.45-7.41 (m, 2H, H-2 and H-1); 7.34-7.31 (m, 4H, H-4"); 7.27-7.24 (m, 2H, H-7 and H-5); 6.69 (dd, 1H, $J_{1',2a'}$ = 5.8, $J_{1',2'b}$ = 9.3 Hz, H-1'); 5.85 (ddd, 1H, $J_{3',4'}$ = 1.8 Hz, $J_{3',2'a}$ = 3.7 Hz, $J_{3',2'b}$ = 5.7 Hz, H-3'); 4.92 (dd, 1H, $J_{4',5'a}$ = 2.7 Hz, $J_{5'a,5'b}$ = 12.2 Hz, H-5'a); 4.83 (dd, 1H, $J_{4',5'b}$ =3.4 Hz, $J_{5'a,5'b}$ = 12.2 Hz, H-5'b); 4.59 (ddd, 1H, $J_{3',4'}$ = 1.8 Hz, $J_{4',5'a}$ = 2.7 Hz $J_{4',5'b}$ = 3.4 Hz, H-4'); 3.18 (ddd, 1H, $J_{2'b,3'}$ = 5.7 Hz, $J_{2'b,1'}$ = 9.3 Hz, $J_{2'b,2'a}$ = 14.5 Hz, H-2'b); 2.51 (ddd, 1H, $J_{2'a,3'}$ = 3.7 Hz, $J_{2'a,1'}$ = 5.8 Hz, $J_{2'a,2'b}$ = 14.5 Hz, H-2'a); 2.50, 2.49 (2s, 6H, CH_3).

^{13}C NMR (125.7 MHz, $CDCl_3$) δ 166.47 (1C, C1"); 166.26 (1C, C1"); 144.63 (1C, C5"); 144.34 (1C, C5"); 139.17 (1C, C11); 138.32 (1C, C10); 134.21 (1C, C7); 129.94 (4C, C3" and C4"); 129.48, 129.46 (4C, C3" and C4"); 129.33 (1C, C4); 127.06 (1C, C2"); 126.69 (1C,

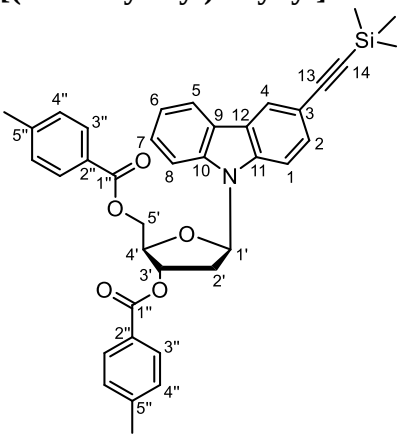
C9); 126.68 (1C, C2); 126.66 (1C, C12); 122.82 (1C, C2''); 120.65 (1C, C5); 112.87 (1C, C1); 110.86 (1C, C8); 85.15 (1C, C1'); 82.98 (1C, C3); 81.29 (1C, C4'), 74.21 (1C, C3'); 63.99 (1C, C5'); 35.67 (1C, C2'); 21.90 (2C, CH_3).

HRMS (ESI) *m/z*: $[M+Na]^+$ Calcd for $C_{33}H_{28}INNaO_5$ 668.0904; Found 668.0895.

3.5.6.3. General procedure for Sonogashira reaction of compound 20 with protected acetylenes

In a 100 mL round-bottom flask, compound **20** (1.0 g, 3.4 mmol), tetrakis(triphenylphosphine)palladium (0) (0.185 g, 0.16 mmol) and CuI (0.06 g, 3.1 mmol) were added. The flask was sealed with septum and carefully degassed *in vacuo* and then purged with argon. Dry DMF (15 mL) was added via syringe, followed by protected acetylene (4 eqv. with reference to compound **20**) and triethylamine (3.6 mmol). The content of the flask was stirred at 60 °C for 2 hr until compound **20** was completely consumed based on TLC analysis of the reaction mixture. The contents of the flask were diluted with EtOAc (30 mL), washed with distilled water (10 mL) followed by brine (10 mL). The organic layer was separated and dried over anhydrous sodium sulfate, filtered, and concentrated *in vacuo*. The crude product was purified by column chromatography over silica gel eluting with 0-15% EtOAc in hexane to afford the desired product as a foam.

3.5.6.3.1. 9-[2-Deoxy-3,5-di-*O*-(4-methylbenzoyl)-β-D-*erythro*-pentofuranosyl]-3-[(trimethylsilyl)ethynyl]-9*H*-carbazole (21a)



R_f = 0.39 (15% EtOAc in hexane); yield: 0.93 g, 97%.

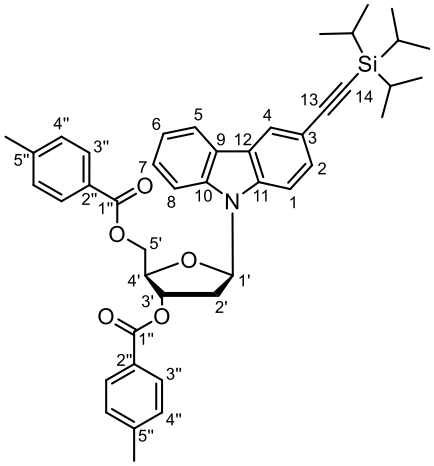
^{1}H NMR (500 MHz, $CDCl_3$) δ 8.19 (d, J =1.2 Hz, 1H, H-4); 8.04-7.99 (m, 5H, H-3'' H-6); 7.65-7.64 (m, 1H, H-8); 7.53 (d, J =8.5 Hz, 1H, H-1); 7.33-7.28 (m, 5H, H-4'' and H-2); 7.23-7.20 (m, 2H, H-7 and H-5); 6.68 (dd, 1H, $J_{1',2a'}$ = 5.8, $J_{1',2'b}$ 9.3 Hz, H-1'); 5.83 (ddd, 1H, $J_{3',4'}$ = 1.8 Hz, $J_{3',2'a}$ = 3.7 Hz, $J_{3',2'b}$ = 5.6 Hz, H-3'); 4.88 (dd, 1H, $J_{4',5'a}$ = 2.8 Hz, $J_{5'a,5'b}$ = 12.2 Hz, H-5'a); 4.81 (dd, 1H, $J_{4',5'b=}$ 3.5 Hz, $J_{5'a,5'b}$ = 12.2 Hz, H-5'b); 4.56 (ddd, 1H, $J_{3',4'}$ = 1.8

Hz, $J_{4',5'a}$ = 2.8 Hz $J_{4',5'b}$ = 3.5 Hz, H-4'); 3.18 (ddd, 1H, $J_{2'b,3'}$ = 5.6 Hz, $J_{2'b,1'=}$ 9.3 Hz, $J_{2'b,2'a}$ = 14.5 Hz, H-2'b); 2.37 (ddd, 1H, $J_{2'a,3'}$ = 3.7 Hz, $J_{2'a,1'}$ = 5.8 Hz, $J_{2'a,2'b}$ = 14.5 Hz, H-2'a); 2.45 (s, 6H, CH_3); 0.29 (s, 9H, Si(CH_3)$_3$).

^{13}C NMR (125.7 MHz, $CDCl_3$) δ 166.50 (1C, C1"; 166.26 (1C, C1"); 144.61 (1C, C5"); 144.28 (1C, C5"); 139.41 (1C, C11); 138.98 (1C, C10); 129.95 (2C, C3"); 129.94 (2C, C4"); 129.47 (5C, C3", C4" and C2); 127.07 (1C, C2"); 126.72 (1C, C2"); 126.41 (1C, C4); 124.54 (1C, C5); 124.00 (1C, C9); 123.74 (1C, C12); 120.69 (1C, C6); 120.65 (1C, C7); 114.65 (1C, C3); 111.14 (1C, C8); 110.48 (1C, C1); 106.33 (1C, C13); 92.29 (1C, C14), 85.18 (1C, C1'); 81.29 (1C, C4'); 74.25 (1C, C3'); 64.05 (1C, C5'); 35.70 (1C, C2'); 21.90, 21.87 (2C, CH_3).

HRMS (ESI) *m/z*: $[M + Na]^+$ Calcd for $C_{38}H_{37}NO_5SiNa$ 638.2333; Found 638.2324.

3.5.6.3.2. 9-[2-Deoxy-3,5-di-*O*-(4-methylbenzoyl)-β-D-*erythro*-pentofuranosyl]-3-[(tri*iso*propylsilyl)ethynyl]-9*H*-carbazole (21b)



R_f = 0.48 (15% EtOAc in hexane); yield: 1.03 g, 95%.

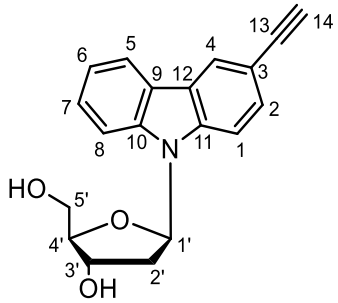
^{1}H NMR (500 MHz, DMSO-d_6) δ 8.32 (s, 1H, H-4); 8.26 (d, *J* =7.6 Hz, 1H, H-5); 8.03-7.99 (m, 4H, H-3"); 7.81 (d, *J* =8.1 Hz, 1H, H-8); 7.75 (d, *J* =8.5 Hz, 1H, H-1); 7.40-7.37 (m, 4H, H-4"); 7.30-7.23 (m, 2H, H-6 and H-7); 7.13 (d, *J* =8.5 Hz, 1H, H-2); 6.90 (dd, 1H, $J_{1',2a'}$ = 6.5, $J_{1',2'b}$ 8.4 Hz, H-1'); 5.88-5.85 (m, 1H, H-3'); 4.86 (dd, 1H, $J_{4',5'a}$ = 2.2 Hz, $J_{5'a,5'b}$ = 12.2 Hz, H-5'a); 4.68 (dd, 1H, $J_{4',5'b}$ = 3.7 Hz, $J_{5'a,5'b}$ = 12.2 Hz, H-5'b); 4.55 (ddd, 1H, $J_{4',5'a}$ = 2.2 Hz, $J_{3',4'}$ = 2.6 Hz, $J_{4',5'b}$ = 3.7 Hz, H-4'); 3.08 (ddd, 1H, $J_{2'b,3'}$ = 7.5 Hz, $J_{2'b,1'=}$ 8.4 Hz, $J_{2'b,2'a}$ = 14.8 Hz, H-2'b); 2.60 (ddd, 1H, $J_{2'a,3'}$ = 2.5 Hz, $J_{2'a,1'}$ = 6.5 Hz, $J_{2'a,2'b}$ = 14.8 Hz, H-2'a); 2.42, 2.41 (s, 6H, CH_3); 1.13 (s, 21H, Si(CHC_2H_6)$_3$).

^{13}C NMR (125.7 MHz, DMSO-d_6) δ 165.55 (2C, C1"); 144.04 (1C, C5"); 143.39 (1C, C5"); 139.08 (1C, C11); 138.22 (1C, C10); 129.60, 129.40; 129.37, 129.29 (9C, C4", C3" and C2); 126.70 (1C, C2") 126.56 (1C, C2"); 126.33 (C6); 123.94 (1C, C9); 123.28 (1C, C4); 122.50 (1C, C12); 120.94 (1C, C5); 120.41 (1C, C7); 113.61 (1C, C3); 111.56 (1C, C8);

111.57 (1C, C1); 108.32 (1C, C13); 87.55 (1C, C14); 84.09 (1C, C1'); 80.18 (1C, C4'); 73.53 (1C, C3'); 63.60 (1C, C5'); 37.79 (1C, C2'); 21.22, 21.17 (2C, CH_3); 18.55 (3C, SiCH); 10.80 (6C, CH_3).

HRMS (ESI) *m/z*: $[M+Na]^+$ Calcd for $C_{44}H_{49}O_5SiNNa$ 722.3272; Found 722.3259.

3.5.6.4. Synthesis of 9-[2-deoxy-β-D-*erythro*-pentofuranosyl]-3-ethynyl-9*H*-carbazole (22a)



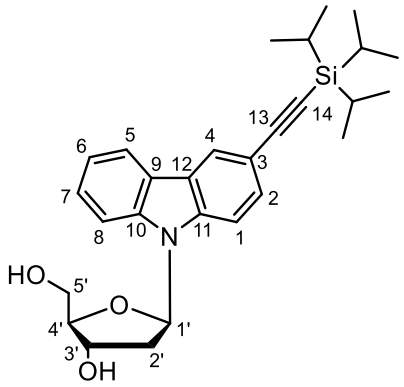
To a solution of compound **21a** (0.96 g, 1.6 mmol) in methanol (30 mL), was added K_2CO_3 (0.47 g, 3.4 mmol) and the mixture was stirred for 2 hr at rt. After the consumption of the starting material as observed by TLC, the solvent was evaporated *in vacuo*, and the residue was dissolved in 100 mL DCM. The DCM solution was washed with distilled water (20 mL) followed by brine (10 mL). The organic layer was dried over anhydrous sodium sulfate, filtered, and concentrated *in vacuo*. The crude product was purified by column chromatography over silica gel eluting with 0-5% MeOH in DCM to afford the desired product as a foam (0.55 g, 86% yield; R_f = 0.29, 5% MeOH in DCM).

^{1}H NMR (500 MHz, DMSO-d_6) δ 8.34 (d, J = 1.4 Hz, 1H, H-4); 8.21 (d, J = 7.7 Hz, 1H, H-5); 7.81 (s, 1H, H-8); 7.80 (s, 1H, H-1); 7.50 (dd, 1H, J = 1.6, 8.5 Hz, H-7); 7.47-7.43 (m, 1H, H2); 7.26-7.24 (m, 1H, H-6); 6.68 (dd, 1H, $J_{1',2a'}$ = 6.5, $J_{1',2'b}$ = 8.5 Hz, H-1'); 5.38 (d, $J_{3'\text{-}OH,3'}$ = 4.6 Hz, 1H, 3'-OH); 5.04 (t, $J_{5'\text{-}OH,5'}$ = 5.2 Hz,1H, 5'-OH); 4.48- 4.44 (m, 1H, H-3'); 4.06 (s, 1H, H-14); 3.87 (ddd, 1H, $J_{4',5'a}$ = 3.3 Hz, $J_{3',4'}$ = 4.1 Hz, $J_{4',5'b}$ = 7.9 Hz, H-4'); 3.79-3.70 (m, 2H, H-5'); 2.64 (ddd, 1H, $J_{2'b,3'}$ = 6.5 Hz, $J_{2'b,1'}$ = 8.5 Hz, $J_{2'b,2'a}$ = 13.5 Hz, H-2'b); 2.12 (ddd, 1H, $J_{2'a,3'}$ = 2.8 Hz, $J_{2'a,1'}$ = 6.5 Hz, $J_{2'a,2'b}$ = 13.5 Hz, H-2'a).

^{13}C NMR (125.7 MHz, DMSO-d_6) δ 139.17 (1C, C11); 138.56 (1C, C10); 129.20 (1C, C7); 126.42 (1C, C2) 124.02 (1C, C5); 123.11 (1C, C9); 122.40 (1C, C12); 120.65 (1C, C4); 120.11 (1C, C6); 112.63 (1C, C3); 111.60 (1C, C8); 111.43 (1C, C1); 86.50 (1C, C4'); 84.64(1C, C13); 84.02 (1C, C1'); 78.78 (1C, C14); 70.08 (1C, C3'); 61.24 (1C, C5'); 37.78 (1C, C2').

HRMS (ESI) m/z: $[M+Na]^+$ Calcd for $C_{19}H_{17}O_3NNa$ 330.1101; Found 330.1097.

3.5.6.5. Synthesis of 9-[2-deoxy-β-D-*erythro*-pentofuranosyl]-3-[(tri*iso*propylsilyl)ethynyl]-9*H*-carbazole (22b)



Compound **21a** was dissolved in MeOH (500 mL) and 28% aq. ammonia (50 mL) was added in one portion. Reaction mixture was stirred at r.t. for 48 h, evaporated *in vacuo*, co-evaporated with H_2O (2 × 200 mL). The crude product was purified by column chromatography over silica gel eluting with 0-5% MeOH in DCM to afford the desired product as a foam (0.65 g, 93% yield; R_f = 0.35, 5% MeOH in DCM).

^{1}H NMR (500 MHz, DMSO-d_6) δ 8.33 (s, 1H, H-4); 8.26 (d, J = 7.7 Hz, 1H, H-5); 7.82-7.80 (m, 2H, H-1 and H-8); 7.50- 7.44 (m, 2H, H-2 and H-7); 7.25 (dd, J =7.5, 7.40 Hz, 1H, H-6); 6.68 (dd, 1H, $J_{1',2a'}$ = 7.1, $J_{1',2'b}$ = 7.7 Hz, H-1'); 5.38 (d, $J_{3'OH,3'}$ = 4.6 Hz, 1H, 3'-OH); 5.03 (t, $J_{5'OH,5'}$ = 5.2 Hz, 1H, 5'-OH); 4.48-4.44 (m, 1H, H-3'); 3.87 (ddd, 1H, $J_{4',5'a}$ = 3.2 Hz, $J_{3',4'}$ = 4.4 Hz, $J_{4',5'b}$ = 9.0 Hz, H-4'); 3.79-3.70 (m, 2H, H-5'); 2.64 (ddd, 1H, $J_{2'b,3'}$ = 6.5 Hz, $J_{2'b,1'}$ = 7.7 Hz, $J_{2'b,2'a}$ = 13.6 Hz, H-2'b); 2.12 (ddd, 1H, $J_{2'a,3'}$ = 2.8 Hz, $J_{2'a,1'}$ = 7.1 Hz, $J_{2'a,2'b}$ = 13.6 Hz, H-2'a); 1.14 (s, 21H, Si(C*H*C$_2$*H*$_6$)$_3$).

^{13}C NMR (125.7 MHz, DMSO-d_6) δ 139.14 (1C, C11); 138.60 (1C, C10); 129.53 (1C, C7); 126.48 (1C, C2) 123.82 (1C, C5); 123.07 (1C, C9); 122.48 (1C, C12); 120.79 (1C, C8); 120.12 (1C, C6); 113.28 (1C, C3); 111.50 (2C, C1 and C4); 108.56 (1C, C13); 87.44 (1C, C14); 86.51 (1C, C4'); 84.01 (1C, C1'); 70.08 (1C, C3'); 61.24 (1C, C5'); 37.43 (1C, C2'); 18.59 (3C, SiCH); 10.84 (6C, CH_3).

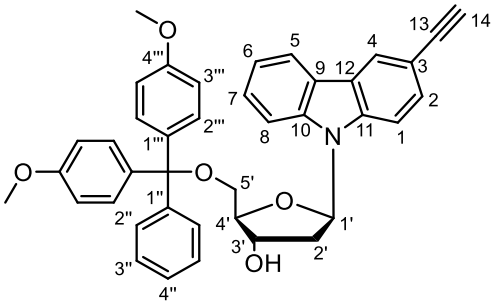
HRMS (ESI) *m/z*: $[M + Na]^+$ Calcd for $C_{28}H_{37}O_3NSiNa$ 486.2435; Found 486.2432.

3.5.6.6. General protocol for 5'-*O*-4,4'-dimethoxytritylation of modified carbazole nucleosides

To a stirring solution of a nucleoside (0.32 g, 1.0 mmol) in dry pyridine (10 mL), 4,4'-dimethoxytrityl chloride (0.67 g, 1.2 mmol) was added at 0 ^{o}C and the mixture was stirred at r.t. overnight under argon. After consumption of the starting nucleoside, water (1 mL) was added. Solvents were evaporated *in vacuo* and the residue was dissolved in 50 mL DCM and

washed with brine (2 × 10 mL). Organic layer was dried over anhydrous sodium sulfate, filtered, and concentrated *in vacuo*. The crude product was purified by column chromatography over silica gel treated with 10% Et_3N in DCM and eluted with 10% acetone in DCM to afford the desired product as a foam.

3.5.6.6.1. 9-[2-Deoxy-5-*O*-(4,4'-dimethoxytrityl)-β-D-*erythro*-pentofuranosyl]-3-ethynyl-9*H*-carbazole (23a)



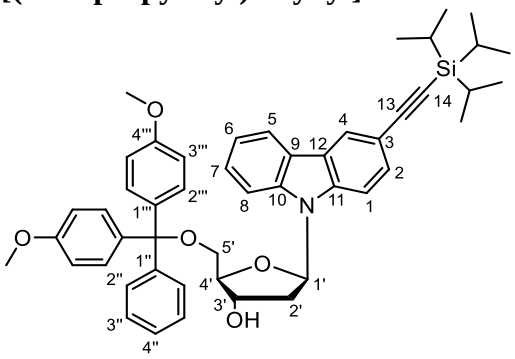
R_f = 0.40 (10% acetone in DCM); yield: 0.5 g, 79%.

^{1}H NMR (500 MHz, $CDCl_3$) δ 8.36 (d, J = 1.3 Hz, 1H, H-4); 8.04- 8.02 (m, 1H, H-5); 7.65- 7.64 (m, 1H, H-8); 7.66-7.63 (m, 1H, H-4''); 7.59 (d, 1H, J = 8.5 Hz, 1H, H-1); 7.51-7.48 (m, 2H, H-2''); 7.40-7.37 (m, 4H, H-2'''); 7.36-7.34 (m, 1H, H-2); 7.30- 7.28 (m, 3H, H-3'' and H-6); 7.25-7.22 (m, 2H, H-2, H-7); 6.84-6.81 (m, 4H, H-3'''); 6.62 (dd, 1H, $J_{1',2a'}$ = 6.5, $J_{1',2'b}$ = 8.5 Hz, H-1'); 4.77 (ddd, 1H, $J_{3',4'}$ = 1.8 Hz, $J_{3',2'a}$ = 3.2 Hz, $J_{3',2'b}$ = 7.1 Hz, H-3'); 4.08 (ddd, 1H, $J_{3',4'}$ = 1.8 Hz, $J_{4',5'a}$ = 2.8 Hz $J_{4',5'b}$ = 3.8 Hz, H-4'); 3.79 (s, 3H, OCH_3); 3.78 (s, 3H, OCH_3); 3.56 (d, 2H, J = 3.8 Hz, H-5'); 3.05 (s, 1H, H-14); 2.90 (ddd, 1H, $J_{2'b,3'}$ = 7.1 Hz, $J_{2'b,1'}$ = 8.5 Hz, $J_{2'b,2'a}$ = 13.9 Hz, H-2'b); 2.22 (ddd, 1H, $J_{2'a,3'}$ = 3.2 Hz, $J_{2'a,1'}$ = 6.5 Hz, $J_{2'a,2'b}$ = 13.9 Hz, H-2'a).

^{13}C NMR (125.7 MHz, $CDCl_3$) δ 158.73 (2C, C4'''); 144.75 (1C, C1''); 139.63 (1C, C10); 139.14 (1C, C11); 135.86 (1C, C1'''); 135.80 (1C, C1'''); 130.35 (4C, C2'''); 129.81 (1C, C7); 128.41 (1C, C2''); 128.09 (1C, C2''); 127.10 (2C, C3''); 126.43 (1C, C4''); 124.55 (1C, C5); 123.96 (1C, C2); 123.93 (1C, 12); 123.41 (1C, C9); 120.48 (1C, C8); 120.44 (1C, C6); 113.38 (4C, C3'''); 113.19 (1C, C3); 111.26 (1C, C1); 111.11 (1C, C4); 86.88 (1C, C-Ar_3); 84.92 (1C, C13); 84.71 (1C, C4'); 84.67 (1C, C1'); 75.49 (1C. C14); 72.11 (1C, C3'); 63.37 (1C, C5'); 55.38 (2C, OCH_3); 38.37 (1C, C2').

HRMS (ESI) *m/z*: $[M + Na]^+$ Calcd for $C_{40}H_{35}O_5NNa$ 632.2407; Found 632.2401.

3.5.6.6.2. 9-[2-Deoxy-5-*O*-(4,4'-dimethoxytrityl)-β-D-*erythro*-pentofuranosyl]-3-[(tri*iso*propylsilyl)ethynyl]-9*H*-carbazole (23b)



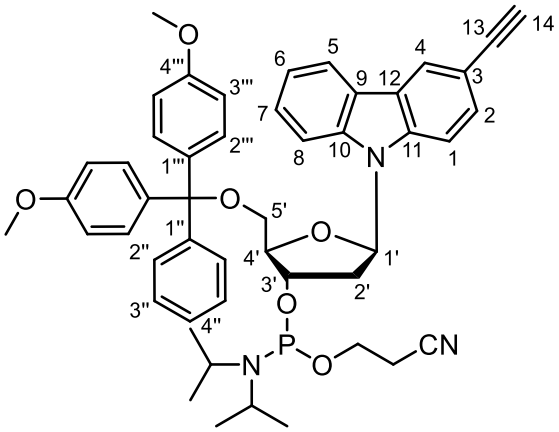
R_f = 0.49 (10% acetone in DCM; yield: 0.45 g, 86%.

^{1}H NMR (500 MHz, $CDCl_3$) δ 8.33 (d, J =1.2 Hz, 1H, H-4); 8.27 (d, J =7.5 Hz, 1H, H-5); 7.81-7.83 (m, 2H, H-1 and H-8); 7.44-7.41 (m, 2H, H-2''); 7.30- 7.25 (m, 7H, H-3'', H-4'' and H-2'''); 7.24-7.20 (m, 3H, H-7, H-2 and H-6); 6.86-6.84 (4H, H-3'''); 6.73 (dd, 1H, $J_{1',2a'}$ = 6.9, $J_{1',2'b}$ = 8.1 Hz, H-1'); 5.48 (d, J = 4.9 Hz, 1H, 3'-OH); 4.62 (ddd, 1H, $J_{3',4'}$ = 1.8 Hz, $J_{3',2'a}$ = 3.7 Hz, $J_{3',2'b}$ = 5.6 Hz, H-3'); 4.0 (ddd, 1H, $J_{3',4'}$ = 1.8 Hz, $J_{4',5'a}$ = 4.9 Hz, $J_{4',5'b}$ = 2.3 Hz, H-4'); 3.72 (s, 6H, OCH_3); 3.41 (dd, 1H, $J_{4',5'a}$ = 4.9 Hz, $J_{5'a,5'b}$ = 10.5 Hz, H-5'a); 3.36 (dd, 1H, $J_{4',5'b}$ = 2.3 Hz, $J_{5'a,5'b}$ = 10.5 Hz, H-5'b); 2.71 (ddd, 1H, $J_{2'b,3'}$ = 5.6 Hz, $J_{2'b,1'}$ = 8.1 Hz, $J_{2'b,2'a}$ = 13.6 Hz, H-2'b); 2.20 (ddd, 1H, $J_{2'a,3'}$ = 3.7 Hz, $J_{2'a,1'}$ = 6.9 Hz, $J_{2'a,2'b}$ = 13.6 Hz, H-2'a); 1.14 (s, 21H, Si(CHC$_2H_6)_3$).

^{13}C NMR (125.7 MHz, $CDCl_3$) δ 158.10 (1C, C4'''); 158.05 (1C, C4'''); 144.86 (1C, C1'') 139.18 (1C, C10); 138.33 (1C, C11); 135.38 (1C, C1'''); 135.34 (1C, C1'''); 129.90 (2C, C2'''); 129.77 (2C, C2'''); 129.45 (1C, C7); 127.81 (4C, C2'' and C3''); 126.65 (1C, C8); 126.33 (1C, C1); 123.87 (1C, C2); 123.19 (1C, C12); 122.44 (1C, C9); 120.86 (1C, C6); 120.17 (1C, C4''); 113.36 (1C, C3); 113.16 (4C, C3'''); 111.75 (1C, C5); 111.20 (1C, C4); 108.49 (1C, C13); 87.45 (1C, C14); 85.67 (C-Ar$_3$); 84.48 (1C, C4'); 83.93 (1C, C1'); 69.81 (1C, C3'); 62.96 (1C, C5'); 54.99 (1C, OCH_3); 54.98 (1C, OCH_3); 37.99 (1C, C2'); 18.57 (3C, SiCH); 10.82 (6C, CH_3).

HRMS (ESI) *m/z*: $[M + Na]^+$ Calcd for $C_{49}H_{55}O_5NSiNa$ 788.3742; Found 788.3732.

3.5.6.7. Synthesis of 9-[3-*O*-(*N*,*N*-di*iso*propylamino-2-cyanoethoxyphosphanyl)-2-deoxy-5-*O*-(4,4'-dimethoxytrityl)-β-D-*erythro*-pentofuranosyl]-3-ethynyl-9*H*-carbazole (24)



Compound **24** was synthesized using the same procedure as for compound **14a**.

R_f = 0.45 (5% acetone in DCM; yield: 0.57 g, 92%.

^{1}H NMR (500 MHz, DMSO-d_6) δ 8.35 (s, 1H, H-4); 8.23-8.22 (m, 1H, H-5); 7.83-7.80 (m, 2H, H-1 and H-8); 7.45-7.41 (m, 2H, H-2"); 7.31-7.20 (m, 10H, H-3", H-4", H-2'", H-7, H-2 and H-6); 6.86-6.84 (m, 2H, H-3'"); 6.83-6.80 (m, 2H, H-3'"); 6.78 (dd, 1H, $J_{1',2a'}$ = 6.8, $J_{1',2'b}$ = 8.2 Hz, H-1'); 4.96-4.83 (m, 1H, H-3'); 4.14-4.09 (m, 1H, H-4'); 4.06 (s, 1H, C*H*); 3.72, 3,70 (2s, 6H, O*CH*$_3$); 3.66-3.60 (m, 2H, C*H*$_2$CH$_2$CN); 3.59-3.50 (m, 2H, NC*H*CH$_3$); 3.46-3.37 (m, 2H, H-5'); 2.86-2.79 (m, 1H, H-2'b); 2.74, 2.66 (2t, *J* =5.9 Hz, 2H, CH$_2$*CH*$_2$CN); 2.46-2.34 (m, 1H, H-2'a); 1.14-1.08 (m, 10H, NCHC*H*$_3$); 0.99 (d, *J* =6.8 Hz, 2H, NCHC*H*$_3$).

^{13}C NMR (125.7 MHz, DMSO-d_6) δ 158.13 (2C, C4'"); 144.69, 144.65 (1C, C1") 139.01, 138.98 (1C, C10); 138.33 (1C, C11); 135.37, 135.27, 135.24, 135.18 (2C, C1'"); 129.84 (4C, C2'"); 129.15 (1C, C7); 127.80 (4C, C2" and C3"); 126.72 (1C, C5); 126.38 (1C, C2); 124.09 (1C, C4); 123.20 (1C, C12); 122.45 (1C, C9); 120.76 (1C, C4"); 120.26 (1C, C6); 118.97, 118.78 (CN);113.14 (4C, C3'"); 112.77 (1C, C13); 112.79, 111.60, 111.54 (1C C1), 111.38, 111.32 (1C, C8); 85.78, 85.75 (*C*-Ar$_3$); 84.50 (1C, C3); 83.77, 83.70 (1C, C1'); 83.12, 82.93 (C4'); 78.84 (1C, C14); 72.24, 72.10, 71.27, 71.13 (1C, C3'); 62.22, 61.97 (1C, C5'); 58.50, 58.34, 58.19 (1C, CH$_2$CH$_2$CN); 55.02, 55.00 (2C, OCH$_3$); 42.63, 42.53 (2C, NCHCH$_3$); 36.80, 36.59 (1C, C2'); 24.37, 24.30, 24.25, 24.19 (4C, NCHCH$_3$); 19.86, 19.81, 19.76 (1C, CH$_2$CH$_2$CN).

HRMS (ESI) *m/z*: [M +Na]$^+$ Calcd for $C_{49}H_{52}O_6N_3Na$ 832.3486; Found 832.3469.

3.5.6.8. Synthesis of 9-[2-deoxy-5-*O*-(4,4'-dimethoxytrityl)-β-D-*erythro*-pentofuranosyl]-3-[(tri*iso*propylsilyl)ethynyl]-9*H*-carbazole bound to CPG support (25)

Compound **23b** was covalently attached to the CPG support using procedure described for compound **4**. The load on the CPG **25** was determined to be 48 μmol/g based on UV absorption of DMT-cation at 504 nm (ε = 76000 $L \cdot mol^{-1} \cdot cm^{-1}$) that was released from an aliquot of compound by treatment with 3% dichloroacetic acid in DCM.

3.5.7. Oligonucleotide synthesis and purification

Oligonucleotides were prepared on a MerMade-4 DNA/RNA synthesizer (BioAutomation) on a 5 μmol scale using standard manufacturer's protocol. Coupling times of modified phosphoramidites were increased from 2 to 10 min. The final detritylated oligos were cleaved from the solid support and deprotected at rt using conc. NH_4OH (for dC and dZ-containing oligos) or 10% Et_2NH in acetonitrile followed by ethylenediamine/toluene (for FdZ-containing oligos and eluted from the column with 1 mL milli-Q water). The deprotected oligos in solution were freeze-dried and dry pellets were dissolved in milli-Q water (1 mL) and purified and isolated by reverse phase HPLC on 250/4.6 mm, 5 μm, 300 Å C18 column (Thermo Fisher Scientific) in a gradient of CH_3CN (0→20% for 20 min, 1.3 mL/min) in 0.1 M TEAA buffer (pH 7.0) with a detection at 260 nm. For the deprotection of TIPS protecting group in oligos, the dry pellets were dissolved in a solution of *o*-nitrophenol (0.050 mmol) in 100 μL THF, followed by 200 μL of 1 M TBAF in THF. The mixture was kept at 22 °C for 30 min. The reaction mixture was quenched by the addition of 2 M TEAA buffer (pH 7.0) and purified by reverse phase HPLC on 250/4.6 mm, 5 μm, 300 Å C18 column (Thermo Fisher Scientific) in a gradient of CH_3CN (0→80% for 14 min, 1.3 mL/min) in 0.1 M TEAA buffer (pH 7.0) with a detection at 260 nm. Oligonucleotides were freeze-dried, pellets were dissolved in milli-Q water (1.5 mL) and desalted by reverse-phase HPLC on 100/10 mm, 5 μm, 300 Å C18 column (Phenomenex) in a gradient of CH_3CN (0→80% for 15 min, 5 mL/min) in milli-Q water with detection at 260 nm. Pure products were quantified by measuring absorbance at 260 nm, analyzed by ESI-MS and concentrated by freeze-drying.

3.5.8. Cross-linking of linear oligonucleotides by CuAAC

The purified linear oligonucleotides were cross-linked by copper(I)-catalyzed azide alkyne cycloaddition (CuAAC) using the following protocol implemented in our laboratory (**Table 3.10** for oligos that did not contain carbazole nucleosides).

Table 3.10. Protocol employed for the CuAAC reaction.

ODs	5<	5-10	10-15	15-20	20-25	25-30	30-35	35-40
Oligonucleotide, µL	30	60	90	120	150	180	210	240
2 M TEAA (pH 7.0), µL	10	20	30	40	50	60	70	80
DMSO, µL	40	80	120	160	200	240	280	320
'Click catalyst', µL*	10	20	30	40	50	60	70	80
Sodium ascorbate (10 mM), µL	5	10	15	20	25	30	35	40

*The 'click catalyst' is prepared fresh by mixing 10 mM of tris[(1-benzyl-1*H*-1,2,3-triazol-4-yl)methyl]amine (TBTA) dissolved in DMSO and 10 mM of Cu(II) sulfate in water (1:1 v/v).

The oligonucleotides (except oligos containing carbazole in the sequence) were dissolved in the required quantity of 2 M TEAA (pH 7.0) followed by addition of DMSO. The bright blue colored 'click catalyst' was then added to the above contents in a tube that was purged with argon, followed by addition of freshly prepared sodium ascorbate solution in water. Reaction mixture was kept overnight at room temperature. A 10 µL aliquot of the reaction mixture was taken and the oligonucleotide was precipitated by adding 25 µL of 2 M $LiClO_4$ and five-fold volume of acetone. The contents were centrifuged using an Eppendorf MiniSpin Plus at 14500 rpm, 14100 ×*g*, for 2 min. The supernatant was discarded, and the pellet was washed carefully with acetone and dried. The dry pellet was dissolved in 10 µL milli-Q water and analyzed by reverse-phase HPLC for the product formation. For oligos containing carbazole in the sequence, 10 mM aq. solution of tris(3-hydroxypropyltriazolylmethyl)amine (THPTA) and 10 mM Cu(II) sulfate in water (1:1 v/v) were used instead of TBTA/Cu(II) complex as a 'click' catalyst. For short oligos, 10-fold excess of 2 M aq. TEAA was used to keep the solution diluted to prevent formation of dimers and multimers. After the completion of the reaction as shown by reverse phase HPLC, the oligos without carbazole in the sequence were precipitated by adding 1 mL of 2 M aq. $LiClO_4$ followed by a ten-fold volume of acetone and mixed well. The contents were centrifuged using a Thermo Fisher Heraeus Multifuge X1R centrifuge with swing bucket rotor at 5000 rpm, 4700 ×*g* for 30 min. The supernatant was discarded, and the pellet was washed carefully with acetone and dried. The dry pellet was dissolved in 10 mL milli-Q water and the product was isolated by reverse-phase HPLC on 250/4.6 mm, 5 µm, 300 Å C18 column (Thermo Fisher Scientific) in a gradient of CH_3CN (0→20% for 20 min, 1.3 mL/min) in 0.1 M TEAA

buffer (pH 7.0) with a detection at 260 nm. Carbazole-containing oligos were purified directly after the 'click' reaction without further treatments. Short oligos were purified by reverse-phase HPLC on 250/4.6 mm, 5 µm, 300 Å C18 column (Thermo Fisher Scientific) in a gradient of CH_3CN (0→80% for 14 min, 1.3 mL/min) in 0.1 M TEAA buffer (pH 7.0) with a detection at 260 nm. Purified oligos were freeze-dried, pellets were dissolved in milli-Q water (1.5 mL) and desalted by reverse-phase HPLC on 100/10 mm, 5 µm, 300 Å C18 column (Phenomenex) in a gradient of CH_3CN (0→80% for 15 min, 5 mL/min) in milli-Q water with detection at 260 nm. Cross-linked oligos were quantified by measuring absorbance at 260 nm, analyzed by ESI-MS and concentrated by freeze-drying.

3.5.9. Kinetic characterization of 5'-T_4**dC**AT as a substrate of A3B$_{CTD}$-QM-ΔL3-AL1swap

For inhibition assays, we decided to use a 7-mer DNA oligo 5'-T_4**dC**AT as a substrate instead of our previously reported procedure with 9-mer oligo 5'-AT_3**dC**AT_3. This was due to the complex ^{1}H NMR spectrum arising from the broad singlets of 2'-deoxyadenosines' amines seen for the 9-mer oligo substrate (**Figure 3.31a**). In contrast, 5'-T_4**dC**AT produced a much cleaner spectrum (**Figure 3.31b**), making it easier for evaluation and quantification of the deamination reaction. A3-catalyzed deamination was evaluated using the NMR-based assay described previously.[54, 124, 126, 140]

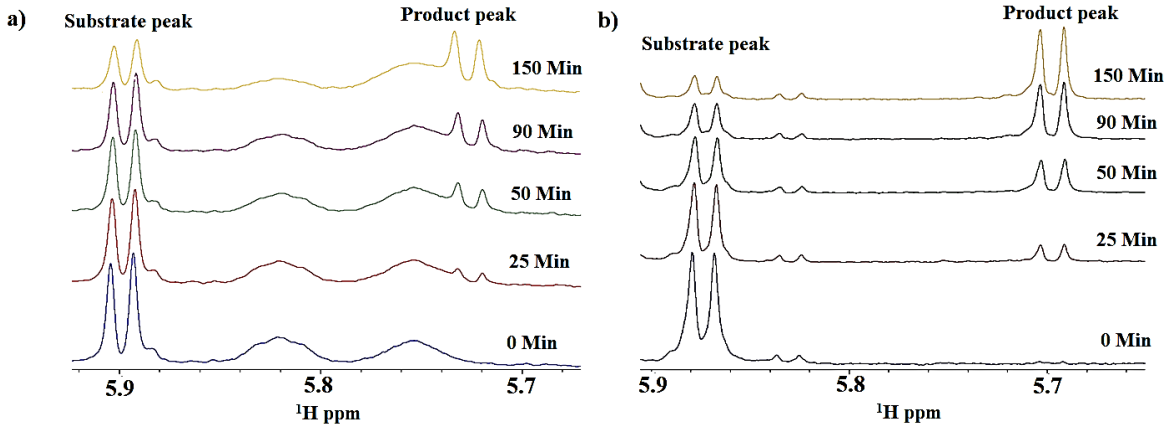


Figure 3.31. Comparison of ^{1}H- NMR spectra of 5'-AT_3**dC**AT_3 vs 5'-T_4**dC**AT.

(a) ^{1}H-NMR spectra for the substrate (5'-AT_3dCAT_3) and its conversion to the product (5'-AT_3dUAT_3) over time in the presence of A3B$_{CTD}$-QM-ΔL3-AL1swap (50 nM) at 298 K in activity assay buffer (pH 5.5) containing 10 % deuterium oxide; 50 mM citrate-phosphate, 200 mM NaCl, 2 mM β-mercaptoethanol, 200 µM 4,4-dimethyl-4-silapentane-1-sulfonic acid (DSS). (b) ^{1}H-NMR spectra for the substrate (5'-T_4dCAT) and its conversion to the product (5'-dT_4dUAT) over time in the presence of A3B$_{CTD}$-QM-ΔL3-AL1swap (300 nM) at 298 K in activity assay buffer (pH 6.0) containing 50 mM sodium phosphate, 100 mM NaCl, 2.5 mM β-mercaptoethanol, 50 µM 3-(trimethylsilyl)-2,2,3,3-tetradeuteropropionic acid (TSP).

Data acquisitions were done on a 700-MHz Bruker NMR spectrometer equipped with a 1.7 mm cryoprobe at 298 K. A series of ^{1}H NMR spectra was recorded of the substrate at various concentrations from 200 to 800 µM with 300 nM of A3B$_{CTD}$-QM-ΔL3-AL1swap[37]

in buffer (pH 6.0) containing 50 mM sodium phosphate, 100 mM NaCl, 2.5 mM β-mercaptoethanol, 50 µM 3-(trimethylsilyl)-2,2,3,3-tetradeuteropropionic acid (TSP) and 20% D_2O. The H-5 proton doublet signal of the cytosine, which appears at 5.88 ppm (J = 7.7 Hz), was baselined and integrated (**Figure 3.31b**). The signal of TSP at 0 ppm was used as an internal standard to determine the concentration of the substrate and its conversion during the reaction to the product, for which the signal appears at 5.71 ppm (J = 8.3 Hz).

The area of the integrated signal was converted to substrate concentration and plotted versus reaction time. Linear regression was used to fit the data to determine the initial speed of the reaction. **Figure 3.32** shows the relationship between rate of the deamination reaction and substrate concentration for $A3B_{CTD}$-QM-ΔL3-AL1swap. The double-reciprocal plot which shows the linear dependence of substrate concentration on the rate of deamination **(Figure 3.33)** was then fitted with linear regression to determine K_m and k_{cat} for $A3B_{CTD}$-QM-ΔL3-AL1swap using the following formula with enzyme concentration [E] = 300 nM:

$$y = V_0 = k_{cat}\,[\mathrm{E}]\frac{[\mathrm{S}]}{K_m + [\mathrm{S}]}$$

$$\frac{1}{V_0} = \frac{K_m + [\mathrm{S}]}{k_{cat}[\mathrm{E}][\mathrm{S}]}$$

$$\frac{1}{V_0} = \frac{K_m}{k_{cat}[\mathrm{E}]}\frac{1}{[\mathrm{S}]} + \frac{1}{k_{cat}[\mathrm{E}]}$$

$$y = ax + b$$

where a and b are, respectively, the slope and intercept obtained from the plot shown in **Figure 3.33**.

$$b = \frac{1}{k_{cat}[\mathrm{E}]} = 17.95\ \mathrm{s\ \mu M^{-1}}$$

$$k_{cat} = 0.19 \pm 0.03\ \mathrm{s^{-1}}$$

$$a = \frac{K_m}{k_{cat}[\mathrm{E}]} = 6335.4\ \mathrm{s}$$

$$K_m = {}^{a}/_{b} = 352 \pm 65\ \mu\mathrm{M}$$

Uncertainties in a and b were calculated from regression fit (**Figure 3.33**) using LINEST function of Excel and uncertainties in K_m and k_{cat} were calculated by standard error propagation, as detailed by us.[54]

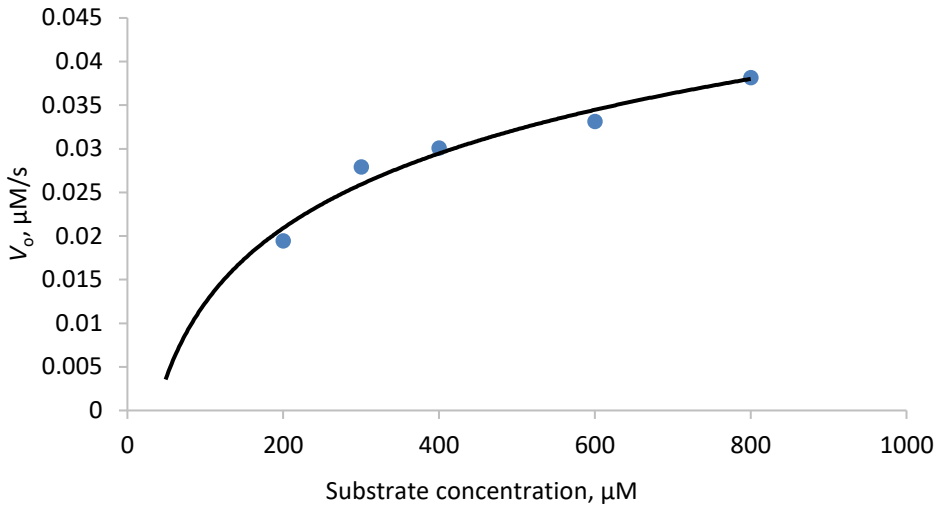


Figure 3.32. Initial rate of deamination 5'-T_4**dC**AT catalyzed by $A3B_{CTD}$-QM-ΔL3-AL1swap (300 nM) as a function of substrate concentration.

The curve is for visual reference only.

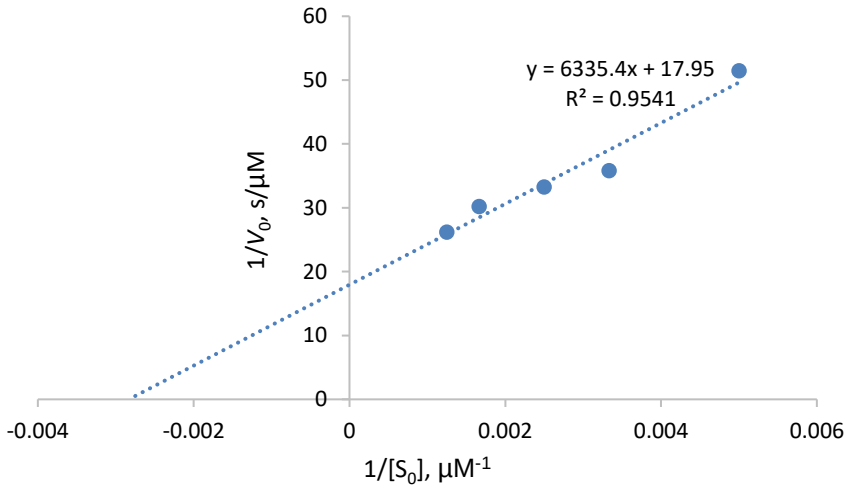


Figure 3.33. Double reciprocal plot showing the linear dependence of substrate concentration on the rate of deamination.

The plot shows inversed rate of deamination catalyzed by $A3B_{CTD}$-QM-ΔL3-AL1swap (300 nM) as a function of the inversed substrate concentration (5'-T_4dCAT), showing the linear dependence of rate of deamination with the concentration of the substrate.

The non-linear regression analysis of the data is presented in **Figure 3.34** and was performed by XLSTAT add-on in Excel (Microsoft) using the Michaelis-Menten enzyme kinetic equation providing the following parameters of the system:

Table 3.11. Model parameters of $A3B_{CTD}$-QM-ΔL3-AL1swap catalyzed deamination of 5'-T_4dCAT obtained using non-linear regression analysis.

Parameters	Value	Standard error
V_{max}, μM/s	0.051	0.005
K_m, μM	292	70

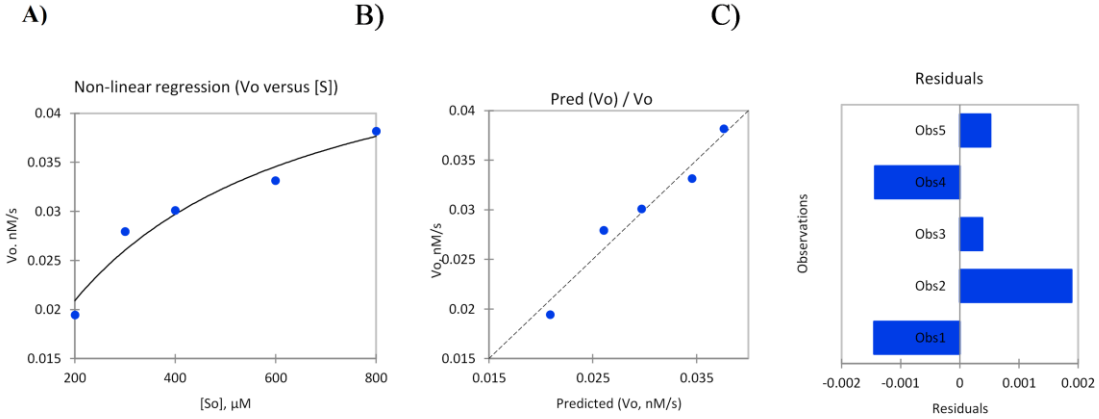


Figure 3.34. Non-linear regression analysis of $A3B_{CTD}$-QM-ΔL3-AL1swap-catalyzed cytosine deamination of 5'-T_4**dC**AT.

A) Plot of initial rate of deamination V_0 as a function of substrate concentration [S_0] fitted with a non-linear regression model to derive K_m and V_{max} reported in **Table 3.11**; B) observed versus predicted values of initial rate of deamination; C) Residuals calculated from the model for each observation, showing random distribution of residuals.

3.5.10. Evaluation of substrate activity of cross-linked oligonucleotides using ^{1}H NMR assay

Substrate preference and deamination activity by $A3B_{CTD}$-QM-ΔL3-AL1swap on the synthesized cross-linked oligonucleotides with **dC** were evaluated using the NMR-based assay as described above using substrates and $A3B_{CTD}$-QM-ΔL3-AL1swap at concentrations shown in **Figure 3.27** in a buffer (pH 6.0) containing 10 % deuterium oxide, 50 mM citrate-phosphate, 200 mM NaCl, 2 mM β-mercaptoethanol, 200 μM 4,4-dimethyl-4-silapentane-1-sulfonic acid (DSS). The H-5 proton doublet signals of the cytosine from different oligonucleotides were baselined and integrated. A doublet of doublets at 2.63 ppm originating from the citrate buffer or peaks of TSP or DSS were used as an internal standard to determine the concentration of the substrate which appears at 5.88 ppm (J = 7.7 Hz) and converted to 5.71 ppm (J = 8.3 Hz) during the reaction over a period of time (**Figure 3.31b**). The integrated signal area was converted to substrate concentration and plotted versus time of the reaction. The data were then fitted with linear regression to determine the initial speed of the reaction.

3.5.11. Qualitative evaluation of inhibitors of A3B$_{CTD}$-catalyzed deamination using the ^{1}H NMR assay

Competitive inhibition of the synthesized oligonucleotides was performed by a similar procedure reported by us previously.[54, 124, 126] A series of ^{1}H NMR spectra was recorded in a similar fashion as that of substrate analysis. Here, we used 400 μM of a standard 7-mer oligonucleotide substrate 5'-T$_4$**dC**AT, 8 or 4 μM of dZ or FdZ-containing cross-linked oligos, 300 or 200 nM of A3B$_{CTD}$-QM-ΔL3-AL1swap in a buffer containing 50 mM sodium phosphate (pH 6.0), 100 mM NaCl, 2.5 mM β-mercaptoethanol, 50 μM TSP and 20% D_2O. Deamination activity was monitored by the same method as for DNA substrates.

3.5.12. Quantitative evaluation of inhibitors of A3B$_{CTD}$-catalyzed deamination using the ^{1}H NMR assay

The best inhibitors were characterized further by varying the concentration of the inhibitor, ranging from 0.25 μM to 8 μM in the presence of 300 nM of A3B$_{CTD}$-QM-ΔL3-AL1swap in activity assay buffer. After the rate of the reaction was determined at various inhibitor concentrations, the data were analyzed by various methods outlined below.

3.5.12.1. Analysis of inhibitors of A3B$_{CTD}$-catalyzed deamination using Dixon plot

In the first instance, the data were analyzed using a plot of inverse speed versus inhibitor concentration, the so-called Dixon plot (**Figure 3.35**), which was then fitted with linear regression to derive the inhibition constant (K_i).

Here, we assumed the competitive character of inhibition and, therefore, Michaelis-Menten expression for deamination velocity v becomes:

$$v = \frac{v_{\max}[\mathrm{S}]}{K_\mathrm{m}\left[1 + \frac{[\mathrm{I}]}{K_\mathrm{i}}\right] + [\mathrm{S}]}$$

$$\frac{1}{v} = \left[\frac{K_\mathrm{m}}{K_\mathrm{i} v_{\max}[\mathrm{S}]}\right][\mathrm{I}] + \frac{K_\mathrm{m} + [\mathrm{S}]}{v_{\max}[\mathrm{S}]}$$

By using K_m and V_{max} derived from the Lineweaver-Burk plot (**Figure 3.33**) and using the trend line for dZ[U^{E}(-2),A^{N3}(+1)]X:

$$y = ax + b$$

$$b = 35.692\ \mu\mathrm{M}^{-1}, a = 24.061\ \mathrm{s}\ \mu\mathrm{M}^{-1}$$

$$K_\mathrm{m} = 353\ \mu\mathrm{M}, [\mathrm{S}] = 400\ \mu\mathrm{M}$$

$$K_\mathrm{i} = \frac{K_\mathrm{m} b k_{\mathrm{cat}}[\mathrm{S}]}{a k_{\mathrm{cat}}[\mathrm{S}](K_\mathrm{m} + [\mathrm{S}])} = \frac{b K_\mathrm{m}}{a(K_\mathrm{m} + [\mathrm{S}])} = 0.66 \pm 0.14\ \mu\mathrm{M}$$

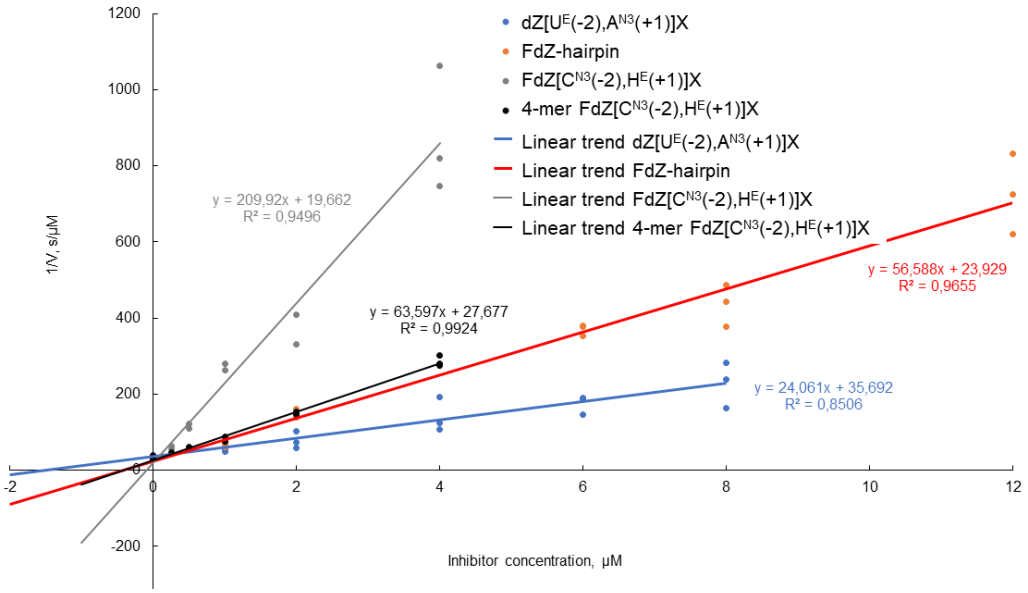


Figure 3.35. Analysis of inhibitors of $A3B_{CTD}$-catalyzed deamination using Dixon plot.

The graph shows inversed rates of $A3B_{CTD}$-QM-ΔL3-AL1swap-catalyzed deamination of 5'-T_4dCAT (400 μM) at various concentrations of cross-linked oligos at 298 K. Enzyme concentration is 300 nM.

Uncertainties in *a* and b were calculated from regression fit (**Figure 3.35**) using LINEST function of Excel; uncertainty of K_i was calculated by standard error propagation, as detailed by us.[54]

3.5.12.2. Analysis of inhibitors of $A3B_{CTD}$-catalyzed deamination using non-linear regression

The second method used the non-linear regression analysis using XLSTAT add-on in Excel (Microsoft) and performing a three-parameter fit (V_{max}, K_m and K_i) in the Michaelis-Menten enzyme kinetic equation for the competitive inhibitor. Statistical parameters and details of non-linear regression analysis are provided below.

Table 3.12. Correlation matrices for non-linear regression analysis of $A3B_{CTD}$-QM-ΔL3-AL1swap-catalyzed deamination of 5'-AT_3dCAT_3 in the presence of varying concentrations of linear inhibitors.

dZ-linear:

Variable	[S]	[I]	Rate
[S]	**1.000**	0.181	0.547
[I]	0.181	**1.000**	-0.604
Rate	0.547	-0.604	**1.000**

FdZ-linear:

Variable	[S]	[I]	Rate
[S]	**1.000**	0.196	0.390
[I]	0.196	**1.000**	-0.693
Rate	0.390	-0.693	**1.000**

Table 3.13. Correlation matrices for non-linear regression analysis of $A3B_{CTD}$-QM-ΔL3-AL1swap-catalyzed deamination of 5'-T_4dCAT in the presence of varying concentrations of cross-linked inhibitors.

dZ[U^{E}(-2),A^{N3}(+1)]X

Variable	[S]	[I]	Rate
[S]	**1.000**	-0.120	0.408
[I]	-0.120	**1.000**	-0.856
Rate	0.408	-0.856	**1.000**

FdZ-hairpin

Variable	[S]	[I]	Rate
[S]	**1.000**	-0.112	0.379
[I]	-0.112	**1.000**	-0.797
Rate	0.379	-0.797	**1.000**

FdZ[C^{N3}(-2),H^{E}(+1)]X

Variable	[S]	[I]	Rate
[S]	**1.000**	-0.100	0.378
[I]	-0.100	**1.000**	-0.759
Rate	0.378	-0.759	**1.000**

4-mer FdZ[C^{N3}(-2),H^{E}(+1)]X

Variable	[S]	[I]	Rate
[S]	**1.000**	-0.100	0.400
[I]	-0.100	**1.000**	-0.853
Rate	0.400	-0.853	**1.000**

Table 3.14. Statistics of fit for non-linear regression analysis $A3B_{CTD}$-QM-ΔL3-AL1swap-catalyzed deamination of 5'-AT_3dCAT_3 in the presence of varying concentrations of linear inhibitors.

Statistic	dZ-linear	FdZ-linear
Observations	12	15
DF	9.000	12.000
R^2	0.975	0.975
SSE	0.000	0.000
MSE	0.000	0.000
RMSE	0.001	0.001
AIC	-177.018	-222.249
AICC	-171.304	-218.249
Iterations	8.000	8.000

Table 3.15. Statistics of fit for non-linear regression analysis $A3B_{CTD}$-QM-ΔL3-AL1swap-catalyzed deamination of 5'-T_4dCAT in the presence of varying concentrations of inhibitors.

Statistic	dZ[U^{E}(2),A^{N3}(+1)]X	FdZ-hairpin	FdZ[C^{N3}(2),H^{E}(+1)]X	4-mer FdZ[C^{N3}(-2),H^{E}(+1)]X
Observations	23	23	19	19
DF	20	20	16	16
R^2	0.964	0.986	0.989	0.979
SSE	0.000	0.000	0.000	0.000
MSE	0.000	0.000	0.000	0.000
RMSE	0.002	0.002	0.001	0.002

AIC	-277.790	-292.151	-242.589	-236.944
AICC	-275.567	-289.929	-239.732	-234.087
Iterations	5.000	5.000	8.000	4.000

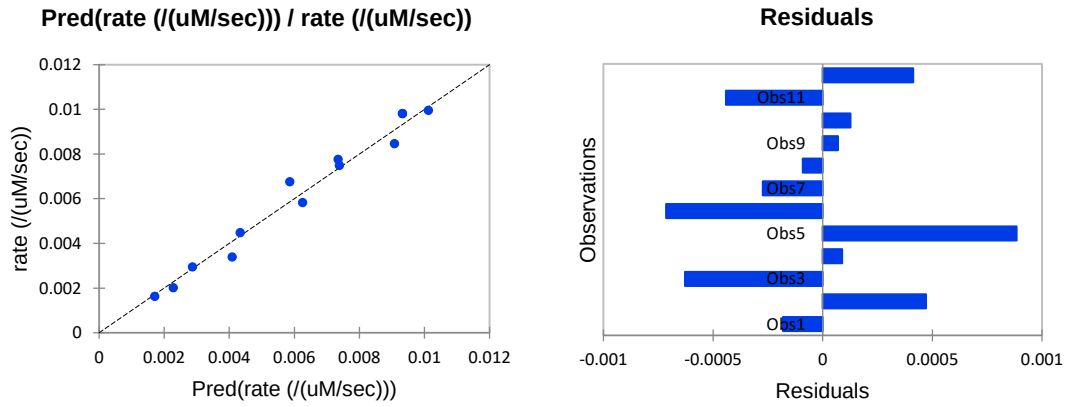
Table 3.16. Parameters of $A3B_{CTD}$-QM-ΔL3-AL1swap-catalyzed deamination of 5'-AT_3dCAT_3 in the presence of varying concentrations of linear inhibitors using non-linear regression analysis.

Parameters	**dZ-linear**		**FdZ-linear**	
	Value	Standard error	Value	Standard error
V_{max}, μM/s	0.0126	0.0008	0.0126	0.0008
K_m, μM	146.	28	147.	28.
K_i, μM	7.0	1.3	4.4	0.7

Table 3.17. Parameters of $A3B_{CTD}$-QM-ΔL3-AL1swap-catalyzed deamination of 5'-T_4dCAT in the presence of varying concentrations of cross-linked inhibitors using non-linear regression analysis.

Parameters	**dZ[U^E(2),A^{N3}(+1)]X**		**FdZ-hairpin**		**FdZ[C^{N3}(2),H^E(+1)]X**		**4-mer FdZ[C^{N3}(-2),H^E(+1)]X**	
	Value	Standard error	Value	Standard error	Value	Standard error	Value	Standard error
V_{max}, μM/s	0.052	0.006	0.052	0.005	0.051	0.004	0.051	0.005
K_m, μM	320	90	310	70	290	60	290	70
K_i, μM	0.69	0.14	0.35	0.06	0.100	0.016	0.28	0.05

dZ-linear:



FdZ-linear:

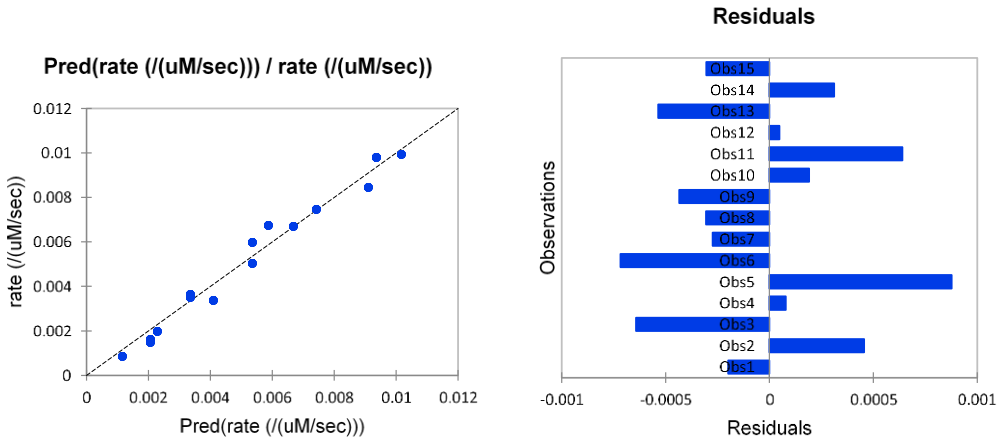
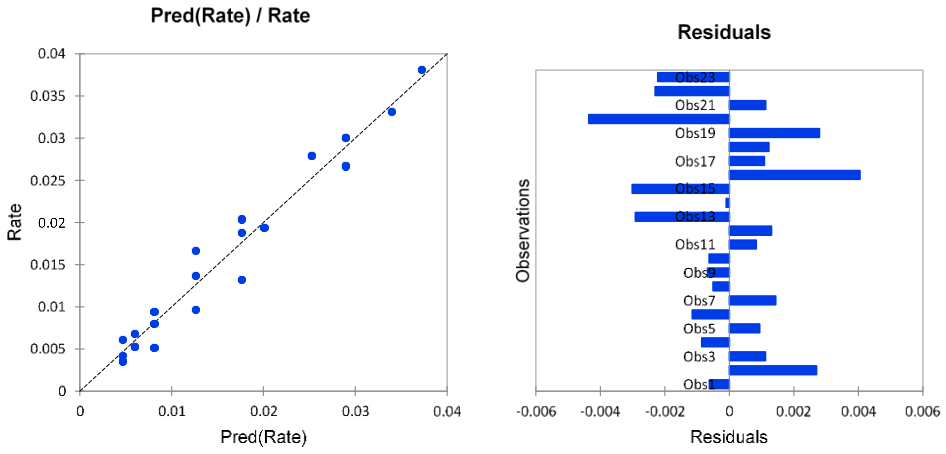


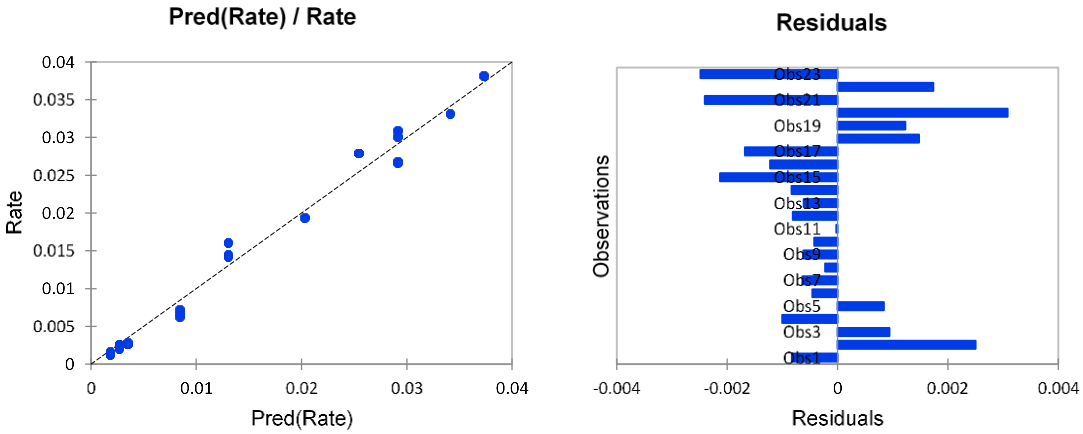
Figure 3.36. Non-linear regression analysis of $A3B_{CTD}$-QM-ΔL3-AL1swap-catalyzed cytosine deamination of AT_3**dC**AT_3 in the presence of varying concentrations of linear inhibitors.

Left column: observed vs predicted values of initial rate of deamination; Right column: residuals calculated from the model for each observation.

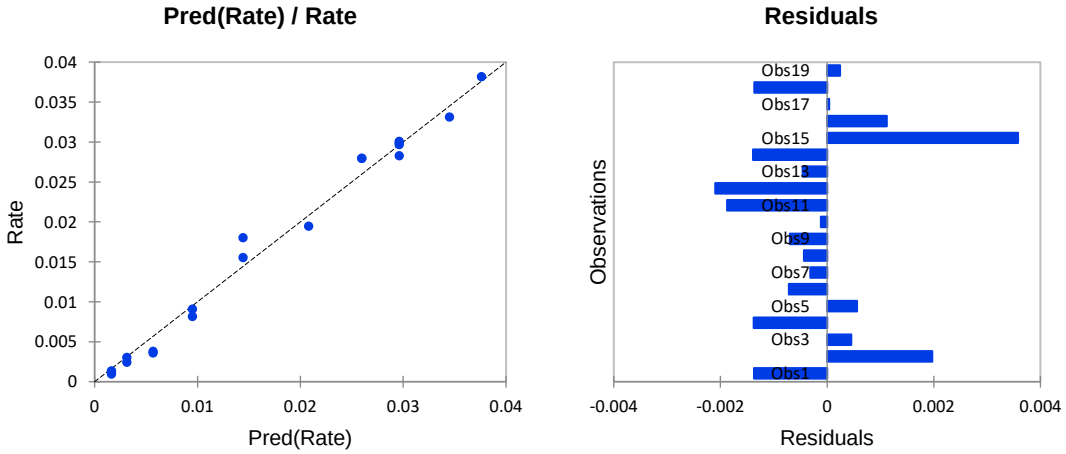
dZ[U^E(-2),A^{N3}(+1)]X:



FdZ-hairpin:



FdZ[C^{N3}(-2),H^{E}(+1)]X:



4-mer FdZ[C^{N3}(-2),H^{E}(+1)]X:

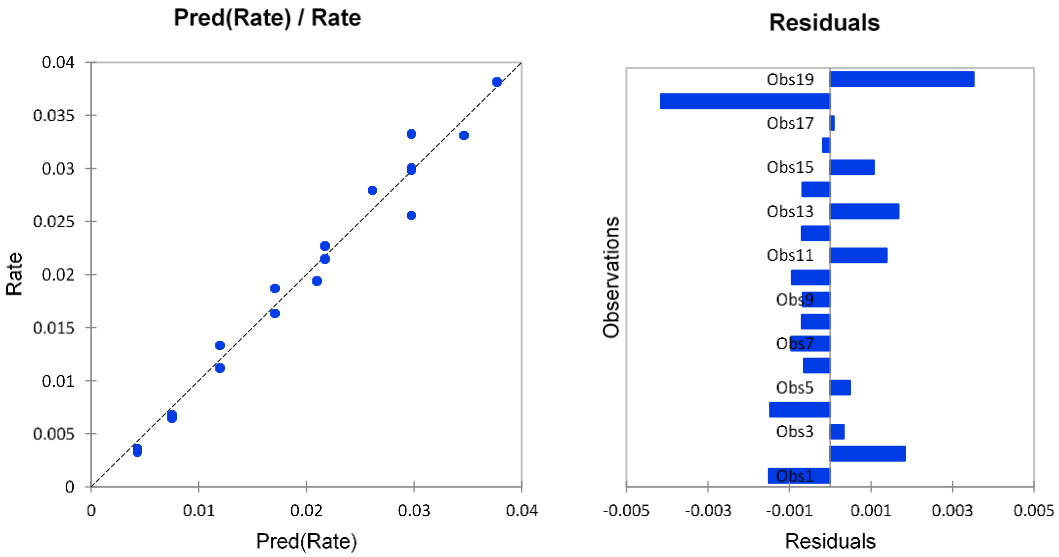


Figure 3.37. Non-linear regression analysis of $A3B_{CTD}$-QM-ΔL3-AL1swap-catalyzed cytosine deamination of 5'-T_4**dC**AT in the presence of varying concentrations of cross-linked inhibitors.

Left column: observed vs predicted values of initial rate of deamination; Right column: residuals calculated from the model for each observation.

3.5.12.3. Evaluation of FdZ[C^{N3}(-2),H^{E}(+1)]X as inhibitor of $A3B_{CTD}$-catalyzed deamination of dC-hairpin using Lambert's W function

Time-resolved ^{1}H NMR kinetics were measured in 50 mM K^+/Na^+ phosphate buffer (pH 6.0) supplemented with 100 mM NaF, 1 mM TCEP, 100 µM sodium trimethylsilylpropanesulfonate (DSS) and 10% D_2O. Substrate (dC-hairpin) was used at several hundred µM concentrations as plotted in **Figure 3.29** and the reaction was performed in the presence of 20 and 27 nM of $A3B_{CTD}$-QM-ΔL3-AL1swap at 298 K. The course of the reaction was followed by ^{1}H NMR until the substrate was consumed. Subsequently the amount of substrate at each time point was calculated by integrating the decreasing substrate peak at 7.66 ppm (singlet) and calibrated by the area of DSS standard peak at 0.0 ppm. Using the known concentration of the standard, the peak was converted to a corresponding substrate concentration. The time at which each spectrum was recorded, as a difference to the first spectrum, was used as the time passed. The substrate concentration *versus* the time of reaction was plotted and fitted using the integrated form of the Michaelis-Menten equation:

$$[S]_t = K_m\, W\left(\frac{[S]_0}{K_m} \cdot e^{\frac{[S]_0 - V_{max}\, t}{K_m}}\right)$$

where W is Lambert's W function, $[S]_t$ is the substrate concentration at specific time, $[S]_0$ is the initial substrate concentration, V_{max} and K_m are the Michaelis-Menten constants and t is the time. The two Michaelis Menten constants, the initial substrate concentration and an offset which corrects for the integration baseline in the NMR spectra were fitted using Lambert's W function in Gnuplot.

3.5.13. Isothermal titration calorimetry

We tested the thermodynamics of binding of our cross-linked oligos with A3 enzymes in comparison to the linear oligos using A3A-E72A and the catalytically competent $A3B_{CTD}$-QM-ΔL3-AL1swap. Desalted unmodified DNA oligonucleotides were purchased (Integrated DNA Technologies) at 1 or 5 µmol synthesis scale and dissolved in one of the buffers described below to give 10 mM solutions. ITC experiments were conducted at 25 °C using a GE MicroCal ITC200 (now Malvern Instruments) isothermal titration calorimeter. Protein A3A-E72A, which is an inactive protein, was diluted in ITC buffer to concentrations of 10-100 µM and titrated with dC oligonucleotides [5'-AT_3**dC**AT_3, 5'-T_4**dC**AT, cross-linked oligos] in ITC buffer. The ratio of protein to oligonucleotide concentration is usually

1:10. Titrations are made up of 30 individual additions. To prevent protein precipitation during the long timescale of the experiment it was necessary to used improved ITC buffers. ITC buffer 1: 50 mM MES, pH 6.0, 100 mM NaCl, 200 μM EDTA, 1 mM β-mercaptoethanol and ITC buffer 2: 50 mM Na^+/K^+ phosphate, pH 6.0, 50 mM NaCl, 50 mM choline acetate, 2.5 mM TCEP, 200 μM EDTA with 30 mg/mL BSA. After preparation, ITC buffer 2 was frozen and defrosted before the experiments. ITC buffer 2* is freshly prepared buffer 2.

Protein $A3B_{CTD}$-QM-ΔL3-AL1swap, which is an active protein, was used for establishing the binding affinity of dZ-containing oligos dZ-linear and dZ[U^E(-2),A^{N3}(+1)]X]. For $A3B_{CTD}$-QM-ΔL3-AL1swap we used ITC buffer 3: 50 mM Na^+/K^+ phosphate, pH 6.0, 200 mM trimethylamine *N*-oxide dihydrate, 2.5 mM TCEP, 200 μM EDTA. Data evaluation was performed with the software provided by the supplier of MicroCal ITC200.

Chapter 4. Synthesis and Evaluation of Seven-Membered Ring-Containing Cytidine Deaminase Inhibitors and Incorporation into ssDNA as Inhibitors of APOBEC3 Enzymes

4.1. Introduction

Cytidine deaminase is the key enzyme which catalyzes the deamination of cytidine or 2'-deoxycytidine to uridine or 2'-deoxyuridine.[195-196] Like APOBEC, CDA also is a zinc^{2+}-dependent deaminase, and it is assumed that all A3 enzymes deaminate the target cytidine by a similar mechanism to CDA because the active site is similar in all of them to CDA. The major difference between them is that CDA accepts only individual nucleosides as substrates, and APOBEC acts only on ssDNA having at least four nucleotides, one of which is cytosine. Spontaneous deamination of cytidine is a very slow process, the accelerated rate of deamination by the enzyme is through the formation of the hydrated species.[197]

We recently demonstrated that, upon incorporation into a specific ssDNA (5'-TCA) sequence, 2'-deoxyzebularine (dZ) and 5-fluoro-2'-deoxyzebularine (5-FdZ) inhibit A3 enzymes.[54, 124, 126] We also demonstrated that dZ as a free nucleoside does not inhibit A3 enzymes, which proves that ssDNA delivers the transition-state analogue into the active site of A3.[54] The above studies prompted us to investigate the synthesis of a series of 2'-deoxy analogues of CDA inhibitors (**Figure 4.2**).[121-123] We proposed to improve the inhibitory potential of our previously reported ssDNA sequences by incorporating these powerful transition state analogues of cytosine deamination. Out of all the reported CDA inhibitors, **ddiazep** is quite different structurally as it has no functionality at C5 to contribute to coordination with Zn^{2+} or to undergo hydration. The ribose form of **ddiazep** has been co-crystallised with human CDA (hCDA), which sheds light on important interactions this inhibitor makes (**Figure 4.1**).[185]

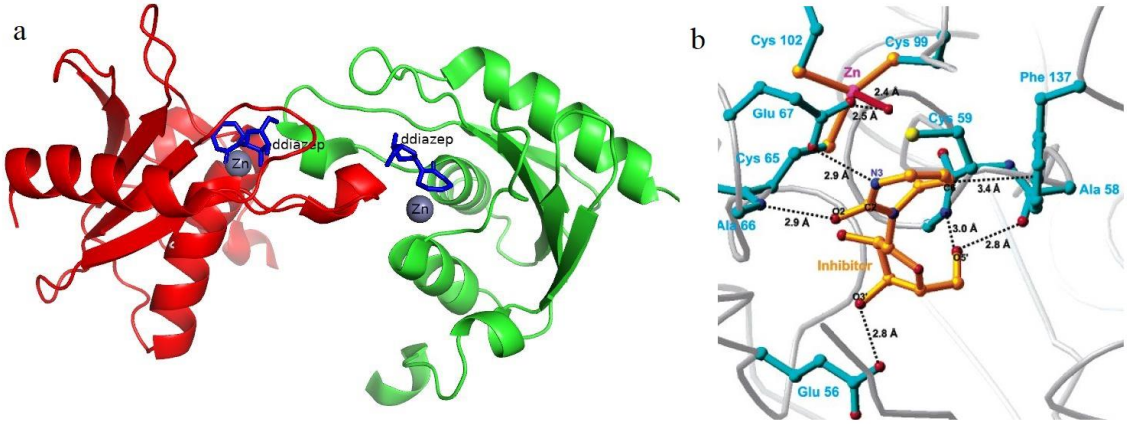


Figure 4.1. Crystal structure of a ribose form of ddiazep that has been co-crystallised with human CDA (hCDA).

a) Ribbon diagram of hCDA and ddiazep co-crystal, where ddiazep is depicted as blue sticks. b) Important interactions ddiazep (orange sticks) makes with the amino acid residues (light blue sticks) in hCDA (PDB ID: 1MQ0). The structures are rendered in PyMol (2.5).

The seven-membered nucleobase is puckered and therefore the C5-C6 part of the molecule is involved in a special π-π interaction with Phe137 residue of hCDA. Interestingly, in A3Bctd, instead of Phe137 in CDA, there is Tyr313 present in the same position making the π-π interaction possible with **ddiazep**. These observations make this compound interesting to be tested as an A3 inhibitor when incorporated into an ssDNA. The ***R*** and ***S*** isomers of a 7-membered ring containing ribose were also reported previously and one of the isomers was shown to be a slow-binding inhibitor, as potent as tetrahydrouridine (THU).[123] Unfortunately, the authors were unable to determine which isomer was more potent. The lack of proper stereochemical information prompted us to synthesize both isomers, isolate and characterise them with relative stereochemistry before incorporation into ssDNA as A3 inhibitors. All the new compounds synthesized were characterised for CDA inhibition using a UV-based assay and the DNA-based inhibitors were tested for their A3 inhibition using our previously described real-time NMR assay.[54, 124, 126]

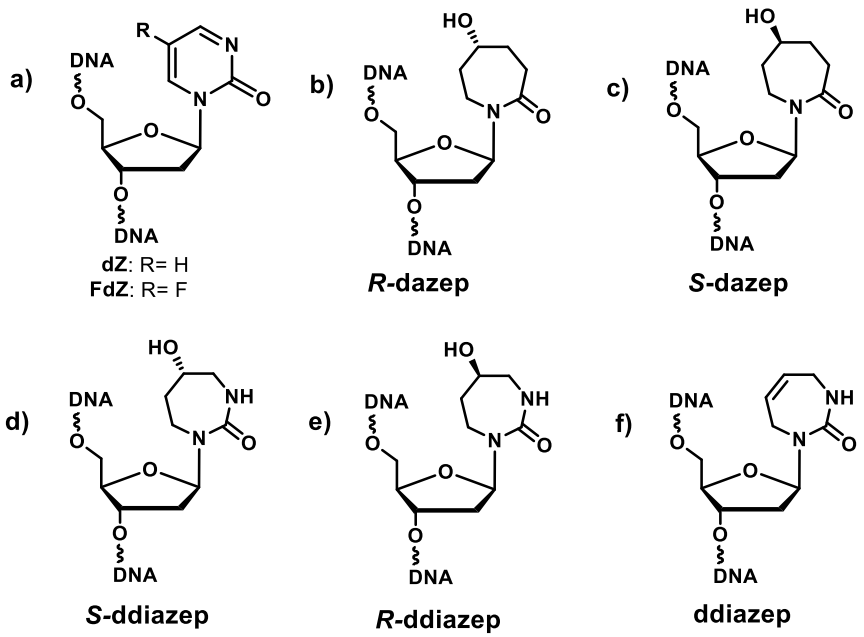


Figure 4.2. 2'- Deoxy analogues of reported and new CDA inhibitors designed to be incorporated into ssDNA in this work.

4.2. Results and discussion

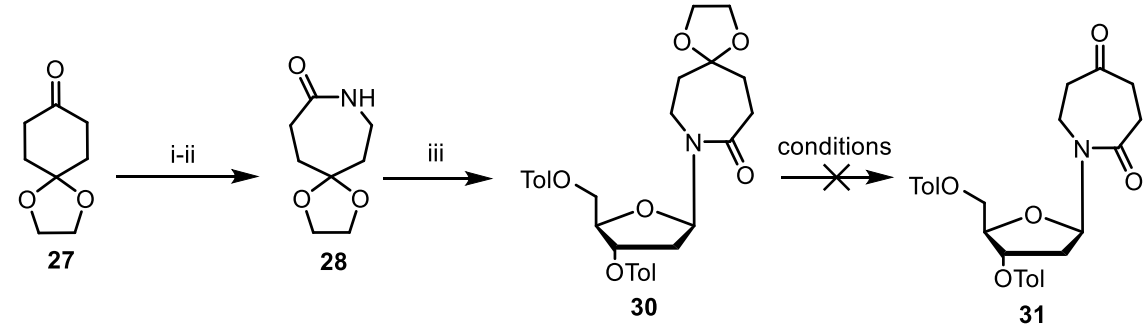
The difference in inhibitory potential of dZ versus FdZ prompted us to investigate the possibility of the synthesis and incorporation of more powerful CDA inhibitors as transition-state analogues. The methodology of synthesis and successful incorporation of the

synthesized compounds into previously described ssDNA sequences[54, 124, 126] along with their evaluation as A3 inhibitors are outlined below.

4.2.1. Synthesis of modified nucleosides, their DMT-protected phosphoramidites and corresponding oligos

4.2.1.1. Synthesis of modified *R*-and *S*-dazep phosphoramidites

The synthesis was performed using the commercially available **27** as starting material. Synthesis of oxime **28** followed by Beckmann rearrangement provided the required protected nucleobase as 7-membered cyclic amide (**28)**. This compound was coupled with Hoffer's chlorosugar[125] (**29**) in the presence of K-^{t}BuO in 1,4-dioxane at room temperature to provide the required protected 7-membered ring containing nucleoside (**30**) with a β:α ratio of 99:1 as determined by NOESY NMR .



Scheme 4.1. Attempted synthesis of ***R***-and ***S*-dazep** phosphoramidites.

Reagents and conditions: i) $NH_2OH{\cdot}HCl$, NaOH, H_2O, MeOH, r.t., 10 min. ii) Tosyl chloride, 4N NaOH, acetone/ H_2O r.t., 2 hr, 4N HCl, r.t., overnight; iii) Hoffer's chlorosugar, K-^{t}BuO, 1,4-dioxane, 60 ºC, 2 hr.

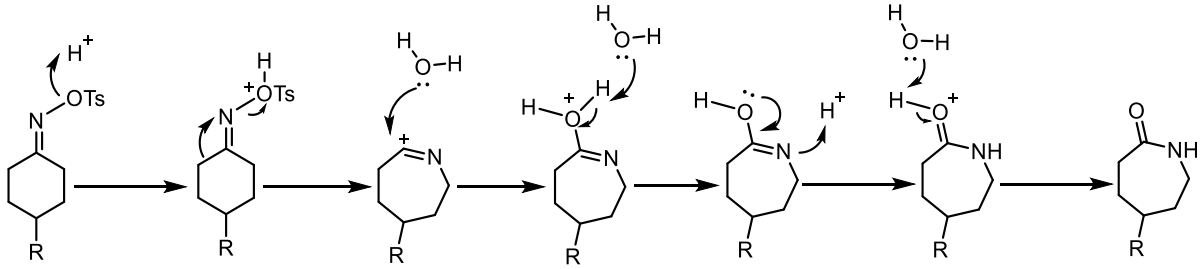


Figure 4.3. Mechanism of Beckmann rearrangement.

Protonation of the tosylated oxime followed by an alkyl migration and the expulsion of the tosyl leaving group produces a nitrilium ion. The water attack on this species followed by acid-catalyzed tautomerism produces the seven-membered cyclic amide (**Figure 4.3**).

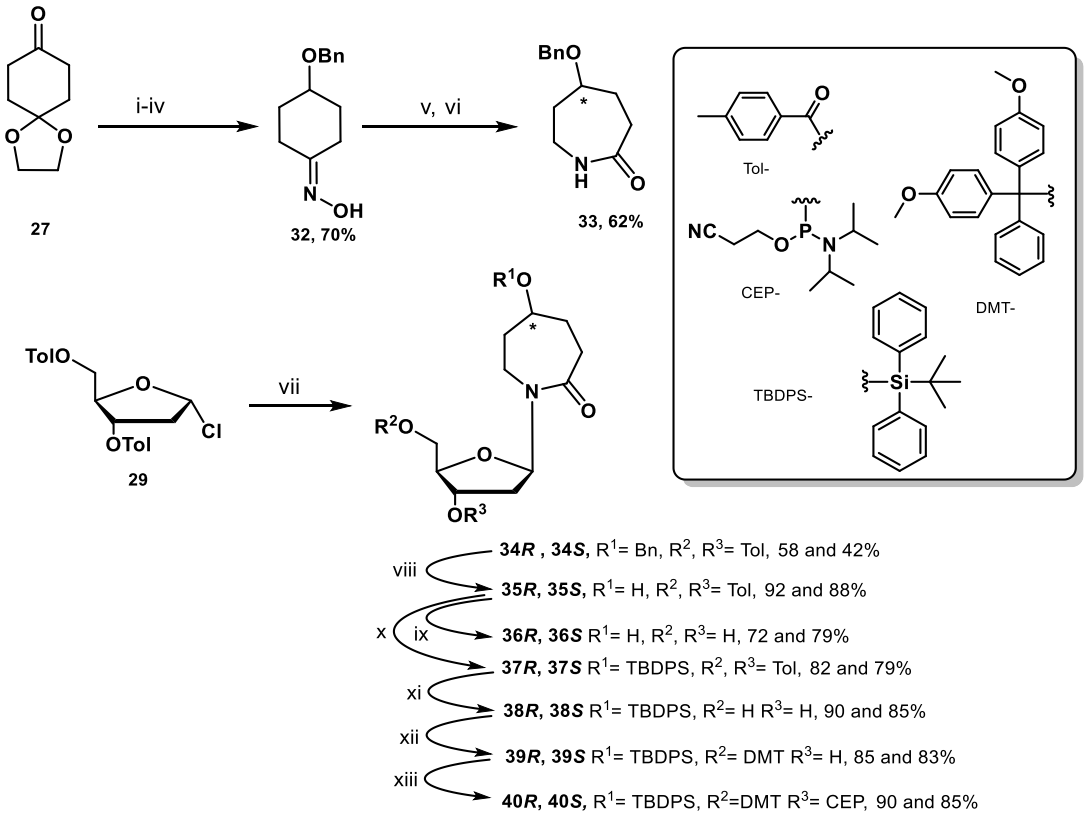
We tried several conditions to deprotect the ketal protection in the compound **30** (**Table 4.1**) but all the reactions failed to produce the required product **31** and the starting material was isolated every time. So, we decided to change the synthetic route (**Scheme 4.2**) where we synthesized a benzyl-protected cyclic amide derivative (**33**).

Table 4.1. Deprotection conditions used to deprotect ketal in the compound **30**.

No.	Conditions
1.	*p*-TsOH (10%), acetone, r.t.
2.	1 N HCl, overnight, r.t.
3.	BF_3.Etherate, *m*-CPBA, toluene, r.t.

A method reported recently described the synthesis of compound **33** from the ketal **32** followed by Beckmann rearrangement to give the cyclic amide (**33**) (**Scheme 4.2**).[198] The cyclic amide (**33**) and Hoffer's chlorosugar **29** were coupled in the presence of K-^{t}BuO in 1,4-dioxane at room temperature to provide a pair of diastereomers **34*R*** and **34*S*** in moderate yields (52 and 48%, respectively) and β:α ratio of 99:1. These diastereomers were purified using silica gel column chromatography. After purification, **34*R*** was crystallized from 30% EtOAc in hexane. The stereochemistry was assigned based on X-ray crystal data (**Figure 4.4.**).

Here onwards, these diastereomers were treated by the same sequence of reactions separately. Benzyl deprotection was performed by catalytic hydrogenation using 10% Pd/C, H_2 at room temperature to give **35*R*** and **35*S*** with 92 and 88%, respectively. One part of the benzyl-deprotected compounds was treated with 30% aq. NH_3 in MeOH at r.t, for 3 days to provide free nucleosides **36*R*** and **36*S*** (72 and 79%, respectively). The rest of the portion was treated with *tert*-butyl diphenylsilyl (TBDPS) chloride in the presence of imidazole as a base in DCM to give **37*R*** and **37*S*** with yields 82 and 79%, respectively. The toluoyl groups on **37*R*** and **37*S*** were cleaved by 30% aq. NH_3, MeOH at r.t. for 3 days to provide **38*R*** and **38*S*** (90 and 85%, respectively). After the toluoyl deprotection, the 5'-hydroxyl group was selectively protected by DMT chloride under standard DMT protection conditions to give **39*R*** and **39*S*** in 85 and 83% yields, respectively. Phosphoramidites **40*R*** and **40*S*** were prepared employing standard phosphitylation conditions with isolated yields of 90 and 85% respectively.



Scheme 4.2. Synthesis of **dazep** phosphoramidites.

Reagents and conditions: i) $NaBH_4$, MeOH, 0 °C→r.t., overnight; ii) benzyl chloride, Et_3N, DMF, 0 °C→r.t., overnight; iii) 1N HCl, THF, r.t., overnight; iv) $NH_2OH{\cdot}HCl$, NaOH, H_2O, MeOH, r.t., v) tosyl chloride, 4N NaOH, acetone/ H_2O r.t., 2 hr vi) 4N HCl, r.t., overnight; vii) Hoffer's chlorosugar, K-^{t}BuO, 1,4-dioxane, 60 ^{o}C, 2 hr; viii) 10% Pd/C, H_2, EtOH, r.t., 6 hr; ix and xi*)* 30% aq. NH_3, MeOH, rt, 3 days; x*)* TBDPS chloride, imidazole, DCM, 0 °C→r.t., overnight; xii*)* 4,4'-dimethoxytrityl chloride (DMTCl), pyridine, 0 °C→r.t., overnight; xii*)* *N*,*N*-di*iso*propylamino-2-cyanoethoxychlorophosphine, Et_3N, CH_2Cl_2, r.t., 20 min.

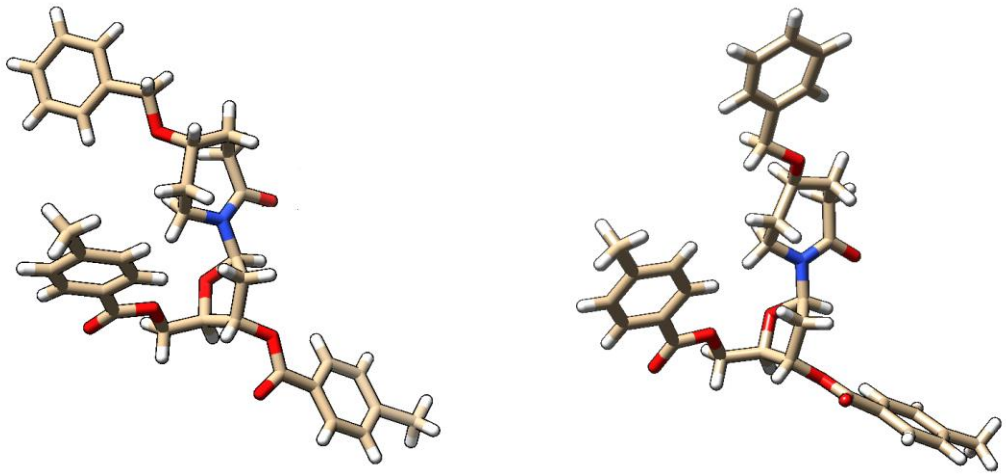


Figure 4.4. X-ray structure of **34*R*** (left) and ***34S*** (right). Ellipsoids are drawn at a 50% probability level.

4.2.1.2. Synthesis of ddiazep phosphoramidite

The synthesis of **ddiazep** started with the formation of diallyl urea (**42**) by the condensation of allylamine (**41**) in presence of *N*,*N*-disuccinimidyl carbonate in THF (**Scheme 4.3**). Coupling of **42** to Hoffer's chlorosugar was performed by previously described silyl modification of the Hilbert-Johnson reaction[199] using $SnCl_4$ as Lewis acid in dichloroethane at -35 °C to furnish **43** with moderate yield (45%) and β:α ratio of 9:1. The β isomer as a major product was confirmed by NOESY NMR experiment. The spectrum shows cross-peaks of the 1' and 4'-protons of the major isomer of compound **43** through space, confirming it as a β-nucleoside, while the minor α-anomer had a cross-peak between the 1' and 3'-protons (**Figure 4.5**).

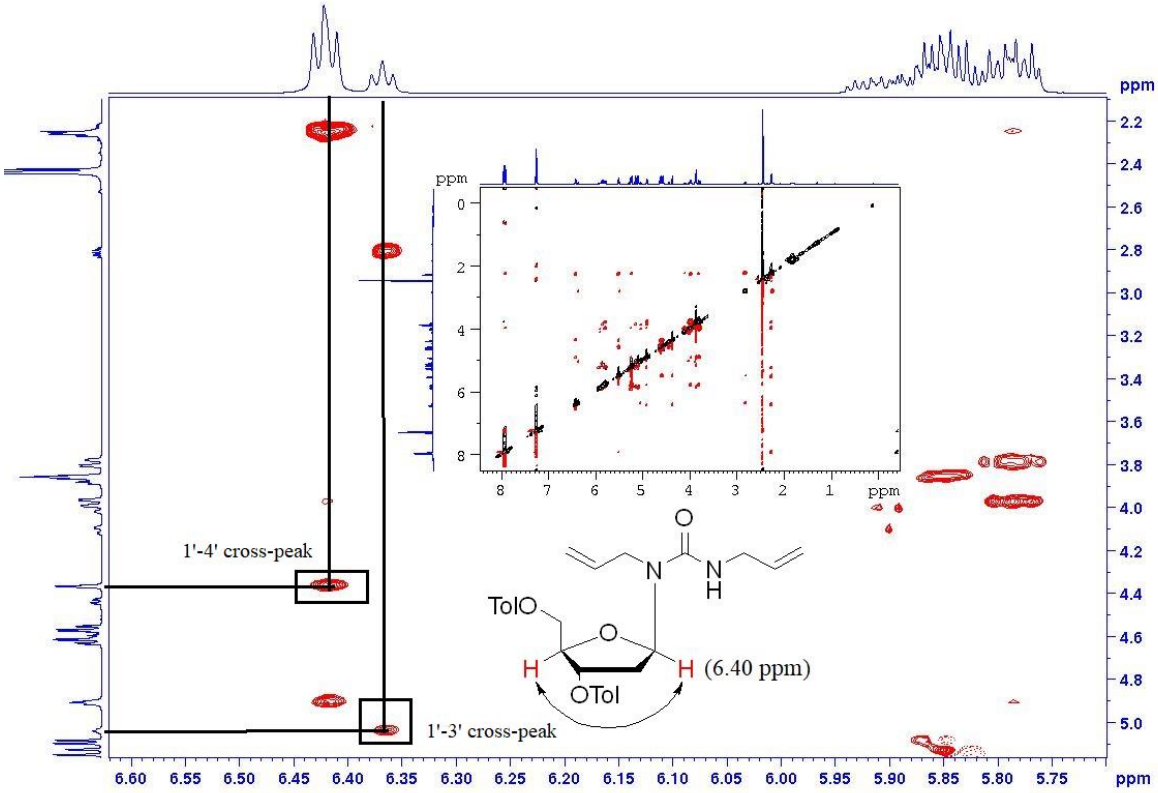


Figure 4.5. NOESY Spectrum of compound **43**.

The spectrum shows the cross-peak between 1'-and 4'-protons of the major isomer of compound **20** through space, confirming it as a β-nucleoside and the cross-peak between 1'-and 3'- protons in the minor α-anomer.

As reported in the literature, alkylation of the amide **43** was required to lock the allyl groups in *cis* orientation (**Figure 4.6**) for the successful ring-closure metathesis (RCM).[120]

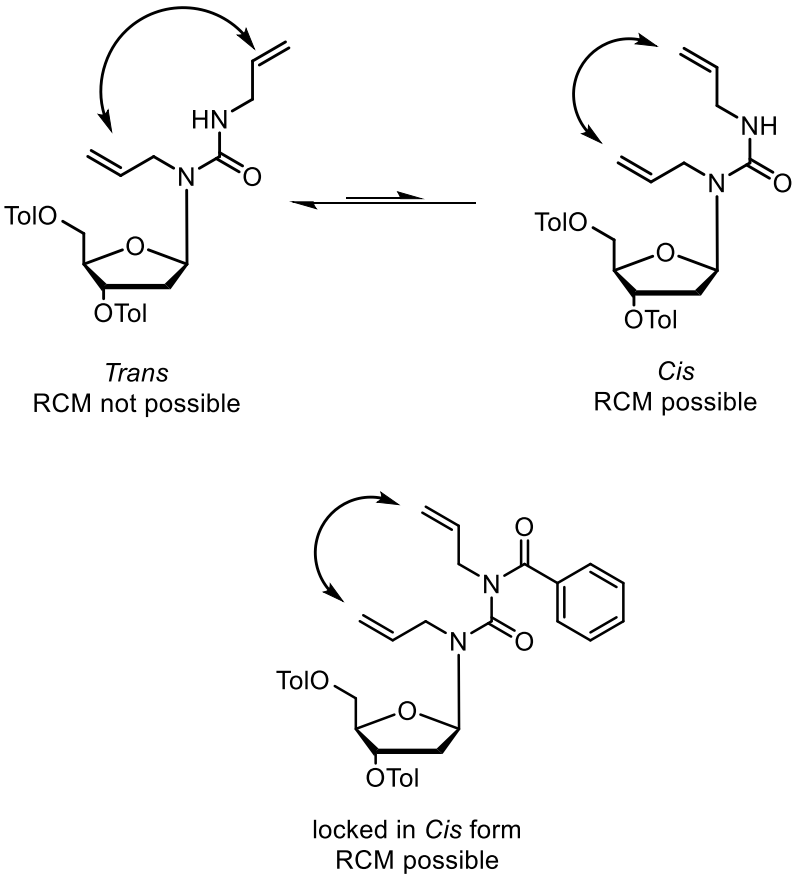
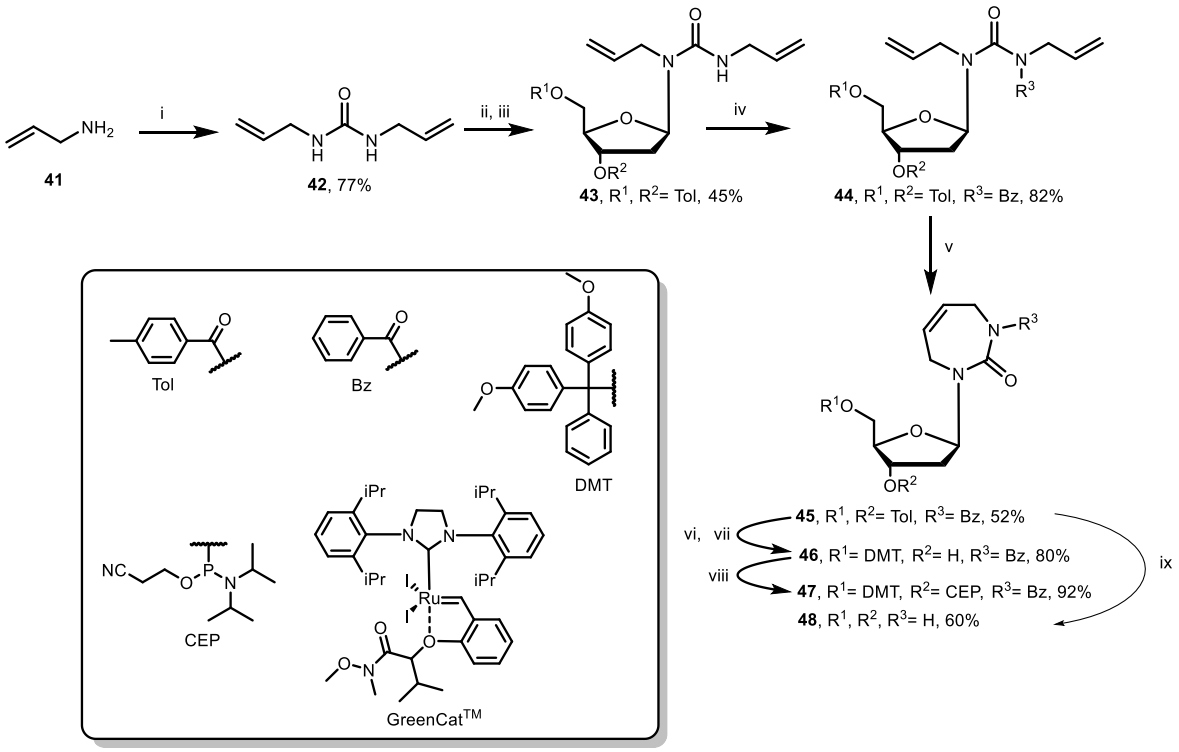


Figure 4.6. Equilibrium between the *cis*- and *trans*-conformations of the unprotected amide bond.

So, we protected the amide (**43**) with benzoyl chloride to provide **44** and performed RCM with GreenCat^{TM}, which provided the required seven-membered ring (**45**) in 52% isolated yield. Resuspending and washing the precipitate in methanol provided >99% pure β anomer.

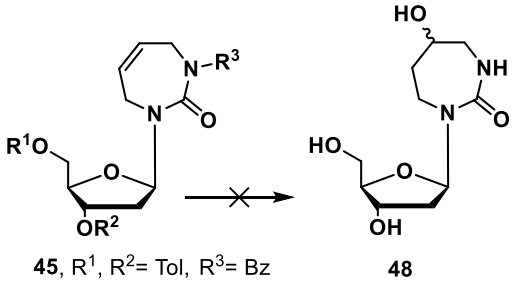
The toluoyl groups on **45** were selectively deprotected by 30% aq. NH_3 in MeOH at r.t. for 3 days followed by selective protection of 5'-hydroxyl group using standard DMT protection conditions to provide **46** (80% yield). Phosphoramidite **47** was prepared employing standard phosphitylation conditions in an isolated yield of 92%.



Scheme 4.3. Synthesis of 2'-deoxy diazepine phosphoramidite

Reagents and conditions: i) *N*,*N'*-Disuccinimidyl carbonate, THF, r.t., 2 hr; ii) TMS chloride, Et_3N, benzene, r.t., overnight; iii) Hoffer's chlorosugar, $SnCl_4$, 1, 2- dichloroethane, -35 °C 1 hr; iv) benzoyl chloride, pyridine, 0 °C→r.t., overnight; v) GreenCat™, dry DCM, reflux, 2hr; vi) 30% aq. NH_3, MeOH, rt, 3 days; vii) 4,4'-dimethoxytrityl chloride (DMTCl), pyridine, 0 °C→r.t., overnight; viii) *N*,*N*-di*iso*propylamino-2-cyanoethoxychlorophosphine (CEP-Cl), Et_3N, dry DCM, 0 °C, 20 min; ix) 30% NH_3, rt, 15 min.

4.2.1.3. Attempted synthesis of *R*-and *S*-ddiazep phosphoramidites

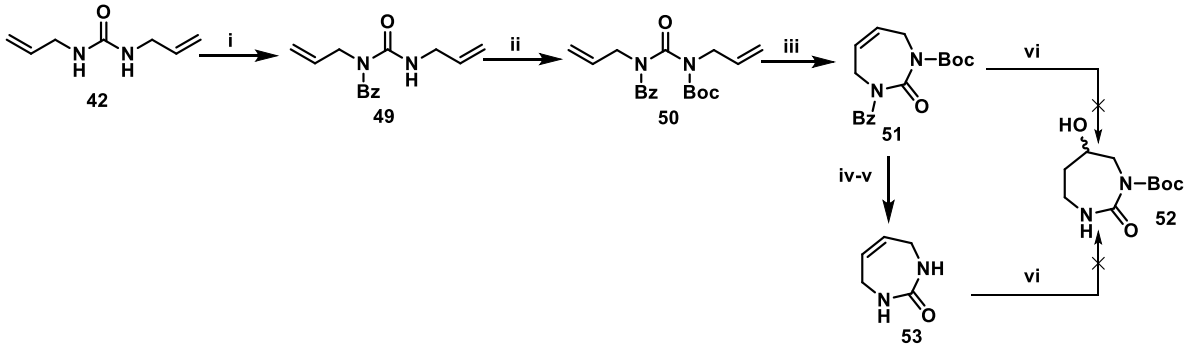


Scheme 4.4. Attempted hydroboration oxidation of the compound **45**.

Reagents and conditions: i) 1 M BH_3.THF, 0 °C→r.t., 2.5 hr, H_2O, 1 N NaOH, 30% H_2O_2, r.t., 1 hr.

After the successful synthesis of compound **45** we planned to synthesize the 2'-deoxy form of the reported compound **48**. For this we investigated the possibility of a one-pot hydroboration oxidation and removal of all the protecting groups. Unfortunately, the standard hydroboration/oxidation with BH_3·THF followed by H_2O_2/NaOH did not yield the required product **48** (**Scheme 4.4**); instead, it resulted in a mixture of the toluoyl and benzoyl-deprotected and sugar-degraded side products.

The above failed reaction on the substrates containing sugar and other protecting groups made us investigate the possibility of hydroboration/oxidation on a protected nucleobase **51**. We started the synthesis with our previously synthesized diallyl urea **42**. Benzoyl protection of the compound **42** followed by *tert*-butyloxy carbonyl (boc) protection allowed the facile RCM of compound **50** producing the protected nucleobase **51** in 68% yield. The attempted hydroboration/oxidation protocol resulted in benzoyl deprotection without any hydroboration oxidation. Next, we tried the same procedure on a free nucleobase **53** which was reported previously.[200] Unfortunately, the reported procedure did not work in our hands (**Scheme 4.5**).



Scheme 4.5. Attempted synthesis of a free **ddiazep** nucleobase and subsequent hydroboration oxidation. Reagents and conditions: i) Benzoyl chloride, pyridine, 0 °C→r.t., overnight; ii) Boc-anhydride, DMAP, Et_3N, THF, reflux, overnight; iii) GreenCat™, dry DCM, r.t., 2hr; iv) 30% ammonia solution in MeOH, 15 min; v) TFA, DCM, 15 min; vi) 1M BH_3·THF, 0 °C→r.t., 2.5 hr, H_2O, 1N NaOH, 30% H_2O_2, r.t., 1 hr.

4.2.2. Oligonucleotide synthesis

Oligodeoxynucleotides (oligos) were synthesized by the phosphoramidite method on a MerMade-4 DNA/RNA synthesizer (BioAutomation) on a 5 μmol scale using the standard manufacturer's protocol for native nucleotides, whereas increased coupling time from 2 to 10 min was used for modified phosphoramidites. Detailed procedures are given in the experimental section. We incorporated the modified nucleosides at the location of dC in the preferred A3 substrate motifs. Oligos were cleaved from the solid support and deprotection of phosphates and nucleobases was accomplished in concentrated aqueous NH_4OH. The TBDPS on dazep oligos were deprotected by TEA·3HF. All oligos were purified by reverse-phase HPLC on 250/4.6 mm, 5 μm, 300 Å C18 column (Thermo Fisher Scientific) using a gradient of CH_3CN (0→20% for 20 min, 1.3 mL/min) in 0.1 M TEAA buffer (pH 7.0) with detection at 260 nm. ESI-MS (**Table 4.2.**) confirmed their compositions.

Table 4.2. List of oligos synthesized as inhibitors of A3 enzymes.

Name	DNA sequence, 5'→3'	Retention time (min)	ESI-MS [Da] found/calculated
7-mer dZ	TTTT**dZ**AT	14.2	2045.3722/2045.3678
9-mer ddiazep	ATTT**ddiazep**ATTT	15.9	2679.4839/2679.5024
7-mer *R*-dazep	TTTT***R*-dazep**AT	13.4	2078.3888/2078.4141
7-mer *S*-dazep	TTTT***S*-dazep**AT	13.4	2078.3824/2078.4141

4.2.3. Evaluation of CDA inhibition by UV- Vis based deamination assay

The seven-membered ring-containing nucleosides synthesized were tested for their inhibition potential against human CDA enzyme by a UV-Vis based deamination assay. All data were obtained by UV-Vis spectroscopy of 2'-deoxycytidine at 286 nm and 25 °C, where the deamination of cytidine leads to a decrease in absorption. For the ***R*-**and ***S*-dazep** nucleosides, the condition of the assay used was 27 nM human CDA and 100 µM dC as a substrate at pH 6.0; an average of 8 measurements were performed for each nucleosides. As the original ***R*-**and ***S*-ddiazep** nucleosides were reported as slow-binding inhibitors, preincubation conditions of an inhibitor with CDA were also tested for the assay. Here, the nucleosides were preincubated with the enzyme for 5 minutes and then 100 µM dC substrate was added to start the measurements (**Figure 4.7**).

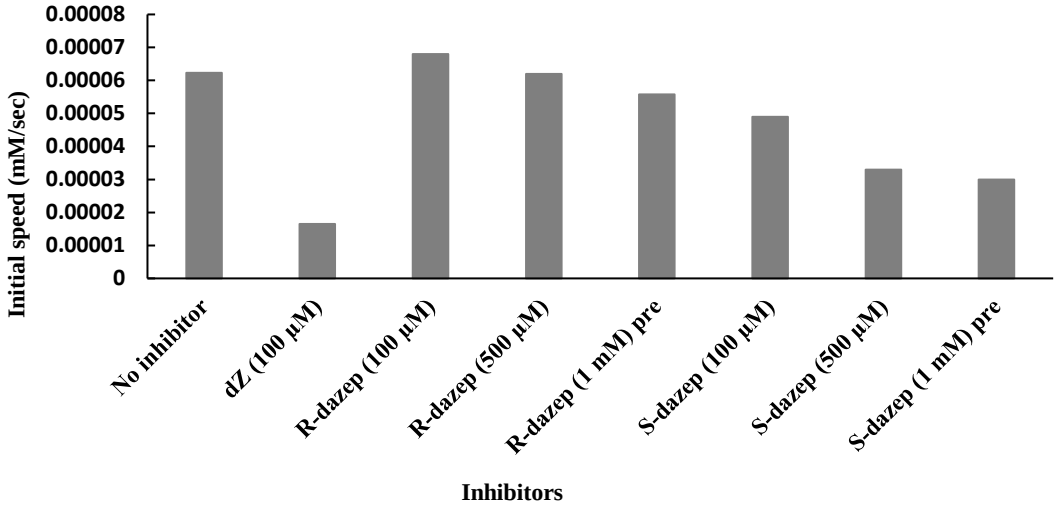


Figure 4.7. Initial rate of human CDA-catalyzed deamination of dC with and without inhibitors.

Conditions: 100 µM dC as a substrate, using different concentrations of inhibitors, 27 nM human CDA, pH 6.0 at 25 °C using UV-Vis. Measurements were taken at 286 nm.

The CDA inhibitor ribose form of **ddiazep** was reported as a very good inhibitor of CDA. Therefore, we decided to test **ddiazep** using our UV-Vis based assay and analyze using Lambert's W function.[141] The deamination of dC was monitored over a period of time after the addition of a known concentration of the inhibitor and the UV absorbance was measured at 286 nm monitoring consumption of dC. The time course of the deamination was recorded until the absorption did no longer change. A blank measurement without the inhibitor and a control measurement with a known concentration of dZ were also performed. Usually approximately 2 mM 2'-deoxycytidine (dC) was deaminated by 27 nM of CDA (**Figure 4.8**).

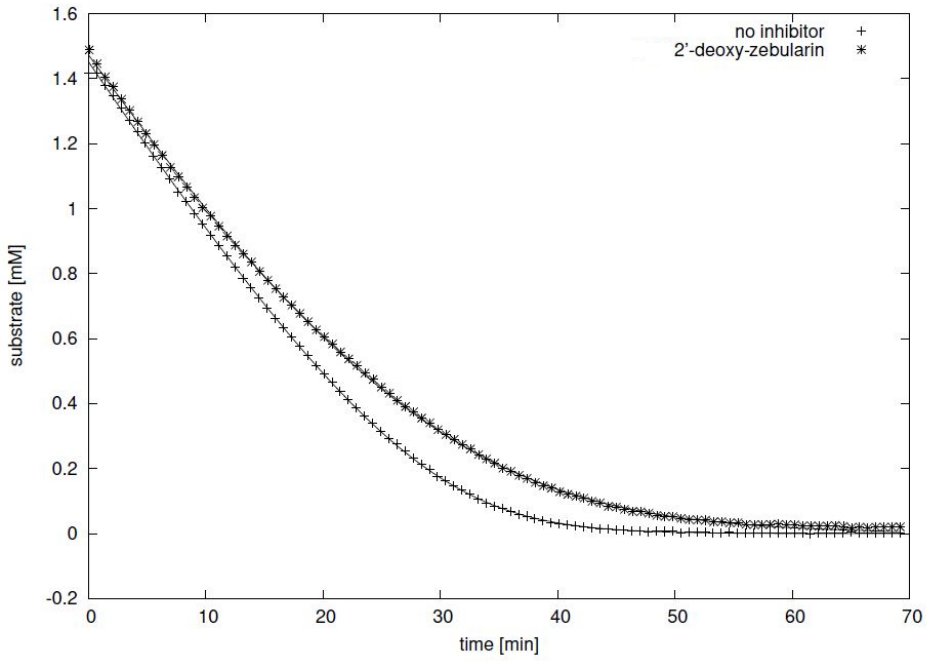


Figure 4.8. Cytidine deaminase-catalyzed deamination of 2'-deoxycytidine.

UV-Vis spectroscopy at 286 nm of a 2 mM solution of 2'-deoxycytidine in the presence of 27 nM CDA. The uninhibited deamination (pluses) and the reaction inhibited by the addition of 5 µM 2'-deoxyzebularine, start at about the same speed. However, as the reaction progresses, the inhibited reaction becomes slower, as its observed K_m is larger than in the uninhibited case. To be able to visualize the fitted curves, only every 5th data point is plotted.

The data were then fitted to the integrated Michaelis Menten equation:

$$[\boldsymbol{mm}]_t = \boldsymbol{K_m W}\left(\frac{\boldsymbol{c}}{\boldsymbol{K_m}}.\boldsymbol{e}^{\frac{\boldsymbol{c}-\boldsymbol{V_{max}t}}{\boldsymbol{K_m}}}\right)+\boldsymbol{d}$$

where W is Lambert's W function, $[mm]_t$ the current substrate concentration with respect to the time t, $[S]_0$ is the initial substrate concentration, V_{max} and K_m are the Michaelis-Menten constants and t is the time, c the initial substrate concentration, and d a baseline correction parameter.

Two exemplary curves of enzyme-catalyzed deamination are shown in **Figure 4.8**. One reaction as control has only substrate and enzyme, while the second reaction contains 5 µM of 2'-deoxyzebularine (dZ). For each inhibitor several time-course reactions were recorded for different inhibitor concentrations. To get an average K_i value, the observed K_m values were plotted versus the inhibitor concentration [I] (**Figure 4.9** and **Figure 4.10**). A linear fit of this data (observed K_m versus [I] as y = a+bx) has the noninhibited CDA K_m as y-axis intersection.

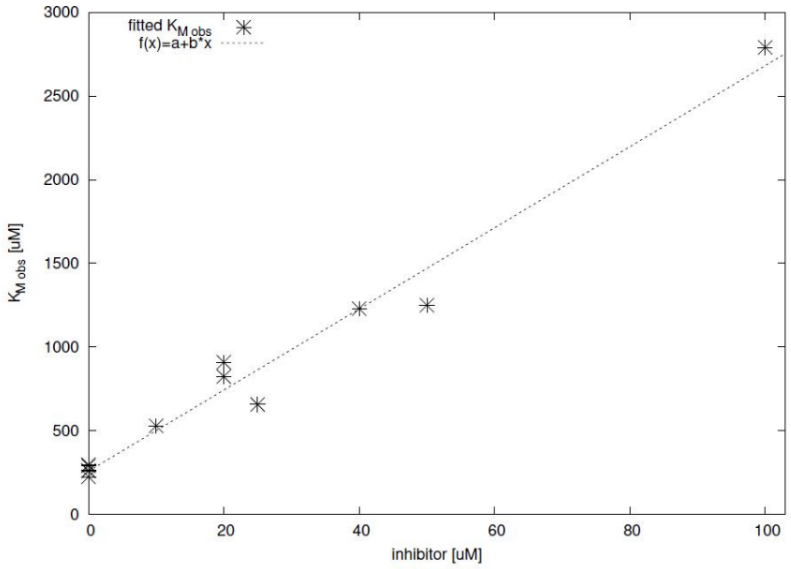


Figure 4.9. Plot of fitted $K_{m\ obs}$ values vs the inhibitor concentration of dZ at pH 7.4.

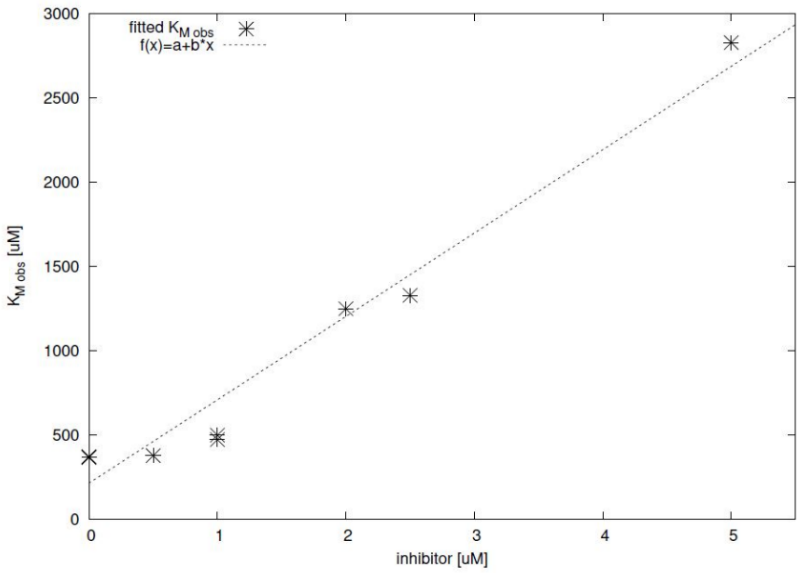


Figure 4.10. Plot of fitted $K_{m\ obs}$ values vs the inhibitor concentration of **ddiazep** at pH 7.4.

The competitive inhibition constant K_i of the inhibitor can be calculated from the fitted parameters (K_i = a/b) and are given in **Table 4.3**.

Table 4.3. K_m of the substrate dC and K_i of dZ and ddiazep determined by UV-vis based assay.

Inhibitor	dC K_m (μM)	K_i (μM)
dZ	260 ± 40	10.7 ± 0.5
ddiazep	220 ± 80	0.43 ± 0.08

4.2.4. Evaluation of A3 inhibitors by NMR assay

We used the NMR assay to compare our new A3 inhibitors with a linear DNA inhibitor containing dZ (5'-T_4ZAT) that was characterized earlier.[54, 124, 126] Residual activity of $A3B_{CTD}$-QM-ΔL3-AL1swap on the unmodified oligo (5'-T_4**C**AT) as a substrate in the presence of a known concentration of inhibitors was measured using the NMR assay (**Figure 4.11**).

The results revealed that although **ddiazep** nucleoside was described as a powerful tight binding inhibitor of CDA, when incorporated into a preferred sequence of ssDNA it was not a better inhibitor of $A3B_{CTD}$-QM-ΔL3-AL1swap compared to 7-mer dZ. It was observed that 4.5 times the concentration of this oligo gave us a comparable inhibition as that of 7-mer dZ. Interestingly, the inhibition by ***R*-dazep** and ***S*-dazep** containing oligos was poor compared to the standard inhibitor (7-mer dZ), but difference in inhibitory potential observed between ***R*** and ***S*** isomers suggested the orientation of -OH which mimics the transition state formed after the attack of H_2O relative to the cytosine in $A3B_{CTD}$-QM-ΔL3-AL1swap. These oligos were not inhibiting the enzyme presumably due to the lack of the N3 atom on the nucleobase, which is present in almost all reported transition-state analogues of CDA providing key hydrogen-bonding interactions. Our efforts to synthesize the precursors containing the seven-membered ring containing inhibitors with the N3 atom present did not work despite trying different schemes and protocols.

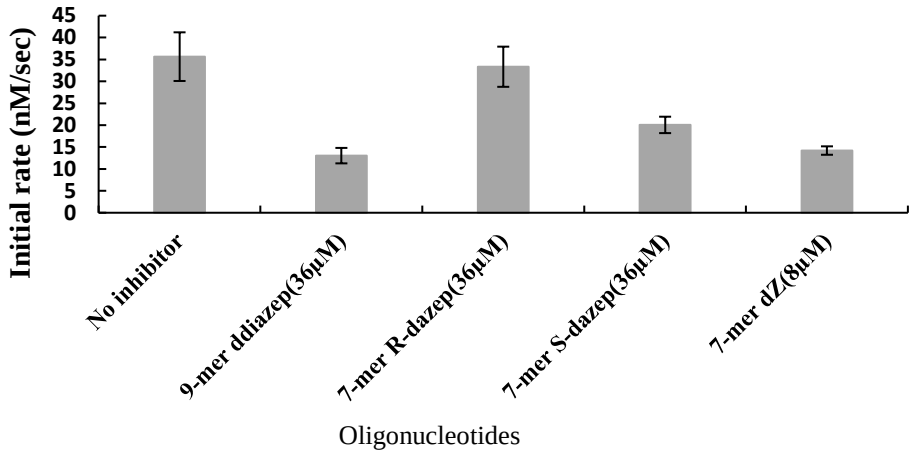


Figure 4.11. Inhibition of $A3B_{CTD}$-QM-ΔL3-AL1swap-catalyzed deamination of 5'-T_4dCAT by transition state analogue containing oligos.

Initial rate of deamination was measured under the following experimental conditions: 400 μM of 5'-T_4CAT, 36 μM of seven-membered ring containing oligos and 8 μM of dZ-containing oligo, 300 nM of $A3B_{CTD}$-QM-ΔL3-AL1swap in a 50 mM sodium phosphate buffer (pH 6.0) containing 100 mM NaCl, 2.5 mM β-mercaptoethanol, 50 μM 3-(trimethylsilyl)-2,2,3,3-tetradeuteropropionic acid (TSP) and 20% D_2O at 25 °C. Error bars are estimated standard deviations from triplicate measurements.

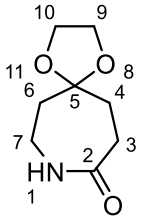
4.3. Conclusions

In this study we developed a synthetic protocol for the synthesis of the reported 2'-deoxy derivatives of a known and a pair of new CDA inhibitors and their facile incorporation into previously described ssDNA sequence as inhibitors of A3 enzymes. Although the synthesized inhibitors were not as promising as expected, the data suggest there is a subtle variation between the mode of binding and therefore the active site of CDA and A3 enzymes. The inhibitor with the double bond (**ddiazep**) was a powerful inhibitor in our experiments with hCDA as reported, but the inhibitory potential was poor against A3, showing the subtle difference in the inhibition of A3 vs CDA. The differences in the inhibitory potential of ***R*** and ***S*** isomers of **dazep** shed light on the relative stereochemistry of the hydroxyl substituent which is important in the inhibition of CDA and A3 enzymes. This study therefore provides useful information for the future development of transition-state analogues of cytidine deamination as powerful inhibitors of A3 enzymes.

4.4. Experimental section

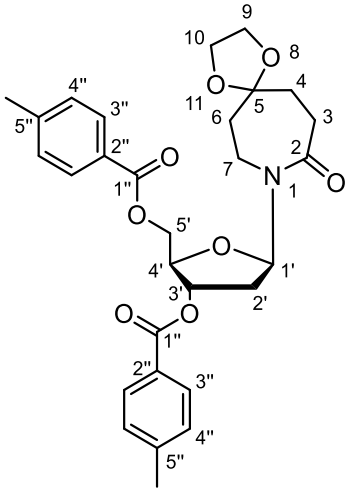
4.4.1. Synthesis of 1,4-dioxa-8-azaspiro[4.6]undecan-9-one (28)

Synthesis of **28** was done from **27** which was prepared by the reported procedure and the ^{1}H NMR was matched with the literature (yield 8 g, 73%).[198]



^{1}H NMR (500 MHz, DMSO-d_6) δ 7.51 (s, 1H, NH); 3.89, 3.88 (2s, 4H, H-9, H-10); 3.09-3.05 (m, 2H, H-7), 2.30-2.26 (m, 2H, H-3); 1.70-1.66 (m, 2H, H-4); 1.65-1.63 (m, 2H, H-6).

4.1.2. Synthesis of 1-[2-deoxy-3,5-bis-*O*-(4-methylbenzoyl)-β-D-*erythro*-pentofuranosyl]-5-azepan-2-one monethylene ketal (30)



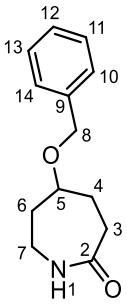
To a solution of compound **28** (7.5 g, 0.99 mmol) in 250 mL of 1,4-dioxane was added K-^{t}BuO (5.8 g, 52 mmol) followed by Hoffer's chlorosugar (14.0 g, 36.0 mmol) and the solution was stirred at rt for 2 hr. After the consumption of the starting materials as observed by TLC, the reaction mixture was diluted with EtOAc (100 mL) and washed with water. Organic layer was separated, dried over anhydrous sodium sulfate, filtered, and concentrated *in vacuo*. The crude mixture was purified by column chromatography over silica gel eluting with 0-30% EtOAc in hexane to afford the desired pure isomers as a foam (R_f = 0.35, 30% EtOAc in hexane, yield 4.5 g, 64%, with purity over 99% for the β-nucleoside as determined by ^{1}H NMR).

^{1}H NMR (500 MHz, DMSO-d_6) δ 7.91-7.89 (m, 4H, H-3"); 7.36-7.34 (m, 4H, H-4"); 6.37 (dd, 1H, J = 9.35, 5.65 Hz, H-1'), 5.50 (app dt, 1H, J = 6.55, 2.20 Hz, H-3'), 4.59 (dd, 1H, J = 11.83, 4.05 Hz, H-5'a), 4.45 (dd, 1H, J = 11.83, 4.75 Hz, H-5'b), 4.31-4.29 (m, 1H,

H-4'); 3.90-3.85 (m, 2H, H-9); 3.83-3.77 (m, 2H, H-10); 3.31-3.19 (m, 2H, H-7); 2.48-2.42 (m, 2H, H-3); 2.39 (s, 6H, CH_3); 2.30-2.23 (m, 1H, H-2'a); 2.17 (ddd, 1H, *J* = 14.10, 5.55, 1.65 Hz, H-2'b), 1.73-1.65 (m, 2H, H-4); 1.53 (t, 2H, *J* = 5.00 Hz, H-6).

^{13}C NMR (125.7 MHz, DMSO-d_6) δ 175.09 (1C, C2), 165.92 (1C, C1"); 165.76 (1C, C1"); 144.46 (1C, C5"); 144.37 (1C, C5"); 129.84 (4C, C3"); 129.79, 129.72 (4C, C4"); 127.13 (1C, C2"); 127.03 (1C, C2"); 108.64 (1C' C5); 83.31 (1C, C1'); 80.34 (1C, C4'); 75.50 (1C,); 64.64 (1C; C5'); 64.43 (1C, C9); 64.41 (1C, C10); 38.95 (1C, C6); 37.12 (1C, C7); 34.27 (1C, C2'); 33.02 (1C, C4); 32.24 (1C, C3); 21.65, 21.61 (2C, CH_3).

4.4.3. Synthesis of 5-benzyloxyazepan-2-one (33)



Synthesis of **33** was performed from **27** by the reported procedures (yield 7.6 g, 70%).[198, 201]

^{1}H NMR (500 MHz, $CDCl_3$) δ 7.36-7.32 (m, 4H, H-10, H-11, H-13, H-14), 7.30-7.25 (m, 1H, H-12), 6.69 (s, 1H, NH), 4.55, 4.51 (2d, 2H, *J* = 11.90 Hz, H-8), 3.70-3.68 (m, 1H, H-5), 3.56-3.50 (m, 1H, H-7), 3.01-2.95 (m, 1H, H7), 2.84-2.79 (m, 1H, H-4), 2.24-2.19 (m, 1H, H-4), 1.97-1.90 (m, 2H, H-6), 1.87-1.80 (m, 2H, H-3).

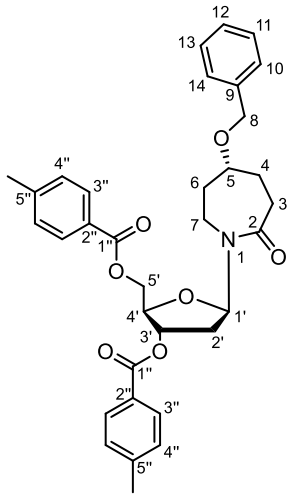
^{13}C NMR (125.7 MHz, $CDCl_3$) δ 178.96 (C2), 138.59 (C9), 128.52 (C10, C14), 127.72 (C12), 127.50 (C11, C13), 75.92 (C8), 70.15 (C5), 37.15 (C7), 34.58 (C3), 30.05 (C4), 27.81 (C6).

HRMS (ESI) *m*/*z*: $[M+H]^+$ Calcd for $C_{13}H_{18}NO_2$ 220.1332; found 220.1331

4.4.4. Synthesis of 1-[2-deoxy-3,5-bis-*O*-(4-methylbenzoyl)-β-D-*erythro*-pentofuranosyl]-(*R*)-5-benzyloxyazepan-2-one (34*R*) and 1-[2-deoxy-3,5-bis-*O*-(4-methylbenzoyl)-β-D-*erythro*-pentofuranosyl]-(*S*)-5-benzyloxyazepan-2-one (34*S*)

To a solution of the compound **33** (7.5 g, 0.99 mmol) in 250 mL of 1,4-dioxane was added K-^{t}BuO (5.8 g, 52 mmol) followed by Hoffer's chlorosugar (14.0 g, 36.0 mmol) and the solution was stirred at rt for 2 hr. After the consumption of the starting materials as observed by TLC, the reaction mixture was diluted with EtOAc (100 mL) and washed with water. The organic layer was separated, dried over anhydrous sodium sulfate, filtered, and

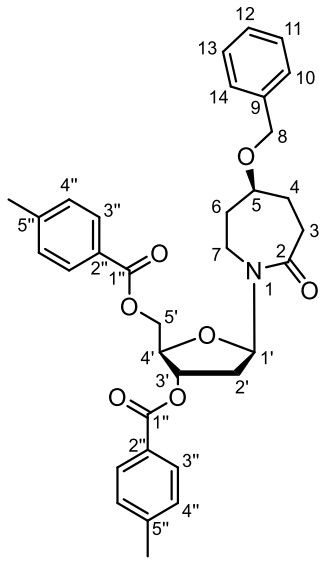
concentrated *in vacuo*. The crude mixture was purified by column chromatography over silica gel eluting with 0-30% EtOAc in hexane to afford the desired pure isomers as a foam (yield 8.5 g, 64% of *R* isomer R_f = 0.45, 30% EtOAc in hexane, and 36% *S* isomer R_f = 0.42, 30% EtOAc in hexane, with purity over 99% as β-nucleosides as determined by ^{1}H NMR).



^{1}H NMR (500 MHz, $CDCl_3$) δ 7.94-7.92 (m, 4H, H-11, H-12, H-14, H-15), 7.35-7.32 (m, 2H, H-3''), 7.29-7.27 (m, 3H, H-3'', H-13), 7.26-7.23 (m, 4H, H-4''), 6.58 (dd, 1H, *J* = 9.60, 5.40 Hz, H-1'), 5.52-5.50 (m, 1H, H-3'), 4.66 (dd, 1H, *J* = 11.89, 3.45 Hz, H-5'), 4.53 (dd, 1H, *J* = 11.89, 3.87 Hz, H-5'), 4.47, 4.43 (2d, 2H, *J* = 11.90 Hz, H-8), 4.34 (td, 1H, *J* = 3.45, 1.36 Hz, H-4'), 3.63 (bs, 1H, H-5), 3.61-3.58 (m, 1H, H-7), 3.14 (dd, 1H, *J* = 15.69, 7.40 Hz, H-7), 2.94 (t, 1H, 12.65 Hz, H-3), 2.42 (s, 3H, CH_3), 2.41 (s, 3H, CH_3), 2.33-2.28 (m, 1H, H-2'a), 2.27-2.23 (m, 1H, H-3), 2.18-2.12 (m, 1H, H-2'b), 1.97-1.94 (m, 1H, H-6), 1.85-1.75 (m, 2H, H-6, H-4), 1.60-1.55 (m, 1H, H-4).

^{13}C NMR (125.7 MHz, $CDCl_3$) δ 176.51 (C2), 166.31 (C1''), 166.25 (C1''), 144.41 (C5''), 144.13 (C5''), 138.56 (C10), 129.92, 129.72 (4C, C3''), 129.37, 129.33 (4C, C4''), 128.52 (C11), 127.72 (C13), 127.44 (C12), 127.10, 126.69 (2C, C2''), 83.36 (C1'), 80.74, (C4'), 75.16 (C3'), 74.79 (C5), 69.90 (C9), 64.55 (C5'), 36.22 (C7), 34.92 (C2'), 34.01 (C6), 31.31 (C3), 28.12 (C4), 21.81 (2C, C6'').

HRMS (ESI) *m*/*z*: $[M+Na]^+$ Calcd for $C_{34}H_{37}NO_7Na$ 594.2462; found 594.2466.



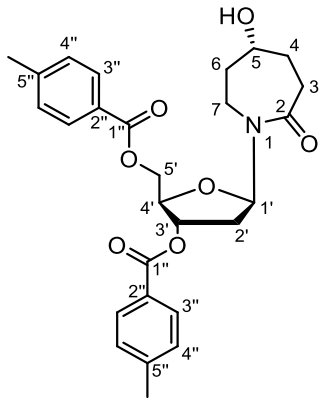
^{1}H NMR (500 MHz, $CDCl_3$) δ 7.94-7.93 (m, 4H, H-11, H-12, H-14, H-15), 7.35-7.33 (m, 2H, H-3''), 7.29-7.24 (m, 7H, H-3'', H-4'', H-13), 6.59 (dd, 1H, *J* = 9.47, 5.46 Hz, H1''), 5.54-5.52 (m, 1H, H-3'), 4.73 (dd, 1H, *J* = 11.88, 3.26 Hz, H-5'), 4.51 (dd, 1H, *J* = 11.88, 3.74 Hz, H-5'), 4.45, 4.41 (2d, 2H, *J* = 11.80 Hz, H-8), 4.32 (td, 1H, *J* = 3.11, 1.92 Hz, H-3'), 3.63-3.60 (m, 1H, H-5), 3.56-3.52 (m, 1H, H-7), 3.08-3.05 (m, 1H, H-7), 2.88-2.85 (m, 1H, H-3), 2.41 (s, 3H, CH_3), 2.41 (s, 3H,CH_3), 2.34-2.31 (m, 1H, H-3), 2.28-2.25 (m, 1H, H-2'a), 2.18-2.14 (m, 1H, H-2'b), 1.93-1.87 (m, 2H, H-4), 1.73-1.67 (m, 2H, H-6).

^{13}C NMR (125.7 MHz, $CDCl_3$) δ 176.27 (C2), 166.34 (C1''), 166.23 (C1''), 144.41 (C5''), 144.24 (C5''), 138.58 (C10), 129.93, 129.70 (4C, C3''), 129.42, 129.33 (4C, C4''), 128.53 (C11), 127.75 (C13), 127.47 (C12), 127.08, 126.69 (2C, C2''), 83.70 (C1'), 81.00, (C4'), 75.14 (C3'), 74.51 (C5), 70.00 (C9), 64.41 (C5'), 36.29 (C7), 34.80 (C2'), 34.06 (C6), 31.49 (C3), 28.19 (C4), 21.83, 21.79 (2C, C6'').

HRMS (ESI) *m/z*: $[M+Na]^+$ Calcd for $C_{34}H_{37}NO_7Na$ 594.2462; found 594.2464.

4.4.5. Synthesis of 1-[2-deoxy-3,5-bis-*O*-(4-methylbenzoyl)-β-D-*erythro*-pentofuranosyl]-(*R*)-5-hydroxyazepan-2-one (35*R*) and 1-[2-deoxy-3,5-bis-*O*-(4-methylbenzoyl)-β-D-*erythro*-pentofuranosyl]-(*S*)-5-hydroxyazepan-2-one (35*S*)

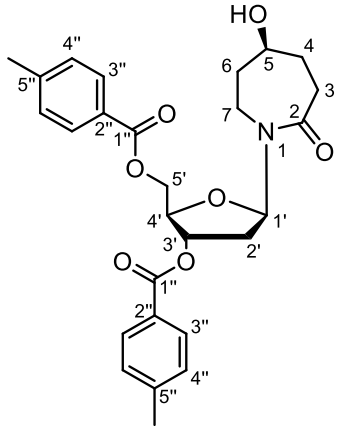
After two vacuum/H_2 cycles to replace air inside the flask with hydrogen, the mixture of the substrate (5 gm 8.8 mmol), 10% Pd/C (10 wt % of the substrate) in EtOH (100 mL) was stirred at room temperature under hydrogen (balloon) for 6 h. The reaction mixture was filtered using celite pad, and washed twice with 50 mL EtOH, the filtrate was concentrated *in vacuo* to provide the crude product. It was purified by column chromatography over silica gel eluting with 0-10% MeOH in DCM to afford the pure products as a foam (yield 3.9 g, 93%).



R_f = 0.47 in 10% MeOH in DCM

^{1}H NMR (500 MHz, $CDCl_3$) δ 7.93-7.90 (m, 4H, H-3''), 7.25-7.22 (m, 4H, H-4''), 6.57 (dd, 1H, *J* = 9.56, 5.40 Hz, H-1'), 5.52-5.50 (m, 1H, H-3'), 4.65 (dd, 1H, *J* = 11.89, 3.36 Hz, H-5'a), 4.52 (dd, 1H, *J* = 11.88, 3.82 Hz, H-5'b), 4.33 (td, *J* = 3.0, 1.60 Hz, H-4'), 3.93 (bs, 1H, H-5), 3.55 (dd, 1H, *J* = 15.49, 9.14 Hz, H-7), 3.12 (dd, 1H, *J* = 15.49, 8.05 Hz, H-7), 2.87-2.82 (m, 1H, H-3), 2.41 (s, 3H, CH_3), 2.40 (s, 3H, CH_3), 2.34-2.29 (m, 1H, H-4), 2.25 (dd, 1H, *J* = 13.80, 5.24 Hz, H-2'a), 2.17-2.11 (m, 1H, H-2'b), 1.92 (t, 1H, *J* = 12.24 Hz, H-3), 1.79-1.76 (m, 1H, H-4), 1.74-1.68 (m, 1H, H-6), 1.64-1.58 (m, 1H, H-6).

^{13}C NMR (125.7 MHz, $CDCl_3$) δ 176.15 (C2), 166.33 (C1''), 166.25 (C1''), 144.45 (C5''), 144.23 (C5''), 129.93, 129.69 (4C, C3''), 129.41, 129.34 (4C, C4''), 127.02, 126.69 (2C, C2''), 83.47 (C1'), 80.84, (C4'), 75.12 (C3'), 69.34 (C5), 64.55 (C5'), 37.30 (C6), 36.35 (C7), 34.99 (C2'), 31.39 (C3), 21.87, 21.82 (2C, 2 x CH_3).



R_f = 0.46 in 10% MeOH in DCM

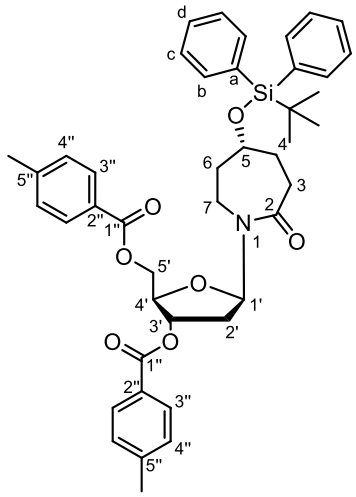
^{1}H NMR (500 MHz, $CDCl_3$) δ 7.93-7.90 (m, 4H, H-3''), 7.25-7.23 (m, 4H, H-4''), 6.57 (dd, 1H, *J* = 9.52, 5.52 Hz, H-1'), 5.52-5.50 (m, 1H, H-3'), 4.67 (dd, 1H, *J* = 11.90, 3.25 Hz, H-5'a), 4.51 (dd, 1H, *J* = 11.90, 3.80 Hz, H-5'b), 4.31 (td, 1H, *J* = 3.20, 1.70 Hz, H-4'), 3.96-3.92 (m, 1H, H-5), 3.54 (dd, 1H, *J* = 15.47, 9.02 Hz, H-7), 3.08 (dd, 1H, *J* = 15.47, 8.32 Hz, H-7), 2.81 (t, 1H, *J* =12.35 Hz, H-3), 2.41 (s, 3H, CH_3), 2.40 (s, 3H, CH_3), 2.37-2.33 (m, 1H,

H-4), 2.26 (dd, 1H, , 13.62, 5.17 Hz, H-2'a), 2.19-2.13 (m, 1H, H-2'b), 1.94 (t, 1H, *J* = 12.10 Hz, H-3), 1.84-1.78 (m, 2H, H-6), 1.75-1.69 (m, 1H, H-4), 1.53-1.47 (m, 1H, H-6).

^{13}C NMR (125.7 MHz, $CDCl_3$) δ 176.04 (C2), 166.34 (C1"), 166.26 (C1"), 144.45 (C5"), 144.29 (C5"), 129.94, 129.66 (4C, C3"), 129.45, 129.35 (4C, C4"), 127.00, 126.69 (2C, C2"), 83.57 (C1'), 80.93, (C4'), 75.12 (C3'), 69.31 (C5), 64.50 (C5'), 37.24 (C6), 36.30 (C7), 34.80 (C2'), 31.46 (2C, C3, C4), 21.84, 21.82 (2C, 2 x CH_3).

4.4.6. Synthesis of 1-{2-deoxy-3,5-di-*O*-(4-methylbenzoyl)-β-D-*erythro*-pentofuranosyl}- [(*tert*-butyldiphenylsilyl)-(*R*)-5-oxy]azepan-2-one (36*R*) and 1-{2-deoxy-3,5-di-*O*-(4-methylbenzoyl)-β-D-*erythro*-pentofuranosyl}-[(*tert*-butyldiphenylsilyl)-(*S*)-5-oxy]azepan-2-one (36*S*)

To a stirring solution of **35*R*** or **35*S*** (1.8 g, 4.0 mmol) in dry DCM (100 mL) at 0 ^{o}C was added imidazole (1.0 g, 14.7 mmol) followed by TBDPS chloride (1.46 mL, 3.6 mmol) dropwise. After the addition, the reaction mixture was slowly brought to room temperature and continued stirring for 6 hr. The reaction was monitored by TLC, after consumption of the starting material, 100 mL DCM was added to the reaction mixture and washed with water followed by brine. The organic layer was separated, dried over anhydrous sodium sulfate, filtered, and concentrated *in vacuo*. The crude product was purified by column chromatography over silica gel eluting with 0-30% EtOAc in hexane to afford the product as a foam (yield 2.3 g, 85%).



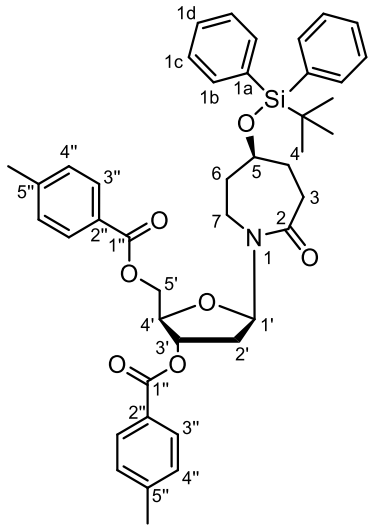
R_f = 0.40 in 30% EtOAc in hexane

^{1}H NMR (500 MHz, $CDCl_3$) δ 7.93-7.89 (m, 4H, H-c), 7.61-7.58 (m, 4H, H-b) 7.44-7.40 (m, 2H, H-d), 7.37-7.32 (m, 4H, H-3"), 7.25-7.18 (m, 4H, H-4"), 6.58 (dd, 1H, *J* = 9.62, 5.32 Hz, H-1'), 5.50 (dt, 1H, *J* = 3.40, 2.55 Hz H-3'); 4.67 (dd, 1H, *J* = 11.90, 3.40 Hz, H-5'a), 4.56 (dd, 1H, *J* = 9.52, 5.52 Hz, H-5'b), 4.35 (td, 1H, *J* = 3.30, 1.80 Hz, H-4'), 4.00 (s, 1H,

H-5), 3.76 (td, 1H, *J* = 14.62, 5.3 Hz, H-7) 3.17-3.08 (m, 2H, H-7, H-3), 2.41 (2s, 6H, CH_3), 2.24 (dd, 1H, *J* = 14.18, 8.46 Hz, H-3), 2.20 (dd, 1H, *J* = 14.55, 6.7 Hz, H-2'a); 2.10 (ddd, 1H, *J* = 16.30, 9.60, 6.7 Hz, H-2'b), 1.78-1.73 (m, 1H, H-4), 1.65-1.59 (m, 2H, H-6, H-4), 1.39 (t, 1H, *J* = 10.97 Hz, H-6), 1.04 (s, 9H, CH_3).

^{13}C NMR (125.7 MHz, $CDCl_3$) δ 176.38 (C2), 166.36 (Ca), 166.21 (Ca), 144.42 (C5''), 144.22 (C5''), 135.74 (2C, Cb), 134.08 (2C, Cb), 133.87 (1C, C1''), 129.95, 129.66 (4C, Cc), 129.41, 129.34 (4C, C4''), 127.80, 126.79 (4C, C3''), 127.05, 126.74 (2C, Cd), 83.78 (1C, C1'), 80.93, (1C, C4'), 75.20 (1C, C3'), 69.31 (C1, C5), 64.48 (1C, C5'), 37.47 (1C, C6), 35.94 (1C, C7), 34.93 (1C, C2'), 31.25 (1C, C3), 31.11 (1C, C4), 27.10 (3C, C-*CH$_3$*), 21.85, 21.82 (2C, C6'''), 19.36 (1C, *C*-CH_3).

HRMS (ESI) *m*/*z*: $[M+H]^+$ Calcd for $C_{43}H_{50}NO_7Si$ 720.3351; found 720.3340.



R_f = 0.42 in 30% EtOAc in hexane

^{1}H NMR (500 MHz, $CDCl_3$) δ 7.94-7.91 (m, 4H, H-c) 7.59-7.54 (m, 4H, H-b), 7.44-7.40 (m, 2H, H-d), 7.35-7.32 (m, 4H, H-3'') 7.25-7.20 (m, 4H, H-4''), 6.57 (dd, 1H, *J* = 9.27, 5.58 Hz, H-1'), 5.54-5.52 (m, 1H, H-3'), 4.67 (dd, 1H, *J* = 11.91, 3.16 Hz, H-5'a), 4.48 (dd, 1H, *J* = 11.91, 3.47 Hz, H-5'b), 4.29 (dt, 1H, *J* = 3.40, 2.0 Hz, H-4'), 3.94 (s, 1H, H-5), 3.67-3.62 (m, 1H, H-7), 3.08-3.03 (m, 1H, H-7), 2.96 (bs, 1H, H-3), 2.41 (s, 3H, CH_3), 2.37 (s, 3H, CH_3), 2.28 (dd, 1H, *J* = 13.85, 5.60 Hz, H-2'a), 2.27 (m, 1H, H-3); 2.25-2.17 (m, 2H, H-2'b), 1.72- 1.68 (m, 2H, H-4) 1.61-1.52 (m, 2H, H-6), 1.03 (s, 9H, CH_3).

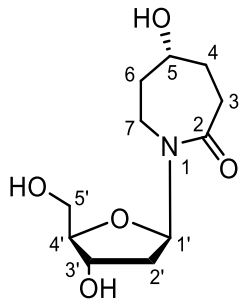
^{13}C NMR (125.7 MHz, $CDCl_3$) δ 176.38 (C2), 166.36, 166.21 (2C, C1''), 144.42, 144.22 (2C, C5''), 135.77, 135.74 (4C, Cb), 134.08, 133.87 (2C, C1a), 129.95, 129.66 (4C, Cc), 129.41, 129.34 (4C, C4''), 127.80, 126.79 (4C, C3''), 127.05, 126.74 (2C, Cd), 83.78 (1C, C1'), 80.93, (1C, C4'), 75.20 (1C, C3'), 69.31 (C1, C5), 64.48 (1C, C5'), 37.47 (1C, C6),

35.94 (1C, C7), 34.93 (1C, C2'), 31.25 (1C, C3), 31.11 (1C, C4), 27.10 (3C, C-*CH*$_3$), 21.85, 21.82 (2C, C6'''), 19.36 (1C, *C*-CH$_3$).

HRMS (ESI) *m*/*z*: $[M+H]^+$ Calcd for $C_{43}H_{50}NO_7Si$ 720.3351; found 720.3350.

4.4.7. Synthesis of 1-[2-deoxy-β-D-*erythro*-pentofuranosyl]-(*R*)-5-hydroxyazepan-2-one (37*R*) and 1-[2-deoxy-β-D-*erythro*-pentofuranosyl]-(*S*)-5-hydroxyazepan-2-one (37*S*)

Toluoyl-protected 2'-deoxynucleoside **35*R*** or **35*S*** (6.10 g, 13.6 mmol) was dissolved in MeOH (500 mL) and aq. ammonia (28%, 50 mL) was added in one portion. Reaction mixture was stirred at room temperature for 48 h, all the volatiles were evaporated *in vacuo*, co-evaporated with H_2O (2 × 200 mL), MeOH (200 mL) several times to get the desired products (yield 3 g, 96%).

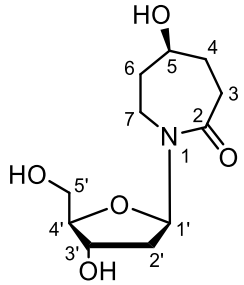


R_f = 0.35 in 30% MeOH in DCM

^{1}H NMR (500 MHz, D_2O) δ 6.33 (dd, 1H, *J* = 8.0, 5.9 Hz, H-1'), 4.30-4.34 (m, 1H, H-3') 4.01-3.95 (m, 1H, H-5), 3.88 (td, 1H, *J* = 4.0, 2.10 Hz, H-4'), 3.76 (dd, 1H, *J* = 12.24, 3.67 Hz, H-5'a), 3.69 (dd, 1H, *J* = 12.24, 5.43 Hz, H-5'b), 3.58 (dd, 1H, *J* = 15.75, 8.53 Hz, H-7), 3.30 (dd, 1H, *J* = 15.75, 9.04 Hz, H-7), 2.71 (t, 1H, *J* = 10.77 Hz, H-3), 2.48-2.41 (m, 1H, H-3), 2.18-2.12 (m, 1H, H-4), 2.07-2.04 (m, 1H, H-2'a), 2.03-1.95 (m, 3H, H-2'b, H4), 1.69-1.57 (m, 2H, H-6).

^{13}C NMR (125.7 MHz, D_2O) δ 179.19 (1C, C2), 85.17 (1C, C4'), 83.61 (1C, C1'), 71.03 (1C, C3'), 69.45 (1C, C5), 61.66 (1C, C5'), 37.47 (1C, C7), 35.94 (1C, C6), 35.72 (1C, C2'), 31.05 (1C, C3), 29.97 (C4).

HRMS (ESI) *m*/*z*: $[M+Na]^+$ Calcd for $C_{11}H_{19}NO_5Na$ 268.1155; found 268.1150.



R_f = 0.35 in 30% MeOH in DCM

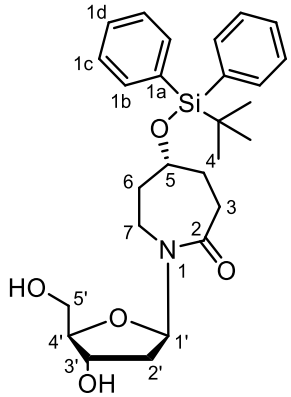
^{1}H NMR (500 MHz, D_2O) δ 6.33 (dd, 1H, J = 8.47, 6.32 Hz, H-1'), 4.37-4.34 (m, 1H, H-3') 3.98-3.94 (m, 1H, H-5), 3.89 (td, 1H, J = 4.3, 2.20 Hz, H-4'), 3.75 (dd, 1H, J = 12.26, 4.03 Hz, H-5'a), 3.68 (dd, 1H, J = 12.26, 5.57 Hz, H-5'b), 3.59 (dd, 1H, J = 15.89, 7.97 Hz, H-7), 3.27 (dd, 1H, J = 15.89, 9.77 Hz, H-7), 2.67-2.51 (m, 2H, H-3), 2.21-2.15 (m, 1H, H-4), 2.08-1.98 (m, 2H, H-2'), 1.63-1.48 (m, 3H, H4, H-6).

^{13}C NMR (125.7 MHz, D_2O) δ 179.23 (1C, C2), 85.11 (1C, C4'), 83.44 (1C, C1'), 71.06 (1C, C3'), 70.00 (1C, C5), 61.66 (1C, C5'), 37.48 (1C, C7), 35.74 (1C, C6), 35.70 (1C, C2'), 31.32 (1C, C3), 30.16 (1C, C4).

HRMS (ESI) *m*/*z*: $[M+Na]^+$ Calcd for $C_{11}H_{19}NO_5Na$ 268.1155; found 268.1151.

4.4.8. Synthesis of 1-[2-deoxy-β-D-*erythro*-pentofuranosyl]-(*R*)-5-[(*tert*-butyldiphenylsilyl)-oxy]azepan-2-one (38*R*) and 1-[-2-deoxy-β-D-*erythro*-pentofuranosyl]-(*S*)-5-[(*tert*-butyldiphenylsilyl)-oxy]azepan-2-one (38*S*)

Compound **37*R*** or **37*S*** (1.2 g, 2.5 mmol) was dissolved in MeOH (100 mL) and aq. ammonia (28%, 15 mL) was added in one portion. Reaction mixture was stirred at room temperature for 48 h, evaporated all the volatiles, co-evaporated with H_2O (2 × 100 mL), MeOH (100 mL) several times to get **38*R*** or **38*S*** as a pure compound (yield 0.72 g, 89%).



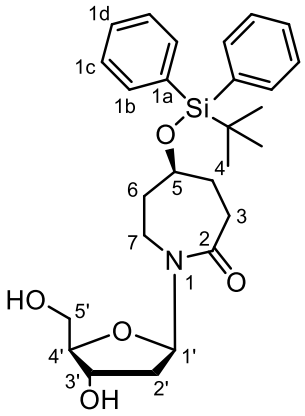
R_f = 0.40 in 20% MeOH in DCM

^{1}H NMR (500 MHz, $CDCl_3$) δ 7.63-7.62 (m, 4H, H-b), 7.44-7.41 (m, 2H, H-d); 7.38-7.36 (m, 4H, H-c); 6.27 (dd, 1H, J = 8.20, 6.50 Hz, H-1'); 4.32-4.29 (m, 1H, H-3'); 4.05 (s,

1H, H-5); 3.82 (td, 1H, *J* = 3.80, 2.80 Hz, H-4'); 3.81-3.79 (m, 1H, H-7); 3.76 (dd, 1H, *J* = 11.60, 3.75 Hz, H-5'a); 3.70 (dd, 1H, *J* = 11.60, 4.05, H-5'b); 3.50 (s, 1H, 3'-OH); 3.10 (d, *J* = 6.95 Hz, 1H, H-7); 3.17 (d, *J* = 6.80 Hz, 1H, H-3); 2.93 (s, 1H, 5'-OH); 2.20 (dd, 1H, *J* = 14.12, 8.47 Hz, H-3); 2.05-1.99 (m, 1H, H-2'a); 1.96- 1.92 (ddd, *J* = 13.30, 6.05, 2.90 Hz, 1H, H-2'b); 1.78-1.72 (m, 2H, H-6, H-4); 1.63 (t, 1H, *J* = 13.20 Hz, H-4); 1.49 (m, 1H, H-6); 1.07 (s, 9H, CH_3).

^{13}C NMR (125.7 MHz, $CDCl_3$) δ 177.25 (C2); 135.82 (2C, Cb); 135.80 (2C, Cb); 133.97 (1C, Ca); 133.92 (1C, C1a); 129.97 (1C, Cd); 129.95 (1C, Cd); 127.83 (4C, Cc); 85.38 (1C, C4'); 84.36 (1C, C1'); 71.98 (1C, C3'); 69.21 (1C, C5); 63.14 (1C, C5'); 37.47 (1C, H7); 37. 39 (1C, C6); 37.01 (1C, C2'); 31.09 (1C, C3); 31.01 (1C, C4); 27.12 (3C, C-*C*H_3); 19.38 (1C, *C*-CH_3).

HRMS (ESI) *m*/*z*: $[M+H]^+$ Calcd for $C_{27}H_{38}NO_5Si$ 484.2514; found 484.2514.



R_f = 0.42 in 20% MeOH in DCM

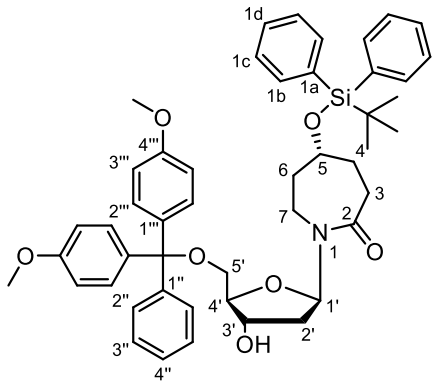
^{1}H NMR (500 MHz, $CDCl_3$) δ 7.63-7.61 (m, 4H, H-b); 7.44-7.41 (m, 2H, H-d); 7.38-7.35 (m, 4H, H-c); 6.27 (dd, 1H, *J* = 7.80, 6.40 Hz, H-1'); 4.31-4.28 (m, 1H, H-3'); 4.00 (s, 1H, H-5); 3.79 (td, 1H, *J* = 4.05, 2.80 Hz, H-4'); 3.76-3.72 (m, 1H, H-7); 3.70 (dd, 1H, *J* = 11.40, 3.30 Hz, H-5'a); 3.64 (dd, 1H, *J* = 11.40, 4.05 Hz, H-5'b); 3.59 (s, 1H, 3'-OH); 3.03 (dd, 1H, *J* = 15.55, 6.90 Hz, H-7); 3.01-2.95 (m, 1H, H-3); 2.93 (s, 1H, 5'-OH); 2.23-2.18 (m, 1H, H-3); 2.10-2.04 (m, 1H, H-2'a); 2.0 (ddd, 1H, *J* = 13.55, 6.30, 3.20 Hz, H-2'b); 1.74-1.61 (m, 4H, H-6, H-4); 1.07 (s, 9H, CH_3).

^{13}C NMR (125.7 MHz, $CDCl_3$) δ 176.87 (1C, C2); 135.80 (2C, Cb); 135.79 (2C, Cb); 133.98 (1C, Ca); 133.97 (1C, Ca); 129.97 (1C, Cd); 129.95 (1C, Cd); 127.83 (4C, Cc); 85.46 (1C, C4'); 84.76 (1C, C1'); 72.02 (1C, C3'); 69.71 (1C, C5); 63.61 (1C, C5'); 37.76 (1C, H7); 37.53 (1C, C2'); 37.31 (1C, C6); 31.27 (1C, C3); 31.20 (1C, C4); 27.13 (3C, C-*C*H3); 19.37 (1C, *C*-CH_3).

HRMS (ESI) *m*/*z*: $[M+H]^+$ Calcd for $C_{27}H_{38}NO_5Si$ 484.2514; found 484.2517.

4.4.9. Synthesis of 1-{2-deoxy-5-*O*-(4,4'-dimethoxytrityl)-β-D-*erythro*-pentofuranosyl}-(*R*)-5-[(*tert*-butyldiphenylsilyl)oxy]azepan-2-one (39*R*) and 1-{2-deoxy-5-*O*-(4,4'-dimethoxytrityl)-β-D-*erythro*-pentofuranosyl}-(*S*)-5-[(*tert*-butyldiphenylsilyl)oxy]azepan-2-one (39*S*)

To a stirring solution of a nucleoside **38*R*** or **38*S*** (0.4 g, 1.16 mmol) in dry pyridine (10 mL), 4,4'-*O*-dimethoxytrityl chloride (0.39 g, 1.15 mmol) was added at 0 °C and the mixture was stirred at r.t. overnight under argon. After consumption of the starting nucleoside, water (1 mL) was added. Solvents were evaporated *in vacuo* and the residue was dissolved in 50 mL DCM and washed with brine (2 × 10 mL). Organic layer was dried over anhydrous sodium sulfate, filtered, and concentrated *in vacuo*. The crude product was purified by column chromatography over silica gel treated with 10% Et_3N in DCM and eluted with 0-5% MeOH in DCM to afford the desired products as a foams (0.68 g, 90%).

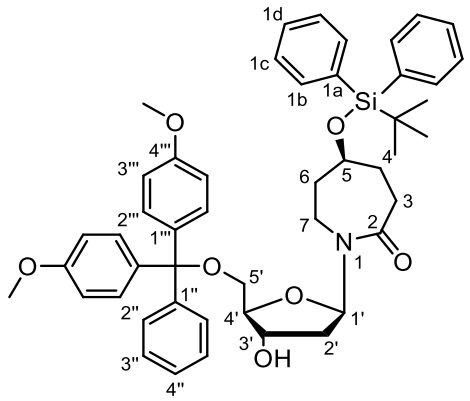


R_f = 0.45 in 5% MeOH in DCM

^{1}H NMR (500 MHz, $CDCl_3$) δ 7.62-7.61 (m, 2H, H-b); 7.57-7.55 (m, 2H, H-b); 7.44-7.40 (m, 3H, H-d, H-4"); 7.37-7.29 (m, 7H, H-c, H-2", H-2'"); 7.27-7.19 (m, 5H, H-c, H-2'"); 6.84- 6.81 (m, 4H, H-3'"); 6.51 (dd, 1H, *J* = 8.10, 6.55 Hz, H-1'); 4.38-4.36 (m, 1H, H-3'); 4.01 (s, 1H, H-5); 3.89 (td, 1H, *J* = 3.90, 2.20 Hz, H-4'); 3.78 (s, 6H, OC*H*$_3$), 3.32 (m, 2H, H-5'); 3.12 (bs, 1H, H-4); 2.81 (m, 2H, H-7); 2.23 (dd, 1H, *J* = 13.99, 8.56 Hz, H-4); 2.03-1.96 (m, 2H, H-2'); 1.80-1.74 (m, 1H, H-3); 1.71-1.63 (m, 2H, H-6, H-3); 1.48-1.40 (m, 1H, H-6); 1.03 (s, 9H, CH_3).

^{13}C NMR (125.7 MHz, $CDCl_3$) δ 176.73 (1C, C2); 158.86 (2C, C4'"), 144.82 (1C, C1"); 135.9 (2C, Ca); 135.77, 135.75 (4C, Cb); 134.04, 133.87 (2C, C1'"); 130.18 (4C, C2'"); 129.91, 129.86 (2C, C2"); 128.25, 128.01 (2C, Cd); 127.79, 127.77 (4C, Cc); 126.97 (2C, C4"); 113.30 (4C, C3'"); 86.57 (1C, *C*-Ar$_3$); 84.35 (1C, C4'); 83.05 (1C, C1'); 73.20 (1C, C3'); 69.20 (1C, C5); 64.29 (1C, C5'); 55.34 (2C, OCH_3); 46.16 (1C, C7); 37.77 (1C, C6); 31.21 (1C, C3); 30.95 (1C, C4); 27.09 (3C, C-*C*H_3); 19.39 (1C, *C*-CH_3).

HRMS (ESI) *m*/*z*: $[M+Na]^+$ Calcd for $C_{48}H_{55}NNaO_7Si$ 808.3640; found 808.3640.



R_f = 0.45 in 5% MeOH in DCM

^{1}H NMR (500 MHz, $CDCl_3$) δ 7.62-7.61 (m, 2H, H-b); 7.58-7.57 (m, 2H, H-b); 7.44-7.39 (m, 3H, H-d, H-2"); 7.38-7.34 (m, 3H, H-2", H-c); 7.31-7.28 (m, 6H, H-2"', H-3"); 7.25-7.18 (m, 3H, H4", H-c); 6.82-6.79 (m, 4H, H-3"'); 6.48 (dd, 1H, J = 8.10, 6.55 Hz, H-1'); 4.40 (m, 1H, H-3'); 4.00 (s, 1H, H-5); 3.85 (td, 1H, J = 3.50, 2.15 Hz, H-4'); 3.78 (2s, 6H, OCH_3), 3.33-3.25 (m, 2H, H-5'); 3.20-3.17 (m, 1H, H-4); 2.99 (bs, 1H, H-4); 2.78 (m, 2H, H-7); 2.28-2.23 (m, 1H, H-6); 2.13- 2.07 (m, 2H, H2'); 1.78-1.73 (m, 1H, H-3); 1.66-1.62 (m, 2H, H-6, H-3); 1.08 (s, 9H, CH_3).

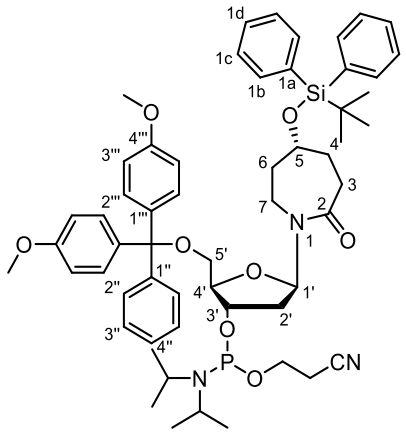
^{13}C NMR (125.7 MHz, $CDCl_3$) δ 176.29 (1C, C2); 158.65 (2C, C4'"), 144.97 (1C, C1"); 135.88 (2C, Ca); 135.79, 135.77 (4C, Cb); 134.01, 133.96 (2C, C1'"); 130.16 (4C, C2'"); 129.92, 129.88 (2C, C2"); 128.15, 127.96 (2C, Cd); 127.98, 127.78 (4C, Cc); 126.96 (1C, C4"); 113.27 (4C, C3'"); 86.43 (1C, *C*-Ar_3); 84.32 (1C, C4'); 83.45 (1C, C1'); 72.98 (1C, C3'); 69.83 (1C, C5); 63.93 (1C, C5'); 55.31 (2C, OCH_3); 46.19 (1C, C7); 38.05 (1C, C6); 31.33 (1C, C3); 31.22 (1C, C4); 27.14 (3C, C-*C*H_3); 19.37 (1C, *C*-CH_3).

HRMS (ESI) *m*/*z*: $[M+Na]^+$ Calcd for $C_{48}H_{55}NNaO_7Si$ 808.3640; found 808.3633.

4.4.10. Synthesis of 1-{3-*O*-(*N*,*N*-di*iso*propylamino-2-cyanoethoxyphosphanyl)-2-deoxy-5-*O*-(4,4'-dimethoxytrityl)-β-D-*erythro*-pentofuranosyl}-(*R*)-5-[(*tert*-butyldiphenylsilyl)oxy]azepan-2-one (40*R*) and 1-{3-*O*-(*N*,*N*-di*iso*propylamino-2-cyanoethoxyphosphanyl)-2-deoxy-5-*O*-(4,4'-dimethoxytrityl)-β-D-*erythro*-pentofuranosyl}-(*R*)-5-[(*tert*-butyldiphenylsilyl)oxy]azepan-2-one (40*S*)

To a stirring solution of the compound **39*R*** or **39*S*** (0.4 g, 0.66 mmol) in dry DCM (10mL) under argon at rt were added Et_3N (0.12 ml, 0.86 mmol) followed by 2-cyanoethyl-*N*,*N*-di*iso*propyl chlorophosphoramidite (0.17 g, 0.71 mmol). After the disappearance of the starting material on TLC in 10 min the reaction mixture was washed with saturated sodium bicarbonate solution (2 × 5 mL) followed by brine (5 mL). The organic layer was dried over anhydrous sodium sulfate, filtered, and the combined fractions were evaporated *in vacuo*.

The crude product was purified by column chromatography over silica gel (60-120 mesh) saturated with Et_3N (10%) and eluted with DCM/acetone (9:1) to give a white foam (0.45 g, 85%).



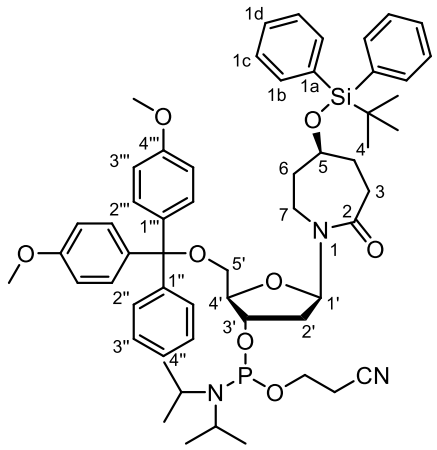
R_f = 0.40 in 10% acetone in DCM

^{1}H NMR (500 MHz, DMSO-d_6) δ 7.59-7.55 (m, 2H, H-b); 7.51-7.48 (m, 2H, H-b); 7.45-7.41 (m, 1H, H-4''); 7.40-7.35 (m, 6H, H-2'' H-c); 7.29-7.19 (m, 8H, H-2''', H-3'', H-d); 6.88-6.83 (m, 4H, H-3'''); 6.25 (dd, 1H, J = 8.45, 6.10 Hz, H-1'); 4.44-4.38 (m, 1H, H-3'); 3.98-4.02 (m, 1H, H-5); 3.92-3.85 (m, 1H, H-4'); 3.70 (s, 6H, OCH_3); 3.62-3.57 (m, 2H, CH_2CH$_2$CN); 3.57-3.55 (m, 1H, H-7); 3.54-3.43 (m, 2H, NCHCH$_3$); 3.29 (bs, 1H, H-4); 3.19-3.09 (m, 2H, H-5'); 2.73, 2.62 (2t, J = 5.90 Hz, CH$_2$CH_2CN); 2.21-2.15 (m, 1H, H-7); 2.08-2.10 (m, 1H, H-2'a); 1.97-1.86 (m, 1H, H-2'b); 1.69-1.59 (m, 2H, H-3); 1.58-1.52 (m, 2H, H-6); 1.19-1.06 (m, 12H, NCHCH_3); 0.96 (s, 9H, CH$_3$).

^{13}C NMR (125.7 MHz, DMSO-d_6) δ 174.86 (1C, C2); 158.12, 158.08 (2C, C4'''), 144.57 (1C, C1''); 135.53, 135.34 (2C, Ca); 135.14, 135.09 (4C, Cb); 133.49, 133.23 (2C, C1'''); 129.84, 129.81, 129.69 (4C, C2'''); 129.60, 129.56 (2C, C2''); 127.78, 127.72 (2C, Cd); 127.98, 127.78 (4C, Cc); 126.66 (1C, C4''); 118.89, 118.72 (1C, CN); 113.11 (4C, C3'''); 85.65, 85.63 (1C, C-Ar$_3$); 82.71, 82.36 (1C, C4'); 82.31 (1C, C1'); 73.8 (1C, C3'); 69.97 (1C, C5); 63.29 (1C, C5'); 58.37, 58.28, 58.22, 58.16 (2C, C-CH$_3$); 54.99 (2C, OCH$_3$); 44.51, 44.66 (CH$_2$CH$_2$CN); 42.54, 42.47, 42.44 (2C, NCHCH$_3$); 36.95 (1C, C6); 36.18 (1C, C7); 36.19 (1C, C2'); 36.02 (1C, C3); 30.73 (1C, C4); 26.70 (3C, C-CH$_3$); 24.32, 24.29, 24.23, 24.17, 24.10 (NCHCH$_3$); 19.80, 19.74 (1C, CH$_2$$CH_2$CN).

^{31}P NMR (202.5 MHz, DMSO-d_6, ref. 85% H_3PO_4) δ 147.10, 146.63 in ~1:1 ratio.

HRMS (ESI) m/z: $[M+H]^+$ Calcd for $C_{57}H_{73}N_3O_8PSi$ 986.4899; found 986.5901.



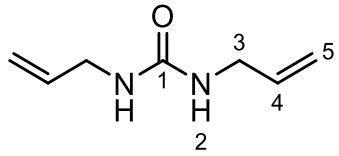
R_f = 0.41 in 10% acetone in DCM

^{1}H NMR (500 MHz, DMSO-d_6) δ 7.57-7.55 (m, 2H, H-b); 7.53-7.51 (m, 2H, H-b); 7.45-7.41 (m, 1H, H-4''); 7.40-7.30 (m, 7H, H-2'', H-c, H-d); 7.25-7.17 (m, 6H, H-2''', H-3''); 6.85-6.81 (m, 4H, H-3'''); 6.23 (dd, 1H, J = 8.00, 6.00 Hz, H-1'); 4.44-4.38 (m, 1H, H-3'); 3.95 (bs, 1H, H-5); 3.87-3.81 (m, 1H, H-4'); 3.70 (s, 6H, OCH_3); 3.61-3.57 (m, 2H, CH_2CH$_2$CN); 3.56-3.44 (m, 2H, NCHCH$_3$); 3.18-3.11 (m, 2H, H-5'); 3.10-3.05 (1H, H-4); 2.75, 2.70 (bs, 1H, H-7); 2.62 (2t, J = 5.84 Hz, CH$_2$CH_2CN); 2.28-2.23 (1H, H-7); 2.14-2.08 (m, 1H, H-2'a); 2.03-1.92 (m, 1H, H-2'b); 1.73-1.68 (m, 1H, H-3); 1.64-1.57 (m, 2H, H-6, H-3); 1.54-1.47 (m, 1H, H-6); 1.12-1.08 (m, 9H, NCHCH_3); 1.00 (s, 9H, CH$_3$); 0.97, 0.96 (2s, NCHCH_3).

^{13}C NMR (125.7 MHz, DMSO-d_6) δ 174.61 (1C, C2); 158.09 (2C, C4'''), 144.73 (1C, C1''); 135.35 (2C, Ca); 135.18, 135.14 (4C, C2b); 133.44, 133.40 (2C, C1'''); 129.85, 129.82, 129.69 (4C, C2'''); 129.68, 129.63 (2C, C2''); 127.79 (2C, Cd); 127.74 (4C, Cc); 127.53 (1C, C4''); 118.72 (1C, CN); 113.11 (4C, C3'''); 85.55 (1C, C-Ar$_3$); 82.62 (1C, C4'); 82.31 (1C, C1'); 73.8 (1C, C3'); 69.97 (1C, C5); 63.29 (1C, C5'); 58.37, 58.28, 58.22, 58.16 (1C, NCH$_2$CH$_2$CN); 54.97 (2C, OCH$_3$); 44.51, 44.66 (CH$_2$CH$_2$CN); 42.55, 42.48, (2C, NCHCH$_3$); 36.95 (1C, C6); 36.19 (1C, C7); 36.02 (1C, C3); 30.73 (1C, C4); 26.77 (3C, C-CH3); 24.35, 24.29, 24.25, 24.19, 24.11 (4C, NCHCH$_3$); 19.75 (1C, CH$_2$$CH_2$CN).

^{31}P NMR (202.5 MHz, DMSO-d_6, ref. 85% H_3PO_4) δ 147.17, 146.72 in ~1:1 ratio.

4.4.11. Synthesis of diallyl urea (42)[202]



To a stirring solution of allylamine (1 g, 17 mmol) in dry THF (15 mL), *N*,*N*-disuccinimidyl carbonate (2.25 g, 9 mmol) was added at r.t. and the mixture was stirred at r.t. for 2 hr under argon. After consumption of the starting amine (TLC analysis R_f = 0.35, 50% EtOAc in

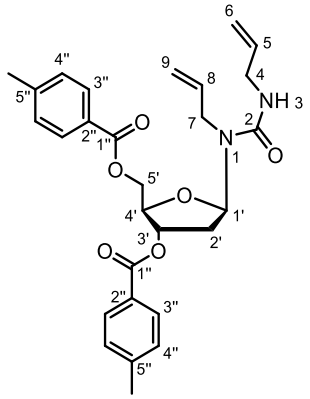
hexane), solvent was evaporated *in vacuo*. The crude product was purified by column chromatography over silica gel eluting with 0-50% EtOAc in hexane to afford the desired product as a white solid (R_f = 0.29, 50% EtOAc in hexane; yield 1.9 g, 77%).

^{1}H NMR (500 MHz, $CDCl_3$) δ 5.84-5.76 (m, 2H, H-4); 5.85 (bs, 2H, NH); 5.14 (dd, 2H, J = 17.20, 1.55 Hz, H-5a); 5.04 (dd, 2H, J = 10.25, 1.55 Hz, H-5b); 3.75-3.72 (m, 4H, H-3).

^{13}C NMR (125.7 MHz, $CDCl_3$) δ 158.68 (1C, C2); 135.59 (2C, C4); 115.60 (2C, C5); 43.00 (2C, C3).

HRMS (ESI) m/z: $[M+H]^+$ Calcd for $C_7H_{13}N_2O$ 141.1022; found 141.1021.

4.4.12. Synthesis of 1-[2-deoxy-3,5-bis-*O*-(4-methylbenzoyl)-β-D-*erythro*-pentofuranosyl]diallylurea (43)



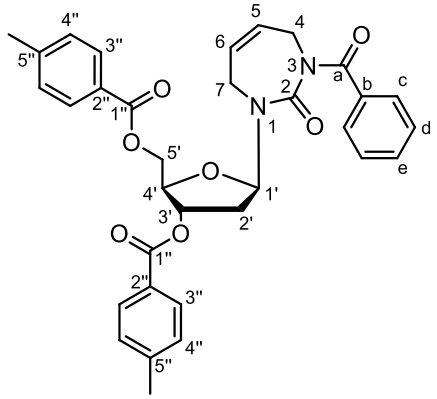
To the solution of the compound **42** (2.0 g, 14.0 mmol) in benzene (30 mL) NEt_3 (3.38 mL, 24 mmol) and trimethylsilyl chloride (2.2 mL, 17 mmol) were sequentially added at r.t. under argon. The reaction mixture was stirred overnight, and then filtered over sintered funnel. Solvents were evaporated *in vacuo* to afford silyl protected diallyl urea derivative. To the solution of silyl-protected diallyl urea (2.5 g, 9.0 mmol) in dichloroethane (60 mL) freshly distilled $SnCl_4$ (3.36 mL, 28 mmol) and Hoffer's chlorosugar (3.36 g, 8.6 mmol) were sequentially added at -35 °C. Reaction mixture was stirred for 1.5 h at -35 °C and after the consumption of starting silylated diallyl amine, pyridine (10 mL) and H_2O (50 mL) were added and reaction mixture was stirred at r.t. for 1h followed by addition of H_2O (100 mL). The resulting mixture was extracted with DCM (4 × 100 mL), combined organic layers were dried over Na_2SO_4, filtered and the combined organic fractions were evaporated *in vacuo*. The crude product was purified by flash chromatography on silica eluting with 0-30% EtOAc in hexane to afford the product as a foam (R_f = 0.38, 30% EtOAc in hexane; yield 2 g, 45%, β/α = 9:1).

^{1}H NMR (500 MHz, $CDCl_3$) δ 7.93-7.89 (m, 4H, H-3"); 7.26-7.21 (m, 4H, H-4"); 6.42 (dd, 1H, J = 8.70, 6.85 Hz, H-1'); 5.87-5.76 (m, 2H, H-5, H-8); 5.50 (dt, 1H, J = 8.70, 6.85

Hz, H-3'); 5.26-5.22 (m, 2H, H-9); 5.14-5.12 (m, 1H, H-6a); 5.09-5.08 m, 1H, H-6b); 4.90 (t, 1H, *J* = 5.28 Hz, H-3); 4.61 (dd, 1H, *J* = 11.90, 3.50 Hz, H-5'a); 4.55 (dd, 1H, *J* = 11.90, 3.64 Hz, H-5'b); 4.36 (m, IH, H-4'); 3.99-3.95 (m, 1H, H-4a); 3.86-3.84 (m, 2H, H-7), 3.80-3.77 (m, 1H, H-4b); 2.43 (2s, 6H, CH_3); 2.26-2.24 (m, 2H, H-2').

^{13}C NMR (125.7 MHz, $CDCl_3$) δ 166.22 (1C, C1"); 166.16 (1C, C1"); 157.89 (1C, C2); 144.26 (1C, C5"); 144.02 (1C, C5"); 135.15 (1C, C2"); 135.02 (1C, C2"); 129.78 (2C, C3"); 129.64 (2C, C3"); 129.21 (4C, C4"); 126.93 (1C, C2"); 126.66 (1C, C2"); 116.48 (1C, C9); 115.58 (1C, C6); 85.60 (1C, C1'); 80.31 (1C, C4'); 75.07 (1C, C3'); 64.67 (1C, C5'); 44.25 (1C, C4); 43.10 (1C, C7); 35.15 (1C, C2'); 21.72, 21.69 (2C, CH_3).

4.4.13. Synthesis of 3-[2-deoxy-3,5-bis-*O*-(4-methylbenzoyl)-β-D-*erythro*-pentofuranosyl]-1-benzoyl-1,3,4,7-tetrahydro-2*H*-1,3-diazepin-2-one (45)



To a stirring solution of the compound **43** (4 g, 8.0 mmol) in pyridine (100 mL) at 0 ^{o}C NEt_3 (4.8 mL, 24 mmol) and benzoyl chloride (2.83 mL, 24 mmol) were added sequentially, and the reaction mixture was allowed to stir overnight under r.t. After the consumption of starting material by observing TLC, pyridine was evaporated *in vacuo.* The residue was resuspended in 100 mL ethyl acetate, washed with brine (2 × 10 mL), the organic layer was dried over anhydrous sodium sulfate, filtered, and concentrated *in vacuo*. The crude product was purified by column chromatography over silica gel eluting with 0-30% EtOAc in hexane to afford the desired product as a foam (R_f = 0.42, 30% EtOAc in hexane; yield 4 g, 82%).

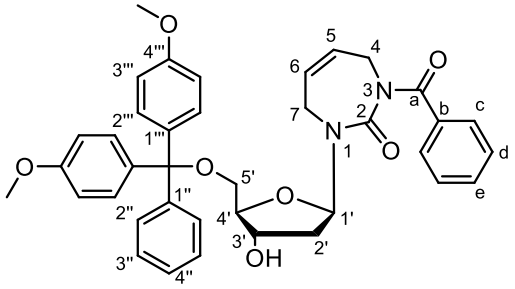
To a solution of the benzoyl-protected compound **44** (5 g, 8.0 mmol), in dry dichloromethane was added GreenCat^{TM} (10 mol%) and refluxed for 2 hr. After the consumption of the starting material, solvent was evaporated *in vacuo* and the crude product was purified by silica gel column chromatography eluting with 0-40% EtOAc in hexane to yield the desired product as a foam (R_f = 0.39, 40% EtOAc in hexane; yield 1.9 g, yield 52%). The material was resuspended and washed with methanol to afford pure β product in >99% purity.

^{1}H NMR (500 MHz, $CDCl_3$) δ 7.98-7.95 (m, 2H, H-3''); 7.89-7.86 (m, 2H, H-3''); 7.56-7.53 (m, 2H, H-c); 7.47-7.44 (m, 1H, H-e); 7.40-7.35 (m, 2H, H-d); 7.31-7.27 (m, 2H, H-4''); 7.23- 7.20 (m, 2H, H-4''); 6.19 (dd, 1H, *J* = 9.25, 5.25 Hz, H-1'); 5.82-5.76 (m, 1H, H-5); 5.70-5.65 (m, 1H, H-6); 5.55 (m, 1H, H3'); 4.74 (dd, 1H, *J* = 15.35, 12.05 Hz, H-5'a); 4.68-4.62 (m, 1H, H7); 4.57 (dd, 1H, *J* = 15.60, 12.05 Hz, H-5'b); 4.35 (td, 1H, *J* = 3.27, 2.45 Hz, H-4'); 4.30-4.23 (m, 1H, H-7); 4.08 (dd, 1H, *J* = 19.15, 16.85 Hz, H-4); 3.96 (dd, 1H, *J* = 21.90, 16.85 Hz, H-4); 2.44, 2.40 (2s, 6H, CH_3); 2.31 (ddd, 1H, *J* = 16.00, 6.70, 5.25 Hz, H-2'b); 22.18 (ddd, 1H, *J* = 15.35, 9.25, 6.70 Hz, H-2'a).

^{13}C NMR (125.7 MHz, $CDCl_3$) δ 171.04 (1C, Ca); 166.22, 166.19 (2C, C1''); 159.70 (1C, C2); 144.52, 144.44 (2C, C5''); 135.21 (1C, C2''); 131.61 (1C, Ce); 129.85, 129.67 (4C, C3''); 129.51, 129.33 (4C, C4''); 128.56 (2C, Cd); 128.30 (1C, C5); 127.37 (2C, Cc); 126.98, 126.52 (2C, C1''); 124.02 (1C, C6); 85.64 (1C, C1'); 81.52 (1C, C4'); 74.93 (1C, C3'); 64.33 (1C, C5'); 43.27 (1C, C7); 39.29 (1C, C4); 35.43 (1C, C2'); 21.84, 21.81 (2C, CH_3).

HRMS (ESI) *m*/*z*: $[M+Na]^+$ Calcd for $C_{33}H_{32}N_2NaO_7$ 591.2102; found 591.2097.

4.4.14. Synthesis of 1-[2-deoxy-5-*O*-(4,4'-dimethoxytrityl)-β-D-*erythro*-pentofuranosyl]-3-benzoyl-1,3,4,7-tetrahydro-2*H*-1,3-diazepin-2-one (46)



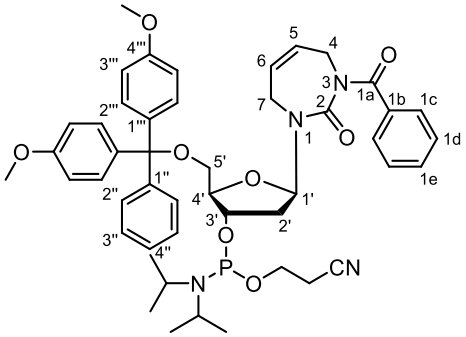
To a stirring solution of the compound **45** (4 g, 7.0 mmol) in 400 mL methanol was added 40 mL of 30% aq. ammonia solution and kept stirring for 3 days. After the disappearance of the starting material, volatiles were removed by rotary vacuum evaporator and residue was co-evaporated again with 100 mL water to remove formed methyl toluate and then freeze dried from water 50 mL to get the deprotected compound (2.25 g). This product was used without further purification to protect it with DMT. To a stirring solution of the deprotected product (2 g, 6.0 mmol) in dry pyridine (40 mL) at 0 °C 4,4'-dimethoxytrityl chloride (3.72 g, 11 mmol) was added and mixture was stirred at r.t. overnight. Pyridine was evaporated *in vacuo*. The residue was dissolved in 50 mL ethyl acetate and washed with brine (2 × 10 mL). The organic layer was dried over anhydrous sodium sulfate, filtered, and evaporated *in vacuo*. The crude product was purified by column chromatography over silica gel treated with 10% Et_3N in DCM and compound was eluted with 0-20% acetone in DCM to afford the desired product **46** as a foam (3.13 g, 80%; R_f = 0.40, 20% acetone in DCM).

^{1}H NMR (500 MHz, $CDCl_3$) δ 7.57-7.54 (m, 2H, H-c); 7.46- 7.42 (m, 3H, H-2", H-e); 7.40-7.36 (m, 2H, H-d); 7.34-7.30 (m, 6H, H-3"', H3"); 7.27-7.24 (m, 1H, H-4"); 6.87-6.83 (m, 4H, H-3"'); 6.05 (app t, 1H, *J* = 7.05 Hz, H-1'); 5.79- 5.75 (m, 1H, H-5); 5.70- 5.66 (m, 1H, H-6); 4.62- 4.55 (m, 1H, H-7); 4.43 (td, 1H, *J* = 4.75, 2.80 Hz, H-4'); 4.40-4.34 (m, 1H, H-7); 4.12 (dd, 1H, *J* = 19.35, 16.80 Hz, H-4); 3.96 (dd, 1H, *J* = 21.65, 16.80 Hz, H-4); 3.87 (td, 1H, *J* = 3.75, 2.20 Hz, H-3'); 3.80 (s, 6H, OC*H*$_3$); 3.40-3.31 (m, 2H, H-5'); 2.05 (dd, 2H, *J* = 6.85, 5.20 Hz, H-2').

^{13}C NMR (125.7 MHz, $CDCl_3$) δ 170.82 (1C, Ca); 159.34 (1C, C2); 158.74 (2C, C4"'); 144.72 (1C, C1"); 135.82, 135.81 (2C, C1"'); 135.28 (1C, Cb); 131.51 (1C, Ce); 130.21, 130.19 (4C, C2"'); 128.51 (2C, C3"); 128.27 (2C, C2"); 128.05 (1C, C5); 128.02 (2C, Cd); 127.44 (2C, Cc); 127.13 (1C, C4"); 124.32 (1C, C6); 113.32 (4C, C3"'); 86.62 (1C, *C*-Ar$_3$); 85.36 (1C, C1'); 84.72 (1C, C4'); 72.45 (1C, C3'); 63.80 (1C, C5'); 55.37 (2C, O*C*H$_3$); 43.18 (1C, C7); 39.65 (1C, C4) ; 38.39 (1C, C2').

HRMS (ESI) *m*/*z*: $[M+Na]^+$ Calcd for $C_{33}H_{32}N_2NaO_7$ 591.2102; found 591.2097.

4.4.15. Synthesis of 1-[3-*O*-(*N*,*N*-di*iso*propylamino-2-cyanoethoxyphosphanyl)-2-deoxy-5-*O*-(4,4'-dimethoxytrityl)-β-D-*erythro*-pentofuranosyl]-3-benzoyl-1,3,4,7-tetrahydro-2*H*-1,3-diazepin-2-one (47)



To a stirring solution of 5'-*O*-DMT protected compound **46** (0.20 g, 0.36 mmol) in dry DCM (10 mL) under argon at r.t., were added Et_3N (0.07 mL, 0.50 mmol) followed by 2-cyanoethyl *N*,*N*-di*iso*propyl chlorophosphoramidite (0.09 g, 0.38 mmol). After the disappearance of the starting material in 10 min the reaction mixture was washed with saturated sodium bicarbonate solution (2 × 5 mL) followed by brine (5 mL). The organic layer was dried over anhydrous sodium sulfate column, filtered and the combined fractions were evaporated *in vacuo*. The crude product was purified by column chromatography over silica gel saturated with Et_3N (10%) in DCM and eluting with DCM/acetone (9:1) to give the desired product as a white foam (R_f = 0.45, 10% acetone in DCM; yield 0.32 g, 92%).

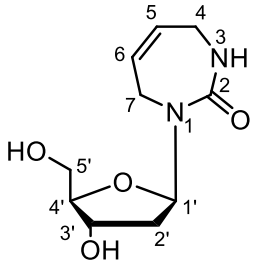
^{1}H NMR (500 MHz, $CDCl_3$) δ 7.62-7.59 (m, 2H, H-c); 7.51-7.46 (m, 3H, H-2", H-e); 7.45- 7.41 (m, 2H, H-d); 7.38-7.31 (m, 6H, H-2"', H-3"); 7.29-7.26 (m, 1H, H-4"); 6.87-6.83 (m, 4H, H-3"'); 6.10 (app t, 1H, *J* = 7.50 Hz, H-1'); 5.83-5.79 (m, 1H, H-5); 5.74-5.68 (m, 1H, H-6); 4.68- 4.63 (m, 1H, H-7); 4.62-4.57 (m, 1H, H-4'); 4.46-4.35 (m, 1H, H-7); 4.25-4.15 (m, 1H, H-4); 4.13- 4.08 (m, 1H, H-4); 4.05, 4.02 (2td, 1H, *J* = 3.50, 1.80 Hz, H-3'); 3.82, 3.83 (2s, 6H, OCH_3); 3.75- 3.67 (m, 1H, NC*H*CH_3); 3.66-3.60 (m, 1H, NC*H*CH_3); 3.58-3.51 (m, 2H, 2H, C*H*$_2$$CH_2CN$); 3.47, 3.43 (2dd, 1H, *J* = 10.35, 2.85 Hz, H-5'a); 3.31-3.26 (m, 1H, H-5'b); 2.60, 2.40 (2t, 2H, *J* = 7.50 Hz, CH_2C*H*$_2$CN); 2.27-2.10 (m, 2H, H-2'); 1.28-1.21 (m, 2H, C-CH_3); 1.19-1.13 (m, 8H, C-CH_3); 1.07-1.03 (m, 2H, C-CH_3).

^{13}C NMR (125.7 MHz, $CDCl_3$) δ 170.88, 170.83 (1C, Ca); 159.42, 159.39 (1C, C2); 158.74, 158.72 (2C, C4'"); 144.75, 144.73 (1C, C1"); 135.88, 135.82 (2C, C1'"); 135.37 (1C, Cb); 131.52 (1C, Ce); 130.29, 130.26, 130.24 (4C, C2'"); 128.56, 128.53 (2C, C2"); 128.39, 128.34 ((2C, Cd),); 128.05 (1C, C5); 127.99 (2C, C3"); 127.46 (2C, Cc); 127.12; 127.08 (1C, C4"); 124.46, 124.42 (1C, C6); 117.72, 117.53 (1C, CN); 113.27 (4C, C3'"); 86.50, 86.49 (1C, *C*-Ar$_3$); 85.51, 85.46 (1C, C1'); 84.45, 84.42, 84.23, 84.18 (1C, C4'); 73.78, 73.65, 73.32, 73.20 (1C, C3'); 63.24, 63.09 (1C, C5'); 58.30, 58.26, 58.16, 58.11 (2C, *C*-CH3); 55.39, 55.37 (2C, O*C*H_3); 43.43, 43.30 (1C, *C*H_2CH_2CN); 43.22, 43.17 (1C, C7); 39.66, 39.63 (1C, C4); 37.60, 37.56 (1C, C2'); 24.74, 24.68, 24.64, 24.60, 24.58, 24.54 (4C, NCH*C*H_3); 20.54, 20.48, 20.32, 20.26 (1C, CH_2*C*H_2CN).

^{31}P NMR (202.5 MHz, $CDCl_3$ ref. 85% H_3PO_4) δ 148.48, 148.25 in ~1:1 ratio.

HRMS (ESI) *m/z*: $[M+H]^+$ Calcd for $C_{47}H_{56}N_4O_8P$ 835.3830; found 835.3836.

4.4.16. Synthesis of 1-[2-deoxy-β-D-erythro-pentofuranosyl]- 1,3,4,7-tetrahydro-2H-1,3-diazepin-2-one (48)

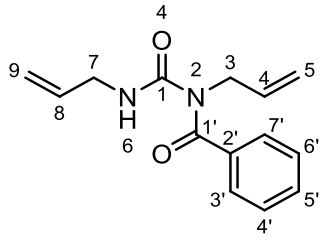


To a stirring solution of the compound **45** (4 g, 7.0 mmol) in 400 mL methanol was added 40 mL of 30% aq. ammonia solution and kept stirring for 3 days. After the disappearance of the starting material, volatiles were removed by rotary vacuum evaporator and residue was co-evaporated again with 100 mL water to remove formed methyl toluate and then freeze dried from water (50 mL) to get the compound 1-[2-deoxy-β-D-erythro-pentofuranosyl]-1,3,4,7-tetrahydro-2H-1,3-diazepin-2-one (2.25 g). To 500 mg of this product was added 3

mL of 30% aq. NH_3 solution and the mixture was kept at rt, for 15 min. After the disappearance of the starting material as evidenced by TLC (R_f = 0.45, 40% MeOH in DCM), ammonia was evaporated *in vacuo* and the crude product was purified by preparative TLC using 20% MeOH in DCM as eluent to afford the required product as a foam (R_f = 0.42, 30% MeOH in DCM; yield: 0.2 g 60%).

^{1}H NMR (500 MHz, D_2O) δ 6.97 (dd, 1H, J = 8.51, 6.62 Hz, H-1'), 5.93- 5.86 (m, 2H, H-5 and H-6); 4.36-4.32 (m, 1H, H-3') 3.86- 3.75 (m, 4H, H-3 and H-7), 3.75- 3.67 (m, 2H, H-5'), 2.20 (ddd, 1H, J = 15.10, 8.20, 7.25 Hz, H-2'b); 2.03 (ddd, 1H, J = 15.10, 6.60, 3.15 Hz, H-2'a).

4.4.17. Synthesis of *N*-allyl-*N*-(allylcarbamoyl)benzamide (49)



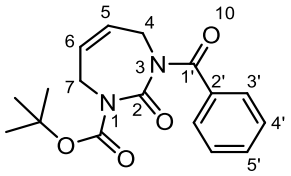
To a stirring solution of the compound **42** (2 g, 14.0 mmol) in pyridine (50 mL) at 0 ^oC, NEt_3 (5.96 mL, 42 mmol) and benzoyl chloride (4.95 mL, 42 mmol) were added sequentially, and the reaction mixture was allowed to stir overnight at r.t. After the consumption of the starting material by observing TLC, pyridine was evaporated *in vacuo.* The residue was resuspended in 50 mL ethyl acetate, washed with brine (2 × 10 mL), organic layer was dried over anhydrous sodium sulfate, filtered, and concentrated *in vacuo*. The crude product was purified by column chromatography over silica gel eluting with 0-30% EtOAc in hexane to afford the desired product as a foam (R_f = 0.38, 30% EtOAc in hexane; yield 3 g, 85%).

^{1}H NMR (500 MHz, $CDCl_3$) δ 9.19 (1H, NH); 7.48-7.45 (m, 1H, H-5'); 7.44-7.40 (m, 4H, H-3', H-4'); 5.96-5.88 (m, 1H, H-8); 5.84-5.76 (m, 1H, H-4); 5.26 (dd, 1H, J = 17.02, 1.26 Hz, H-9a); 5.16 (dd, 1H, J = 10.25, 1.26 Hz, H-9b); 5.11 (dd, 1H, J = 10.40, 1.26 Hz, H-5b); 4.96 (d, 1H, J = 17.33, 1.26 Hz, H-5a); 4.33-4.29 (m, 2H, H-3); 4.01-3.97 (m, 2H, H-7).

^{13}C NMR (125.7 MHz, $CDCl_3$) δ 175.20 (1C, C1'); 154.53 (1C, C1); 136.17 (1C, C2'); 133.98 (1C, C8); 133.95 (1C, C4); 133.65 (1C, C5'); 130.50 (1C, C6'); 130.17 (1C, C7'); 128.48 (2C, C3'); 126.07 (2C, C4'); 116.49 (1C, C5); 116.16 (1C, C9); 49.07 (1C, C3); 43.01 (1C, C7).

HRMS (ESI) *m/z*: $[M+Na]^+$ Calcd for $C_{14}H_{16}N_2O_2Na$ 267.1104; found 267.1098.

4.4.18. Synthesis of 1-(*tert*-butyloxy carbonyl)-3-benzoyl-1,3,4,7-tetrahydro-2*H*-1,3-diazepine-2-one (51)



To a stirring solution of compound **49** (4.0 g, 16.4 mmol) in 100 mL of dry THF at 0 ^{o}C was added Et_3N (3 mL, 22.0 mmol) and stirred at the same temperature for 0.5 h, then boc anhydride (5.70 mL, 25 mmol) was added dropwise. The reaction mixture was refluxed overnight under argon. The reaction mixture was diluted with EtOAc (3 × 30 mL) and washed with water (10.0 mL). The combined extracts were washed with brine (2 ×10 mL), the organic layer was dried over anhydrous sodium sulfate, filtered, concentrated *in vacuo*, and purified by silica gel column chromatography (10% EtOAc in hexane) to afford the desired compound **50** as an oil (R_f = 0.42, 10% EtOAc in hexane: Yield 8.6 g, 89%).

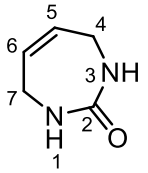
To a solution of the compound **50** (4 g, 11.6 mmol), in dry DCM (100 mL) was added GreenCat^{TM} (10 mol%) and the reaction mixture was refluxed for 2 hr. After the consumption of the starting material, solvent was evaporated *in vacuo* and the crude product was purified by silica gel column chromatography eluting with 0-40% EtOAc in hexane to yield the desired product as a foam (R_f = 0.40, 30% EtOAc in hexane; yield 2.65 g, 72%).

^{1}H NMR (500 MHz, $CDCl_3$) δ 7.61-7.57 (m, 2H, H-3'); 7.52-7.49 (m, 1H, H-5'); 7.45-7.41 (m, 2H, H-4'); 5.90 (s, 2H, H5 and H6); 4.61-4.55 (m, 4H, H-4 and H-7); 1.48 (s, 9H, CH_3).

^{13}C NMR (125.7 MHz, $CDCl_3$) δ 171.19 (1C, C1'); 157.30 (1C, C2'); 131.76 (1C, C5'); 128.47 (2C, C3'); 127.60 (2C, C4'); 126.28 (1C, C5); 126.05 (1C, C6); 83.50 (1C, *C*-CH_3); 43.82 (1C, C7); 42.31 (1C, C4); 27.97 (3C, CH_3).

HRMS (ESI) *m*/*z*: $[M+Na]^+$ Calcd for $C_{17}H_{20}N_2O_4Na$ 339.1315; found 339.1305.

4.4.19. Synthesis of 1,3,4,7-tetrahydro-2*H*-1,3-diazepin-2-one (53)



To a stirring solution of **51** in methanol (100 mL) was added 10 mL of 30% aq. ammonia. After stirring at r.t. for 10 min, the starting material was completely consumed as evidenced by TLC. The volatiles were evaporated *in vacuo* and the residue was dissolved in 20 mL

DCM. Trifluoroacetic acid (10 mL) was added at r.t., and kept stirring for 10 min. The reaction mixture was concentrated *in vacuo* and co evaporated several times with DCM. The crude product was purified by silica gel column chromatography, eluting with 0-10% MeOH in DCM to afford the desired product as an off-white solid (R_f = 0.29, 10% MeOH in DCM; yield 0.9 g, 85%).

^{1}H NMR (500 MHz, DMSO-d_6) δ 5.99 (bs, 2H, NH); 5.77 (s, 2H, H5, H6); 3.53 (s, 4H, H4, H7).

^{13}C NMR (125.7 MHz, $CDCl_3$) δ 164.90 (1C, C2); 127.84 (2C, C4, C7); 40.80 (2C, C5, C6).

HRMS (ESI) *m*/*z*: $[M+H]^+$ Calcd for $C_5H_9N_2O$ 113.0709; found 113.0709.

4.4.20. Oligonucleotide synthesis and purification

Oligonucleotides were prepared on a MerMade-4 DNA/RNA synthesizer (BioAutomation) on a 5 μmol scale using standard manufacturer's protocol. Coupling times of modified phosphoramidites were increased from 2 to 10 min. The final detritylated oligos were cleaved from the solid support and deprotected at r.t. using conc. NH_4OH. The deprotected oligos in solution were freeze-dried and dry pellets were dissolved in milli-Q water (1 mL) and purified and isolated by reverse-phase HPLC on 250/4.6 mm, 5 μm, 300 Å C18 column (Thermo Fisher Scientific) in a gradient of CH_3CN (0→20% for 20 min, 1.3 mL/min) in 0.1 M TEAA buffer (pH 7.0) with a detection at 260 nm. For the deprotection of the TBDPS protecting group in oligos, the dry pellets were dissolved in a solution of 100 μL THF, followed by 200 μL of 1 M TBAF in THF. The mixture was kept at 22 °C for 30 min. The reaction mixture was quenched by the addition of 2 M TEAA buffer (pH 7.0) and oligos were purified by reverse-phase HPLC on 250/4.6 mm, 5 μm, 300 Å C18 column (Thermo Fisher Scientific) in a gradient of CH_3CN (0→80% for 14 min, 1.3 mL/min) in 0.1 M TEAA buffer (pH 7.0) with a detection at 260 nm. Oligonucleotides were freeze-dried, pellets were dissolved in milli-Q water (1.5 mL) and desalted by reverse-phase HPLC on 100/10 mm, 5μm, 300 Å C18 column (Phenomenex) in a gradient of CH_3CN (0→80% for 15 min, 5 mL/min) in milli-Q water with detection at 260 nm. Pure products were quantified by measuring absorbance at 260 nm, analyzed by ESI-MS and concentrated by freeze-drying.

4.4.21. Qualitative evaluation of inhibitors of A3B$_{CTD}$-catalyzed deamination using ^{1}H NMR assay

Competitive inhibition of the synthesized oligonucleotides was performed by a similar procedure reported by us previously.[54, 124, 126] A series of ^{1}H NMR spectra was recorded in a

similar fashion as that of substrate analysis. Here, we used 400 μM of a standard 7-mer oligonucleotide substrate 5'-T_4**dC**AT, 36 μM of inhibitor containing cross-linked oligos, 300 or 200 nM of $A3B_{CTD}$-QM-ΔL3-AL1swap in a buffer containing 50 mM sodium phosphate (pH 6.0), 100 mM NaCl, 2.5 mM β-mercaptoethanol, 50 μM TSP. Deamination activity was monitored by the same method as for DNA substrates.

Chapter 5. Summary and Future Directions

5.1. Summary

A3-induced mutagenic signatures from various cancer cell lines suggest involvement of A3A as well as A3B in driving cancer evolution.[7-8, 13, 27, 66, 94, 96] It has been proven that these mutations are one of the important factors in the development of drug resistance and cancer metastasis leading to poor treatment outcomes.[8-9, 27, 87-89, 96, 203-205] The selective inhibition of both A3A and A3B may be key in halting A3-induced cancer mutagenesis, while keeping other A3 enzymes active to fight against various viruses. Therefore, we focused on the design of A3A/A3B selective inhibitors which can be potentially used as a conjuvant in cancer chemo and immuno-therapies making them effective for the longer periods of time which should result in higher rates of remission. This research capitalized on our previous work on substrate-like inhibitors of A3 enzymes and the information from various crystal and NMR structural studies suggesting an unusual U-shaped conformation of ssDNA upon binding to A3.[36, 54, 60, 124, 126, 140] This work provided us with the first low nano-molar inhibitor of A3A/A3B by a cross-linking of ssDNA. We also tried to synthesize novel transition-state analogues of cytosine deamination for their potential use as A3 inhibitors. Unfortunately, this methodology did not provide any improvement to inhibition of A3 enzymes yet.

In **Chapter 3** we created a series of cross-linked oligonucleotides as substrates and inhibitors of A3 enzymes by CuAAC. The methodology was inspired from previous research on the synthesis of terminally cross-linked oligonucleotides using CuAAC.[206] The internal cross-linking was challenging due to the difficulties associated with incorporation of an azido-modified phosphoramidite in the ssDNA. Here we described a facile synthesis of a series of azido-modified nucleoside phosphoramidites and their incorporation in the DNA sequence using an automated DNA synthesizer. The DNA sequences synthesized with the incorporation of various alkyne-and azido-modified nucleosides were cross-linked with CuAAC efficiently. The proof of a successful cross-linking reaction since molecular masses of both linear and cross-linked materials are the same presented us an opportunity to use NMR spectrometry in analyzing both linear and cross-linked materials. NMR provided us sufficient evidence of a successful cross-linking reaction, which along with reverse-phase HPLC, were quite useful in monitoring the cross-linking reaction.

The substrate activity of our cross-linked oligos was analyzed by the previously described real-time NMR assay.[54, 124] In line with our rationale of incorporating a transition-state analogue (dZ/FdZ) in a substrate instead of the target dC provided us superior inhibitors. The cross-link number 3 is the first nano-molar potent inhibitor of A3. The overall inhibition

effect was improved from 20-fold (**dZ**-linear) and 35-fold (**FdZ**-linear) to more than 1500-fold for FdZ[C^{N3}(-2),H^{E}(+1)]X. When this best inhibitor was evaluated against the preferred substrate of $A3B_{CTD}$, dC-hairpin, a similar inhibition effect of ~1000-fold was observed. To design an inhibitor, which is more drug-like, we also synthesized short 3-and 4-mer oligos as inhibitors. The reduced inhibitory potential of these oligos as evidenced from real-time NMR assays suggests a loss of the important interactions between nucleotides (originally present in the 9-mer sequence) and amino-acid residues of $A3B_{CTD}$.

The differential inhibition of dZ vs FdZ as shown by us in recent research, prompted us to synthesize novel inhibitors of CDA.[124] This was based on the rationale that incorporating transition-state analogues of cytosine deamination in the place of target dC in a preferred ssDNA leads to inhibition of the A3 enzyme. We reasoned of a better inhibition of A3 when a more potent CDA inhibitor was incorporated in the sequence. In **Chapter 4** we successfully synthesized and characterized a 2'-deoxy analogue of a reported CDA inhibitor (**ddiazep**) and two new compounds based on seven-membered ring nucleobase (***R***-and ***S***-**dazep**). We developed a facile route for the synthesis of these nucleoside phosphoramidites and incorporated them in an A3-preferred linear ssDNA. Unfortunately, the inhibition of A3 enzymes observed by real-time NMR assays was comparatively less than that by a standard linear inhibitor containing dZ. However, there were differences in the level of inhibition by the ***R*** and ***S*** isomers. The ***S*** isomer of **dazep** produced a higher level of inhibition compared to the ***R*** isomer suggesting that the relative orientation of the hydroxyl group is important in the inhibition of A3.

5.2. Future directions

The low nanomolar inhibition of A3 exhibited by our cross-link 3 encouraged us to develop inhibitors based on cross-linked ssDNAs as potential drug candidates to enhance and prolong the effectiveness of cancer chemo and immuno-therapies. This goal can be achieved by a further structural optimization of the cross-links and addressing the innate hurdles of using oligonucleotides having natural nucleotides in the sequence. For example, poor cellular uptake and degradation by nucleases are typical for oligonucleotide drugs before reaching a target organ/tissue.

5.2.1. Evaluation of inhibitors with wild-type A3A and A3B

Our substrate activity, inhibition, and binding experiments (cross-linked oligos containing dZ/FdZ) relied on the enzyme $A3B_{CTD}$-QM-ΔL3-AL1swap, which is a chimeric A3A mimic. We believed that these experiments were a very good starting point because expression and purification of wild type A3A/A3B is problematic. In our binding

experiments of substrates containing dC, we used the enzyme A3A-E72A, where the active-site glutamic acid is replaced with alanine so that deamination can be avoided. Even though this mutation does cause a small but significant diminution in ligand binding through loss of hydrogen bonding, this experiment provides solid evidence that our cross-linked substrates can bind to wild-type A3A as well. However, this should be verified for our substrates and inhibitors using wild-type A3A enzyme. Initial experiments performed by Dr. Stefan Harjes showed that our oligos inhibit wild-type A3A. Inhibition of wild-type A3B is still a question, and these experiments would provide a proof as to whether we can develop selective inhibitors against A3A and A3B.

5.2.2. Structural studies on A3 cross-linked ssDNA complex

X-Ray crystallographic data on A3 enzymes bound to ssDNA provided detailed information of how the ssDNA is bound to it. This was a starting point for this research. We saw a clear difference in substrate activity and inhibitory potential between different generations of cross-links. We assumed that in cross-link 1, there is only a triazole in the place of the dT -2 and the difference between cross-link 2 and 3 is the presence of a highly hydrophobic carbazole base in cross-link 3 instead of adenosine at the +1 position. To understand why exactly cross-link 3 showed us superior inhibition in comparison with others, the crystal structures of A3 bound to cross-linked ssDNAs are essential. The information from the crystal structure can be used to design more powerful inhibitors of A3 enzymes.

5.2.3. Design of more potent inhibitors of CDA/A3

Based on our previous observation of improved inhibition of A3A/A3B by FdZ-containing oligos vs dZ-containing oligos, we can inhibit A3 enzymes by employing more potent inhibitors of CDA to mimic the transition state of cytosine deamination. The difference in CDA inhibition by the ***R*** and ***S*** isomers of **dazep** provides important information for the design of powerful CDA inhibitors that could potentially be incorporated into a ssDNA sequence to inhibit A3 enzymes. The new and more powerful transition-state analogues embedded in a cross-linked ssDNA with optimized structural features would potentially provide picomolar inhibitors.

5.2.4. Cross-linked oligonucleotides with enhanced stability against digestion by nucleases

Oligonucleotides containing natural nucleosides are prone to degradation by nucleases which limit their applications *in vivo.* Experiments with snake venom phosphodiesterase showed that the cross-link provides some degree of protection against the nucleases at low

concentrations of the enzyme, but the stability of these oligos in biological system is still a question. Phosphodiesterase enzymes in biological systems are very active in degrading natural nucleoside phosphate backbones. One of the most widely used transformations of the phosphate backbone in clinically approved oligonucleotide drugs is a phosphothioate (P-S) modification. Here, one of the oxygens attached to the phosphorus is replaced by sulfur.

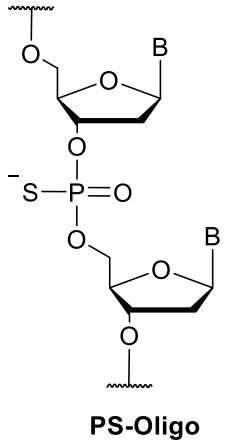


Figure 5.1. PS-modification used in oligonucleotides.

This modification is relatively easy to introduce into DNA/RNA during automated nucleic acid synthesis. PS-modified oligos show high resistance to nucleases.[207] However, PS-oligos contain a chiral center at each phosphorus atom and so it give a mixture of isomers. Takeshi Wada and co-workers developed a stereo-controlled solid-phase synthesis of PO/PS chimeric oligodeoxyribonucleotides on an automated synthesizer using an oxazaphospholidine-phosphoramidite method.[208] Stereo-controlled oligonucleotide synthesis with iterative capping and sulfurization protocol developed by Wave Life Sciences addressed the disadvantages of Wada's method, which suffers from the challenge associated with post-synthetic removal of the chiral auxiliary.[209]

It has been reported recently that a zwitterionic modification (N+) and a tosyl modification (Ts) of the phosphate backbone has shown significantly increased nuclease stability even at 20-fold excess of snake venom phosphodiesterase.[210] These modifications can be used as well to protect inhibitors from enzymatic degradation.

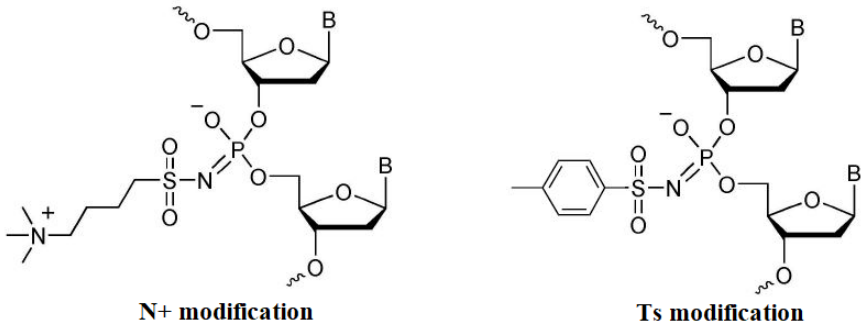


Figure 5.2. N+ and Ts phosphate modifications used to increase nuclease stability of oligonucleotides.

5.2.5. Targeted delivery of cross-linked oligonucleotides in the body

As different A3 enzymes have varying degree of localization inside of the cells, targeting them with oligonucleotides is a challenge. One of the major contributors of host genomic mutations is caused by A3B, which is predominantly nuclear, whereas A3A is mainly cytoplasmic. Selective inhibition can be achieved with appropriately modified inhibitors which can localize selectively to the nucleus of the cell or the cytoplasm. Native oligonucleotides pass biological membranes very poorly due to their negatively charged backbone. The chemistry of either the backbone or other structural features of the nucleotides can be modified to achieve better pharmacokinetics of these inhibitors to show their effect. It has been shown that PS-modified oligos can achieve desired concentrations in cells to elicit their therapeutic properties. Most of the antisense oligonucleotides approved for clinical use are based on this modification and they induce RNase H activity on premature mRNA in the cell nucleus. Recently, it has been shown that apart from nuclease stability, N+-and Ts-modified oligos can also be used to deliver oligos into either the cell nucleus or the cytoplasm. This study proved that Ts oligos were present in the nucleus of the cell whereas N+ oligos were present in the cytoplasm near the nucleus in vesicles.[210]

The membrane penetration and localization of oligonucleotides are also based on the length of the sequences used. To achieve the desired concentration of the oligo in the body, the length of the oligo can be reduced to the minimum which is optimum for binding with A3 enzymes. We found a reduced inhibitory potential of our 4-mer and 3-mer cross-linked oligonucleotides in comparison with 9-mer cross-linked oligo. This suggests further optimization of the length is necessary to make it more drug-like without compromising the inhibitory potential. Also, the localization of oligos can be modulated using a proper use of hydrophilic and hydrophobic residues in the sequence.

Thus, a cross-linked oligonucleotide containing a more powerful transition-state analogue of cytosine deamination, having an optimum U-shaped configuration, nuclease

resistant and selectively targeting different compartments of the body would be the best A3 inhibitor candidate for clinical use.

5.2.6. H-Phosphonate precursors of azide-containing nucleoside modifications for large scale synthesis

The synthesis of an azide-containing phosphoramidite is challenging due to the presence of a trivalent phosphorus and azide in the same compound. The compounds synthesized were used without any purifications, and the minute amounts of water present in the phosphoramidite during work up of the reaction reduces the coupling efficiency of these phosphoramidites. The storage of these compounds is not viable due to possible Staudinger reaction, and the phosphoramidite was used in the DNA synthesis right after the synthesis without any purification. These limitations are the main challenges in the large-scale production of the cross-linked ssDNAs with an azido-phosphoramidite. Testing the cross-linked DNAs *in vivo* in animal models to establish the pharmacokinetic, pharmacodynamic parameters and toxicity profile requires a large amount of material. The possibility of using H-phosphonates of nucleoside precursors can potentially eliminate the challenges associated with the large-scale synthesis of cross-linked ssDNAs.

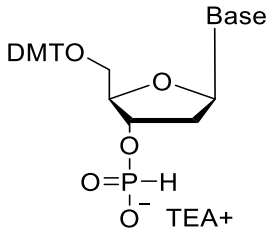


Figure 5.3. The chemical structure of nucleoside H-phosphonate precursor used for DNA synthesis.

We have been testing the DNA synthesis using a mixed protocol employing phosphoramidite and H-phosphonates in the automated DNA synthesizer, and the protocol proved to be quite successful to give a good overall yield of final oligo-containing an azide and an alkyne in the sequence. Further optimization of oligo synthesis is required to minimize losses during the synthesis and multiple handling steps,including three HPLC purifications.

Chapter 6. References

1. Housman, G.; Byler, S.; Heerboth, S.; Lapinska, K.; Longacre, M.; Snyder, N.; Sarkar, S., Drug resistance in cancer: an overview. *Cancers* **2014**, *6* (3), 1769-1792.

2. Mansoori, B.; Mohammadi, A.; Davudian, S.; Shirjang, S.; Baradaran, B., The Different Mechanisms of Cancer Drug Resistance: A Brief Review. *Adv. Pharm. Bull.* **2017**, *7* (3), 339-348.

3. Borst, P.; Evers, R.; Kool, M.; Wijnholds, J., A family of drug transporters: the multidrug resistance-associated proteins. *J. Natl. Cancer Inst.* **2000**, *92* (16), 1295-1302.

4. Cole, S.; Bhardwaj, G.; Gerlach, J.; Mackie, J.; Grant, C.; Almquist, K.; Stewart, A.; Kurz, E.; Duncan, A.; Deeley, R. G., Overexpression of a transporter gene in a multidrug-resistant human lung cancer cell line. *Science* **1992**, *258* (5088), 1650-1654.

5. Doyle, L. A.; Yang, W.; Abruzzo, L. V.; Krogmann, T.; Gao, Y.; Rishi, A. K.; Ross, D. D., A multidrug resistance transporter from human MCF-7 breast cancer cells. *Proc. Natl. Acad. Sci.* **1998**, *95* (26), 15665-15670.

6. Komatani, H.; Kotani, H.; Hara, Y.; Nakagawa, R.; Matsumoto, M.; Arakawa, H.; Nishimura, S., Identification of breast cancer resistant protein/mitoxantrone resistance/placenta-specific, ATP-binding cassette transporter as a transporter of NB-506 and J-107088, topoisomerase I inhibitors with an indolocarbazole structure. *Cancer Res.* **2001**, *61* (7), 2827-2832.

7. Roberts, S. A.; Lawrence, M. S.; Klimczak, L. J.; Grimm, S. A.; Fargo, D.; Stojanov, P.; Kiezun, A.; Kryukov, G. V.; Carter, S. L.; Saksena, G., An APOBEC cytidine deaminase mutagenesis pattern is widespread in human cancers. *Nat. Genet.* **2013**, *45* (9), 970.

8. Jarvis, M. C.; Ebrahimi, D.; Temiz, N. A.; Harris, R. S., Mutation Signatures Including APOBEC in Cancer Cell Lines. *JNCI Cancer Spectr.* **2018**, *2* (1).

9. Swanton, C.; McGranahan, N.; Starrett, G. J.; Harris, R. S., APOBEC enzymes: mutagenic fuel for cancer evolution and heterogeneity. *Cancer Discov.* **2015**, *5* (7), 704-712.

10. Jarmuz, A.; Chester, A.; Bayliss, J.; Gisbourne, J.; Dunham, I.; Scott, J.; Navaratnam, N., An anthropoid-specific locus of orphan C to U RNA-editing enzymes on chromosome 22. *Genomics* **2002**, *79* (3), 285-296.

11. Navaratnam, N.; Morrison, J. R.; Bhattacharya, S.; Patel, D.; Funahashi, T.; Giannoni, F.; Teng, B.; Davidson, N.; Scott, J., The p27 catalytic subunit of the apolipoprotein B mRNA editing enzyme is a cytidine deaminase. *J. Biol. Chem.* **1993**, *268* (28), 20709-20712.

12. Chan, L., Apolipoprotein B, the major protein component of triglyceride-rich and low density lipoproteins. *J. Biol.Chem.* **1992**, *267* (36), 25621-25624.

13. Alce, T. M.; Popik, W., APOBEC3G is incorporated into virus-like particles by a direct interaction with HIV-1 Gag nucleocapsid protein. *J. Biol. Chem.* **2004**, *279* (33), 34083-34086.

14. Bogerd, H. P.; Cullen, B. R., Single-stranded RNA facilitates nucleocapsid: APOBEC3G complex formation. *RNA* **2008**, *14* (6), 1228-1236.
15. Haché, G.; Liddament, M. T.; Harris, R. S., The retroviral hypermutation specificity of APOBEC3F and APOBEC3G is governed by the C-terminal DNA cytosine deaminase domain. *J. Biol. Chem.* **2005**, *280* (12), 10920-10924.
16. Yu, Q.; König, R.; Pillai, S.; Chiles, K.; Kearney, M.; Palmer, S.; Richman, D.; Coffin, J. M.; Landau, N. R., Single-strand specificity of APOBEC3G accounts for minus-strand deamination of the HIV genome. *Nat. Struct. Mol. Biol.* **2004**, *11* (5), 435-442.
17. Harris, R. S.; Bishop, K. N.; Sheehy, A. M.; Craig, H. M.; Petersen-Mahrt, S. K.; Watt, I. N.; Neuberger, M. S.; Malim, M. H., DNA deamination mediates innate immunity to retroviral infection. *Cell* **2003**, *113* (6), 803-809.
18. Mangeat, B.; Turelli, P.; Caron, G.; Friedli, M.; Perrin, L.; Trono, D., Broad antiretroviral defence by human APOBEC3G through lethal editing of nascent reverse transcripts. *Nature* **2003**, *424* (6944), 99-103.
19. Baumert, T. F.; Rösler, C.; Malim, M. H.; von Weizsäcker, F., Hepatitis B virus DNA is subject to extensive editing by the human deaminase APOBEC3C. *Hepatology* **2007**, *46* (3), 682-689.
20. Köck, J.; Blum, H. E., Hypermutation of hepatitis B virus genomes by APOBEC3G, APOBEC3C and APOBEC3H. *J. Gen. Virol.* **2008**, *89* (5), 1184-1191.
21. Shapiro, M.; Krug, L. T.; MacCarthy, T., Mutational pressure by host APOBEC3s more strongly affects genes expressed early in the lytic phase of herpes simplex virus-1 (HSV-1) and human polyomavirus (HPyV) infection. *PLoS Path.* **2021**, *17* (4), e1009560.
22. Warren, C. J.; Xu, T.; Guo, K.; Griffin, L. M.; Westrich, J. A.; Lee, D.; Lambert, P. F.; Santiago, M. L.; Pyeon, D., APOBEC3A functions as a restriction factor of human papillomavirus. *J. Virol.* **2015**, *89* (1), 688-702.
23. Friedman, N.; Jacob-Hirsch, J.; Drori, Y.; Eran, E.; Kol, N.; Nayshool, O.; Mendelson, E.; Rechavi, G.; Mandelboim, M., Transcriptomic profiling and genomic mutational analysis of Human coronavirus (HCoV)-229E-infected human cells. *PLoS One* **2021**, *16* (2), e0247128.
24. Simmonds, P., Rampant C→ U hypermutation in the genomes of SARS-CoV-2 and other coronaviruses: causes and consequences for their short-and long-term evolutionary trajectories. *Msphere* **2020**, *5* (3), e00408-20.
25. Koning, F. A.; Newman, E. N.; Kim, E.-Y.; Kunstman, K. J.; Wolinsky, S. M.; Malim, M. H., Defining APOBEC3 expression patterns in human tissues and hematopoietic cell subsets. *J. Virol.* **2009**, *83* (18), 9474-9485.
26. Refsland, E. W.; Stenglein, M. D.; Shindo, K.; Albin, J. S.; Brown, W. L.; Harris, R. S., Quantitative profiling of the full APOBEC3 mRNA repertoire in lymphocytes and tissues: implications for HIV-1 restriction. *Nucleic Acids Res.* **2010**, *38* (13), 4274-4284.

27. Burns, M. B.; Lackey, L.; Carpenter, M. A.; Rathore, A.; Land, A. M.; Leonard, B.; Refsland, E. W.; Kotandeniya, D.; Tretyakova, N.; Nikas, J. B., APOBEC3B is an enzymatic source of mutation in breast cancer. *Nature* **2013**, *494* (7437), 366.

28. Conticello, S. G., The AID/APOBEC family of nucleic acid mutators. *Genome Biol.* **2008**, *9* (6), 229.

29. Harris, R. S.; Liddament, M. T., Retroviral restriction by APOBEC proteins. *Nat. Rev. Immunol.* **2004**, *4* (11), 868-877.

30. Bohn, M.-F.; Shandilya, S. M.; Albin, J. S.; Kouno, T.; Anderson, B. D.; McDougle, R. M.; Carpenter, M. A.; Rathore, A.; Evans, L.; Davis, A. N., Crystal structure of the DNA cytosine deaminase APOBEC3F: the catalytically active and HIV-1 Vif-binding domain. *Structure* **2013**, *21* (6), 1042-1050.

31. Byeon, I.-J. L.; Ahn, J.; Mitra, M.; Byeon, C.-H.; Hercík, K.; Hritz, J.; Charlton, L. M.; Levin, J. G.; Gronenborn, A. M., NMR structure of human restriction factor APOBEC3A reveals substrate binding and enzyme specificity. *Nat. Commun.* **2013**, *4* (1), 1-11.

32. Byeon, I.-J. L.; Byeon, C.-H.; Wu, T.; Mitra, M.; Singer, D.; Levin, J. G.; Gronenborn, A. M., Nuclear magnetic resonance structure of the APOBEC3B catalytic domain: structural basis for substrate binding and DNA deaminase activity. *Biochemistry* **2016**, *55* (21), 2944-2959.

33. Chen, K.-M.; Harjes, E.; Gross, P. J.; Fahmy, A.; Lu, Y.; Shindo, K.; Harris, R. S.; Matsuo, H., Structure of the DNA deaminase domain of the HIV-1 restriction factor APOBEC3G. *Nature* **2008**, *452* (7183), 116-119.

34. Holden, L. G.; Prochnow, C.; Chang, Y. P.; Bransteitter, R.; Chelico, L.; Sen, U.; Stevens, R. C.; Goodman, M. F.; Chen, X. S., Crystal structure of the anti-viral APOBEC3G catalytic domain and functional implications. *Nature* **2008**, *456* (7218), 121-124.

35. Kitamura, S.; Ode, H.; Nakashima, M.; Imahashi, M.; Naganawa, Y.; Kurosawa, T.; Yokomaku, Y.; Yamane, T.; Watanabe, N.; Suzuki, A., The APOBEC3C crystal structure and the interface for HIV-1 Vif binding. *Nat. Struct. Mol. Biol.* **2012**, *19* (10), 1005-1010.

36. Shi, K.; Carpenter, M. A.; Banerjee, S.; Shaban, N. M.; Kurahashi, K.; Salamango, D. J.; McCann, J. L.; Starrett, G. J.; Duffy, J. V.; Demir, Ö., Structural basis for targeted DNA cytosine deamination and mutagenesis by APOBEC3A and APOBEC3B. *Nat. Struct. Mol. Biol.* **2017**, *24* (2), 131-139.

37. Shi, K.; Carpenter, M. A.; Kurahashi, K.; Harris, R. S.; Aihara, H., Crystal structure of the DNA deaminase APOBEC3B catalytic domain. *J. Biol. Chem.* **2015**, *290* (47), 28120-28130.

38. Bonvin, M.; Greeve, J., Effects of point mutations in the cytidine deaminase domains of APOBEC3B on replication and hypermutation of hepatitis B virus in vitro. *J. Gen. Virol.* **2007**, *88* (12), 3270-3274.

39. Navarro, F.; Bollman, B.; Chen, H.; König, R.; Yu, Q.; Chiles, K.; Landau, N. R., Complementary function of the two catalytic domains of APOBEC3G. *Virology* **2005**, *333* (2), 374-386.

40. Siriwardena, S. U.; Guruge, T. A.; Bhagwat, A. S., Characterization of the catalytic domain of human APOBEC3B and the critical structural role for a conserved methionine. *J. Mol. Biol.* **2015**, *427* (19), 3042-3055.

41. Harris, R. S.; Dudley, J. P., APOBECs and virus restriction. *Virology* **2015**, *479*, 131-145.

42. Refsland, E. W.; Harris, R. S., The APOBEC3 family of retroelement restriction factors. *Intrinsic Immunity* **2013**, 1-27.

43. Harjes, E.; Gross, P. J.; Chen, K.-M.; Lu, Y.; Shindo, K.; Nowarski, R.; Gross, J. D.; Kotler, M.; Harris, R. S.; Matsuo, H., An extended structure of the APOBEC3G catalytic domain suggests a unique holoenzyme model. *J. Mol. Biol.* **2009**, *389* (5), 819-832.

44. Shandilya, S. M.; Nalam, M. N.; Nalivaika, E. A.; Gross, P. J.; Valesano, J. C.; Shindo, K.; Li, M.; Munson, M.; Royer, W. E.; Harjes, E., Crystal structure of the APOBEC3G catalytic domain reveals potential oligomerization interfaces. *Structure* **2010**, *18* (1), 28-38.

45. Lu, X.; Zhang, T.; Xu, Z.; Liu, S.; Zhao, B.; Lan, W.; Wang, C.; Ding, J.; Cao, C., Crystal structure of DNA cytidine deaminase ABOBEC3G catalytic deamination domain suggests a binding mode of full-length enzyme to single-stranded DNA. *J. Biol. Chem.* **2015**, *290* (7), 4010-4021.

46. Xiao, X.; Li, S.-X.; Yang, H.; Chen, X. S., Crystal structures of APOBEC3G N-domain alone and its complex with DNA. *Nat. Commun.* **2016**, *7* (1), 1-11.

47. Chelico, L.; Sacho, E. J.; Erie, D. A.; Goodman, M. F., A model for oligomeric regulation of APOBEC3G cytosine deaminase-dependent restriction of HIV. *J. Biol. Chem.* **2008**, *283* (20), 13780-13791.

48. Yang, H.; Ito, F.; Wolfe, A. D.; Li, S.; Mohammadzadeh, N.; Love, R. P.; Yan, M.; Zirkle, B.; Gaba, A.; Chelico, L., Understanding the structural basis of HIV-1 restriction by the full length double-domain APOBEC3G. *Nat. Commun.* **2020**, *11* (1), 1-11.

49. Maiti, A.; Myint, W.; Delviks-Frankenberry, K. A.; Hou, S.; Kanai, T.; Balachandran, V.; Rodriguez, C. S.; Tripathi, R.; Yilmaz, N. K.; Pathak, V. K., Crystal Structure of a Soluble APOBEC3G Variant Suggests ssDNA to Bind in a Channel that Extends between the Two Domains. *J. Mol. Biol.* **2020**, *432* (23), 6042-6060.

50. LaRue, R. S.; Andrésdóttir, V.; Blanchard, Y.; Conticello, S. G.; Derse, D.; Emerman, M.; Greene, W. C.; Jónsson, S. R.; Landau, N. R.; Löchelt, M., Guidelines for naming nonprimate APOBEC3 genes and proteins. *J. Virol.* **2009**, *83* (2), 494-497.

51. Conticello, S. G.; Thomas, C. J.; Petersen-Mahrt, S. K.; Neuberger, M. S., Evolution of the AID/APOBEC family of polynucleotide (deoxy) cytidine deaminases. *Mol. Biol. Evol.* **2005**, *22* (2), 367-377.

52. Silvas, T. V.; Schiffer, C. A., APOBEC3s: DNA-editing human cytidine deaminases. *Protein Sci.* **2019**, *28* (9), 1552-1566.

53. Chelico, L.; Pham, P.; Calabrese, P.; Goodman, M. F., APOBEC3G DNA deaminase acts processively 3'→ 5' on single-stranded DNA. *Nat. Struct. Mol. Biol.* **2006**, *13* (5), 392-399.

54. Kvach, M. V.; Barzak, F. M.; Harjes, S.; Schares, H. A.; Jameson, G. B.; Ayoub, A. M.; Moorthy, R.; Aihara, H.; Harris, R. S.; Filichev, V. V., Inhibiting APOBEC3 Activity with Single-Stranded DNA Containing 2'-Deoxyzebularine Analogues. *Biochemistry* **2018**, *58* (5), 391-400.

55. Beale, R. C.; Petersen-Mahrt, S. K.; Watt, I. N.; Harris, R. S.; Rada, C.; Neuberger, M. S., Comparison of the differential context-dependence of DNA deamination by APOBEC enzymes: correlation with mutation spectra in vivo. *J. Mol. Biol.* **2004**, *337* (3), 585-596.

56. Kohli, R. M.; Maul, R. W.; Guminski, A. F.; McClure, R. L.; Gajula, K. S.; Saribasak, H.; McMahon, M. A.; Siliciano, R. F.; Gearhart, P. J.; Stivers, J. T., Local sequence targeting in the AID/APOBEC family differentially impacts retroviral restriction and antibody diversification. *J. Biol. Chem.* **2010**, *285* (52), 40956-40964.

57. Suspene, R.; Guétard, D.; Henry, M.; Sommer, P.; Wain-Hobson, S.; Vartanian, J.-P., Extensive editing of both hepatitis B virus DNA strands by APOBEC3 cytidine deaminases in vitro and in vivo. *Proc. Natl. Acad. Sci.* **2005**, *102* (23), 8321-8326.

58. Liu, M.; Mallinger, A. l.; Tortorici, M.; Newbatt, Y.; Richards, M.; Mirza, A.; van Montfort, R. L.; Burke, R.; Blagg, J.; Kaserer, T., Evaluation of APOBEC3B recognition motifs by NMR reveals preferred substrates. *ACS Chem. Biol.* **2018**, *13* (9), 2427-2432.

59. Harjes, S.; Solomon, W. C.; Li, M.; Chen, K.-M.; Harjes, E.; Harris, R. S.; Matsuo, H., Impact of H216 on the DNA binding and catalytic activities of the HIV restriction factor APOBEC3G. *J. Virol.* **2013**, *87* (12), 7008-7014.

60. Silvas, T. V.; Hou, S.; Myint, W.; Nalivaika, E.; Somasundaran, M.; Kelch, B. A.; Matsuo, H.; Yilmaz, N. K.; Schiffer, C. A., Substrate sequence selectivity of APOBEC3A implicates intra-DNA interactions. *Sci. Rep.* **2018**, *8* (1), 1-11.

61. Neuberger, M. S.; Harris, R. S.; Di Noia, J.; Petersen-Mahrt, S. K., Immunity through DNA deamination. *Trends Biochem. Sci.* **2003**, *28* (6), 305-312.

62. Kouno, T.; Silvas, T. V.; Hilbert, B. J.; Shandilya, S. M.; Bohn, M. F.; Kelch, B. A.; Royer, W. E.; Somasundaran, M.; Yilmaz, N. K.; Matsuo, H., Crystal structure of APOBEC3A bound to single-stranded DNA reveals structural basis for cytidine deamination and specificity. *Nat. Commun.* **2017**, *8* (1), 1-8.

63. Maiti, A.; Myint, W.; Kanai, T.; Delviks-Frankenberry, K.; Rodriguez, C. S.; Pathak, V. K.; Schiffer, C. A.; Matsuo, H., Crystal structure of the catalytic domain of HIV-1 restriction factor APOBEC3G in complex with ssDNA. *Nat. Commun.* **2018**, *9* (1), 1-11.

64. Henderson, S.; Fenton, T., APOBEC3 genes: retroviral restriction factors to cancer drivers. *Trends Mol. Med.* **2015**, *21* (5), 274-284.

65. Roberts, S. A.; Gordenin, D. A., Hypermutation in human cancer genomes: footprints and mechanisms. *Nat. Rev. Cancer* **2014**, *14* (12), 786.

66. Roberts, S. A.; Sterling, J.; Thompson, C.; Harris, S.; Mav, D.; Shah, R.; Klimczak, L. J.; Kryukov, G. V.; Malc, E.; Mieczkowski, P. A., Clustered mutations in yeast and in human cancers can arise from damaged long single-strand DNA regions. *Mol. Cell* **2012**, *46* (4), 424-435.

67. Law, E. K.; Sieuwerts, A. M.; LaPara, K.; Leonard, B.; Starrett, G. J.; Molan, A. M.; Temiz, N. A.; Vogel, R. I.; Meijer-van Gelder, M. E.; Sweep, F. C., The DNA cytosine deaminase APOBEC3B promotes tamoxifen resistance in ER-positive breast cancer. *Sci. Adv.* **2016**, *2* (10), e1601737.

68. Desimmie, B. A.; Delviks-Frankenberrry, K. A.; Burdick, R. C.; Qi, D.; Izumi, T.; Pathak, V. K., Multiple APOBEC3 restriction factors for HIV-1 and one Vif to rule them all. *J. Mol. Biol.* **2014**, *426* (6), 1220-1245.

69. Hultquist, J. F.; Harris, R. S., Leveraging APOBEC3 proteins to alter the HIV mutation rate and combat AIDS. *Future Neurol.* **2009**, *4* (6), 605-619.

70. Bishop, K. N.; Verma, M.; Kim, E.-Y.; Wolinsky, S. M.; Malim, M. H., APOBEC3G inhibits elongation of HIV-1 reverse transcripts. *PLoS Path.* **2008**, *4* (12), e1000231.

71. Newman, E. N.; Holmes, R. K.; Craig, H. M.; Klein, K. C.; Lingappa, J. R.; Malim, M. H.; Sheehy, A. M., Antiviral function of APOBEC3G can be dissociated from cytidine deaminase activity. *Curr. Biol.* **2005**, *15* (2), 166-170.

72. Schumacher, A. J.; Haché, G.; MacDuff, D. A.; Brown, W. L.; Harris, R. S., The DNA deaminase activity of human APOBEC3G is required for Ty1, MusD, and human immunodeficiency virus type 1 restriction. *J. Virol.* **2008**, *82* (6), 2652-2660.

73. Bishop, K. N.; Holmes, R. K.; Malim, M. H., Antiviral potency of APOBEC proteins does not correlate with cytidine deamination. *J. Virol.* **2006**, *80* (17), 8450-8458.

74. Lecossier, D.; Bouchonnet, F.; Clavel, F.; Hance, A. J., Hypermutation of HIV-1 DNA in the absence of the Vif protein. *Science* **2003**, *300* (5622), 1112-1112.

75. Zhang, H.; Yang, B.; Pomerantz, R. J.; Zhang, C.; Arunachalam, S. C.; Gao, L., The cytidine deaminase CEM15 induces hypermutation in newly synthesized HIV-1 DNA. *Nature* **2003**, *424* (6944), 94-98.

76. Gabuzda, D. H.; Lawrence, K.; Langhoff, E. e. a.; Terwilliger, E.; Dorfman, T.; Haseltine, W.; Sodroski, J., Role of vif in replication of human immunodeficiency virus type 1 in CD4+ T lymphocytes. *J. Virol.* **1992**, *66* (11), 6489-6495.

77. Sheehy, A. M.; Gaddis, N. C.; Choi, J. D.; Malim, M. H., Isolation of a human gene that inhibits HIV-1 infection and is suppressed by the viral Vif protein. *Nature* **2002**, *418* (6898), 646-650.

78. Conticello, S. G.; Harris, R. S.; Neuberger, M. S., The Vif protein of HIV triggers degradation of the human antiretroviral DNA deaminase APOBEC3G. *Curr. Biol.* **2003**, *13* (22), 2009-2013.

79. Sheehy, A. M.; Gaddis, N. C.; Malim, M. H., The antiretroviral enzyme APOBEC3G is degraded by the proteasome in response to HIV-1 Vif. *Nat. Med.* **2003**, *9* (11), 1404-1407.
80. Mehle, A.; Strack, B.; Ancuta, P.; Zhang, C.; McPike, M.; Gabuzda, D., Vif overcomes the innate antiviral activity of APOBEC3G by promoting its degradation in the ubiquitin-proteasome pathway. *J. Biol. Chem.* **2004**, *279* (9), 7792-7798.
81. Willems, L.; Gillet, N. A., APOBEC3 interference during replication of viral genomes. *Viruses* **2015**, *7* (6), 2999-3018.
82. Turelli, P.; Mangeat, B.; Jost, S.; Vianin, S.; Trono, D., Inhibition of Hepatitis B Virus Replication by APOBEC3G. *Science* **2004**, *303* (5665), 1829-1829.
83. Stenglein, M. D.; Burns, M. B.; Li, M.; Lengyel, J.; Harris, R. S., APOBEC3 proteins mediate the clearance of foreign DNA from human cells. *Nat. Struct. Mol. Biol.* **2010**, *17* (2), 222-229.
84. Nik-Zainal, S.; Alexandrov, L. B.; Wedge, D. C.; Van Loo, P.; Greenman, C. D.; Raine, K.; Jones, D.; Hinton, J.; Marshall, J.; Stebbings, L. A., Mutational processes molding the genomes of 21 breast cancers. *Cell* **2012**, *149* (5), 979-993.
85. Barbieri, C. E.; Baca, S. C.; Lawrence, M. S.; Demichelis, F.; Blattner, M.; Theurillat, J.-P.; White, T. A.; Stojanov, P.; Van Allen, E.; Stransky, N., Exome sequencing identifies recurrent SPOP, FOXA1 and MED12 mutations in prostate cancer. *Nat. Genet.* **2012**, *44* (6), 685-689.
86. Stransky, N.; Egloff, A. M.; Tward, A. D.; Kostic, A. D.; Cibulskis, K.; Sivachenko, A.; Kryukov, G. V.; Lawrence, M. S.; Sougnez, C.; McKenna, A., The mutational landscape of head and neck squamous cell carcinoma. *Science* **2011**, *333* (6046), 1157-1160.
87. Sieuwerts, A. M.; Willis, S.; Burns, M. B.; Look, M. P.; Meijer-Van Gelder, M. E.; Schlicker, A.; Heideman, M. R.; Jacobs, H.; Wessels, L.; Leyland-Jones, B., Elevated APOBEC3B correlates with poor outcomes for estrogen-receptor-positive breast cancers. *Horm. Cancer* **2014**, *5* (6), 405-413.
88. Yan, S.; He, F.; Gao, B.; Wu, H.; Li, M.; Huang, L.; Liang, J.; Wu, Q.; Li, Y., Increased APOBEC3B predicts worse outcomes in lung cancer: a comprehensive retrospective study. *J. Cancer* **2016**, *7* (6), 618-625.
89. Glaser, A. P.; Fantini, D.; Wang, Y.; Yu, Y.; Rimar, K. J.; Podojil, J. R.; Miller, S. D.; Meeks, J. J., APOBEC-mediated mutagenesis in urothelial carcinoma is associated with improved survival, mutations in DNA damage response genes, and immune response. *Oncotarget* **2018**, *9* (4), 4537-4548.
90. Harris, R. S., Molecular mechanism and clinical impact of APOBEC3B-catalyzed mutagenesis in breast cancer. *Breast Cancer Res.* **2015**, *17* (1), 1-10.
91. Henderson, S.; Chakravarthy, A.; Su, X.; Boshoff, C.; Fenton, T. R., APOBEC-mediated cytosine deamination links PIK3CA helical domain mutations to human papillomavirus-driven tumor development. *Cell Rep.* **2014**, *7* (6), 1833-1841.
92. Warren, C. J.; Pyeon, D., APOBEC3 in papillomavirus restriction, evolution and cancer progression. *Oncotarget* **2015**, *6* (37), 39385-39386.

93. Lackey, L.; Demorest, Z. L.; Land, A. M.; Hultquist, J. F.; Brown, W. L.; Harris, R. S., APOBEC3B and AID have similar nuclear import mechanisms. *J. Mol. Biol.* **2012**, *419* (5), 301-314.

94. Cortez, L. M.; Brown, A. L.; Dennis, M. A.; Collins, C. D.; Brown, A. J.; Mitchell, D.; Mertz, T. M.; Roberts, S. A., APOBEC3A is a prominent cytidine deaminase in breast cancer. *PLoS Genet.* **2019**, *15* (12), e1008545.

95. Komatsu, A.; Nagasaki, K.; Fujimori, M.; Amano, J.; Miki, Y., Identification of novel deletion polymorphisms in breast cancer. *Int. J. Oncol.* **2008**, *33* (2), 261-270.

96. Buisson, R.; Langenbucher, A.; Bowen, D.; Kwan, E. E.; Benes, C. H.; Zou, L.; Lawrence, M. S., Passenger hotspot mutations in cancer driven by APOBEC3A and mesoscale genomic features. *Science* **2019**, *364* (6447), eaaw2872.

97. Olson, M. E.; Harris, R. S.; Harki, D. A., APOBEC Enzymes as Targets for Virus and Cancer Therapy. *Cell Chem. Biol.* **2018**, *25* (1), 36-49.

98. Harris, R. S., Enhancing immunity to HIV through APOBEC. *Nat. Biotechnol.* **2008**, *26* (10), 1089-1090.

99. Nathans, R.; Cao, H.; Sharova, N.; Ali, A.; Sharkey, M.; Stranska, R.; Stevenson, M.; Rana, T. M., Small-molecule inhibition of HIV-1 Vif. *Nat. Biotechnol.* **2008**, *26* (10), 1187-1192.

100. Ali, A.; Wang, J.; Nathans, R. S.; Cao, H.; Sharova, N.; Stevenson, M.; Rana, T. M., Synthesis and structure-activity relationship studies of HIV-1 virion infectivity factor (Vif) inhibitors that block viral replication. *ChemMedChem* **2012**, *7* (7), 1217-1229.

101. Mohammed, I.; Parai, M. K.; Jiang, X.; Sharova, N.; Singh, G.; Stevenson, M.; Rana, T. M., SAR and lead optimization of an HIV-1 Vif-APOBEC3G axis inhibitor. *ACS Med. Chem. Lett.* **2012**, *3* (6), 465-469.

102. Kao, S.; Miyagi, E.; Khan, M. A.; Takeuchi, H.; Opi, S.; Goila-Gaur, R.; Strebel, K., Production of infectious human immunodeficiency virus type 1 does not require depletion of APOBEC3G from virus-producing cells. *Retrovirology* **2004**, *1* (1), 1-12.

103. Caride, E.; Brindeiro, R. M.; Kallas, E. G.; de Sá, C. A. M.; Eyer-Silva, W. A.; Machado, E.; Tanuri, A., Sexual transmission of HIV-1 isolate showing G → A hypermutation. *J. Clin. Virol.* **2002**, *23* (3), 179-189.

104. Kieffer, T. L.; Kwon, P.; Nettles, R. E.; Han, Y.; Ray, S. C.; Siliciano, R. F., G→ A hypermutation in protease and reverse transcriptase regions of human immunodeficiency virus type 1 residing in resting CD4+ T cells in vivo. *J. Virol.* **2005**, *79* (3), 1975-1980.

105. Kim, E.-Y.; Bhattacharya, T.; Kunstman, K.; Swantek, P.; Koning, F. A.; Malim, M. H.; Wolinsky, S. M., Human APOBEC3G-mediated editing can promote HIV-1 sequence diversification and accelerate adaptation to selective pressure. *J. Virol.* **2010**, *84* (19), 10402-10405.

106. Mulder, L. C.; Harari, A.; Simon, V., Cytidine deamination induced HIV-1 drug resistance. *Proc. Natl. Acad. Sci.* **2008**, *105* (14), 5501-5506.

107. Jern, P.; Russell, R. A.; Pathak, V. K.; Coffin, J. M., Likely role of APOBEC3G-mediated G-to-A mutations in HIV-1 evolution and drug resistance. *PLoS Pathog.* **2009**, *5* (4), e1000367.

108. Li, M.; Shandilya, S. M.; Carpenter, M. A.; Rathore, A.; Brown, W. L.; Perkins, A. L.; Harki, D. A.; Solberg, J.; Hook, D. J.; Pandey, K. K., First-in-class small molecule inhibitors of the single-strand DNA cytosine deaminase APOBEC3G. *ACS Chem. Biol.* **2012**, *7* (3), 506-517.

109. Olson, M. E.; Li, M.; Harris, R. S.; Harki, D. A., Small molecule APOBEC3G DNA cytosine deaminase inhibitors based on a 4-amino-1,2,4-triazole-3-thiol scaffold. *ChemMedChem* **2013**, *8* (1), 112-117.

110. Wentworth, D. F.; Wolfenden, R., [52] Cytidine deaminases (from Escherichia coli and human liver). In *Methods Enzymol.*, Elsevier: **1978**, Vol. 51, pp 401-407.

111. Lauter, C. B.; Bailey, E. J.; Lerner, A. M., Assessment of cytosine arabinoside as an antiviral agent in humans. *Antimicrob. Agents Chemother.* **1974**, *6* (5), 598-602.

112. Robak, T.; Wierzbowska, A., Current and emerging therapies for acute myeloid leukemia. *Clin. Ther.* **2009**, *31*, 2349-2370.

113. Kantarjian, H.; Issa, J. P. J.; Rosenfeld, C. S.; Bennett, J. M.; Albitar, M.; DiPersio, J.; Klimek, V.; Slack, J.; De Castro, C.; Ravandi, F., Decitabine improves patient outcomes in myelodysplastic syndromes: results of a phase III randomized study. *Cancer.* **2006**, *106* (8), 1794-1803.

114. Laliberte, J.; Marquez, V. E.; Momparler, R. L., Potent inhibitors for the deamination of cytosine arabinoside and 5-aza-2'-deoxycytidine by human cytidine deaminase. *Cancer Chemother. Pharmacol.* **1992**, *30* (1), 7-11.

115. Mahfouz, R. Z.; Jankowska, A.; Ebrahem, Q.; Gu, X.; Visconte, V.; Tabarroki, A.; Terse, P.; Covey, J.; Chan, K.; Ling, Y., Increased CDA expression/activity in males contributes to decreased cytidine analog half-life and likely contributes to worse outcomes with 5-azacytidine or decitabine therapy. *Clin. Cancer. Res.* **2013**, *19* (4), 938-948.

116. McCormack, J. J.; Marquez, V. E.; Liu, P. S.; Vistica, D. T.; Driscoll, J. S., Inhibition of cytidine deaminase by 2-oxopyrimidine riboside and related compounds. *Biochem. Pharmacol.* **1980**, *29* (5), 830-832.

117. Betts, L.; Xiang, S.; Short, S. A.; Wolfenden, R.; Carter Jr, C. W., Cytidine Deaminase. The 2· 3 Å crystal structure of an enzyme: transition-state analog complex. *J. Mol. Biol.* **1994**, *235* (2), 635-656.

118. Kim, C. H.; Marquez, V. E.; Mao, D. T.; Haines, D. R.; McCormack, J. J., Synthesis of pyrimidin-2-one nucleosides as acid-stable inhibitors of cytidine deaminase. *J. Med. Chem.* **1986**, *29* (8), 1374-1380.

119. Camiener, G. W., Studies of the enzymatic deamination of ara-cytidine—V: Inhibition in vitro and in vivo by tetrahydrouridine and other reduced pyrimidine nucleosides. *Biochem. Pharmacol.* **1968**, *17* (9), 1981-1991.

120. Ludek, O. R.; Schroeder, G. K.; Liao, C.; Russ, P. L.; Wolfenden, R.; Marquez, V. E., Synthesis and Conformational Analysis of Locked Carbocyclic Analogues of 1,3-Diazepinone Riboside, a High-Affinity Cytidine Deaminase Inhibitor. *J. Org. Chem.* **2009**, *74* (16), 6212-6223.

121. Kim, M.; Gajulapati, K.; Kim, C.; Jung, H. Y.; Goo, J.; Lee, K.; Kaur, N.; Kang, H. J.; Chung, S. J.; Choi, Y., A facile synthetic route to diazepinone derivatives via ring closing metathesis and its application for human cytidine deaminase inhibitors. *Chem. Commun.* **2012**, *48* (93), 11443-11445.

122. Liu, P. S.; Marquez, V. E.; Kelley, J. A.; Driscoll, J. S., Synthesis of 1, 3-diazepin-2-one nucleosides as transition-state inhibitors of cytidine deaminase. 2. *J. Org. Chem.* **1980**, *45* (25), 5225-5227.

123. Marquez, V. E.; Liu, P. S.; Kelley, J. A.; Driscoll, J. S.; McCormack, J. J., Synthesis of 1,3-diazepin-2-one nucleosides as transition-state inhibitors of cytidine deaminase. *J. Med. Chem.* **1980**, *23* (7), 713-715.

124. Kvach, M. V.; Barzak, F. M.; Harjes, S.; Schares, H. A. M.; Kurup, H. M.; Jones, K. F.; Sutton, L.; Donahue, J.; D'Aquila, R. T.; Jameson, G. B.; Harki, D. A.; Krause, K. L.; Harjes, E.; Filichev, V. V., Differential Inhibition of APOBEC3 DNA-Mutator Isozymes by Fluoro- and Non-Fluoro-Substituted 2'-Deoxyzebularine Embedded in Single-Stranded DNA. *ChemBioChem* **2020**, *21* (7), 1028-1035.

125. Rolland, V.; Kotera, M.; Lhomme, J., Convenient preparation of 2-deoxy-3, 5-di-O-p-toluoyl-α-D-erythro-pentofuranosyl chloride. *Synth. Commun.* **1997**, *27* (20), 3505-3511.

126. Barzak, F. M.; Harjes, S.; Kvach, M. V.; Kurup, H. M.; Jameson, G. B.; Filichev, V. V.; Harjes, E., Selective inhibition of APOBEC3 enzymes by single-stranded DNAs containing 2'-deoxyzebularine. *Org. Biomol. Chem.* **2019**, *17* (43), 9435-9441.

127. Kidd, J. M.; Newman, T. L.; Tuzun, E.; Kaul, R.; Eichler, E. E., Population stratification of a common APOBEC gene deletion polymorphism. *PLoS Genet* **2007**, *3* (4), e63.

128. Bouillon, C.; Meyer, A.; Vidal, S.; Jochum, A.; Chevolot, Y.; Cloarec, J.-P.; Praly, J.-P.; Vasseur, J.-J.; Morvan, F., Microwave assisted “click” chemistry for the synthesis of multiple labeled-carbohydrate oligonucleotides on solid support. *J. Org. Chem.* **2006**, *71* (12), 4700-4702.

129. El-Sagheer, A. H.; Brown, T., Click chemistry with DNA. *Chem. Soc. Rev.* **2010**, *39* (4), 1388-1405.

130. Géci, I.; Filichev, V. V.; Pedersen, E. B., Stabilization of parallel triplexes by twisted intercalating nucleic acids (TINAs) incorporating 1,2,3-triazole units and prepared by microwave‐accelerated click chemistry. *Chem.- Eur. J.* **2007**, *13* (22), 6379-6386.

131. Gierlich, J.; Burley, G. A.; Gramlich, P. M.; Hammond, D. M.; Carell, T., Click chemistry as a reliable method for the high-density postsynthetic functionalization of alkyne-modified DNA. *Org. Lett.* **2006**, *8* (17), 3639-3642.

132. Astakhova, I. K.; Kumar, T. S.; Campbell, M. A.; Ustinov, A. V.; Korshun, V. A.; Wengel, J., Branched DNA nanostructures efficiently stabilised and monitored by novel pyrene-perylene 2'-α-l-amino-LNA FRET pairs. *Chem. Commun.* **2013**, *49* (5), 511-513.
133. Hou, S.; Silvas, T. V.; Leidner, F.; Nalivaika, E. A.; Matsuo, H.; Kurt Yilmaz, N.; Schiffer, C. A., Structural analysis of the active site and DNA binding of human cytidine deaminase APOBEC3B. *J. Chem. Theory Comput.* **2018**, *15* (1), 637-647.
134. Graham, D.; Parkinson, J. A.; Brown, T., DNA duplexes stabilized by modified monomer residues: synthesis and stability. *J. Chem. Soc. Perkin Trans. 1* **1998**, (6), 1131-1138.
135. Mitchell, A. R., Bruce Merrifield and solid-phase peptide synthesis: A historical assessment. *J. Pept. Sci.* **2008**, *90* (3), 175-184.
136. Kolb, H. C.; Finn, M.; Sharpless, K. B., Click chemistry: diverse chemical function from a few good reactions. *Angew. Chem. Int. Ed.* **2001**, *40* (11), 2004-2021.
137. Beckman, K. B.; Ames, B. N., Oxidative decay of DNA. *J. Biol. Chem.* **1997**, *272* (32), 19633-19636.
138. Chan, T. R.; Hilgraf, R.; Sharpless, K. B.; Fokin, V. V., Polytriazoles as copper(I)-stabilizing ligands in catalysis. *Org. Lett.* **2004**, *6* (17), 2853-2855.
139. Donnelly, P. S.; Zanatta, S. D.; Zammit, S. C.; White, J. M.; Williams, S. J., 'Click'cycloaddition catalysts: copper(I) and copper(II) tris(triazolylmethyl) amine complexes. *Chem. Commun.* **2008**, (21), 2459-2461.
140. Harjes, S.; Jameson, G. B.; Filichev, V. V.; Edwards, P. J.; Harjes, E., NMR-based method of small changes reveals how DNA mutator APOBEC3A interacts with its single-stranded DNA substrate. *Nucleic Acids Res.* **2017**, *45* (9), 5602-5613.
141. Paar, M.; Schrabmair, W.; Mairold, M.; Oettl, K.; Reibnegger, G., Global Regression Using the Explicit Solution of Michaelis-Menten Kinetics Employing Lambert's W Function: High Robustness of Parameter Estimates. *ChemistrySelect* **2019**, *4* (6), 1903-1908.
142. Burnouf, D.; Ennifar, E.; Guedich, S.; Puffer, B.; Hoffmann, G.; Bec, G.; Disdier, F. o.; Baltzinger, M.; Dumas, P., kinITC: a new method for obtaining joint thermodynamic and kinetic data by isothermal titration calorimetry. *J. Am. Chem. Soc.* **2011**, *134* (1), 559-565.
143. Freire, E.; Mayorga, O. L.; Straume, M., Isothermal titration calorimetry. *Anal. Chem.* **1990**, *62* (18), 950A-959A.
144. Pethica, B. A., Misuse of thermodynamics in the interpretation of isothermal titration calorimetry data for ligand binding to proteins. *Anal. Biochem.* **2015**, *472*, 21-29.
145. Castellano, B. M.; Eggers, D. K., Experimental support for a desolvation energy term in governing equations for binding equilibria. *J. Phys. Chem.* **2013**, *117* (27), 8180-8188.
146. Borchers, C. H.; Marquez, V. E.; Schroeder, G. K.; Short, S. A.; Snider, M. J.; Speir, J. P.; Wolfenden, R., Fourier transform ion cyclotron resonance MS reveals the presence of a water molecule in an enzyme transition-state analogue complex. *Proc. Natl. Acad. Sci. U. S. A.* **2004**, *101* (43), 15341-15345.

147. Wada, T.; Mochizuki, A.; Higashiya, S.; Tsuruoka, H.; Kawahara, S.-i.; Ishikawa, M.; Sekine, M., Synthesis and properties of 2-azidodeoxyadenosine and its incorporation into oligodeoxynucleotides. *Tetrahedron Lett.* **2001**, *42* (52), 9215-9219.

148. Jawalekar, A. M.; Meeuwenoord, N.; Cremers, J. G. O.; Overkleeft, H. S.; van der Marel, G. A.; Rutjes, F. P. J. T.; van Delft, F. L., Conjugation of Nucleosides and Oligonucleotides by [3+2] Cycloaddition. *J. Org. Chem.* **2008**, *73* (1), 287-290.

149. Heindl, D. Preparation of LightCycler Red 640-labeled DNA. WO2007059912A1, 2007.

150. Kupryushkin, M. S.; Pyshnyi, D. V.; Stetsenko, D. A., Phosphoryl guanidines: a new type of nucleic Acid analogues. *Acta Naturae* **2014**, *6* (4), 116-118.

151. Prokhorova, D. V.; Chelobanov, B. P.; Burakova, E. A.; Fokina, A. A.; Stetsenko, D. A., New oligodeoxyribonucleotide derivatives bearing internucleotide N-tosyl phosphoramidate groups: Synthesis and complementary binding to DNA and RNA. *Russ. J. Bioorgan. Chem.* **2017**, *43* (1), 38-42.

152. Su, Y.; Fujii, H.; Burakova, E. A.; Chelobanov, B. P.; Fujii, M.; Stetsenko, D. A.; Filichev, V. V., Neutral and Negatively Charged Phosphate Modifications Altering Thermal Stability, Kinetics of Formation and Monovalent Ion Dependence of DNA G-Quadruplexes. *Chem. - Asian J.* **2019**, *14* (8), 1212-1220.

153. Miroshnichenko, S. K.; Patutina, O. A.; Burakova, E. A.; Chelobanov, B. P.; Fokina, A. A.; Vlassov, V. V.; Altman, S.; Zenkova, M. A.; Stetsenko, D. A., Mesyl phosphoramidate antisense oligonucleotides as an alternative to phosphorothioates with improved biochemical and biological properties. *Proc. Natl. Acad. Sci. U. S. A.* **2019**, *116* (4), 1229-1234.

154. Su, Y.; Edwards, P. J. B.; Stetsenko, D. A.; Filichev, V. V., The importance of phosphates for DNA G-quadruplex formation: evaluation of zwitterionic G-rich oligodeoxynucleotides. *ChemBioChem* **2020**, *21*, 2455-2466.

155. Kiviniemi, A.; Virta, P.; Lonnberg, H., Utilization of intrachain 4'-C-azidomethylthymidine for preparation of oligodeoxyribonucleotide conjugates by click chemistry in solution and on a solid support. *Bioconjugate Chem.* **2008**, *19* (8), 1726-1734.

156. Brown, S. D.; Graham, D., Conjugation of an oligonucleotide to Tat, a cell-penetrating peptide, via click chemistry. *Tetrahedron Lett.* **2010**, *51* (38), 5032-5034.

157. Pourceau, G.; Meyer, A.; Vasseur, J. J.; Morvan, F., Azide Solid Support for 3'-Conjugation of Oligonucleotides and Their Circularization by Click Chemistry. *J. Org. Chem.* **2009**, *74* (17), 6837-6842.

158. Santner, T.; Hartl, M.; Bister, K.; Micura, R., Efficient Access to 3'-Terminal Azide-Modified RNA for Inverse Click-Labeling Patterns. *Bioconjugate Chem.* **2014**, *25* (1), 188-195.

159. Jentzsch, E.; Mokhir, A., A Fluorogenic, Nucleic Acid Directed "Click" Reaction. *Inorg. Chem.* **2009**, *48* (20), 9593-9595.

160. Fomich, M. A.; Kvach, M. V.; Navakouski, M. J.; Weise, C.; Baranovsky, A. V.; Korshun, V. A.; Shmanai, V. V., Azide Phosphoramidite in Direct Synthesis of Azide-Modified Oligonucleotides. *Org. Lett.* **2014**, *16* (17), 4590-4593.

161. Yang, H.; Tang, J. A.; Greenberg, M. M., Synthesis of Oligonucleotides Containing the N(6)-(2-Deoxy-α,β-D-erythropentofuranosyl)-2,6-diamino-4-hydroxy-5-formamidopy rimidine (FapydG) Oxidative Damage Product Derived from 2'-Deoxyguanosine. *Chem.—Eur. J.* **2020**, *26* (24), 5441-5448.

162. Štimac, A.; Kobe, J., Stereoselective synthesis of 1,2-cis- and 2-deoxyglycofuranosyl azides from glycosyl halides. *Carbohydr. Res.* **2000**, *329* (2), 317-324.

163. Damha, M. J.; Giannaris, P. A.; Zabarylo, S. V., An improved procedure for derivatization of controlled-pore glass beads for solid-phase oligonucleotide synthesis. *Nucleic Acids Res.* **1990**, *18* (13), 3813-3821.

164. Tararov, V. I.; Kolyachkina, S. V.; Alexeev, C. S.; Mikhailov, S. N., N6-Acetyl-2',3',5'-tri-*O*-acetyladenosine; a convenient, ‘missed out’ substrate for regioselective N6-alkylations. *Synthesis* **2011**, (15), 2483-2489.

165. Ottria, R.; Casati, S.; Baldoli, E.; Maier, J. A.; Ciuffreda, P., N6-Alkyladenosines: Synthesis and evaluation of in vitro anticancer activity. *Biorg. Med. Chem.* **2010**, *18* (23), 8396-8402.

166. Sugai, N.; Heguri, H.; Ohta, K.; Meng, Q.; Yamamoto, T.; Tezuka, Y., Effective Click Construction of Bridged- and Spiro-Multicyclic Polymer Topologies with Tailored Cyclic Prepolymers (kyklo-Telechelics). *J. Am. Chem. Soc.* **2010**, *132* (42), 14790-14802.

167. Miah, A.; Reese, C. B.; Song, Q., Convenient intermediates for the preparation of C-4 modified derivatives of pyrimidine nucleosides. *Nucleosides Nucleotides Nucl. Acids* **1997**, *16* (1-2), 53-65.

168. Budhathoki-Uprety, J.; Jena, P. V.; Roxbury, D.; Heller, D. A., Helical polycarbodiimide cloaking of carbon nanotubes enables inter-nanotube exciton energy transfer modulation. *J. Am. Chem. Soc.* **2014**, *136* (44), 15545-15550.

169. Mouscadet, J.-F.; Ketterle, C.; Goulaouic, H.; Carteau, S.; Subra, F.; Le Bret, M.; Auclair, C., Triple helix formation with short oligonucleotide-intercalator conjugates matching the HIV-1 U3 LTR end sequence. *Biochemistry* **1994**, *33* (14), 4187-4196.

170. Nakamura, S.; Kawabata, H.; Fujimoto, K., Double duplex invasion of DNA induced by ultrafast photo-cross-linking using 3-cyanovinylcarbazole for antigene methods. *Chem. Commun.* **2017**, *53* (54), 7616-7619.

171. Fujimoto, K.; Hiratsuka-Konishi, K.; Sakamoto, T.; Ohtake, T.; Shinohara, K.-i.; Yoshimura, Y., Specific and reversible photochemical labeling of plasmid DNA using photoresponsive oligonucleotides containing 3-cyanovinylcarbazole. *Mol. Biosyst.* **2012**, *8* (2), 491-494.

172. Gouda, A. S.; Przypis, Ł.; Walczak, K.; Jørgensen, P. T.; Wengel, J., Carbazole modified oligonucleotides: synthesis, hybridization studies and fluorescence properties. *Org. Biomol. Chem.* **2020**, *18* (35), 6935-6948.

173. Matteucci, M. D.; von Krosigk, U., Hybridization properties of oligonucleotides bearing a tricyclic 2'-deoxycytidine analog based on a carbazole ring system. *Tetrahedron Lett.* **1996**, *37* (29), 5057-5060.

174. S Tsutsumi, L.; Gündisch, D.; Sun, D., Carbazole scaffold in medicinal chemistry and natural products: a review from 2010-2015. *Curr. Top. Med. Chem.* **2016**, *16* (11), 1290-1313.

175. Głuszyńska, A., Biological potential of carbazole derivatives. *Eur. J. Med. Chem.* **2015**, *94*, 405-426.

176. Issa, S.; Prandina, A.; Bedel, N.; Rongved, P.; Yous, S.; Le Borgne, M.; Bouaziz, Z., Carbazole scaffolds in cancer therapy: a review from 2012 to 2018. *J. Enzyme Inhib. Med. Chem.* **2019**, *34* (1), 1321-1346.

177. Ingale, S. A.; Mei, H.; Leonard, P.; Seela, F., Ethynyl side chain hydration during synthesis and workup of "clickable" oligonucleotides: bypassing acetyl group formation by triisopropylsilyl protection. *J. Org. Chem.* **2013**, *78* (22), 11271-11282.

178. Fatthalla, M. I.; Pedersen, E. B., Unexpected Hydration of a Triple Bond During DNA Synthesis: Conjugating 3-(Pyren-1-ylethynyl) indole to DNA for Triplex Studies. *Eur. J. Org. Chem.* **2016**, *2016* (21), 3528-3535.

179. Ingale, S. A.; Leonard, P.; Yang, H.; Seela, F., 5-Nitroindole oligonucleotides with alkynyl side chains: Universal base pairing, triple bond hydration and properties of pyrene "click" adducts. *Org. Biomol. Chem.* **2014**, *12* (42), 8519-8532.

180. Barchi, J. J., Jr.; Musser, S.; Marquez, V. E., The decomposition of 1-(β-D-ribofuranosyl)-1,2-dihydropyrimidin-2-one (zebularine) in alkali: mechanism and products. *J. Org. Chem.* **1992**, *57* (2), 536-41.

181. Myers, A. G.; Goldberg, S. D., Synthesis of the Kedarcidin Core Structure by a Transannular Cyclization Pathway. *Angew. Chem. Int. Ed. Engl.* **2000**, *39* (15), 2732-2735.

182. Barchi, J. J.; Haces, A.; Marquez, V. E.; McCormack, J. J., Inhibition of Cytidine Deaminase by Derivatives of 1-(β-D-Ribofuranosyl)-Dihydropyrimidin-2-One (Zebularine). *Nucleosides Nucleotides* **1992**, *11* (10), 1781-1793.

183. Marquez, V. E.; Barchi, J. J., Jr.; Kelley, J. A.; Rao, K. V.; Agbaria, R.; Ben-Kasus, T.; Cheng, J. C.; Yoo, C. B.; Jones, P. A., Zebularine: a unique molecule for an epigenetically based strategy in cancer chemotherapy. The magic of its chemistry and biology. *Nucleosides Nucleotides Nucleic Acids* **2005**, *24* (5-7), 305-318.

184. Ashley, G. W.; Bartlett, P., Inhibition of Escherichia coli cytidine deaminase by a phosphapyrimidine nucleoside. *J. Biol. Chem.* **1984**, *259* (21), 13621-13627.

185. Chung, S. J.; Fromme, J. C.; Verdine, G. L., Structure of human cytidine deaminase bound to a potent inhibitor. *J. Med. Chem.* **2005**, *48* (3), 658-660.

186. Losey, H. C.; Ruthenburg, A. J.; Verdine, G. L., Crystal structure of Staphylococcus aureus tRNA adenosine deaminase TadA in complex with RNA. *Nat. Struct. Mol. Biol.* **2006**, *13* (2), 153-159.

187. Dickey, T. H.; McKercher, M. A.; Wuttke, D. S., Nonspecific recognition is achieved in Pot1pC through the use of multiple binding modes. *Structure* **2013**, *21* (1), 121-132.
188. Dickey, T. H.; Wuttke, D. S., The telomeric protein Pot1 from *Schizosaccharomyces* pombe binds ssDNA in two modes with differing 3' end availability. *Nucleic Acids Res.* **2014**, *42* (15), 9656-9665.
189. Sachs, R.; Max, K. E.; Heinemann, U.; Balbach, J., RNA single strands bind to a conserved surface of the major cold shock protein in crystals and solution. *RNA* **2012**, *18* (1), 65-76.
190. Max, K. E.; Zeeb, M.; Bienert, R.; Balbach, J.; Heinemann, U., T-rich DNA single strands bind to a preformed site on the bacterial cold shock protein Bs-CspB. *J. Mol. Biol.* **2006**, *360* (3), 702-714.
191. Vives, M.; Eritja, R.; Tauler, R.; Marquez, V.; Gargallo, R., Synthesis, stability, and protonation studies of a self-complementary dodecamer containing the modified nucleoside 2'-deoxyzebularine. *Biopolymers* **2004**, *73* (1), 27-43.
192. Fall, A.; Seck, I.; Diouf, O.; Gaye, M.; Seck, M.; Gomez, G.; Fall, Y., Synthesis of new triazolium-based ionic liquids and their use in the Morita–Baylis–Hillman reaction. *Tetrahedron Lett.* **2015**, *56* (36), 5128-5131.
193. Miah, A.; Reese, C. B.; Song, Q., Convenient Intermediates for the Preparation of C-4 Modified Derivatives of Pyrimidine Nucleosides. *Nucleosides Nucleotides* **1997**, *16* (1-2), 53-65.
194. Bonesi, S. M.; Erra‐Balsells, R., Synthesis and isolation of iodocarbazoles. Direct iodination of carbazoles by N-iodosuccinimide and N-iodosuccinimide-silica gel system. *J. Heterocycl. Chem.* **2001**, *38* (1), 77-87.
195. Cacciamani, T.; Vita, A.; Cristalli, G.; Vincenzetti, S.; Natalini, P.; Ruggieri, S.; Amici, A.; Magni, G., Purification of human cytidine deaminase: molecular and enzymatic characterization and inhibition by synthetic pyrimidine analogs. *Arch. Biochem. Biophys.* **1991**, *290* (2), 285-292.
196. Ko, T.-P.; Lin, J.-J.; Hu, C.-Y.; Hsu, Y.-H.; Wang, A. H.-J.; Liaw, S.-H., Crystal structure of yeast cytosine deaminase: insights into enzyme mechanism and evolution. *J. Biol. Chem.* **2003**, *278* (21), 19111-19117.
197. Schroeder, G. K.; Wolfenden, R., Rates of spontaneous disintegration of DNA and the rate enhancements produced by DNA glycosylases and deaminases. *Biochemistry* **2007**, *46* (47), 13638-13647.
198. Tarkin-Tas, E.; Mathias, L. J., Synthesis and Ring-Opening Polymerization of 5-Azepane-2-one Ethylene Ketal: A New Route to Functional Aliphatic Polyamides. *Macromolecules* **2010**, *43* (2), 968-974.
199. Kotick, M. P.; Szantay, C.; Bardos, T. J., Synthesis of 5-s-substituted 2'-deoxyuridines. Study of the factors influencing the stereoselectivity of the silyl modification of the Hilbert-Johnson reaction. *J. Org. Chem.* **1969**, *34* (12), 3806-3813.

200. Marquez, V. E.; Liu, P. S.; Kelley, J. A.; Driscoll, J. S., 1, 3-Diazepinones. 1. Synthesis of 5-hydroxyperhydro-1, 3-diazepin-2-one. *J. Org. Chem.* **1980**, *45* (3), 485-489.
201. Schunk, S.; Linz, K.; Frormann, S.; Hinze, C.; Oberbörsch, S.; Sundermann, B.; Zemolka, S.; Englberger, W.; Germann, T.; Christoph, T., Discovery of spiro [cyclohexane-dihydropyrano [3, 4-b] indole]-amines as potent NOP and opioid receptor agonists. *ACS Med. Chem. Lett.* **2014**, *5* (8), 851-856.
202. Perveen, S.; Abdul Hai, S. M.; Khan, R. A.; Khan, K. M.; Afza, N.; Sarfaraz, T. B., Expeditious Method for Synthesis of Symmetrical 1, 3-Disubstituted Ureas and Thioureas. *Synth. Commun.* **2005**, *35* (12), 1663-1674.
203. Venkatesan, S.; Rosenthal, R.; Kanu, N.; McGranahan, N.; Bartek, J.; Quezada, S.; Hare, J.; Harris, R.; Swanton, C., Perspective: APOBEC mutagenesis in drug resistance and immune escape in HIV and cancer evolution. *Ann. Oncol.* **2018**, *29* (3), 563-572.
204. Wang, S.; Jia, M.; He, Z.; Liu, X.-S., APOBEC3B and APOBEC mutational signature as potential predictive markers for immunotherapy response in non-small cell lung cancer. *Oncogene* **2018**, *37* (29), 3924-3936.
205. Zou, J.; Wang, C.; Ma, X.; Wang, E.; Peng, G., APOBEC3B, a molecular driver of mutagenesis in human cancers. *Cell Biosci.* **2017**, *7* (1), 22-29.
206. Pourceau, G.; Meyer, A.; Vasseur, J.-J.; Morvan, F., Azide solid support for 3'-conjugation of oligonucleotides and their circularization by click chemistry. *J. Org. Chem.* **2009**, *74* (17), 6837-6842.
207. Sands, H.; Gorey-Feret, L. J.; Cocuzza, A. J.; Hobbs, F. W.; Chidester, D.; Trainor, G. L., Biodistribution and metabolism of internally 3H-labeled oligonucleotides. I. Comparison of a phosphodiester and a phosphorothioate. *Mol. Pharmacol.* **1994**, *45* (5), 932-943.
208. Nukaga, Y.; Oka, N.; Wada, T., Stereocontrolled Solid-Phase Synthesis of Phosphate/Phosphorothioate (PO/PS) Chimeric Oligodeoxyribonucleotides on an Automated Synthesizer Using an Oxazaphospholidine–Phosphoramidite Method. *J. Org. Chem.* **2016**, *81* (7), 2753-2762.
209. Iwamoto, N.; Butler, D. C.; Svrzikapa, N.; Mohapatra, S.; Zlatev, I.; Sah, D. W.; Standley, S. M.; Lu, G.; Apponi, L. H.; Frank-Kamenetsky, M., Control of phosphorothioate stereochemistry substantially increases the efficacy of antisense oligonucleotides. *Nat. Biotechnol.* **2017**, *35* (9), 845-851.
210. Su, Y.; Bayarjargal, M.; Hale, T. K.; Filichev, V. V., DNA with zwitterionic and negatively charged phosphate modifications: Formation of DNA triplexes, duplexes and cell uptake studies. *Beilstein J. Org. Chem.* **2021**, *17* (1), 749-761.

Chapter 7. Electronic Supplementary Information

7.1. Experimental data for compounds synthesized

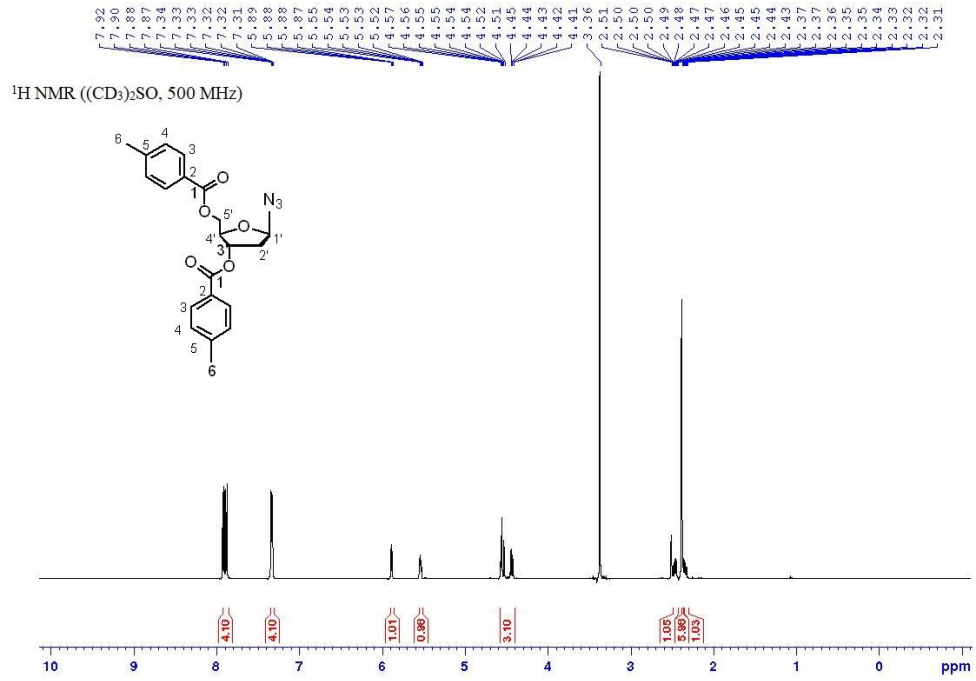


Figure S1. ^{1}H NMR spectrum of compound **2**.

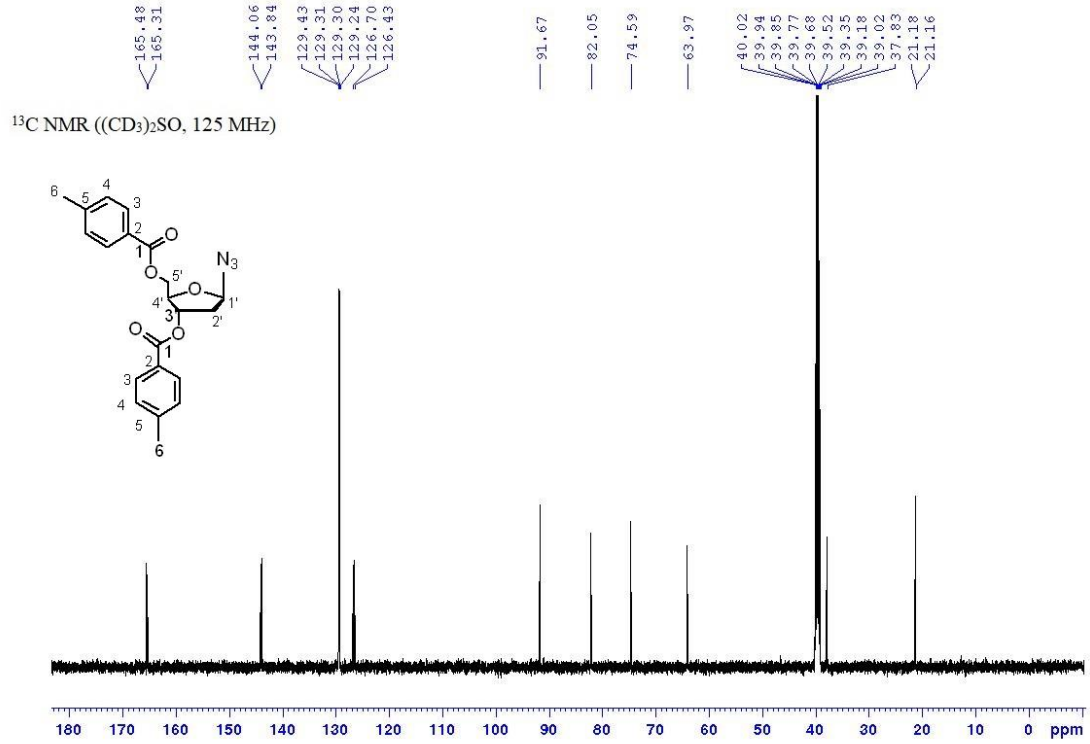


Figure S2. ^{13}C NMR spectrum of compound **2**.

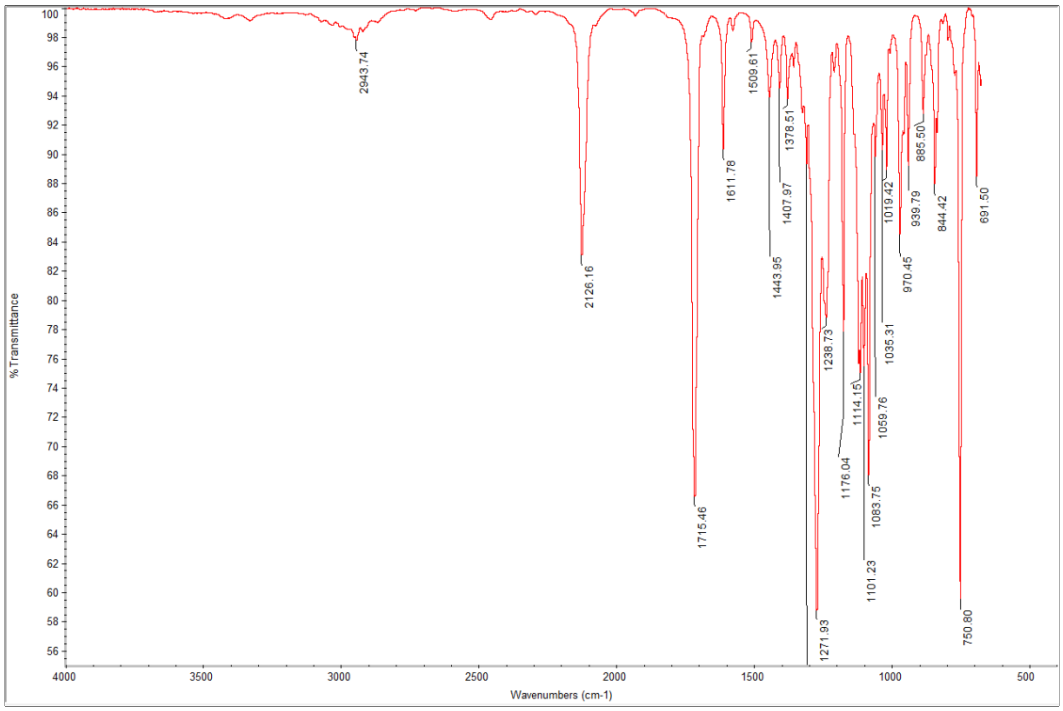


Figure S3. IR spectrum of compound **2**.

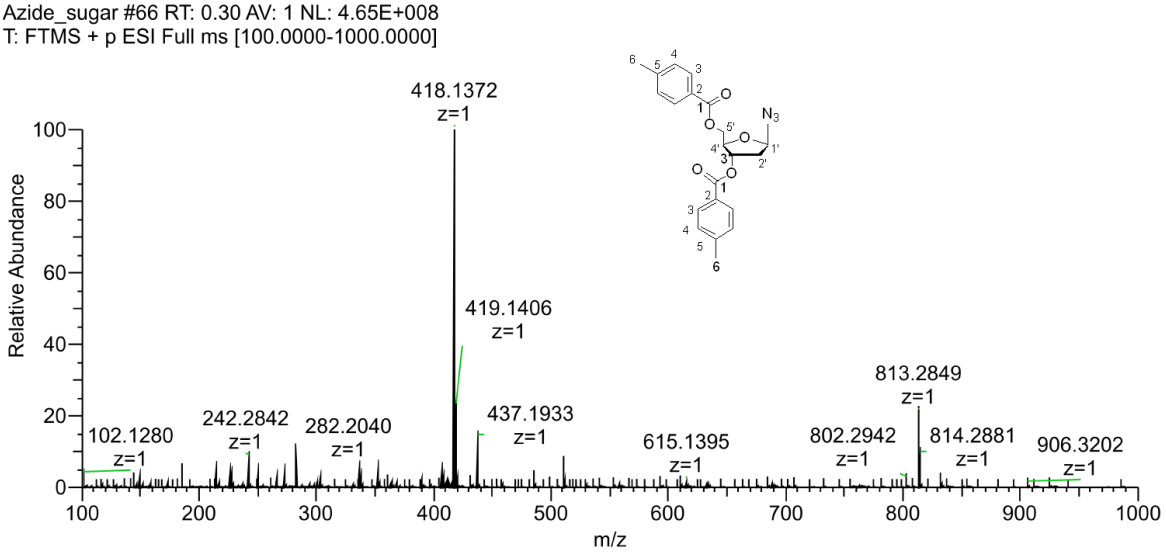


Figure S4. HRMS (ESI) of compound **2**.

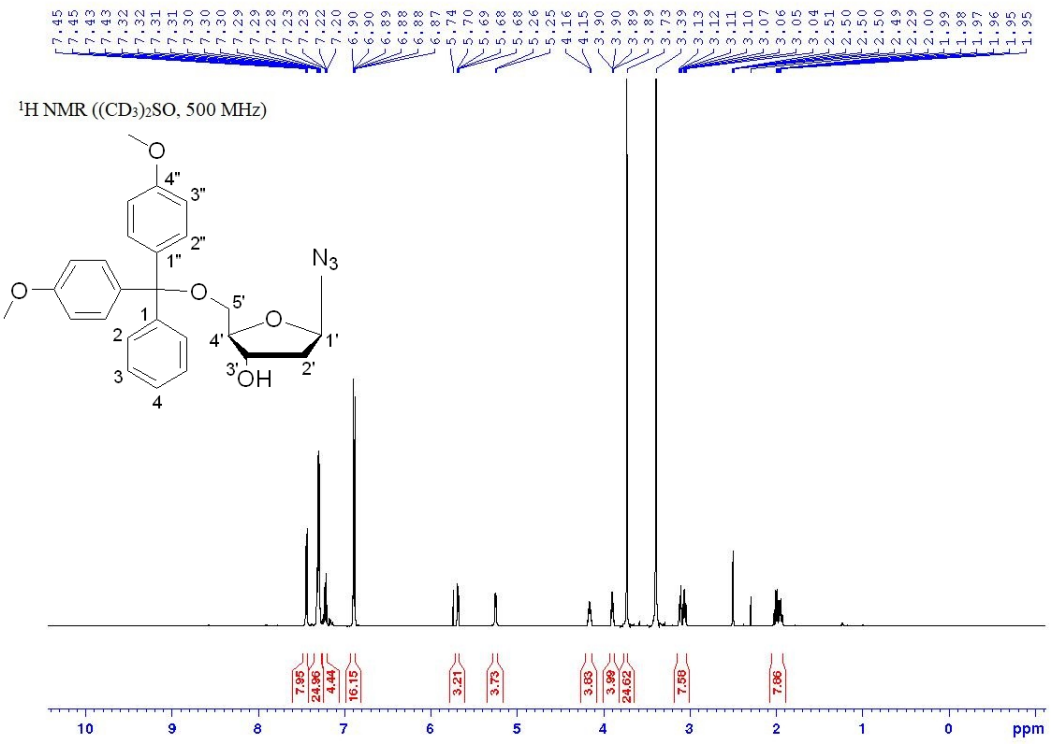


Figure S5. ^{1}H NMR of compound **3**.

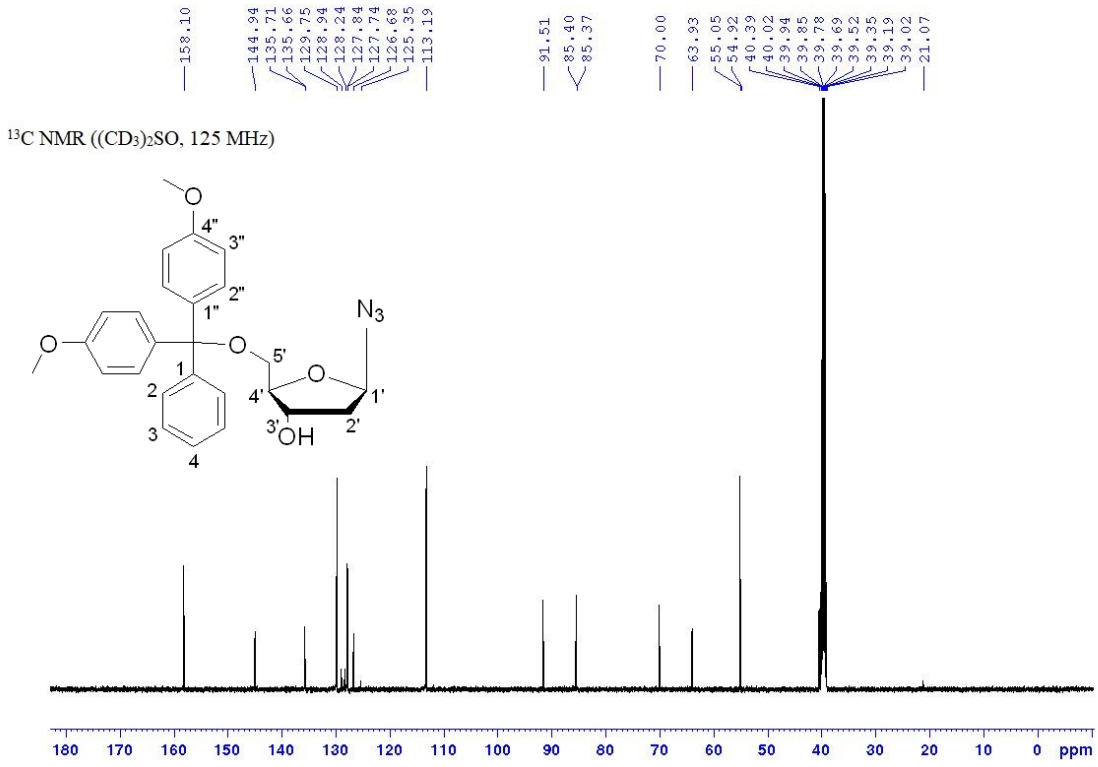


Figure S6. ^{13}C NMR of compound **3**.

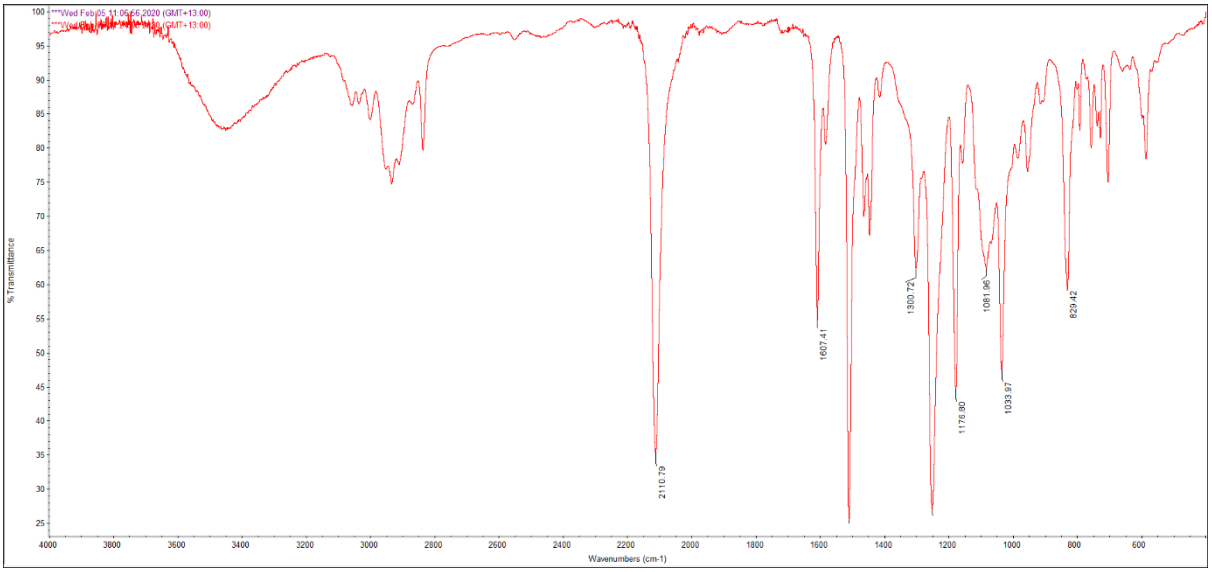


Figure S7. IR (ATR) spectrum of compound **3**.

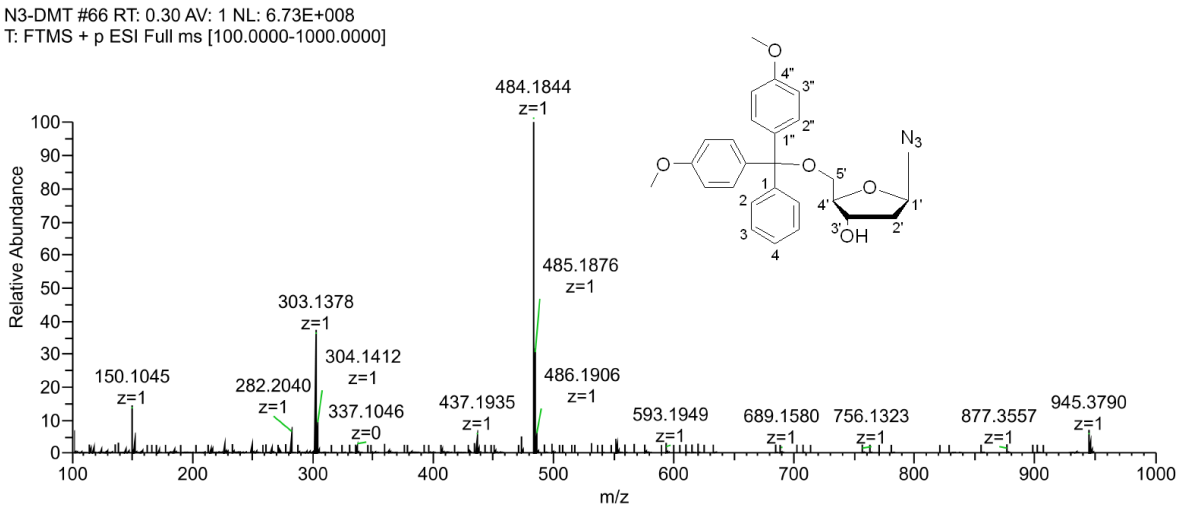


Figure S8. HRMS (ESI) of compound **3**.

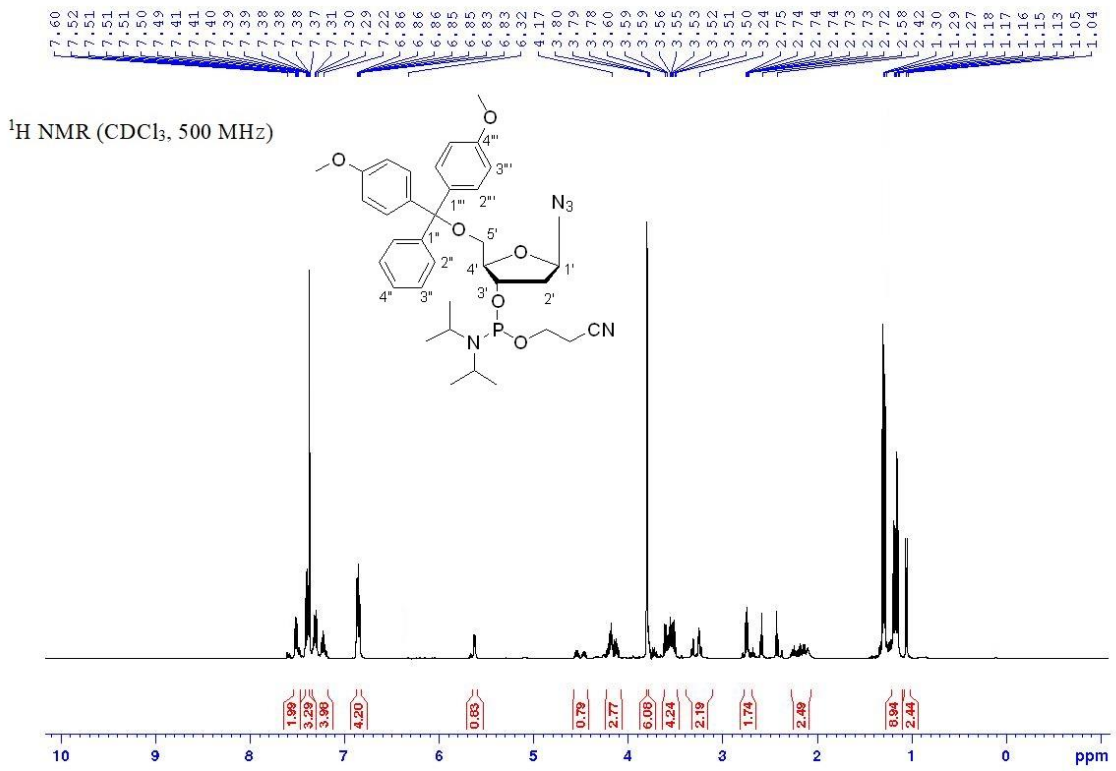


Figure S9. ^{1}H NMR of compound **5**.

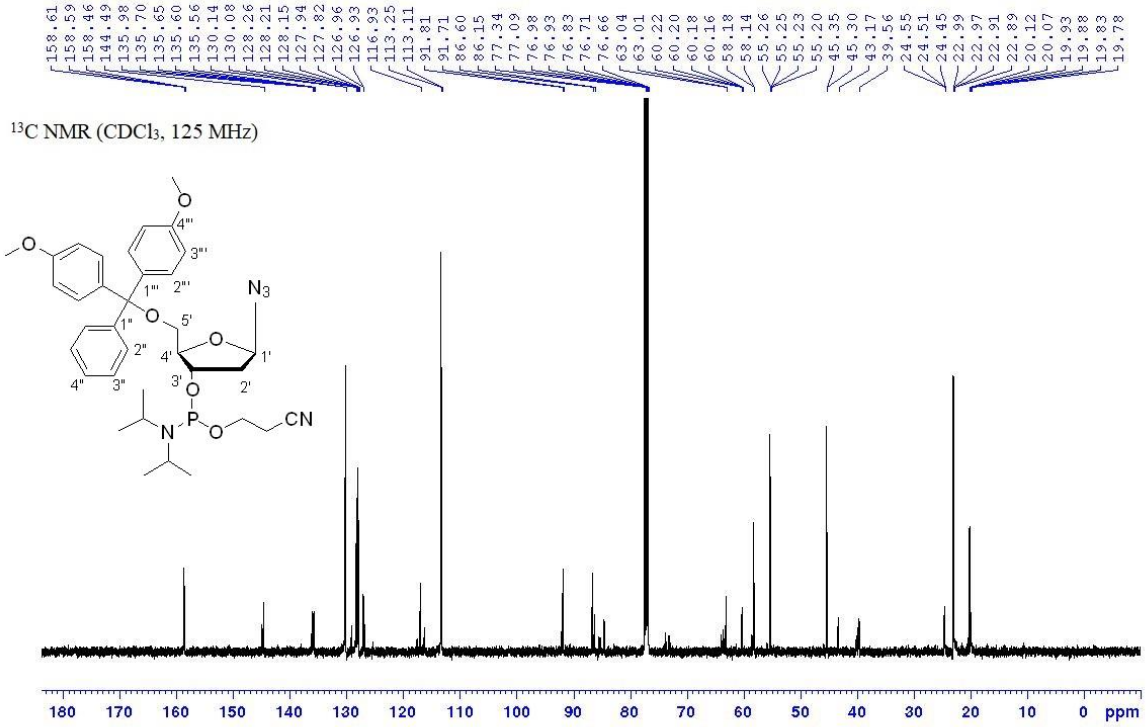


Figure S10. ^{13}C NMR of compound **5**.

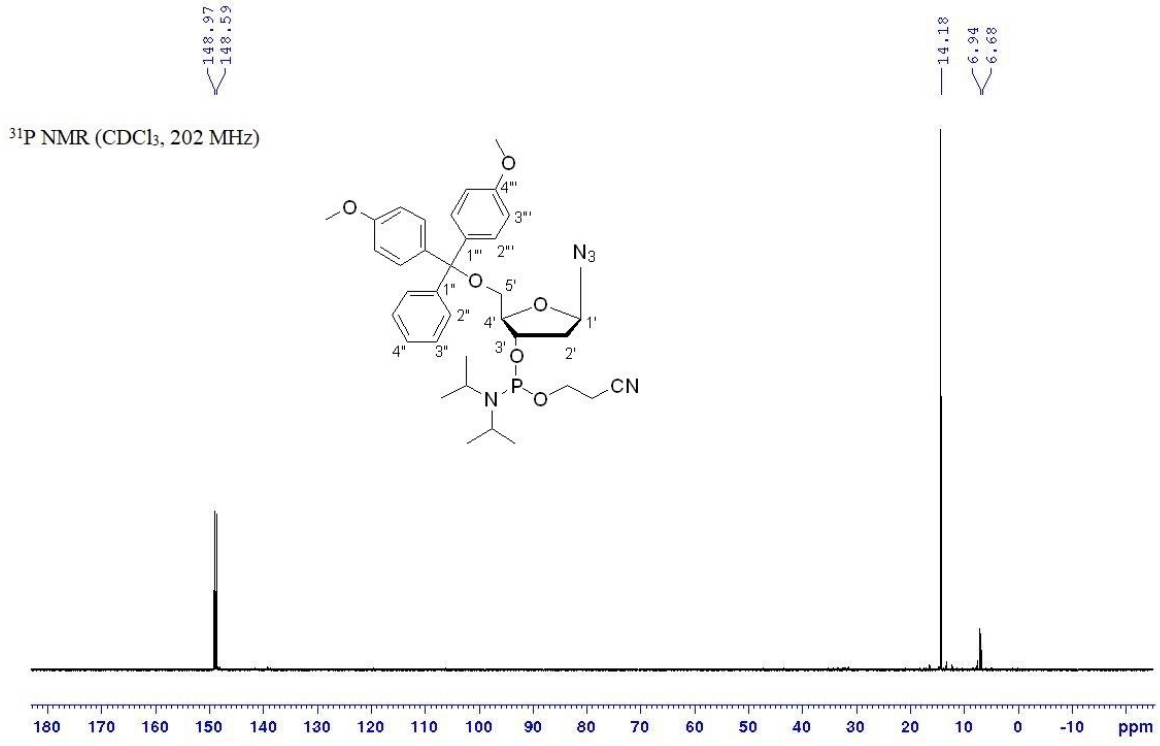


Figure S11. ^{31}P NMR of compound **5**.

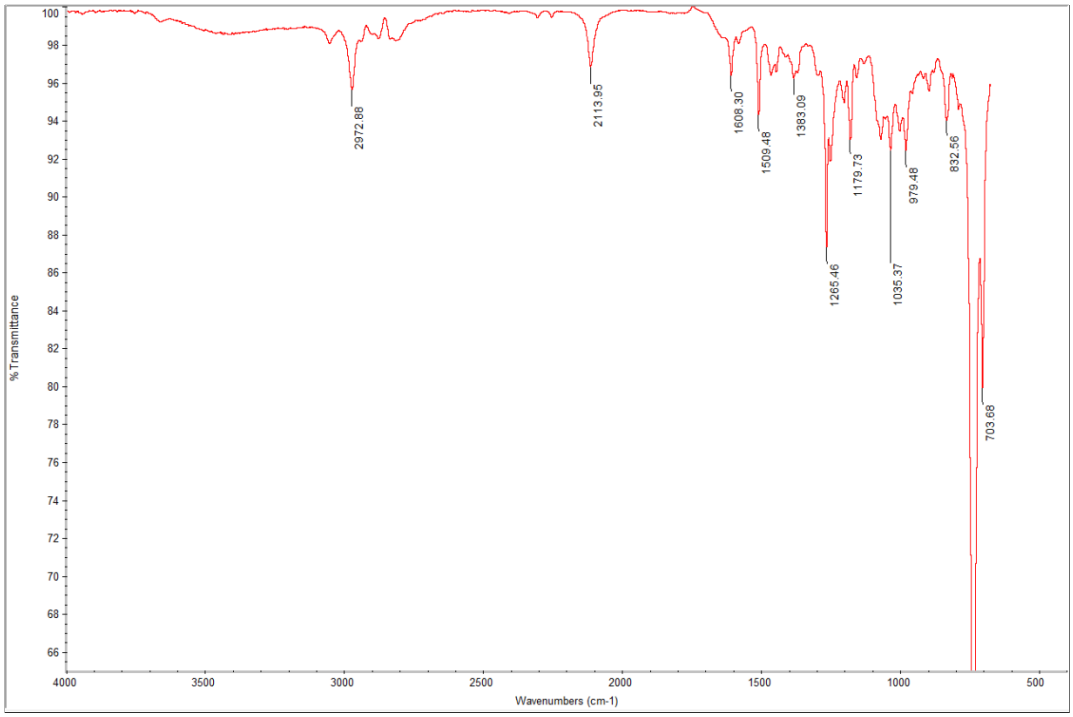


Figure S12. IR spectrum of compound **5**.

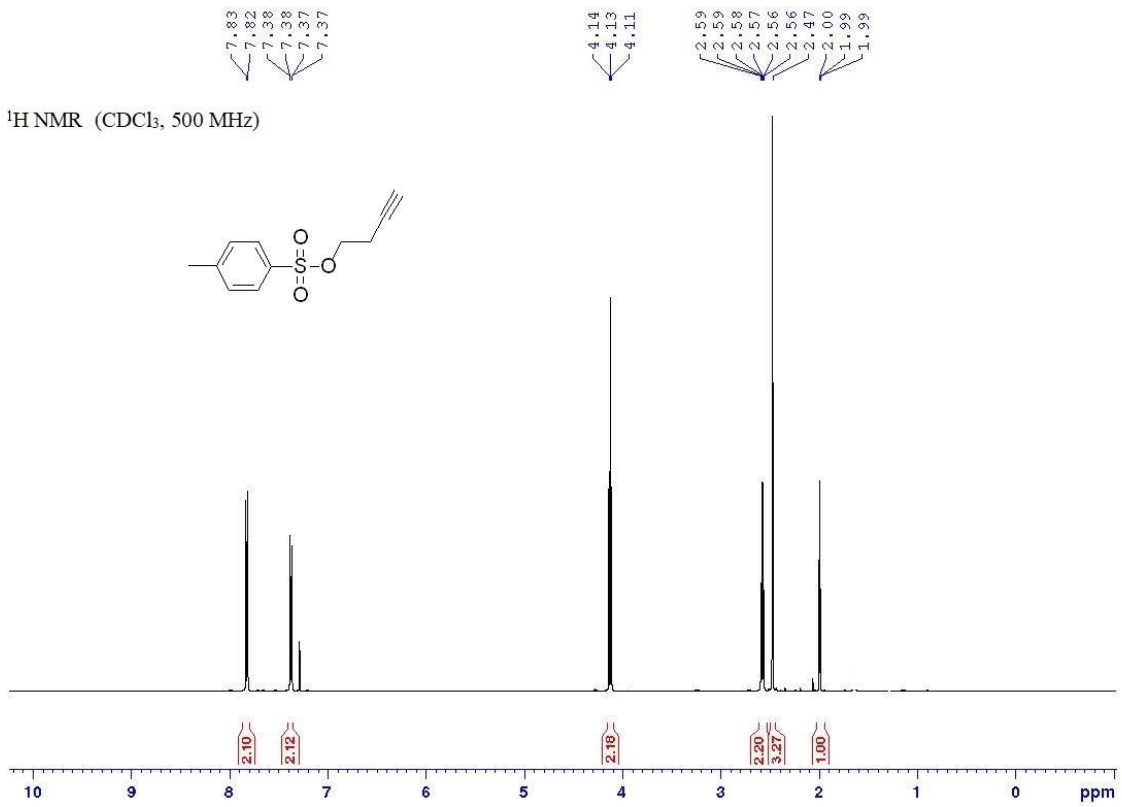


Figure S13. ^{1}H NMR of compound **6**.

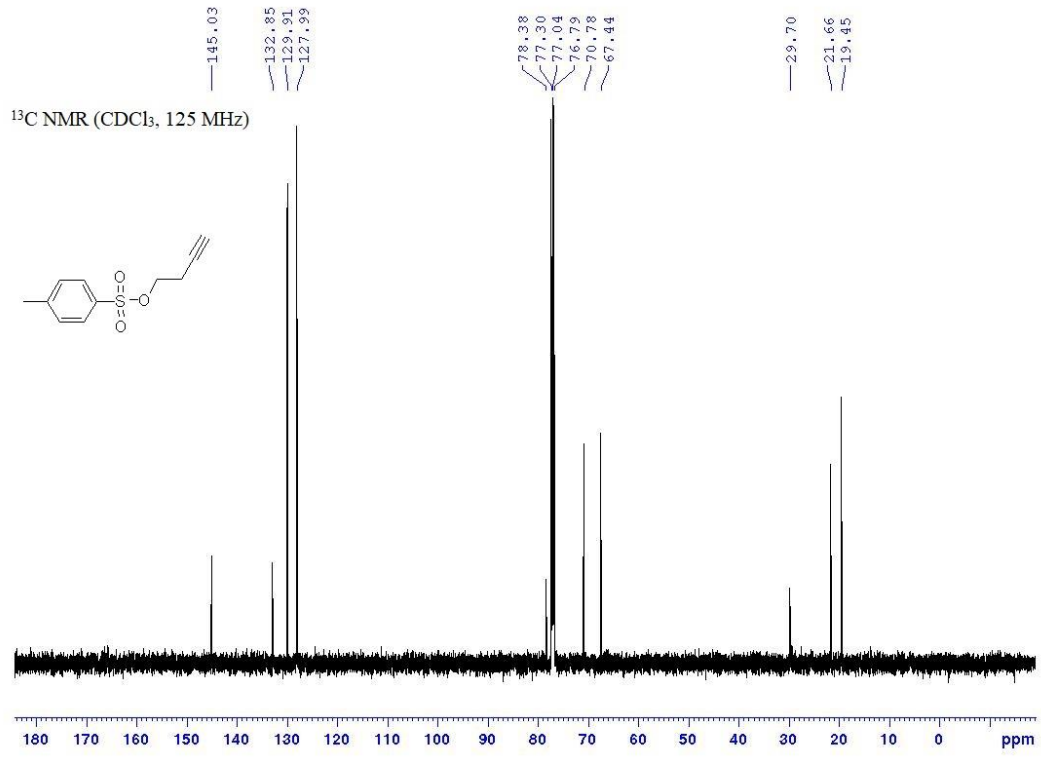


Figure S14. ^{13}C NMR of compound **6**.

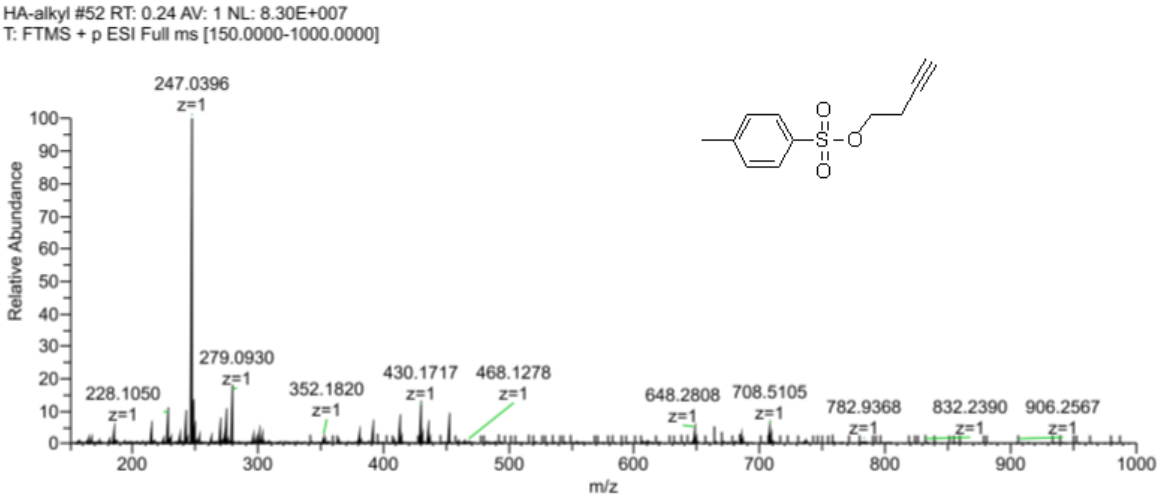


Figure S15. HRMS (ESI) of compound **6**.

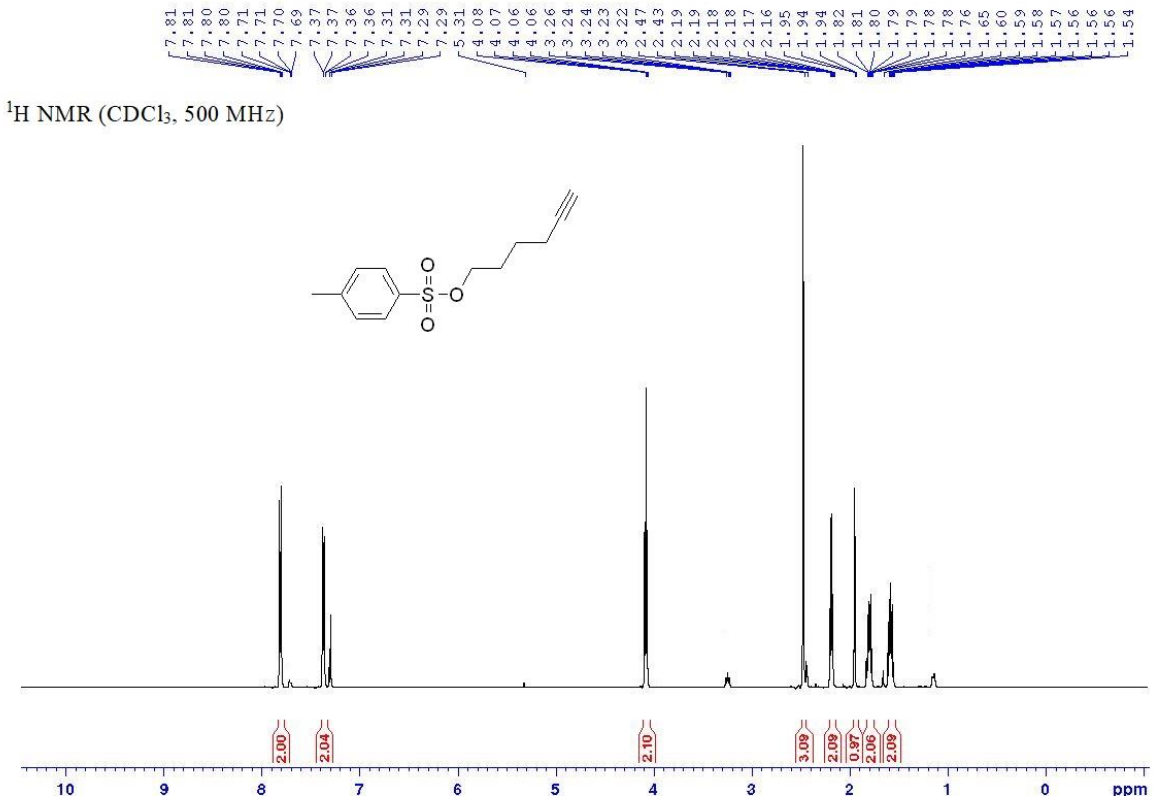


Figure S16. ^{1}H NMR of compound **7**.

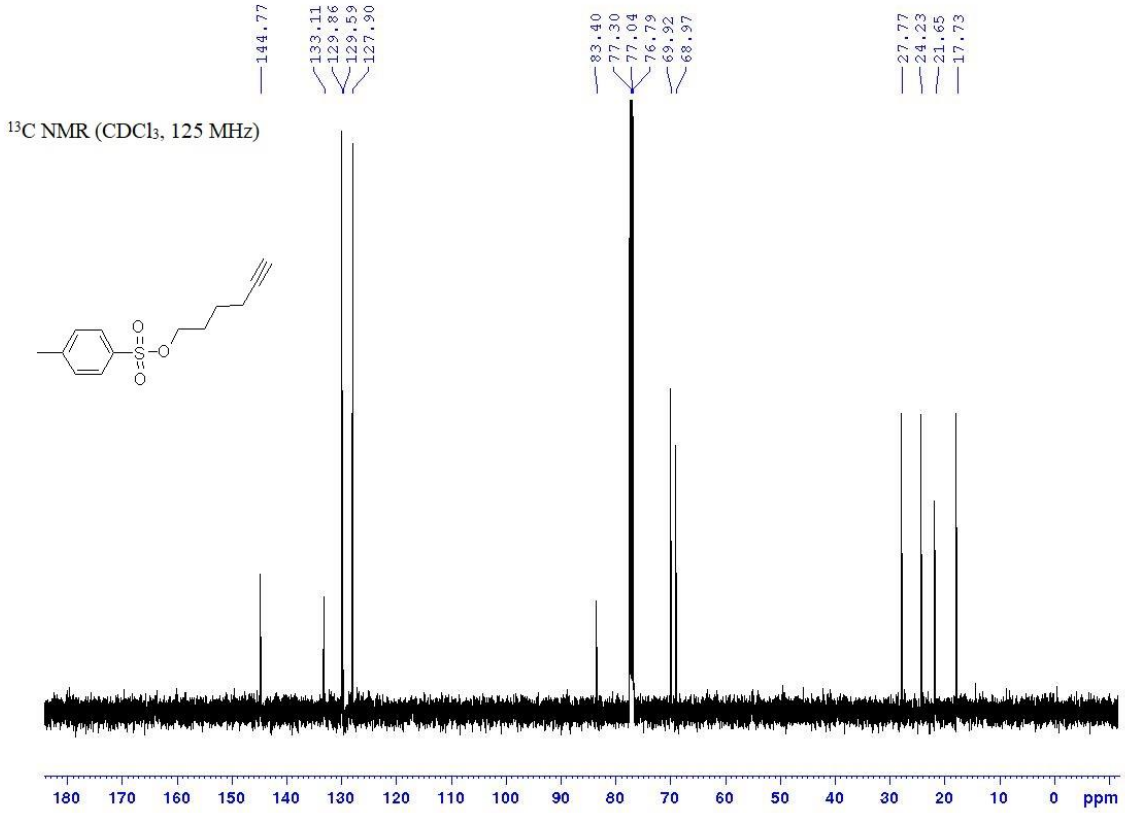


Figure S17. ^{13}C NMR of compound **7.**

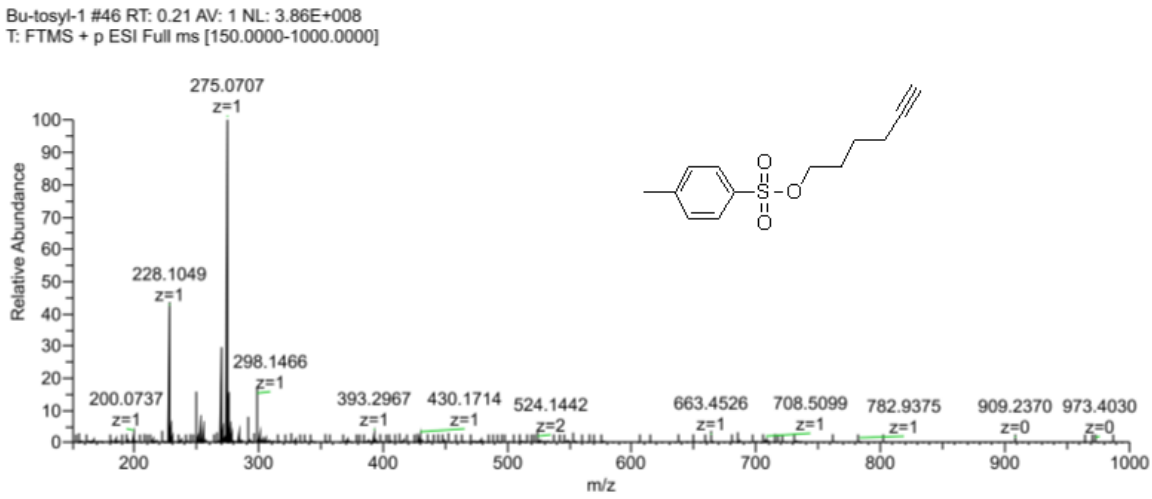


Figure S18. HRMS (ESI) of compound **7.**

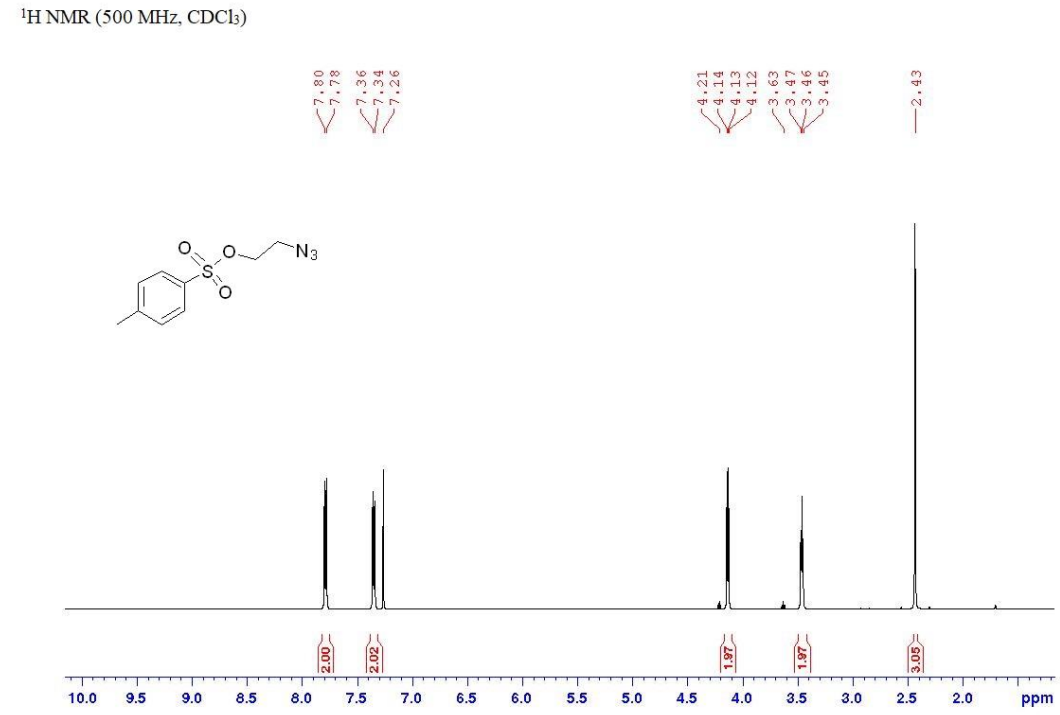


Figure S19. ^{1}H NMR spectrum of compound **8**.

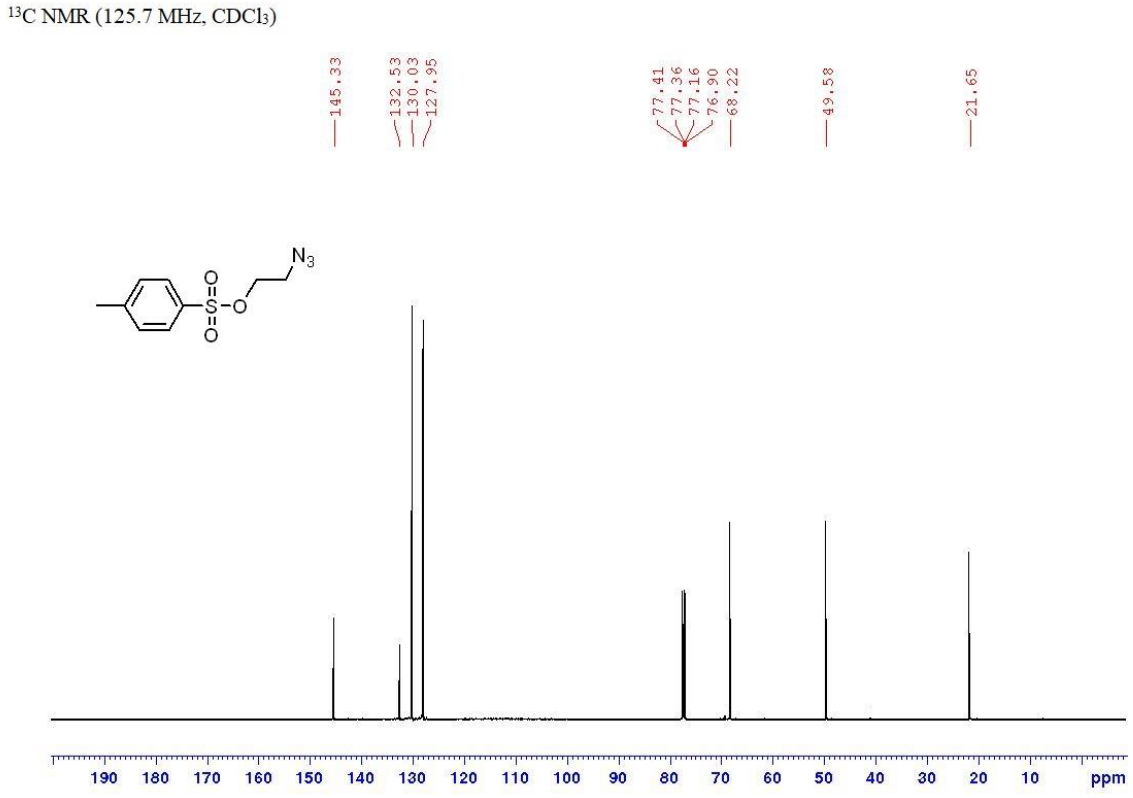


Figure S20. ^{13}C NMR spectrum of compound **8**.

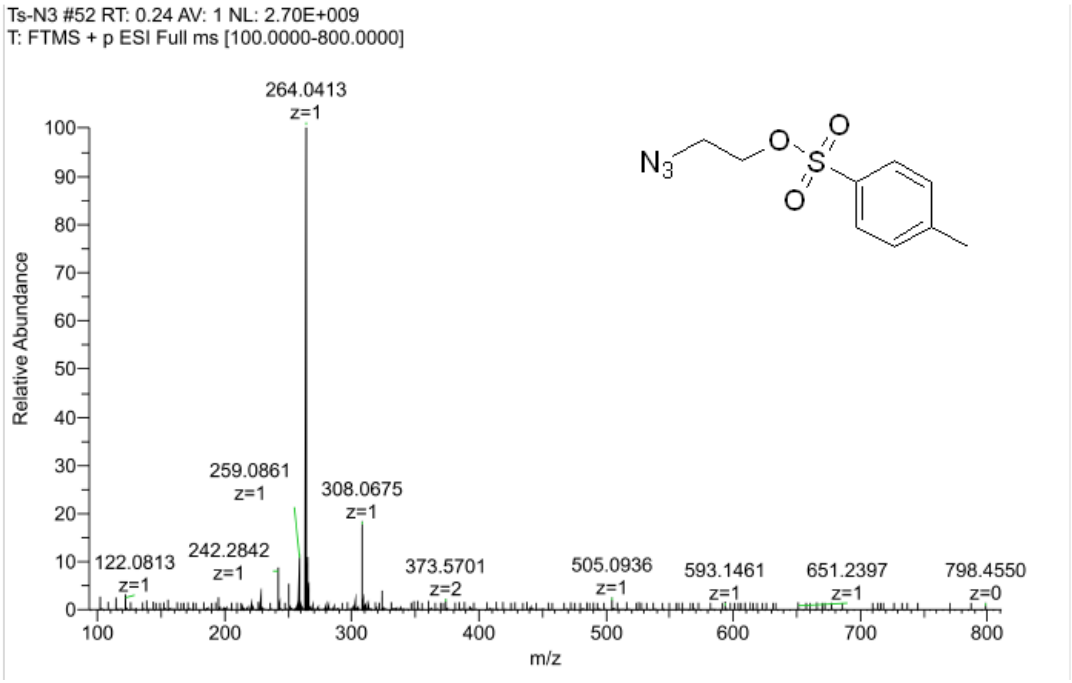


Figure S21. HRMS (ESI) of compound **8**.

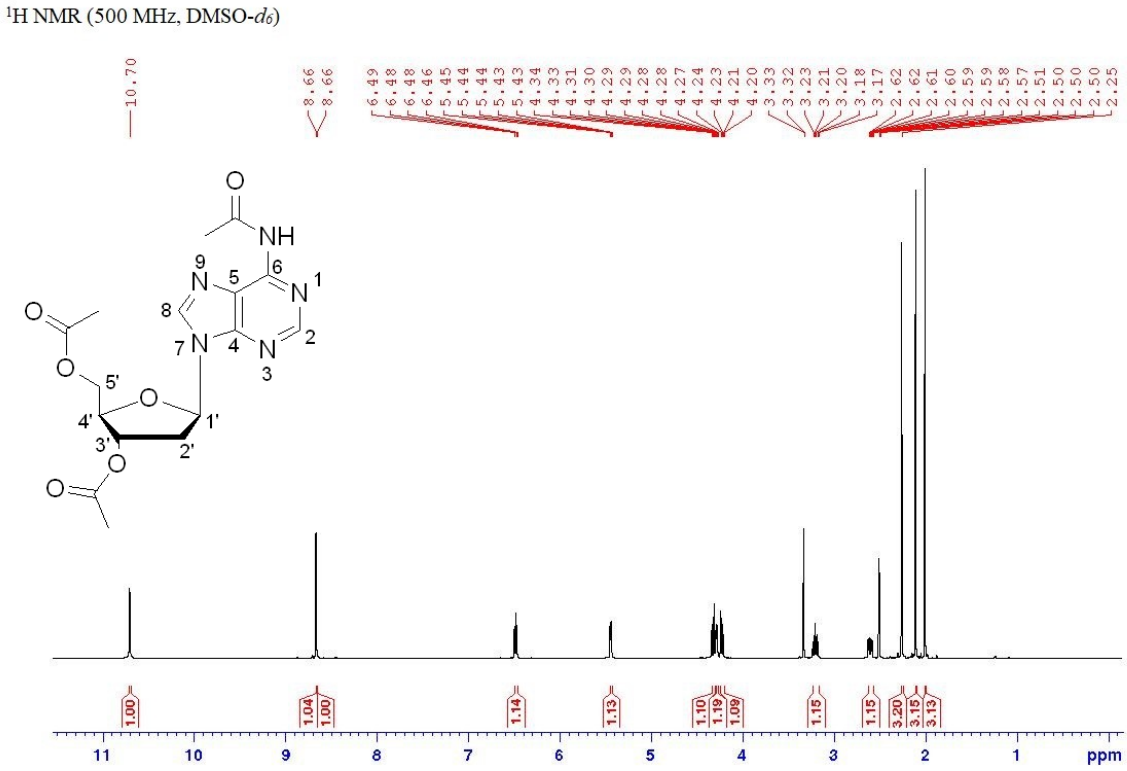


Figure S22. ^{1}H NMR spectrum of compound **10**.

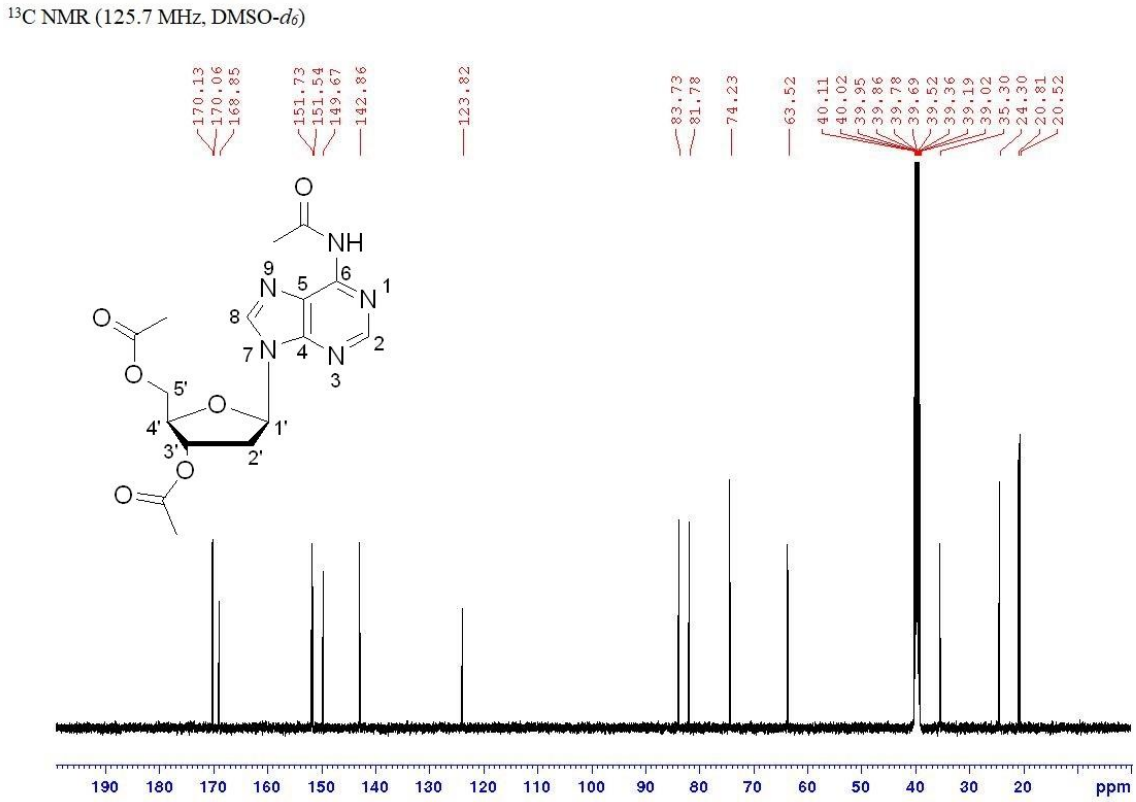


Figure S23. ^{13}C NMR spectrum of compound **10**.

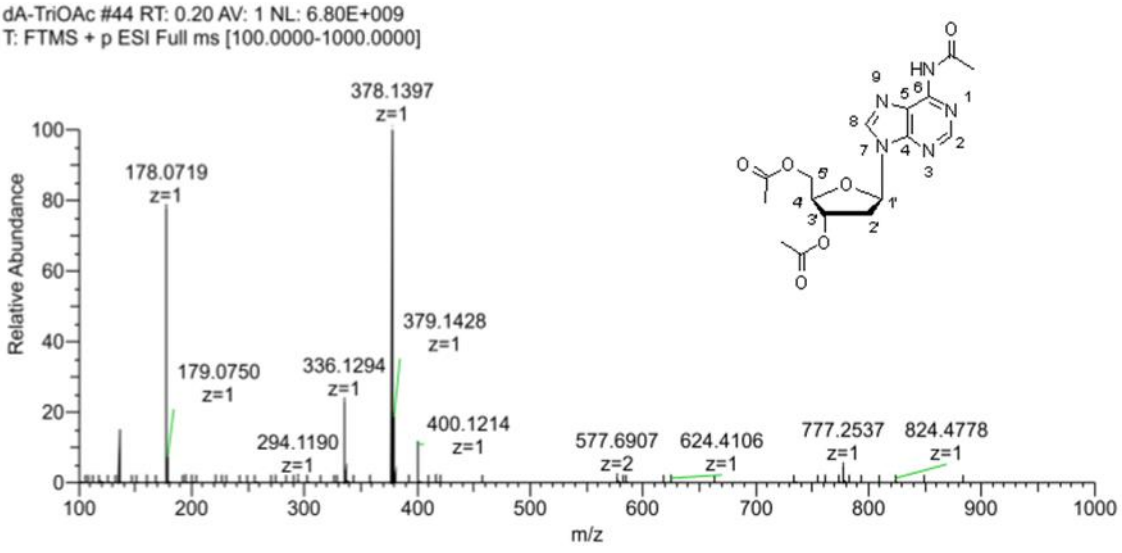


Figure S24. HRMS (ESI) of compound **10**.

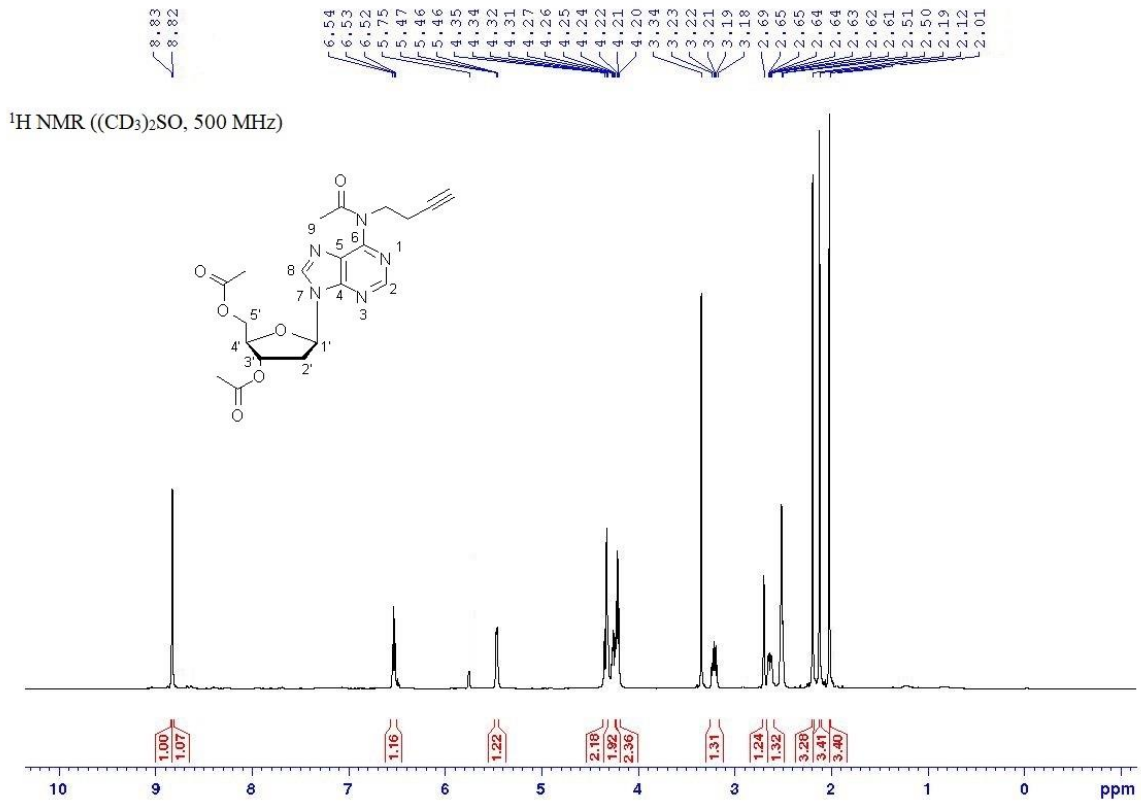


Figure S25. ^{1}H NMR of compound **11a**.

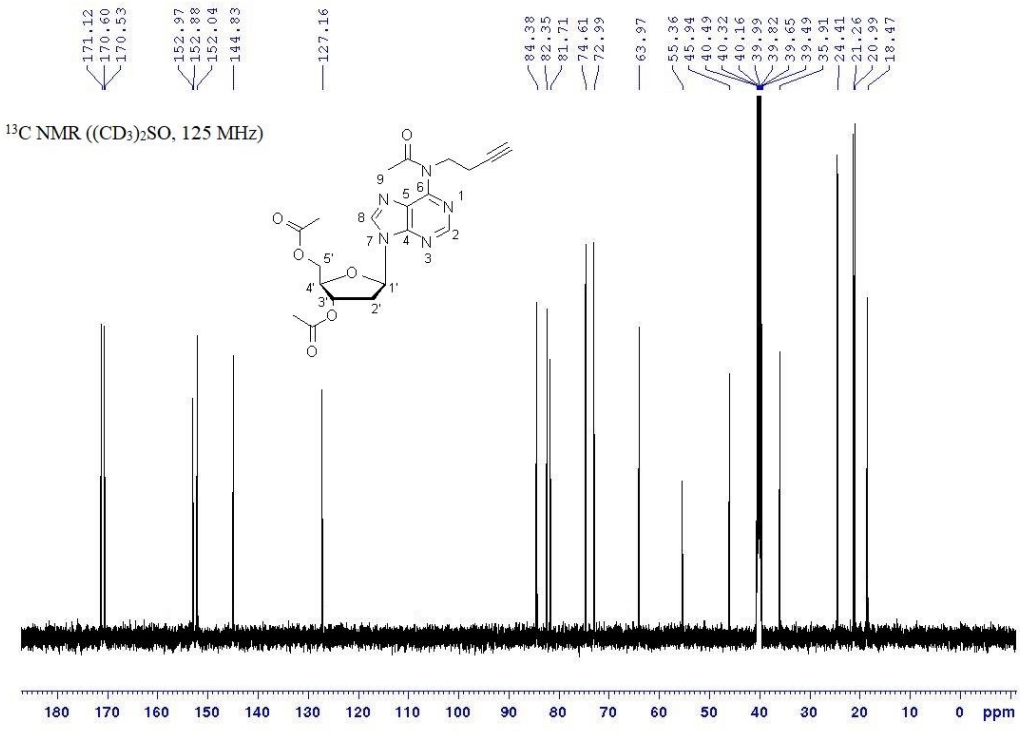


Figure S26. ^{13}C NMR of compound **11a**.

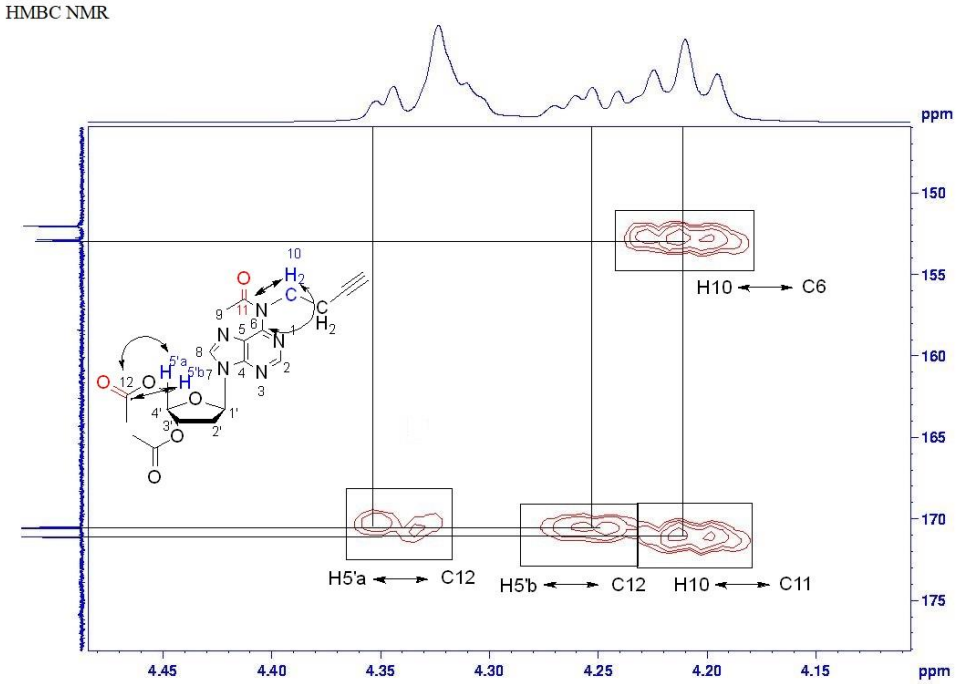


Figure S27. HMBC NMR of compound **11a**.

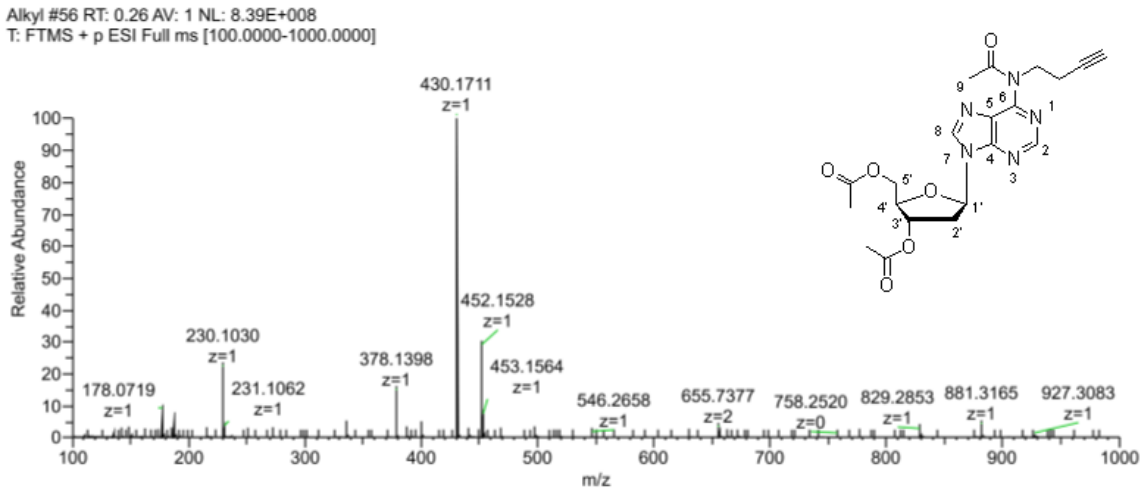


Figure S28. HRMS (ESI) of compound **11a**.

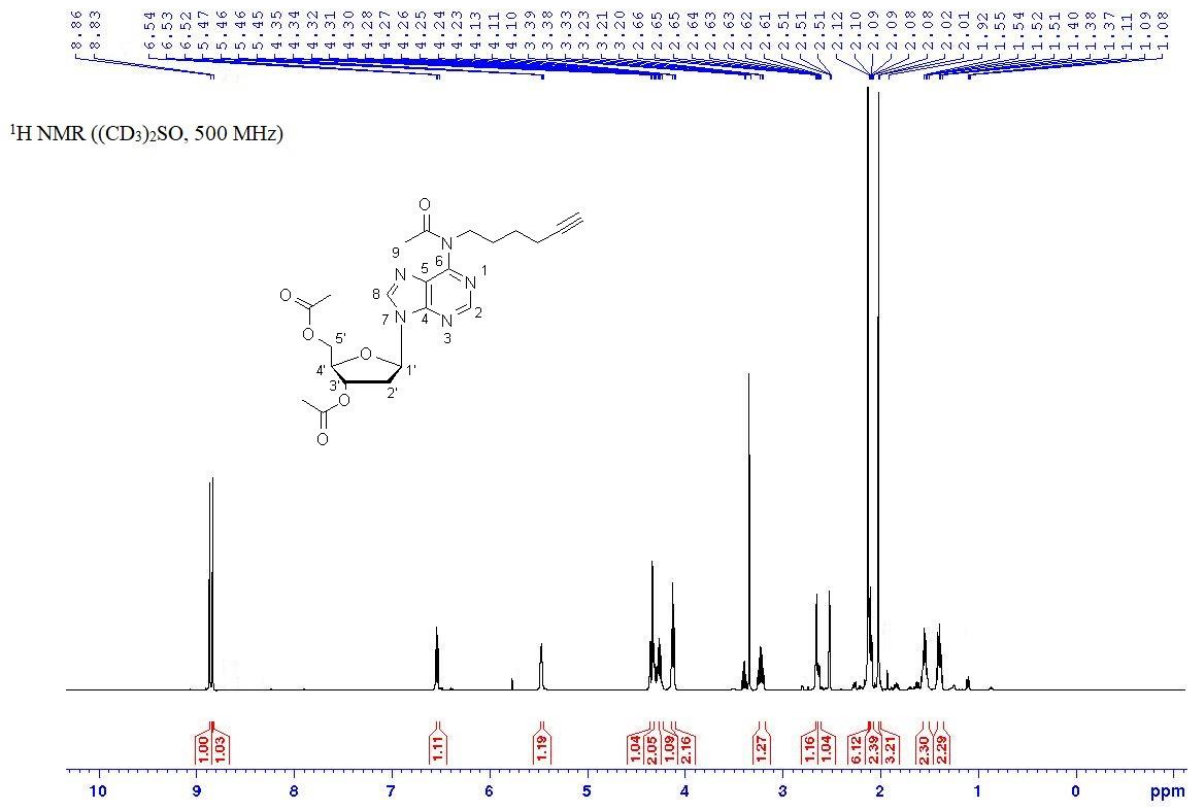


Figure S29. ^{1}H NMR of compound **11b**.

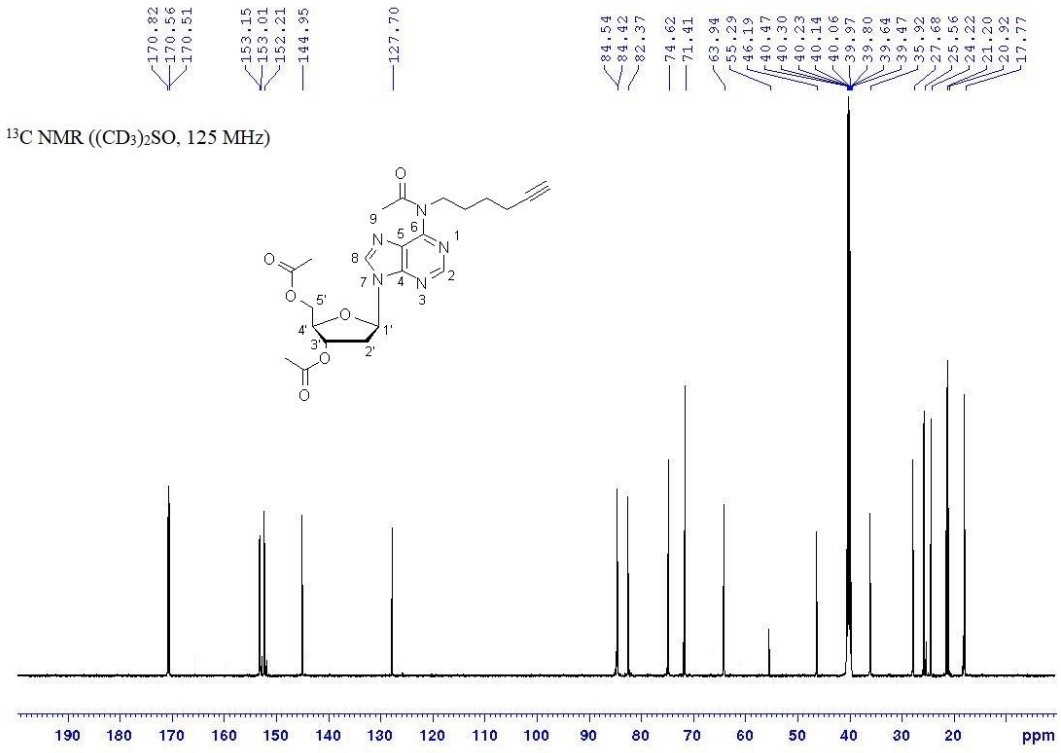


Figure S30. ^{13}C NMR of compound **11b**.

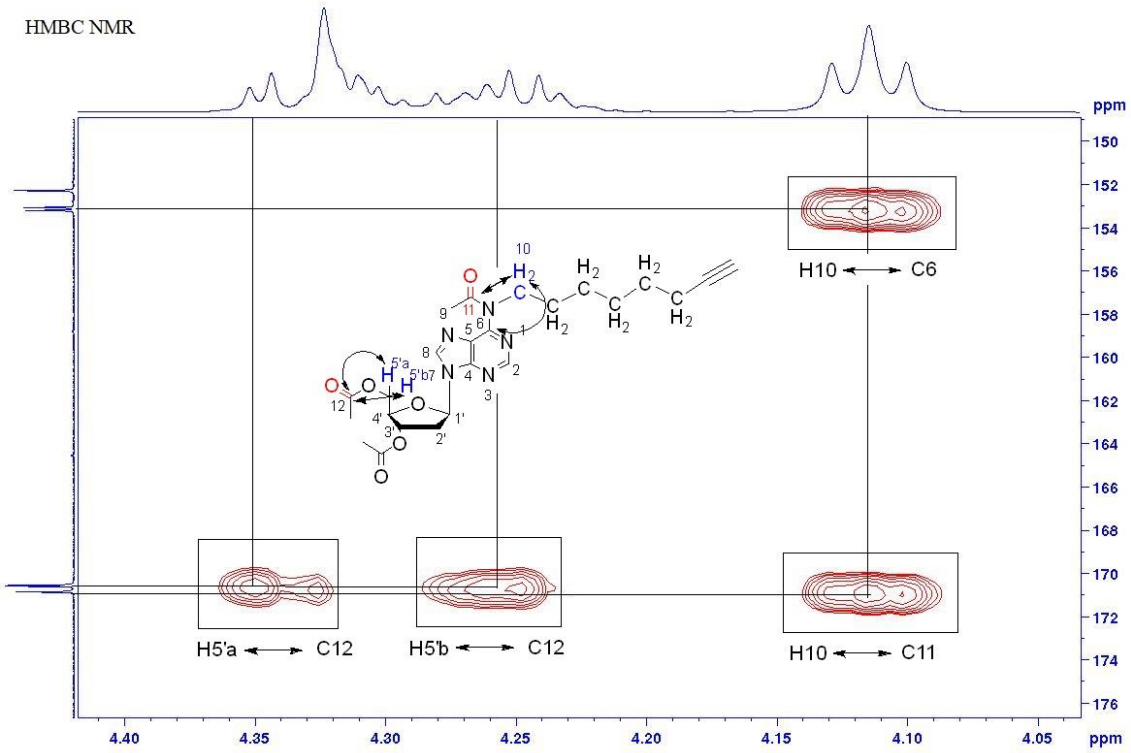


Figure S31. HMBC NMR of compound **11b**.

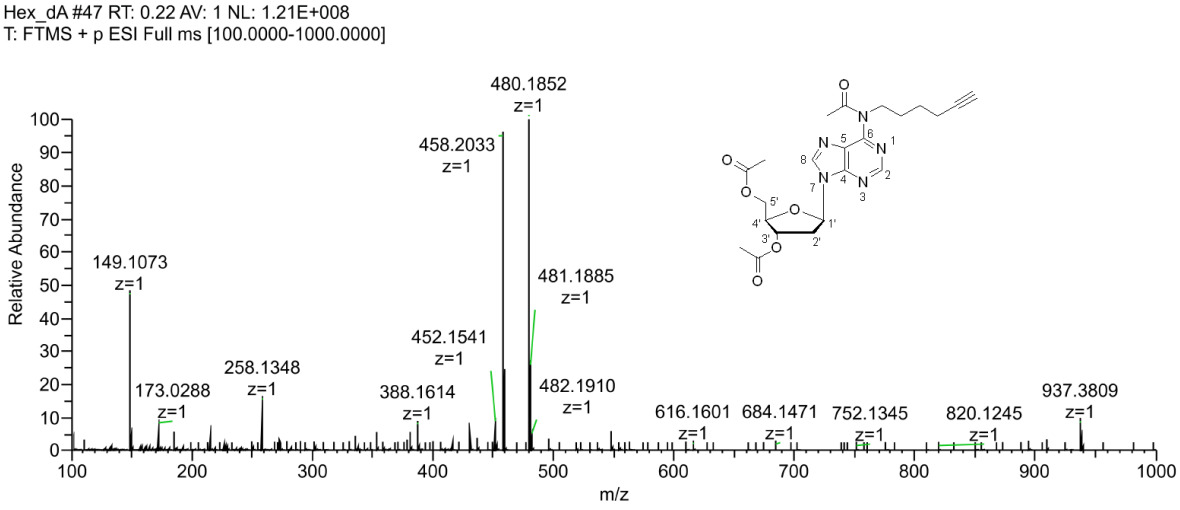


Figure S32. HRMS (ESI) of compound **11b**.

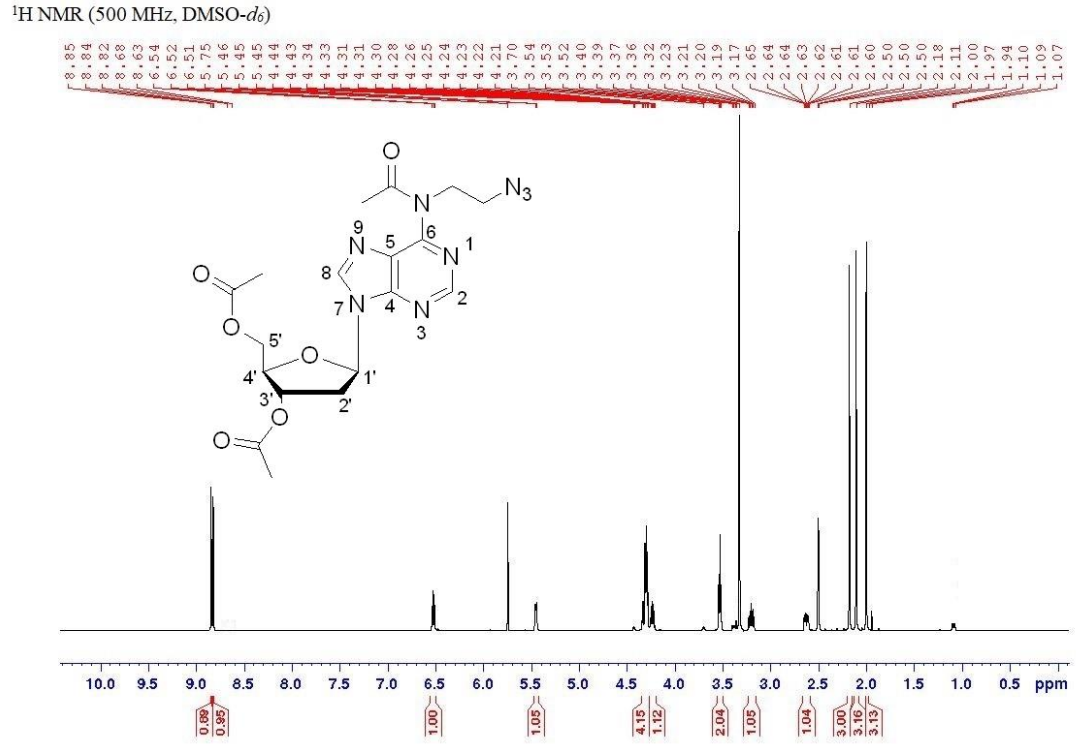


Figure S33. ^{1}H NMR spectrum of compound **11c**.

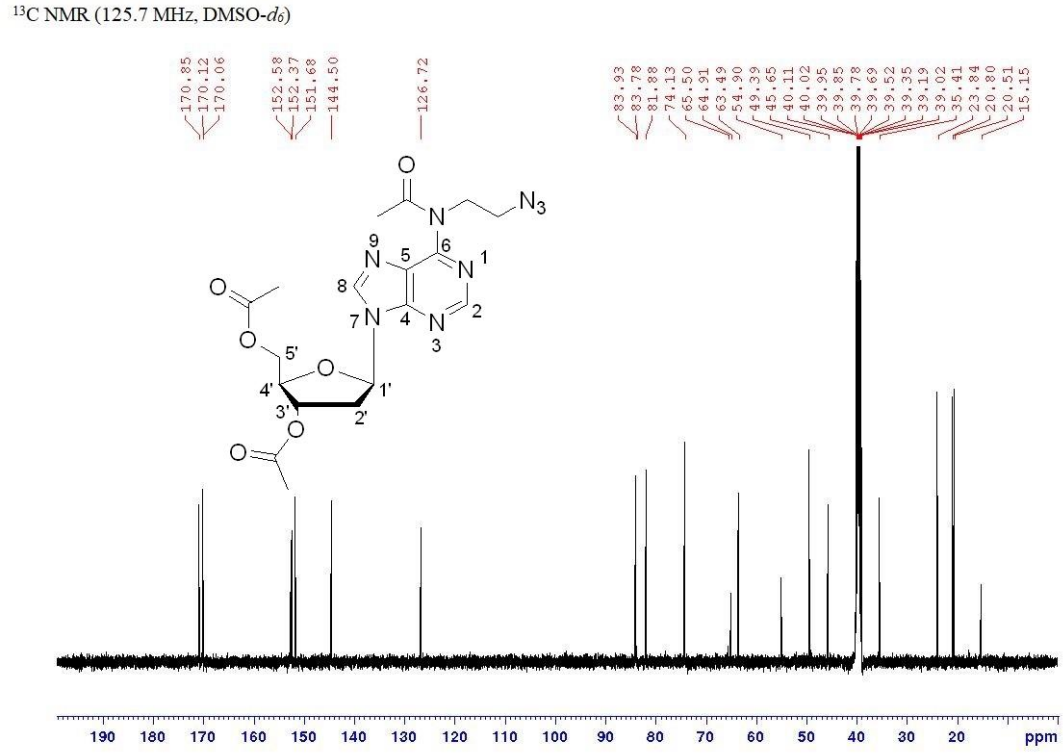


Figure S34. ^{13}C NMR spectrum of compound **11c**.

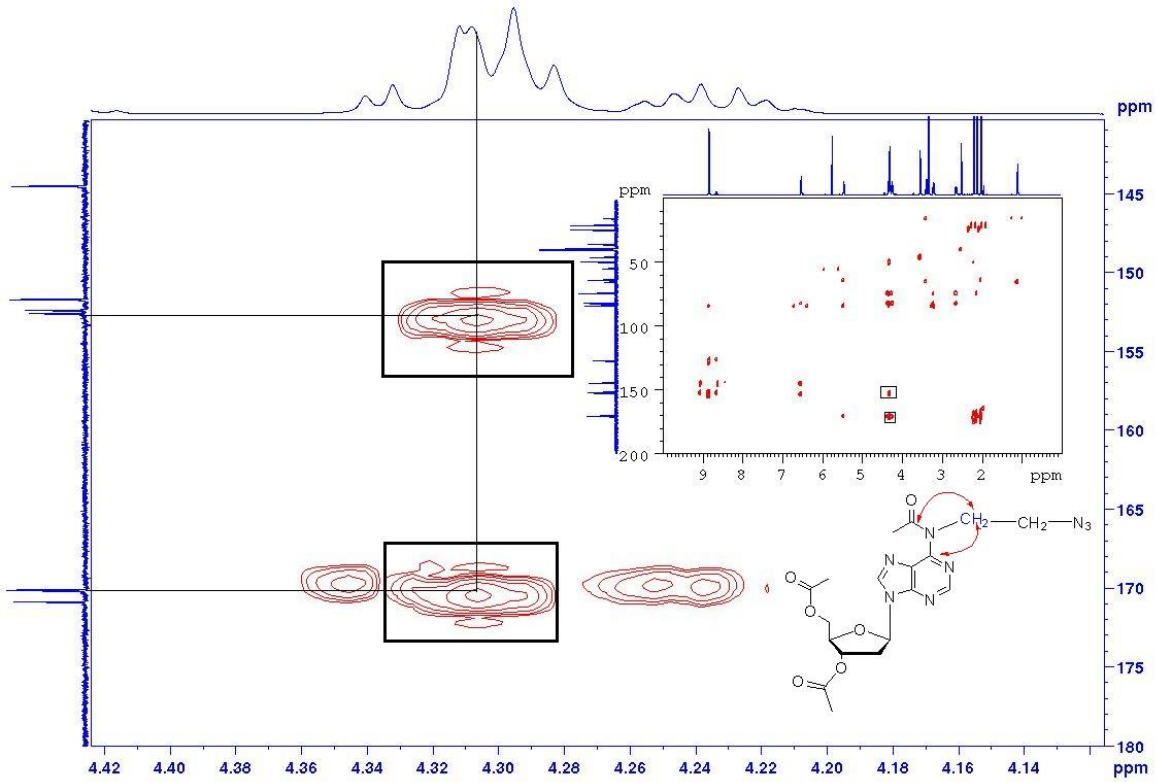


Figure S35. HMBC NMR spectrum of compound **11c**.

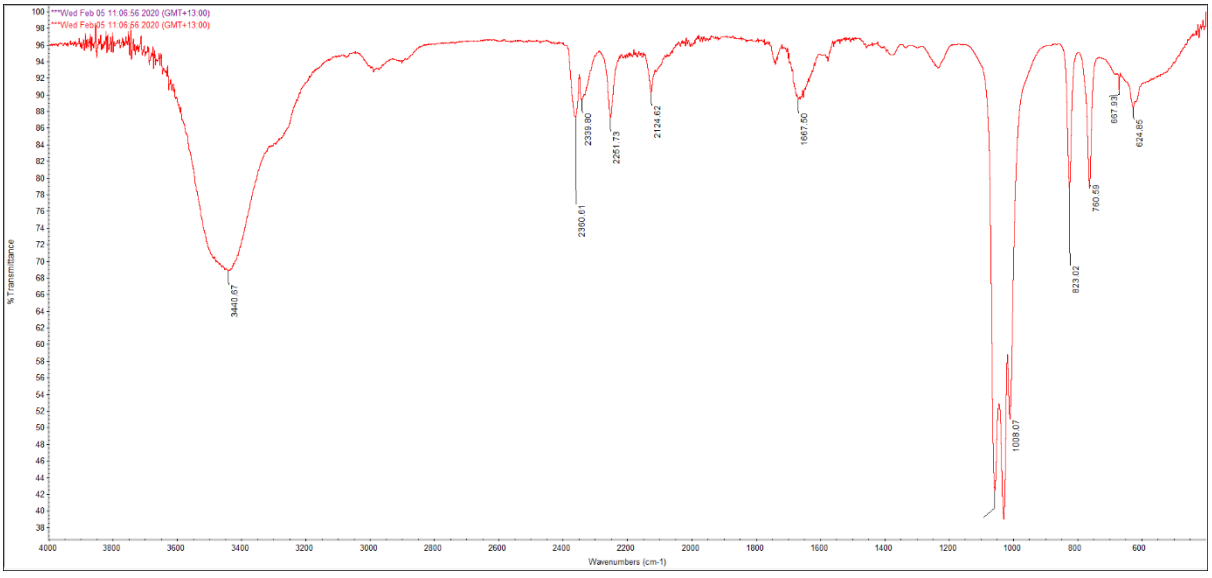


Figure S36. IR (ATR) spectrum of compound **11c**.

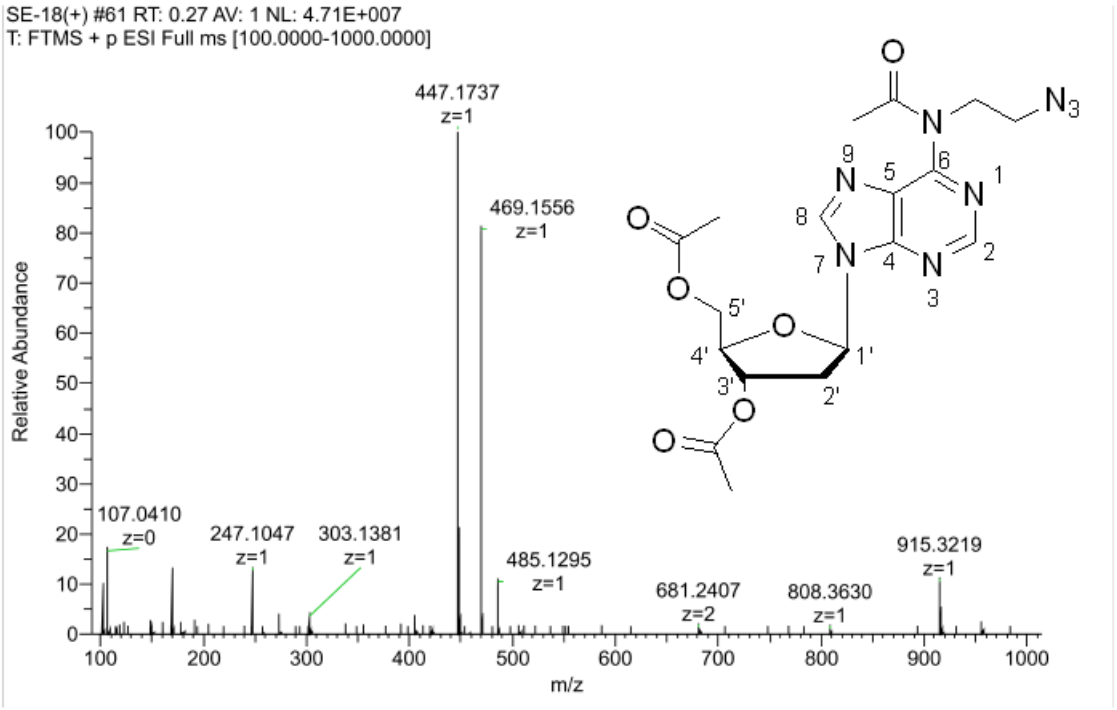


Figure S37. HRMS (ESI) of compound **11c**.

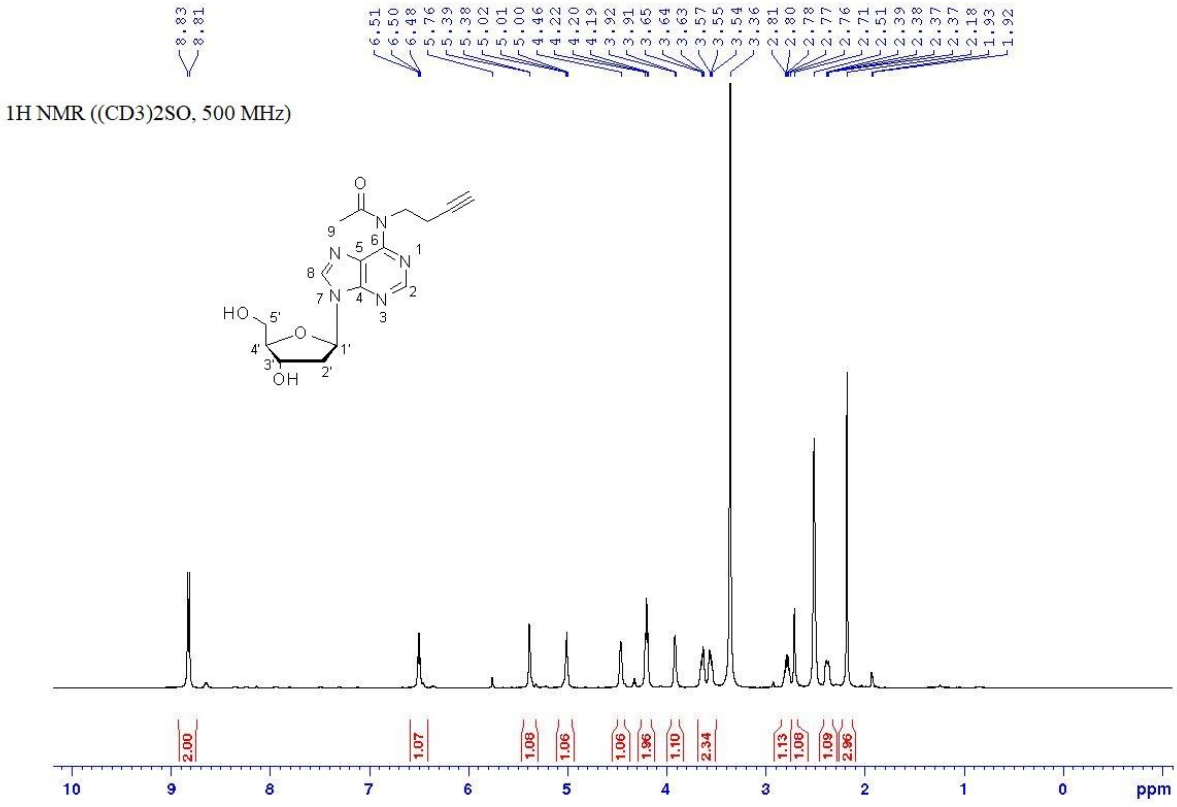


Figure S38. [1]H NMR of compound **12a**.

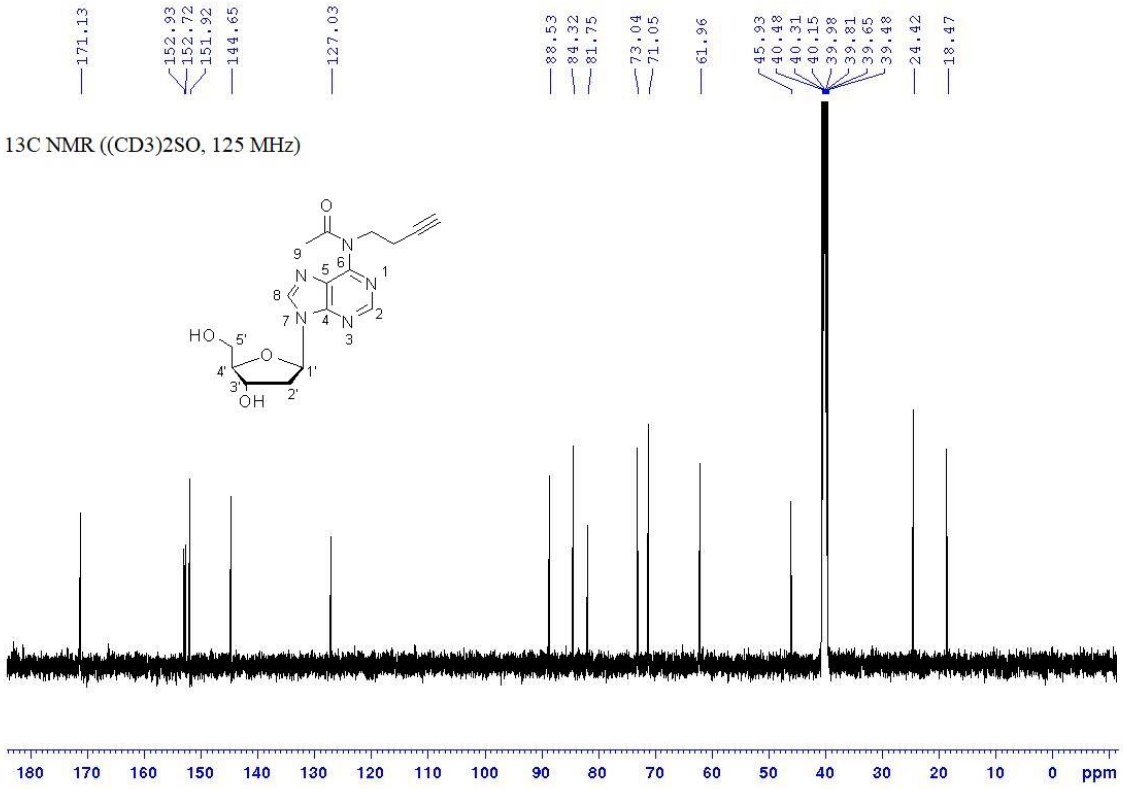


Figure S39. ^{13}C NMR of compound **12a**.

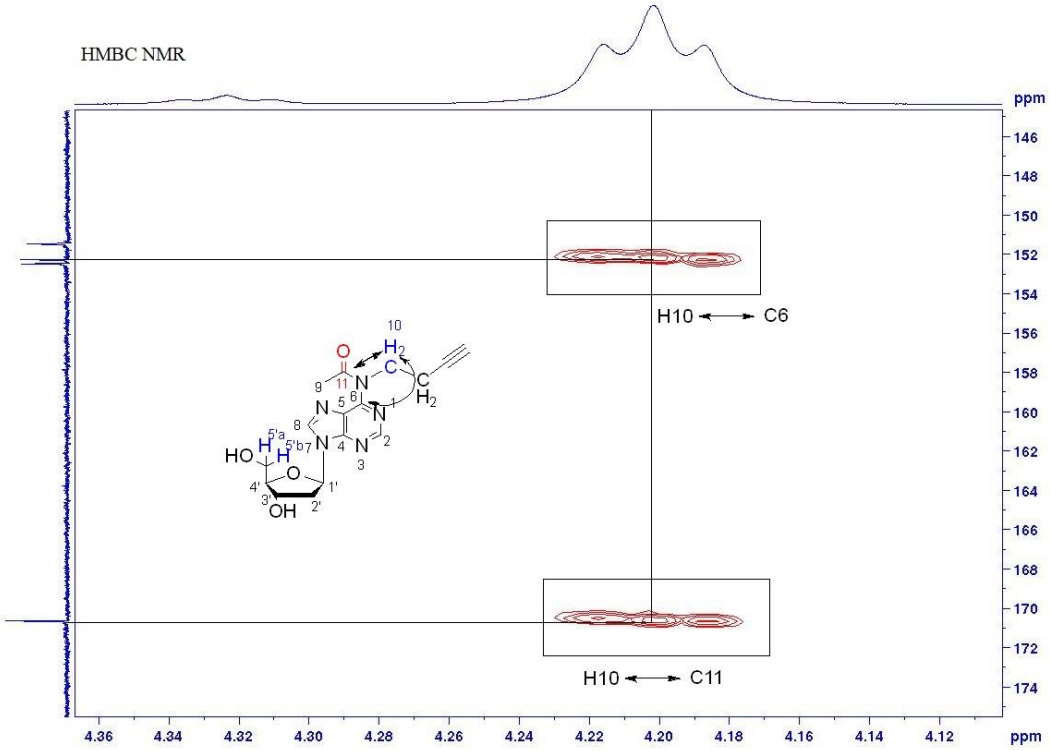


Figure S40. HMBC NMR of compound **12a**.

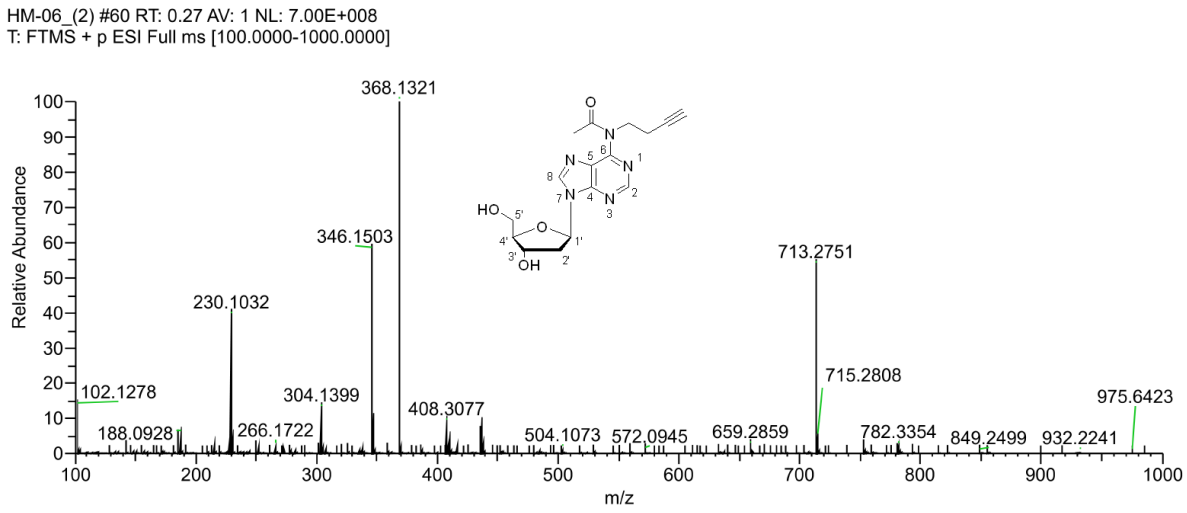


Figure S41. HRMS (ESI) of compound **12a**.

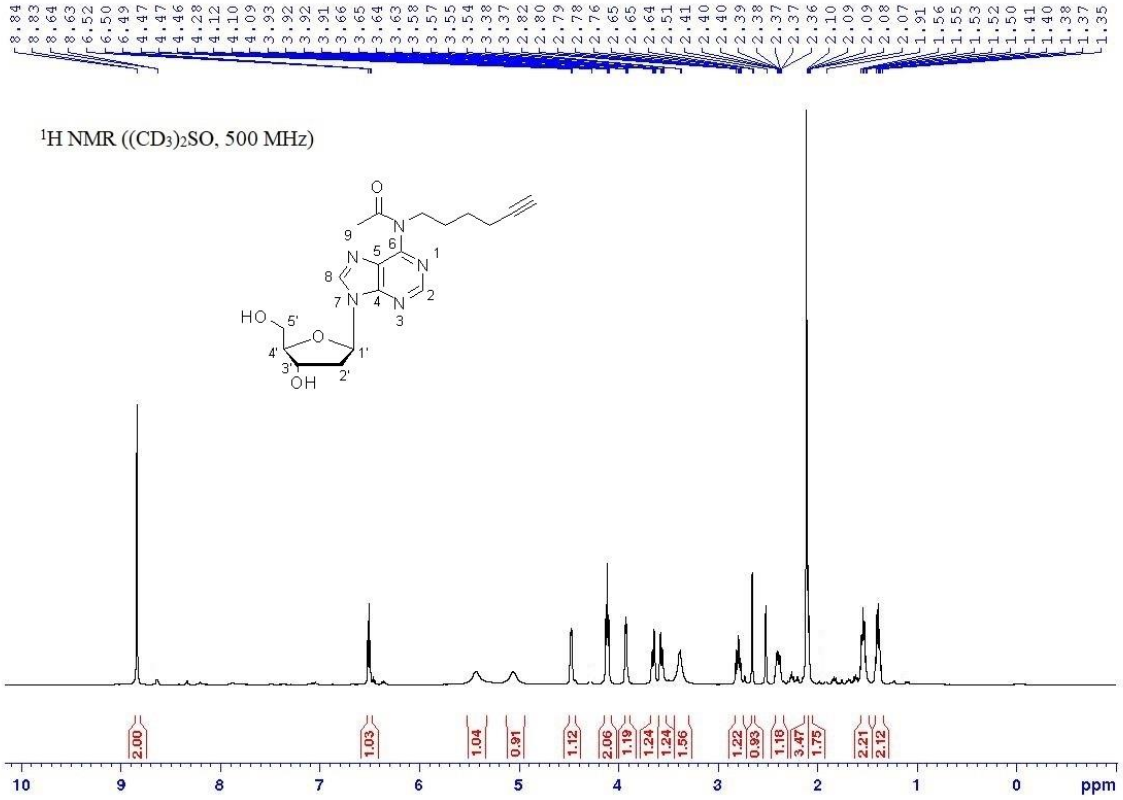


Figure S42. ^{1}H NMR of compound **12b.**

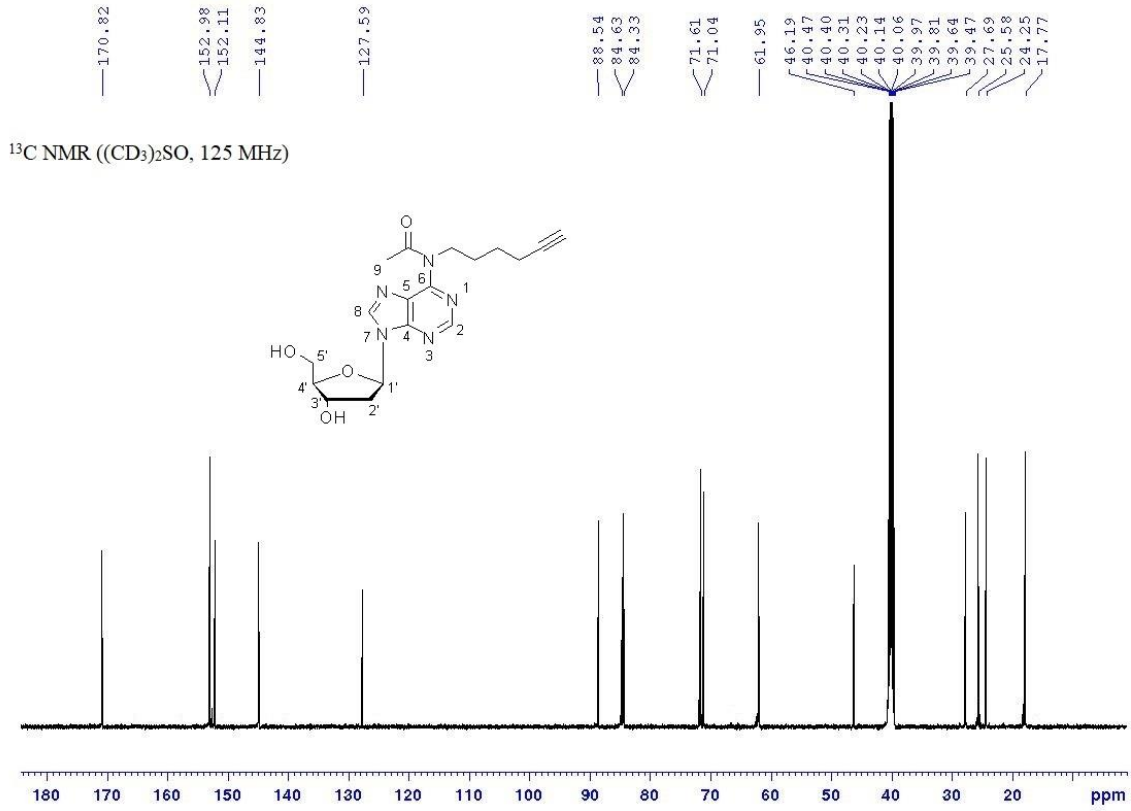


Figure S43. ^{13}C NMR of compound **12b**.

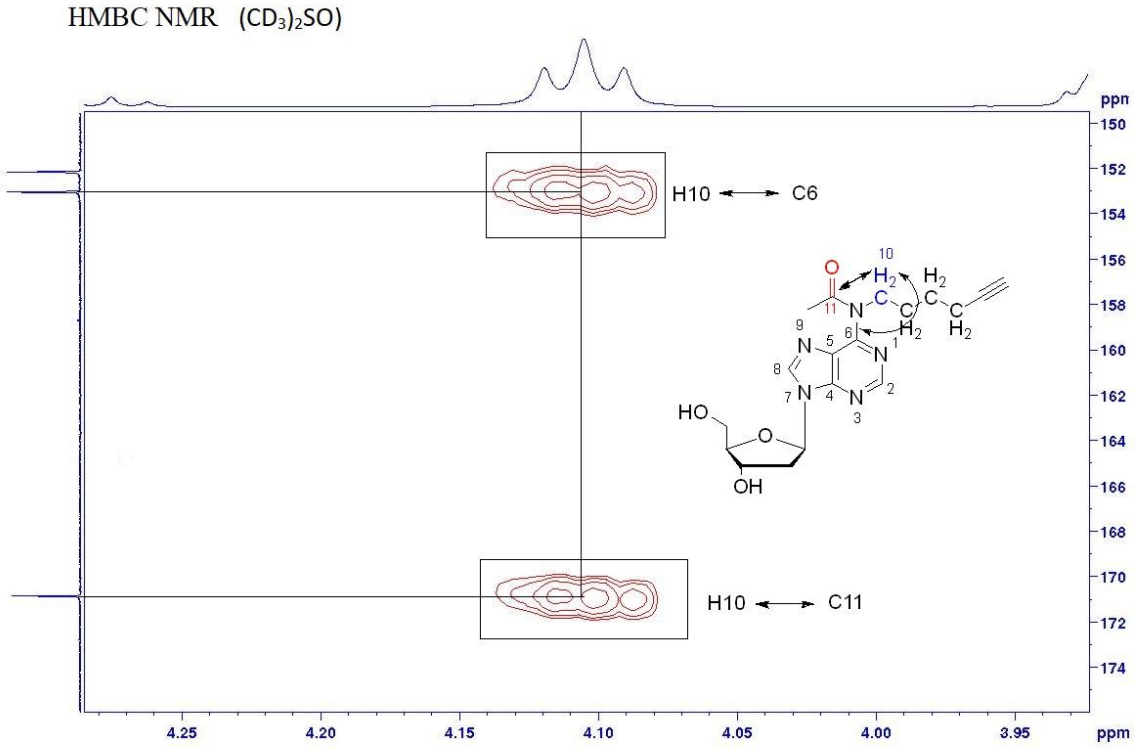


Figure S44. HMBC NMR of compound **12b**.

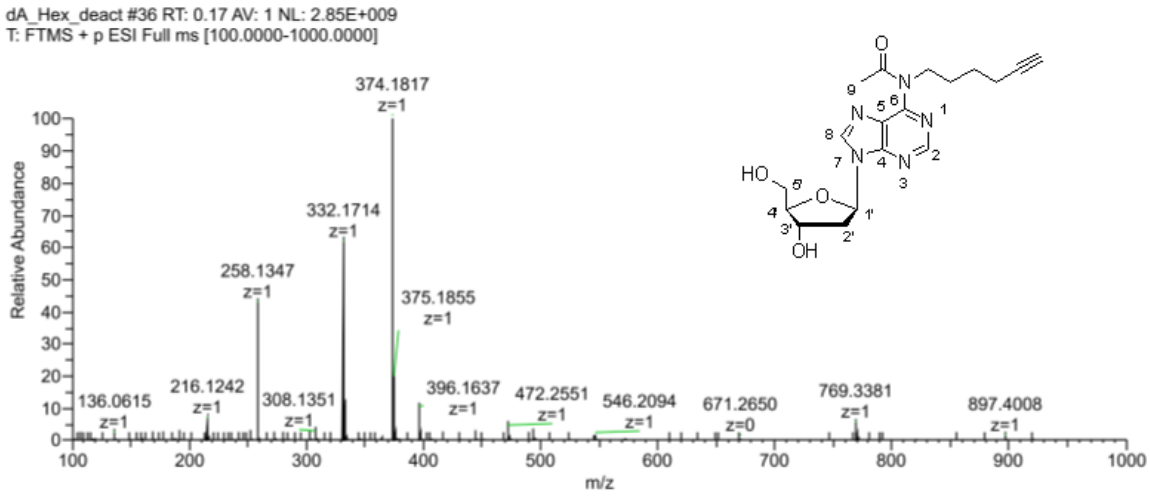


Figure S45. HRMS (ESI) of compound **12b**.

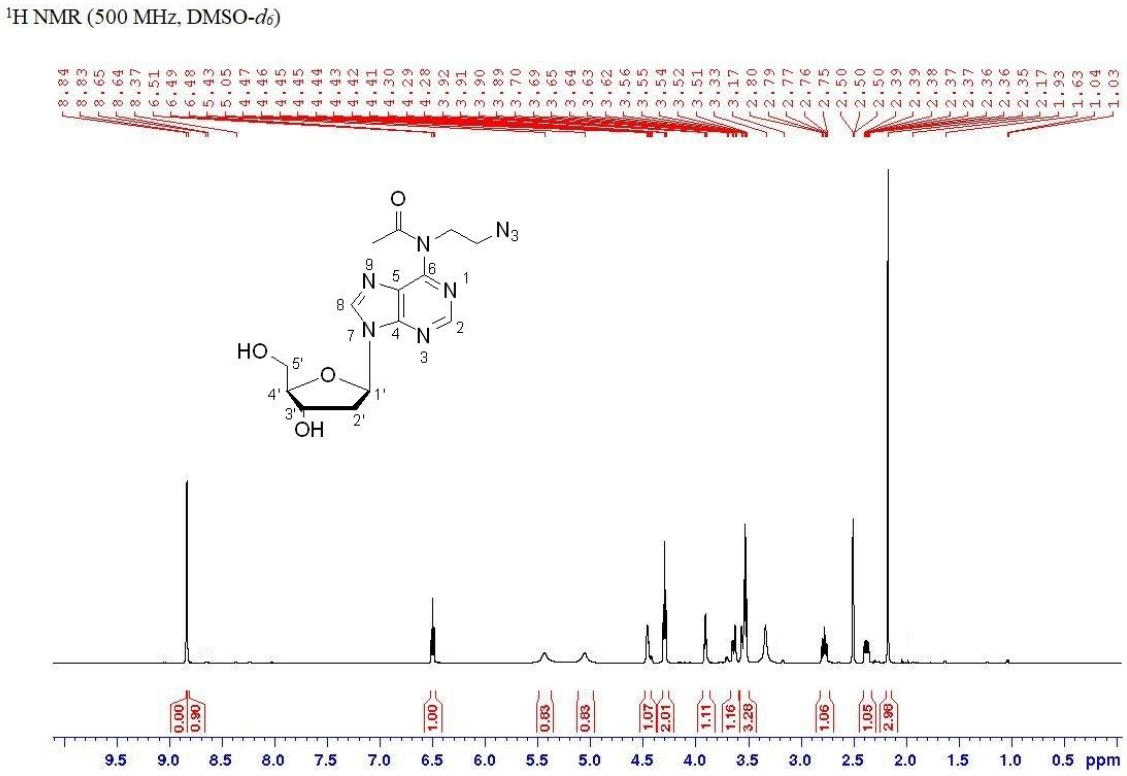


Figure S46. ^{1}H NMR spectrum of compound **12c**.

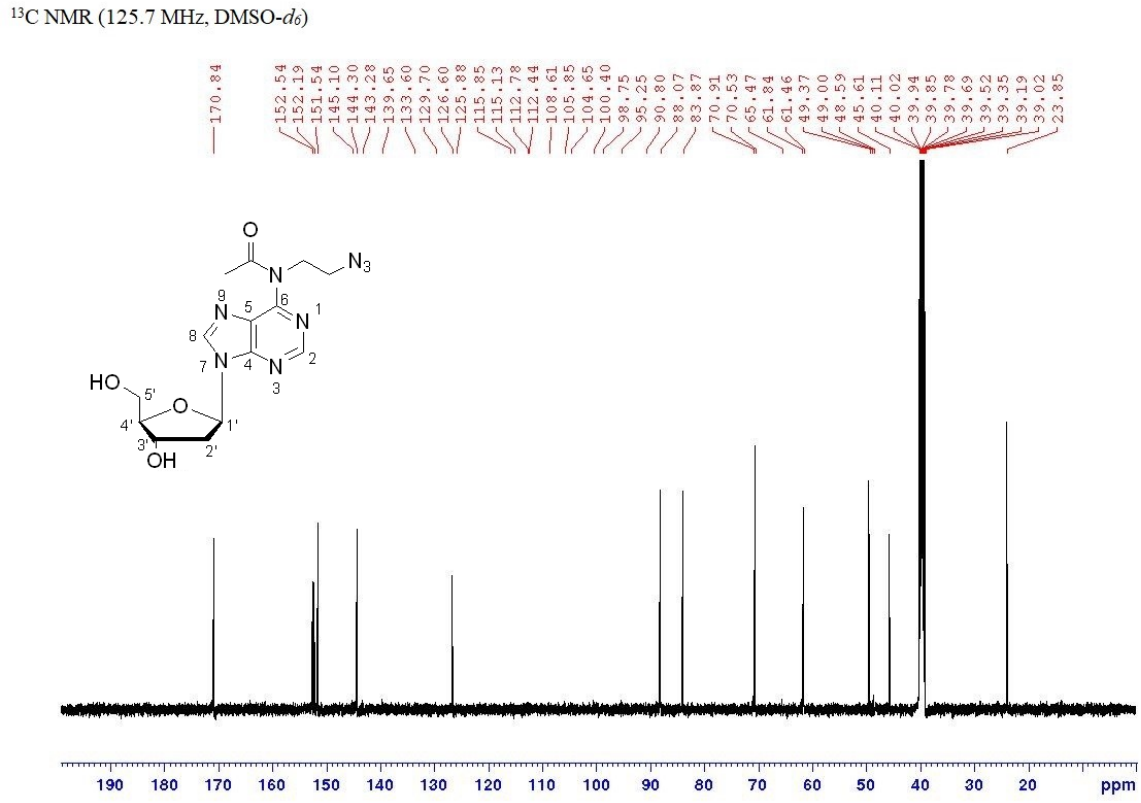


Figure S47. ^{13}C NMR spectrum of compound **12c**.

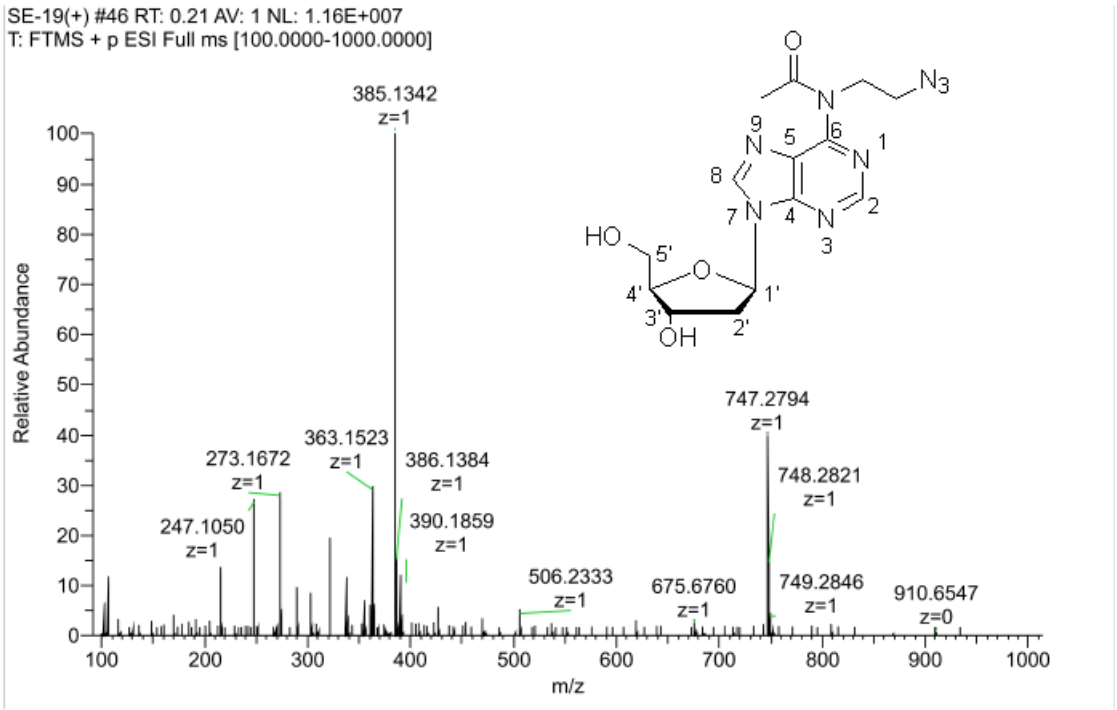


Figure S48. HRMS (ESI) of compound **12c**.

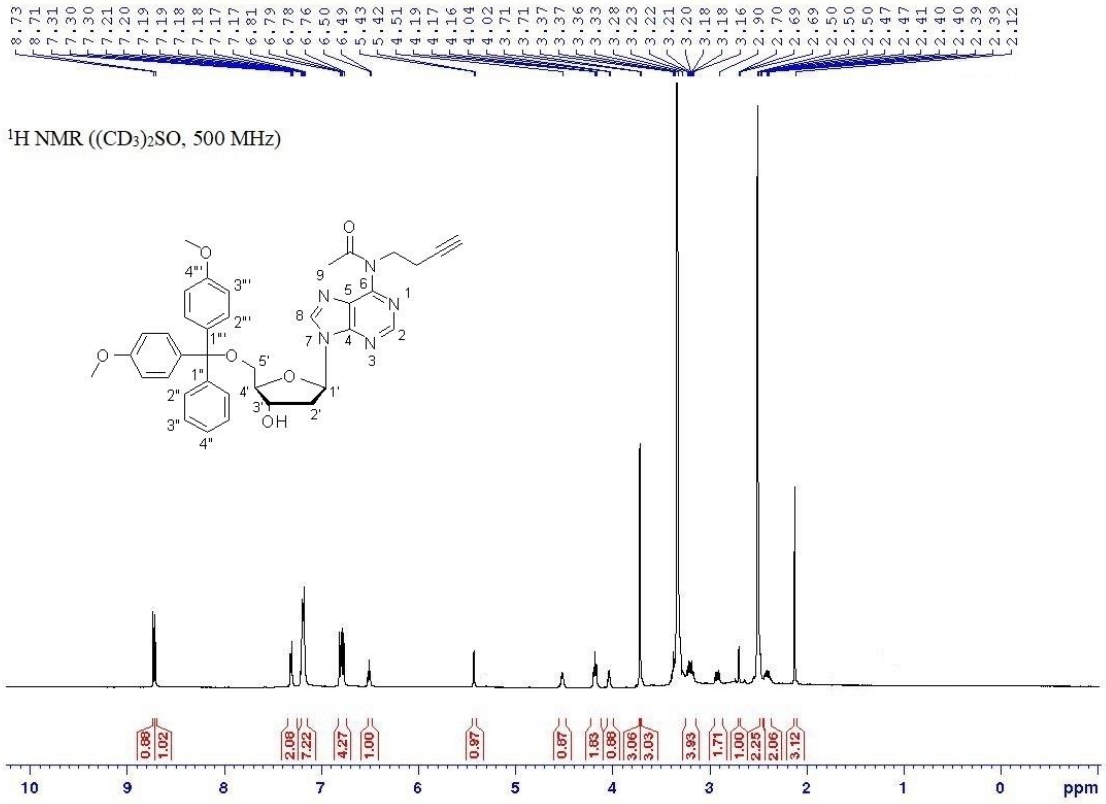


Figure S49. ^{1}H NMR of compound **13a**

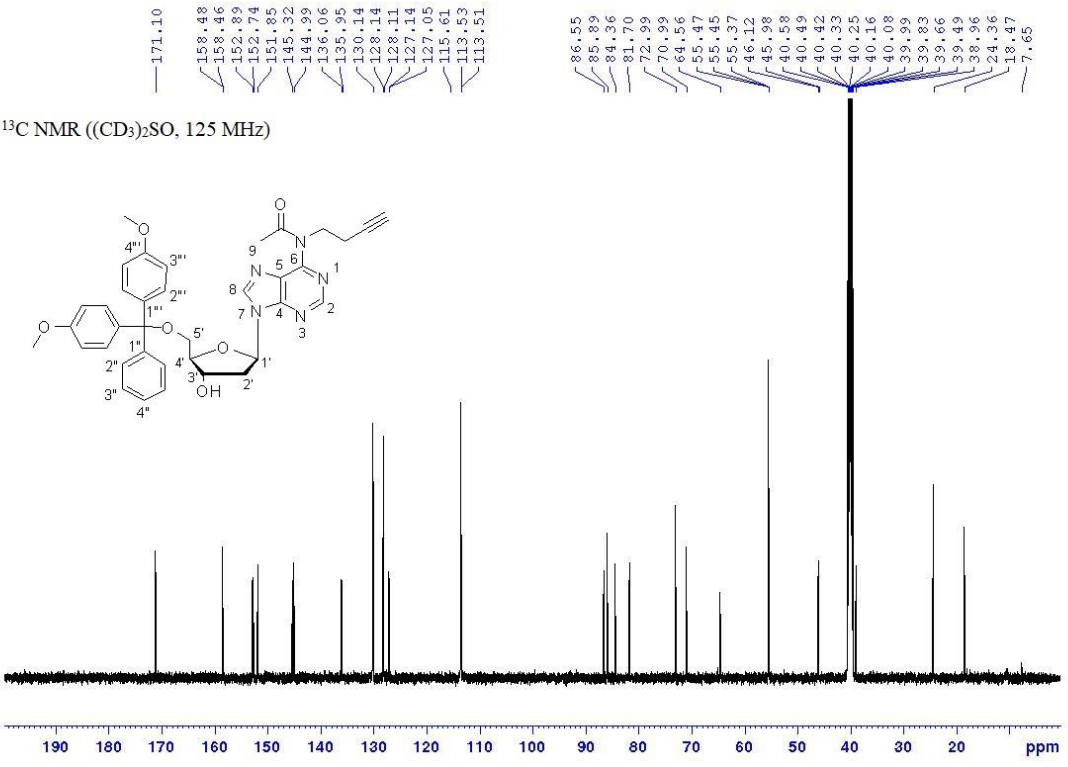


Figure S50. ^{13}C NMR of compound **13a**.

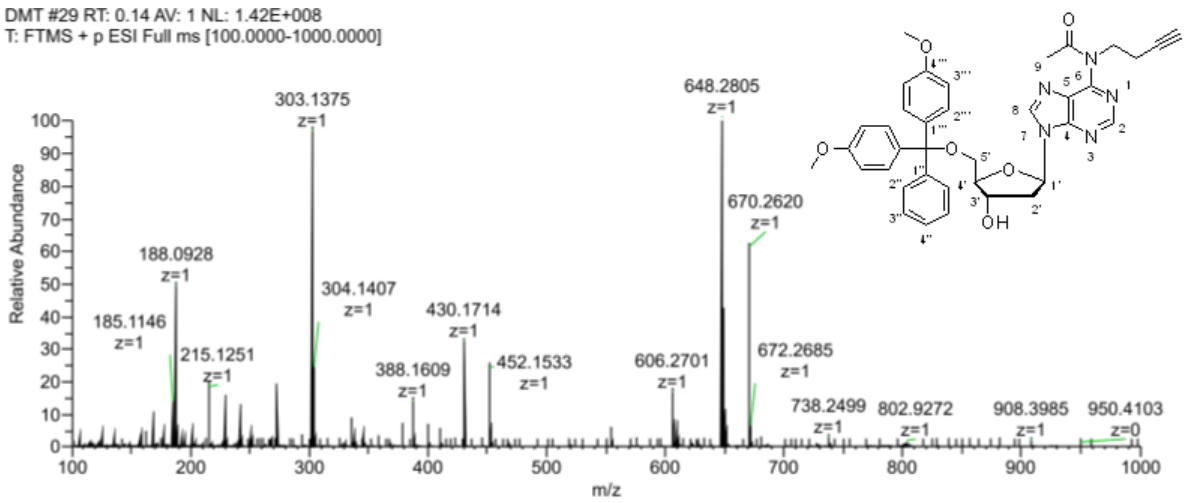


Figure S51. HRMS (ESI) of compound **13a.**

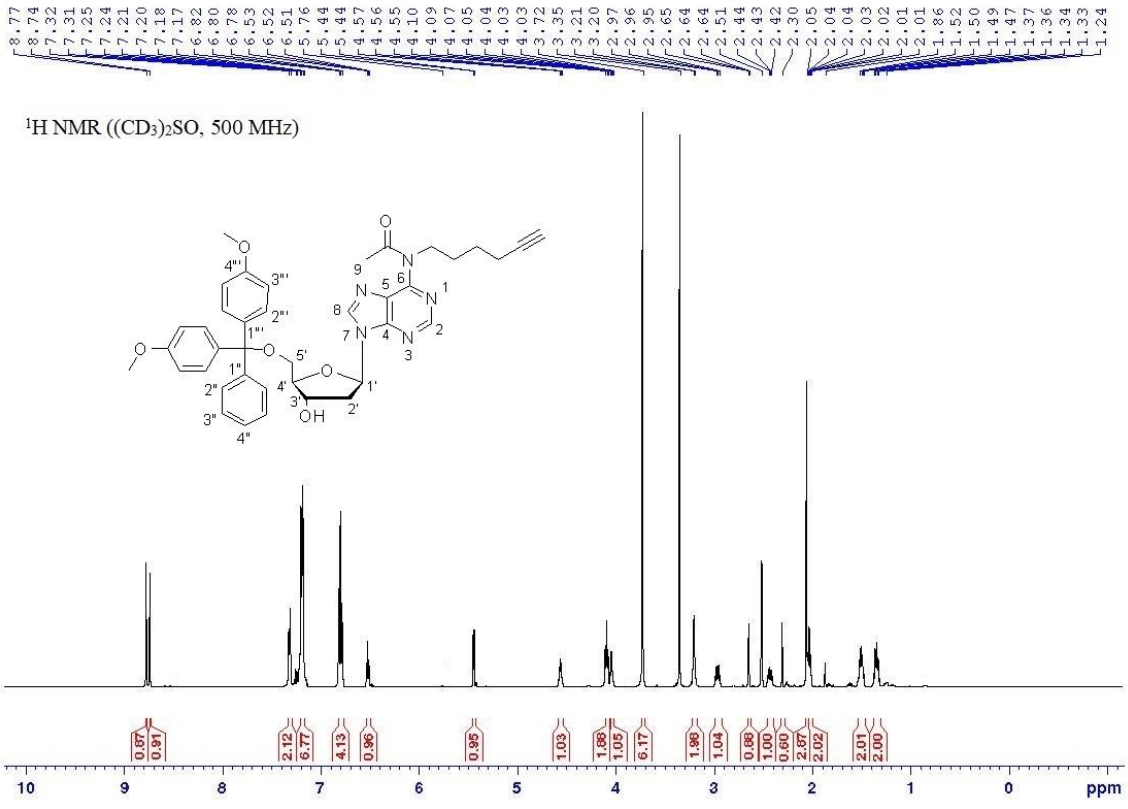


Figure S52. ^{1}H NMR of compound **13b**.

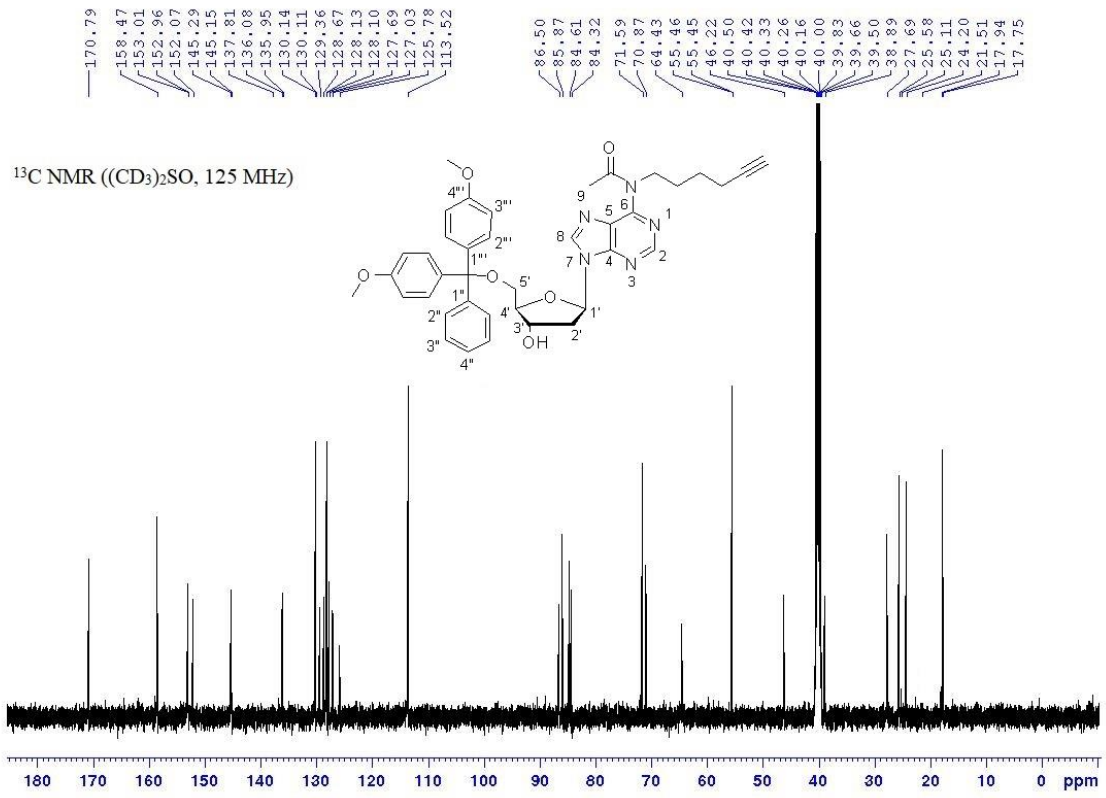


Figure S53. ^{13}C NMR of compound **13b**

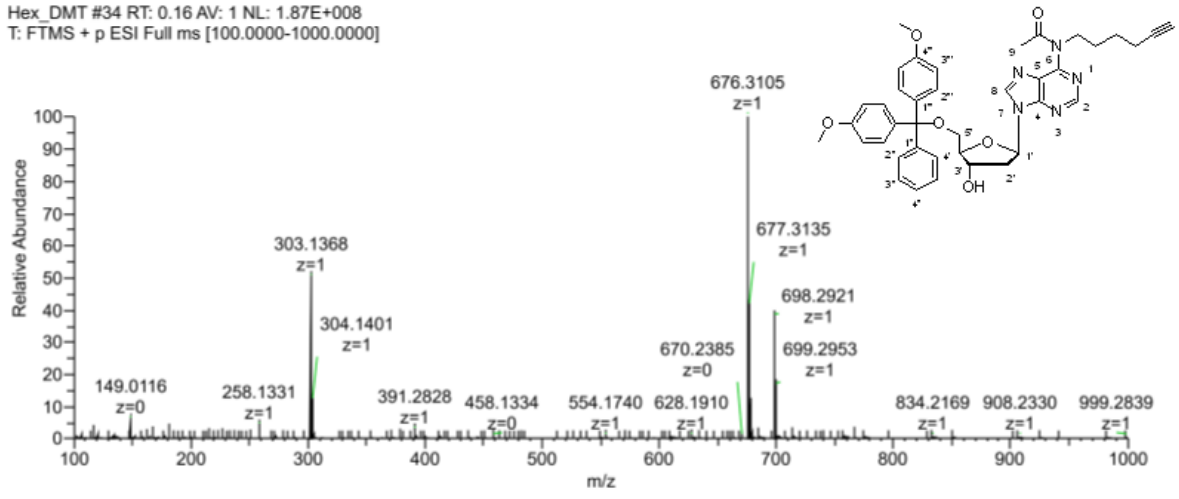


Figure S54. HRMS (ESI) of compound **13b.**

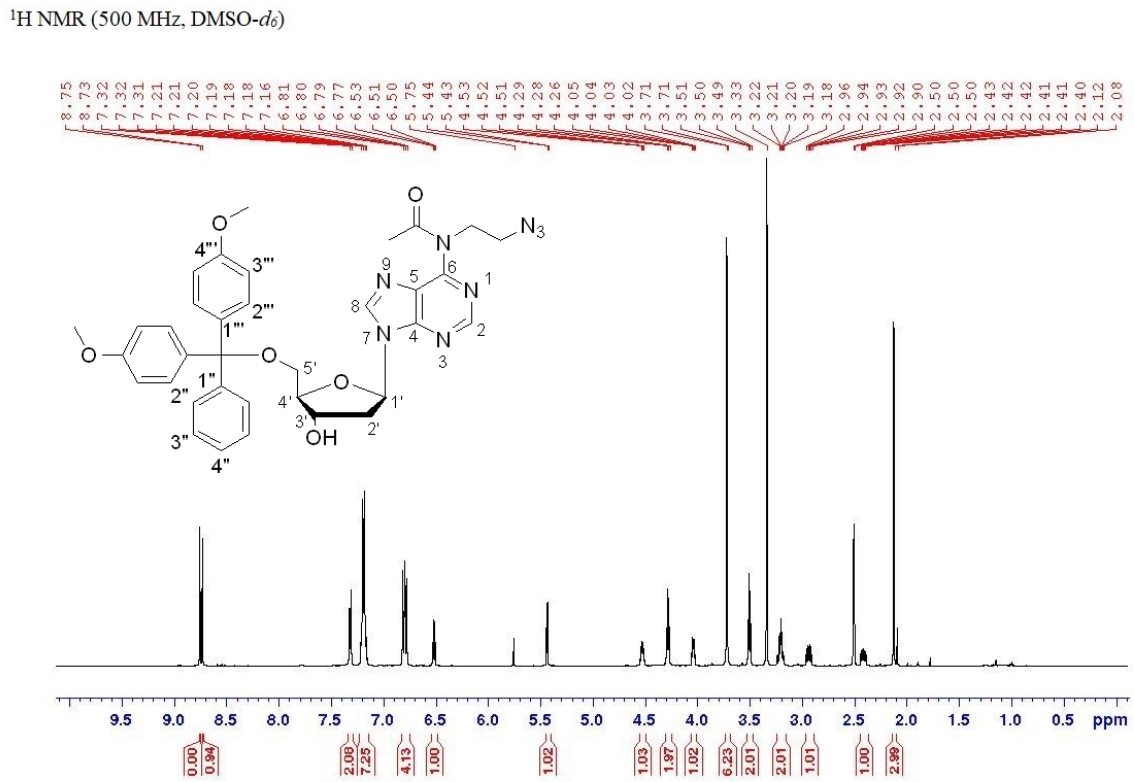


Figure S55. ^{1}H NMR spectrum of compound **13c**.

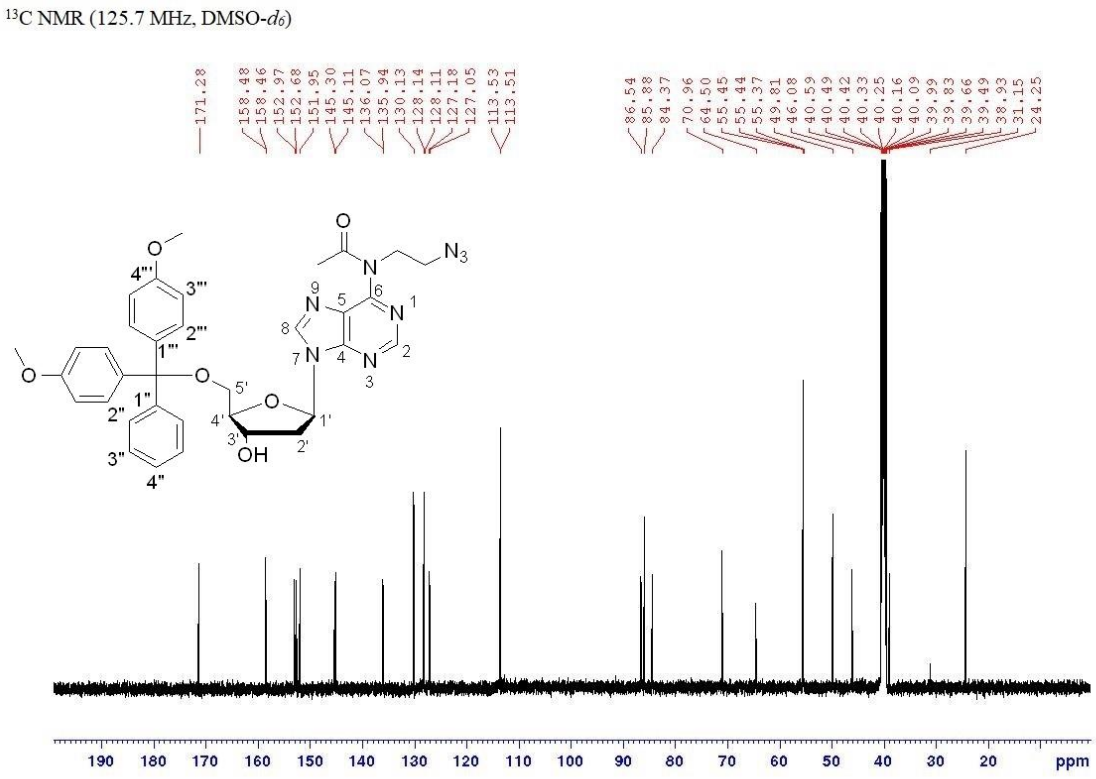


Figure S56. ^{13}C NMR spectrum of compound **13c**.

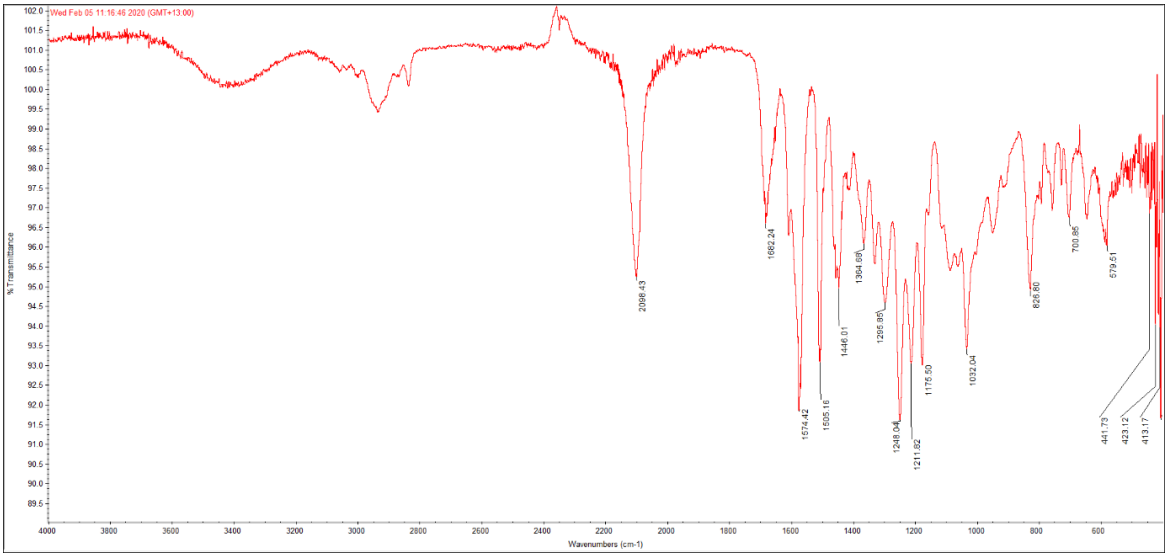


Figure S57. IR (ATR) spectrum of compound **13c**.

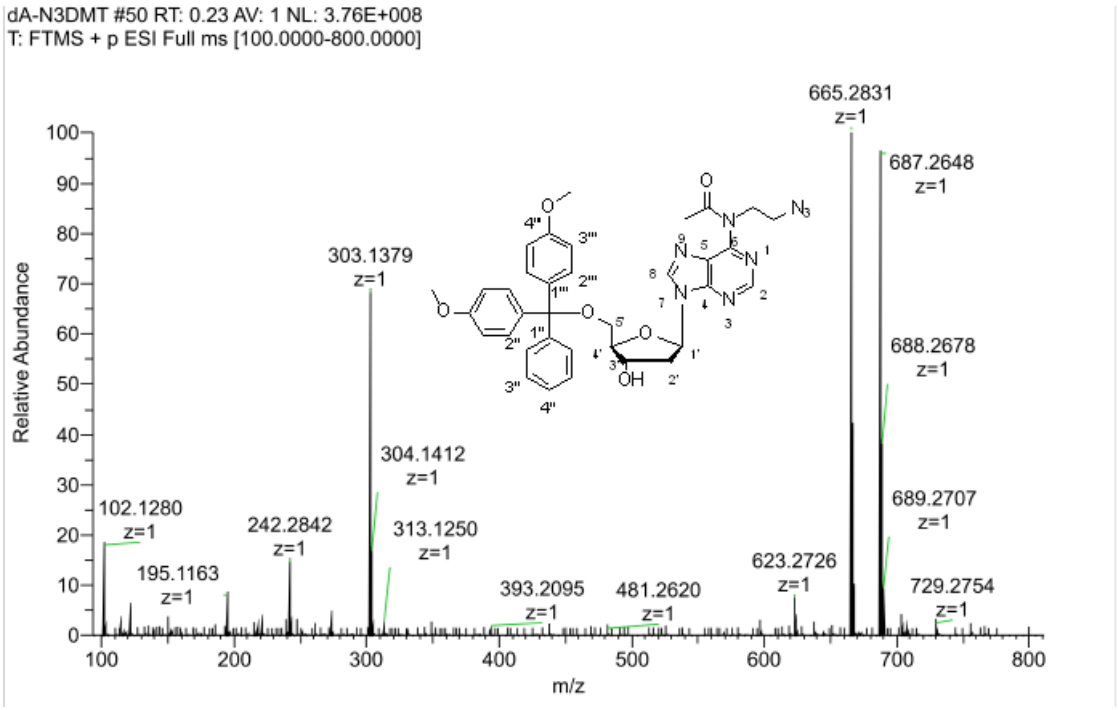


Figure S58. HRMS (ESI) of compound **13c**.

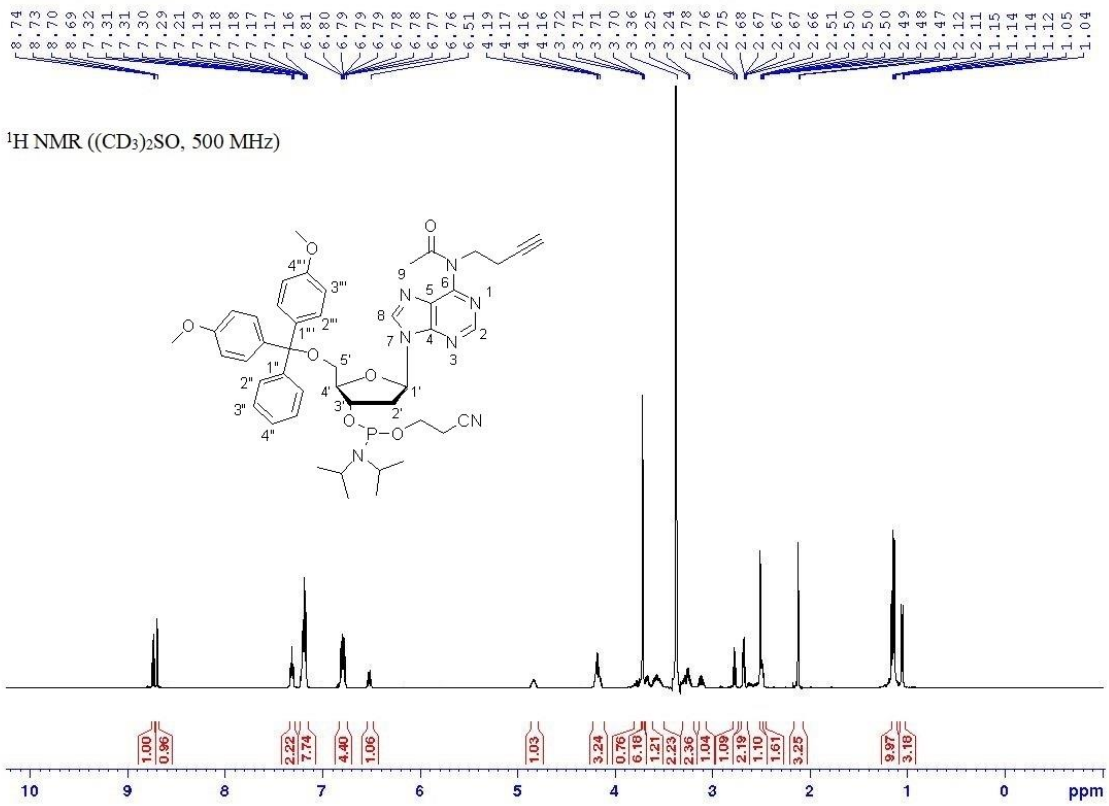


Figure S59. ^{1}H NMR of compound **14a.**

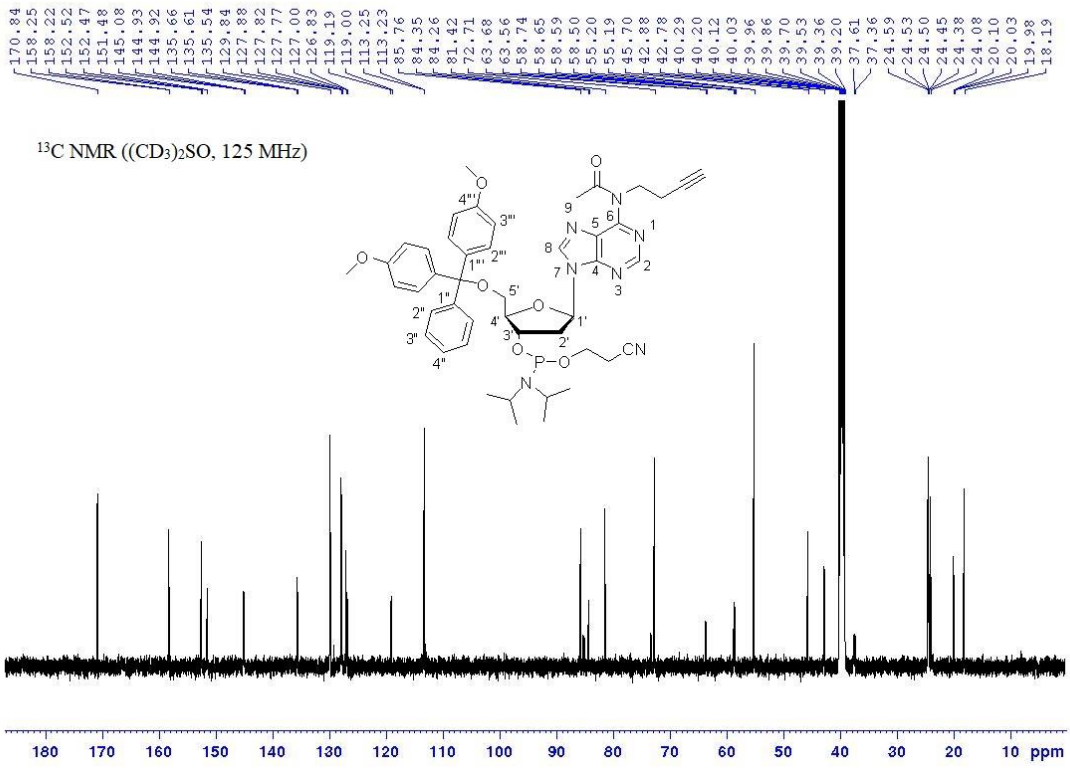


Figure S60. ^{13}C NMR of compound **14a**

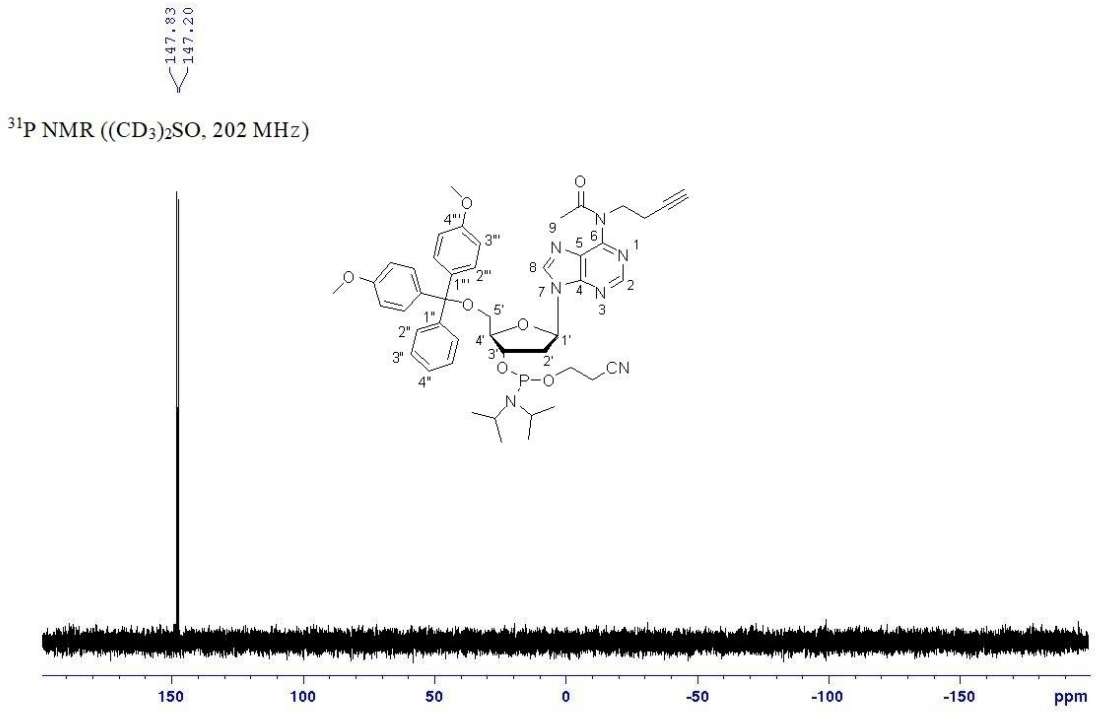


Figure S61. ^{31}P NMR of compound **14a**.

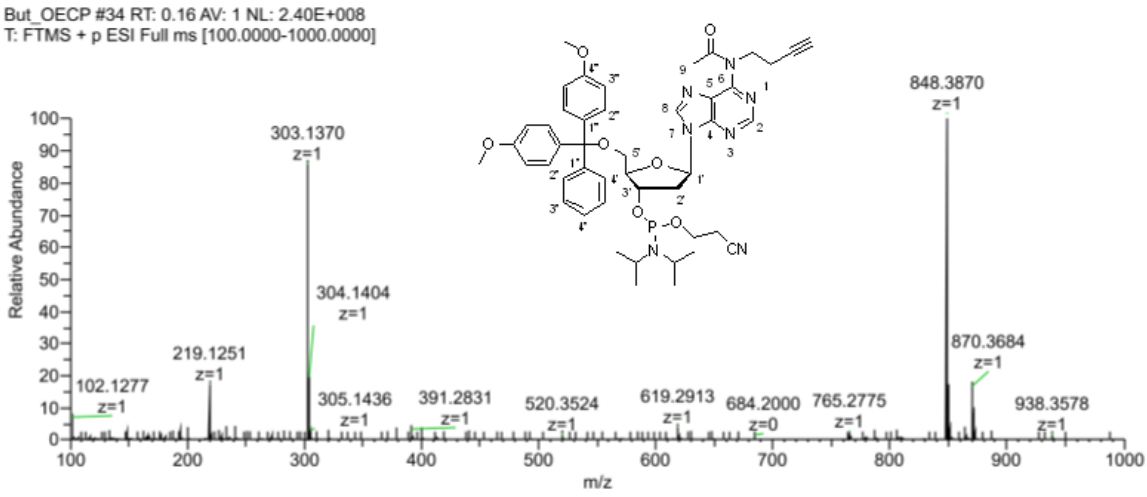


Figure S62. HRMS (ESI) of compound **14a**.

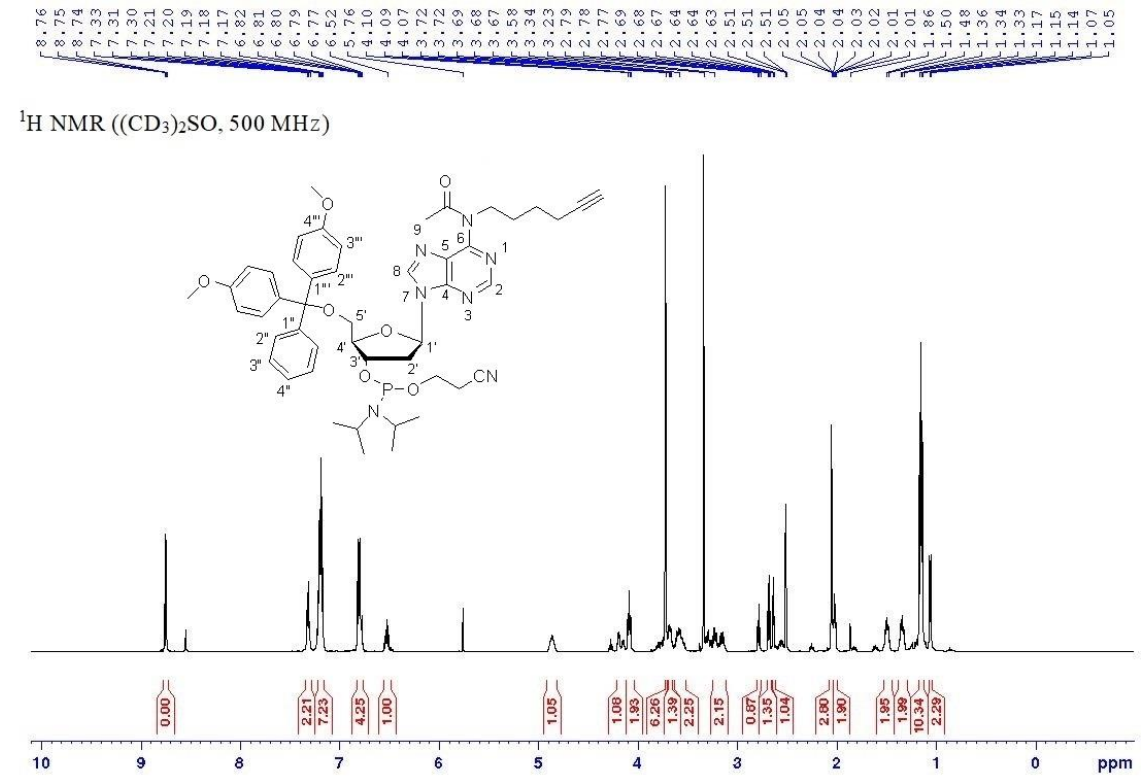


Figure S63. ^{1}H NMR of compound **14b**.

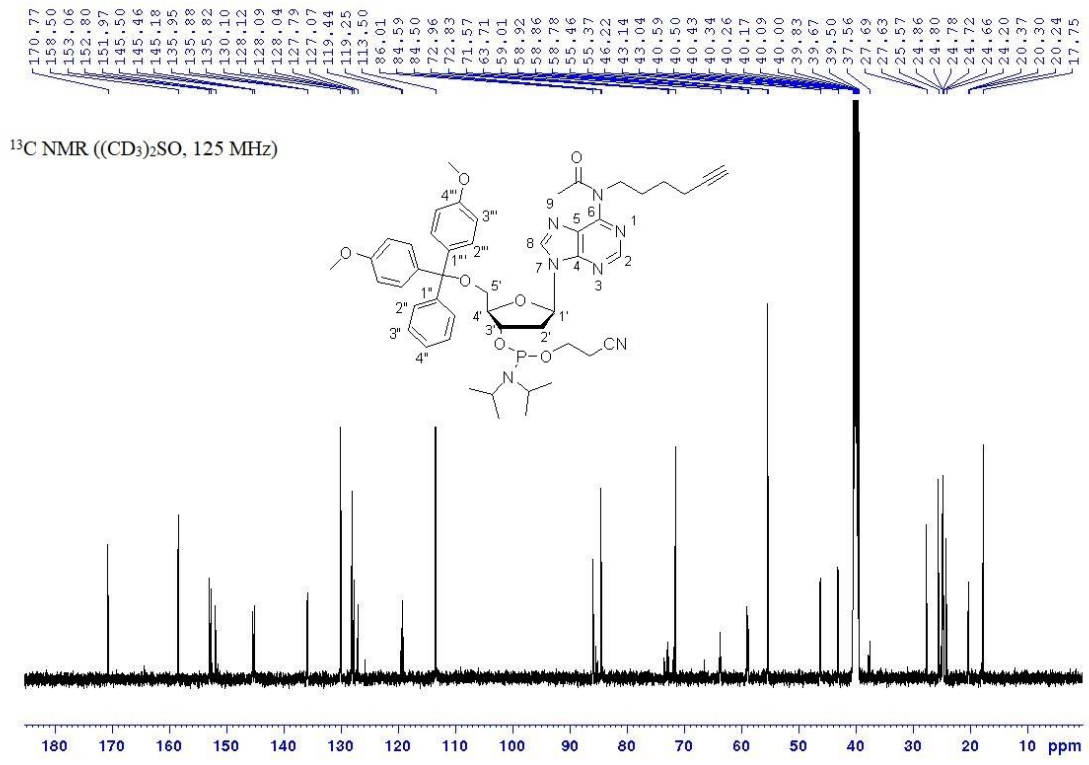


Figure S64. ^{13}C NMR of compound **14b**.

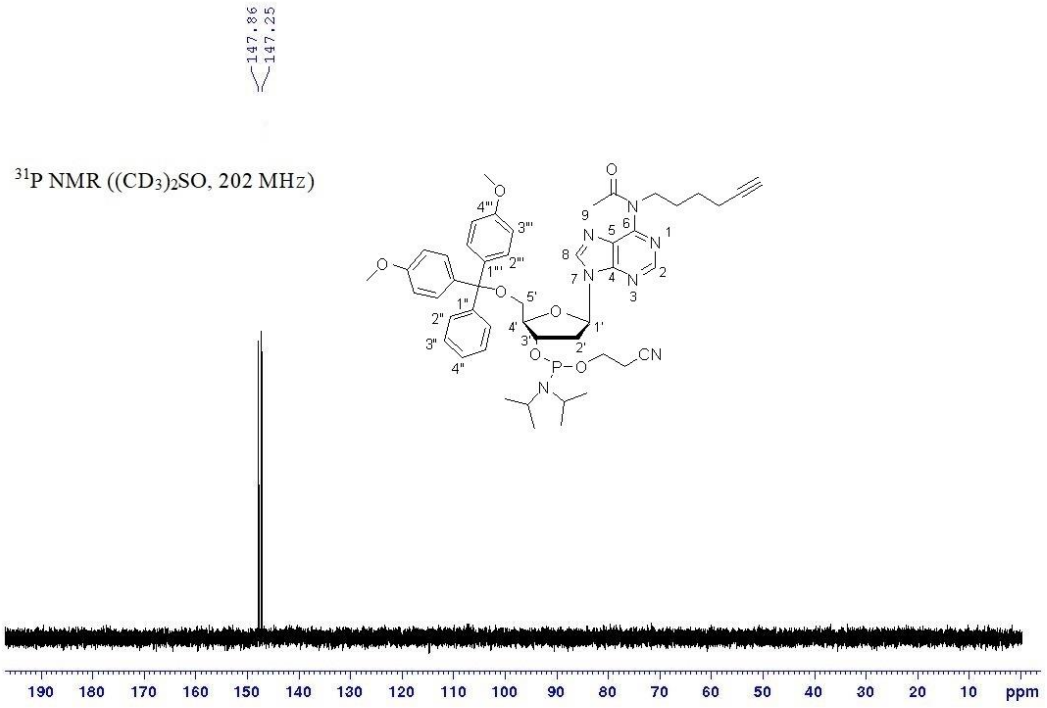


Figure S65. ^{31}P NMR of compound **14b**.

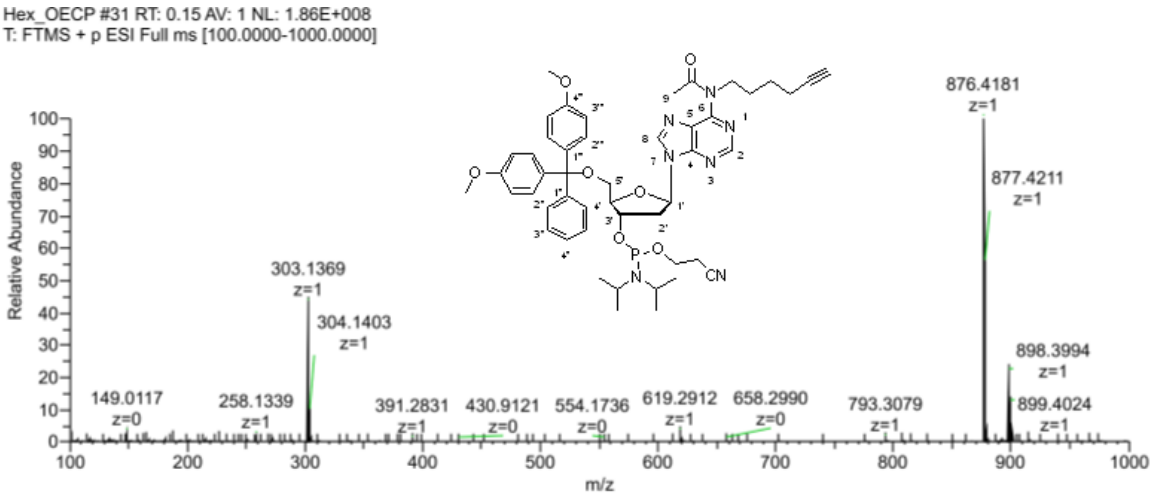


Figure S66. HRMS (ESI) of compound **14b**.

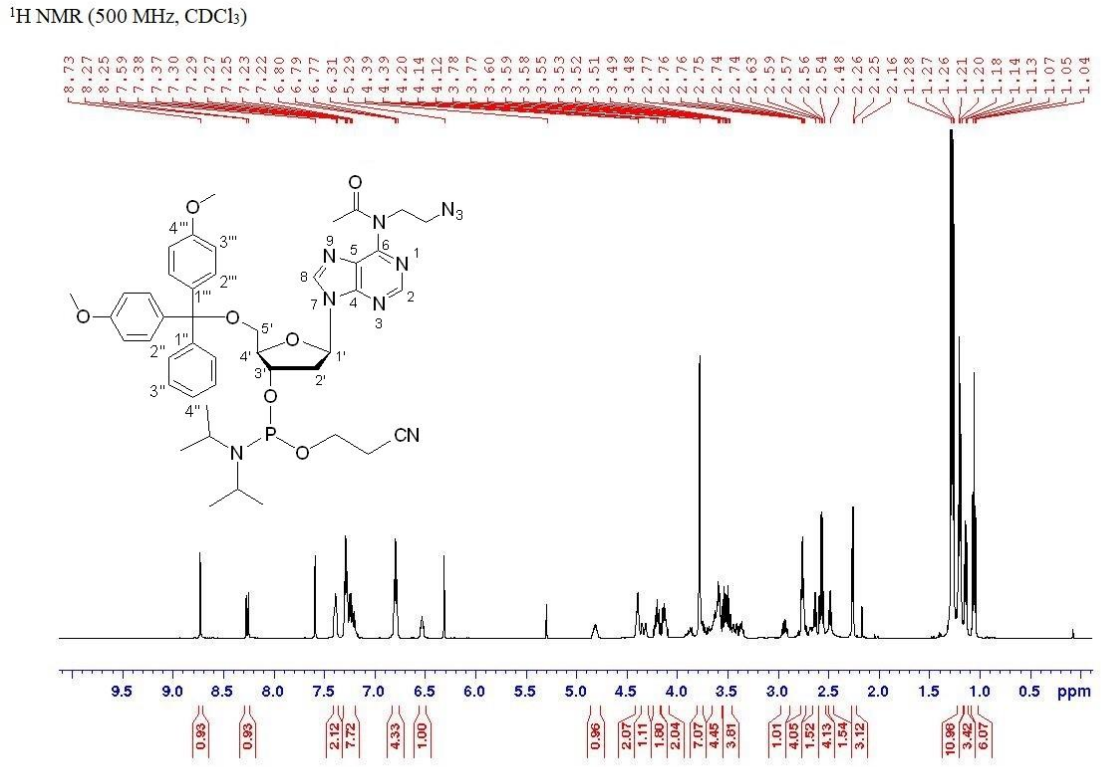


Figure S67. ^{1}H NMR spectrum of compound **14c**.

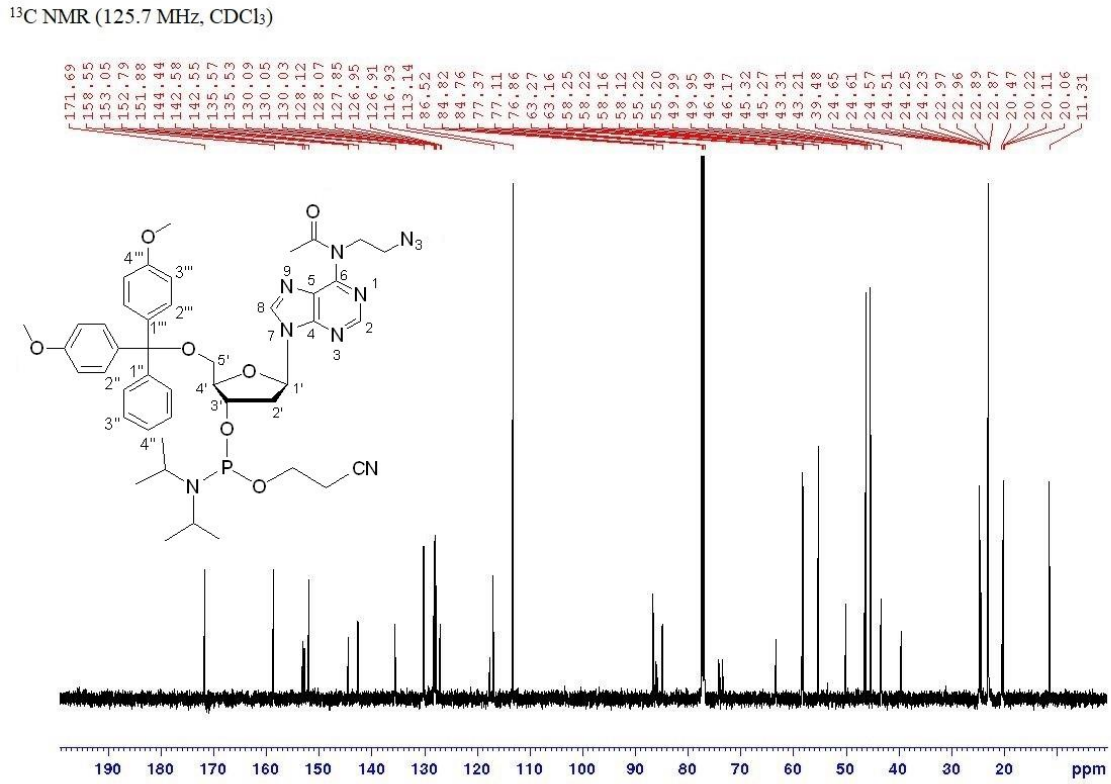


Figure S68. ^{13}C NMR spectrum of compound **14c**.

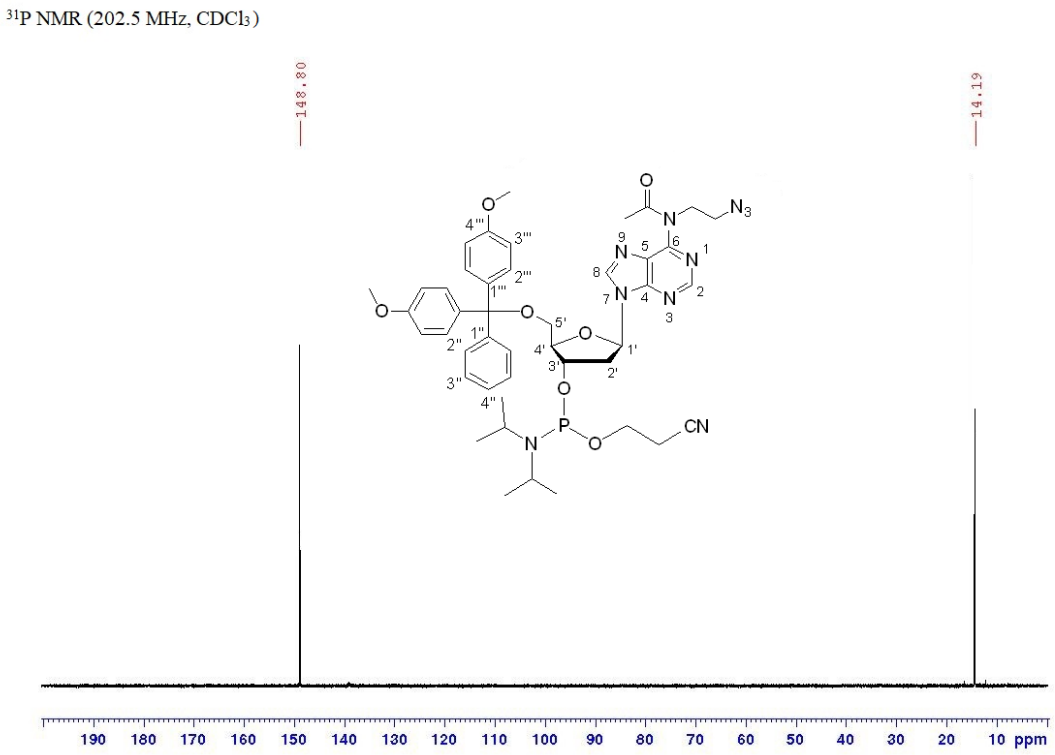


Figure S69. ^{31}P NMR spectrum of compound **14c**.

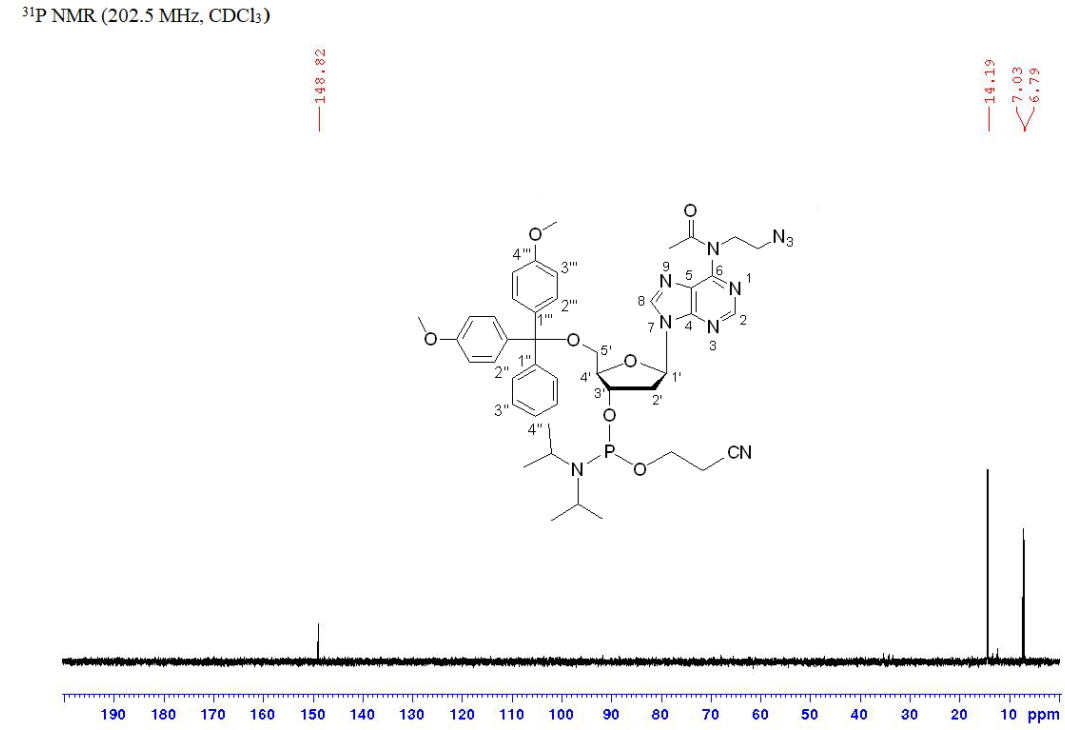


Figure S70. ^{31}P NMR spectrum of compound **14c** after 24 hours at rt showing significant degradation of phosphoramidite **14c**.

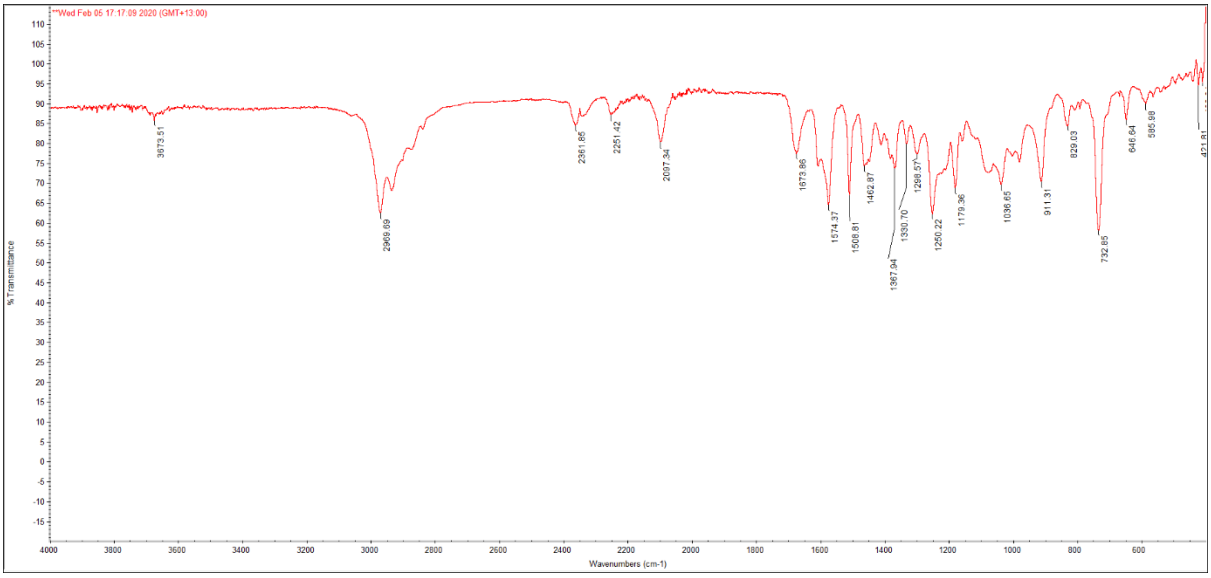


Figure S71. IR(ATR) spectrum of compound **14c**.

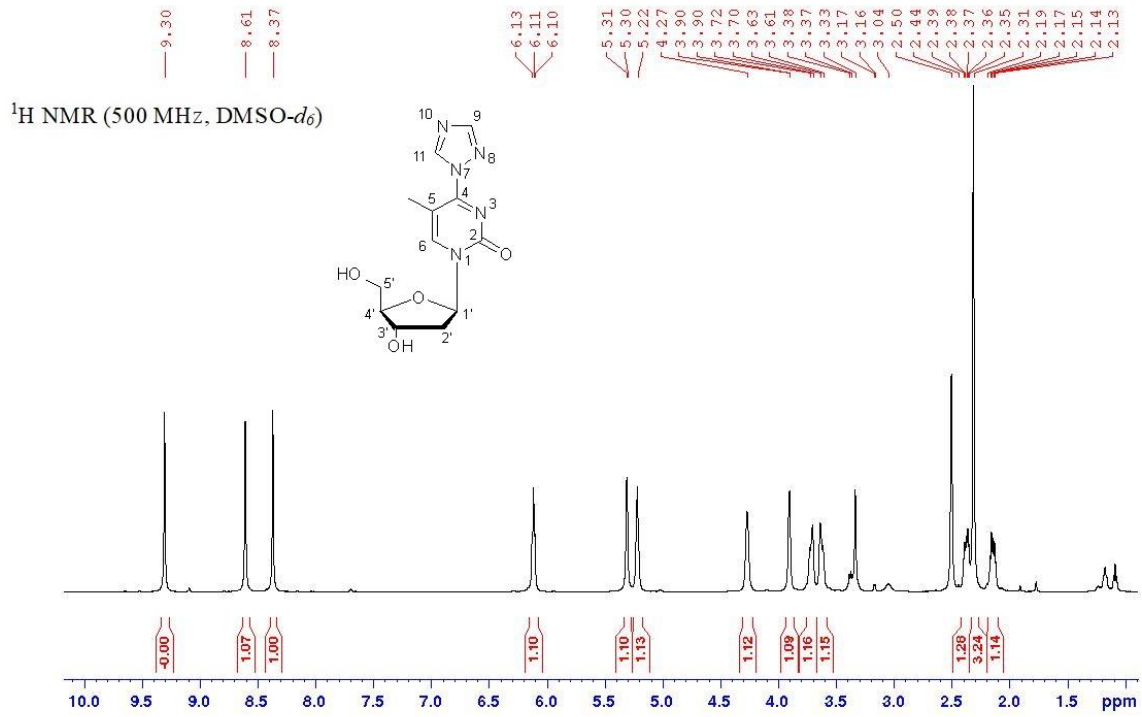


Figure S72. ^{1}H NMR spectrum of compound **15**.

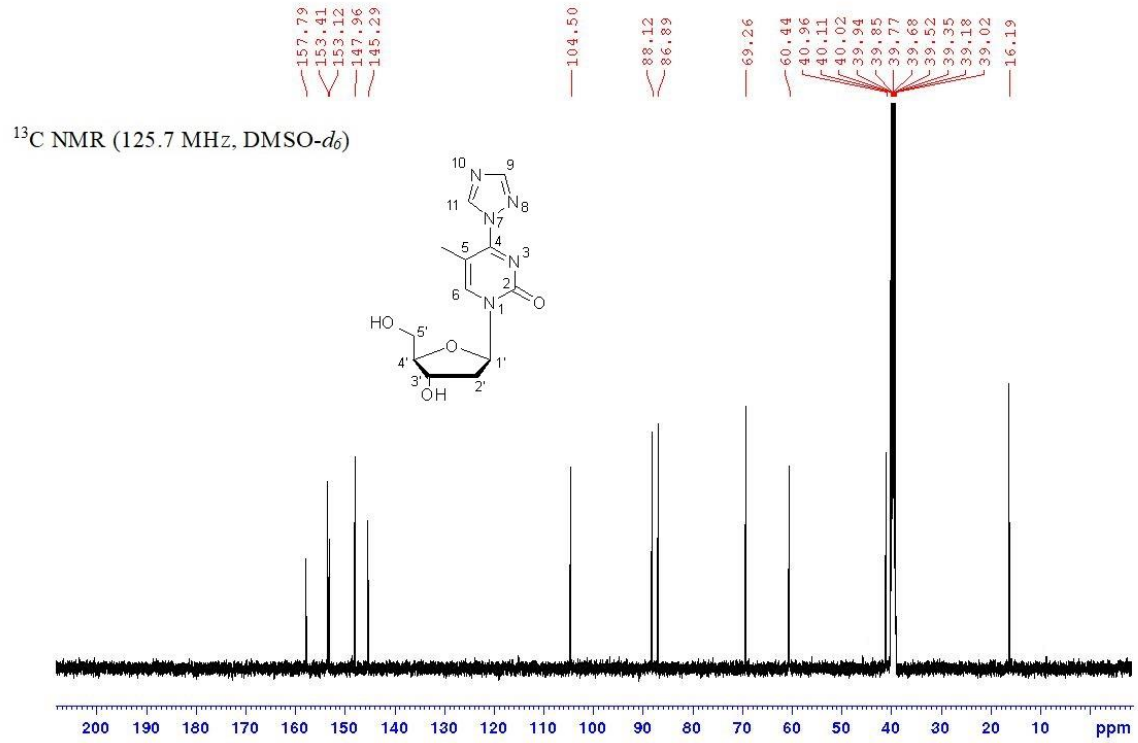


Figure S73. ^{13}C NMR spectrum of compound **15**.

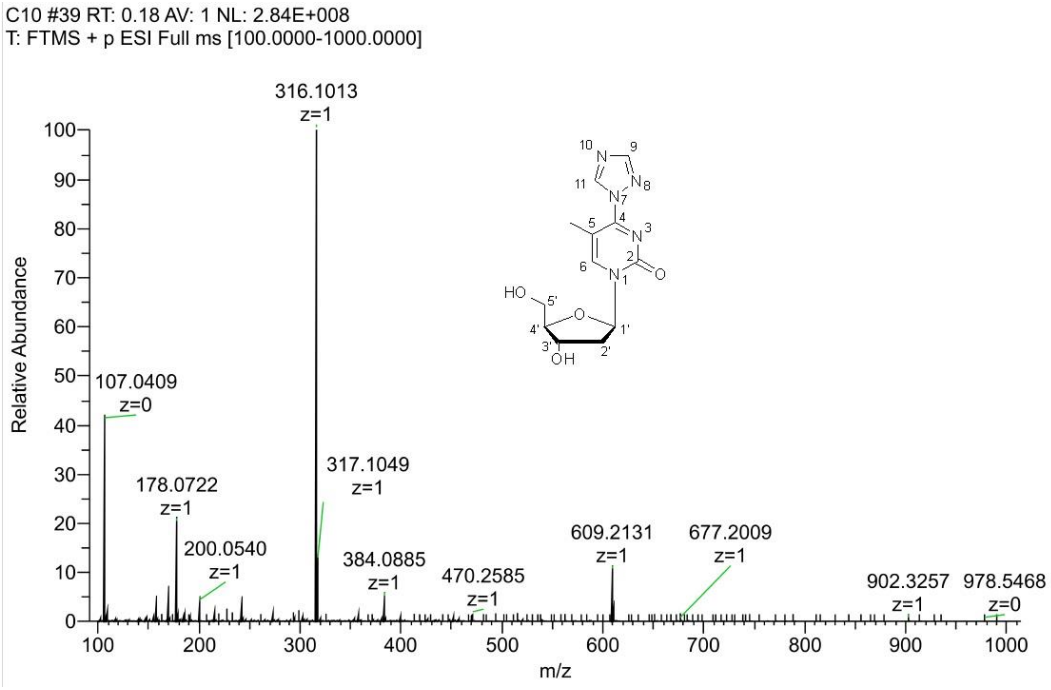


Figure S74. HRMS (ESI) of compound **15**.

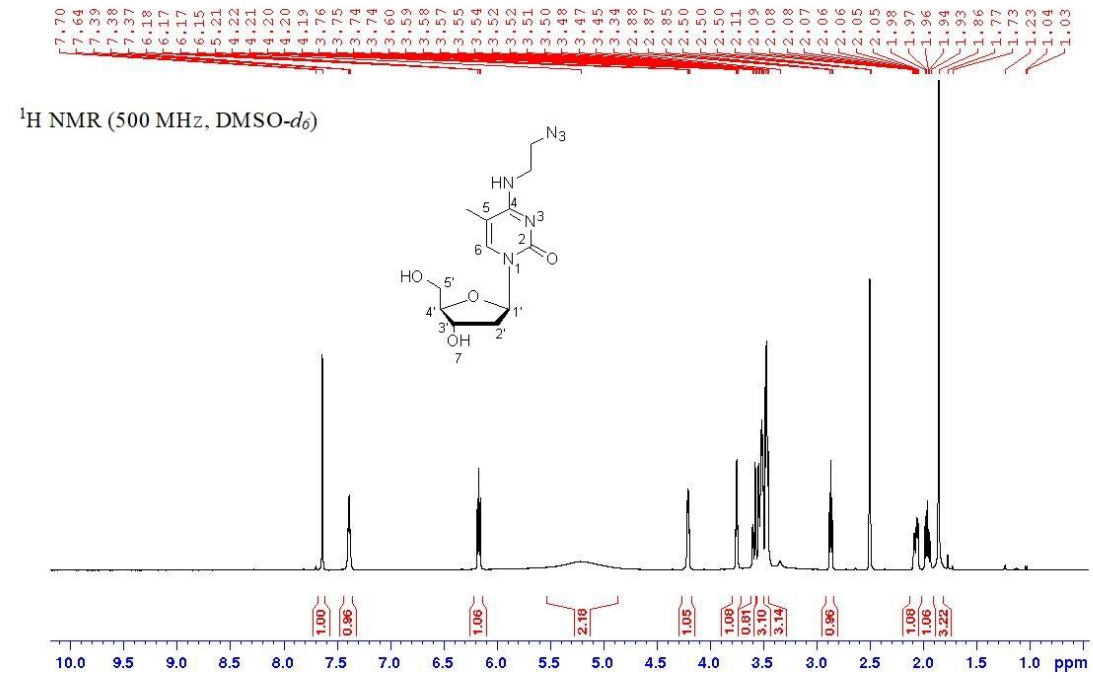


Figure S75. ^{1}H NMR spectrum of compound **16**.

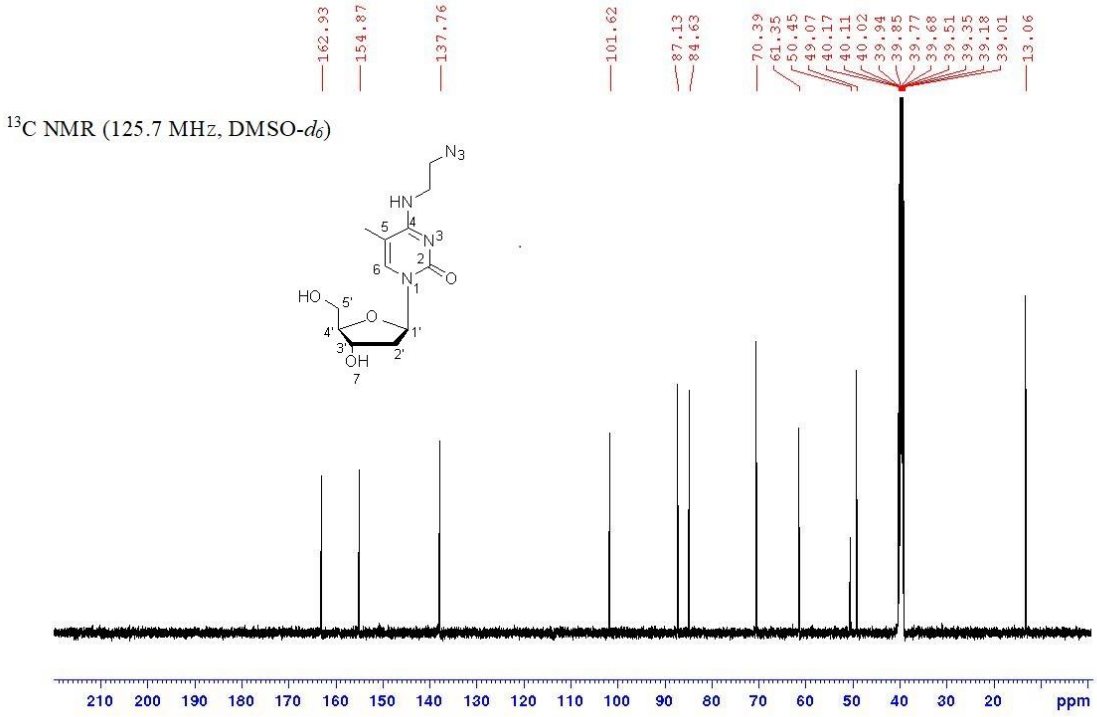


Figure S76. ^{13}C NMR spectrum of compound **16**.

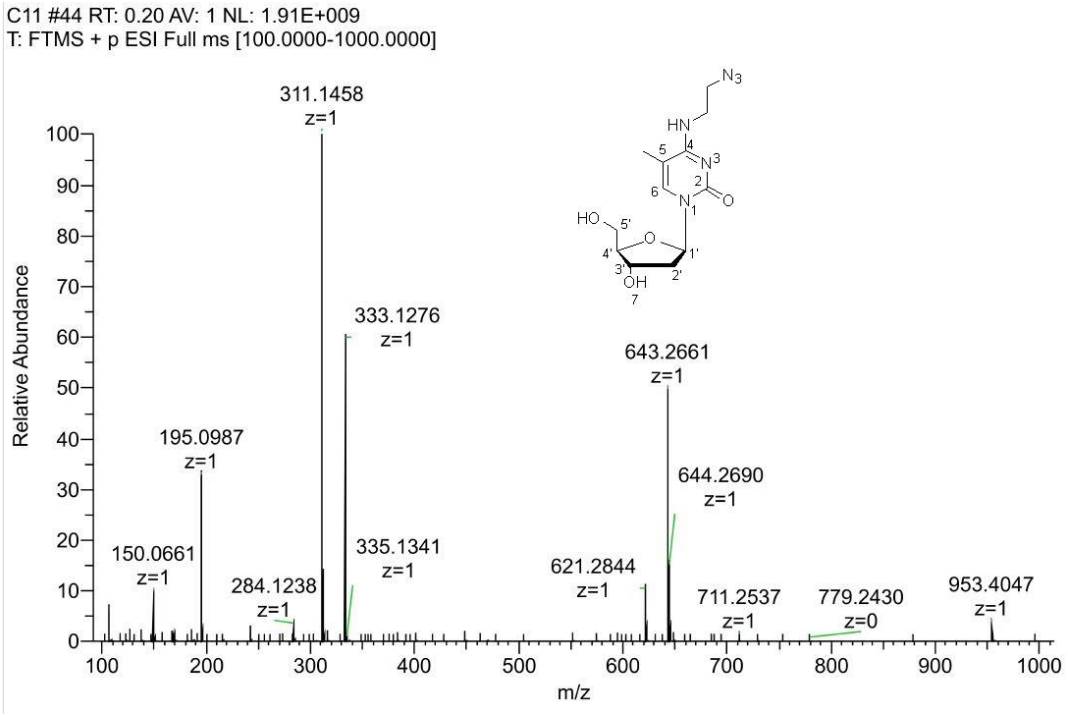


Figure S77. HRMS (ESI) of compound **16**.

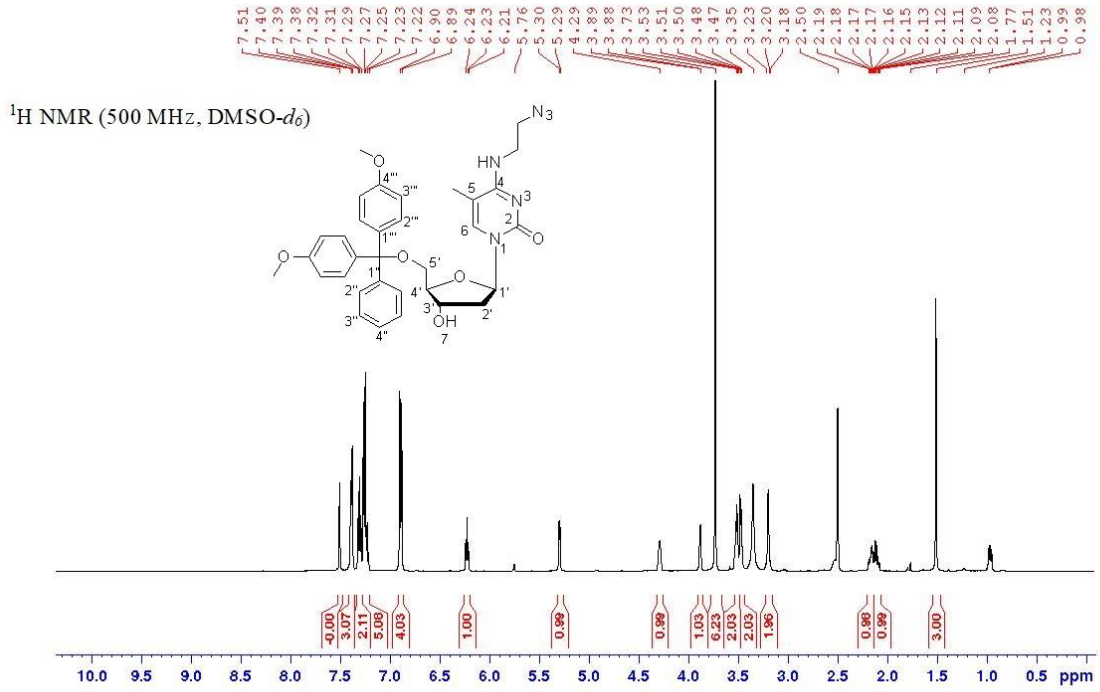


Figure S78. ^{1}H NMR spectrum of compound **17**.

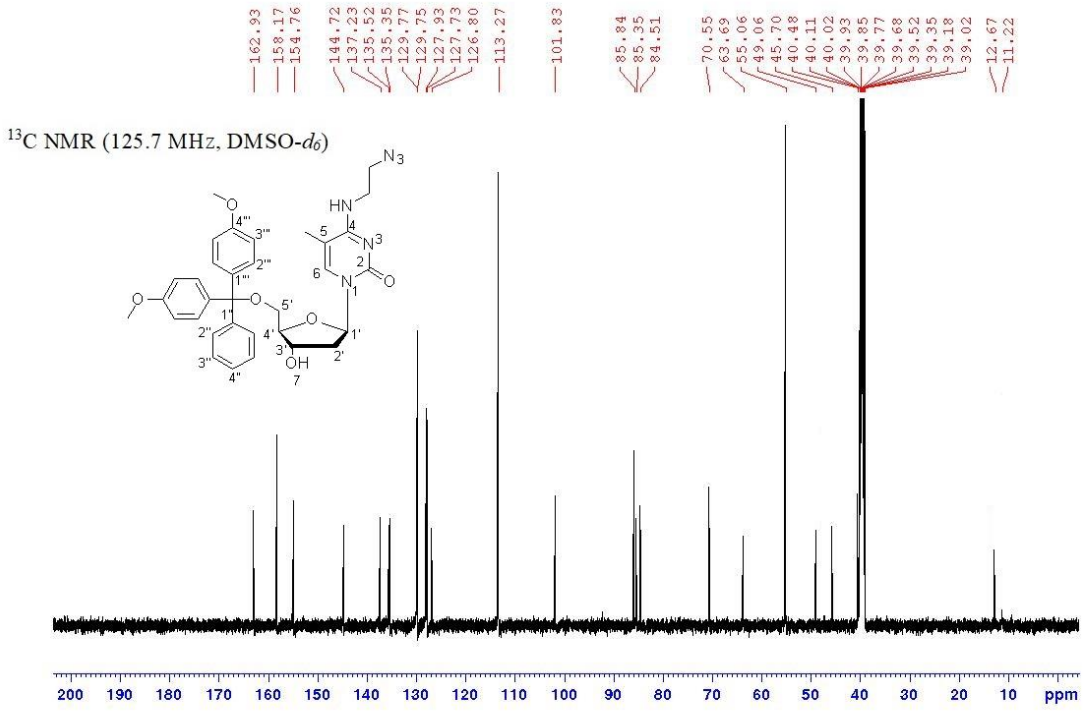


Figure S79. ^{13}C NMR spectrum of compound **17**.

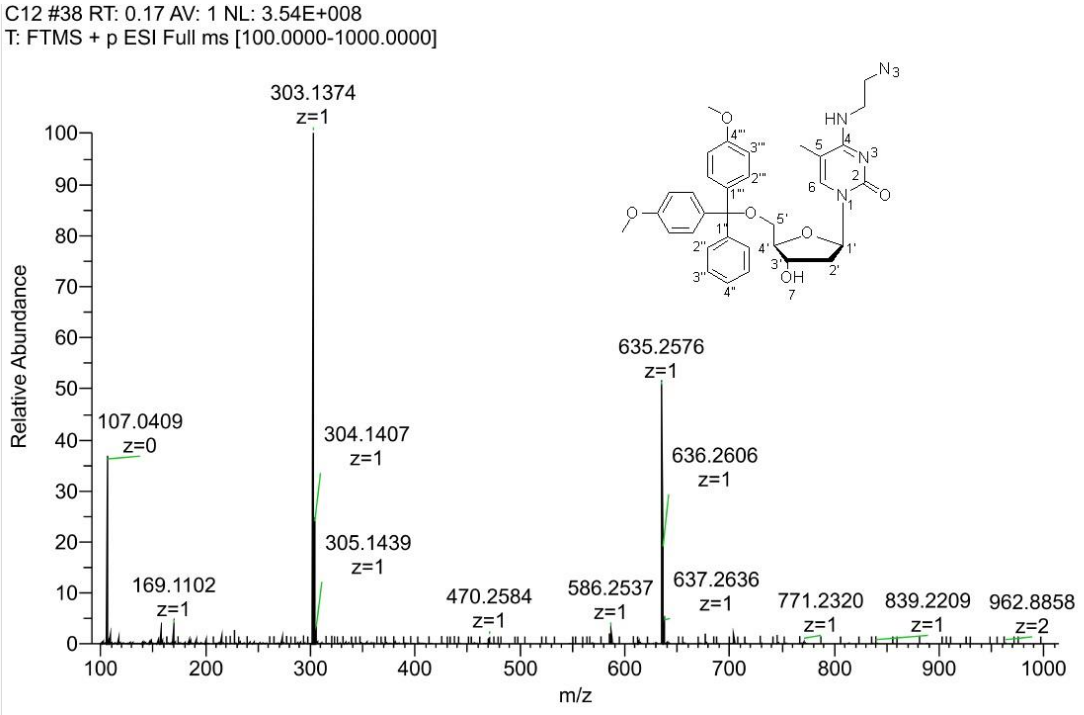


Figure S80. HRMS (ESI) of compound **17**.

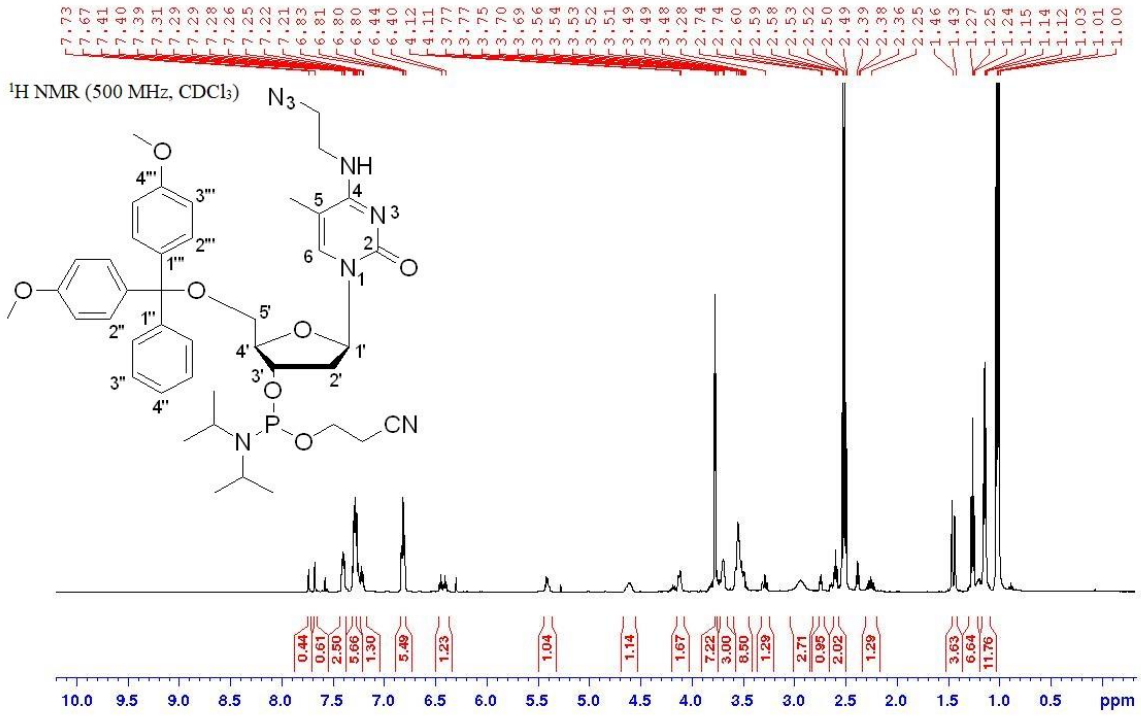


Figure S81. ^{1}H NMR spectrum of compound 18.

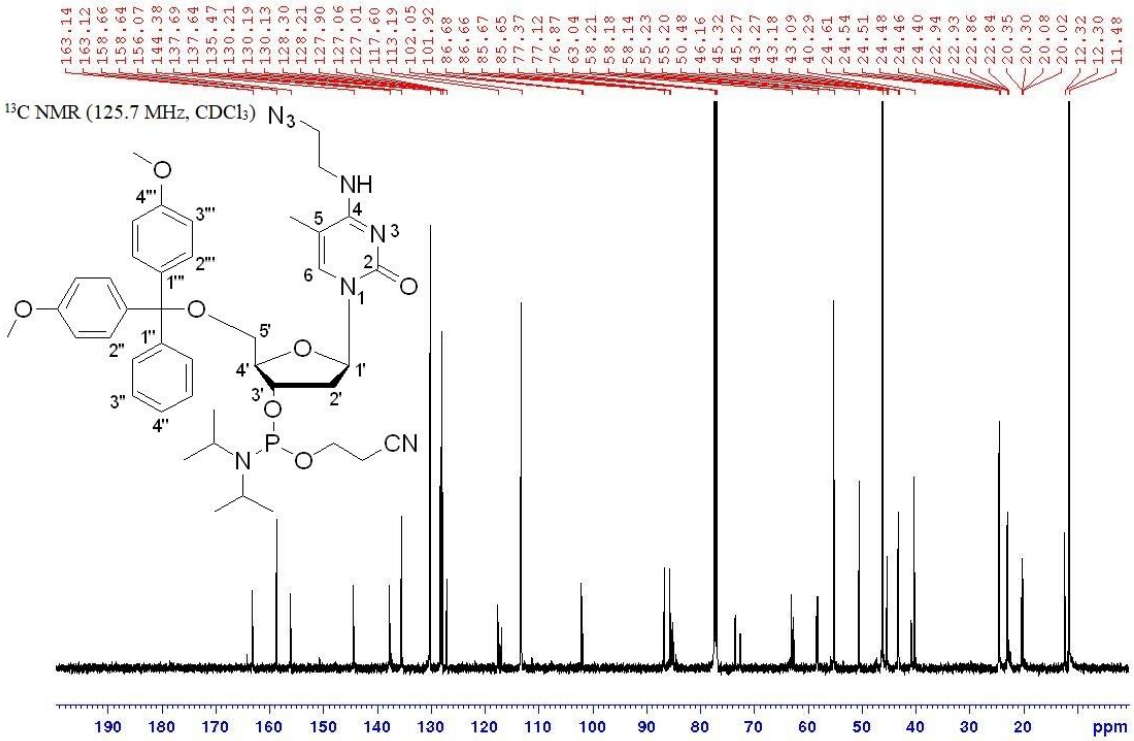


Figure S82. ^{13}C NMR spectrum of compound 18.

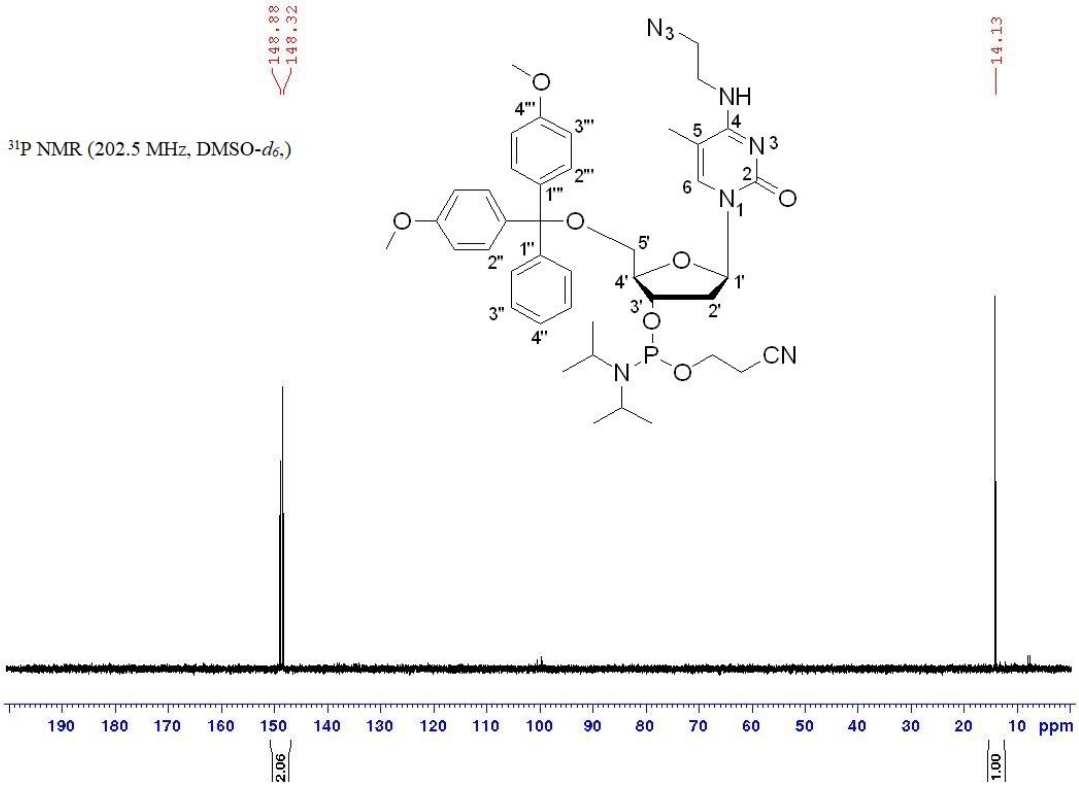


Figure S83. ^{31}P NMR spectrum of compound 18.

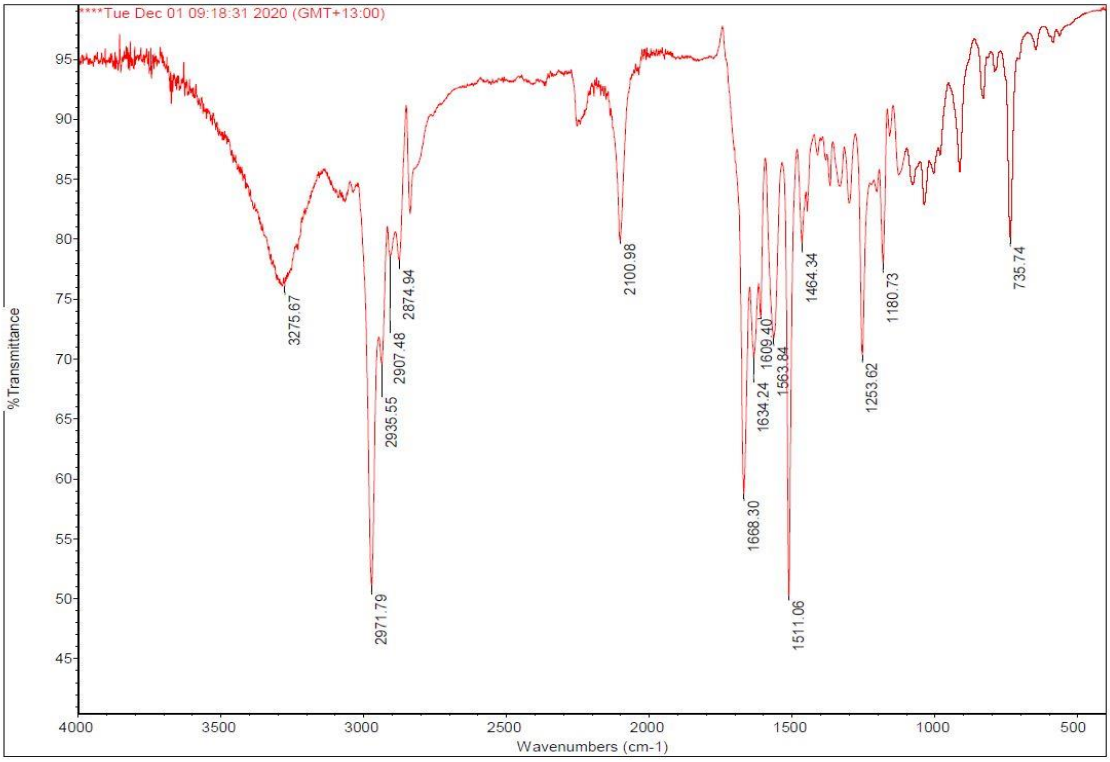


Figure S84. IR spectrum of compound 18.

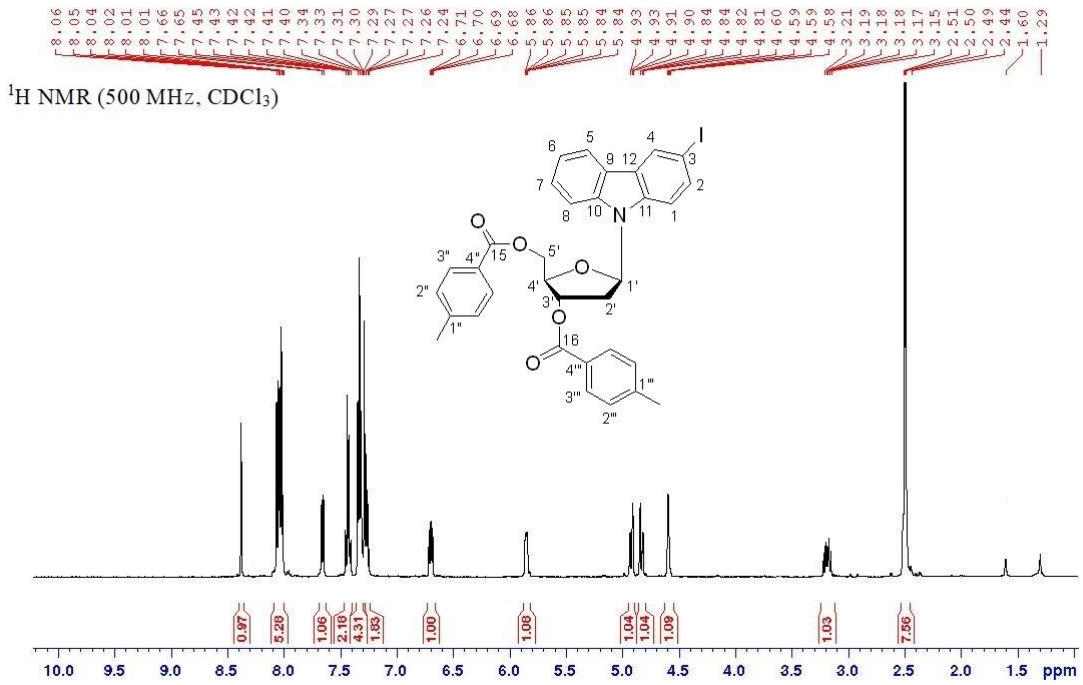


Figure S85. ^{1}H NMR spectrum of compound **20**.

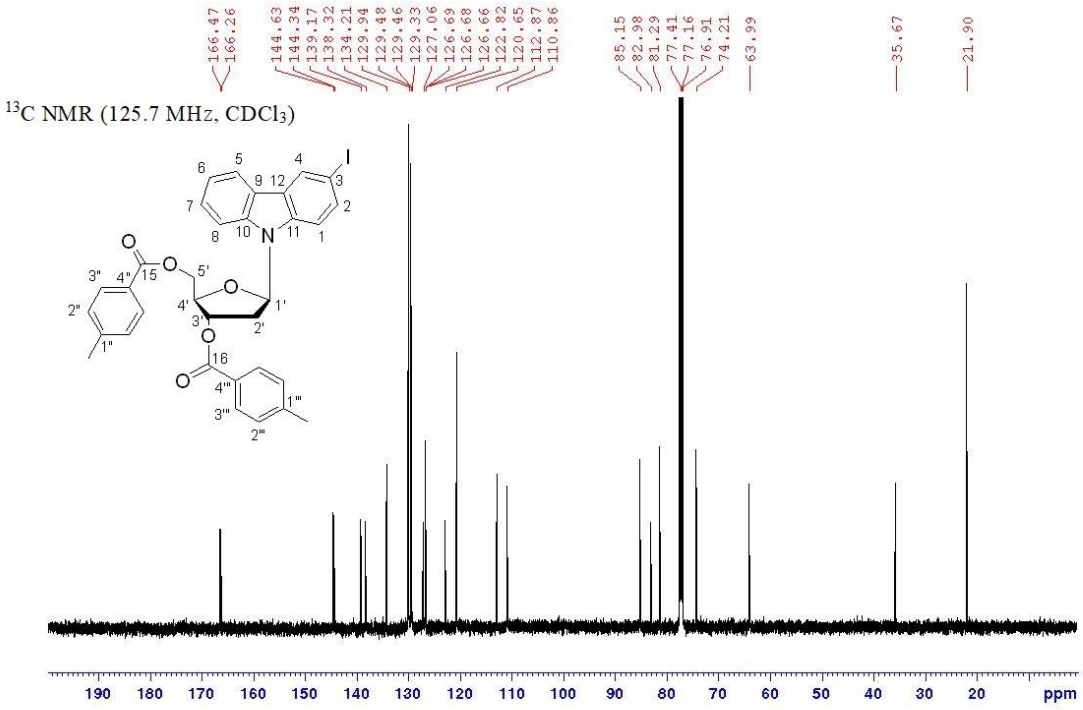


Figure S86. ^{13}C NMR spectrum of compound **20**.

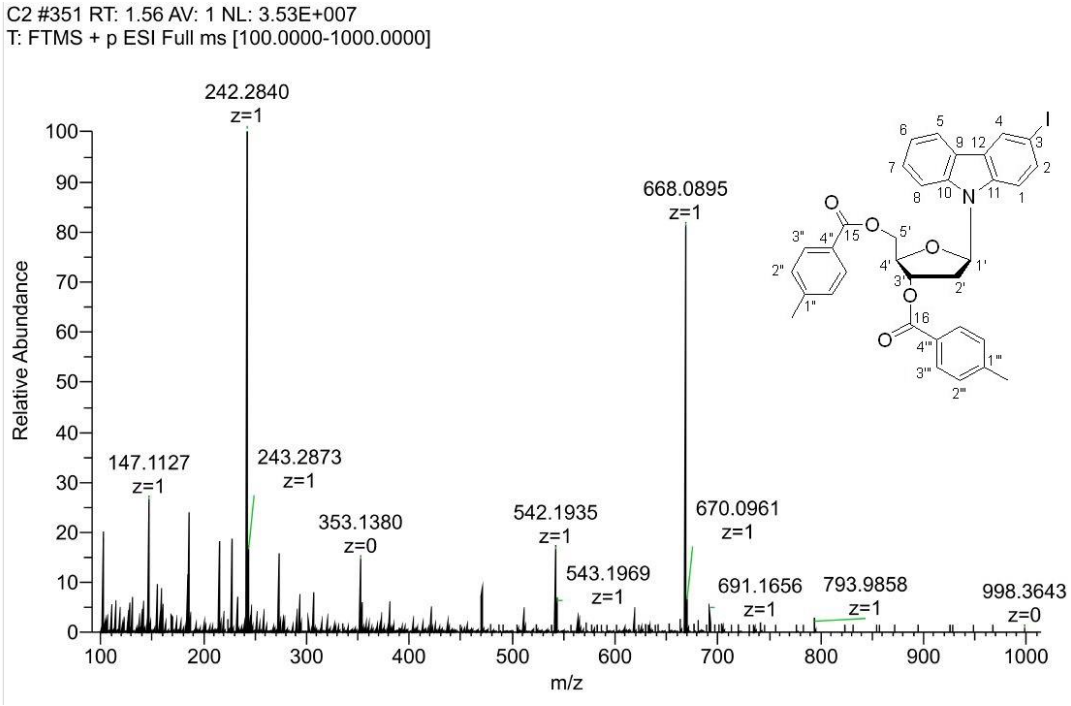


Figure S87. HRMS (ESI) of compound **20**.

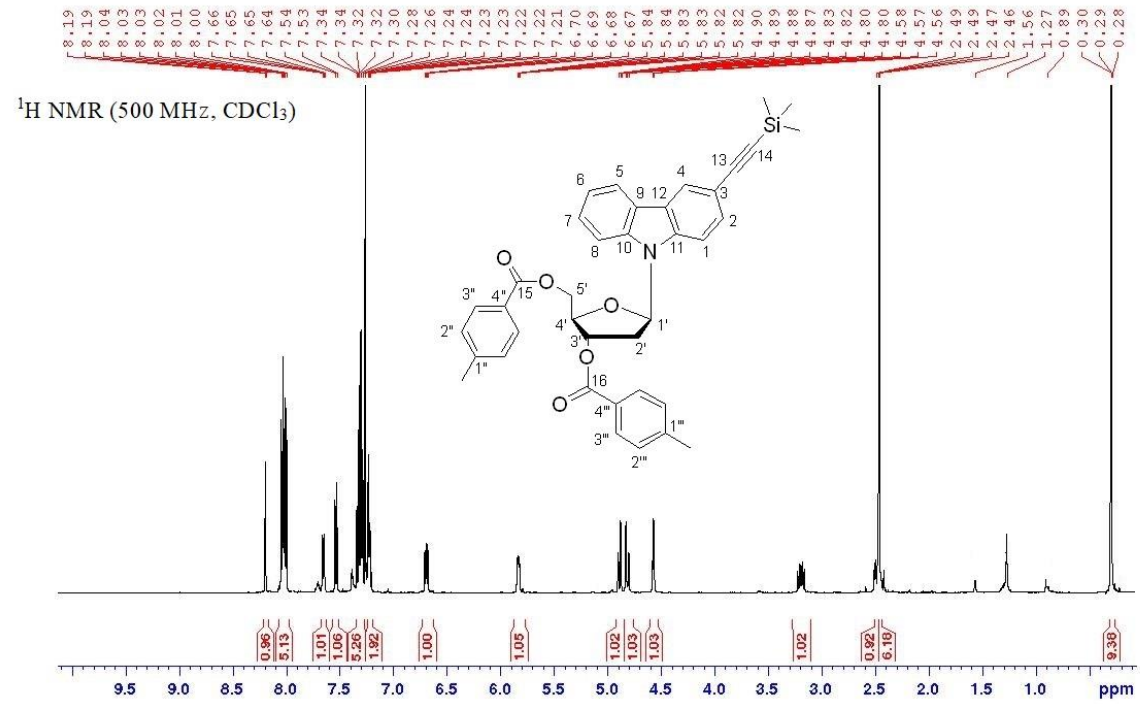


Figure S88. ^{1}H NMR spectrum of compound **21a**.

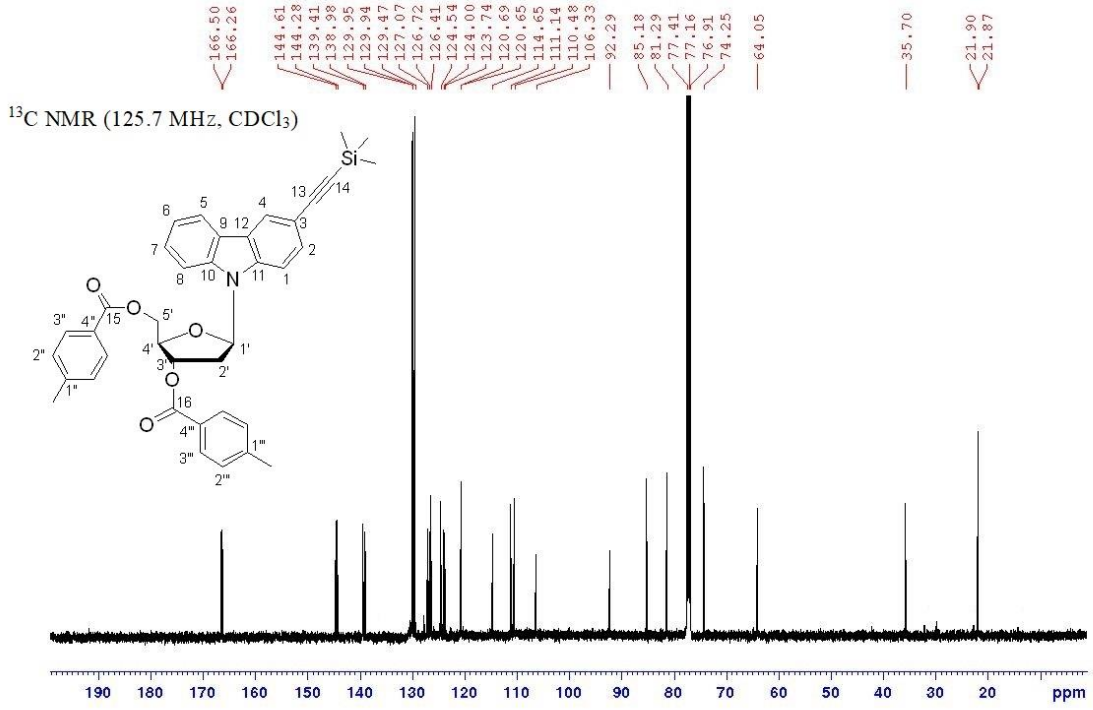


Figure S89. ^{13}C NMR spectrum of compound **21a**.

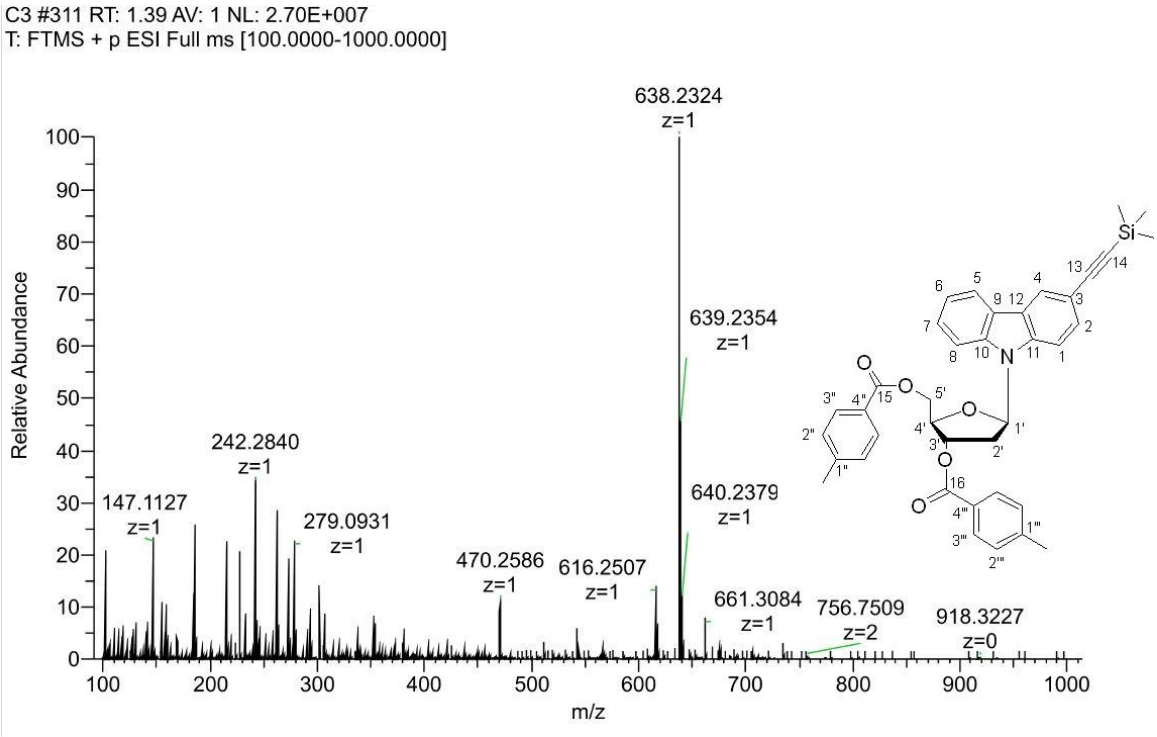


Figure S90. HRMS (ESI) of compound **21a**.

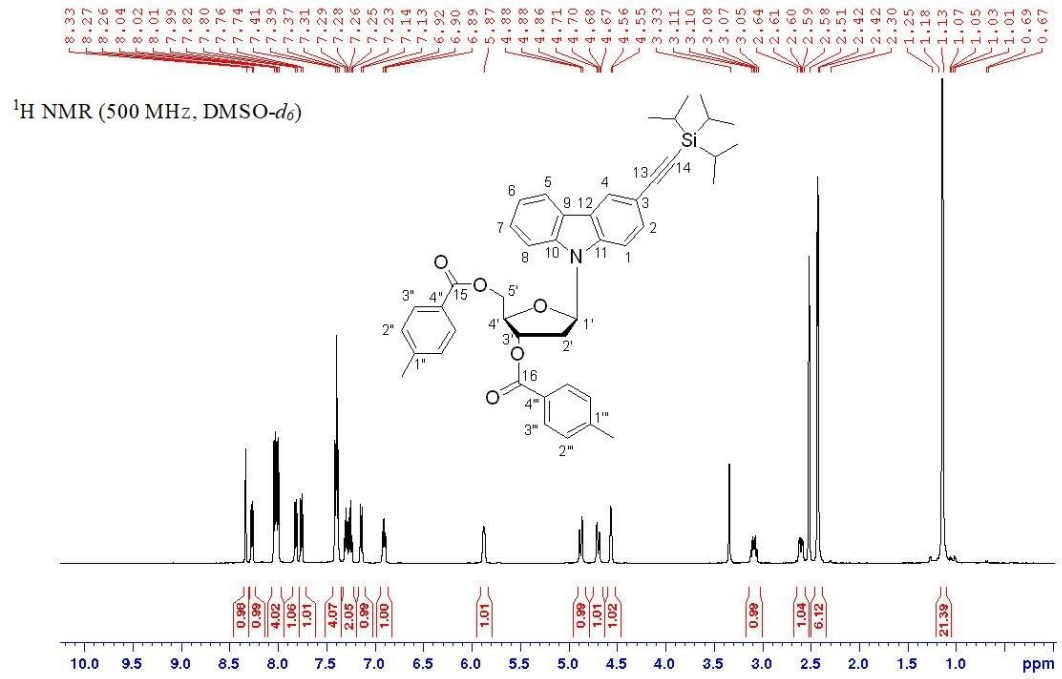


Figure S91. ^{1}H NMR spectrum of compound **21b**.

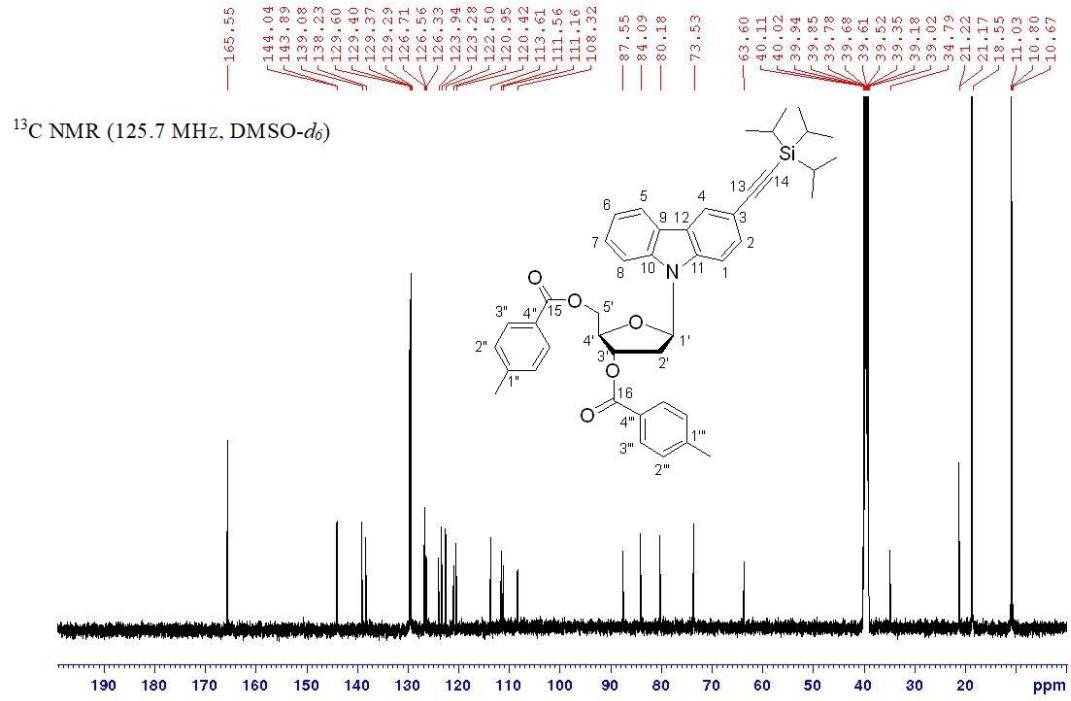


Figure S92. ^{13}C NMR spectrum of compound **21b**.

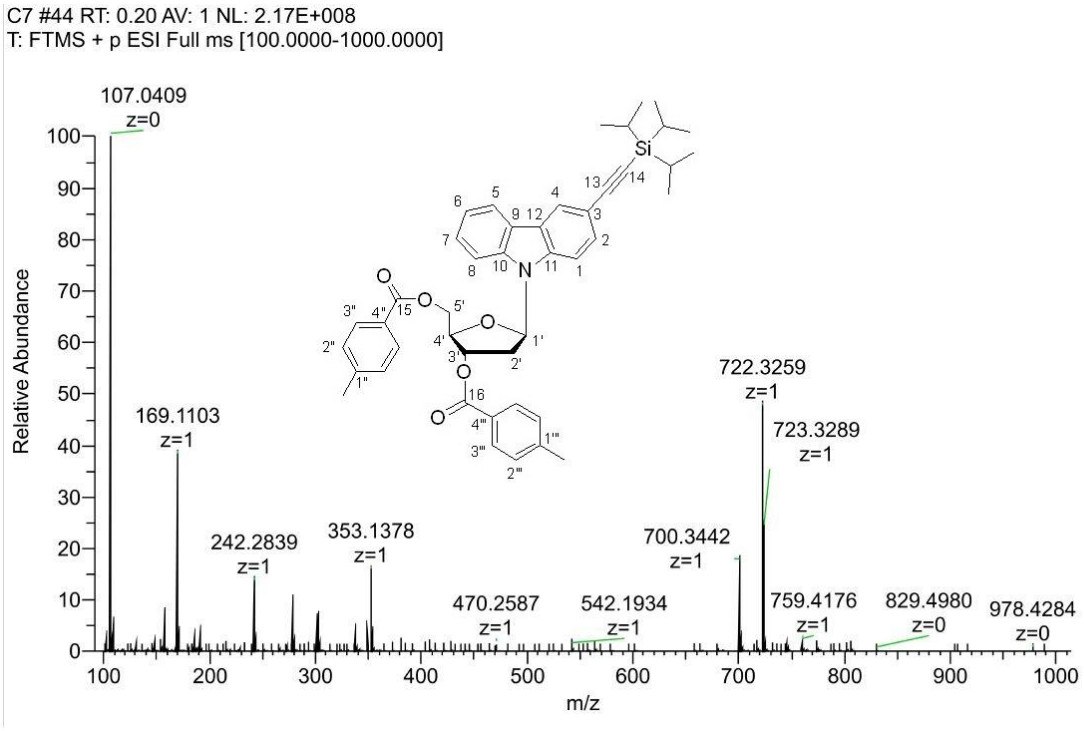


Figure S93. HRMS (ESI) of compound **21b**.

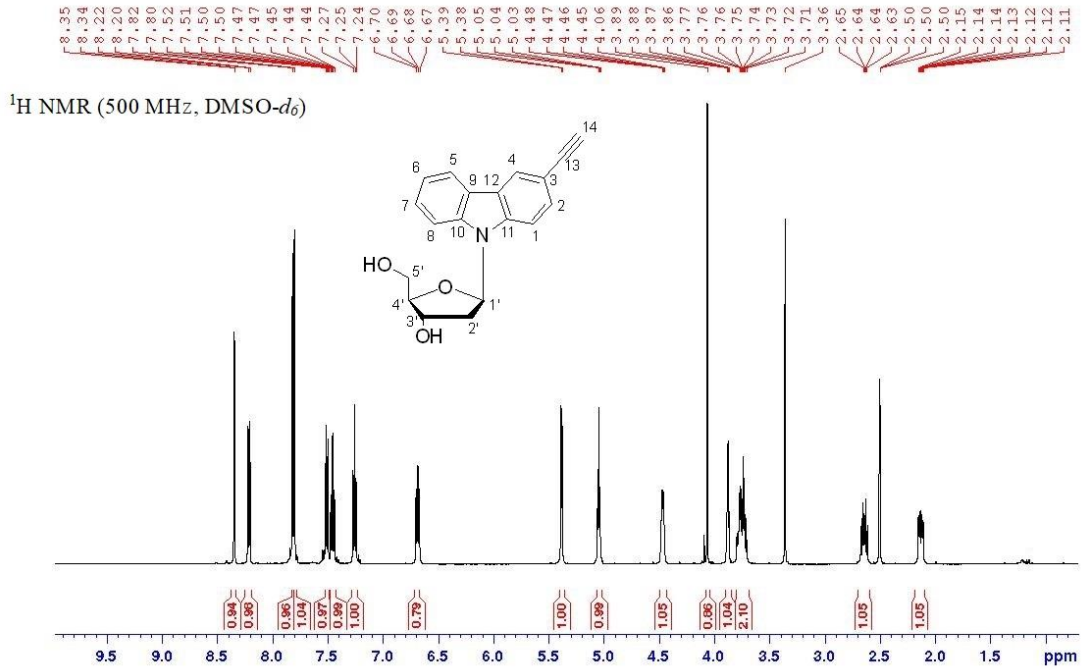


Figure S94. ^{1}H NMR spectrum of compound **22a**.

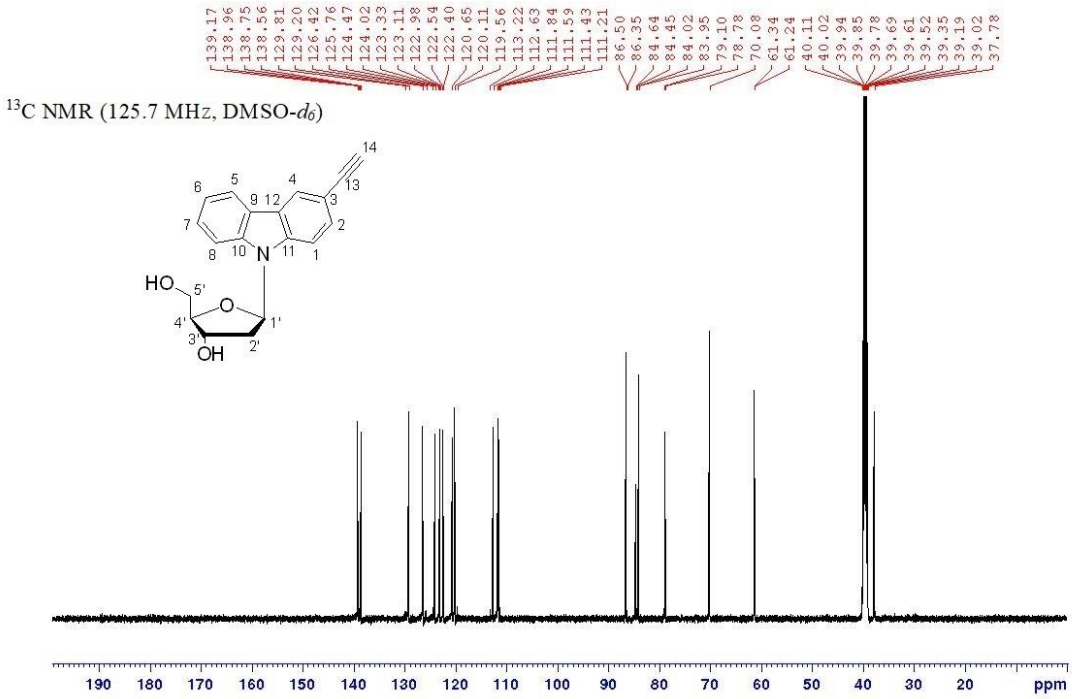


Figure S95. ^{13}C NMR spectrum of compound **22a**.

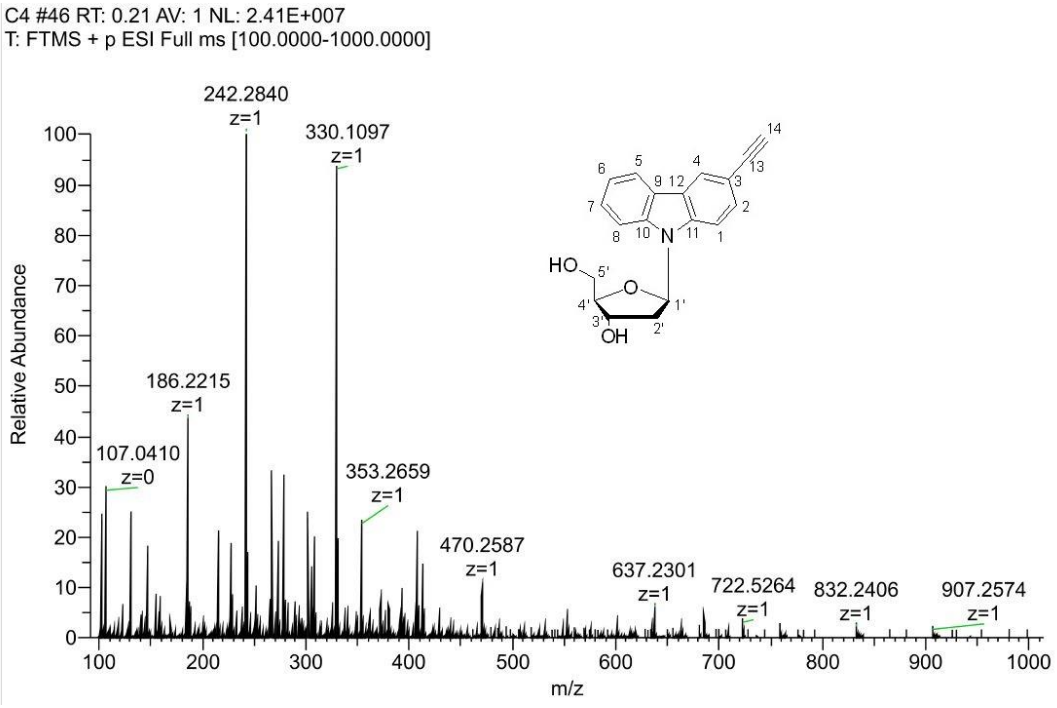


Figure S96. HRMS (ESI) of compound **22a**.

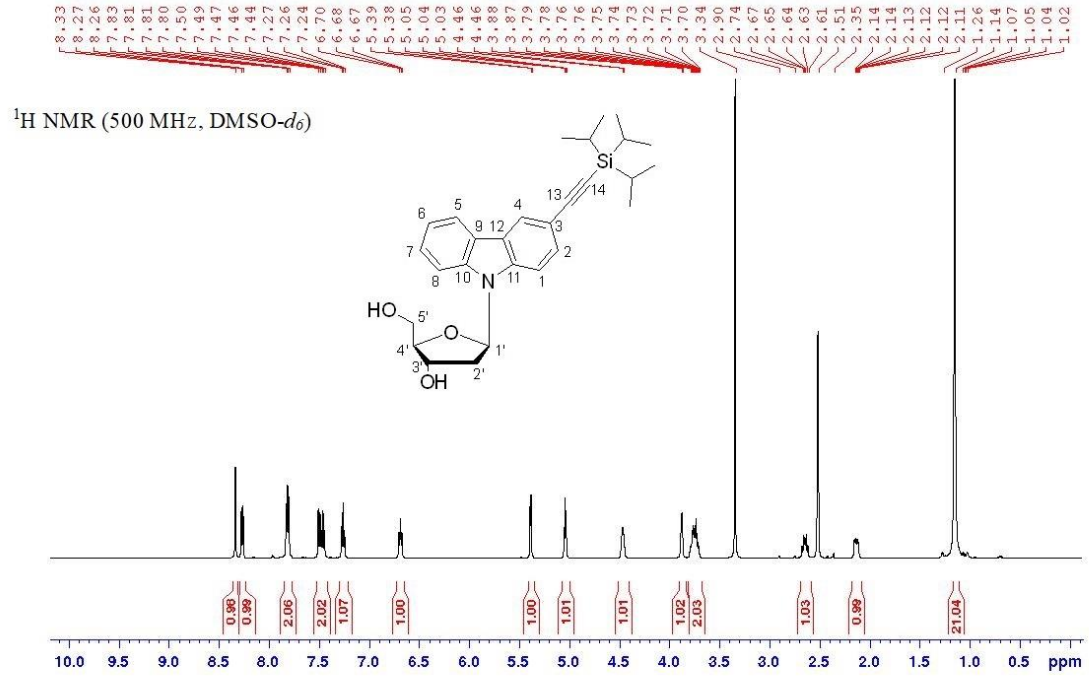


Figure S97. ^{1}H NMR spectrum of compound **22b**.

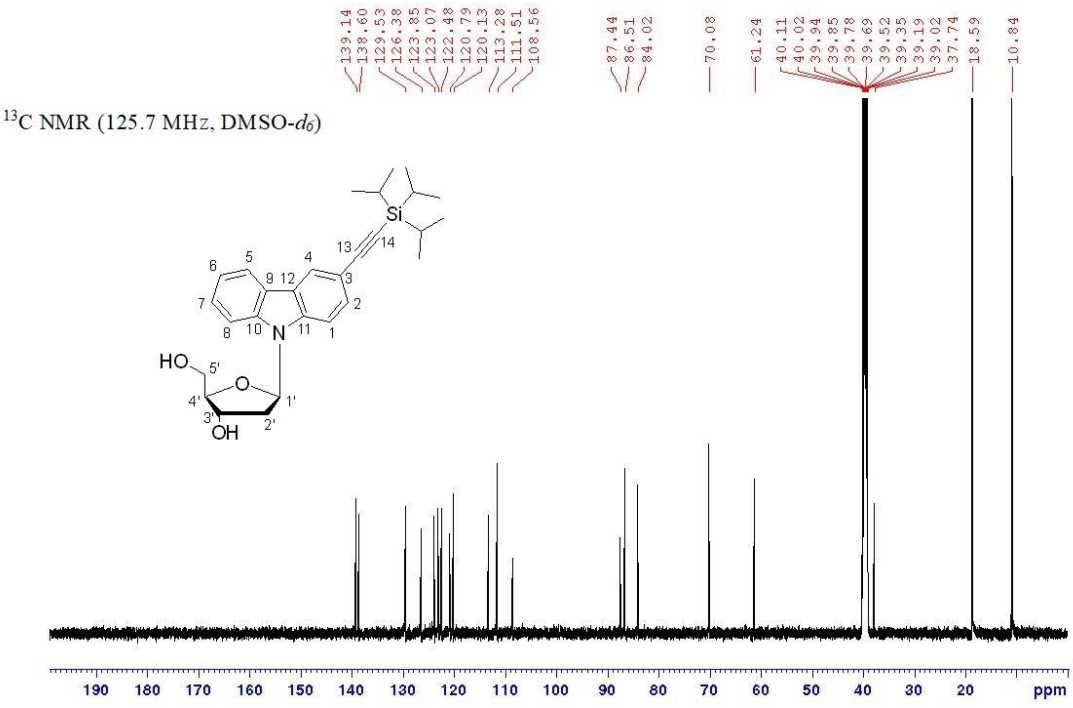


Figure S98. ^{13}C NMR spectrum of compound **22b**.

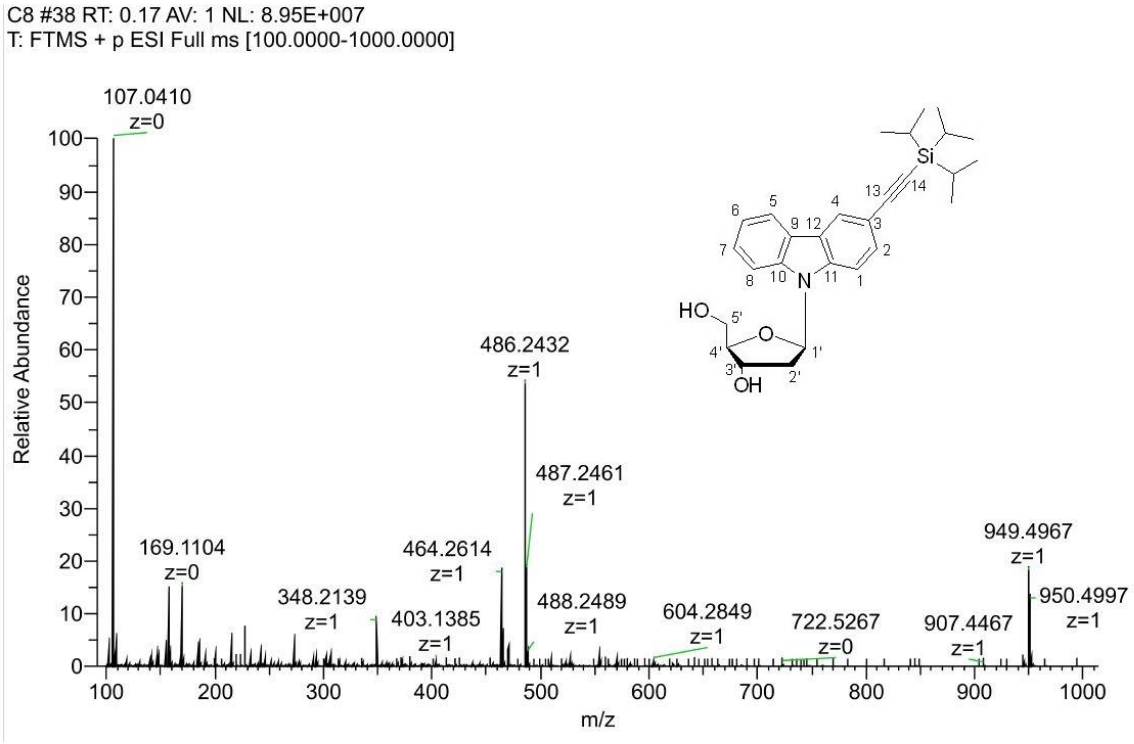


Figure S99. HRMS (ESI) of compound **22b**.

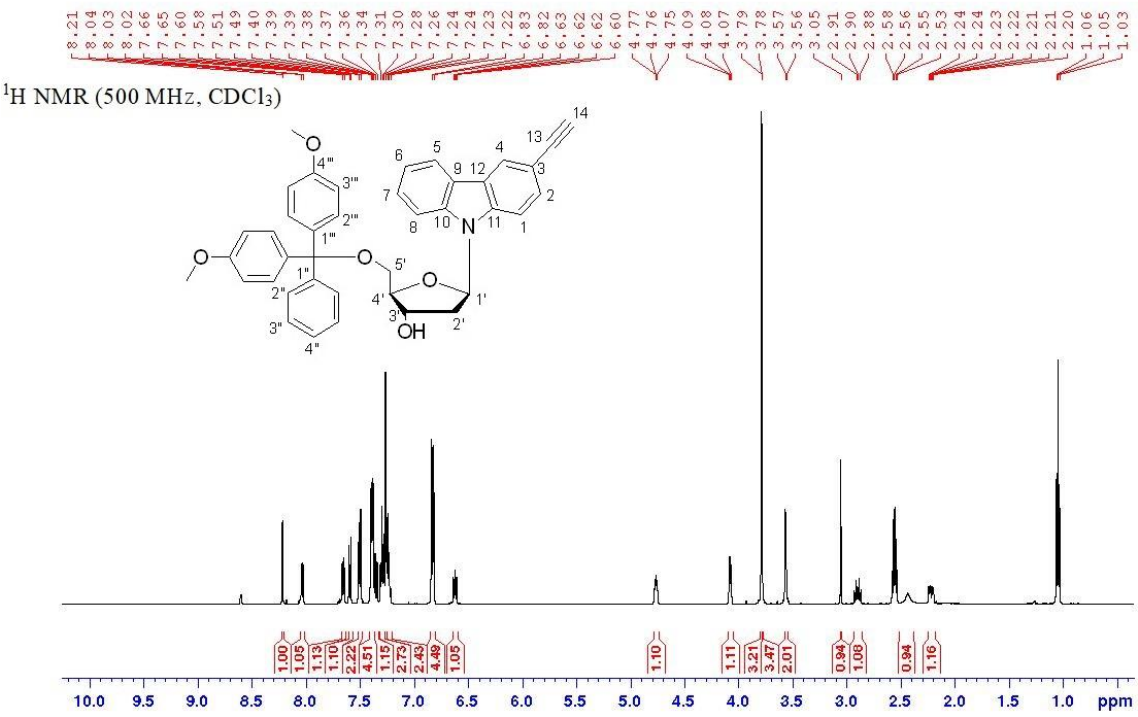


Figure S100. ^{1}H NMR spectrum of compound **23a**.

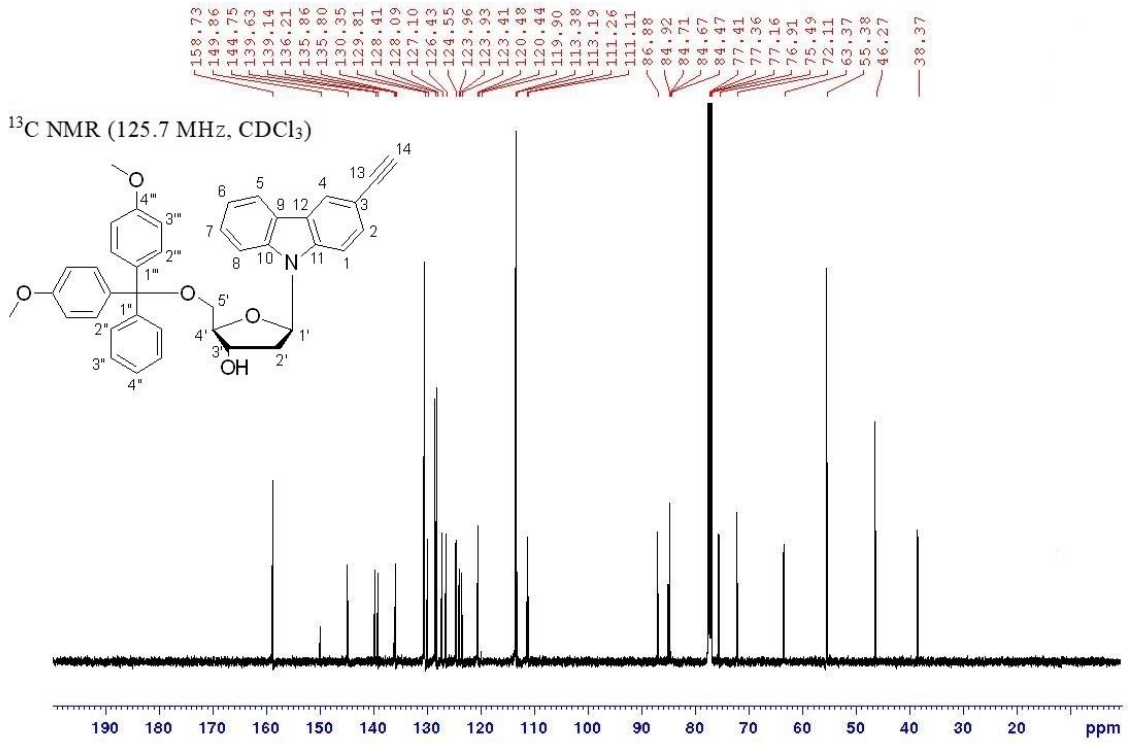


Figure S101. ^{13}C NMR spectrum of compound **23a**.

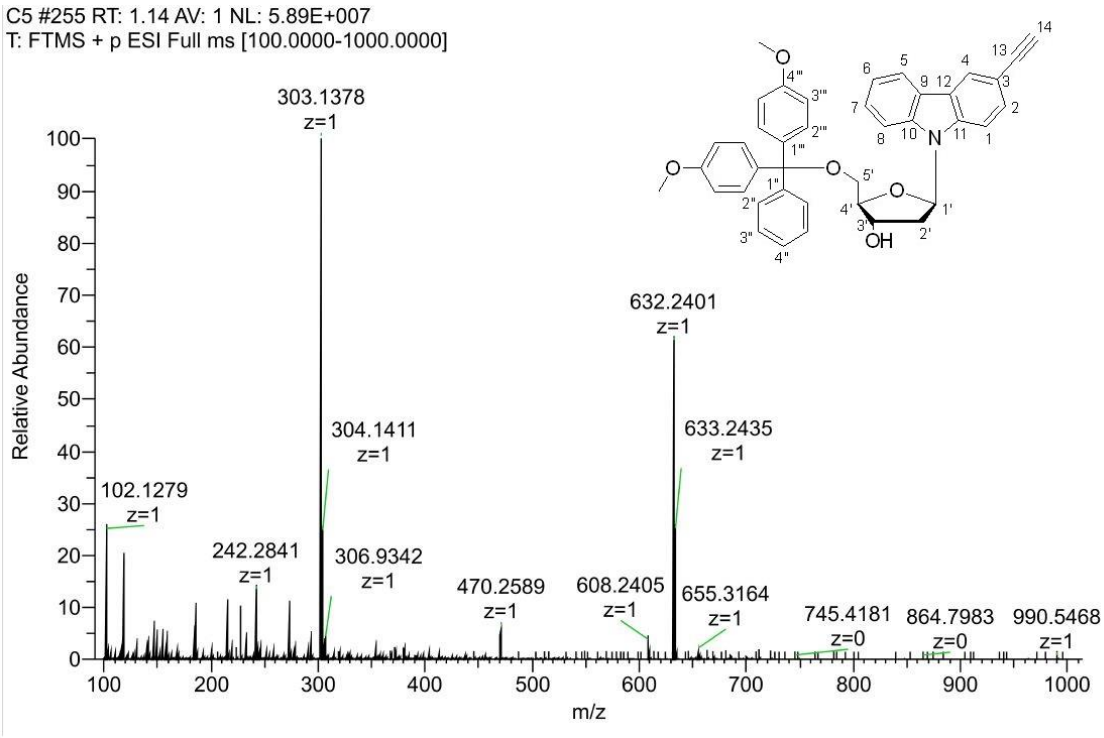


Figure S102. HRMS (ESI) of compound **23a**.

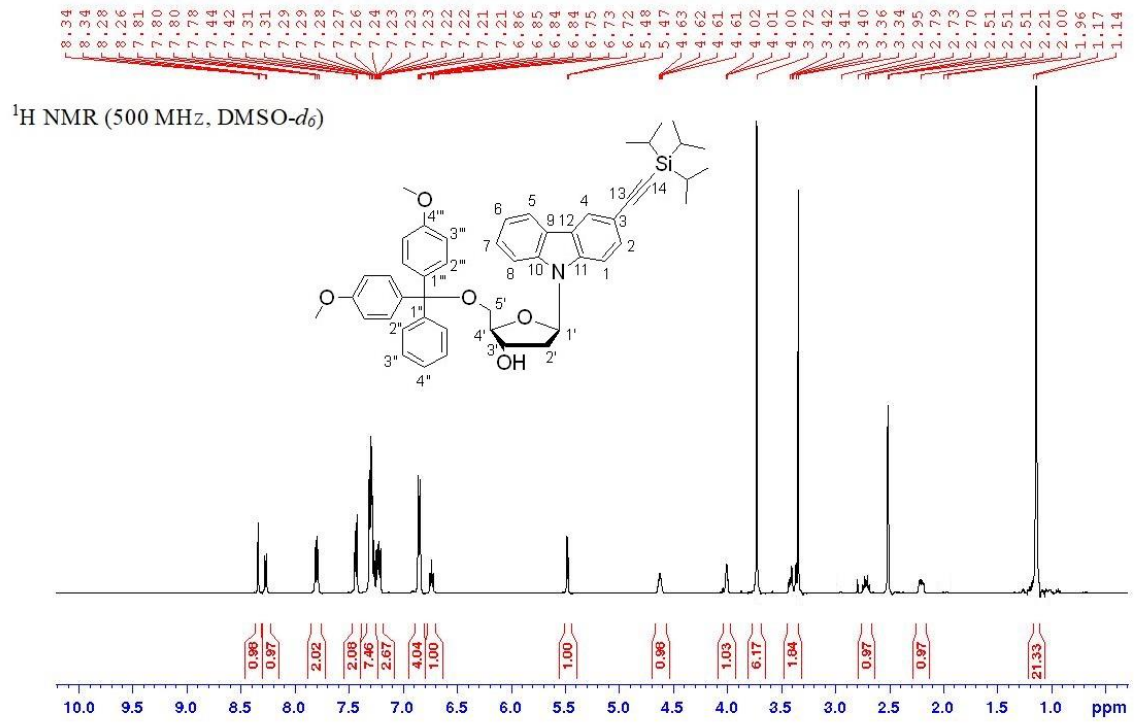


Figure S103. ^{1}H NMR spectrum of compound **23b**.

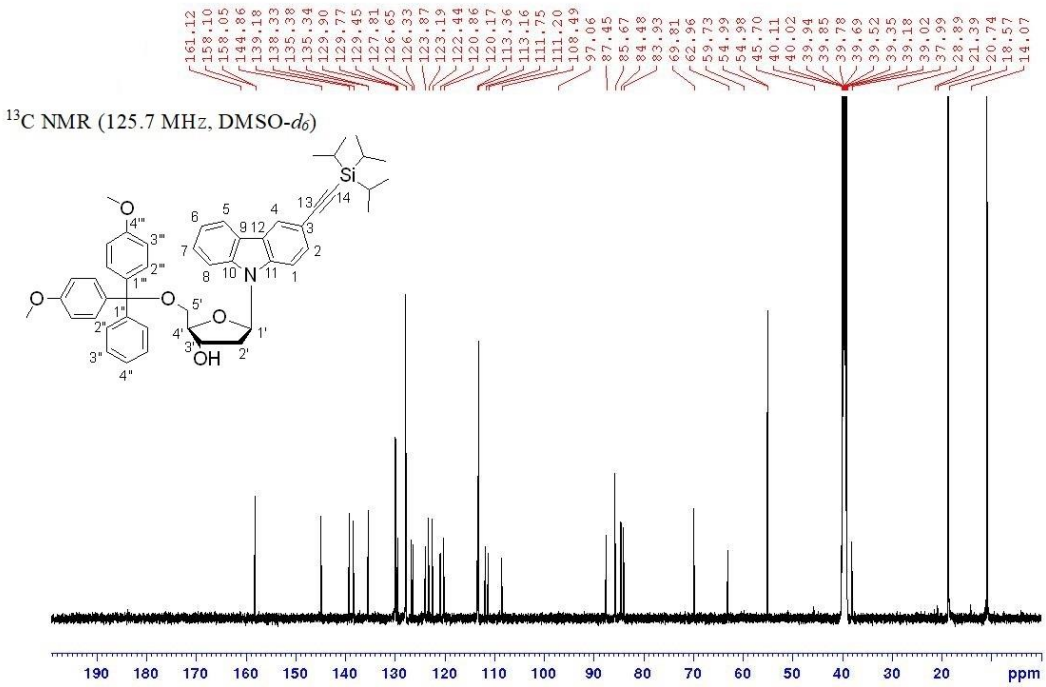


Figure S104. ^{13}C NMR spectrum of compound **23b**.

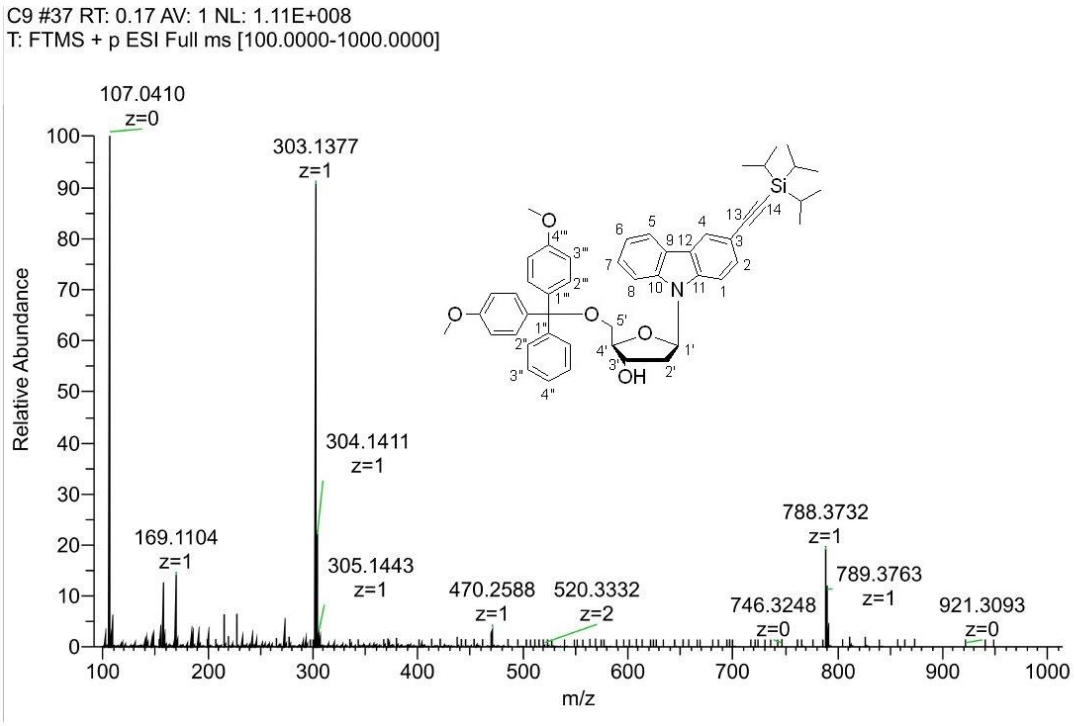


Figure S105. HRMS (ESI) of compound **23b**.

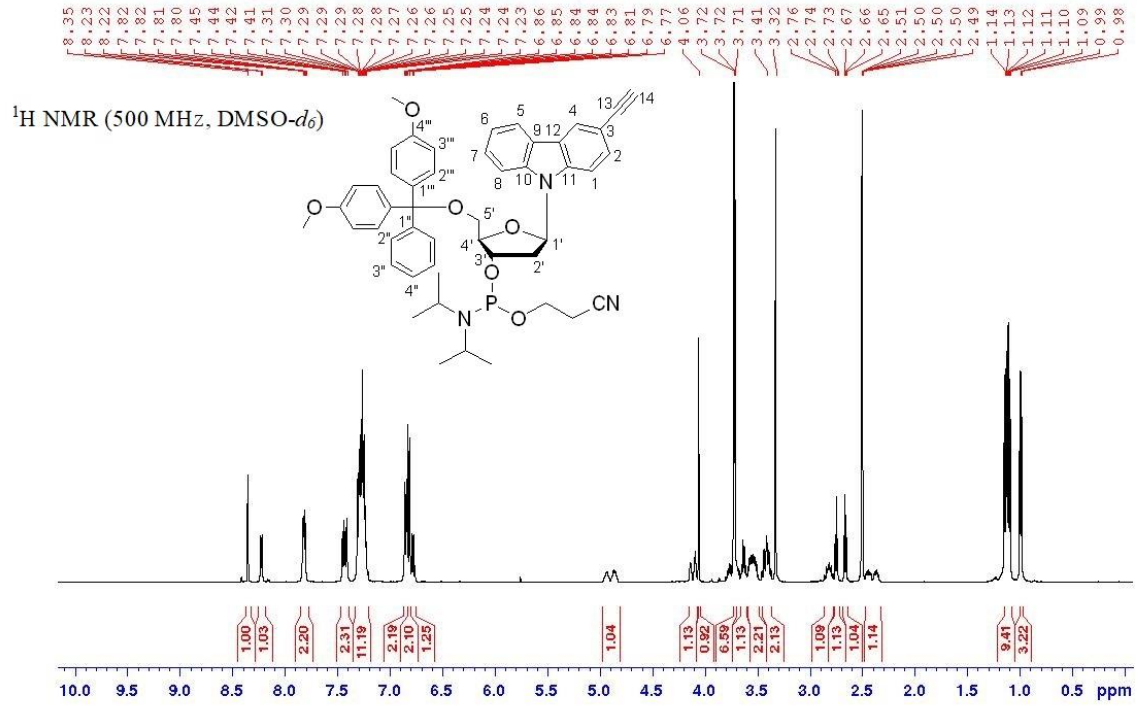


Figure S106. ^{1}H NMR spectrum of compound **24**.

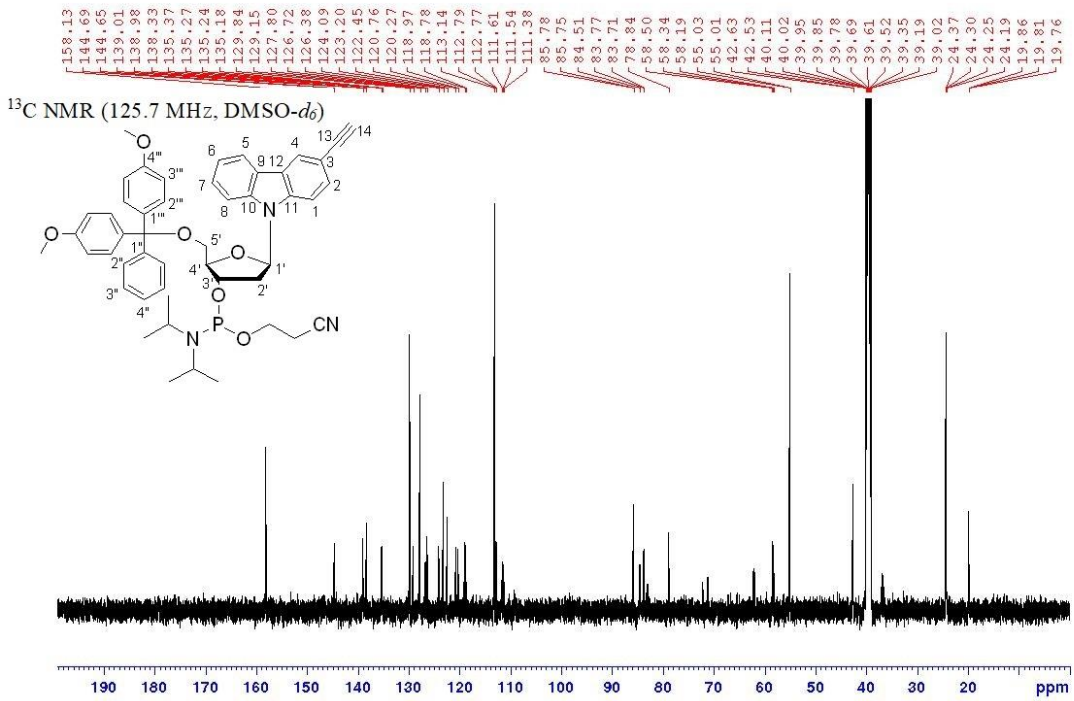


Figure S107. ^{13}C NMR spectrum of compound **24**.

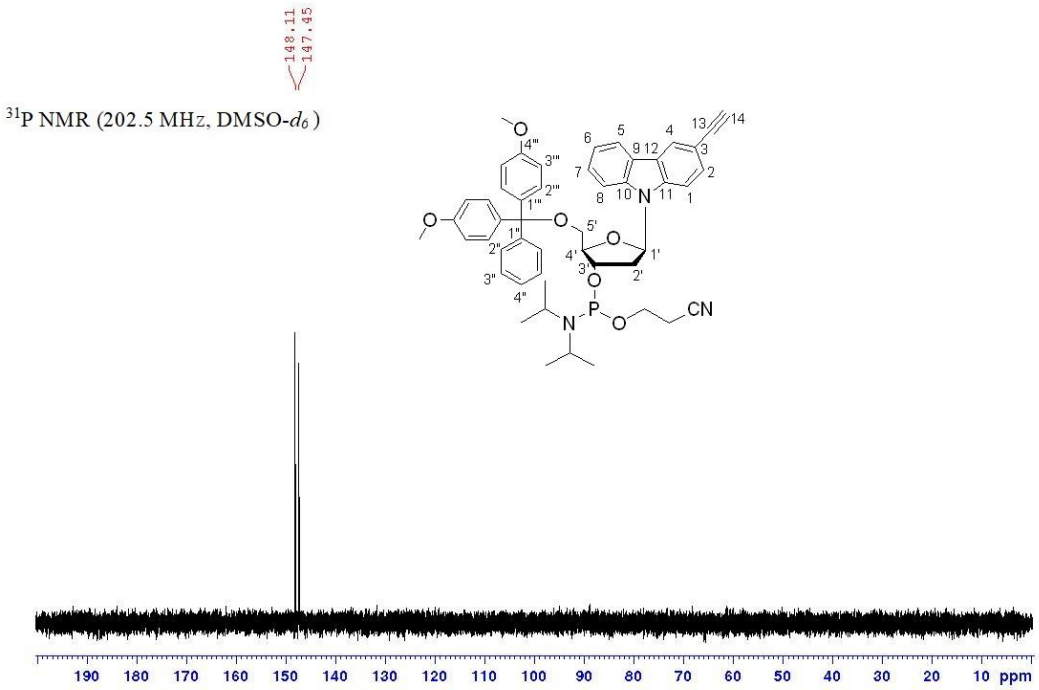


Figure S108. ^{31}P NMR spectrum of compound **24**.

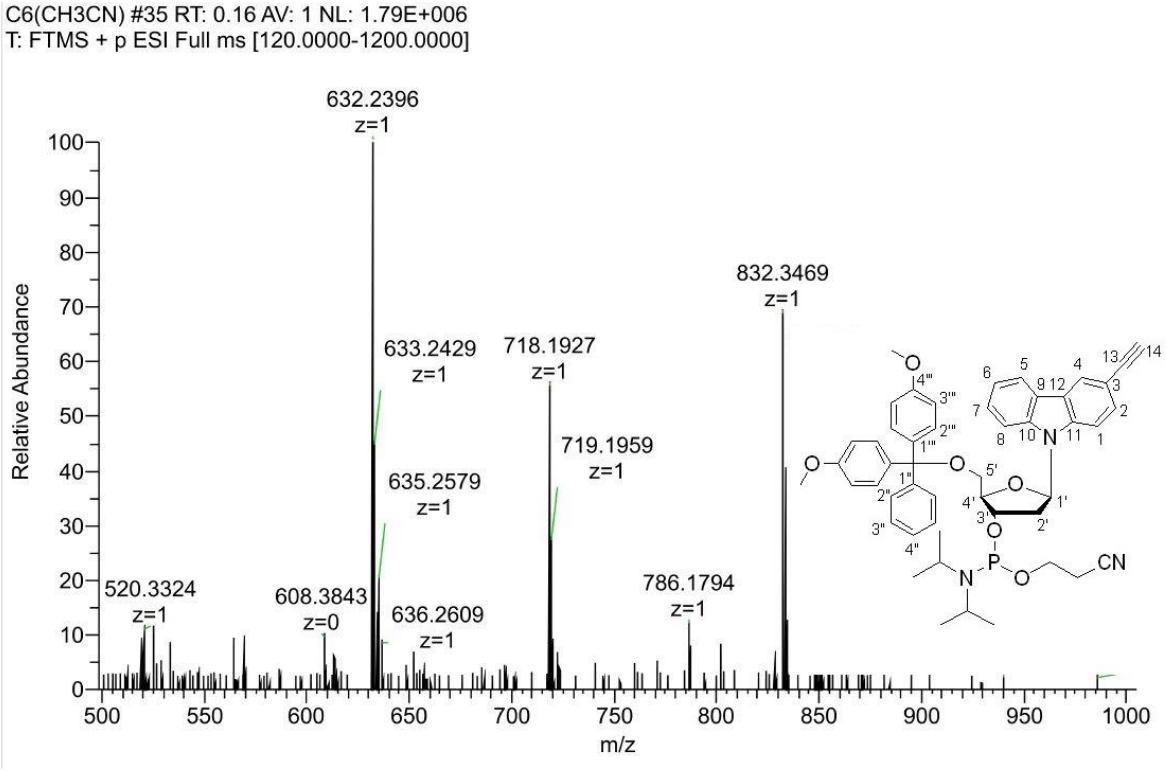


Figure S109. HRMS (ESI) of compound **24**.

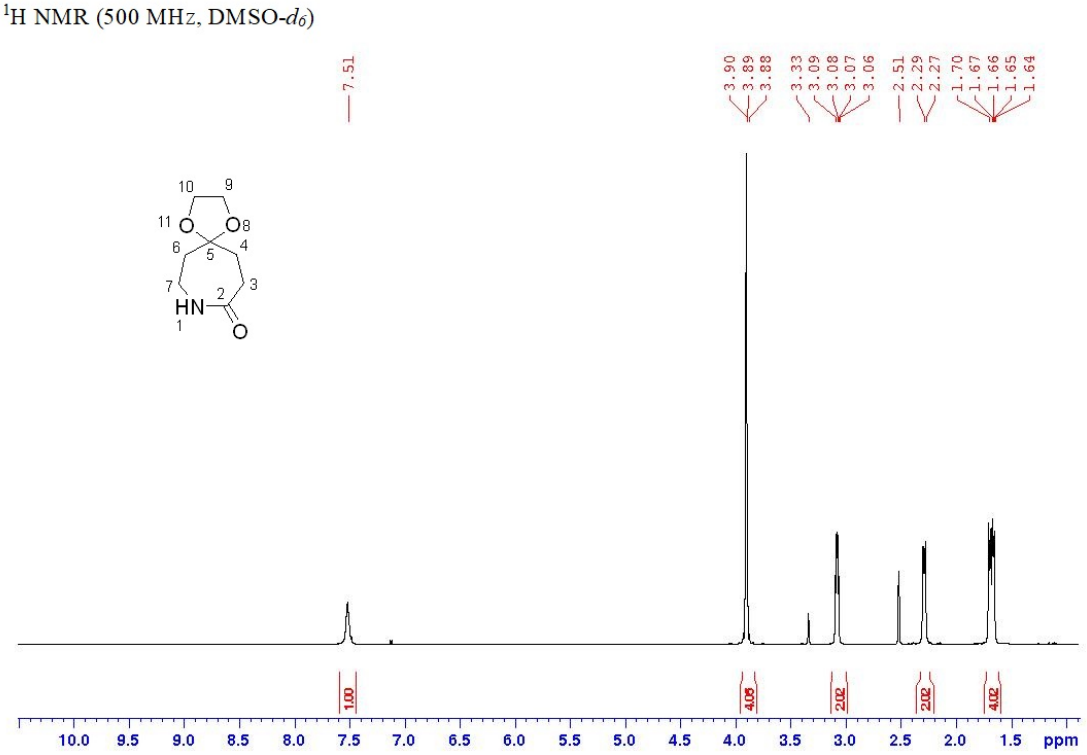


Figure S110. ^{1}H NMR spectrum of compound **28**.

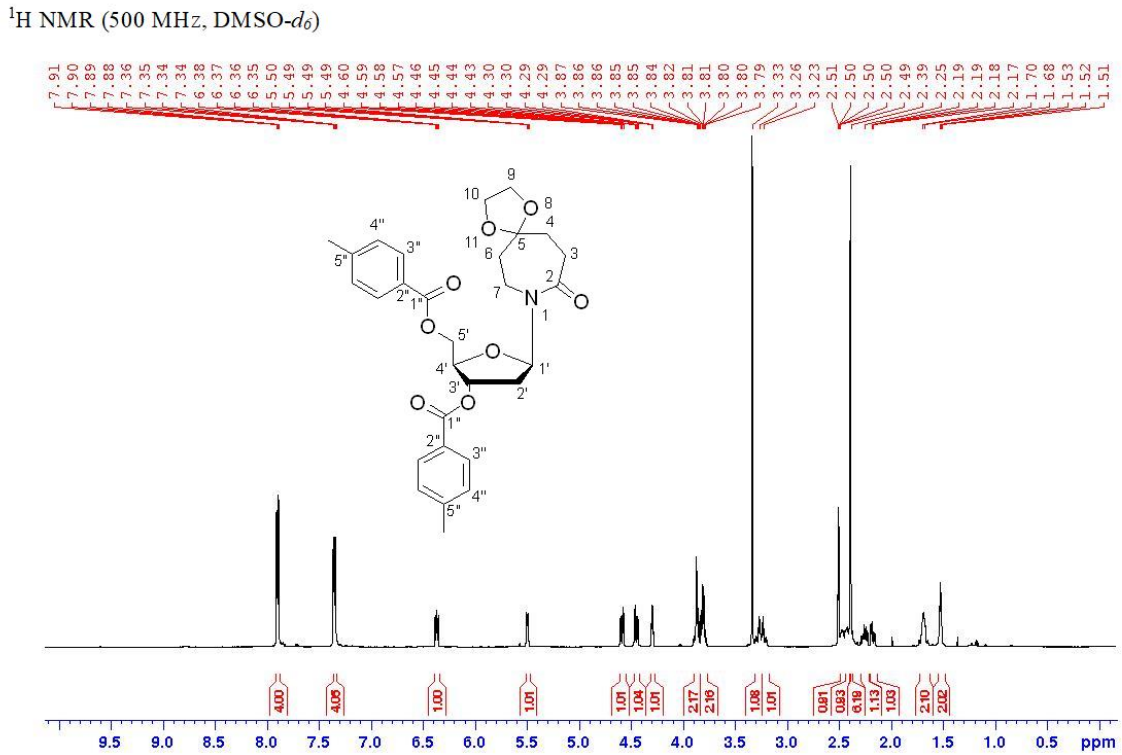


Figure S111. ^{1}H NMR spectrum of compound **30**.

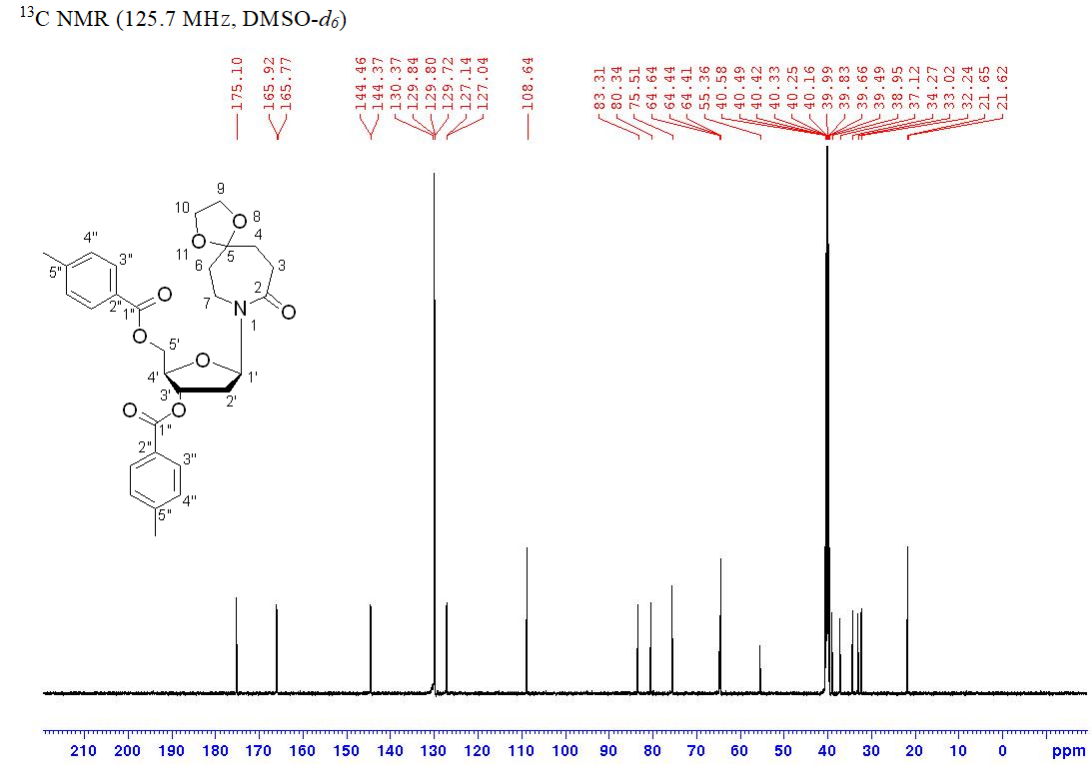


Figure S112. ^{13}C NMR spectrum of compound **30**.

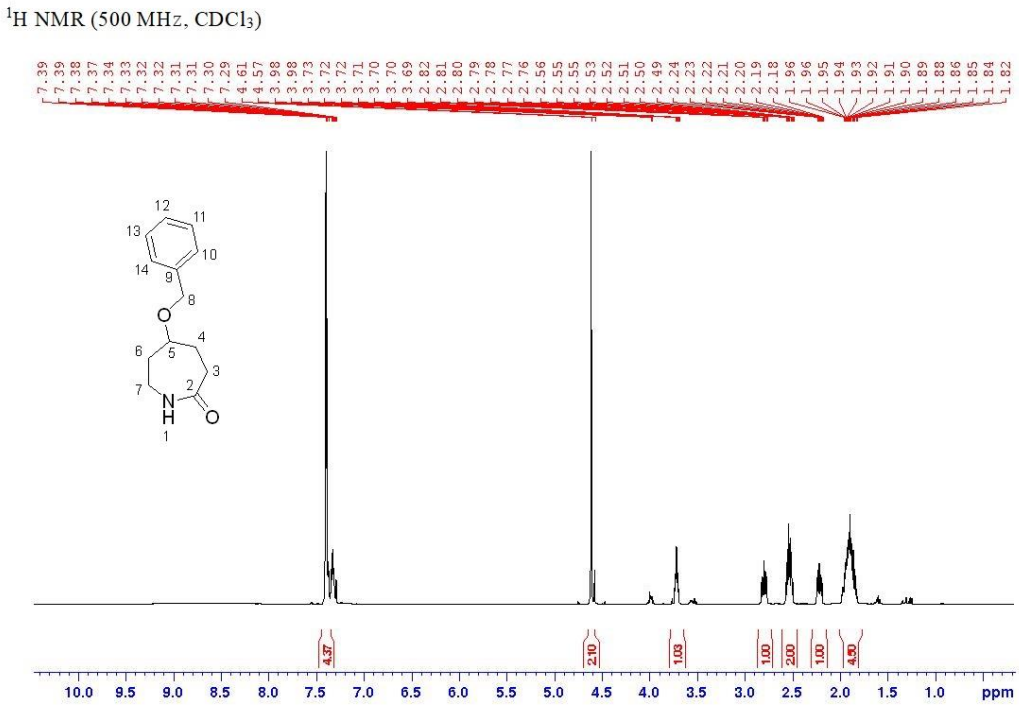


Figure S113. ^{1}H NMR spectrum of compound **33**.

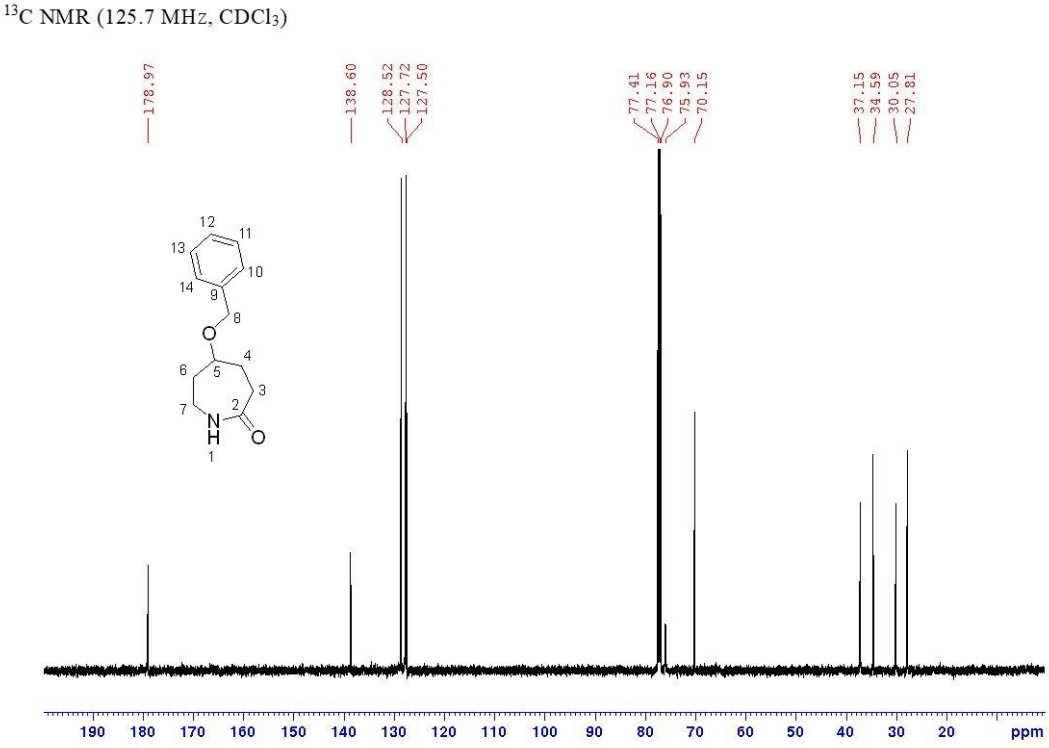


Figure S114. ^{13}C NMR spectrum of compound **33**.

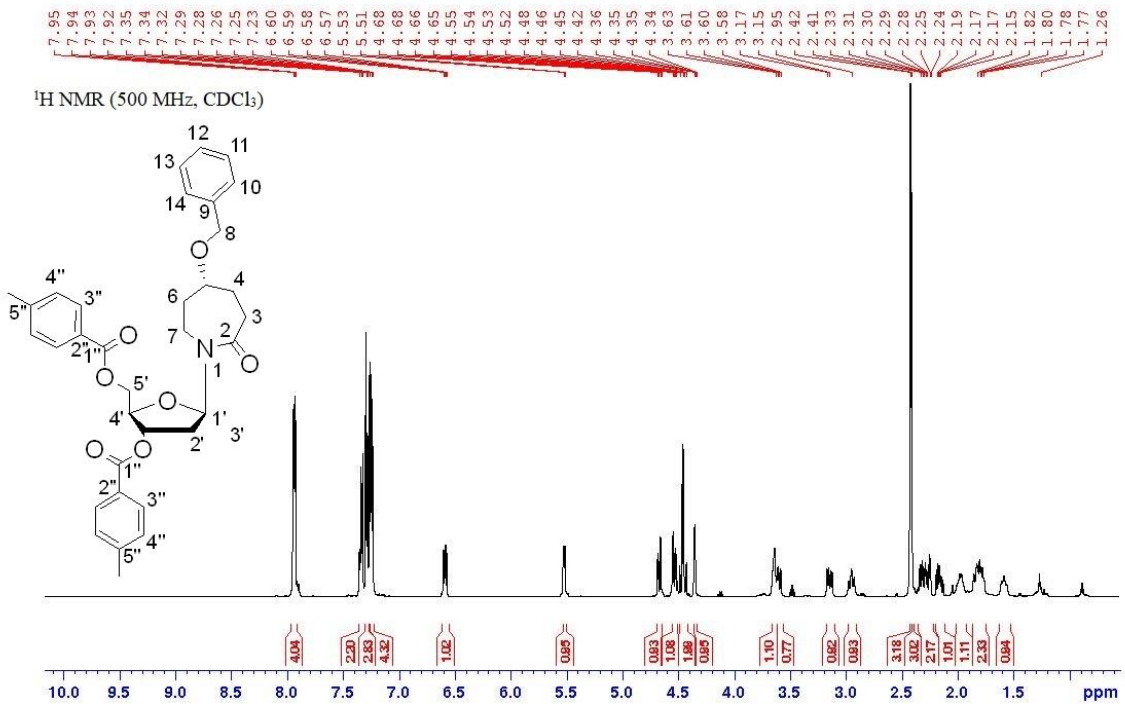


Figure S115. ^{1}H NMR spectrum of compound **34*R***.

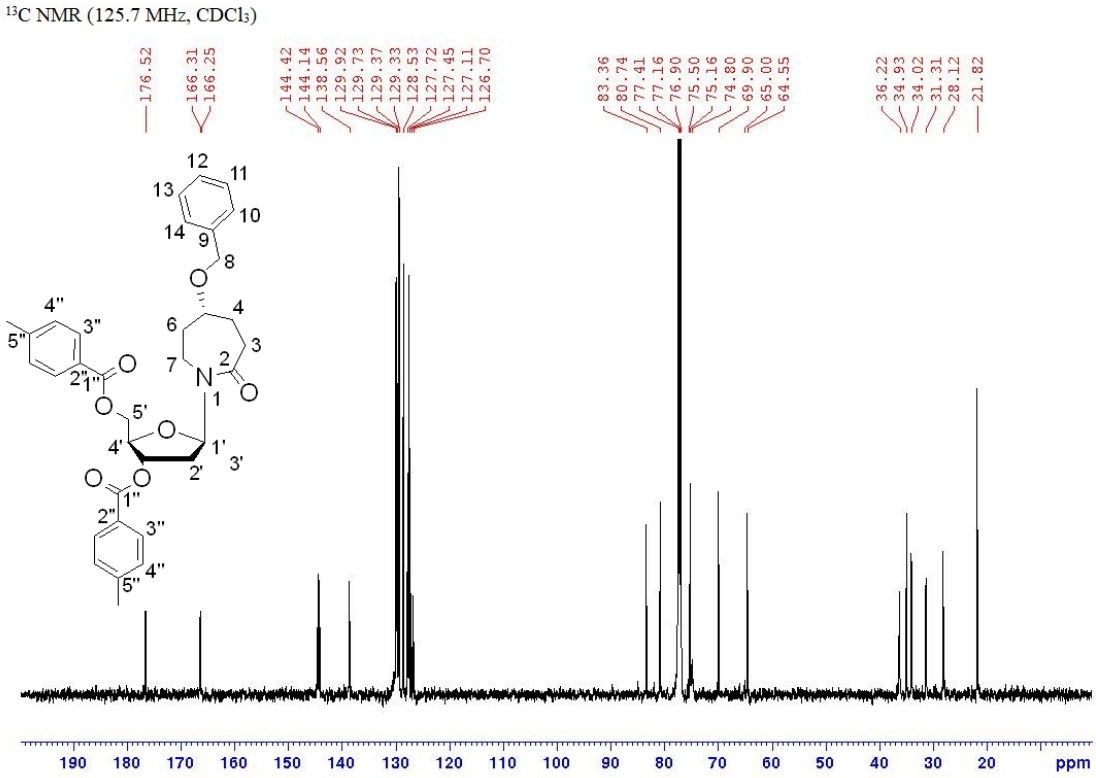


Figure S116. ^{13}C NMR spectrum of compound **34*R***.

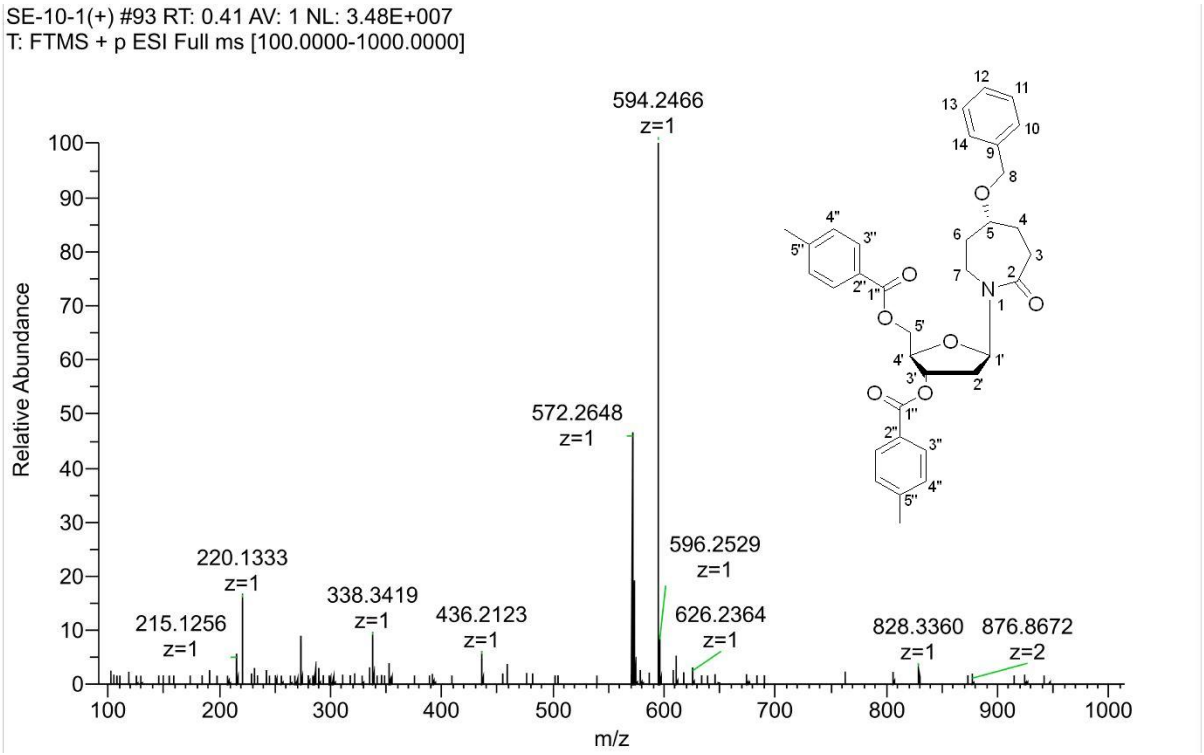


Figure S117. HRMS (ESI) of compound **34*R***.

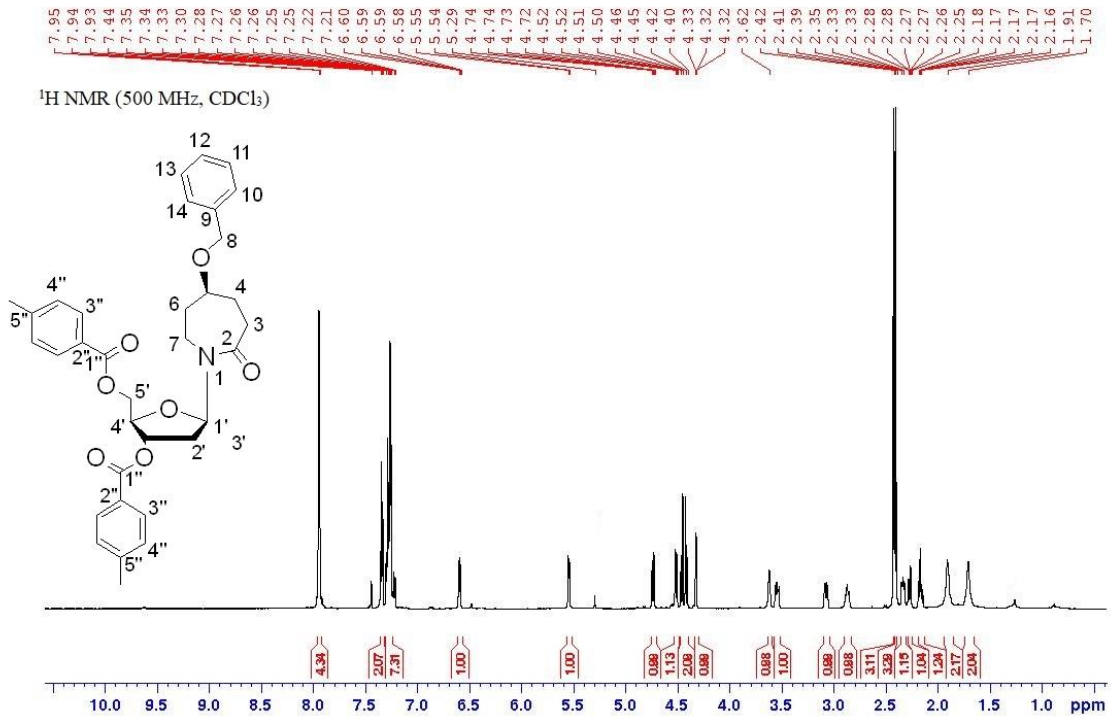


Figure S118. ^{1}H NMR spectrum of compound **34*S***.

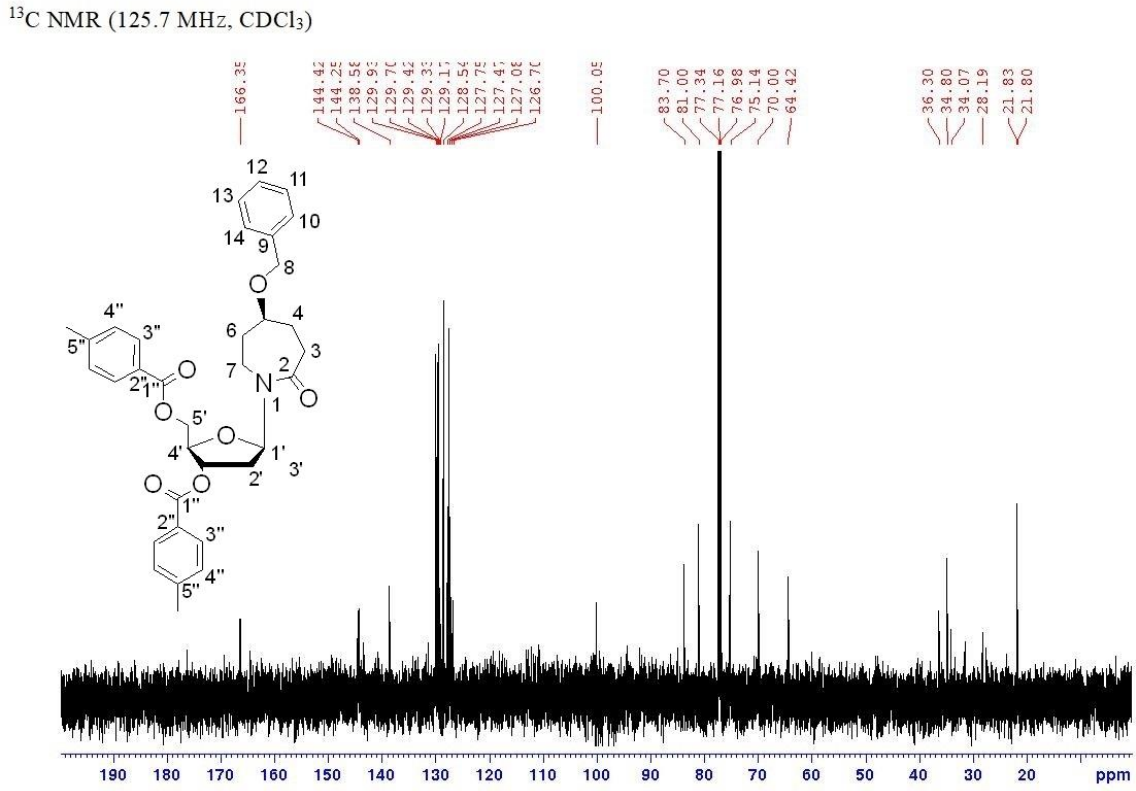


Figure S119. ^{13}C NMR spectrum of compound **34*S***.

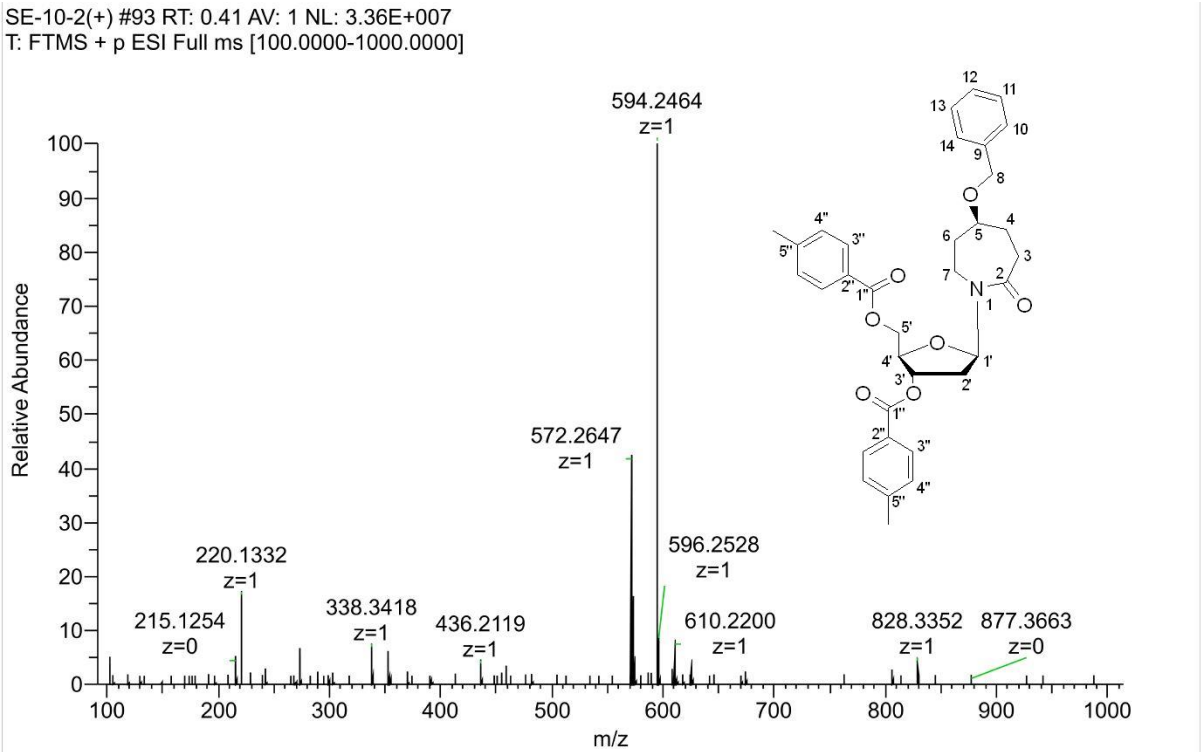


Figure S120. HRMS (ESI) of compound **34*S***.

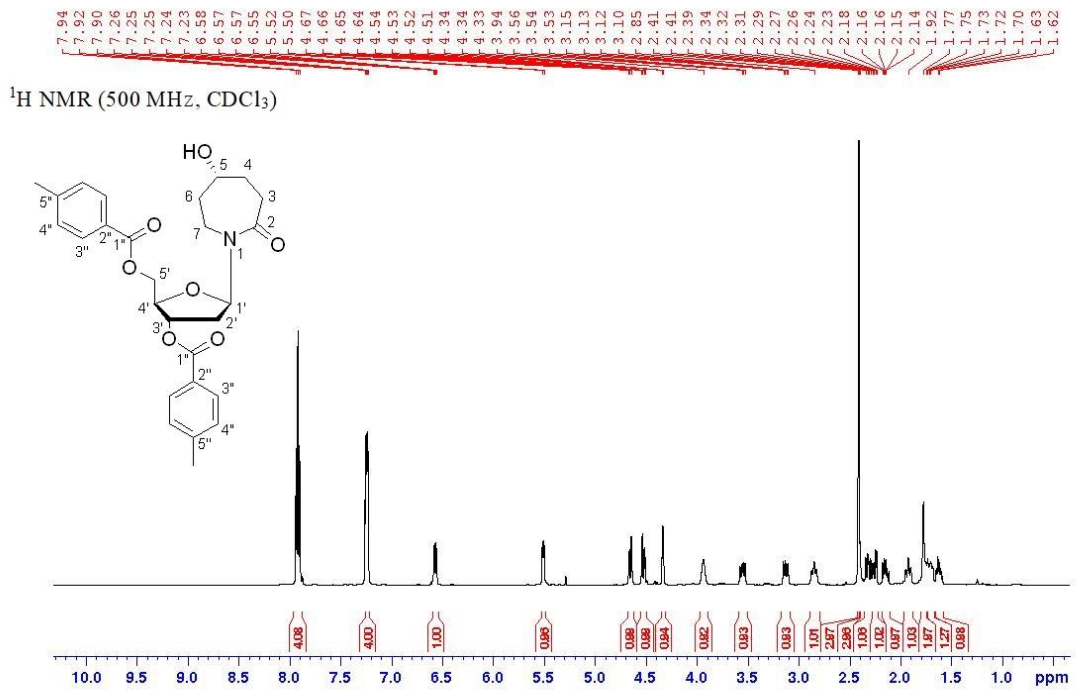


Figure S121. ^{1}H NMR spectrum of compound **35*R***.

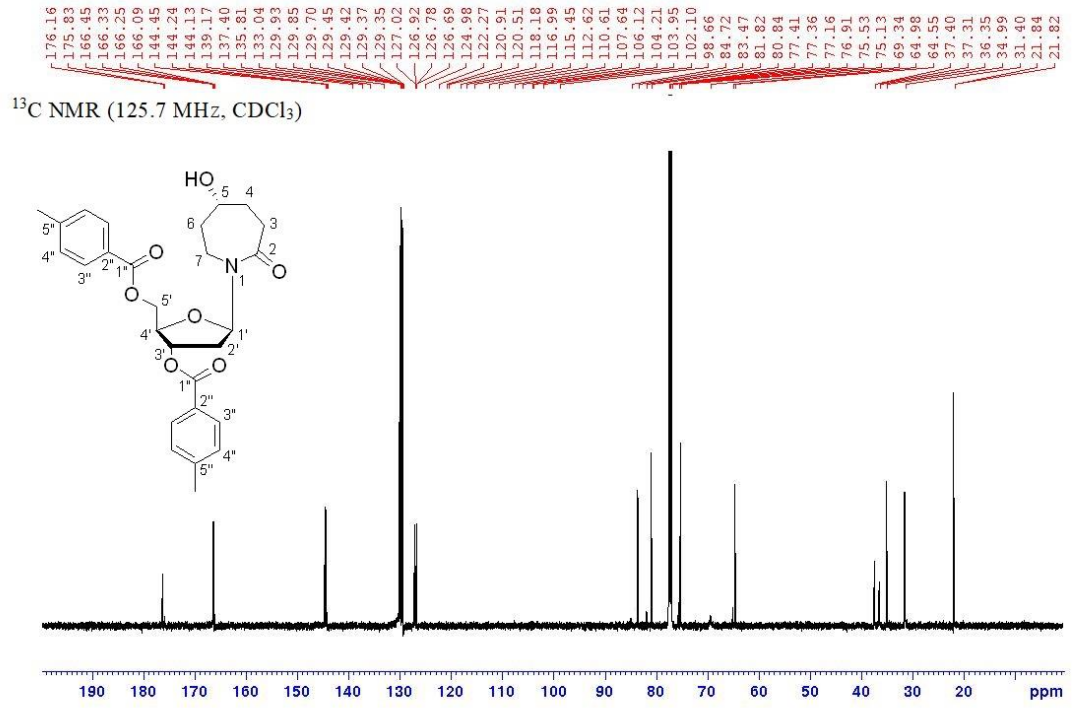


Figure S122. ^{13}C NMR spectrum of compound **35*R***.

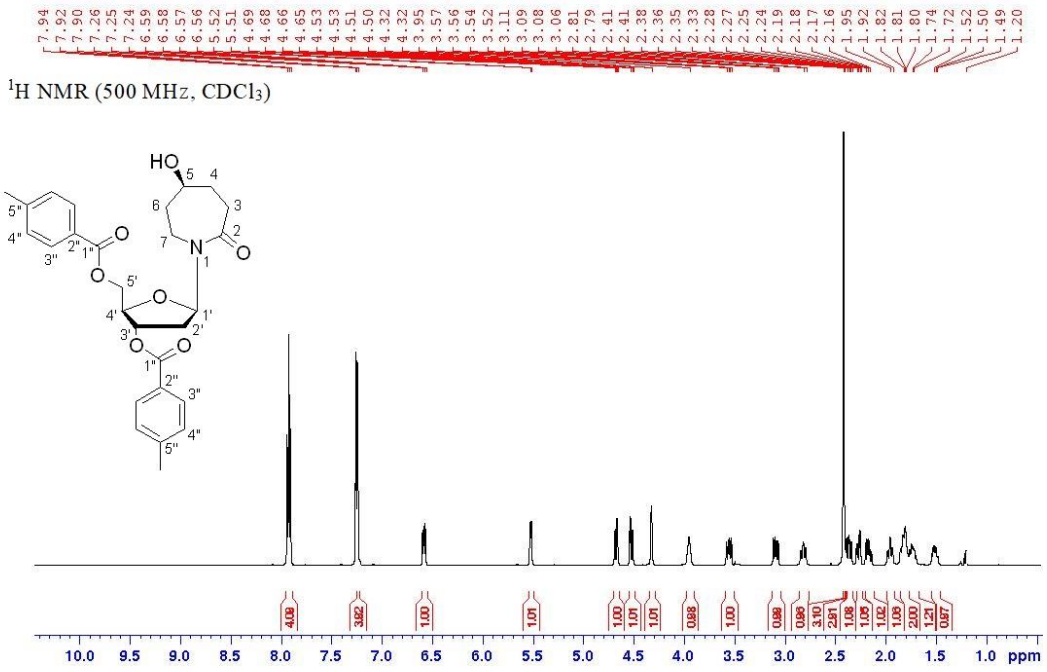


Figure S123. ^{1}H NMR spectrum of compound **35*S***.

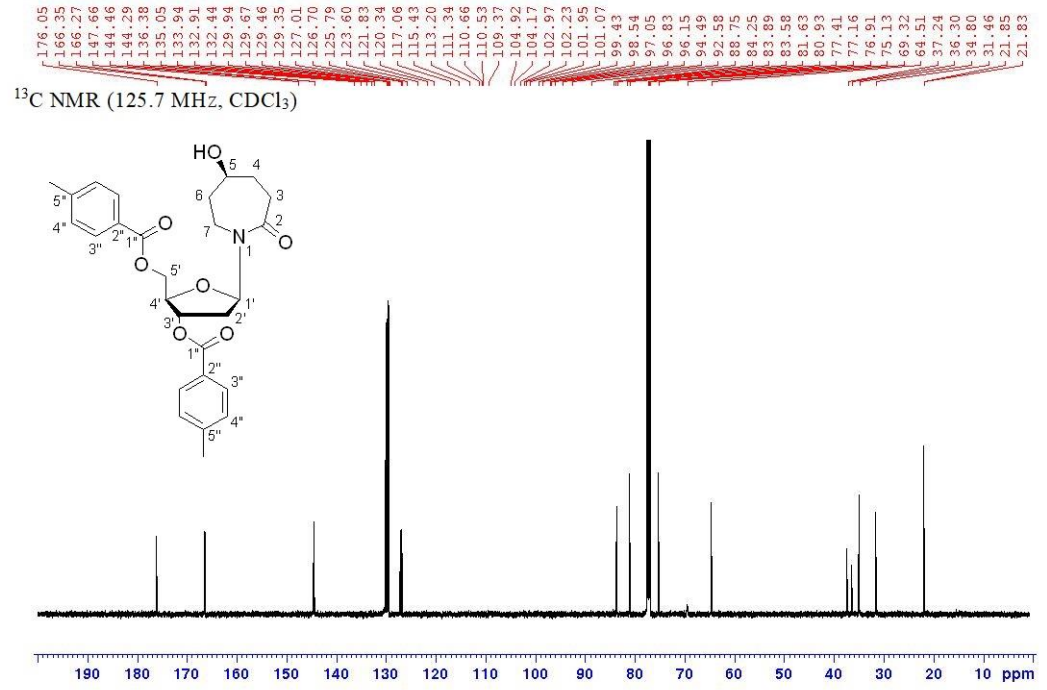


Figure S124. ^{13}C NMR spectrum of compound **35*S***.

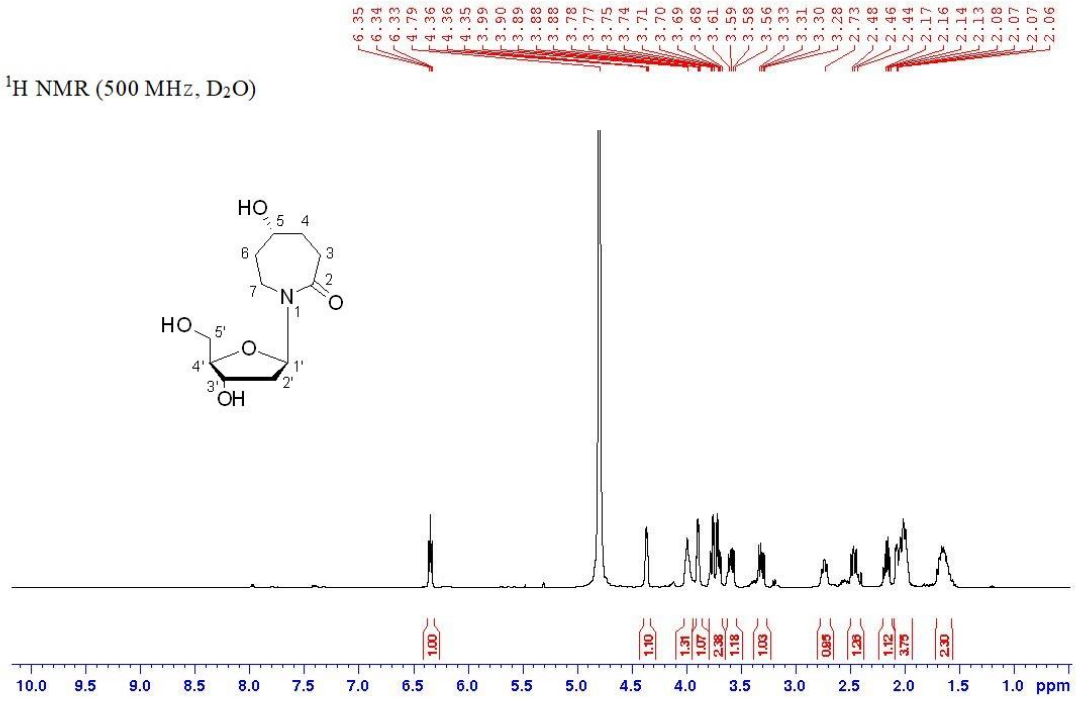


Figure S125. ^{1}H NMR spectrum of compound **36*R***.

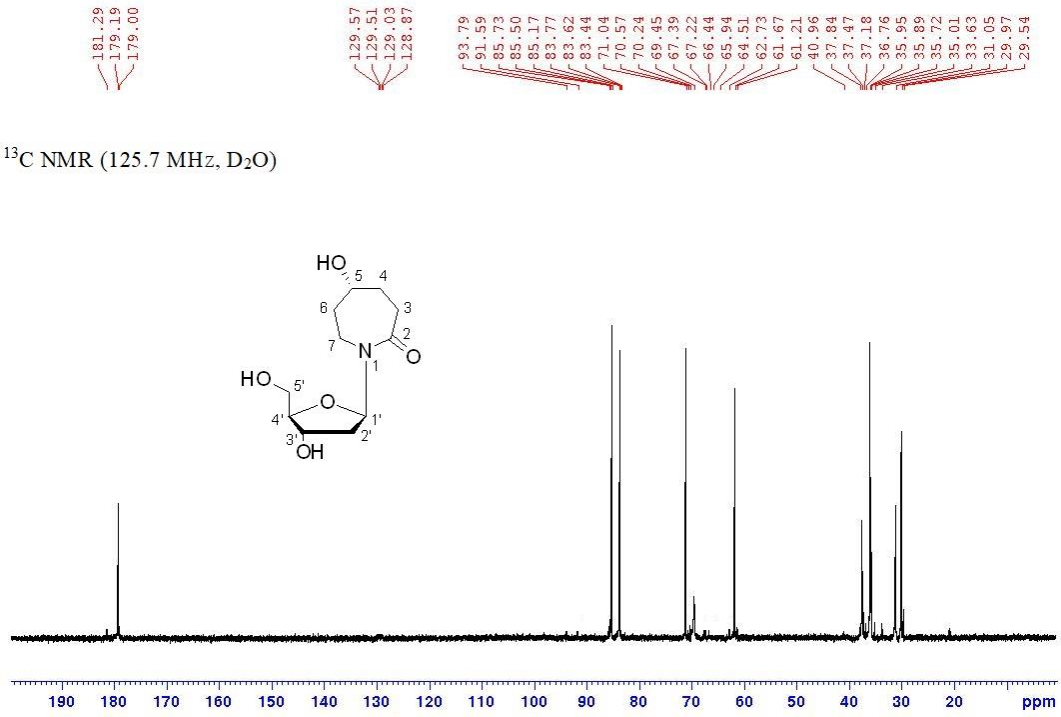


Figure S126. ^{13}C NMR spectrum of compound **36*R***.

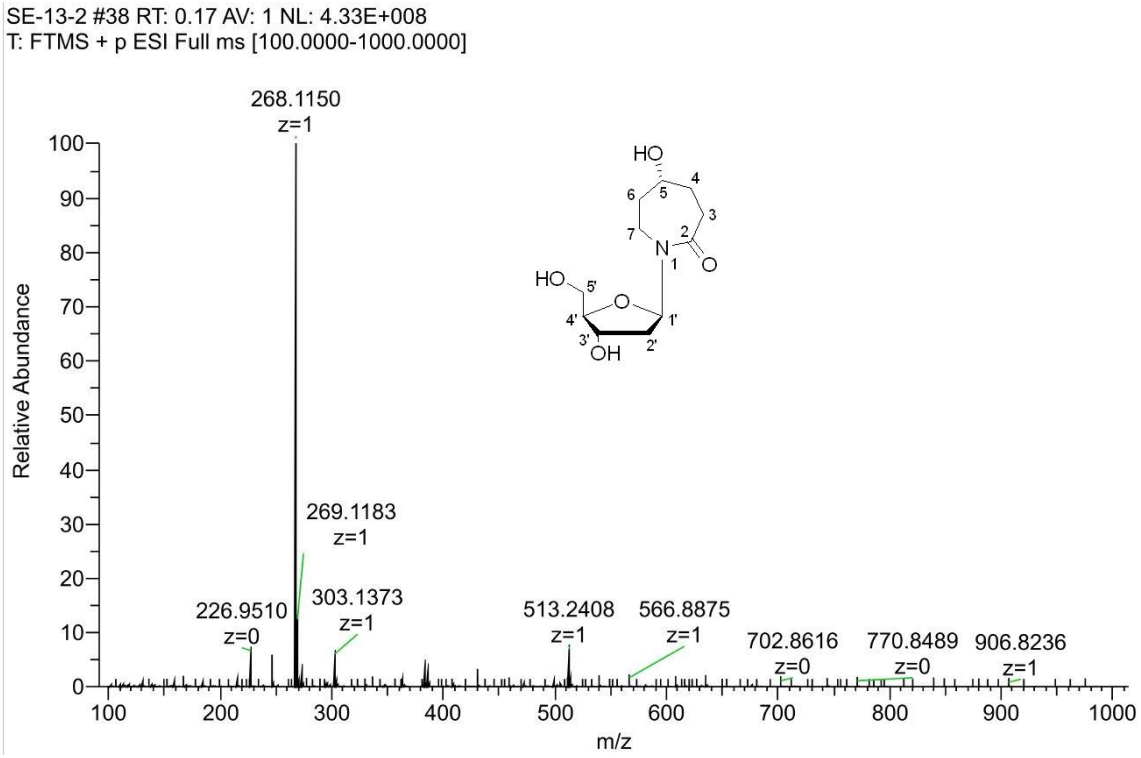


Figure S127. HRMS (ESI) of compound **36*R*.**

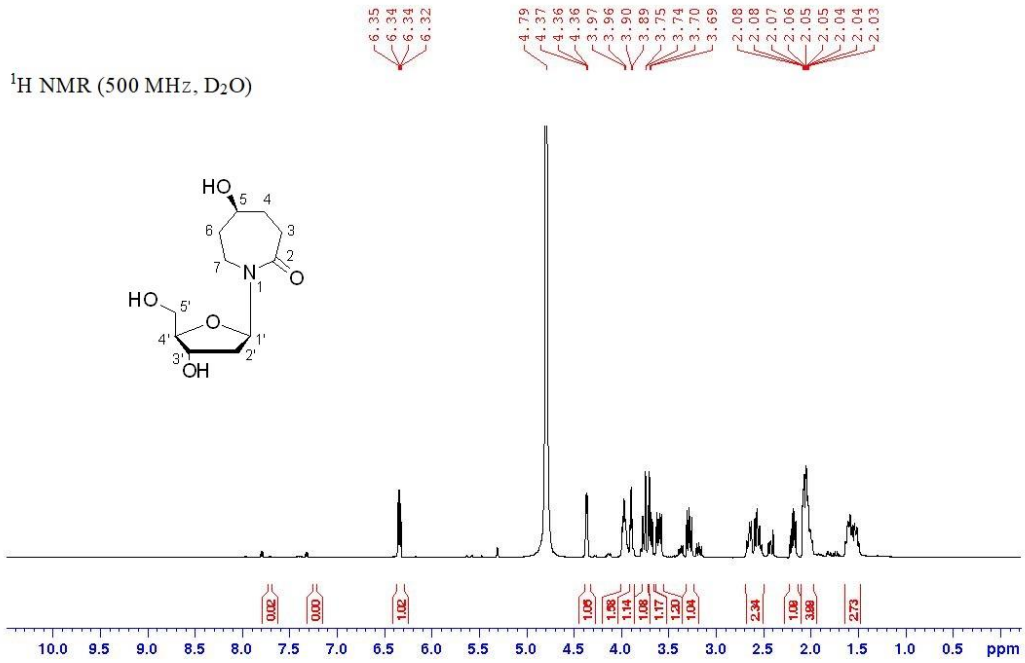


Figure S128. ^{1}H NMR spectrum of compound **36*S*.**

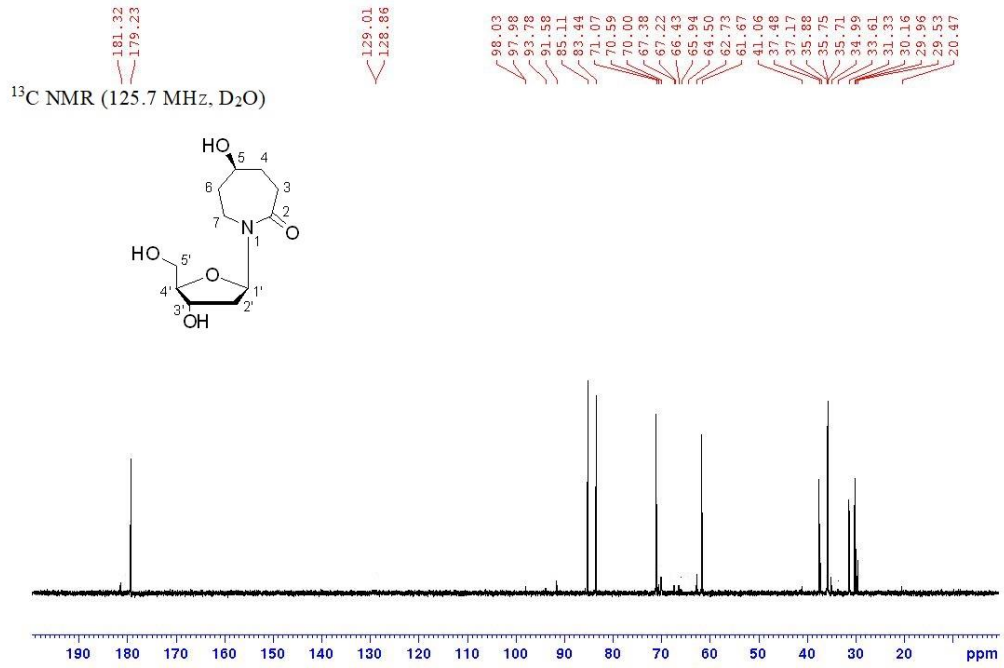


Figure S129. ^{13}C NMR spectrum of compound **36*S***.

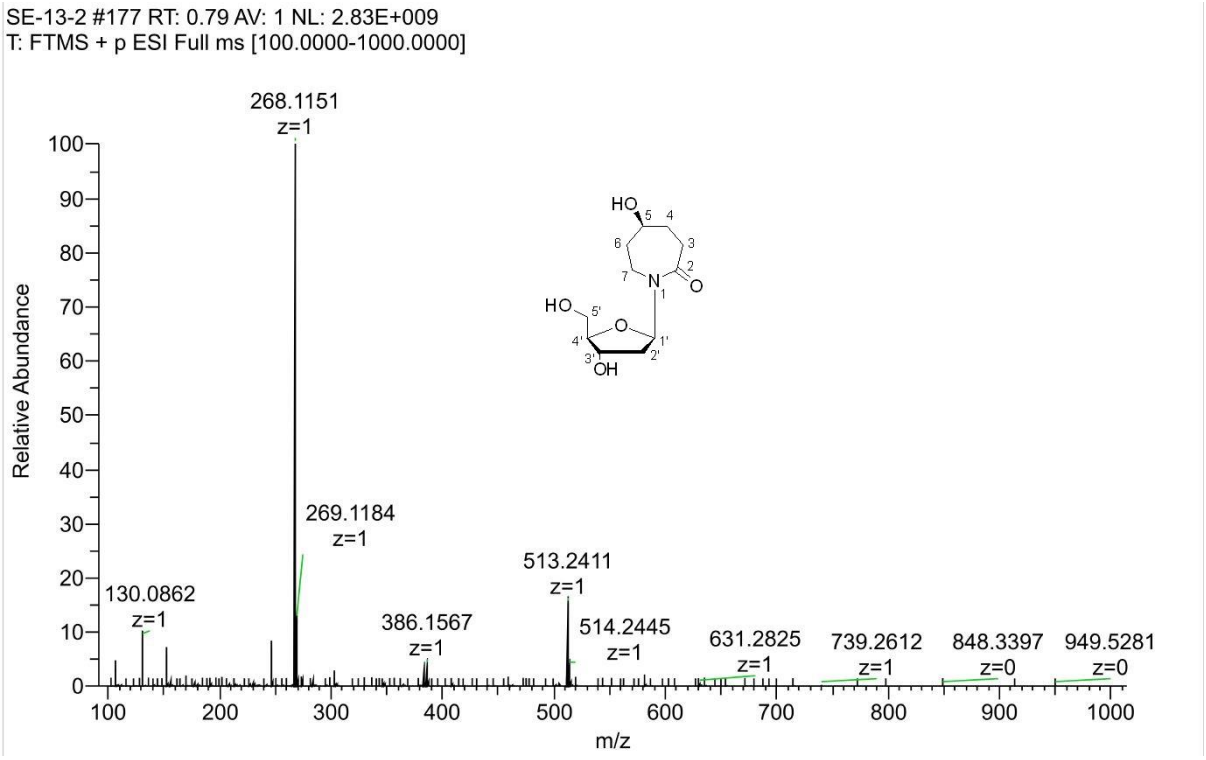


Figure S130. HRMS (ESI) of compound **36*S***.

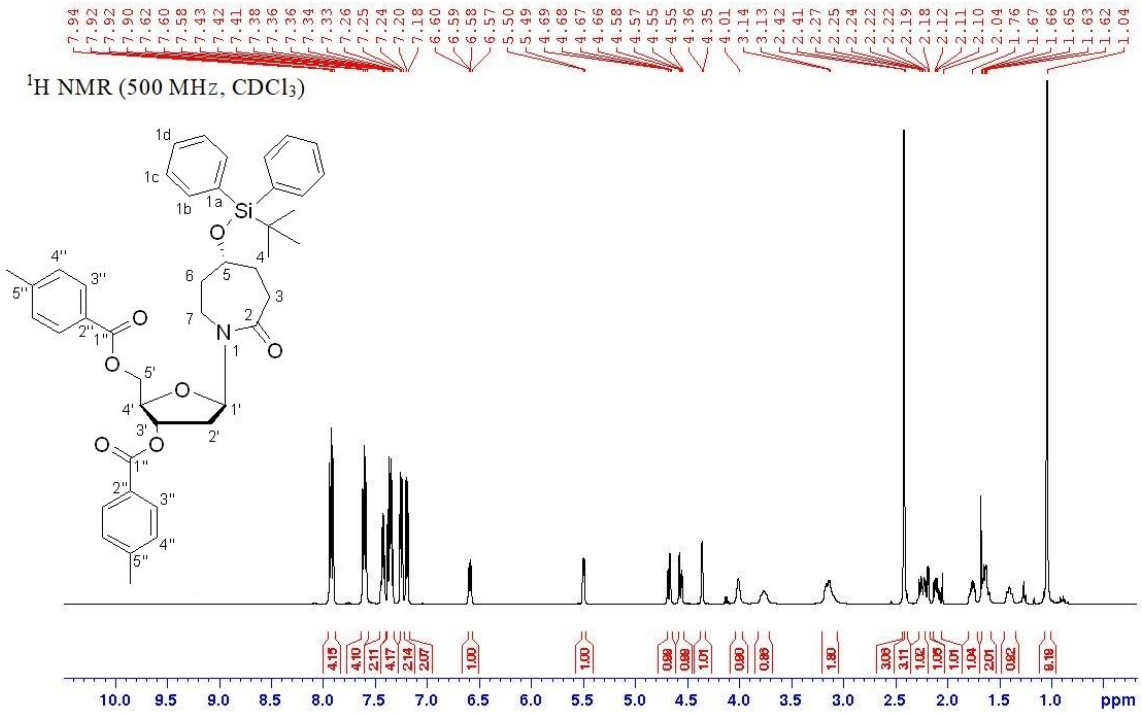


Figure S131. ^{1}H NMR spectrum of compound **37*R***.

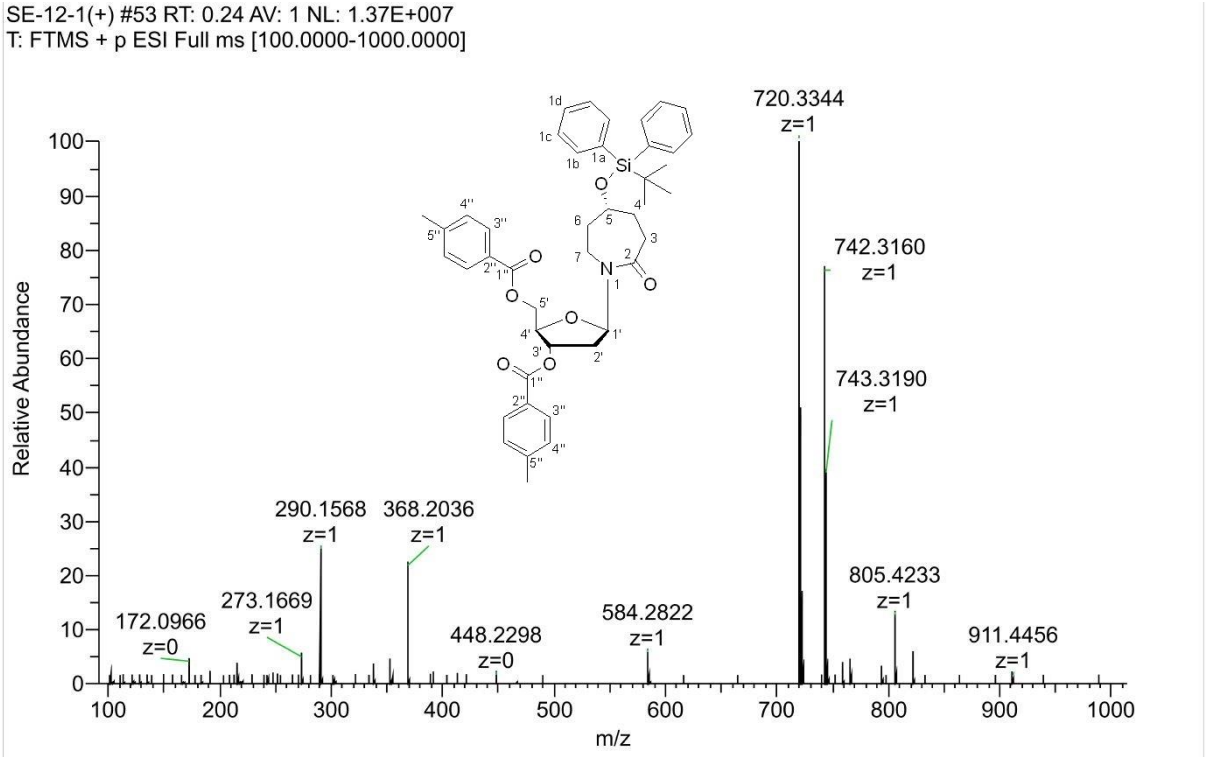


Figure S132. HRMS (ESI) of compound **37*R***.

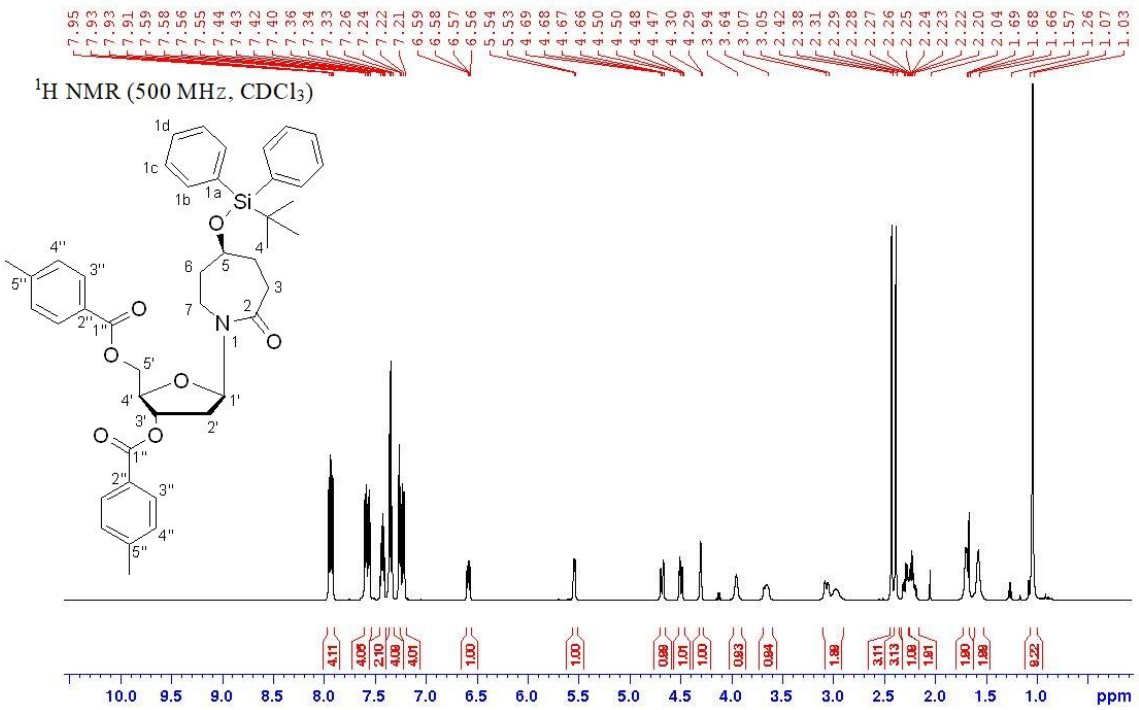


Figure S133. ^{1}H NMR spectrum of compound **37*S***.

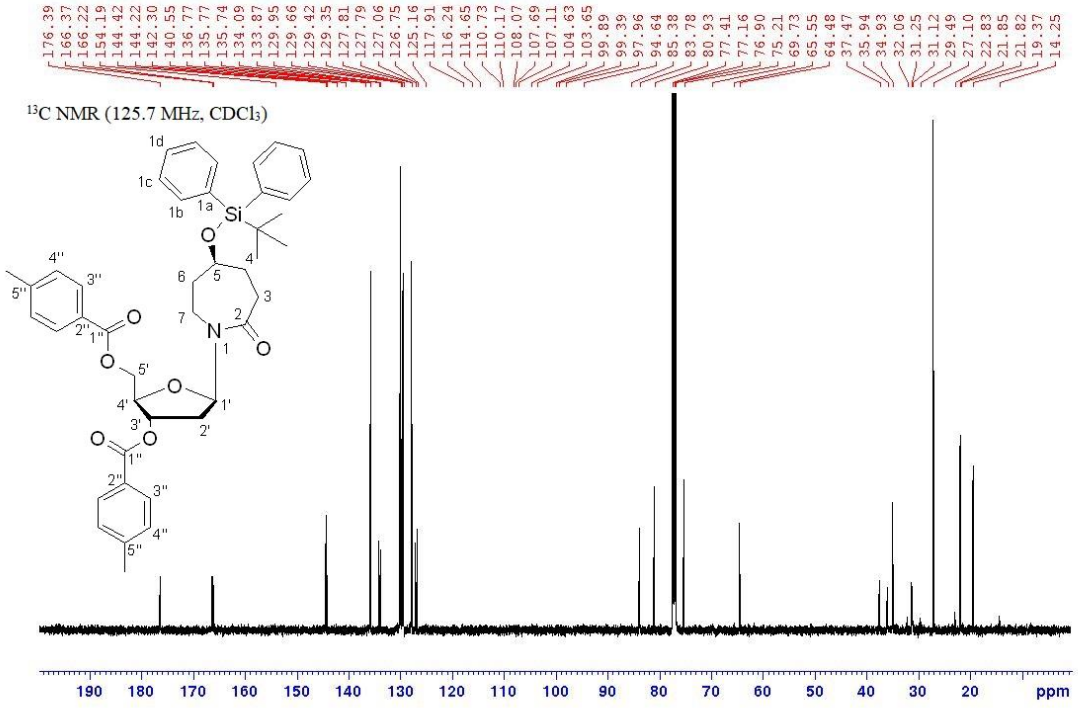


Figure S134. ^{13}C NMR spectrum of compound **37*S***.

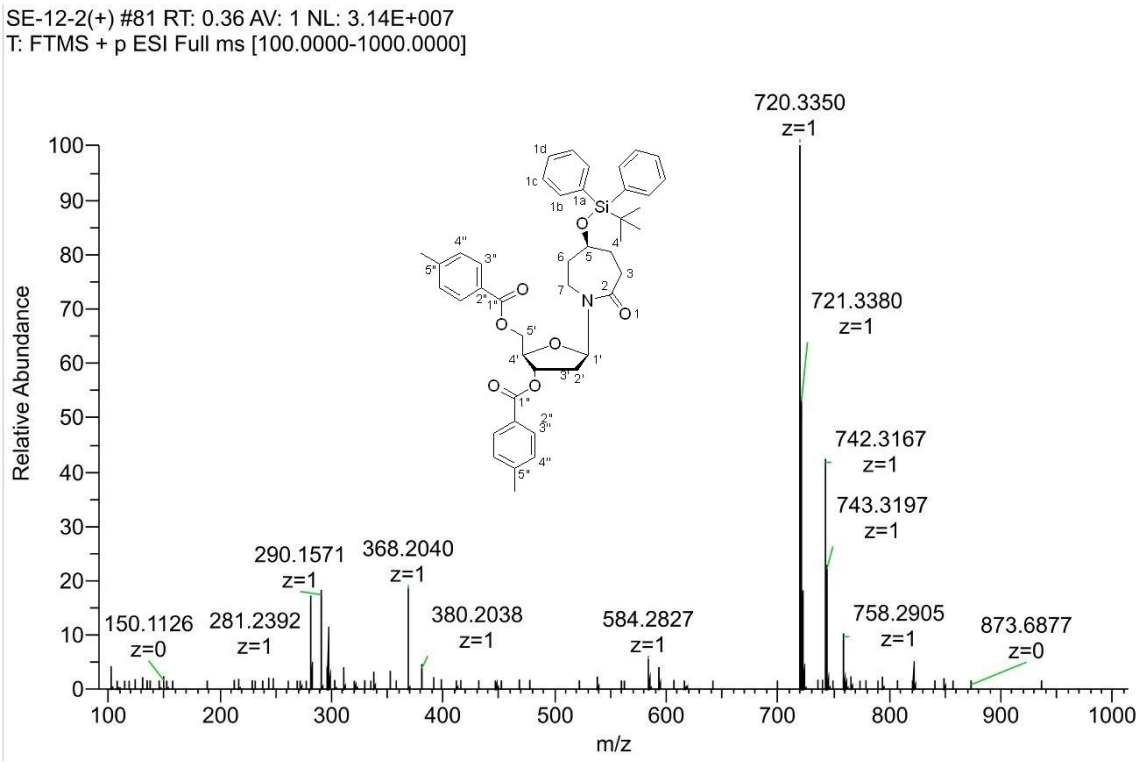


Figure S135. HRMS (ESI) of compound **37*S***.

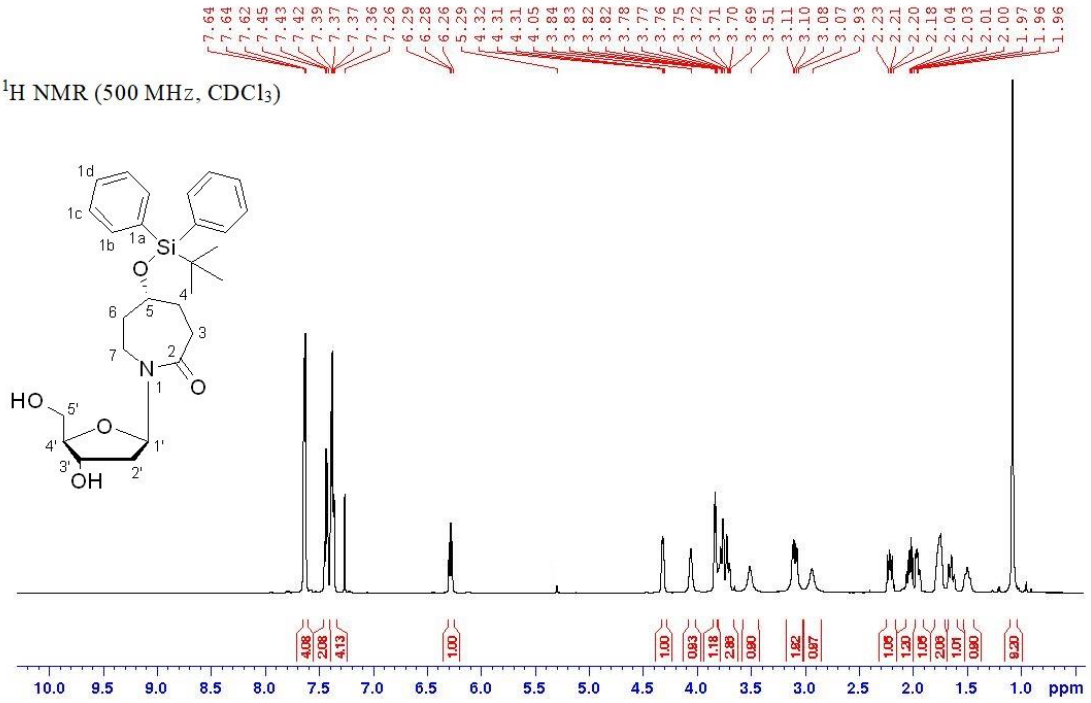


Figure S136. ^{1}H NMR spectrum of compound **38*R***.

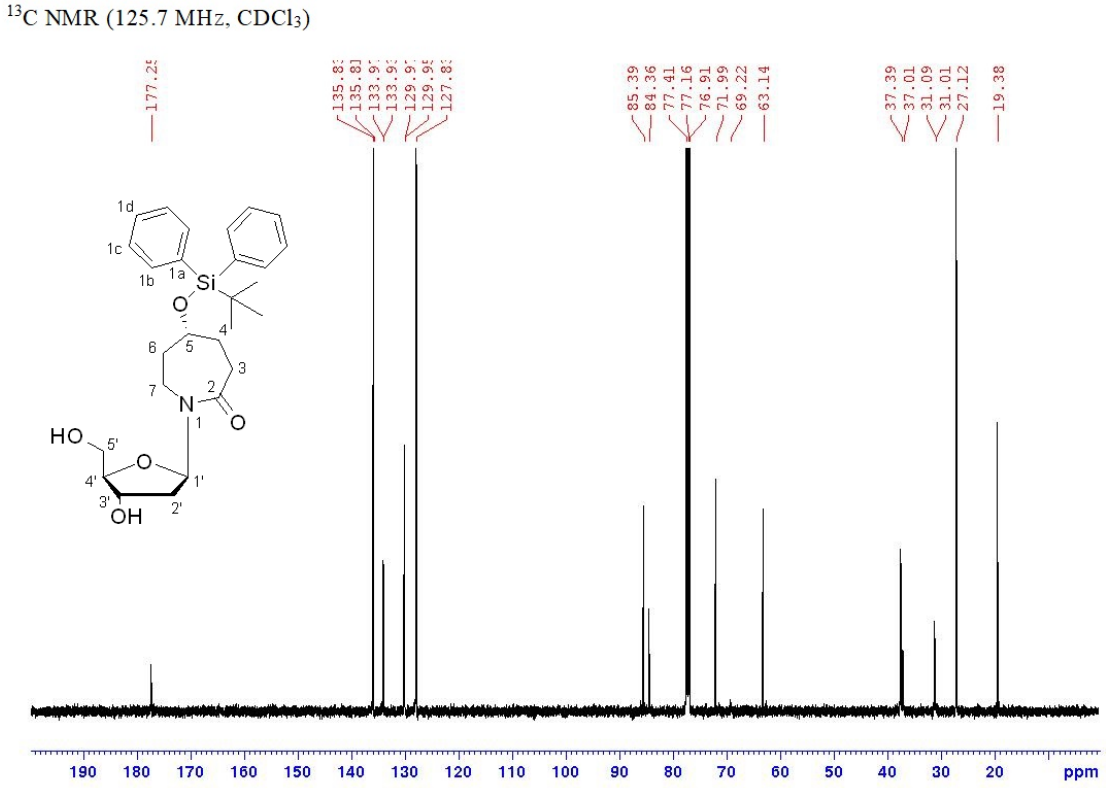


Figure S137. ^{13}C NMR spectrum of compound **38*R***.

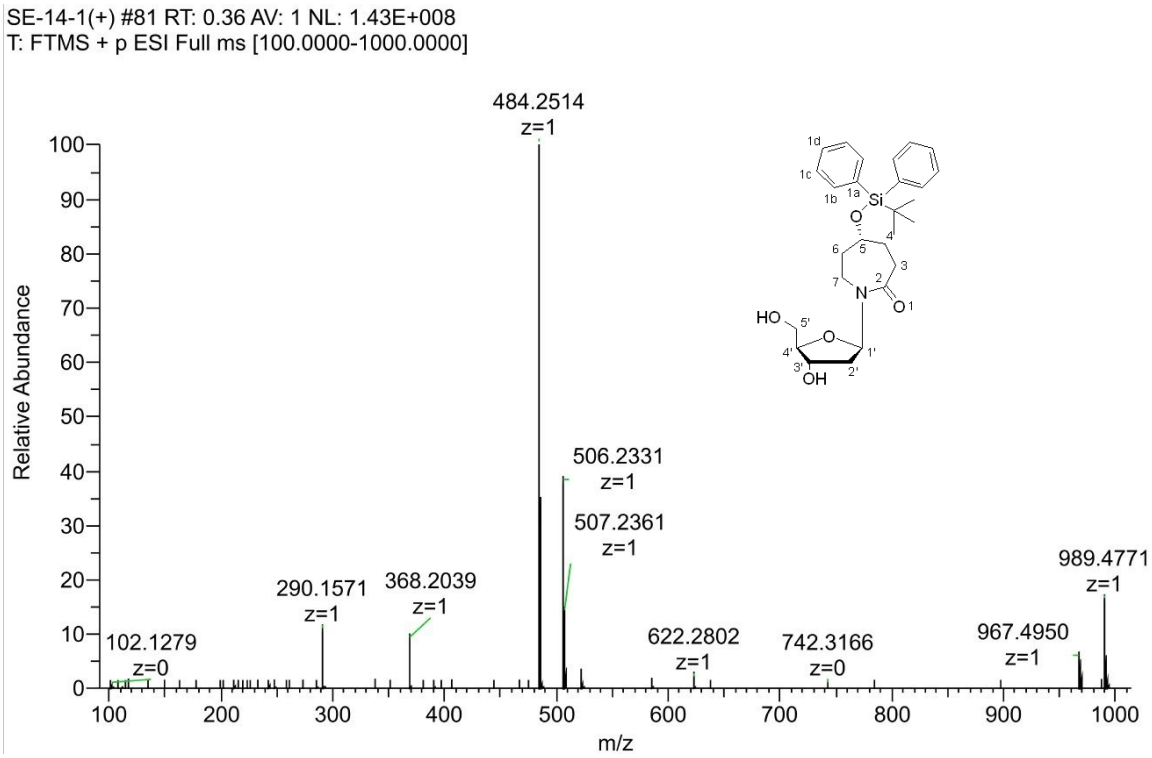


Figure S138. HRMS (ESI) of compound **38*R***.

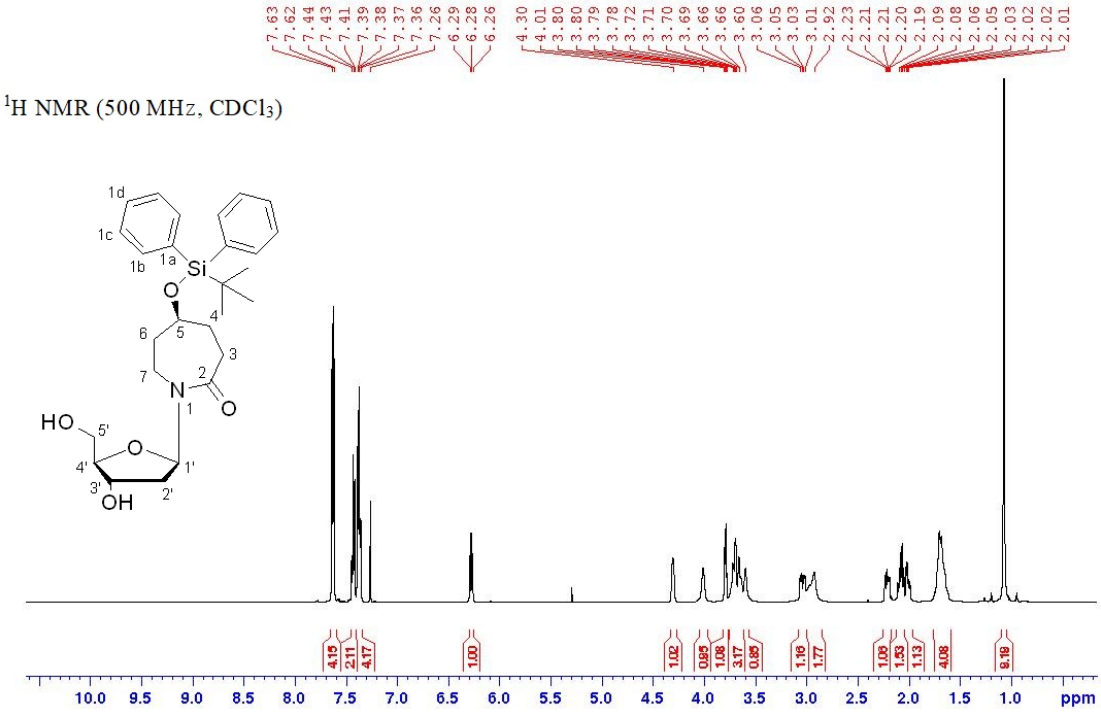


Figure S139. ^{1}H NMR spectrum of compound **38*S***.

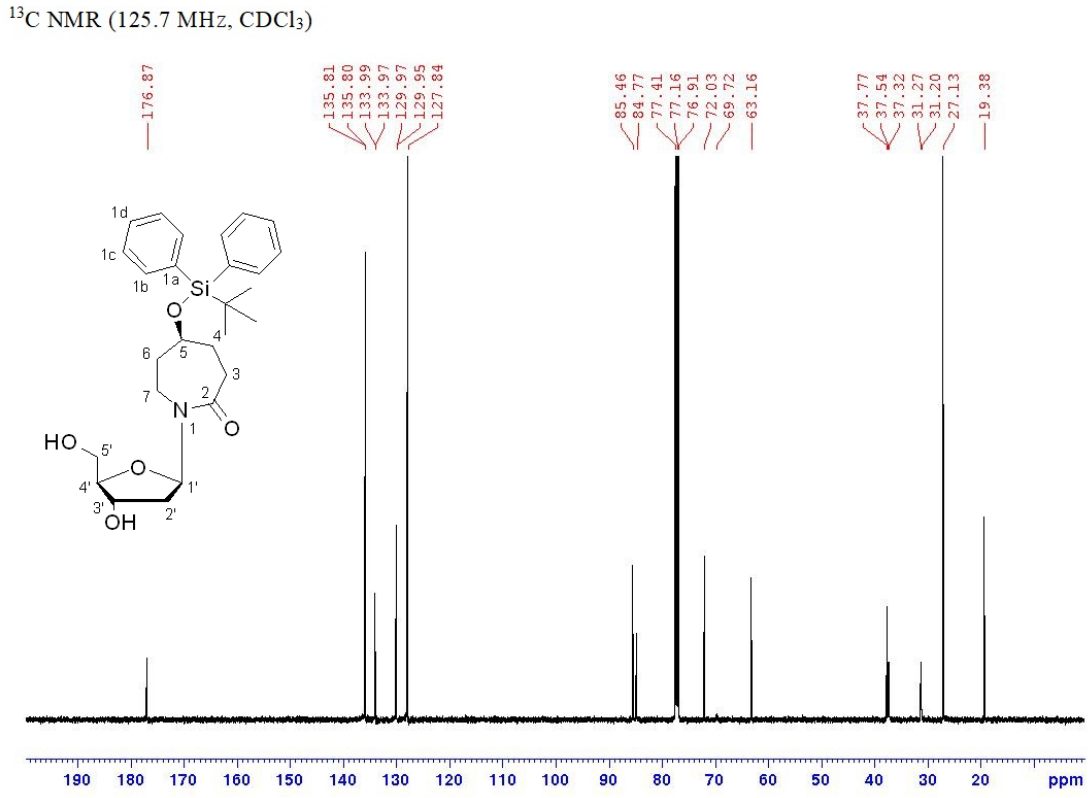


Figure S140. ^{13}C NMR spectrum of compound **38*S***.

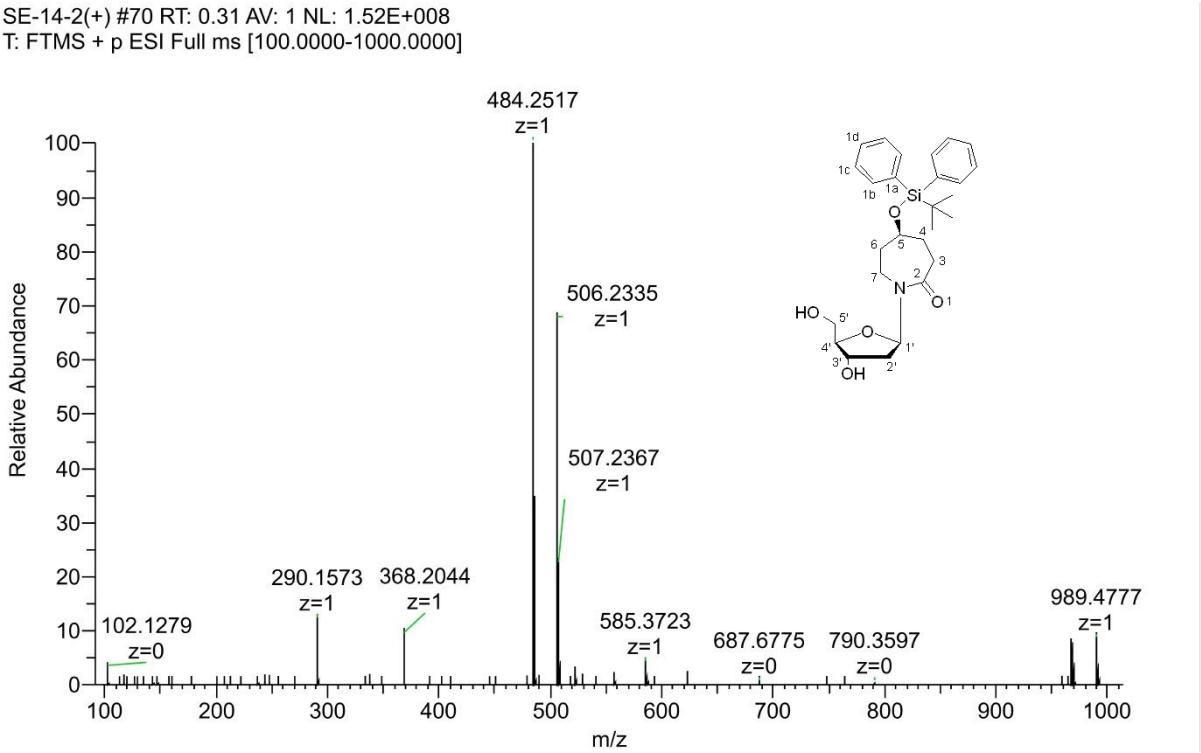


Figure S141. HRMS (ESI) of compound **38*S***.

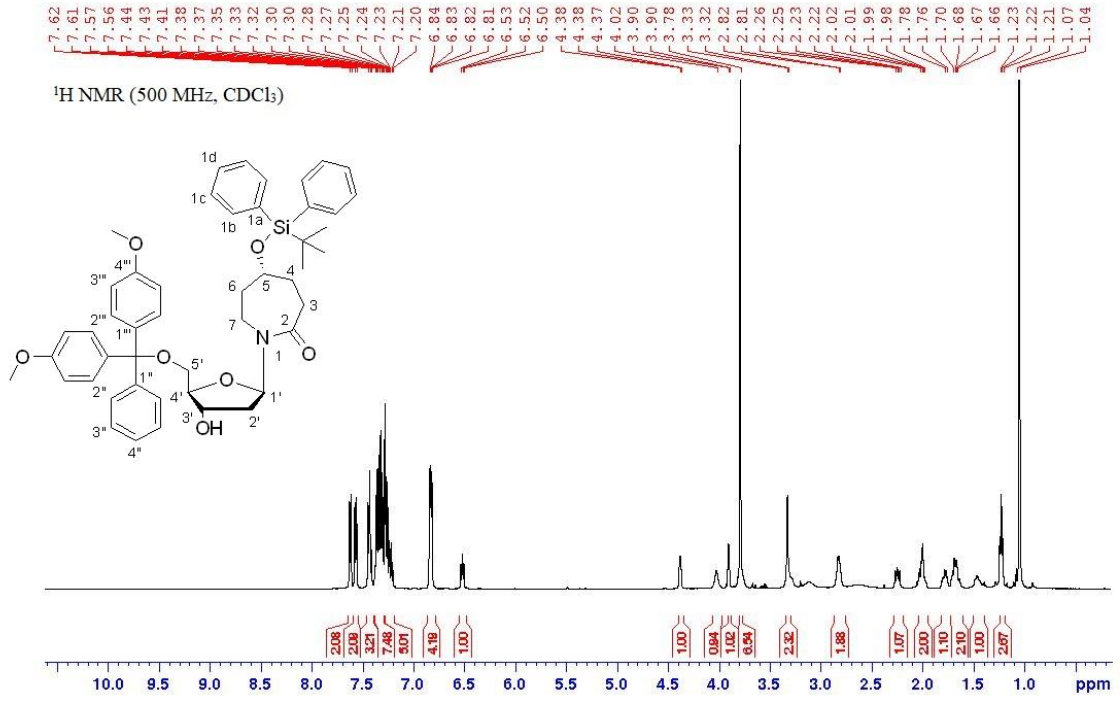


Figure S142. ^{1}H NMR spectrum of compound **39*R***.

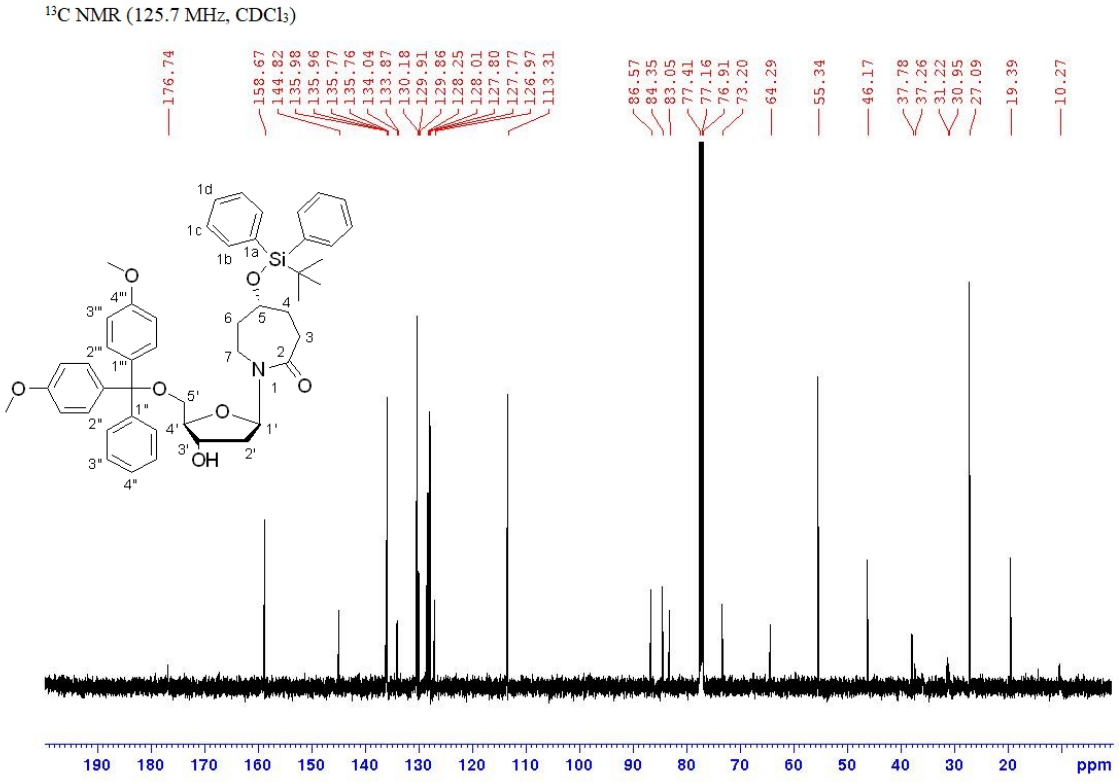


Figure S143. ^{13}C NMR spectrum of **compound 39*R***.

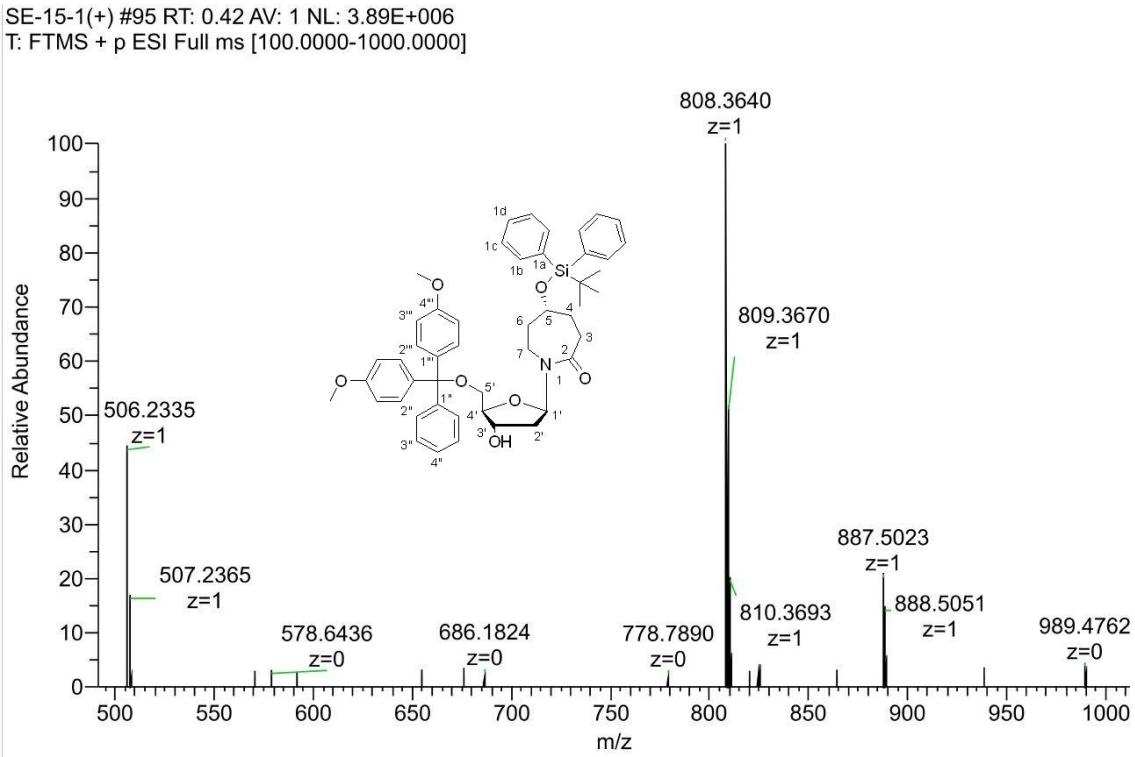


Figure S144. HRMS (ESI) of compound **39*R***.

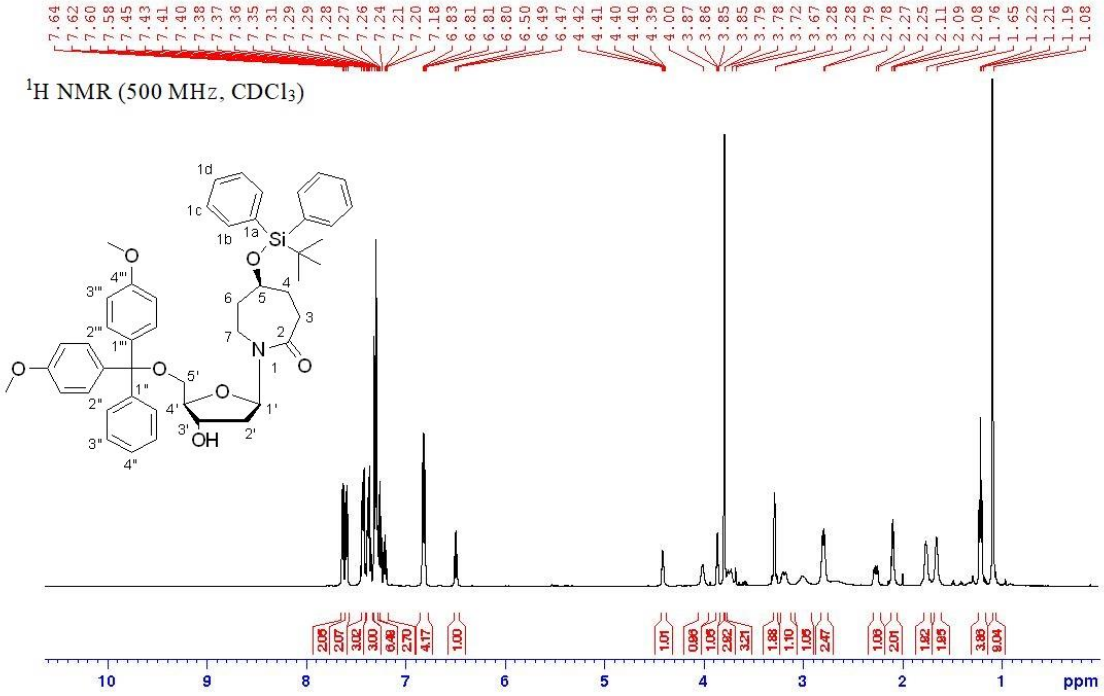


Figure S145. ^{1}H NMR spectrum of compound **39*S***.

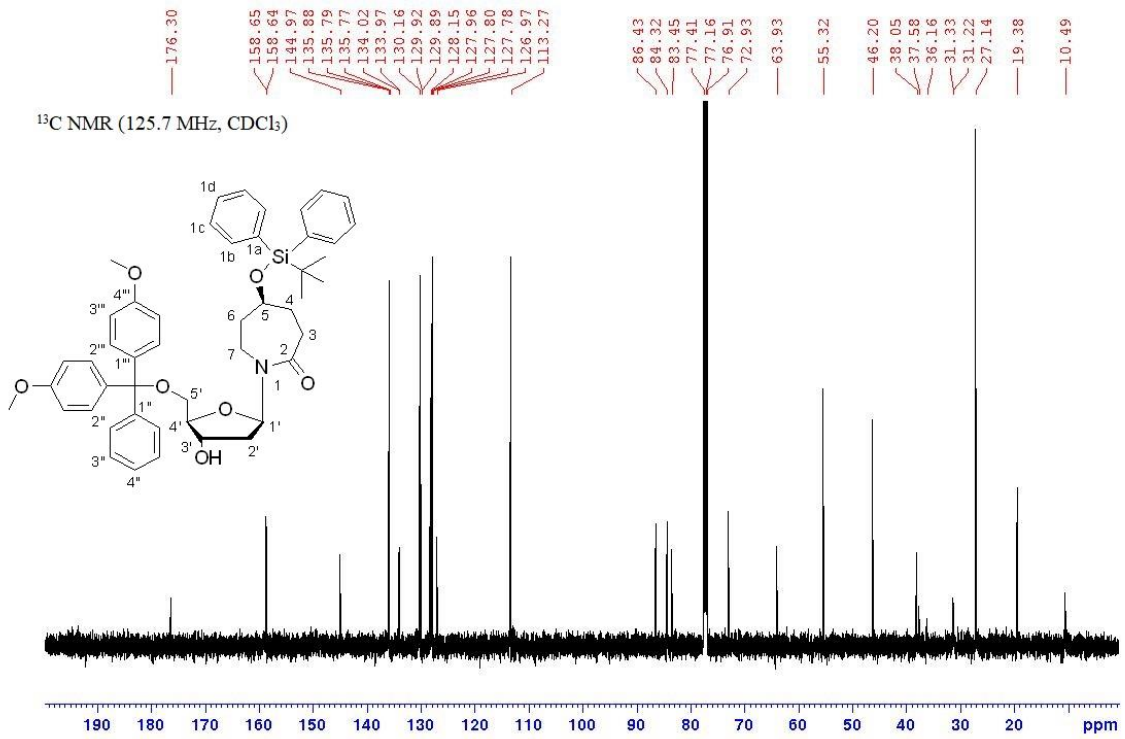


Figure S146. ^{13}C NMR spectrum of compound **39*S***.

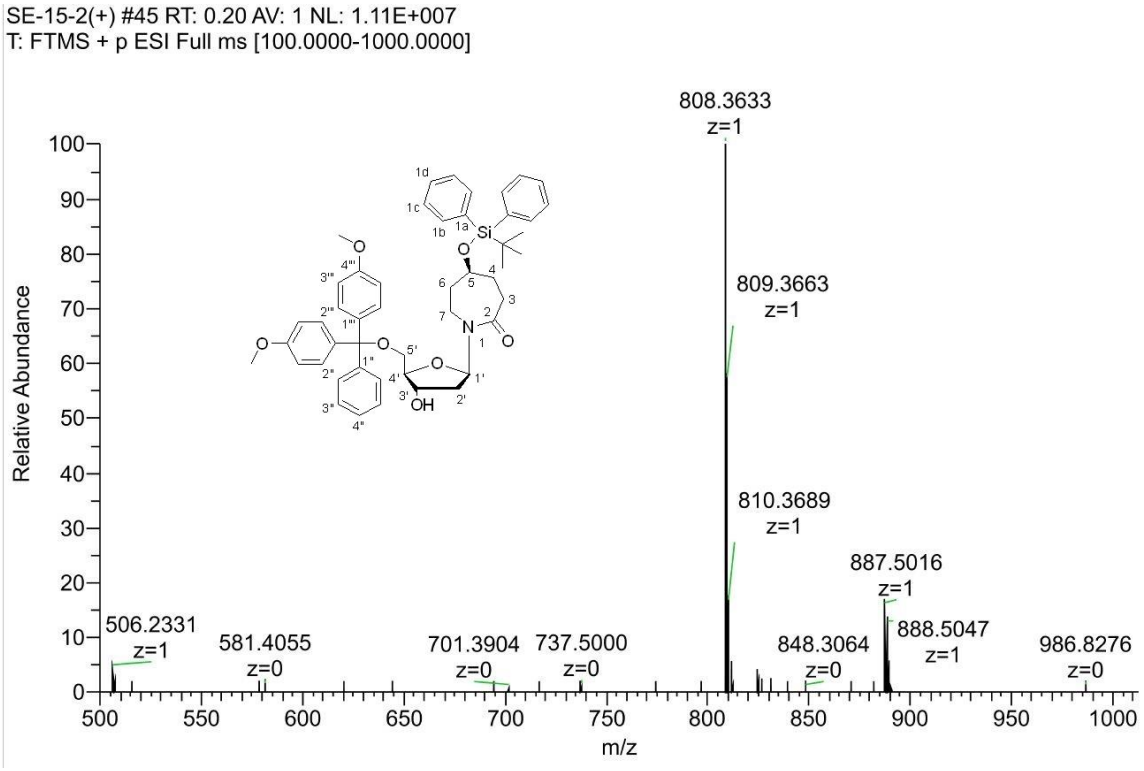


Figure S147. HRMS (ESI) of compound **39*S***.

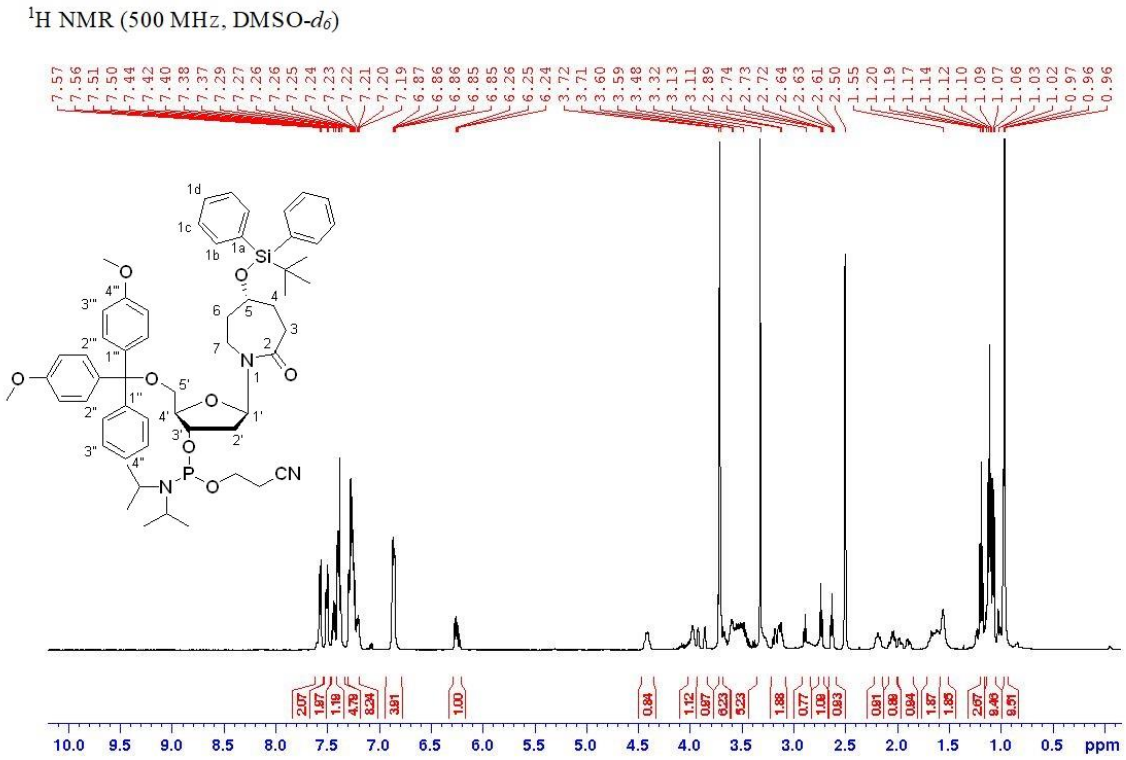


Figure S148. ^{1}H NMR spectrum of compound **40*R***.

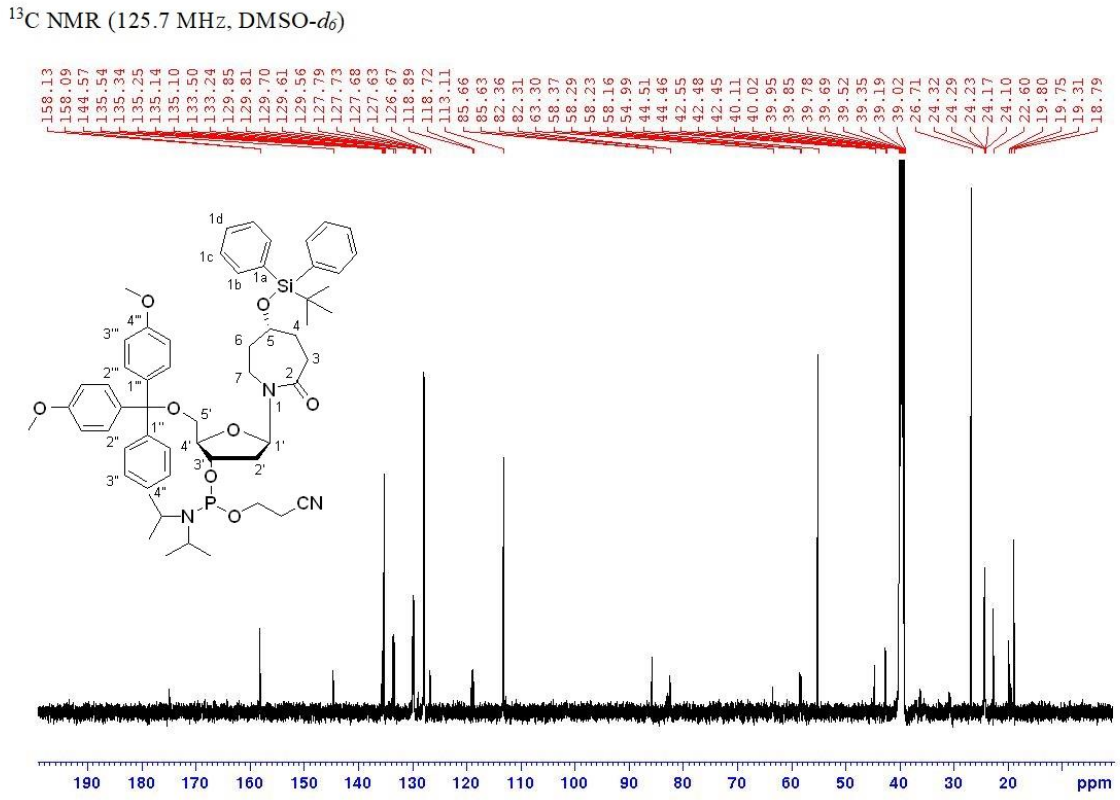


Figure S149. ^{13}C NMR spectrum of compound **40*R***.

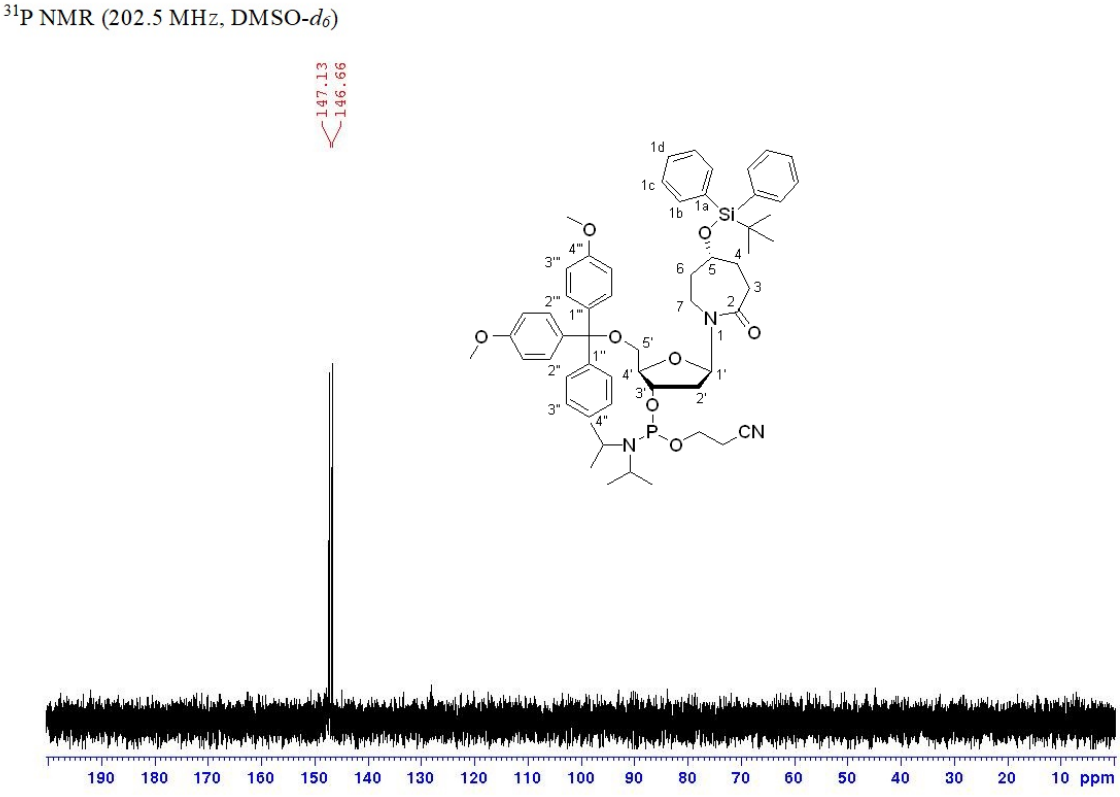


Figure S150. ^{31}P NMR spectrum of compound **40*R***.

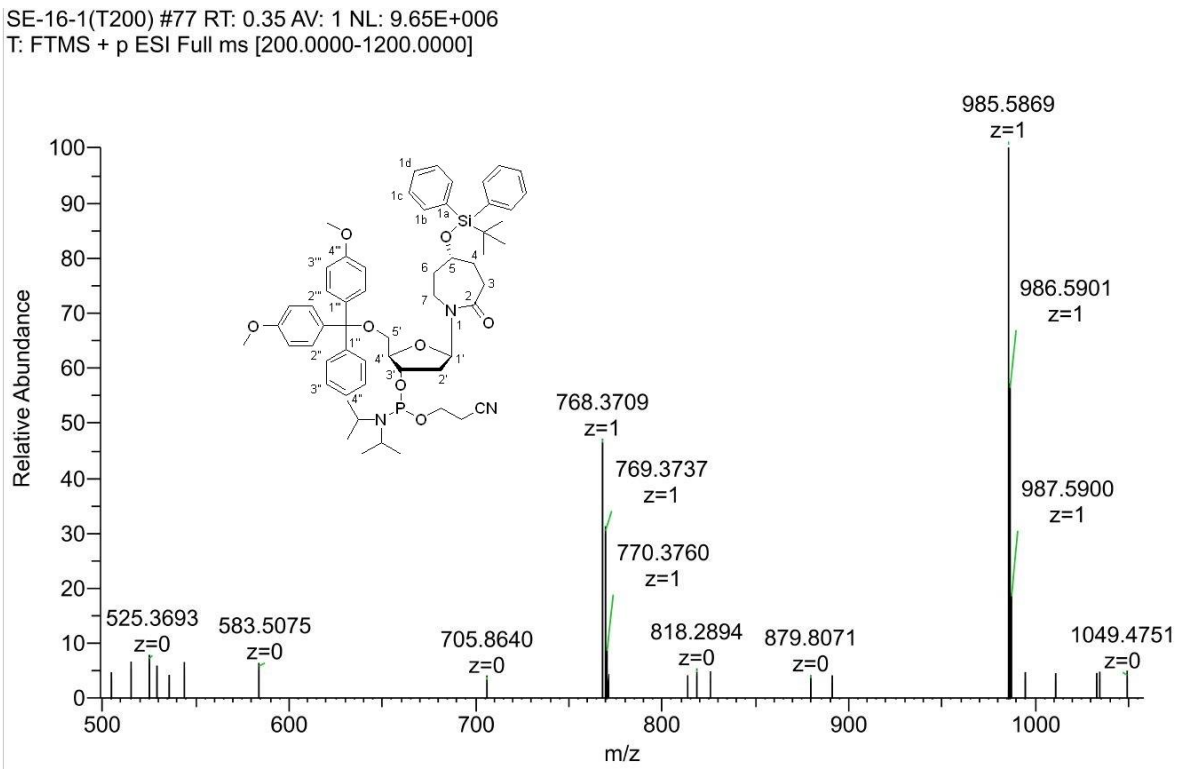


Figure S151. HRMS (ESI) of compound **40*R***.

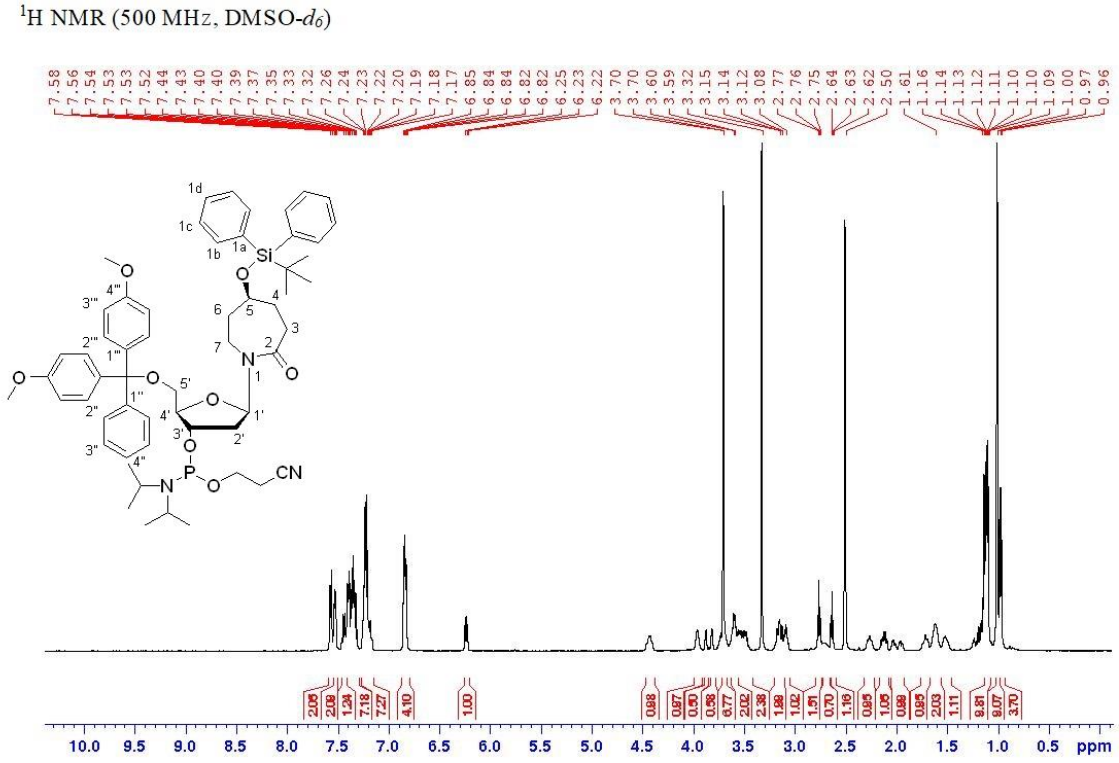


Figure S152. ^{1}H NMR spectrum of compound **40*S***.

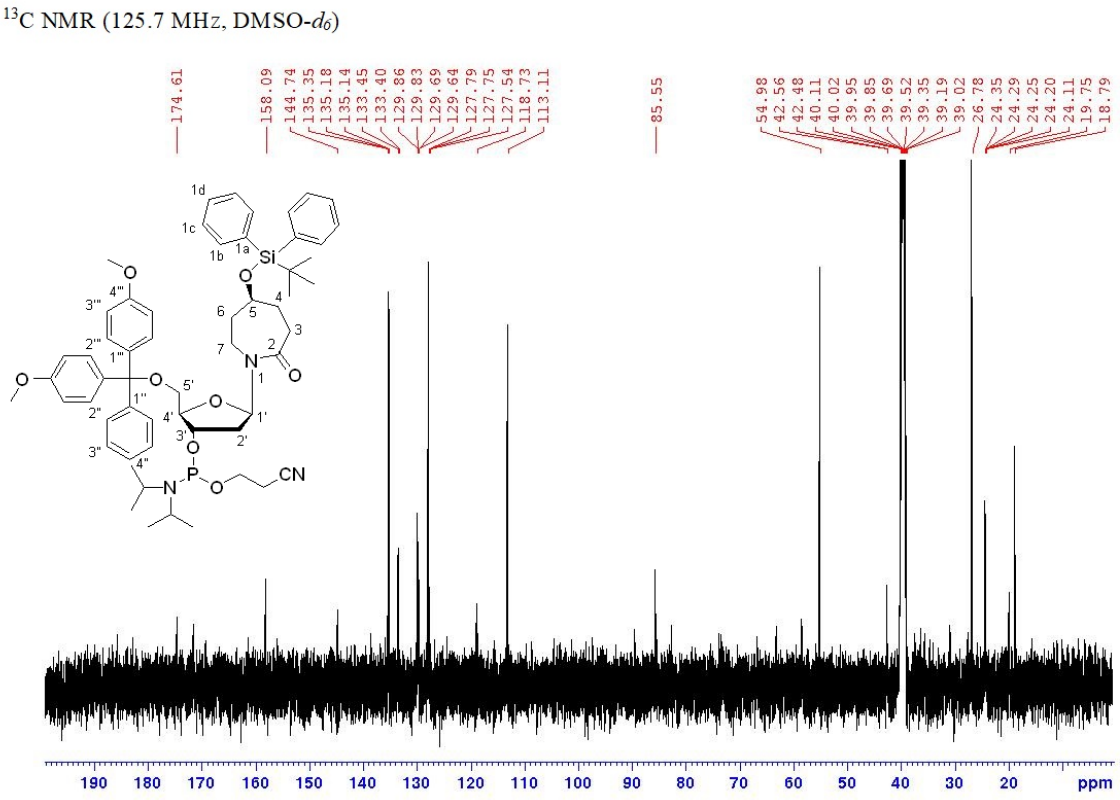


Figure S153. ^{13}C NMR spectrum of compound **40*S***.

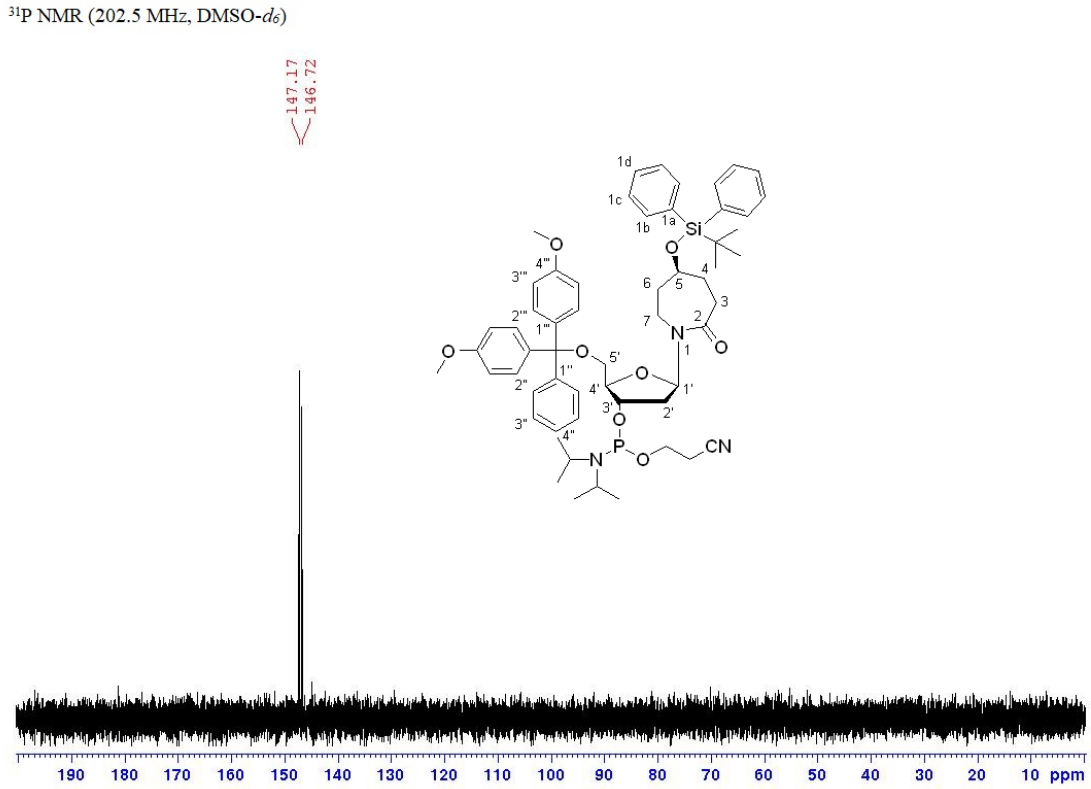


Figure S154. ^{31}P NMR spectrum of compound **40*S***.

^{1}H NMR (500 MHz, $CDCl_3$)

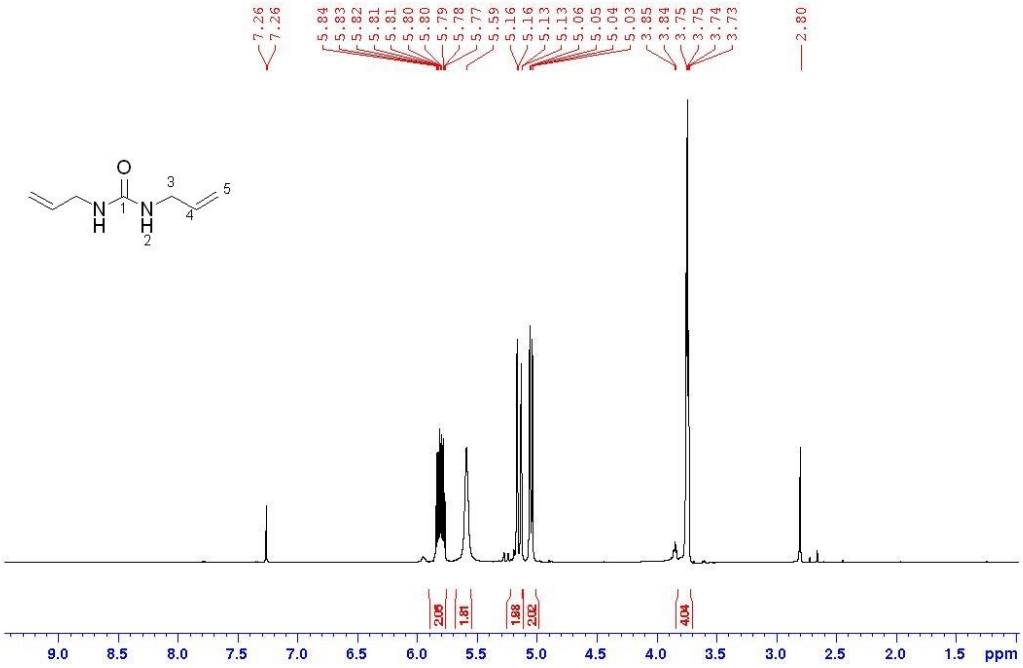


Figure S155. ^{1}H NMR spectrum of compound **42**.

^{13}C NMR (125.7 MHz, $CDCl_3$)

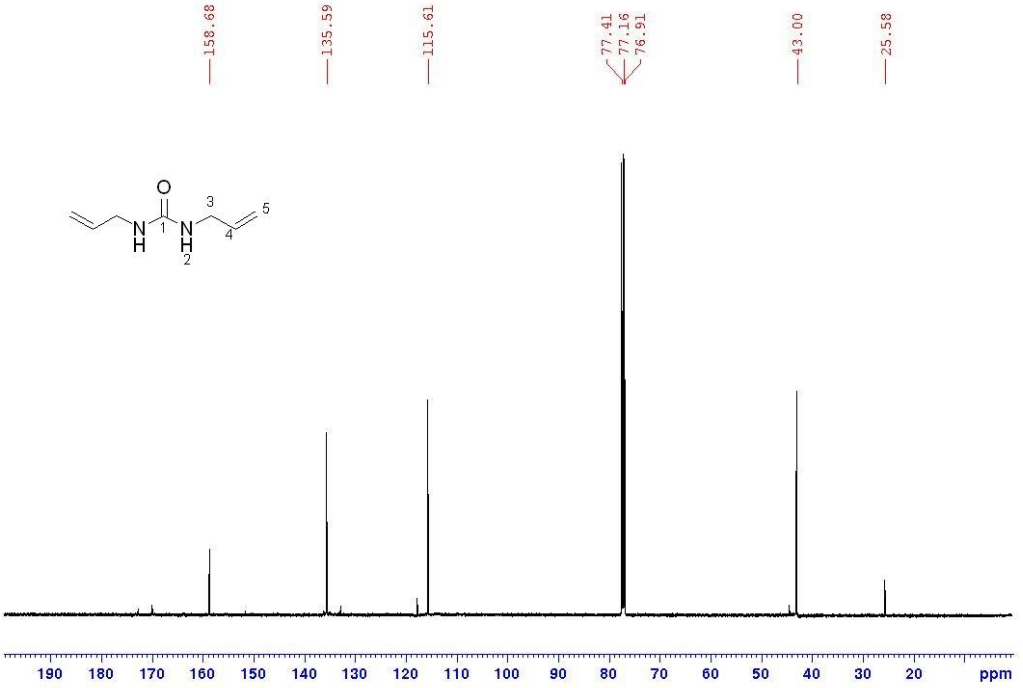


Figure S156. ^{13}C NMR spectrum of compound **42**.

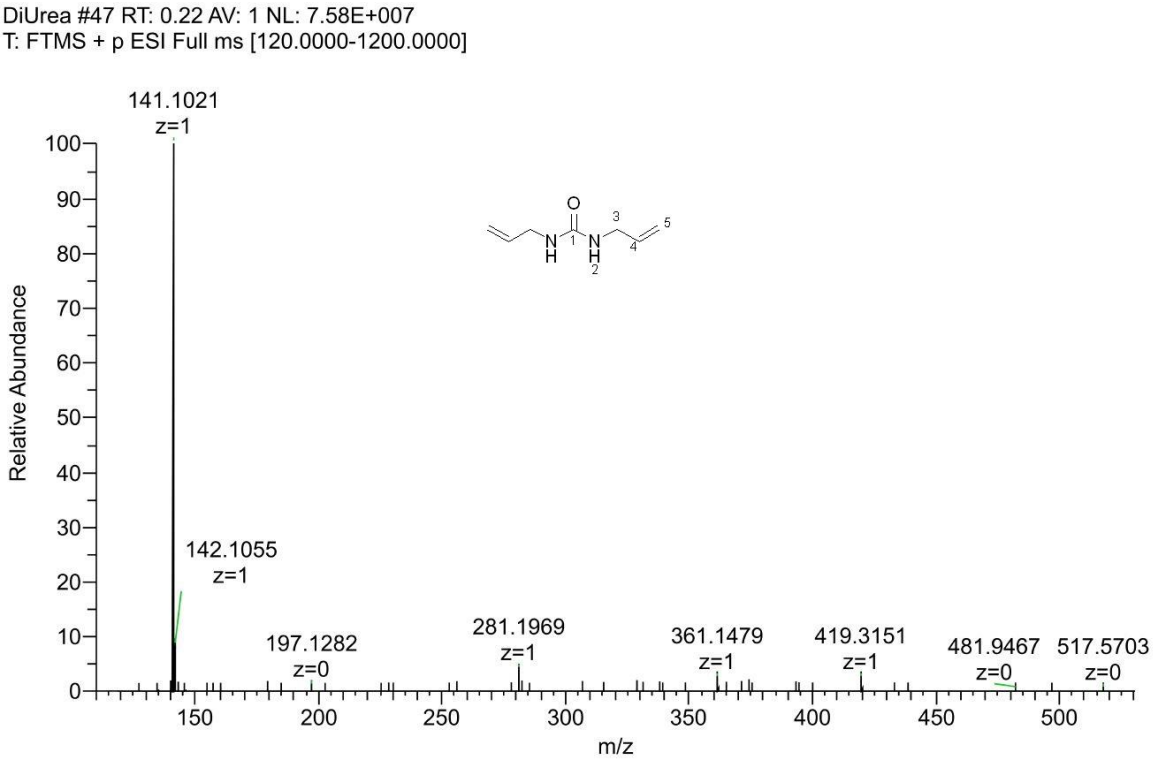


Figure S157. HRMS (ESI) of compound **42**.

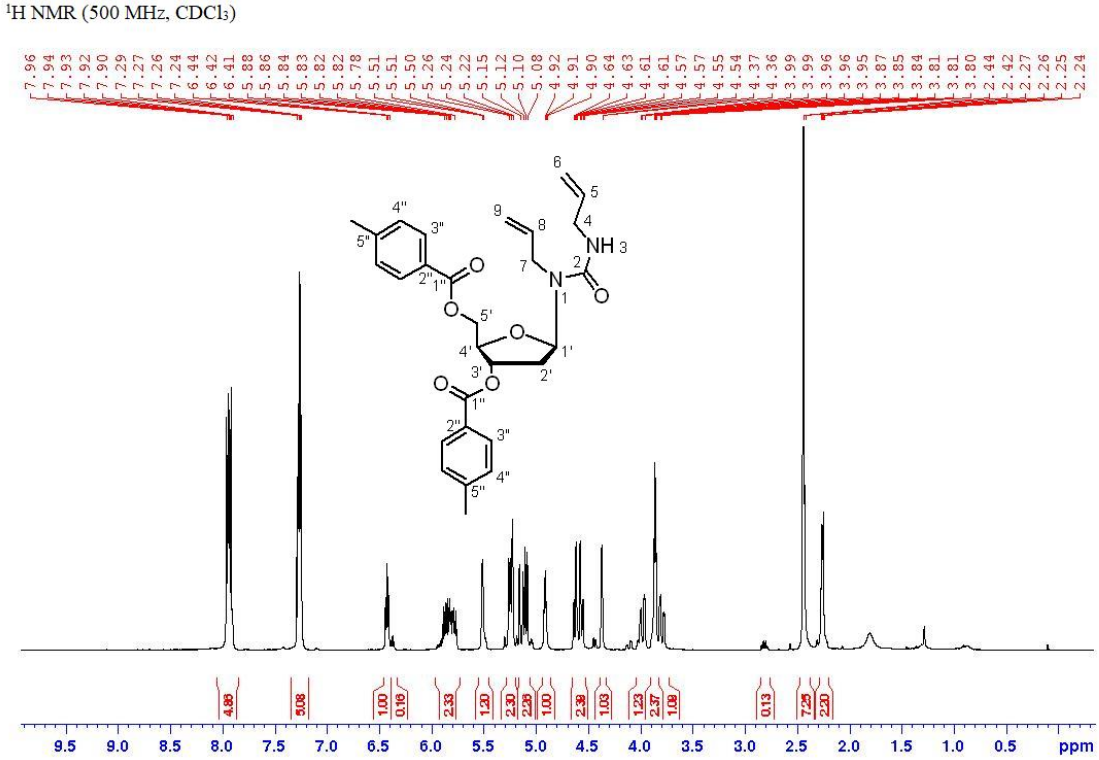


Figure S158. ^{1}H NMR spectrum of compound **43**.

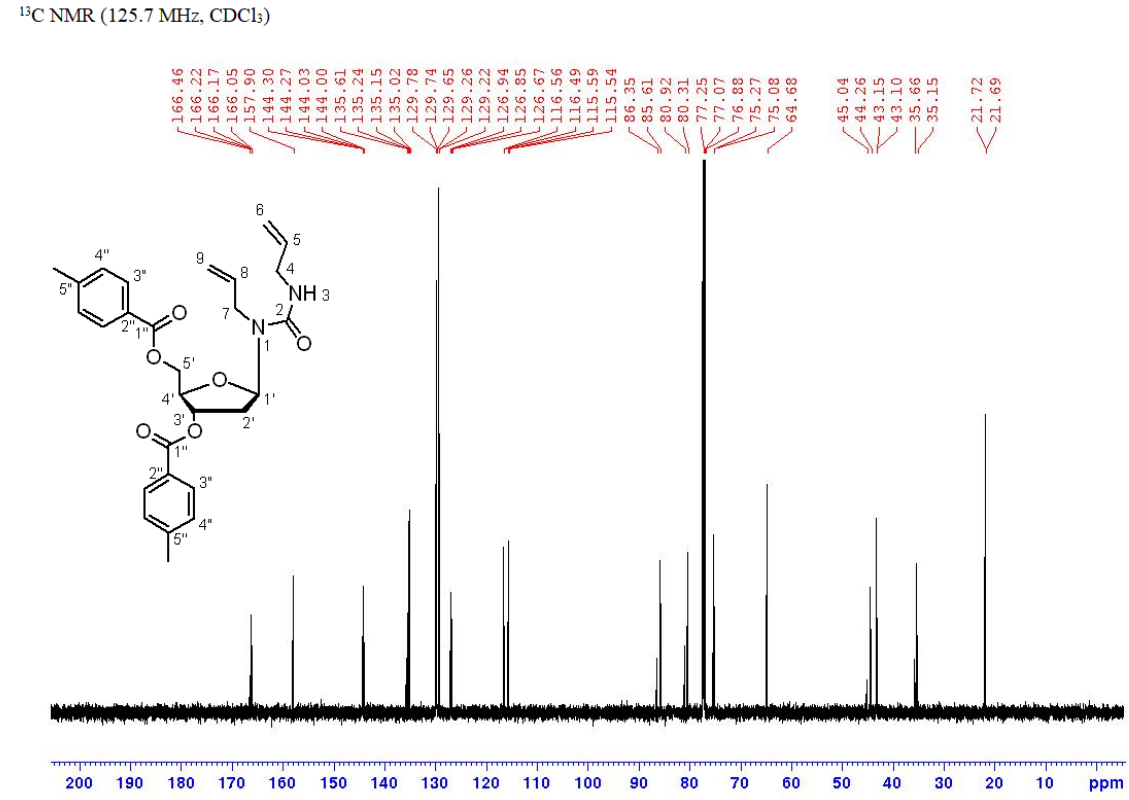


Figure S159. ^{13}C NMR spectrum of compound **45**.

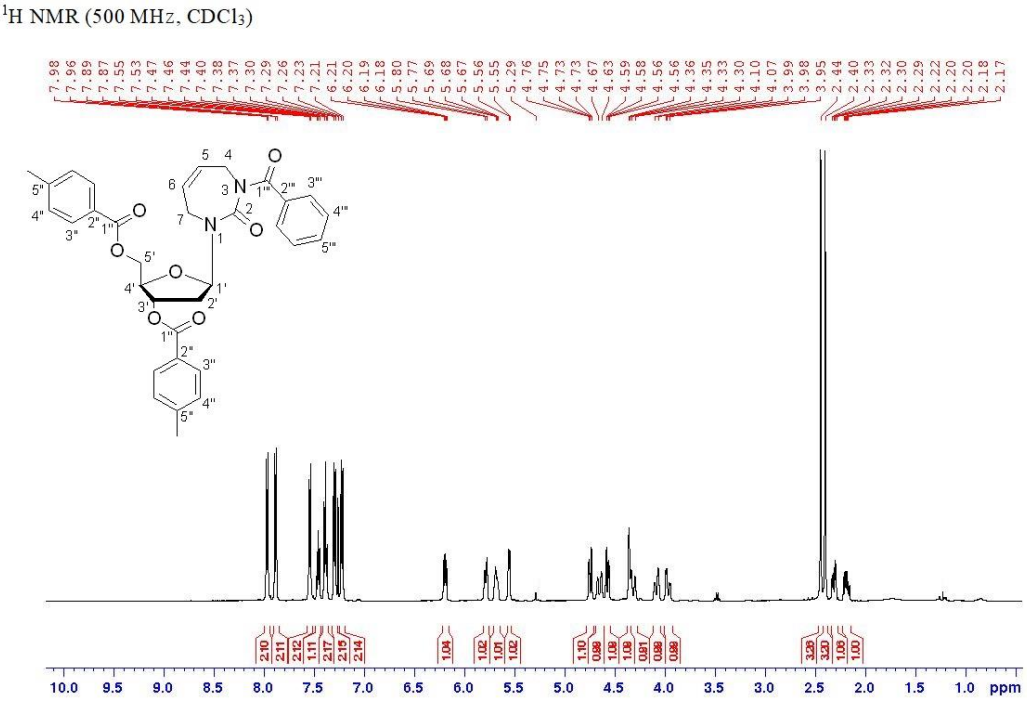


Figure S160. ^{1}H NMR spectrum of compound **45**.

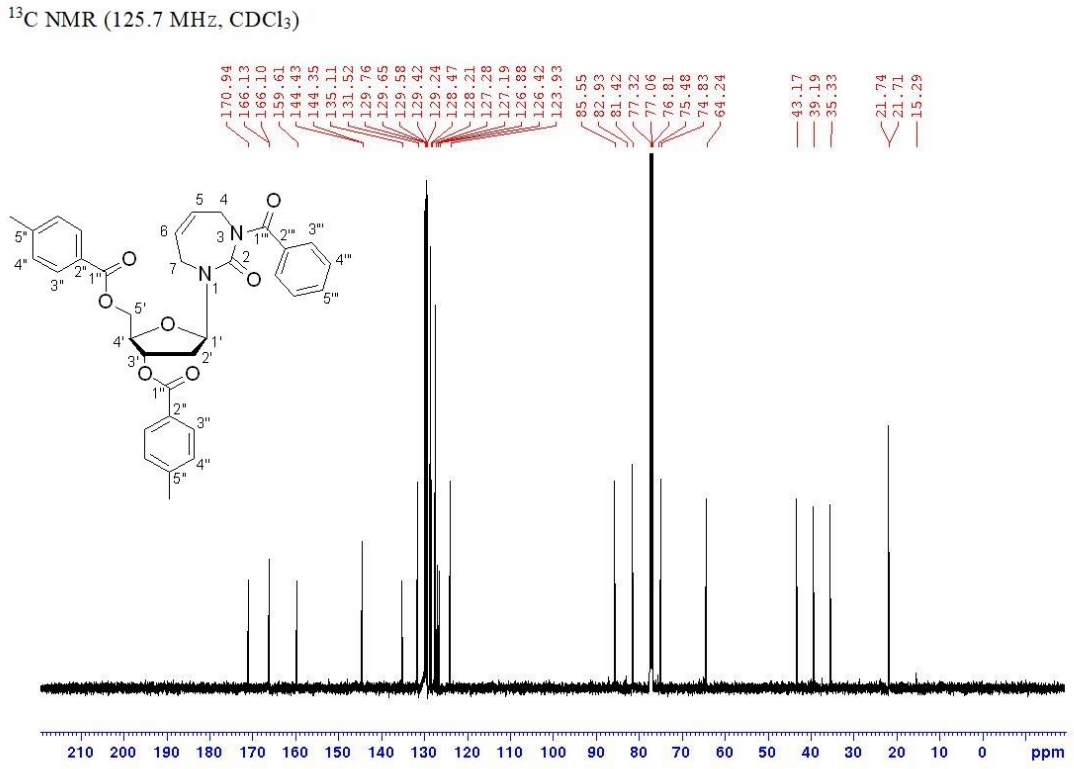


Figure S161. ^{13}C NMR spectrum of compound **45**.

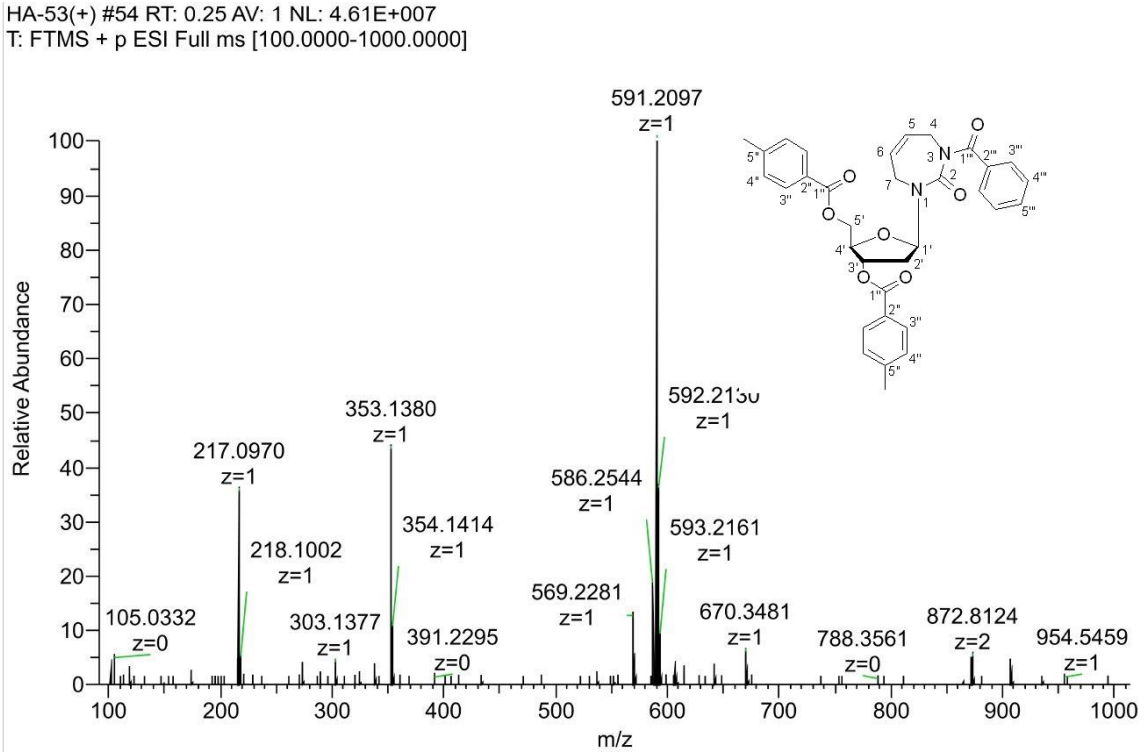


Figure S162. HRMS (ESI) of compound **45**.

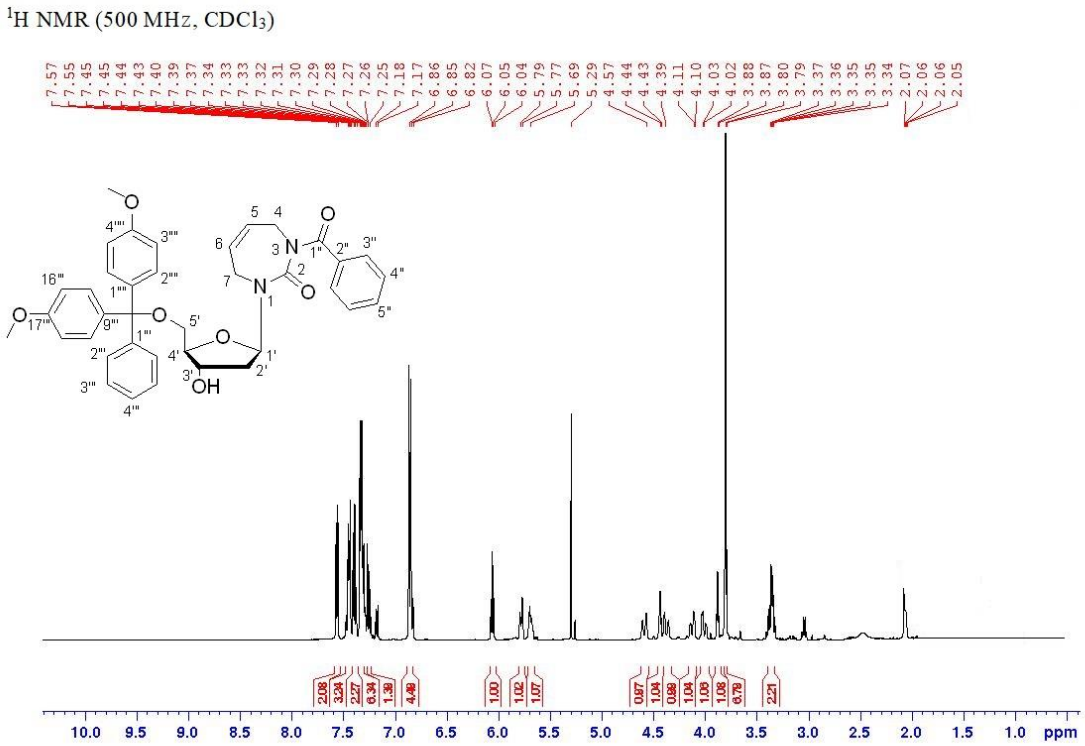


Figure S163. ^{1}H NMR spectrum of compound **46**.

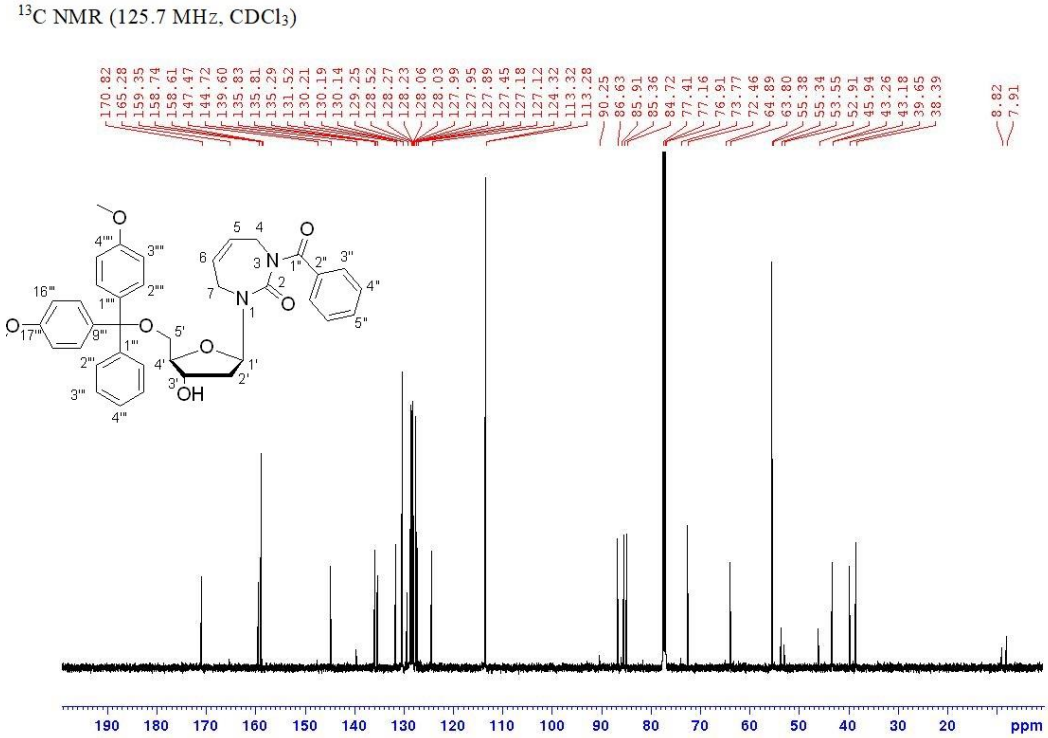


Figure S164. ^{13}C NMR spectrum of compound **46**.

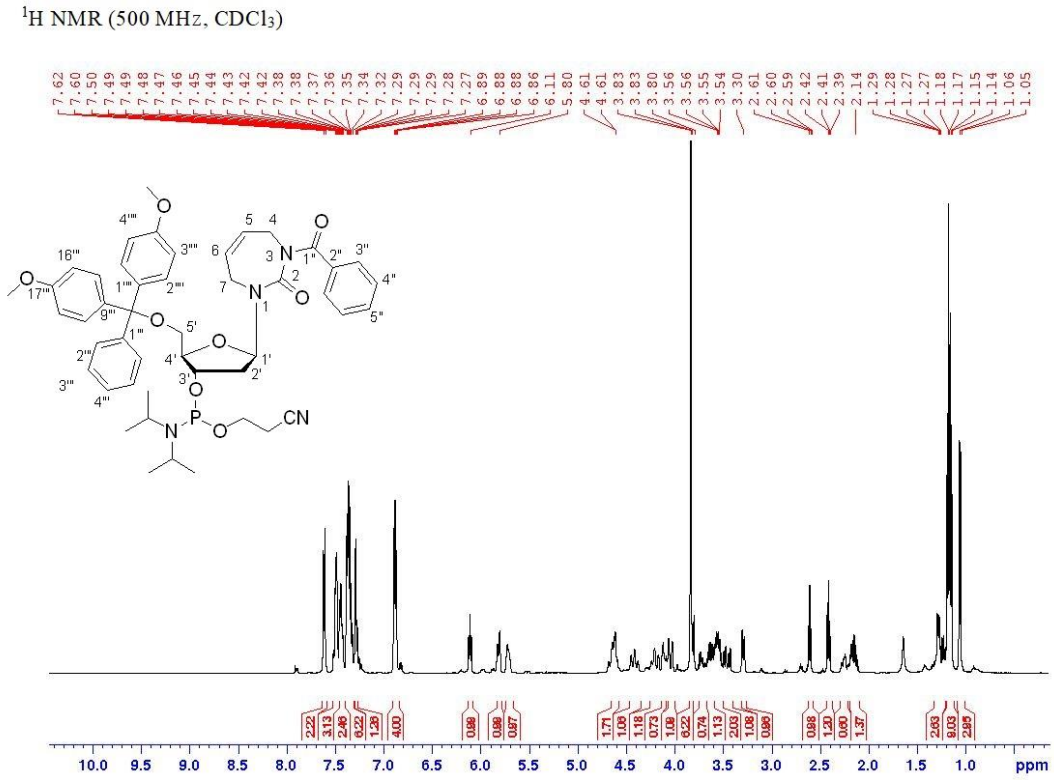


Figure S165. ^{1}H NMR spectrum of compound **47**.

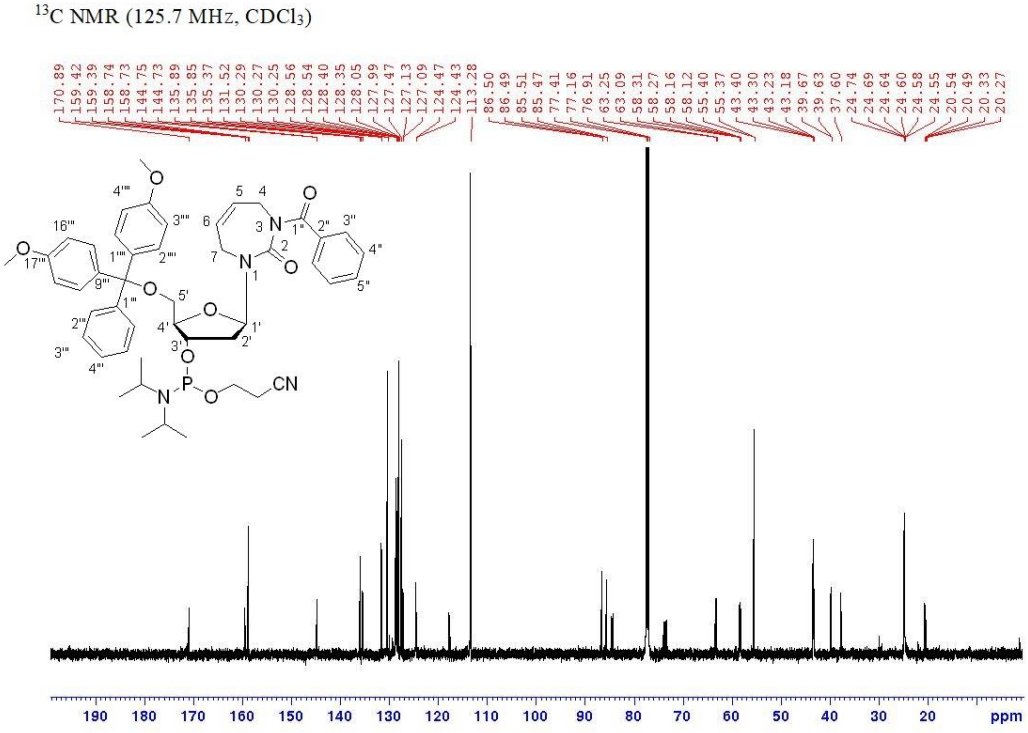


Figure S166. ^{13}C NMR spectrum of compound **47**.

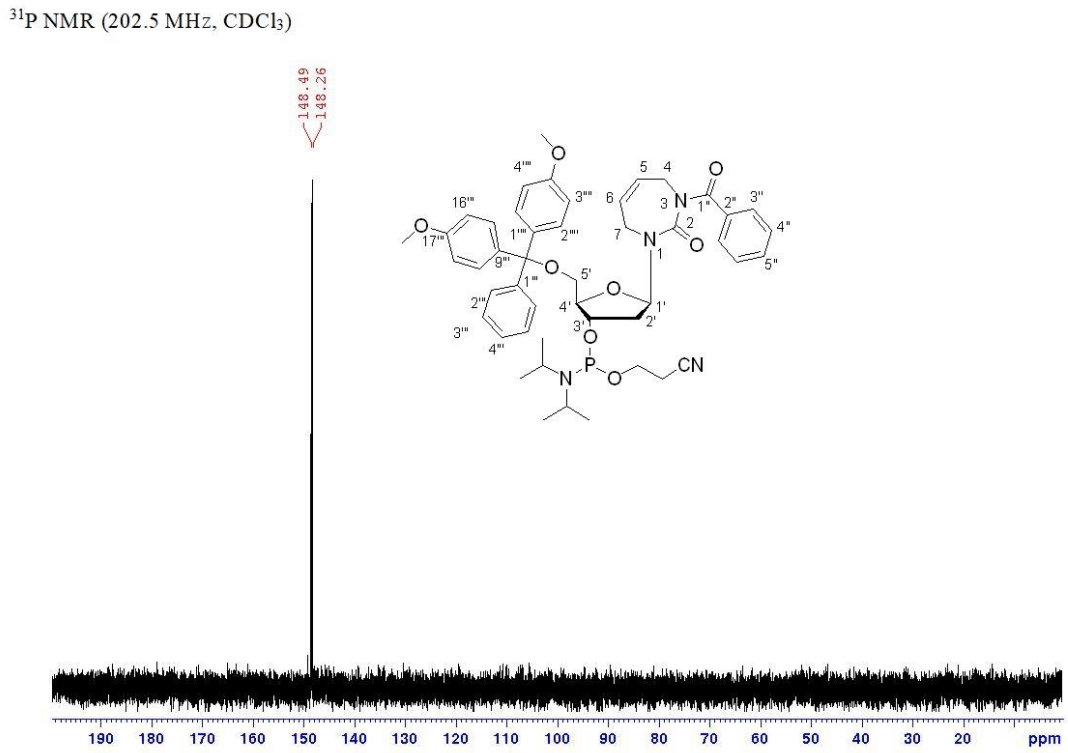


Figure S167. ^{31}P NMR spectrum of compound **47**.

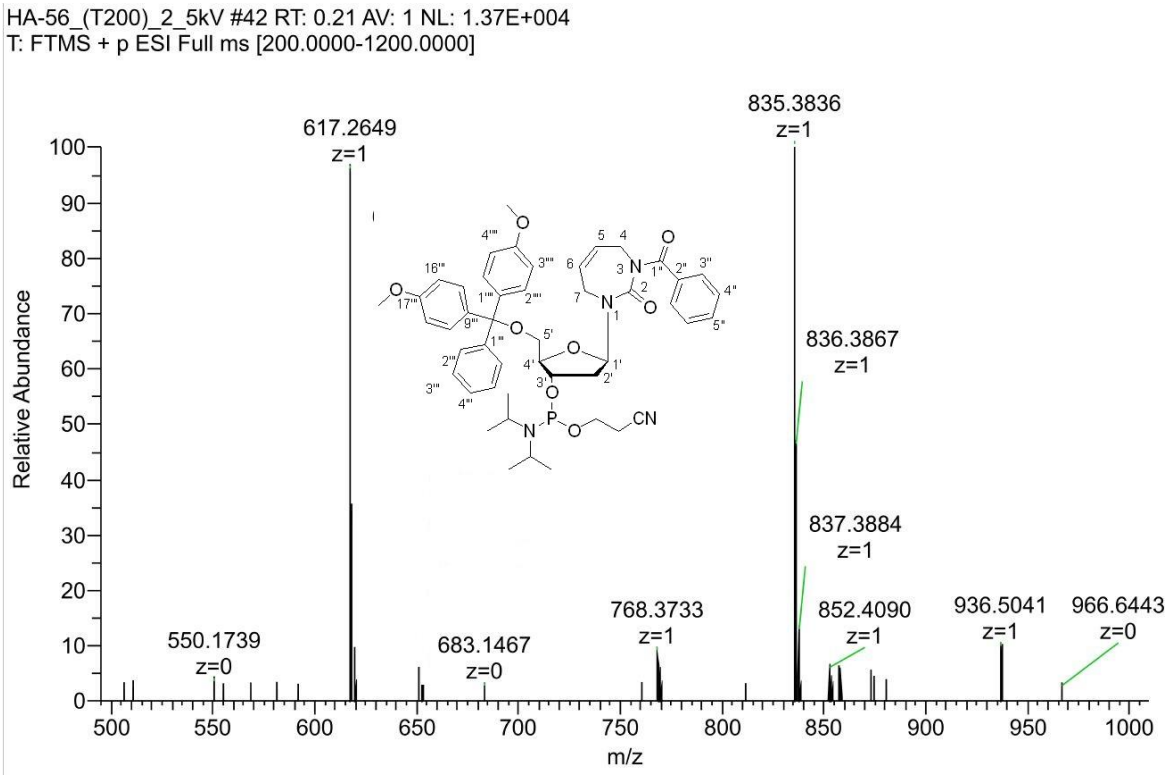


Figure S168. HRMS (ESI) of compound **47**.

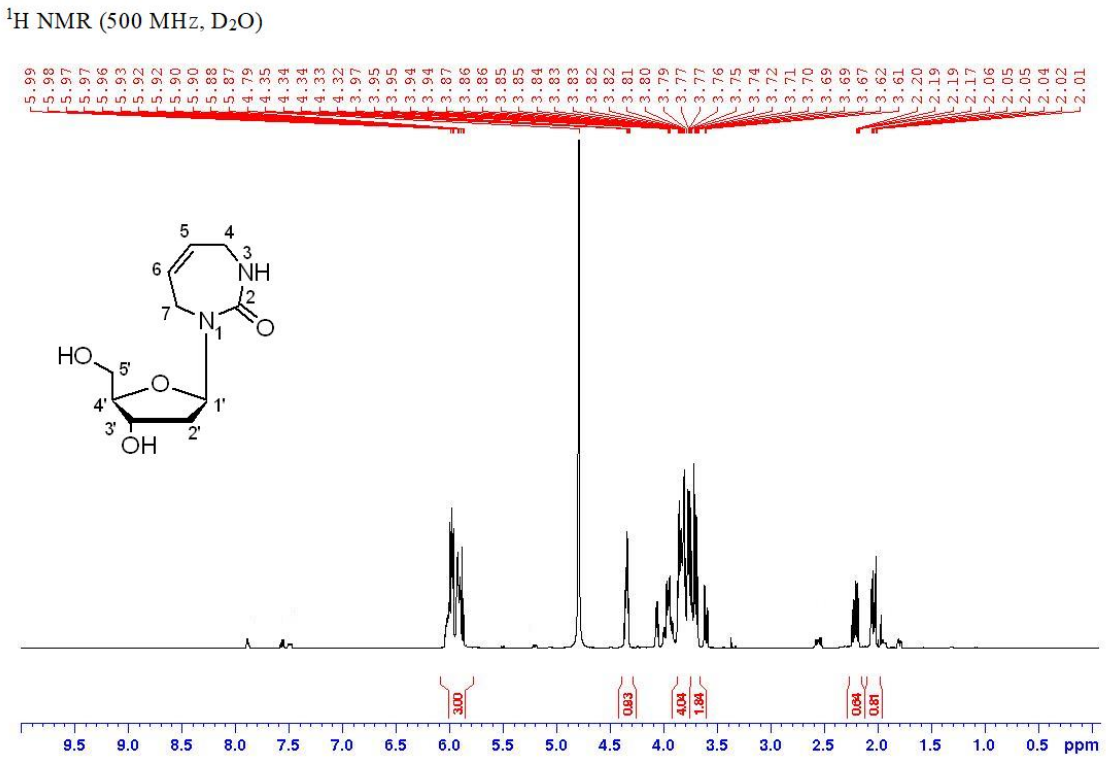


Figure S169. ^{1}H NMR spectrum of compound 48.

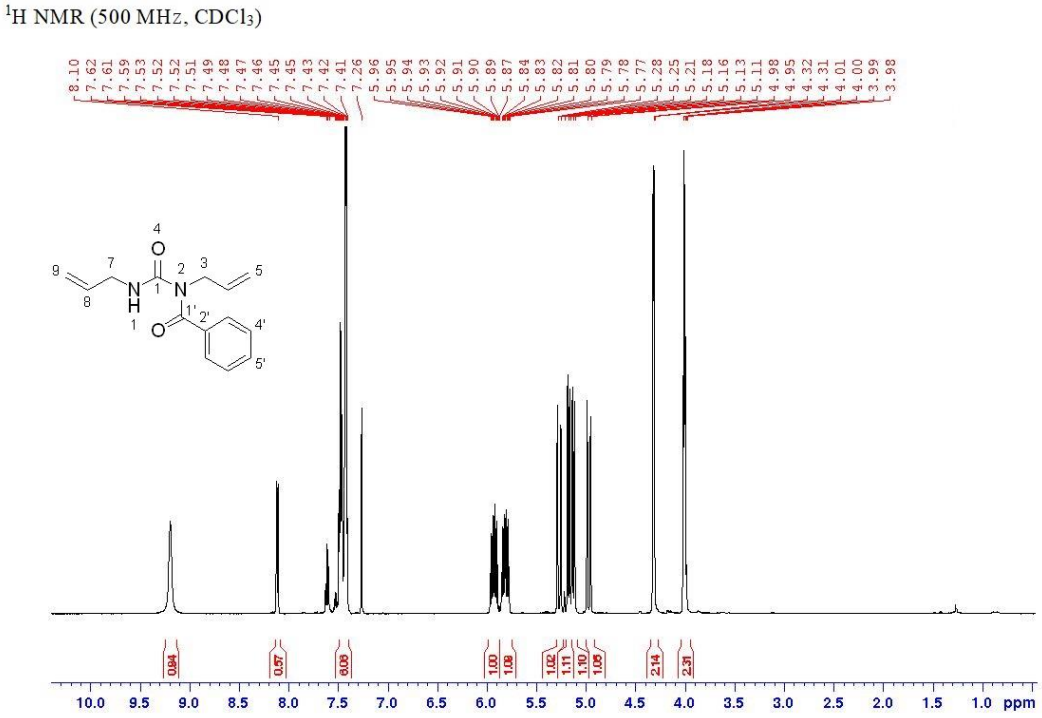


Figure S170. ^{1}H NMR spectrum of compound **49**.

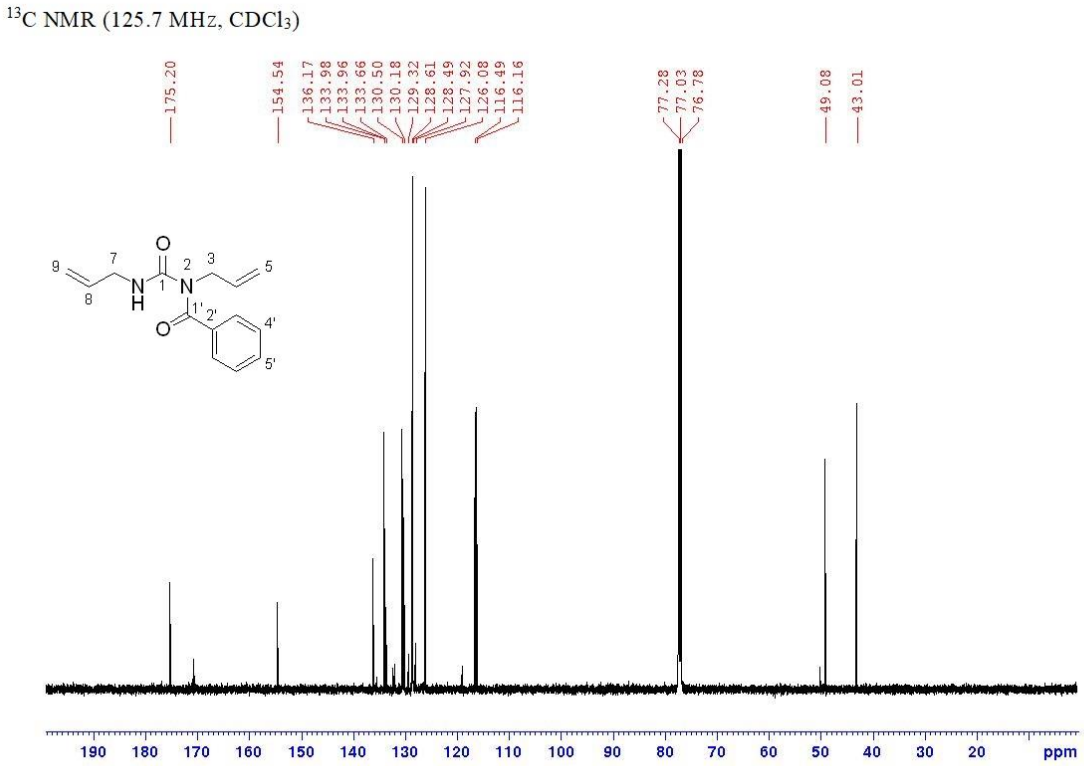


Figure S171. ^{13}C NMR spectrum of compound **49**.

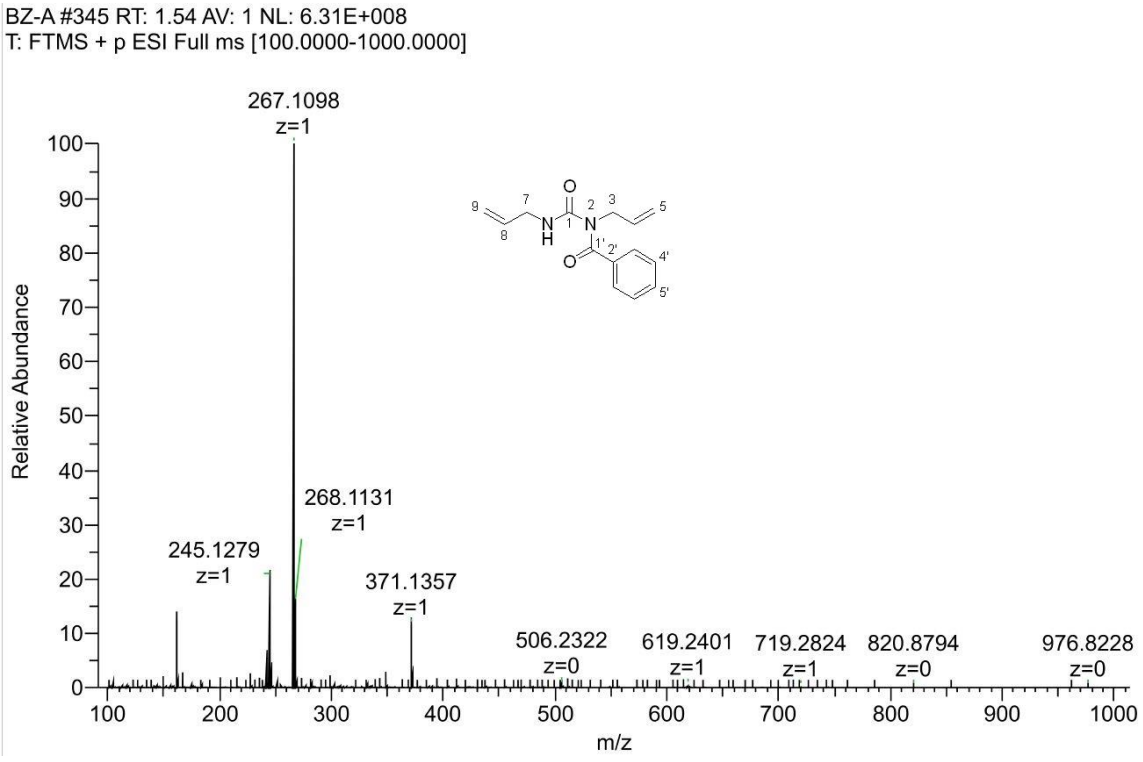


Figure S172. HRMS (ESI) of compound **49**.

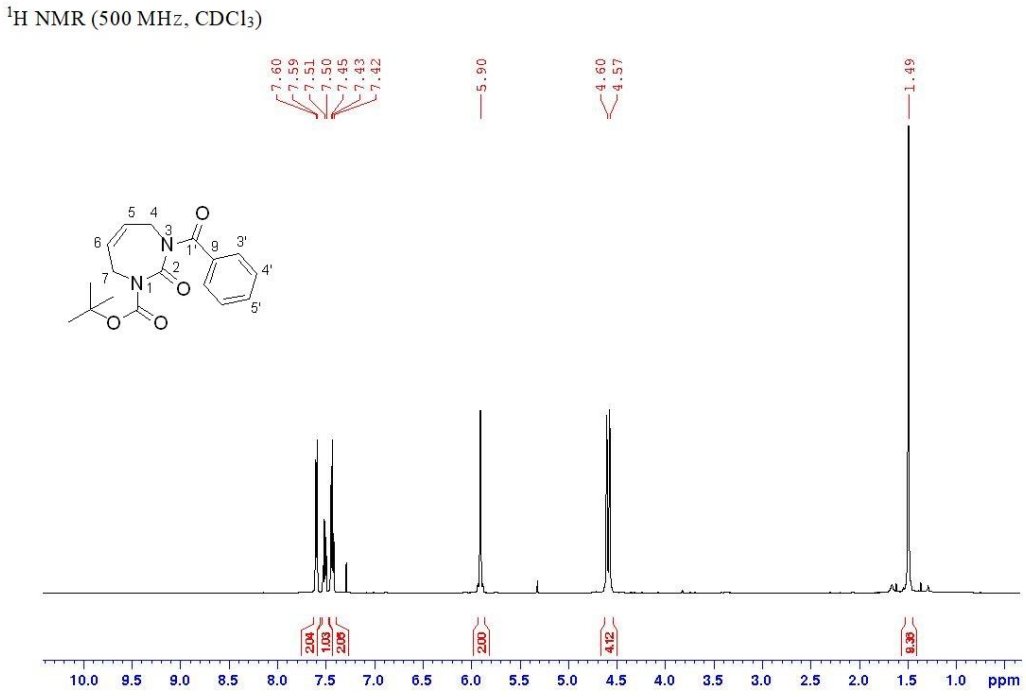


Figure S173. ^{1}H NMR spectrum of compound **51**.

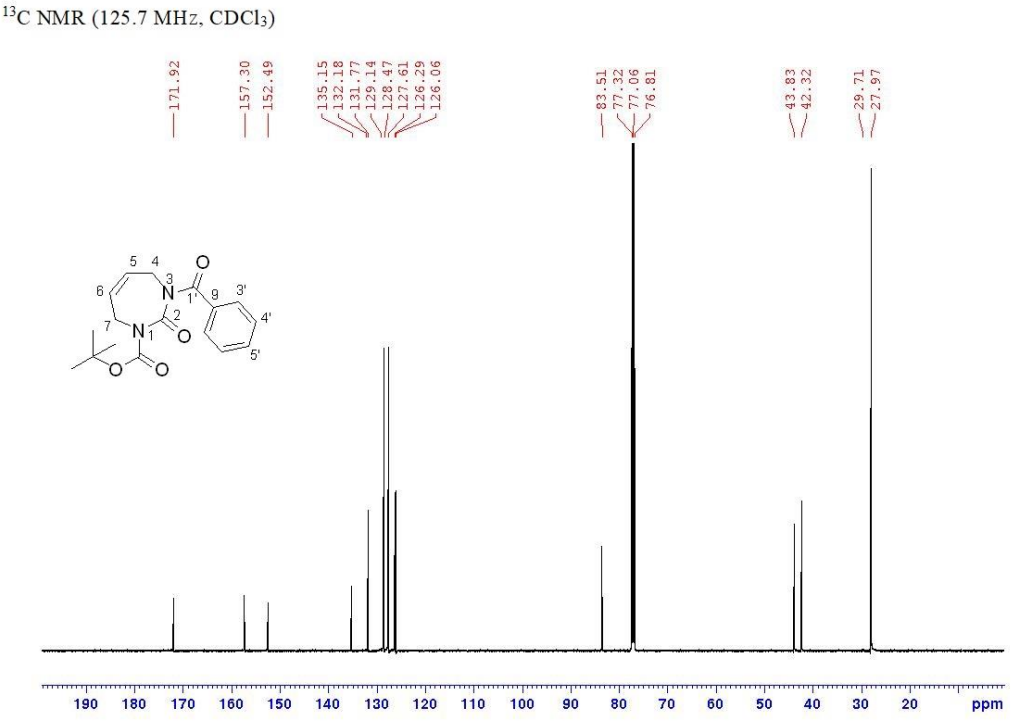


Figure S174. ^{13}C NMR spectrum of compound **51**.

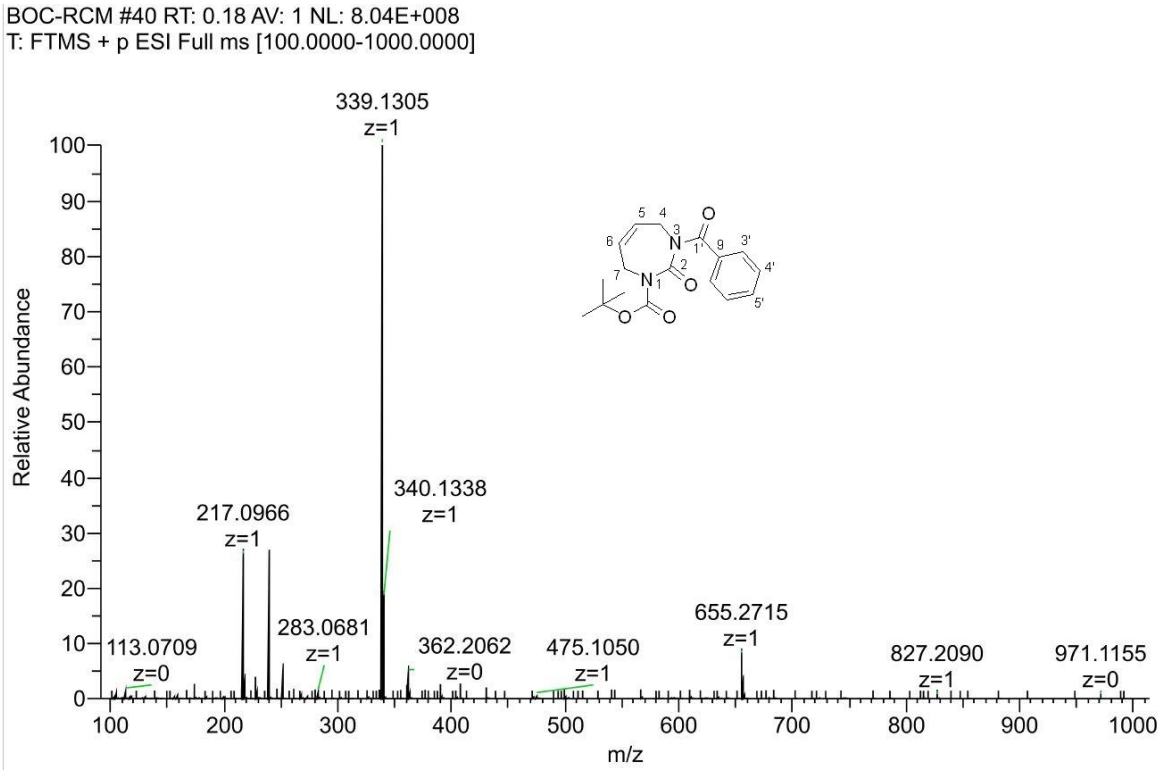


Figure S175. HRMS (ESI) of compound **51**.

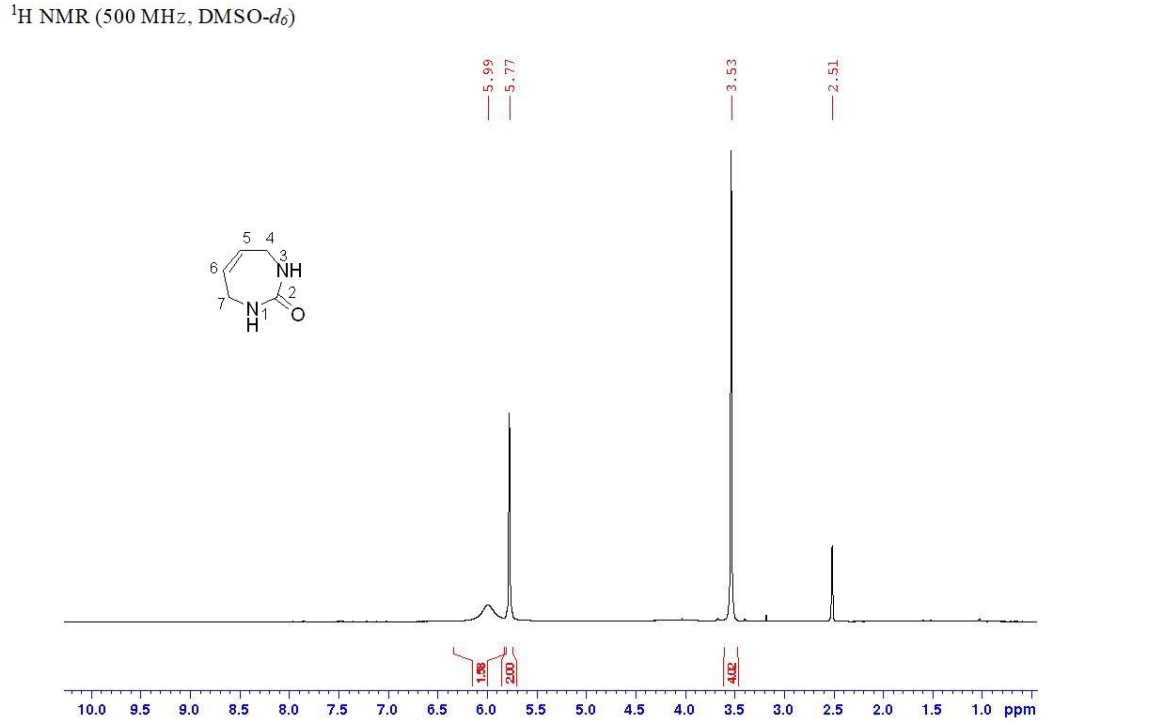


Figure S176. ^{1}H NMR spectrum of compound **53**.

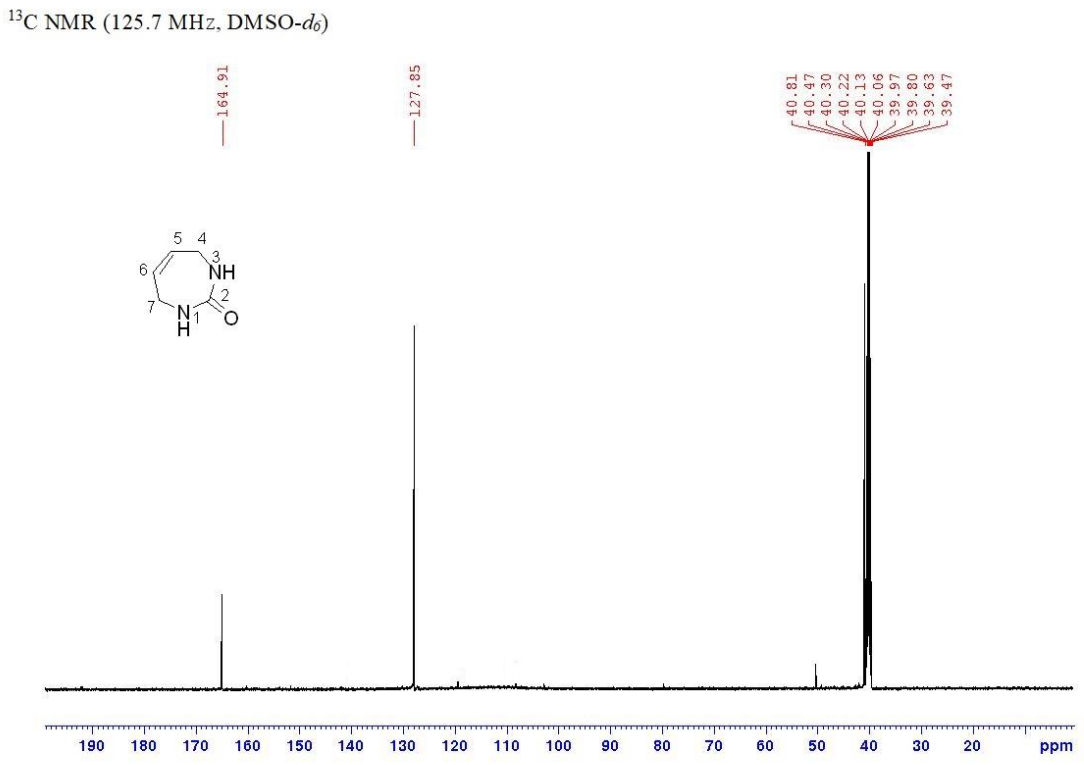


Figure S177. ^{13}C NMR spectrum of compound **53**.

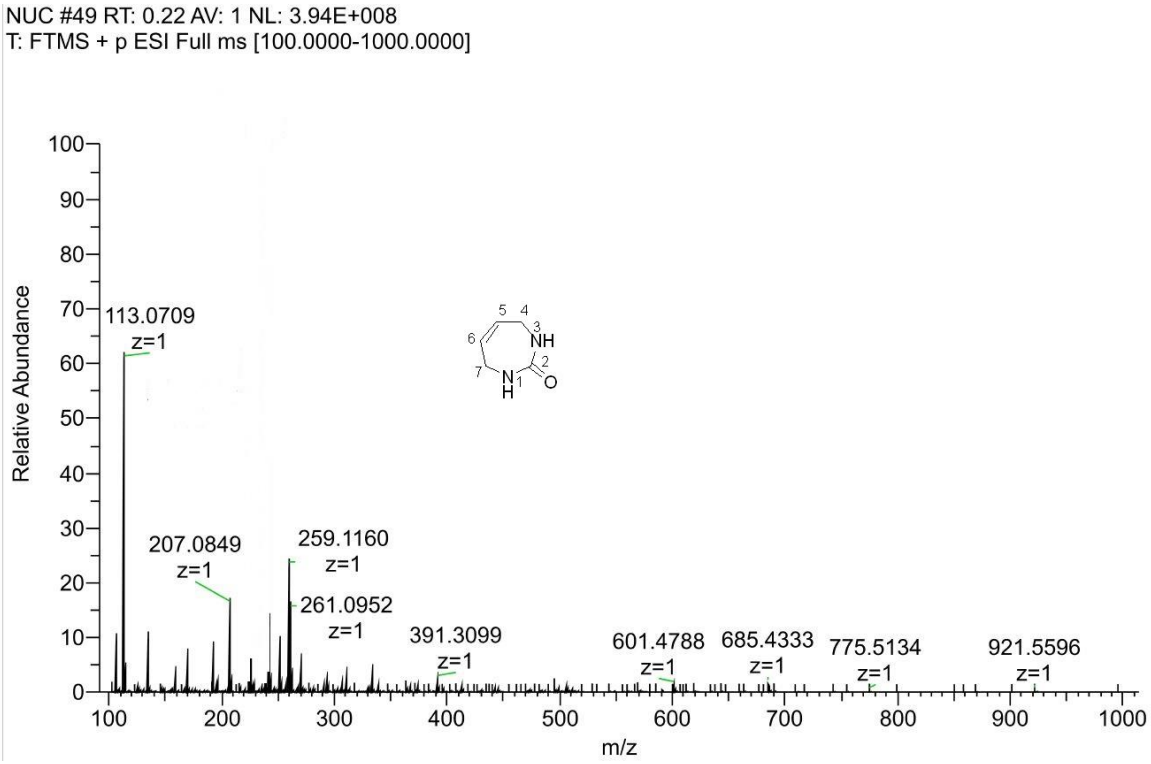


Figure S178. HRMS (ESI) of compound **53**.

Experimental data for oligonucleotides synthesized

Table S 1. Terminally cross-linked oligos and their ε260s.

Oligo abbreviation	ε260 ($L \cdot mol^{-1} \cdot cm^{-1}$)
dC-9mer-X	66300
dC-7mer-X	48900
dC-5mer-X	31500

Table S 2. Internally cross-linked oligos and their ε260s.

Oligo abbreviation	ε260 ($L \cdot mol^{-1} \cdot cm^{-1}$)
Cross-link 1	
dC[R^{N3}(-3),A^{Y2}(+1)]X	81700
dC[R^{N3}(-2),A^{Y2}(+1)]X	81700
dC[R^{N3}(-1),A^{Y2}(+1)]X	81700
dC[R^{N3}(-2),A^{Y4}(+1)]X	81700
dZ[R^{N3}(-3),A^{Y2}(+1)]X	76100
dZ[R^{N3}(-2),A^{Y2}(+1)]X	76100
dZ[R^{N3}(-1),A^{Y2}(+1)]X	76100
dZ[R^{N3}(-2),A^{Y4}(+1)]X	76100
Cross-link 2	
dC[U^{E}(-3),A^{N3}(+1)]X	90400
dC[U^{E}(-2),A^{N3}(+1)]X	90400
dZ[U^{E}(-3),A^{N3}(+1)]X	84800
dZ[U^{E}(-2),A^{N3}(+1)]X	84800
Cross-link 3	
dC[C^{N3}(-2),H^{E}(+1)]X	96000
dZ[C^{N3}(-2),H^{E}(+1)]X	90400

FdZ[C^{N3}(-2),H^{E}(+1)]X	90400
4-mer FdZ[C^{N3}(-2),H^{E}(+1)]X	40200
4-mer dZ[C^{N3}(-2),H^{E}(+1)]X	40200
3-mer dZ[C^{N3}(-2),H^{E}(+1)]X	31500

Table S 3. Linear oligos and their ε260s.

Oligo abbreviation	ε260 ($L \cdot mol^{-1} \cdot cm^{-1}$)
FdZ-Hairpin	94800
dZ linear	84800
FdZ linear	60700
9-mer ddiazep	83000
7-mer *R*-dazep	58900
7-mer S-dazep	58900

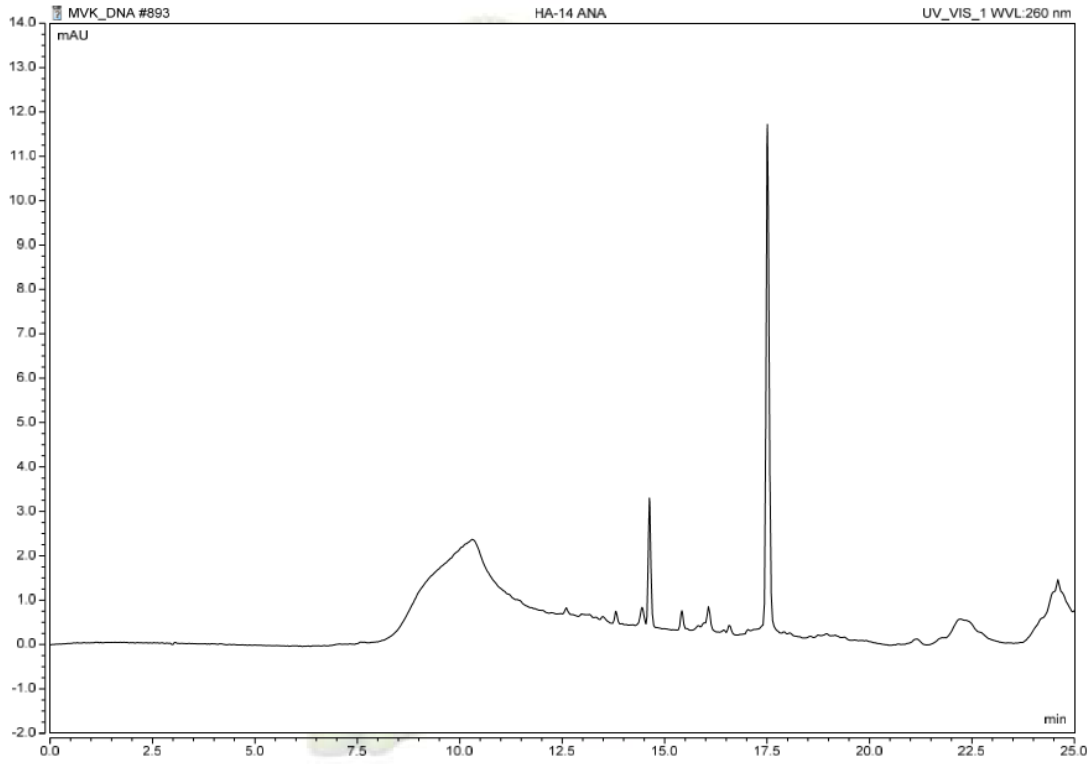


Figure S179. Reverse phase HPLC profile of dC-9mer.

Note that the broad peak at 10 min is an artefact of the column.

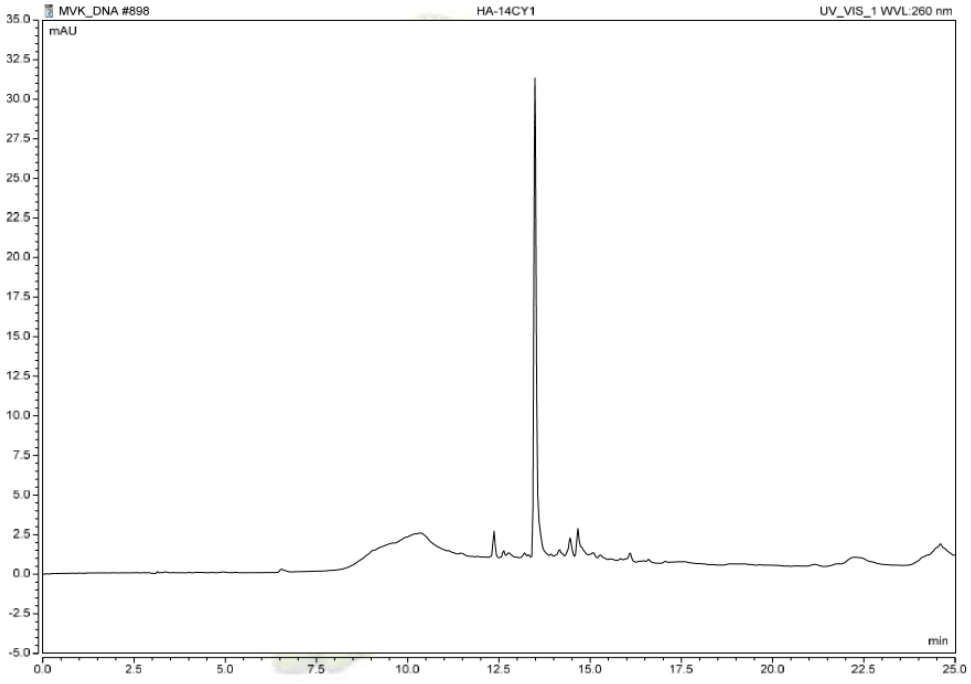


Figure S180. Reverse phase HPLC profile of dC-9mer-X**.**

Note that the broad peak at 10 min is an artefact of the column.

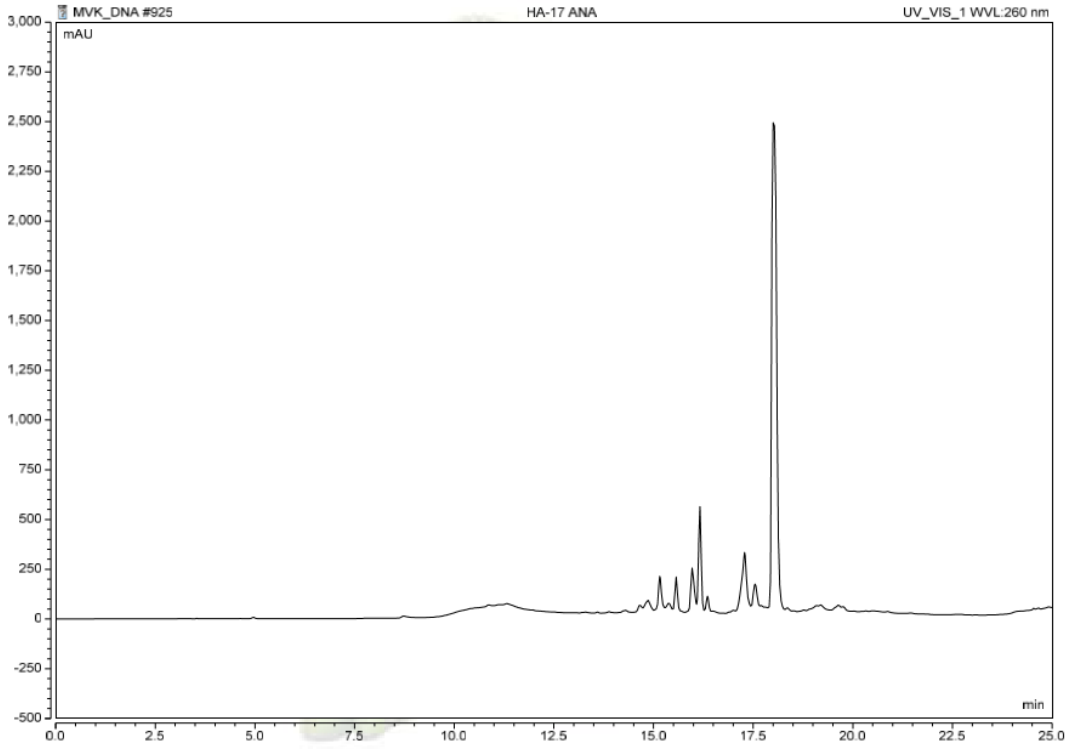


Figure S181. Reverse phase HPLC profile of dC-7mer.

Note that the broad peak at 10 min is an artefact of the column.

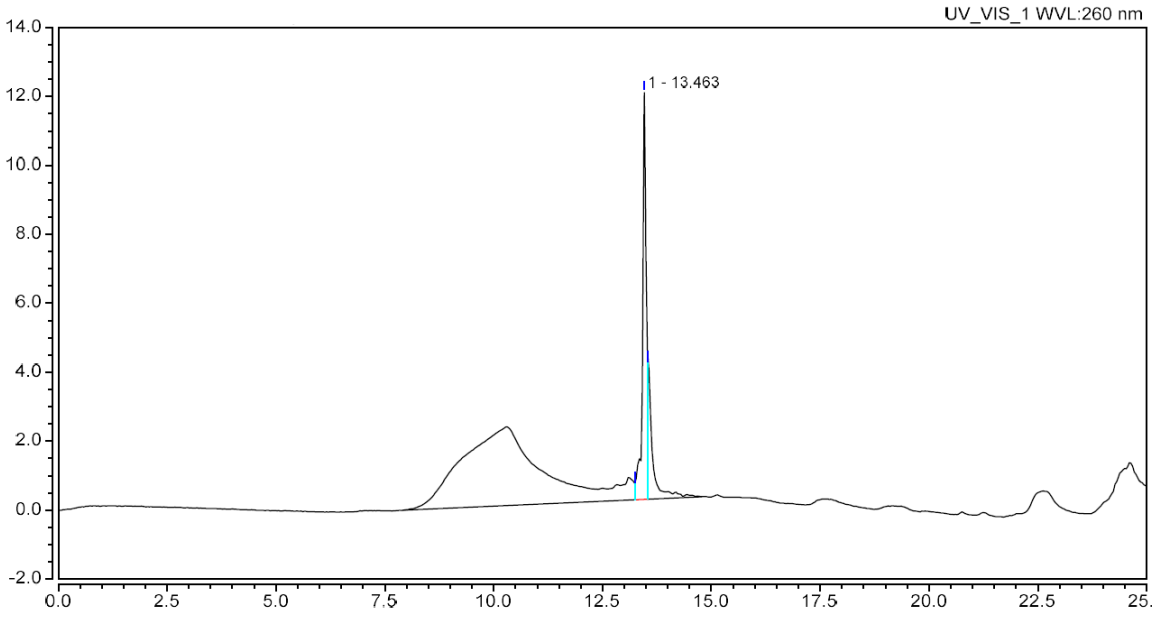


Figure S182. Reverse phase HPLC profile of dC-7mer-X.

Note that the broad peak at 10 min is an artefact of the column.

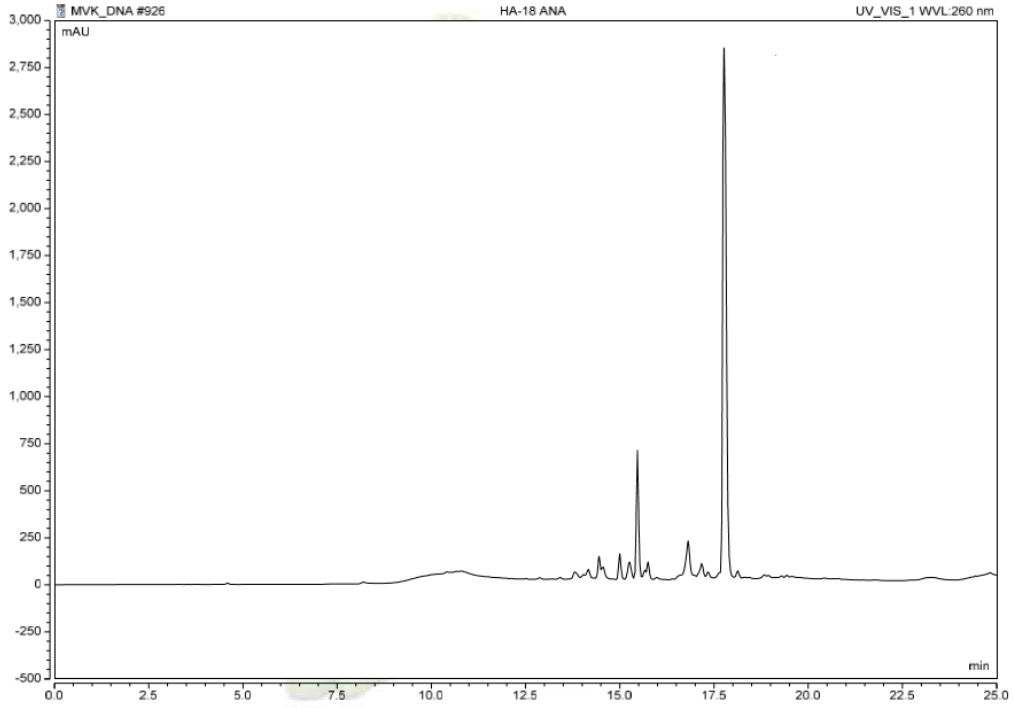


Figure S183. Reverse phase HPLC profile of dC-5mer.

Note that the broad peak at 10 min is an artefact of the column.

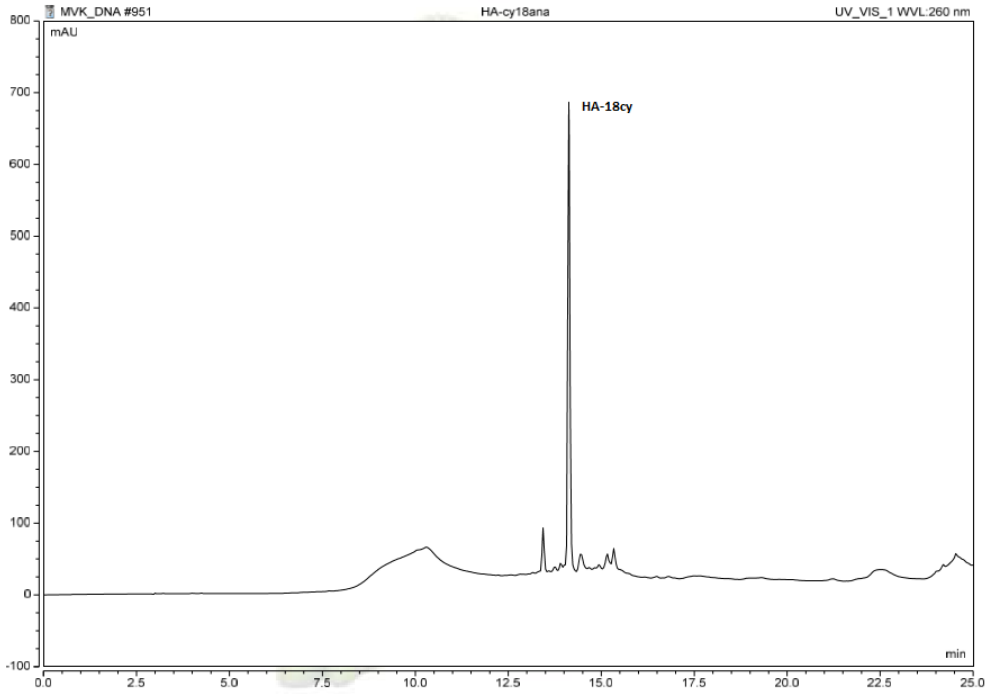


Figure S184. Reverse phase HPLC profile of dC-5mer-X.

Note that the broad peak at 10 min is an artefact of the column.

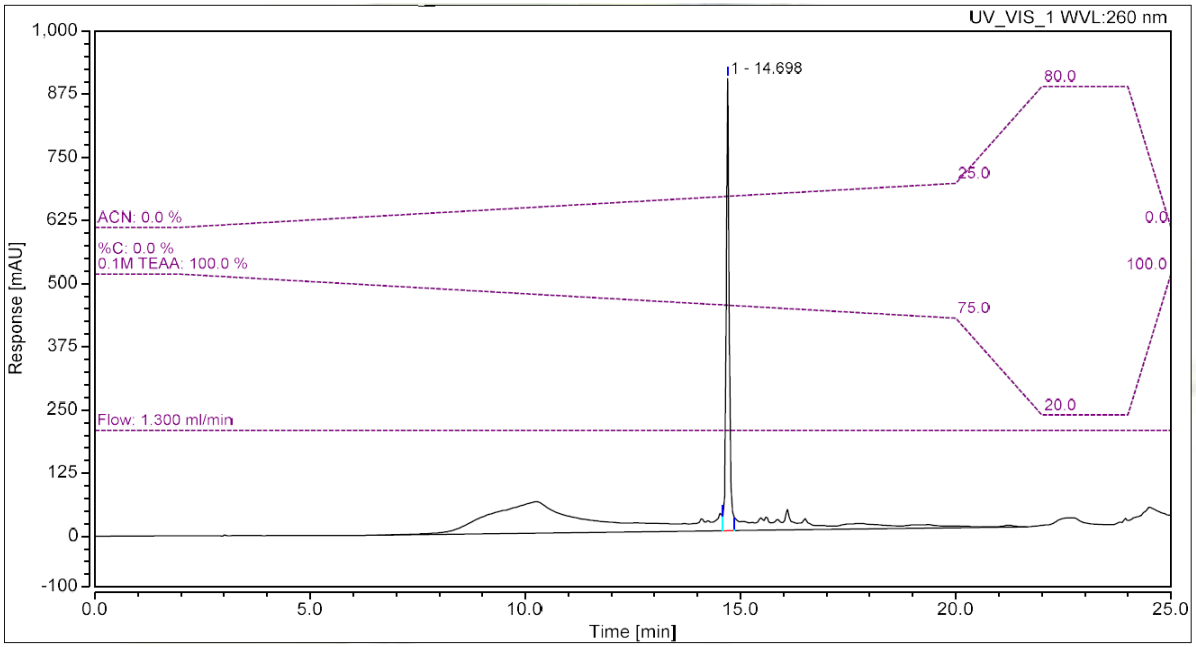


Figure S185. Reverse phase HPLC profile of dC[R^{N3}(-3),A^{Y2}(+1)]X.

Note that the broad peak at 10 min is an artefact of the column.

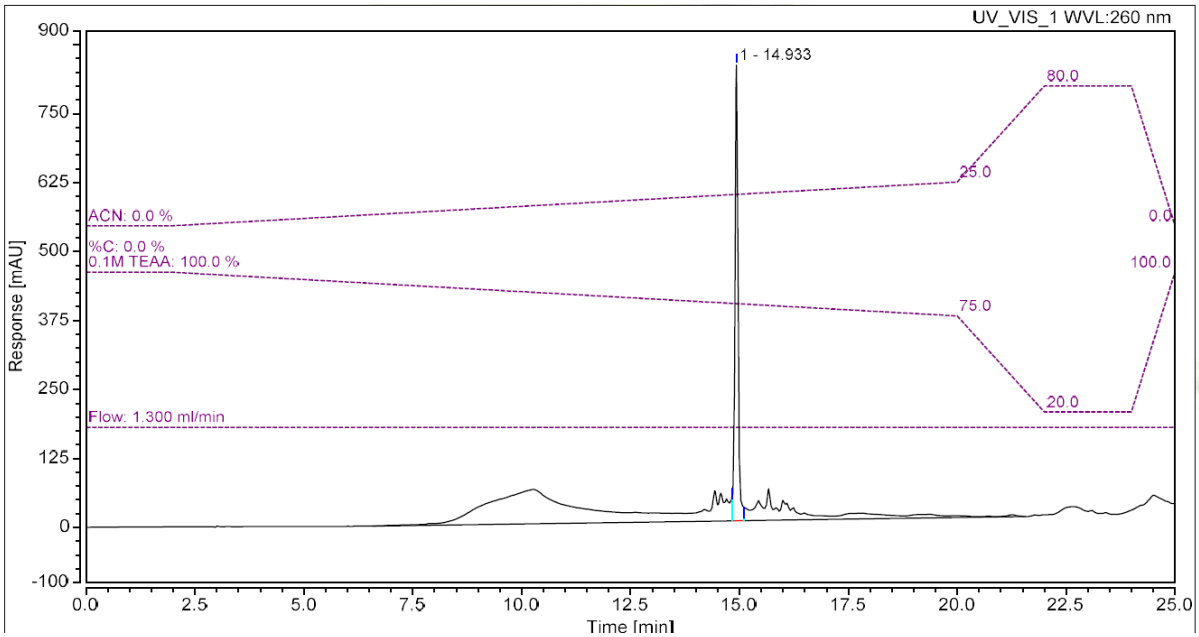


Figure S186. Reverse phase HPLC profile of dC[R^{N3}(-2),A^{Y2}(+1)]X.

Note that the broad peak at 10 min is an artefact of the column.

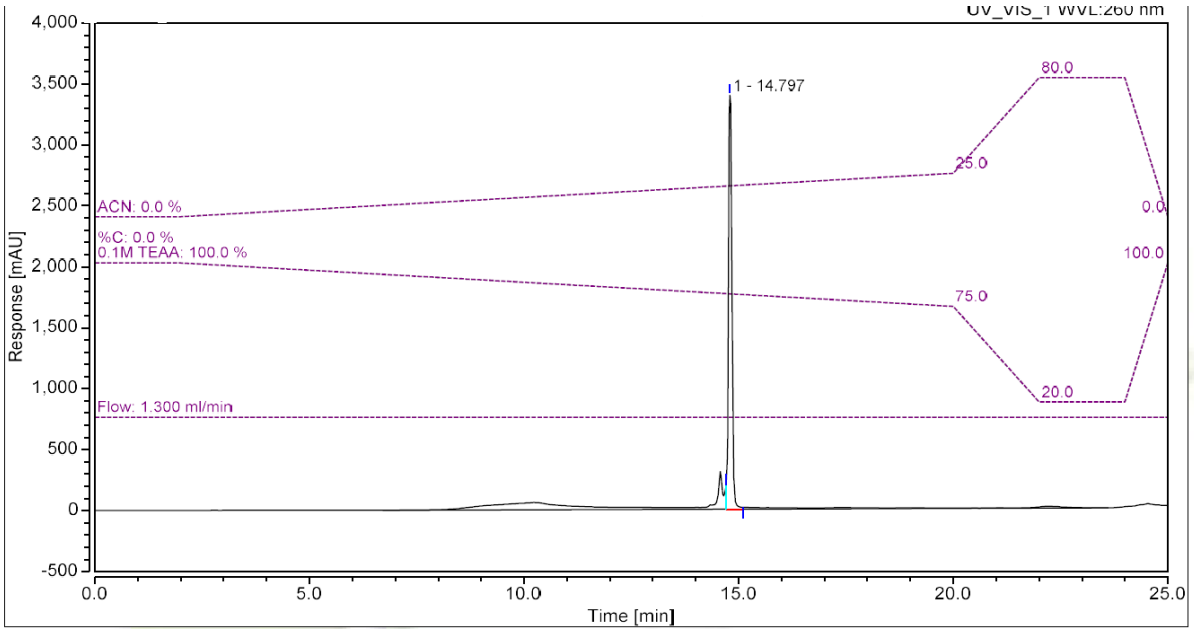


Figure S187. Reverse phase HPLC profile of dC[R^{N3}(-1),A^{Y2}(+1)]X.

Note that the broad peak at 10 min is an artefact of the column.

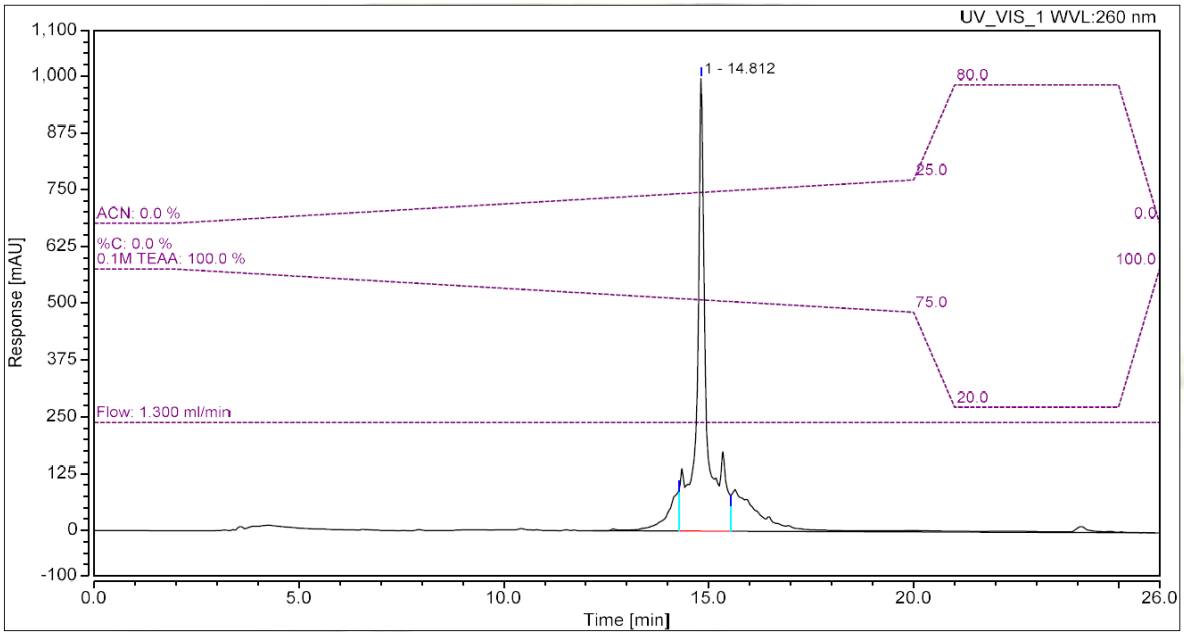


Figure S188. Reverse phase HPLC profile of dZ[R^{N3}(-3),A^{Y2}(+1)]X.

Note that the broad peak at 4 min is an artefact of the column.

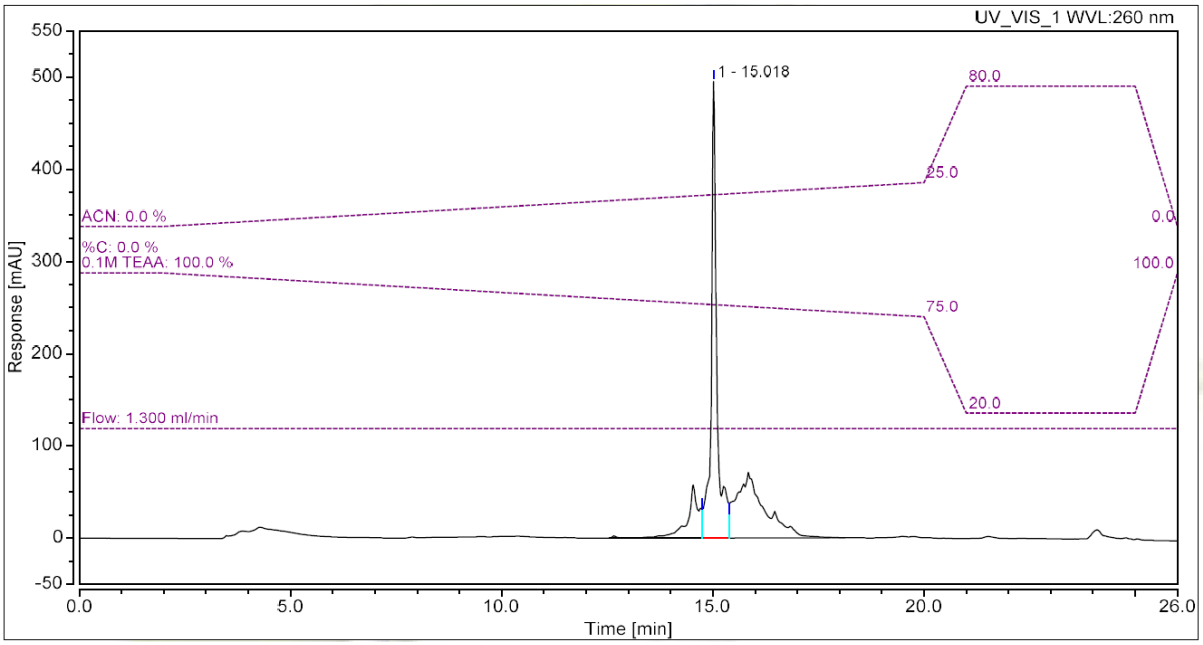


Figure S189. Reverse phase HPLC profile of dZ[R^{N3}(-2),A^{Y2}(+1)]X.

Note that the broad peak at 4 min is an artefact of the column.

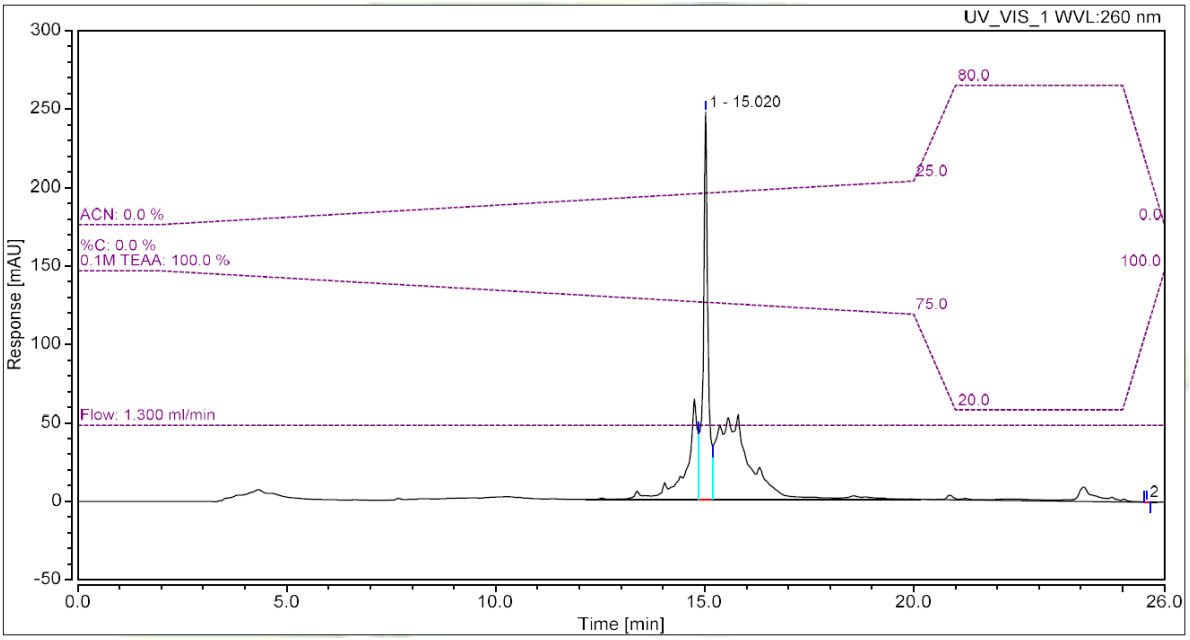


Figure S190. Reverse phase HPLC profile of dZ[R^{N3}(-1),A^{Y2}(+1)]X.

Note that the broad peak at 4 min is an artefact of the column.

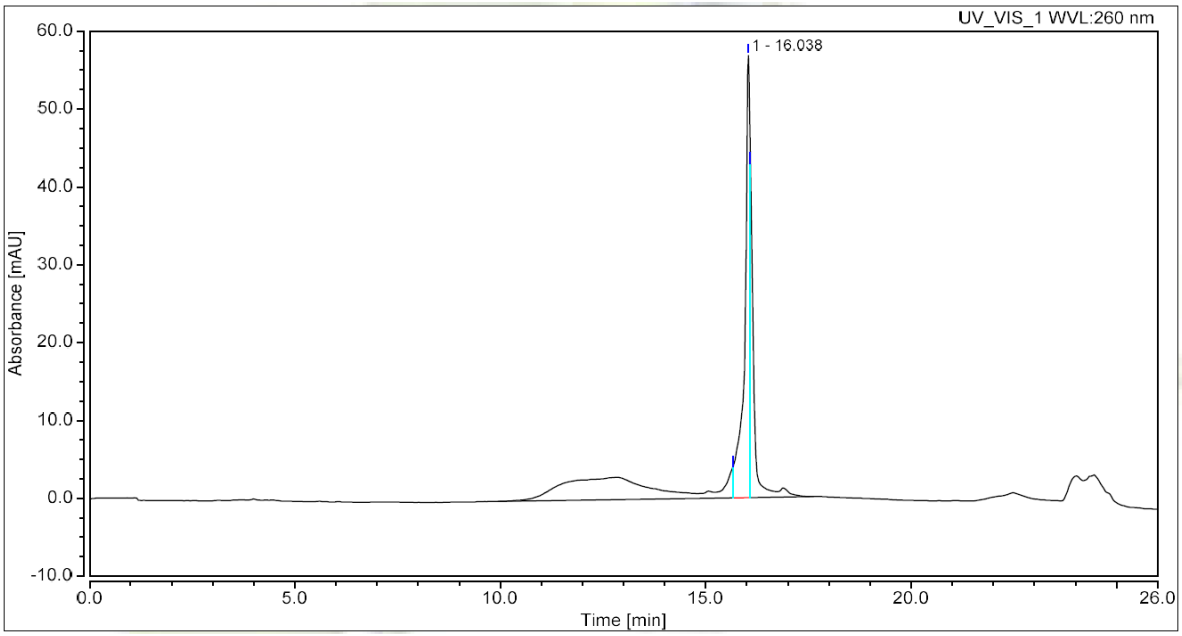


Figure S191. Reverse phase HPLC profile of dC[R^{N3}(-2),A^{Y4}(+1)]X.

Note that the broad peak at 12 min is an artefact of the column.

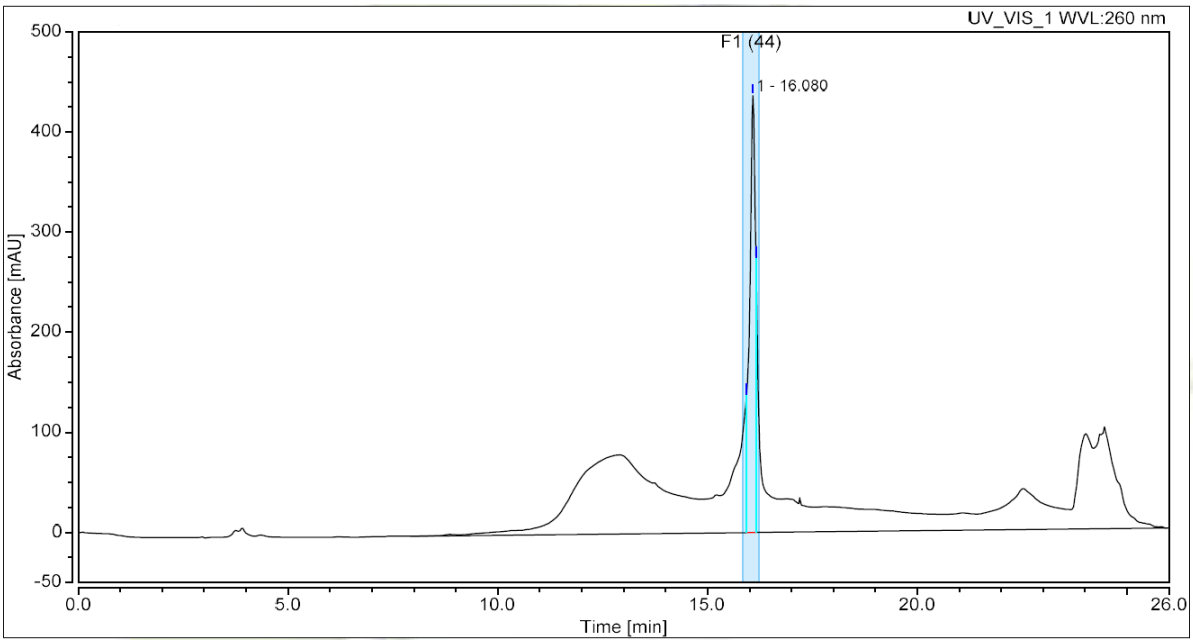


Figure S192. Reverse phase HPLC profile of dZ[R^{N3}(-2),A^{Y4}(+1)]X.

Note that the broad peak at 12 min is an artefact of the column.

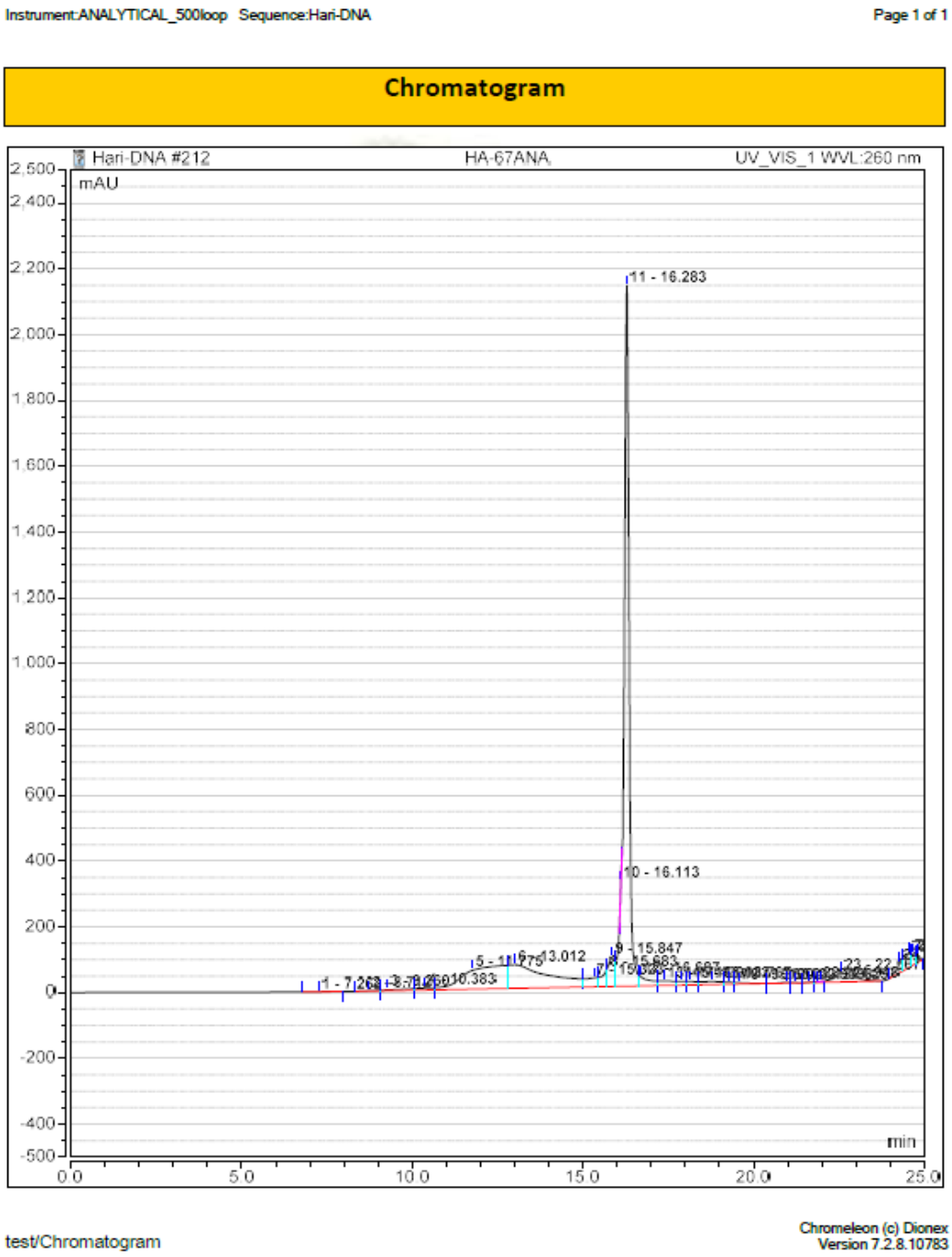


Figure S193. Reverse phase HPLC profile of dC[U^{E}(-2),A^{N3}(+1)].

Note that the broad peak at 12 min is an artefact of the column.

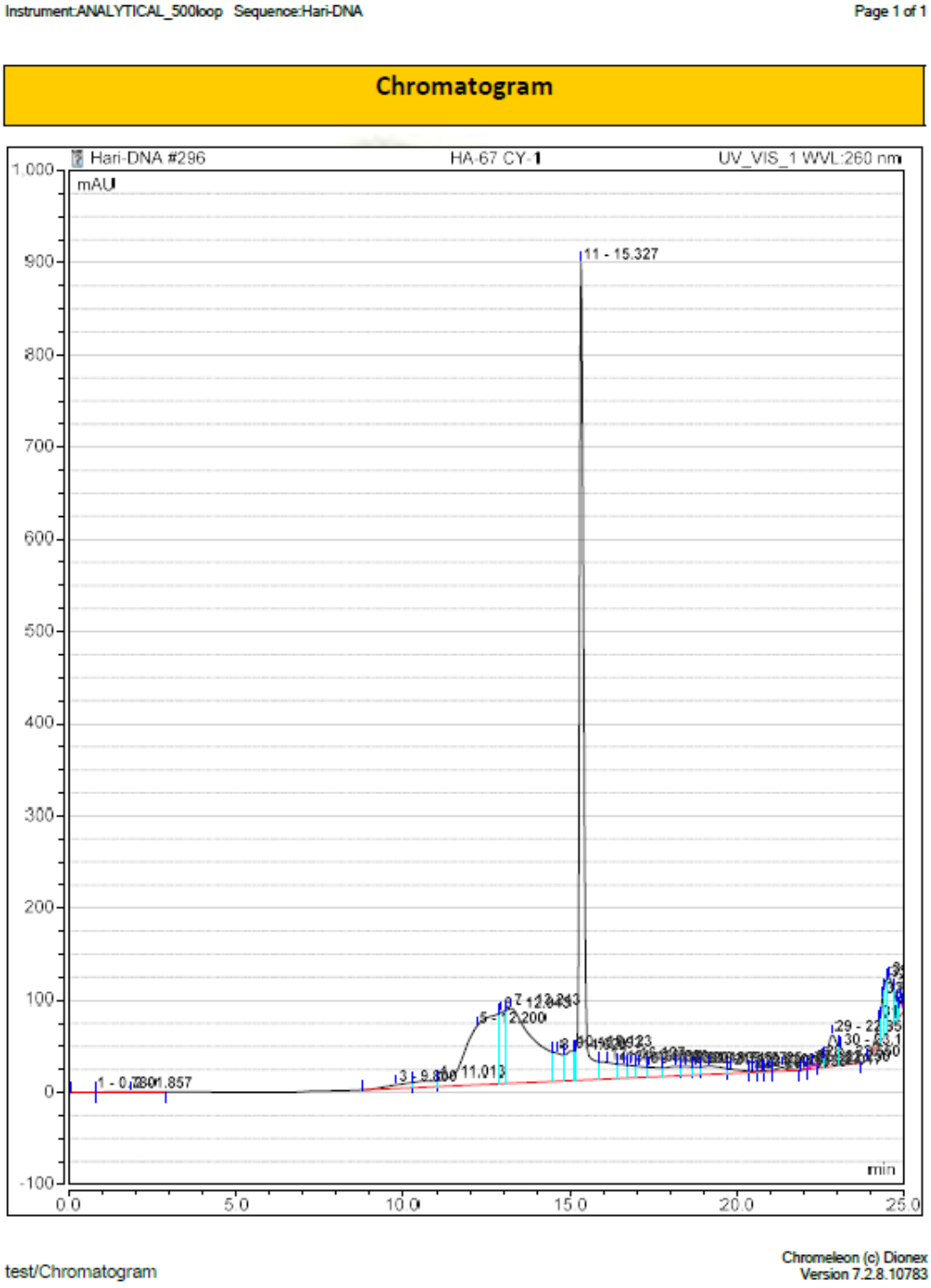


Figure S194. Reverse phase HPLC profile of dC[U^{E}(-2),A^{N3}(+1)]X.

Note that the broad peak at 12 min is an artefact of the column.

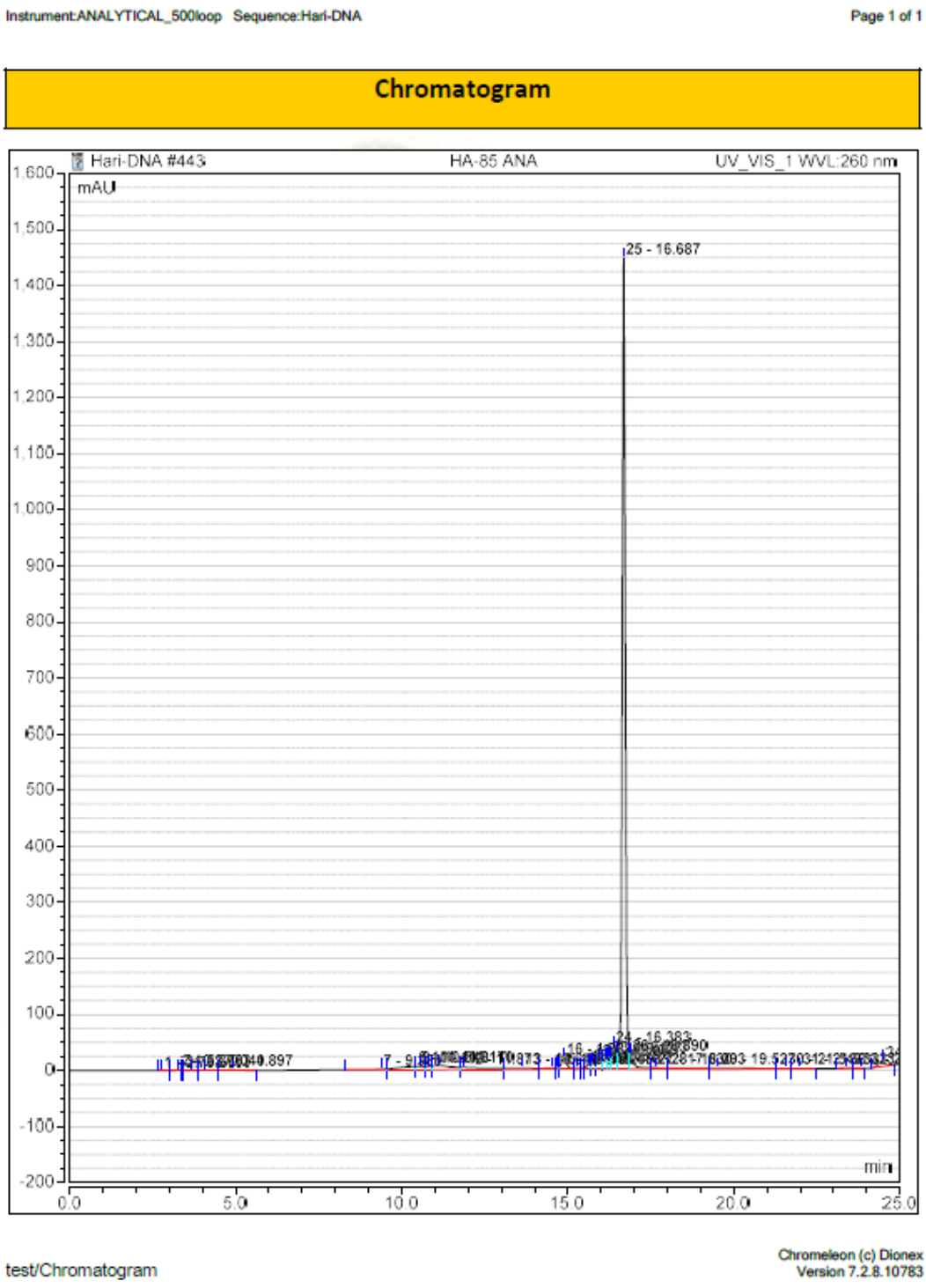


Figure S195. Reverse phase HPLC profile of dC[U^{E}(-3),A^{N3}(+1)].

Note that the broad peak at 12 min is an artefact of the column.

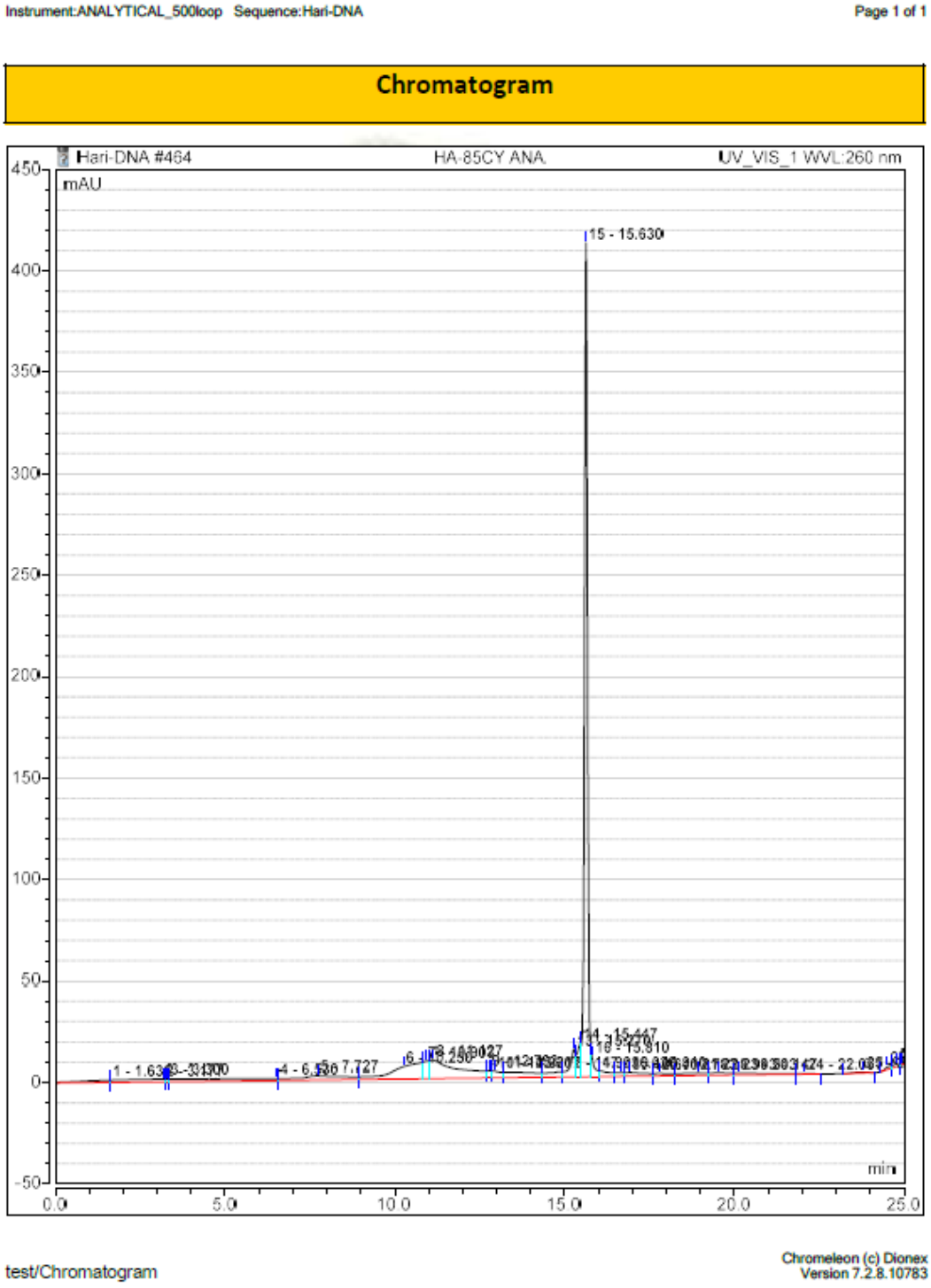


Figure S196. Reverse phase HPLC profile of dC[U^{E}(-3),A^{N3}(+1)]X.

Note that the broad peak at 12 min is an artefact of the column.

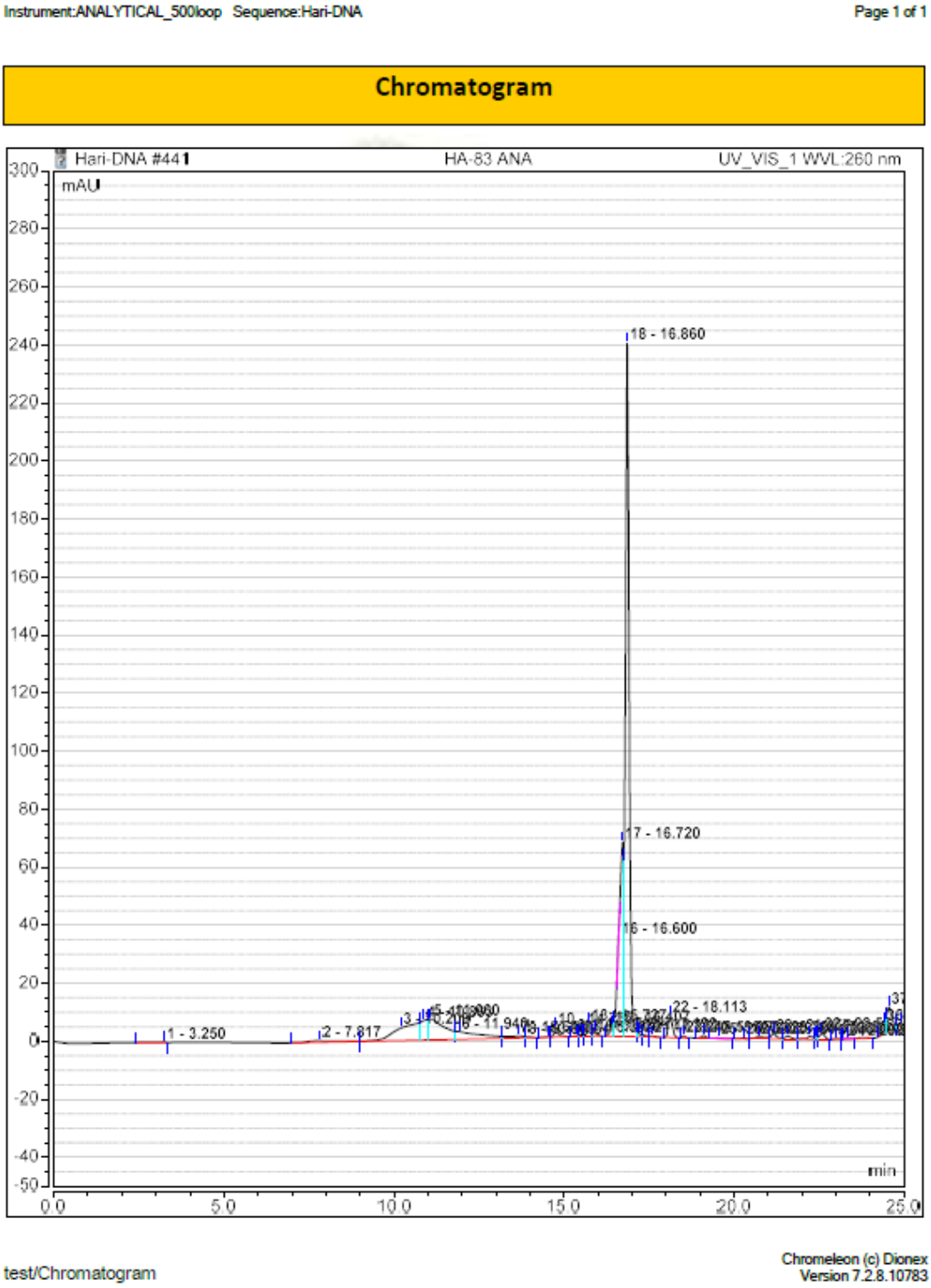


Figure S197. Reverse phase HPLC profile of dZ[U^{E}(-3),A^{N3}(+1)].

Note that the broad peak at 11 min is an artefact of the column.

Figure S198. Reverse phase HPLC profile of dZ[U^{E}(-3),A^{N3}(+1)]X.

Note that the broad peak at 11 min is an artefact of the column.

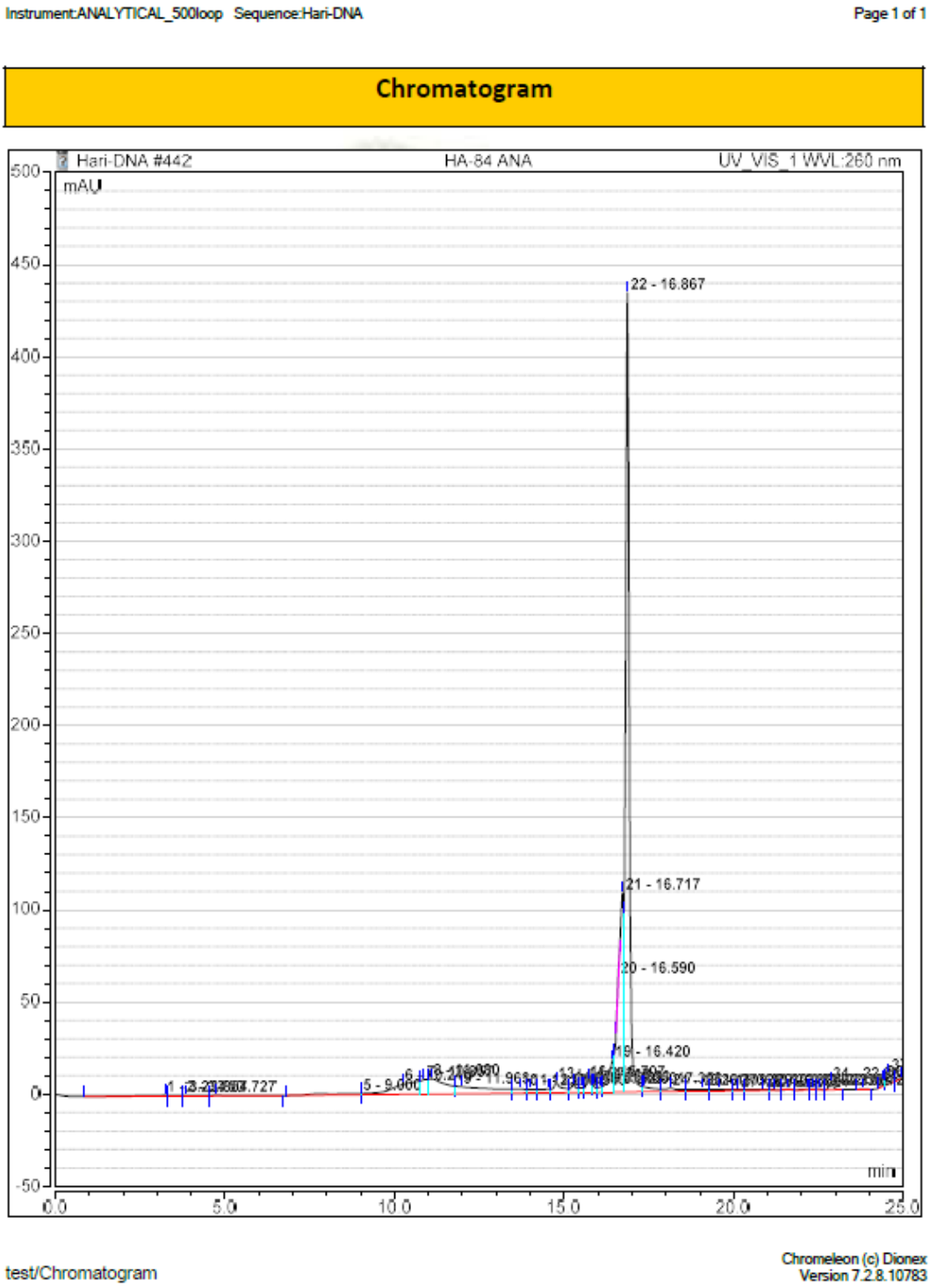


Figure S199. Reverse phase HPLC profile of dZ[U^{E}(-2),A^{N3}(+1)].

Note that the broad peak at 11 min is an artefact of the column.

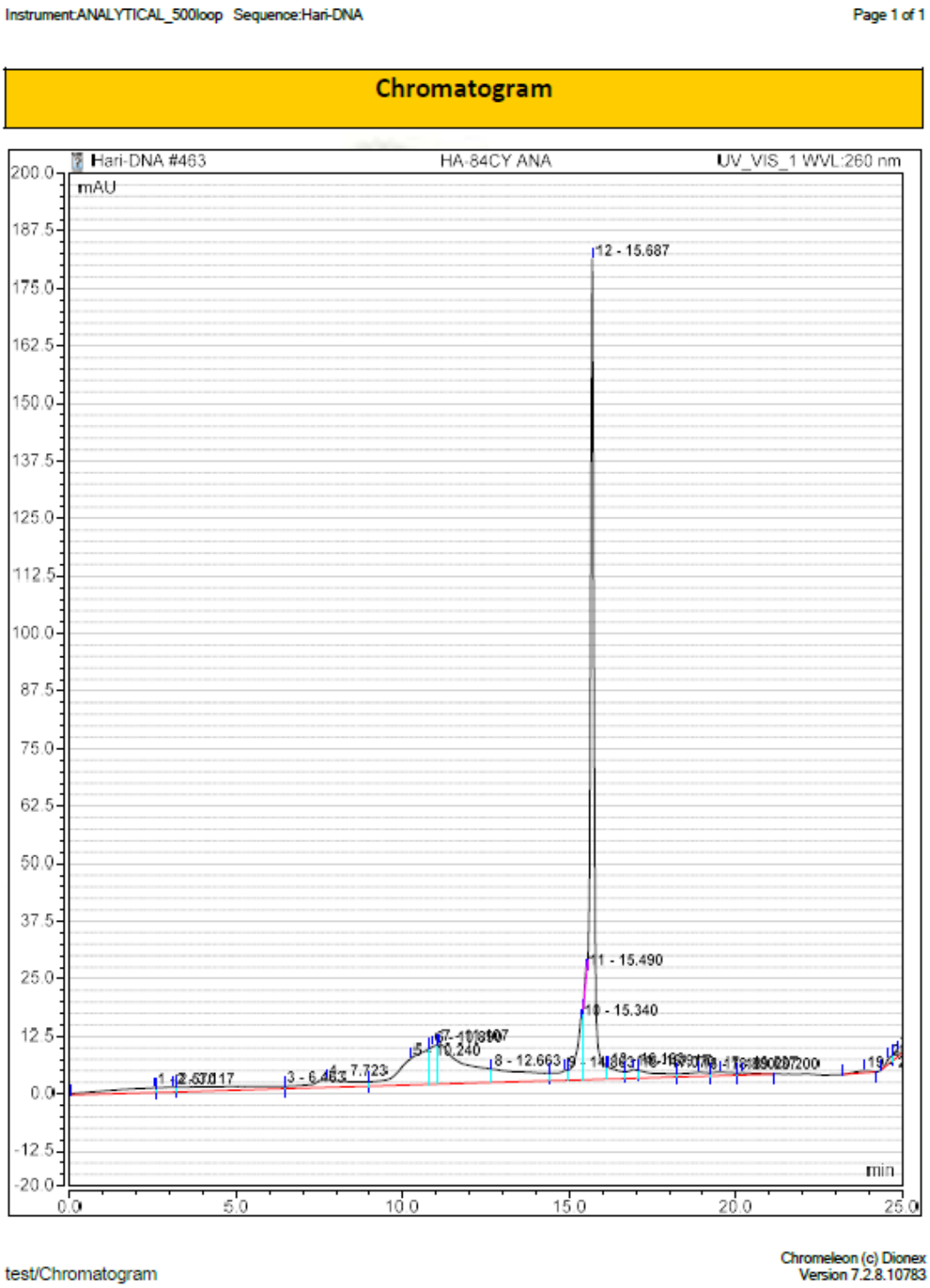


Figure S200. Reverse phase HPLC profile of dZ[U^{E}(-2),A^{N3}(+1)]X.

Note that the broad peak at 11 min is an artefact of the column.

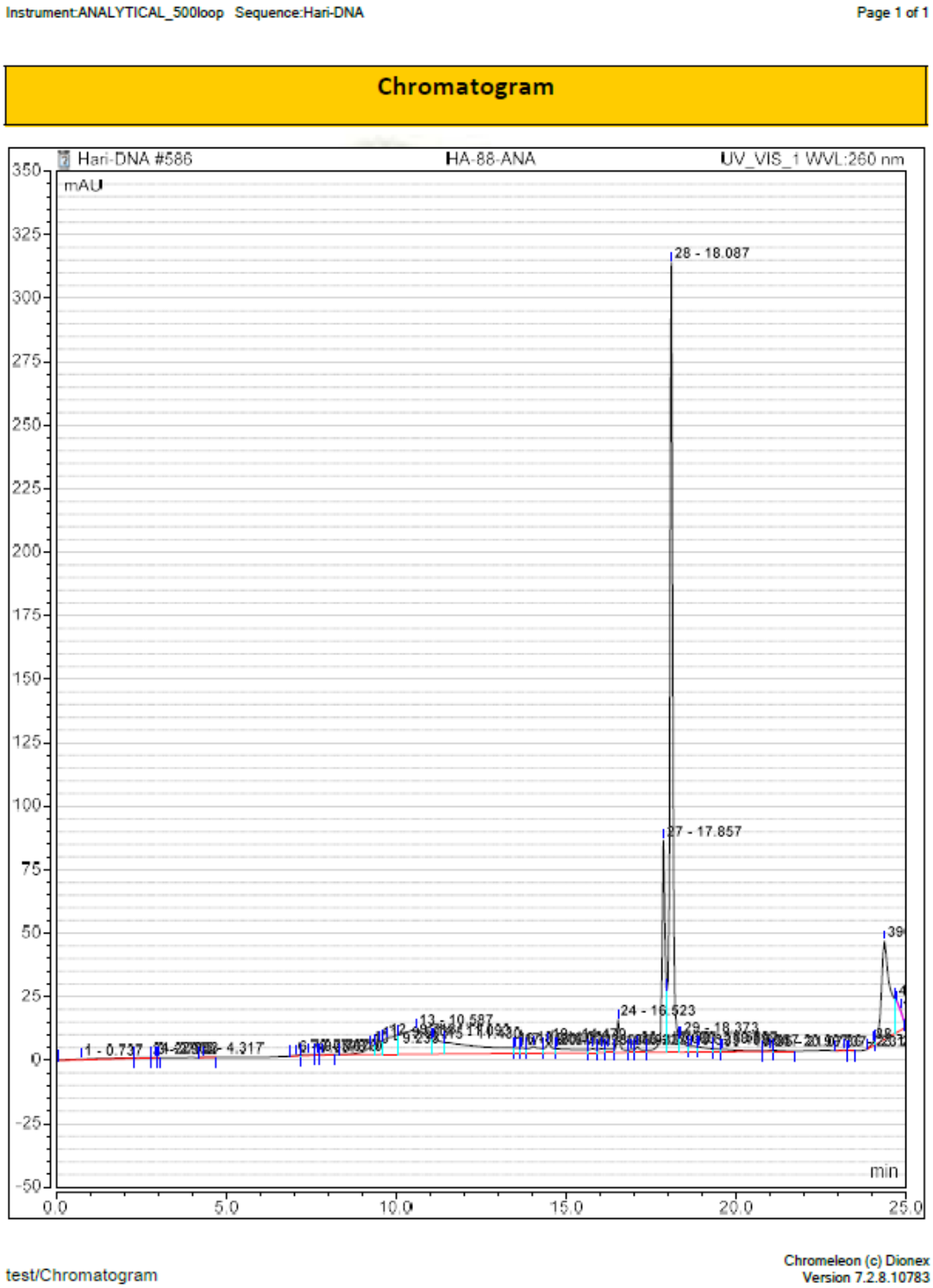


Figure S201. Reverse phase HPLC profile of dC[C^{N3}(-2),H^{E}(+1)].

Note that the broad peaks at 10 and 24.5 min are artefacts of the column.

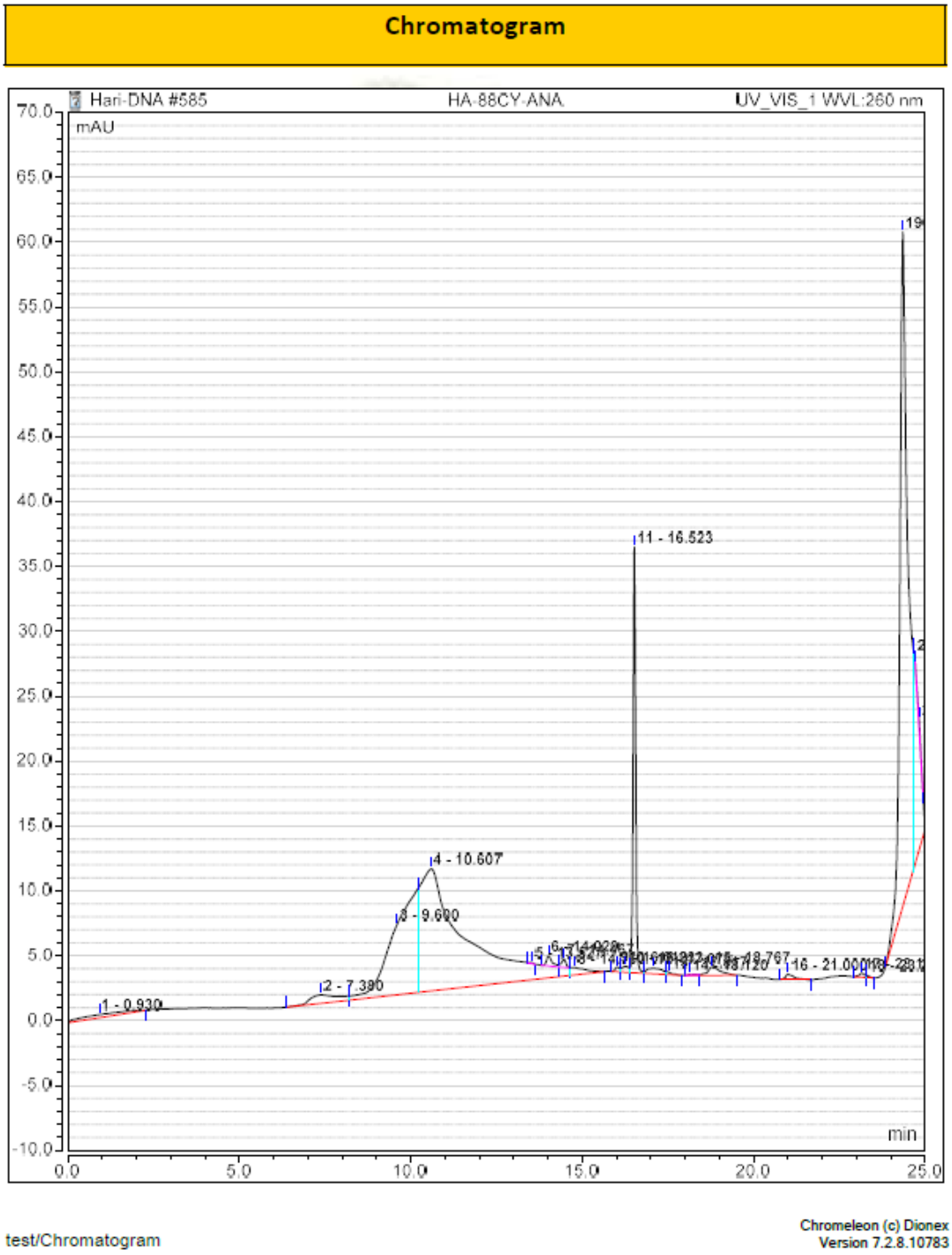


Figure S202. Reverse phase HPLC profile of dC[C^{N3}(-2),H^{E}(+1)]X.

Note that the broad peaks at 10 and 24.5 min are artefacts of the column.

Figure S203. Reverse phase HPLC profile of dZ[C^{N3}(-2),H^{E}(+1)].

Note that the broad peaks at 10 and 24.5 min are artefacts of the column.

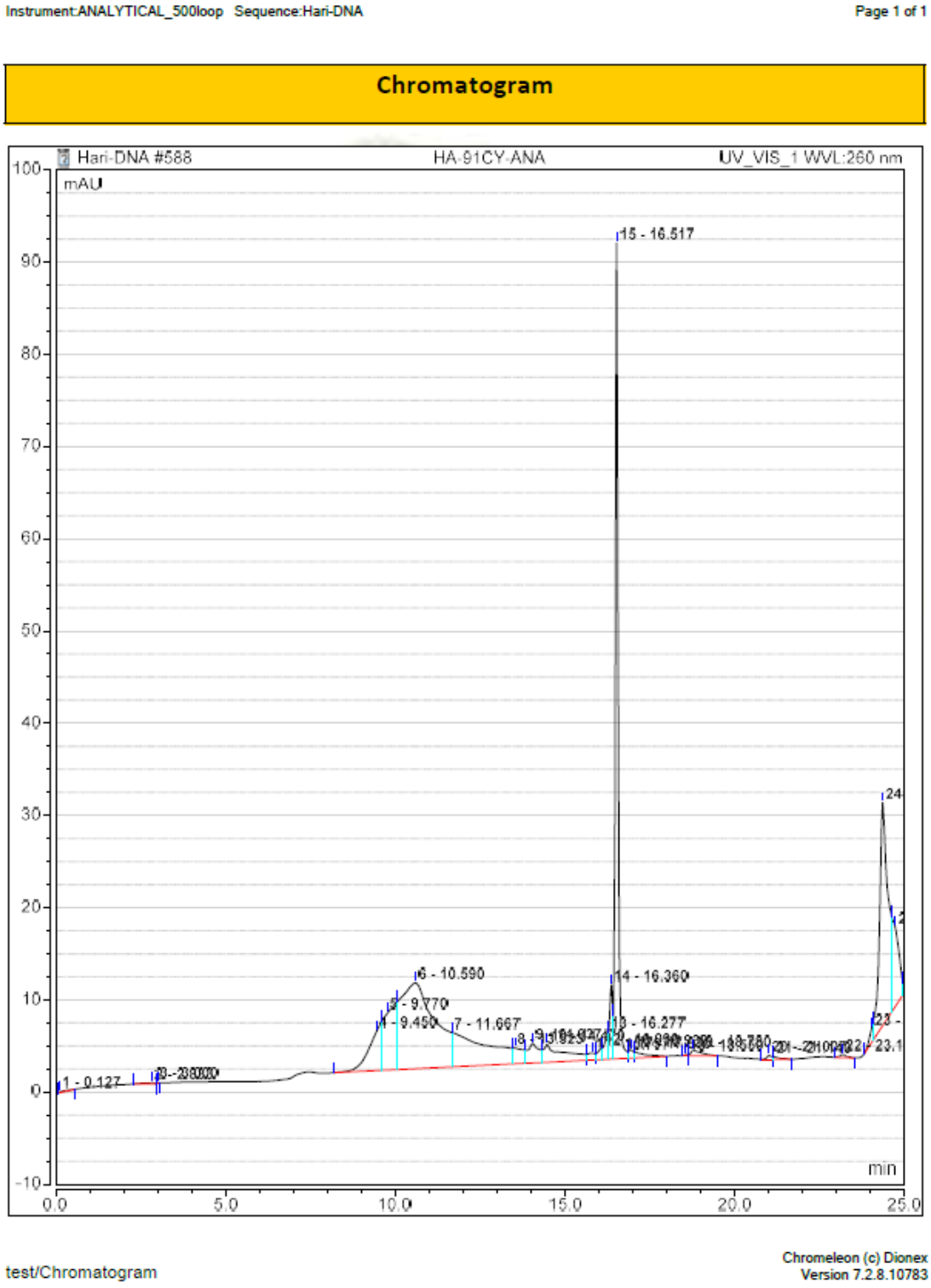


Figure S204. Reverse phase HPLC profile of dZ[C^{N3}(-2),H^{E}(+1)]X.

Note that the broad peaks at 10 and 24.5 min are artefacts of the column.

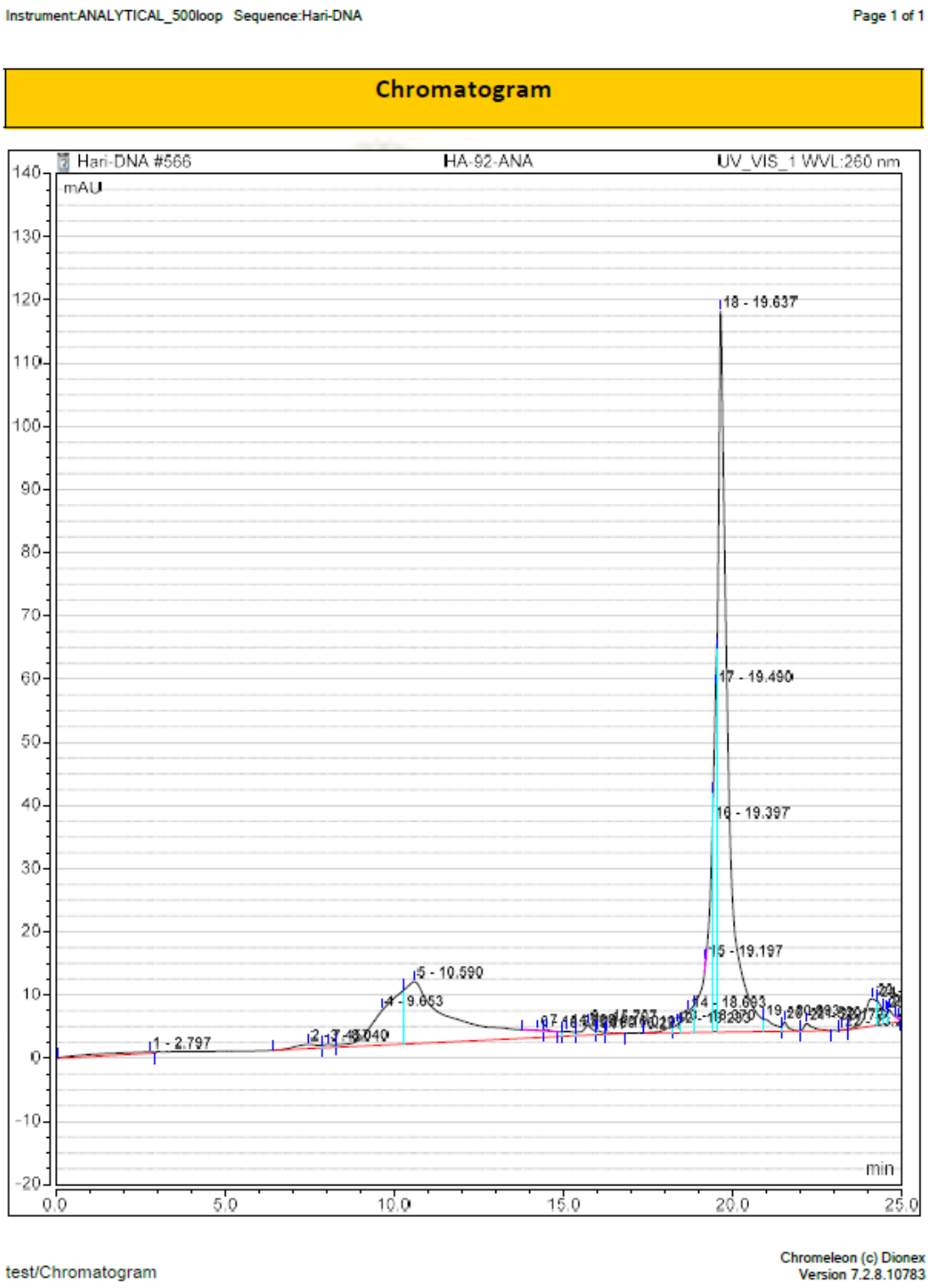


Figure S205. Reverse phase HPLC profile of FdZ[C^{N3}(-2),H^{E}(+1)].

Note that the broad peaks at 10 and 24.5 min are artefacts of the column.

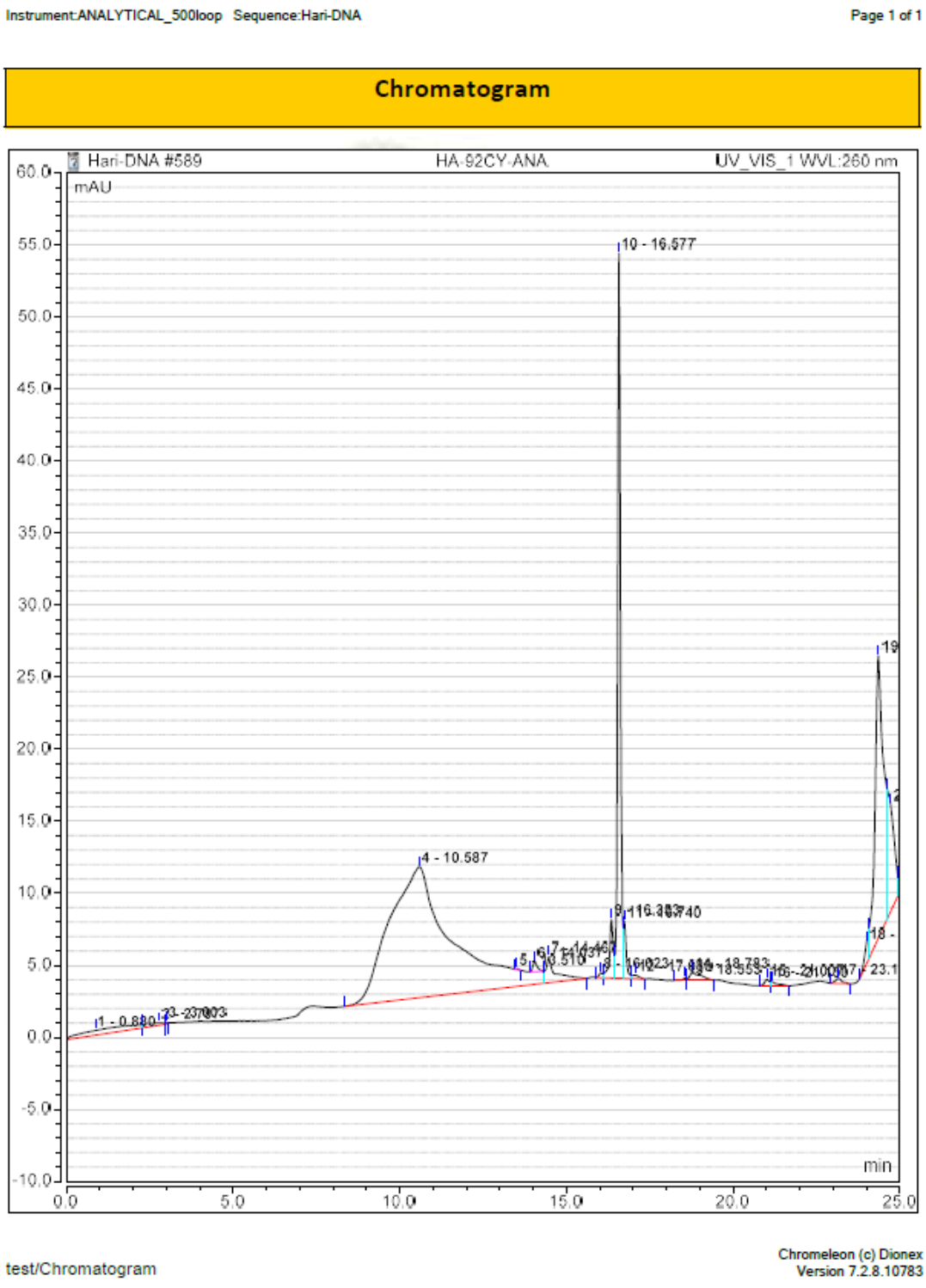


Figure S206. Reverse phase HPLC profile of FdZ[C^{N3}(-2),H^{E}(+1)]X.

Note that the broad peaks at 10 and 24.5 min are artefacts of the column.

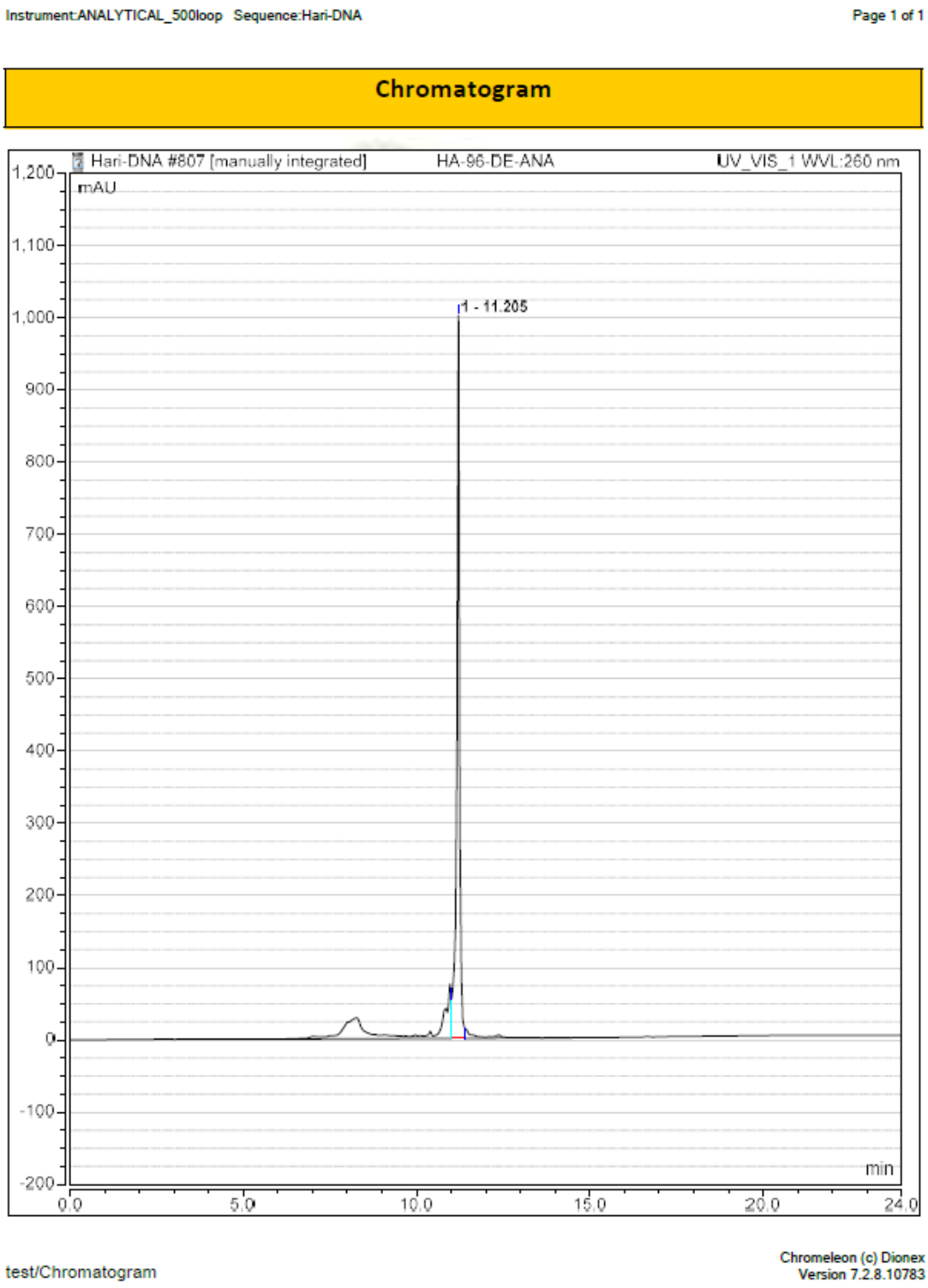


Figure S207. Reverse phase HPLC profile of 4-mer dZ[C^{N3}(-2),H^{E}(+1)].

Note that the broad peak at 8 min is an artefact of the column.

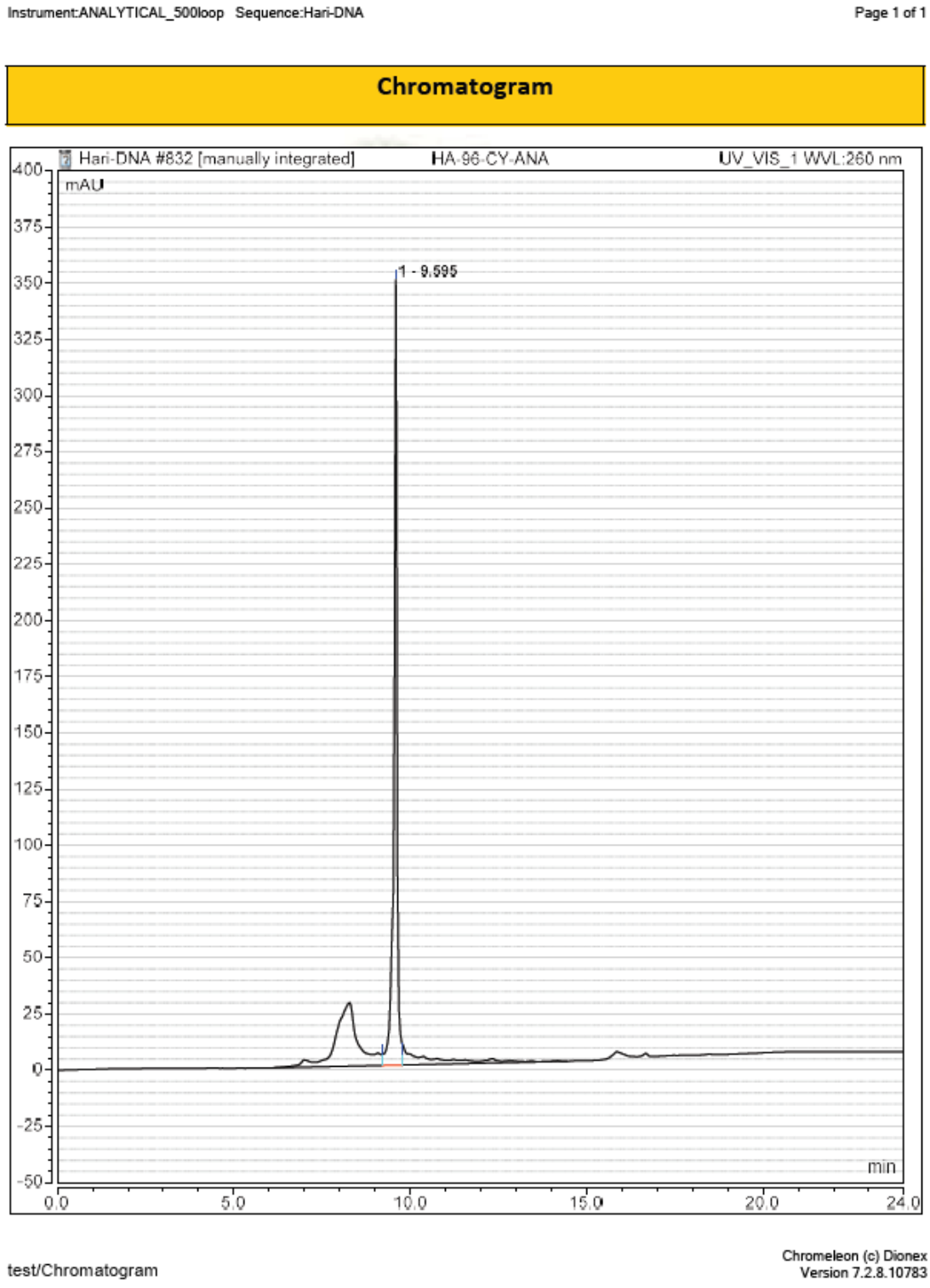


Figure S208. Reverse phase HPLC profile of 4-mer dZ[C^{N3}(-2),H^{E}(+1)]X.

Note that the broad peak at 8 min is an artefact of the column.

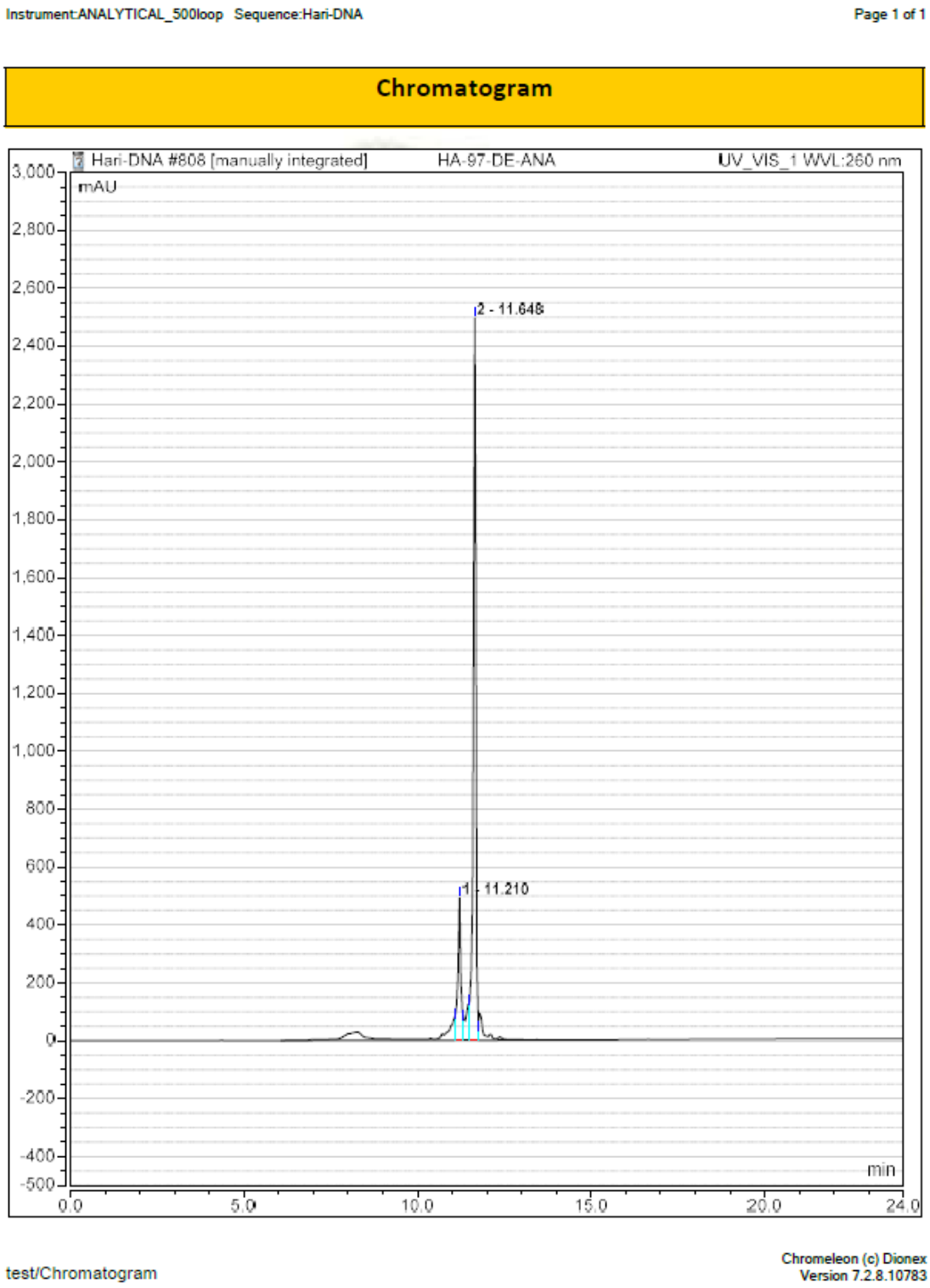


Figure S209. Reverse phase HPLC profile of 3-mer dZ[C^{N3}(-1),H^{E}(+1)].

Note that the broad peak at 8 min is an artefact of the column.

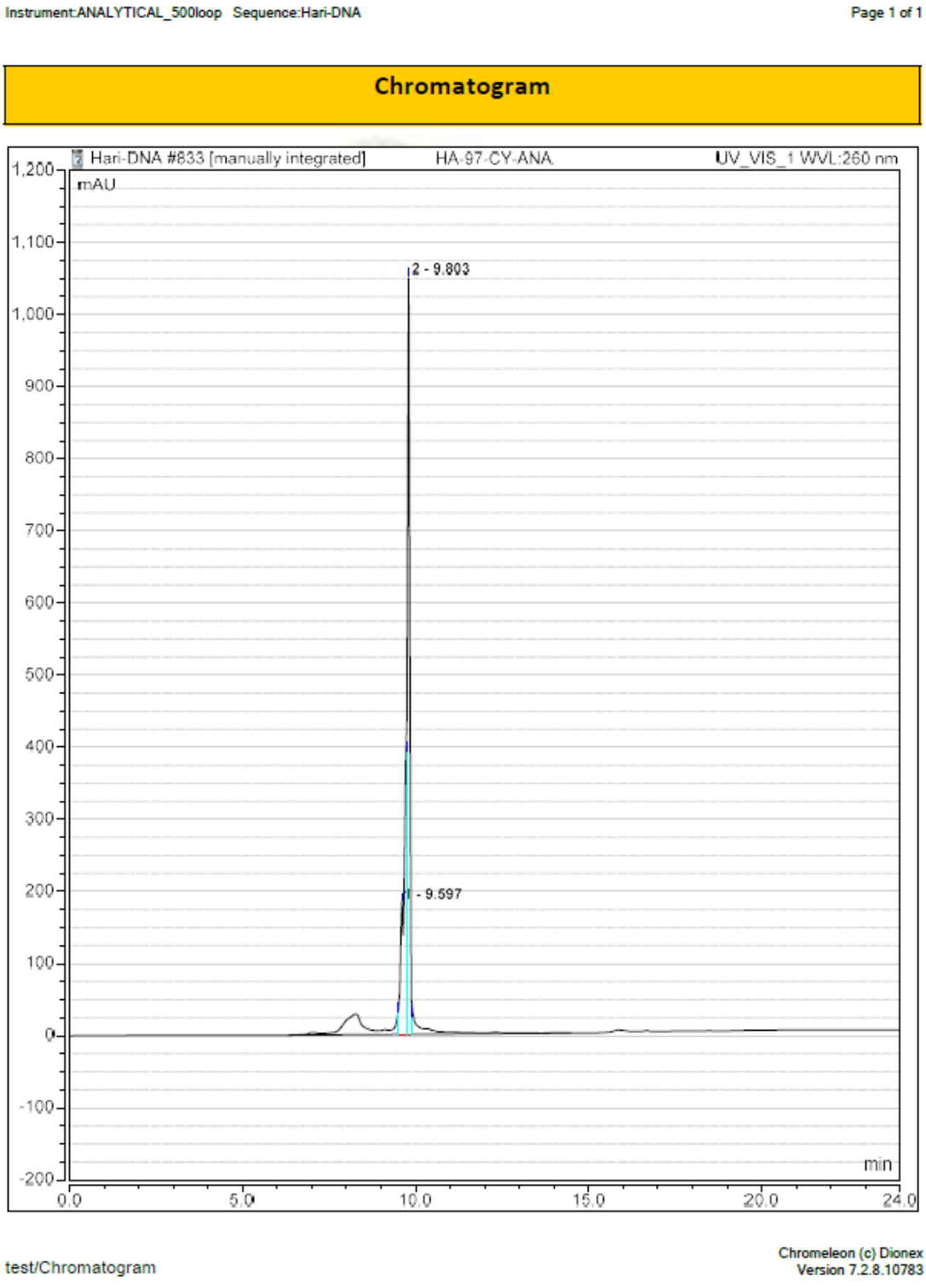


Figure S210. Reverse phase HPLC profile of 3-mer dZ[C^{N3}(-1),H^{E}(+1)]X.

Note that the broad peak at 8 min is an artefact of the column.

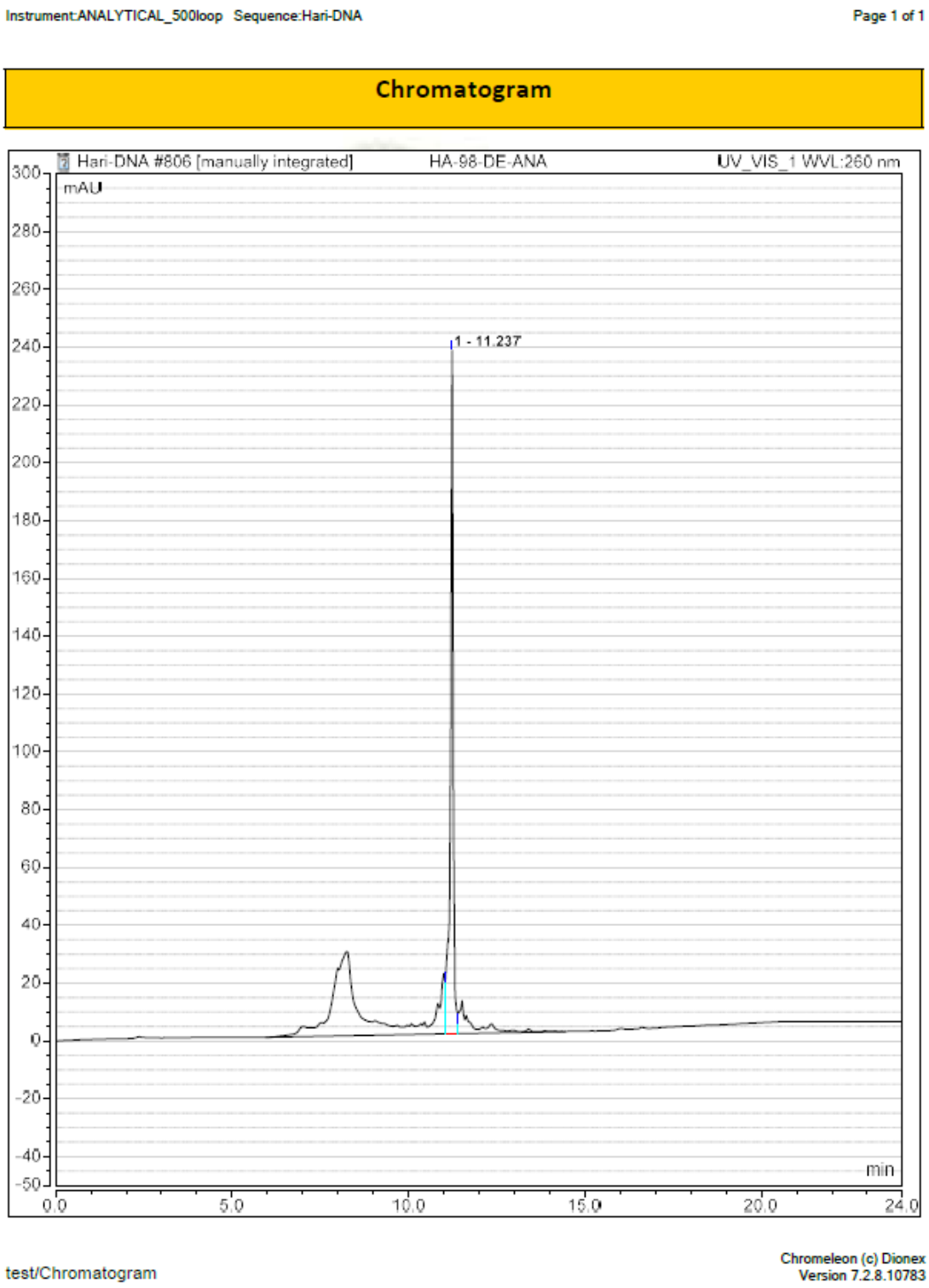


Figure S211. Reverse phase HPLC profile of 4-mer FdZ[C^{N3}(-2),H^{E}(+1)].

Note that the broad peak at 8 min is an artefact of the column.

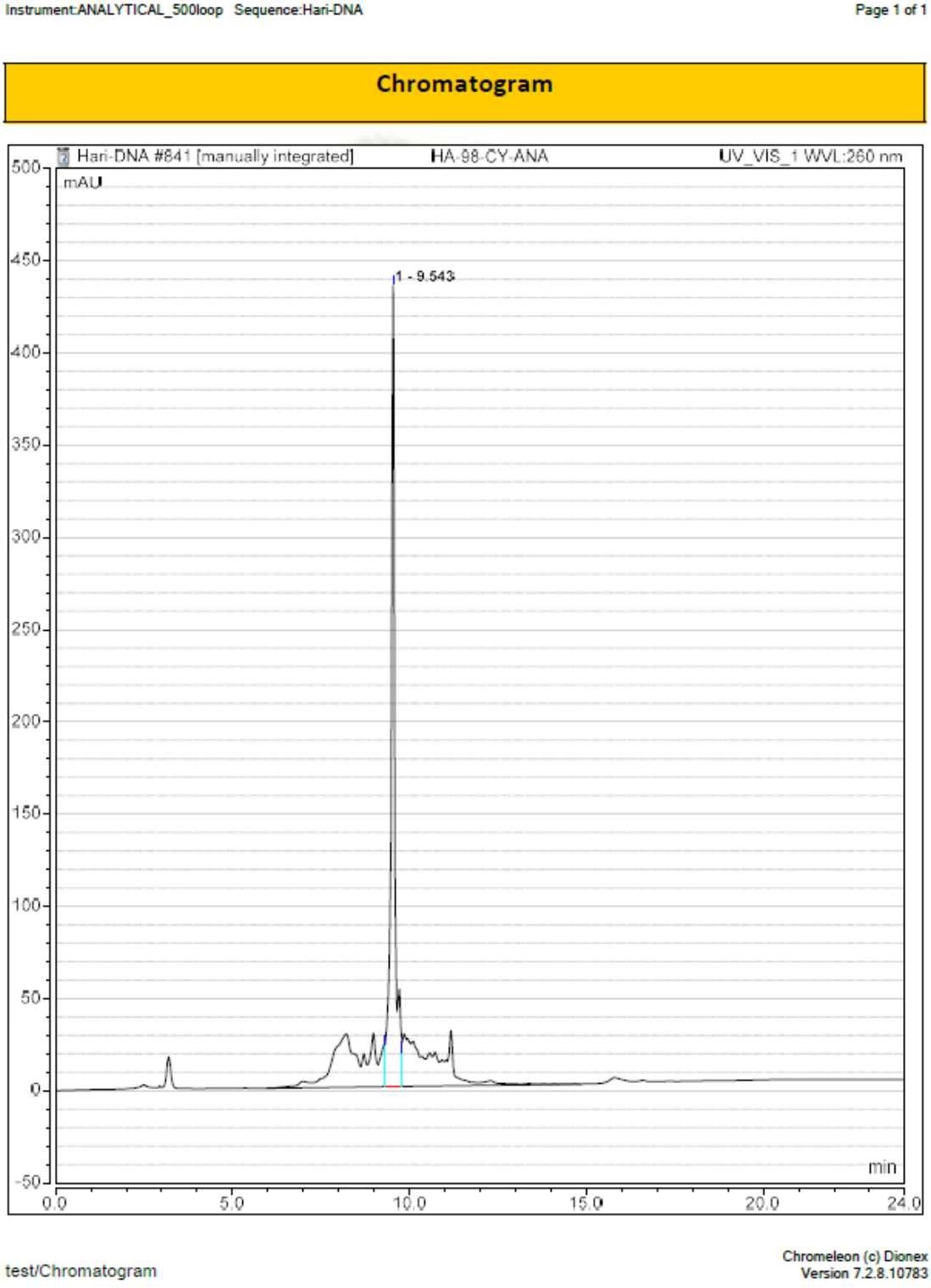


Figure S212. Reverse phase HPLC profile of 4-mer FdZ[C^{N3}(-2),H^{E}(+1)]X.

Note that the broad peak at 8 min is an artefact of the column.

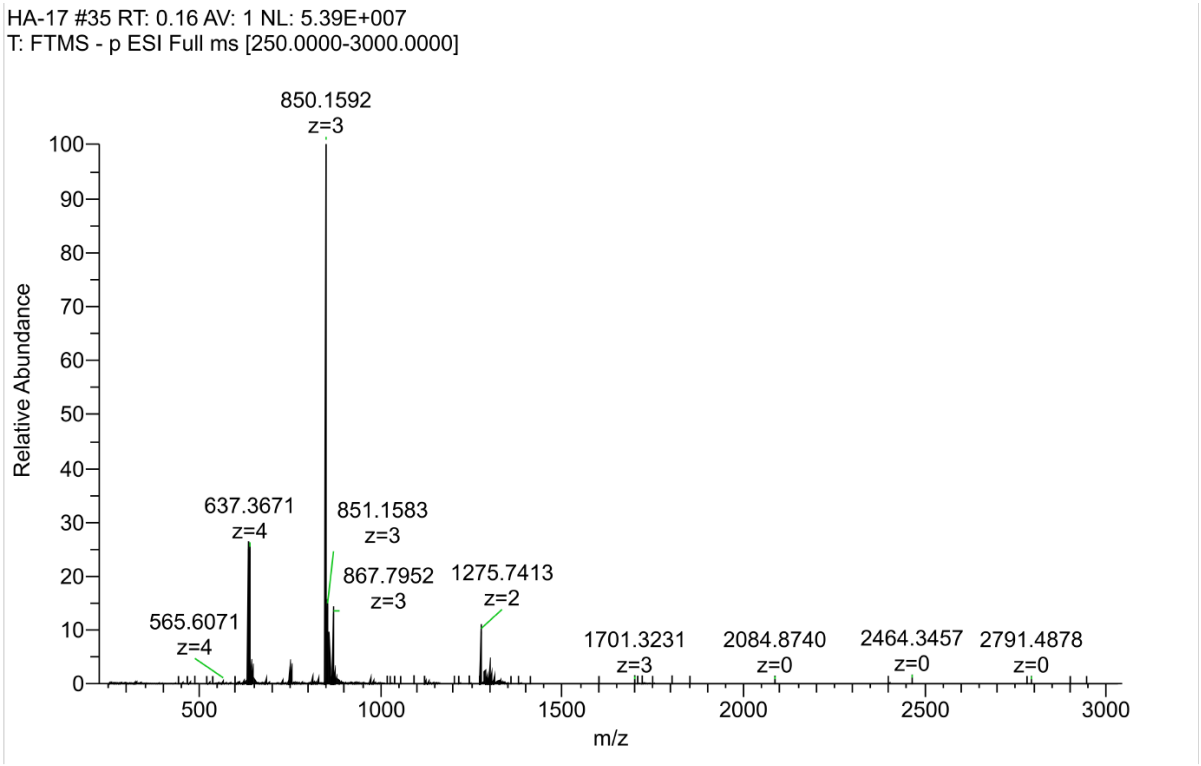


Figure S213. HRMS (ESI) of dC-9mer.

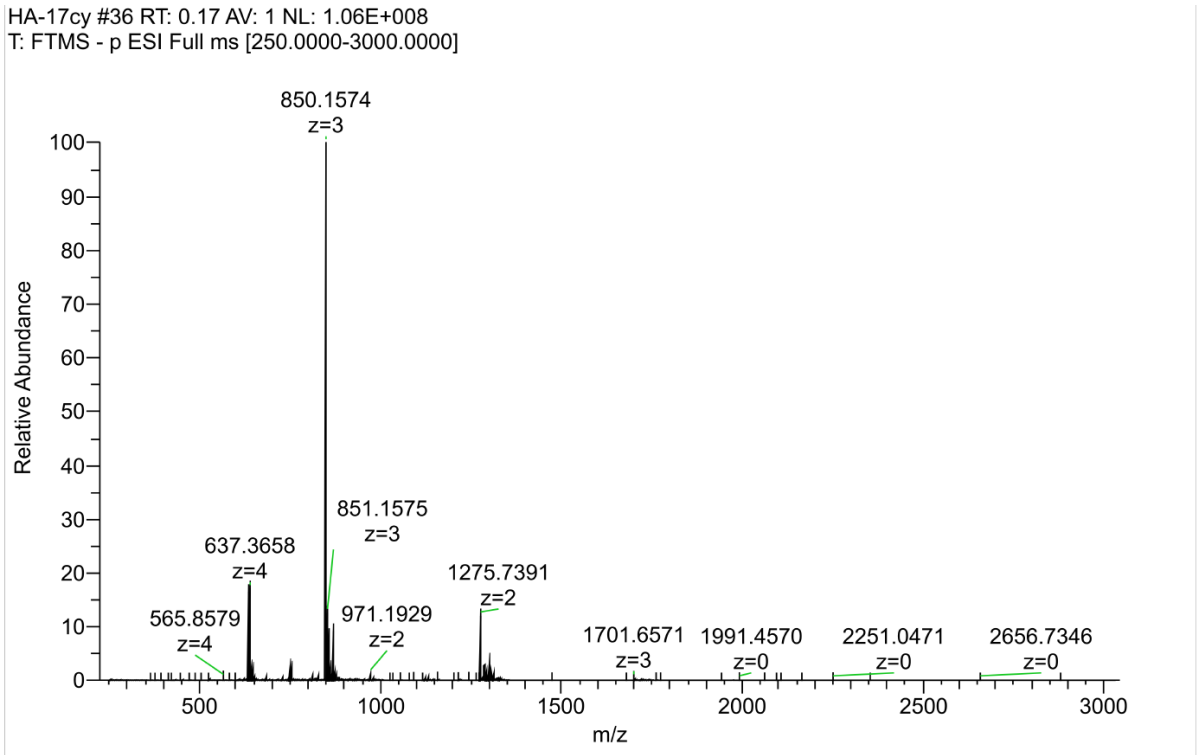


Figure S214. HRMS (ESI) of dC-9mer-X.

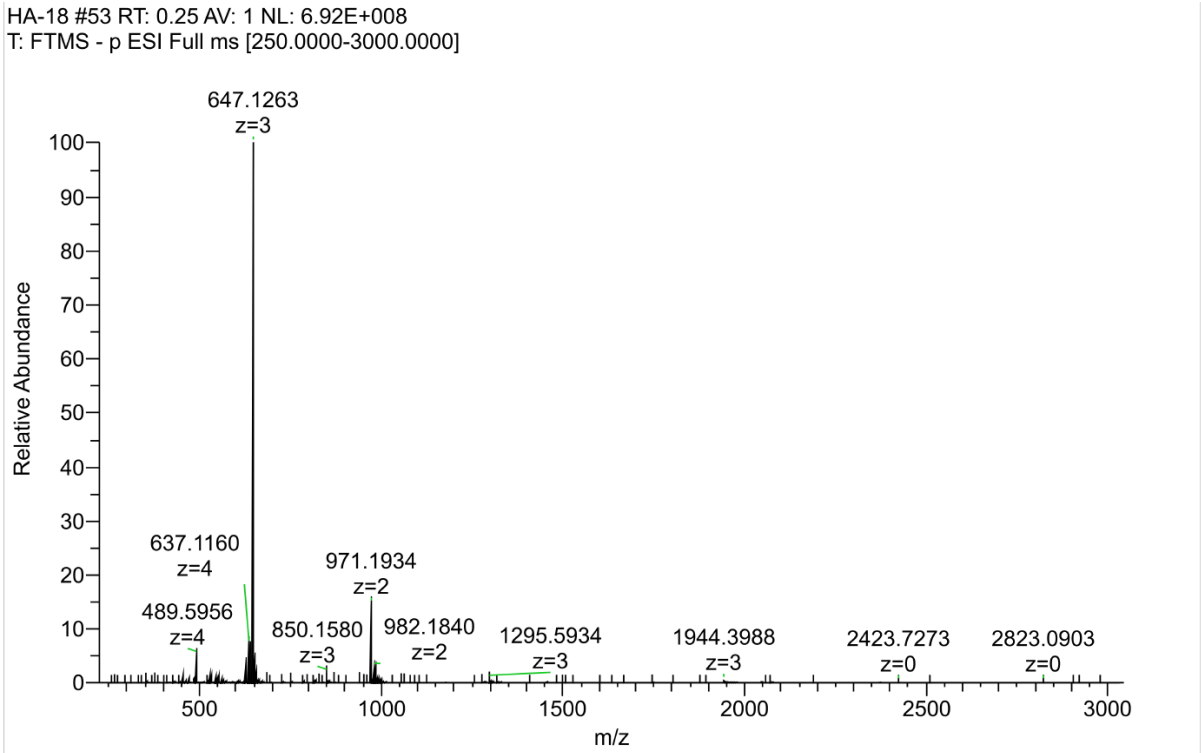


Figure S215. HRMS (ESI) of dC-7mer.

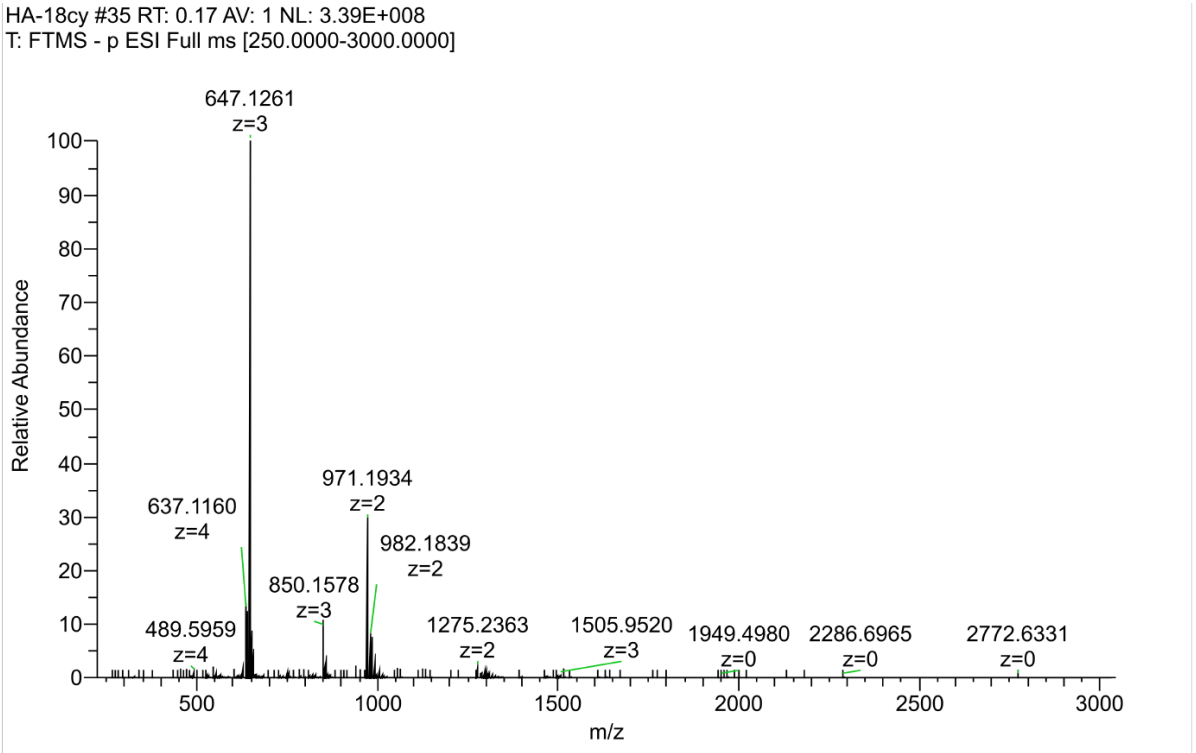


Figure S216. HRMS (ESI) of dC-7mer-X.

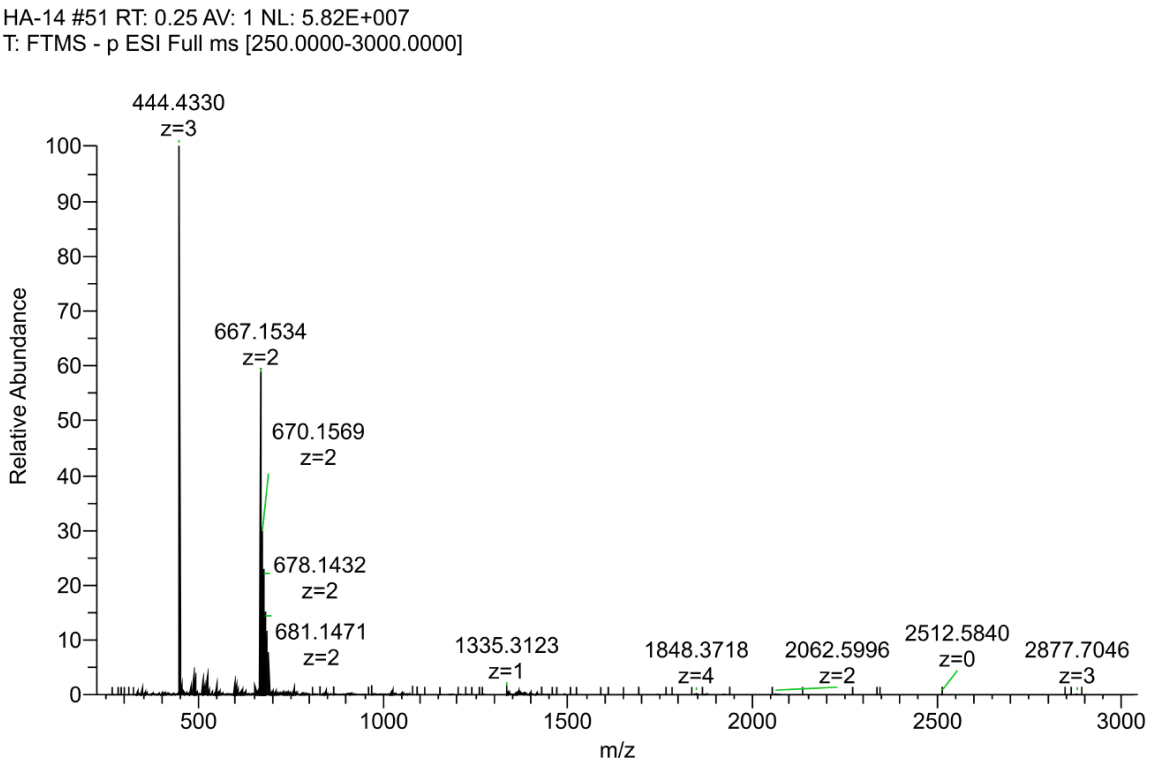


Figure S217. HRMS (ESI) of dC-5mer.

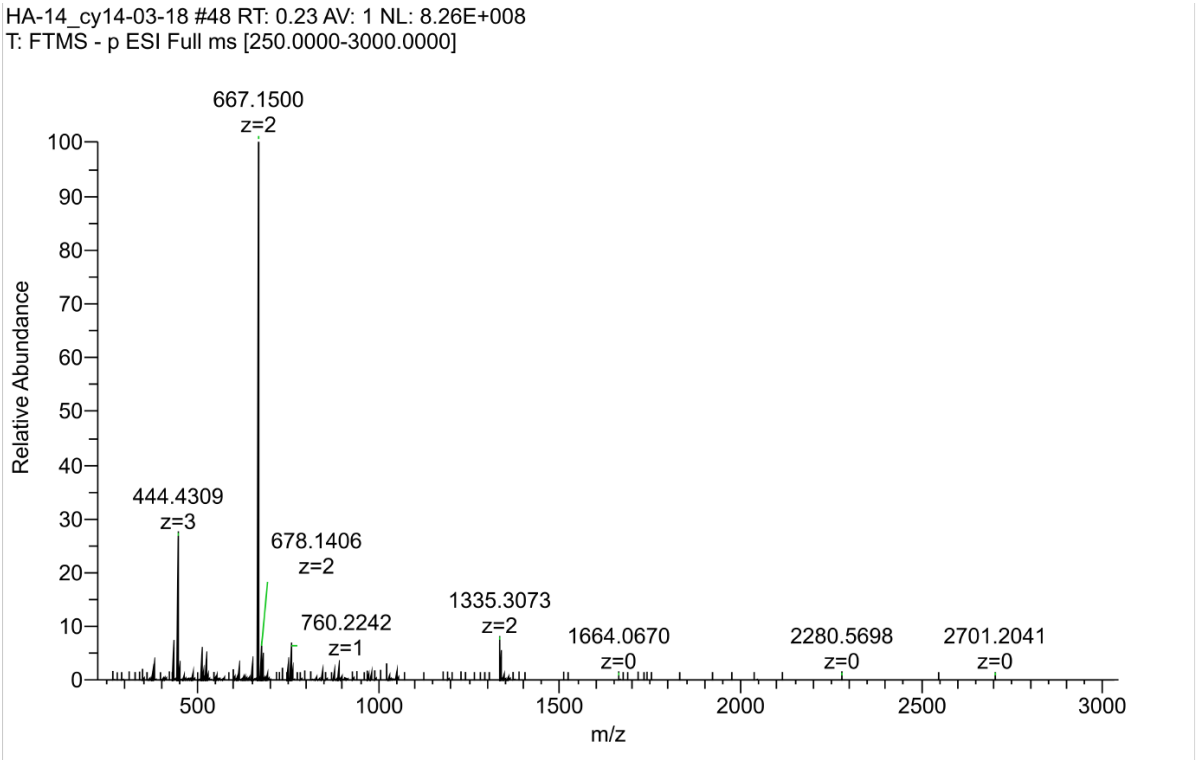


Figure S218. HRMS (ESI) of dC-5mer-X.

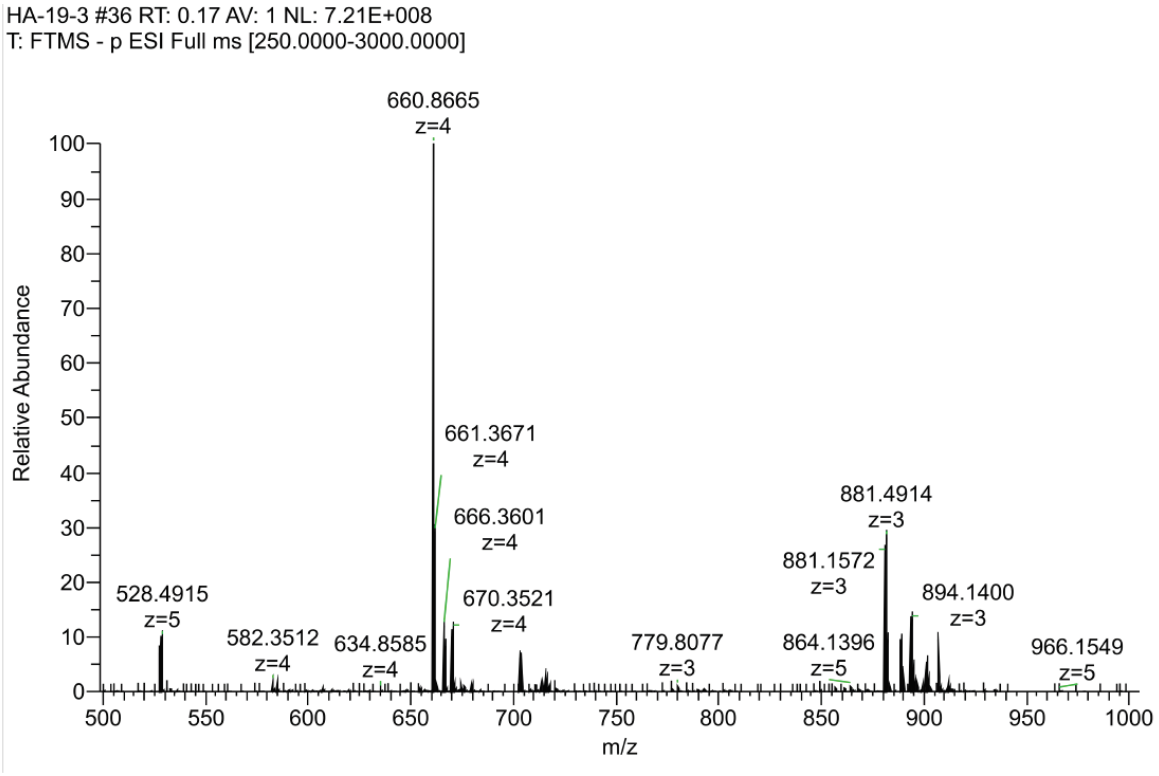


Figure S219. HRMS (ESI) of dC[R^{N3}(-3),A^{Y2}(+1)].

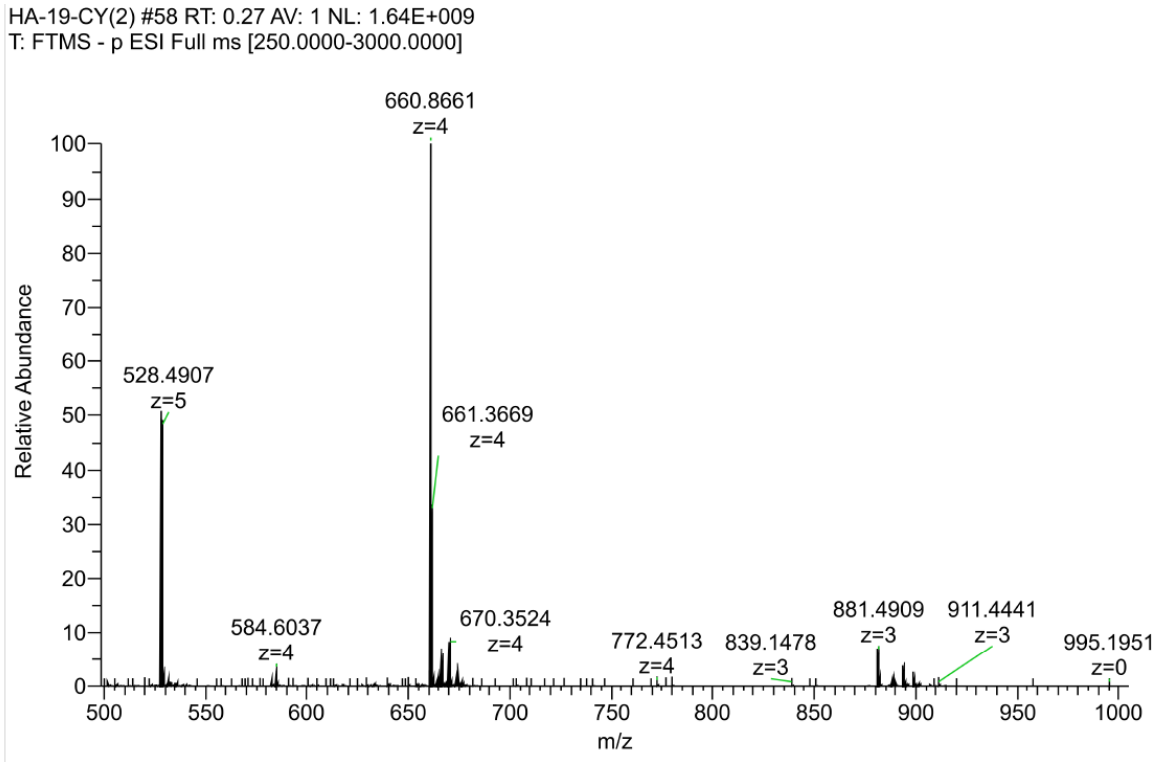


Figure S220. HRMS (ESI) of dC[R^{N3}(-3),A^{Y2}(+1)]X

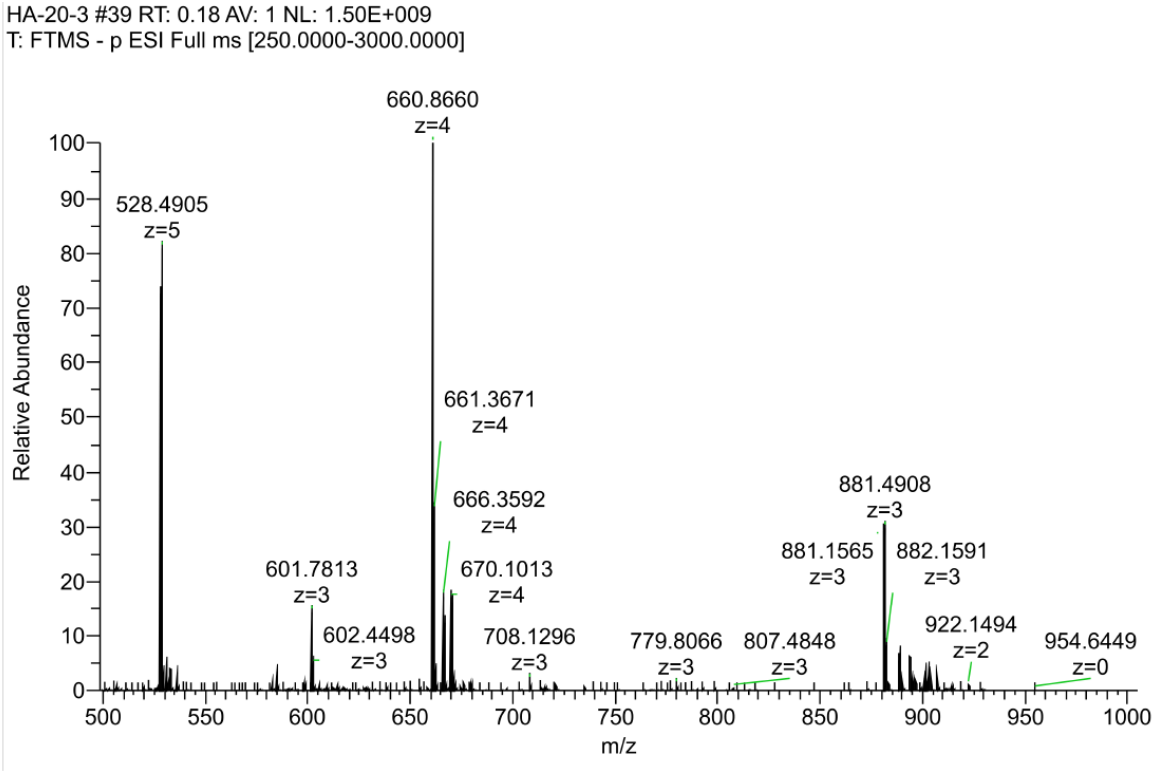


Figure S221. HRMS (ESI) of dC[R^{N3}(-2),A^{Y2}(+1)].

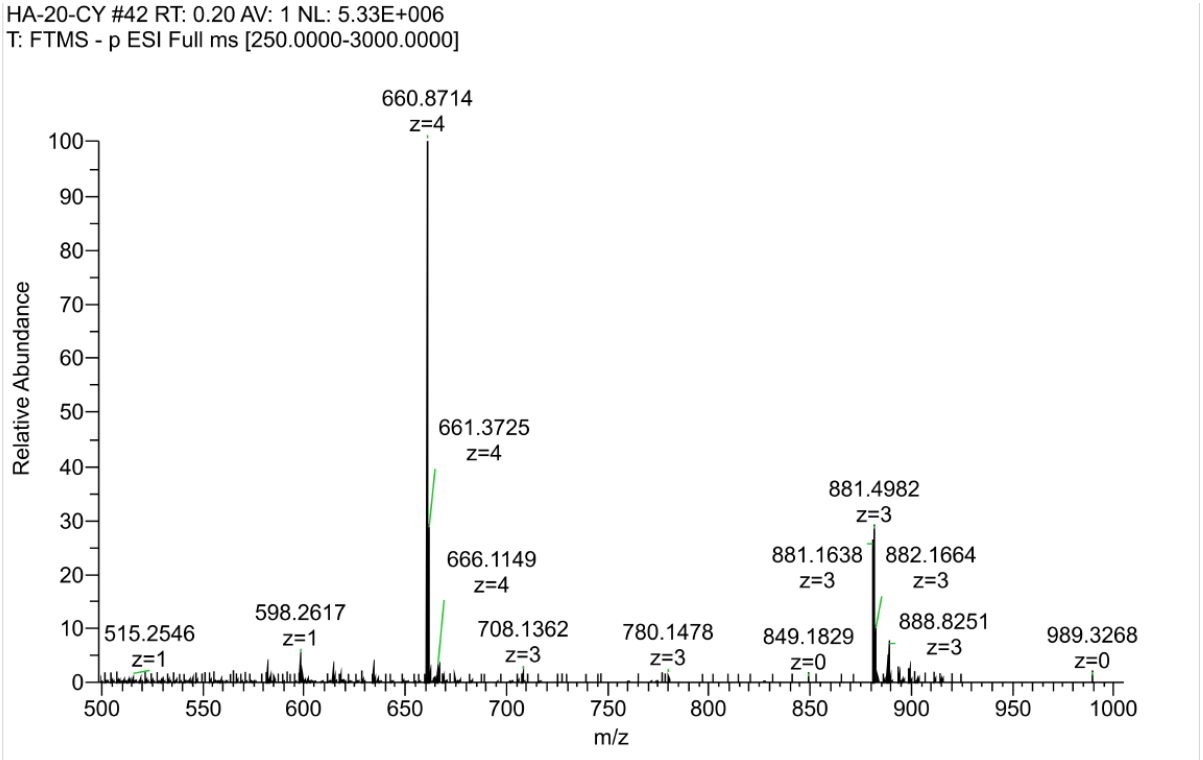


Figure S222. HRMS (ESI) of dC[R^{N3}(-2),A^{Y2}(+1)]X.

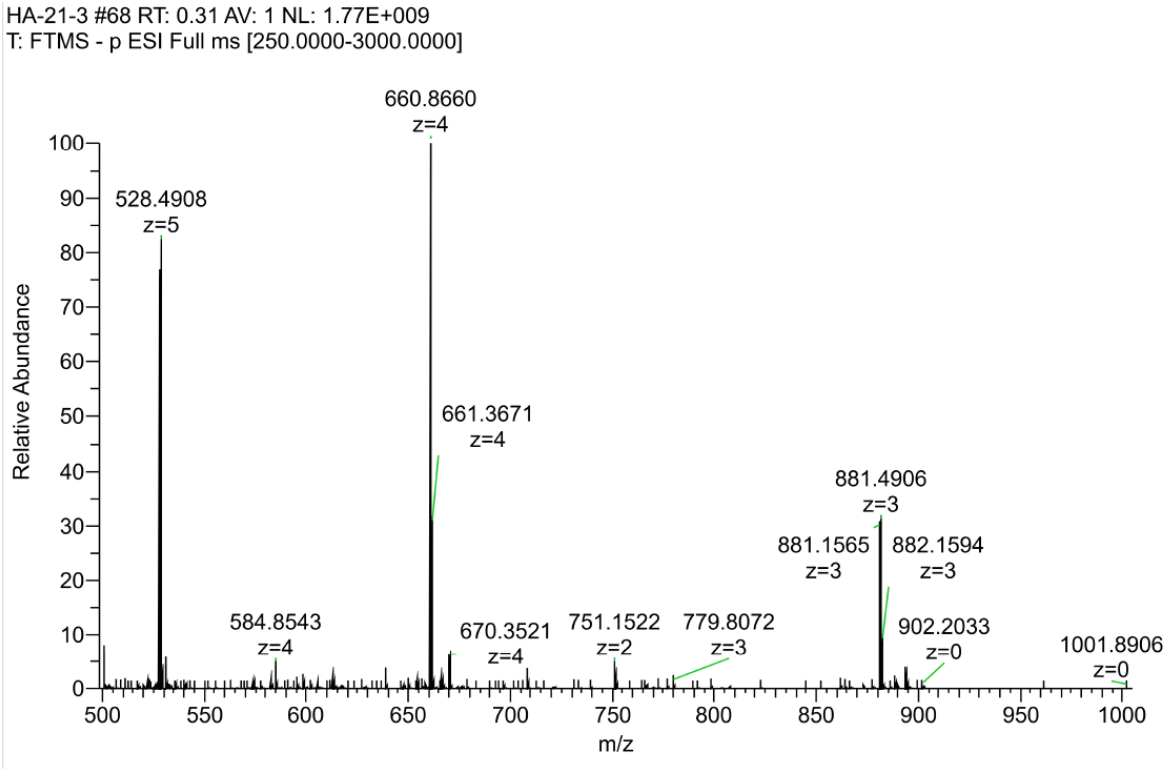


Figure S223. HRMS (ESI) of dC[R^{N3}(-1),A^{Y2}(+1)].

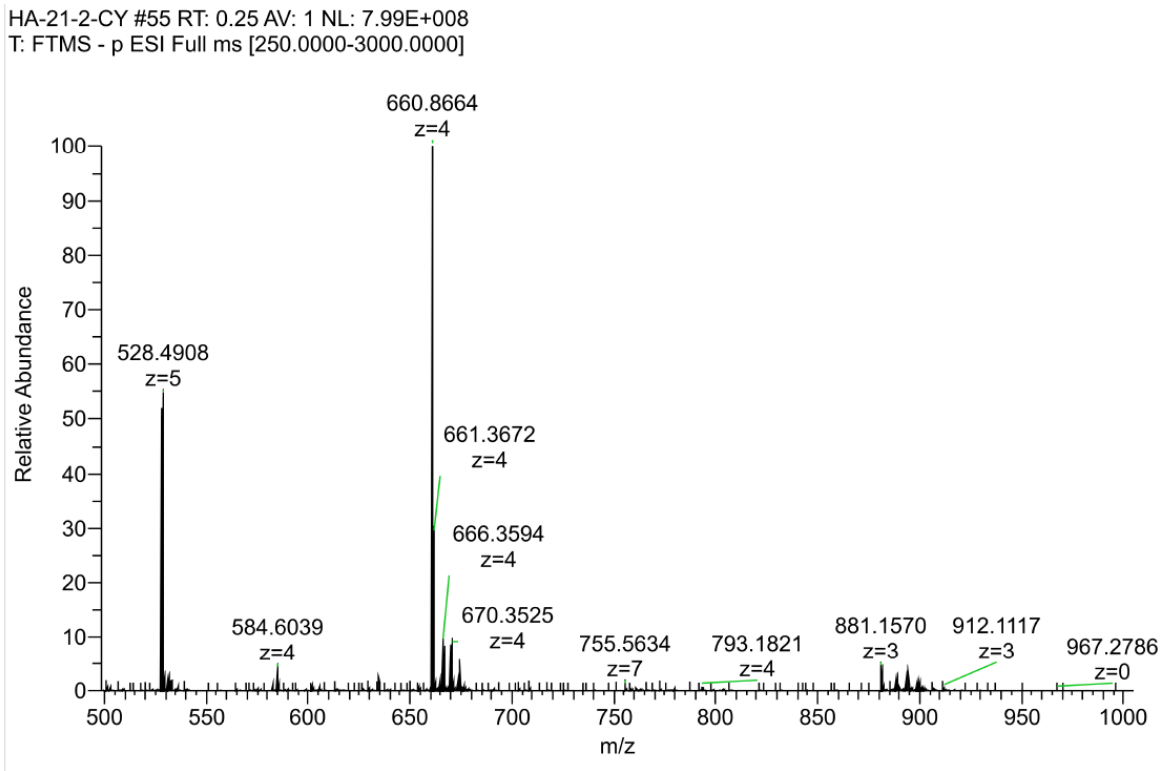


Figure S224. HRMS (ESI) of dC[R^{N3}(-1),A^{Y2}(+1)]X.

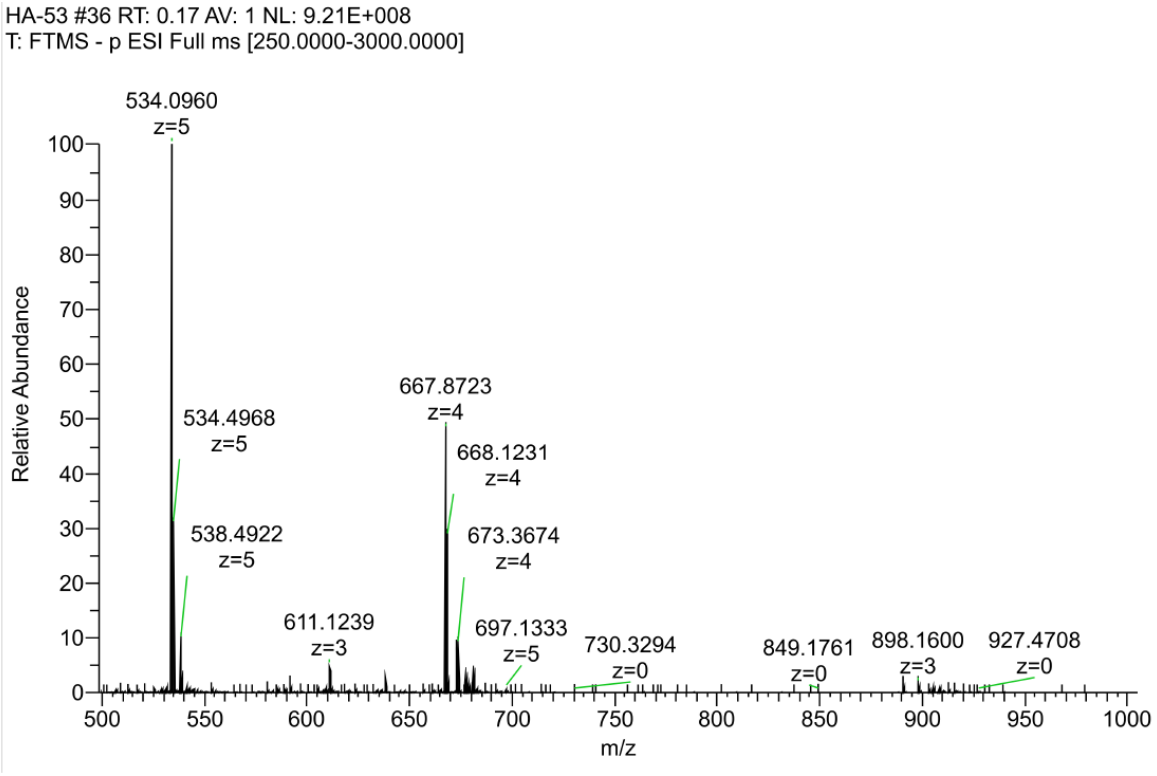


Figure S225. HRMS (ESI) of dC[R^{N3}(-2),A^{Y4}(+1)].

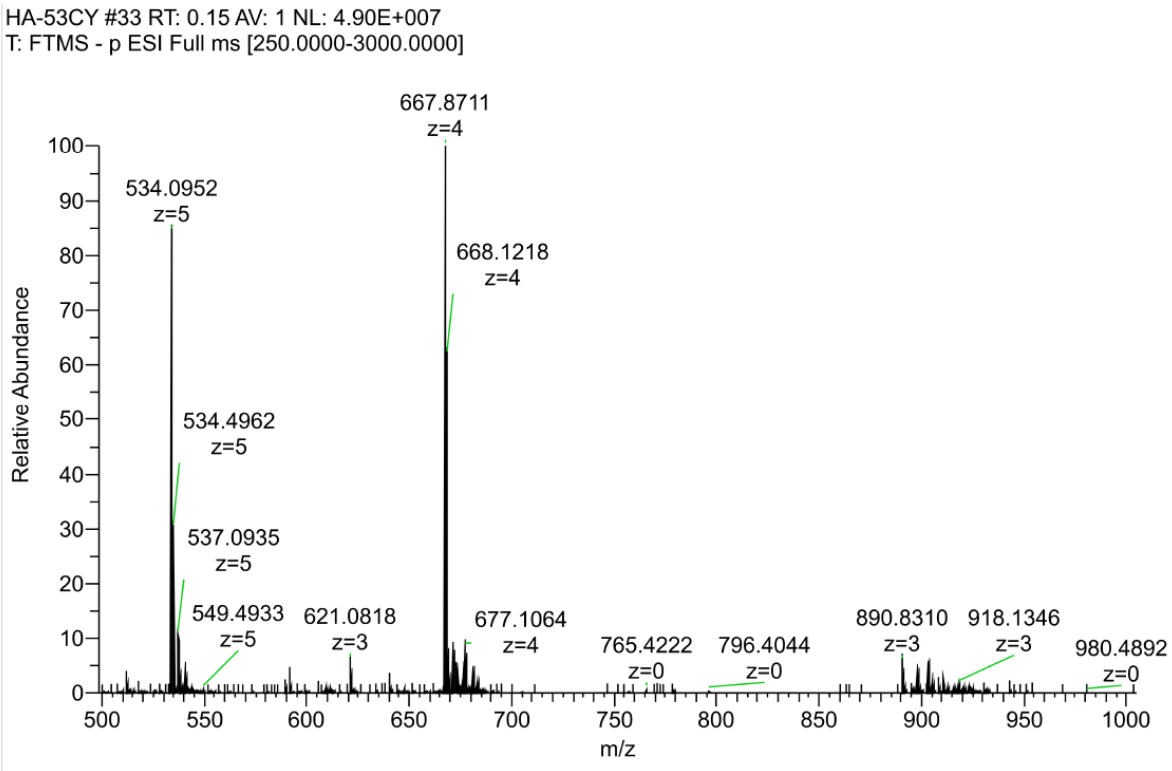


Figure S226. HRMS (ESI) of dC[R^{N3}(-2),A^{Y4}(+1)]X.

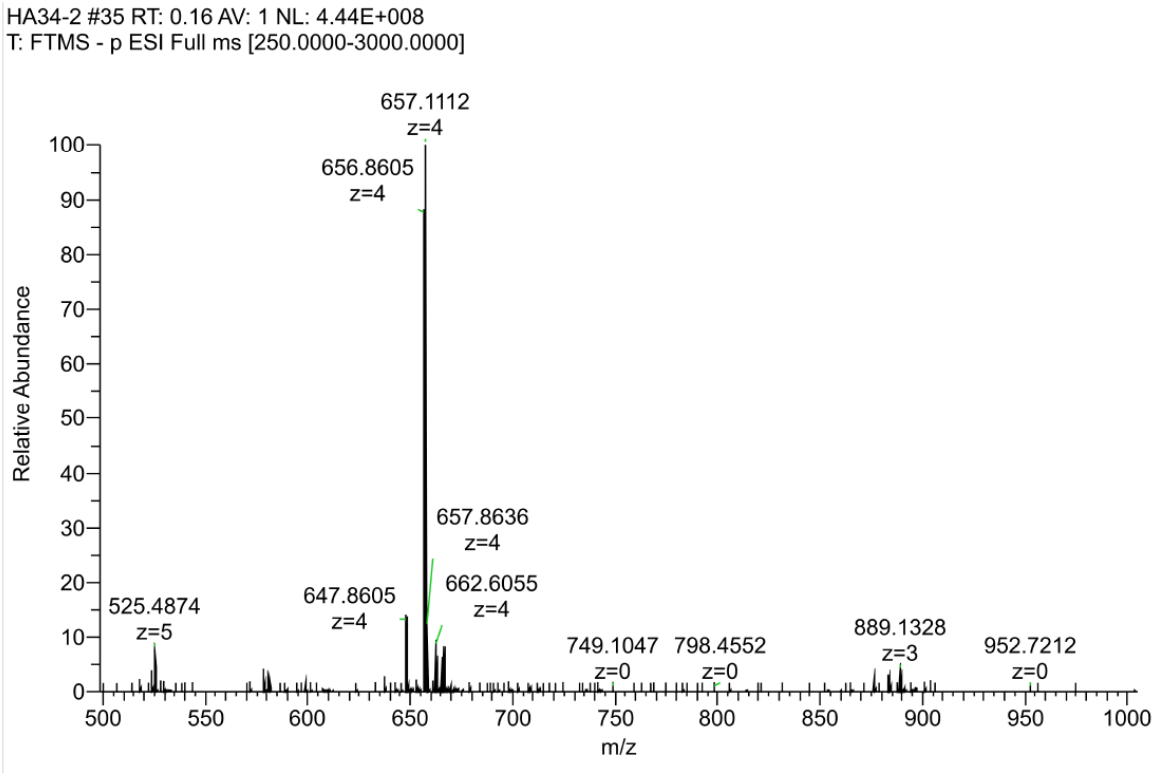


Figure S227. HRMS (ESI) of dZ[R^{N3}(-3),A^{Y2}(+1)].

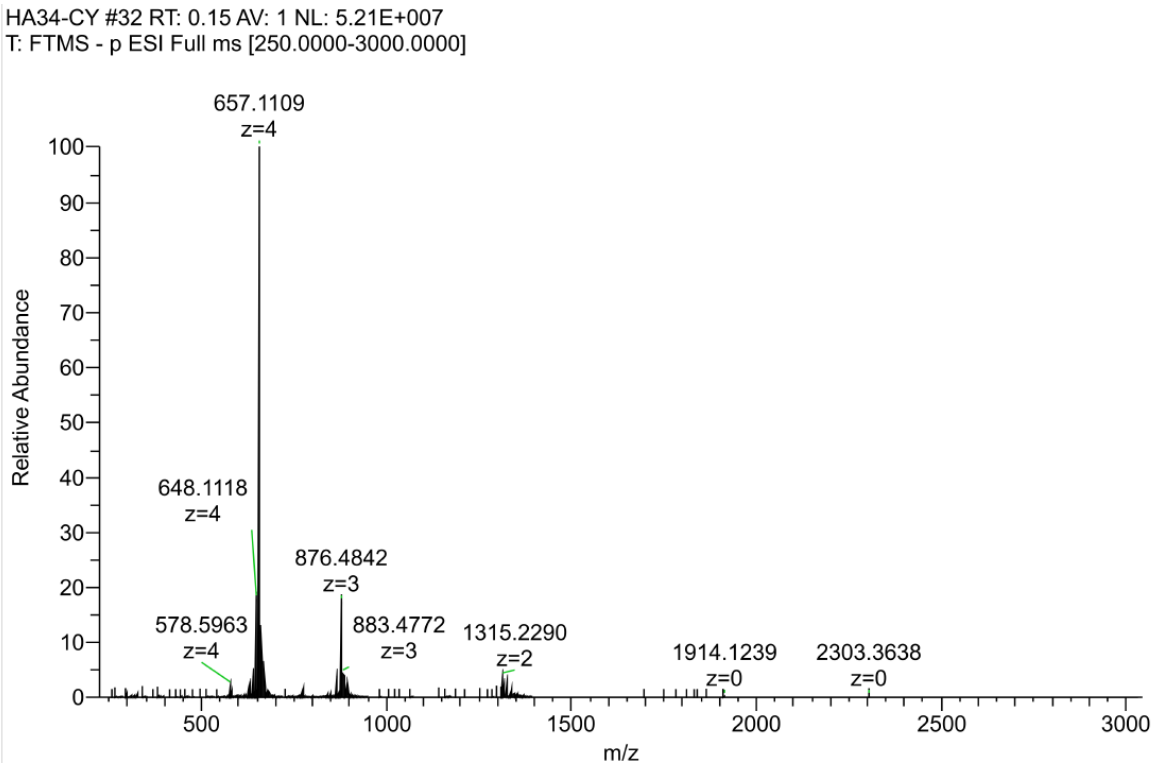


Figure S228. HRMS (ESI) of dZ[R^{N3}(-3),A^{Y2}(+1)]X.

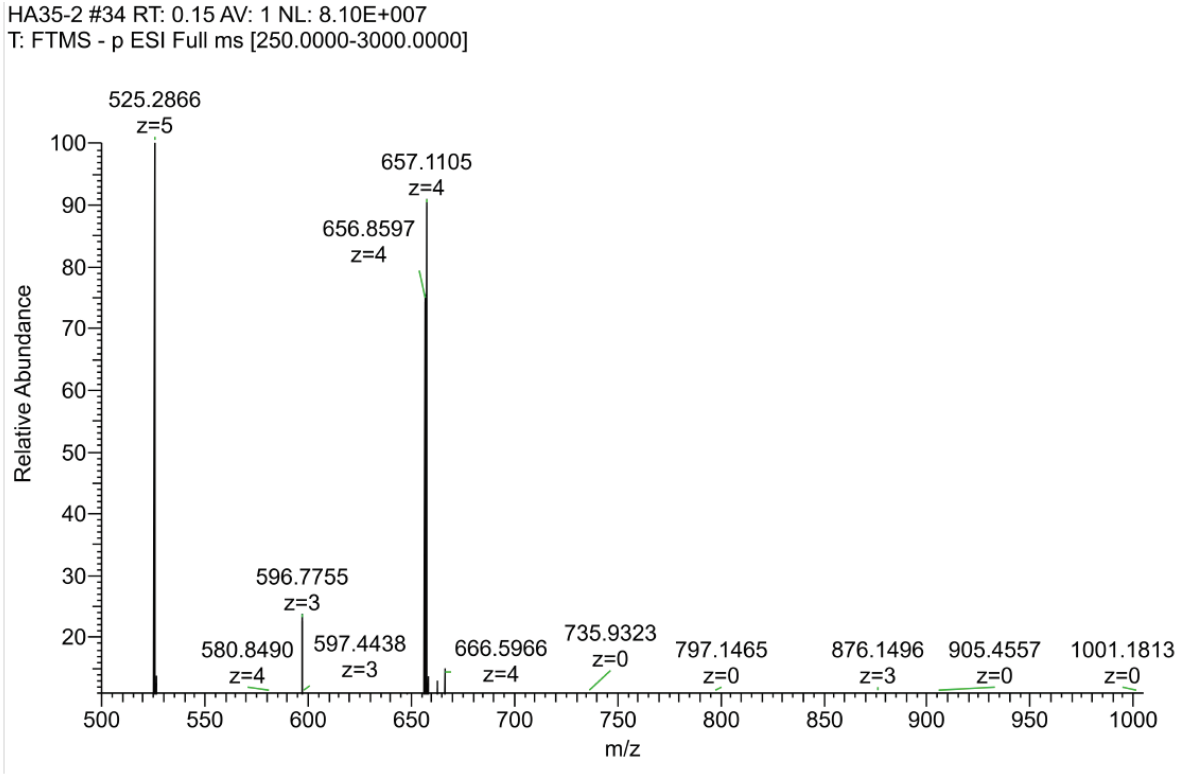


Figure S229. HRMS (ESI) of dZ[R^{N3}(-2),A^{Y2}(+1)].

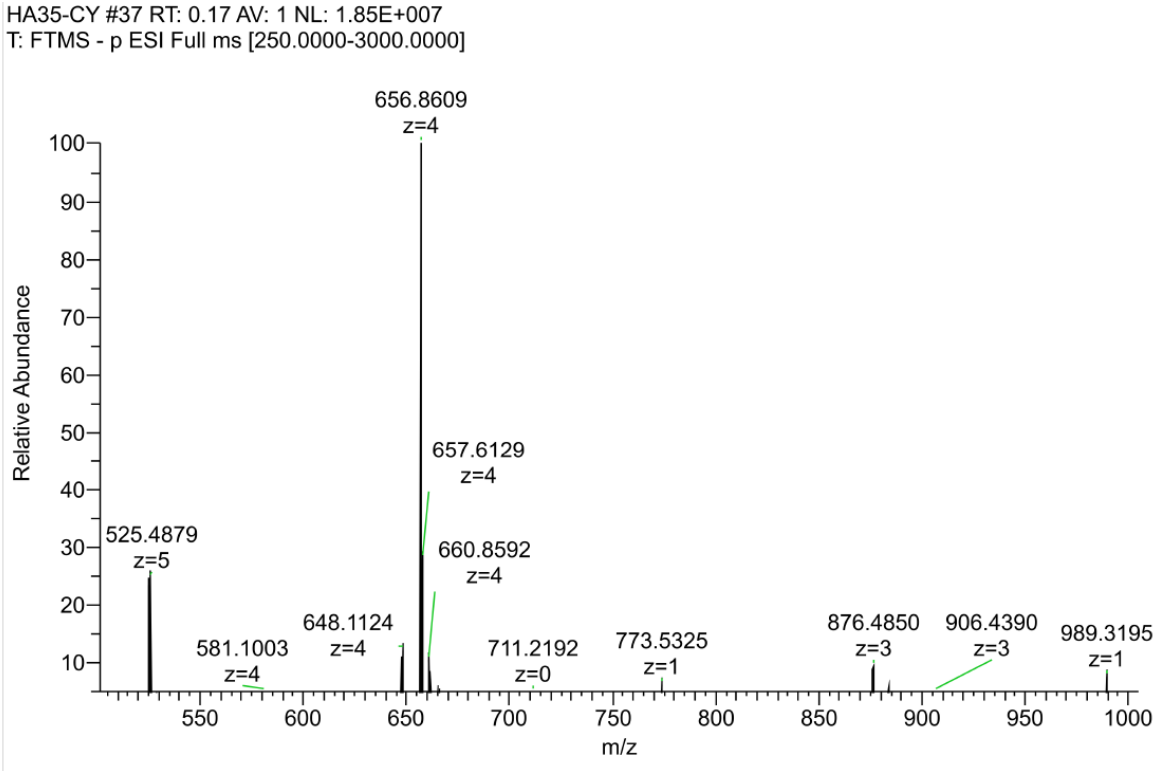


Figure S230. HRMS (ESI) of dZ[R^{N3}(-2),A^{Y2}(+1)]X.

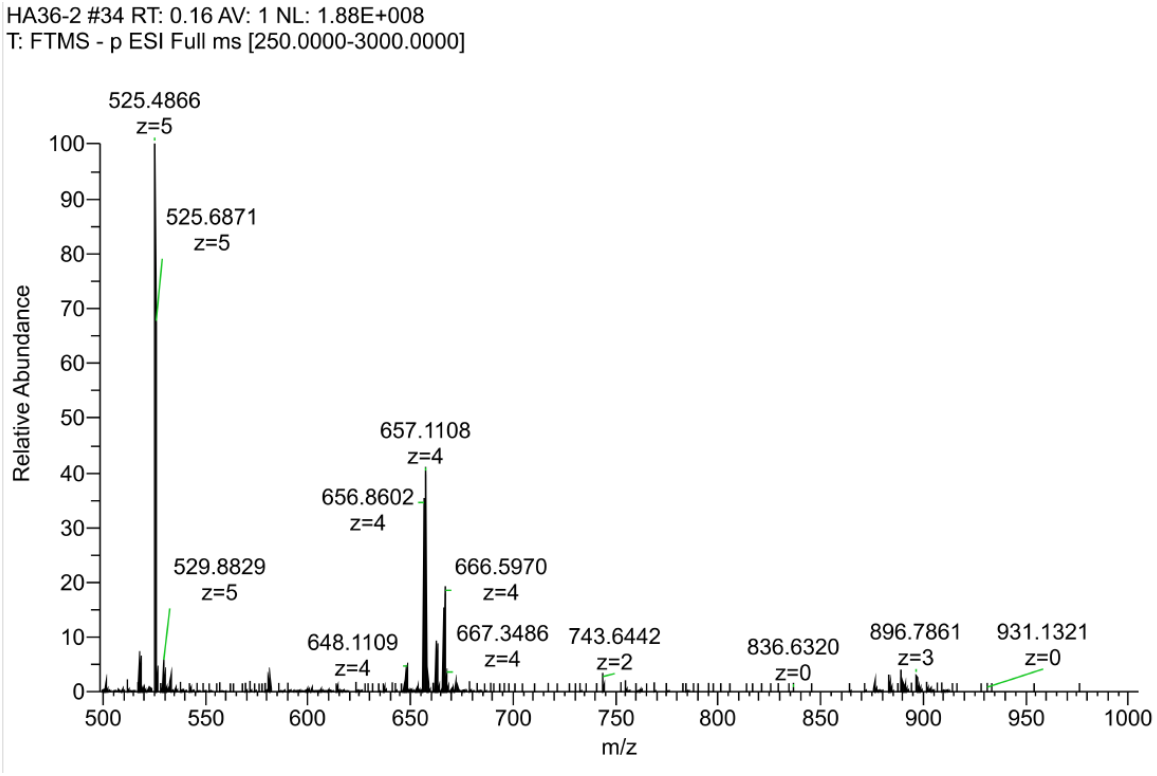


Figure S231. HRMS (ESI) dZ[R^{N3}(-1),A^{Y2}(+1)].

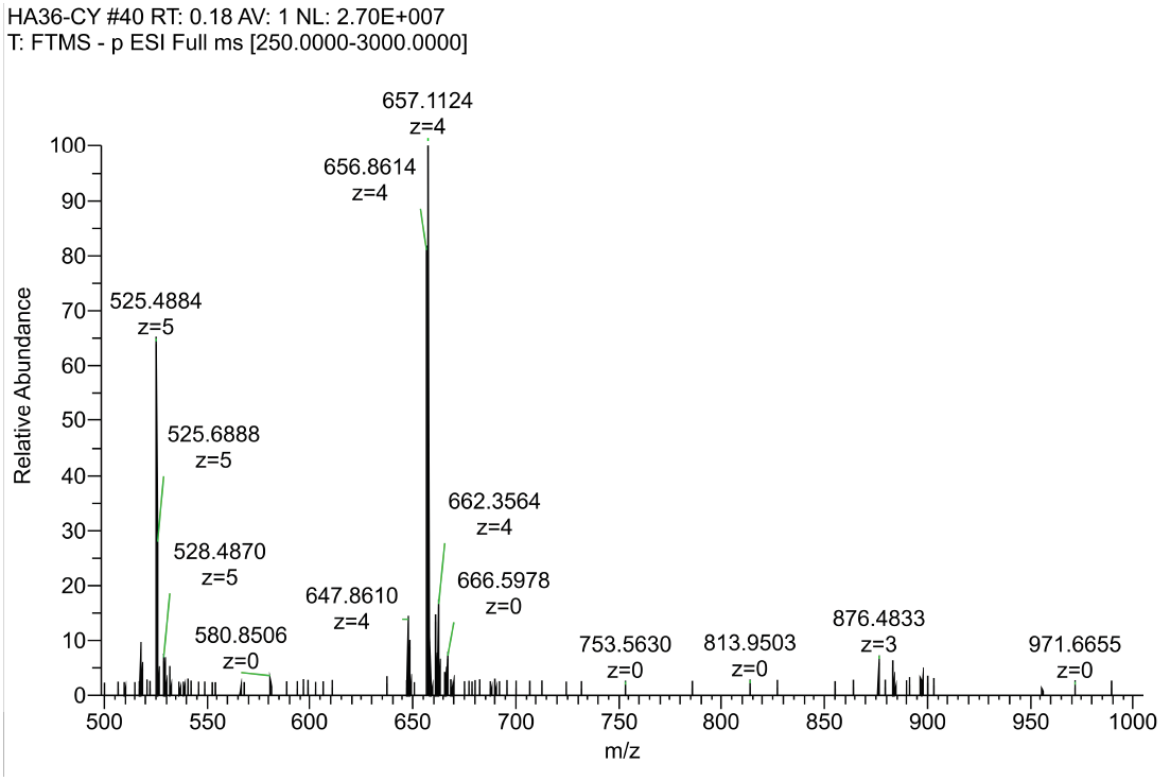


Figure S232. HRMS (ESI) dZ[R^{N3}(-1),A^{Y2}(+1)]X.

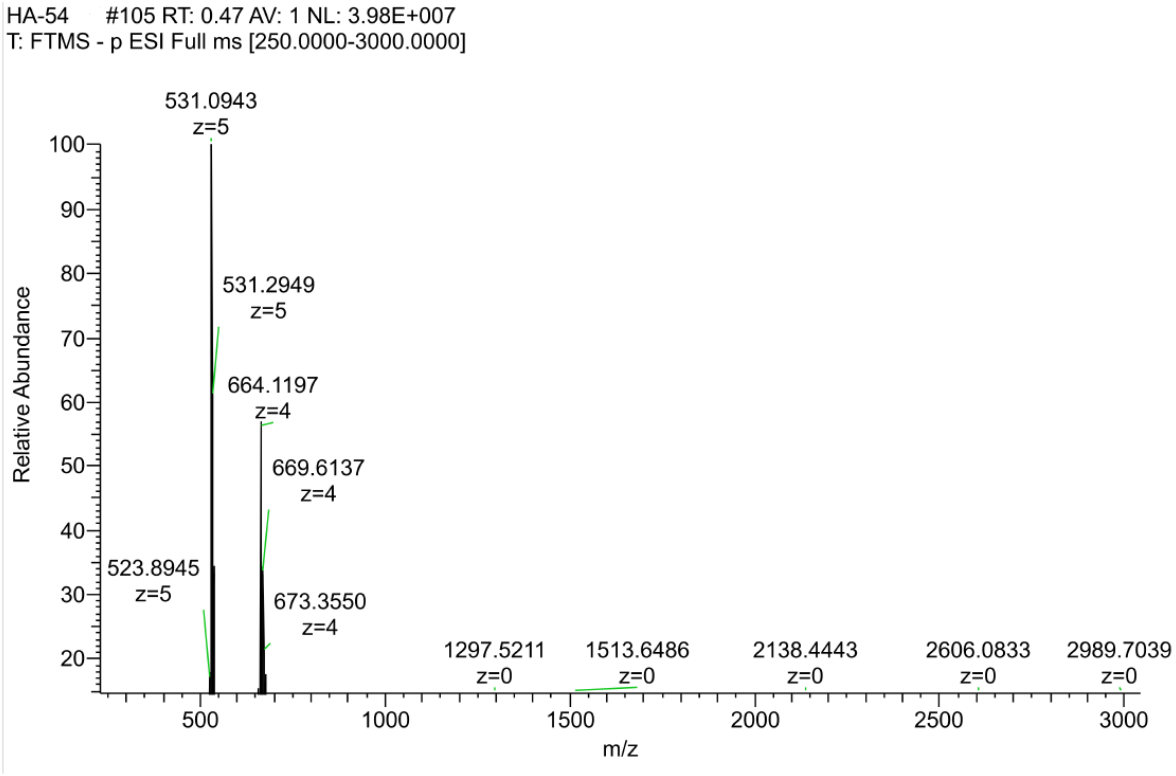


Figure S233. HRMS (ESI) of dZ[R^{N3}(-2),A^{Y4}(+1)].

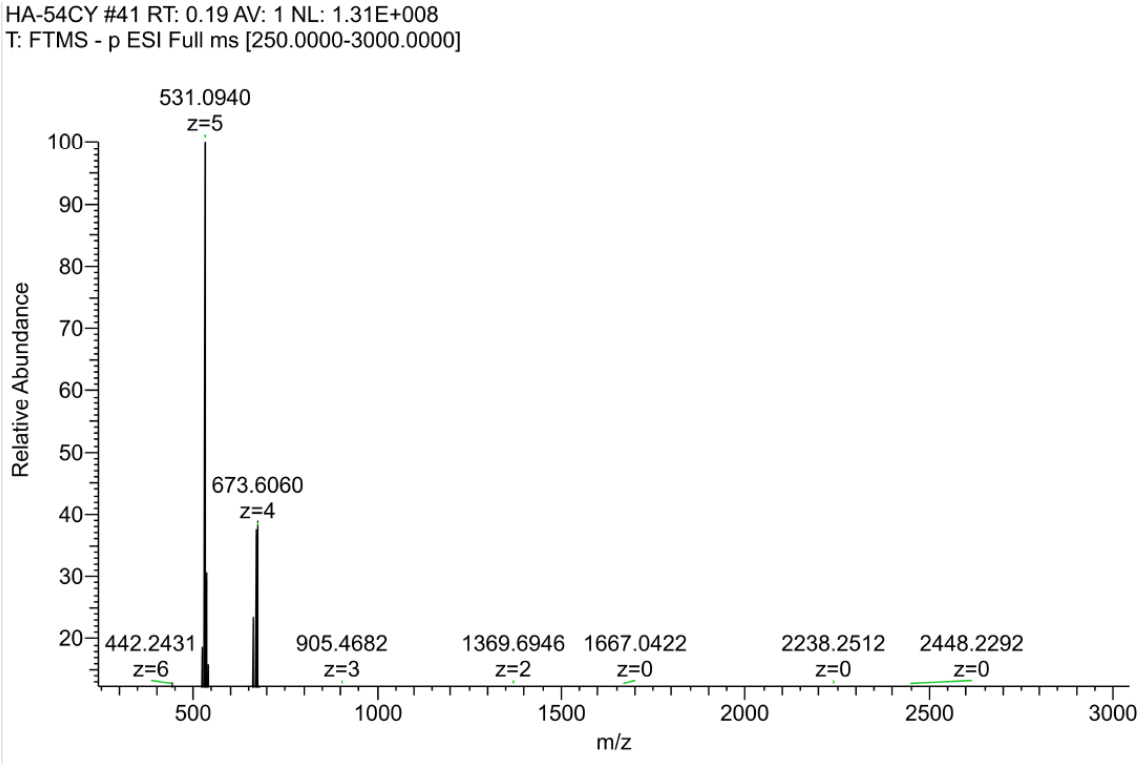


Figure S234. HRMS (ESI) of dZ[R^{N3}(-2),A^{Y4}(+1)]X.

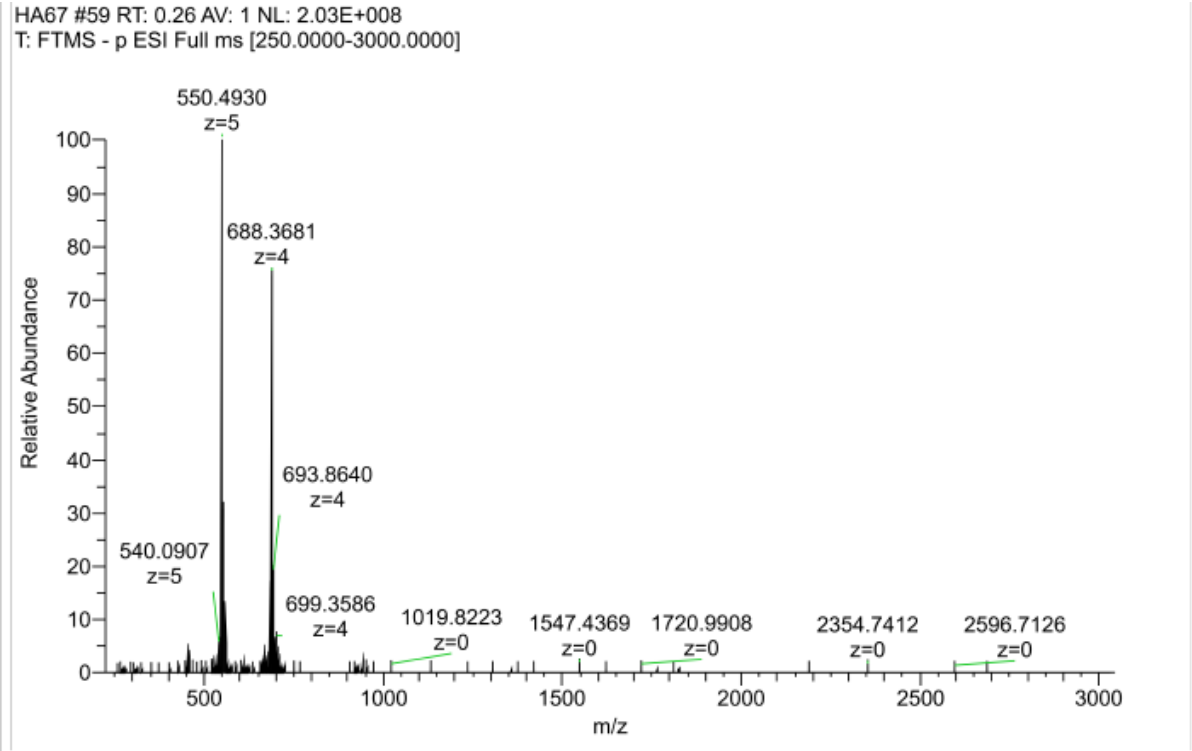


Figure S235. HRMS (ESI) of dC[U^{E}(-2),A^{N3}(+1)].

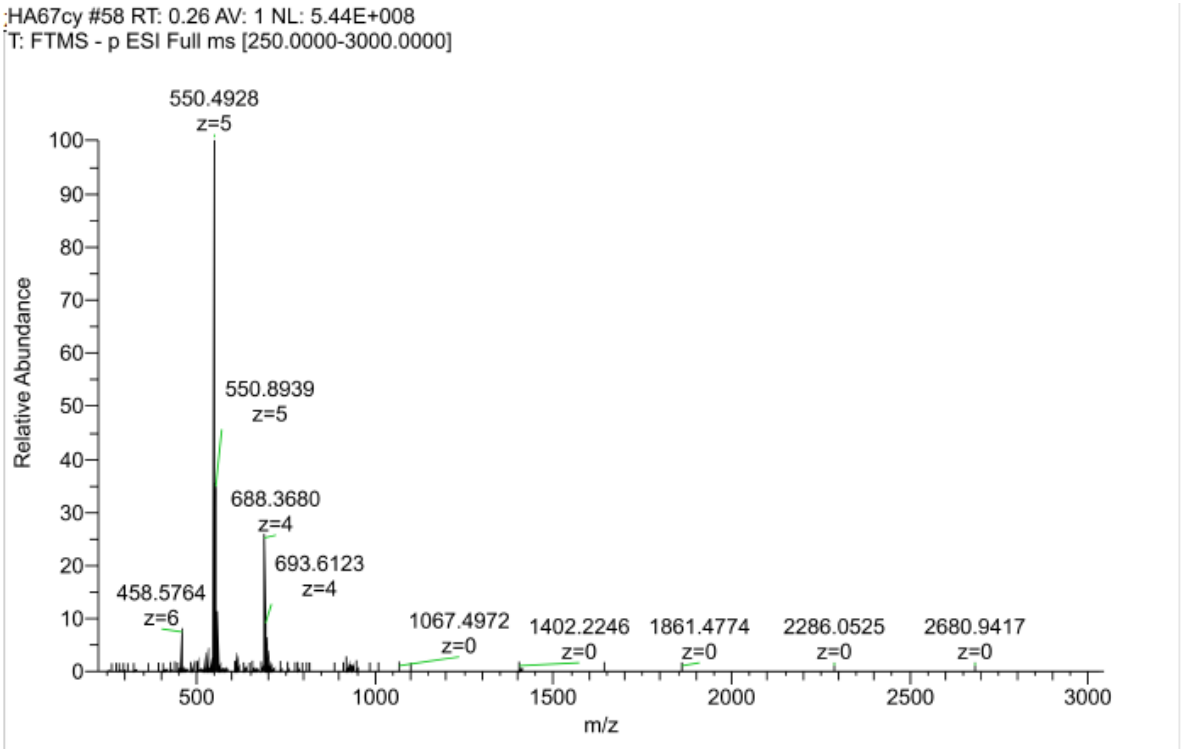


Figure S236. HRMS (ESI) of dC[U^{E}(-2),A^{N3}(+1)]X.

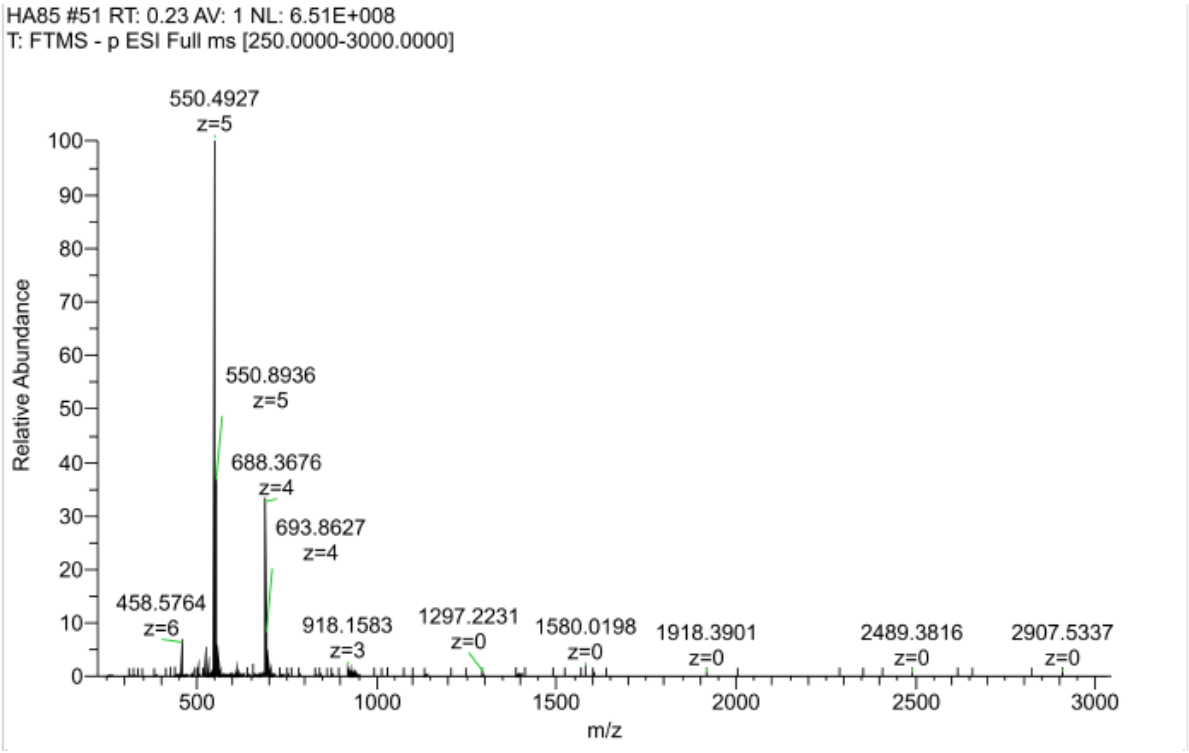


Figure S237. HRMS (ESI) of dC[U^{E}(-3),A^{N3}(+1)].

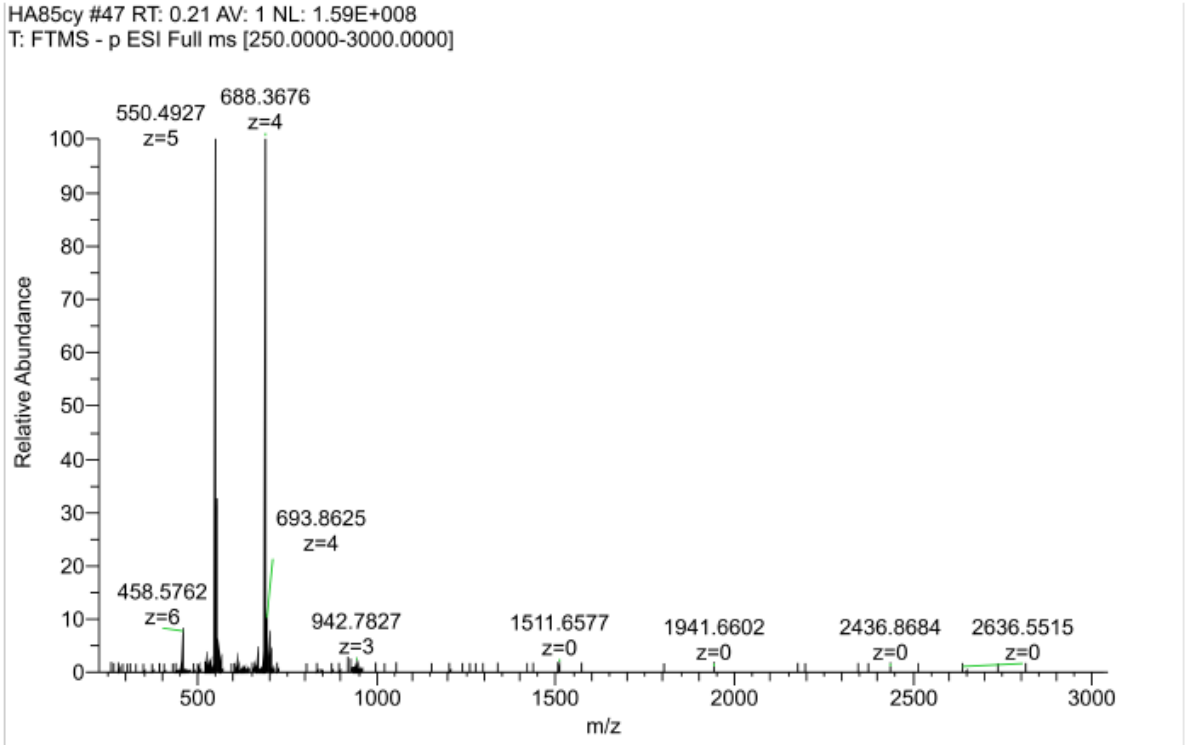


Figure S238. HRMS (ESI) of dC[U^{E}(-3),A^{N3}(+1)]X.

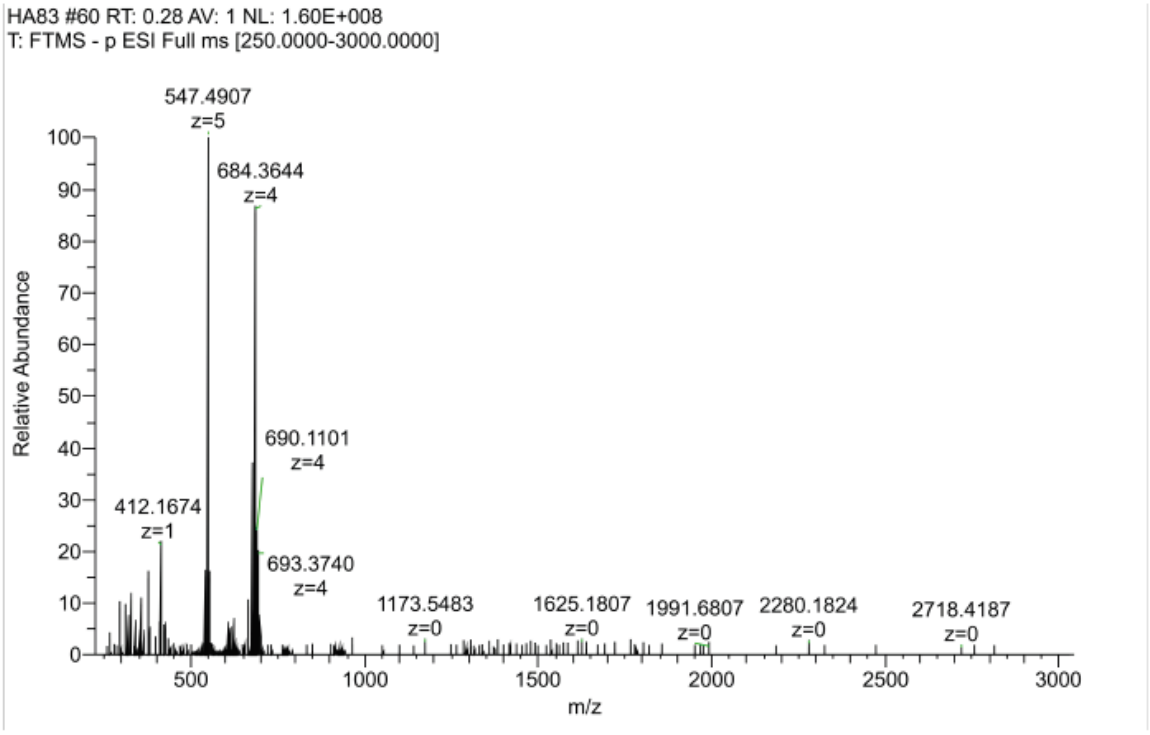


Figure S239. HRMS (ESI) of dZ[U^{E}(-3),A^{N3}(+1)].

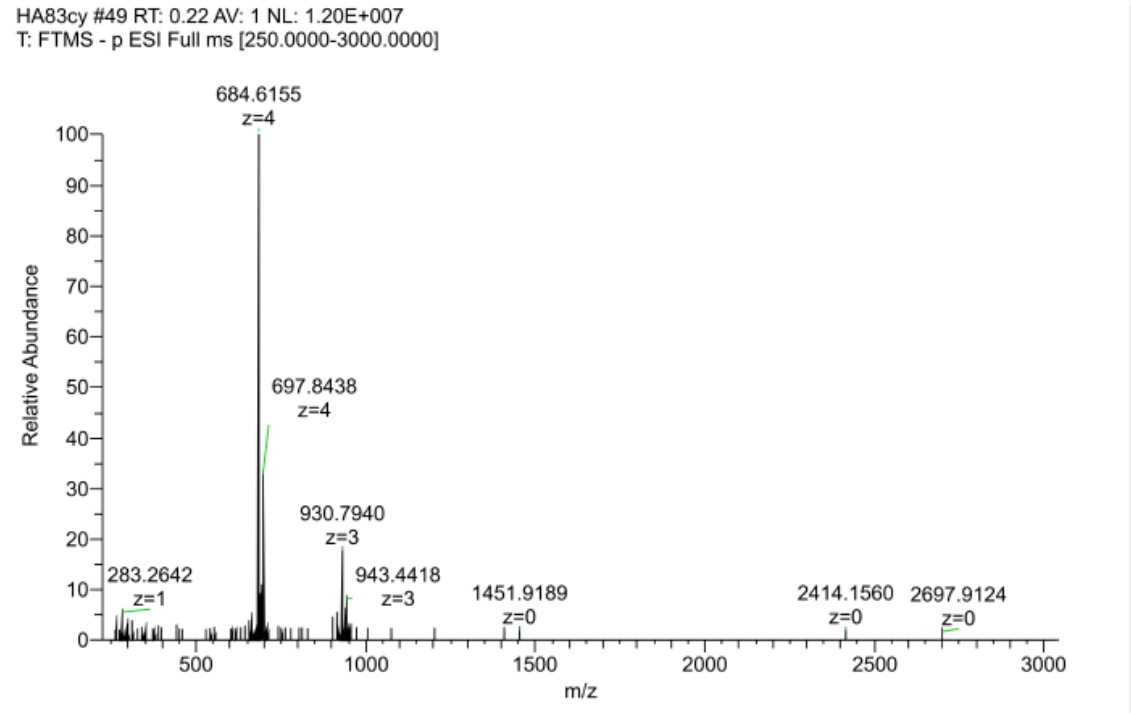


Figure S240. HRMS (ESI) of dZ[U^{E}(-3),A^{N3}(+1)]X.

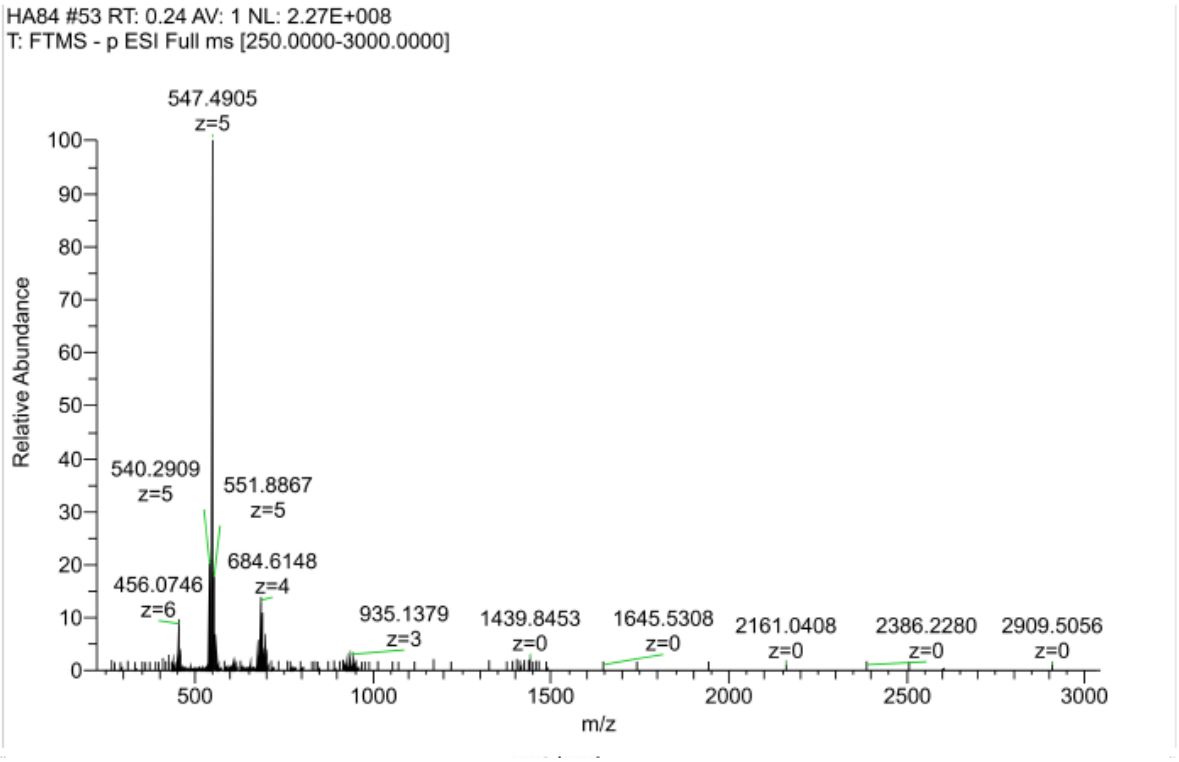


Figure S241. HRMS (ESI) of dZ[U^{E}(-2),A^{N3}(+1)].

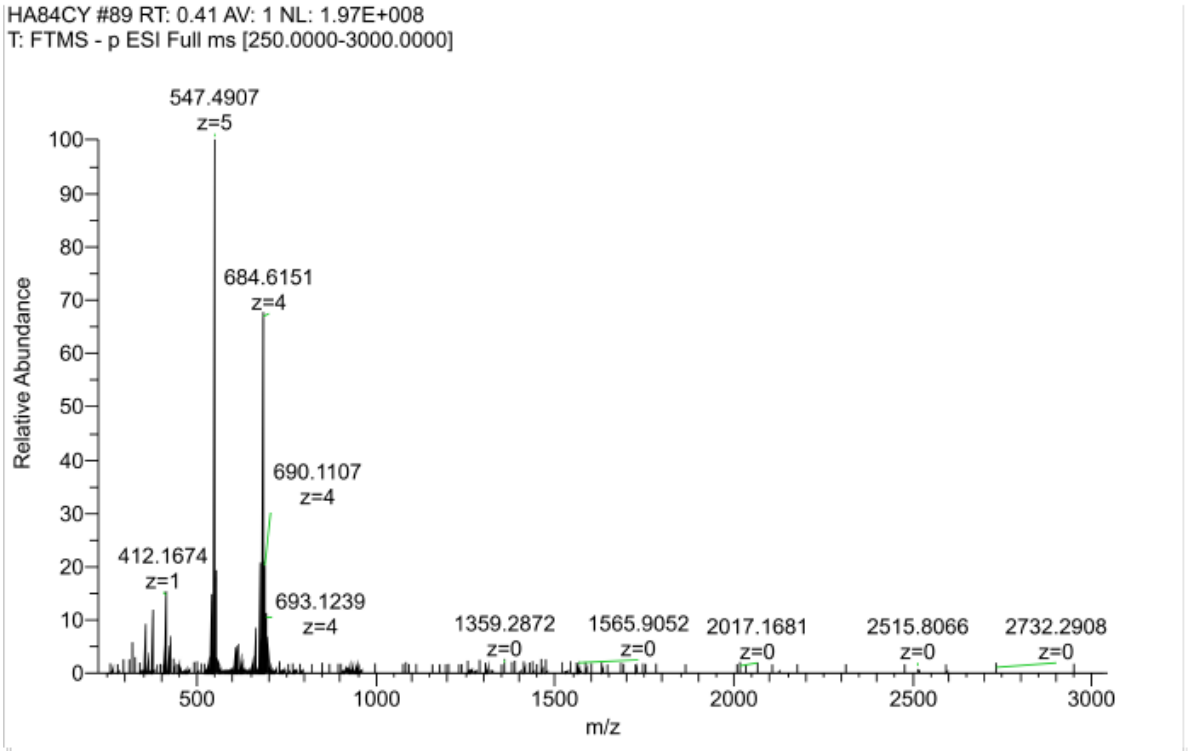


Figure S242. HRMS (ESI) of dZ[U^{E}(-2),A^{N3}(+1)]X.

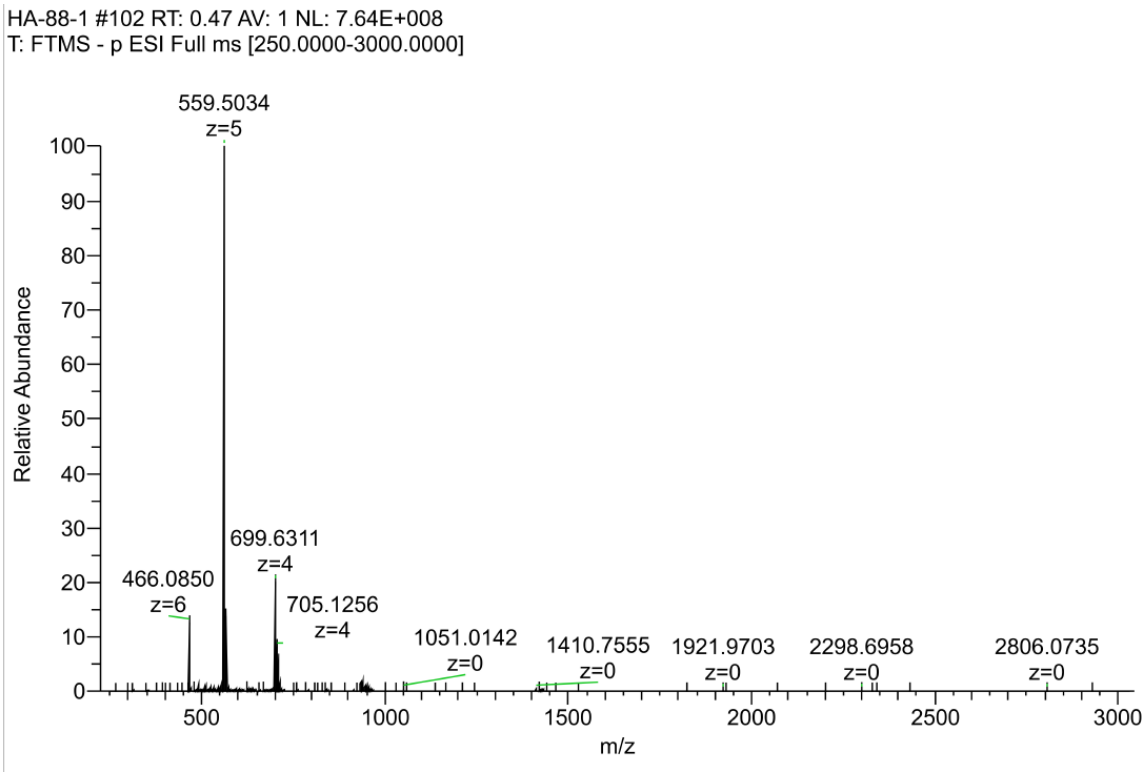


Figure S243. HRMS (ESI) of dC[C^{N3}(-2),H^{E}(+1)].

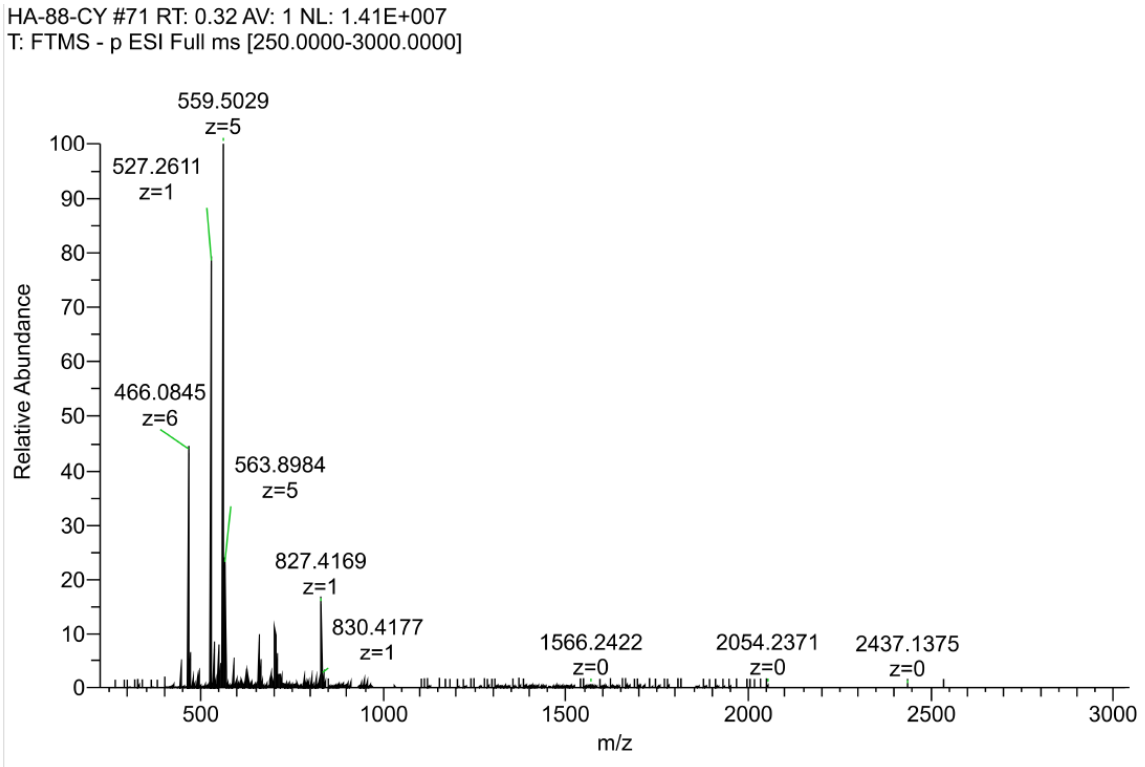


Figure S244. HRMS (ESI) of dC[C^{N3}(-2),H^{E}(+1)]X.

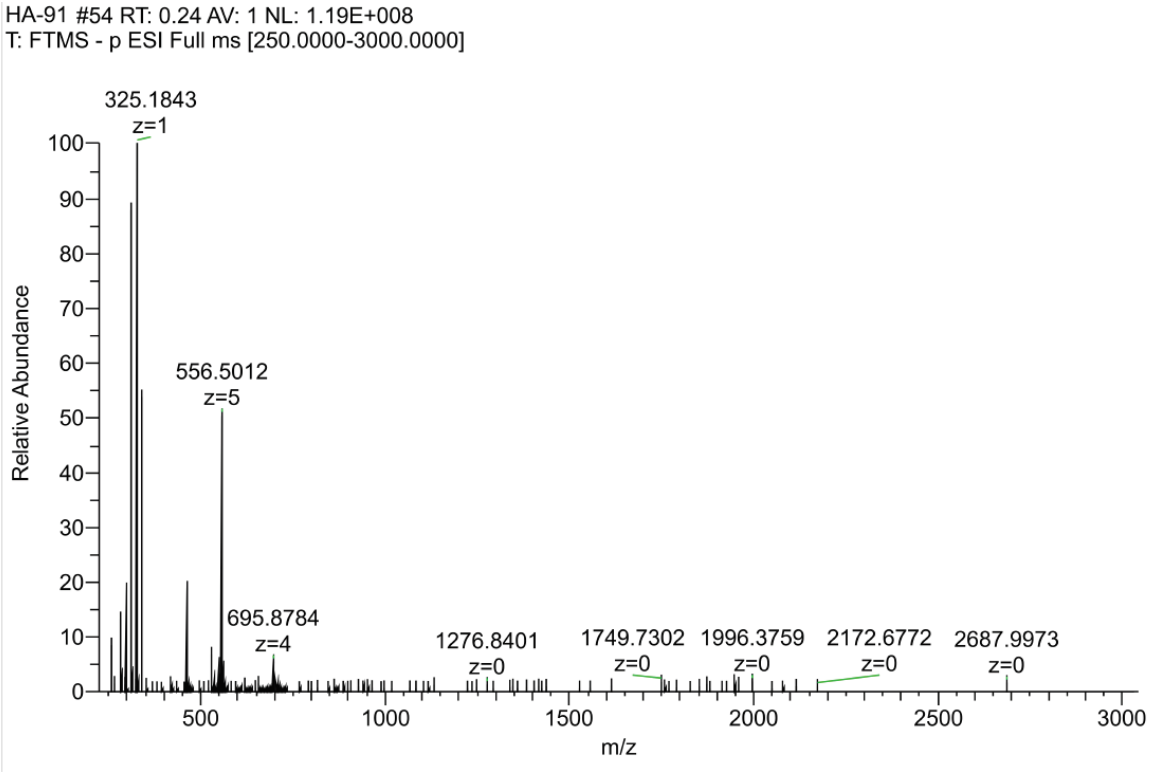


Figure S245. HRMS (ESI) of dZ[C^{N3}(-2),H^{E}(+1)].

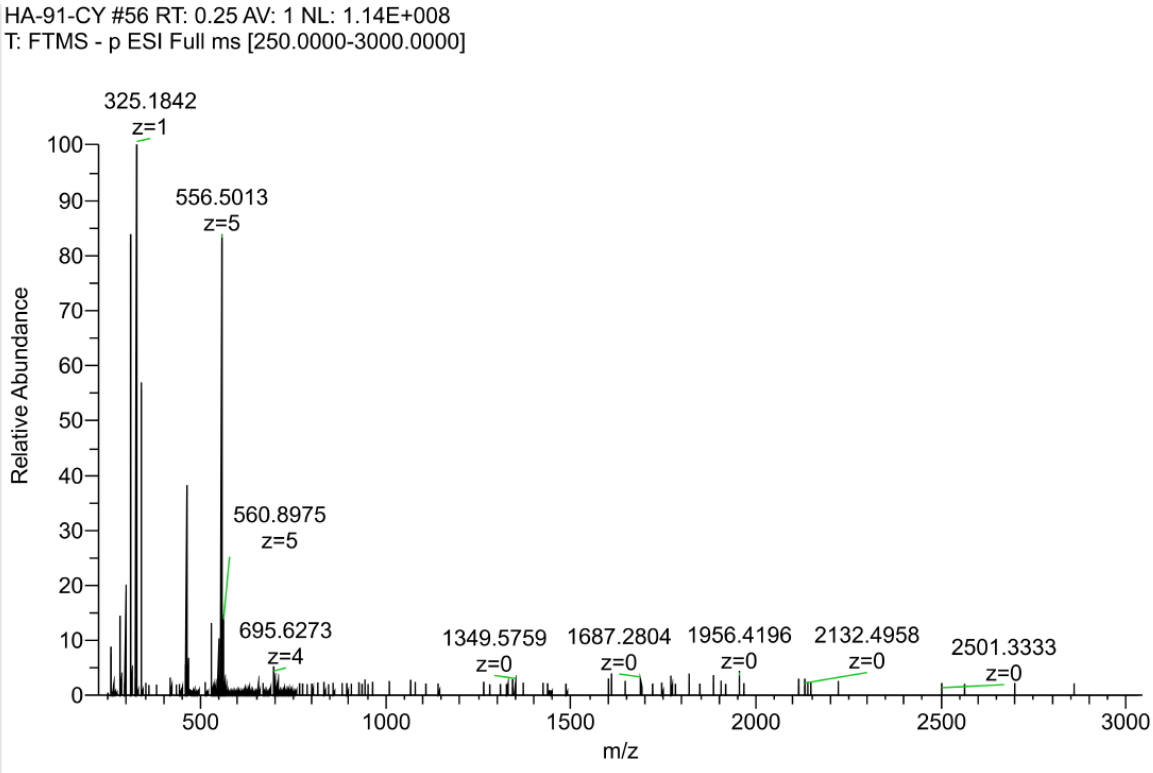


Figure S246. HRMS (ESI) of dZ[C^{N3}(-2),H^{E}(+1)]X.

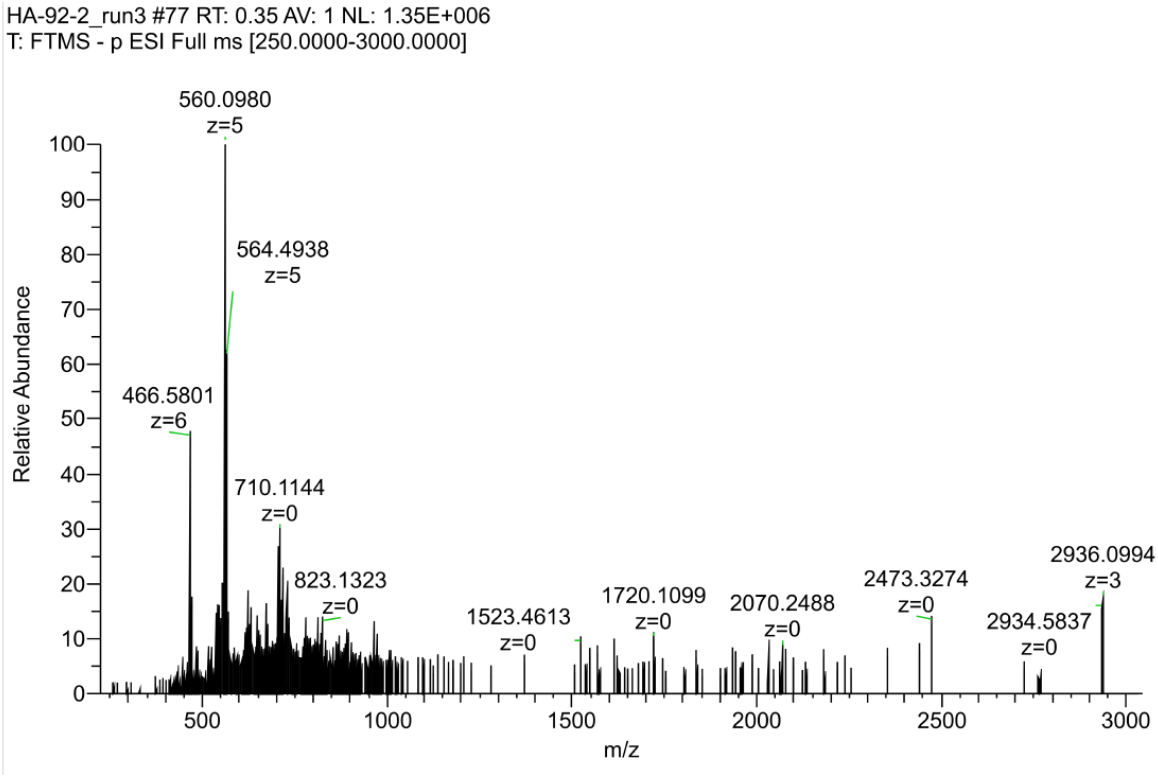


Figure S247. HRMS (ESI) of FdZ[C^{N3}(-2),H^{E}(+1)].

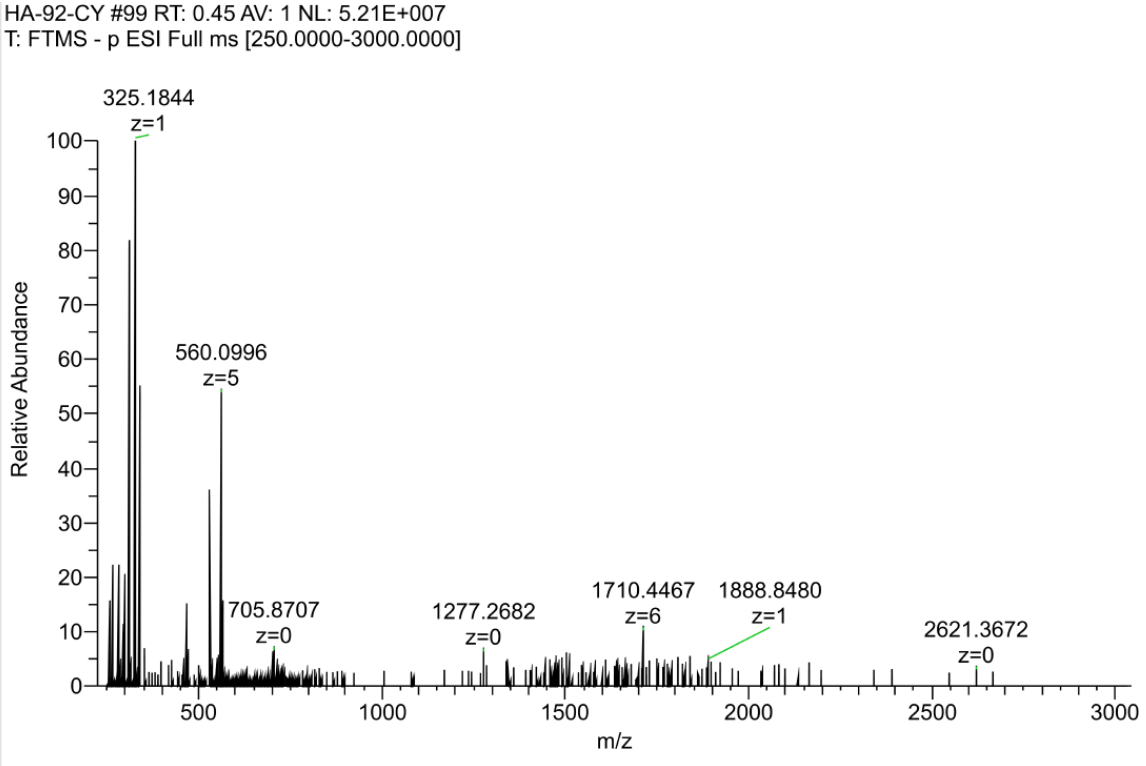


Figure S248. HRMS (ESI) of FdZ[C^{N3}(-2),H^{E}(+1)]X.

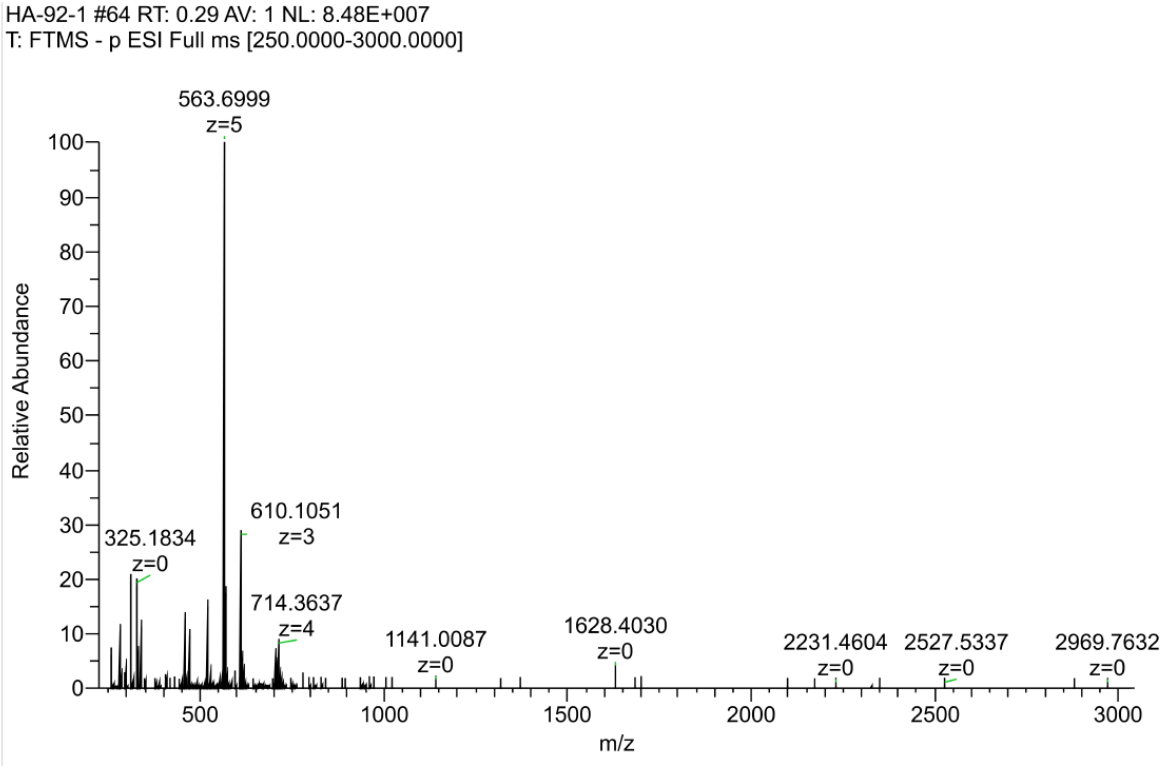


Figure S249. HRMS (ESI) of hydrated FdZ[C^{N3}(-2),H^{E}(+1)].

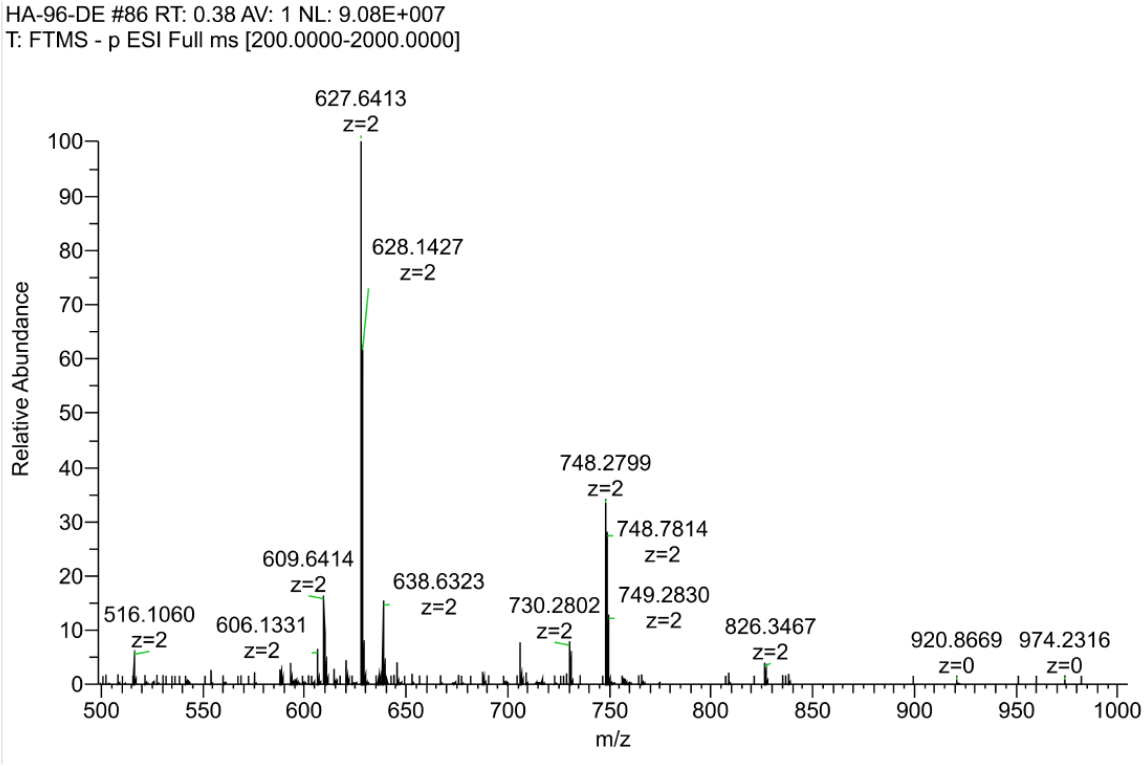


Figure S250. HRMS (ESI) of 4-mer dZ[C^{N3}(-2),H^{E}(+1)].

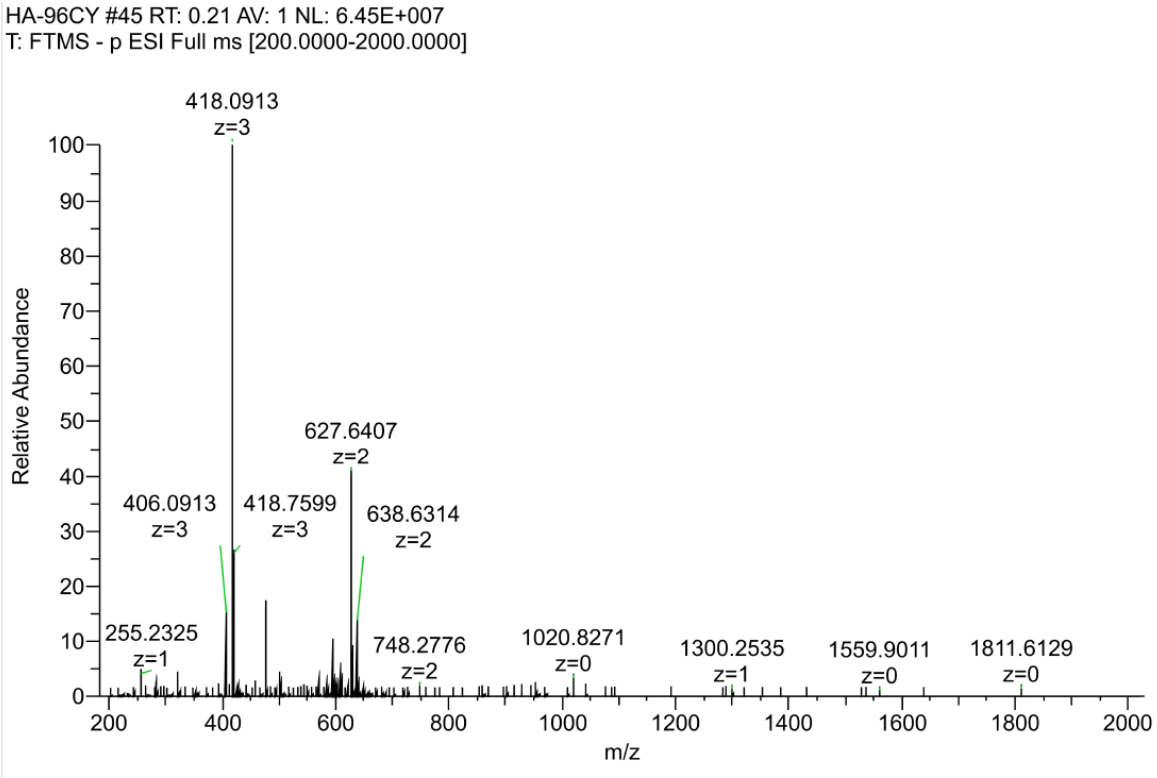


Figure S251. HRMS (ESI) of 4-mer dZ[C^{N3}(-2),H^{E}(+1)]X.

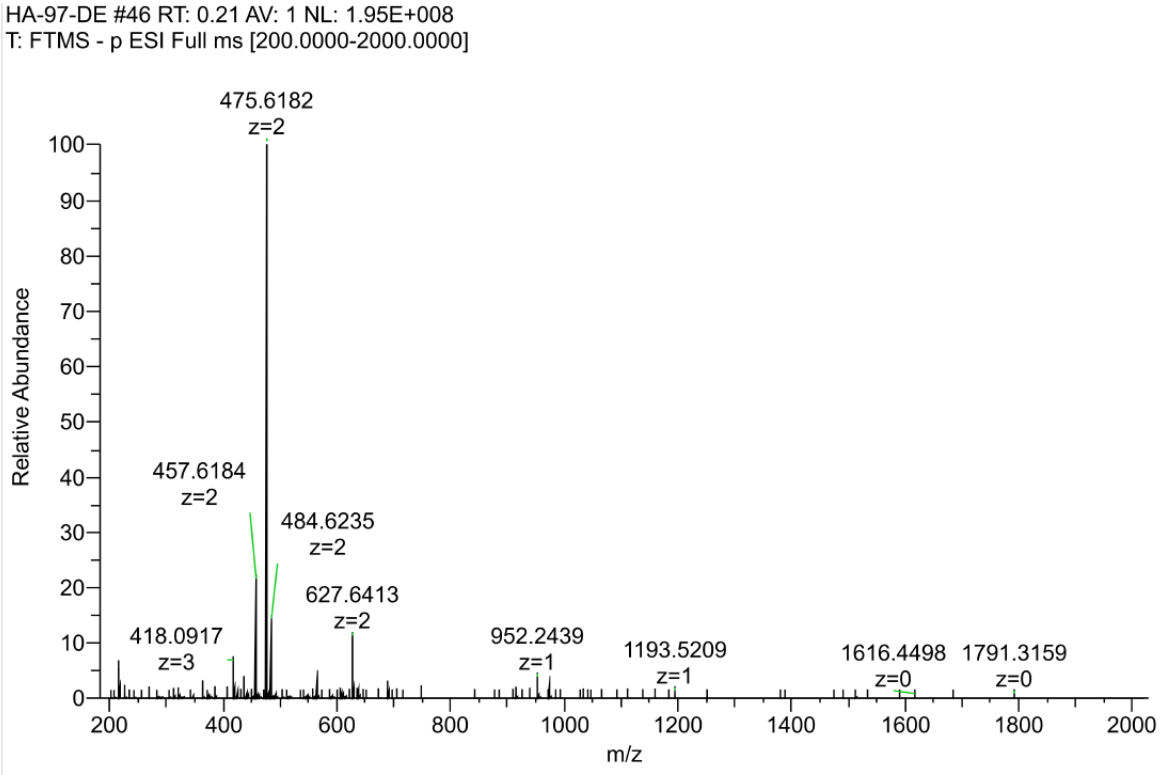


Figure S252. HRMS (ESI) of 3-mer dZ[C^{N3}(-1),H^{E}(+1)].

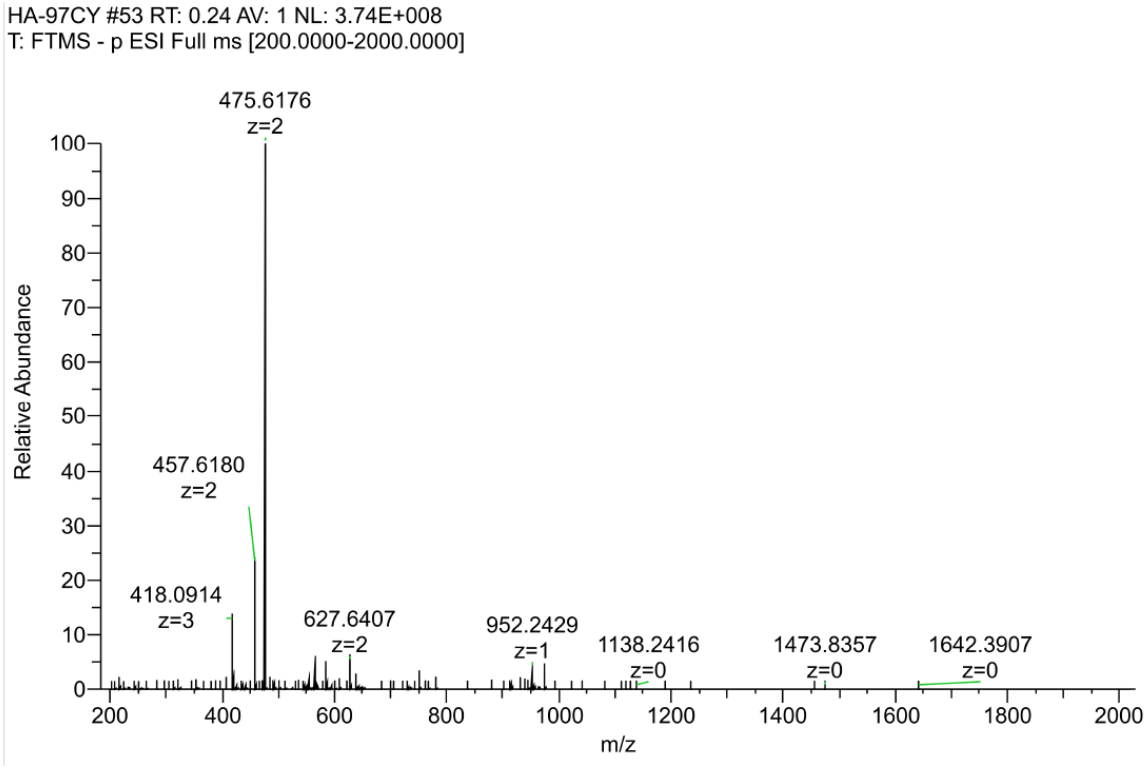


Figure S253. HRMS (ESI) of 3-mer dZ[C^{N3}(-1),H^{E}(+1)]X.

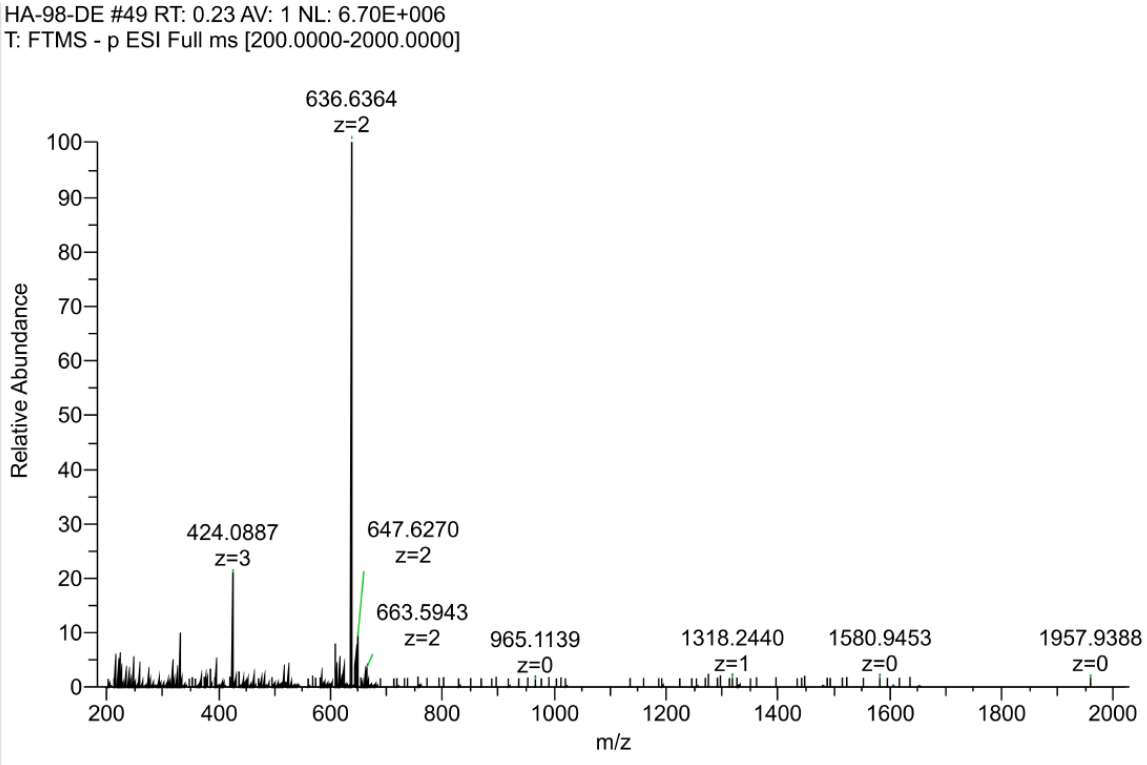


Figure S254. HRMS (ESI) of 4-mer FdZ[C^{N3}(-2),H^{E}(+1)].

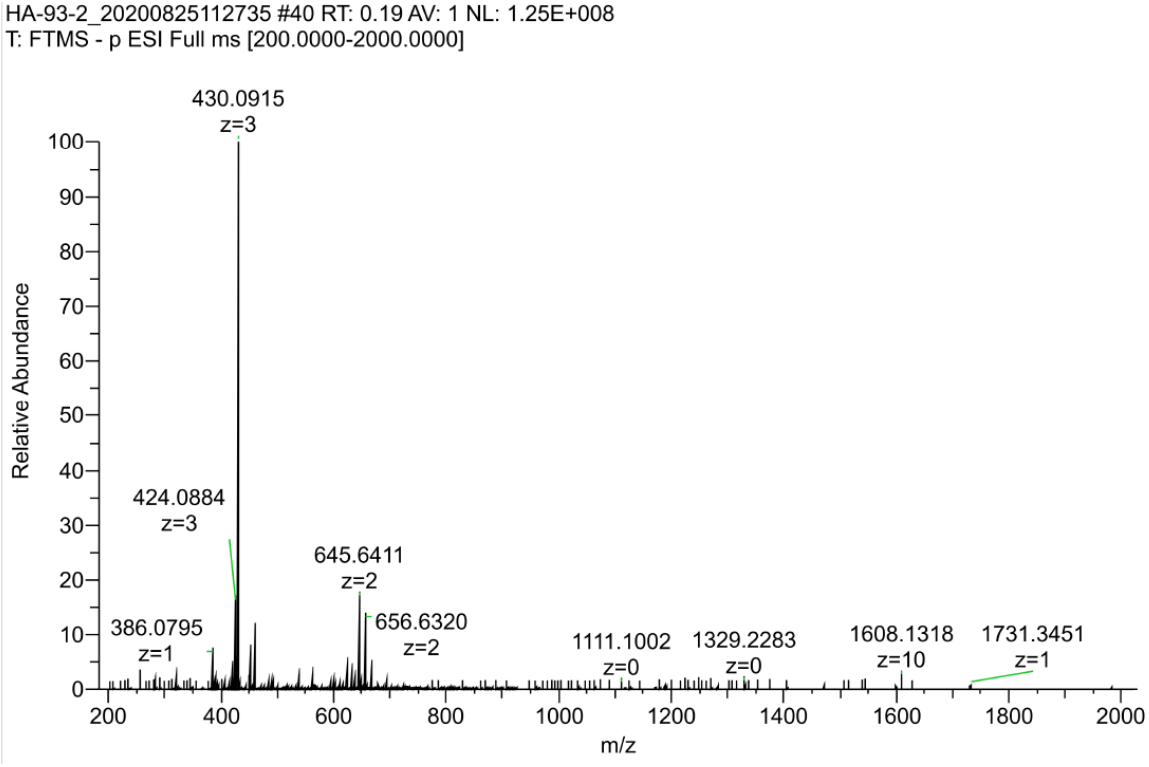


Figure S255. HRMS (ESI) of FdZ-hydration of the oligo 4-mer FdZ[C^{N3}(-2),H^{E}(+1)].

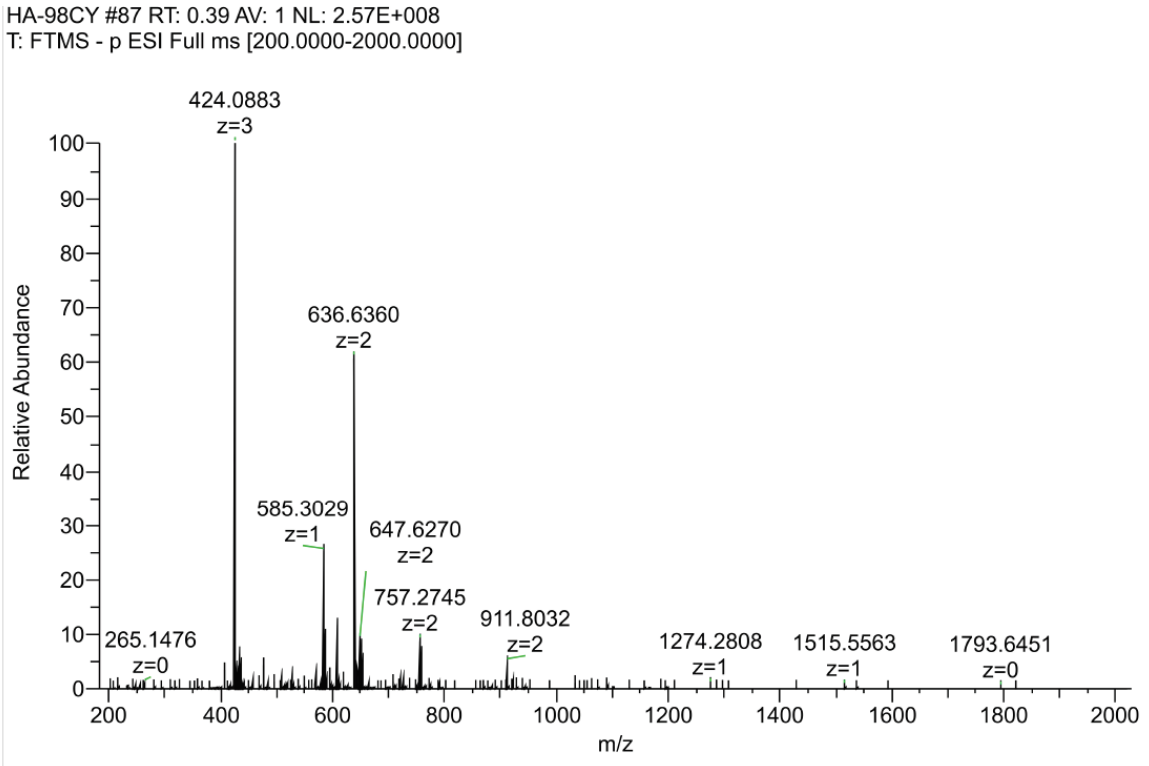


Figure S256. HRMS (ESI) of 4-mer FdZ[C^{N3}(-2),H^{E}(+1)]X.

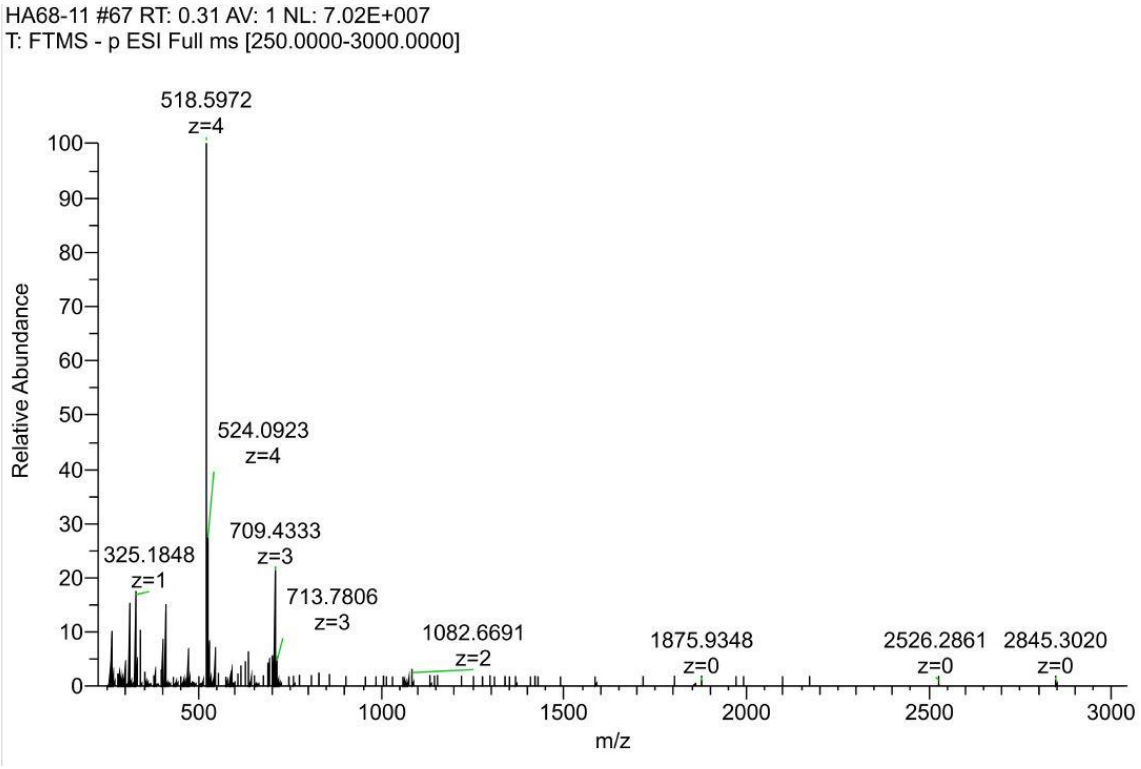


Figure S257. HRMS (ESI) of 9-mer ddiazep.

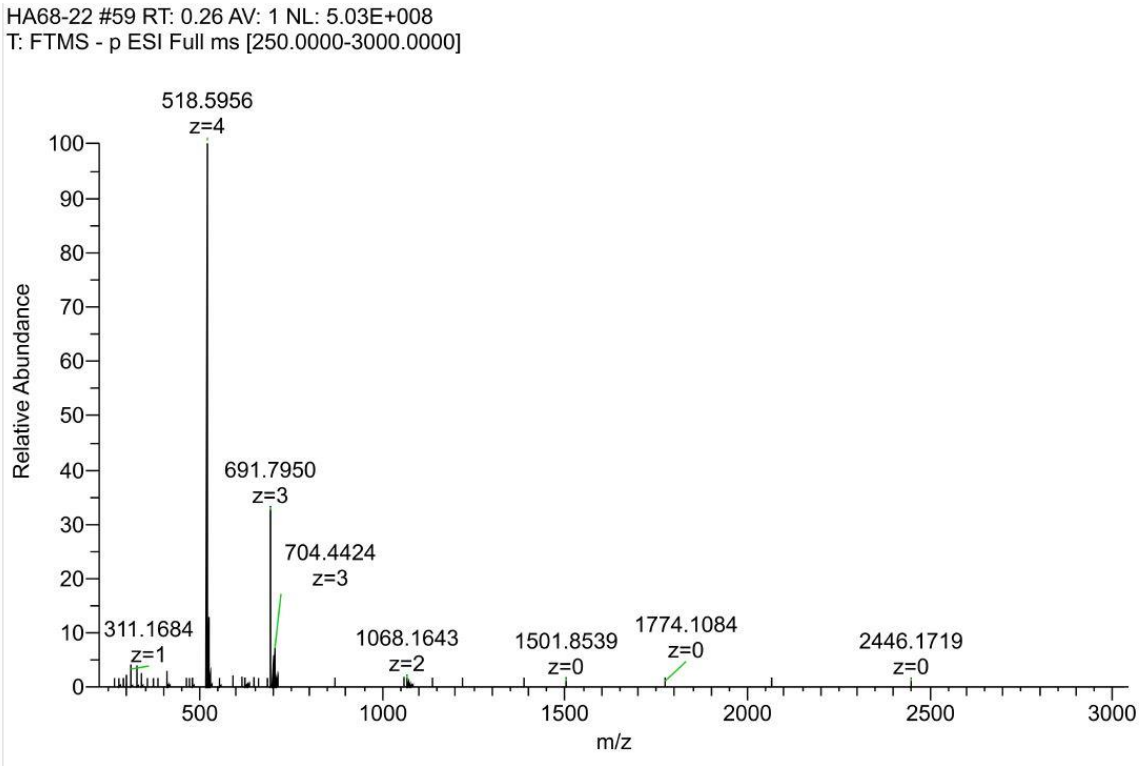


Figure S258. HRMS (ESI) of 7-mer ***R*-dazep**.

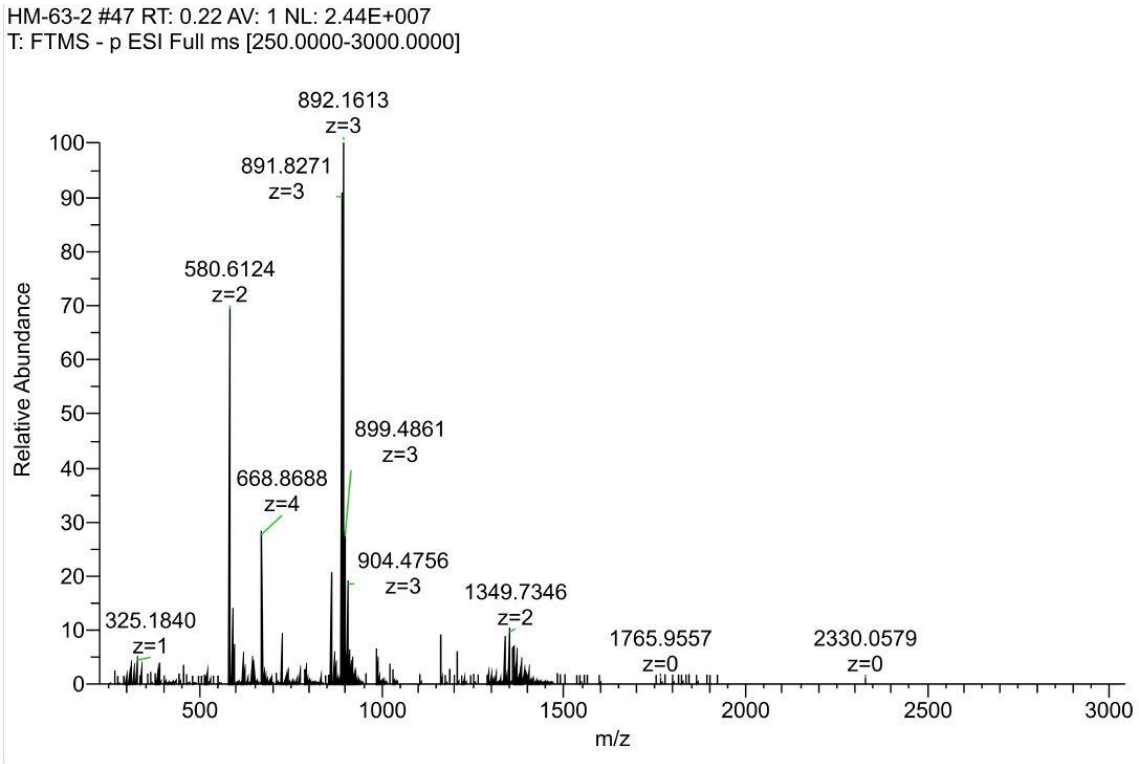


Figure S259. HRMS (ESI) of 7-mer ***S*-dazep**.