Repair of modified O^4-alkyl-pyrimidines by O^6-Alkylguanine DNA alkyltransferases

Lauralicia Sacre

A Thesis
in
The Department
of
Chemistry and Biochemistry

Presented in Partial Fulfillment of the Requirements
for the Degree of Master of Science (Chemistry) at
Concordia University
Montréal, Québec, Canada

March 2017

CONCORDIA UNIVERSITY

School of Graduate Studies

This is to certify that the thesis prepared

By: Lauralicia Sacre

Entitled: Repair of modified O^4-alkyl-pyrimidines by O^6-Alkylguanine DNA alkyltransferases

and submitted in partial fulfillment of the requirements for the degree of

Master of Science (Chemistry)

complies with the regulations of the University and meets the accepted standards with respect to originality and quality.

Signed by the final examining committee:

Dr. Guillaume Lamoureux_____ Chair

Dr. Sebastien Robidoux _____ Examiner

Dr. Joanne Turnbull_____ Examiner

Dr. Christopher Wilds_____ Supervisor

Approved by _____

Chair of Department or Graduate Program Director

April 14 2017

_____

Dean of Faculty

ABSTRACT

Repair of modified O^4-alkyl-pyrimidines by
O^6-Alkylguanine DNA alkyltransferases

Lauralicia Sacre

O^4-Alkyl lesions are a form of DNA damage that occurs due to exposure to some alkylating agents, which can be repaired by O^6-Alkylguanine DNA alkyltransferases (AGTs). AGTs from different species differ in their substrate specificity and proficiency towards repair of these lesions. O^4-alkyl-pyrimidines with C5-modifications were prepared by using a combination of small molecule and solid phase synthesis in order to investigate their repair by different AGT homologues such as human AGT (hAGT), *E. coli* OGT and Ada-C and a hAGT/OGT chimera (hOGT). Studies have shown that hAGT is less efficient at removing O^4 lesions than the *E. coli* variants. Previous work from our group revealed that replacing the C5-methyl group with hydrogen results in an increase in the proficiency of repair of O^4-alkyl lesions observed by the human homologue. It was observed that the presence of a C5-fluorine atom resulted in a futher increase in repair of a O^4-methyl lesion, particularly by hAGT and the chimera. Repair of the O^4-methyl lesion increased about 20-fold after 1 min with 5 equivalents of AGT when fluorine rather than hydrogen was at the C5-position. Repair of larger lesions at the O^4 position such as ethyl, benzyl, 4-hydroxybutyl and 7-hydroxyheptyl were also evaluated for analogs with fluorine at the C5 position. Interestingly, all substrates were repaired efficiently. Ada-C was able to repair larger adducts such as O^4-benzyl and O^4-4-hydroxybutyl. Human AGT showed a preference for the benzyl adduct, as observed previously in the literature. Smaller adducts (such as O^4-ethyl) were repaired faster than the larger ones by the various AGTs with O^4-7-hydroxyheptyl repair observed to occur more efficiently relative to its 4 carbon analogue presumably due to a more flexible chain allowing for a more optimal accommodation for the removal of the lesion. Other halogens such as chlorine and bromine at the C5 position were

also analyzed to observed their influence relative to fluorine. Repair of O^4-methyl lesions with chlorine at the C5-position occurred faster relative to bromine. The larger C5-bromine did influence repair with a maximum of 55% removal of the O^4-methyl lesion by the AGTs observed. The presence of chlorine at the C5-position had a positive influence on removal of the O^4-methyl group by human AGT with slightly faster repair observed compared to the C5-fluorine analog. Overall, the presence of the C5-fluorine group was shown to have a positive influence on O^4-methyl repair by AGT, distinct from the influence of chlorine or bromine at the C5-position. These studies have expanded our knowledge of the range of subtrates that can be repaired by AGTs and may be useful in guiding the design of hAGT inhibitors as therapeutics against uncontrolled cellular proliferation seen in cancer.

DEDICATION

First, I would like to thank my supervisor, Dr. Wilds, for giving me the opportunity to carry out research and work within his group. Dr. Wilds has been a phenomenal mentor and a source of support over the last two years. His passion for science, his encouragement and his kindness create a lab environment that is perfect for students to grow both personally and professionally. I am proud to say that thanks to my experience at his lab, I am a more confident, fearless, judicious and passionate scientist. Dr. Wilds, thank you for being an excellent researcher, professor and human being.

I would also like to thank my committee members, Dr. Sébastien Robidoux and Dr. Joanne Turnbull, who consistently kept up with my research and gave me helpful and critical suggestions. Thank you to all of the lab members, both former and current, for your support, help and advice. The lab would not have been the same without each of you. I am also grateful to my fellow graduate students and the many friends who have provided irreplaceable support and true friendship. Special thanks are due to Dr. Anne Noronha for her help with the ABI 3400 synthesizer, for all of our interesting scientific discussions and for her friendship. The same appreciation goes to Dr. Derek O'Flaherty, who taught and trained me from day one. Thank you for the brainstorming, the advice, the inspiration and the confidence you provided throughout my journey. It has been a great honour to work with and to get to know you all.

Finally, I would like to dedicate my thesis to my family, particularly to my parents. Thank you dad for all of your unconditional support and love. You are, clearly, my biggest inspiration and role model. I hope to one day grow to be as passionate and successful as you are, in life and in work. I also hope to always make you proud. Gracias mami, por todo tu apoyo y por todos tus consejos. Gracias por aguantar mis loqueras y por apoyarme en mis decisiones. Te amo mami. Thank you to my siblings: Ricardo, Lourdes and Greta, for all of the support and wonderful memories. Los amo familia.

ACKNOWLEDGEMENTS

I am grateful to Alain Tesser for his assistance with the mass spectrometry analysis of the different oligomers, to Dr. Marcos Di Falco for his help with the mass spectrometry analysis of the small molecules, to Dr. Alexey Denisov for his training and help with the NMR instrumentation, to Philippe Archambault for his help in the computational studies regarding the geometry optimization models, to Dr. Enrique Pedroso and Dr. Anna Grandas for the opportunity to do an exchange in their lab in Barcelona and for the compound described in appendix 4, to Concordia University, FRQNT and the NSERC CREATE Bionanomachines program for their respective fellowships.

Contribution of Authors

CHAPTER 2

Sacre, Lauralicia: Wrote the first draft of the manuscript and will make the suggested corrections. Performed all the experiments.

Wilds, Christopher J.: Project supervisor responsible for correcting the manuscript.

CHAPTER 3

Sacre, Lauralicia: Wrote the first draft of the manuscript and will make the suggested corrections. Performed the remainder of experiments, which include small molecule synthesis and full characterization of compounds 2a, 2b, 3a, 3b, 4a, and 5a; purification and snake venom analysis of oligonucleotides; biophysical studies (T_m's and CD); repair assays.

O'Flaherty, Derek K.: Perfomed the small molecule synthesis and full characterization of compounds 4b and 5b.

Wilds, Christopher J.: Project supervisor responsible for correcting the manuscript.

CHAPTER 4

Sacre, Lauralicia: Wrote the first draft of the manuscript and will make the suggested corrections. Performed the remainder of experiments, which include small molecule synthesis and full characterization of compounds 3a, 4a, 5a, and 6a; purification and snake venom analysis of oligonucleotides; biophysical studies (T_m's and CD); repair assays

Bahsoun, Yehya: Perfomed the small molecule synthesis 5-bromo-O^4-methyl-2'-deoxyuridine and characterization of 5-bromo-O^4-methyl-2'-deoxyuridine and its intermediates (3b, 4b, 5b and 6b).

Wilds, Christopher J.: Project supervisor responsible for correcting the manuscript.

Table of Contents

List of figures……..…..xi

List of schemes.....……xii

List of tables.....……xiii

List of abbreviations……..….xiv

CHAPTER 1. Introduction……..1

1.1 DNA Alkylation……1

1.1.1 Alkylating Agents……1

1.1.2 Sources of Alkylation……2

1.1.3 Predominant sites and Consequences of DNA Alkylation……3

1.2 DNA Repair Mechanism.……4

1.2.1 Direct Repair……5

1.2.2 Direct Repair by AGTs……5

1.3 *O*^{6}-Alkylgunanie DNA Alkyltransferase……6

1.3.1 hAGT structure and mechanism ……7

1.3.2 AGT homologues and their substrate specificity……8

1.3.3 DNA damage detection and binding to DNA……9

1.3.4 Key AGT residues……10

1.4 C5 Modifications: CH_3 vs H……11

1.5 Introduction of fluorine in DNA……12

1.5.1 Propierties of fluorine……12

1.5.2 Inductive and resonance effects……13

1.5.3 Fluorine substitutions……14

1.5.4 Potential Interactions of fluorine with proteins……15

1.5.5 Other Halogens: Fluorine vs Chlorine vs Bromine……15

1.6 Objectives……17

CHAPTER 2……

2. Fluorine at the C5 position of 2'-deoxyuridine enhances repair of *O*^{4}-methyl adducts by human *O*^{6}-alkylguanine DNA Alkyltransferase……**18**

2.1 Abstract……**18**

2.2 Introduction……**18**

2.3 Results and Discussion……**21**

2.3.1 Synthesis of nucleosides and oligonucleotides……21

2.3.2 UV thermal denaturation and CD spectroscopy……23

2.3.3 Repair Assays……25

2.4 Conclusion……**29**

CHAPTER 3……

3. Influence of fluorine at the C5 position of various *O*^{4}-alkylated 2'-deoxyuridines on repair by *O*^{6}-alkylguanine-DNA alkyltranferases……**30**

3.1 Abstract……**30**

3.2 Introduction……**30**

3.3 Results and Discussion……**33**

3.3.1 Synthesis of nucleosides and oligonucleotides……33

3.3.2 UV thermal denaturation ……36

3.3.3 CD spectroscopy……37

3.3.4 Repair Assays……38

3.4 Conclusion……**45**

CHAPTER 4……

4. Influence of halogens at the C5-position of O^4-alkyl pyrimidines towards repair by O^6-alkylguanine-DNA alkyltransferases……..47

4.1 Abstract……..47

4.2 Introduction……47

4.3 Results and Discussion……50

4.3.1 Synthesis of nucleosides and oligonucleotides……50

4.3.2 UV thermal denaturation and CD spectroscopy……52

4.3.3 Repair Assays……54

4.4 Conclusion……60

CHAPTER 5: Conclusions and Future Directions……62

5.1 Conclusions……62

5.2 Future Directions……68

Refecences……70

Appendix I: Supporting Information for CHAPTER 2……80

Appendix II: Supporting Infromation for CHAPTER 3……108

Appendix III: Supporting Information for CHAPTER 4……167

Appendix IV: Formation of small circular oligonucleotides by using a removable crosslink……212

List of figures

Figure 1.1 Examples of Alkylating Agents……2
Figure 1.2 *O*-Alkylations in dG and T……3
Figure 1.3 Watson-Crick base pairing of dA-T, dA-O^4MeT and dG-O^4MeT……4
Figure 1.4 hAGT Domains……6
Figure 1.5 hAGT secondary structure……7
Figure 1.6 Repair mechanism by hAGT……8
Figure 1.7 Structures of O^4-Alkyl-pyrimidines……12
Figure 1.8 Structures of C5-Halogenated pyrimidines……17
Figure 2.1 Structures of dFU-Me, dU-Me, T-Me and the DNA sequence……21
Figure 2.2 T_m values of O^4-methylated pyrimidines……24
Figure 2.3 CD profiles of O^4-methylated pyrimidines…… 25
Figure 2.4 Total repair of dFU and dFU-Me ……26
Figure 2.5 Time course repair assay of dFU-Me by hAGT, OGT, Ada-C and hOGT ……27
Figure 3.1 Structures of modified 5-fluoro-O^4-alkyl-2'-deoxyuridine and the DNA sequence…33
Figure 3.2 T_m values of dFU-Et, dFU-Bn, dFU-4COH, and dFU-7COH……37
Figure 3.3 CD profiles of dFU-Et, dFU-Bn, dFU-4COH, and dFU-7COH…… 38
Figure 3.4 Time course repair assay of dFU-Et, dFU-Bn, dFU-4COH, and dFU-7COH……41
Figure 3.5 Favourable orientation of Pro140 and Tyr158 with respect to C5-F……42
Figure 3.6 Distance from C145 sulphur with respect to C5-F and O^4 atom……43
Figure 3.7 Repair of dFU-Et, dFU-Bn, dFU-4COH, and dFU-7COH at specific time points……45
Figure 4.1 Structures of dClU-Me, dBrU-Me, and the DNA sequence……50
Figure 4.2 T_m values of dClU, dClU-Me, dBrU, and dBrU-Me……53
Figure 4.3 CD profiles of dClU, dClU-Me, dBrU, and dBrU-Me ……54
Figure 4.4 Time course repair assays of dClU-Me and dBrU-Me……55
Figure 4.5 Possible interactions of C5-Cl with hAGT's active site residues……57
Figure 4.6 Possible interactions of C5-Br with hAGT's active site residues……58
Figure 4.7 Possible favourable interaction of C5-Cl with Tyr158……59
Figure 4.8 Comparison of the repair between dFU-Me, dClU-Me and dBrU-Me……60
Figure 5.1 T_m values of unmodified duplexes containing T, dU, dFU, dClU and dBrU……63
Figure 5.2 Possible interaction of C5-F in hAGT's active site pyrimidines……65
Figure 5.3 Structure of 5-iodo-O^4-methyl-2'-deoxyuridine and 5-trifluoromethyl-O^4-methyl-2'-deoxyuridine……68
Figure 5.4 Structure of FFAU and FMAU……69

List of schemes

Scheme 2.1 Synthesis of 5-fluoro-O^4-methyl-2'-deoxyuridine……23

Scheme 3.1 Synthesis of 5-fluoro-O^4-alkyl-2'-deoxyuridine adducts……36

Scheme 4.1 Synthesis of dClU-Me and dBrU-Me adducts……52

List of tables

Table 5.1. Relative repairs of different T substrates by different AGTs……..66

Table 5.2. Relative repairs of different dU substrates by different AGTs……67

Table 5.3. Relative repairs of different dFU substrates by different AGTs……67

List of abreviations

AcOH - Acetic Acid
AGT - O^6-Alkylguanine-DNA Alkyltransferase
Ar - Argon
ATP - Adenosine Triphosphate
BCNU - 1,3-bis(2-chloroethyl)-1-nitrosourea
BER - Base Excision Repair
CD - Circular Dichroism
CENU - Chloroethylnitrosourea
dA - 2'-deoxyadenosine
DBU - 1,8-Diazabicycloundec-7-ene
dBrU - 5-bromo- 2'-deoxyuridine
dBrU-Me - 5-bromo-O^4-methyl-2'-deoxyuridine
dC - 2'-deoxycytidine
DCM - Dichloromethane
dClU - 5-chloro-2'-deoxyuridine
dClU-Me - 5-chloro-O^4-methyl-2'-deoxyuridine
dFU - 5-fluoro-2'-deoxyuridine
dFU-Bn - 5-fluoro-O^4-benzyl-2'-deoxyuridine
dFU-C2 - 5-fluoro-O^4-ethyl-2'-deoxyuridine
dFU-4COH - 5-fluoro-O^4-(4-hydroxybutyl)-2'-deoxyuridine
dFU-7COH - 5-fluoro- O^4-(7-hydroxyheptyl)- 2'-deoxyuridine
dFU-Me - 5-fluoro-O^4-methyl-2'-deoxyuridine
dG - 2'-deoxyguanosine
DIPEA - Diisopropylethylamine
DMAP - N,N'-Dimethylaminopyridine
DMTr - 4,4'-Dimethoxytrityl
dsDNA - double stranded 2'-deoxyribonucleic acid
dT - deoxythymidine
ENNG - N-Ethyl-N'-nitro-N-nitrosoguanidine
ESI-MS - Electrospray Ionization Mass Spectrometry
Esu - Electrostatic unit of charge
2'-F'ANA - 2'-deoxy-2'-fluoro- β-D-arabinonucleic acid
FFAU - 2'-fluoro-5-fluoro-1-β-D-arabinofuranosyluracil
FIAU - 2'-fluoro-5-iodo-1-β-D-arabinofuranosyluracil
FMAU - 2'-fluoro-5-methyl -1-β-D-arabinofuranosyluracil
HCl - Hydrochloric Acid
MeOH - Methanol
$MgCl_2$ - Magnesium Chloride
$MgSO_4$ - Magnesium Sulfate
MMNG - *N*-Methyl-*N*'-nitro-*N*-nitrosoguanidine
MMR - Mismatch Repair

NaOMe - Sodium methoxide
$NaHCO_3$ - Sodium Carbonate
NER - Nucleotide Exicision Repair
nt - nucleotide
Pac-Cl -Phenoxyacetyl chloride
PAGE - Polyacrylamide gel electrophoresis
PNK - Polinucleotide Kinase
$POCl_3$ - Phosphoryl chloride
SAX-HPLC - Strong anion exchange High Performance Liquid Chromatography
S_N1 - Nucleophilic Substitution, First Order
S_N2 - Nucleophilic Substitution, Second Order
ssDNA - single stranded 2'-deoxyribonucleic acid
T - Thymidine
TBAF - Tetrabutylammonium fluoride
TBS - t-butyldimethylsilyl
TEA - Triethylamine
THF - Tetrahydrofuran
T_m - Thermal melt
UV - Ultraviolet
vdWr - van der Waals radius

Chapter 1. Introduction

1.1 DNA Alkylation

DNA is a complex macromolecule that carries our genetic information. Every day, DNA is exposed to endogenous and exogenous agents that can cause damage. Depending on the damaging agent, different lesions such as alkylations can be produced. These lesions could be detrimental if they do not get repaired as they could eventually lead to mutations or modifications in the DNA. DNA alkylation can result from different sources and it could occur at the atoms that are more nucleophilic. Most of the resulting adducts are mutagenic or cytotoxic, and therefore organisms have developed different mechanisms to repair the vast range of alkyl lesions.

1.1.1 Alkylating Agents

Alkylating agents allow the formation of different covalent adducts between their electrophilic moieties and the nucleophilic atoms in DNA. Alkylating agents can be classified as monofunctional or bifunctional agents which undergo one or two alkylating events, respectively.[1] They can also be divided according to the mechanism of the alkylation reaction, which can be either S_N1 (unimolecular) or S_N2 (bimolecular).[2] This final classification depends on the specific alkylating agent. In these nucleophilic substitutions, the S_N2 rate depends on the concentration of both the electrophile and the nucleophile (nucleophilic centers in DNA), while the S_N1 rate only depends on the electrophile (alkylating agent). Therefore S_N2 alkylating agents are more selective with regards to the positions in DNA with which they react. Monofunctional alkylating agents predominantly act upon the N-atoms of the nucleobases, capable of introducing a range of lesions including methyl, ethyl and other adducts. Examples of these are N-methyl-N'-nitro-N-nitrosoguanidine (MNNG), N-ethyl-N'-nitro-N-nitrosoguanidine (ENNG), N-methyl-N'-nitrosourea, which operate via a S_N1 mechanism whereas dimethyl sulfate and methylmethane sulfonate reaction by a S_N2 process.[3–6]

Bifunctional alkylating agents tend to be more damaging and lethal for cells as they can potentially generate numerous products including interstrand (involving complementary DNA strands) and intrastrand (occurring at the same DNA strand) cross-links (secondary product) in addition to mono-adducts (primary product).[1,7] Examples of bifunctional alkylating agents include derivatives of chloroethylating agents like chloroethylnitrosourea (CENU) and alkyl sulfonates such as busulfan and hepsulfam (S_N2 reagents).[1]

1.1.2 Sources of Alkylation

Alkylating agents can be endogenous or exogenous to the cell. Endogenous sources include S-adenosylmethionine, which acts as a methyl donor in many cellular reactions (Figure 1.1 A).[3] Studies have shown that S-adenosylmethionine could produce O^6MedG (about 10-30 events per day).[8] Exogenous sources involve a long list of nitrosoureas and nitrosamines that can be found in alcohol and tobacco.[3,9,10] A high percentage of these exogenous sources can also introduce lesions at the oxygen atoms of the nucleobases such as O^6 and O^4 atoms of 2'-deoxyguanosine and thymidine, respectively.

The alkylation of DNA can be exploited to provide a therapeutic benefit. This is the case of some cancer therapeutics, where the DNA alkylation is desired. The introduction of alkyl lesions in order to introduce DNA damage can lead to the death of cancer cells. Temozolamide, which is a methylating agent that generates DNA mono adducts, is employed in treatment of brain tumors (Figure 1.1 C). Bisalkylating agents are also used in treatments against cancer. One example of these is the alkyl sulfonate busulfan. [11,12]

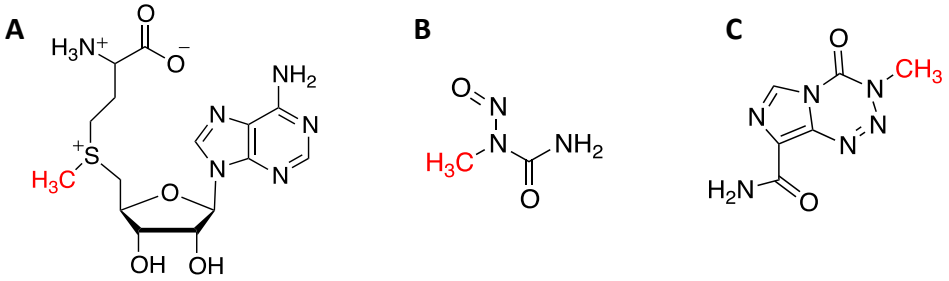


Figure 1.1: Alkylating agents such as (**A**) S-Adenosylmethionine, (**B**) *N*-methyl-*N*'-nitrosourea and (**C**) Temozolamide. The electrophilic methyl group is shown in red.

1.1.3 Predominant sites and consequences of DNA alkylation

Alkylation with electrophiles occurs on the electron rich atoms (nucleophilic atoms) of DNA; therefore it can occur at the DNA backbone or the nucleobases especially at the N- and O-atoms. Certain atoms and positions are more susceptible than others to alkylation. Alkyl lesions can have mutagenic, cytotoxic and even carcinogenic consequences.[13] For instance, lesions at the atoms involved in Watson-Crick base pairing tend to be more mutagenic.[14,15]

The most vulnerable atom for alkylation is N7 of guanine.[16] For instance when DNA is treated with simple methylating agents, around 80% of the methylation occurs at the N7 position.[1,4] Alkylations at the N7 position predominantly undergo depurination.[4] The second most predominant site for alkylation is N3 of adenine. This type of lesion has cytotoxic effects as it could cause the blockage of DNA replication.[4,17] The repair of these N-alkylated purines occurs normally through base excision repair (BER).[1]

Other types of alkyl lesions tend to have a higher mutagenicity effect. This is the case of the *O*-alkylations, particularly O^6MedG and O^4MeT (Figure 1.2). O^6 alkylation tends to occur in more significant amounts compared to alkylation at the O^4-atom. Analysis of DNA exposed to different methylating agents such as N-methyl-N'-nitrosourea (MNU) revealed that among the total DNA adducts about 8% and less than 0.4% corresponded to O^6MedG and O^4MeT respectively.[16,18,19]

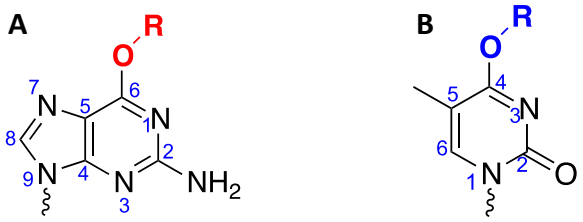


Figure 1.2: Alkylations in (**A**) dG and (**B**) T at the O^6 and O^4 positions, respectively.

The mutagenicity of the lesions not only depends on how often they occur but in how efficiently they get repaired by the organism and their persistence. In the case of O^6MedG, if it

is not removed by hAGT (to be introduced later on in this thesis), it has the possibility to cause a dG → dA transition where now it can base pair with T (T → dC transition shown in Figure 1.3). This O^6MedG-T pair can be recognized by the MMR (Mismatch Repair) pathway and two possible scenarios could develop; (1) If the lesion occurs on the non-template strand, it can be removed and a point mutation will be generated as now the template strand has a thymidine rather than a 2'-deoxycytidine or (2) If the lesion is on the template-strand, the MMR machinery will keep incorporating T and recognizing it again as a mismatch entering into a futile cycle that eventually signals for apoptosis.[5]

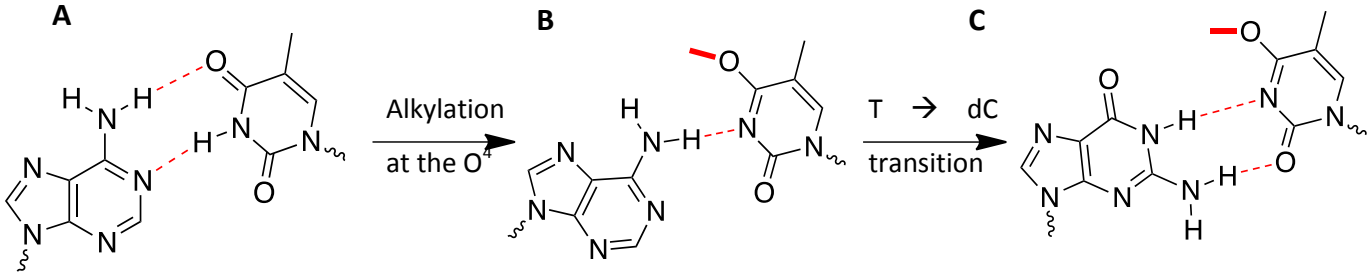


Figure 1.3: Watson-Crick base pairing of (**A**) dA•T, (**B**) dA• O^4-MeT and (**C**) dG•O^4-MeT

Different studies in mammalian cells have shown that O^4MeT lesions persist, positioning this lesion in a higher level of toxicity (which could generate T → dC transitions) compared to O^6MedG.[14] This is due to the fact that mammalian DNA repair mechanisms are not efficient at removing this lesion. Studies have revealed that O^4MeT is neutral to the presence or absence of hAGT and does not get excised by the MMR pathway. On the other hand, the NER (Nucleotide Excision Repair) pathway can restore DNA containing this lesion.[20–23]

1.2 DNA Repair mechanisms

Organisms have evolved different repair pathways in order to fight against modifications to DNA. Depending on the type of lesion, different DNA damage repair mechanism would be used. The pathways include base excision repair (BER), nucleotide excision repair (NER), mismatch repair (MMR), homologous recombination, non-homologous end joining and direct repair.

The importance of research in this field was recognized by the scientific community with the 2015 Nobel prize in chemistry awarded to Lindahl, Modrich and Sancar for their respective work and mechanistic studies in BER, NER and MMR pathways. BER is in charge of removing small and non-bulky modifications on the heterocycle. This involves repair of some N-alkylated purines (as previously mentioned) and it plays a backup role for repairing O^6MedG.[1] NER is responsible for the removal of bulky and small adducts including O^4-alkyl-T lesions.[24–26] Finally, MMR focuses on detecting and correcting the improper incorporation of nucleotides. As stated before, MMR is able to recognize the mismatch which occurs with O^6-alkyl-dG.[5]

1.2.1 Direct Repair

O^6-alkyl-dG lesions, and to a lesser extent O^4-alkyl-T lesions, can be processed in mammalian systems by the direct repair pathway. Direct repair acts upon relatively specific forms of damage and does not involve excision of the DNA strand where the lesion is present. In humans, there are two direct repair proteins. The first class corresponds to the homologs of AlkB protein, ABH2 and ABH3, which perform the direct dealkylation through an oxidative process.[4,27,28] These Fe^{II}-dependent dioxygenases are specialized in repairing alkyl lesions (particularly methylation) at *N*1 of adenine and *N*3 of cytosine yielding the repaired base and formaldehyde as the other product.[4,27,28]

The second class are the O^6-Alkylguanine DNA Alkyltansferases (AGTs). AGTs are the proteins that the different projects in this thesis focus on, specifically human AGT (hAGT), the two variants from *E. coli* (OGT and Ada-C) and a chimera hOGT (discussed later on text).

1.2.2 Direct repair by AGTs

O^6-Alkylguanine DNA Alkyltansferases (AGTs) perform a direct covalent transfer of an alkyl group from the damaged base to a nucleophilic cysteine residue present in the active site of the protein (Figure 1.6). This direct transfer is irreversible and at the end it generates an alkylated protein (which now is deactivated) and the restored DNA. AGTs execute a stoichiometric reaction because one AGT protein is required for the removal of one alkyl lesion.

They are specialized in repairing the alkylations occurring at the O^6-position of 2'-deoxyguanosine and, to a variable extent, the O^4-position of thymidine.[28–30]

1.3 O^6-Alkylguanine DNA Alkyltransferases

AGTs are relatively small proteins, normally around 19-21kDa.[31] They are present in almost all kingdoms of life. Even though AGTs from different species have a similar overall structure, their sequence identity is low. Nonetheless, these proteins have conserved important features such as a PCHR motif present in the active site. They have a helix-turn-helix DNA domain (Figure 1.5) and differ from other DNA binding proteins in the way that they use the minor groove of DNA for the binding interaction.[29,32]

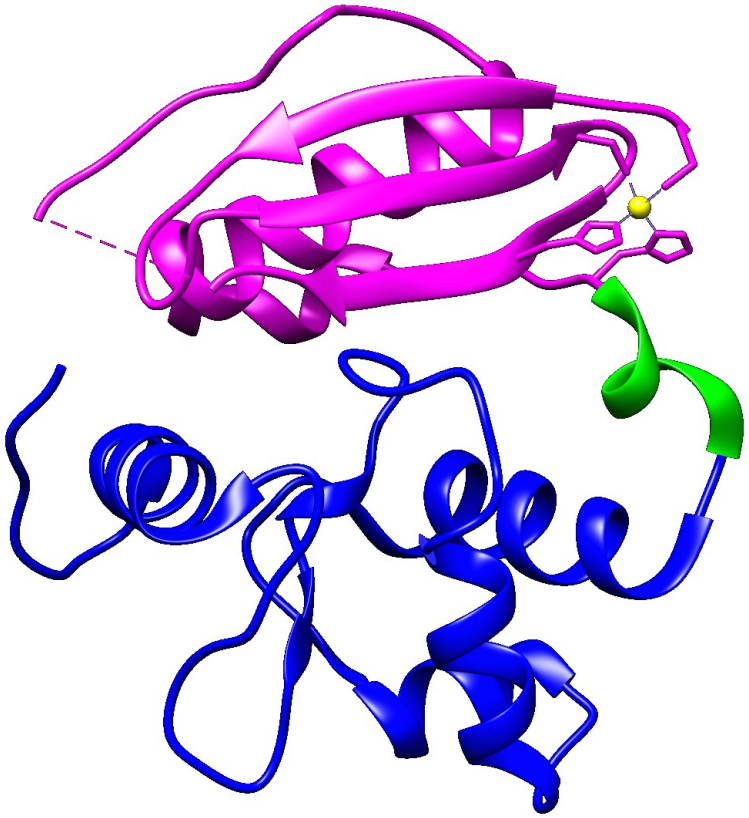


Figure 1.4: hAGT N-terminal (magenta), Zinc ion (yellow), Helix-3 connecting N- and C-terminal (green) and C-terminal (blue) (Model prepared by PyMol using entry PDB 1EH6)[33]

1.3.1 hAGT structure and mechanism

The AGT protein consists of N- and C-terminal domains. The N-terminal domain has a structural role and in humans it contains a zinc ion (shown in Figure 1.4), which aids to stabilize the interface between the two domains.[33,34] The C-terminal domain is where the active site is located as well as the DNA binding region. The DNA binding motif is known as the helix-turn-helix motif (HTH). In hAGT, the first helix of the domain has residues Tyr114-Ala121 and the second helix comprises Ala127-Gly136. This second helix is involved in recognition and interacts with the DNA.[34,35]

The active site and the HTH domain are linked by residue Asn137. The active site pocket, containing the conserved PCHR sequence, is situated near the bottom of the solvent accessible groove forming the binding channel (Figure 1.5). This grove is approximately 8 Å wide, 9 Å deep and 14 Å long, giving it a tunnel-like shape.[33,35] Once in the hAGT active site, the damaged base is surrounded by Pro140, Tyr 158 and Ser159, which constitutes part of the pocket's walls.[34]

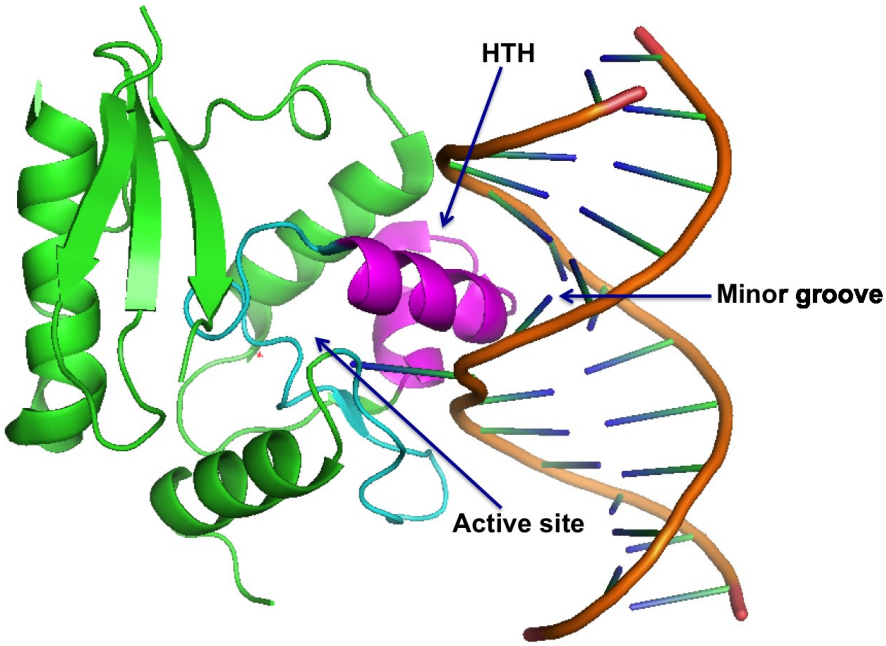


Figure 1.5: hAGT secondary structure showing the HTH domain (magenta), the active site (cyan) and the binding to the DNA (orange) at the minor groove (Model prepared by PyMol using entry PDB 1T38)[33]

The repair reaction performed by AGTs occurs in a S_N2 fashion (Figure 1.6). It begins with the damaged nucleotide being flipped out of the DNA duplex through the rotation of the 3' phosphate.[29,35] There is a Glu-His-water-Cys hydrogen bond network that increases the reactivity of Cys145 (in hAGT) in the active site via a proton transfer, so the second step of nucleophilic attack of the thiolate on the alpha-carbon attached to the nucleobase can occur. Two products are produced; the irreversibly alkylated protein and the restored DNA.[29,35]

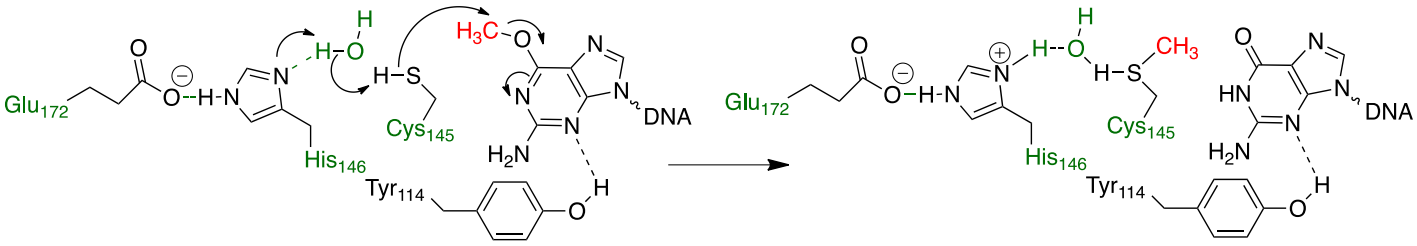


Figure 1.6: Repair mechanism of O^6-Methyl-dG by hAGT (reproduced from Pegg 2011 Toxicol.)

Alkylated AGT, which now is inactive, can interfere with the task of other active AGTs, as the ability to bind DNA does not get affected even when inactivated. Consequently, they are substrates for ubiquitin ligases causing them to get ubiquitinated and later on degraded by the proteasome.[29,36–38]

1.3.2 AGT homologues and their substrate specificity

E. coli has two AGT variants. The first one to be characterized and with a known crystal structure was Ada (39kDa), an adaptive response protein that is highly inducible to alkylation damage.[29,39] Only its C-terminal domain (19 kDa) has direct repair activity and is selective towards the removal of alkyl lesions.[39] Ada-C has a preference for repairing smaller adducts such as methyl groups due to its relatively small active site pocket compared to other AGTs. The presence of tryptophan 161 in the active site causes steric exclusion of larger groups.[31] Ada-C also contains 2 cysteine residues that can perform direct repair but it is Cys321 that is responsible for removal of the alkyl groups from the O^6- and O^4-atoms of dG and T respectively, favoring O^6-dG lesions.[13,31,40]

OGT is the second variant and in contrast to Ada, it is not inducible and normally present in small amounts (about 30 molecules per cell).[39,41–43] Currently, there is no crystal structure available of this 19 kDa protein.[39] OGT is able to remove O^6MedG and O^4MeT lesions efficiently, giving preference to the latter. This variant has demonstrated in multiple studies to be more tolerant of bulkier substrates than Ada-C and more efficient at repairing both smaller adducts and larger alkyl chains (at the O^4-position of T) relative to other AGTs. [30,43,44]

Given the proficient repair of O^4-MeT observed by OGT compared to the human analogue hAGT, studies to replace certain amino acids by site directed mutagenesis to make hAGT more OGT-like have been performed. McManus prepared a chimera of hAGT and OGT (termed hOGT) mutating residues 139-159 of hAGT with those present in OGT (residues 133-153) while maintaining hAGT's Pro140 residue (aiding in solubility of the protein).[45] This work was inspired by a previous study of the Pegg group, who mutated 8 different residues of hAGT to the analogous residues from OGT in order to generate hAGT variants capable of repairing O^4-MeT lesions however it presented some solubility issues.[46] Our group's hOGT chimera demonstrated the ability to repair the same adducts as OGT as well as efficiency of repairing larger adducts in a manner similar to hAGT.

Finally, it has been reported that hAGT is capable of repairing small adducts and demonstrates a pronounced preference (35-fold higher) for repairing O^6-dG over O^4-T lesions.[43] These studies have been performed on modified bases and substrates incorporated in a DNA sequence. In general these AGTs have an inclination for the lesions when they are in dsDNA but still they are able to repair when they are in ssDNA.[40]

1.3.3 DNA damage detection and binding to DNA

How AGTs are able to detect the damaged base has been a question of interest and multiple hypotheses have been proposed. AGT binds to DNA and scans through it in a 5' to 3' direction preferentially searching for places where the native structure might be destabilized due to base modifications.[29,34] There are three major contacts made by hAGT; one to a

phosphate backbone in the strand containing the flipped base (by the first helix of the HTH), a second one to a phosphate backbone in the complementary strand (by some residues such as Thr95 and Asn123) and the third along the groove (by the second helix of the HTH).[34]

AGTs bind to DNA by using a HTH domain at the minor groove of the duplex. The recognition sequence has hydrophobic residues with small side chains that allow less specific and tight hydrophobic interactions with DNA.[29] The compact size of AGT allows it to occupy an area of approximately 7 base pairs when it binds.[37,47,48] Studies of crystal structures have shown that there is minimal to no conformational change upon the binding with DNA.[33–35,49] Binding of AGTs to the damaged DNA can also have consequences such as interference of the NER pathway.[13]

1.3.4 Key AGT residues and examples of their importance

Various residues are involved in the repair reaction performed by AGT in addition to contributing to substrate specificity and positioning. For example, Tyr114 through steric clashes and charge repulsion might play a role in promoting nucleotide flipping via the rotation of the 3' phosphodiester into the active site. Crystallographic studies illustrate that its hydroxyl moiety interacts via a hydrogen bond with the O^2 atom in cytosine and perhaps the same interaction might be occurring with thymidine.[34,35]

The Arg128 residue provides a cationic finger, which intercalates with DNA through the minor groove, stabilizing the duplex by hydrogen bonding with the unpaired base. One of the protagonist residues is the active site Cys145 which has a low pKa (about 4.5 in hAGT) and demonstrates high reactivity.[50,51] The Cys145 residue performs the nucleophilic attack removing the alkyl lesion from the nucleobase.

Other residues have a less defined function but still play a role in helping to obtain the right positioning of the substrate and shape of the active site. Water molecules interact via hydrogen bonds with the Cys145, Tyr158, His 146 (normally acting as a general base) and Ile141

residues.[33] In addition, proline residues (Pro138 and Pro140) in the human protein allow bigger adducts to enter on its active pocket as they help to defined it and to provide a greater accessibility.[31]

Many residues vary between AGTs and in some instances this plays a role in determining which substrates are repaired. One of the big differences between hAGT and Ada-C is the residues found in the active site. As mentioned previously, the presence of Pro138 and Pro140 allows hAGT's active pocket to tolerate the entrance of large adducts, but in the case of Ada-C, the absence of these residues (replaced by Lys139 and Ala41) causes a relaxation in the helix provoking the pocket to close more and therefore blocking the access to bulky adducts.[33,52] One of substrates that hAGT is proficient in reacting with is O^6-BndG. One of the reasons why this substrate is preferred amongst others is largely due to the hydrophobic stacking interactions that occur between the benzyl moiety and the Pro140 ring. Tyr158 side chain (which tends to stack against smaller O^6-alkyl-dG adducts as well) and the C_β of Ser159 also contribute to these hydrophobic interactions.[33,35] Repair of larger adducts and even crosslinks at the O^6-atom of dG has been observed by hAGT.[53]

Due to steric reasons, Ada-C cannot accomodate large adducts such as O^6-BndG in its active site. The presence of the Trp161 side chain occupies part of the space that larger and bulkier adducts such as a benzyl group would require in order to accommodate or even to enter the active site. Hence, there is only room available for smaller adducts, preferentially repaired by Ada-C.[31] In the case of OGT, it has shown to weakly react with O^6-BndG. The absence of Trp161 in OGT leaves sufficient space in the active site to accommodate the benzyl group. However, the absence of Pro140 removes the favorable hydrophobic interactions between the two moieties. The presence of the second proline residue in OGT, may contribute to a wider active site entrance relative to Ada-C.[52]

1.4 C5 modifications: CH_3 vs H

The priority and efficient repair of hAGT towards O^6-alkyl lesions and its poor capability to deal with O^4-alkyl lesions has come into the attention of some researchers. Recent studies

have shown that the C5 methyl group in T possibly causes some steric effects in the active site of hAGT and other variants, affecting repair. These studies employed the investigation of O^4-alkylated 2'-deoxyuridine series where a hydrogen atom is attached to the C5 position.[54]

Repair of O^4-adducts in 2'-deoxyuridine by the AGTs in this study was dramatically increased. Some of the highlights were that removal of the O^4-methyl lesion by hAGT occurred 25-fold faster when H was present at the C5 rather than CH_3. Similar results were obtained with the other substrates, for example O^4-benzyl dU was repaired efficiently by all AGTs except for Ada-C (repair of 4 carbon and 7 carbon alkylene linkers was also inhibited). OGT mediated repair was the most efficient and it was able to now repair larger adducts such as 4-hydroxybutyl and 7-hydroxyheptyl (showing a 30-fold increase compared to its T analogue), suggesting that the C5 methyl group might not be allowing the proper introduction or accommodation of these substrates.[54]

The differences observed in repair between the AGT homologues rely on their respective active site shapes and sizes. The negative effect of the C5 methyl group was more pronounced for hAGT and the chimera hOGT. The modeling studies with hAGT implies that the C5 methyl group in T may encounter steric clashes with Arg135 and Ser158 which might not allow the proper accommodation and orientation of the lesion.[54]

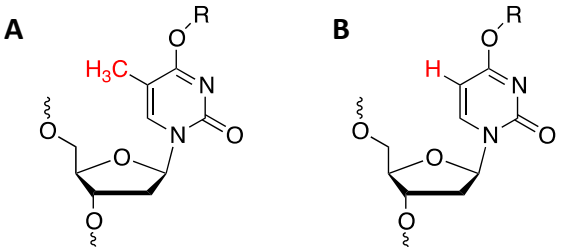


Figure 1.7: Structures of (**A**) O^4-alkyl-thymidine and (**B**) O^4-alkyl-2'-deoxyuridine

1.5 Introduction of fluorine in DNA

1.5.1 Properties of fluorine

Fluorine belongs to the halogen group and is well known for being the most electronegative element. Fluorine has a van der Waals radius of 1.47 Å while hydrogen corresponds to 1.2 Å and oxygen is 1.52 Å.[55–58] Even though fluorine's size more closely resembles oxygen, its relatively small size has made it an attractive candidate to substitute for hydrogen atoms for a number of applications.[59]

Other characteristics of fluorine is that it has a low polarizability, it can increase the acidity of compounds and it has a large electronic effect on residues that are nearby.[60] Hydrogen bonds with fluorine (H---F-C bond) are possible but they are much weaker (2.4 kcal/mol) than those occurring with other heteroatoms such as O and N (~ 5.0 kcal/mol).[56,59] Fluorine has been observed to form short contacts with other atoms for example in proteins. These short contacts are dependent on the orientation and distance of both atoms.[61]

1.5.2 Inductive and resonance effects

The electronic properties of fluorine can be categorized as either inductive or resonance effects. According to some Hammett type substituent parameters (σ) determined in various studies, it has been determined that fluorine is always electron withdrawing by the inductive effect (σ_i of 0.52) but also acts as electron donating due to the resonance effect (σ_r of -0.34).[56]

The effect of fluorine also depends on its position on the molecule. For example, in fluorotoluenes, β-fluorination has show to increase the acidity of C-H bonds through both inductive and resonance influence. In the case of α-fluorination, it has shown to decrease C-H acidity when the carbanion is almost in a planar geometry due to the π- π interaction between both lone-pair of electrons causing a destabilization.[56,62]

Studies of different pentafluorobenzene compounds have determined the influence of fluorine towards certain groups according to whether the fluorines are located ortho, meta or para with respect to the substituents. In most of the cases ortho-fluorine with respect to substituents like amino, hydroxy and carboxyl groups, contributes the most to the acidity of the molecule. The preferential positions towards nucleophilic reactions in the fluorobenzene system follow ortho > meta > para and it is mainly governed by the inductive effect (electron-

accepting) with a slight influence of the resonance effect (π-electron donating). Analysis of fluorobenzene also suggests that ortho-fluorine enhances the acidity of the alpha C-H bond as well as meta-fluorine, which displays a greater inductive effect than ortho-fluorine in this case.[62]

1.5.3 Fluorine Substitutions

The use of fluorine in medicinal chemistry and in biological systems has been of great interest as it increases the lipophilicity (like aromatic fluorinations) of molecules as well as, in certain cases, enhancing their biological activity.[56] Hydrophobic effects normally play a major role in protein interactions. It has been demonstrated that the presence of fluorine can cause changes in the basicity and acidity of ligand molecules, especially by affecting the pKa of neighboring atoms.[55,63]

A substitution commonly observed in certain amino acids is the change of a methyl for a trifluoromethyl group. The presence of CF_3 clearly increases the hydrophobic parameter (derived from octanol-water partition coefficients) of that specific residue and gives some increase in the conformational flexibility.[56,59] When CF_3 is compared to CH_3 in multiple crystal structures it has been observed that fluorination increases the steric size of the alkyl group. In this case, the size of CF_3 more closely resembles more to an isopropyl group.[56] For instance when CF_3 is placed in benzene, it has a π -inductive effect (electron-withdrawing by inductive and conjugation effects).[64]

Fluorinated nucleoside analogues have been under study and some of them have been exploited as chemotherapeutics. Fluorine has shown advantageous effects when it is present in DNA. Examples of sites where fluorine has been introduced include the nucleobase and the sugar. Some of the most efficient radiosensitizers are 5-fluoropyrimidines such as 5-fluorouracil (used for treating colorectal and other cancers) and various produgs, with one of their main mechanisms of action being the inhibition of the enzyme thymidylate synthase disrupting DNA biosynthesis and downstream causing cell death.[65–70] A number of fluorine modified oligonucleotides have been developed including the sugar modified 2'-deoxy-2'-fluoro-

arabinonucleic acids (2'-F-ANA) which has found applications in gene silencing therapies and exhibits stability towards nucleases.[71–73]

1.5.4 Potential interactions of fluorine with proteins

The interactions that fluorine can have with different atoms in proteins strongly depend on their alignments and orientation, which compensate for potential electrostatic repulsions. The times fluorine is in short contact with other atoms in proteins have been studied. The fluorophilic group that shows the highest contact with fluorine is sulfur. All of the fluorine-sulfur contacts come from cysteine, whose potentiality for repulsion is perhaps counteracted by an extra weak interaction with the sulfhydryl hydrogen.[61]

The guanidinium group (specifically at the sp2 carbon) of arginine as well as the hydroxyl groups (through sp3 oxygen) of tyrosine and serine (normally in a linear arrangement) have also been shown to be fluorophilic moieties. Other groups that also show important contacts with fluorine are the sp2 N atoms in the aromatic ring of histidine and surprisingly the sp2 hybridized O atom of carboxyl groups, which adopt an orthogonal alignment to minimize repulsion, and non polar hydrogens.[61,74] Alternatively there are groups that do not interact as much with fluorine, probably due to repulsion, such as the N atom of amide linkages.

1.5.5 Other Halogens: Fluorine vs Chlorine vs Bromine

Even though the main focus of the present work is the influence of fluorine, it is of interest to analyze other halogens such as chlorine and bromine. Bromine and chlorine (with van der Waals radii of 1.85 and 1.75 Å, respectively) are larger than fluorine and at the same time they exhibit a increased polarizability, imparting this effect to the molecule to which they are part of.[56,75] When the halogen atoms are linked to an aromatic system, they can participate in both electron accepting (inductive) and electron donating (through conjugation) events. The electronic properties of these halogens suggest that fluorine exhibits a more pronounced electron donating resonance effect (conjugation) compared to the other halogens

(σ_r: -0.23 for Cl and σ_r: -0.19 for Br) probably triggered by the overlapping of the π orbitals as it is the smallest of the series.[64] The inductive effect is highest for fluorine (σ_i: 0.52) but chlorine (σ_i: 0.47) and bromine (σ_i: 0.44) are not far behind.[64]

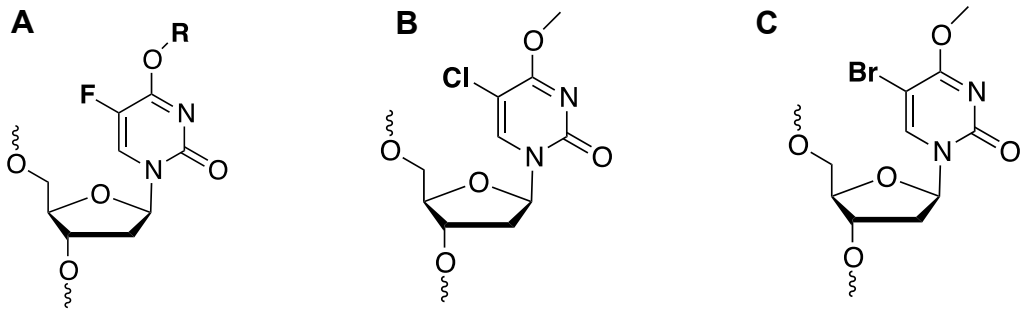
Studies to investigate the effect of different halogens at the C5 of cytosine have shown that the electronic effects can be transmitted through bonds during base pairing. This is the case of C5-fluoro in cytosine, which through its O^2 disfavors the protonation, which can be caused by electrophilic carcinogens, at the N^2 of guanine in the CG base pair.[75] In this particular experiment, the proton transfer from the N^2 atom of G to the O^2 atom of C (which may occur during carcinogen nucleophilic attacks) was used to measure the strength of the C5 group effect. If the electronic effects were dominating this interaction, Br would be less destabilizing than Cl, which turns out to be less destabilizing than F, and with all of these halogens behaving as electron withdrawing substituents.[75]

Computational studies on 5-halouracils have shown that the C5-X bond length increases in going from F (1.34 Å), Cl (1.73 Å) to Br (1.89 Å). On the other hand, the length for the C4-O^4 bond decreases very slightly. Substitution by halogen atoms causes relatively small changes in H-bond distances when it base pairs with adenine, but is still within the ranges for optimal bonding.[65] The charges on the substituted halogens vary from -0.326 esu for fluorine to 0.099 esu for bromine, where it is only F that carries a negative charge.[65]

Fluorine has been said to increase the lipophilic character of molecules. The lipophilic character in halogens also relates to their hydrophobic parameter which increases in the following order: F < Cl < Br.[56,63] Taking advantage of all their properties, the use of halogens in chemotherapeutic agents has gained significant popularity. For example during radiotherapies, halogens are really useful as they are able to localize the influence of radiation due to their large cross-sectional area.[65]

1.6 Objectives

The general objective of this thesis is to study and explore the electronic effects of the C5 group in AGT-mediated repair of various O^4-alkyl modified pyrimidines. The objective for the first project shown in chapter 2 involves studying the repair by different AGTs (hAGT, OGT, Ada-C and the chimera hOGT) of the methyl adduct at the O^4-atom when fluorine is present at the C5 position of 2'-deoxyuridine. Chapter 3 focuses on the effect of C5-fluoro in repair by AGTs of different and larger alkyl lesions at the O^4-position of 2'-deoxyuridine. Finally, chapter 4 evaluates and investigates the electronic influence of other halogens such as chloro and bromo at the C5 position in repair of O^4-methylated pyrimidines by AGTs.



R = Me, Et, Bn, 4COH, or 7COH

Figure 1.8: Structure of (**A**) 5-fluoro-O^4-alkyl-2'-deoxyuridine, (**B**) 5-chloro-O^4-methyl-2'-deoxyuridine and (**C**) 5-bromo-O^4-methyl-2'-deoxyuridine.

Chapter 2. Fluorine at the C5 position of 2'-deoxyuridine enhances repair of O^4-methyl adducts by human O^6-alkylguanine DNA Alkyltransferase

2.1 Abstract

Alkylations at the O^6- and O^4-atoms of 2'-deoxyguanosine (dG) and thymidine (T), respectively, can be repaired by O^6-alkylguanine-DNA alkyltransferases (AGTs). Previous studies have shown that human AGT (hAGT) repairs small adducts poorly at the O^4-atom of T, in comparison to the *E. coli* variants (OGT and Ada-C) of this protein. The C5 methyl group of the thymine nucleobase is suspected to reduce repair efficiencies via possible steric effects. This study examines the "electronic effects" of the substitutent at the C5 position on repair by various AGTs using oligonucleotides containing a 5-fluoro-O^4-methyl-2'-deoxyuridine (**dFU-Me**) insert. The ability of hAGT, *E. coli* variants (OGT and Ada-C) and a chimeric hAGT/OGT protein (hOGT) to mediate repair of the O^4-methyl adduct was investigated. Compared to previous studies with O^4-methyl-2'-deoxyuridine (**dU-Me**) and O^4-methyl-thymidine (**T-Me**) substrates, the presence of the *C*5-fluorine atom notably increased repair of the lesion, particularly by hAGT and the hOGT chimera.

2.2 Introduction

DNA damage can occur from different environmental factors that can arise from within and external to the cell.[76,77] For example alkylating agents such as *N*-methyl-*N*'-nitrosourea has been shown to introduce alkyl lesions at the O^6 and O^4 atoms of 2'-deoxyguanosine and thymidine respectively.[54,76] Such alkylations have been implicated in severely disrupting proper functioning of the cellular machinery, a feature that has been exploited in the treatment of cancer by designing drugs, which introduce alkyl lesions. Examples of chemotherapeutic drugs include Temozolamide and BCNU (1,3-bis (2-chloroethyl)-1-nitrosourea) which are known to target DNA as their main mode of action via the introduction of DNA damage.[78]

The detrimental effects of the naturally produced O^6-alkyl dG and O^4-alkyl T to hinder DNA polymerases and cause transitional events (G:C to A:T) have been well documented.[78–80] Their mutagenic influence arises mainly from the persistence of these lesions in addition to how often they occur.[43,81] O^6MedG incidences occur more frequently compared to O^4MeT in both *in vitro* and *in vivo* experiments where DNA is subjected to alkylating agents such as *N*-methyl-*N*'-nitrosourea and methyl methanesulfonate.[20,82,83] O^4MeT, is however said to have a higher level of cytoxicity due to its lower repair efficiency by the cellular repair machinery.[20,84] The O^4MeT lesion is detrimental as it readily forms a non-wobble base pair with the incorrect dG nucleotide, leading to a possible mutagenic outcome if left unrepaired.[80]

O^6-alkylguanine DNA alkyltransferase proteins (AGTs) are able to identify and remove these aforementioned alkylated lesions, via a direct reversal mechanism.[37] The alkylated nucleotide is flipped out of the duplex into the active site of the protein, where an activated cysteine residue (Cys145 in the human variant) reacts with the alkyl group. In doing so, the alkyl group is irreversibly transferred to the protein.[54,78] The protein product is then degraded rapidly via the ubiquitin pathway.[36] AGT proteins are found in most kingdoms of life, and crystal structures of different AGT variants depict a similar overall structure, despite high or low homology.[37] Certain AGT variants, however, display large substrate specificity, as well as adduct repair rates. For example, Ada-C and OGT from *E. coli* have been shown to repair small adducts at either the O^6-atom of dG or O^4-atom of T, with O^4-methyl T analogues being repaired more efficiently.[84–86] Human AGT (hAGT) has shown a preference for a range of adducts of various sizes at the O^6-position of dG.[84,85,87–89]

Among the substrates hAGT is reported capable of repairing are a wide array of mono-adducts (like ethyl, benzyl, just for mentioning a few) and O^6-alkyl(ene)-dG linked DNA intra- and interstrand cross-links. Moreover, O^6-benzylguanine (**Figure A1.1**) as the free base is also a very efficient pseudo-substrate for hAGT.[90–92] Studies have also revealed that while Ada-C and OGT are efficient at repairing O^4MeT DNA adducts, hAGT has a poor activity towards it.[78,84,85] Despite hAGT having the ability to recognize and bind to DNA containing O^4MeT, it does not efficiently remove it. The inability of repair of the mutagenic and cytotoxic O^4MeT lesion by

MMR and hAGT[13,78] and the fact that there are no available crystal structures of hAGT- O^4MeT DNA (Figure 2.1) warrants systematic and in-depth studies of this lesion by researchers in the field.

In previous work from our lab, the role of the *C*5 methyl group on O^4-alkyl T repair was studied.[54] It was believed that a steric clash could be occurring between the *C*5 methyl group (van der Waals radius of 2.0 Å) of the nucleobase and Arg135 of hAGT, preventing successful repair. Therefore, DNA containing O^4MedU (**dU-Me**, Figure 2.1) was investigated and repair analysis concluded that the steric contributions of the *C*5 methyl group may be responsible, at least in part for hAGT substrate selectivity since the dU substrates were processed more efficiently relative to their T counterparts.[54,93] An AGT chimera consisting in the hAGT protein with residues 139-159 replaced with those from the OGT active site was also analyzed. The chimera exhibited a twenty fold repair enhancement of **dU-Me** compared to hAGT.[54]

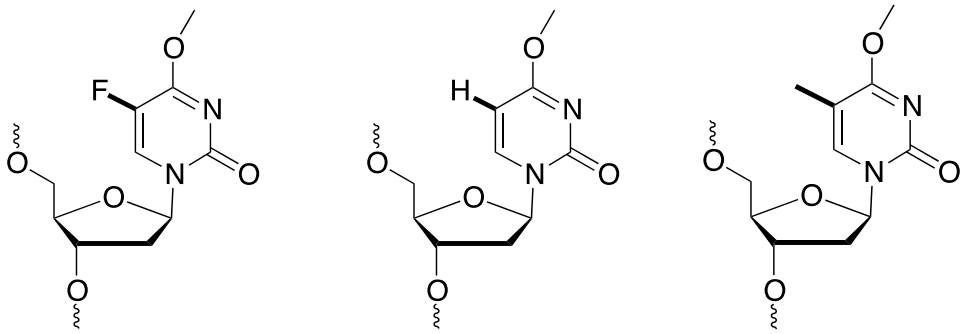
Given the effect of the *C*5 methyl group of O^4-alkyl T on AGT-mediated repair, we wished to further explore the electronic effects by replacing this group with the electronegative fluorine atom at the *C*5-position. The similar van der Waals radius of fluorine (1.47 Å) and hydrogen (1.20 Å) atoms enabled us to evaluate the electronic effects without introducing steric additional effects (in comparison to **dU-Me**).[59,94]

Another attractive aspect of studying the replacement of hydrogen with fluorine has been its role in drug discovery and presence in therapeutic agents. Fluorine has shown to increase the lipophilicity and fat solubility of various compounds.[59,94,95] 5-fluorouracil (**Figure A1.1**) and flunitrazepam are examples of drugs that demonstrate how the presence fluorine enhances therapeutic properties relative to the parent compounds.[27-29]

Fluorine modified oligonucleotides containing 2'-deoxy-2'-fluororibose (2'F-RNA) and 2'-deoxy-2'-fluoroarabinose (2'F-ANA) (**Figure A1.1**) have been shown to influence siRNA recognition by RISC, an important feature in strand selectivity and gene silencing traits.[72] In this chapter we have investigated the influence of fluorine at the C5 position on an O^4-

methylated pyridimine nucleobase examining the influence on duplex stability and structure in addition to susceptibility towards repair by AGT variants.

A



dFU-Me (*O*^{4}MedFU) **dU-Me (*O*^{4}MedU)** **T-Me (*O*^{4}MeT)**

B

$^{5'}$ GGC T**X**G ATC ACC AG **X** = *O*^{4}MedFU, dFU-Me

Figure 2.1: (**A**) Structures of the modified *O*^{4}-methylated pyrimidines, and (**B**) DNA sequence where X corresponds to the modified nucleoside.

2.3 Results and Discussion

2.3.1 Synthesis of nucleosides and oligonucleotides

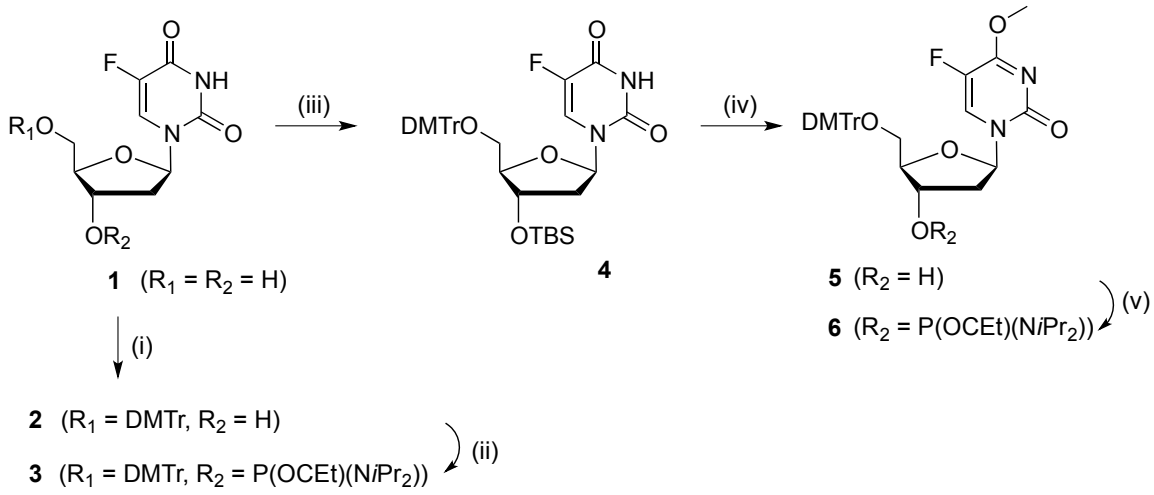
The structure of **dFU-Me** and the DNA sequence containing this modification are shown in **Figure 2.1**. Rather than a methyl group attached to the C5-position (**T-Me**), the **dFU-Me** substrate has a fluorine atom, which allows for exploring the influence of an electron withdrawing group at this position. Recently we found that the CH_3 group at C5 of *O*^{4}MeT affects AGT mediated repair.[54] Specifically, an *O*^{4}MeT containing oligonucleotide was observed to be repaired less efficiently relative to *O*^{4}MedU.[54] A logical extension of this study was to

examine the electronic effect exerted by substituents at this position *i.e* replacing a H atom by an electronic withdrawing group such as fluorine.

The oligomers containing the modified inserts were prepared by a combination of small molecules and automated solid-phase synthesis. The synthetic pathway is shown in **Scheme 2.1** and it begins with the commercially available 5-fluoro-2'deoxyuridine (**dFU**), in which the 5'-hydroxyl was protected with a 4,4'-dimethoxytrityl group, followed by silylation of the 3'-hydroxyl functionality using previously published procedures.[97,98] A convertible nucleoside was prepared by the addition of a 1,2,4-triazole group at the *C*4 position of the nucleoside. This reaction was optimized for the dFU analog since the procedure commonly used for the T and dU series resulted in a number of undesired by-products including detriylated nucleosides.[54,97] This may have been due to the presence of the fluorine, which could increase the reactivity at the *C*4 position. The formation of convertible intermediate was accomplished by the slow addition of 1,2,4-triazole, triethylamine and phosphoryl chloride in three portions at 30 min intervals with stirring at 0^{o}C to a solution containing compound **4**. After four hours, the reaction was complete and the convertible nucleoside was then transformed to compound **5** by treatment with sodium methoxide in methanol, followed by the removal of the 3'-O-*tert*-butyldimethylsilyl group by a fluoride treatment. Finally, compounds **2** and **5** were phosphitylated to generate phosphoramidites **3** and **6**, respectively, in good yields, using previously published procedures.[54] Purification was achieved by using short flash column chromatography and the purity of the compounds are demonstrated by their ^{31}P NMR spectra (**Figure A1.5** and **A1.12**). The ^{31}P NMR spectra revealed two sharp signals in the 147-149 ppm region characteristic of phosphoramidites.[54]

Solid-phase synthesis of the oligonucleotide was performed using "fast-deprotecting" commercially available 3'-*O*-phosphoramidites due to the labile nature of the **dFU-Me** adduct.[88] Phenoxyacetic anhydride was employed as the capping reagent to avoid transamidation.[99] Total deprotection and cleavage of the oligomers from solid support was accomplished with an anhydrous solution of potassium carbonate in methanol (0.05 M) for 4 h at 22^{o}C under gentle rocking.[88,97] Excess base was neutralized with an equimolar amount of

acetic acid and the solvent removed in a speed-vacuum concentrator. The oligonucleotide containing **dFU-Me** was purified by SAX-HPLC. Characterization by ESI-MS of DNA containing the **dFU-Me** adduct confirms the presence of the modification after the deprotection process and it was in agreement with the expected mass (**Figure A1.14**). Enzymatic digestion of modified oligonucleotides using snake venom phosphodiesterase and calf intestinal phosphatase enzymes also confirmed nucleoside composition consistent with expected ratios (**Figure A1.15**).



Scheme 2.1: Reagents and conditions: (i) DMTr-Cl, pyridine, DMAP, 16 h, 21 ^{o}C; (ii) *N,N*-diisopropylaminocyanoethylphosphonamidic chloride, DIPEA, THF, 30 min; (iii) DMTr-Cl, pyridine, DMAP, 16 h, 21 ^{o}C; TBS-Cl, Imidazole, DCM, 16 h, 21 ^{o}C; (iv) 1,2,4-Triazole, triethylamine, $POCl_3$, MeCN, 16 h, 0 ^{o}C; MeOH, NaOMe, 4 h, 21 ^{o}C; TBAF (1M in THF), 30 min; (v) *N,N*-diisopropylaminocyanoethylphosphonamidic chloride, DIPEA, THF, 30 min.

2.3.2 UV thermal denaturation and CD spectroscopy

In order to evaluate the effect of the fluorine at the *C*5 position and the methyl adduct at the O^4-atom in the stability of the DNA duplex, UV thermal denaturation studies were conducted. The results obtained are summarized in **Figure 2.2**. The introduction of the methyl group at the O^4 position caused a decrease in thermal stability by approximately 11 ^{o}C relative

to the unmodified controls. The same relative decrease in T_m upon methylation was observed whether the DNA contains the **T**, **dU** or the **dFU** O^4-Me lesions. These results suggest that the decrease in duplex stability after the incorporation of the methyl group was independent of the identity of the nucleobase (**T**, **dU** or **dFU**). The decrease in thermal stability may correspond to the disruption of the H-bonding between the O^4-methylated **dFU** and its base-paired 2'-deoxyadenosine, as well as less optimal base stacking with the adjacent bases. The results were in agreement with the data observed previously by our group.[54] Coincidentally, the **dU** and **dFU** substrates displayed the same T_m value of 58 °C, as well as their respective modifications (dFU-Me and dU-Me), implying that the presence of the fluorine at the *C*5 position had minimal influence on the DNA duplex stability.

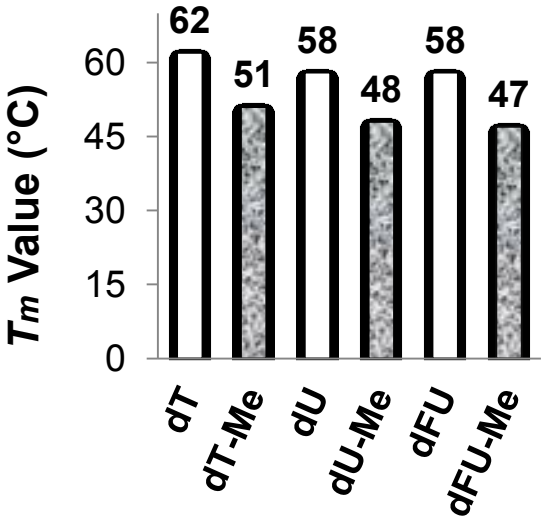


Figure 2.2 : T_m values (°C) of duplexes containing **T, T-Me, dU, dU-Me, dFU, dFU-Me**. Colorless and grey bars represent the unmodified controls and methylated adducts, respectively. Hyperchromicity change (A_{260}) *versus* temperature (°C) profiles are shown in **Figure A1.17.**

The CD analysis of the profiles from **dFU**, **dFU-Me**, and the control **T** display spectroscopic signatures consistent with the B-form DNA family, exhibiting a maximal positive signal at 280 nm, a cross-over near 256 nm, and a maximal negative signal at 245 nm. The CD signatures of the duplex containing **dFU-Me** was similar to the unmodified control DNA duplex profile, which suggested that the O^4-methyl group did not cause any major effect on the overall

DNA duplex structure, as previously reported for the dU series.[54] Basic molecular modeling was performed on duplexes containing **dFU**, **dFU-Me** and the unmodified control using the AMBER force field (**Figure A1.20)**. The structures of the duplexes generated from molecular modeling supported the minor deviations of the signatures observed for the CD experiments, displaying minimal global distortions in the DNA duplexes upon introduction of the *C*5-fluoro or both the *C*5-fluoro and O^4-Me groups.

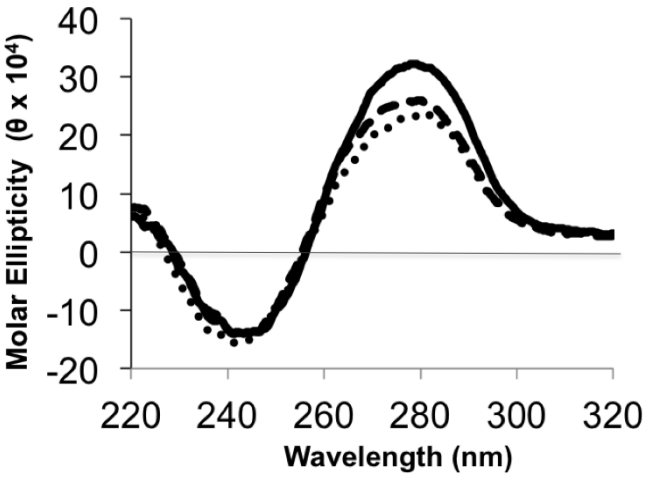


Figure 2.3: Circular dichroism spectra of DNA duplexes containing **dFU•dA** (———), **dFU-Me•dA** (—) and unmodified control **T•dA** DNA (•••).

2.3.3 Repair Assays

The repair of **dFU-Me** substrate was studied with three recombinant AGT proteins (hAGT, Ada-C and OGT), and a chimera protein (hOGT). The repair assays were performed as described previously by our group, where we monitored production of a product generated from *BclI* cleavage with DNA duplexes designed to contain a cut site.[54] The **dFU-Me** modification is located in a position close to this cut site and if repair of the modified sequence to the native (repaired) sequence occurs, *BclI* cuts the oligomer into two smaller DNA fragments (5 and 9 mers) that are separated by PAGE. The strand containing the modification is radiolabeled and a shorter 5mer can be visualized by autoradiography.[54] If the O^4-methyl group is retained, no repair has occurred and the intact 14mer is observed.

First, a total repair assay was conducted with the **dFU-Me** substrate using the various AGTs, allowing the reactions to proceed overnight at 37 ^{o}C. Denaturing PAGE was used to resolve the different species after the repair reaction. Repair assay results revealed that **dFU-Me** was efficiently repaired by all four AGTs evaluated at five-molar equivalences (reaching approximately 80 % repair). These results mirrored trends observed for the total repair assays of O^4-methyl-dU but not the O^4-methyl-T series.[54,88]

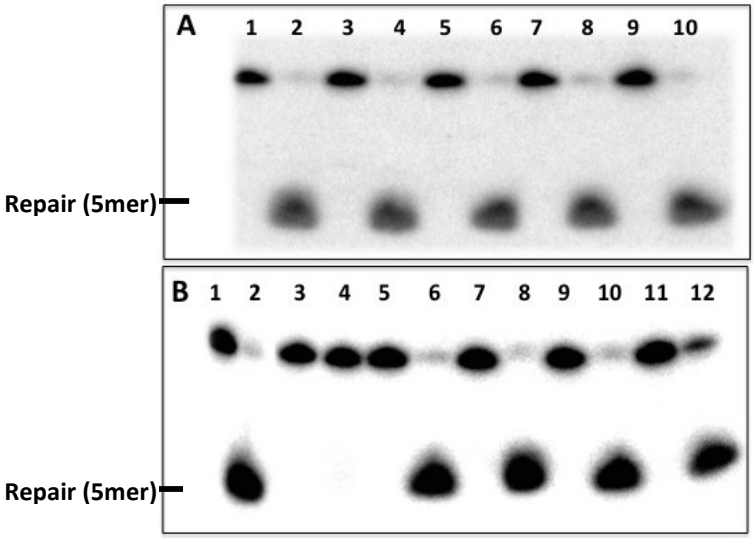


Figure 2.4: Total Repair of (**A**) **dFU** (2 pmol) and (**B**) **dFU-Me** (2 pmol) by hAGT (10 pmol), OGT (10 pmol), Ada-C (10 pmol), and hOGT (10 pmol), for 2.5 h at 37^{o}C. Panel **A**; Lane 1, **dFU** DNA; lane 2, **dFU** + *BclI*; lane 3, **dFU** + hAGT; lane 4, **dFU** + hAGT + *BclI*; lane 5, **dFU** + OGT; lane 6, **dFU** + OGT+ *BclI*; lane 7, **dFU** + Ada-C; lane 8, **dFU** + Ada-C + *BclI*; lane 9, **dFU** + hOGT; lane 10, **dFU** + hOGT + *BclI*. Panel **B**; Lane 1, **dFU-Me** DNA; lane 2, **dFU-Me** + *BclI*; Lane 3, **dFU-Me**; lane 4, **dFU-Me** + hAGT; lane 5, **dFU-Me** + hAGT; lane 6, **dFU-Me** + hAGT + *BclI*; lane 7, **dFU-Me** + OGT; lane 8, **dFU-Me** + OGT+ *BclI*; lane 9, **dFU-Me** + Ada-C; lane 10, **dFU-Me** + Ada-C + *BclI*; lane 11, **dFU-Me** + 10 hOGT; lane 12, **dFU-Me** + hOGT + *BclI*.

The time course assays revealed that **dFU-Me** was repaired faster compared to the **dU-Me** and **T-Me** analogues. Time course repair studies were performed with the four AGTs variants conducted at room temperature (Figure 2.5 and A1.19). Both *E. coli* variants (OGT and Ada-C) required less than 15 s to achieve full repair of **dFU-Me**. These results were in

agreement with the previous studies with **dU-Me** and **T-Me** where total repair by OGT and Ada-Cl occurred in less than 15 s.[54]

Remarkably, repair of **dFU-Me** by hAGT only required approximately 90 s for the reaction to reach completion. It was surprising to observe complete repair of **dFU-Me** by hAGT to occur more quickly relative to the **dU-Me** and **T-Me** analogues. It was previously shown that **dU-Me** undergoes 25% repair by hAGT in 2.5 min, whereas similar repair levels were obtained for **T-Me** after 30 min. **dU-Me** hAGT-mediated repair was approximately 12-fold faster compared to **T-Me** in terms of time to achieve 25% substrate repair.[54] **dFU-Me** total repair by hAGT required 20 times less time relative to the **dU-Me** total repair profile, and more than a 100 times less time to repair compared to **T-Me**.

Complete repair of **dFU-Me** by the chimera hOGT occurred in less than 1.5 min, displaying a similar repair profile to hAGT. Repair by the hOGT chimera of **dFU-Me** was similar compared to **dU-Me**. When compared to results from previous studies, **dFU-Me** appears to require 3 times less time to achieve 80% repair than the dU-Me adduct.[54]

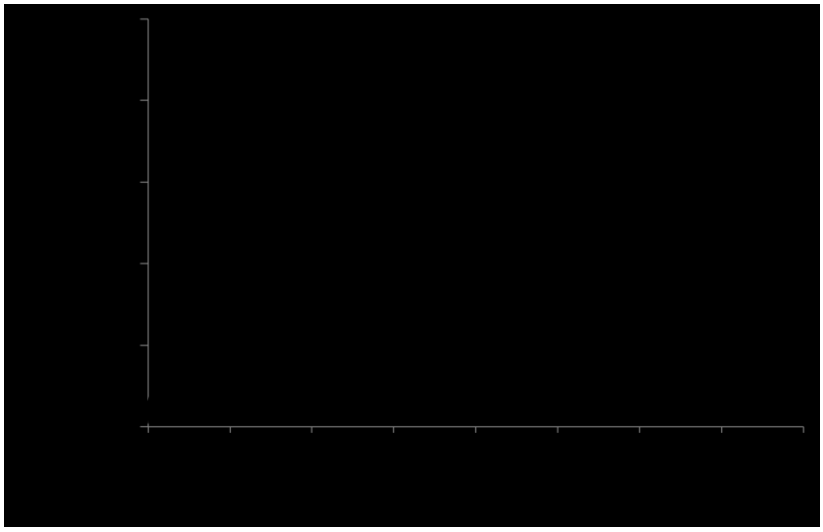


Figure 2.5: Time course repair assay at room temperature of **dFU-Me** by OGT (—), Ada-C (—•—), hAGT (—••—), and hOGT (———) at 5-molar equivalence protein. Graphical illustration displays **dFU-Me** repair [%] over time (min).

Our study shows that the presence of the fluorine at the *C*5-position not only enhances repair by hAGT but also reduces the difference in repair efficiency between hAGT and OGT. It appears that the presence of the fluorine at the *C*5 position weakens the ether O^4-C_α bond making it more susceptible to nucleophilic attack by Cys145. Towards this end, computational studies using Gaussian 09 and B3LYP/6-31+G(2d,2p) mode were performed with the assistance of Dr. Derek O'Flaherty and Philippe Archambault in order to gain a better understanding of the effect of fluorine on the electron density of the O^4- C_α bond. The calculations provided the O^4-C_α bond length from the geometrically optimized models of the corresponding *N*1-methylated modified pyrimidines (see **Figure A1.21)**. The results revealed that the O^4-C_α bond in **dFU** may be weaker compared to **dU** and **T** as it is slightly longer. This may explain why this lesion is susceptible to repair by AGTs. Alternatively, the protein's active site could be participating in favourable hydrophobic interactions in the presence of the fluorine atom and computational studies are currently underway. The presence of fluorine might also have a substantial effect on weak dipolar interactions and basicity of the active site.[59,61] Crystallographic structures of the protein complexed with damaged DNA will aid in elucidating the critical interactions, which promote hAGT-mediated repair. Moreover, high-resolution structures of duplexes containing **dFU-Me** are currently being investigated via a combination of molecular dynamics and high-field NMR experiments to gain insights on the impact of this modification on DNA structure.

The importance and role of fluorine in biological applications and for drug regimens has been well documented.[59,63] The enhanced repair of **dFU-Me** containing oligomers by hAGT reported here may be due to some allosteric interactions occurring with the protein. Moreover, the increase in hAGT levels has been correlated to an increase in cellular resistance to certain chemotherapeutic agents. Future studies are currently underway with other **dFU** substrates for AGT-mediated repair, which may find potential uses as AGT inhibitors in certain combinatory regimens.[90,91]

2.4 Conclusion

Our study describes the synthesis of the **dFU-Me** phosphoramidite and its incorporation into oligonucleotides by solid-phase synthesis in scales and purity sufficient for biochemical and biophysical studies. A decrease of 11 ^{o}C in the thermal stability was observed for duplexes containing the modification relative to the control duplex. The CD data revealed no major change on the global structure of DNA containing the methylated dFU insert relative to the control duplex. Repair of **dFU-Me** by the different AGTs occurred in less than 2 min, with the *E. coli* variants displaying greatest proficienty (with repair occurring in less than 15 sec), which was observed previously with substrates containing hydrogen rather than a methyl group at the C5-position.[54,78] hAGT shows a remarkable increase in its repair efficiency towards **dFU-Me** compared to the **dU** and **T** analogues. The presence of fluorine at the *C*5 position increases susceptibility towards hAGT repair. We suspect that the electronic effect of the fluorine on the O^4-methyl lesion of the pyrimidine enhances susceptiblity towards thiolate nucleophilic attack by AGT. The **dFU-Me** substrate may find applications for anticancer therapies.[91] Given the fast repair of **dFU-Me**, studies using this substrate as a free nucleoside could help in guiding the design of hAGT inhibitors. Further investigation of different **dFU** substrates to undergo repair by hAGT is underway.

Chapter 3. Influence of fluorine at the C5 position of various O^4-alkylated 2'-deoxyuridines on repair by O^6-alkylguanine-DNA alkyltranferases

3.1 Abstract

Oligonucleotides containing various alkyl groups at the O^4-atom of 5-fluoro-O^4-alkyl-2'deoxyuridine (dFU) were prepared by solid-phase synthesis. UV thermal denaturation studies of duplexes containing O^4-ethyl-dFU, O^4-benzyl-dFU, O^4-(4-hydroxybutyl)-dFU, and O^4-(7-hydroxyheptyl)-dFU were found to decrease duplex stability by 10 ^{o}C compared to the control duplex containing unmodified dFU. Circular dichroism profiles of the modified duplexes all resembled a canonical B-form DNA. Repair of smaller adducts was observed for human and *E coli.* O^6-alkylguanine DNA alkyltransferases (Ada-C and OGT). Surprisingly, human AGT and the *E. coli* variant Ada-C displayed proficiency in the repair of larger O^4-adducts. Repair of duplexes containing an 7-hydroxyheptyl linker at the O^4-position of dFU resulted in more efficient repair compared to the 4-hydroxybutyl lesion. In contrast to most trends observed, Ada-C was capable of repairing the 4-hydroxybutyl adducts but was unable to repair the 7-hydroxyheptyl linker. These results reveal the positive influence of the C5-fluorine atom on repair of larger O^4-alkyl adducts expanding our knowledge of the range of substrates that are able to be repaired by AGT.

3.2 Introduction

DNA is continuously exposed to different modifying agents either within the cell or coming from the environment. Examples of these agents include various metabolites such as nitroso-compounds and the methylating agent S-adenosylmethionine. In addition, certain chemotherapeutic regimens employ methylating or chloroethylating agents.[76,77,79,100] O^6-Alkylguanine-DNA-alkyltransferases contribute to the maintenance of genomic integrity by repairing alkyl lesions occurring at the O^6- and O^4- atoms of 2'-deoxyguanosine and thymidine, respectively. The alkyl group is transferred from the DNA to the protein via a non-reversible

nucleophilic attack by a cysteine thiolate (Cys145 in human AGT (hAGT)) in the active site. These alkylations can be detrimental if they are not repaired as they can cause transitions that could lead to mutations and carcinogenesis.[43,80,100]

Alkylating agents can vary in the size and nature of the lesions introduced in DNA. These lesions can range from methyl and ethyl groups, to much larger adducts such as crosslinks between residues within DNA (intrastrand or interstrand) or to other biomolecules such as proteins.[32,87,88,100,101] The capability of repair by AGTs is related to the accommodation of lesions within the protein active site. [52,100] Removal of alkyl lesions occurs via transfer to an activated cysteine residue located in the active site of AGT. Crystal structures of human AGT and the *E. coli* AGT variant Ada-C have provided insights about the roles that various active site residues play in mediating critical interactions for successful repair. For example, the poor repair of O^4-alkyl lesions of T by hAGT has been proposed to be linked to less than optimal positioning of the target lesion with respect to the activated cysteine responsible for damage removal.[33,40,46,102]

The identity of residues and the shape of the active site varies between AGT homologues. These differences determine in great part the substrate specificity. For instance hAGT's active site has a tunnel-like shape (8 Å wide, 9 Å deep and 14 Å long) that allows for accommodation of larger adducts like benzyl and cross-links.[33,35,53,89] The Pro140 residue of hAGT stacks with the benzyl group generating favorable hydrophobic interactions. The active site of Ada-C is said to be smaller, containing residue Trp161 that impedes larger adducts from being accomodated into the active site. The lack of two prolines which are found on hAGT (Lys 139 and Ala141 instead of Pro138 and Pro140 respectively) may also contribute to the narrowing of the active site.[33,35,40,102]

Understanding the basis for specificity and discovering new substrates that could be repaired more efficiently by AGTs has been investigated by our group and others.[46,54,88,100] McManus *et al.* reported that the C5 methyl group of O^4-alkyl-T lesions may contribute to the reduced repair observed by hAGT, possibly via steric interactions.[54] Substituting the methyl group for hydrogen was shown to drastically improve the efficiency of hAGT in removing alkyl

groups at the O^4-position. In addition, it has been demonstrated that replacing the C5-methyl group of thymidine with fluorine further enhanced susceptibility towards repair of an O^4-methyl adduct by hAGT. It is believed that the presence of the fluorine atom at this position may weaken the O^4-C_α bond enhancing susceptibility to undergo transfer to the activated thiolate in AGT. In addition, the presence of fluorine may contribute to favourable interactions in the AGT active site pocket, with reduced repulsion or steric effects with amino acid side chains.

Fluorine substitutions have widely been investigated in studies involving proteins and nucleic acids to understand biological properties and mechanisms. Incorporation of fluorine in proteins can be used to evaluate the effects of altering the dipole moment of a molecule or altering the acidity or basicity of groups that are in close proximity.[59,63] There are numerous examples of fluorinated nucleoside analogues, which have demonstrated beneficial properties in the treatment of various diseases including cancer. For example, fluoropyrimidine-based prodrugs whose mechanism primarily inhibits thymidylate synthase have been used for cancer treatment.[67] 5-Fluorouracil, 5-fluoro-2'-deoxyuridine and Capecitabine have also been studied as therapeutic agents. [67,69,70] In nucleic acid chemistry, the introduction of fluorine at the sugar has produced molecules with interesting properties. One example is 2'-deoxy-2'-fluoro-β-D-arabinonucleic acid (2'-F-ANA) which exhibits significant resistance towards degradation by various nucleases.[71] This modified nucleic acid has been explored as a gene silencing agent for RNA interference.[72,73]

In the present study, we have investigated the effect of replacing the C5-methyl group of thymidine with fluorine on alkyl lesions larger than a methyl group. Fluorine has a dual electronic influence due to inductive and resonance effects. In fluorobenzene, fluorine is said to have an electron-withdrawing inductive effect and is also electron-donating by resonance on the aromatic system.[56,64]

The positive influence of fluorine observed on repair by various AGTs towards the O^4-methyl lesion has evoked our interest in investigating bulkier/larger adducts. In the present study ethyl, benzyl, 4-hydroxybutyl and 7-hydroxyheptyl adducts at the O^4-position of 5-fluoro-2'-deoxyuridine were synthesized, incorporated into 14-mer oligonucleotides and their

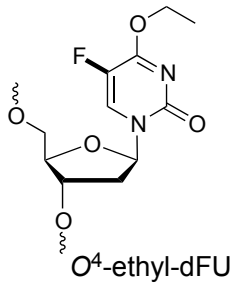
influence on duplex stability and structure evaluated by UV thermal denaturation and circular dichroism spectroscopy. Susceptibility towards repair of these bulkier adducts by hAGT, in comparison to their T (C5-methyl) or dU (C5-hydrogen) counterparts was also investigated.

3.3 Results and Discussion

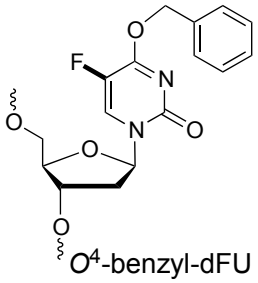
3.3.1 Synthesis of nucleosides and oligonucleotides

There have been numerous reports describing how fluorine modifications lead to enhancement of the properties of proteins and nucleic acids.[59,63] Previous studies from our group have demonstrated the positive influence of replacing the C5 hydrogen atom of 2'-deoxyuridine with fluorine on the repair of O^4-methyl adducts by various AGT homologues. Thus, the objective of this study was to examine whether the influence of the C5-fluoro group extends to AGT-mediated repair of larger adducts at the O^4-atom. The adducts evaluated include O^4-ethyl-dFU (**dFU-Et**), O^4-benzyl-dFU (**dFU-Bn**), O^4-(4-hydroxybutyl)-dFU (**dFU-C4OH**), and O^4-(7-hydroxyheptyl)-dFU (**dFU-C7OH**). Their chemical structures and the DNA sequences highlighting the position of the modification are shown in **Figure 3.1**.

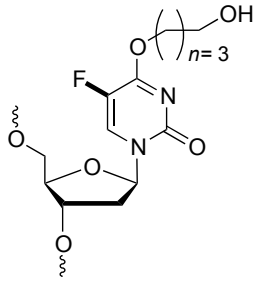
A



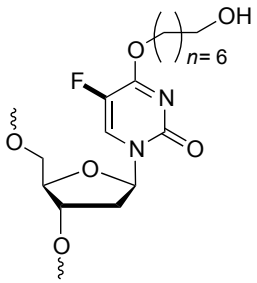
O^4-ethyl-dFU
(**dFU-Et**)



O^4-benzyl-dFU
(**dFU-Bn**)



O^4-(4-hydroxybutyl)-dFU
(**dFU-C4OH**)



O^4-(7-hydroxyheptyl)-dFU
(**dFU-C7OH**)

B 5' GGC T**X**G ATC ACC AG

X = dFU-Et

dFU-Bn

dFU-C4OH

dFU-C7OH

Figure 3.1: (**A**) Structures of the modified 5-fluoro-2'-deoxyuridine adducts at the O^4 position. (**B**) DNA sequence where X corresponds to the modified nucleoside.

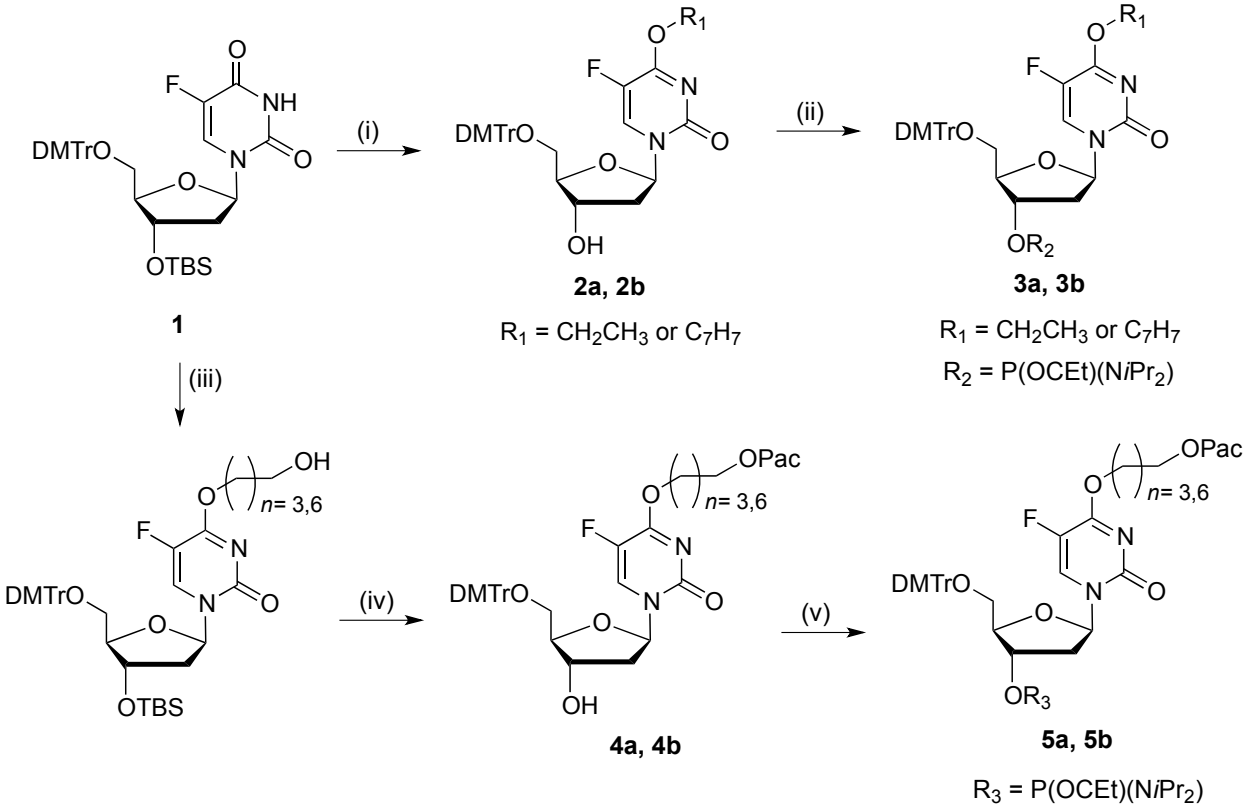
dFU-Et was chosen as it is slightly larger than a methyl group and is a lesion which can occur in genomic DNA as a result of exposure to ethylating agents.[26,90,103] **dFU-Bn** was inspired from the well-characterized inhibitor, O^6-benzyl-guanine, which has been said to be one of the preferred substrates of hAGT.[35,44,100] Although, preference towards the benzyl adduct is altered at the O^4-atom of T relative to the O^6-atom of dG, recent studies showed that hAGT was capable of efficiently repairing O^4-benzyl-dU.[54] O^4-(4-hydroxybutyl)-dFU and O^4-(7-hydroxyheptyl)-dFU corresponded to the potential hydrolysis adducts of busulfan (sulfonate) and hepsulfam (sulfamate) alkylating agents, which can generate a variety of DNA damage including mono-adducts and DNA cross-links.[88,89,100] These substrates also allowed for comparison with the results of analogues studied previously by our group modified at the O^4 atom of T and dU in the same DNA sequence context. [45,54,88]

The synthetic pathway for the target phosphoramidites is shown in **Scheme 3.1.** The synthesis begins with the protection of the 5'- and 3'-hydroxyl groups of commercially available 5-fluoro-2'-deoxyuridine (**dFU**) according to previously published procedures.[97,98] The formation of a convertible intermediate was accomplished by the slow addition of triethylamine followed by phosphoryl chloride to a solution containing 1,2,4-triazole and 3'-*O*-(*tert*-butyldimethylsilyl)-5'-*O*-(4,4'-dimethoxytrityl)- 5-fluoro-2'-deoxyuridine stirring at 0^{o}C.[54] In order to obtain compounds **2a** or **2b**, the convertible nucleoside was treated with sodium ethoxide in ethanol or benzyl alcohol and DBU, respectively, in MeCN. Fluoride treatment was used to remove the 3'-*O*-*tert*-butyldimethylsilyl protecting group. Finally compounds **2a** and **2b** were phosphitylated to yield phosphoramidites **3a** and **3b** using previously published procedures from our group.[54,88,89,104] For the adducts containing the terminal hydroxyl functionalities, the convertible nucleoside was treated with 1,4-butanediol or 1,7-heptanediol in MeCN. Protection of the terminal alcohol on the adducts was performed with phenoxyacetyl chloride (also employed as the capping agent during solid phase synthesis).[54,88,99] Removal of the 3'-*O*-*tert*-butyldimethylsilyl group was achieved by fluoride treatment to yield compounds **4a** and **4b**. Finally, phosphitylation of these compounds was performed to furnish compounds

5a and **5b**, respectively. Purification of all phosphoramidites was achieved by using short flash column chromatography and the purity and identity are confirmed by their ^{31}P NMR spectra (**Figures A2.8**, **A2.15**, **A2.22** and **A2.29**), all of which have two signals in the region of 148-150 ppm, characteristic of phosphoramidites.

Due to the labile nature of the O^4-alkyl modifications and our previous observation that this is enhaced by the C5-fluoro group, "fast-deprotecting" commercially available 3'-*O*-phosphoramidites were employed for solid-phase synthesis.[54,88] Deprotection after solid phase synthesis was achieved by treating the oligomers bound to the solid support with a solution of 10% (by volume) DBU in the corresponding alcohol employed during the synthesis of each respective modified phosphoramidite (*e.g.* ethanol for **dFU-Et** and benzyl alcohol for **dFU-Bn**). Using the respective alcohols ensured the retention of the desired O^4-modification given the labile nature of alkyl lesions at this position.[88,105] 1,7- Heptanediol was particularly difficult to remove, which necessitated extractions with DCM in order to eliminate the remnant diol and to avoid interference with the desalting process, as observed previously.[54]

The oligonucleotides **dFU-Et**, **dFU-Bn**, **dFU-C4OH**, and **dFU-C7OH** were purified by SAX-HPLC. Characterization by ESI-MS of DNA containing each of the above mentioned modifications confirmed the retention of the desired modifications after the deprotection step (**Figures A2.30 - A2.33**). Moreover, enzymatic digestion of the modified oligonucleotides using snake venom phosphodiesterase and calf intestinal phosphatase enzymes was performed and the resulting nucleoside mixtures were analyzed by RP-HPLC (amounts, retention times and nucleoside ratios are shown in **Figures A2.34 - A2.37**). The presence of the additional modified and four standard nucleosides, in the correct ratio, also serves to confirm the identity of the sequences and their composition.



Scheme 3.1: Reagents and conditions: i) 1,2,4-Triazole, triethylamine, $POCl_3$, MeCN, 16 h, 0 ^oC; EtOH (**a**-series compounds) or BnOH (**b**-series compounds), DBU, MeCN, 16 h; TBAF (1M in THF), 30 min; ii) N,N-diisopropylaminocyanoethylphosphonamidic chloride, DIPEA, THF, 30 min; iii) 1,2,4-Triazole, triethylamine, $POCl_3$, MeCN, 16 h, 0 ^oC; 1,4-butanediol (**a**-series compounds) or heptane-1,7-diol (**b**-series compounds), DBU, MeCN, 16 h; iv) phenoxyacetyl chloride, triethylamine, THF, 30 min; TBAF (1M in THF), 30 min; v) N,N-diisopropylaminocyanoethylphosphonamidic chloride, DIPEA, THF, 30 min.

3.3.2 UV thermal denaturation studies

The influence of the various alkyl adducts at the O^4 atom of dFU on duplex stability was evaluated. As observed previously, the **dFU** control DNA duplex had a T_m of 58 °C. The presence of a methyl group at O^4-position of dFU (**dFU-Me**) was shown to decrease the thermal stability of the duplex by approximately 10^oC. A similar decrease in thermal stability was observed for the other alkyl adducts at the O^4-atom including ethyl (**dFU-Et**), benzyl (**dFU-Bn**), 4-hydroxybutyl (**dFU-C4OH**) and 7-hydroxyheptyl (**dFU-C7OH**) (as shown in Figure 3.2). Such a

decrease in stability can be attributed to the altered hydrogen bond interactions between the bases (Figure 1.3 A and B) and stacking interactions, similar to O^4MeT.[80] The results demonstrate that the decrease in thermal stability is similar regardless of the adduct size for the modifications studied. Moreover, having fluorine versus a methyl or hydrogen atom at the C5 position has minimal effect on the thermal stability of the DNA duplexes containing adducts at the O^4-atom.

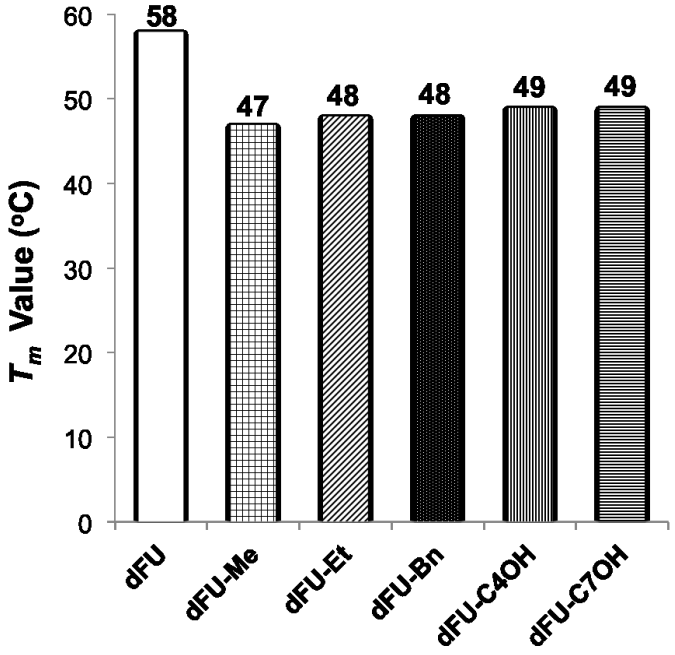


Figure 3.2 : T_m values (°C) of duplexes containing the control **dFU** (colorless)**, dFU-Me** (grey squares), **dFU-Et** (grey diagonal lines), **dFU-Bn** (black)**, dFU-C4OH** (grey vertical lines), and **dFU-C7OH** (grey horizontal lines).

3.3.3 Circular dichroism

Circular dichroism studies were performed in order to evaluate the global structure of the DNA duplexes containing the O^4-modifications. The CD profiles for duplexes containing **dFU-Et**, **dFU-Bn**, **dFU-C4OH** and **dFU-C7OH** (shown in Figure 3.3) exhibit spectroscopic signatures characteristic of B-form DNA, with positive and negative signals at around 280 and 245nm, respectively. The CD profiles are similar to the control duplex containing **dFU**, suggesting that

the introduction of these modifications at the O^4-atom, in spite of their size, has a minimal effect on the global structure of the DNA duplex.

Figure 3.3: Circular dichroism spectra of duplexes containing **dFU** control (—), **dFU-Et** (•••), **dFU-Bn** (———), **dFU-C4OH** (—•—) and **dFU-C7OH** (—••—).

3.3.4 Repair Assays

Repair of these adducts was evaluated with 4 different AGT homologues (in a 5:1 protein to DNA ratio); human (hAGT), *E. coli* (Ada-C and OGT), and a chimera protein "hOGT" (hAGT where residues 139–159 are replaced with residues 133-153 from OGT but maintaining Pro140)[45]. We were interested in evaluating the repair of these adducts at the O^4-atom, as substrate preference and repair efficiency varies between AGTs. Of particular interest is the influence of the fluoro group at the C5 position, compared to the C5-methyl of T or C5-hydrogen of dU, on repair efficiency with adducts larger than an O^4-methyl group.

hAGT is known to have a preference for repair of larger adducts such as the benzyl group, especially at the O^6-atom of dG.[23,30,106] Of the 4 novel substrates we evaluated, **dFU-Bn** was observed to be repaired the fastest by hAGT, with complete repair achieved in 30 seconds (Figure 3.4 A). Surprisingly, hAGT was incapable of repairing the same adduct in thymidine,

which differs only in the presence of a methyl group instead of the aforementioned fluoro group at the C5-position of the nucleobase.[107] Complete repair of **dFU-Et** required approximately 45 min, slower than **dFU-Bn**, suggesting that even though the ethyl adduct is smaller than Bn the latter adapts better to the active site of hAGT. This may be due to a stacking interaction on the benzyl group with Pro140 which may facilitate an optimal orientation for binding and thiolate attack.[31,33,35,100] For the alkylene adducts containing terminal hydroxyl groups, it was observed (Figure A2.49) that **dFU-C7OH** (150 min to reach about 70% repair) was more efficiently repaired over **dFU-C4OH** (40% repair after overnight incubation). Perhaps the 7 carbon linker may have greater flexability adopting a conformation that evades steric clashes in the active site of hAGT. Similar results were found in repair of intrastrand and interstrand cross-links, where the 7 carbon linker is repaired preferentially over the 4 carbon analog.[7,11,12,89] Studies of crystal structures of hAGT have revealed that its binding channel has a tunnel-like shape that allows the entrance and accommodation of larger adducts with the reactive Cys145 located at the end of the tunnel (9 Å deep).[33,35,49] Surprisingly, hAGT displays no repair of the 4-hydroxybutylene and 7-hydroxyheptylene lesions at the O^4-atom of T (C5-methyl) or dU (C5-hydrogen),[54] demonstrating that the C5-fluoro modification of dFU contributes to increased susceptibility of damage removal at this position.

OGT is known to repair a range of alkyl lesions at the O^4-position of thymidine.[43,52,54,100] The **dFU-Et** lesion was repaired by OGT in under 15 sec and the **dFU-Bn** adduct was fully repaired in 15 sec (Figure 3.4 B). Previously, it was observed that OGT repair of **dU-Bn** occurred in 15 sec.[54] Larger groups at the O^4-position of T such as 4-hydroxybutylene and 7-hydroxyheptylene are more of a challenge for repair by OGT. Surprisingly after 10 min, **dFU-C7OH** was almost fully repaired while less than 50% of **dFU-C4OH** was repaired (Figure 3.7). For the dU series, OGT repair of the 7-hydroxyheptylene and 4-hydroxybutylene adducts occurred efficiently within 10 and 2 min respectively, with similar repair profiles observed within the first minute of the reaction.[54] The reduced efficiency observed for repair of **dFU-C4OH** by OGT is the first example we have encountered of an O^4-alkylated lesion where the C5-fluoro analog (of 5-fluoro-2'-deoxyuridine) is less efficiently repaired relative to the C5-hydrogen (of 2'-deoxyuridine) analog. This highlights a balance between the positive influence of the electronic

effects of fluorine at the C5-position on AGT repair of O^4-alkyl groups with factors such as adduct size and flexibility.

The other *E. coli* variant, Ada-C, has a limited range of substrates it can repair. For instance it is proficient at repairing smaller adducts such as methyl and ethyl groups at the O^6- and O^4-atoms of dG and T, respectively.[84,108–110] A preference in repair is observed when the adduct is a methyl rather than an ethyl group at the O^6 position of dG.[108] Ada-C was able to reach full repair of **dFU-Et** in less than 15 sec (Figure 3.4 C) with repair of the O^4-ethyl group observed to be efficient as the O^4-methyl group with a fluorine atom at the C5 position.

It has been proposed that larger adducts, such as a benzyl group, might not be repaired by Ada-C due to steric obstruction in the protein's active site.[107,108] Unexpectedly, Ada-C demonstrated activity towards **dFU-Bn** with about 40% repair observed in the first 15 sec and 60% repair after 150 min. Remarkably, Ada-C also repaired 40% of **dFU-C4OH** after 150 min (Figure A2.49), whereas the 2'-deoxyuridine analogs **dU-Bn** and **dU-4COH** were resistant to repair.[54] No repair was observed on **dFU-C7OH** by Ada-C, which may be due to the active site not being capable of accommodating the longer linker. Recently, our group showed moderate repair of an O^4-thymidine-heptylene-O^6-2'-deoxyguanosine intrastrand cross-link adduct lacking the intradimer phosphodiester group by Ada-C. The repair occurred solely at the O^6-position of the 2'-deoxyguanosine moiety and marked one of the first examples where Ada-C demonstrated proficiency in the repair of larger alkyl adducts. Interestingly, the analogous butylene cross-link was inert towards Ada-C repair. An inverted trend is observed for the pyrimidyl series described herein for Ada-C.

Finally, the chimera hOGT behaved similar to OGT, being more effective at fully repairing the smaller **dFU-Et** adduct within 30 sec (Figure 3.4 D). Full repair of **dFU-Bn** was achieved in about 1 min, which is similar to the results obtained for OGT and hAGT. The combination of the optimal accommodation in the active site in addition to the favourable effect of fluorine at the C5 position likely contributes to **dFU-C4OH** and **dFU-C7OH** repair to be higher with hOGT, relative to OGT and hAGT, achieving approximately 70% repair after 45 min and 1 min, respectively (Figure 3.7). Again, the 5-fluoro-2'-deoxyuridine series exhibited

enhanced susceptibility towards repair of the adducts at the O^4-position relative to the 2'-deoxyuridine series.

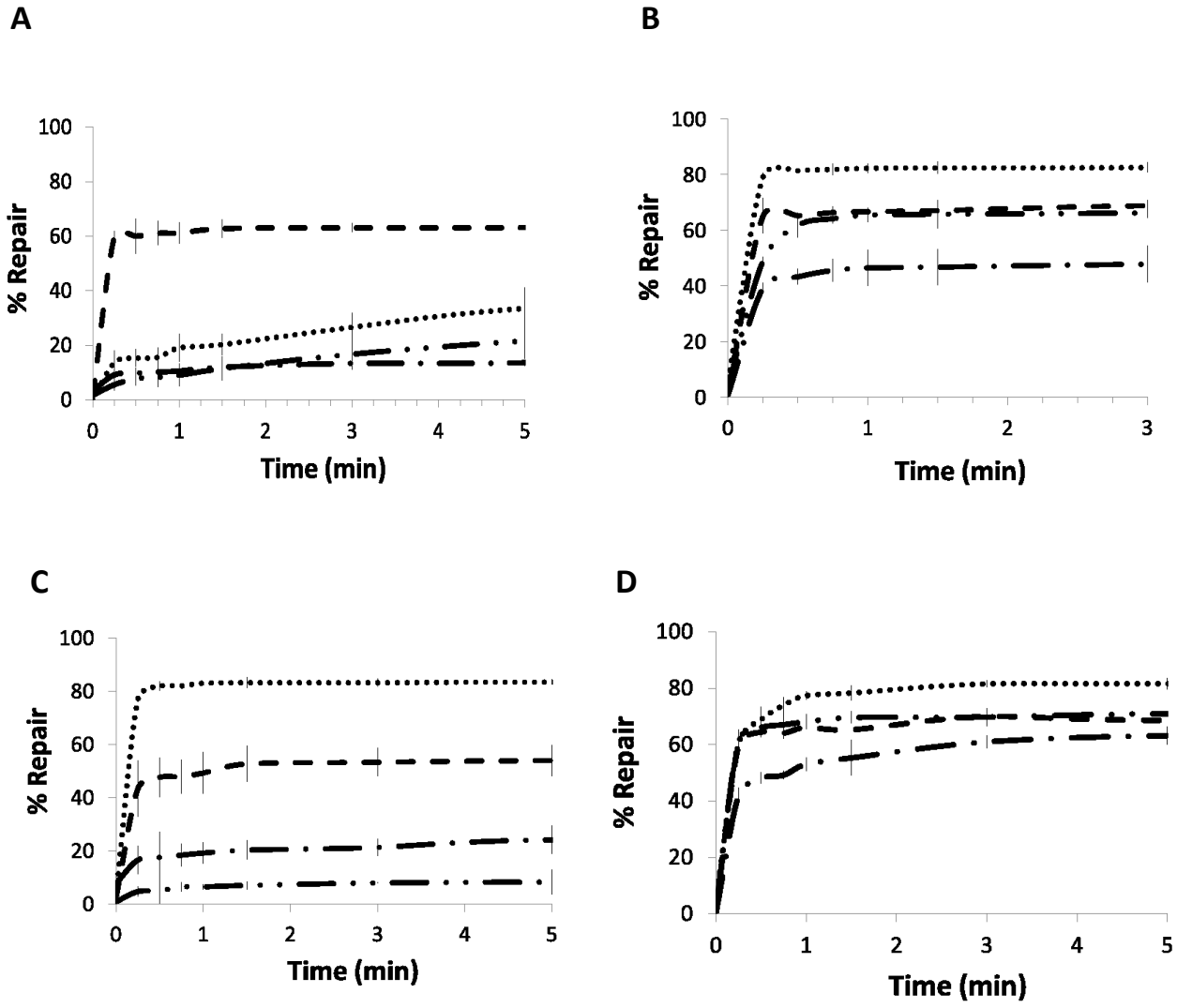


Figure 3.4: Time course of repair of **dFU-Et** (•••), **dFU-Bn** (———), **dFU-C4OH** (—•—) and **dFU-C7OH** (—••—) at room temperature: **A)** by hAGT, **B)** by OGT, **C)** by AdaC and **D)** by the chimera hOGT.

One of the hypotheses as to why the C5-fluorine group has a positive influence on AGT repair is that it may exert an effect on the O^4-C_α bond of alkyl group attached to the nucleobase rendering it more labile for facile transfer to the thiolate anion in the active site. There could be an electronic effect occurring within the nucleobase bonds causing the O^4-C_α weakening. This

has been observed in C5-fluoro cytosine, in where the C5-fluoro group through the O^2 and N3 atoms disrupts the CG basepair.[75,111,112] The presence of fluorine in a biomolecule can impart interesting properties and allow for probing of the role various residues play in maintaining structure and catalyzing reactions.[59,63] There has been much debate on the capability of fluorine to participate in hydrogen bonding.[59,74,113,114] Interaction with fluorine depends on the distance and position relative to the atoms involved.[61,115] The literature contains numerous examples where fluorine influences the basicity or acidity of certain chemical groups.[63] It can also generate specific polar and hydrophobic interactions with some protein residues (Figure 3.5). In the case of a more desolvated cavity, the conditions may be suitable for hydrogen bonding to occur with fluorine.[59]

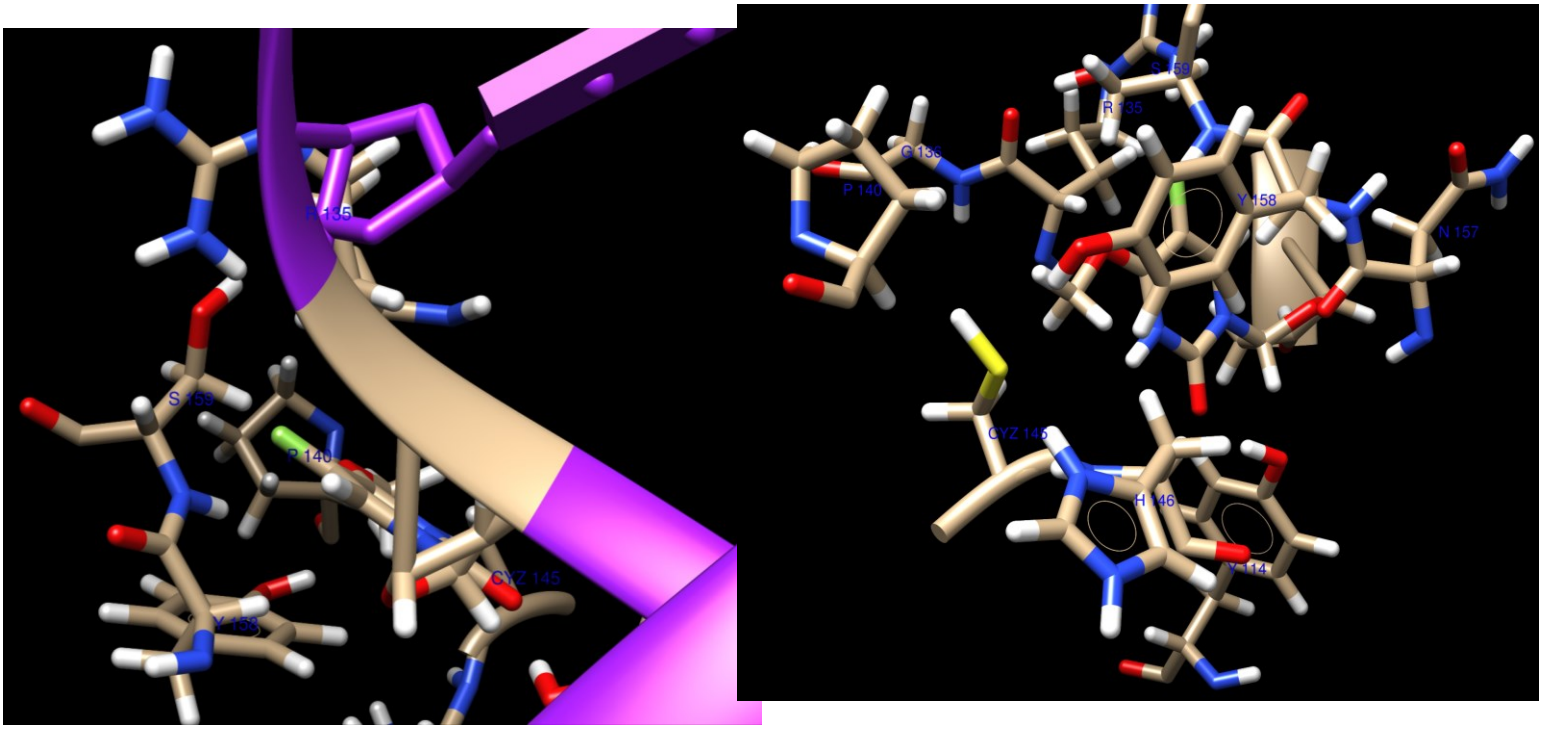


Figure 3.5: Favourable orientation of C5-F (green atom) for hydrophobic and π-sytem interaction with Pro140 (left) and Tyr158 (right) respectively. (Modified from PDB 1T38 using UCSF-Chimera)

The existence of short contacts between a fluorine atom and different amino acids have been reported.[61,116] Short contacts between the thiol from cysteine and fluorine have been described. [61] Histidine has been reported to be able to interact with fluorine via its imidazole ring.[74]

The high reactivity of Cys145 in AGT has been studied by several groups. Reports have suggested that Cys145 has a low pKa that allows it to be in the thiolate form at neutral pH.[50] It is possible that the fluorine atom at the C5 position of the O^4-alkylated nucleobase, interacts with key residues such as Cys145 (C-F bond and thiolate short contacts or even H-bonding) and His146 (C-F bond and π-system interaction) in the AGT active site, enhancing the reactivity of the thiolate anion (Figure 3.6).[61,63,74]

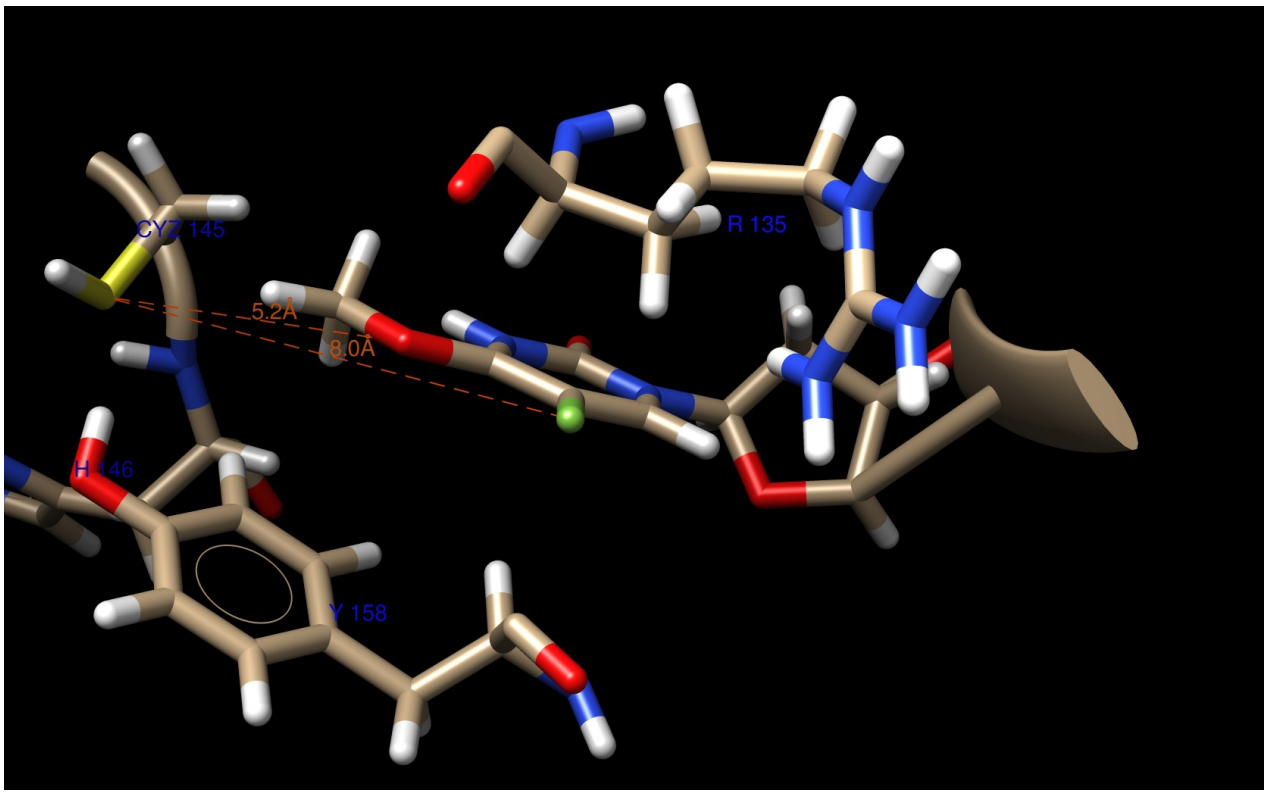


Figure 3.6: hAGT active site with 5-fluoro-O^4-methyl-2'-deoxyuridine displaying distance from C145 sulphur (yellow atom) to the O^4 (red atom) and C5-F (green atom) respectively. (Modified from PDB 1T38 using UCSF-Chimera)

To test this hypothesis, it may be worthwhile to prepare AGT mutants to probe these possible interactions. In other studies, mutations at key residues have shown to reduce AGT repair activity. For example, Y114F or Y114A mutations highlight the importance of Y114 in flipping of the nucleoside and positioning of the alkyl lesion.[53,93] Likewise, the P140K mutation revealed the importance of hydrophobic interactions between this residue with benzyl moieties at the nucleobase.[35] Although the shape and properties of some amino acids in the active site are different between AGTs, it may be interesting to explore electrostatic interactions between the C5-fluoro group of the nucleobase and key conserved residues. Hydroxyl and guanidinium groups have shown to be fluorophilic and amino acids such as tyrosine (Tyr114 and Tyr158) and arginine (Arg135) found in the AGT active site contain these groups (Figure 3.6).[61,116]

In this study it was observed that the different AGTs deviated from the patterns observed regarding repair proficiency of various O^4-alkylated T and dU lesions. It is clear that the presence of fluorine at the C5-position of the nucleobase generally aids in achieving efficient repair (graphic examples shown in Figure 3.7). We hypothesize that eventually a limit is reached where there is insufficient space available in the active site cavity to accommodate bulky lesions. It is proposed that the C5-fluorine of the alkylated nucleobase could be acting like an allosteric effector with possible interactions with AGT active site residues that may contribute to enhance activity towards the α-carbon. We are currently synthesizing intra- and inter-strand crosslinks with C5-fluoro pyrimidines in order to determine whether enhanced repair is also observed for these lesions.

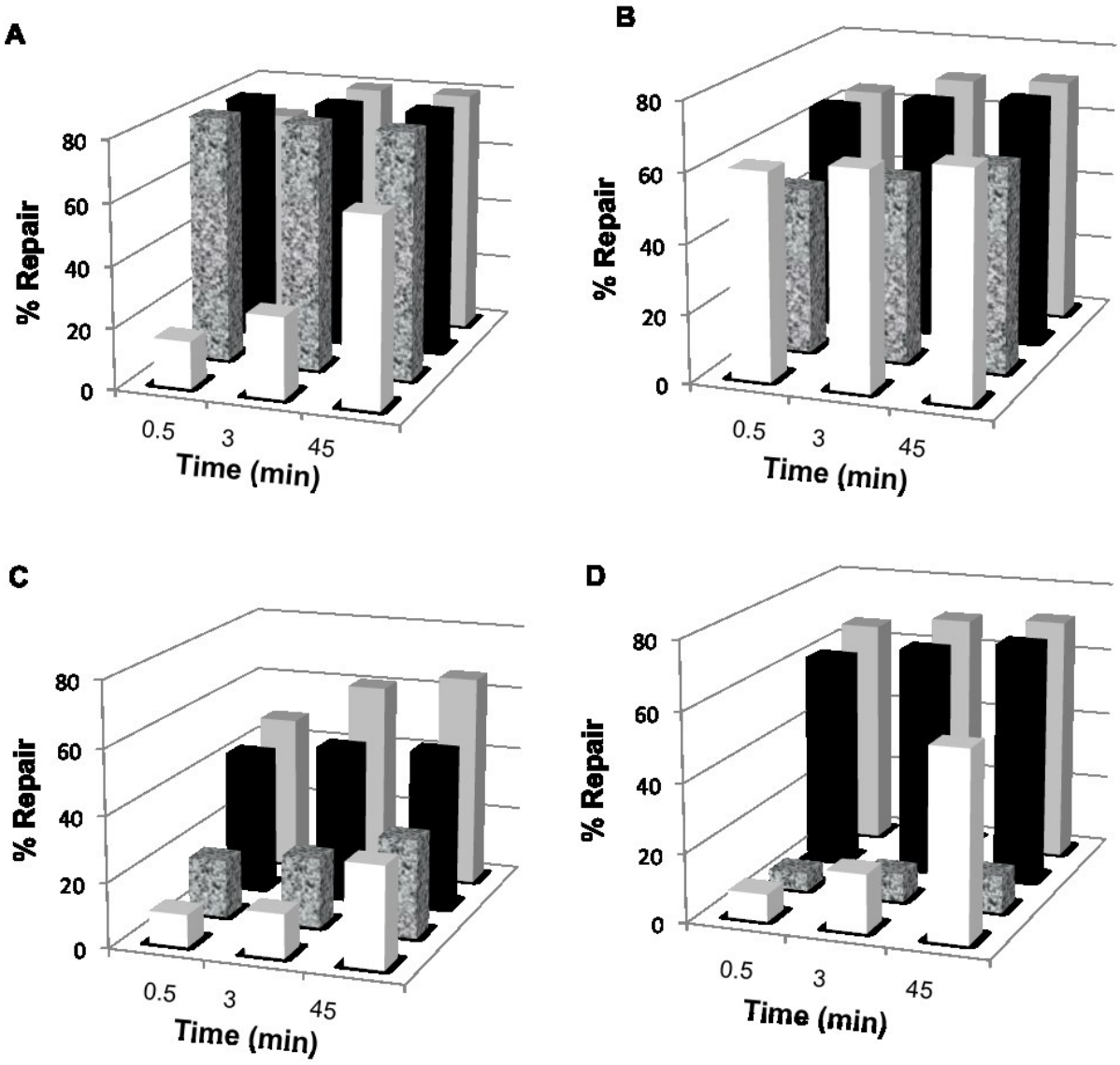


Figure 3.7: Repair of **A)** dFU-Et, **B)** dFU-Bn, **C)** dFU-4COH and **D)** dFU-7COH at time points by **hAGT** (white bars), **Ada-C** (granite bars), **OGT** (black bars), and **hOGT** (light grey bars).

3.4 Conclusion

The synthesis of O^4-ethyl-dFU, O^4-benzyl-dFU, O^4-(4-hydroxybutyl)-dFU, and O^4-(7-hydroxyheptyl)-dFU phosphoramidites as well as their incorporation into oligonucleotides by solid-phase synthesis is described. The UV thermal denaturation studies revealed a decrease of approximately 10 ^{o}C upon the introduction of the alkyl groups at the O^4-atom for the various

adducts investigated. CD spectra of the DNA duplexes containing these O^4-modifications suggest minimal differences from a standard B-form structure.

Generally, the repair studies showed an enhancement in AGT-mediated repair due to the C5-fluoro group. Surprisingly, hAGT demonstrated the ability to repair 4-hydroxybutyl and 7-hydroxyheptyl lesions at the O^4-position of 5-fluoro-2'-deoxyuridine, whereas this protein is incapable of repairing these adducts for thymidine and 2'-deoxyuridine. To the extent of our knowledge, we have shown the first examples of Ada-C repairing larger O^4-adducts like benzyl and 4-hydroxybutyl for 5-fluoro-2'-deoxyuridine. Repair by OGT occurred towards all O^4-lesions evaluated, except for 4-hydroxybutyl which occurred slower for 5-fluoro-2'-deoxyuridine compared to 2'-deoxyuridine. The chimera hOGT achieved fast repair for all O^4-adducts similar to OGT. It was noted that repair of 7-hydroxyheptyl and 4-hydroxybutyl groups at the O^4-position by hOGT occurred about 2x and 40x faster respectively for 5-fluoro-2'-deoxyuridine relative to 2'-deoxyuridine. Overall, it was observed that repair of the 7-hydroxyheptyl adduct was preferred over 4-hydroxybutyl, with exception of Ada-C (where no repair of 7-hydroxyheptyl was observed). The longer adduct is inherently more flexibile, which may allow for proper orientation of the O^4-C_α bond in order for alkyl transfer to the activated thiolate group. The active site of Ada-C may be too small to accommodate the 7-hydroxyheptyl adduct.

In general, the C5-fluoro group enhanced AGT repair of O^4-alkyl pyrimidinyl lesions. This modification enabled lesions at the O^4-position that were otherwise inert, such as the 4-hydroxybutyl and 7-hydroxyheptyl, to be acted on by certain AGTs (hAGT for instance). This may be attributed in part to favourable interactions of fluorine and the AGT active site enabling repair of a wider scope of substrates. Further investigation of the basis of the interactions occurring with the dFU substrates will enhance our understanding of AGT-mediated repair.

Chapter 4. Influence of halogens at the C5-position of O^4-alkyl pyrimidines towards repair by O^6-alkylguanine-DNA alkyltransferases.

4.1 Abstract

Oligonucleotides containing 5-chloro-O^4-methyl-2'-deoxyuridine (dClU-Me) or 5-bromo-O^4-methyl-2'-deoxyuridine (dBrU-Me) were prepared by solid-phase synthesis. Biophysical studies of the DNA duplexes containing these modifications revealed an 11 ^{o}C decrease in thermal stability compared to the duplexes containing 5-chloro-2'-deoxyuridine (dClU) or 5-bromo-2'-deoxyuridine (dBrU). Circular dichroism profiles demonstrated that the modified duplexes maintained an overall B-form DNA structure. Despite the larger size of these atoms compared to fluorine or hydrogen, their presence at the C5-position did not impair the ability of various AGTs to remove O^4-methyl group from these modified 2'-deoxyuridines. Surprisingly, dClU-Me was repaired efficiently by human AGT (hAGT) while other AGT proteins (AdaC and OGT from *E. coli* and a chimera hAGT/OGT protein "hOGT") only displayed moderate repair towards this lesion. For dBrU-Me, all AGTs repair similarly with hAGT demonstrating the least proficiency towards this adduct. These results suggest a combination of factors contributing to the repair proficiency observed between (1) the influence of inductive and resonance effects of C5-chloro and C5-bromo groups on the heterocycle and (2) the possible favourable interactions (halogen-bonding, short contacts, etc.) of the C5-halogens with AGT's active site residues.

4.2 Introduction

DNA damage in the form of halogenated nucleobases can result from different sources including some inflammatory processes or the presence of hypobromous and hypochlorous acids.[111,117–119] One of the sites most susceptible to modification is the C5 position of the pyrimidine nucleobases. The presence of 5-halogenated pyrimidines can have biological consequences which affect DNA-protein interactions and processes such as C5-methylation.[111,120]

The steric and electronic effects resulting from halogenation are dependent on the identity of the halogen. It is well documented that fluorine is commonly used as a substitute for hydrogen atoms, although the former is slightly larger whereas chlorine and bromine are closer to the size of a methyl group.[111] Recently, the presence of halogens in many biological systems has been found to exert interesting properties. Halogen bonding, which is the non covalent interaction between an electrophilic region (σ-hole) in the halogen atom with the nucleophilic part of another molecule, and short contacts are examples of interaction of halogen atoms with proteins and nucleic acids.[121]

The influence of replacing the methyl group of 5-methylcytidine with various halogens on endonuclease activity has been evaluated. It was shown that *MspI* cleavage activity was unaffected by the presence of halogens at the C5 position whereas *HpaII* activity was reduced in the presence of a fluoro group and completely absent in the presence of either a chloro or bromo group.[111] Some halogens have a minimal effect on DNA structure whereas larger ones may introduce a structural deformation. It was concluded that steric effects mainly impacted the protein-DNA interactions and that the presence of chloro or bromo groups could simulate the effect of methylation.[111] In other studies it was noticed that only 5-fluorocytosine could act as a methyltransferase trap, whereas the chlorinated and brominated analogues were not able to function in the same way probably due to steric reasons.[111]

Halogen substituents in pyrimidines have been studied for decades and found applications as therapeutics. They have been employed in chemotherapy as radiosensitizers due to their large cross-sectional area that enables the absorption of radiation and their Auger effect that can induce single and double strand breaks in DNA.[65,122,123]

O^6-alkylguanine-DNA alkyltransferases (AGTs) are known for being capable of repairing alkylation at the O^4-position of thymidine, with some variants displaying greater proficiency relative to others. It has been proposed that C5-methyl group of thymidine may contribute to the reduced repair activity of human AGT towards O^4-alkyl lesions.[54] The electron-donating nature of the methyl group has made our group interested in analyzing the effect of electron withdrawing groups at the C5-position. It was observed in studies performed by our group that

the presence of fluorine dramatically increased AGTs efficiency to repair O^4-methyl adducts. Hence, the purpose this study is to analyze the effect of other halogens including chloro and bromo groups at the C5 position, taking into account electronic and steric effects, on repair of O^4-methylated pyrimidines by AGTs.

The presence of halogens on the nucleobases have been shown to influence DNA duplex stability. Halogens can alter the pKa of different groups and can influence stabilization of the DNA duplex. Studies performed with 5-halogenated cytosine revealed that the pKa of the N3 atom decreases.[111,124] This change in pKa reduces the strength or even removes one of the three hydrogen bonds in CG base pairing.[111,112] Nevertheless, C5-halogenations increase the pyrimidine's polarizability, which leads to an increase of base-stacking interactions, possibly contributing to a more stable DNA duplex (higher thermal stability).[111,125]

Halogen bonding and short contacts (short distances between a pair of atoms) have been observed in nucleic acids-protein structures.[116] The versatility of these halogens allows them to have short contacts with O and N atoms such as the backbone carbonyl groups in proteins.[115,116] Even though halogen bonds are weaker than hydrogen bonds, they are involved in hydrophobic interactions and even form halogen-water bridges, contributing to stability.[115,116] The strength of halogen bonding has been observed to increase from chlorine to bromine due to the higher polarizability of the latter.[116,126,127] The moderate strength of the bromine bond could be one of the reasons why nucleic acids have a tendency to more readily be brominated rather than chlorinated.[116]

The distance required for hydrogen-bond formation to halides varies depending on the donor. Normally this distance increases going through the series from fluoride to iodide.[128] Substitutions of hydrogens and hydroxyl groups for halogens in proteins have shown to increase their thermal and structural stability. For this and other reasons halogens have been also employed for molecular imaging and structural studies.[115]

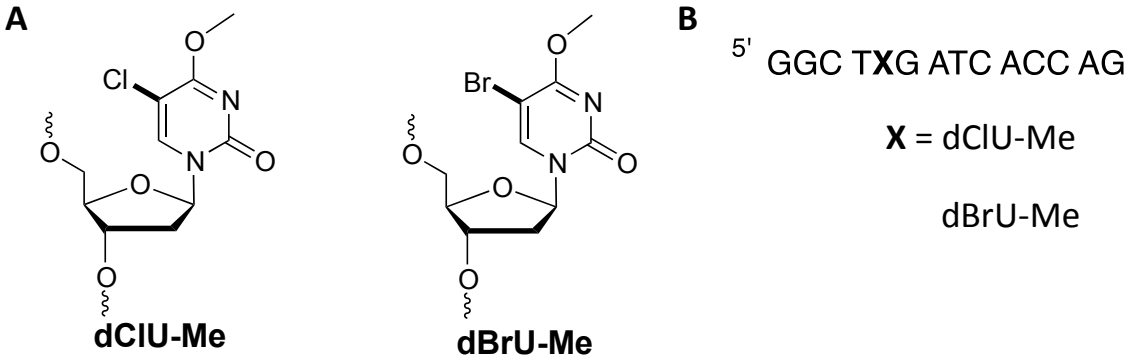


Figure 4.1: (**A**) Structures of 5-chloro-O^4-methyl-2'-deoxyuridine (dClU-Me) and 5-bromo-O^4-methyl-2'-deoxyuridine (dBrU-Me). (**B**) DNA sequence where X corresponds to the modified nucleoside.

4.3 Results and Discussion

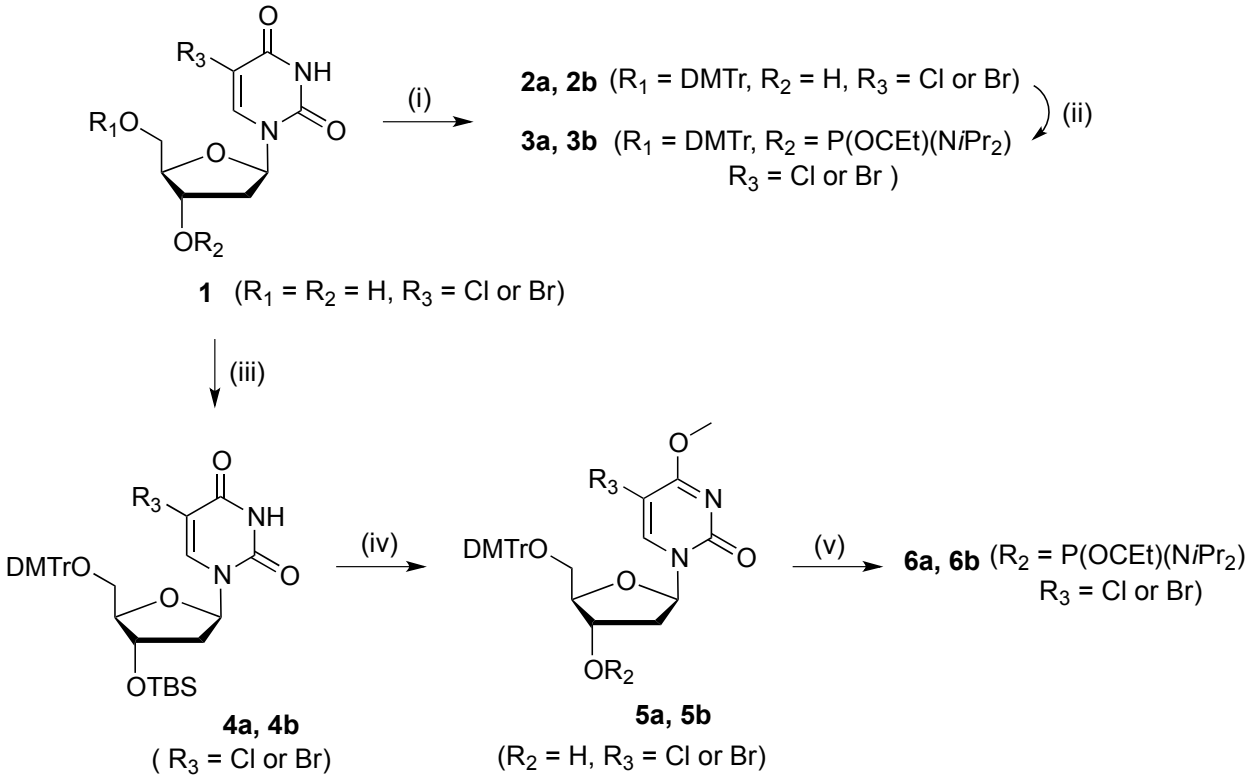
4.3.1 Synthesis of nucleosides and oligonucleotides

The structure of dClU-Me, dBrU-Me and the DNA sequence where they are introduced are shown in **Figure 4.1**. These nucleobase modified oligonucleotides were use to study the size and electronic influence of the C5-group relative to their T (C5-methyl), dU (C5-hydrogen) and dFU (C5-fluorine) counterparts.[54] Previous studies in our group demonstrated the positive influence of fluorine at the C5-position on repair by AGTs possibly due to favorable interactions occurring at the protein's active site as well as the influence that the fluorine atom may have on the heterocycle making the O^4-C_α bond more susceptible for the thiolate attack. (Chapters 2 and 3) The investigation of the halide series, in hand with our studies involving the fluorine modification, will give us a better understanding on the contributions of size and/or the electronic effects of groups at the C5-position on AGT repair.

The modified oligomers were obtained after small molecule and solid-phase synthesis. The synthetic pathway shown in **Scheme 4.1** begins with the protection of the 5'- and 3'-end of the commercially available 5-chloro-2'-deoxyuridine (**dClU**) and 5-bromo-2'-deoxyuridine (**dBrU**) using previously published procedures.[97,98] The preparation of the convertible

intermediate was accomplished by the slow addition of 1,2,4-triazole, triethylamine and phosphoryl chloride while stirring at 0°C to a solution containing 3'-*O*-(*tert*-butyldimethylsilyl)-5'-*O*-(4,4'-dimethoxytrityl)-5-chloro (or bromo)-2'-deoxyuridine.[54] The convertible nucleosides were then converted to compounds **5a** and **5b** by treatment with sodium methoxide in methanol, followed by the removal of the 3'-*O*-*tert*-butyldimethylsilyl group by a fluoride treatment. Finally compounds **2a**, **2b**, **5a** and **5b** were phosphitylated to yield phosphoramidites **3a**, **3b**, **6a** and **6b** according to previously published procedures.[54] Purification of all the phosphoramidites was achieved by using short flash column chromatography and the compounds were characterized by ^{31}P NMR spectroscopy (**Figures A3.3**, **A3.10**, **A3.13** and **A3.20**). The ^{31}P NMR spectra of each phosphoramidite revealed two signals near 147-149 ppm region, which is diagnostic for a phosphoramidite.[54]

Due to the labile nature of dFU-Me and other O^4-adducts, mild deprotection conditions were employed for oligonucleotides containing these halogen analogues.[88] "Fast-deprotecting" commercially available 3'-*O*-phosphoramidites were used for solid-phase synthesis and phenoxyacetic anhydride was used as the capping reagent.[99] Total deprotection and cleavage of the oligomers from the solid support was accomplished with an anhydrous solution of potassium carbonate in methanol (0.05 M) for four hours at 21°C under gentle rocking, followed by neutralization of the excess base with an equimolar amount of acetic acid.[88,97] Oligonucleotides containing the different modifications (**dClU**, **dClU-Me**, **dBrU** and **dBrU-Me**) were purified by SAX-HPLC and characterized by ESI-MS. The ESI-MS results showed the presence of the modifications and the observed masses are in agreement with the expected masses (**Figures A3.21**, **A3.22**, **A3.23** and **A3.24**). Enzymatic digestion of modified oligonucleotides using snake venom phosphodiesterase and calf intestinal phosphatase enzymes, followed by analysis using reversed phase HPLC, confirmed the nucleoside composition of the oligomers to be consistent with expected ratios (**Figures A3.25 and A3.26**).



Scheme 4.1: Reagents and conditions: (i) DMTr-Cl, pyridine, DMAP, 16 h, 21 ^{o}C; (ii) *N,N*-diisopropylaminocyanoethylphosphonamidic chloride, DIPEA, THF, 30 min; (iii) DMTr-Cl, pyridine, DMAP, 16 h, 21 ^{o}C.; TBS-Cl, Imidazole, DCM, 16 h, 21 ^{o}C; (iv) 1,2,4-Triazole, triethylamine, $POCl_3$, MeCN, 16 h, 0 ^{o}C; MeOH, NaOMe, 4 h, 21 ^{o}C.; TBAF (1M in THF), 30 min; (v) *N,N*-diisopropylaminocyanoethylphosphonamidic chloride, DIPEA, THF, 30 min.

4.3.2 UV thermal denaturation and CD spectroscopy

UV thermal denaturation studies were conducted in order to assess the influence of chlorine or bromine at the C5-position of the O^4-methylated adducts on stability of the DNA duplexes. The results are summarized in **Figure 4.2** and suggest that upon the methylation at the O^4-position, there is a decrease in thermal stability of 11 ^{o}C relative to the unmodified controls. A similar decrease was observed for dFU-Me and the T and dU analogues.[54] The origin of the decrease in thermal stability is most likely attributed to the disruption of hydrogen

bonding between the dClU-Me:dA and dBrU-Me:dA base pairs and their influence on diminished stacking with the flanking base pairs. It was observed that the dClU and dBrU controls had a moderately higher T_m value than the dFU (and dU) control (Figures 4.2 and A3.31). The difference may be attributed to the higher polarizability of chlorine and bromine, relative to fluorine, on the pyrimidine ring that can improve base-stacking.[111,125] The dClU and dBrU containing duplexes have a thermal stability slightly lower relative to the T control (62 ^{o}C). The close resemblance to the T control suggests that C5-chlorine and C5-bromine behave more like a methyl than fluorine in terms of their contribution to duplex stability.[111] Overall, the presence of chlorine and bromine at the C5 position of 2'-deoxyuridine exhibits a minimal influence on the DNA duplex stability.[112]

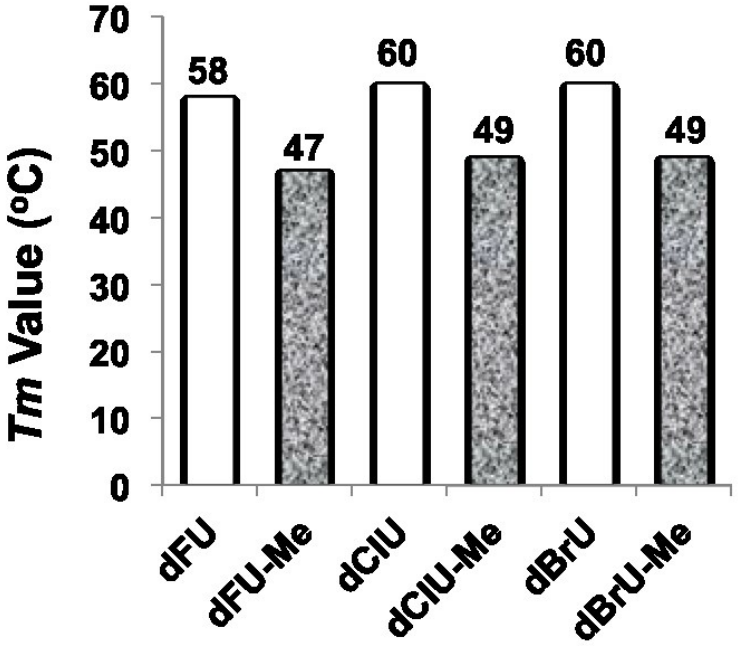


Figure 4.2 : T_m values (^{o}C) of duplexes containing **dFU, dFU-Me, dClU, dClU-Me, dBrU, dBrU-Me**. Colorless and grey bars represent the unmodified controls and methylated adducts, respectively.

The CD profiles shown in **Figure 4.3** display spectroscopic signatures consistent with a B-form DNA family (a maximum at around 280 nm and minimum at approximately 245 nm). The CD signatures of **dClU-Me** and **dBrU-Me** were similar to their unmodified counterparts (**dClU** and **dBrU**) and the DNA duplex containing T at the same position. The results suggest that the

presence of the O^4-methyl group did not cause any major alteration in the overall structure of the DNA, as previously observed for dFU, dU and T series.[54] The small variations observed in the CD data and the difference in molar absorptivity of the signatures are related to the change in the ellipticities due to the 5-halogen substitutions. From the different 5-halogen substitutions studied by our group, the duplex containing the C5-fluorine group has the highest ellipticity while the lowest ellipticity is observed for the C5-bromine modification which has a higher polarizability.[129] The electronic effect of each halogen has an influence on the ellipticities observed. Overall, the introduction of the C5-chlorine or C5-bromine group did not induce a significant distortion to the global B-form DNA structure.[112]

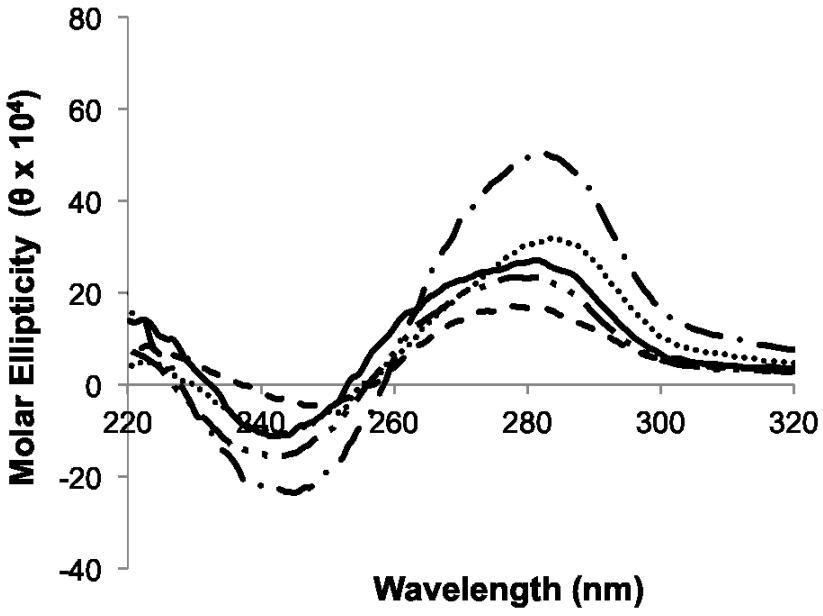


Figure 4.3: Circular dichroism spectra of duplexes containing **dClU** (—), **dClU-Me** (•••), **dBrU** (———), **dBrU-Me** (—•—) and **T** (—••—) .

4.3.3 Repair Assays

Repair of the C5-halogenated O^4-methyl adducts by different AGTs including human AGT (hAGT), *E. coli* OGT and Ada-C in addition to a hAGT / OGT chimera (hOGT) was evaluated. These proteins exhibit a wide variety in substrate preference and active site residues between

them. Exploring their proficiency at removing O^4-methyl lesions allows us to examine if the C5-halogens and their associated steric and electronic effects can influence repair.

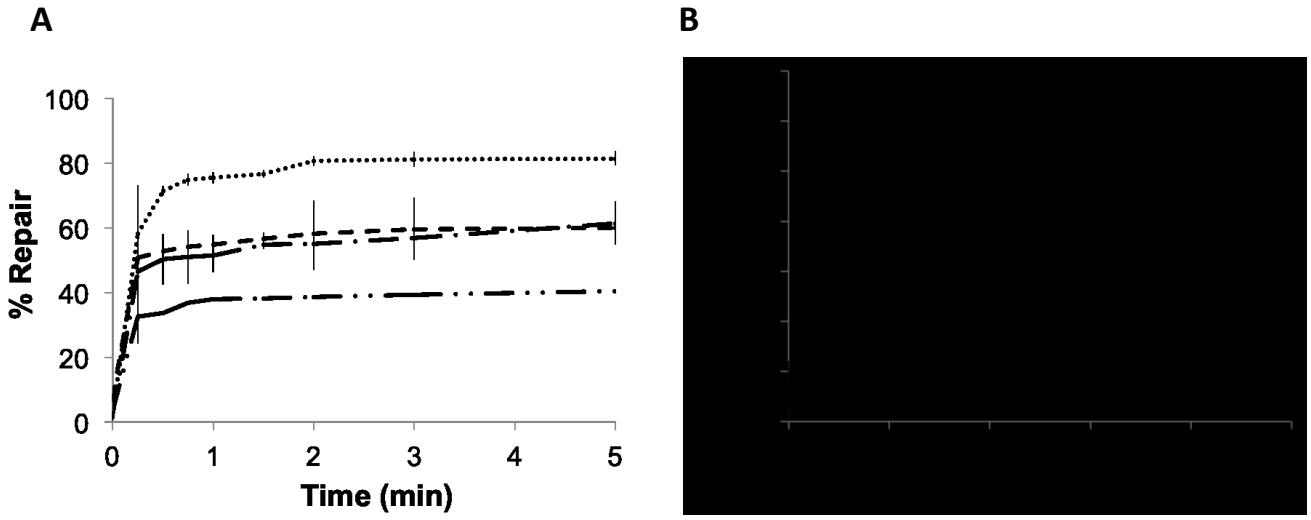


Figure 4.4: Time course repair assay of (**A**) **dClU-Me** and (**B**) **dBrU-Me** by hAGT (•••), OGT (———), Ada-C (—•—) and hOGT (—••—) at 5-molar equivalence of protein.

All four AGTs in this study were found to repair **dClU-Me** (**Figure 4.1 A and Figure A3.38**). hAGT was the most efficient, demonstrating full repair of **dClU-Me** in less than 2 min. Surprisingly, repair of **dClU-Me** by hAGT appears to be more efficient than the **dFU-Me** analogue. This was not expected as it was assumed that the bulkier chlorine atom could encounter steric clashes in hAGT's active site. Clearly the C5-chlorine group is readily accommodated and may be participating in favorable interactions with active site residues. The chimera hOGT protein, where the hAGT protein has twenty active site residues replaced by these found in OGT (except for the active site Pro140 residue), displayed reduced proficiency towards the **dClU-Me** substrate achieving almost 50% repair after reaction overnight.

The *E. coli* variants (OGT and Ada-C) were found to have similar proficiency towards **dClU-Me**. These variants reached about 65% repair, which remains at a constant level after 45 min (**Figure A3.38**). Although OGT and Ada-C are known to efficiently repair small lesions such as methyl groups, the presence of the C5-chlorine group limits a high level of repair efficiency.

Ada-C has been described to have a small and crowded active site, which may not be able to accommodate the C5-chlorine group of the O^4-methylated nucleobase. Some experiments have shown the favorable interactions of chlorine with tryptophan. Therefore even though the pocket might be small, it may be compensated by possible interactions between C5-chloro group and Trp161.[115,121]

For the larger **dBrU-Me** adduct, moderate repair by the various AGTs was observed. hAGT was found to be the least efficient of the AGT proteins at repairing **dBrU-Me** achieving 45% repair. Ada-C and OGT were slightly more efficient with approximately 55% repair observed. The hOGT chimera's repair efficiency towards the **dBrU-Me** adduct was only slightly better than hAGT (a similar trend was observed for **dFU-Me** whereas the opposite occurred for **dClU-Me**). Perhaps with a group the size of bromine at the C5-position, a point is reached where the influence is the same for all the AGT proteins evaluated. If this is the case, the repair observed will come mostly from favorable interactions of bromine with some proximal residues as well as its own electronic effects on the heterocycle.

From the repair outcomes observed between the two substrates, the most significant difference between the AGT variants was observed for hAGT, which performs full repair of **dClU-Me** and only 46% repair of **dBrU-Me**. For the AGTs investigated in this study, the human variant has been described to contain a slightly larger active site relative to the others, which better accomodates a C5-chlorine versus C5-bromine group.

Despite chlorine and bromine having a reduced inductive (electron withdrawing) and conjugation (electron donating) effect relative to fluorine, they have been shown to interact with different residues in proteins.[56] Halogens prefer to interact with hydrophobic amino acids (such as leucine and phenylalanine), carbonyl groups of the protein backbone and with other H-bonding residues such as serine.[115,121] It is proposed that that the C5-chloro and C5-bromo groups of 2'-deoxyuridine could be interacting with the Ser159, which forms part of the active site wall of hAGT.[51,115] Studies in other protein systems (such as hydronitrile lyase) have shown the formation of hydrogen bonds between chlorine and the residues tryptophan and serine.[115] Hence, the significant repair efficiency observed by hAGT on **dClU-Me** could be attributed to

the formation of halogen bonds between the C5-chloro group and residues Tyr158 and Ser159. In the case of Ada-C, the C5-chloro group could be participating in similar bonds with Trp161. The average distance required for the formation of hydrogen-bonds increases as the halogen gets larger and it is normally around 3.0 Å to 3.3 Å.[115]

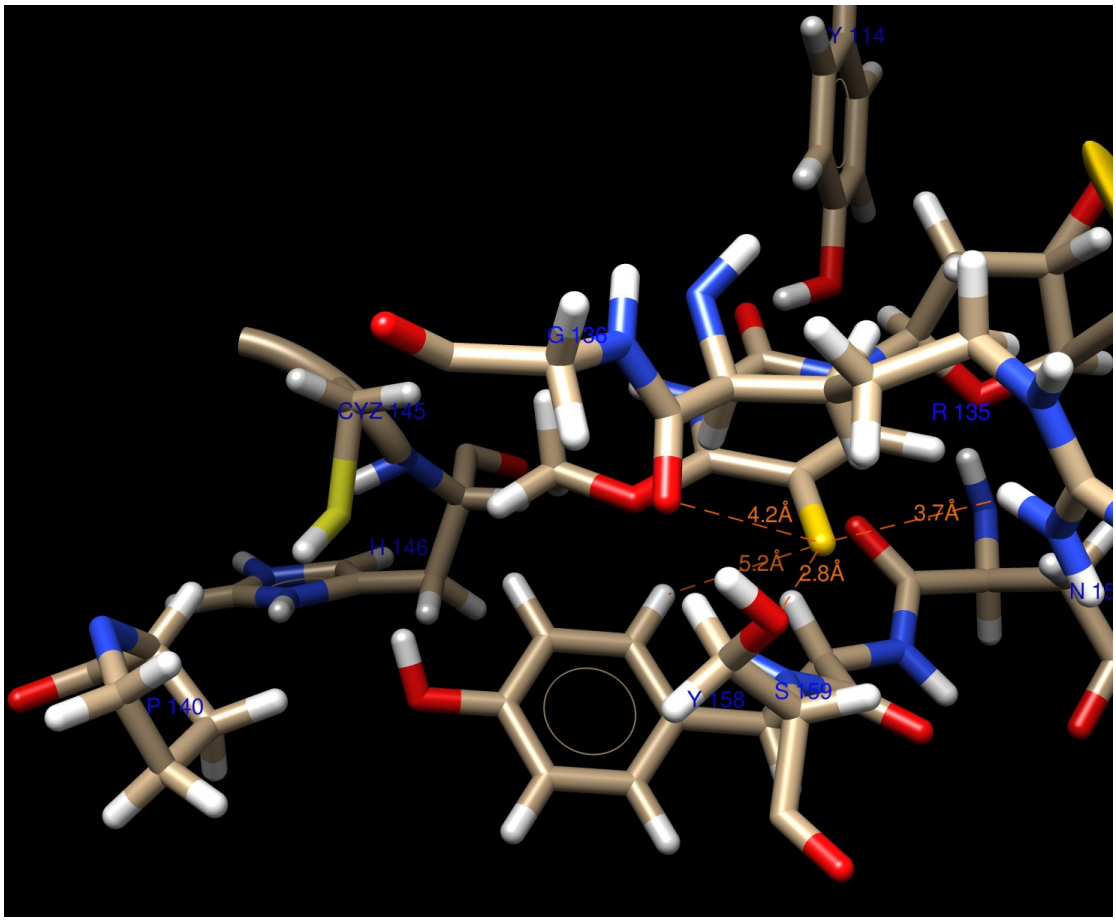


Figure 4.5: Interactions that might occur between C5-Chloro (gold atom) and the residues from hAGT active site. (Modified from PDB 1T38 using UCSF-Chimera)

Other studies have shown that chlorine has an inclination to interact with leucine and bromine with arginine.[115] Arg160 is present in the active site of Ada-C and despite having a smaller pocket than hAGT, the presence of C5-bromo group does not impare the ability of Ada-C to remove the O^4-methyl adduct.[31] hAGT contains Arg135, which in previous studies has been suggested to sterically clash the aliphatic portion of its side chain with the C5-methyl

group of T. Perhaps the C5-bromo group might be interacting with Arg135, despite the size of the halogen.[34] Tyr158 also forms part of Ada-C's active site pocket and thus it could possibly hydrogen bond with the C5-bromo group or even to form a halogen bond with the tyrosine's π-system (Figure 4.7).[31,115,121]

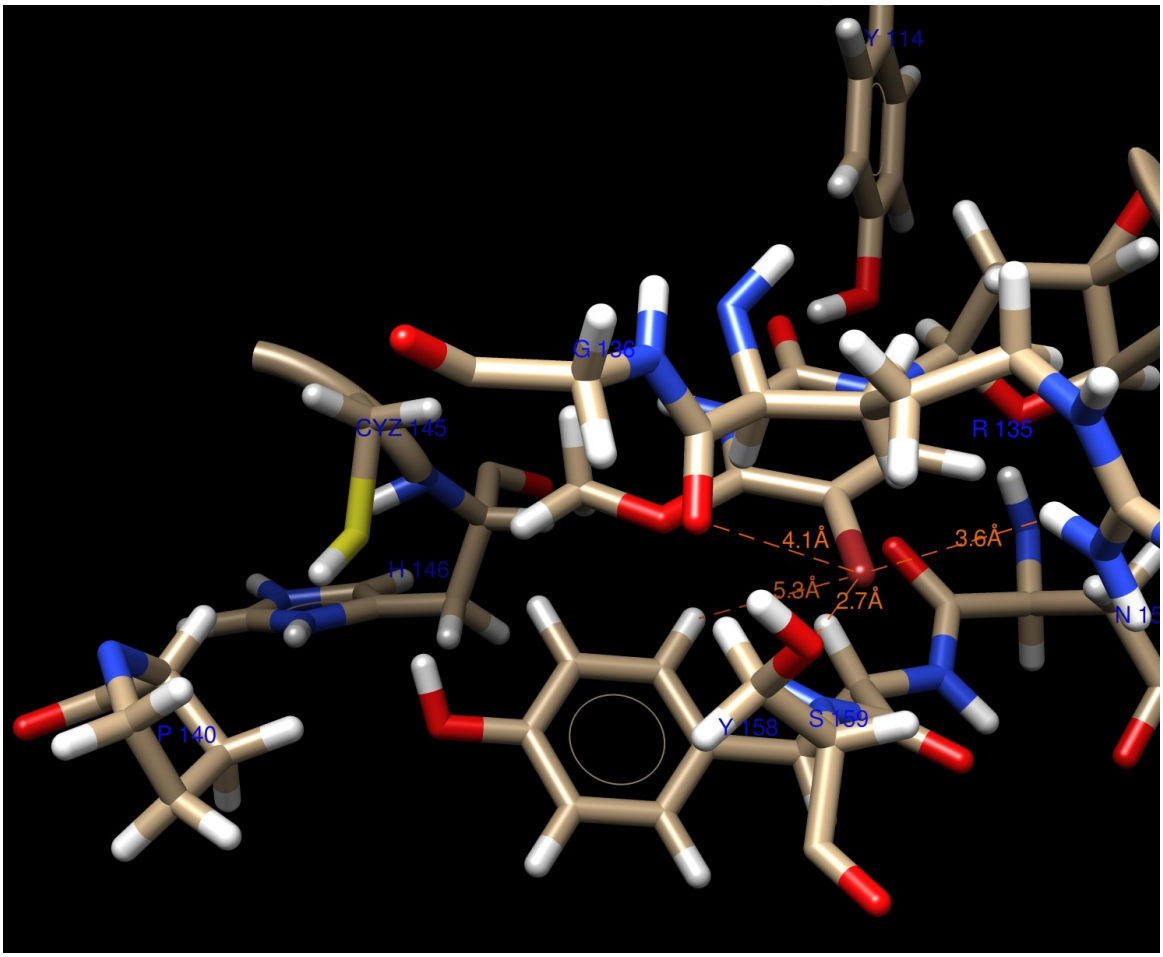


Figure 4.6: Interactions that might occur between C5-Bromo (darker red atom) and the residues from hAGT active site. (Modified from PDB 1T38 using UCSF-Chimera)

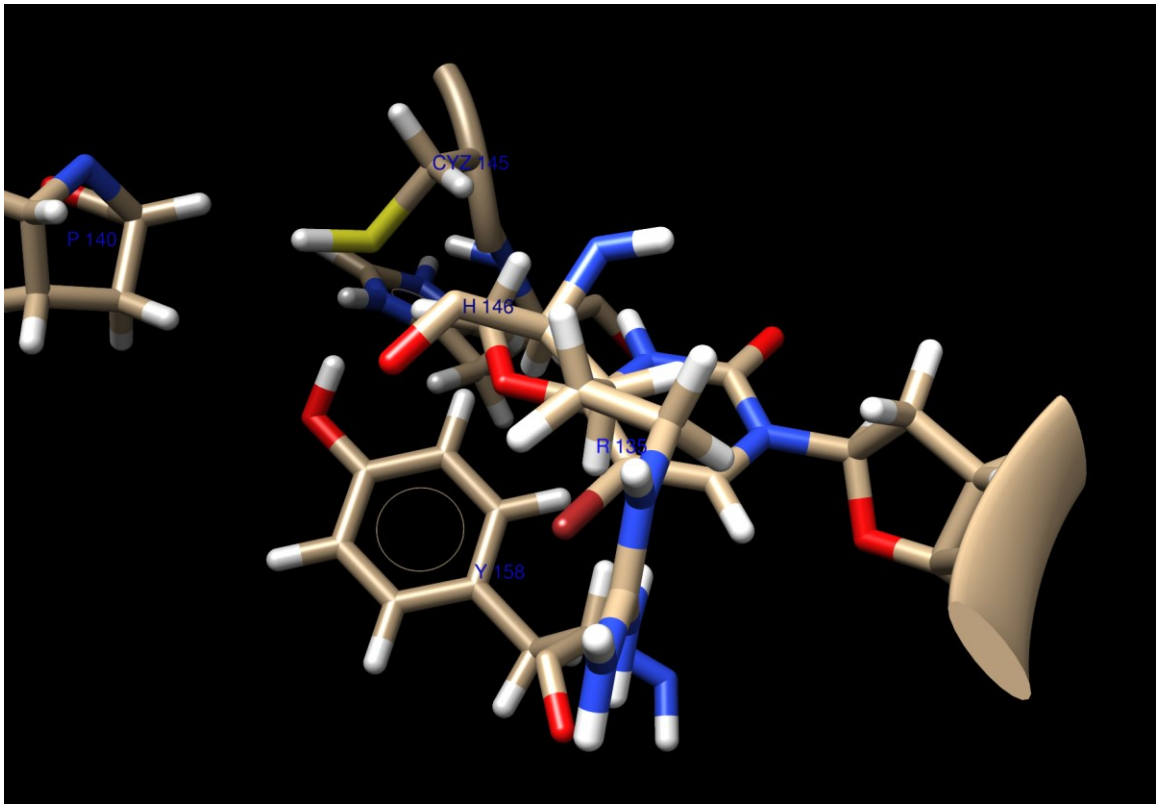


Figure 4.7: Possible favourable interaction of C5-Bromo (darker red atom) with Tyr158. (Modified from PDB 1T38 using UCSF-Chimera)

Both OGT and hOGT have residues in their active site that have been reported to interact with chloride and bromide. Residues such as Ser134, conserved Arg141, Arg145, Thr148, Thr150 and Tyr152 (the numbering corresponds to the OGT protein) might potentially be interacting with the C5-chloro and C5-bromo groups via H-bonds if they are at an optimal distance or they may engage in hydrophobic interactions.[130] Unfortunately, to our knowledge, no crystal structures of these proteins are available at the moment to aid in this hypothesis.

Other examples of interactions related to the present study are those which occur between aryl halides and aromatic groups.[115,121] In addition, sulfur containing amino acids have also been found to act as halogen bonding partners.[121] Water molecules can also engage in halogen bonding between the oxygen atom and the halogen (with optimal distances between 2.6 Å to 3.6 Å).[115] The presence of halogen-water bridges with different residues has been observed.[116,126] A modification of interest to expand upon the present study is 5-iodo-O^4-

methyl-2'-deoxyuridine. 5-iodouracil is used clinically in order to enhance the efficacy of radiotherapy.[65] It would be interesting to investigate repair of 5-iodo-O^4-methyl-2'-deoxyuridine by AGTs.

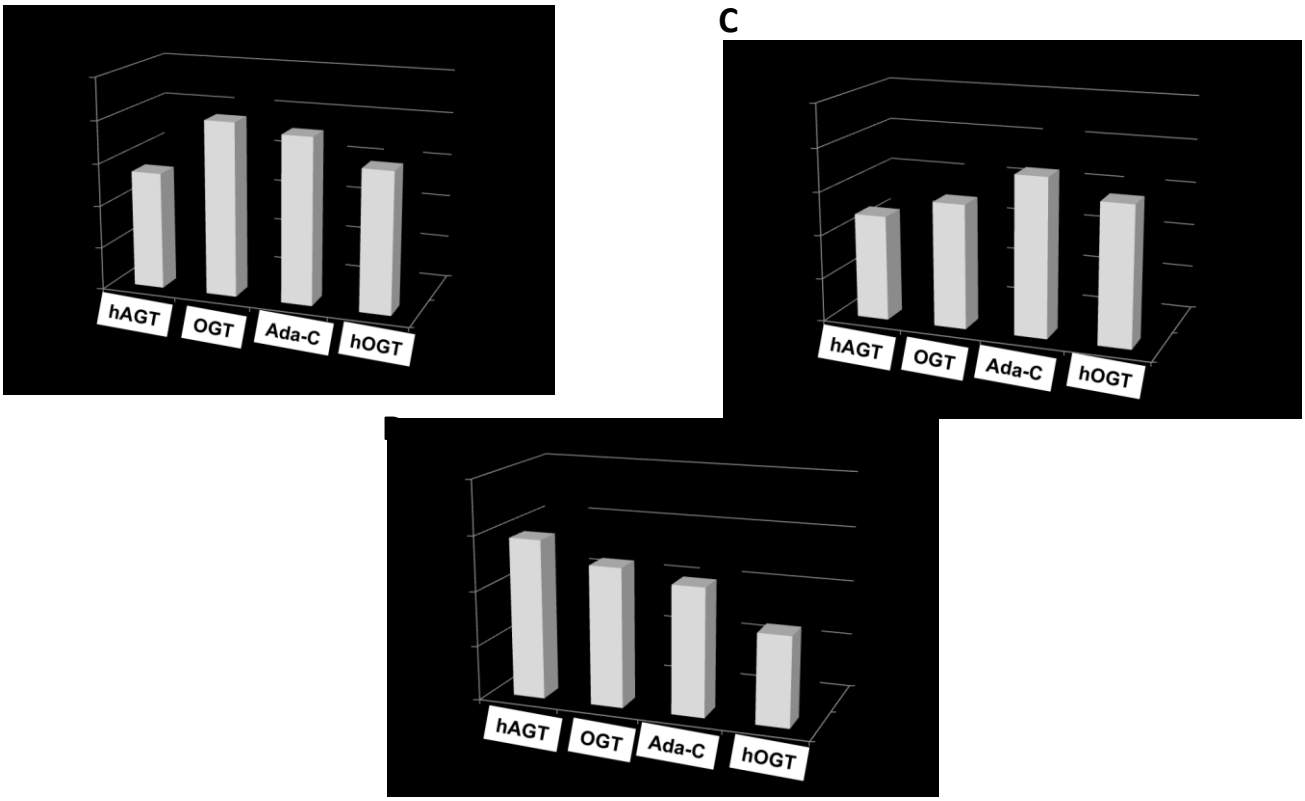


Figure 4.8: Comparison of the repair at 15 sec and 1 min of (**A**) **dFU-Me**, (**B**) **dClU-Me** and (**C**) **dBrU-Me** by hAGT, OGT, Ada-C and hOGT.

4.4 Conclusion

The synthesis and incorporation of 5-chloro-O^4-methyl-2'-deoxyuridine and 5-bromo-O^4-methyl-2'-deoxyuridine phosphoramidites into oligonucleotides by solid-phase synthesis was accomplished. The UV thermal denaturation and circular dichroism studies revealed a decrease of 11 ^{o}C upon the methylation at the O^4-atom for both the 5-chloro and 5-bromo derivatives and with no major alteration on the global structure of the duplexes relative to B-form DNA.

The repair studies of **dClU-Me** revealed that hAGT was efficient (in fact, slightly better relative to the fluorinated analogue) at removing the O^4-methyl lesion. AGT proteins described as having a smaller active site (*E. coli* Ada-C and OGT and the chimera hOGT) were able to repair **dClU-Me** with 42-65% repair observed (Figure 4.8 B).

Despite the fact that chlorine is larger than fluorine and hydrogen, it demonstrates favourable influence on repair efficiency of an O^4-methyl group by hAGT. The size of chlorine (smaller than methyl group) and its properties might be optimal to (1) engage in interactions with residues in the hAGT's active site and assist in orienting the lesion towards the thiolate in Cys145 and (2) not introduce steric clashes with residues in the hAGT active site.

In the case of **dBrU-Me**, the larger size of bromide did influence repair, with hAGT activity the most affected in spite of its slightly larger active site relative to the other AGTs studied (Figure 4.8 C). For the O^4-methyl adduct only 55% repair was observed by hAGT. Regardless of their larger size and potential steric effects, chloride and bromide have the ability to interact via halogen bonds with residues such as tyrosine, tryptophan, arginine and serine, which might facilitate removal of alkyl lesions at the O^4-position. Furthermore, the inductive and conjugation effects of these halogens, despite being lower than in fluorine, can influence the O^4-C_α bond making it more susceptible to removal by AGTs. Further investigation of C5-halogenated-O^4-alkylated pyrimidines will contribute to our understanding of AGT-mediated repair.

Chapter 5. Conclusions and Future Directions

5.1 Conclusions

Starting from 5-fluoro-2'-deoxyuridine, 5-chloro-2'-deoxyuridine and 5-bromo-2'-deoxyuridine various phosphoramidite derivates of O^4-alkyl adducts were successfully prepared by multistep synthesis. These modified nucleosides were incorporated into oligonucleotides (14 nucleotides in length) by solid-phase synthesis, deprotected and purified. ESI-MS and nuclease digests confirmed the identity and composition of these sequences, ensuring that the deprotection conditions employed did not compromise the adducts at the O^4-position.

UV thermal denaturation studies revealed a decrease of approximately 10^{o}C with the introduction of alkylation at the O^4-position for the lesions prepared. The unmodified analogues exhibited the following trend in duplex stability: T (C5-methyl) > dClU (C5-chloro) ≈ dBrU (C5-bromo) > dU (C5-hydrogen) ≈ dFU (C5-fluoro) (displayed in Figure 5.1). This trend in stabilization most likely arises from differences in stacking interactions of the C5-group of the 2'-deoxyuridines in the duplex. Circular dichroism spectra of all the O^4-modified substrates revealed an overall B-form DNA structure with profiles similar to those of the unmodified (lacking an O^4-alkyl group) duplexes. The small differences in the molar ellipticity likely result from minimal perturbations induced by larger lesions in addition to the influence of the halogens on the molar ellipticity (in the series fluoro > chloro > bromo).[129]

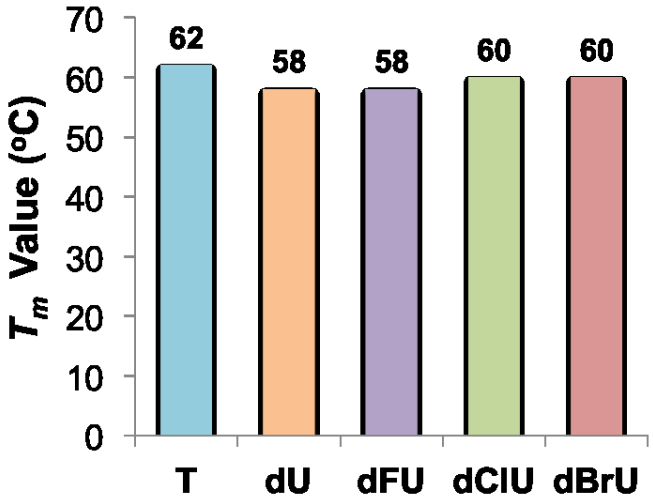


Figure 5.1: T_m values (^{o}C) of unmodified duplexes containing **T**, **dU**, **dFU, dClU,** and **dBrU**.

The presence of the C5-fluorine enhances susceptibility towards repair by different AGTs. The presence of the C5-fluoro group resulted in a 20-fold increase in repair of an O^4-methyl adduct by hAGT relative to 2'-deoxyuridine (containing a C5-hydrogen) and 100-fold incease compared to thymidine (containing a C5-methyl group). The presence of the C5-fluoro group also had a positive influence on AGT repair of larger O^4-adducts. hAGT displayed more repair activity towards dFU-Et, relative to the 2'-deoxyuridine and thymidine analogs. Moreover, hAGT was able to repair the larger dFU-C4OH and dFU-C7OH adducts, which was not observed for the corresponding 2'-deoxyuridine and thymidine adducts.

Surprisingly, Ada-C was able to repair the larger dFU-Bn and dFU-C4OH adducts. For the bulky 4-hydroxybutyl and 7-hydroxyheptyl adducts at the O^4-position, all AGTs displayed greater proficiency towards the latter except Ada-C (where no repair of 7-hydroxyheptyl occurred). The longer adduct is believed to be more capable of adopting an optimal orientation for the alkyl transfer as it is possible that its longer carbon chain could allow it to reach a closer position towards the reactive cysteine. Ada-C's active site might be too small due to the presence of certain residues such as Trp161 preventing accommodation of the 7-hydroxyheptyl adduct. It is proposed that C5-fluorine atom does not encounter steric clashes in the active site and that the electronic properties of fluorine (inductive and resonance effects) might affect the O^4-C_α bond allowing it to be more susceptible towards nucleophilic attack in the AGT active site.

In addition, favorable interactions of the C5-fluorine with active site residues may contribute to the nucleobase adopting an optimal orientation for transfer of the alkyl group.

The influence of size and electronic effects of the C5 group of 2'-deoxyuridine on susceptibility to undergo repair by AGTs was explored with substrates containing O^4-methyl-5-chloro-2'-deoxyuridine (**dClU-Me**) and 5-bromo-O^4-methyl-2'-deoxyuridine (**dBrU-Me**). Unexpectedly (despite to the bigger size of both atoms where the van der Waals radius of chlorine is 1.75 Å and bromine is 1.85 Å compared to 1.47 Å for fluorine), both substrates were repaired by the 4 AGTs. The **dClU-Me** was efficiently repaired by hAGT. Moreover, **dClU-Me** repair by hAGT occurred faster relative to **dFU-Me**. This suggests that C5-chloro group might be interacting with the active site residues positioning the nucleobase to adopt an optimal orientation for the removal of the O^4-methyl group. It is assumed that the bigger size of **dBrU-Me** presented more of a challenge to the AGTs to repair the O^4-methyl group. All the AGT proteins displayed similar repair proficiency towards **dBrU-Me** with hAGT and Ada-C achieving 45% and 55% repair, respectively.

Fluorine has inductive (σ-withdrawing electrons) and conjugation (π-donating electrons) effects that can alter the heterocycle contributing towards making the O^4-C_α bond more susceptible for the removal of lesions by AGT. These effects are weaker for chlorine followed by bromine. In similar systems, such as in fluorotoluenes or fluorobenzenes, it has been observed that an ortho-fluoro group or β-fluorination with respect to amino, hydroxy or carboxyl groups increases their acidity mostly due to the inductive effect. The ortho-fluoro position with respect to these groups also tends to favor nucleophilic reactions (stabilized through the inductive effect).[56,62]

Studies have shown that fluorine, chlorine and bromine are able to engage in interactions with different residues. These halogens can be involved in hydrophobic interactions, which has been shown to occur with benzyl moieties.[35,59,63] Halogens can also decrease the pKa of neighbouring groups.[59,63] It may be that the fluorine enhances specific interactions in the AGT active site leading to a pKa shift by engaging interactions with certain active site residues such as Cys145 and His146. Short contacts between fluorine and the thiol

from Cys145 and the imidazole ring His146 could enhance the formation of the nucleophile required for alkyl transfer resulting in the enhanced efficiency in repair.[61,74,116] Other contributions may arise from fluorophilic moieties such as hydroxyl (present in Tyr114 and Tyr158) and guanidinium groups (found in Arg135) (Figure 5.2).[61,116]

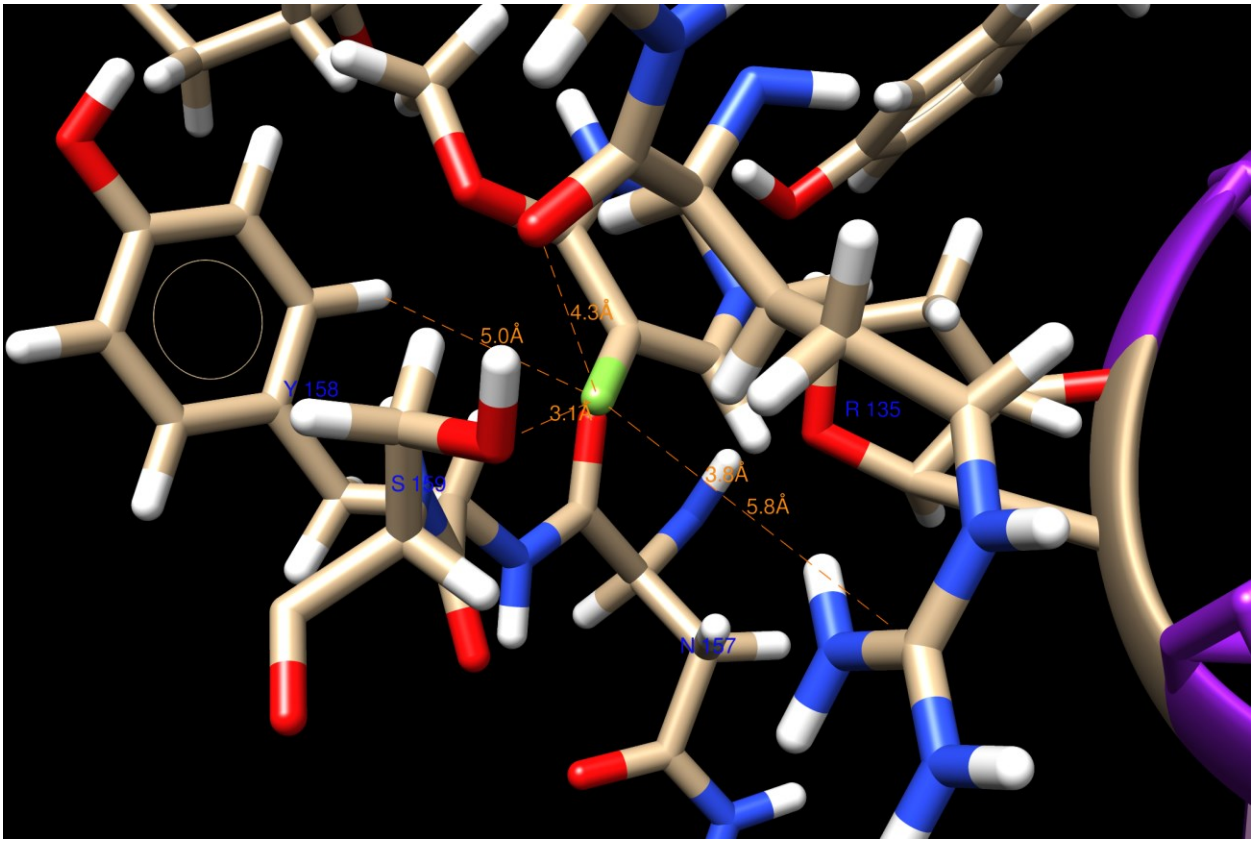


Figure 5.2: Possible interactions of C5-Fluoro (green atom) with residues in hAGT active site. (Modified from PDB 1T38 using UCSF-Chimera)

Halogens can participate in hydrogen bonds but it is not that common and optimal conditions and orientation are required for this to occur. Any interaction with fluorine or other halogens depends on the distance and position relative to the interacting atoms.[61,115] Chloride and bromide's ability to interact via halogen bonds with residues such as tyrosine, tryptophan, arginine and serine might promote a more optimal accommodation and positioning of the lesion for removal.

For example in human AGT, the C5-chloro group could be interacting with Tyr158 and Ser159 possible through halogen or hydrogen bonds.[115] Chlorine might also achieve similar interactions with Trp161 and Tyr159 (located on Ada-C's active site).[115] Bromine has shown to have an inclination to interact with arginine.[115] Therefore it is possible that apart from interacting with the previous mentioned residues, this halogen could be interacting with Arg135 (in hAGT), whose aliphatic region of its side chain is in proximity to the C5-methyl group in T, despite some steric issues.[34] Perhaps the underlying cause of why **dBrU-Me** is repaired by Ada-C in spite of having a small active pocket could be a possible interaction with Arg160.[102,115]

A summary and comparison of the repair of all the different substrates with hAGT, OGT, Ada-C and hOGT are shown in Tables 5.1.1, 5.1.2 and 5.1.3. There still exist multiple questions for future studies. First, is there space available in the active site of the AGTs for these modified substrates? Second, how does the inductive and resonance effects of fluorine (or other halogens) make the α carbon of the lesion more prone to the nucleophilic attack by the AGT? Finally, once the substrate can fit in the active site, is fluorine able to interact with the active site residues and (1) generate certain favorable interactions that allow the nucleophilic attack to occur faster? and/or (2) generate interactions that might orient and accommodate the lesion better for removal? Many of these effects also depend on the environment. Crystal structures of these halogenated modified DNA-AGT complexes would contribute to understanding these interactions. It is evident that the C5-fluoro and (to a lesser extent) C5-chloro and C5-bromo groups of modified 2'-deoxyuridine have an influence on the repair of O^4-alkylated adducts by the AGT variants.

Table 5.1 Relative repairs of different T substrates by different AGTs

AGTs	T-Me	T-Et	T-Bn	T-4COH	T-7COH
hAGT	+	-	+++	-	-
OGT	+++++	++++	++++	+++	++
Ada-C	+++++	+++	-	-	-
hOGT	+++	-	++++	-	-

Table 5.2 Relative repairs of different dU substrates by different AGTs

AGTs	dU-Me	dU-Et	dU-Bn	dU-4COH	dU-7COH
hAGT	+++	+	+++++	-	-
OGT	+++++	++++	+++++	+++++	++++
Ada-C	+++++	++++	-	-	-
hOGT	+++++	+++	+++++	+++/++	++++

Table 5.3 Relative repairs of different dFU, dClU and dBrU substrates by different AGTs

AGTs	dFU-Me	dFU-Et	dFU-Bn	dFU-4COH	dFU-7COH	dClU-Me	dBrU-Me
hAGT	+++++	+++	+++++	++	+++/++	+++++	++++
OGT	+++++	+++++	+++++	+++++/++++	+++++	+++++	+++++/++++
Ada-C	+++++	+++++	+++++	++	-	+++++	+++++/++++
hOGT	+++++	+++++	+++++	+++++/++++	+++++	+++++/++++	+++++/++++

+++++ (25% repair observed in < 15sec)

++++ (25% repair observed between 15sec to1 min)

+++ (25% repair observed between 1 min to 15 min)

++ (25% repair observed between 15 min to 30 min)

+ (25% repair observed in more than 30 min)

- (no repair observed)

*Columns shaded in light orange correspond to the analyses I performed

*Columns that are not shaded correspond to the analyses performed by Dr. F. McManus (PhD Thesis 2014)

5.2 Future Directions

An interesting extension of this study would be to prepare and evaluate repair of 5-iodo-O^4-alkyl-2'-deoxyuridine by the various AGTs. Iodine has a van der Waal radius of 1.95 Å and these results would contribute to our understanding of the steric and electronic effects of the C5-group on AGT repair of O^4-alkyl groups. 5-Iodouracil is a compound that is used as a therapeutic agent that enhances radiosensitivity.[65]

Another interesting modification to investigate at the C5-position would be the trifluoromethyl (CF_3) group. CF_3 has been used to replace methyl groups in residues such as in lysine. Studies have shown that it might have the opposite conjugation effect relative to fluorine. For instance in a benzene ring system, CF_3 is mainly an electron withdrawing substituent with a negative conjugation effect and a dominant inductive effect.[64] Moreover from a structural perspective the CF_3 group corresponds in size to an isopropyl group.[59]

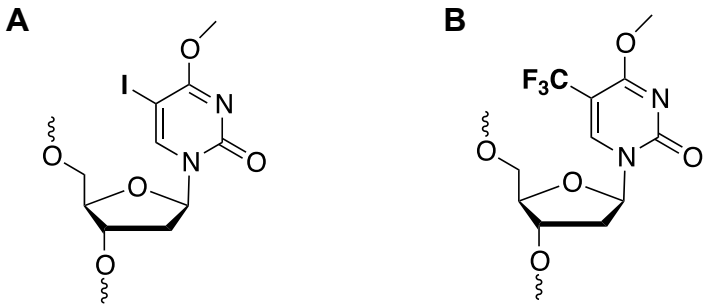


Figure 5.3: Structure of (**A**) 5-iodo-O^4-methyl-2'-deoxyuridine and (**B**) 5-trifluoromethyl-O^4-methyl-2'-deoxyuridine.

The introduction of fluorine at the nucleobase of intra- and inter-strand crosslinks could be of interest to observe whether enhanced repair is observed for these bulky AGT substrates. Moreover, the enhaced reactivity observed with the C5-fluoro pyrimidine adducts could be exploited to direct the AGT reaction to produce well defined AGT-DNA cross-linked species.

Finally, it may be interesting to explore sugar-modified nucleic acids on AGT repair. One example of an analog are the 2'-deoxy-2'-fluoroarabinoses (2'F-ANA) which have shown

promise as gene silencing agents and have a global structure similar to DNA. O^4-Alkylated analogues of 2'F-ANA that contain modifications at the C5-position, such as 2'-fluoro-5-iodo-1-β-D-arabinofuranosyluracil (FIAU) and 2'-fluoro-5-fluoro-1-β-D-arabinofuranosyluracil (FFAU), could be interesting to investigate as AGT substrates. Moreover, the use of these backbone modified and base alkylated oligonucleotides may be exploited as a combined hAGT inhibitor and gene silencing agent.[72,131]

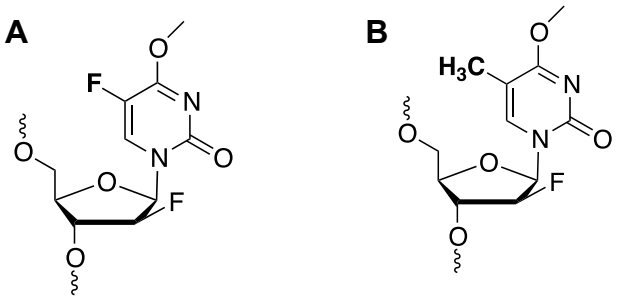


Figure 5.4: Structures of (**A**) FFAU and (**B**) FMAU

Chapter 6. References

[1] N. Kondo, A. Takahashi, K. Ono, T. Ohnishi, *Journal of Nucleic Acids* **2010**, *2010*, 1–7.

[2] Y. Mishina, E. M. Duguid, C. He, *Chem. Rev.* **2008**, *106*, 215–232.

[3] M. Bouziane, F. Miao, N. Ye, G. Holmquist, G. Chyzak, T. R. O'Connor, *Acta Biochimica Polonica* **1998**, *45*, 192–202.

[4] Y. Mishina, C. He, *Chem. Rev.* **2006**, *100*, 670–678.

[5] G. Sun, S. and Jin, R. Baskaran, *Exp. Cell Res.* **2010**, *315*, 3163–3175.

[6] K. V. Shooter, R. Howse, S. A. Shah, P. D. Lawley, *Biochem. J* **1974**, *137*, 303–312.

[7] D. K. O'Flaherty, C. J. Wilds, *Org. Biomol. Chem* **2016**.

[8] B. Rydberg, T. Lindahl, *The EMBO Journal* **1982**, *1*, 211–216.

[9] S. S. Hecht, *Chemical Research in Toxicology* **1998**, *11*, 561–590.

[10] T. Iwakuma, K. Sakumi, Y. Nakatsuru, H. Kawate, H. Igarashi, A. Shiraishi, T. Tsuzuki, T. Ishikawa, M. Sekiguchi, *Carcinogenesis* **1997**, *18*, 1631–1635.

[11] D. K. O. Flaherty, C. J. Wilds, *Chem. Asian Journal* **2016**, 576–583.

[12] D. K. O. Flaherty, C. J. Wilds, *Chem. Eur. J.* **2015**, *21*, 10522–10529.

[13] L. Samson, S. Han, J. C. Marquis, *Carcinogenesis* **1997**, *18*, 919–924.

[14] B. D. Preston, B. Singert, L. A. Loeb, *Biochemistry* **1986**, *83*, 8501–8505.

[15] E. L. Loechler, C. L. Green, J. M. Essigmannt, *Biochemistry* **1984**, *81*, 6271–6275.

[16] D. T. Beranek, *Mutation Research/Fundamental and Molecular Mechanisms of*

Mutagenesis **1990**, *231*, 11–30.

[17] M. D. Wyatt, J. M. Allan, A. Y. Lau, T. E. Ellenberger, L. D. Samson, *BioEssays* **1999**, *21*, 668–676.

[18] R. Goth-goldstein, *Cancer Research* **1980**, *40*, 2623–2624.

[19] B. Kaina, A. A. van Zeeland, A. de Groot, A. T. Natarajan, *Mutation Research Letters* **1990**, *243*, 219–224.

[20] G. T. Pauly, R. C. Moschel, *Chem. Res. Toxicol.* **2001**, *14*, 894–900.

[21] P. Wang, N. J. Amato, Q. Zhai, Y. Wang, *Nucleic Acids Research* **2015**, *43*, 10795–10803.

[22] M. J. Hickman, L. D. Samson, *Mol. Cell* **2004**, *14* , 105–116.

[23] P. Zak, K. Kleibhll, F. Laval, *The Journal of Biological Chemistry* **1994**, *269*, 730–733.

[24] A. Sancar, *Annu. Rev. Biochem* **1996**, *65*, 43–81.

[25] K. Theis, M. Skorvaga, M. Machius, N. Nakagawa, B. Van Houten, C. Kisker, *Mutation Research* **2000**, *460*, 277–300.

[26] G. P. Margison, M. F. Santiba, *BioEssays* **2002**, *24*, 255–266.

[27] Y. Mishina, C. J. Lee, C. He, *Nucleic Acids Research* **2004**, *32*, 1548–1554.

[28] J. L. Tubbs, J. A. Tainer, *Encyclopedia of Biological Chemistry* **2013**, *2*, 9–15.

[29] A. E. Pegg, *Chemical Research in Toxicology* **2011**, 618–639.

[30] S. M. O. Toole, A. E. Pegg, J. A. Swenberg, *Cancer Research* **1993**, *53*, 3895–3898.

[31] J. E. A. Wibley, A. E. Pegg, P. C. E. Moody, *Nucleic Acids Research* **2000**, *28*, 393–401.

[32] C. Yi, C. He, *Cold Spring Harbor Perspectives in Biology* **2013**, 1–18.

[33] D. S. Daniels, C. D. Mol, A. S. Arvai, S. Kanugula, A. E. Pegg, J. A. Tainer, *The EMBO Journal* **2000**, *19*, 1719–1730.

[34] E. M. Duguid, P. A. Rice, C. He, *Journal of Molecular Biology* **2005**, *350*, 657–666.

[35] J. L. Tubbs, A. E. Pegg, J. A. Tainer, *DNA Repair* **2007**, 1100–1115.

[36] K. S. Srivenugopal, X. Yuan, H. S. Friedman, F. Ali-osman, *Biochemistry* **1996**, *35*, 1328–1334.

[37] A. E. Pegg, *Mutation Research* **2000**, 83–100.

[38] Q. Fang, S. Kanugula, A. E. Pegg, *Biochemistry* **2005**, *44*, 15396–15405.

[39] G. W. Rebeck, C. M. Smith, D. L. Goad, L. Samson, *J. Bacteriol.* **1989**, *171*, 4563–4568.

[40] M. H. Moore, J. M. Gulbis, E. J. Dodson, B. Demple, P. C. E. Moody, *The EMBO Journal* **1994**, *13*, 1495–1501.

[41] P. M. Potter, M. C. Wilkinson, J. Fitton, F. J. Carr, J. Brennad, D. P. and Cooper, G. P. Margison, *Nucleic Acids Research* **1987**, *15*, 9177–9193.

[42] S. Mitra, B. C. Pal, R. S. Foote, B. Division, O. Ridge, O. Ridge, *Journal of Bacteriology* **1982**, *152*, 534–537.

[43] N. Shrivastav, D. Li, J. M. Essigmann, *Carcinogenesis* **2010**, 59–70.

[44] R. H. Elder, G. P. Margison, J. A. Rafferty, *Biochem. J* **1994**, *235*, 231–235.

[45] F. P. Mcmanus, C. J. Wilds, *Toxicol. Res* **2013**, 158–162.

[46] Q. Fang, S. Kanugula, J. L. Tubbs, J. A. Tainer, A. E. Pegg, *Journal of Biological Chemistry* **2010**, DOI 10.1074/jbc.M109.045518.

[47] D. S. Daniels, T. T. Woo, K. X. Luu, D. M. Noll, N. D. Clarke, A. E. Pegg, J. A. Tainer, *Nat Struct Mol Biol* **2004**, *11*, 714–720.

[48] A. E. Pegg, M. E. Dolan, R. C. Moschel, in *Progress in Nucleic Acid Research and Molecular Biology* (Ed.: W.E.C. and K.M.B.T.-P. in N.A.R. and M. Biology), Academic Press, **1995**, pp. 167–223.

[49] D. S. Daniels, J. A. Tainer, *Mutat. Res.* **2000**, *460* , 151–163.

[50] F. P. Guengerich, Q. Fang, L. Liu, D. L. Hachey, A. E. Pegg, *Biochemistry* **2003**, *42* , 10965–10970.

[51] R. Coulter, M. Blandino, J. M. Tomlinson, G. T. Pauly, M. Krajewska, R. C. Moschel, L. A. Peterson, A. E. Pegg, T. E. Spratt, *Chem. Res. Toxicol.* **2007**, *20*, 1966–1971.

[52] K. Goodtzova, S. Kanugula, S. Edara, G. T. Pauly, R. C. Moschel, A. E. Pegg, *The Journal of Biological Chemistry* **1997**, *272*, 8332–8339.

[53] Q. Fang, A. M. Noronha, S. P. Murphy, C. J. Wilds, J. L. Tubbs, J. A. Tainer, G. Chowdhury, F. P. Guengerich, A. E. Pegg, *Biochemistry* **2008**.

[54] F. P. Mcmanus, C. J. Wilds, *ChemBioChem* **2014**, *15*, 1966–1977.

[55] B. C. Buer, E. N. G. Marsh, *Protein Science* **2012**, *21*, 453–462.

[56] B. E. Smart, *Journal of Fluorine Chemistry* **2001**, *109*, 3–11.

[57] A. Bondi, *The Journal of Physical Chemistry* **1965**, *68*, 441–451.

[58] K. L. Kirk, *Journal of Fluorine Chemistry* **2006**, *127*, 1013–1029.

[59] D. O. Hagan, H. S. Rzepa, *ChemComm* **1997**, 645–652.

[60] J. Swinson, *Halocarbon Products Corporation* **2005**, 1–5.

[61] P. Zhou, J. Zou, F. Tian, Z. Shang, *J. Chem. Inf. Model* **2009**, *49*, 2344–2355.

[62] V. M. Vlasov, G. G. Yakobson, *Russian Chemical Reviews* **1974**, *43*, 1642–1668.

[63] H. Bohm, D. Banner, S. Bendels, M. Kansy, B. Kuhn, M. Klaus, U. Obst-sander, M. Stahl, *ChemBioChem* **2004**, *5*, 637–643.

[64] L. M. Yagupol'skii, A. Y. Ll'chenko, N. V. Kondratenko, *Russian Chemical Reviews* **1974**, *43*, 64–94.

[65] E. Heshmati, P. Abdolmaleki, H. Mozdarani, A. Sabet, *Bioorganic & Medicinal Chemistry Letters* **2009**, *19*, 5256–5260.

[66] E. B. Lamont, R. L. Schilsky, *Clinical Cancer Research* **1999**, *5*, 2289–2296.

[67] R. Malet-Martino, R. Martino, *The Oncologist* **2002**, 288–323.

[68] D. G. Haller, J. Cassidy, S. J. Clarke, D. Cunningham, E. Van Cutsem, P. M. Hoff, M. L. Rothenberg, L. B. Saltz, H. Schmoll, C. Allegra, et al., *Journal of Clinical Oncology* **2008**, *26*, 2118–2123.

[69] D. Baraniak, D. Baranowski, P. Ruszkowski, J. Boryski, D. Baraniak, D. Baranowski, P. Ruszkowski, J. Boryski, *Nucleosides, Nucleotides and Nucleic Acids* **2016**, *35*, 178–194.

[70] J. Voorde, S. Liekens, C. Mcguigan, P. G. S. Murziani, M. Slusarczyk, J. Balzarini, *Biochemical Pharmacology* **2011**, *82*, 441–452.

[71] G. R. Rettig, M. A. Behlke, *Molecular Therapy* **2009**, *20*, 483–512.

[72] T. Dowler, D. Bergeron, A. Tedeschi, L. Paquet, N. Ferrari, M. J. Damha, *Nucleic Acids Research* **2006**, *34*, 1669–1675.

[73] A. Kalota, L. Karabon, C. R. Swider, E. Viazovkina, M. Elzagheid, M. J. Damha, A. M. Gewirtz, *Nucleic Acids Res.* **2006**, *34*, 451–461.

[74] E. Carosati, S. Sciabola, G. Cruciani, *J. Med. Chem.* **2004**, 5114–5125.

[75] A. Moser, R. Guza, D. M. York, *Theor Chem Acc.* **2009**, *122*, 179–188.

[76] S. S. Hecht, *Toxicology* **1998**, *11*, 560–603.

[77] P. D. Lawley, D. J. Orr, S. A. Shah, P. B. Farmer, M. Jarman, *Biochem. J.* **1973**, *135*, 193–201.

[78] A. E. Pegg, *Chem. Res. Toxicol.* **2011**, *24*, 618–39.

[79] A. K. Basu, J. M. Essigmann, *Chem. Res. Toxicol.* **1988**, *1*, 1–13.

[80] D. J. Kalnik, M. W., Kouchakdjian, M., Li, B., Swann, P. F., and Patel, *Biochemistry* **1988**, 108–115.

[81] T. P. Brent, M. E. Dolan, R. Montesano, A. E. Pegg, P. M. Potter, B. Singer, J. A. Swenberg, D. B. Yarosw, *Biochem. J* **1988**, *85*, 1759–1762.

[82] K. B. Altshuler, C. S. Hodes, J. M. Essigmann, *Chem. Res. Toxicol.* **1996**, *9*, 980–987.

[83] G. T. Pauly, S. H. Hughes, R. C. Moschel, *Carcinogenesis* **1998**, *19*, 457–461.

[84] S. R. Paalman, C. Sung, N. D. Clarke, *Biochemistry* **1997**, *36*, 11118–11124.

[85] H. Takinowaki, Y. Matsuda, T. Yoshida, Y. Kobayashi, T. Ohkubo, *Protein Science* **2006**, *15*, 487–497.

[86] G. W. Rebeck, S. Coons, P. Carroll, L. Samson, *Genetics* **1988**, 3039–3043.

[87] F. P. McManus, A. Khaira, A. M. Noronha, C. J. Wilds, *Bioconjugate Chem.* **2013**, *24*, 224–233.

[88] F. P. Mcmanus, D. K. O 'flaherty, A. M. Noronha, C. J. Wilds, *Org. Biomol. Chem* **2012**, *10*, 7078–7090.

[89] F. P. Mcmanus, Q. Fang, J. D. M. Booth, A. M. Noronha, A. E. Pegg, C. J. Wilds, *Org. Biomol. Chem* **2010**, *8*, 4414–4426.

[90] W. J. Bodell, T. Aida, M. S. Berger, M. L. Rosenblum, *Environmental Health Perspectives*

1985, *62*, 119–126.

[91] Y. Cai, M. H. Wu, M. Xu-welliver, A. E. Pegg, S. M. Ludeman, M. E. Dolan, *Cancer Research* **2000**, *60*, 5464–5469.

[92] R. Loeber, E. Michaelson, Q. Fang, C. Campbell, A. Pegg, N. and Tretyakova, *Chem. Res. Toxicol.* **2010**, *21*, 612–626.

[93] E. M. Duguid, P. A. Rice, C. He, *Journal of Molecular Biology* **2005**, *350*, 657–666.

[94] P. Shah, A. D. Westwell, *Journal of Enzyme Inhibition and Medicinal Chemistry* **2016**, *6366*, 527–540.

[95] J. Wang, J. Luis, C. Pozo, A. E. Sorochinsky, S. Fustero, V. A. Soloshonok, H. Liu, *Chem. Rev.* **2014**, *114*, 2432–2506.

[96] L. Heidelberger, C., Chaudruri, N.K., Dannerberg, P., Mooren, D. and Griesbach, *Nature* **1957**, *4561*, 663–666.

[97] Y. Xu, P. F. Swann, *Nucleic Acids Research* **1990**, *18*, 4061–4066.

[98] R. Stolarski, W. Egan, T. L. James, *Biochemistry* **1992**, *31*, 7027–7042.

[99] Q. Zhu, M. O. Delaney, M. M. Greenberg, *Bioorganic and Medicinal Chemsitry Letters* **2001**, *11*, 1105–1107.

[100] A. E. Pegg, *Chem. Res. Toxicol.* **2011**, 618–639.

[101] N. Shrivastav, D. Li, J. M. Essigmann, *Carcinogenesis* **2010**, *31*, 59–70.

[102] J. E. Wibley, A. E. Pegg, P. C. Moody, *Nucleic Acids Res.* **2000**, *28* , 393–401.

[103] M. Sriram, G. A. van der Marel, H. Roelen, J. van Boom, A. Wang, *Biochemistry* **1992**, *31*, 11823–11834.

[104] D. K. O'Flaherty, C. J. Wilds, *Chem. Eur. J.* **2015**, *21*, 10522–10529.

[105] D. K. O. Flaherty, F. P. Mcmanus, A. M. Noronha, C. J. Wilds, *Current Protocols in Nucleic Acid Chemistry* **2013**, 5.13.1–5.13.19.

[106] M. Sassanfar, M. K. Dosanjhllii, J. M. Essigmannv, L. Samson, *The Journal of Biological Chemistry* **1991**, *266*, 2767–2771.

[107] A. E. Pegg, M. Boosalis, L. Samson, R. C. Mosche, T. L. Byers, K. Swenn, M. E. Dolanl, *Biochemistry* **1993**, *32*, 11998–12006.

[108] P. F. S. and G. P. M. M.C. Wilkinson, P. M. Potter, L. Cawkwell, P. Georgiadis, D. Patel, *Nucleic Acids Research* **1989**, *17*, 8475–8484.

[109] B. Sedgwicks, P. Robins, N. Tottyqll, T. Lindahl, *Journal of Biological Chemistry* **1988**, *263*, 4430–4433.

[110] R. J. Graves, B. F. L. Li, P. F. Swann, *Carcinogenesis* **1989**, *10*, 661–666.

[111] V. Valinluck, W. Wu, P. Liu, J. W. Neidigh, L. C. Sowers, *Chem. Res. Toxicol.* **2006**, *19*, 556–562.

[112] B. Puffer, C. Kreutz, U. Rieder, M. Ebert, R. Konrat, R. Micura, *Nucleic Acids Research* **2009**, *37*, 7728–7740.

[113] J. Graton, Z. Wang, A. Brossard, D. G. Monteiro, Y. Le Questel, B. Linclau, *Angewandte Communications* **2012**, 6176–6180.

[114] N. Martín-Pintado, M. Yahyaee-Anzahaee, G. F. Deleavey, G. Portella, M. Orozco, M. J. Damha, C. González, *Journal of the American Chemical Society* **2013**, *135*, 5344–5347.

[115] S. Kortagere, S. Ekins, W. J. Welsh, *Journal of Molecular Graphics and Modelling* **2008**, *27*, 170–177.

[116] P. Zhou, J. Lv, J. Zou, F. Tian, Z. Shang, *Journal of Structural Biology* **2010**, *169*, 172–182.

[117] J. P. Henderson, J. Byun, J. W. Heinecke, *J. Biol. Chem.* **1999**, *274*, 33440–33448.

[118] J. P. Henderson, J. Byun, M. V. Williams, D. M. Mueller, M. L. McCormick, J. W. Heinecke, *Journal of Biological Chemistry* **2001**, *276*, 7867–7875.

[119] Q. Jiang, B. C. Blount, B. N. Ames, *Journal of Biological Chemistry* **2003**, *278*, 32834–32840.

[120] R. Ivarie, *Nucleic Acids Research* **1987**, *15*, 9975–9983.

[121] S. Sirimulla, J. B. Bailey, R. Vegesna, M. Narayan, *Journal of Chemical Information and Modeling* **2013**, *53*, 2781–2791.

[122] P. Bolognesi, P. O. Keeffe, Y. Ovcharenko, M. Coreno, L. Avaldi, V. Feyer, O. Plekan, K. C. Prince, W. Zhang, V. Carravetta, et al., *The Journal of Chemical Physics* **2014**, *133*.

[123] L. Storchi, F. Tarantelli, S. Veronesi, P. Bolognesi, E. Fainelli, L. Avaldi, *The Journal of Biological Chemistry* **2008**, *129*.

[124] L. C. Sowers, *Journal of Biomolecular Structure and Dynamics* **2000**, *17*, 713–723.

[125] L. C. Sowers, B. R. Shaw, W. D. Sedwick, *Biochemical and Biophysical Research Communications* **1987**, *148*, 790–794.

[126] S. Z. Borozan, S. D. Stojanović, *Computational Biology and Chemistry* **2013**, *47*, 231–239.

[127] M. Fourmigué, *Current Opinion in Solid State & Materials Science* **2009**, *13*, 36–45.

[128] T. Steiner, È. Kristallographie, F. Universita, È. Berlin, D.- Berlin, *Acta Crystallographica* **1998**, 456–463.

[129] P. A. Hart, J. P. Davis, *Journal of the American Chemical Society* **1972**, *139*, 2572–2577.

[130] J. Poznański, D. Shugar, *Biochimica et biophysica acta* **2013**, *1834*, 1381–6.

[131] H. Cai, Z. Li, P. S. Conti, *Nuclear Medicine and Biology* **2011**, *38*, 659–666.

[132] D. M. Noll, a M. Noronha, P. S. Miller, *J. Am. Chem. Soc.* **2001**, *123*, 3405–3411.

[133] J. D. Puglisi, I. J. Tinoco, *Methods Enzymol.* **1989**, *180*, 304–325.

[134] S. Cohen, M. R. Loeb, J. Lichtenstein, *Biochemistry* **1958**, *44*, 1004–1012.

[135] D. Baraniak, D. Baranowski, P. Ruszkowski, J. Boryski, D. Baraniak, D. Baranowski, P. Ruszkowski, J. Boryski, *Nucleosides, Nucleotides and Nucleic Acids* **2016**, *35*, 178–194.

[136] H. Zheng, M. Xiao, Q. Yan, Y. Ma, S. Xiao, *J. Am. Chem. Soc.* **2014**, 10194–10197.

[137] V. P. Antao, I. Tinoco, *Nucleic Acids Res.* **1991**, *19*, 5901–5905.

[138] P. C. Bevilacqua, J. M. Blose, *Annu. Rev. Phys. Chem.* **2008**, *59*, 79–103.

[139] J. M. Blose, K. P. Lloyd, P. C. Bevilacqua, *Biochemistry* **2009**, *48*, 8787–8794.

[140] B. F. Eichman, J. M. Vargason, B. H. Mooers, P. S. Ho, *Proceedings of the National Academy of Sciences of the United States of America* **2000**, *97*, 3971–3976.

Appendix I: Supporting Information for Chapter 2

Contents **Page**

Supporting Methods

A1.1 Synthesis and characterization of dFU nucleosides and oligonucleotides **81**
A1.2 UV thermal denaturation **85**
A1.3 Circular dichroism (CD) spectroscopy **86**
A1.4 AGT repair assay of dFU-Me DNA duplexes **86**
A1.5 Molecular modelling **87**
A1.6 Geometry Optimization **88**

Supporting Figures

Figure A1.1 - Different adducts and substrates mentioned in Chapter 2 **88**
Figure A1.2 - 500 MHz ^{1}H NMR spectrum of compound (**3**) (in d_6-acetone) **89**
Figure A1.3 - 125.7 MHz ^{13}C NMR spectrum of compound (**3**) (in d_6-acetone) **90**
Figure A1.4 - 470.4 MHz ^{19}F NMR spectrum of compound (**3**) (in d_6-acetone) **91**
Figure A1.5 - 202.3 MHz ^{31}P NMR spectrum of compound (**3**) (in d_6-acetone) **92**
Figure A1.6 - 500 MHz ^{1}H NMR spectrum of compound (**5**) (in $CDCl_3$) **93**
Figure A1.7 - 125.7 MHz ^{13}C NMR spectrum of compound (**5**) (in $CDCl_3$) **94**
Figure A1.8 - 470.4 MHz ^{19}F NMR spectrum of compound (**5**) (in $CDCl_3$) **95**
Figure A1.9 - 500 MHz ^{1}H NMR spectrum of compound (**6**) (in d_6-acetone) **96**
Figure A1.10 - 125.7 MHz ^{13}C NMR spectrum of compound (**6**) (in d_6-acetone) **97**
Figure A1.11 - 470.4 MHz ^{19}F NMR spectrum of compound (**6**) (in d_6-acetone) **98**
Figure A1.12 - 202.3 MHz ^{31}P NMR spectrum of compound (**6**) (in d_6-acetone) **99**
Figure A1.13 - ESI MS spectrum of oligonucleotide **dFU** **100**
Figure A1.14 - ESI MS spectrum of oligonucleotide **dFU-Me** **101**
Figure A1.15 - C-18 RP-HPLC profile of nuclease digested **dFU-Me** **102**
Figure A1.16 - SAX-HPLC profile of crude and pure **dFU-Me** **103**
Figure A1.17 - UV thermal denaturation experiments for unmodified DNA duplex (T), **dFU**, and **dFU-Me** **103**
Figure A1.18 - Circular dichroism spectra for unmodified DNA duplex (T), **dFU** and **dFU-Me** **104**
Figure A1.19 - Time course repair gel of duplexes containing **dFU-Me** by (**A**) hAGT, (**B**) OGT, (**C**) Ada-C and (**D**) hOGT **105**
Figure A1.20 - Molecular models of unmodified control duplex (**T**) and duplexes containing **dFU** and **dFU-Me** that were geometry optimized using the AMBER forcefield **106**
Figure A1.21 - Geometry optimization of unmodified control duplex (**T**) and duplexes containing **dFU** and **dFU-Me** using Gaussian 09 and **107**

B3LYP/6-31+G(2d,2p) mode

A1.1 Synthesis and characterization of nucleosides and oligonucleotides

A1.1.1 General

5-fluoro-2'-deoxyuridine (compound **1**) was purchased from Berry Associates (Dexter, Michigan). "Fast deprotecting" 5'-*O*-Dimethoxytrityl-2'-deoxyribonucleoside-3'-*O*-(β-cyanoethyl-N,N-diisopropyl)phosphoramidites and protected 2'-deoxyribonucleoside–CPG supports were purchased from Glen Research (Sterling, Virginia). Compounds **2** and **4** were prepared according to previously published procedures.[54,97,98] All other chemicals and solvents were purchased from the Aldrich Chemical Company (Milwaukee, WI) or EMD Chemicals Inc. (Gibbstown, NJ). Flash column chromatography was performed using silica gel 60 (230–400 mesh) purchased from Silicycle (Quebec City, QC). Thin layer chromatography (TLC) was carried out with precoated TLC plates (Merck, Kieselgel 60 F_{254}, 0.25 mm) purchased from EMD Chemicals Inc. (Gibbstown, NJ). NMR spectra were recorded on a Varian 500 MHz NMR spectrometer at room temperature. ^{1}H NMR spectra were recorded at a frequency of 500.0 MHz and chemical shifts were reported in parts per million (ppm) downfield from tetramethylsilane. ^{13}C NMR spectra (^{1}H decoupled) were recorded at a frequency of 125.7 MHz and chemical shifts were reported in ppm with tetramethylsilane as a reference. ^{19}F NMR spectra were recorded at a frequency of 470.4 MHz and chemical shifts were reported in ppm with trichlorofluoromethane as a reference. ^{31}P NMR spectra (^{1}H decoupled) were recorded at a frequency of 202.3 MHz and chemical shifts were reported in ppm with H_3PO_4 used as an external standard. High resolution mass spectrometry of modified nucleosides were obtained using an 7T-LTQ FT ICR instrument (Thermo Scientific), at the Concordia University Centre for Structural and Functional Genomics. The mass spectrometer was operated in full scan, positive ion detection mode. ESI mass spectra for oligonucleotides were obtained at the Concordia University Centre for Biological Applications of Mass Spectrometry (CBAMS) using a Micromass Qtof2 mass spectrometer (Waters) equipped with a nanospray ion source. The mass spectrometer was operated in full scan, negative ion detection mode. T4 polynucleotide kinase

(PNK) was obtained from New England BioLabs (NEB). [γ-^{32}P]ATP was purchased from PerkinElmer (Woodbridge, ON). The *BclI* restriction enzyme was purchased from New England Biolabs (Ipswich, MA).

3'-*O*-(2-Cyanoethoxy(diisopropylamino)-phosphino)-5'-*O*-(4,4'-dimethoxytrityl)- 5-fluoro-2'-deoxyuridine (3) DIPEA (0.19 mL, 1.12 mmol) was added to a solution of 5'-*O*-(4,4'-dimethoxytrityl)-5-fluoro-2'-deoxyuridine (0.204 g, 0.372 mmol) in THF (3.7 mL), followed by the dropwise addition of N,N-diisopropylamino cyanoethyl phosphonamidic chloride (0.2 mL, 0.893 mmol). After 30 min, the solvent was evaporated *in vacuo*, the crude product was taken up in ethyl acetate (40 mL), the solution was washed with 3% (w/v) solution of $NaHCO_3$ (2 x 50mL) and once with brine (50 mL). The organic layer was dried over anhydrous Na_2SO_4 (~ 4 g), and concentrated *in vacuo*. The crude product was purified by flash column chromatography using a hexanes/ethyl acetate (2:8) solvent system to afford 0.187 g (67%) of a colorless foam. R_f (SiO_2 TLC): 0.87 (100% EtOAc). $\lambda_{max(MeCN)}$ = 268 nm. HRMS (ESI-MS) *m/z* calculated for $C_{39}H_{47}FN_4O_8P^+$: 749.3110; found 749.3104 $[M + H]^+$. ^{1}H NMR (500 MHz, d_6-acetone, ppm): 7.93 (d, J= 6.3 Hz, 0.5H; H6), 7.91 (d, J= 6.3 Hz, 0.5H; H6), 7.23–7.53 (m, 9H; Ar), 6.90-6.93 (m, 4H; Ar), 6.28-6.32 (m, 1H; H1'), 4.71-4.78 (m, 1H; H3'), 4.17-4.24 (m, 1H; H4'), 3.59-3.93 (m, 10H; NCH, $ArOCH_3$, CH_2OP), 3.39-3.50 (m, 2H; H5', H5''), 2.77-2.79 (t, J= 6 Hz, 1H; CH_2CN), 2.65-2.67 (t, J= 6 Hz, 1H; CH_2CN), 2.47-2.52 (m, 1H; H2'), 2.06-2.07 (m, 1H; H2''), 1.20-1.22 (m, 10H; CH_3), 1.13-1.14 (m, 2H; CH_3). ^{13}C NMR (125.7 MHz, d6-acetone, ppm): 160.99, 158.82, 156.61, 148.65, 144.96, 139.64, 137.15, 135.65, 135.55, 130.12, 130.09, 128.07, 128.03, 127.81, 126.80, 126.77, 117.84, 113.11, 86.67, 86.61, 85.42, 85.39, 85.08, 85.01, 63.29, 63.15, 58.71, 58.56, 54.64, 51.09, 43.11, 43.01, 39.33, 24.02, 23.96, 23.93, 19.85, 19.79. ^{19}F NMR (470.4 MHz, d_6-acetone, ppm): -167.84, -167.85, -167.90, -167.91. ^{31}P NMR (202.3 MHz, d_6-acetone, ppm): 148.31, 148.34. IR (thin film); ν_{max} (cm^{-1}) = 3195, 3066, 2967, 2932, 2836, 2362, 2336, 2252, 1717, 1607, 1508, 1464, 1251, 1179, 1125, 829, 736.

5'-*O*-(4,4'-Dimethoxytrityl)-5-fluoro-*O*^{4}-methyl-2'-deoxyuridine (5): To a solution of triazole (0.28 g, 4.07 mmol) in anhydrous MeCN (6 mL) at 0^{o}C under stirring, was added dropwise $POCl_3$ (0.084 ml, 0.91 mmol) followed by triethylamine (0.543 mL, 3.89 mmol). A solution of 3'-*O*-

(*tert*-butyldimethylsilyl)-5'-O-(4,4'-dimethoxytrityl)-5-fluoro-2'-deoxyuridine (0.300 g, 0.45 mmol) in anhydrous MeCN (6 mL) was added to the triazole solution. After 40 min, an additional solution of triazole (0.28 g, 4.07 mmol), $POCl_3$ (0.084 mL, 0.91 mmol) and triethylamine (0.543 mL, 3.89 mmol) in anhydrous MeCN (3 mL) was added dropwise with stirring at 0^oC to the solution containing the nucleoside. This extra addition was repeated for a second time after 40 min. After 1 h, the solvent was evaporated *in vacuo*, the crude was taken up in MeOH (7 mL) and a solution of NaOMe (0.086 g, 1.6 mmol) in MeOH (7 mL) was added. After 30 min, an additional solution of NaOMe (0.086 g, 1.6 mmol) in MeOH (3 mL) was added to the nucleoside. After 16 h, the solvent was evaporated *in vacuo* and the crude product was taken up in CH_2Cl_2 (40 mL), washed with a 3% (w/v) solution of $NaHCO_3$ (2 x 50 ml), dried over anhydrous Na_2SO_4 (~ 4 g), and concentrated *in vacuo*. The crude product was taken up in THF (4.5 mL) and TBAF (1 M in THF) (0.543 mL, 0.54 mmol) was added drop-wise. After 30 min, the solvent was evaporated *in vacuo* and the crude product was taken up in CH_2Cl_2 (40 mL), washed with a 3% (w/v) solution of $NaHCO_3$ (2 x 50 ml), dried over anhydrous Na_2SO_4 (~ 4 g), and concentrated *in vacuo*. The crude product was purified by flash column chromatography using a gradient of CH_2Cl_2/MeOH (49.5:0.5 → 49:1) as eluent to afford 0.189 g (74%) of a colorless foam. R_f(SiO_2 TLC): 0.63 (100% EtOAc). $\lambda_{max(MeCN)}$ = 283 nm. HRMS (ESI-MS) *m/z* calculated for $C_{31}H_{31}FN_2NaO_7^+$: 585.2008; found 585.2006 $[M + Na]^+$. ^{1}H NMR (500 MHz, $CDCl_3$, ppm): 8.06 (d, J= 5.6 Hz, 1H; H6), 7.21–7.41 (m, 9H; Ar), 6.83– 6.85 (m, 4H; Ar), 6.25-6.27 (m, 1H; H1'), 4.55-4.57 (m, 1H; H3'), 4.15-4.17 (m, 1H; H4'), 4.04 (s, 3H; OCH_3) 3.79 (s, 6H; $ArOCH_3$), 3.42 (m, 2H; H5', H5''), 3.18 (s, 1H; OH), 2.72-2.76 (m, 1H; H2'), 2.23-2.28 (m, 1H; H2''). ^{13}C NMR (125.7 MHz, $CDCl_3$, ppm): 162.68, 162.58, 158.63, 153.55, 144.31, 137.68, 135.72, 135.42, 135.25, 129.98, 129.96, 127.98, 127.93, 127.68, 127.43, 127.02, 113.31, 113.29, 87.10, 87.02, 86.59, 63.11, 55.22, 55.07, 42.03. ^{19}F (470.4 MHz, $CDCl_3$, ppm): -168.60, -168.62. IR (thin film); ν_{max} (cm^{-1}) = 3423, 2362, 2335, 1652, 1635, 1 506, 1497, 1403, 1251, 1176, 1035, 828.

3'-*O*-(2-Cyanoethoxy(diisopropylamino)-phosphino)-5'-*O*-(4,4'-dimethoxytrityl)-5-fluoro-*O*^{4}-methyl-2'-deoxyuridine (6): DIPEA (0.14 mL, 0.80 mmol) was added to a solution of (**5**) (0.150 g, 0.266 mmol) in THF (2.7 mL), followed by the dropwise addition of N,N-diisopropylamino cyanoethyl phosphonamidic chloride (0.14 mL, 0.64 mmol). After 30 min, the solvent was

evaporated *in vacuo*, the crude product was taken up in ethyl acetate (40 mL), then the solution was washed with a 3% (w/v) solution of $NaHCO_3$ (2 x 50 ml) and once with brine (50 mL). The organic layer was dried over anhydrous Na_2SO_4 (~ 4 g) and concentrated *in vacuo*. The crude product was purified by flash column chromatography using a hexanes/ethyl acetate (2:8) solvent system to afford 0.173 g (86%) of a colorless foam. R_f (SiO_2 TLC): 0.11, 0.26 (1:1 Hexanes/EtOAc). $\lambda_{max(MeCN)}$ = 283 nm. HRMS (ESI-MS) *m/z* calculated for $C_{40}H_{49}FN_4O_8P^+$: 763.3267; found 749.3104 $[M + H]^+$. ^{1}H NMR (500 MHz, $CDCl_3$, ppm): 8.16 (d, J= 6.3 Hz, 0.5H; H6), 8.13 (d, J= 6.3 Hz, 0.5H; H6), 7.23–7.53 (m, 9H; Ar), 6.89– 6.93 (m, 4H; Ar), 6.18-6.23 (m, 1H; H1'), 4.70-4.78 (m, 1H; H3'), 4.22-4.29 (m, 1H; H4'), 3.97-3.98 (s, 3H; OCH_3), 3.63-3.88 (m, 10H; NCH, $ArOCH_3$, CH_2OP), 3.43-3.52 (m, 2H; H5', H5''), 2.77-2.80 (t, J= 5.9 Hz, 1H; CH_2CN), 2.60-2.70 (t, J= 5.9 Hz, 1H; CH_2CN), 2.39-2.47 (m, 1H; H2''), 2.06-2.07 (m, 1H; H2'') 1.20-1.23 (m, 9H; CH_3), 1.12-1.14 (m, 3H; CH_3). ^{13}C NMR (125.7 MHz, $CDCl_3$, ppm): 162.27, 162.17, 158.83, 152.36, 144.93, 137.15, 137.12, 135.62, 135.60, 135.51, 135.47, 135.21, 135.18, 130.11, 130.08, 128.06, 127.82, 127.81, 127.68, 127.65, 126.81, 126.79, 118.07, 117.94, 113.13, 113.11, 86.70, 86.65, 86.58, 86.51, 85.73, 85.70, 85.46, 85.41, 73.18, 73.05, 72.71, 72.58, 63.00, 62.80, 58.73, 58.67, 58.58, 58.52, 54.65, 54.63, 53.98, 43.13, 43.11, 43.03, 43.01, 40.36, 40.34, 40.17, 40.13,19.90, 19.86, 19.84, 19.81. ^{19}F (470.4 MHz, d_6-acetone, ppm): -171.63, -171.64, -171.68, -171.70. ^{31}P NMR (202.3 MHz, d_6-acetone, ppm): 148.35, 148.21. IR (thin film); ν_{max} (cm^{-1}) =3071, 2966, 2869, 2836, 2362, 2335, 2252, 1683, 1652, 1607, 1541, 1506, 1457, 1401, 1336, 1252, 1179, 1116, 1035, 977, 829, 734.

A1.1.2 Preparation, purification and characterization of the dFU oligonucleotides

The sequences containing the dFU adducts, which are shown in **Figure 2.1**, were assembled with an Applied Biosystems Model 3400 synthesizer on a 1.5 μmol scale using β-cyanoethylphosphoramidite chemistry supplied by the manufacturer with slight modifications to coupling times. The nucleoside phosphoramidites protected with "fast-deprotecting" groups were prepared in anhydrous MeCN at a concentration of 0.1 M for the 3'-*O*-

deoxyphosphoramidites. Oligomer sequence assembly was carried out as previously described.[54,97] The capping step of the assembly was carried out using phenoxyacetic anhydride/pyridine/tetrahydrofuran 1:1:8 (v/v/v; solution A) and 1-methyl-imidazole/ tetrahydrofuran 16:84 (w/v; solution B). Coupling wait times for phosphoramidites **3** and **7** were extended to 10 min (compared to 2 min for the commercially available phosphoramidites). Protecting group removal and cleavage from the solid support was carried out by treatment with 0.05M K_2CO_3 in MeOH for 4 h at room temperature with mild rocking in 2 mL screw-cap microfuge tubes fitted with teflon lined caps. The base was neutralized with an equimolar amount of AcOH and crude oligomers were transferred and lyophilized in a Speedvac concentrator. Purification was achieved by strong anion exchange HPLC using a Dionex DNAPAC PA-100 column (0.4 cm x 25 cm) purchased from Dionex Corp, Sunnyvale, CA using a linear gradient of 0–52% buffer B over 24 min (buffer A: 100 mm Tris HCl, pH 7.5, 10% MeCN and buffer B: 100 mm Tris HCl, pH 7.5, 10% MeCN, 1 M NaCl) at 55 ^{o}C. The columns were monitored at 260 nm for analytical runs or 280 nm for preparative runs. The purified oligomers were desalted using C-18 SEP PAK cartridges (Waters Inc.) as previously described.[132] The molecular mass of the modified oligomers was identified by deconvolution of the ESI-MS and the measured values were in agreement with the expected masses (please see **Figures A1.13** and **A1.14** for MS spectra). **dFU-Me** (0.05 A_{260} units) was also characterized by enzymatic digestion (snake venom phosphodiesterase: 0.28 units and calf intestinal phosphatase: 5 units, in 10 mM Tris, pH 8.1 and 2 mM $MgCl_2$) for 48 h at 37 ^{o}C. The resulting nucleoside mixture was analyzed by reversed phase HPLC carried out using a Symmetry® C-18 5 µm column (0.46 x 15 cm) via a linear gradient of 0-70% buffer B over 30 min (buffer A: 50 mM sodium phosphate, pH 5.8, 2% MeCN and buffer B: 50 mM sodium phosphate, pH 5.8, 50% MeCN). Results and nucleoside ratios from digestion are shown in **Figure A1.15.**

A1.2 UV thermal denaturation

Molar extinction coefficients for the unmodified and dFU modified oligonucleotides were calculated from those of the mononucleotides and dinucleotides using the nearest-neighbor approximations (M^{-1} cm^{-1}). All duplexes were prepared by mixing equimolar amounts

of the interacting strands and lyophilizing the mixture to dryness. The resulting pellet was then dissolved in 90 mM sodium chloride, 10 mM sodium phosphate, 1 mM EDTA buffer (pH 7.0) to give a final duplex concentration of 3.6 µM. Prior to the thermal run, samples were degassed by placing them in a speed- vac concentrator for 2 min. Annealing curves were acquired at 260 nm starting at 95 ^{o}C and decreasing temperature at a rate of cooling of 0.5 ^{o}C min^{-1} until 15 ^{o}C, using a Varian CARY Model 3E spectrophotometer fitted with a 6-sample thermostated cell block and a temperature controller. The samples were then denatured by heating from 15 ^{o}C to 95 ^{o}C at an increasing temperature rate of 0.5 ^{o}C min^{-1} to show reversibility. Processing of the UV thermal denaturation data was carried out as described by Puglisi and Tinoco[133] and transferred to Microsoft Excel™ for viewing.

A1.3 Circular dichroism (CD) spectroscopy

Circular dichroism spectra were obtained on a Jasco J-815 spectropolarimeter equipped with a Julaba F25 circulating bath. Samples were allowed to equilibrate for 5 min at 15 ^{o}C in 90 mM NaCl, 10 mM sodium phosphate, 1 mM EDTA (pH 7.0), at a final duplex concentration of 3.6 µM. Each spectrum was an average of 3 scans, collecting at a rate of 50 nm min^{-1}, with a bandwidth of 1 nm and sampling wavelength of 0.2 nm using fused quartz cells (Starna 29-Q-10). The CD spectra were recorded from 350 to 220 nm at 15 ^{o}C. The molar ellipticity (ϕ) was calculated from the equation $\phi = \varepsilon/Cl$, where ε is the relative ellipticity (mdeg), C is the molar concentration of oligonucleotides (moles/L), and l is the path length of the cell (cm). The data were processed using software supplied by the manufacturer (JASCO, Inc.) and transferred into Microsoft Excel™ for viewing.

A1.4 AGT repair assay of the Modified DNA duplexes

DNA substrates were labeled at the 5'-end using γ-[^{32}P]-ATP as previously described.[5] Briefly, a 20 µM solution of DNA was made in 1X PNK buffer along with 1 µL γ-[^{32}P]-ATP (10 µCi µL^{-1}) and 5 units of T4 PNK. The labeling reaction was performed for 1 h at 37 °C after which the reaction was terminated by boiling the sample for 5 min. 100 pmol of labeled DNA was added to 110 pmol of the complement strand in a total volume of 50 µL of water making a 2

μM dsDNA solution with 10% excess of the non-damaged/non-modified strand. The solution was boiled for 2 min, cooled slowly to room temperature and kept in a refrigerator at 4 °C overnight to ensure proper annealing of the duplex. The repair reaction mixtures were constituted of 2 pmol of the DNA duplex and 10 pmol of AGT in a total volume of 15 μL of Activity Buffer [10 mM Tris–HCl (pH 7.6), 100 mM NaCl and 1 mM DTT] and allowed to react at 37 °C for 2.5 hours and overnight. The samples were boiled for 5 min and 1.66 μL of 0.1 M $MgCl_2$ followed by 7.5 units of *BclI* enzyme were added. Samples were incubated for at least 45 min at 37 °C. The reaction was terminated by the addition of 24 μL of stop buffer [81 mM Tris–HCl, 81 mM boric acid, 1.8 mM EDTA and 1% SDS (sodium dodecyl sulfate) (pH 8.0) in 80% formamide] followed by boiling for 2 min. Samples (10 μL) were loaded on a 14 cm × 16 cm, 20% 7 M urea denaturing polyacrylamide gel (19:1) for separation. The gels were run using 1X TBE for 1 h at 450 V and the gels exposed to a storage phosphoroscreen. The image was captured on a Typhoon 9400 (GE Healthcare, Piscataway, NJ) and the autoradiography counts obtained by Image-Quant™ (Amersham Biosciences). For the repair time course assays, master mixes of 75 μL composed of 100 pmol (5-fold) of AGT and 20 pmol of DNA substrate were prepared. Each sample was placed at 37 ^{O}C and at each time point, 7.5 μL was removed from the master mix and placed in a tube with 0.8 μL 0.1 M $MgCl_2$. 7.5 units of *BclI* enzyme were then added to each tube and they were incubated for at least 45 min at 37 °C. 12 μL of stop reaction buffer were then added and analysis was performed as described above. All reactions were analyzed by polyacrylamide gel electrophoresis (using 1X TBE for 1 h at 450 V) and visualized as described above.

A1.5 Molecular modelling

Molecular modeling was performed by using the Hyperchem 7.5 software package from Hypercube utilizing the AMBER force field. Hybridized oligomers containing a T·dA, dFU·dA, and dFU-Me·dA base pair were constructed from the nucleic acid template option using a B-form duplex. Sequence contexts were the following 5’-GGC TXG ATC ACC AG and 3’-CCG AYC TAG TGG TC for proper solvation. Duplexes were solvated with water using a periodic box. Standard Amber99 parameters were used with the dielectric set to constant “one to four scale factors”

and non-bonded interactions were set to 0.5 (both electrostatic and van der Waals). Cutoffs were applied to "switched" to an outer and inner radius of 14.5 and 10.5 Å. All structures were geometry optimized using Polak-Ribiere conjugate gradient until the RMS gradient was less than 0.1 kcal/(Å mol) using the periodic boundary condition option.

A1.6 Geometry optimization

GaussView was the visualization software employed and **Gaussian 09** was the software used for the calculation of the geometrical optimized structures under the B3LYP/6-31+G(2d,2p) mode. [136-139]

Figure A1.1: Structures of (**A**) O^6-benzylguanine (hAGT inhibitor).[90–92] (**B**) 5-fluorouracil.[96,134] (**C**) 2'-deoxy-2'-fluoro-arabinonucleic acid (F-ANA)[72]

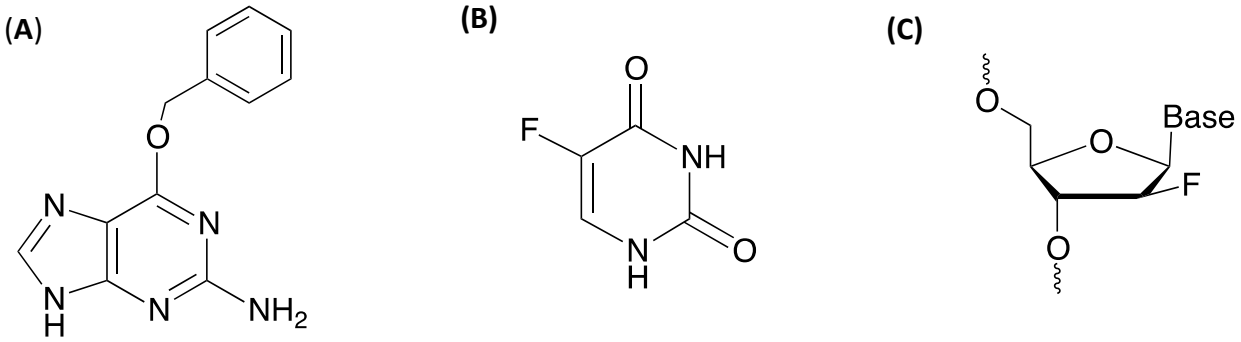


Figure A1.2 - 500 MHz ^{1}H NMR spectrum of compound (**3**) (in d_6-acetone)

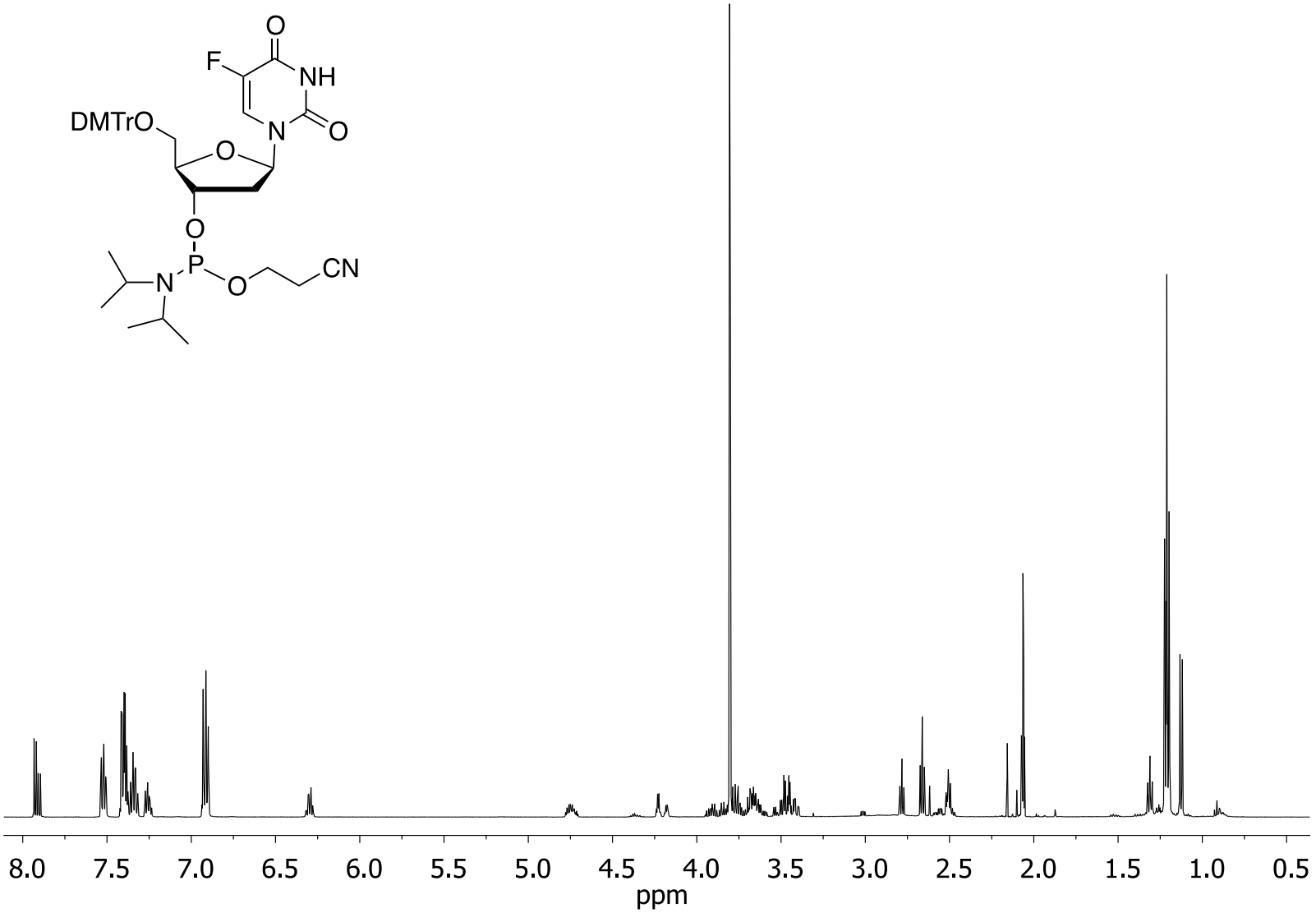


Figure A1.3 - 125.7 MHz ^{13}C NMR spectrum of compound (**3**) (in d_6-acetone)

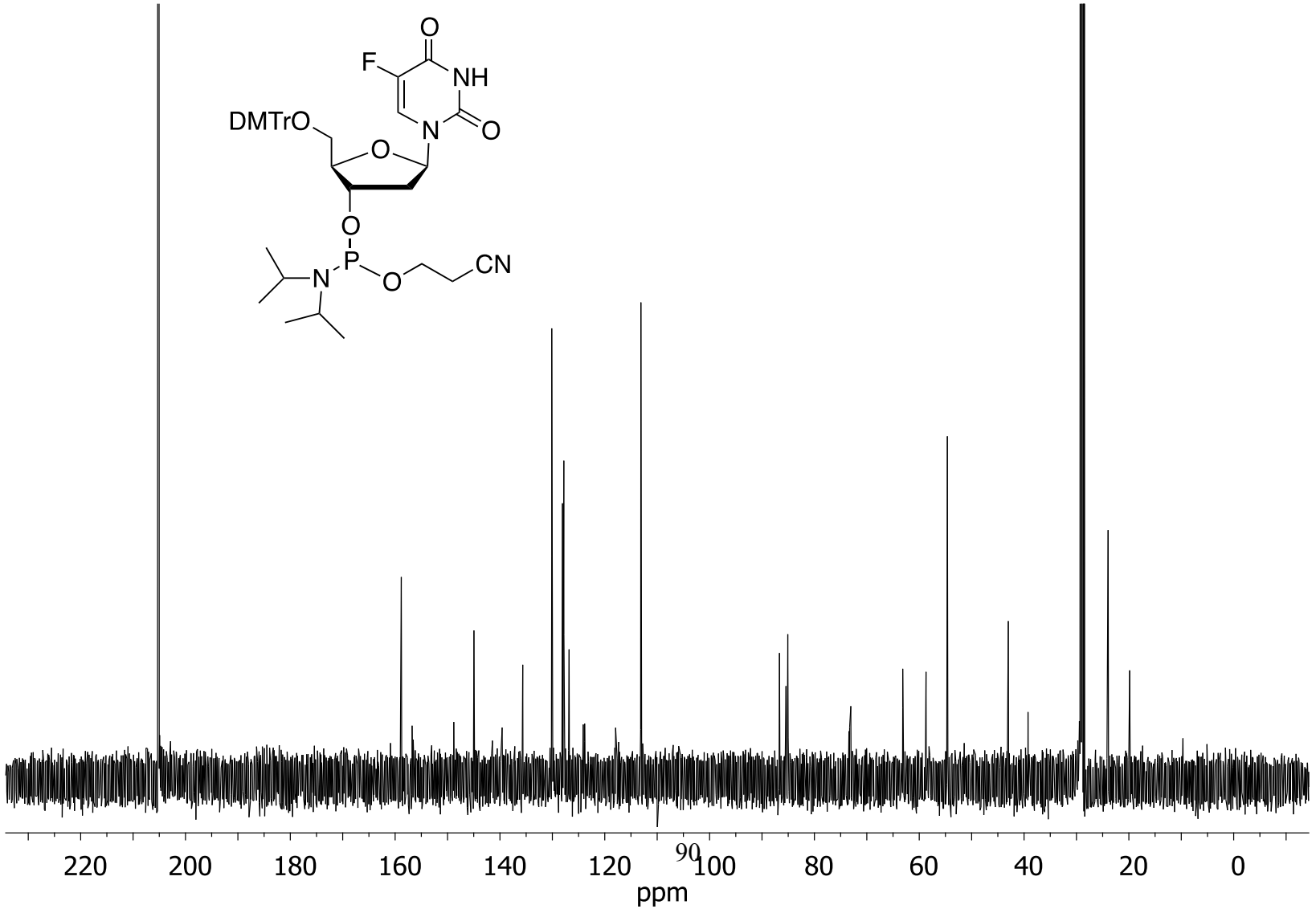


Figure A1.4 - 470.4 MHz ^{19}F NMR spectrum of compound (**3**) (in d_6-acetone)

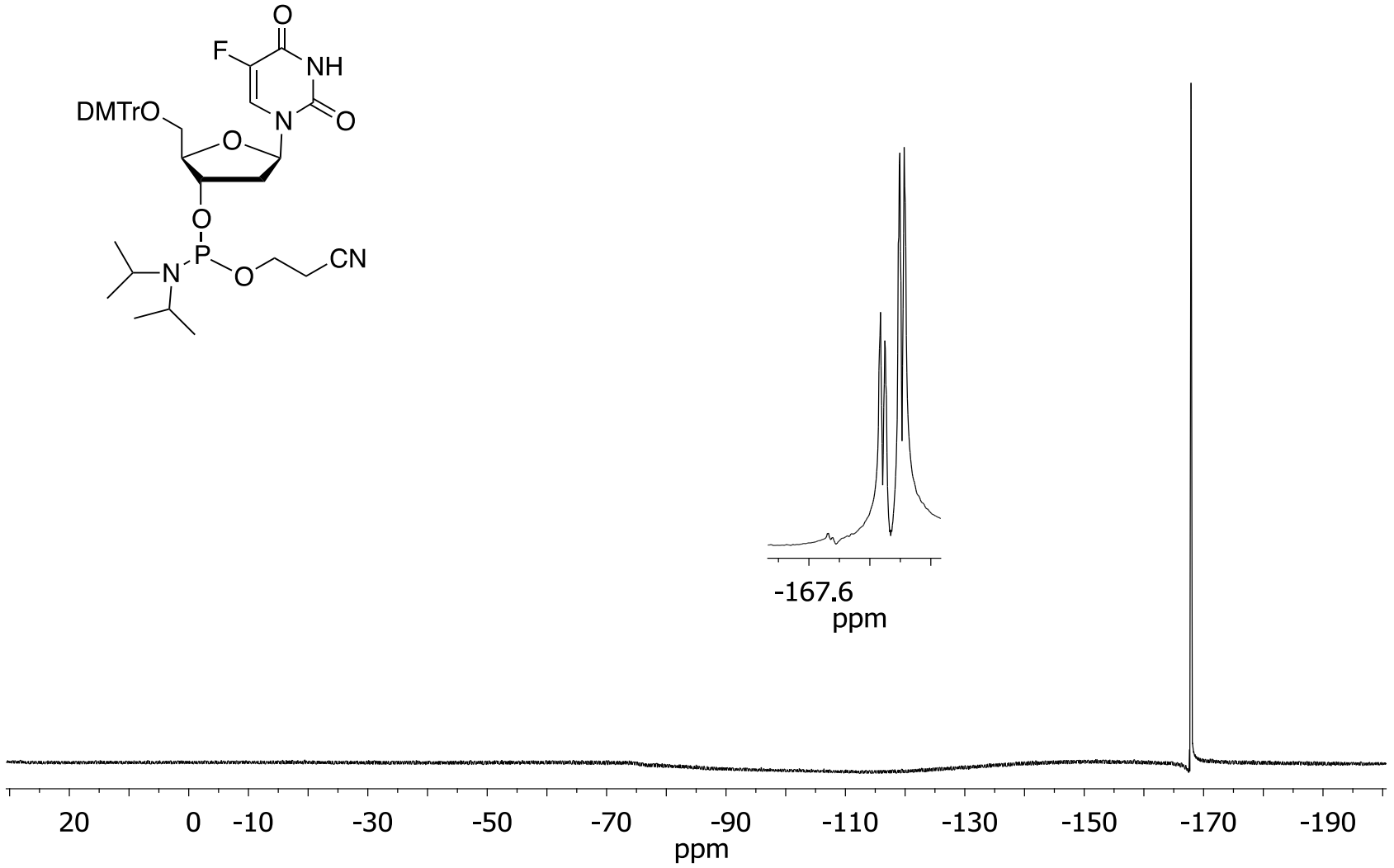


Figure A1.5 - 202.3 MHz ^{31}P NMR spectrum of compound (**3**) (in d_6-acetone)

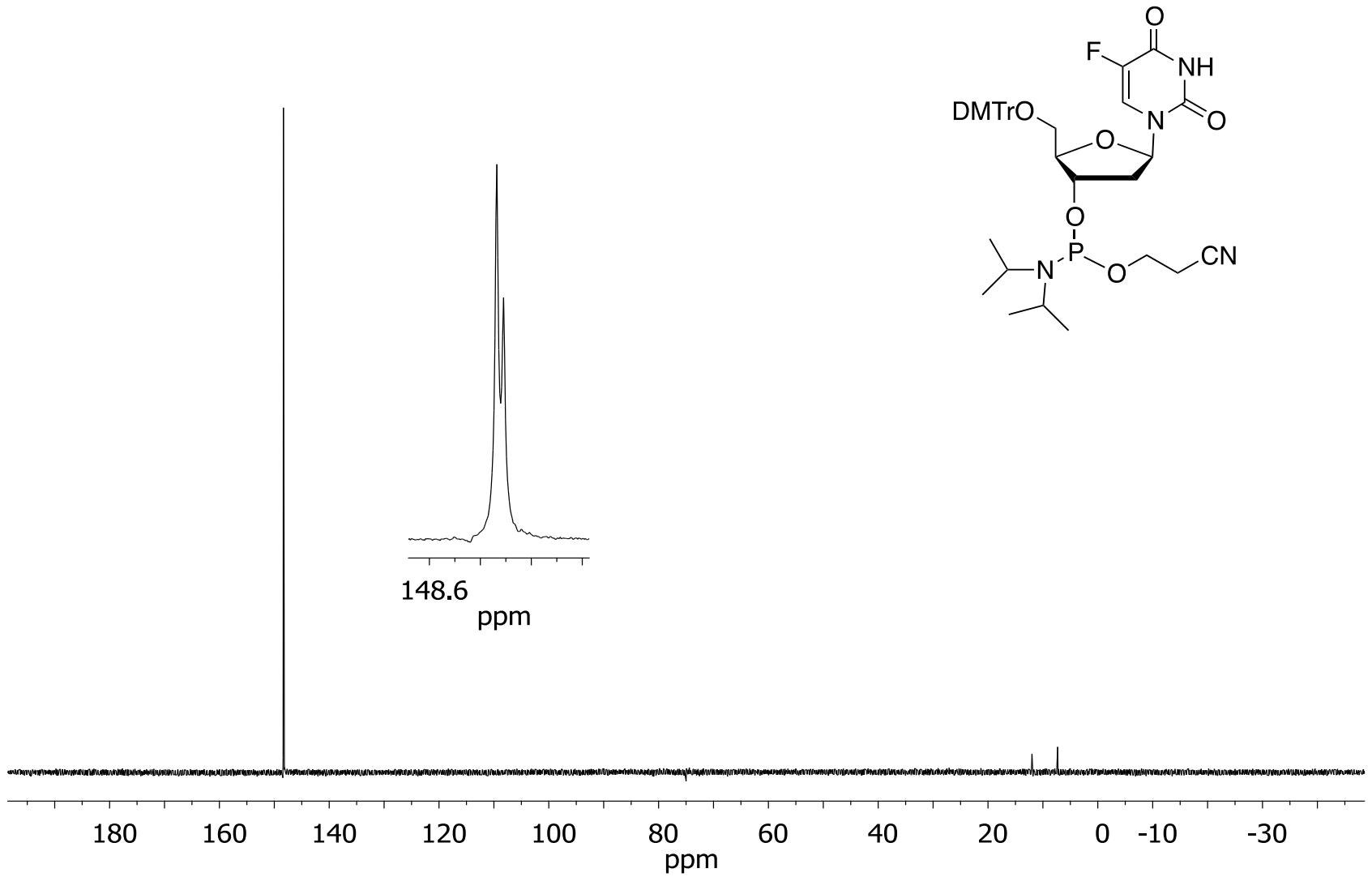


Figure A1.6 - 500 MHz ^{1}H NMR spectrum of compound (**5**) (in $CDCl_3$)

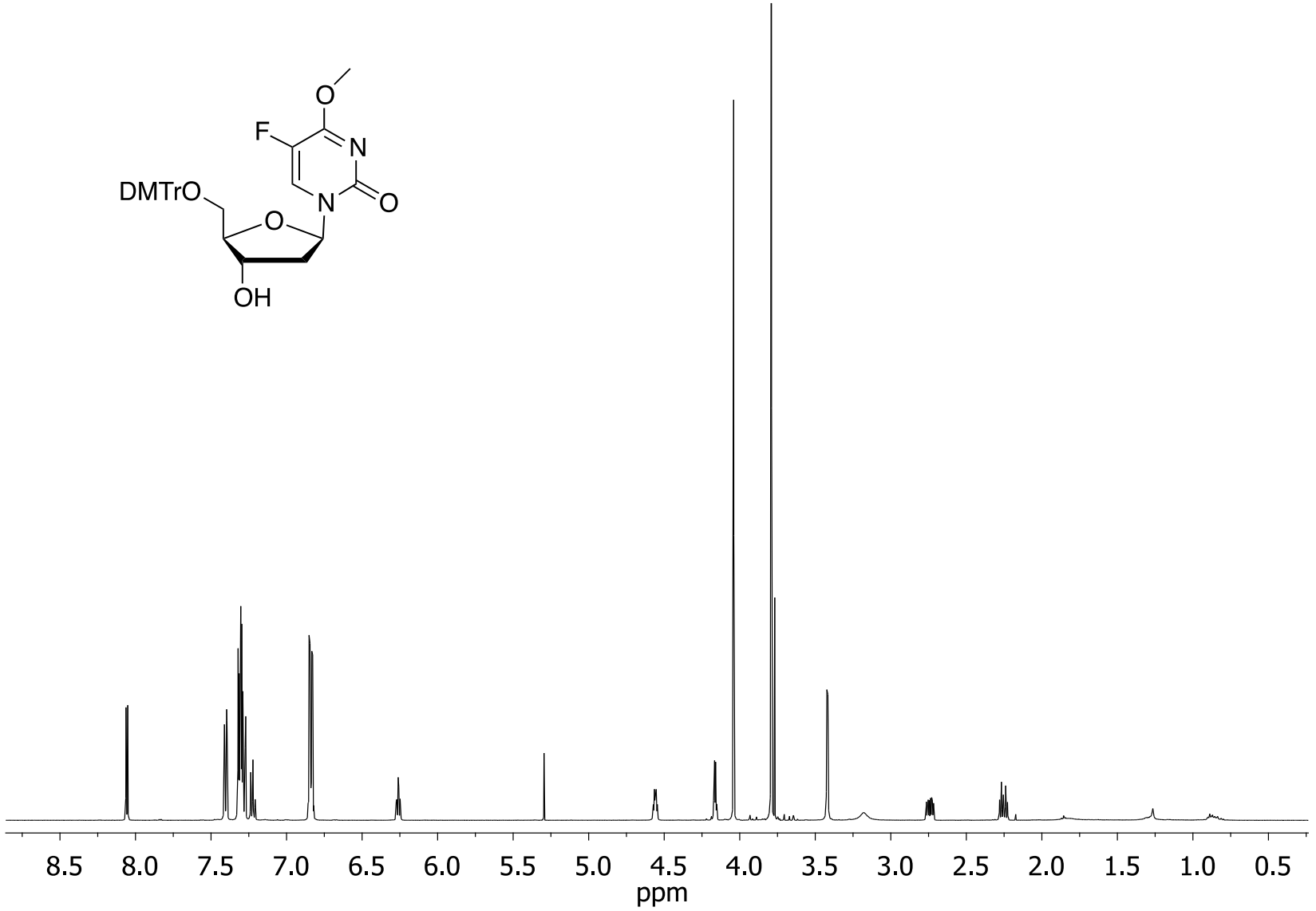


Figure A1.7 - 125.7 MHz ^{13}C NMR spectrum of compound (**5**) (in $CDCl_3$)

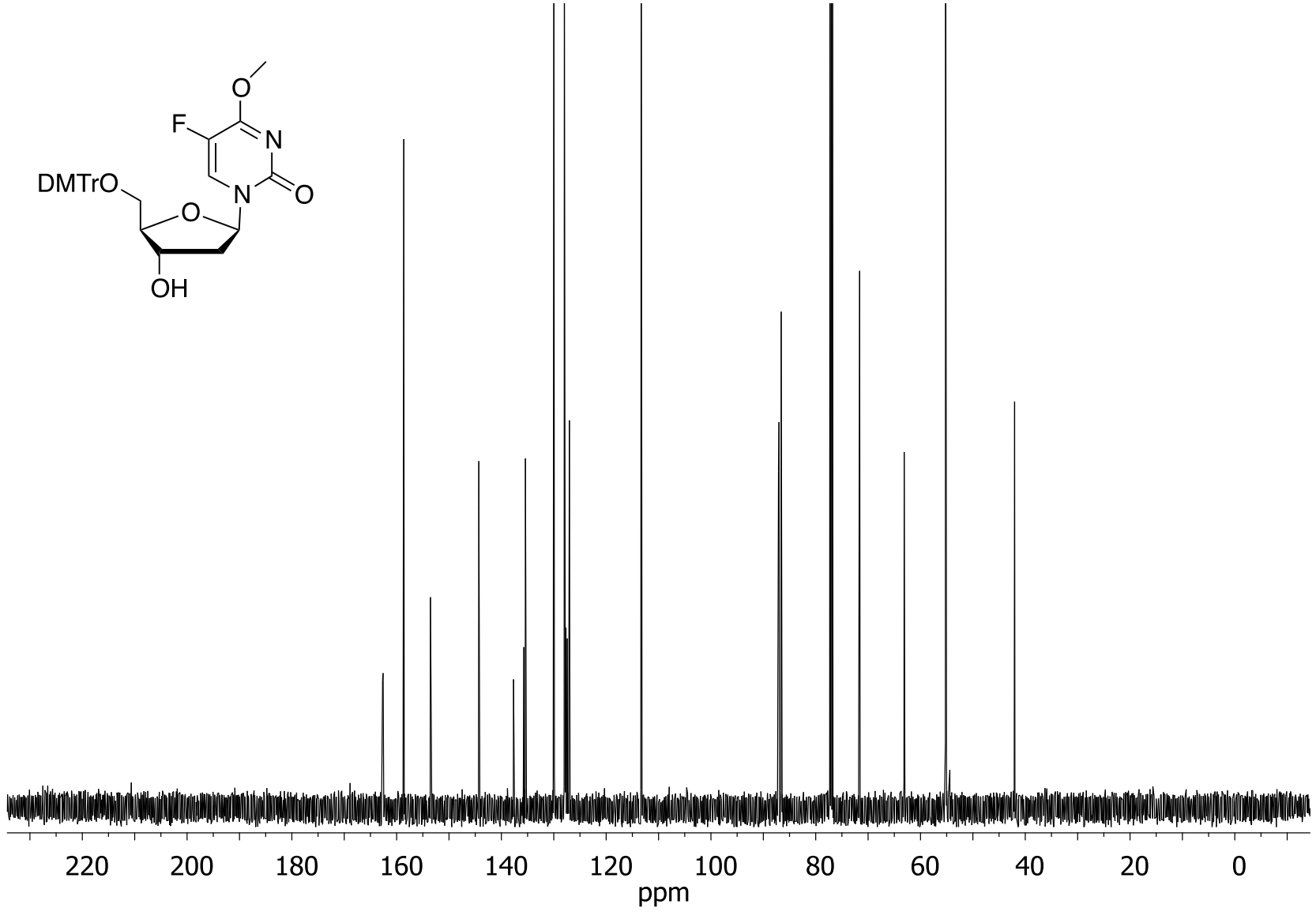


Figure A1.8 - 470.4 MHz ^{19}F NMR spectrum of compound (**5**) (in $CDCl_3$)

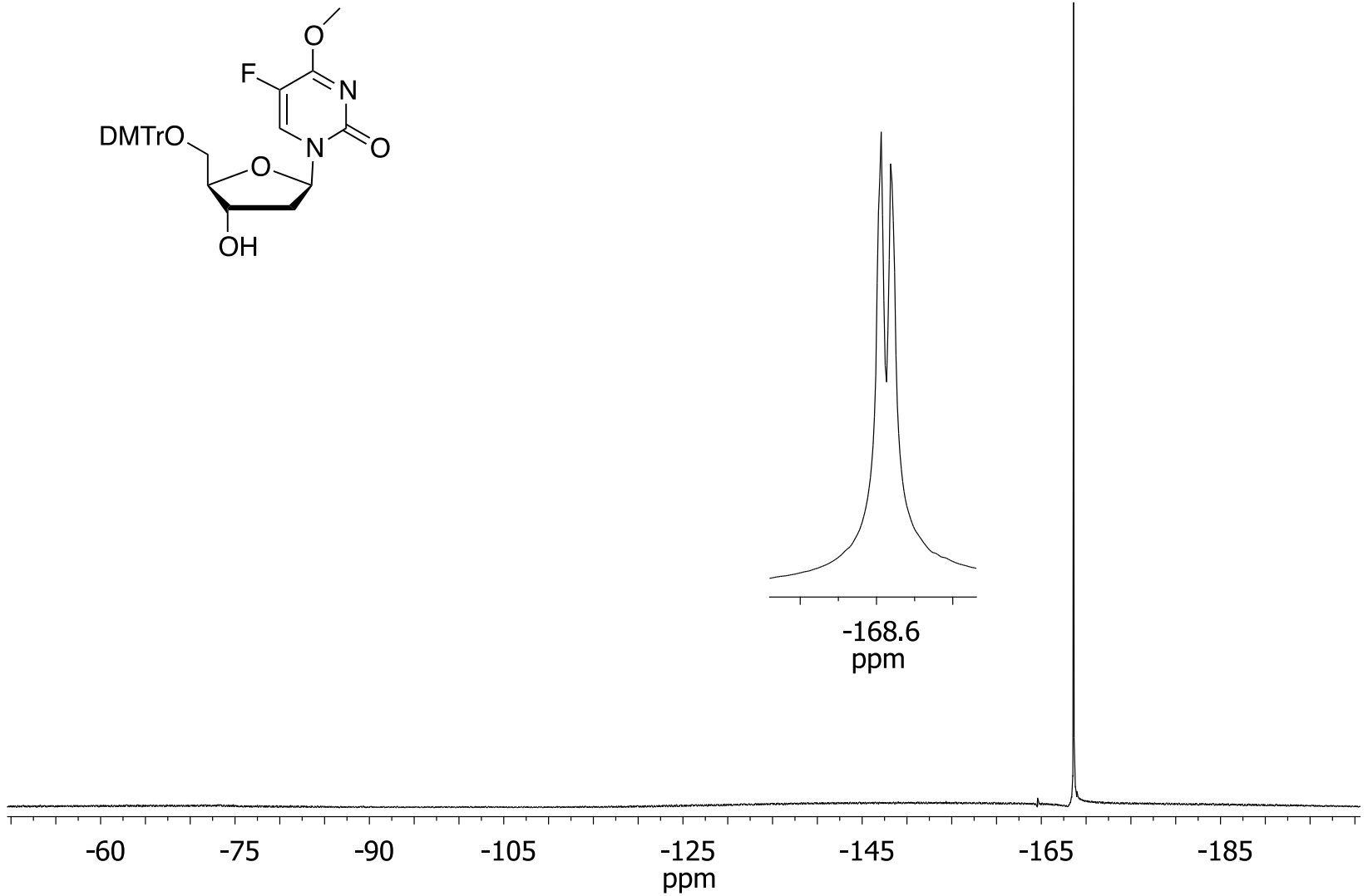


Figure A1.9 - 500 MHz ^{1}H NMR spectrum of compound (**6**) (in d_6-acetone)

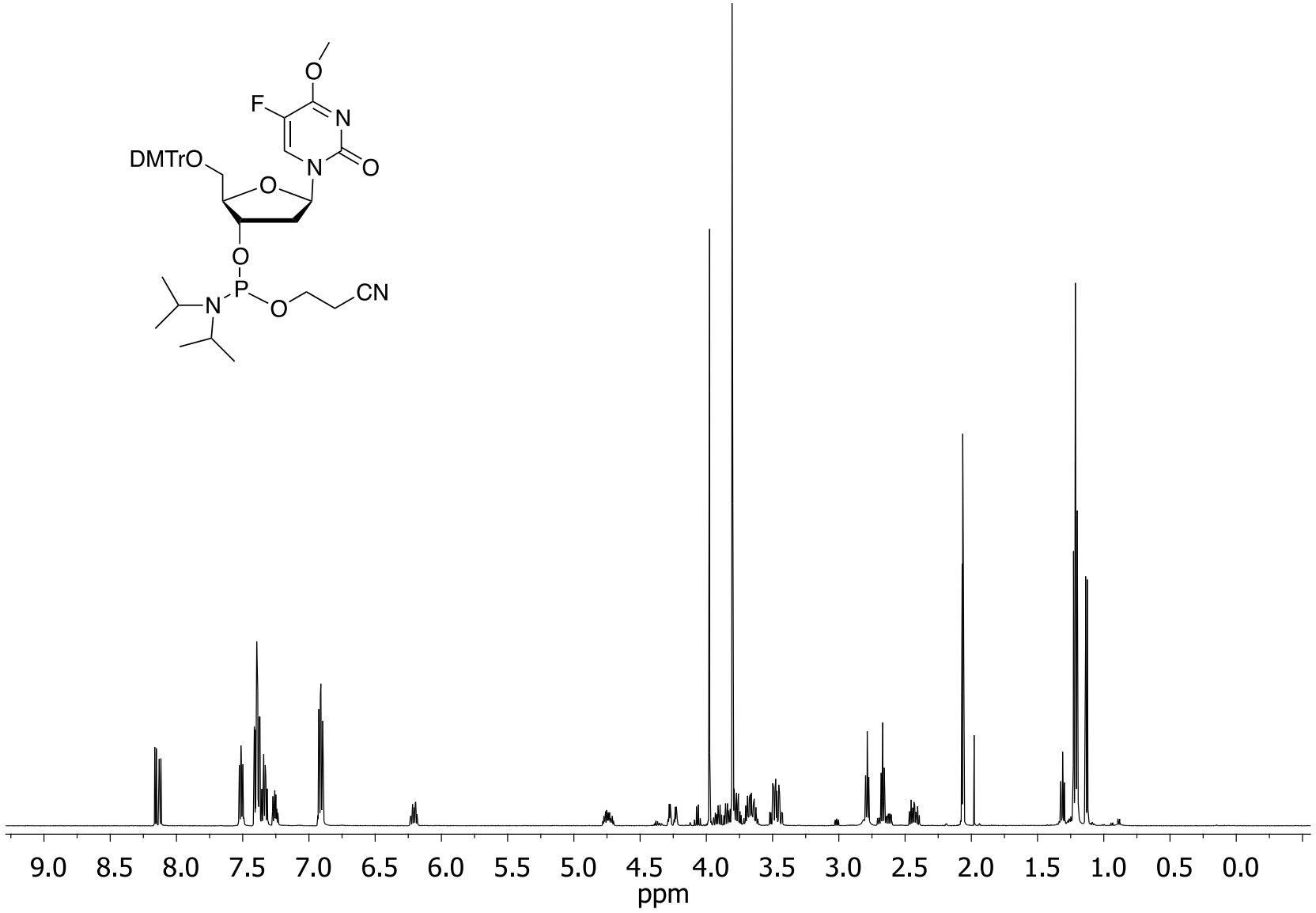


Figure A1.10 - 125.7 MHz ^{13}C NMR spectrum of compound (**6**) (in d_6-acetone)

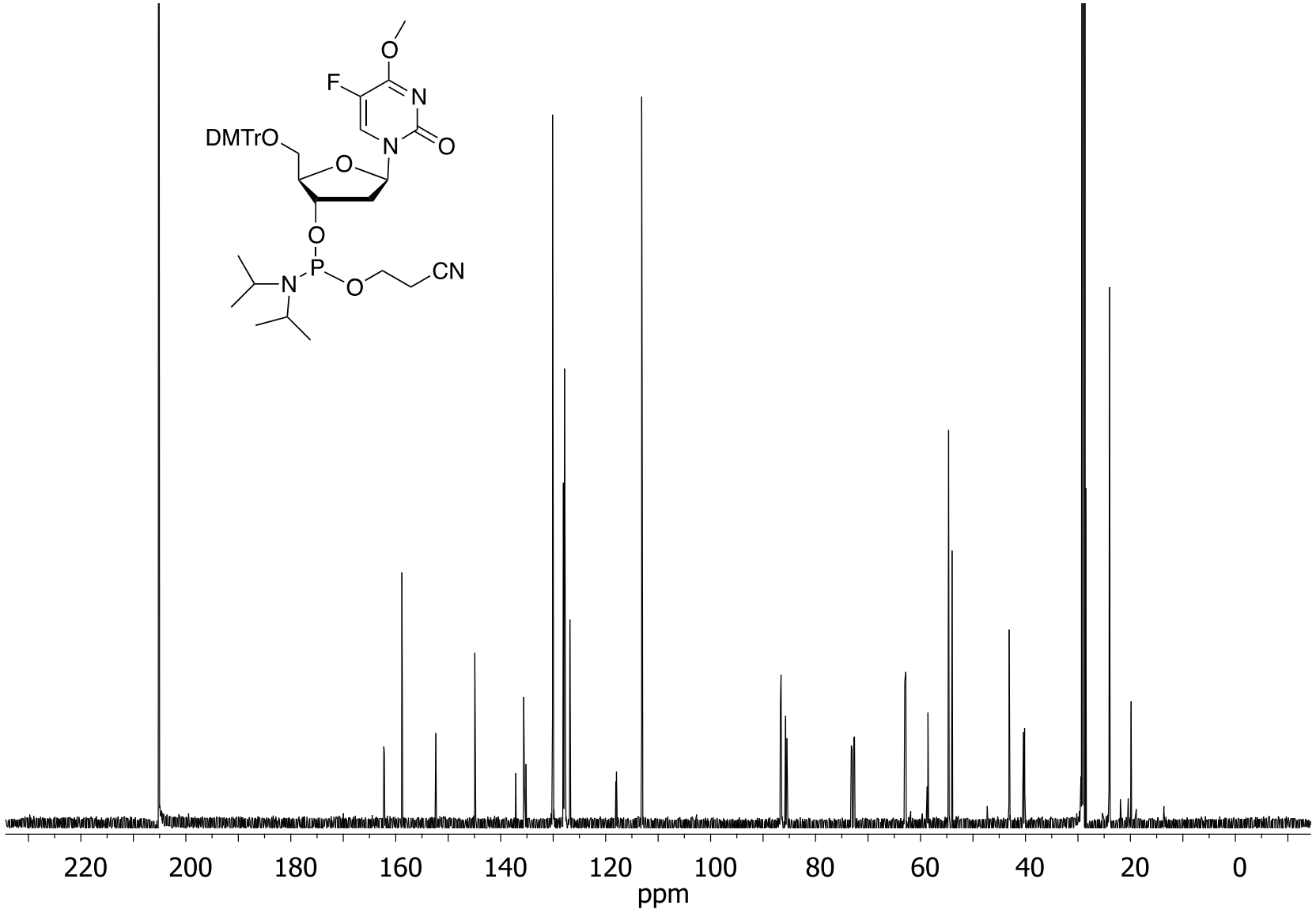


Figure A1.11 - 470.4 MHz ^{19}F NMR spectrum of compound (**6**) (in d_6-acetone)

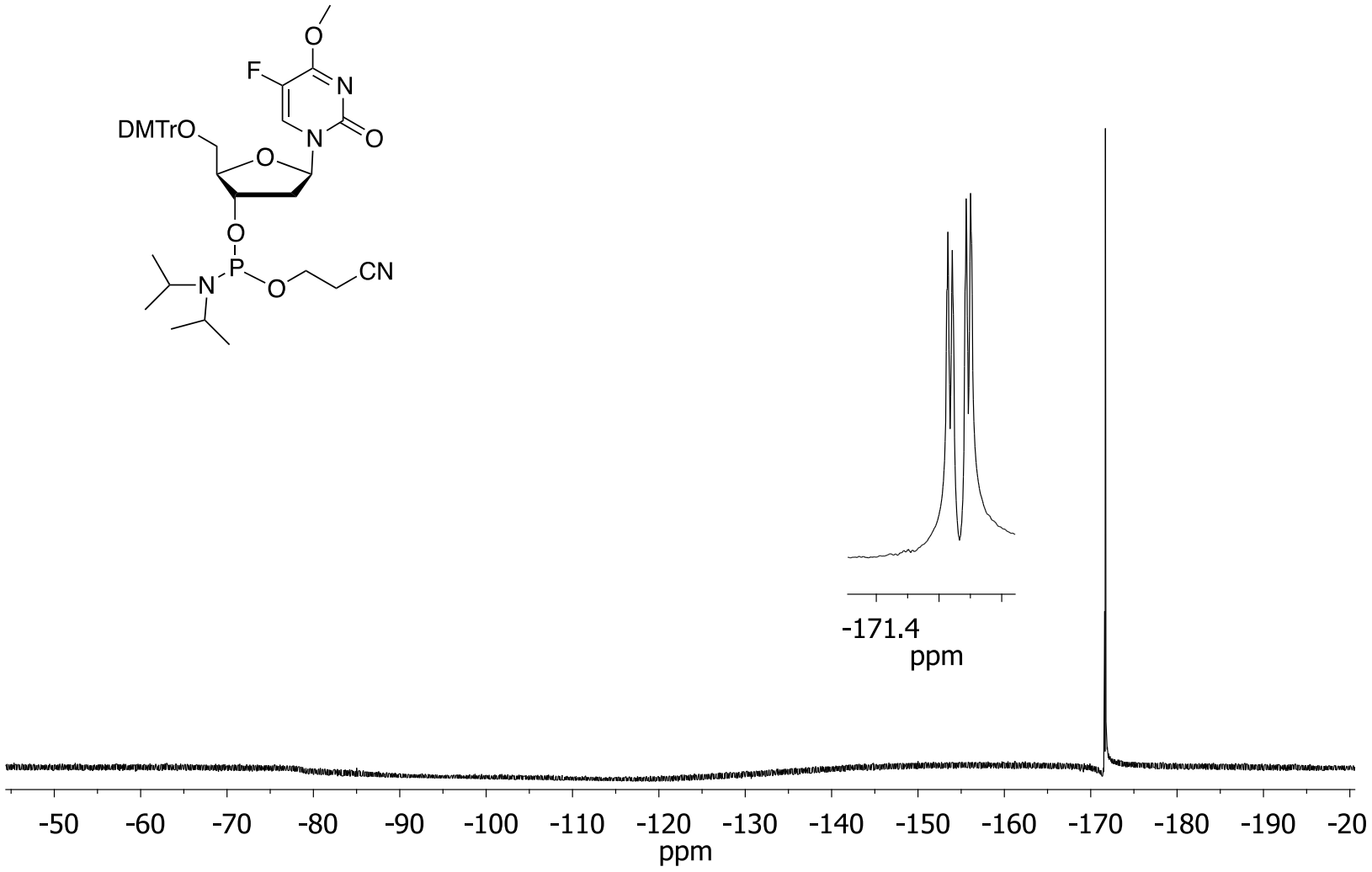


Figure A1.12 - 202.3 MHz ^{31}P NMR spectrum of compound (**6**) (in d_6-acetone)

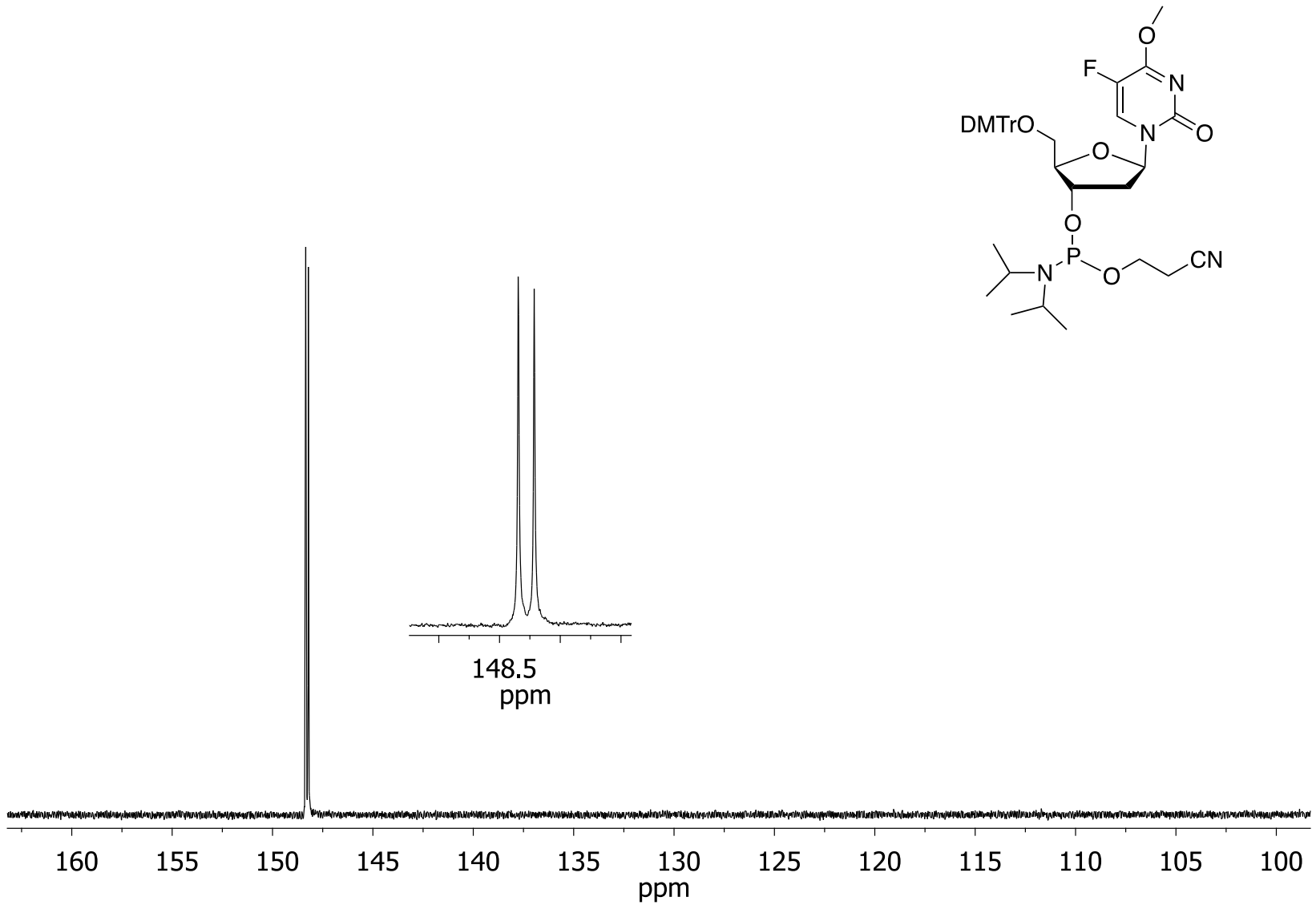


Figure A1.13 - ESI MS spectrum of oligonucleotide **dFU** (expected mass of 4268.5)

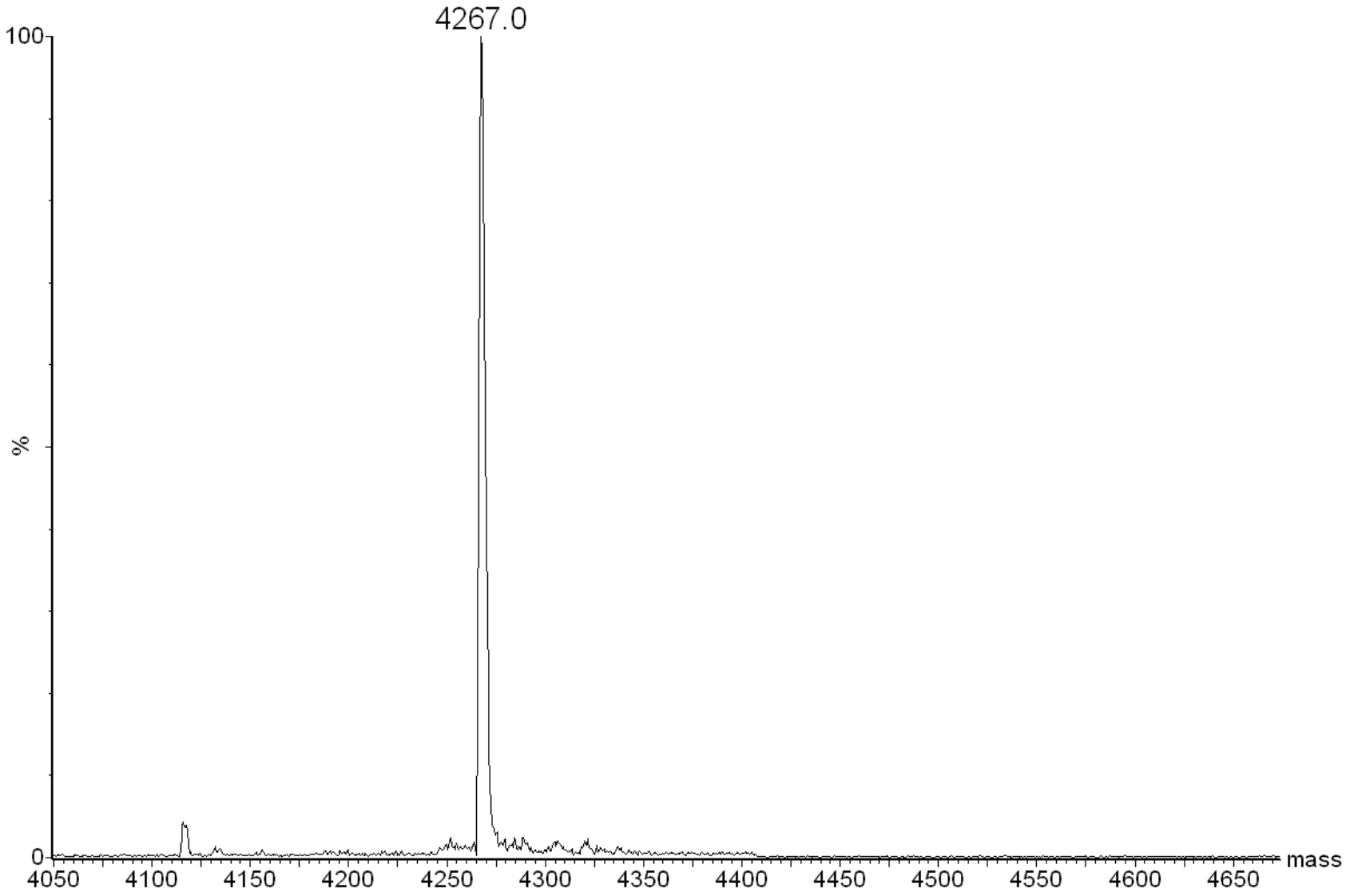


Figure A1.14 - ESI MS spectrum of oligonucleotide **dFU-Me** (expected mass of 4282.1)

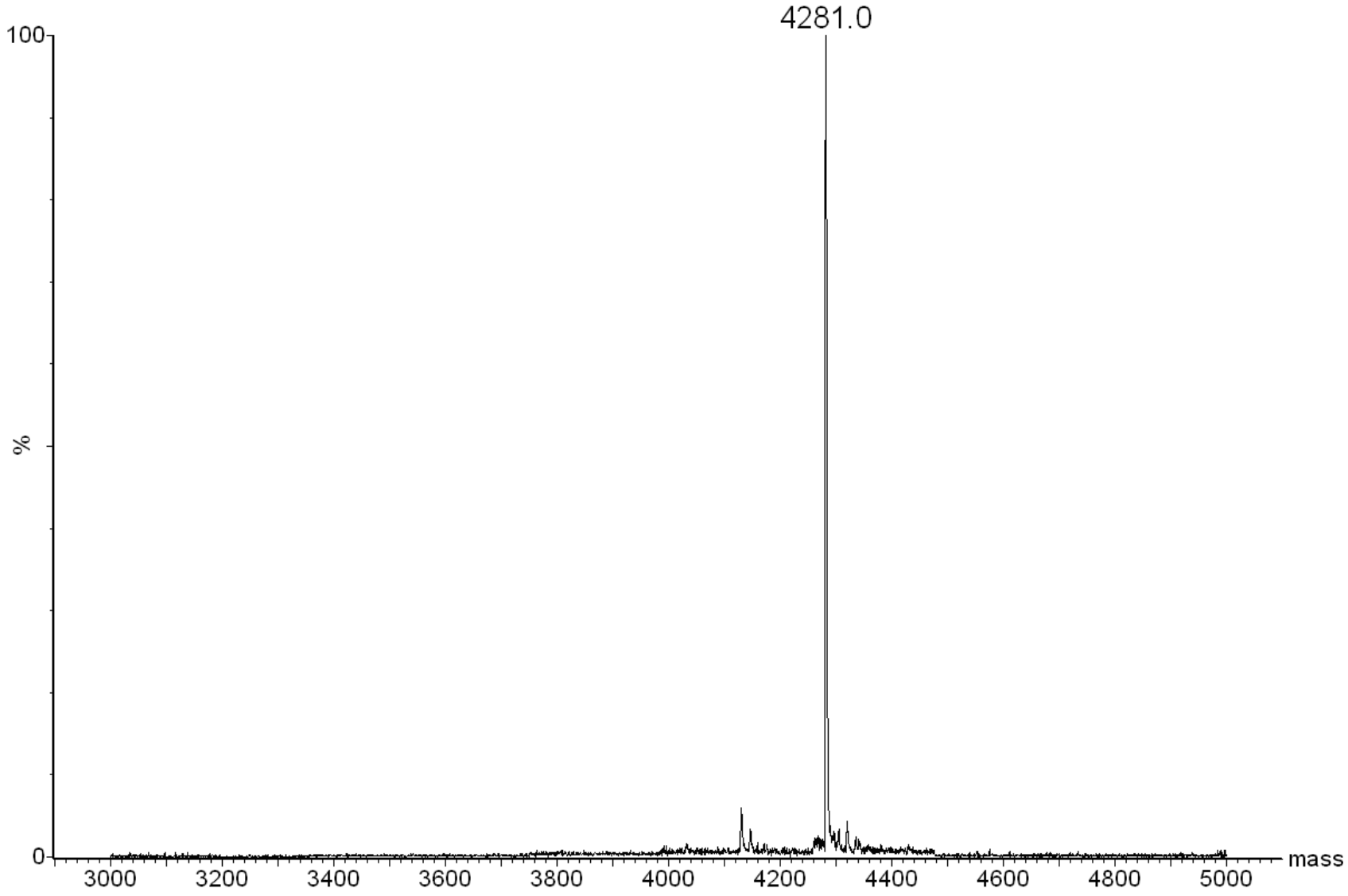
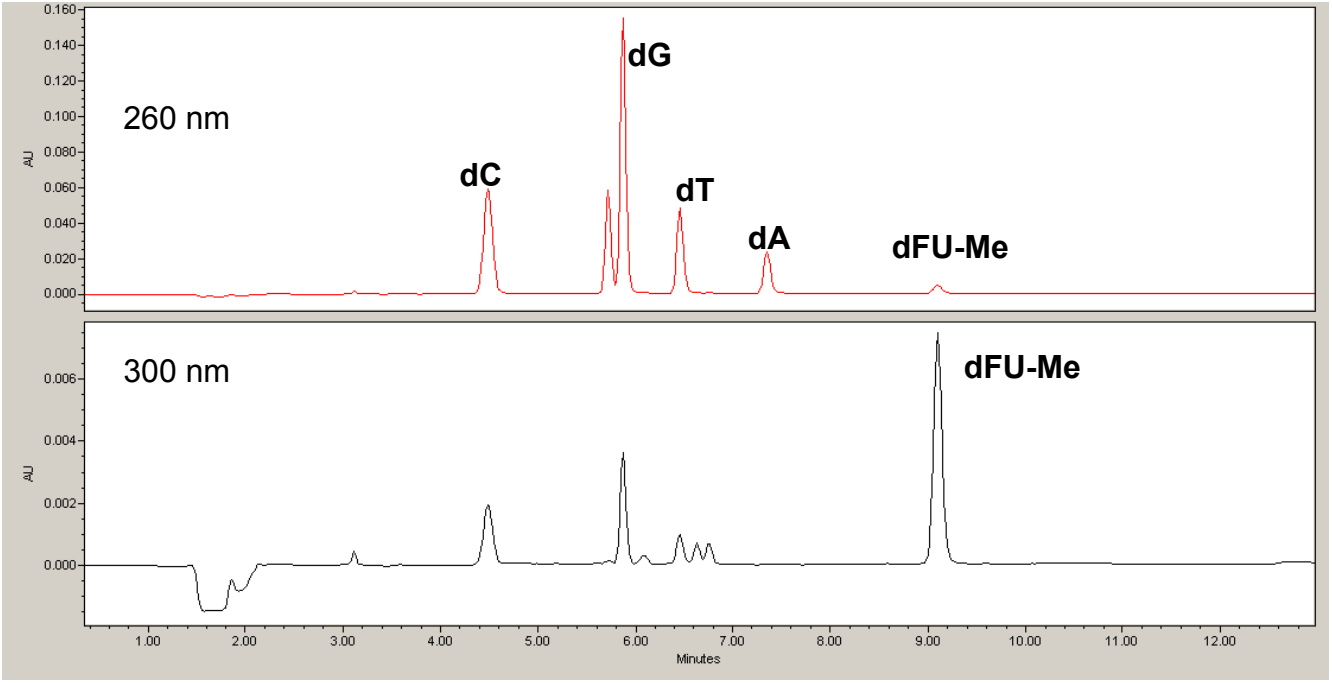


Figure A1.15- C-18 reversed phase HPLC profile of nuclease-digested **dFU-Me**. The column was eluted with a linear gradient of 0-70% buffer B over 30 min (buffer A: 50 mM sodium phosphate, pH 5.8, 2% MeCN and buffer B: 50 mM sodium phosphate, pH 5.8, 50% MeCN).



Oligomer	Nucleoside composition	Nucleoside Ratio	
		Expected	Observed
dFU-Me	dC	4	4.0
	dG	4	4.1
	dA (dI)	3	2.4
	dT	2	2.0
	dFU-Me	1	0.5

Figure A1.16- SAX-HPLC profile of **dFU-Me** crude **(A)** and pure **(B)**. The column (Dionex DNAPAC PA- 100) was eluted using a linear gradient of 0-70% buffer B over 30 min (buffer A: 100 mM Tris HCl, pH 7.5, 10% MeCN and buffer B: 100mM Tris HCl, pH 7.5, 10% MeCN, 1 M NaCl)

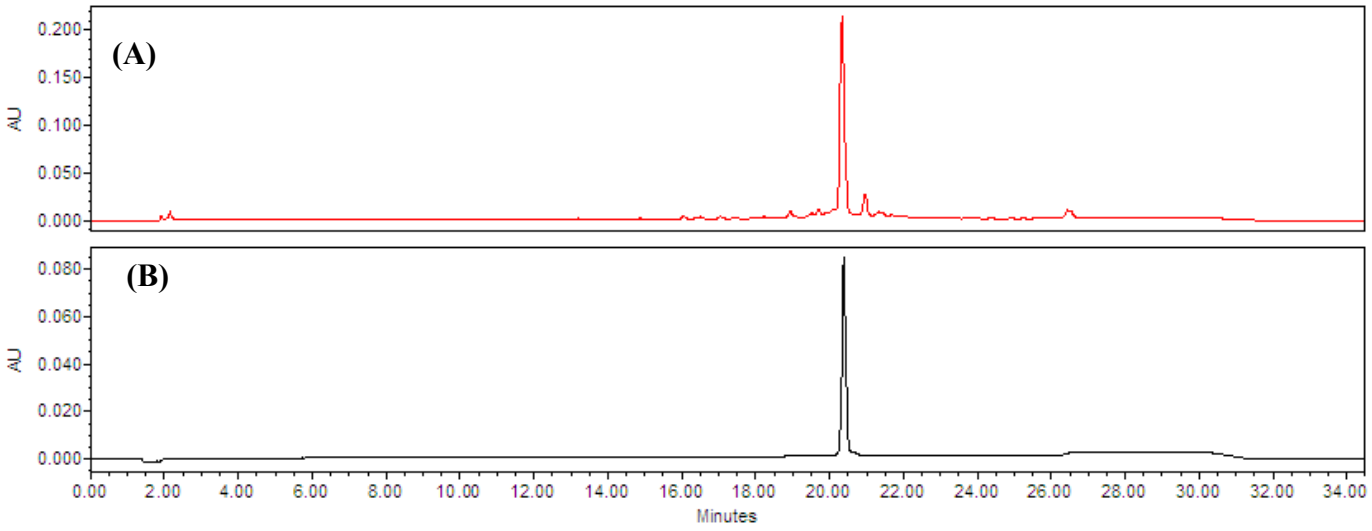


Figure A1.17 - Hyperchromicity change (A_{260}) *versus* temperature (^{o}C) profiles of duplexes containing **dFU** (—), **dFU-Me** (———) and unmodified **T** control DNA (•••).

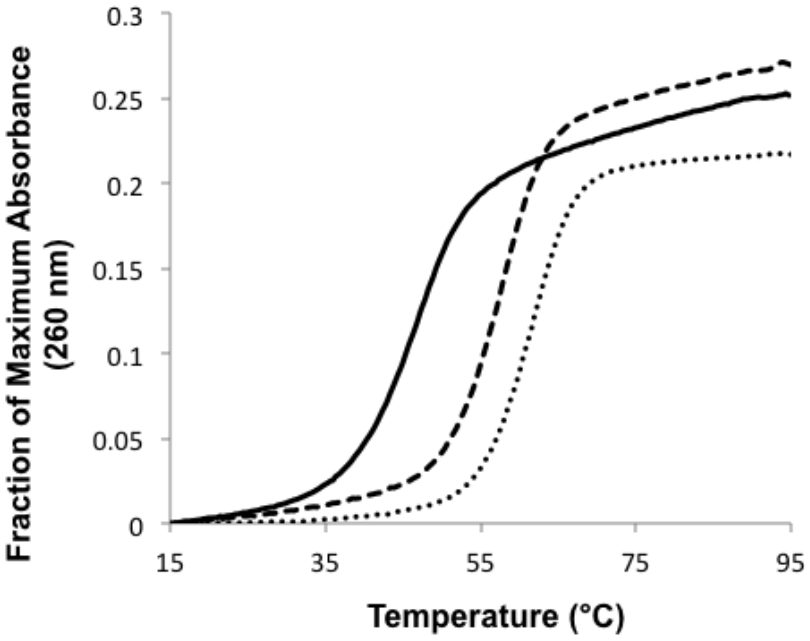


Figure A1.18 - Circular dichroism spectra of duplexes containing **dFU** (—), **dFU-Me** (———) and unmodified control **T** DNA(•••).

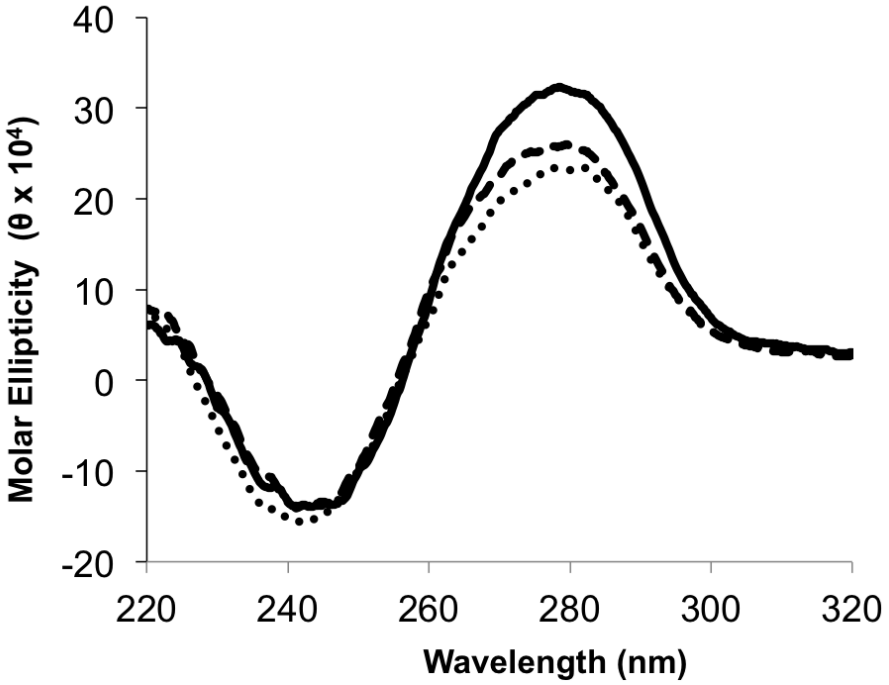


Figure A1.19 - Time course repair gel of duplexes containing **dFU-Me** by (**A**) **hAGT**, (**B**) **OGT**, (**C**) **Ada-C** and (**D**) **hOGT**. (**A**) Denaturing gel of the repair of 2 pmol of **dFU-Me** by 10 pmol hAGT as a function of time: lane 1, 2 pmol Control; lanes 2-10, 2 pmol + 10 pmol hAGT incubated for 0.25, 0.5, 0.75, 1, 1.5, 2, 5, 15, 45 min, respectively. (**B**) Denaturing gel of the repair of 2 pmol of **dFU-Me** by 10 pmol OGT as a function of time: lane 1, 2 pmol Control; lanes 2-10, 2 pmol + 10 pmol OGT incubated for 0.25, 0.5, 0.75, 1, 1.5, 2, 5, 15, 45 min, respectively. (**C**) Denaturing gel of the repair of 2 pmol of **dFU-Me** by 10 pmol Ada-C as a function of time: lane 1, 2 pmol Control; lanes 2-10, 2 pmol + 10 pmol Ada-C incubated for 0.25, 0.5, 0.75, 1, 1.5, 2, 3, 5, 15, 45, 150 min, respectively. (**D**) Denaturing gel of the repair of 2 pmol of **dFU-Me** by 10 pmol hOGT as a function of time: lane 1, 2 pmol Control; lanes 2-10, 2 pmol + 10 pmol hOGT incubated for 0.25, 0.5, 0.75, 1, 1.5, 2, 3, 5, 15, 45, 150 min, respectively. All the time courses were performed at room temperature.

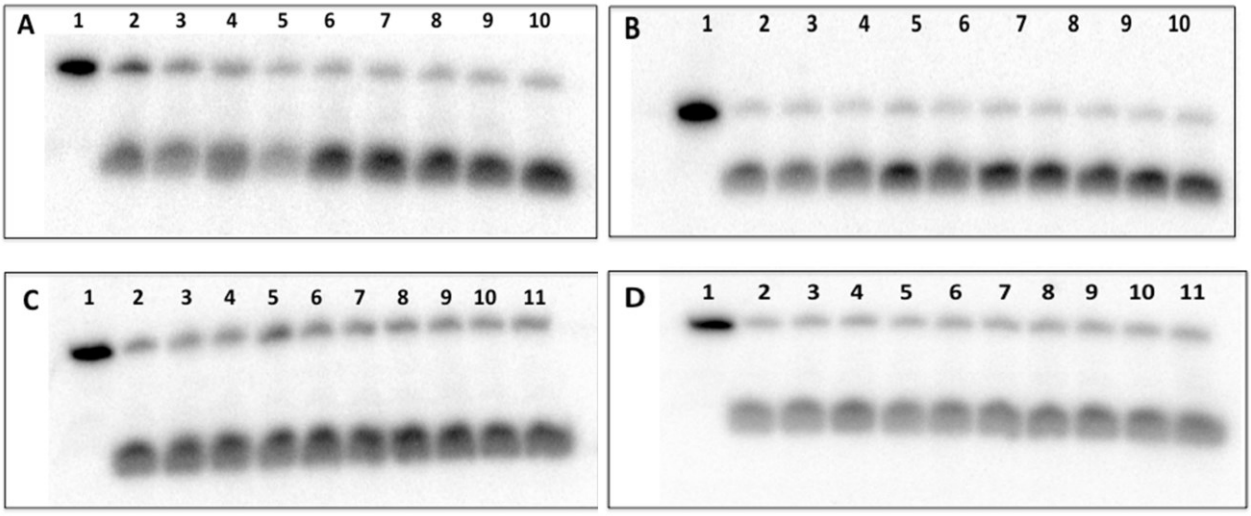
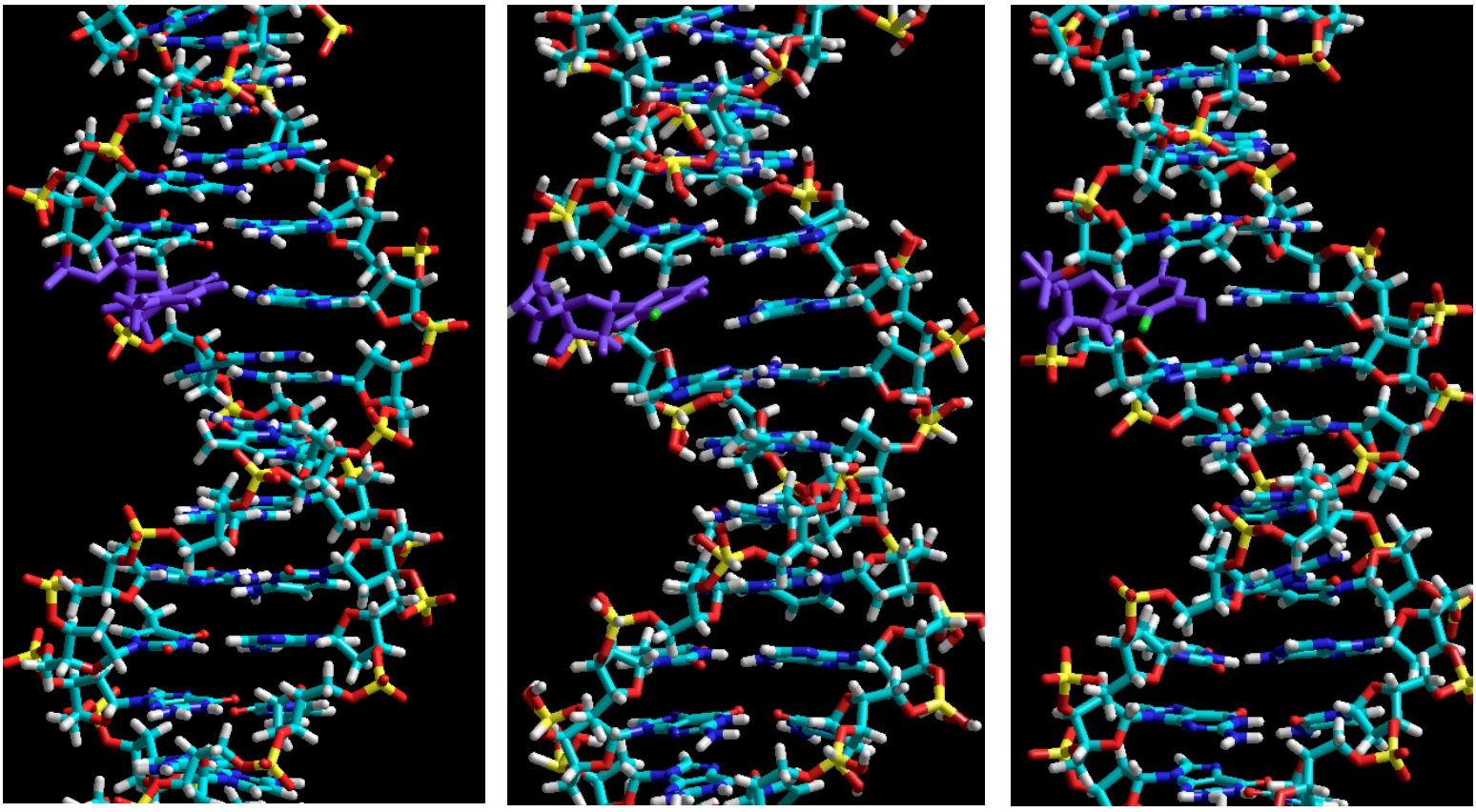
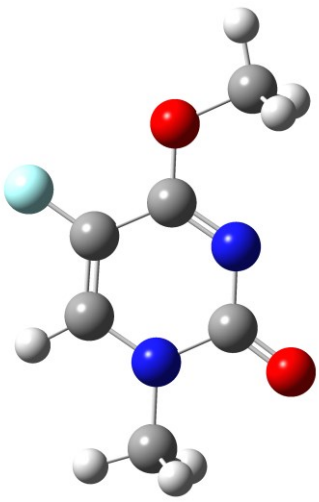


Figure A1.20 – Molecular models of unmodified control duplex (**T**) and duplexes containing **dFU** and **dFU-Me** that were geometry optimized using the AMBER forcefield



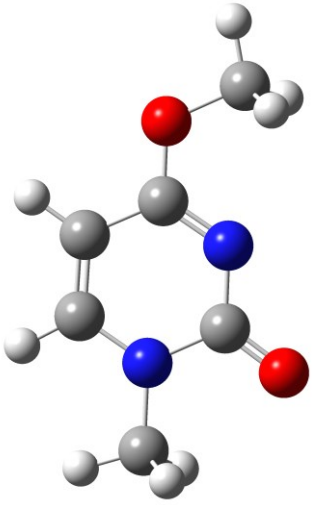
Unmodified Control **dFU** **dFU-Me**

Figure A1.21– Geometry optimization of the **dFU-Me, dU-Me** and **T-Me** nucleobases with a methyl group at the *N*1 atom using Gaussian 09 and B3LYP/6-31+G(2d,2p) mode (Red sphere: oxygen; White sphere: hydrogen; Grey sphere: carbon; Blue sphere: nitrogen; Light blue sphere: fluorine)



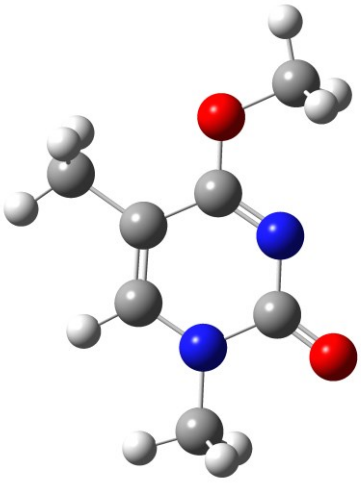
dFU-Me

O^4-C_α length: 1.440 Å



dU-Me

O^4-C_α length: 1.438 Å



T-Me

O^4-C_α length: 1.438 Å

Appendix II: Supporting Information for Chapter 3

Contents **Page**

Supporting Methods

A2.1 Synthesis and characterization of **dFU** nucleosides and oligonucleotides **110**
A2.2 UV thermal denaturation **119**
A2.3 Circular dichroism (CD) spectroscopy **120**
A2.4 AGT repair assay of **dFU-C2**, **dFU-Bn**, **dFU-C4OH** and **dFU-C7OH** DNA duplexes **120**

Supporting Figures

Figure A2.1 - Examples of different DNA adducts mentioned in the main text **120**
Figure A2.2 - 500 MHz ^{1}H NMR spectrum of compound (**2a**) (in $CDCl_3$) **121**
Figure A2.3 - 125.7 MHz ^{13}C NMR spectrum of compound (**2a**) (in $CDCl_3$) **122**
Figure A2.4 - 470.4 MHz ^{19}F NMR spectrum of compound (**2a**) (in $CDCl_3$) **123**
Figure A2.5 - 500 MHz ^{1}H NMR spectrum of compound (**3a**) (in d_6-acetone) **124**
Figure A2.6 - 125.7 MHz ^{13}C NMR spectrum of compound (**3a**) (in d_6-acetone) **125**
Figure A2.7 - 470.4 MHz ^{19}F NMR spectrum of compound (**3a**) (in d_6-acetone) **126**
Figure A2.8 - 202.3 MHz ^{31}P NMR spectrum of compound (**3a**) (in d_6-acetone) **127**
Figure A2.9 - 500 MHz ^{1}H NMR spectrum of compound (**2b**) (in $CDCl_3$) **128**
Figure A2.10 - 125.7 MHz ^{13}C NMR spectrum of compound (**2b**) (in $CDCl_3$) **129**
Figure A2.11 - 470.4 MHz ^{19}F NMR spectrum of compound (**2b**) (in $CDCl_3$) **130**
Figure A2.12 - 500 MHz ^{1}H NMR spectrum of compound (**3b**) (in d_6-acetone) **131**
Figure A2.13 - 125.7 MHz ^{13}C NMR spectrum of compound (**3b**) (in d_6-acetone) **132**
Figure A2.14 - 470.4 MHz ^{19}F NMR spectrum of compound (**3b**) (in d_6-acetone) **133**
Figure A2.15 - 202.3 MHz ^{31}P NMR spectrum of compound (**3b**) (in d_6-acetone) **134**
Figure A2.16 - 500 MHz ^{1}H NMR spectrum of compound (**4a**) (in $CDCl_3$) **135**
Figure A2.17 - 125.7 MHz ^{13}C NMR spectrum of compound (**4a**) (in $CDCl_3$) **136**
Figure A2.18 - 470.4 MHz ^{19}F NMR spectrum of compound (**4a**) (in $CDCl_3$) **137**
Figure A2.19 - 500 MHz ^{1}H NMR spectrum of compound (**5a**) (in d_6-acetone) **138**
Figure A2.20 - 125.7 MHz ^{13}C NMR spectrum of compound (**5a**) (in d_6-acetone) **139**
Figure A2.21 - 470.4 MHz ^{19}F NMR spectrum of compound (**5a**) (in d_6-acetone) **140**
Figure A2.22 - 202.3 MHz ^{31}P NMR spectrum of compound (**5a**) (in d_6-acetone) **141**
Figure A2.23 - 500 MHz ^{1}H NMR spectrum of compound (**4b**) (in $CDCl_3$) **142**
Figure A2.24 - 125.7 MHz ^{13}C NMR spectrum of compound (**4b**) (in $CDCl_3$) **143**
Figure A2.25 - 470.4 MHz ^{19}F NMR spectrum of compound (**4b**) (in $CDCl_3$) **144**
Figure A2.26 - 500 MHz ^{1}H NMR spectrum of compound (**5b**) (in d_6-acetone) **145**

Figure A2.27 - 125.7 MHz ^{13}C NMR spectrum of compound (**5b**) (in d_6-acetone) **146**
Figure A2.28 - 470.4 MHz ^{19}F NMR spectrum of compound (**5b**) (in d_6-acetone) **147**
Figure A2.29 - 202.3 MHz ^{31}P NMR spectrum of compound (**5b**) (in d_6-acetone) **148**
Figure A2.30 - ESI MS spectrum of oligonucleotide **dFU-Et** **149**
Figure A2.31 - ESI MS spectrum of oligonucleotide **dFU-Bn** **150**
Figure A2.32 - ESI MS spectrum of oligonucleotide **dFU-C4OH** **151**
Figure A2.33 - ESI MS spectrum of oligonucleotide **dFU-C7OH** **152**
Figure A2.34 - C-18 RP-HPLC profile of nuclease digested **dFU-Et** **153**
Figure A2.35 - C-18 RP-HPLC profile of nuclease digested **dFU-Bn** **154**
Figure A2.36 - C-18 RP-HPLC profile of nuclease digested **dFU-C4OH** **155**
Figure A2.37 - C-18 RP-HPLC profile of nuclease digested **dFU-C7OH** **156**
Figure A2.38 - SAX-HPLC profile of crude and pure **dFU-Et** **157**
Figure A2.39 - SAX-HPLC profile of crude and pure **dFU-Bn** **157**
Figure A2.40 - SAX-HPLC profile of crude and pure **dFU-C4OH** **158**
Figure A2.41 - SAX-HPLC profile of crude and pure **dFU-C7OH** **158**
Figure A2.42 - UV thermal denaturation experiments for unmodified DNA duplex (**dF**), **dFU-Et**, **dFU-Bn**, **dFU-C4OH** and **dFU-C7OH** **159**
Figure A2.43 - Circular dichroism spectra for unmodified DNA duplex (**dF**), **dFU-Et**, **dFU-Bn**, **dFU-C4OH** and **dFU-C7OH** **159**
Figure A2.44 - Total repair gel of duplexes containing (**A**) **dFU-Et**, (**B**) **dFU-Bn**, (**C**) **dFU-C4OH** and (**D**) **dFU-C7OH** by hAGT, OGT, Ada-C, and hOGT for 2.5h at 37^{o}C **160**
Figure A2.45 - Time course repair gel of duplexes containing **dFU-Et** by (A) hAGT, (B) OGT, (C) Ada-C and (D) hOGT **161**
Figure A2.46 - Time course repair gel of duplexes containing **dFU-Bn** by (A) hAGT, (B) OGT, (C) Ada-C and (D) hOGT **162**
Figure A2.47 - Time course repair gel of duplexes containing **dFU-C4OH** by (A) hAGT, (B) OGT, (C) Ada-C and (D) hOGT **163**
Figure A2.48 - Time course repair gel of duplexes containing **dFU-C7OH** by (A) hAGT, (B) OGT, (C) Ada-C and (D) hOGT **164**
Figure A2.49 - Time course of repair graphs of **dFU-Et**, **dFU-Bn**, **dFU-C4OH** and **dFU-C7OH** at room temperature: **(A)** by hAGT, **(B)** by OGT, **(C)** by AdaC and **(D)** the chimera hOGT. **165**
Figure A2.50 - Repair in 3D bar graphs of **dFU-Et**, **dFU-Bn**, **dFU-C4OH** and **dFU-C7OH** by (**A**) hAGT, (**B**) OGT, (**C**) Ada-C, and (**D**) hOGT at 3 different time points. **166**

A2.1 Synthesis and characterization of nucleosides and oligonucleotides

A2.1.1 General

5-fluoro-2'-deoxyuridine was purchased from Berry Associates (Dexter, Michigan). "Fast deprotecting" 5'-*O*-dimethoxytrityl-2'-deoxyribonucleoside-3'-*O*-(β-cyanoethyl-N,N-diisopropyl)phosphoramidites and protected 2'-deoxyribonucleoside–CPG supports were purchased from Glen Research (Sterling, Virginia). Compound **1** was prepared according to previously published procedures. [54,97,98] All other chemicals and solvents were purchased from the Aldrich Chemical Company (Milwaukee, WI) or EMD Chemicals Inc. (Gibbstown, NJ). Flash column chromatography was performed using silica gel 60 (230–400 mesh) purchased from Silicycle (Quebec City, QC). Thin layer chromatography (TLC) was carried out with precoated TLC plates (Merck, Kieselgel 60 F_{254}, 0.25 mm) purchased from EMD Chemicals Inc. (Gibbstown, NJ). NMR spectra were recorded on a Varian 500 MHz NMR spectrometer at room temperature. ^{1}H NMR spectra were recorded at a frequency of 500.0 MHz and chemical shifts were reported in parts per million (ppm) downfield from tetramethylsilane. ^{13}C NMR spectra (^{1}H decoupled) were recorded at a frequency of 125.7 MHz and chemical shifts were reported in ppm with tetramethylsilane as a reference. ^{19}F NMR spectra were recorded at a frequency of 470.4 MHz and chemical shifts were reported in ppm with trichlorofluoromethane as a reference.^{31}P NMR spectra (^{1}H decoupled) were recorded at a frequency of 202.3 MHz and chemical shifts were reported in ppm with H_3PO_4 used as an external standard. High resolution mass spectrometry of modified nucleosides were obtained using an 7T-LTQ FT ICR instrument (Thermo Scientific), at the Concordia University Centre for Structural and Functional Genomics. The mass spectrometer was operated in full scan, positive ion detection mode. ESI mass spectra for oligonucleotides were obtained at the Concordia University Centre for Biological Applications of Mass Spectrometry (CBAMS) using a Micromass Qtof2 mass spectrometer (Waters) equipped with a nanospray ion source. The mass spectrometer was operated in full scan, negative ion detection mode. T4 polynucleotide kinase (PNK) was obtained from New England BioLabs (NEB). [γ-^{32}P]ATP was purchased from PerkinElmer (Woodbridge, ON). *BclI* restriction enzyme was obtained from New England Biolabs (Ipswich, MA).

5'-O-(4,4'-Dimethoxytrityl)-5-fluoro-O^4-ethyl-2'-deoxyuridine (2a): To a solution of triazole (0.28 g, 4.07 mmol) in anhydrous MeCN (5.6 mL) at 0°C under stirring, was added dropwise $POCl_3$ (0.084 ml, 0.91 mmol) followed by triethylamine (0.543 mL, 3.89 mmol). A solution of 3'-*O*-(*tert*-butyldimethylsilyl)-5'-O-(4,4'-dimethoxytrityl) -5-fluoro-2'-deoxyuridine (0.300g, 0.45mmol) in anhydrous MeCN (5.6mL) was added to the triazole solution. After 40 min, an additional solution of triazole (0.28 g, 4.07 mmol), $POCl_3$ (0.084 mL, 0.91 mmol) and triethylamine (0.543 mL, 3.89 mmol) in anhydrous MeCN (3 mL) was added dropwise with stirring at 0°C to the solution containing the nucleoside. This extra addition was repeated for a second time after 40 min. After 1 h, the solvent was evaporated *in vacuo*, the crude was taken up in EtOH (6 mL) and a solution of NaOEt (0.108 g, 1.1 mmol) in EtOH (6 mL) was added. After 30 min, an additional solution of NaOEt (0.108 g, 1.1 mmol) in EtOH (3 mL) was added to the nucleoside. After 16 h, the solvent was evaporated *in vacuo* and the crude product was taken up in CH_2Cl_2 (40 mL), washed with a 3% (w/v) solution of $NaHCO_3$ (2 x 50 ml), dried over anhydrous Na_2SO_4 (~ 4 g), and concentrated *in vacuo*. The crude product was taken up in THF (4.5 mL) and TBAF (1 M in THF) (0.543 mL, 0.54 mmol) was added drop-wise. After 30min, the solvent was evaporated *in vacuo* and the crude product was taken up in CH_2Cl_2 (40 mL), washed with a 3% (w/v) solution of $NaHCO_3$ (2 x 50 ml), dried over anhydrous Na_2SO_4 (~ 4 g), and concentrated *in vacuo*. The crude product was purified by flash column chromatography using a EtOAc/hexanes (6:4 →7:3) gradient as eluent to afford 0.154 mg (59%) of a colorless foam. R_f (SiO_2 TLC): 0.60 (100% EtOAc). $\lambda_{max(MeCN)}$ = 283 nm. HRMS (ESI-MS) *m/z* calculated for $C_{32}H_{33}FN_2NaO_7^+$: 599.2164; found 599.2165 $[M + Na]^+$. ^{1}H NMR (500 MHz, $CDCl_3$, ppm): 8.00-8.01 (d, J= 8.0 Hz, 1H; H6), 7.24–7.42 (m, 9H; Ar), 6.84– 6.86 (m, 4H; Ar), 6.22-6.25 (dd, J= 6.24 Hz, 1H; H1'), 4.49-4.55 (m, 3H; H3', CH_2), 4.08-4.15 (m, 1H; H4'), 3.80 (s, 6H; $ArOCH_3$), 3.41-3.48 (m, 2H; H5', H5''), 2.64-2.69 (m, 1H; H2'), 2.24-2.29 (m, 1H; H2''), 1.41-1.44 (t, J= 7.4 Hz, 3H; CH_3). ^{13}C NMR (125.7 MHz, $CDCl_3$, ppm): 158.65, 135.37, 135.25, 129.98, 128.01, 127.94, 127.05, 113.31, 87.07, 86.72, 86.20, 77.25, 76.99, 76.74, 71.53, 64.22, 62.91, 58.23, 41.83, 14.11. ^{19}F (470.4 MHz, $CDCl_3$, ppm): -168.43, -168.44. IR (thin film); ν_{max} (cm^{-1}) = 3420, 3054, 2966, 2837, 2362, 2331, 1675, 1652, 1549, 1495, 1457, 1385, 1332, 1250, 1177, 1093, 1034, 828, 736.

5'-O-(4,4'-Dimethoxytrityl)-5-fluoro-O^4-benzyl-2'-deoxyuridine (2b): To a solution of triazole (0.188 g, 2.72 mmol) in anhydrous MeCN (3 mL) at 0°C under stirring, was added dropwise $POCl_3$ (0.056 mL, 0.60 mmol) followed by triethylamine (0.362 mL, 2.6 mmol). A solution of 3'-*O*-(*tert*-butyldimethylsilyl)-5'-*O*-(4,4'-dimethoxytrityl)-5-fluoro-2'-deoxyuridine (0.200 g, 0.30 mmol) in anhydrous MeCN (3 mL) was added to the triazole solution. After 40 min, an additional solution of triazole (0.188 g, 2.72 mmol), $POCl_3$ (0.056 mL, 0.60 mmol) and triethylamine (0.362 mL, 2.6 mmol) in anhydrous MeCN (2 mL) was added dropwise with stirring at 0°C to the solution containing the nucleoside. This extra addition was repeated for a second time after 40 min. After 1 h, the solvent was evaporated *in vacuo* and the crude was taken up in MeCN (4 mL). BnOH (0.320 mL, 3.1 mmol) was added dropwise followed by the addition of DBU (0.104 mL, 1.02 mmol). After 16 h, the solvent was evaporated *in vacuo* and the crude product was taken up in CH_2Cl_2 (40 mL), washed with a 3% (w/v) solution of $NaHCO_3$ (2 x 50 mL), dried over anhydrous Na_2SO_4 (~ 4 g), and concentrated *in vacuo*. The crude product was taken up in THF (4 mL) and TBAF (1 M in THF) (0.36 mL, 0.36 mmol) was added drop-wise. After 30min, the solvent was evaporated *in vacuo* and the crude product was taken up in CH_2Cl_2 (40 mL), washed with a 3% (w/v) solution of $NaHCO_3$ (2 x 50 ml), dried over anhydrous Na_2SO_4 (~ 4g), and concentrated *in vacuo*. The crude product was purified by flash column chromatography using CH_2Cl_2/MeOH (49:1) as eluent to afford 0.111 mg (58%) of a colorless foam. R_f (SiO_2 TLC): 0.66 (CH_2Cl_2/MeOH (47.5:2.5)). $\lambda_{max(MeCN)}$ = 283 nm. HRMS (ESI-MS) *m/z* calculated for $C_{37}H_{35}FN_2NaO_7^+$: 661.2321; found 661.2322 $[M + Na]^+$. ^{1}H NMR (500 MHz, $CDCl_3$, ppm): 8.06-8.07 (d, J= 8.06 Hz, 1H; H6), 7.18–7.47 (m, 15H; Ar, Bn), 6.83– 6.86 (m, 4H; Ar), 6.26-6.28 (dd, J= 6.27 Hz, 1H; H1'), 5.49 (s, 2H; OCH_2 Bn) 4.53-4.55 (m, 1H; H3'), 4.13-4.15 (m, 1H; H4'), 3.79 (s, 6H; $ArOCH_3$), 3.41-3.47 (m, 2H; H5', H5''), 2.70-2.76 (m, 1H; H2'), 2.25-2.31 (m, 1H; H2''). ^{13}C NMR (125.7 MHz, $CDCl_3$, ppm): 162.12, 162.01, 158.66, 153.36, 144.29, 137.56, 135.59, 135.37, 135.23, 134.80, 129.98, 129.96, 129.12, 128.62, 128.57, 127.92, 127.58, 127.04, 113.33, 113.30, 113.17, 87.08, 86.45, 85.98, 77.26, 77.01, 76.75, 71.63, 69.50, 63.00, 55.23, 41.96, 0.01. ^{19}F (470.4 MHz, $CDCl_3$, ppm): -168.09, -168.10. IR (thin film); ν_{max} (cm^{-1}) = 3384, 2929, 2836, 2361, 2337, 1683, 1645, 1540, 1507, 1487, 1456, 1360, 1251, 1177, 1094, 1034, 828, 735, 699.

3'-*O*-(2-Cyanoethoxy(diisopropylamino)-phosphino)-5'-*O*-(4,4'-dimethoxytrityl)- 5-fluoro-O^4-ethyl- 2'-deoxyuridine (3a) : DIPEA (0.14 mL, 0.80 mmol) was added to a solution of (2a) (0.154 g, 0.266 mmol) in THF (2.7 mL), followed by the dropwise addition of N,N-diisopropylamino cyanoethyl phosphonamidic chloride (0.14 mL, 0.64 mmol). After 30 min, the solvent was evaporated *in vacuo*, the crude product was taken up in ethyl acetate (40 mL), the solution washed with 3% (w/v) solution of $NaHCO_3$ (2 x 50 mL) followed by brine (50 mL). The organic layer was dried over anhydrous Na_2SO_4 (~ 4 g), and concentrated *in vacuo*. The crude product was purified by flash column chromatography using a hexanes/ethyl acetate (2:8) solvent system to afford 0.163 g (79%) of a colorless foam. R_f (SiO_2 TLC): 0.18, 0.36 (1:1 Hexanes/EtOAc). $\lambda_{max(MeCN)}$ = 283 nm. HRMS (ESI-MS) *m/z* calculated for $C_{41}H_{50}FN_4NaO_8P^+$: 799.3243; found 799.3238 $[M + Na]^+$. ^{1}H NMR (500 MHz, $CDCl_3$, ppm): 8.11-8.15 (m, 1H; H6), 7.24–7.54 (m, 9H; Ar), 6.90– 6.93 (m, 4H; Ar), 6.19-6.24 (m, 1H; H1'), 4.70-4.78 (m, 1H; H3'), 4.42-4.47 (m, 2H; CH_2CH_3), 4.22-4.29 (m, 1H; H4'), 3.61-3.94 (m, 10H; 2 x NCH, 2 x $ArOCH_3$, CH_2OP), 3.43-3.52 (m, 2H; H5', H5''), 2.76-2.81 (t, J= 6.0 Hz, 2H; CH_2CN), 2.61-2.70 (m, 4H; H2',CH_2CH_3), 2.39-2.46 (m, 1H; H2''), 1.20-1.23 (m, 9H; CH_3), 1.12-1.14 (m, 3H; CH_3). ^{13}C NMR (125.7 MHz, $CDCl_3$, ppm): 161.91, 161.81, 158.83, 152.42, 144.93, 137.11, 135.62, 135.47, 135.20, 130.11, 128.06, 127.88, 127.81, 127.63, 127.59, 126.84, 126.75, 118.06, 117.96, 113.11, 86.69, 86.65, 86.55, 86.47, 85.68, 85.47, 85.44, 85.39, 73.23, 73.09, 72.76, 72.63, 63.37, 63.03, 62.83, 58.72, 58.68, 58.57, 58.52, 54.64, 54.62, 43.12, 43.11, 43.02, 43.01, 40.36, 40.33, 40.17, 40.14, 24.04, 24.01, 23.98, 23.95, 23.93, 19.89, 19.86, 19.83, 19.80, 13.44. ^{19}F (470.4 MHz, d_6-acetone, ppm): -171.28, -171.29, -171.34, -171.35. ^{31}P NMR (202.3 MHz, d_6-acetone, ppm): 148.34, 148.21. IR (thin film); ν_{max} (cm^{-1}) = 3421, 2966, 2930, 2361, 2336, 1684, 1652, 1558, 1540, 1506, 1495, 1457, 1251, 1178, 1034, 977, 828.

3'-*O*-(2-Cyanoethoxy(diisopropylamino)-phosphino)-5'-*O*-(4,4'-dimethoxytrityl)-5-fluoro-O^4-benzyl-2'-deoxyuridine (3b): DIPEA (0.091 mL, 0.52 mmol) was added to a solution of (2b) (0.111 g, 0.174 mmol) in THF (2 mL), followed by the dropwise addition of N,N-diisopropylamino cyanoethyl phosphonamidic chloride (0.092 mL, 0.42 mmol). After 30 min, the solvent was evaporated *in vacuo*, the crude product was taken up in ethyl acetate (40 mL), the solution was washed with a 3% (w/v) solution of $NaHCO_3$ (2 x 50 mL) and once with brine

(50 mL). The organic layer was dried over anhydrous Na_2SO_4 (~ 4 g), and concentrated *in vacuo*. The crude product was purified by flash column chromatography using hexanes/ethyl acetate (2:8) as eluent to afford 0.102 g (70%) of a colorless foam. R_f (SiO_2 TLC): 0.36, 0.51 (1:1 Hexanes/EtOAc). $\lambda_{max(MeCN)}$ = 283 nm. HRMS (ESI-MS) *m/z* calculated for $C_{46}H_{52}FN_4NaO_8P^+$: 861.3399; found: 861.3395 $[M + Na]^+$. ^1H NMR (500 MHz, d_6-acetone, ppm): 8.15-8.20 (m, 1H; H6), 7.21–7.54 (m, 14H; Ar, Bn), 6.89– 6.93 (m, 4H; Ar), 6.19-6.25 (m, 1H; H1'), 5.46 (s, 2H; CH_2Bn), 4.71-4.79 (m, 1H; H3'), 4.23-4.30 (m, 1H; H4'), 3.62-3.94 (m, 10H; 2 x NCH, 2 x $ArOCH_3$, CH_2OP), 3.43-3.53 (m, 2H; H5', H5''), 2.78-2.80 (t, J= 6.0 Hz, 1H; CH_2CN), 2.62-2.72 (m, 2H; H2', CH_2CN), 2.41-2.49 (m, 1H; H2''), 1.20-1.23 (m, 9H; 3 x CH_3), 1.13-1.14 (m, 3H; CH_3). ^{13}C NMR (125.7 MHz, $CDCl_3$, ppm): 161.72, 161.57, 158.87, 152.26, 144.94, 144.93, 137.05, 137.03, 135.61, 135.60, 135.58, 135.49, 135.45, 135.11, 135.09, 130.11, 130.09, 130.08, 128.57, 128.49, 128.38, 128.29, 128.27, 128.04, 128.01, 127.82, 126.81, 126.78, 113.13, 113.11, 86.71, 86.66, 86.58, 85.77, 85.74, 85.49, 85.45, 73.21, 73.08, 72.75, 72.62, 68.86, 63.01, 62.81, 58.73, 58.68, 58.58, 58.53, 54.64, 54.62, 43.13, 43.11, 43.03, 43.02, 40.39, 40.36, 40.20, 40.16, 24.04, 23.99, 23.93, 19.90, 19.84, 19.81. ^{19}F (470.4 MHz, d_6-acetone, ppm): -171.22, -171.23, -171.28, -171.29. ^{31}P NMR (202.3 MHz, d_6-acetone, ppm): 148.36, 148.23. IR (thin film); ν_{max} (cm^{-1}) = 3423, 2361, 2336, 1652, 1506, 1456, 1251, 1178, 1035, 735.

5'-*O*-(4,4'-Dimethoxytrityl)-5-fluoro-*O*^{4}-(phenoxyacetyloxybutyl)-2'-deoxyuridine (4a): A convertible nucleoside was prepared by slowly adding triethylamine (1.45 mL, 10.4 mmol) followed by dropwise addition of $POCl_3$ (0.105 mL, 1.13 mmol) to a solution containing triazole (0.703 g, 10.8 mmol) and compound 1 in anhydrous MeCN (4.5 mL) at 0^{o}C under stirring. After 2 hours the reaction was over and the solvent was evaporated *in vacuo*. To a solution of convertible nucleoside (0.32 g, 0.45 mmol) and 1,4-butanediol (0.29 g, 3.2 mmol) in MeCN (3 mL) was added dropwise DBU (0.15 mL, 1.0 mmol). After 16 h, the solvent was evaporated *in vacuo* and the crude product was taken up in CH_2Cl_2 (40 mL), washed with a 3% (w/v) solution of $NaHCO_3$ (2 x 50 mL), dried over anhydrous Na_2SO_4 (~ 4 g), and concentrated *in vacuo*. To a solution of this intermediate and triethylamine (0.11 mL, 0.81 mmol) in THF (5 mL) was added dropwise Pac-Cl (0.094 mL, 0.68 mmol) while stirring at 0^{o}C. After 30 min, the solvent was evaporated *in vacuo* and the crude product was taken up in CH_2Cl_2 (40 mL), washed with a 3%

(w/v) solution of $NaHCO_3$ (2 x 50 mL), dried over anhydrous Na_2SO_4 (~ 4 g), and concentrated *in vacuo*. The resulting residue was taken up in THF (5 mL) and TBAF (1 M in THF) (0.68 mL, 0.68 mmol) was added dropwise. After 30 min, the solvent was evaporated *in vacuo* and the crude product was taken up in CH_2Cl_2 (40 mL), washed with a 3% (w/v) solution of $NaHCO_3$ (2 x 50 mL), dried over anhydrous Na_2SO_4 (~ 4 g), and concentrated *in vacuo*. The crude product was purified by flash column chromatography using EtOAc / Hex (4 : 1) as eluent to afford 0.16 g (47 % over 3 steps) of a colorless foam. R_f (SiO_2 TLC): 0.32 (100% EtOAc). $\lambda_{max(MeCN)}$ = 276 nm. HRMS (ESI-MS) *m/z* calculated for $C_{42}H_{43}FN_2NaO_{10}^+$: 777.2794; found 777.2807 $[M + Na]^+$. ^1H NMR (500 MHz, $CDCl_3$, ppm): 8.03 (d, 1H; H6, J = 5.5 Hz), 7.41–7.38 (m, 2H; Ar), 7.32-7.20 (m, 9H; Ar), 6.98 (m, 1H; Ar), 6.92– 6.89 (m, 2H; Ar), 6.85-6.82 (m, 4H; Ar), 6.24 (m, 1H; H1'), 4.64 (s, 2H; $COCH_2OPh$), 4.53 (m, 1H; H3'), 4.44-4.42 (m, 2H; $ArOCH_2$), 4.26-4.24 (m, 2H; $COOCH_2$), 4.12 (m, 1H; H4'), 3.78 (2 s, 6H; 2 x OCH_3), 3.41-3.42 (m, 2H; H5', H5''), 2.72-2.67 (m, 2H; H2', OH), 2.24 (m, 1H; H2''), 1.81-1.77 (m, 4H; CH_2CH_2). ^{13}C NMR (125.7 MHz, $CDCl_3$, ppm): 171.14, 168.99, 162.27, 162.17, 158.64, 157.75, 153.43, 144.30, 137.56, 135.59, 135.38, 135.23, 129.97, 129.95, 129.55, 127.99, 127.92, 127.67, 127.42, 127.02, 121.74, 114.60, 113.31, 113.29, 87.04, 86.95, 86.45, 71.61, 67.45, 65.29, 64.64, 63.03, 60.38, 55.22, 41.95, 25.08, 24.95, 21.03, 14.18. ^{19}F (470.4 MHz, $CDCl_3$, ppm): -168.40, -168.41. IR (thin film); ν_{max} (cm^{-1}) = 3391, 3066, 2956, 2931, 2837, 2362, 2336, 1760, 1679, 1542, 1510, 1492, 1458, 1379, 1251, 1176, 1090, 1034, 829, 755.

5'-*O*-(4,4'-Dimethoxytrityl)-5-fluoro-*O*^{4}-(phenoxyacetyloxyheptyl)-2'-deoxyuridine (4b): A convertible nucleoside was prepared by slowly adding triethylamine (1.45 mL, 10.4 mmol) followed by dropwise addition of $POCl_3$ (0.105 mL, 1.13 mmol) to a solution containing triazole (0.703 g, 10.8 mmol) and compound 1 in anhydrous MeCN (4.5 mL) at 0^{o}C under stirring. After 2 h the reaction was over and the solvent was evaporated *in vacuo*. To a solution of convertible nucleoside (0.32 g, 0.45 mmol) and 1,7-heptanediol (0.072 g, 0.54 mmol) in MeCN (6 mL) was added dropwise DBU (0.088 mL, 0.59 mmol). After 16 h, the solvent was evaporated *in vacuo* and the crude product was taken up in CH_2Cl_2 (40 mL), washed with a 3% (w/v) solution of $NaHCO_3$ (2 x 50 ml), dried over anhydrous Na_2SO_4 (~ 4 g), and concentrated *in vacuo*. To a solution of the intermediate and triethylamine (0.11 mL, 0.81 mmol) in THF (5 mL) was added dropwise Pac-Cl (0.094 mL, 0.68 mmol) while stirring at 0^{o}C. After 30 min, the solvent was

evaporated *in vacuo* and the crude product was taken up in CH_2Cl_2 (40 mL), washed with a 3% (w/v) solution of $NaHCO_3$ (2 x 50 ml), dried over anhydrous Na_2SO_4 (~ 4 g), and concentrated *in vacuo*. The resulting residue was taken up in THF (5 mL) and TBAF (1 M in THF) (0.68 mL, 0.68 mmol) was added dropwise. After 30 min, the solvent was evaporated *in vacuo* and the crude product was taken up in CH_2Cl_2 (40 mL), washed with a 3% (w/v) solution of $NaHCO_3$ (2 x 50 ml), dried over anhydrous Na_2SO_4 (~ 4 g), and concentrated *in vacuo*. The crude product was purified by flash column chromatography using EtOAc / Hex (7 : 3) as eluent to afford 0.084 g (23 % over 3 steps) of a colorless foam. R_f(SiO_2 TLC): 0.54 (100% EtOAc). $\lambda_{max(MeCN)}$ = 276 nm. HRMS (ESI-MS) *m/z* calculated for $C_{45}H_{49}FN_2NaO_{10}^+$ 819.3263; found 819.3266 $[M + Na]^+$. ^1H NMR (500 MHz, $CDCl_3$, ppm): 8.01 (d, 1H; H6, J = 6 Hz), 7.41–7.38 (m, 2H; Ar), 7.30-7.20 (m, 9H; Ar), 6.98 (m, 1H; Ar), 6.92– 6.89 (m, 2H; Ar), 6.85-6.82 (m, 4H; Ar), 6.24 (m, 1H; H1'), 4.63 (s, 2H; $COCH_2OPh$), 4.51 (m, 1H; H3'), 4.43-4.41 (m, 2H; $ArOCH_2$), 4.20 (t, 2H; $COOCH_2$, J = 7 Hz), 4.11 (m, 1H; H4'), 3.78 (2 s, 6H; 2 x OCH_3), 3.41-3.42 (m, 2H; H5', H5''), 2.68 (m, 1H; H2'), 2.41 (bs, 1H; OH), 2.24 (m, 1H; H2''), 1.76 (m, 2H; CH_2), 1.64 (m, 2H; CH_2), 1.39-1.32 (m, 6H; $CH_2CH_2CH_2$). ^{13}C NMR (125.7 MHz, $CDCl_3$, ppm): 169.07, 162.45, 162.35, 158.64, 157.80, 153.48, 144.30, 137.65, 135.68, 135.38, 135.24, 129.97, 129.95, 129.52, 127.99, 127.92, 127.41, 127.16, 127.02, 121.69, 114.60, 113.31, 113.29, 87.04, 86.84, 86.33, 71.60, 68.21, 65.35, 65.29, 62.99, 55.22, 41.91, 28.76, 28.43, 28.26, 25.64, 25.60. ^{19}F (470.4 MHz, $CDCl_3$, ppm): -168.19, -168.20. IR (thin film); ν_{max} (cm^{-1}) = 3394, 3065, 2935, 2836, 2362, 2336, 1760, 1679, 1607, 1542, 1510, 1492, 1458, 1377, 1251, 1176, 1090, 1034, 829, 755.

3'-*O*-(2-Cyanoethoxy(diisopropylamino)-phosphino)-5'-*O*-(4,4'-dimethoxytrityl)-5-fluoro-O^4-(phenoxyacetyloxybutyl)-2'-deoxyuridine (5a): DIPEA (52 µL, 0.30 mmol) was added to a solution of (**4a**) (0.075 g, 0.098 mmol) in THF (1.5 mL), followed by the dropwise addition of N,N-diisopropylamino cyanoethyl phosphonamidic chloride (53 µL, 0.24 mmol). After 30 min, the solvent was evaporated *in vacuo*, the crude product was taken up in ethyl acetate (40 mL), the solution was washed with 3% (w/v) solution of $NaHCO_3$ (2 x 50 ml) and once with brine (50 mL). The organic layer was dried over anhydrous Na_2SO_4 (~ 4 g), and reconcentrated *in vacuo*. The crude product was purified by flash column chromatography using an ethyl acetate solvent system to afford 0.80 g (82%) of a colorless foam. R_f (SiO_2 TLC): 0.73, 0.85 (EtOAc). $\lambda_{max(MeCN)}$ =

276 nm. HRMS (ESI-MS) *m/z* calculated for $C_{51}H_{60}FN_4NaO_{11}P^+$ 977.3872; found 977.3868 [M + Na]$^+$. ^{1}H NMR (500 MHz, d_6-acetone, ppm): 8.16-8.11 (m, 1H, H6), 7.52–7.49 (m,2H; Ar), 7.38-7.24 (m, 9H; Ar), 6.95-6.88 (m, 7H; Ar), 6.23-6.18 (m, 1H; H1'), 4.77-4.70 (m, 3H; H3', $COCH_2OPh$), 4.43-4.31 (m, 2H; $ArOCH_2$), 4.28-4.21 (m, 3H; H4', $COOCH_2$), 3.93-3.60 (m, 10H; $POCH_2$, 2 x NCH, 2 x OCH_3), 3.51-3.42 (m, 2H; H5', H5''), 2.78-2.75 (m, 1H; CHCN), 2.70-2.59 (m, 2H; CHCN, H2'), 2.46-2.38 (m, 1H; H2''), 1.87-1.77 (m, 4H; CH_2CH_2), 1.31-1.29 (m, 3H; CH_3), 1.20-1.19 (m, 6H; 2 x CH_3), 1.13-1.11 (m, 3H; CH_3). Residual EtOAc solvent peaks are observed. ^{13}C NMR (125.7 MHz, $CDCl_3$, ppm): 169.98, 168.59, 161.93, 161.83, 158.82, 158.20, 152.36, 144.93, 144.92, 137.12, 137.09, 135.61, 135.58, 135.49, 135.46, 135.17, 135.15, 130.11, 130.09, 129.39, 128.05, 128.01, 127.97, 127.94, 127.82, 127.72, 127.69, 126.82, 126.79, 121.19, 118.07, 117.95, 114.51, 113.13, 113.12, 86.70, 86.65, 86.58, 86.50, 85.73, 85.69, 85.44, 85.39, 73.18, 73.05, 72.74, 72.61, 66.91, 64.78, 64.14, 63.00, 62.81, 61.90, 61.86, 59.64, 58.71, 58.66, 58.56, 58.51, 54.66, 54.64, 47.31, 47.27, 43.12, 43.11, 43.02, 43.01, 40.37, 40.35, 40.18, 40.14, 24.95, 24.84, 24.05, 24.02, 24.00, 23.96, 23.94, 21.92, 21.90, 20.43, 20.41, 19.95, 19.90, 19.86, 19.84, 19.81, 18.95, 18.88, 13.63.^{19}F (470.4 MHz, d_6-acetone, ppm): -171.19, -171.21, -171.25, -171.26. ^{31}P NMR (202.3 MHz, d_6-acetone, ppm): 148.32, 148.19. IR (thin film); ν_{max} (cm^{-1}) = 3069, 2968, 2831, 2872, 2252, 1760, 1683, 1648, 1542, 1510, 1492, 1459, 1409, 1364, 1252, 1180, 1087, 1035, 977, 830, 755.

3'-*O*-(2-Cyanoethoxy(diisopropylamino)-phosphino)-5'-*O*-(4,4'-dimethoxytrityl)-5-fluoro-*O*^{4}-(phenoxyacetyloxyheptyl)-2'-deoxyuridine (5b): DIPEA (49 µL, 0.28 mmol) was added to a solution of (**4b**) (0.075 g, 0.094 mmol) in THF (1.5 mL), followed by the dropwise addition of N,N-diisopropylamino cyanoethyl phosphonamidic chloride (50 µL, 0.23 mmol). After 30 min, the solvent was evaporated *in vacuo*, the crude product was taken up in ethyl acetate (40 mL), the solution was washed with 3% (w/v) solution of $NaHCO_3$ (2 x 50 ml) and once with brine (50 mL). The organic layer was dried over anhydrous Na_2SO_4 (~ 4 g), and concentrated *in vacuo*. The crude product was purified by flash column chromatography using ethyl acetate as eluent to afford 0.74 g (78%) of a colorless foam. R_f (SiO_2 TLC): 0.83, 0.90 (EtOAc). $\lambda_{max(MeCN)}$ = 277 nm. HRMS (ESI-MS) *m/z* calculated for $C_{54}H_{66}FN_4NaO_{11}P^+$ 1019.4342; found 1019.4332 [M + Na]$^+$. ^{1}H NMR (500 MHz, d_6-acetone, ppm): 8.15-8.10 (m, 1H, H6), 7.52–7.48 (m, 2H; Ar), 7.41-7.22 (m,

9H; Ar), 6.97-6.88 (m, 7H; Ar), 6.23-6.18 (m, 1H; H1'), 4.77-4.70 (m, 3H; H3', $COCH_2OPh$), 4.39-4.32 (m, 2H; $ArOCH_2$), 4.28-4.21 (m, 1H; H4'), 4.18-4.16 (m, 2H; $COOCH_2$), 3.93-3.60 (m, 10H; $POCH_2$, 2 x NCH, 2 x OCH_3), 3.51-3.42 (m, 2H; H5', H5''), 2.78-2.75 (m, 1H; CHCN), 2.70-2.59 (m, 2H; CHCN, H2'), 2.46-2.38 (m, 1H; H2''), 1.80-1.74 (m, 2H; CH_2), 1.68-1.62 (m, 2H; CH_2), 1.46-1.35 (m, 6H; $CH_2CH_2CH_2$), 1.32-1.29 (m, 3H; CH_3), 1.21-1.19 (m, 6H; 2 x CH_3), 1.13-1.11 (m, 3H; CH_3). Signals for residual EtOAc solvent peaks are also observed. ^{13}C NMR (125.7 MHz, $CDCl_3$, ppm): 169.97, 168.62, 162.03, 161.92, 158.81, 158.21, 152.39, 144.93, 144.92, 137.15, 137.12, 135.61, 135.58, 135.50, 135.46, 135.21, 135.18, 130.11, 130.08, 129.39, 128.05, 128.01, 127.89, 127.86, 127.82, 127.64, 127.61, 126.81, 126.78, 121.19, 118.06, 117.94, 114.48, 113.13, 113.11, 86.69, 86.64, 86.56, 86.48, 85.72, 85.68, 85.43, 85.38, 73.20, 73.07, 72.75, 72.62, 67.40, 64.80, 64.53, 63.01, 62.82, 61.90, 61.86, 59.64, 58.71, 58.66, 58.56, 58.51, 54.65, 54.63, 47.31, 47.27, 43.12, 43.10, 43.02, 43.01, 40.36, 40.34, 40.17, 40.14, 25.50, 25.49, 24.05, 24.02, 23.99, 23.96, 23.94, 21.92, 21.90, 20.42, 20.41, 19.95, 19.89, 19.86, 19.84, 19.80, 18.95, 18.88, 13.63. ^{19}F (470.4 MHz, d_6-acetone, ppm): -171.20, -171.22, -171.26, -171.28. ^{31}P NMR (202.3 MHz, d_6-acetone, ppm): 148.32, 148.19. IR (thin film); ν_{max} (cm^{-1}) = 3068, 2966, 2834, 2864, 2252, 1759, 1683, 1647, 1607, 1541, 1510, 1492, 1458, 1409, 1365, 1252, 1179, 1086, 1035, 978, 829, 755.

A2.1.2 Preparation, purification and characterization of the dFU oligonucleotides

The sequences containing the different dFU adducts, which are shown in **Figure 3.1**, were assembled with an Applied Biosystems Model 3400 synthesizer on a 1.5 μmol scale using β-cyanoethylphosphoramidite chemistry supplied by the manufacturer with slight modifications to coupling times. The nucleoside phosphoramidites protected with "fast-deprotecting" groups were dissolved in anhydrous MeCN at a concentration of 0.1 M for the 3'-*O*-deoxyphosphoramidites, and 0.15 M for the modified phosphoramidites (**3a**, **3b**, **5a** and **5b**). Oligomer sequence assembly was carried out as previously described.[104] The capping step of the assembly was carried out using phenoxyacetic anhydride/pyridine/tetrahydrofuran 1:1:8 (v/v/v; solution A) and 1-methyl-imidazole/ tetrahydrofuran 16:84 (w/v; solution B). Coupling wait times for phosphoramidites **3a, 3b, 5a** and **5b** were extended to 10 min (compared to 2

min for the commercially available phosphoramidites). Protecting group removal and cleavage from the solid support was carried out by treatment with 10% DBU in their corresponding alcohol (ethanol for the ethyl adduct, benzyl alcohol for the benzyl adduct, etc.) overnight at 50 ^{o}C in 2 mL screw-cap microfuge tubes fitted with Teflon lined caps.[54] The base was neutralized with an equimolar amount of AcOH and crude oligomers were transferred and lyophilized in a Speedvac concentrator. Purification was achieved by strong anion exchange HPLC using a Dionex DNAPAC PA-100 column (0.4 cm x 25 cm) purchased from Dionex Corp, Sunnyvale, CA using a linear gradient of 0–52% buffer B over 24 min (buffer A: 100 mm Tris HCl, pH 7.5, 10% MeCN and buffer B: 100 mm Tris HCl, pH 7.5, 10% MeCN, 1M NaCl) at 55 ^{o}C. The columns were monitored at 260 nm for analytical runs or 280 nm for preparative runs. The purified oligomers were desalted using C-18 SEP PAK cartridges (Waters Inc.) as previously described.[132] The molecular mass of the modified oligomers was identified by deconvolution of the ESI-MS and the measured values were in agreement with the expected masses (**Figures A2.30, A2.31, A2.32** and **A2.33** for MS spectra). Oligonucleotides **dFU-Et**, **dFU-Bn**, **dFU-C4OH** and **dFU-C7OH** (0.05 A_{260} units each) were also characterized by enzymatic digestion (snake venom phosphodiesterase: 0.28 units and calf intestinal phosphatase: 5 units, in 10 mM Tris, pH 8.1 and 2 mM $MgCl_2$) for 48 h at 37 ^{o}C. The resulting nucleoside mixture was analyzed by reversed phase HPLC carried out using a Symmetry® C-18 5 µm column (0.46 x 15 cm) via a linear gradient of 0-70% buffer B over 30 min (buffer A: 50 mM sodium phosphate, pH 5.8, 2% MeCN and buffer B: 50 mM sodium phosphate, pH 5.8, 50% MeCN). Results and nucleoside ratios from digestion are shown in **Figures A2.34, A2.35, A2.36** and **A2.37.** The molecular mass of the modified oligomers was identified by deconvolution of the ESI-MS and the measured values were in agreement with the expected masses

A2.2 UV thermal denaturation

Please refer to section A1.2 for the experimental details. The results obtained are shown in **Figure A2.42**.

A2.3 Circular dichroism (CD) spectroscopy

Please refer to section A1.3 for the experimental details. The results obtained are shown in **Figure A2.43**.

A2.4 AGT repair assay of DNA duplexes

Please refer to section A1.4 for the experimental details.

Figure A2.1: (**A**) 2'-deoxy-2'-fluoro-arabinonucleic acid (F-ANA),[71–73] (**B**) 5-fluorouracil,[67,135] (**C**) Capecitabine[67,135] and (**D**) Fluorobenzene.[56]

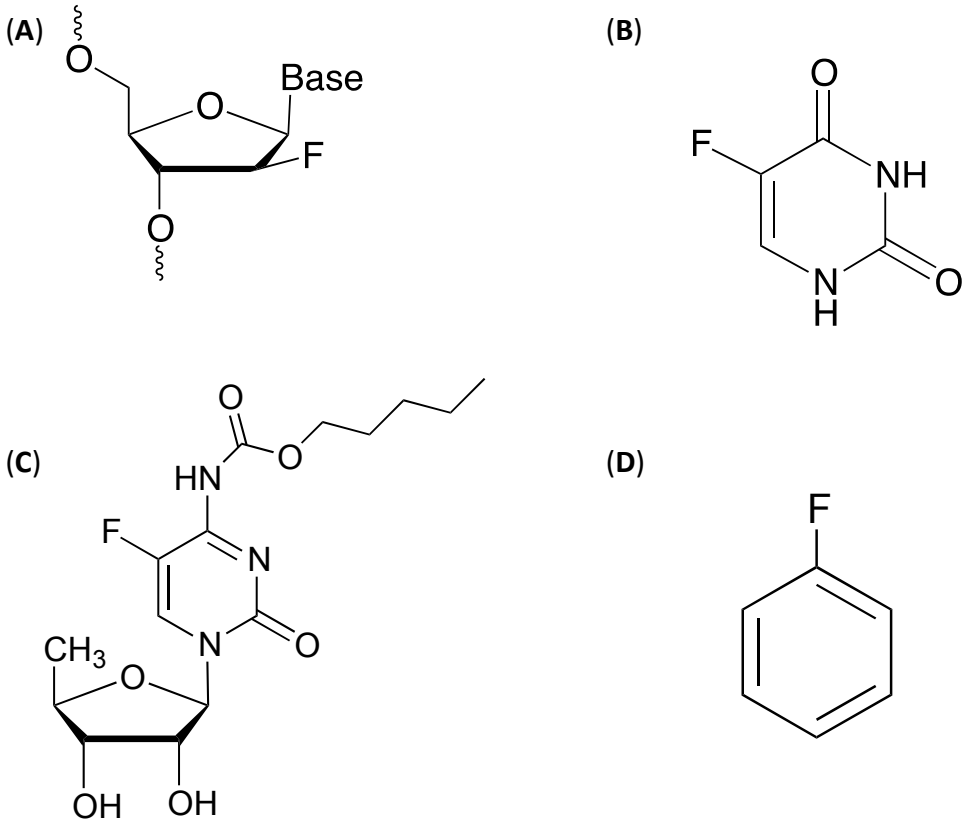


Figure A2.2 - 500 MHz ^{1}H NMR spectrum of compound (**2a**) (in $CDCl_3$)

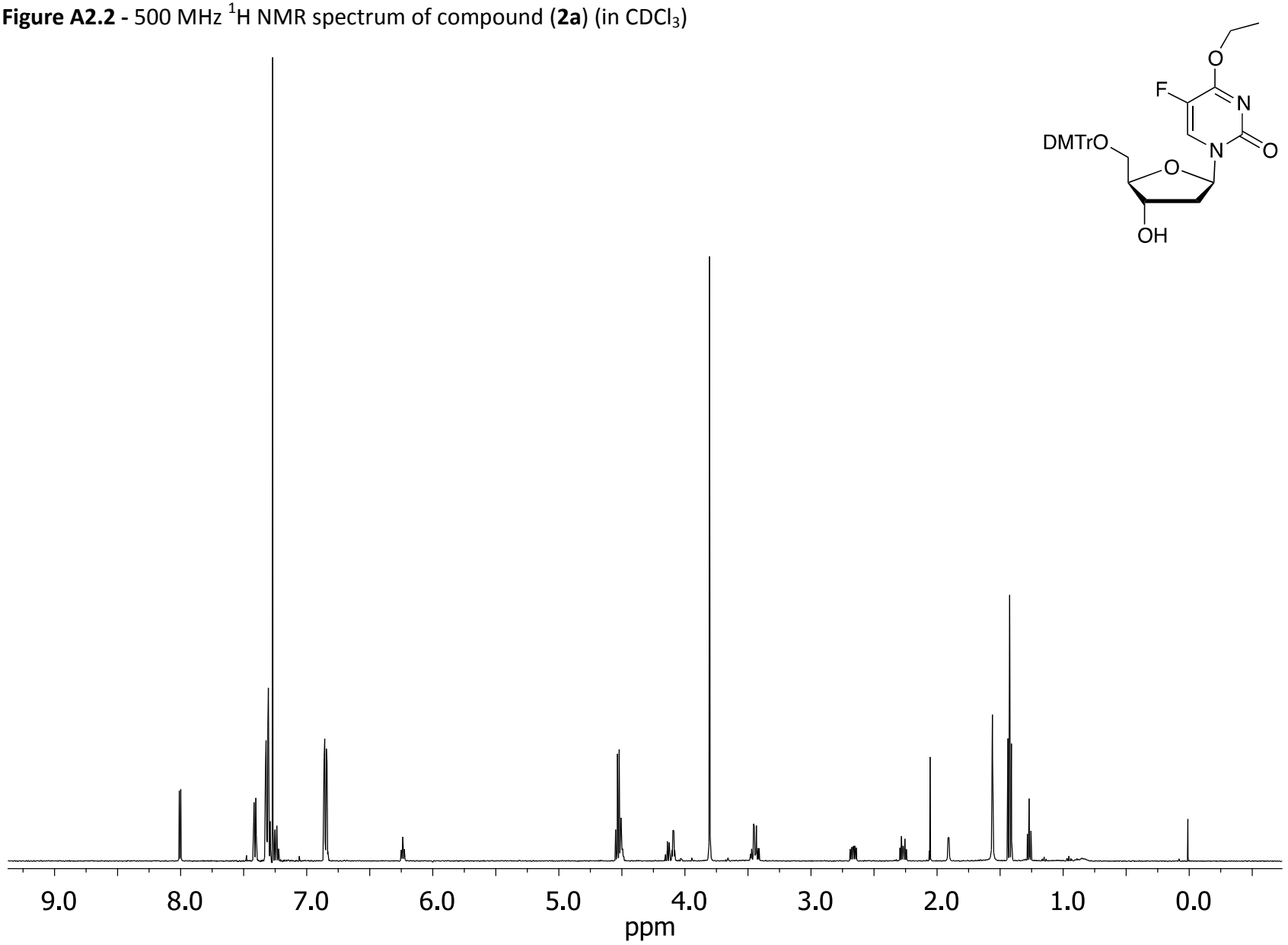


Figure A2.3 - 125.7 MHz ^{13}C NMR spectrum of compound (**2a**) (in $CDCl_3$)

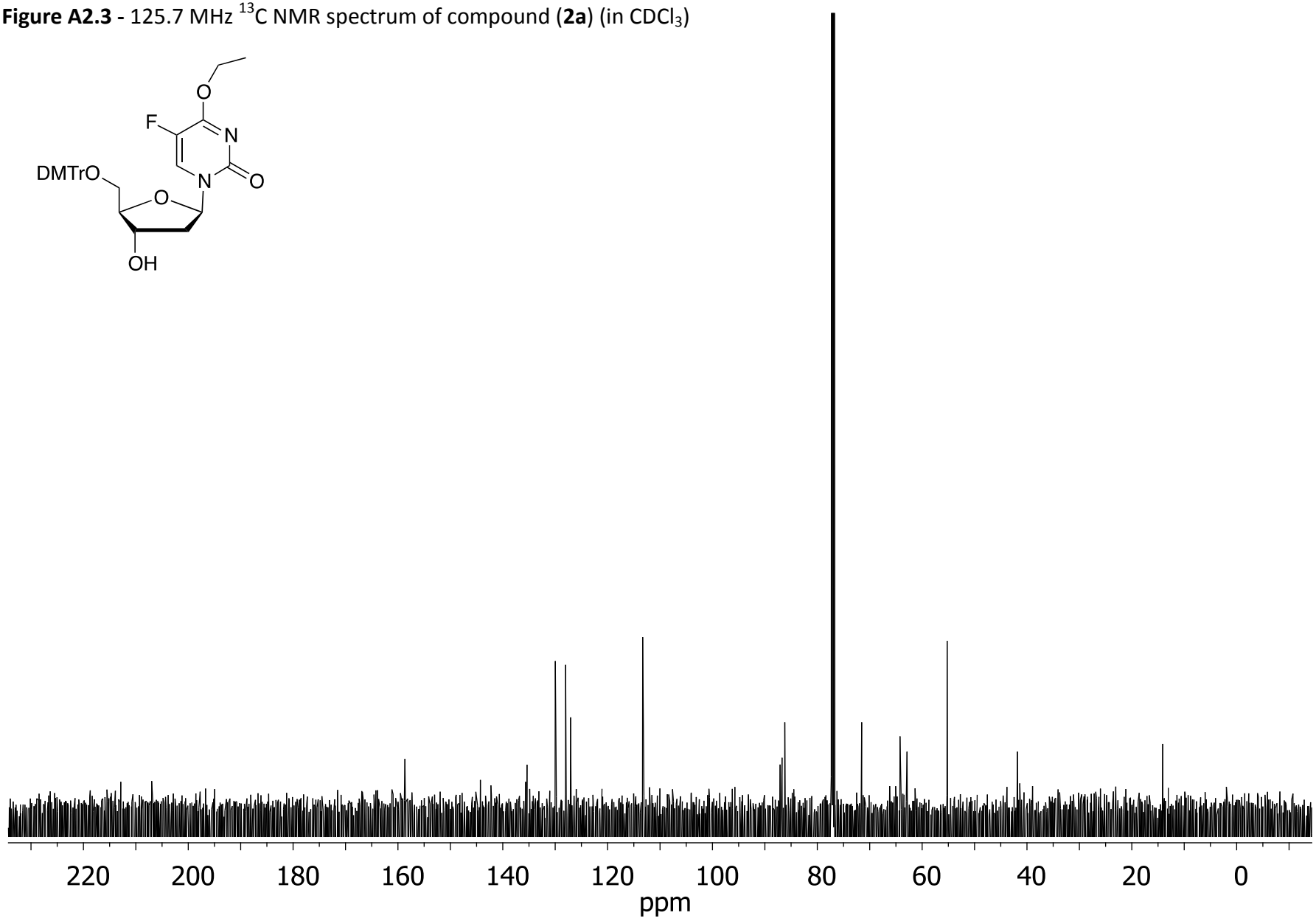


Figure A2.4 - 470.4 MHz ^{19}F NMR spectrum of compound (**2a**) (in $CDCl_3$)

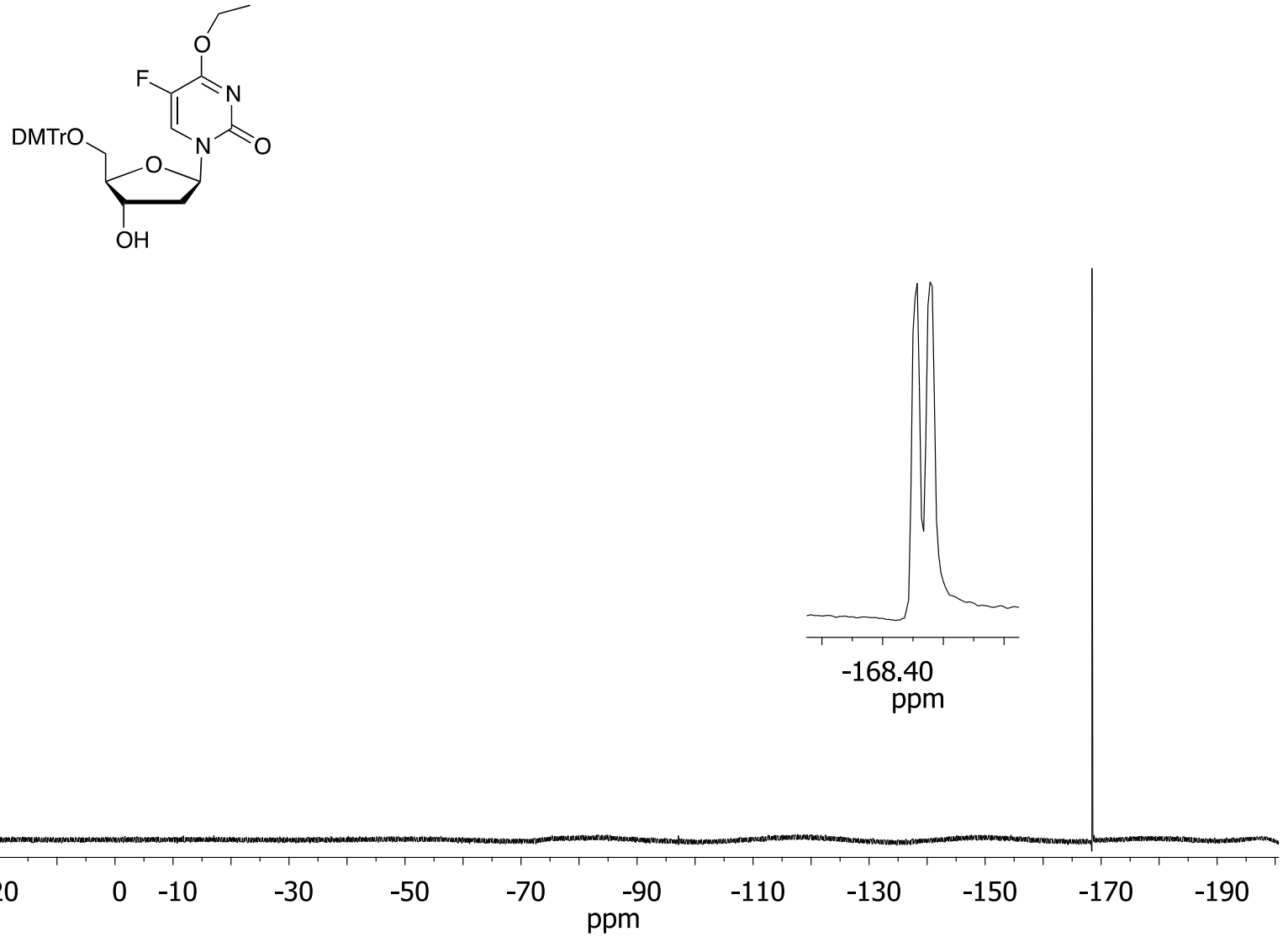


Figure A2.5 - 500 MHz ^{1}H NMR spectrum of compound (**3a**) (in d_6-acetone)

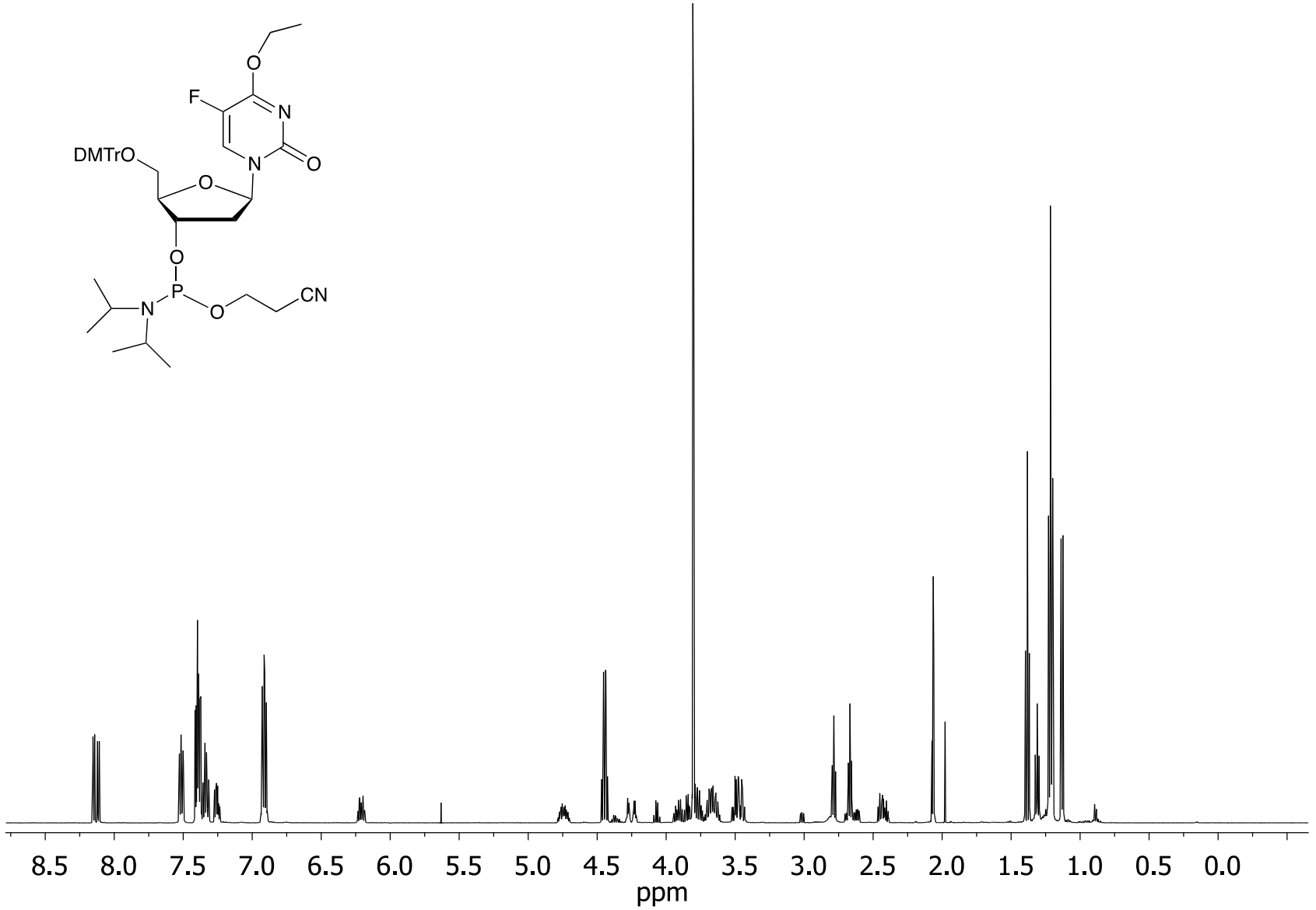


Figure A2.6 - 125.7 MHz ^{13}C NMR spectrum of compound (**3a**) (in d_6-acetone)

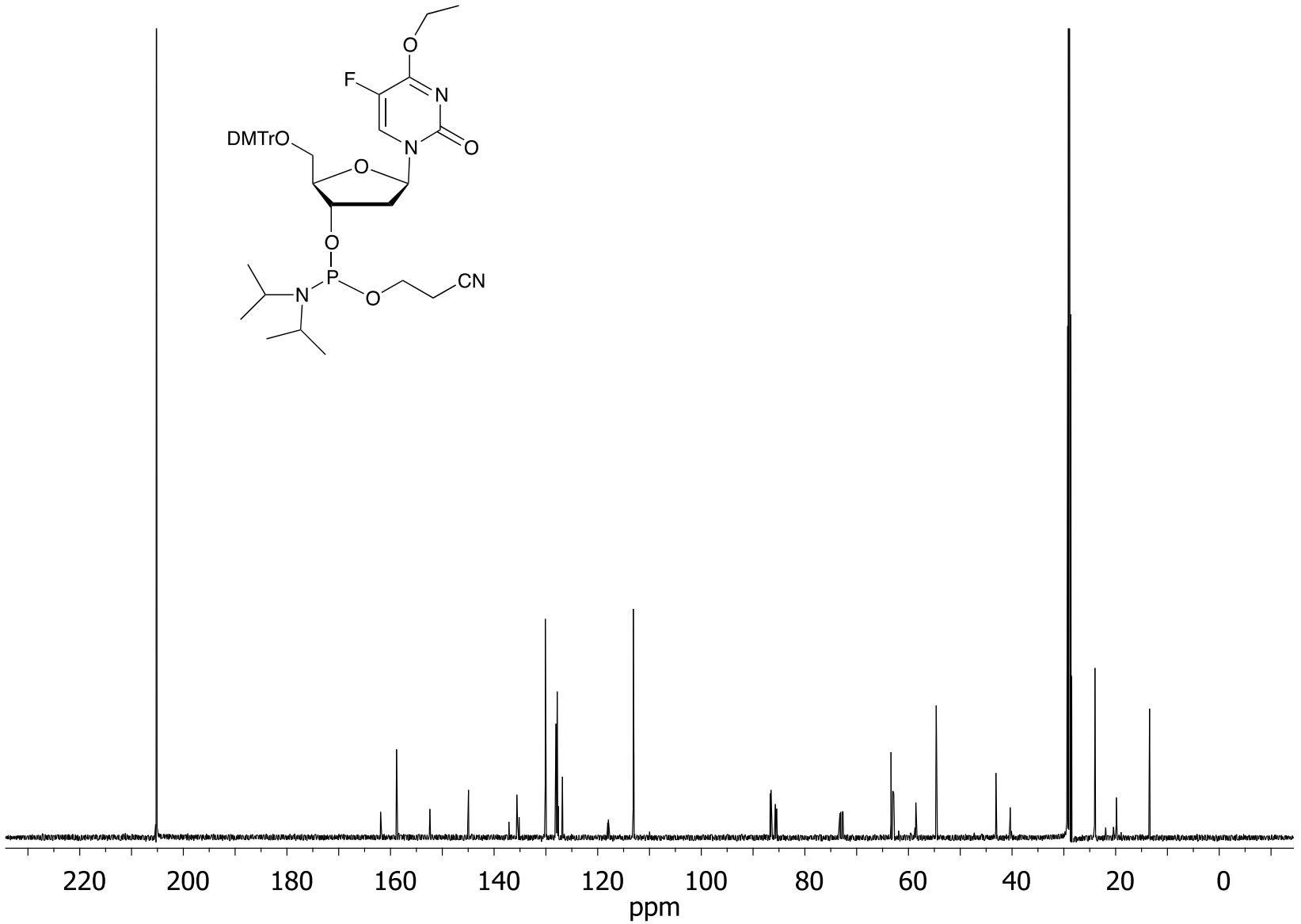


Figure A2.7 - 470.4 MHz ^{19}F NMR spectrum of compound (**3a**) (in d_6-acetone)

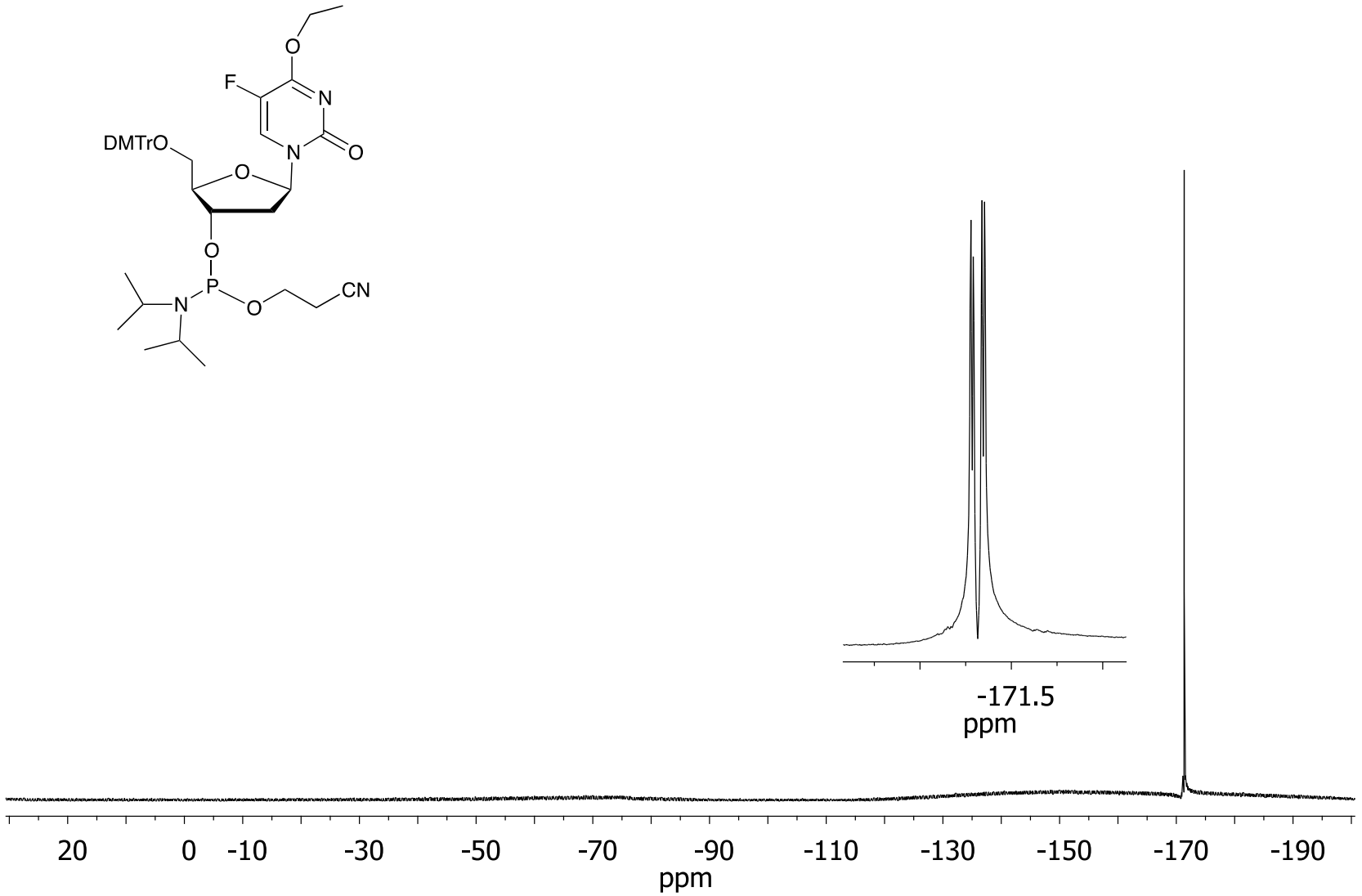


Figure A2.8 - 202.3 MHz ^{31}P NMR spectrum of compound (**3a**) (in d_6-acetone)

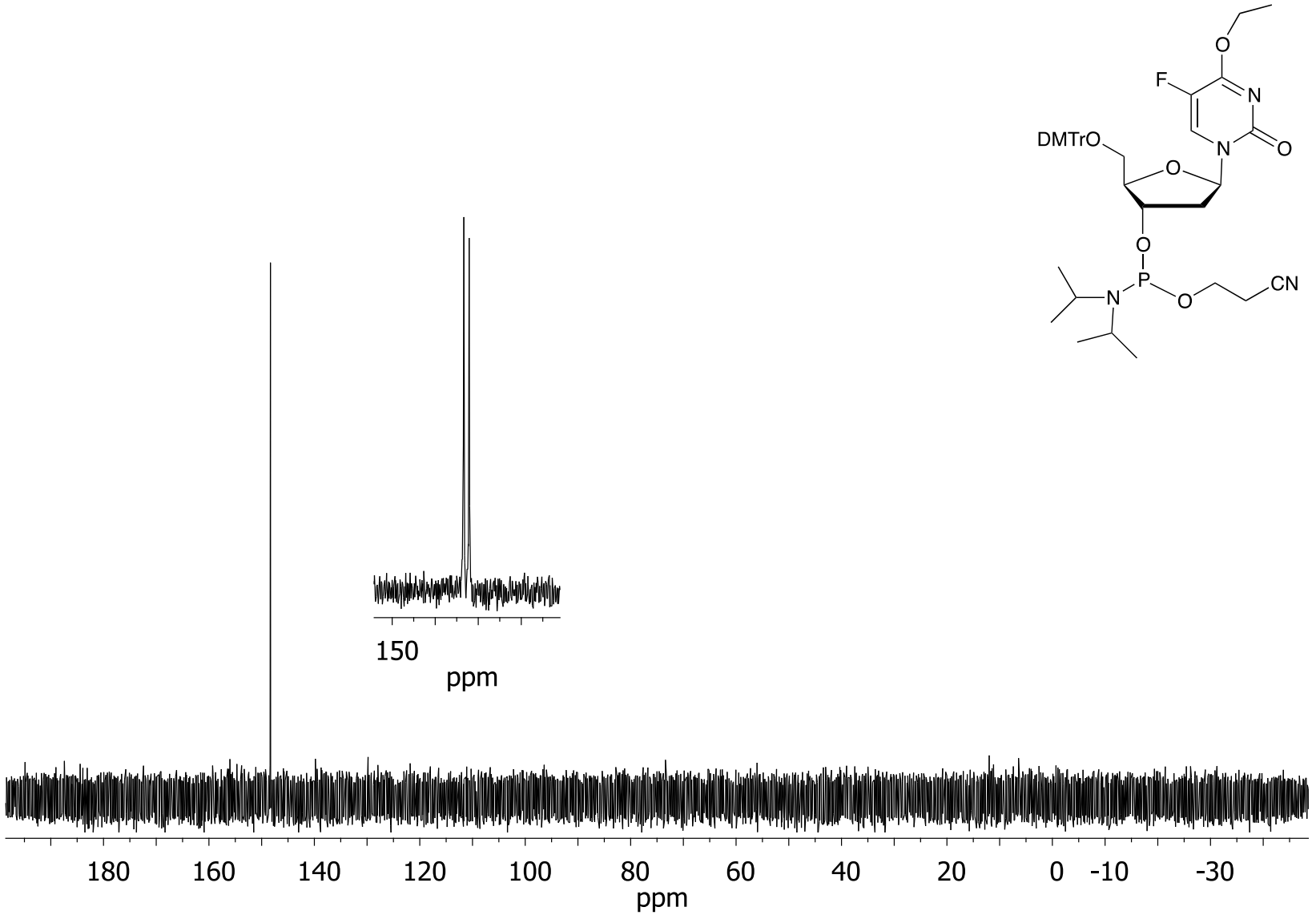


Figure A2.9 - 500 MHz ^{1}H NMR spectrum of compound (**2b**) (in $CDCl_3$)

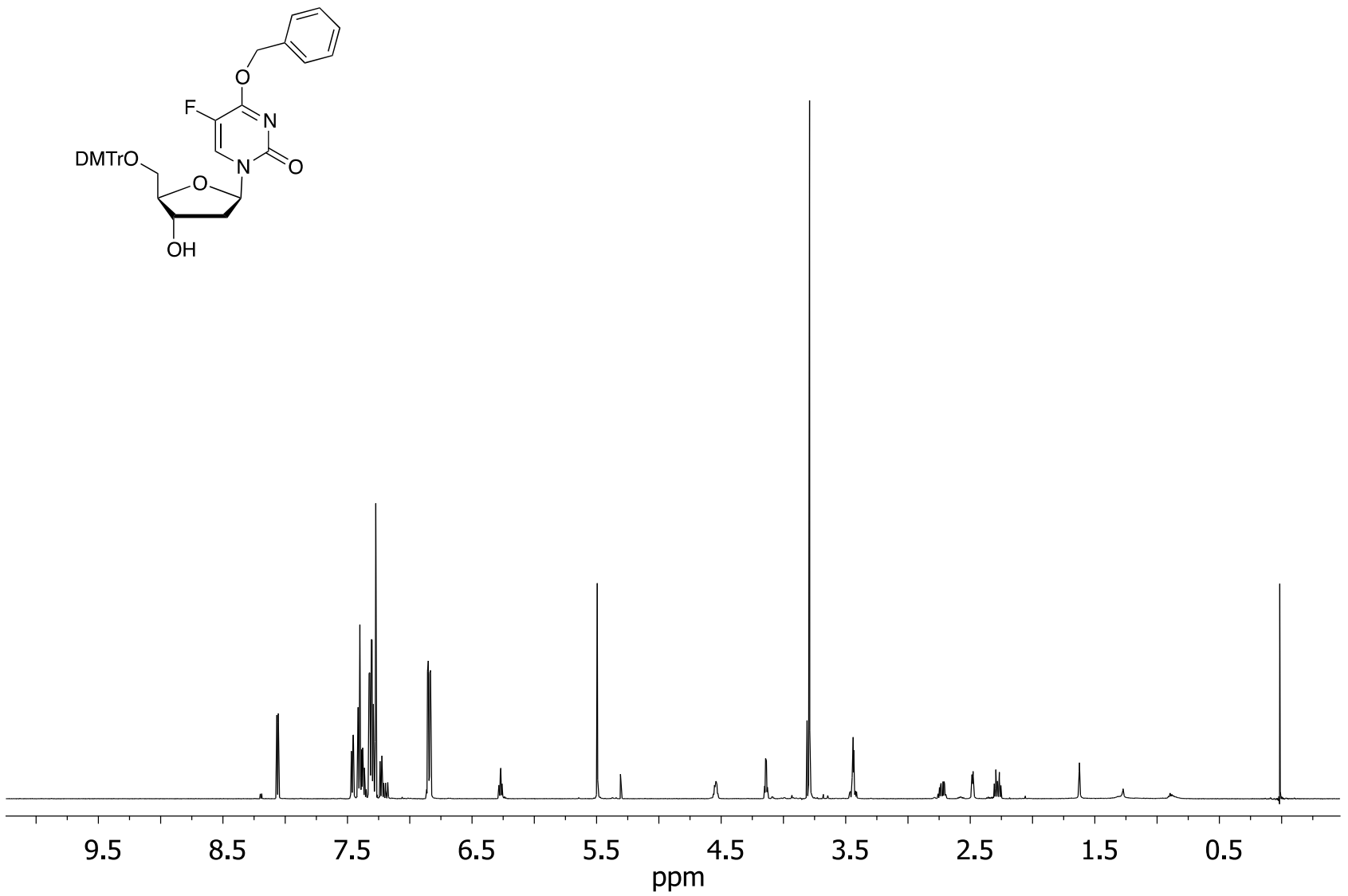


Figure A2.10 - 125.7 MHz ^{13}C NMR spectrum of compound (**2b**) (in $CDCl_3$)

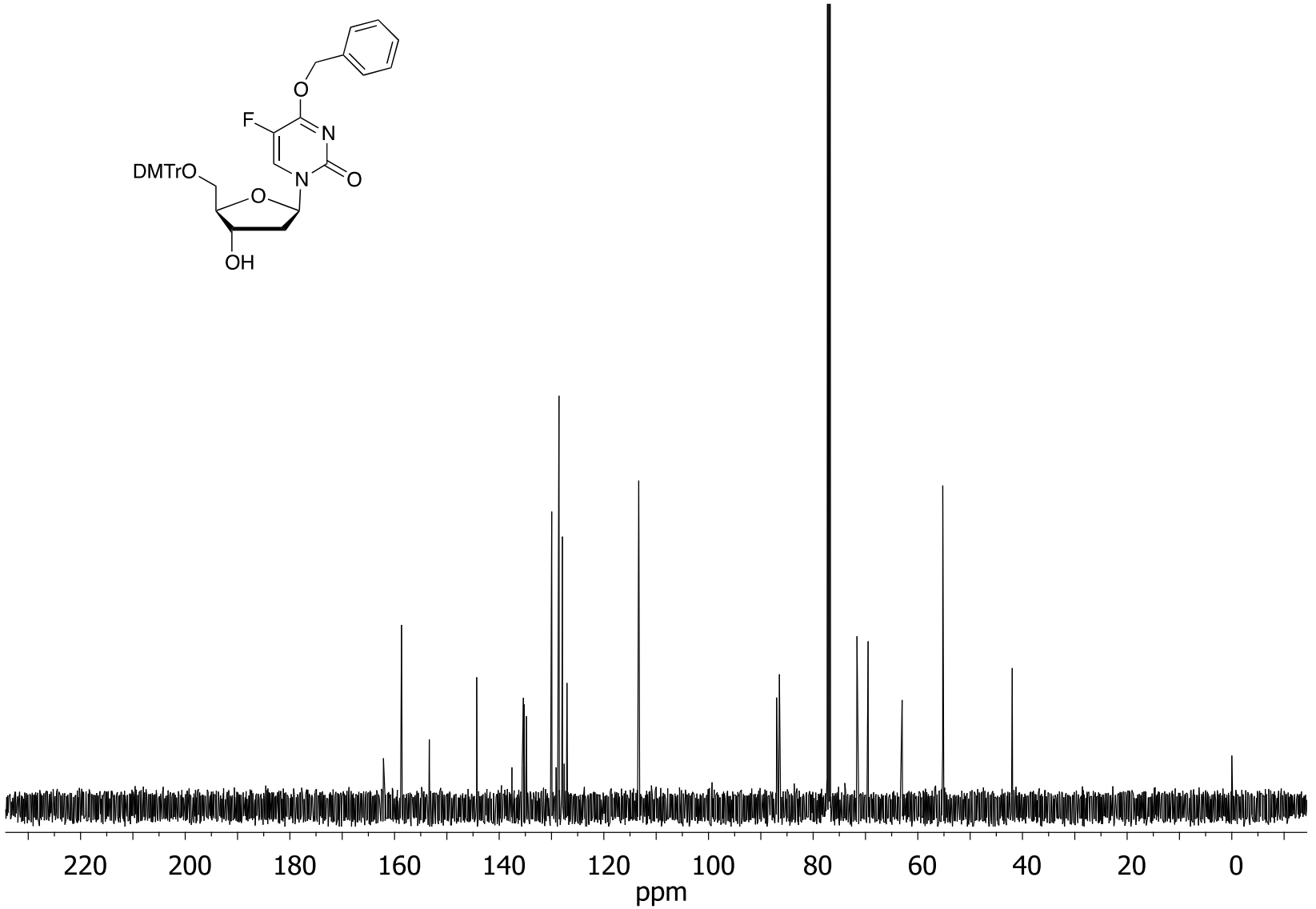


Figure A2.11 - 470.4 MHz ^{19}F NMR spectrum of compound (**2b**) (in $CDCl_3$) (in d_6-acetone)

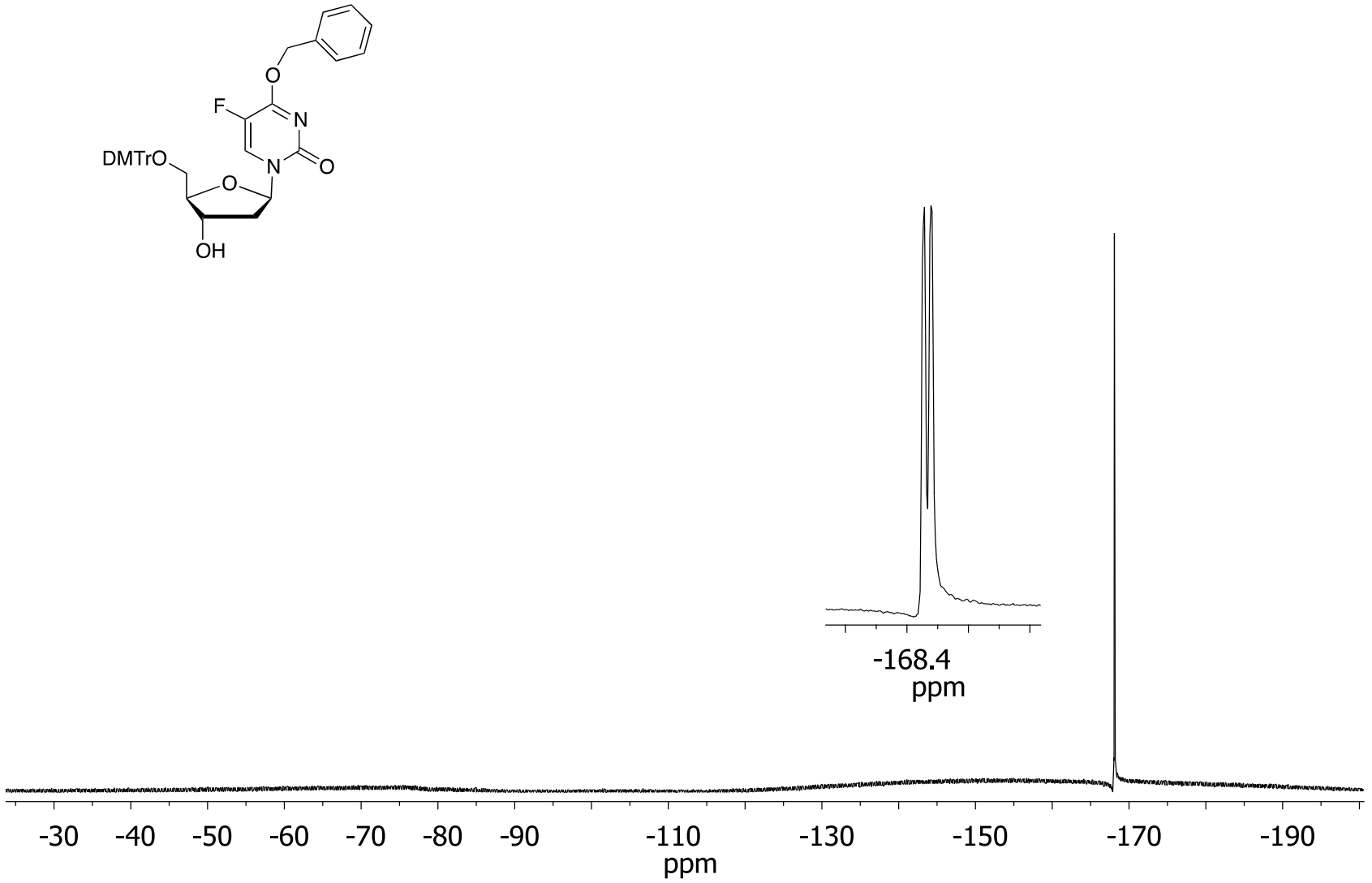


Figure A2.12 - 500 MHz ^{1}H NMR spectrum of compound (**3b**) (in d_6-acetone)

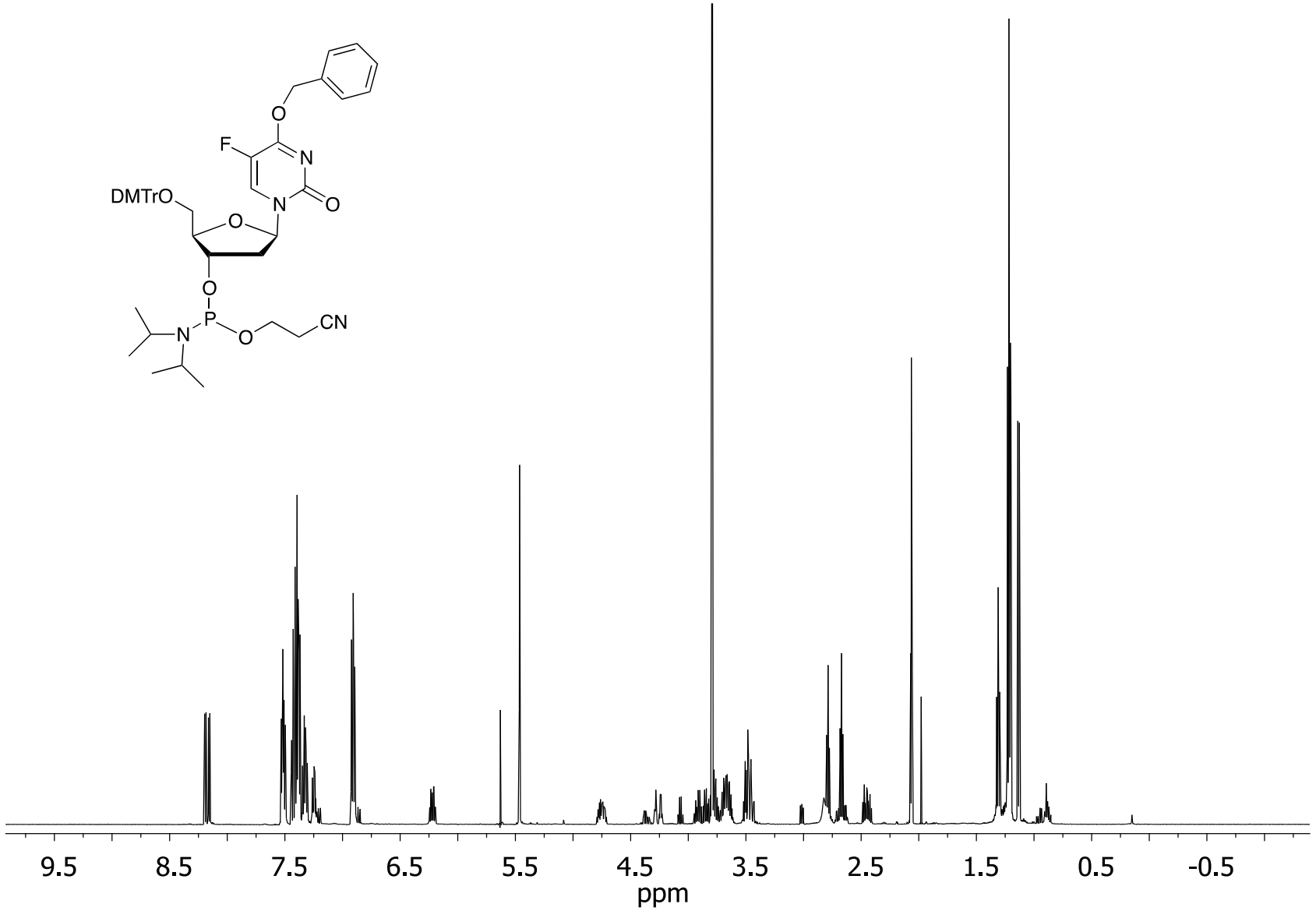


Figure A2.13 - 125.7 MHz ^{13}C NMR spectrum of compound (**3b**) (in d_6-acetone)

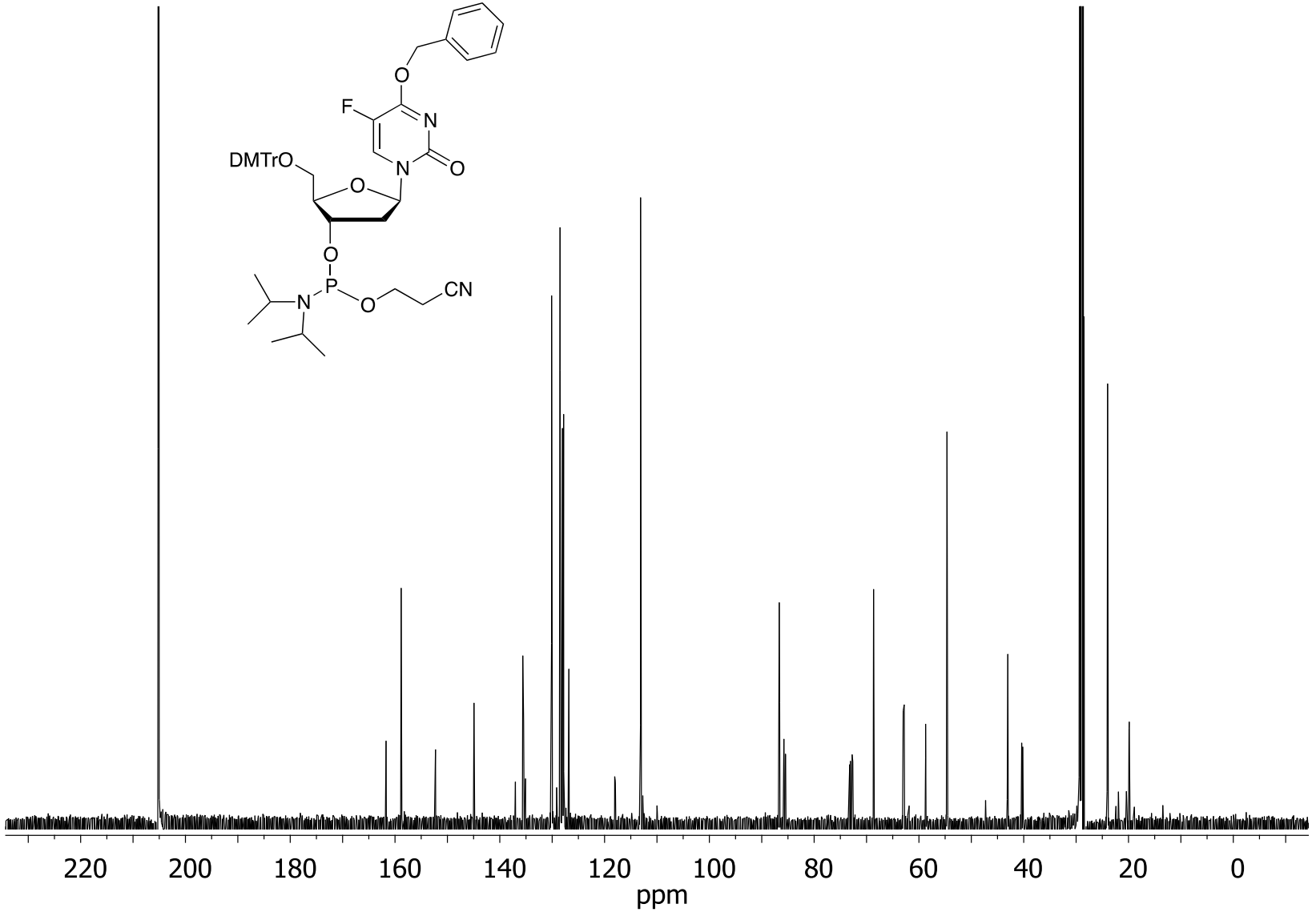


Figure A2.14 - 470.4 MHz ^{19}F NMR spectrum of compound (**3b**) (in d_6-acetone)

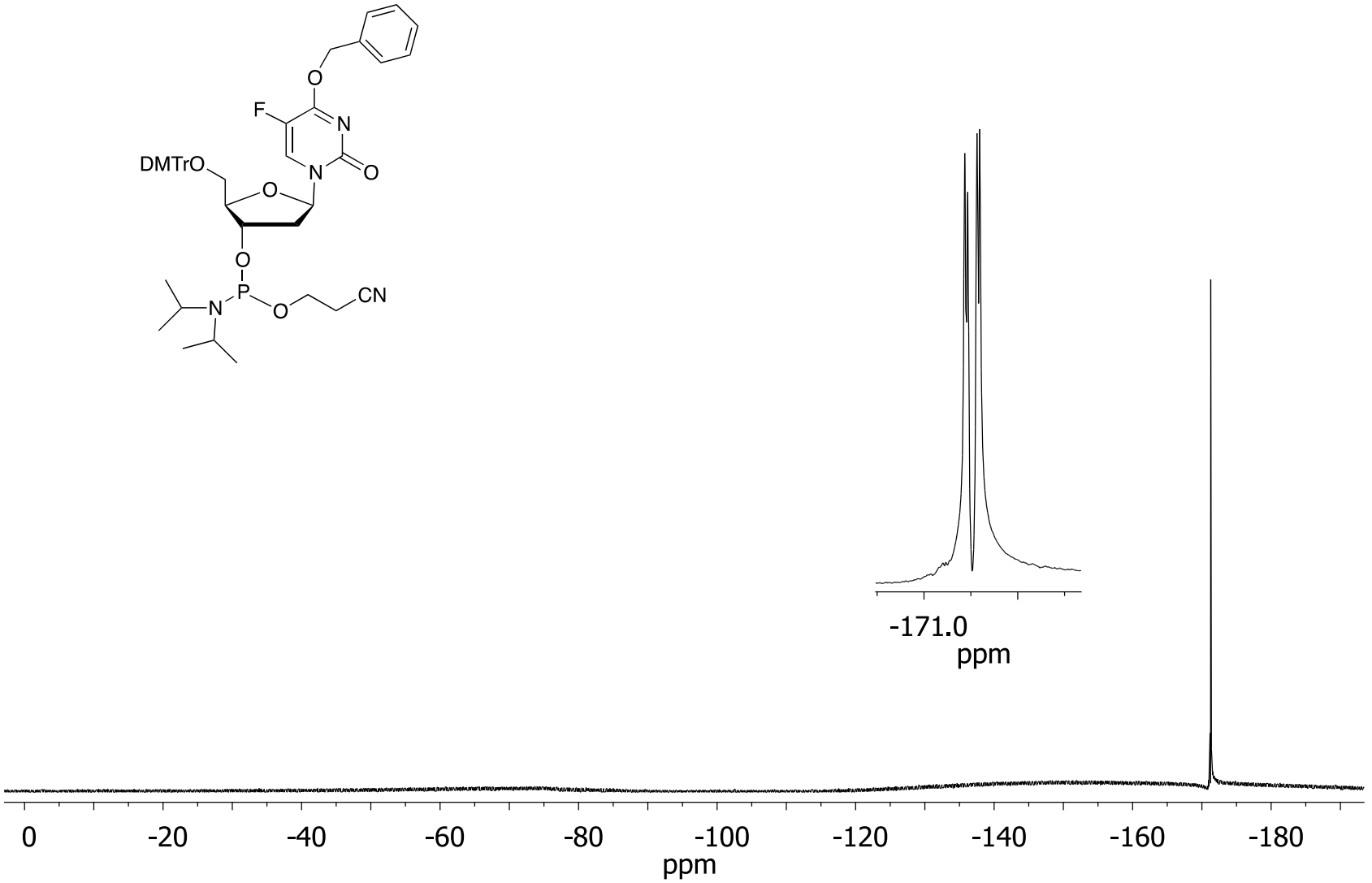


Figure A2.15 - 202.3 MHz ^{31}P NMR spectrum of compound (**3b**) (in d_6-acetone)

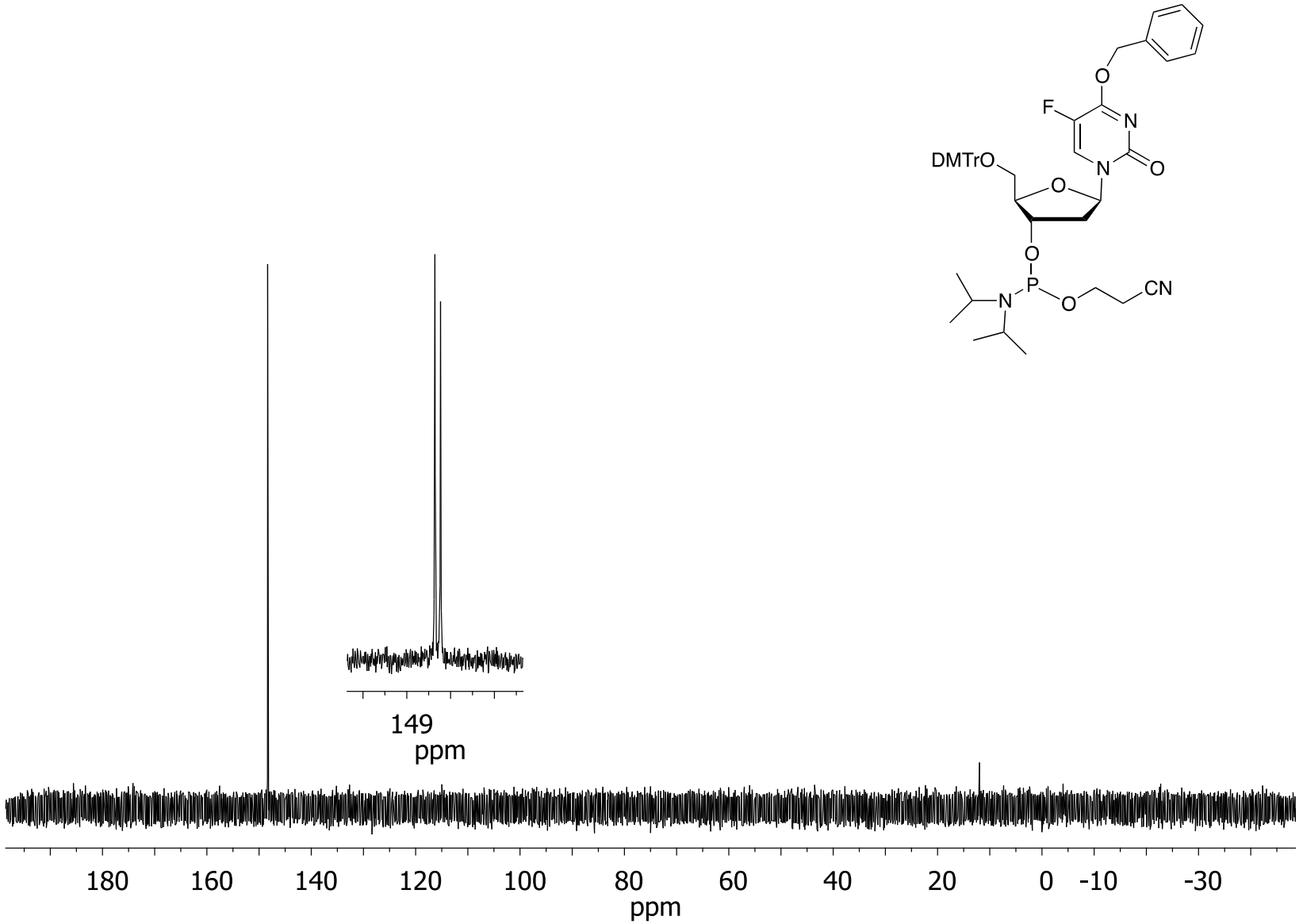


Figure A2.16 - 500 MHz ^{1}H NMR spectrum of compound (**4a**) (in $CDCl_3$)

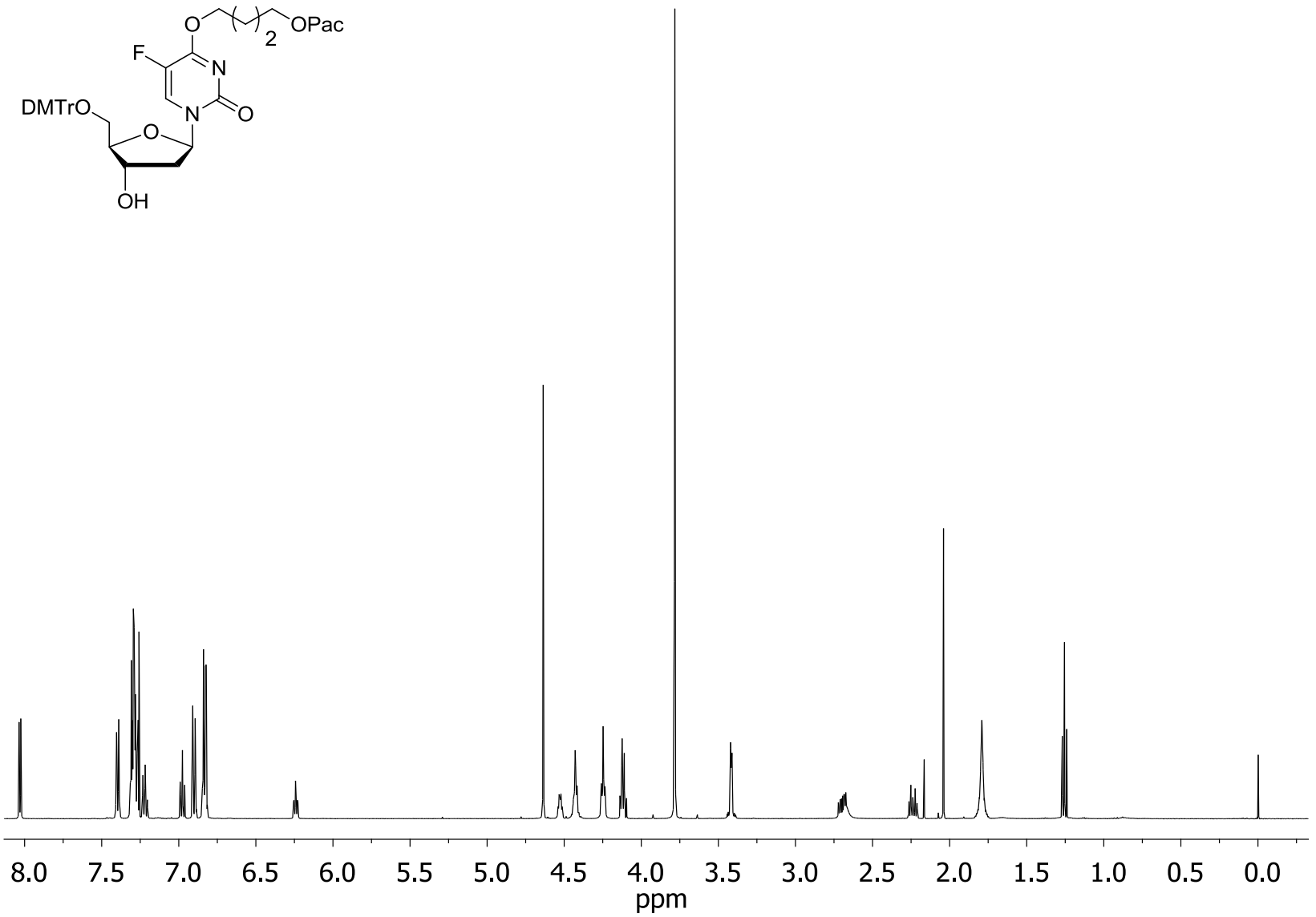


Figure A2.17 – 125.7 MHz ^{1}H NMR spectrum of compound (**4a**) (in $CDCl_3$)

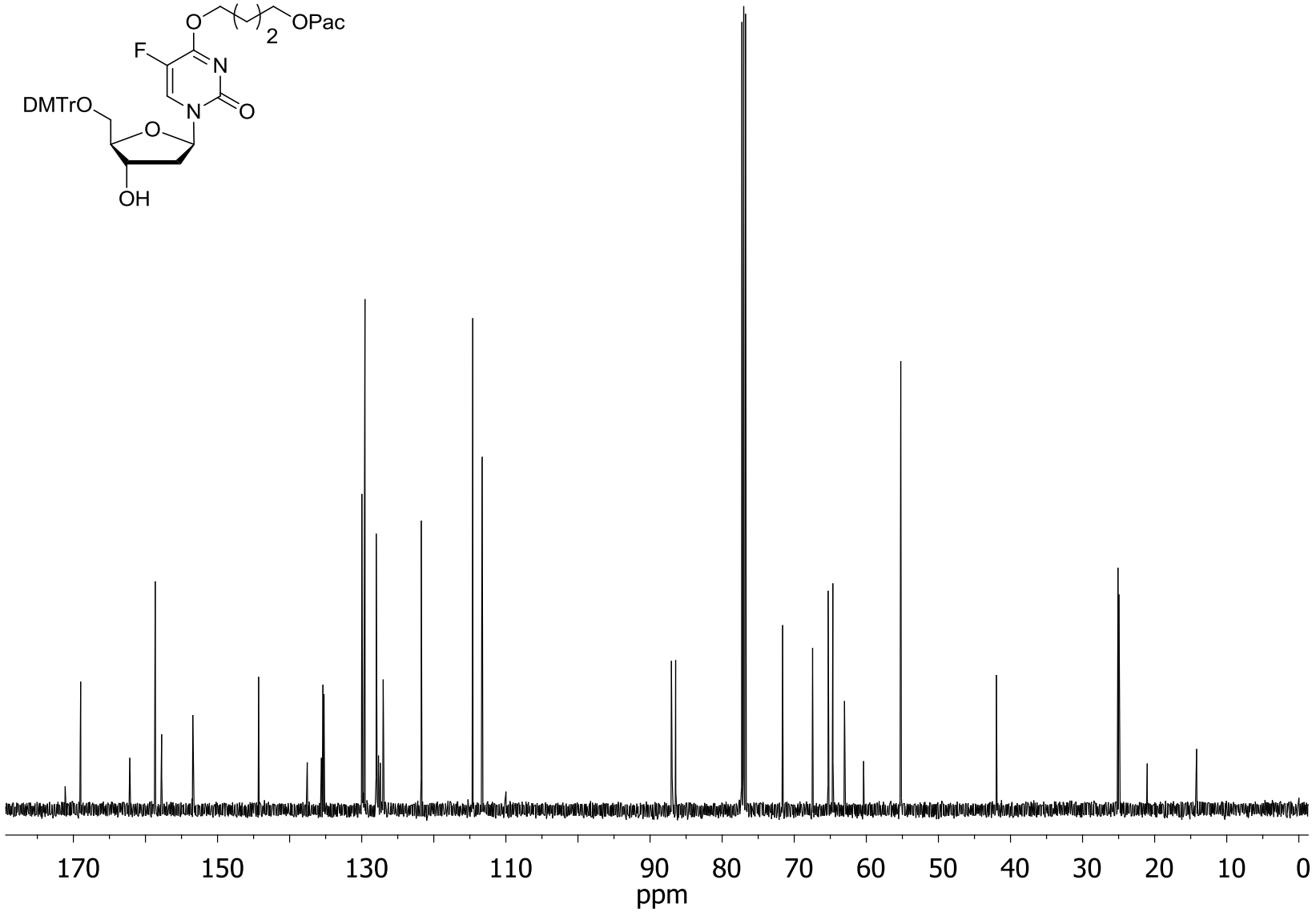


Figure A2.18 – 470.4 MHz ^{19}F NMR spectrum of compound (**4a**) (in $CDCl_3$)

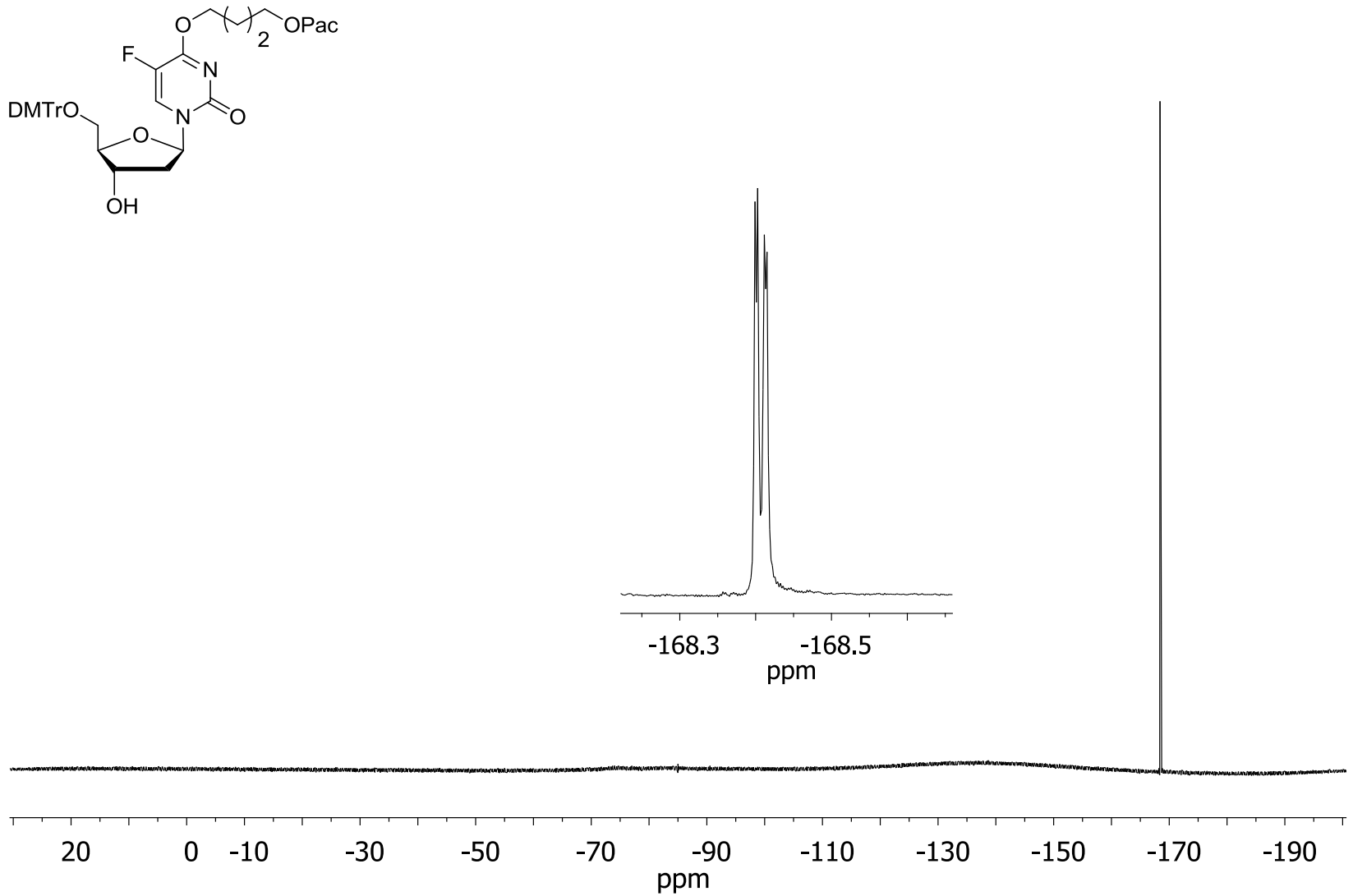


Figure A2.19 - 500 MHz ^{1}H NMR spectrum of compound (**5a**) (in d_6-acetone)

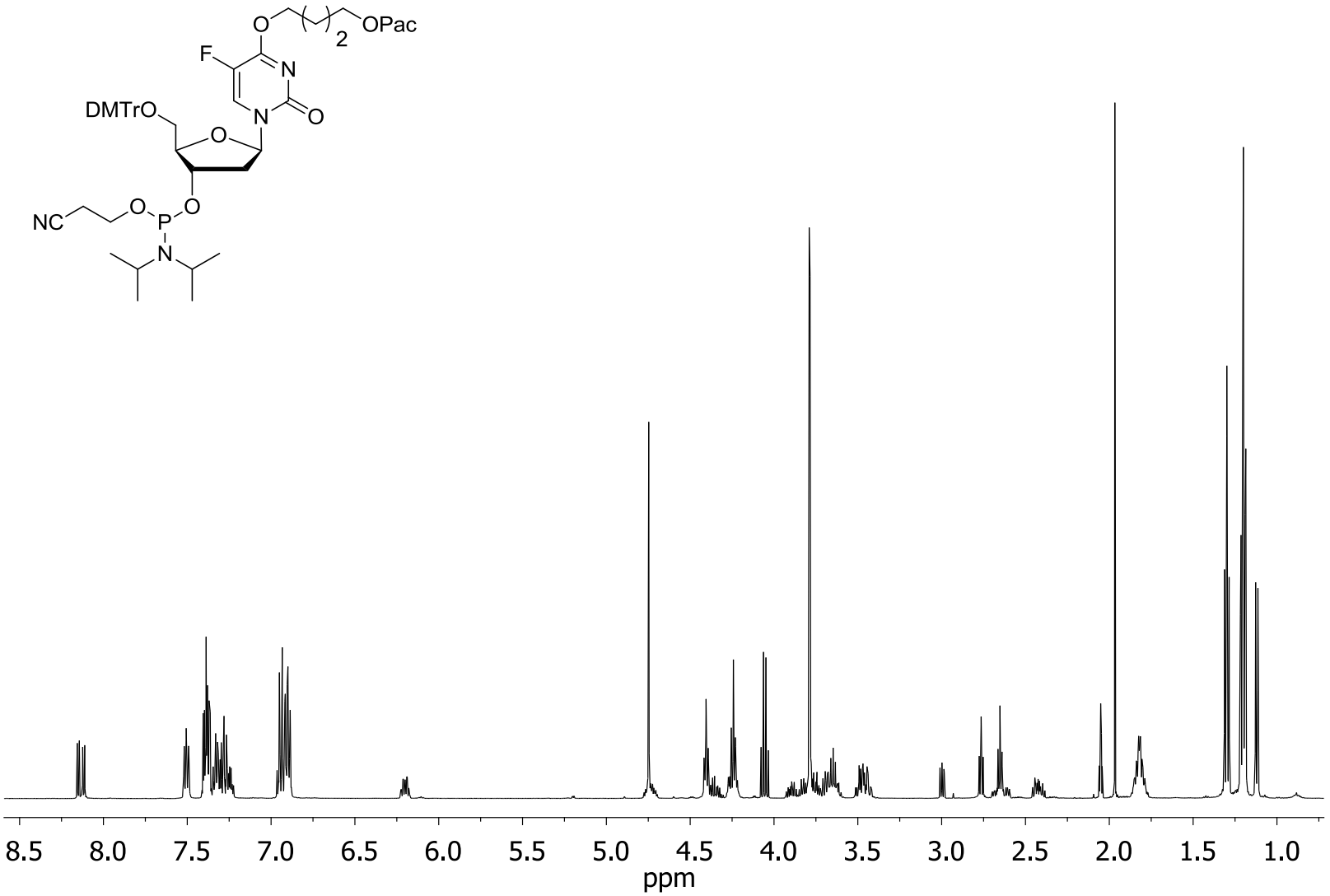


Figure A2.20 – 125.7 MHz ^{1}H NMR spectrum of compound (**5a**) (in d_6-acetone)

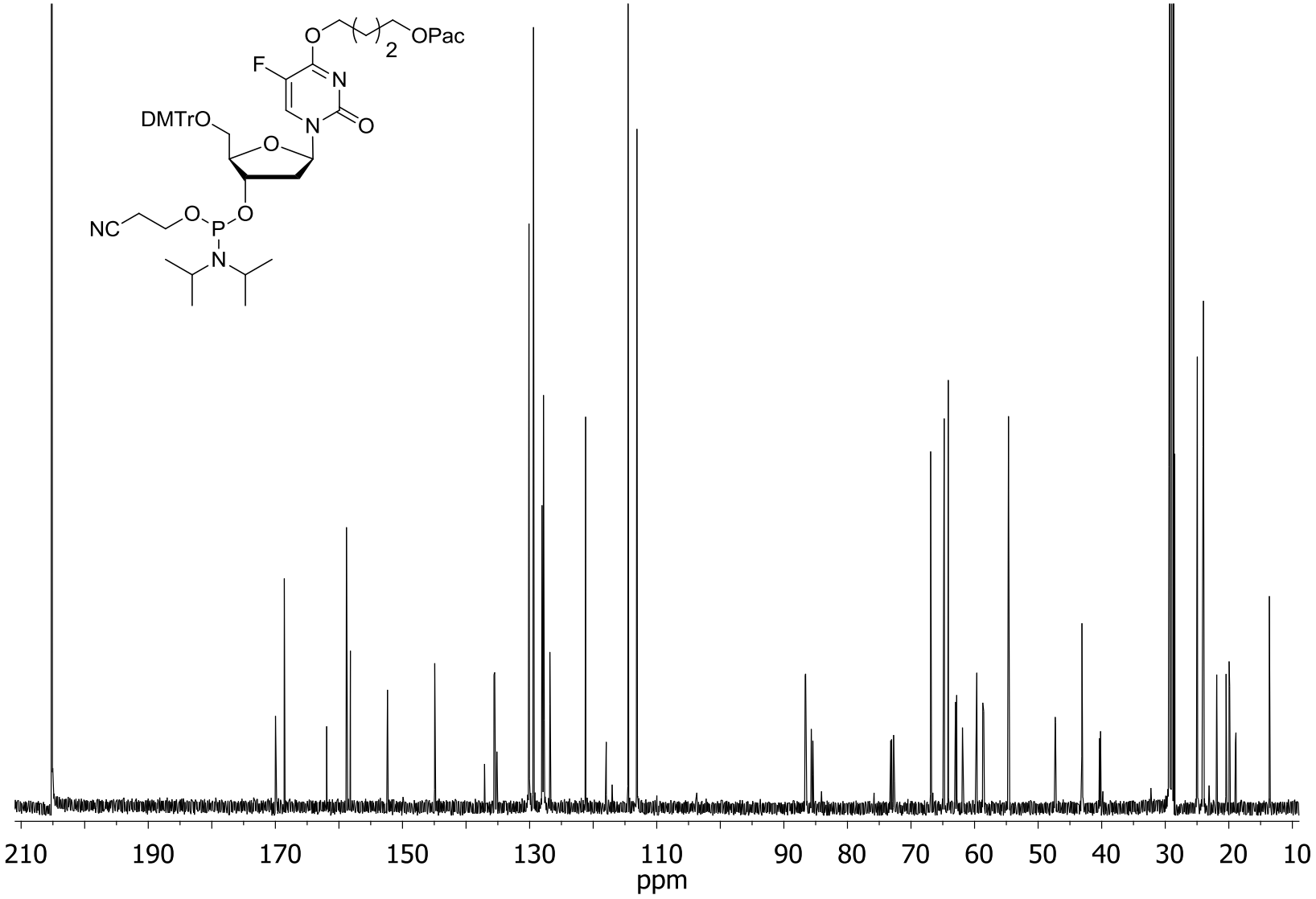


Figure A2.21 – 470.4 MHz ^{19}F NMR spectrum of compound (**5a**) (in d_6-acetone)

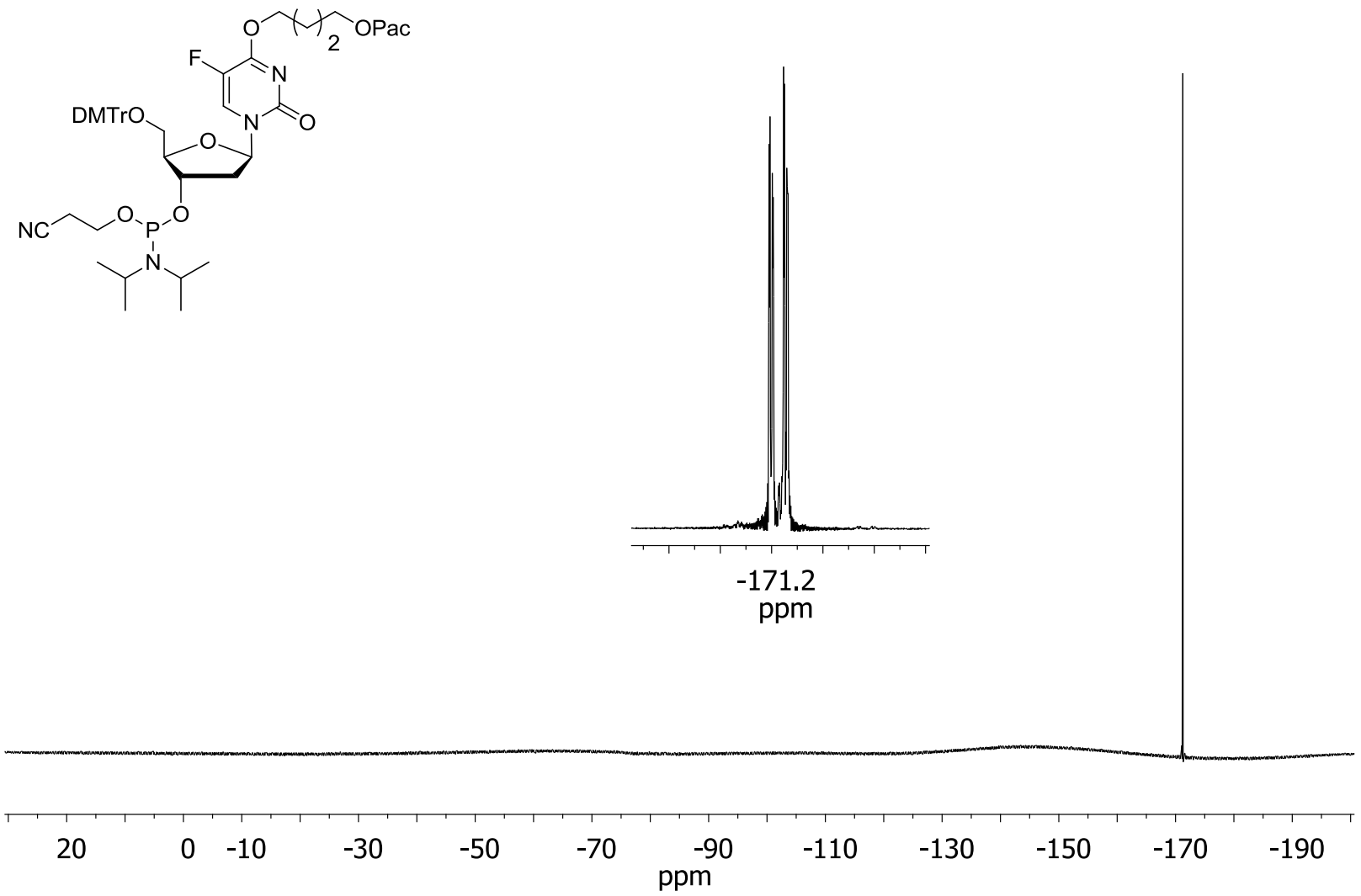


Figure A2.22 - 202.3 MHz ^{31}P NMR spectrum of compound (**5a**) (in d_6-acetone)

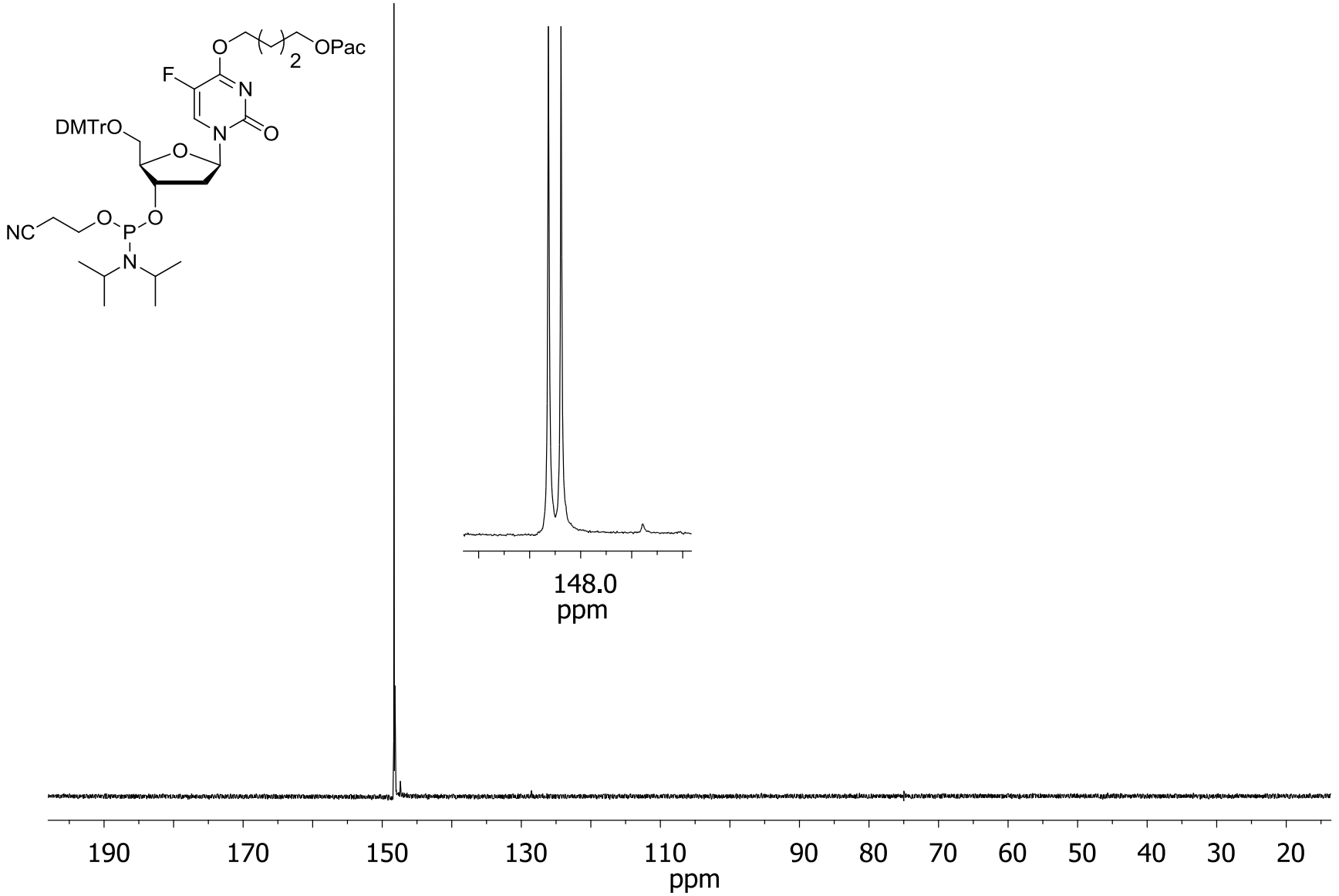


Figure A2.23 - 500 MHz ^{1}H NMR spectrum of compound (**4b**) (in $CDCl_3$)

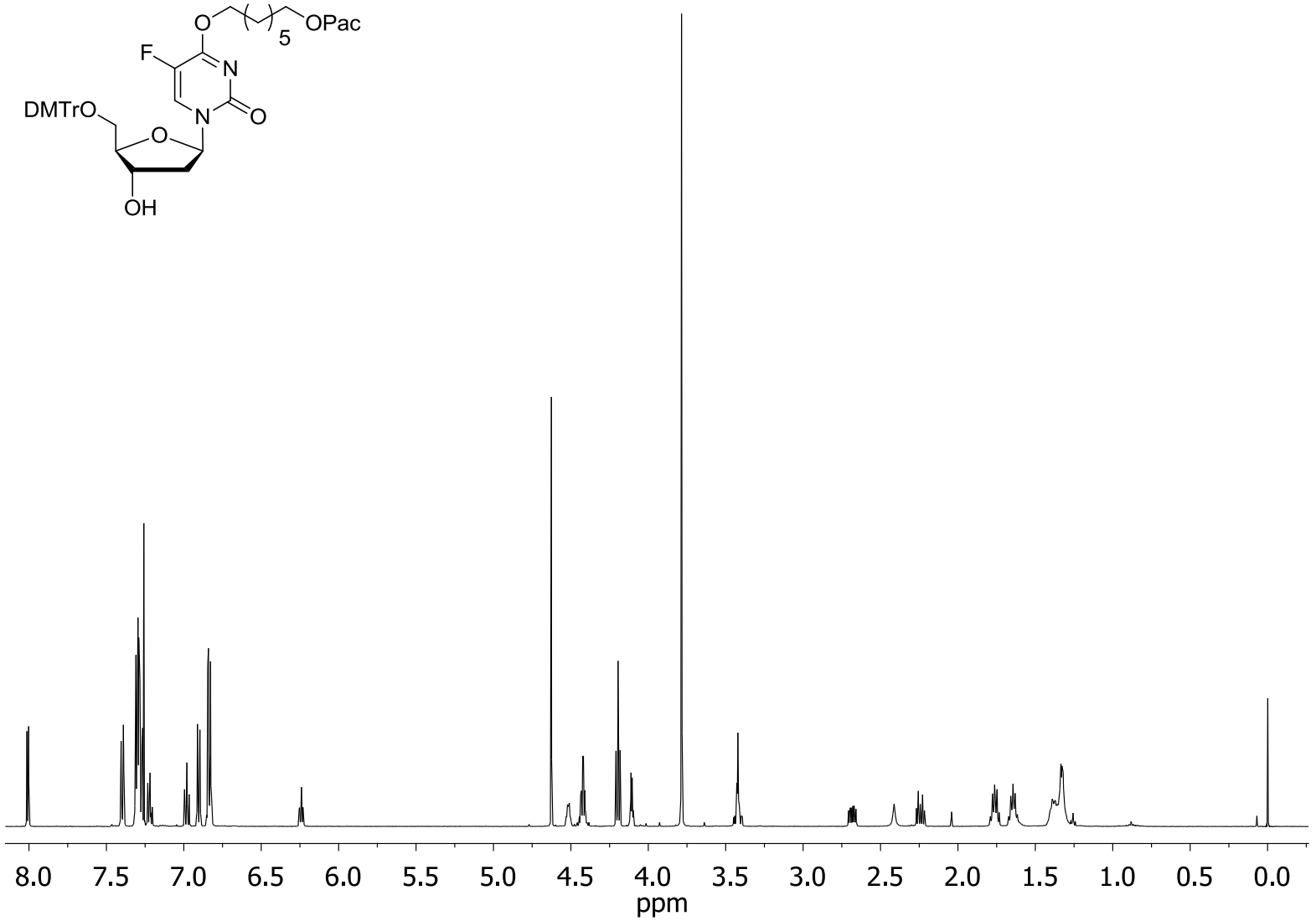


Figure A2.24 – 125.7 MHz ^{1}H NMR spectrum of compound (**4b**) (in $CDCl_3$)

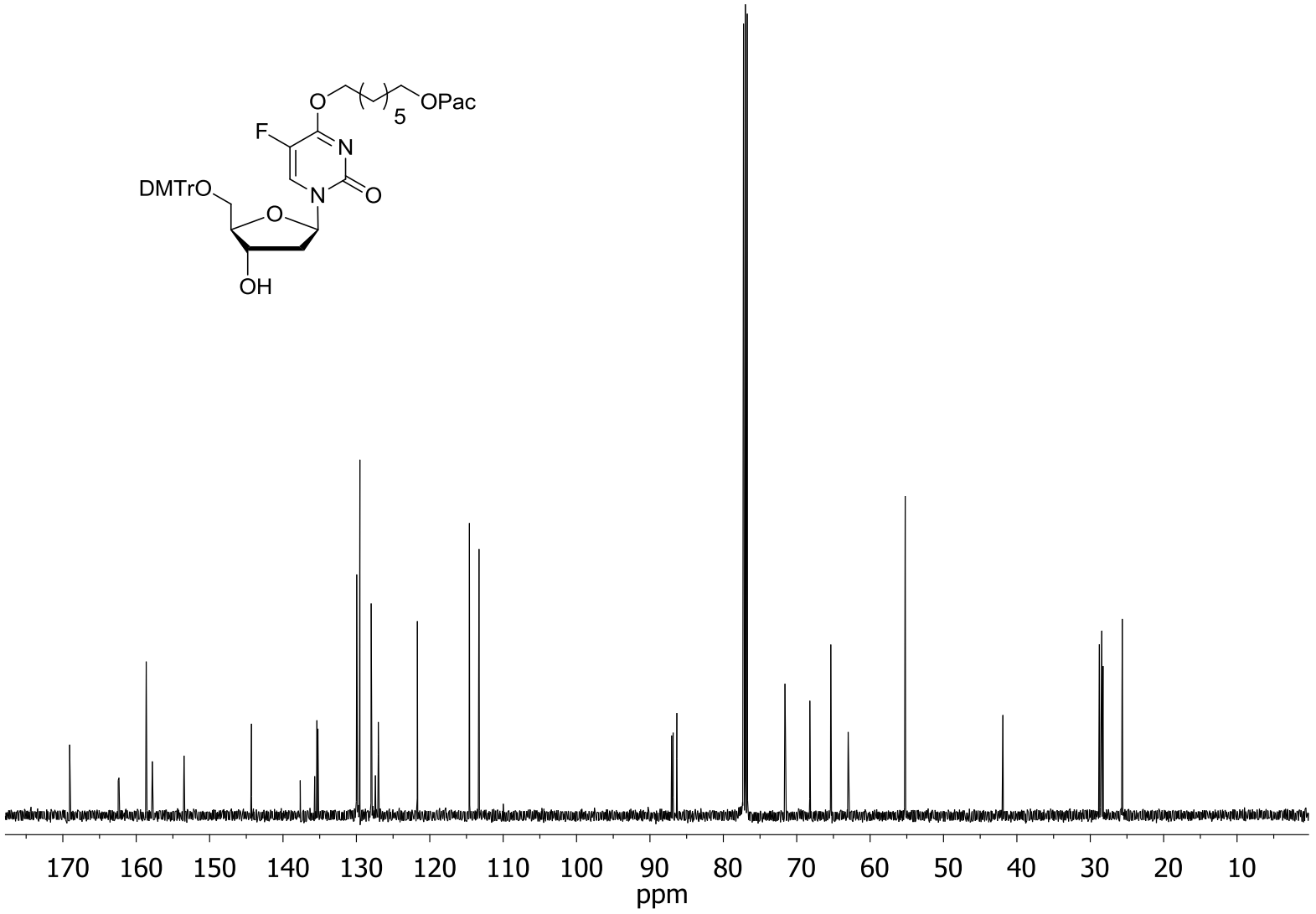


Figure A2.25 – 470.4 MHz ^{19}F NMR spectrum of compound (**4b**) (in $CDCl_3$)

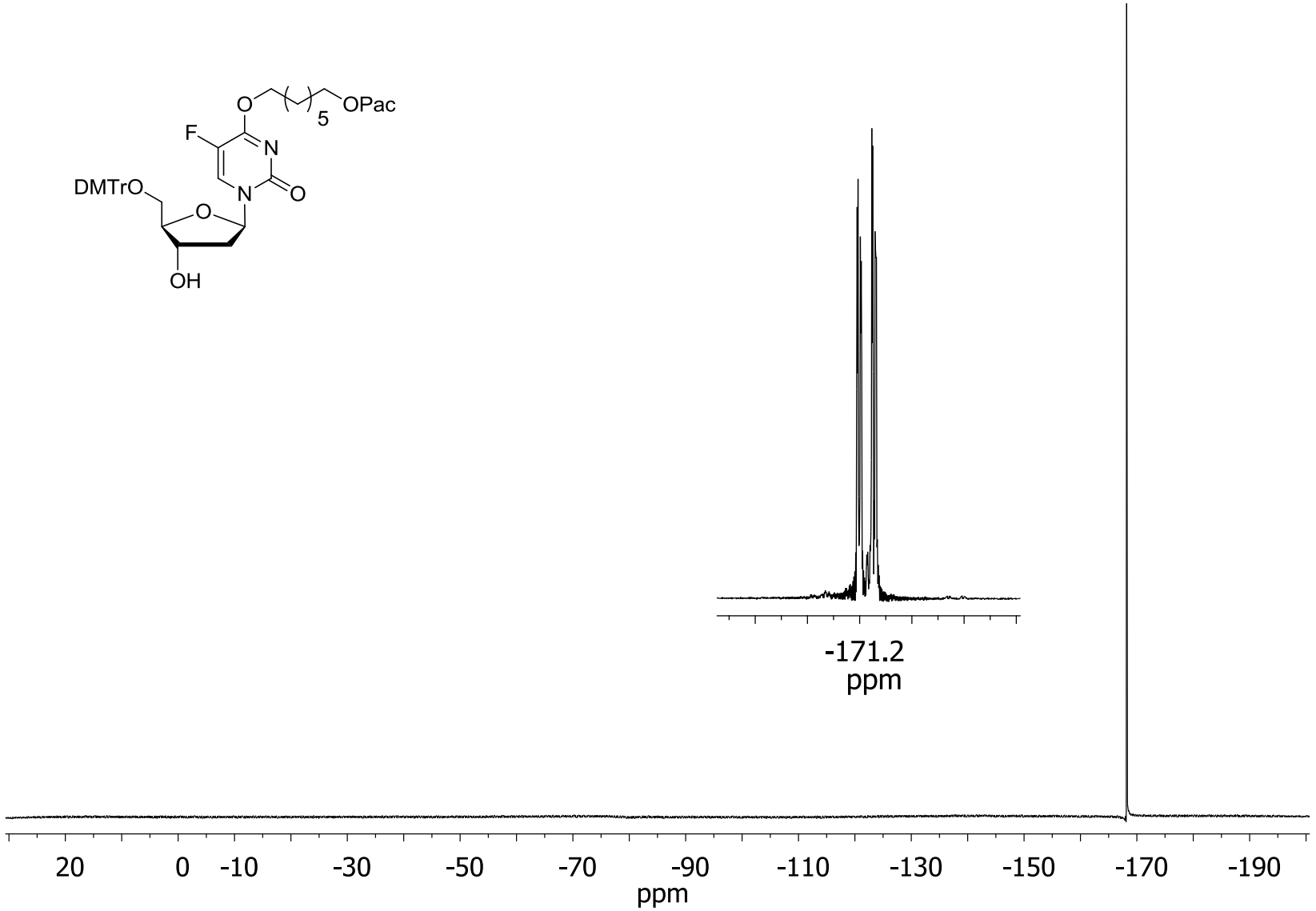


Figure A2.26 - 500 MHz ^{1}H NMR spectrum of compound (**5b**) (in d_6-acetone)

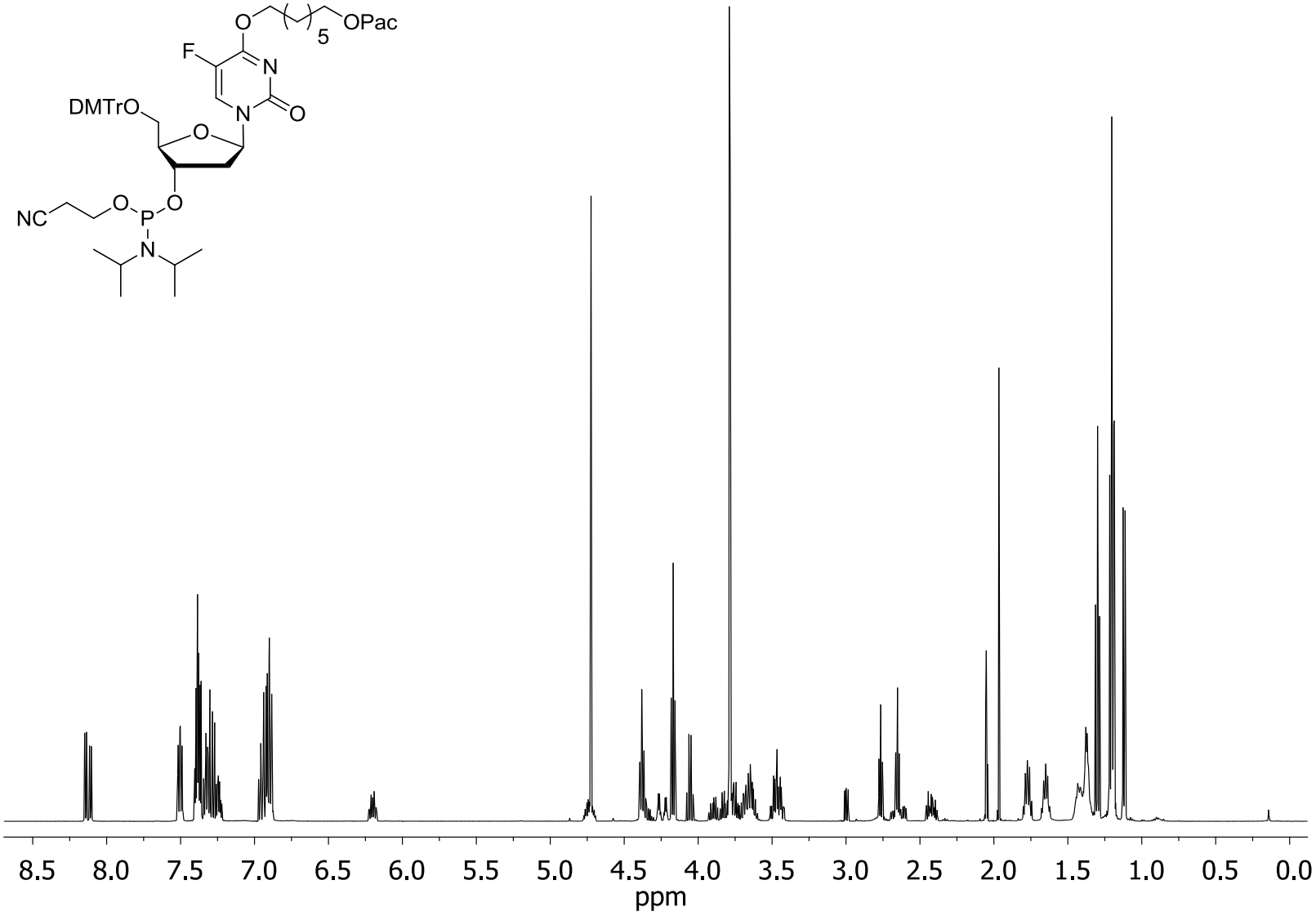


Figure A2.27 – 125.7 MHz ^{1}H NMR spectrum of compound (**5b**) (in d_6-acetone)

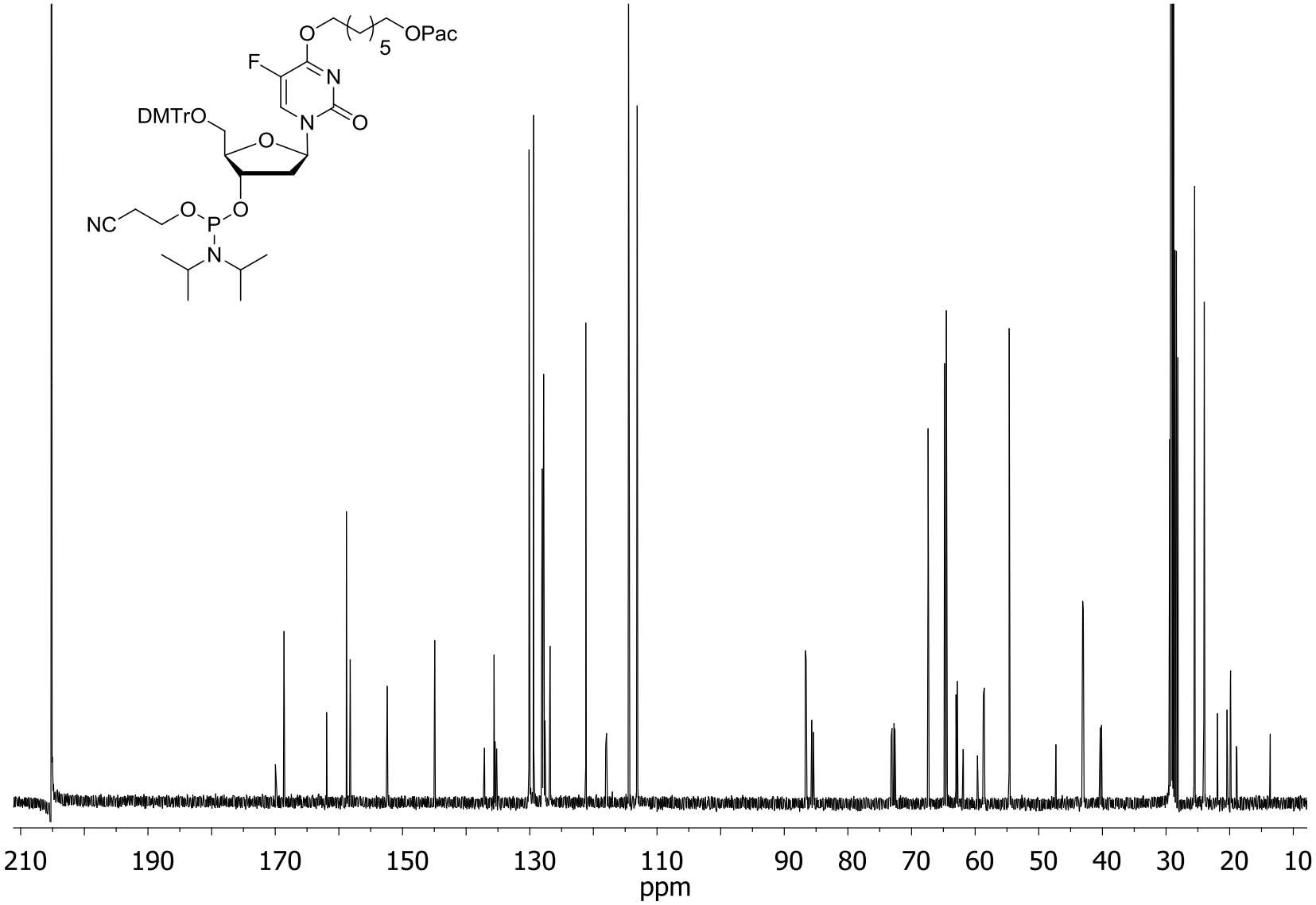


Figure A2.28 – 470.4 MHz ^{19}F NMR spectrum of compound (**5b**) (in d_6-acetone)

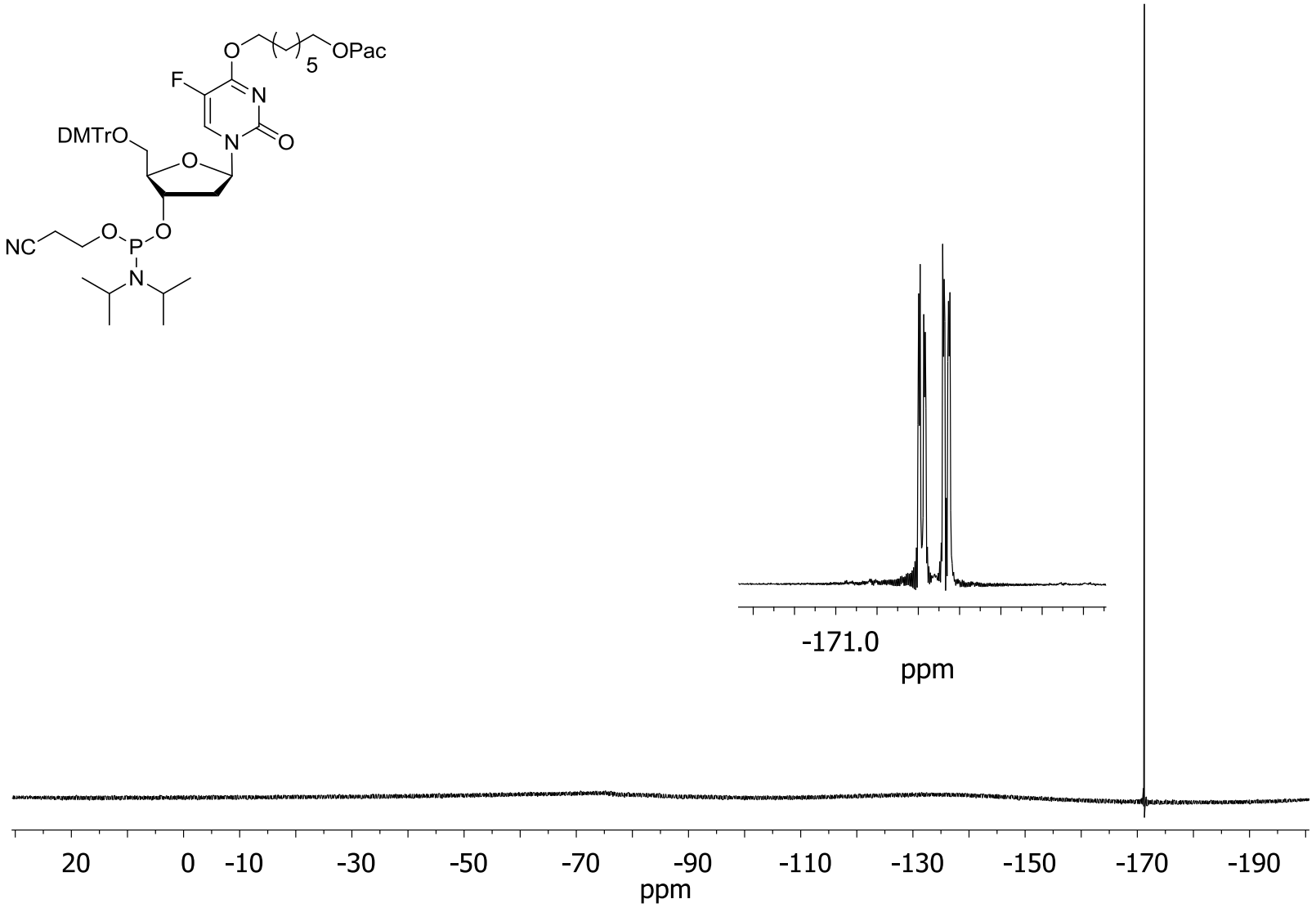


Figure A2.29 - 202.3 MHz ^{31}P NMR spectrum of compound (**5b**) (in d_6-acetone)

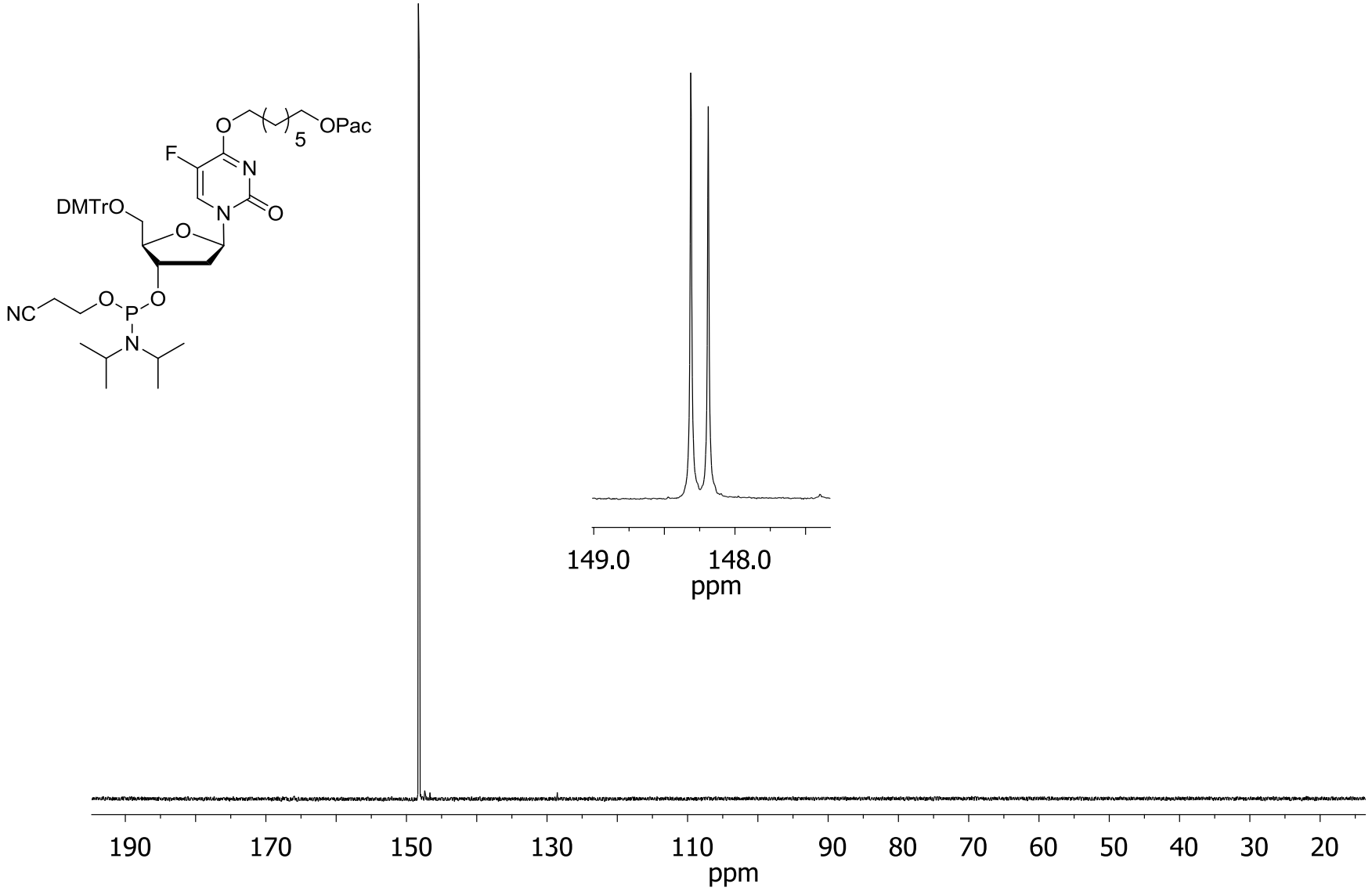


Figure A2.30 - ESI MS spectrum of oligonucleotide **dFU-Et** (expected mass of 4295.1)

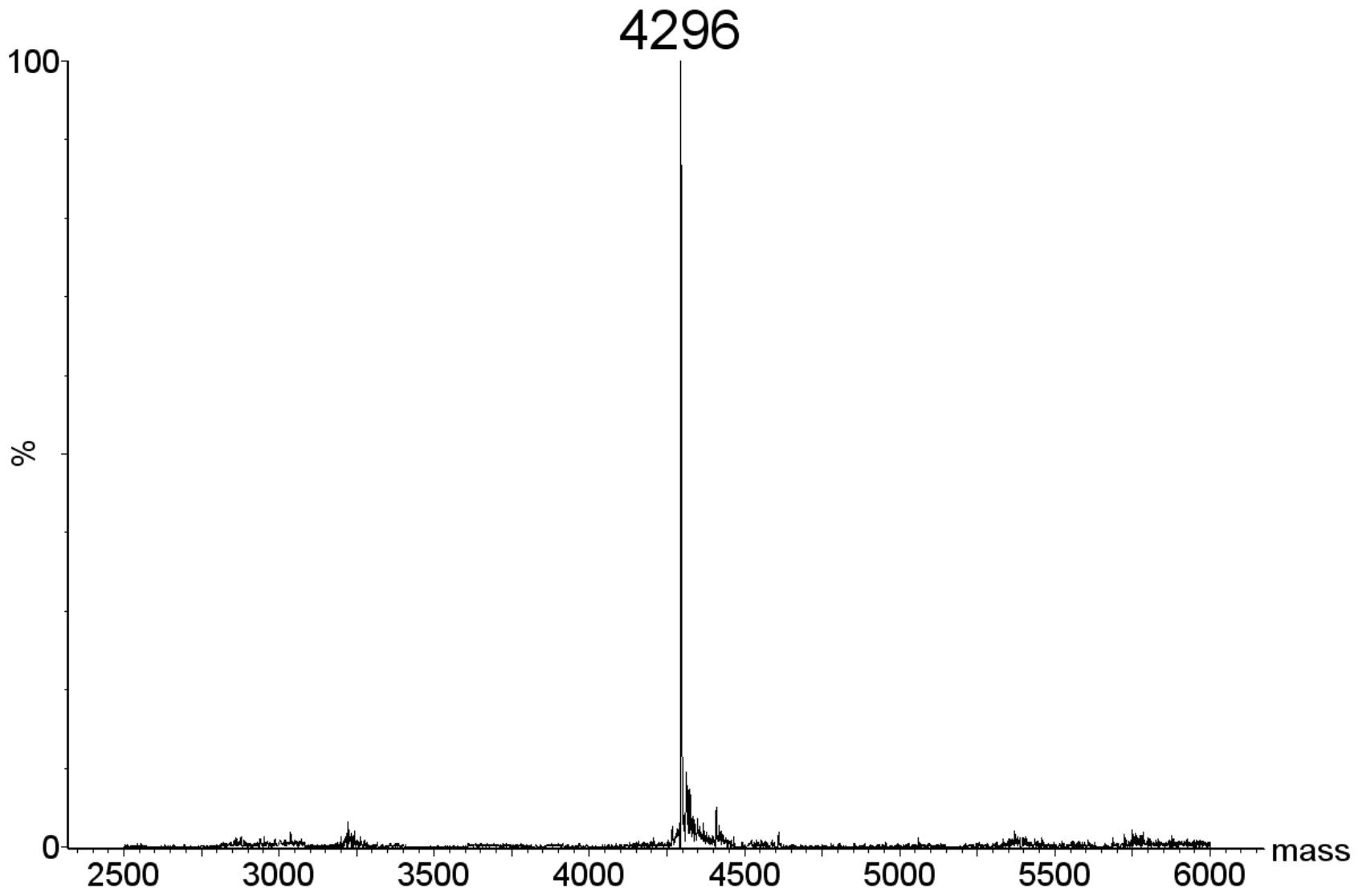


Figure A2.31 - ESI MS spectrum of oligonucleotide **dFU-Bn** (expected mass of 4358.1)

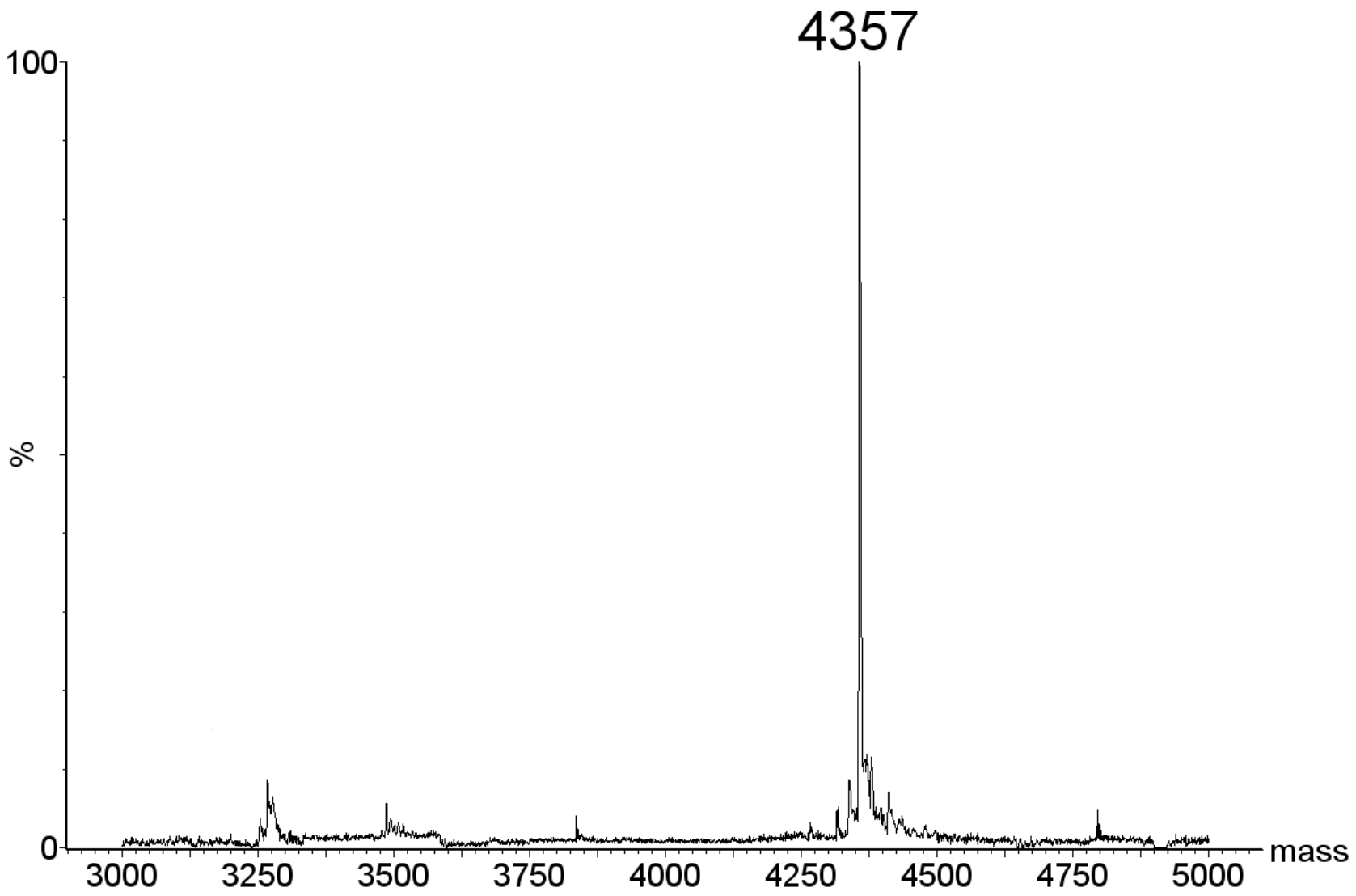


Figure A2.32 - ESI MS spectrum of oligonucleotide **dFU-C4OH** (expected mass of 4341)

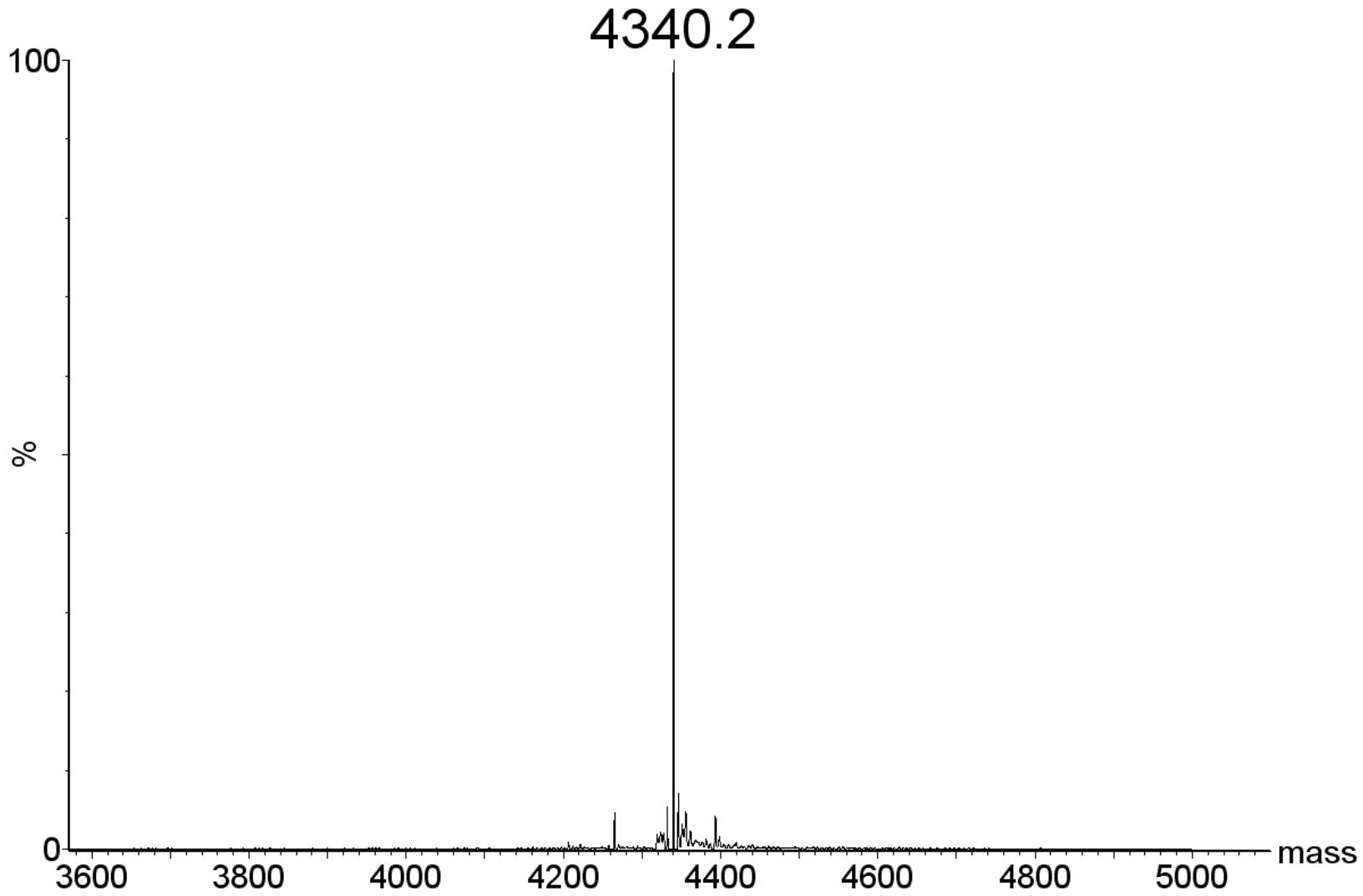


Figure A2.33 - ESI MS spectrum of oligonucleotide **dFU-C7OH** (expected mass of 4382.9)

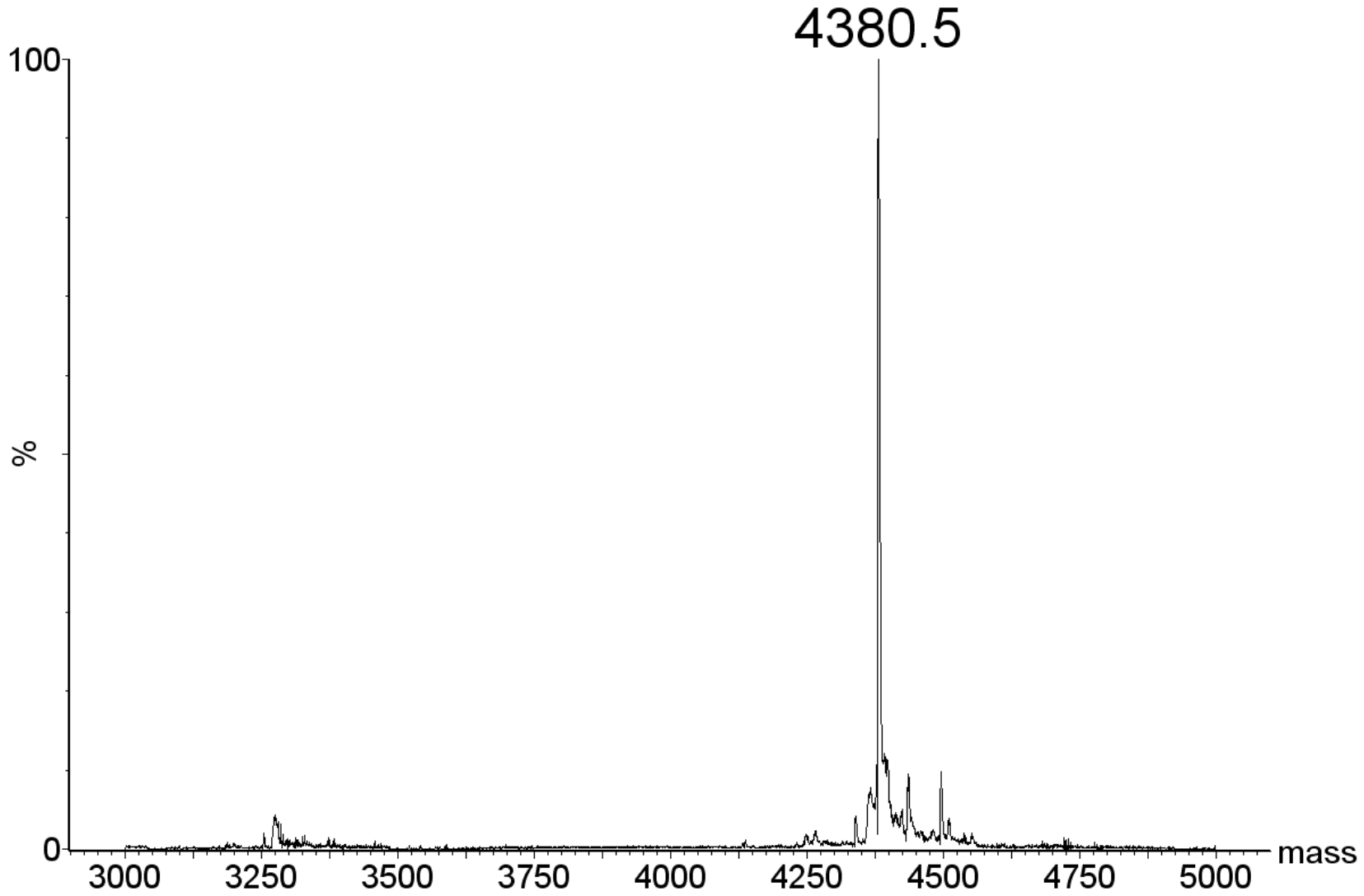
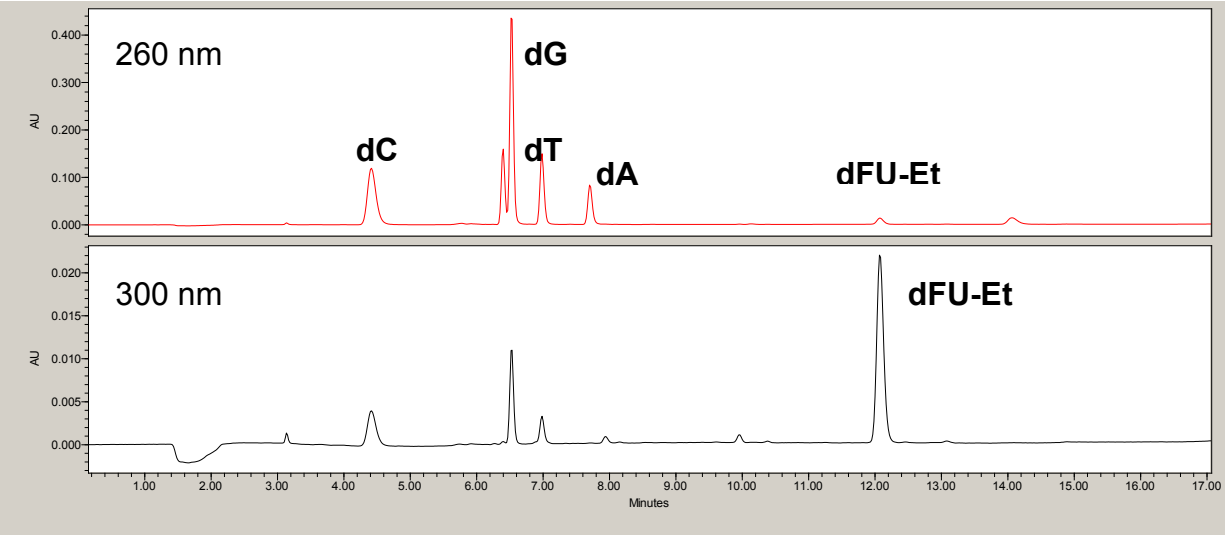
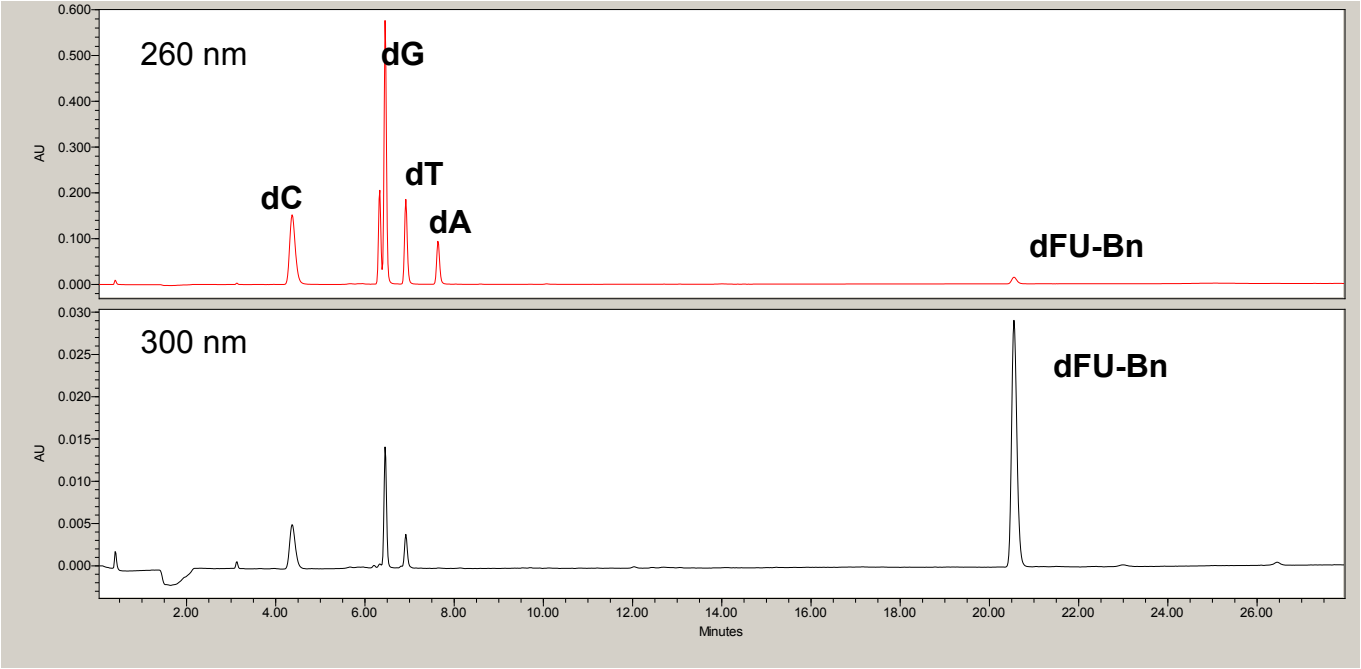


Figure A2.34- C-18 reversed phase HPLC profile of nuclease-digested **dFU-Et**. The column was eluted with a linear gradient of 0-70% buffer B over 30 min (buffer A: 50 mM sodium phosphate, pH 5.8, 2% MeCN and buffer B: 50 mM sodium phosphate, pH 5.8, 50% MeCN).



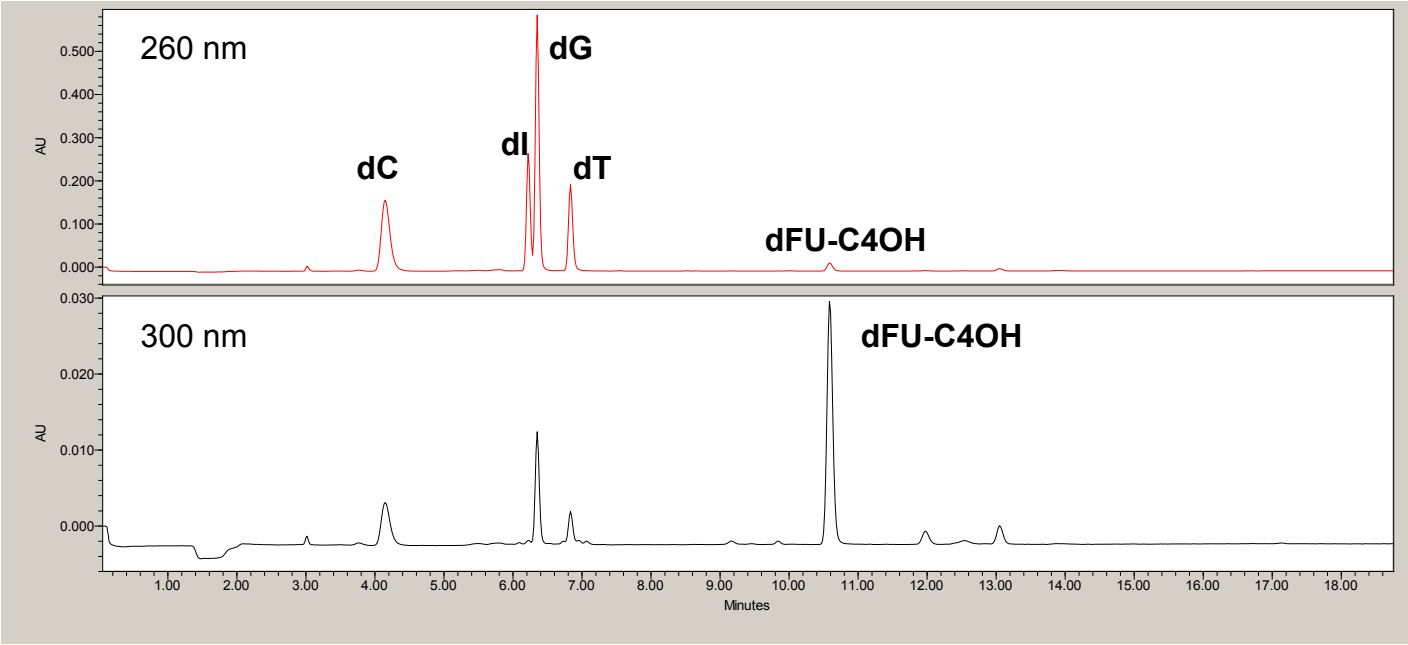
Oligomer	Nucleoside composition	Nucleoside Ratio	
		Expected	Observed
dFU-Et	dC	4	3.8
	dG	4	4.2
	dA (dI)	3	2.8
	dT	2	1.8
	dFU-Et	1	0.4

Figure A2.35- C-18 reversed phase HPLC profile of nuclease-digested **dFU-Bn**. The column was eluted with a linear gradient of 0-70% buffer B over 30 min (buffer A: 50 mM sodium phosphate, pH 5.8, 2% MeCN and buffer B: 50 mM sodium phosphate, pH 5.8, 50% MeCN).



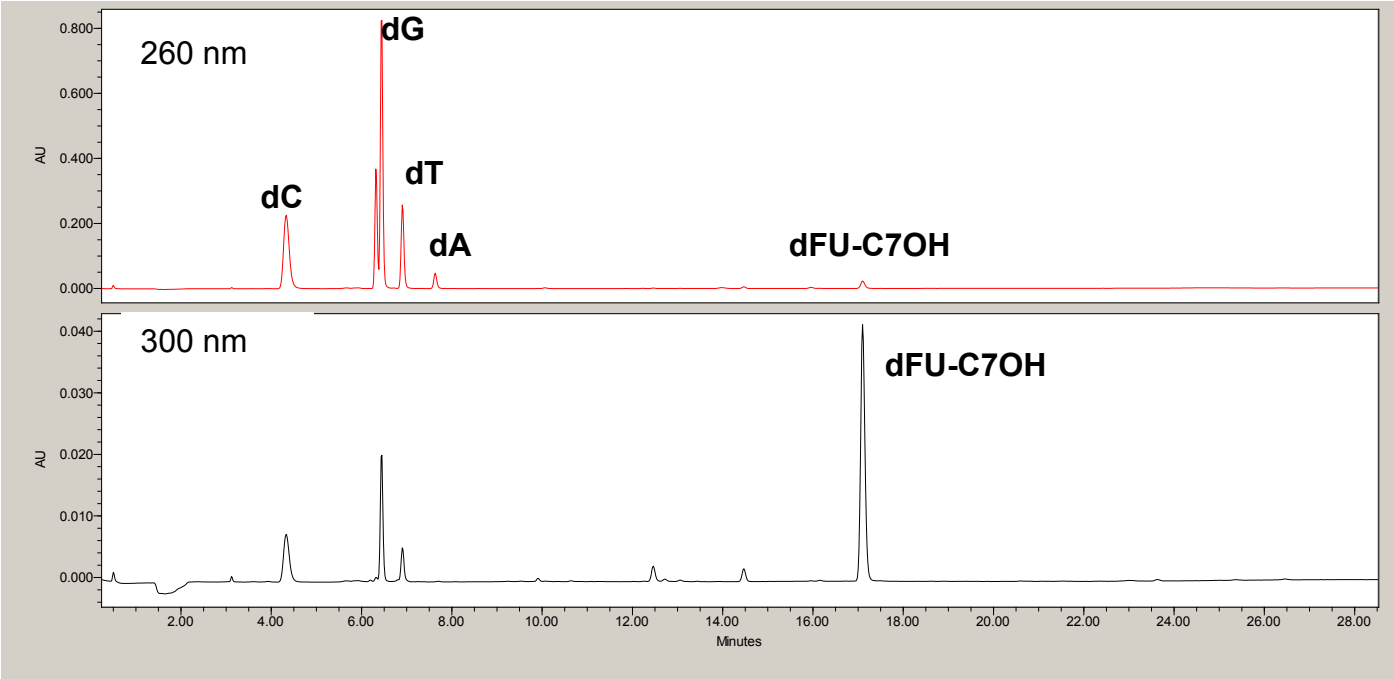
Oligomer	Nucleoside composition	Nucleoside Ratio	
		Expected	Observed
dFU-Bn	dC	4	4.0
	dG	4	4.3
	dA (dI)	3	2.5
	dT	2	1.8
	dFU-Bn	1	0.4

Figure A2.36- C-18 reversed phase HPLC profile of nuclease-digested **dFU-C4OH**. The column was eluted with a linear gradient of 0-70% buffer B over 30 min (buffer A: 50 mM sodium phosphate, pH 5.8, 2% MeCN and buffer B: 50 mM sodium phosphate, pH 5.8, 50% MeCN).



Oligomer	Nucleoside composition	Nucleoside Ratio	
		Expected	Observed
dFU-4COH	dC	4	4.1
	dG	4	3.3
	dA (dI)	3	2.4
	dT	2	2.0
	dFU-C4OH	1	0.3

Figure A2.37- C-18 reversed phase HPLC profile of nuclease-digested **dFU-C7OH**. The column was eluted with a linear gradient of 0-70% buffer B over 30 min (buffer A: 50 mM sodium phosphate, pH 5.8, 2% MeCN and buffer B: 50 mM sodium phosphate, pH 5.8, 50% MeCN).



Oligomer	Nucleoside composition	Nucleoside Ratio	
		Expected	Observed
dFU-7COH	dC	4	3.9
	dG	4	4.1
	dA (dI)	3	2.1
	dT	2	2.3
	dFU-C7OH	1	0.3

Figure A2.38- SAX-HPLC profile of **dFU-Et** crude **(A)** and pure **(B)**. The column (Dionex DNAPAC PA- 100) was eluted using a linear gradient of 0-70% buffer B over 30 min (buffer A: 100 mM Tris HCl, pH 7.5, 10% MeCN and buffer B: 100 mM Tris HCl, pH 7.5, 10% MeCN, 1 M NaCl)

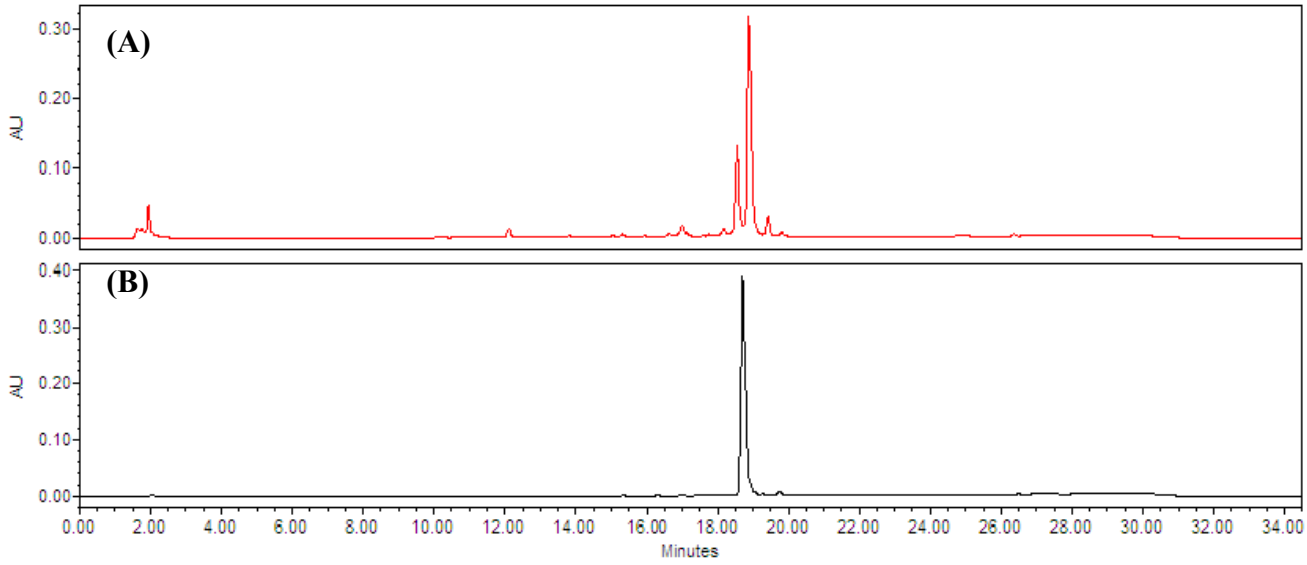


Figure A2.39- SAX-HPLC profile of **dFU-Bn** crude **(A)** and pure **(B)**. The column (Dionex DNAPAC PA- 100) was eluted using a linear gradient of 0-70% buffer B over 30 min (buffer A: 100 mM Tris HCl, pH 7.5, 10% MeCN and buffer B: 100 mM Tris HCl, pH 7.5, 10% MeCN, 1 M NaCl)

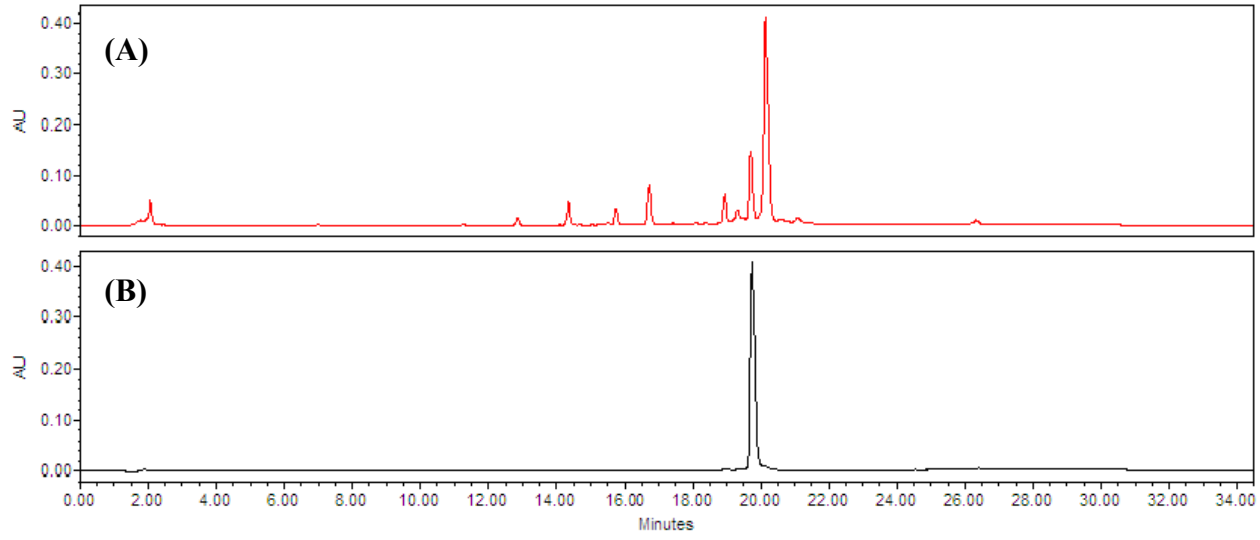


Figure A2.40- SAX-HPLC profile of **dFU-C4OH** crude **(A)** and pure **(B)**. The column (Dionex DNAPAC PA- 100) was eluted using a linear gradient of 0-70% buffer B over 30 min (buffer A: 100 mM Tris HCl, pH 7.5, 10% MeCN and buffer B: 100 mM Tris HCl, pH 7.5, 10% MeCN, 1 M NaCl)

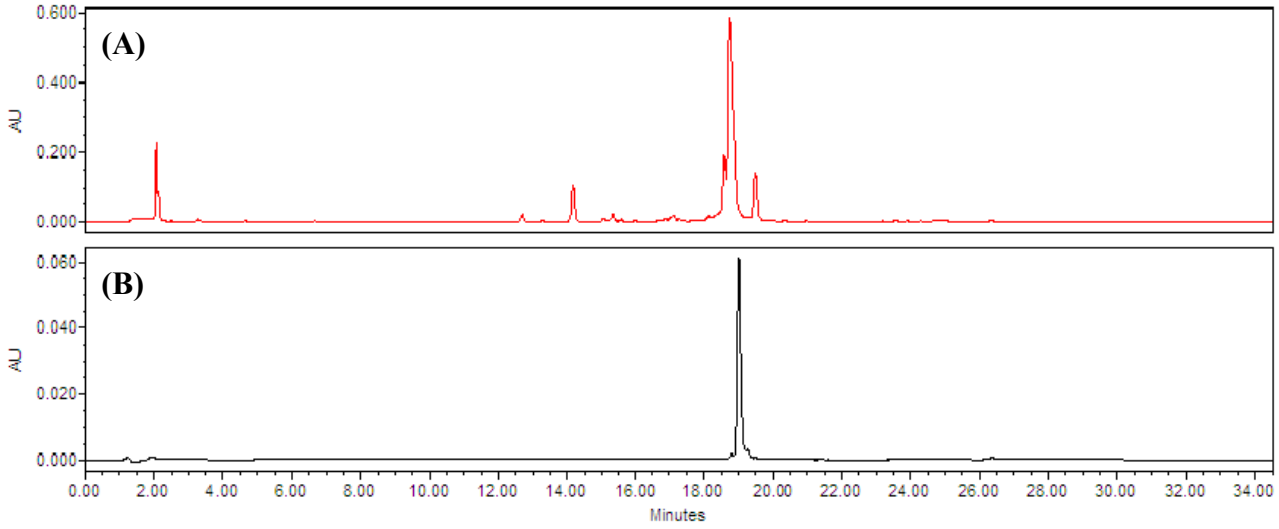


Figure A2.41- SAX-HPLC profile of **dFU-C7OH** crude **(A)** and pure **(B)**. The column (Dionex DNAPAC PA- 100) was eluted using a linear gradient of 0-70% buffer B over 30 min (buffer A: 100 mM Tris HCl, pH 7.5, 10% MeCN and buffer B: 100 mM Tris HCl, pH 7.5, 10% MeCN, 1 M NaCl)

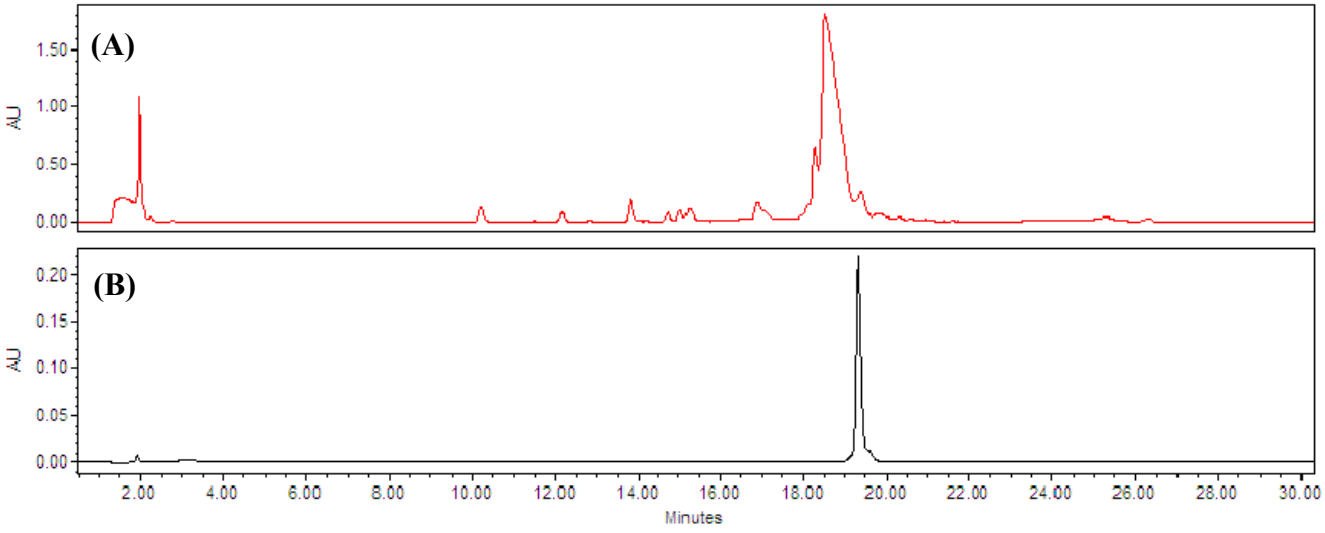


Figure A2.42 - Hyperchromicity change (A_{260}) *versus* temperature (^{o}C) profiles of duplexes containing **dFU** control (—), **dFU-Et** (•••), **dFU-Bn** (———), **dFU-C4OH** (—•—) and **dFU-C7OH** (—••—).

Figure AII.43 - Circular dichroism spectra of duplexes containing **dFU** control (—), **dFU-Et** (•••), **dFU-Bn** (———), **dFU-C4OH** (—•—) and **dFU-C7OH** (—••—).

Figure A2.44 - Repair of (**A**) **dFU-**Et, (**B**) **dFU-Bn**, (**C**) **dFU-C4OH** and (**D**) **dFU-C7OH** by hAGT, OGT, Ada-C, and hOGT for 2.5 h at 37^{o}C. Denaturing PAGE of repair reactions as described in the experimental section with 5-fold protein concentration. Panel **A**; Lane 1, 2 pmol **dFU-Et** DNA; lane 2, 2 pmol **dFU-Et** + *BclI*; lane 3, 2 pmol **dFU- Et** + 10 pmol hAGT; lane 4, 2 pmol **dFU-Et** + 10 pmol hAGT + *BclI*; lane 5, 2 pmol **dFU-Et** + 10 pmol OGT; lane 6, 2 pmol **dFU-Et** + 10 pmol OGT+ *BclI*; lane 7, 2 pmol **dFU-Et** + 10 pmol Ada-C; lane 8, 2 pmol **dFU-Et** + 10 pmol Ada-C + *BclI*; lane 9, 2 pmol **dFU-Et** + 10 pmol hOGT; lane 10, 2 pmol **dFU-Et** + 10 pmol hOGT + *BclI*. Panel **B**; Identical to panel **A** except with **dFU-Bn** instead of **dFU-Et**. Panel **C**; Identical to panel **A** except with **dFU-C4OH** instead of **dFU-Et**. Panel **D**; Identical to panel **A** except with **dFU-C7OH** instead of **dFU-Et**.

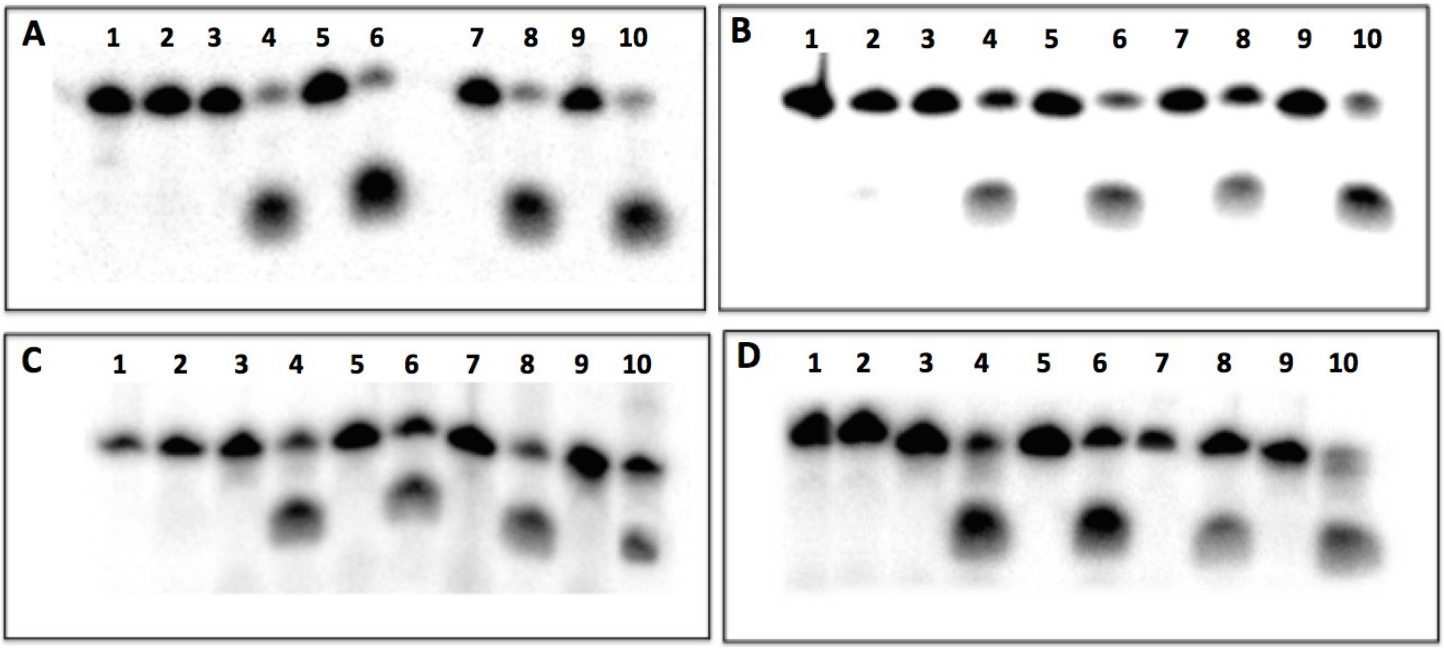


Figure A2.45 - Time course repair gel of duplexes containing **dFU-Et** by (**A**) **hAGT**, (**B**) **OGT**, (**C**) **Ada-C** and (**D**) **hOGT**. (**A**) Denaturing gel of the repair of 2 pmol of **dFU-Et** by 10 pmol hAGT as a function of time: lane 1, 2 pmol Control; lanes 2-10, 2 pmol + 10 pmol hAGT incubated for 0.25, 0.5, 0.75, 1, 1.5, 3, 5, 15, 45 min, respectively. (**B**) Denaturing gel of the repair of 2 pmol of **dFU-Et** by 10 pmol OGT as a function of time: lane 1, 2 pmol Control; lanes 2-10, 2 pmol + 10 pmol OGT incubated for 0.25, 0.5, 0.75, 1, 1.5, 2, 5, 15, 45 min, respectively. (**C**) Denaturing gel of the repair of 2 pmol of **dFU-Et** by 10 pmol Ada-C as a function of time: lane 1, 2 pmol Control; lanes 2-10, 2 pmol + 10 pmol Ada-C incubated for 0.25, 0.5, 0.75, 1, 1.5, 2, 3, 5, 15, 45 min, respectively. (**D**) Denaturing gel of the repair of 2 pmol of **dFU-Et** by 10 pmol hOGT as a function of time: lane 1, 2 pmol Control; lanes 2-10, 2 pmol + 10 pmol hOGT incubated for 0.25, 0.5, 0.75, 1, 1.5, 2, 3, 5, 15, 45 min, respectively. All the time courses were performed at room temperature.

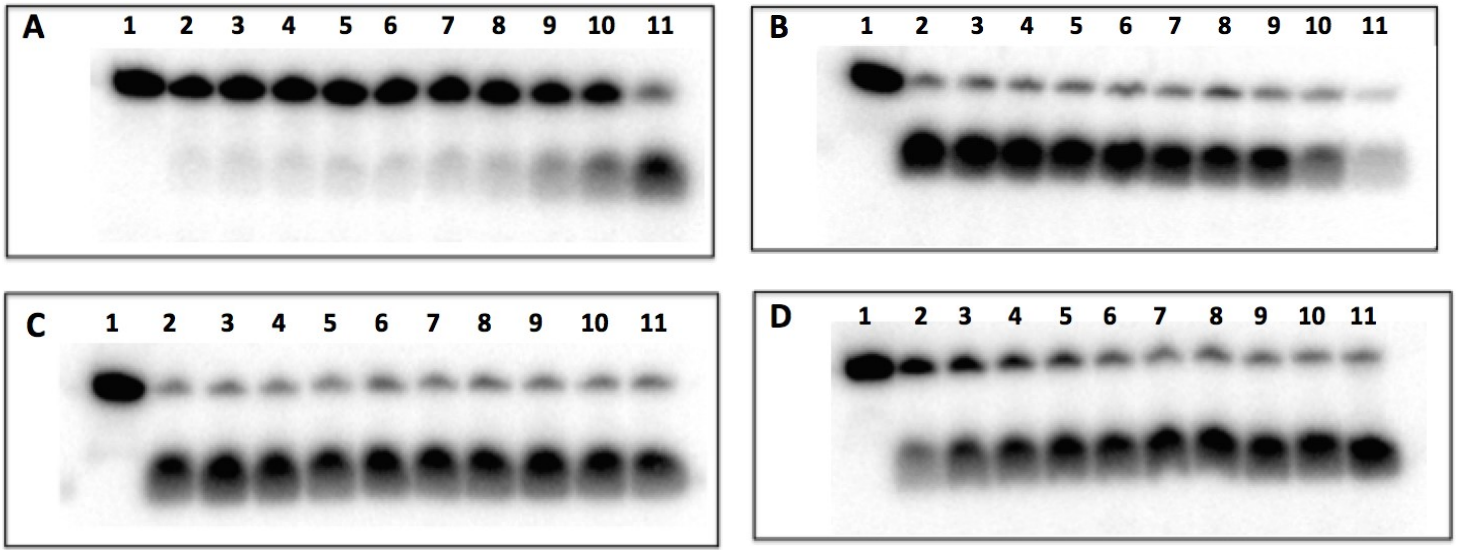


Figure A2.46 - Time course repair gel of duplexes containing **dFU-Bn** by (**A**) **hAGT**, (**B**) **OGT**, (**C**) **Ada-C** and (**D**) **hOGT**. (**A**) Denaturing gel of the repair of 2 pmol of **dFU-Bn** by 10 pmol hAGT as a function of time: lane 1, 2 pmol Control; lanes 2-10, 2 pmol + 10 pmol hAGT incubated for 0.25, 0.5, 0.75, 1, 1.5, 2, 5, 15, 45 min, respectively. (**B**) Denaturing gel of the repair of 2 pmol of **dFU-Bn** by 10 pmol OGT as a function of time: lane 1, 2 pmol Control; lanes 2-10, 2 pmol + 10 pmol OGT incubated for 0.25, 0.5, 0.75, 1, 1.5, 3, 5, 15, 45 min, respectively. (**C**) Denaturing gel of the repair of 2 pmol of **dFU-Bn** by 10 pmol Ada-C as a function of time: lane 1, 2 pmol Control; lanes 2-10, 2 pmol + 10 pmol Ada-C incubated for 0.25, 0.5, 0.75, 1, 1.5, 2, 3, 5, 15, 45 min, respectively. (**D**) Denaturing gel of the repair of 2 pmol of **dFU-Bn** by 10 pmol hOGT as a function of time: lane 1, 2 pmol Control; lanes 2-10, 2 pmol + 10 pmol hOGT incubated for 0.25, 0.5, 0.75, 1, 1.5, 2, 3, 5, 15, 45 min, respectively. All the time courses were performed at room temperature.

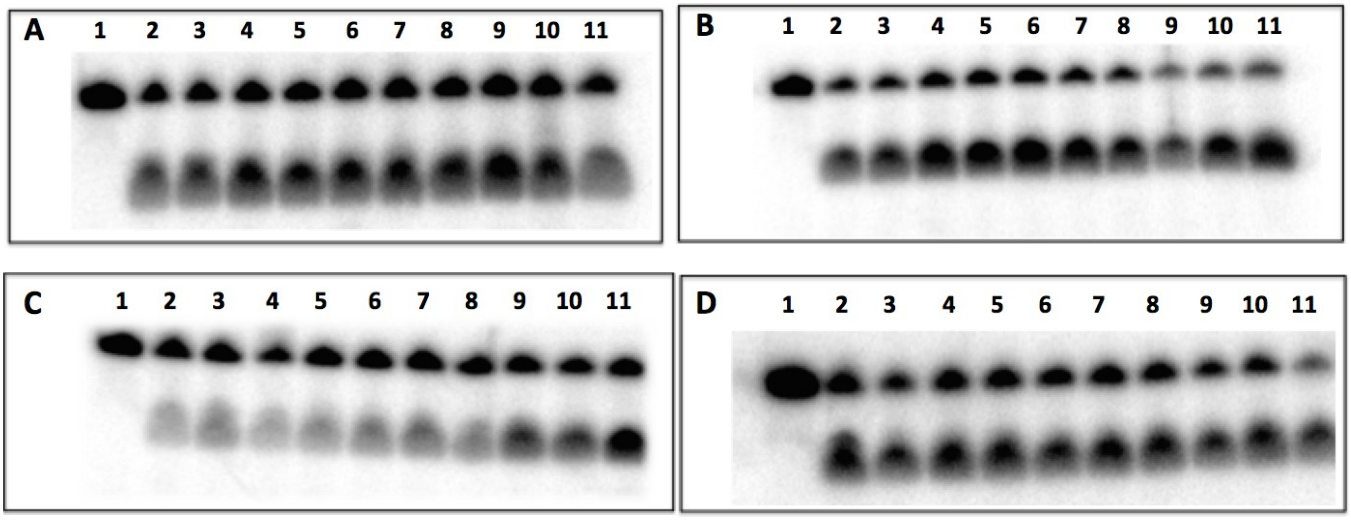


Figure A2.47 - Time course repair gel of duplexes containing **dFU-C4OH** by (**A**) **hAGT**, (**B**) **OGT**, (**C**) **Ada-C** and (**D**) **hOGT**. (**A**) Denaturing gel of the repair of 2 pmol of **dFU-C4OH** by 10 pmol hAGT as a function of time: lane 1, 2 pmol Control; lanes 2-10, 2 pmol + 10 pmol hAGT incubated for 0.25, 0.5, 0.75, 1, 1.5, 3, 5, 15, 45 min, respectively. (**B**) Denaturing gel of the repair of 2 pmol of **dFU-C4OH** by 10 pmol OGT as a function of time: lane 1, 2 pmol Control; lanes 2-10, 2 pmol + 10 pmol OGT incubated for 0.25, 0.5, 0.75, 1, 1.5, 2, 5, 15, 45 min, respectively. (**C**) Denaturing gel of the repair of 2 pmol of **dFU-C4OH** by 10 pmol Ada-C as a function of time: lane 1, 2 pmol Control; lanes 2-10, 2 pmol + 10 pmol Ada-C incubated for 0.25, 0.5, 0.75, 1, 1.5, 2, 3, 5, 15, 45 min, respectively. (**D**) Denaturing gel of the repair of 2 pmol of **dFU-C4OH** by 10 pmol hOGT as a function of time: lane 1, 2 pmol Control; lanes 2-10, 2 pmol + 10 pmol hOGT incubated for 0.25, 0.5, 0.75, 1, 1.5, 3, 5, 15, 45, 150 min, respectively. All the time courses were performed at room temperature.

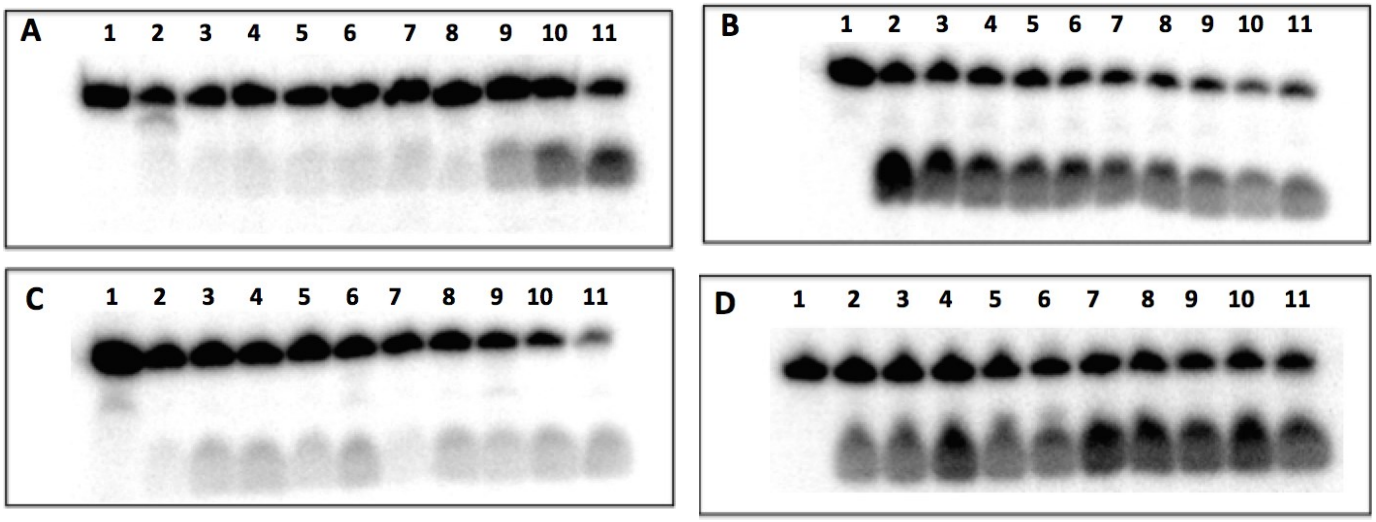


Figure A2.48 - Time course repair gel of duplexes containing **dFU-C7OH** by (**A**) **hAGT**, (**B**) **OGT**, (**C**) **Ada-C** and (**D**) **hOGT**. (**A**) Denaturing gel of the repair of 2 pmol of **dFU-C7OH** by 10 pmol hAGT as a function of time: lane 1, 2 pmol Control; lanes 2-10, 2 pmol + 10 pmol hAGT incubated for 0, 0.25, 0.5, 0.75, 1, 1.5, 3, 5, 15, 45, 150 min, respectively. (**B**) Denaturing gel of the repair of 2 pmol of **dFU-C7OH** by 10 pmol OGT as a function of time: lane 1, 2 pmol Control; lanes 2-10, 2 pmol + 10 pmol OGT incubated for 0, 0.25, 0.5, 0.75, 1, 1.5, 3, 5, 15, 45, 150 min, respectively. (**C**) Denaturing gel of the repair of 2 pmol of **dFU-C7OH** by 10 pmol Ada-C as a function of time: lane 1, 2 pmol Control; lanes 2-10, 2 pmol + 10 pmol Ada-C incubated for 0, 0.25, 0.5, 0.75, 1, 1.5, 3, 5, 15, 45, 150 min, respectively. (**D**) Denaturing gel of the repair of 2 pmol of **dFU-C7OH** by 10 pmol hOGT as a function of time: lane 1, 2 pmol Control; lanes 2-10, 2 pmol + 10 pmol hOGT incubated for 0, 0.25, 0.5, 0.75, 1, 1.5, 3, 5, 15, 45, 150 min, respectively. All the time courses were performed at room temperature.

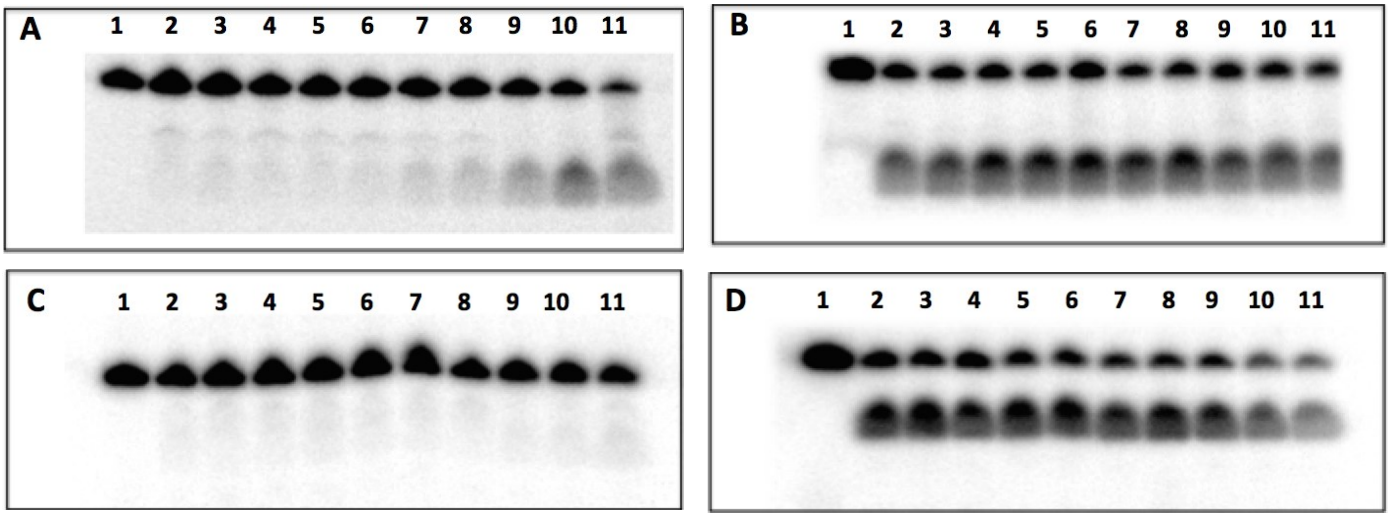


Figure A2.49 – Time course of repair (2.5 h) graphs of **dFU-Et** (•••), **dFU-Bn** (———), **dFU-C4OH** (—•—) and **dFU-C7OH** (—••—) at room temperature: **A)** by hAGT, **B)** by OGT, **C)** by Ada-C and **D)** the chimera hOGT.

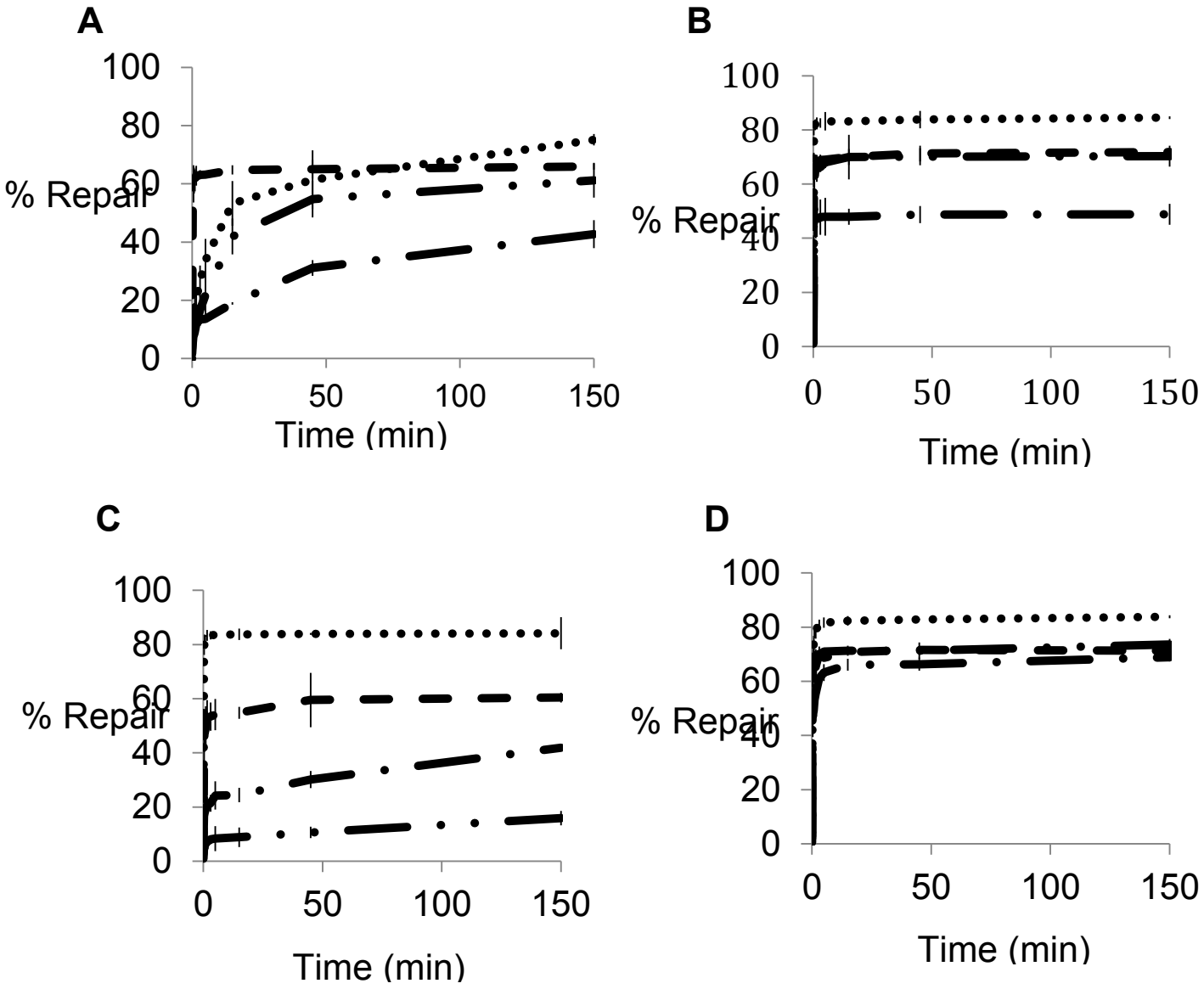
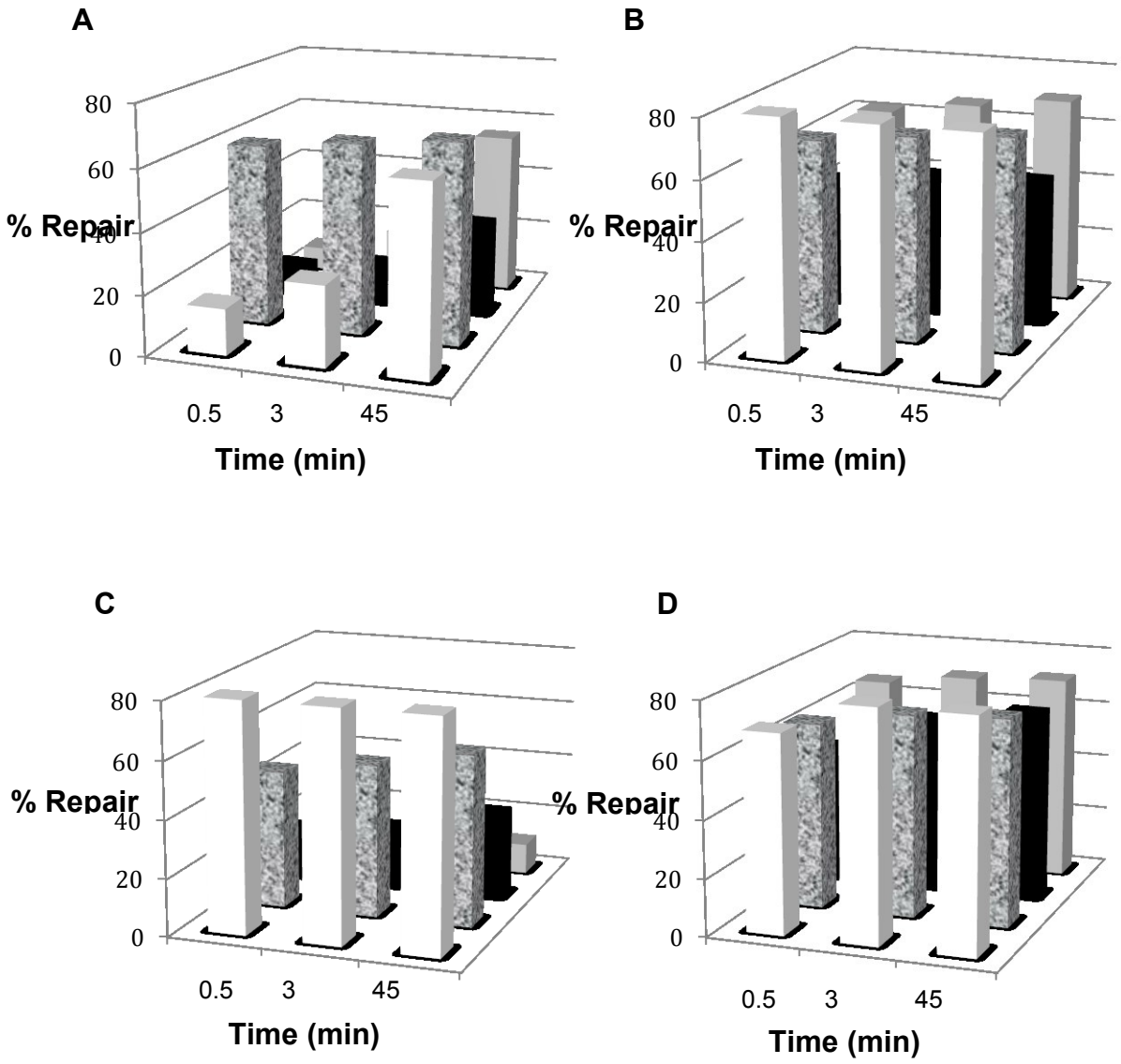


Figure A2.50 – Repair of **dFU-Et** (white bars), **dFU-Bn** (granite bars), **dFU-C4OH** (black bars) and **dFU-C7OH** (light grey bars) by (**A**) hAGT, (**B**) OGT, (**C**) Ada-C, and (**D**) hOGT for different time points.



Appendix III: Supporting Information for Chapter 4

Contents **Page**

Supporting Methods

A3.1 Synthesis and characterization of **dClU** and **dBrU** nucleosides and oligonucleotides **169**
A3.2 UV thermal denaturation **177**
A3.3 Circular dichroism (CD) spectroscopy **177**
A3.4 AGT repair assay of **dClU-Me** and **dBrU-Me** DNA duplexes **177**

Supporting Figures

Figure A3.1- 500 MHz ^{1}H NMR spectrum of compound (**3a**) (in d_6-acetone) **178**
Figure A3.2- 125.7 MHz ^{13}C NMR spectrum of compound (**3a**) (in d_6-acetone) **178**
Figure A3.3 - 202.3 MHz ^{31}P NMR spectrum of compound (**3a**) (in d_6-acetone) **180**
Figure A3.4 - 500 MHz ^{1}H NMR spectrum of compound (**4a**) (in $CDCl_3$) **181**
Figure A3.5 - 125.7 MHz ^{13}C NMR spectrum of compound (**4a**) (in $CDCl_3$) **182**
Figure A3.6 - 500 MHz ^{1}H NMR spectrum of compound (**5a**) (in $CDCl_3$) **183**
Figure A3.7 - 125.7 MHz ^{13}C NMR spectrum of compound (**5a**) (in $CDCl_3$) **184**
Figure A3.8 - 500 MHz ^{1}H NMR spectrum of compound (**6a**) (in d_6-acetone) **185**
Figure A3.9- 125.7 MHz ^{13}C NMR spectrum of compound (**6a**) (in d_6-acetone) **186**
Figure A3.10- 202.3 MHz ^{31}P NMR spectrum of compound (**6a**) (in d_6-acetone) **187**
Figure A3.11- 500 MHz ^{1}H NMR spectrum of compound (**3b**) (in d_6-acetone) **188**
Figure A3.12 - 125.7 MHz ^{13}C NMR spectrum of compound (**3b**) (in d_6-acetone) **189**
Figure A3.13 - 202.3 MHz ^{31}P NMR spectrum of compound (**3b**) (in d_6-acetone) **190**
Figure A3.14- 500 MHz ^{1}H NMR spectrum of compound (**4b**) (in $CDCl_3$) **191**
Figure A3.15- 125.7 MHz ^{13}C NMR spectrum of compound (**4b**) (in $CDCl_3$) **192**
Figure A3.16- 500 MHz ^{1}H NMR spectrum of compound (**5b**) (in $CDCl_3$) **193**
Figure A3.17– 125.7 MHz ^{13}C NMR spectrum of compound (**5b**) (in $CDCl_3$) **194**
Figure A3.18- 500 MHz ^{1}H NMR spectrum of compound (**6b**) (in d_6-acetone) **195**
Figure A3.19- 125.7 MHz ^{13}C NMR spectrum of compound (**6b**) (in d_6-acetone) **196**
Figure A3.20- 202.3 MHz ^{31}P NMR spectrum of compound (**6b**) (in d_6-acetone) **197**
Figure A3.21- ESI MS spectrum of oligonucleotide **dClU** **198**
Figure A3.22 - ESI MS spectrum of oligonucleotide **dClU-Me** **199**
Figure A3.23 - ESI MS spectrum of oligonucleotide **dBrU** **200**
Figure A3.24 - ESI MS spectrum of oligonucleotide **dBrU-Me** **201**
Figure A3.25- C-18 RP-HPLC profile of nuclease digested **dClU-Me** **202**
Figure A3.26- C-18 RP-HPLC profile of nuclease digested **dBrU-Me** **203**
Figure A3.27- SAX-HPLC profile of crude and pure **dClU** **204**

Figure A3.28- SAX-HPLC profile of crude and pure **dClU-Me** **204**
Figure A3.29- SAX-HPLC profile of crude and pure **dBrU** **205**

Figure A3.30- SAX-HPLC profile of crude and pure **dBrU-Me** **205**

Figure A3.31 - UV thermal denaturation experiments for DNA duplex containing **dClU**, **dClU-Me**, **dBrU** and **dBrU-Me** **206**
Figure A3.32 - Circular dichroism spectra for DNA duplex containing **dClU**, **dClU-Me**, **dBrU** and **dBrU-Me** **206**
Figure A3.33 - Total repair gel of duplexes containing (**A**) **dClU** and (**B**) **dClU-Me** by hAGT, OGT, Ada-C, and hOGT for 2.5h at 37^{o}C **207**
Figure A3.34 - Total repair gel of duplexes containing (**A**) **dBrU** and (**B**) **dBrU-Me** by hAGT, OGT, Ada-C, and hOGT for 2.5h at 37^{o}C **208**
Figure A3.35 - Time course repair gel of duplexes containing **dClU-Me** by (A) hAGT, (B) OGT, (C) Ada-C and (D) hOGT **209**
Figure A3.36 - Time course repair gel of duplexes containing **dBrU-Me** by (A) hAGT, (B) OGT, (C) Ada-C and (D) hOGT **210**
Figure A3.37- Time course repair graphs of (**A**) **dClU-Me** and (**B**) **dBrU-Me** by hAGT, OGT, Ada-C, and hOGT for 45 min. **211**

A3.1 Synthesis and characterization of nucleosides and oligonucleotides

A3.1.1 General

5-chloro-2'-deoxyuridine and 5-bromo-2'-deoxyuridine were purchased from the Cayman Chemical Company (Ann Arbor, Michigan) and Carbosynth Limited (Berkshire, UK) respectively. "Fast deprotecting" 5'-*O*-dimethoxytrityl-2'-deoxyribonucleoside-3'-*O*-(β-cyanoethyl-N,N-diisopropyl)phosphoramidites and protected 2'-deoxyribonucleoside–CPG supports were purchased from Glen Research (Sterling, Virginia). Compounds **2a** and **2b** were prepared according to previously published procedures.[54,97,98] All other chemicals and solvents were purchased from the Aldrich Chemical Company (Milwaukee, WI) or EMD Chemicals Inc. (Gibbstown, NJ). Flash column chromatography was performed using silica gel 60 (230–400 mesh) purchased from Silicycle (Quebec City, QC). Thin layer chromatography (TLC) was carried out with precoated TLC plates (Merck, Kieselgel 60 F_{254}, 0.25 mm) purchased from EMD Chemicals Inc. (Gibbstown, NJ). NMR spectra were recorded on a Varian 500 MHz NMR spectrometer at room temperature. ^{1}H NMR spectra were recorded at a frequency of 500.0 MHz and chemical shifts were reported in parts per million (ppm) downfield from tetramethylsilane. ^{13}C NMR spectra (^{1}H decoupled) were recorded at a frequency of 125.7 MHz and chemical shifts were reported in ppm with tetramethylsilane as a reference. ^{31}P NMR spectra (^{1}H decoupled) were recorded at a frequency of 202.3 MHz and chemical shifts were reported in ppm with H_3PO_4 used as an external standard. High resolution mass spectrometry of modified nucleosides were obtained using an 7T-LTQ FT ICR instrument (Thermo Scientific), at the Concordia University Centre for Structural and Functional Genomics. The mass spectrometer was operated in full scan, positive ion detection mode. ESI mass spectra for oligonucleotides were obtained at the Concordia University Centre for Biological Applications of Mass Spectrometry (CBAMS) using a Micromass Qtof2 mass spectrometer (Waters) equipped with a nanospray ion source. The mass spectrometer was operated in full scan, negative ion detection mode. T4 polynucleotide kinase (PNK) was obtained from New England BioLabs (NEB). [γ-^{32}P]ATP was purchased from PerkinElmer (Woodbridge, ON). The *BclI* restriction enzyme was obtained from New England Biolabs (Ipswich, MA).

3'-*O*-(2-Cyanoethoxy(diisopropylamino)-phosphino)-5'-*O*-(4,4'-dimethoxytrityl)-5-chloro-2'-deoxyuridine (3a) DIPEA (0.12 mL, 0.69 mmol) was added to a solution of 5'-*O*-(4,4'-dimethoxytrityl)-5-chloro-2'-deoxyuridine (0.130 g, 0.23 mmol) in THF (2.0 mL), followed by the dropwise addition of N,N-diisopropylamino cyanoethyl phosphonamidic chloride (0.123 mL, 0.55 mmol). After 30 min, the solvent was evaporated *in vacuo*, the crude product was taken up in ethyl acetate (40 mL) and washed with a 3% (w/v) solution of $NaHCO_3$ (2 x 50 ml) followed by brine (50 mL). The organic layer was dried over anhydrous Na_2SO_4 (~ 4 g), and concentrated *in vacuo*. The crude product was purified by flash column chromatography using hexanes/ethyl acetate (2:8) as eluent to afford 0.103 g (81%) of a colorless foam. R_f (SiO_2 TLC): 0.88 (100% EtOAc). $\lambda_{max(MeCN)}$ = 275 nm. HRMS (ESI-MS) *m/z* calculated for $C_{39}H_{46}ClN_4NaO_8P^+$: 787.2634; found 787.2644 $[M + Na]^+$. ^{1}H NMR (500 MHz, d_6-acetone, ppm): 8.01-8.03 (2 s, 1H; H6), 7.50-7.53 (m, 2H; Ar), 7.23-7.41 (m, 7H; Ar), 6.89-6.92 (m, 4H; Ar), 6.26-6.30 (m, 1H; H1'), 4.70-4.77 (m, 1H; H3'), 4.17-4.24 (m, 1H; H4'), 3.88-3.92 (m, 2H; CH_2OP), 3.79 (2 s, 6H; ArOMe), 3.74-3.77 (m, 1H; NCH), 3.62-3.69 (m, 1H; NCH), 3.38-3.48 (m, 2H; H5', H5''), 2.76-2.78 (t, J= 6.1 Hz, 1H; CH_2CN), 2.63-2.66 (t, J= 6.1 Hz, 1H; CH_2CN), 2.50-2.58 (m, 2H; H2', H2''), 1.18-1.21 (m, 9H; CH_3), 1.10-1.12 (m, 3H; CH_3). ^{13}C NMR (125.7 MHz, d_6-acetone, ppm): 160.97, 158.83, 156.60, 149.27, 148.65, 144.89, 139.64, 137.15, 137.12, 136.89, 135.52, 135.44, 130.12, 130.09, 128.07, 127.83, 126.76, 117.82, 113.12, 109.98, 108.14, 86.77, 85.58, 85.22, 73.53, 73.14, 64.54, 63.38, 63.18, 58.81, 58.52, 54.63, 51.09, 43.10, 43.00, 39.33, 23.94, 19.76. IR (thin film); ν_{max} (cm^{-1}): 3716, 3208, 2966, 2361, 2340, 1709, 1607, 1548, 1509, 1461, 1445, 1364, 1250, 1178, 1154, 1079, 1033, 977, 827, 790.

3'-*O*-(*tert*-butyldimethylsilyl)-5'-*O*-(4,4'-dimethoxytrityl)-5-chloro-2'-deoxyuridine (4a) Imidazole (0.174 g, 2.55 mmol) and DMAP (approximately 1 mg) were added to a solution of 5'-*O*-(4,4'-dimethoxytrityl)-5-chloro-2'-deoxyuridine (0.400 g, 0.708 mmol) in DCM (15 mL), followed by *tert*-butyldimethylsilyl chloride (0.192 mg, 1.27 mmol). After 4 h, the solvent was evaporated *in vacuo* and the crude product was taken up in DCM (20 mL). The solution was washed with a 3% (w/v) solution of $NaHCO_3$ (3 x 50 ml). The organic layer was dried over anhydrous Na_2SO_4 (~ 4 g), and concentrated *in vacuo*. The crude product was purified by flash column chromatography using hexanes/ethyl acetate (6:4) as eluent to afford 0.430 g (90%) of a

colorless foam. R_f (SiO_2 TLC): 0.62 (1:1 hexanes/EtOAc). $\lambda_{max(MeCN)}$ = 275 nm. HRMS (ESI-MS) *m/z* calculated for $C_{36}H_{43}ClN_2N_aO_7S_i^+$: 701.2420; found 701.2419 $[M + Na]^+$. ^{1}H NMR (500 MHz, $CDCl_3$, ppm): 8.60 (s, 1H; NH), 8.03 (s, 1H; H6), 7.22-7.42 (m, 9H; Ar), 6.82-6.85 (m, 4H; Ar), 6.26-6.29 (m, 1H; H1'), 4.47-4.50 (m, 1H; H3'), 3.99-4.01 (m, 1H; H4'), 3.79 (s, 6H; ArOMe), 3.30-3.40 (m, 2H; H5', H5''), 2.37-2.41 (m, 1H; H2'), 2.17-2.22 (m, 1H; H2''), 0.85 (s, 9H; t-Bu), 0.04 (s, 3H; CH_3), 0.01 (s, 3H; CH_3). ^{13}C NMR (125.7 MHz, $CDCl_3$, ppm): 158.68, 149.02, 144.22, 136.79, 135.25, 129.95, 128.01, 127.95, 127.06, 113.31, 109.22, 87.31, 87.08, 85.83, 76.99, 76.74, 72.28, 62.95, 60.37, 55.23, 41.93, 25.69, 21.02, 17.93, 14.18. IR (thin film); ν_{max} (cm^{-1}) = 3648, 3420, 3060, 2953, 2930, 2856, 2361, 1700, 1608, 1508, 1457, 1252, 1177, 1104, 1034, 832, 778, 736, 701.

5'-*O*-(4,4'-dimethoxytrityl)-5-chloro-*O*^{4}-methyl-2'-deoxyuridine (5a): To a solution of compound (**4a**) in anhydrous MeCN (4.5 mL) was added triazole (0.713 g, 10.34 mmol) and triethylamine (1.48 ml, 10.56 mmol) while stirring at 0^{o}C. Then, $POCl_3$ (0.107 mL, 1.15 mmol) was added dropwise. After 16 h, the solvent was evaporated *in vacuo* and the crude product was taken up in CH_2Cl_2 (40 mL), washed with a 3% (w/v) solution of $NaHCO_3$ (2 x 50 ml), dried over anhydrous Na_2SO_4 (~ 4 g), and concentrated *in vacuo*. The crude was taken up in MeOH (7 mL) and a solution of NaOMe (0.087 g, 1.61 mmol) in MeOH (7 mL) was added. After 2 h, the solvent was evaporated *in vacuo* and the crude product was taken up in CH_2Cl_2 (30 mL), washed with a 3% (w/v) solution of $NaHCO_3$ (3 x 50 ml), dried over anhydrous Na_2SO_4 (~ 4 g), and concentrated *in vacuo*. The crude product was taken up in THF (2.5 mL) and TBAF (1 M in THF) (0.144 mL, 0.551 mmol) was added drop-wise. After 30min, the solvent was evaporated *in vacuo* and the crude product was taken up in CH_2Cl_2 (40 mL), washed with a 3% (w/v) solution of $NaHCO_3$ (2 x 50 ml), dried over anhydrous Na_2SO_4 (~ 4 g), and concentrated *in vacuo*. The crude product was purified by flash column chromatography using a gradient of hexanes/EtOAc (3:7 → 100% EtOAc) as eluent to afford 0.087 g (34%) of colorless foam. R_f(SiO_2 TLC): 0.64 (100% EtOAc). $\lambda_{max(MeCN)}$ = 282 nm. HRMS (ESI-MS) *m/z* calculated for $C_{31}H_{31}ClN_2NaO_7^+$: 601.1712; found 601.1732 $[M + Na]^+$. ^{1}H NMR (500 MHz, $CDCl_3$, ppm): 8.19 (s, 1H; H6), 7.20-7.40 (m, 9H; Ar), 6.80-6.83 (m, 4H; Ar), 6.27-6.29 (m, 1H; H1'), 4.55-4.57 (m, 1H; H3'), 4.17-4.19 (m, 1H; H4'), 4.01 (s, 3H; OCH_3), 3.77 (s, 6H; ArOMe), 3.38-3.39 (m, 2H; H5', H5''), 2.73-2.77 (m, 1H; H2'), 2.21-2.28 (m, 1H; H2''). ^{13}C NMR (125.7 MHz, $CDCl_3$, ppm): 166.79, 158.61, 154.47, 144.38,

140.35, 135.36, 129.99, 127.98, 127.93, 126.97, 113.31, 113.29, 102.54, 87.41, 86.97, 86.78, 77.05, 76.79, 71.95, 63.36, 60.41, 55.50, 55.22, 53.42, 42.25, 21.03, 20.68, 14.17, 14.00. IR (thin film); ν_{max} (cm^{-1}) = 3439, 2753, 2364, 2348, 2335, 2283, 1645, 1562, 1543, 1510, 1250, 1176, 663.

3'-*O*-(2-Cyanoethoxy(diisopropylamino)-phosphino)-5'-*O*-(4,4'-dimethoxytrityl)-5-chloro-O^4-methyl-2'-deoxyuridine (6a): DIPEA (0.078 mL, 0.50 mmol) was added to a solution of (**5a**) (0.087 g, 0.15 mmol) in THF (1.5 mL), followed by the dropwise addition of N,N-diisopropylamino cyanoethyl phosphonamidic chloride (0.080 mL, 0.36 mmol). After 30 min, the solvent was evaporated *in vacuo* and the crude product was taken up in ethyl acetate (40 mL). The solution was washed with a 3% (w/v) solution of $NaHCO_3$ (2 x 50 ml) and then brine (50 mL). The organic layer was dried over anhydrous Na_2SO_4 (~ 4 g), and concentrated *in vacuo*. The crude product was purified by flash column chromatography using hexanes/ethyl acetate (2:8) as eluent to afford 0.094 g (76%) of a colorless foam. R_f (SiO_2 TLC): 0.71, 0.79 (100% EtOAc). $\lambda_{max(MeCN)}$ = 283 nm. HRMS (ESI-MS) *m/z* calculated for $C_{40}H_{49}ClN_4O_8P^+$: 779.2971; found 779.2994 $[M + H]^+$. ^{1}H NMR (500 MHz, d_6-acetone, ppm): 8.23-8.20 (2 s, 1H; H6), 7.22-7.52 (m, 9H; Ar), 6.88-6.92 (m, 4H; Ar), 6.18-6.22 (m, 1H; H1'), 4.68-4.76 (m, 1H; H3'), 4.22-4.29 (m, 1H; H4'), 3.96 (s, 3H; OCH_3), 3.81-3.92 (m, 2H; CH_2OP), 3.78-3.79 (s, 6H; $ArOCH_3$), 3.61-3.69 (m, 2H; NCH,), 3.40-3.49 (m, 2H; H5', H5''), 2.70-2.72 (m, 2H; CH_2CN), 2.20-2.27 (m, 2H; H2', H2''), 1.18-1.21 (m, 9H; CH_3), 1.11-1.12 (m, 3H; CH_3). ^{13}C NMR (125.7 MHz, d_6-acetone, ppm): 166.78, 158.62, 154.42, 144.36, 140.30, 135.35, 129.99, 127.99, 127.93, 126.98, 113.31, 113.30, 102.48, 87.35, 86.99, 86.71, 77.02, 76.76, 71.96, 65.82, 63.31, 55.50, 55.22, 52.01, 42.21, 20.43, 15.24, 13.80. ^{31}P NMR (202.3 MHz, d_6-acetone, ppm): 148.34, 148.13. IR (thin film); ν_{max} (cm^{-1}) =3683, 2966, 2361, 2251, 1710, 1607, 1509, 1461, 1445, 1365, 1250, 1178, 1079, 1033, 827, 726, 702.

3'-*O*-(2-Cyanoethoxy(diisopropylamino)-phosphino)-5'-*O*-(4,4'-dimethoxytrityl)-5-bromo-2'-deoxyuridine (3b): DIPEA (0.297 mL, 1.686 mmol) was added to a solution of 5'-*O*-(4,4'-dimethoxytrityl)-5-bromo-2'-deoxyuridine (0.3433 g, 0.562 mmol) in THF (4 mL), followed by the dropwise addition of N,N-diisopropylamino cyanoethyl phosphonamidic chloride (0.298 mL, 1.348 mmol). After 30 min, the solvent was evaporated *in vacuo* and the crude product was

taken up in ethyl acetate containing few drops of triethylamine (40 mL). The solution was washed with a 3% (w/v) solution of $NaHCO_3$ (3 x 50 mL) followed by brine (50 mL). The organic layer was dried over anhydrous Na_2SO_4 (~ 4 g), and reconcentrated *in vacuo*. The crude product was purified by flash column chromatography with hexanes/ethyl acetate (2:8) as eluent to afford 0.3425 g (80%) of a colorless foam. R_f (SiO_2 TLC): 0.833 (8:2 EtOAc/Hexane). $\lambda_{max(MeCN)}$ = 276 nm. HRMS (ESI-MS) *m/z* calculated for $C_{39}H_{46}BrKN_4O_8P^+$: 847.1868; found 847.2090 $[M + K]^+$. ^1H NMR (500 MHz, d_6-acetone, ppm): 8.12 (s, 0.51H; H6), 8.10 (s, 0.43H; H6), 7.39-7.42 (m, 9H; Ar), 6.90-6.92 (m, 4H; Ar), 6.27-6.31 (m, 1H; H1'), 4.71-4.74 (m, 1H; H3'), 4.19-4.25 (m, 1H; H4'), 3.61-3.89 (m, 10H; NCH, $ArOCH_3$, CH_2OP), 3.41-3.45 (m, 2H; H5', H5''), 2.75-2.77 (t, J= 6.0 Hz, 1H; CH_2CN), 2.63-2.65 (t, J= 6.0 Hz, 1H; CH_2CN), 2.52-2.54 (m, 1H; H2'), 2.04-2.06 (m, 1H; H2''), 1.18-1.21 (m, 10H; CH_3), 1.10-1.12 (m, 2H; CH_3). ^{13}C NMR (125.7 MHz, d_6-acetone, ppm): 170.01, 158.83, 158.81, 158.64, 158.62, 149.61, 144.95, 139.47, 139.41, 135.70, 135.67, 135.60, 135.55, 130.15, 130.14, 130.12, 130.11, 128.10, 128.05, 127.87, 127.86, 126.80, 126.78, 118.05, 117.91, 113.16, 96.32, 96.27, 86.72, 86.67, 85.67, 85.62, 85.45, 85.39, 85.37, 7370, 73.57, 73.38, 73.24, 63.38, 63.22, 59.66, 58.74, 58.70, 58.59, 58.55, 54.69, 54.67, 54.66, 43.12, 43.02, 39.74, 39.71, 39.68, 39.65, 29.45, 29.29, 29.14, 28.98 28.83, 28.68, 28.52, 24.05, 24.02, 24.01, 23.99, 23.96, 23.95, 19.97, 19.90, 19.85, 19.79, 13.64. ^{31}P NMR (202.3 MHz, d_6-acetone, ppm): 148.18, 148.33. IR (thin film); ν_{max} (cm^{-1}) = 3436, 2965, 2364, 2348, 2362, 1648, 1580, 1547, 1444, 1250, 1178, 1032, 662.

3'-*O*-(*tert*-Butyldimethylsilyl)-5'-*O*-(4,4'-dimethoxytrityl)-5-bromo-2'-deoxyuridine (4b)

Imidazole (0.141 g, 2.05 mmol) and DMAP were added to a solution of 5'-*O*-(4,4'-dimethoxytrityl)-5-bromo-2'-deoxyuridine (0.370 g, 0.572 mmol) in DCM (8 mL), followed by the addition of *tert*-butyldimethylsilyl chloride (0.155 mg, 1.03 mmol). After 4 h, the solvent was evaporated *in vacuo* and the crude product was taken up in DCM (20 mL). The solution was washed with a 3% (w/v) solution of $NaHCO_3$ (3 x 50 ml). The organic layer was dried over anhydrous Na_2SO_4 (~ 4 g), and concentrated *in vacuo*. The crude product was purified by flash column chromatography using hexanes/ethyl acetate (7:3) as eluent to afford 0.387 g (85%) of a colorless foam. R_f (SiO_2 TLC): 0.68 (1:1 hexanes/EtOAc). $\lambda_{max(MeCN)}$ = 275 nm. HRMS (ESI-MS) *m/z* calculated for $C_{36}H_{43}BrN_2NaO_7Si^+$: 745.1915; found 745.1914 $[M + Na]^+$. ^1H NMR (500 MHz,

$CDCl_3$, ppm): 8.89 (s, 1H; NH), 8.13 (s, 1H; H_6), 7.21-7.43 (m, 9H; Ar), 6.84-6.85 (m, 4H; Ar), 6.27-6.30 (m, 1H; H1'), 4.47-4.49 (m, 1H; H3'), 3.99-4.01 (m, 1H; H4'), 3.79 (s, 6H; ArOMe), 3.39-3.41 (m, 1H; H5'), 3.29-3.32 (m, 1H; H5''), 2.37-2.41 (m, 1H; H2'), 2.17-2.23 (m, 1H; H2''), 0.85 (s, 9H; t-Bu), 0.03 (s, 3H; CH_3), 0.01 (s, 3H; CH_3). ^{13}C NMR (125.7 MHz, $CDCl_3$, ppm): 158.89, 158.67, 149.40, 144.26, 139.33, 135.31, 130.03, 128.03, 127.98, 127.05, 113.33, 113.32, 96.97, 87.34, 87.05, 85.86, 77.01, 76.75, 72.37, 62.98, 55.24, 41.98, 29.67, 29.64, 29.34, 25.70, 25.63, 22.67, 17.94, 14.10. IR (thin film); ν_{max} (cm^{-1}) = 3056, 2955, 2928, 2849, 2361, 2336, 1716, 1695, 1653, 1608, 1576, 1456, 1265, 1252, 1177, 1099, 1033, 832, 737, 703.

5'-*O*-(4,4'-Dimethoxytrityl)-5-bromo-*O*^{4}-methyl-2'-deoxyuridine (5b): Compound **4b** (2.2733 g, 3.136 mmol) was dissolved in anhydrous MeCN (30 mL). Triazole (4.871 g, 70.57 mmol) was added to the solution at 0^{o}C under stirring followed by triethylamine (10 mL, 72.128 mmol). Then, $POCl_3$ (0.730 mL, 7.840 mmol) was added dropwise. After 4 h, the solvent was evaporated and the reaction was co-evaporated twice with methanol. The crude was taken up in MeOH (7 mL) and a solution of NaOMe (0.592 g, 10.98 mmol) in MeOH (7 mL) was added. After 16 h, the solvent was evaporated *in vacuo* and the crude product was taken up in CH_2Cl_2 (40 mL), washed with a 3% (w/v) solution of $NaHCO_3$ (3 x 50 ml), dried over anhydrous Na_2SO_4 (~ 4 g), and concentrated *in vacuo*. The crude product was taken up in THF (20 mL) and TBAF (1 M in THF) (3.76 mL, 3.76 mmol) was added drop-wise. After 30 min, the solvent was evaporated *in vacuo* and the crude product was taken up in CH_2Cl_2 (40 mL), washed with a 3% (w/v) solution of $NaHCO_3$ (2 x 50 ml), dried over anhydrous Na_2SO_4 (~ 4 g), and concentrated *in vacuo*. The crude product was purified by flash column chromatography using EtOAc/Hexane (7:3) as eluent to afford 1.11 g (71.12%) of a colorless foam. R_f(SiO_2 TLC): 0.261 (7:3 EtOAc/Hexane). $\lambda_{max(MeCN)}$ = 283 nm. HRMS (ESI-MS) *m/z* calculated for $C_{31}H_{31}BrN_2NaO_7^+$: 646.4803; found: 646.4833 [M + Na]$^+$. ^1H NMR (500 MHz, $CDCl_3$, ppm): 8.27 (s, 1H; H6), 7.26–7.41 (m, 9H; Ar), 6.81– 6.84 (m, 4H; Ar), 6.28-6.30 (m, 1H; H1'), 4.55-4.56 (m, 1H; H3'), 4.17-4.18 (m, 1H; H4'), 4.00 (s, 3H; OCH_3) 3.77 (s, 6H; $ArOCH_3$), 3.37-3.38 (m, 2H; H5', H5''), 2.73-2.77 (m, 1H; H2'), 2.25-2.28 (m, 1H; H2''). ^{13}C NMR (125.7 MHz, $CDCl_3$, ppm): 167.24, 158.59, 154.76, 144.43, 142.92, 135.58, 135.42, 130.02, 129.99, 127.99, 127.96, 126.95, 113.31, 88.72, 87.49, 86.93, 86.87, 77.36, 77.11, 76.85,

72.01, 63.45, 55.66, 55.23, 53.46, 42.32, 20.55, 13.91. IR (thin film); ν_{max} (cm^{-1}) = 3396, 2955, 2361, 2336, 1652, 1616, 1507, 1473, 1386, 1301, 1250, 1176, 1096, 1035, 827, 701.

3'-*O*-(2-Cyanoethoxy(diisopropylamino)-phosphino)-5'-*O*-(4,4'-dimethoxytrityl)-5-bromo-*O*^{4}-methyl-2'-deoxyuridine (6b): DIPEA (0.181 mL, 0. 1044 mmol) was added to a solution of (**5b**) (0.2231 g, 0.3483 mmol) in THF (4 mL), followed by the dropwise addition of N,N-diisopropylamino cyanoethyl phosphonamidic chloride (0.154 mL, 0.697 mmol). After 30 min, the solvent was evaporated *in vacuo* and the crude product was taken up in ethyl acetate (40 mL). The solution was washed with a 3% (w/v) solution of $NaHCO_3$ (3 x 50 ml) followed by brine (50 mL). The organic layer was dried over anhydrous Na_2SO_4 (~ 4 g), and concentrated *in vacuo*. The crude product was purified by flash column chromatography using hexanes/ethyl acetate (2:8) solvent as eluent to afford 0.2431 g (83%) of a colorless foam. R_f (SiO_2 TLC): 0.908, 0.857 (2:8 Hexanes/EtOAc). $\lambda_{max(MeCN)}$ = 283 nm. HRMS (ESI-MS) *m/z* calculated for $C_{40}H_{48}BrN_4NaO_8P^+$: 846.6981; found 846.7110[M + Na]$^+$. ^{1}H NMR (500 MHz, $CDCl_3$, ppm): 8.31 (s, 0.43H; H6), 8.26 (s, 0.5H; H6), 7.33–7.51 (m, 9H; Ar), 6.89– 6.92 (m, 4H; Ar), 6.19-6.24 (m, 1H; H1'), 4.68-4.75 (m, 1H; H3'), 4.24-4.33 (m, 1H; H4'), 3.92-3.97 (s, 3H; OCH_3), 3.61-3.91 (m, 10H; NCH, $ArOCH_3$, CH_2OP), 3.44-3.47 (m, 2H; H5', H5''), 2.75-2.77 (t, J= 6.0 Hz, 1H; CH_2CN), 2.63-2.66 (t, J= 6.0 Hz, 1H; CH_2CN), 2.41-2.49 (m, 1H; H2''), 2.04-2.06 (m, 1H; H2'') 1.19-1.22 (m, 9H; CH_3), 1.11-1.12 (m, 3H; CH_3). ^{13}C NMR (125.7 MHz, $CDCl_3$, ppm): 169.99, 168.68, 158.82, 153.30, 144.96, 144.94, 143.25, 143.20, 135.68, 135.65, 135.59, 135.54, 130.15, 130.14, 130.12, 130.10, 128.08, 138.04, 127.92, 127.88, 126.81, 126.79, 118.05, 117.93, 113.18, 87.03, 86.96, 86.89, 86.84, 86.74, 86.70, 85.99, 85.95, 85.72, 85.67, 73.59, 73.45, 73.16, 73.03, 63.20, 62.98, 59.66, 58.73, 58.58, 54.78, 54.70, 54.69, 54.67, 54.64, 43.13, 43.03, 40.65, 40.62, 40.50, 40.47, 29.47, 29.32, 29.26, 29.16, 29.01, 28.85, 28.75, 28.70, 28.55, 28.70, 28.55, 24.08, 24.05, 24.02, 23.99, 23.97, 21.96, 21.94, 20.48, 20.45, 20.00, 19.92, 19.87, 19.81, 13.66. ^{31}P NMR (202.3 MHz, d_6-acetone, ppm): 148.34, 148.08. IR (thin film); ν_{max} (cm^{-1}) =3421, 2966, 2361, 1733, 1717, 1674, 1616, 1559, 1536, 1472, 1457, 1386, 1301, 1251, 1222, 11179, 1118, 1082, 1037, 977, 828, 735.

A3.1.2 Preparation, purification and characterization of the dClU and dBrU oligonucleotides

The sequences containing the modified dClU and dBru nucleosides, which are shown in **Figure 4.1**, were assembled with an Applied Biosystems Model 3400 synthesizer on a 1.5 μmol scale using β-cyanoethylphosphoramidite chemistry supplied by the manufacturer with slight modifications to coupling times. The nucleoside phosphoramidites protected with "fast-deprotecting" groups were dissolved in anhydrous MeCN at a concentration of 0.1 M for the 3'-*O*-deoxyphosphoramidites. Oligomer sequence assembly was carried out as previously described.[54,97] The capping step of the assembly was carried out using phenoxyacetic anhydride/pyridine/tetrahydrofuran 1:1:8 (v/v/v; solution A) and 1-methyl-imidazole/ tetrahydrofuran 16:84 (w/v; solution B). Coupling wait times for phosphoramidites **3a**, **3b**, **6a** and **6b** were extended to 10 min (compared to 2 min for the commercially available phosphoramidites). Protecting group removal and cleavage from the solid support was carried out by treatment with 0.05M K_2CO_4 in MeOH for 4 h at room temperature with mild rocking in 2 mL screw-cap microfuge tubes fitted with Teflon lined caps. The base was neutralized with an equimolar amount of AcOH and crude oligomers were transferred and solvent removed in a Speedvac concentrator. Purification was achieved by strong anion exchange HPLC using a Dionex DNAPAC PA-100 column (0.4 cm x 25 cm) purchased from Dionex Corp, Sunnyvale, CA using a linear gradient of 0–52% buffer B over 24 min (buffer A: 100 mm Tris HCl, pH 7.5, 10% MeCN and buffer B: 100 mm Tris HCl, pH 7.5, 10% MeCN, 1M NaCl) at 55 ^{o}C. The columns were monitored at 260 nm for analytical runs or 280 nm for preparative runs. The purified oligomers were desalted using C-18 SEP PAK cartridges (Waters Inc.) as previously described.[132] The molecular mass of the modified oligomers was identified by deconvolution of the ESI-MS and the measured values were in agreement with the expected masses (**Figures A3.21**, **A3.22**, **A3.23** and **A3.24** for MS spectra). **dClU-Me** and **dBrU-Me** (0.05 A_{260} units each) were also characterized by enzymatic digestion (snake venom phosphodiesterase: 0.28 units and calf intestinal phosphatase: 5 units, in 10 mM Tris, pH 8.1 and 2 mM $MgCl_2$) for 48 h at 37 ^{o}C. The resulting nucleoside mixture was analyzed by reversed phase HPLC carried out using a

Symmetry® C-18 5 µm column (0.46 x 15 cm) via a linear gradient of 0-70% buffer B over 30 min (buffer A: 50 mM sodium phosphate, pH 5.8, 2% MeCN and buffer B: 50 mM sodium phosphate, pH 5.8, 50% MeCN). Results and nucleoside ratios from the digestion are shown in **Figures A3.25** and **A3.26.**

A3.2 UV thermal denaturation

Please refer to section A1.2 for the experimental details.

A3.3 Circular dichroism (CD) spectroscopy

Please refer to section A1.3 for the experimental details.

A3.4 AGT repair assay

Please refer to section A1.4 for the experimental details.

Figure A3.1 - 500 MHz ^{1}H NMR spectrum of compound (**3a**) (in d_6-acetone)

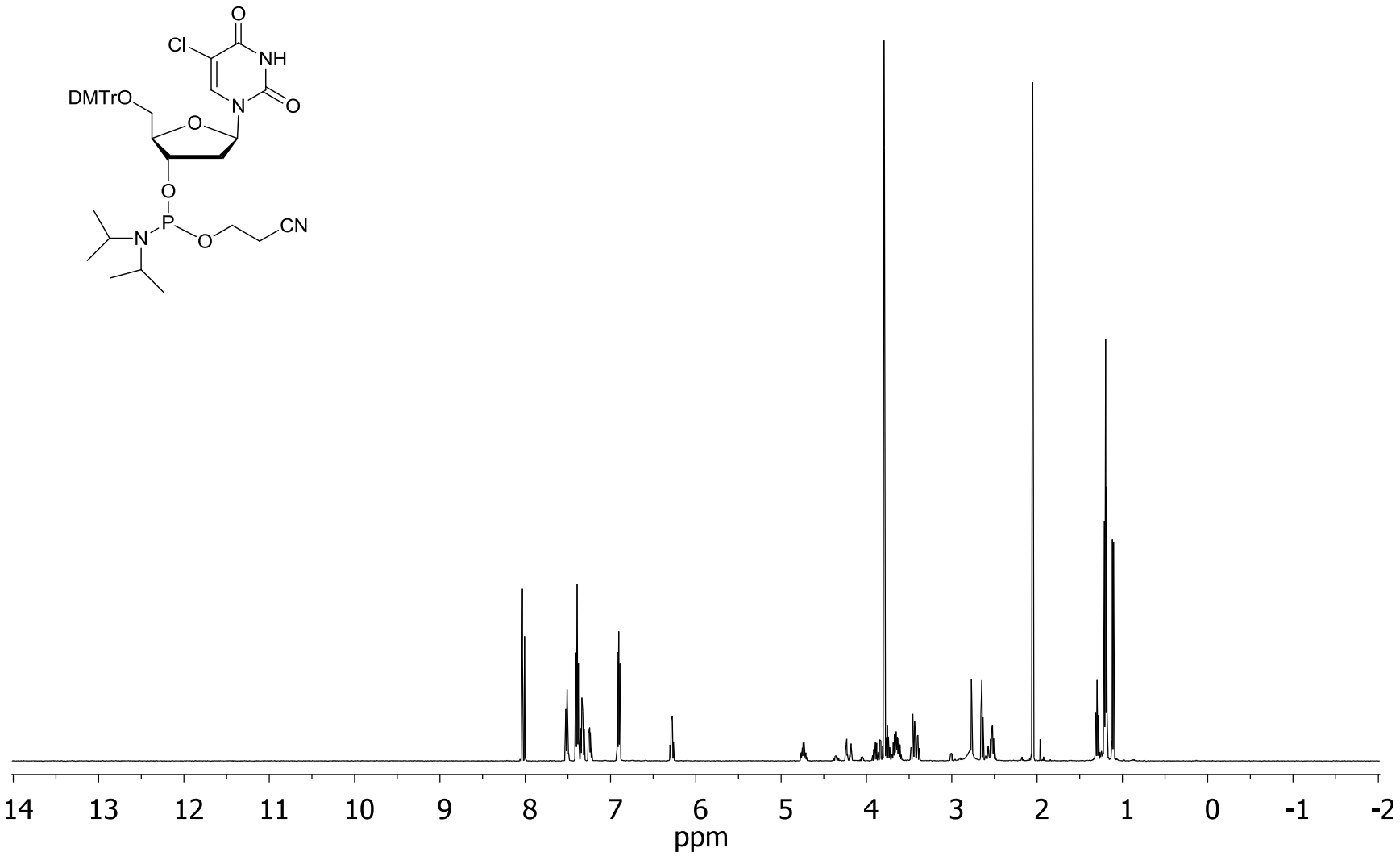


Figure A3.2 -125.7 MHz ^{13}C NMR spectrum of compound (**3a**) (in d_6-acetone)

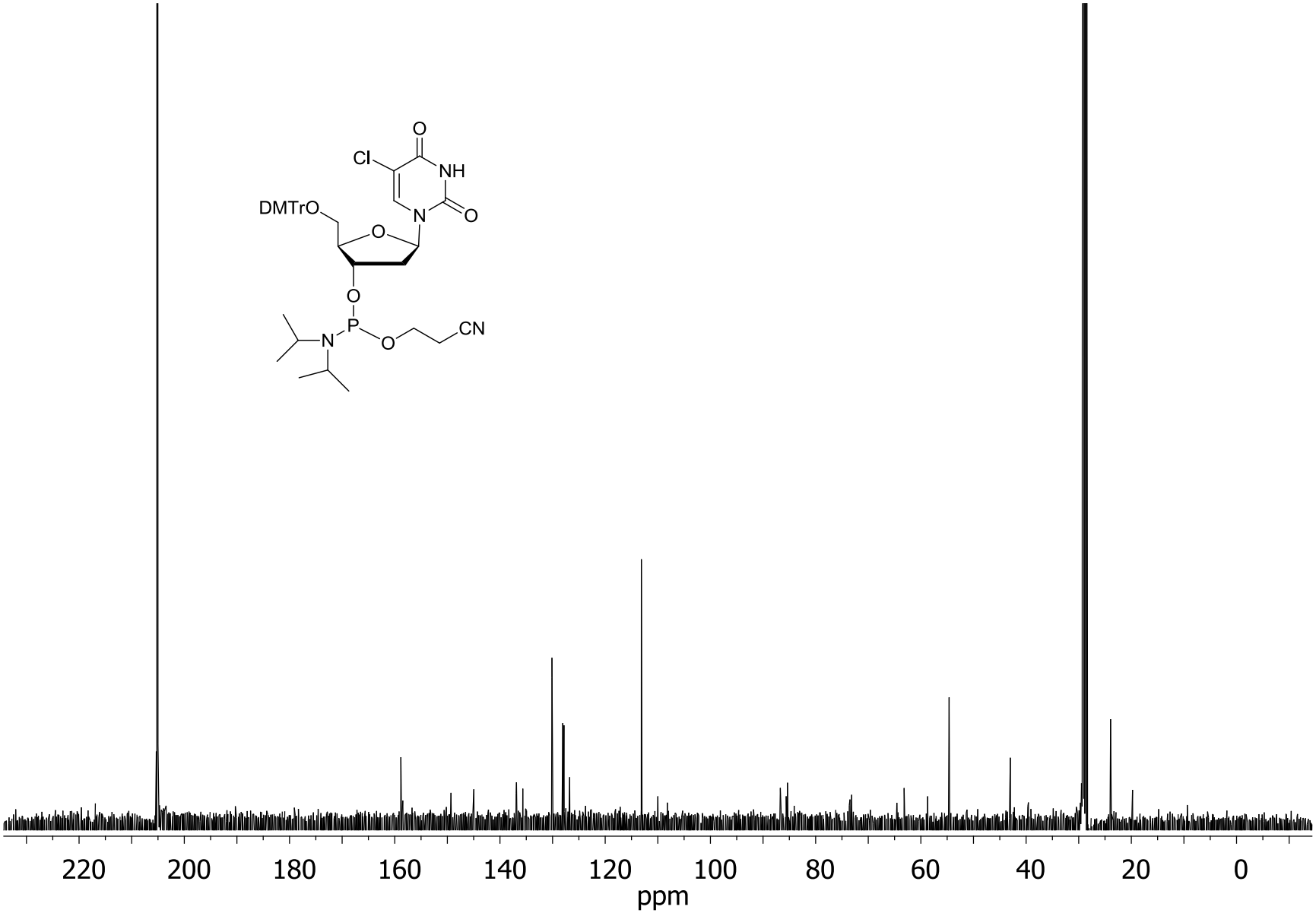


Figure A3.3 - 202.3 MHz ^{31}P NMR spectrum of compound (**3a**) (in d_6-acetone)

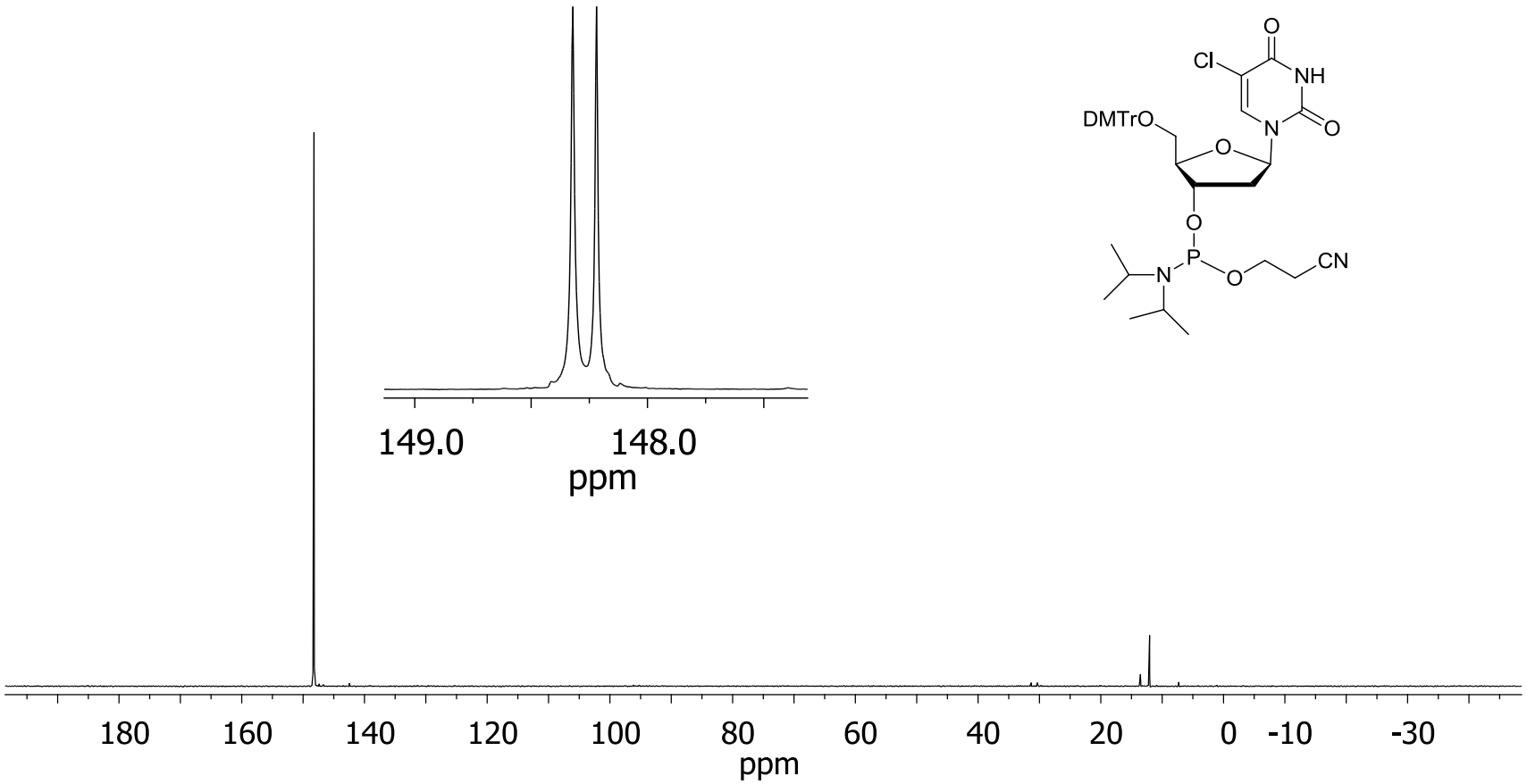


Figure A3.4 - 500 MHz ^{1}H NMR spectrum of compound (**4a**) (in $CDCl_3$)

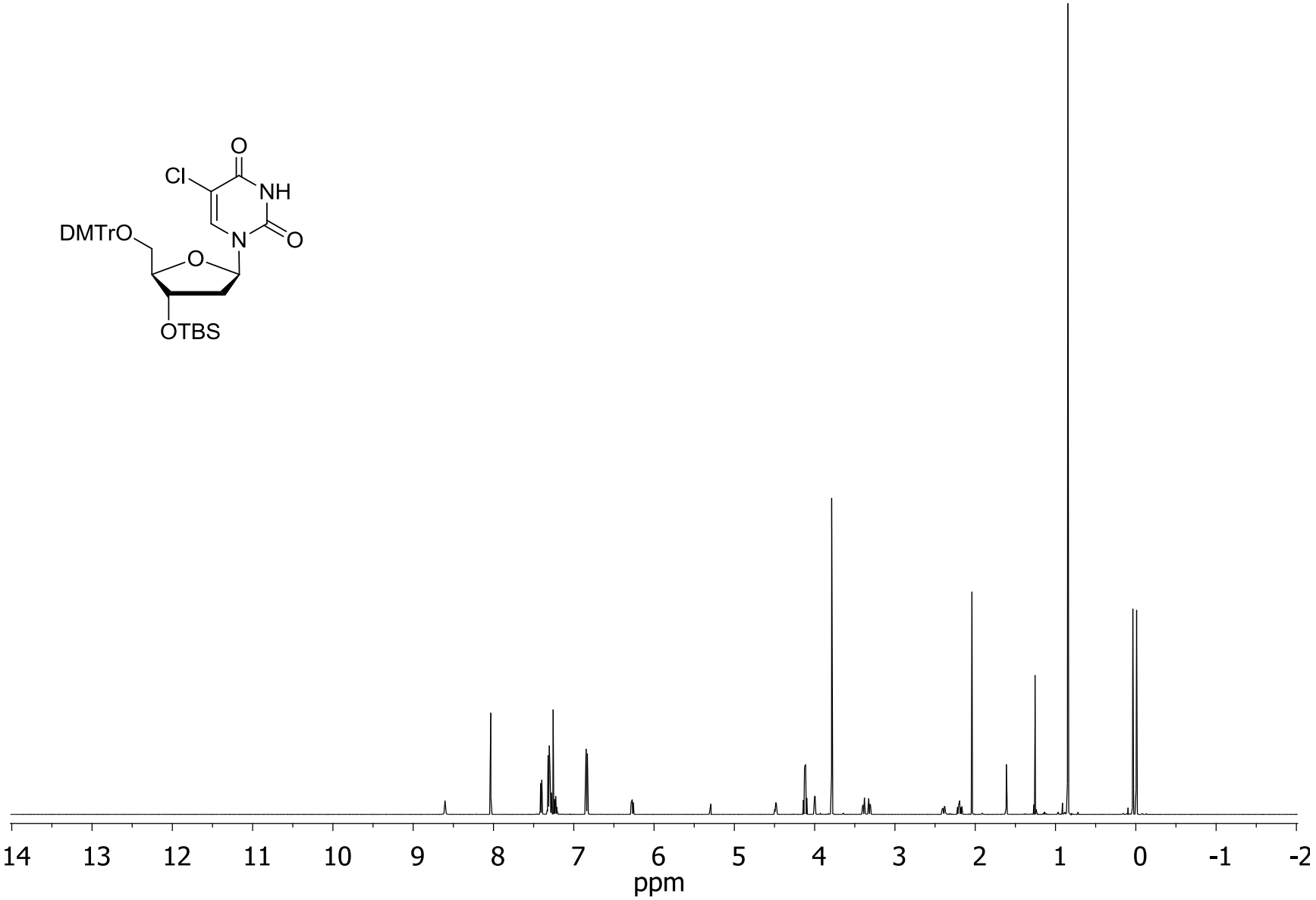


Figure A3.5 - 125.7 MHz ^{13}C NMR spectrum of compound (**4a**) (in $CDCl_3$)

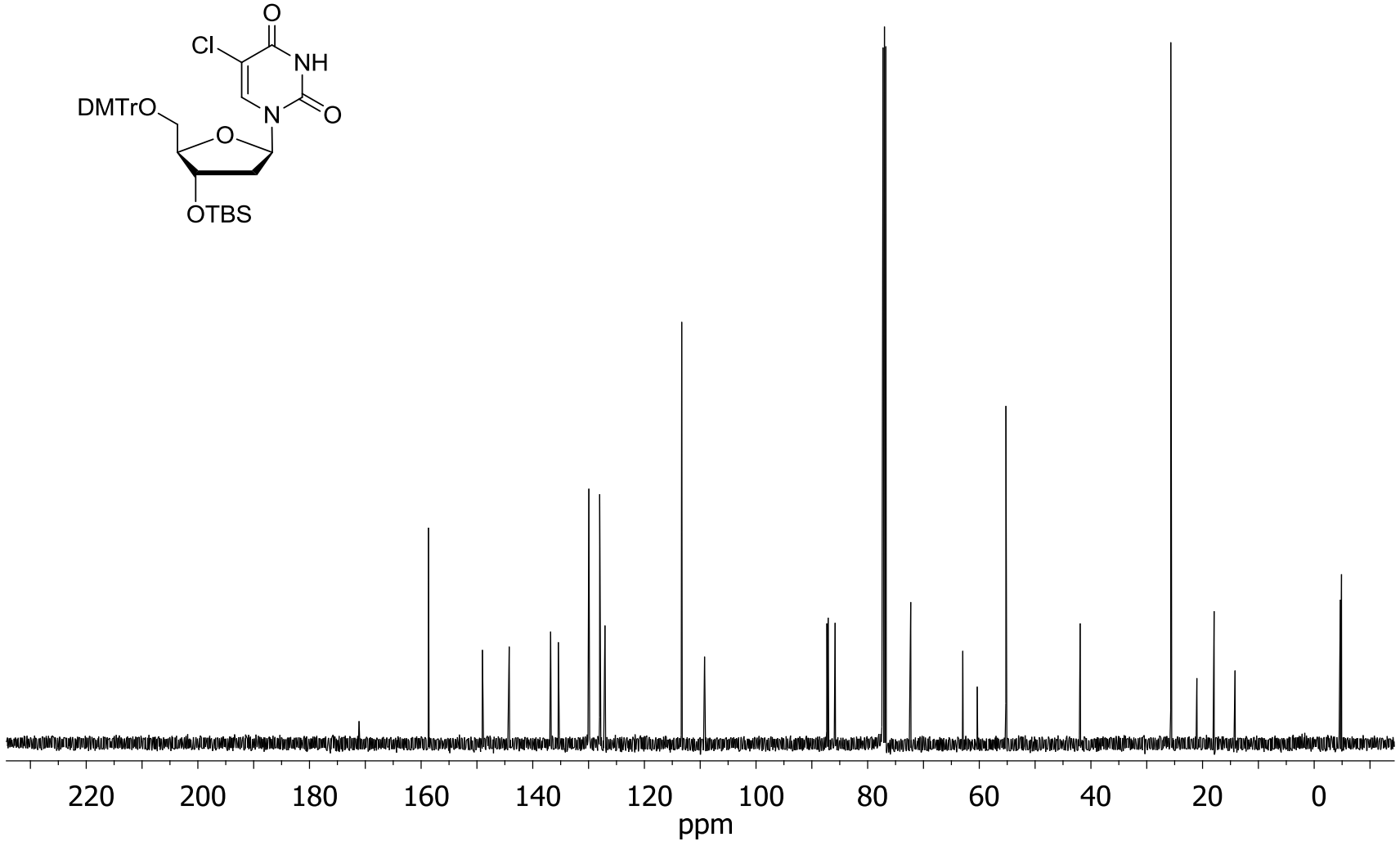


Figure A3.6 - 500 MHz ^{1}H NMR spectrum of compound (**5a**) (in $CDCl_3$)

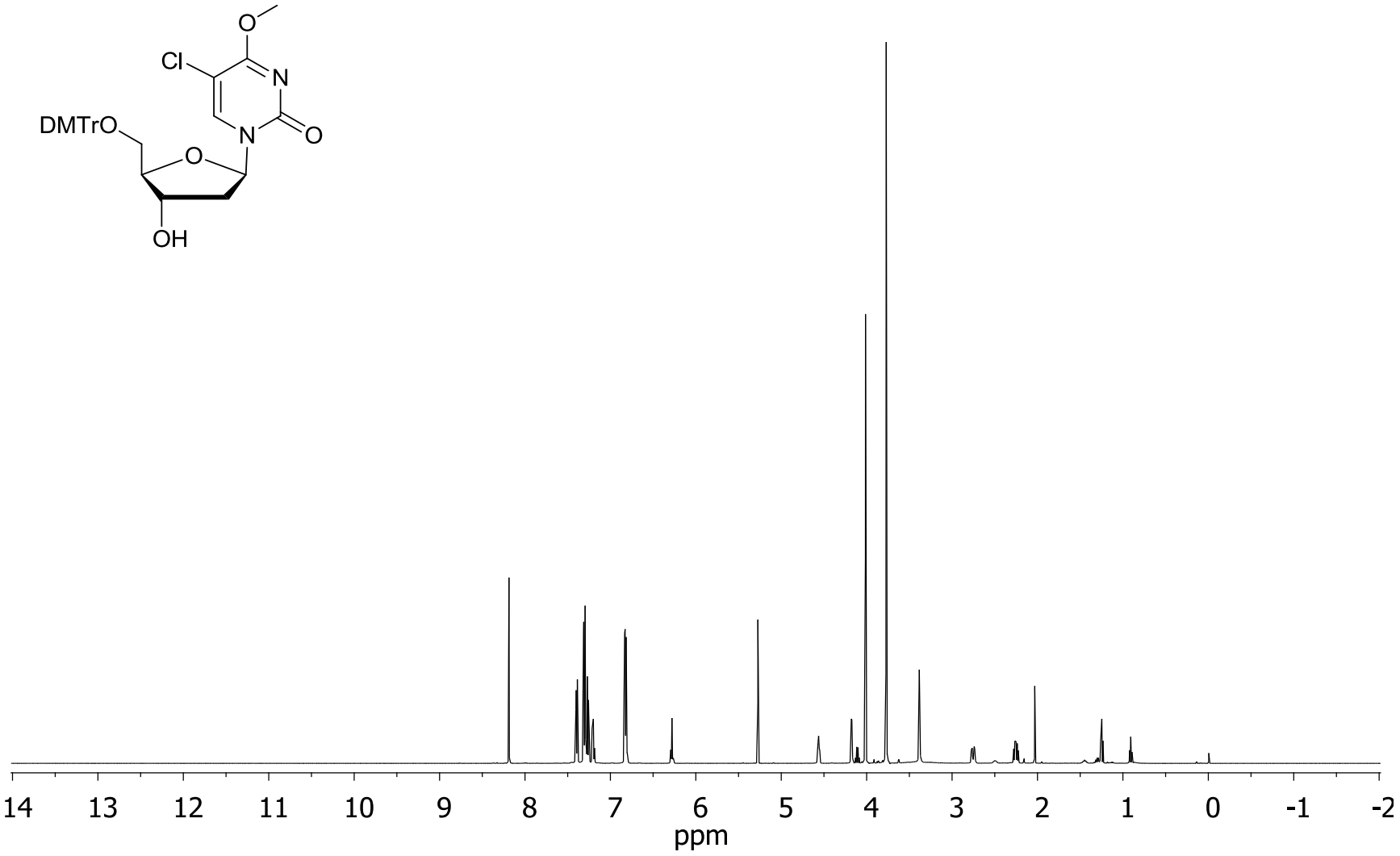


Figure A3.7 - 125.7 MHz ^{13}C NMR spectrum of compound (**5a**) (in $CDCl_3$)

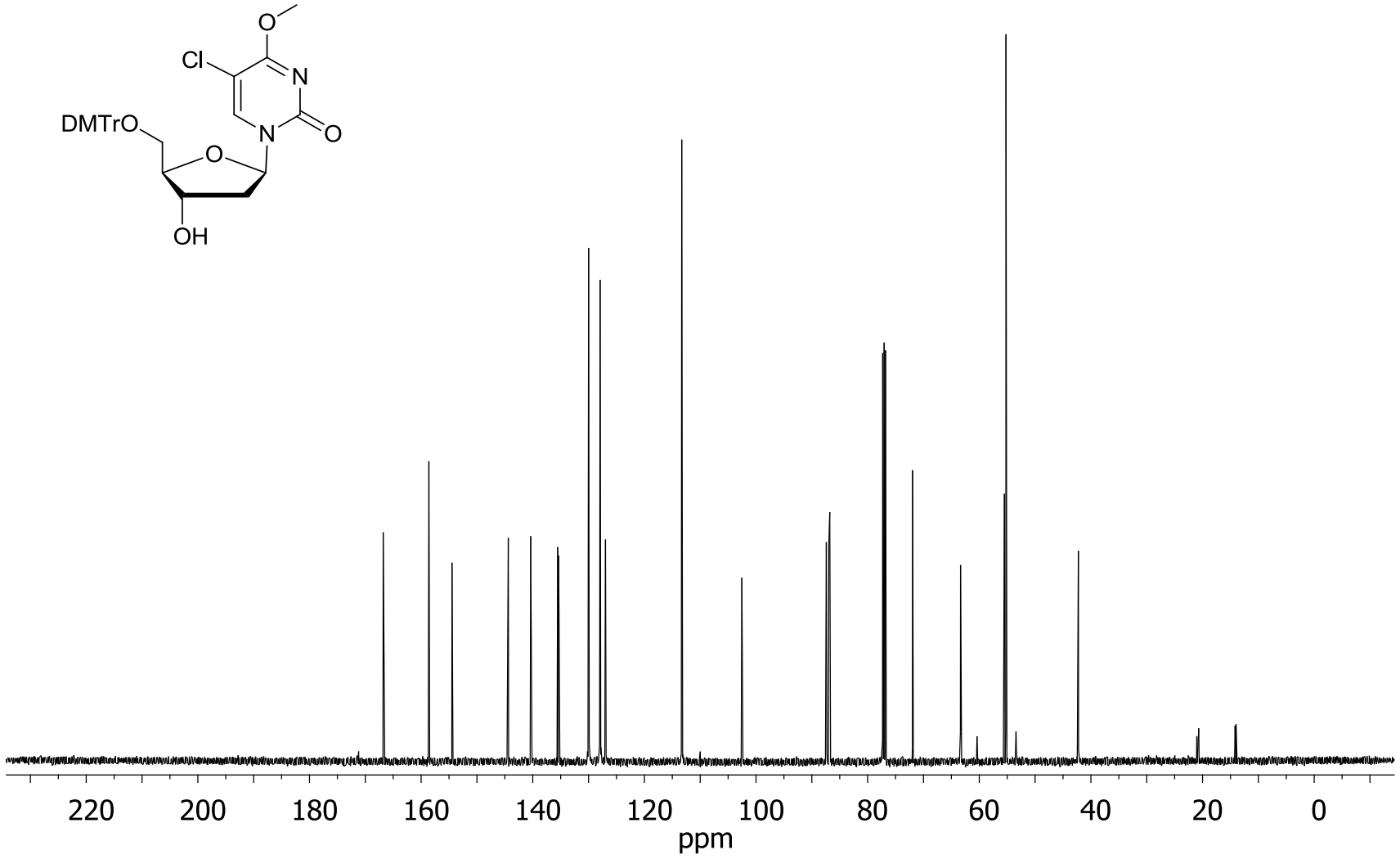


Figure A3.8 - 500 MHz ^{1}H NMR spectrum of compound (**6a**) (in d_6-acetone)

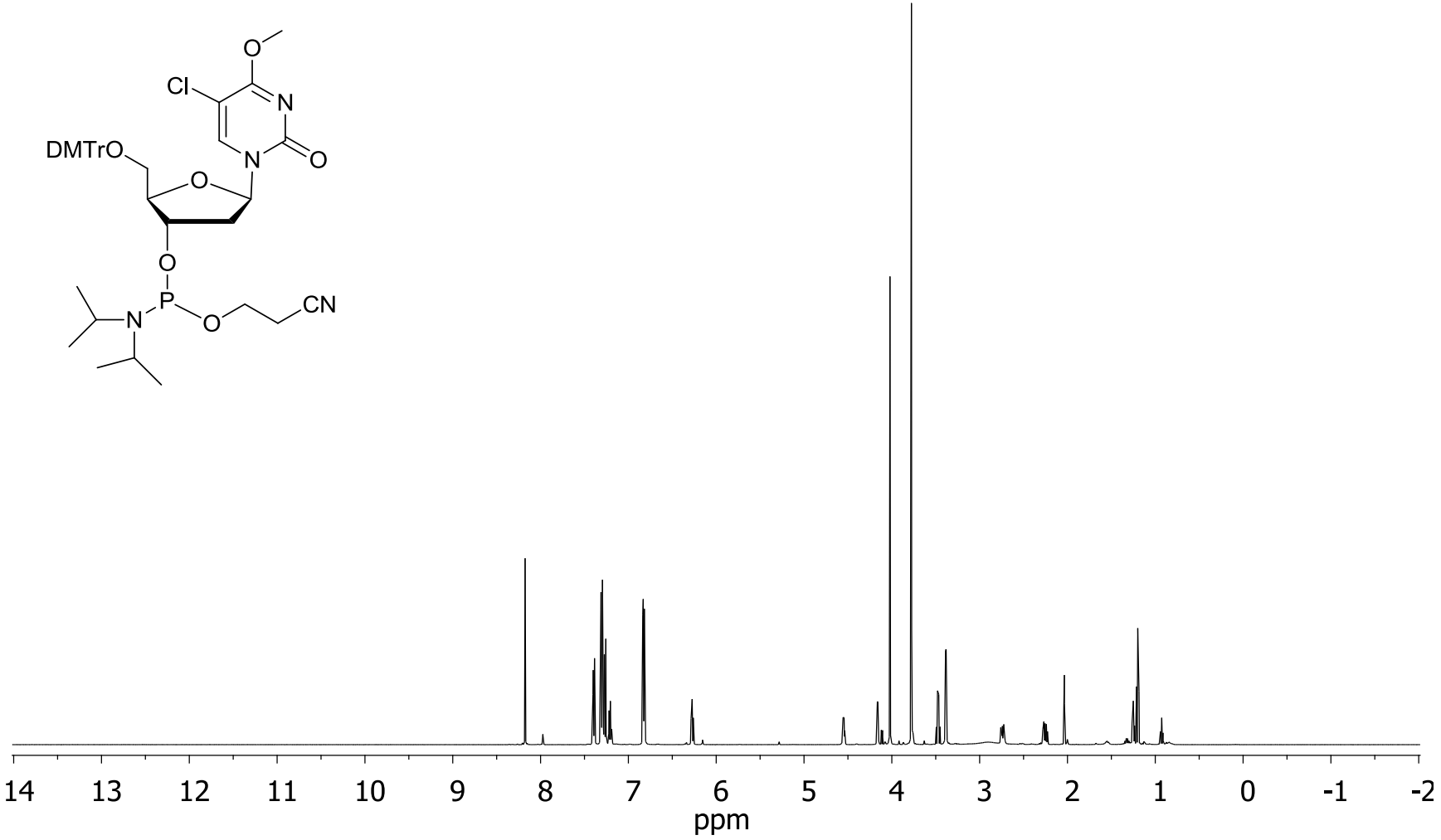


Figure A3.9 - 125.7 MHz ^{13}C NMR spectrum of compound (**6a**) (in d_6-acetone)

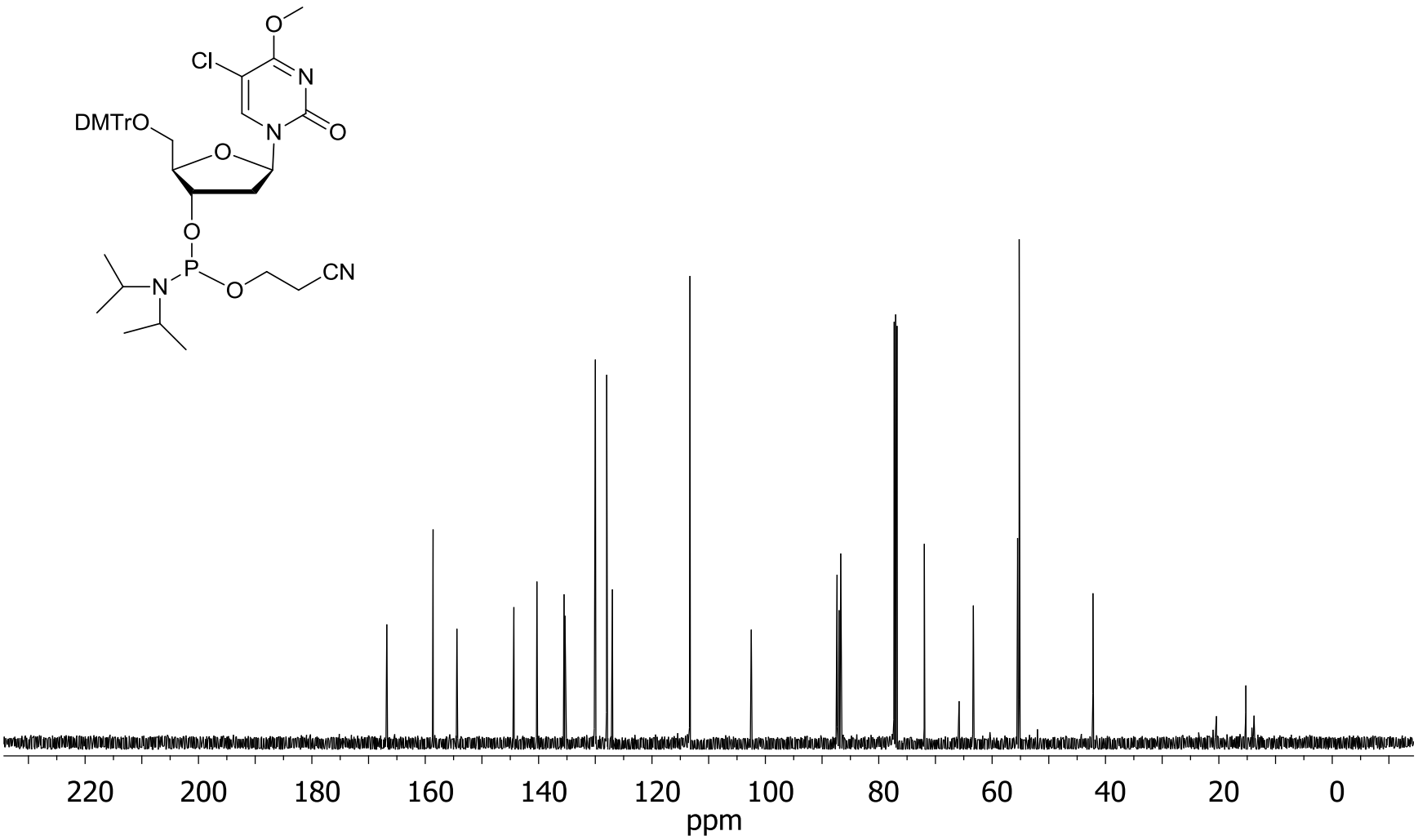


Figure A3.10 - 202.3 MHz ^{31}P NMR spectrum of compound (**6a**) (in d_6-acetone)

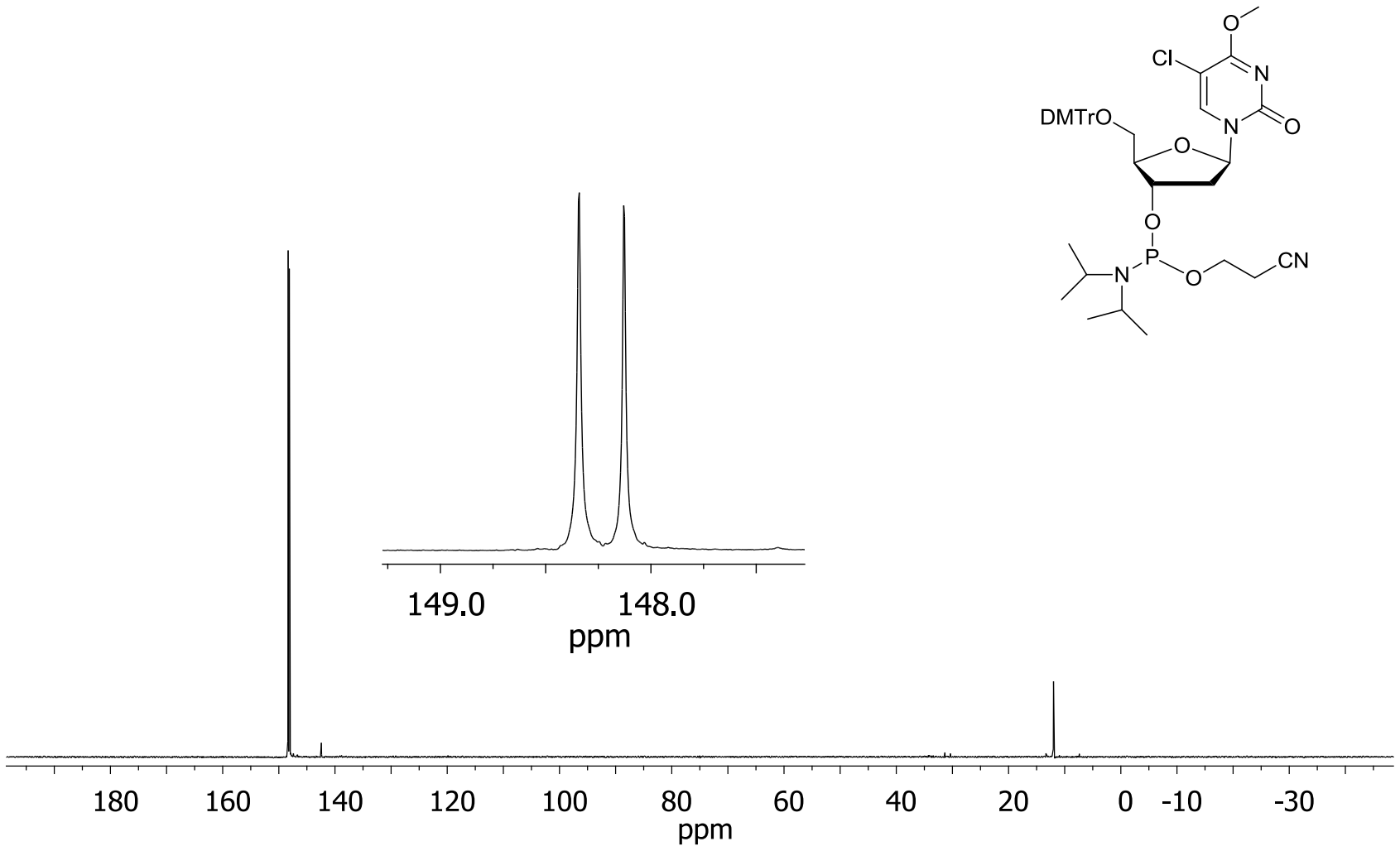


Figure A3.11: 500 MHz ^{1}H NMR spectrum of compound (**3b**) (in d_6-acetone)

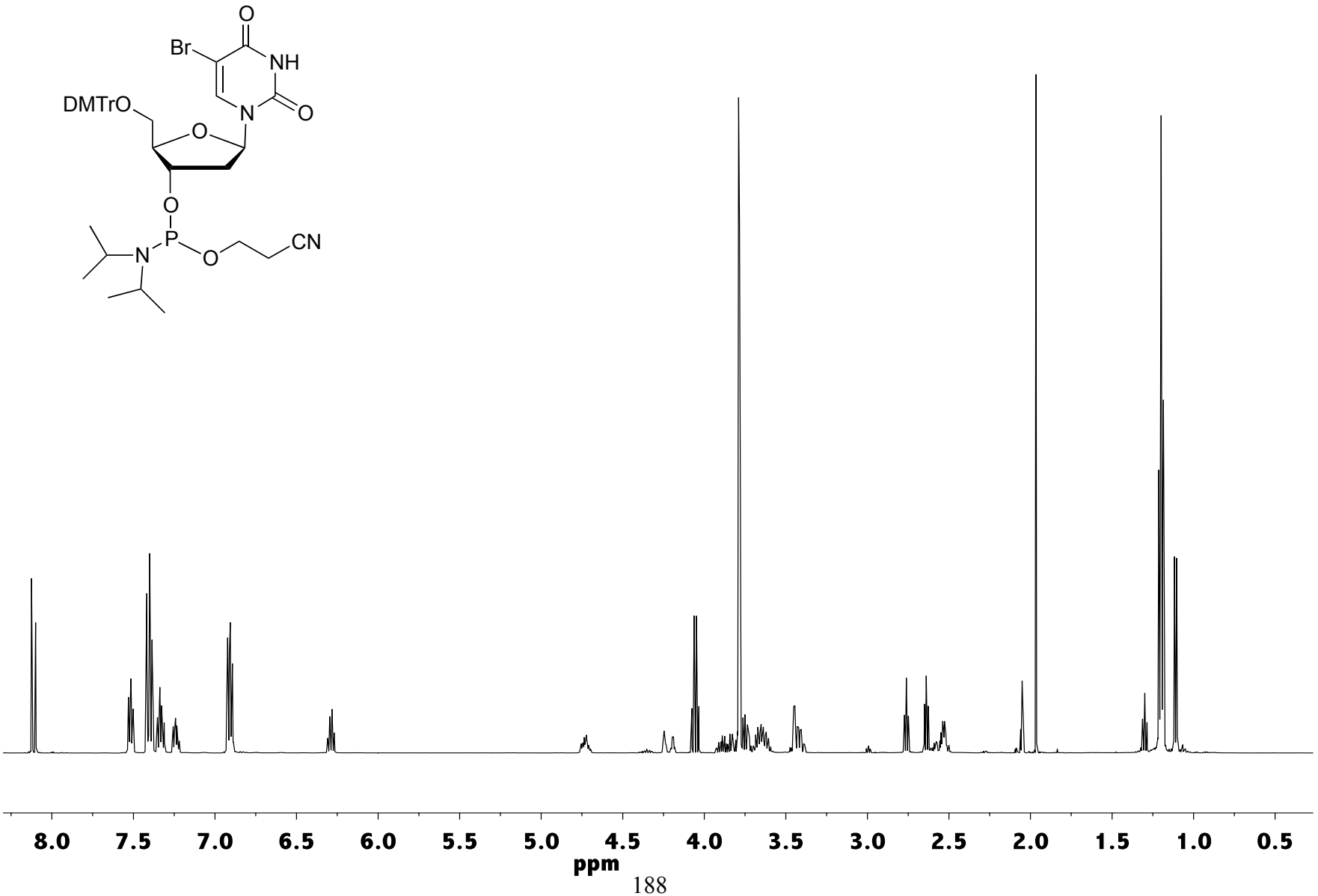


Figure A3.12: 125.7 MHz ^{13}C NMR spectrum of compound (**3b**) (in d_6-acetone)

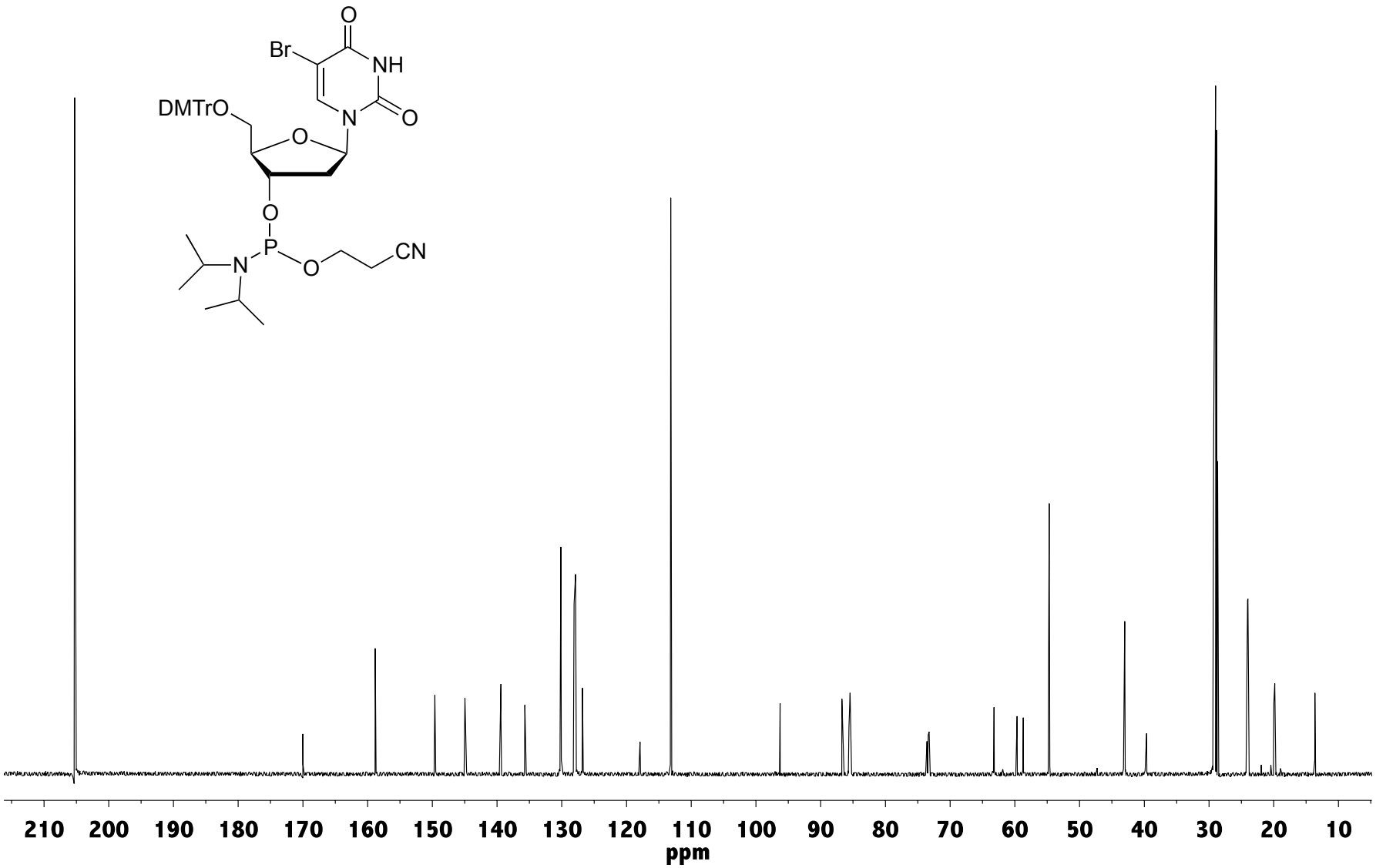


Figure A3.13- 202.3 MHz ^{31}P NMR spectrum of compound (**3b**) (in d_6-acetone)

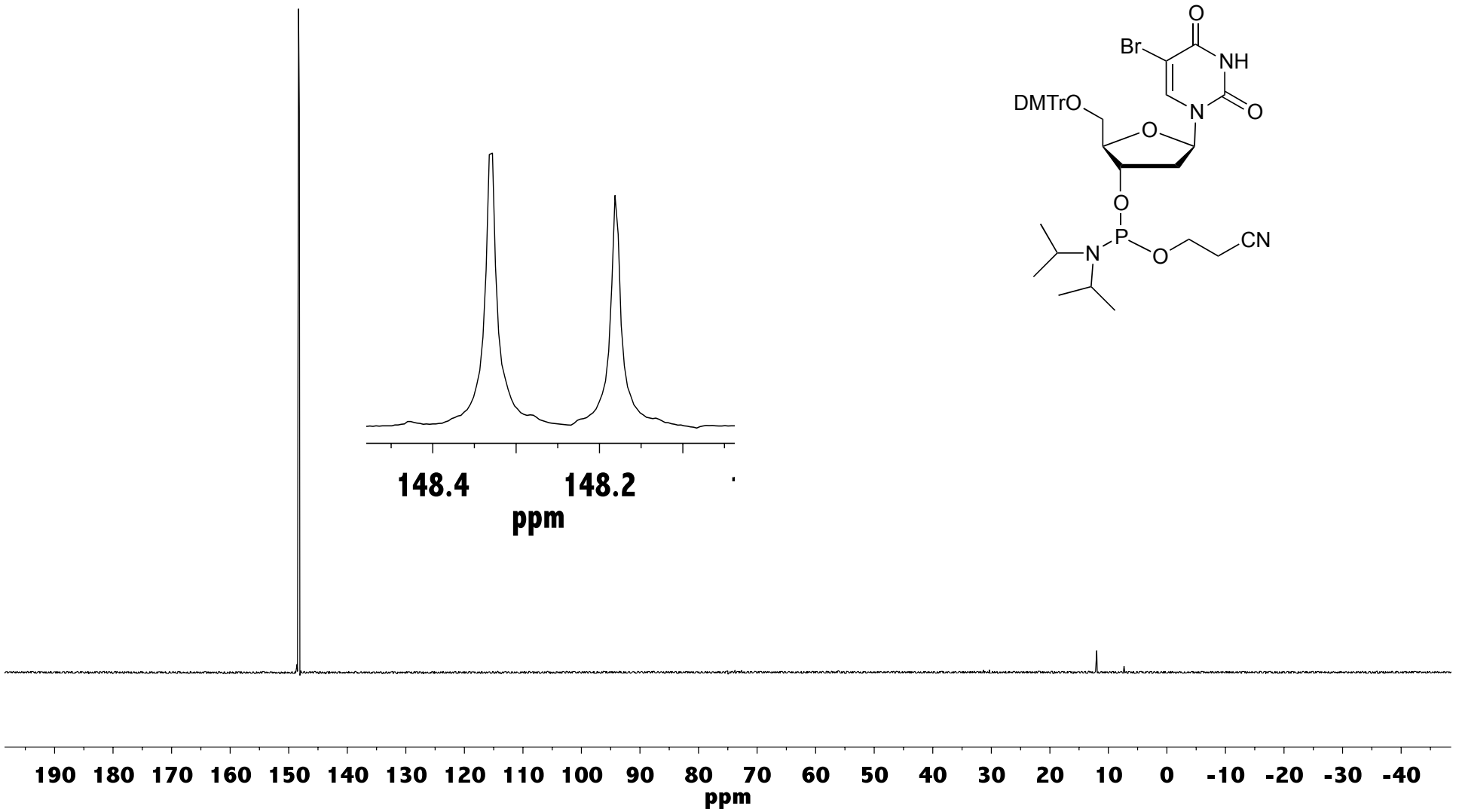


Figure A3.14: 500 MHz ^{1}H NMR spectrum of compound (**4b**) (in $CDCl_3$)

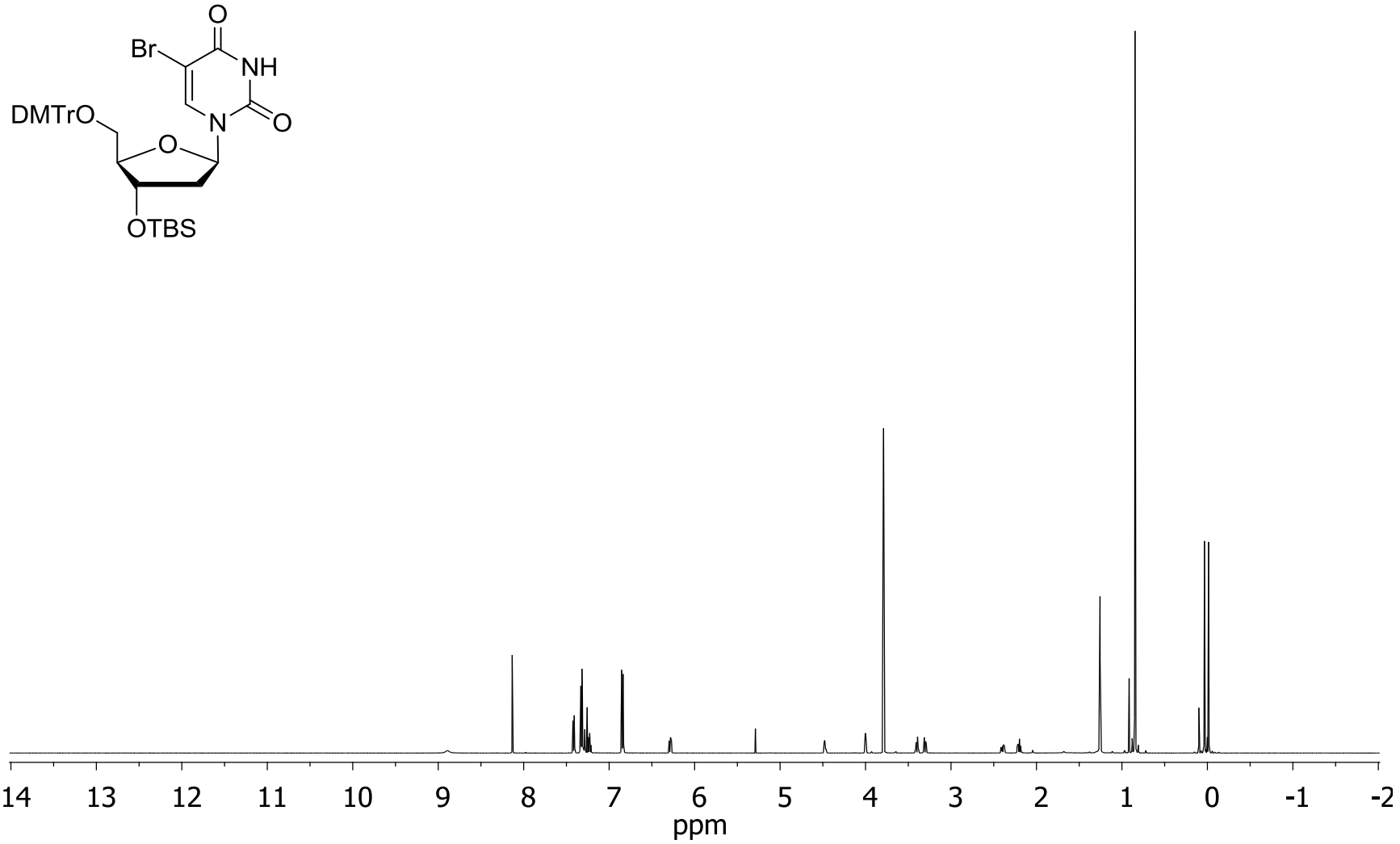


Figure A3.15: 125.7 MHz ^{13}C NMR spectrum of compound (**4b**) (in $CDCl_3$)

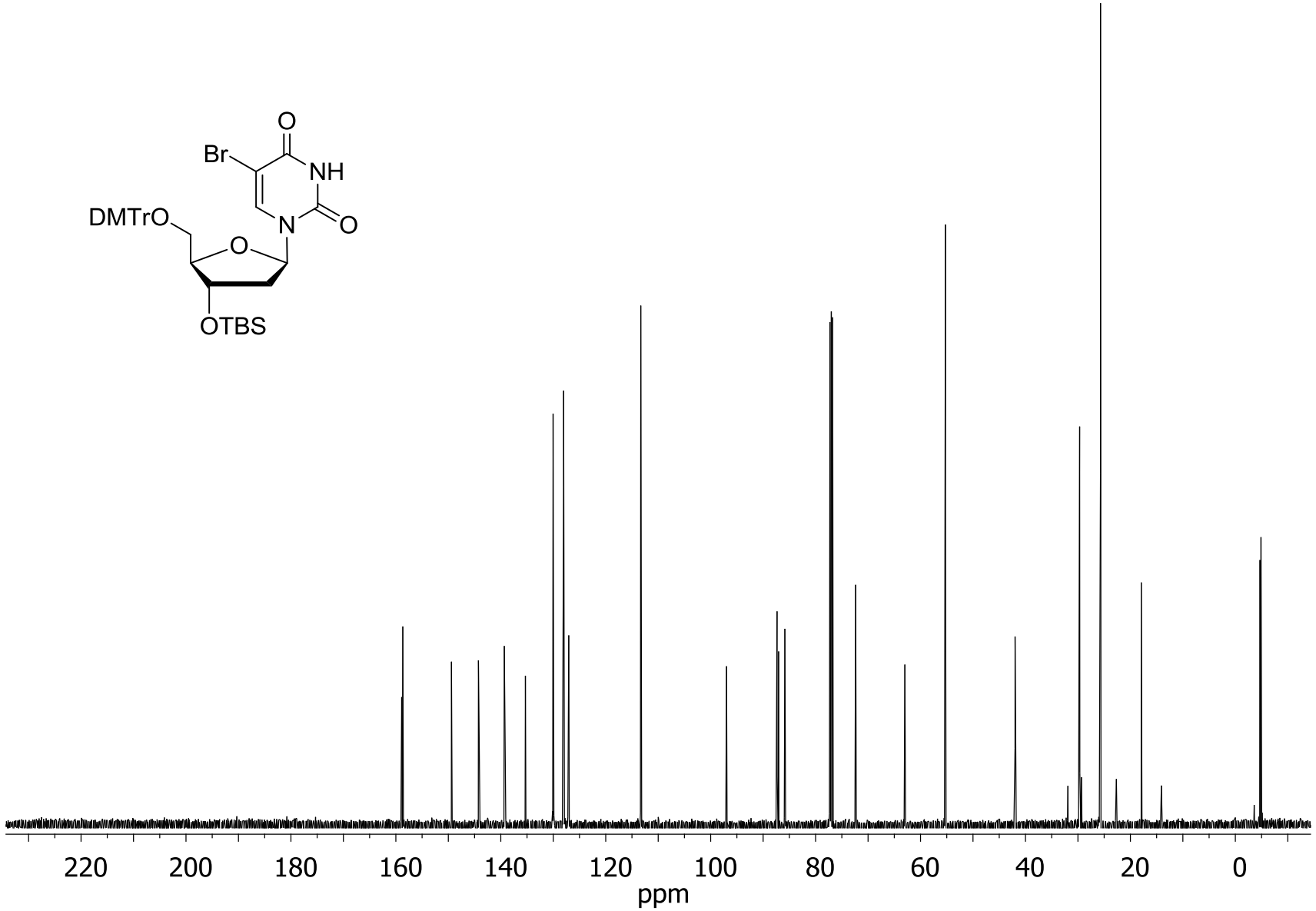


Figure A3.16: 500 MHz ^{1}H NMR spectrum of compound (**5b**) (in $CDCl_3$)

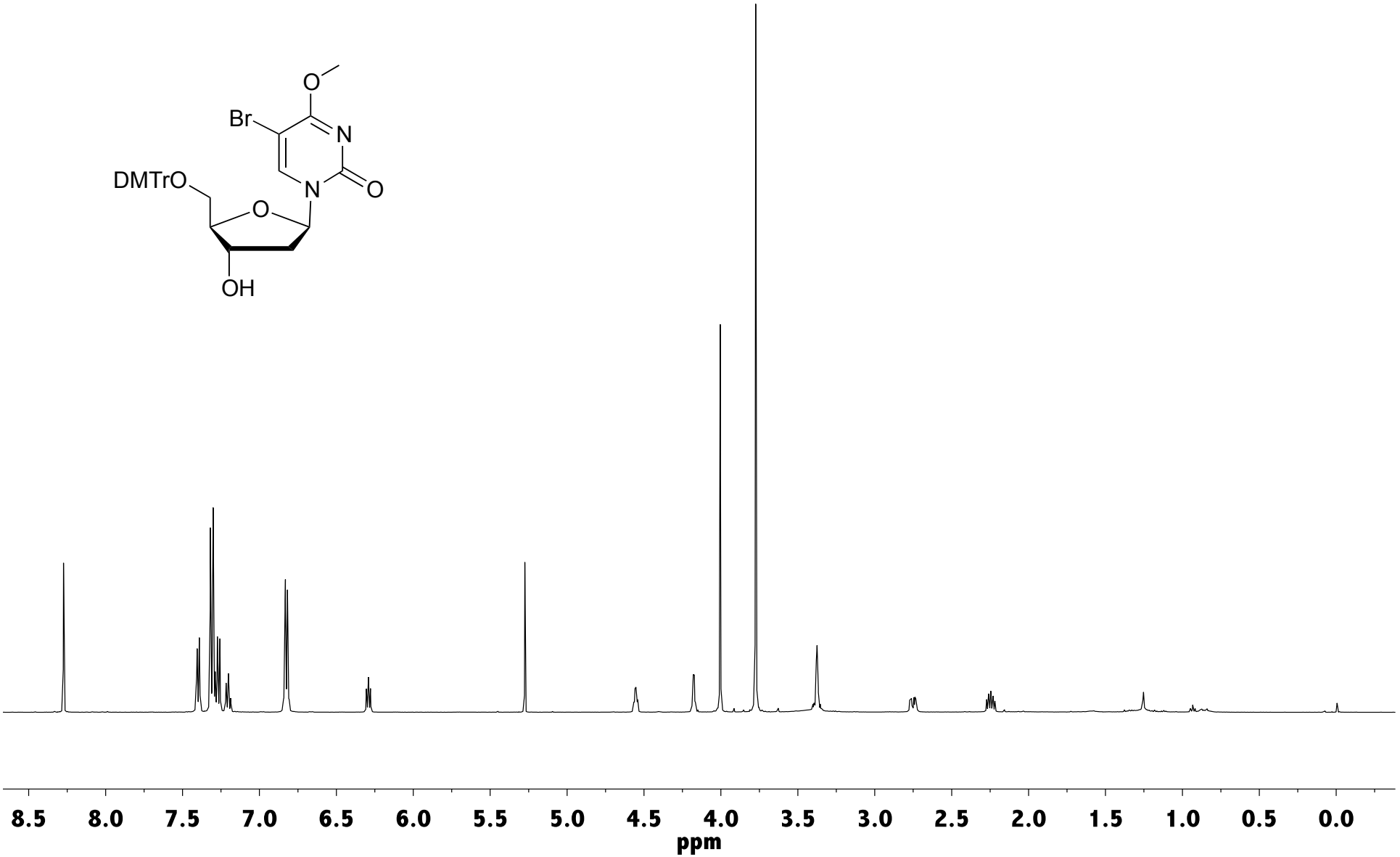


Figure A3.17: 125.7 MHz ^{13}C NMR spectrum of compound (**5b**) (in $CDCl_3$)

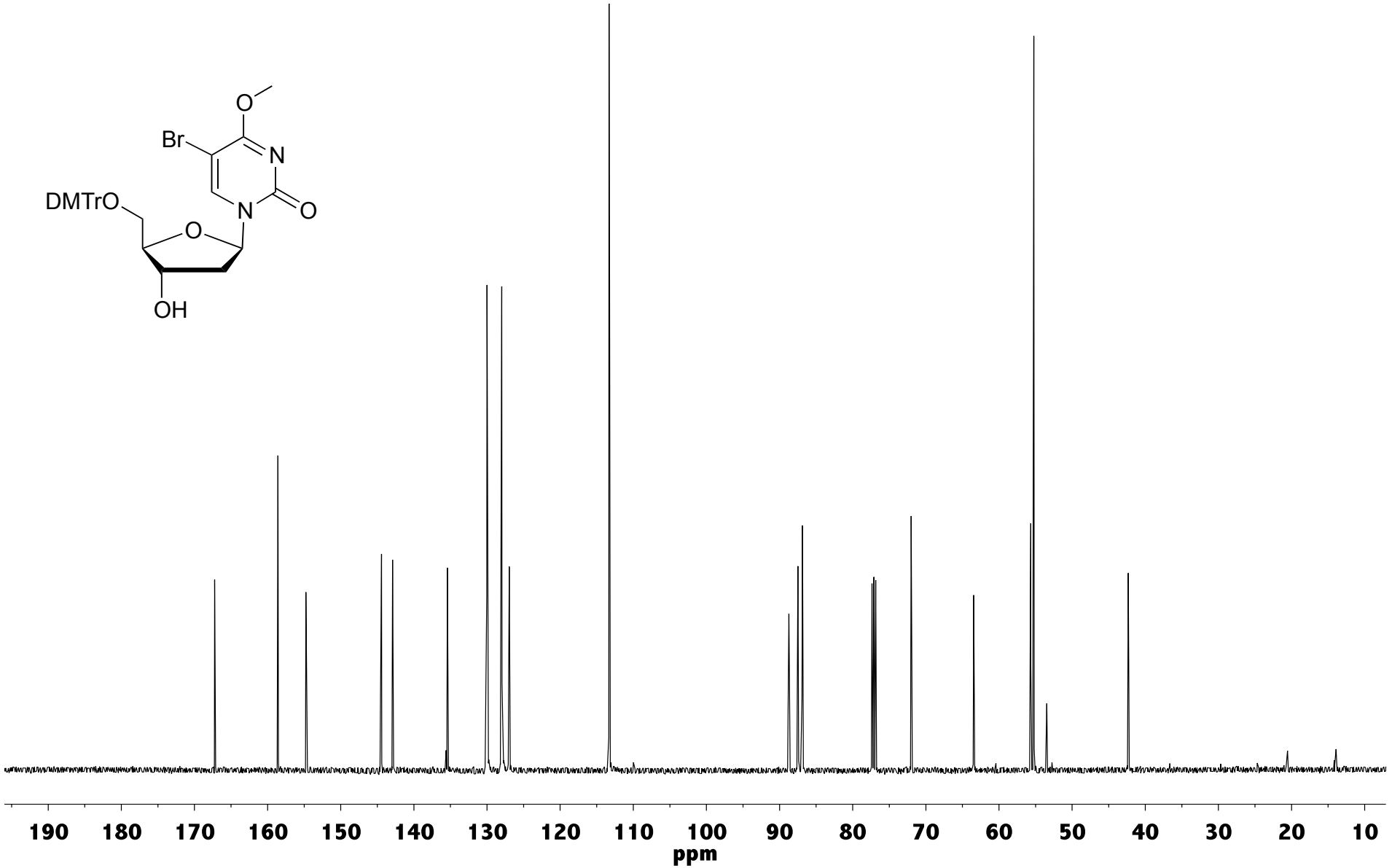


Figure A3.18 - 500 MHz ^{1}H NMR spectrum of compound (**6b**) (in d_6-acetone)

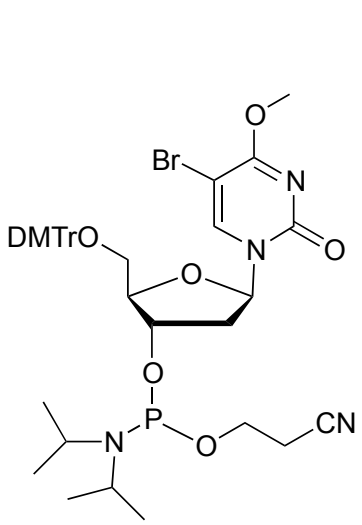


Figure A3.19- 125.7 MHz ^{13}C NMR spectrum of compound (**6b**) (in d_6-acetone)

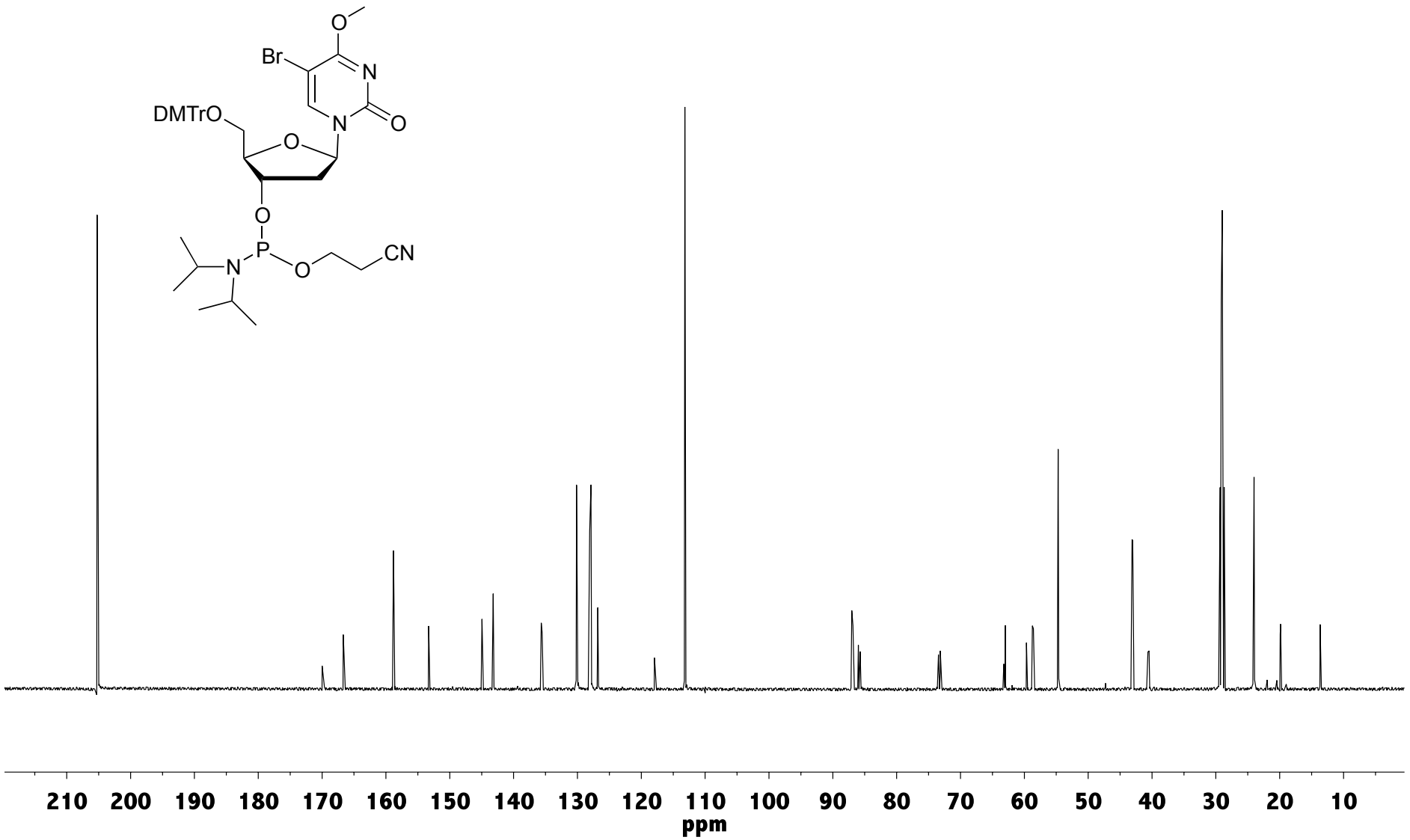


Figure A3.20 - 202.3 MHz ^{31}P NMR spectrum of compound (**6b**) (in d_6-acetone)

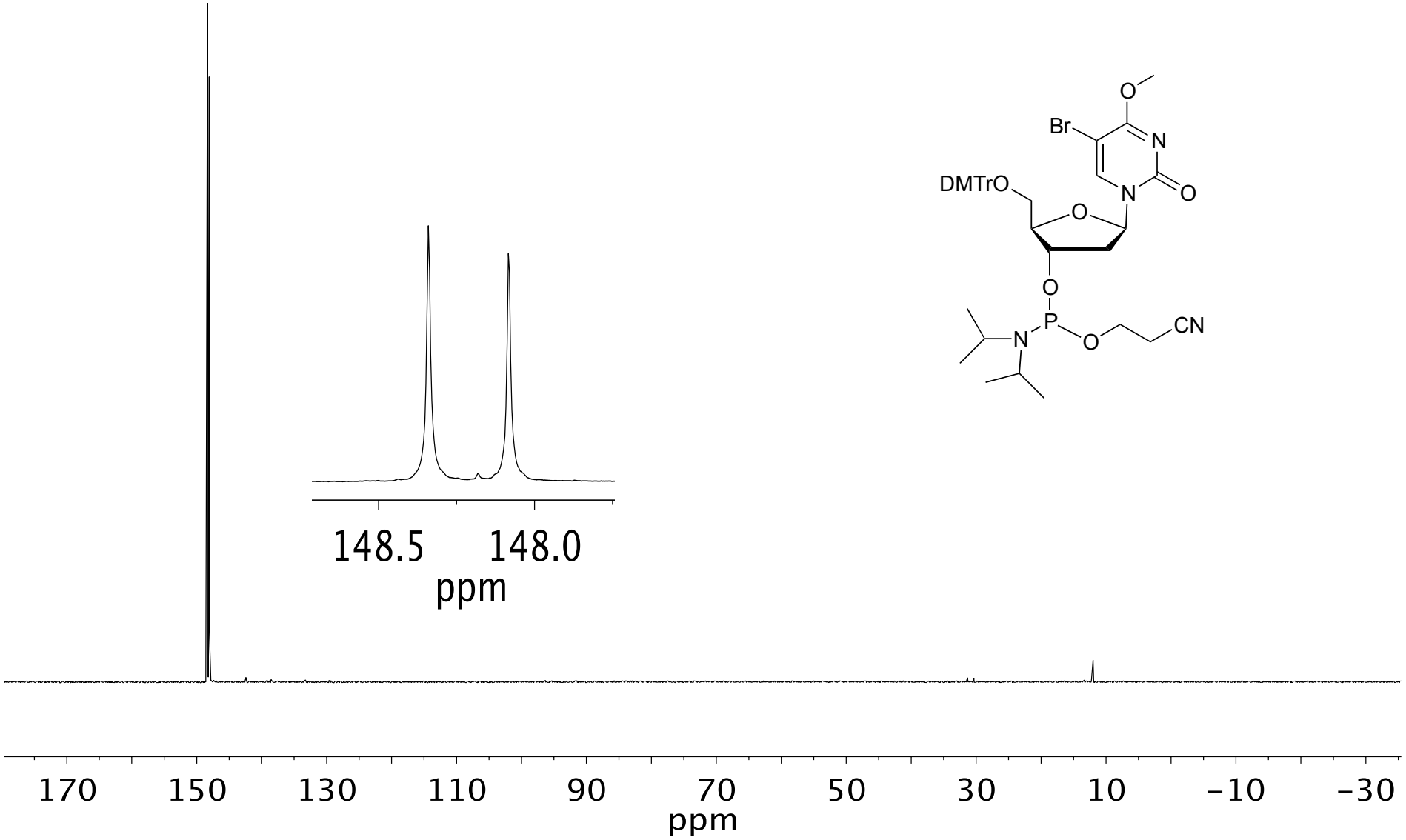


Figure A3.21 - ESI MS spectrum of oligonucleotide **dClU** (expected mass of 4284.9)

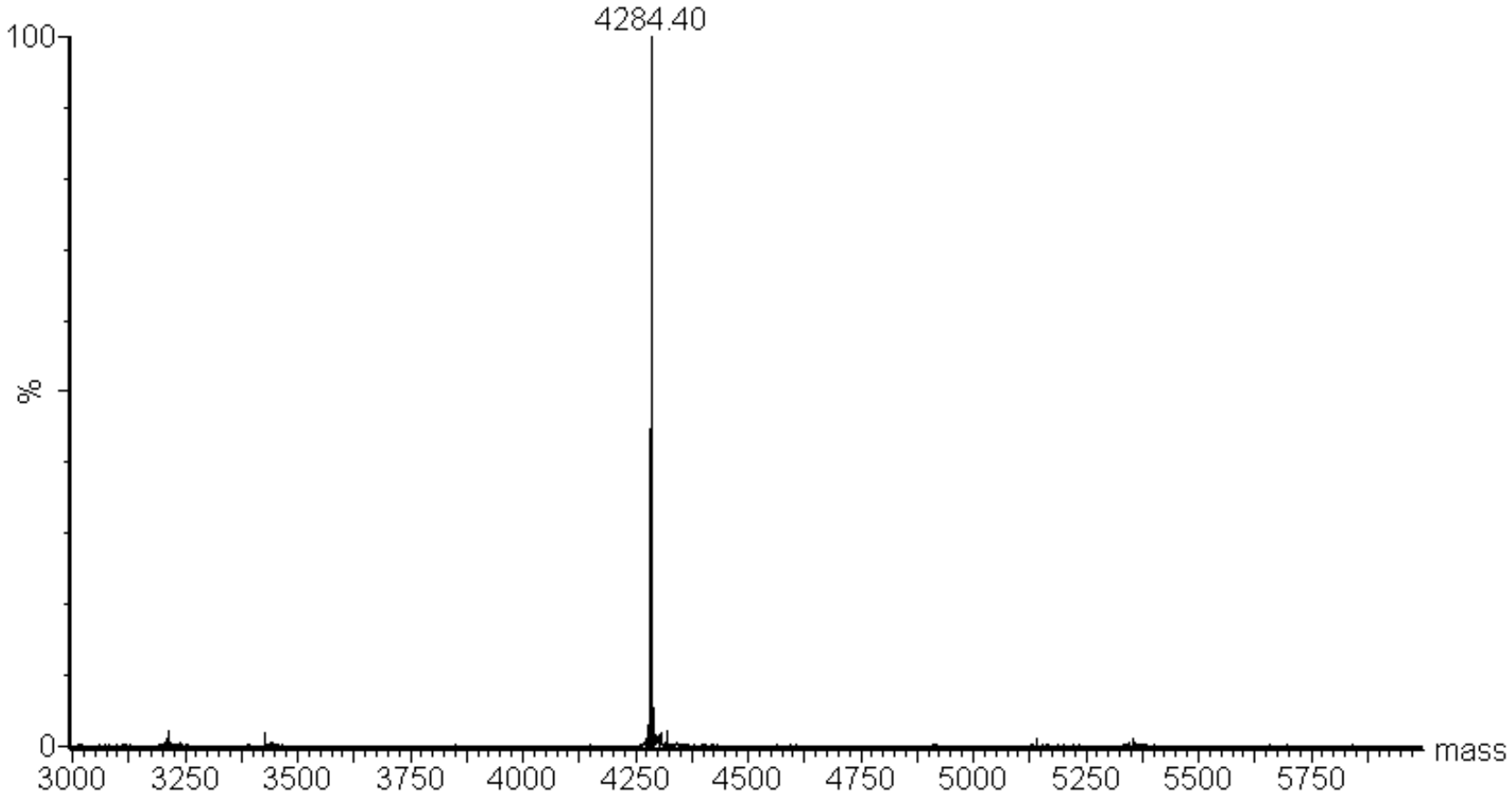


Figure A3.22 - ESI MS spectrum of oligonucleotide **dClU-Me** (expected mass of 4298.9)

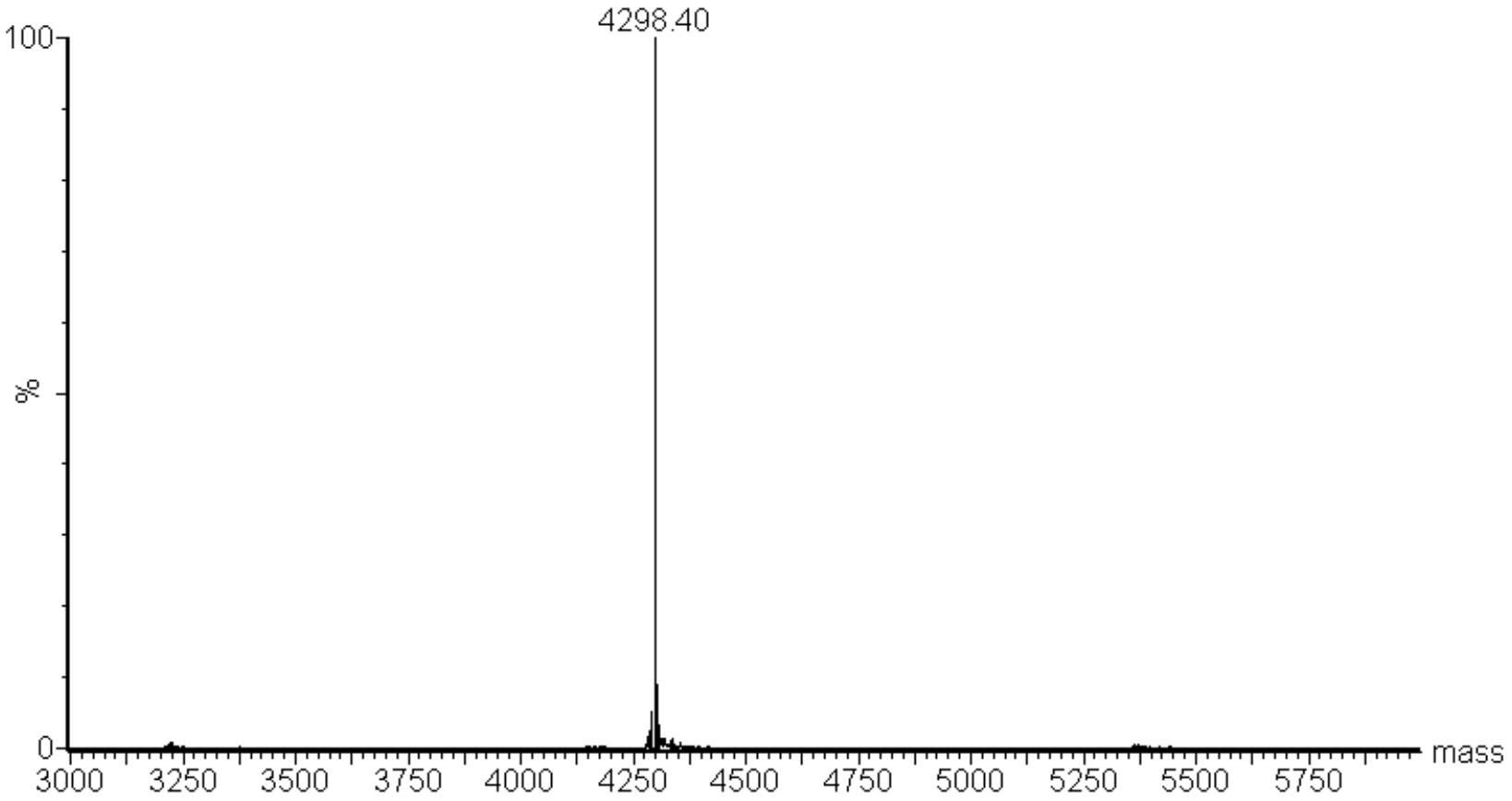


Figure A3.23 - ESI MS spectrum of oligonucleotide **dBrU** (expected mass of 4328)

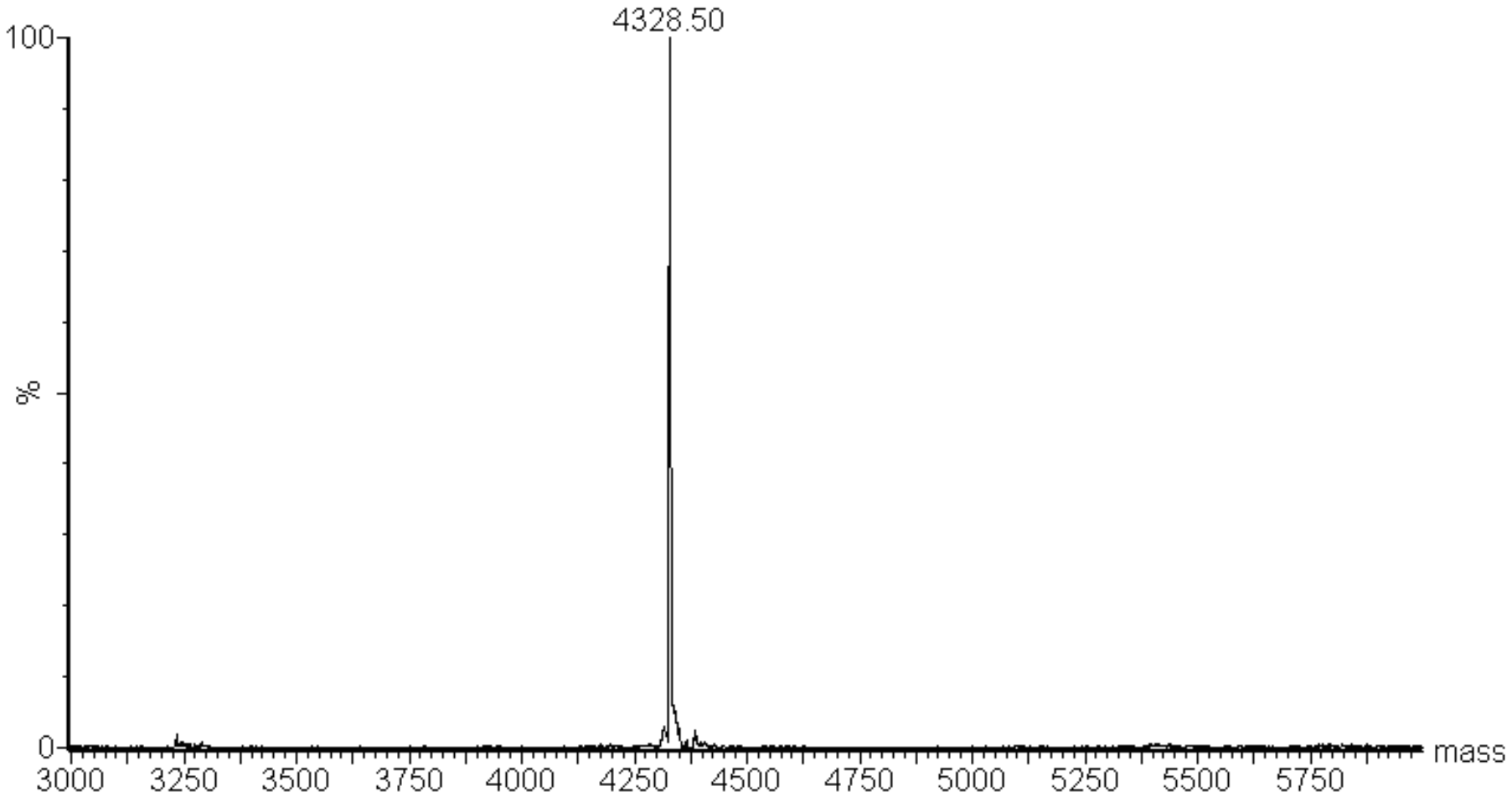


Figure A3.24 - ESI MS spectrum of oligonucleotide **dBrU-Me** (expected mass of 4342.5)

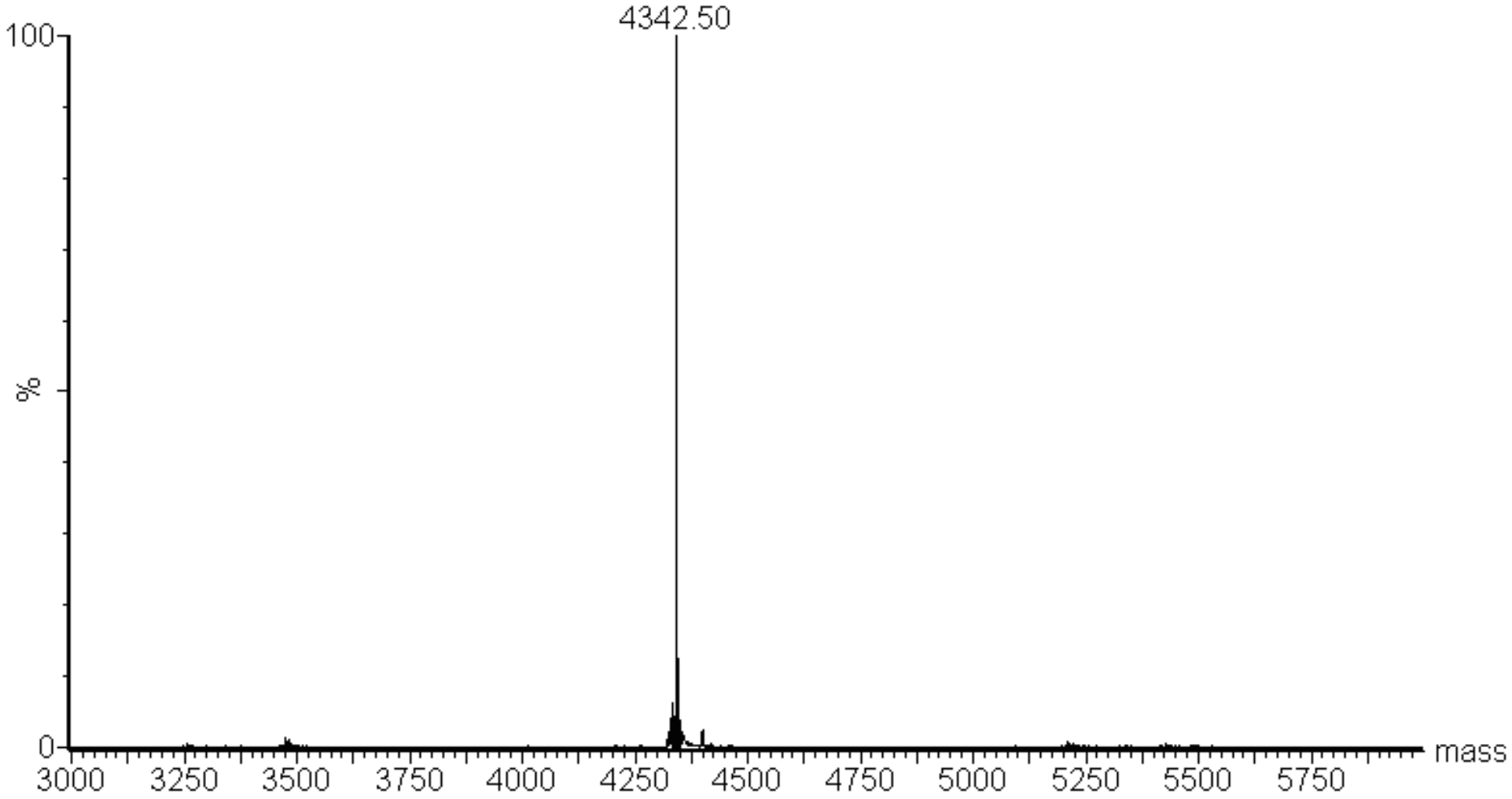
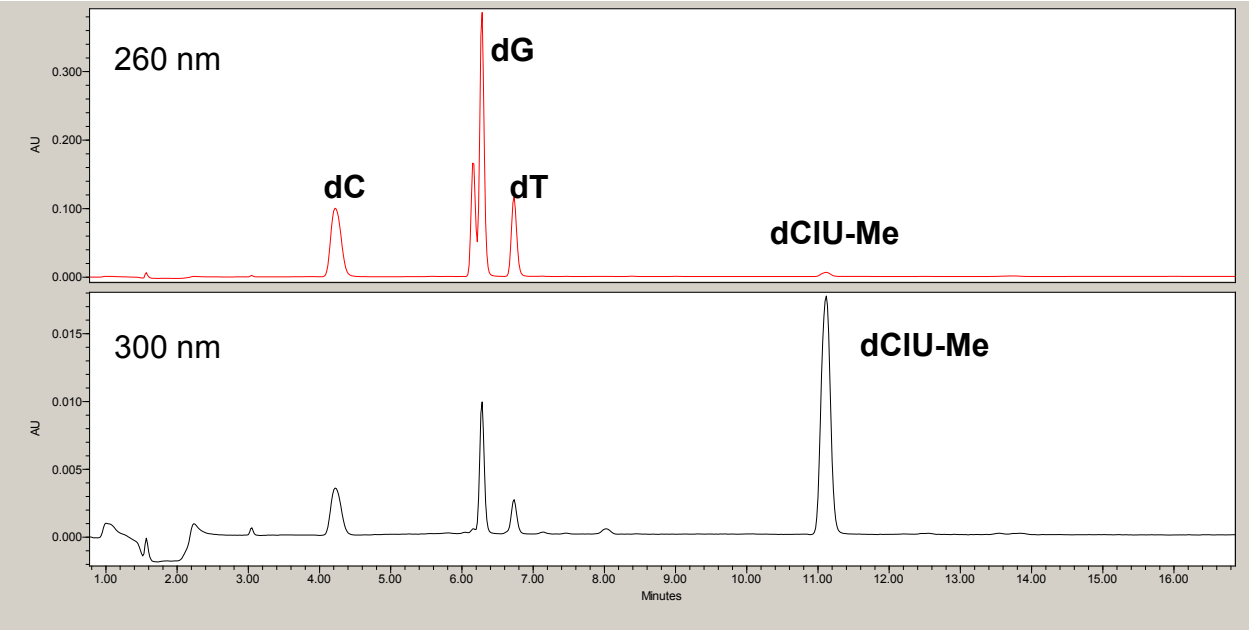
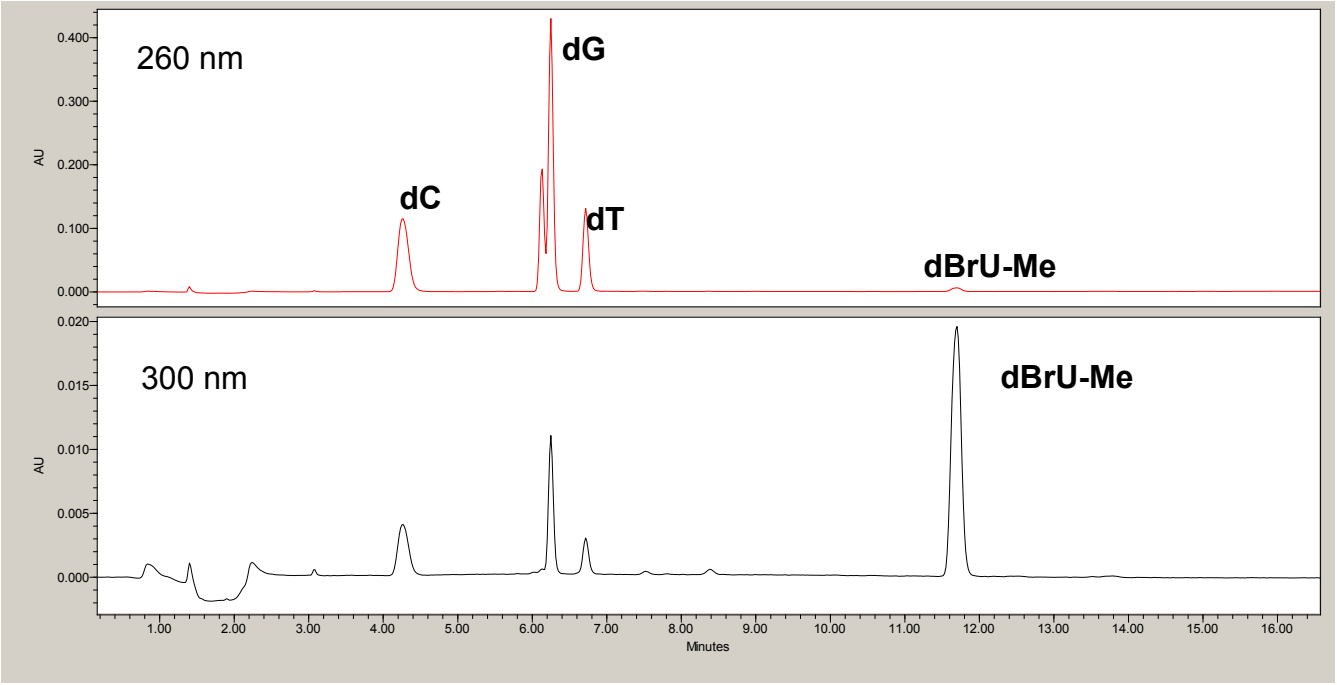


Figure A3.25- C-18 reversed phase HPLC profile of nuclease-digested **dClU-Me**. The column was eluted with a linear gradient of 0-70% buffer B over 30 min (buffer A: 50 mM sodium phosphate, pH 5.8, 2% MeCN and buffer B: 50 mM sodium phosphate, pH 5.8, 50% MeCN).



Oligomer	Nucleoside composition	Nucleoside Ratio	
		Expected	Observed
dClU-Me	dC	4	4.2
	dG	4	4.0
	dA (dI)	3	2.8
	dT	2	2.1
	dClU-Me	1	0.6

Figure A3.26- C-18 reversed phase HPLC profile of nuclease-digested **dBrU-Me**. The column was eluted with a linear gradient of 0-70% buffer B over 30 min (buffer A: 50 mM sodium phosphate, pH 5.8, 2% MeCN and buffer B: 50 mM sodium phosphate, pH 5.8, 50% MeCN).



Oligomer	Nucleoside composition	Nucleoside Ratio	
		Expected	Observed
dBrU-Me	dC	4	4.0
	dG	4	3.8
	dA (dI)	3	2.7
	dT	2	2.1
	dBrU-Me	1	0.6

Figure A3.27- SAX-HPLC profile of **dClU** crude **(A)** and pure **(B)**. The column (Dionex DNAPAC PA-100) was eluted using a linear gradient of 0-70% buffer B over 30 min (buffer A: 100 mM Tris HCl, pH 7.5, 10% MeCN and buffer B: 100 mM Tris HCl, pH 7.5, 10% MeCN, 1 M NaCl)

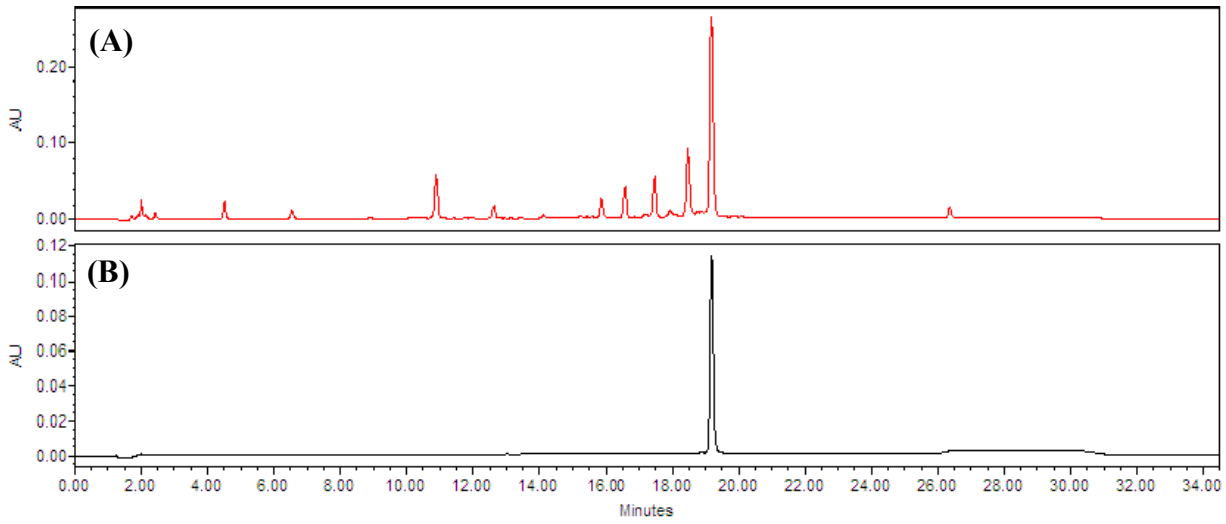


Figure A3.28- SAX-HPLC profile of **dClU-Me** crude **(A)** and pure **(B)**. The column (Dionex DNAPAC PA- 100) was eluted using a linear gradient of 0-70% buffer B over 30 min (buffer A: 100 mM Tris HCl, pH 7.5, 10% MeCN and buffer B: 100 mM Tris HCl, pH 7.5, 10% MeCN, 1 M NaCl)

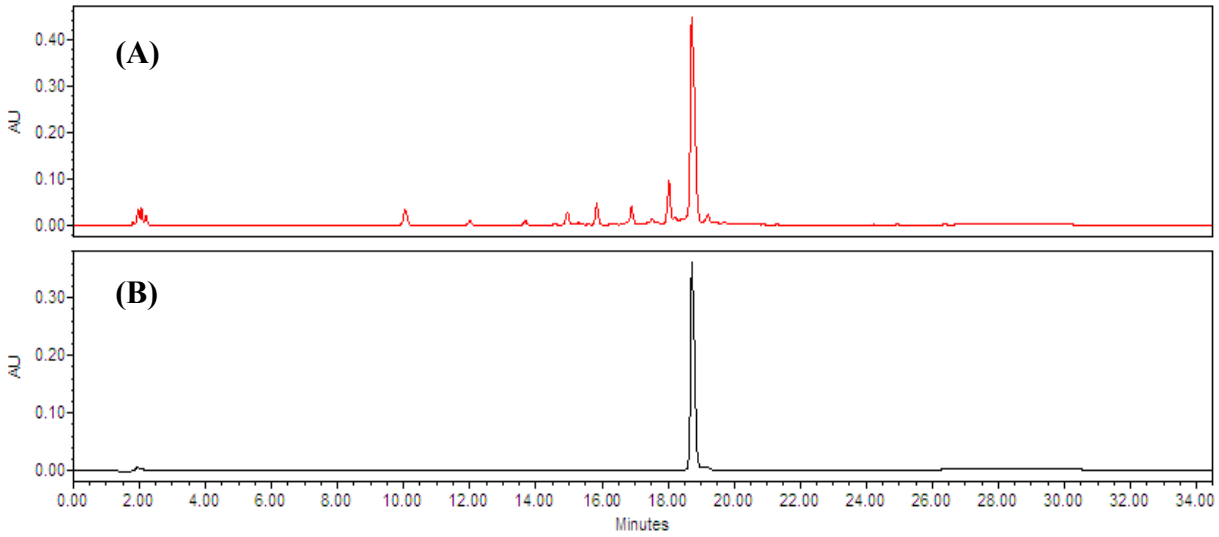


Figure A3.29- SAX-HPLC profile of **dClU** crude **(A)** and pure **(B)**. The column (Dionex DNAPAC PA-100) was eluted using a linear gradient of 0-70% buffer B over 30 min (buffer A: 100 mM Tris HCl, pH 7.5, 10% MeCN and buffer B: 100 mM Tris HCl, pH 7.5, 10% MeCN, 1 M NaCl)

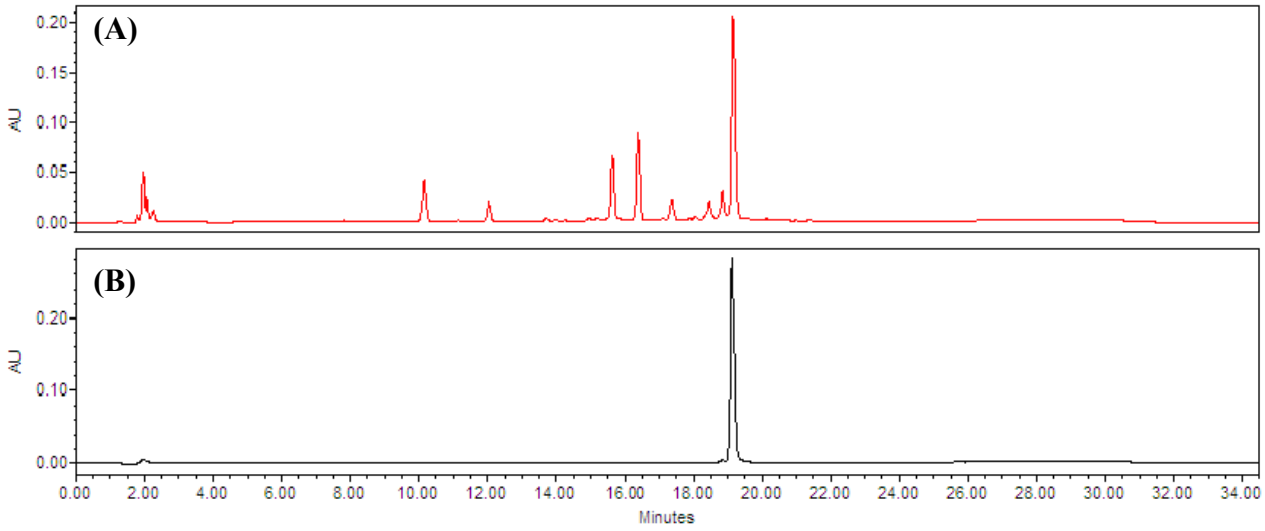


Figure A3.30- SAX-HPLC profile of **dBrU-Me** crude **(A)** and pure **(B)**. The column (Dionex DNAPAC PA- 100) was eluted using a linear gradient of 0-70% buffer B over 30 min (buffer A: 100 mM Tris HCl, pH 7.5, 10% MeCN and buffer B: 100 mM Tris HCl, pH 7.5, 10%

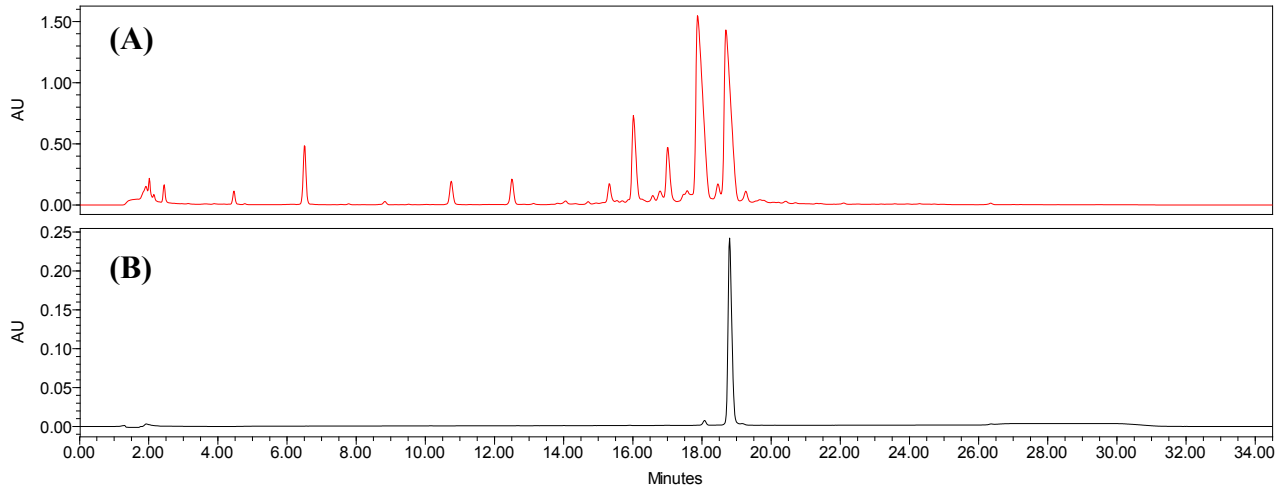


Figure A3.31 - Hyperchromicity change (A_{260}) *versus* temperature (^{o}C) profiles of duplexes containing **dClU** (—), **dClU-Me** (•••), (———), **dBrU** (———), and **dBrU-Me** (—•—).

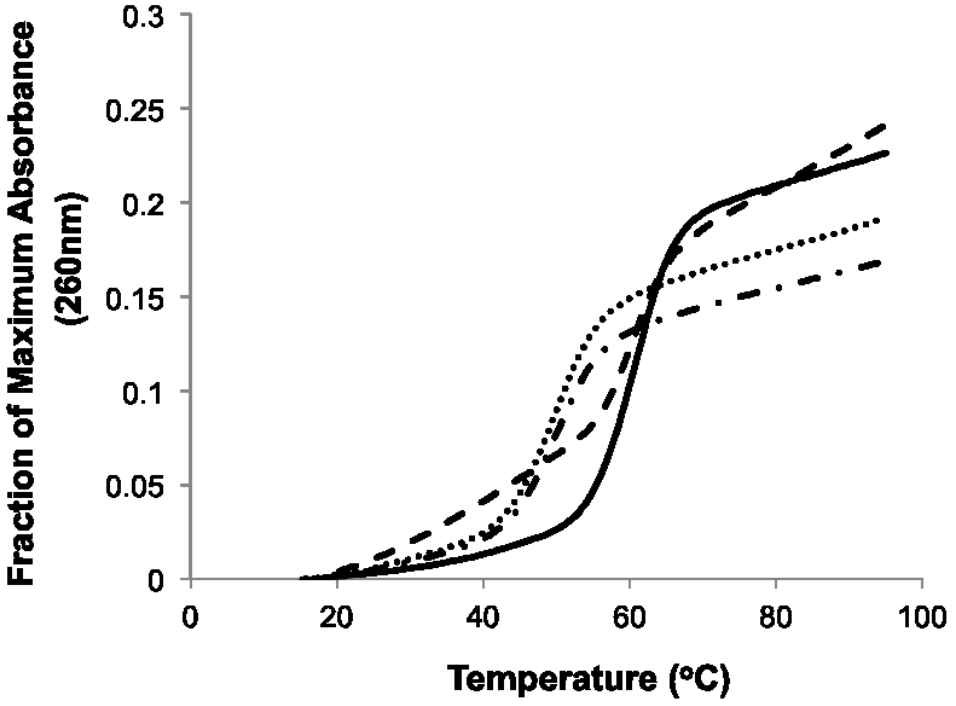


Figure A3.32 - Circular dichroism spectra of duplexes containing **dClU** (—), **dClU-Me** (•••), **dBrU** (———), and **dBrU-Me** (—•—).

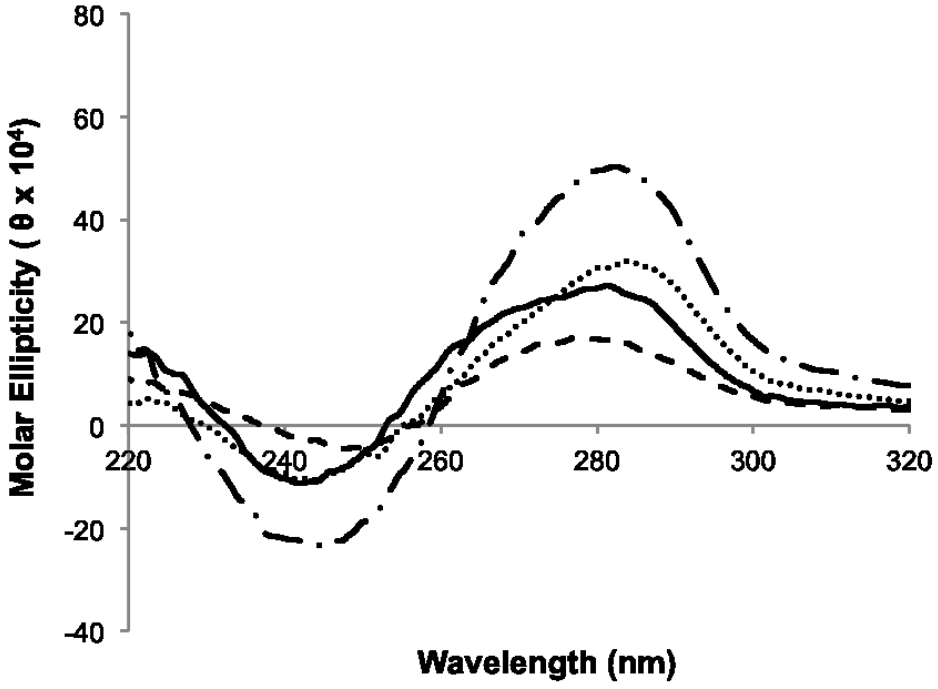


Figure A3.33: Repair of (**A**) **dClU** (2 pmol) and (**B**) **dClU-Me** (2 pmol) by hAGT (10 pmol), OGT (10 pmol), Ada-C (10 pmol), and hOGT (10 pmol), for 2.5 h at 37^{o}C. Panel **A**; Lane 1, **dFU** DNA; lane 2, **dClU** + *Bcll*; lane 3, **dClU** + hAGT; lane 4, **dClU** + hAGT + *Bcll*; lane 5, **dClU** + OGT; lane 6, **dClU** + OGT+ *Bcll*; lane 7, **dClU** + Ada-C; lane 8, **dClU** + Ada-C + *Bcll*; lane 9, **dClU** + hOGT; lane 10, **dClU** + hOGT + *BclI*. Panel **B**; Lane 1, **dClU-Me** DNA; lane 2, **dClU-Me** + *Bcll*; Lane 3, **dClU-Me**; lane 4, **dClU-Me** + hAGT; lane 5, **dClU-Me** + hAGT; lane 6, **dClU-Me** + hAGT + *Bcll*; lane 7, **dClU-Me** + OGT; lane 8, **dClU-Me** + OGT+ *Bcll*; lane 9, **dClU-Me** + Ada-C; lane 10, **dClU-Me** + Ada-C + *Bcll*; lane 11, **dClU-Me** + 10 hOGT; lane 12, **dClU-Me** + hOGT + *Bcll*.

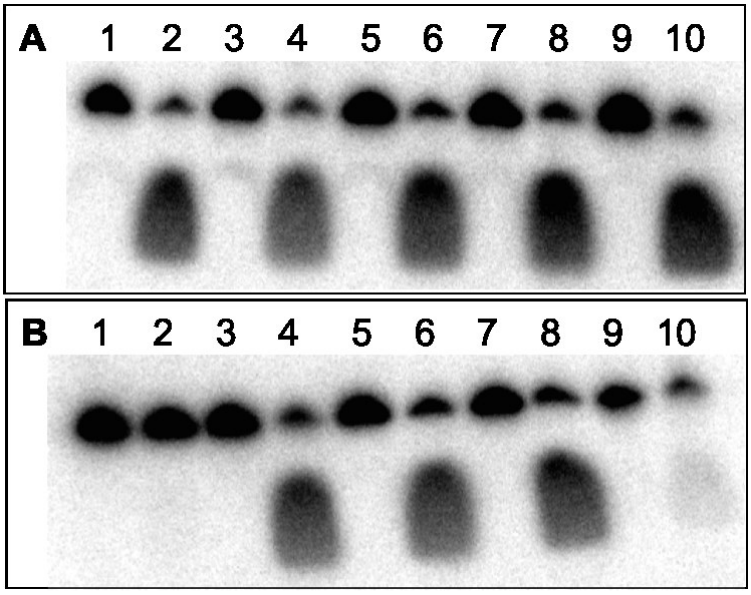


Figure A3.34: Repair of (**A**) **dBrU** (2 pmol) and (**B**) **dBrU-Me** (2 pmol) by hAGT (10 pmol), OGT (10 pmol), Ada-C (10 pmol), and hOGT (10 pmol), for 2.5 h at 37^{o}C. Panel **A**; Lane 1, **dBrU** + hAGT; lane 2, **dBrU** + hAGT + *BclI*; lane 3, **dBrU** + OGT; lane 4, **dBrU** + OGT+ *BclI*; lane 5, **dBrU** + Ada-C; lane 6, **dBrU** + Ada-C + *BclI*; lane 7, **dBrU** + hOGT; lane 8, **dBrU** + hOGT + *BclI*. Panel **B**; Lane 1, **dBrU-Me** DNA; lane 2, **dBrU-Me** + *BclI*; Lane 3, **dBrU-Me**; lane 4, **dBrU-Me** + hAGT; lane 5, **dBrU-Me** + hAGT; lane 6, **dBrU-Me** + hAGT + *BclI*; lane 7, **dBrU-Me** + OGT; lane 8, **dBrU-Me** + OGT+ *BclI*; lane 9, **dBrU-Me** + Ada-C; lane 10, **dBrU-Me** + Ada-C + *BclI*; lane 11, **dBrU-Me** + 10 hOGT; lane 12, **dBrU-Me** + hOGT + *BclI*.

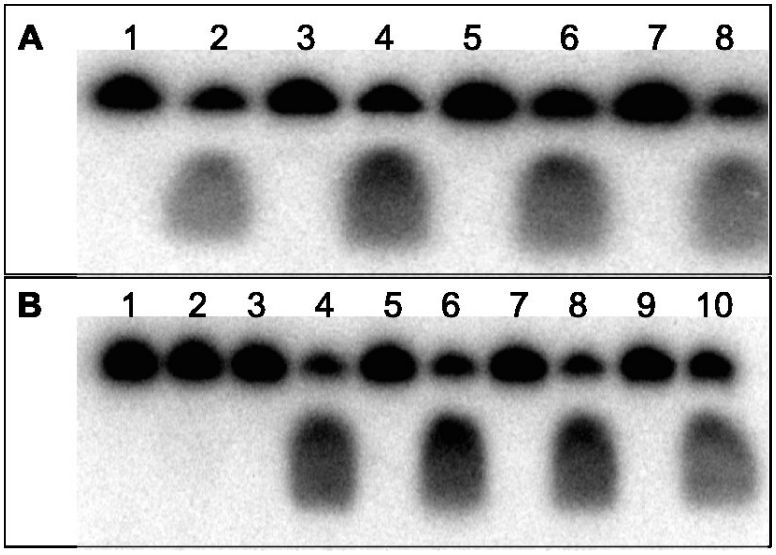


Figure A3.35 - Time course repair gel of duplexes containing **dClU-Me** by (**A**) **hAGT**, (**B**) **OGT**, (**C**) **Ada-C** and (**D**) **hOGT**. (**A**) Denaturing gel of the repair of 2 pmol of **dClU-Me** by 10 pmol hAGT as a function of time: lane 1, 2 pmol Control; lanes 2-10, 2 pmol + 10 pmol hAGT incubated for 0.25, 0.5, 0.75, 1, 1.5, 3, 5, 15, 45 min, respectively. (**B**) Denaturing gel of the repair of 2 pmol of **dClU-Me** by 10 pmol OGT as a function of time: lane 1, 2 pmol Control; lanes 2-10, 2 pmol + 10 pmol OGT incubated for 0.25, 0.5, 0.75, 1, 1.5, 3, 5, 15, 45, 90 min, respectively. (**C**) Denaturing gel of the repair of 2 pmol of **dClU-Me** by 10 pmol Ada-C as a function of time: lane 1, 2 pmol Control; lanes 2-10, 2 pmol + 10 pmol Ada-C incubated for 0.25, 0.5, 0.75, 1, 1.5, 3, 5, 15, 45, 90, 150 min, respectively. (**D**) Denaturing gel of the repair of 2 pmol of **dClU-Me** by 10 pmol hOGT as a function of time: lane 1, 2 pmol Control; lanes 2-10, 2 pmol + 10 pmol hOGT incubated for 0.25, 0.5, 0.75, 1, 1.5, 3, 5, 15, 45, 90, 150 min, respectively. All the time courses were performed at room temperature.

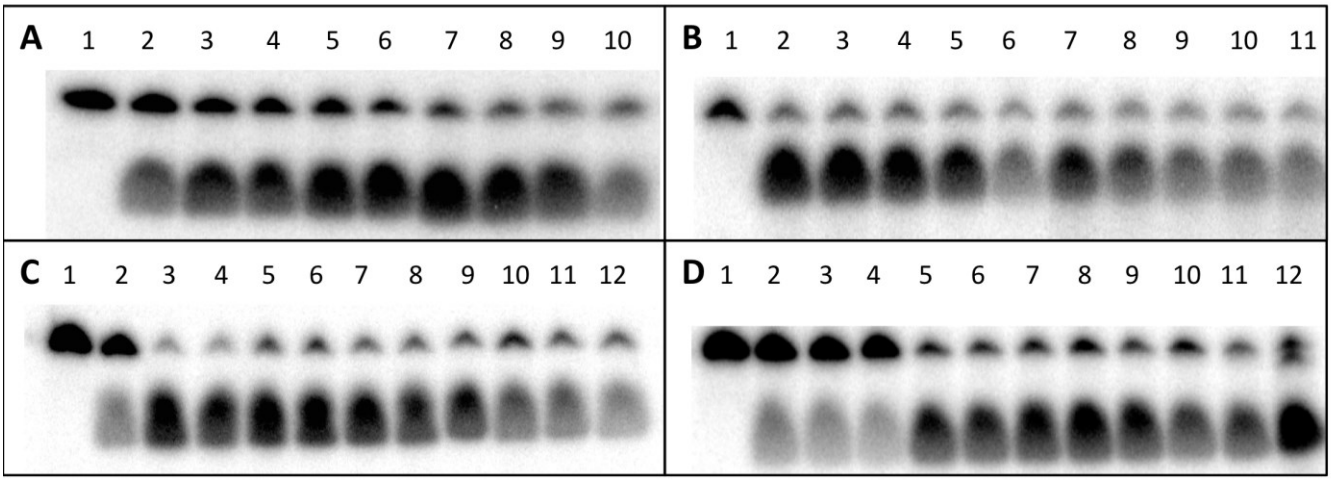


Figure A3.36 - Time course repair gel of duplexes containing **dBrU-Me** by (**A**) **hAGT**, (**B**) **OGT**, (**C**) **Ada-C** and (**D**) **hOGT**. (**A**) Denaturing gel of the repair of 2 pmol of **dBrU-Me** by 10 pmol hAGT as a function of time: lane 1, 2 pmol Control; lanes 2-10, 2 pmol + 10 pmol hAGT incubated for 0.25, 0.5, 0.75, 1, 1.5, 3, 5, 15, 45, 90 min, respectively. (**B**) Denaturing gel of the repair of 2 pmol of **dBrU-Me** by 10 pmol OGT as a function of time: lane 1, 2 pmol Control; lanes 2-10, 2 pmol + 10 pmol OGT incubated for 0.25, 0.5, 0.75, 1, 1.5, 3, 5, 15, 45, 90 min, respectively. (**C**) Denaturing gel of the repair of 2 pmol of **dBrU-Me** by 10 pmol Ada-C as a function of time: lane 1, 2 pmol Control; lanes 2-10, 2 pmol + 10 pmol Ada-C incubated for 0.25, 0.5, 0.75, 1, 1.5, 3, 5, 15, 45, 90, 150 min, respectively. (**D**) Denaturing gel of the repair of 2 pmol of **dBrU-Me** by 10 pmol hOGT as a function of time: lane 1, 2 pmol Control; lanes 2-10, 2 pmol + 10 pmol hOGT incubated for 0.25, 0.5, 0.75, 1, 1.5, 3, 5, 15, 45, 90, 150 min, respectively. . All the time courses were performed at room temperature.

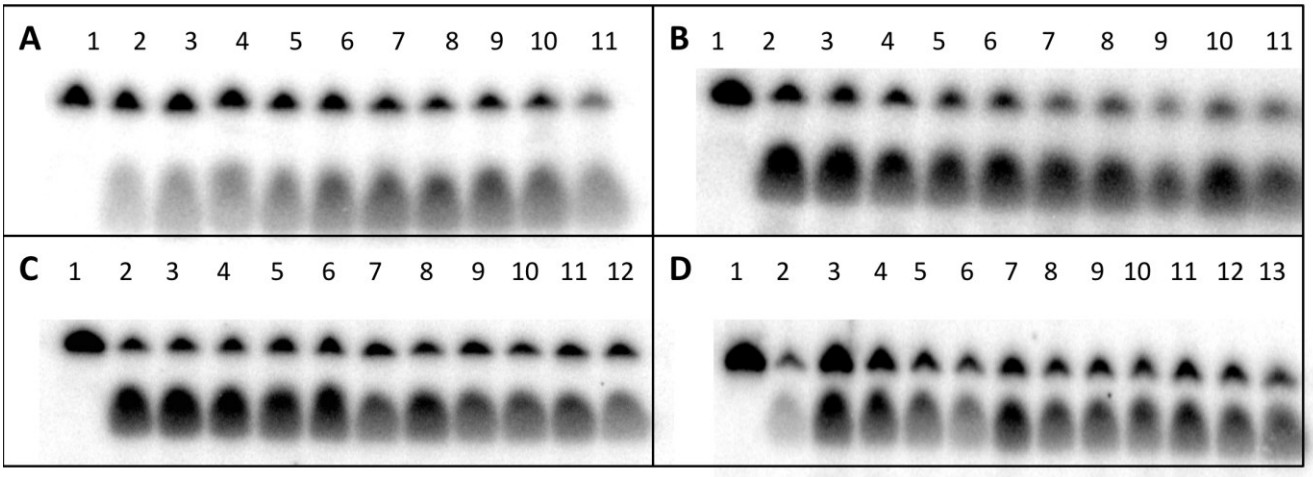
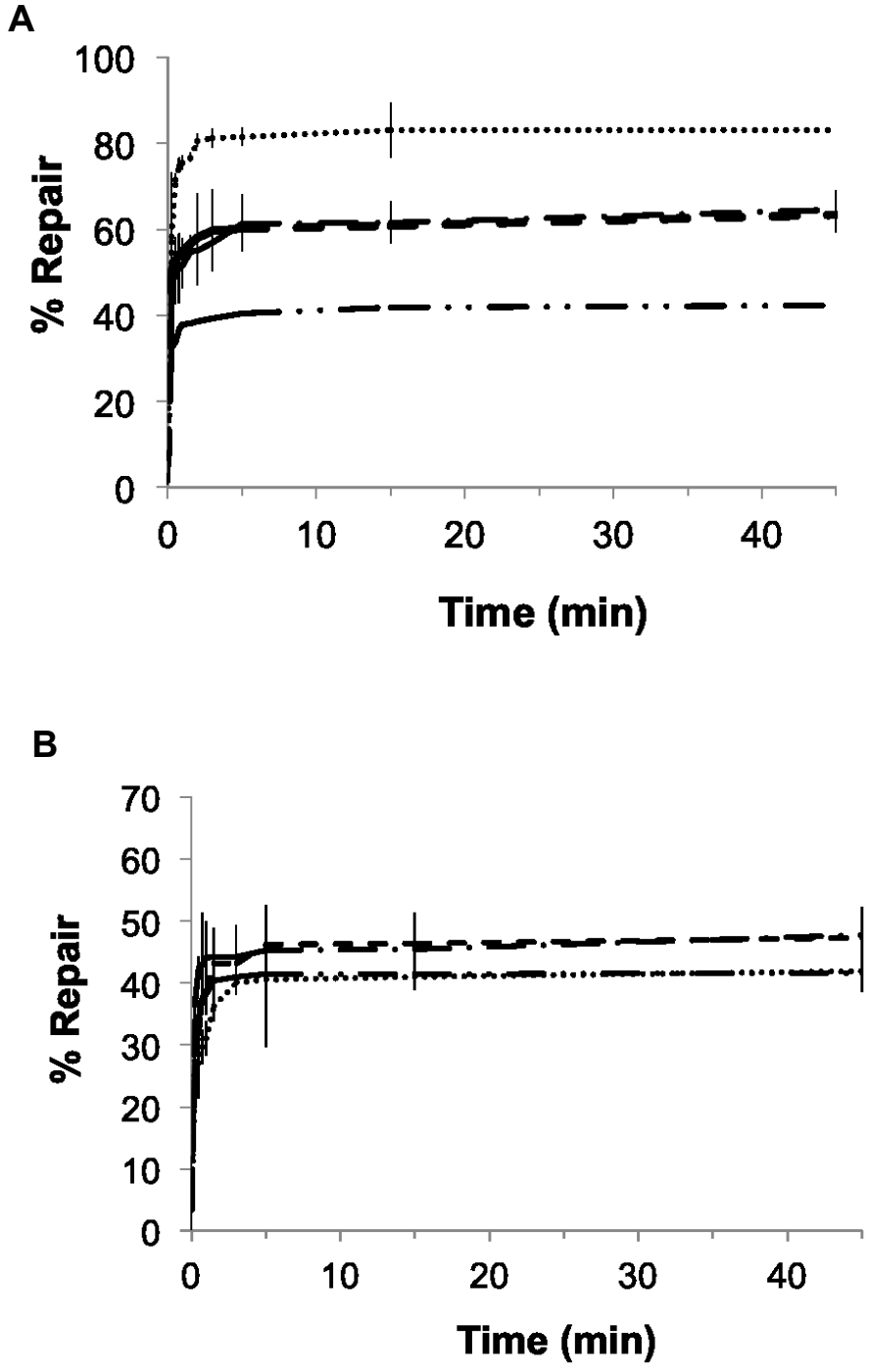


Figure A3.37 –Time course repair assay graphs of (**A**) **dClU-Me** and (**B**) **dBrU-Me** by hAGT (•••), OGT (———), Ada-C (—•—) and hOGT (—••—).



Appendix IV: Formation of Small Circular Oligonucleotides by using a Removable Crosslink

Contents **Page**

Supporting Methods

A4.1 Abstract **212**
A4.2 Main Text **213**

Supporting Figures

Figure A4.1- Oligonucleotides used for the study **213**
Figure A4.2- 40mer sequence as a holiday juction **214**
Figure A4.3 - Bisamidite with divinylsulfone linker **215**
Figure A4.4 - First attempts of phosphorylation and cyclization **216**
Figure A4.5 - Scheme of phosphorylation, cyclization and removal of crosslink **217**
Figure A4.6 - Results of the 40mer cyclization reaction **217**
Figure A4.7 - Removal of the divinylsulfone crosslink **218**
Figure A4.8 - S1 digestion of the phosphorylated and cyclic 40mer sequence **219**

A4.1 Abstract

The use of crosslinks has been explored in order to help generate small circular single stranded DNA. Different sequences have been explored in order to achieve circularization. In this specific case, the advantage of being able to remove the crosslink once the oligonucleotide is circularized plays an attractive role in therapeutic and diagnostic applications using small circular oligonucleotides as molecular recognition agents. Recent studies have also shown the use of small circular DNAs as a base for the generation of DNA nanotubes.[136]

A4.2 Main text

Another type of DNA damage that can occur is the formation of crosslinks. In this section we are using a crosslink as a splint in order to generate a single-stranded circular DNA. The crosslink used corresponds to divinylsulfone, which can be removed under basic conditions (0.05M K_2CO_3 in methanol for more than 48 h). Different sequences with different length and base composition were studied. For instance, the sequences have a tendency to form tetraloops (GATA) or a triloop (GAA) in the case of the smallest sequence. They also form a CG loop-closing base pair.[137–139] 20, 28, 32 and 40 nucleotide long sequences were studied. In the preliminary ligation studies, it was observed that the smaller sequences had difficulty to undergo ligation. It was thought that the circularization in the smaller sequences was more challenging due to their reduced flexibity and the greater influence of the cross-link, which may induce structural deformations closer to the 5' phosphate end affecting the ligation.

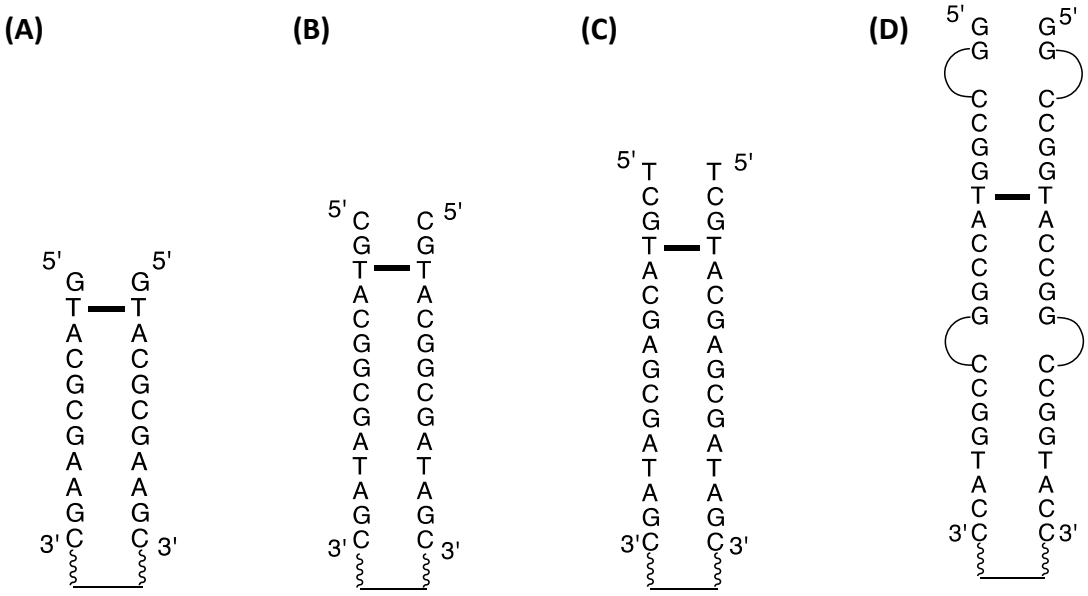


Figure A4.1: Oligonucleotide sequences in the presence of the divinylsulfone crosslink; (A) 20mer, (B) 28mer, (C) 32mer and (D) 40mer.

Out the sequences studied, the 40mer showed the most promise to undergo ligation with the formation of some multimeric species. Therefore, the conditions for the ligation were optimized in order to reduce the production of multimeric species and favour the intramolecular ligation. It is important to mention that the 40mer sequence also has the ability to form a Holiday Junction which could have facilitated the circularization.[140]

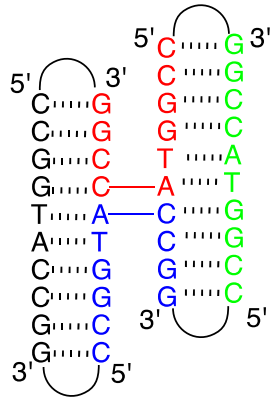


Figure A4.2: 40mer sequence representing the possible formation of a Holiday Junction

A4.2.1 Bisamidite of the nucleoside dimer

The following procedure and characterization of the bisamidite (Figure A4.3) was obtained from Roger Tresánchez, PhD Thesis, University of Barcelona, 2013.

To a solution of the nucleoside dimer (1.96 g, 1.62 mmol, 1 eq) in anhydrous DCM (25 mL) (in a dry 10 mL-round bottomed flask) DIPEA (1.15 mL, 6.59 mmol, 4 eq) was added under an Ar atmosphere. After 5 min stirring, Cl-P(OCNE)$NiPr_2$ (1.00 g, 4.22 mmol, 2.6 eq) was added, and the mixture was reacted for 2.5 h (Ar). The reaction mixture was diluted with DCM (100 mL) and washed with 10% aq. $NaHCO_3$ (3 × 100 mL). The organic fraction was dried over anhydrous $MgSO_4$ and taken to dryness. Purification: silica gel column chromatography eluting with hexanes/AcOEt/TEA mixtures (from 48:50:2 to 28:70:2). 2.00 g of product were obtained (77 % yield).

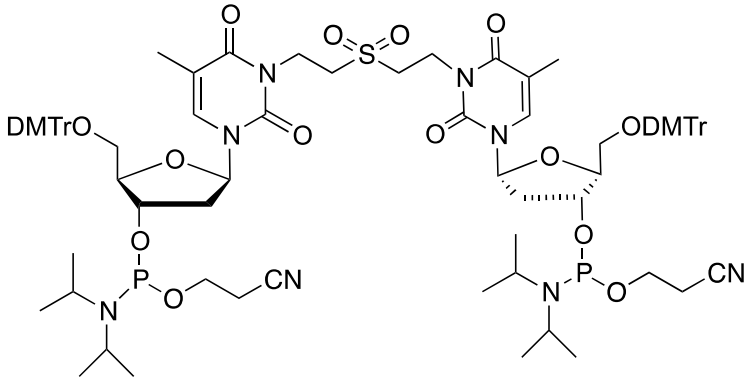


Figure A4.3: Bisamidite with divinylsulfone linker

White solid; **R_f:** 0.74 (AcOEt/Hexanes 7:3); **^{1}H NMR** (400 MHz, $CDCl_3$) δ (ppm): 7.67, 7.63, (2 s, 1H, H6), 7.41–7.23 (m, 9H, Ar-DMT), 6.85–6.6.82 (m, 4H, Ar-DMT), 6.39 (q, *J*=8 Hz, 1H, H1'), 4.67-4.61 (m,1H,H3'),4.17-4.13 (2q, *J*=4 Hz, 1H, H4'), 4.46–4.41 (m, 2H, $CH_2\alpha$), 3.79 (s, 6H, OMe), 3.82–3.69 (m, 2H, $CH_2\alpha'$), 3.63–3.47 (m, 4H, $CH_2\beta$, CH ^{i}Pr), 3.33-3.28 (m, 2H, H5' i H5''), 2.62 (t, *J*=6 Hz) & 2.41 (t, *J*=6 Hz) (2H, $CH_2\beta'$), 2.57-2.46 (m) & 2.37–2.30 (m) (2H, H2', H2''), 1.43 (s, 3H, CH_3), 1.17, 1.16, 1.05 (3d, *J*=4 Hz, 12H, CH_3 ^{i}Pr); **^{13}C NMR** (101 MHz, $CDCl_3$) δ (ppm): 163.33 (C4), 158.88 (Ar-DMT), 150.74 (C2), 144.39, 133.52 , 133.46 (Ar-DMT), 134.36 (C6), 130.34, 128.41, 128.13, 127.28 , 113.40 (Ar-DMT), 110.46 (C5), 87.04 (C1'), 85.84, 84.65 (C4', Ar-DMT), 73.23 (C3'), 63.28 (C5'), 58.56 ($CH_2\alpha'$), 55.43 (OMe), 49.06 ($CH_2\beta$), 43,43 (CH ^{i}Pr), 40.40 (C2'), 35.22 ($CH_2\alpha$), 24.75 (CH_3 ^{i}Pr), 20.53 ($CH_2\beta'$), 12.55 (CH_3); **^{31}P NMR** (121 MHz, $CDCl_3$): 148.95, 148.93, 148.49 (3 diastereoisomers); **HRMS** *m/z*: 1624.6958 $[M+NH_4^+]^+$; (M_{calc} $C_{84}H_{108}N_9O_{18}P_2S^+$=1624.7003).

A4.2.2 Synthesis of cross-linked oligonucleotides.

All oligonucleotides were automatically synthesized with an Applied Biosystems Model 3400 synthesizer in a 1.5 µmol scale. Protecting group removal and cleavage from the solid support was carried out by treatment with 0.05M K_2CO_3 in MeOH for 4 h at room temperature with mild rocking in 2 mL screw-cap microfuge tubes fitted with teflon lined caps. The base was neutralized with an equimolar amount of AcOH and crude oligomers were transferred and lyophilized in a Speedvac concentrator. Purification was achieved by 20% polyacrylamide (19:1) denaturing gel electrophoresis, at least for the 40mer sequence.

A4.2.3 Phosphorylation of cross-linked oligonucleotides.

The required amount of cross-linked oligonucleotide was introduced into a microcentrifuge tube, and a solution (the volume necessary to yield a 10 µM solution of cross-linked oligonucleotide) containing 70 mM Tris·HCl pH=7.6, 10 mM $MgCl_2$, 5 mM DTT, 0.5 mM ATP and T4 PNK (NE BioLabs, 1 U/µL) was added. The mixture was incubated at 37 ^{o}C for 1 h. The reaction was quenched by heating at 70 ^{o}C for about 10 min, and this reaction mixture was used in the cyclization experiments.

Chemical phosphorylation was also attempted by solid-phase synthesis with the Applied Biosystems Model 3400 DNA synthesizer by using a commercially available chemical phosphorylating agent II ([3-(4,4'-Dimethoxytrityloxy)-2,2-dicarboxyethyl]propyl-(2-cyanoethyl)-(N,N-diisopropyl)-phosphoramidite from Glen Research).

A4.2.4 Cyclization of cross-linked oligonucleotides.

The mixture resulting from the phosphorylation reaction was heated at 95 ^{o}C for 10 min, and allowed to cool down slowly. A solution (the volume necessary to yield a 6 µM solution of cross-linked oligonucleotide) containing 50 mM Tris·HCl, 10 mM $MgCl_2$, 5 mM DTT, and T4 DNA ligase (NE BioLabs, 1 U/µL) was added, and the mixture reacted at 30 ^{o}C. After 20 minutes, the reaction was then quenched by heating at 70 ^{o}C for 15 min. Cyclization crudes were analyzed by analytical PAGE (450 V, 4 h) and/or purified by preparative PAGE (450 V, 4 h).

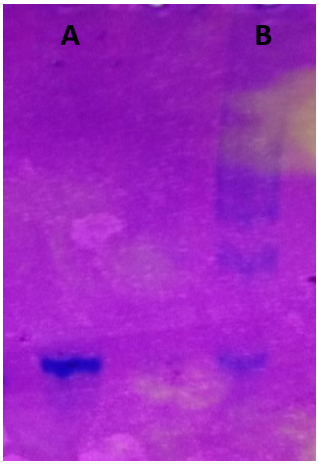


Figure A4.4: First attempts of the (A) Phosphorylation reaction of the 40mer and (B) Cyclization reaction of the 40mer (10% PAGE)

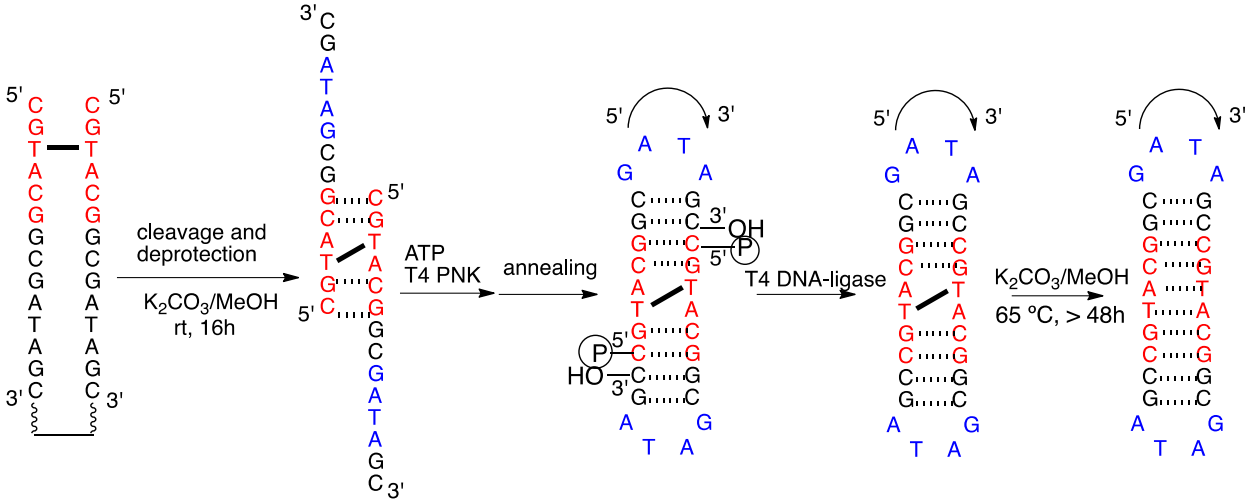


Figure A4.5: Scheme of the phosphorylation and cyclization of the oligonucleotides and the site of the divinylsulfone crosslink.

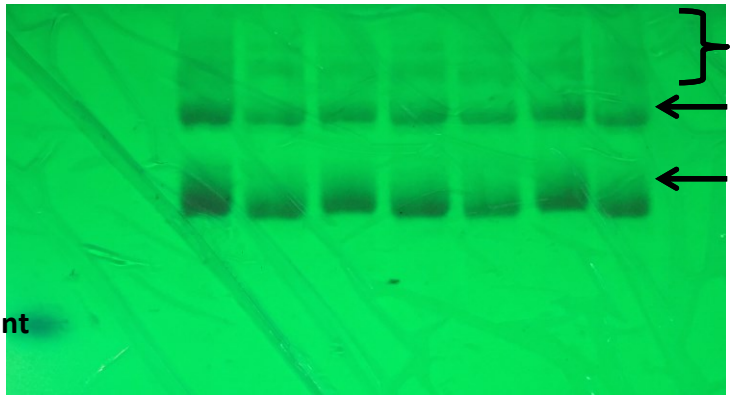


Figure A4.6: Results after the cyclization reaction of the 40mer sequence (20% PAGE) for cyclic product recovery.

From the results obtained in figure A4.4 and A4.6 it can be summarized that the phosphorylated product and the ligated product migrated at different levels and that the reaction did not proceed 100%, leaving about 50% of phosphorylated product unreacted.

A4.2.5 Removal of cross-link

About 0.5 OD of the cyclic oligonucleotide was dissolved with a solution of 0.05 M K_2CO_3 in methanol and left at 65^{o}C for 72 hours with some gently shaking in between. After the 72 hours, an equivalent amount of AcOH was added to neutralize the reaction. As it can be observed in figure A4.7, it seems that 72 hours were not enough to completely remove the crosslink. This suggests that the sequence might be adopting a structure that could be minimizing the exposure of the crosslink. The results obtained also showed that after the removal of the crosslink, the cyclic oligonucleotide seemed to adopt a more flexible structure (like a supercoiled) allowing it to travel a bit further in the gel. It is important to mention that only two products were observed, which implies that the cyclic product was successfully ligated as no smaller products (20mer), which would have been generated as a result of the removal of the crosslink in the non-ligated sequences, appeared in the gel.

Further work has to be performed; for example: trying a more concentrated sample (minimum 1 OD), higher temperatures such as 80 ^{o}C, going over 72 hours and adding extra solution of 0.05 M K_2CO_3 in methanol in order to compensate for the possible MeOH evaporation, could be good conditions to try.

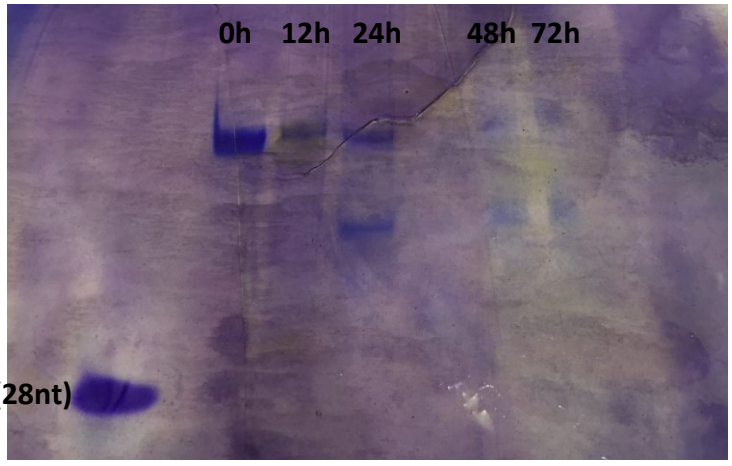


Figure A4.7: Removal of the divinylsulfone crosslink

Due to the high molecular weight of the 40mer sequence, the circular product was not able to be analyzed by mass spectrometry. Another way to confirm the circularity was to perform S1 endonuclease digestion.

A4.2.6 S1 Endonuclease Digestion

About 0.05 OD of the cyclic oligonucleotide was dried and 3 µL of 5X S1 buffer containing 200 mM sodium acetate (pH 4.5 at 25°C), 1.5 M NaCl and 10 mM $ZnSO_4$, 0.5 µL of S1 nuclease (Thermo Scientific 100U/µL) and 10.5 µL of deionized water were added. The reaction mixture was incubated at 37 ^{o}C for 10 min and it was later stopped by heating to 70 ^{o}C for another 10 min. The challenge of this digestion is to stop the reaction right on the first cleavage without letting further cleavage of the sample. The results obtained are shown in figure A4.8 where it can be observed that the first product of the digestion of the cyclic 40mer traveled similarly to the intact phosphorylated 40mer.

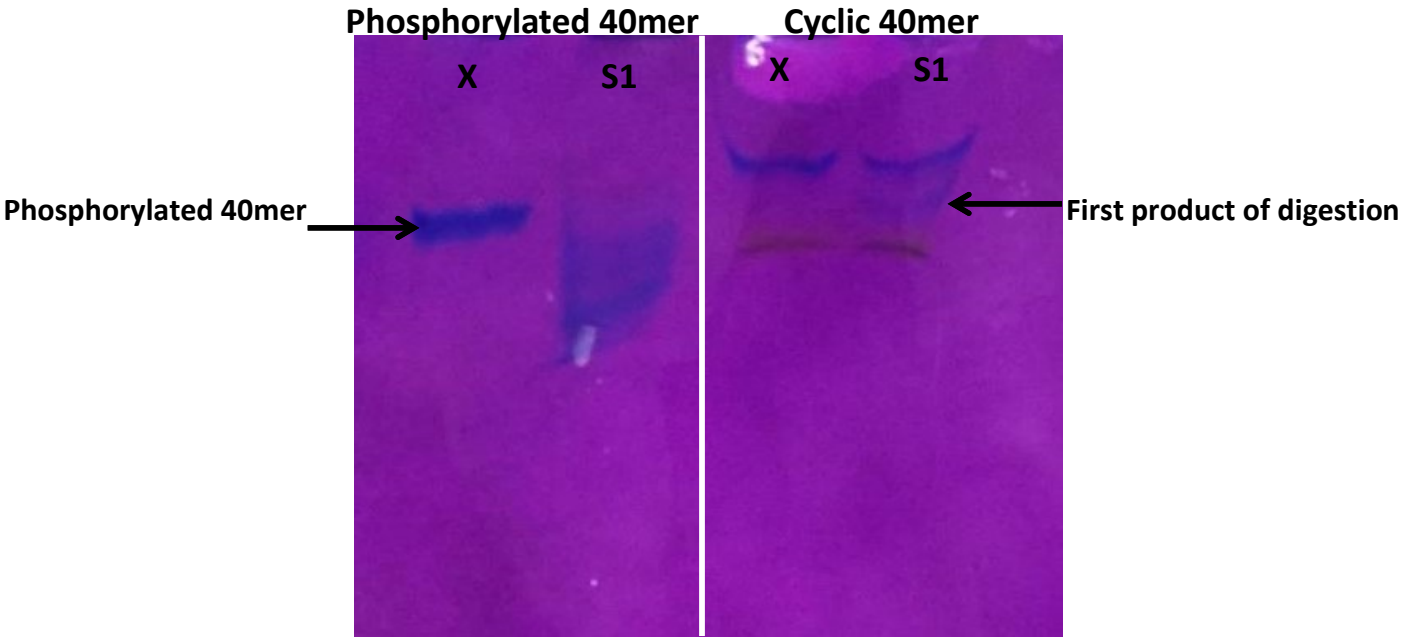


Figure A4.8: Digestion of the phosphorylated and cyclic 40mer sequence, both in the absence (X) and presence (S1) of S1 endonuclease.

After performing the S1 nuclease digestion and removing the crosslink of the 40mer sequence, the results suggested the possibility of successful ligation. Future work will required to collect more cyclic 40mer and repeat the steps of crosslink removal and S1 endonuclease digestion to confirm ligation.