Repair of Intrastrand Alkylene Cross-links in Various DNA Structures by O^6-Alkylguanine DNA Alkyltransferases

William Copp

A Thesis

In the Department

of

Chemistry and Biochemistry

Presented in Partial Fulfillment of the Requirements

for the Degree of

Doctor of Philosophy (Chemistry) at

Concordia University

Montréal, Québec, Canada

August 2021

CONCORDIA UNIVERSITY SCHOOL OF GRADUATE STUDIES

This is to certify that the thesis prepared

By: William Copp

Entitled: Repair of Intrastrand Alkylene Cross-links in Various DNA Structures by O^6-alkylguanine DNA Alkyltransferases

and submitted in partial fulfillment of the requirements for the degree of
Doctor of Philosophy (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:

_____ Chair
Dr. Rachael (Ré) Mansbach

_____ External Examiner
Dr. Jean- Paul Desaulniers

_____ External to Program
Dr. David Kwan

_____ Examiner
Dr. Louis Cuccia

_____ Examiner
Dr. Alisa Piekny

_____ Supervisor
Dr. Chris Wilds

Approved by: _____
Dr. Yves Gelinas – Graduate Program Director

August 24, 2021
Date of Defence

_____
Dr. Pascale Sicotte – Dean, Arts and Science

Abstract

Repair of Intrastrand Alkylene Cross-links in Various DNA Structures by O^6-Alkylguanine DNA Alkyltransferases

William Copp, PhD
Concordia University, 2021

O^6-alkylguanine DNA alkyltransferases (AGTs) are proteins found in all kingdoms of life except plants. The protein's primary role is the removal of methyl adducts from the O^6 position of 2'-deoxyguanosine (dG). However, the human protein (hAGT) has been shown to remove bulkier lesions such as benzyl groups as well inter- and intrastrand alkylene cross-links (ICLs and IaCLs, respectively) from the O^6 position of dG. Repair in canonical duplex DNA has been well studied however DNA can show vast structural polymorphism. Due to its excellent programmability, ease of chemical synthesis and low cost it has been used as a nanomaterial. While initial structures were confined to two dimensions, many polyhedra have now been constructed such as tetrahedra, cubes and icosahedra based on DNA's simple base-pairing rules. Applications of these structures have primarily been focused as a drug delivery vehicle and for biosensing. While these structures consist of the DNA double helix, DNA can fold into a variety of non-canonical structurers such as *i*-motifs, G-quadruplexes and triplexes. These structures have found use in nanotechnology but have also been shown to play a biological role. Modified nucleic acids that contain alternate nucleobases, backbones and sugars, which can be prepared by chemical synthesis, have been evaluated for various applications including as therapeutic agents. One example of a sugar modified nucleic acids, the threose nucleic acids (TNA) which contain a four instead of five carbon sugar have generated interest as a possible precursor to RNA in the study of origins of life.

This thesis explores four separate facets of repair in unusual nucleic acid based structures. It was investigated whether a chimeric AGT (human AGT with the active site of that of *E. coli* OGT) could act on IaCLs lacking a phosphate backbone as a means to covalently capture *E. coli* OGT's active site. Site specific DNA-protein cross-links were prepared with dG-O^6-alkylene-O^6-

dG and dT-O^6-alkylene-O^6-dG IaCLs in good yield as determined by denaturing polyacrylamide gel electrophoresis (PAGE) and electrospray ionization-mass spectrometry. A DNA tetrahedra was constructed containing dG-O^6-alkylene-O^6-dG three-way junctions at the vertices. These structures assembled in quantitative yield with either butylene or heptylene linkers as was assessed by PAGE under non-denaturing conditions. Furthermore, these DNA tetrahedra were well processed by human AGT at lower protein concentrations and essentially completely consumed at higher hAGT equivalents. Next, the effect of alkyl lesions at the O^6 position of dG on the biophysical properties of the G-quadruplex formed by the *c-myc* gene promoter region was explored. Repair of these lesions by human and *E. coli* AGTs was also evaluated. All alkyl lesions were shown to be thermally destabilizing to the G-quadruplex independent on chain length but damage on the base-paired guanine was more destabilizing than the guanine in the loop. The *E. coli* AGTs were able to act on methyl lesions however struggled to process the bulkier hydroxyheptylene and IaCLs lacking a phosphate backbone. hAGT was able to process all lesions however depending on the location of the lesions variable repair efficiency was observed: Alkyl damage on the internal guanine was more poorly repaired than the external. Lastly, the synthesis of a phosphoramidite derivative of a thymine threose nucleic acid containing an O^4 methyl adduct is reported. UV thermal denaturation studies of oligodeoxynucleotides containing this modification demonstrated a similar destabilizing effect to those observed with O^4-methyl dT when base-paired with adenine. A similar destabilization in duplex stability was observed when mismatched with a guanine residue. Unfortunately, repair was inconclusive and further assay development is required.

These results show promise for the novel DNA tetrahedra as a means for drug delivery, cellular entry and encapsulation with further investigation ongoing. The repair of alkylation damage in the promoter region G-quadruplex of an oncogene would provide important insights with ramifications in gene regulation and structural elucidation. Conformational changes of the G-quadruplexes after being acted upon by AGTs has additional potential in nanotechnology. Future assay development on repair by AGTs on O^4-methyl thymine in TNA to determine if DNA repair proteins act on TNA will provide interesting knowledge towards the chemical etiology of nucleic acid structures.

Acknowledgements

I would first and foremost like to thank my supervisor Dr. Christopher Wilds. Dr. Wilds accepted me as an MSc student in 2014 while I was working in industry and has been an exceptional mentor since. Dr. Wilds has consistently been there when I was struggling with research issues and was always happy to have a conversation. Through the guidance of Dr. Wilds I have learned to love and appreciate the building blocks of life, and have now found a career passion. The seven years spent under the guidance of Dr. Wilds have turned me from a student that was scared to give a group meeting presentation to having a relative air of confidence giving a research seminar. Dr. Wilds has strongly encouraged me over the years to present my research and I have been to several fruitful international conferences thanks to this. The countless reference letters which I did not take for granted have been crucial in me achieving my career goals and I am extremely grateful. It will have always been an honour to have worked for Dr. Wilds.

I am very appreciative of my committee members Dr. Louis Cuccia and Dr. Alisa Piekny. They have given me excellent guidance throughout my PhD, and their suggestions have been instrumental in my research projects' directions.

I would like to thank Dr. Anne Noronha for her training and without whom I would not have been able to do any oligonucleotide work.

The technical staff at Concordia has been exceptional. Alain Tessier and Dr. Heng Jiang's work at CBAMS has promptly given me great results (and when it wasn't great it was 100 % always my sample). I would like to thank Dr. Alexey Denisov and Dr. Costantin Yannopolous for their excellent help with the NMR work. I have extremely appreciative of Dr. Findlay for graciously allowing me to use his lab space for all the radioactivity work.

Collegial labmates past and present have created an excellent work environment and it has been a pleasure to have met them. Coming to the lab has always been a pleasure and I would like to thank Dr. Derek O'Flaherty, Laura Sacre, Chris Liczner, Gianna DiCenso, Alex Pontarelli, Gabrielle Juneau, Kieran Duke, Andrew Anderson and others over the past 7 years.

Outside of the lab I would like to lend a special thanks to Dr. Justin DiTrani who I have had many thoughtful conversations with. Dr. Robert Harkness the Fifth was very helpful with my G-quadruplex studies of which I am quite grateful. I would also like to thank Donald Campbell, Fadil Taj and Daniel Mangel.

Perhaps most importantly I would like to thank my family, David, Martine and Sebastian for their relentless support. Throughout this degree I experienced lot of highs and lows and they were always supportive unconditionally. The look on their faces whenever I achieved (mostly minor) successes was a major driving force for me. Je tiens à reconnaître mon Grandpère, qui a toujours encouragé et soutenu mes etudes dès mon enfance. Ma Grandmère a été une des plus grandes influences dans ma vie, et je ne serais pas ici aujourd'hui sans elle. Je sais qu'elle serait fière de moi.

Table of Contents

List of Figures

List of Tables

List of Schemes

List of Abbreviations

ABI: Applied Biosystems Inc.

AFM: Atomic force microscopy

AGT: O^6-alkylguanine DNA alkyltransferase

Alloc: Allyloxycarbonyl

ASO: Antisense oligonucleotide

BPB: Bromophenol blue

BER: Base excision repair

bp: Base-pair

CD: Circular dichroism

CE: Cyanoethyl

CFL: Confocal laser scanning microscopy

CPG: Controlled pore glass

C_t: Single strand concentration

dA: 2'-Deoxyadenosine

dC: 2'-Deoxycytidine

DCM: Dichloromethane

dG: 2'-Deoxyguanosine

DIAD: Diisopropyl azodicarboxylate

DIPEA: N,N-Diisopropylethylamine

DOX: Doxorubicin

dT: Thymidine

DMAP: Dimethylaminopyridine

DMTr: Dimethoxytrityl

DNA: 2'-Deoxyribonucleic Acid

DPC: DNA-protein cross-link

DTT: Dithiolthreitol

dU: 2'-Deoxyuridine

EDAC: 1-Ethyl-3-(3-dimethylaminopropyl)carbodiimide

EDTA: Ethylenediaminetetraacetic acid

EtOAc: Ethyl acetate

ETT: 5-(Ethylthio)-1H-tetrazole

FBS: Fetal bovine serum

FRET: Fluorescence resonance energy transfer

GFP: Green fluorescent protein

IaCL: Intrastrand cross-link

ICL: Interstrand cross-linnk

IEX: Ion exchange

Im: Imidazole

Lev: Levulinyl

HPLC: High performance liquid chromatography

HRMS: High-resolution mass spectrometry

HTH: Helix-turn-helix

Pac: Phenoxyacetyl

PAGE: Polyacrylamide gel electrophoresis

Py: Pyridine

MeCN: Acetonitrile

MeOH: Methanol

miRNA: MicroRNA

MNU: *N*-methyl-*N*′-nitrosourea

MMR: Mismatch repair

MMS: Methylmethane sulfonate

MP: Median product

MS: Mass spectrometry

NaOMe: Sodium methoxide

NER: Nucleotide excision repair

NMR: Nuclear magnetic resonance

nt: Nucleotide

OD: Optical density

ODN: Oligonucleotide

PAGE: Polyacrylamide gel electrophoresis

PS: Phosphorothioate

siRNA: small interfering RNA

SDS: Sodium dodecyl sulphate

RNA: Ribonucleic acid

RP: Reverse phase

RP: Repair product

r.t.: Room temperature

TBAF: Tetrabutylammonium fluoride

TBS: *tert*-Butyldimethylsilyl

TBE: Tris-boric acid-EDTA

TCA: Trichloroacetic acid

TEA: Triethylamine

TEMED: Tetramethylethylenediamine

THF: Tetrahydrofuran

TLC: Thin layer chromatography

tT: TNA thymine

TWJ: Three-way junction

qTOF-MS: Quadrupole-time of flight mass spectrometry

T_m: Melting temperature

UV: Ultraviolet

XC: Xylene cyanol

Contribution of authors

Chapter 1 Section 1.3.3

"*Recent Advances of DNA Tetrahedra for Therapeutic Applications and Biosensing*"

Excerpt from W. Copp, A. Pontarelli, C. J. Wilds, *ChemBioChem* **2021**, *6*, 1–11.

Copp, W.: Wrote section on DNA tetrahedra therapeutics.

Pontarelli, A.: Wrote section on DNA tetrahedra diagnostics.

Wilds, C.J.: Project supervisor responsible for reviewing and editing manuscript.

Chapter 2

"*Covalent Capture of OGT's Active Site Using Engineered Human-E. coli Chimera and Intrastrand DNA Cross-links*"

Copp, W.: Performed denaturing PAGE repair assays, SDS-PAGE and prepared samples for ESI-MS. Wrote 1st draft of manuscript, prepared figures and processed data.

O'Flaherty, D.K.: Synthesized and characterized all small molecules and oligonucleotides. Edited manuscript.

Wilds, C.J.: Project supervisor responsible for reviewing and editing manuscript.

Chapter 3

"*O^6-Alkylguanine DNA Alkyltransferase Mediated Disassembly of a DNA Tetrahedron*"

Copp, W.: Performed all experiments. Wrote 1st draft of manuscript, prepared figures and processed data.

Wilds, C.J.: Project supervisor responsible for reviewing and editing manuscript.

Chapter 4

"Repair of DNA Alkylation in the G-quadruplex Formed by the c-myc Promoter Region by O^6-Alkylguanine DNA Alkyltransferases"

Copp, W.: Performed all experiments. Wrote first draft of manuscript.

Wilds, C.J.: Project supervisor responsible for reviewing and editing manuscript.

Chapter 1

Introduction

1.1 Preamble

DNA is a polymer consisting of repeating nucleotides. Each nucleotide consists of a phosphate group, a 2'-deoxyribose sugar and a nucleobase. The dominant tertiary structure of DNA was elucidated by Watson and Crick in 1953 and was shown to be a double helix stabilized by a highly specific hydrogen bonding network as well as base stacking (**Figure 1.1A**).[1] The genetic information is contained in the identity of the nucleobase and the specific base-pairing pattern: guanine with cytosine and adenine with thymine (**Figure 1.1B**). Nucleoside numbering is shown in **Figure 1.2** and will be used throughout this thesis.

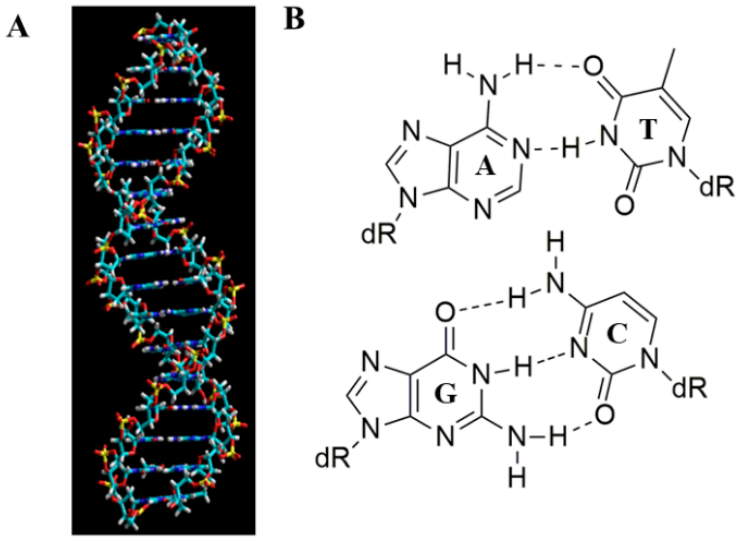


Figure 1.1 DNA structure. A) B-DNA double helix. Produced with HyperChem™. B) Watson-Crick base-pairing.

The high-fidelity incorporation of the appropriate nucleotide by DNA polymerases is crucial to cellular function as misincorporation can lead to disease. DNA is subject to damage from various sources such as electromagnetic radiation, oxidative stress and alkylation.[2] The assortment of lesions caused by this damage can be harmful to maintaining the integrity of the genome. Fortunately, the cellular machinery has various mechanisms to remove DNA damage through several repair pathways. Direct repair, base excision repair (BER), nucleotide excision repair (NER) and mismatch repair (MMR) are often employed to remove a wide array of damage.

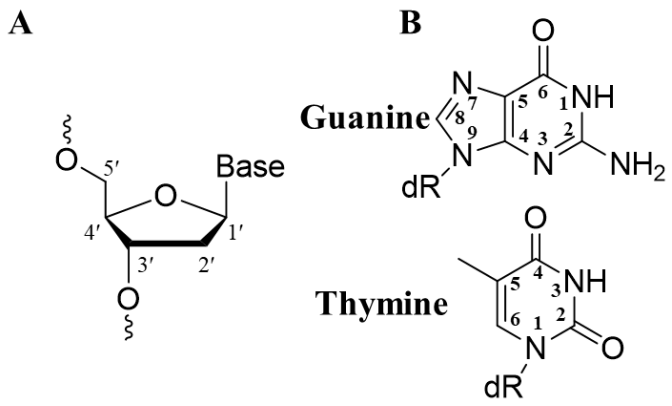


Figure 1.2 Sugar and nucleobase numbering A) 2'-Deoxyribose sugar B) Thymine and guanine nucleobases

A common form of damage is DNA alkylation which commonly occurs on the nucleobase. Damage at the O^6 position of guanine is particularly mutagenic and can lead to G:C→ A:T transitions. O^6-Alkylguanine DNA alkyltransferase (AGT) is a protein found in most organisms that irreversibly removes alkylation damage. It is a suicide protein as it acts only once and is then degraded by the ubiquitin pathway. The human protein (hAGT) has been shown to act on a variety of lesions at the O^6 position of dG including monoadducts, inter- and intrastrand cross-links (ICL and IaCL, respectively). IaCL and ICL can form in cells when treated with bifunctional alkylating agents such as the anticancer drugs busulfan and hepsulfam.

While the role of DNA in biology has been well established, the advent of automated solid-phase synthesis has enabled use of DNA as a nanomaterial. The ease of synthesis, cost effectiveness and programmability of DNA has led to its use to self-assemble into a wide variety of 2D and 3D shapes. These shapes have shown promise in therapeutics due to their ability to host guest molecules and enter cells. Furthermore, the polymorphism of DNA is further exhibited by the fact is not restricted to the double helix with Watson-Crick base-pairing and a plethora of other structures have been discovered such as DNA triplexes, *i*-motifs and G-quadruplexes. The applications of these alternate DNA structures have been considered for DNA nanotechnology. In this work we study how DNA alkylation on these structures, particularly using intrastrand cross-links (IaCLs) lacking a phosphate backbone, can be acted upon by AGTs. The potential for understanding the biology, as well as applications for drug delivery are herein examined.

1.2 DNA Alkylation and Direct Repair by O^6-Alkylguanine DNA Alkyltransferases

1.2.1 DNA Alkylation

DNA is constantly subject to damage from various exogenous and endogenous agents. Alkylation of DNA is particularly damaging and if left unrepaired is mutagenic and can potentially lead to cancer. It has been determined that S-adenosyl methionine is a common source of nonenzymatic methylation, ie damage, which can cause methylation on positions of DNA such as the N^7 and O^6 atoms of guanine [3] Examples of exogenous mutagenic agents include methylmethane sulfonate

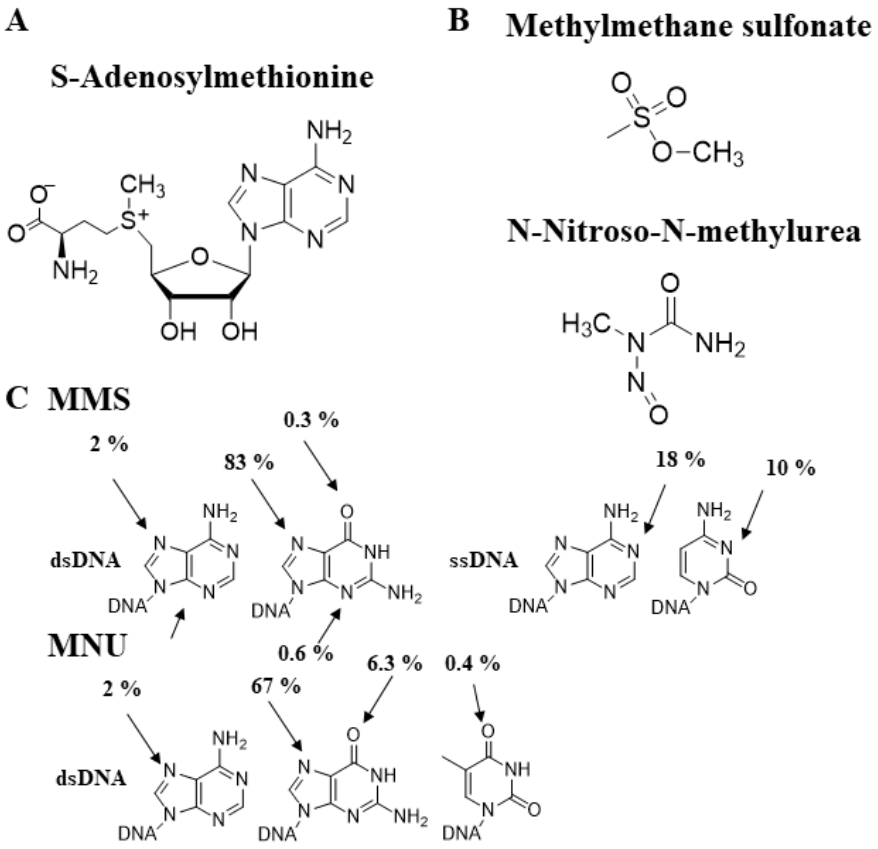


Figure 1.3 Methylating agents A) S-adenosyl methionine B) Exogeneous alkylation agents MMS and MNU C) Some sites and frequency of alkylation on DNA caused by monofunctional alkylating agents MMS and MNU. Adapted with permission from ref [4].

(MMS) and *N*-methyl-*N*′-nitrosourea (MNU) (**Figure 1.3A-B**) which primarily methylate the N^7 position of guanine with ~70-80 % of the methylation occuring there.[4] However, this lesion rapidly depurinates resulting in the formation of an abasic site which is repaired by cellular enzymes and is thus not thought to be mutagenic. The O^6 position of guanine and O^4 position of thymine however are quite stable and require cellular machinery to restore the integrity of the genome. As depicted in **Figure 1.3C** damage can occur on the nucleobase as well as the DNA backbone (depending on the alkylating agent and context). DNA is subject to a wide variety of damage which commonly involves the direct repair, BER and NER repair pathways.[5] Of note is that the broad range of alkylation damage is often able to be repaired by direct repair pathways.

While there is a large amount of types of DNA damage that can occur as a result of exposure to alkylating agents resulting in lesions that are mutagenic and potentially carcinogenic, paradoxically many chemotherapeutics operate by purposely damaging DNA. This damage, if left unrepaired, signals to the cell to stop diving and the cell undergoes apoptosis.[6] This is an effective strategy for arresting tumour cell proliferation. There are many classes of anticancer alkylating agents but bifunctional alkylating agents are believed to be the most cytotoxic.[7] These electrophilic species can form covalent lesions with DNA such as monoadducts (hydrolysis product), interstrand cross-links (ICLs), intrastrand cross-links (IaCLs) and DNA-protein cross-links (DPCs) (**Figure 1.4**).[8] After World War II, it was noted that the nitrogen mustards, created for use as a chemical weapon, could actually be used against lymphomas. These are less toxic than

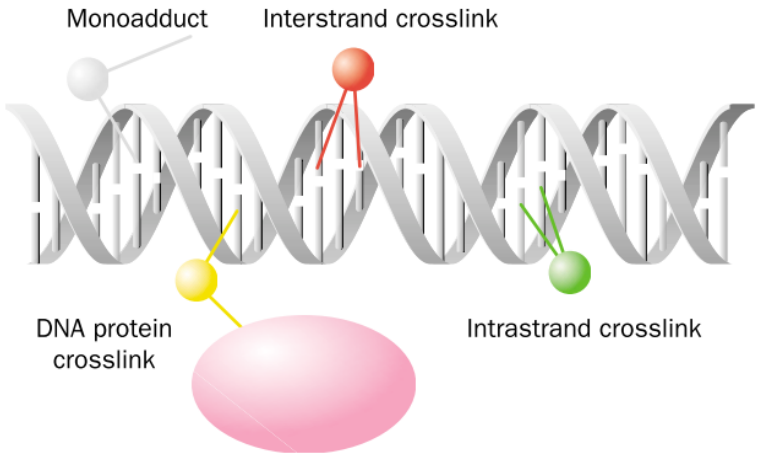


Figure 1.4 Possible outcomes upon treatment with bifunctional alkylating agents. Reproduced with permission from ref [8].

the World War I sulfur mustards and regression of the tumours was noted in back 1946 in patients treated with mechlorethamine.[9] This prototypical drug is still in use today and has spurred research into other ways to cross-link DNA. Another class of bifunctional cross-linking agents are alkyl sulfonates and alkyl sulfamates. In particular, busulfan and hepsulfam (**Figure 1.5**) have drawn interest due to their antitumour activity. Busulfan, the first such agent, was first shown to be effective against chronic myeloid leukemia in the early 1950s.[10]

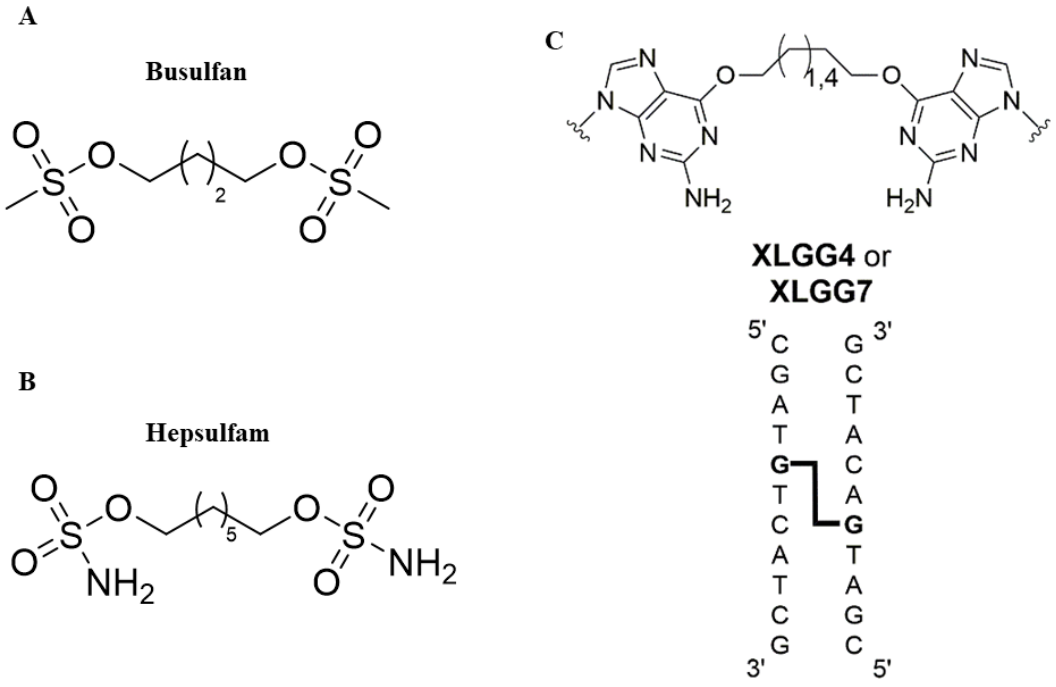


Figure 1.5 Bifunctional alkylating agents A) Busulfan and B) Hepsulfam. C) Butylene and heptylene intrastrand cross-links. Reproduced with permission from ref [36].

Hepsulfam, containing a seven carbon linker as opposed to a four carbon, showed even more promise due to its effectiveness against more tumour types in preclinical trials; presumably because they showed it was forming more cross-links.[11] Unfortunately, in clinical trials busulfan and hepsulfam showed similar effectiveness. The primary mode of cytotoxicity by these bifunctional alkylating agents is believed to be by acting on the N^7 atom of guanine residues. For example, it has been shown by mass spectrometry that DNA treated with busulfan resulted in 7-diguanidyl butane, as well as the aforementioned monoadduct 7-(hydroxylbutyl) guanine.[12] Whilst alkylene cross-links at the O^6 position has not been observed to date, it is highly plausible

that it occurs due to the fact that many monofunctional alkylating agents damage this position in DNA.

1.2.2 O^6-Alkylguanine DNA Alkyltransferases (AGTs)

Alkylation of the N^7 position of guanine while prominent is difficult to study due its instability. However, the O^6 position of guanine has been well researched due to its stability and toxicity. This damage is mutagenic as it leads to G:C→A:T mutations due to preferential insertion of thymine across from guanine by DNA polymerases.[13,14] Fortunately, the cell has developed mechanisms to remove these lesions through various pathways including direct repair. O^6-Alkylguanine DNA alkyltransferases are proteins found in most organisms whose role is removal of alkylation damage

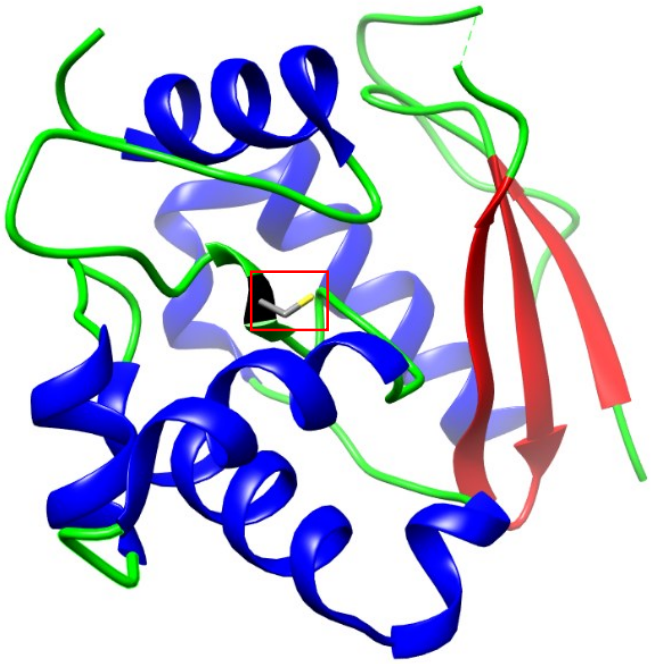


Figure 1.6 Crystal structure of hAGT. Nucleophilic sulfur is depicted in yellow. PDB:1QNT

at the O^6 position of guanine and the O^4 position of thymine. AGTs are small proteins (~21 kDa) whose topology is largely conserved throughout organisms. Human AGT (hAGT) binds DNA *via* a helix-turn-helix (HTH) motif and unusually this occurs on the minor groove of DNA (**Figure 1.7A**).[15] This sequence-independent binding has supported the hypothesis that hAGT scan DNA with no energetic cost close to the diffusion limit.[16] The mechanism involves a Glu-His-water-Cys catalytic triad (**Figure 1.7B**.), similar to the Asp-His-Ser observed in proteases,[17] which serves to activate the thiol residue. Due to a hydrogen bonding network, Cys145 (human

numbering) is essentially a thiolate anion. Cys145 is highly reactive and can thus remove the alkylation damage from guanine *via* an S_N2 mechanism. Important residues are Tyr114 (Phe in some species) which promotes the crucial nucleotide flipping step by probing the 3' phosphate either by steric or electrostatic interactions.[15] Once the damaged base is in the hAGT's active site, the DNA duplex is stabilized by an Arg128 finger which inserts across from the cytosine. After the damage removal has been completed the altered conformation of hAGT is recognized and the suicide protein is degraded by the ubiquitin pathway.[18,19] hAGT binds in a cooperative manner with approximately 1 protein/4 bp.[20]

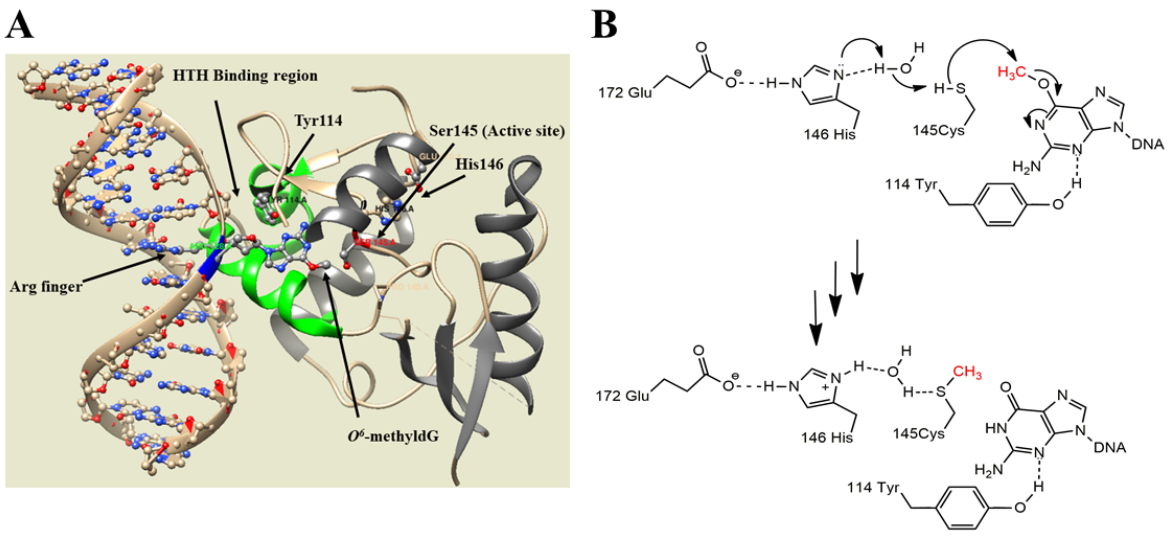


Figure 1.7 A) Crystal structure of hAGT bound to DNA. PDB:1T38 (Model prepared with UCSF Chimera^{TM}). B) hAGT's mechanism of repair.

Upregulation of hAGT has been implicated in chemotherapeutic resistance to alkylating agents. This is unsurprising as they act to cause lesions which promote apoptosis in cells as removal of alkylation would allow for the tumour cell to proliferate. Research has shown that levels of hAGT vary depending on the tissue and whether it is healthy *vs.* tumour.[21] For example, malignant lung cells contain approximately 4-fold hAGT activity. Liver contains the highest amounts of hAGT activity and brain or lung the lowest in healthy tissue. Furthermore, studies have shown that tumour cells, which often express high levels of hAGT, can show 4-10x more resistance against alkylating agents such a procarbazine.[22,23] Thus, hAGT's attractiveness as a drug target has spurred interest in developing substrates to inhibit its activity. O^6-benzylguanine has shown great potential for use in tandem with alkylating agents such as temozolomide due to it

permanently inactivating hAGT[24] however it had little benefit in clinical trials.[25] This effort was not fruitless however as research is ongoing into derivatives and other substrates.[26,27]

1.2.3 Substrate Scope of hAGT

While originally termed methylguanine DNA methyltransferase, it was noted in the early 1980s that hAGT could act upon O^6-ethyl adducts in tissues such as human and rat liver.[28,29] Shortly afterwards hAGT was shown to repair longer propyl and butyl adducts, with the bulkier *iso*-alkyl adducts being repaired at a slower rate.[30] Since then it has been shown to repair a plethora of adducts such as O^6-benzylguanine[31] 4-oxo-4-(3-pyridyl)butyl-guanine,[32] and cyclic 1,O^6-ethanoxanthosine (**Figure 1.8**).[33] The latter was found to form a DPC with hAGT and assisted in elucidating the crystal structure of hAGT. As can be seen by the alkylation damage caused by the alkylating agents MMS and MNU, the O^4 position of thymine is alkylated at a significantly lower frequency that the other nucleobase positions. However, this lesion is believed to be a more potent

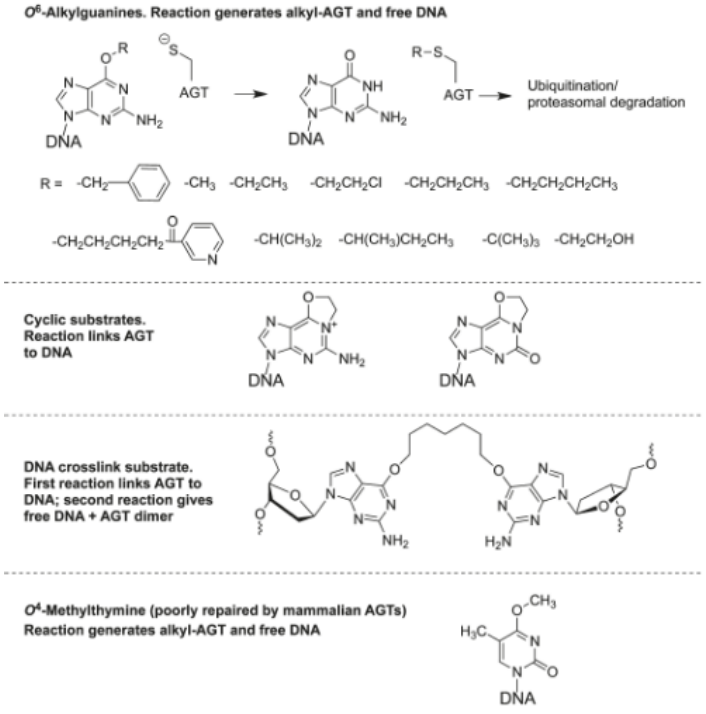


Figure 1.8 Some of the scope of substrates able to be acted on by AGTs. Reproduced with permission from ref [151].

inducer of mutagenesis (A:T→G:C transition mutations) as it is a poor substrate for hAGT and furthermore binding of the protein to DNA inhibits repair by the NER pathway.[34]

As mentioned, the bifunctional alkylating agents busulfan and hepsulfam are capable of cross-linking DNA, preferentially at the N^7 position of guanine. Possibly at the O^6 position guanine and to a lesser extent the O^4 position of thymine as witnessed with monofunctional alkylating agents. Oligonucleotide duplexes were synthesized with dG-O^6-alkylene-O^6-dG interstrand cross-links with either butylene and heptylene linkages and it was shown that the four carbon substrate was a poor substrate for hAGT while repair was witnessed for the longer seven carbon linker.[35] It should be noted that for the DNA to be restored, two repair events must occur with these cross-linked species: the first involves the formation of a DNA-hAGT cross-link (Median product) followed by a second hAGT protein that then acts upon the DPC at the α carbon of the cross-link at the second O^6 position of dG. In experiments done in Chinese hamster ovary cells deficient in hAGT there was more cytotoxicity experienced when treated with hepsulfam than when hAGT was expressed.[36] No differences were observed when the +hAGT and -hAGT cells were treated with busulfam which corroborates that hAGT cannot act upon the DNA lesion introduced by this agent in a cellular context. Various ICLs with butylene or heptylene tethers were investigated with alkylation at the O^4 position of thymine however all those examined were refractory to repair.[37,38]

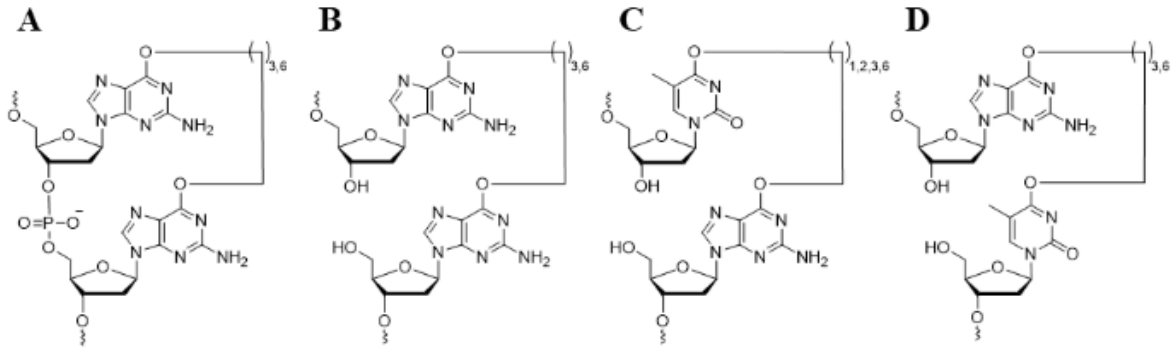


Figure 1.9 IaCLs with and without phosphate backbones a) dG-O^6-alkylene-O^6-dG b) dT-O^4-alkylene-O^6-dG c) dG-O^6-alkylene-O^4-dT and with a phosphate backbone d) dG-O^6-alkylene-O^6-dG.

When a dG-O^6-alkylene-O^6-dG was on the same DNA strand (**Figure 1.9A**), an IaCL, its repair was similar to the butylene and heptylene ICL analogues.[39] When a phosphate backbone was not present however, the IaCL substrates were virtually completely consumed at lower hAGT equivalents.[40] Oligonucleotides with all combinations of a butylene or alkylene cross-link at the

O^6 and O^4 position of guanine and thymine, respectively (5' *vs.* 3' variations*)* were synthesized lacking the phosphodiester backbone (**Figure 1.9B-D**) however in no instances was repair at the O^4 position of thymine observed.[41,42]

1.2.4 AGTs in different species

Thus far this chapter has been focusing on the human homologue hAGT however AGTs are a protein found in most species. The first AGT discovered was the adaptive response protein *Ada* from *E. coli* which the cell increases production of as the suicide protein is consumed due to a positive feedback loop.[43] It was determined that the C-terminal domain (Ada-C) of the protein was responsible for the activity on O^6-methylguanine.[44] It was reported afterwards another protein in *E. coli*, OGT, could also repair methylation at the O^6 position of guanine however it is expressed constitutively.[45] While the primary sequences often differ between species,[46] the overall fold remains similar between Ada-C and hAGT (among other homologues).[47] Key conserved residues are highlighted in **Figure 1.10** with residues in the minor groove binding recognition helix and active site residues PCHR, clearly important from an evolutionary standpoint. Wide disparities in repair are observed between the AGTs found in *E. coli* and human AGT. hAGT is much more adept at repairing bulkier lesions such as O^6-benzylguanine, IaCLs and ICLs meanwhile Ada-C is incapable of repairing any larger lesions and OGT repairs O^6-benzylguanine poorly[48,49] with no

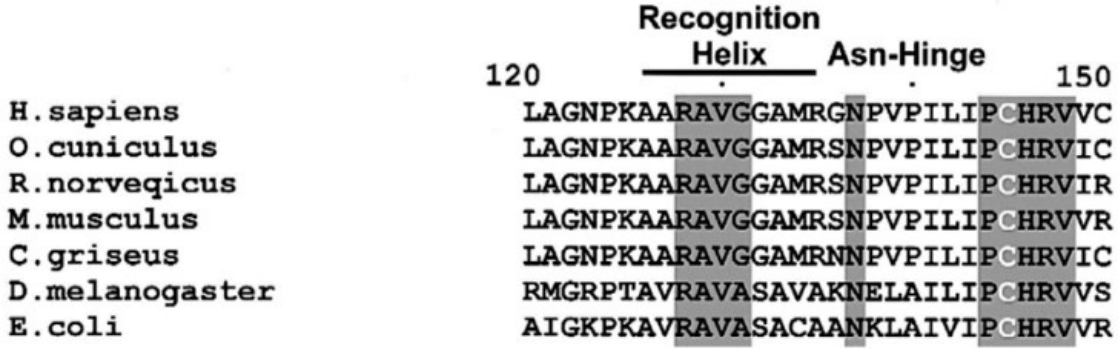


Figure 1.10 Conserved amino acid residues across species. Reproduced with permission from ref [46].

repair of alkylene cross-links observed. This has been largely attributed to hAGT's Pro140 residue (absent in Ada-C and OGT) widening its active side to accommodate the larger lesions.[31] Ada-C has a very compact active site which would explain its inability to repair any lesions larger than a

methyl group and unfortunately there are no known crystal structures of OGT to date. It is proposed due to Ada-C's buried active site a conformational change is likely to be required to instigate the repair event. In contrast to hAGT, the *E. coli* proteins are exceptional at removing damage at the O^4 position of thymine[50] whereas as stated above the human homologue struggles to remove the methyl adduct.[34] Through site-directed mutagenesis, Fang *et al.* engineered the active site of hAGT to create various mutants to gain mechanistic information as to why OGT can repair O^4-methyl dT but not hAGT. Their top mutant, hAGT-03, had 8 active site substitutions of OGT and they hypothesized (through modeling) that the mutations allow for a more flexible loop residue (containing Cys145). This would be due to glycine mutations and the the protein is able to orient properly for several small chain alkyl transfers. Work by McManus *et al.* from our lab engineered an hAGT protein with 20 active site residue (except Pro140) mutated to those of OGT and it demonstrated an amalgam of the two species' AGTs repair abilities.[51] Remarkably, it was able to act on O^4-methyl dT, O^4-hydroxyalkylene dT as well as various ICLs.

1.2.5 Applications of hAGT in Nanotechnology

In addition to its important role in biology, AGT has found use a tool in nanotechnology. Detecting protein expression and location of proteins is crucial to understanding their function. Classic techniques such fluorescent proteins have several limitations such as low brightness and stability.[52] Keppler *et al.* exploited the promiscuousness of hAGT and synthesized fluorescent guanine derivatives tethered by alkyl moieties at the O^6 position which were shown to be substrates to G160W hAGT (**Figure 1.11A**).[53] They showed that the fusion of multiple proteins with G160W hAGT allowed monitoring of their activity with fluorescent probes in living cells. Of the same ilk, through clever protein engineering of Gauthier *et al.* involved the design of an hAGT mutant which was able to act on the O^2 position of fluorescently benzylated cytosine derivatives. Because their fusion-protein "CLIP-Tag" operates on a different substrate, they showed they could monitor two different proteins simultaneously in a cell. These versatile "SNAP"® and "CLIP"® tags for protein detection have had commercial success and are available on the market.

Fabrega have developed several probes for hAGT detection. One approach involves the thrombin binding aptamer which cannot bind its substrate when the O^6 position is methylated (**Figure 1.11B**).[54] The aptamer (a nucleic acid which adopts a conformation which recognizes and binds a target with high affinity with applications such as biosensing and delivery) was

tethered to a sheet of DNA origami. They were able to detect by atomic force microscopy (AFM) that after the aptamer had been incubated with hAGT, an increase in height was observed. This height increase correlated well to thrombin binding because the methyl damage had been repaired

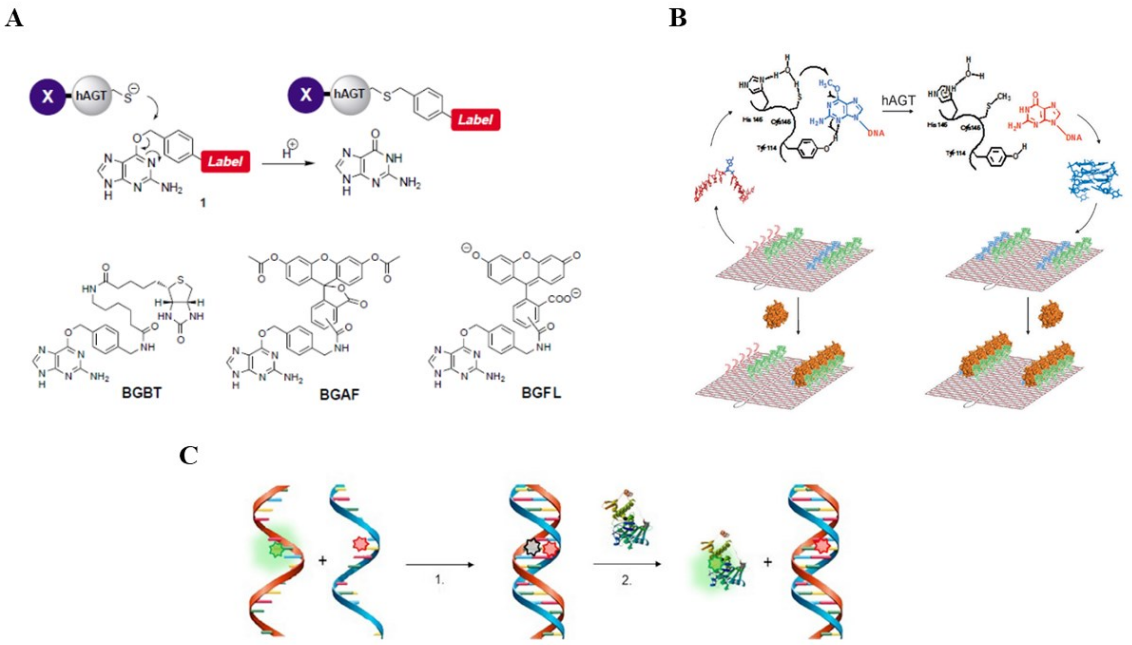


Figure 1.11 Some applications of hAGT in nanotechnology. A) SNAP-tag fusion protein and fluorescent O^6 benzyl guanine probe allows detection of fused protein *in vivo*. Protein is depicted by X and BGBY, BGAF and BGFL are fluorescent probes. Reproduced with permission from ref [53] B) O^6 methyl guanine renders thrombin binding aptamer inactive however upon hAGT removing methyl lesion thrombin binding aptamer can bind its substrate. Reproduced with permission from ref [54] C) Fluorescent O^6 benzyl guanine derivative's (depicted in green) fluorescence is restored upon hAGT repair event. Reproduced with permission from ref [55].

allowing the thrombin binding aptamer to fold properly and bind its substrate. Another hAGT probe was designed by having a quencher and a fluorophore in duplex DNA (**Figure 1.11C**).[55] The fluorescent O^6 benzylguanine probe (similar to the SNAP tag concept) was a substrate for hAGT and utilized as a means for measuring repair activity of dsDNA.

1.3 DNA as a Nanomaterial

1.3.1 DNA Nanostructures

While the molecular biology role of DNA had been well established, Dr. Nadrian Seeman noted that due to the predicable self-assembly properties of DNA, a wide variety of shapes could be constructed based on simple Watson-Crick base pairing rules. The development of automated solid-phase DNA synthesis fueled the rise of this field as DNA could be synthesized easily in a

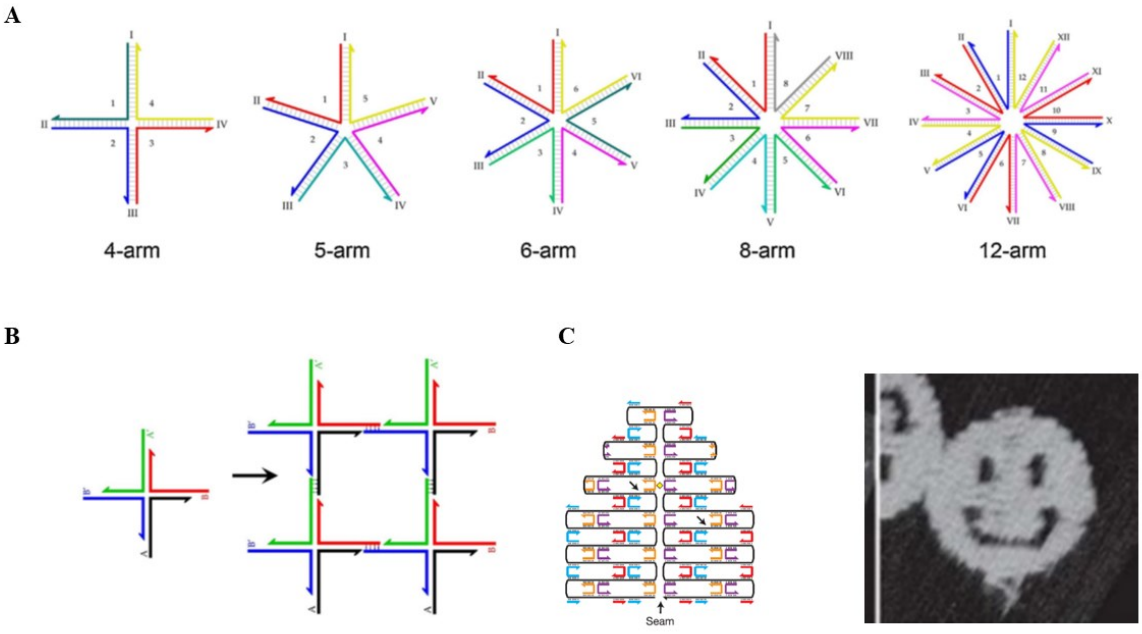


Figure 1.12 Two dimensional architectures constructed with DNA. A) Multiarmed junctions. Reproduced with permission from ref [58] B) Use of sticky ends to expand in two dimensions. Reproduced with permission from ref [59] C) DNA origami to staple large DNA strands into well defined shapes. Structure visualized by AFM. Reproduced with permission from ref [60].

cost-effective manner. Furthermore, the tunability of DNA is well specified as the length of the DNA helical turn is very well defined thus size is easily controlled. In 1983, Seeman and coworkers constructed a Holliday junction constructed with four oligonucleotides as a proof of concept (**Figure 1.12A**).[56] Seeman then continued based on these basic base-pairing principles to build junctions with five, six, eight and 12 arms.[57,58] The use of "sticky ends", DNA duplexes with

unpaired single stranded overhangs allowed the construction of tile-like networks as illustrated in **Figure 1.12B**. This work has spurred various complex patterns to be woven using sticky ends and multi-armed junctions.[59] DNA origami, for example, involves the usage of many staple strands

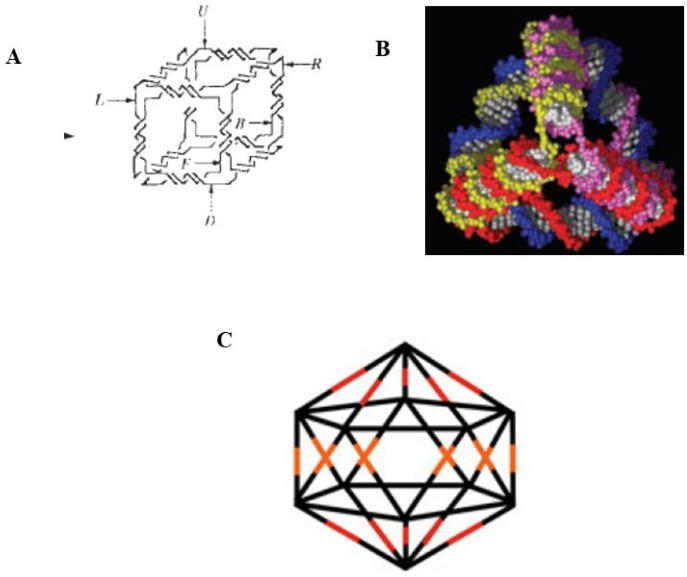


Figure 1.13 Prototypical three-dimensional structures constructed with DNA. A) The first shape constructed was a cube. Reproduced with permission from ref [61] B) DNA tetrahedra are one of the most studied 3D shapes. Reproduced with permission from ref [62] C) DNA icosahedra. Reproduced with permission from ref [236].

to assemble a large DNA sequence into shapes from simple geometries to more intricate designs such as a smiley face (**Figure 1.12C**).[60]

This work was not limited to 2D structures with the first three-dimensional shape being a DNA cube designed by Seeman which could be assembled using the same rules as above however complexities arise due to steps of ligation and purification (**Figure 1.13A**).[61] In 2005, Turberfield was able to build a novel polyhedron from DNA and is credited with constructing the first DNA tetrahedron (**Figure 1.13B**).[62] Since then, many other polyhedral structures have been assembled with DNA such as octahedra,[63] icosahedra[64] and buckyballs[65] to name a few. The assembly of these structures can be more complex than the tile structures and can require purification, ligation and carefully controlled annealing. Nonetheless, research labs have cunningly been able to synthesize a large range of geometries which have included complex wireframes[66] and DNA suitcases.[67]

1.3.2 DNA Nanoswitches

The polymorphism of DNA will be discussed more in section 1.4 however due to the fact that DNA is not limited to a standard "B-form" double helix (the more canonical duplex), groups have looked at methods to induce conformational changes upon external stimuli. Nucleic acid structures are highly pH dependent due to the presence of several basic nitrogen atoms on each nucleobases. For example, DNA triplexes can be formed when tracts of nucleotides containing polypurine and polypyrimidine bases are present. The C+:G:C motif has a pH dependence and thus does not form at neutral pH where as the T:A:T triplex disassembles in the higher pH range. Idili *et al.* utilized this phenomenon to engineer two unimolecular triplex pH sensors: one could register changes in pH in the low pH region and the other sensor in the high (**Figure 1.14A**).[68] A pH-dependent reversible oscillator was constructed with a tethered *i*-motif (discussed in more detail in section 1.4) and demonstrated excellent reversibility (**Figure 1.14B**).[69] Other factors that can influence conformational changes in nucleic acid structures explored include ions[70,71] and the addition of DNA strands.[72]

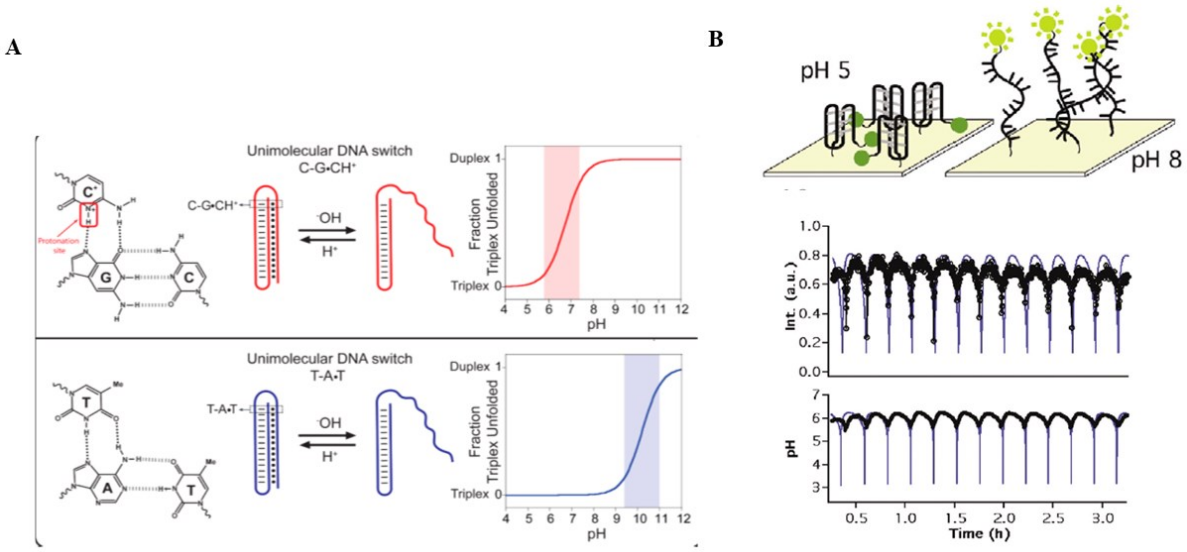


Figure 1.14 pH sensitive nanodevices A) DNA triplex nanosensor. Reproduced with permission from ref [68] B) *i*-Motif reversible oscillator. Reproduced with permission from ref [69].

1.3.3 Applications of DNA Nanostructures

DNA nanostructures have found applications due to their cellular compatibility and also their ability to undergo conformational changes upon external stimuli. The DNA nanocages contain an open cavity and their ability to house guest molecules has been investigated. Turberfield eloquently showed that the tetrahedron could house cytochrome c and internalization was strongly dependent on the position of the DNA-protein tether.[73] Gold nanoparticles have been shown to be quite compatible to being housed by DNA nanocages particularly due to thiol moieties affinity

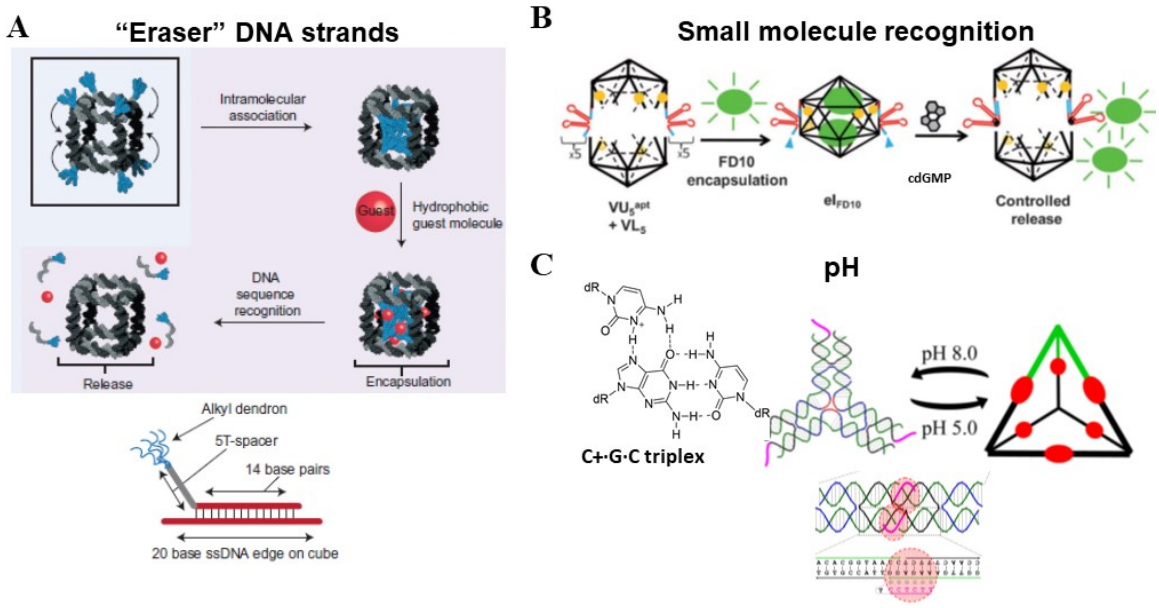


Figure 1.15 Methods of dismantling DNA nanostructures and cargo release. A) DNA eraser strands fully complementary to toeholds on the edges allow for hydrophobic guest molecule release. Reproduced with permission from ref [75] B) Aptamer at the interface undergoes conformational change on binding of target molecules releasing FD10 dye. Reproduced with permission from ref [76] C) pH sensitive triplex forming oligonucleotide disassemble a tetrahedron into constituent oligonucleotides upon raising the pH. Reproduced with permission from ref [77].

to them.[74] One of the driving forces for this research is to deliver molecules to a cell for therapeutic applications, however one of the issues with these structures is they are too structurally stable. Furthermore, hydrophobic drug molecules present challenges due to the relatively large core of DNA nanostructures (the molecule would exit by passive diffusion) however groups such as the Sleiman lab have circumvented this problem.[75] They designed a DNA cube with dendritic

DNA (D-DNA), which had hydrophobic alkyl chains tethered to the DNA strands on the edges of the cube (**Figure 1.15A**). The alkyl chains would then assemble in the hydrophobic core of the cube which could host hydrophobic guest molecules. The D-DNA however was not fully complementary to the edges of the cube, and thus when the fully complementary "eraser strand" was added, the D-DNA was displaced and the hydrophobic cargo was released. Another strategy was employed by the Krishnan group who designed an icosahedra to contain an aptamer forming oligonucleotide at the interface of its two constituent halves.[76] This aptamer when bound to cyclic di-GMP (cdGMP) undergoes a conformational change and then the icosahedra is no longer capable of remaining intact which was monitored by the release of a dye molecule (**Figure 1.15B**). Liu *et al.* designed a DNA tetrahedron with triplex forming sequences at the interface which would anneal at the interfaces of the tetrahedron at low pH (**Figure 1.15C**).[77] The protonation of the third strand's cytosine allows it to hydrogen bond with the Hoogsteen face of guanine. This pH dependence allows the DNA tetrahedron to disassemble upon experiencing basic conditions. Another strategy for nanostructure disassembly is the use of light.[78]

DNA nanostructures have found another application in biosensing. In particular DNA tetrahedra (TDNs) due to their ability to enter cells without transfection agents and their serum stability.[79] Fluorescence resonance energy transfer (FRET) is a technique commonly employed in cells for imaging. It operates on the principle that when a fluorescent moiety is excited an energy transfer can occur to a molecule in close proximity and the consequence can be of an emission wavelength shift or quenching. The Tan group designed a DNA tetrahedron whose fluorescent moieties were tethered by DNA sequences and were off in the resting conformation (**Figure 1.16A**).[80] These two strands were each complementary to an mRNA (TK1) associated with tumour proliferation. It was designed as such that when hybridized to their complementary strand the fluorophores are forced into close proximity and FRET is observed. In another example, DNA tetrahedra were used to detect the DNA repair protein APE1. This protein is involved in the base excision repair pathway and its misregulation has been correlated with several diseases including cancer.[81] dsDNA containing an apyrimidic/apurinic site, that can undergo cleavage by APE1, was tethered to the DNA tetrahedron (**Figure 1.16B**).[82] The dsDNA was designed with a fluorophore and a quencher thus no fluorescence was observed except when cut by APE1. The DNA tetrahedron was even able to detect high levels of cytoplasmic APE1 in several cancer cell lines but not in healthy cells. They used their FRET probe in a related DNA tetrahedron as a means of

inhibiting, as well as monitoring, APE1. Other devices examined can map the internal pH of cells such as an "*i*-motif nanomachine"[83] and the aforementioned icosahedron.[84]

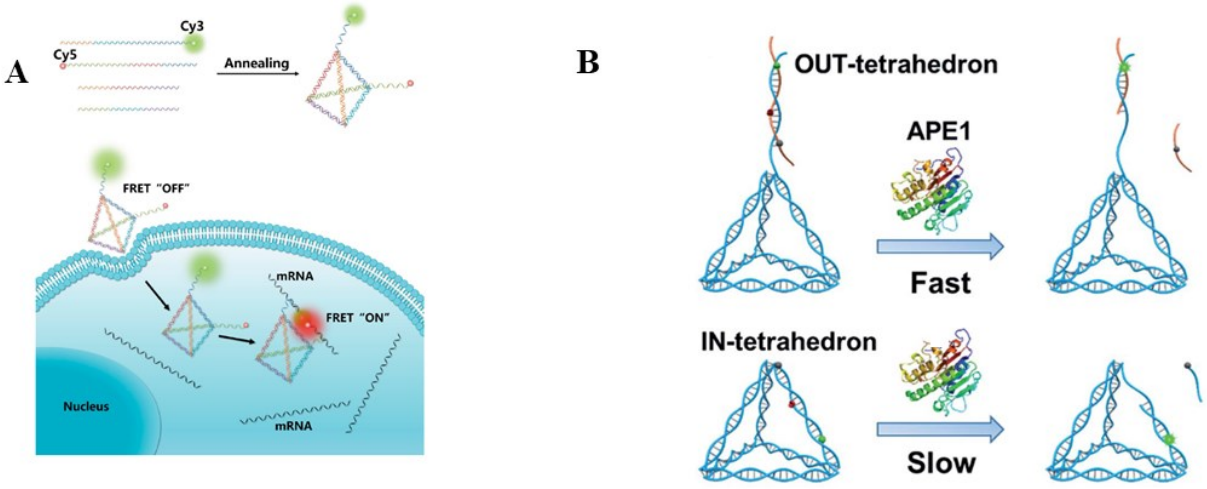


Figure 1.16 Biosensing applications of DNA nanostructures A) mRNA associated with cancer can be detected with FRET-functionalized TDN. Reproduced with permission from ref [80] B) TDNs are capable of both inhibiting and detecting the DNA repair protein APE1. Reproduced with permission from ref [82] C) Icosahedra containing dye can monitor endosome maturing *in vivo*. Dye is depicted in yellow. Reproduced with permission from ref [84].

1.3.3 Recent Advances of DNA Tetrahedra for Therapeutic Applications

The ability of DNA tetrahedra (TDNs) to enter cells without chemical modification in conjunction with their ease of synthesis, serum stability and biocompatibility provides an excellent route for the delivery of therapeutic agents. This active area of investigation comprises numerous creative approaches, with some recent examples discussed below.

1.3.3.1 Small molecules

The targeted delivery of therapeutic agents to the mitochondria is attractive due to its ability to induce apoptosis. The peptide D-$(KLAKLAK)_2$ possesses cationic character and an α-helical geometry which effectively disrupts the negatively charged mitochondrial membrane and can be internalized.[85] Yan and co-workers reported employing "click-chemistry" to conjugate three $(KLAKLAK)_2$ peptides to a 55-mer ssDNA to form a DNA tetrahedron (3KLA-TDN) and monitored cellular internalization, organelle targeting and subsequent cell death in 4T1 cells

(**Figure 1.17A**).[86] As a dual-functional therapeutic TDN they successfully loaded ~390 doxorubicin (DOX) molecules per DNA tetrahedron which was ~75 % more efficient than other delivery systems such as polymer nanoparticles. Confocal laser scanning microscopy (CLM) revealed that free DOX was located in the nucleus, but large quantities were also located in the mitochondria when present on 3KLA-TDNs which was corroborated by flow cytometry on mitochondrial extracts. Lastly, activation of the apoptotic pathway was monitored by cytochrome *c* release. Cytochrome *c* signals were the highest with DNA tetrahedra loaded with three $(KLAKLAK)_2$ peptides relative to DNA tetrahedra with reduced DOX loading and controls, and unsurprisingly showed the highest cytotoxicity in several cell lines.

Bacteria have shown to have effective defences against antimicrobial peptides (AMPs); a class of antibiotics that typically disrupt bacterial membranes. While AMPs have been successful, problems inherently occur in the presence of proteases. Recently Liu and co-workers demonstrated that a peptide (GL13K) conjugated DNA tetrahedra caused bacterial membrane disruption.[87] This was particularly of interest as they tested their peptide-DNA tetrahedra conjugate against *Porphyromonas gingivalis*, which is known to degrade free GL13K. These examples emphasize the potential of DNA tetrahedra to navigate around classical issues in therapeutic delivery.

1.3.3.2 Antisense and siRNA

It has long been known that short strands of DNA, known as antisense oligonucleotides (ASOs), can be employed for the targeted cleavage of mRNA by RNase H-mediated degradation.[88] More recently, it was discovered that short double-stranded RNA (small interfering (si) RNA) could induce gene knockdown through incorporation into the RNA-induced silencing complex.[89] Despite initial promise, few drugs are on the market due to hurdles such as poor cell permeation, degradation by nucleases and off-target effects. Chemical modification and conjugation have been instrumental in overcoming the aforementioned issues and delivery with DNA tetrahedra have also been explored. In one pivotal experiment, the Anderson group showed excellent gene knockdown in tumour cells in live mice when the siRNA was hybridized to a DNA tetrahedron.[90] Despite this ground-breaking study, there has been a relative hiatus in DNA tetrahedra as a vehicle for siRNA delivery and the field is still in its infancy. In 2018, the Ding group showed that a self-assembled "double-bundle" DNA tetrahedron efficiently delivered ASOs (**Figure 1.17B**).[91] The term "double-bundle" refers to an engineered DNA tetrahedron with increased sites for modification

and drug loading. They also were able to orient the ASOs via this strategy into the core of DNA tetrahedra for protection from nucleases. In their work they showed a successful knockdown of *c-raf*, a gene involved in cell proliferation, in lung cells. Recently, they have employed these double-bundled DNA tetrahedras in a synergistic approach by simultaneously knocking down a key protein involved in mitochondrial membrane potential and carrying KillerRed protein to promote apoptosis in tumour cells.[92] Damage to the mitochondrial membrane is an effective way of

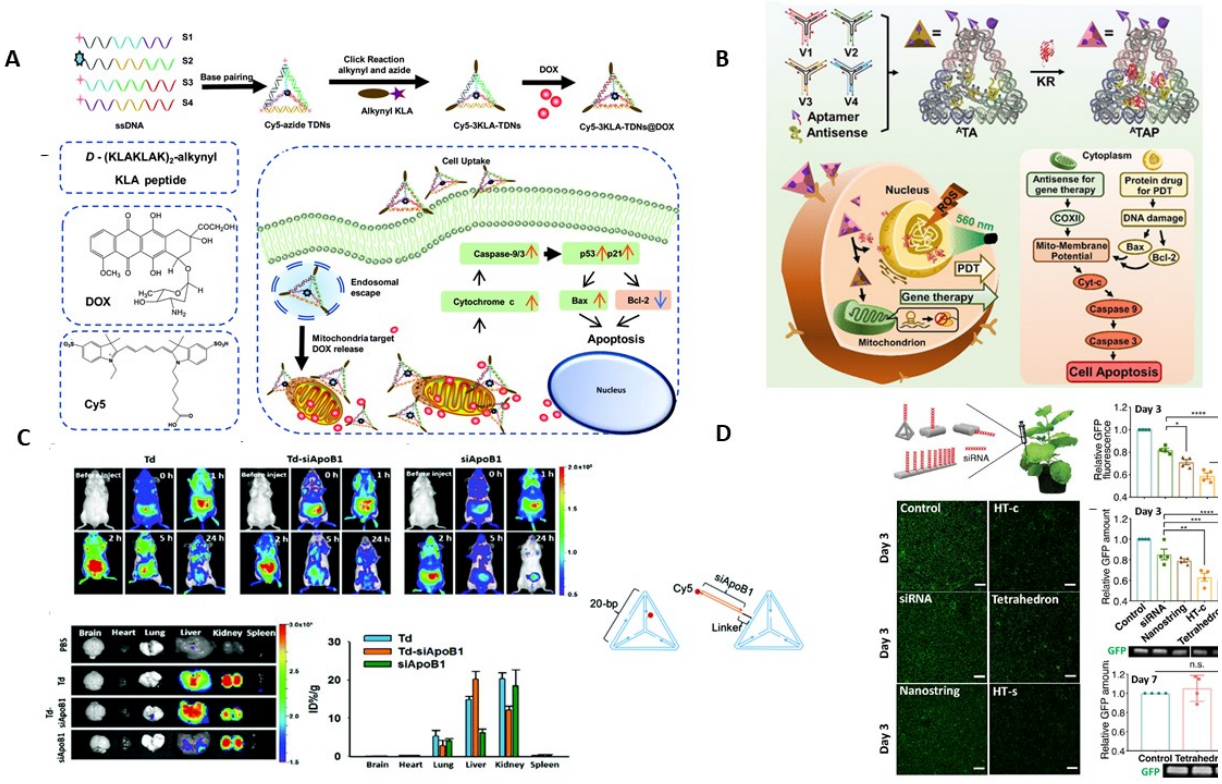


Figure 1.17 TDN strategies for delivering small molecules and oligonucleotide therapeutics. **A**) Peptide-DOX DNA tetrahedron disrupts mitochondrial cell membrane and triggers apoptosis. Reproduced with permission from ref [86] **B**) Three-pronged use of a DNA tetrahedron with aptamer-targeting, oncogene knockdown and small molecule drug delivery. Reproduced with permission from ref [92] **C**) siRNA-DNA tetrahedron conjugates accumulating in liver of live mice increase mouse survival. Reproduced with permission from ref [95] **D**) GFP levels in plant cells are most efficiently diminished when siRNA is tethered to a DNA tetrahedron. Reproduced with permission from ref [96].

promoting apoptosis.[93] In addition, due to the overexpression of nucleolin on the surface of cancer cells,[94] nucleolin binding aptamer AS1411 was tethered to the modified TDNs for targeted delivery.

One concern of oligonucleotide-based therapeutics is that upon intravenous injection, they may be degraded with the products sent to the kidneys to be excreted. The Ahn group constructed a DNA tetrahedron decorated with siRNA that exhibited liver-targeted reduction of ApoB1 mRNA in living mice (**Figure 1.17C**).[95] Higher accumulation of the intact DNA tetrahedron in the liver was achieved by intraperitoneal injection. Overexpression of ApoB1 is a factor in hypercholesterolemia, thus it is an attractive drug target. They demonstrated effective gene silencing in real-time by western blot analysis. Furthermore, they measured serum lipid (HDL and LDL cholesterol and triglycerides) and demonstrated that a 20-30 % decrease only occurred with the DNA tetrahedron-conjugated siRNA. This success in live animals should lead to a plethora of studies using tetrahedral geometries for the delivery of oligonucleotide therapeutics.

Synthetic oligonucleotide mediated gene silencing in plants is a promising tool in agricultural technology but is hindered primarily by cellular permeability. Recently, Zhang *et al*. assembled an siRNA loaded DNA tetrahedron capable of infiltrating mature plant cells.[96] While TDNs are often defined generally, there are important factors to consider when attempting to internalize them in these cells. Size of the DNA tetrahedra was quite important with a decline in cellular uptake observed between a diameter of 8.8 and 12.6 nm suggesting a cut-off of ~10 nm. Other important factors contributing to cellular intake were compactness and flexibility. Investigation of the location of the siRNA on the DNA tetrahedron structure provided design and mechanistic insights. The optimized DNA tetrahedra were tethered to siRNA targeting green fluorescent protein (GFP) and introduced into *Nicotiana benthamiana* leaves. **Figure 1.17D** depicts efficient gene silencing as monitored by a decrease in fluorescence and western blots were correlated with cellular uptake. As observed in mammalian cells, the DNA tetrahedron-conjugated siRNA are more effective than the naked siRNA controls. This study has promising implications in the structural design of gene-silencing DNA tetrahedra capable of infiltrating plant cells for purposes in plant bioengineering.

Interest has been generated in using ASOs for antibiotic-resistant bacteria; however, cellular entry is difficult for naked oligonucleotides due to their relatively high molecular weights and the bacterial cell wall.[97] Recently, Hu *et al*. looked at ASO-mediated gene silencing in several bacterial species targeting GFP and the acyl carrier protein ACP mRNA.[98] Despite initial promise through CLM, flow cytometry and DNase digests, it was shown that the DNA tetrahedron were simply bound to the bacterial membrane. Different concentrations of lipofectamine 2000 (LP2000)

were explored to increase cellular uptake and improve stability to DNase digestion. Rapid cellular uptake was observed in an energy-independent manner (less than 10 minutes) for various strains of bacteria when LP2000 was present for the DNA tetrahedron, but not for single-stranded DNA (**Figure 1.18A**). Lastly, phosphorothioate (PS) conjugated ASOs showed successful knockdown of GFP expression as well as *acpP* as determined by bacterial growth. This work with therapeutically relevant bacteria such as multidrug-resistant *Pseudomonas aeruginosa*, shows promise in the development of DNA-based antibiotics. The necessity of the use of a transfection agent emphasizes the challenges involved in entering bacterial cells.

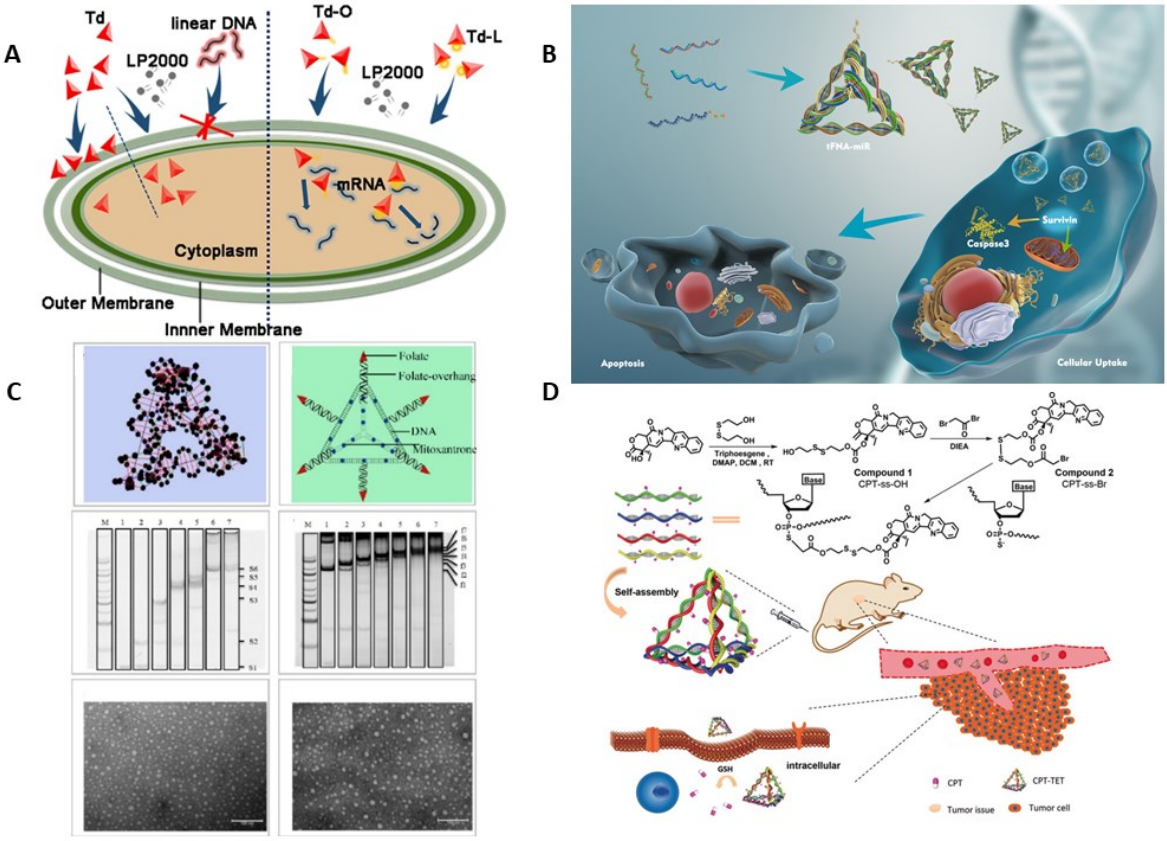


Figure 1.18 Bacterial antisense delivery and targeted therapeutics utilizing DNA tetrahedra. **A**) Linkage to DNA tetrahedra allows successful entry through bacterial cell walls. Reproduced with permission from ref [98] **B**) DNA tetrahedra-miRNA conjugates result in cell apoptosis by targeting survivin. Reproduced with permission from ref [100] **C**) Folate overhangs on DNA tetrahedra specifically target and kill leukemia cells. Reproduced with permission from ref [106] **D**) Reducible disulfide bridges on the DNA backbone target mouse tumours with generic hydrophobic drug. Reproduced with permission from ref [107].

MicroRNAs (miRNAs) are abundant in the cell and play a role in many regulatory processes and diseases due to their interaction with mRNA.[99] The instability of unmodified miRNA, however, has limited its use as a therapeutic agent. Exciting work done by the Cai lab demonstrated increased miRNA stability and was the first example of the application through TDN conjugation (**Figure 1.18B**).[100] Survivin is an inhibitor protein of apoptosis thus its overexpression is implicated in tumour cell proliferation. Furthermore, its low concentrations in healthy cells renders it a suitable candidate for cancer-targeting therapies. MiRNA-214-3p is capable of binding and inactivating survivin mRNA, therefore it has vast potential in killing tumour cells.[101] After demonstrating 5'-MiRNA-214-oligonucleotides could form a DNA tetrahedron, it was shown to enter cells by CLM after 8 h. Promisingly, the DNA tetrahedron stabilized miRNA remained entirely intact after 24 h in fetal bovine serum (FBS). Due to the ability of it to enter A549, suppress survivin mRNA and enable apoptosis, they envision tethering an abundance of therapeutically relevant miRNAs to their DNA tetrahedron.

CpG oligonucleotides (ODNs) are of interest due to their immunostimulatory properties via the activation of Toll-like receptor 9.[102] The Fan group found that several cytokine levels were highest when four CpG ODNs were tethered to a DNA tetrahedron, however even for DNA tetrahedra modified with one CpG strand, significant potency was observed relative to the single-stranded CpG ODNs.[103] A follow up study by Ohtsuki *et al* showed that CpG DNA tetrahedron induced large amounts of tumour necrosis factor-α in RAW264.7 cells.[104] These studies demonstrate another application of the DNA tetrahedron as opposed to traditional methods of combating antigens.

1.3.3.3 Anticancer therapeutics

Drugs that target tumour cells are of obvious interest to deter the extremely damaging and fatal events that occur due to off-target effects. For example, folate ligand conjugation is common due to the folate receptor being overexpressed on the surface of numerous cancer cells and is therefore used for targeted tumour delivery.[105] Mitoxantrone (MTX), similar to DOX, is a powerful DNA intercalating agent. However, when administered, its off-target effects are severe due to substantial biodistribution. This problem was addressed by Bu *et al.* who combined the targeting effects of folate and the anticancer properties of mitoxantrone with the excellent nuclease stability and cellular internalization of DNA tetrahedra[106] Interestingly, their TDN was assembled by 6

oligonucleotides with folate overhangs producing a mixture of products (0-6 folate residues). They focused on the completely folated 25 nm DNA tetrahedron (fTDN) and demonstrated excellent MTX loading with minimal leaking (**Figure 1.18C**). The MTX loaded fTDN was readily taken up by target leukemic cells and increased apoptosis was observed relative to DNA tetrahedra lacking folates in multiple cell lines. Moreover, diminished tumour growth and increased survival rates were observed in live mice containing xenografted tumours with minimal safety concerns. This folate-modified DNA tetrahedron approach could be easily extended to other drug molecules to improve cellular targeting and uptake.

While DNA nanostructures such as DNA tetrahedra are an excellent guest system for hydrophobic drugs, the Zhang lab recognized that there remains challenges for therapeutics with different properties and little control over guest molecule loading.[107] To address this, they synthesized oligonucleotides for DNA tetrahedron formation with PS moieties at well-defined positions in the structure. Following, carbonethylbromide-modified camptothecin (CPT), a chemotherapeutic agent, was conjugated to the PS DNA backbone through redox sensitive disulfide bonds (**Figure 1.18D**). Their yields were characterized to be 100 % by UV/Vis spectroscopy, mass spectrometry and gel electrophoresis with lower quantities of PS sites (<6).

After analysis of the CPT-DNA in reducing conditions, it was observed that release of the drug was accomplished with redox conditions similar to cancer cells, but not with physiological glutathione (GSH) concentrations. Lastly, they labelled their construct with a Cy5.5 fluorescent dye and showed it was capable of targeting mouse tumours. As shown in their in vitro experiments, tumour growth is greatly delayed relative to the controls including single-stranded CPT-grafted DNA. As in previous examples, tumour targeting is particularly important with camptothecin, an agent known to cause liver and kidney damage. In this case, the mice suffered minimal damage to their major organs as determined by histology and immunohistochemical analysis.

1.3.3.4 Protein delivery

There is great interest in the delivery of proteins in medicine to tune cellular enzymatic activity, however various factors inhibit their entry into cells. Laboratories including the Mirkin group have exploited DNA structures to allow functional enzymes to permeate into cells.[108] Interestingly, while the first molecule encapsulated was the protein cytochrome *c* by Turberfield in 2006,[73] TDNs have not been exploited as protein delivery systems until recently. Li *et al.* tethered the

RNase A protein to a DNA tetrahedron via reducible disulfide bridges (termed asdTDF-RNase A).[109] As performed by Turberfield, they engineered internalization of the protein by specifying the position of the disulfide tether. Additionally, internalization of RNase A was confirmed by protection against degradation by proteinase K, emphasizing the DNA tetrahedron's ability to shield the protein cargo. The DNA tetrahedron was further modified with two sgc8 tumour-targeting aptamers which bind protein tyrosine kinase-7 expressed on HeLa cells. The asdTDF-RNase A nanostructures were shown to internalize well, and the cancer cells' reducing environment allowed release of the RNase A guest. Lastly, their cell viability studies showed that degradation of the target RNA by RNase A promoted apoptosis which highlights the exciting potential for DNA tetrahedra as a means for protein therapy.

1.3.3.5 Outlook for DNA tetrahedra

The DNA tetrahedron is an inspired structure that continues to accelerate important discoveries and applications related to health including the delivery of therapeutic agents and detection of species within the cell. Its straightforward assembly, nuclease stability and susceptibility for modification to introduce various functionalities will continue to inspire and enable new platforms for nanotechnology. Despite excellent preliminary potential, this literature survey reveals two major challenges to resolve, inhibiting the use of DNA tetrahedra in a clinical setting: (1) serum stability and (2) cellular uptake. In all cases, DNA tetrahedra show increased nuclease resistance when compared to single stranded controls which has been suggested to be attributed to DNase's inability to form the required protein-DNA intermediate complexes due to the rigidity of the DNA tetrahedron structures.[110] However, one recent literature review has revealed a lack of experimental data regarding serum stability of DNA tetrahedron, and for those that do contain such data, large variations in DNA tetrahedron stability in biologically relevant media such as FBS are reported.[111] Moreover, efficient cellular uptake of unmodified nanostructures remains unclear as common methods for visualization have been shown to produce misleading results if proper controls are not performed.[111,112] Thus, it is difficult to draw concurrent conclusions based on experimental variations found within the literature and an accurate general set of control experiments are required. Nonetheless, several strategies such as oligomer conjugation,[79,113] chemical modification,[113,114] and enzymatic ligation of nicks[115] have been shown to increase serum stability. In addition, there is an inherent lack of high-resolution structural information

regarding the DNA tetrahedron. To date there are no known crystal structures of DNA tetrahedra. However cryo-EM has shown promise despite few recent advancements.[116] Future efforts in elucidating the structure of the TDN could provide vital information in the development of ideal DNA tetrahedron based therapeutics and sensors.

While DNA tetrahedra have not been established on the market to date, their biocompatibility could offer promising results in humans. Their versatility is emphasized by overcoming hurdles in other eukaryotes such as plants, and also in bacteria. Future research efforts should focus on alternative protection strategies towards increased serum stability and prolonged blood circulation for the development of DNA tetrahedra for diagnostic and therapeutic applications in a clinical setting.

1.4 G-Quadruplexes

1.4.1 Polymorphism of nucleic acids

Nucleic acids are not limited to the canonical double helix. Functional RNA in a biological context generally contains a wide variety of secondary structures such as hairpins, bulges and pseudoknots.[117] While B-DNA is the dominant double helix observed, a variety of other double helices such as A-DNA and Z-DNA also exist. As mentioned before the pH dependence of DNA triplexes can be used as a nanoswitch (Surface model and T:A:T base-pairing depicted in **Figure 1.19A**). DNA triplexes have also been implicated in various biological functions such as gene regulation.[118] Another well studied structure is the *i*-motif which is formed in cytosine rich DNA strands at low pH (**Figure 1.19B**). In 1993 Gehring *et al.* showed by NMR that a tetrameric structure was forming stabilized by hemiprotonated C^+:C base-pairs composed of two duplexes intercalating.[119] G-quadruplexes are one of the most studied alternate structures of DNA due to

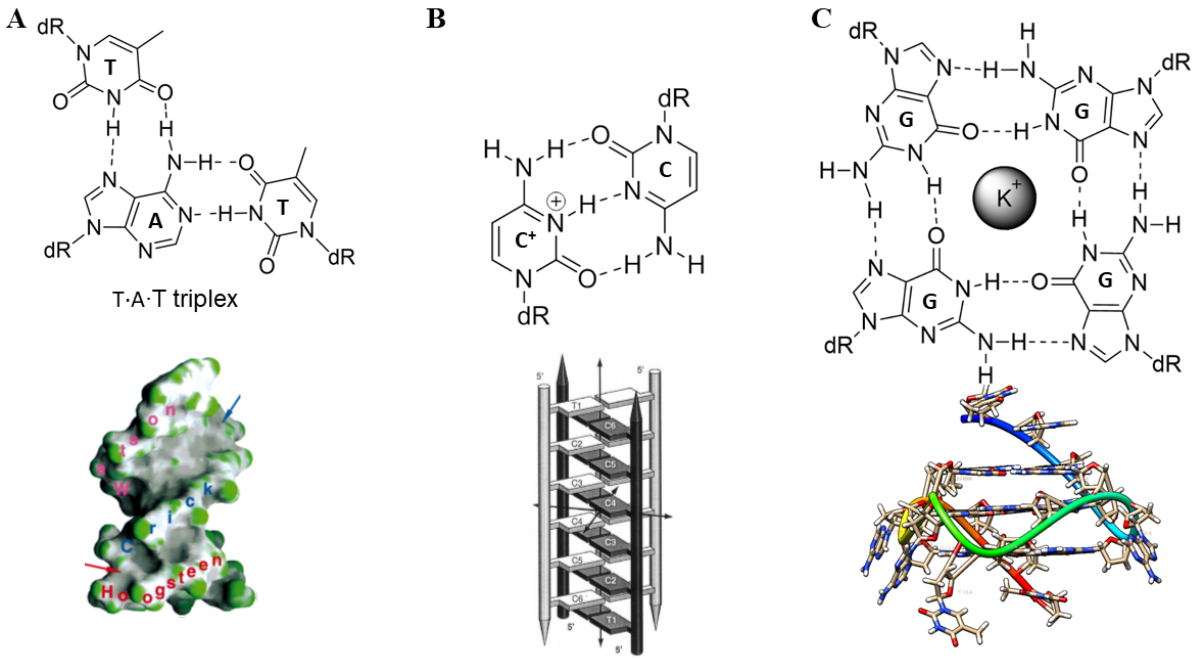


Figure 1.19 Polymorphism of DNA. Hydrogen bonding patterns and tertiary structures. A) DNA triplexes Reproduced with permission from ref [237] B) *i*-motif. Reproduced with permission from ref [119] C) G-quadruplexes. Model produced with UCSF Chimera™. PDB ID: 2LBY

the fact they can form at neutral pH and in physiological conditions. G-quadruplexes form in guanine rich sequences of DNA either intra or intermolecularly and their hydrogen bonding pattern involves both the Watson-Crick and Hoogsteen faces of the guanine nucleobases (**Figure 1.19C**). The G-tetrads are stacked on top of one another and the quantity of tetrads can vary. G-quadruplexes are stabilized by large monovalent cations and potassium is particularly stabilizing.[120] G-quadruplexes are structurally diverse and properties such as molecularity and directionality can depend on sequence and environment among other variables.

1.4.2 Role of G-quadruplexes in Biology

Telomeres are important regions of the genome in many organisms that are found at the ends of chromosomes for protection and contain a long $(TTAGGG)_n$ overhang repeat in humans.[121] The shortening of telomeres is believed to be associated with aging and telomerases are responsible for their extension. Telomerases are misregulated in approximately 85% of cancers and thus telomeres have drawn attention as a drug target.[122] Due to the G-rich nature of this DNA sequence it can fold into a variety of intramolecular G-quadruplexes however the dominant conformation appears to be a "hybrid" with strands running both parallel and antiparallel (in potassium buffer).[123] As depicted in **Figure 1.20A**, interest has been drawn in stabilizing these G-quadruplexes with ligands as the dominant G-quadruplex form inhibits telomerase's ability to extend the telomeres, and halts tumour cell proliferation. Therefore, much research has gone into stabilization of telomeric G-quadruplexes with ligands such as telomestatin[124] and anthraquinones.[125] A further interplay, and challenge, involves the protein POT1 (Protection of telomeres 1) which has been shown to destabilize the G-quadruplexes and restore telomerase function.[126]

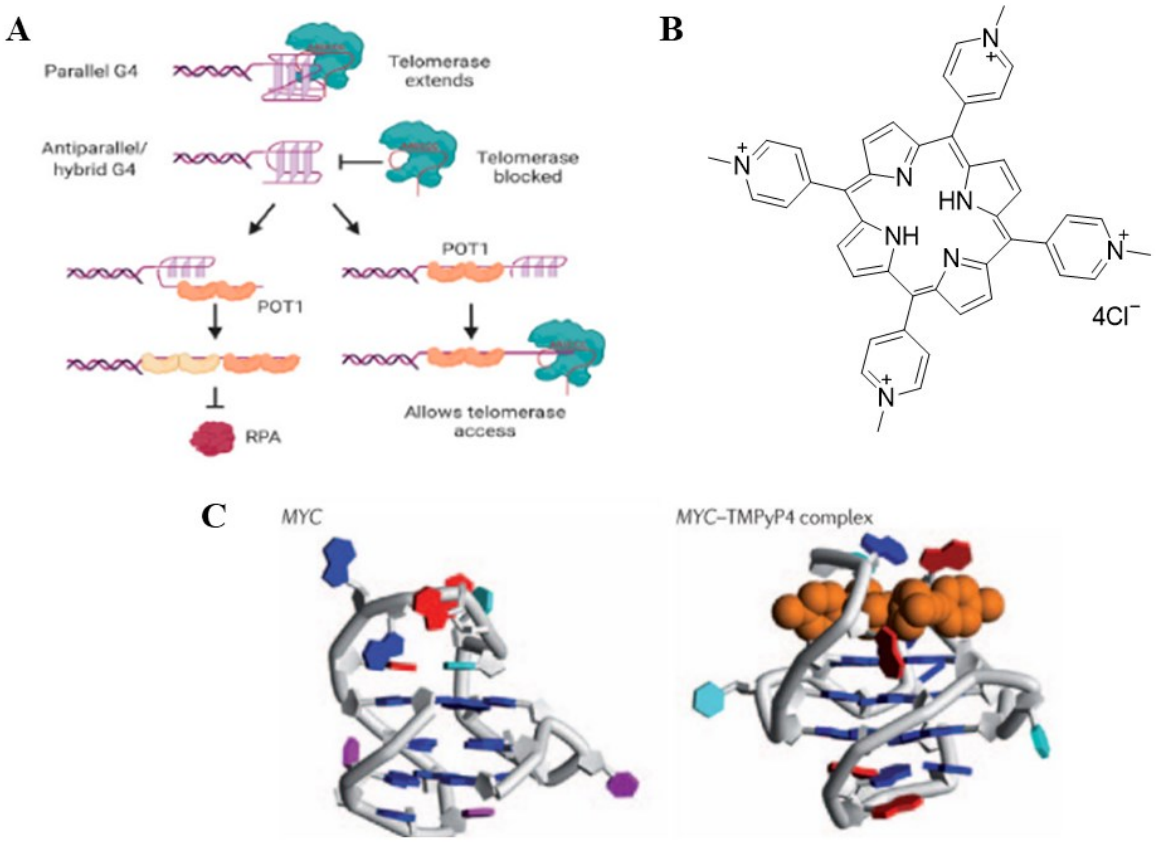


Figure 1.20 G-quadruplexes as a therapeutic target A) Illustration of telomerase being stalled by parallel stranded G-quadruplexes but blocked by antiparallel. POT1 protein is capable of clearing out the G-quadruplex and allowing telomere extension to occur. Reproduced with permission from ref [238] B) G-quadruplex stabilizing ligand TMPyP4 C) NMR structures of *c-myc* G-quadruplex and when complexed with TMPyP4. TMPyP4 is depicted with space filling in orange. Adapted with permission from ref [239].

Other notable regions of the genome are the promoter regions of oncogenes such as *c-myc*, KRAS and VEGF which have also been shown to form G-quadruplexes.[127] The complexity of the G-quadruplex *vs*. duplex equilibrium is not fully understood to date but plays an important role in gene expression.[128,129]. Analogous to telomerase inhibition stabilization of these G-quadruplexes prevents transcription of the oncogene. This research was spurred by targeting the nuclease hypersensitive element NHE of *c-myc*, which is highly implicated in cancer through several pathways.[130] The first ligand was a cationic porphyrin TMPyP4 (**Figure 1.20B**) which exhibited strong reduction of RNA polymerase activity.[131] **Figure 1.20C** depicts the NMR structure of the TMPyP4-G-quadruplex complex providing insights on the mechanism of

stabilization. Since then other small molecules types have been utilized for stabilizing the promoter region G-quadruplexes such as trisubstituted isoalloxazines[132] and platinum complexes.[133] This is a particularly attractive strategy of reducing *c-myc* expression as the protein is generally considered undruggable.[134] Due to the fact there are an estimated 376,000 possible G-quadruplex forming sequences in the genome[135] whose existence or roles are generally undetermined, selectivity for specific G-quadruplexes remains a challenge.

As a brief note, DNA damage can also occur G-quadruplexes and examples studied include O^6-methylguanine and 8-oxo-7,8-dihydroguanine (OG). Fried and coworkers have looked at varying the position of O^6-methyl dG in the human telomeric repeat on stability as well as repair by hAGT.[136] They showed that it was a substrate for hAGT however repair proficiencies depended on its position in the G-quadruplex. An extensive study by Fleming *et al.* look at the oxidative damage OG (and its downstream oxidation products) and their repair by the BER pathway NEIL proteins.[137] What was fascinating is damage in the G-quadruplex studied, *VEGF*, which contains four G tetrads was not a substrate for NEILs. However when they studied the sequence containing the downstream run of guanines, repair occurred which was hypothesized to be due to this additional G tract flipping in to replace the damaged base (which is subsequently flipped out) allowing the protein to repair it.

1.4.3 Biotechnological Applications of G-Quadruplexes

One characteristic of aptamers is they often contain a G-quadruplex for proper folding. Additionally, G-quadruplexes are exceptionally thermally stable, and this has shown to be one of their positive attributes when present in aptamers. Others include nuclease stability and cellular uptake.[138] Thrombin-binding aptamer, the first and most studied, was already discussed in section 1.2.5. Its substrate thrombin is involved in blood coagulation and entered into clinical trials for coronary artery bypass graft surgery (however further success was not achieved).[139] G-quadruplexes containing aptamers contain anticancer properties as well: the nucleolin targeting aptamer discussed previously contains a G-quadruplex. While the earlier example exploited this structure to target tumour cells, the aptamer itself can have antitumour properties and was shown to be effective against human glioma cells.[140] G-quadruplex aptamers have targeted a variety of proteins involved in HIV infection, for example AR177 inhibits HIV integrase and thus stops HIV

from inserting itself into the DNA genome (also tested in clinical trials).[141] Other applications of G-quadruplexes include, but are not limited to, biosensing[142,143] and antibacterial therapeutics.[144]

1.5 Objectives of the thesis and arrangement

1.5.1 Objectives of the thesis

The role of AGT has been well established for maintaining cellular integrity and for chemotherapeutic resistance. The differences between organisms raise evolutionary questions. Why do bacterial *E. coli* AGTs process O^4 methyl dT well but not mammalian? Modeling data has given insights, however, without a crystal structure concrete conclusions are difficult to draw. **Chapter 2** attempts to explore if the active site of OGT can act on various IaCLs for covalent capture for structural determination by high-resolution X-ray crystallography.

DNA nanostructures have garnered much interest due to their ease of synthesis, cost-effectiveness and ability to enter cells intact. Due to hAGT's ability to unhook IaCLs lacking a phosphate backbone, **Chapter 3** seeks to 1) Synthesize three-way junction oligonucleotides linked by either a butylene and heptylene linker 2) Construct a novel DNA tetrahedron with three-way junctions at the vertices and 3) Explore if it could be disassembled in the presence of hAGT.

Non-canonical DNA structures have been gathering attention to their increasing role in biology as well as harnessing their attributes for therapeutic and diagnostic applications. This thesis aims to explore the substrate scope to non-canonical structures such as nanocages or in biologically relevant motifs such as G-quadruplexes. G-quadruplexes are stable structures which are found in parts of the human genome who have a role in gene regulation. Disruption of their stability has effects on gene expression. Methylation has been shown to be destabilizing in certain regions of the genome however larger adducts have not been explored. **Chapter 4** aims to study the biophysical effects of alkylation in the parallel-stranded *c-myc* promoter region G-quadruplex, as well as repair of larger lesions such as IaCLs

Threose nucleic acids contain a four carbon sugar instead of the typical five carbon found in DNA and RNA. Due the simplicity of the molecule, TNA is believed to be a possible precursor to RNA in terms of origins of life. This is supported by their ability to base-pair with DNA and RNA and also be incorporated by polymerases, **Chapter 5** looks at the synthesis of O^4 methylated TNA thymine (tT-OMe), its biophysical properties and repair by AGTs.

1.5.2 Thesis Arrangement

Chapter 2

This chapter investigates how mutations in the active site of hAGT to those of *E. coli* OGT affect unhooking of IaCLs lacking a phosphate backbone. This chapter aims to covalently link IaCLs to OGT's active site and examine the kinetics of the process. We showed the dG-O^6-alkylene-O^6-dG and dT-O^4-alkylene-O^6-dG IaCLs lacking a phosphate backbone were an effective way to covalently capture OGT's active site by PAGE and ESI-MS.

Chapter 3

In this chapter, three-way junction oligonucleotides were synthesized with either a butylene or heptylene IaCL and it was examined whether four three-way junction oligonucleotides could self-assemble into a TDN. This was performed by synthesizing nucleoside phosphoramidites with a fluoride labile TBS group on the internal 5' hydroxyl group. Three-way junction orthogonal extension by oligonucleotide synthesis occurred in approximately 50 % yield. Native PAGE displayed a slow migrating tetrahedron species relative to the controls for both butylene and heptylene linkers and UV thermal denaturation studies confirmed that both assemblies were stable at 37 ºC. Repair assays showed virtually complete consumption of both the three-way junctions and DNA tetrahedra substrates at high protein equivalent and moderate consumption at lower equivalents. This study expands the toolbox of DNA structures that can be disassembled in response to external stimuli.

Chapter 4

In this chapter the effects of O^6 methylation of dG of the *c-myc* promoter region, the potential hydrolysis product formed by busulfan and hepsulfam at the O^6 position of guanine as well as IaCLs lacking a phosphodiester linkage was examined. All O^6 alkyl damage was shown to be destabilizing to the G-quadruplex, and especially when the alkylation was located in the core guanine residue as opposed to a guanine in the loop. Two sites of alkylation were particularly destabilizing to G-quadruplex formation and possibly promoted folding into an alternate G-quadruplex as assessed by CD. All lesions were shown to be substrates for AGTs from human AGT and the hybrid hAGT/OGT (hOGT). The *E. coli* AGTs (Ada-C and OGT) were not able to process the larger lesions as observed previously with duplex DNA. This work elucidates light on

how alkylation possibly affects gene expression, and potential applications in biotechnology (especially for the IaCLs lacking a phosphodiester linkage).

Chapter 5

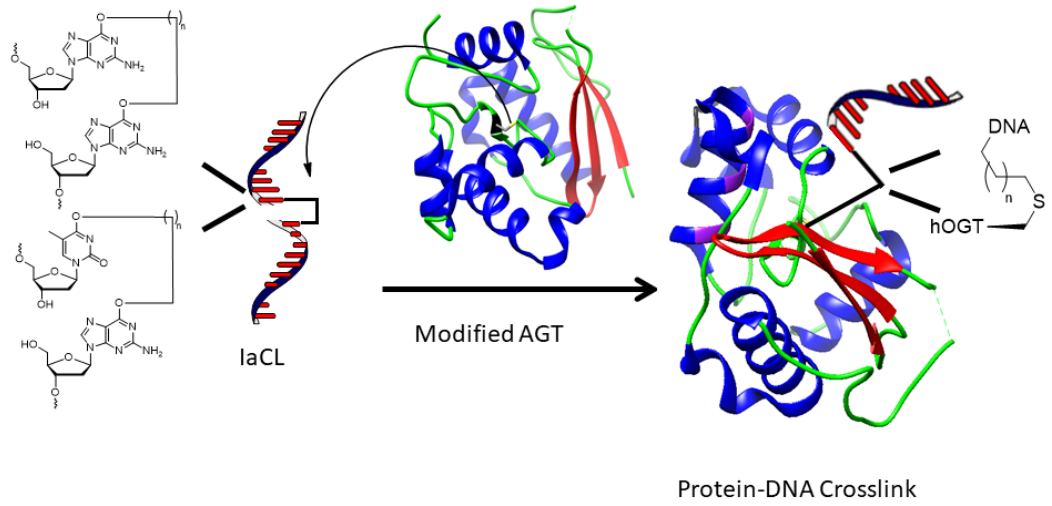
Lastly, the effects of methylation of the O^4 position of thymine in TNA were studied. Using a convertible nucleoside methodology the O^4 position of TNA thymine was methylated, converted to a phosphoramidite and incorporated into a DNA oligonucleotide with minor modifications to automated solid-phase olionucleotide synthesis protocols employing standard β-cyanoethylphosphoramidite chemistry, deprotection and purification. It was shown to act similarly to its O^4-methyl dT counterpart by UV thermal denaturation and circular dichroism studies. Unfortunately, repair assays generally employed by our lab were inconclusive with the TNA analogue and further method development is required.

Chapter 6

This chapter briefly summarizes the results from the thesis and future work explores the future directions for the project.

Chapter 2

Covalent Capture of OGT's Active Site Using Engineered Human-E. coli Chimera and Intrastrand DNA Cross-links



Published as: W. Copp, D. K. O'Flaherty, C. J. Wilds, *Org. Biomol. Chem.* **2018**, *16*, 9053–9058.

Department of Chemistry and Biochemistry, Concordia University, Montréal, Québec, Canada, H4B1R6.

Abstract

O^6-Alkylguanine DNA alkyltransferases (AGTs) are proteins found in most organisms whose role is to remove alkylation damage from the O^6- and O^4-positions of 2'-deoxyguanosine (dG) and thymidine (dT), respectively. Variations in active site residues between AGTs from different organisms leads to differences in repair proficieny: The human variant (hAGT) has a proclivity for removal of alkyl groups at the O^6-position of guanine and the *E. coli* OGT protein has activity towards the O^4-position of thymine. A chimeric protein (hOGT) that our laboratory has engineered with twenty of the active site residues mutated in hAGT to those found in OGT, exhibited activity towards a broader range of substrates relative to native OGT. Among the substrates that the hOGT protein was found to act upon was interstrand cross-linked DNA connected by an alkylene linkage at the O^6-position of dG to the complementary strand. In the present study the activity of hOGT towards DNA containing alkylene intrastrand cross-links (IaCL) at the O^6- and O^4-positions respectively of dG and dT, which lack a phosphodiester linkage between the connected residues, was evaluated. The hOGT protein exhibited proficiency at removal of an alkylene linkage at the O^6-atom of dG but the O^4-position of dT was refractory to protein activity. The activity of the chimeric hOGT protein towards these IaCLs to prepare well defined DNA-protein cross-linked conjugates will enable mechanistic and high resolution structural studies to address the differences observed in the repair adeptness of O^4-alkylated dT by the OGT protein relative to other AGT variants.

2.1 Introduction

DNA is not a chemically inert molecule and subject to damage from various endogenous and exogenous agents.[5] Modifications introduced on the nucleobases of DNA can result in pairing errors during DNA replication, leading to mutagenesis and cancer.[145] For example, alkylation of the O^6 atom of guanine is a mutagenic lesion as O^6-methylguanine preferentially base-pairs with thymine leading to point mutations, namely G:C → A:T transitions, during DNA replication.[146] The last line of cellular defence is apoptosis as O^6-methylguanine is recognized to trigger a cascade of events to ensure the genomic integrity is left uncorrupted.[147] Conversely, some anticancer therapeutics operate by damaging DNA. These include alkylating agents, which can be mono- or bifunctional in nature. The latter can introduce alkylene cross-links between the complementary strands (interstrand cross-links or ICLs) or same strand (intrastrand cross-links or IaCLs) of a DNA duplex. Busulfan and hepsulfam are examples of cross-linking agents which result in butylene or heptylene linkages, respectively, introduced onto DNA.[148] Other exogenous chemical agents, such as 1,2-dihaloalkanes, are capable of introducing ethylene and propylene cross-links in biological settings.[149] These cross-linking agents have a propensity to form their lesions at the N7 atom of guanine.[4]

Maintenance of the integrity of DNA is essential and several cellular repair pathways are invoked to remove lesions.[150] O^6-Alkylguanine DNA alkyltransferases (AGTs) are responsible for removing alkyl groups at the O^6 atom of 2'-deoxyguanosine, and to a lesser extent the O^4 atom of thymidine in mammals.[151] The first AGTs studied were the *E. coli* proteins Ada-C and OGT, which remove a methyl group irreversibly from the O^6-position of guanine consequently inactivating the protein.[152] Once the protein is alkylated, it is then targeted for degradation by the ubiquitin pathway.[19] While difficulties were encountered in purification, it was noted that a similar protein existed in mammalian tissues and which led to the discovery that AGTs were found in most eukaryotes, mammals and archea.[151,153] Essentially all AGTs studied contain the conserved PCHR amino acids in the active site where the sulfur atom of Cys145 (human numbering) is responsible for the nucleophilic attack on the alkyl lesion's α carbon. The unusually high reactivity of the sulfur atom is believed to be due to a coordinated hydrogen bond network between Glu172, His146 and a bridging water molecule. Unusually, AGTs bind DNA's minor groove *via* a helix-turn-helix motif showing minimal sequence specificity.[15] Flipping of the damaged nucleotide by human AGT (hAGT) is mediated by Tyr114, which serves to perturb the

phosphate group 3' to the alkyl lesion *via* steric[15] and/or electrostatic interactions.[154] The extrahelical guanine nucleobase conformation is stabilized by Arg128 which intercalates the duplex and forms hydrogen bonds with the opposing strand's cytosine.

Despite the conserved residues described above, there are stark contrasts in AGTs ability to repair alkyl lesions across species. In addition to methyl adducts, hAGT has been shown to be proficient at removing other lesions at the O^6-position of dG such as benzyl groups.[31] Conversely, unlike hAGT, OGT from *E. coli* has a proclivity for repairing O^4-methyl T.[155] While O^6-methyl dG is a more prevalent lesion, research suggests that O^4-methyl T could be more mutagenic.[156,157] Initial investigations of the origins of the disparity were shown not to be due to binding, with little discrimination with AGTs for unmodified *vs.* alkylated DNA.[158] *E. coli* AGTs are believed to interact with DNA *via* a scanning mechanism close to the diffusion limit;[16] a more cooperative mechanism has been observed in hAGT with a ratio of ~4 base-pairs/hAGT molecule.[20]

hAGT, adept at removing damage at the O^6-position of dG, has demonstrated the ability to act upon dG-O^6-alkylene-O^6-dG ICL.[35,36] However, a dT-O^4-alkylene-O^4-dT ICL was shown to be refractory to repair by both hAGT and OGT.[37] Counterintuitively, while hAGT is capable of repairing dG-O^6-alkylene-O^6-dG ICLs whereas OGT is not, OGT binds cross-linked DNA with a higher affinity than hAGT.[159] This suggests binding alone does not explain the enhanced repair by hAGT and in fact it has been proposed the tight binding might inhibit the conformational change (nucleotide flipping) required for repair.[159] The active sites of the variants, while having the aforementioned conserved residues, do differ in amino acid sequence (see **Section 3.4.2 Table S2.1**). Various mutants of hAGT have been prepared to investigate why hAGT weakly repairs the potentially more mutagenic O^4-methyl dT damage.[160] Initial studies suggested the reason that the O^4-methyl lesion evades repair by hAGT is that it is not positioned closely enough to the reactive active site cysteine. To investigate the propensity of hAGT and OGT to repair O^6-alkyl dG and O^4-alkyl dT lesions, Pegg *et al.*[156] performed site-directed mutagenesis to elucidate the influence of the species-dependent amino acids. Their most successful mutant, hAGT-03, had 8 hAGT active site residues mutated to those of OGT. Their models attribute an additional glycine residue as well as a rearrangement in the active site loops allowing the active cysteine to orient itself sufficiently

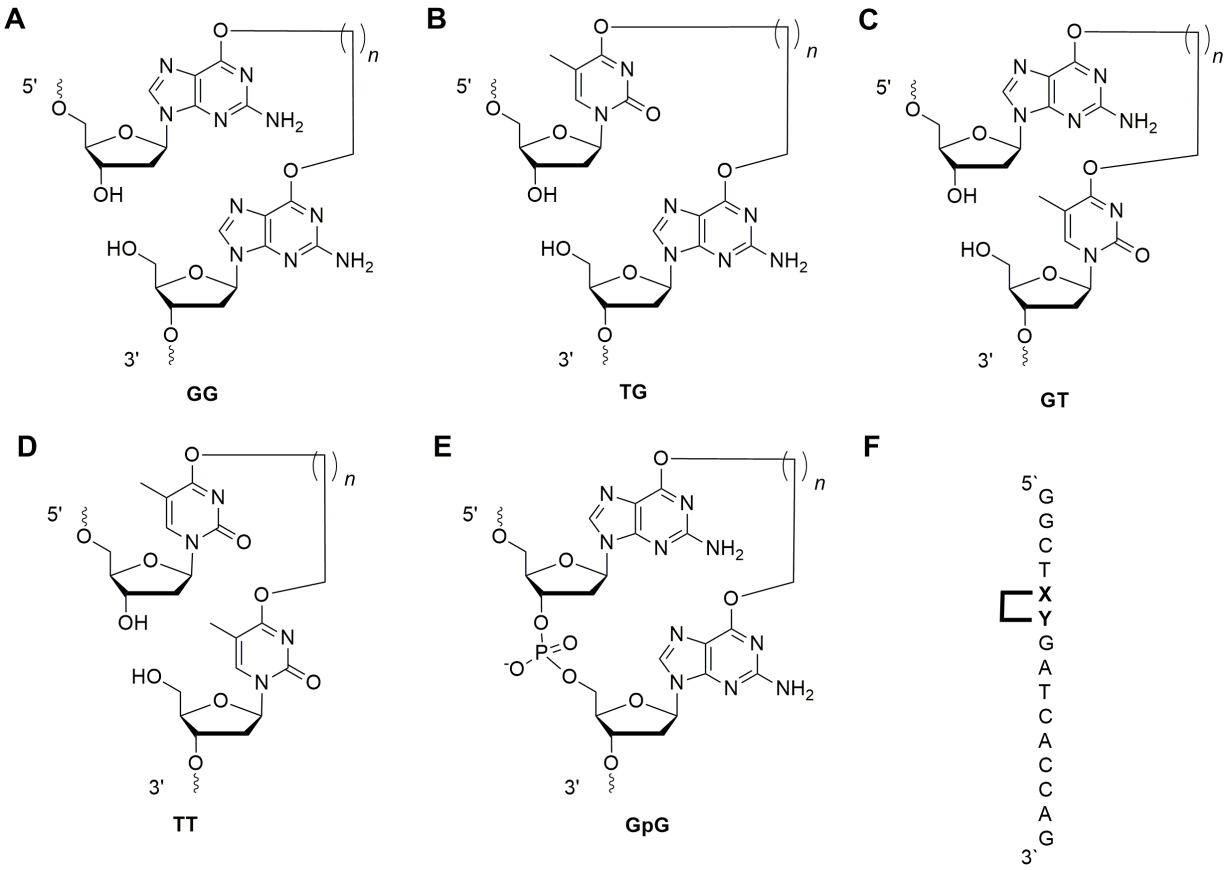


Figure 2.1 Structures of IaCLs in this study. A) O^6-2'-deoxyguanosine-alkylene-O^6-2'-deoxyguanosine (n = 3, 6), B) O^4-thymidine-alkylene-O^6-2'-deoxyguanosine (n = 1, 2, 3, 6), C) O^6-2'-deoxyguanosine-alkylene-O^4-thymidine (n = 3, 6), D) O^4-thymidine-alkylene-O^4-thymidine residues lacking a phosphodiester backbone linkage (n = 3, 6) and E) O^6-2'-deoxyguanosine-alkylene-O^6-2'-deoxyguanosine (n = 3, 6) containing a phosphodiester backbone linkage. F) The oligonucleotide sequence indicating the position of the IaCLs (shown by XY).

close to the O^4-methyl group of dT facilitating alkyl transfer. Our lab has recently engineered a chimera of hAGT containing active site residues 139-159 of OGT (with the exception of Pro 140 for folding and solubility purposes) to investigate its ability to repair various O^4-alkyl-dT and O^6-alkyl-dG lesions.[51] The hybrid nature of the chimera (hOGT) led to an amalgamation of hAGT and OGT's abilities to repair both O^6- and O^4-alkyl lesions. Impressively, the chimera could repair bulky O^4 lesions such as a heptylene mono-adduct and dG-O^6-alkylene-O^6-dG ICLs. However, hOGT was unable to repair dT-O^4-alkylene-O^4-dT ICLs. The broader substrate range was

attributed to more flexible loops as well as Arginine 135's steric clash with the C5 methyl group in the hAGT-DNA interaction; where glycine is the corresponding residue in OGT.

To date the exact roles of these key active site residues in OGT have not been confirmed and unfortunately there is no crystal structure of this protein. Preparation of DNA-protein cross-links (DPCs) has stimulated interest for various purposes such as protein detection,[161–164] studying protein-DNA interactions,[165,166] nanotechnology,[167] therapeutics[168] as well as structural studies.[169,170] To date, no DNA-OGT covalent conjugates have been prepared, which have potential for structural determination by X-ray crystallography that would offer valuable insights as to the repair proficiency of the OGT protein towards O^4-alkylated dT. The covalent capture of hAGT with N^1,O^6-ethanoxanthosine provided one of the first crystal structures of hAGT whilst providing important mechanistic information.[33] DNA-hOGT cross-links have been prepared with DNA containing an O^6-alkylene dG ICL, however the reaction was inefficient (20-30 % repair after 8 h for butylene ICL and 5 h for heptylene ICL) and required large quantities of protein (30-fold protein-DNA excess). Furthermore, while the role of active site amino acids has been thoroughly investigated in O^6-alkylene-dG and O^4-alkylene-dT ICL, little is known about the effect of modifying active site residues on hAGT's ability to repair DNA IaCLs (**Figure 2.1**). Here we present the first studies on hOGT and its activity towards various IaCLs by denaturing PAGE with ^{32}P labelled oligonucleotides, SDS-PAGE and ESI-MS studies to prepare DPCs. These IaCLs lack a phosphodiester linkage between the residues connected by an alkylene linkage at the O^6- or O^4-positions of the guanine or thymine nucleobases, respectively.[40,42] hAGT demonstrated proficient activity towards unhooking of the O^6-2'-deoxyguanosine-alkylene-O^6-2'-deoxyguanosine (**GG**) and O^4-thymidine-alkylene-O^6-2'-deoxyguanosine (**TG**) IaCL. The demonstrated activity towards these IaCL, which may be attributed in part to the greater flexibility of the backbone without the phosphodiester linkage, expands the repertoire of nucleic acids structures, including G-quadruplexes containing O^6-methyl-dG adducts,[136] that the hAGT protein can act upon. In the present study, it was demonstrated that both the **GG** and **TG** IaCL are also processed by hOGT. However, the IaCL DNA consisting of an O^6-2'-deoxyguanosine-alkylene-O^4-thymidine (**GT**), O^4-thymidine-alkylene-O^4-thymidine (**TT**) and O^6-2'-deoxyguanosine-alkylene-O^6-2'-deoxyguanosine containing a phosphodiester backbone linkage (**GpG**) were not acted upon by the hOGT protein. Covalent capture is a well-validated method of studying DNA-protein interactions.[171] This strategy to prepare specific DNA-hOGT cross-linked species may

ultimately be applied for structural studies to elucidate the basis for differences observed in OGT's repair activity relative to hAGT.

2.2 Results and Discussion

2.2.1 Repair of IaCLs by hOGT

All substrates investigated herein have been prepared according to previously reported procedures.[39–42] The various IaCL substrates contain alkylene linkers of different lengths (see **Section 2.4.2 Table S2.2**) for the various sequences investigated). For example, **GG4** refers to the O^6-2'-deoxyguanosine-butylene-O^6-2'-deoxyguanosine IaCL lacking the intradimer phosphodiester group. The strand containing the IaCL was 5'-^{32}P radiolabelled and annealed to its complement. The ability of hOGT to repair 5'-^{32}P labelled IaCLs in duplex DNA was investigated by monitoring repair products by denaturing PAGE. IaCL DNA (2 pmol) was incubated with hOGT (10 or 60 pmol) overnight at 37 °C, reactions were quenched, and repair products loaded on the denaturing gel (**Figure 2.2**). Lanes 1 and 2 (**Figure 2.2**) contained unmodified DNA with and without hOGT as negative controls, and lane 7 (**Figure 2.2**) contained both possible repair products as a positive control. The 10mer control was radiolabelled as the presence of an internal 5'-OH at the IaCL presents a second potential site for phosphorylation; therefore, repair of the oligonucleotide can produce two detectable products. It should be noted that repair product percentages are reported from time-courses when 5-fold protein abundance was tested, and 18h for total repair when 30-fold.

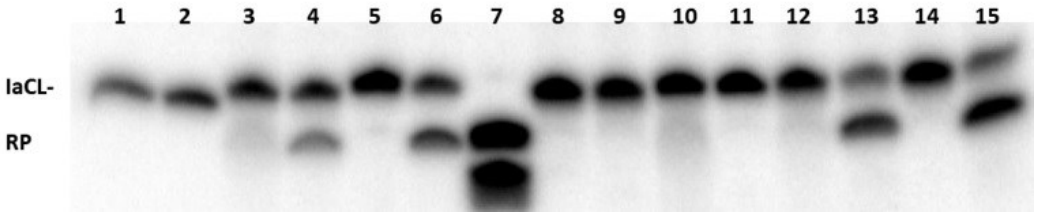


Figure 2.2 Repair of IaCL duplexes by hOGT. 1) 2 pmol Control-1, 2) 2 pmol Control-1 + 10 pmol hOGT, 3) 2 pmol **GG4**, 4) 2 pmol **GG4** + 10 pmol hOGT, 5) 2 pmol **GG7**, 6) 2 pmol **GG7** + 10 pmol hOGT, 7) 5mer and 10mer repair product controls (5'-GGCTG and 5'-GGATCACCAG), 8) 2 pmol **TG2**, 9) 2 pmol **TG2** + 10 pmol hOGT, 10) 2 pmol **TG3**, 11) 2 pmol **TG3** + 10 pmol hOGT, 12) 2 pmol **TG4**, 13) 2 pmol **TG4** + 10 pmol hOGT, 14) 2 pmol **TG7** 15) 2 pmol **TG7** + 10 pmol hOGT.

Inspection of lanes 4 and 6 (**Figure 2.2**) showed moderate repair of **GG4** and **GG7**, with 69.1±3.9 % and 81.3±0.3 % repair observed respectively with 5-fold protein abundancy. The unhooking of the IaCLs was slightly less efficient relative to hAGT, which showed repair product accumulations of 76 % and 91 % for **GG4** and **GG7** respectively.[40] The more efficient repair of the heptylene linked IaCL relative to butylene is consistent with trends observed for both ICLs[35,36] and IaCLs.[40] This is thought to be due to the increased conformational flexibility of the heptylene linker furthermore the mechanism of cross-link repair is thought to involve two AGT molecules, where the second AGT acts on the DPC resulting in an AGT-alkylene-AGT product. Modelling studies suggest the tunnel-like structure of the AGT's active site allows a second hAGT protein to act on the more accessible hAGT-heptylene-O^6-DNA adduct than the shorter butylene cross-link.[35] Repair with 30-fold protein was significantly more proficient, with 87.5±0.8 % repair observed for **GG4** and 93.2±3.1 % for **GG7** (see **Section 2.4.2 Figure S2.1A**) as expected. Lanes 9 and 11 (**Figure 2.2**) containing **TG2** and **TG3** respectively with 10 pmol hOGT revealed no evidence of repair products, which were shown to be poorly repaired by hAGT.[41] For hAGT and hOGT, repair of the GG IaCLs with 5-fold protein concentration is significantly better than those of ICLs at 30-fold concentration presumably due to the flexible backbone allowing the adducted base to be flipped into the protein's active site. At 30-fold protein concentration, slight repair of **TG3** was observed (see **Section 2.4.2 Figure S2.1C**) consistent with previous observations. The weak repair of **TG2** and **TG3** by hOGT is likely due to the shorter alkylene tether restricting rotation of the modified nucleobase for optimal positioning in the protein active site. Repair of **TG4** (lane 13, **Figure 2.2**) and **TG7** (lane 15, **Figure 2.2**) followed the same trend as GG cross-links albeit less efficiently, with repair of 71.9±1.3 % and 74.0±5.8 % respectively with 5-fold excess hOGT. Similar repair was again observed with 30-fold excess of protein, with repair efficiency of 65.3±1.2 % for **TG4** and 76.5±1.6 % for **TG7** (see **Section 2.4.2 Figure S2.1B**). hOGT was envisioned as a construct of hAGT capable of enhanced O^4-methyl dT repair and it was disappointing that it was unable to repair dT-O^4-alkylene-O^4-dT ICLs.[37] Oligonucleotides containing TT IaCLs lacking a phosphodiester linkage were immune to unhooking despite the enhanced flexibility, which would presumably allow the flipping of the damaged nucleobase into hOGT's active site. This is despite the attributes of a larger active site for hOGT and the efficient removal of O^4-alkylation observed by native OGT. Unfortunately, both **TT4** and **TT7** showed no repair by hOGT even with increased protein levels (see **Section 2.4.2 Figure S2.2**), contributing

to the list of AGTs unable to unhook these cross-links. Modelling has suggested these larger alkyl lesions promote the incorrect rotamer of Cys145 in the active site of OGT, which prevents nucleophilic attack.[37] **GT4** and **GT7** where the O^6-alkylated dG residue is 5' to the alkylated thymine were examined next. At both 5-fold and 30-fold hOGT concentrations no repair products were observed, as was seen with hAGT (see **Section 2.4.2 Figure S2.3**).[42] This is presumed to be due the absence of a 3' phosphate to the O^6-alkylene dG lesion which prevents Tyr114 from sterically probing the phosphate, which promotes the damaged nucleobase to flip into the protein's active site. The last set of oligonucleotides examined were **GpG4** and **GpG7** which contained O^6-dG-alkylene-O^6-dG IaCLs with the presence of a phosphodiester backbone (see **Section 2.4.2 Figure S2.4**). No repair products were observed for either lesion at 5-fold nor at 30-fold protein concentration. In contrast, hAGT was able to show weak activity at 30-fold protein concentration when the intradimer phosphate backbone was present, with 26 % repair towards **GpG4** and 36 % repair towards **GpG7**; slightly less repair than that observed for the GG ICLs.[36,39] Rationale for lack of repair was presumably due to a combination of the phosphate group and alkylene cross-link restraining nucleobase flipping. The active site of OGT present in hOGT further impedes unhooking as it is not optimal (relative to hAGT's) for O^6-alkyl dG repair; this could explain the absence of GpG unhooking. Repair of IaCLs which hOGT showed activity towards were then monitored as a function of time.

2.2.2 Times to reach maximum repair for hOGT on IaCLs

The repair of duplexes formed between **GG4**, **GG7**, **TG4** and **TG7** with the complementary strand were monitored as a function of time by incubating 2 pmol of dsDNA with 10 pmol of hOGT, quenching the reactions at intervals and subjecting the aliquots to denaturing PAGE (see **Section 2.4.2 Figures S2.5** and **S2.6**). Maximum repair by hOGT was reached after 45 minutes for **GG4** and 10 minutes for **GG7** (**Figure 2.3**). This trend was previously observed by hAGT acting on the GG IaCLs and repair occurs more quickly on **GG7** likely due to a more accessible α carbon due to more flexible heptylene linkage. Additionally, the longer length of the alkylene tether allows a second AGT molecule to access the DPC which likely allows quicker accumulation of repair products. Similarly, unhooking by hOGT of **TG7** reached maximum repair in one-third the time it took for **TG4**, in agreement with all previous studies demonstrating more efficient repair for cross-links containing longer alkylene linkages between nucleotides. However, while hAGT and hOGT

have similar proficiency of unhooking dG-O^6-alkylene-O^6-dG ICLs,[51] hOGT activity towards the IaCLs studied proceeded more efficiently than hAGT (**Figure 2.3A-C**) for **GG4**, **GG7** and **TG4**: maximum repair occurred in approximately 11, 3.5 and 2-fold less time respectively. The role of Pro140, which is retained from hAGT, could allow the chimera to accommodate larger substrates such as ICLs and can partially explain the ability of the chimera's ability to act on the IaCLs.[35,36,159] It has been suggested that the C150G and N157G mutations in hAGT-03 (and therefore hOGT) causes a longer, and more flexible active site loop allowing the reactive cysteine to access the O^6-alkyl damage which could explain the faster repair.[156] The time required for maximum repair was nearly identical for **TG7** by hAGT and hOGT. Whereas hAGT reached the plateau of maximum repair of the TG series more quickly than the GG series, hOGT had similar rates for **GG4** and **TG4**, and peak repair was reached for **GG7** 5 minutes before **TG7**. Evidently, the 5' nucleotide's identity is not of crucial importance when hOGT is processing the IaCLs, and the slight improvement of

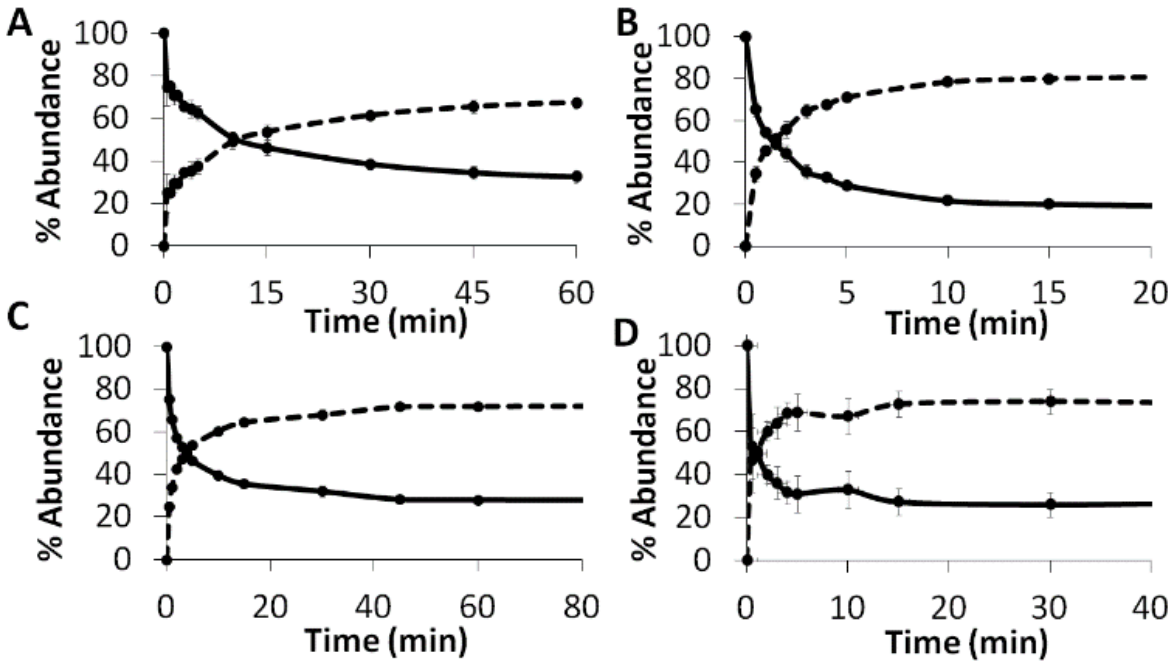


Figure 2.3 Time course repair assay of IaCLs by hOGT. A) **GG4**, B) **GG7**, C) **TG4**, D) **TG7** by hOGT at 5-fold excess protein. Depicting substrate (—) and repair product (– –) abundance (%) as a function of time (min).

hOGT's repair of **GG7** or **TG7** could be from the improved flexibility about the glycosidic bond of purines over pyrimidines.[172] A crystal structure revealed the methyl lesions on the O^6 atom of

guanine needs to be oriented in a syn position to be in sufficient proximity to the reactive Cys145 therefore some conformational flexibility is required.[15] It should be noted that hOGT also had similar rates of repair for both GG and TG ICLs.[51] The characterization of the site of repair, and the formation of DPCs were subsequently examined.

2.2.3 Identification of DPCs by SDS-PAGE and ESI-MS

While ^{32}P labelled oligonucleotides were effective for the monitoring of the formation of various DNA products of hOGT activity, direct monitoring of DPCs was accomplished by SDS-PAGE and ESI-MS. Equimolar ratios of ssDNA oligonucleotides and hOGT were incubated for 2 h at 37 °C, quenched and then subject to gel electrophoresis. Lanes 2 and 3 (**Figure 2.4**), which contained hOGT and hOGT with unmodified DNA control respectively, had a similar migration relative to the 25 kDa molecular weight marker (lane 1). The lanes containing hOGT and **GG4** or **GG7** (lanes 4 and 5, respectively in **Figure 2.4**) showed the presence of a second species with lower mobility presumed to be the higher molecular weight DPC, as shown previously with hAGT.[39] Lanes 6 and 7 (**Figure 2.4**) containing **TG4** and **TG7**, also contained a second band with lower mobility presumed to be a DPC species. Consistent with trends observed in the repair assays, a more intense DPC band was noted in lanes 5 and 7 (**Figure 2.4**) which contained heptylene IaCL incubation products. Lanes 8 and 9 (**Figure 2.4**) with the GpG IaCL series containing a phosphodiester linkage showed one band corresponding to the hOGT protein alone; consistent with the lack of reaction observed, as demonstrated above.

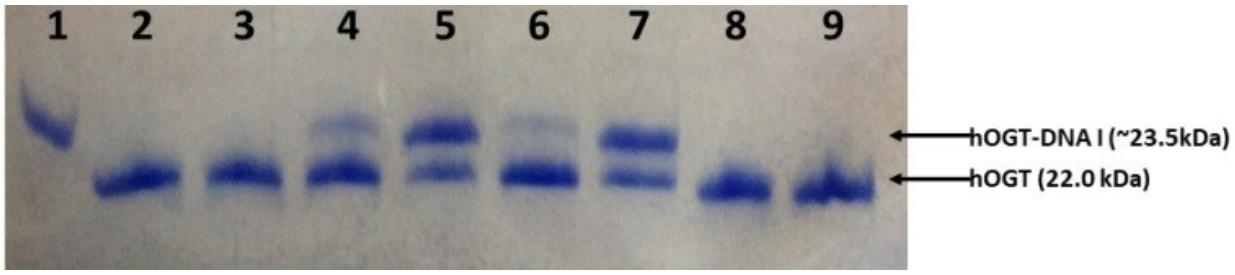


Figure 2.4 15% SDS-PAGE of DNA-hOGT cross-links. Lane 1) 25 kDa molecular weight marker, 2) hOGT, 3) hOGT+DNA, 4) hOGT+**GG4**, 5) hOGT+**GG7**, 6) hOGT+**TG4**, 7) hOGT+**TG7**, 8) hOGT+**GpG4**, 9) hOGT+**GpG7**.

While an effective technique for qualitative observation of DPCs, we turned to mass spectrometry for more accurate characterization as well as structural information (e.g. which DNA residue reacted with the protein) of the lower mobility species observed by gel electrophoresis. As with the SDS-PAGE samples, 300 pmol of hOGT was incubated with the IaCLs which showed evidence of repair and subjected to ESI-qTOF mass spectrometry. For all four oligonucleotides containing intrastrand cross-links two peaks were observed (**Figure 2.5**). In all four spectra the species with lower *m/z* was in close agreement with unmodified hOGT. Mass spectrometry of reaction mixtures of **GG4** and **GG7** when incubated with hOGT (**Figure 2.5A** and **B**) showed the presence of a higher *m/z* species, which was in agreement with the corresponding to butylene and

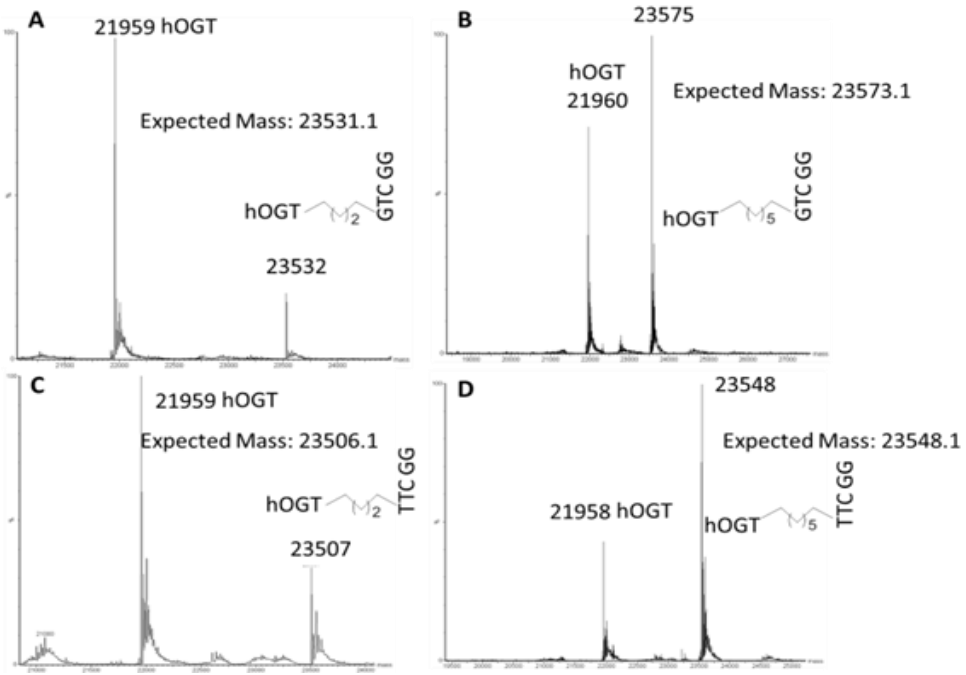


Figure 2.5 Deconvoluted ESI-MS data of DNA-hOGT. DPC products formed with A) **GG4**, B) **GG7**, C) **TG4**, D) **TG7**. Expected masses are for the respective DPC products.

heptylene linked 5'-fragment conjugate, respectively. The preference of hOGT's thiolate anion to act on the α carbon of the O^6-alkylated dG 3' to the lesion does not deviate from previous data, for both **GG4** and **GG7**, concerning hAGT. Similarly, hOGT-alkylene-O^4-thymidine DPC were observed (**Figure 2.5C** and **D**) when the protein was incubated with **TG4** and **TG7**. As observed previously hOGT selectively repairs at the O^6-position of dG. Unhooking of the 3' end of the lesion

by hOGT is less remarkable considering hOGT was not capable of repairing the O^4-position of dT for **TT4** and **TT7**.

2.3 Conclusions

The present study describes a methodology to covalently cross-link the active site of *E. coli* OGT, replacing the corresponding residues in hAGT, to oligonucleotides *via* alkylene linkages. hOGT had a similar repair efficiency to hAGT towards dG-O^6-alkylene-O^6-dG IaCL. Remarkably, the maximum repair was reached more quickly by hOGT, relative to hAGT, especially when processing **GG4**. The improved repair is thought to be due to the larger active site due to Pro140 as well as a more flexible loop. Repair of **GG** IaCLs possessing a phosphodiester linkage between the cross-linked nucleosides by hOGT was not observed. Repair of 5'-dT-O^4-alkylene-O^6-dG (but not 5'-dG-O^6-alkylene-O^4-dT) IaCL occurred to similar levels and timescales by hOGT as hAGT. hOGT had issues unhooking the shorter ethylene linked **TG2**, and some repair was observed with **TG3** at increased protein levels. DPC formation by gel electrophoresis was observed with a new higher molecular weight species in lanes where repair occurred. The observed mass of DPCs was in agreement with the expected mass by ESI-MS, and the site of repair was determined to be exclusive to the O^6-position of guanine. The formation of an OGT-DNA covalent complex has potential for structure determination by X-ray crystallography to offer valuable insights as to the active site of OGT, as was accomplished with the human AGT structure. This is of crucial importance as there is no structure available of an AGT capable of repairing O^4-methyl dT.[155] Moreover, the tethering of alkylene linked DNA provides may provide insight regarding the accommodation of adducts in the OGT active site.

The preparation of DPCs with hOGT *via* IaCLs occurs more rapidly and efficiently than with previously described ICLs whilst using 6-fold less protein. Promisingly, until this study hAGT was the only known protein to repair the aforementioned IaCLs. We believe our methodology for covalent capture of the active site of OGT shows great potential in the preparation of DPC species and mechanistic studies to study mutations in the active site of hAGT.

2.4 Supporting Information

2.4.1 General Methods

2.4.1.1 General

Most reagents for biochemical assays, and protein purification, were purchased from Bioshop Canada Inc (Burlington, ON). EDTA-free protease inhibitor cocktail tablets were purchased from Roche (Laval, QC) and Ni-NTA Superflow Resin from Qiagen (Mississauga, ON). T4 polynucleotide kinase was bought from Fermentas (Burlington, ON), and [γ-^{32}P]ATP from PerkinElmer (Woodbridge, ON). DNA primers were acquired from Biocorp (Montreal, QC) and Phusion Polymerase from New England Biolabs (Ipswich, MA). Stratagene (Cedar Creek, TX) provided XL-10 Gold and Bl21(DE3) *E. coli* cells.

2.4.1.2 Synthesis, purification and characterization of modified oligonucleotides

All oligonucleotides were synthesized, deprotected and characterized as described previously.[39–42]

2.4.1.3 Protein purification and overexpression

Protein was purified and overexpressed as described previously.[36,51] Protein masses were verified by ESI-MS run in positive ion mode on a Micromass Q-Tof Ultima API.

2.4.1.4 Denaturing PAGE repair assays

100 pmol DNA was incubated with 1 µL of γ-[^{32}P]-ATP (10µCi/µL), 5 units of T4 PNK in 10 µL of 1xPNK buffer for 1 hour at 37 °C for radiolabelling. After 1 h of incubation the samples were boiled for 10 mins, after which 110 pmol of complement strand was added, the total solution volume was brought to 50 µL of H_2O. The solution was boiled for 5 minutes, slowly cooled to room temperature and incubated at 4 °C overnight to give a concentration of 2 µM of dsDNA. For total repair assays 2 pmol of DNA was then aliquoted with 10 (or 60 pmol) of hOGT and the total volume was brought to 15 µL with activity buffer [10 mM Tris–HCl (pH 7.6), 100 mM NaCl and 1 mM DTT]. The samples were incubated at 37 °C overnight, after which the reactions were quenched by adding 18.2 µL of stop buffer [81 mM Tris–HCl, 81 mM boric acid, 1.8 mM EDTA and 1 % SDS (sodium dodecyl sulfate)] and boiling for 5 minutes. The samples were loaded on a 14 cm × 16 cm, 20 % (19:1 acrylamide bis-acrylamide), 7M urea PAGE which was run for 45

minutes in 1X TBE (longer run times were required for **GpG4** and **GpG7** to separate repair products from substrate). Time course assays were performed by preparing a master mix containing 80 pmol of protein, 16 pmol of DNA and a total volume of 120 µL activity buffer. The samples were incubated at 37 °C and at defined time points 7.5 µL aliquots were removed and quenched by boiling for 2 minutes followed by addition of 9.1 µL of stop buffer. The samples were subject to denaturing PAGE as described above. To visualize the oligonucleotides, gels were exposed to a storage phosphor screen, images captured on a Typhoon 9400 (GE Healthcare, Piscataway, NJ) and data was analyzed with ImageJ (National Institutes of Health, Bethesda, MD). Relative quantities of substrate and repair products were calculated by summing the integrated bands in each lane and dividing the corresponding band by the total.

2.4.1.5 SDS-PAGE of DPCs

300 pmol of hOGT was incubated with 300 pmol of DNA for 2.5 h in activity buffer [10 mM Tris–HCl (pH 7.6), 100 mM NaCl and 1 mM DTT] at 37 °C. The total reaction volume was 25 µL. To the samples was added 6 µL of stop buffer [200 mM Tris-HCl (pH 7.6), 3.2 % SDS, 48 % glycerol and 0.4 M DTT], and 15 µL of sample was loaded on a 15 % resolving SDS-PAGE. The samples were electrophoresed for 2 h at 120 V on a 7.5 cm x 10 cm plates. hOGT and DPCs were visualized by Coomassie blue stain.

2.4.1.6 ESI-MS of DPCs

300 pmol of hOGT was incubated with 300pmol of DNA for 30 min in activity buffer [10 mM Tris–HCl (pH 7.6), 100 mM NaCl and 1 mM DTT] at 37 °C in a total reaction volume of 25 µL. A 5 µL aliquot was diluted in 0.1 % (v/v) aqueous formic acid and subject to HPLC on an Agilent 1200 Series system using a Grace™ Vydac™ C4 (214MS) column (Fisher Scientific) (100 mm x 2.1 mm). Ionization occurred *via* 3.5 kV electrospray source and masses were scanned in positive ion mode. ESI mass spectra were acquired on a Micromass Qtof2 mass spectrometer (Waters) at the CBAMS.

2.4.2 Supporting Figures and Tables

Table S2.1 Active site residues of hAGT, OGT and hOGT.

Protein	Amino Acid Sequence (139-159) [a]
hAGT	VPILIIPCHRVVCSSGAVGNYS
OGT	ISIVVPCHRVIGRNGTMTGYA
hOGT [b]	I**P**IVVPCHRVIGRNGTMTGYA

[a] Human numbering.

[b] Proline140, retained from hAGT, is bolded and underlined.

Table S2.2 Oligonucleotide sequences studied in this paper.

Oligonucleotide	Sequence
Control-1	5′-GGC TGG GAT CAC CAG
GG4	5′-GGC T**G4G** GAT CAC CAG
GG7	5′-GGC T**G7G** GAT CAC CAG
GpG4	5′-GGC T**Gp4G** GAT CAC CAG
GpG7	5′-GGC T**Gp7G** GAT CAC CAG
Control-2	5′-GGC TTG GAT CAC CAG
TG2	5′-GGC T**T2G** GAT CAC CAG
TG3	5′-GGC T**T3G** GAT CAC CAG
TG4	5′-GGC T**T4G** GAT CAC CAG
TG7	5′-GGC T**T7G** GAT CAC CAG
Control-3	5′-GGC TGT GAT CAC CAG
GT4	5′-GGC T**G4T** GAT CAC CAG
GT7	5′-GGC T**G7T** GAT CAC CAG

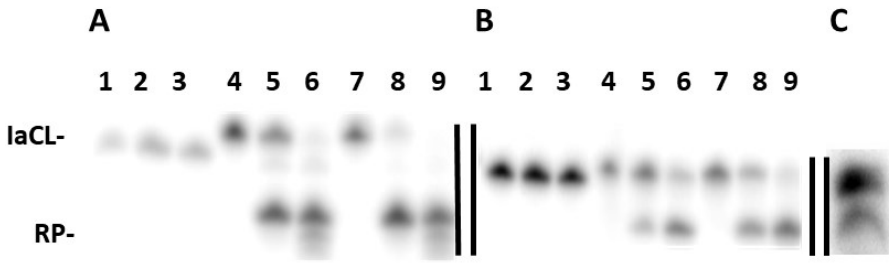


Figure S2.1 – Total repair of **GG4**, **GG7**, **TG4** and **TG7** IaCLs by hOGT. **A)** Lane 1) 2 pmol DNA + 0 pmol hOGT; Lane 2) 2 pmol DNA + 10 pmol hOGT; Lane 3) 2 pmol DNA + 60 pmol hOGT; Lane 4) 2 pmol **GG4** + 0 pmol hOGT; Lane 5) 2 pmol **GG4** + 10 pmol hOGT; Lane 6) 2 pmol **GG4** + 60 pmol hOGT; Lane 7) 2 pmol **GG7** + 0 pmol hOGT; Lane 8) 2 pmol **GG7** + 10 pmol hOGT; Lane 9) 2 pmol **GG7** + 60 pmol hOGT. **B)** Lane 1) 2 pmol DNA + 0 pmol hOGT; Lane 2) 2 pmol DNA + 10 pmol hOGT; Lane 3) 2 pmol DNA + 60 pmol hOGT; Lane 4) 2 pmol **TG4** + 0 pmol hOGT; Lane 5) 2 pmol **TG4** + 10 pmol hOGT; Lane 6) 2 pmol **TG4** + 60 pmol hOGT; Lane 7) 2 pmol **TG7** + 0 pmol hOGT; Lane 8) 2 pmol **TG7** + 10 pmol hOGT; Lane 9) 2 pmol **TG7** + 60 pmol hOGT. **C)** 2 pmol **TG3** + 60 pmol hOGT.

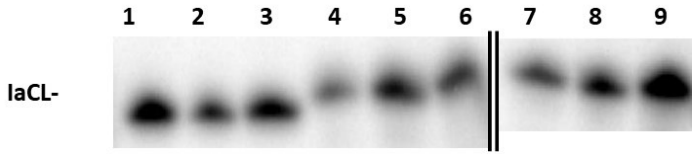


Figure S2.2 – Total repair of **TT** IaCLs by hOGT. Lane 1) 2 pmol DNA + 0 pmol hOGT; Lane 2) 2 pmol DNA + 10 pmol hOGT; Lane 3) 2 pmol DNA + 60 pmol hOGT; Lane 4) 2 pmol **TT4** + 0 pmol hOGT; Lane 5) 2 pmol **TT4** + 10 pmol hOGT; Lane 6) 2 pmol **TT4** + 60 pmol hOGT; Lane 7) 2 pmol **TT7** + 0 pmol hOGT; Lane 8) 2 pmol **TT7** + 10 pmol hOGT; Lane 9) 2 pmol **TT7** + 60 pmol hOGT.

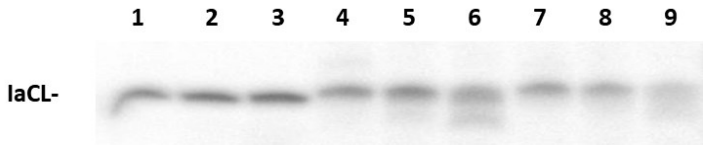


Figure S2.3 – Total repair of **GT** IaCLs by hOGT. Lane 1) 2 pmol DNA + 0 pmol hOGT; Lane 2) 2 pmol DNA + 10 pmol hOGT; Lane 3) 2 pmol DNA + 60 pmol hOGT; Lane 4) 2 pmol **GT4** + 0 pmol hOGT; Lane 5) 2 pmol **GT4** + 10 pmol hOGT; Lane 6) 2 pmol **GT4** + 60 pmol hOGT; Lane 7) 2 pmol **GT7** + 0 pmol hOGT; Lane 8) 2 pmol **GT7** + 10 pmol hOGT; Lane 9) 2 pmol **GT7** + 60 pmol hOGT.

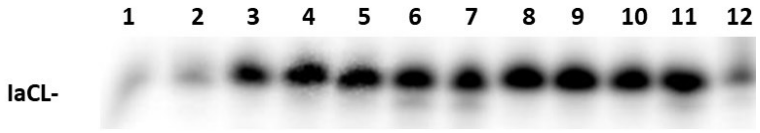


Figure S2.4 – Total repair of **GpG** IaCLs by hOGT; Lane 1) 2 pmol DNA + 0 pmol hOGT; Lane 2) 2 pmol DNA + 60 pmol hOGT; Lane 3) 2 pmol **GpG4** + 0 pmol hOGT; Lane 4&5) 2 pmol **GpG4** + 10 pmol hOGT; Lane 6&7) 2 pmol **GpG4** + 60 pmol hOGT; Lane 8) 2 pmol **GpG7** + 0 pmol hOGT; Lane 9&10) 2 pmol **GpG7** + 10 pmol hOGT; Lane 11&12) 2 pmol **GpG7** + 60 pmol hOGT.

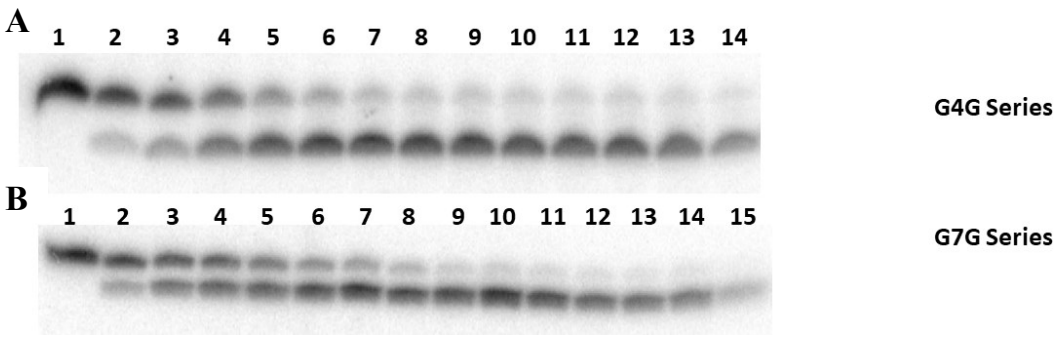


Figure S2.5 Denaturing PAGEs of repair as a function of time of 2 pmol **GG** IaCL duplexes incubated with 10 pmol hOGT. A) **GG4** Lane 1) Control; Lanes 2-14) 1, 2.5, 5, 10, 15, 25, 35, 45, 60, 90, 120, 180, 240 min respectively. B) **GG7** Lane 1) Control; Lanes 2-15) 0.5, 1, 1.5, 2, 3, 4, 5, 10, 15, 30, 45, 60, 90, 120 min respectively.

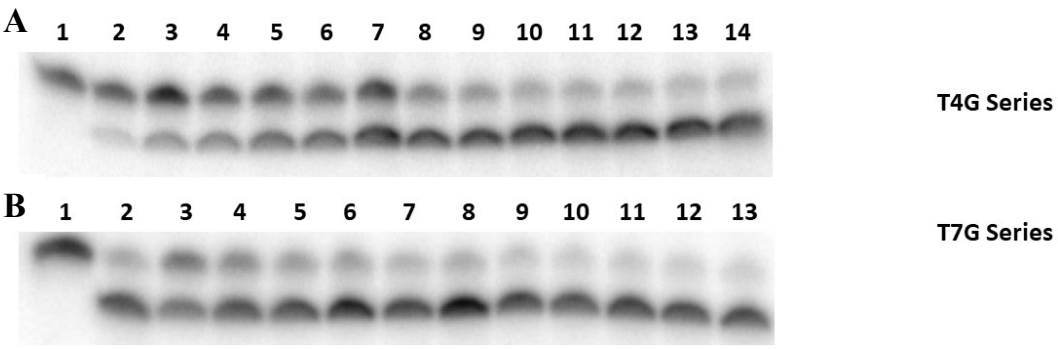
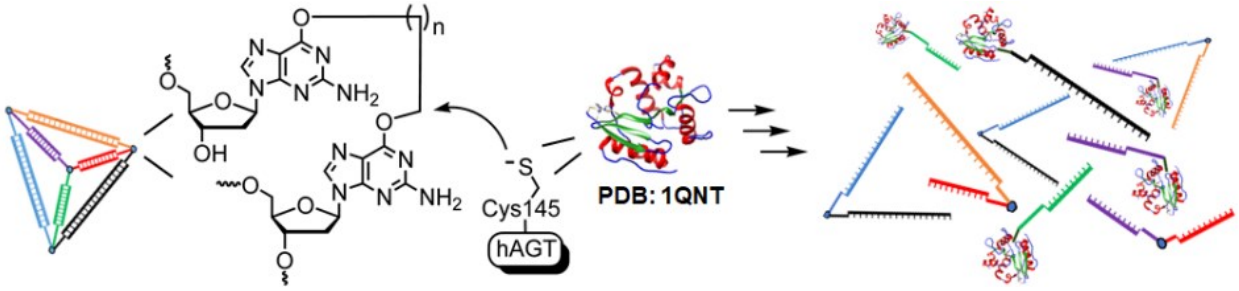


Figure S2.6 Denaturing PAGEs of repair as a function of time of 2 pmol **TG** IaCL duplexes incubated with 10 pmol hOGT. A) **TG4** Lane 1) Control; Lanes 2-14) 0.5, 1, 1.5, 2, 3, 4, 5, 10, 15, 30, 45, 60, 90, 120, 180 min, respectively. B) **TG7** Lane 1) Control; Lanes 2-15) 0.5, 1, 1.5, 2, 3, 4, 5, 10, 15, 30, 45, 60, 90, 120 min, respectively.

Chapter 3

O^6-Alkylguanine DNA Alkyltransferase Mediated Disassembly of a DNA Tetrahedron



Published as: W. Copp, C. J. Wilds, *Chem. - A Eur. J.* **2020**, *26*, 14802–14806.

Department of Chemistry and Biochemistry, Concordia University, Montréal, Québec, Canada, H4B1R6.

Abstract

Tetrahedron DNA structures were formed by the assembly of three-way junction oligonucleotides containing O^6-2'-deoxyguanosine-alkylene-O^6-2'-deoxyguanosine (butylene and heptylene linked) intrastrand cross-links (IaCLs) lacking a phosphodiester group between the 2'-deoxyribose residues. The DNA tetrahedra containing three-way junctions were shown to undergo an unhooking reaction by the human DNA repair protein O^6-alkylguanine DNA alkyltransferase (hAGT) resulting in structure disassembly. The unhooking reaction of hAGT towards the DNA tetrahedra was observed to be moderate to virtually complete depending on the protein equivalents. DNA tetrahedron structures have been explored as drug delivery platforms that release their payload in response to triggers such as light, chemical agents or hybridization of release strands. The dismantling of DNA tetrahedron structures by a DNA repair protein contributes to the armamentarium of approaches for drug release employing DNA nanostructures.

3.1 Introduction

O^6-Alkyguanine DNA alkyltransferases (AGTs) are proteins found in numerous organisms whose role is removal of alkyl adducts at the O^6 position of 2'-deoxyguanosine (dG).[151,153] AGTs bind DNA at the minor groove *via* a recognition helix-turn-helix motif, resulting in groove widening.[15] AGT's mechanism of DNA repair invokes a nucleophilic cysteine in the protein's active site (residue Cys145 in the human protein variant) to act upon the alkyl lesion's α carbon. The alkyl group is irreversibly transferred from the O^6 atom of dG to this cysteine residue and the resulting alkylated protein is subsequently degraded by the ubiquitin pathway.[173] hAGT has been shown to repair an array of alkyl lesions at the O^6-position of dG.[151] There are increasing reports demonstrating that the specific repair reaction of hAGT can occur with various structures including G-quadruplexes and DNA origami, and has been utilized for applications such as protein detection.[53,54,136]

The self-assembly properties of DNA have been harnessed to produce a plethora of DNA architectures such as cubes,[61] tetrahedra,[62,174] and various other polyhedra[175–177] which can self-assemble due to Watson-Crick base pairing. The aforementioned nanoarchitectures have been employed as cages for guest molecules such as proteins,[73,178] gold nanoparticles,[64,74] and small molecules.[75] The remarkable stability of these structures poses challenges for controlled release of their nanocargo upon reaching their target. Thus, various strategies have been explored where stimuli can induce a conformational change or disassembly of the nanocage upon reaching its destination. For example, Sleiman and coworkers have demonstrated that the use of an eraser strand (an oligonucleotide sequence which hybridizes to a toehold sequence in the construct) leads to the release of the payload.[67,75,179] Other successful stimuli to release the guest molecules from DNA nanocarriers include temperature,[56] pH,[77] small molecule recognition[76] and light.[78]

Inspired by these elegant approaches, we sought to engineer a DNA nanocarrier that could be dismantled in response to hAGT activity to contribute to the growing list of techniques for nanostructure disassembly. We selected a tetrahedral architecture due to its simplicity and the observation that such structures have been shown to enter cells without external functionalization such as endocytic ligands or transfection agents.[180] To this end, we adapted the DNA tetrahedron designed by von Kiedrowski *et al*.[181] composed of four three-way junction (**TWJ**) oligonucleotides bound by a tris alkylene-phenyl scaffold at the vertices (see **Section 3.5.2 Figure S3.1**). In our construct, the vertices of the three-way junction oligonucleotides contain an alkylene

intrastrand cross-link (IaCL) connecting adjacent dG residues at the O^6-position (**Figure 3.1A**) by butylene or heptylene linkages.[40] This particular IaCL, lacking a phosphodiester backbone linkage between the nucleosides, has been shown to be efficiently unhooked at the 3'-dG residue leading to the formation of a hAGT-DNA conjugate and a DNA strand. We demonstrate harnessing of the repair reaction of hAGT towards an O^6-alkylene dG cross-linked containing DNA tetrahedron, to disassemble the structure (**Figures 3.1B** and **C**).

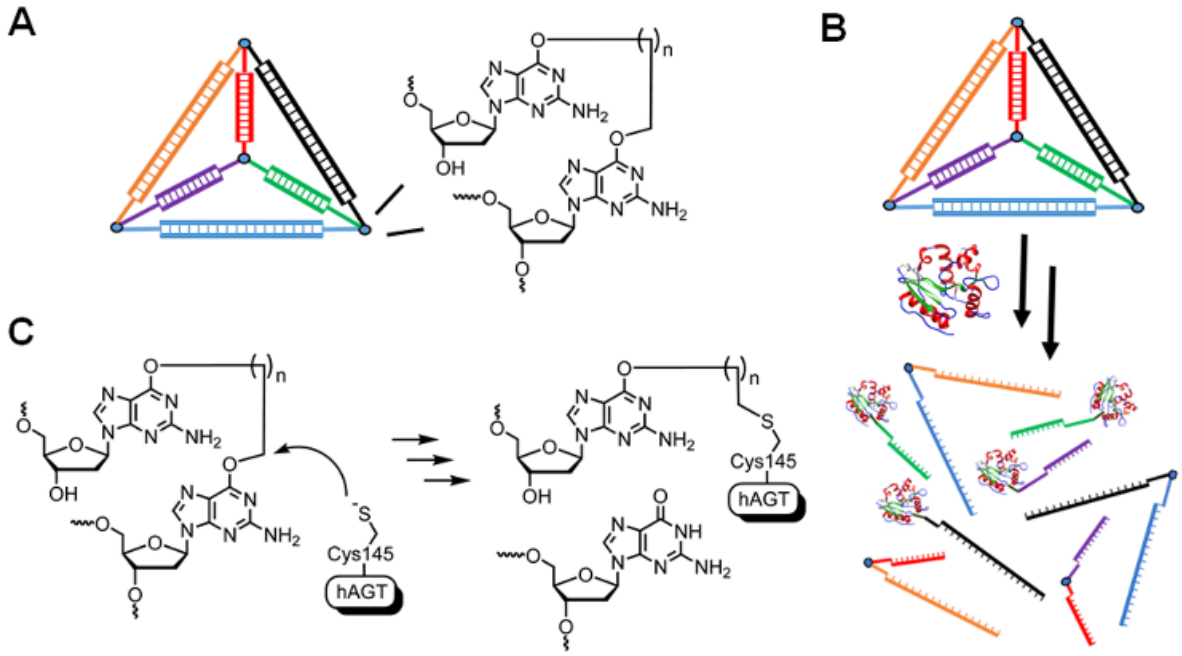


Figure 3.1 Self-assembly of three-way junction oligonucleotides containing IaCL at the vertices into a tetrahedral architecture. (**A**) Depiction of four three-way junctions containing O^6-alkylene dG IaCL at the vertices. (**B**) Incubation with hAGT (**PDB: 1QNT**) leads to unhooking of each vertex resulting in tetrahedra dissociation. (**C**) Abridged mechanism of hAGT unhooking of O^6-alkylene dG IaCL.

3.2 Results and Discussion

The synthesis of the **TWJ** oligonucleotides required modification of the internal 5'-hydroxyl group of a dG-O^6-alkylene-O^6-dG phosphoramidite has been described previously.[40] Full details of this modified synthesis can be found in **Section 3.5.2**. Specifically, the internal 5'-hydroxyl group of

the dG-O^6-alkylene-O^6-dG dimers were protected with a fluoride labile *tert*-butyldimethylsilyl (TBS) group (see **Section 3.5.2 Schemes S3.1** and **S3.2**). Removal of the TBS group followed by resumption of solid-phase oligonucleotide assembly was performed to produce the three-way junctions. Phosphoramidite dimers **3.8a** and **3.8b**, containing a butylene and heptylene linkage were incorporated into oligonucleotides (**C4-** and **C7TWJ1-4**, respectively) with slight modifications to oligonucleotide synthesis procedures (see **Section 3.5.2 Scheme S3.2** and **Figures S3.1-3.29**). The synthesis of the three-way junction oligonucleotides involves protection of the 3'-OH of the branched dimer and 5'-OH of the first assembled strand with a levulinoyl group, which can be selectively removed with hydrazine if additional functionalization of the three-way junction oligonucleotides is desired.[182] The **TWJ** sequences **C4TWJ1-4** and **C7TWJ1-4** were deprotected then analyzed and purified by denaturing preparatory PAGE (**Section 3.5.2 Figure S3.30**). Although the deprotection and extension reactions did not proceed quantitatively, the linear precursors and desired three-way junction products were well resolved by denaturing PAGE (see **Section 3.5.2 Figure S3.31**). The calculated and expected masses determined by ESI-qTOF-MS were in close agreement (see **Section 3.5.2 Figures S3.32**, **S3.33** and **Table S3.1**).

The ability of hAGT to unhook the single-stranded three-way junctions was investigated before evaluating their self-assembly properties (**Figure 3.2** and **Section 3.5.2 Figure S3.34**). Each three-way junction oligonucleotide (2 pmol) was incubated with hAGT (either 20 pmol or 60 pmol) for 18 h, quenched and analyzed by 20 % denaturing PAGE. hAGT acting on these IaCLs can generate 15- and 30-mer repair products (both of which are ^{32}P-radiolabelled and designated as RPs). The center lane of the gel contains 15- and 30-mer controls. At the top of the gel, the species with lowest mobility are hAGT-DNA cross-links (designated as DPC) which have been reported.[40] The proficiency of hAGT to unhook the individual **C4TWJ** oligonucleotides were comparable to the linear dG-dG IaCLs oligonucleotide series evaluated.[40] The 30-mer unhooking product was observed in greater abundance due to the tendency of hAGT to act upon the 3' nucleotide of the IaCL.[40] This has been attributed to the Tyr114 residue of hAGT which plays a role in the base flipping of the alkylated guanine.[15,154] The **C4TWJ1** and **C4TWJ4** oligonucleotides were completely unhooked with 20 pmol of hAGT in 18 h. The reaction was slightly less proficient towards the **C4TWJ2** with 26±5 % and 19±7 % of substrate remaining with 20 and 60 pmol of hAGT, respectively after 18 h. Similarly, 22±6 % and 11±5 % of **C4TWJ3** remained after being incubated with 20 and 60 pmol of hAGT. Similar to previous reports on DNA

containing O^6-dG-alkylene-O^6-dG interstrand cross-links, the reaction of hAGT with the **TWJ** oligonucleotides does not go to completion.[36] This may be attributed to the alkylated hAGT binding with the DNA whereas *in vivo* the alkylated protein would undergo ubiquitination and degradation.[183,184] Another possibility is that the cross-link in the tetrahedron may adopt a less reactive orientation, as suggested for other lesions at the O^6-atom of dG.[185] The variance in unhooking efficiency by hAGT may also be attributed to differences in nucleotide composition around the IaCL, observed before for repair of methyl lesions at the O^6-atom of dG.[186] Unhooking of **C7TWJ** by hAGT was evaluated with virtually complete consumption of **C7TWJ1** observed with 20 pmol of hAGT. **C7TWJ2** was processed to a greater extent relative to **C4TWJ2**, consistent with observations that longer, more conformationally flexible cross-links are acted upon more proficiently by the hAGT protein.[35]

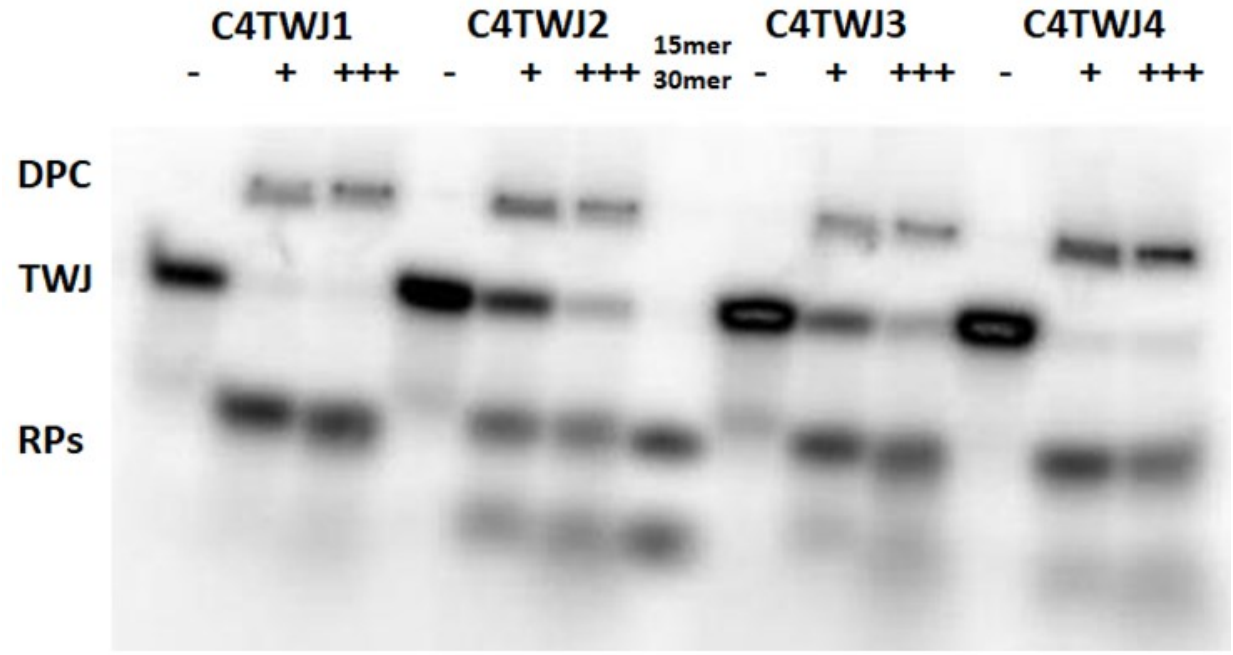


Figure 3.2 20 % Denaturing PAGE to monitor unhooking of the single stranded **C4TWJ** series by hAGT. Each **TWJ** was ^{32}P-labelled in activity buffer. The **TWJ** (2 pmol) of DNA were incubated with 0 (-), 20 (+) or 60 (+++) pmol of hAGT for 18 h. **RPs** (repair products) are the result of the unhooking reaction by hAGT and **DPC** (DNA-protein cross-link) are hAGT-DNA cross-linked species.

Having demonstrated that hAGT was capable of unhooking the single-stranded three-way junctions, assembly of the three-way junctions into DNA tetrahedra was then evaluated. UV thermal melting experiments were performed on samples containing all four three-way junctions for both the **C4** and **C7** series and displayed a monophasic transition with a T_m of 59 °C (see

Section 3.5.2 Figure S3.35). For both the **C4TWJ** and **C7TWJ** series, each possible combination of the oligonucleotides were annealed to produce dimer, trimer and tetramer assemblies and the products analyzed by non-denaturing PAGE in TAMg buffer. The 10 % native PAGE revealed a distinct stepwise trend with products of decreasing mobility, depending on the number of three-way junctions incubated, for both the **C4** and **C7**-series (**Figure 3.3** and **Section 3.5.2 Figure S3.37**). Inspection of the lane labelled T, where all four **C4**- or **C7TWJ** to form the tetrahedron were present, one major band was observed with reduced mobility suggesting the novel DNA tetrahedra were formed in high yield. To corroborate tetrahedron formation, mung bean nuclease (MBN) assays were performed.[181] As MBN is known to digest single but not double-stranded DNA, the DNA tetrahedron would be expected to be refractory to its activity. Analysis of products by denaturing PAGE revealed that the **C4** and **C7** ss**TWJ**s are degraded by MBN whereas the assembled tetrahedra remained intact (see **Section 3.5.2 Figure S3.38**). This suggests that the complementary sequences for both the **C4** and **C7** DNA tetrahedra are double stranded.

With evidence of the self-assembly and thermal stability of the IaCL containing DNA tetrahedra, the ability of hAGT to act upon these structures were determined by denaturing PAGE. The differences in unhooking efficiency of hAGT on each vertex of the tetrahedra were monitored separately by radiolabelling of each individual TWJs which were then annealed with the non-radiolabelled complements. It should be noted that the integrity of the **C4**- and **C7**-tetrahedral structures in hAGT activity buffer was verified by native PAGE with a reduced mobility relative to the assembly formed by three of the four TWJ oligonucleotides (see **Section 3.5.2 Figure S3.39**). Varying amounts of hAGT (0, 12.5 or 37.5 pmol) were added to the tetrahedra (1.25 pmol) at 37 °C for 18 h, after which samples were quenched by the addition of stop buffer. Boiling of the samples was avoided to ensure no repair product (RP) artefacts due to hAGT's activity towards denatured tetrahedra species.

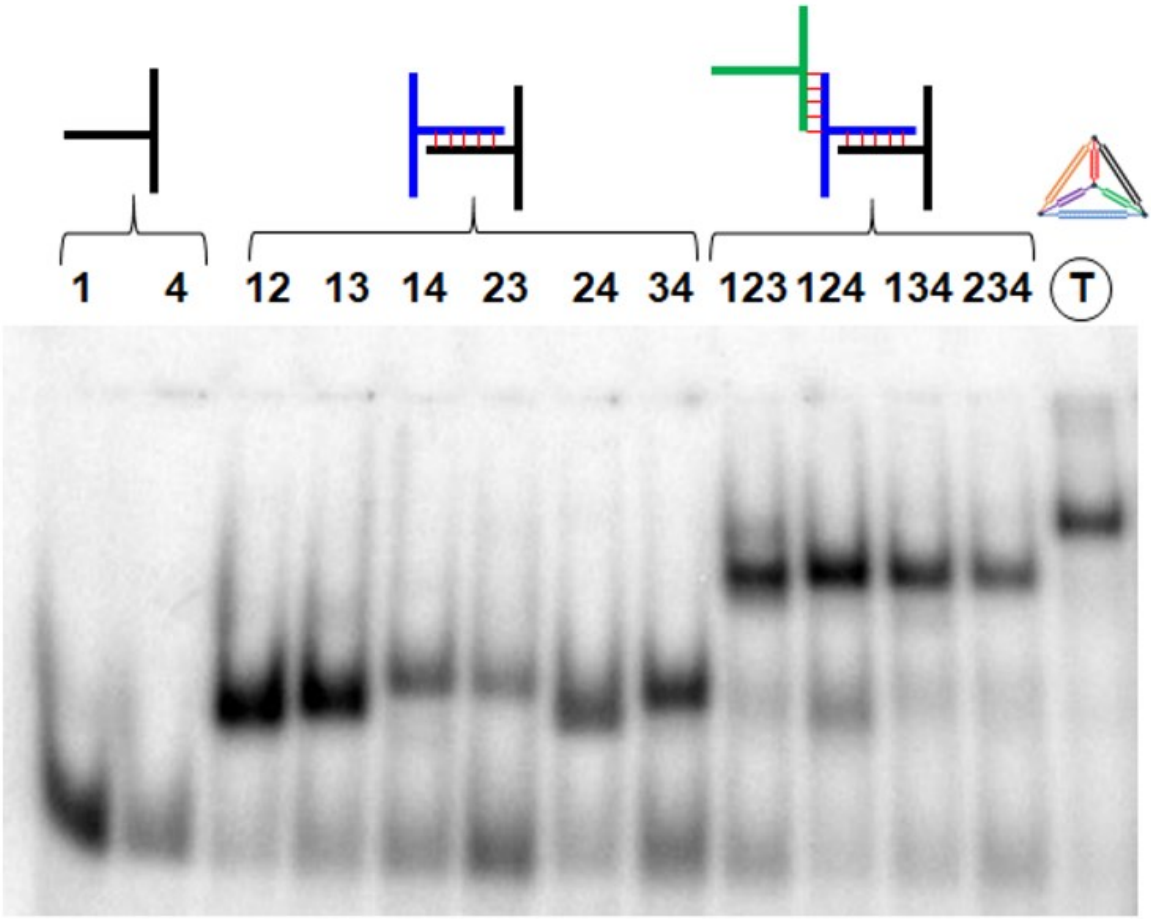


Figure 3.3 10 % Non-denaturing PAGE of each combination of **C4TWJ** oligonucleotides to produce dimer, trimer and tetrahedra (final lane) species.

The reactions were then analyzed by denaturing PAGE for both the **C4** and **C7** series (**Figure 3.4** and **Section 3.5.2 Figure S3.40**). For the **C4TWJ** series, at each of four labelled vertices, the substrates underwent an efficient reaction with 10 equivalents of hAGT with depletion of the tetrahedra observed after 18 h. Tetrahedra labelled at vertices **C4TWJ1** and **C4TWJ4** revealed most efficient reaction at these sites with 19±3 % and 22±1 % of the species remaining, respectively. Similar to the three-way junction series, efficient unhooking by hAGT was observed at the **C4TWJ2** and **C4TWJ3** labelled vertices with 40±3 % and 48±3 % remaining for each. The similar trends in unhooking for hAGT towards the IaCL in individual three-way junction or in the tetrahedral assembly, illustrate the importance of flanking nucleobase composition on reaction efficiency. With 30 equivalents of hAGT unhooking after 18 h was practically complete at each vertex.

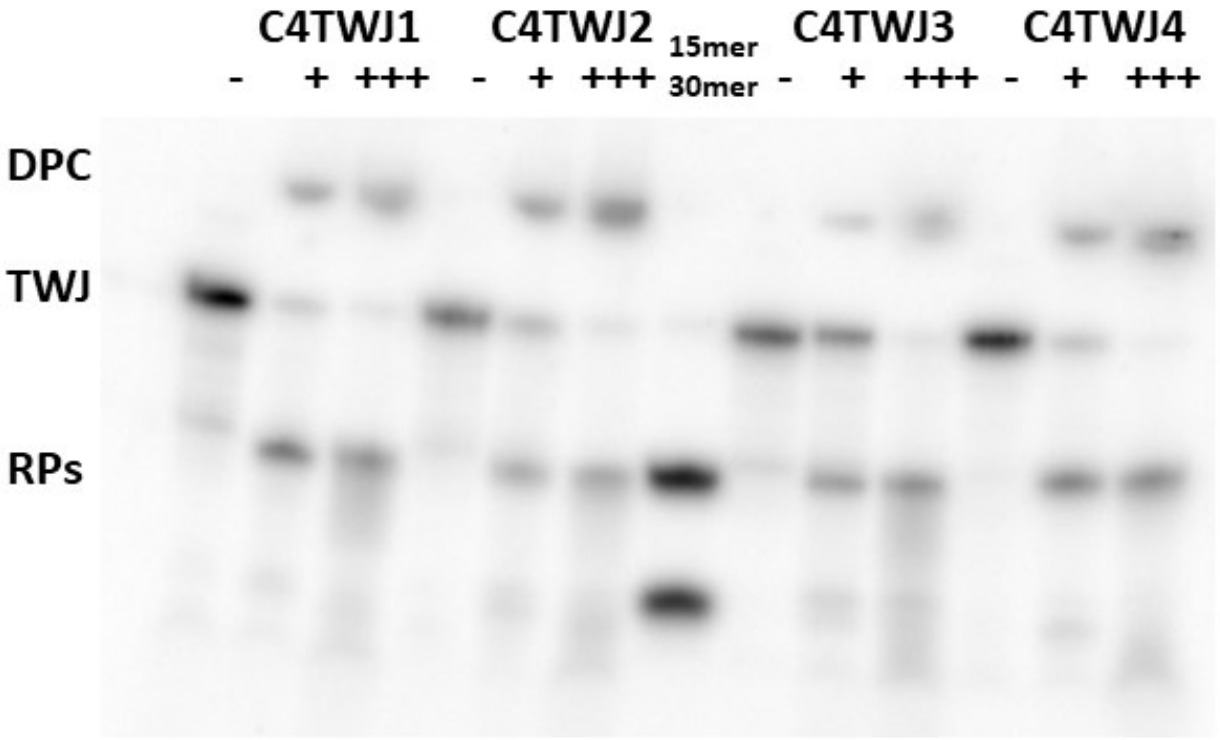


Figure 3.4 20 % denaturing polyacrylamide gel to monitor unhooking of the **C4TWJ** series tetrahedra by hAGT. Each labelled vertex (indicated at the top of the gel) was separately ^{32}P-labelled and annealed with the other complementary (not radiolabelled) **TWJ**s in TAMg buffer. Tetrahedra (1.25 pmol) were incubated with either 0 (-), 12.5 (+) or 37.5 (+++) pmol of hAGT at 37 °C for 18 h. Lane 7 contains ssDNA 15- and 30-mer controls corresponding in length to the products resulting from the unhooking reaction of hAGT.

For the **C7** series, at 10 equivalents of hAGT with the labelled **C7TWJ1** vertex 21±3 % of substrate was remaining after 18 h similar to the **C4** counterpart. For the **C7TWJ4** labelled vertex, 28±10 % of substrate remained unconsumed, slightly less efficient relative to the **C4** analog. Unhooking by hAGT at labelled vertices **C7TWJ2** and **C7TWJ3** were the least efficient in the series with 51±5 % and 46±5 % remaining, respectively. Similar to the **C4** series, little substrate remained for each of the labelled **C7TWJ** tetrahedra after incubation with 30 equivalents of hAGT.

Potentially, the less efficient unhooking observed may be attributed to reduced binding affinity of hAGT to the non-canonical tetrahedron structure itself or sequence composition. Moreover, it is plausible that the orientation of the α-carbon that is transferred from the O^6-atom of dG to the hAGT protein may not be in the optimal *syn* orientation at the vertex of the tetrahedron

for the reaction. Nevertheless, tetrahedral disassembly is significant, and represents an unusual nucleic acid substrate for hAGT. In addition, given the dynamic cellular concentration of hAGT and that it undergoes ubiquitination and degradation *in vivo* when alkylated, hAGT mediated disassembly of the DNA tetrahedron may be more efficient in the cell.[183,184,187]

3.3 Conclusions

Protein-DNA nanotechnology is a field which encompasses diverse applications from DNA amplification to FRET based detection systems.[188] In this work we exploit the specific activity of hAGT to trigger disassembly of a unique DNA tetrahedra containing an IaCL at the O^6-atom of dG. Given the intracellular presence of hAGT and its upregulation in certain cancer cells, harnessing it as a specific endogenous trigger for dismantling a DNA tetrahedron offers promise for drug delivery systems. It is envisioned that this modified DNA tetrahedron can be functionalized for the delivery of hydrophobic molecules and nucleic acid based therapeutic agents, building upon design elements reported for nucleic acid based drug carriers.[75,90] Furthermore, due to reports of DNA tetrahedra to enter cells without transfecting agents, we believe this may be an appropriate vehicle for drug delivery. The mechanism of cancer treatment could have dual purposes as the dG-dG IaCLs could act as suicide inhibitors of hAGT in chemotherapeutic resistant cells as well. Moreover, nuclear localization of the DNA tetrahedra is not required as hAGT can be found in the cytosol. The increased levels of hAGT in malignant cells, in conjunction with the ability of hAGT to unhook dG-dG intrastrand cross-links in DNA tetrahedra could function as a bifunctional therapeutic agent.

3.4 Acknowledgements

The authors would like to thank Dr. Anne Noronha for her assistance with solid-phase oligonucleotide synthesis. We are grateful to Dr. Anthony Pegg for providing the plasmid encoding the hAGT gene. This research was supported by a grant from the Natural Sciences and Engineering Research Council of Canada (NSERC). W.C. was the recipient of a Concordia Graduate Fellowship and is a trainee of the NSERC CREATE Programmed Molecules for Therapeutics, Sensing and Diagnostics program.

3.5 Supporting Information

3.5.1 Supporting Methods

3.5.1.1 General

5'-*O*-(4,4'-Dimethoxytrityl)-N^2-phenoxyacetyl-2'-deoxyguanosine and 2-cyanoethyl N,N-diisopropylchloro phosphoramidite were purchased from ChemGenes Corporation (Wilmington, MA). 5'-*O*-(4,4'-Dimethoxytrityl)-2'-deoxyribonucleoside-3'-*O*-(*β*-cyanoethyl-N,N-diisopropyl) phosphoramidites and protected 2'-deoxyribonucleoside-controlled pore glass (CPG) supports were acquired from Glen Research (Sterling, VA). Chemicals and solvents were obtained from MilliporeSigma (St. Louis, MO). Flash column chromatography silica gel 60 (230-400 mesh) was purchased from Silicycle (Québec City, QC). Thin layer chromatography was performed with precoated TLC plates (Merck, Kieselgel 60 F_{254}, 0.25 mm) obtained from MilliporeSigma (St. Louis, MO). NMR spectra were acquired on a Varian 500 MHz NMR spectrometer at room temperature and all chemical shifts were reported in ppm downfield from an internal standard. ^{1}H NMR spectra were recorded at a frequency of 500 MHz and ^{13}C spectra (^{1}H decoupled) at 125.7 MHz with tetramethylsilane as a reference. ^{31}P NMR spectra (^{1}H decoupled) were acquired at a frequency of 202.3 MHz with 85 % H_3PO_4 as an external reference. Acquisition of high-resolution mass spectrometry data of small molecules was performed using a 7T-LTQ FT ICR mass spectrometer from ThermoScientific (Waltham, MA) at the Concordia University Centre for Structural and Functional Genomics operated in positive ion scan mode. Most reagents for biochemical assays, PAGE and protein purification, were purchased from Bioshop Canada Inc (Burlington, ON). EDTA-free protease inhibitor cocktail tablets were obtained from Roche (Laval, QC) and Ni-NTA Superflow Resin from Qiagen (Mississauga, ON). T4 polynucleotide kinase was bought from Fermentas (Burlington, ON), and [γ-^{32}P]-ATP from PerkinElmer (Woodbridge, ON). DNA primers were acquired from Biocorp (Montreal, QC) and Phusion Polymerase from New England Biolabs (Ipswich, MA). Stratagene (Cedar Creek, TX) provided XL-10 Gold and Bl21 (DE3) *E. coli* cells. Mung Bean nuclease was purchased from New England BioLabs Inc. (Whitby, ON).

3.5.1.2 Chemical synthesis of nucleosides

5'-O-(4,4'-dimethoxytrityl)-3'-O-levulinoyl-N^2-phenoxyacetyl-2'-deoxyguanosine ***(3.1)***

To a solution of 5'-*O*-(4,4'-dimethoxytrityl)-N^2-phenoxyacetyl-2'-deoxyguanosine (1.7 g, 2.3 mmol), EDAC·HCl (1.1 g, 5.9 mmol) and DMAP (cat. 2 mg) in dioxane (33 mL) was added levulinic acid (0.685 g, 5.9 mmol) while stirring. After 18 h the solvent was removed *in vacuo* and the crude colorless gum was taken up in DCM (100 mL). The organic layer was washed with $NaHCO_3$ (*aq*, 3 % w/v) (2 x 75 mL), brine (1 x 75 mL), dried over anhydrous Na_2SO_4 (~4 g), decanted and the solvent removed *in vacuo*. The crude compound was purified by silica gel flash column chromatography using a gradient of MeOH:DCM (1 %, 1.25 %, 1.5 %, 2 %) to afford 1.38 g (73 %) of colorless foam. R_f (SiO_2, TLC): 0.32 MeOH:DCM (5%). $\lambda_{max(MeCN)}$ = 239 nm. ^{1}H NMR (500 MHz, $CDCl_3$, ppm): 11.76 (s, 1H, NH), 9.00 (s, 1H, NH), 7.80 (s, 1H, H8), 7.27 - 7.25 (m, 3H, Ar), 7.23 (m, 1H, Ar), 7.19 (m, 1H, Ar), 7.09 (m, 1H, Ar), 6.98 (m, 2H, Ar), 6.78 (m, 4H, Ar), 6.26 (dd, J = 8.4, 5.8 Hz, 1H, H1'), 5.52 (dd, J = 3.7, 2.2 Hz, 1H, H3'), 4.64 (s, 2H, $PhOCH_2CO$), 4.25 (m, 1H, H4'), 3.77 (s, 6H, 2 x OCH_3), 3.41 (dd, J = 10.4, 4.1 Hz, 1H, H5'), 3.32 (dd, J = 10.4, 3.6 Hz, 1H, H5'), 2.90 - 2.73 (m, 3H, H2' & CH_2 [lev]), 2.67 - 2.51 (m, 3H, H2' & CH_2 [lev]), 2.19 (s, 3H, CH_3). ^{13}C NMR (125.7 MHz, $CDCl_3$, ppm): 206.41, 172.13, 169.46, 158.59, 156.38, 155.24, 147.64, 146.15, 144.31, 137.17, 135.41, 135.31, 130.04, 130.01, 129.99, 127.97, 128.04, 127.90, 126.99, 86.73, 84.14, 83.72, 75.20, 67.06, 66.96, 63.57, 37.90, 37.83, 29.78, 27.92. HRMS (ESI-MS) *m/z* calculated for $C_{44}H_{44}N_5O_{10}^+$ $[M+H]^+$: 802.3088; found: 802.3096.

5'-O-(4,4'-dimethoxytrityl)-3'-O-levulinoyl-N^2-phenoxyacetyl-O^6-(4-(tert-butyldiphenylsilyloxy) butyl)-2'-deoxyguanosine ***(3.2a)***

To a solution of **3.1** (1.0 g, 1.2 mmol), 4-(*tert*-butyldiphenylsilyloxy)butan-1-ol (0.49 g, 1.2 mmol) and PPh_3 (0.50 g, 1.9 mmol) in dioxane (30mL) was added DIAD (366 µL, 1.9 mmol) while stirring on ice. The reaction was allowed to warm up to room temperature. After 18 h the solvent was removed *in vacuo* and the crude colorless gum was taken up in DCM (100 mL). The organic layer was washed with $NaHCO_3$ (*aq*, 3 % w/v) (2 x 75 mL), brine (1 x 75 mL), dried over anhydrous Na_2SO_4 (~4 g), decanted and the solvent removed *in vacuo*. The crude compound was purified by silica gel flash column chromatography using EtOAc:hexanes (6:4) to afford 0.91 g (66 %) of a colorless foam. R_f (SiO_2, TLC): 0.35 EtOAc:hexanes (7:3). $\lambda_{max(MeCN)}$ = 270 nm. ^{1}H

NMR (500 MHz, $CDCl_3$, ppm): 8.64 (s, 1H, NH), 7.99 (s, 1H, C8), 7.67 (m, 5H, Ar), 7.57 – 7.51 (m, 1H, Ar), 7.46 (m, 1H, Ar), 7.36 (m, 9H, Ar), 7.27 (m, 2H, Ar), 7.25 – 7.14 (m, 5H, Ar), 7.10 – 6.96 (m, 3H, Ar), 6.83 – 6.70 (m, 4H, Ar), 6.45 (d, *J* = 2.7 Hz, 1H, H1'), 5.54 (d, *J* = 6.1 Hz, 1H, H3'), 4.76 (s, 2H, $PhOCH_2CO$), 4.59 (t, *J* = 6.7 Hz, 2H, Ar-O-CH_2-), 4.27 (d, *J* = 1.7 Hz, 1H,H4'), 3.75 (s, 6H, 2 x OCH_3), 3.73 (t, *J* = 6.2 Hz, 2H, -CH_2-O-Si), 3.47 (d, *J* = 4.3 Hz, 1H, H5'), 3.37 (d, *J* = 3.5 Hz, 1H, H5'), 2.94 (ddd, *J* = 14.4, 8.5, 6.2 Hz, 1H, H2'), 2.77 (dd, *J* = 6.7, 2.8 Hz, 2H, CH_2 [lev]), 2.68 – 2.57 (m, 3H, H2' & CH_2 [lev]), 2.20 (s, 3H, CH_3 [lev]), 2.03 – 1.96 (m, 2H,-CH_2-), 1.76 (m, 2H, -CH_2-), 1.04 (s, 9H, Si-$(CH_3)_3$). ^{13}C NMR (125.7 MHz, $CDCl_3$, ppm) 206.28, 172.13, 161.22, 158.52, 152.49, 151.17, 144.41, 139.80, 135.54, 135.46, 133.89, 132.12, 132.04, 131.88, 130.02, 129.78, 129.52, 128.52, 128.42, 128.07, 127.85, 127.59, 126.91, 122.28, 114.93, 113.16, 86.69, 84.36, 84.00, 75.47, 68.01, 67.80, 63.86, 63.43, 55.17, 38.18, 37.84, 29.79, 28.93, 27.99, 26.84, 25.47, 19.19, -0.02. HRMS (ESI-MS) *m/z* calculated for $C_{64}H_{70}N_5O_{11}Si^+$ $[M+H]^+$: 1112.4841; found: 1112.4839.

5'-O-(4,4'-dimethoxytrityl)-3'-O-levulinoyl-N^2-phenoxyacetyl-O^6-(7-(tert-heptyldiphenylsilyloxy) heptyl)-2'-deoxyguanosine ***(3.2b)***

To a solution of **3.1** (0.750 g, 0.94 mmol), 7-(*tert*-butyldiphenylsilyloxy)heptyl-1-ol (0.55 g, 1.3 mmol) and PPh_3 (0.43 g, 1.6 mmol) in dioxane (33 mL) was added DIAD (320 µL, 1.6 mmol) dropwise while stirring on ice. After addition the reaction was allowed to come to room temperature. After 18 h the solvent was removed *in vacuo* and the crude colorless gum was taken up in DCM (100 mL). The organic layer was washed with $NaHCO_3$ (*aq*, 3 % w/v) (2 x 75 mL), brine (1 x 75 mL), dried over anhydrous Na_2SO_4 (~4 g), decanted and the solvent removed *in vacuo*. The crude compound was purified by silica gel flash column chromatography using EtOAc:hexanes (6:4) to afford 0.83 g (75 %) of a colorless foam. R_f (SiO_2, TLC): 0.57 EtOAc:hexanes (7:3). $\lambda_{max(MeCN)}$ = 270 nm. ^{1}H NMR (500 MHz, $CDCl_3$,ppm): 8.66 (s, 1H, NH), 7.99 (s, 1H, C8), 7.66 (m, 4H, Ar), 7.43 – 7.26 (m, 12H, Ar), 7.25 – 7.14 (m, 4H, Ar), 7.07 – 6.98 (m, 3H, Ar), 6.77 (dd, *J* = 8.9, 2.0 Hz, 4H, Ar), 6.45 (dd, *J* = 8.5, 5.8 Hz, 1H, H1') , 5.55 (d, J = 6.0 Hz, 1H, H3'), 4.77 (s, 2H, $PhOCH_2CO$), 4.57 (t, *J* = 6.7 Hz, 2H, Ar-O-CH_2-), 4.27 (m, 1H, H4'), 3.76 (s, 6H, 2 x OCH_3), 3.65 (t, J = 6.5 Hz, 2H, -CH_2-O-Si), 3.49 (dd, J = 10.3, 4.2 Hz, 1H, H5'), 3.36 (dd, J = 10.4, 3.4 Hz, 1H, H5'), 2.99 - 2.86 (m, 1H, H2'), 2.78 (td, J = 6.5, 3.2 Hz, 2H, CH_2 [lev]), 2.61 (m, 3H, H2' & CH_2 [lev]), 2.20 (s, 3H, CH_3 [lev]), 1.86 (m, 2H, -CH_2-), 1.62 -

1.31 (m, 8H, ,-CH_2-), 1.04 (s, 9H, Si-$(CH_3)_3$). ^{13}C NMR (125.7 MHz, $CDCl_3$, ppm): 206.32, 172.14, 161.28, 158.52, 152.46, 151.19, 144.41, 139.81, 135.54, 135.46, 134.13, 132.93, 132.11, 132.03, 131.92, 131.90, 130.02, 129.78, 129.46, 128.53, 128.43, 128.07, 127.85, 127.54, 126.91, 122.27, 114.92, 113.16, 86.68, 84.35, 84.02, 67.95, 63.93, 63.86, 55.17, 38.14, 37.84, 32.52, 29.78, 29.16, 28.80, 27.99, 26.87, 25.95, 25.73, 19.20. HRMS (ESI-MS) *m/z* calculated for $C_{67}H_{76}N_5O_{11}Si^+$ $[M+H]^+$: 1154.5311; found: 1154.5264.

5'-O-(4,4'-dimethoxytrityl)-3'-O-levulinoyl-N^2-phenoxyacetyl-O^6-(hydroxybutyl)-2'-deoxyguanosine ***(3.3a)***

To a solution of **3.2a** (0.89 g, 0.80 mmol) in THF (8.6 mL) was added TBAF (5.75 mL of 0.25 M in THF, 1.2 mmol) dropwise while stirring. After 1.5 h the reaction was taken up in DCM (100 mL) and the organic layer was washed with $NaHCO_3$ (*aq*, 3 % w/v) (2 x 75 mL), brine (1 x 75 mL), dried over anhydrous Na_2SO_4 (~4 g), decanted and the solvent removed *in vacuo*. The crude compound was purified by silica gel flash column chromatography using a gradient of MeOH:DCM (1 %, 1.25 %, 1.5 %, 1.75 %) to afford 0.47 g (68 %) of a colorless foam. R_f (SiO_2, TLC): 0.35 MeOH:DCM (5 %). $\lambda_{max(MeCN)}$ = 269 nm. ^{1}H NMR (500 MHz, $CDCl_3$, ppm): 8.72 (s, 1H, NH), 7.99 (s, 1H, C8), 7.41 – 7.31 (m, 4H, Ar), 7.28 – 7.26 (m, 2H, Ar), 7.25 – 7.16 (m, 5H, Ar), 7.08 – 6.93 (m, 3H, Ar), 6.77 (dd, *J* = 8.8, 1.6 Hz, 4H, Ar), 6.44 (dd, *J* = 8.4, 5.8 Hz, 1H, H1'), 5.54 (d, *J* = 6.1 Hz, 1H, H3'), 4.73 (s, 2H, $PhOCH_2CO$), 4.66 (t, *J* = 6.7 Hz, 2H, Ar-O-CH_2-), 4.27 (d, *J* = 1.8 Hz, 1H, H4'), 3.76 (s, *J* = 0.7 Hz, 6H, 2 x OCH_3), 3.73 (m, 2H, -CH_2-OH), 3.47 (dd, *J* = 10.4, 4.3 Hz, 1H, H5'), 3.36 (dd, *J* = 10.4, 3.4 Hz, 1H, H5'), 3.02 – 2.88 (m, 1H, H2'), 2.77 (m, 2H, CH_2 [lev]), 2.70 – 2.51 (m, 3H, H2' & CH_2 [lev]), 2.19 (s, 3H, CH_3 [lev]), 2.00 (m, 2H, -CH_2-), 1.77 (m, 2H, -CH_2-).^{13}C NMR (125.7 MHz, $CDCl_3$, ppm) 206.28, 172.12, 161.11, 158.52, 157.16, 152.50, 151.15, 144.38, 139.90, 135.50, 135.44, 130.01, 129.80, 128.06, 127.85, 126.91, 122.31, 118.79, 114.92, 113.16, 86.69, 84.36, 84.02, 75.43, 67.95, 67.58, 63.82, 62.23, 55.17, 38.13, 37.83, 29.78, 29.01, 27.97, 25.17, -0.03. HRMS (ESI-MS) *m/z* calculated for $C_{48}H_{52}N_5O_{11}^+$ $[M+H]^+$: 874.3663; found: 874.3644.

5'-O-(4,4'-dimethoxytrityl)-3'-O-levulinoyl-N^2-phenoxyacetyl-O^6-(hydroxyheptyl)-2'-deoxyguanosine ***(3.3b)***

To a solution of **3.2b** (0.70 g, 0.60 mmol) in THF (7 mL) was added TBAF (2.8 mL of 0.25 M in THF) dropwise while stirring. After 1.5 h the reaction was taken up in DCM (100 mL) and the organic layer was washed with $NaHCO_3$ (*aq*, 3 % w/v) (2 x 75 mL), brine (1 x 75 mL), dried over anhydrous Na_2SO_4 (~4 g), decanted and the solvent removed *in vacuo*. The crude compound was purified by silica gel flash column chromatography using a gradient of MeOH:DCM (0.75 % to 2 %) to afford 0.43 g (81 %) of a colorless foam. R_f (SiO_2, TLC): 0.55 (5 % MeOH:DCM). $\lambda_{max(MeCN)}$ = 269 nm. ^{1}H NMR (500 MHz, $CDCl_3$, ppm): 8.70 (s, 1H, NH), 7.99 (s, 1H, C8), 7.41 – 7.26 (m, 6H, Ar), 7.25 – 7.16 (m, 5H, Ar), 7.08 – 6.98 (m, 3H, Ar), 6.82 – 6.73 (m, 4H, Ar), 6.44 (dd, *J* = 8.4, 5.8 Hz, 1H, H1'), 5.54 (d, *J* = 6.0 Hz, 1H, H3'), 4.77 (s, 2H, $PhOCH_2CO$), 4.58 (t, *J* = 6.6 Hz, 2H, Ar-O-CH_2-), 4.26 (dd, *J* = 5.6, 3.9 Hz, 1H, H4'), 3.76 (t, *J* = 2.4 Hz, 6H, 2 x OCH_3), 3.63 (t, *J* = 6.0 Hz, 2H, -CH_2-OH), 3.48 (dd, *J* = 10.4, 4.3 Hz, 1H, H5'), 3.35 (dd, *J* = 10.4, 3.4 Hz, 1H, H5'), 2.94 (ddd, *J* = 14.4, 8.5, 6.2 Hz, 1H, H2'), 2.77 (m, 2H, CH_2 [lev]), 2.71 – 2.54 (m, 3H, H2' & CH_2 [lev]), 2.19 (s, 3H, CH_3 [lev]), 1.89 (m, 2H, -CH_2-), 1.66 – 1.46 (m, 6H, -CH_2-), 1.40 (m, 4H, -CH_2-).^{13}C NMR (125.7 MHz, $CDCl_3$, ppm): 206.29, 172.13, 161.25, 158.52, 152.45, 151.20, 144.40, 139.81, 135.52, 135.45, 130.01, 129.78, 128.07, 127.85, 126.91, 122.27, 114.92, 113.16, 86.68, 84.34, 84.01, 75.44, 68.00, 67.90, 63.84, 62.84, 55.17, 38.14, 37.84, 32.65, 29.78, 28.96, 28.53, 27.98, 25.86, 25.49, -0.03. HRMS (ESI-MS) *m/z* calculated for $C_{51}H_{58}N_5O_{11}^+$ $[M+H]^+$: 916.4133; found: 916.4134.

3'-O-alloxycarbonyl-N^2-phenoxyacetyl-2'-deoxyguanosine ***(3.4)***

To a solution of 5'-*O*-(4,4'-dimethoxytrityl)-N^2-phenoxyacetyl-2'-deoxyguanosine (1.0 g, 1.4 mmol) and DMAP (0.03 g, 0.1 mmol) in THF/pyridine (22.5/2.5 mL) was added allyl 1-benzo-6-chlorotriazolyl carbonate (0.55 g, 2.13 mmol) while stirring. After 18 h the solvent was removed *in vacuo* and the crude yellow gum was taken up in DCM (100 mL). The organic layer was washed with $NaHCO_3$ (*aq*, 3 % w/v) (2 x 75mL), brine (1 x 75mL), dried over anhydrous Na_2SO_4 (~4 g), decanted and the solvent removed *in vacuo*. To a solution of the crude product in DCM:MeOH (25:8 mL) was added *p*-TsOH•H_2O (0.89 g, 4.7 mmol) while stirring. After 30 minutes TEA (170 µL, 1.2 mmol) was added and total volume was brought to 100 mL with DCM. The organic layer was washed with $NaHCO_3$ (*aq*, 3 %) (2 x 75 mL), brine (1 x 75 mL), dried over anhydrous Na_2SO_4

(~4 g) and decanted. The crude yellow gum was purified by silica gel flash column chromatography using a gradient of MeOH:DCM (0 %, 1 %, 2 %, 3 %) to afford 0.45 g (67 % over two steps) of a yellow foam. R_f (SiO_2, TLC): 0.36 MeOH:DCM (5 %). $\lambda_{max(MeCN)}$ = 259 nm. ^{1}H NMR (500 MHz, $CDCl_3$, ppm): 7.81 (s, 1H, C8), 7.35 (m, 2H, Ar), 7.17 – 6.96 (m, 3H, Ar), 6.23 (dd, *J* = 9.6, 5.5 Hz, 1H, H1'), 5.95 (ddt, *J* = 16.4, 10.5, 5.9 Hz, 1H, allyl), 5.49 – 5.27 (m, 3H, allyl +H3'), 4.66 (m, 4H, vinylCH_2O + PhOCH_2CO), 4.33 (m, 1H, H4'), 3.95 (m, 2H, H5'), 3.12 – 2.90 (m, 1H, H2'), 2.53 (dd, *J* = 14.2, 5.5 Hz, 1H, H2'). ^{13}C NMR (125.7 MHz, $CDCl_3$, ppm) 169.17, 156.37, 154.19, 146.95, 146.88, 138.40, 131.05, 129.94, 123.07, 122.88, 119.59, 114.81, 86.60, 86.55, 79.49, 68.93, 66.67, 63.10, 45.86, 38.10, 10.36, -0.03. *m/z* calculated for $C_{22}H_{23}N_5NaO_8^+$ $[M+Na]^+$: 508.1444; found: 508.1434.

3'-O-alloxycarbonyl-5'-O-(tert-butyldimethylsilyl)-N^2-phenoxyacetyl-2'-deoxyguanosine ***(3.5)***

To a solution of **3.4** (1.1 g, 2.4 mmol), imidazole (0.65 g, 4.8 mmol) and DMAP (cat. 2 mg) in DCM (30 mL) was added TBS-Cl (0.72 g, 4.8 mmol) while stirring. After 18 h the solvent was removed *in vacuo* and the crude colorless powder was taken up in DCM (100 mL). The organic layer was washed with $NaHCO_3$ (*aq*, 3 % w/v) (2 x 75 mL), brine (1 x 75 mL), dried over anhydrous Na_2SO_4 (~4 g), decanted and the solvent removed *in vacuo*. The crude compound was purified by silica gel flash column chromatography using a gradient of MeOH:DCM (0 %, 0.5 %, 1 %, 1.5 %, 2 %) to afford 1.1 g (79 %) of a colorless powder. R_f (SiO_2, TLC): 0.4 MeOH:DCM (5 %). $\lambda_{max(MeCN)}$ = 252 nm. ^{1}H NMR (500 MHz, $CDCl_3$, ppm) 11.74 (s, 1H, NH), 9.02 (s, 1H, NH), 8.04 (s, 1H, C8), 7.38 (dd, *J* = 8.7, 7.5 Hz, 2H, Ar), 7.11 (m, 1H, Ar), 7.03 (dd, *J* = 8.7, 0.9 Hz, 2H, Ar), 6.34 (s, 1H, H1'), 5.96 (m, 1H, allyl), 5.41 (m, 1H, allyl), 5.37 – 5.33 (m, 1H, H3'), 5.33 – 5.26 (m, 1H, allyl), 4.70 (s, 2H, PhOCH_2CO), 4.68 (m, 2H, vinylCH_2O), 4.28 (d, *J* = 1.6 Hz, 1H, H4'), 3.89 (dd, *J* = 6.9, 2.6 Hz, 2H, H5'), 2.64 (m, 2H, H2'), 1.09 – 0.70 (s, 9H, Si-$(CH_3)_3$), 0.11 (s, *J* = 1.4 Hz, 6H, 2 x Si-CH_3). ^{13}C NMR (125.7 MHz, $CDCl_3$, ppm) 169.33, 156.30, 155.21, 154.23, 147.52, 146.12, 136.77, 131.04, 130.04, 123.09, 121.85, 119.65, 114.85, 85.48, 83.73, 78.93, 68.95, 66.85, 63.60, 39.00, 25.92, 18.36, -0.03, -5.40, -5.60. HRMS (ESI-MS) *m/z* calculated for $C_{28}H_{38}N_5O_8Si^+$ $[M+H]^+$: 600.2490; found: 600.2487.

1-{O^6-[5'-O-(4,4'-dimethoxytrityl)-3'-O-levulinoyl-N^2-phenoxyacetyl-2'-deoxyguanidinyl]}-4-{O^6-[3'-O-alloxycarbonyl-5'-O-(tert-butyldimethylsilyl)-N^2-phenoxyacetyl-2'-deoxyguanidinyl]}-butane ***(3.6a)***

To a solution of **3.3a** (0.46 g, 0.53 mmol), **3.5** (0.35 g, 0.59 mmol) and PPh_3 (0.14 g, 0.54 mmol) in dioxane (12 mL) was added DIAD (106 µL, 0.54 mmol) in dioxane (3 mL) dropwise while stirring on ice. The reaction was then allowed to warm up to room temperature. After 18 h to the reaction was added PPh_3 (0.070 g, 0.27 mmol) and DIAD (55 µL, 0.27 mmol) dropwise while stirring on ice. After a total of 24 h the solvent was removed *in vacuo* and the crude colorless gum was taken up in DCM (100 mL). The organic layer was washed with $NaHCO_3$ (*aq*, 3 % w/v) (2 x 75 mL), brine (1 x 75 mL), dried over anhydrous Na_2SO_4 (~4 g) decanted and the solvent removed *in vacuo*. The crude compound was purified by silica gel flash column chromatography using a gradient of EtOAc:hexanes (1:1, 6:4, 7:3, 8:2, 9:1) to afford 0.49 g (62 %) of a colorless foam. R_f (SiO_2, TLC): 0.13 EtOAc:hexanes (7:3). $\lambda_{max(MeCN)}$ = 269 nm. ^{1}H NMR (500 MHz, $CDCl_3$, ppm): 8.83 (s, 1H, NH), 8.75 (s, 1H, NH), 8.20 (s, 1H, C8), 8.00 (s, 1H, C8), 7.40 – 7.30 (m, 6H, C8), 7.27 (m, 2H, C8), 7.25 – 7.24 (m, 2H, C8), 7.24 – 7.19 (m, 2H, C8), 7.18 (d, *J* = 7.2 Hz, 1H, Ar), 7.01 (m, 6H, Ar), 6.80 – 6.70 (m, 4H, Ar), 6.49 (dd, *J* = 8.0, 6.3 Hz, 1H, H1'), 6.43 (dd, *J* = 8.4, 5.8 Hz, 1H, H1'), 6.02 – 5.87 (m, 1H, allyl), 5.53 (m, 1H, H3'), 5.40 (m, 1H, allyl), 5.37 – 5.34 (m, 1H, H3'), 5.31 (m, 1H, allyl)., 4.77 (s, 4H, 2 x $PhOCH_2CO$), 4.73 – 4.62 (m, 6H, vinylCH_2O, 2 x Ar-O-CH_2-), 4.28 (m, 2H, 2 x H4'), 3.92 (d, *J* = 2.7 Hz, 2H, 2 x H5'), 3.75 (d, *J* = 0.9 Hz, 6H, 2 x OCH_3), 3.48 (dd, *J* = 10.4, 4.4 Hz, 1H, H5'), 3.35 (dd, *J* = 10.4, 3.5 Hz, 1H, H5'), 2.94 (m, 1H, H2'), 2.83 – 2.69 (m, 4H, 2 x H2' + CH_2 [lev]), 2.68 – 2.56 (m, 3H, H2' & CH_2 [lev]), 2.19 (s, 3H, CH_3 [lev]), 2.12 (m, 4H, -CH_2-), 0.96 – 0.85 (s, 9H, $Si(CH_3)_3$), 0.10 (s,6H, 2 x Si-CH_3). ^{13}C NMR (125.7 MHz, $CDCl_3$, ppm): 206.25, 172.11, 161.11, 158.51, 154.32, 152.53, 144.41, 135.52, 135.46, 131.12, 130.01, 129.75, 128.06, 127.84, 126.89, 119.43, 114.91, 114.88, 113.15, 86.66, 85.47, 84.35, 84.04, 78.88, 75.46, 68.84, 67.06, 63.70, 55.17, 37.84, 29.78, 27.99, 25.94, 25.36, 18.36, -0.03, -5.39, -5.56. HRMS (ESI-MS) *m/z* calculated for $C_{76}H_{87}N_{10}O_{18}Si^+$ $[M+H]^+$:1455.5969; found: 1455.5969.

1-{O^6-[5'-O-(4,4'-dimethoxytrityl)-3'-O-levulinoyl-N^2-phenoxyacetyl-2'-deoxyguanidinyl]}-7-{O^6-[3'-O-alloxycarbonyl-5'-O-(tert-butyldimethylsilyl)-N^2-phenoxyacetyl-2'-deoxyguanidinyl]}-heptane ***(3.6b)***

To a solution of **3.3b** (0.28 g, 0.31 mmol), **3.5** (0.21 g, 0.36 mmol) and PPh_3 (0.080 g, 0.30 mmol) in dioxane (6.6 mL) was added DIAD (0.060 mL, 0.30 mmol) in dioxane (3 mL) dropwise while stirring on ice. The reaction was allowed to come to room temperature. After 18 h to the reaction was added PPh_3 (0.080 g, 0.30 mmol) and DIAD (0.060 mL, 0.30 mmol) dropwise while stirring on ice. After a total of 48 h the solvent was removed *in vacuo* and the crude colorless gum was taken up in DCM (100 mL). The organic layer was washed with $NaHCO_3$ (*aq*, 3 % w/v) (2 x 75 mL), brine (1 x 75 mL), dried over anhydrous Na_2SO_4 (~4 g), decanted and the solvent removed *in vacuo*. The crude compound was purified by silica gel flash column chromatography using a gradient of EtOAc:hexanes (1:1, 6:4, 7:3, 8:2) to afford 0.24 g (53 %) of a colorless foam. R_f (SiO_2, TLC): 0.19 EtOAc:hexanes (7:3). $\lambda_{max(MeCN)}$ = 269 nm. ^{1}H NMR (500 MHz, $CDCl_3$, ppm): 8.67 (s, 2H, 2 x NH), 8.18 (s, 1H, C8), 7.98 (s, 1H, C8), 7.34 (m, 6H, Ar), 7.26 (m, 2H, Ar), 7.25 – 7.14 (m, 6H, Ar), 7.08 – 6.96 (m, 6H, Ar), 6.76 (m, 4H, Ar), 6.50 (m, 1H, H1'), 6.44 (dd, *J* = 8.4, 5.9 Hz, 1H, H1'), 5.95 (ddt, *J* = 16.3, 10.4, 5.8 Hz, 1H, allyl), 5.54 (m, 1H, H3'), 5.40 (dd, *J* = 17.2, 1.4 Hz, 1H, allyl), 5.36 (m, 1H, H3'), 5.31 (dd, *J* = 10.4, 1.2 Hz, 1H, allyl), 4.76 (s, 4H, 2 x $PhOCH_2CO$), 4.66 (d, *J* = 5.8 Hz, 2H, vinylCH_2O), 4.58 (m, 4H, 2 x Ar-O-CH_2-), 4.35 – 4.22 (m, 2H, H4'), 3.91 (d, *J* = 2.7 Hz, 2H, 2 x H5'), 3.75 (s, 6H, 2 x OCH_3), 3.47 (m, 1H, H5'), 3.36 (m, 1H, H5'), 3.01 – 2.85 (m, 1H, H2'), 2.74 (m, 4H, 2 x H2' + CH_2 [lev]), 2.68 – 2.51 (m, 3H, H2' + CH_2 [lev]), 2.19 (s, 3H, CH_3 [lev]), 1.98 – 1.81 (m, 4H, 2 x -CH_2-), 1.52 (m, 6H, 3 x -CH_2-), 0.90 (s, 9H, $Si(CH_3)_3$), 0.10 (s, 6H, 2 x Si-CH_3). ^{13}C NMR (125.7 MHz, $CDCl_3$, ppm): 206.27, 172.12, 161.25, 158.51, 154.32, 152.47, 152.41, 144.40, 139.62, 135.50, 135.45, 131.11, 130.01, 129.78, 128.06, 127.84, 126.90, 122.24, 119.43, 114.91, 114.88, 113.15, 86.66, 85.44, 84.35, 84.00, 78.89, 75.47, 68.83, 67.98, 67.82, 63.71, 55.17, 38.97, 38.14, 37.83, 29.79, 29.09, 28.76, 27.98, 25.94, 25.86, 18.36, 14.18, -0.03, -5.39, -5.56. HRMS (ESI-MS) *m/z* calculated for $C_{79}H_{93}N_{10}O_{18}Si^+$ $[M+H]^+$:1497.6439; found: 1497.6439.

1-{O^6-[5'-O-(4,4'-dimethoxytrityl)-3'-O-levulinoyl-N^2-phenoxyacetyl-2'-deoxyguanidinyl]}-4-{O^6-[5'-O-(tert-butyldimethylsilyl)-N^2-phenoxyacetyl-2'-deoxyguanidinyl]}-butane ***(3.7a)***

To a solution of **3.6a** (0.49 g, 0.34 mmol), $Pd(PPh_3)_4$ (0.016 g, 0.013 mmol) and PPh_3 (0.023 g, 0.087 mmol) was added a sonicated solution of formic acid (28 µL, 0.74 mmol): *n*-butylamine (67 µL, 0.67 mmol) in THF (2 mL). The round bottomed flask was rinsed and sonicated with an additional 2 mL of THF. After 20 min, the solvent was removed *in vacuo* and the crude compound

was resuspended in DCM (100 mL). The organic layer was washed with $NaHCO_3$ (*aq*, 3 % w/v) (2 x 75 mL), brine (1 x 75 mL), dried over anhydrous Na_2SO_4 (~4 g), decanted and the solvent removed *in vacuo*. The crude compound was purified by silica gel flash column chromatography using a gradient of EtOAc:hexanes:MeOH (8:2:0, 9:1:0, 10:0:0, 9:0:1) to afford 0.34 g (74 %) of a colorless foam. R_f (SiO_2, TLC): 0.088 EtOAc (100). $\lambda_{max(MeCN)}$ = 269 nm. ^{1}H NMR (500 MHz, $CDCl_3$, ppm): 8.85 (s, 1H, NH), 8.76 (s, 1H, NH), 8.17 (s, 1H, C8), 7.99 (s, 1H, C8), 7.38 – 7.26 (m, 7H, Ar), 7.25 – 7.14 (m, 5H, Ar), 7.11 – 6.96 (m, 6H, Ar), 6.79 – 6.73 (m, 4H, Ar), 6.55 (t, 1H, H1'), 6.43 (dd, *J* = 8.3, 5.9 Hz, 1H, H1'), 5.54 (d, *J* = 6.1 Hz, 1H, H3'), 4.95 – 4.72 (m, 5H, H3' + 2 x $PhOCH_2CO$), 4.69 (m, 4H, 2 x Ar-O-CH_2-), 4.26 (m, 1H, H4'), 4.11 (m, 1H, H4'), 3.99 – 3.82 (m, 2H, 2 x H5'), 3.75 (s, 6H, 2 x OCH_3), 3.47 (m, 1H, H5'), 3.35 (m, 1H, H5'), 2.94 (m, 1H, H2'), 2.82 – 2.51 (m, 8H,3 x H2', 2 x CH_2 [lev]), 2.19 (s, 3H, s, 3H, CH_3 [lev]), 2.12 (m, 4H, -CH_2-), 0.89 (s, 9H, $Si(CH_3)_3$), 0.08 (s, 6H, 2 x Si-CH_3). ^{13}C NMR (125.7 MHz, $CDCl_3$, ppm): 206.30, 172.13, 161.09, 161.03, 158.51, 152.49, 152.32, 144.41, 135.52, 135.46, 130.01, 129.78, 129.75, 128.06, 127.84, 126.89, 122.18, 114.90, 113.15, 87.31, 86.66, 84.34, 84.17, 84.04, 75.46, 72.55, 68.00, 67.13, 63.88, 63.74, 55.17, 41.19, 38.11, 37.84, 29.78, 27.99, 25.94, 25.34, 18.39, -0.03, -5.38, -5.51. HRMS (ESI-MS) *m/z* calculated for $C_{72}H_{83}N_{10}O_{16}Si^+$ [M+H$^+$]: 1371.5758; found: 1371.5751.

1-{O^6-[5'-O-(4,4'-dimethoxytrityl)-3'-O-levulinoyl-N^2-phenoxyacetyl-2'-deoxyguanidinyl]}-7-{O^6-[5'-O-(tert-butyldimethylsilyl)-N^2-phenoxyacetyl-2'-deoxyguanidinyl]}-heptane ***(3.7b)***

To a solution of **3.6b** (0.24 g, 0.16 mmol), $Pd(PPh_3)_4$ (0.0074 g, 0.0064 mmol) and PPh_3 (0.011 g, 0.042 mmol) was added a sonicated solution of formic acid (12 µL, 0.32 mmol) : butylamine (32 µL, 0.32 mmol) in THF (2 mL). The round bottomed flask was rinsed and sonicated with an additional 2 mL of THF. After 20 min, the solvent was removed *in vacuo* and the crude compound was resuspended in DCM (100 mL). The organic layer was washed with $NaHCO_3$ (*aq*, 3 % w/v) (2 x 75 mL), brine (1 x 75 mL), dried over anhydrous Na_2SO_4 (~4 g), decanted and the solvent removed *in vacuo*. The crude compound was purified by silica gel flash column chromatography using a gradient of EtOAc:hexanes (1:1, 6:4, 7:3. 8:2, 9:1, 10:0) to afford 0.34 g (62 %) of a colorless foam. R_f (SiO_2, TLC): 0.15 EtOAc (100 %). $\lambda_{max(MeCN)}$ = 269 nm. ^{1}H NMR (500 MHz, $CDCl_3$, ppm): 8.77 (s, 1H, NH), 8.67 (s, 1H, NH), 8.16 (s, 1H, C8), 7.98 (s, 1H, C8), 7.40 – 7.31 (m, 6H, Ar), 7.26 (m, 2H, Ar), 7.25 – 7.14 (m, 5H, Ar), 7.08 – 6.95 (m, 6H, Ar), 6.76 (m, 4H, Ar),

6.58 (m, 1H, H1'), 6.44 (dd, *J* = 8.4, 5.8 Hz, 1H, H1'), 5.53 (m, 1H, H3'), 4.74 (m, 5H, H3' + 2 x PhOCH$_2$CO), 4.57 (t, *J* = 6.2 Hz, 4H, 2 x Ar-O-CH$_2$-), 4.26 (m, 1H, H4'), 4.15 – 4.05 (m, 2H, H4'), 3.86 (m, 2H, 2 x H5'), 3.75 (s, 6H, 2 x OCH$_3$), 3.47 (m, 1H, H5'), 3.36 (m, 1H, H5'), 2.93 (m, 1H, H2'), 2.76 (m, 3H, H2' + CH$_2$ [lev]), 2.62 (m, 4H, 2 x H2' + CH$_2$ [lev]), 2.19 (s, 3H, CH$_3$ [lev]), 1.96 – 1.81 (m, 4H, -CH$_2$-), 1.59 – 1.38 (m, 6H, -CH$_2$-), 0.89 (s, 9H, Si(CH$_3$)$_3$), 0.08 (s, 6H, 2 x Si-CH$_3$). ^{13}C NMR (125.7 MHz, CDCl$_3$, ppm): 206.29, 172.12, 161.25, 161.20, 158.51, 152.46, 152.28, 151.17, 144.40, 140.19, 139.79, 135.51, 135.46, 130.01, 129.80, 129.77, 128.06, 127.84, 126.90, 122.25, 114.91, 113.15, 87.37, 86.67, 84.34, 84.17, 84.00, 75.46, 72.53, 67.99, 67.84, 67.75, 63.86, 63.76, 55.17, 41.27, 38.14, 37.83, 29.78, 29.05, 28.74, 27.98, 25.95, 25.84, 18.40, 14.18, -0.03, -5.38, -5.51. HRMS (ESI-MS) *m/z* calculated for $C_{75}H_{89}N_{10}O_{16}Si^+$ $[M+H]^+$:1413.6227; found: 1413.6208.

1-{O^6-[5'-O-(4,4'-dimethoxytrityl)-3'-O-levulinoyl-N^2-phenoxyacetyl-2'-deoxyguanidinyl]}-4-{O^6-[5'-O-(tert-butyldimethylsilyl)-3'-O-(2-cyanoethoxy(diisopropylamino)-phosphino)-N^2-phenoxyacetyl-2'-deoxyguanidinyl]}-butane ***(3.8a)***

To a solution of **3.7a** (0.30 g, 0.22 mmol) and DIPEA (68 µL, 0.39 mmol) in DCM (2 mL) was added Cl-POCEN*i*Pr$_2$ (68 µL, 0.26 mmol) dropwise while stirring. After 45 min, the solvent was removed *in vacuo* and the crude compound was resuspended in EtOAc (50 mL). The organic layer was washed with NaHCO$_3$ (*aq*, 3 % w/v) (2 x 35 mL), brine (1 x 35 mL), dried over anhydrous Na$_2$SO$_4$ (~2 g), decanted and the solvent removed *in vacuo*. The crude compound was purified by silica gel flash column chromatography (conditioned with 0.5 % TEA:EtOAc) using MeCN:EtOAc (1:9) to afford 0.23 g (67 %) of a colorless foam. R_f (SiO$_2$, TLC): 0.29, 0.36 EtOAc (100 %). $\lambda_{max(MeCN)}$ = 269 nm. ^{1}H NMR (500 MHz, d$_6$-acetone, ppm): 9.44 (d, *J* = 2.3 Hz, 1H), 9.34 (s, 1H), 8.25 (d, *J* = 5.3 Hz, 1H), 8.21 (s, 1H), 7.39 (dd, *J* = 8.2, 1.4 Hz, 2H), 7.34 – 7.23 (m, 8H), 7.22 – 7.10 (m, 3H), 7.04 – 6.99 (m, 4H), 6.99 – 6.91 (m, 2H), 6.82 – 6.74 (m, 2H), 6.72 – 6.65 (m, 2H), 6.47 (dd, *J* = 8.4, 5.5 Hz, 2H), 5.57 – 5.47 (m, 1H), 5.10 (s, 2H), 5.04 (s, 2H), 4.84 (ddd, *J* = 12.9, 7.7, 4.6 Hz, 1H), 4.72 (d, *J* = 4.7 Hz, 4H), 4.33 – 4.27 (m, 1H), 4.23 (t, *J* = 8.0 Hz, 1H), 4.19 – 4.12 (m, 1H), 4.02 – 3.78 (m, 4H), 3.74 (d, *J* = 2.3 Hz, 6H), 3.73 – 3.58 (m, 3H), 3.43 – 3.33 (m, 1H), 3.31 (dd, *J* = 10.3, 3.3 Hz, 1H), 3.08 – 2.93 (m, 1H), 2.90 (t, *J* = 6.0 Hz, 1H), 2.83 – 2.67 (m, 6H), 2.65 – 2.48 (m, 4H), 2.13 – 2.08 (m, 4H), 1.35 – 1.10 (m, 15H), 0.97 – 0.85 (m, 9H), 0.14 – 0.04 (m, 6H). ^{13}C NMR (125.7 MHz, d$_6$-acetone, ppm): 171.87, 160.88, 158.60,

158.52, 158.23, 152.75, 152.69, 151.54, 145.27, 141.39, 135.88, 135.75, 130.10, 129.89, 129.42, 129.40, 128.02, 127.50, 126.51, 121.17, 121.13, 118.70, 114.74, 114.69, 112.83, 112.77, 86.03, 84.98, 84.74, 84.30, 75.30, 67.92, 66.84, 66.73, 64.64, 63.36, 63.20, 58.65, 54.56, 43.15, 43.11, 43.05, 43.01, 37.34, 36.02, , 25.46, 25.45, 25.33, 25.25, 24.05, 24.01, 23.98, 23.96, 18.04, -6.06, -6.08, -6.11. ^{31}P NMR (202.3 MHz, d_6-acetone, ppm) 148.25, 148.14. HRMS (ESI-MS) *m/z* calculated for $C_{81}H_{100}N_{12}O_{17}PSi^+$ $[M+H]^+$:1571.6836; found: 1571.6832.

1-{O^6-[5'-O-(4,4'-dimethoxytrityl)-3'-O-levulinoyl-N^2-phenoxyacetyl-2'-deoxyguanidinyl]}-7-{O^6-[5'-O-(tert-butyldimethylsilyl)-3'-O-(2-cyanoethoxy(diisopropylamino)-phosphino)-N^2-phenoxyacetyl-2'-deoxyguanidinyl]}-heptane ***(3.8b)***

To a solution of **3.7b** (0.31 g, 0.22 mmol) and DIPEA (57 µL, 0.33 mmol) in DCM (2 mL) was added Cl-POCEN*i*Pr$_2$ (58 µL, 0.26 mmol) dropwise while stirring. After 30 min, the solvent was removed *in vacuo* and the crude compound was resuspended in EtOAc (75 mL). The organic layer was washed with $NaHCO_3$ (*aq*, 3 % w/v) (2 x 50 mL), brine (1 x 50 mL), dried over anhydrous Na_2SO_4 (~4 g), decanted and the solvent removed *in vacuo*. The crude compound was purified by silica gel flash column chromatography (conditioned with 0.5 % TEA:EtOAc) using MeCN:EtOAc (1:9) to afford 0.25 g (70 %) of a colorless foam. R_f (SiO_2, TLC): 0.59, 0.64 EtOAc (100 %). $\lambda_{max(MeCN)}$ = 269 nm. ^{1}H NMR (500 MHz, d_6-acetone) δ 9.37 (2x s, 2H), 9.30 (s, 2H)f, 8.25 (s, 1H), 8.24 (s, 1H), 8.20 (s, 2H), 7.40-7.38 (m, 4H), 7.35 – 7.24 (m, 15H), 7.22 – 7.10 (m, 6H), 7.02 (m, 7H), 7.01-7.00 (m, 4H), 6.79 – 6.73 (m, 4H), 6.73 – 6.67 (m, 4H), 6.49-6.45 (m, 4H), 5.59 – 5.46 (m, 2H), 5.09 (s, 4H), 5.04 (s, 3H), 4.90 – 4.79 (m, 2H), 4.62-4.57 (m, 8H), 4.31 – 4.29 (m, 2H), 4.25-4.22 (m, 1H), 4.16-4.14 (m, 1H), 4.05 (q, *J* = 7.1 Hz, 4H), 4.02 – 3.80 (m, 8H), 3.74 (s, 5H), 3.74 (s, 5H), 3.71 – 3.60 (m, 5H), 3.47 – 3.36 (m, 2H), 3.31 (dd, *J* = 10.3, 3.3 Hz, 2H), 3.10 – 2.92 (m, 2H), 2.85 – 2.73 (m, 10H), 2.59 – 2.54 (m, 5H), 2.13 (s, 6H), 2.05 (dt, *J* = 4.4, 2.2 Hz, 16H), 1.96 (s, 7H), 1.94 – 1.83 (m, 8H), 1.57-1.51 (m, 12H), 1.29 – 1.16 (m, 32H), 0.90(d, J = 6.8 Hz, 17H), 0.09-0.07 (d, J = 9.1 Hz, 11H). ^{13}C NMR (125.7 MHz, d_6-acetone, ppm): 160.99, 158.61, 158.52, 145.27, 141.28, 135.88, 135.76, 130.11, 129.90, 129.43, 129.41, 128.03, 127.49, 126.51, 121.21, 114.74, 114.69, 112.82, 112.76, 86.03, 84.95, 84.75, 75.31, 67.90, 67.21, 67.11, 64.65, 54.55, 43.00, 37.33, 36.01, 25.70, 25.67, 25.46, 25.45, 24.01, 23.98, 23.95, 19.91, 13.59, -6.08. ^{31}P NMR (202.3 MHz, d_6-acetone, ppm) 148.24, 148.12. HRMS (ESI-MS) *m/z* calculated for $C_{84}H_{106}N_{12}O_{17}PSi^+$ $[M+H]^+$:1613.7306; found: 1613.7306.

3.5.1.3 Synthesis, purification and characterization of three-way junction oligonucleotides

The solid-phase synthesis of the three-way junction oligonucleotides was performed using an ABI Model 3400 synthesizer on a 1.5 μmol scale using standard β-cyanoethylphosphoramidite chemistry with polystyrene solid-supports from ChemGenes (Wilmington, MA). The DNA tetrahedron of von Kiedrowski, containing phenyl rings at the vertices, was assembled in a 5' to 3' direction with 5'-*O*-phosphoramidites. Modification of the directionality and location of the oligonucleotide sequences in the IaCL containing three-way junction was necessary relative to the design of von Kiedrowski. Therefore, the design of the IaCL containing three-way junction (see **Section 3.5.2 Scheme S3.1**) and that of von Kiedrowski is that oligonucleotide assembly of the former proceeds entirely in a 3' to 5' direction with 3'-*O*-phosphoramidites. Finally, in the tetrahedron design of von Kiedrowski, a 2'-deoxycytidine (dC) residue was adjacent to the phenyl ring at the vertices. Given that the IaCL is composed of two cross-linked dG residues, a C to G substitution was performed. Concentrations were 0.1 M and 0.15 M (in MeCN) for standard 3'-*O*-2'-deoxyphosphoramidites and compounds **3.8a** or **3.8b**, respectively. Coupling times of compounds **3.8a** and **3.8b** were increased to 10 minutes to maximize coupling efficiency. The 5'-termini of the oligonucleotides were coupled 5'-*O*-levulinoyl-3'-*O*-2'-deoxyphosphoramidites for 3 minutes. DMT groups were removed over a period of 45 seconds using 3 % trichloroacetic acid (in dichloromethane). Phenoxyacetic anhydride was used as the capping agent. Orthogonal chain extension was achieved by removal of the cyanoethyl groups from the support bound oligomers with 1 mL of anhydrous triethylamine (TEA) for a minimum of 12 h. The solid supports were then washed with 30 mL of MeCN followed by 30 mL of THF and then dried *via* high vacuum for a minimum of 30 minutes. The 5'-*O*-TBS group was then removed by treating the solid supports with 2 x 1 mL of triethylamine trihydrofluoride (TEA·3HF) for 2 x 30-minute increments. The supports were then washed with 30 mL of THF and 30 mL of MeCN followed with drying by high vacuum for 18 h. The deprotected support bound oligonucleotides were extended orthogonally using 3'-*O*-2'-deoxyphosphoramidites with coupling times of 4.5 minutes for **C4TWJs** and 3.5 minutes for **C7TWJs** with no further modifications to the oligonucleotide synthesis methodology. Cleavage of the three-way junctions from the polystyrene support and simultaneous deprotection was achieved with 1 mL of NH_4OH (28 %):EtOH (3:1) for a minimum of 4 h at 55°C. The supernatant was transferred and the polystyrene supports were washed with 2 x 200 μL of 50 % MeCN then dried down with a speed vacuum concentrator. The quality of synthesis was assessed

by loading ~0.1 OD_{260} of crude sample on a 20 % (19:1 acrylamide:bisacrylamide) denaturing PAGE run at 200 V for 45 minutes and visualized by UV shadowing. Three-way junctions were purified by preparatory denaturing PAGE. The preparatory denaturing PAGE consisted of a 20 % acrylamide solution (19:1 acrylamide:bis-acrylamide) in 1× TBE running buffer on standard 16 x 18 cm glass plates at 450 V until sufficient separation was achieved (approximately 4 h). The desired band was excised and the pure oligonucleotide extracted with 8-10 mL of 0.1 M sodium acetate (NaOAc) solution on a mixer overnight. The purified oligonucleotides were then desalted with a C-18 SEP PAK cartridge. The cartridge was prepared by washing with 10 mL of HPLC grade MeCN, followed by 50 % MeCN (*aq*) and equilibrated with 0.1 M NaOAc. The three-way junctions were adsorbed to the C-18 column then the salt was removed with 2 x 10 mL of H_2O. The sample was eluted with MeOH:H_2O:MeCN (2:1:1). Purity was assessed to be greater than 90% for all oligonucleotides as determined by analytical denaturing PAGE. All oligonucleotides were quantitated using a Varian Cary Model 3E spectrophotometer. Single strand concentrations were calculated using the Beer-Lambert law and the absorbance was measured at 260 nm. Molar extinction coefficients were calculated by the nearest neighbour approximation.[189] The identity of the oligonucleotides were confirmed by mass spectrometry at the Concordia University Centre for Biological Applications of Mass Spectrometry (see **Section 3.5.2 Table S3.1**). 0.15 OD of oligonucleotide was dried down for ESI-qTOF MS analysis on a Micromass qTOF Ultima API. The mass spectrometer was run in full scan, negative ion detection mode.

3.5.1.4 UV thermal denaturation studies

300 pmol of each three-way junction (either **C4** or **C7** series) was dried down, resuspended in 1 mL of TAMg buffer (40 mM Tris, 30 mM acetic acid, 20 mM $MgCl_2$ pH 7.6) to a final concentration of 3 μM of complex, heated at 95 °C for 5 minutes, slowly cooled to room temperature and incubated at 4 °C overnight. The samples were then degassed on the speed-vacuum concentrator for 2 minutes. UV thermal denaturation data was acquired with a Varian CARY Model 3E Spectrophotometer fitted with a 12-sample thermostated cell block and a temperature controller, with absorbance measured at 260 nm and a heating rate at of 0.5 °C/min. The melting temperature (T_m) was calculated as the maximum of the first derivative according to the method of Puglisi and Tinoco.[189] All data analysis was performed with Microsoft Excel™.

3.5.1.5 Circular dichroism spectroscopy

Circular dichroism spectra were acquired at the Concordia University Integrated Platform for Biomolecular Function, Interactions and Structure (BIOFINS) using a Jasco J-815 spectropolarimeter equipped with a Julaba F25 circulating bath. Samples were prepared by drying down 300 pmol of each three-way junction (either **C4** or **C7** series) which was then dissolved in 1 mL of TAMg buffer (40 mM Tris, 30 mM acetic acid, 20 mM $MgCl_2$ pH 7.6) to a final concentration of 3 µM of complex. The samples were heated on a sand bath to 95 °C for 5 minutes, allowed to cool to room temperature and stored at 4 °C overnight. The samples were transferred to quartz cells (Starna 29-Q-10) and spectra were recorded from 220 to 350 nm at 15 °C. Each spectrum was an average of 5 scans, collected at a rate of 50 nm min^{-1} with a 1 nm bandwidth and a 0.2 nm sampling wavelength. The molar ellipticity (φ) was calculated by the equation $\varphi = \varepsilon/Cl$, where ε is the relative ellipticity (mdeg), C is the oligonucleotide concentration (moles/L), and l is the cell path length (cm). The data were processed using the software provided by Jasco and transferred to Microsoft Excel™ to prepare the spectrum.

3.5.1.6 Mung bean nuclease (MBN) digest

150 pmol of three-way junctions (individual oligonucleotides or all four combined) were annealed in 90 µL TAMg buffer by heating at 90°C for 5 minutes, then cooled to room temperature over 18 h. To each solution was added 10 units of MBN and 9 µL of 10 x mung bean buffer, and samples were incubated at 37 °C for 18 h. Each sample was then evaporated to dryness on a speed vacuum concentrator and resuspended in 10 µL of formamide. Samples were then loaded on a 20 % denaturing PAGE (19:1 acrylamide:bis-acrylamide) in 1 x TBE run at 200 V for 45 minutes. Gels were cast on 7.5 cm x 10 cm plates and were visualized by UV shadowing.

3.5.1.7 Native PAGE of TWJs

5 pmol of each ^{32}P-radiolabelled three-way junction oligonucleotide were added in every possible combination to a total volume of 200 µL of TAMg buffer (40 mM Tris-HCl, 30 mM acetic acid, 20 mM $MgCl_2$, pH 7.6) in a 200 µL PCR tube. Samples were heated to 90 °C for 5 minutes and cooled to 25 °C overnight. 10 % (19:1 acrylamide: bis-acrylamide) non-denaturing PAGE were cast on standard 16 x 18 cm glass plates and gels were prerun at 200 V for 20 minutes at 4 °C. 5

μL of samples were added to 5 μL of loading buffer (20 % glycerol, 1x TAMg) before being loaded on the gel. Gels were run at 200 V for 2.5 h at 4 °C before being processed as described above.

3.5.1.8 Protein purification and overexpression

The hAGT protein was purified and overexpressed as described previously.[36] The mass of the hAGT protein was verified by ESI-MS run in positive ion mode on a Micromass Q-Tof Ultima API.

3.5.1.9 Denaturing PAGE AGT unhooking assays of ssTWJs

200 pmol of the ss**TWJ**s were incubated with 2 μL of γ-[^{32}P]-ATP (10μCi/μL), 10 units of T4 PNK in 20 μL of 1x PNK buffer for 1 hour at 37 °C for radiolabelling. After 1 h of incubation the samples were boiled for 10 mins, and 30 μL of 18 MΩ H_2O was added to give stock concentrations of 4 μM. Repair of ss**TWJ**s was examined in duplicate by incubating 2 pmol of DNA with either 0, 20 or 60 pmol of hAGT in 15 μL activity buffer (10 mM Tris–HCl, 100 mM NaCl, 1 mM DTT, pH 7.6) at 37 °C for 18 h, after which the reactions were quenched by adding 18.2 μL of stop buffer [81 mM Tris–HCl, 81 mM boric acid, 1.8 mM EDTA and 1% SDS (sodium dodecyl sulfate) in 80 % formamide] and boiling for 5 minutes. The samples were loaded on a 16 cm × 18 cm, 20 % (19:1 acrylamide:bis-acrylamide), 7 M urea PAGE which was run for 1-1.5 h at 450 V in 1x TBE. Visualization of the gels was achieved by exposure to a storage phosphor screen, images captured on a Typhoon 9400 (GE Healthcare, Piscataway, NJ) and data was analyzed with ImageJ (National Institutes of Health, Bethesda, MD). Relative quantities of substrate and repair products were calculated by summing the integrated bands in each lane and dividing the corresponding band by the total.

3.5.1.10 Denaturing PAGE AGT unhooking assays of tetrahedron samples

5 pmol of each ^{32}P radiolabelled oligonucleotide was annealed with 5 pmol of the other three non-radiolabelled oligonucleotides as described above. 50 μL (2 pmol) was then aliquoted to which either 10- or 30-fold of hAGT was added. Samples were incubated at 37 °C for 18 h, after which reactions were quenched by the addition of 50 μL of stop buffer (81 mM Tris-HCl, 81 mM boric acid, 1.8 mM EDTA and 1% SDS (sodium dodecyl sulfate) in 80 % formamide) and loaded on a 16 cm × 18 cm, 20 % (19:1 acrylamide:bis-acrylamide), 7 M urea PAGE which was run for 1-1.5

h at 450 V in 1x TBE. Gels were imaged and quantified as described above. Integrity of the DNA tetrahedra was verified by adding 20 µL of protein activity buffer to control samples, incubated at 37 °C for 18 h, and electrophoresed in non-denaturing conditions as described above (see **Section 3.5.2 Figure S3.39**).

3.5.1.11 Preliminary serum stability assays

Samples were prepared with 10 µL of FBS, 60 µL of DMEM and 20 µL of either ss**TWJ** or the DNA tetrahedra (to give 5 nM C_T). Aliquots were removed at defined time intervals and quenched by boiling for 5 minutes. It should be noted that solubility issues were observed thus some lanes are void of sample. As depicted in **Figure S3.41**, the ss**TWJ**s for both the **C4** and **C7** series were well digested after 4-24 h. However, looking at the TDN series, it appears to be mostly intact after 24 h. Future optimization of this assay will involve quenching on dry ice, and supplementing samples with Proteinase K.

3.5.1.12 Unhooking of TDNs as a function of time

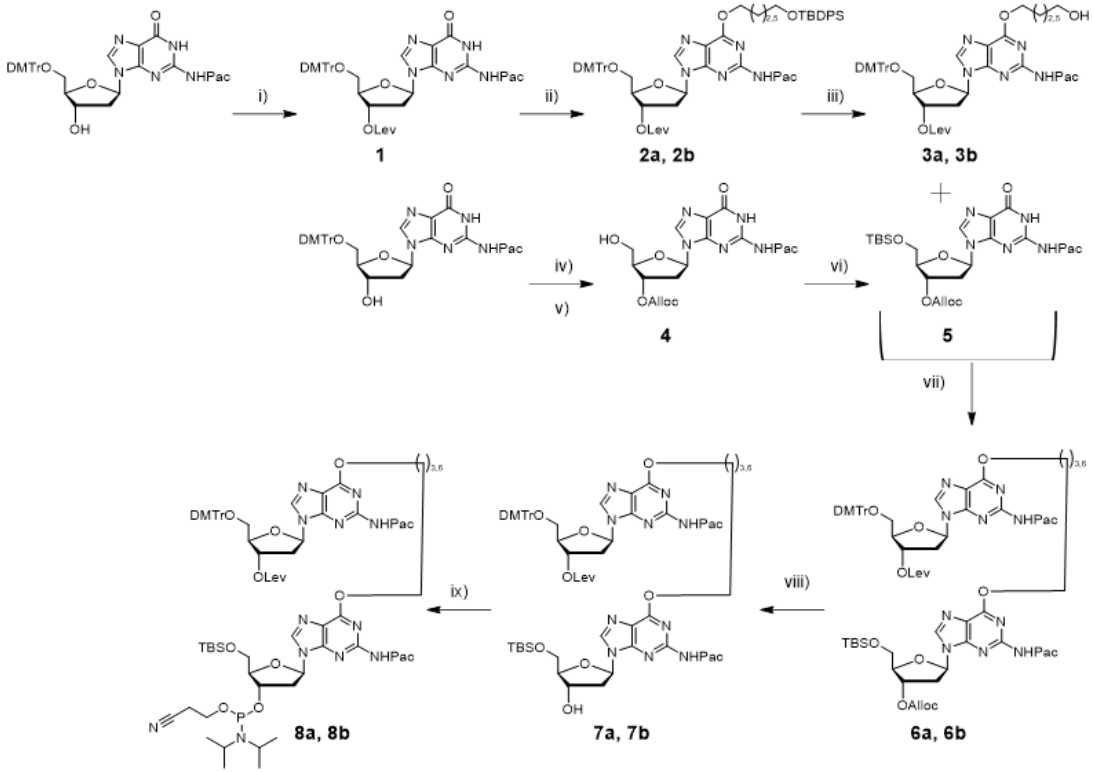
We also looked at unhooking of TDNs as a function of time (**Figure S3.42**). As opposed to the main text assays, all vertices were radiolabeled for these experiments. While the butylene series did not display a particularly smooth curve, both series appear to have reached a maximum repair around the 120 minute mark. This is more sluggish than when in a linear sequence (maximum repair reached before 60 minutes) however the butylene and heptylene linkages are comparable in the case of the DNA tetrahedra: which was not the case with linear series.

3.5.1.13 Magnesium dependence of DNA tetrahedra formation

The magnesium dependence of DNA tetrahedra dependence was examined. As shown in **Figure S3.43**, DNA tetrahedra formation appears to be quantitative at concentrations above 5 mM $MgCl_2$ for both **C4** and **C7** TDNs. At 0 mM $MgCl_2$ a substantial amount appears to not be leaving the well which is likely due to aggregation, and at 2 mM most DNA tetrahedra seem to be not aggregating.

3.5.2 Supporting Figures, Tables and Schemes

Scheme S3.1 Synthesis of compounds **8a** and **8b**. i) Lev-OH, EDAC-HCl, DMAP, dioxane, r.t., 18 h, 73 % 1 ii) 4-(tert-butyldiphenylsilyloxy)butan-1-ol (**a-series** compounds) or 7-(tert-butyldiphenylsilyloxy)heptan-1-ol (**b-series** compounds), DIAD, PPh_3, dioxane, r.t., 18 h, 66 % **2a**, 75% **2b** iii) TBAF (0.25 M in THF), THF, r.t., 1.5 h, 68 % **3a**, 81 % **3b** iv) Alloc-OBt-Cl, DMAP, THF:Py (9:1) v) p-TsOH, MeOH:DCM (1:3), r.t., 30 min, 67 % **4** vi) TBS-Cl, Im, DMAP, DCM, r.t., 16 h 79 % 5 vii) DIAD, PPh_3, dioxane, r.t., 18 h, 62% **6a**, 53 % **6b** viii) $Pd(PPh_3)_4$, PPh_3, HCOOH:$C_4H_9NH_2$ (1:1), THF, r.t., 74 % **7a**, 62 % **7b** ix) Cl-P(OCE)N*i*Pr_2, DIPEA, DCM, r.t., 30-45 min, 67 % **8a**, 70% **8b**.



Scheme S3.2 Orthogonal approach of synthesizing **TWJ1**. All **TWJs** were synthesized in identical manner. *i*) coupling of tripartite dG-O^6-alkylene-O^6-dG phosphoramidites, *ii*) chain extension and capping with a 5'-*O*-levulinoyl protected nucleoside, *iii*) removal of internal 5'-*O*-TBS with TEA·3HF, *iv*) orthogonal extension of third arm with standard 3'-*O*-phosphoramidites and *v*) simultaneous cleavage from solid support and deprotection with NH_3(*aq*):EtOH.

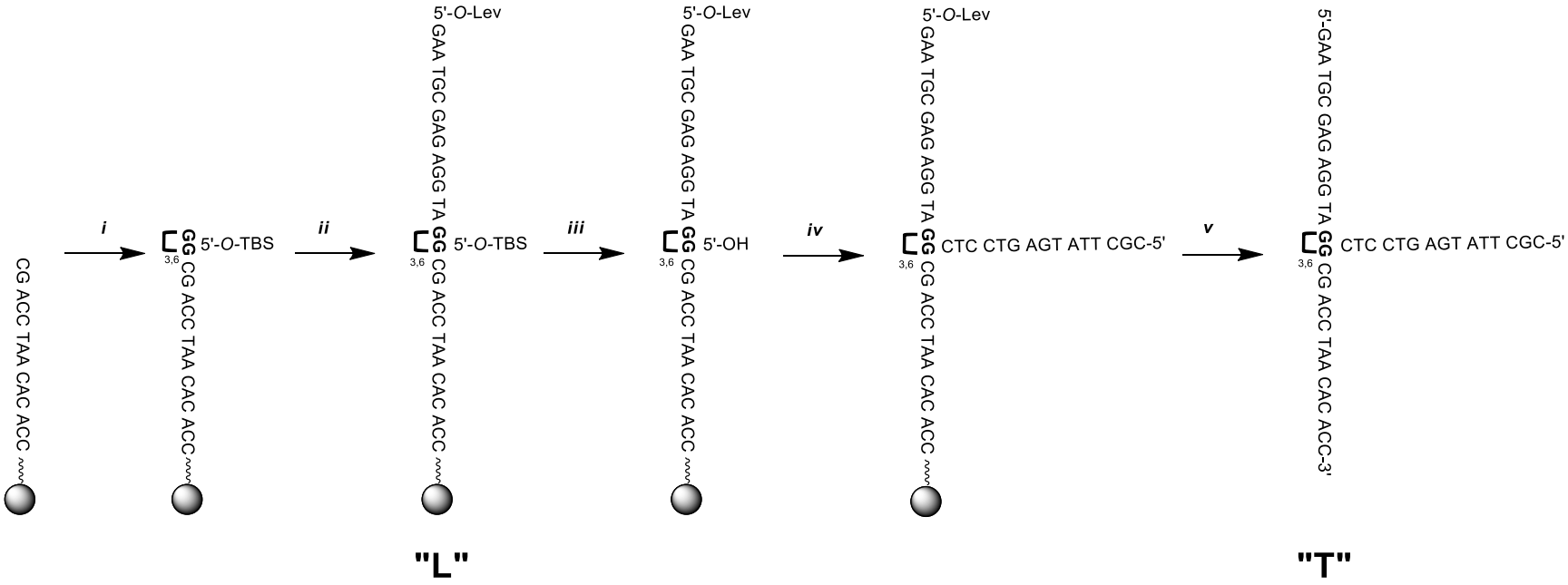


Figure S3.1 (**A**) Sequence composition of **C4/C7 TWJ1-4**. Complementary regions are denoted by the letters **A-F** with the location of the intrastrand cross-link at the vertex of each sequence denoted by $[^{G}_{G}$. (**B**) Overall topology of the assembly of **C4/C7 TWJ1-4** to form a tetrahedron and position of the intrastrand cross-links.

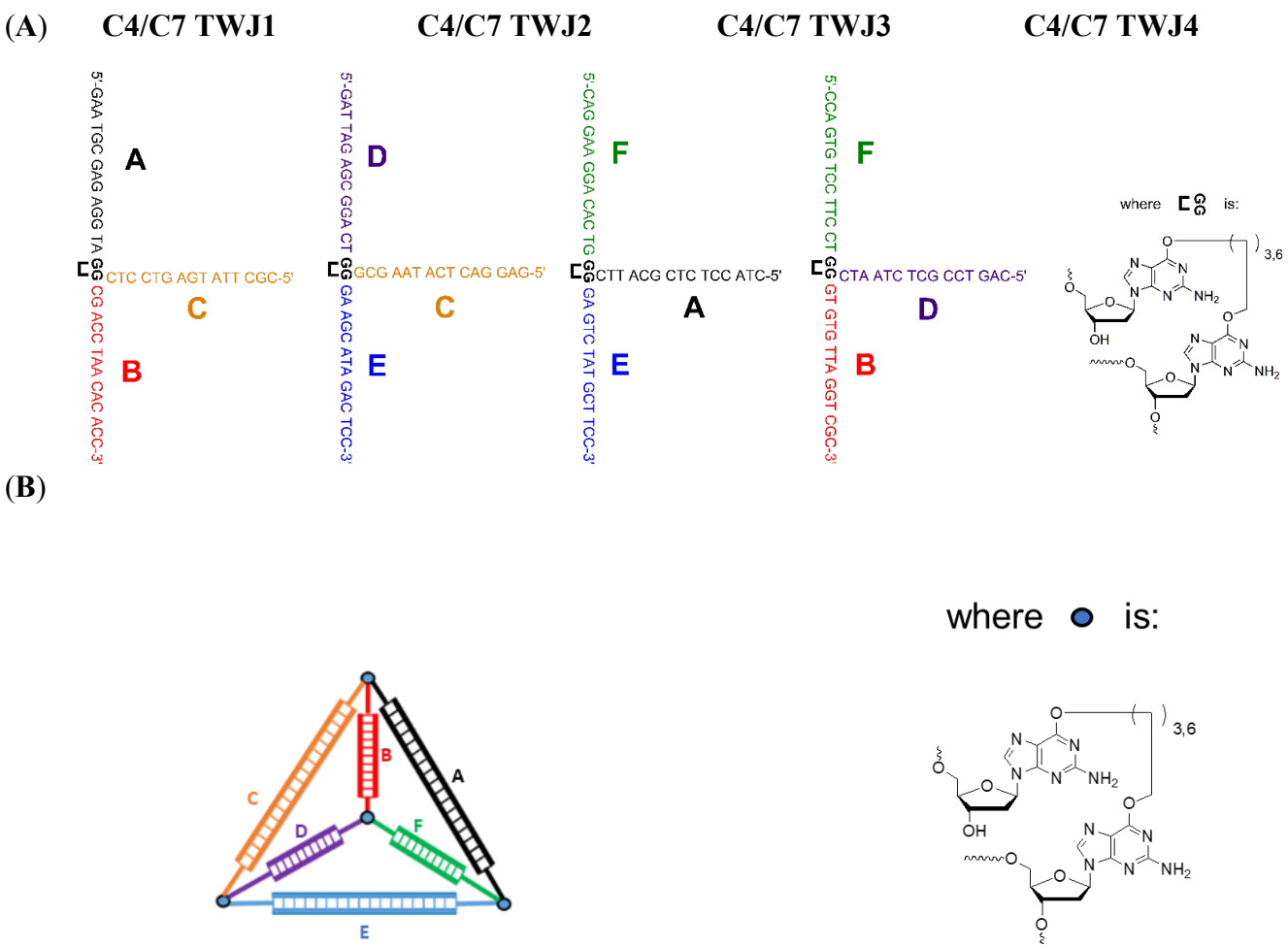


Figure S3.2 500 MHz ^{1}H NMR spectrum of compound **3.1** (in $CDCl_3$)

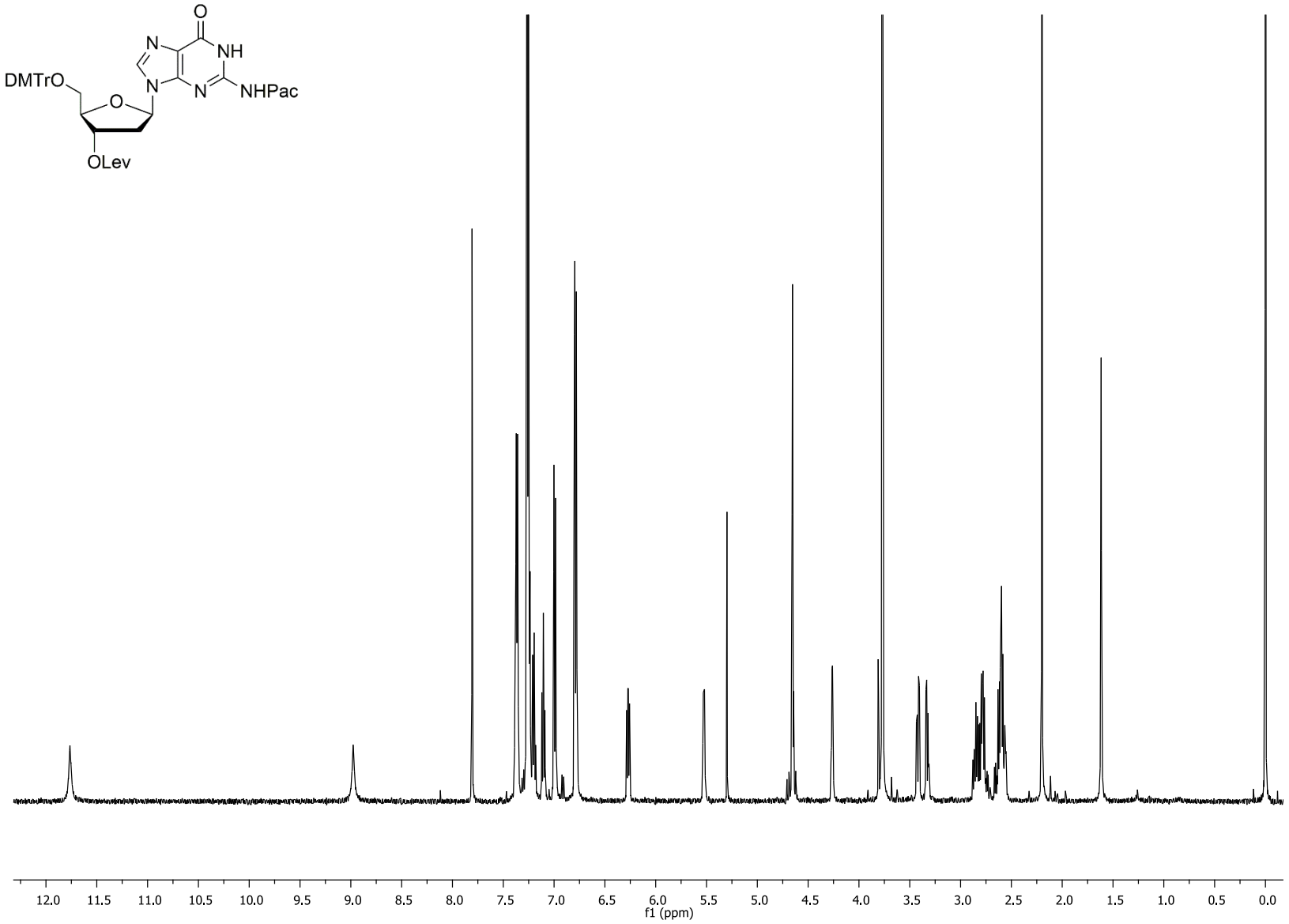


Figure S3.3 125.7 MHz ^{13}C NMR spectrum of compound **3.1** (in $CDCl_3$)

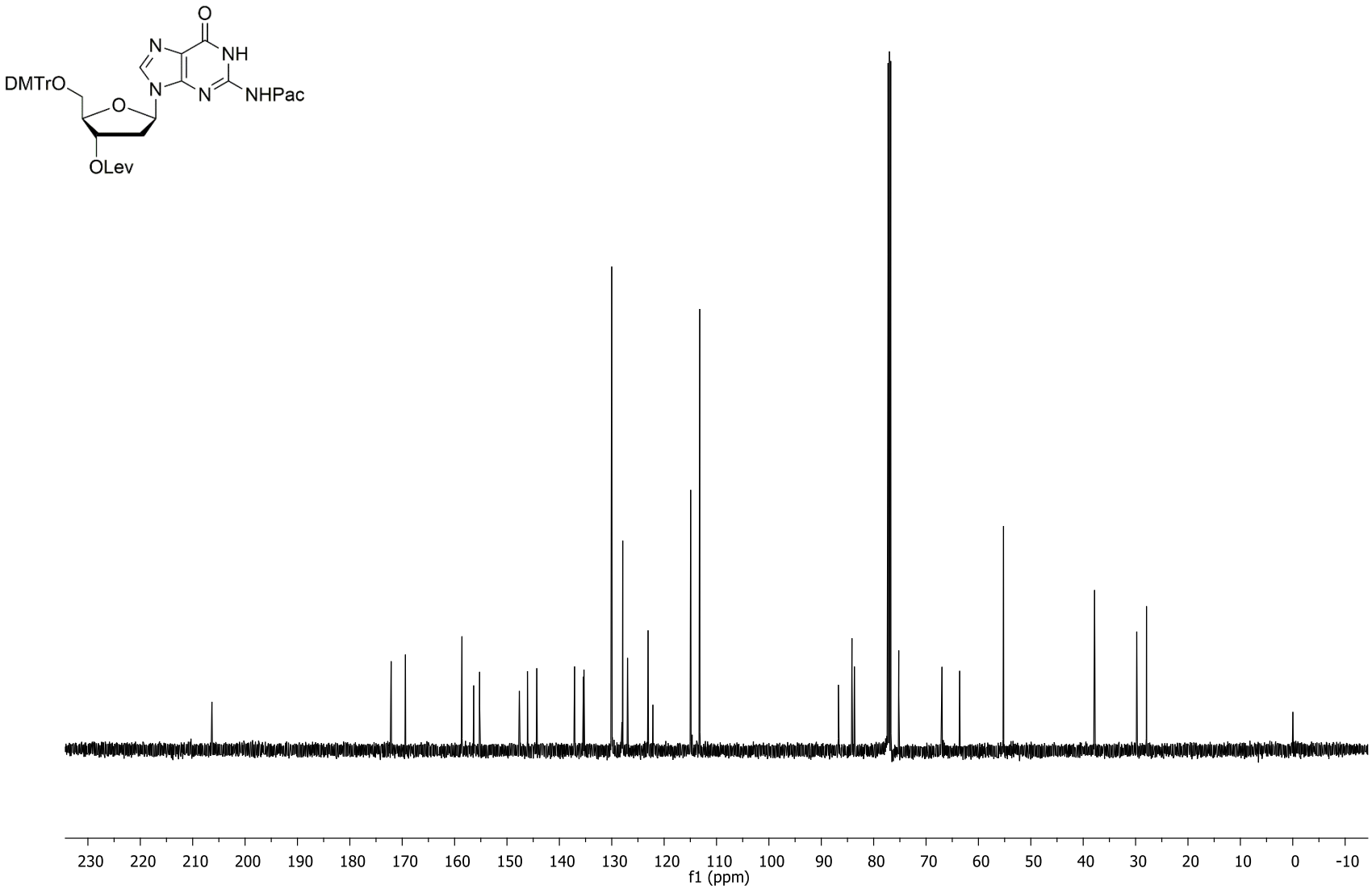


Figure S3.4 500 MHz ^{1}H NMR spectrum of compound **3.2a** (in $CDCl_3$)

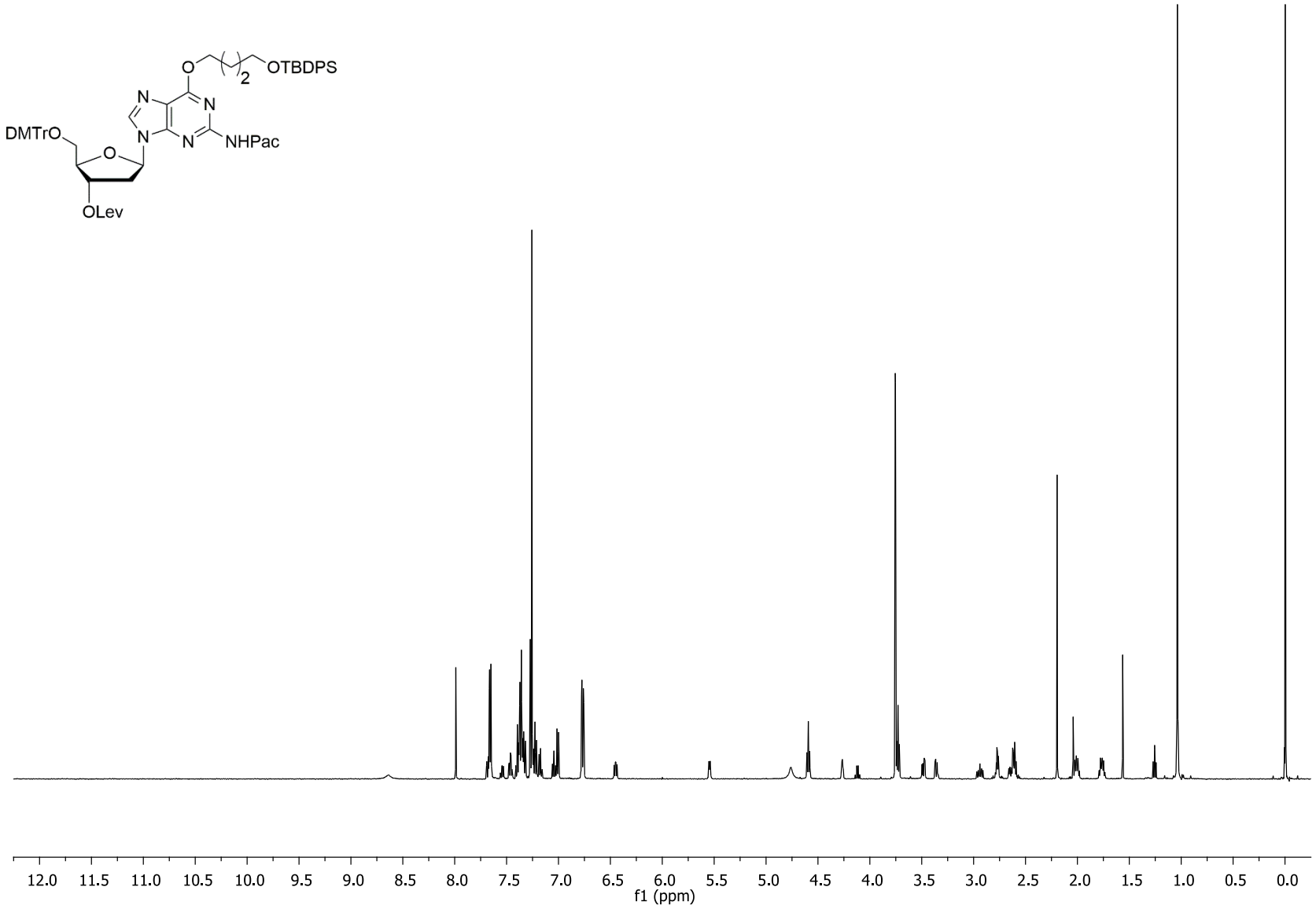


Figure S3.5 125.7 MHz ^{13}C spectrum of compound **3.2a** (in $CDCl_3$)

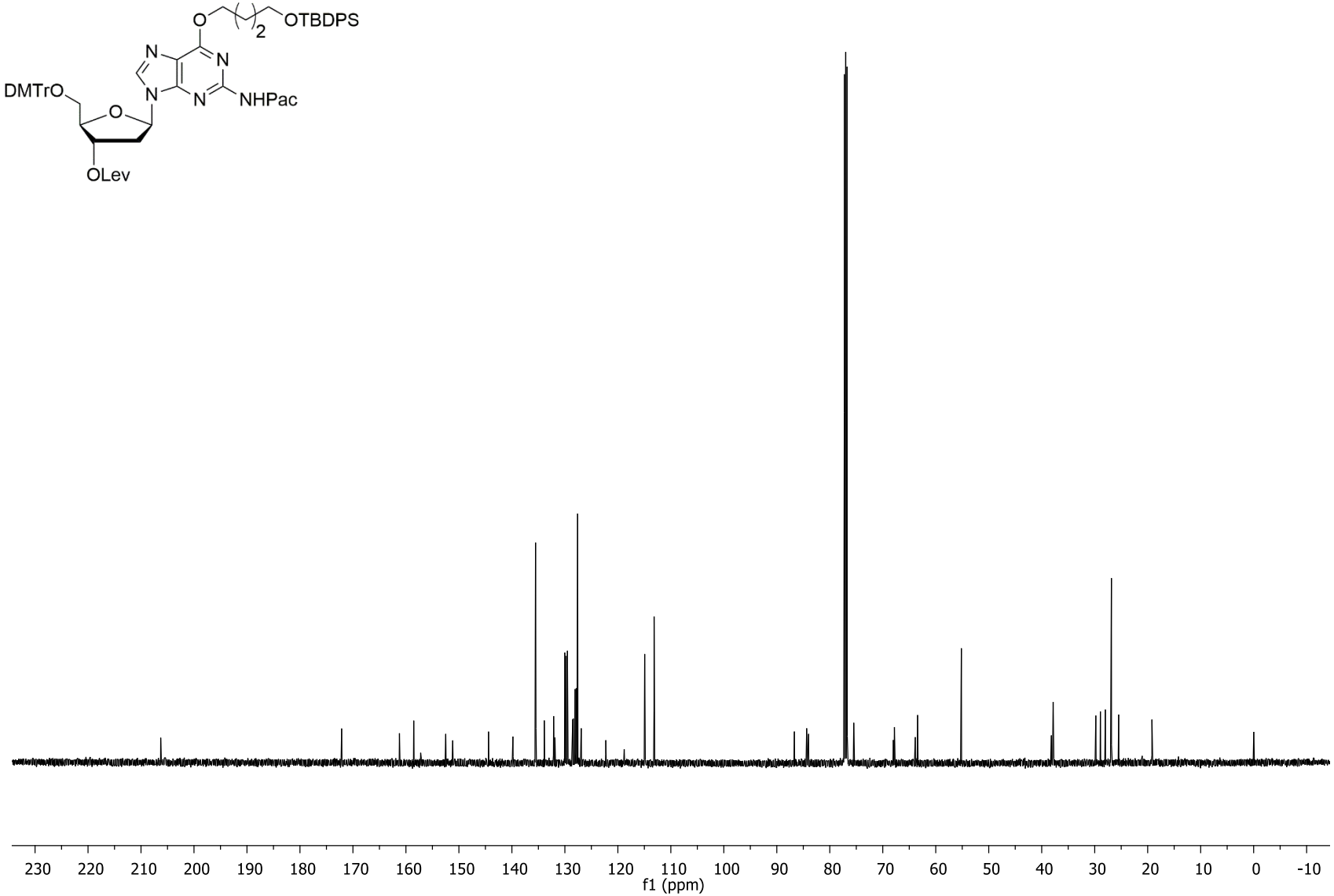


Figure S3.6 500 MHz ^{1}H NMR spectrum of compound **3.2b** (in $CDCl_3$)

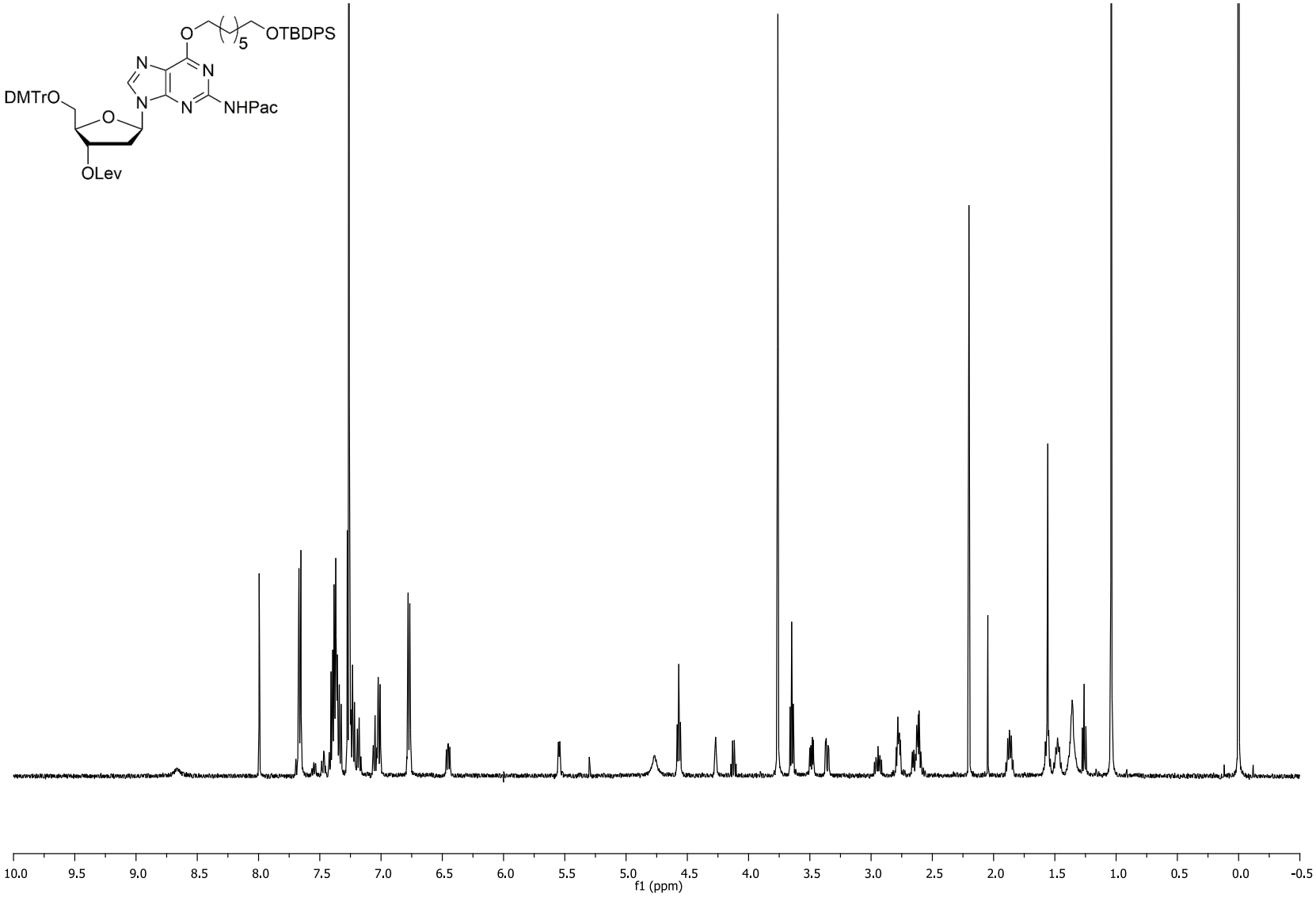


Figure S3.7 125.7 MHz ^{13}C NMR spectrum of compound **3.2b** (in $CDCl_3$)

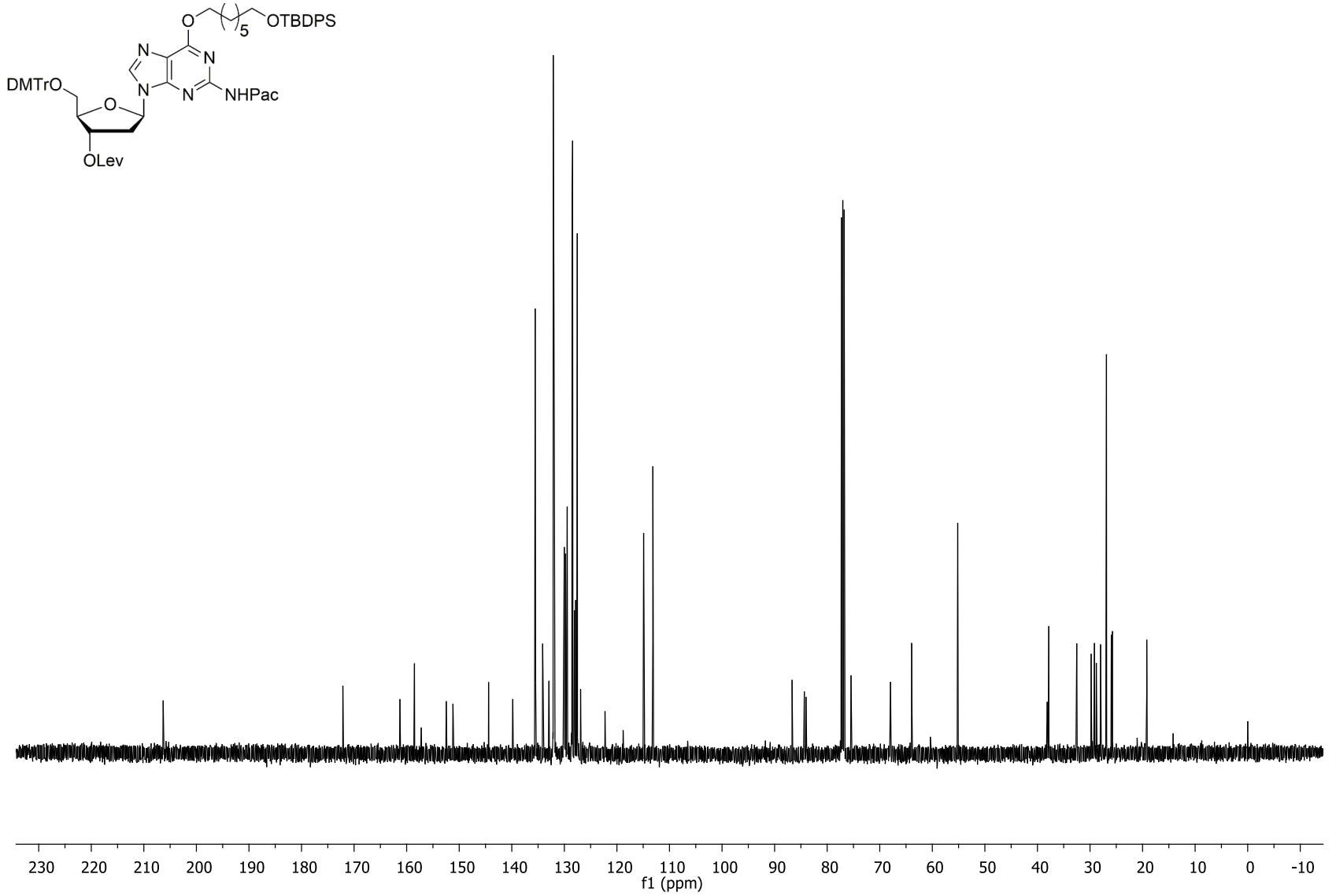


Figure S3.8 500 MHz ^{1}H NMR spectrum of compound **3.3a** (in $CDCl_3$)

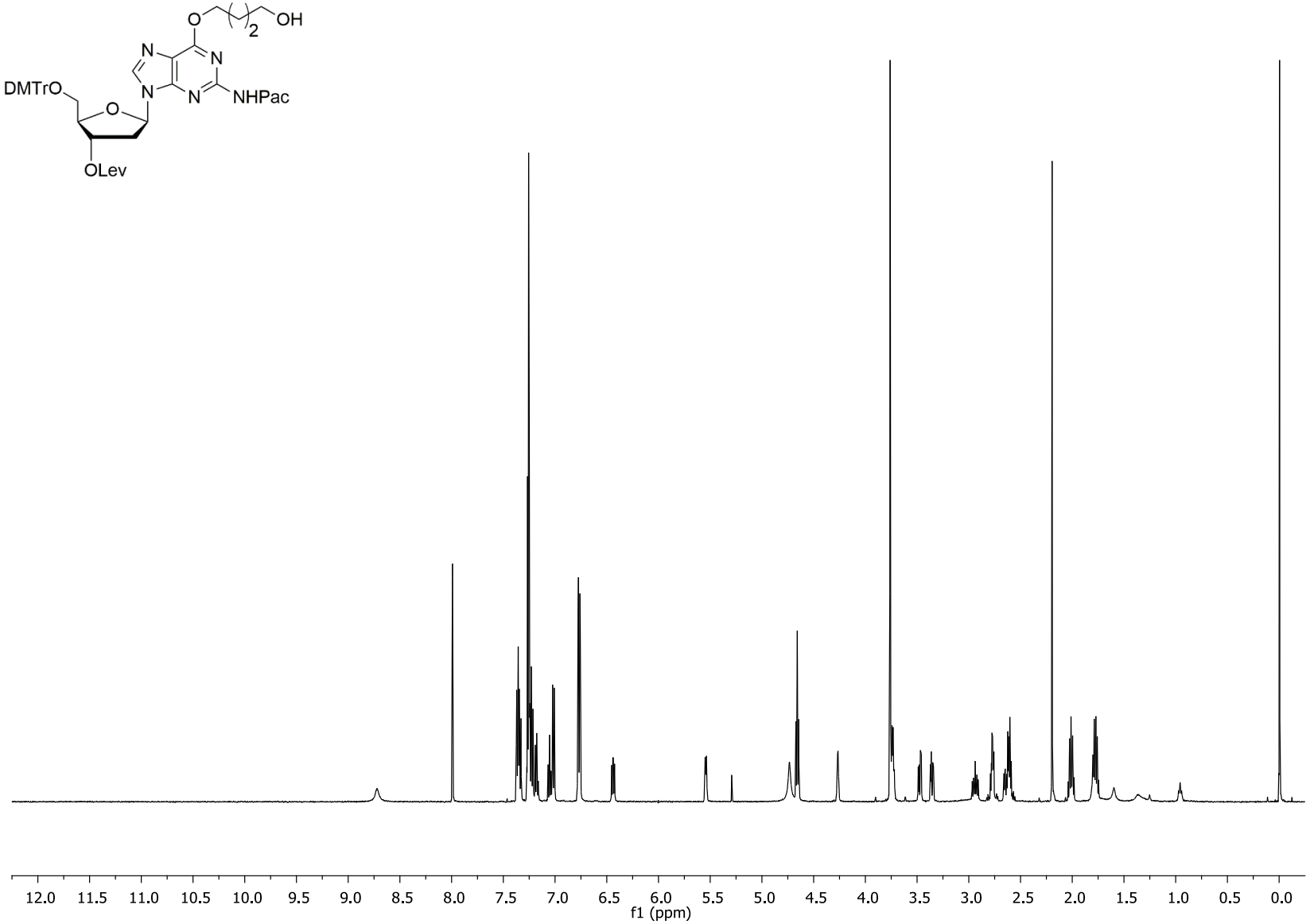


Figure S3.9 125.7 MHz ^{13}C NMR spectrum of compound **3.3a** (in $CDCl_3$)

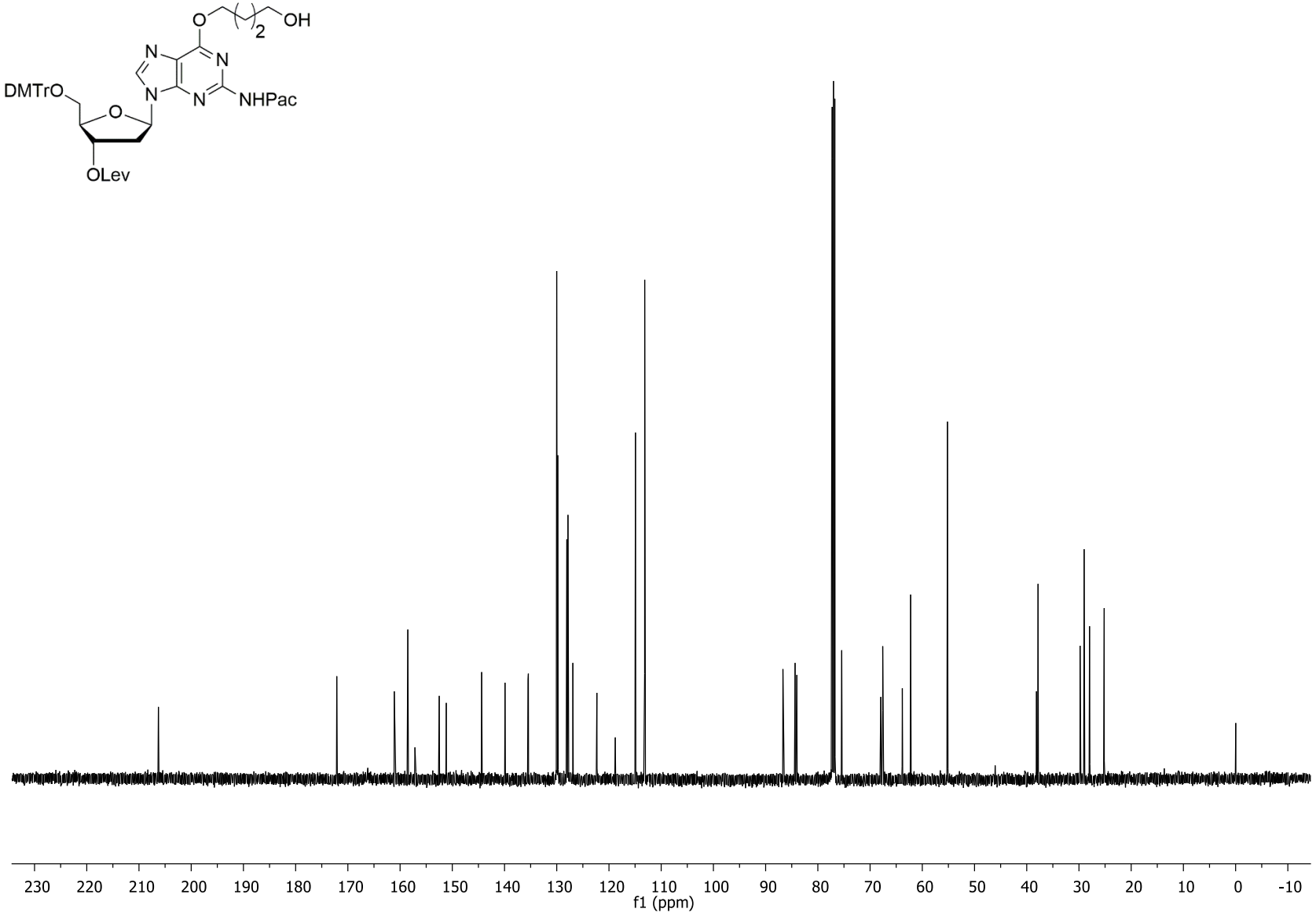


Figure S3.10 500 MHz ^{1}H NMR spectrum of compound **3.3b** (in $CDCl_3$)

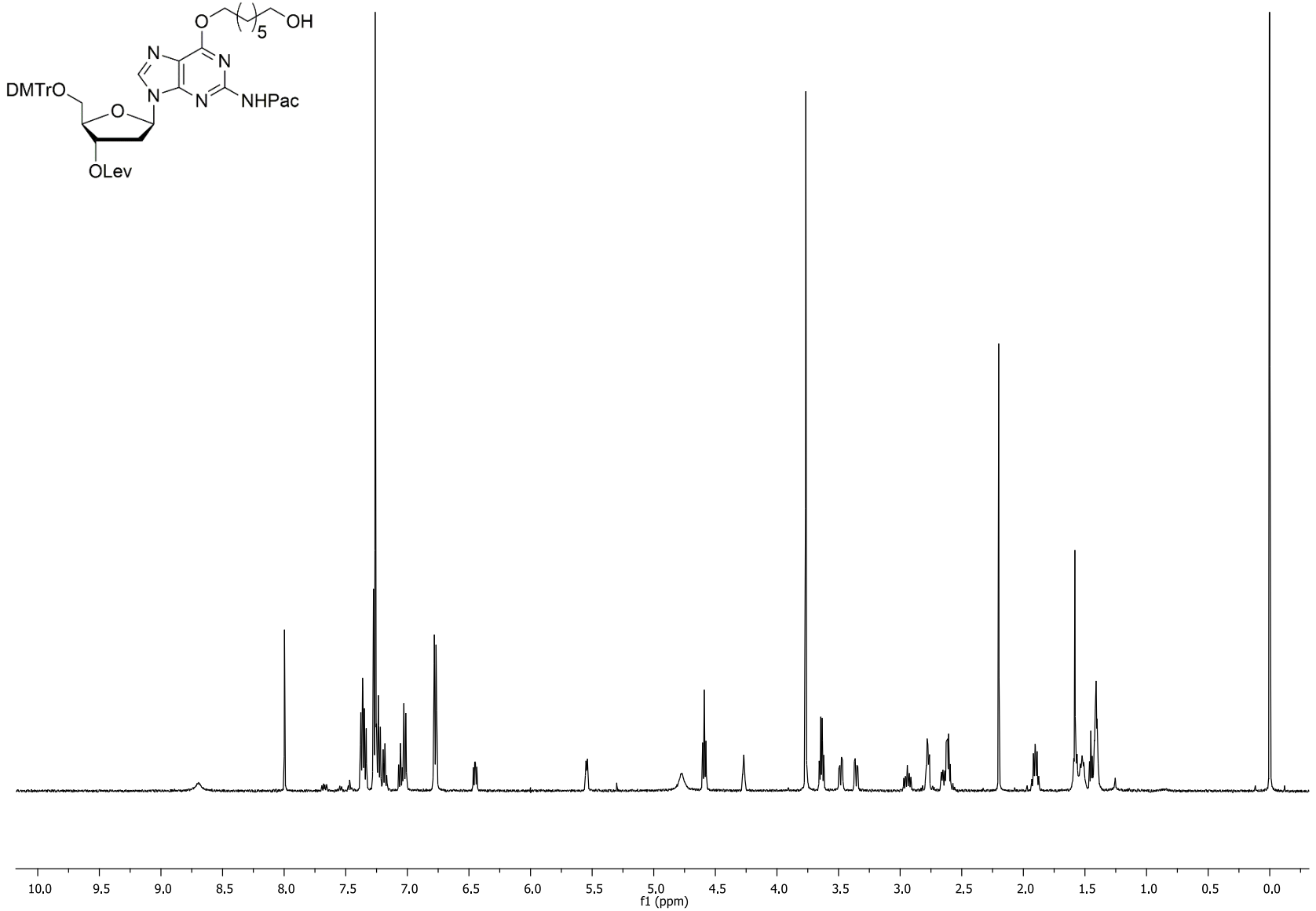


Figure S3.11 125.7 MHz ^{13}C NMR spectrum of compound **3.3b** (in $CDCl_3$)

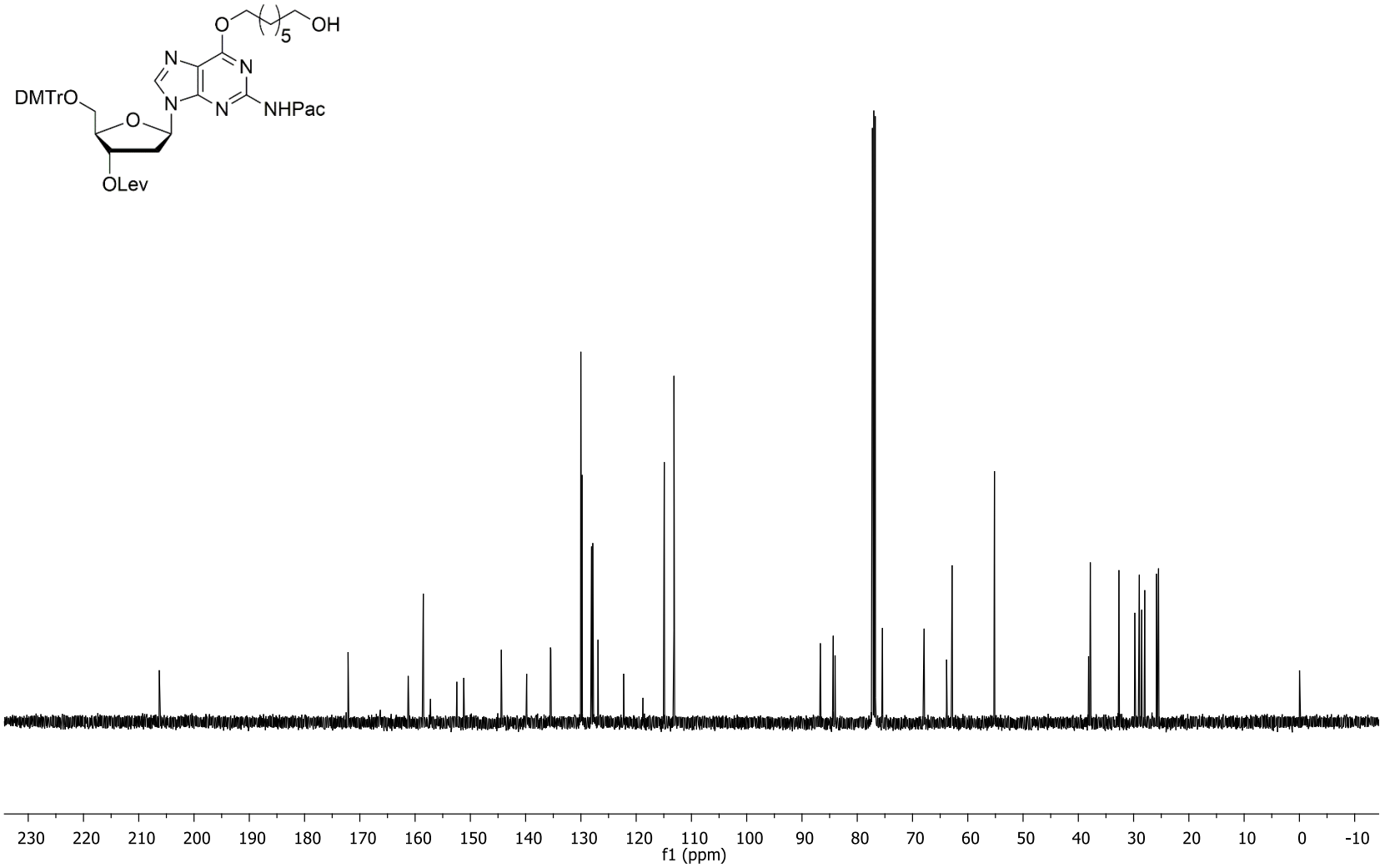


Figure S3.12 500 MHz ^{1}H NMR spectrum of compound **3.4** (in $CDCl_3$)

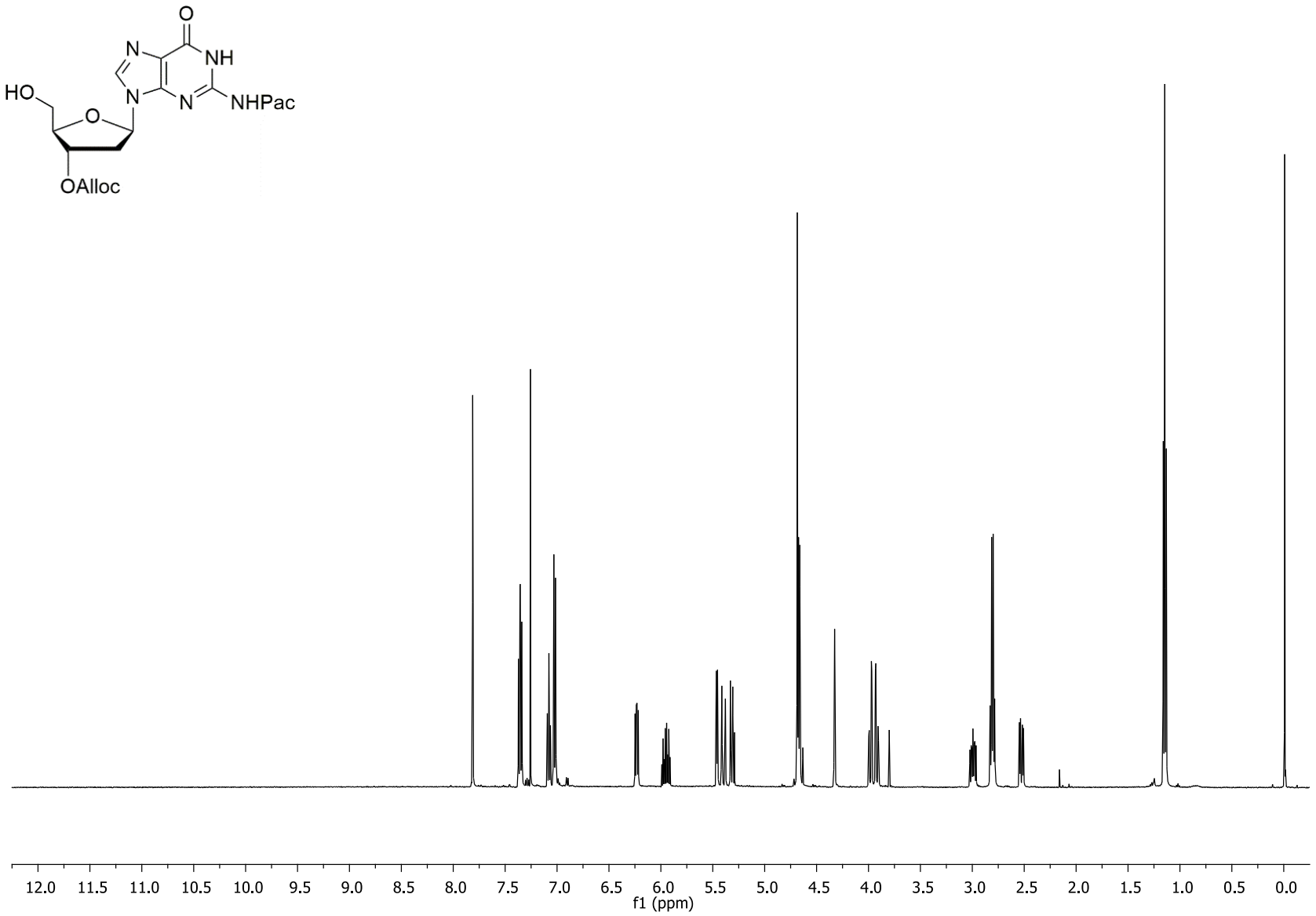


Figure S3.13 125.7 MHz ^{13}C NMR spectrum of compound **3.4** (in $CDCl_3$)

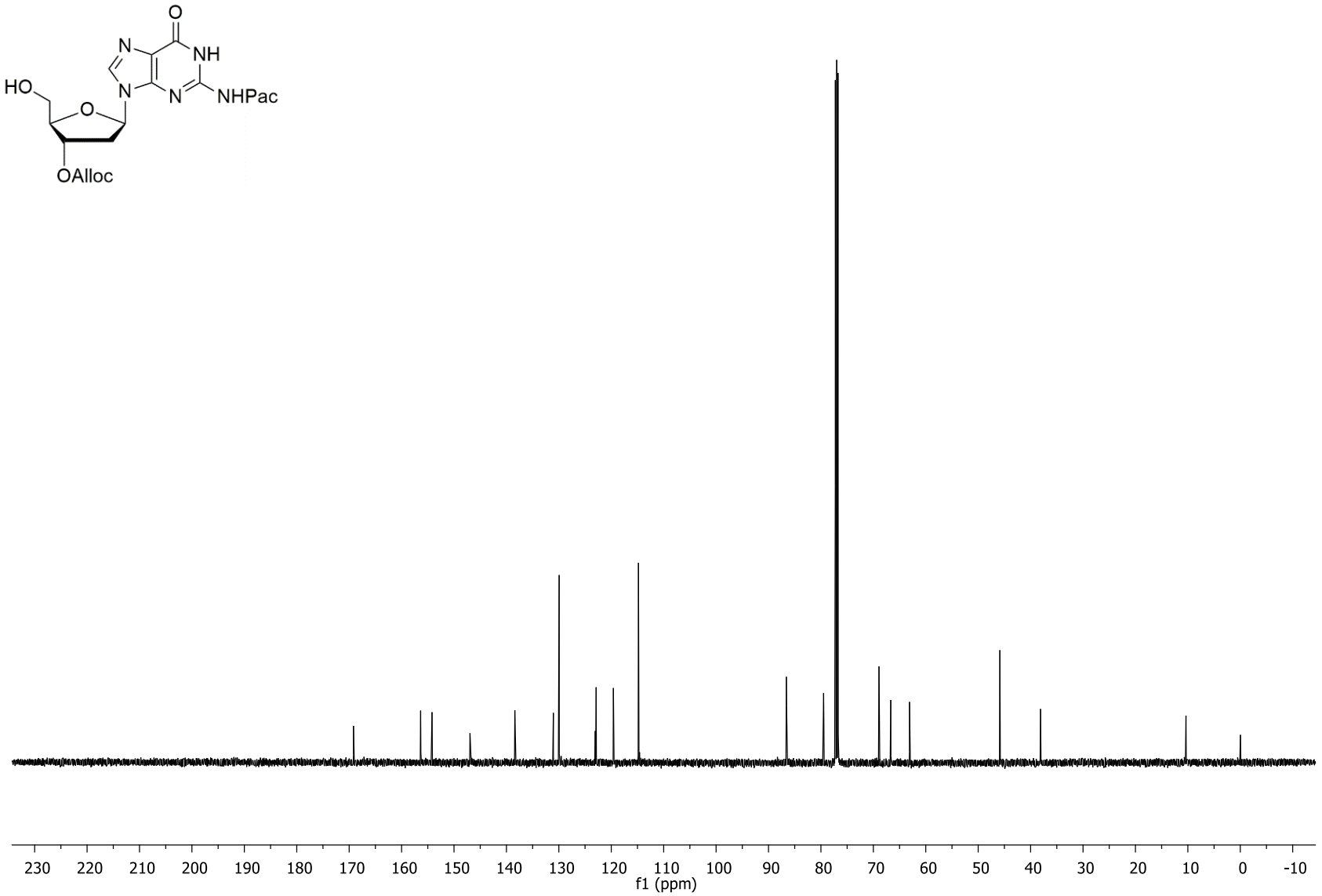


Figure S3.14 500 MHz ^{1}H NMR spectrum of compound **3.5** (in $CDCl_3$)

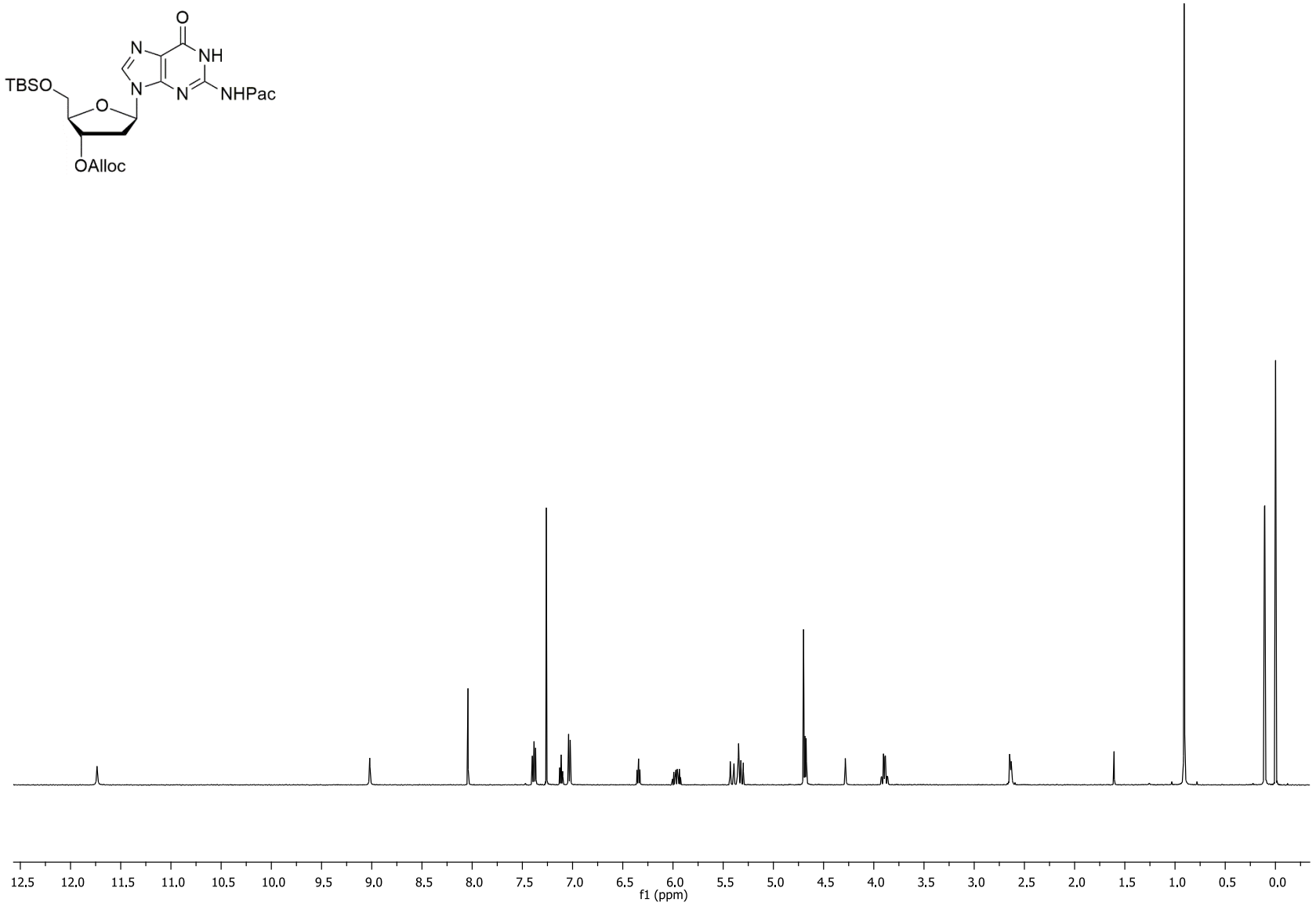


Figure S3.15 125.7 MHz ^{13}C NMR spectrum of compound **3.5** (in $CDCl_3$)

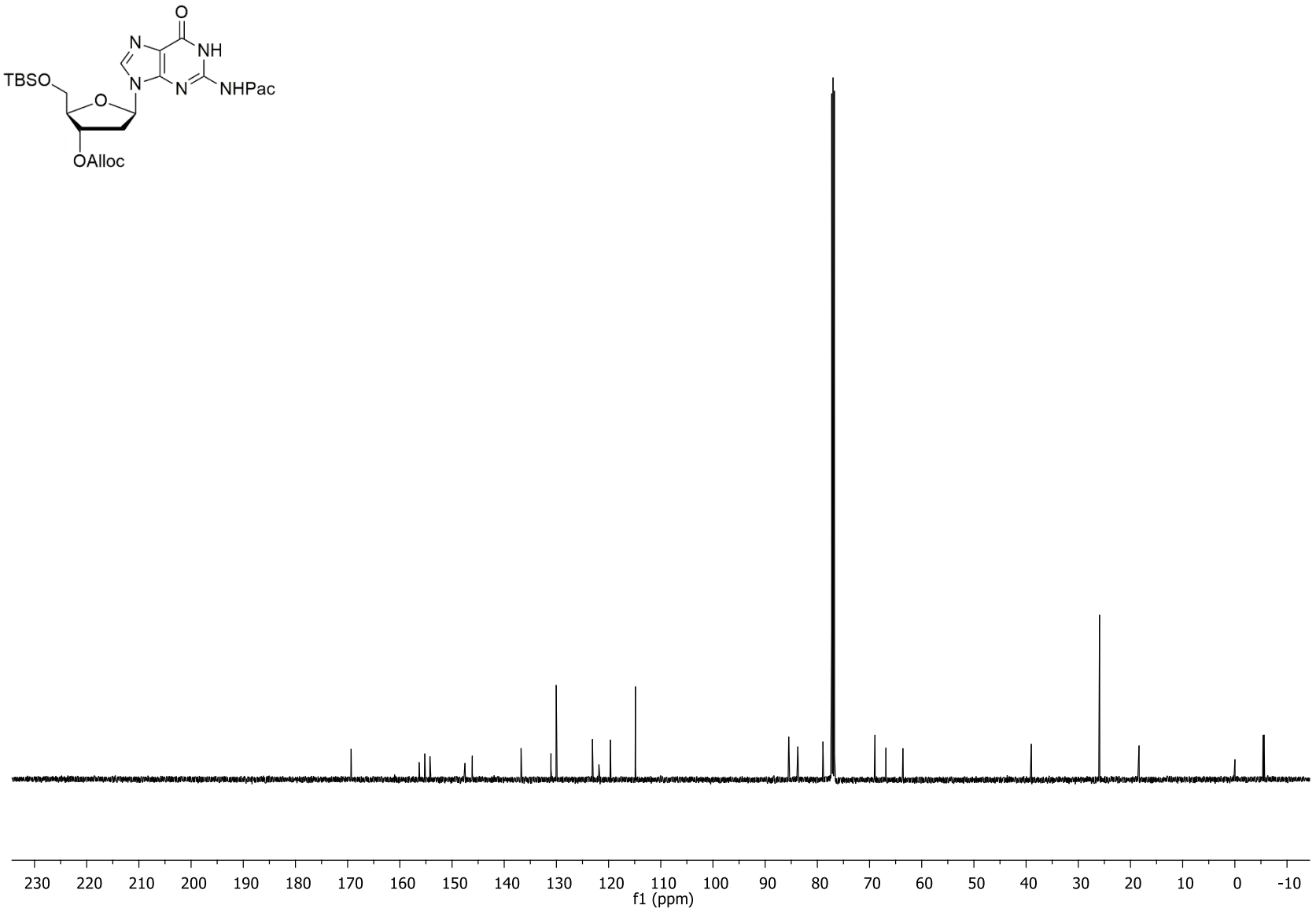


Figure S3.16 500 MHz ^{1}H NMR spectrum of compound **3.6a** (in $CDCl_3$)

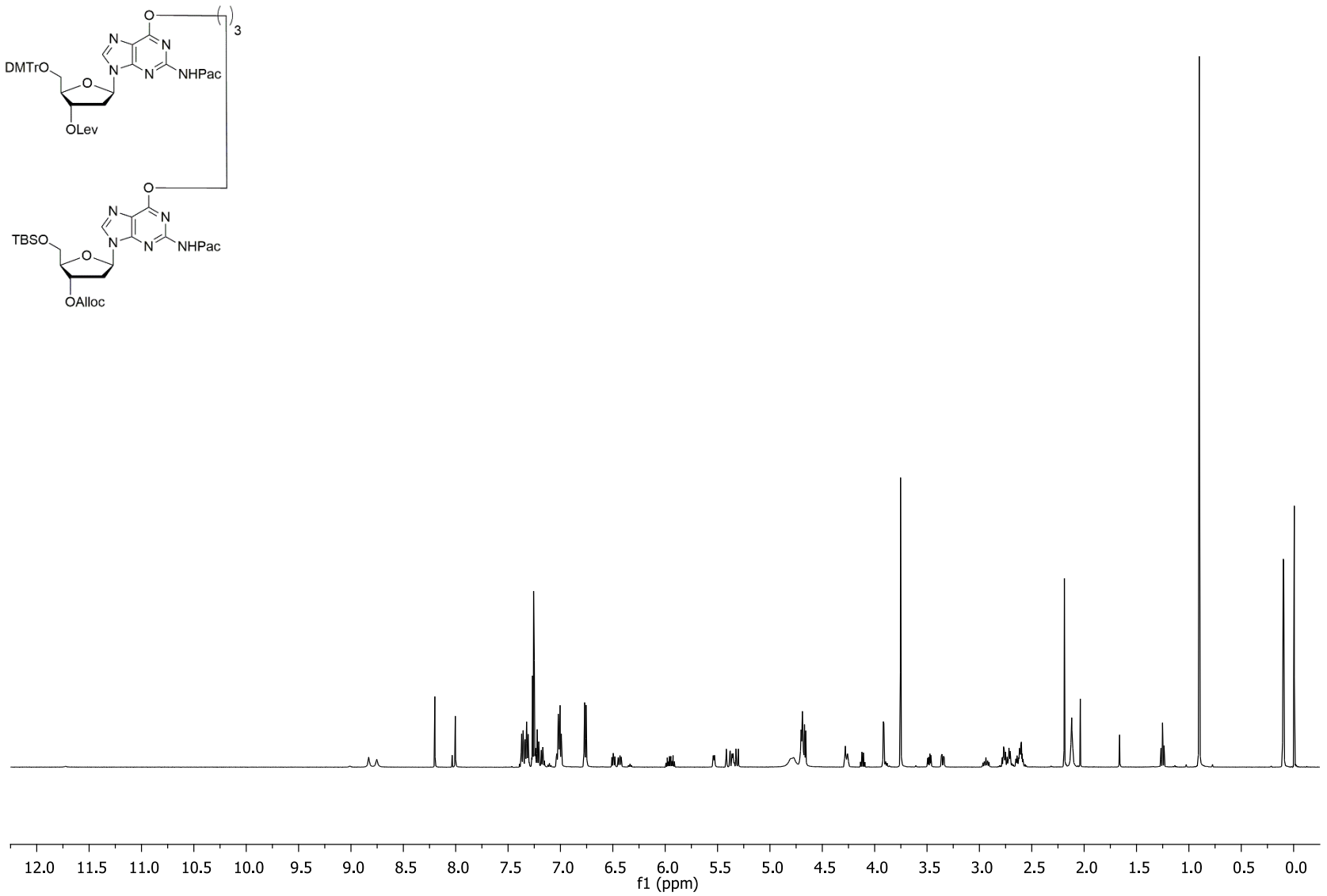


Figure S3.17 125.7 MHz ^{13}C NMR spectrum of compound **3.6a** (in $CDCl_3$)

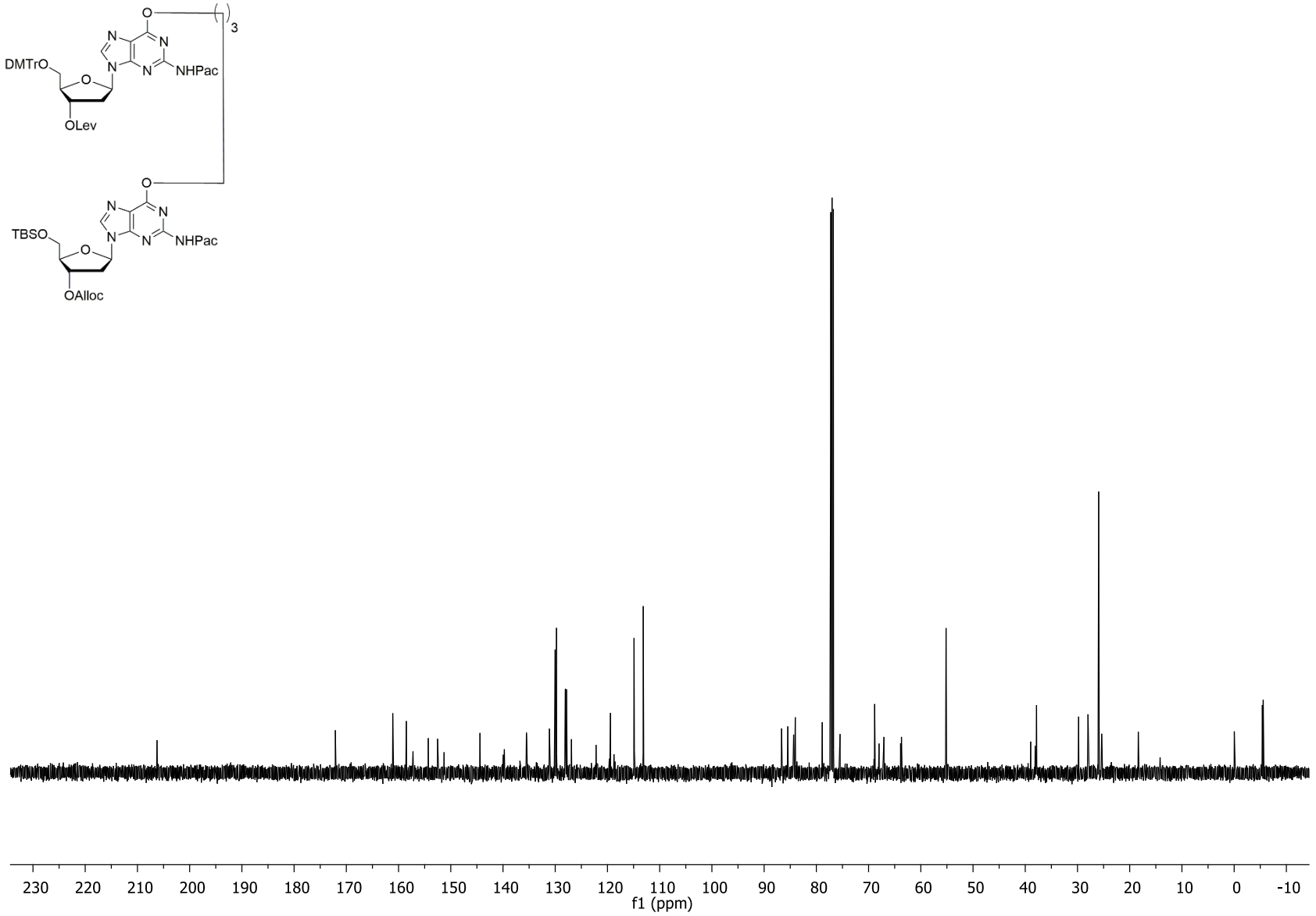


Figure S3.18 500 MHz ^{1}H NMR spectrum of compound **3.6b** (in $CDCl_3$)

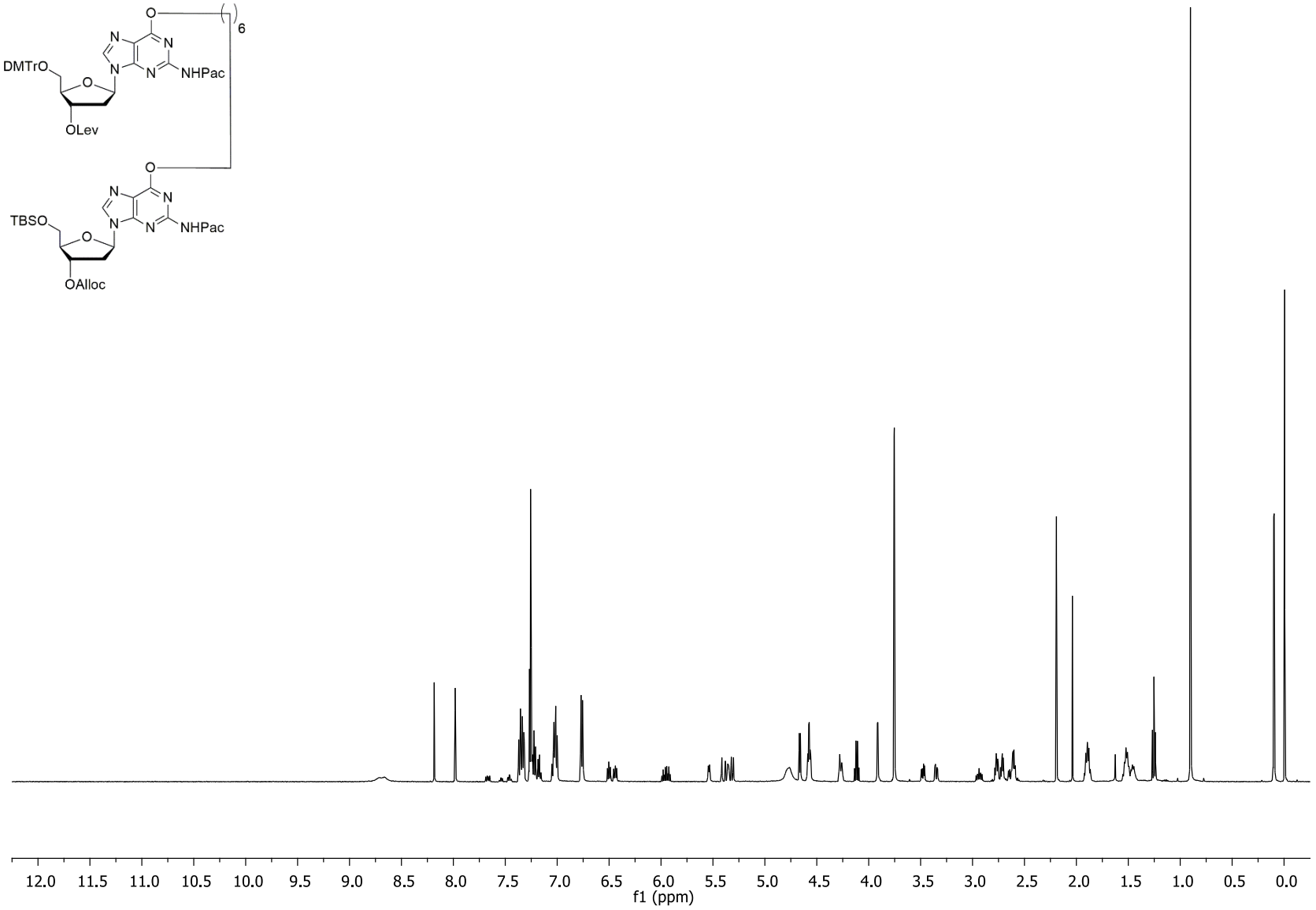


Figure S3.19 125.7 MHz ^{13}C NMR spectrum of compound **3.6b** (in $CDCl_3$)

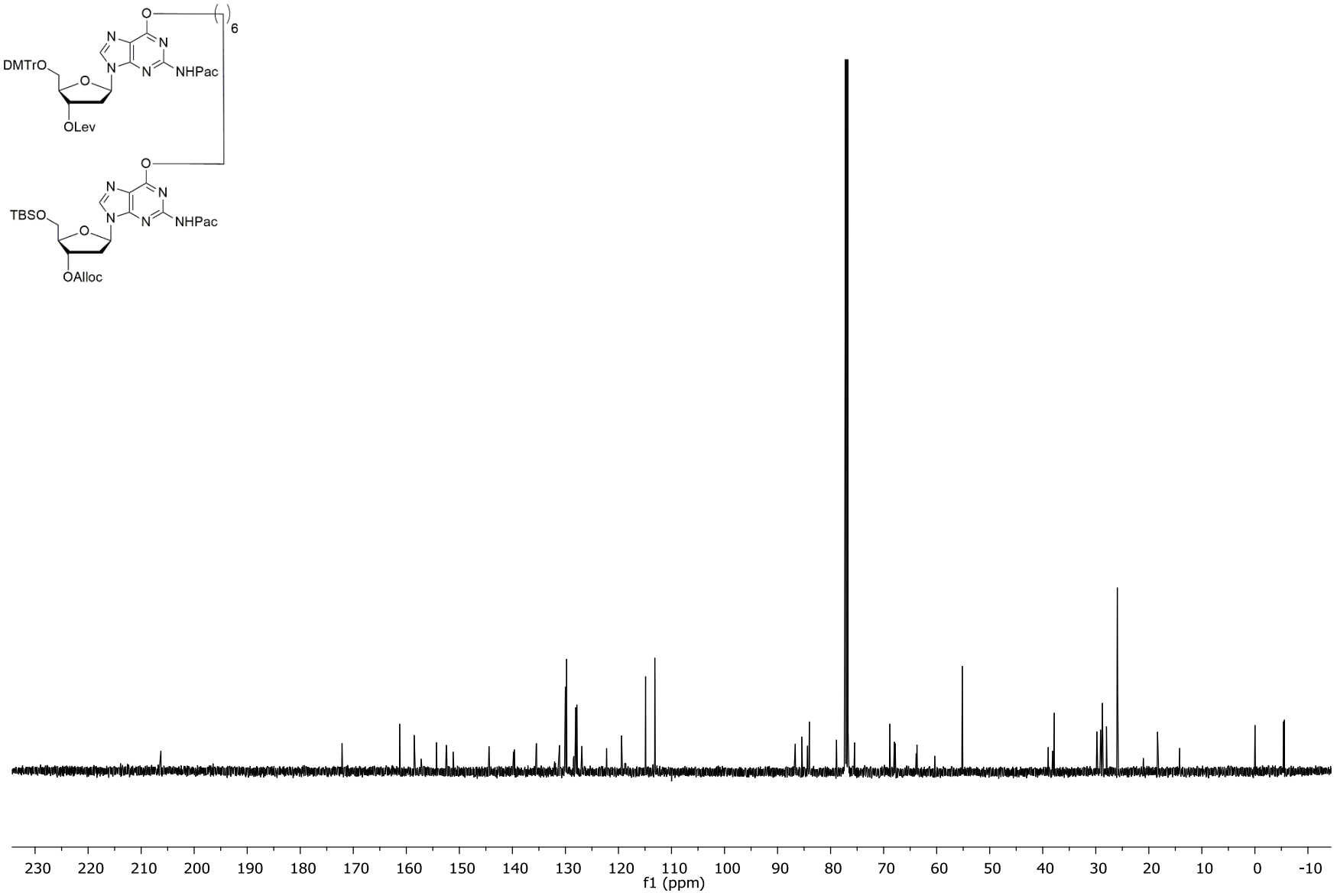


Figure S3.20 500 MHz ^{1}H NMR spectrum of compound **3.7a** (in $CDCl_3$)

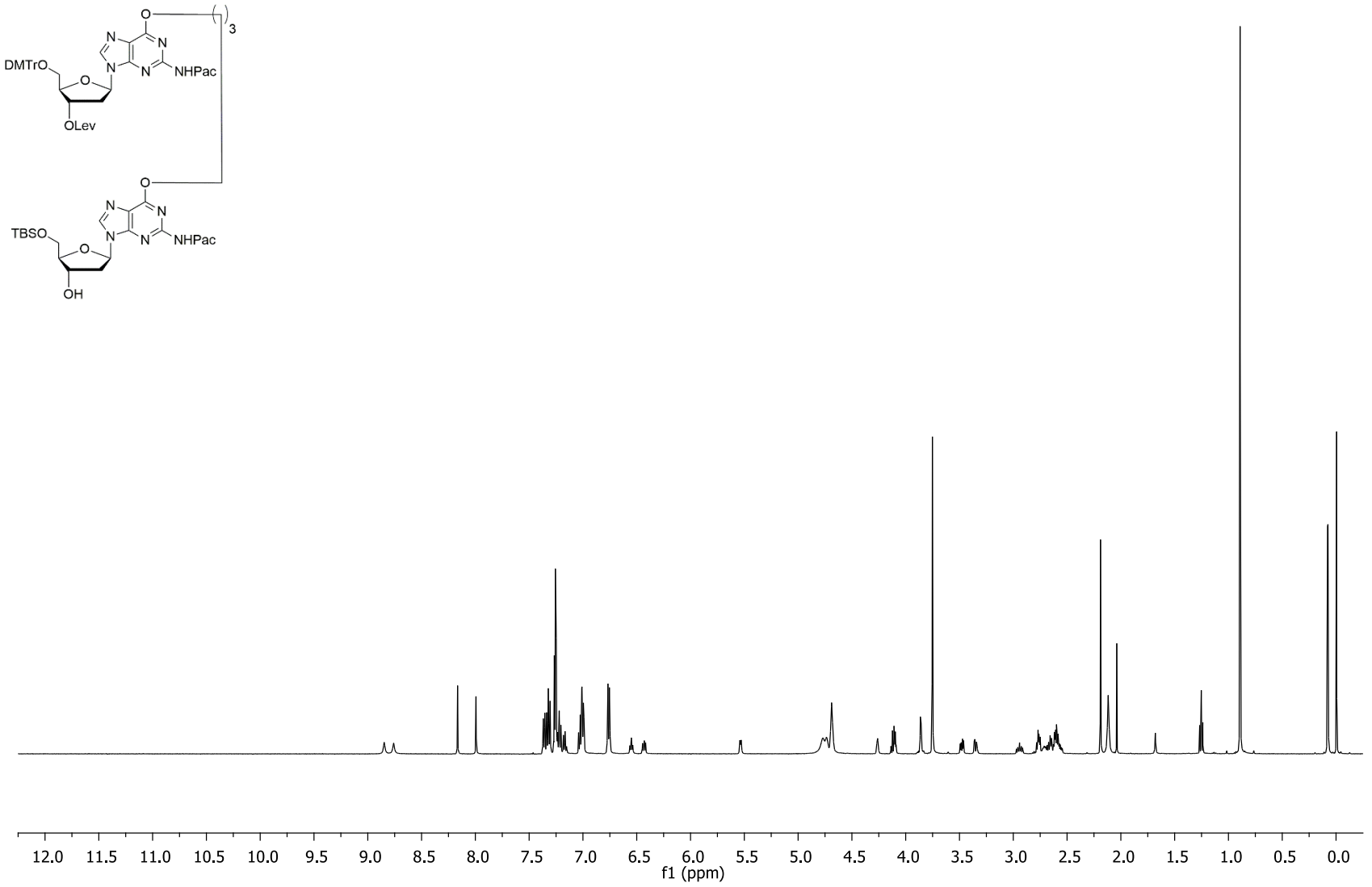


Figure S3.21 125.7 MHz ^{13}C NMR spectrum of compound **3.7a** (in $CDCl_3$)

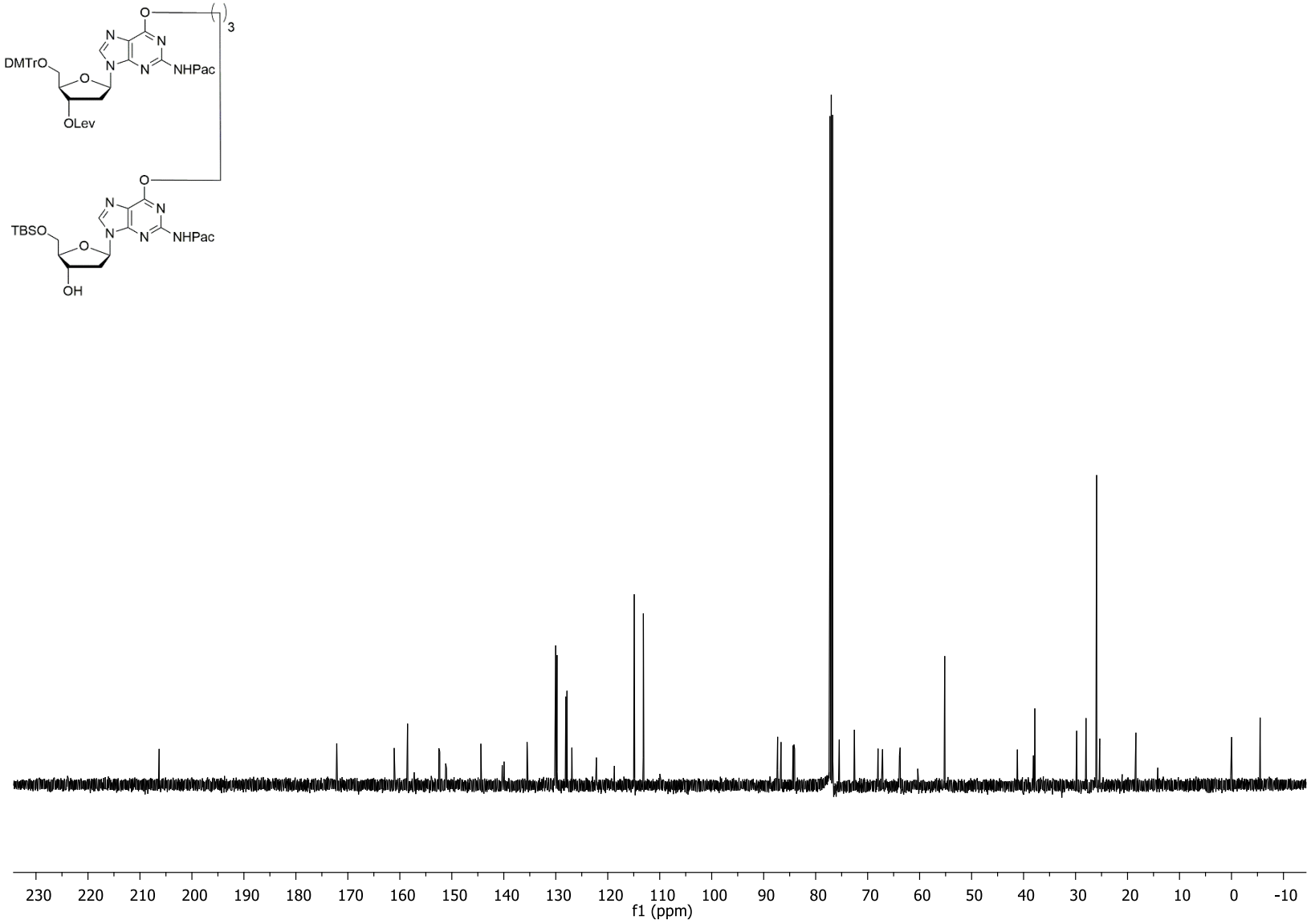


Figure S3.22 500 MHz ^{1}H NMR spectrum of compound **3.7b** (in $CDCl_3$)

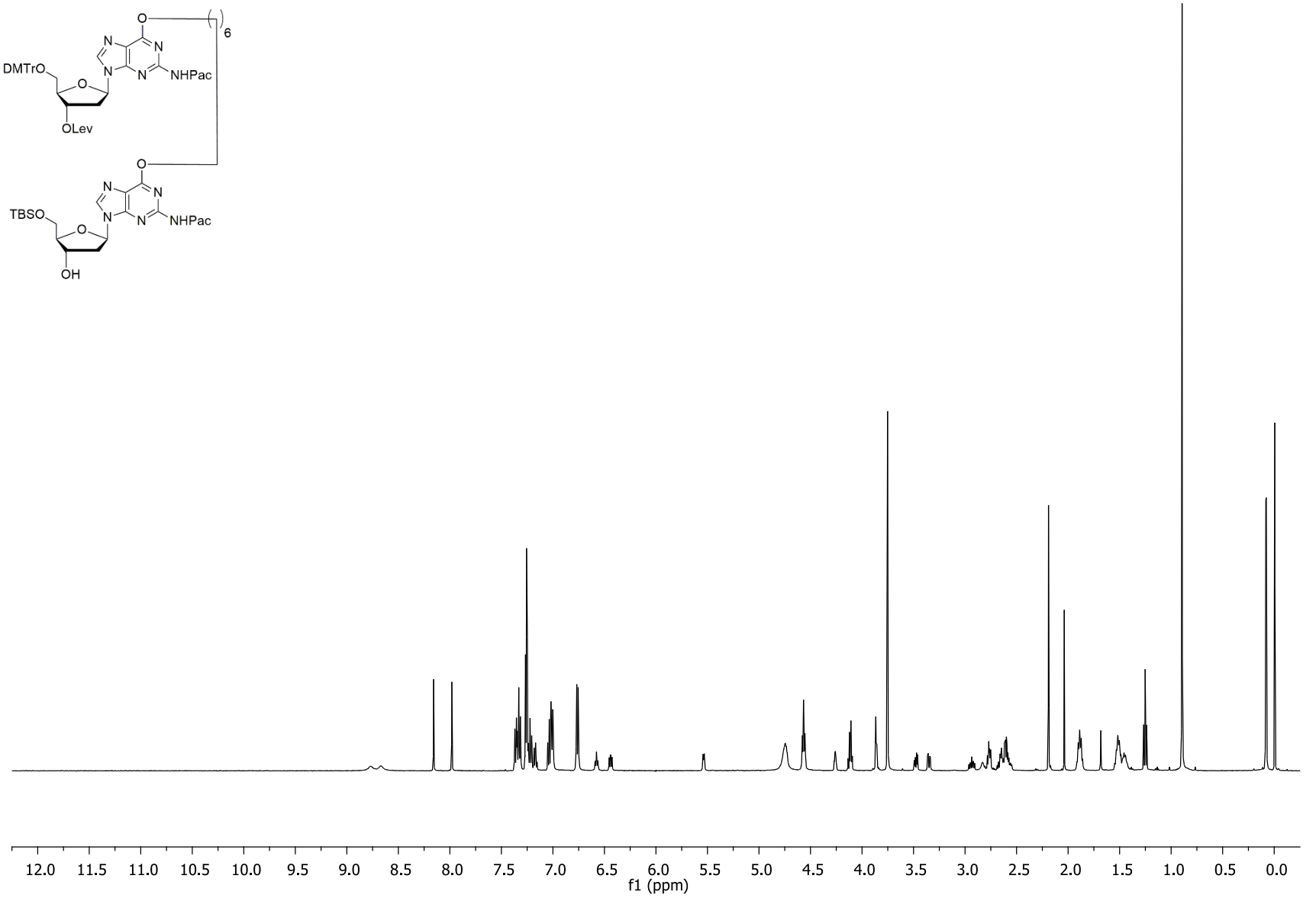


Figure S3.23 125.7 MHz ^{13}C NMR spectrum of compound **3.7b** (in $CDCl_3$)

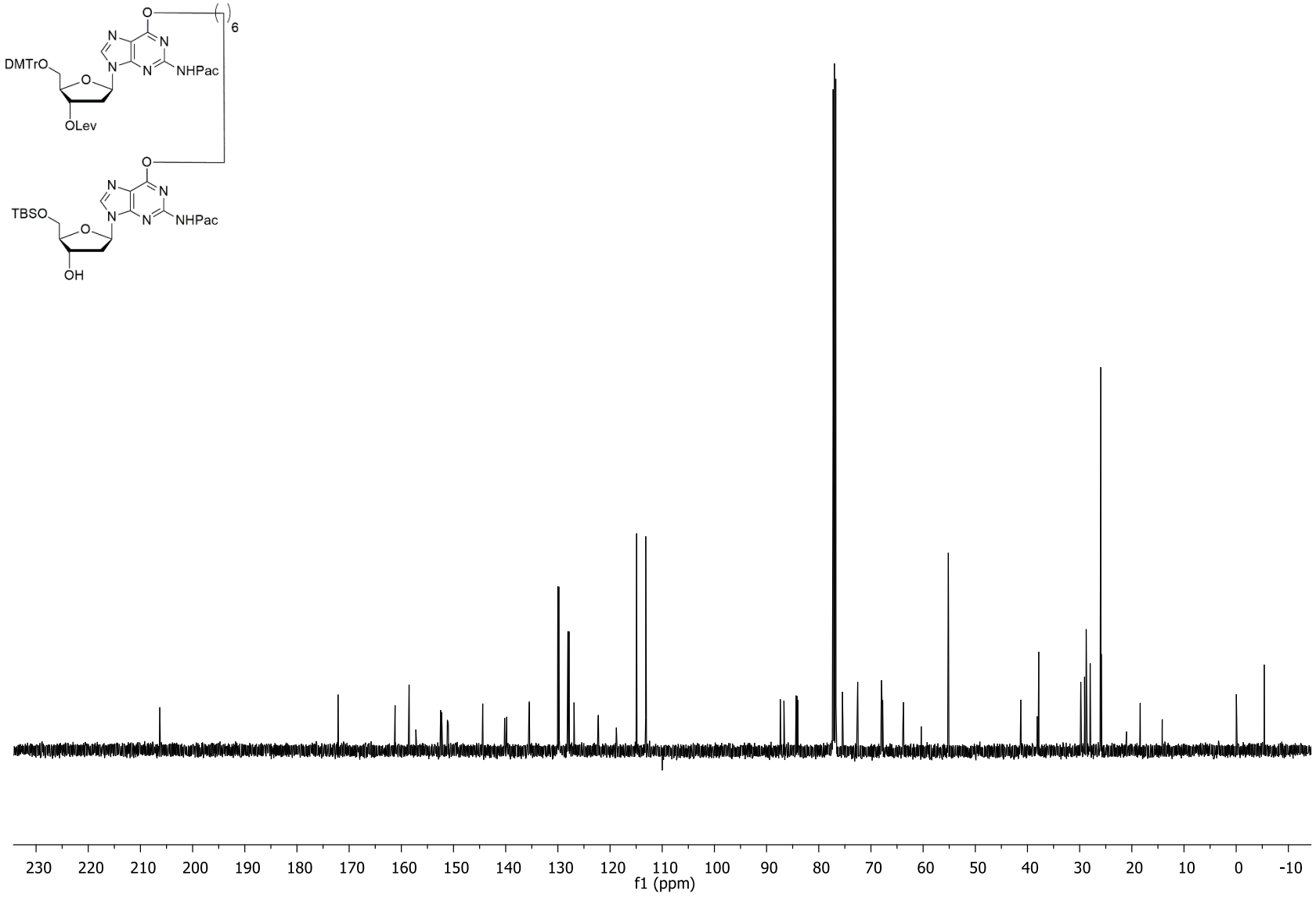


Figure S3.24 500 MHz ^{1}H NMR spectrum of compound **3.8a** (in d_6-acetone)

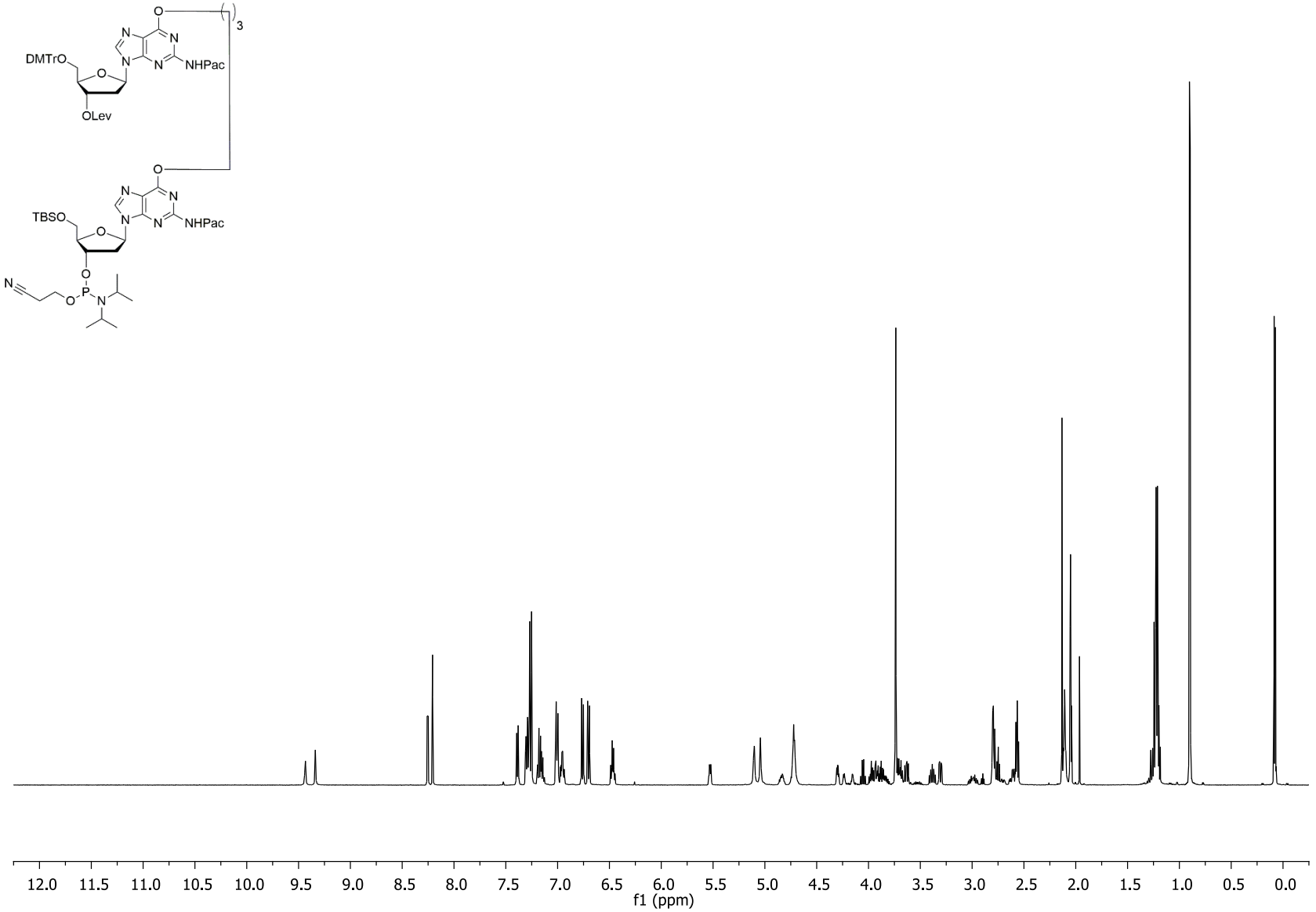


Figure S3.25 202.3 MHz ^{31}P NMR spectrum of compound **3.8a** (d_6-acetone)

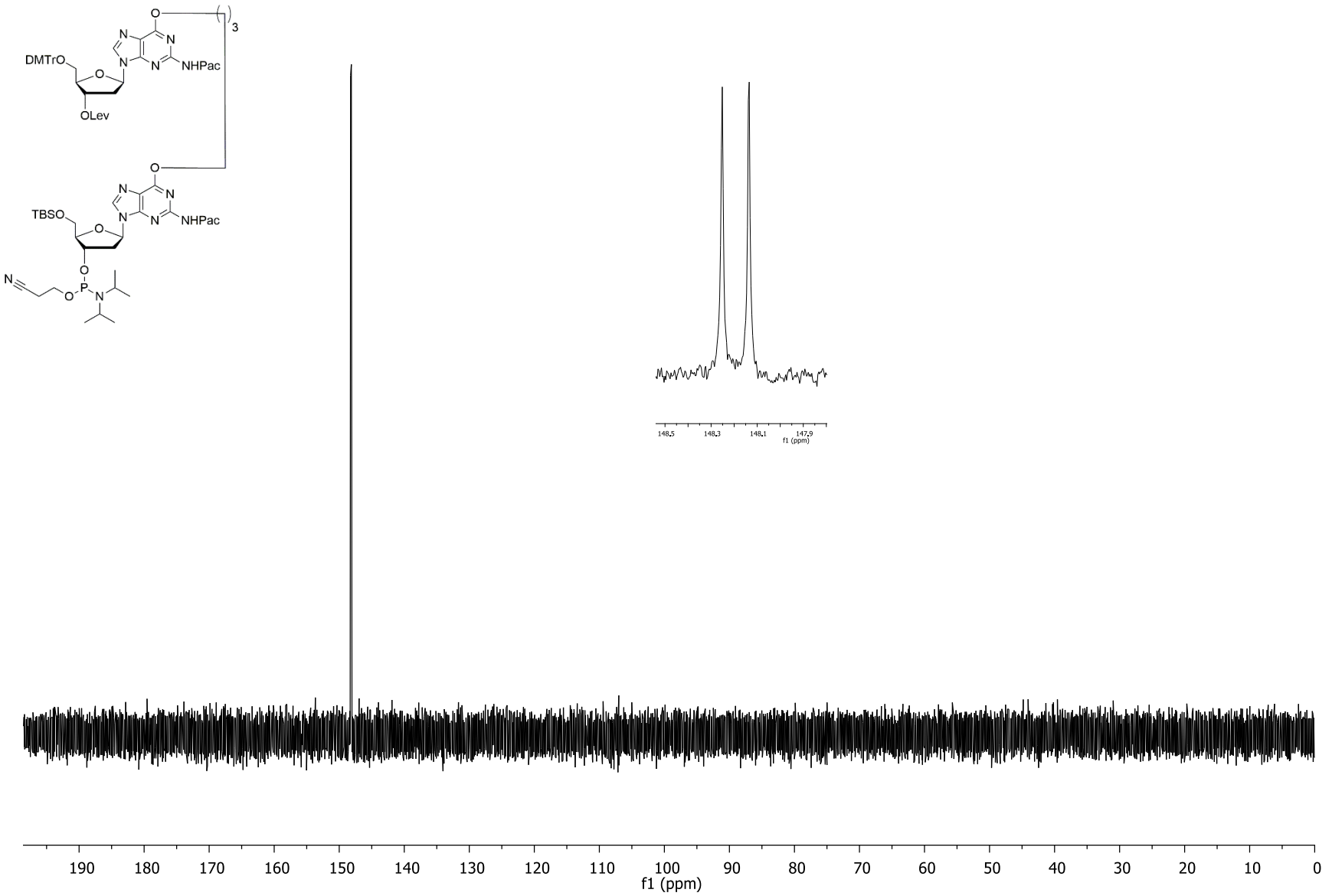


Figure S3.26 125.7 MHz ^{13}C NMR spectrum of compound **3.8a** (in d_6-acetone)

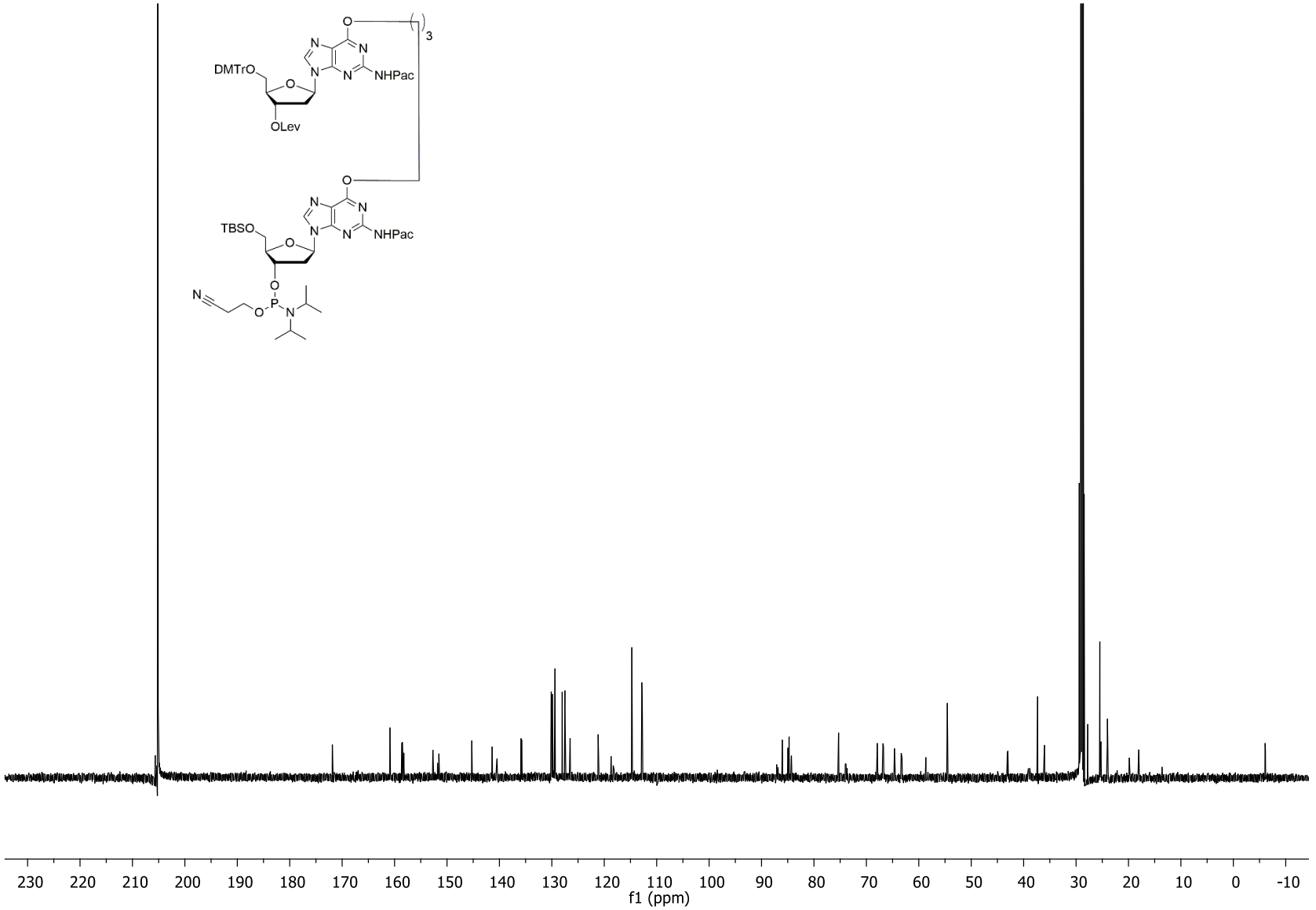


Figure S3.27 500 MHz ^{1}H NMR spectrum of compound **3.8b** (in d_6-acetone)

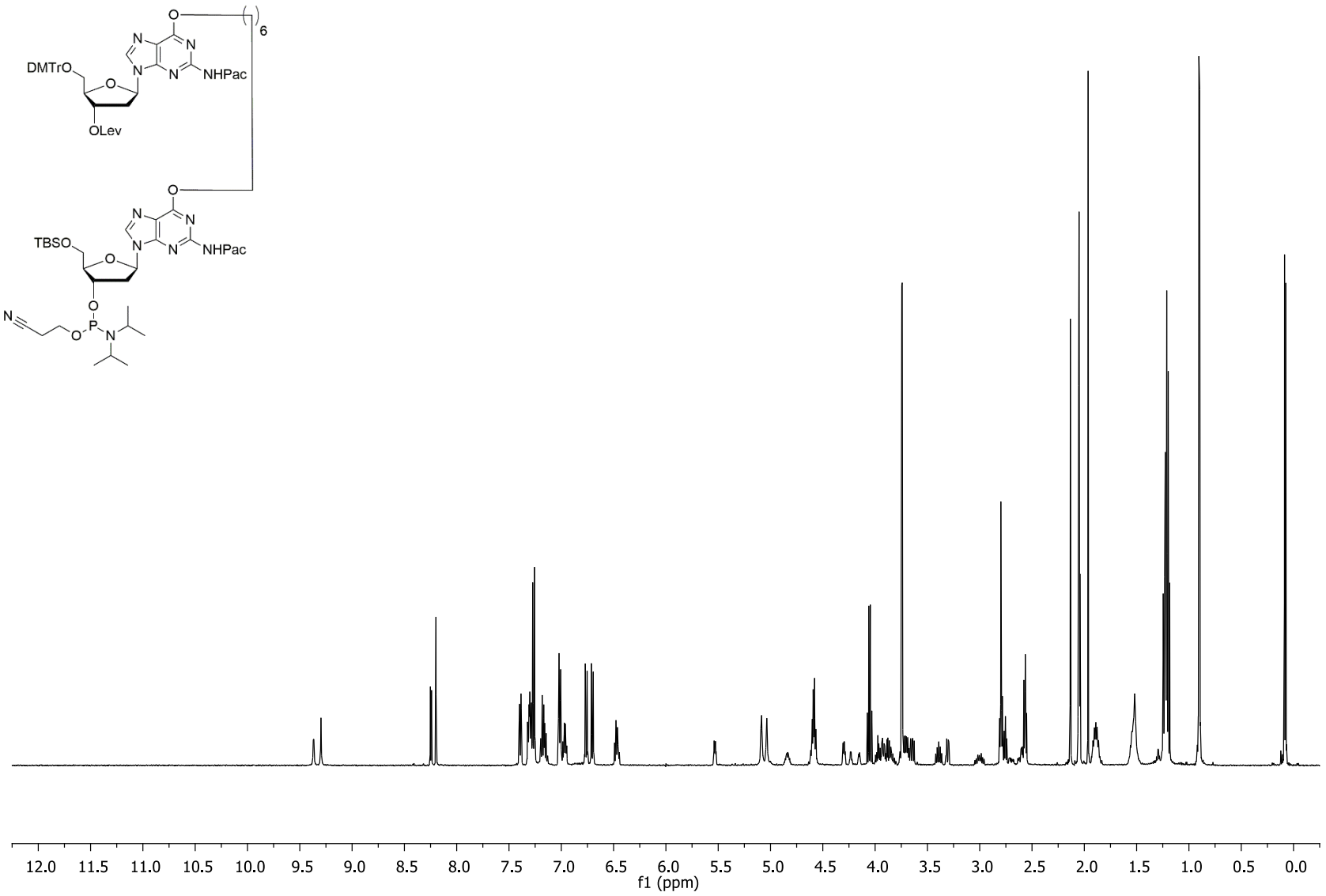


Figure S3.28 202.3 MHz ^{31}P NMR spectrum of compound **3.8b** (d_6-acetone)

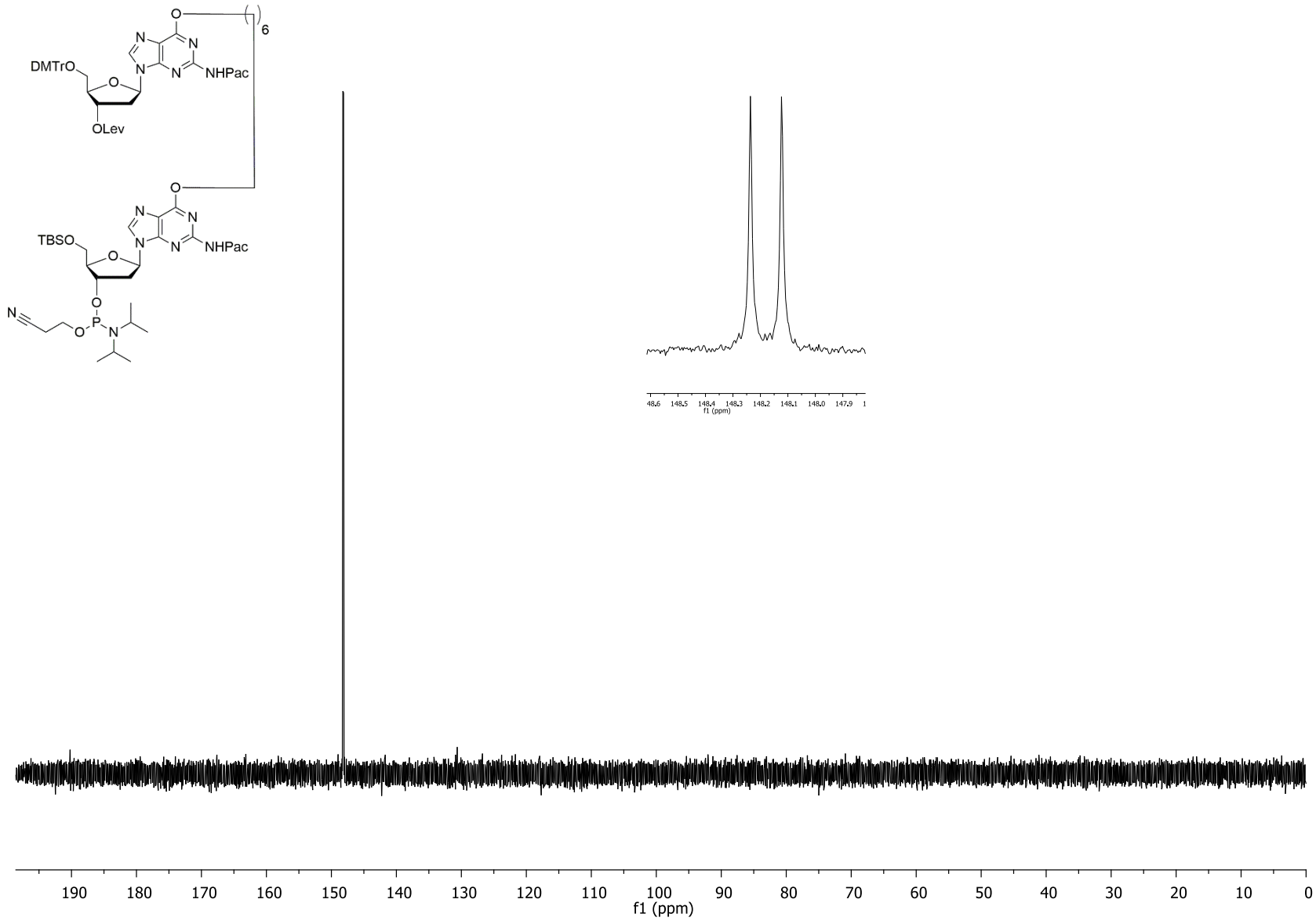


Figure S3.29 125.7 MHz ^{13}C NMR spectrum of compound **3.8b** (d_6-acetone)

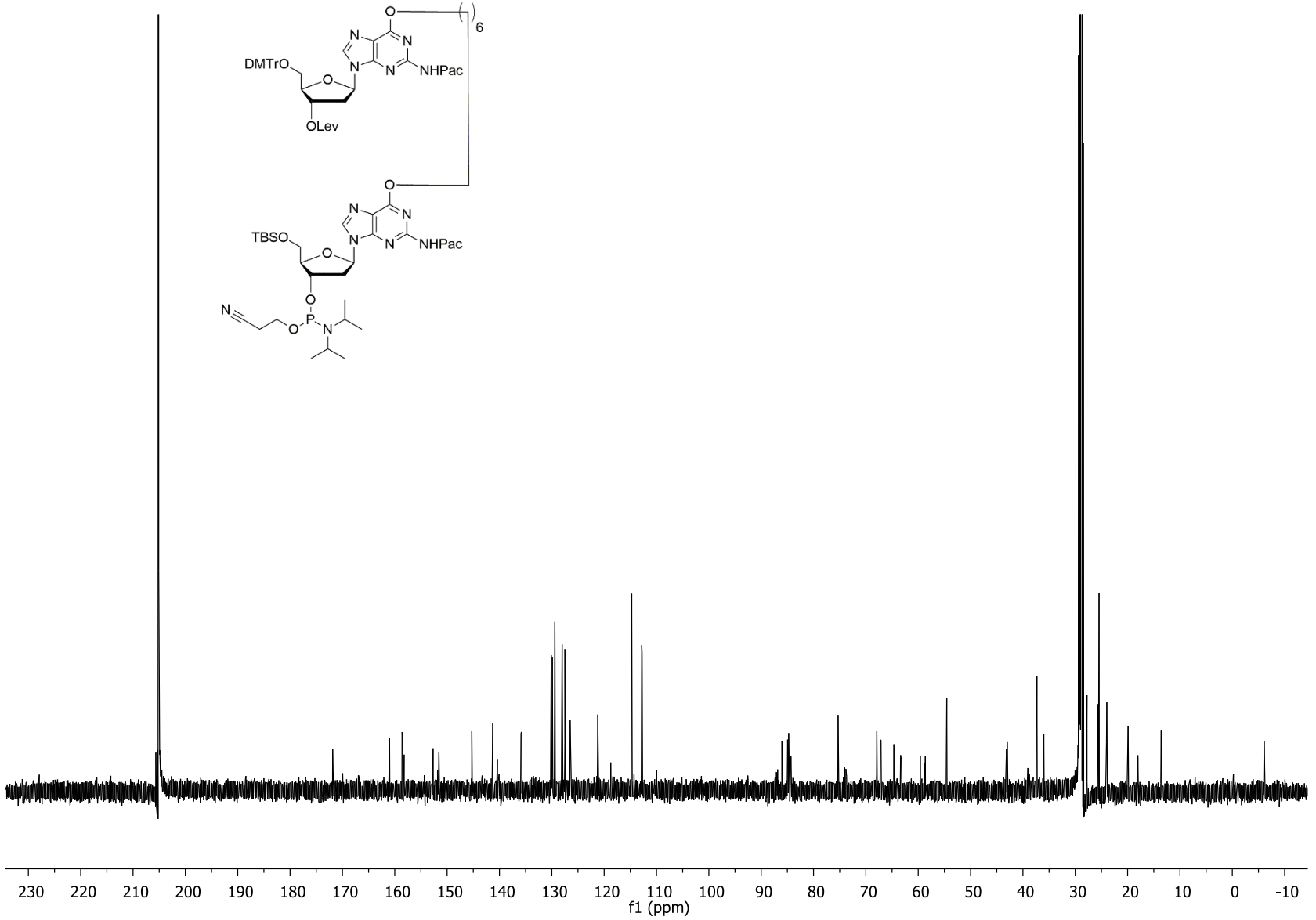


Figure S3.30 20 % Denaturing PAGE of crude (**A**) **C4TWJ**s series: lane 1) Xylene cyanol (XC) and bromophenol blue (BPB) marker dyes, 2) **C4L1**, 3) **C4TWJ1**, 4) **C4L2**, 5) **C4TWJ2**, 6) **C4L3**, 7) **C4TWJ3**, 8) **C4L4** and 9) **C4TWJ4** and (**B**) **C7TWJ**s series: lane 1) **C7L1**, 2) **C7TWJ1**, 3) **C7L2**, 4) **C7TWJ2**, 5) **C7L3**, 6) **C7TWJ3**, 7) **C7L4** and 8) **C7TWJ4**.

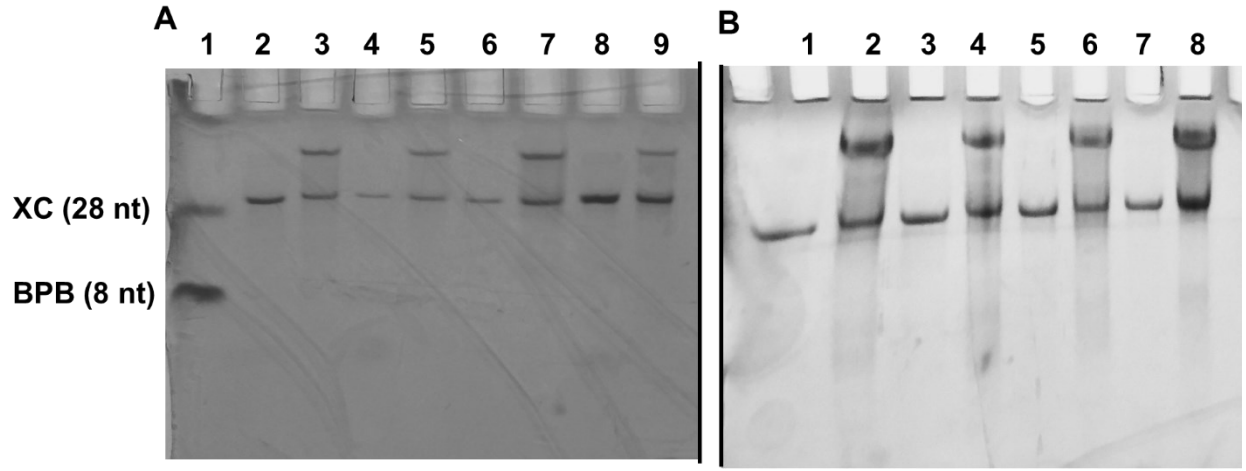


Figure S3.31 20 % Denaturing PAGE of pure **TWJ**s. 1) Xylene cyanol (XC) and bromophenol blue (BPB) marker dyes, 2) **C7TWJ1**, 3) **C7TWJ2**, 4) **C7TWJ3**, 5) **C7TWJ4**, 6) **C4TWJ1**, 7) **C4TWJ2**, 8) **C4TWJ3**, 9) **C4TWJ4.**

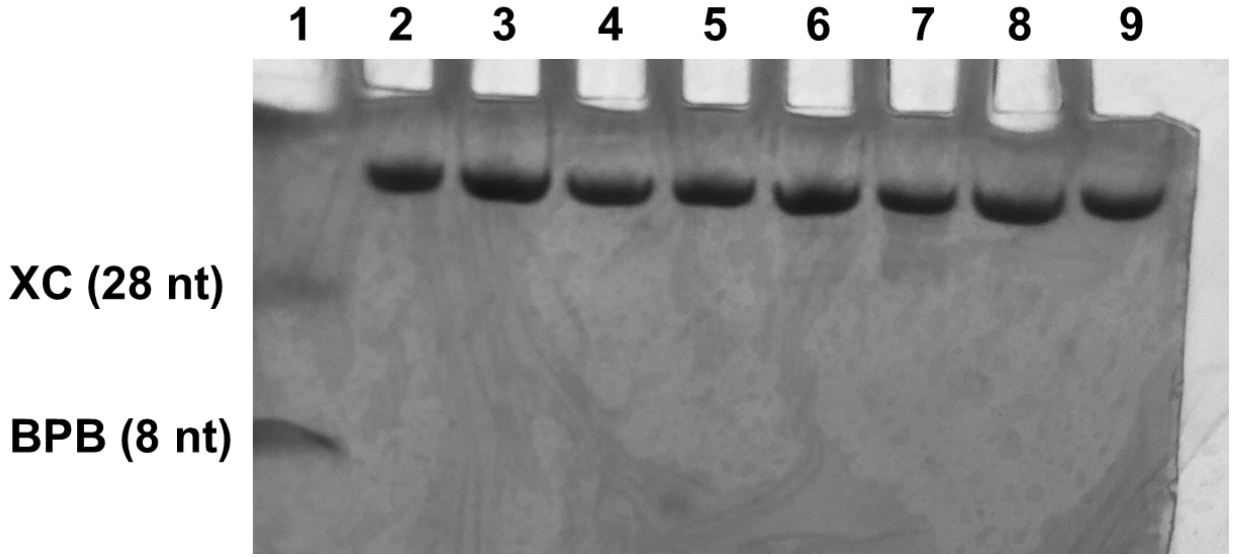


Figure S3.32 ESI-MS Spectra of (**A**) **C4TWJ1**, (**B**) **C4TWJ2**, (**C**) **C4TWJ3** and (**D**) **C4TWJ4**.

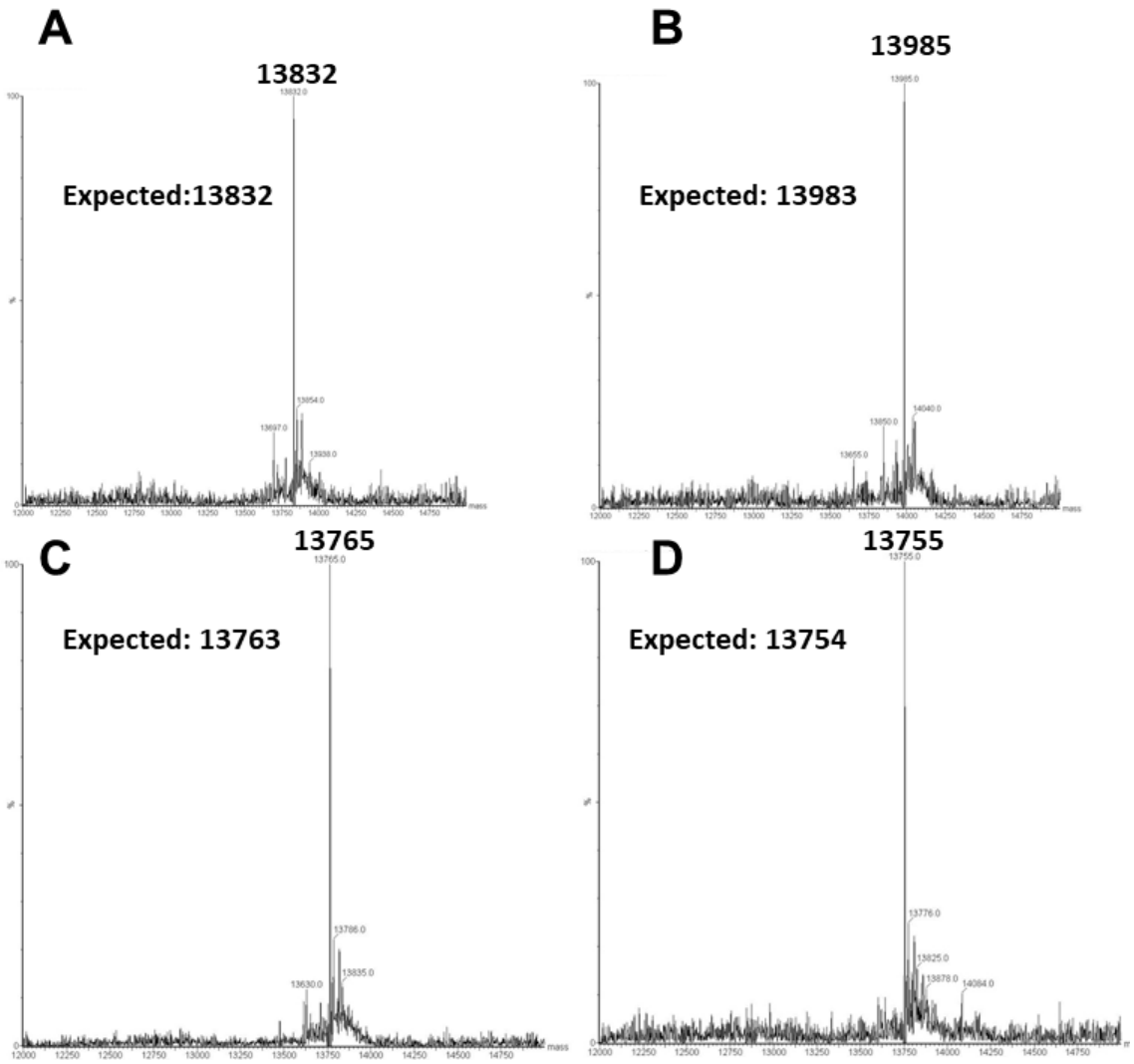


Figure S3.33 ESI-MS spectra of (**A**) **C7TWJ1**, (**B**) **C7TWJ2**, (**C**) **C7TWJ3** and (**D**) **C7TWJ4**.

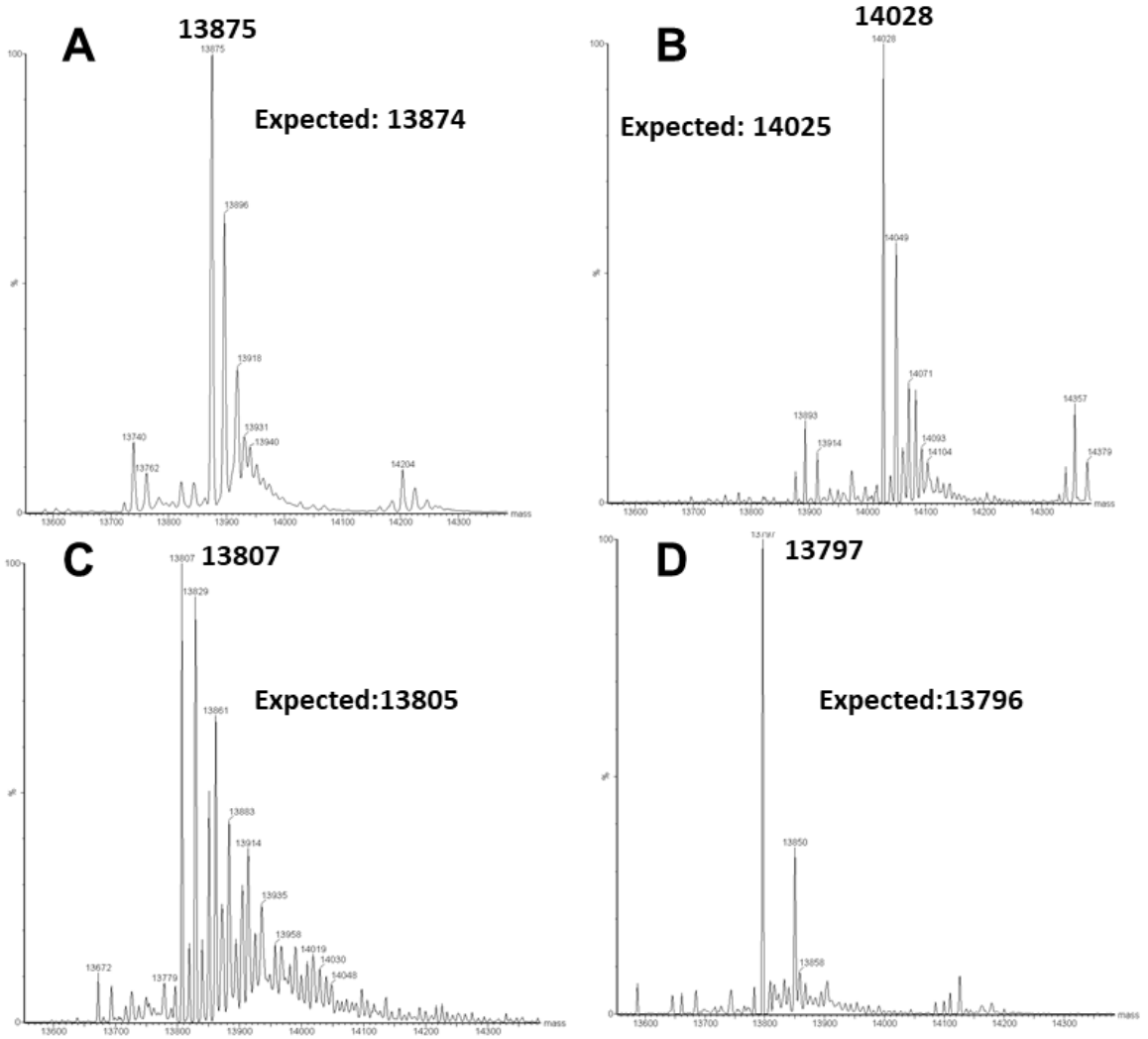


Figure S3.34 Total unhooking of ss**C7TWJ**s. 2 pmol of ss**TWJ** was incubated with either 0 pmol (-), 20 pmol (+) or 60 pmol (+++) of hAGT for 18 h at 37 °C.

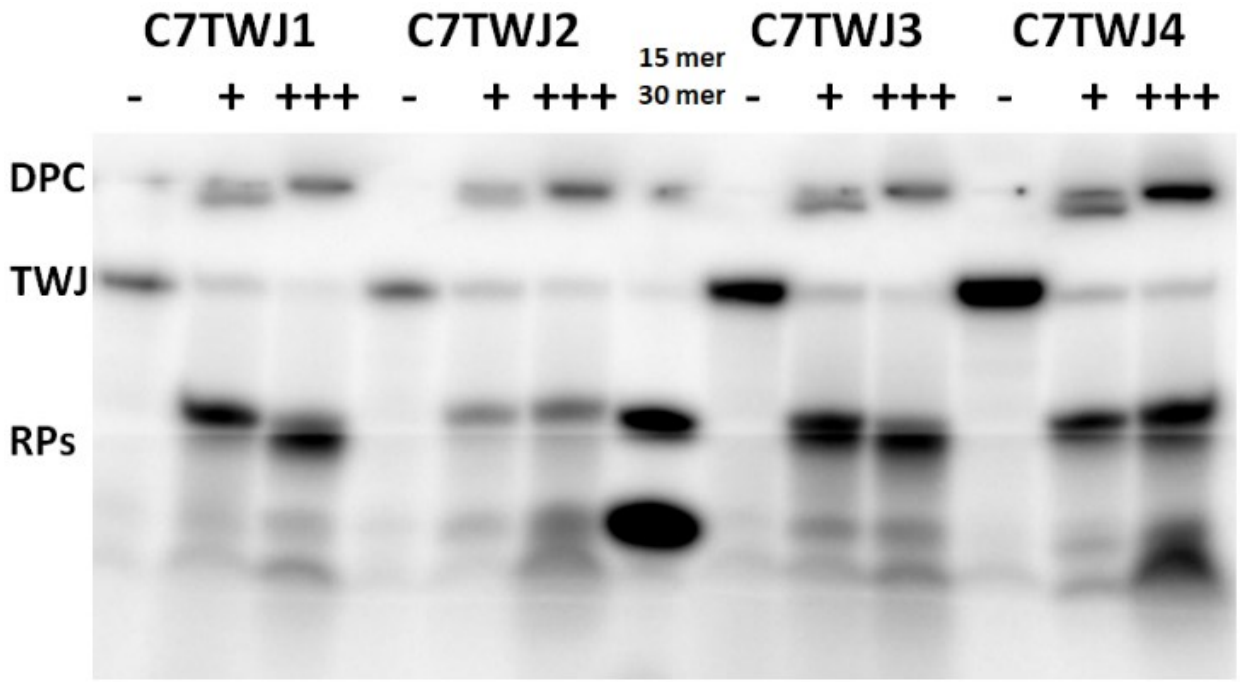


Figure S3.35 UV thermal denaturation profiles for (**A**) **C4** and (**B**) **C7** tetrahedra.

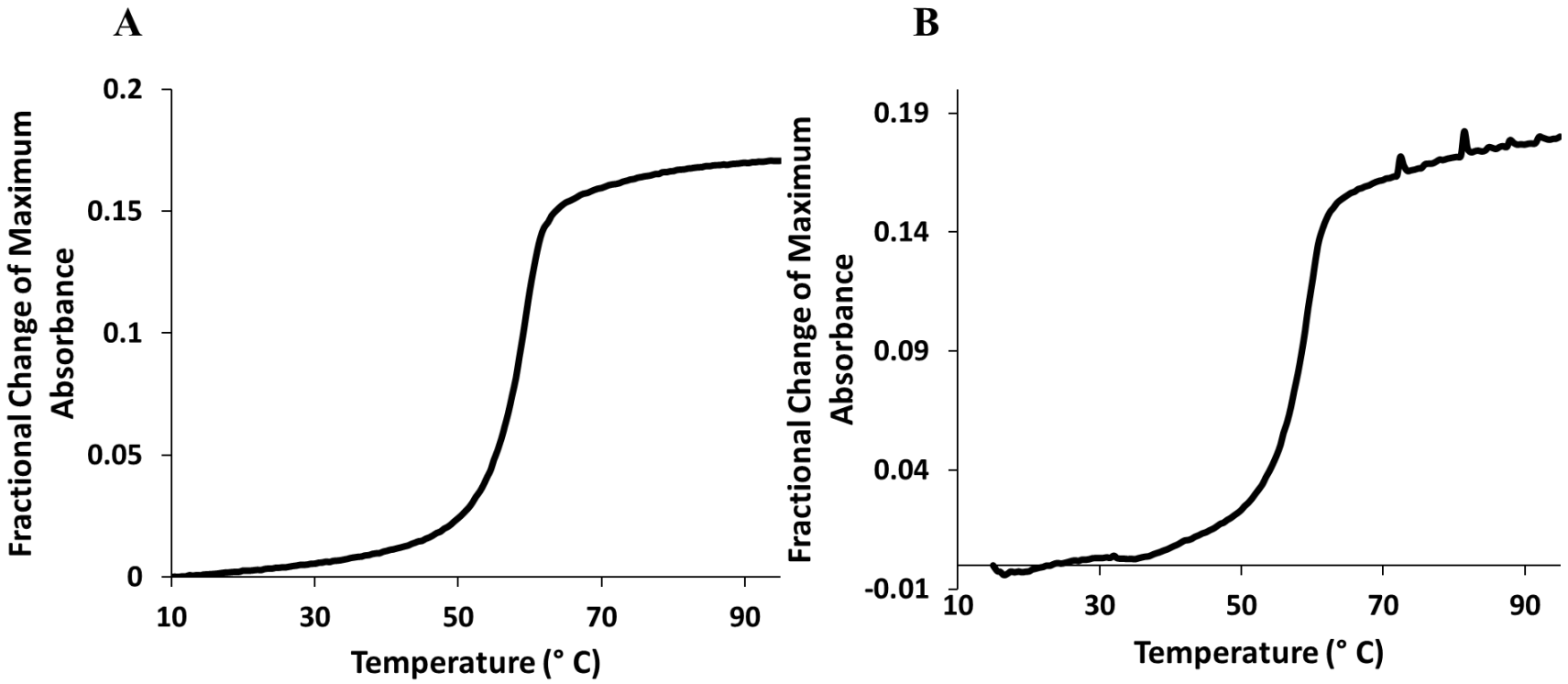


Figure S3.36 CD spectra for **C4** (- - - - -) and **C7** (———) tetrahedra.

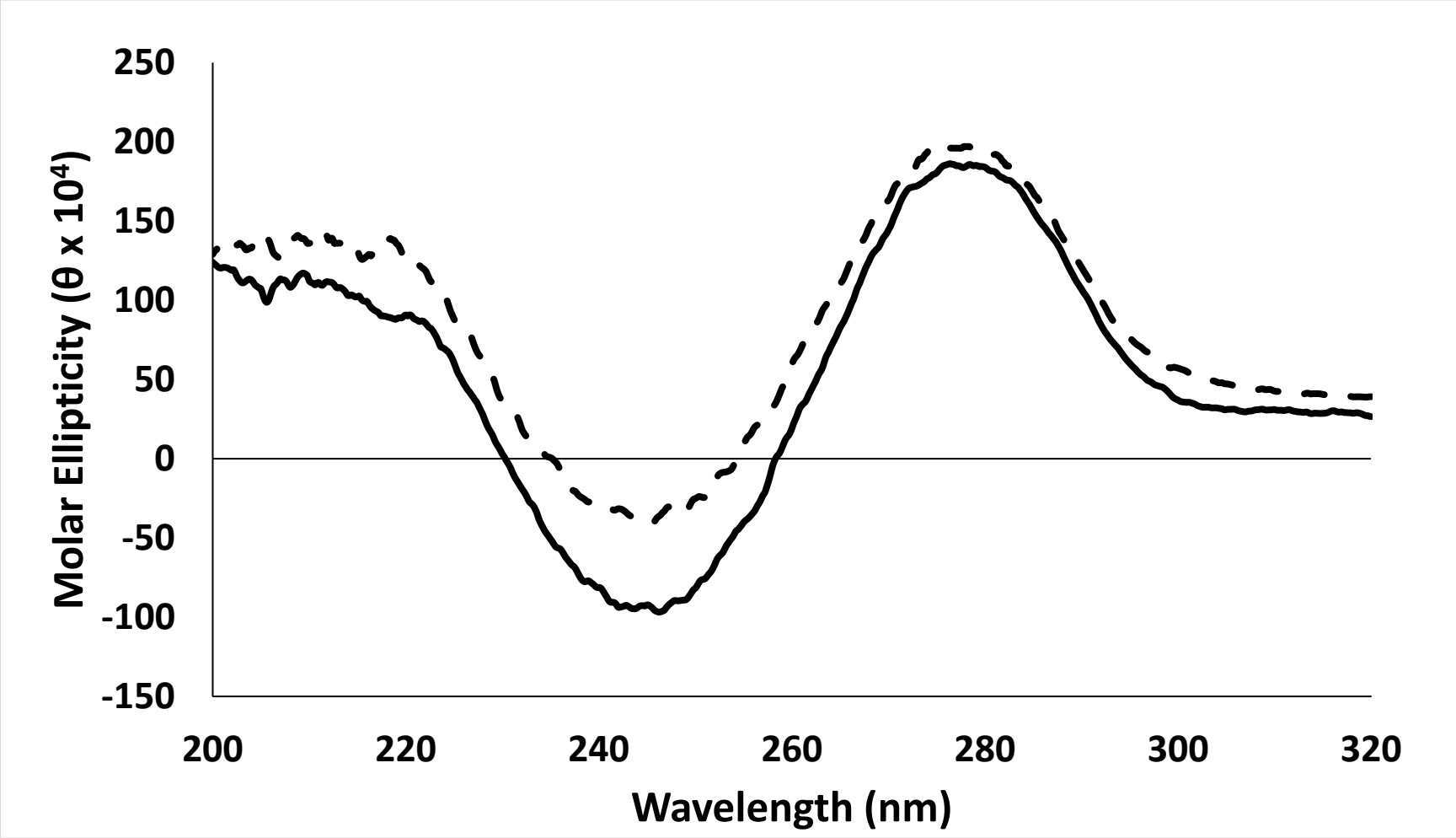


Figure S3.37 10 % Non-denaturing PAGE of each combination of **C7TWJ** oligonucleotides.

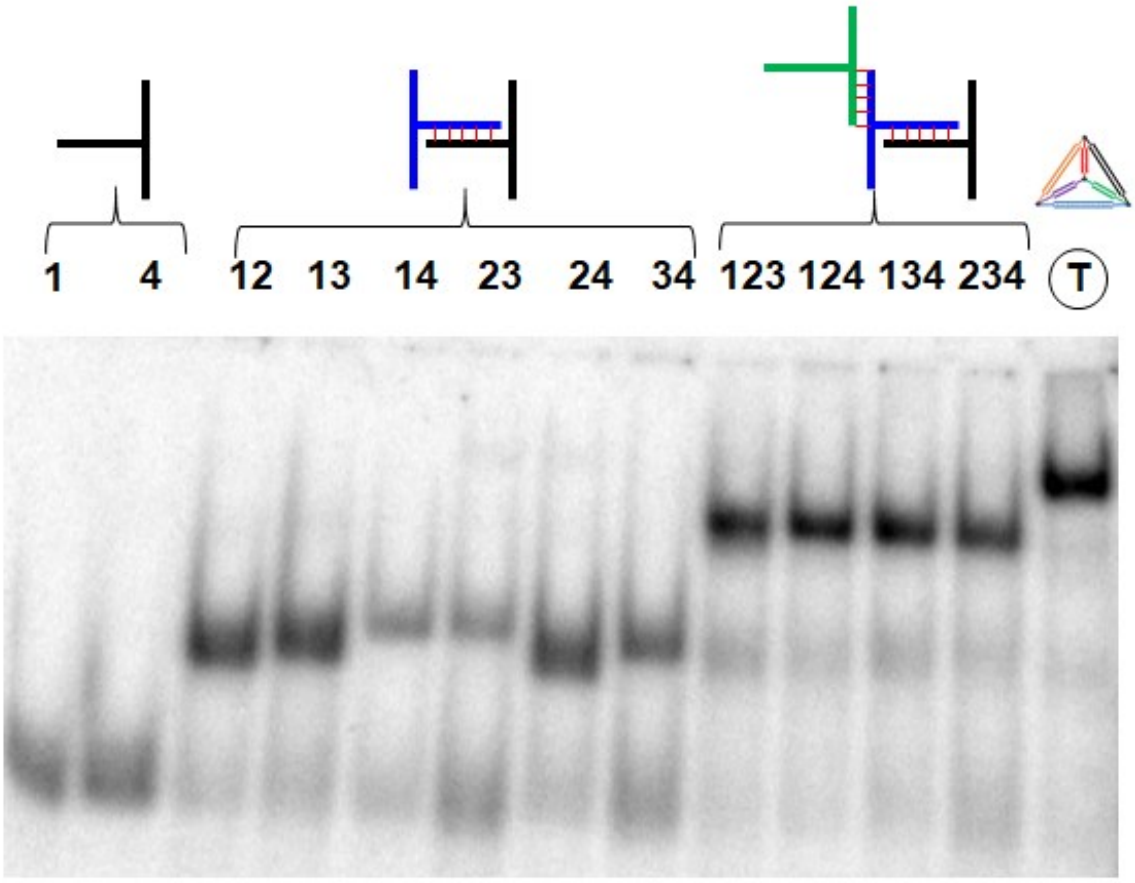


Figure S3.38 Mung bean nuclease (MBN) digest of ss *vs* self-assembled **TWJ**s. Lane 1) ss**C4TWJ**, 2) ss**C4TWJ** + MBN, 3) **C4TWJ1-4**, 4) **C4TWJ1-4** + MBN, 5) ss**C7TWJ**, 6) ss**C7TWJ** + MBN, 7) **C7TWJ1-4** and 8) **C7TWJ1-4** + MBN.

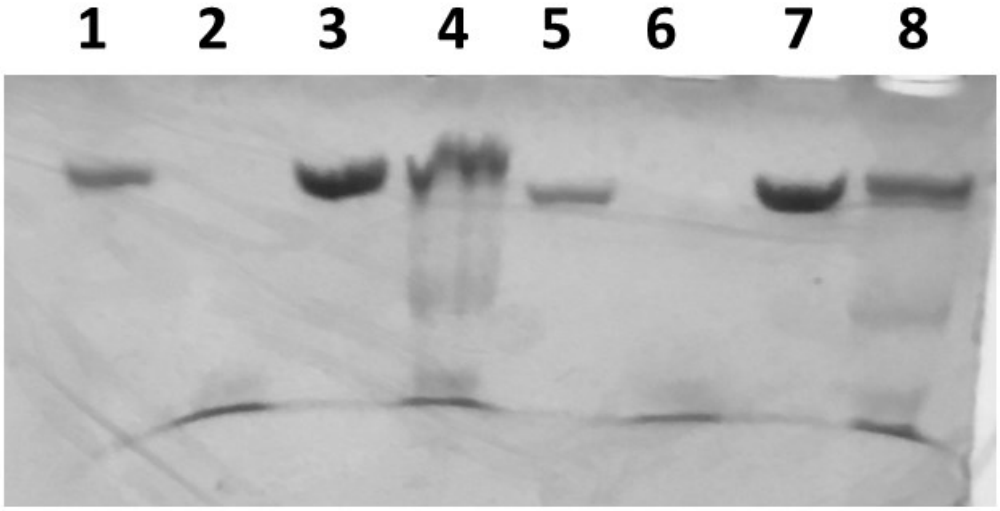


Figure S3.39 Native PAGE of DNA tetrahedra for hAGT unhooking in activity buffer. 20 µL of protein activity buffer was added and samples were incubated at 37 °C overnight. Only one **TWJ** vertex was labelled and hybridized with the other three **TWJs** (unlabelled). In the first lane on the left, three **TWJ** oligonucleotides to form a trimer structure (labelled as **TWJ trimer control**) was included as a reference species with slightly greater electrophoretic mobility by native PAGE relative to the tetrahedron. Lane 1) **C7 TWJ** tetrahedron control, lane 2) **C7TWJ1**, lane 3) **C7TWJ2**, lane 4) **C7TWJ3**, lane 5) **C7TWJ4**, lane 6) **C4TWJ** tetrahedron control, lane 7) **C4TWJ1**, lane 8) **C4TWJ2**, lane 9) **C4TWJ3**, lane 10) **C4TWJ4**. 10 % non-denaturing PAGE was run for 2 h at 200 V at 4 °C.

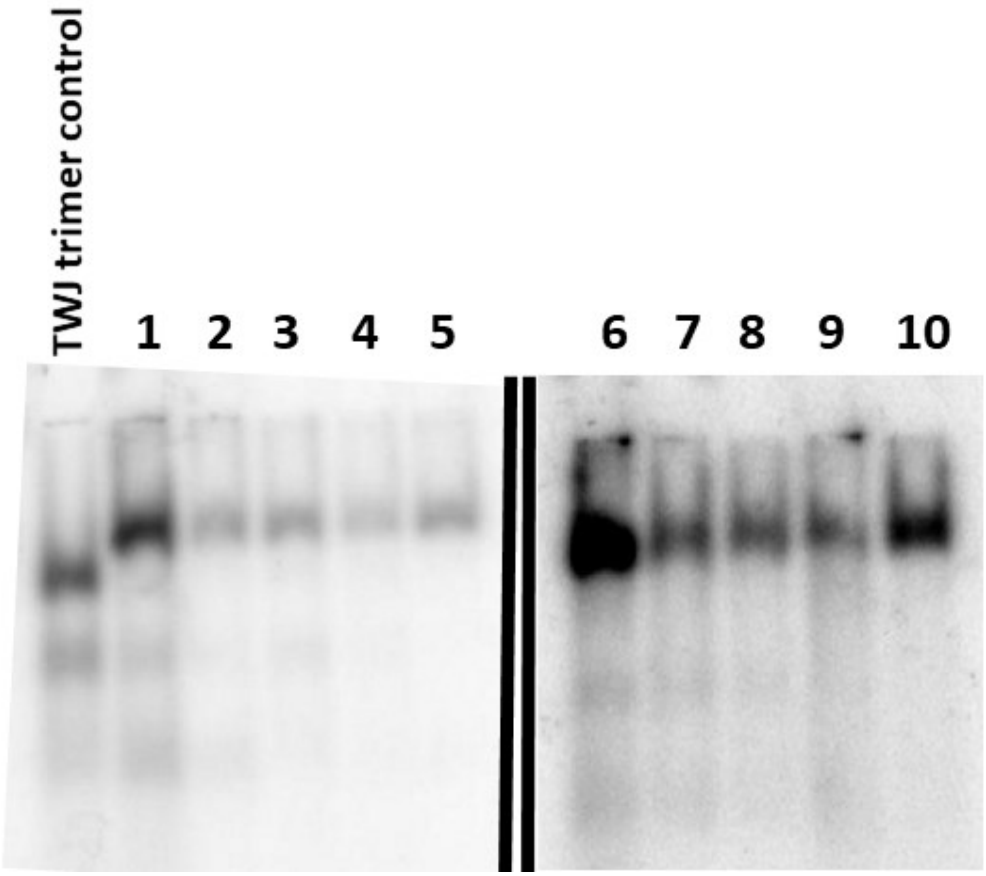


Figure S3.40 Total unhooking of **C7TWJ** tetrahedra by hAGT. Each vertex was separately ^{32}P labelled and incubated with non-radiolabelled complementary **TWJ**s in TAMg buffer. 50 µL (1.25 pmol) of tetrahedron stock solution was incubated with either 0 pmol (-), 12.5 pmol (+) or 37.5 pmol (+++) of hAGT at 37 °C for 18 h. To each reaction was added 50 µL of stop buffer and samples were loaded onto a 20 % denaturing PAGE. ssDNA 45-mers were loaded as positive controls (left pane) and 30mer and 15mer positive control repair products were loaded in the center lane. The labels on the left refer to repair products (**RPs**), three-way junctions (**TWJ**) and DNA-protein cross-linked (**DPC**) species produced by the action of the hAGT upon the **TWJ**s.

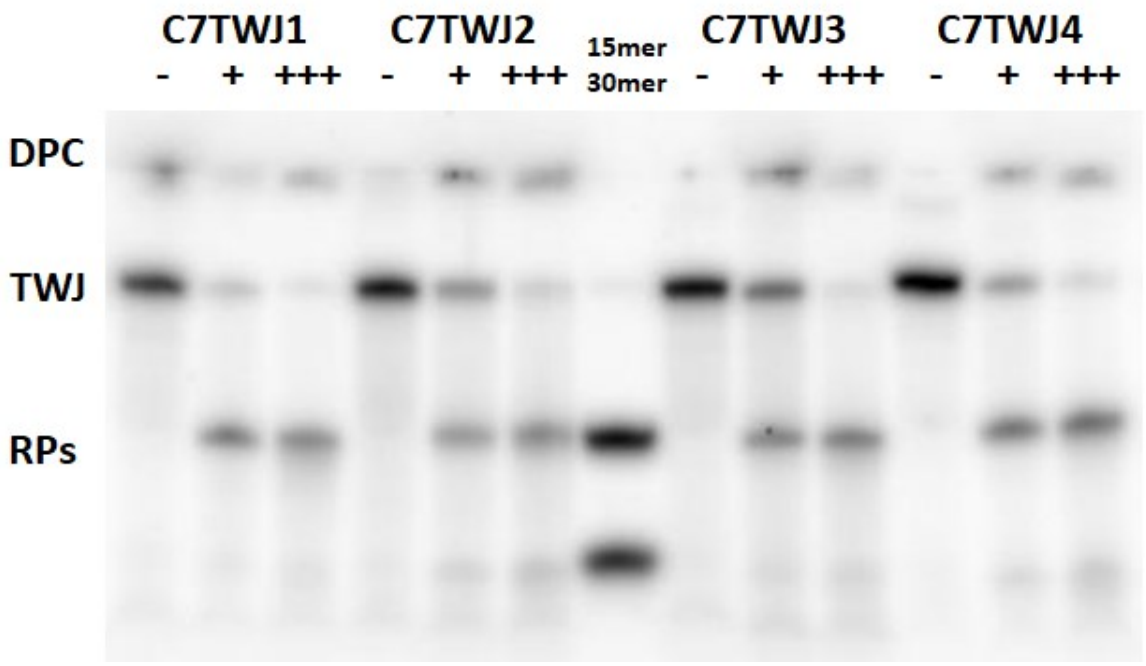


Figure S3.41 FBS assays on ss**C4TWJ** and ss**C7TWJ** and their respective TDNs.

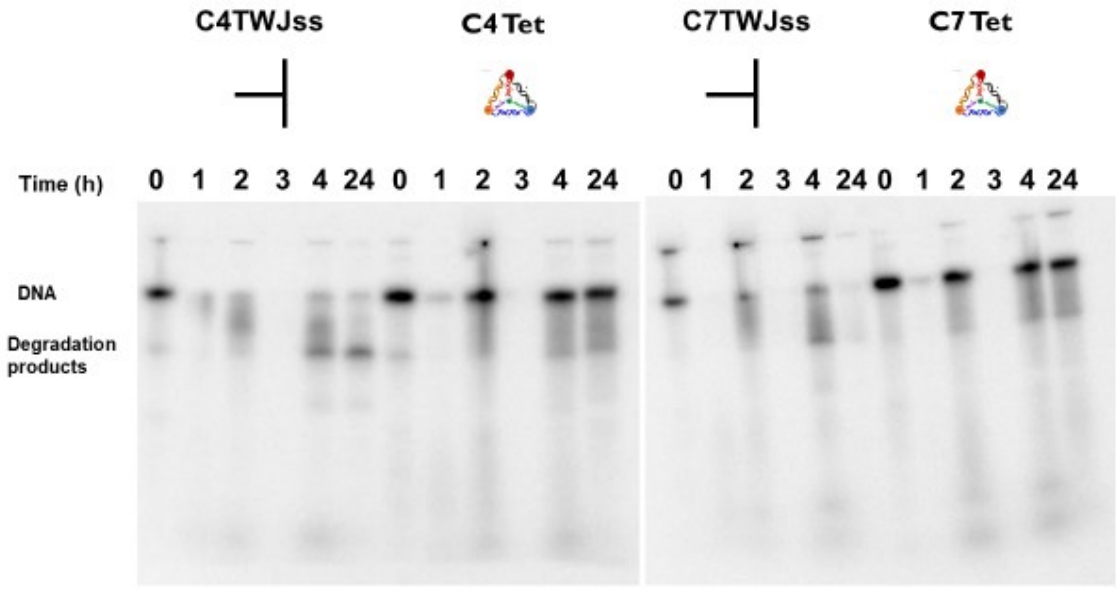


Figure S3.42 Time course assays of DNA tetrahedra prepared with **TWJ**s. A) Butylene series B) Heptylene series. A Master mix of 133 uL of 25 nM oligo stocks (3.3 pmol) with 33.3 pmol hAGT was prepared and 10 uL aliquots were boiled for 5 mins, to which was added 10 uL of stop buffer**.** 20 % PAGE (19:1) run at 450 V for 1.5 h.

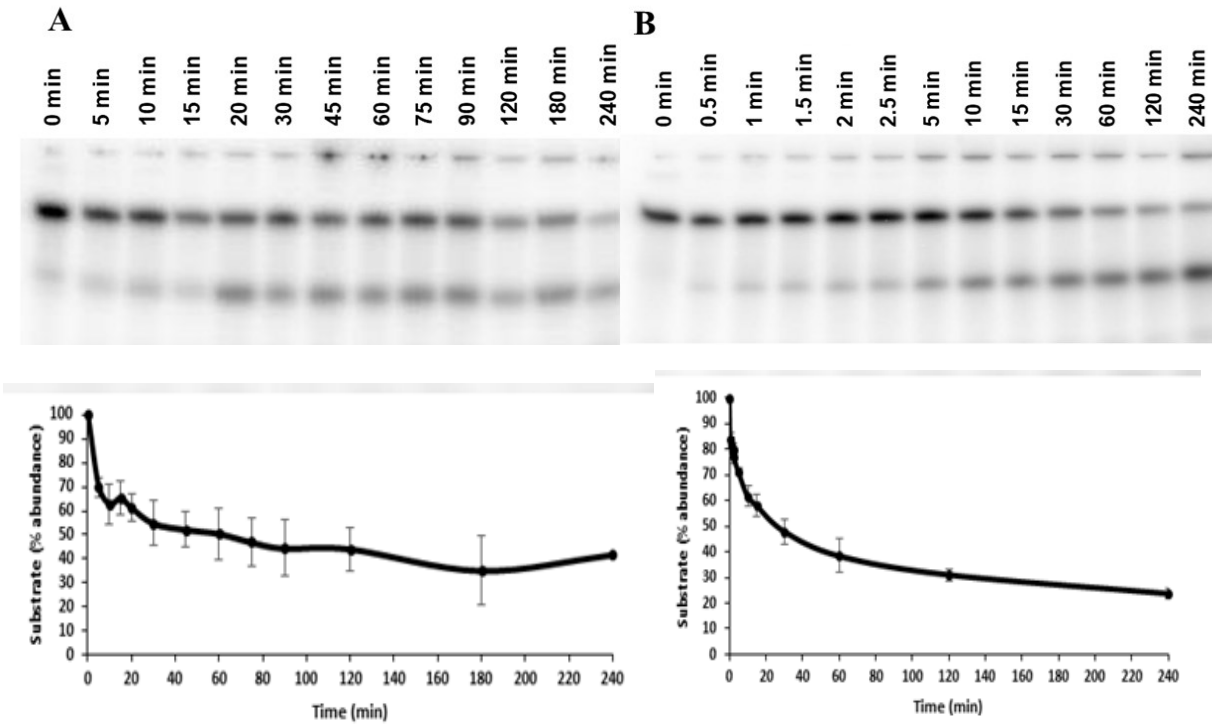


Figure S3.43 Magnesium dependence of DNA tetrahedra formation.

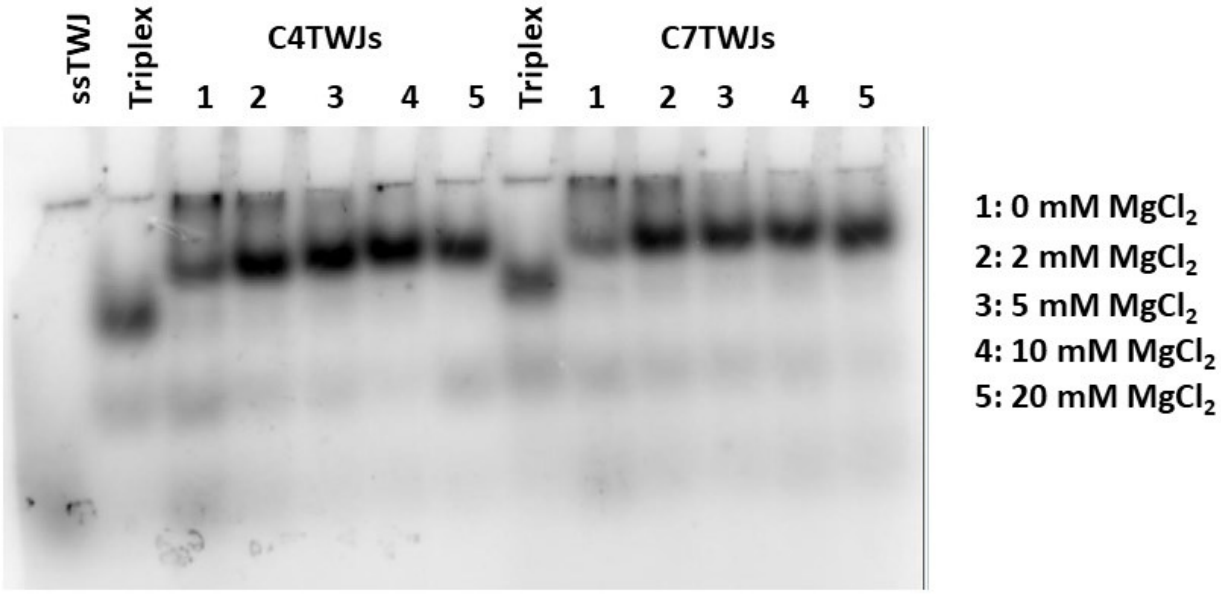


Table S3.1 Mass spectrometry of **TWJ** oligonucleotides

Oligonucleotide	Expected Mass (Da)	Observed Mass (Da)
C4TWJ1	13832	13832
C4TWJ2	13983	13985
C4TWJ3	13763	13765
C4TWJ4	13754	13755
C7TWJ1	13874	13875
C7TWJ2	14025	14028
C7TWJ3	13805	13807
C7TWJ4	13796	13797

Chapter 4

Repair of DNA Alkylation in the G-quadruplex Formed by the *c-myc* Promoter Region by O^6-Alkylguanine DNA Alkyltransferases

Copp, W. and Wilds, C. J.*

To be submitted to:

ACS Chemical Biology, 2021

Abstract

O^6-alkylguanine DNA alkyltransferase (AGT) is a protein found in most organisms whose role is removal of alkylation at the *O*^{6} position of 2'-deoxyguanosine (dG) which if left unrepaired can be mutagenic. The promiscuous human homologue is capable of repairing larger lesions such as *O*^{6}-benzyl guanine, 4-oxo-4-(3-pyridyl)butyl-guanine, interstrand and intrastrand alkylene cross-links at the *O*^{6} position of dG however repair abilities vary between AGT variants from different species. G-quadruplexes are a nucleic acid tetrameric structure which can be formed in guanine-rich sequences in vivo such as the *c-myc* oncogene promoter region. In this study we investigated the thermal stability and the effects on global structure of various alkyl(ene) damage at the *O*^{6} position of dG residues on the parallel-stranded *c-myc* G-quadruplex by UV thermal denaturation and circular dichroism spectroscopy, respectively. All lesions were destabilizing to different extents, and if there were multiple lesions present the global topology was affected. We also evaluated the susceptibility to dealkylation at the *O*^{6} position of dG by human, *E. coli* and hybrid AGTs. Removal of the alkyl lesions was observed by all proteins studied, with different proficiencies depending on the nature of the lesion with respect to the protein acting on it. This study has consequences for *c-myc* gene regulation as well a potential for a G-quadruplex conformational switch in with respect to DNA-protein nanotechnology.

4.1 Introduction

DNA is constantly subject to damage from various endogenous and exogenous alkylating agents.[4] Alkylation at the O^6 position of guanine is particularly mutagenic and potentially carcinogenic due to preferential insertion of thymine across from guanine by DNA polymerases.[146] O^6-alkylguanine DNA alkyltransferase (AGT) is a protein found in most organisms whose role is irreversible transfer of alkyl lesions to Cys145 (human numbering) by acting on the alkyl group's α carbon. After removal of the alkylation damage, the deactivated AGT protein is then degraded by the ubiquitin pathway.[19] Human AGT (hAGT) is a protein which can act upon an assortment of alkyl damage from longer alkyl chains, bulky lesions as well dG-O^6-alkylene-O^6-dG inter or intrastrand cross-links (ICLs and IaCLs, respectively).[35,39,151] These proteins are capable of acting on the O^4 position of thymine however repair proficiency varies between AGT homologues. Differences in repair are notable between hAGT and those found in *E. coli*. Despite AGT's largely conserved amino acid residues among species[151], *E coli*'s two AGT proteins OGT and Ada-C have demonstrated different repair abilities relative to hAGT. This has been attributed to stark differences in active site residues: Ada-C has a very compact active site[47] and there is no structural information on OGT to date. It is understood that the large accommodating tunnel-like active site of hAGT allows it to act upon larger alkyl lesions such as O^6-benzylguanine and 4-oxo-4-(3-pyridyl)butyl-guanine, ICLs and IaCLs due to the presence of pro140 (human numbering). In contrast, for reasons to be elucidated *E. coli*'s AGT's are efficient at repairing damage at the O^4 position of thymine; a poor substrate for hAGT. Our lab has engineered a chimera (hOGT), which has 20 residues from OGT's active site replacing the corresponding ones from the human protein with retention of Pro140 from hAGT. hOGT exhibits a hybrid of hAGT and OGT's repair abilities for mechanistic insights.[51]

Some chemotherapeutic regimens employ bi-functional alkylating agents which introduce intra- or interstrand cross-links in DNA of cancer cells as a means to halt their proliferation. It is possible to observe the hydrolysis product as well resulting in the formation of a monoadduct. Various repair pathways contribute to restoring the information content of DNA by removal of damage, allowing the cell to survive. Paradoxically, while DNA damage can lead to cancer, hAGT has actually been implicated in chemotherapeutic resistance to alkylating agents that act upon the O^6-atom of dG.[190] In this case, removal of DNA damage by hAGT allows the tumour cells to proliferate as opposed to undergoing apoptosis.

The polymorphism of DNA is exhibited by its ability to form a wide array of noncanonical structures such as triplexes,[191] *i*-motifs,[119] and parallel-stranded duplexes.[192] G-quadruplexes are of interest due to their presence in regions of the human genome such as telomeres and

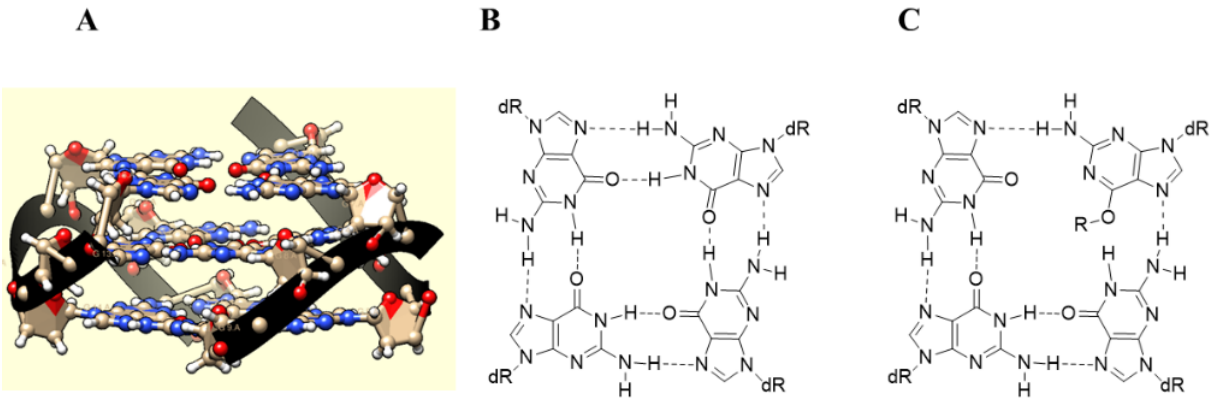


Figure 4.1 Depiction of G-quadruplexes. A) Parallel-stranded *c-myc* G-quadruplex topology. PDB: 2LBY. Coordinates were used from ref [205] B) Hydrogen bonding of G-quadruplexes invokes both the Watson-Crick and Hoogsteen faces of guanine and is stabilized by monovalent cations such as potassium. C) Methylation of the O^6 of guanine is destabilizing due to interruption of G-quadruplex's hydrogen-bond network. dR = 2'-deoxyribose.

promoter region of oncogenes.[127,193] G-quadruplexes themselves demonstrate vast polymorphism examples of which include i) directionality (parallel *vs*. antiparallel), ii) molecularity and iii) base conformation (**Figure 4.1A**). G-quadruplexes are stabilized by stacked G-quartets which are formed in guanine rich sequences. The G-quadruplex tetrad's hydrogen bonding pattern invokes both the Watson-Crick and Hoogsteen faces of the guanine nucleobases (**Figure 4.1B**) and is stabilized by monovalent cations. The intramolecular G-quadruplex formed by the promoter region of the *c-myc* gene has garnered interest due to its implication in cancer because of its role in expression of transcription activators involved in cell growth and repressors for growth arrest as well as apoptosis.[130] The interplay of these promoter regions between duplex and G-quadruplex has an important role on gene regulation.[194] The highly conserved NHE III_1 region has been shown to be involved in 80-95 % of *c-myc* gene expression.[195,196] Truncated Pu27, *c-myc*-2345 18mer, has been shown to be the primary G-quadruplex *in vitro* and it forms a propeller-type parallel-stranded as determined by NMR.[197] Hellman *et al.* have shown that hAGT is able to remove methyl adducts at the O^6 position of guanine in telomeric DNA G-quadruplexes at various positions (**Figure 4.1C**).[136] It is believed that the role of alkylation by endogenous and exogenous

agents has a role in oncogene expression: destabilization of the *c-myc* G-quadruplex would promote gene expression.[128,129]

While the aforementioned lesions have been studied in duplex DNA, to date little is known on their biophysical effects nor the ability of AGTs to repair the bulkier alkylene and IaCL lesions at the O^6 position of dG in G-quadruplexes. Thus, in this study the biophysical characterization and reaction with AGTs of these lesions was investigated in the parallel-stranded *c-myc* G-quadruplex (**Table 4.1** and **Figure 4.2**) which has been well characterized and has applications in biology as well as nanotechnology. It was found that when there is one alkyl(ene) lesion present (monoadduct), it is thermally destabilizing to the G-quadruplex essentially independent of the length of the lesion. The effect of alkylating the internal G_7 however is significantly more destabilizing to the *c-myc* G-quadruplex than when the external G_6 is modified. Those containing two alkylene monoadducts and the IaCLs were shown to be unstable at low K^+ concentratraions

Table 4.1 Sequences studied in this project. X= O^6-Methyl dG, Y= O^6-4-hydroxybutyl dG, Z= O^6-7-hydroxyheptyl dG. G4G and G7G are butylene and heptylene linked IaCLs lacking a phosphate linkage, respectively.

Oligonucleotide	Sequence (5'-3')
ODN-4.1	5'-$A_1G_2G_3$ $G_4T_5G_6$ $G_7G_8G_9$ $A_{10}G_{11}G_{12}$ $G_{13}T_{14}G_{15}$ $G_{16}G_{17}G_{18}$-3'
ODN-4.2	5'-AGG GT**X** GGG AGG GTG GGG-3' X=dG^{OMe}
ODN-4.3	5'-AGG GTG **X**GG AGG GTG GGG-3' X=dG^{OMe}
ODN-4.4	5'-AGG GT**X** **X**GG AGG GTG GGG-3' X=dG^{OMe}
ODN-4.5	5'-AGG GT**Y** GGG AGG GTG GGG-3' Y=dG^{C4OH}
ODN-4.6	5'-AGG GTG **Y**GG AGG GTG GGG-3' Y=dG^{C4OH}
ODN-4.7	5'-AGG GT**Y** **Y**GG AGG GTG GGG-3' Y=dG^{C4OH}
ODN-4.8	5'-AGG GT**Z** GGG AGG GTG GGG-3' Z=dG^{C7OH}
ODN-4.9	5'-AGG GTG **Z**GG AGG GTG GGG-3' Z=dG^{C7OH}
ODN-4.10	5'-AGG GT**Z** **Z**GG AGG GTG GGG-3' Z=dG^{C7OH}
ODN-4.11	5'-AGG GT **G4G** GG AGG GTG GGG-3'
ODN-4.12	5'-AGG GT **G7G** GG AGG GTG GGG-3'

however stable at physiological levels. The structural integrity of the propeller-type parallel-stranded G-quadruplex was determined to be intact when a single lesion was present as assessed by CD spectroscopy. A hybrid G-quadruplex is believed to be forming when there were two alkylated sites. All sequences studied underwent the dealkylation reaction with the AGT variants however the nature of the alkylation damage determined which proteins were active towards the substrate. The IaCLs were more resistant to repair relative to their duplex DNA counterparts however moderate repair was still observed.

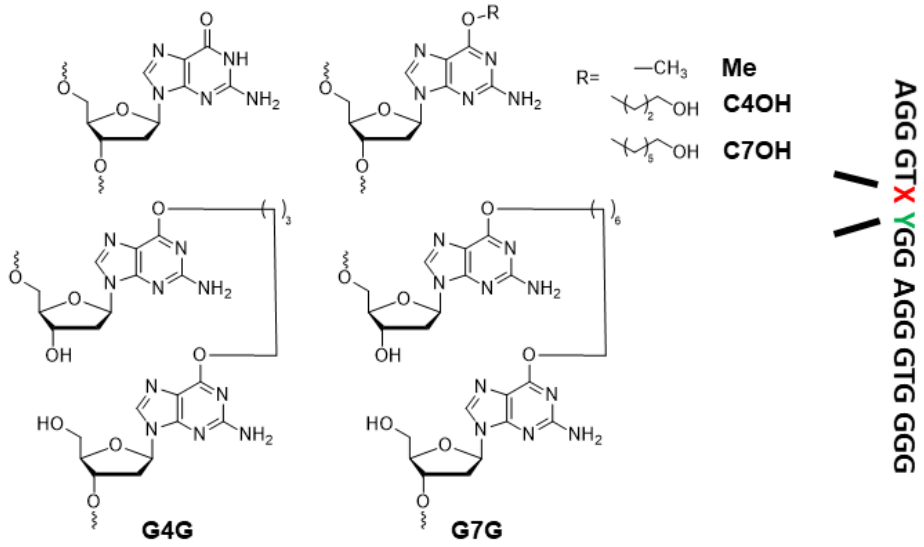


Figure 4.2 Alkyl(ene) lesions explored in this study.

4.2 Results and Discussion

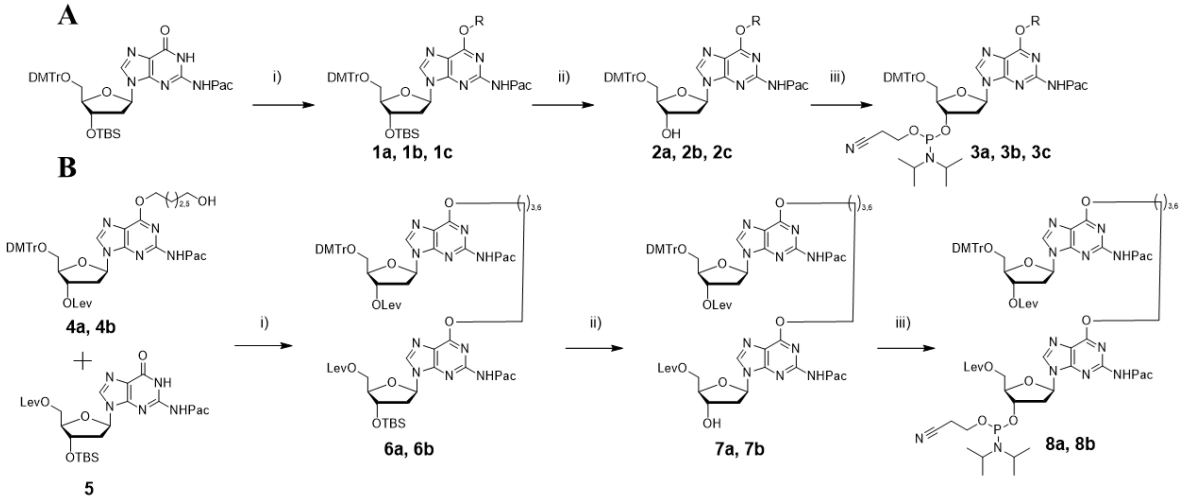
4.2.1 Synthesis of G-quadruplex forming oligonucleotides containing alkylene lesions.

The synthesis of O^6-alkylated 2'-dG phosphoramidites containing methyl and hydroxyalkylene adducts is depicted in **Scheme 4.1A**. Divergent synthesis was performed commencing with 3'-*O*-(*tert*-butyldimethylsilyl)-5'-*O*-(4,4'-dimethoxytrityl)-N^2-phenoxyacetyl-2'-deoxyguanosine which was synthesized as described previously by a silylation reaction of commercially available 5'-*O*-(4,4'-dimethoxytrityl)-N^2-phenoxyacetyl-2'-deoxyguanosine.[40] A Mitsunobu reaction was then performed with methanol, or phenoxyacetyl protected alkylene diols with suitable carbon chain lengths to mimic the hydrolysis adducts formed chemotherapeutic agents busulfan and hepsulfam to yield compounds **4.1a**-**4.1c** with yields of 43 %, 67 % and 60 % respectively. The *tert*-butyldimethylsilyl (TBS) groups were then removed with a tetrabutylammonium fluoride (TBAF)

treatment to obtain compounds **4.2a**-**4.2c**. The free 3'-hydroxyl groups were then reacted with phosphitylating agent to gives nucleoside phosphoramidites **4.3a**-**4.3c** in moderate yields. IaCL phosphoramidites were synthesized utilizing similar methodology previously employed by our group (**Scheme 4.1B**). Compounds **4.4a** and **4.4b** were acquired by Mitsunobu reactions as above however an extra deprotection step to unmask the nucleophilic alcohol by removal of *tert*-butyldiphenylsilyl groups was necessary. The second monomer was synthesized by removal of the dimethoxytrityl group of 5'-*O*-(4,4'-dimethoxytrityl)-N^2-phenoxyacetyl-2'-deoxyguanosine with *para*-toluenesulfonic acid followed by protection of the 5'-hydroxyl group with a levulinyl group to produce compound **5** (see **Section 4.4.2 Scheme S.1**). The levulinyl moiety was chosen due to its chemical robustness as a protecting group for the subsequent synthetic steps. Coupling of the monomeric species **4.4a, 4.4b** and **4.5** was done by a second Mitsunobu reaction to acquire compounds **4.6a** and **4.6b** with yields of 50 % and 60 % respectively. The TBS groups were removed with TBAF and the nucleosides phosphitylated to yield IaCL phosphoramidites **4.8a** and **4.8b**. ^{31}P NMR confirmed the presence of nucleoside phosphoramidites for compounds **4.3a-c** and **4.8a-b** in the diagnostic 148-150 ppm region of the spectra. Interestingly, while two signals were observed for the methyl and hydroxyalkylene adduct series, characteristic of the nucleoside diastereomers, only one signal was observed for the dimeric phosphoramidites.

Oligonucleotides ODNs **4.1-4.12** were then synthesized by standard β-cyanoethylphosphoramidite chemistry with increased concentration and coupling times (5-10 minutes) for the modified nucleoside phosphoramidites. The oligonucleotides were then cleaved from solid support and deprotected with NH_3(*aq*):EtOH (3:1) at 55 °C for minimum of 4 hours. The crude oligonucleotides were purified by preparatory PAGE and the desired band corresponding to the full length product, once well resolved from the failure sequences, was then extracted and desalted. The oligonucleotides were characterized by ESI-MS and their expected masses were well correlated with the deconvoluted masses observed (see **Section 4.4.2 Table S4.1**).

Scheme 4.1 A) Synthesis of compounds **4.3a, 4.3b and 4.3c** where R= Me, -$(CH_2)_4$-OPac, -$(CH_2)_7$-OPac. B) Synthesis of compounds **4.8a** (**G4 IaCL**) and **4.8b** (**G7 IaCL**).



Reagents and conditions: i) MeOH **(a-series** compounds)**,** 4-(phenoxyacetyl)butan-1-ol (**b-series** compounds), or 7-(phenoxyacetyl)heptan-1-ol (**c-series** compounds), DIAD, PPh_3, dioxane, r.t., 18 h iii) TBAF (1 M in THF), THF, r.t., 20 min iv) Cl-P(OCE)N*i*Pr_2, DIPEA, DCM, r.t., 30 min. Reagents and conditions: i) DIAD, PPh_3, dioxane, r.t., 18 ii) TBAF (1 M in THF), THF, r.t., 30 min iii) Cl-P(OCE)N*i*Pr_2, DIPEA, DCM, r.t., 30 min.

4.2.2 Influence of O^6-alkylated guanines on G-quadruplex thermal stability.

Next, the influence of O^6-alkylated guanines on G-quadruplex stability was evaluated by UV thermal denaturation experiments. Positions G_6 and G_7 of the *c-myc* promoter regions were selected non-arbitrarily, as we were interested in whether the 4th guanine when alkylated (which has shown to entropically stabilize the G-quadruplex)[198] would influence the G-quadruplex's stability. Absorbances at 260 nm and 295 nm were measured as a function of temperature, with ramping the sample temperature upwards and then downwards at 0.5 ºC/min, using the same buffer system as Harkness *et al.*[198] with a low potassium content to obtain melting profiles with transitions in a recordable temperature range. First, we evaluated the effect of O^6-methyl dG in G-quadruplexes (**ODNs 4.2-4.4**, **Figure 4.3A**). It should be noted that this modification has been shown to be thermally destabilizing to G-quadruplexes[136,199] however to our knowledge not in the context of this parallel-stranded propeller type G-quadruplex, nor a G-quadruplex with the ability to insert an auxiliary guanine to further conformationally stabilize the G-quadruplex. Whilst

at both wavelengths we witnessed a hypochromic transition upon heating, we report the 260 nm curves in **Figures 4.3A-C** and T_ms are shown in **Table 2**. O^6-methylation of the internal G_7 (**ODN-4.3**) showed a destabilization of -20 ºC: unsurprising considering disruption to hydrogen bonding on both the Watson-Crick and Hoogsteen face of the guanine residues. When the external G_6 position (not participating in the hydrogen bonding) was methylated (**ODN-4.2**), a notable drop in T_m of 8 ºC was observed: presumably due to the disruption of G-register exchange dynamics. O^6-methylation of external G_6 and internal G_7 of the G-quadruplex (**ODN-4.4**) resulted in significant reduction of thermal stability with an inconclusive transition at lower temperatures when measuring at 260 nm and a significant destabilizing effect of ~ -48 ºC (approximated due to no well-defined lower plateau) when measured at 295 nm (**Figure 4.3D**).

We next looked at the lesions that mimic the hydrolysis products from the anticancer alkylating agents busulfan and hepsulfam (**ODNs 4.5-4.7**). Inspection of those with hydroxybutylene linkers showed a destabilizing effect comparable to those containing methyl

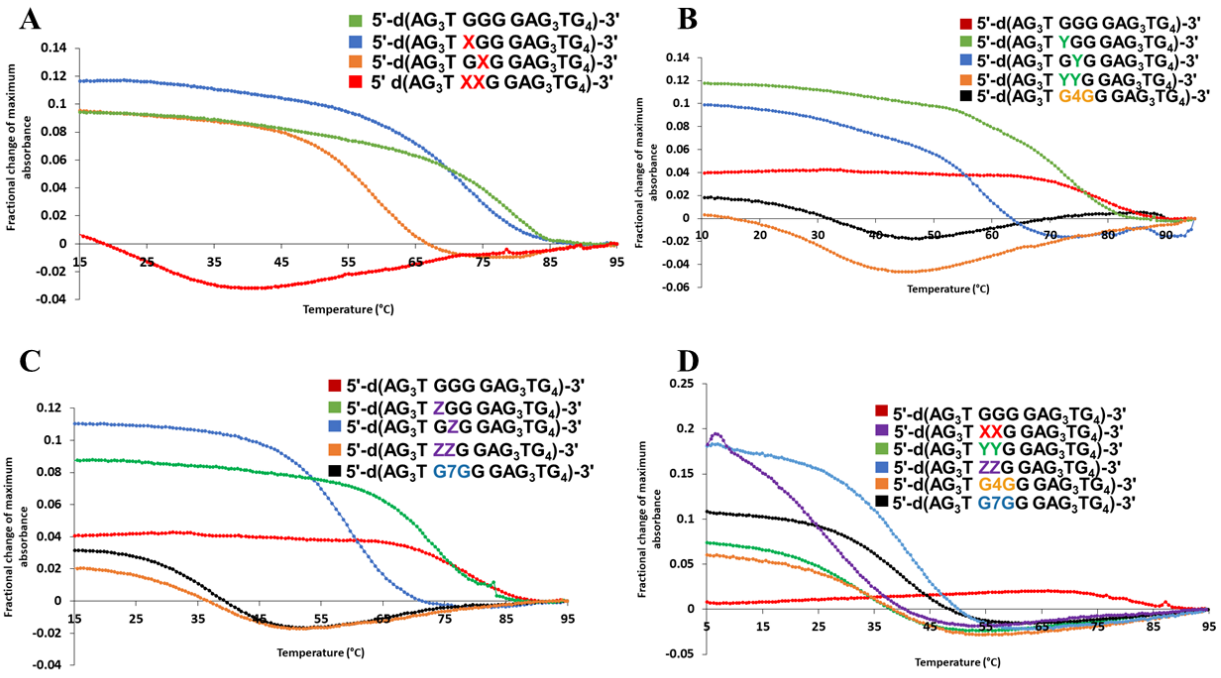


Figure 4.3 UV thermal denaturation experiments for alkyl modified G-quadruplexes. A) Methyl B) Butylene C) Heptylene and D) Di-alkylated oligonucleotides. Temperature was ramped at 0.5 ºC/min and absorbance measured at 260 nm for the monoalkylated species and 295 nm for the dialkylated). X= O^6-Methyl dG, Y= O^6-4-hydroxybutyl dG, Z= O^6-7-hydroxyheptyl dG. Buffer: 2.5 mM K_2HPO_4, 2.5 mM KH_2PO_4, 3.2 mM KOH, pH 7.2.

lesions: -5 ºC when located on the external guanine, G_6, and -21 ºC on the internal, G_7, (**Figure 4.3B** and **Table 2**). When two lesions were present a subtle transition was noted in the lower temperature range at 260 nm.

A more distinct sigmoidal transition was observed at 295 nm for the hydroxybutylene series than the methyl series and it was destabilizing by -44 ºC relative to the control G-quadruplex. The hydroxyheptylene series (**ODNs 4.8-4.10**) was also destabilizing, with a decrease of -5 ºC and -19

Table 4.2 T_ms of G-quadruplexes containing alkyl lesions. T_ms for dialkylated species were calculated for UV thermal denaturation experiments at 295 nm.

Oligonucleotide	T_m (ºC)	ΔT_m (ºC)
ODN-4.1	79	-
ODN-4.2	72	-7
ODN-4.3	59	-20
ODN-4.4	31	-48
ODN-4.5	74	-5
ODN-4.6	58	-21
ODN-4.7	35	-44
ODN-4.8	74	-5
ODN-4.9	60	-19
ODN-4.10	40	-39
ODN-4.11	35	-44
ODN-4.12	40	-39

ºC when located in the external and internal guanines, respectively. It was interesting to note the dialkylated hydroxyheptylene G-quadruplex was more stable than the series described above, when absorbance was measured at 295 nm, with a decrease in T_m of -39 ºC. While it was expected the presence of bulkier lesions would result in lower thermal stabilities, the opposite trend is observed which may be attributed to the formation of a stabilizing hydrogen bond interaction of the hydroxyl group attached to the longer more flexible alkyl chain which can allow for orientation to a more energetically favoured conformation within the G-quadruplex.

We then evaluated the IaCL series, **ODNs 4.11** and **4.12**, and noted a similar non-definitive transition in the lower temperature range as those containing multiple lesions described above when monitoring absorbance at 260 nm. However at 295 nm a destabilization of -44 ºC and -39 ºC for the butylene and heptylene IaCL G-quadruplex species, respectively, was observed. This decrease in T_m was comparable to the decreases observed for the G-quadruplexes containing the corresponding multiple monoalkylated lesions with the heptylene IaCL exhibiting slightly higher stability relative to the butylene IaCL G-quadruplex. Both G-quadruplexes containing the IaCLs are more stable than the dimethylated G-quadruplex **ODN-4.4**. Despite the increased steric bulk relative to the dimethylated series, it is presumed the lack of phosphate backbone between the crosslinked dG residues confers added flexibility which may allow for improved stacking and other stabilizing interactions resulting the higher T_ms observed.

While the low salt content buffer gave useful information with the effect of alkylation on the thermal stability when only one lesion was present, we were interested in whether repair would occur in the more physiologically relevant intracellular potassium content of 140-150 mM K^+.[200] As expected, no transition was observed for the unmodified control and those containing single alkyl(ene) adducts (data not shown) due to their exceptional thermal stability which was also seen at low potassium concentrations. Upon heating the annealed strands it was noted that the four sequences (**ODN**s **4.4, 4.7, 4.11, 4.12**) at 37 ºC are still stable G-quadruplexes (see **Section 4.4.2 Figure S4.38**) when monitoring absorption at 295 nm; with T_ms of 54 ºC and higher (see **Section 4.4.2 Table S2**). It should be noted that the annealing curve is reported for the species containing two sites of alkylation and physiological potassium concentration (**Figure 4.3D**) due to minor hysteresis in the IaCL series (data not shown). Similar trends were observed as at low potassium concentration with the longer alkylene chain (**ODN-4.10**) and IaCLs being the most stable. While the characteristic wavelength to measure change of absorbance as a function of temperature for G-quadruplexes is 295 nm,[201,202] a large change in hypochromicity was observed upon heating for the mono-alkylated G-quadruplexes at 260 nm (note at both wavelengths a hypochromic transition was noted). However, at 260 nm for the di-alkylated G-quadruplexes an increase in absorbance with inconclusive sigmoidal character was observed when heated and then cooled (see **Section 4.4.2 Figure S4.39**). It is possible that these sequences may adopt a different G-quadruplex conformation, thus circular dichroism experiments were performed to gain insight on the modified G-quadruplexes' global structure.

4.2.3 Global topology of the modified G-quadruplexes as assessed by circular dichroism.

Circular dichroism is a powerful technique for the study of G-quadruplexes as they are quite structurally diverse which due to the tetrameric base-pairing pattern can adopt numerous topologies such as i) differing directionalities (antiparallel *vs.* parallel) ii) loop lengths ii) molecularity (inter *vs.* intra) and iv) base orientations (*anti vs. syn*) quartets.[203,204] The global topology of the G-quadruplexes were assessed at both high and low K^+ concentration (**Figure 4.4A-C** and **Section 4.4.2 Figure S4.40**). At 130 mM KCl, pH 7.2 and 37 ºC, all sequences studied exhibited positive maxima at ~260 nm, negative minima at ~240 nm with a second maxima around

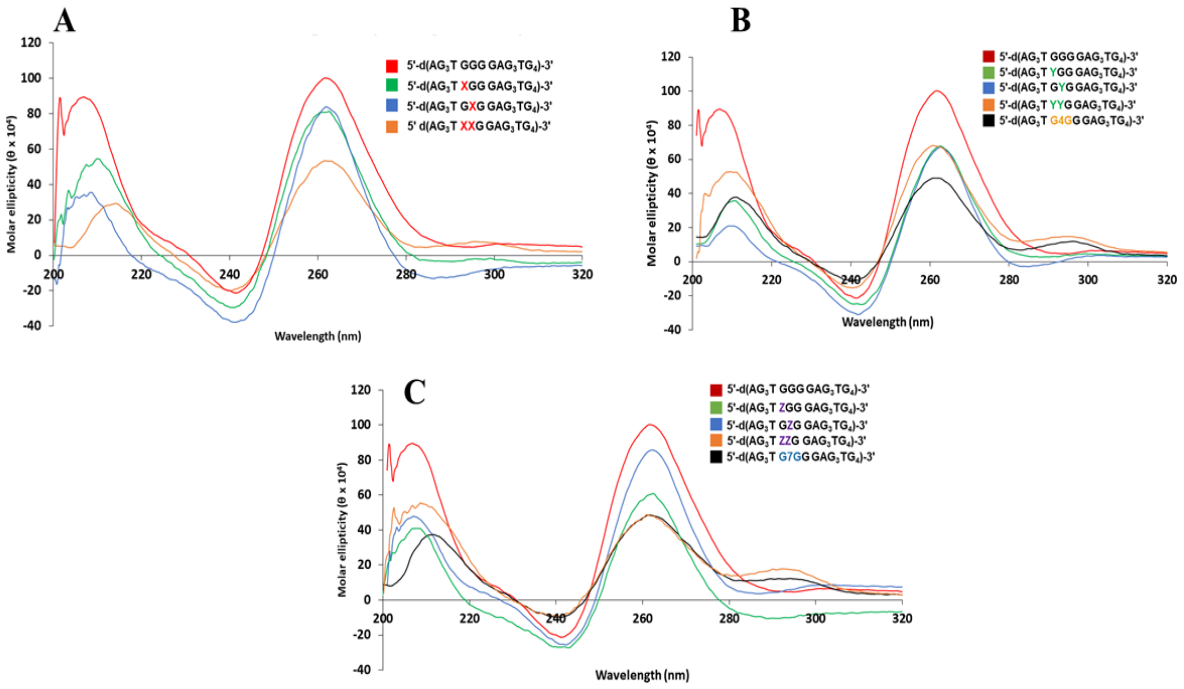


Figure 4.4 Circular dichroism spectra acquired at 37 ºC for **ODN**s studied. A) Methyl B) Hydroxybutylene and C) Hydroxyheptylene where X= O^6-Methyl dG, Y= O^6-4-hydroxybutyl dG, Z= O^6-7-hydroxyheptyl dG. Buffer: 130 mM KCl, 5 mM KH_2PO_4, 5 mM K_2HPO_4, 5.4 mM KOH, pH 7.2.

~210 nm. The CD spectra observed of the G-quadruplex containing one site of alkylation at the O^6 position are consistent with the parallel-stranded propeller type G-quadruplex of the *c-myc* promoter sequence reported in the literature, with guanines oriented in the *anti* conformation.[194,205] Due to intrinsic differences in UV melting profiles for those containing two sites of alkylation, specifically no distinct sigmoidal transition at 260 nm with an increase in

absorbance upon heating (a decrease in absorbance was noted for the other G-quadruplexes), it was suspected they may be adopting a different conformation. For the CD spectra of G-quadruplex containing two sites of alkylation (**ODNs 4.4**, **4.7** and **4.10**) an additional positive peak around 295 nm is observed for all G-quadruplex species at both low and high K^+ concentrations. Furthermore, it is significantly more pronounced for these G-quadruplexes at low salt relative to their IaCL counterparts (**ODNs 4.11** and **4.12**); potentially emphasizing the important of ions on these G-quadruplexes' conformations. The importance of cations on G-quadruplex structure has been shown in the literature and particularly the nature of the cation such as Na^+ or K^+.[206,207] Silva *et al.*[203] describe an antiparallel "Group III-2KOW" G-quadruplex which possessed similar CD spectra displaying positive peaks at 270 nm and 295 nm, and a negative peak at 240 nm. The CD spectra's dichroic peaks were also characteristic of a [3+1] or "hybrid" topology[208] which merits structure elucidation by NMR or X-ray crystallography.

4.2.4 Repair of O^6-alkyl lesions in G quadruplexes by AGT variants.

Having characterized the G-quadruplexes by UV thermal denaturation experiments and CD, the susceptibility of these G-quadruplex to undergo dealkylation by the various AGT variants was assessed in physiological relevant potassium concentration to ensure proper folding of the G-quadruplex structures. 2 pmol of ^{32}P radiolabelled modified DNA was incubated with either hAGT, Ada-C, OGT or hOGT with varying protein equivalents (6, 20 or 60 pmol) at 37 ºC for 18 h and analyzed by denaturing PAGE. Of particular interest was whether those containing multiple lesions could act as a molecular switch where the integrity of the proper *c-myc* fold (parallel-stranded propeller motif) could be restored (**ODNs 4.4, 4.7, 4.10**) or the G-quadruplex could be disassembled (IaCL **ODNs 4.11** and **4.12**) once incubated with the proteins.

First, the parallel-stranded G-quadruplex containing O^6-methylguanine was assessed as a substrate for the AGTs. As a control, in lane 6 both the substrate and its repair product to verify sufficient resolution for each gel. Both **ODNs** containing one O^6-Me lesion (See **Section 4.4.2 Figures S4.42-4.43**) were substrates for all four AGT proteins studied as observed by denaturing PAGE. At high protein concentrations virtually all substrate was consumed except when incubated with hOGT. It was noteworthy that at the lower equivalents of protein (6 pmol), all were less efficient when the lesion was located on the internal G_7 position. This is not surprising as presumably the O^6 methyl lesion is more accessible when located on the G_6 loop residue. With

lower protein concentrations hAGT was the most proficient at repair, with ~25 % of **ODN-4.2** remaining after incubation for 18 h. Examination of the species containing two O^6-methyl lesions for **ODN-4.4** (**Figure 4.5A**) revealed that hAGT, OGT and hOGT were superior at repairing the two alkyl lesions than when only one was present as the case with **ODNs 4.2** and **4.3**. This data supports our hypothesis that the G-quadruplex containing two adjacent lesions are adopting an alternate conformation that facilitates access to the damaged site allowing for enhanced repair.

Next, it was examined whether repair would occur on the G-quadruplexes containing single, or two hydroxyalkylene lesions. Repair on the G-quadruplexes containing a single O^6 hydroxybutylene or hydroxyheptylene adduct for **ODNs 4.5, 4.6, 4.8** and **4.9** (shown in **Section 4.4.2 Figures S4.44-4.47**) revealed that hAGT and hOGT were both able to act upon the lesions and **ODNs 4.5-4.6** and **4.8-4.9** exhibited higher repair than when the lesions were located in the external G_7 position which follows the same trend observed for **ODNs 4.2-4.3**. **ODNs 4.5-4.10** were all refractory to activity by both OGT and Ada-C. Inability of Ada-C and OGT was expected as they are unable to act on bulkier lesions such as O^6-benzylguanine; presumably due to a more

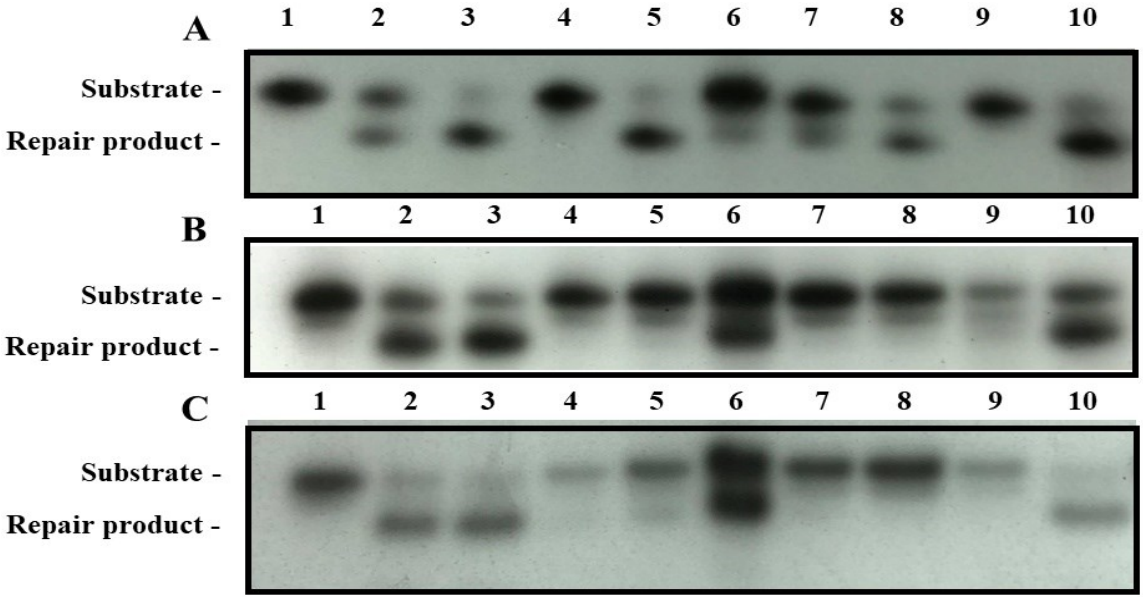


Figure 4.5 Repair assays on oligonucleotides containing two alkyl(ene) lesions. A) **ODN-4** (Methyl) B) **ODN-7** (hydroxybutylene) and C) **ODN-10** hydroxyheptylene lesions by AGTs. Lane 1) 2 pmol DNA 2) 2 pmol DNA+ 6 pmol hAGT 3) 2 pmol DNA + 60 pmol hAGT 4) 2 pmol DNA + 6 pmol Ada-C 5) 2 pmol DNA + 60 pmol Ada-C 6) Positive control: DNA + repair product 7) 2 pmol DNA + 6 pmol OGT 8) 2 pmol DNA + 60 pmol OGT 9) 2 pmol + 6 pmol hOGT 10) 2 pmol + 60 pmol hOGT.

compact, unaccommodating active site.[31] Increased levels of substrate consumption were observed for the hydroxyheptylene *vs.* the hydroxybutylene series for all ODNs studied by both hAGT and hOGT. Specifically, while small quantities of substrate remained when **ODN-4.10** was incubated with 60 pmol of hAGT or hOGT (7 % and 10 %), more substantial quantities remained with **ODN-4.7** (31 % with hAGT and 42 % with hOGT). This could be explained by the hydrophobic tunnel-like active site of hAGT being more accommodating and energetically favourable for the longer heptylene chain as the case with ICLs.[35] Presumably, the hydroxyheptylene cross-link adopts a conformation with allows the α carbon to be correctly positioned for the AGT's activated cysteine residue. The repair by the chimera protein hOGT has been rationalized by the retention of human active site's Pro140 residue which widens the active site of hOGT, whilst various *E. coli* mutations such as N157G and S159A allow the more flexible loop to orient the nucleophilic cysteine towards the alkyl(ene) chain's α carbon.[51]

Use of IaCLs lacking a phosphate backbone has shown potential in nanotechnology as the DNA strand is cleaved at the site of the alkylene linkages.[209] Therefore we were interested if the various AGTs were capable of cleaving G-quadruplexes containing IaCLs lacking a phosphate backbone (**ODN**s **4.11** and **4.12**) by monitoring the unhooking of the lesion by these proteins (**Figure 4.6**). There are two potential full repair products (as well a DNA-AGT cross-link) for these IaCLs lacking a phosphate backbone due to the presence of two 5'-OH groups (which can

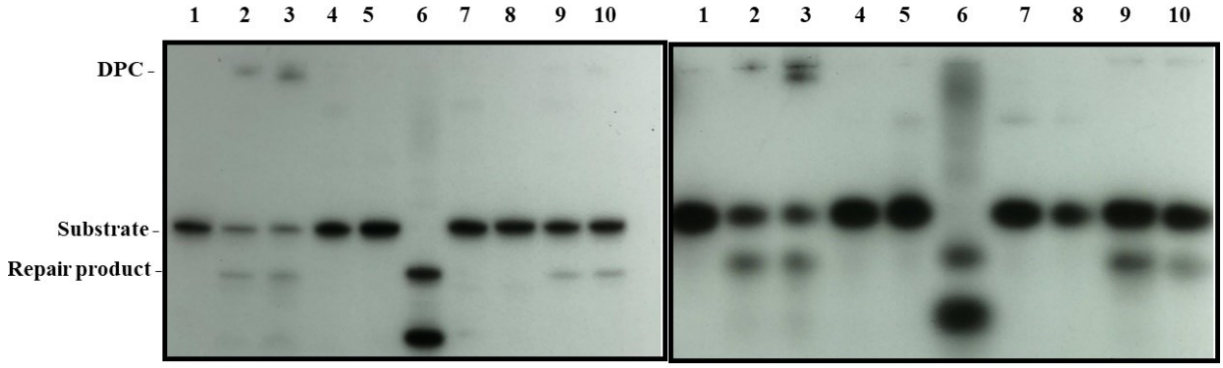


Figure 4.6 Repair assays on oligonucleotides containing IaCLs. A) **ODN-11 (**Butylene) and B) **ODN-12** (Heptylene) IACLs by AGTs. Lane 1) 2 pmol DNA 2) 2 pmol DNA+ 20 pmol hAGT 3) 2 pmol DNA + 60 pmol hAGT 4) 2 pmol DNA + 20 pmol Ada-C 5) 2 pmol DNA + 60 pmol Ada-C 6) Positive control: DNA + repair product 7) 2 pmol DNA + 20 pmol OGT 8) 2 pmol DNA + 60 pmol OGT 9) 2 pmol DNA + 20 pmol hOGT 10) 2 pmol DNA + 60 pmol hOGT.

each be radiolabeled) and which were loaded in lane 6. Due to no repair observed with 3-fold protein concentration (data not shown), for the low concentration data point we incubated with 10-fold AGT. Unsurprisingly, *E. coli*'s OGT and Ada-C proteins were unable to act upon both IaCLs as was observed with duplex DNA (**Figure 4.6** lanes 4-5, 7-9).[40] hAGT was able to repair both butylene and heptylene linked IaCLs (**ODNs 4.11**, **4.12**) but oddly in an essentially non-concentration dependent manner. We observed ~55-60 % substrate remaining for the butylene IaCL (**ODN-4.11**) with both 10-fold and 30-fold protein concentration after 18h. At higher protein concentrations hAGT showed increased repair acting on the heptylene IaCL (**ODN-4.12**) 55 % substrate was remaining with 10-fold and 40 % with 30-fold excess. Curiously, the decrease in substrate at high hAGT concentration was attributed to an increase in the formation of the more slowly migrating DNA-protein crosslink (DPC) band. Evidently, once hAGT is covalently bound to the G-quadruplex, a second hAGT has difficulty accessing the α carbon at the O^6 position of the second dG. hOGT acted on both butylene and heptylene IaCLs with a similar proficiency and concentration independence, with 75-80 % substrate remaining. The diminished activity of hAGT towards IaCLs in G-quadruplex *vs.* duplex DNA, where very little substrate remained for butylene and heptylene IaCLs,[40] is likely due to the inability of hAGT to flip out the 3' O^6-alkylated dG. ESI-MS data shows that hAGT acts upon the 3' dG likely due to the presence of its 3' phosphate group which is probed by Tyr 114 (either by sterics or charge repulsion) to induce nucleotide flipping.[15] It is plausible that the combination of the alkylene tether hindering flipping coupled with the inability to locally denature the stable hydrogen bonding network diminishes hAGT's unhooking of the IaCLs. AGTs stabilize the DNA duplex with an arginine finger which base-pairs with the guanines opposing cytosine.[15] In the case of the G-quadruplex tetrad, this would be unlikely to have the same stabilizing effect. It is postulated that lack of significant concentration-dependent repair could be due the kinetics of the reaction: perhaps the protein is denaturing before it has time to react further with the G-quadruplex substrate. Further studies will be performed to determine the kinetics of repair by AGTs with G-quadruplex alkylation compared to that in duplex DNA.

4.3 Conclusions

In this work, we have examined the biophysical properties and susceptibility towards reaction with AGT proteins of O^6-alkylene dG adducts and IaCL incorporated into the G-quadruplex formed by a truncated *c-myc* 18mer. This study expands on previous work by Fried who varied O^6 methyl dG's position in the telomeric repeat $(TTAGGG)_x$ and showed the G-quadruplexes to be substrates of hAGT. The truncated *c-myc* 18mer was selected due to its biological relevance, having been well characterized and demonstrating a low degree of polymorphism. The conformational dynamics described by Harkness *et al*. emphasized a non-hydrogen bonding alkylated guanine residue being destabilizing to G-quadruplexes (independent of its length). G-quadruplexes have merited extensive research due to their structural diversity as well as implications in regions of the genome and their importance for aptamer folding. We have shown that we can add to the toolbox and modulate the stability of a G-quadruplex depending on the lesion location, or quantity of lesions. DNA aptamers are sequences of DNA which when properly folded are capable of binding a substrate with high affinity and selectivity.[210] We believe these sequences, in particular the IaCLs, would have potential applications in AGT triggered aptamer unfolding as a key characteristic of aptamers is they often require a G-quadruplex for proper folding.[211] Groups have looked at sugar[212,213] and base modifications[214] which affect both sugar pucker and base conformation to tune G-quadruplex topology with applications in nanotechnology.[215] The dG-O^6-alkylene-O^6-dG intrastrand cross-links lacking the phosphate backbone could also be a useful tool for cleavage induced inactivation of DNA aptamers upon unhooking by AGTs. hAGT has been utilized previously in protein-DNA nanotechnology as a means of detection thus there is a precedent for further exploitation of AGT in other contexts.[53,54] Chemical modification has shown promise for the therapeutic properties of aptamers: for example anticancer efficacy for the thrombin binding aptamer have shown to be enhanced.[216,217]

Differences with bacterial (OGT and Ada-C) and human (hAGT) proteins could be pertinent in the production of antibiotics as those containing hydroxyalkylene and IaCLs were not substrates for bacteria. Only O^6 methyl groups were removed by the *E. coli* proteins and all bulky lesions were refractory to repair by however were repaired by the human variant. Thus, one can envision differences in aptamer folding depending on whether they are in a bacterial cell, where they maintain a certain fold, or if located in a human cell they maintain a different conformation. upon being acted upon by hAGT. This has potential for specific targeting of bacterial cells by

aptamers. Bacterial targeting aptamers have been identified such as the PPK2 G2 aptamer against *Mycobacteriuam tuberculosis*.[144]

From a molecular biology perspective, the dimethylation data suggests it is possible when multiple sites of methylation, or alkylation when treated with chemotherapeutics lead to destabilization of the G-quadruplex or conformational changes and thus allows transcription of the *c-myc* oncogene to occur. One may extrapolate that more lesions cause even further destabilization and may play a role in cancer by increasing the expression of transcription activators with genes involved in cellular proliferation and repressors for those with growth arrest.[130] Herein, we show that these lesions in G-quadruplexes are substrates for hAGT even at low protein concentrations. Methylation has particular relevance as there are many endogenous and exogenous agents that we have shown could play a role in gene expression; their removal by AGTs stabilizes the *c-myc* promoter region G-quadruplex and leads to decreased RNA transcription. Mechanistic insights from structures of AGT's inactive mutants bound to the G-quadruplexes would give important results with respect to the phenomena observed in this study. Thus, the work performed here should have implications for both biological and nanotechnology applications.

4.4 Supporting Information

4.4.1 Supporting Methods

4.4.1.1 General

5'-*O*-(4,4'-Dimethoxytrityl)-N^2-phenoxyacetyl-2'-deoxyguanosine and 2-cyanoethyl N,N-diisopropylchloro phosphoramidite were purchased from ChemGenes Corporation (Wilmington, MA). 5'-*O*-(4,4'-Dimethoxytrityl)-2'-deoxyribonucleoside-3'-*O*-(*β*-cyanoethyl-N,N-diisopropyl) phosphoramidites and protected 2'-deoxyribonucleoside-controlled pore glass (CPG) supports were acquired from Glen Research (Sterling, VA). Chemicals and solvents were obtained from MilliporeSigma (St. Louis, MO). Flash column chromatography silica gel 60 (230-400 mesh) was purchased from Silicycle (Québec City, QC). Thin layer chromatography was performed with precoated TLC plates (Merck, Kieselgel 60 F_{254}, 0.25 mm) obtained from MilliporeSigma (St. Louis, MO). NMR spectra were acquired on a Varian 500 MHz NMR spectrometer at room temperature and all chemical shifts were reported in ppm downfield from an internal standard. ^{1}H NMR spectra were recorded at a frequency of 500 MHz and ^{13}C spectra (^{1}H decoupled) at 125.7 MHz with tetramethylsilane as a reference. ^{31}P NMR spectra (^{1}H decoupled) were acquired at a

frequency of 202.3 MHz with 85 % H_3PO_4 as an external reference. Acquisition of high-resolution mass spectrometry data of small molecules was performed using a 7T-LTQ FT ICR mass spectrometer from ThermoScientific (Waltham, MA) at the Concordia University Centre for Structural and Functional Genomics operated in positive ion scan mode. Most reagents for biochemical assays, PAGE and protein purification, were purchased from Bioshop Canada Inc (Burlington, ON). EDTA-free protease inhibitor cocktail tablets were obtained from Roche (Laval, QC) and Ni-NTA Superflow Resin from Qiagen (Mississauga, ON). T4 polynucleotide kinase was bought from Fermentas (Burlington, ON), and [γ-^{32}P]-ATP from PerkinElmer (Woodbridge, ON). DNA primers were acquired from Biocorp (Montreal, QC) and Phusion Polymerase from New England Biolabs (Ipswich, MA). Stratagene (Cedar Creek, TX) provided XL-10 Gold and Bl21 (DE3) *E. coli* cells. Mung Bean nuclease was purchased from New England BioLabs Inc. (Whitby, ON). X-ray film was purchased from Harvard Apparatus Canada (St. Laurent, Québec).

4.4.1.2 Chemical synthesis of nucleosides

*3'-O-(tert-butyldimethylsilyl)-5'-O-(4,4'-dimethoxytrityl)-O^6-methyl-N^2-phenoxyacetyl-2'-deoxyguanosine (***4.1a***)*

To a solution of 3'-*O*-(*tert*-butyldimethylsilyl)-5'-*O*-(4,4'-dimethoxytrityl)-N^2-phenoxyacetyl-2'-deoxyguanosine (0.50 g, 0.61 mmol), methanol (29 µL, 0.73 mmol) and PPh_3 (0.25 g, 0.92 mmol) in anhydrous dioxane (23 mL) was added DIAD (180 µL, 0.92 mmol) dropwise while stirring in ice/water bath under argon. After 24 h the solvent was removed *in vacuo* and the crude compound redissolved in 100 mL of DCM. The crude compound was washed $NaHCO_3$ (*aq*, 3 % w/v) (2 x 75mL), brine (1 x 75 mL), dried over Na_2SO_4 (~4 g), decanted and the solvent removed *in vacuo*. The crude compound was subsequently purified by silica gel flash column chromatography with EtOAc:hexanes (3:7→4:6) to yield 220 mg (43 %) of **4.1a** as a colourless foam. R_f (SiO_2, TLC: 0.47 (5 % MeOH:DCM). $\lambda_{max(MeCN)}$ = 267 nm. ^{1}H NMR (500 MHz, $CDCl_3$) δ 8.68 (s, 1H, NH), 8.03 (s, 1H, C8), 7.48 – 6.72 (m, 18H, Ar), 6.42. (t, 1H, H1'), 4.76 (s, 2H, $PhOCH_2CO$), 4.64 – 4.51 (m, 1H, H3'), 4.19 (s, 3H, OCH_3), 4.10-4.07 (m, 1H, H4'), 3.76 (s, 6H, 2 x OCH_3), 3.33 (d, *J* = 3.8 Hz, 2H, H5' and H5''), 2.73-2.68 (m, 1H, H2'), 2.46-2.42 (m, 1H, H2''), 0.86 (s, 9H, $Si(CH_3)_3$), 0.05 (s, 3H, Si-CH_3), 0.01 (s, 3H, Si-CH_3). ^{13}C NMR (126 MHz, $CDCl_3$) δ 161.32, 158.48, 152.27, 144.51, 140.21, 135.71, 135.69, 132.98, 132.15, 132.12, 132.04, 131.90, 131.88, 129.99, 129.78, 128.51, 128.42, 128.10, 127.80, 126.84, 122.26, 114.91, 113.10, 86.97, 86.45, 84.30, 72.52, 68.03,

63.47, 55.17, 54.59, 40.89, 25.73, 17.94, -0.02, -4.69, -4.83. HRMS (ESI-MS) *m/z* calculated for $C_{46}H_{53}N_5O_8Si^+$ $[M+H]^+$: 832.3742; found: 832.3776.

*5'-O-(4,4'-dimethoxytrityl)-O^6-methyl-N^2-phenoxyacetyl-2'-deoxyguanosine (***4.2a***)*

To a solution of **4.1a** (0.11 g, 0.13 mmol) in anhydrous THF (3 mL) was added TBAF (150 µL of 1 M in THF, 0.15 mmol) dropwise while stirring under argon. After 20 minutes the volume was brought to 100 mL with DCM, the crude compound was washed with $NaHCO_3$ (*aq*, 3 % w/v) (2 x 75mL), brine (1 x 75 mL), dried over Na_2SO_4 (~4 g), decanted and the solvent removed *in vacuo*. The crude compound was purified by silica gel flash column chromatography with MeOH:DCM (1.5→1.75 %) to obtain 85 mg (91 %) of **4.2a** as a colourless foam. R_f (SiO_2, TLC: 0.26 (5 % MeOH:DCM). $\lambda_{max(MeCN)}$ = 282 nm. ^{1}H NMR (500 MHz, $CDCl_3$) δ 8.81 (s, 1H, NH), 8.03 (s, 1H, C8), 7.46 – 7.12 (m, 11H, Ar), 7.11 – 6.93 (m, 3H, Ar), 6.81 – 6.73 (m, 4H, Ar), 6.68 (t, J=6.3 Hz, 1H, H1'), 4.85 – 4.74 (m, 1H, H3'), 4.65 (s, 2H, $PhOCH_2CO$), 4.28-4.25 (m, 1H, H4'), 4.14 (s, 3H, OCH_3), 3.73 (s, 6H, 2 x OCH_3), 3.45-3.42 (m, 1H, H5'), 3.37-3.31 (m, 1H, H5''), 2.78 – 2.60 (m, 2H, H2' and H2''). ^{13}C NMR (126 MHz, $CDCl_3$) δ 161.25, 158.44, 157.05, 152.26, 150.89, 144.57, 140.45, 135.77, 135.68, 130.03, 130.00, 129.81, 128.12, 127.80, 126.83, 122.37, 118.99, 114.90, 113.08, 86.66, 86.42, 84.22, 77.28, 77.02, 76.77, 72.49, 67.86, 64.18, 55.15, 55.14, 54.55, 45.91, 40.82, 29.67, 8.67, -0.03.HRMS (ESI-MS) *m/z* calculated for $C_{40}H_{39}N_5O_8^+$ $[M+H]^+$: 718.2877; found: 718.2888.

*3'-O-(2-cyanoethoxy(diisopropylamino)-phosphino)-5'-O-(4,4'-dimethoxytrityl)-O^6-methyl-N^2-phenoxyacetyl-2'-deoxyguanosine (***4.3a***)*

Compound **4.2a** was co-evaporated with anhydrous MeCN (1 x 10 mL). To a solution of **4.2a** (0.088 g, 0.12 mmol) and DIPEA (32 µL, 0.18 mmol) in anhydrous DCM (2 mL) was added Cl-POCEN*i*Pr_2 (33 µL, 0.15 mmol) dropwise while stirring under argon. After 20 minutes the solvent was removed *in vacuo* and the crude compound was redissolved in 50 mL of EtOAc. The compound was washed with $NaHCO_3$ (*aq*, 3 %) (2 x 35 mL), brine (1 x 35 mL), dried over Na_2SO_4 (~2 g), decanted and the solvent removed *in vacuo*. The crude compound was purified by silica gel flash column chromatography (conditioned with 0.5 % TEA/EtOAc) with EtOAc (100 %) to obtain 73 mg (66 %) of **4.3a** as a colourless foam. R_f (SiO_2, TLC): 0.56, 0.64 (100 % EtOAc). $\lambda_{max(MeCN)}$ = 269 nm. ^{1}H NMR (500 MHz, d_6-acetone) δ 9.30 (s, 2H), 8.22 (s, 2H), 7.49 – 7.06 (m,

22H), 7.07 – 6.93 (m, 6H), 6.90 – 6.64 (m, 8H), 6.51-6.49 (m, 2H), 5.20-5.15 (m, 2H), 5.09-5.03 (m, 4H), 5.00-4.95 (m, 2H), 4.94-4.89 (m, 1H), 4.35 – 4.23 (m, 2H), 4.14 (s, 6H), 3.94 – 3.80 (m, 2H), 3.78 – 3.72 (m, 14H), 3.72 – 3.47 (m, 8H), 3.39-3.29 (m, 2H), 3.29 – 3.13 (m, 2H), 2.77 – 2.70 (m, 2H), 2.65-2.59 (m, 2H), 1.30 (t, J = 7.0 Hz, 3H), 1.22 – 1.17 (m, 12H), 1.12 (d, J = 6.8 Hz, 4H). ^{13}C NMR (126 MHz, d_6-acetone) δ 161.19, 158.61, 158.54, 158.20, 152.59, 152.53, 151.60, 145.31, 145.29, 141.30, 141.23, 135.98, 135.91, 135.82, 135.78, 130.15, 130.10, 129.95, 129.93, 129.45, 128.10, 128.04, 127.52, 126.53, 121.23, 121.20, 114.72, 112.84, 112.82, 112.78, 112.77, 86.21, 86.17, 86.00, 85.93, 84.77, 84.70, 74.30, 74.16, 73.97, 73.83, 67.87, 64.22, 58.72, 58.66, 58.57, 58.51, 54.56, 54.55, 54.53, 53.83, 43.10, 43.00, 38.35, 24.01, 23.95, 19.86, 19.81, 19.75. ^{31}P NMR (202 MHz, d_6-acetone) δ 148.43, 148.22. HRMS (ESI-MS) *m/z* calculated for $C_{49}H_{56}N_7O_9P^+$ $[M+H]^+$: 918.3955; found: 918.3917.

*3'-O-(tert-butyldimethylsilyl)-5'-O-(4,4'-dimethoxytrityl)-N^2-phenoxyacetyl-O^6-(4-phenoxyacetyloxy butylene)-2'-deoxyguanosine (***4.1b***)*

To a solution of 3'-*O*-(*tert*-butyldimethylsilyl)-5'-*O*-(4,4'-dimethoxytrityl)-N^2-phenoxyacetyl-2'-deoxyguanosine (0.50 g, 0.61 mmol), 4-(phenoxyacetyl)-butyl-1-ol (0.20 g, 0.77 mmol) and PPh_3 (0.16 g, 0.61 mmol) in anhydrous dioxane (17 mL) was added DIAD (120 µL, 0.61 mmol) dropwise while stirring in an ice/water bath under argon. After 4 h, more PPh_3 (0.16 g, 0.61 mmol) and DIAD (120 µL, 0.61 mmol) were added. After 18 h, the solvent was removed *in vacuo* and the crude gum was redissolved in DCM (100mL). The crude compound was washed $NaHCO_3$ (*aq*, 3 % w/v) (2 x 75mL), brine (1 x 75 mL), dried over Na_2SO_4 (~4 g), decanted and the solvent removed *in vacuo*. The crude compound was purified by silica gel flash column chromatography EtOAc:hexanes (3:7→ 4:6) to yield 421 mg (67 %) of **4.1b** as a colourless foam. R_f (SiO_2, TLC): 0.62 (5 % MeOH:DCM). $\lambda_{max(MeCN)}$ = 269 nm. ^{1}H NMR (500 MHz, $CDCl_3$) δ 8.69 (s, 1H, NH), 8.03 (s, 1H, C8), 7.43 – 7.37 (m, 2H), 7.37 – 7.14 (m, 11H), 7.08 – 6.93 (m, 4H), 6.93 – 6.85 (m, 2H), 6.82 – 6.73 (m, 4H), 6.42 (t, J = 6.6 Hz, 1H, H1'), 4.75 (s, 2H, $PhOCH_2CO$), 4.64 (s, 2H, $PhOCH_2CO$), 4.63 – 4.55 (m, 3H, H3' + CH_2-Ar), 4.29 (t, *J* = 6.1 Hz, 2H, CH_2-OPac), 4.08 (d, *J* = 3.6 Hz, 1H, H4'), 3.76 (s, 6H, 2 x OCH_3), 3.33 (d, *J* = 4.3 Hz, 2H, H5' and H5''), 2.73-2.68 (m, 1H, H2'), 2.47-2.42 (m, 1H, H2''), 1.97-1.86 (m, 4H, -$(CH_2)_2$-), 0.86 (s, 9H, $Si(CH_3)_3$), 0.05 (s, 3H, Si-CH_3), 0.00 (s, 3H, Si-CH_3). ^{13}C NMR (126 MHz, $CDCl_3$) δ 169.00, 160.97, 158.48, 157.77, 152.42, 151.07, 144.50, 140.13, 139.43, 135.71, 135.68, 132.12, 132.04, 131.90, 131.87, 129.99,

129.97, 129.82, 129.79, 129.53, 129.10, 128.51, 128.42, 128.09, 127.81, 127.74, 126.85, 122.43, 122.29, 121.69, 118.89, 114.91, 114.61, 113.15, 113.11, 112.98, 89.78, 86.96, 86.46, 84.24, 73.37, 72.50, 67.99, 66.92, 65.30, 64.83, 64.78, 63.42, 55.23, 55.18, 40.95, 25.77, 25.73, 25.36, 25.31, 25.19, 25.15, 17.95, -0.02, -4.69, -4.71, -4.77, -4.83. HRMS (ESI-MS) *m/z* calculated for $C_{57}H_{65}N_5O_{11}Si^+$ $[M+H]^+$: 1024.4528; found: 1024.4517.

*3'-O-(tert-butyldimethylsilyl)-5'-O-(4,4'-dimethoxytrityl)-N^2-phenoxyacetyl-O^6-(7-phenoxyacetyloxy heptylene-2'-deoxyguanosine (***4.1c***)*

To a solution of 3'-*O*-(*tert*-butyldimethylsilyl)-5'-*O*-(4,4'-dimethoxytrityl)-N^2-phenoxyacetyl-2'-deoxyguanosine (0.45 g, 0.54 mmol), 7-(phenoxyacetyl)-heptyl-1-ol (0.17 g, 0.65 mmol) and PPh_3 (0.30 g, 1.1 mmol) in anhydrous dioxane (15 mL) was added DIAD (220 µL, 1.1 mmol) dropwise while stirring on ice under argon. After 18 h, the solvent was removed *in vacuo* and the crude gum was redissolved in DCM (100mL). The crude compound was washed $NaHCO_3$ (*aq*, 3 % w/v) (2 x 75mL), brine (1 x 75 mL), dried over Na_2SO_4 (~4 g), decanted and the solvent removed *in vacuo*. The crude compound was purified by silica gel flash column chromatography with EtOAc:hexanes (4:6) to yield 350 mg (60 %) of **4.1c** as a colourless foam. R_f (SiO_2, TLC): 0.71 (5 % MeOH:DCM). $\lambda_{max(MeCN)}$ = 269 nm. ^{1}H NMR (500 MHz, $CDCl_3$) δ 8.65 (s, 1H, NH), 8.03 (s, 1H, C8), 7.43 – 7.37 (m, 2H, Ar), 7.38 – 7.14 (m, 10H, Ar), 7.06-6.96 (m, 4H, Ar), 6.94 – 6.87 (m, 3H, Ar), 6.82 – 6.75 (m, 4H, Ar), 6.42 (t, J = 6.6 Hz, 1H, H1'), 4.78 (s, 2H, $PhOCH_2CO$), 4.63 (s, 2H, $PhOCH_2CO$), 4.60-4.57 (m, 3H, $-CH_2$-Ar+H3'), 4.19 (t, J = 6.6 Hz, 2H, $-CH_2$-O), 4.09-4.06 (m, 1H, H4'), 3.76 (s, 6H, 2 x OCH_3), 3.33 (d, *J* = 4.3 Hz, 2H, H5' and H5''), 2.72-2.67 (m, 1H, H2'), 2.46-2.41 (m, 1H, H2''), 1.94 – 1.83 (m, 2H, $-CH_2$-), 1.68-1.62 (m, 2H, $-CH_2$-), 1.52-1.46 (m, 2H, $-CH_2$-), 1.39-1.30 (m, 4H, $(-CH_2-)_2$), 0.86 (s, 9H, $Si(CH_3)_3$), 0.05 (s, 3H, Si-CH_3), 0.00 (s, 2H, Si-CH_3). ^{13}C NMR (126 MHz, $CDCl_3$) δ 169.05, 161.20, 158.48, 157.80, 152.35, 151.12, 144.52, 140.03, 135.73, 135.69, 129.99, 129.97, 129.78, 129.52, 128.09, 127.81, 126.84, 122.25, 121.67, 114.91, 114.61, 113.10, 86.95, 86.45, 84.25, 72.52, 68.04, 67.75, 65.34, 63.44, 55.18, 40.92, 28.89, 28.72, 28.45, 25.83, 25.73, 25.69, 17.95, -0.02, -4.69, -4.83. HRMS (ESI-MS) *m/z* calculated for $C_{60}H_{71}N_5O_{11}Si^+$ $[M+H]^+$: 1066.4998; found: 1066.4882.

*5'-O-(4,4'-dimethoxytrityl)-N^2-phenoxyacetyl-O^6-(4-phenoxyacetyloxybutylene)-2'-deoxyguanosine (***4.2b***)*

To a solution of **4.1b** (0.39 g, 0.38 mmol) in anhydrous THF (8 mL) was added TBAF (150 µL of 1 M in THF, 0.46 mmol) dropwise while stirring under argon. After 30 minutes the volume was brought to 100 mL with DCM, the crude compound was washed with $NaHCO_3$ (*aq*, 3 % w/v) (2 x 75 mL), brine (1 x 75 mL), dried over Na_2SO_4 (~4 g), decanted and the solvent removed *in vacuo*. The crude compound was purified by silica gel flash column chromatography with MeOH:DCM (1.5 %) to obtain 268 mg (78 %) of **4.2b** as a colourless foam. R_f (SiO_2, TLC): 0.28 (5 % MeOH:DCM). $\lambda_{max(MeCN)}$ = 269 nm. ^{1}H NMR (500 MHz, $CDCl_3$) δ 8.73 (s, 1H, NH), 7.99 (s, 1H, C8), 7.74 – 6.70 (m, 23H, Ar), 6.52 (t, *J* = 6.5 Hz, 1H, H1'), 4.80 (s, 1H, H3'), 4.68 (s, 2H, $PhOCH_2CO$), 4.63 (s, 2H, $PhOCH_2CO$), 4.59 (t, *J* = 6.0 Hz, 2H, CH_2-Ar), 4.29 (t, *J* = 6.0 Hz, 2H, CH_2-OPac), 4.16 (d, *J* = 4.0 Hz, 1H, H4'), 3.75 (s, 6H, 2 x OCH_3), 3.45 (dd, *J* = 10.2, 4.8 Hz, 1H, H5'), 3.34 (dd, *J* = 10.3, 4.4 Hz, 1H, H5''), 2.84 – 2.75 (m, 1H, H2'), 2.62 – 2.54 (m, 1H, H2'), 1.91 (m, 4H, -$(CH_2)_2$-). ^{13}C NMR (126 MHz, $CDCl_3$) δ 169.02, 160.95, 158.48, 157.76, 152.42, 150.88, 144.56, 140.35, 135.75, 135.67, 132.10, 132.02, 131.95, 130.03, 130.00, 129.83, 129.53, 128.54, 128.45, 128.11, 127.83, 126.86, 122.41, 121.70, 118.90, 114.92, 114.60, 113.11, 86.53, 86.48, 84.15, 72.57, 67.86, 66.88, 65.29, 64.79, 64.10, 55.16, 46.08, 40.78, 25.34, 25.17, -0.01. HRMS (ESI-MS) *m/z* calculated for $C_{51}H_{51}N_5O_{11}^+$ $[M+H]^+$: 910.3663; found:910.3648.

*5'-O-(4,4'-dimethoxytrityl)-N^2-phenoxyacetyl-O^6-(7-phenoxyacetyloxyheptylene)-2'-deoxyguanosine (***4.2c***)*

To a solution of **4.1c** (0.34 g, 0.31 mmol) in anhydrous THF (4 mL) was added TBAF (1.5 mL of 0.25 M in TBAF, 0.38 mmol) dropwise while stirring under argon. After 1 h the volume was brought to 100 mL with DCM, the crude compound was washed with $NaHCO_3$ (*aq*, 3 % w/v) (2 x 75 mL), brine (1 x 75 mL), dried over Na_2SO_4 (~4 g), decanted and the solvent removed *in vacuo*. The crude compound was purified by silica gel flash column chromatography with MeOH:DCM (1 → 2 %) to obtain 210 mg (69 %) of **4.2c** as a colourless foam. R_f (SiO_2, TLC): 0.45 (5 % MeOH:DCM). $\lambda_{max(MeCN)}$ = 269 nm. ^{1}H NMR (500 MHz, $CDCl_3$) δ 8.72 (s, 1H, NH), 7.99 (s, 1H, C8), 7.41 – 7.21 (m, 12H, Ar), 7.19-7.16 (m, 1H, Ar), 7.07-7.04 (m, 1H, Ar), 7.02 – 6.96 (m, 3H, Ar), 6.93 – 6.87 (m, 2H, Ar), 6.84 – 6.71 (m, 4H, Ar), 6.52 (t, J = 6.4 Hz 1H, H1'), 4.83-4.79 (m, 1H, H3'), 4.70 (s, 2H, $PhOCH_2CO$), 4.62 (s, 2H, $PhOCH_2CO$), 4.57 (t, *J* = 6.6 Hz,

2H, CH_2-Ar), 4.19 (t, *J* = 6.7 Hz, 2H, -CH_2-OPac), 3.76 (s, 6H, 2 x OCH_3), 3.45 (dd, *J* = 10.2, 4.8 Hz, 1H, H5'), 3.34 (dd, *J* = 10.2, 4.5 Hz, 1H, H5''), 2.82-2.77 (m, 1H, H2'), 2.66 – 2.52 (m, 1H, H2'), 1.87 (s, 2H, -CH_2-), 1.65 (s, 2H, -CH_2-), 1.53 – 1.44 (m, 2H, -CH_2-), 1.40 – 1.31 (m, 4H, -$(CH_2)_2$-). ^{13}C NMR (126 MHz, $CDCl_3$) δ 169.05, 161.17, 158.46, 157.78, 152.32, 150.91, 144.54, 140.22, 135.74, 135.66, 130.01, 129.98, 129.81, 129.50, 128.09, 127.81, 126.84, 122.37, 121.66, 118.95, 114.90, 114.58, 113.09, 86.47, 84.11, 77.24, 76.99, 76.74, 72.60, 67.88, 67.72, 65.31, 64.09, 55.15, 40.70, 28.87, 28.69, 28.43, 25.80, 25.66, -0.03. HRMS (ESI-MS) *m/z* calculated for $C_{54}H_{57}N_5O_{11}^+$ $[M+H]^+$: 952.4133; found: 952.4069.

*3'-O-(2-cyanoethoxy(diisopropylamino)-phosphino-5'-O-(4,4'-dimethoxytrityl)-N^2-phenoxyacetyl-O^6-(4-phenoxyacetyloxybutylene)-2'-deoxyguanosine (***4.3b***)*

Compound **4.2b** was co-evaporated with anhydrous MeCN (1 x 10 mL). To a solution of **4.2b** (0.24 g, 0.27 mmol) and DIPEA (71 µL, 0.4 mmol) in anhydrous DCM (2 mL) was added Cl-POCEN*i*Pr_2 (71 µL, 0.40 mmol) dropwise while stirring under argon. After 30 minutes the solvent was removed *in vacuo* and the crude compound was redissolved in 50 mL of EtOAc. The compound was washed with $NaHCO_3$ (*aq*, 3 %) (2 x 35 mL), brine (1 x 35 mL), dried over Na_2SO_4 (~2 g), decanted and the solvent removed *in vacuo* The crude compound was purified by silica gel flash column chromatography (conditioned with 0.5 % TEA/EtOAc) with EtOAc (100 %) to obtain 236 mg (79 %) of **4.3b** as a colorless foam. R_f (SiO_2, TLC): 0.74, 0.82 (100 % EtOAc). $\lambda_{max(MeCN)}$ = 269 nm. ^{1}H NMR (500 MHz, d_6-acetone) δ 9.26 (s, 1H), 8.22 (s,1H), 7.46 – 6.67 (m, 22H), 6.51-6.48 (m, 1H), 5.04-5.02 (m, 2H), 5.00 – 4.94 (m, 1H), 4.94-4.89 (m, 1H), 4.74 (s, 2H), 4.68-4.59 (m, 2H), 4.33 – 4.20 (m, 3H), 3.94 – 3.84 (m, 1H), 3.84 – 3.78 (m, 1H), 3.78 – 3.73 (m, 7H), 3.65 (m, 2H), 3.53 (m, 1H), 3.41 – 3.31 (m, 1H), 3.24 – 3.11 (m, 1H), 2.80 (s, 2H), 2.77 – 2.67 (m, 2H), 2.65-2.59 (m, 1H), 2.08 – 2.02 (m, 6H), 2.00 – 1.92 (m, 2H), 1.20-1.17 (m, 10H), 1.12 (d, J = 5.0 Hz 3H). ^{13}C NMR (126 MHz, d_6-acetone) δ 168.62, 160.86, 158.62, 158.56, 158.18, 151.59, 145.28, 141.19, 141.12, 135.95, 135.88, 135.84, 135.81, 130.13, 130.08, 129.97, 129.95, 129.46, 129.38, 128.10, 128.04, 127.54, 126.55, 121.23, 121.16, 114.73, 114.48, 112.86, 112.82, 112.80, 86.04, 84.69, 84.62, 67.87, 66.64, 64.76, 64.25, 58.65, 54.57, 43.11, 43.01, 38.40, 25.25, 25.14, 24.02, 24.00, 23.96, 19.82, 0.21. ^{31}P NMR (202 MHz, d_6-acetone) δ 148.44, 148.24. HRMS (ESI-MS) *m/z* calculated for $C_{60}H_{68}N_7O_{12}P_1^+$ $[M+H]^+$: 1110.4742; found: 1110.4792.

*3'-O-(2-cyanoethoxy(diisopropylamino)-phosphino-5'-O-(4,4'-dimethoxytrityl)-N^2-phenoxyacetyl-O^6-(7-phenoxyacetyloxyheptylene)-2'-deoxyguanosine (***4.3c***)*

Compound **4.2c** was co-evaporated with anhydrous MeCN (1 x 10 mL). To a solution of **4.2c** (0.21 g, 0.22 mmol) and DIPEA (60. µL, 0.35 mmol) in anhydrous DCM (2 mL) was added Cl-POCEN*i*Pr$_2$ (64 µL, 0.28 mmol) dropwise while stirring under argon. After 1 h, additional portions of DIPEA (23 µL, 0.13 mmol) and Cl-POCEN*i*Pr$_2$ (24 µL, 0.11 mmol) were added. After 1.5 h the solvent was removed *in vacuo* and the crude compound was redissolved in 50 mL of EtOAc. The compound was washed with $NaHCO_3$ (*aq*, 3 %) (2 x 35 mL), brine (1 x 35 mL), dried over Na_2SO_4 (~2 g), decanted and the solvent removed *in vacuo* The crude compound was purified by silica gel flash column chromatography (conditioned with 0.5 % TEA/EtOAc) with EtOAc (100 %) to obtain 180 mg (73 %) of **4.3c** as a colorless foam. R_f (SiO_2, TLC): 0.85, 0.90 (100 % EtOAc). $\lambda_{max(MeCN)}$ = 269 nm. ^{1}H NMR (500 MHz, d$_6$-acetone) δ 9.29 (s, 2H), 8.22 (s, 1H), 8.22 (s, 1H), 7.44 – 7.36 (m, 4H), 7.36 – 7.11 (m, 24H), 7.09 – 6.87 (m, 12H), 6.84 – 6.68 (m, 9H), 6.51-6.48 (m, 2H), 5.06-5.05 (m, 4H), 4.99-4.94 (m, 1H), 4.93 – 4.87 (m, 1H), 4.71 (s, 4H), 4.61-4.56 (m, 5H), 4.33 – 4.28 (m, 1H), 4.28 – 4.23 (m, 1H), 4.16 (t, *J* = 6.5 Hz, 5H), 3.93 – 3.85 (m, 1H), 3.84 – 3.70 (m, 16H), 3.70 – 3.60 (m, 4H), 3.57-3.50 (m, 2H), 3.41 – 3.30 (m, 2H), 3.22-3.16 (m, 2H), 2.06-2.04 (m, 9H), 1.93 – 1.78 (m, 5H), 1.77 – 1.59 (m, 5H), 1.60 – 1.47 (m, 4H), 1.45 – 1.32 (m, 9H), 1.22 – 1.16 (m, 16H), 1.12 (d, J = 6.8 Hz, 6H).^{13}C NMR (126 MHz, d$_6$-acetone) δ 168.62, 166.71, 161.00, 158.62, 158.56, 158.22, 158.20, 152.67, 152.62, 151.63, 151.58, 145.28, 145.27, 141.12, 141.05, 135.95, 135.88, 135.86, 135.82, 130.13, 130.08, 129.98, 129.95, 129.45, 129.38, 128.11, 128.04, 127.54, 126.54, 126.52, 121.25, 121.22, 121.19, 118.71, 118.67, 118.12, 117.96, 114.73, 114.45, 112.86, 112.85, 112.81, 112.80, 86.16, 86.12, 86.04, 86.03, 85.93, 85.88, 84.69, 84.63, 74.29, 74.15, 73.94, 73.81, 67.90, 67.15, 64.77, 64.54, 64.18, 58.72, 58.66, 58.57, 58.51, 54.58, 54.57, 43.11, 43.01, 38.39, 38.35, 25.67, 25.55, 24.02, 23.99, 23.96, 23.94, 22.26, 19.87, 19.82, 19.76. ^{31}P NMR (202 MHz, d$_6$-acetone) δ 148.39, 148.18. HRMS (ESI-MS) *m/z* calculated for $C_{63}H_{74}N_7O_{12}P^+$ [M+H]: 1152.5211; found: 1152.5193.

*3'-O-(tert-butyldimethylsilyl)-5'-O-levulinoyl-N^2-phenoxyacetyl-2'-deoxyguanosine (***4.5***)*

To a solution of *3'-O-(tert-butyldimethylsilyl)-N^2-phenoxyacetyl-2'-deoxyguanosine* (0.38 g, 0.73 mmol), EDAC-HCl (0.21 g, 1.8 mmol) and DMAP (cat. 2 mg) in anhydrous dioxane: DMF (10 mL: 1.8 mL) was added levulinic acid (0.26 g, 1.8 mmol) while stirring under argon. After 18 h,

the solvent was removed *in vacuo* and the crude compound redissolved in 100 mL of DCM. The crude compound was washed with $NaHCO_3$ (*aq*, 3 % w/v) (2 x 75 mL), brine (1 x 75 mL), dried over Na_2SO_4 (~4 g), decanted and the solvent removed *in vacuo* The crude compound was purified by silica gel flash column chromatography with MeOH:DCM (1.5 %) to obtain 243 mg (54 %) of **4.5** as a colourless foam. R_f (SiO_2, TLC): 0.22 (5 %, MeOH:DCM). $\lambda_{max(MeCN)}$ = 253 nm. ^{1}H NMR (500 MHz, $CDCl_3$) δ 11.80 (s, 1H, NH), 9.13 (s, 1H, NH), 7.91 (s, 1H, C8), 7.40-7.32 (m, 2H, Ar), 7.08-7.05 (m, 3H, Ar), 6.27 (t, *J* = 6.5 Hz, 1H, H1'), 4.70 (s, 2H, $PhOCH_2CO$), 4.55- (m, 1H, H3'), 4.32 (dd, *J* = 13.9, 4.4 Hz, 2H, H5' and H5''), 4.15-4.13 (m, 1H, H4'), 2.80 – 2.64 (m, 3H, -CH_2- [Lev] + H2'), 2.58-2.51 (m, 2H, -CH_2- [Lev]), 2.43-2.39 (m, 1H, H2''), 2.13 (s, 3H, CH_3 [Lev]), 0.92 (s, 9H, $Si(CH_3)_3$), 0.12 (s, 6H, 2 x Si-CH_3). ^{13}C NMR (126 MHz, $CDCl_3$) δ 206.27, 172.56, 169.54, 156.42, 155.26, 147.47, 146.08, 137.46, 130.00, 123.04, 122.48, 114.92, 85.18, 84.20, 72.09, 67.02, 63.19, 40.49, 37.80, 29.68, 27.75, 25.69, 17.95, -0.02, -4.75, -4.88. HRMS (ESI-MS) *m/z* calculated for $C_{29}H_{39}N_5O_8Si^+$ [M+H]: 614.2646; found: 614.2650.

*1-{O^6-[5'-O-(4,4'-dimethoxytrityl)-3'-O-levulinoyl-N^2-phenoxyacetyl-2'-deoxyguanidinyl]}-4-{O^6-[3'-O-(tert-butyldimethylsilyl)-5'-O-levulinoyl-N^2-phenoxyacetyl-2'-deoxyguanidinyl]}-butane (***4.6a***)*

Compounds **4a** and **5** were co-evaporated with anhydrous MeCN (1 x 10 mL). To a solution of compound **4a** (0.33 g, 0.39 mmol), compound **5** (0.24 g, 0.39 mmol) and PPh_3 (0.10 g, 0.39 mmol) in anhydrous dioxane (12 mL) was added DIAD (80. µL, 0.39 mmol) dropwise while stirring in an ice/water bath under argon. After 18 h, the solvent was removed *in vacuo* and the crude gum was redissolved in DCM (100mL). The crude compound was washed $NaHCO_3$ (*aq*, 3 % w/v) (2 x 75mL), brine (1 x 75 mL), dried over Na_2SO_4 (~4 g), decanted and the solvent removed *in vacuo*. The crude compound was purified by silica gel flash column chromatography using a gradient of EtOAc: hexanes (1:1, 6:4, 7:3, 8:2, 9:1, 9.5:0.5) to yield 285 mg (50 %) of **4.6a** as a colourless foam. R_f (SiO_2, TLC): 0.31 (5 % MeOH:DCM). $\lambda_{max(MeCN)}$ = 269 nm. ^{1}H NMR (500 MHz, $CDCl_3$) δ 8.85 (s, 2H, NH), 8.06 (s, 1H, C8), 8.00 (s, 1H, C8), 7.38-7.31 (m, 6H, Ar), 7.28 – 7.16 (m, 7H, Ar), 7.06-7.00 (m, 6H), 6.78-6.76 (m, 4H, Ar), 6.44 (dd, J = 8.4, 5.8 Hz, 1H, H1'), 6.36 (t, J = 6.4 Hz, 1H, H1'), 5.54-5.53 (m, 1H, H3'), 4.77 (s, 4H, 2 x $PhOCH_2CO$), 4.72-4.68 (m, 5H, H3' + 2 x Ar-CH_2-), 4.38 (d, *J* = 4.8 Hz, 2H, 2 x H5′), 4.28-4.26 (m, 1H, H4′), 4.17 – 4.14 (m, 1H, H4′), 3.76 (s, 6H, 2 x OCH_3), 3.49 (dd, J = 14.8, 4.4 Hz, 1H, H5′), 3.36 (dd, J = 13.8, 3.5 Hz, 1H, H5′), 2.94

(m, H, H2′), 2.86 – 2.51 (m, 10H, 2 x -$(CH_2)_2$- [Lev] + 2 x H2′), 2.46-2.41 (m, 1H, H2′), 2.19 (s, 3H, CH_3 [Lev]), 2.15 (s, 3H, CH_3 [Lev]), 2.13 – 2.07 (m, 4H, -$(CH_2)_2$-), 0.92 (s, 9H, $Si(CH_3)_3$), 0.12 (s, 6H, 2 x Si-CH_3). ^{13}C NMR (126 MHz, $CDCl_3$) δ 172.42, 161.29, 158.50, 144.40, 135.50, 135.44, 130.00, 129.77, 128.05, 127.84, 122.25, 114.92, 114.90, 113.14, 86.66, 55.16, 37.83, 29.67, 29.09, 28.75, 27.97, 27.77, 25.85, 25.70, 17.92, -4.77, -4.91. HRMS (ESI-MS) *m/z* calculated for $C_{77}H_{88}N_{10}O_{18}Si^+$ [M+H]: 1469.6126; found: 1469.6085.

*1-{O^6-[5'-O-(4,4'-dimethoxytrityl)-3'-O-levulinoyl-N^2-phenoxyacetyl-2'-deoxyguanidinyl]}-7-{O^6-[3'-O-(tert-butyldimethylsilyl)-5'-O-levulinoyl-N^2-phenoxyacetyl-2'-deoxyguanidinyl]}-heptane (***4.6b***)*

Compounds **4.4b** and **4.5** were co-evaporated with anhydrous MeCN (1 x 10 mL). To a solution of compound **4.4b** (0.39 g, 0.43 mmol), compound **4.5** (0.26 g, 0.43 mmol) and PPh_3 (0.11 g, 0.43 mmol) in anhydrous dioxane (13 mL) was added DIAD (83 µL, 0.43 mmol) dropwise while stirring in an ice/water bath under argon. After 18 h, additional PPh_3 (0.11 g, 0.43 mmol) and DIAD (83 µL, 0.43 mmol) were added dropwise while stirring in an ice/water bath under argon. After 22 h the solvent was removed *in vacuo* and the crude gum was redissolved in DCM (100mL). The crude compound was washed $NaHCO_3$ (*aq*, 3 % w/v) (2 x 75mL), brine (1 x 75 mL), dried over Na_2SO_4 (~4 g), decanted and the solvent removed *in vacuo*. The crude compound was purified by silica gel flash column chromatography using a gradient of EtOAc: hexanes (1:1, 6:4, 7:3, 8:2, 9:1) to yield 410 mg (64 %) of **4.6b** as a colourless foam. R_f (SiO_2, TLC): 0.30 (5 % MeOH:DCM). $\lambda_{max(MeCN)}$ = 269 nm. ^{1}H NMR (500 MHz, $CDCl_3$) δ 8.75 (s, 1H, NH), 8.66 (s, 1H NH), 8.00 (s, 1H, C8), 7.98 (s, 1H, C8), 7.40 – 7.29 (m, 6H, Ar), 7.27-7.16 (m, 7H, Ar), 7.06 – 6.95 (m, 6H, Ar), 6.79-6.75 (m, 4H, Ar), 6.46 – 6.40 (dd, J = 8.4 Hz. 5.9 Hz, 1H, H1'), 6.35 (t, *J* = 6.5 Hz, 1H, H1'), 5.55-5.53 (m, 1H, H3'), 4.75 (m, 4H, 2 x $PhOCH_2CO$), 4.72-4.69 (m, 1H, H3'), 4.59-4.56 (m, 4H, 2 x Ar- CH_2-), 4.37 (d, *J* = 4.8 Hz, 2H, 2 x H5'), 4.27-4.25 (m, 1H, H4′), 4.15-4.13 (m, 1H, H4'), 3.75 (s, 6H, 2 x OCH_3), 3.48 (dd, J = 10.5, 4.4 Hz, 1H, H5'), 3.34 (dd, J = 10.4, 3.4 Hz, 1H, H5'), 2.99 – 2.88 (m, 2H, 2 x H2'), 2.80 – 2.51 (m, 9H, -$(CH_2)_4$- [Lev], H2'), 2.45 – 2.37 (m, 1H, H2'), 2.19 (s, 3H, -CH_3 [Lev]), 2.14 (s, 3H , -CH_3 [Lev]), 1.93-1.86 (m, 4H, -$(CH_2)_2$-), 1.55-1.43 (m, 6H, - $(CH_2)_3$ -), 0.90 (s, 9H, $Si(CH_3)_3$), 0.11 (s, 6H, 2 x Si-(CH_3)). ^{13}C NMR (126 MHz, $CDCl_3$) δ 206.34, 206.29, 172.43, 172.13, 161.29, 161.25, 158.51, 157.21, 152.47, 152.14, 151.18, 151.02, 144.41, 140.53, 139.80, 135.51, 135.45, 130.01, 129.78, 128.06, 127.85, 126.90, 122.26, 114.92,

114.91, 113.15, 86.67, 85.16, 84.78, 84.35, 83.99, 75.47, 72.16, 68.01, 67.85, 67.80, 63.86, 63.74, 55.17, 40.11, 38.14, 37.84, 29.79, 29.76, 29.68, 29.09, 28.76, 27.98, 27.78, 25.86, 25.71, 17.93, -0.02, -4.76, -4.90. HRMS (ESI-MS) *m/z* calculated for $C_{80}H_{94}N_{10}O_{18}Si^+$ [M+H]: 1511.6595; found: 1511.6586.

*1-{O^6-[5'-O-(4,4'-dimethoxytrityl)-3'-O-levulinoyl-N^2-phenoxyacetyl-2'-deoxyguanidinyl]}-4-{O^6-[5'-O-levulinoyl-N^2-phenoxyacetyl-2'-deoxyguanidinyl]}-butane (***4.7a***)*

To a solution of **4.6a** (0.28 g, 0.19 mmol) in anhydrous THF (4 mL) was added TBAF (190 µL of 1 M in THF, 0.19 mmol) dropwise while stirring under argon. After 30 minutes the volume was brought to 100 mL with DCM, the crude compound was washed with $NaHCO_3$ (*aq*, 3 % w/v) (2 x 75 mL), brine (1 x 75 mL), dried over Na_2SO_4 (~4 g), decanted and the solvent removed *in vacuo*. The crude compound was purified by silica gel flash column chromatography with MeOH:DCM (1.5 %) to obtain 171 mg (67 %) of **4.7a** as a colourless foam. R_f (SiO_2, TLC): 0.19 (5 % MeOH:DCM). $\lambda_{max(MeCN)}$ = 269 nm. ^{1}H NMR (500 MHz, $CDCl_3$) δ 8.91 (s, 1H, NH), 8.81 (s, 1H, NH), 8.06 (s, 1H, C8), 7.98 (s, 1H, C8), 7.39 – 7.11 (m, 13H, Ar), 7.01-6.97 (m, 6H), 6.75-6.73 (m, 4H, Ar), 6.48 (t, *J* = 6.3 Hz, 1H, H1'), 6.45 – 6.38 (dd, J = 7.9 Hz, 7.9 Hz, 1H, H1'), 5.52-5.51 (m, 1H, H3'), 4.83-4.82 (m, 1H, H3'), 4.77-4.74 (m, 4H, 2 x $PhOCH_2CO$), 4.70 – 4.61 (m, 4H, 2 x $–(CH_2)_2$-), 4.40-4.34 (m, 2H, 2 x H5'), 4.25-4.21 (m, 2H, 2 x H4'), 3.73 (s, 6H, 2 x OCH_3), 3.46 (dd, *J* = 10.4, 4.4 Hz, 1H, H5'), 3.33 (dd, *J* = 10.3, 3.1 Hz, 1H, H5'), 2.95-2.89 (m 1H, H2'), 2.88 – 2.45 (m, 11H, 3 x H2', 4 x $–(CH_2)_2$ [Lev]), 2.16 (s, 3H, $-CH_3$ [Lev]), 2.11 (s, 3H, $-CH_3$ [Lev]), 2.11-2.09 (m, 4H, $-(CH_2)_2$). ^{13}C NMR (126 MHz, $CDCl_3$) δ 206.85, 206.36, 172.63, 172.13, 161.06, 161.05, 158.49, 157.14, 152.46, 152.21, 151.21, 150.95, 144.40, 140.64, 139.92, 135.49, 135.43, 129.99, 129.75, 129.73, 128.04, 127.83, 126.89, 122.21, 122.14, 119.04, 118.67, 114.89, 114.87, 113.13, 86.64, 84.78, 84.46, 84.33, 84.03, 75.42, 71.51, 67.92, 67.13, 64.16, 63.85, 55.15, 46.02, 39.95, 38.09, 37.92, 37.82, 29.77, 29.74, 29.66, 27.96, 27.84, 25.41, 25.28, 10.14. HRMS (ESI-MS) *m/z* calculated $C_{71}H_{74}N_{10}O_{18}^+$ [M+H]: 1355.5261; found: 1355.5291.

*1-{O^6-[5'-O-(4,4'-dimethoxytrityl)-3'-O-levulinoyl-N^2-phenoxyacetyl-2'-deoxyguanidinyl]}-7-{O^6-[5'-O-levulinoyl-N^2-phenoxyacetyl-2'-deoxyguanidinyl]}-heptane (***4.7b***)*

To a solution of **4.6b** (0.39 g, 0.19 mmol) in anhydrous THF (6 mL) was added TBAF (310 µL of 1 M in THF, 0.31 mmol) dropwise while stirring under argon. After 45 minutes the volume was

brought to 100 mL with DCM, the crude compound was washed with $NaHCO_3$ (*aq*, 3 % w/v) (2 x 75 mL), brine (1 x 75 mL), dried over Na_2SO_4 (~4 g), decanted and the solvent removed *in vacuo*. The crude compound was purified by silica gel flash column chromatography using a gradient of MeOH:DCM (1, 1.25, 1.5 %) to obtain 327 mg (91 %) of **4.7b** as a colourless foam. R_f (SiO_2, TLC): 0.21 (5 % MeOH:DCM). $\lambda_{max(MeCN)}$ = 268 nm. ^{1}H NMR (500 MHz, $CDCl_3$) δ 8.81 (s, 1H, NH), 8.66 (s, 1H, NH), 7.98 (s, 2H, 2 x C8), 7.38-7.32 (m, 6H, Ar), 7.29 – 7.12 (m, 3H, Ar), 7.06 – 6.97 (m, 6H, Ar), 6.79 – 6.68 (m, 4H, Ar), 6.45 (dd, J = 8.4 Hz, 5.9 Hz, 1H, H1'), 6.40 (t, J = 6.2 Hz, 1H, H1'), 5.55-5.53 (m, 1H, H3'), 4.98-4.95 (m, 1H. H3'), 4.75 (2x s, 4H, 2 x $PhOCH_2CO$), 4.59-4.55 (m, 4H, 2 x Ar–(CH_2)), 4.42-4.41 (m, 2H, 2 x H5'), 4.27-4.26 (m, 1H, H4'), 4.21-4.18 (m, 1H, H4'), 3.75 (s, 6H, 2 x OCH_3), 3.52 – 3.43 (m, 1H, H5'), 3.43 – 3.29 (m, 1H, H5'), 2.97-2.90 (m, 2H, H2', H2''), 2.82 – 2.67 (m, 4H, -(CH_2)$_2$- [Lev]), 2.66 – 2.42 (m, 6H, -(CH_2)$_2$- [Lev] and 2 x H2'), 2.19 (s, 3H, -CH_3 [Lev]), 2.15 (s, 3H, -CH_3 [Lev]), 1.92-1.86 (m, 4H, -(CH_2)$_2$-), 1.48 (m, 6H, –(CH_2)$_3$–). ^{13}C NMR (126 MHz, $CDCl_3$) δ 207.01, 206.35, 172.70, 172.14, 161.28, 161.24, 158.51, 152.45, 152.08, 151.17, 150.94, 144.40, 140.54, 139.82, 135.51, 135.45, 130.01, 129.80, 129.78, 128.06, 127.85, 126.90, 122.33, 122.25, 118.78, 114.92, 114.90, 113.15, 86.67, 84.68, 84.47, 84.34, 84.00, 75.46, 71.59, 67.97, 67.84, 67.80, 64.01, 63.86, 55.17, 39.77, 38.14, 37.98, 37.84, 29.79, 29.76, 29.68, 29.05, 28.73, 27.98, 27.87, 25.84, -0.02. HRMS (ESI-MS) *m/z* calculated $C_{74}H_{80}N_{10}O_{18}^+$ [M+H]: 1397.5730; found: 1397.5773.

*1-{O^6-[5'-O-(4,4'-dimethoxytrityl)-3'-O-levulinoyl-N^2-phenoxyacetyl-2'-deoxyguanidinyl]}-4-{O^6-[3'-O-(2-cyanoethoxy(diisopropylamino)-phosphino)-5'-O-levulinoyl-N^2-phenoxyacetyl-2'-deoxyguanidinyl]}-butane (***4.8a***)*

Compound **4.7a** was co-evaporated with anhydrous MeCN(1 x 10 mL). To a solution of **4.7a** (0.11 g, 0.078 mmol) and DIPEA (22 µL, 0.13 mmol) in anhydrous DCM (2 mL) was added Cl-POCEN*i*Pr$_2$ (23 µL, 0.10 mmol) dropwise while stirring under argon. After 45 minutes, additional DIPEA (11 µL, 0.063 mmol) and Cl-POCEN*i*Pr$_2$ (12 µL, 0.050 mmol) was added. After 1 h, the solvent was removed *in vacuo* and the crude compound was redissolved in 50 mL of EtOAc. The compound was washed with $NaHCO_3$ (*aq*, 3 %) (2 x 35 mL), brine (1 x 35 mL) and dried over Na_2SO_4 (~2 g). The crude compound was purified by silica gel flash column chromatography (conditioned with 0.5 % TEA/EtOAc) with MeCN:EtOAc (2:8) to obtain 67 mg (55 %) of **4.8a** as a colourless foam. R_f (SiO_2, TLC): 0.095 (100 % EtOAc). $\lambda_{max(MeCN)}$ = 268 nm. ^{1}H NMR (500

MHz, d_6-acetone) δ 9.44 (s, 1H), 9.36 (s, 1H), 8.24 (m, 2H), 8.20 (s, 1H), 7.39 – 6.64 (m, 26H), 6.52 – 6.43 (m, 2H), 5.54-5.51 (m, 1H), 5.07-5.05 (m, 5H), 5.00-4.90 (m, 1H), 4.78 – 4.68 (m, 5H), 4.50 – 4.23 (m, 5H), 3.98 – 3.90 (m, 1H), 3.89-3.80 (m, 1H), 3.74 (s, 6H), 3.72 – 3.59 (m, 4H), 3.41-3.36 (m, 1H), 3.32-3.29 (m, 1H), 3.25 – 3.10 (m, 2H), 2.84 – 2.46 (m, 16H), 2.12-2.10 (m, 2H), 2.09-2.08 (m, 3H), 2.06-2.04 (m, 4H), 1.97-1.96 (m, 1H), 1.27 – 1.18 (m, 16H). ^{13}C NMR (126 MHz, d_6-acetone) δ 171.88, 160.87, 158.60, 158.52, 152.68, 151.55, 145.26, 141.36, 135.87, 135.75, 130.10, 129.89, 129.42, 128.02, 127.50, 126.52, 121.16, 114.73, 114.72, 112.83, 112.76, 86.03, 84.97, 84.73, 75.29, 67.91, 66.84, 66.78, 64.64, 54.55, 43.14, 37.33, 36.02, 27.77, 27.57, 25.27, 24.02, 23.97, 23.91. ^{31}P NMR (202 MHz, d_6-acetone) δ 148.49. HRMS (ESI-MS) *m/z* calculated $C_{80}H_{91}N_{12}O_{19}P^+$ [M+H]: 1555.6339; found: 1555.6380.

*1-{O^6-[5'-O-(4,4'-dimethoxytrityl)-3'-O-levulinoyl-N^2-phenoxyacetyl-2'-deoxyguanidinyl]}-7-{O^6-[3'-O-(2-cyanoethoxy(diisopropylamino)-phosphino)-5'-O-levulinoyl--N^2-phenoxyacetyl-2'-deoxyguanidinyl]}-heptane (***4.8b***)*

Compound **4.7b** was co-evaporated with anhydrous MeCN (1 x 10 mL). To a solution of **4.7b** (0.11 g, 0.080 mmol) and DIPEA (22 µL, 0.13 mmol) in anhydrous DCM (2 mL) was added Cl-POCEN*i*Pr$_2$ (23 µL, 0.10 mmol) dropwise while stirring under argon. After 30 minutes, additional DIPEA (18 µL, 0.11 mmol) and Cl-POCEN*i*Pr$_2$ (17 µL, 0.07 mmol) was added. After 45 minutes the solvent was removed *in vacuo* and the crude compound was redissolved in 50 mL of EtOAc. The compound was washed with $NaHCO_3$ (*aq*, 3 %) (2 x 35 mL), brine (1 x 35 mL) and dried over Na_2SO_4 (~2 g). The crude compound was purified by silica gel flash column chromatography (conditioned with 0.5 % TEA/EtOAc) with MeCN:EtOAc (2:8) to obtain 71 mg (55 %) of **4.8b** as a colourless foam. R_f (SiO_2, TLC): 0.17 (100 % EtOAc). $\lambda_{max(MeCN)}$ = 269 nm. ^{1}H NMR (500 MHz, d_6-acetone) δ 9.41 (s, 1H), 9.33 (s, 1H), 8.24 (s, 1H), 8.21 (s, 1H), 7.80 – 6.62 (m, 27H), 6.48 (m, 2H), 5.70 – 5.42 (m, 1H), 5.06-5.04 (m, 4H), 5.02 – 4.88 (m, 2H), 4.68 – 4.52 (m, 5H), 4.51 – 4.26 (m, 5H), 4.01 – 3.80 (m, 3H), 3.73 (s, 3H), 3.73 – 3.59 (m, 2H), 3.42-3.36 (m, 1H), 3.32-3.30 (m, 1H), 3.22-3.14 (m, 1H), 2.87 – 2.69 (m, 10H), 2.63 – 2.38 (m, 7H), 2.14 (s, 3H), 2.10-2.09 (m, 3H), 2.06-2.05 (m, 3H), 1.97 (s, 3H), 1.94 – 1.79 (m, 5H), 1.55-1.48 (m, 7H), 1.25-1.19 (m, 19H). ^{13}C NMR (126 MHz, d_6-acetone) δ 171.89, 160.98, 158.52, 158.18, 152.67, 145.27, 141.28, 135.88, 135.76, 130.11, 129.90, 129.44, 128.03, 127.51, 126.52, 121.21, 114.72, 112.83, 112.76, 86.03, 84.94, 84.74, 75.31, 67.89, 67.15, 54.57, 37.34, 36.02, 27.77, 25.69, 23.98. ^{31}P NMR (202

MHz, d_6-acetone) δ 148.47. HRMS (ESI-MS) *m/z* calculated $C_{83}H_{97}N_{12}O_{19}P^+$ [M+H]: 1597.6809; found: 1597.6792.

4.4.1.3 Synthesis, purification and characterization of oligonucleotides

The solid-phase synthesis of modified oligonucleotides was performed using an ABI Model 3400 synthesizer on a 1.5 μmol scale using standard β-cyanoethylphosphoramidite chemistry with controlled pore glass (CPG) from ChemGenes (Wilmington, MA). Concentrations were 0.1 M (in MeCN) for standard 3'-*O*-2'-deoxyphosphoramidites and ranged from 0.1 to 0.15 M (in MeCN) for the synthesized nucleoside phosphoramidites. Coupling times of compounds **4.8a** and **4.8b** were increased to 5 minutes for monoalkylated nucleoside phosphoramidite and 10 minutes for IaCLs to maximize coupling efficiency. DMT groups were removed over a period of 45 seconds using 3 % trichloroacetic acid (in dichloromethane). Phenoxyacetic anhydride was used as the capping agent. Cleavage of the oligonucleotides from the CPG support and simultaneous deprotection was achieved with 1 mL of NH_4OH (28 %):EtOH (3:1) for a minimum of 4 h at 55 °C. The supernatant was transferred and the CPG supports were washed with 2 x 200 μL of 50 % MeCN then dried down with a speed vacuum concentrator. The quality of synthesis was assessed by loading ~0.1 OD_{260} of crude sample on a 20 % (19:1 acrylamide:bisacrylamide) denaturing PAGE run at 200 V for 45 minutes and visualized by UV shadowing. The oligonucleotides were purified by preparatory denaturing PAGE. The preparatory denaturing PAGE consisted of a 20 % acrylamide solution (19:1 acrylamide:bis-acrylamide) in 1× TBE running buffer on standard 16 x 18 cm glass plates at 450 V until sufficient separation was achieved (approximately 4 h). The desired band was excised and the pure oligonucleotide extracted with 8-10 mL of 0.1 M sodium acetate (NaOAc) solution on a mixer overnight. The purified oligonucleotides were then desalted with a C-18 SEP PAK cartridge. The cartridge was prepared by washing with 10 mL of HPLC grade MeCN, followed by 50 % MeCN (*aq*) and equilibrated with 0.1 M NaOAc. The oligonucleotides were adsorbed to the C-18 column then the salt was removed with 2 x 10 mL of H_2O. The sample was eluted with MeOH:H_2O:MeCN (2:1:1). Purity was assessed to be greater than 90% for all oligonucleotides as determined by analytical denaturing PAGE. All oligonucleotides were quantitated using a Varian Cary Model 3E spectrophotometer. Single strand concentrations were calculated using the Beer-Lambert law and the absorbance was measured at 260 nm. Molar extinction coefficients were calculated by the nearest neighbour approximation.[1]

The identity of the oligonucleotides were confirmed by mass spectrometry at the Concordia University Centre for Biological Applications of Mass Spectrometry (see **Section 4.4.2 Table S4.1**). 0.15 OD of oligonucleotide was dried down for ESI-MS analysis. LC-MS analyses were performed on an Agilent 1100 LC system coupled to a Thermo LTQ Orbitrap Velos mass spectrometer equipped with a heated electrospray ion source at negative mode. A Spursil C18-EP column (50 x 2.1 mm and 3 μM particle diameter, Dikma Technologies) was used and oligonucleotides were eluted using a 20-min gradient at an initial flow rate of 250 μL/min with mobile phase A (10 mM ammonium acetate and 1 mM ammonium fluoride water solution) and B (MeCN). The gradient started at 2% B and held for 3 min, linear gradients were achieved to 50% B at 8 min, to 90% B at 10 min, then kept isocratic with 90% B for 2 min. The column was reconditioned from 13 min with 2% B at a flow rate of 400 μL/min for 5 min and at 250 μL/min for extra 2 min. Dried samples were reconstituted in 50 μL of mobile phase A and the injection volume was 10 μL. The divert valve was set at 0 min to the waste, and at 4.0 min to the detector. MS spectra (*m/z* 300-2000) were acquired in the Orbitrap at a resolution of 60,000. The uncharged monoisotopic mass of nucleotides were calculated using Thermo FreeStyle™ software (v1.7 SP1).

4.4.1.4 UV Thermal denaturation studies

0.5 OD of oligonucleotide was dissolved in 1 mL of buffer (in 2.5 mM KH_2PO_4, 2.5 mM K_2HPO_4, 3.2 mM KOH, pH 7.2 or 130 mM KCl, 2.5 mM KH_2PO_4, 2.5 mM K_2HPO_4, 5.4 mM KOH, pH 7.2), heated at 95 °C for 5 minutes then slowly cooled to room temperature overnight. The samples were then degassed on a speed-vacuum concentrator for 2 minutes. Samples were equilibrated at 10 °C or 15 °C for 5 minutes. UV thermal denaturation data was acquired with a Varian CARY Model 3E Spectrophotometer fitted with a 12-sample thermostated cell block and a temperature controller, with absorbance measured at 260 nm and 295 nm and a heating rate at of 0.5 °C/min. The melting temperature (T_m) was calculated as the maximum of the first derivative according to the method of Puglisi and Tinoco.[189] All data analysis was performed with Microsoft Excel™.

4.4.1.5 Circular dichroism spectroscopy

Circular dichroism (CD) spectra were acquired at the Concordia University Integrated Platform for Biomolecular Function, Interactions and Structure (BIOFINS) using a Jasco J-815 spectropolarimeter equipped with a Julaba F25 circulating bath. The aforementioned T_m samples

(cooled to 5-15 °C at a rate of 0.5 °C/min) were analyzed by CD. The samples were transferred to quartz cells (Starna 29-Q-10) and spectra were recorded from 220 to 350 nm at 20 or 37 ºC. Each spectrum was an average of 5 scans, collected at a rate of 20 or 50 nm min^{-1} with a 1 nm bandwidth and a 0.2 nm sampling wavelength. The molar ellipticity (φ) was calculated by the equation $\varphi = \varepsilon/Cl$, where ε is the relative ellipticity (mdeg), C is the oligonucleotide concentration (moles/L), and l is the cell path length (cm). The data were processed using the software provided by Jasco and transferred to Microsoft Excel™ to prepare the spectrum.

4.4.1.6 Protein purification and overexpression

AGT proteins were purified and overexpressed as described previously.[36] The masses of the proteins were verified by ESI-MS run in positive ion mode on a Micromass Q-Tof Ultima API.

4.4.1.7 Repair assays

100 pmol DNA was incubated with 1 µL of γ-[^{32}P]-ATP (10 µCi/µL), 5 units of T4 polynucleotide kinase (PNK) in 10 µL of 1X PNK buffer for 1 h at 37 °C for radiolabelling. After 1 h of incubation, the total solution volume was brought to 50 µL (130mM KCl, 2.5 mM KH_2PO_4, 2.5 mM K_2HPO_4, 5.4 mM KOH buffer, pH 7.2) and the samples were heated at 95 °C for 5 minutes and then slowly cooled to room temperature. For total repair assays 2 pmol of DNA was then incubated with 6, 20 or 60 pmol of protein and the total volume was brought to 15 µL with activity buffer [10 mM Tris–HCl (pH 7.6), 100 mM NaCl and 1 mM DTT]. The reactions were incubated at 37 °C overnight, after which the reactions were quenched by adding 18.2 µL of stop buffer (0.1 M LiOH in formamide) and boiling for 5 minutes. The samples were loaded on a 14 cm × 16 cm, 20 % (19:1 acrylamide bis-acrylamide), 7M urea PAGE which was run for 1h30m for the IaCL series at room temperature and 3-3.5 h with 16 % (19:1 acrylamide bis-acrylamide), 7M urea PAGE for the other oligonucleotides in 1X TBE at 20 °C. To visualize the oligonucleotides, gels were exposed to X-ray film. Repair was quantified using ImageJ software. Relative quantities of substrate and repair products were calculated by summing the integrated bands in each lane and dividing the corresponding band by the total.

4.4.1.8 On repair assays of ODNs containing single hydroxyalkylene adducts

It has been shown that both Ada-C and OGT are not capable of repairing bulkier lesions such as longer alkylene chains, ICLs and IaCLs. Thus, no repair by these AGT proteins at high protein concentrations of the butylene and heptylene in G-quadruplexes was unsurprising (see **Section 4.4.2 Figures S4.44-4.47**, lanes 4-5, 7-8). We noted a general trend that less repair occurred when the internal G_7 was alkylated at the O^6 position for both hydroxybutylene and hydroxyheptylene containing **ODN**s.

4.4.2 Supporting Figures, Tables and Schemes

Figure S4.1 500 MHz ^{1}H NMR spectrum of compound **4.1a** (in $CDCl_3$)

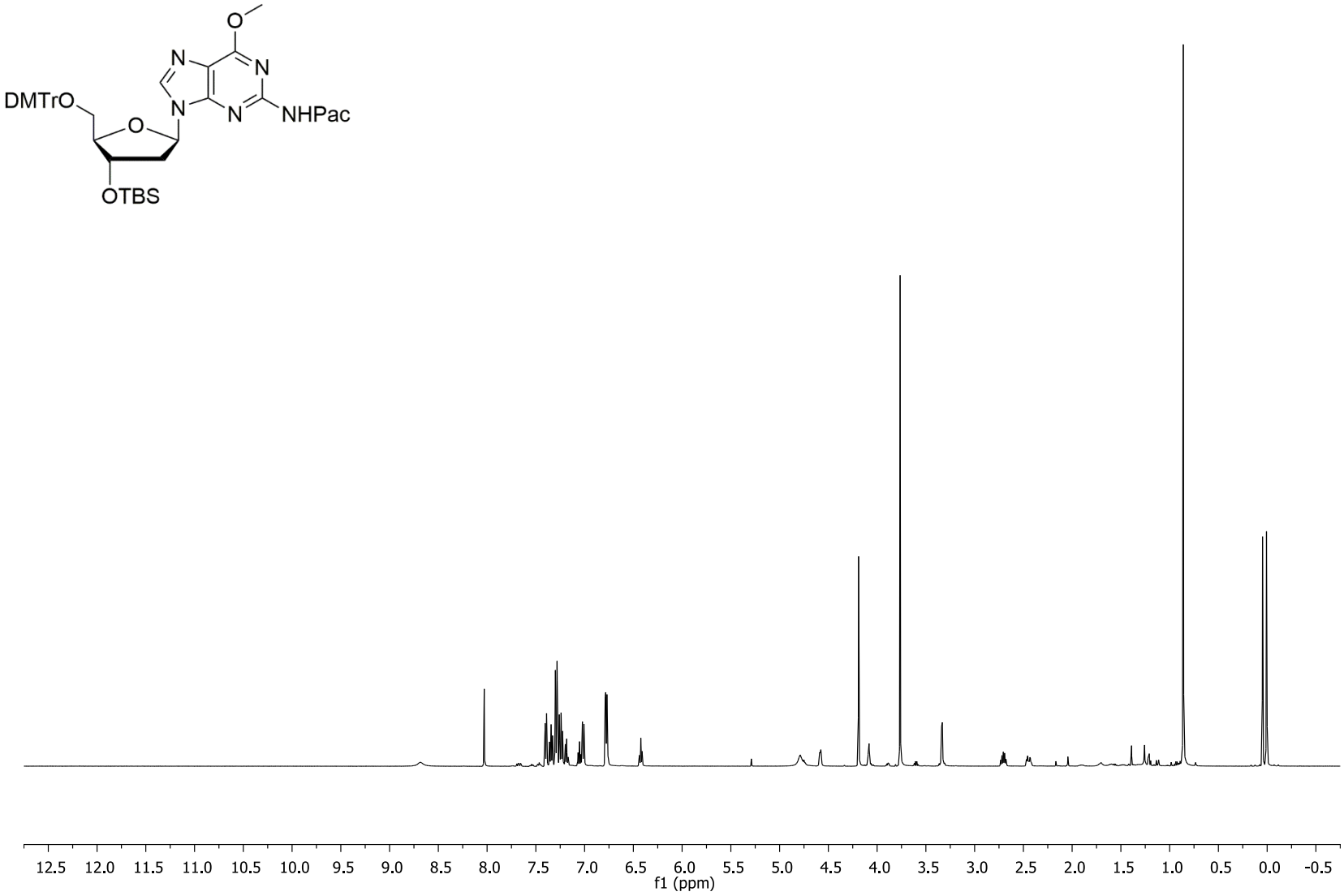


Figure S4.2 125.7 MHz ^{13}C NMR spectrum of compound **4.1a** (in $CDCl_3$)

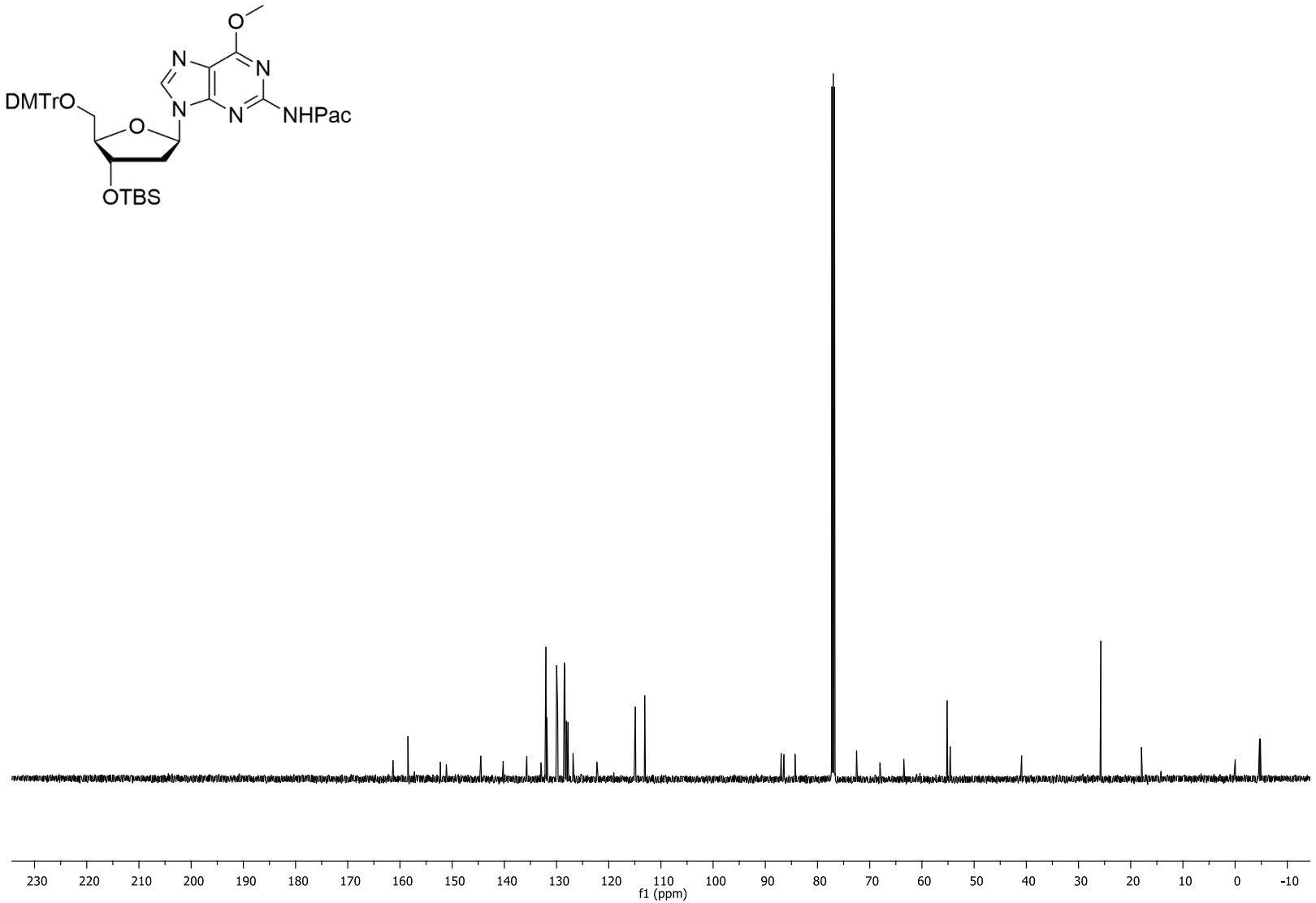


Figure S4.3 500 MHz ^{1}H NMR spectrum of compound **4.2a** (in $CDCl_3$)

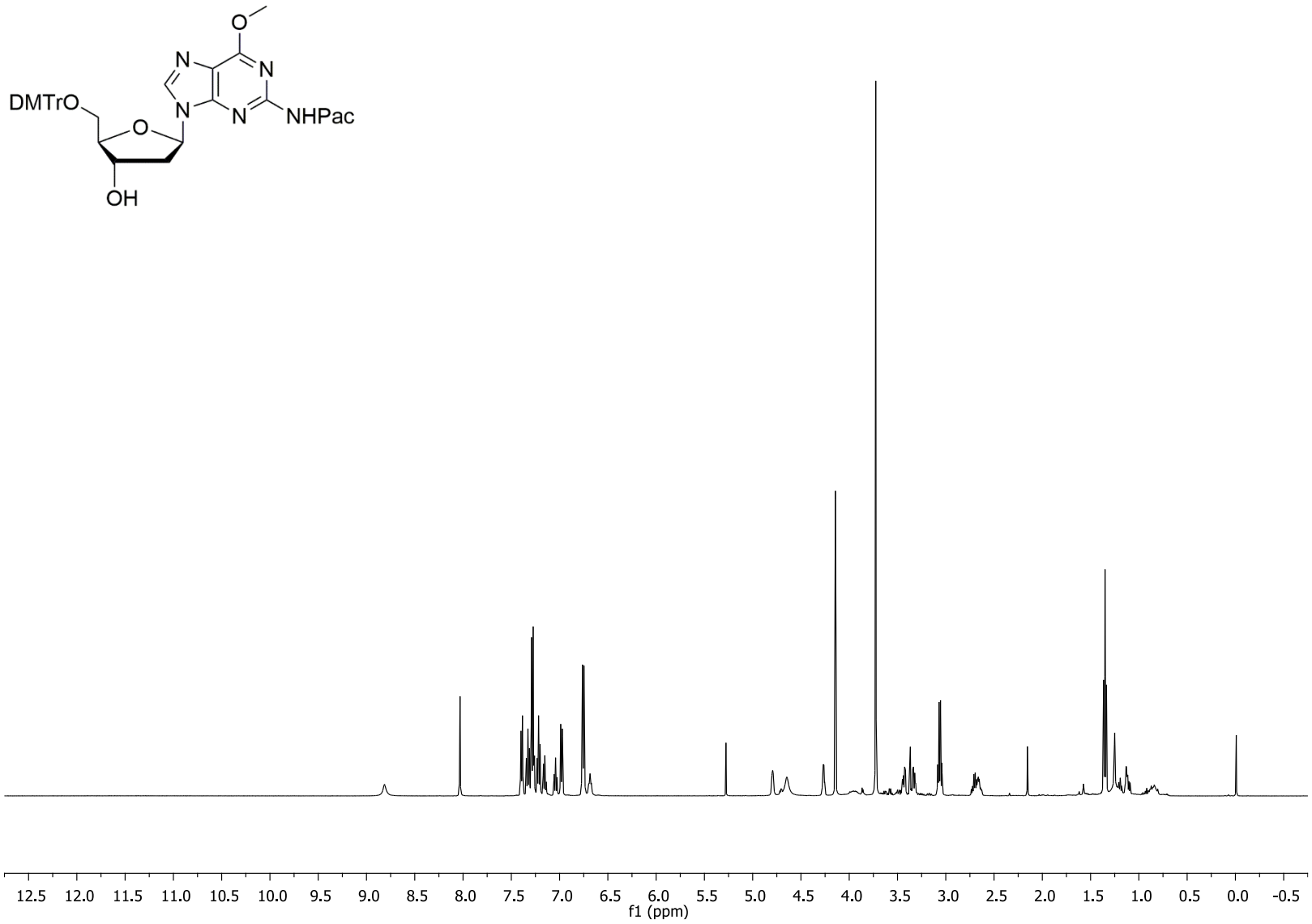


Figure S4.4 125.7 MHz ^{13}C NMR spectrum of compound **4.2a** (in $CDCl_3$)

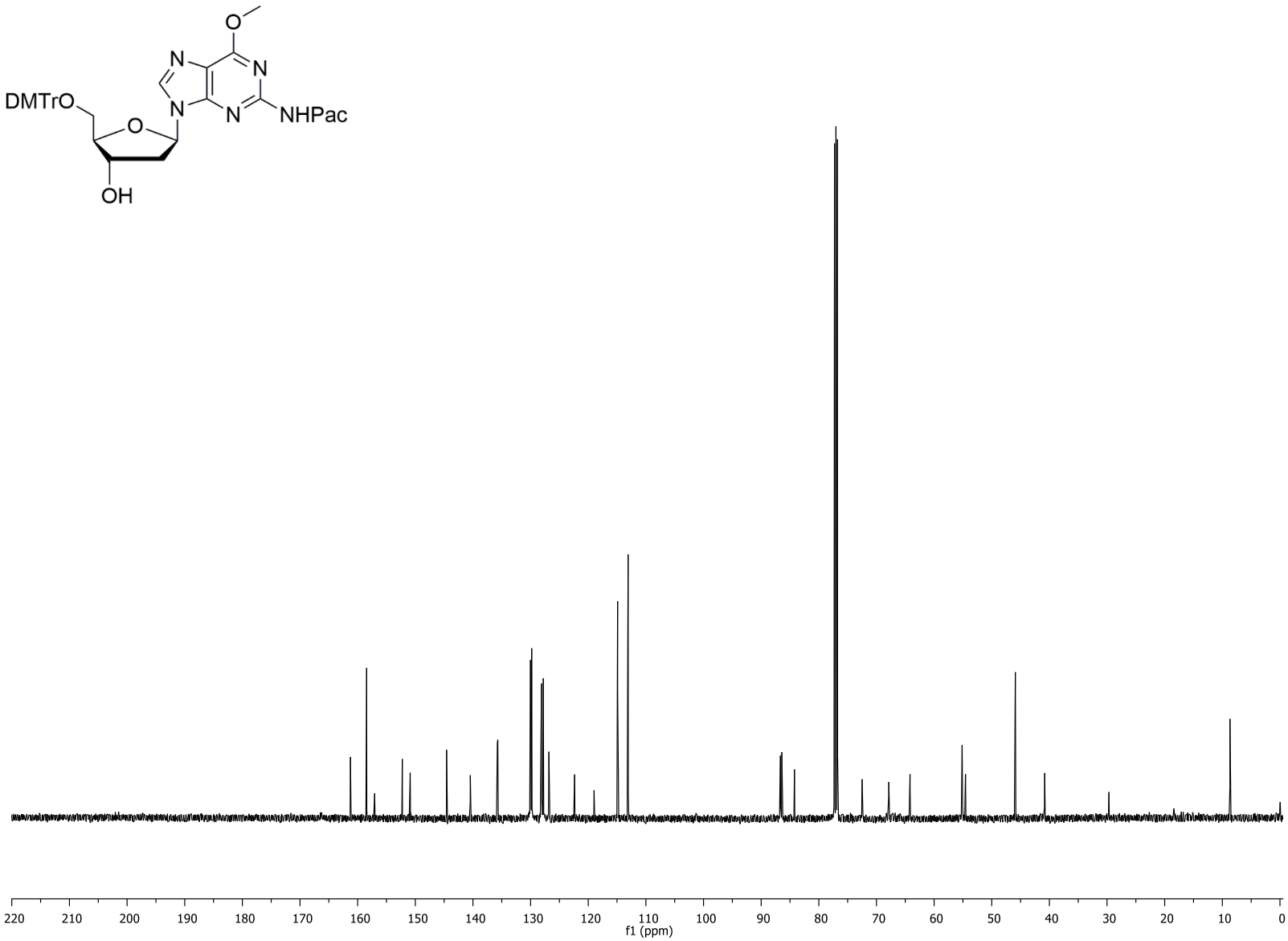


Figure S4.5 500 MHz ^{1}H NMR spectrum of compound **4.3a** (in d_6-acetone)

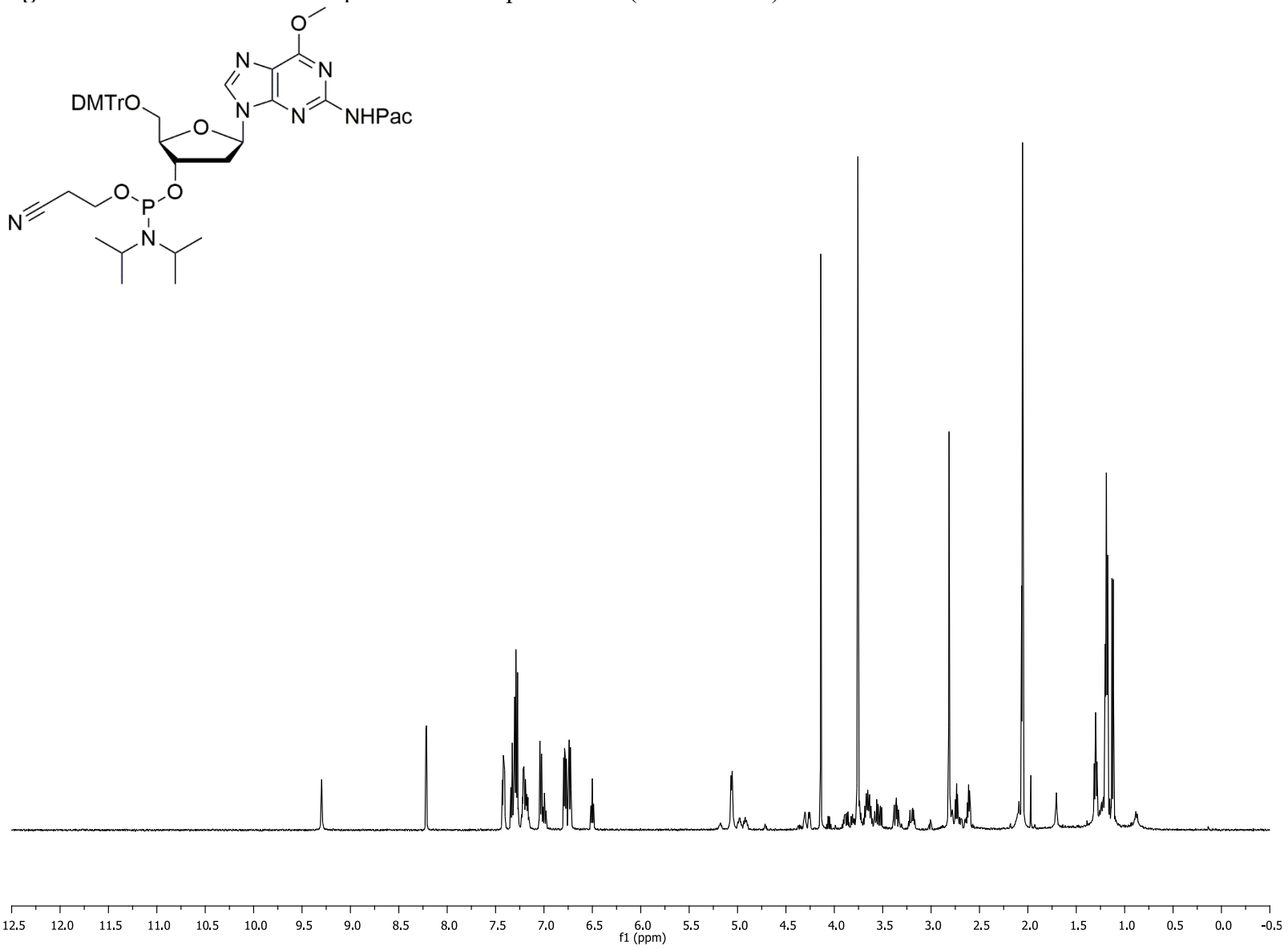


Figure S4.6 125.7 MHz ^{13}C NMR spectrum of compound **4.3a** (in d_6-acetone)

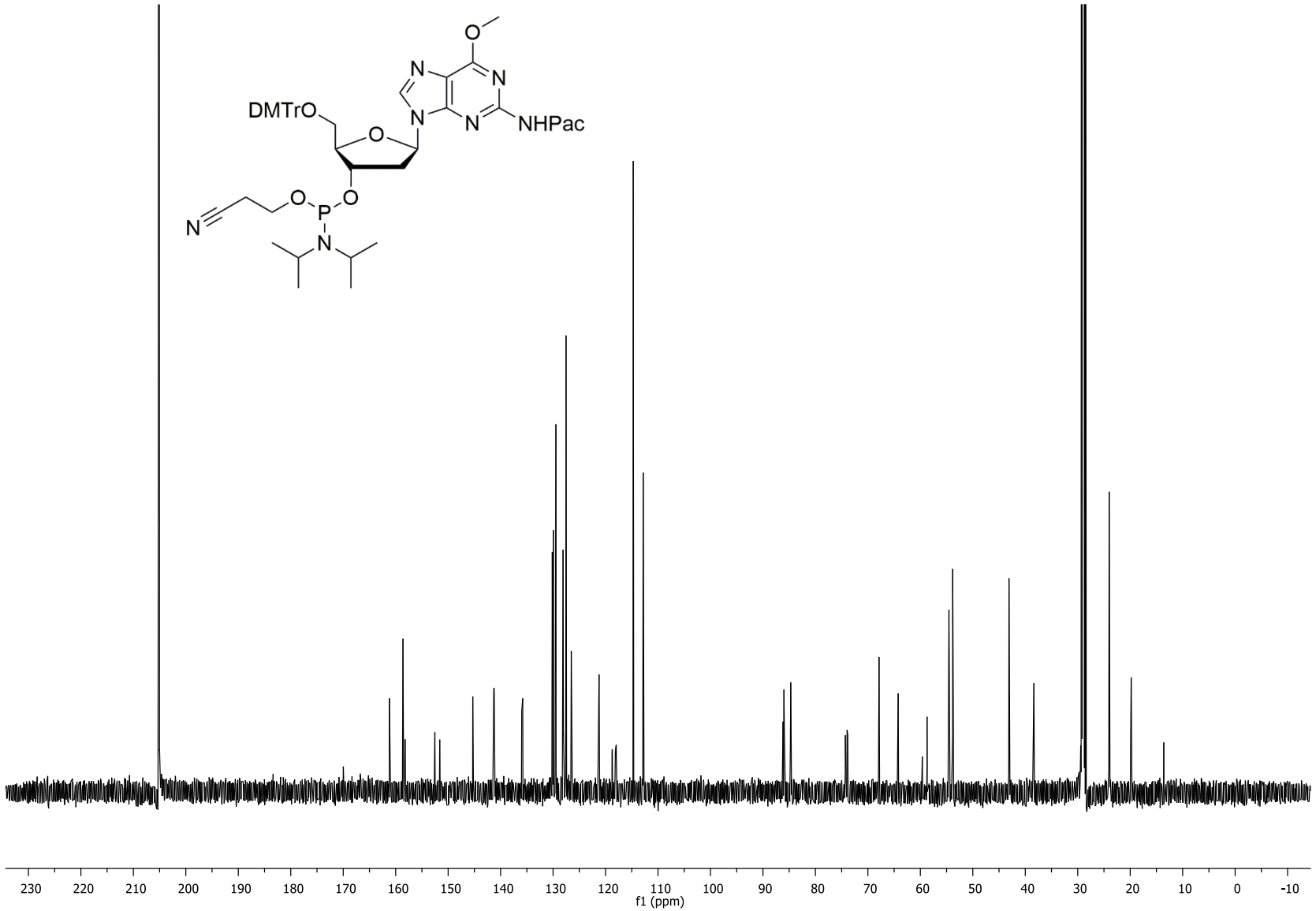


Figure S4.7 202.3 MHz ^{31}P NMR spectrum of compound **4.3a** (in d_6-acetone)

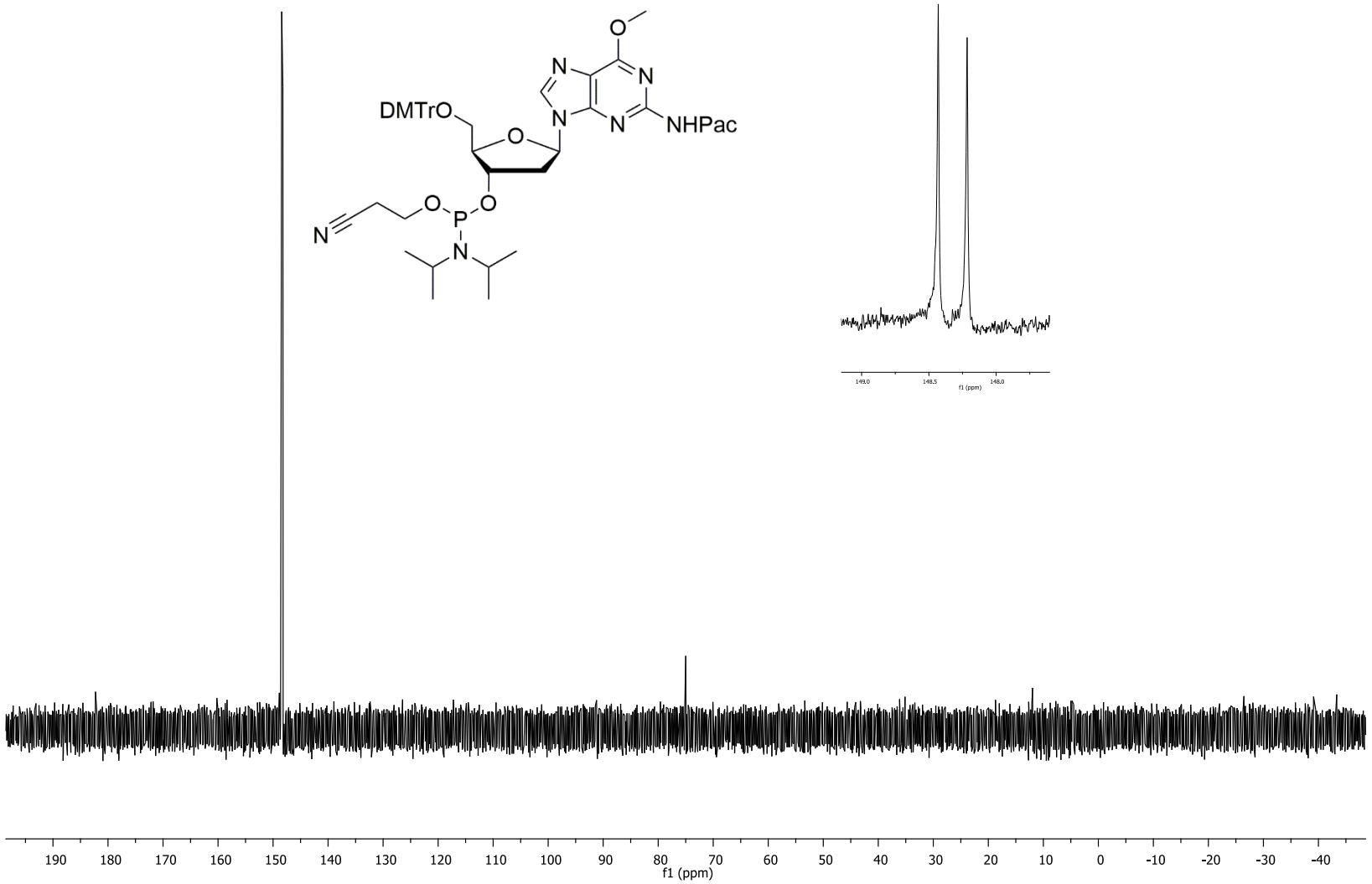


Figure S4.8 500 MHz ^{1}H NMR spectrum of compound **4.1b** (in $CDCl_3$)

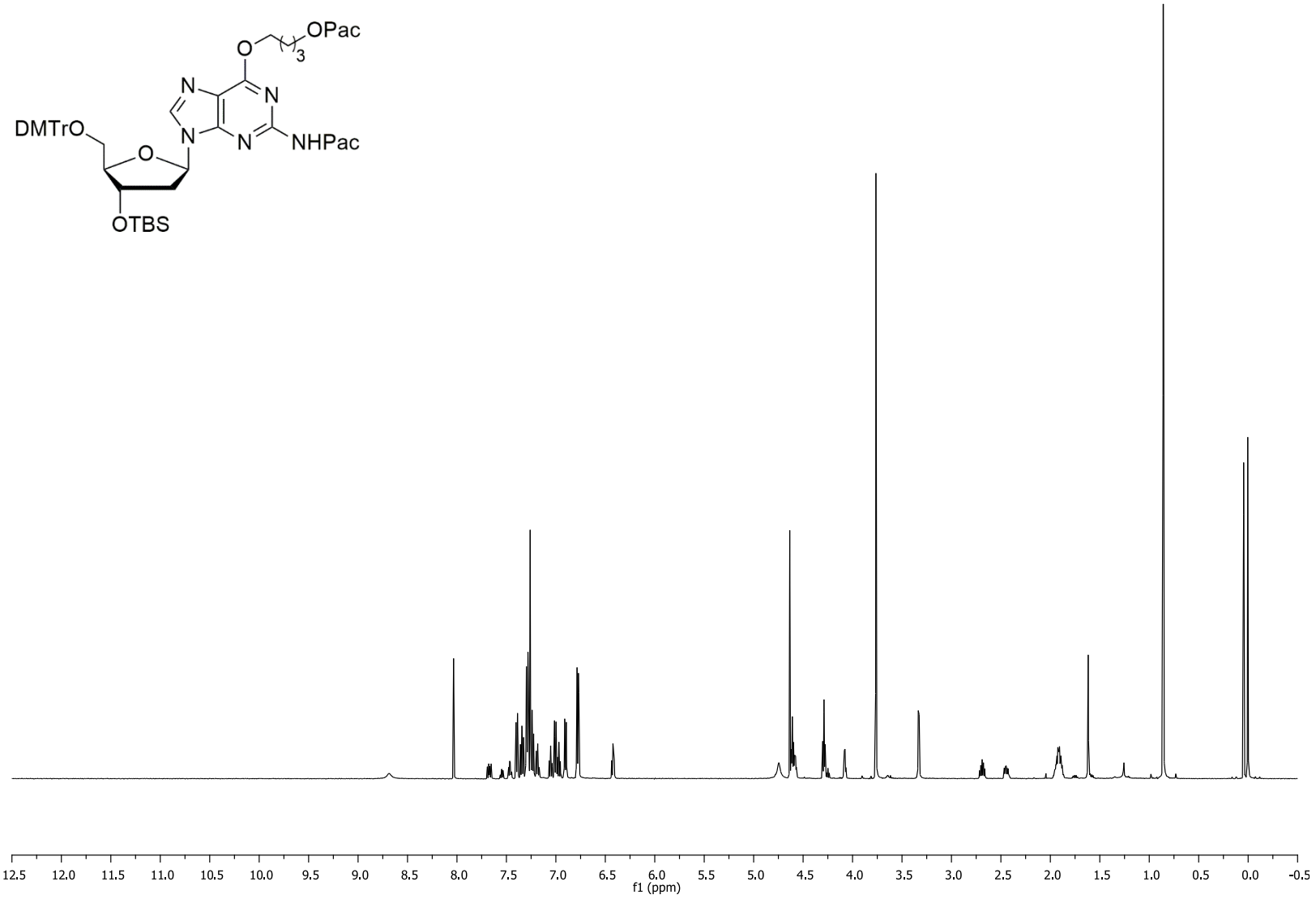


Figure S4.9 125.7 MHz ^{13}C NMR spectrum of compound **4.1b** (in $CDCl_3$)

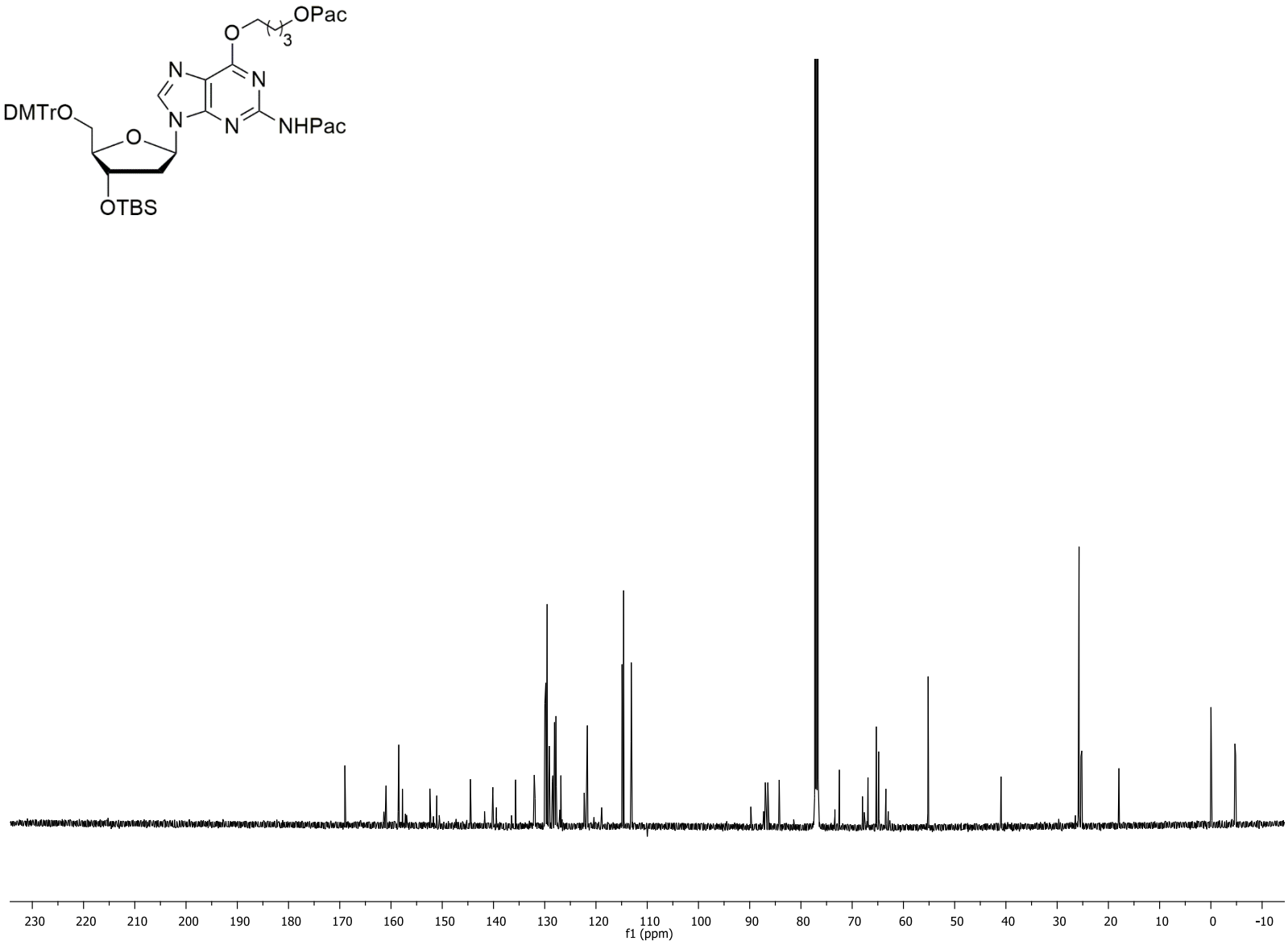


Figure S4.10 500 MHz ^{1}H NMR spectrum of compound **4.1c** (in $CDCl_3$)

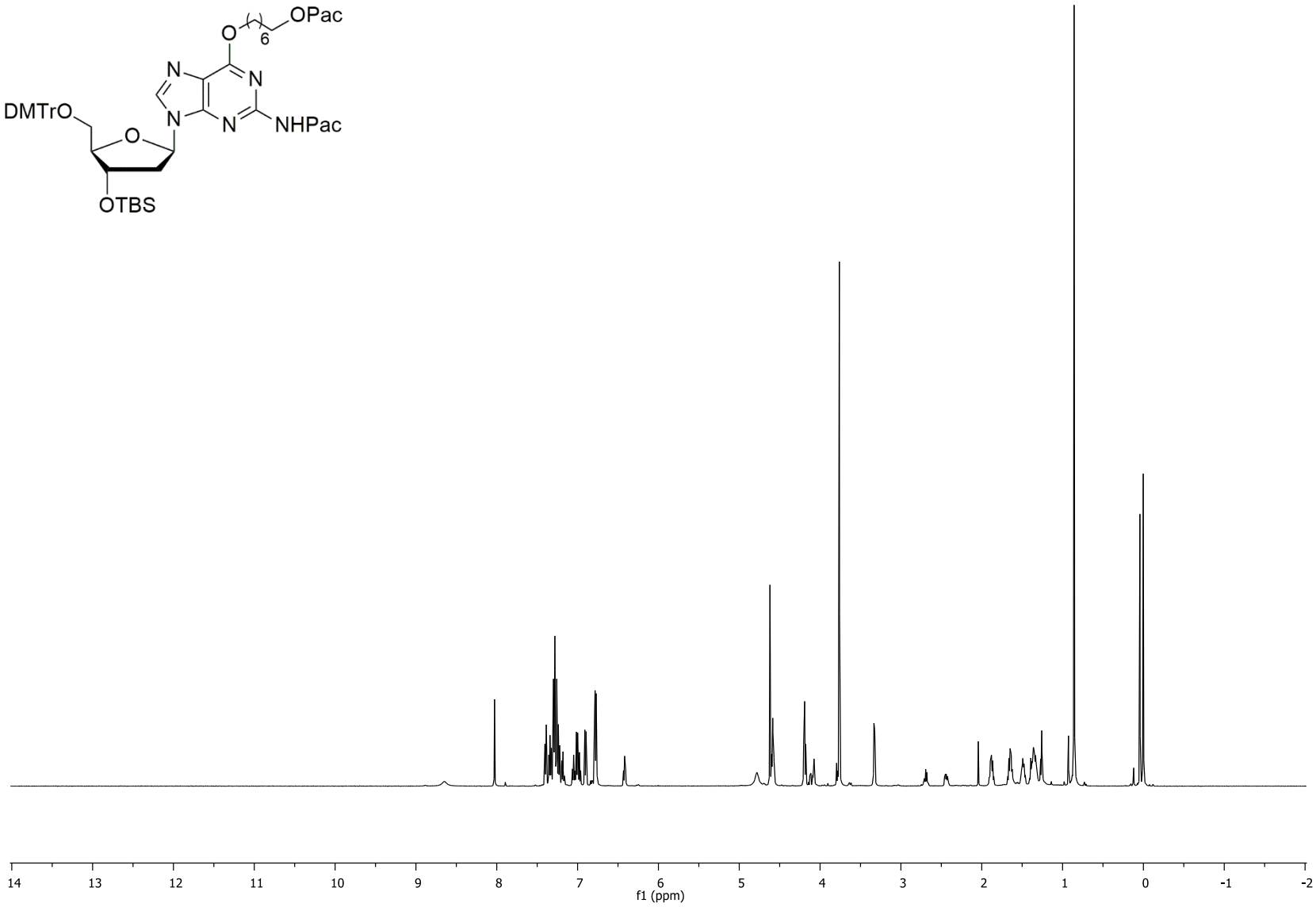


Figure S4.11 125.7 MHz ^{13}C NMR spectrum of compound **4.1c** (in $CDCl_3$)

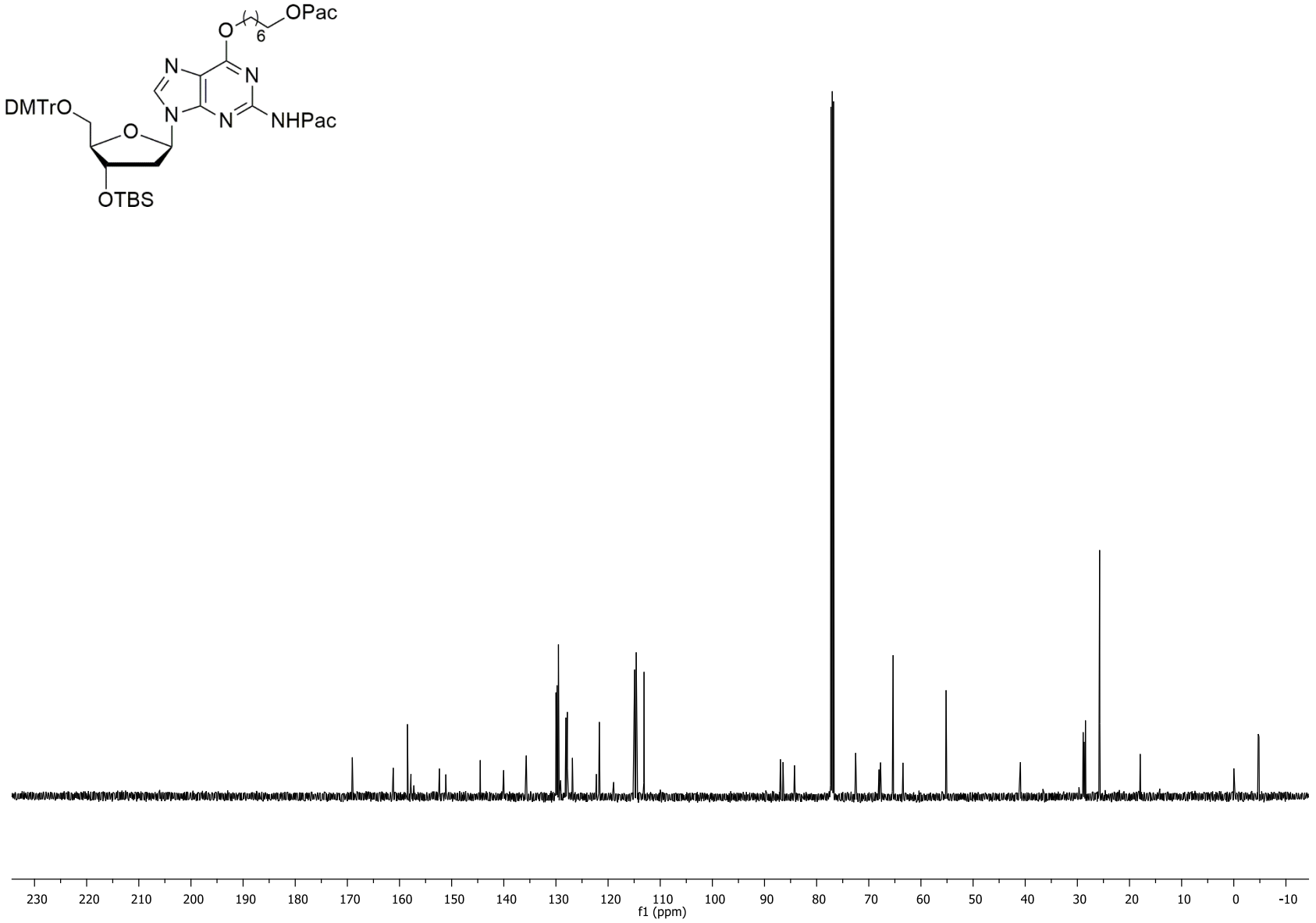


Figure S4.12 500 MHz ^{1}H NMR spectrum of compound **4.2b** (in $CDCl_3$)

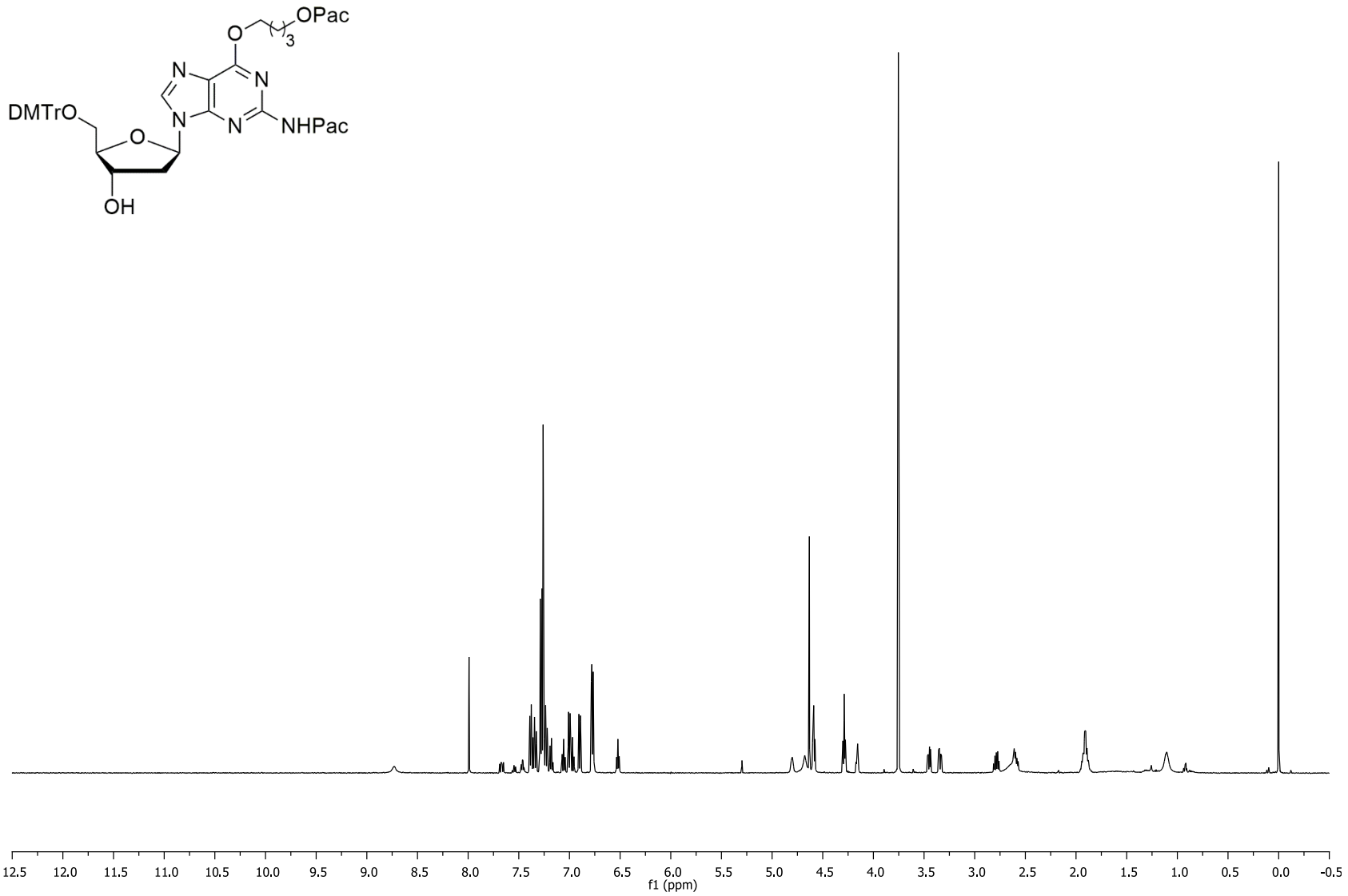
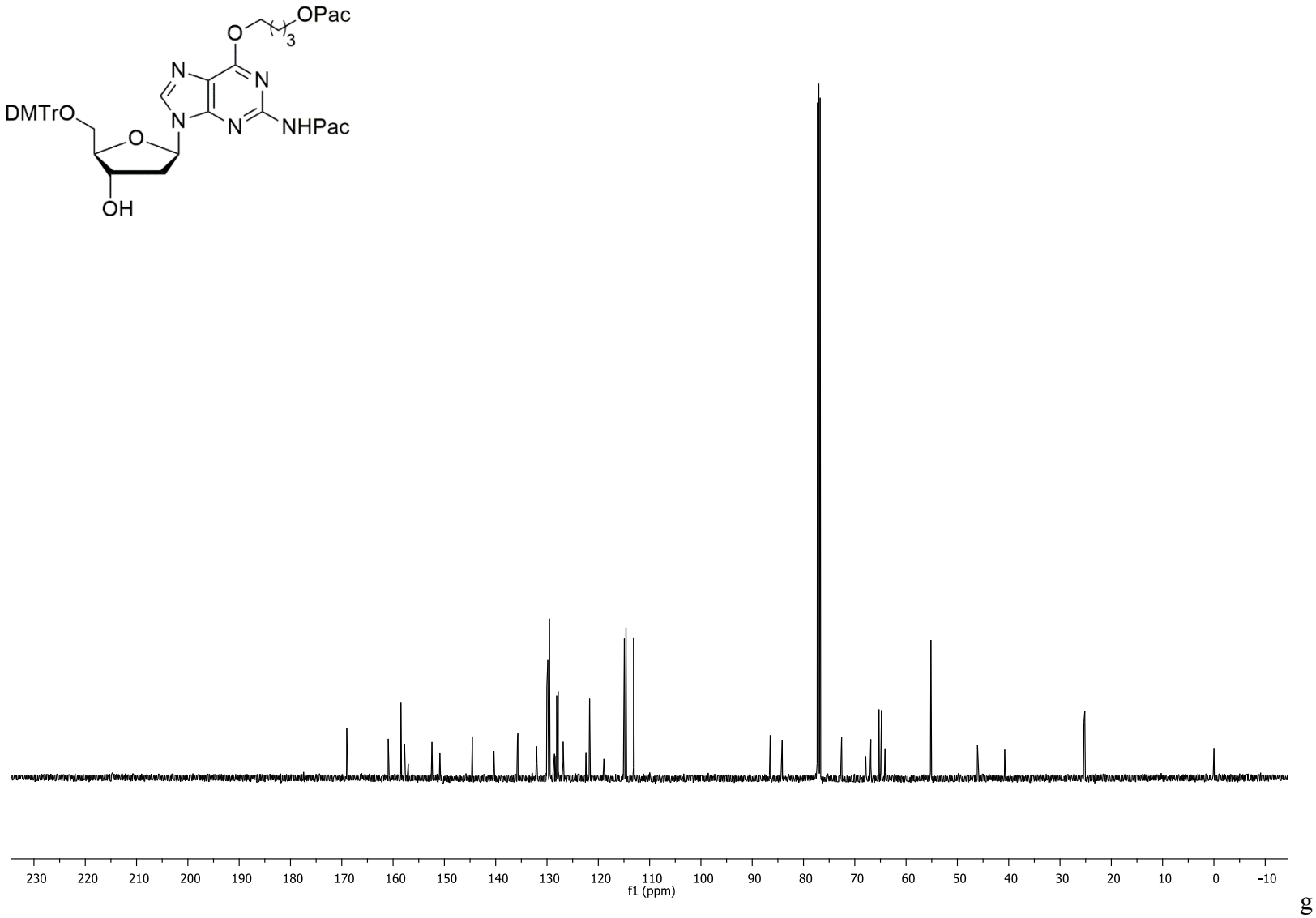


Figure S4.13 125.7 MHz ^{13}C NMR spectrum of compound **4.2b** (in $CDCl_3$)



g

Figure S4.14 500 MHz ^{1}H NMR spectrum of compound **4.2c** (in $CDCl_3$)

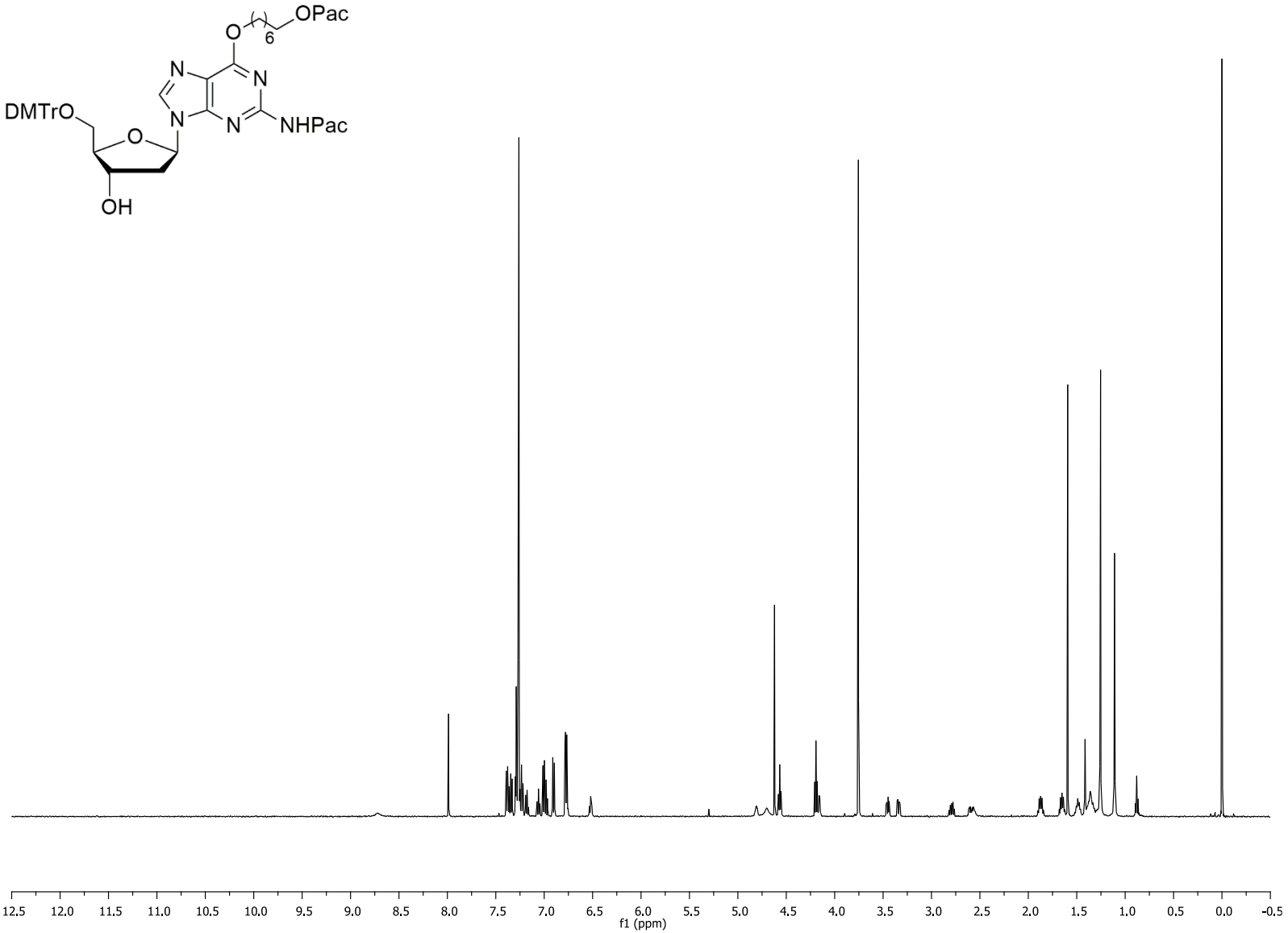


Figure S4.15 125.7 MHz ^{13}C NMR spectrum of compound **4.2c** (in $CDCl_3$)

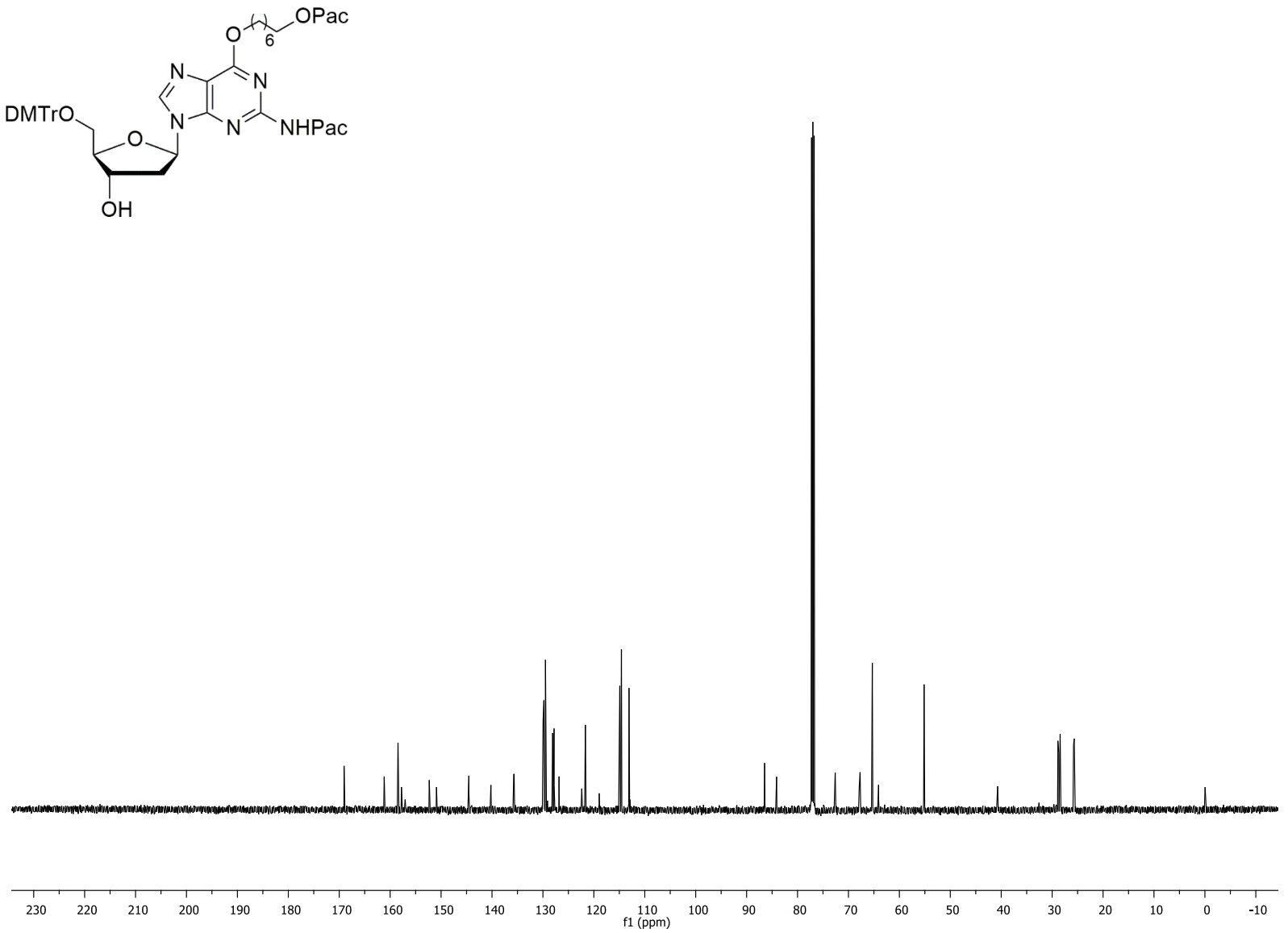


Figure S4.16 500 MHz ^{1}H NMR spectrum of compound **4.3b** (in d_6-acetone)

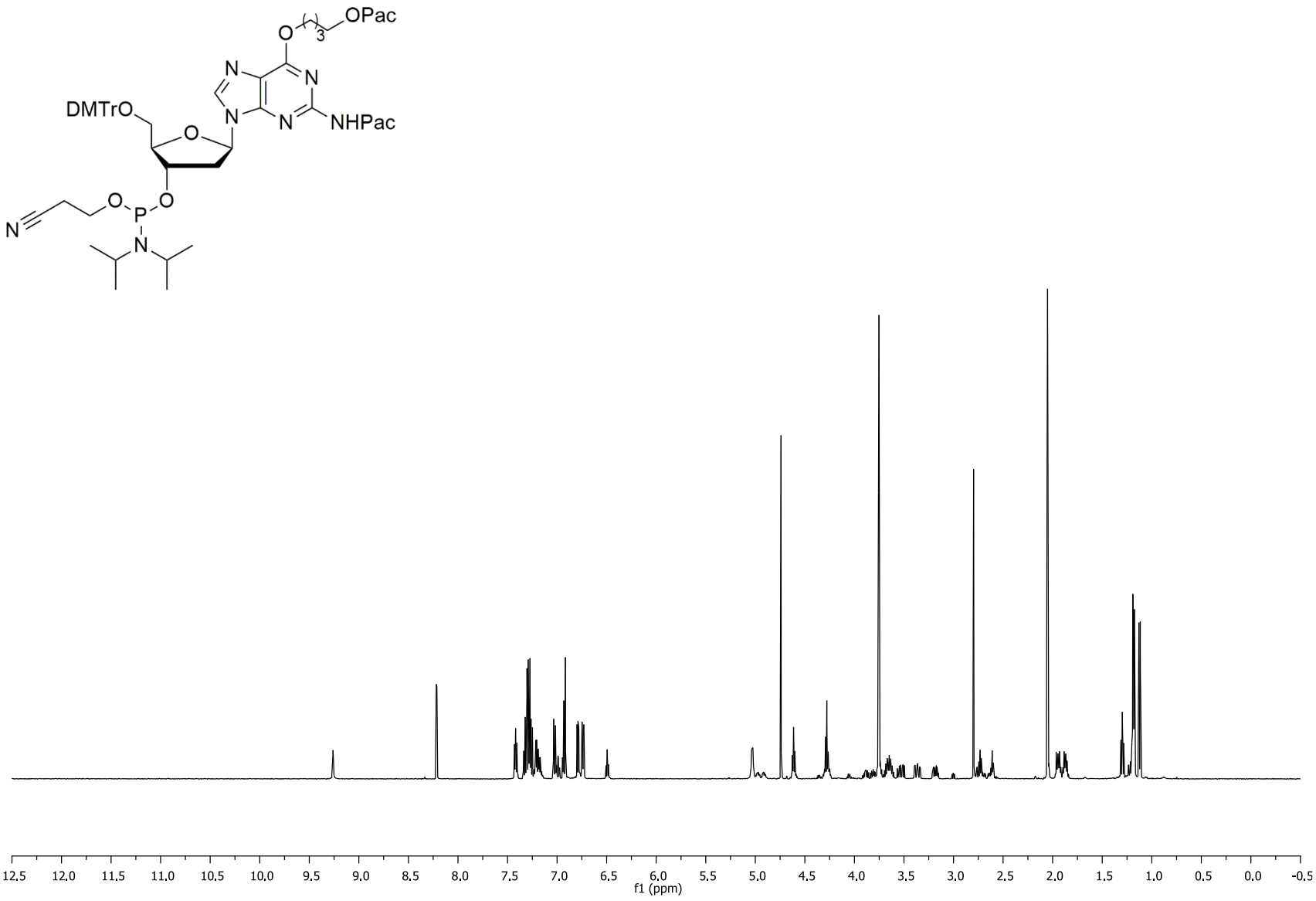


Figure S4.17 125.7 MHz ^{13}C NMR spectrum of compound **4.3b** (in d_6-acetone)

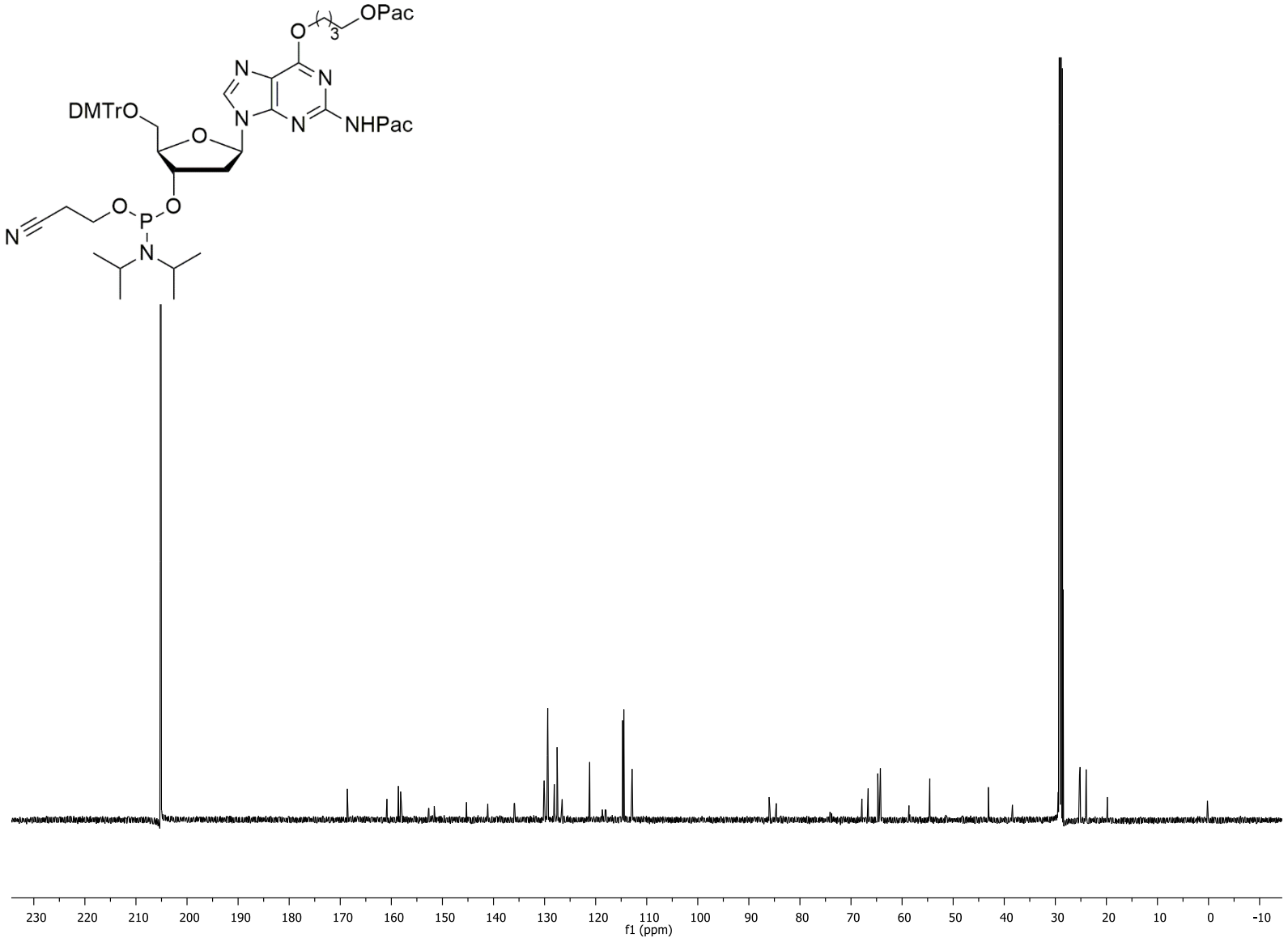


Figure S4.18 202.3 MHz ^{31}P NMR spectrum of compound **4.3b** (in d_6-acetone)

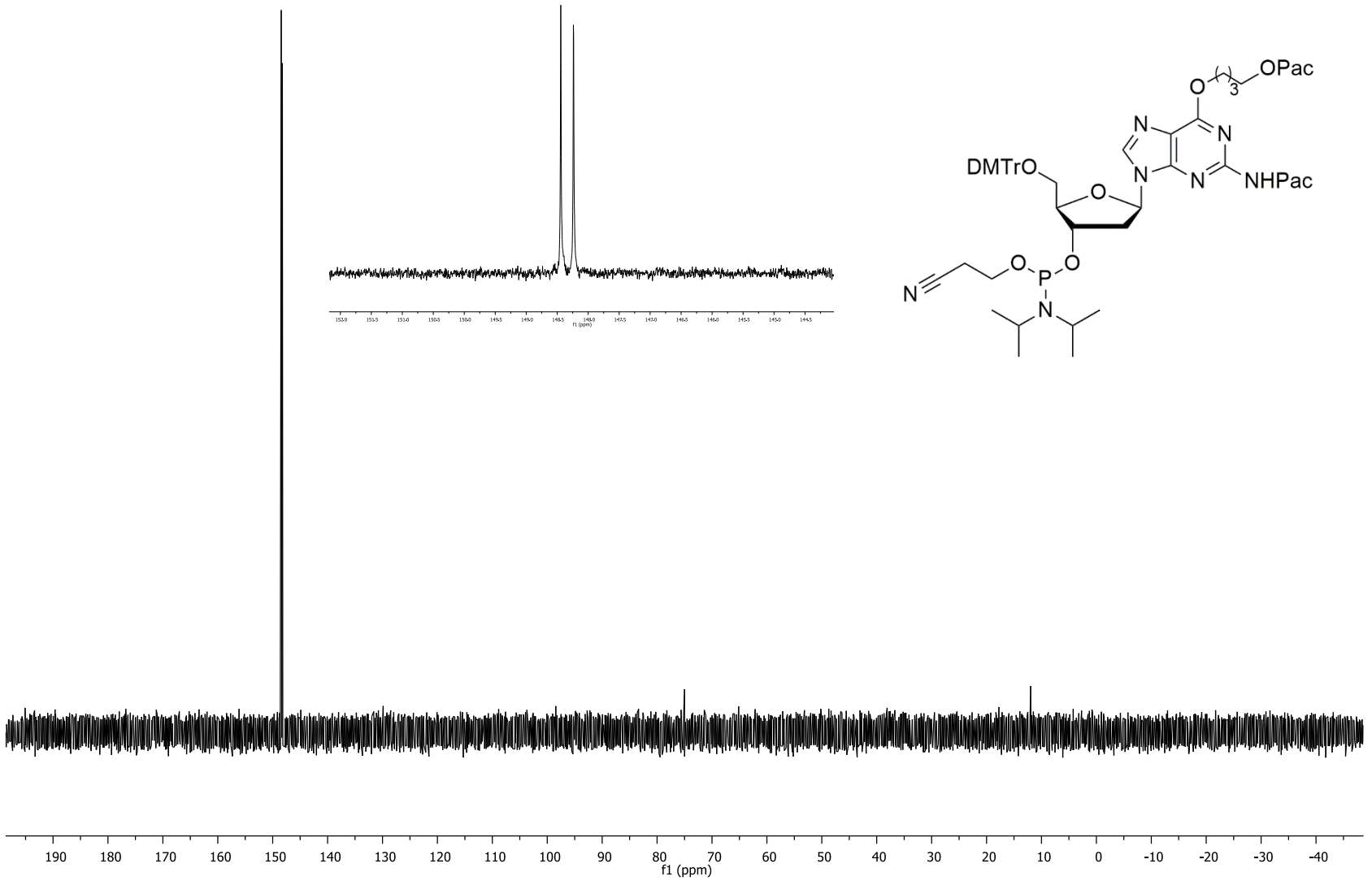


Figure S4.19 500 MHz ^{1}H NMR spectrum of compound **4.3c** (in d_6-acetone)

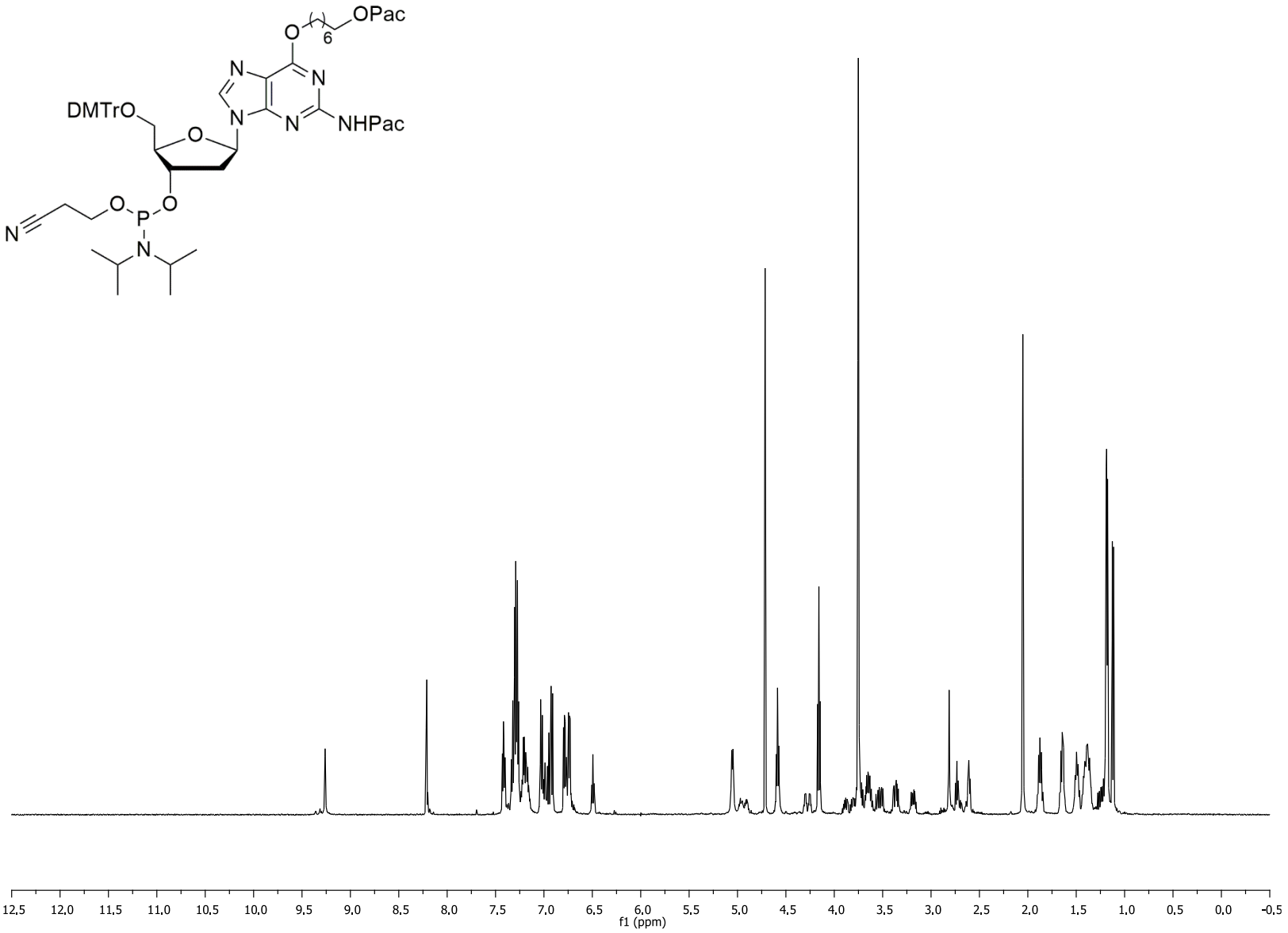


Figure S4.20 125.7 MHz ^{13}C NMR spectrum of compound **4.3c** (in d_6-acetone)

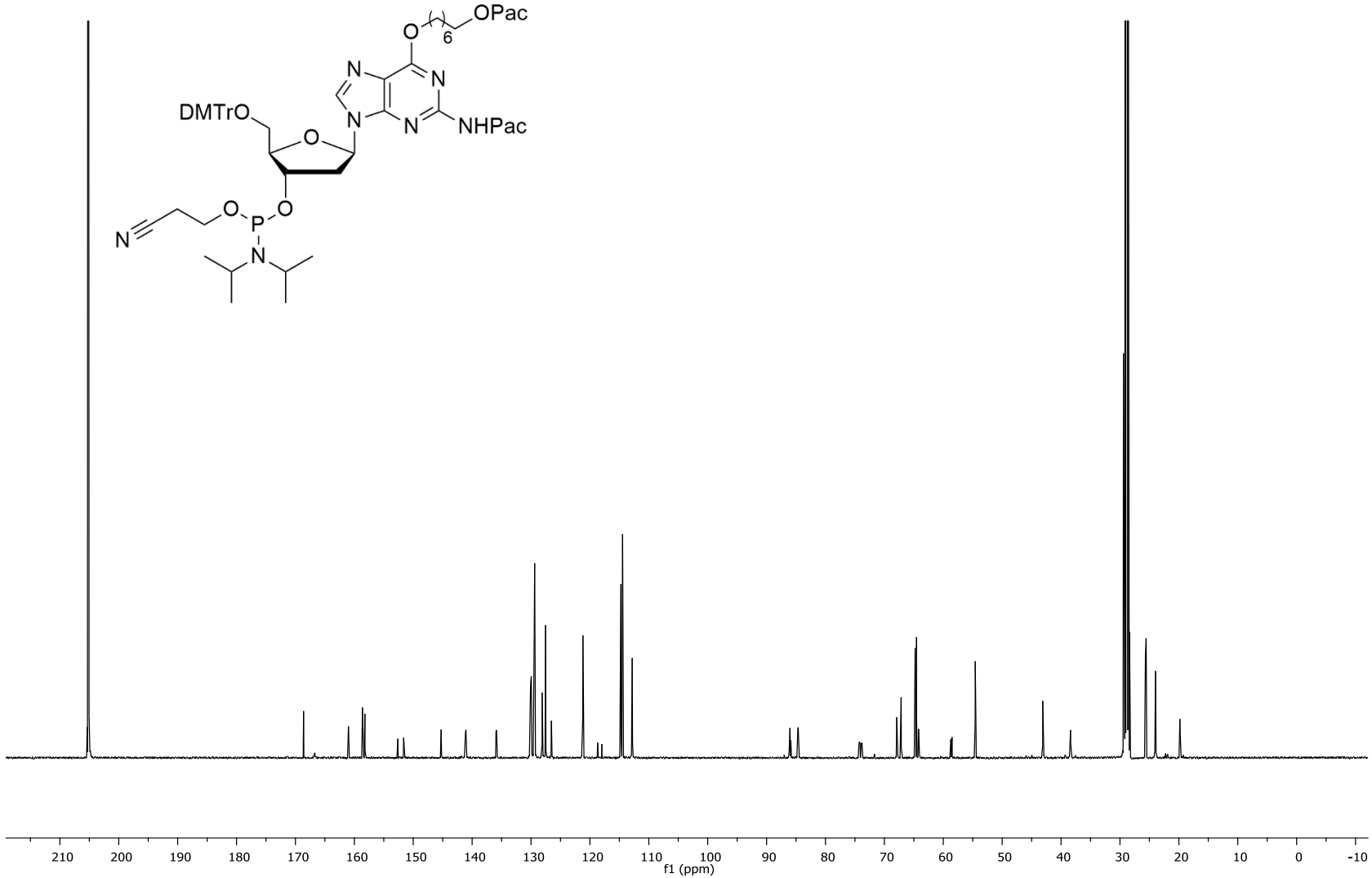


Figure S4.21 202.3 MHz ^{31}P NMR spectrum of compound **4.3c** (in d_6-acetone)

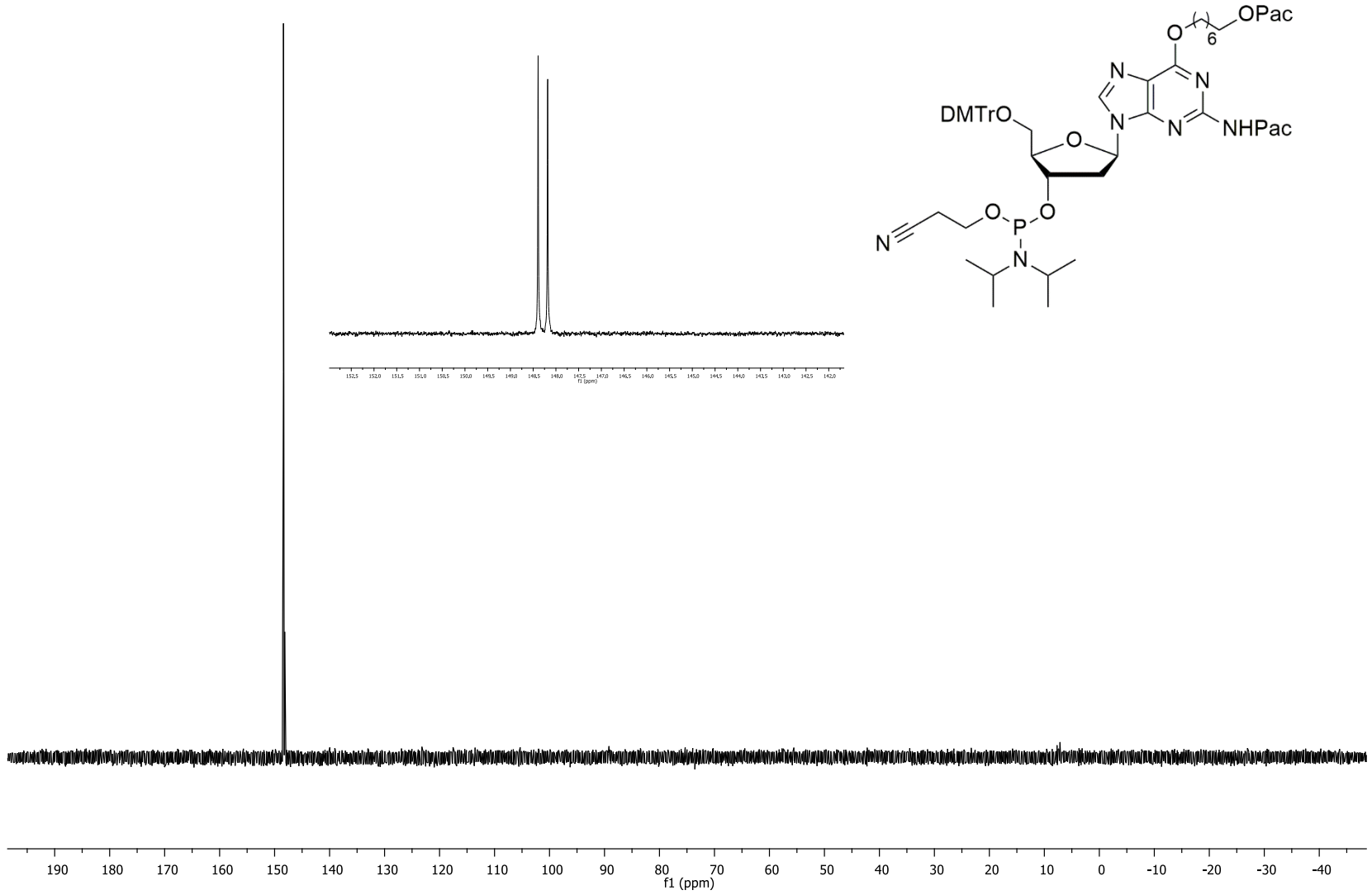


Figure S4.22 500 MHz ^{1}H NMR spectrum of compound **4.5** (in $CDCl_3$)

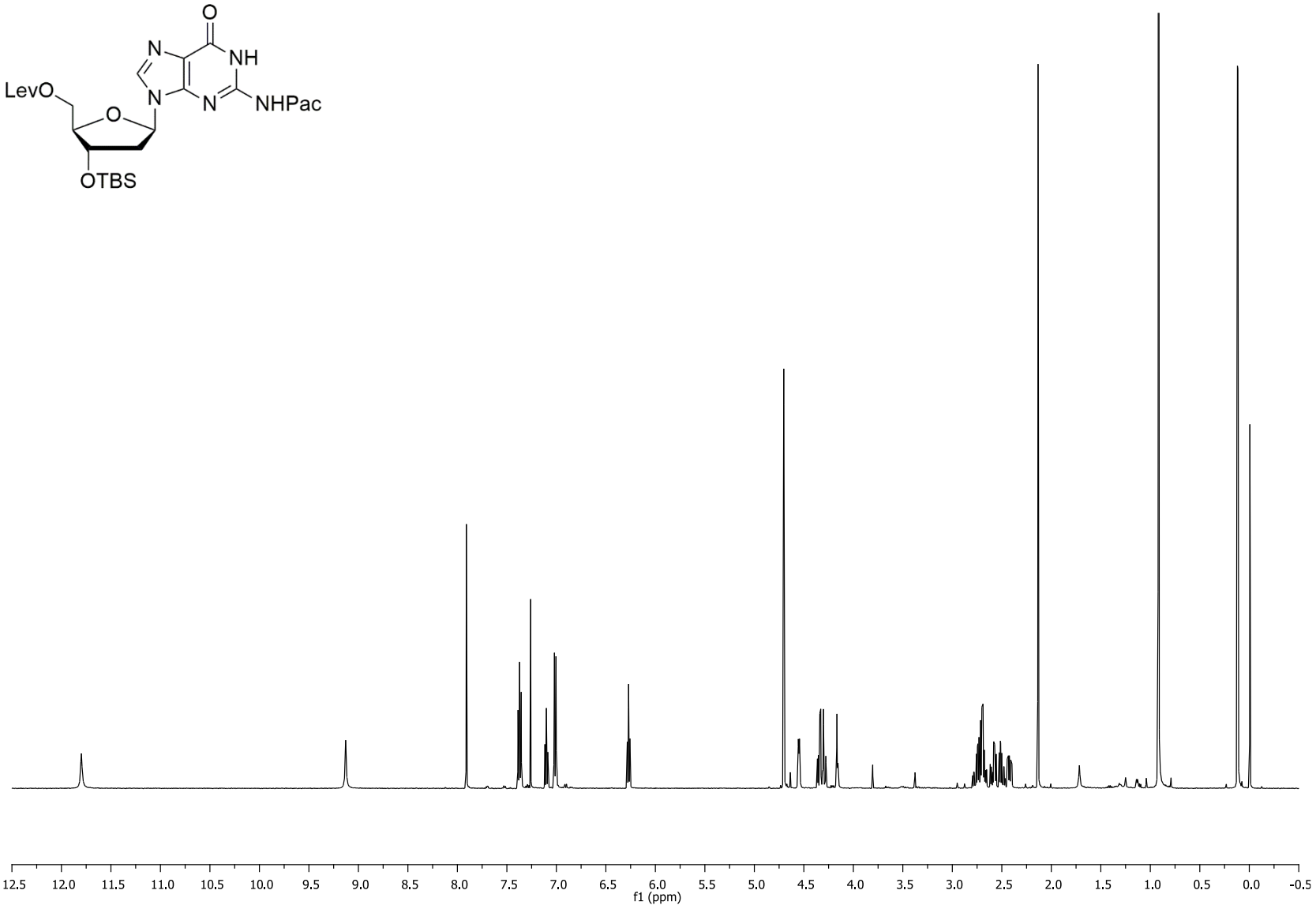


Figure S4.23 125.7 MHz ^{13}C NMR spectrum of compound **4.5** (in $CDCl_3$)

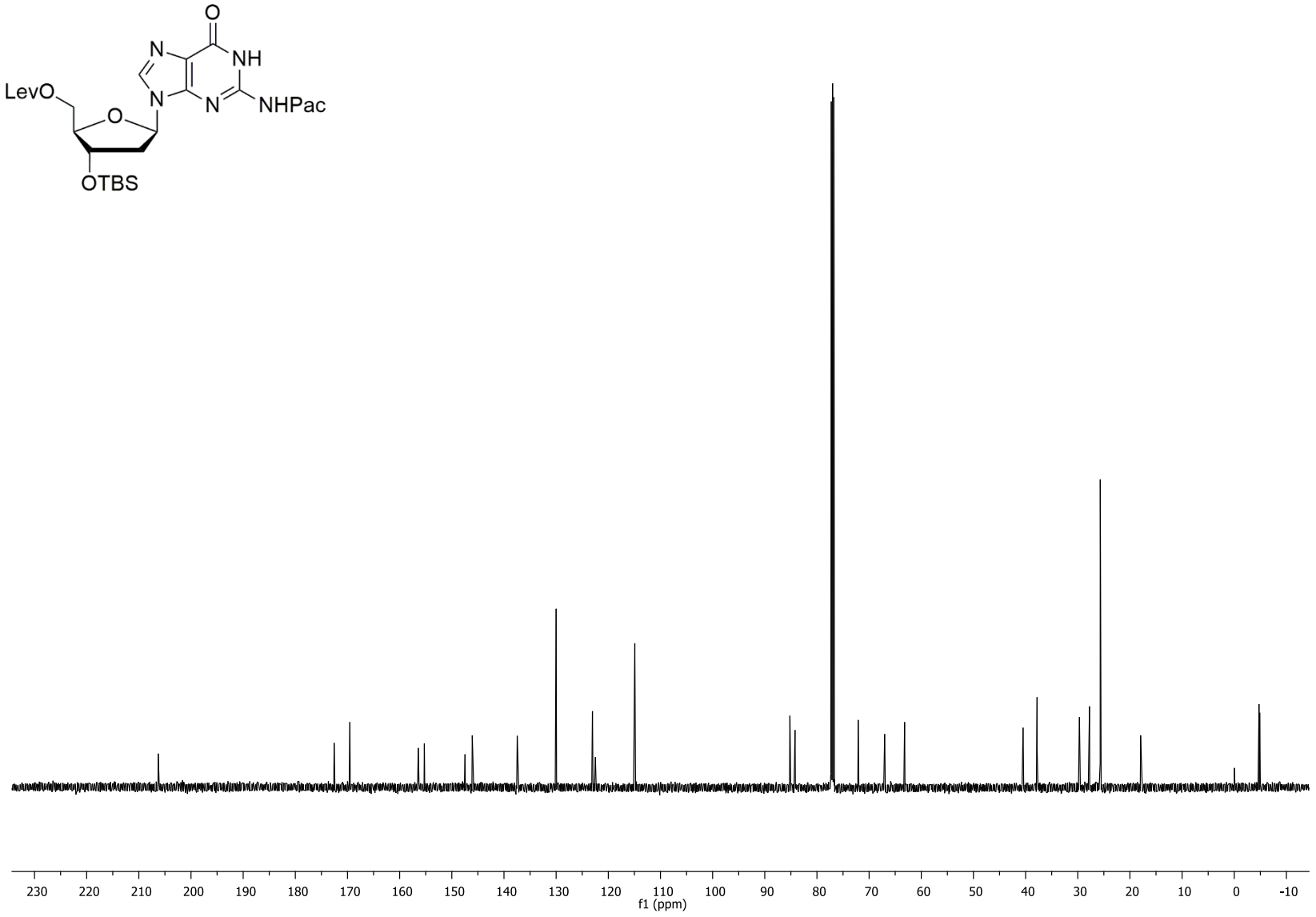


Figure S4.24 500 MHz ^{1}H NMR spectrum of compound **4.6a** (in $CDCl_3$)

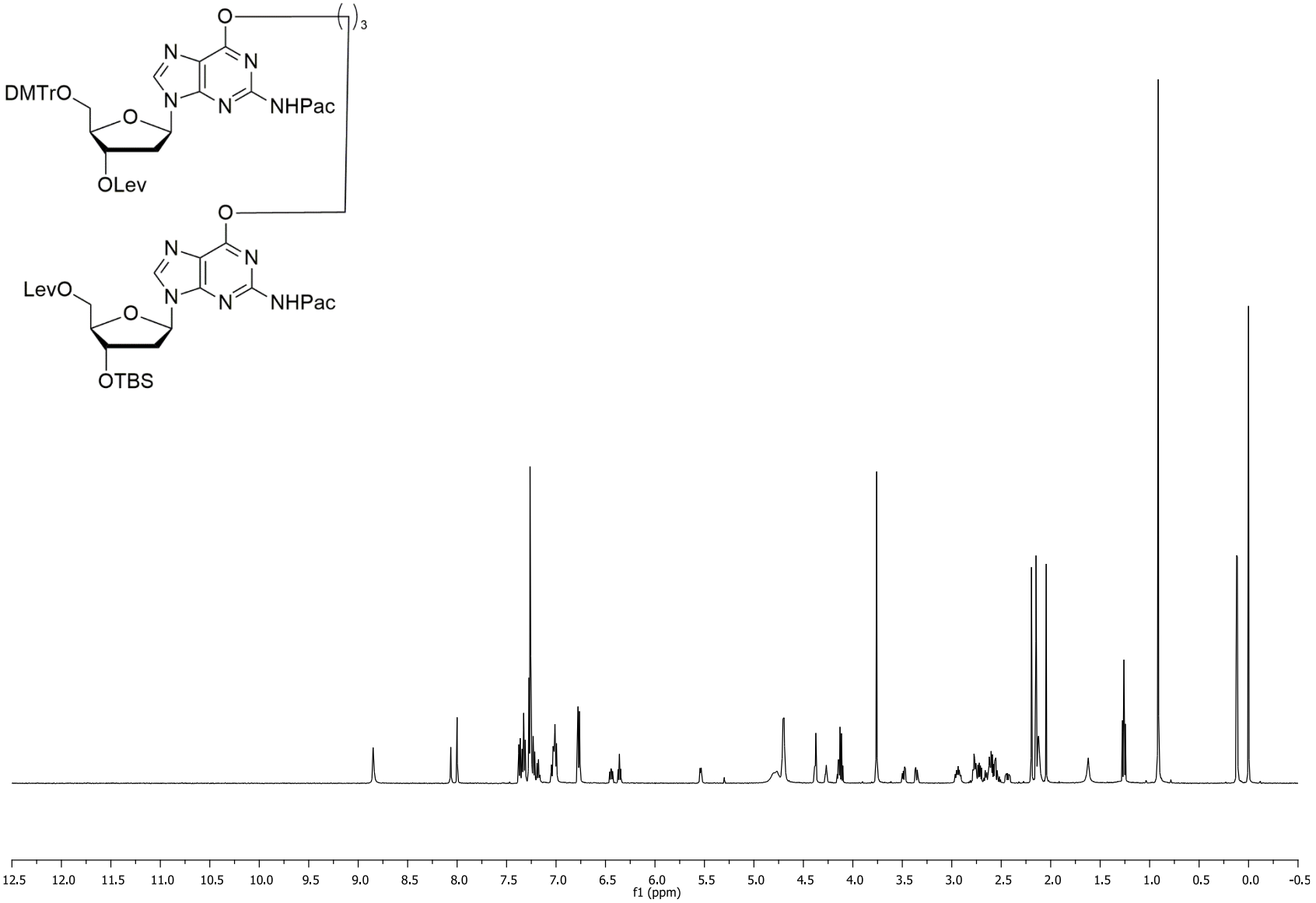


Figure S4.25 125.7 MHz ^{13}C NMR spectrum of compound **4.6a** (in $CDCl_3$)

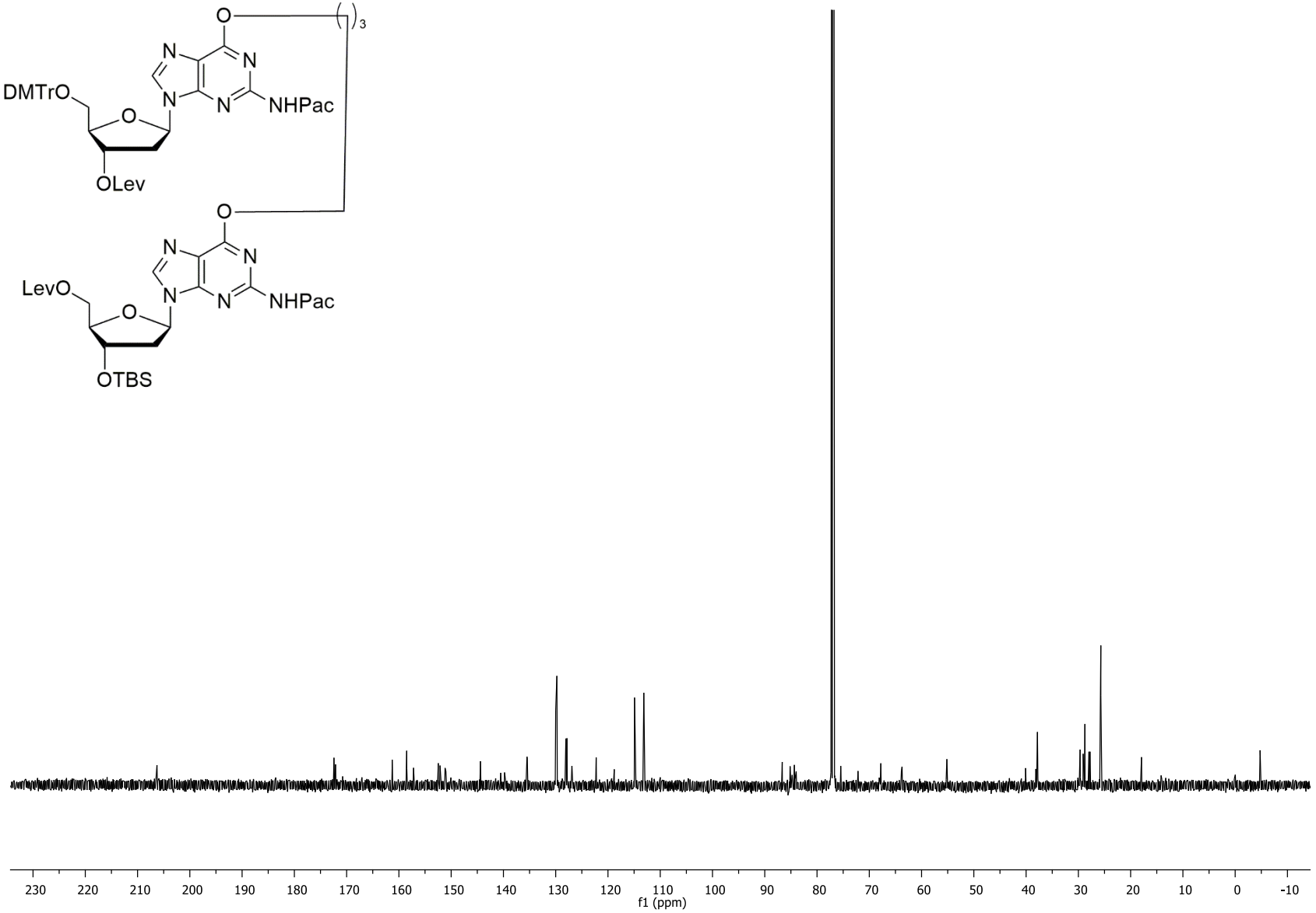


Figure S4.26 500 MHz ^{1}H NMR spectrum of compound **4.6b** (in $CDCl_3$)

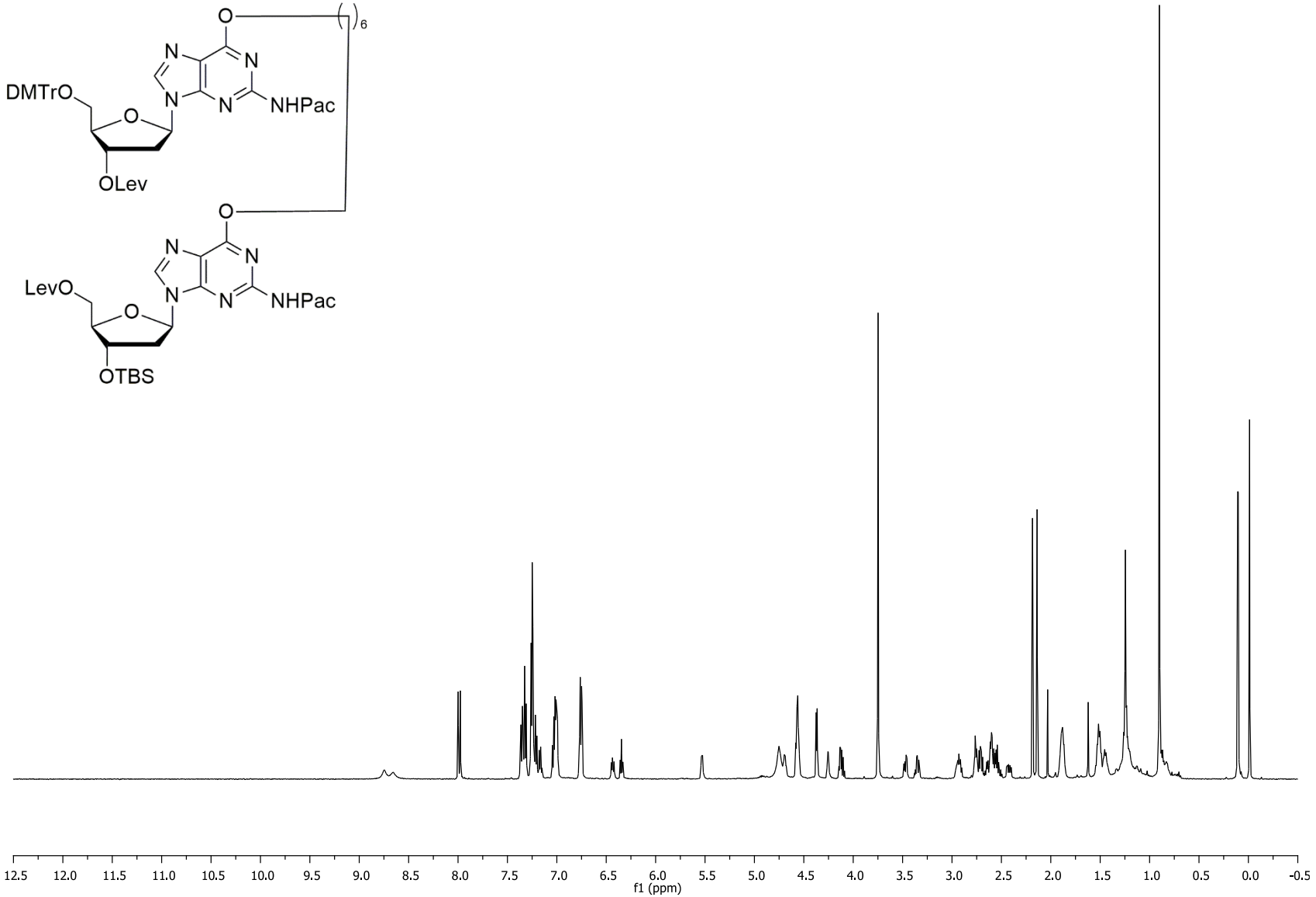


Figure S4.27 125.7 MHz ^{13}C NMR spectrum of compound **4.6b** (in $CDCl_3$)

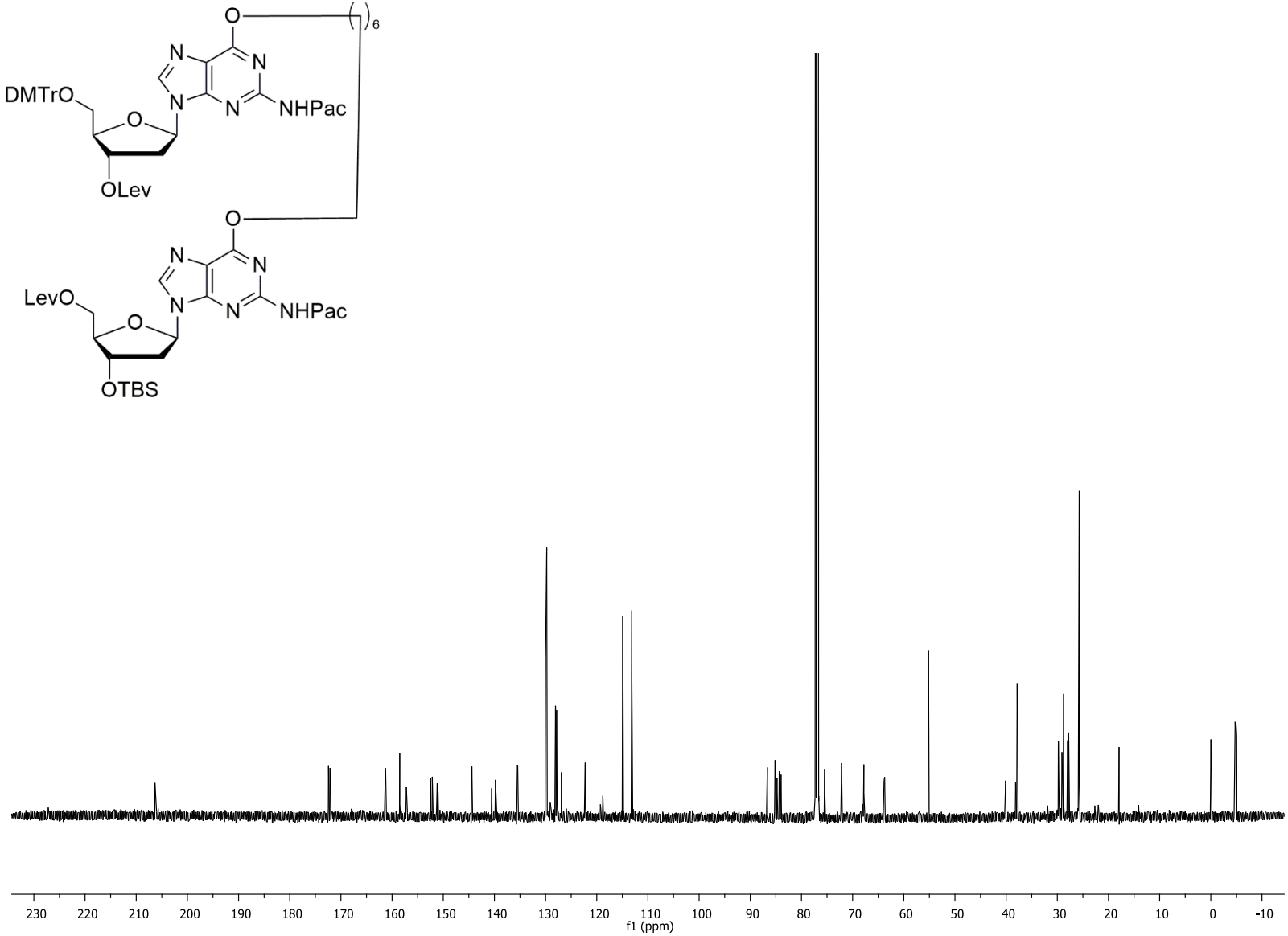


Figure S4.28 500 MHz ^{1}H NMR spectrum of compound **4.7a** (in $CDCl_3$)

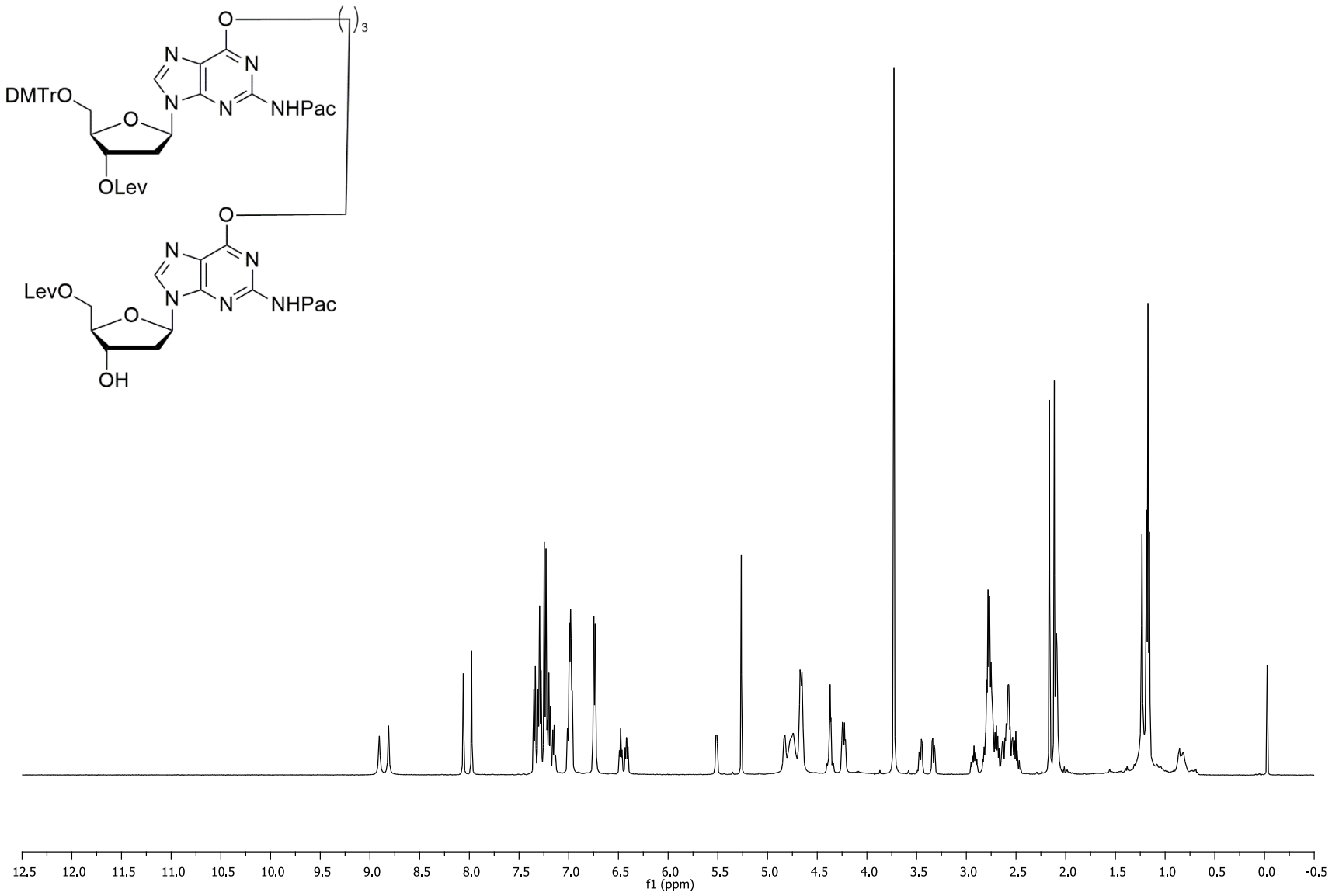


Figure S4.29 125.7 MHz ^{13}C NMR spectrum of compound **4.7a** (in $CDCl_3$)

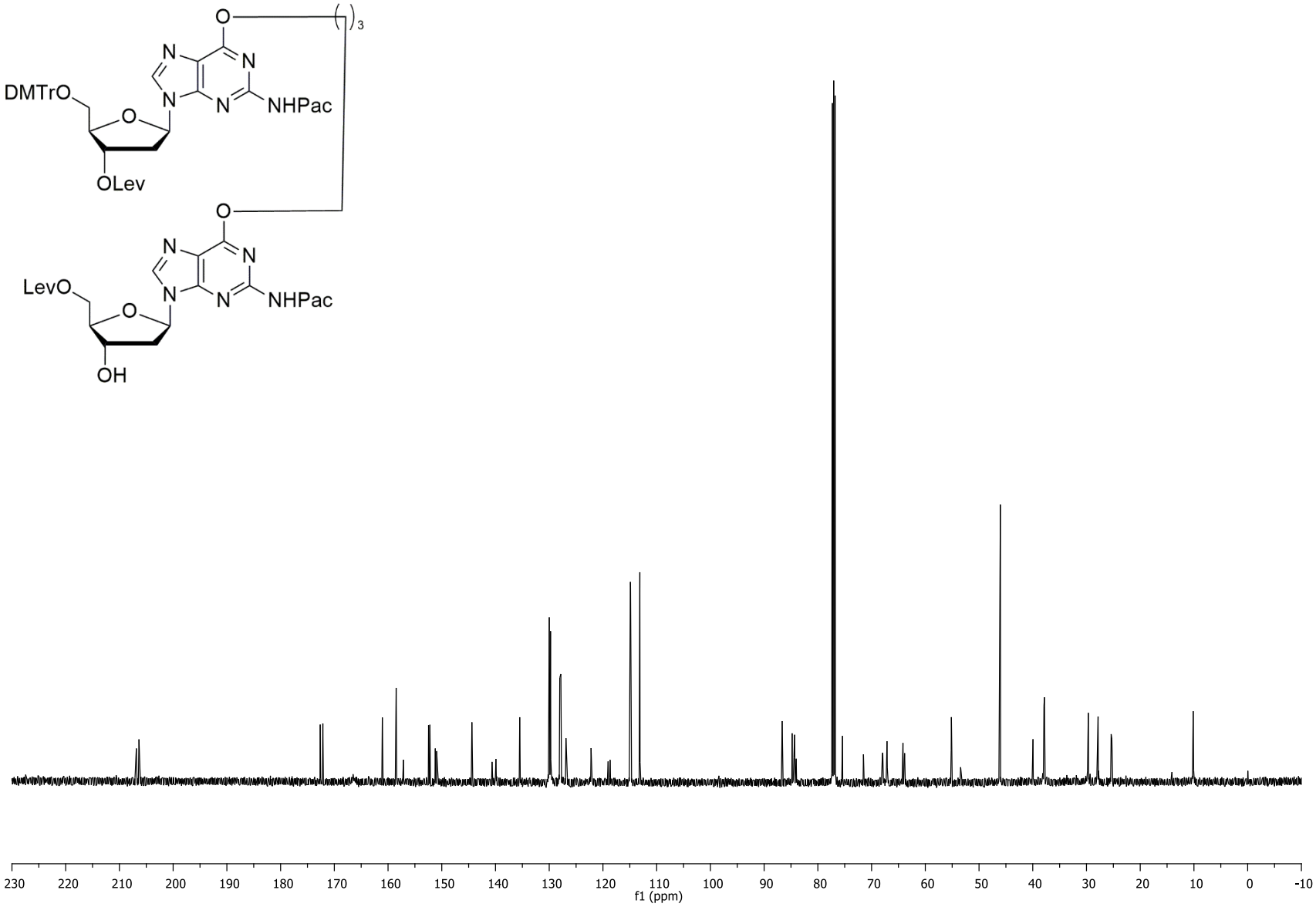


Figure S4.30 500 MHz ^{1}H NMR spectrum of compound **4.7b** (in $CDCl_3$)

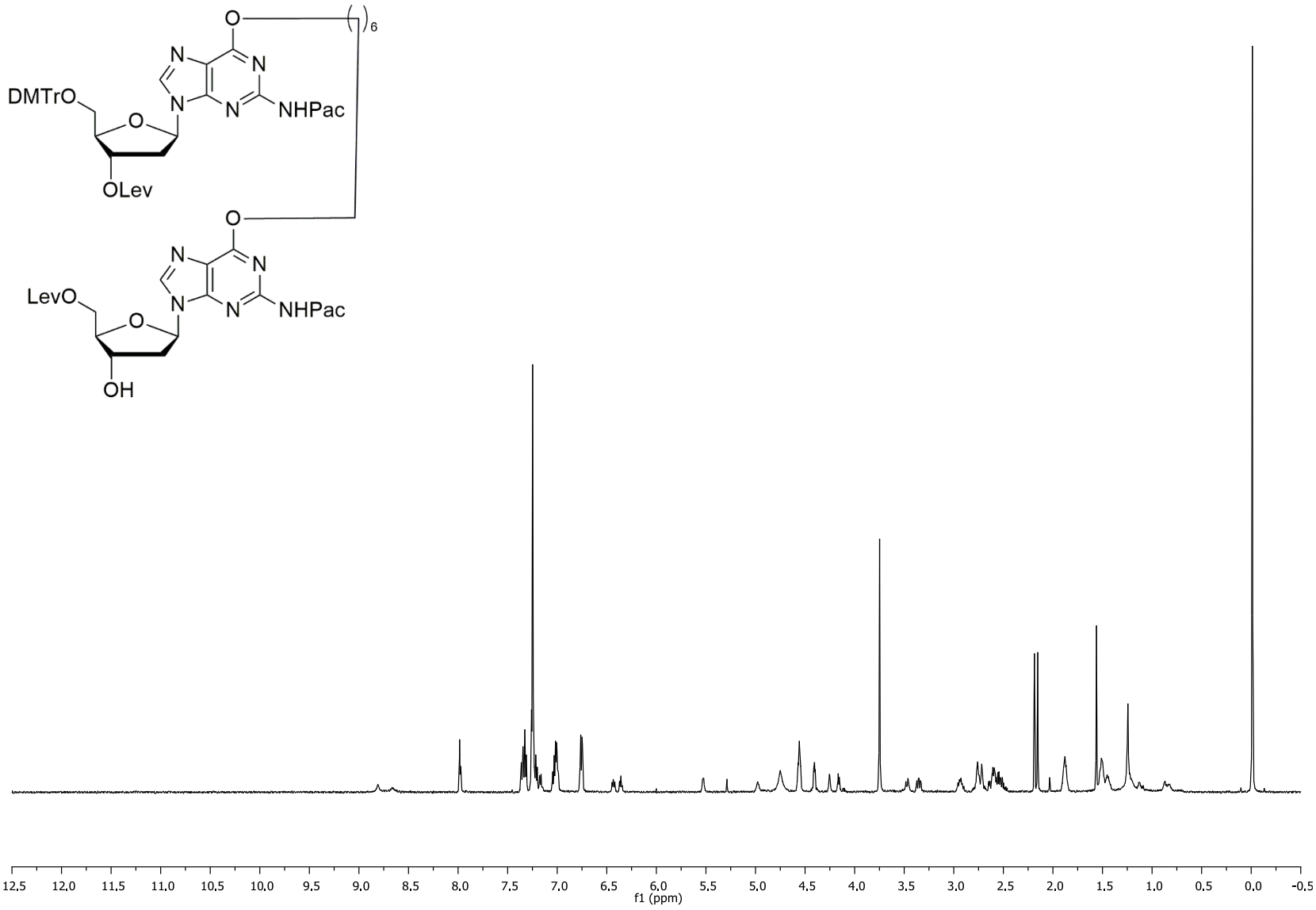


Figure S4.31 125.7 MHz ^{13}C NMR spectrum of compound **4.7b** (in d_6-$CDCl_3$)

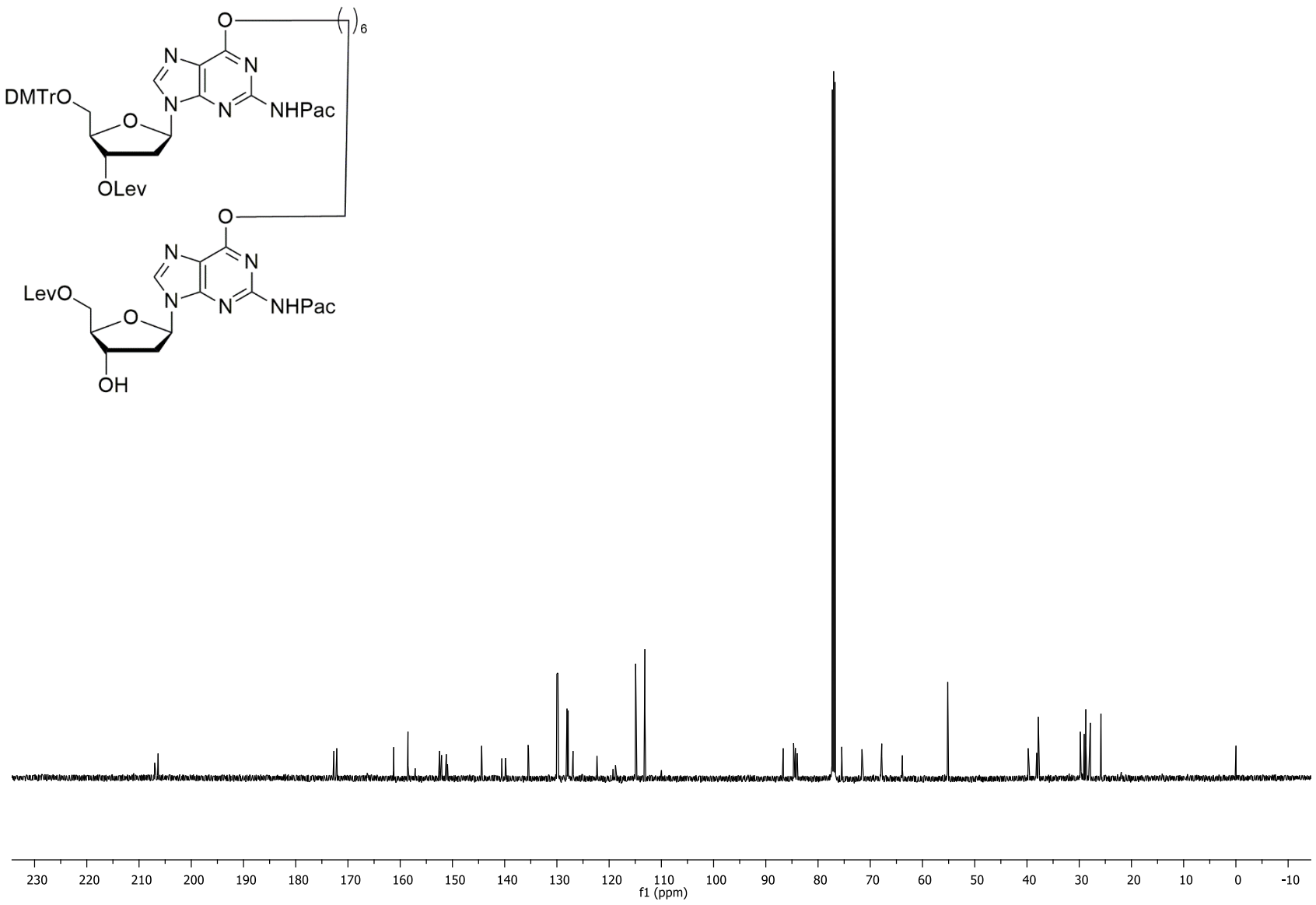


Figure S4.32 500 MHz ^{1}H NMR spectrum of compound **4.8a** (in d_6-acetone)

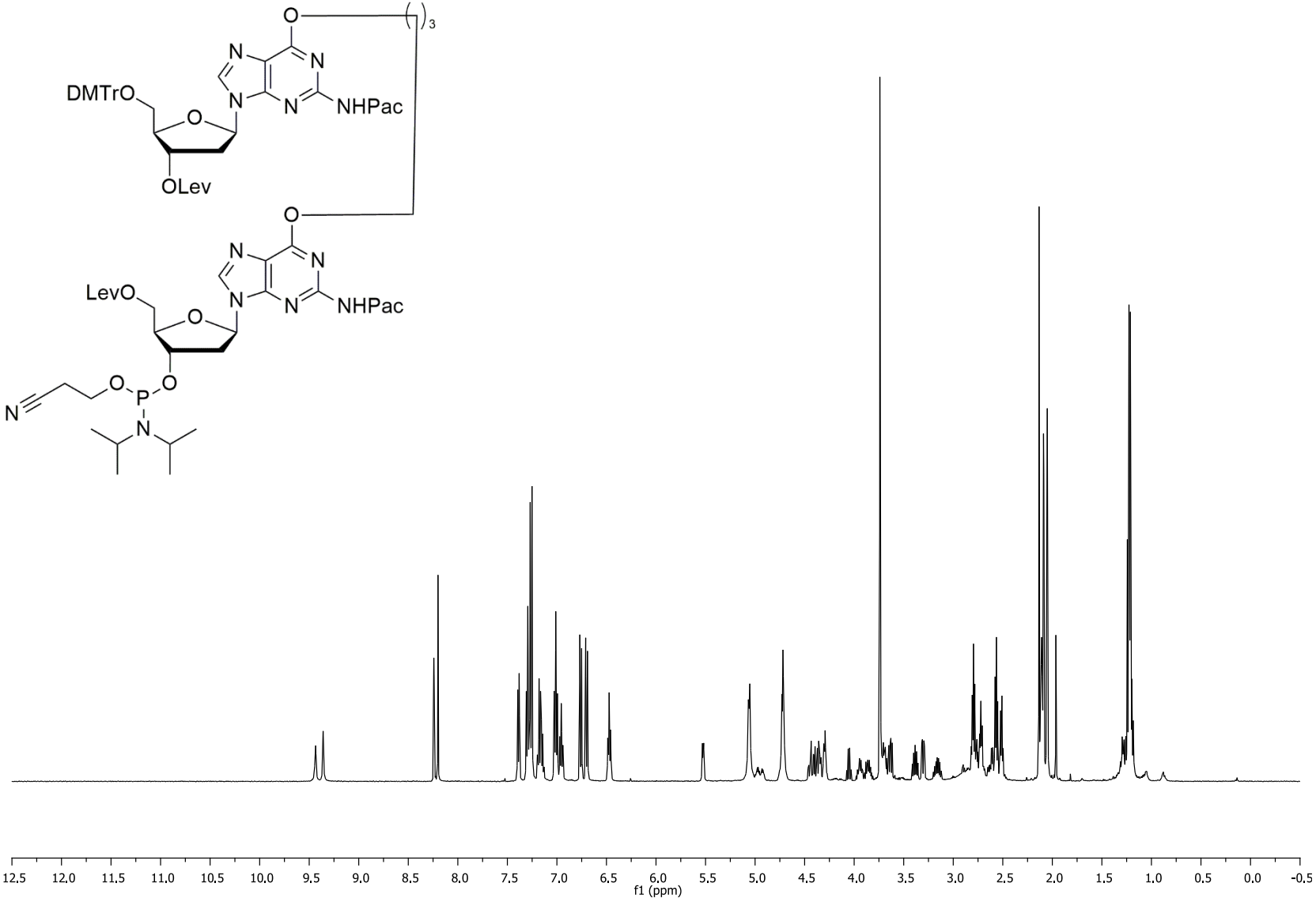


Figure S4.33 125.7 MHz ^{13}C NMR spectrum of compound **4.8a** (in d_6-acetone)

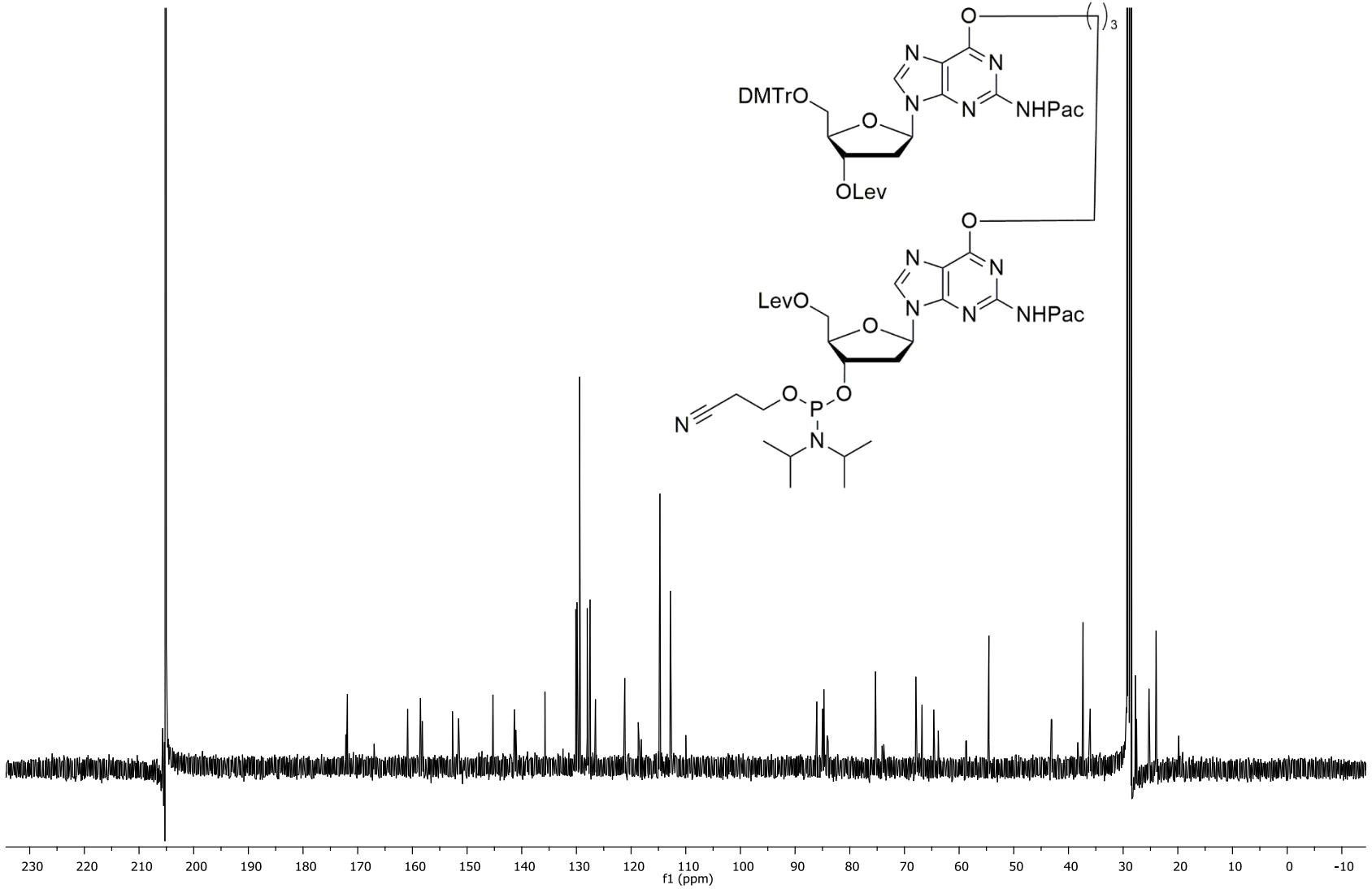


Figure S4.34 202.3 MHz ^{31}P NMR spectrum of compound **4.8a** (in d_6-acetone)

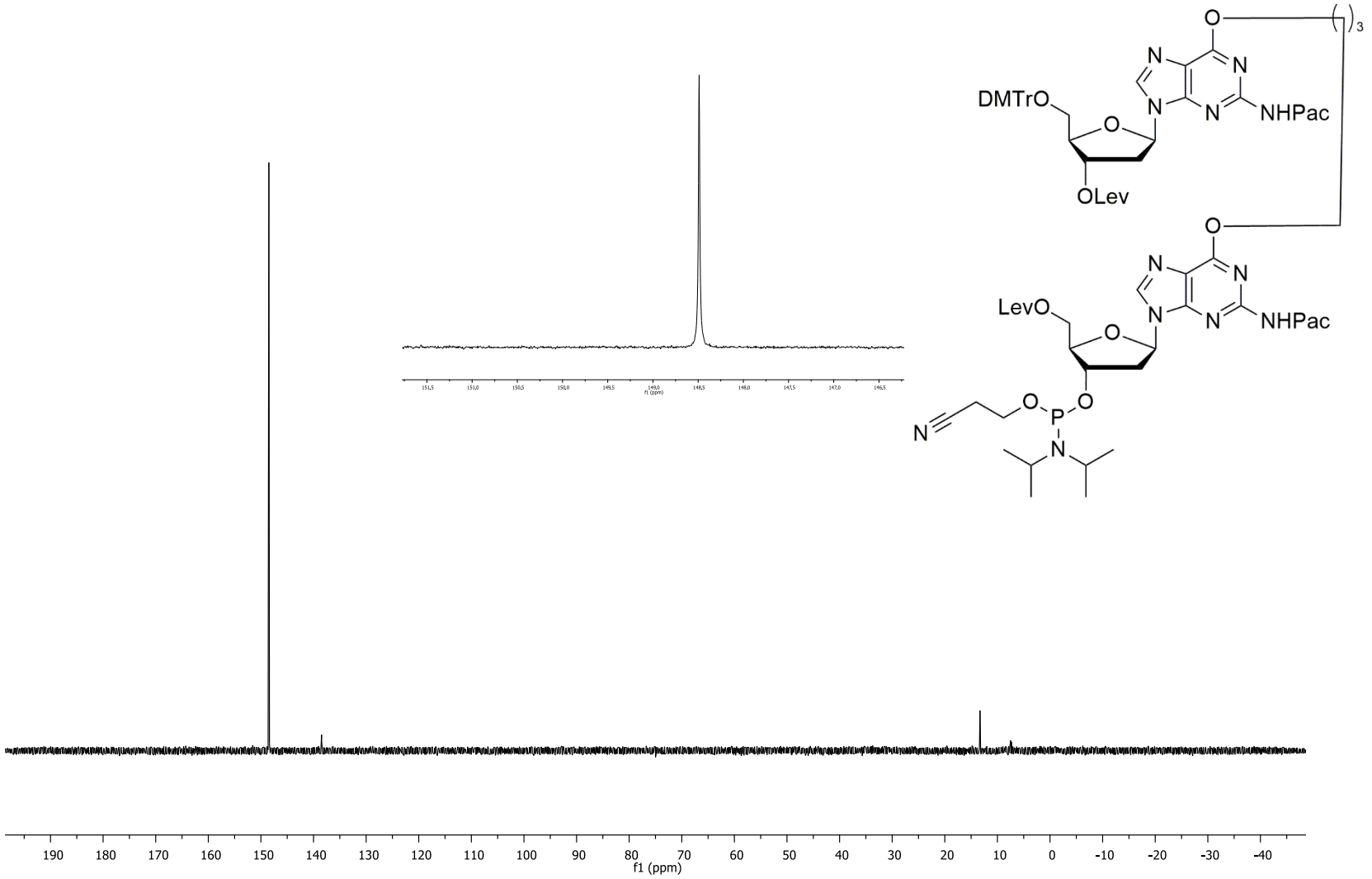


Figure S4.35 500 MHz ^{1}H NMR spectrum of compound **4.8b** (in d_6-acetone)

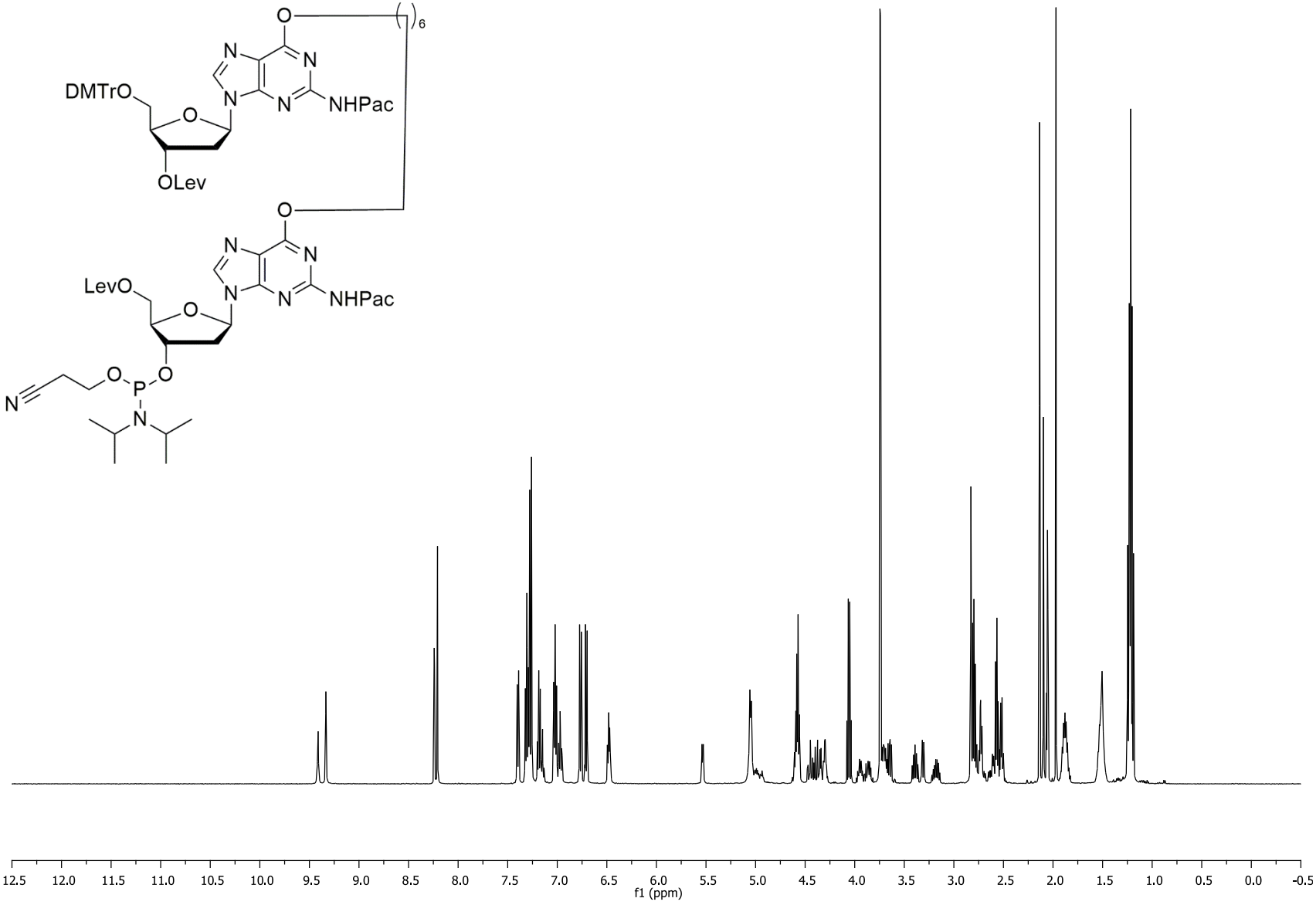


Figure S4.36 125.7 MHz ^{13}C NMR spectrum of compound **4.8b** (in d_6-acetone)

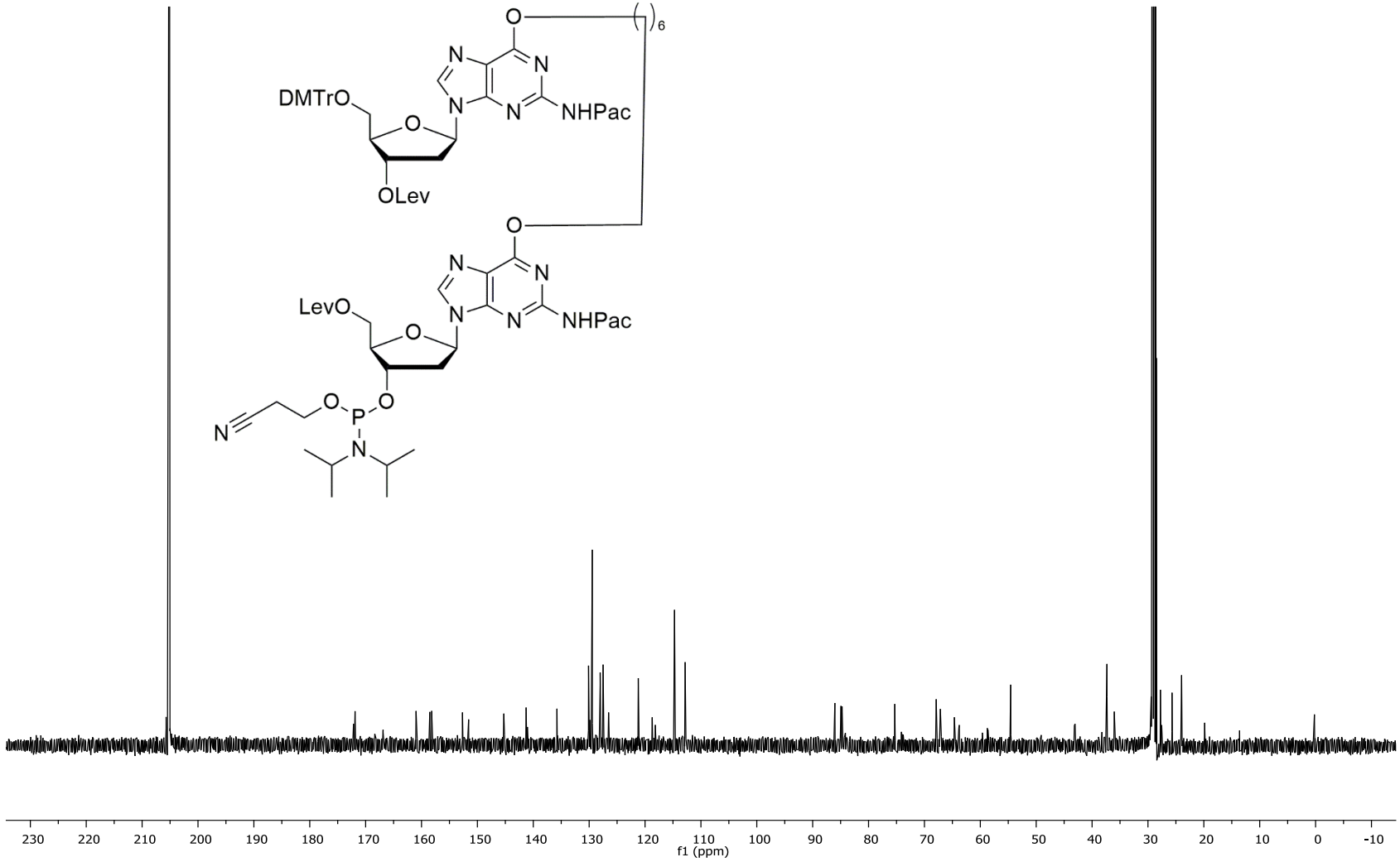


Figure S4.37 202.3 MHz ^{31}P NMR spectrum of compound **4.8b** (in d_6-acetone)

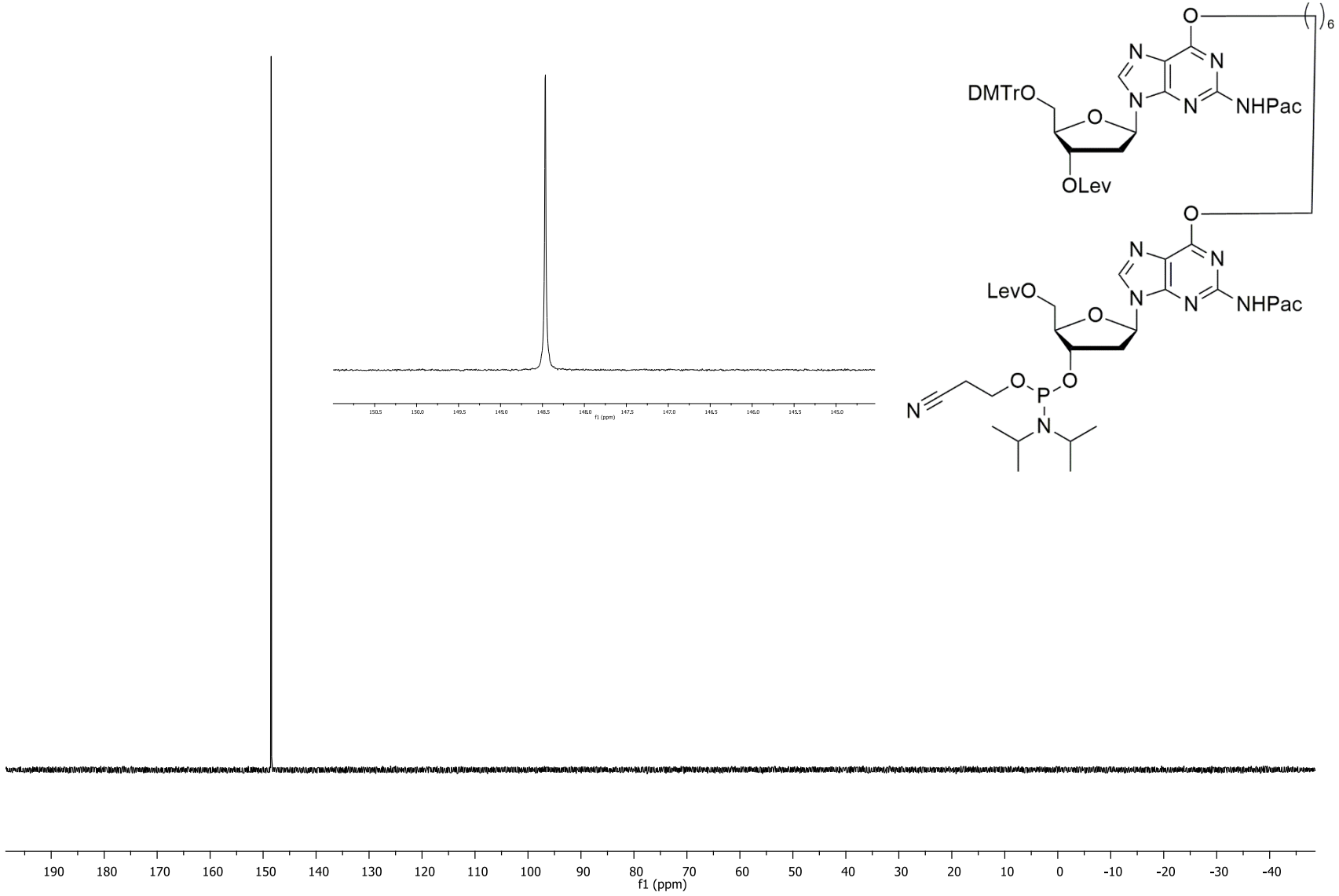


Figure S4.38 UV Thermal denaturation profiles of G-quadruplexes containing two sites of alkylation measured at 295 nm and high [K^+] Buffer: 130 mM KCl, 5 mM KH_2PO_4, 5 mM K_2HPO_4, 5.4 mM KOH, pH 7.2. X= O^6-Methyl dG, Y= O^6-4-hydroxybutyl dG, Z= O^6-7-hydroxyheptyl dG.

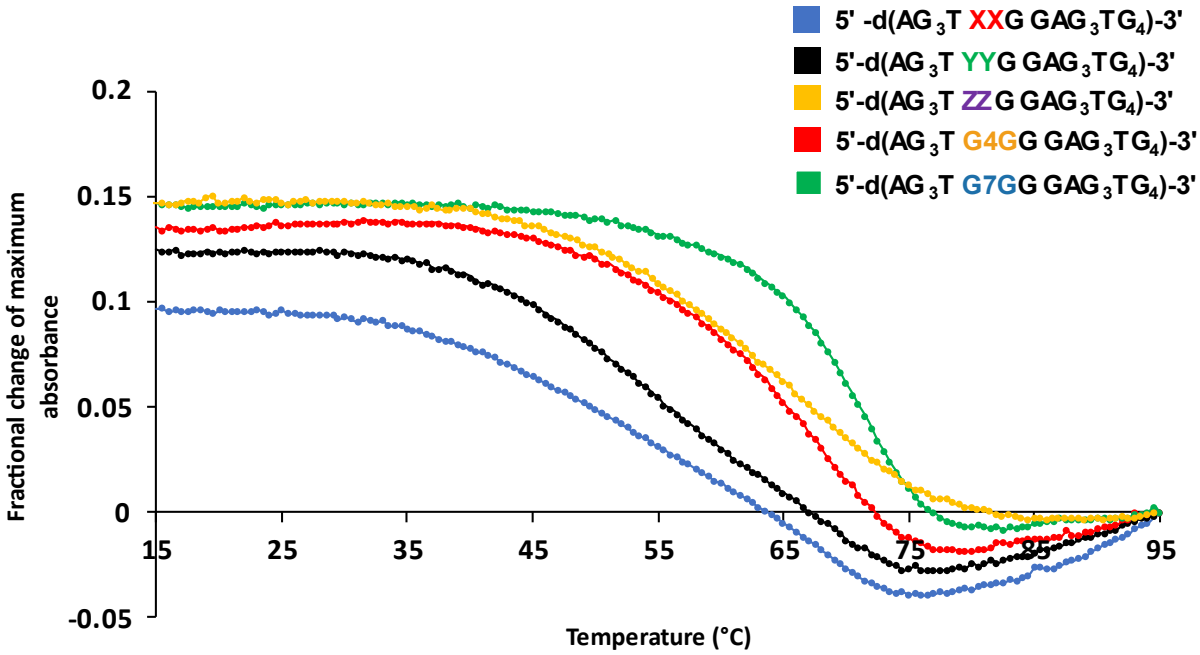


Figure S4.39 UV Thermal denaturation profiles of G-quadruplexes with two sites of alkylation measured at 260 nm and high [K^+]. Buffer: 130 mM KCl, 5 mM KH_2PO_4, 5 mM K_2HPO_4, 5.4 mM KOH, pH 7.2. X= O^6-Methyl dG, Y= O^6-4-hydroxybutyl dG, Z= O^6-7-hydroxyheptyl dG.

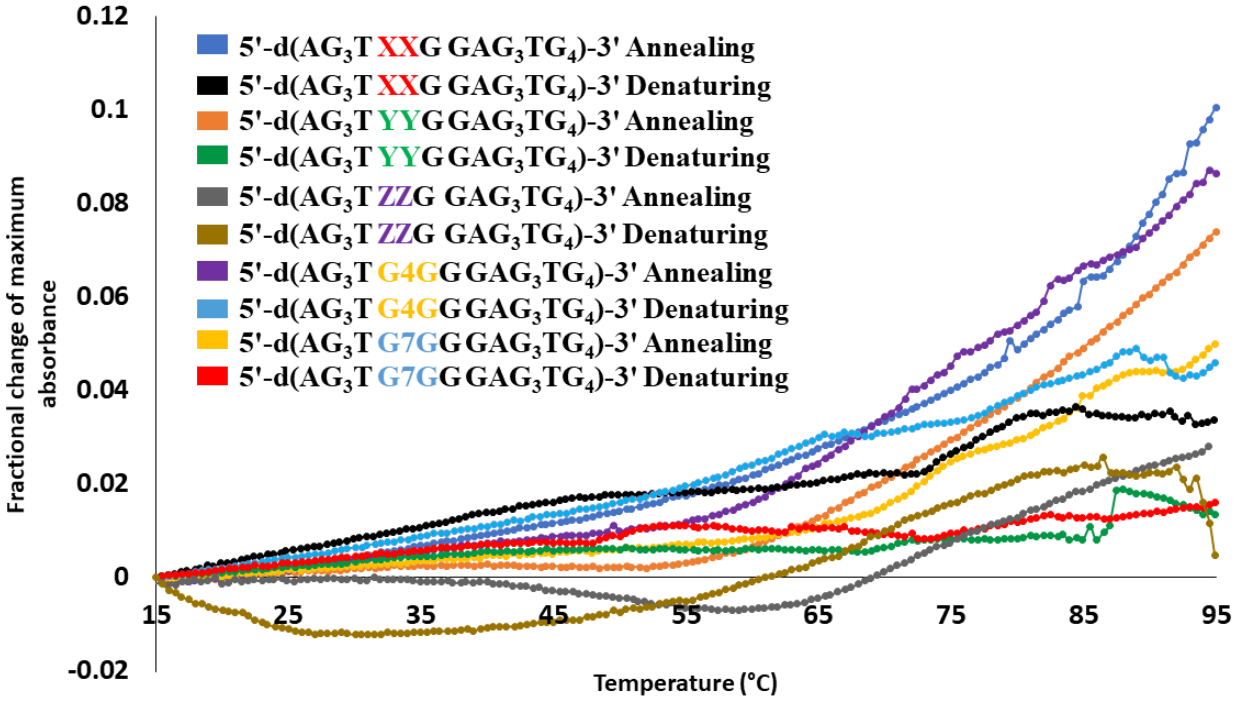


Figure S4.40 Circular dichroism spectra of G-quadruplexes at low salt concentrations containing A) methyl B) butylene C) heptylene lesions and D) IaCLs. X= O^6-Methyl dG, Y= O^6-4-hydroxybutyl dG, Z= O^6-7-hydroxyheptyl dG.

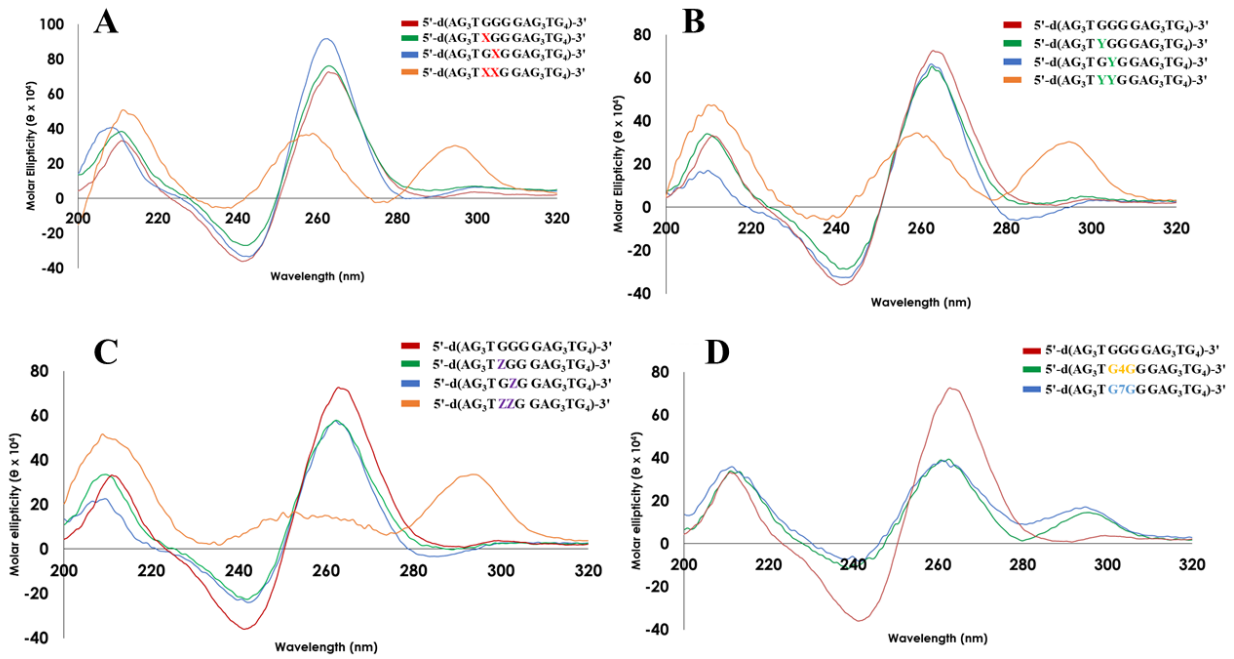


Figure S4.41 Denaturing PAGE of control **ODN-4.1** when incubated with AGTs. Lane 1) 2 pmol **ODN-4.1**, 2) 2 pmol **ODN-4.1** + 20 pmol hAGT, 3) 2 pmol **ODN-4.1** + 60 pmol hAGT, 4) 2 pmol **ODN-4.1** + 20 pmol Ada-C, 5) 2 pmol **ODN-4.1** + 60 pmol Ada-C, 6) 2 pmol **ODN-4.1** + 20 pmol OGT, 7) 2 pmol **ODN-4.1** + 60 pmol OGT, 8) 2 pmol **ODN-4.1** + 20 pmol hOGT, 9) 2 pmol **ODN-4.1** + 60 pmol hOGT.

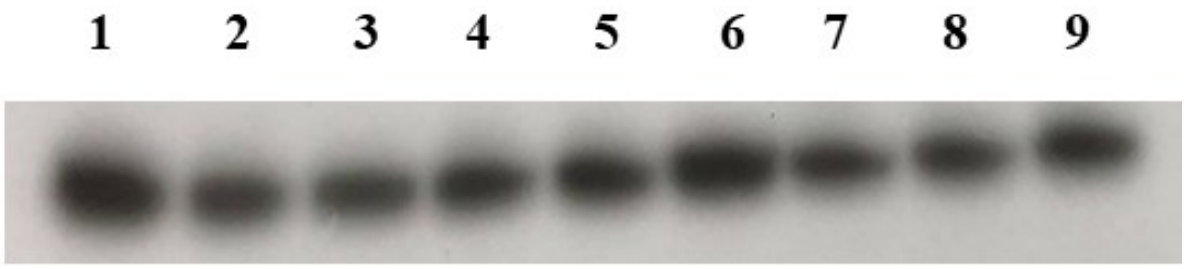


Figure S4.42 Denaturing PAGE of **ODN-4.2** when incubated with AGTs. Lane 1) 2 pmol **ODN-4.2**, 2) 2 pmol **ODN-4.2** + 6 pmol hAGT, 3) 2 pmol **ODN-4.**2 + 60 pmol hAGT, 4) 2 pmol **ODN-4.2** + 6 pmol Ada-C, 5) 2 pmol **ODN-4.2** +60 pmol Ada-C, 6) **ODN-4.2** + repair product, 7) 2 pmol **ODN-4.2** + 6 pmol OGT, 8) 2 pmol **ODN-4.2** + 60 pmol OGT, 9) 2 pmol **ODN-4.2** + 6 pmol hOGT, 10) 2 pmol **ODN-4.2** + 60 pmol hOGT.

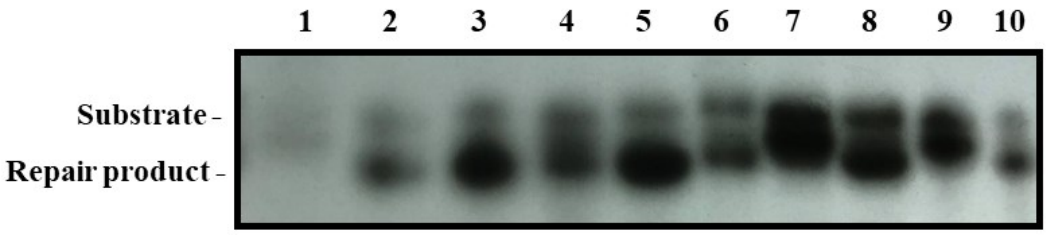
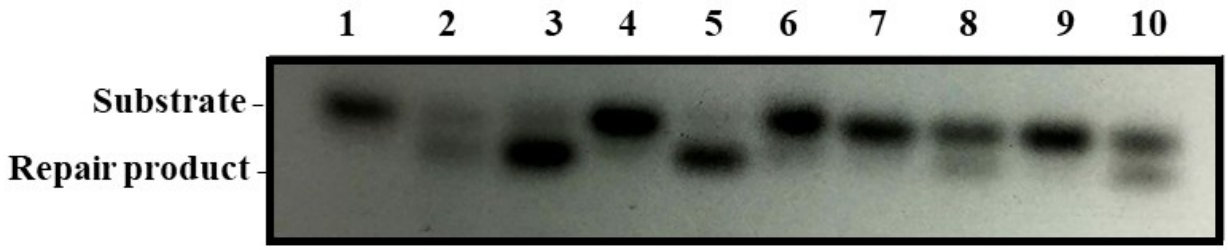


Figure S4.43 Denaturing PAGE of **ODN-4.3** when when incubated with AGTs. Lane 1) 2 pmol **ODN-4.3**, 2) 2 pmol **ODN-4.3** + 6 pmol hAGT, 3) 2 pmol **ODN-4.3** + 60 pmol hAGT, 4) 2 pmol **ODN-4.3** + 6 pmol Ada-C, 5) 2 pmol **ODN-4.3** + 60 pmol Ada-C, 6) **ODN-4.3** + repair product, 7) 2 pmol **ODN-4.3** + 6 pmol OGT, 8) 2 pmol **ODN-4.3** + 60 pmol OGT, 9) 2 pmol **ODN-4.3** + 6 pmol hOGT, 10) 2 pmol **ODN-4.3** + 60 pmol hOGT.



.

Figure S4.44 Denaturing PAGE of **ODN-4.5** when incubated with AGTs. Lane 1) 2 pmol **ODN-4.5**, 2) 2 pmol **ODN-4.5** + 6 pmol hAGT, 3) 2 pmol **ODN-4.5** + 60 pmol hAGT, 4) 2 pmol **ODN-4.5** + 6 pmol Ada-C, 5) 2 pmol **ODN-4.5** + 60 pmol Ada-C, 6) **ODN-4.5** + repair product, 7) 2 pmol **ODN-4.5** + 6 pmol OGT, 8) 2 pmol **ODN-4.5** + 60 pmol OGT, 9) 2 pmol **ODN-4.5** + 6 pmol hOGT, 10) 2 pmol **ODN-4.5** + 60 pmol hOGT.

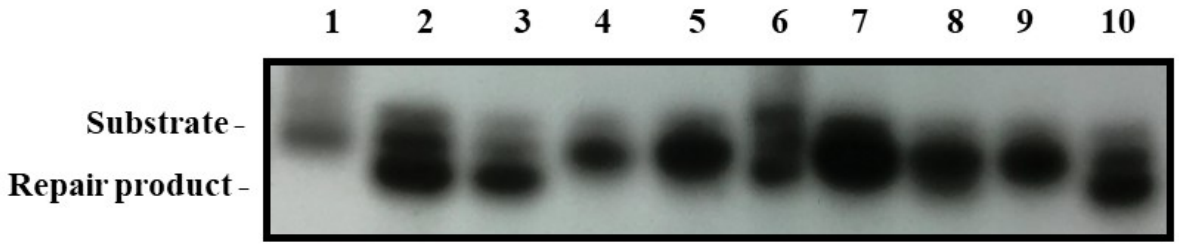


Figure S4.45 Denaturing PAGE of **ODN-4.6** when incubated with AGTs. Lane 1) 2 pmol **ODN-4.6**, 2) 2 pmol **ODN-4.6** + 6 pmol hAGT, 3) 2 pmol **ODN-4.6** + 60 pmol hAGT, 4) 2 pmol **ODN-4.6** + 6 pmol Ada-C, 5) 2 pmol **ODN-4.6** + 60 pmol Ada-C, 6) **ODN-4.6** + repair product, 7) 2 pmol **ODN-4.6** + 6 pmol OGT, 8) 2 pmol **ODN-4.6** + 60 pmol OGT, 9) 2 pmol **ODN-4.6** + 6 pmol hOGT, 10) 2 pmol **ODN-4.6** + 60 pmol hOGT.

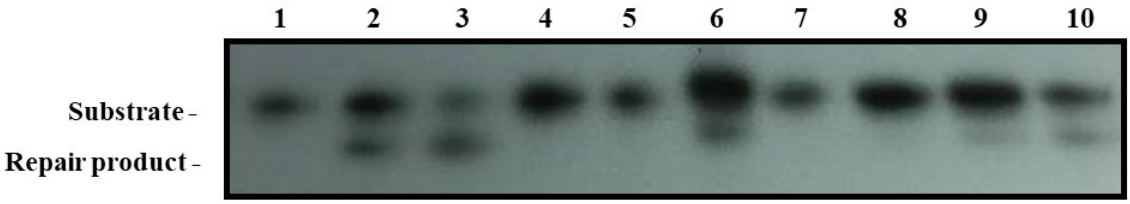


Figure S4.46 Denaturing PAGE of **ODN-4.8** when incubated with AGTs. Lane 1) 2 pmol **ODN-4.8**, 2) 2 pmol **ODN-4.8** + 6 pmol hAGT, 3) 2 pmol **ODN-4.8** + 60 pmol hAGT, 4) 2 pmol **ODN-4.8** + 6 pmol Ada-C, 5) 2 pmol **ODN-4.8** + 60 pmol Ada-C, 6) **ODN-4.8** + repair product, 7) 2 pmol **ODN-4.8** + 6 pmol OGT, 8) 2 pmol **ODN-4.8** + 60 pmol OGT, 9) 2 pmol **ODN-4.8** + 6 pmol hOGT, 10) 2 pmol **ODN-4.8** + 60 pmol hOGT.

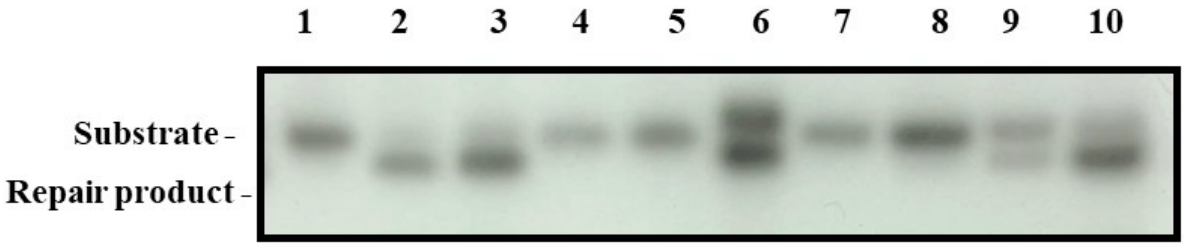


Figure S4.47 Denaturing PAGE of **ODN-4.9** when incubated with AGTs. Lane 1) 2 pmol **ODN-4.9**, 2) 2 pmol **ODN-4.9** + 6 pmol hAGT, 3) 2 pmol **ODN-4.9** + 60 pmol hAGT, 4) 2 pmol **ODN-4.9** + 6 pmol Ada-C, 5) 2 pmol **ODN-4.9** + 60 pmol Ada-C, 6) **ODN-4.9** + repair product, 7) 2 pmol **ODN-4.9** + 6 pmol OGT, 8) 2 pmol **ODN-4.9** + 60 pmol OGT, 9) 2 pmol **ODN-4.9** + 6 pmol hOGT, 10) 2 pmol **ODN-4.9** + 60 pmol hOGT.

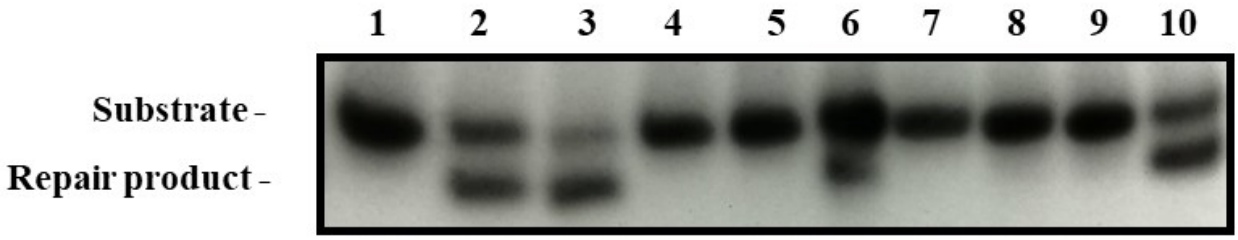
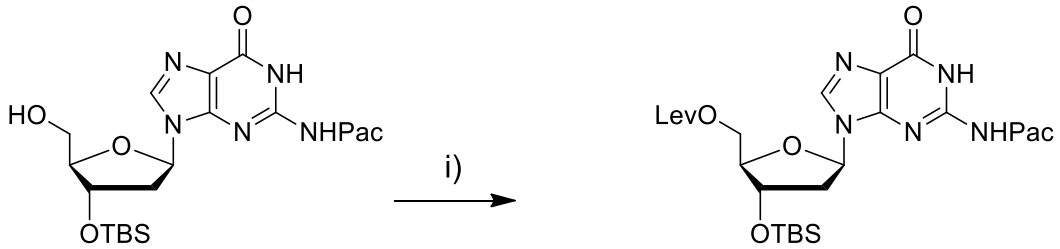


Table S4.1 Mass spectrometry of oligonucleotides in this study.

Oligonucleotide	Expected Mass (Da)	Observed Mass (Da)
ODN-4.1	5779	5779
ODN-4.2	5793	5793
ODN-4.3	5793	5793
ODN-4.4	5807	5807
ODN-4.5	5851	5852
ODN-4.6	5851	5851
ODN-4.7	5900	5923 (Na^+)
ODN-4.8	5893	5896
ODN-4.9	5893	5896
ODN-4.10	6007	6007
ODN-4.11	5771	5771
ODN-4.12	5813	5813

Table S4.2 *T_m*s of G-quadruplexes containing multiple sites of alkylation with high [K^+]. Buffer: 130 mM KCl, 5 mM KH_2PO_4, 5 mM K_2HPO_4, 5.4 mM KOH, pH 7.2.

Oligonucleotide	**T_m (°C)**
ODN-4.4	54
ODN-4.7	54
ODN-4.10	67
ODN-4.11	68
ODN-4.12	72



5

Scheme S4.1. Synthesis of compound **5**. Reagents and conditions: i) Levulinic acid, EDAC •HCl, DMAP in dioxane for 18 h at r.t.

Chapter 5

Synthesis, Characterization and Biophysical Studies on DNA Containing an O^4-Methyl Thymine Threose Nucleic Acid

5.1 Introduction

It has long been postulated that RNA was the precursor to DNA due to its ability to both carry genetic information as well as catalyze reactions.[218] This theory is further supported by the observation that synthetic RNA can self-replicate.[219] Why nature chose DNA however, has been the subject of debate. Interestingly, the six membered pyranose sugars form more stable duplexes however their inability of crosspairing with DNA and RNA deems them unlikely precursors.[220] Furthermore, a lack of base-pairing fidelity is noted in some cases. Six carbon sugars such as glucose and allose have been deemed unlikely precursors due to weak base-pairing and specifity.[221] Threose nucleic acids (TNA) are composed of four carbon sugars and the simplicity of the molecule (simpler in that in can assemble from two identical molecules) makes it a more likely candidate to have been formed in the prebiotic conditions of earth (**Figure 5.1)**. Extensive work by Eschenmoser to support this theory showed that TNA could base-pair with both DNA and RNA in an antiparallel fashion which strengthens the hypothesis that TNA could have been able to pass on genetic information to the more complex macromolecules.[222] The solution structure of a TNA-TNA duplex (**Figure 5.2A)** revealed it adopts a duplex similar to A-form RNA. Furthermore, a single tT insert in a B-form DNA duplex is well tolerated and has very minor influence on the global topology of the structure.[223] tT adopts a sugar pucker of C4′ exo in the duplex context as is the case with the nucleotide. A crystal structure of an an A-form duplex DNA duplex containing TNA-A inserts could explain why TNA base-pairs more strongly with RNA than DNA based on the intranucleotide phosphate backbone distances.[224] TNA is one of many xeno nucleic acids (XNAs) studied many others have been shown to base-pair with DNA and RNA such as locked nucleic acids,[225] peptide nucleic acids[226] and glycerol nucleic acids.[227] The varying structures and stereochemistries lead to vast differences in base-pairng ability and duplex topologies (for example A-form *vs*. B-form, handedness). The XNAs have generated extensive research with due to characteristics such as nuclease stability, thermal stability, ability to fold into functional aptamers and engage in gene silencing.

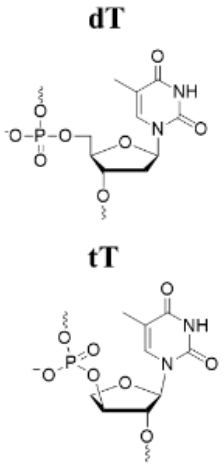


Figure 5.1 Five carbon 2′-deoxyribose *vs.* four carbon threose sugars.

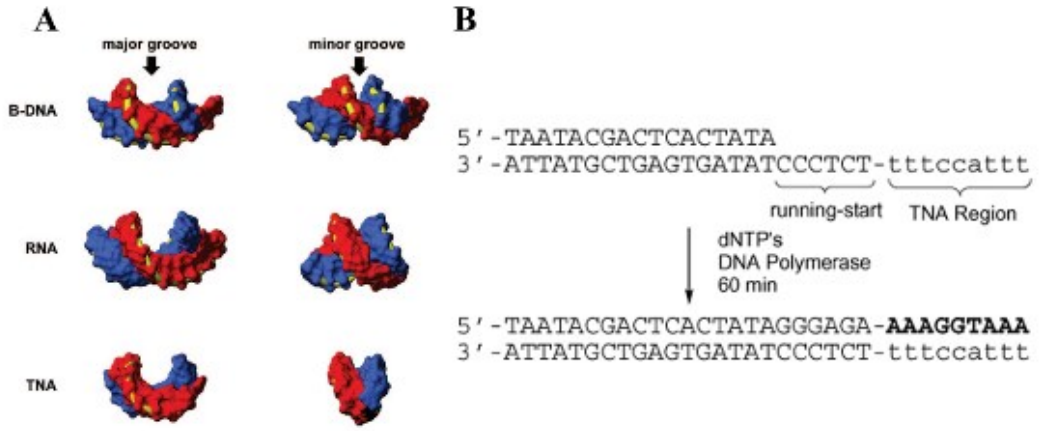


Figure 5.2 Structure of TNA duplex and extension of DNA on a TNA template. A) Comparison of TNA-TNA duplex with DNA and RNA. Reproduced with permission from ref [240] B) Extension of DNA on a short TNA template. TNA is shown in lower case. Reproduced with permission from ref [228].

Chaput *et al.* showed the ability of range of DNA polymerases to add deoxynucleotide triphosphates (dNTPs) across a short TNA template (**Figure 5.2B**) (achievement of full-length product was polymerase dependent). Importantly, the error rate was comparable to that the dNTPs across from a DNA template.[228] The converse was subsequently shown where tT triphosphates (tTTPs) were able to be added to a DNA primer strand by several polymerases (and reverse transcriptases).[229] The reduced catalytic efficient could be due to slight distortion of the backbone with a single tT insert in B-DNA as shown from the high resolution crystal structure which may have a cumulative effect as more tTs are inserted.[223] This has sparked research into examination of other polymerases to improve TNA synthesis on DNA templates.[230,231] We have become accustomed to pyrophosphate as the leaving group during DNA/RNA synthesis however this was not necessarily the case in a prebiotic world. Groups have been working on leaving groups such as 2-methylimidazole which have been shown to extend RNA oligomers on a TNA template without any enzymes.[232] On a final note, if TNA is the precursor to RNA it would likely have to be able to function as an enzyme as well. Through *in vitro* evolution, Yu *et al.* developed a TNA aptamer capable of binding thrombin with a similar affinity as RNA aptamers which has led to further studies.[233]

While we may only speculate informatively about the conditions of the prebiotic world, it is plausible that proteins and TNA coexisted. If methylating agents existed, TNA damage would have been a natural consequence. This lesion would have had relevance to the stability of duplexes, and if repair events occurred it would be pertinent to processing by enzymes such as DNA polymerases. Moreover, given the interest of AGT in biotechnology (highlighted in Chapter 1) understanding the range of substrates could be of relevance. Herein, we look at the effects of a methylated of tT (tT-OMethyl) insertion on the biophysical properties of a DNA duplex as well as attempts to decipher if AGTs from human and *E.coli* can repair the damaged base. DNA sequences were synthesized (**Table 5.1**) containing a dU insert as the 5mer repair product from the sequence containing the O^4 methy lesion, generated by the action of uracil-DNA glycosylase, would be easier to resolve by gel electrophoresis than those in the full-length sequence.

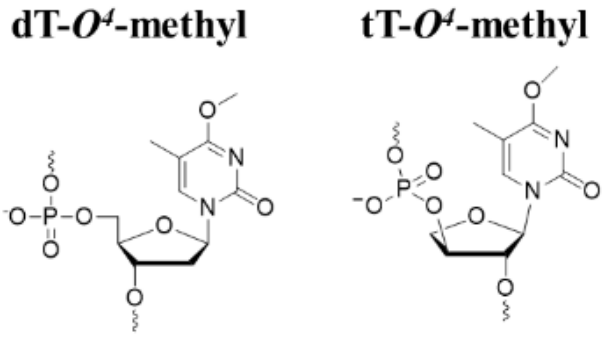


Figure 5.3 O^4-methylated dT and tT examined in this study.

Table 5.1 Sequences examined in this study.

Oligonucleotide	**Sequence (5'→3')**
ODN-5.1:	CG**T** AT dUAG CAC CAG
ODN-5.2:	CG**tT** AT dUAG CAC CAG
ODN-5.3	CG**dT-O^{4}Me** AT dUAG CAC CAG
ODN-5.4	CG**tT-*O*^{4}-OMe** AT dUAG CAC CAG

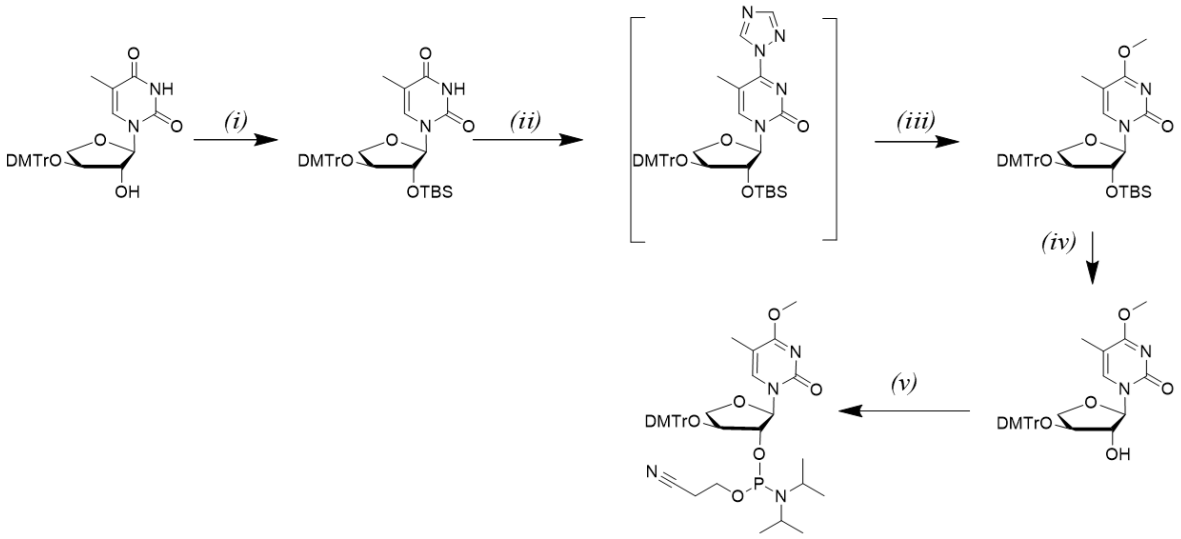
5.2 Results and Discussions

5.2.1 Synthesis of O4-methyl tT monomer and oligonucleotides

1-{[2'-*O*-tert-butyldimethylsilyl-3'-*O*-[4,4'-dimethoxytrityl]-α-L-threofuranosyl}thymidine (Compound **5.1**) was acquired from the lab of Dr. Chaput (University of California, Irvine). The initial step in the synthesis was to protect the 2' hydroxyl position with a *tert*-butyldimethyl silyl group which required an unusually high quantity of equivalents relative to DNA nucleosides to give compound **5.2** (**Scheme 5.1**). The convertible nucleoside was successfully synthesized in

quantitative yield by a methodology previously employed by our group followed by alkylation of the nucleobase with sodium methoxide in methanol to afford compound **5.3**. The TBS group was then removed with a fluoride treatment in moderate yield (compound **5.4**) and lastly phosphitylation was accomplished to give the synthesizer ready nucleoside phosphoramidite compound **5.5**.

Oligonucleotide **ODN-5.4** containing O^4-methyl tT was prepared by solid-phase synthesis. Increased coupling time was performed for the modified tT nucleoside phosphoramidite **5.5**, and due to the reduced nucleophilicity of the secondary nucleophilic alcohol the subsequent



Scheme 5.1 Synthesis of O^4 methylated thymine TNA. i) TBS-Cl (6 eq), Im (12 eq), DMAP (cat. 2 mg), DCM, r.t., 24 h ii) 1,2,4 triazole (27 eq), $POCl_3$ (2.7 eq), TEA (27 eq), MeCN:DCM, r.t, 5 h iii) NaOMe (3.6 eq), MeOH iv) TBAF (1.2 eq) (1 M in THF), THF, r.t, 5 min v) Cl-P(OCE)N(iPr)$_2$ (1.3 eq), DIPEA (1.6 eq), DCM, r.t., 30 min.

unmodified nucleobases also had an extended coupling time. Standard fast deprotecting conditions with 0.05 M K_2CO_3/MeOH and purification by IEX HPLC were successful in acquiring the desired oligonucleotide as confirmed by ESI-MS. Control sequences were also synthesized with unmodified TNA and alkylated dT (**Table 5.1**).

5.2.2 UV Thermal Denaturation Studies

In order to assess the effects of methylation on duplex thermal stability of the O^4 position of tT UV thermal denaturation studies were performed. The duplexes exhibited a single sigmoidal transition (**Figure 5.4A**) and a similar destabilizing effect of ~8 ºC for O^4methylated dT as that for tT relative

to the controls was observed. Mismatch studies against a guanine were also investigated. Guanine was chosen for the mismatch as it is a purine commonly incorporated across from O^4-methyl dT by high fidelity DNA polymerases.[234,235] While a 6 ºC drop was noted for control DNA against a

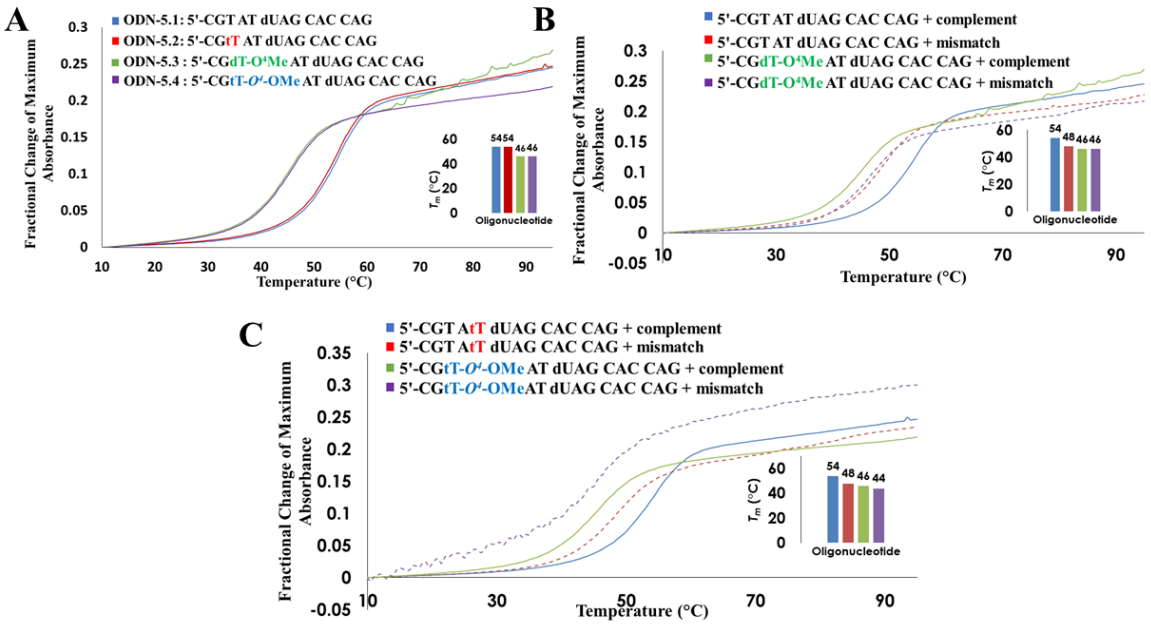


Figure 5.4 UV Thermal denaturation profiles of TNA modified oligonucleotides. Mismatches are represented with dotted lines. A) O^4-methylation on in a DNA and TNA context against an adenine. B) Mismatched O^4-methyl dT in a DNA strand. C) Mismatched O^4-methyl tT in a DNA strand.

mismatch, there was no difference in T_m if O^4-methy dT was paired against a guanine or an adenine with values of 46 ºC observed for the DNA series (**Figure 5.4B**). A 6 ºC destabilizing effect was also noted for the unmodified TNA oligonucleotides. When O^4-methyl tT was mismatched against a guanine it was 2 ºC more destabilizing than an adenine which merits further investigation (**Figure 5.4C**). Evidently, O^4-methylated TNA despited the sugar containing one less carbon behaves similarly to the thermal stability as the DNA counterpart.

5.2.3 Circular Dichroism

Analysis of the circular dichroism spectra revealed that the introduction of tT or O^4-methyl tT whether properly base paired or across from a guanine residue did not influence major structural

perturbations (**Figure 5.5**). A characteristic B-form DNA spectrum was observed with positive maxima around 275 nm and negative maxima at ~245 nm for all duplexes studied.

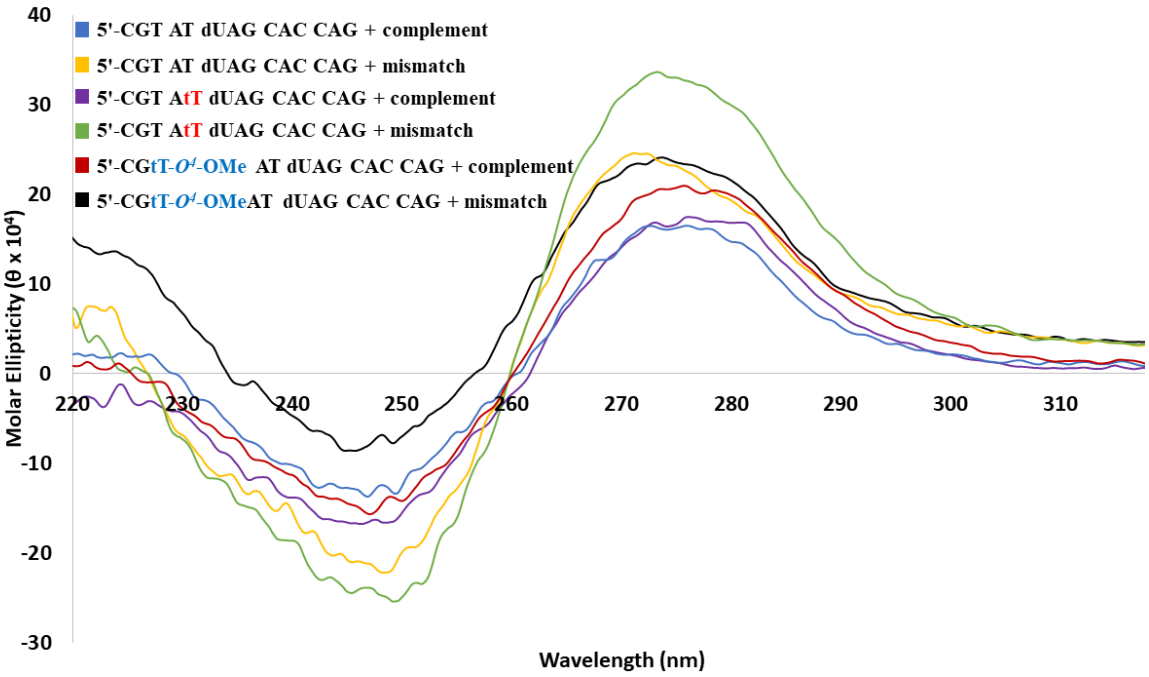


Figure 5.5 Circular dichroism spectra of TNA modified DNA duplexes.

5.2.4 Attempts to Evaluate Repair by AGTs

In our AGT assays with monoadducts on the nucleobases, the substrates are duplexes containing sequences with a site that can be cleaved by a restriction enzyme. If the alkyl adduct is present (indicating no repair) then there is no cleavage by the restriction enzyme and the full-length oligonucleotide is present after treatement with a restriction enzyme. If the alkyl adduct is removed (indicating repair) the restriction enzyme cleaves the repaired oligonucleotide and a shorter oligonucleotide is produced. This allows for the cleavage products to be resolved by gel electrophoresis. Our lab normally employs this assay with the restriction enzymes *Bcl1* and *PvuII*. Repair assays with the duplexes containing tT were challenging to accomplish. Testing various parameters including incubation time, equivalents and tT location, the restriction enzymes were unable to cleave the unmodified strands rendering this assay inadequate (see **Section 5.4.2 Figure S10**). Next we attempted to resolve repair products of single stranded oligonucleotide 14mers by RP HPLC however were unsuccessful in a buffer system of NaH_2PO_4 with the non-polar mobile

phase composed of MeCN (a variety of gradients were tested) and a C-18 column. With extended run times poor peak shape was observed. Monitoring repair by UV active species by HPLC is further complicated if duplex DNA, the preferred substrate of AGTs, containing a tT insert is to be studied as the complement is another oligonucleotide to be separated.

New sequences were studied (those in the above table) containing a 2'-deoxyuridine present which is recognized and cleaved by uracil-DNA glycosylase (UDG). A solvent system was found to resolve the 5mer repair product and the O^4-methylated DNA which was promising. Trial attempts were performed with DNA sequences **ODNs 5.1** and **5.3** to conserve material. However when the assay was done with AGTs and then a UDG treatment there were many UV active peaks such as DTT and the peak shape and resolution of oligonucleotides was poor (data not shown). Attempts to separate the 5mers by PAGE by varying parameters such as PAGE percentage and run times were also unsuccessful. Chloroform-phenol extractions were attempted however HPLC traces were inconclusive. Evidently further assay development is required and techniques such as ESI-MS would be a suitable semi-quantitative alternative.

5.3 Conclusion

In conclusion, O^4-methyl tT can be synthesized in good yield *via* convertible nucleoside chemistry which has shown much success with its DNA counterpart. A single O^4-methyl tT insert in a DNA strand is similarly thermally destabilizing to the DNA duplex as O^4-methylated thymine whether across from an adenine or guanine base. Circular dichroism experiments reveal that a single O^4-methyl tT insert has minimal effects on the global topology of the B-form DNA. Repair studies will be done examining different stationary phases (IEX for example) and UDG digests to optimize resolution with different denaturing PAGE percentages and acrylamide/bis-acrylamide ratios to resolve the 5mer repair products.

5.4 Experimental

5.4.1 Supporting Methods

5.4.1.1 General

General information on the suppliers of reagents and chemicals can be found in **Chapters 2-4**. Likewise the same experimental techniques and instrumentation were used with minor

modifications discussed below. tT precusor nucleoside was generously provided by Dr. John Chaput (University of California, Irvine).

Automated solid-phase synthesis was performed with "fast deprotecting" nucleoside phosphoramidites which we deprotected with 0.05 M K_2CO_3 for 4 h with gentle rocking. The solution was neutralized with 5 µL of glacial acetic acid and then standard purification conditions were employed.

Modifications to sample preparation for UV thermal denaturation studies and circular dichroism was that the single stranded concentration was 3.8 µM and the buffer used was 90 mM NaCl, 10 mM NaH_2PO_4, 1 mM EDTA, pH 7.

5.4.1.2 Chemical synthesis of nucleosides

*1-{[2'-O-tert-butyldimethylsilyl-3'-O-[4,4'-dimethoxytrityl]-α-L-threofuranosyl}thymidine (***5.2***)*

To a solution of **5.1** (0.17 g, 0.32 mmol), imidazole (0.13 g, 1.92 mmol) and DMAP (cat. 2 mg) in anhydrous DCM (3.4 mL) was added TBS-Cl (0.14 g, 0.96 mmol) while stirring. After 6 h, additional imidazole (0.13 g, 0.1.92 mmol) and TBS-Cl (0.14 g, 0.96 mmol) was added. After 24 h, the volume was diluted to 100 mL with DCM, and the organic layer was washed with $NaHCO_3$ (3 % w/v, *aq*) (2 x 75 mL), brine (1 x 75 mL), dried over Na_2SO_4 (~4 g), decanted and the solvent removed *in vacuo*. The crude compound was purified by silica gel flash column chromatography with a gradient of EtOAc:hexanes (1:9→3:7) to obtain 165 mg (80 %) of **5.2** as a colourless foam. R_f (SiO_2, TLC: 0.40 (5 % MeOH:DCM). $\lambda_{max(MeCN)}$ = 238 nm. ^{1}H NMR (500 MHz, $CDCl_3$) δ 8.30 (s, 1H, NH), 7.38 (d, *J* = 1.1 Hz, 1H, H6), 7.35 – 7.16 (m, 9H, Ar), 6.85 – 6.75 (m, 4H, Ar, H6), 5.57 (s, 1H, H1'), 4.19 (s, 1H, H3'), 4.03 (d, *J* = 3.4 Hz, 1H, H2'), 3.81 (d, *J* = 3.5 Hz, 1H, H4'), 3.80 – 3.76 (s, 6H, 2 x OCH_3), 3.21 (m, 1H, H4'), 1.78 (d, *J* = 0.9 Hz, 3H, -CH_3), 0.82 (d, *J* = 2.9 Hz, 9H, -(Si-CH_3)$_3$), 0.16 (s, 3H, Si-CH_3), 0.03 (s, 3H, Si-CH_3).^{13}C NMR (126 MHz, $CDCl_3$) δ 164.02, 158.77, 158.75, 150.20, 144.60, 137.07, 135.77, 135.57, 129.79, 129.70, 128.15, 127.66, 127.11, 113.59, 113.53, 108.76, 92.83, 88.08, 80.69, 78.37, 75.30, 55.25, 25.64, 25.58, 17.73, 12.70, -0.02, -3.59, -4.59, -5.07. HRMS (ESI-MS) *m/z* calculated for $C_{36}H_{44}N_2NaO_7Si^+$ $[M+Na]^+$: 667.2815; found: 667.2805.

*1-{[2'-O-tert-butyldimethylsilyl-3'-O-[4,4'-dimethoxytrityl-O^4-methyl]-α-L-threofuranosyl}thymidine (***5.3***)*

Compound **5.2** was co-evaporated with 10 mL of anhydrous MeCN. To a solution of **2** (0.25 g, 0.39 mmol), 1,2,4-triazole (0.60 g, 8.7 mmol) and TEA (1.3 mL, 8.9 mmol) in MeCN:DCM (3.8:3.8 mL) was added $POCl_3$ (89 µL, 0.97 mmol) dropwise while stirring on ice. After 3.5 h, additional 1,2,4-triazole (0.12 g, 1.7 mmol), TEA (250 µL, 1.8 mmol) and $POCl_3$ (9.1 µL, 0.098 mmol) were added while stirring on ice. After 5 h, the solvent was removed *in vacuo* and the volume was brought to 100 mL with DCM. The organic layer was washed with $NaHCO_3$ (3 % w/v, *aq*) (2 x 75 mL), brine (1 x 75 mL), dried over Na_2SO_4 (~4 g), decanted and the solvent removed *in vacuo*. To a solution of the crude compound in MeOH (5 mL) was added NaOMe (0.073 g, 1.4 mmol). After 18 h, the solvent was removed *in vacuo* and the volume was brought to 100 mL with DCM. The organic layer was washed with $NaHCO_3$ (3 % w/v, *aq*) (2 x 75 mL), brine (1 x 75 mL), dried over Na_2SO_4 (~4 g), decanted and the solvent removed *in vacuo*. The crude compound was purified by silica gel flash column chromatography with a gradient of MeOH:DCM (0.5→ 1 %) to obtain 150 mg (60 % over two steps) of **5.3** as a colourless foam. R_f (SiO_2, TLC: 0.48 (5 % MeOH:DCM). $\lambda_{max(MeCN)}$ = 282 nm. ^{1}H NMR (500 MHz, $CDCl_3$) δ 7.48 (d, *J* = 0.7 Hz, 1H, H6), 7.32 – 7.06 (m, 9H, Ar), 6.86 – 6.69 (m, 4H, Ar), 5.62 (s, 1H, H1'), 4.46 (s, 1H, H3'), 4.04 (s, 3H, -O^4-CH_3), 3.99 (d, *J* = 3.4 Hz, 1H, H2'), 3.81 (dd, *J* = 9.9, 3.5 Hz, 1H, H4'), 3.78 (s, 6H, 2 x -OCH_3), 3.06 (d, *J* = 9.8 Hz, 1H, H4'), 1.80 (d, *J* = 0.5 Hz, 3H, -CH_3), 0.82 (s, 9H, (Si-CH_3)$_3$), 0.26 (s, 3H, Si-CH_3), 0.09 (s, 3H, Si-CH_3). ^{13}C NMR (126 MHz, $CDCl_3$) δ 170.92, 158.69, 158.66, 156.14, 144.70, 140.90, 135.96, 135.65, 129.79, 129.71, 128.00, 127.69, 126.96, 113.48, 113.38, 102.75, 93.82, 87.85, 80.63, 78.43, 75.35, 55.22, 55.21, 54.44, 25.67, 17.76, 12.39, -4.46, -5.23. HRMS (ESI-MS) *m/z* calculated for $C_{37}H_{46}N_2NaO_7Si^+$ $[M+Na]^+$: 681.2972; found: 681.3050.

*1-{[3'-O-[4,4'-dimethoxytrityl-O^4-methyl]-α-L-threofuranosyl}thymidine (***5.4***)*

To a solution of **5.3** (0.15 g, 0.22 mmol) in THF (3 mL) was added TBAF (1 M in THF) (260 µL, 0.26 mmol) dropwise while stirring. After 5 min, the solvent was removed *in vacuo* and the volume was brought to 100 mL with DCM. The organic layer was washed with $NaHCO_3$ (3 % w/v, *aq*) (2 x 75 mL), brine (1 x 75 mL), dried over Na_2SO_4 (~4 g), decanted and the solvent removed *in vacuo*. The crude compound was purified by silica gel flash column chromatography with 5 %

MeOH:DCM to obtain 150 mg (60 % over two steps) of **5.4** as a colourless foam. R_f (SiO_2, TLC: 0.30 (5 % MeOH:DCM). $\lambda_{max(MeCN)}$ = 282 nm. ^{1}H NMR (500 MHz, $CDCl_3$) δ 7.55 (s, 1H, H6), 7.44 – 7.05 (m, 9H, Ar), 6.91 – 6.66 (m, 4H, Ar), 5.61 (d, *J* = 1.1 Hz, 1H, H1′), 4.25 – 4.19 (m, 2H, H2' + H3'), 4.02 (s, 3H, O^4-CH_3), 3.78 (d, *J* = 1.0 Hz, 6H, 2 x OCH_3), 3.54 (dd, *J* = 9.8, 4.4 Hz, 1H, H4'), 3.25 (dd, *J* = 9.8, 2.6 Hz, 1H, H4'), 1.89 (s, 3H, - CH_3). ^{13}C NMR (126 MHz, $CDCl_3$) δ 171.10, 158.74, 158.72, 156.68, 144.80, 139.80, 136.06, 135.78, 129.92, 129.88, 127.99, 127.87, 127.00, 113.44, 113.38, 103.80, 93.84, 87.63, 81.44, 77.79, 74.28, 55.23, 54.67, 46.04, 12.36, -0.02. HRMS (ESI-MS) *m/z* calculated for $C_{31}H_{32}N_2NaO_7^+$ $[M+Na]^+$: 567.2107; found: 567.2189.

*1-{[2'-O-(2-cyanoethoxy(diisopropylamino)-phosphino)-3'-O-[4,4'-dimethoxytrityl-O^4-methyl]-α-L-threofuranosyl}thymidine (***5.5***)*

To a solution of **5.4** (95 mg, 0.17 mmol) and DIPEA (48 µL, 0.28 mmol) in DCM (2 mL) was added Cl-POCEN*i*Pr_2 (52 µL, 0.23 mmol) while stirring. After 30 minutes, the solvent was removed *in vacuo* and the crude compound was dissolved in 50 mL of EtOAc. The compound was washed with $NaHCO_3$ (*aq*, 3 %) (2 x 35 mL), brine (1 x 35 mL) and dried over Na_2SO_4 (~2 g). The crude compound was purified by silica gel flash column chromatography (conditioned with 0.5 % TEA/EtOAc) with EtOAc:hexanes (8:2) to obtain 107 mg (84 %) of **5.5** as a colourless foam. R_f (SiO_2, TLC: 0.62, 0.71 (100 % EtOAc). $\lambda_{max(MeCN)}$ = 282 nm. ^{1}H NMR (500 MHz, d_6-acetone) δ 7.69 (s, 1H), 7.64 (m, 1H), 7.52 – 7.13 (m, 17H), 7.05 – 6.75 (m, 8H), 5.78 (s, 1H), 5.75 (d, J= 1.3 Hz, 1H), 4.56 (d, J=9.5 Hz, 1H), 4.36 (d, J=8.5 Hz, 1H), 4.30 (d, J=3.6 Hz, 1H), 4.18 (d, J=3.5 Hz, 1H), 4.05-3.99 (m, 1H), 3.96 (s, 6H), 3.92 – 3.81 (m, 3H), 3.81 – 3.78 (s, 11H), 3.74-3.69 (m, 2H), 3.66 – 3.54 (m, 4H), 3.48 (d, J= 9.9 Hz, 1H), 3.32 (d, J= 10.0 Hz, 1H), 2.91 – 2.73 (m, 7H), 2.61-2.56 (m, 1H), 2.52-2.46 (m, 1H), 1.23 (d, *J* = 6.8 Hz, 7H), 1.19 (d, *J* = 6.8 Hz, 7H), 1.15 (d, *J* = 6.8 Hz, 11H), 1.03 (d, *J* = 6.8 Hz, 11H). ^{13}C NMR (126 MHz, d_6-acetone) δ 170.71, 158.95, 155.04, 145.11, 141.23, 141.12, 135.85, 135.71, 129.99, 129.92, 129.87, 129.84, 127.99, 127.94, 127.80, 127.70, 126.84, 126.82, 116.74, 113.54, 113.46, 113.44, 113.36, 102.08, 102.06, 92.42, 92.39, 88.03, 87.96, 81.85, 81.68, 77.72, 77.68, 75.07, 74.91, 59.09, 58.93, 54.69, 53.54, 43.49, 43.39, 43.21, 43.11, 24.02, 23.97, 23.92, 23.86, 23.79, 23.73, 19.73, 19.67, 11.63, 0.22. HRMS (ESI-MS) *m/z* calculated for $C_{40}H_{49}N_4NaO_8P^+$ $[M+Na]^+$: 767.3186; found: 767.3258.

5.4.2 Supplementary Figures

Figure S5.1 500 MHz ^{1}H NMR spectrum of compound **5.1** (in $CDCl_3$)

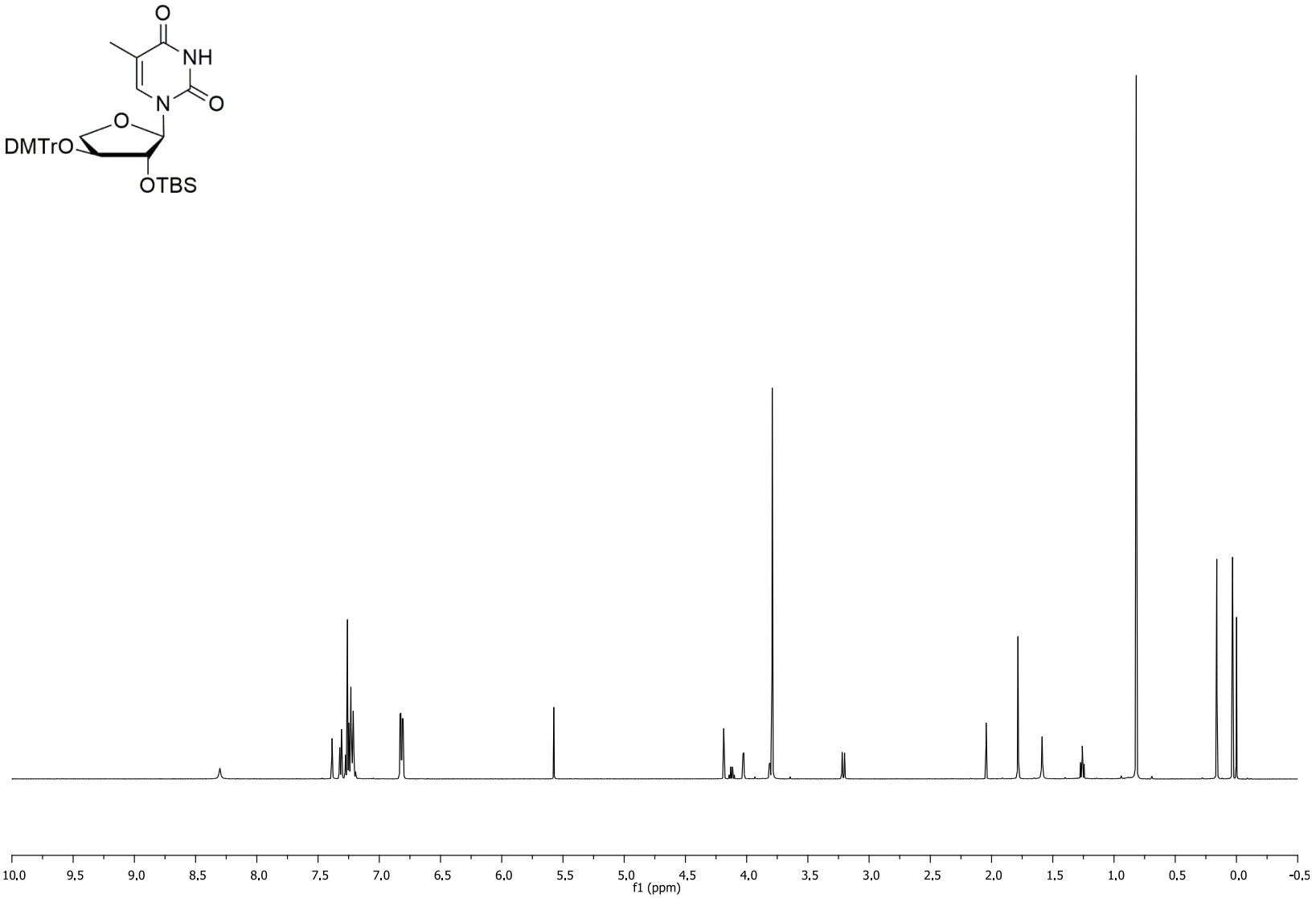


Figure S5.2 125.7 MHz ^{13}C NMR spectrum of compound **5.1** (in $CDCl_3$)

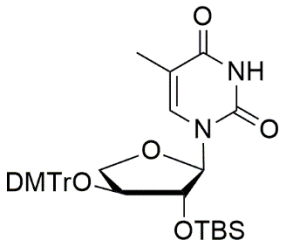


Figure S5.3 500 MHz ^{1}H NMR spectrum of compound **5.2** (in $CDCl_3$)

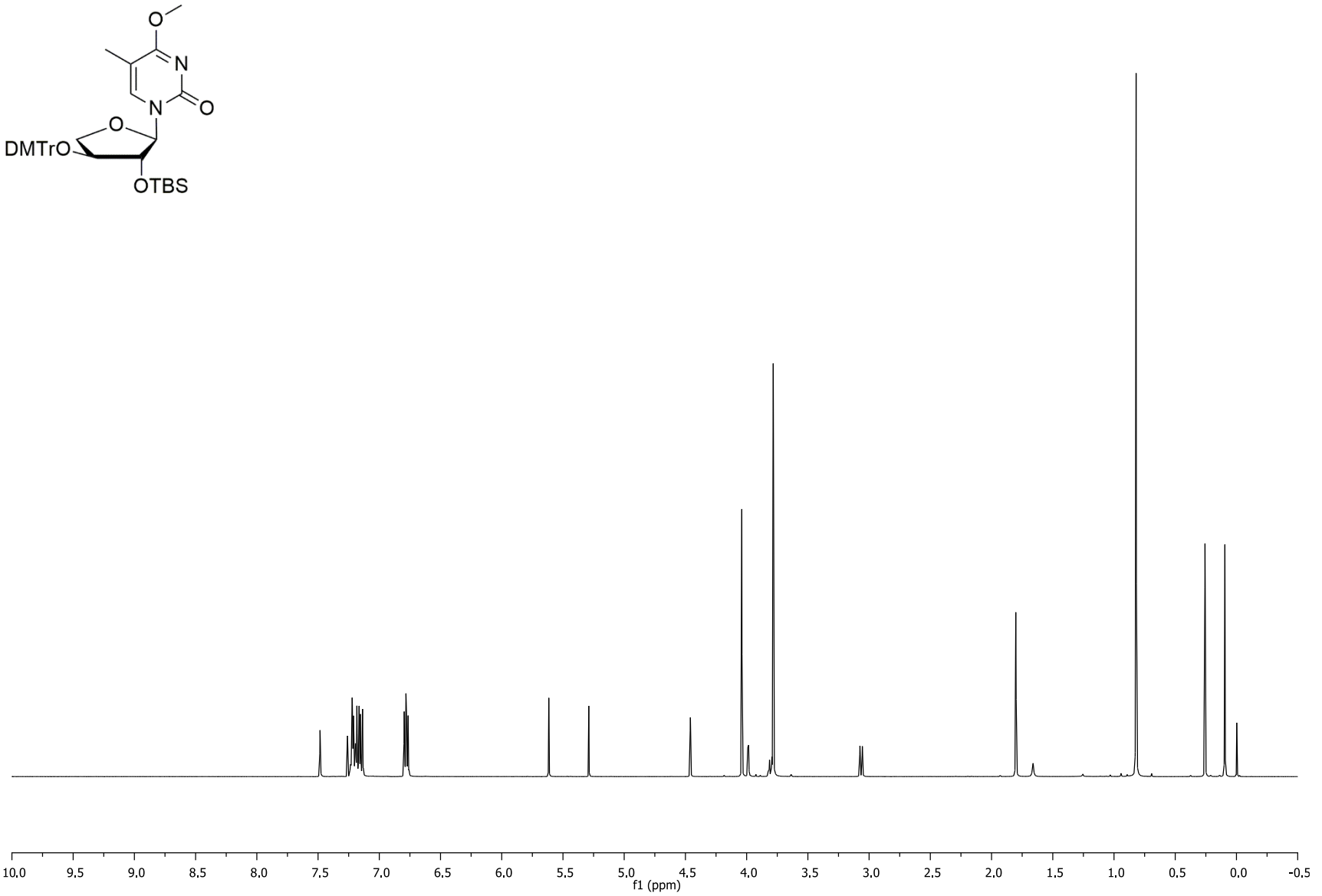


Figure S5.4 125.7 MHz ^{13}C NMR spectrum of compound **5.2** (in $CDCl_3$)

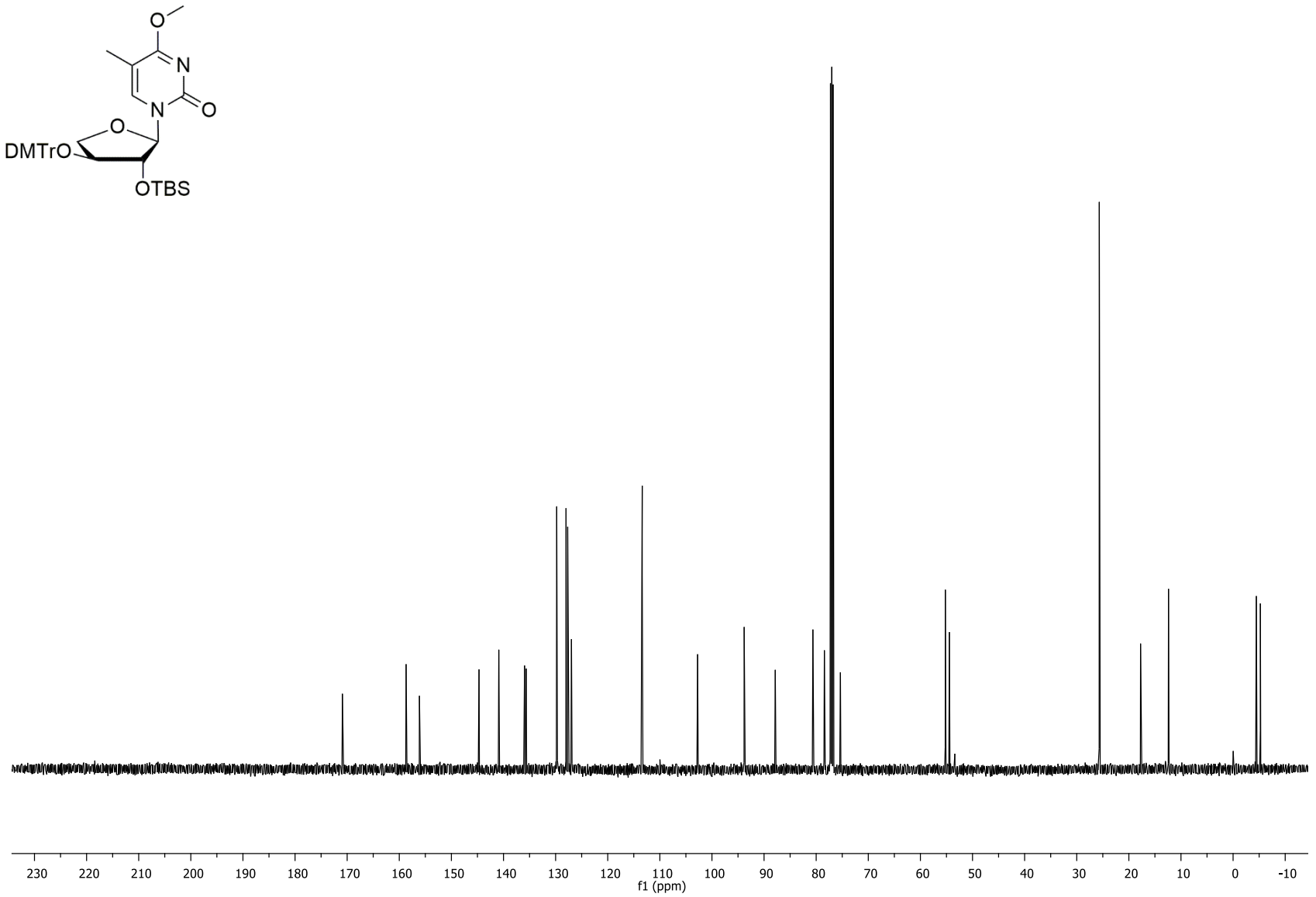


Figure S5.5 500 MHz ^{1}H NMR spectrum of compound **5.3** (in $CDCl_3$)

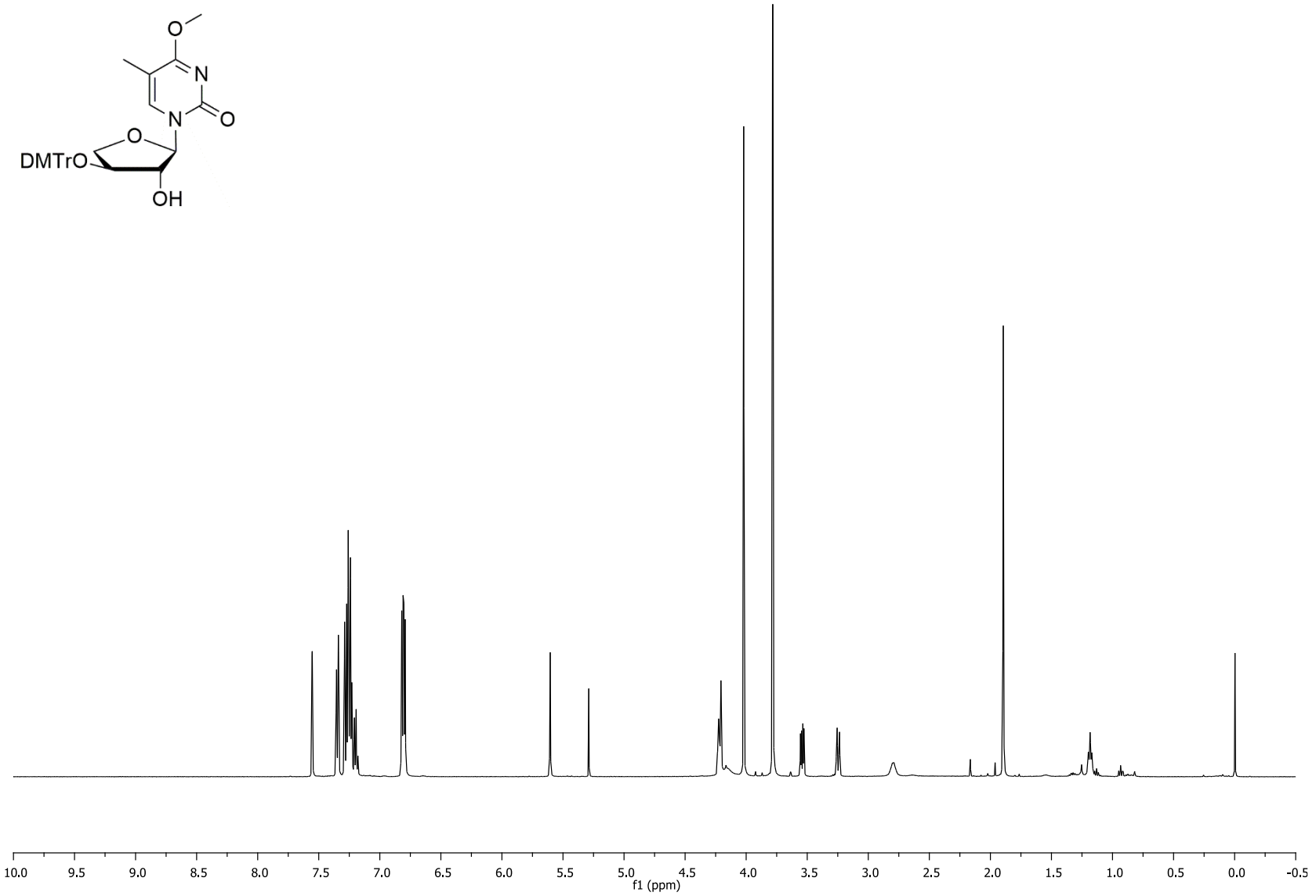


Figure S5.6 125.7 MHz ^{13}C NMR spectrum of compound **5.3** (in $CDCl_3$)

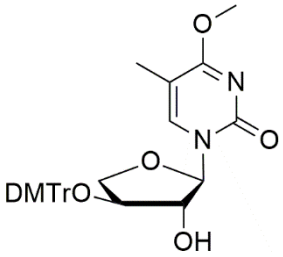


Figure S5.7 500 MHz ^{1}H NMR spectrum of compound **5.4** (in $CDCl_3$)

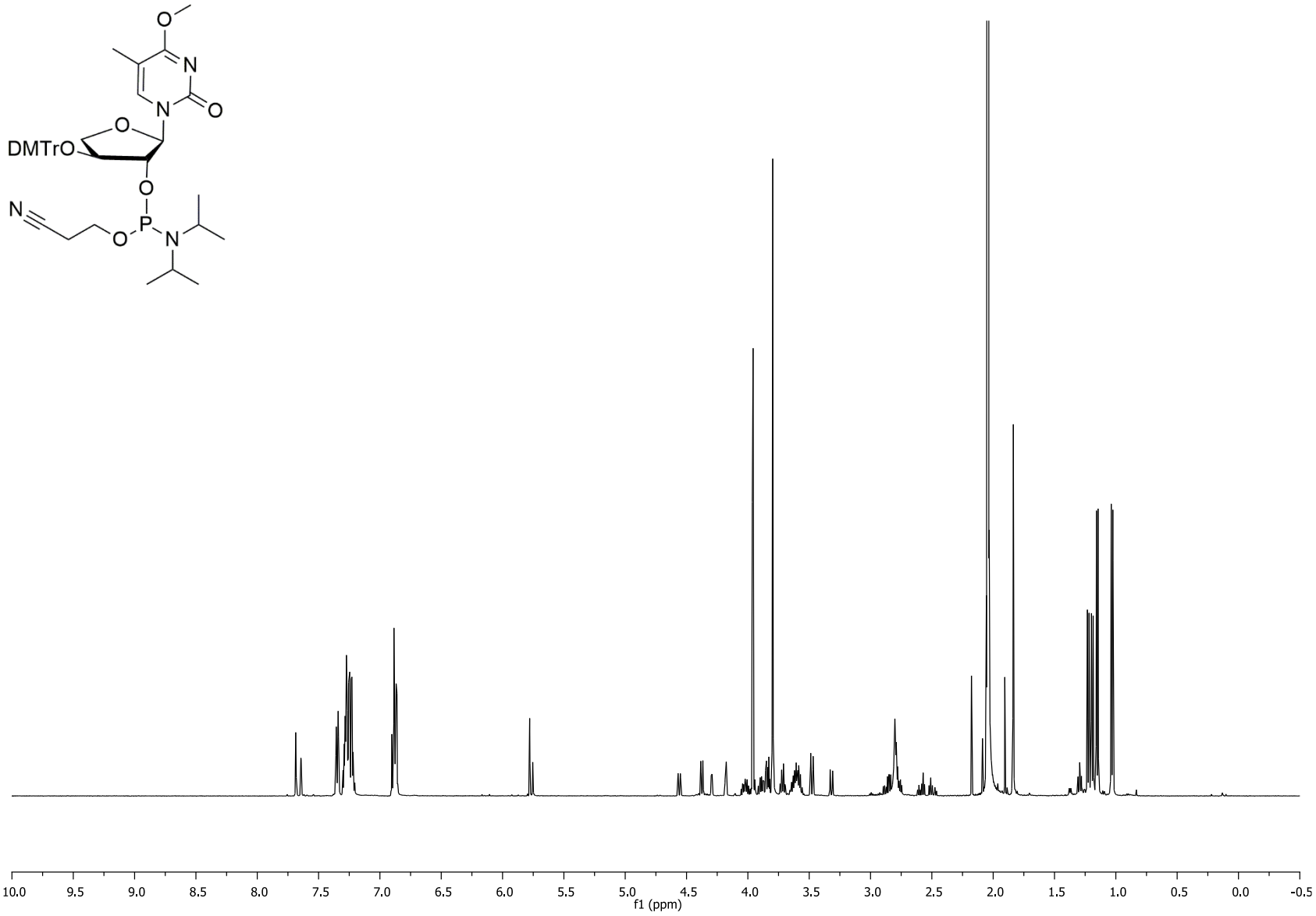


Figure S5.8 125.7 MHz ^{13}C NMR spectrum of compound **5.4** (in $CDCl_3$)

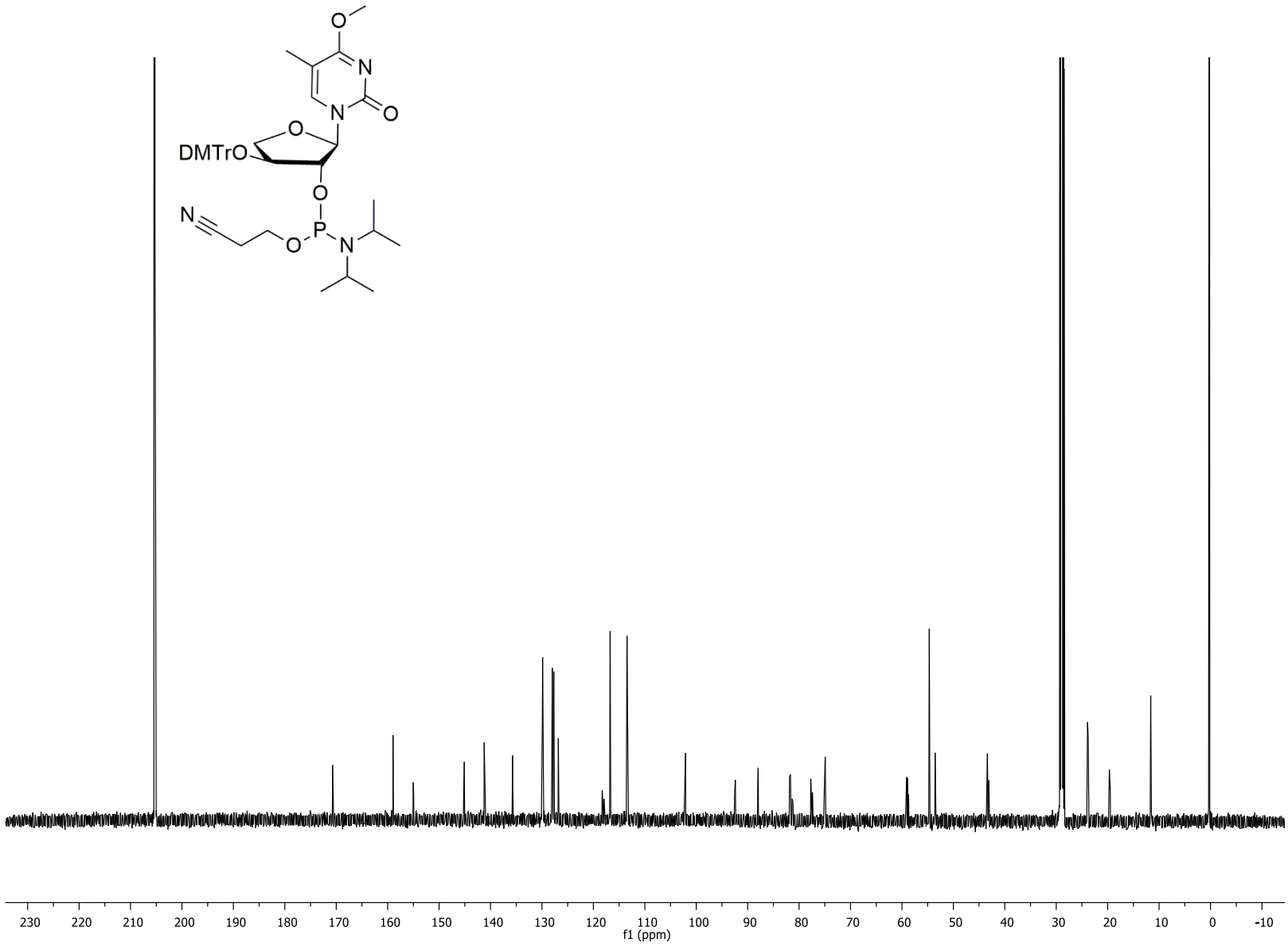


Figure S5.9 202.3 MHz ^{31}P NMR spectrum of compound **5.4** (in d_6-acetone)

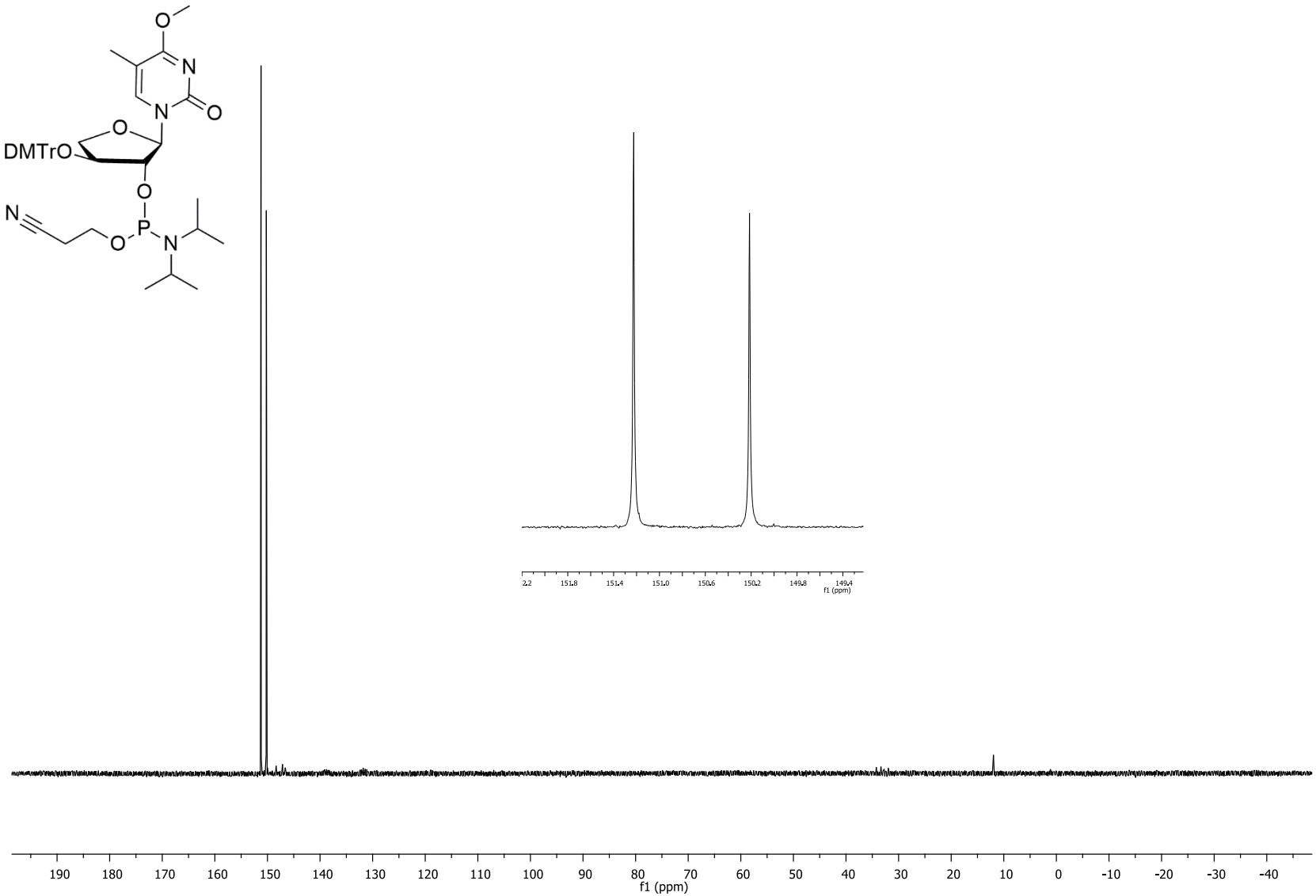


Figure S5.10 Restriction enzyme digests of DNA sequences containing unmodified tT near the cut site.

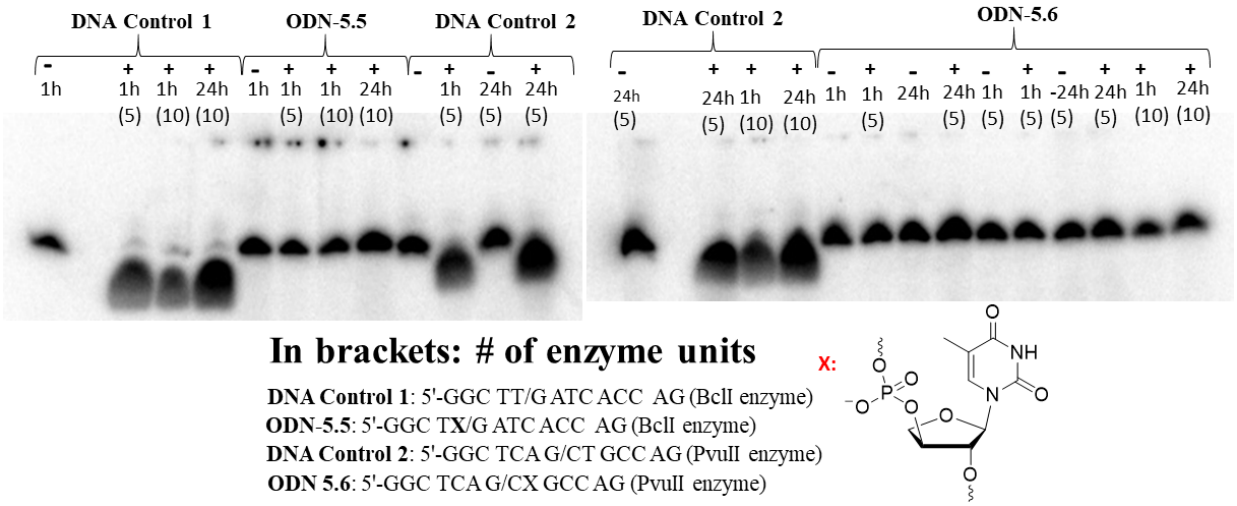


Table S5.1 ESI-MS data for ODNs in this study.

Oligonucleotide	Expected Mass (Da)	Observed Mass (*m/z*)
ODN-5.1	4234	4232
ODN-5.2	4220	4218
ODN-5.3	4248	4248
ODN-5.4	4234	4232

Chapter 6

6.1 Contribution to Knowledge

In addition to the well characterized double helix, DNA is capable of self-assembling into many structures. The ability of proteins to interact with these structures has drawn interest from a molecular biology perspective as well as applications in DNA nanotechnology. In **Chapter 2** we developed a strategy to covalently trap the active site of OGT. The methodology involved utilizing substrates for hOGT (IaCLs in this case) as an effective way to form DNA-protein cross-links. It acted on the same lesions as hAGT, and unfortunately was unable to react with the O^4 position of thymine as determined by ESI-MS. While modeling and site-directed mutagenesis has revealed important insights into why some AGTs can remove damage from the O^4 position of thymine, the lack of structural information is crucial. It has been suggested from an evolutionary standpoint perhaps the fact since this damage is a minor lesion, human AGT adapted to deal with more prevalent, bulkier lesions at the O^6 position of dG.[151]

Chapter 3 examined if TWJs containing IaCLs lacking a phosphate backbone could self-assemble into a DNA tetrahedron. This work showed that hAGT was able to disassemble novel DNA tetrahedra with both butylene and heptylene linkages at the vertices. Unhooking was essentially independent of the alkylene chain length however appears to be sequence dependent based on the ^{32}P radiolabelling the vertices.

G-quadruplexes have gathered interested especially due to their presence in the genome as well as their importance of proper folding of aptamers. The *c-myc* promoter region G-quadruplex has been implicated in various cancers and alkyl damage to it could affect gene regulation. In **Chapter 4**, we show that alkyl lesions were thermally destabilizing and when located in the internal guanine position were more destabilizing than the external. Thermal denaturation appears to be independent of chain length for the monoadducts, and the IaCLs were similarly destabilizing to the sequences containing two sites of alkylation. Circular dichroism confirmed the presence of G-quadruplexes at physiological pH however when two sites of alkylation were present it appears a different species was forming. All lesions were repaired by AGTs variants with the *E. coli* AGTs repairing the small methyl lesions but not the bulkier hydroxyalkylene nor IaCLs as has been observed with duplex DNA. Longer heptylene hydroxyalkylene monoadducts and IaCLs were better substrates for hAGT and hOGT than the shorter butylene lesions.

Chapter 5 sought to examine the effect of methylation of thymine on the simpler four carbon threose nucleic acid in a DNA sequence. Unfortunately our assays to assess demethylation at the O^4 position of thymine in threose nucleic acids could not decipher if repair was occurring. The biophysical assays revealed that it was similarly destabilizing as the DNA counterpart when methylated, and no structural perturbation was observed by circular dichroism. The prebiotic world is poorly understood however important deductions have been made. Did AGTs exist at the same time as TNA, and did repair have an influence on the genomes of early organisms? Groups such as the Szotak and Chaput labs have examined thoroughly how proteins such as polymerases process TNA and it is plausible DNA repair proteins likely had a relevance to the maintenance of genetic information. Another reason why DNA is better suited as the respository of genetic information is due to more facile repair of lesions to restore structure and function relative to TNA.

6.2 Future work

This thesis has left the door open for several directions to expand upon. **Chapter 2** established an effective strategy for cross-linking the active site of OGT to DNA: A strategy employed to solve one of the crystal structures of hAGT with a DNA substrate. We are pursuing studies on crystallizing these DPCs currently to gain insight on why OGT is superior at repairing damage at the O^4 position of thymine relative to hAGT. To date modeling and site-directed mutagenesis has provided theories but no structural information is available for confirmation. Due to hOGT's ability to effectively repair a wide range of lesions (especially O^4-methyl dT), it has been proposed as a possible candidate for protein therapy as well.

We believe the DNA tetrahedra synthesized in **Chapter 3** could be used as a drug delivery vehicle. Traditionally our lab has studied butylene and heptylene cross-links as they are a product produced by therapeutically relevant cross-linking agents. In this study, the tetrahedron design contained IaCLs with alkylene linkages (tetra- and heptamethylene) that we had experience with. Future studies will involve examining various linker lengths to enhance the stability of the DNA tetrahedra. The **C4TWJ** and **C7TWJ** were well processed by hAGT at lower concentrations *in vitro*, perhaps it can be modulated to last longer in the cell for timed delivery with **TG** linkages as were discussed in **Chapter 2**. Preliminary results (see **Section 3.5.2 Figure S3.41**) suggest that the DNA tetrahedron is stable in fetal bovine serum (FBS) but must be repeated. Moieties such as hexaethylene glycol (HEG) could also be used to link the O^6 positions of the DNA tetrahedra as

they are commonly used to improve serum stability. Other preliminary research has suggested these TDNs can enter HeLa cells and we are interested in pursuing this work. These TDN are equipped with various 5' hydroxyl groups at the vertices which can be futher modified to increase functionality. Inspired by the DNA cube of Edwardson *et al.* discussed in **Chapter 1**, we are interested if i) D-DNA (**Figure 6.1**) is also compatible with these TDNs, ii) Can they house hydrophobic guest molecules such as dyes or drugs and, if so, iii) Can controlled release be

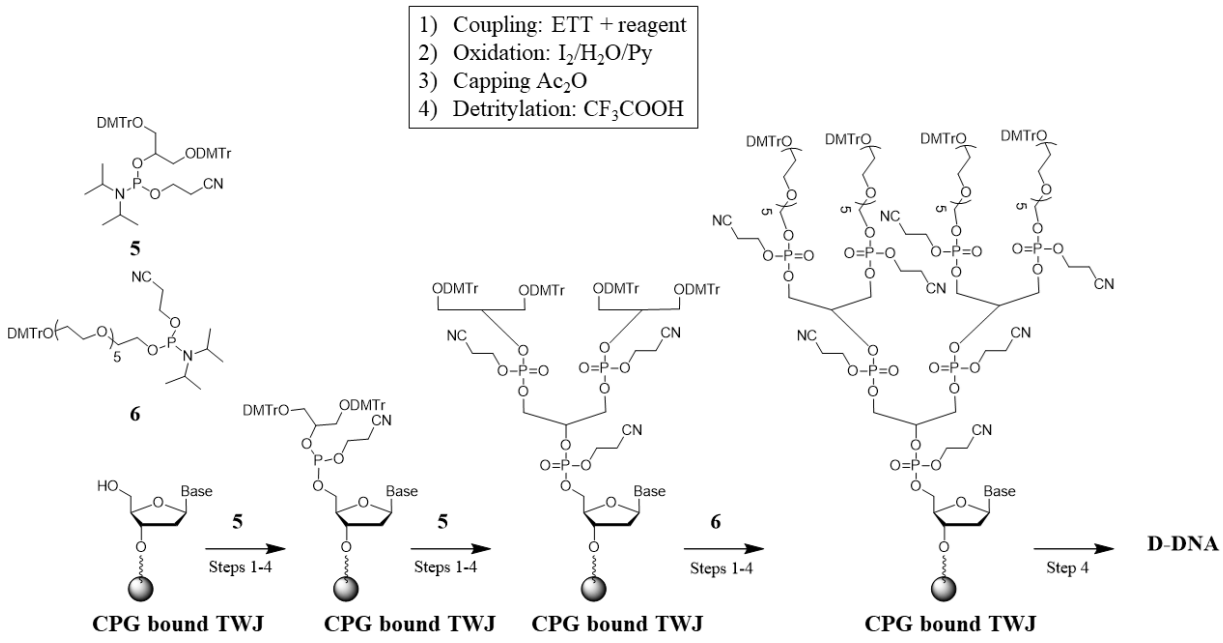


Figure 6.1 Synthetic scheme for attachment of dendritic DNA to the 5' hydroxyl groups of the **TWJ**s. A) Dendritic phosphoramidites for branching. B) Automated solid-phase synthesis of D-DNA scheme Adapted with permission from ref [241].

achieved in the presence of hAGT. Functionalization such as tethering siRNA for gene knockdown and aptamers such as AS1411 for targeting cancer cells should also be examined. Lastly, we are interested in elucidating structural information on the TDNs as the native PAGE and mung bean digests strongly confirm the presence of self-assembly. We are interested in determining detailed structural properties which will aid in future structural design. Atomic force microscopy and cryogenic electron microscopy are commonly used techniques and could be used to determine the structure of the IaCL TDNs.

We determined that AGTs are capable repairing methyl and larger alkyl lesions in the *c-myc* promoter region in **Chapter 4**. The kinetics of repair has also not been determined thus far and time-course assays could determine the difference in rates between proteins as well as lesions. As the G-quadruplexes with two sites of alkylation appear to be adopting a different conformation, structure determination by NMR spectroscopy could be used to confirm this. While the biology of how these physiological relevant lesions is of interest due to oncogene regulation, we saw promise for other applications due to the UV thermal denaturation, CD and repair experiments. Alkylation at two sites would be interesting to explore if it affected aptamers such AS1411 to bind their substrates and if incubated with AGTs their activity restored with applications as a conformational switch in nanotechnology. On the converse, could those containing IaCLs lacking a phosphate backbone fold into a functional aptamer and their function eliminated upon encountering AGT. Additionally, the presence of a restriction enzyme cut site in an aptamer would assist with the difficulties experienced in resolving repair products (no such site was found in the *c-myc* promoter region G-quadruplex).

5′-d(GGTGGTGGTGGTTGTGGTGGTGGTGG)

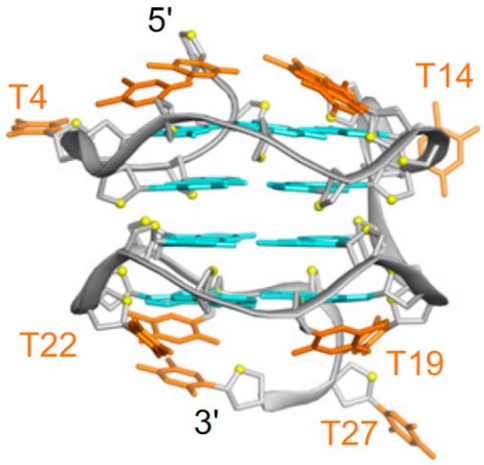


Figure 6.2 Crystal structure of AS1411. There are numerous consecutive guanines to explore the effect of IaCLs. Adapted with permission from ref [242].

Unfortunately repair of methylation at the O^4 position of thymine in TNA was undetermined due to assay development. ESI-MS is a technique that would give insight if repair was occurring. Snake venom phosphodiesterase digests and analysis by reverse phase HPLC could

allow for easier resolution of repair products by looking at the nucleosides. While methylation is believed to be the most prominent lesion now, larger alkyl damage could have been prominent at the dawn of life. Synthesis of TNA with larger alkyl lesions is of interest and furthermore, their resolution by gel electrophoresis is more achievable. We are also interested of the biophysical effects and repair of O^6 methylation of guanine in TNA (**Figure 6.3**). TNA has found applications in aptamers and modulating the stability upon an external stimulus (AGT) would be an interesting avenue to pursue.

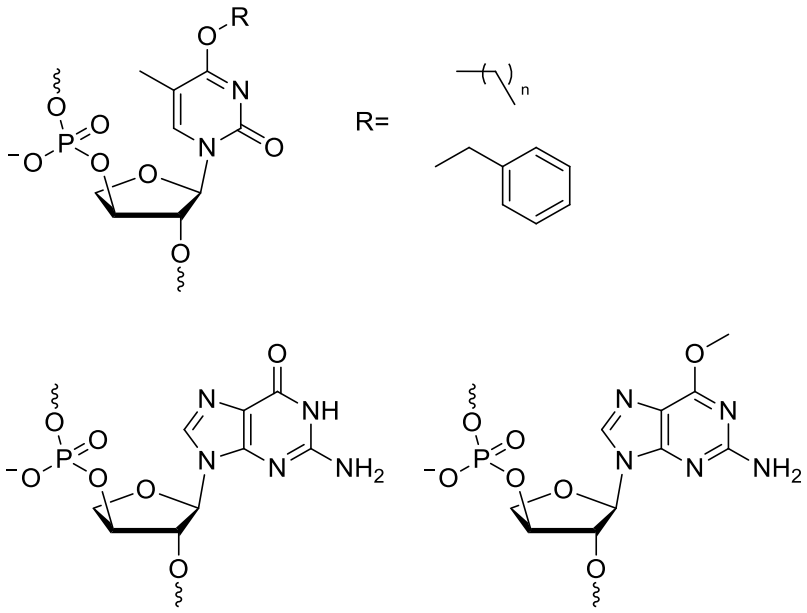


Figure 6.3 Modifications to tT and TNA-G to be explored.

References

[1] J. Watson, F. H. F. Crick, *Nature* **1953**, *171*, 737–8.

[2] B. Liu, Q. Xue, Y. Tang, J. Cao, F. P. Guengerich, H. Zhang, *Mutat. Res. - Rev. Mutat. Res.* **2016**, *768*, 53–67.

[3] L. R. Barrows, P. N. Magee, *Carcinogenesis* **1982**, *3*, 349–351.

[4] Y. Mishina, E. M. Duguid, C. He, *Chem. Rev.* **2006**, *106*, 215–232.

[5] W. P. Roos, A. D. Thomas, B. Kaina, *Nat. Rev. Cancer* **2016**, *16*, 20–33.

[6] J. A. Hickman, *Cancer Metastasis Rev.* **1992**, *11*, 121–139.

[7] M. Colvin, in *Holland-Frei Cancer Med.* (Eds.: D.W. Kufe, R.E. Pollock, R.R. Weichselbaum, R.C. Bast, T.S. Gansler, J.F. Holland, E. Frei III), Bc Decker, Hamilton, **2003**.

[8] P. J. McHugh, V. J. Spanswick, J. A. Hartley, *Lancet Oncol.* **2001**, *2*, 483–490.

[9] G. L. S, M. M. Wintrobe, W. Dameshek, M. J. Goodman, A. Gilman, M. T. McLennan, *J. Am. Med. Assoc.* **1946**, *132*, 126–132.

[10] D. A. Galton, M. Till, *Lancet* **1955**, *268*, 425–430.

[11] D. Y. Pacheco, N. K. Stratton, N. W. Gibson, *Cancer Res.* **1989**, *49*, 5108–5110.

[12] W. P. Tong, D. B. Ludlum, *Biochim. Biophys. Act - Nucleic Acids Protein Synth.* **1980**, *608*, 174–181.

[13] G. P. Margison, M. F. Santibáñez-Koref, *BioEssays* **2002**, *24*, 255–266.

[14] J. Y. Choi, G. Chowdhury, H. Zang, K. C. Angel, C. C. Vu, L. A. Peterson, F. P. Guengerich, *J. Biol. Chem.* **2006**, *281*, 38244–38256.

[15] 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.

[16] Y. Lin, T. Zhao, X. Jian, Z. Farooqui, X. Qu, C. He, A. R. Dinner, N. F. Scherer, *Biophys. J.* **2009**, *96*, 1911–1917.

[17] G. Dodson, A. Wlodawer, *Trends Biochem. Sci.* **1998**, *23*, 347–352.

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

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

[20] M. Melikishvili, J. J. Rasimas, A. E. Pegg, M. G. Fried, *Biochemistry* **2008**, *47*, 13754–

13763.

[21] M. Christmann, B. Verbeek, W. P. Roos, B. Kaina, *Biochim. Biophys. Acta* **2011**, *1816*, 179–190.

[22] S. C. Schold, T. P. Brent, E. von Hofe, H. S. Friedman, S. Mitra, D. D. Bigner, J. A. Swenberg, P. Kleihues, *J. Neurosurg.* **1989**, *70*, 573–577.

[23] L. Liu, S. L. Gerson, *Clin. Cancer Res.* **2006**, *12*, 328–331.

[24] M. E. Dolan, R. C. Moschel, A. E. Pegg, *Proc. Natl. Acad. Sci. U. S. A.* **1990**, *87*, 5368–5372.

[25] D. T. Blumenthal, C. Rankin, K. J. Stelzer, A. M. Spence, A. E. Sloan, D. F. Moore, G. D. A. Padula, S. B. Schulman, M. L. Wade, E. J. Rushing, *Int. J. Clin. Oncol.* **2015**, *20*, 650–658.

[26] H. A. Tawbi, L. Villaruz, A. Tarhini, S. Moschos, M. Sulecki, F. Viverette, J. Shipe-Spotloe, R. Radkowski, J. M. Kirkwood, *Br. J. Cancer* **2011**, *105*, 773–777.

[27] W. Yu, L. Zhang, Q. Wei, A. Shao, *Front. Oncol.* **2020**, *9*, 1547.

[28] J. Hall, H. Bresil, R. Montesano, *Carcinogenesis* **1985**, *6*, 208–211.

[29] B. Sedgwick, T. Lindhal, *J. Mol. Biol.* **1982**, *154*, 169–175.

[30] K. Morimoto, M. E. Dolan, D. Scicchitano, A. E. Pegg, *Carcinogenesis* **1985**, *6*, 1027–1031.

[31] K. Goodtzova, S. Kanugula, S. Edara, G. T. Pauly, R. C. Moschel, A. E. Pegg, *J. Biol. Chem.* **1997**, *272*, 8332–8339.

[32] R. S. Mijal, N. M. Thomson, R. C. Moschel, S. Kanugula, Q. Fang, A. E. Pegg, L. A. Peterson, *Chem. Res. Toxicol.* **2004**, *17*, 424–434.

[33] D. M. Noll, N. D. Clarke, *Nucleic Acids Res.* **2001**, *29*, 4025–4034.

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

[35] 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**, *47*, 10892–10903.

[36] 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.

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

[38] F. P. Mcmanus, A. Khaira, A. M. Noronha, C. J. Wilds, *Bioconjug. Chem.* **2013**, *24*, 224–

233.

[39] D. K. O'Flaherty, C. J. Wilds, *Chem. - An Asian J.* **2016**, *11*, 576–583.

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

[41] D. K. O'Flaherty, C. J. Wilds, *ChemBioChem* **2017**, *18*, 2351–2357.

[42] D. K. O'Flaherty, C. J. Wilds, *Org. Biomol. Chem.* **2017**, *15*, 189–196.

[43] I. Teo, B. Sedgwick, B. Demple, B. Li, T. Lindahl, *EMBO J* **1984**, *3*, 2151–2157.

[44] T. P. Brent, M. E. Dolan, H. Fraenkel-Conrat, J. Hall, P. Karran, F. Laval, G. P. Margison, R. Montesano, A. E. Pegg, P. M. Potter, B. Singer, J. A. Swenberg, D. B. Yarosh, *Proc. Natl. Acad. Sci. U. S. A.* **1988**, *85*, 1759–1762.

[45] P. M. Potter, M. C. Wilkinson, J. Fitton, F. J. Carr, J. Brennand, D. P. Cooper, G. P. Margison, *Nucleic Acids Res.* **1987**, *15*, 9177–9193.

[46] D. S. Daniels, J. A. Tainer, *Mutat. Res. - DNA Repair* **2000**, *460*, 151–163.

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

[48] R. H. Elder, G. P. Margison, J. A. Rafferty, *Biochem. J.* **1994**, *298*, 231–235.

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

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

[51] F. P. McManus, C. J. Wilds, *Toxicol. Res. (Camb).* **2013**, *2*, 158–162.

[52] G. Crivat, J. W. Taraska, *Trends Biotechnol.* **2012**, *30*, 8–16.

[53] A. Keppler, S. Gendreizig, T. Gronemeyer, H. Pick, H. Vogel, K. Johnsson, *Nat. Biotechnol.* **2003**, *21*, 86–89.

[54] M. Tintoré, I. Gállego, B. Manning, R. Eritja, C. Fàbrega, *Angew. Chemie - Int. Ed.* **2013**, *52*, 7747–7750.

[55] M. Tintoré, S. Grijalvo, R. Eritja, C. Fàbrega, *Bioorganic Med. Chem. Lett.* **2015**, *25*, 5208–5211.

[56] N. R. Kallenback, R.-I. Ma, N. C. Seeman, *Nature* **1983**, *305*, 829–831.

[57] Y. Wang, J. E. Mueller, B. Kemper, N. C. Seeman, *Biochemistry* **1991**, *30*, 5667–5674.

[58] X. Wang, N. C. Seeman, *J. Am. Chem. Soc.* **2007**, *129*, 8169–8176.

[59] A. R. Chandrasekaran, R. Zhuo, *Appl. Mater. Today* **2016**, *2*, 7–16.

[60] P. W. K. Rothemund, *Nature* **2006**, *440*, 297–302.

[61] J. Chen, N. C. Seeman, *Nature* **1991**, *350*, 631–633.

[62] R. P. Goodman, I. A. T. Schaap, C. F. Tardin, C. M. Erben, R. M. Berry, C. F. Schmidt, A. J. Turberfield, *Science* **2005**, *310*, 1661–1666.

[63] W. M. Shih, J. D. Quispe, G. F. A. Joyce, *Nature* **2004**, *427*, 618–621.

[64] D. Bhatia, S. Mehtab, R. Krishnan, S. S. Indi, A. Basu, Y. Krishnan, *Angew. Chemie Int. Ed.* **2009**, *48*, 4134–4137.

[65] Y. He, T. Ye, M. Su, C. Zhang, A. E. Ribbe, W. Jiang, C. Mao, *Nature* **2008**, *452*, 198–201.

[66] W. Wang, S. Chen, B. An, K. Huang, T. Bai, M. Xu, G. Bellot, Y. Ke, Y. Xiang, B. Wei, *Nat. Commun.* **2019**, *10*, 1067.

[67] K. E. Bujold, J. C. C. Hsu, H. F. Sleiman, *J. Am. Chem. Soc.* **2016**, *138*, 14030–14038.

[68] A. Idili, A. Vallée-Bélisle, F. Ricci, *J. Am. Chem. Soc.* **2014**, *136*, 5836–5839.

[69] T. Liedl, M. Olapinski, F. C. Simmel, *Angew. Chem. Int. Ed. Engl.* **2006**, *45*, 5007–10.

[70] N. Safaee, A. M. Noronha, D. Rodionov, G. Kozlov, C. J. Wilds, G. M. Sheldrick, K. Gehring, *Angew. Chem. Int. Ed. Engl.* **2013**, *52*, 10370–10373.

[71] C. Mao, W. Sun, Z. Shen, N. C. Seeman, *Nature* **1999**, *397*, 144–146.

[72] B. Yurke, A. J. Turberfield, A. P. Mills, F. C. Simmel, J. L. Neumann, *Nature* **2000**, *406*, 605–608.

[73] C. M. Erben, R. P. Goodman, A. J. Turberfield, *Angew. Chemie Int. Ed.* **2006**, *45*, 7414–7417.

[74] C. Zhang, X. Li, C. Tian, G. Yu, Y. Li, W. Jiang, C. Mao, *ACS Nano* **2014**, *8*, 1130–1135.

[75] T. G. W. Edwardson, K. M. M. Carneiro, C. K. McLaughlin, C. J. Serpell, H. F. Sleiman, *Nat. Chem.* **2013**, *5*, 868–875.

[76] A. Banerjee, D. Bhatia, A. Saminathan, S. Chakraborty, S. Kar, Y. Krishnan, *Angew. Chemie - Int. Ed.* **2013**, *52*, 6854–6857.

[77] Z. Liu, Y. Li, C. Tian, C. Mao, *Biomacromolecules* **2013**, *14*, 1711–1714.

[78] A. T. Veetil, K. Chakraborty, K. Xiao, M. R. Minter, S. S. Sisodia, Y. Krishnan, *Nat. Nanotechnol.* **2017**, *12*, 1183–1189.

[79] J. W. Conway, C. K. Mc Laughlin, K. J. Castor, H. Sleiman, *Chem. Commun.* **2013**, *49*, 1172–1174.

[80] L. He, D. Q. Lu, H. Liang, S. Xie, C. Luo, M. Hu, L. Xu, X. Zhang, W. Tan, *ACS Nano*

2017, *11*, 4060–4066.

[81] R. Abbotts, S. Madhusudan, *Cancer Treat. Rev.* **2010**, *36*, 425–435.

[82] Y. Zhang, Y. Deng, C. Wang, L. Li, L. Xu, Y. Yu, X. Su, *Chem. Sci.* **2019**, *10*, 5959–5966.

[83] S. Modi, S. M G, D. Goswami, G. D. Gupta, S. Mayor, Y. Krishnan, *Nat. Nanotechnol.* **2009**, *4*, 325–330.

[84] D. Bhatia, S. Surana, S. Chakraborty, S. P. Koushika, Y. Krishnan, *Nat. Commun.* **2011**, *2*, 339.

[85] M. M. Javadpour, M. M. Juban, W. C. J. Lo, S. M. Bishop, J. B. Alberty, S. M. Cowell, C. L. Becker, M. L. McLaughlin, *J. Med. Chem.* **1996**, *39*, 3107–3113.

[86] J. Yan, J. Chen, N. Zhang, Y. Yang, W. Zhu, L. Li, B. He, *J. Mater. Chem. B* **2020**, *8*, 492–503.

[87] Y. Liu, Y. Sun, S. Li, M. Liu, X. Qin, X. Chen, Y. Lin, *Nano Lett.* **2020**, *20*, 3602–3610.

[88] P. C. Zamecnik, M. L. Stephenson, *Proc. Natl. Acad. Sci. U. S. A.* **1978**, *75*, 280–284.

[89] A. Fire, S. Xu, M. K. Montgomery, S. A. Kostas, S. E. Driver, C. C. Mello, *Nature* **1998**, *391*, 806–811.

[90] H. Lee, A. K. R. Lytton-Jean, Y. Chen, K. T. Love, A. I. Park, E. D. Karagiannis, A. Sehgal, W. Querbes, C. S. Zurenko, M. Jayaraman, C. G. Peng, K. Charisse, A. Borodovsky, M. Manoharan, J. S. Donahoe, J. Truelove, M. Nahrendorf, R. Langer, D. G. Anderson, *Nat. Nanotechnol.* **2012**, *7*, 389–393.

[91] J. Yang, Q. Jiang, L. He, P. Zhan, Q. Liu, S. Liu, M. Fu, J. Liu, C. Li, B. Ding, *ACS Appl. Mater. Interfaces* **2018**, *10*, 23693–23699.

[92] T. Wu, Q. Liu, Y. Cao, R. Tian, J. Liu, B. Ding, *ACS Appl. Mater. Interfaces* **2020**, *in press*, 32461–32467.

[93] J. D. Ly, D. R. Grubb, A. Lawen, *Apoptosis* **2003**, *8*, 115–128.

[94] F. Mongelard, P. Bouvet, *Trends Cell Biol.* **2007**, *17*, 80–86.

[95] K.-R. Kim, H. Jegal, J. Kim, D. R. Ahn, *Biomater. Sci.* **2020**, *8*, 586–590.

[96] H. Zhang, G. S. Demirer, H. Zhang, T. Ye, N. S. Goh, A. J. Aditham, F. J. Cunningham, C. Fan, M. P. Landry, *Proc. Natl. Acad. Sci. U. S. A.* **2019**, *116*, 7543–7548.

[97] L. Good, R. Sandberg, O. Larsson, P. E. Nielsen, C. Wahlestedt, *Microbiology* **2000**, *146*, 2665–2670.

[98] Y. Hu, Z. Chen, X. Mao, M. Li, Z. Hou, J. Meng, X. Luo, X. Xue, *J. Nanobiotechnology* **2020**, *18*, 109.

[99] R. C. Friedman, K. K. H. Farh, C. B. Burge, D. P. Bartel, *Genome Res.* **2009**, *19*, 92–105.

[100] S. Li, Y. Sun, T. Tian, X. Qin, S. Lin, T. Zhang, Q. Zhang, M. Zhou, X. Zhang, Y. Zhou, H. Zhao, B. Zhu, X. Cai, *Cell Prolif.* **2020**, *15*, DOI: 10.1186/s11671-020-03377-y.

[101] P. Phatak, K. A. Byrnes, D. Mansour, L. Liu, S. Cao, R. Li, J. N. Rao, D. J. Turner, J. Y. Wang, J. M. Donahue, *Oncogene* **2016**, *35*, 2087–2097.

[102] E. Latz, A. Verma, A. Visintin, M. Gong, C. M. Sirois, D. C. G. Klein, B. G. Monks, J. C. McKnight, M. S. Lamphier, P. W. Duprex, T. Espevik, D. T. Golenbock, *Nat. Immunol.* **2007**, *8*, 772–779.

[103] J. Li, H. Pei, B. Zhu, L. Liang, M. Wei, Y. He, N. Chen, D. Li, Q. Huang, C. Fan, *ACS Nano* **2011**, *5*, 8783–8789.

[104] S. Ohtsuki, N. Matsuzaki, K. Mohri, M. Endo, T. Emura, K. Hidaka, H. Sugiyama, Y. Takahashi, K. Ishiyama, N. Kadowaki, Y. Takakura, M. Nishikawa, *Nucleic Acid Ther.* **2015**, *25*, 245–253.

[105] J. Sudimack, R. J. Lee, *Adv. Drug Deliv. Rev.* **2000**, *41*, 147–162.

[106] Y. Z. Bu, J. R. Xu, Q. Luo, M. Chen, L. M. Mu, W. L. Lu, *Nanomaterials* **2020**, *10*, DOI DOI 10.3390/nano10050951.

[107] J. Zhang, Y. Guo, F. Ding, G. Pan, X. Zhu, C. Zhang, *Angew. Chemie - Int. Ed.* **2019**, *58*, 13794–13798.

[108] J. D. Brodin, A. J. Sprangers, J. R. McMillan, C. A. Mirkin, *J. Am. Chem. Soc.* **2015**, *137*, 14838–14841.

[109] D. Li, X. Li, F. Yang, R. Yuan, Y. Xiang, *ACS Appl. Mater. Interfaces* **2020**, *12*, 54489–54496.

[110] J. W. Keum, H. Bermudez, *Chem. Commun.* **2009**, 7036–7038.

[111] C. M. Green, D. Mathur, I. L. Medintz, *J. Mater. Chem. B* **2020**, *8*, 6170–6178.

[112] A. Lacroix, E. Vengut-Climent, D. De Rochambeau, H. F. Sleiman, *ACS Cent. Sci.* **2019**, *5*, 882–891.

[113] H. Bila, E. E. Kurisinkal, M. M. C. Bastings, *Biomater. Sci.* **2019**, *7*, 532–541.

[114] K. R. Kim, T. Lee, B. S. Kim, D. R. Ahn, *Chem. Sci.* **2014**, *5*, 1533–1537.

[115] S. Raniolo, G. Vindigni, V. Unida, A. Ottaviani, E. Romano, A. Desideri, S. Biocca,

Nanoscale **2018**, *10*, 12078–12086.

[116] Y. He, Y. Ye, M. Su, C. Zhang, R. A. E, W. Jiang, C. Mao, *Nature* **2008**, *452*, 198–201.

[117] W. Saenger, *Principles of Nucleic Acid Structure*, Springer-Verlag, Berlin, **1984**.

[118] A. Bacolla, G. Wang, K. M. Vasquez, *PLoS Genet.* **2015**, *11*, e1005696.

[119] K. Gehring, J.-L. Leroy, M. Gueron, *Nature* **1993**, *363*, 561–565.

[120] S. Burge, G. N. Parkinson, P. Hazel, A. K. Todd, S. Neidle, *Nucleic Acids Res.* **2006**, *34*, 5402–15.

[121] R. K. Moyzis, J. M. Buckingham, L. S. Cram, M. Dani, L. L. Deaven, M. D. Jones, J. Meyne, R. L. Ratliff, J. R. Wu, *Proc. Natl. Acad. Sci.* **1988**, *85*, 6622–6626.

[122] J. Shay, S. Bacchetti, *Eur. J. Cancer* **1997**, *33*, 787–791.

[123] A. Ambrus, D. Chen, J. Dai, T. Bialis, R. A. Jones, D. Yang, *Nucleic Acids Res.* **2006**, *34*, 2723–2735.

[124] M. Kim, H. Vankayalapati, K. Shin-ya, K. Wierzba, L. H. Hurley, *J. Am. Chem. Soc.* **2002**, *124*, 2098–2099.

[125] M. Read, R. J. Harrison, B. Romagnoli, F. A. Tanious, S. H. Gowan, A. P. Reszka, W. D. Wilson, L. R. Kelland, S. Neidle, *Proc. Natl. Acad. Sci. U. S. A.* **2001**, *98*, 4844–4849.

[126] A. J. Zaug, E. R. Podell, T. R. Cech, *Proc. Natl. Acad. Sci. U. S. A.* **2005**, *102*, 10864–10869.

[127] S. Balasubramanian, L. H. Hurley, S. Neidle, *Nat Rev Drug Discov* **2011**, *10*, 261–275.

[128] T. A. Brooks, S. Kendrick, L. Hurley, *FEBS J.* **2010**, *277*, 3459–3469.

[129] D. Rhodes, H. J. Lipps, *Nucleic Acids Res.* **2015**, *43*, 8627–8637.

[130] L. M. Facchini, L. Z. Penn, *FASEB J.* **1998**, *12*, 633–651.

[131] A. Siddiqui-Jain, C. L. Grand, D. J. Bearss, L. H. Hurley, *Proc. Natl. Acad. Sci. U. S. A.* **2002**, *99*, 11593–11598.

[132] M. Bejugam, S. Sewitz, P. S. Shirude, R. Rodriguez, R. Shahid, S. Balasubramanian, **n.d.**

[133] P. Wang, C.-H. Leung, D.-L. Ma, S.-C. Yan, C.-M. Che, *Chem. A Eur. J.* **2010**, *16*, 6900–6911.

[134] M. A. Gregory, S. R. Hann, *Mol. Cell. Biol.* **2000**, *20*, 2423–2435.

[135] J. L. Huppert, S. Balasubramanian, *Nucleic Acids Res.* **2005**, *33*, 2908–2916.

[136] L. M. Hellman, T. J. Spear, C. J. Koontz, M. Melikishvili, M. G. Fried, *Nucleic Acids Res.* **2014**, *42*, 9781–9791.

[137] A. M. Fleming, J. Zhou, S. S. Wallace, C. J. Burrows, *ACS Cent. Sci.* **2015**, *1*, 226–233.

[138] T. Chang, C. Qi, J. Meng, N. Zhang, T. Bing, X. Yang, Z. Cao, D. Shangguan, *PLoS One* **2013**, *8*, 1–10.

[139] A. Avino, C. Fabrega, M. Tintore, R. Eritja, *Curr. Pharm. Des.* **2012**, *18*, 2036–2047.

[140] Y. Cheng, G. Zhao, S. Zhang, F. Nigim, G. Zhou, Z. Yu, Y. Song, Y. Chen, Y. Li, *PLoS One* **2016**, *11*, 1–20.

[141] N. Jing, C. Marchand, J. Liu, R. Mitra, M. E. Hogan, Y. Pommier, *J. Biol. Chem.* **2000**, *275*, 21460–21467.

[142] S. Srinivasan, V. Ranganathan, M. C. DeRosa, B. M. Murari, *Anal. Bioanal. Chem.* **2019**, *411*, 1319–1330.

[143] A. Bahreyni, M. Ramezani, M. Alibolandi, P. Hassanzadeh, K. Abnous, S. M. Taghdisi, *Anal. Biochem.* **2019**, *575*, 1–9.

[144] K. T. Shum, E. L. H. Lui, S. C. K. Wong, P. Yeung, L. Sam, Y. Wang, R. M. Watt, J. A. Tanner, *Biochemistry* **2011**, *50*, 3261–3271.

[145] A. Vassilev, M. L. DePamphilis, *Genes (Basel).* **2017**, *8*, 45.

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

[147] W. P. Roos, B. Kaina, *Cancer Lett.* **2013**, *332*, 237–248.

[148] D. P. Berger, B. R. Winterhalter, W. A. Dengler, H. H. Fiebig, *Anticancer Drugs* **1992**, *3*, 531–539.

[149] S. R. Rajski, R. M. Williams, *Chem. Rev.* **1998**, *98*, 2723–2795.

[150] C. Yi, C. He, *Cold Spring Harb. Perspect. Biol.* **2013**, *5*, a012575.

[151] A. E. Pegg, *Chem. Res. Toxicol.* **2011**, *24*, 618–639.

[152] T. Lindahl, B. Demple, P. Robins, *EMBO J.* **1982**, *1*, 1359–1363.

[153] G. Perugino, A. Vettone, G. Illiano, A. Valenti, M. C. Ferrara, M. Rossi, M. Ciaramella, *J. Biol. Chem.* **2012**, *287*, 4222–4231.

[154] J. Hu, A. Ma, A. R. Dinner, *Proc. Natl. Acad. Sci.* **2008**, *105*, 4615–4620.

[155] M. Sassanfar, M. K. Dosanjh, J. M. Essigmann, L. Samson, *J. Biol. Chem.* **1991**, *266*, 2767–2771.

[156] Q. Fang, S. Kanugula, J. L. Tubbs, J. A. Tainer, A. E. Pegg, *J. Biol. Chem.* **2010**, *285*, 8185–8195.

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

[158] M. G. Fried, S. Kanugula, J. L. Bromberg, A. E. Pegg, *Biochemistry* **1996**, *35*, 15295–15301.

[159] N. M. Schoonhoven, D. K. O'Flaherty, F. P. McManus, L. Sacre, A. M. Noronha, M. Judith Kornblatt, C. J. Wilds, *Molecules* **2017**, *22*, 1948.

[160] L. P. Encell, L. A. Loeb, *Biochemistry* **1999**, *38*, 12097–12103.

[161] R. Loeber, M. Rajesh, Q. Fang, A. E. Pegg, N. Tretyakova, *Chem. Res. Toxicol.* **2006**, *19*, 645–654.

[162] R. Loeber, M. Rajesh, Q. Fang, A. E. Pegg, *Chem. Res. Toxicol.* **2008**, *21*, 787–795.

[163] L. Liu, D. L. Hachey, G. Valadez, K. M. Williams, F. P. Guengerich, N. A. Loktionova, S. Kanugula, A. E. Pegg, *J. Biol. Chem.* **2004**, *279*, 4250–4259.

[164] G. Chowdhury, S.-H. Cho, A. E. Pegg, F. P. Guengerich, *Angew. Chemie - Int. Ed.* **2013**, *52*, 12879–12882.

[165] T. Aoki, D. Wolle, E. P.-B. Noon, Q. Dai, E. C. Lai, P. Schedl, *Fly (Austin).* **2014**, *8*, 43–51.

[166] E. A. Hoffman, B. L. Frey, L. M. Smith, D. T. Auble, *J. Biol. Chem.* **2015**, *290*, 26404–26411.

[167] O. I. Wilner, Y. Weizmann, R. Gill, O. Lioubashevski, R. Freeman, I. Willner, *Nat. Nanotechnol.* **2009**, *4*, 249–254.

[168] R. L. Loeber, E. D. Michaelson-Richie, S. G. Codreanu, D. C. Liebler, C. R. Campbell, N. Y. Tretyakova, *Chem. Res. Toxicol.* **2009**, *22*, 1151–1162.

[169] T. S. Furey, *Nat. Rev. Genet.* **2012**, *13*, 840–852.

[170] N. Y. Tretyakova, A. Groehler, S. Ji, *Acc. Chem. Res.* **2015**, *48*, 1631–1644.

[171] G. L. Verdine, D. P. G. Norman, *Annu. Rev. Biochem.* **2003**, *72*, 337–366.

[172] N. Foloppe, B. Hartmann, L. Nilsson, A. D. MacKerell, *Biophys. J.* **2002**, *82*, 1554–1569.

[173] M. Xu-Welliver, A. E. Pegg, *Carcinogenesis* **2002**, *23*, 823–830.

[174] Z. Li, B. Wei, J. Nangreave, C. Lin, Y. Liu, Y. Mi, H. Yan, *J. Am. Chem. Soc.* **2009**, *131*, 13093–13098.

[175] Y. Zhang, N. C. Seeman, *J. Am. Chem. Soc.* **1994**, *116*, 1661–1669.

[176] F. F. Andersen, B. Knudsen, C. L. P. Oliveira, R. F. Frøhlich, D. Krüger, J. Bungert, M. Agbandje-McKenna, R. Mckenna, S. Juul, C. Veigaard, J. Koch, J. L. Rubinstein, B. Guldbrandtsen, M. S. Hede, G. Karlsson, A. H. Andersen, J. S. Pedersen, B. R. Knudsen,

Nucleic Acids Res. **2008**, *36*, 1113–1119.

[177] Y. He, M. Su, P. A. Fang, C. Zhang, A. E. Ribbe, W. Jiang, C. Mao, *Angew. Chemie - Int. Ed.* **2010**, *49*, 748–751.

[178] S. Juul, F. Iacovelli, M. Falconi, S. L. Kragh, B. Christensen, R. Frøhlich, O. Franch, E. L. Kristoffersen, M. Stougaard, K. W. Leong, Y. Ho, E. S. Sørensen, V. Birkedal, A. Desideri, B. R. Knudsen, *ACS Nano* **2013**, *7*, 9724–9734.

[179] P. K. Lo, P. Karam, F. A. Aldaye, C. K. McLaughlin, G. D. Hamblin, G. Cosa, H. F. Sleiman, *Nat. Chem.* **2010**, *2*, 319–328.

[180] A. S. Walsh, H. Yin, C. M. Erben, M. J. A. Wood, A. J. Turberfield, *ACS Nano* **2011**, *5*, 5427–5432.

[181] C. Panagiotidis, S. Kath-Schorr, G. Von Kiedrowski, *ChemBioChem* **2016**, *17*, 254–259.

[182] S. N. Mikhailov, E. V. Efimtseva, A. A. Rodionov, A. A. Shelkunova, J. Rozenski, G. Emmerechts, G. Schepers, A. Van Aerschot, P. Herdewijn, *Chem. Biodivers.* **2005**, *2*, 1153–1163.

[183] J. J. Rasimas, A. E. Pegg, M. G. Fried, *J. Biol. Chem.* **2003**, *278*, 7973–7980.

[184] J. J. Rasimas, P. A. Dalessio, I. J. Ropson, A. E. Pegg, M. G. Fried, *Protein Sci.* **2004**, *13*, 301–305.

[185] 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.

[186] R. Guza, A. E. Pegg, N. Tretyakova, *Effects of Sequence Context on O6-Alkylguanine DNA Alkyltransferase Repair of O6-Alkyl-Deoxyguanosine Adducts*, ACS Symposium Series, Vanderbilt University, Nashville, TN, **2010**.

[187] S. Mitra, B. Kaina, *Prog. Nucleic Acid Res. Mol. Biol.* **1993**, *44*, 109–142.

[188] B. Saccà, C. M. Niemeyer, *Chem. Soc. Rev.* **2011**, *40*, 5910–5921.

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

[190] S. L. Gerson, *J. Clin. Oncol.* **2002**, *20*, 2388–2399.

[191] M. D. Frank-Kamenetskii, S. M. Mirkin, *Annu. Rev. Biochem.* **1995**, *64*, 65–95.

[192] N. B. Ramsing, T. M. Jovin, *Nucleic Acids Res.* **1988**, *16*, 6659–6676.

[193] C. Lin, D. Yang, *Methods Mol. Biol.* **2017**, *1587*, 171–196.

[194] D. Yang, L. Hurley, *Nucleosides, Nucleotides and Nucleic Acids* **2006**, *25*, 951–968.

[195] S. J. Berberich, E. J. Postel, *Oncogene* **1995**, *10*, 2343–2347.

[196] T. Simonsson, P. Pecinka, M. Kubista, *Nucleic Acids Res.* **1998**, *26*, 1167–1172.

[197] A. T. Phan, Y. S. Modi, D. J. Patel, *J. Am. Chem. Soc.* **2004**, *126*, 8710–8716.

[198] R. W. Harkness, A. K. Mittermaier, *Nucleic Acids Res.* **2016**, *44*, 3481–3494.

[199] C. S. Mekmaysy, L. Petraccone, N. C. Garbett, P. A. Ragazzon, R. Gray, J. O. Trent, J. B. Chaires, J. Graham, *J. Am. Chem. Soc.* **2008**, *130*, 6710–6711.

[200] M. Zacchia, M. L. Abategiovanni, S. Stratigis, G. Capasso, *Kidney Dis.* **2016**, *2*, 72–79.

[201] J. L. Mergny, L. Lacroix, *Curr. Protoc. Nucleic Acid Chem.* **2009**, *37*, 17.1.1-17.1.15.

[202] J. L. Mergny, J. Li, L. Lacroix, S. Amrane, J. B. Chaires, *Nucleic Acids Res.* **2005**, *33*, e158.

[203] A. I. Karsisiotis, N. M. A. Hessari, E. Novellino, G. P. Spada, A. Randazzo, M. Webba Da Silva, *Angew. Chemie - Int. Ed.* **2011**, *50*, 10645–10648.

[204] M. Kaushik, S. Kaushik, A. Bansal, S. Saxena, S. Kukreti, *Curr. Mol. Med.* **2011**, *11*, 744–769.

[205] R. I. Mathad, E. Hatzakis, J. Dai, D. Yang, *Nucleic Acids Res.* **2011**, *39*, 9023–9033.

[206] T. Miura, J. M. Benevides, G. J. Thomas, *J. Mol. Biol.* **1995**, *248*, 233–238.

[207] M. Kaushik, A. Bansal, S. Saxena, S. Kukreti, *Biochemistry* **2007**, *46*, 7119–7131.

[208] R. del Villar-Guerra, J. O. Trent, J. B. Chaires, *Angew. Chemie - Int. Ed.* **2018**, *57*, 7171–7175.

[209] W. Copp, C. J. Wilds, *Chem. - A Eur. J.* **2020**, *26*, 14802–14806.

[210] A. D. Ellington, J. W. Szostak, *Nature* **1990**, *346*, 818–822.

[211] C. Roxo, W. Kotkowiak, A. Pasternak, *Molecules* **2019**, *24*, 3781.

[212] C. J. Lech, Z. Li, B. Heddi, A. T. Phan, *Chem. Commun.* **2012**, *48*, 11425–11427.

[213] Z. Li, C. J. Lech, A. T. Phan, *Nucleic Acids Res.* **2014**, *42*, 4068–79.

[214] P. L. Thao Tran, A. Virgilio, V. Esposito, G. Citarella, J.-L. Mergny, A. Galeone, *Biochimie* **2011**, *93*, 399–408.

[215] L. Haase, B. Karg, K. Weisz, *ChemBioChem* **2019**, *20*, 985–993.

[216] W. Kotkowiak, J. Lisowiec-Wachnicka, J. Grynda, R. Kierzek, J. Wengel, A. Pasternak, *Mol. Ther. - Nucleic Acids* **2018**, *10*, 304–316.

[217] A. Pasternak, F. J. Hernandez, L. M. Rasmussen, B. Vester, J. Wengel, *Nucleic Acids Res.* **2011**, *39*, 1155–1164.

[218] B. Johnson, A. Johnson, J. Lewis, M. Raff, K. Roberts, P. Walter, *RNA World and the*

Origins of Life, Garland Science, New York, **2002**.

[219] M. P. Robertson, G. F. Joyce, *Chem. Biol.* **2014**, *21*, 238–245.

[220] O. Jungmann, H. Wippo, M. Stanek, H. K. Huynh, R. Krishnamurthy, A. Eschenmoser, *Org. Lett.* **1999**, *1*, 1527–1530.

[221] A. Eschenmoser, *Chemical Synthesis: Gnosis to Prognosis*, Kluwer Academic, Dordrecht, Netherlands, **1996**.

[222] A. Eschenmoser, *Science* **1999**, *284*, 1347–1351.

[223] C. J. Wilds, Z. Wawrzak, R. Krishnamurthy, A. Eschenmoser, M. Egli, *J. Am. Chem. Soc.* **2002**, *124*, 13716–13721.

[224] P. S. Pallan, C. J. Wilds, Z. Wawrzak, R. Krishnamurthy, A. Eschenmoser, M. Egli, *Angew. Chemie - Int. Ed.* **2003**, *42*, 5893–5895.

[225] A. Eichert, K. Behling, C. Betzel, V. A. Erdmann, J. P. Fürste, C. Förster, *Nucleic Acids Res.* **2010**, *38*, 6729–6736.

[226] J. I. Yeh, E. Pohl, D. Truan, W. He, G. M. Sheldrick, S. Du, C. Achim, *Chemistry (Easton).* **2010**, *16*, 11867–11875.

[227] A. T. Johnson, M. K. Schlegel, E. Meggers, L.-O. Essen, O. Wiest, *J. Org. Chem.* **2011**, *76*, 7964–7974.

[228] J. C. Chaput, J. K. Ichida, J. W. Szostak, *J. Am. Chem. Soc.* **2003**, *125*, 856–857.

[229] J. C. Chaput, J. W. Szostak, *J. Am. Chem. Soc.* **2003**, *125*, 9274–9275.

[230] V. Kempeneers, K. Vastmans, J. Rozenski, P. Herdewijn, *Nucleic Acids Res.* **2003**, *31*, 6221–6226.

[231] J. K. Ichida, A. Horhota, K. Zou, L. W. McLaughlin, J. W. Szostak, *Nucleic Acids Res.* **2005**, *33*, 5219–5225.

[232] B. O. Heuberger, C. Switzer, *Org. Lett.* **2006**, *8*, 5809–5811.

[233] H. Yu, S. Zhang, J. C. Chaput, *Nat. Chem.* **2012**, *4*, 183–187.

[234] P. F. Swann, *Mutat. Res. - Fundam. Mol. Mech. Mutagen.* **1990**, *233*, 81–94.

[235] M. W. Kalnik, M. Kouchakdjian, B. F. L. Li, P. F. Swann, D. J. Patel, *Biochemistry* **1988**, *27*, 108–115.

[236] D. Bhatia, S. Chakraborty, S. Mehtab, Y. Krishnan, *Methods Mol. Biol.* **2013**, *991*, 65–80.

[237] S. Rhee, Z. Han, K. Liu, H. T. Miles, D. R. Davies, *Biochemistry* **1999**, *38*, 16810–16815.

[238] T. M. Bryan, *Molecules* **2020**, *25*, 1–22.

[239] S. Neidle, *Curr. Opin. Struct. Biol.* **2009**, *19*, 239–250.

[240] M. O. Ebert, C. Mang, R. Krishnamurthy, A. Eschenmoser, B. Jaun, *J. Am. Chem. Soc.* **2008**, *130*, 15105–15115.

[241] K. M. M. Carneiro, N. Avakyan, H. F. Sleiman, *J. Am. Chem. Soc.* **2010**, *132*, 679–685.

[242] W. J. Chung, B. Heddi, E. Schmitt, K. W. Lim, Y. Mechulam, A. T. Phan, *Proc. Natl. Acad. Sci. U. S. A.* **2015**, *112*, 2729–2733.